THE UNIVERSITY OF CHICAGO

CATALYST-CONTROLLED, SITE-SELECTIVE HALOGENATION USING FLAVIN-DEPENDENT HALOGENASES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

MARY CATHERINE ANDORFER

CHICAGO, ILLINOIS

AUGUST 2017

TABLE OF CONTENTS

CHAPTER 1: AN INTRODUCTION TO FLAVIN-DEPENDENT HALOGENASES AND DIRECTED EVOLUTION **1**

1.1 The Importance of halogenation **1**

1.2 Flavin-dependent halogenases: mechanism and characterization **3**

1.3 Directed evolution as an approach for protein engineering **12**

1.4 Conclusions **14**

1.5 References **15**

CHAPTER 2: CHARACTERIZING THE SUBSTRATE SCOPE OF WILD-TYPE REBH AND EVOLVING REBH VARIANTS WITH INCREASED THERMOSTABILITY **19**

2.1 Introduction **19**

2.2 Results and Discussion **21**
2.2.1 Optimizing biocatalysis with RebH 21
2.2.2 Initial exploration of RebH site-selectivity on non-native substrates 24
2.2.3 Directed evolution of RebH for increased thermostability 27

2.3 Conclusions **32**

2.4 Experimental **33**
2.4.1 General experimental procedures 33
2.4.2 Specific experimental procedures 35
2.4.3 Detailed isolation and characterization I 48
2.4.4 Detailed isolation and characterization II 55
2.4.5 Additional data 58

2.5 References **62**

CHAPTER 3: DEVELOPING MUTAGENESIS AND SCREENING METHODS FOR DIRECTED EVOLUTION OF FDHS **65**

3.1 Introduction **65**

3.2 Results and Discussion **69**
3.2.1 A Simple, combinatorial codon mutagenesis method for targeted protein engineering 69
3.2.2 A Colorimetric screen for rapid analysis of FDH activity on anilines 78
3.2.3 A MALDI-MS screen for directly probing FDH site-selectivity 85

3.3 Conclusions **92**

3.4 Experimental **93**
3.4.1 General experimental procedures 93
3.4.2 Specific experimental procedures 97
3.4.3 Detailed isolation and characterization 112
3.4.4 Additional data 117

3.5 References **125**

CHAPTER 4: ALTERING THE SITE SELECTIVITY OF REBH THROUGH DIRECTED EVOLUTION **128**

4.1 Introduction **128**

4.2 Results and Discussion **130**
4.2.1 Using a MALDI MS screen to alter the site selectivity of RebH 130
4.2.2 Characterization of variants: site selectivity, kinetics, substrate scope, and docking simulations 143
4.2.3 Characterization of variants: X-ray crystallography, mutations reversions, and molecular dynamics 148

4.3 Conclusions **158**

4.4 Experimental **158**
4.4.1 General experimental procedures 158
4.4.2 Specific experimental procedures 168
4.4.3 Detailed isolation and characterization 174
4.4.4 Additional data 179

4.5 References **183**

CHAPTER 5: UNDERSTANDING FLAVIN-DEPENDENT HALOGENASE REACTIVITY VIA SUBSTRATE ACTIVITY PROFILING **187**

5.1 Introduction **187**

5.2 Results and Discussion **189**
5.2.1 Constructing FDH substrate activity profiles 189
5.2.2 Selectivity of FDH halogenation 194
5.2.3 The Effect of substrate binding and substrate electronics on FDH halogenation 199

5.3 Conclusions **203**

5.4 Experimental **204**
5.4.1 General experimental procedures 204
5.4.2 Specific experimental procedures 209
5.4.3 Detailed isolation and characterization 212
5.4.4 Additional data 232

5.5 References **255**

CHAPTER 6: DISCOVERING NEW FLAVIN-DEPENDENT HALOGENASES THROUGH GENOME MINING **258**

6.1 Introduction **258**

6.2 Results and Discussion **259**
6.2.1 Halogenase sequence similarity network and analysis 259
6.2.2 Initial results from halogenase library 261

6.3 Conclusions **265**

6.4 Experimental **265**
6.4.1 General experimental procedures 265
6.4.2 Specific experimental procedures 266

6.5 References **270**

APPENDIX I **271**

NMR Spectra for Compounds from Chapter 2

APPENDIX II **308**

NMR Spectra for Compounds from Chapter 3

APPENDIX III **321**

NMR Spectra for Compounds from Chapter 4

APPENDIX IV **331**

NMR Spectra for Compounds from Chapter 5

LIST OF FIGURES

CHAPTER 1

Figure 1.1: Structures of halogenated compounds and corresponding bioactivities 2
Figure 1.2: Overview of FAD(C4a)-OOH formation and hydroxylation of hydroxybenzoate by a flavin-dependent monooxygenase 4
Figure 1.3: General scheme for FDH-catalyzed halogenation. Mechanism for generation of proposed halenium for EAS within FDHs 6
Figure 1.4: Substrate binding in RebH and PyrH active sites is consistent with observed 7- and 5-chlorination, respectively 8
Figure 1.5: FDH-catalyzed halogenations 9
Figure 1.6: Comparison of chemocatalytic vs. biocatalytic routes to obtain the chiral amine in sitagliptin phosphate 14

CHAPTER 2

Figure 2.1: Cofactor regeneration system used in RebH bioconversions 23
Figure 2.2: Yields and selectivities for preparative RebH catalyzed halogenation reactions 26
Figure 2.3: Halogenation conversion of RebH variants after heat treatment 28
Figure 2.4: Thermal denaturation curves and conversion-temperature profiles for RebH variants 29
Figure 2.5: FDH halogenation is affected by salt concentration and reaction shaking 30
Figure 2.6: Time courses for RebH and 3LSR for 2-methyltryptamine at 40 °C 32
Figure 2.7: Hanes-Woolf plot for RebH conversion of L-tryptophan 58
Figure 2.8: Hanes-Woolf plot for RebH conversion of tryptamine 58
Figure 2.9: Hanes-Woolf plot for RebH conversion of 2-aminonaphthalene 59
Figure 2.10: Hanes-Woolf plot for RebH conversion of tryptoline 59
Figure 2.11: Hanes-Woolf plot for RebH conversion of 2-methyltryptamine at 40 °C 60
Figure 2.12: Hanes-Woolf plot for 3LSR conversion of 2-methyltryptamine at 40 °C 60
Figure 2.13: Hanes-Woolf plot for RebH conversion of 2-methyltryptamine at 21 °C 61
Figure 2.14: Hanes-Woolf plot for 3LSR conversion of 2-methyltryptamine at 21 °C 61

CHAPTER 3

Figure 3.1: Overview of directed evolution and common mutagenesis methods 66
Figure 3.2: Overview of CCM method 70
Figure 3.3: Mutation frequency per gene with different numbers of PCR fragmentation cycles and frequency of mutation at targeted residues 71
Figure 3.4: Mutation frequency per gene with different numbers of rounds of CCM and different numbers of PCR fragmentation cycles for POP 73
Figure 3.5: Frequency of mutation at targeted residues for a representative POP library 73
Figure 3.6: Overview of a BM3 library constructed using CCM 74

Figure 3.7: Frequency of mutation at targeted residues for a RebH library targeting the FAD-binding pocket 76
Figure 3.8: Summary of mutation locations for BM3, POP, and RebH libraries, mutational frequency for each library and kinetic characterization of hits 77
Figure 3.9: Hits from POP-E7 deconvolution library 78
Figure 3.10: Overview of directed evolution approach to generate FDHs capable of halogenating less activated sites 80
Figure 3.11: Colorimetric assay used to measure conversion of anthranilic acid with purified RebH variant 0K 82
Figure 3.12: Calibration curves using the colorimetric assay in cell lysate, with different amounts of added lysate (25 and 50 μL) 83
Figure 3.13: Comparison between colorimetric HTS and HPLC in cell lysate 83
Figure 3.14: Initial rates for chlorination of d_5-L-trp and L-trp 87
Figure 3.15: Chlorination competition reactions for d_5-L-trp and L-trp 87
Figure 3.16: Overview of mass spectrometry assay for FDH selectivity using Probe 1 88
Figure 3.17: RebH time course analyzed by MALDI MS and UPLC 91
Figure 3.18: Distinguishing Probe 1 from tryptamine by MALDI MS 92
Figure 3.19: Screening using HME 117
Figure 3.20: Screening using BME 118
Figure 3.21: Activity of purified protein variants with HME 118
Figure 3.22: Activity of purified protein variants with BME 119
Figure 3.23: Saturation plot for 1DF conversion of BME 119
Figure 3.24: Hanes-Woolf plot for 1DF conversion of BME 120
Figure 3.25: Saturation plot for BM3 conversion of BME 120
Figure 3.26: Hanes-Woolf plot for BM3 conversion of BME 121
Figure 3.27: Saturation plot for 1K conversion of anthranilic acid 121
Figure 3.28: Hanes-Woolf plot for 1K conversion of anthranilic acid 122
Figure 3.29: Saturation plot for 0K conversion of anthranilic acid 122
Figure 3.30: Hanes-Woolf plot for 0K conversion of anthranilic acid 123
Figure 3.31: Saturation plot for 0K conversion of 5-methyl-anthranilic acid 123
Figure 3.32: Hanes-Woolf plot for 0K conversion of 5-methyl-anthranilic acid 124

CHAPTER 4

Figure 4.1: Selective installation of functional groups (FG) on indoles *via* C–H bond activation using different catalyst directing groups (DG) or enzyme catalysis 128
Figure 4.2: Total conversion of tryptamine to chlorinated product using potential RebH starting points for directed evolution (Round 0) 131
Figure 4.3: Total conversion of tryptamine to chlorinated product using variant 1P (parent) and variants from round 2 132
Figure 4.4: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variants from round 0-3 133
Figure 4.5: Total conversion of tryptamine to chlorinated product and selectivity for halogenation at C5/C6 using variant 3W (parent) and variants from round 4 134
Figure 4.6: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variant 5LS (parent) and variants from round 6 136

Figure 4.7: General overview of evolutionary lineage for rounds 6-10 of mutagenesis and screening 138
Figure 4.8: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variant 6TL (parent) and variants from round 7 139
Figure 4.9: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variant variants from round 7 and 8 140
Figure 4.10: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variant 7M and 8F (7M+L111F) 141
Figure 4.11: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variants 3W-10S 142
Figure 4.12: Lineage diagram summarizing mutagenesis methods and mutations found in selected variants above and below lineage arrows, respectively 142
Figure 4.13: Chlorination of tryptamine using engineered halogenases 144
Figure 4.14: Key residues in the RebH and PyrH active sites. Interactions involved in L-tryptophan binding in RebH and PyrH 147
Figure 4.15: Overview of docking studies with 0S, 10S, and 8F computationally minimized structures 148
Figure 4.16: Active site of 0S crystal structure with tryptamine bound overlaid with wild-type RebH with bound L-tryptophan 150
Figure 4.17: Conversion of tryptamine to 7-, 6-, and 5-chlorotyrptamines for 10S point mutants 152
Figure 4.18: Active site of 10S crystal structure with docked tryptamine overlaid with wild-type RebH with bound L-tryptophan 154
Figure 4.19: Conversion of tryptamine to 7-, 6-, and 5-chlorotyrptamines for 8F point mutants 156
Figure 4.20: Location of 8F mutations that affect selectivity and activity and active site of 8F crystal structure overlaid with wild-type RebH with bound L-tryptophan 157
Figure 4.21: Overview of LCMS method to analyze site selectivity of chlorination 171
Figure 4.22: Hanes-Woolf plot for 0S conversion of tryptamine 179
Figure 4.23: Saturation curve for 0S conversion of tryptamine 180
Figure 4.24: Hanes-Woolf plot for 10S conversion of tryptamine 180
Figure 4.25: Saturation curve for 10S conversion of tryptamine 181
Figure 4.26: Hanes-Woolf plot for 8F conversion of tryptamine 181
Figure 4.27: Saturation curve for 8F conversion of tryptamine 182

CHAPTER 5

Figure 5.1: Previously characterized substrate scope of Rdc2 and GsfI 187
Figure 5.2: Phylogenetic tree of FDHs examined 189
Figure 5.3: Substrate activity profiles in heat map form for eight FDHs on substrates in panels 1 (indoles, pyrroles, azoles, anilines, anildes) and 2 (phenols) 192
Figure 5.4: Substrate activity profiles and structures of selected substrates 193
Figure 5.5: Conversion data from initial substrate activity profiles and structures for compounds in panel 2 194
Figure 5.6: Site selectivity of RebH and Rdc2 on representative substrates 195

Figure 5.7: Conversion vs. halenium affinity (HalA) for all sp^2 carbons selected substrates in panels 1 and 2 200
Figure 5.8: Docking poses for selected substrates observed for ROCS pharmacophore overlay 203

CHAPTER 6

Figure 6.1: Sequence similarity network for putative flavin-dependent halogenases 260
Figure 6.2: Overview of putative FDH sequence selection method 261
Figure 6.3: LCMS method using in initial assessment of putative FDH library 262
Figure 6.4: Homology model of FDH #65 aligned with RebH crystal structure 263
Figure 6.5: FDH #65 demonstrates higher bromination activity than chlorination 264

APPENDIX I

NMR Spectra for Compounds from Chapter 2

Figure AI.1: ^{1}HNMR spectrum of 1. 271
Figure AI.2: ^{13}CNMR spectrum of 1. 272
Figure AI.3: NOSEY spectrum of 1. 273
Figure AI.4: ^{1}HNMR spectrum of 2. 274
Figure AI.5: ^{13}CNMR spectrum of 2. 275
Figure AI.6: COSY spectrum of 2. 276
Figure AI.7: HMQC spectrum of 2. 277
Figure AI.8: ^{1}HNMR spectrum of 3. 278
Figure AI.9: ^{13}CNMR spectrum of 3. 279
Figure AI.10: NOSEY spectrum of 3. 280
Figure AI.11: ^{1}HNMR spectrum of 4. 281
Figure AI.12: ^{13}CNMR spectrum of 4. 282
Figure AI.13: NOSEY spectrum of 4. 283
Figure AI.14: ^{1}HNMR spectrum of 5. 284
Figure AI.15: ^{13}CNMR spectrum of 5. 285
Figure AI.16: NOSEY spectrum of 5. 286
Figure AI.17: ^{1}HNMR spectrum of 6. 287
Figure AI.18: ^{13}CNMR spectrum of 6. 288
Figure AI.19: NOSEY spectrum of 6. 289
Figure AI.20: NOSEY spectrum of 6 (zoomed). 290
Figure AI.21: ^{1}HNMR spectrum of 7. 291
Figure AI.22: ^{13}CNMR spectrum of 7. 292
Figure AI.23: NOSEY spectrum of 7. 293
Figure AI.24: HMQC spectrum of 7. 294
Figure AI.25: HMBC spectrum of 7. 295
Figure AI.26: ^{1}HNMR spectrum of 8a and 8b. 296
Figure AI.27: ^{13}CNMR spectrum of 8a and 8b. 297
Figure AI.28: ^{1}HNMR spectrum of 9. 298

Figure AI.29: ^{13}CNMR spectrum of 9. 299
Figure AI.30: ^{1}HNMR spectrum of 10. 300
Figure AI.31: ^{13}CNMR spectrum of 10. 301
Figure AI.32: NOSEY spectrum of 10. 302
Figure AI.33: ^{1}HNMR spectrum of 11. 303
Figure AI.34: ^{13}CNMR spectrum of 11. 304
Figure AI.35: ^{1}HNMR spectrum of 12. 305
Figure AI.36: ^{13}CNMR spectrum of 12. 306
Figure AI.37: NOSEY spectrum of 12. 307

APPENDIX II

NMR Spectra for Compounds from Chapter 3
Figure AII.1: ^{1}HNMR spectrum of 1a. 308
Figure AII.2: ^{1}HNMR spectrum of 1b. 309
Figure AII.3: ^{13}CNMR spectrum of 1b. 310
Figure AII.4: ^{1}HNMR spectrum of 1c. 311
Figure AII.5: ^{13}CNMR spectrum of 1c. 312
Figure AII.6: ^{1}HNMR spectrum of Probe 1. 313
Figure AII.7: ^{13}CNMR spectrum of Probe 1. 314
Figure AII.8: ^{1}HNMR spectrum of 2a. 315
Figure AII.9: ^{13}CNMR spectrum of 2a. 316
Figure AII.10: ^{1}HNMR spectrum of 2b. 317
Figure AII.11: ^{13}CNMR spectrum of 2b. 318
Figure AII.12: ^{1}HNMR spectrum of Probe 2. 319
Figure AII.13: ^{13}CNMR spectrum of Probe 2. 320

APPENDIX III

NMR Spectra for Compounds from Chapter 4
Figure AIII.1: ^{1}HNMR spectrum of 3. 321
Figure AIII.2: ^{1}HNMR spectrum of 5. 322
Figure AIII.3: ^{1}HNMR spectrum of 6. 323
Figure AIII.4: ^{1}HNMR spectrum of 7. 324
Figure AIII.5: ^{1}HNMR spectrum of 7 and 8. 325
Figure AIII.6: ^{1}HNMR spectrum of 9. 326
Figure AIII.7: ^{1}HNMR spectrum of 10. 327
Figure AIII.8: ^{1}HNMR spectrum of 11. 328
Figure AIII.9: ^{1}HNMR spectrum of 12. 329
Figure AIII.10: ^{1}HNMR spectrum of 13. 330

APPENDIX IV

NMR Spectra for Compounds from Chapter 5

Figure AIV.1: ^{1}HNMR spectrum of 29. 331
Figure AIV.2: ^{1}HNMR spectrum of 29a. 332
Figure AIV.3: NOSEY spectrum of 29a. 333
Figure AIV.4: ^{1}HNMR spectrum of 30. 334
Figure AIV.5: ^{1}HNMR spectrum of 30a. 335
Figure AIV.6: ^{1}HNMR spectrum of 31. 336
Figure AIV.7: ^{1}HNMR spectrum of 31a. 337
Figure AIV.8: NOSEY spectrum of 31a. 338
Figure AIV.9: ^{1}HNMR spectrum of 32. 339
Figure AIV.10: ^{1}HNMR spectrum of 32a. 340
Figure AIV.11: NOSEY spectrum of 32a. 341
Figure AIV.12: ^{1}HNMR spectrum of 33. 342
Figure AIV.13: ^{1}HNMR spectrum of 33a. 343
Figure AIV.14: NOSEY spectrum of 33a. 344
Figure AIV.15: ^{1}HNMR spectrum of 33b. 345
Figure AIV.16: ^{1}HNMR spectrum of 34. 346
Figure AIV.17: ^{1}HNMR spectrum of 34a. 347
Figure AIV.18: NOSEY spectrum of 34a. 348
Figure AIV.19: ^{1}HNMR spectrum of 35. 349
Figure AIV.20: ^{1}HNMR spectrum of 35a. 350
Figure AIV.21: NOSEY spectrum of 35a. 351
Figure AIV.22: Zoomed NOSEY spectrum of 35a. 352
Figure AIV.23: ^{1}HNMR spectrum of 36. 353
Figure AIV.24: ^{1}HNMR spectrum of 36a. 354
Figure AIV.25: ^{1}HNMR spectrum of 37. 355
Figure AIV.26: NOSEY spectrum of 37. 356
Figure AIV.27: ^{1}HNMR spectrum of 37a. 357
Figure AIV.28: ^{1}HNMR spectrum of 37b. 358
Figure AIV.29: ^{1}HNMR spectrum of 38. 359
Figure AIV.30: ^{1}HNMR spectrum of 38a. 360
Figure AIV.31: ^{1}HNMR spectrum of 39. 361
Figure AIV.32: ^{1}HNMR spectrum of 39a. 362
Figure AIV.33: ^{1}HNMR spectrum of 40. 363
Figure AIV.34: ^{1}HNMR spectrum of 40a. 364
Figure AIV.35: NOSEY spectrum of 40a. 365
Figure AIV.36: ^{1}HNMR spectrum of 41. 366
Figure AIV.37: ^{1}HNMR spectrum of 41a. 367
Figure AIV.38: ^{1}HNMR spectrum of 42. 368
Figure AIV.39: ^{1}HNMR spectrum of 42a. 369
Figure AIV.40: NOSEY spectrum of 42a. 370
Figure AIV.41: ^{1}HNMR spectrum of 43. 371
Figure AIV.42: ^{1}HNMR spectrum of 43a. 372
Figure AIV.43: NOSEY spectrum of 43a. 373
Figure AIV.44: ^{1}HNMR spectrum of 44. 374
Figure AIV.45: ^{1}HNMR spectrum of 44a. 375

Figure AIV.46: ^{1}HNMR spectrum of 44b. 376
Figure AIV.47: NOSEY spectrum of 44b. 377
Figure AIV.48: ^{1}HNMR spectrum of 45. 378
Figure AIV.49: ^{1}HNMR spectrum of 45a. 379
Figure AIV.50: NOSEY spectrum of 45a. 380
Figure AIV.51: ^{1}HNMR spectrum of 46. 381
Figure AIV.52: ^{1}HNMR spectrum of 46a. 382
Figure AIV.53: NOSEY spectrum of 46a. 383
Figure AIV.54: ^{1}HNMR spectrum of 46b. 384
Figure AIV.55: ^{1}HNMR spectrum of 48. 385
Figure AIV.56: ^{1}HNMR spectrum of 48a. 386
Figure AIV.57: NOSEY spectrum of 48a. 387
Figure AIV.58: ^{1}HNMR spectrum of 52. 388
Figure AIV.59: ^{1}HNMR spectrum of 52a. 389
Figure AIV.60: NOSEY spectrum of 52a. 390
Figure AIV.61: ^{1}HNMR spectrum of 62. 391
Figure AIV.62: ^{1}HNMR spectrum of 62a. 392
Figure AIV.63: NOSEY spectrum of 62a. 393
Figure AIV.64: ^{1}HNMR spectrum of 64. 394
Figure AIV.65: ^{1}HNMR spectrum of 64a. 395
Figure AIV.66: ^{1}HNMR spectrum of 67. 396
Figure AIV.67: NOSEY spectrum of 67. 397
Figure AIV.68: ^{1}HNMR spectrum of 67a. 398
Figure AIV.69: ^{1}HNMR spectrum of 67b. 399
Figure AIV.70: ^{1}HNMR spectrum of 77. 400
Figure AIV.71: ^{1}HNMR spectrum of 77a. 401
Figure AIV.72: NOSEY spectrum of 77a. 402
Figure AIV.73: ^{1}HNMR spectrum of 82. 403
Figure AIV.74: ^{1}HNMR spectrum of 82a. 404
Figure AIV.75: NOSEY spectrum of 82a. 405
Figure AIV.76: ^{1}HNMR spectrum of 23a and 23b. 406
Figure AIV.77: ^{1}HNMR spectrum of 23c. 407
Figure AIV.78: ^{1}HNMR spectrum of 24a. 408
Figure AIV.79: ^{1}HNMR spectrum of 25a. 409
Figure AIV.80: ^{13}CNMR spectrum of 25a. 410
Figure AIV.81: ^{1}H-^{13}C HSQC spectrum of 25a. 411
Figure AIV.82: ^{1}H-^{13}C HMBC spectrum of 25a. 412
Figure AIV.83: ^{1}HNMR spectrum of 28a. 413

LIST OF SCHEMES

CHAPTER 1

Scheme 1.1: Chlorination of indole occurs at the more electronically activated pyrrolo positions 3

CHAPTER 2

Scheme 2.1: RebH or PrnA catalyzed 7-chlorination of L-tryptophan 20
Scheme 2.2: General scheme for RebH catalyzed halogenation 25
Scheme 2.3: General scheme for RebH catalyzed halogenation and substrates used to explore scope 31

CHAPTER 3

Scheme 3.1: General scheme for high-throughput colorimetric assay 79
Scheme 3.2: Deuterated arenes as substrates and mass spectrometry as a screen to directly analyze site selectivity of chlorination reactions 86
Scheme 3.3: Synthesis of Probe 1 89
Scheme 3.4: Synthesis of Probe 2 90

CHAPTER 4

Scheme 4.1: Potential scheme for 2-oxytryptamine formation using FDH variants 137

CHAPTER 5

Scheme 5.1: General scheme for analytical FDH bioconversions used to construct substrate activity profiles 190

LIST OF TABLES

CHAPTER 2

Table 2.1: Yields of RebH and RebF under different conditions 23
Table 2.2: Catalytic parameters for RebH halogenation of select substrates 27
Table 2.3: Representative yields for preparative 3L(S)R catalyzed[a] halogenation reactions and comparison to wild-type RebH reactions 31
Table 2.4: Catalytic parameters for halogenation of 2-methyltryptamine by RebH and 3LSR at 21 °C and 40 °C 32
Table 2.5: Table of primers for section 2.2.1 36

CHAPTER 3

Table 3.1: Catalytic parameters for anthranilic acid analogs 80

CHAPTER 4

Table 4.1: Kinetic parameters for RebH, 0S, 8F, and 10S 145
Table 4.2: Conversion and selectivity for halogenation of different substrates using 8F and 10S 146
Table 4.3: Summary of data collection and refinement statistics for crystal structures of RebH variants 0S, 8F and 10S 149
Table 4.4: Chlorinated product distribution for 10S point mutants 152
Table 4.5: Chlorinated product distribution for 8F point mutants 156

CHAPTER 5

Table 5.1: Relative product distributions for halogenation of representative substrates using FDHs and NCS 197
Table 5.2: Initial conversion data from panel 1 and 2 232
Table 5.3: Isolated products from bioconversions 243

ACKNOWLEDGEMENTS

First, I would like to thank those who have contributed to the research in this dissertation. My advisor, Prof. Jared Lewis, has influenced me significantly throughout my time in graduate school and has helped me to grow as a scientist and mentor. I thank everyone in the Lewis lab, both past and present, as well as those I interacted with at Novartis Institutes for Biomedical Research, specifically Dr. James Payne, Dr. Catherine Poor, Jaylie Vergara-Coll, Anna Girlich, Julia Chael, Dr. Ketaki Belsare, Dr. Landon Durak, Dr. Ken Ellis-Guardiola, Dr. Hyun June Park, Yifan Gu, Alan Swartz, Dr. Brian Fisher, Natalie Chan, David Upp, Paul Butkovich, Harrison Snodgrass, Atreyi Bhattacharya, Christian Gomez, Dr. Kian Tan, Jonathan Grob, Dr. Piro Siuti, Jeremiah Lilly, and Dr. Christine Hajdin. I am especially grateful to James Payne, who served as an excellent mentor, coworker, and friend.

Additionally, I would like to thank my friends at the University of Chicago who have helped and supported me throughout the years, including Judith Kamm, Dr. Ketaki Belsare, Dr. Ken Ellis-Guardiola, Dr. Charles William Heaps III, Clare Heaps, Sam Tazzia, Alan Swartz, Dennis Tam, Dr. Landon Durak, Dr. James Payne, Dr. Wayne Lau, Hunter Vibbert, Joe Gair, Jeff Montgomery, and Erica Sturm. This is by no means an exhaustive list, as this community has enabled many relationships, both personal and professional.

I would also like to thank my committee members Prof. Joseph Piccirilli and Prof. Chuan He for their helpful insight and guidance. Prof. Piccirilli and Prof. He have played crucial roles in my career development through the CBI and for this I am most grateful. I thank Dr. Antoni Jurkiewicz and Dr. Jin Qin for helping with NMR and mass spectrometry equipment, without which this research would not have been possible.

Lastly, I would like to thank my friends and family outside of the UChicago community. My parents, Tony and Marie Andorfer, have loved and supported me no matter what, and for this I am immensely grateful. I thank my sister, Kristy Andorfer, whom I can always count on to make holidays way too much fun. In addition, I am grateful to my grandparents - Joan Teders, Duane Teders, Bill Andorfer, and Judi Andorfer - as well as my aunt and uncle - Dan and Becky Herr. I thank Laurel Heckman, who has been like a sister to me since our first year in college. She has been crucial in my development as a person and scientist, and I look forward to continuing our careers together. I would also like to thank my amazing friends from high school and college – Jean Longsworth, Courtney Christensen, Megan Welling, Dr. Kristin Crider, Kelly Crider, Kelly Tippmannn, Rebecca Pacchiano, Stephanie Densberger, Krista Jensen, and Annie Mullowney. I am also grateful for Alicia Drier and Jamie Burier who are ridding me of horrifyingly long grad school hair. Finally, I would like to thank Joe Gair, who has been a constant source of support, both personal and professional.

ABSTRACT

This dissertation describes the use and engineering of flavin-dependent halogenases (FDHs) as mild, selective halogenation catalysts. Unlike conventional aromatic halogenation that relies on substrate electronic properties to govern site selectivity, FDHs use catalyst control to halogenate at relatively less electronically activated sites. Moreover, FDH bioconversions can be conducted using water as solvent, NaCl or NaBr as a halide source, and oxygen as the terminal oxidant – making them a milder, greener alternative. Prior to the work in this dissertation, structural and mechanistic work had provided insight into this class of enzymes; however, their use in biocatalysis was largely limited to analytical-scale reactions on native substrates. Furthermore, protein engineering efforts using FDHs was minimal. Herein, the first directed evolution of an FDH for altered site selectivity is described.

Chapter 1 provides a background of aromatic halogenation, FDHs, and directed evolution. Although critical in the development of pharmaceutical and agrochemicals, aromatic halogenation methods often suffer from harsh reaction conditions and poor site selectivity. Tryptophan-FDHs (Trp-FDHs) overcome limitations of substrate-controlled selectivity by using catalyst control. Thus, Trp-FDHs are able to halogenate the benzo positions of L-tryptophan in the presence of more electronically activated pyrrolo position. The mechanism by which this selectivity is conferred is thought to arise from positioning substrate within the active site proximal to a reactive halenium species. Although much structural and mechanistic work has developed our understanding of this class of enzyme, the limited success of protein engineering efforts using targeted point mutations reveals the limitations to this understanding. Directed evolution is a powerful approach to protein engineering which does not require previous knowledge of enzyme structure or mechanism.

Chapter 2 discusses the development of robust protocols for the use of 7-Trp-FDH RebH in biocatalysis. Bioconversions were conducted using these protocols to establish site selectivity of halogenation using RebH on a variety of aromatic substrates. In addition to the native substrate L-tryptophan, many other indole derivatives were chlorinated by RebH at benzo positions. Subsequently, directed evolution was used to increase the thermostability of RebH. Three rounds of mutagenesis and screening produced variants with melting temperatures up to 18 °C higher than that of RebH. After further characterization, these variants were found to have increased catalyst lifetimes, making them valuable biocatalysts.

Chapter 3 describes the development of mutagenesis and screening methods that can be applied to FDH directed evolution efforts. A new targeted, combinatorial codon mutagenesis method is developed and used to generate high quality libraries for several enzyme targets. In addition, a colorimetric screen used to rapidly detect increased chlorination is described and used in an evolution campaign. Lastly, a MALDI MS screen using deuterium-substituted probes was developed as a means of directly observing changes in site selectivity of FDH halogenation in a high-throughput manner.

The MALDI MS screen detailed in Chapter 3 was used in an evolution campaign to alter the site selectivity of RebH on the substrate tryptamine. This campaign, which led to RebH variants capable of chlorinating substituted indoles ortho-, meta-, and para- to the indole nitrogen, is described in Chapter 4. The x-ray structures of these variants were elucidated and have provided insight into the mechanism by which site selectivity was altered.

Chapter 5 describes the generation and analysis of substrate activity profiles for 4 native FDHs and 4 engineered variants on 93 low molecular weight aromatic compounds. These profiles provided fundamental insights into how substrate class, functional group substitution,

electronic activation, and binding impact FDH activity and selectivity. Together, these data, provide a useful framework for understanding and exploiting FDH reactivity in organic synthesis.

Chapter 6 briefly introduces an ongoing project that aims to discover new, natural FDHs through a genome mining approach. Selection of putative FDH genes, screening methods, and initial results are described within this chapter.

Chapter 1: An Introduction to flavin-dependent halogenases and directed evolution

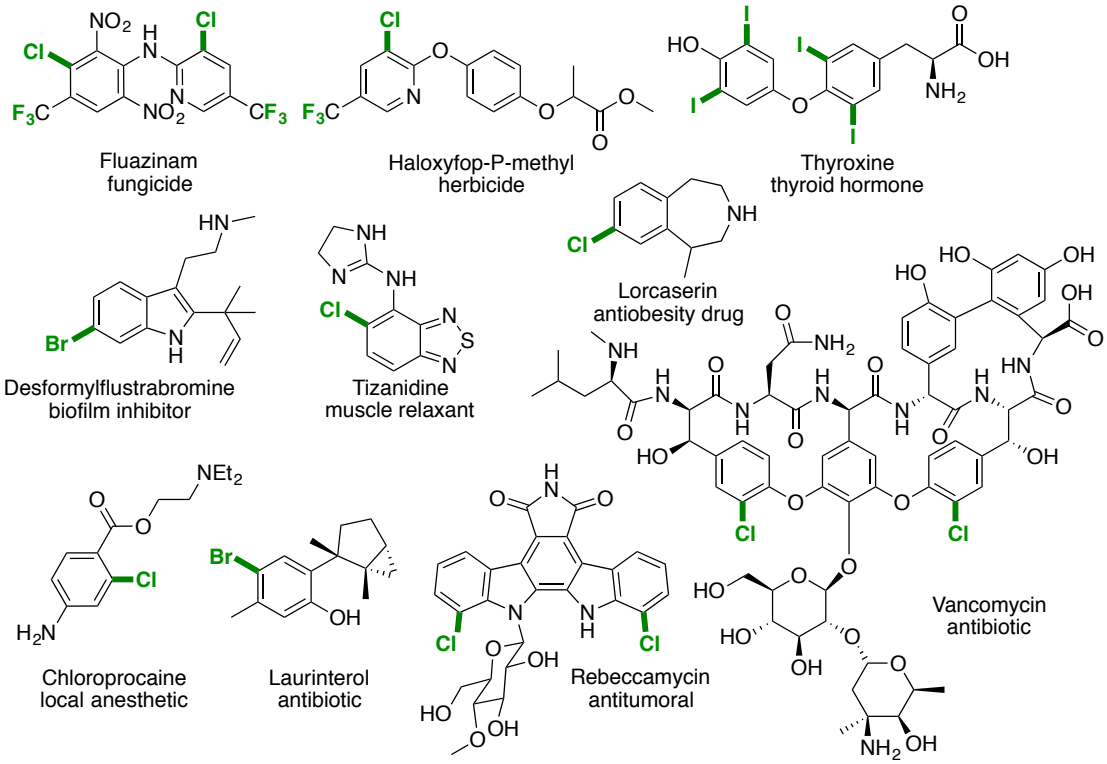
1.1 The Importance of halogenation

Over 4,500 halogenated natural products, many of which exhibit desirable functionality, have been reported to date.[1] This functionality has led to widespread use of halogenated compounds as pharmaceuticals, including the antibiotic vancomycin and the antitumoral rebeccamycin (Fig. 1.1).[2] Compounds such as laurinterol and isolaurinterol have even demonstrated potent antibacterial activity against penicillin- and vancomycin-resistant antibacterial strains, which have become increasingly problematic as bacteria evolve to withstand commonly administered drugs.[3] Additionally, halogen substitution has been shown to raise the potency of insecticide, herbicide, and fungicide activity, making it invaluable in agrochemical design.[4] For this reason, 78.5% of newly approved agrochemicals from 1998-2008 contained a halogen substituent (Fig. 1.1).[4]

The wide extent of halogenation in biologically active compounds does not appear coincidental[5,6] – halogen substitution has often been demonstrated to directly affect biological activity.[7-10] For example, in 2012, the FDA approved an anti-obesity drug, Lorcaserin (Fig. 1.1). The desired weight-loss effects of Lorcaserin are thought to arise from its highly specific interaction with serotonin receptor $5HT_{2C}$, which was also the target of the previously prescribed weight-loss drug, fenfluramine/dexfenfluramine. Although fenfluramine/dexfenfluramine and Lorcaserin are both effective weight-loss drugs, fenfluramine/dexfenfluramine was found to have substantial off-target interactions with the $5HT_{2B}$ receptor, located in the cardiovascular system.

These undesired interactions led to serious defects in cardiac valves and pulmonary hypertension in some patients taking fenfluramine/dexfenfluramine. These side effects are not observed with Lorcaserin. Structure activity relationship studies determined that the aromatic chlorine substituent plays a significant role in selectivity for the $5HT_{2C}$ serotonin receptor over $5HT_{2B}$.[9] In another recent example, the biofilm inhibition activity of desformylflustrabromine (Fig. 1.1) is lost when its bromine substituent is removed.[10]

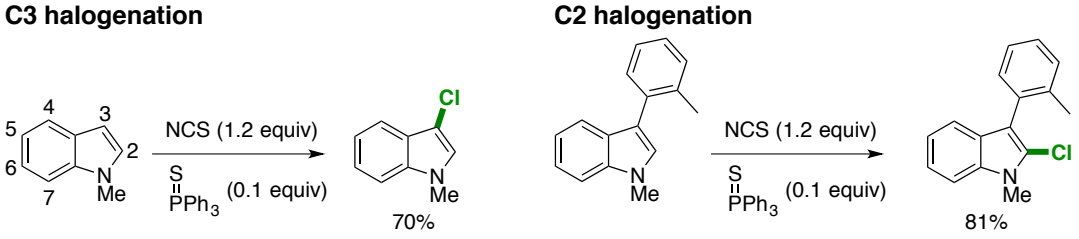
Figure 1.1: Structures of halogenated compounds and corresponding bioactivities.



Halogenated compounds not only serve as desirable target compounds, but are also invaluable building blocks in organic synthesis.[11,12] The biaryl motif commonly found in pharmaceutical compounds[13] can be accessed through cross-coupling reactions, in which aryl halides are used as starting materials.[11,12,14] Although halogenated compounds can be readily constructed using electrophilic aromatic substitution, radical chemistry, electrophilic addition,

etc., site selectivity in these reactions remains challenging. In most cases, site selectivity is determined by substrate control (e.g. substrate electronics). For example, halogenation though an electrophilic aromatic substitution mechanism usually directs halogen substitution to the most electronically activated sites on a target compound.[15] Electronic activation at multiple sites often affords product mixtures, limiting the overall yield and efficiency of the reaction.[15] Furthermore, most halogenation strategies cannot access less electronically activated sites on an aromatic compound in the presence of more activated sites. An example of this is clearly demonstrated with indole compounds. All pyrrolo- (Scheme 1.1, C2-3) and benzo-(Scheme 1.1, C4-7) positions of indole are activated for EAS; however, because the pyrrolo-sites are significantly more nucleophilic, halogenation is only observed at these sites (Scheme 1.1).[16] Halogenation catalysts capable of overriding substrate electronics and site selectively halogenating would be valuable to the chemical community.

Scheme 1.1: Chlorination of indole occurs at the more electronically activated pyrrolo positions.[a]



[a] Examples of C3 and C2 chlorination of indolyl compounds from Gustafson and coworkers.[16] Reactions were conducted at room temperature in dichloromethane and isolated yields are provided for products. Standard numbering of indoles is shown for *N*-methyl-indole.

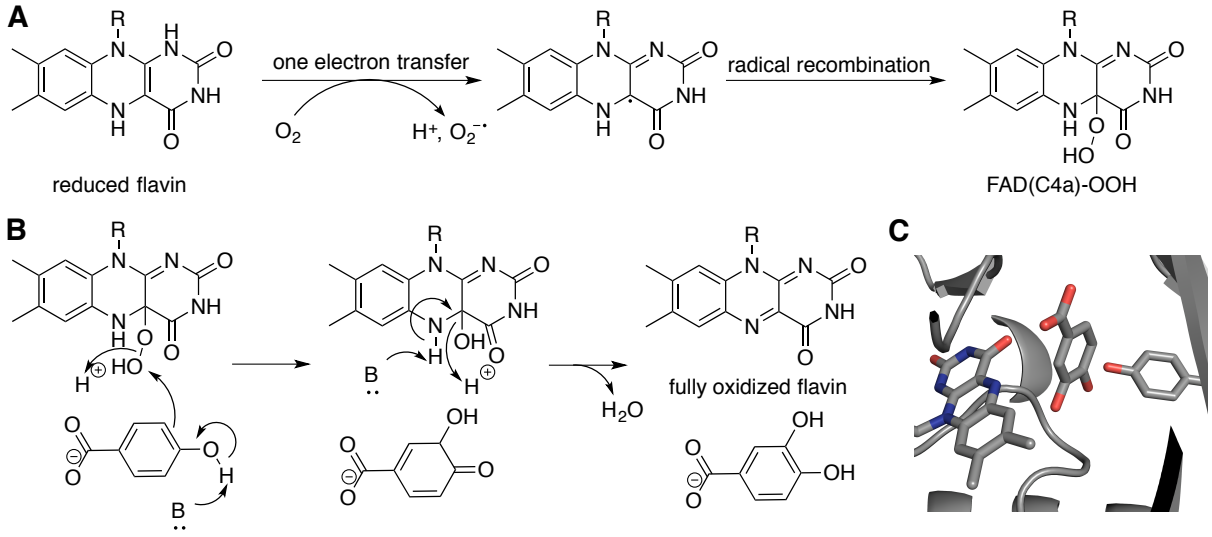
1.2 Flavin-dependent halogenases: mechanism and characterization

To accomplish halogenation in nature, enzymes capable of halogenating a broad diversity of compounds have evolved. These enzymes are broadly categorized as haloperoxidases,

nonheme iron-dependent halogenases, *S*-adenosyl-L-methionine-dependent halogenases, and flavin-dependent halogenases.[17] Although the former three categories have been the focus of much exciting work,[17-19] this thesis will focus on flavin-dependent halogenases (FDHs).

FDHs closely resemble flavin-dependent oxygenases both mechanistically and structurally. In flavin-dependent oxygenases, bound $FADH_2$ reacts with O_2 to form a FAD(C4a)-OOH intermediate (Fig. 1.2A).[20,21] This intermediate can subsequently oxidize organic substrates. For example, the flavin-dependent oxygenase *p*-hydroxybenzoate hydroxylase (PHBH) uses FAD(C4a)-OOH to hydroxylate the proximally bound aromatic substrate via electrophilic aromatic substitution (Fig. 1.2B-C).[21,22]

Figure 1.2: Overview of FAD(C4a)-OOH formation (A) and hydroxylation of hydroxybenzoate by a flavin-dependent monooxygenase (B, C[a]).

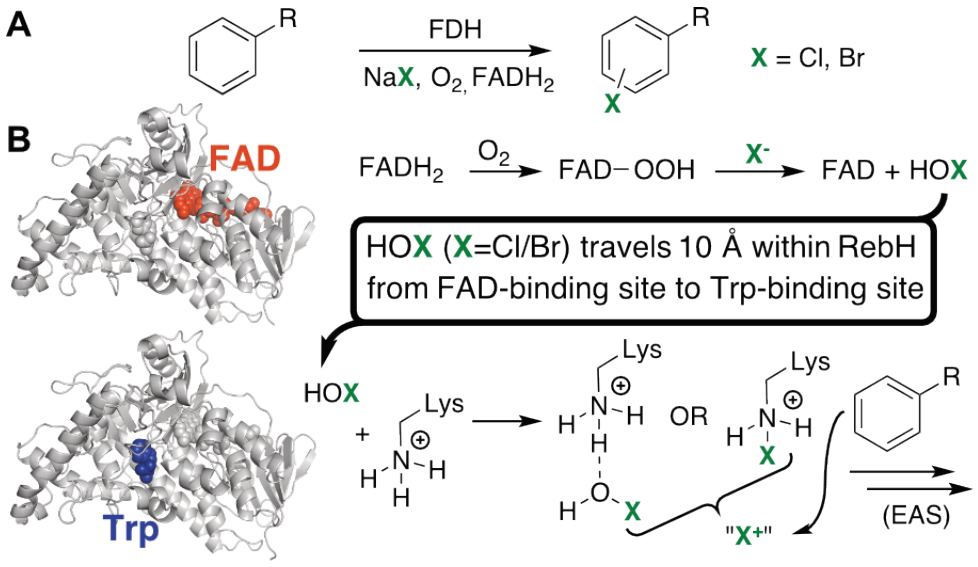


[a] X-ray crystal structure of PHBH (PDB ID: 1PHH).[23] Isoalloxazine ring of FAD, dihydroxybenzoate, and residue Y201 are shown as sticks.

Walsh and coworkers have used stopped-flow kinetics to characterize the FAD(C4a)-OOH intermediate in the 7-tryptophan FDH (Trp-FDH) RebH as well;[24] however, in contrast to

monooxygenases such as PHBH, the FAD(C4a)-OOH intermediate does not directly oxidize substrate. As shown in Fig. 1.2C, PHBH positions the C4a of the isoalloxazine ring 4.2 Å from the functionalized site of substrate, close enough for subsequent electrophilic aromatic substitution to occur directly.[23] The crystal structures of Trp-FDHs show a lengthening of this distance to 11.4 Å, indicating the FAD-peroxide cannot interact directly with substrate without large conformational changes.[25-27] No such conformational changes are observed in structures with and without substrate and/or FAD bound.[25-27] When chloride is present in FDH solutions during crystallization and/or soaked into FDH crystals, a chloride-binding site proximal to the isoalloxazine is observed.[25-27] Whether this chloride-binding site is relevant to FDH halogenation activity or simply an artifact of crystallization is still uncertain.[28] From these structural studies, it is proposed that, instead of organic substrate, chloride (perhaps that bound near the isoalloxazine ring) reacts with the FAD(C4a)-OOH intermediate to form hypochlorous acid (HOCl, Fig. 1.3). This HOCl species is thought to travel through a tunnel to the active site where it interacts with an active site lysine residue, either via simple hydrogen bonding[29] or via a chloramine intermediate.[19] The proximally bound L-tryptophan can subsequently undergo electrophilic aromatic substitution (EAS) with either the bound HOCl or chloramine species (Fig. 1.3).

Figure 1.3: (A) General scheme for FDH-catalyzed halogenation. (B) Mechanism for generation of proposed halenium for EAS within FDHs.

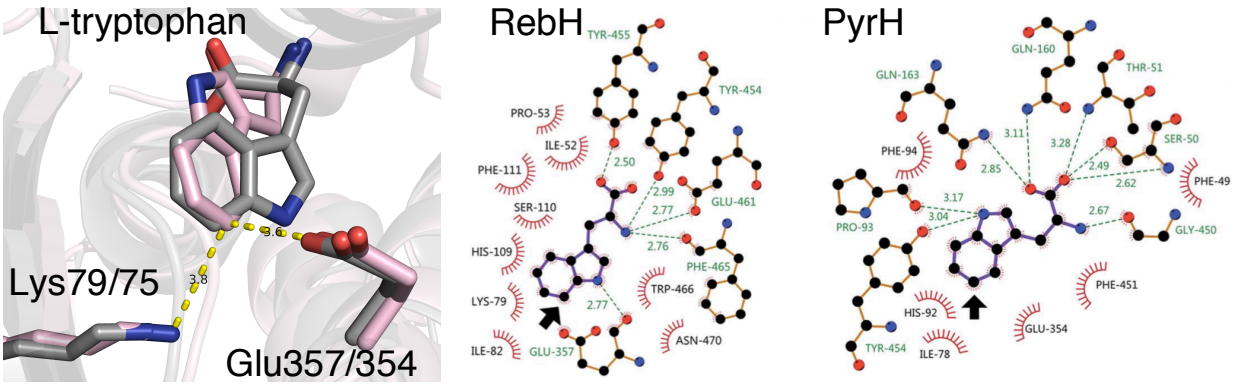


The nature of the interaction between HOCl and the active site lysine residue has been debated. Work from the Walsh and Drennan labs has demonstrated the presence of a long-lived chlorinating intermediate in RebH.[26] This intermediate was formed when RebH, O_2, reduced flavin, and chloride react in the absence of L-tryptophan. Flavin was subsequently removed through standard desalting procedures, and the resulting RebH intermediate retained chlorinating ability for up to 63 hours.[26] Lys79 was found to be crucial to the formation of this intermediate, and thus it was proposed that HOCl reacts with Lys79 to form the more stable (with respect to HOCl) chloramine species.[26] On the other hand, work from the van Peé lab suggests that HOCl may be the chlorinating species. This claim is based on the observation that active site residue Glu346 in PrnA, another 7-Trp-FDH, is crucial for high rates of catalysis. Instead of Lys-εNH-Cl serving as the chlorinating species, van Peé and coworkers propose that it could act as a stable halenium reservoir that, upon substrate binding, could regenerate HOCl. Subsequently, HOCl could interact with both Lys79 and Glu346 to increase its electrophilicity.[29] Further experiments

and, likely, simulations will be necessary to clarify the nature of the active chlorinating species in FDHs.

Whether a covalent or non-covalent interaction is occurring between HOCl and Lys79, this interaction is crucial for site selective halogenation. By anchoring the halogenating species within the enzyme active site, Trp-FDHs can halogenate a single C–H bond on L-tryptophan. In RebH and PrnA, L-tryptophan binds in the active site such that the C7–H bond is positioned proximal to the halenium species and undergoes electrophilic aromatic substitution.[26,27] In addition to 7-Trp-FDHs,[30,31] both 5-[32-34] and 6-Trp-FDHs[35-38] have been characterized. As previously noted in section 1.1, site-selectivity for a single benzo-position, especially in the presence of the more electronically activated pyrrolo-position, is remarkable. A crystal structure of the 5-Trp-FDH PyrH with bound L-tryptophan shows that the active site lysine residue is positioned proximal to C5–H bond of L-tryptophan, consistent with the observed selectivity.[39] To achieve this differential L-tryptophan binding with respect to the 7-Trp-FDHs RebH and PrnA, L-tryptophan is flipped in the active site (Fig. 1.4). Structures of 6-Trp-FDHs SttH[29] and Th-Hal[30] have also been elucidated; however, neither contains bound L-tryptophan. Thus, further structural or computational analysis will be necessary to determine the mode of substrate binding in 6-Trp-FDHs.

Figure 1.4: Substrate binding in RebH and PyrH active sites is consistent with observed 7- and 5-chlorination, respectively.[a]



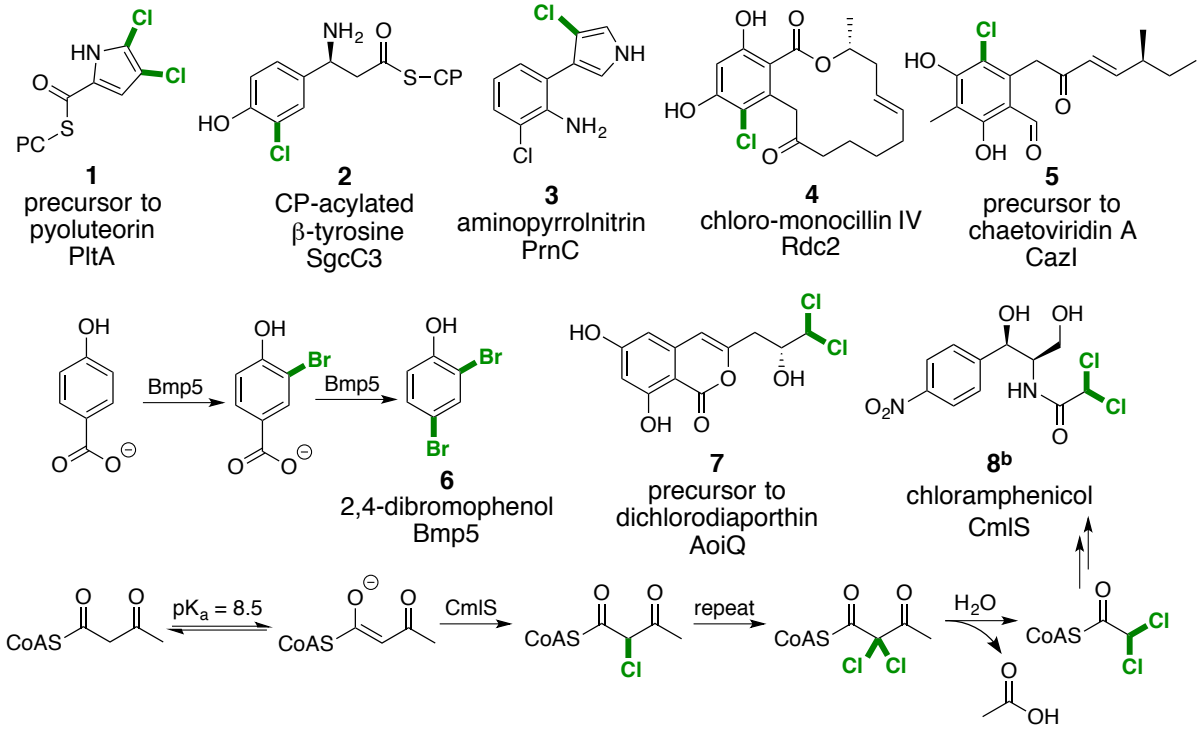
[a] Aligned X-ray crystal structures of 7-Trp-FDH RebH (PDB 2OA1, grey)[25] and 5-Trp-FDH PyrH (PDB 2WEU, pink).[39] Interactions involved in tryptophan binding in RebH and PyrH are also shown to the right. Arrow denotes chlorination site.

Much of our current understanding of the structure and mechanism of flavin-dependent halogenases has been developed through studies on Trp-FDHs; however, numerous FDHs involved in the chlorination of substrates containing phenol,[40-44] pyrrole,[45-47] and even alkyl[48,49] moieties have also been identified. Several of these FDHs have been shown to halogenate free substrate; however, many accept only substrates tethered to carrier proteins (CP). Walsh and coworkers reported the first example of such an FDH.[45] PltA was determined to be the halogenase responsible for chlorination of pyoluteorin (Fig. 1.5, **1**). *In vitro* activity of PltA was reconstituted only when the pyrrolyl substrate was tethered to a CP. Similarly, the FDH SgcC3 was found to chlorinate CP-acylated β-tyrosine (Fig. 1.5, **2**), whereas no chlorination activity was observed using the free amino acid.[50] Many more FDHs have since been reported to halogenate CP-bound substrates.

In addition to Trp-FDHs, *in vitro* FDH-catalyzed halogenation has been demonstrated with a diversity of other free substrates. Unlike PltA, the FDH PrnC halogenates the pyrrole ring of monodechloroaminopyrrolnitrin (Fig. 1.5, **3**) without the need for a CP.[47] Several FDHs,

including Rdc2,[42] CazI,[44] and Bmp5[41], have also been found to halogenate free phenol-containing compounds (Fig. 1.5, **4**-**6**). Interestingly, Bmp5 is the first and, prior to work later performed in the Lewis lab,[51] only reported single component FDH. Like single component flavin-dependent oxygenases, Bmp5 is capable of generating reduced flavin, removing the need for a separate flavin reductase.[41] Also unique to FDHs currently characterized, Bmp5 catalyzes a two-step reaction, in which bromination occurs at the 3-position of 4-hydroxybenzoate, followed by subsequent decarboxylative-bromination (Fig. 1.5, **6**).

Figure 1.5: FDH-catalyzed halogenations.[a]



[a] FDH responsible for incorporating bolded, green halogen on each compound is listed. [b] The exact substrate for CmlS is unknown. One proposed mechanism involving chlorination of acetoacetyl-CoA is shown.

Interestingly, FDHs capable of halogenating alkyl substrates have recently been described.[48,49] The FDH CmlS was identified as crucial for the production of the dichloroacetyl moiety of chloramphenicol (Fig. 1.5, **8**); however, the substrate of CmlS has not been

established.[48] It is thought that a CoA thioester form of the dichloroacetyl moiety is added to a chloramphenicol precursor by other enzymes in this cluster (Fig. 1.5),[52] but it remains unclear whether halogenation or thioester formation occurs first. The identity of the acyl group of this compound prior to halogenation is also unknown. Although acetate, acetoacetate, and malonate could serve as the acyl source, the latter two would serve as more reactive substrates, should chlorination proceed through an enolate mechanism (potential mechanism for acetoacetyl-CoA shown in Fig. 1.5).[48] Should these dicarbonyl substrates be involved instead of acetate directly, a retro-Claisen reaction (for acetoacetyl-CoA) or decarboxylation (for malonyl-CoA) could form the acyl donor dichloroacetyl-CoA. Regardless of the exact substrate chlorinated by CmlS, this FDH marks an interesting case of non-aromatic halogenation. The pKa for malonate (13.5) and acetoacetate (8.5) are such that deprotonation followed by chlorination with hypochlorous acid or chloramine are reasonable, thus CmlS could indeed follow a similar mechanism as Trp-FDHs.[48]

Remarkably, a fungal FDH has very recently been reported that appears to chlorinate an unactivated aliphatic carbon.[49] Hertweck and coworkers have established *in vitro* halogenation activity of FDH AoiQ (Fig. 1.5). For *in vitro* reactions, the deschloro precursor of **7** (orthosporin) was added as substrate to buffer containing purified AoiQ, flavin reductase, FAD, NaCl, and NADH. Formation of chlorinated products was monitored by LCMS, but product from these reactions could not be isolated due to small scale. Instead, authentic standards for the expected mono and dichlorinated products (Fig. 1.5, **7**) were synthesized and shown to coelute with reaction product peaks when analyzed by LCMS. Furthermore, when mono chlorinated orthosporin is used as substrate, higher yields of **7** are observed. Taken together, these results suggest AoiQ is halogenating an unactivated, sp^3-hybridized carbon; however, further product

characterization is required to fully verify this impressive reaction. The authors note that the mechanism of this FDH likely deviates from previously characterized FDHs (perhaps proceeding through a radical pathway), but further studies will be required to determine how this unique reactivity is achieved. In addition to substantially expanding upon the known scope of FDHs, AoiQ is also the first reported FDH to contain a functional *O*-methylation domain.

As mentioned previously, FDHs are strikingly similar to flavin-dependent monooxygenases. When the structures of CmlS, CndH[53] (an FDH thought to halogenate a phenolic substrate), PltA[54] and PrnA[27] are compared, the structurally conserved regions are also found in the oxygenase PHBH.[48] Structural analysis displays high conservation of the FAD-binding domain and halogenation active site (including the catalytically crucial lysine residue); however, the C-terminus varies between CmlS, CndH, PltA and Trp-FDHs. This region is thus thought to impart substrate specificity and site selectivity between FDHs. In the crystal structure of CndH, the C-terminus is less structured than Trp-FDHs and does not completely envelop the active site as is the case for Trp-FDHs. This observation led to the hypothesis that CndH halogenated a CP-bound substrate, and that a more unstructured C-terminus and open active site was common to this type of FDH. Based on the constricted active site in the CmlS crystal structure reported in 2010, CmlS was originally hypothesized to function on free substrate; however, the structure of PltA, reported in 2015,[54] demonstrated that this need not be the case. The C-terminal region of PltA was well ordered and constructs the putative active site. Because PltA is known to halogenate a CP-bound substrate,[45] the C-terminus must undergo a large conformational change to allow substrate access. Further studies are necessary to identify the nature of these conformational changes, as well as whether CndH and CmlS function on free or CP-bound substrate.

1.3 Directed evolution as an approach for protein engineering

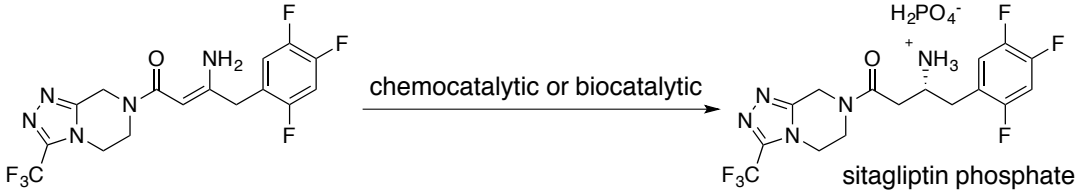
As the number of characterized FDHs continues to grow and diversify, it is increasingly apparent these biocatalysts could be useful to the chemical community as environmentally-friendly, site-selective halogenation catalysts. Though the known diversity of FDHs is impressive, challenges associated with using these catalysts in synthetic organic chemistry exist. As noted above, many FDHs do not halogenate free substrate, but instead carrier-protein bound substrate, thus limiting the feasibility of their use in synthesis. FDHs that are known to halogenate free substrate have been reported to express poorly in heterologous hosts such as *E. coli*,[30,42] which is important for large-scale production. In addition to this, low enzyme stability, slow kinetic parameters (e.g. RebH k_{cat} = 1.4 min^{-1}),[30] and limited substrate scope limit the utility and generality of FDH biocatalysis. Some progress has been made to improve these aspects of FDH halogenation through structure-guided point mutagenesis.[55,56] For example, O'Connor and coworkers were able to switch RebH specificity from L-tryptophan to tryptamine through a single point mutation.[55] Micklefield and coworkers expanded the known substrate scope of PrnA and PyrH, as well as improved efficiency of PrnA chlorination of benzoic acids.[55] Ultimately, if FDHs were capable of halogenating diverse sets of arenes on preparative scale, the ability to alter site selectivity on any given arene would be of great interest. Prior to the work described in this thesis, van Peé and coworkers demonstrated that a single point mutation in PrnA resulted in a mixture of 7- and 5-bromotryptophans (ratio of 2:1); however, a severe decrease in activity was observed.[57] Similarly, in 2014, Micklefield and coworkers used point mutagenesis of PrnA to shift the selectivity of halogenation for 2-aminobenzoic acid (from 16% to 62% selectivity para to the amine).[56] Two years later, they created a variant of SttH (6-Trp-FDH) that had altered selectivity for 3-indolepropionic acid (10% to 75% selectivity for C5).[58]

While structure-guided mutagenesis has successfully been used in certain cases to alter substrate specificity and selectivity of FDHs, directed evolution of FDHs for new and/or improved functions could significantly expand engineering efforts of this enzyme class. Through the use of iterative rounds of mutagenesis and screening, desired functionality could be obtained in a systematic manner, without relying on information regarding enzyme structure and mechanism.[59] Even with structural and mechanistic information for a target protein, beneficial mutations for a desired function can be hard to predict, especially those far away from the active site.[60] Moreover, mutations accumulated throughout a directed evolution campaign can inform our understanding of protein function.[59]

Directed evolution has proven to be a powerful and general approach to protein engineering. Using this approach, scientists have engineered enzymes for numerous new functions including, but by no means limited to, higher thermostabilities,[61] changes in substrate scope and specificity,[62] and increased catalytic efficiencies[63]. A particularly powerful example of directed evolution is illustrated by the combined efforts of Codexis and Merck to improve the industrial synthesis of sitagliptin, a drug used to treat type 2 diabetes.[63] A final step to produce sitagliptin involves an asymmetric hydrogenation of an enamine (Fig. 1.6). Previously, this step had been accomplished by using a chiral rhodium catalyst, which requires specialized high-pressure hydrogenation equipment. Through 11 rounds of mutagenesis and screening, a transaminase capable of performing the same reaction with even higher e.e. than when rhodium is used as a catalyst (99.95% e.e. and 97% e.e., respectively). Remarkably, the undesired enantiomer was never detected using evolved transaminase variants, which eliminated the need for a final recrystallization step that is required when using rhodium as a catalyst. The best transaminase variant catalyzes this transformation with an improvement of 4 orders of magnitude

compared to the initial prositagliptin transaminase, demonstrating that directed evolution can be used to dramatically increase catalytic efficiencies. Compared to the rhodium-catalyzed reaction, a 10% increase in yield is observed with a 53% increase in productivity and a 19% reduction in total waste. Moreover, using a transaminase biocatalyst for this transformation removes all need for heavy metals and specialized hydrogenation equipment, thus lowering manufacturing cost.

Figure 1.6: Comparison of chemocatalytic vs. biocatalytic routes to obtain the chiral amine in sitagliptin phosphate.[63]



Chemocatalytic route

1. Rh[Josiphos]/H_2 (250 psi) - 97% e.e.
2. carbon treatment to remove Rh
3. Recrystallization, heptane/*i*-PrOH
4. H_3PO_4

Biocatalytic route

1. transaminase, *i*-$PrNH_2$ - 99.95% e.e.
2. H_3PO_4

1.4 Conclusions

Flavin-dependent halogenases (FDHs) offer a green alternative to aromatic halogenation and are capable of achieving incredible site-selectivities. In some cases, as is observed with Trp-FDHs, sites on an arene are selectively halogenated in the presence of significantly more activated sites using catalyst control. Barriers to FDH use on a preparative scale for a wide range of substrates could be overcome through directed evolution. Moreover, the ability to tune the site selectivity of FDHs on non-native substrates could enable halogenation of sites on arenes not currently accessible by other means.

1.5 References

1. Chen, X. & van Pée, K.-H. Catalytic mechanisms, basic roles, and biotechnological and environmental significance of halogenating enzymes. *Acta Biochim Biophys Sinica* **40,** 183–193 (2008).
2. Herrera-Rodriguez, L. N., Khan, F., Robins, K. T. & Meyer, H.-P. Perspectives on biotechnological halogenation Part I: Halogenated products and enzymatic halogenation. *Chimica Oggi-Chemistry Today* **29,** 31–33 (2011).
3. Vairappan, C. S., Kawamoto, T., Miwa, H. & Suzuki, M. Potent Antibacterial Activity of Halogenated Compounds against Antibiotic-Resistant Bacteria. *Planta med* **70,** 1087–1090 (2004).
4. Jeschke, P. The unique role of halogen substituents in the design of modern agrochemicals. *Pest. Manag. Sci.* **66,** 10–27 (2009).
5. Wilcken, R., Zimmermann, M. O., Lange, A., Joerger, A. C. & Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **56,** 1363–1388 (2013).
6. Auffinger, P., Hays, F. A., Westhof, E., Ho, P. S. & Van Holde, K. E. Halogen Bonds in Biological Molecules. *Proc. Natl. Acad. Sci.* **101,** 16789–16794 (2004).
7. Harris, C. M., Kannan, R. & Kopecka, H. The role of the chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono-and didechlorovancomycin. *J. Am. Chem. Soc.* **107,** 6652–6658 (1985).
8. Groll, M., Huber, R. & Potts, B. C. M. Crystal Structures of Salinosporamide A (NPI-0052) and B (NPI-0047) in Complex with the 20S Proteasome Reveal Important Consequences of β-Lactone Ring Opening and a Mechanism for Irreversible Binding. *J. Am. Chem. Soc.* **128,** 5136–5141 (2006).
9. Smith, B. M. *et al.* Discovery and Structure–Activity Relationship of (1 R)-8-Chloro-2,3,4,5-tetrahydro-1-methyl-1 H-3-benzazepine (Lorcaserin), a Selective Serotonin 5-HT 2CReceptor Agonist for the Treatment of Obesity. *J. Med. Chem.* **51,** 305–313 (2008).
10. Bunders, C. A. *et al.* Intercepting Bacterial Indole Signaling with Flustramine Derivatives. *J. Am. Chem. Soc.* **133,** 20160–20163 (2011).
11. Littke, A. F. & Fu, G. C. A Convenient and General Method for Pd-Catalyzed Suzuki Cross-Couplings of Aryl Chlorides and Arylboronic Acids. *Angew. Chem. Int. Ed. Engl.* **37,** 3387–3388 (1998).
12. Old, D. W., Wolfe, J. P. & Buchwald, S. L. *A highly active catalyst for palladium-catalyzed cross-coupling reactions: room-temperature Suzuki couplings and amination of unactivated aryl chlorides. Journal of the American Chemical Society* **120,** 9722–9723 (Journal of the American …, 1998).
13. Horton, D. A., Bourne, G. T. & Smythe, M. L. The Combinatorial Synthesis of Bicyclic Privileged Structures or Privileged Substructures. *Chem. Rev.* **103,** 893–930 (2003).
14. *Metal-catalyzed cross-coupling reactions*. (Wiley-VCH, 1998). doi:10.1002/9783527612222
15. Podgoršek, A., Zupan, M. & Iskra, J. Oxidative Halogenation with 'Green' Oxidants: Oxygen and Hydrogen Peroxide. *Angew. Chem. Int. Ed.* **48,** 8424–8450 (2009).

16. Maddox, S. M., Nalbandian, C. J., Smith, D. E. & Gustafson, J. L. A Practical Lewis Base Catalyzed Electrophilic Chlorination of Arenes and Heterocycles. *Org. Lett.* **17,** 1042–1045 (2015).
17. Agarwal, V. *et al.* Enzymatic Halogenation and Dehalogenation Reactions: Pervasive and Mechanistically Diverse. *Chem. Rev.* acs.chemrev.6b00571–56 (2017). doi:10.1021/acs.chemrev.6b00571
18. Neumann, C. S., Fujimori, D. G. & Walsh, C. T. Halogenation Strategies In Natural Product Biosynthesis. *Chem. Biol.* **15,** 99–109 (2008).
19. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneau-Tsodikova, S. & Walsh, C. T. Nature's Inventory of Halogenation Catalysts: Oxidative Strategies Predominate. *Chem. Rev.* **106,** 3364–3378 (2006).
20. Massey, V. Activation of molecular oxygen by flavins and flavoproteins. *J. Biol. Chem.* **269,** 22459–22462 (1994).
21. Walsh, C. T. & Wencewicz, T. A. Flavoenzymes: Versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.* **30,** 175–200 (2013).
22. Entsch, B. & van Berkel, W. J. Structure and mechanism of para-hydroxybenzoate hydroxylase. *FASEB J.* **9,** 476–483 (1995).
23. Schreuder, H. A., van der Laan, J. M., Hol, W. G. & Drenth, J. Crystal structure of p-hydroxybenzoate hydroxylase complexed with its reaction product 3,4-dihydroxybenzoate. *J. Mol. Biol.* **199,** 637–648 (1988).
24. Yeh, E. *et al.* Flavin Redox Chemistry Precedes Substrate Chlorination during the Reaction of the Flavin-Dependent Halogenase RebH †. *Biochemistry* **45,** 7904–7912 (2006).
25. Bitto, E. *et al.* The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins* **70,** 289–293 (2007).
26. Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases †,‖. *Biochemistry* **46,** 1284–1292 (2007).
27. Dong, C. J. *et al.* Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* **309,** 2216–2219 (2005).
28. Blasiak, L. C. & Drennan, C. L. Structural Perspective on Enzymatic Halogenation. *Acc. Chem. Res.* **42,** 147–155 (2009).
29. Flecks, S. *et al.* New Insights into the Mechanism of Enzymatic Chlorination of Tryptophan. *Angew. Chem. Int. Ed.* **47,** 9533–9536 (2008).
30. Yeh, E., Garneau, S. & Walsh, C. T. Robust in vitro Activity of RebF and RebH, a Two-Component Reductase/Halogenase, Generating 7-Chlorotryptophan during Rebeccamycin Biosynthesis. *Proc. Natl. Acad. Sci.* **102,** 3960–3965 (2005).
31. Keller, S. *et al.* Purification and Partial Characterization of Tryptophan 7-Halogenase (PrnA) from Pseudomonas fluorescens. *Angew. Chem. Int. Ed. Engl.* **39,** 2300–2302 (2000).
32. Zehner, S. *et al.* A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in Streptomyces rugosporus LL-42D005. *Chem. Biol.* **12,** 445–452 (2005).
33. Chang, F.-Y. & Brady, S. F. Cloning and Characterization of an Environmental DNA-Derived Gene Cluster That Encodes the Biosynthesis of the Antitumor Substance BE-54017. *J. Am. Chem. Soc.* **133,** 9996–9999 (2011).

34. Ryan, K. S. Biosynthetic Gene Cluster for the Cladoniamides, Bis-Indoles with a Rearranged Scaffold. *PLoS ONE* **6,** e23694–12 (2011).
35. Chang, F. Y. & Brady, S. F. Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proc. Natl. Acad. Sci.* (2013). doi:10.1073/pnas.1218073110/-/DCSupplemental
36. Seibold, C. *et al.* A flavin-dependent tryptophan 6-halogenase and its use in modification of pyrrolnitrin biosynthesis. *Biocatalysis and Biotransformation* **24,** 401–408 (2006).
37. Zeng, J. & Zhan, J. Characterization of a tryptophan 6-halogenase from Streptomyces toxytricini. *Biotechnol Lett* **33,** 1607–1613 (2011).
38. Menon, B. R. K. *et al.* Structure and biocatalytic scope of thermophilic flavin-dependent halogenase and flavin reductase enzymes. *Org. Biomol. Chem.* **14,** 9354–9361 (2016).
39. Zhu, X. *et al.* Structural Insights into Regioselectivity in the Enzymatic Chlorination of Tryptophan. *J. Mol. Biol.* **391,** 74–85 (2009).
40. Chankhamjon, P. *et al.* Biosynthesis of the Halogenated Mycotoxin Aspirochlorine in Koji Mold Involves a Cryptic Amino Acid Conversion. *Angew. Chem. Int. Ed.* **53,** 13409–13413 (2014).
41. Agarwal, V. *et al.* Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat. Chem. Biol.* **10,** 640–647 (2014).
42. Zeng, J. & Zhan, J. A Novel Fungal Flavin-Dependent Halogenase for Natural Product Biosynthesis. *ChemBioChem* **11,** 2119–2123 (2010).
43. Wang, S. *et al.* Functional Characterization of the Biosynthesis of Radicicol, an Hsp90 Inhibitor Resorcylic Acid Lactone from Chaetomium chiversii. *Chem. Biol.* **15,** 1328–1338 (2008).
44. Sato, M. *et al.* Combinatorial Generation of Chemical Diversity by Redox Enzymes in Chaetoviridin Biosynthesis. *Org. Lett.* **18,** 1446–1449 (2016).
45. Dorrestein, P. C., Yeh, E., Garneau-Tsodikova, S., Kelleher, N. L. & Walsh, C. T. Dichlorination of a pyrrolyl-S-carrier protein by FADH(2)-dependent halogenase PltA during pyoluteorin biosynthesis. *Proc. Natl. Acad. Sci.* **102,** 13843–13848 (2005).
46. Yamanaka, K., Ryan, K. S., Gulder, T. A. M., Hughes, C. C. & Moore, B. S. Flavoenzyme-Catalyzed Atropo-Selective N,C-Bipyrrole Homocoupling in Marinopyrrole Biosynthesis. *J. Am. Chem. Soc.* **134,** 12434–12437 (2012).
47. Hohaus, K. *et al.* NADH-Dependent Halogenases Are More Likely To Be Involved in Halometaolite Biosynthesis Than Haloperoxidases. *Angewandte Chemie International Edition in English* **36,** 2012–2013 (1997).
48. Podzelinska, K. *et al.* Chloramphenicol Biosynthesis: The Structure of CmlS, a Flavin-Dependent Halogenase Showing a Covalent Flavin–Aspartate Bond. *J. Mol. Biol.* **397,** 316–331 (2010).
49. Chankhamjon, P. *et al.* Regioselective Dichlorination of a Non-Activated Aliphatic Carbon Atom and Phenolic Bismethylation by a Multifunctional Fungal Flavoenzyme. *Angew. Chem.* **128,** 12134–12138 (2016).
50. Lin, S., Van Lanen, S. G. & Shen, B. Regiospecific Chlorination of (S)-β-Tyrosyl- S-Carrier Protein Catalyzed by SgcC3 in the Biosynthesis of the Enediyne Antitumor Antibiotic C-1027. *J. Am. Chem. Soc.* **129,** 12432–12438 (2007).

51. Andorfer, M. C., Belsare, K. D., Girlich, A. M. & Lewis, J. C. Aromatic Halogenation Using Bifunctional Flavin Reductase-Halogenase Fusion Enzymes. In preparation.
52. Pacholec, M., Sello, J. K., Walsh, C. T. & Thomas, M. G. Formation of an aminoacyl-S-enzyme intermediate is a key step in the biosynthesis of chloramphenicol. *Org. Biomol. Chem.* **5,** 1692–3 (2007).
53. Buedenbender, S., Rachid, S., Müller, R. & Schulz, G. E. Structure and Action of the Myxobacterial Chondrochloren Halogenase CndH: A New Variant of FAD-dependent Halogenases. *J. Mol. Biol.* **385,** 520–530 (2009).
54. Pang, A. H., Garneau-Tsodikova, S. & Tsodikov, O. V. Crystal structure of halogenase PltA from the pyoluteorin biosynthetic pathway. *Journal of Structural Biology* **192,** 349–357 (2015).
55. Glenn, W. S., Nims, E. & O'Connor, S. E. Reengineering a Tryptophan Halogenase To Preferentially Chlorinate a Direct Alkaloid Precursor. *J. Am. Chem. Soc.* **133,** 19346–19349 (2011).
56. Shepherd, S. A. *et al.* Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chem. Sci.* **6,** 3454–3460 (2015).
57. Lang, A. *et al.* Changing the Regioselectivity of the Tryptophan 7-Halogenase PrnA by Site-Directed Mutagenesis. *Angew. Chem. Int. Ed.* **50,** 2951–2953 (2011).
58. Shepherd, S. A. *et al.* A Structure-Guided Switch in the Regioselectivity of a Tryptophan Halogenase. *ChemBioChem* **17,** 821–824 (2016).
59. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10,** 866–876 (2009).
60. Shimotohno, A., Oue, S., Yano, T., Kuramitsu, S. & Kagamiyama, R. Demonstration of the importance and usefulness of manipulating non-active-site residues in protein design. *J. Biochem.* **129,** 943–948 (2001).
61. Reetz, M. T., Carballeira, J. D. & Vogel, A. Iterative Saturation Mutagenesis on the Basis of B Factors as a Strategy for Increasing Protein Thermostability. *Angew. Chem. Int. Ed.* **45,** 7745–7751 (2006).
62. Fasan, R., Chen, M. M., Crook, N. C. & Arnold, F. H. Engineered Alkane-Hydroxylating Cytochrome P450BM3 Exhibiting Nativelike Catalytic Properties. *Angew. Chem. Int. Ed.* **46,** 8414–8418 (2007).
63. Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329,** 305–309 (2010).

Chapter 2: Characterizing the substrate scope of wild-type RebH and evolving RebH variants with increased thermostability

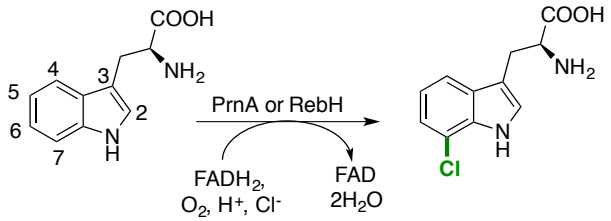
2.1 Introduction

Halogenated aromatic compounds often exhibit unique bioactivities and are thus widely used as agrochemicals and pharmaceuticals.[1-4] It has been demonstrated that these bioactivities are often directly influenced by the presence of halogen substituents.[5-7] In addition to target compounds, halogenated arenes are frequently used as building blocks in organic synthesis, for example in cross-coupling reactions.[8] Although the utility of halogenated arenes cannot be overstated, methods of generating these compounds, namely electrophilic aromatic substitution (EAS) methods, often require harsh reaction conditions.[9] Moreover, the site-selectivity of these reactions is determined by substrate electronics, whereby the most electronically activated C-H bonds are converted to C-Cl bonds.[10] In cases where multiple sites are equivalently activated, halogenation occurs at all sites and gives product mixtures (e.g. ortho and para halogenation of phenol). Thus, the ability to control the selectivity of halogenation reactions would significantly contribute to the synthetic community.

In nature, selective arene halogenation is accomplished by flavin-dependent halogenases (FDHs).[11] Among the most well characterized FDHs are tryptophan halogenases (Trp-FDHs), which catalyze chlorination of L-tryptophan at the less electronically favored 5-,[12], 6-,[13] or 7-positions,[14,15] instead of the more electronically activated 2-position (Scheme 2.1). Biochemical and structural characterization of two 7-Trp-FDHs, PrnA and RebH, have provided insight into the mechanism by which this novel site-selectivity is conveyed.[15-20] Reduced FAD is used to

reduce O_2 and form a flavin peroxide within the FDH. This flavin peroxide can oxidize a closely bound chloride ion, thus forming hypochlorous acid (HOCl). This species migrates to the FDH active site where it is stabilized by an active site lysine residue, either through covalent modification[17] or hydrogen bonding[16]. This lysine residue is positioned proximal to C7 of L-tryptophan, and thus EAS at this site is enabled, producing 7-Cl-tryptophan (Scheme 2.1).

Scheme 2.1: RebH or PrnA catalyzed 7-chlorination of L-tryptophan.



Despite the mild reaction conditions and exquisite catalyst-controlled selectivity, FDH use in preparative synthesis remained largely unexplored at the outset of this work. Most reports prior to the work described herein explored catalysis on the analytical scale with the native substrate L-tryptophan; however, van Pee and coworkers had investigated the substrate scope of PrnA using purified enzyme from its native host *Pseudomonas fluorescens*.[21] Although PrnA was reported to halogenate a variety of substituted indoles, halogenation occurred at the most electronically activated site (C2) for all non-native substrates. In contrast, O'Connor and coworkers demonstrated that RebH exclusively halogenates the non-native substrate tryptamine at C7.[22] Inspired by these examples, we sought to develop an FDH halogenation system capable of preparative, environmentally friendly, selective catalysis. Using this system, we aimed to characterize the scope and site-selectivity of the FDH RebH on non-native substrates. In addition, we sought to develop a robust protocol for the directed evolution of FDHs, which could subsequently be used to broaden native FDH substrate scope as well as alter site-selectivity. Papers have been published summarizing these findings.[23,24]

Authorship

This work was performed with Dr. James Payne and Dr. Catherine Poor. Dr. Payne conducted most of the experiments described in section 2.2.1 and contributed to the results detailed in section 2.2.2. Dr. Poor conducted the directed evolution campaign described in section 2.2.3 as well as characterized melting temperatures and conversion-temperature profiles. These results have been included for clarity.

2.2 Results and Discussion

2.2.1 Optimizing biocatalysis with RebH

With the goal of catalyst-controlled, site-selective halogenation in mind, we selected Trp-FDHs as targets for biocatalysis efforts. As previously discussed, Trp-FDHs use catalyst control to override substrate electronic bias to halogenate the 5-, 6- and 7-positions of L-tryptophan, as opposed to the more electronically activated 2-position. Among the Trp-FDHs, the tryptophan-7-halogenase RebH was chosen for these studies because of its high catalytic efficiency compared with other Trp-FDHs and because it had been previously over-expressed in *E. coli*.[15]

In order to halogenate L-Trp, RebH requires reduced FAD, which is used to form an FAD-peroxide intermediate. FAD-peroxide is thought to generate hypohalous acid (HOX) within RebH, which travels to the active site where it interacts with a lysine residue and halogenates L-Trp. To generate reduced FAD, it has been demonstrated that the chemical reductant dithiotheritol (DTT) can be used;[22] however, we observed lower total turnover numbers when compared with flavin reductases. Several reports have shown that *E. coli* reductases, such as SsuE[14] and Fre[12], can be used to generate reduced flavin in these halogenation systems. RebF, a reductase located in the same biosynthetic gene cluster as RebH, had been previously over-

expressed in E. coli and was used to determine kinetic parameters for RebH.[15] We chose to investigate RebF for our biocatalysis efforts.

When Dr. Payne initially expressed RebH and RebF, low protein yields were observed after purification on Ni-NTA resin (15 and 3 mg/L, respectively). Although both enzymes were sufficiently over-expressed, nearly all RebH and RebF expressed as insoluble inclusion bodies. It has been shown that fusing the target protein to a well-folded protein can increase solubility.[25] Encouraged by these reports, Dr. Payne constructed fusions of RebH and RebF with four solubility partners. When RebH and RebF were fused with maltose binding protein (MBP), increased solubility and purified protein yields were observed.

Upon testing the activities of MBP-RebH and MBP-RebF, we discovered that while MBP-RebH retains no halogenation activity, the turnover number of MBP-RebF (MBP-F) exceeds that of RebF ($k_{cat, RebF}$ = 32 min^{-1}, $k_{cat, MBP\text{-}RebF}$ = 80 min^{-1}). After expression optimization, 33 mg/L of purified MBP-RebF could be readily obtained and used to regenerate reduced flavin for halogenation reactions.

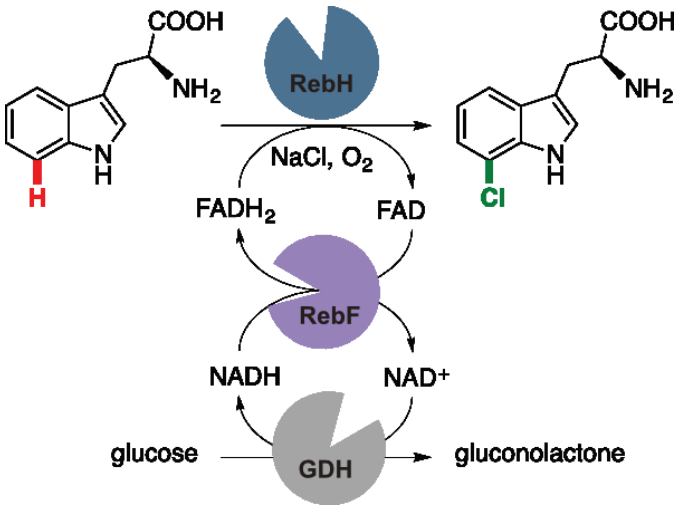
Since the MBP-RebH fusion did not retain halogenation activity, co-expression with chaperone proteins was employed as a means to increase solubility.[26] Five commercially available chaperone plasmids were transformed into RebH-containing *E. coli* BL21 (DE3) and the chaperone and halogenase genes were co-expressed. Cells expressing the GroEL-GroES chaperone proteins (encoded on pGro7), were found to produce higher concentrations of soluble RebH. After expression optimization of this system, we were able to obtain 111 mg/L of soluble RebH and could store this as either a solution (25 mM HEPES buffer, 10% glycerol, pH = 7.4) or a lyophilized powder.

Table 2.1: Yields of RebH and RebF under different conditions.

Conditions	RebH (mg L^{-1})	RebF (mg L^{-1})
Literature conditions[15]	15	3
With pGro7 or MBP fusion	111	33

With large quantities of purified RebH and MBP-F in hand, we sought to complete our cofactor regeneration cycle (Fig. 2.1). A third enzyme component capable of regenerating NADH from NAD would allow us to use catalytic amounts of NAD and FAD in halogenation reactions. Glucose dehydrogenases are commonly used in biocatalysis for this purpose, whereby NADH can be regenerated through oxidation of glucose. Commercially available glucose dehydrogenase (GDH-105) was purchased from Codexis and was found compatible with this system. All concentrations of reaction components were optimized and a robust cofactor regeneration system was developed (Fig. 2.1). Thus, halogenation reactions with RebH consist of catalytic amounts of NAD, FAD, MBP-F and GDH, and stoichiometric amounts of NaCl, glucose and O_2.

Figure 2.1: Cofactor regeneration system used in RebH bioconversions.



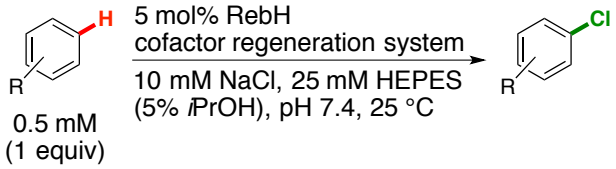
2.2.2 Initial exploration of RebH site-selectivity on non-native substrates

The scope of wild-type RebH was explored through analytical bioconversions on the 75 μL scale using the cofactor regeneration system outlined in in section 2.2.1. Many indoles, as well as some simple arenes, were tested. Significant conversion to chlorinated product was detected by LCMS for several of these compounds. Interestingly, some arenes, including gramine, *N*-methyl-tryptophan, and indole, that were reportedly not halogenated by PrnA, a close FDH homolog, were halogenated by RebH.[21] Very recently, a new FDH was also reported to have activity on many 3-substituted indoles, but, like PrnA, not indole itself.[27] In addition to these interesting differences in substrate scope, high conversions for several tryptamines, tryptophol, naphthol, and 2-aminonaphthalene were observed. Since these studies, a more comprehensive study of the scope of wild-type RebH has been performed in collaboration with Novartis Institutes for Biomedical Research and is described in Chapter 5.[28]

Although PrnA had previously been demonstrated to halogenate the more electronically activated C2 position of all 3-substituted indoles tested,[21] the O'Connor lab showed that RebH selectively halogenates tryptamine at C7.[22] Encouraged by this result, we decided to characterize the selectivity of halogenation by RebH for a set of non-native substrates. To determine site-selectivities, as well as to demonstrate the feasibility of preparative scale reactions with this enzyme class, bioconversions were conducted for several non-native substrates on the 10 mg scale (Scheme 2.2). Bioconversions were carried out in crystallization dishes and were open to air, using buffer as solvent. All reagents and enzymes were added as stock solutions to the crystallization dish. Bioconversions were monitored by HPLC until maximum conversion to mono-chlorinated product was observed. To lessen emulsions formed during liquid-liquid extractions, protein was precipitated with acid and filtered through a pad of Celite. After liquid-

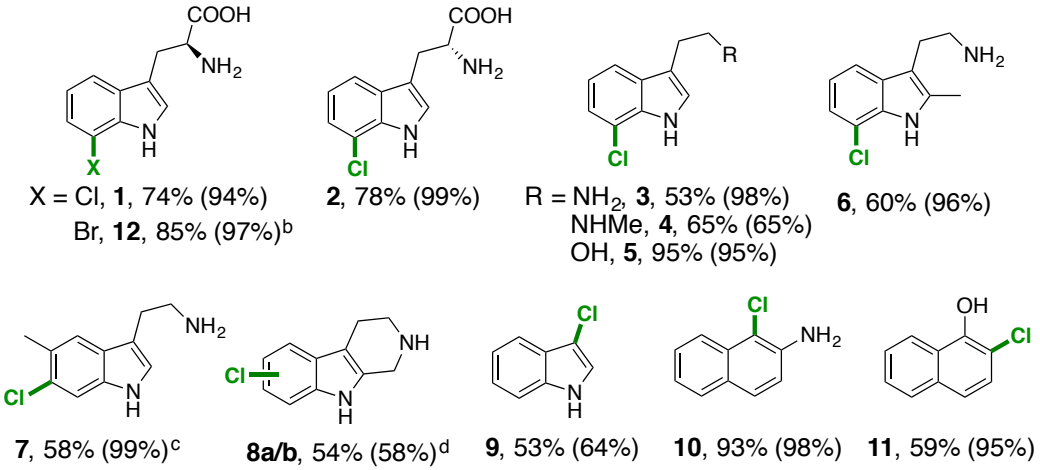
liquid extraction into dichloromethane, products were purified by either normal or reverse phase chromatography.

Scheme 2.2: General scheme for RebH catalyzed halogenation.



After pure mono-chlorinated product was obtained for several non-native substrates, the site of halogenation on these products was determined by NMR spectroscopy. RebH was found to halogenate several non-native substrates at sites other than the most electronically activated site. Examples of this include 7-chlorination of D-tryptophan, tryptamine, *N*-methyltrytpamine, and tryptophol (Fig. 2.2, **2**-**5**). Surprisingly, 5-methyltryptamine was exclusively chlorinated at C6 and tryptoline at C6 and C7 (Fig. 2.2, **7**-**8**). Although the C2 of the substrate 2-methyltryptamine is blocked from EAS, it is nonetheless remarkable that exclusive 7-chlorination is observed when the 4-, 5-, and 6-positions are all similarly activated (Fig. 2.2, **6**). As we moved away from 3-substituted indoles, we saw halogenation at only the most electronically activated sites (i.e. those that would be chlorinated by a reagent such as *N*-chlorosuccinimide, Fig. 2.2, **9**-**11**). Later work has demonstrated that RebH does indeed halogenate some non-indole compounds at disfavored sites; this work is described in Chapter 5.[28]

Figure 2.2: Yields and selectivities for preparative RebH catalyzed halogenation reactions.[a]



[a] Yields of isolated products and HPLC conversions (in parenthesis) are provided. The cofactor regeneration system consisted of 0.2 equiv FAD, 0.2 equiv NAD, 0.5% RebF, 50 U mL^{-1} glucose dehydrogenase, and 40 equiv glucose. [b] 100 mM NaBr was used in place of 10 mM NaCl. [c] 6,7-dichloro-5-methyltryptamine was also isolated in 23% yield. [d] 10 mol% RebH loading was used. A nearly 1:1 mixture of 6- and 7-halogenation was observed.

In addition to determining site-selectively on key substrates, we explored bromination reactions as well. RebH was found to brominate non-native substrates with similar efficiency as the corresponding chlorination reactions. A 10 mg bromination was conducted for L-tryptophan, showing brominated products can also be isolated from FDH-catalyzed bromination reactions (Fig. 2.2, **12**). We were also able to demonstrate 100 mg chlorination of L-tryptophan using cell lysates instead of purified enzyme. Lastly, we kinetically characterized RebH on several non-native substrates, including tryptamine, 2-aminonaphthalene, and tryptoline.

Table 2.2: Catalytic parameters for RebH halogenation of select substrates.[a]

Substrate	RebH (mol%)	Conv. (%)	TTN	k_{cat}[b] (min^{-1})	K_m[b] (μM)
L-tryptophan[c]	0.5	75	165	1.1	7.3
tryptamine	1	15	19	0.023	9.0
tryptoline	5	30	3.6	0.027	216
2-aminonaphthalene	1	32	26	0.59	14

[a] 75 μL reactions were conducted as described in Figure 2.2 using the indicated RebH loading and 0.5 mM phenol as an internal standard, quenched with an equal volume of methanol, and analyzed by HPLC. [b] Initial rate data were used to construct Hanes-Woolf plots to determine K_m and k_{cat}. [c] Reported K_m and k_{cat} values for RebH on L-tryptophan are 2.0 μM and 1.4 min^{-1}, respectively.[15]

2.2.3 Directed evolution of RebH for increased thermostability

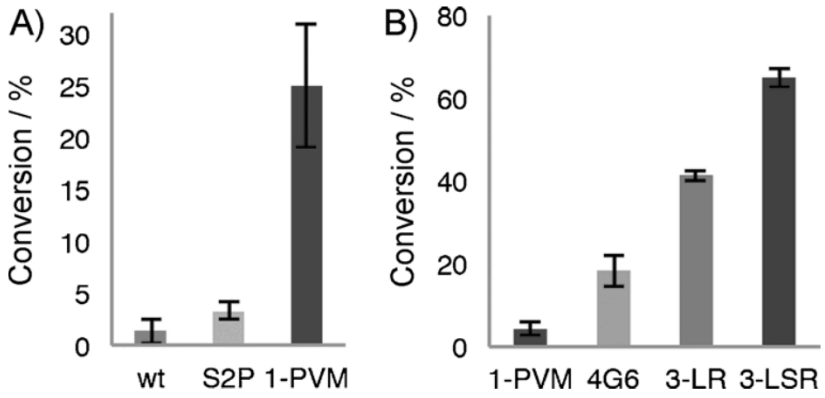
Excited by the broad substrate scope and interesting site-selectivities observed with wild-type RebH, we explored the feasibility of directed evolution with this enzyme. To screen RebH variants each round of evolution, halogenation activity must be observable using RebH cell free extracts grown in 96-well plates. When Dr. Payne and I first tried expressing RebH as 1 mL cultures in 96-deep-well plates, we again observed low expression of soluble RebH, even when co-expressing with chaperone proteins. After optimization, we discovered that inducing RebH expression in 1 mL cultures with 10-fold less isopropyl β-D-1-thiogalactopyranoside (IPTG, 10 μM) produces ~10-fold more soluble RebH. When the RebH gene is expressed in plates using our optimized conditions, cell free extract from these cultures can be used to halogenate tryptophan with good conversions (full conversion to chlorinated product is observed when entire cell extract is used).

Using this expression procedure, Dr. Catherine Poor began using directed evolution to increase the thermostability of RebH. We were interested in pursuing thermostability initially because it has been linked to longer catalyst lifetime and higher optimum reaction

temperatures.[29,30] Stable enzymes can also better tolerate potentially destabilizing mutations introduced to alter enzyme properties such as scope or selectivity.[31] For these reasons, we speculated that a more thermostable variant could prove useful in subsequent engineering efforts.

To create more thermostable RebH variants, three rounds of random mutagenesis and screening were performed. Dr. Poor screened cell extracts for conversion of L-tryptophan to chlorinated product after submitting the extracts to heat treatment. In this way, variants capable of withstanding higher temperatures that retain halogenation activity could be discovered. With each progressive round, variants capable of withstanding higher temperatures were found (Fig. 2.3, round 1 - S2P, 1PVM; round 2 - 4G6; round 3 - 3LR, 3LSR).

Figure 2.3: Halogenation conversion of RebH variants after heat treatment.[a]

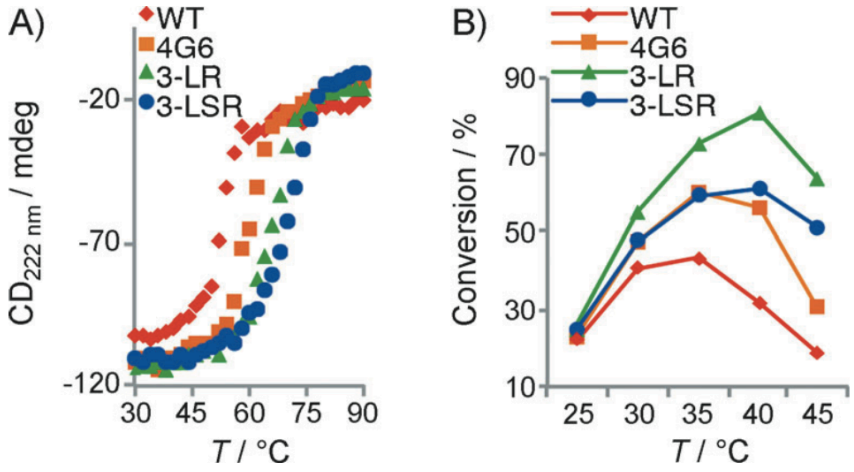


[a] RebH and RebH variants were heat treated at 49 °C for2 hours before conducting halogenation reactions. Reactions were performed on L-tryptophan with A) 2 mol% and B) 0.5 mol% FDH loading.

One of the final variants from the third round of mutagenesis and screening, 3LSR, displayed a melting temperature 18 °C higher than that of wild type RebH (Fig. 2.4). In addition, this variant displayed a higher optimal reaction temperature, as judged by total conversion to

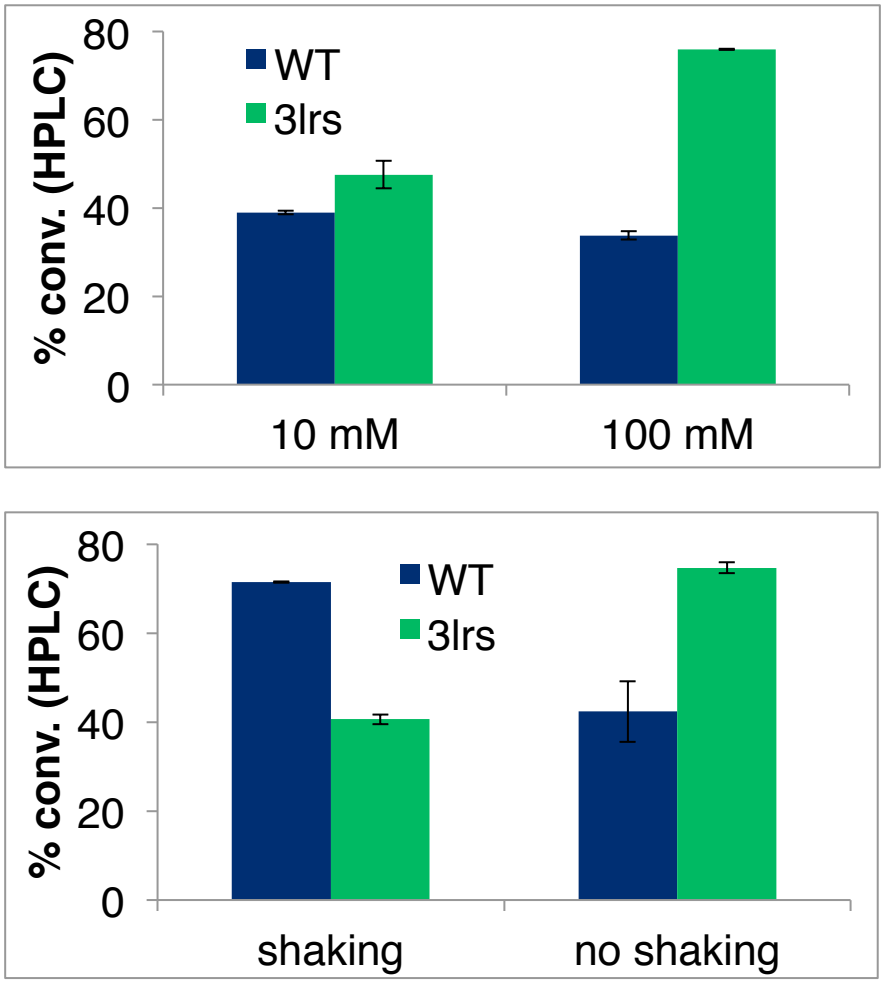
halogenated product (Fig. 2.4). These results signify that 3LSR is a more stable enzyme, enabling the use of higher reaction temperatures in FDH halogenations.

Figure 2.4: Thermal denaturation curves (A) and conversion-temperature profiles (B) for RebH variants.



Upon further reaction optimization, I found that 3LSR is more sensitive to concentration of NaCl than wild-type RebH. While little difference in conversion of 2-methyltryptamine is observed when increasing the [NaCl] from 10 mM to 100 mM, a nearly two-fold increase is observed for 3LSR (Fig. 2.5). In addition to this, whether or not halogenation reactions are shaken has opposite impacts for RebH and 3LSR. When reactions with RebH are not shaken, a significant decrease in total conversion is observed. Surprisingly, roughly the same decrease in total conversion is observed for 3LSR if reactions are shaken (Fig. 2.5). As oxygen is a reagent in FDH reactions, both of these differences could be due to differences in optimal O_2 concentrations between FDHs. With no reaction shaking, less oxygen can diffuse into the full reaction mixture. Similarly, increased NaCl concentration leads to lower oxygen solubility in the reaction mixture. Although this hypothesis has not been thoroughly explored, differences in oxygen sensitivity seem a plausible explanation for these observations. Researchers in the Merck Biocatalysis group are currently exploring this hypothesis.

Figure 2.5: FDH halogenation is affected by salt concentration and reaction shaking.[a]



[a] Halogenation reactions were conducted on the substrate 2-methyltryptamine similarly to those in section 2.2.2. Top) Reactions were not shaken. 10 and 100 mM NaCl concentrations were tested. Bottom) 100 mM NaCl was use in these reactions.

Using optimal reaction conditions for each enzyme (including optimal [NaCl], reaction shaking, and temperature), I conducted preparative-scale bioconversions for L-tryptophan and three non-native substrates (Scheme 2.3, **1**-**4**). Increased product yields (1.7-4.1 fold) were observed in all cases for 3L(S)R (Table 2.3). Interestingly, a change in selectivity for the 6- vs. 7-position was observed upon isolation of the mono-chlorinated tryptoline products (**4**) - from 60% selectivity for C6 with RebH to 88% with 3LSR.

Scheme 2.3: General scheme for RebH catalyzed halogenation and substrates used to explore scope.

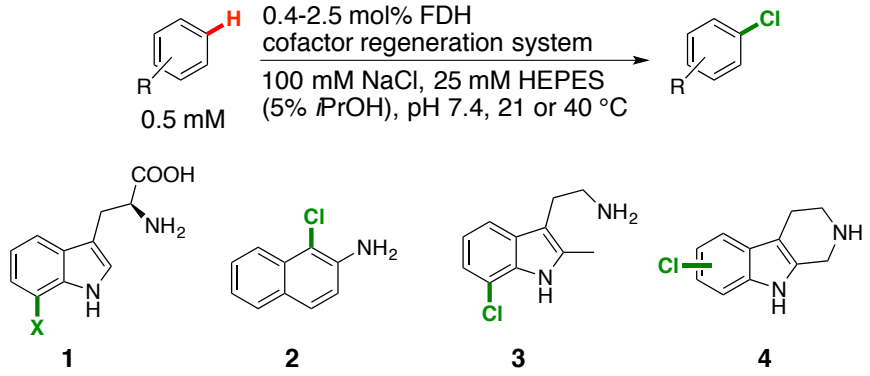


Table 2.3: Representative yields for preparative 3L(S)R catalyzed[a] halogenation reactions and comparison to wild-type RebH reactions.

Product	FDH (mol%)	Temp. (°C)	Time (h)	Yield[a] (min^{-1})	Fold-improvement[b]
1	3LR (0.4)	40	16	69	2.8
2	3LSR (0.8)	21	30	62	2.3
3	3LSR (1.0)	40	36	56	4.1
4	3LSR (2.5)	21	48	67[c]	1.7

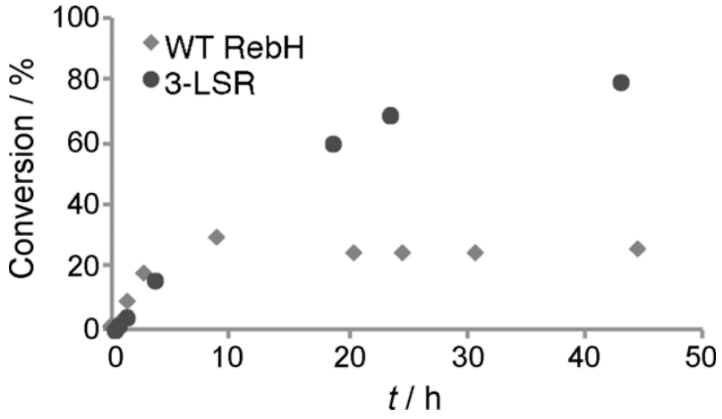
[a] Isolated product yields. [b] Ratio of product concentrations relative to internal standard from HPLC analysis of crude reaction mixtures when using 3L(S)R or wild-type RebH. [c] 88:12 ratio of 6-/7-halogenation products.

Steady state kinetic parameters were measured and compared with RebH at two temperatures (21 and 40 °C) to further characterize 3LSR. At both temperatures, the catalytic efficiency of RebH is higher than that of 3LSR, indicating that higher product yields with 3LSR are not a result of a faster enzyme (Table 2.4). Instead, it was found that 3LSR has a substantially increased catalyst lifetime (Fig. 2.6). These results are consistent with the notion that increasing an enzyme's stability can reduce conformational flexibility and thus decrease catalytic activity.[32,33] Similar reports for stabilized enzymes have been reported;[34,35] however, stabilized enzymes with unchanged[36,37] or even increased[38,39] catalytic efficiencies have also been reported.

Table 2.4: Catalytic parameters for halogenation of 2-methyltryptamine by RebH and 3LSR at 21 °C and 40 °C.

FDH	Temp. (°C)	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} μM^{-1})
RebH	40	0.25	280.1	8.8×10^{-4}
3LSR	40	0.15	202.5	7.9×10^{-4}
RebH	21	0.060	16.8	3.6×10^{-3}
3LSR	21	0.013	40.2	3.3×10^{-4}

Figure 2.6: Time courses for RebH and 3LSR for 2-methyltryptamine at 40 °C.



Until this work, no halogenase from a thermophilic organism had been characterized. It should be noted, however, that Micklefield and coworkers recently discovered a 6-Trp-FDH from a thermophilic bacteria.[40] Although this provides another potentially useful biocatalyst or starting point for further engineering efforts, the most thermostable variant from our directed evolution lineage displays a 22.2 °C higher melting temperature than the native FDH (native FDH: 47.8 °C; 3LSR: 70.0 °C). Thus, although genome mining is a powerful approach to biocatalyst discovery, directed evolution can complement and even exceed the limits of this approach.

2.3 Conclusions

The flavin-dependent halogenase RebH was explored for use in biocatalysis and directed evolution efforts. The wild-type scope of RebH on various 3-substituted indoles as well as a

limited set of simple arenes was determined. Optimization of RebH expression and a cofactor regeneration system to supply RebH with reduced FAD enabled preparative-scale halogenation reactions for some of these substrates. Interesting site-selectivities contrary to expected EAS selectivity were observed for several non-native substrates upon isolation of products from these reactions. This was in contrast with the previously characterized scope of the homologous FDH PrnA, with which EAS selectivity was observed for all non-native substrates. In addition to characterization of wild-type RebH, the first directed evolution of an FDH was demonstrated, whereby the thermostability of RebH was increased through three rounds of random mutagenesis and screening. Developing a robust procedure for FDH evolution as well as stabilized FDH variants sets the stage for further engineering efforts of FDHs.

2.4 Experimental

2.4.1 General experimental procedures

Materials:

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Deuterated solvents were obtained from Cambridge Isotope labs. Silicycle silica gel plates (250 mm, 60 F254) were used for analytical TLC, and preparative chromatography was performed using SiliCycle SiliaFlash silica gel (230-400 mesh). Oligonucleotides were purchased from Integrated DNA Technologies (San Diego, CA), and the sequences of the primers used in this study are reported below.

Plasmids pET-28a/RebF and pET-28a/RebH in BL-21 DE3 *E. coli* were provided by the Walsh group of Harvard Medical School, Boston, MA.[15] The pLIC-MBP plasmid was provided by the Bottomley group of Monash University, Clayton, Australia.[41] The pGro7 plasmid encoding the groES and groEL chaperone set was purchased from Takara (Otsu, Shiga, Japan).

DH5α and BL21(DE3) *E. coli* cells were purchased from Invitrogen (Carlsbad, CA). SacII restriction enzyme, T4 DNA polymerase, and Phusion HF polymerase were purchased from New England Biolabs (Ipswitch, MA). Luria broth (LB) and Terrific broth (TB) media were purchased from Research Products International (Mt. Prospect, IL). Qiagen Miniprep Kits were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. All genes were confirmed by sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility (900 E. 57th Street, Room 1230H, Chicago, IL 60637). Electroporation was carried out on a Bio-Rad MicroPulser using method Ec3. Ni-nitrilotriacetic acid (Ni-NTA) resin and Pierce® BCA Protein Assay Kits were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were following when using both products (for Ni-NTA resin, 5 mL resin were used, with buffers delivered by a peristaltic pump at a rate of 1 mL/min, in a 4 °C cold cabinet). Amicon® 30 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C. The glucose dehydrogenases GDH-105 and CDX-901, FAD, and NAD were purchased from Codexis (Redwood City, CA). NADH was purchased from Chem-Impex International (Wood Dale, IL). DOWEX™ 50WX8 strong cation exchange resin was purchased from Sigma-Aldrich. Biotage reverse phase columns (SNAP-KP-C18-HS) and amine-modified silica columns (SNAP-KP NH) were purchased from Biotage.

General Procedures:

Standard molecular cloning procedures were followed.[42] Reactions were monitored using HPLC (Agilent 1200 UHPLC with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN; 0-10 min, B = 15%; 10-20 min, B

= 15-100%; 20-24 min, B = 100%). Reverse phase preparative chromatography was carried out using a Biotage Isolera One. ^{1}H and ^{13}C NMR spectra were recorded at 500 MHz and 126 MHz, respectively, on a Bruker DMX-500 or DRX-500 spectrometer, and chemical shifts are reported relative to residual solvent peaks.[43] High-resolution mass spectra were obtained at the University of Chicago Mass Spectrometry Service Center on an Agilent Technologies 6224 TOF LC/MS. Low resolution mass spectra were obtained at the University of Chicago Mass Spectrometry Service Center on a Varian Saturn 2000, CP-3800 GC-MS/MS. The T_m's of enzymes were found using a AVIV 202 CD Spectrometer with Peltier temperature controller and SigmaPlot (Systat Software, San Jose, CA).

Sonication Conditions:

5 mL cultures: sonication was performed on a Qsonica S-4000 Sonicator with a microtip using the following procedure: 6x30s with 30s rests, 30% duty cycle delivering 15 W, while cooling on a circulating ice-water bath.

750 mL cultures: sonication was performed on a Qsonica S-4000 Sonicator with a 0.5" horn using the following procedure: 5x1m with 1m rests, 20% duty cycle delivering 40-50 W, while cooling on a circulating ice-water bath.

2.4.2 Specific experimental procedures

Cloning of MBP-RebF Fusion Plasmid: pET-28a/RebF was isolated from the cell stock obtained from the Walsh group using a Qiagen Miniprep Kit. Chemically competent DH5α *E. coli* stocks were prepared using the calcium chloride method described in Sambrook and Russell,

p. 1.116.[42] RebF was amplified from the isolated pET-28/RebF described using the following oligonucleotides:

Table 2.5: Table of primers for section 2.2.1.

#	Primer Name	Sequence
1	RebFFusNdeF	5' – CCA GGG AGC AGC CTC GCA TAT GAC GAT CGA GTT CGA CAG ACC CG – 3'
2	RebFFusHindR	5' – GCA AAG CAC CGG CCT CGT TAC ATA TGT CCC TCC GGT GTC CAC ACG GCG – 3'

The PCR conditions were as follows: 1x Phusion HF buffer, 0.2 mM dNTPs each, 0.2 μM forward primer, 0.2 μM reverse primer, and 0.02 U/μl Phusion polymerase. The final PCR reaction volumes were 50 μL each, and were carried out in an Eppendorf Mastercycler pro S thermal cycler. The PCR procedure was as follows: 95 °C for 1 minute; 94 °C for 3 minutes; 30 cycles of 98 °C for 30 seconds, 50 °C for 1 minute, and 72 °C for 3 minutes; and ending with 72 °C for 5 minutes. Formation of pLIC-MBP/RebF was performed as described in Cabrita *et al.*,[41] with the ligation product transformed into the chemically competent DH5α *E. coli* described above and plated onto LB/agarose plates containing 0.1 mg/mL ampicillin. A glycerol stock was prepared from 3.0 mL of overnight culture in LB with 0.1 mg/mL ampicillin, which was spun down (3,000g, 10 min) and resuspended in 350 μL fresh LB with 0.1 mg/mL ampicillin, then diluted with 500 μL 50% v/v water/glycerol. Glycerol stocks were all stored at -80 °C. The pLIC-MBP/RebF plasmid was isolated from this DH5α *E. coli* using a Qiagen Miniprep Kit. Electrocompetent BL21(DE3) *E. coli* stocks were prepared using the method described in Sambrook and Russell, p. 1.120.[42] Isolated pLIC-MBP/RebF plasmid was transformed by electroporation into electrocompetent BL21(DE3) *E. coli* and plated onto LB/agarose plates containing 0.1 mg/mL ampicillin. A glycerol stock was prepared as described above and stored at -80 °C.

Cloning of RebH with pGro7 Chaperone Plasmid: Electrocompetent DH5α *E. coli* stocks were prepared as described above. The pGro7 plasmid was transformed by electroporation into electrocompetent DH5α *E. coli* and then plated on LB/agarose plates containing 0.02 mg/mL chloramphenicol. A glycerol stock was prepared from 3.0 mL of overnight culture in LB with 0.02 mg/mL chloramphenicol, which was spun down (3,000g, 10 min) and resuspended in 350 µL fresh LB with 0.02 mg/mL chloramphenicol, then diluted with 500 µL 50% v/v water/glycerol. pGro7 plasmid was then isolated from this cell stock using a Qiagen Miniprep Kit. Electrocompetent pET-28a/RebH BL21(DE3) *E. coli* stocks were prepared as described above. The isolated pGro7 plasmid was transformed by electroporation into electrocompetent pET-28a/RebH BL21(DE3) *E. coli* and plated on LB/agarose plates containing 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin. A glycerol stock was prepared from 3.0 mL of overnight culture in LB with 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin, which was spun down (3,000g, 10 min) and resuspended in 350 µL fresh LB with 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin, then diluted with 500 µL 50% v/v water/glycerol.

Small-Scale Expression of MBP-RebF and RebH: For an initial comparison of improvements to solubility, RebH with and without pGro7 and RebF with and without the MBP fusion tag were expressed in BL21(DE3) *E. coli* overnight at 37 °C in 5 mL LB medium containing the requisite antibiotic(s) (100 µg/mL ampicillin for the fusion tag, 50 µg/mL kanamycin and 20 µg/mL chloramphenicol for the chaperone coexpression, and 50 µg/mL kanamycin for the controls). 5 mL fresh media were inoculated with 50 µL of overnight culture, and growth was continued at 37 °C until the cultures reached an OD_{600} of 0.6. At this point, 100 µM isopropyl β-D-thiogalactopyranoside (IPTG) and 2 mg/mL L-arabinose (for pGro7 chaperone coexpression)

were added to induce expression, and incubation was continued for 20 hours at 37 °C. The cells were harvested by centrifugation at 3,000 g for 10 minutes at 4 °C and resuspended in 1 mL of 25 mM HEPES pH 7.4. The cells were lysed by sonication and pelleted by centrifugation at 16,100 g for 10 minutes at 4 °C. The supernatant was decanted, the pellets were each resuspended in 1 mL of 25 mM HEPES pH 7.4, and the soluble (supernatant) and insoluble (resuspended pellet) proteins for each sample were examined by SDS-PAGE.

Large-Scale Expression and Purification of MBP-RebF: Primary cultures of pLIC-MBP/RebF BL21(DE3) *E. coli* were grown (7.5 mL LB containing 0.1 mg/mL ampicillin in 15 mL Falcon culture tubes) from the glycerol stock described above at 37 °C and 250 rpm overnight. 7.5 mL of primary culture were used to inoculate 750 mL of LB containing 0.1 mg/mL ampicillin in a 2.8 L, non-beveled glass flask. The cultures were then grown at 37 °C and 250 rpm to OD_{600} = 0.6, at which point isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 100 μM. The cultures were then incubated at 37 °C and 250 rpm for an additional three hours, at which time the cells were pelleted by centrifugation (4,000g, 30 min, 4 °C) and stored at -20 °C until further use.

The pellets were each resuspended in 30 mL of HEPES (25 mM, pH 7.4), transferred to 50 mL conical tubes, and lysed by sonication. The lysates were then clarified by centrifugation (24,000g, 30 min, 4 °C). The clarified lysate was then purified by Ni-NTA affinity chromatography. The elution fractions were then concentrated to 3 mL using an Amicon® 15 mL 30 kD cutoff centrifugal filter and exchanged five times into HEPES/glycerol buffer (25 mM, pH 7.5, 10% v/v glycerol). Protein concentrations were measured using the Pierce® BCA Protein Assay Kit and protein stocks were then stored at -20 °C until use.

Large-Scale Expression and Purification of RebH: Primary cultures of pET-28a/RebH & pGro7 BL21(DE3) *E. coli* were grown (7.5 mL LB containing 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin in 15 mL Falcon culture tubes) from the glycerol stock described above at 37 °C and 250 rpm overnight. 3.75 mL of primary culture were used to inoculate 750 mL of TB containing 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin in a 2.8 L, non-beveled glass flask. The cultures were then grown at 37 °C and 250 rpm to OD_{600} = 0.6, at which point isopropyl β-D-1-thiogalactopyranoside (IPTG) and L(+)-arabinose were added to final concentrations of 100 μM and 2 mg/mL, respectively. The temperature was then reduced to 30 °C and the shake rate left at 250 rpm for an additional 20 hours, at which time the cells were pelleted by centrifugation (4,000g, 30 min, 4 °C) and stored at -20 °C until further use.
The pellets were each resuspended in 30 mL of HEPES (25 mM, pH 7.4), transferred to 50 mL conical tubes, and lysed by sonication. The lysates were then clarified by centrifugation (24,000g, 30 min, 4 °C). The clarified lysate was then purified by Ni-NTA affinity chromatography. The elution fractions were then concentrated to 3 mL using an Amicon® 15 mL 30 kD cutoff centrifugal filter and exchanged five times into HEPES/glycerol buffer (25 mM, pH 7.5, 10% v/v glycerol). Protein concentrations were measured using the Pierce® BCA Protein Assay Kit and protein stocks were then stored at -20 °C until use.

NADH oxidation by MBP-RebF: The activities of RebF and the MBP-RebF fusion were measured by monitoring the decrease in absorbance at 340 nM, which is caused by oxidation of NADH to NAD. For kinetic characterization, 1 μM RebF or MBP-RebF was incubated with 0.01-1.0 mM NADH, 50 μM FAD, and 10 mM NaCl in HEPES (25 mM, pH 7.5) in a 96-well microtiter plate. Reactions were monitored at 340 nm using a Tecan infinite M200pro plate

reader and the resulting data were fit to the equation $y = (B_{max} * x) / (K_d + x)$ using SigmaPlot to obtain values for k_{cat}.

Determination of Kinetic Parameters for RebH: Rates were determined by monitoring the conversion of 75-215 μM substrate in the presence of NAD (100 μM final concentration), FAD (100 μM final concentration), NaCl (10 mM final concentration), MBP-RebF (2.5 μM final concentration), glucose dehydrogenase (50 U/mL final concentration CDX-901), glucose (20 mM final concentration), and phenol as an internal standard (0.5 mM final concentration) at a final volume of 75 μL in a microtiter plate. RebH was added at a final concentration of either 1 μM for L-tryptophan and 2-aminonaphthalene or 25 μM for tryptamine and tryptoline. The reactions were left shaking at 600 rpm at room temperature, then quenched at 5-120 minutes (all time points were collected in triplicate) by addition of 75 μL of MeOH. The precipitated protein was then removed by centrifugation and the reactions were filtered and analyzed by HPLC using the method described in the General Procedures. Product formation was quantitated by calculating the ratio of product to internal standard and fitting that value to a calibration curve prepared from known concentrations of each material. The kinetic parameters (K_m and k_{cat}) for each substrate were determined using the Hanes-Woolf plots (section 2.4.5) constructed from the substrate concentrations and the observed initial rates.

Determination of Kinetic Parameters for RebH and 3-LSR with 2-methyltryptamine: Kinetic parameters were determined for WT RebH and 3-LSR at 21 °C and 40 °C with the unnatural substrate 2-methyltryptamine. Rates were determined by monitoring the conversion of 75-215 μM substrate in the presence of NAD (100 μM final concentration), FAD (100 μM final concentration), NaCl (100 mM final concentration), MBP-RebF (2.5 μM final concentration),

glucose dehydrogenase (50 U/mL final concentration GDH), glucose (20 mM final concentration), and phenol as an internal standard (0.5 mM final concentration) at a final volume of 75 µL in a microtiter plate. WT RebH or 3-LSR was added at a final concentration of 4-25 µM (20 µM WT RebH at 21 °C; 25 µM 3-LSR at 21 °C; 4 µM WT RebH at 40 °C; 8 µM 3-LSR at 40 °C). The reactions were left shaking at 600 rpm in an IKA® microtiter shaker at room temperature, then quenched at 30-105 minutes (all time points were collected in triplicate) by addition of 75 µL of MeOH. The precipitated protein was then removed by centrifugation and the reactions were filtered and analyzed by HPLC (Agilent 1200 UHPLC with an Agilent Eclipse Plus C18 4.6 x 50 mm column, 3.5 µM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN) using the following method: 0-2 min, B = 15%; 2-2.5 min, B = 15-20%; 2.5-4.5 min, B = 20%; 4.5-5.5 min, B = 20-30%; 5.5-6 min, B = 30-70%; 6-7 min, B = 70%. Product formation was quantitated by calculating the ratio of product to internal standard and fitting that value to a calibration curve prepared from known concentrations of each material. The kinetic parameters (K_m and k_{cat}) for each substrate were determined using the Hanes-Woolf plots (section 2.4.5) constructed from the substrate concentrations and the observed initial rates.

Time Courses: Conversion vs. time courses for halogenation by WT RebH and 3-LSR were plotted for the substrate 2-methyltryptamine. 10 mg bioconversions were conducted according to the general procedures for 1 and 10 mg bioconversions. WT RebH and 3-LSR bioconversions were each conducted at a final concentration of 5 µM RebH. The WT bioconversion was mixed at 110 rpm. 3-LSR was not agitated. Aliquots (75 µL) were pulled from 10 mg bioconversions at various times, were quenched with MeOH (75 µL) and were analyzed using the HPLC method described in the general procedures for 1 and 10 mg bioconversions.

Determination of Total Turnover Numbers (TTNs) for RebH: Reactions were assembled in triplicate containing substrate (1 equiv., 0.5 mM final concentration), NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), MBP-RebF (0.005 equiv., 2.5 μM final concentration), glucose dehydrogenase (50 U/mL final concentration CDX-901), glucose (40 equiv., 20 mM final concentration), and phenol as an internal standard (1 equiv., 0.5 mM final concentration) at a final volume of 75 μL in a microtiter plate. RebH was added at a final concentration of either 2.5 μM for L-tryptophan, 5 μM for tryptamine and 2-aminonaphthalene, or 25 μM for tryptoline. The reactions were left shaking at 600 rpm at room temperature, then quenched after 12 hours by addition of 75 μL of MeOH. The precipitated protein was then removed by centrifugation and the reactions were filtered and analyzed by HPLC using the method described in the General Procedures. Product formation was quantitated by calculating the ratio of product to internal standard and fitting that value to a calibration curve prepared from known concentrations of each material.

General Procedure for 10 mg Bioconversions for wild-type RebH scope studies: A solution of 10 mg substrate (1 equiv., 0.5 mM final concentration) in 5 mL HEPES buffer (25 mM, pH 7.4) or isopropanol was added to a crystallization dish (125 x 65 or 100 x 50 mm). Solutions in HEPES buffer (25 mM, pH 7.4) of NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), and a glucose dehydrogenase (50 U/mL final concentration CDX-901) were added. Solutions in HEPES/glycerol buffer (25 mM, pH 7.5, 10% glycerol v/v) of RebH (0.05 equiv., 25 μM final concentration) and MBP-RebF (0.005 equiv., 2.5 μM final concentration) were added. The

bioconversion was diluted with HEPES buffer (25 mM, pH 7.4) to the appropriate reaction volume and an aqueous solution of glucose (40 equiv., 20 mM final concentration) was added to initiate the cofactor regeneration cycle. The dish was covered with perforated aluminum foil and stirred with a magnetic stir bar at 60 rpm.

Reactions were monitored by HPLC as described in the General Procedures and were quenched with aqueous HCl (5.0 M, until pH < 2) upon completion. NaCl was added to saturation. Precipitated protein was filtered out through a pad of Celite and was washed with water and/or CH_2Cl_2 until all product had been rinsed from the Celite pad, as indicated by HPLC. The filtrate was either extracted into CH_2Cl_2 or submitted to strong cation exchange chromatography. The crude material thus obtained was purified by either normal or reverse phase chromatography.

General Procedure for 1 and 10 mg Bioconversions of variants from thermostability lineage: Substrate (1 or 10 mg) was added to a beaker (100 mL) or crystallization dish (100 x 50 mm) as a solution in HEPES buffer (25 mM HEPES, pH = 7.4) or isopropanol (5% v/v). Solutions of NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (200 equiv., 100 mM final concentration), and a glucose dehydrogenase (50 U/mL final concentration GDH) were added to the reaction. This was diluted to the appropriate volume with HEPES buffer, and RebH (0.002-0.1 equiv., 1-50 μM final concentration) and MBP-RebF (0.005 equiv., 2.5 μM final concentration) were added as solutions of HEPES/glycerol buffer (25 mM HEPES, pH 7.5, 10% glycerol v/v). The reaction was initiated with a solution of 1 M glucose (40 equiv., 20 mM final concentration), sealed with an AeraSeal film, and placed in incubator at the desired temperature. During initial screens of substrates at different temperatures, it was found that wild-type RebH halogenates in higher yield when

agitated, whereas 3-LSR has optimal activity in the absence of shaking. Because of this, bioconversions with wild-type enzyme were agitated with 110 rpm; those with 3-LR or 3-LSR were not agitated.

To the 1 mg bioconversions, an internal standard (phenol, 0.5 mM final concentration) was added as a 10 mM solution in DMSO after quenching with one volume of methanol. These reactions were analyzed using the HPLC (Agilent 1200 UHPLC with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN). The following method was used for all substrates: 0-10 min, B = 15%; 10-20 min, B = 15-100%; 20-24 min, B = 100%. The area of product to internal standard was used to compare the activity of the mutant enzymes to wild-type.

The 10 mg bioconversions were quenched with HCl (1 M, until pH<2) and saturated with NaCl. Precipitated protein was filtered out through a pad of Celite. The filtrate was extracted into CH_2Cl_2 or submitted to strong cation exchange chromatography. The crude material was purified by normal or reverse phase chromatography.

Procedure for 100 mg L-Tryptophan Bioconversion: A solution of 100 mg L-tryptophan (1 equiv., 0.5 mM final concentration) in 10 mL HEPES buffer (25 mM, pH 7.4) was added to a crystallization dish (190 x 100 mm). Solutions in HEPES buffer (25 mM, pH 7.4) of NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), and CDX-901 (50 U/mL final concentration) were added. Sixteen 750 mL cultures of RebH coexpressing pGro7 were grown, expressed, spun down, resuspended, lysed, and clarified as described above, then pooled to furnish approximately 500 mL of clarified lysate in HEPES buffer (25 mM, pH 7.4). The rate of conversion of L-tryptophan was monitored for a volume of this lysate and compared to the rate of conversion of L-

tryptophan using a known concentration of purified RebH. From this comparison, an effective RebH concentration in the lysate was calculated. Lysate was added such that the effective final concentration of RebH in the reaction was 15 μM (0.03 equiv.). A solution in HEPES/glycerol buffer of MBP-RebF (0.005 equiv., 2.5 μM final concentration) was added. The bioconversion was diluted with HEPES buffer (25 mM, pH 7.4) to the appropriate reaction volume and an aqueous solution of glucose (40 equiv., 20 mM final concentration) was added to initiate the cofactor regeneration cycle. The reaction was split evenly into two identical crystallization dishes (190 x 100 mm). The dishes were covered with perforated aluminum foil and stirred with magnetic stir bars at 60 rpm.

The reaction was monitored by HPLC as described in the General Procedures. Product concentration was not observed to increase between 20 and 22 hours, and the reaction was quenched with aqueous HCl (5.0 M, until pH < 2) and submitted to centrifugation (3,000g, 10 min) for initial protein removal. The supernatant was collected, and the pellet was washed through three rounds of re-suspension in acidic water (1 M HCl used to adjust pH < 2) and centrifugation (3,000g, 10 min). The supernatants were combined and passed through a Celite pad for further protein removal. The filtrate was submitted to strong cation exchange, following the procedure described below. Elution fractions from ion exchange were pooled, dry loaded onto Celite, and purified by reverse phase chromatography (Biotage SNAP-KP-C18-HS, gradient from pure H_2O to 15% CH_3CN/H_2O).

Procedure for Strong Cation Exchange: After protein precipitation and filtration through Celite, aqueous filtrate was submitted to strong cation exchange to remove salts if the product could not be extracted into organic solvent. DOWEX™ 50WX8 resin was slurry-packed with

methanol in a 250 mL chromatography column. The resin was washed with ~300 mL of methanol and ~300 mL of deionized water. The resin was acidified with HCl (1 M) until the pH of flow through was less than 2. The resin was washed with deionized water until the pH was between 6-7. Acidic filtrate (pH < 2) was added to resin. The resin was washed with ~500 mL of deionized water. The product was eluted with NH_4OH (1 M) until product was no longer eluting from the column, as indicated by HPLC. Product-containing fractions were concentrated to dryness using a rotary evaporator under high vacuum.

Library construction, expression, and screening: All genes encoding RebH were cloned into pET-28a between the NdeI and HindIII digestion sites. Mutant libraries were constructed by error-prone PCR, using *Taq* polymerase with 150 μM $MnCl_2$ (round 1) or 100 μM $MnCl_2$ (rounds 2 and 3). PCR was performed in a volume of 50 μL with conditions of 95 °C 30 s, (95 °C 30 s, 55 °C 30 s, 72 °C 90 s) for 20 cycles, 72 °C 10 min. Beneficial mutations were recombined via overlap extension[44] with PCR conditions of 98 °C 30 s, (98 °C 10 s, 72 °C 50 s) for 35 cycles, 72 °C 10 min. Plasmids were transformed by electroporation into *E. coli* containing the chaperone pGro7. Library colonies were picked using an automated colony picker (Norgren Systems) and arrayed in 1-ml 96-well plates containing 300 μL LB with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol. Cells were grown overnight at 37 °C, 250 rpm, and 50-100 μL of overnight culture was used to inoculate 1 mL TB (with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol) in 2-mL 96-well plates. Following growth at 37 °C, 250 rpm, to an OD_{600} = 0.9-1, enzyme expression was induced with IPTG and arabinose to final concentrations of 10 μM and 0.2 mg/mL, respectively. Protein expression continued for ~20 h at 30 °C, 250 rpm, after which cultures were harvested by centrifugation and stored at -80 °C until use.

Cell pellets were thawed and suspended in 100 µL HEPES buffer (25 mM, pH 7.4) containing 0.75 mg/mL lysozyme. After incubation at 37 °C, 250 rpm, cells were flash frozen in liquid nitrogen and thawed in a 37 °C water bath. Ten microliters of DNaseI at 1 mg/mL were added and the cells incubated at 37 °C, 250 rpm, for 15 min. After centrifugation, 50 µL of supernatant were transferred to a microtiter plate for screening.

Libraries were sealed (AeraSeal, Research Products International), incubated at 42 °C for 2 h (round 1), 51 °C for 2 h (round 2), or 54 °C for 3 h (round 3) and then immediately cooled in an ice water bath. Similar to what has been described previously for halogenation reactions, tryptophan (0.5 mM final concentration) was added to 50 µL lysate as a solution of 25 mM HEPES (pH 7.4). NaCl (10 mM final concentration), FAD (100 µM final concentration), NAD (100 µM final concentration), MBP-RebF (2.5 µM final concentration) and glucose dehydrogenase (50 U/mL final concentration) were also added to reactions as solutions of 25 µM HEPES buffer. A solution of glucose (20 mM final concentration) was added to the reaction mixtures to initiate. Reactions were mixed, the plates sealed, and left overnight on the benchtop. Reactions were quenched with an equal volume of methanol and centrifuged, and the supernatant was filtered and analyzed for 7-chlorotryptophan production via HPLC (Agilent 1200 UHPLC with an Agilent Eclipse Plus C18 2.1 x 50 mm column, 1.8 µM particle size; solvent A = H_2O/0.1 % TFA, solvent B = CH_3CN; 0-0.5 min, B = 16 %; 0.5-1.5 min, B = 16-80 %).

Residual activity determination: The residual activity was determined following incubation of 50 µL of pure protein at 49 °C for 2 h in 1.5-mL microcentrifuge tubes. Tryptophan halogenation reactions consisted of the same reagents used during library screening with the following exceptions: pure protein was substituted for lysate, and the buffer was 20 mM HEPES (pH 7.4),

6.7% glycerol, and 100 mM NaCl. Reactions were conducted overnight on the benchtop, and quenched and analyzed the following day just as the library reactions were.

T_m and T_{opt} analyses: Melting temperature measurements were conducted in 20 mM HEPES (pH 7.4), 150 mM NaCl, and 10 % glycerol, with a protein concentration of 20 μM. Thermal denaturation was irreversible and monitored by circular dichroism spectroscopy using an AVIV 202 CD Spectrometer with Peltier temperature controller. Denaturation was monitored at 222 nm in 2 °C increments from 20-90 °C with 2 min equilibration at each temperature. The midpoint of the denaturation curve was determined with SigmaPlot (Systat Software, San Jose, CA) after fitting to a 4-parameter sigmoid.

Activity-temperature profiles were constructed using purified enzyme with 75 μL reactions in 1.5-mL microcentrifuge tubes. Reactions were conducted in a buffer of 20 mM HEPES (pH 7.4), 6.7% glycerol, and 100 mM NaCl, with 0.5 mM L-tryptophan, 20 mM DTT, and 100 μM FAD. Reactions were conducted overnight at temperatures ranging from 21-45 °C. These were quenched and analyzed by the same methods as the library reactions.

2.4.3 Detailed isolation and characterization I

Detailed isolation and characterization for compounds in section 2.2.2:

7-chloro-L-tryptophan (1): The bioconversion was conducted according to the general procedure in a 100 x 50 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, submitted to strong cation exchange chromatography, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude

material was purified by reverse phase chromatography (gradient from 100% H_2O to 15% CH_3CN/H_2O) to afford **1** in 74% yield (8.6 mg, 0.036 mmol). ^1H NMR (500 MHz, Deuterium Oxide with one drop TFA) δ 7.23 (d, J = 8.0 Hz, 1H), 7.01 (s, 1H), 6.92 (d, J = 7.6 Hz, 1H), 6.76 (t, J = 7.8 Hz, 1H), 4.03 (dd, J = 7.3, 5.4 Hz, 1H), 3.18 – 3.04 (m, 2H). ^{13}C NMR (126 MHz, Deuterium Oxide) δ 177.45, 133.29, 128.61, 125.37, 121.24, 120.13, 117.34, 116.50, 109.97, 55.51, 27.84. HRMS (ESI-TOF) calcd for $C_{11}H_{12}N_2O_2Cl$ $[M + H]^+$: 239.0529 and 241.0502, found: 239.0534 and 241.0501.

7-chloro-D-tryptophan (2): The bioconversion was conducted according to the general procedure in a 100 x 50 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, submitted to strong cation exchange chromatography, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography (gradient from 100% H_2O to 15% CH_3CN/H_2O) to afford 2 in 78% yield (9.1 mg, 0.038 mmol). ^1H NMR (500 MHz, Deuterium Oxide with one drop TFA) δ 7.22 (d, J = 8.0 Hz, 1H), 6.99 (s, 1H), 6.93 (d, J = 7.6 Hz, 1H), 6.77 (t, J = 7.8 Hz, 1H), 4.04 (dd, J = 7.1, 5.4 Hz, 1H), 3.15-3.03 (m, 2H). ^{13}C NMR (126 MHz, Deuterium Oxide) δ 133.40, 128.53, 125.59, 121.40, 120.28, 117.33, 116.61, 109.37, 55.22, 27.04. HRMS (ESI-TOF) calcd for $C_{11}H_{12}N_2O_2Cl$, $[M + H]^+$: 239.0587 and 241.0561, found: 239.0529 and 241.0502. The carboxylate carbon was not detected because of low sample concentration. In the NOESY spectrum of this product, resonances between the singlet at 6.99 ppm and other aryl protons are present. When examining the cross-peaks of these resonances, it

was found they could not be phased properly and thus were not reliable. COSY and HMQC spectra were collected to ensure that these resonances were not true NOEs, and that the proposed structure was thus consistent with the acquired spectra. This is observed for all tryptophan products as well as 7-chlorotryptamine.

7-chloro-tryptamine (3): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP- C18-HS). The crude material was purified by reverse phase chromatography (gradient from water 0.1% TFA to 20% CH_3CN/water 0.1%TFA) to afford **3** in 53% yield (10.2 mg of **3**·TFA, 0.033 mmol). ^{1}H NMR (500 MHz, Deuterium Oxide) δ 7.48 (d, J = 8.0 Hz, 1H), 7.23 (s, 1H), 7.17 (d, J = 7.6 Hz, 1H), 7.01 (t, J = 7.8 Hz, 1H), 3.19 (t, J = 7.0 Hz, 2H), 3.02 (t, J = 7.0 Hz, 2H). ^{13}C NMR (126 MHz, Deuterium Oxide) δ 133.48, 128.15, 125.00, 121.46, 120.27, 117.05, 116.73, 110.36, 39.63, 22.64. HRMS (ESI-TOF) calcd for $C_{10}H_{12}N_2Cl$ $[M + H]^+$: 195.0689 and 197.0661, found: 195.0627 and 197.0664.

7-chloro-*N*-Ω-methyltryptamine (4): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse

phase column (Biotage SNAP-KP- C18-HS). The crude material was purified by reverse phase chromatography (gradient from water 0.1% TFA to 25% CH_3CN/water 0.1%TFA) to afford **4** in 65% yield (12.0 mg of **4**·TFA, 0.037 mmol). ^{1}H NMR (500 MHz, 75:25 Deuterium Oxide:Acetonitrile-d3) δ 7.91 (d, J = 7.9 Hz, 1H), 7.65 (s, 1H), 7.57 (d, J = 7.5 Hz, 1H), 7.43 (t, J = 7.8 Hz, 1H), 3.60 (t, J = 7.5 Hz, 2H), 3.46 (t, J = 7.4 Hz, 2H), 2.99 (s, 3H). ^{13}C NMR (126 MHz, 75:25 Deuterium Oxide:Acetonitrile-d3) δ 132.88, 127.94, 124.19, 120.75, 119.63, 116.67, 115.96, 109.76, 48.51, 32.09, 21.01. HRMS (ESI-TOF) calcd for $C_{10}H_{13}N_2Cl$ $[M + H]^+$: 209.0846 and 211.0818, found: 209.0808 and 211.0782.

7-chlorotryptophol (5): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite and extracted into CH_2Cl_2. Purification by flash column chromatography (SiO_2, 30% ethyl acetate/hexanes) afforded **5** in 95% yield (11.5 mg, 0.059 mmol). ^{1}H NMR (500 MHz, Chloroform-d) δ 8.28 (s, 1H), 7.53 (d, J = 7.9 Hz, 1H), 7.21 (d, J = 7.6 Hz, 1H), 7.15 (s, 1H), 7.07 (t, J = 7.8 Hz, 1H), 5.3 (s, 1H), 3.91 (t, J = 6.4 Hz, 2H), 3.03 (t, J = 6.4 Hz, 2H). ^{13}C NMR (126 MHz, Chloroform-d) δ 133.66, 128.91, 123.03, 121.54, 120.26, 117.51, 116.72, 113.64, 62.61, 28.79. HRMS (ESI-TOF) calcd for $C_{10}H_{11}NOCl$ $[M + H]^+$: 196.0529 and 198.0502, found: 196.0489 and 198.0465.

7-chloro-2-methyltryptamine (6): The bioconversion was conducted according to the general procedure in a 100 x 50 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by

HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2 and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-NH). The crude material was purified by normal phase chromatography (gradient from 100% CH_2Cl_2 to 5% MeOH/ CH_2Cl_2) to afford **6** in 60% yield (7.2 mg, 0.034 mmol). ^{1}H NMR (500 MHz, Acetonitrile-d3) δ 9.25 (s, 1H), 7.42 (d, J = 7.8 Hz, 1H), 7.06 (d, J = 8.3 Hz, 1H), 6.96 (t, J = 7.7 Hz, 1H), 2.81 – 2.72 (m, 4H), 2.39 (s, 3H). ^{13}C NMR (126 MHz, DMSO-d6) δ 133.89, 131.93, 130.39, 119.29, 119.08, 116.47, 114.88, 109.65, 42.71, 28.28, 11.26. HRMS (ESI-TOF) calcd for $C_{11}H_{14}N_2Cl$ $[M + H]^+$: 209.0846 and 211.0818, found: 209.0788 and 211.0762.

6-chloro-5-methyltryptamine (7): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After the reaction was quenched with acid as described above, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP- C18-HS). The crude material was purified by reverse phase chromatography (gradient from water 0.1% TFA to 25% CH_3CN/water 0.1%TFA) to afford **7** in 58% yield (10.7 mg of **7**·TFA, 0.033 mmol). ^{1}H NMR (500 MHz, DMSO-d6) δ 11.00 (s, 1H), 7.87 (s, 3H), 7.49 (s, 1H), 7.41 (s, 1H), 7.23 (s, 1H), 3.06 (s, 2H), 2.98 – 2.90 (m, 2H), 2.39 (s, 3H). ^{13}C NMR (126 MHz, MeOD–d4) δ 137.36, 129.48, 127.41, 127.17, 125.24, 120.23, 112.63, 110.06, 41.14, 24.39, 20.52. HRMS (ESI-TOF) calcd for $C_{11}H_{13}N_2Cl$ $[M + H]^+$: 209.0846 and 211.0818, found: 208.9847 and 210.9836.

5-chloro-tryptoline (8-a) and 6-chloro-tryptoline (8-b): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After maximum conversion to monohalogenated products was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP- C18-HS). The crude material was purified by reverse phase chromatography (gradient from water 0.1% TFA to 30% CH_3CN/water 0.1%TFA) to afford a mixture of **8-a** and **8-b** in 54% yield (10.1 mg of **8-a**·TFA and **8-b**·TFA, 0.031 mmol). ^{1}H NMR (500 MHz, Methanol-d4) δ 7.46 (d, J = 1.9 Hz, 0.56H), 7.43 (d, J = 8.4 Hz, 0.39H), 7.35 (d, J = 1.8 Hz, 0.37H), 7.31 (d, J = 8.6 Hz, 0.56H), 7.10 (dd, J = 8.6, 2.0 Hz, 0.57H), 7.03 (dd, J = 8.4, 1.8 Hz, 0.39H), 4.43 (m, 2H), 3.57 (t, J = 6.1 Hz, 2H), 3.05 (m, 2H). ^{13}C NMR (126 MHz, Methanol-d4) δ 138.55, 136.56, 129.17, 128.60, 128.28, 127.52, 126.26, 126.21, 123.44, 121.04, 120.02, 118.47, 113.51, 112.18, 107.29, 106.86, 43.59, 42.11, 42.08, 19.32, 19.30. HRMS (ESI-TOF) calcd for $C_{11}H_{12}N_2Cl$ $[M + H]^+$: 207.0689 and 209.0662, found: 207.0633 and 209.0610.

3-chloroindole[45] **(9):** The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite and extracted into CH_2Cl_2. Purification by flash column chromatography (SiO_2, 5% ethyl acetate/hexanes) afforded **9** in 53% yield (6.9 mg, 0.046 mmol). ^{1}H NMR (500 MHz, Methylene Chloride-d2) δ 8.25 (s, 1H), 7.61 (d, J = 7.9 Hz, 1H), 7.41 (d, J = 8.2 Hz, 1H), 7.25 (t, J = 8.2 Hz, 1H), 7.22 (d, J = 2.6 Hz, 1H), 7.19 (t, J = 8.0 Hz, 1H). ^{13}C

NMR (126 MHz, Methylene Chloride-d2) δ 135.38, 125.61, 123.42, 121.43, 120.76, 118.28, 111.92, 106.40. GCMS (EI) calcd for C_8H_6NCl [M]: 151.0 and 153.0, found: 151.0 and 153.0.

1-chloro-2-aminonaphthalene (10): The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite and extracted into CH_2Cl_2. Purification by flash column chromatography (SiO_2, 1% ethyl acetate/hexanes) afforded **10** in 93% yield (11.5 mg, 0.065 mmol). ^{1}H NMR (500 MHz, Methylene Chloride-d2) δ 8.03 (d, J = 8.5 Hz, 1H), 7.72 (d, J = 8.1 Hz, 1H), 7.61 (d, J = 8.7 Hz, 1H), 7.51 (t, J = 8.3 Hz, 1H), 7.30 (t, J = 8.1 Hz, 1H), 7.04 (d, J = 8.7 Hz, 1H), 4.37 (s, 3H). ^{13}C NMR (126 MHz, Methylene Chloride-d2) δ 141.18, 132.13, 128.76, 128.43, 128.11, 127.78, 123.13, 122.38, 118.15, 111.38. HRMS (ESITOF) calcd for $C_{10}H_9NCl$ $[M + H]^+$: 178.0424 and 180.0396, found: 178.0376 and 180.0350.

2-chloronaphthol[46] **(11):** The bioconversion was conducted according to the general procedure in a 125 x 65 mm crystallization dish. Substrate was added in a solution of isopropanol (5% v/v). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite and extracted into CH_2Cl_2. Purification by flash column chromatography (SiO_2, hexanes) afforded **11** in 59% yield (7.3 mg, 0.041 mmol). ^{1}H NMR (500 MHz, Methylene Chloride-d2) δ 8.24 – 8.16 (m, 1H), 7.80-7.82 (m, 1H), 7.57 – 7.48 (m, 2H), 7.44 – 7.35 (m, 2H), 6.10 (s, 2H). ^{13}C NMR (126 MHz, Methylene Chloride-d2) δ 133.63, 127.97, 127.05, 126.48, 126.28, 124.76, 122.21, 121.30. GCMS (EI) calcd for $C_{10}H_7OCl$ [M]: 178.0 and 180.0, found: 178.0 and 180.0.

7-bromo-L-tryptophan (12): The bioconversion was conducted according to the general procedure in a 100 x 50 mm crystallization dish. Substrate was added in a solution of HEPES buffer (25 mM, pH 7.4). After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, submitted to strong cation exchange chromatography, and dry-loaded onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography (gradient from 100% H_2O to 15% CH_3CN/H_2O) to afford **12** in 85% yield (11.8 mg, 0.042 mmol). ^{1}H NMR (500 MHz, Deuterium Oxide) δ 7.54 (d, J = 7.9 Hz, 1H), 7.28 (d, J = 7.6 Hz, 1H), 7.19 (s, 1H), 6.93 (t, J = 7.8 Hz, 1H), 3.80 (dd, J = 7.8, 4.8 Hz, 1H), 3.27-3.23 (m, 1H), 3.11-3.07 (m, 1H). ^{13}C NMR (126 MHz, Deuterium Oxide) δ 173.23, 136.44, 129.78, 127.50, 126.23, 122.48, 119.33, 109.55, 106.14, 54.81, 27.45. HRMS (ESI-TOF) calcd for $C_{11}H_{12}N_2O_2Br$ $[M + H]^+$: 283.0082 and 285.0063, found: 283.0017 and 285.0015.

2.4.4 Detailed isolation and characterization II

Detailed isolation and characterization for compounds in section 2.2.3:

7-chloro-L-tryptophan (1): The 10 mg bioconversion was conducted according to the general procedure. Substrate was added in a solution of HEPES (25 mM HEPES, pH = 7.4). RebH was added to a final concentration of 2 μM (0.004 equiv.) and the reaction was incubated at 40°C. After reaction completion, the reaction mixture was filtered through Celite, submitted to strong cation exchange chromatography, and dry-loaded onto Celite. This was packed into a Biotage samplet, loaded into a reverse phase column (Biotage SNAP-KP- C18-HS) and purified by

reverse phase chromatography (gradient from water 0.1% TFA to 15% CH_3CN/water 0.1%TFA) to afford **1** in 69% yield (11.9 mg of **1**·TFA, 0.034 mmol). ^{1}HNMR spectrum was consistent with previous reports of this compound.[23] ^{1}H NMR (500 MHz, Deuterium Oxide with one drop TFA). δ 7.35 (d, J = 5.7 Hz, 1H), 7.13 (s, 1H), 7.04 (d, J = 7.6 Hz, 1H), 6.88 (t, J = 7.8 Hz, 1H), 4.15 (t, J = 4.8 Hz, 1H), 3.23 (m, 2H).

7-chloro-2-methyltryptamine (2): The 10 mg bioconversion was conducted according to the general procedure. Substrate was added in a solution of isopropanol (5% v/v). RebH was added to a final concentration of 5 μM (0.01 equiv.) and the reaction was incubated at 40°C. After reaction completion, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dry-loaded onto Celite. This was packed into a Biotage samplet, loaded into a reverse phase column (Biotage SNAP-KP- C18-HS) and purified by reverse phase chromatography (gradient from water 0.1% TFA to 15% CH_3CN/water 0.1% TFA) to afford **2** in 56% yield (10.4 mg of **2**·TFA, 0.032 mmol). ^{1}HNMR spectrum was consistent with previous reports of this compound.[23] ^{1}H NMR (500 MHz, Methanol-d4). δ 7.40 (d, J = 7.8 Hz, 1H), 7.06 (d, J = 7.6 Hz, 1H), 6.98 (t, J = 7.7 Hz, 1H), 3.16-3.10 (m, 2H), 3.07-3.02 (m, 2H), 2.43 (s, 3H).

1-chloro-2-aminonaphthalene (3): The 10 mg bioconversion was conducted according to the general procedure. Substrate was added in a solution of isopropanol (5% v/v). RebH was added to a final concentration of 4 μM (0.008 equiv.) and the reaction was incubated at 21°C. After reaction completion, the reaction mixture was filtered through Celite and extracted into CH_2Cl_2. The crude material was purified by flash column chromatography (SiO_2, 1% ethyl acetate/hexanes) to afford **3** in 62% yield (7.7 mg of **3**, 0.043 mmol). ^{1}HNMR spectrum was

consistent with previous reports of this compound.[23] ^{1}H NMR (500 MHz, Methylene Chloride-d2). δ 8.00 (d, J = 8.5 Hz, 1H), 7.70 (d, J = 8.1 Hz, 1H), 7.59 (d, J = 8.7 Hz, 1H), 7.49 (t, J = 7.7 Hz, 1H), 7.27 (t, J = 8.0 Hz, 1H), 7.02 (d, J = 8.7 Hz, 1H), 3.58 (s, 3H).

5-chloro-tryptoline and 6-chloro-tryptoline (4): The 10 mg bioconversion was conducted according to the general procedure. Substrate was added in a solution of isopropanol (5% v/v). RebH was added to a final concentration of 50 μM (0.1 equiv.) and the reaction was incubated at 21°C. After reaction completion, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dry-loaded onto Celite. This was packed into a Biotage samplet, loaded into a reverse phase column (Biotage SNAP-KP- C18-HS) and purified by reverse phase chromatography (gradient from water 0.1% TFA to 30% CH_3CN/water 0.1%TFA) to afford **4** in 67% yield (12.5 mg of **4**·TFA, 0.039 mmol). ^{1}HNMR spectrum was consistent with previous reports of this compound.[23] ^{1}H NMR (500 MHz, Methanol-d4). δ 7.47 (s, 0.80H), 7.44 (d, J = 8.4 Hz, 0.12H), 7.36 (s, 0.10H), 7.32 (d, J = 8.6 Hz, 0.81H), 7.11 (dd, J = 8.6, 2.0 Hz, 0.80H), 7.04 (dd, J = 8.4, 1.8 Hz, 0.10H), 4.44 (s, 2H), 3.58 (t, J = 6.1 Hz, 2H), 3.05 (t, J = 6.0 Hz, 2H).

2.4.5 Additional data

Hanes-Woolf plots for kinetic values in Table 2.2:

Figure 2.7: Hanes-Woolf plot for RebH conversion of L-tryptophan.

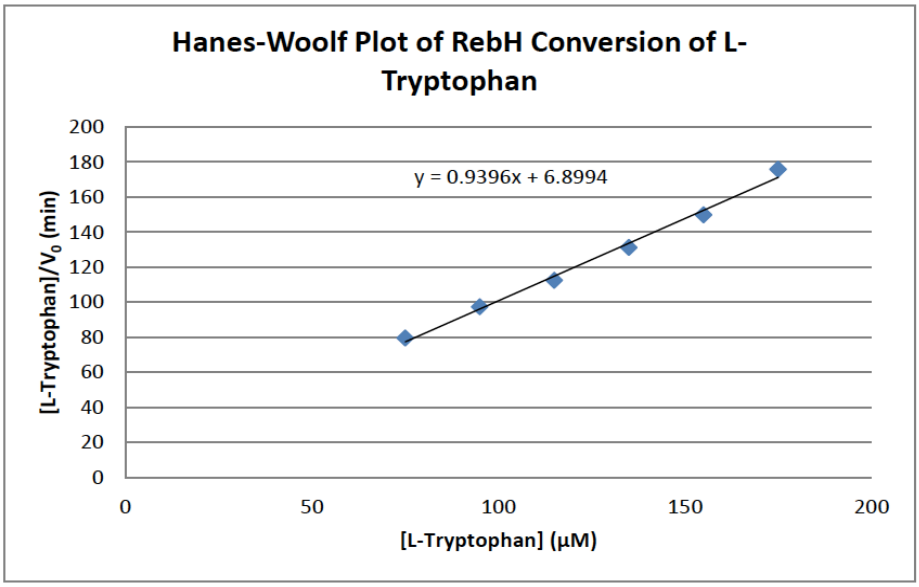


Figure 2.8: Hanes-Woolf plot for RebH conversion of tryptamine.

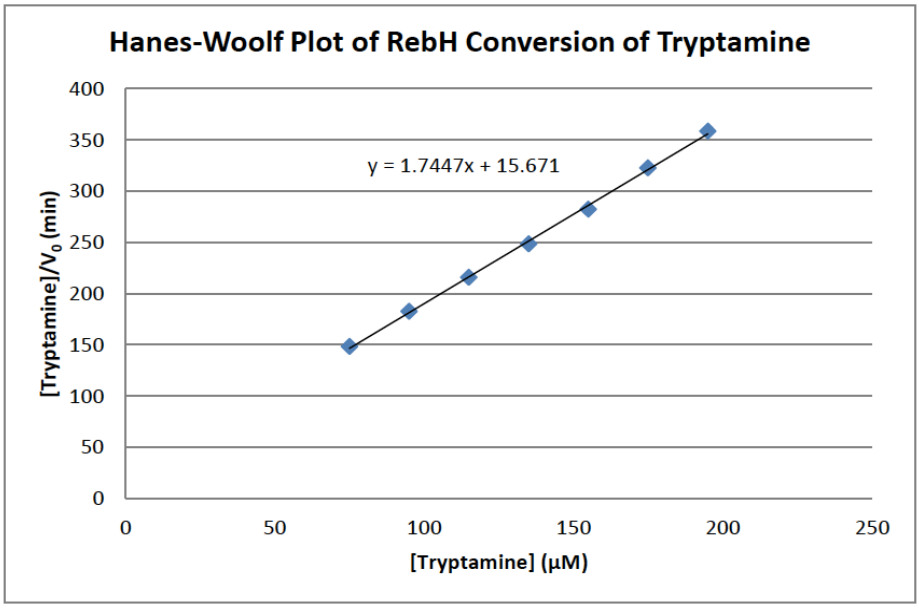


Figure 2.9: Hanes-Woolf plot for RebH conversion of 2-aminonaphthalene.

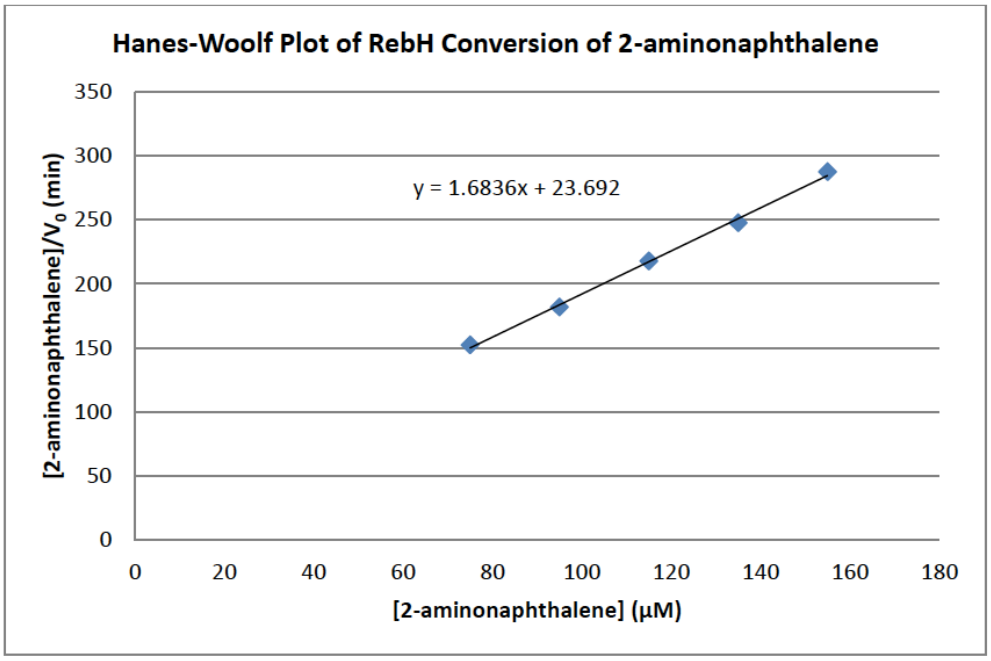
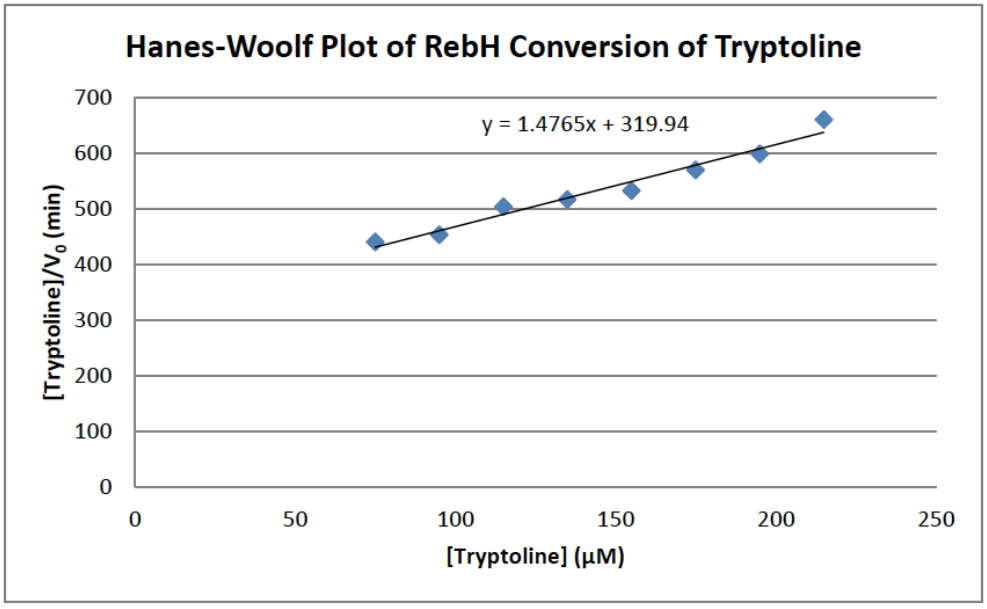


Figure 2.10: Hanes-Woolf plot for RebH conversion of tryptoline.



Hanes-Woolf plots for kinetic values in Table 2.4:

Figure 2.11: Hanes-Woolf plot for RebH conversion of 2-methyltryptamine at 40 °C.

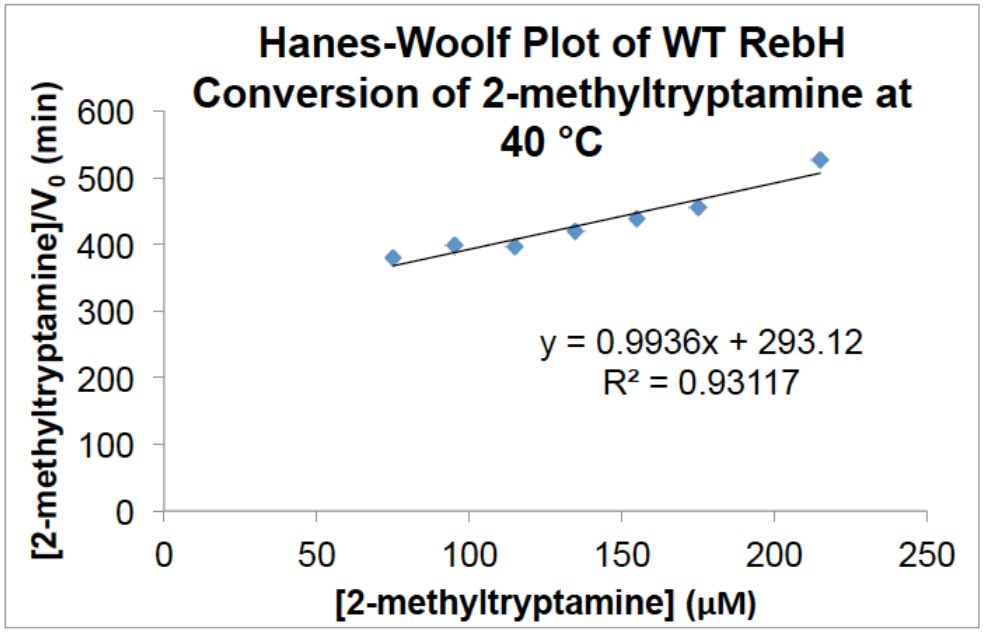


Figure 2.12: Hanes-Woolf plot for 3LSR conversion of 2-methyltryptamine at 40 °C.

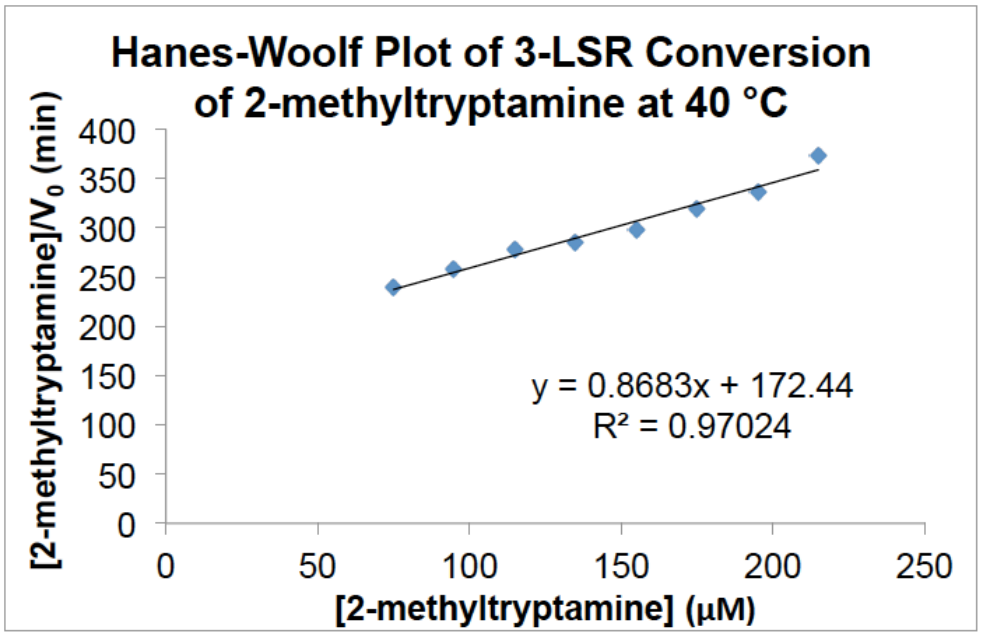


Figure 2.13: Hanes-Woolf plot for RebH conversion of 2-methyltryptamine at 21 °C.

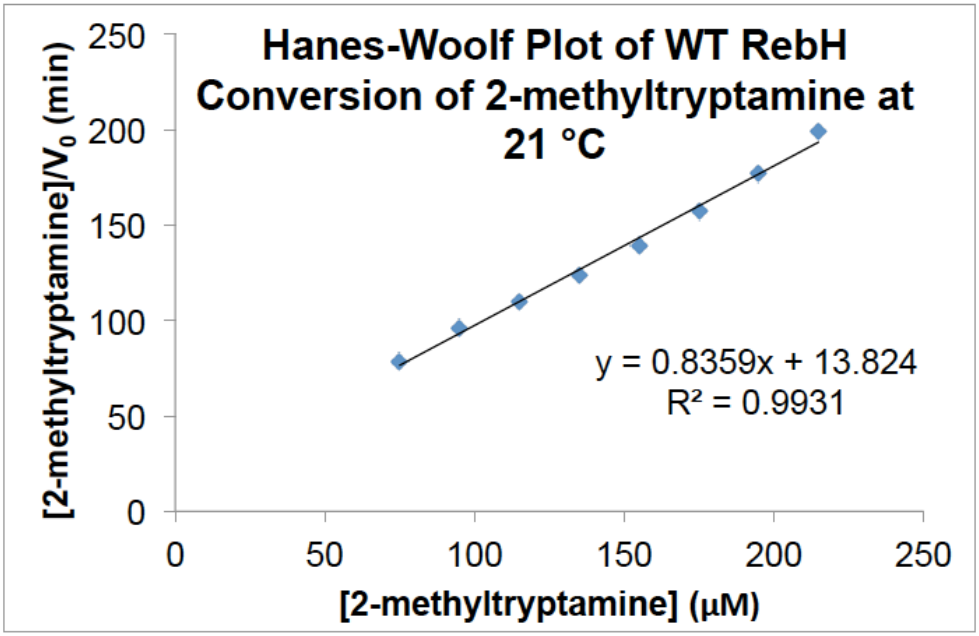
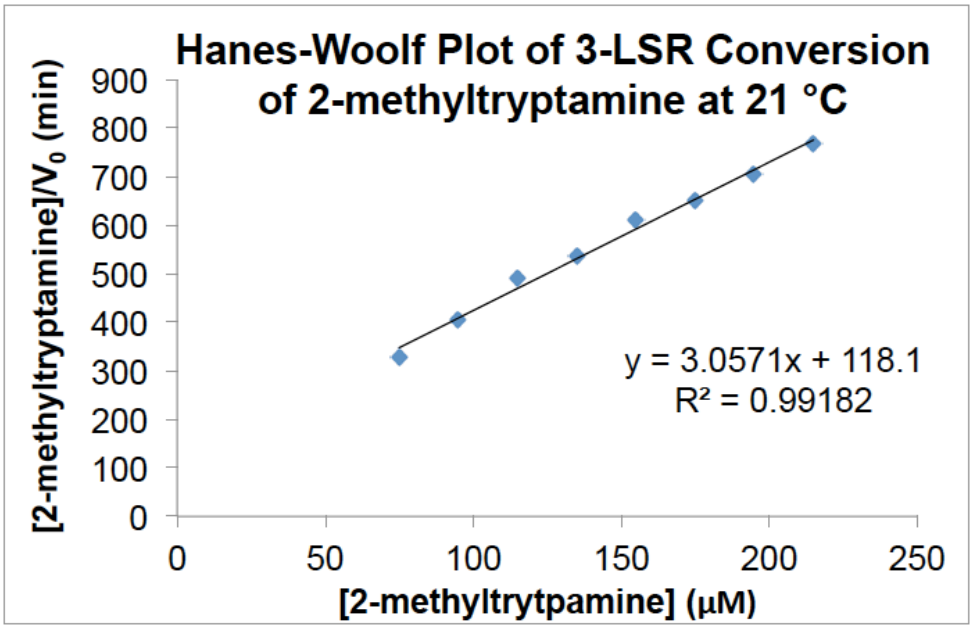


Figure 2.14: Hanes-Woolf plot for 3LSR conversion of 2-methyltryptamine at 21 °C.



Acknowledgements

We would like to thank Dr. Landon Durak for advice on NMR experiments. We would also like to thank the Walsh group for RebH and RebF plasmids and the Bottomley group for fusion plasmids.

2.5 References

1. Jeschke, P. The unique role of halogen substituents in the design of modern agrochemicals. *Pest. Manag. Sci.* **66,** 10–27 (2009).
2. Hernandes, M. Z., Cavalcanti, S. M. T., Moreira, D. R. M., de Azevedo Junior, W. F. & Leite, A. C. L. Halogen atoms in the modern medicinal chemistry: hints for the drug design. *Curr Drug Targets* **11,** 303–314 (2010).
3. Naumann, K. Influence of chlorine substituents on biological activity of chemicals: a review. *Pest. Manag. Sci.* **56,** 3–21 (2000).
4. Herrera-Rodriguez, L. N., Khan, F., Robins, K. T. & Meyer, H.-P. Perspectives on biotechnological halogenation Part I: Halogenated products and enzymatic halogenation. *Chimica Oggi-Chemistry Today* **29,** 31–33 (2011).
5. Harris, C. M., Kannan, R. & Kopecka, H. The role of the chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono-and didechlorovancomycin. *J. Am. Chem. Soc.* **107,** 6652–6658 (1985).
6. Bunders, C. A. *et al.* Intercepting Bacterial Indole Signaling with Flustramine Derivatives. *J. Am. Chem. Soc.* **133,** 20160–20163 (2011).
7. Smith, B. M. *et al.* Discovery and Structure–Activity Relationship of (1 R)-8-Chloro-2,3,4,5-tetrahydro-1-methyl-1 H-3-benzazepine (Lorcaserin), a Selective Serotonin 5-HT 2CReceptor Agonist for the Treatment of Obesity. *J. Med. Chem.* **51,** 305–313 (2008).
8. *Metal-catalyzed cross-coupling reactions*. (Wiley-VCH, 1998). doi:10.1002/9783527612222
9. Smith, K. & El-Hiti, G. A. Regioselective control of electrophilic aromatic substitution reactions. *Current Organic Synthesis* **1,** 253–274 (2004).
10. Podgoršek, A., Zupan, M. & Iskra, J. Oxidative Halogenation with 'Green' Oxidants: Oxygen and Hydrogen Peroxide. *Angew. Chem. Int. Ed.* **48,** 8424–8450 (2009).
11. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneau-Tsodikova, S. & Walsh, C. T. Nature's Inventory of Halogenation Catalysts: Oxidative Strategies Predominate. *Chem. Rev.* **106,** 3364–3378 (2006).
12. Zehner, S. *et al.* A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in Streptomyces rugosporus LL-42D005. *Chem. Biol.* **12,** 445–452 (2005).
13. Seibold, C. *et al.* A flavin-dependent tryptophan 6-halogenase and its use in modification of pyrrolnitrin biosynthesis. *Biocatalysis and Biotransformation* **24,** 401–408 (2006).
14. Keller, S. *et al.* Purification and Partial Characterization of Tryptophan 7-Halogenase (PrnA) from Pseudomonas fluorescens. *Angew. Chem. Int. Ed. Engl.* **39,** 2300–2302 (2000).

15. Yeh, E., Garneau, S. & Walsh, C. T. Robust in vitro Activity of RebF and RebH, a Two-Component Reductase/Halogenase, Generating 7-Chlorotryptophan during Rebeccamycin Biosynthesis. *Proc. Natl. Acad. Sci.* **102,** 3960–3965 (2005).
16. Flecks, S. *et al.* New Insights into the Mechanism of Enzymatic Chlorination of Tryptophan. *Angew. Chem. Int. Ed.* **47,** 9533–9536 (2008).
17. Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases †,‖. *Biochemistry* **46,** 1284–1292 (2007).
18. Yeh, E. *et al.* Flavin Redox Chemistry Precedes Substrate Chlorination during the Reaction of the Flavin-Dependent Halogenase RebH †. *Biochemistry* **45,** 7904–7912 (2006).
19. Bitto, E. *et al.* The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins* **70,** 289–293 (2007).
20. Dong, C. J. *et al.* Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* **309,** 2216–2219 (2005).
21. Hölzer, M., Burd, W., Reißig, H.-U. & van Pée, K.-H. Substrate Specificity and Regioselectivity of Tryptophan 7-Halogenase from Pseudomonas fluorescens BL915. *Adv. Synth. Catal.* **343,** 591–595 (2001).
22. Glenn, W. S., Nims, E. & O'Connor, S. E. Reengineering a Tryptophan Halogenase To Preferentially Chlorinate a Direct Alkaloid Precursor. *J. Am. Chem. Soc.* **133,** 19346–19349 (2011).
23. Payne, J. T., Andorfer, M. C. & Lewis, J. C. Regioselective Arene Halogenation using the FAD-Dependent Halogenase RebH. *Angew. Chem. Int. Ed.* **52,** 5271–5274 (2013).
24. Poor, C. B., Andorfer, M. C. & Lewis, J. C. Improving the Stability and Catalyst Lifetime of the Halogenase RebH By Directed Evolution. *ChemBioChem* **15,** 1286–1289 (2014).
25. Esposito, D. & Chatterjee, D. K. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17,** 353–358 (2006).
26. Sørensen, H. P. & Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microb. Cell Fact.* **4,** 1 (2005).
27. Smith, D. R. M. *et al.* An unusual flavin-dependent halogenase from the metagenome of the marine sponge Theonella swinhoei WA. *ACS Chem. Biol.* acschembio.6b01115–12 (2017). doi:10.1021/acschembio.6b01115
28. Andorfer, M. C., Grob, J. E., Hajdin, C. E. & Chael, J. R. Understanding Flavin-Dependent Halogenase Reactivity via Substrate Activity Profiling. *ACS Catal.* 1897–1904 (2017). doi:10.1021/acscatal.6b02707
29. Wu, I. & Arnold, F. H. Engineered thermostable fungal Cel6A and Cel7A cellobiohydrolases hydrolyze cellulose efficiently at elevated temperatures. *Biotechnol. Bioeng.* **110,** 1874–1883 (2013).
30. Zhao, H. & Arnold, F. H. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng. Des. Sel.* **12,** 47–53 (1999).
31. Pakula, A. Genetic Analysis Of Protein Stability And Function. *Annual Review of Genetics* **23,** 289–310 (1989).
32. Zavodszky, P., Kardos, J., Svingor, A. & Petsko, G. A. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci.* **95,** 7406–7411 (1998).

33. Wolf-Watz, M. *et al*. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol* **11,** 945–949 (2004).
34. Pei, X.-Q., Yi, Z.-L., Tang, C.-G. & Wu, Z.-L. Three amino acid changes contribute markedly to the thermostability of Î²-glucosidase BglC from Thermobifida fusca. *Bioresource Technology* **102,** 3337–3342 (2011).
35. McDaniel, A., Fuchs, E., Liu, Y. & Ford, C. Directed evolution of Aspergillus nigerglucoamylase to increase thermostability. *Microbial Biotechnology* **1,** 523–531 (2008).
36. Zhang, D. *et al*. Gradually accumulating beneficial mutations to improve the thermostability of N-carbamoyl-d-amino acid amidohydrolase by step-wise evolution. *Appl Microbiol Biotechnol* **90,** 1361–1371 (2011).
37. Wang, Y. *et al*. Improvement ofAspergillus niger Glucoamylase Thermostability by Directed Evolution. *Starch - Stärke* **58,** 501–508 (2006).
38. Kumar, R., Sharma, M., Singh, R. & Kaur, J. Characterization and evolution of a metagenome-derived lipase towards enhanced enzyme activity and thermostability. *Mol Cell Biochem* **373,** 149–159 (2012).
39. Maté, D. *et al*. Laboratory Evolution of High-Redox Potential Laccases. *Chem. Biol.* **17,** 1030–1041 (2010).
40. Menon, B. R. K. *et al*. Structure and biocatalytic scope of thermophilic flavin-dependent halogenase and flavin reductase enzymes. *Org. Biomol. Chem.* **14,** 9354–9361 (2016).
41. Cabrita, L. D., Dai, W. & Bottomley, S. P. A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol* **6,** 12 (2006).
42. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning: a laboratory manual, 2nd edn*. (Cold Spring Laboratory Press, 1989).
43. Gottlieb, H. E., Kotlyar, V. & Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *J. Org. Chem.* **62,** 7512–7515 (1997).
44. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2,** 924–932 (2007).
45. Wagger, J., Grošelj, U., Svete, J. & Stanovnik, B. Synthesis of Racemic, N-Benzylated Neoechinulin A and Isoneoechinulin A[1]. *Synlett* **2010,** 1197–1200 (2010).
46. Ding, Z., Xue, S. & Wulff, W. D. A Succinct Synthesis of the Vaulted Biaryl Ligand Vanol via a Dienone-Phenol Rearrangement. *Chem. Asian J.* **6,** 2130–2146 (2011).

Chapter 3: Developing mutagenesis and screening methods for directed evolution of FDHs

3.1 Introduction

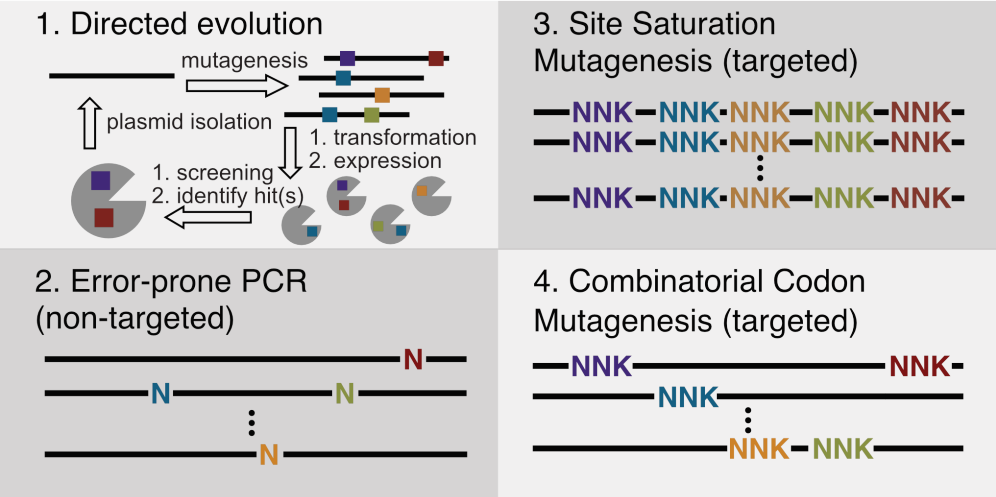
Directed evolution is a powerful approach to protein engineering whereby enzyme function is systematically improved via iterative rounds of mutagenesis and screening (Fig. 3.1-1).[1] A successful directed evolution campaign requires methods to create sufficient diversity in an enzyme library (mutagenesis) and efficient methods for identifying improved enzyme variants (screen).[2] The first directed evolution of an FDH, described in Chapter 2, used random mutagenesis to create diverse libraries and UPLC conversions of starting material to product to identify variants that were more active than the parent enzyme for each round.[3] These methods are robust, general, and frequently used for directed evolution efforts. Indeed, Dr. James Payne and Dr. Catherine Poor used these mutagenesis and screening methods to expand the substrate scope of RebH to larger, biologically active substrates.[4] Although random mutagenesis and screening by UPLC have emerged as the workhorses of directed evolution efforts in the Lewis lab, exploring alternatives can enable more direct and/or efficient library generation and analysis.

Mutagenesis methods

Many mutagenesis strategies have been used in directed evolution efforts. Error prone PCR, involving incorporation of random mutations into a gene of interest by amplifying the gene in the presence of $MnCl_2$ (Fig. 3.1-2), remains one of the most common and simplest of these methods.[2,5] Thus, a diverse set of genes can be rapidly created without any prior structural knowledge; however, the low probability of mutating contiguous base pairs leads to codon bias,

stop codons can be introduced, and many enzymes within the library contain no mutations.[5] In contrast to error prone PCR, targeted mutagenesis can be employed if structural information is available. Ideally, targeted approaches reduce screening effort while still improving function; however, predicting the impact of specific mutations on an enzyme's function remains very challenging. To partially overcome this difficulty, degenerate codons can be introduced to examine the effect of several mutations at a certain site (Fig. 3.1-3).[6,7] Both single and multiple site saturation mutagenesis can be accomplished by splicing by overlap extension (SOE) PCR[8] using oligonucleotides containing degenerate codons or mixtures of codons at targeted positions.

Figure 3.1: Overview of directed evolution (1) and common mutagenesis methods (2-4).



In addition to the challenge of selecting appropriate residues, several technical issues arise when using saturation methods. These typically require a low number of sites to be simultaneously targeted, as higher mutation frequencies generally lead to unfolded proteins. Thus, in order to target a desired region of the enzyme (i.e. active site), small subsets of residue are mutated in separate libraries.[9] With this method, beneficial epistasis between mutations in different subsets will be missed and cloning of these libraries is tedious and repetitive.

Combinatorial codon mutagenesis (CCM) methods can mitigate challenges faced by other mutagenesis methods. CCM can be used to replace, with tunable frequency, each codon within a set of targeted sites with a desired codon (Fig. 3.1-4). CCM methods have previously been described, but these are often technically difficult and/or require specialized reagents (i.e. biotinylated primers, phosphorylated primers, thermostable ligase, etc.).[10-13] Standard SOE PCR has been used to encode mixtures of wild-type and degenerate codons; however, cloning becomes increasingly tedious as more sites are targeted.[14,15] Recently, Bloom described a CCM method that is only a slight variation of SOE PCR.[16] In this method, standard oligonucleotides containing the degenerate NNN codon were used to create random variants of influenza nucleoprotein for experimentally determined evolutionary models.[16] We envisioned that a modified version of this CCM method could be used to merge the best features of random mutagenesis and site saturation mutagenesis as a tool for protein engineering efforts.

Screening methods

Once high quality libraries of variants have been created through mutagenesis of the parent gene, variants with improved function must be identified. It is often said of directed evolution that "you get what you screen for," indicating the assay used to analysis variants is crucial to success.[17] The more relevant the screen is to the desired function, the more likely the evolution campaign will be successful and efficient. Often, screening directly for desired function is at odds with practical time limitations. For example, screening FDHs for improved conversion to product by UPLC can be used to identify more active catalysts; however, UPLC methods can be lengthy. This narrows the practical limit for how many variants can be screened each round.

Colorimetric assays to determine enzyme function are simple and rapid, and thus decrease the amount of time spent screening.[18,19] To date, directed evolution of FDHs using a colorimetric screen has not been reported. Recently, however, Micklefield and coworkers have reported a colorimetric assay for analyzing FDH halogenation of anilines.[20] In this method, 4-methyl-1,2-benzoquinone and aniline substrates are coupled through a Michael addition. These compounds are highly colored and, in several cases, significant differences are observed between the UV/Vis spectra of their chlorinated analogues. Micklefield and coworkers demonstrate that chlorination of these anilines can be quantitated through spectrophotometric analysis.[20] Although this approach could prove valuable for screening larger libraries of FDHs, it has only been established for reactions using purified enzyme. In order for this to be a practical method for directed evolution of FDHs, it will have to function using products in cell lysates.

While colorimetric assays could substantially decrease screening time as compared to UPLC, neither screen can readily be used to directly observe altered site selectivity. Chlorinated isomers often elute closely by chromatography, thus methods capable of separating them are typically lengthy. In some cases, as with 6- and 7-chlorotryptopline, chlorinated isomers cannot be separated using standard methods and conditions.

To improve efficiency and speed of FDH evolution campaigns, we sought to design a simple, general CCM method and assess compatibility of the aforementioned colorimetric screen with cell extracts. We also aimed to develop a screen that could directly assess site selectivity of chlorination reactions. Such a screen could allow the directed evolution of FDHs for altered site selectivities. Papers describing the majority of the work in sections 3.2.1 and 3.2.3 have been published.[21,22]

Authorship

The work described in section 3.2.1 was performed with Dr. Ketaki Belsare. Experiments with $P450_{BM3}$ and prolyloligopeptidase (POP) were conducted by Dr. Belsare, aided by an undergraduate student in the Lewis lab, Frida Cardenas. Julia Chael, another undergraduate student in the Lewis lab, helped in the characterization of RebH variants obtained using our CCM method. Initial rate comparisons between d_5-L-tryptophan and L-tryptophan in section 3.2.3 were conducted by an undergraduate in the Lewis lab, Jaylie Vergara-Coll.

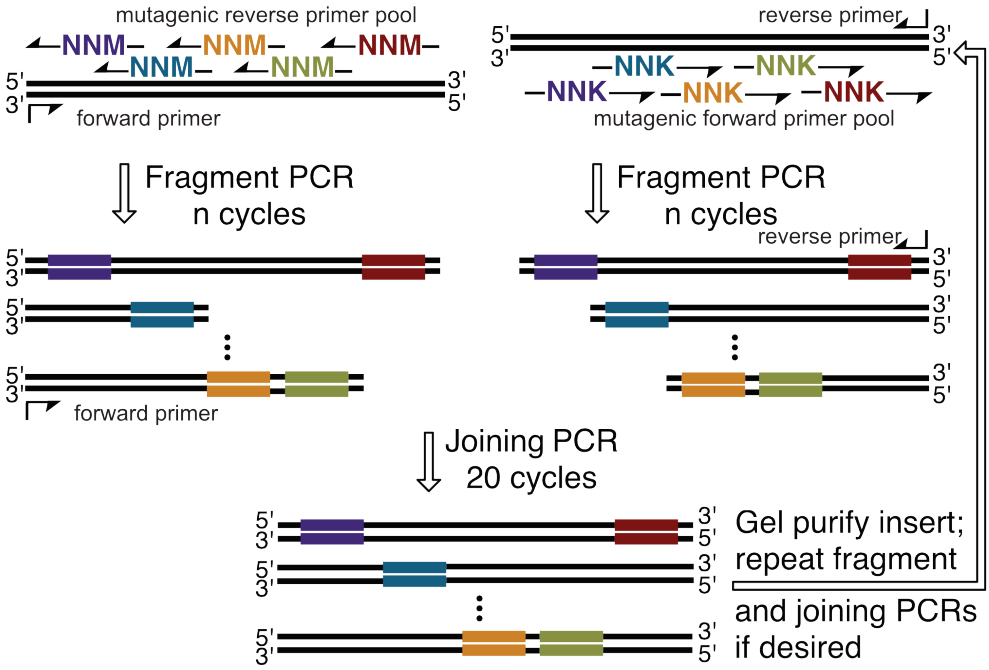
3.2 Results and Discussion

3.2.1 A Simple, combinatorial codon mutagenesis method for targeted protein engineering

As described in section 3.1, a recent report by Bloom and coworkers detailed a combinatorial codon mutagenesis method (CCM) to create random mutants of influenza nucleoprotein for experimentally determined evolutionary models.[16] This method is similar to SOE PCR,[8] but instead of using a single mutagenic oligonucleotide to create gene fragments, pools of short mutagenic oligonucleotides (21-27 nucleotides) are used during the fragment PCR (Fig. 3.2). After many diverse fragments have been created, they can be assembled to form the full length gene in a joining PCR (Fig. 3.2). After the joining PCR, the full-length gene can be gel purified and ligated into vector for transformation. In this way, a targeted set of codons within a gene sequence can be combinatorially mutated (Fig. 3.1-4). This method could serve as a powerful library design approach for protein engineering efforts, especially if the mutation frequency was readily tunable. Mutation frequency should vary with number of fragmentation PCR cycles as well as the number of rounds of fragmentation and joining PCRs conducted. In

this work, we aimed to demonstrate that mutation frequency was readily tunable through altering these two parameters and to apply this method to several protein engineering efforts.

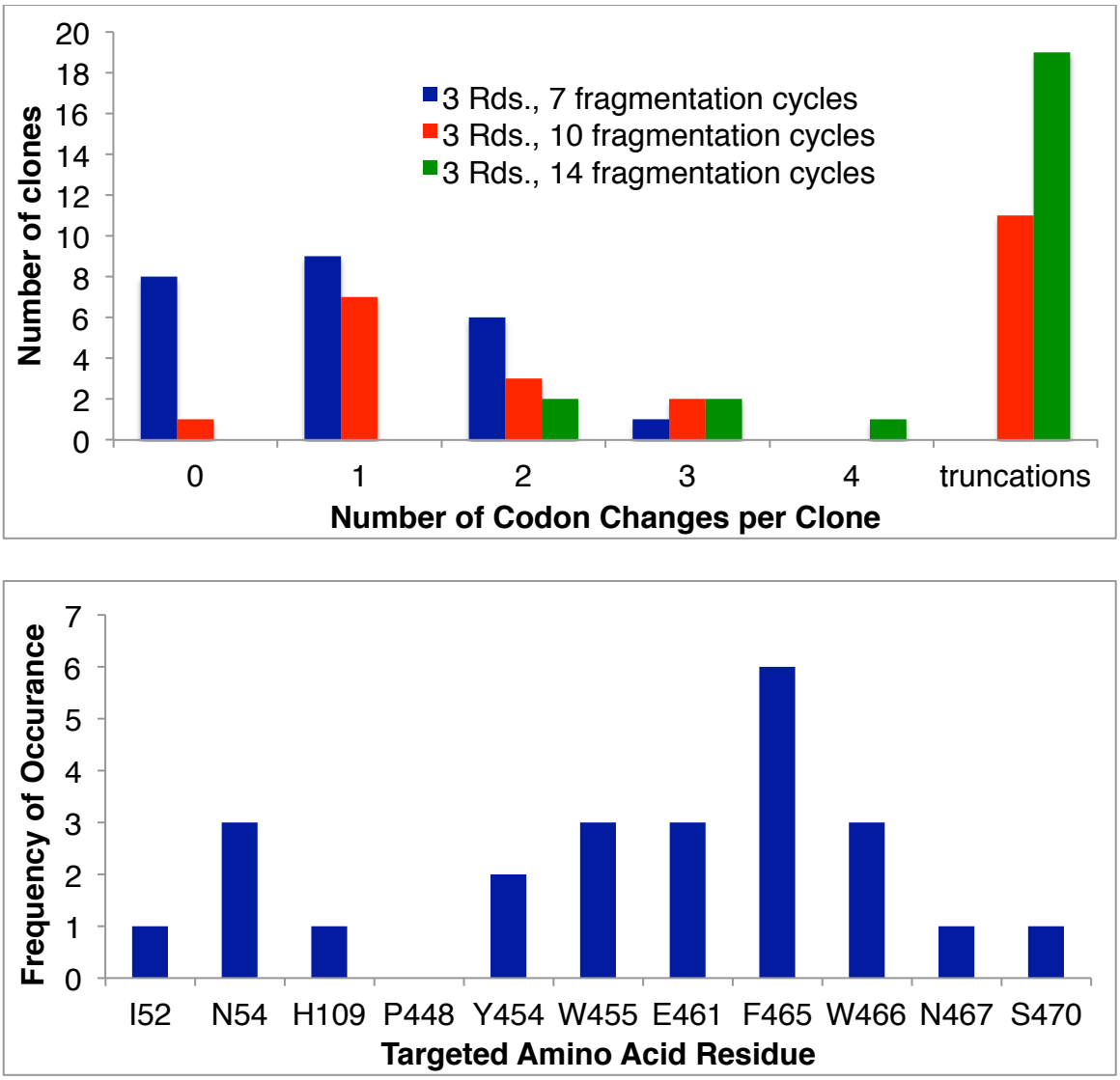
Figure 3.2: Overview of CCM method.



I first began exploring the method using the RebH gene as a template. Eleven amino acid residues, which, based on results discussed in chapter 4, were thought to impact RebH selectivity, were targeted and primers containing the degenerate NNN codon for each of these were designed. The number of rounds of fragmentation and joining PCRs was tested (3 and 6 rounds) as well as the number of fragmentation cycles per round (7, 10, and 14 fragmentation cycles). Although higher mutation frequency was observed when more rounds were performed (data not shown) and when more fragmentation cycles were used (Fig. 3.3), a substantial amount of genes showed truncations. Most of these truncations were caused by the global reverse primer non-specifically binding to the nucleotides encoding for residues near 465-467. All three of these residues were targeted in this study, and when mutated, the global reverse primer bound to this region often. In an attempt to remedy this problem, I used different global reverse primers,

explored various additives, and tested different PCR conditions. While some of these changes did improve the ratio of full gene to truncated genes, many truncations were always observed at higher mutation frequency. In addition to gene truncations, biases in the distribution of mutations were observed (Fig. 3.3). For example, residue F465 was mutated more often than other residues, while residue P448 was rarely mutated.

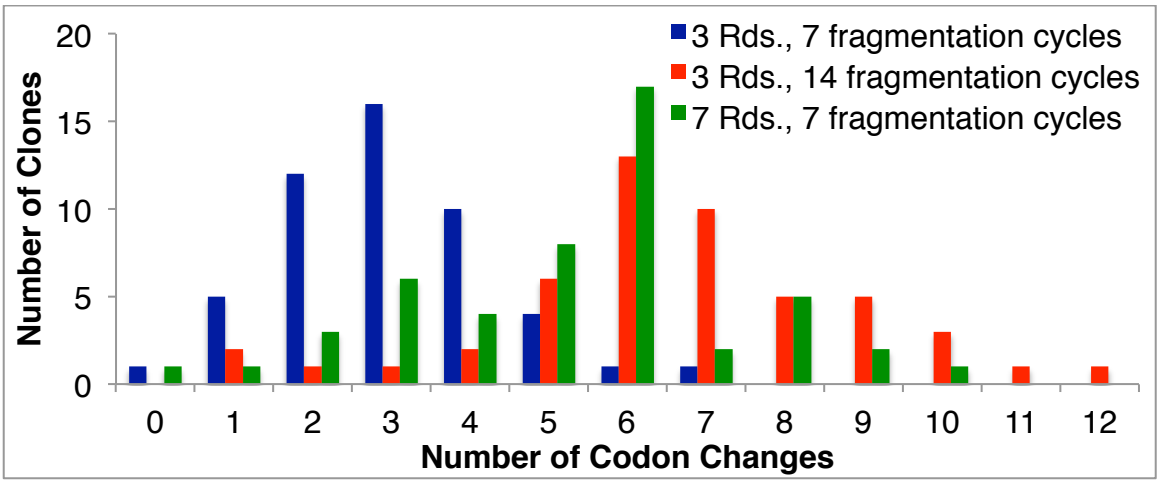
Figure 3.3: Mutation frequency per gene with different numbers of PCR fragmentation cycles (top) and frequency of mutation at targeted residues (bottom).[a]



[a] Twenty-four random clones from each library were sequenced. The bottom graph displays frequency of mutation at targeted residues for the library constructed by 3 rounds of CCM, each with 7 PCR fragmentation cycles.

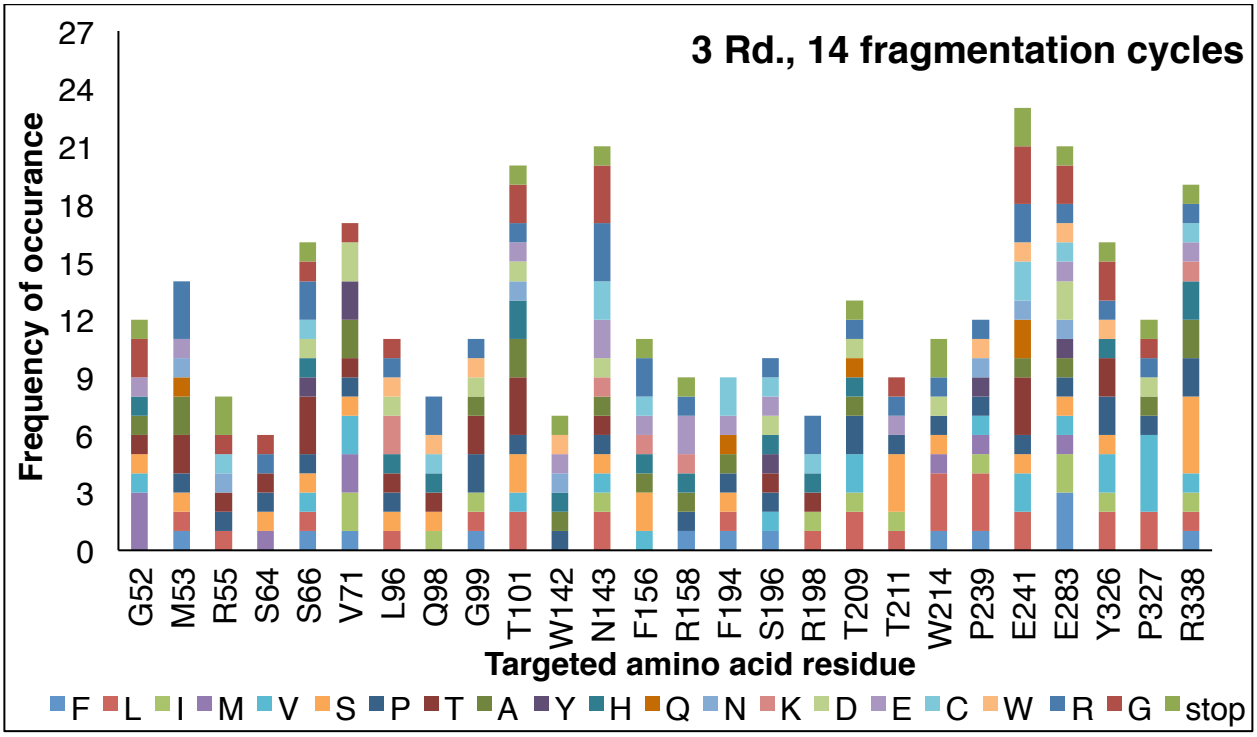
Although our initial results with RebH appeared discouraging, we decided to try this method for other enzymes. RebH is a relatively GC-rich gene (65%) and has not been codon optimized, which could cause the problems described above. Because the enzyme prolyl oligopeptidase (POP), which is currently being explored in the Lewis lab as a scaffold for generating artificial metalloenzymes,[23] has a lower GC content (49%) and has been codon optimized, we switched to using this gene as a template for CCM. Using the POP gene, we again tested how the number of rounds of fragmentation and joining PCRs and the number of fragmentation cycles per round affected the mutation frequency. Twenty-six active site residues were targeted and primers containing the degenerate codon NNN were designed for each residue. Three separate trials were performed by Dr. Belsare to test mutation frequency - 3 rounds with 7 fragmentation cycles per round, 3 rounds with 14 fragmentation cycles per round, and 7 rounds with 7 fragmentation cycles per round. Again, the mutation frequency was observed to be readily tunable by changing either of these parameters (3, 5, and 7 codon mutations per gene for the three libraries, respectively, Fig. 3.4), and no truncations of the POP gene were observed. Moreover, all targeted sites were mutated with similar efficiencies, unlike RebH (Fig. 3.5).

Figure 3.4: Mutation frequency per gene with different numbers of rounds of CCM and different numbers of PCR fragmentation cycles for POP.[a]



[a] Fifty random clones from each library were sequenced.

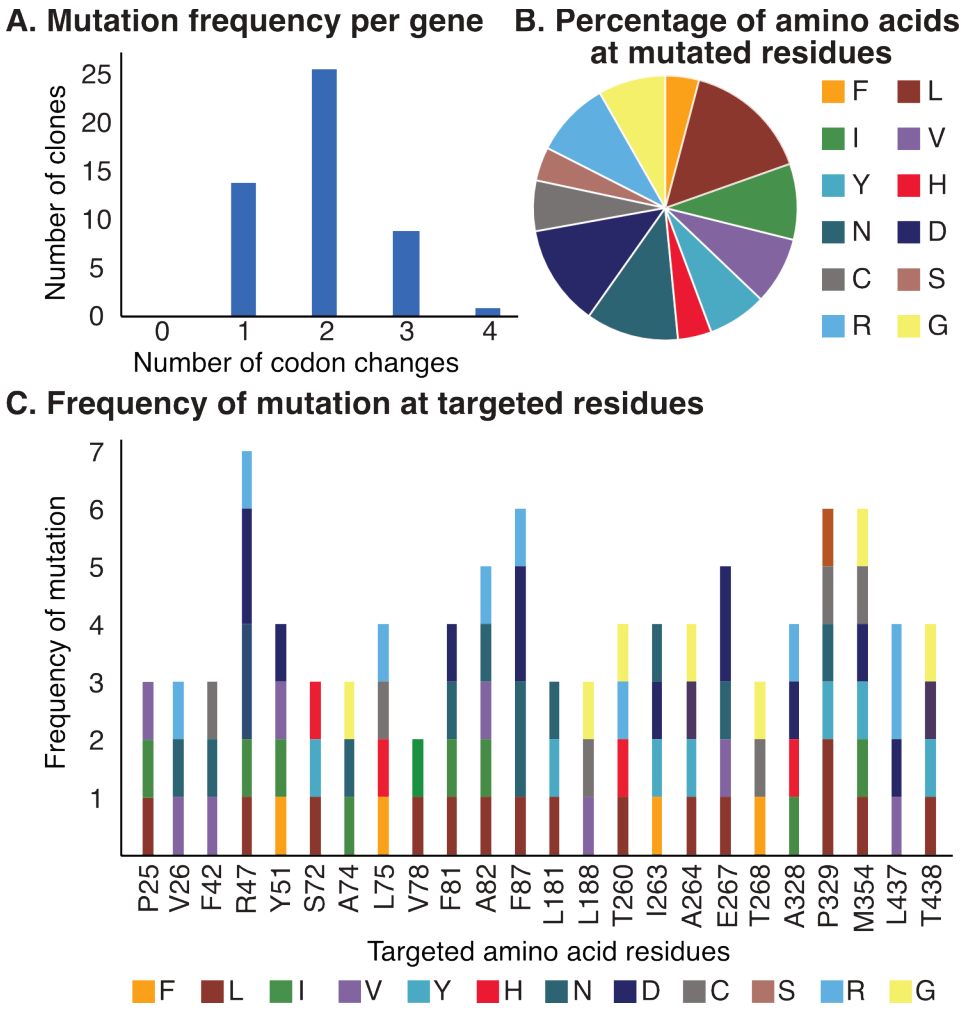
Figure 3.5: Frequency of mutation at targeted residues for a representative POP library.



We next sought to test this method with another enzyme as well as to demonstrate the utility of CCM libraries in protein engineering efforts. We first chose the cytochrome $P450_{BM3}$

(BM3), which has been extensively engineered using a variety of mutagenesis strategies.[14,24,25] We envisioned the use of CCM to generate a single library combinatorially targeting sites found throughout these previous engineering efforts. With this goal in mind, Dr. Belsare chose twenty-two such sites from the literature and designed primers for each containing a degenerate NDT codon. For this library, a mutation frequency of 2 mutations per gene was targeted, which was achieved using two rounds with 7 fragmentation cycles for this particular gene. As observed with POP, no gene truncations were seen and all sites were mutated with similar efficiency (Fig. 3.6).

Figure 3.6: Overview of a BM3 library constructed using CCM.[a]

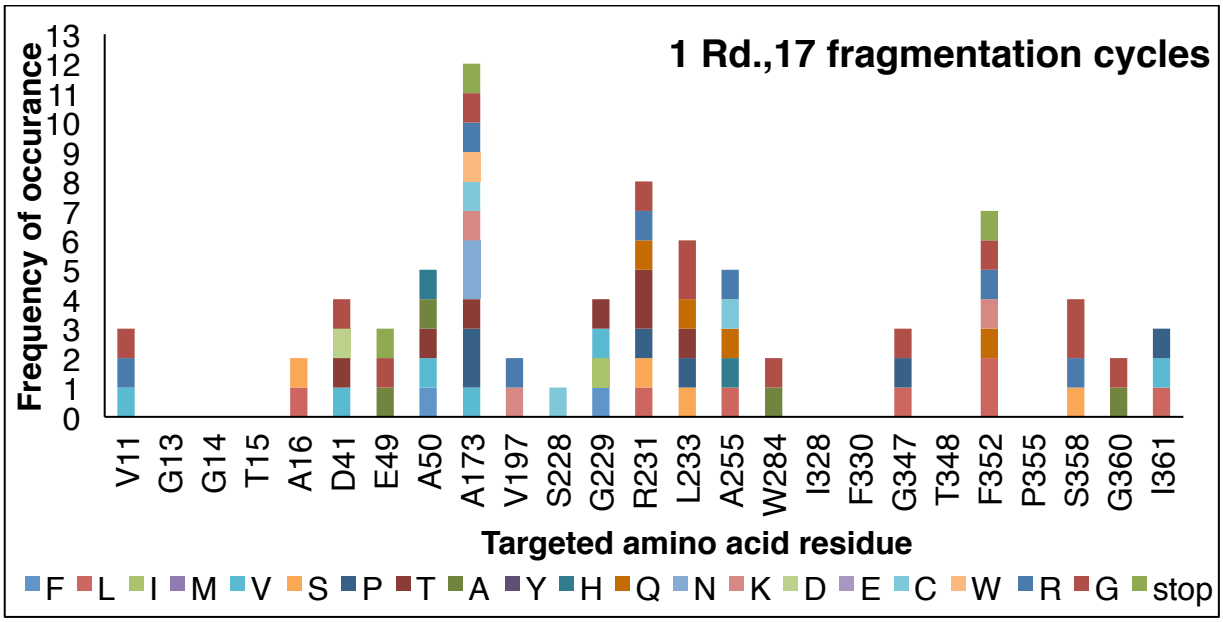


[a] Fifty random clones were sequenced to generate this data.

Using this BM3 library, Dr. Belsare screened 1,000 variants for demethylation activity on methyl ether (HME) and benzyl methyl ether (BME). Various combinations of codon mutations at F87, T268, P239, and L437 improved activity with respect to wild-type BM3. The variant (1DF) with the highest activity on BME was kinetically characterized, and a 3.1-fold improvement in catalytic activity was observed (Fig. 3.8-C). Thus, CCM can be used to target residues known to impact enzyme function in order to create new variants with improved catalytic efficiencies on multiple substrates.

After CCM had been established for BM3 and POP, we sought to demonstrate use of this method for RebH, showing that CCM statistics can be variable depending on the particular gene used as template. Instead of targeting active site mutations, as had been previously done, we chose to target twenty-five residues in the FAD cofactor binding pocket. Recent reports using other enzymes had shown that mutations within highly conserved redox cofactor binding sites could improve catalytic efficiencies in a substrate independent manner.[26,27] We envisioned that CCM could be used to find such mutations in RebH. The ease of mutation frequency tuning was important in this case, as mutation of many residues at once within this highly conserved binding pocket was expected to lead to nonfunctional enzymes. RebH variant 0K[28] (RebH-E461K, discussed further in section 3.2.2) was used as a template, and one round with 17 fragmentation cycles of CCM was performed on this gene. Upon sequencing random clones from this library, a bias in the distribution of mutations was again observed (Fig. 3.7); however, because the region between 465-467 was not targeted, no truncations were observed.

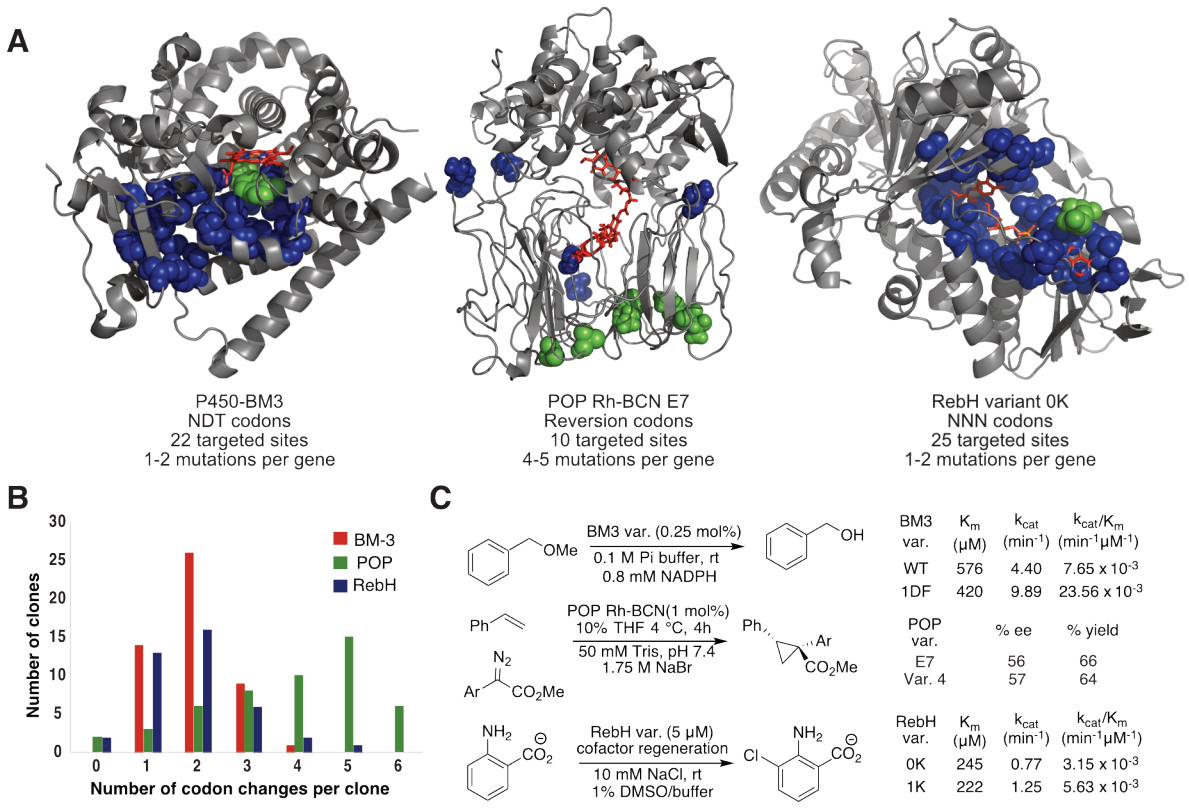
Figure 3.7: Frequency of mutation at targeted residues for a RebH library targeting the FAD-binding pocket.



[a] Fifty random clones were sequenced to generate this data.

Even though this library displayed uneven distribution of mutations, 1,000 variants were screened for halogenation activity on anthranilic acid. Halogenation reactions were analyzed by UPLC and, as expected, a significant proportion of variants displayed no activity. However, several variants showed improved activity and upon further characterization of these, variant 1K (0K+R231K) demonstrated a 1.7-fold improvement in activity. Upon kinetic characterization, it was discovered that much of this improvement was due to an increase in k_{cat} from 0.77 min^{-1} for 0K to 1.25 min^{-1} for 1K (Fig. 3.8-C). The K_m's for these enzymes remain largely the same (Fig. 3.80-C). For further use of CCM with RebH or any other GC-rich genes, we will explore the use of codon-optimized genes. We have recently obtained the codon-optimized gene for RebH and are excited to test whether truncation issues and distribution biases are eliminated by this.

Figure 3.8: Summary of mutation locations for BM3, POP, and RebH libraries (A), mutational frequency for each library (B) and kinetic characterization of hits (C).



Lastly, we hoped this CCM method would serve as a rapid way to deconvolute or recombining mutations accumulated throughout an evolution lineage. Instead of individually cloning point mutations to deconvolute or recombine mutations in a lineage, combinations of variants could be rapidly constructed through CCM with primer pools containing all mutations/reversions. Initially, deconvolution of mutations in RebH variant 10S (discussed in Chapter 4) was tested; however, because amino acid residues at 465 and 470 must be included in these experiments, truncations were again observed. Instead, we switched to demonstrating deconvolution feasibility on an artificial metalloenzyme POP variant termed E7.[29] This variant contained 10 mutations relative to the parent POP variant (ZA2). Reversion primers for all 10 sites were designed and one round of CCM with 17 fragmentation cycles was performed. Two hundred variants from this library were screened using a cyclopropanation reaction, and enzymes

that showed the same conversion and enantioselectivity for this reaction were discovered. Sequencing these variants revealed that five of the ten mutations accumulated during this lineage were not necessary for improved conversion or enantioselectivity. Good convergence on these five reversions was observed in the sequenced variants, and one variant actually contained all five reversions (Fig. 3.9).

Figure 3.9: Hits from POP-E7 deconvolution library.[a]

		Targeted Residues										e.e. (st.dev.)
		84	99	146	161	166	202	301	308	330	335	
	POP-ZA2	S	G	F	K	V	D	S	T	K	V	32.0 (3.0)
	E7	N	S	A	N	A	A	G	S	I	A	57.0 (1.4)
Variants	1	N	S	F	N	A	D	G	S	I	A	56.6 (0.7)
	2	N	S	F	N	A	D	G	S	I	A	56.6 (0.7)
	3	N	S	F	N	A	A	G	S	I	V	57.0 (1.4)
	4	N	S	F	N	A	D	G	T	K	V	56.0 (1.4)
	5	N	S	A	N	A	A	G	T	K	A	56.0 (1.4)

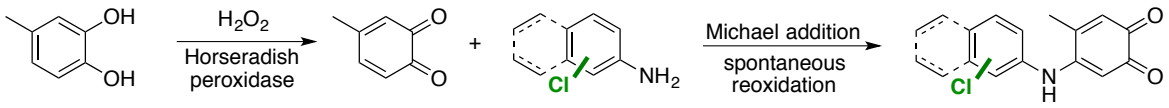
[a] Residues at targeted sites in POP-ZA2 (parent from evolution lineage) and E7 are shown in gray. Residues shown in blue were targeted but not reverted; residues shown in green were reverted.

3.2.2 A Colorimetric screen for rapid analysis of FDH activity on anilines

In addition to new, efficient mutagenesis methods, we are also interested in developing FDH screening methods for directed evolution campaigns. Although RebH is capable of halogenating a diverse array of arenes, we have found that substrates must be sufficiently electronically activated for halogenation to occur (discussed more in chapter 5). For example, while aniline is readily chlorinated by RebH and variants, phenol, even though still electronically activated, appears too poorly nucleophilic for significant conversion to be observed. Thus, extending the halogenating ability of RebH to less activated arenes could enable halogenation of new substrate classes by FDHs.

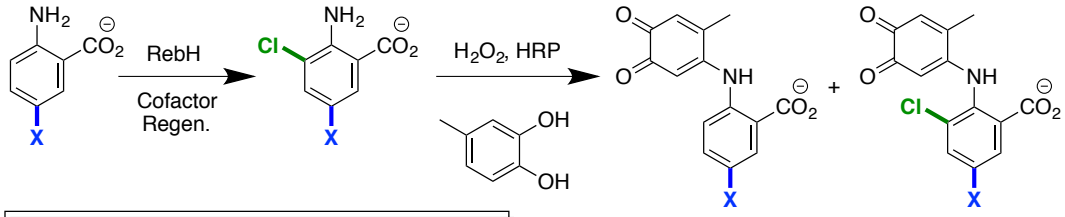
As described in section 3.1, Micklefield and coworkers describe a colorimetric screen for rapidly assaying halogenation of anilines.[20] After conducting halogenation reactions, 4-methylcatechol can be oxidized by H_2O_2 and horseradish peroxidase and can subsequently react with aryl amines (Scheme 3.1). The resulting compounds are brightly colored, and in many cases the wavelength of the maximum absorbance in the UV/Vis spectra of these compounds are significantly different when they are halogenated. We envisioned using this method to enable the rapid directed evolution of RebH on more electron deficient substrates.

Scheme 3.1: General scheme for high-throughput colorimetric assay.



In addition to describing this colorimetric screen, Micklefield and coworkers reported a point mutation in the 7-tryptophan halogenase PrnA that increased the activity 8-fold on the substrate anthranilic acid.[30] When we made the corresponding mutation in RebH (E461K), we observed a similar increase in activity. This variant, termed 0K, also demonstrated high regioselectivity for chlorination of the 3-position of anthranilic acid (ortho to the amine). Because this was demonstrated as an effective substrate for the colorimetric assay, we decided to use anthranilic acid as an initial substrate and 0K as a parent for directed evolution. By altering the substituent at the 5 position of anthranilic acid (X in Fig. 3.10) to increasingly electron-withdrawing groups (substrate walking), variants of RebH could be evolved with activity on less electronically activated sites.

Figure 3.10: Overview of directed evolution approach to generate FDHs capable of halogenating less activated sites.



X	% Yield WT	% Yield 0K	Van der Waals radius
H	40%	>100%	1.2
Me	1.1%	43%	2
Cl	0%	4%	1.8
CF_3	0%	0%	2

Examining the scope of 0K on 5-substituted anthranilic acid derivatives revealed that diminished activity is observed even for the methyl-substituted compound, which should be electronically activated relative to the unsubstituted compound. Steric bulk appears to be poorly tolerated at the 5-position as evidenced by the large increase in the K_M of the methyl-substituted substrate (Table 3.1). Interestingly, however, trace activity and no activity are observed for the chlorine- and trifluoromethyl-substituted compounds, respectively (Fig. 3.10). These substituents are similar in size but increasingly electron withdrawing relative to methyl. Thus, while steric differences eliminate the possibility of comparing electronic effects between H and larger substituents, within the series of Me, Cl, and CF_3, such comparisons appear to be possible.

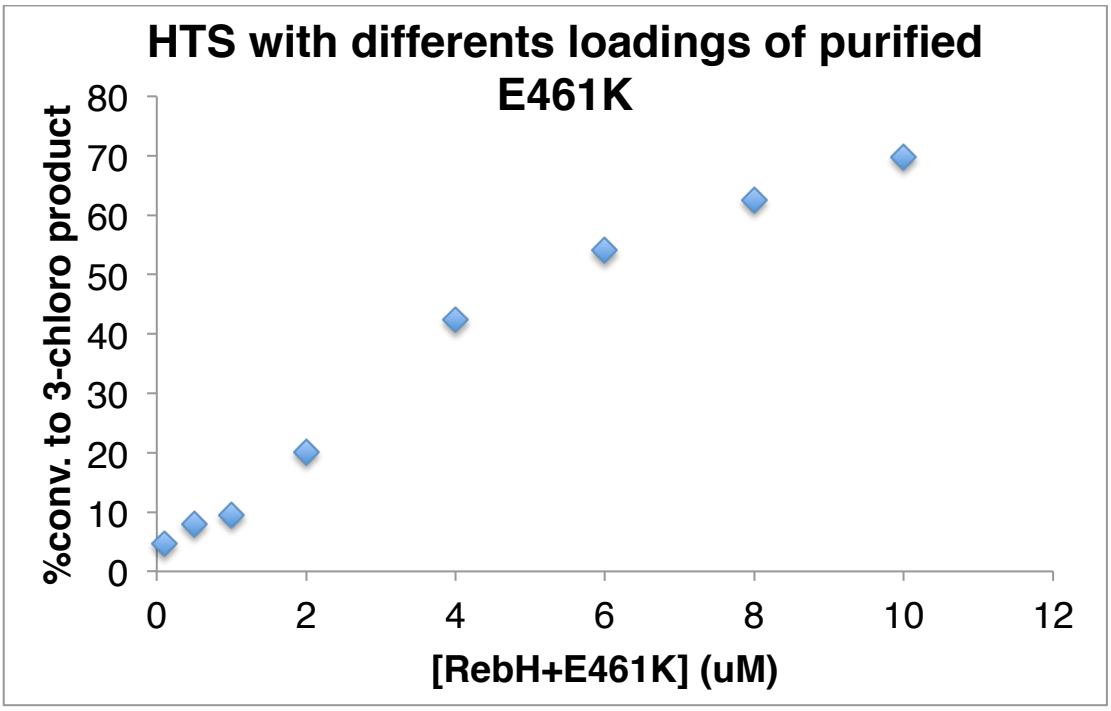
Table 3.1: Catalytic parameters for anthranilic acid analogs.[a]

FDH	Substrate	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} μM^{-1})
PrnA-E450K[b]	X = H	0.93	384	2.4×10^{-3}
RebH-E461K	X = H	0.77	245	3.2×10^{-3}
RebH-E461K	X = Me	0.61	4070	1.5×10^{-4}

[a] 5-X-2-aminobenzoic acids were used as substrates. [b] Previously reported literature values.[30]

The previous report of this assay demonstrated its use with purified enzyme; however, its use in the presence of cell lysates was not explored.[20] Reliable signal must be observed in cell lysates to use this assay for directed evolution. To begin optimization in cell lysates, we first wanted to demonstrate that this method is able to colorimetrically measure product formation using our bioconversion conditions. To test this, several bioconversions were conducted on anthranilic acid with different loadings of purified 0K (0.1-10 μM). Chlorination reactions were heat quenched after one hour, high-throughput screen (HTS) components were added, and disappearance of λ_{max} = 535 nm was measured. The results from this test demonstrate the effectiveness of this HTS - higher concentrations of 0K show higher conversion to product (Fig. 3.11). It should be noted that while several arylamines can be distinguished from their chlorinated analogues using this assay, we found early on that this is not true in every case. For example, while there is a shift in λ_{max} when anthranilic acid is chlorinated at the 3 position (ortho to the amine), chlorination at the 5 position (para to the amine) results in no observable shift and thus cannot be detected using this screen. In addition, no color change is observed upon chlorination of the substrates 2-aminobenzoic acid methyl ester.

Figure 3.11: Colorimetric assay used to measure conversion of anthranilic acid with purified RebH variant 0K.



We quickly moved to using this assay with cell lysate. Bioconversions were conducted in lysate with two different loadings of 0K-containing lysate (25 and 50 µL). Several bioconversions were not initiated with glucose, and different ratios of anthranilic acid and 3-chloroanthranilic acid were added instead. These served as calibration curves for two loadings of lysate. Based on these plots, it appears the amount of lysate used in bioconversions does not have a significant impact on signal (Fig. 3.12). Bioconversions were subjected to the same method for HTS analysis as those with purified enzyme. Results from this were compared to HPLC conversions for the same reactions. Good agreement was observed between the two analysis methods; however, higher standard deviations were seen when using the colorimetric assay (Fig. 3.13). Because hits from all rounds of directed evolution are fully characterized, the standard deviations were not concerning. This colorimetric assay will serve as an initial screen for library variants and hits will be expressed on a larger scale, purified, and re-screened by HPLC.

Figure 3.12: Calibration curves using the colorimetric assay in cell lysate, with different amounts of added lysate (25 and 50 µL).

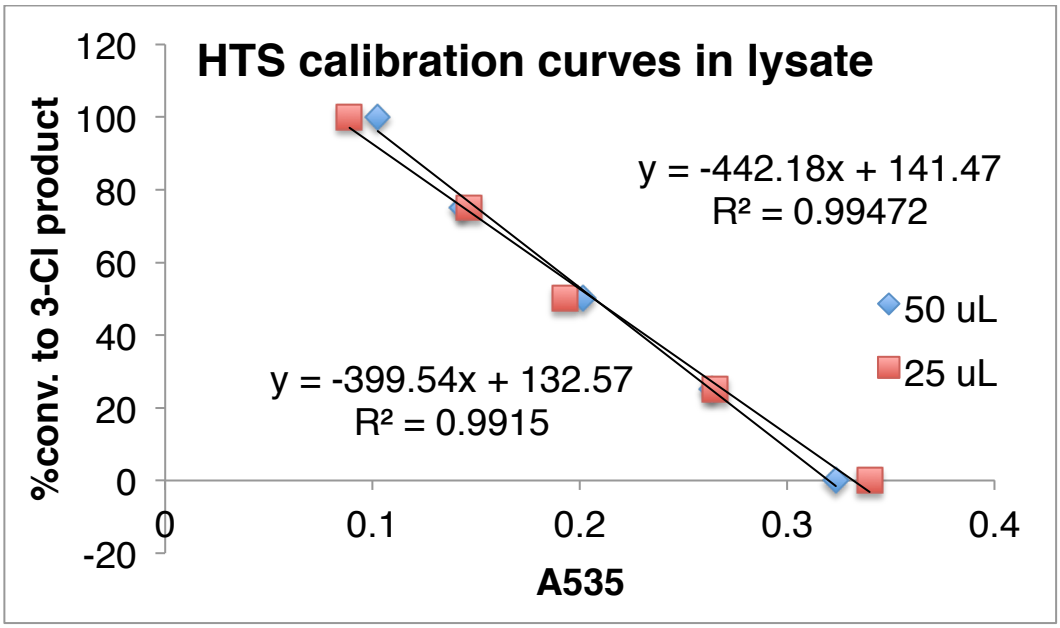
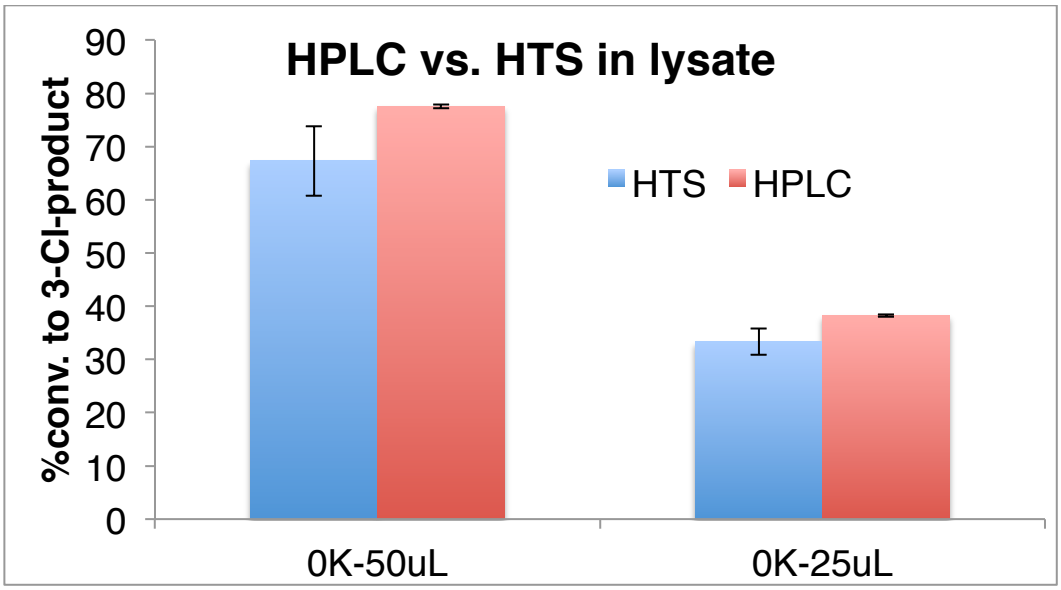


Figure 3.13: Comparison between colorimetric HTS and HPLC in cell lysate.



With a functional method for measuring chlorination of anthranilic acid in cell lysate in hand, we began mutagenesis and screening on the parent FDH 0K. Diversity in the 0K gene was generated using error-prone PCR similar to the method described in Chapter 2. Approximately

850 variants were expressed and screened using the colorimetric assay. Five hits were identified, but upon further characterization, these were all found to be false positives on anthranilic acid. Because the ultimate goal of these evolution efforts is to achieve chlorination on less electronically activated substrates, these five enzymes were tested for activity on 5-chloroanthranilic acid (Fig. 3.10). Surprisingly, one variant showed a nearly 2-fold increase in conversion for this substrate, even though it halogenated anthranilic acid with similar efficiency. This variant contained two mutations - Y116C and D435G - and was thus named 1CG.

Variant 1CG was subjected to a second round of error-prone mutagenesis and screening. Two thousand variants were expressed and screened. Eight hits were obtained, and three of these were found to improve activity on 5-chloroanthranilic acid. Each of these hits contained two mutations. Upon deconvolution and recombination of these mutations, it was found that variant 2SS (G14S and G504S) showed the highest conversion of 5-chloroanthranilic acid (2-fold higher than 1CG). Using 2SS as parent, another 2,000-member library was expressed and screened. One hit was discovered from this round (3TS, S110T and N512S), but showed only a 1.4-fold improvement in conversion of 5-chloroanthranilic acid.

Concurrently with these studies, other natural halogenases were being explored (see chapters 5–6). In particular, an FDH from the fungus *Pochonia chlamydosporia* was found to halogenate several free phenol substrates, which are significantly less electronically activated than their aniline counterparts (discussed further in chapter 5). Inspired by this result, we decided to pursue a genome mining approach to developing halogenase biocatalysts capable of halogenating less electronically activated substrates. While these efforts have temporarily halted the evolution efforts described in this section, we have clearly demonstrated that this approach can be used to access FDHs capable of more efficiently halogenating less electronically activated

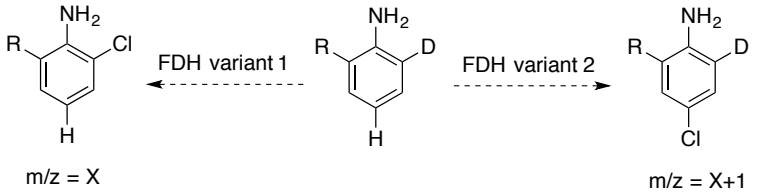
substrates. In addition, the development of this colorimetric screen as a practical tool for directed evolution campaigns can be used in future efforts with aniline substrates. It should be noted that very recently, Sewald and coworkers developed a fluorescence screen for FDH halogenation and demonstrated its use in cell lysate.[31] This screen could also prove useful in future engineering efforts.

3.2.3 A MALDI-MS screen for directly probing FDH site-selectivity

A previous evolution campaign from the Lewis lab to widen the substrate scope of RebH had led to a variant, 3SS, that halogenated tryptoline with higher site selectivity for C6 than wild-type RebH, which halogenates C6 and C7 with similar efficiency.[4] This altered selectivity was not screened for directly, but instead came as a consequence of improving activity on tryptoline. We sought to develop a screen by which we could directly screen for altered site-selectivity, instead of determining site selectivity *post hoc*.

While HPLC could potentially be used to resolve different chlorinated isomers, such separations are often challenging and require extended run times. In some cases, such as with 6- and 7-chlorotryptoline, the isomers cannot be separated using standard HPLC conditions. Previously, Li and coworkers reported a method for directly observing changes in enantioselectivity of hydroxylation reactions on deuterated substrates by analysis with mass spectrometry.[32] Inspired by this work, we envisioned that the use of site-selectively deuterated arenes as FDH substrates followed by mass spectrometry screening could allow for direct observation of changes in selectivity of chlorination (Scheme 3.2).

Scheme 3.2: Deuterated arenes as substrates and mass spectrometry as a screen to directly analyze site selectivity of chlorination reactions.



In reactions proceeding through an electrophilic aromatic substitution (EAS) mechanism, deprotonation occurs after the rate-limiting step. Because halogenation with FDHs proceeds through an EAS mechanism, the ratio of halogenated products should not be affected by a kinetic isotope effect. To confirm this, we determined the initial rates of RebH halogenation of d_5-L-tryptophan and L-tryptophan (Fig. 3.14). No significant difference was observed. We also conducted competition experiments in which bioconversions were conducted with equal concentrations of d_5-L-tryptophan and L-tryptophan. Again, no significant difference was observed between the rates of conversion of the two substrates (Fig. 3.15). Based on these experiments, we concluded that no significant isotope effect exists for RebH-catalyzed halogenation reactions and that the ratio of halogenated products could be used to directly evaluate site selectivity without correcting for any isotope effect (as was required in the work by Li and co-workers).

Figure 3.14: Initial rates for chlorination of d_5-L-trp and L-trp.

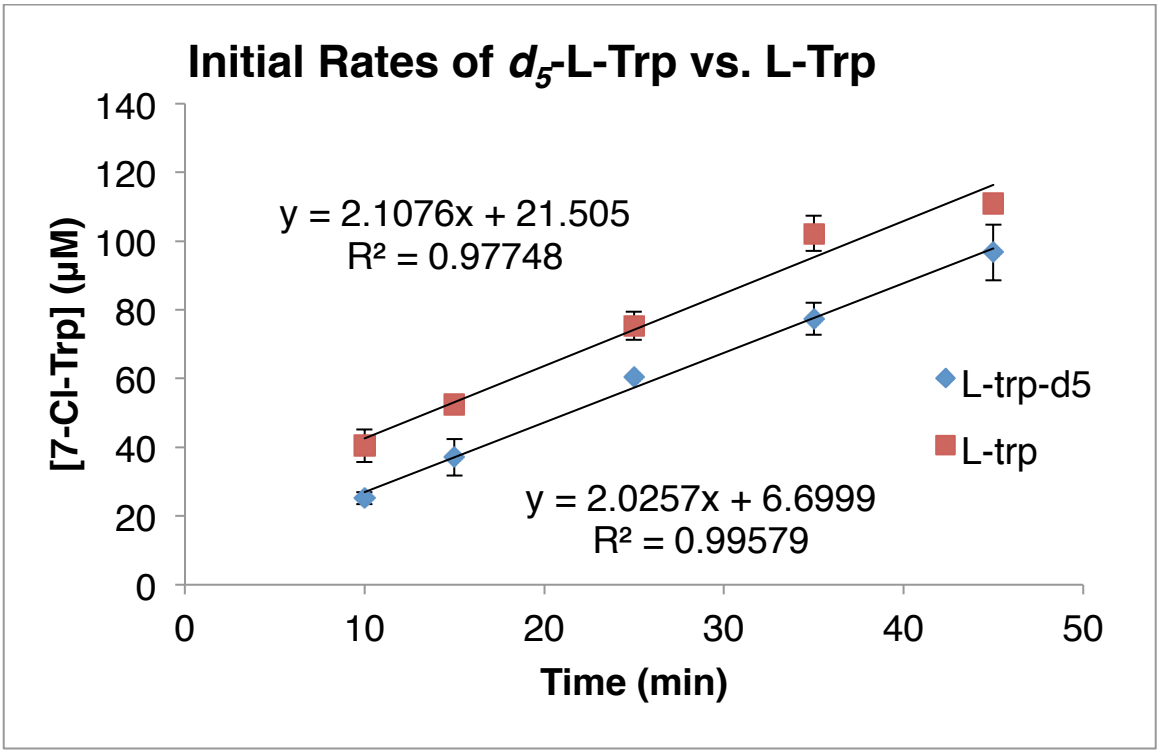
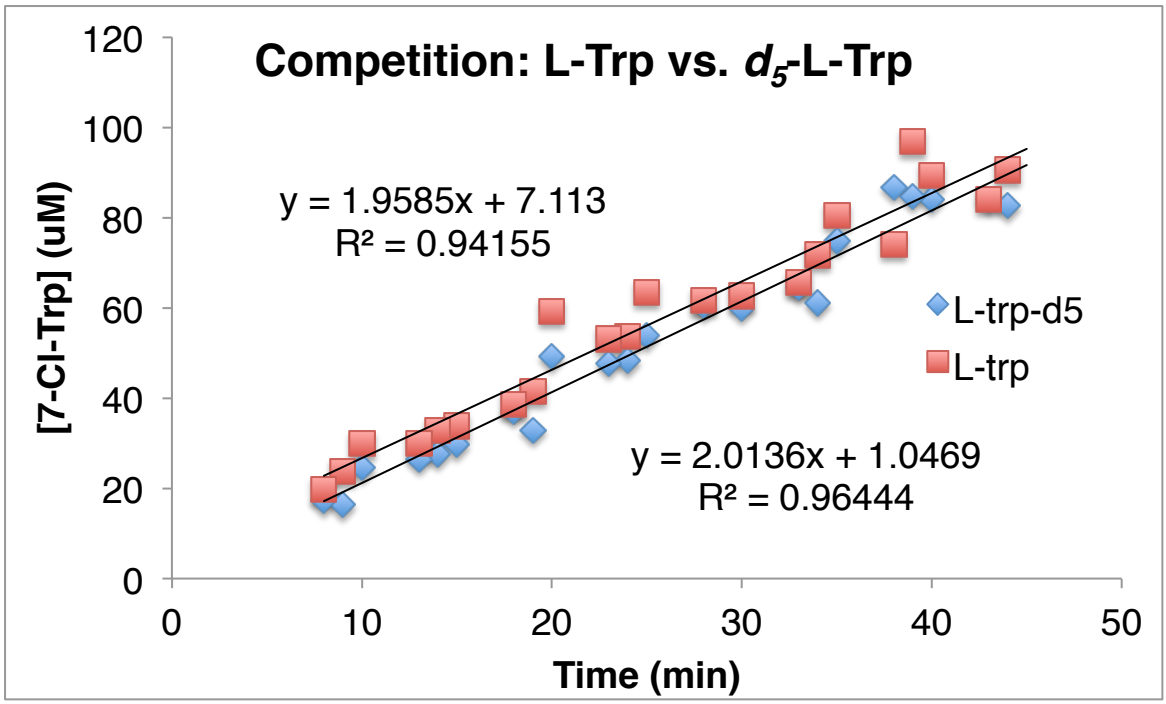
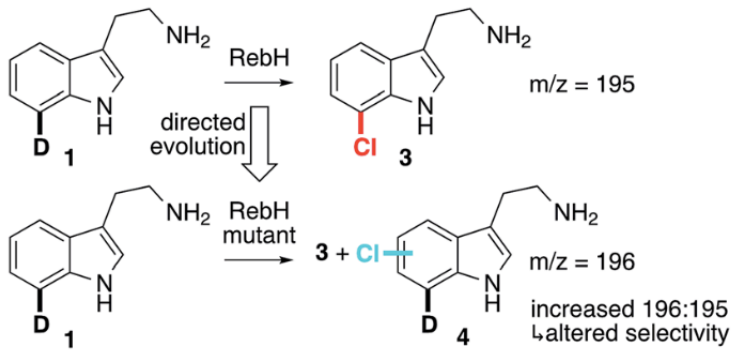


Figure 3.15: Chlorination competition reactions for d_5-L-trp and L-trp.



To establish whether mass spectrometry analysis of reactions conducted on deuterated substrates could be used as an assay for halogenase selectivity, we chose tryptamine as a test substrate. In previous work, we had demonstrated that good yields of exclusively 7-chlorotryptamine could be obtained using RebH as a catalyst.[33] In addition, deuterated tryptamines could be more readily prepared than tryptophans. For these reasons, we chose to alter site selectivity on the substrate tryptamine. We initially made 7-deutrotryptamine (Probe **1**) so that any change in selectivity could be readily identified during screening (Fig. 3.16).

Figure 3.16: Overview of mass spectrometry assay for FDH selectivity using Probe 1.

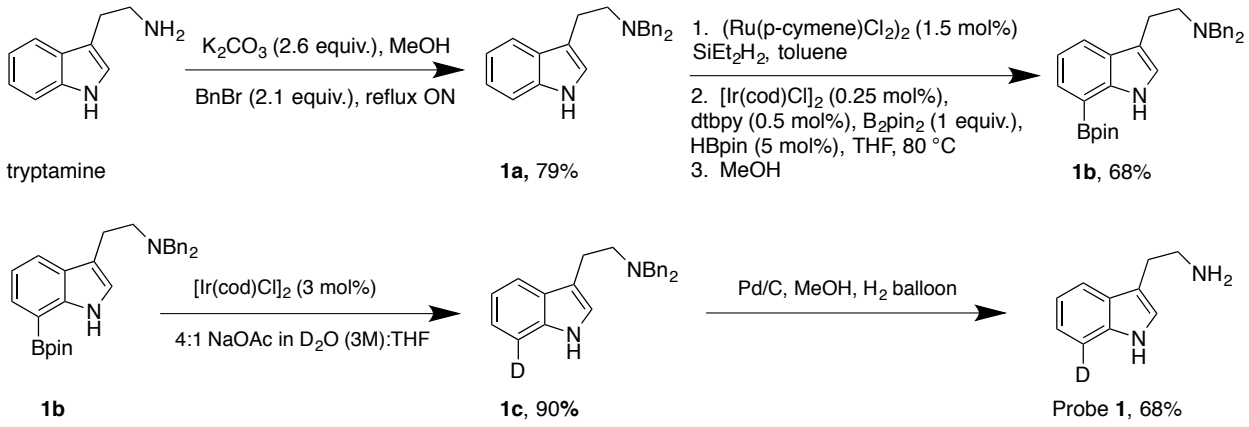


Selective C7 functionalization required for Probe **1** synthesis was achieved using a one-pot indole borylation procedure reported by Hartwig and coworkers.[34] The indole NH was silylated, and the silyl group was used to direct an iridium catalyst to the proximal C7-H bond. Oxidative addition into this bond followed by reductive elimination occurs to form a new C-B bond. The silyl group can then be removed, affording the selectively borylated product. This sequence was carried out on benzyl-protected tryptamine (**1a**, Scheme 3.3). While silylation and borylation occurred as desired, removal of the silyl group with high concentrations of NaOAc in water also removed the previously installed boryl group. In the absence of NaOAc, selective removal of the silyl group was achieved, suggesting that protodeborylation could be achieved in

the presence of Ir, NaOAc, and water. In practice, this proved to be the case, and full protodeborylation was achieved in a matter of hours under these conditions.

Prior to establishing the protodeborylation conditions above, I attempted to use deuterodeborylation for selective deuterium incorporation into **1b** by boiling this compound in d_4-acetic acid. In practice, however, H/D exchange occurred at all sp^2 C-H bonds. I reasoned that the Ir-catalyzed deborylation conditions identified above could be used for selective deuterodeborylation of **1b** by simply using D_2O as solvent instead of H_2O. Indeed, under these conditions, **1b** was converted to **1c** with high isotopic purity at the 7 position (>95% by HNMR). Deprotecting the benzyl groups on the primary amine afforded Probe **1** (Scheme 3.3). Notably, a very similar Ir-catalyzed method for selective deuteration was published by Smith and coworkers after our report.[35]

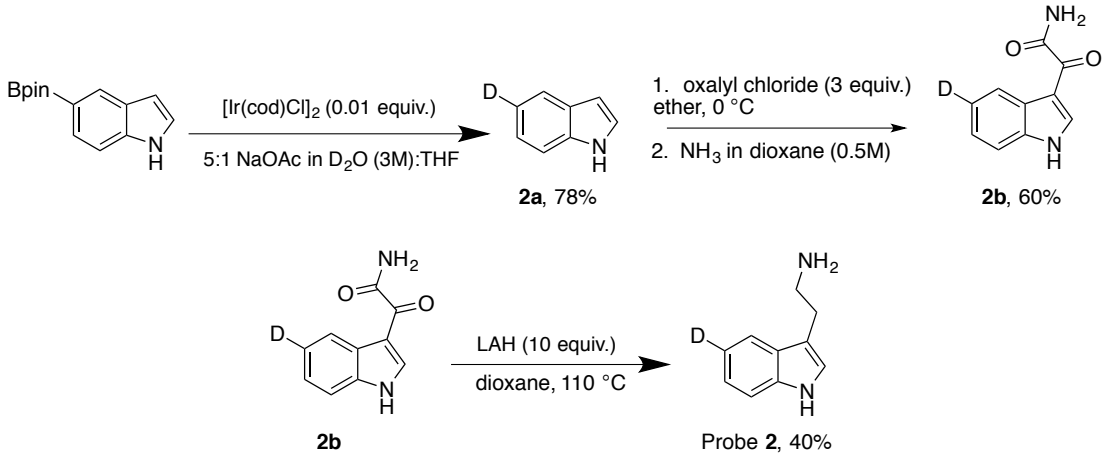
Scheme 3.3: Synthesis of Probe 1.



In addition to a 7-deuterotryptamine probe, we also anticipated that other probes might be necessary during the course of evolution. For example, should a halogenase capable of functionalizing both the C5 and C6 positions arise, we would need another probe, deuterated at one of these positions, to directly screen for increased selectivity at one of these sites. This did in fact occur (discussed in Chapter 4), and as such, a 5-deuterated probe was also developed (Probe

2). No procedure existed for selective functionalization of the 5-position of indole compounds, and thus 5-Bpin-indole was purchased. This compound was deborylated using the $[Ir(cod)Cl]_2$ catalyst in D_2O, affording 5-deuteroindole (**2a**, Scheme 3.4). To this compound was added oxalyl chloride and ammonia to form **2b**. Both carbonyls were reduced with lithium aluminum hydroxide, thus forming 5-deuterotryptamine (Probe **2**, Scheme 3.4).

Scheme 3.4: Synthesis of Probe 2.



We next developed a mass spectrometry screen that could reliably measure conversion to product as well as H:D product ratio. MALDI MS was chosen as a screen because collecting and analyzing data could be easily automated. After spot conditions were optimized to ensure homogeneity, the ability to accurately measure conversion to product was tested. RebH-catalyzed halogenation of tryptamine was conducted and analyzed by MALDI MS and UPLC. It was found that, while the MALDI MS response was not linear, it could be used to determine relative conversions (Fig. 3.17). This allows MALDI MS analysis to function as an initial screen of libraries. A secondary HPLC/LC-MS analysis of conversion and selectivity could be performed for hits found using this MALDI MS screen.

Once deuterated Probe **1** had been synthesized, the reliability of MALDI MS to distinguish tryptamine from Probe **1** was explored. Solutions of a constant concentration of

tryptamine (0.5 mM) and varying concentrations of Probe **1** (0.01-0.5 mM final concentration) were spotted onto a MALDI target plate and tested with an automated MS acquisition method. The ratio of the intensity of the m/z peak corresponding to Probe **1** (162) and the intensity of the m/z peak corresponding to tryptamine (161) was calculated and plotted against the known concentration of Probe **1** (Fig. 3.18). This plot demonstrates that the MALDI MS screen can be used to determine the relative ratio of Probe **1**:tryptamine.

Figure 3.17: RebH time course analyzed by MALDI MS and UPLC.

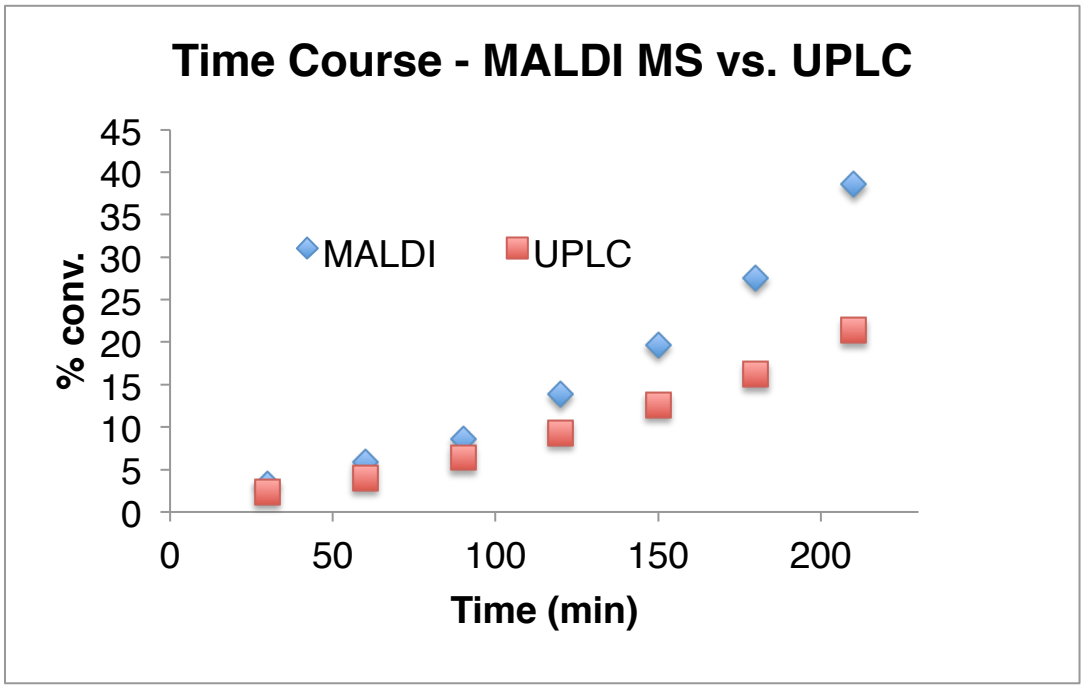
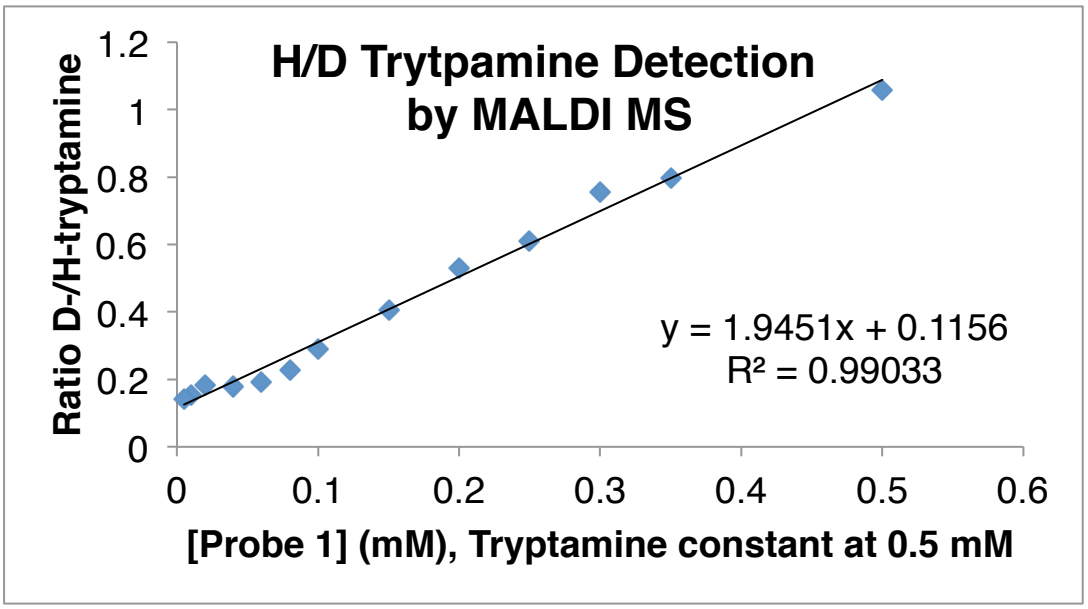


Figure 3.18: Distinguishing Probe 1 from tryptamine by MALDI MS.



With deuterated probes synthesized and a reliable MALDI MS screen capable of measuring conversion and ratio of M:M+1, we began using directed evolution to alter the site selectivity of RebH on the substrate tryptamine. This directed evolution campaign is described in Chapter 4.

3.3 Conclusions

Alternate mutagenesis and screening methods for efficient directed evolution of FDHs were developed. A codon mutagenesis method capable of targeting many amino acid residues combinatorially with tunable frequency was developed. The use of this method was demonstrated for multiple protein engineering applications. In addition, a high-throughput colorimetric assay was optimized for use in cell lysate and was subsequently used for FDH evolution. Lastly, a MALDI MS assay was designed that allows FDH libraries to be screened directly for altered site-selectivity. This screen was used to alter the selectivity of RebH on the substrate tryptamine, described in Chapter 4.

3.4 Experimental

3.4.1 General experimental procedures

Materials:

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Plasmid pEVOL-pAzF was provided by the Schultz group of the Scripps Research Institute. Plasmids pET-28a/RebF and pET-28a/RebH in BL-21 DE3 E. coli were provided by the Walsh group of Harvard Medical School, Boston, MA.[36] The pGro7 plasmid encoding the groES and groEL chaperone set was purchased from Takara (Otsu, Shiga, Japan). *E. coli* DH5α, BL21 (DE3), *BamH I, Nde I, EcoRI* restriction enzyme, T4 DNA ligase, Taq DNA polymerase and Phusion HF polymerase were purchased from New England Biolabs (Ipswitch, MA). KOD hot start master mix (product number 71842) was purchased from EMD Millipore (Billerica MA). Luria broth (LB), rich medium (2YT), L-arabinose and agar were purchased from Research Products International (Mt. Prospect, IL). Qiagen DNA extraction kit (Cat# 28706) and plasmid isolation kit (Cat# 27106) were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. DNA purification kit (Zymo, Cat# D4004) was purchased from Zymo research (Irvine, CA) and used as recommended. Lysozyme (L6876), DNAseI (DN-25, 1g) were purchased from sigma Aldrich. Sequencing of codon mutagenesis libraries was done at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility (900 E. 57th Street, Room 1230H, Chicago, IL 60637). Electroporation was carried out on a Bio-Rad MicroPulser using method Ec2. Ni-nitrilotriacetic acid (Ni-NTA) resin (# 88222) and Pierce® BCA Protein Assay Kits (Cat# 23225) were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were followed when using both products. Amicon® 30

kD and 50 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C. Oligonucleotides were purchased from Integrated DNA Technologies in a 96- well format (San Diego, CA). Glucose dehydrogenase (GDH-105), FAD, NAD, and NADPH (# 41939) were purchased from Codexis (Redwood City, CA).

Acetonitrile, Tetrahydrofuran, Ethanol, and DMSO for other applications were purchased from Fisher Chemical, HPLC grade. 4-methoxy styrene (cat# B22553) was purchased from Alfa Aesar. 4-Azido-l-phenylalanine (# F-3075.0001) was purchased from Bachem (Bubendorf, Germany). Hexyl methyl ether (# A17207), benzyl methyl ether (# B23490), and 2-aminobenzoic acid (# A12715) were purchased from Alfa Aesar. 5-aminolevulinic acid hydrochloride (103920050) was purchased from Acros organics. Isopropyl β-D-1-thiogalactopyranoside (# I2481C25) was purchased from Gold biotechnology. HiTrap™ Q FF (17-5156-01) and HisTrap HP (17-52-48-02) were purchased from GE Healthcare. $[Ir(cod)Cl]_2$ and $[Ru(p\text{-cymene})Cl_2]_2$ were purchased from Strem. Deuterated solvents were obtained from Cambridge Isotope labs. Silicycle silica gel plates (250 mm, 60 F254) were used for analytical TLC, and preparative chromatography was performed using SiliCycle SiliaFlash silica gel (230-400 mesh).

24 well deep well plate (CS15124) and 96 well filter plates (201009-100 PP, 0.45) for expression and screening of prolyl oligopeptidase libraries were purchased from Fisher Scientific and Agilent Seahorse, respectively. 96 well plate cap mats (cat#AWSM-1003S) and Aeraseal adhesive sealing films (cat# 202505) were purchased from Arctic white LLC and Research Products International (Mt. Prospect, IL). Azide Agarose was purchased from Click Chemistry Tools LLC. Labquake Tube Shaker/Rotators were purchased from Thermo Scientific (Catalogue

(Cat)# 4002110Q). Dirhodium BCN cofactor and aryldiazo acetate were synthesized as previously described.[23]

Complete analytical data have been reported in the literature for: *N,N-bis(phenylmethyl)-1H-Indole-3-ethanamine* (**1a**).[37]

General Procedures:

Standard molecular biology and cloning procedures for preparation of competent cells, restriction endonuclease digestion, and T4 DNA ligation reaction were followed.[38] Reactions were analyzed using an Agilent Technologies 1200 UHPLC, Agilent Technologies 6130 LC-MS or Agilent Technologies 1290 SFC. RebH and prolyl oligopeptidase protein concentrations were determined by A280 measurements taken on a Tecan Infinite M200 pro microplate reader. Concentration of purified P450 BM3 was determined as previously reported using carbon monoxide binding assay.[39] Reverse phase preparative chromatography was carried out using a Biotage Isolera One or Agilent Technologies 1100 HPLC. ^{1}H and ^{13}C NMR spectra were recorded at 500 MHz and 126 MHz, respectively, on a Bruker DMX-500 or DRX-500 spectrometer, and chemical shifts are reported relative to residual solvent peaks.[40] All mass spectra were collected at the University of Chicago Mass Spectrometry Service Center. High-resolution mass spectra were obtained on an Agilent Technologies 6224 TOF LC-MS. High-throughput screening was performed using a Bruker Ultraflextreme MALDI-TOF-TOF with a Bruker MTP384 steel massive target plate. A Microlab® Nimbus liquid handling robot was used in MALDI target plate preparation, and library expression and screening. Library colonies were picked using a Norgen Systems colony-picking robot.

MALDI target preparation and screening: A 384-well MALDI target plate was spotted with 2 μL of a solution of α-cyano-4-hydroxycinnamic acid by a liquid handling robot. The matrix solution contained 7.5 mg/mL of α-cyano-4-hydroxycinnamic acid in 1:1 THF:H_2O. This was dried in a vacuum oven for 15 minutes. Once dry, 2 μL of the filtered bioconversions were spotted onto the MALDI target plate by a liquid handling robot. This was dried in a vacuum oven for 15-30 minutes. The plate was loaded into a Bruker Ultraflextreme MALDI-TOF-TOF and an automated method was developed. Spectra were generated with the reflectron positive (RP) mode. The detector range was set at 160-200 Da. Final mass spectra were produced by averaging 500 raster shots taken at 50 random positions within each spot, which amounted to 25-30 seconds per spot. Targets were shot using the AutoXecute tool of the Flex Control acquisition software.

MALDI MS data analysis and HPLC screening: MALDI MS peaks were analyzed in Flex Analysis software. All peaks within the detector range besides those corresponding to D-tryptamine, chlorotryptamine, and D-chlorotryptamine (m/z = 162, 195, and 196) were set as background. Data from the spectra for these three peaks were then exported into Excel. An Excel macro was developed to insert the value "0" for spectra that did not contain peaks 195 and/or 196. From this list, the data could be easily arrayed into conversions and selectivities in 96-well-plate format in Excel. Hits were identified as variants that showed higher conversion relative to parent or higher selectivity relative to parent. When using Probe **1**, the ratios 196/(162+196+195) (for tryptamine conversion) and 196/(196+195) (for selectivity) were calculated. When using Probe **2**, 195/(162+196+195) and 195/(196+195) were calculated. These ratios were compared to those of the parent reactions, and the highest hits were re-

screened by UPLC. Hits confirmed by UPLC were sequenced and verified by enzyme purification and re-analysis. For rounds 7 and 8, 5-chlorination hits were directly identified by MALDI MS using 195/(162+196+195) and 195/(196+195). Variants with high values for 196/(162+196+195) and 196/(196+195) were re-screened by UPLC Method 2 (see below) to distinguish 6- from 7-chlorination.

UHPLC/LC-MS methods:

UHPLC/LC-MS Method 1: Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN; 0-10 min, B = 15%; 10-17 min, B = 15-22%; 17-20 min, B = 22-30%; 20-21 min, B = 30%. Absorbance at 280 nm was measured.

UHPLC/LC-MS Method 2: Agilent Eclipse Plus C18 4.6 x 50 mm column, 3.5 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN; 0-4 min, B = 15%; 4-4.5 min, B = 15-20%; 4.5-6.5 min, B = 20-30%; 6.5-7 min, B = 100%. Absorbance at 280 nm was measured.

3.4.2 Specific experimental procedures

Construction of combinatorial codon mutagenesis libraries: Combinatorial codon mutagenesis libraries were prepared based on a procedure described by J. Bloom.[16] The first step was to generate a template for the library. Oligonucleotides used in this step were gene specific non-mutagenic oligonucleotides. Oligonucleotides for template generation for RebH contained overhangs for NdeI and HindIII and BM-3 had overhangs for BamHI and EcoRI restriction sites. In the case of POP, only 900 base pairs containing the β-domain were amplified and the resulting fragment was cloned into a vector containing the rest of the POP gene via Gibson assembly. Mutagenic oligonucleotides for CCM libraries were obtained in a 96 well plate format and diluted to a concentration of 100 μM using sterile molecular biology grade water. A forward

mutagenic primer pool was prepared by adding equimolar quantities of individual forward primer to obtain a final concentration of 4.5 μM. Reverse mutagenic primer pool was prepared in a similar way.

1. Template generation

DNA Plasmid	1 μl (10 ng/ μl stock)
2X KOD hot start master mix	25 μl
F.P	1.5 μl (10 μM stock)
R.P	1.5 μl (10 μM stock)
Molecular grade H_2O	21 μl

PCR protocol
1. 95 °C for 2 minutes
2. 95 °C for 20 seconds
3. 70 °C for 1 second
4. 50 °C for 30 seconds cooling to 50 °C at 0.5 °C per second.
5. 70 °C for 40 seconds
6. Repeat steps 2 through 5 for 24 additional cycles
7. Hold 4 °C

The PCR product from the template generation step was run on a 1% agarose gel and, the band corresponding to the template was excised and purified using Qiagen gel extraction kit, which was used to set up forward and reverse fragmentation reactions.

2. Fragmentation reaction

<u>Forward fragmentation reaction</u>	
2X KOD hot start	15 μl
Forward mutagenic primer pool	2 μl (4.5 μM stock)
Insert specific non mutagenic R.P	2 μl (4.5 μM stock)
Template (from step 1)	4 μl (3 ng/ μl)
Molecular grade H_2O	7 μl
<u>Reverse fragmentation reaction</u>	
2X KOD hot start master mix	15 μl
Reverse mutagenic primer pool	2 μl (4.5 μM stock)
Insert specific non mutagenic F.P	2 μl (4.5 μM stock)
Template (from step 1)	4 μl (3 ng/ μl)
Molecular grade H_2O	7 μl

The PCR program for these fragmentation reactions was identical to the template PCR program except that it utilized a total of 7-17 rather than 25 thermal cycles. The products from

the fragment PCR reactions were diluted 1:4 in water. These dilutions were then used for the joining PCR reactions.

3. Joining reaction

2X KOD hot start master mix	15 μl
1:4 dilution of forward fragment reaction	4 μl
1:4 dilution of reverse fragment reaction	4 μl
F.P.	2 μl
R.P.	2 μl
Molecular grade H_2O	3 μl

The products from these joining PCRs were purified over 1 % agarose gels using a Qiagen gel extraction kit. The purified products of the first joining PCR reactions were used as templates for a second round of fragmentation reactions followed by joining PCR. These second-round products were used as templates for a third round.

Cloning of combinatorial codon mutagenesis libraries: RebH and BM-3 inserts from the joining reactions were gel purified and digested with the restriction enzymes HindIII (0.33 U/μl) and NdeI (0.33 U/μl) for RebH and HindIII HF (0.33 U/μl) and EcoRI HF (0.33 U/μl) for BM-3 in 10x Cutsmart buffer in a final reaction volume of 60 μl. The digestion was conducted at 37 °C for 12-16 hours, after which digestion reactions were gel purified. This insert was ligated into previously digested pET-28a vector (insert:pET-28 ratio of 7:1) using T4 DNA ligase (1 U/ml). Ligations were conducted for 20 hours at 16 °C. Ligations were cleaned with Zymo DNA cleaning and concentrating kits and were transformed by electroporation into *E. coli* containing a plasmid encoding the chaperone pGro7 for RebH and *E.coli* BL-21(DE3) for BM-3. For cloning of POP CCM libraries, insert was gel purified following joining PCR. Vector backbone for POP CCM libraries was amplified and gel purified. Gibson assembly reactions were set up at 50 °C for 1 h in a thermocycler with insert:vector ratio of 7:1. The Gibson assembly reaction was

cleaned with Zymo DNA cleaning and concentrating kits and was transformed by electroporation into *E. coli* containing a plasmid encoding the pEVOL-pAzF plasmid.

Library colonies were picked using an automated colony picker (Norgren Systems) and arrayed into 96-deep-well plates (1 mL) containing 300 μL LB with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol for RebH & POP libraries and with 50 μg/mL kanamycin for BM-3 library. In each 96-well plate, 6 wells were left blank as a control for contamination, and 6 wells were parent cultures. Cells were grown overnight at 37 °C, 250 rpm and used for expression.

Expression and screening of codon mutagenesis libraries in 96 deep well plates:

Cytochrome P450 BM-3

Expression and lysis[39]

120 µL aliquots of overnight cultures were transferred into 2 mL, deep-well plates containing 1 mL terrific broth (TB) (100 μg/mL ampicillin and 1 µL/mL trace element solution, 1mM 5-aminlevulinic acid). The cultures were incubated at 37 °C for 4 h, and 30 min after reducing the incubation temperature to 25°C (225 rpm), 50 µL isopropyl β-D-1-thiogalactopyranoside (IPTG, 4.5 mM) was added, and the cultures were allowed to grow for another 24 h at 25 °C (225 rpm). Cells were then pelleted (3,600 rpm, 30 min, 4°C) and stored at -20 °C until further use. For cell lysis, plates were allowed to thaw for 30 min at room temperature and then cell pellets were resuspended in 300 µL phosphate buffer (0.1 M, pH = 8, 1 mg/mL lysozyme and 20 U/mL DNase I). The cells were incubated at 37 °C for 1 h, 250 rpm and then cell debris was separated by centrifugation at 3,600 rpm and 4 °C for 20 min. The resulting crude lysates were then transferred to 96-well microtiter plates for screening. For large scale expression, a colony for the selected mutant was inoculated in 5 mL LB medium with 50 µg/mL kanamycin and incubated overnight at 37 °C, 250 rpm. On the following day, 5 mL of the

overnight cultures was used to inoculate 500 mL of fresh TB media having the same antibiotics, in a 2.8L Fernbach flask. The culture was incubated at 37 ºC, 250 rpm, and protein expression was induced by adding 0.1mM IPTG when OD_{600}=1.0. The induced culture was allowed to grow for 22 hours, and then the cells were harvested by centrifugation at 4 ºC, 3600 x g for 15 minutes. Cell pellets were re-suspended in 30 mL of 25 mM Tris-HCl pH 8.0 and sonicated (40 amplitude, 30 second burst, 10 minute total time). Lysed culture was clarified by centrifugation at 16000 x g, 4 ºC for 30 minutes.

Screening [37]

Cell lysates for screening were prepared as described above. For direct measurement of product formation, a screen based on the demethylation of benzyl methyl ether (BME) and hexyl methyl ether (HME) was used. To 30 μL of crude lysate, 115 μL of phosphate buffer (0.1 M, pH 8) and 5 μL of substrate (0.4 mM in ethanol, final concentration for both HME and BME) was added. After 2 min of incubation at room temperature, NADPH (50 μL, 0.8 mM final concentration) was added. Purpald (168 mM in 2 M NaOH) was added 15 min after initiating the reaction to form a purple product with the formaldehyde that was generated upon demethylation of the substrate. The purple color was read approximately 15 min later at 550 nm using Tecan Infinite M200 pro microplate reader. Mutants were selected from these libraries based upon their absorbance at 550 nM compared to BM-3 WT.

Purification and kinetic characterization

Supernatant obtained (described above) from expression of WT and mutant in 500 mL TB media was purified by loading onto a HiTrap™ Q FF column and was washed with three column volumes (cv) of 25 mM Tris-HCl, 150 mM NaCl, pH 8.0. Bound protein was eluted with 25 mM Tris-HCl, 340 mM NaCl, pH 8.0, concentrated and buffer exchanged with 100 mM KPi,

pH 8.0 using Millipore spin filters tubes. After buffer exchange, protein samples were frozen and stored at -80 °C until further use. K_{cat} and K_m values were determined for WT and improved mutant 1DF using the substrate hexyl methyl ether (HME). Rates were determined by monitoring formation of formaldehyde which was detected using Purpald assay. Reactions contained WT (0.5 μM) or 1DF (0.25 μM), 250-950 μM HME, 0.1 M phosphate buffer, pH 8. Reactions were initiated by addition of NADPH (0.8 mM final concentration) at a final concentration of 150 μL in a microtiter plate. Reaction mixtures were quenched at 5-20 minutes by addition of 75 μL of isopropanol. All time points were collected in triplicate. Another microtiter plate with 100-500 μM formaldehyde was set up simultaneously as a calibration plate. After the last time point, purpald (168 mM in 2 M NaOH) was added to form a purple product with the formaldehyde that was generated upon demethylation of the substrate. The purple color was read 15 min later at 550 nm using Tecan Infinite M200 pro microplate reader. Product formation was determined by calculating the ratio of product from the calibration curve prepared from known concentrations of formaldehyde. The kinetic parameters (K_{cat} and K_m) for each mutant were determined using the Hanes-Woolf plots constructed from the substrate concentrations and the observed initial rates (Fig. 3.24, 3.26).

Prolyl oligopeptidase

Expression and lysis[23]

60 µL aliquots of overnight cultures were transferred into 6 mL of 2YT media in 24 well deep-well containing 50 μg/mL kanamycin and 20 μg/mL chloramphenicol. The cultures were incubated at 37 °C, 200 rpm for 6 h, and 30 min to reach OD_{600}= 1.2. Protein expression was induced by addition of 42 μL of induction solution (248 mg 4-Azido-l-phenylalanine, 5mL 20 % arabinose, 1mL 1M IPTG, 500 μL of 5 M NaOH) and the cultures were allowed to grow for

another 17 h. Cells were then pelleted (3,600 rpm, 20 min, 4°C), washed with 4 mL Tris HCl buffer (50 mM, pH 7.4) and stored at -20 °C until further use but at least for 2 h. Lysis was performed by adding 630 μL of 50 mM Tris HCl buffer containing 1 mg/mL of lysozyme and incubating the plates at 37 °C, 200 rpm, 1 h. Cells were then flash frozen in liquid nitrogen and thawed in a 37 °C water bath. DNaseI (70 μL of 1 mg/mL, 5o mM Tris HCl buffer, pH 7.4) was added, and the cells were incubated at 37 °C, 200 rpm, for 15 min followed by incubation in a 75 °C water bath. After centrifugation (3600 rpm 4 °C), 480 μL of cell lysates was transferred into 96 well filter plates for screening.

For expression in 500 mL media, single colonies for E7 and variant 4 were inoculated in 5 mL 2YT medium with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol and incubated overnight at 37 ºC , 250 rpm. On the following day, 5 mL of the overnight cultures was used to inoculate 500 mL of fresh 2YT media having the same antibiotics, in a 2.8L Fernbach flask. The culture was incubated at 37 ºC, 250 rpm, and protein expression was induced by adding 1mM IPTG, 2mM 4-Azido-l-phenylalanine and 1% (w/v) l-arabinose when OD_{600}=1.0. The induced culture was allowed to grow for 12 hours, and then the cells were harvested by centrifugation at 4 ºC, 3000 x g for 20 minutes. Cell pellets were re-suspended in 30 mL of buffer (20 mM phosphate buffer pH 7.4, containing 20 mM imidazole, and 50 mM NaCl) and sonicated (40 amplitude, 30 second burst, 10 minute total time). Lysed culture was clarified by centrifugation at 16000 x g, 4 ºC for 30 minutes and supernatant obtained was used for purification.

Screening

Screening was performed using 480 μL of crude cell lysate (obtained as described above) in a 96 well filter plate.100 μL Ni-NTA resin was added to each well of the plate, and the plate was incubated at 4 °C, 4 h, 600 rpm. Following incubation at 4 °C, 120 μL of Rh-BCN co-factor

was added and the plate was incubated overnight at 4 °C, 600 rpm. The following day, a combined stock solution containing 4-methoxystyrene (2.64 mM final concentration) and methyl 4-methoxyphenyldiazoacetate (0.53 mM final concentration) was added to initiate biocatalysis. The plate was sealed with a cap mat and incubated for 4 °C 4 h, 600 rpm. The reaction was quenched by adding 300 μL ethyl acetate and incubated at 600 rpm, 10 mins. The biphasic mixture was clarified by centrifugation at 2000 rpm, 10 mins, after which the plate was flash frozen in liquid nitrogen for 10 mins and kept at -20 °C for 10 mins. 200 μL from the top ethyl acetate layer was carefully transferred to a 200-μL 96-well polypropylene plate without interrupting the frozen aqueous layer. Ethyl acetate was evaporated in a vacuum desiccator at room temperature and 120 μL 10% isopropanol/hexane was added to redissolve the mixture and analyzed for enatioselectivity for cyclopropanation on SFC.

Purification, bioconjugation and determination of enatioselectivity

Supernatant obtained (described above) from expression of E7 and variant 4 in 500 mL TB media was purified using HisTrap HP columns and an ÄKTA protein purification system (GE healthcare). E7 and variant 4 were purified by loading supernatants onto a HisTrap™ column. The column was washed with three column volumes (cv) of buffer (20 mM phosphate buffer pH 7.4, containing 20 mM imidazole, and 50 mM NaCl). Bound protein was eluted with 20 mM phosphate buffer pH 7.4, containing 500 mM imidazole, and 50 mM NaCl. The purified protein was concentrated and buffer exchanged with 25 mM Tris HCl pH 7.4 using Millipore spin filters tubes. After buffer exchange protein samples were frozen and stored at -80 °C until further use. To set up bioconjugation for the POP-Z mutants (E7 and variant 4), purified protein (75 mM in 50mM Tris-HCl buffer, pH 7.4) and a solution of cofactor dirhodium BCN (120 μL, 0.75mM in acetonitrile, 0.655 mg ml^{-1}) were added to a 1.5-ml microcentrifuge tube and shaken

at 750 rpm, 4 °C for 4 h. The final concentrations were: 60 mM of E7 or variant 4, 150 mM dirhodium BCN, 20 vol% acetonitrile/Tris buffer. The resulting solution was treated with 100 μL azide agarose resin and rotated on a Labquake Tube Shaker/Rotator in a 4 °C cold cabinet for 24 h to remove excess cofactor. The suspension was then centrifuged at 5,000 rpm for 3 min and the supernatant was transferred to a new microcentrifuge tube. The resin was rinsed twice with 600 μL 50mM Tris HCl buffer and centrifuged at 5,000 rpm for 3 min. These supernatants were combined with the first supernatant and buffer exchanged using buffer (50 mM PIPES buffer pH 7.4 with 1.75 M NaBr) to determine percentage of bioconjugated protein scaffold. ESI-MS was used to characterize the bioconjugated protein scaffold as previously described.[5] The total protein concentration was calculated based on its absorbance at 280 nm (A_{280}) and extinction coefficient for the protein (109,210 $M^{-1}cm^{-1}$ from ExPASy). The cofactor absorbance at 280 nm is negligible relative to POP in aqueous solution under the concentrations used. Efficiency of dirhodium incorporation to form POP artificial metalloenzyme (ArM) was calculated based on the ratio of the high resolution ESI-MS peak intensity of the ArM and scaffold ($I_{ArM}/ (I_{ArM}+I_{scaffold})$). The effective ArM concentration was calculated by multiplying the total protein concentration by the efficiency of dirhodium incorporation ([ArM] = [Total protein]*($I_{ArM}/ (I_{ArM}+I_{scaffold})$). The effective ArM loading was adjusted to 1 mol% with respect to the dirhodium cofactor in bioconversion reactions. To set up bioconversion reaction, a solution of aryldiazoacetate (25 μL, 96mM, in THF), styrene (25 μL, 485mM, in THF) and E7 and variant 4 ArM solution (500 μL, 48 mM) were added to a 1.5-ml microcentrifuge tube. The final concentrations of the reagents were as follows: 22mM olefin, 4.4mM aryldiazoacetate, 44 mM ArM (E7 or variant 4). The resulting mixture was incubated at 4 °C, 750 rpm for 4 h. The reactions were quenched by adding 20 ml 1,2,4-trimethoxybenzene solution (30mM, in THF) and 600 μL ethyl acetate. The

mixture was vortexed and centrifuged (15,000g, 3 min). The top organic layer was collected and the bottom aqueous layer was extracted with 600 μL ethyl acetate twice. The organic layers were combined, evaporated and re-dissolved in 200 μL 90/10 hexanes/isopropanol, and analyzed on NP-HPLC to determine conversions and enatioselectivity. The conversions and enatioselectivity are reported as average of three trials from the same batch of ArM reactions set up in parallel. NP-HPLC to determine enatioselectivity was performed on Agilent 1100 series HPLC system using a Phenomenex Lux® 3u Cellulose-1 column (1000 Å, 3 μM, 4.6 mm i.d. x 250 mm), with a flow rate of 1 mL/min and detection wavelength set at 230 nm. The following gradient was used: 10 % to 70 % B from 0-10 min, 70 % B from 10-15 min, 70 % to 100 % B from 15-18 min, 100 % B from 18-22 min, 4 min post-run (solvent A: 2-propanol; solvent B: hexanes).

Tryptophan halogenase RebH

Library expression, lysis, and screening[3]

100 μL of overnight culture was used to inoculate 1 mL TB (with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol) in 2-mL 96-well plates. Following growth at 37 °C, 250 rpm, to an OD_{600} = 0.8-1, enzyme expression was induced with IPTG and arabinose to final concentrations of 10 μM and 0.2 mg/mL, respectively. Protein expression continued for ~20 h at 30 °C, 250 rpm, after which cultures were harvested by centrifugation and stored at -20 °C until use. Cell pellets were thawed and suspended in 100 μL HEPES buffer (25 mM, pH 7.4) containing 0.75 mg/mL lysozyme. After incubation at 37 °C, 250 rpm, cells were flash frozen in liquid nitrogen and thawed in a 37 °C water bath. Ten microliters of DNaseI at 1 mg/mL were added and the cells incubated at 37 °C, 250 rpm, for 15 min. After centrifugation, 50 μL of supernatant were transferred to a microtiter plate for screening.

Similar to what has been described previously for halogenation reactions,[6] MBP-RebF (0.0017 equiv., 2.5 μM final concentration) and glucose dehydrogenase (9 U/mL final concentration) were added as solutions (25 mM HEPES, pH 7.4) to the RebH lysate. A solution containing 2-aminobenzoic acid (1 equiv., 1.5 mM final concentration), NAD (0.067 equiv., 100 μM final concentration), FAD (0.067 equiv., 100 μM final concentration), NaCl (66.7 equiv., 100 mM final concentration), and glucose (13.3 equiv., 20 mM final concentration) was added via multichannel pipette to simultaneously initiate the reactions (final reaction volume of 75 μL). Reactions were shaken for approximately 2 hours after which they were quenched with an equal volume of methanol and centrifuged. The supernatant was filtered and analyzed via HPLC (Agilent 1200 UHPLC with an Agilent Eclipse Plus C18 2.1 x 50 mm column, 1.8 μM particle size; solvent A = H_2O/0.1 % TFA, solvent B = CH_3CN; 0-1 min, B = 10-16 %; 1-2 min, B = 16-80 %; 2-2.5 min, B = 80-100%).

Expression, purification and kinetic characterization [3]

For expression and purification procedures for analytical bioconversions and kinetic characterization, an overnight starter culture was used to inoculate 50 mL TB (with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol). Following growth at 37 °C, 250 rpm, until OD_{600} = 0.6-0.8, enzyme expression was induced with IPTG and arabinose to final concentrations of 100 μM and 2 mg/mL, respectively. Protein expression continued for ~20 h at 30 °C, 250 rpm, after which cultures were harvested by centrifugation and stored at -20 °C until use. Cell pellets were thawed, suspended in 10 mL 25 mM HEPES (pH 7.4), and lysed by sonication while kept on ice (Qsonica S-4000 with a 0.5" horn; 8 x 30 s with 45 s rests, 20 % duty cycle delivering 40-50 W).

After clarification by centrifugation 0K and 1K were purified by Ni-NTA affinity chromatography and exchanged into HEPES buffer (25 mM, pH 7.4, 10% v/v glycerol). Protein

concentration was determined using A_{280} and extinction coefficients calculated based on amino acid composition (Protein Calculator v3.3,http://www.scripps.edu/~cdputnam/protcalc.html)

For kinetic characterization, initial rates were determined by monitoring the conversion of 100-800 μM 2-aminobenzoic acid in the presence of NAD (100 μM final concentration), FAD (100 μM final concentration), NaCl (100 mM final concentration), MBP-RebF (2.5 μM final concentration), glucose dehydrogenase (9 U/mL final concentration GDH), glucose (20 mM final concentration), and benzoic acid as an internal standard (0.5 mM final concentration) at a final volume of 75 μL in a microtiter plate. RebH was added at a final concentration of 5 μM. Plates were shaken at 650 rpm at room temperature. Reaction mixtures were quenched at 5-20 minutes by addition of 75 μL of MeOH. All time points were collected in triplicate. The precipitated protein was then removed by centrifugation and the reactions were filtered and analyzed by HPLC (Agilent 1200 UHPLC with an Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H_2O/0.1 % TFA, solvent B = CH_3CN; 0-5 min, B = 15 %; 5-9 min, B = 15-45 %; 9-12 min, B = 45-100%). Product formation was determined by calculating the ratio of product to internal standard and fitting that value to a calibration curve prepared from known concentrations of chlorinated 2-aminobenzoic acid. The kinetic parameters (K_m and k_{cat}) for each mutant were determined using the Hanes-Woolf plots constructed from the substrate concentrations and the observed initial rates (Fig. 3.28, 3.30).

Specific procedures for colorimetric assay in lysate

FDH libraries were expressed as previously reported.[3] Bioconversions were conducted as previously reported, using anthranilic acid (1 mM final concentration) as a substrate and 10-50 μL cell lysate as the FDH source.[3] For calibration curves, reaction components were not added to some wells. After 1 hour, protein in halogenation reactions was quenched by heat treatment at 95

°C for 10 min. Different ratios of anthranilic acid and 3-chloro-anthranilic acid were added to the calibration curve wells and mixed. Reactions were centrifuged for 10 min at 3,600 rpm. 40 μL of the supernatants were transferred to new 96-well microtiter plates. To these plates was added 4-methylcatechol (0.667 mM final concentration) as a solution in HEPES buffer (25 mM, pH = 7.4). Coupling of 4-methylcatechol to anthranilic acids was initiated by adding horseradish peroxidase (4.2 μg/mL final concentration) and H_2O_2 (3.75 mM final concentration) as a solution in HEPES buffer (25 mM, pH = 7.4). Final reaction volume was 60 μL. Reactions were mixed at 650 rpm at room temperature for 10-30 minutes. Bubbles were removed by misting reactions with 70% EtOH and the absorbance was measured at 535 nm. Wells with lower absorbance at 535 nm were considered hits. Hits were expressed, lysed, and purified as previously reported.[3] Purified enzymes were used for bioconversions on 5-X-2-aminobenzoic acids (Fig. 3.10). Bioconversions were analyzed by UPLC Method 1, as described in the general procedures.

Specific procedures for MALDI MS screen

Kinetic isotope effect of deuterium labeling: In order to affirm deprotonation of the arene occurs after the rate-limiting step, and thus has little to no contribution to the observed selectivity on the deuterated tryptamine probes, initial rates of L-tryptophan and d_5-L-tryptophan chlorination were compared. RebH variant 0S (0.002 equiv., 1 μM final concentration), MBP-RebF (0.005 equiv., 2.5 μM final concentration), glucose dehydrogenase (9 U/mL final concentration), and FAD (0.2 equiv., 100 μM final concentration) were added as solutions (25 mM HEPES, pH 7.4) to an Eppendorf tube. A solution containing L-tryptophan or d_5-L-tryptophan (1 equiv., 0.5 mM final concentration), NAD (0.2 equiv., 100 μM final concentration), NaCl (200 equiv., 100 mM final concentration), phenol (internal standard, 1 equiv., 0.5 mM final concentration) and glucose (40 equiv., 20 mM final concentration) was added to this tube to initiate reaction (1200 μL final

reaction volume for each substrate). This reaction was then split into 75 μL aliquots in a 96-well microtiter plate. The L-tryptophan reaction was initiated and pipetted into the microtiter plate first, followed by *d5*-L-tryptophan. The reactions were mixed at 650 rpm on top of an Eppendorf air bath and were quenched with 1 volume (75 μL) methanol at various time points from 10-45 min. Reactions were conducted in triplicate. The precipitated protein was removed by centrifugation and the reactions were filtered and analyzed by UPLC Method 2 from the General Procedures. Product formation was obtained by fitting the ratio of product to internal standard to a calibration curve prepared from known concentrations of chlorinated tryptophan. The initial rates of L-tryptophan and d_5-L-tryptophan were found to be 2.1076 and 2.0257 μM/min, respectively (Fig. 3.14). This suggests there is no significant isotope effect for RebH halogenation of H/D tryptophan ($k_H/k_D = 1.04$). Note: Because of the time difference between the two reaction initiations, the conversions of L-tryptophan in Fig. 3.14 are slightly higher than *d5*-L-tryptophan. The number of minutes in between time points, however, is identical in both cases, thus the initial rates can be accurately determined from the slopes.

In a second experiment, d_5-L-tryptophan and L-tryptophan were added in equal amounts (1 equiv., 0.5 mM final concentration each) to reactions using the same method as described above. Reactions were quenched with methanol at times between 9-45 min. The precipitated protein was removed by centrifugation. The reactions were filtered and analyzed by LC-MS using Method 2 described in the General Procedures. Product formation was obtained by fitting the ratio of product to internal standard to a calibration curve prepared from known concentrations of chlorinated tryptophan. The ratio of H/D-chlorotryptophan was determined by finding the relative intensities of m/z = 239 and m/z = 243 in the mass spectrum of the product peak. Again, no significant isotope effect was observed ($k_H/k_D = 0.97$, Fig. 3.15).

Using MALDI MS to analyze RebH bioconversions: To determine whether reaction conversion could be measured by MALDI MS, a time course of wtRebH chlorination of tryptamine was conducted and analyzed by UPLC and MALDI MS. wtRebH (0.05 equiv., 25 μM final concentration), MBP-RebF (0.005 equiv., 2.5 μM final concentration), and glucose dehydrogenase (9 U/mL final concentration) were added as solutions (25 mM HEPES, pH 7.4) to an Eppendorf tube. A solution containing tryptamine (1 equiv., 0.5 mM final concentration), FAD (0.2 equiv., 100 μM final concentration), NAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), and glucose (40 equiv., 20 mM final concentration) was added to this tube to initiate reaction (600 μL final reaction volume). This reaction was then split into 75 μL aliquots in Eppendorf tubes. These were agitated at 650 rpm in an air bath at 25 °C. Reactions were quenched with 1 volume (75 μL) methanol at various time points from 30-210 min. Precipitated protein was removed by centrifugation, and the reactions were filtered according to the method found in the General Procedures. Ten microliters of 75 mM HCl was added to each reaction to slightly acidify it, which allowed spotting onto a MALDI target without re-dissolving the pre-spotted matrix.

A MALDI target was spotted by hand with matrix, followed by the acidified bioconversions using the method described in the General Procedures. Spots were analyzed using the MALDI MS automatic method outlined in the General Procedures. Percent conversion was calculated by taking the intensity of the m/z peak for chlorotryptamine (195) divided by the added intensities of the chlorotryptamine peak (195) and tryptamine (161). After the MALDI target plate had been spotted with reaction mixtures, these same reactions were analyzed by UPLC using Method 1 described in the General Procedures. Percent conversion was calculated

by taking the area of the product peak divided by the area of the combined product and starting material peaks (Fig. 3.17).

Detection of different H/D-tryptamine ratios: Once deuterated Probe **1** had been synthesized, the reliability of MALDI MS to distinguish tryptamine from Probe **1** was explored. Solutions containing tryptamine (0.5 mM final concentration), FAD (100 μM final concentration), NAD (100 μM final concentration), NaCl (10 mM final concentration), and glucose (20 mM final concentration) were made in 25 mM HEPES, pH 7.4. Each solution contained a different amount of Probe **1** (0.01-0.5 mM final concentration). These solutions (75 μL each) were acidified (10 uL of 75 mM HCl) and 1 volume of methanol was added to each to mimic bioconversion preparation. The acidified solutions were spotted by hand on a MALDI target plate and shot with the automatic method described in the General Procedures. The ratio of the intensity of the m/z peak corresponding to Probe **1** (162) and the intensity of the m/z peak corresponding to tryptamine (161) was calculated and plotted against the known concentration of Probe **1** (Fig. 3.18).

3.4.3 Detailed isolation and characterization

Synthesis and characterization of deuterated probes

*N,N-bis(phenylmethyl)-1H-Indole-3-ethanamine (**1a**):* Tryptamine (2 g, 12.48 mmol, 1 equiv.) and K_2CO_3 (4.5 g, 32.56 mmol, 2.6 equiv.) were added to a round bottom flask under nitrogen. Dry methanol (60 mL) was added, followed by benzyl bromide (3.12 mL, 26.24 mmol, 2.1 equiv.). The reaction mixture was heated to reflux for 12 hours. The crude reaction mixture was passed through filter paper and the filtrate was concentrated onto silica gel. Purification by flash chromatography (SiO_2, 20% ethyl acetate/hexanes) afforded the known compound[37] **1a** in 79% yield (3.36 g, 9.88 mmol). ^{1}H NMR (500 MHz; $CDCl_3$): δ 7.88 (s, 1H), 7.47-7.42 (m, 5H), 7.38-

7.35 (m, 5H), 7.29 (t, J = 7.3, 2H), 7.21 (t, J = 7.6, 1H), 7.09 (td, J = 7.5, 0.9, 1H), 6.94 (d, J = 2.1, 1H), 3.76 (s, 4H), 3.04 (t, J = 7.9, 2H), 2.87 (t, J = 7.8, 2H). HRMS (ESI-TOF) calcd for $C_{24}H_{24}N_2$ $[M + H]^+$: 341.2012, found: 341.0735.

*7-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-N,N-bis(phenylmethyl)-1H-Indole-3-ethanamine **(1b)**:* The procedure for this reaction was adapted from a previous report.[34] **1a** (1.8651 g, 5.49 mmol, 1 equiv.), $[Ru(p\text{-cymene})Cl_2]_2$ (50 mg, 0.0823 mmol, 0.015 equiv.), diethylsilane (1.07 mL, 8.235 mmol, 1.5 equiv.) and toluene (2.75 mL) were added to a dry Teflon-sealed Schlenk flask inside a glovebox. The reaction mixture was stirred at room temperature until full conversion to the *N*-silyltryptamine was observed by ^{1}H NMR (~10 hours). The solvent was removed under vacuum. $[Ir(cod)Cl]_2$ (9.2 mg, 0.0137 mmol, 0.0025 equiv.), dtbpy (7.9 mg, 0.0275 mmol, 0.005 equiv.), B_2pin_2 (1.394 g, 5.49 mmol, 1 equiv.), HBpin (39.9 μL, 0.275 mmol, 0.05 equiv.) and THF (5.49 mL) were added to the flask. The reaction was removed from the glovebox and stirred at 80 °C until full conversion was observed by ^{1}H NMR. After the reaction was cooled, MeOH (~5.5 mL) was added. This was stirred until complete desilylation was observed by ^{1}H NMR. The reaction mixture was concentrated onto silica gel. Purification by flash chromatography (SiO_2, 8% ethyl acetate/hexanes) afforded **1b** in 68% yield (1.733 g, 3.733 mmol). ^{1}H NMR (500 MHz; $CDCl_3$): δ 8.95 (s, 1H), 7.60 (d, *J* = 7.0 Hz, 1H), 7.48 (d, *J* = 7.8 Hz, 1H), 7.39 (d, *J* = 7.4 Hz, 4H), 7.30 (t, *J* = 7.5 Hz, 4H), 7.23 (t, *J* = 7.2 Hz, 2H), 7.03 (t, *J* = 7.5 Hz, 1H), 6.97 (s, 1H), 3.69 (s, 4H), 2.99 (t, *J* = 7.8 Hz, 2H), 2.80 (t, *J* = 7.8 Hz, 2H), 1.38 (s, 12H). ^{13}C NMR (126 MHz; $CDCl_3$): δ 141.61, 140.14, 129.34, 129.02, 128.37, 126.97, 126.69, 122.67, 121.57, 118.78, 114.25, 83.92, 58.59, 54.45, 25.20, 23.42. ^{11}B NMR (160 MHz;

$CDCl_3$): δ 31.33. HRMS (ESI-TOF) calcd for $C_{30}H_{35}N_2O_2B$ [M + H]+: 467.2870, found: 467.1362.

*7-deutero- N,N-bis(phenylmethyl)-1H-Indole-3-ethanamine (**1c**):* In a glovebox, **1b** (1 g, 2.15 mmol, 1 equiv.) and $[Ir(cod)Cl]_2$ (44 mg, 0.0655 mmol, 0.03 equiv.) were added to a dry round bottom flask. The flask was taken out of the glovebox and placed under nitrogen. Dry THF (9.5 mL) and D_2O/NaOAc (38.8 mL, 3M) were added. The reaction mixture was stirred at room temperature until full conversion of the starting material was observed by TLC (~48 hours). Purification by flash chromatography (SiO_2, 20% ethyl acetate/hexanes) afforded **1c** in 90% yield (659 mg, 1.93 mmol). ^{1}H NMR (500 MHz; $CDCl_3$): δ 7.88 (s, 1H), 7.43-7.36 (m, 5H), 7.30 (t, J = 7.2, 4H), 7.23 (t, J = 7.1, 2H), 7.16 (d, J = 7.0, 1H), 7.06-7.01 (m, 1H), 6.92 (d, J = 1.0, 1H), 3.70 (s, 4H), 2.99 (t, J = 7.7, 2H), 2.82 (t, J = 7.7, 2H). ^{13}C NMR (126 MHz; $CDCl_3$): δ 139.90, 136.22, 128.96, 128.29, 127.64, 126.96, 121.79, 121.53, 119.17, 118.91, 114.52, 110.78 (t), 58.40, 53.94, 23.10. HRMS (ESI-TOF) calcd for $C_{24}H_{23}DN_2$ [M + H]+: 342.2075, found: 342.0788.

*7-deutero-1H-Indole-3-ethanamine (Probe **1**):* **1c** (250 mg, 0.733 mmol, 1 equiv.) and Pd/C (150 mg, 10 wt. % loading on activated carbon) were added to a round bottom flask under nitrogen. Dry MeOH (25 mL) was added as solvent. The flask was purged with hydrogen for 15 minutes, after which the reaction was kept under 1 atm of hydrogen by balloon. The reaction was stirred at room temperature and monitored by HPLC until full conversion to the debenzylated product was observed. The reaction mixture was filtered through Celite and subsequently concentrated onto Celite. The Celite was packed into a Biotage samplet, which

was then loaded into a reverse phase column (Biotage SNAP-KP- C18-HS). The crude material was purified by reverse phase chromatography (gradient from water 0.1% TFA to 15% CH_3CN/water 0.1%TFA). When concentrated by rotovap, the deuterium label is partially lost due to proton exchange. Instead, 7-deutero-tryptamine-containing fractions were pooled, basified and extracted into DCM. The DCM was concentrated by rotovap to afford Probe **1** in 68% yield (80.2 mg, 0.499 mmol). ^{1}H NMR (500 MHz; MeOD): δ 7.56 (d, J = 7.8, 1H), 7.16 (s, 1H), 7.12 (d, J = 6.9, 1H), 7.04 (t, J = 7.5, 1H), 3.22 (t, J = 7.1, 2H), 3.11 (t, J = 7.2, 2H). ^{13}C NMR (126 MHz; MeOD): δ 138.21, 128.75, 123.54, 122.22, 119.56, 119.26, 113.40, 111.99 (t), 43.07, 29.52. HRMS (ESI-TOF) calcd for $C_{10}H_{11}DN_2$ [M + H]+: 162.1136, found: 162.2032.

*5-deutero-indole (**2a**):* In a glovebox, 5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1H-indole (5 g, 20.6 mmol, 1 equiv.) and $[Ir(cod)Cl]_2$ (138 mg, 0.206 mmol, 0.01 equiv.) were added to a dry round bottom flask. The flask was taken out of the glovebox and placed under nitrogen. Dry THF (20 mL) and D_2O/NaOAc (100 mL, 3M) were added, and the mixture was stirred at room temperature until full conversion of starting material was observed by TLC (~48 hours). Purification by flash chromatography (SiO_2, 5% ethyl acetate/hexanes) afforded **2a** in 78% yield (1.886 g, 16.0 mmol). ^{1}H NMR (500 MHz; $CDCl_3$): δ 8.01 (s, 1H), 7.74 (s, 1H), 7.41 (d, J = 8.2, 1H), 7.28 (d, J = 8.2, 1H), 7.20 (t, J = 2.8, 1H), 6.63 (m, 1H). ^{13}C NMR (126 MHz; CD_2Cl_2): δ 136.37, 128.41, 124.71, 122.23, 120.90, 119.89 (t), 111.47, 102.81. HRMS (ESI-TOF) calcd for C_8H_6DN [M + H]+: 119.0714, found: 118.9099.

*It should be noted that some amount of deuteration at the C-3 position was occasionally observed at longer reaction times. Because the C-3 position is deprotonated during alkylation with oxalyl chloride, this was not an issue for the synthesis of this tryptamine probe. If

necessary, the 3-deuteration can be removed by reaction of product with 0.01 M HCl at 60 °C for 4 hours.[41]

*7-deutero-α-oxo-H-Indole-3-acetamide (**2b**):* The procedure for this reaction was adapted from a previous report.[42] Oxalyl chloride (304 μL, 3.54 mmol, 3 equiv.) was added dropwise to a stirring solution of **2a** (140 mg, 1.19 mmol, 1 equiv.) in dry ether (10.5 mL) under nitrogen at 0 °C. The reaction mixture was then warmed to room temperature and stirred until full conversion of indole was observed by TLC. Stirring was stopped, and the reaction mixture was allowed to settle. The supernatant was removed and discarded. The precipitant was placed back under nitrogen, and ammonia in dioxane (14 mL, 0.5 M) was added. This mixture was stirred at room temperature for 14 hours. $KHCO_3$ was added, and the reaction mixture was filtered. Filter paper was rinsed with hot acetone. The combined filtrates were concentrated onto silica gel. Purification by flash chromatography (SiO_2, 50% ethyl acetate/hexanes) afforded **2b** in 60% yield (136 mg, 0.72 mmol). ^{1}H NMR (500 MHz; DMSO): δ 12.19 (s, 1H), 8.68 (d, J = 3.0, 1H), 8.22 (s, 1H), 8.07 (s, 1H), 7.70 (s, 1H), 7.53 (d, J = 8.1, 1H), 7.26 (d, J = 8.1, 1H). ^{13}C NMR (126 MHz; DMSO): δ 183.43, 166.48, 138.66, 136.77, 126.62, 123.72, 122.65 (t), 121.61, 112.97, 112.59. HRMS (ESI-TOF) calcd for $C_{10}H_7DN_2O_2$ [M + H]+: 190.0721, found: 190.0727.

*5-deutero-1H-Indole-3-ethanamine (Probe **2**):* The procedure for this reaction was adapted from a previous report.[42] In a glovebox, lithium aluminum hydride (271.4 mg, 7.2 mmol, 10 equiv.) was added to a solution of **2b** (136 mg, 0.72 mmol, 1 equiv.) in dry dioxane (44 mL). The reaction was removed from the glovebox and refluxed at 110 °C under nitrogen until full conversion to 5-deutero-tryptamine was observed by HPLC. To quench the excess LAH, H_2O

(272 µL) was added slowly, followed by addition of NaOH (272 µL 15% w/v) and H_2O (816 µL). The reaction mixture was stirred for 1 hour, then filtered over Celite through a coarse fritted glass filter and washed with methanol. The filtrate was concentrated onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography (gradient from water 0.1% TFA to 15% CH_3CN/water 0.1%TFA). When concentrated by rotovap, the deuterium label is partially lost due to proton exchange. Instead, 5-deutero-tryptamine-containing fractions were pooled, basified and extracted into DCM. The DCM was concentrated by rotovap to afford Probe **2** in 40% yield (46.2 mg, 0.288 mmol). ^{1}H NMR (500 MHz; $CDCl_3$): δ 8.58 (s, 1H), 7.63 (s, 1H), 7.35 (d, J = 8.1, 1H), 7.21 (d, J = 8.2, 1H), 7.00 (s, 1H), 3.05 (t, J = 6.6, 2H), 2.93 (t, J = 6.6, 2H). ^{13}C NMR (126 MHz; $CDCl_3$): δ 136.61,127.62, 122.20, 121.92, 118.82, 113.69, 111.28, 42.44, 29.57. HRMS (ESI-TOF) calcd for $C_{10}H_{11}DN_2$ [M + H]+: 162.1136, found: 162.1142.

3.4.4 Additional data

Figure 3.19: Screening using HME.

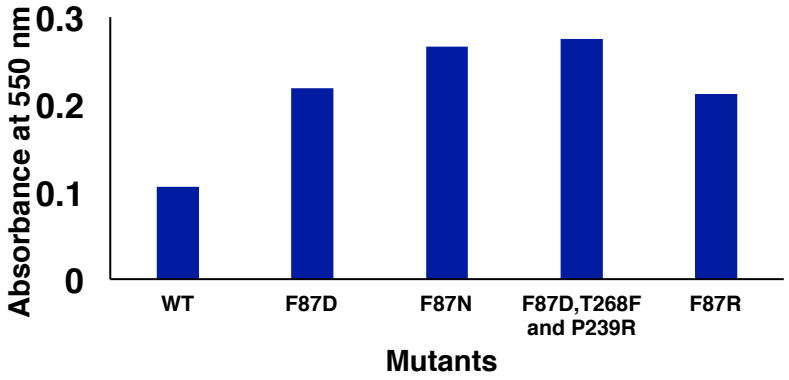


Figure 3.20: Screening using BME.

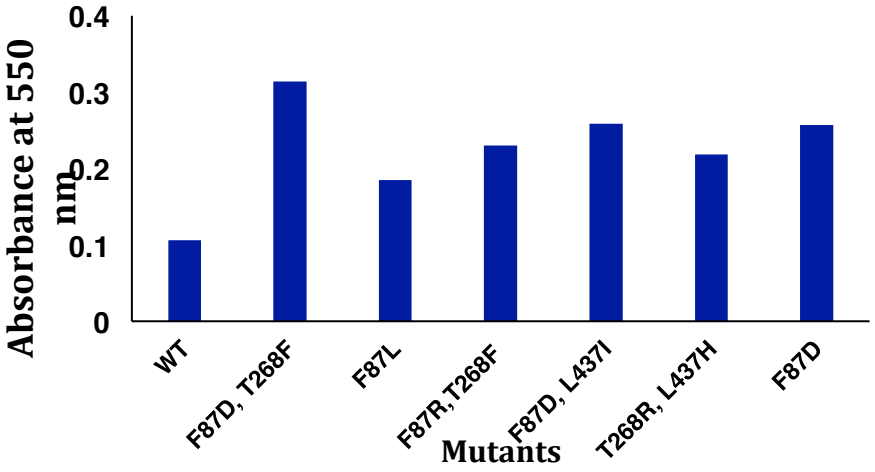


Figure 3.21: Activity of purified protein variants with HME.

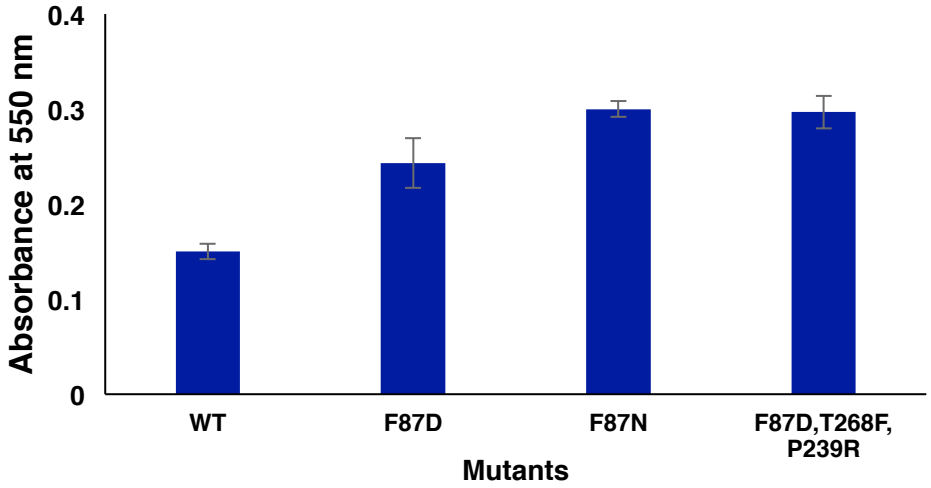


Figure 3.22: Activity of purified protein variants with BME.

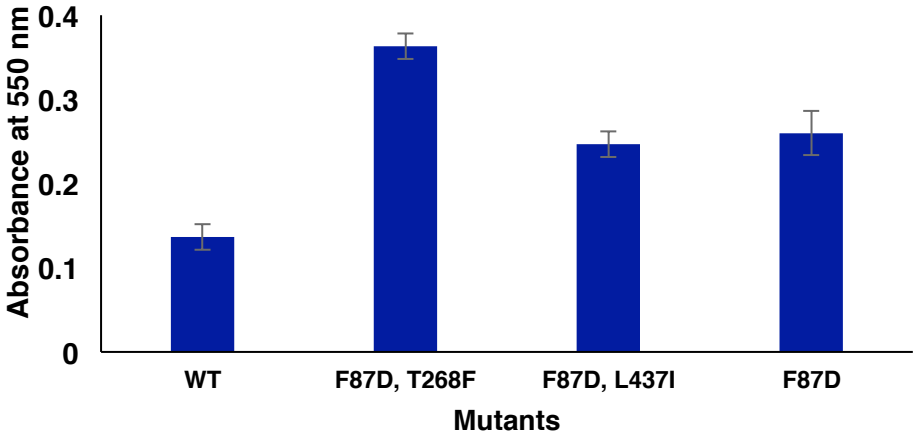


Figure 3.23: Saturation plot for 1DF conversion of BME.

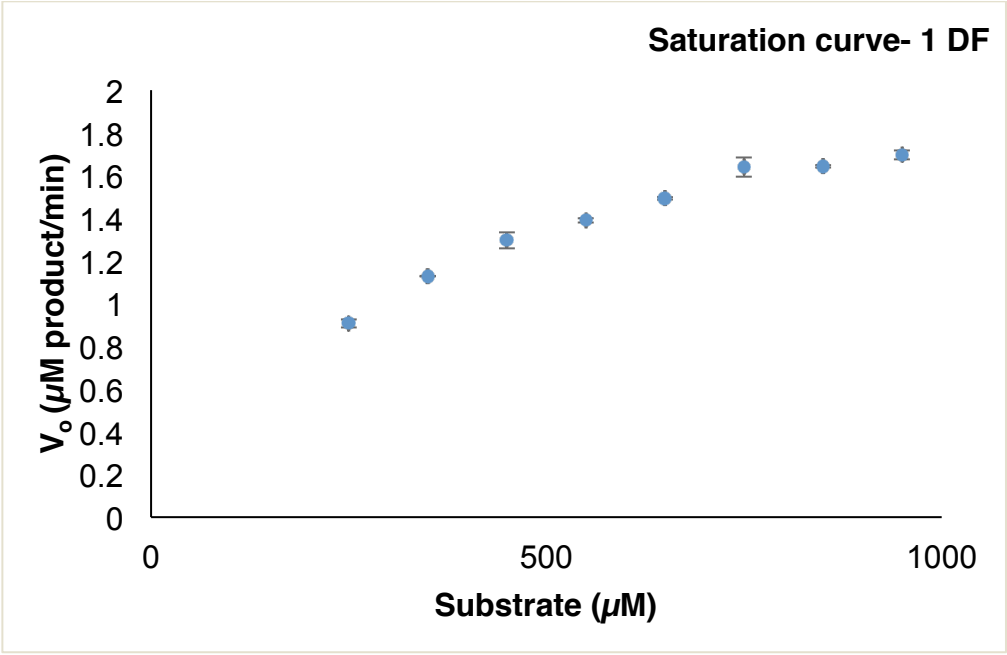


Figure 3.24: Hanes-Woolf plot for 1DF conversion of BME.

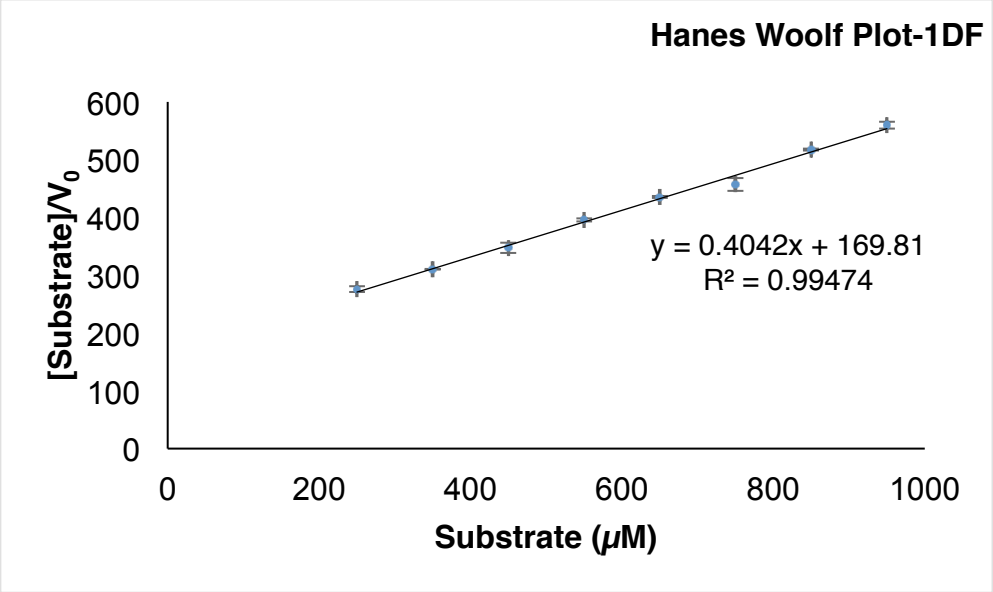


Figure 3.25: Saturation plot for BM3 conversion of BME.

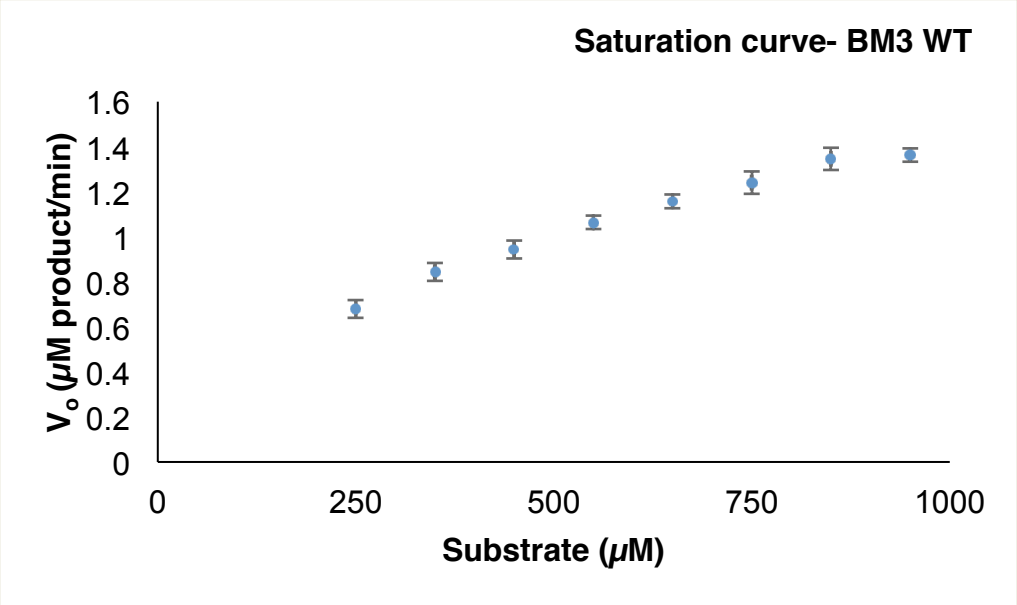


Figure 3.26: Hanes-Woolf plot for BM3 conversion of BME.

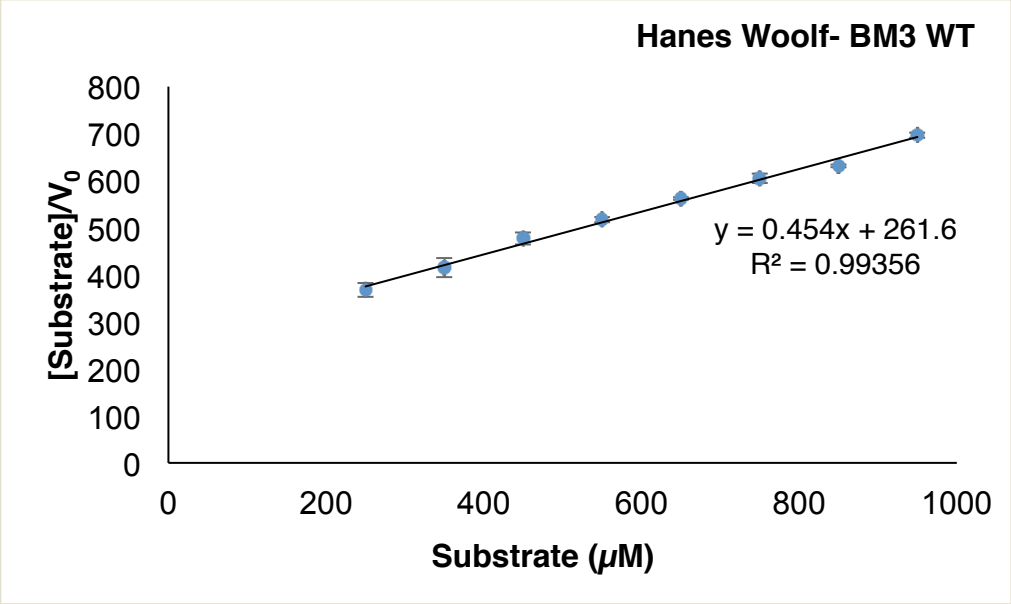


Figure 3.27: Saturation plot for 1K conversion of anthranilic acid.

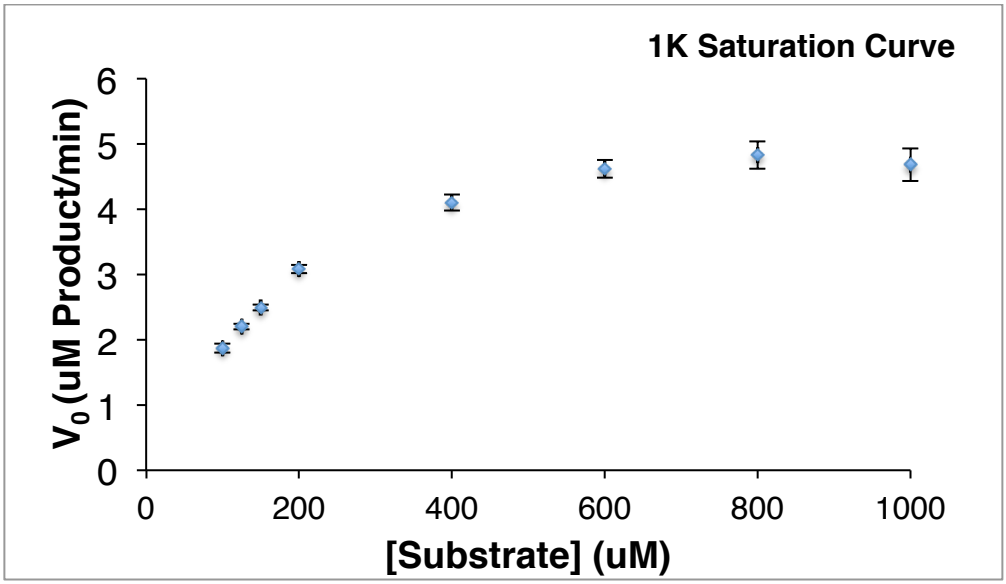


Figure 3.28: Hanes-Woolf plot for 1K conversion of anthranilic acid.

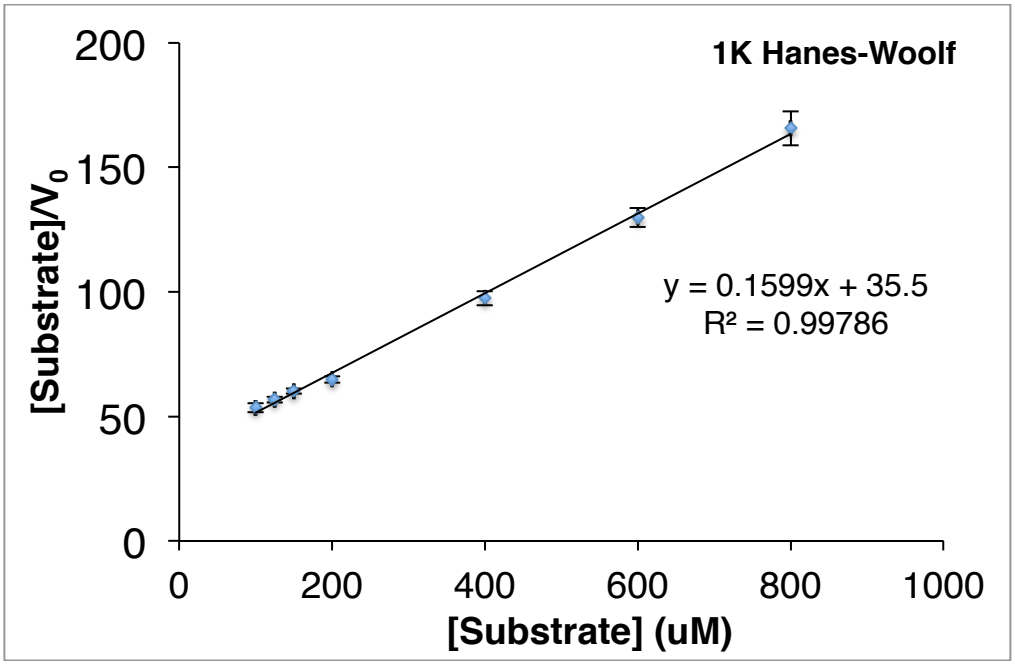


Figure 3.29: Saturation plot for 0K conversion of anthranilic acid.

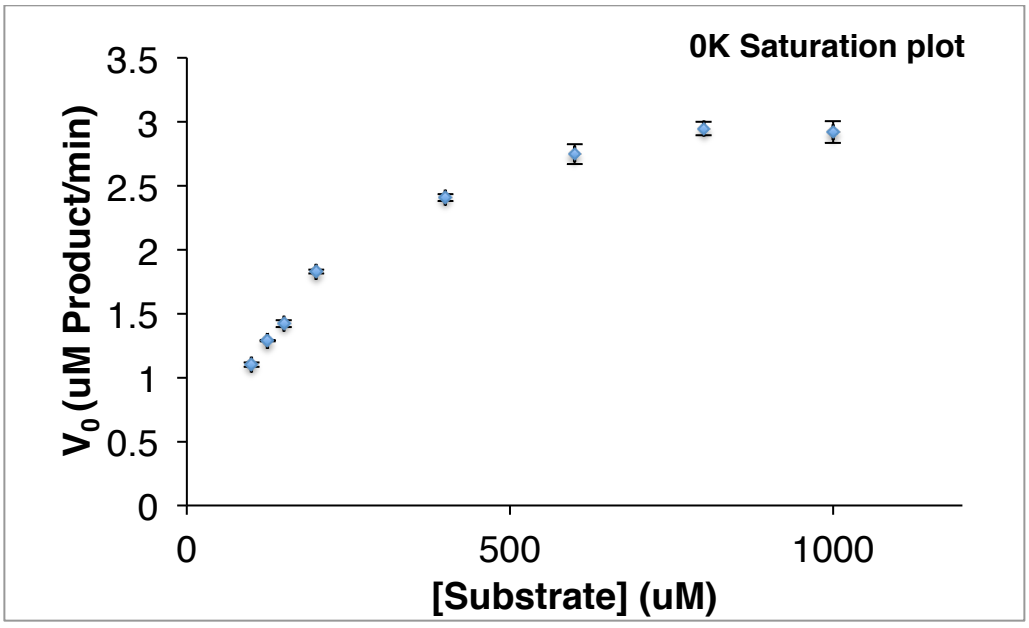


Figure 3.30: Hanes-Woolf plot for 0K conversion of anthranilic acid.

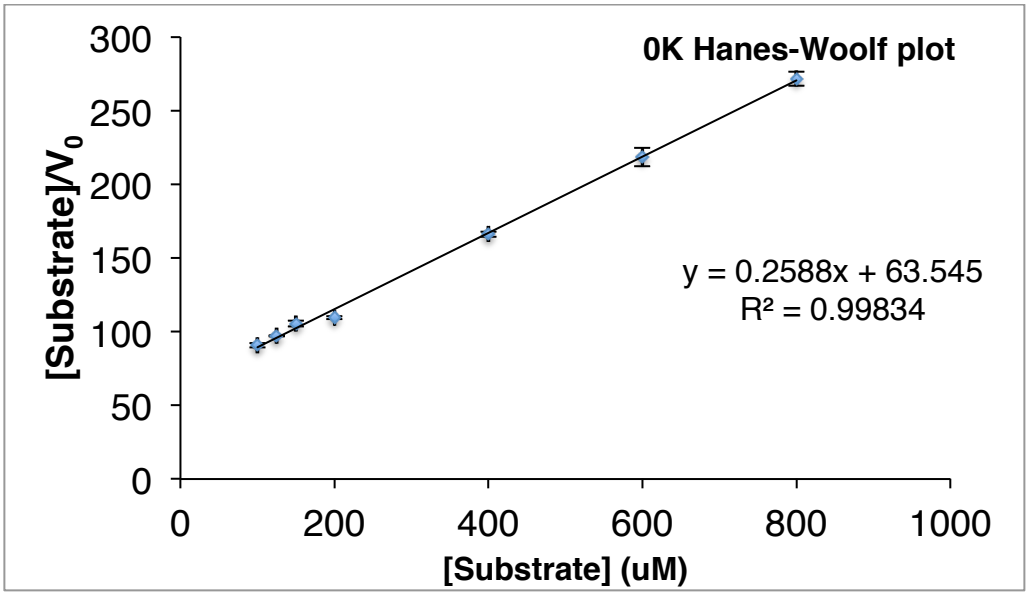


Figure 3.31: Saturation plot for 0K conversion of 5-methyl-anthranilic acid.

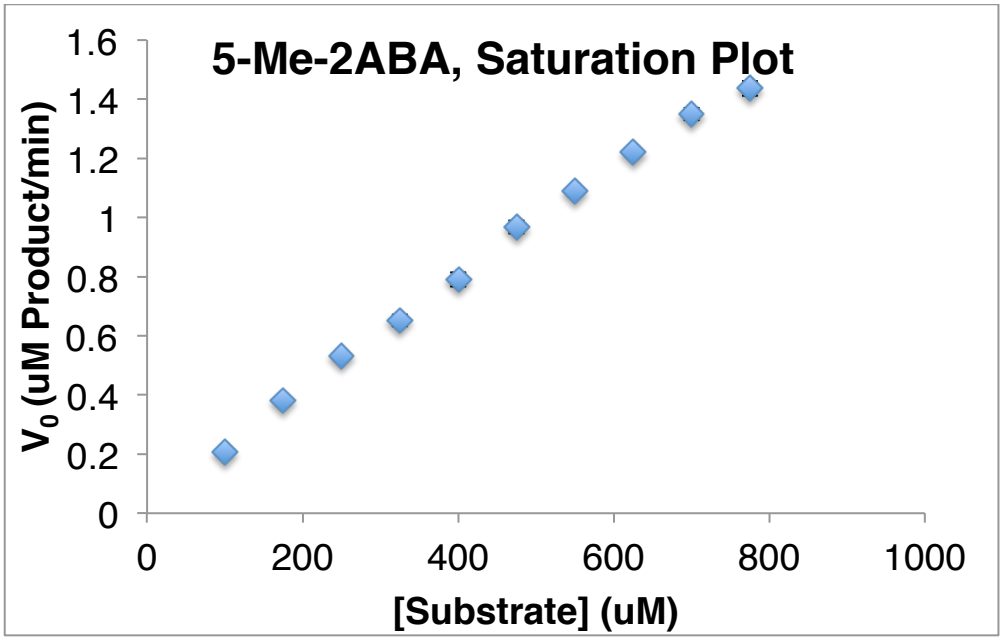
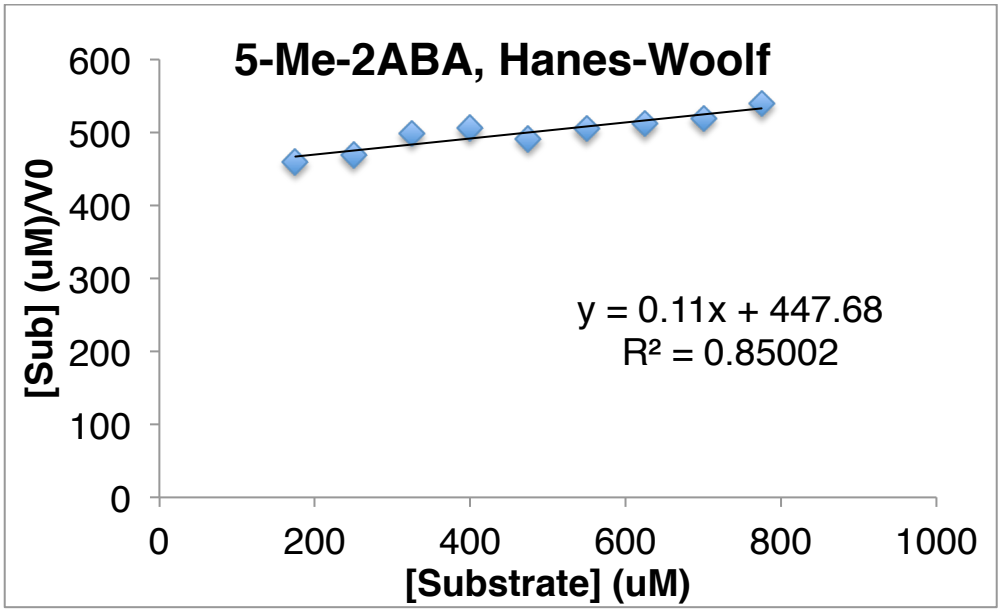


Figure 3.32: Hanes-Woolf plot for 0K conversion of 5-methyl-anthranilic acid.



Acknowledgements

We would like to Dr. Hyun June Park for development of the POP artificial metalloenzyme screen and Dr. Jaime Cabrera-Pardo for advice on automation of the Bruker Ultraflextreme MALDI-TOF-TOF assay.

3.5 References

1. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10,** 866–876 (2009).
2. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16,** 379–394 (2015).
3. Poor, C. B., Andorfer, M. C. & Lewis, J. C. Improving the Stability and Catalyst Lifetime of the Halogenase RebH By Directed Evolution. *ChemBioChem* **15,** 1286–1289 (2014).
4. Payne, J. T., Poor, C. B. & Lewis, J. C. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angew. Chem.* **127,** 4300–4304 (2015).
5. Cirino, P. C., Mayer, K. M. & Umeno, D. Generating mutant libraries using error-prone PCR. *Methods Mol. Biol.* **231,** 3–9 (2003).
6. Siloto, R. M. P. & Weselake, R. J. Site saturation mutagenesis Methods and applications in protein engineering. *Biocatal. Agric. Biotechnol.* **1,** 181–189 (2012).
7. Denard, C. A., Ren, H. & Zhao, H. Improving and repurposing biocatalysts via directed evolution. *Curr. Opin. Biotechnol.* **25,** 55–64 (2015).
8. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2,** 924–932 (2007).
9. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2,** 891–903 (2007).
10. Stemmer, W. P. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci.* **91,** 10747–10751 (1994).
11. Herman, A. & Tawfik, D. S. Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.* **20,** 219–226 (2007).
12. Firnberg, E. & Ostermeier, M. PFunkel: Efficient, Expansive, User-Defined Mutagenesis. *PLoS ONE* **7,** e52031–10 (2012).
13. Jin, P. *et al.* Combinatorial Evolution of Enzymes and Synthetic Pathways Using One-Step PCR. *ACS Synth. Biol.* **5,** 259–268 (2016).
14. Lewis, J. C. *et al.* Combinatorial Alanine Substitution Enables Rapid Optimization of Cytochrome P450BM3 for Selective Hydroxylation of Large Substrates. *ChemBioChem* **11,** 2502–2505 (2010).
15. Morrison, K. L. & Weiss, G. A. Combinatorial alanine-scanning. *Curr. Opin. Biotechnol.* **5,** 302–307 (2001).
16. Bloom, J. D. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol. Biol. Evol.* **31,** 1956–1978 (2014).
17. Schmidt-Dannert, C. & Arnold, F. H. Directed evolution of industrial enzymes. in **17,** 135–136 (1999).
18. Johannes, T. W., Woodyer, R. D. & Zhao, H. in *Enzyme Assays High-throughput Screening, Genetic Selection and Fingerprinting* (ed. Reymond, J.-L.) 77–93 (2006).
19. Mayer, K. M. & Arnold, F. H. A Colorimetric Assay to Quantify Dehydrogenase Activity in Crude Cell Lysates. *Journal of Biomolecular Screening* **7,** 135–140 (2002).

20. Hosford, J., Shepherd, S. A., Micklefield, J. & Wong, L. S. A High-Throughput Assay for Arylamine Halogenation Based on a Peroxidase-Mediated Quinone-Amine Coupling with Applications in the Screening of Enzymatic Halogenations. *Chem. Eur. J.* **20,** 16759–16763 (2014).
21. Belsare, K. D. *et al.* A Simple Combinatorial Codon Mutagenesis Method for Targeted Protein Engineering. *ACS Synth. Biol.* **6,** 416–420 (2017).
22. Andorfer, M. C., Park, H. J., Vergara-Coll, J. & Lewis, J. C. Directed evolution of RebH for catalyst-controlled halogenation of indole C–H bonds. *Chem. Sci.* **7,** 3720–3729 (2016).
23. Srivastava, P., Yang, H., Ellis-Guardiola, K. & Lewis, J. C. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat Commun* **6,** 7789 (2015).
24. Kille, S., Zilly, F. E., Acevedo, J. P. & Reetz, M. T. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.* **3,** 738–743 (2011).
25. Agudo, R., Roiban, G.-D. & Reetz, M. T. Achieving Regio- and Enantioselectivity of P450-Catalyzed Oxidative CH Activation of Small Functionalized Molecules by Structure-Guided Directed Evolution. *ChemBioChem* **13,** 1465–1473 (2012).
26. Cahn, J. K. B., Baumschlager, A., Brinkmann-Chen, S. & Arnold, F. H. Mutations in adenine-binding pockets enhance catalytic properties of NAD(P)H-dependent enzymes. *Protein Eng. Des. Sel.* gzv057–8 (2015). doi:10.1093/protein/gzv057
27. Osborne, R. *et al.* Novel p450-bm3 variants with improved activity.
28. Andorfer, M. C., Grob, J. E., Hajdin, C. E. & Chael, J. R. Understanding Flavin-Dependent Halogenase Reactivity via Substrate Activity Profiling. *ACS Catal.* **7,** 1897–1904 (2017).
29. Srivastava, P., Yang, H., Ellis-Guardiola, K. & Lewis, J. C. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat Commun* **6,** 7789 (2015).
30. Shepherd, S. A. *et al.* Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chem. Sci.* **6,** 3454–3460 (2015).
31. Schnepel, C., Minges, H., Frese, M. & Sewald, N. A High-Throughput Fluorescence Assay to Determine the Activity of Tryptophan Halogenases. *Angew. Chem. Int. Ed.* **55,** 14159–14163 (2016).
32. Chen, Y., Tang, W. L., Mou, J. & Li, Z. High-Throughput Method for Determining the Enantioselectivity of Enzyme-Catalyzed Hydroxylations Based on Mass Spectrometry. *Angew. Chem. Int. Ed.* **49,** 5278–5283 (2010).
33. Payne, J. T., Andorfer, M. C. & Lewis, J. C. Regioselective Arene Halogenation using the FAD-Dependent Halogenase RebH. *Angew. Chem. Int. Ed.* **52,** 5271–5274 (2013).
34. Robbins, D. W., Boebel, T. A. & Hartwig, J. F. Iridium-Catalyzed, Silyl-Directed Borylation of Nitrogen-Containing Heterocycles. *J. Am. Chem. Soc.* **132,** 4068–4069 (2010).
35. Kallepalli, V. A. *et al.* Harnessing C–H Borylation/Deborylation for Selective Deuteration, Synthesis of Boronate Esters, and Late Stage Functionalization. *J. Org. Chem.* **80,** 8341–8353 (2015).

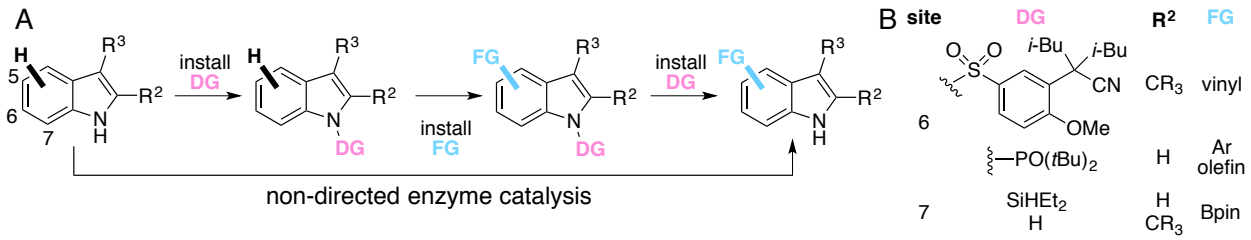
36. Yeh, E., Garneau, S. & Walsh, C. T. Robust in vitro Activity of RebF and RebH, a Two-Component Reductase/Halogenase, Generating 7-Chlorotryptophan during Rebeccamycin Biosynthesis. *Proc. Natl. Acad. Sci.* **102,** 3960–3965 (2005).
37. Parsons, R. L., Berk, J. D. & Kuehne, M. E. Total synthesis of strychnan and aspidospermatan alkaloids. 3. The total synthesis of (. -.)-strychnine. *J. Org. Chem.* **58,** 7490–7497 (1993).
38. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning: a laboratory manual, 2nd edn.* (Cold Spring Laboratory Press, 1989).
39. Lewis, J. C. *et al.* Chemoenzymatic elaboration of monosaccharides using engineered cytochrome P450BM3 demethylases. *Proc. Natl. Acad. Sci.* **106,** 16550–16555 (2009).
40. Gottlieb, H. E., Kotlyar, V. & Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *J. Org. Chem.* **62,** 7512–7515 (1997).
41. Ibaceta-Lizana, J. S. L., Jackson, A. H., Prasitpan, N. & Shannon, P. V. R. Electrophilic substitution in indoles. Part 13. The synthesis and rearrangement of 2-deuteriospiro[cyclopentane-3′-indolenine]. *J. Chem. Soc., Perkin Trans. 2* 1221–1226 (1987). doi:10.1039/P29870001221
42. Maresh, J. J. *et al.* Strictosidine Synthase: Mechanism of a Pictet–Spengler Catalyzing Enzyme †. *J. Am. Chem. Soc.* **130,** 710–723 (2008).

Chapter 4: Altering the site selectivity of RebH through directed evolution

4.1 Introduction

Catalytic C–H functionalization has the potential to directly convert ubiquitous C–H bonds to desired functionality without the need for functional group manipulation.[1,2] Organometallic catalysts are typically used for C–H functionalization,[1] but, although incredibly powerful and synthetically useful, site selective functionalization with these catalysts remains largely dictated by substrate electronics,[3-5] sterics,[6] or directing groups.[7-9] Functionalization of the indole core exemplifies site-selectivity challenges. As discussed in chapter 1, selective functionalization (including aromatic halogenation) can be achieved at the C3 position due to its reactivity relative to other positions.[10-14] Additionally, site selectivity at the C2 position can be readily achieved,[15] primarily through blocking of the C3 position[10,16] (see Scheme 1.1) or through installing a directing group on the indole nitrogen.[14,17,18] Few examples of site selective functionalization at the benzo positions (C4-C7) have been reported, and all of these require directing groups and/or prior functionalization at C2 (Fig. 4.1).[11,19-27]

Figure 4.1: Selective installation of functional groups (FG) on indoles *via* C–H bond activation using different catalyst directing groups (DG) or enzyme catalysis.



Enzymes are able to overcome these site selectivity challenges through substrate binding such that a single C–H bond is positioned proximal to reactive species.[28] For example, tryptophan halogenases (Trp-FDHs) use catalyst control to selectively halogenate the 5-, 6- and 7-positions of L-tryptophan (Fig. 4.1, FG = Cl, Br).[29] Although wild-type enzymes may not possess suitable activity or process compatibility for biocatalysis, directed evolution has proven to be a general approach for altering enzyme function.[30-32] In certain cases, site selectivity has been changed through directed evolution; however, no selective pressure was applied to accomplish this.[33-36] Instead, active variants were identified in a high-throughput screen, and their selectivity was determined *post hoc*. The ability to systematically tune the site selectivity of enzymes such as Trp-FDHs through directed evolution could significantly expand the reach of selective C–H functionalization.

Previous work has demonstrated that the site-selectivity of Trp-FDHs can indeed be altered through mutagenesis. In 2011, van Pee and coworkers introduced point mutation F103A into PrnA, a variant that deviated from its native 7-halogenation to provide a 1:2 mixture of 5- and 7-chlorotryptophans.[37] Site directed mutagenesis was used in the Micklefield lab to alter the selectivity of PrnA toward 2-aminobenzoic acid (16% to 62% selectivity para to the amine)[38] as well as SttH toward 3-indolepropionic acid (10% to 75% selectivity for C5).[39] Expanding upon these efforts, we sought to demonstrate that the MALDI MS screen described in Chapter 3 could be used to systematically alter the site selectivity of RebH through rounds of mutagenesis and screening. This could provide new FDH biocatalysts as well as a general method for tuning site selectivity of FDHs on a given substrate. The work described in sections 4.2.1 and 4.2.2 has been published.[40]

Authorship

Docking studies described in section 4.2.2 were conducted by Dr. Hyun June Park. Crystallographic characterization described in section 4.2.3 was performed with two undergraduate students, Jaylie Vergara-Coll and Anna Girlich. Data processing and refinement of structures were aided by Dr. Narayanasami Sukumar, a staff scientist at the NE-CAT beamlines at the Advanced Photon Source, Argonne National Labs. All computational studies in section 4.2.3 were performed by Song Yang, Cyndi He, and Marc Garcia-Borras of Prof. Kendall Houk's group at UCLA.

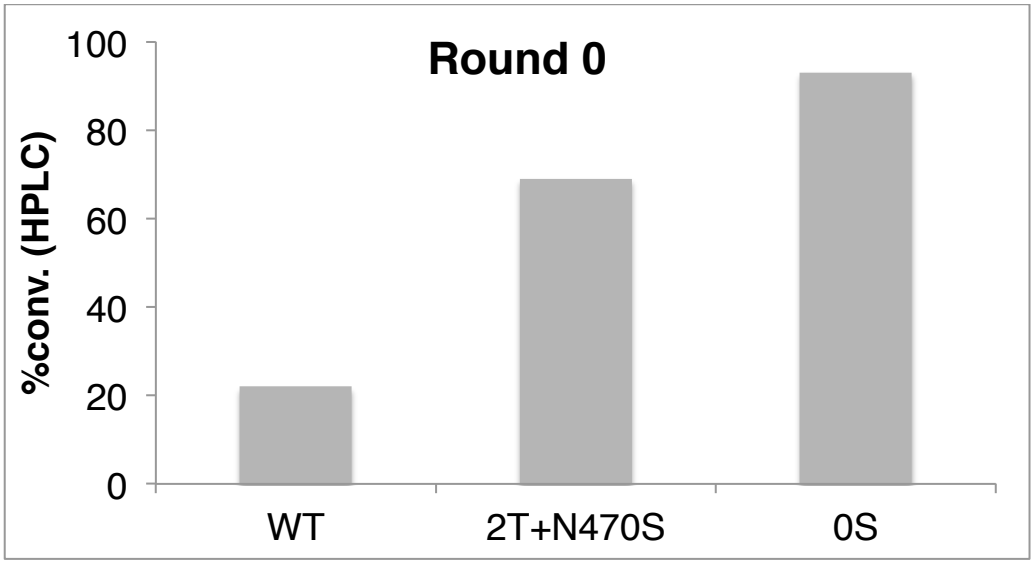
4.2 Results and Discussion

4.2.1 Using a MALDI MS screen to alter the site selectivity of RebH

The FDH RebH site selectively functionalizes the 7 position of the non-native substrate tryptamine (FG = Cl, Br),[41,42] which as discussed in section 4.1, is a challenging transformation for organometallic C–H functionalization catalysts. Moreover, we envisioned that directed evolution could be used to tune the site selectivity of RebH for functionalization of the other benzo positions of tryptamine using the MALDI MS screen from chapter 3. We anticipated altering selectivity could lead to variants with reduced total conversion.[43] We therefore tested RebH variants previously engineered for increased thermostability[44] and widened substrate scope[36] for increased conversion of tryptamine. All variants tested favored chlorination at C7 (>98% selectivity), and variant 2T+N470S displayed the highest total conversion (Fig. 4.2). Because variant 2T halogenated tryptamine to a lesser extent than wild-type RebH, we hypothesized that N470S was the mutation responsible for increased activity on tryptamine. To test this, I used site directed mutagenesis to create variant 0S (RebH+N470S). This variant did

give higher conversion to 7-chlorotryptamine than 2T+N470S and thus was used as parent for the first round of mutagenesis and screening.

Figure 4.2: Total conversion of tryptamine to chlorinated product using potential RebH starting points for directed evolution (Round 0).[a]



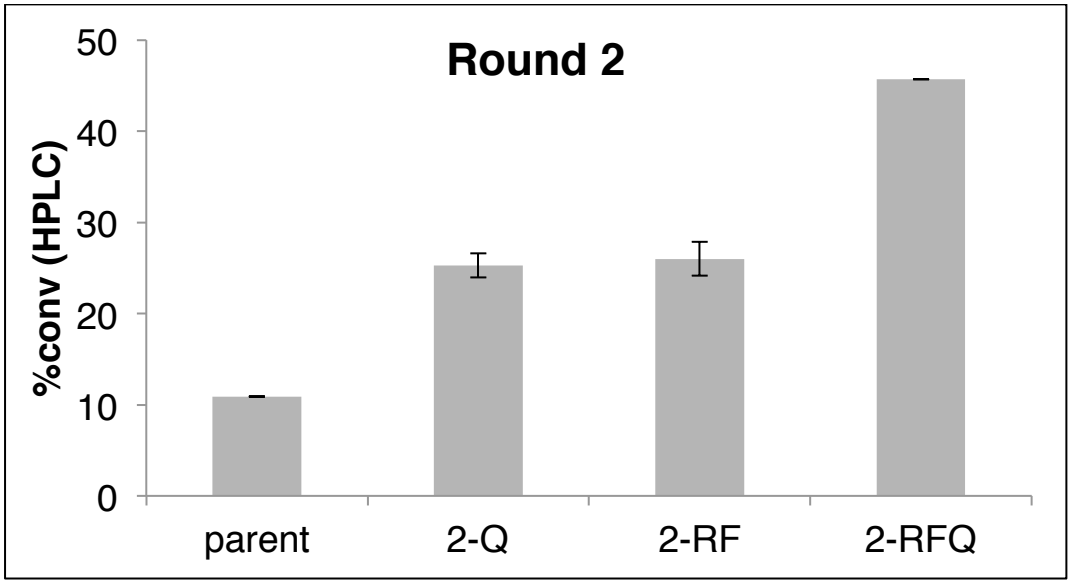
[a] Reactions were conducted using 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 25 mM HEPES buffer pH 7.4, 25 °C, 0.5 mM tryptamine, 10 μM RebH variant.

For the first round of directed evolution, error-prone PCR was performed on the template 0S. One thousand 0S variants were expressed and bioconversions were conducted using Probe **1** (7-deuterotryptamine) as a substrate. The MALDI MS screen described in chapter 3 was used to determine conversion and selectivities relative to parent for all variants. From this library, variant 1P (0S+S448P) was identified as a hit. 1P displayed a 4.5-fold increase in 5/6-chlorination of tryptamine over 0S (from 0.9% to 4.2% of total product, Fig. 4.4). Although 1P has higher activity for 5- and 6-chlorotryptamine, total conversion to 7-chlorotryptamine was reduced by approximately 2-fold (Fig. 4.4). If this library had been screened for total conversion to chlorinated product instead of selectivity, this hit would not have been identified, demonstrating the importance of screening directly for altered selectivity. We continued to observe this trend at

various points along the evolution lineage, whereby variants with altered selectivities often displayed lower total conversion of tryptamine.

Continuing to round 2 of mutagenesis and screening, I used 1P as a template for error-prone PCR. Again, a 1,000 member library was screened by MALDI MS on Probe **1**. From this library, no variants with altered selectivity were identified; however, variants 2RF and 2Q (1P+Q494R+L380F and 1P+R509Q) had increased conversion of chlorotryptamine without a decrease in selectivity for 5- and 6-chlorination. The mutation R509Q was introduced into 2RF via overlap extension PCR to test whether the beneficial mutations would be additive. The resulting 2RFQ variant gave the highest conversion to chlorotryptamine, without a decrease in selectivity for 5- and 6-chlorination (Fig. 4.3). Mutations beneficial to total conversion such as these increased the signal for the MALDI MS screen and thus were carried onto subsequent rounds.

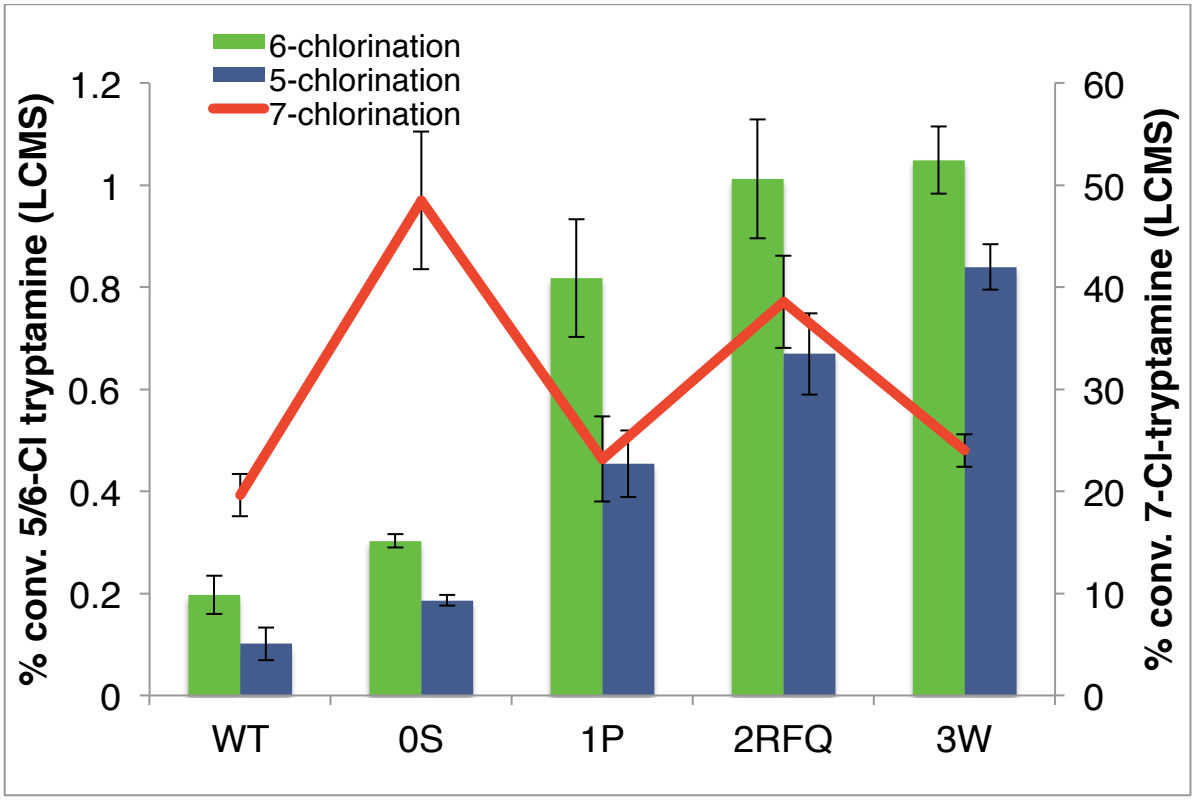
Figure 4.3: Total conversion of tryptamine to chlorinated product using variant 1P (parent) and variants from round 2.[a]



[a] Reactions were conducted using 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 25 mM HEPES buffer pH 7.4, 37 °C, 2 mM tryptamine, 10 μM RebH variant.

O'Connor and coworkers had previously reported the point mutation Y455W to increase substrate specificity of RebH for tryptamine over tryptophan.[42] To probe the effect of Y455W on variant 2RFQ, overlap extension PCR was used to create variant 3W (2RFQ+Y455W). Surprisingly, Y455W did not increase conversion to 7-chlorotryptamine in the 2RFQ template as it had done when incorporated into wild-type RebH. Instead, a significant drop in conversion to 7-chlorotryptamine was observed with 3W without significantly affecting 5/6-chlorination activity. Thus the overall selectivity for 5/6 chlorination of tryptamine was increased 1.7-fold (Fig. 4.4).

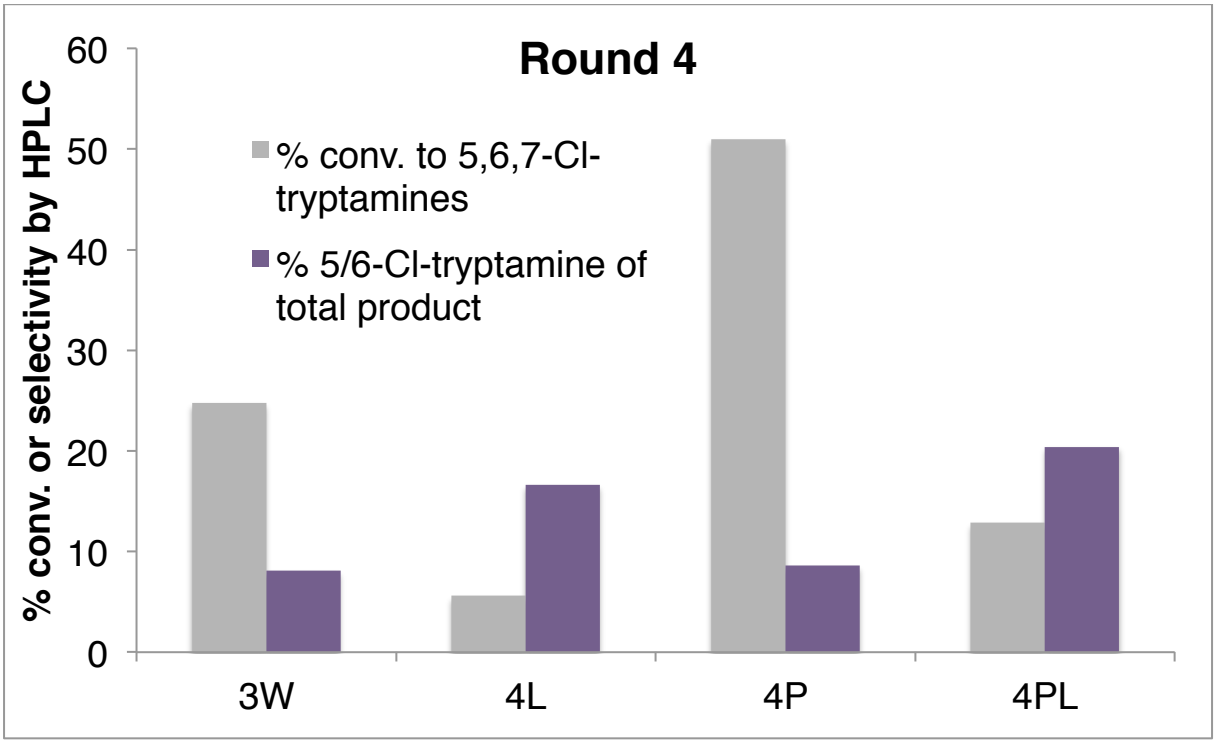
Figure 4.4: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variants from round 0-3.[a]



[a] Yield of 7- (left y-axis) and 6- and 5-cholorotryptamine (right y-axis) for RebH and variants along the halogenase lineage. Reactions were conducted using 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 100 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 0.5 mM phenol, 0.5% v/v *i*-PrOH/25 mM HEPES buffer pH 7.4, 25 °C, 1.5 mM tryptamine probe **2**, 15 μM RebH variant. Reactions were analyzed by LCMS. A complete description of analytical method is provided in section 4.4.2.

3W was used as template for error-prone PCR to generate another library. Probe **1** was used as substrate for a 1,000-member library and bioconversions were analyzed using the MALDI MS screen. Two variants, 4P and 4L (3W+S110P and 3W+F111L), were identified as hits. Bioconversions with 4P resulted in higher conversion to 5/6-chlorotryptamine without a decrease in selectivity, while 4L increased the selectivity for 5/6-chlorination of tryptamine. Addition of S110P to 4L by overlap extension PCR resulted in variant 4PL, which increased the total conversion of 4L while retaining increased selectivity for 5/6-chlorination of tryptamine (Fig. 4.5).

Figure 4.5: Total conversion of tryptamine to chlorinated product and selectivity for halogenation at C5/C6 using variant 3W (parent) and variants from round 4.[a]



[a] Reactions were conducted using 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 25 mM HEPES buffer pH 7.4, 25 °C, 1.5 mM tryptamine, 10 μM RebH variant.

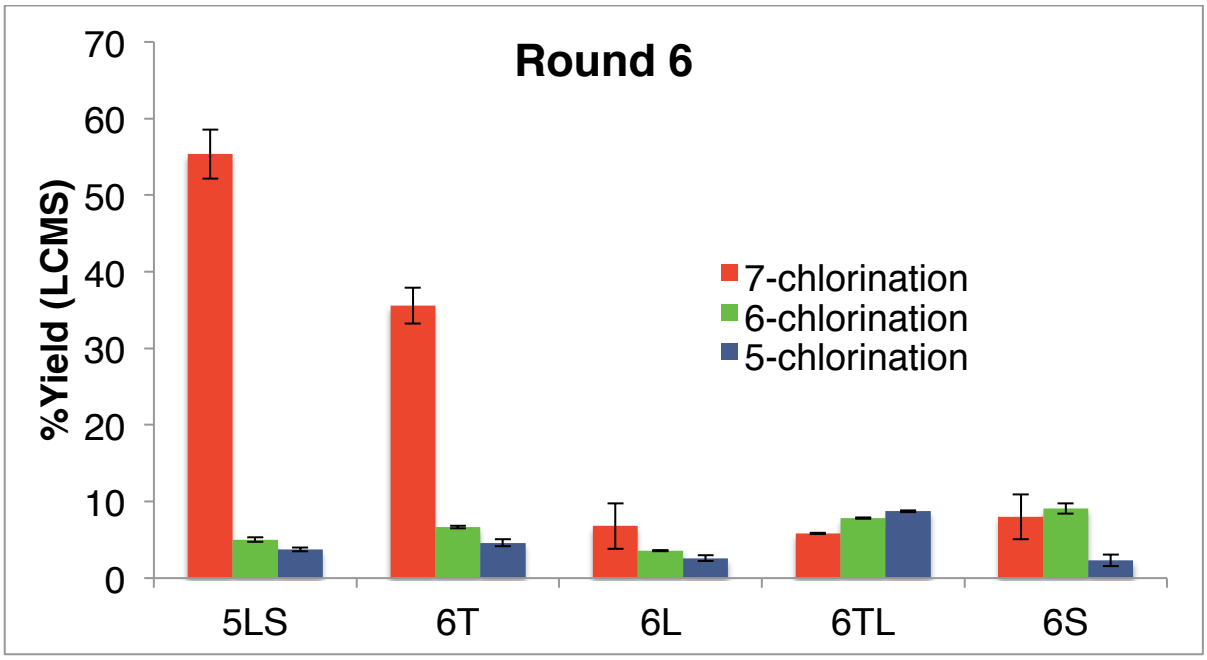
In an attempt to increase the total conversion and/or catalyst lifetime, I individually incorporated mutations from the thermostability lineage described in chapter 2 into variant 4PL.

Indeed two mutations, N166S and S130L, were found to increase total conversion without affecting selectivity. These were combined to form variants 5LS (Fig. 4.11).

Variants 5LS was used as a template for another round of error-prone mutagenesis and screening. One thousand variants were screened using Probe **1** and the MALDI MS screen. Three hits were identified that displayed significantly increased selectivity for 5-/6-chlorination of tryptamine - 6ILLA (5LS+T322I+F458L+F465L+V481A), 6TA (5LS+I52T+T496A), and 6VS (5LS+A58V+L111S). Because each of these variants had more than one mutation, some mutations were probably innocuous or even slightly detrimental. By examining the wild-type RebH crystal structure, the residues I52T, L111S and F465L were identified as active site residues. Instead of creating individual reversions of all 8 mutations found in this round, point mutations 5LS+I52T (6T), 5LS+L111S (6S), and 5LS+F465L (6L) were cloning via overlap extension PCR. Indeed, these residues were responsible for the observed changes in selectivity, and eliminating the additional mutations in each of the variants identified provided similar or higher conversions.

To find the optimal combinations of these three active site mutations, all possible double mutants as well as the triple mutant were cloned by overlap extension PCR. Most combinations gave significant 5-/6-chlorination of tryptamine; however, conversions were too low for screening in plates. One combination, variant 5LS+I52T+F465L (6TL), halogenated the 5-, 6- and 7-positions with similar efficiency while maintaining high enough conversion for screening (Fig. 4.6). At this point in the evolutionary lineage, variants displayed higher selectivity for the 5- or 6-positions than for the native 7-position (Fig. 4.6, 6TL, 6S). Thus, we decided to branch the lineage and begin screening for 5- and 6-halogenases independently, instead of screening with Probe **1** for collective increased 5-/6-chlorination.

Figure 4.6: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variant 5LS (parent) and variants from round 6.[a]

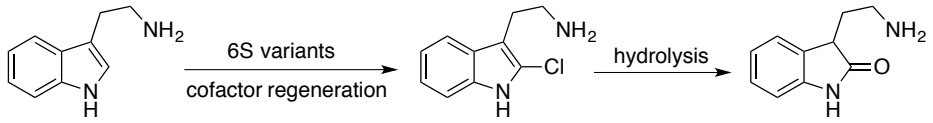


[a] Reactions were conducted using 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 100 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 0.5 mM phenol, 0.5% v/v *i*-PrOH/25 mM HEPES buffer pH 7.4, 25 °C, 0.5 mM tryptamine probe **2**, 25 μM RebH variant. Reactions were analyzed by LCMS. A complete description of analytical method is provided in section 4.4.2.

We initially chose 6S as a starting point for the 6-chlorination branch of the lineage. 6S contains the active site mutation F111S. Interestingly, this residue was also mutated in round 4 (F111L, variant 4L). Previous targeted mutagenesis studies have demonstrated that the corresponding residue in 7-tryptophan halogenase PrnA, F103, can be mutated to A103 to provide a 1:3 mixture of 5- and 7-bromotryptophan with greatly reduced activity.[37] Since this residue did appear to contribute significantly to altered selectivity in our system as well, we used site saturation mutagenesis to target this site as well as neighboring residues 112 and 113. Using 6S as a parent, residues 111-113 were randomized through overlap extension PCR using degenerate NDT codons. From this, a library of 1,000 variants was generated, expressed and screened on Probe **1**, using the procedure outlined in the General Procedures in section 4.4.1 (in

this case, MALDI MS, followed by rescreen of hits by HPLC). Although this corresponded to only 45% library coverage, hits with high selectivity for 6-chlorination from this library were initially identified. Upon further analysis it was found that these hits produced high amounts of 2-oxytryptamine. 2-Oxytryptamine could be formed by chlorination at C2 of tryptamine, followed by hydrolysis (Scheme 4.1).[45] To confirm that this process could happen readily, a chemical chlorination of tryptamine with NCS was conducted. Upon reverse phase Biotage purification using conditions similar to both UPLC analytical methods, only 2-oxytryptamine was recovered. MALDI MS data for variants along the selectivity lineage were re-evaluated for the presence of 2-oxytryptamine. Only mutants with the F111S mutation generated substantial amounts of this product. Because of this, the branch of the lineage containing 6S was no longer pursued, and efforts were focused on the I52T and F465L mutations (Fig. 4.6).

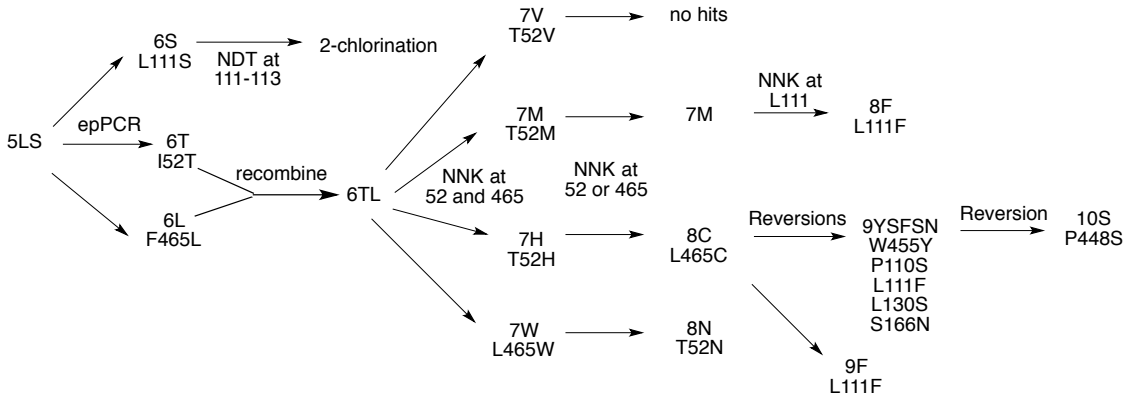
Scheme 4.1: Potential scheme for 2-oxytryptamine formation using FDH variants.[45]



Because variant 6TL produced similar amounts of 5-, 6-, and 7-chlorotryptamines (Fig. 4.6), it was a promising candidate for both the 5- and 6-chlorination branches of the lineage. To differentiate between the 5- and 6-chloro-isomers, Probe **2** (5-deuterotryptamine) was synthesized (see Chapter 3). Variants selective for C5 could be directly identified by screening libraries with Probe **2,** and the pool of potential 6-halogenases could be narrowed. To identify 6-halogenases, variants screened with Probe **2** that retained halogenation activity (many library members lose function) but lower selectivity for C5 were rescreened by UPLC, which is able to separate 7-chlorotryptamine from 5-/6-chlorotryptamines. In this way, C5 and C6 selective variants could be discovered. An alternative approach of using 6-deuterotryptamine to screen for

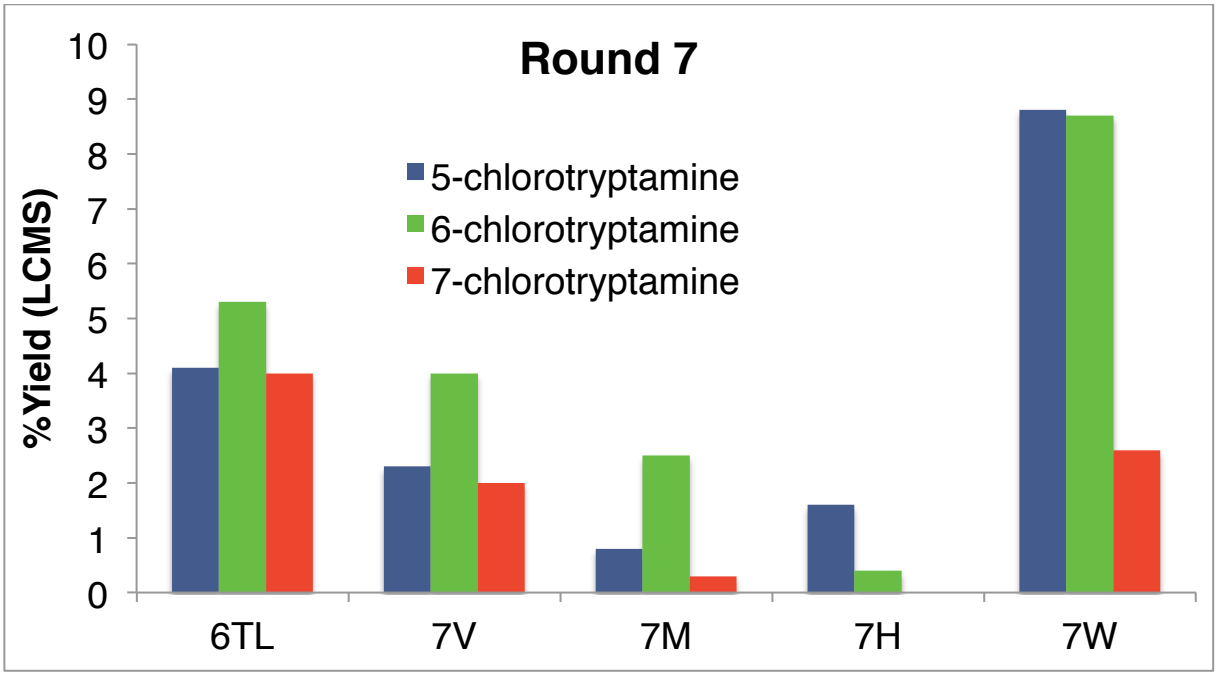
C6 halogenation could have also been employed, but we found this unnecessary for this particular lineage.

Figure 4.7: General overview of evolutionary lineage for rounds 6-10 of mutagenesis and screening.



We decided to optimize the residues at sites 52 and 465 (the key active sites that were mutated in round 6). Two libraries were generated by individual randomization of residues 52 and 465 in 6TL by site directed mutagenesis with NNK codons. One to two hundred variants were screened for each library, corresponding to over 95% coverage.[46] Libraries were screened for activity on Probe **2** using sequential MALDI-MS/UPLC, as described in section 4.4.1. Many hits were identified (Fig. 4.7, 4.8), including 7M (T52M), 7H (T52H), 7W (L465W) and 7V (T52V).

Figure 4.8: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variant 6TL (parent) and variants from round 7.[a]

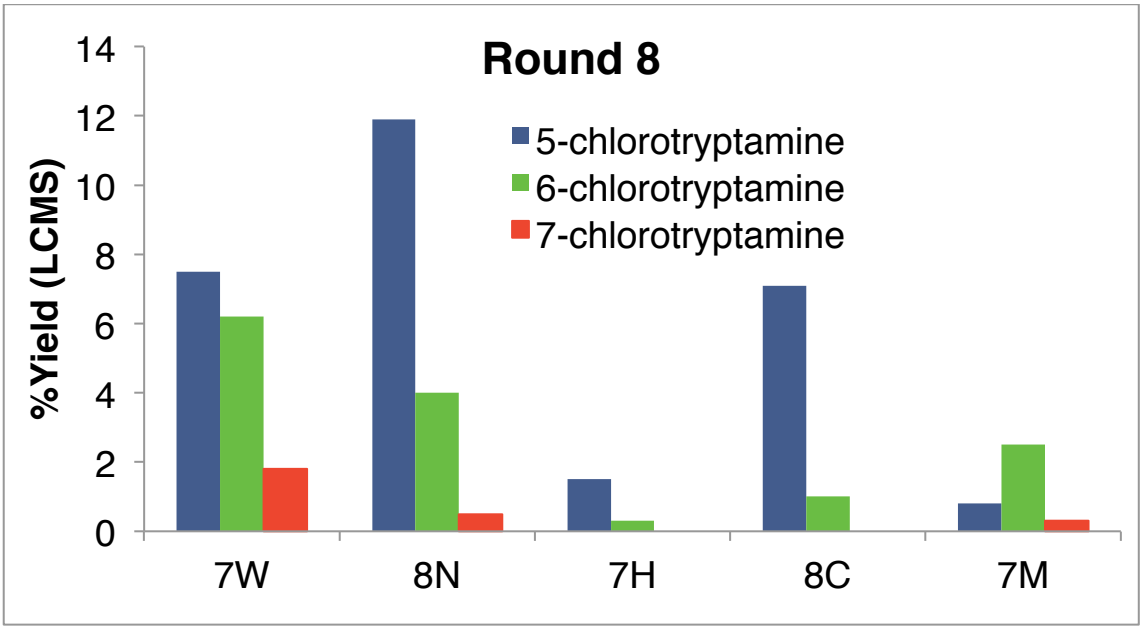


[a] Reactions were conducted using 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 25 mM HEPES buffer pH 7.4, 25 °C, 1.5 mM tryptamine probe **2**, 10 μM RebH variant.

The four hits, 7M, 7H, 7W, and 7V, were used as templates for the next round of mutagenesis and screening. Three libraries were generated by individual randomization of residue 465 in 7M, 7H and 7V by site directed mutagenesis with NNK codons. A fourth library was generated by randomization of residue 52 in 7W by site directed mutagenesis with an NNK codon. In this way, beneficial mutations at these sites arising from epistasis between sites 52 and 465 could be discovered. One to two hundred variants were screened for each library, corresponding to over 95% coverage.[46] Libraries were screened for activity on Probe **2** using sequential MALDI MS/UPLC, as described in section 4.4.1. No hits were identified for the 7V and 7M libraries. Mutant 8N (T52N) from the 7W library and mutant 8C (L465C) from the 7H

library were identified (Fig. 4.7). The mutants with the highest selectivities for 5- and 6-chlorination of tryptamine were 8C and 7M, respectively (Fig. 4.9).

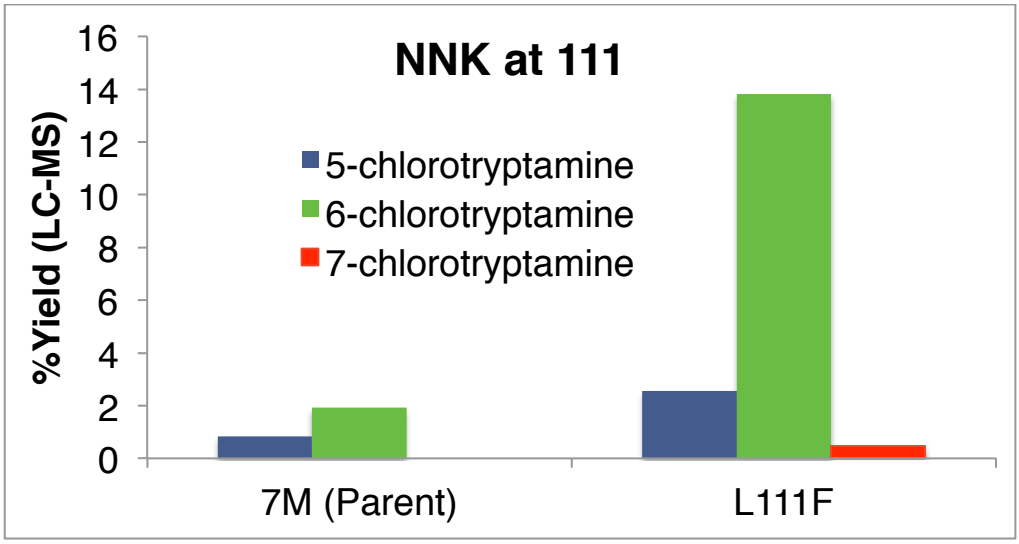
Figure 4.9: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variant variants from round 7 and 8.[a]



[a] Reactions were conducted using 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 25 mM HEPES buffer pH 7.4, 25 °C, 1.5 mM tryptamine probe **2**, 10 μM RebH variant.

Although variant 7M chlorinated tryptamine at C6, it suffered from low activity. Since residue 111 had previously affected both the selectivity and activity so substantially (F111L, Fig. 4.5 and L111S, Fig. 4.6), a library was generated by randomization of residue 111 in 7M by site directed mutagenesis with the degenerate NNK codon. One hundred and fifty variants were screened, corresponding to over 95% coverage.[46] Libraries were screened for activity on Probe **2** using sequential MALDI MS/UPLC. One hit, 8F (7M+L111F) was identified (Fig. 4.10). As discussed earlier, mutation at site 111 has been linked to altered selectivity; however, in this particular enzyme (7M), the wild-type residue F111 is optimal for chlorination at C6. Context-dependent mutations such as this will be discussed in more depth in section 4.2.2 and 4.2.3.

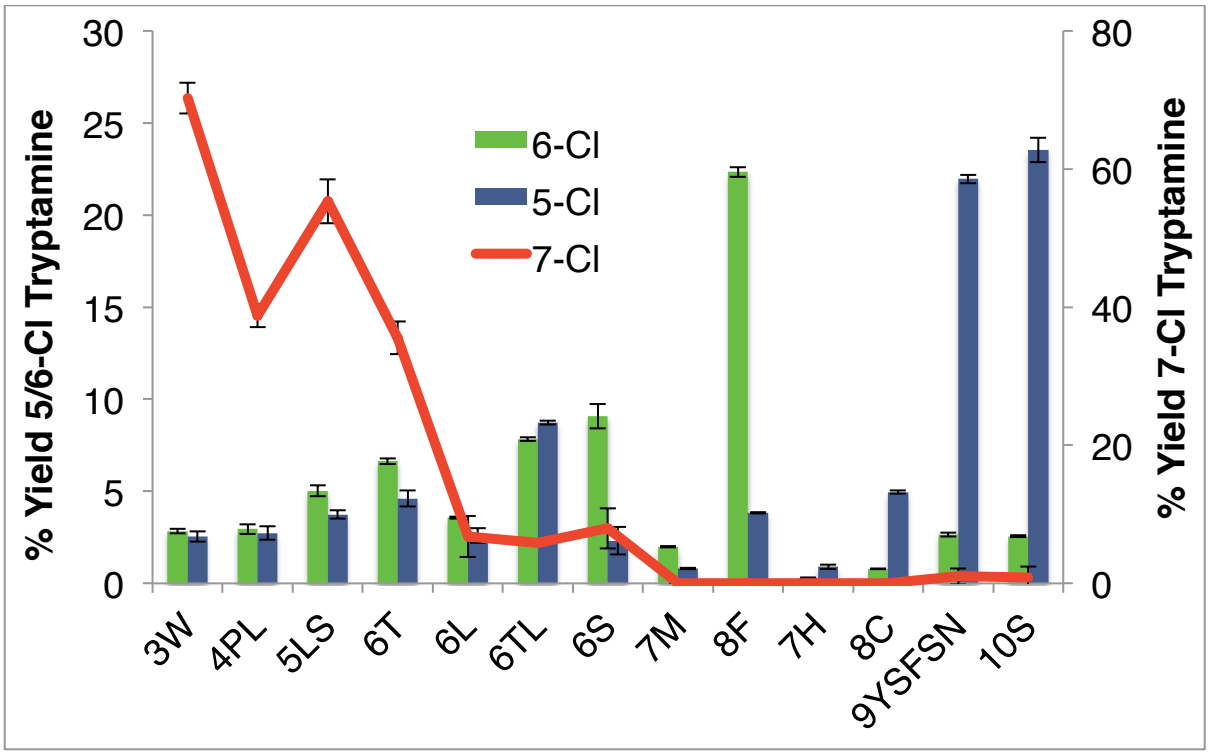
Figure 4.10: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variant 7M and 8F (7M+L111F).[a]



[a] Reactions were conducted using 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 25 mM HEPES buffer pH 7.4, 25 °C, 1.5 mM tryptamine probe **2**, 10 μM RebH variant.

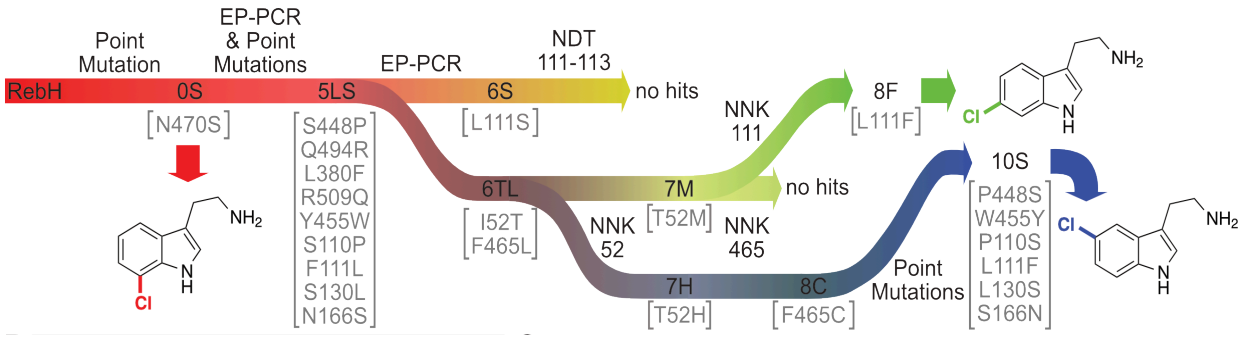
The L111F mutation was introduced by overlap extension PCR into the 5-halogenase 8C in an effort to increase conversion. While this did slightly increase the conversion, it still remained too low for preparative reactions. For this reason, other reversions were made. The I52H and F465C mutations (from rounds 7-8) were introduced into every parent along the lineage (0S, 1P, 2RFQ, 3W and 4PL). From these variants, 9YSFSN (2RFQ+I52H+F465C) gave the highest conversion with no loss of selectivity. The S448P mutation led to a loss of conversion when introduced into 0S (forming 1P); therefore, this mutation was reverted in 9YSFSN. The resulting variant, 10S, had increased conversion of tryptamine without loss of selectivity (Fig. 4.11).

Figure 4.11: Total conversion of tryptamine to 5-, 6-, and 7-chlorotryptamine using variants 3W-10S.[a]



[a] Yield of 7- (left y-axis) and 6- and 5-cholorotryptamine (right y-axis) for RebH variants along the halogenase lineage. Reactions were conducted using 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 100 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 0.5 mM phenol, 0.5% v/v *i*-PrOH/25 mM HEPES buffer pH 7.4, 25 °C, 0.5 mM tryptamine probe **2**, 25 μM RebH variant. Reactions were analyzed by LCMS. A complete description of analytical method is provided in section 4.4.2.

Figure 4.12: Lineage diagram summarizing mutagenesis methods and mutations found in selected variants above and below lineage arrows, respectively.

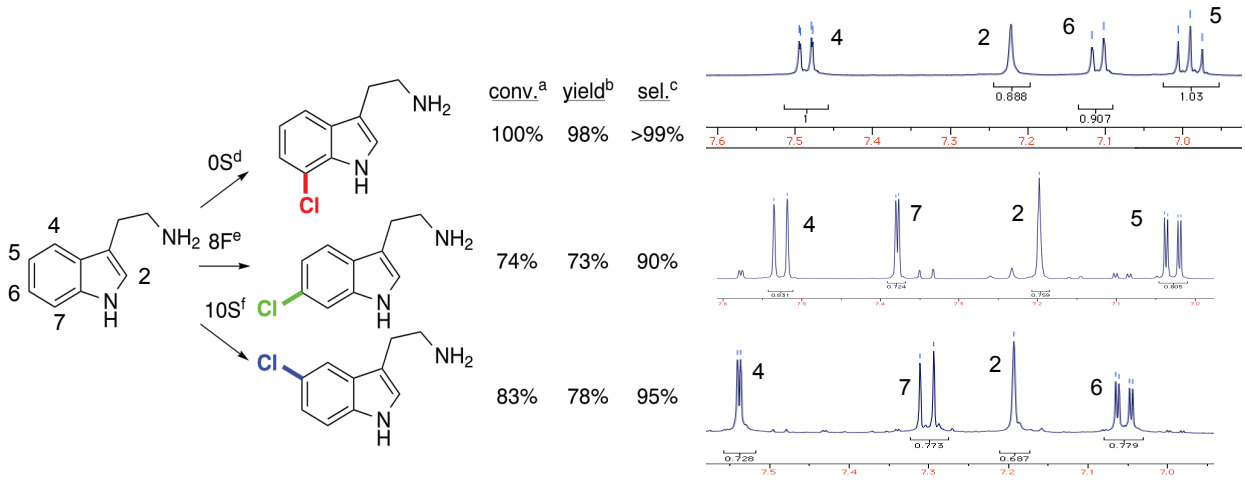


Throughout this evolutionary lineage (Fig. 4.12), it is interesting to note that instead of switching directly to selectivity for another C–H bond on tryptamine, we first loosened the selectivity to give a mixture of isomers (e.g. variant 6TL). From an unselective variant such as 6TL, variants that halogenate specific C–H bonds were then evolved, which is both a common finding in other directed evolution campaigns and is believed to be a common mechanism by which new function emerges via natural selection.[47,48] This example highlights the power of the iterative directed evolution approach to enzyme design.

4.2.2 Characterization of variants: site selectivity, kinetics, substrate scope, and docking simulations

After the ten rounds of mutagenesis and screening detailed in section 4.2.1, halogenases capable of halogenating tryptamine with high selectivity for the 5-position (10S) and the 6-position (8F) were identified. To assess the yield and site selectivity on a preparative scale, I optimized reaction conditions for 10S and 8F (all variants in Fig. 4.4 and 4.11 were conducted using a set of standard conditions). Conducting bioconversions at lower temperatures (10 and 16 °C) produced higher conversions for both 10S and 8F. 10S conversion to chlorinated product also increased when higher NaCl concentrations were used (100 mM), as has been previously observed with other engineered RebH variants.[44] 10 mg bioconversions were conducted for 10S and 8F using optimal conditions for each enzyme. In addition, the yield and selectivity for the new 7-halogenase 0S was determined through a preparative 10 mg bioconversion. Good isolated yields (73-98%) and high site selectivities (90-100%) were observed (Fig. 4.13).

Figure 4.13: Chlorination of tryptamine using engineered halogenases.



[a] Conversion of starting material determined by UPLC analysis of crude reaction mixtures. [b] Isolated yield of pure product. [c] Selectivity determined by NMR analysis of a purified mixture of isomers (inseparable by preparative chromatography). The aryl region of each corresponding ^{1}H NMR spectrum is shown to the right. [d] 10 μM 0S, 0.5 mM tryptamine (10 mg), 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 0.5% v/v *i*-PrOH/25 mM HEPES buffer pH 7.4, 25 °C. [e] As in [d] but 50 μM 8F, 16 °C. [f] As in [d] but 50 μM 10S, 100 mM NaCl, 10 °C.

The enzyme loading requirements to achieve good isolated yields with all three variants differed, with 0S achieving higher yields with lower enzyme loading (Fig. 4.13). The kinetic parameters for these variants were determined, and the lower kcat/K_m obtained for 8F and 10S was consistent with their reduced catalytic efficiency on preparative scale (Table 4.1). Interestingly, 0S has a similar K_m to wild-type RebH, but a 6-fold increase in k_{cat}. The k_{cat} values for both 8F and 10S are significantly lower than 0S, but they are slightly higher than that of wild-type (Table 4.1). Higher K_m values for both 10S and 8F were observed (Table 4.1).

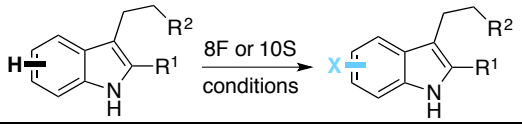
Table 4.1: Kinetic parameters for RebH, 0S, 8F, and 10S.[a]

Enzyme	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m $(min\ \mu M)^{-1}$	Number of Mutations
RebH[b]	9	0.023	2.6×10^{-3}	0
0S	10.6	0.135	2.6×10^{-2}	1
8F	1747	0.037	2.1×10^{-5}	11
10S	160	0.028	1.8×10^{-4}	6

[a] 2-4500 μM tryptamine, 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 100 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 0.5 mM phenol, 2.5% v/v DMSO/25 mM HEPES buffer pH 7.4, 25 °C. 0.1 μM 0S, 25 μM 10S, 25 μM 8F. Time points collected from 10–60 minutes. [b] Values taken from a previous study.[41]

Given that wild-type RebH halogenates several 3-substituted indoles with high selectivity, we wanted to determine if 8F and 10S also demonstrated broad substrate scope. Indeed, several indole derivatives, including 2-methyltryptamine (entry 3), *N*-methyltryptamine (entry 4), and tryptophol (entry 5), were halogenated by both enzymes (Table 4.2). 8F provided similar or better selectivity on these substrates than on tryptamine (entry 1). On the other hand, 10S was completely selective for the 5-position of *N*-methyltryptamine (entry 4), fairly selective for the 5-position of tryptophol (entry 5), and poorly selective for the 5-position of 2-methyltryptamine (entry 3). Notably, neither 8F nor 10S catalyze chlorination of the native RebH substrate L-tryptophan. In addition to chlorination, 8F and 10S are also able to brominate tryptamine, albeit with lower site-selectivity (entry 2).

Table 4.2: Conversion and selectivity for halogenation of different substrates using 8F and 10S.

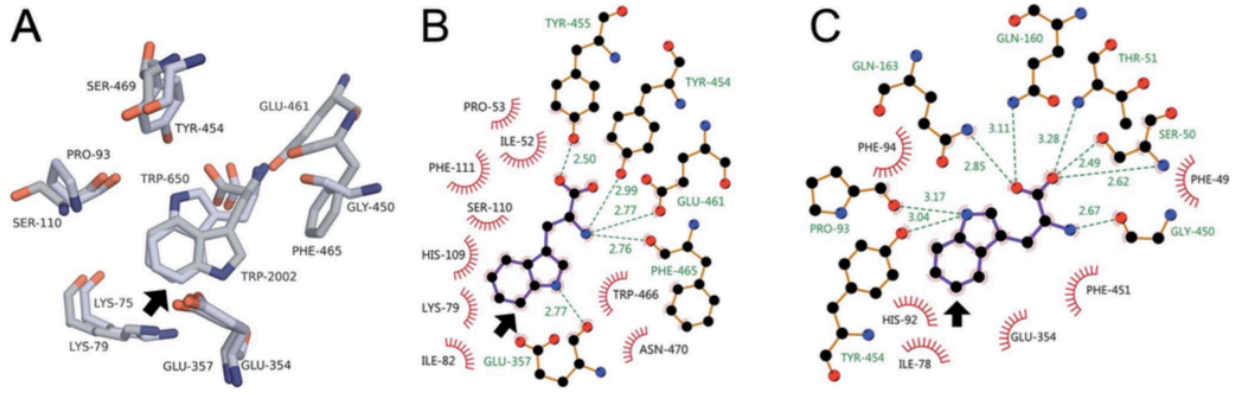


				8F (6-halogenase)[a]		10S (5-halogenase)[b]	
Entry	**R1**	**R2**	**X**	**Conv. (%)[c]**	**6-X (%)[d]**	**Conv. (%)[c]**	**5-X (%)[d]**
1[e]	H	NH_2	Cl	74	90	83	95
2	H	NH_2	Br	84	69	35	59
3	Me	NH_2	Cl	97	99	77	24
4	H	NHMe	Cl	54	98	74	>99
5	H	OH	Cl	48	89	48	84

[a] 50 μM 8F, 0.5 mM substrate (1–2 mg), 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 0.5% v/v *i*-PrOH/25 mM HEPES buffer pH 7.4, 16 °C. [b] As in [a] but with 50 μM 10S and 100 mM NaCl. [c] Conversion determined by UPLC. [d] Selectivity determined by NMR analysis of a purified mixture of inseparable isomers (X = Cl) or by LCMS analysis of reactions conducted using probe **2** (X = Br). [e] Data from preparative reactions (Fig. 4.13).

As previously described, Trp-FDHs contain an active site lysine residue that interacts, either in a covalent or noncovalent manner, with the reactive halenium species. The unique site selectivity of Trp-FDHs is accomplished by substrate positioning in the binding pocket whereby a single C–H bond is located proximal to this lysine (Lys79 in RebH). To better understand how substrate binding in 0S, 8F, and 10S might control site selectivity, the structures of PyrH (5-Trp-FDH)[49] and RebH (7-Trp-FDH)[50] were aligned (Fig. 4.14A). The C7–H bond of L-trp in the RebH structure is closely aligned with the C5–H bond of L-trp in PyrH, and both are positioned proximal to an active site lysine residue thought to position the chlorinating species.[51,52] These substrate-binding modes provide a rationale for the observed selectivity.[53] Thus, interactions between substrate and individual residues within the active site (i.e. hydrogen bonding, π-stacking, electrostatic interactions) are central to controlling halogenation selectivity (Fig. 4.14B-C).

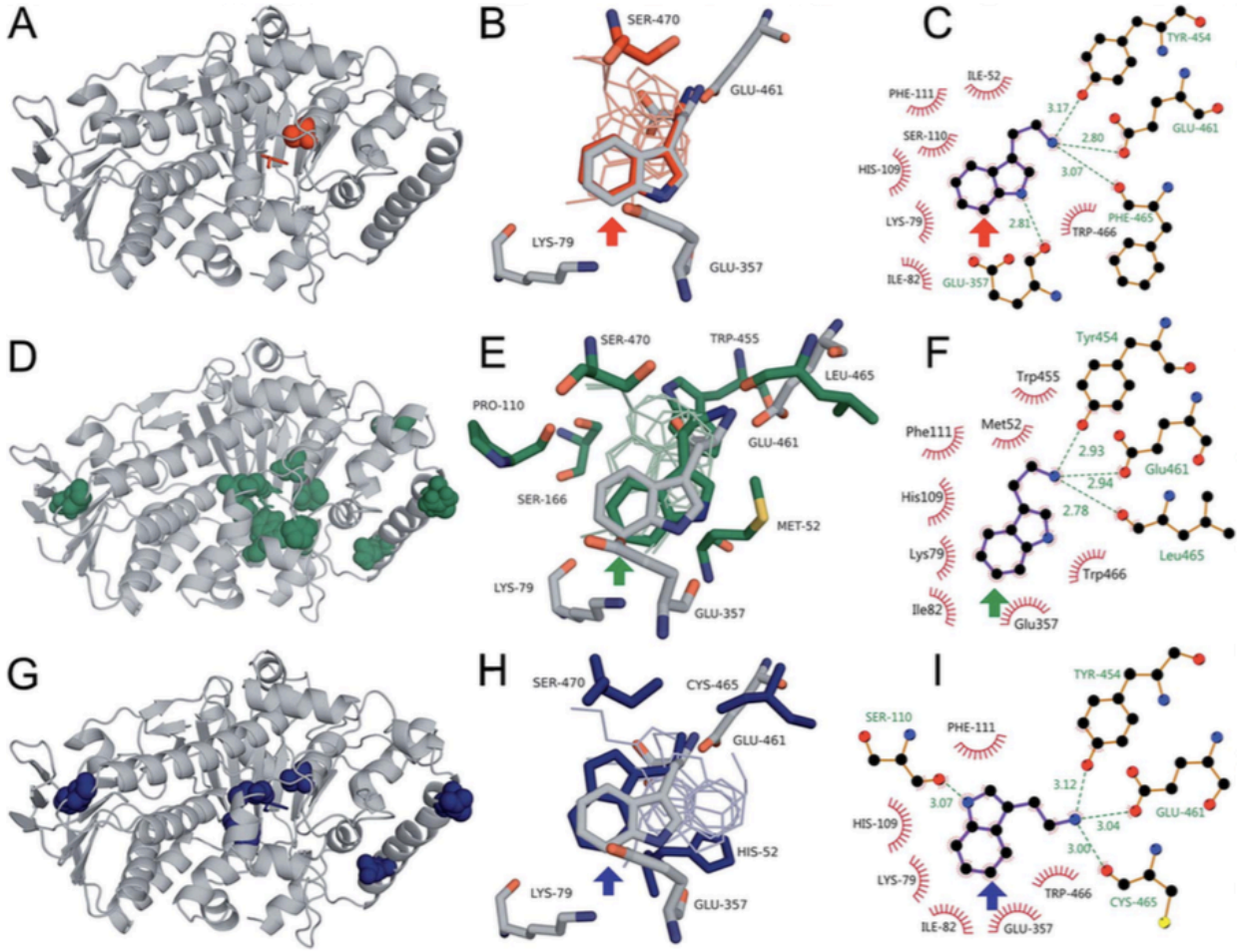
Figure 4.14: Key residues in the RebH and PyrH active sites (A). Interactions involved in L-tryptophan binding in RebH (B) and PyrH (C).[a]



[a] Arrow denotes chlorination site.

To investigate how site selectivity is conferred by our engineered variants, Dr. Hyun June Park minimized the structures of 0S, 8F, and 10S using Swiss-PDBViewer, and subsequently docked tryptamine into each of these structures. For each variant, poses consistent with the observed selectivities were obtained; however, poses consistent with other selectivities were also obtained for 0S and 8F. The pose consistent with the observed 7-selectivity of 0S is very similar to L-tryptophan binding in RebH (Fig. 4.15B-C). A pose consistent with 6-chlorination was obtained for 8F, whereby tryptamine is rotated so that C6–H is proximal to the active site lysine (Fig. 4.15E). In contrast, tryptamine is flipped in the 10S pose consistent with 5-selectivity, reminiscent of tryptophan binding in PyrH (Fig. 4.15H). Unfortunately, it was unclear from these computational data how any of the mutations in our evolutionary campaign had influenced the change in selectivity.

Figure 4.15: Overview of docking studies with 0S, 10S, and 8F computationally minimized structures.[a]



[a] Location of mutations and tryptamine poses for 0S (red), 8F (green), and 10S (blue) mapped onto the RebH structure (grey). (A, D, and G) Location of mutations (spheres) and tryptamine poses (sticks). (B, E, and H) Active site mutations, conserved residues, tryptamine poses consistent with observed selectivity, and native tryptophan pose (sticks) and additional tryptamine poses (lines). (C, F, and I) Binding interactions in poses consistent with observed selectivity. Colored arrows indicate the chlorination site.

4.2.3 Characterization of variants: X-ray crystallography, mutations reversions, and molecular dynamics

Although the computational studies in section 4.2.2 provided insight into tryptamine positioning within the active sites of 0S, 8F, and 10S, they did not illustrate how individual mutations were altering selectivity. For this reason, we obtained X-ray crystal structures for 0S,

8F, and 10S. Crystal structures were determined to 2.6, 2.05, and 2.55 Å for 0S, 8F, and 10S, respectively. Phases were obtained through molecular replacement by using wild-type RebH (PDB ID: 2OA1), and the search model and subsequent models were refined to the R and R_{free} values found in Table 4.3.

Table 4.3: Summary of data collection and refinement statistics for crystal structures of RebH variants 0S, 8F and 10S.[a]

	0S	8F	10S
Data collection			
Space group	P62	P62	P62
Resolution (Å)	50 – 2.6 (2.69 – 2.6)	50 – 2.05 (2.12 – 2.05)	50 – 2.55 (2.64 – 2.55)
No. unique reflections	52,474 (5,219)	108,332 (10,827)	55,848 (5,525)
R_{sym} (%)	28.4 (123.8)	10.1 (109.0)	23.3 (139.8)
R_{pim} (%)	11.6 (60.0)	4.8 (56.2)	9.1 (54.8)
$CC_{1/2}$	(62.4)	(52.3)	(66.9)
<I/σ(I)>	6.6 (1.8)	15.5 (1.5)	10.8 (1.6)
Completeness (%)	99.7 (100)	99.8 (99.8)	100 (100)
Multiplicity	6.4 (5.1)	5.3 (4.6)	7.7 (7.6)
Refinement			
R_{work}/R_{free}	0.165/0.203	0.179/0.207	0.183/0.214
RMSD			
bond lengths (Å)	0.015	0.012	0.003
bond angles (°)	1.32	1.14	0.58
Ramachandran statistics (%)			
favored	97.4	96.6	97.8
allowed	2.6	3.0	2.2
disallowed	0	0.4	0

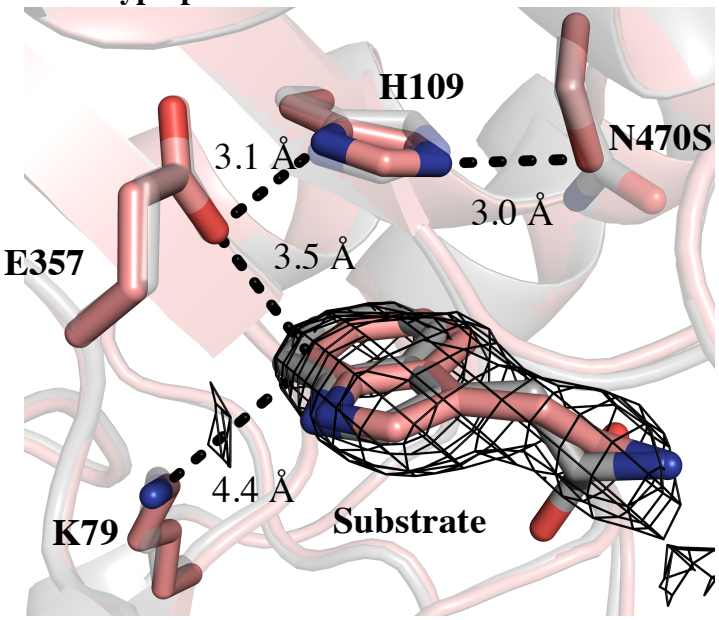
[a] Values in parenthesis correspond to those for the highest resolution bin.

Although all crystals were soaked in a solution containing tryptamine, good occupancy for this substrate was only observed for 0S. This is consistent with the much lower K_m of 0S (10.6 μM) relative to 0S and 8F (160 and 1747 μM, respectively) for tryptamine. Tryptamine was found to bind in the active site of 0S in a very similar manner as L-tryptophan does in RebH (Fig. 4.16). No large conformational changes are observed when comparing structures of wild-type RebH and 0S. The only point mutation in 0S, N470S, is located within the active site;

however, it is too far away from substrate to engage in any type of bonding interaction. Notably, the N470S mutation does increase the size of the binding pocket. We hypothesized this could be leading to the increased k_{cat} (approx. 6-fold) that is observed for tryptamine chlorination with 0S relative to wild-type RebH. To test this, we introduced N470A into RebH and compared this variant's ability to chlorinate tryptamine with RebH and 0S. Variant N470A did not increase tryptamine chlorination rates relative to wild-type RebH as N470S did.

Although N470A was not found to improve chlorination rate in RebH, we verified that this is also the case when introduced into 10S. When 10S + S470A was tested, we observed similar activity and selectivity of this variant relative to 10S + S470N, which again indicated that the identity of serine is crucial for increased activity at this position. The same Ser-His-Glu triad found in 0S is also observed in 10S, thus understanding the effect of N470S in 0S could carry over to 10S as well.

Figure 4.16: Active site of 0S crystal structure with tryptamine bound overlaid with wild-type RebH with bound L-tryptophan.[a]



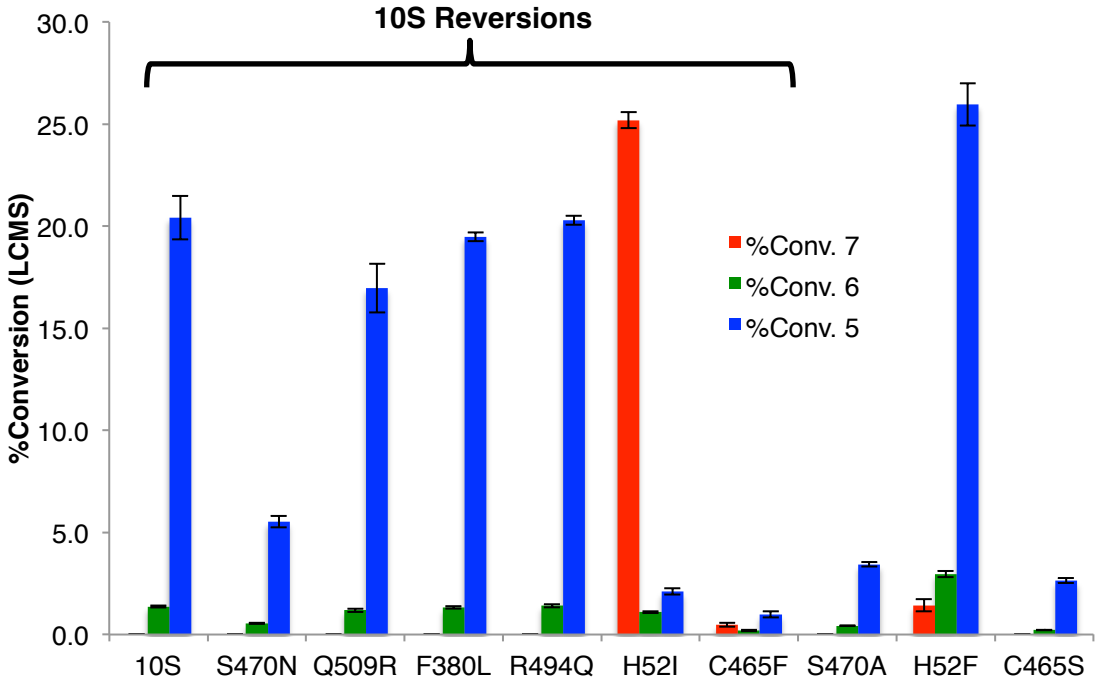
[a] 0S and RebH (PDB ID: 2OA1) ribbon and carbon atoms are shown in salmon and grey, respectively. Heteroatoms shown as sticks are colored according to atom identity. The 2Fo-Fc electron density composite omit map is shown for tryptamine bound in 0S, contoured to 1σ.

Upon further analysis of the 0S crystal structure, we discovered a hydrogen-bonding network between S470, H109, and E357, which is reminiscent of the catalytic triad found in serine hydrolases (Fig. 4.16). Our computational collaborators in Prof. Houk's group at UCLA have found that this hydrogen-bonding network is maintained during molecular dynamics simulations. H109 and E357 are critical residues for chlorination of tryptamine by RebH. H109 forms π-stacking interactions with the indole core of tryptamine. The role of E357 in FDH catalysis is still debated; however, it has been shown to be important for halogenation activity.[52] E357 has been proposed to act as a general base to deprotonate the Wheland intermediate, but this is likely not its only function in catalysis, as deprotonation occurs after the rate-determining step (no observed KIE, see chapter 3). It has also been proposed that E357 could help to stabilize the transition state or could interact with free HOCl thus increasing its electrophilicity.[52,53] We are currently investigating the interactions between S470, H109, and E357 through further mutagenesis studies as well as molecular dynamics simulations. Understanding how S470 increases the rate of chlorination of tryptamine could also further illuminate the role of E357 in FDH catalysis.

In an attempt to better understand how mutations in 10S impact the switch in selectivity from 99% selective for C7 in RebH to 95% selective for C5 (selectivities determined by HNMR),[40] I made individual reversions mutations in 10S. The activity and selectivity of each of these was analyzed by LCMS to determine the role of each point mutation (Fig. 4.17). It's clear that I52H and F465C, both of which are located in the active site, primarily contribute to altering the selectivity of chlorination (Fig. 4.17 and Table 4.4). Surprisingly, when I52H and F465C are introduced into wild-type RebH, the resulting variant is inactive. Mutations N470S and R509Q, the latter of which is *not* in the RebH active site, were found to improve the conversion of

chlorination without significantly affecting the selectivity (Fig. 4.17 and Table 4.4). By increasing total chlorination activity of 10S, these two residues appear to enable detection of chlorinated product.

Figure 4.17: Conversion of tryptamine to 7-, 6-, and 5-chlorotyrptamines for 10S point mutants.[a]



[a] Reactions were conducted using 2.5 µM MBP-RebF, 9 U mL^{-1} GDH, 100 mM NaCl, 20 mM glucose, 100 µM NAD and FAD, 0.5 mM phenol, 1.7% v/v DMSO/25 mM HEPES buffer pH 7.4, 25 °C, 0.5 mM tryptamine probe **2**, 25 µM 10S variant. Reactions were analyzed by LCMS.

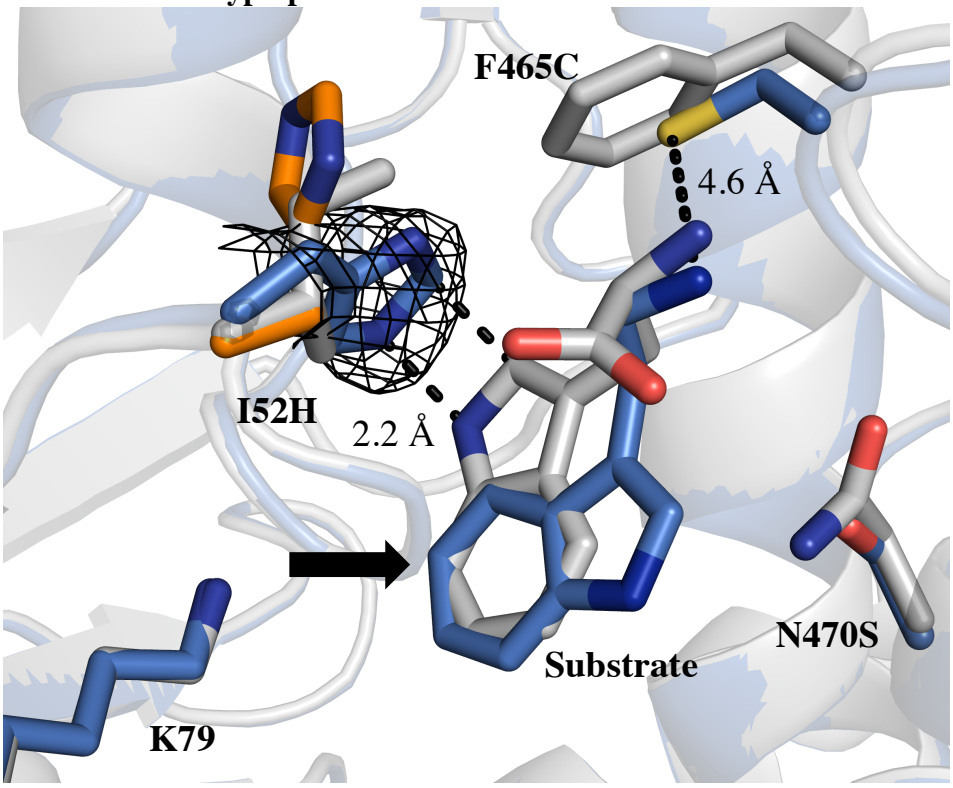
Table 4.4: Chlorinated product distribution for 10S point mutants.[a]

Enzyme	%7Cl	%6Cl	%5Cl
10S	0	6.3±0.3	93.7±0.3
S470N	0	9.0±0.1	91.0±0.1
Q509R	0	6.5±0.1	93.5±0.1
F380L	0	6.4±0.2	93.6±0.2
R494Q	0	6.5±0.3	93.5±0.3
H52I	88.7±0.6	3.8±0.2	7.4±0.5
C465F	28.9±4.4	11.9±0.8	59.2±3.8
S470A	0	11.0±0.4	89.0±0.4
H52F	4.7±1.0	9.8±0.3	85.5±0.6
C465S	0	7.6±0.4	92.4±0.4

[a] Data obtained from Fig. 4.17.

To learn more about why I52H, F465C, and N470S are important for 10S activity and selectivity, we turned to our crystal structure of 10S. As mentioned above, we were not able to obtain a structure of 10S with tryptamine bound. When tryptamine is docked into the crystal structure, a single mechanistically relevant pose, consistent with halogenation at the 5-position, is observed. This pose is very similar to that described in section 4.2.2, in which tryptamine was docked into a computationally minimized 10S model; however, residue H52 adopts a different conformation in the crystal structure than in the minimized structure (Fig. 4.18, blue = crystal structure, orange = minimized structure). While analysis of the computational structure discussed in section 4.2.2 provided no insight into how I52H could be altering site selectivity, a clear explanation can be observed from the crystal structure. When site 52 is Ile, as in wild-type RebH, there is space in the active site for L-tryptophan and tryptamine to bind in a manner that positions C7 proximal to Lys79. When in this binding pose, the pyrrolo ring is situated near the side chain of residue 52. This tryptamine pose is not feasible in the 10S active site since H52 is located only 2.2 Å from the pyrrolo ring of substrate (Fig. 4.18). Instead, docking simulations predict that tryptamine is flipped relative to its native binding pose, thus removing the steric clash and positioning the C5–H bond towards Lys79.

Figure 4.18: Active site of 10S crystal structure with docked tryptamine overlaid with wild-type RebH with bound L-tryptophan.[a]



[a] 10S and RebH (PDB ID: 2OA1) ribbon and carbon atoms are shown in blue and grey, respectively. Heteroatoms shown as sticks are colored according to atom identity. The 2Fo-Fc electron density map is shown for 10S side chain H52, contoured to 1σ. Arrow denotes chlorination site.

Based on these results, we questioned whether another bulky residue such as phenylalanine would impart the same selectivity when introduced into 10S. Upon making this variant, we found that high selectivity for C5 is indeed observed, but this selectivity is lower than that of 10S (85.5% and 93.7%, respectively - Fig. 4.17 and Table 4.4). In the 10S crystal structure, the side chain of H52 can hydrogen bond with nearby main chain carbonyl. We hypothesize that this hydrogen bond, which cannot be formed with the F52 side chain, could favor the observed side chain conformation. Interestingly, all three known 5-Trp-FDHs contain a Phe residue at this position while the 7-Trp-FDHs PrnA and RebH both contain an Ile residue. Although this further points to the importance of this site in Trp-FDHs for site selectivity,

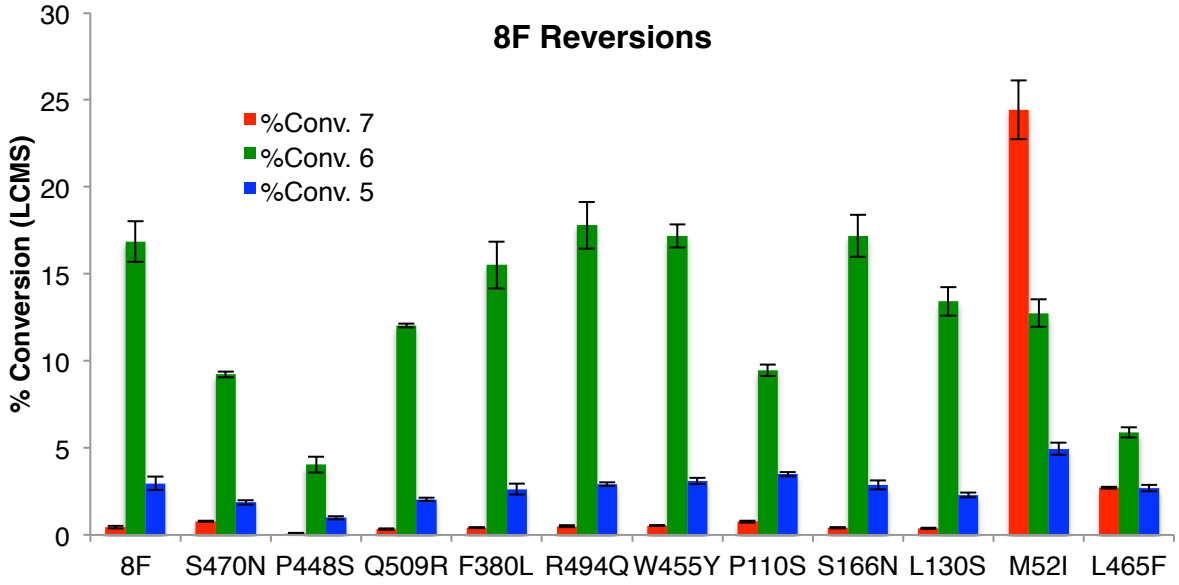
mutation at this site is not sufficient to switch site selectivity. In our case, RebH is inactive when I52H and F465C are introduced into wild-type RebH. Similarly, O'Connor and coworkers have shown that upon introducing I52F into RebH, the resulting variant is inactive.[42]

From the crystal structure of 10S, it is still unclear how F465C are impacting activity and selectivity so dramatically. Interestingly, when C465S is introduced into 10S, site selectivity is not affected by this mutation, but a substantial decrease in conversion is observed relative to 10S (nearly 8-fold, Fig. 4.17). Our computational collaborators are investigating potential reasons why cysteine might be more favorable at this position than serine.

Reversions of all 11 point mutations in 8F were also created to determine which were impacting activity and selectivity. Although mutations at residues 52 and 465 influence selectivity to the greatest extent, as was observed for 10S, P110S was also found to impact site selectivity (Fig. 4.19 and Table 4.5). In addition, S470N, P448S, Q509R, P110S, and L465F were found to decrease conversion. Surprisingly, P448S was found to be the most important residue for total conversion. This mutation, discovered in the first round of mutagenesis and screening, was originally identified because it altered site selectivity; however, significant decrease in total conversion was observed. This mutation was found to be unnecessary to 10S activity and selectivity, and thus was removed during the evolutionary campaign; however, in 8F this mutation is very important for higher conversion. It is located distal to the active site, at the end of a flexible, alpha-helical loop (Fig. 4.20). Residues A450-Y455 in this loop are highly disordered in our 10S and 8F structures, and in structures deposited in the PDB by others as well.[51] Many residues within this loop are located within the active site. Because of the high level of disorder in our 8F crystal structure, it is unclear how this distal mutation at 448 is affecting

chlorination activity; however, our computational collaborators in Prof. Houk's group are currently investigating this observation.

Figure 4.19: Conversion of tryptamine to 7-, 6-, and 5-chlorotyrptamines for 8F point mutants.[a]



[a] Reactions were conducted using 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 0.5 mM phenol, 1.7% v/v DMSO/25 mM HEPES buffer pH 7.4, 25 °C, 0.5 mM tryptamine probe **2**, 25 μM 8F variant. Reactions were analyzed by LCMS.

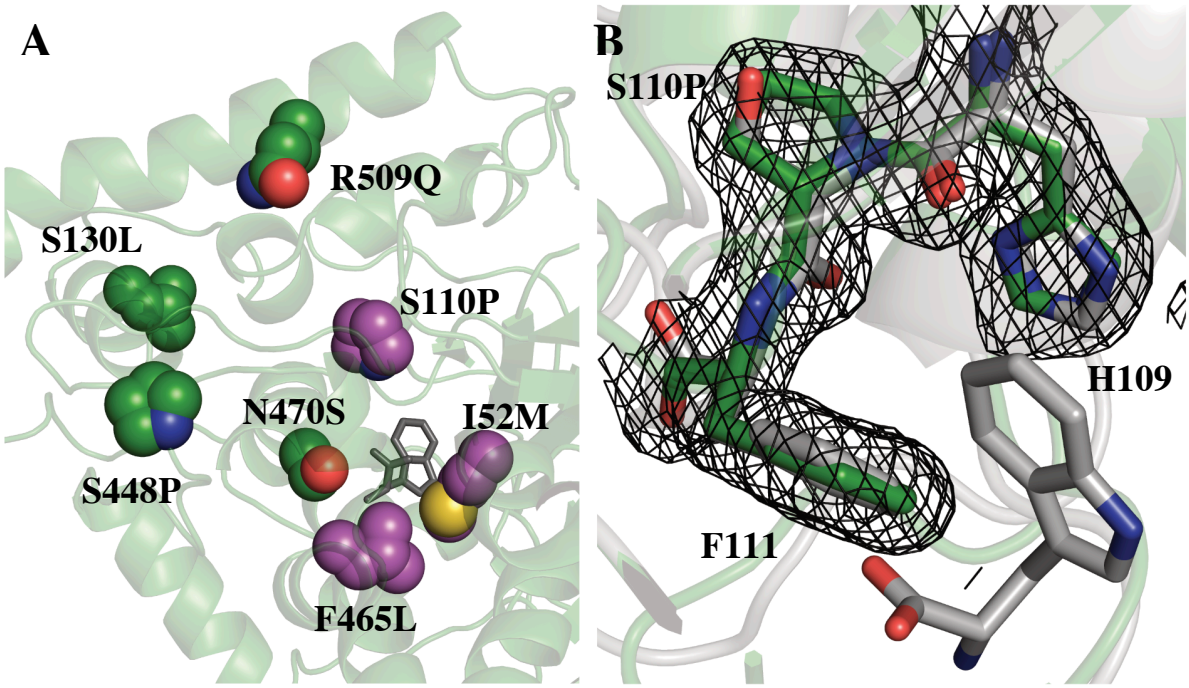
Table 4.5: Chlorinated product distribution for 8F point mutants.[a]

Enzyme	%7Cl	%6Cl	%5Cl
8F	2.2±0.5	83.2±1.1	14.6±1.0
S470N	6.7±0.3	77.6±0.6	15.7±0.9
P448S	1.8±0.4	78.8±0.5	19.4±0.8
Q509R	2.3±0.2	83.6±0.5	14.1±0.6
F380L	2.3±0.3	83.6±0.5	14.1±0.6
R494Q	2.3±0.2	83.9±0.5	13.8±0.5
W455Y	2.6±0.2	82.6±0.5	14.9±0.6
P110S	5.6±0.6	69.0±0.5	25.4±0.2
S166N	2.1±0.3	83.9±0.1	14.0±0.3
L130S	2.4±0.4	83.4±0.3	14.3±0.1
M52I	58.0±0.2	30.3±0.2	11.8±0.1
L465F	24.1±0.6	52.1±0.1	23.7±0.5

[a] Data obtained from Fig. 4.19.

It was also surprising to us that P110S decreased not only total conversion but also site selectivity for chlorination at the 6-position. The S110P mutation was identified during the fourth round of mutagenesis and screening for improved conversion, not for altered selectivity (Fig. 4.5). Moreover, S110P was removed from 10S without a change in site selectivity (round 9). P110 is flanked by two aromatic residues, H109 and F111, which can π-stack with aromatic substrate (Fig. 4.20). Proline can enforce a beta turn, which can in turn enforce interaction of flanking residues.[54] Indeed, our collaborators at UCLA performed MD simulations to determine whether P110 increases the rigidity of the flanking His and Phe residues and found that when the identity of site 110 is Pro as opposed to Ser, much less motion is observed for His109 and F111.

Figure 4.20: Location of 8F mutations that affect selectivity and activity (A)[a] and active site of 8F crystal structure overlaid with wild-type RebH with bound L-tryptophan (B).[b]



[a] 8F crystal structure is shown in green. Side chains of residues found to affect the site selectivity (magenta) and activity (green) are shown as spheres. Heteroatoms shown as spheres are colored according to atom identity. [b] 8F and RebH (PDB ID: 2OA1) ribbon and carbon atoms are shown in green and grey, respectively. Heteroatoms shown as sticks are colored according to atom identity. The 2Fo-Fc electron density map is shown for 8F residues H109–F111, contoured to 1σ.

As was the case for 10S, we were not able to obtain a tryptamine-bound 8F structure. When AutoDock was used to dock tryptamine into the 8F crystal structure, no mechanistically relevant poses were obtained. This is consistent with the very high K_M of 8F (1747 μM); however, it is problematic when assessing how site selectivity is altered. Through further molecular dynamics studies, our collaborators are working to find relevant binding poses for tryptamine in 8F as well as analyze how M52 and L465 could be altering site selectivity.

4.3 Conclusions

RebH variants capable of chlorinating ortho, meta, and para to the indole nitrogen of tryptamine were developed through directed evolution using the novel MALDI MS screen described in chapter 3. The RebH variants created in these studies provide excellent examples of the power in enzyme control over site selectivity and how this selectivity can be systematically tuned. Using similar techniques, the site selectivity of other C–H functionalization enzymes could also be tuned, thus overcoming a major challenge in current C–H functionalization. Crystal structures for engineered variants were obtained and, in combination with reversion data, we are beginning to shed light on the mechanism by which selectivity was altered. Further structural analysis as well as molecular dynamics simulations using these structures are currently ongoing and will hopefully provide further understanding of this important class of enzyme.

4.4 Experimental

4.4.1 General experimental procedures

Materials:

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Deuterated solvents were obtained from Cambridge Isotope labs.

Silicycle silica gel plates (250 mm, 60 F254) were used for analytical TLC, and preparative chromatography was performed using SiliCycle SiliaFlash silica gel (230-400 mesh). Oligonucleotides were purchased from Integrated DNA Technologies (San Diego, CA).

Plasmids pET-28a/RebF and pET-28a/RebH in BL-21 DE3 E. coli were provided by the Walsh group of Harvard Medical School, Boston, MA.[55] The pLIC-MBP plasmid was provided by the Bottomley group of Monash University, Clayton, Australia.[56] The pGro7 plasmid encoding the groES and groEL chaperone set was purchased from Takara (Otsu, Shiga, Japan). BL21(DE3) E. coli cells were purchased from Invitrogen (Carlsbad, CA). NdeI and HindIII restriction enzymes, T4 DNA polymerase, and Phusion HF polymerase were purchased from New England Biolabs (Ipswitch, MA). Luria broth (LB) and Terrific broth (TB) media were purchased from Research Products International (Mt. Prospect, IL). Qiagen Miniprep Kits were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. All genes were confirmed by sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility (900 E. 57th Street, Room 1230H, Chicago, IL 60637). Electroporation was carried out on a Bio-Rad MicroPulser using method Ec2. Ni-nitrilotriacetic acid (Ni-NTA) resin and Pierce® BCA Protein Assay Kits were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were followed when using both products. Amicon® 30 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C. The glucose dehydrogenase (GDH-105), FAD, and NAD were purchased from Codexis (Redwood City, CA). Biotage reverse phase columns (SNAP-KP-C18-HS) were purchased from Biotage.

Complete analytical data have been reported in the literature for: *7-chloro-tryptamine* (**3**),[41] *6-chloro-tryptamine* (**5**),[57] *5-chloro-tryptamine* (**6**),[58] and *2-oxytryptamine* (**13**).[59]

General Procedures:

Standard molecular cloning procedures were followed.[60] Reactions were analyzed using an Agilent Technologies 1200 UHPLC or Agilent Technologies 6130 LC-MS. Reverse phase preparative chromatography was carried out using a Biotage Isolera One or Agilent Technologies 1100 HPLC. ^{1}H and ^{13}C NMR spectra were recorded at 500 MHz and 126 MHz, respectively, on a Bruker DMX-500 or DRX-500 spectrometer, and chemical shifts are reported relative to residual solvent peaks.[61] All mass spectra were collected at the University of Chicago Mass Spectrometry Service Center. High-resolution mass spectra were obtained on an Agilent Technologies 6224 TOF LC-MS. High-throughput screening was performed using a Bruker Ultraflextreme MALDI-TOF-TOF with a Bruker MTP384 steel massive target plate. RebH protein concentrations were determined by A_{280} measurements taken on a Tecan Infinite M200 pro microplate reader. A Microlab® Nimbus liquid handling robot was used in MALDI target plate preparation, and library expression and screening. Library colonies were picked using a Norgen Systems colony-picking robot.

Error-prone library construction and expression[44]*:* Error-prone PCR was used to construct libraries for rounds 1, 2, 4 and 6. Forward and reverse primers used were 5' - TTAATACATATGTCCGGCAAGATTGACAAGATCCTC - 3' and 5' – TATTAAAAGCTTTCAGCGGCCGTGCTGTTGCCTCAG – 3', respectively. The PCR conditions were as follows: 1 ng/µl parent template, 10x Taq buffer, 0.2 mM dNTPs each, 0.2

μM forward primer, 0.2 μM reverse primer, 0.025 U/μl Taq polymerase, and 100 μM $MnCl_2$. PCR was performed in a volume of 50 μL with the following procedure: 95 °C 30 s, (95 °C 30 s, 55 °C 30 s, 72 °C 90 s) for 20 cycles, 72 °C 10 min. The resulting RebH insert was gel purified and digested with the restriction enzymes HindIII (0.33 U/μl) and NdeI (0.33 U/μl) in 10x Cutsmart buffer in a final reaction volume of 60 μl. The digestion was conducted at 37 °C for 12-16 hours, after which it was gel purified. This insert was ligated into digested pET-28a (insert:plasmid ratio of 7.5:1) using T4 DNA ligase. Ligations were conducted for 20 hours at 16 °C. Ligations were cleaned with Zymo DNA Cleaning and Concentrating kits and were transformed by electroporation into *E. coli* containing a plasmid encoding the chaperone pGro7. Library colonies were picked using an automated colony picker (Norgren Systems) and arrayed into 96-deep-well plates (1 mL) containing 300 μL LB with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol. In each 96-well plate, 6 wells were left blank as a control for contamination, and 6 wells were parent cultures. Cells were grown overnight at 37 °C, 250 rpm, and 50-100 μL of overnight culture were used to inoculate 1 mL TB (with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol) in 96-deep-well plates (2 mL) using a liquid handling robot. Following growth at 37 °C, 250 rpm, to an OD_{600} = 0.8-1, enzyme expression was induced with IPTG and arabinose (added using a multichannel pipette) to final concentrations of 10 μM and 0.2 mg/mL, respectively. Protein expression continued for ~20 h at 30 °C, 250 rpm, after which cells were pelleted by centrifugation, the supernatants were discarded, and cell pellets were stored at -20 °C until use.

Targeted library construction and expression: A targeted library approach was used to construct libraries for rounds 7 and 8. Mutations were introduced via overlap extension technique.[62] The

fragment PCR conditions were as follows: 1 ng/µl parent template, 5x Phusion GC buffer, 0.2 mM dNTPs each, 0.5 µM forward primer, 0.5 µM reverse primer, 0.02 U/µl Phusion polymerase, and 5% v/v DMSO. Fragments were gel purified. The assembly PCR conditions were as follows: 1:1 ratio of fragments, 5x Phusion GC buffer, 0.2 mM dNTPs each, 0.5 µM forward primer, 0.5 µM reverse primer, 0.02 U/µl Phusion polymerase, and 5% v/v DMSO. Fragment and assembly PCR were performed in a volume of 50 µL with the following procedure: 98 °C 30 s, (98 °C 20 s, 55 °C 30 s, 72 °C 90 s) for 28 cycles, 72 °C 10 min. DMSO addition was found to enhance specificity and yield of PCR products.

The resulting RebH insert was gel purified and digested with the restriction enzymes HindIII (0.33 U/µl) and NdeI (0.33 U/µl) in 10x Cutsmart buffer in a final reaction volume of 60 µl. The digestion was conducted at 37 °C for 12-16 hours, after which it was gel purified. This insert was ligated into digested pET-28a (insert:plasmid ratio of 7.5:1) using T4 DNA ligase. Ligations were conducted for 20 hours at 16 °C. Ligations were cleaned with Zymo DNA Cleaning and Concentrating kits and were transformed by electroporation into *E. coli* containing a plasmid encoding the chaperone pGro7. This protocol was also used to create all point mutations in this work.

Library colonies were picked using an automated colony picker (Norgren Systems) and arrayed in 1-ml 96-well plates containing 300 µL LB with 50 µg/mL kanamycin and 20 µg/mL chloramphenicol. For libraries in which only a single residue was randomized, a sufficient number of colonies were picked to ensure 95% library coverage[46] (100-150 colonies). For the library in which residues 111-113 were randomized with NDT codons, 1,000 colonies were picked, which corresponds to 44% library coverage.[46] In each 96-well plate, 6 wells were left blank as a control for contamination, and 6 wells were parent cultures. Cells were grown

overnight at 37 °C, 250 rpm, and 50-100 μL of overnight culture were used to inoculate 1 mL TB (with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol) in 2-mL 96-well plates using a liquid handling robot. Following growth at 37 °C, 250 rpm, to an OD_{600} = 0.8-1, enzyme expression was induced with IPTG and arabinose (added using a multichannel pipette) to final concentrations of 10 μM and 0.2 mg/mL, respectively. Protein expression continued for ~20 h at 30 °C, 250 rpm, after which cells were pelleted by centrifugation, the supernatants were discarded, and cell pellets were stored at -20 °C until use.

Library Lysis and Screening[44]*:* Cell pellets in 2-mL 96-well plates were thawed and suspended in 300 μL HEPES buffer (25 mM, pH 7.4). Cells were pelleted by centrifugation and the supernatant was discarded to wash away residual culture media. Cell pellets were then suspended in 100 μL HEPES buffer (25 mM, pH 7.4) containing 0.75 mg/mL lysozyme. After incubation at 37 °C, (250 rpm, for 30 min.) cells were flash frozen in liquid nitrogen and thawed in a 37 °C water bath. DNaseI (10 μL of 1 mg/mL, 25 mM HEPES buffer, pH 7.4) was added, and the cells were incubated at 37 °C, 250 rpm, for 15 min. After centrifugation, 50 μL of supernatant containing cell lysates was transferred to a microtiter plate for screening using a liquid handling robot.

Halogenation reaction set-up: Similar to what has been described previously for halogenation reactions,[41,44] MBP-RebF (0.0017 equiv., 2.5 μM final concentration) and glucose dehydrogenase (9 U/mL final concentration) were added as solutions (25 mM HEPES, pH 7.4) to the RebH lysate. A solution containing D-tryptamine (1 equiv., 1.5 mM final concentration), NAD (0.067 equiv., 100 μM final concentration), FAD (0.067 equiv., 100 μM final

concentration), NaCl (66.7 equiv., 100 mM final concentration), and glucose (13.3 equiv., 20 mM final concentration) was added via multichannel pipette to simultaneously initiate the reactions (final reaction volume of 75 μL). Probe **1** (7-D-tryptamine) was used as substrate in rounds 1, 2, 4 and 6. Probe **2** (5-D-tryptamine) was used as substrate in rounds 7 and 8 to directly determine 5-halogenation hits, and to narrow the pool of potential 6-halogenation hits. The smaller size of the NNK libraries allowed for the use of UPLC Method 2 as a secondary screen for variants found by MALDI MS with conversion and good selectivity for 6/7-chlorotryptamine. If larger libraries were needed, the 6-deutrotryptamine probe could have been synthesized, but for this study, it was not required. The microtiter plates were sealed (ALPS 3000, Thermo Scientific) and incubated at 37 °C (rounds 1, 2, and 4) or 25 °C (rounds 6,7, and 8) in a vertical incubator at 150 rpm for 12-16 hours. The next morning, reactions were quenched with 75 μL of methanol. Ten microliters of 75 mM HCl was added to each reaction using a multichannel pipette, which was required for spotting mixtures onto a MALDI target without re-dissolving the pre-spotted matrix, but was done regardless of how the reactions were analyzed. Precipitated protein was then pelleted by centrifugation, and the supernatant was transferred to a 96-well 0.45 μm filter plate using a liquid handling robot. The filter plate was loaded on top of a new 96-well microtiter plate and centrifuged at 2,453 g for 10 minutes, or until all supernatant had passed through the filter. A liquid handling robot was then used to transfer filtered reaction mixtures onto a MALDI target plate (detailed MALDI target preparation shown below) or a microtiter plate that could be sealed (ALPS 3000, Thermo Scientific) for subsequent HPLC analysis if necessary.

MALDI target preparation and screening: A 384-well MALDI target plate was spotted with 2 μL of a solution of α-cyano-4-hydroxycinnamic acid by a liquid handling robot. The matrix solution contained 7.5 mg/mL of α-cyano-4-hydroxycinnamic acid in 1:1 THF:H_2O. This was dried in a vacuum oven for 15 minutes. Once dry, 2 μL of the filtered bioconversions were spotted onto the MALDI target plate by a liquid handling robot. This was dried in a vacuum oven for 15-30 minutes. The plate was loaded into a Bruker Ultraflextreme MALDI-TOF-TOF and an automated method was developed. Spectra were generated with the reflectron positive (RP) mode. The detector range was set at 160-200 Da. Final mass spectra were produced by averaging 500 raster shots taken at 50 random positions within each spot, which amounted to 25-30 seconds per spot. Targets were shot using the AutoXecute tool of the Flex Control acquisition software.

MALDI MS data analysis and HPLC screening: MALDI MS peaks were analyzed in Flex Analysis software. All peaks within the detector range besides those corresponding to D-tryptamine, chlorotryptamine, and D-chlorotryptamine (m/z = 162, 195, and 196) were set as background. Data from the spectra for these three peaks were then exported into Excel. An Excel macro was developed to insert the value "0" for spectra that did not contain peaks 195 and/or 196. From this list, the data could be easily arrayed into conversions and selectivities in 96-well-plate format in Excel. Hits were identified as variants that showed higher conversion relative to parent or higher selectivity relative to parent. When using Probe **1**, the ratios 196/(162+196+195) (for tryptamine conversion) and 196/(196+195) (for selectivity) were calculated. When using Probe **2**, 195/(162+196+195) and 195/(196+195) were calculated. These ratios were compared to those of the parent reactions, and the highest hits were re-

screened by UPLC. Hits confirmed by UPLC were sequenced and verified by enzyme purification and re-analysis. For rounds 7 and 8, 5-chlorination hits were directly identified by MALDI MS using 195/(162+196+195) and 195/(196+195). Variants with high values for 196/(162+196+195) and 196/(196+195) were re-screened by UPLC Method 2 (see below) to distinguish 6- from 7-chlorination.

UHPLC/LC-MS methods:

UHPLC/LC-MS Method 1: Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 µM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN; 0-10 min, B = 15%; 10-17 min, B = 15-22%; 17-20 min, B = 22-30%; 20-21 min, B = 30%. Absorbance at 280 nm was measured.

UHPLC/LC-MS Method 2: Agilent Eclipse Plus C18 4.6 x 50 mm column, 3.5 µM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN; 0-4 min, B = 15%; 4-4.5 min, B = 15-20%; 4.5-6.5 min, B = 20-30%; 6.5-7 min, B = 100%. Absorbance at 280 nm was measured.

Expression and Purification of MBP-RebF and RebH: For preparative bioconversions and kinetic studies, large-scale cultures (750 mL) of MBP-RebF and RebH variants were grown, expressed and purified as previously reported.[41] MBP-RebF concentrations were measured using the Pierce BCA Protein Assay Kit. RebH concentrations were determined using A_{280} and extinction coefficients calculated based on amino acid composition (Protein Calculator v3.3, http://www.scripps.edu/~cdputnam/protcalc.html).

Smaller cultures (50 mL) of RebH variants were grown to compare conversion and selectivity between variants. Primary culture (500 µL) was used to inoculate 50 mL TB (with 50 µg/mL kanamycin and 20 µg/mL chloramphenicol). Following growth at 37 °C, 250 rpm, until

OD_{600} = 0.6-0.8, enzyme expression was induced with IPTG and arabinose to final concentrations of 100 μM and 2 mg/mL, respectively. Protein expression continued for ~20 h at 30 °C, 250 rpm, after which cultures were harvested by centrifugation. Cell pellets were suspended in 10 mL 25 mM HEPES (pH 7.4) in 50 mL conical tubes, and lysed by sonication while kept on ice.

Sonication Conditions:

50 mL cultures: sonication was performed on a Qsonica S-4000 Sonicator with a 0.5" horn using the following procedure: 8 x 30 s with 45 s rests, 20% duty cycle delivering 40-50 W. To keep the sample from over-heating, the conical tube was submerged in a circulating ice-water bath.

750 mL cultures: sonication was performed on a Qsonica S-4000 Sonicator with a 0.5" horn using the following procedure: 5 x 1 min with 1 min rests, 20% duty cycle delivering 40-50 W. To keep the sample from over-heating, the conical tube was submerged was submerged in a circulating ice-water bath.

After sonication, cell debris was pelleted by centrifugation and the clarified lysate was passed over a Ni-NTA affinity chromatography column. RebH was eluted from the Ni-NTA with 250 mM imidazole. RebH containing fractions were pooled and exchanged into a buffer of 25 mM HEPES (pH 7.4), 10 % glycerol. RebH concentrations were determined using A_{280} and extinction coefficients were again calculated based on amino acid composition using Protein Calculator v3.3. Protein stocks were then stored at -20 °C until use.

Preparative tryptamine bioconversions: Preparative bioconversions were conducted similarly for all enzymes. A solution of 1, 2 or 10 mg substrate (1 equiv., 0.5 mM final concentration) in

0.5% v/v isopropanol was added to a reaction vessel (crystallization dish - 100 x 50 mm, or 50 mL Erlenmeyer flask). Solutions in HEPES buffer (25 mM, pH 7.4) of NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (20 or 200 equiv., 10 or 100 mM final concentration), and a glucose dehydrogenase (9 U/mL final concentration GDH) were added. Solutions in HEPES/glycerol buffer (25 mM, pH 7.5, 10% glycerol v/v) of RebH (0.02-0.1 equiv., 10-50 μM final concentration) and MBP-RebF (0.005 equiv., 2.5 μM final concentration) were added. The bioconversion was diluted with HEPES buffer (25 mM, pH 7.4) to the appropriate reaction volume and an aqueous solution of glucose (40 equiv., 20 mM final concentration) was added to initiate the cofactor regeneration cycle. The dish was covered with perforated aluminum foil and agitated in a vertical incubator at 90 rpm at 10-25 °C.

Reactions were monitored by HPLC Method 1 as described above and were quenched with aqueous HCl (5.0 M, until pH ~1-2) upon completion. NaCl was added to saturation. Precipitated protein was filtered out through a pad of Celite and was washed with water. The filtrate was basified (pH ~9-12) and extracted into CH_2Cl_2. The crude material was purified by reverse phase chromatography, on a Biotage or a preparative HPLC (Agilient 1100).

Reported conversions are the unadjusted final conversions of substrate to chlorinated products. Selectivities of preparative reactions were calculated by comparing analogous protons in the aryl region of ^{1}H NMR spectra.

4.4.2 Specific experimental procedures

Conversion and selectivity determination of lineage (Figures 4.4 and 4.11): In order to compare the conversions and selectivities for 5-, 6- and 7-chlorination of tryptamine of key mutants along the lineage, 50 mL expression cultures were grown for the following variants: wtRebH, 0S, 1P,

2RFQ, 3W, 4PL, 5LS, 6L, 6T, 6TL, 6S, 7M, 8F, 7H, 8C, 9YSFSN, 10S. These cultures were expressed, lysed, and purified according to the protocol found in the General Procedures.

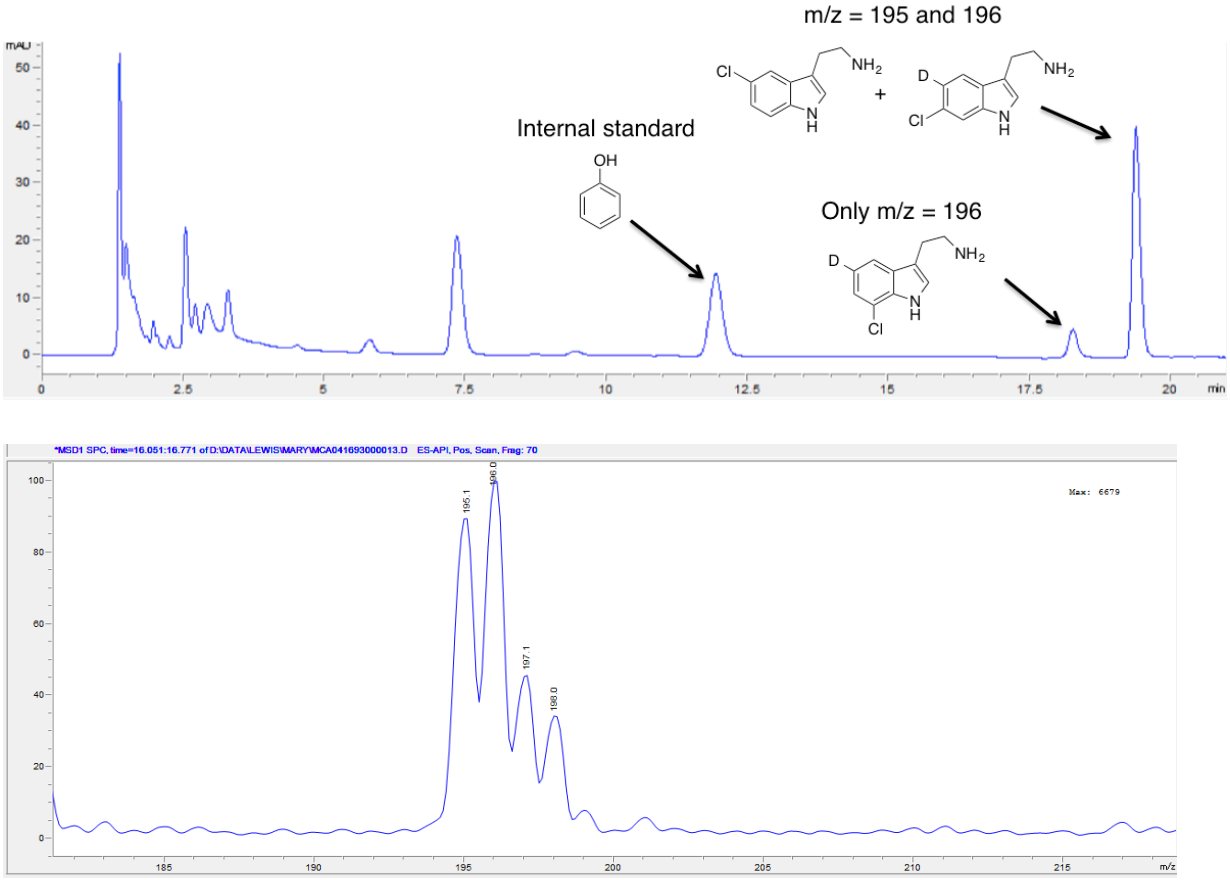
Reaction set-up for wtRebH, 0S, 1P, 2RFQ, 3W: RebH variants (0.05 equiv., 15 μM final concentration) were arrayed into a 96-well microtiter plate in triplicate. MBP-RebF (0.0017 equiv., 2.5 μM final concentration) and glucose dehydrogenase (9 U/mL final concentration) were added as solutions (25 mM HEPES, pH 7.4) to the RebH. A solution containing Probe **2** (1 equiv., 1.5 mM final concentration), NAD (0.067 equiv., 100 μM final concentration), FAD (0.067 equiv., 100 μM final concentration), NaCl (66.7 equiv., 100 mM final concentration), phenol (internal standard, 0.33 equiv., 0.5 mM final concentration) and glucose (13.3 equiv., 20 mM final concentration) were added via multichannel pipette to simultaneously initiate the reactions (final reaction volume of 75 μL).

Reaction set-up for 3W, 4PL, 5LS, 6L, 6T, 6TL, 6S, 7M, 8F, 7H, 8C, 9YSFSN, 10S: RebH variants (0.05 equiv., 25 μM final concentration) were arrayed into a 96-well microtiter plate in triplicate. MBP-RebF (0.005 equiv., 2.5 μM final concentration) and glucose dehydrogenase (9 U/mL final concentration) were added as solutions (25 mM HEPES, pH 7.4) to the RebH. A solution containing Probe **2** (1 equiv., 0.5 mM final concentration), NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (200 equiv., 100 mM final concentration), phenol (internal standard, 1 equiv., 0.5 mM final concentration) and glucose (40 equiv., 20 mM final concentration) were added via multichannel pipette to simultaneously initiate the reactions (final reaction volume of 75 μL).

Reactions for all mutants were mixed at 650 rpm on top of an Eppendorf air bath and were quenched with 1 volume (75 μL) methanol after 16 hours. The precipitated protein was removed by centrifugation and the reactions were filtered and analyzed by LC-MS using Method

1 from the General Procedures. 7-chlorotryptamine separates from 5- and 6-chlorotryptamine using this LC-MS method. Because of this, concentrations of 7-chlorotryptamine were determined by calculating the ratio of product to internal standard and fitting that value to a calibration curve. 5- and 6-chlorotryptamines do not separate by LC-MS. To find the ratio of 5:6 chlorinated product, the mass intensities m/z = 195 and 196 for the 5/6-chlorotryptamine peak were analyzed. The 196 peak was adjusted according to the predicted spectrum for 5-chlorotryptamine (the intensity of m/z = 196 – (the intensity of m/z = 195 * 0.12)). The ratio of 195:196 was then calculated using these values (intensity of 195 and adjusted intensity of 196). This ratio was then used to find the area of the UV peak corresponding to the individual 5- and 6-chlorotryptamines. These two values were divided by the area for internal standard and fit to a calibration curve.

Figure 4.21: Overview of LCMS method to analyze site selectivity of chlorination.[a]



[a] Top: An example LC-MS trace for a RebH bioconversion. 7-chlorotryptamine is chromatographically separated from the 5- and 6-chlorotryptamines. Bottom: Mass spectrum of peak for 5- and 6-chlorotryptamines. When Probe **2** is used, the difference between the masses of 5- and 6-chlorotryptamines can be used to calculate the ratio of 5:6.

Determination of kinetic parameters: Rates were determined by monitoring the conversion of 2-4500 μM tryptamine in the presence of NAD (100 μM final concentration), FAD (100 μM final concentration), NaCl (100 mM final concentration), MBP-RebF (2.5 μM final concentration), glucose dehydrogenase (9 U/mL final concentration GDH), glucose (20 mM final concentration), and phenol as an internal standard (0.5 mM final concentration) at a final volume of 75 μL in a microtiter plate. RebH was added at a final concentration of either 0.1 μM (0S) or 25 μM (10S, and 8F). Plates were sealed using a plate sealer and shaken at 650 rpm at room temperature.

Reaction mixtures were quenched at 10-60 minutes by addition of 75 μL of MeOH. All time points were collected in triplicate. The precipitated protein was then removed by centrifugation and the reactions were filtered and analyzed by HPLC method 2 described in the General Procedures. Product formation was determined by calculating the ratio of product to internal standard and fitting that value to a calibration curve prepared from known concentrations of each chlorinated isomer of tryptamine. The kinetic parameters (K_m and k_{cat}) for each mutant were determined using the Hanes-Woolf plots constructed from the substrate concentrations and the observed initial rates.

In silico mutation and molecular docking studies: The holo form of RebH structure (resolution of 2.15Å, PDB ID: 2OA1) was used for molecular simulations.[50] FAD, Cl^- and Trp were removed for obtaining receptor structure. Using Swiss-PDBViewer, the 8F, 10S, and 0S variants were modulated for each multiple mutation sites.[63] Furthermore, energy minimization was proceed for repair distorted geometry (wild type structure was also minimized for further simulation). The Swiss-PDBViewer uses GROMOS 43B1 force field for minimization. Molecular docking simulation was carried out with AutoDock Vina, which has been successfully used for detecting binding poses, was used for molecular docking.[64] The target ligand, tryptamine, was set to rotate three rotatable bonds in the aminoethyl substituent. All protein structures including ligand were converted into PDBQT format by AutoDock tools to run the simulation. Polar hydrogen bonds were added to receptors (protein structure) and Gasteiger charges were typed to the structures. A cubic grid box (grid spacing = 1.000 Å; 24 × 20 × 20 grid points) was centered on the coordinates of the tryptophan in original PDB. Exhaustiveness and number of binding modes were set to 15. Obtained docking poses were analyzed based on

binding energies and the mechanism of halogenase that could ensure feasibility of each reactive position.

Crystallization and structure determination: Following standard Ni-NTA protein purification, RebH variants were further purified by gel filtration chromatography using a HiLoad 16/600 Superdex 200 column (GE Healthcare Life Sciences) into a buffer of 20 mM HEPES (pH 7.4) for crystallography. Protein concentration was determined using A_{280} and extinction coefficients calculated based on amino acid composition (Protein Calculator v3.3, http://www.scripps.edu/~cdputnam/protcalc.html). Purified protein was concentrated to 8-11 mg/mL, and crystals were grown at 20 °C using the hanging drop vapor diffusion method with a reservoir solution of 1.2-1.3 M Na/K phosphate buffer (pH 6.8). Single 0S and 10S crystals grew in 1-3 weeks under these conditions. Microcyrstalline showers were observed with 8F. 8F microcrystals were used to seed new drops, which produced large, single crystals. Single crystals for 0S, 10S, and 8F were soaked in a solution containing 10 mM tryptamine, 30 mM NaCl, and 5 mM FAD for 5-15 minutes and were subsequently flash frozen in liquid nitrogen following cryoprotection with the reservoir solution supplemented with 16 % glycerol. Data were collected at NE-CAT beamline 24-ID-C at the Advanced Photon Source at Argonne National Laboratory using the Pilatus detector. Data were processed using HKL2000.[65] Phases were determined via molecular replacement using Phaser[66] and wild-type RebH (PDB ID 2OA1) as the search model. Manual model building was performed in Coot,[67] and the structures were refined with PHENIX.[68]

4.4.3 Detailed isolation and characterization

Detailed isolation and characterizations:

7-chloro-tryptamine (3): The bioconversion was conducted in a crystallization dish according to the general procedure, using 0.02 equiv. of 0S (10 μM final concentration) at 25 °C, with 10 mg of substrate and 10 mM NaCl. After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and concentrated onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography (gradient from 100% H_2O to 15% CH_3CN/H_2O) to afford the known compound[41] **3** in 98% yield (19.0 mg of **3**·TFA, 0.0617 mmol). ^{1}H NMR (500 MHz; MeOD): δ 7.49 (dd, J = 7.9, 0.8, 1H), 7.22 (s, 1H), 7.11 (d, J = 7.6, 1H), 6.99 (t, J = 7.8, 1H), 3.19 (t, J = 7.4, 2H), 3.08 (t, J = 7.4, 2H). HRMS (ESI-TOF) calcd for $C_{10}H_{11}N_2Cl$ $[M + H]^+$: 195.0684 and 197.0654, found: 195.0682 and 197.0658.

6-chloro-tryptamine (5): The bioconversion was conducted in a crystallization dish according to the general procedure, using 0.1 equiv. of 8F (50 μM final concentration) at 16 °C, with 10 mg of substrate and 10 mM NaCl. After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and concentrated onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography (gradient from 100% H_2O to 15% CH_3CN/H_2O) to afford the known compound[57] **5** in 73% yield (14.1 mg of **5**·TFA, 0.0456 mmol). ^{1}H NMR (500 MHz; MeOD): δ 7.53 (d, J = 8.4, 1H), 7.38 (d, J = 1.8, 1H), 7.20 (s, 1H), 7.03 (dd, J = 8.5, 1.9, 1H),

3.22 (t, J = 7.4, 2H), 3.10 (t, J = 7.4, 2H). HRMS (ESI-TOF) calcd for $C_{10}H_{11}N_2Cl$ $[M + H]^+$: 195.0684 and 197.0654, found: 195.0736 and 197.0669.

5-chloro-tryptamine (6): The bioconversion was conducted according to the general procedure using 0.1 equiv. of 10S (50 μM final concentration) at 10 °C, with 10 mg of substrate and 100 mM NaCl. After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and concentrated onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography (gradient from 100% H_2O to 15% CH_3CN/H_2O) to afford the known compound[58] **6** in 78% yield (15.0 mg of **6**·TFA, 0.0487 mmol). ^{1}H NMR (500 MHz; MeOD): δ 7.54 (d, J = 2.0, 1H), 7.30 (d, J = 8.6, 1H), 7.19 (s, 1H), 7.05 (dd, J = 8.6, 2.0, 1H), 3.17 (t, J = 7.4, 2H), 3.04 (t, J = 7.4, 2H). HRMS (ESI-TOF) calcd for $C_{10}H_{11}N_2Cl$ $[M + H]^+$: 195.0684 and 197.0654, found: 195.0687 and 197.0655.

2-oxytryptamine (13): The prodecure was adapted form a previous report.[69] Tryptamine (100 mg, 0.625 mmol, 1 equiv.) and *N*-chlorosuccinimide (83.4 mg, 0.625 mmol, 1 equiv.) were added to a 25 mL round bottom flask. A mixture of glacial acetic acid and formic acid (10:3, 7.9 mL) was added and the reaction was stirred for 20 minutes at room temperature. The reaction was neutralized with NaOH, extracted into CH_2Cl_2, and concentrated onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography (gradient from 100% H_2O to 15% CH_3CN/H_2O) to afford the known compound **13**.[59] No 2-

chlorotryptamine was observed upon purification. ^{1}H NMR (500 MHz; MeOD): δ 7.30 (dd, J = 7.4, 0.6, 1H), 7.25 (tt, J = 7.7, 1.0, 1H), 7.06 (td, J = 7.6, 1.0, 1H), 6.92 (d, J = 7.8, 1H), 3.64 (dd, J = 8.4, 5.1, 1H), 3.14 (t, J = 7.6, 2H), 2.35-2.29 (m, 1H), 2.14-2.07 (m, 1H). HRMS (ESI-TOF) calcd for $C_{10}H_{12}N_2O$ $[M + H]^+$: 177.1022, found: 177.1011.

Substrate scope characterization:

^{1}H NMR and high-resolution mass spectrometry were used to characterize products isolated from 1-2 mg reactions. Because of the small amounts of isolated product,^{13}C NMR was not obtained. Reported conversion values were obtained from raw HPLC data. No isolated yields were obtained. Selectivities were determined by comparing ^{1}H NMR integrations of similar aryl protons for the compounds.

6-chloro-2-methyl-tryptamine (7): The bioconversion was conducted in a flask according to the general procedure using 0.08 equiv. of 8F (40 μM final concentration) at 16 °C, with 1 mg of substrate and 10 mM NaCl. After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dried under high vacuum. The crude material was re-suspended in 200 μL of MeOH, filtered, and purified by preparative reverse phase HPLC in 50 μL increments according to HPLC method 1. Product-containing fractions were pooled to afford **7.** ^{1}H NMR (500 MHz; MeOD): δ 7.41 (d, J = 8.4, 1H), 7.27 (s, 1H), 6.99 (dd, J = 8.4, 1.8, 1H), 3.13 (t, J = 7.3, 2H), 3.04 (t, J = 7.6, 2H), 2.40 (s, 3H). HRMS (ESI-TOF) calcd for $C_{11}H_{13}N_2Cl$ $[M + H]^+$: 209.0840 and 211.0811, found: 209.0851 and 211.0824.

5-chloro-2-methyl-tryptamine (8) and 6-chloro-2-methyl-tryptamine (7): The bioconversion was conducted in a flask according to the general procedure using 0.08 equiv. of 10S (40 μM final concentration) at 16 °C, with 1 mg of substrate and 100 mM NaCl. After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dried under high vacuum. The crude material was re-suspended in 200 μL of MeOH, filtered, and purified by preparative reverse phase HPLC in 50 μL increments according to HPLC method 1. Product-containing fractions were pooled to afford a mixture of **7** and **8.** ^{1}H NMR (500 MHz; MeOD): δ 7.47 (d, J = 1.8, 0.168H), 7.41 (d, J = 8.4, 0.642H), 7.27 (d, J = 1.9, 0.476H), 7.23 (d, J = 8.6, 0.158H), 7.02 (dd, J = 8.32, 2.26, 0.209H), 6.99 (ddd, J = 8.4, 1.9, 0.7, 0.529H), 3.12 (m, 2H), 3.03 (m, 2H), 2.40 (m, 3H). HRMS (ESI-TOF) calcd for $C_{11}H_{13}N_2Cl$ $[M + H]^+$: 209.0840 and 211.0811, found: 209.0844 and 211.0840.

6-chloro-N-methyl-tryptamine (9): The bioconversion was conducted in a flask according to the general procedure using 0.1 equiv. of 8F (50 μM final concentration) at 16 °C, with 2 mg of substrate and 10 mM NaCl. After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dried under high vacuum. The crude material was re-suspended in 400 μL of MeOH, filtered, and purified by preparative reverse phase HPLC in 50 μL increments according to HPLC method 1. Product-containing fractions were pooled to afford **9.** ^{1}H NMR (500 MHz; MeOD): δ 7.54 (d, J = 8.5, 1H), 7.39 (d, J = 1.8, 1H), 7.22 (m, 1H), 7.05 (dd, J = 8.5, 1.9, 1H), 3.13 (t, J = 7.4, 2H), 2.71 (s, 3H). HRMS (ESI-TOF) calcd for $C_{11}H_{13}N_2Cl$ $[M + H]^+$: 209.0840 and 211.0811, found: 209.0850 and 211.0818.

5-chloro-N-methyl-tryptamine (10): The bioconversion was conducted in a flask according to the general procedure using 0.1 equiv. of 10S (50 μM final concentration) at 16 °C, with 2 mg of substrate and 100 mM NaCl. After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dried under high vacuum. The crude material was re-suspended in 400 μL of MeOH, filtered, and purified by preparative reverse phase HPLC in 50 μL increments according to HPLC method 1. Product-containing fractions were pooled to afford **10.** ^{1}H NMR (500 MHz; MeOD): δ 7.60 (d, J = 1.9, 1H), 7.35 (d, J = 8.6, 1H), 7.25 (s, 1H), 7.11 (dd, J = 8.6, 1.9, 1H), 3.29 (m, 2H), 3.12 (t, J = 7.4, 2H). HRMS (ESI-TOF) calcd for $C_{11}H_{13}N_2Cl$ $[M + H]^+$: 209.0840 and 211.0811, found: 209.0855 and 211.0823.

6-chloro-tryptophol (11): The bioconversion was conducted in a flask according to the general procedure using 0.1 equiv. of 8F (50 μM final concentration) at 16 °C, with 2 mg of substrate and 10 mM NaCl. After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dried under high vacuum. The crude material was re-suspended in 400 μL of MeOH, filtered, and purified by preparative reverse phase HPLC in 50 μL increments according to HPLC method 1. Product-containing fractions were pooled to afford **11.** ^{1}H NMR (500 MHz; MeOD): δ 7.49 (dd, J = 8.4, 0.5, 1H), 7.32 (dd, J = 1.9, 0.5, 1H), 7.09 (m, 1H), 6.97 (dd, J = 8.4, 1.9, 1H), 3.79 (t, J = 7.2, 2H), 2.94 (td, J = 7.2, 0.8, 2H). HRMS (ESI-TOF) calcd for $C_{10}H_{10}NOCl$ $[M + H]^+$: 196.0524 and 198.0495, found: 196.0509 and 198.0531.

5-chloro-tryptophol (12): The bioconversion was conducted in a flask according to the general

procedure using 0.1 equiv. of 10S (50 μM final concentration) at 16 °C, with 2 mg of substrate and 100 mM NaCl. After maximum conversion to monohalogenated product was observed by HPLC, the reaction mixture was filtered through Celite, extracted into CH_2Cl_2, and dried under high vacuum. The crude material was re-suspended in 400 μL of MeOH, filtered, and purified by preparative reverse phase HPLC in 50 μL increments according to HPLC method 1. Product-containing fractions were pooled to afford **12.** ^{1}H NMR (500 MHz; MeOD): δ 7.52 (s, 1H), 7.29 (d, J = 9.2, 1H), 7.13 (s, 1H), 7.04 (d, J = 8.6, 1H), 3.79 (t, J = 7.2, 2H), 2.93 (t, J = 7.1, 2H). HRMS (ESI-TOF) calcd for $C_{10}H_{10}NOCl$ $[M + H]^+$: 196.0524 and 198.0495, found: 196.0504 and 198.0530.

4.4.4 Additional data

Hanes-Woolf plots and saturation curves for kinetic values in Table 4.1:

Figure 4.22: Hanes-Woolf plot for 0S conversion of tryptamine.

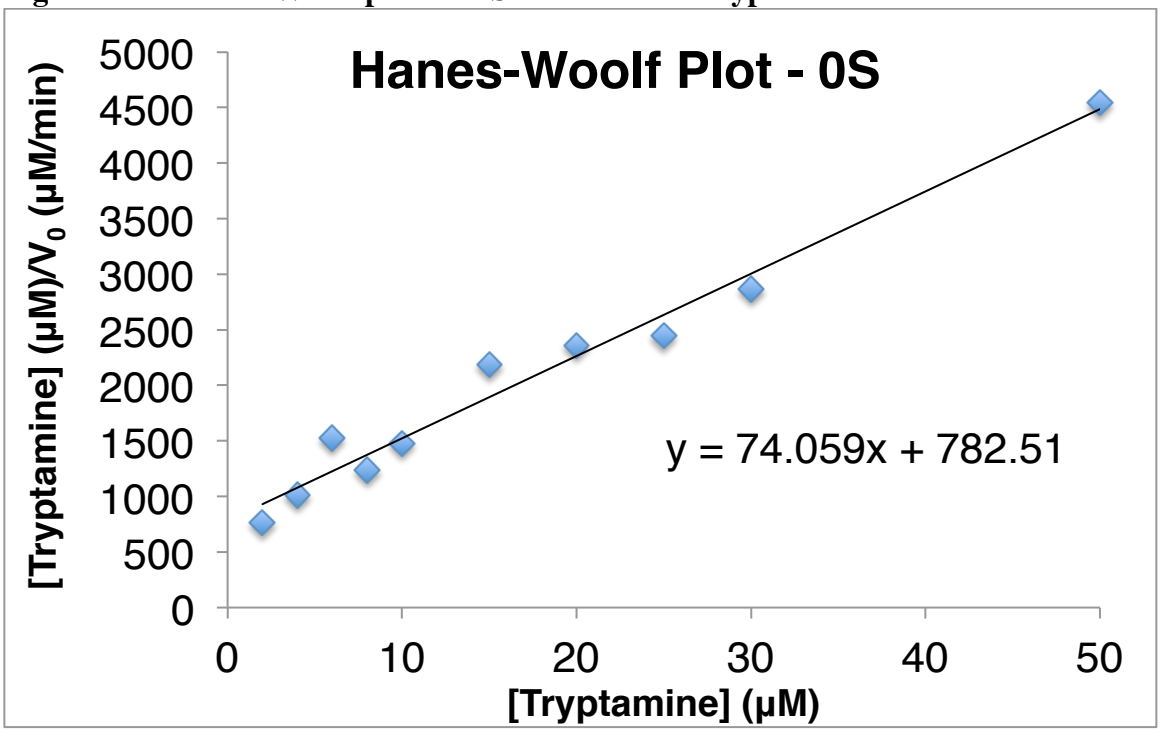


Figure 4.23: Saturation curve for 0S conversion of tryptamine.

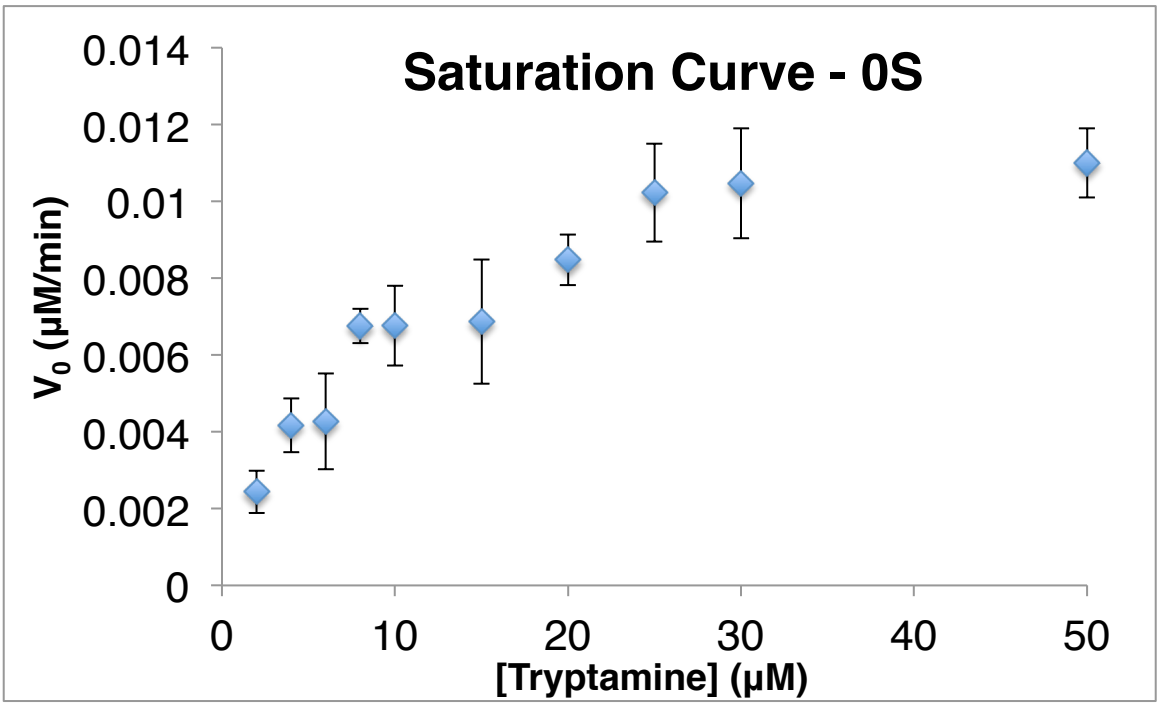


Figure 4.24: Hanes-Woolf plot for 10S conversion of tryptamine.

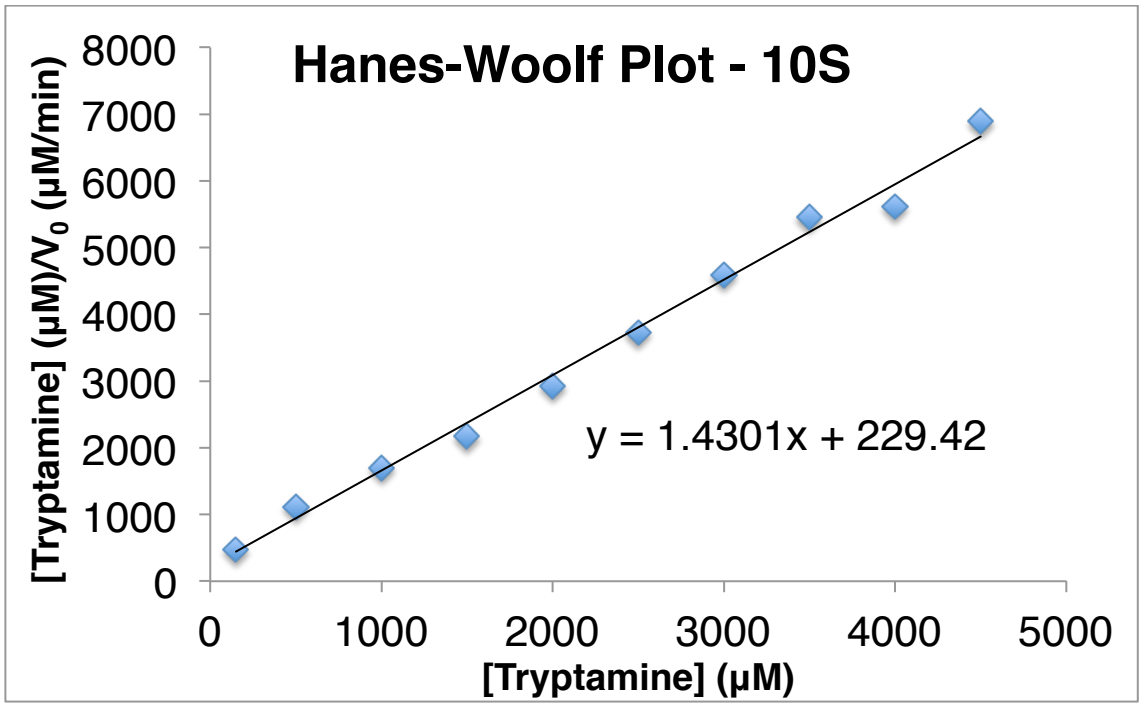


Figure 4.25: Saturation curve for 10S conversion of tryptamine.

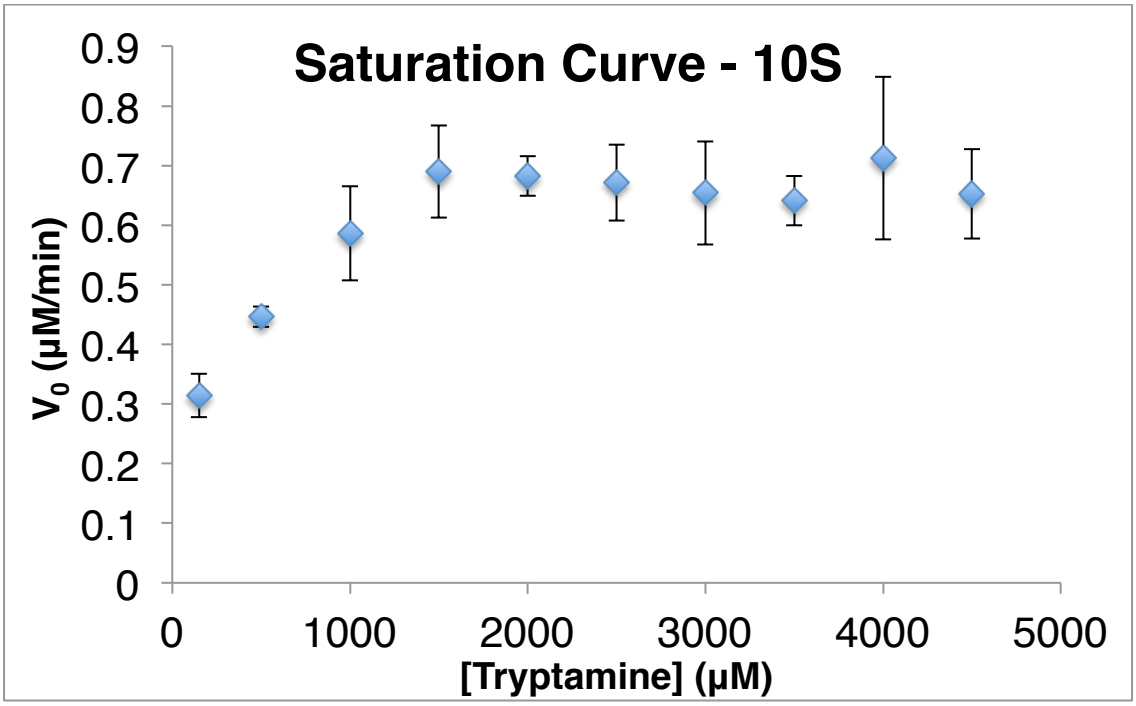


Figure 4.26: Hanes-Woolf plot for 8F conversion of tryptamine.

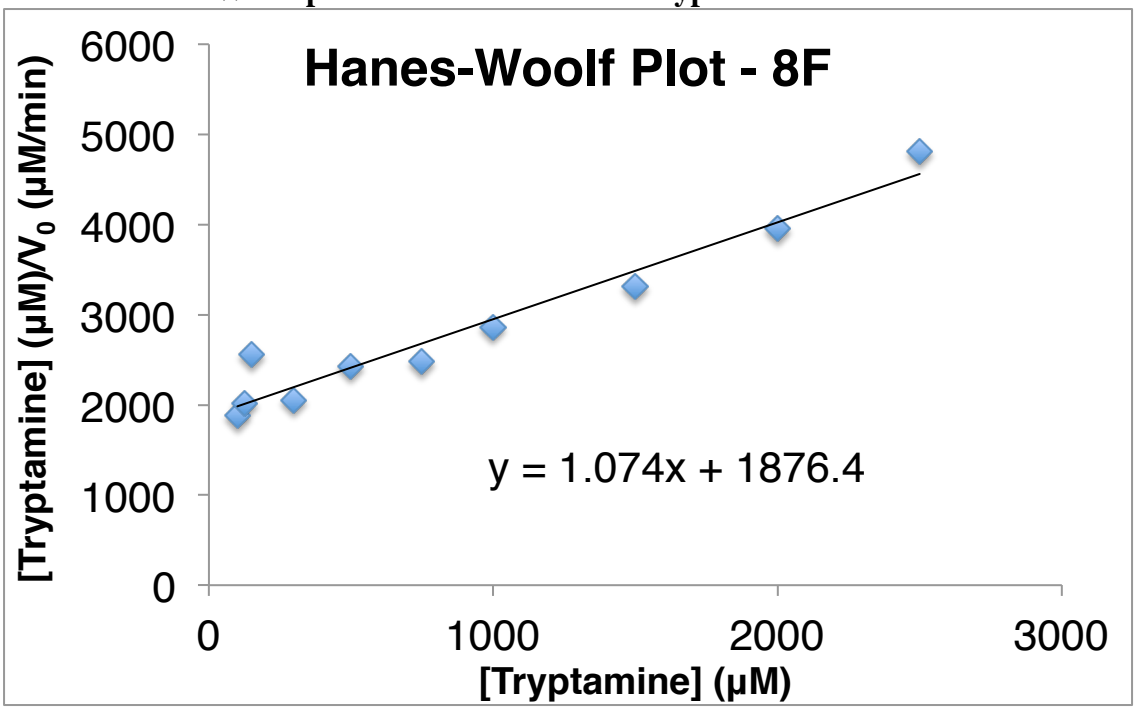
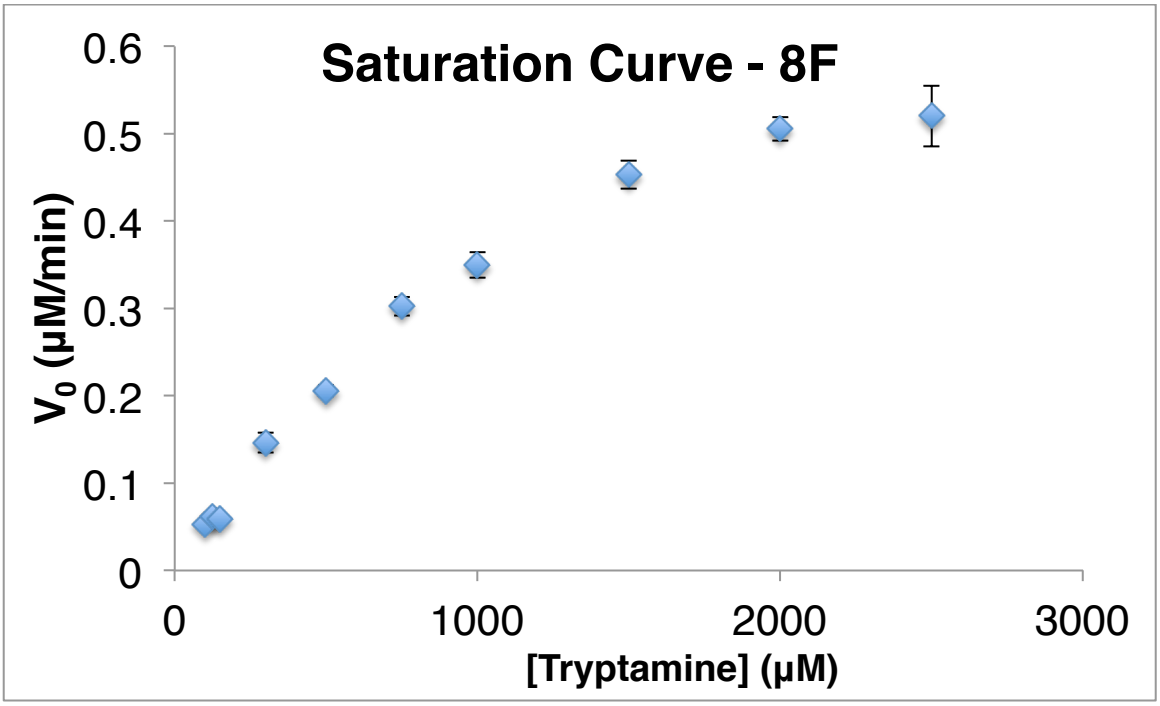


Figure 4.27: Saturation curve for 8F conversion of tryptamine.



Acknowledgements

We would like to thank James Payne for conducting the tryptamine/tryptophan competition reactions, Prof. Benoit Roux and Dr Yilin Meng for assistance with docking simulations, and members of the NSF CCHF for helpful discussions. We would also like to thank Prof. Jennifer Bridwell-Rabb for advice on protein crystal soaking experiments and Ken Ellis-Guardiola for advice on protein structure determination.

4.5 References

1. Bergman, R. G. C–H activation. *Nature* (2007).
2. Chen, D. Y. K. & Youn, S. W. C–H Activation: A Complementary Tool in the Total Synthesis of Complex Natural Products. *Chem. Eur. J.* **18,** 9452–9474 (2012).
3. O'Hara, F., Blackmond, D. G. & Baran, P. S. Radical-Based Regioselective C–H Functionalization of Electron-Deficient Heteroarenes: Scope, Tunability, and Predictability. *J. Am. Chem. Soc.* **135,** 12122–12134 (2013).
4. Ackermann, L. Carboxylate-Assisted Transition-Metal-Catalyzed C–H Bond Functionalizations: Mechanism and Scope. *Chem. Rev.* **111,** 1315–1345 (2011).
5. Jia, C. G. *et al*. Efficient activation of aromatic C-H bonds for addition to C-C multiple bonds. *Science* **287,** 1992–1995 (2000).
6. Hartwig, J. F. Borylation and Silylation of C–H Bonds: A Platform for Diverse C–H Bond Functionalizations. *Acc. Chem. Res.* **45,** 864–873 (2012).
7. Lyons, T. W. & Sanford, M. S. Palladium-Catalyzed Ligand-Directed C–H Functionalization Reactions. *Chem. Rev.* **110,** 1147–1169 (2010).
8. Colby, D. A., Bergman, R. G. & Ellman, J. A. Rhodium-Catalyzed C–C Bond Formation via Heteroatom-Directed C–H Bond Activation. *Chem. Rev.* **110,** 624–655 (2010).
9. Arockiam, P. B., Bruneau, C. & Dixneuf, P. H. Ruthenium(II)-Catalyzed C–H Bond Activation and Functionalization. *Chem. Rev.* **112,** 5879–5918 (2012).
10. Maddox, S. M., Nalbandian, C. J., Smith, D. E. & Gustafson, J. L. A Practical Lewis Base Catalyzed Electrophilic Chlorination of Arenes and Heterocycles. *Org. Lett.* **17,** 1042–1045 (2015).
11. Leitch, J. A., McMullin, C. L., Mahon, M. F., Bhonoah, Y. & Frost, C. G. Remote C6-Selective Ruthenium-Catalyzed C–H Alkylation of Indole Derivatives via σ-Activation. *ACS Catal.* **7,** 2616–2623 (2017).
12. Seregin, I. V. & Gevorgyan, V. Direct transition metal-catalyzed functionalization of heteroaromatic compounds. *Chem. Soc. Rev.* **36,** 1173–21 (2007).
13. Zhu, Y. & Rawal, V. H. Palladium-Catalyzed C3-Benzylation of Indoles. *J. Am. Chem. Soc.* **134,** 111–114 (2012).
14. Phipps, R. J., Grimster, N. P. & Gaunt, M. J. Cu(II)-Catalyzed Direct and Site-Selective Arylation of Indoles Under Mild Conditions. *J. Am. Chem. Soc.* **130,** 8172–8174 (2008).
15. Deprez, N. R., Kalyani, D., Krause, A. & Sanford, M. S. Room Temperature Palladium-Catalyzed 2-Arylation of Indoles. *J. Am. Chem. Soc.* **128,** 4972–4973 (2006).
16. Nadres, E. T., Lazareva, A. & Daugulis, O. Palladium-Catalyzed Indole, Pyrrole, and Furan Arylation by Aryl Chlorides. *J. Org. Chem.* **76,** 471–483 (2011).
17. Lebrasseur, N. & Larrosa, I. Room Temperature and Phosphine Free Palladium Catalyzed Direct C-2 Arylation of Indoles. *J. Am. Chem. Soc.* **130,** 2926–2927 (2008).
18. Li, J. & Ackermann, L. Cobalt-Catalyzed C H Cyanation of Arenes and Heteroarenes. *Angew. Chem. Int. Ed.* **54,** 3635–3638 (2014).

19. Yang, Y., Qiu, X., Zhao, Y., Mu, Y. & Shi, Z. Palladium-Catalyzed C–H Arylation of Indoles at the C7 Position. *J. Am. Chem. Soc.* **138,** 495–498 (2016).
20. Hartung, C. G., Fecher, A., Chapell, B. & Snieckus, V. Directed orthoMetalation Approach to C-7-Substituted Indoles. Suzuki–Miyaura Cross Coupling and the Synthesis of Pyrrolophenanthridone Alkaloids. *Org. Lett.* **5,** 1899–1902 (2003).
21. Paul, S. *et al.* Ir-Catalyzed Functionalization of 2-Substituted Indoles at the 7-Position: Nitrogen-Directed Aromatic Borylation. *J. Am. Chem. Soc.* **128,** 15552–15553 (2006).
22. Robbins, D. W., Boebel, T. A. & Hartwig, J. F. Iridium-Catalyzed, Silyl-Directed Borylation of Nitrogen-Containing Heterocycles. *J. Am. Chem. Soc.* **132,** 4068–4069 (2010).
23. Yang, G. *et al.* Pd(II)-Catalyzed meta-C–H Olefination, Arylation, and Acetoxylation of Indolines Using a U-Shaped Template. *J. Am. Chem. Soc.* **136,** 10807–10813 (2014).
24. Feng, Y. *et al.* Total Synthesis of Verruculogen and Fumitremorgin A Enabled by Ligand-Controlled C-H Borylation. *J. Am. Chem. Soc.* **137,** 10160–10163 (2015).
25. Yang, Y., Li, R., Zhao, Y., Zhao, D. & Shi, Z. Cu-Catalyzed Direct C6-Arylation of Indoles. *J. Am. Chem. Soc.* **138,** 8734–8737 (2016).
26. Liu, H., Zheng, C. & You, S.-L. Catalytic C6 Functionalization of 2,3-Disubstituted Indoles by Scandium Triflate. *J. Org. Chem.* **79,** 1047–1054 (2014).
27. Simonetti, M. *et al.* Ruthenium-Catalyzed C–H Arylation of Benzoic Acids and Indole Carboxylic Acids with Aryl Halides. *Chem. Eur. J.* **23,** 549–553 (2016).
28. Lewis, J. C., Coelho, P. S. & Arnold, F. H. Enzymatic functionalization of carbon–hydrogen bonds. *Chem. Soc. Rev.* **40,** 2003–2021 (2011).
29. Weichold, V., Milbredt, D. & van Pée, K.-H. Specific Enzymatic Halogenation-From the Discovery of Halogenated Enzymes to Their Applications In Vitro and In Vivo. *Angew. Chem. Int. Ed. Engl.* **55,** 6374–6389 (2016).
30. Arnold, F. H. Design by directed evolution. *Acc. Chem. Res.* (1998).
31. Schmidt-Dannert, C. & Arnold, F. H. Directed evolution of industrial enzymes. in **17,** 135–136 (1999).
32. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485,** 185–194 (2012).
33. Fishman, A., Tao, Y., Rui, L. & Wood, T. K. Controlling the regiospecific oxidation of aromatics via active site engineering of toluene para-monooxygenase of Ralstonia pickettii PKO1. *J. Biol. Chem.* **280,** 506–514 (2005).
34. Lewis, J. C. *et al.* Chemoenzymatic elaboration of monosaccharides using engineered cytochrome P450BM3 demethylases. *Proc. Natl. Acad. Sci.* **106,** 16550–16555 (2009).
35. Fasan, R. Tuning P450 Enzymes as Oxidation Catalysts. *ACS Catal.* **2,** 647–666 (2012).
36. Payne, J. T., Poor, C. B. & Lewis, J. C. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angew. Chem.* **127,** 4300–4304 (2015).
37. Lang, A. *et al.* Changing the Regioselectivity of the Tryptophan 7-Halogenase PrnA by Site-Directed Mutagenesis. *Angew. Chem. Int. Ed.* **50,** 2951–2953 (2011).

38. Shepherd, S. A. *et al*. Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chem. Sci*. **6,** 3454–3460 (2015).
39. Shepherd, S. A. *et al*. A Structure-Guided Switch in the Regioselectivity of a Tryptophan Halogenase. *ChemBioChem* **17,** 821–824 (2016).
40. Andorfer, M. C., Park, H. J., Vergara-Coll, J. & Lewis, J. C. Directed evolution of RebH for catalyst-controlled halogenation of indole C–H bonds. *Chem. Sci*. **7,** 3720–3729 (2016).
41. Payne, J. T., Andorfer, M. C. & Lewis, J. C. Regioselective Arene Halogenation using the FAD-Dependent Halogenase RebH. *Angew. Chem. Int. Ed*. **52,** 5271–5274 (2013).
42. Glenn, W. S., Nims, E. & O'Connor, S. E. Reengineering a Tryptophan Halogenase To Preferentially Chlorinate a Direct Alkaloid Precursor. *J. Am. Chem. Soc*. **133,** 19346–19349 (2011).
43. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol*. **10,** 866–876 (2009).
44. Poor, C. B., Andorfer, M. C. & Lewis, J. C. Improving the Stability and Catalyst Lifetime of the Halogenase RebH By Directed Evolution. *ChemBioChem* **15,** 1286–1289 (2014).
45. PHILLIPS, R. S. & COHEN, L. A. Intramolecular General Acid and General Base Catalyzes in the Hydrolysis of 2-Halotryptophans and Their Analogs. *J. Am. Chem. Soc*. **108,** 2023–2030 (1986).
46. Caster, Library Evaluation Tool.
47. Matsumura, I. & Ellington, A. D. In vitro Evolution of Beta-glucuronidase into a Beta-galactosidase Proceeds Through Non-specific Intermediates. *J. Mol. Biol*. **305,** 331–339 (2001).
48. Lewis, J. C. & Arnold, F. H. Catalysts on Demand: Selective Oxidations by Laboratory-Evolved Cytochrome P450 BM3. *CHIMIA* **63,** 309–312 (2009).
49. Zhu, X. *et al*. Structural Insights into Regioselectivity in the Enzymatic Chlorination of Tryptophan. *J. Mol. Biol*. **391,** 74–85 (2009).
50. Bitto, E. *et al*. The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins* **70,** 289–293 (2007).
51. Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases †,‖. *Biochemistry* **46,** 1284–1292 (2007).
52. Flecks, S. *et al*. New Insights into the Mechanism of Enzymatic Chlorination of Tryptophan. *Angew. Chem. Int. Ed*. **47,** 9533–9536 (2008).
53. Dong, C. J. *et al*. Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* **309,** 2216–2219 (2005).
54. Zhang, C. *et al*. π-Clamp-mediated cysteine conjugation. *Nat. Chem*. **8,** 120–128 (2015).
55. Yeh, E., Garneau, S. & Walsh, C. T. Robust in vitro Activity of RebF and RebH, a Two-Component Reductase/Halogenase, Generating 7-Chlorotryptophan during Rebeccamycin Biosynthesis. *Proc. Natl. Acad. Sci*. **102,** 3960–3965 (2005).
56. Cabrita, L. D., Dai, W. & Bottomley, S. P. A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol* **6,** 12 (2006).

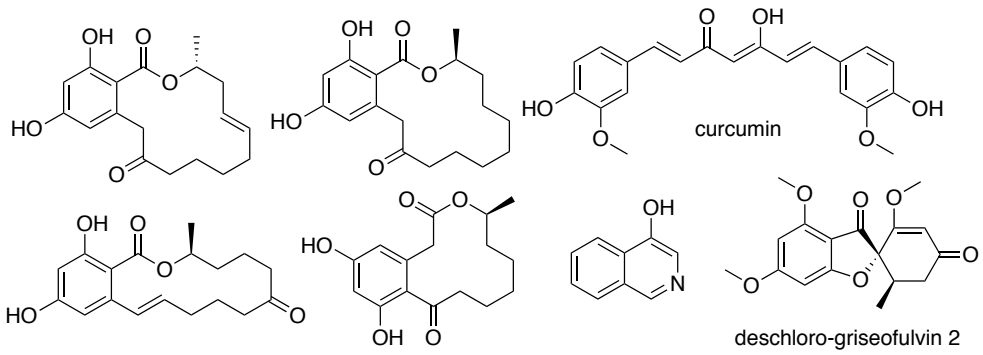
57. Hara, T., Durell, S. R., Myers, M. C. & Appella, D. H. Probing the Structural Requirements of Peptoids That Inhibit HDM2–p53 Interactions. *J. Am. Chem. Soc.* **128,** 1995–2004 (2006).
58. Nicolaou, K. C., Krasovskiy, A., Trépanier, V. É. & Chen, D. Y. K. An Expedient Strategy for the Synthesis of Tryptamines and Other Heterocycles. *Angew. Chem. Int. Ed.* **47,** 4217–4220 (2008).
59. Pham, V. C., Ma, J., Thomas, S. J., Xu, Z. & Hecht, S. M. Alkaloids from Alangiumjavanicumand Alangiumgrisolleoidesthat Mediate Cu 2+-Dependent DNA Strand Scission. *J. Nat. Prod.* **68,** 1147–1152 (2005).
60. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning: a laboratory manual, 2nd edn.* (Cold Spring Laboratory Press, 1989).
61. Gottlieb, H. E., Kotlyar, V. & Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *J. Org. Chem.* **62,** 7512–7515 (1997).
62. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2,** 924–932 (2007).
63. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *ELECTROPHORESIS* **30,** S162–S173 (2009).
64. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **65,** NA–NA (2009).
65. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276,** 307–326 (1997).
66. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr.* **61,** 458–464 (2005).
67. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
68. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).
69. Miyake, F. Y., Yakushijin, K. & Horne, D. A. Preparation and Synthetic Applications of 2-Halotryptamines: Synthesis of Elacomine and Isoelacomine. *Org. Lett.* **6,** 711–713 (2004).

Chapter 5: Understanding flavin-dependent halogenase reactivity via substrate activity profiling

5.1 Introduction

Flavin-dependent halogenases (FDHs) can be grouped into two broad classes based on whether they act on free substrate and or acyl-carrier-protein (ACP) bound substrates.[1,2] Studies on FDHs that halogenate free substrates have largely been limited to L-tryptophan halogenases (Trp-FDHs).[1] Recently, Rdc2, a fungal FDH with activity on several free substrates, including resorcylic acid lactone substrates, curcumin, and hydroxyisoquinolines (Fig. 5.1), was reported.[3-5] In addition, GsfI, a homolog of Rdc2, catalyzes a selective chlorination in the biosynthesis of griseofulvin (Fig. 5.1).[6] At the time of this work, no information comparing the substrate scopes of Trp-FDHs and these fungal halogenases had been reported.

Figure 5.1: Previously characterized substrate scope of Rdc2 and GsfI.



Furthermore, a through investigation of the substrate scope of natural and engineered Trp-FDH on a variety of different substrate classes had not yet been demonstrated. Section 2.2.2 of this thesis describes our initial work towards characterizing the substrate scope of the 7-

tryptophan halogenase, RebH.[7] Although several non-native substrates were accepted as substrates and halogenated by RebH, most of these contain an indole core, like the native substrate, L-tryptophan. The scope of RebH has been extended by the Lewis lab to larger, biologically active substrates through development of new RebH variants such as 4V; however, halogenation was demonstrated on only indole-containing compounds.[8] As described in Chapter 4 of this thesis, RebH variants 8F and 10S display altered site selectivity on a variety of 3-substituted indoles.[9] In addition to indole-containing compounds, the activity of RebH variant 0K[10] as well as variants of PrnA[11] was explored on a handful of aniline compounds, most containing a carboxylic acid moiety. Beyond RebH and variants, substantially less information is known about the substrate scope of the 6-tryptophan halogenase, Thal, which has demonstrated halogenation activity on both L- and D-tryptophan.[12]

Given that Trp-FDHs have demonstrated great potential as biocatalysts for site selective aromatic halogenation, we sought to establish detailed profiles of FDH activity on diverse panels of small molecule substrates.[13,14] With these profiles, we hoped to compare substrate scope and specificities for natural Trp-FDHs, engineered Trp-FDHs, as well as the fungal FDHs Rdc2 and GsfI. Using these substrate activity profiles, we hoped to gain insight into FDH specificity and site selectivity through a combination of techniques, including halenium affinity calculations[15] and substrate docking. A paper has been published summarizing these findings in collaboration with Novartis Institutes for Biomedical Research through the Center for Selective C-H Functionalization.[16]

Authorship

Docking studies in section 5.2.3 were conducted by Dr. Christine E. Hajdin, an investigator at Novartis Institutes for Biomedical Research (NIBR). Cloning of Rdc2 and GsfI

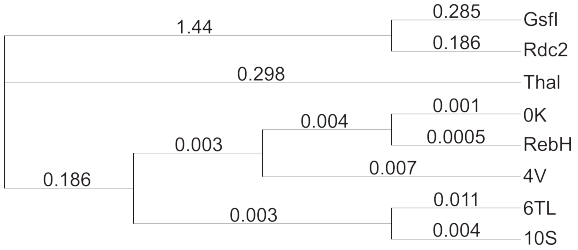
genes was performed by Dr. Piro Siuti and Jeremiah Lilly, investigators at NIBR. Characterization of chlorinated products was aided by Jonathan Grob, an investigator at NIBR, and Julia Chael, an undergraduate student in the Lewis lab.

5.2 Results and Discussion

5.2.1 Constructing FDH substrate activity profiles

Substrate activity profiles were generated by evaluating the activity of 8 FDHs on 93 substrates. To evaluate the differences between known wild-type tryptophan halogenases (Trp-FDHs) and engineered RebH variants, we selected RebH, Thal, and RebH variants 0K, 4V, 6TL, and 10S for analysis. Rdc2 and GsfI were also analyzed to assess differences in substrate scope between Trp-FDHs and other FDHs. Phylogenetic analysis of these 8 FDHs indicates that they largely fall into two categories according to organism of origin (bacterial and fungal, Figure 5.2). RebH and Thal, which are both bacterial Trp-FDHs, have higher sequence homology to one another (61% identity, 99% query coverage) than to RebH and the fungal FDH Rdc2 (43% identity, 48% query coverage). We hypothesized that this difference in primary sequence could result in significant differences in substrate scope.

Figure 5.2: Phylogenetic tree of FDHs examined.[a]



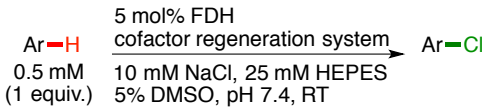
[a] Branch labels display amino acid substitutions per site.

The 93 substrates selected for evaluation were grouped into two panels based on their structures. One panel consisted of 86 *N*-containing aromatic substrates (indoles, pyrroles, azoles, anilines, and anilides). The second panel contained 7 phenols and anisoles, taking into

consideration the native substrates of the fungal FDHs. Because FDH halogenation proceeds through electrophilic aromatic substitution,[17,18] all substrates in both panels contained at least one electronically activated site. Substrates were also chosen to probe significant steric and electronic variation, as well as functional group substitution. Moreover, the compounds in these panels were acquired from the NIBR fragment compound library and are therefore representative of those commonly used in fragment-based drug design.[19] These studies were thus expected to provide valuable insight into FDH reactivity toward motifs common in pharmaceuticals.

To obtain sufficient quantities of enzyme for constructing substrate activity profiles for these 8 FDHs on 93 aromatic substrates, I co-expressed the FDH genes with the pGro7 chaperone system as described in Chapter 2.[7] High yields of purified enzyme were obtained for all 8 FDHs, demonstrating that optimal RebH expression conditions are can also produce high quantities of soluble Rdc2 and GsfI. With purified FDHs in hand, I set up two analytical bioconversions on each substrate for each FDH (1,488 reactions, Scheme 5.1). One set of bioconversions proceeded overnight while the other set was quenched with one volume of methanol after one hour. All bioconversions were analyzed by LCMS and conversion of starting material to chlorinated product(s) was calculated. Substrate activity profiles in heat map form were generated from this conversion data (Fig. 5.3). Full profiles containing quantitative data can be found in section 5.4.3 (Table 5.2).

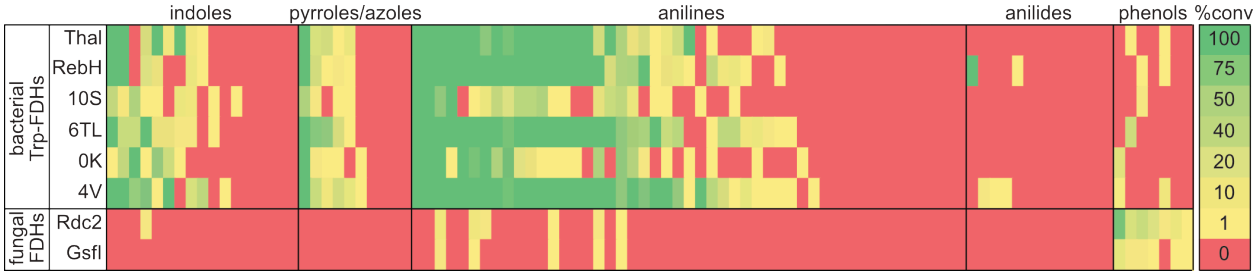
Scheme 5.1: General scheme for analytical FDH bioconversions used to construct substrate activity profiles.



Analysis of these substrate activity profiles indicated that 69% of panel 1 (59 substrates) was halogenated by at least one of the Trp-FDHs (Fig. 5.3). Remarkable functional group

tolerance of these FDHs was displayed, with substrates containing amines, alcohols, amides, esters, nitriles, thioesters, sulfonamides, pyridines, azoles, quinolones, and pyrroles all undergoing halogenation (all structures tested can be found in Table 5.2). Quantitatively, Thal, RebH, and 6TL demonstrated similarly broad substrate scopes, with halogenation of 40-47% of panel 1 substrates with >1% conversion. 0K and 10S displayed much narrower substrate scope, with only 23-24% of panel 1 substrates undergoing halogenation with >1% conversion. As discussed in Chapter 3, 0K was engineered for higher activity on benzoic acid-containing substrates by replacing active site Glu461 with Lys. The narrower scope observed in this study could have been due to the limited number of substrates that contained acidic functionality. The narrower substrate scope of 10S compared to 6TL is consistent with previous examples in which evolutionary intermediates serve as more general catalysts than final variants with high activity/selectivity on a target substrate.[20,21] 10S was the final variant in a directed evolution campaign to site-selectively halogenate the 5-position of tryptamine, whereas 6TL was a variant discovered along this campaign that halogenated the 5-, 6-, and 7-positions of tryptamine with similar efficiencies.[9] Interestingly, the FDH with the broadest substrate scope on panel 1 substrates was 4V. This variant had been previously engineered for increased RebH substrate scope to larger, biologically active indole-containing compounds.[8] It appears that widening the substrate scope to certain target compounds enabled a broader, more general scope as well.

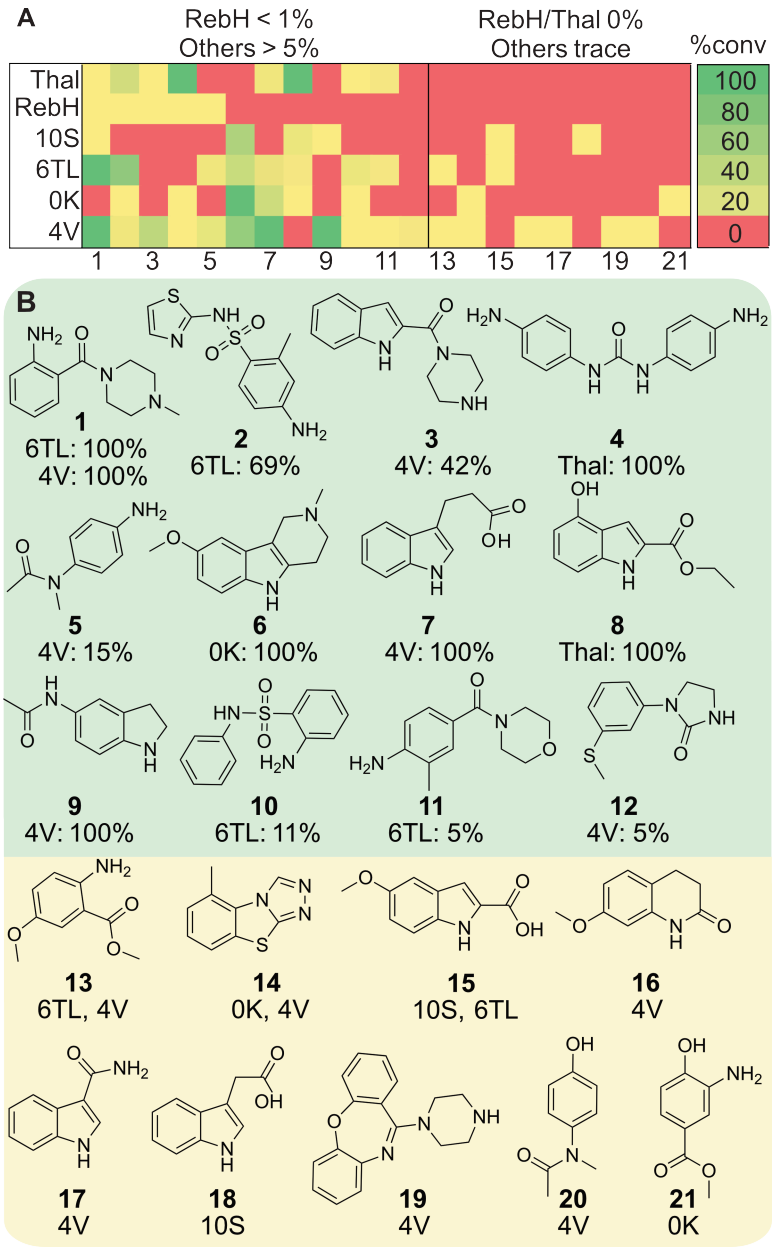
Figure 5.3: Substrate activity profiles in heat map form for eight FDHs on substrates in panels 1 (indoles, pyrroles, azoles, anilines, anildes) and 2 (phenols).[a]



[a] Reactions conditions are described in Scheme 5.1. Maximum conversion is shown for each enzyme-substrate pair. For complete data sets at both time points, see Table 5.2.

Next, we more closely examined the scope of RebH compared to Thal and the engineered variants. Several substrates that were not halogenated or were only halogenated to a trace amount by RebH were halogenated by Thal or engineered variants to a significant extent (Fig. 5.4, **1**-**12**). In addition, many more substrates that were not halogenated by either wild-type Trp-FDH were halogenated to trace amounts by engineered variants (Fig. 5.4, **13**-**21**). This trace activity, determined by LCMS using chlorine isotope patterns, can be used as a starting point for evolution. Indeed, we have demonstrated that good activities and selectivities can be evolved from FDHs with only trace activity[8] or selectivity[9] for a particular substrate or site. Moreover, these findings highlight that minor changes in primary sequence can have substantial impact on FDH specificity and scope; thus, it is important to evaluate multiple FDHs on a substrate(s) to find active catalysts, be it for preparative reactions or starting points for evolution.

Figure 5.4: Substrate activity profiles (A) and structures (B) of selected substrates.[a]

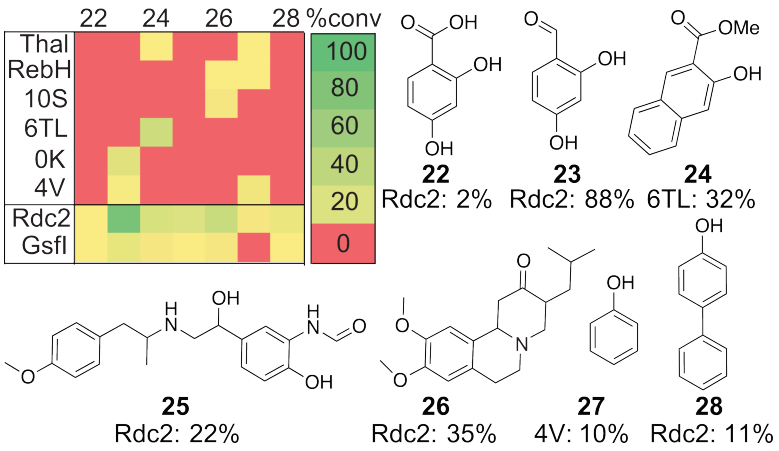


[a] Compounds for which >5% conversion was observed are highlighted in green. % Conversion of substrate to chlorinated product is provided for these substrates (**1**-**12**). Compounds halogenated to a trace extent are highlighted in yellow.

Although the Trp-FDHs chlorinated the majority of substrates in panel 1, very few were accepted as substrates by the fungal FDHs (8%, 7 substrates, Fig. 5.3). For this reason, a second panel of phenols and anisoles, substrates more closely resembling these enzymes native

substrates, was analyzed with all 8 FDHs. The Trp-FDHs were not able to halogenate these less nucleophilic substrates[22] to as great of an extent as the more electronically activated aniline, indole, pyrrole, and azole compounds; however, all panel 2 substrates were chlorinated by at least one of the fungal FDHs (Fig. 5.5). This complementarity in substrate scope between the Trp-FDHs and fungal FDHs was particularly interesting to us. As described in section 3.2.2, a key aim in the Lewis lab is to halogenate less electronically activated positions on a compound with FDHs. While the directed evolution campaign described in section 3.2.2 is a valid option to accomplish this, the substrate scope of Rdc2 and GsfI have led us to pursue genome mining for natural FDHs as an alternate approach to discovering new halogenation reactivity.

Figure 5.5: Conversion data from initial substrate activity profiles and structures for compounds in panel 2.[a]



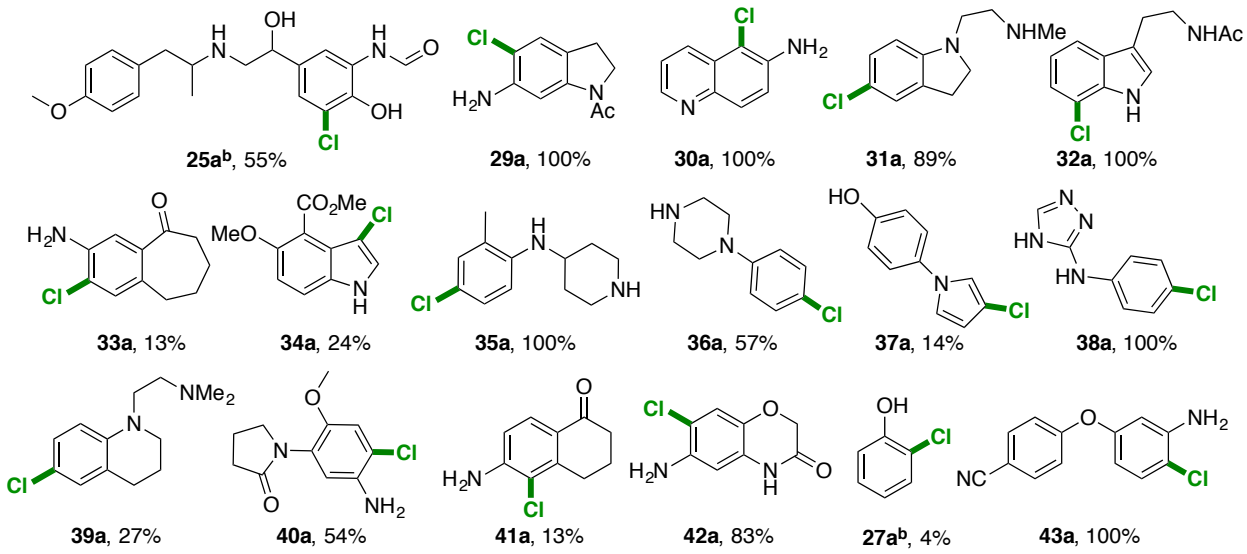
[a] Conversion data from initial activity profile and structures for substrates **22**-**28**.

5.2.2 Selectivity of FDH halogenation

After evaluating the substrate scope for 8 FDHs, we next determined the site selectivity of FDH halogenation, beginning with RebH. Preparative bioconversions (1-10 mg) were performed for all substrates that RebH halogenated with more than trace conversion. Chlorinated

products were purified by preparative LCMS and characterized by NMR and HRMS. Notably, only one monochlorinated isomer was observed for the majority of these compounds (22 of 27 bioconversions, Fig. 5.6 and Table 5.3). Moreover, for many of these compounds, more than one site was activated for electrophilic aromatic substitution (Fig. 5.6). For example, **35** is highly activated for EAS at the ortho and para positions to he aniline substituent; however, only chlorination at the para position (**35a**) is observed.

Figure 5.6: Site selectivity of RebH and Rdc2 on representative substrates.[a]



[a] Conversions for substrates with multiple electronically activated sites that provided a single halogenated product. Reactions were conducted on 1-10 mg scale using either RebH (**29a**-**43a**) or Rdc2[b] (**25a**, **27a**). See Table 5.3 for full product list. Conversion and selectivity for **27a** were determined by comparison with authentic standard.

Given that site-selective RebH-catalyzed chlorination was observed even on non-native substrates differing significantly from L-tryptophan (Fig. 5.6), we next compared this selectivity with that of the small-molecule chlorinating agent *N*-chlorosuccinimide (NCS). Several substrates for which RebH-catalyzed chlorination provided a single isomer were chlorinated at several positions with NCS (Table 5.1, entries 2-4, 6). A remarkable example of this is substrate

43 (Table 5.1, entry 4), which has three unique, electronically activated positions - two unique positions ortho and one para to the aniline substituent. Although RebH chlorinates nearly exclusively (98% selectivity) at one of the ortho positions, NCS gives a product mixture of all three chlorinated isomers.

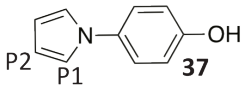
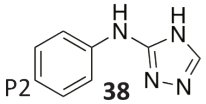
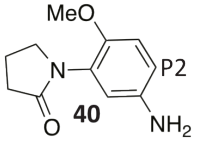
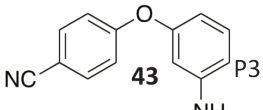
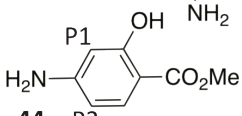
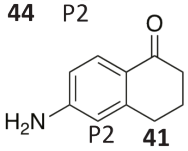
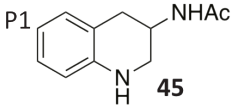
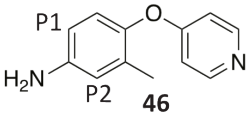
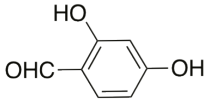
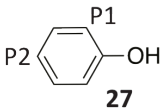
Excitingly, one case arose for which both NCS and RebH produced a single chlorinated product; however, upon NMR characterization of both products, they were determined to be different isomers. NCS halogenated substrate **37** selectively at the 2-position of the pyrrole (i.e. the most electronically activated site) while RebH only catalyzed chlorination of the less electronically activated 3-position (Table 5.1, entry 1). Although complete overriding of substrate electronics has been observed with Trp-FDHs for 3-substituted indoles, such high levels of catalyst control on a substrate so different from L-tryptophan is notable.

In some cases, RebH chlorinated substrates at multiple sites, producing product mixtures. When compared with the chlorination product mixtures obtained using NCS as the halogenating species, different product ratios were often observed (Table 5.1, entries 7-8). This suggests that even when RebH produces product mixtures, catalyst control is still responsible in part for observed selectivities and could be further improved through directed evolution.

Because RebH site selectivity clearly involves factors other than substrate electronic activation, we next wondered if the same was true for the fungal FDHs. Substrate **25**, a long-acting β_2 agonist known as formoterol,[23] contains three sites activated for EAS. Rdc2 produced higher conversion of **25** compared to GsfI, so a preparative-scale Rdc2-catalyzed chlorination of **25** was conducted to determine site selectivity. From this bioconversion, only one chlorinated isomer was observed (Fig. 5.1, **25a**). Subsequently, reaction of **25** with NCS was tested and found to produce a large mixture of compounds. A single chlorinated isomer could not be

isolated from this mixture using standard purification techniques (Biotage reverse phase chromatography and preparative HPLC).

Table 5.1: Relative product distributions for halogenation of representative substrates using FDHs and NCS.[a]

Entry	Substrate	Species	%1	%2	%3	% di	% conv.
1	 **37**	NCS	100	0	-	-	97
		RebH	0	100	-	-	55
2	 **38**	NCS	18	41	29	13	87
		RebH	0	100	0	0	32
		6TL	17	83	0	0	51
3	 **40**	NCS	7	65	-	28	86
		RebH	0	100	-	0	35
4	 **43**	NCS	41	20	30	9	80
		RebH	2	0	98	0	94
5	 **44**	NCS	63	37	-	-	94
		6TL	3	97	-	-	100
6	 **41**	NCS	10	90	-	-	98
		RebH	0	100	-	-	15
		6TL	93	7	-	-	81
		4V	70	30	-	-	22
7	 **45**	NCS	2	25	-	73	65
		RebH	50	0	-	50	50
		6TL	58	42	-	0	44
8	 **46**	NCS	52	43	-	5	94
		RebH	30	70	-	0	74
		4V	77	23	-	0	81
9	 **23**	NCS	37	52	-	11	62
		Rdc2	75	18	-	7	44
		GsfI	91	9	-	0	2
10	 **27**	NCS	23	75	-	2	94
		Rdc2	100	0	-	-	4
		4V	0	100	-	-	10

[a] Known sites of halogenation are indicated for select substrates (P1-P3).

Like formoterol, phenol (**27**) has more than one electronically activated site and is halogenated by Rdc2. Thus, we sought to determine the site selectivity on this substrate as well. An analytical, Rdc2-catalyzed bioconversion was conducted with phenol as substrate and analyzed by GCMS. A single peak was observed, and when compared with authentic standards of ortho- and para-chlorophenol, it was determined to be ortho chlorinated phenol (Fig. 5.6, **27a**). Thus, Rdc2, like RebH, is able to override substrate electronics and site selectively halogenate non-native substrates.

Lastly, we sought to compare the site selectivity of chlorination between FDHs. A representative subset of substrates was tested with several FDHs, and LCMS was used to determine site selectivities. Interestingly, complete switches in selectivity of chlorination were observed in some cases (Table. 5.1, entries 6, 10). **41** is halogenated exclusively at one of the positions ortho to the aniline substituent with RebH; however, high selectivity (93%) for the other ortho position is observed with 6TL. As expected, NCS chlorinates both of these ortho positions. Similarly, a complete switch in selectivity is observed for chlorination of phenol when 4V is used as the catalyst as opposed to Rdc2. As mentioned above, Rdc2 halogenates the ortho position of phenol; however, only para chlorination is observed with 4V. This particular example highlights the power of FDH catalyst control for aromatic halogenation, whereby even a simple non-native substrate can be site selectively halogenated.

For many substrates, different product ratios were observed when different FDHs were used for chlorination (Table 5.1, entries 2, 6-9). For example, **46** is primarily chlorinated ortho to the methyl substituent by RebH (70% selectivity); however, 4V primarily chlorinates para to the methyl substituent (77% selectivity). Reactions with NCS produce a fairly even mixture of these two isomers (Table 5.1, entry 8). Halogenation with FDHs presents the unique opportunity to

tune this site selectivity, as evidenced by the work described in Chapter 4. Even though product mixtures are observed for substrates such as **46**, RebH and 4V favor chlorination at different sites and could be engineered to have high site selectivities for these positions if desired.

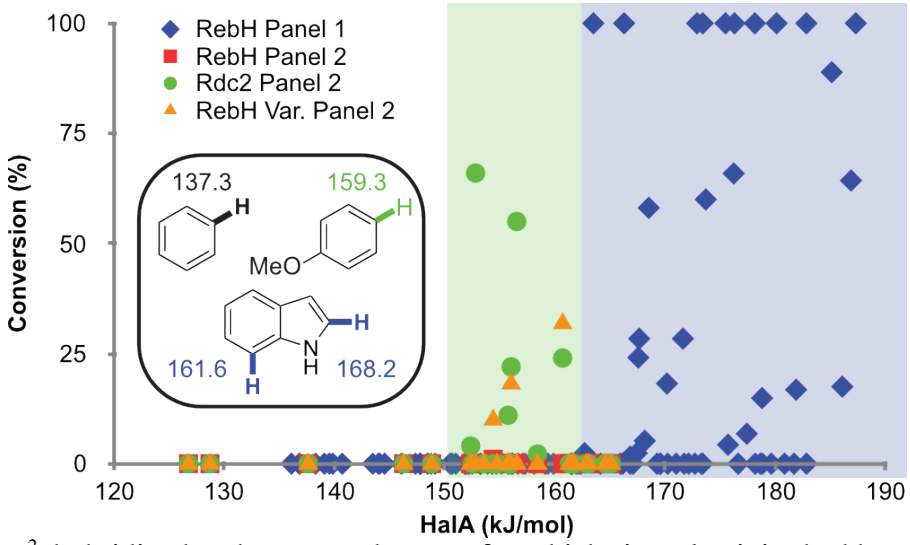
5.2.3 The Effect of substrate binding and substrate electronics on FDH halogenation

With large amounts of conversion data and selectivity data from sections 5.2.1 and 5.2.2 in hand, we next sought to better understand factors underlying the observed substrate specificities and selectivities.[24] From the results described in section 5.2.2, it was clear to us that substrate binding influenced the observed site selectivities, even for non-native substrates, since the results were not consistent with substrate electronics alone. Nonetheless, because FDH halogenation proceeds through an EAS mechanism,[17,25] substrate electronics should have some effect on specificity and selectivity as well. To look for correlations between halogenase activity and electronic effects, the electronic activation of each sp^2-hybridized C-H bond on a given compound was determined via calculation of halenium affinity (HalA) values.[15] Similar methods have been used to rationalize site selectivities of electrophilic aromatic substitution reactions for small molecule reagents.[22,26-28]

Initially, HalA values were calculated for all sp^2-hybridized C-H bonds on all panel 1 substrates for which RebH site selectivity had been determined. These values were subsequently plotted against the observed chlorination at each site (Fig. 5.7, blue diamonds). As expected from site selectivity comparisons with NCS, several sites that were highly electronically activated (175-185 kJ/mol) were not chlorinated by RebH. Although no clear trend existed between conversion at a particular site and HalA, there appeared to be a HalA cutoff, below which trace to no conversion to product was observed for RebH (~160 kJ/mol, Fig. 5.7).

Next, we sought to understand the substrate specificity of RebH on panel 1 substrates. For this, we calculated the HalA values for all panel 1 substrates and compared the highest HalA value (max HalA) of each substrate to conversion of that particular substrate. Substrate binding by the enzyme appeared to be playing a key role in not only substrate site selectivity, but also specificity, as many substrates with high max HalA values were not halogenated. Notably, the anilide substrates in panel 1 (Fig. 5.7) had lower max HalA values than the anilines, indoles, pyrroles, and azoles (although still sufficiently activated, most were between 160-172 kJ/mol). This could account for the low levels of RebH activity observed for these substrates (Fig. 5.7). All results taken together, it appears that a given C-H bond must be sufficiently electronically activated for RebH halogenation (>160 kJ/mol) as well as able to access a position in the active site proximal to the catalytic Lys79.

Figure 5.7: Conversion vs. halenium affinity (HalA) for all sp^2 carbons selected substrates in panels 1 and 2.[a]



[a] HalA for sp^2-hybridized carbons on substrates for which site selectivity had been determined was plotted vs. conversion to chlorinated product at that particular site. HalA ranges for panel 1 substrates halogenated by RebH and panel 2 substrates halogenated by Rdc2 are highlighted in blue and green, respectively. The inset shows representative HalA values for common arenes (not substrates).

We then calculated HalA values for panel 2 substrates. Consistent with other reports, these phenol and anisole compounds had lower HalA values than the anilines, indoles, pyrroles, and azoles in panel 1 (Fig. 5.7). Conversion to chlorinated product with RebH (red squares), Rdc2 (green circles), and engineered RebH variants (orange triangles) was plotted against HalA for panel 2 substrates (Fig. 5.7). From this, it was clear that Rdc2 is able to halogenate these less nucleophilic phenol substrates to a much greater extent than RebH. It is striking that Rdc2 is able to accomplish this, but halogenates very few of the substantially more activated aniline substrates (Fig. 5.3). This interesting substrate specificity could be due to differences in substrate binding of phenol vs. aniline substrates within the Rdc2 active site or due to a slightly different mechanism for Rdc2 chlorination. For example, it is possible that residues within the Rdc2 active site are able to partially or fully deprotonate the alcohol substituent of a phenol substrate, rendering it more electronically activated than would be expected from the HalA values of the neutral compound. In addition to Rdc2, engineered RebH variants were able to chlorinate panel 2 substrates to a greater extent than RebH. These results suggest that both directed evolution and genome mining for novel, natural FDHs are viable methods for discovering FDHs with the ability to halogenate compounds with lower HalA values.

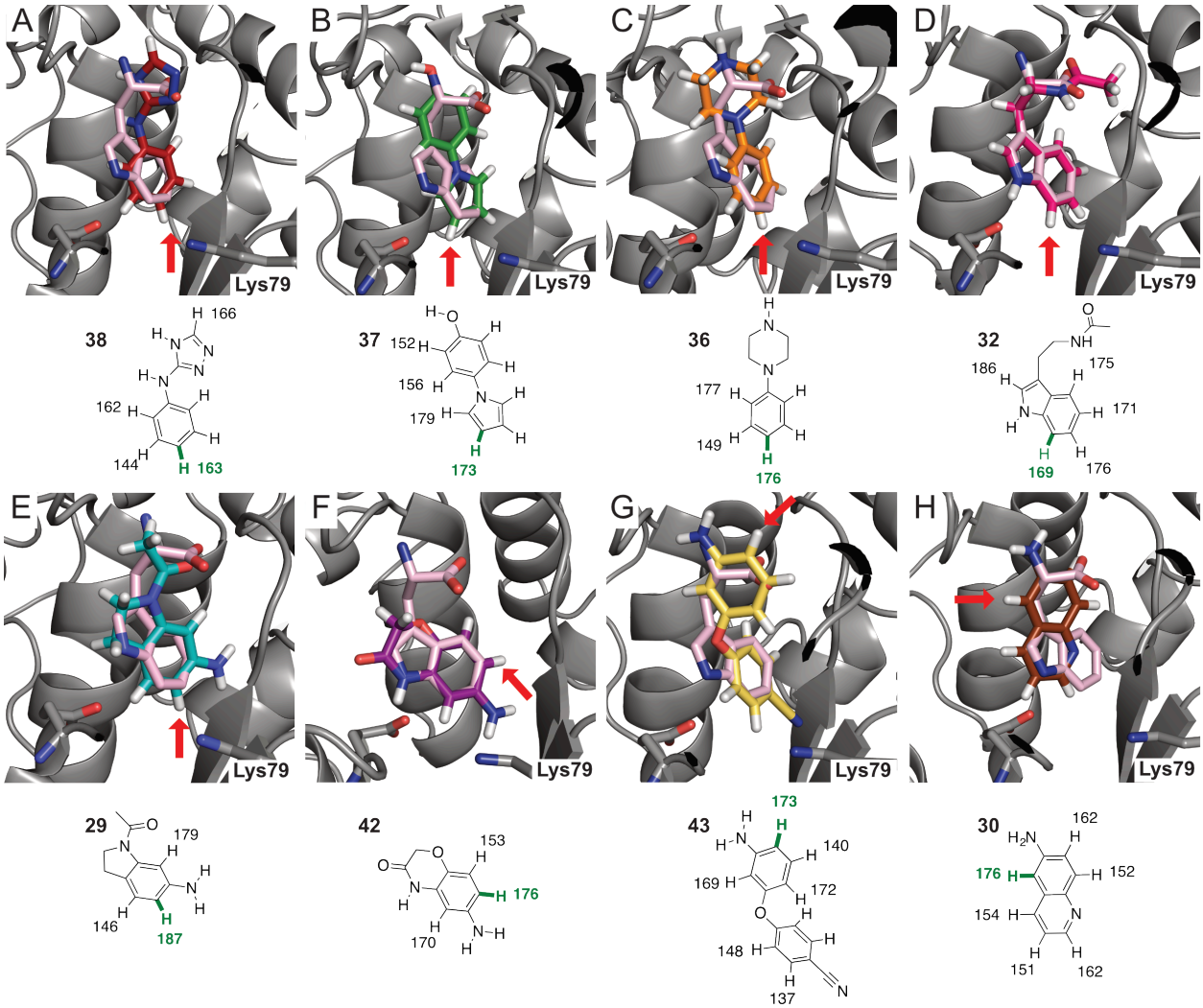
To further understand how substrate specificity and selectivity is conferred by FDHs, we evaluated substrate binding in RebH through docking simulations. Dr. Christine Hajdin used two docking methods, ICM Molsoft docking[29] and ROCS pharmacophore overlay[30], to evaluate binding for panel 1 substrates. The ICM Molsoft docking method is able to return multiple energetically favorable binding poses of a substrate within an enzyme target; however, it is blind to known substrate-enzyme interactions in the active site. ROCS pharmacophore overlay, on the

other hand, aligns a given substrate with L-tryptophan in the RebH active site based on molecular shape and charge.

Neither method was found to reliably predict substrate specificity - catalytically relevant poses were returned for several compounds that were not accepted as substrates by RebH. In addition, many poses from the ICM Molsoft docking method were not catalytically relevant (i.e. no sp^2-hybridized C-H bond proximal to Lys 79). On the other hand, although ROCS pharmacophore overlay did not provide insight into substrate specificity, it did provide a number of interesting results concerning the selectivity on substrates that were halogenated. For example, **32**, **36**, **37**, and **38** were all selectively halogenated at a site other than the most electronically activated (Fig. 5.8, A-D). When docked into the RebH active site using the pharmacophore model, the site for which halogenation was known to occur was positioned proximal to Lys79 (Fig. 5.8, A-D, site of halogenation indicated with red arrows). Polar functional groups distal to the site of halogenation appear to bind in the same pocket as the amino acid moiety of L-tryptophan. The unique site selectivities observed for these substrates could result from this anchoring effect.[31] In cases such as **29**, both HalA values and docking results support halogenation at the indicated site (i.e. the site of halogenation corresponded with the highest HalA as well as was predicted based on docking).

Although these cases provide insight into manner in which binding could be influencing selectivity, many substrates docked so that the C-H bond proximal to Lys79 was not consistent with experimentally determined site selectivity (Fig. 5.8, F-H). Thus, although this computationally inexpensive docking method can provide insight into FDH selectivity, more computationally rigorous methods such as molecular dynamics simulations[32] could improve predictive abilities for FDH substrate specificity and selectivity.

Figure 5.8: Docking poses for selected substrates observed for ROCS pharmacophore overlay.[a]



[a] Tryptophan in the RebH-tryptophan complex (PDB 2OA1) is shown in pink. Lys79 is labeled in each model, and the site of observed halogenation is indicated with a red arrow. Structures for each substrate are shown with HalA values for each site and the halogenated site in green.

5.3 Conclusions

In collaboration with Novartis Institutes for Biomedical Research through the Center for Selective C-H Functionalization, substrate activity profiles were constructed for eight FDHs (four wild-type and four engineered) for 93 substrates. These substrates contained a diversity of functional groups and sampled a range of steric and electronic parameters. Trp-FDHs were able

to halogenate the majority of tested anilines, indoles, azoles, and pyrroles; however, the fungal FDHs halogenated only a handful. On the other hand, the fungal FDHs chlorinated phenol substrates with higher efficiency than the Trp-FDHs. For many substrates, the site selectivity of chlorination with RebH or Rdc2 was determined. From these data, it is clear that FDHs are capable of overriding substrate electronics and site selectively halogenating a variety of substrates. Further insight into FDH regioselectivity was gained through quantitation of substrate electronics (HalA) and docking simulations.

5.4 Experimental

5.4.1 General experimental procedures

Materials:

All reagents were obtained from commercial suppliers and used without further purification. Deuterated solvents were obtained from Cambridge Isotope labs. Silicycle silica gel plates (250 mm, 60 F254) were used for analytical TLC, and preparative chromatography was performed using SiliCycle SiliaFlash silica gel (230-400 mesh). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, Iowa), and the sequences of the primers used in this study are reported below. All PCR was done using New England Biolabs (Ipswich, Massachusetts) Q5 high fidelity 2x mix catalog #M0494. Vector pET303-Chis was purchased from Thermo Fisher/Invitrogen catalog #K630203 (Cambridge, Massachusetts). NEBuilder HiFi DNA Assembly Master Mix catalog #E2621L was purchased from NEB. Dh5 alpha cells catalog #C2987on and BL21DE3 cells catalog #C2527 were purchased from NEB. LB Carbenicillin 100 plates were purchsaed from Teknova (Hollister, California). Plasmids pET-28a/RebF and pET-28a/RebH in BL-21 DE3 E. coli were provided by the Walsh group of

Harvard Medical School, Boston, MA.[33] The pLIC-MBP plasmid was provided by the Bottomley group of Monash University, Clayton, Australia.[34] The pGro7 plasmid encoding the groES and groEL chaperone set was purchased from Takara (Otsu, Shiga, Japan). The Thal gene was provided by Prof. Karl-Heinz van Pée.[12] BL21(DE3) *E. coli* cells were purchased from Invitrogen (Carlsbad, CA). Luria broth (LB) and Terrific broth (TB) media were purchased from Research Products International (Mt. Prospect, IL). All FDH genes were confirmed by sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility (900 E. 57th Street, Room 1230H, Chicago, IL 60637). Electroporation was carried out on a Bio-Rad MicroPulser using method Ec2. Ni-nitrilotriacetic acid (Ni-NTA) resin and Pierce® BCA Protein Assay Kits were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were followed when using both products. Amicon® 30 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C. The glucose dehydrogenase (GDH-105), FAD, and NAD were purchased from Codexis (Redwood City, CA). Biotage reverse phase columns (SNAP-KP-C18-HS) were purchased from Biotage.

General Procedures:

Reactions were analyzed using an Agilent Technologies 1200 BGM LCMS or an Agilent Technologies 6130 LC-MS. Reverse phase preparative chromatography was carried out using a Biotage Isolera One or a Waters Autopure 2545 BGM preparative LCMS. ^{1}H spectra were recorded at either 400 or 500 MHz and ^{13}C NMR spectra were recorded at 126 MHz on a Bruker DMX-500 at the University of Chicago Mass Spectrometry Service Center, or were recorded at either 400 or 600 MHz on a Bruker cryoprobe at the Global Discovery Chemistry at Novartis

Institutes for Biomedical Research. High-resolution mass spectra were obtained on an Agilent Technologies 6224 TOF LC-MS or a Water LCT Premier. FDH protein concentrations were determined by Pierce BCA Protein Assay Kit.

Expression and purification of MBP-RebF and FDHs: For all bioconversions in these studies, large-scale cultures (750 mL) of MBP-RebF and RebH were grown, expressed and purified as previously reported.[7] RebH variants, Thal, Rdc2, and GsfI were grown, expressed and purified according to the procedure described for RebH. This procedure produced more than 35 mg/L of purified enzyme for all FDHs. Concentrations of enzymes were measured using the Pierce BCA Protein Assay Kit. Following purification, MBP-RebF and FDHs were buffer exchanged into HEPES buffer (25 mM, pH 7.4), concentrated to 7 mg/mL, and lyophilized.

LCMS Methods:

LCMS method 1: Agilent Technologies 1200 BGM LCMS. Waters Sunfire C18 3 x 30mm column, 3.5 μM particle size; solvent A: H_2O/0.1% formic acid, solvent B: acetonitrile with 0.1% formic acid; 0-0.3 min, B = 95%; 0.3-3.7 min, B = 95-5%; 3.7-4.5 min, B = 5%; 4.5-4.75 min, B = 5-95%; 4.75-5 min, B = 95%. Flow rate 2.1 mL/min. Absorbance at 254 nm was measured.

LCMS method 2: Agilent Technologies 6130 LC-MS. Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN; 0-10 min, B = 15%; 10-17 min, B = 15-80%; 17-20 min, B = 80-100%; 20-24 min, B = 100%. Flow rate 1 mL/min. Absorbance at 280 nm and 254 nm was measured.

LCMS method 3a-d: Waters Autopure 2545 BGM preparative LCMS. Waters Sunfire C18 10 x 100mm column, 5 μM particle size; solvent A: H_2O/0.1% formic acid, solvent B: acetonitrile with 0.1% formic acid. Flow rate 15 mL/min. Absorbance at 254 nm was measured. Fraction collection was triggered by mass.

3a. 0-0.35 min, B = 2%; 0.35-4.3 min, B = 2-20%; 4.3-5.3 min, B = 20%; 5.3-5.5 min, B = 20-95%; 5.5-6.1 min, B = 95%; 6.1-6.2 min, B = 95-2%; 6.2-7.5 min, B = 2%.

3b. 0-0.35 min, B = 2%; 0.35-4.3 min, B = 2-5%; 4.3-5.3 min, B = 5%; 5.3-5.5 min, B = 5-95%; 5.5-6.1 min, B = 95%; 6.1-6.2 min, B = 95-2%; 6.2-7.5 min, B = 2%.

3c. 0-0.35 min, B = 12%; 0.35-4.3 min, B = 12-30%; 4.3-5.3 min, B = 30%; 5.3-5.5 min, B = 30-95%; 5.5-6.1 min, B = 95%; 6.1-6.2 min, B = 95-12%; 6.2-7.5 min, B = 12%.

3d. 0-0.35 min, B = 22%; 0.35-4.3 min, B = 22-40%; 4.3-5.3 min, B = 40%; 5.3-5.5 min, B = 40-95%; 5.5-6.1 min, B = 95%; 6.1-6.2 min, B = 95-22%; 6.2-7.5 min, B = 22%.

Preparative RebH bioconversions: Preparative bioconversions were conducted similarly for all substrates tested. A solution of 2, 4 or 10 mg substrate (1 equiv., 1.0 mM final concentration) in 5% v/v methanol was added to a reaction vessel (beaker, 25 or 50 mL). Solutions in HEPES buffer (25 mM, pH 7.4) of NAD (0.1 equiv., 100 μM final concentration), FAD (0.1 equiv., 100 μM final concentration), NaCl (10 equiv., 10 mM final concentration), and a glucose dehydrogenase (9 U/mL final concentration GDH) were added. Solutions in HEPES/glycerol buffer (25 mM, pH 7.5, 10% glycerol v/v) of RebH (0.025-0.05 equiv., 25-50 μM final concentration) and MBP-RebF (0.0025 equiv., 2.5 μM final concentration) were added. The bioconversion was diluted with HEPES buffer (25 mM, pH 7.4) to the appropriate reaction

volume and an aqueous solution of glucose (20 equiv., 20 mM final concentration) was added to initiate the cofactor regeneration cycle. The beaker was agitated on a stir plate at 25 °C until quenched.

Reactions were monitored by LCMS method 1 as described in the General Procedures and were quenched with one volume of dichloromethane upon completion. The reaction mixture was transferred to a 50 mL falcon tube and centrifuged to separate the aqueous and organic layers. The organic layer was collected and the aqueous layer was extracted twice more with dichloromethane. Organic layers were combined and concentrated. The crude material was then resuspended in DMSO (250 μL) and passed through a 0.45 μm filter. The crude material was purified by reverse phase chromatography on by preparative LCMS, methods 3a-d. Purified fractions were concentrated overnight using a Genevac.

Preparative NCS reactions: Preparative reactions with the chlorinating agent N-chlorosuccinimde (NCS) were conducted for several substrates in Panel 2. Substrate (50 mg, 1 equiv.) and NCS (1 equiv.) were added as solids to a dry Teflon-sealed Schlenk flask. Acetonitrile was added as solvent (5 mL) and the reaction mixture was stirred at 80 °C. Reactions were monitored by either TLC or LCMS method 2. The crude material was purified by normal or reverse phase chromatography.

Isolated mono-chlorinated products from these reactions were then used as standards to determine the site-selectivity of Rdc2 on Panel 2 substrates. Notable exceptions were substrates **25** and **26**. Following this reaction procedure, mono-chlorinated isomers could not be obtained for either of these substrates. For this reason, bioconversions were conducted with Rdc2 on these two substrates. While the product of **26** could not be isolated from this reaction either, a single

mono-chlorinated isomer was obtained for **25**. Since the selectivity of 26 could not be determined, this substrate was not included in Figure 5.7.

5.4.2 Specific experimental procedures

Cloning of FDHs: The open reading frames for Rdc2 and Gsf1 were ordered as gBlocks from Integrated DNA Technologies. The ORF's were codon optimized by IDT for *E. coli* expression. The gblocks served as PCR templates for Gibson assembly fragment generation. The PCR program used in these studies is as follows: 98 °C for 2 min; 98 °C 20 seconds; 65 °C 20 seconds; 72 °C for an appropriate amount of time depending on fragment size; 25 cycles. Primer pairs are as follows:

Rdc2:

5'-TTTTGTTTAACTTTAAGAAGGAGGTCTAGAATGTCAGTGCCCAAATCTTGCACAATCC-3'
5'-ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGAACCTTGTTCAGGCCCAGTTTGCCATG-3'

Gsf1:

5'-TTTTGTTTAACTTTAAGAAGGAGGTCTAGAATGGCGATTCCCCAGTCTTGCACGGTC-3'
5'-ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGCATCTGTAAGTCACCTGCGTGATCGCGG-3'

Using vector pET303-Chis as template, the vector was amplified by PCR using the primer pair:

5'-TCTAGACCTCCTTCTTAAAGTTAAACAAAATTATTAC-3'
5'-CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGC-3'

The vector PCR was digested with DpnI to eliminate carry over circular vector plasmid contamination. The gene fragments and vector DNA were Gibson assembled using NEBuilder HiFi DNA Assembly Master Mix at 50 °C for 1 hour. Products were transformed into NEB Dh5 alpha cells and plated on LB Carbenicillin 100 plates. Colonies were picked and mini prepped using a Zymo classic miniprep kit. Clones were sequence verified by sequencing at Genewiz (Cambridge, Massachusetts). Verified clones were transformed into electrocompetent cells

(BL21(DE3) *E. coli*) containing the pGro7 plasmid, as previously described for RebH.[7] All Trp-FDHs were cloned into a pET28a vector and transformed into electrocompetent cells (BL21(DE3) *E. coli*) containing the pGro7 plasmid, as previously described for RebH.[7]

Phylogenetic Tree (Figure 5.2): The amino acid sequences of all 8 FDHs were obtained through sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility. These sequences were exported into Geneious and a tree was built using a global alignment and the cost matrix Blosum62. The genetic distance model was Jukes-Cantor and the tree build method was neighbor-joining. The sequence identities of RebH vs. Thal, and RebH vs. Rdc2 was calculated using an alignment of the amino acid sequences in protein-protein Blast (Blastp).

Bioconversions of Panel 1 and Panel 2: Bioconversions of substrates in Panel 1 and 2 were conducted in 96-well microtiter plates. Substrates were added to each well (3.75 μL of a 10 mM stock in DMSO, 1 equiv., 0.5 mM final concentration). To each of these wells, a solution containing FAD (0.2 equiv., 100 μM final concentration), NAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), and glucose (40 equiv., 20 mM final concentration) was added. To initiate reaction, each FDH (0.05 equiv., 25 μM final concentration), MBP-RebF (0.005 equiv., 2.5 μM final concentration), and glucose dehydrogenase (9 U/mL final concentration) were added as solutions (25 mM HEPES, pH 7.4). The final reaction volume of all bioconversions was 75 μL. Reaction plates were sealed using a plate sealer and were either agitated overnight at 650 rpm on a plate shaker (RebH, Thal, RebH variants) or left unshaken on the benchtop (Rdc2, GsfI) at room temperature. These were

quenched with 75 μL of methanol, centrifuged to remove precipitated protein, filtered, and analyzed by LCMS Method 1, as described in the General Procedures.

Analytical reactions with NCS and FDHs (Table 5.1): Analytical reactions with the chlorinating agent N-chlorosuccinimde (NCS) were conducted for several substrates in Panel 1 and 2. Substrate (5 mg, 1 equiv.) and NCS (1 equiv.) were added as solids to an eppendorf tube. Acetonitrile was added as solvent (500 μL) and the reaction mixture was agitated at 650 rpm at 80°C. Bioconversions with these same substrates were conducted according to the procedure described in the section entitled *Bioconversions of Panel 1 and Panel 2.* Reactions with both NCS and FDHs were analyzed by LCMS Method 2 (substrates **23**, **38**, **40**, **41**, **43**, **44**, **45**, **46**) or GCMS (substrate **27**). Reported selectivity differences for all substrates in Table 1, with the exception of **37,** were determined by comparison of the areas of mono- or di-chlorinated product peaks at different retention times, either by LCMS or GCMS. The selectivity of **37** was determined through preparative reaction and isolation of the mono-chlorinated products from NCS and RebH.

Halenium affinity calculations: Halenium affinity (HalA) refers to the energy associated with the reaction of a substrate with a halenium ion (X^+). In this study, X = Cl, and the reaction involves addition of Cl^+ at all sp^2-hybridized carbons on all substrates in panels 1 and 2.

Geometry optimizations and frequency calculations were conducted using Gaussian (DFT, B3LYP, LANL2DZ). Halenium affinity values for different positions were calculated using equations (1) and (2) according a literature procedure.[15] Electronic energies ($E_{(elec)}$), zero point energies (*ZPE*), and frequencies (ν) were obtained directly from Gaussian output files

(Sum of electronic and zero-point Energies, Zero-point correction, and Vibrational temperatures, respectively).

$$HalA = -\Delta E_{(elec)} - \Delta ZPE - \Delta E'_{(vib)} + \frac{5}{2}RT \quad (1)$$

$$E'_{(vib)}(T) = \sum_{i=1}^{3n-6} \frac{Nhv_i}{e^{Nhv_i} - 1} \quad (2)$$

Docking studies: To identify the potential poses of the compounds in the halogenase enzyme, compounds were docked via two approaches (1) ICM Molsoft docking and (2) ROCS pharmacophoric overlay. Due to the availability of structural information available on RebH (PDB ID: 2OA1, 2O9Z, 2OAM, 2E4G, 2OAL), this enzyme was chosen as the sample system. For the first method, molecules were docked using MolSoft ICM version 3.8.4. A multi-conformer approach with a thoroughness score of 10 was utilized and the top 10 poses were scored. Scored poses with energetically favorable scores were considered as potential binding modes and were utilized in identifying sights of potential halogenation.

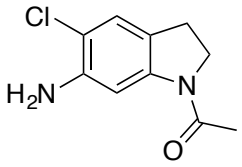
For the second method, the crystalized tryptophan was used as the template for pharmacophoric searching (ROCS OpenEye version 3.2). The molecule with the best Tanimoto similarity score was considered as a potential binding mode and was used for consideration in identifying sights of potential halogenation.

5.4.3 Detailed isolation and characterization

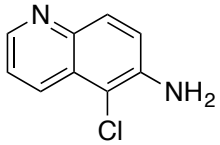
Detailed isolation and characterizations - bioconversions:

For each compound isolated, the ^{1}HNMR and HRMS of the chlorinated compound is shown. In some cases, impurity peaks were observed in the ^{1}HNMR of the chlorinated products,

which is due to the small scale of these reactions and the use of formic acid during prep LC/MS purification. For this reason, the ^{1}HNMR of the starting material for each is included as well as a reference and yields are not reported (only % conversions based on LCMS UV traces).

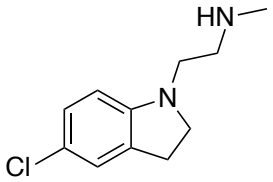


1-(6-Amino-5-chloro-2,3-dihydro-1H-indol-1-yl)ethanone (29a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude reaction mixture was concentrated onto Celite, packed into a Biotage samplet, and loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography to afford compound **29a** as the TFA salt. ^{1}H NMR (500 MHz; DMSO): δ 7.68 (s, 1H), 7.04 (s, 1H), 4.04 (t, J = 8.4, 2H), 2.98 (t, J = 8.3, 2H), 2.13 (s, 3H). HRMS (ESI-MS) calcd for $C_{10}H_{11}ClN_2O$ $[M + H]^+$: 211.0633 and 213.0604, found: 211.0604 and 213.0577.

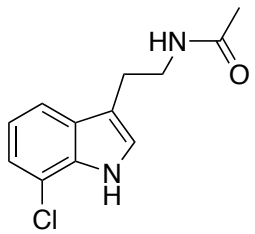


6-Amino-5-chloroquinoline (30a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to

quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3a) to afford compound **30a**. ^{1}H NMR (400 MHz; DMSO): δ 8.59 (dd, J = 4.2, 1.6, 1H), 8.22 (ddd, J = 8.5, 1.5, 0.8, 1H), 7.74 (d, J = 9.1, 1H), 7.48 (dd, J = 8.5, 4.2, 1H), 7.37 (d, J = 9.1, 1H), 5.99 (s, 2H). HRMS (ESI-MS) calcd for $C_9H_7ClN_2$ $[M + H]^+$: 179.0371 and 181.0341, found: 179.0343 and 181.0319.

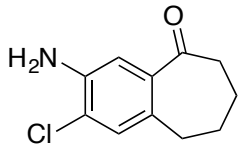


5-chloro-2,3-dihydro-N-methyl-1H-Indole-1-ethanamine (31a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3b) to afford compound **31a**. ^{1}H NMR (400 MHz; DMSO): δ 7.08-7.03 (m, 2H), 6.56 (dd, J = 8.2, 3.4, 1H), 3.36 (dd, J = 8.0, 4.9, 2H), 3.28 (s, 2H), 3.05 (s, 2H), 2.91 (t, J = 3.8, 2H), 2.54 (s, 3H). HRMS (ESI-MS) calcd for $C_{11}H_{15}ClN_2$ $[M + H]^+$: 211.0997 and 213.0967, found: 211.0974 and 213.0959.



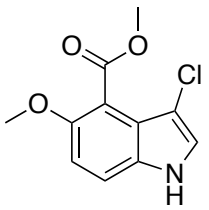
***N*-[7-chloro-2-(1H-indol-3-yl)ethyl]-acetamide (32a):** The bioconversion was conducted in a

beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3d) to afford compound **32a**. ^{1}H NMR (400 MHz; DMSO): δ 11.21 (s, 1H), 7.95 (t, J = 5.5, 1H), 7.51 (d, J = 7.9, 1H), 7.23 (d, J = 2.4, 1H), 7.15 (d, J = 7.5, 1H), 7.00 (t, J = 7.7, 1H), 3.30 (t, J = 6.6, 2H), 2.80 (t, J = 7.3, 2H), 1.79 (s, 3H). HRMS (ESI-MS) calcd for $C_{12}H_{13}ClN_2O$ $[M + H]^+$: 237.0789 and 239.0670, found: 237.0799 and 239.0787.

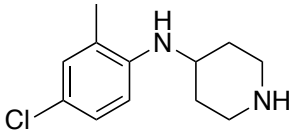


3-amino-2-chloro-6,7,8,9-tetrahydro-*H*-benzocyclohepten-5-one (33a): The bioconversion was conducted in a beaker according to the general procedure, using 0.05 equiv. of RebH (50 μM final concentration) at 25 °C, with 4 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3d) to afford compound **33a**. ^{1}H NMR (400 MHz; DMSO): δ 7.16 (s, 1H), 7.07 (s, 1H), 5.42 (s, 2H), 2.76 (t, J = 6.1, 2H), 2.63-2.60 (m, 2H), 1.68 (td, J = 12.6, 6.4, 4H). HRMS (ESI-MS) calcd for $C_{11}H_{12}ClNO$ $[M + H]^+$: 210.0680 and 212.0651, found: 210.0737 and 212.0709. Note: Partial reduction of the ketone to an alcohol was observed. This side product, 3-amino-2-chloro-6,7,8,9-tetrahydro-5H-benzocyclohepten-5-ol was isolated and characterized as well. ^{1}H NMR (400 MHz; DMSO): δ 6.94 (s, 1H), 6.87 (s, 1H),

5.15 (d, J = 4.0, 1H), 5.06 (s, 2H), 4.57 (dd, J = 9.5, 3.9, 1H), 2.62 (dd, J = 13.8, 7.2, 1H), 1.86 (t, J = 9.3, 2H), 1.65 (dd, J = 20.2, 9.7, 2H), 1.43 (q, J = 10.2, 1H), 1.17 (dt, J = 22.1, 10.9, 1H). HRMS (ESI-MS) calcd for $C_{11}H_{14}ClNO$ $[M + H]^+$: 212.0837 and 214.0808, found: 212.0827 and 214.0806.

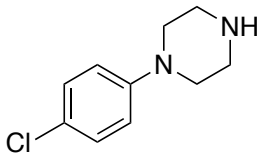


Methyl 3-chloro-5-methoxy-1H-indole-4-carboxylate (34a): The bioconversion was conducted in a beaker according to the general procedure, using 0.05 equiv. of RebH (50 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3d) to afford compound **34a**. ^{1}H NMR (400 MHz; DMSO): δ 11.51 (s, 1H), 7.57 (d, J = 2.7, 1H), 7.47 (d, J = 9.0, 1H), 7.04 (d, J = 9.0, 1H), 3.83 (s, 3H), 3.79 (s, 3H). HRMS (ESI-MS) calcd for $C_{11}H_{10}ClNO_3$ $[M + H]^+$: 240.0422 and 242.0393, found: 240.0442 and 242.0224.

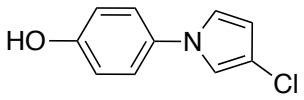


N-(4-chloro-2-methylphenyl)-4-piperidinamine (35a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction

mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3c) to afford compound **35a**. ^{1}H NMR (400 MHz; DMSO): δ 7.02-7.00 (m, 2H), 6.62 (d, J = 9.3, 1H), 3.51 (s, 1H), 3.26 (d, J = 12.3, 2H), 2.95-2.86 (m, 2H), 2.08 (s, 3H), 1.99 (d, J = 12.1, 2H), 1.60 (q, J = 11.0, 2H). HRMS (ESI-MS) calcd for $C_{12}H_{17}ClN_2$ $[M + H]^+$: 225.1153 and 227.1124, found: 225.1160 and 227.1131.

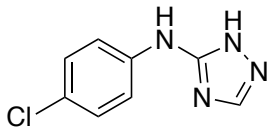


para-chlorophenylpiperazine (36a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude reaction mixture was concentrated onto Celite, packed into a Biotage samplet, and loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography to afford compound **36a** as the TFA salt. ^{1}H NMR (500 MHz; DMSO): δ 8.91 (s, 2H), 7.31-7.28 (m, 2H), 7.03-7.00 (m, 2H), 3.34 (m, J = 5.1, 4H), 3.23 (s, 4H). HRMS (ESI-MS) calcd for $C_{10}H_{13}ClN_2$ $[M + H]^+$: 197.0840 and 199.0811, found: 197.1111 and 199.1042.

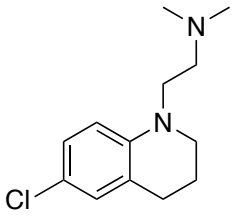


4-(3-chloro-1H-pyrrol-1-yl)phenol (37a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at

25 °C, with 1 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3c) to afford compound **37a**. ^{1}H NMR (500 MHz; DMSO): δ 7.36-7.33 (m, 3H), 7.21 (dd, J = 3.0, 2.6, 1H), 6.84-6.81 (m, 2H), 6.25 (dd, J = 3.0, 1.8, 1H). HRMS (ESI-MS) calcd for $C_{10}H_8ClNO$ $[M - H]^-$: 192.0211 and 194.0182, found: 192.0247 and 194.0278.

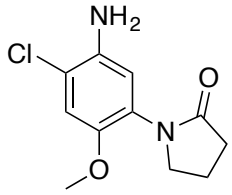


4-chloro-N-phenyl-1H-1,2,4-Triazol-5-amine (38a): The bioconversion was conducted in a beaker according to the general procedure, using 0.05 equiv. of RebH (50 μM final concentration) at 25 °C, with 4 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3d) to afford compound **38a**. ^{1}H NMR (400 MHz; DMSO): δ 13.41 (s, 1H), 9.42 (s, 1H), 8.21 (s, 1H), 7.59-7.55 (m, 2H), 7.27-7.23 (m, 2H). HRMS (ESI-MS) calcd for $C_8H_7ClN_4$ $[M + H]^+$: 195.0432 and 197.0403, found: 195.0431 and 197.0409.



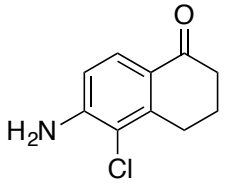
6-chloro-3,4-dihydro-N,N-dimethyl-1(2H)-quinolineethanamine (39a): The bioconversion was conducted in a beaker according to the general procedure, using 0.05 equiv. of RebH (50 μM

final concentration) at 25 °C, with 10 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3b) to afford compound **39a**. ^{1}H NMR (400 MHz; DMSO): δ 6.97 (dd, J = 8.8, 2.7, 1H), 6.91 (d, J = 2.6, 1H), 6.56 (d, J = 8.8, 1H), 3.38 (t, J = 7.3, 2H), 3.26 (t, J = 5.6, 2H), 2.65 (t, J = 6.3, 2H), 2.55 (t, J = 7.3, overlapping solvent peak), 2.32 (s, 6H), 1.84-1.78 (m, 2H). HRMS (ESI-MS) calcd for $C_{13}H_{19}ClN_2$ $[M + H]^+$: 239.1310 and 241.1281, found: 239.1317 and 241.1299. Note: A peak with a mass corresponding to 2 amu less than the mass of the starting material was consistently observed for this reaction. This "$M_{SM} - 2$" side product could not be isolated.

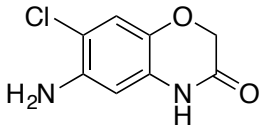


1-(5-amino-4-chloro-2-methoxyphenyl)-2-Pyrrolidinone (40a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude reaction mixture was concentrated onto Celite, packed into a Biotage samplet, and loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography to afford compound **40a** as the TFA salt. ^{1}H NMR (500 MHz; DMSO): δ 7.05 (s, 1H), 6.80 (s,

1H), 3.71 (s, 3H), 3.62 (t, J = 7.0, 2H), 2.36 (t, J = 8.1, 2H), 2.06 (quintet, J = 7.6, 2H). HRMS (ESI-MS) calcd for $C_{11}H_{13}ClN_2O_2$ $[M + H]^+$: 241.0738 and 243.0709, found: 241.0739 and 243.0727.

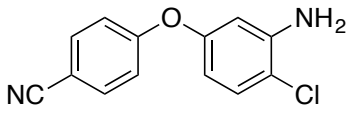


6-amino-5-chloro-3,4-dihydro-1(2*H*)-naphthalenone (41a): The bioconversion was conducted in a beaker according to the general procedure, using 0.05 equiv. of RebH (50 μM final concentration) at 25 °C, with 4 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3d) to afford compound **41a**. ^{1}H NMR (400 MHz; DMSO): δ 7.62 (d, J = 8.6, 1H), 6.71 (d, J = 8.6, 1H), 6.32 (s, 2H), 2.87 (t, J = 6.1, 2H), 2.43 (t, J = 6.5, 2H), 1.99 (dt, J = 12.7, 6.3, 2H). HRMS (ESI-MS) calcd for $C_{10}H_{10}ClNO$ $[M + H]^+$: 196.0524 and 198.0495, found: 196.0710 and 198.1047.

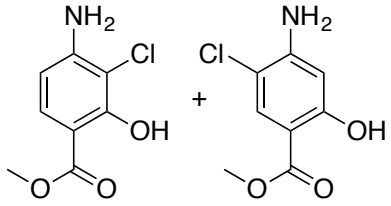


6-Amino-7-chloro-2H-1,4-benzoxazin-3(4H)-one (42a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane

according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3c) to afford compound **42a**. ^{1}H NMR (400 MHz; DMSO): δ 10.64 (s, 1H), 6.84 (s, 1H), 6.38 (s, 1H), 5.09 (s, 2H), 4.42 (s, 2H). HRMS (ESI-MS) calcd for $C_8H_7ClN_2O_2$ $[M + H]^+$: 199.0269 and 201.0240, found: 199.0321 and 200.9913.

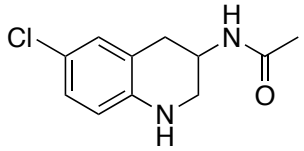


4-(3-amino-4-chloro-phenoxy)-benzonitrile (43a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3d) to afford compound **43a**. ^{1}H NMR (400 MHz; DMSO): δ 7.85-7.82 (m, 2H), 7.25 (d, J = 8.6, 1H), 7.13-7.10 (m, 2H), 6.50 (d, J = 2.8, 1H), 6.29 (dd, J = 8.6, 2.8, 1H), 5.61 (s, 2H). HRMS (ESI-MS) calcd for $C_{13}H_9ClN_2O$ $[M + H]^+$: 245.0476 and 247.0447, found: 245.0401 and 247.4741.

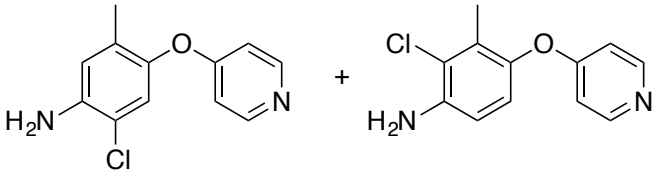


Methyl 4-amino-3-chlorosalicylate and **Methyl 4-amino-5-chlorosalicylate (44a** and **44b):** The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of

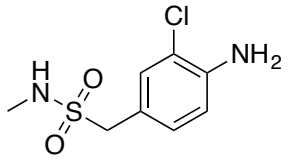
dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3d) to afford compound **44a** and **44b**. **44a:** ^{1}H NMR (400 MHz; DMSO): δ 11.49 (s, 1H), 7.46 (d, J = 8.9, 1H), 6.46 (s, 1H), 6.34 (d, J = 8.9, 1H), 3.83 (s, 2H). HRMS (ESI-MS) calcd for $C_8H_8ClNO_3$ $[M + H]^+$:, found: 202.0550 and 204.0287 **44b:** ^{1}H NMR (400 MHz; DMSO): δ 10.57 (s, 1H), 7.56 (s, 1H), 6.36 (s, 2H), 6.24 (s, 1H), 3.81 (s, 3H). HRMS (ESI-MS). calcd for $C_8H_8ClNO_3$ $[M + H]^+$:, found: 202.0483 and 204.0473.



***N*-(6-chloro-1,2,3,4-tetrahydro-3-quinolinyl)-acetamide (45a):** The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3c) to afford compound **45a**. **45a**: ^{1}H NMR (500 MHz; DMSO): δ 7.82 (d, J = 7.1, 1H), 6.91-6.88 (m, 2H), 6.47 (d, J = 8.3, 1H), 5.92 (s, 1H), 3.91-3.99 (m, 1H), 3.23 (d, J = 9.8, 1H), 2.92 (t, J = 9.0, 1H), 2.84 (dd, J = 16.0, 4.7, 1H), 2.59 (dd, J = 16.2, 8.5, 1H), 1.82 (s, 3H). HRMS (ESI-MS) calcd for $C_{11}H_{13}ClN_2O$ $[M + H]^+$: 225.0789 and 227.0760, found: 225.0764 and 227.0745.

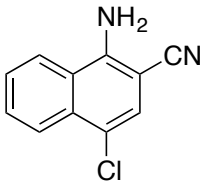


6-chloro-3-methyl-4-(4-pyridinyloxy)-benzenamine and **2-chloro-3-methyl-4-(4-pyridinyloxy)-benzenamine (46a** and **46b):** The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3b) to afford compounds **46a** and **46b**. **46a**: ^{1}H NMR (500 MHz; DMSO): δ 8.72 (d, J = 7.1, 2H), 7.34 (d, J = 7.1, 2H), 7.20 (s, 1H), 6.79 (s, 1H), 1.97 (s, 3H). HRMS (ESI-MS) calcd for $C_{12}H_{11}ClN_2O$ $[M + H]^+$: 235.0633 and 237.0604, found: 235.0643 and 237.0607. **46b**: ^{1}H NMR (500 MHz; DMSO): δ 8.70 (d, J = 7.1, 2H), 7.31 (d, J = 7.1, 2H), 6.96 (d, J = 8.8, 1H), 6.80 (d, J = 8.8, 1H), 2.07 (s, 3H). HRMS (ESI-MS) calcd for $C_{12}H_{11}ClN_2O$ $[M + H]^+$: 235.0633 and 237.0604, found: 235.0622 and 237.0600.

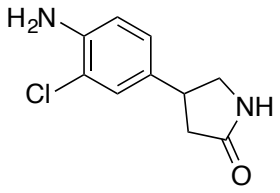


4-Amino-3-chlorophenyl-N-methyl methanesulfonamide (48a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into

dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3c) to afford compound **48a**. ^{1}H NMR (400 MHz; DMSO): δ 7.17 (d, J = 1.9, 1H), 6.99 (dd, J = 8.3, 2.0, 1H), 6.83 (q, J = 4.9, 1H), 6.75 (d, J = 8.3, 1H), 5.44 (s, 2H), 4.13 (s, 2H), 2.52 (m, overlapping solvent peak). HRMS (ESI-MS) calcd for $C_8H_{11}ClN_2O_2S$ $[M + H]^+$: 235.0302 and 237.0273, found: 235.0638 and 237.0078.

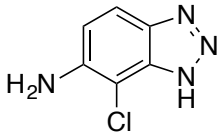


1-amino-4-chloro-2-naphthalenecarbonitrile (52a): The bioconversion was conducted in a beaker according to the general procedure, using 0.05 equiv. of RebH (50 μM final concentration) at 25 °C, with 4 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3d) to afford compound **52a**. ^{1}H NMR (400 MHz; DMSO): δ 8.44 (d, J = 8.4, 1H), 8.08 (dd, J = 8.4, 0.9, 1H), 7.79 (ddd, J = 8.3, 7.0, 1.2, 1H), 7.64 (ddd, J = 8.4, 7.0, 1.3, 1H), 7.58 (s, 1H), 7.13 (s, 2H). HRMS (ESI-MS) calcd for $C_{11}H_{10}ClNO_3$ $[M + H]^+$: 240.0422 and 242.0393, found: 240.0217 and 241.9712.



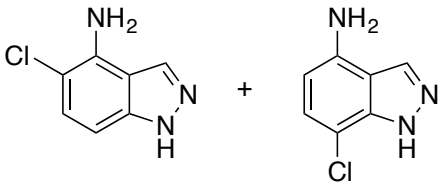
4-(4-amino-3-chlorophenyl)-2-pyrrolidinone (62a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final

concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude reaction mixture was concentrated onto Celite, packed into a Biotage samplet, and loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography. Product containing fractions were pooled, basified with K_2CO_3, and extracted into dichloromethane to afford compound **62a**. ^{1}HNMR (500 MHz; DMSO): δ 7.65 (s, 1H), 7.13 (d, J = 1.9, 1H), 6.97 (dd, J = 8.4, 1.9, 1H), 6.75 (d, J = 8.3, 1H), 3.55-3.51 (m, overlapping with solvent), 3.48-3.41 (m, overlapping with solvent), 3.10 (t, J = 8.5, 1H), 2.42 (dd, J = 16.2, 8.6, 1H), 2.22 (dd, J = 16.3, 9.4, 1H), 1.24 (s, 2H). HRMS (ESI-MS) calcd for $C_{10}H_{11}ClN_2O$ $[M + H]^+$: 211.0633 and 213.0604, found: 211.0712 and 213.0650.

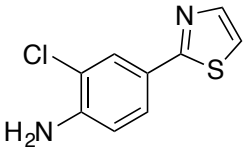


7-chloro-1H-benzotriazol-6-amine (64a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3a) to afford compound **64a**. Note: Upon concentration of LCMS-purified product with formic acid, the formamide partially formed (can be observed in HNMR and HRMS). ^{1}H NMR (400 MHz; DMSO): δ 7.62 (d, J = 8.8, 1H), 6.92 (d, J = 8.8, 1H), 5.74 (s, 2H). HRMS (ESI-MS)

calcd for $C_6H_5ClN_4$ $[M + H]^+$: 169.0276 and 171.0246, found: 169.1359 and 170.9829.

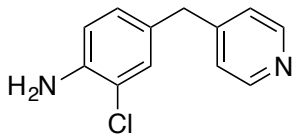


5-chloro-1H-indazol-4-amine and **7-chloro-1H-indazol-4-amine (67a** and **67b):** The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3b) to afford compounds **67a** and **67b**. **67a:** ^{1}H NMR (400 MHz; DMSO): δ 13.07 (s, 1H), 8.18 (s, 1H), 7.01 (d, J = 8.0, 1H), 6.11 (d, J = 8.0, 1H), 5.93 (s, 2H). HRMS (ESI-MS) calcd for $C_7H_6ClN_3$ $[M + H]^+$: 168.0323 and 170.0294, found: 168.2418 and 170.1853. **67b:** ^{1}H NMR (400 MHz; DMSO): δ 12.85 (s, 1H), 8.19 (s, 1H), 7.08 (d, J = 8.7, 1H), 6.66 (d, J = 8.6, 1H), 5.98 (s, 2H). HRMS (ESI-MS) calcd for $C_7H_6ClN_3$ $[M + H]^+$: 168.0323 and 170.0294, found: 168.0796 and 170.0637.

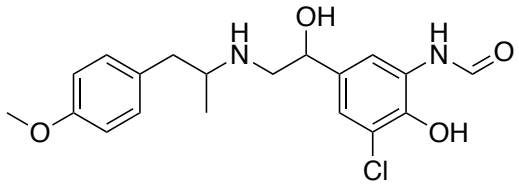


2-chloro-4-(2-thiazolyl)-benzenamine (77a): The bioconversion was conducted in a beaker according to the general procedure, using 0.05 equiv. of RebH (50 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the

general procedure. The crude material was purified by reverse phase chromatography (LCMS method 3d) to afford compound **77a**. ^{1}H NMR (400 MHz; DMSO): δ 7.78 (d, J = 3.3, 1H), 7.76 (d, J = 2.1, 1H), 7.61-7.59 (m, 2H), 6.85 (d, J = 8.5, 1H), 5.93 (s, 2H). HRMS (ESI-MS) calcd for $C_9H_7ClN_2S$ $[M + H]^+$: 211.0091 and 213.0062, found: 211.0114 and 213.0093.



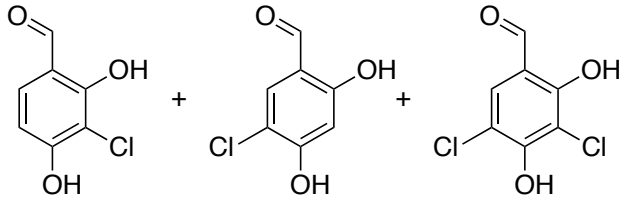
2-chloro-4-(4-pyridinylmethyl)-benzenamine (82a): The bioconversion was conducted in a beaker according to the general procedure, using 0.025 equiv. of RebH (25 μM final concentration) at 25 °C, with 2 mg of substrate. After maximum conversion to monohalogenated product was observed by LCMS, one volume of dichloromethane was added to the reaction mixture and vortexed to quench. The reaction mixture was extracted into dichloromethane according to the general procedure. The crude reaction mixture was concentrated onto Celite, packed into a Biotage samplet, and loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography. Product containing fractions were pooled, basified with K_2CO_3, and extracted into dichloromethane to afford compound **82a**. ^{1}H NMR (500 MHz; DMSO): δ 8.44 (d, J = 5.6, 2H), 7.21 (d, J = 5.9, 2H), 7.09 (d, J = 1.9, 1H), 6.91 (dd, J = 8.2, 1.9, 1H), 6.73 (d, J = 8.2, 1H), 5.23 (s, 2H), 3.79 (s, 2H). HRMS (ESI-MS) calcd for $C_{12}H_{11}ClN_2$ $[M + H]^+$: 219.0684 and 221.0655, found: 219.0647 and 221.0627.



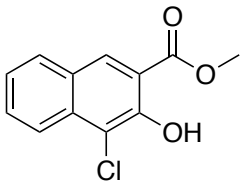
***N*-[3-chloro-2-hydroxy-5-[1-hydroxy-2-[[2-(4-methoxyphenyl)-1-methylethyl]amino]ethyl]phenyl]-formamide (25a):** A solution of 10 mg of **25** (1 equiv., 0.5 mM final concentration) in 5% v/v methanol was added to a crystallization dish (90 x 50 mm). Solutions in HEPES buffer (25 mM, pH 7.4) of NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl (20 equiv., 10 mM final concentration), and a glucose dehydrogenase (9 U/mL final concentration GDH) were added. Solutions in HEPES/glycerol buffer (25 mM, pH 7.5, 10% glycerol v/v) of Rdc2 (0.05 equiv., 25 μM final concentration) and MBP-RebF (0.0025 equiv., 2.5 μM final concentration) were added. The bioconversion was diluted with HEPES buffer (25 mM, pH 7.4) to the appropriate reaction volume and an aqueous solution of glucose (40 equiv., 20 mM final concentration) was added to initiate the cofactor regeneration cycle. The dish was covered with aluminum foil and left on the benchtop. The reaction was monitored by LCMS Method 1 and was quenched with aqueous HCl (5.0 M, until pH ~1-2) upon completion (55% conv.). The quenched reaction mixture was transferred to 2 x 50 mL falcon tubes and was centrifuged at 15,000 rpm for 20 minutes. The aqueous supernatant was filtered through a 45 μm filter and was concentrated by rotovap. The product was extracted from reaction salts through a solid-liquid extraction with MeOH (4 extractions with ~5 mL per extraction). The MeOH was filtered through filter paper and was concentrated onto Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography to afford **25a.** ^{1}H NMR (500 MHz; DMSO): δ 9.99 (s, 1H), 8.33

(d, J = 1.9, 1H), 8.05 (d, J = 2.0, 1H), 7.19 (d, J = 2.0, 1H), 7.15 (d, J = 8.7, 2H), 6.91 (d, J = 8.7, 2H), 4.85 (dd, J = 10.2, 2.5, 1H), 3.74 (s, 3H), 3.42 (m, 1H), 3.17-3.02 (m, 3H), 2.59 (dd, J = 13.1, 10.7, 1H), 1.09 (d, J = 6.5, 3H). ^{13}C NMR (126 MHz; DMSO): δ 161.06, 158.63, 143.02, 134.48, 130.71, 129.08, 128.87, 122.56, 121.59, 118.05, 114.53, 68.43, 55.52, 55.50, 51.04, 38.25, 15.21. HRMS (ESI-MS) calcd for $C_{19}H_{23}ClN_2O_4$ $[M + H]^+$: 379.1419 and 381.1391, found: 379.1449 and 381.1443.

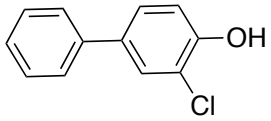
Detailed isolation and characterizations – NCS reactions:



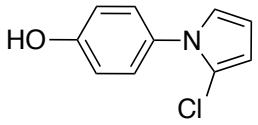
3-chloro-2,4-dihydroxy-benzaldehyde, **5-chloro-2,4-dihydroxy-benzaldehyde** and **3,5-dichloro-2,4-dihydroxy-benzaldehyde (23a, 23b,** and **23c):** The reaction was conducted according to the general procedure. After maximum conversion to monohalogenated product was observed by LCMS Method 2, the crude reaction mixture was dry-loaded into Celite. The Celite was packed into a Biotage samplet, which was then loaded into a reverse phase column (Biotage SNAP-KP-C18-HS). The crude material was purified by reverse phase chromatography to afford a purified mixture of **23a** and **23b** in 55% yield (34 mg, 0.198 mmol) and **23c** in 7% yield (5.5 mg, 0.027 mmol). **23a** and **23b**: ^{1}H NMR (500 MHz; CD_3CN): δ 9.73 (s, 1H), 9.69 (s, 1H), 7.65 (s, 1H), 7.51 (d, J = 8.6, 1H), 6.70 (d, J = 8.6, 1H), 6.52 (s, 1H). HRMS (ESI-MS) calcd for $C_7H_5ClO_3$ $[M - H]^-$: 170.9843 and 172.9814, found: 170.9874 and 172.9840. **23c**: ^{1}H NMR (500 MHz; CD3CN): δ 9.73 (s, 1H), 7.71 (s, 1H). HRMS (ESI-MS) calcd for $C_7H_4Cl_2O_3$ $[M - H]^-$: 204.9454 and 206.9425, found: 204.9450 and 206.9470.



4-chloro-3-hydroxy-2-naphthalenecarboxylic acid, methyl ester (24a): The reaction was conducted according to the general procedure. After maximum conversion to monohalogenated product was observed by LCMS Method 2, the reaction mixture was concentrated. Because nearly complete conversion to one monohalogenated product was observed (94% by LCMS), it was not necessary to purify the crude reaction mixture to determine selectivity. **24a**: ^{1}H (500 MHz; CD_2Cl_2): δ 11.08 (s, 1H), 8.41 (s, 1H), 8.14 (dd, J = 8.6, 0.7, 1H), 7.83 (dd, J = 8.2, 0.4, 1H), 7.67 (ddd, J = 8.5, 7.0, 1.3, 1H), 7.43 (ddd, J = 8.1, 6.9, 1.1, 1H), 4.05 (s, 3H). HRMS (ESI-MS) calcd for $C_{12}H_9ClO_3$ $[M - H]^-$: 235.0156 and 237.0128, found: 235.0214 and 237.0140.



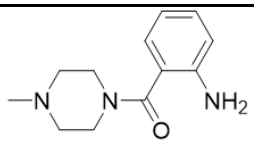
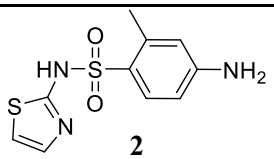
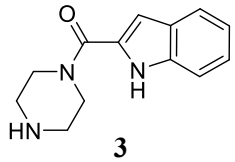
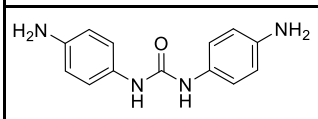
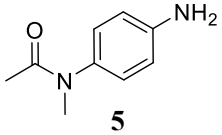
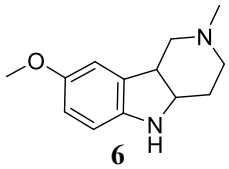
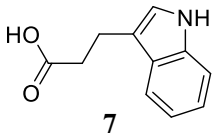
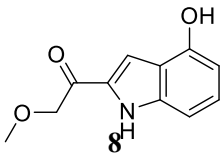
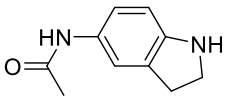
3-chloro-[1,1'-biphenyl]-4-ol (28a): The reaction was conducted according to the general procedure. After maximum conversion to monohalogenated product was observed by GCMS, the reaction mixture was concentrated onto silica gel. Purification by flash chromatography (SiO_2, 5% ethyl acetate/hexanes) afforded **28a** in 69% yield (41.1 mg, 0.201 mmol). ^{1}H NMR (500 MHz; CD_2Cl_2): δ 7.64 (d, J = 2.2, 1H), 7.59-7.57 (m, 2H), 7.50-7.46 (m, 3H), 7.40-7.37 (m, 1H), 7.14 (d, J = 8.4, 1H). HRMS (ESI-MS) calcd for $C_{12}H_9ClO$ $[M - H]^-$: 203.0258 and 205.0229, found: 203.0266 and 205.0275.



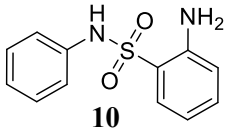
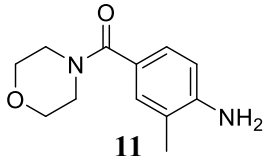
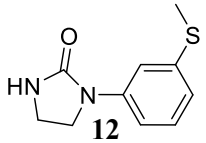
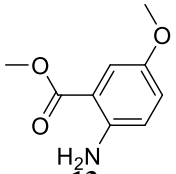
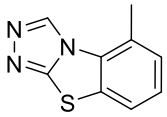
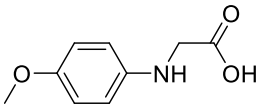
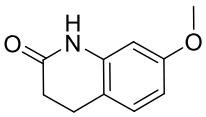
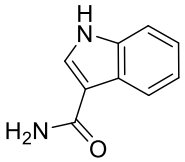
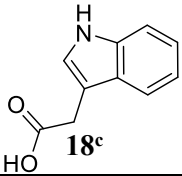
4-(2-chloro-1H-pyrrol-1-yl)phenol (37b): The reaction was conducted according to the general procedure. After maximum conversion to monohalogenated product was observed by LCMS Method 2, the reaction mixture was concentrated onto silica gel. Purification by flash chromatography (SiO_2, 4% ethyl acetate/hexanes) afforded **37b.** ^{1}H NMR (500 MHz; DMSO): δ 7.19-7.17 (m, 2H), 6.93 (m, 1H), 6.87-6.85 (m, 2H), 6.20 (m, 2H). HRMS (ESI-MS) calcd for $C_{10}H_8ClNO$ $[M + H]^-$: 194.0367 and 196.0338, found: 194.0484 and 196.0932. Note: polymerization of this product occurred within hours of isolation. For this reason, spectra were collected immediately after purification.

5.4.4 Additional data

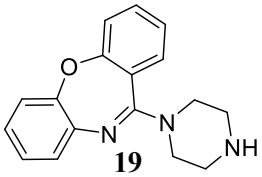
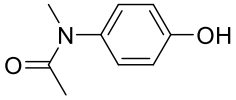
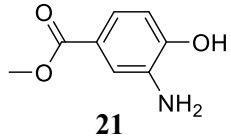
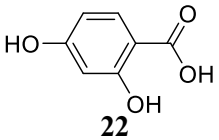
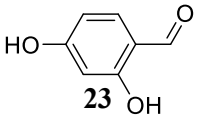
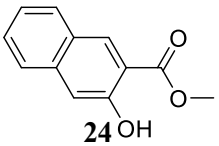
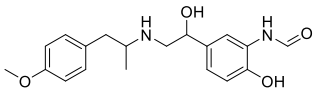
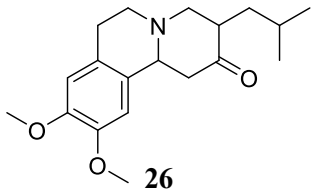
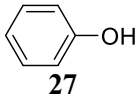
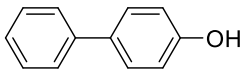
Table 5.2: Initial conversion data from panel 1 and 2.

Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
	% Conversion (LCMS)								
 **1**	1h	1	0	0	54	0	100	0	0
	ON	1	1	1	100	0	100	0	0
 **2**	1h	21	0	0	22	0	0	0	0
	ON	30	1	0	69	1	9	0	0
 **3**	1h	0	0	0	0	0	11	0	0
	ON	1	1	0	0	0	42	0	0
 **4**[c]	1h	-	-	-	-	-	-	-	-
	ON	100	1	0	0	1	1	0	0
 **5**	1h	0	0	0	9	0	15	0	0
	ON	0	1	0	1	0	1	0	0
 **6**	1h	0[a]	0	0[a]	0[a]	100	1	0	0
	ON	0	0	50[a]	35[ab]	50[a]	65[ab]	0	0
 **7**	1h	8	0	0	0	0	1	0	0
	ON	9	0	0	14	33	100	0	0
 **8**	1h	0	0	0	6	1	0	0	0
	ON	100	0	12	0	0	0	0	0
 **9**	1h	0	0	0	0	0	0	0	0
	ON	0	0	1[b]	0	0	100[b]	0	0

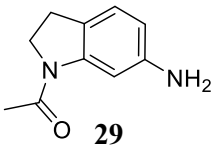
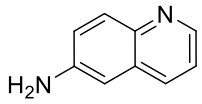
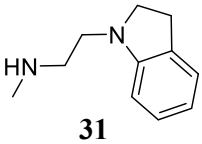
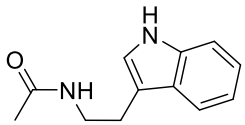
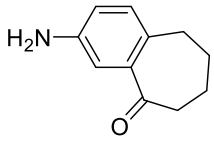
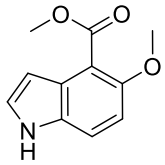
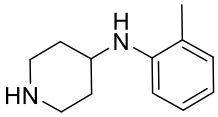
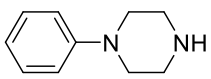
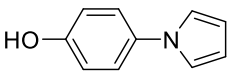
(Table 5.2: Continued from page 232)

Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
 **10**	1h	0	0	0	1	1	1	0	0
	ON	1	0	0	11	1	1	0	0
 **11**	1h	4	0	0	0	0	0	0	0
	ON	4	0	0	5	0	1	0	0
 **12**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	5	0	0
 **13**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	1	0	1	0	0
 **14**	1h	0	0	0	0	0	1	0	0
	ON	0	0	0	0	1	0	0	0
 **15**	1h	0	0	1	1	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **16**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	1	0	0
 **17**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	1	0	0
 **18[c]**	1h	-	-	-	-	-	-	-	-
	ON	0	0	1	0	0	0	0	0

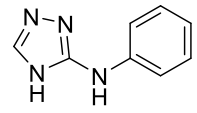
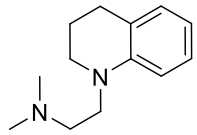
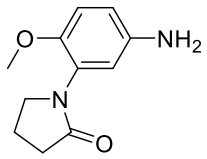
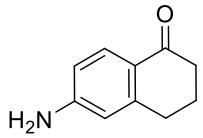
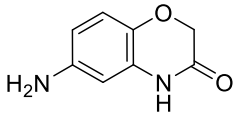
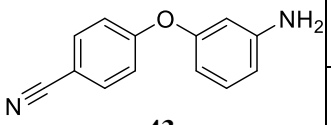
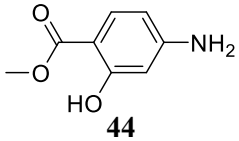
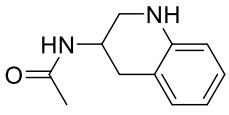
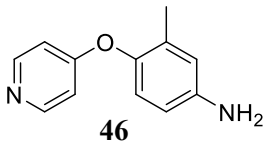
(Table 5.2: Continued from page 232)

Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
 **19**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	1	0	0
 **20**	1h	0	0	0	0	0	1	0	0
	ON	0	0	0	0	0	0	0	0
 **21**	1h	0	0	0	0	1	0	0	0
	ON	0	0	0	0	0	0	0	0
 **22**	ON	0	0	0	0	0	0	2	2
 **23**	ON	0	0	0	0	18	4	88	15
 **24**	ON	1	0	0	32	0	0	24	7
 **25**	ON	0	0	0	0	0	0	22	2
 **26**	ON	0	0	2	5	0	0	35	6
 **27**	ON	1	1	0	0	0	10	4	0
 **28**	ON	0	0	0	0	0	0	11	1

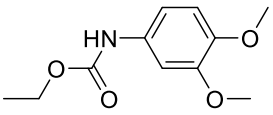
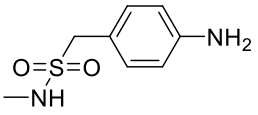
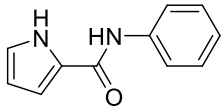
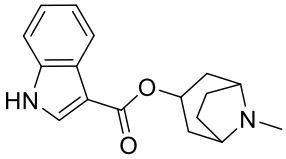
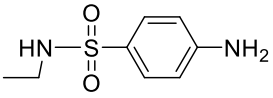
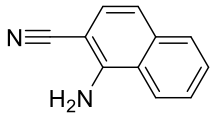
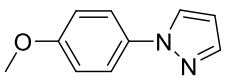
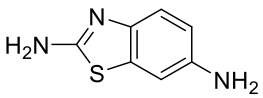
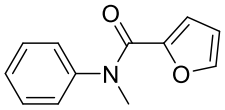
(Table 5.2: Continued from page 232)

Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
 **29**	1h	27	100	10	85	22	0	5	1
	ON	23	100	16	100	50	100	5	1
 **30**	1h	38	100	0	28	25	66	0	0
	ON	88	90	20	85	89	90	0	0
 **31**	1h	100	100	19	100	10	100	0	0
	ON	100	100	44	100	14	100	0	0
 **32**	1h	100	100	0	1	1	0	0	0
	ON	51	100	7	36	37	>100[d] (36%)	0	0
 **33**	1h	0	1	0	1	0	1	0	0
	ON	32	13	0	44	0	67	0	0
 **34**	1h	26	19	1	24	0	0	4	0
	ON	23	20	1	100	1	80	4	0
 **35**	1h	100	100	31	100	12	100	0	0
	ON	100	100	44	100	19	100	0	0
 **36**	1h	100	100	1	72	1	100	0	0
	ON	100	100	50	98	0	100	0	0
 **37**	1h	100	100	14	100	18	100	0	0
	ON	100	0	50	100	100	100	0	0

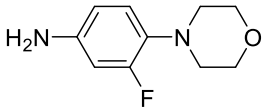
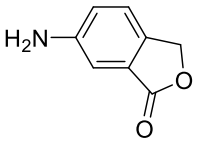
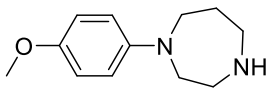
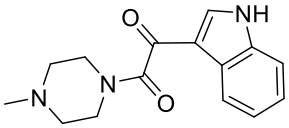
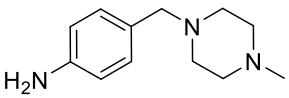
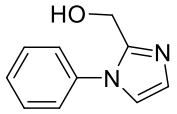
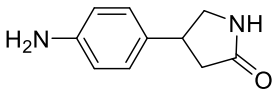
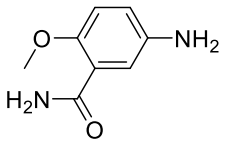
(Table 5.2: Continued from page 232)

Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
 **38**	1h	9	1	1	1	0	13	0	0
	ON	12	14	1	35	1	100	0	0
 **39**	1h	100	0[a]	0	100	0	0	0	0
	ON	100[b]	25[b]	35[b]	100[b]	0	100[b]	0	0
 **40**	1h	0	100	0	1	0	34	0	0
	ON	25	50	0	54	5	80	0	0
 **41**	1h	0	0	0	1	0	0	0	0
	ON	0	13	1	42	0	1	0	0
 **42[c]**	1h	100	100	0	100	0	100	0	0
	ON	-	-	-	-	-	-	-	-
 **43**	1h	26	100	0	33	35	29	4	0
	ON	73	100	13	100	100	100	4	0
 **44**	1h	14	55	0	50	1	31	0	0
	ON	8	0	11	56	46	100	0	0
 **45**	1h	46	39	15	26	50	25	1	1
	ON	49	50	41	70	70	70	1	1
 **46**	1h	1	100	0	1	0	100	0	0
	ON	100	100	1	100	1	100	0	0

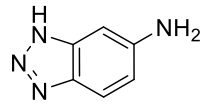
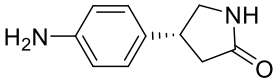
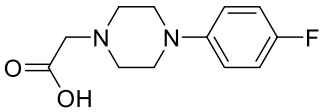
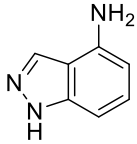
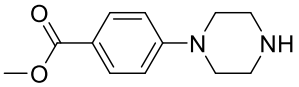
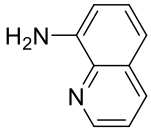
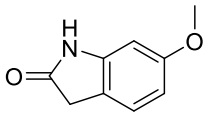
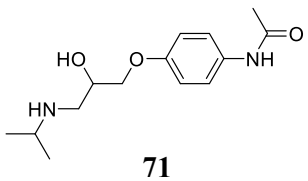
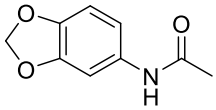
(Table 5.2: Continued from page 232)

Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
 **47**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **48**	1h	35	100	0	9	0	100	0	0
	ON	100	100	0	100	1	100	0	0
 **49**	1h	0	1	0	1	1	1	0	0
	ON	1	6	1	26	1	30	0	0
 **50**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **51**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **52**	1h	0	0	0	0	0	0	0	0
	ON	0	10	0	42	0	26	0	0
 **53**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **54**	1h	100	1	1	100	1	1	1	1
	ON	100	100	0	100	0	100	1	1
 **55**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0

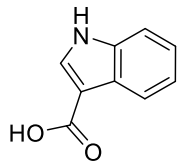
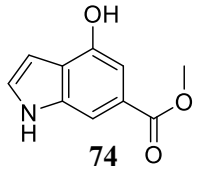
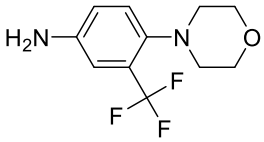
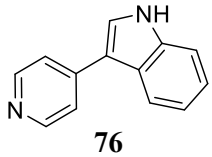
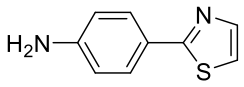
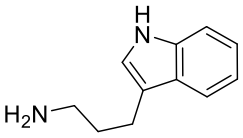
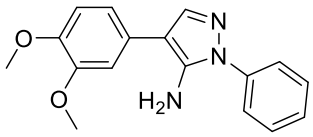
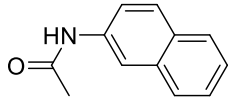
(Table 5.2: Continued from page 232)

Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
 **56**	1h	0	12	0	1	1	1	0	0
	ON	1	38	0	1	1	43	0	0
 **57**[c]	1h	-	-	-	-	-	-	-	-
	ON	0	0	0	0	0	0	0	0
 **58**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **59**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **60**[c]	1h	-	-	-	-	-	-	-	-
	ON	0	0	0	0	0	0	0	0
 **61**[c]	1h	0	0	0	0	0	0	0	0
	ON	-	-	-	-	-	-	-	-
 **62**	1h	100	>100[d] (54%)	100	100	100	>100[d] (N/A)	0	0
	ON	>100[d] (7%)	>100[d] (59%)	100	100	>100[d] (22%)	>100[d] (62%)	0	0
 **63**[c]	1h	-	-	-	-	-	-	-	-
	ON	0	1	0	1	0	1	0	0

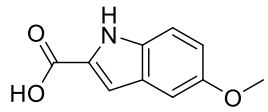
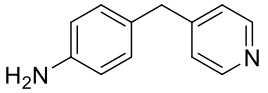
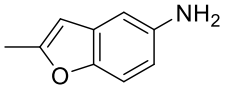
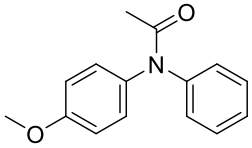
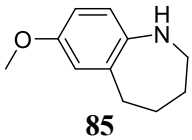
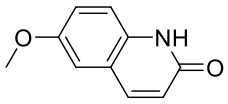
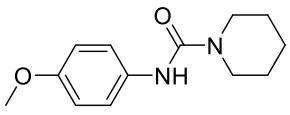
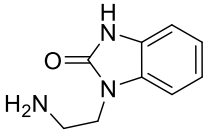
(Table 5.2: Continued from page 232)

Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
 **64**	1h	100	100	0	1	0	100	0	0
	ON	100	100	0	100	100	30	0	0
 **65**	1h	100	>100[d] (57%)	100	100	100	>100[d] (N/A)	0	0
	ON	100	>100[d] (99%)	100	100	>100[d] (N/A)	>100[d] (99%)	0	0
 **66[c]**	1h	0	0	0	0	0	0	0	0
	ON	-	-	-	-	-	-	-	-
 **67**	1h	0	>100[d] (99%)	1	0	100	>100[d] (N/A)	10	5
	ON	>100[d] (99%)	>100[d] (100%)	69	>100[d] (100%)	>100[d] (99%)	>100[d] (100%)	10	5
 **68**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **69**	1h	0	100	1	1	51	0	0	0
	ON	100	100	37	100	0[a]	100	0	0
 **70**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **71**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **72**	1h	0	0	0	0	0	0	0	0
	ON	0	1	0	0	0	0	0	0

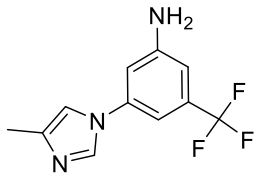
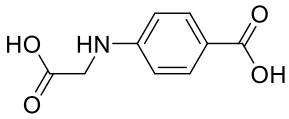
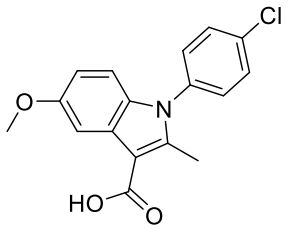
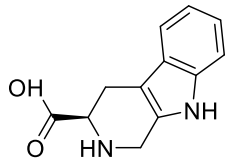
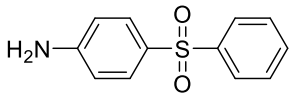
(Table 5.2: Continued from page 232)

Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
 **73**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **74**	1h	100	14	0	14	11	1	0	0
	ON	55	1	1	1	67	1	0	0
 **75**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **76**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **77**	1h	17	20	1	28	1	35	0	0
	ON	24	28	5	77	6	26	0	0
 **78**	1h	100	100	1	100	1	100	0	0
	ON	82	64	39	>100[d] (99%)	1	>100[d] (100%)	0	0
 **79**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **80**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0

(Table 5.2: Continued from page 232)

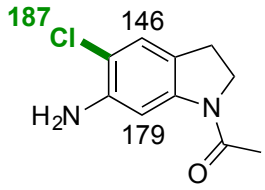
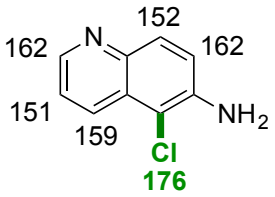
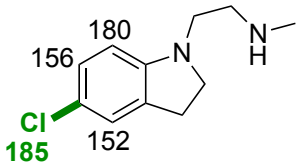
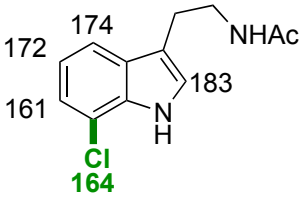
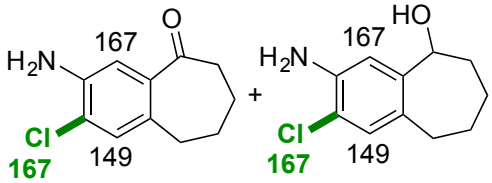
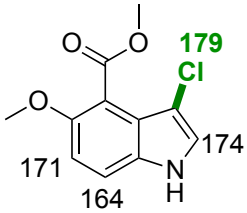
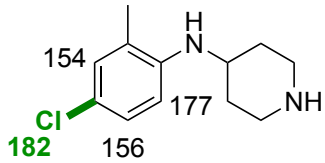
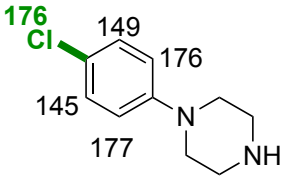
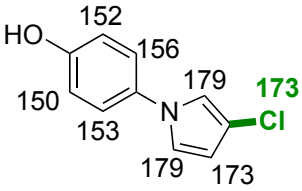
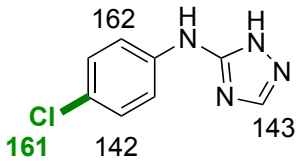
Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
 **81**	1h	0	0	1	1	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **82**	1h	100	100	100	100	1	100	0	0
	ON	100	100	1	100	1	100	0	0
 **83**	1h	100	84	1	100	25	100	10	1
	ON	37	100	1	1	86	100	10	1
 **84**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **85**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **86**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **87**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **88**	1h	0	0	0	0	0	0	0	0
	ON	0	100	0	0	0	0	0	0

(Table 5.2: Continued from page 232)

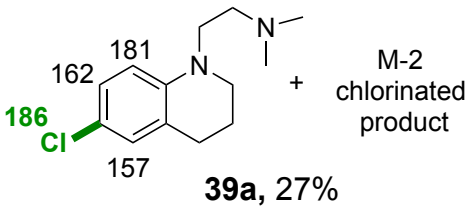
Compound	Time	Thal	RebH	10S	6TL	0K	4V	Rdc2	GsfI
 **89**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **90[c]**	1h	0	0	0	0	0	0	0	0
	ON	-	-	-	-	-	-	-	-
 **91**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0
 **92**	1h	0	0	0	0	0	8	0	0
	ON	10	11	10	6	0	24	0	0
 **93**	1h	0	0	0	0	0	0	0	0
	ON	0	0	0	0	0	0	0	0

[a] Species with a molecular weight "M-2" was observed.
[b] Species with a molecular weight "M-2+Cl" was observed. In this table, the listed conversion data includes both "M+Cl" as well as the "M-2+Cl" product.
[c] For these substrates, neither starting material nor product was observed.
[d] When a substrate was completely halogenated at one site, and began dichlorination at another site, this is indicated by ">100%." Values in parentheses underneath this are % conversions to dichlorinated product.

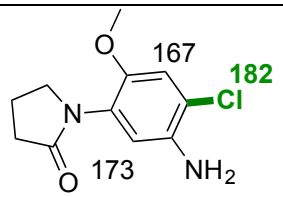
Table 5.3: Isolated products from bioconversions.[a]

 **29a,** 100%	 **30a,** 100%
 **31a,** 89%	 **32a,** 100%
 **33a, 33b,** 6%	 **34a,** 24%
 **35a,** 100%	 **36a,** 100%
 **37a,** 57%	 **38a,** 14%

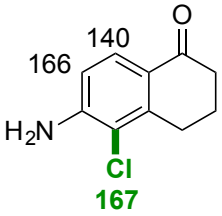
(Table 5.2: Continued from page 243)



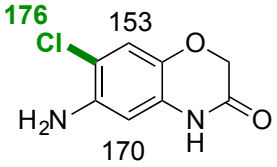
39a, 27%



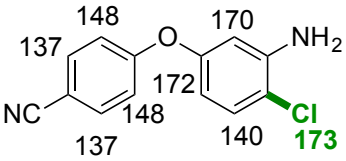
40a, 54%



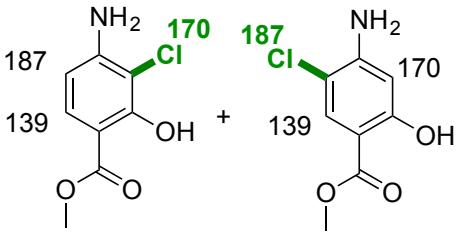
41a, 15%



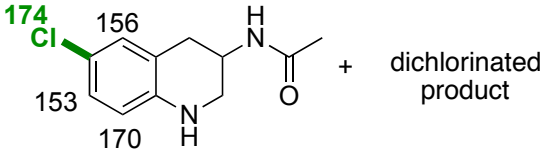
42a, 95%



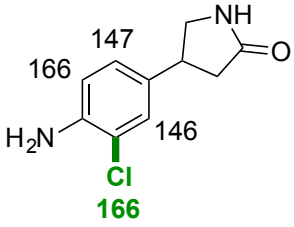
43a, 100%



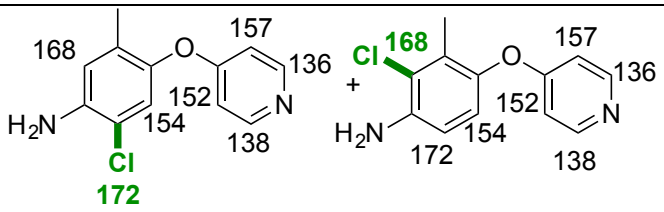
44a, 5% **44b**, 81%



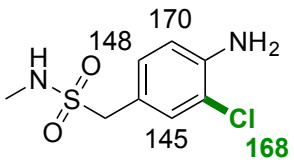
45a, 41% **45b**, 23%



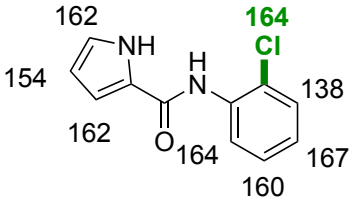
62a, 100%

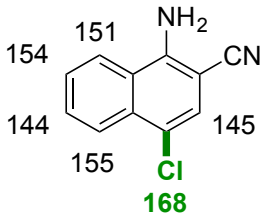


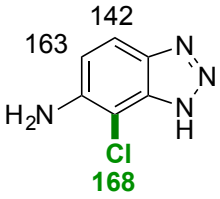
46a, 25% **46b,** 60%

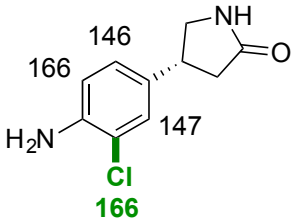


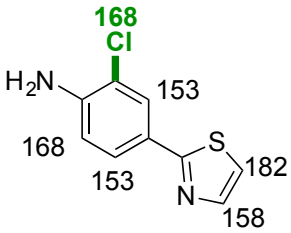
48a, 33%

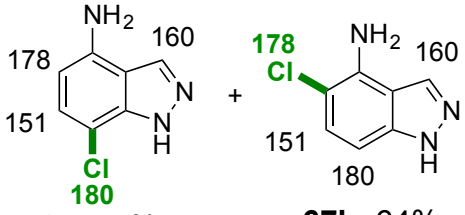
<table>
<tr>
<td>(Table 5.2: Continued from page 243)
 

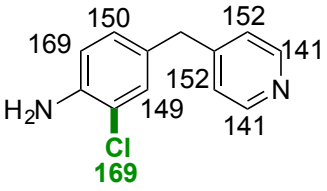
+ 2 different isomers on pyrrole ring
49a, 6%</td>
<td> 

52a, 10%</td>
</tr>
<tr>
<td> 

64a, N/A</td>
<td> 

65a, 100%</td>
</tr>
<tr>
<td> 

77a, 33%</td>
<td> 

67a, 4%
67b, 64%</td>
</tr>
<tr>
<td> 

82a, 79%</td>
<td></td>
</tr>
</table>

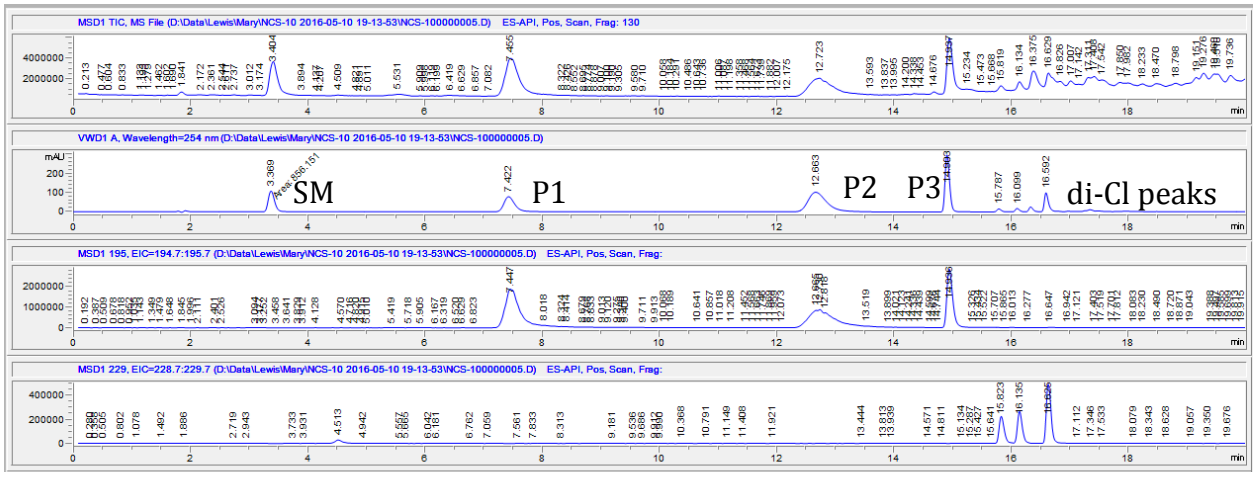
[a] HalA values for all sp^2 hybridized C-H bonds are shown next to the respective C-H bond. The HalA value for the C-H bond that was chlorinated in each case is shown in green. Substrate numbers are bolded, followed by the % conversion to product (LCMS).

Selectivity comparison with NCS: Data for selectivities shown in Table 5.1. Numbering refers to Table 5.1 entries.

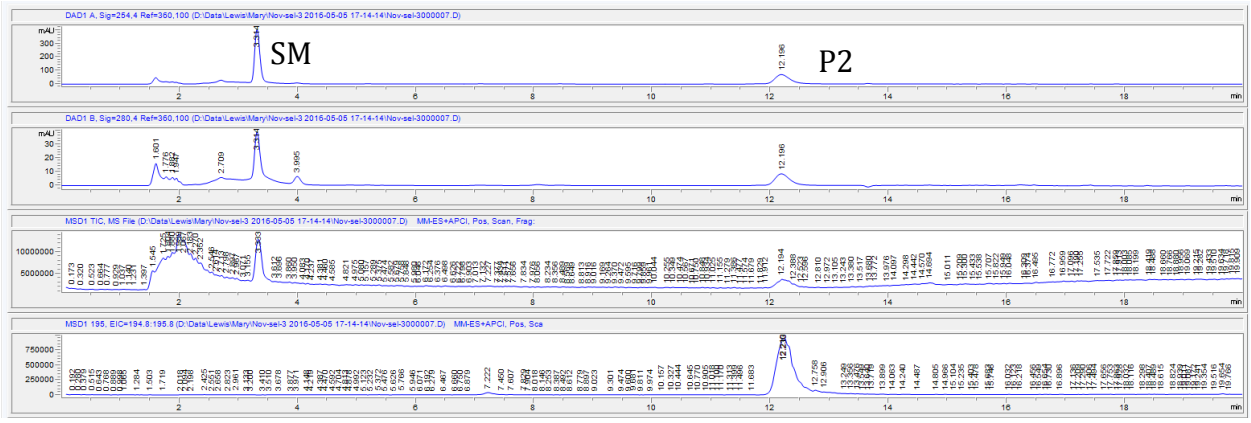
1. HNMR characterization of both RebH and NCS products for comparison.

2. LCMS characterization. Shown below are UV traces at 254 nm, total ion count, and extracted ion for the mono-chlorinated product, as well as the di-chlorinated product when necessary.

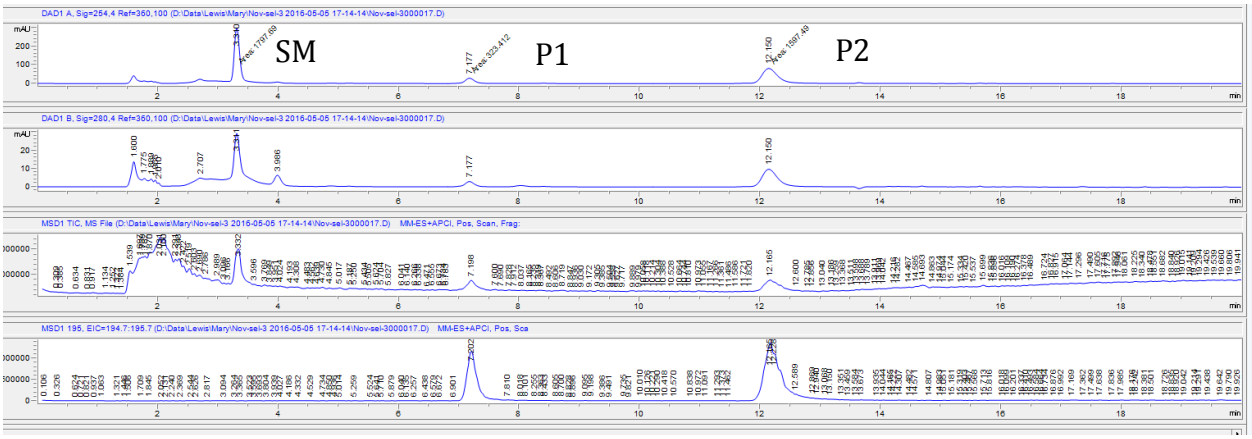
NCS



RebH

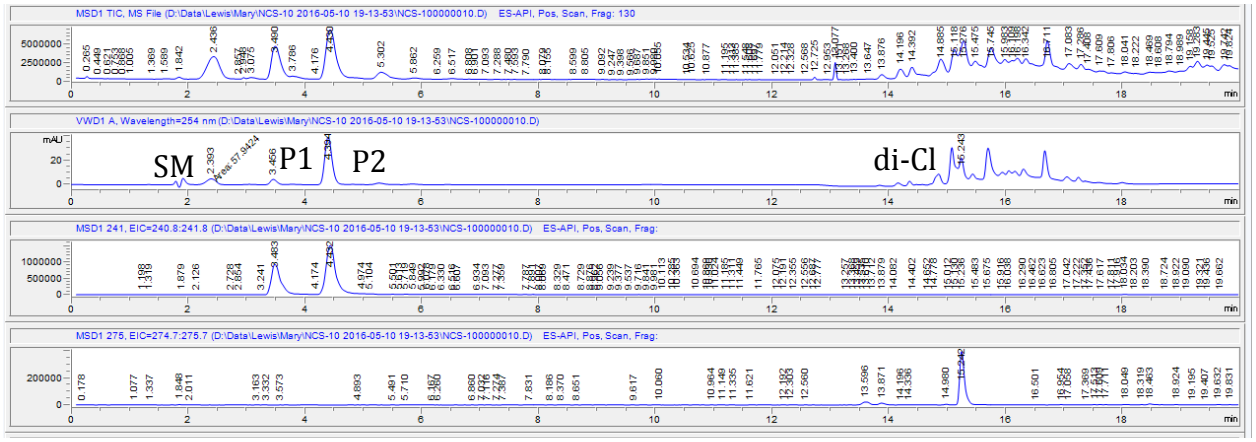


6TL



3. LCMS characterization. Shown below are UV traces at 254 nm, total ion count, and extracted ion for the mono-chlorinated product, as well as the di-chlorinated product when necessary.

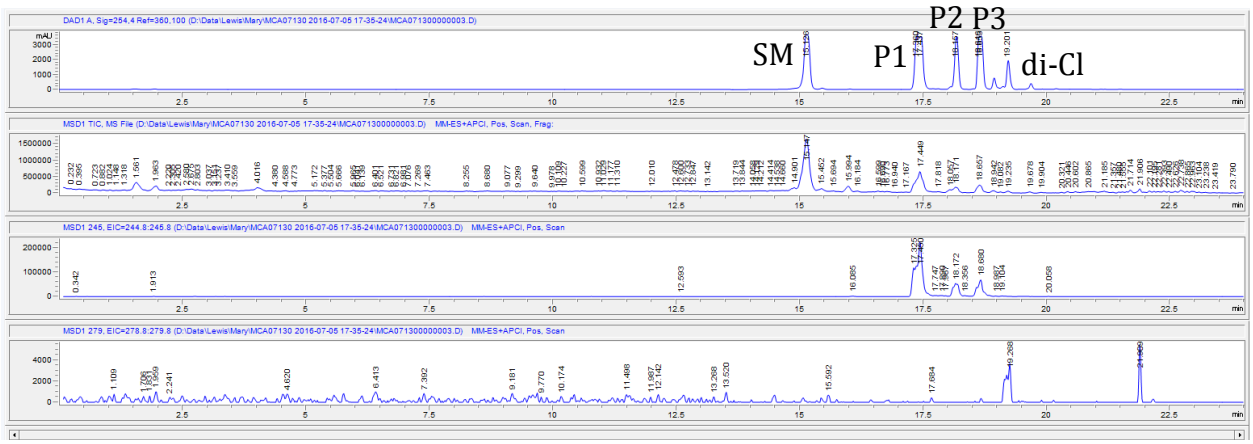
NCS



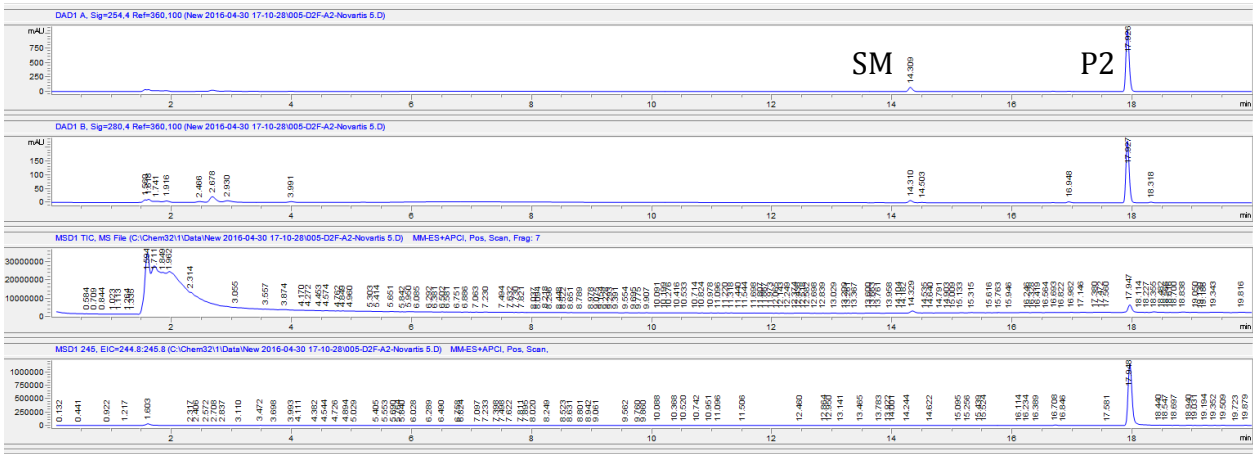
RebH

4. LCMS characterization. Shown below are UV traces at 254 nm, total ion count, and extracted ion for the mono-chlorinated product, as well as the di-chlorinated product when necessary.

NCS

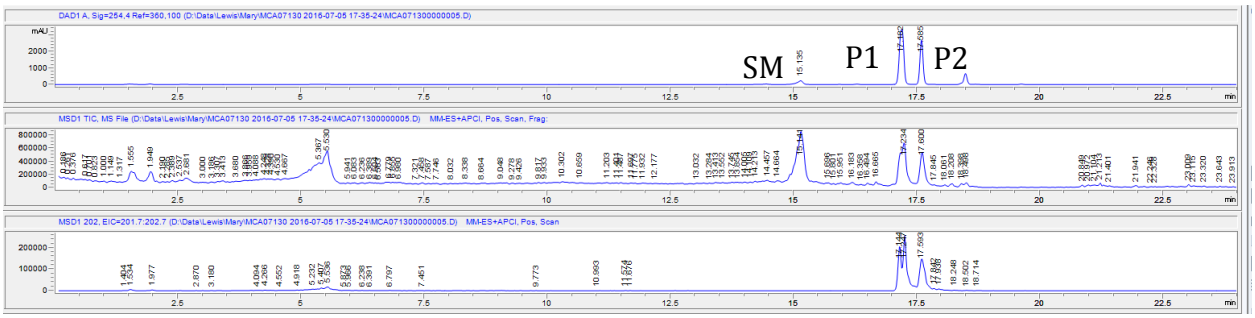


RebH



5. LCMS characterization. Shown below are UV traces at 254 nm, total ion count, and extracted ion for the mono-chlorinated product.

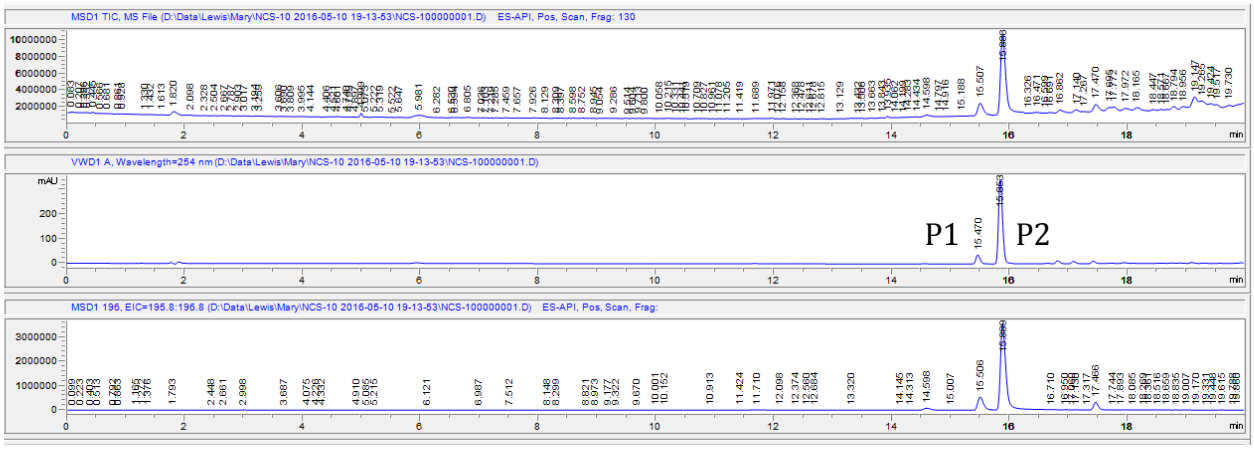
NCS



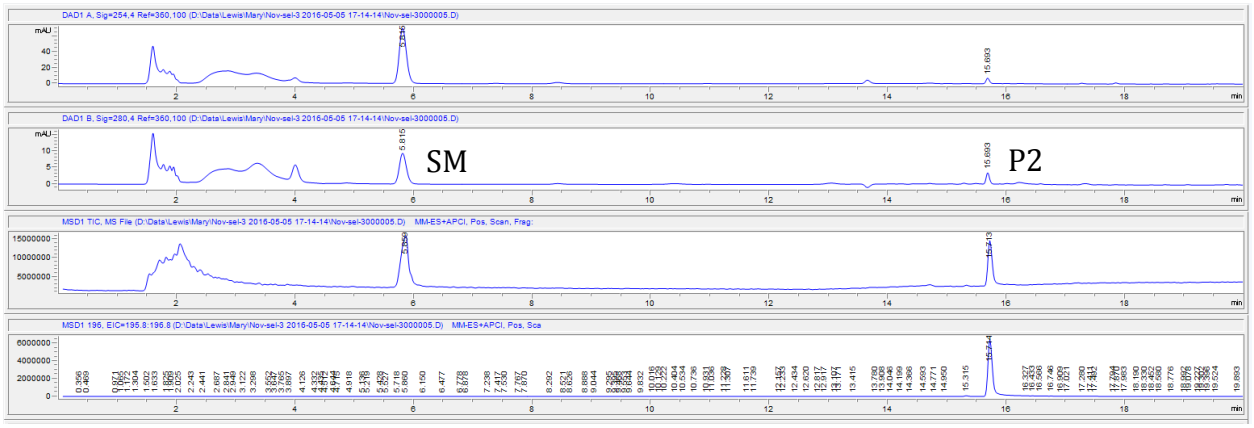
6TL

6. LCMS characterization. Shown below are UV traces at 254 nm, total ion count, and extracted ion for the mono-chlorinated product.

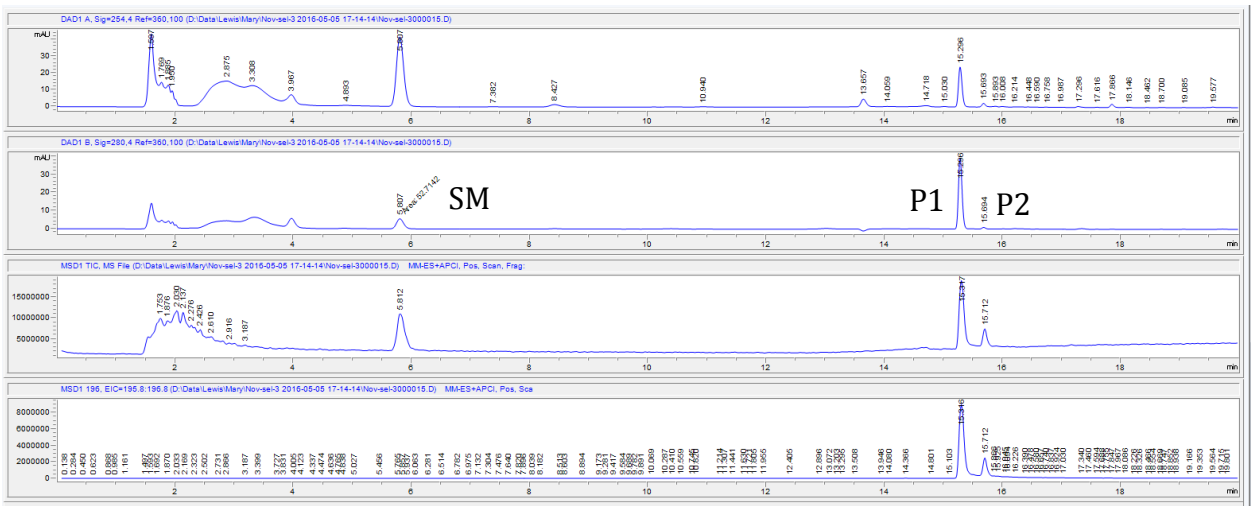
NCS



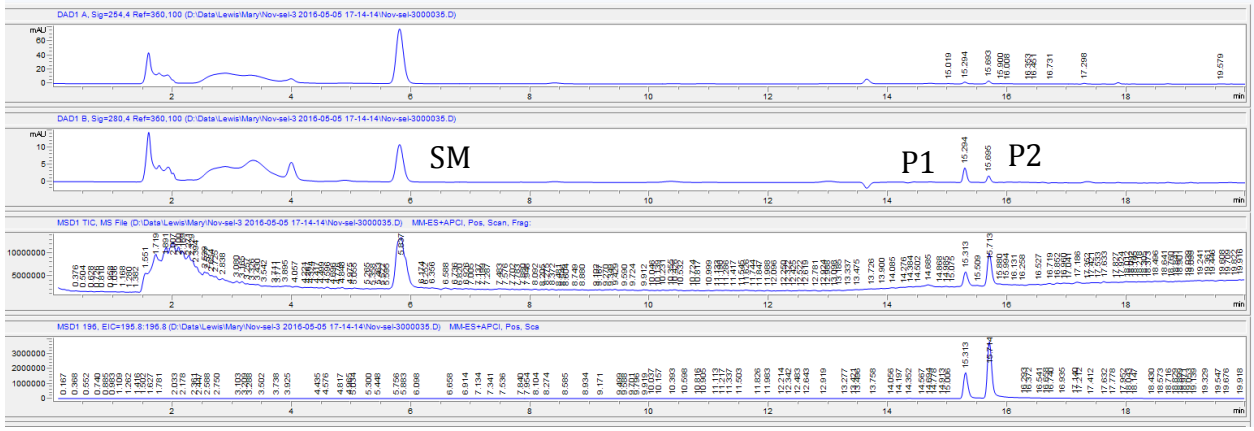
RebH



6TL

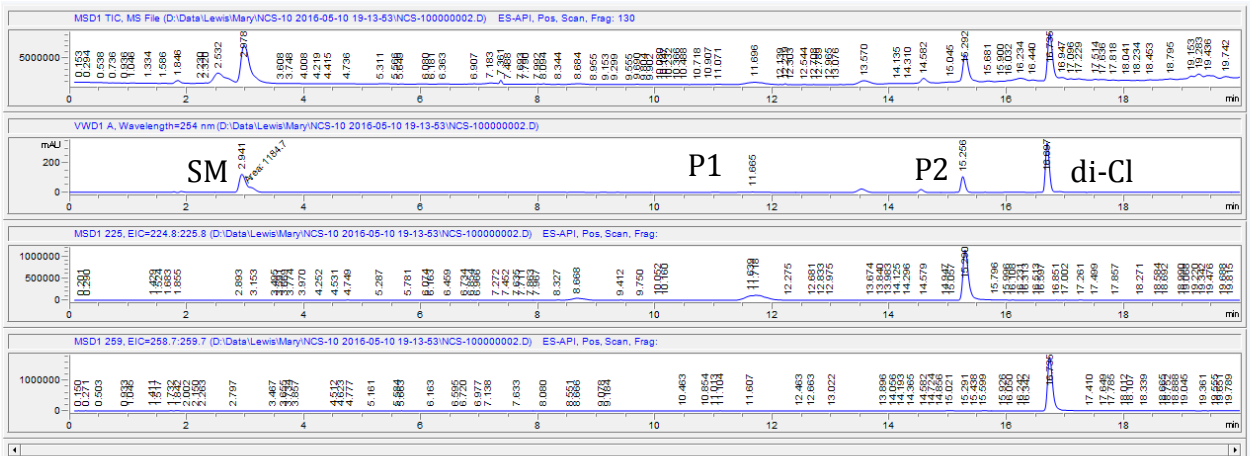


4V

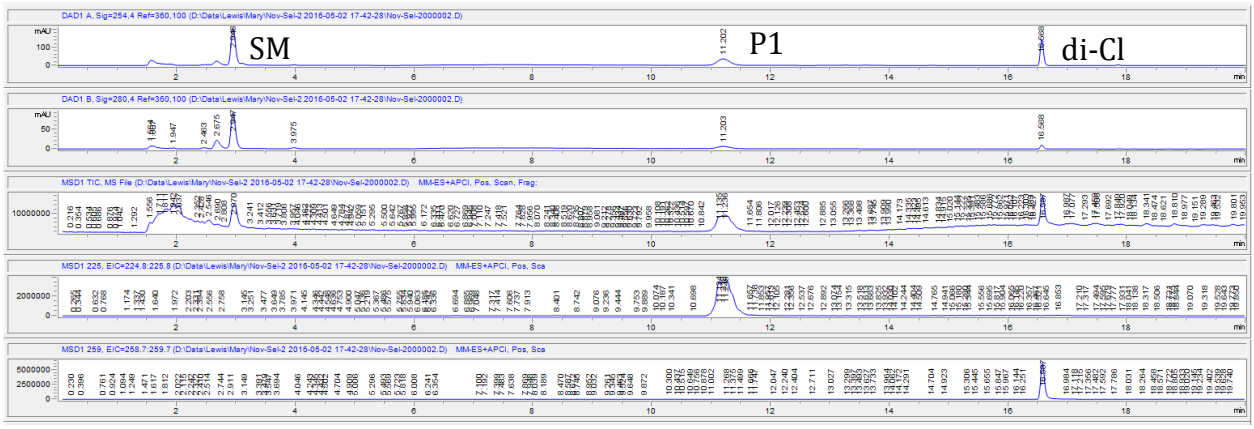


7. LCMS characterization. Shown below are UV traces at 254 nm, total ion count, and extracted ion for the mono-chlorinated product, as well as the di-chlorinated product when necessary.

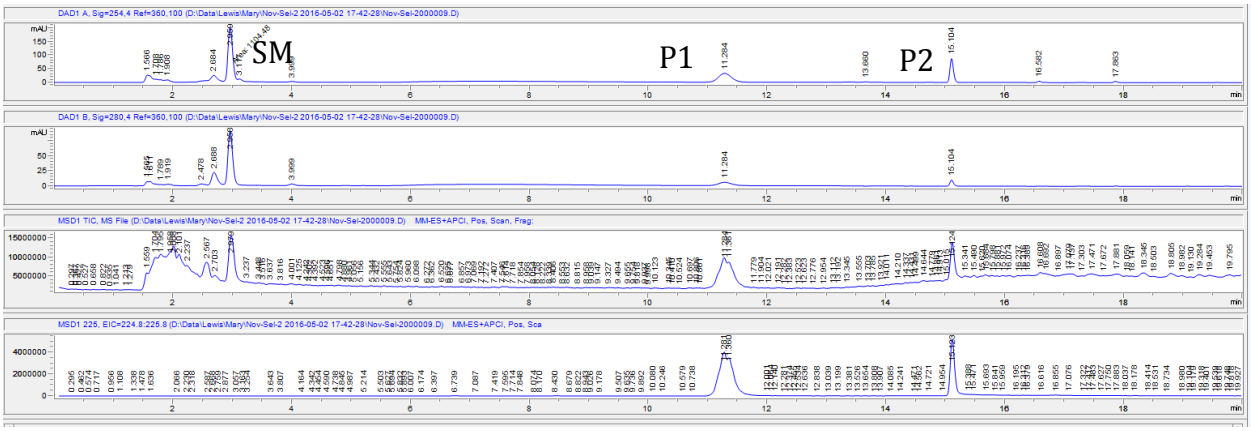
NCS



RebH

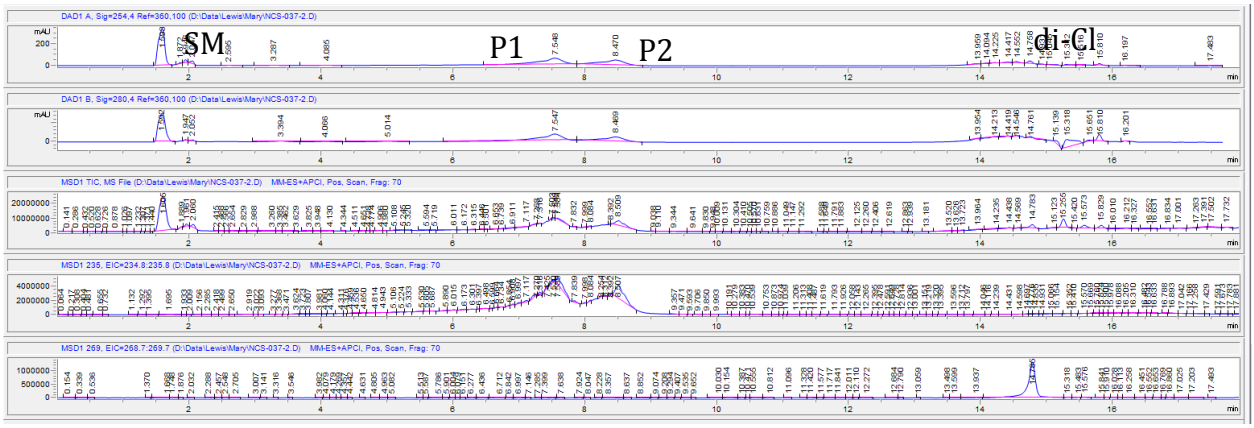


6TL

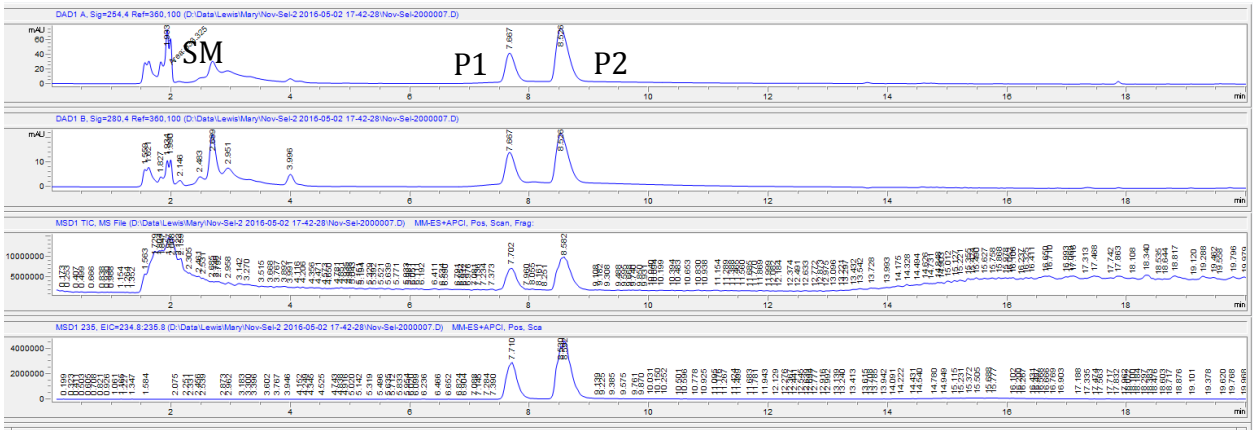


8. LCMS characterization. Shown below are UV traces at 254 nm, total ion count, and extracted ion for the mono-chlorinated product, as well as the di-chlorinated product when necessary.

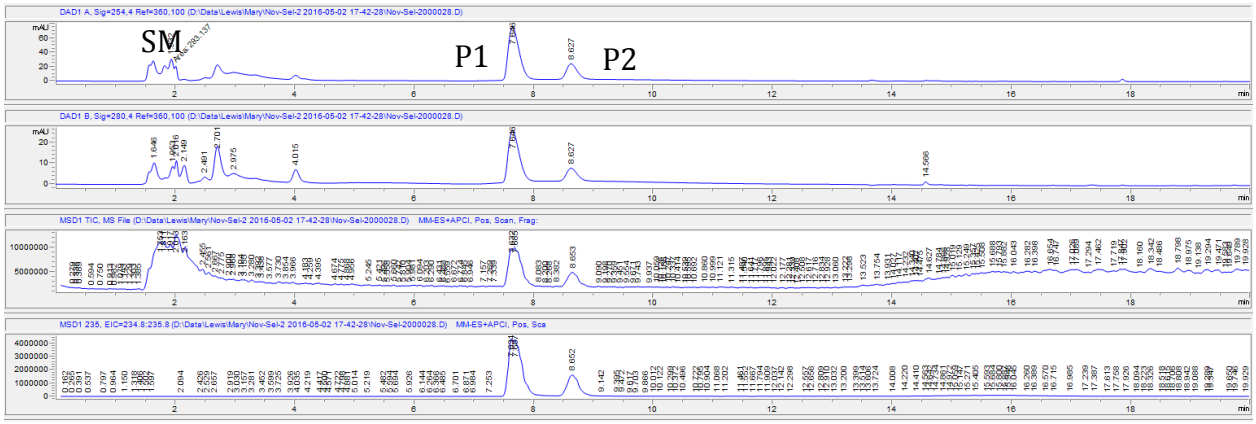
NCS



RebH

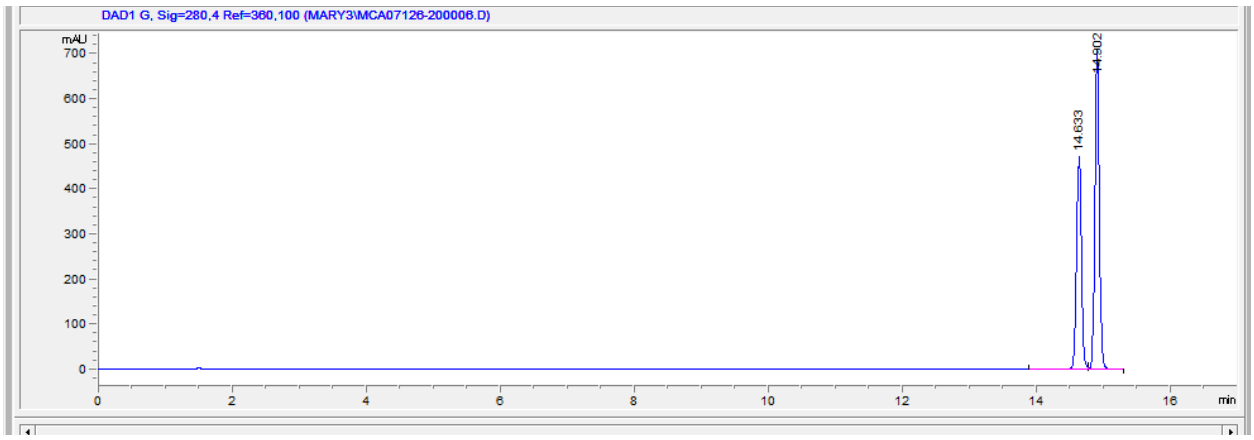


4V

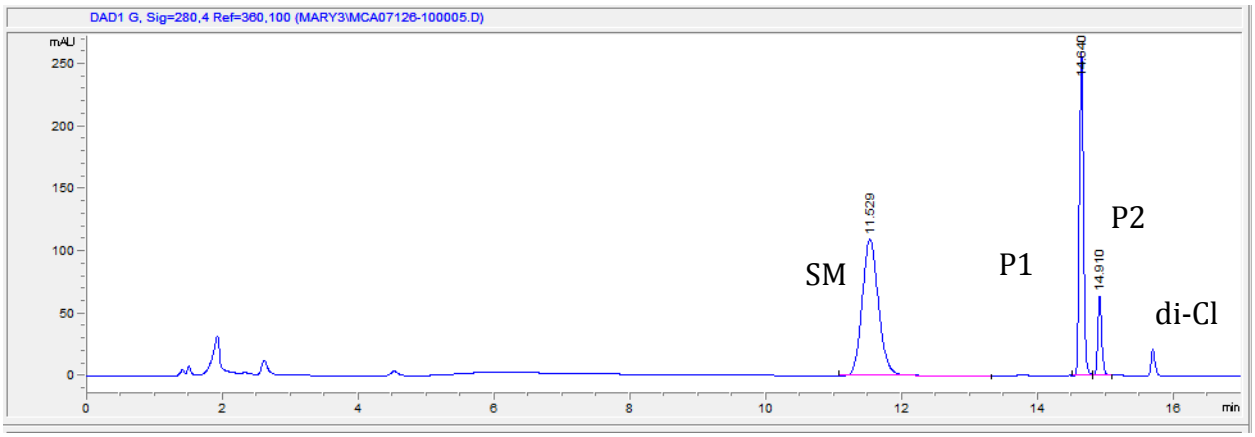


9. UHPLC characterization. The NCS product trace shown below was isolated and characterized by HNMR.

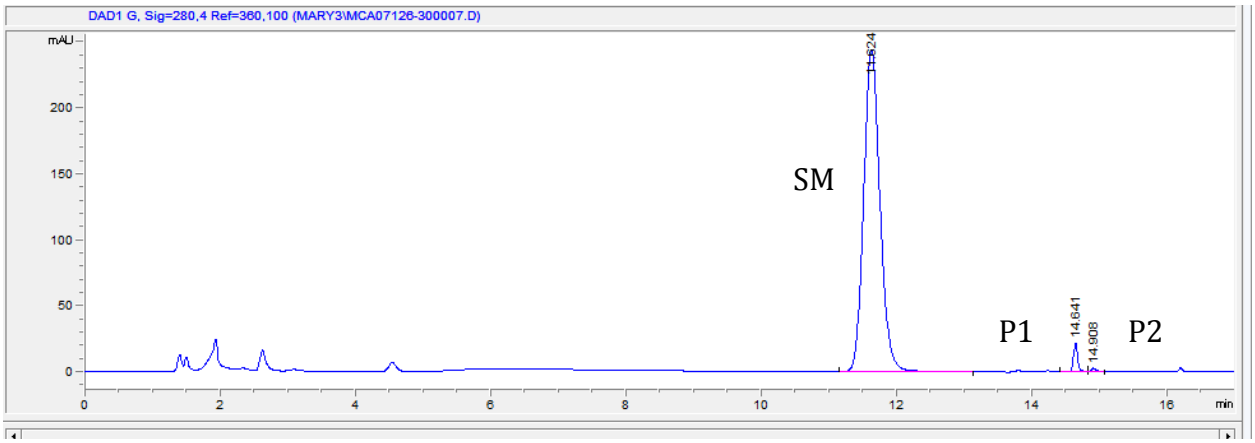
NCS



Rdc2

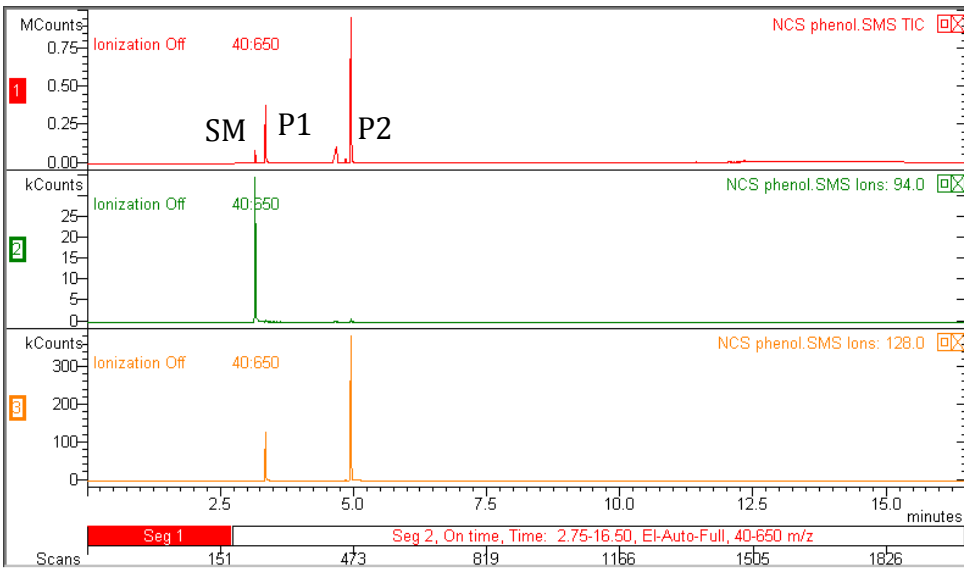


GsfI

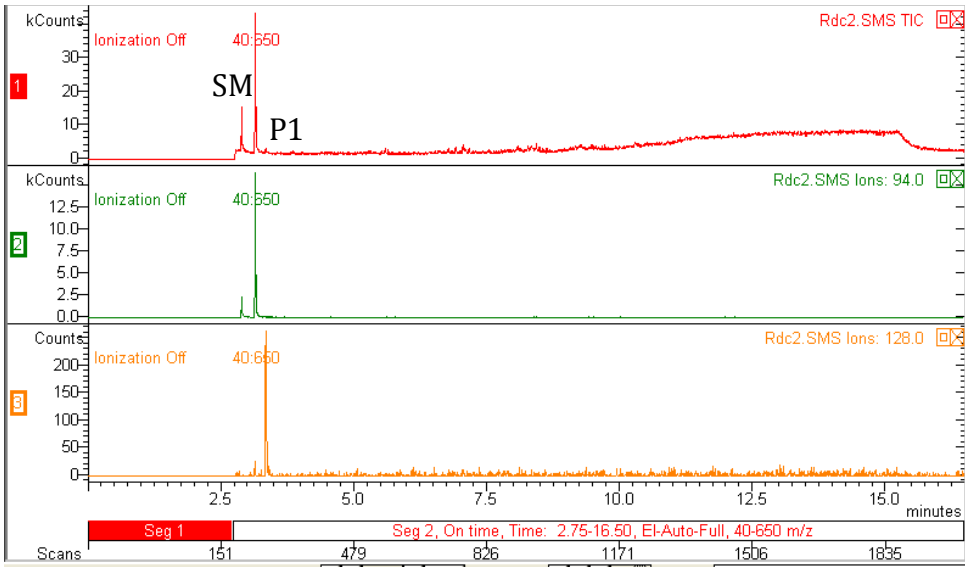


10. GCMS characterization. Shown below are the total ion count and extracted ion traces for the starting material and mono-chlorinated product.

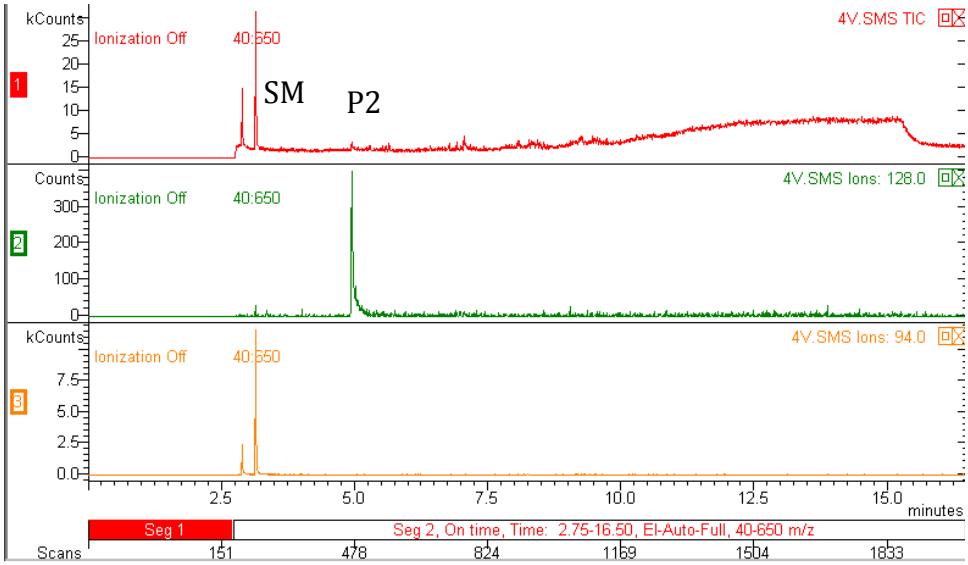
NCS



Rdc2



4V



Acknowledgements

We would like to thank the Center for Selective C-H Functionalization for enabling and funding this collaboration as well as Novartis Institutes for Biomedical Research for supplying substrates, facilities, and other materials. We would like to thank Clayton Springer for making models of the halenium adducts for substrates in panel 1. This work was completed in part with

resources provided by the University of Chicago Research Computing Center, with assistance from Dr. Jonathan Skone. We also thank Prof. Karl-Heinz van Pée for supplying the Thal gene. Lastly, we thank Laurel Heckman and Randy Richter for housing me during my time at NIBR.

5.5 References

1. Agarwal, V. *et al*. Enzymatic Halogenation and Dehalogenation Reactions: Pervasive and Mechanistically Diverse. *Chem. Rev*. acs.chemrev.6b00571–56 (2017). doi:10.1021/acs.chemrev.6b00571
2. Podzelinska, K. *et al*. Chloramphenicol Biosynthesis: The Structure of CmlS, a Flavin-Dependent Halogenase Showing a Covalent Flavin–Aspartate Bond. *J. Mol. Biol*. **397,** 316–331 (2010).
3. Zeng, J. & Zhan, J. A Novel Fungal Flavin-Dependent Halogenase for Natural Product Biosynthesis. *ChemBioChem* **11,** 2119–2123 (2010).
4. Zeng, J., Lytle, A. K., Gage, D., Johnson, S. J. & Zhan, J. Specific chlorination of isoquinolines by a fungal flavin-dependent halogenase. *Bioorg. Med. Chem. Lett*. **23,** 1001–1003 (2013).
5. Xu, F., Merkley, A., Yu, D. & Zhan, J. Selective biochlorination of hydroxyquinolines by a flavin-dependent halogenase. *Tetrahedron Letters* **57,** 5262–5265 (2016).
6. Chooi, Y.-H., Cacho, R. & Tang, Y. Identification of the Viridicatumtoxin and Griseofulvin Gene Clusters from Penicillium aethiopicum. *Chem. Biol*. **17,** 483–494 (2010).
7. Payne, J. T., Andorfer, M. C. & Lewis, J. C. Regioselective Arene Halogenation using the FAD-Dependent Halogenase RebH. *Angew. Chem. Int. Ed*. **52,** 5271–5274 (2013).
8. Payne, J. T., Poor, C. B. & Lewis, J. C. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angew. Chem*. **127,** 4300–4304 (2015).
9. Andorfer, M. C., Park, H. J., Vergara-Coll, J. & Lewis, J. C. Directed evolution of RebH for catalyst-controlled halogenation of indole C–H bonds. *Chem. Sci*. **7,** 3720–3729 (2016).
10. Belsare, K. D. *et al*. A Simple Combinatorial Codon Mutagenesis Method for Targeted Protein Engineering. *ACS Synth. Biol*. **6,** 416–420 (2017).
11. Shepherd, S. A. *et al*. Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chem. Sci*. **6,** 3454–3460 (2015).
12. Seibold, C. *et al*. A flavin-dependent tryptophan 6-halogenase and its use in modification of pyrrolnitrin biosynthesis. *Biocatalysis and Biotransformation* **24,** 401–408 (2006).
13. Landwehr, M., Carbone, M., Otey, C. R., Li, Y. & Arnold, F. H. Diversification of Catalytic Function in a Synthetic Family of Chimeric Cytochrome P450s. *Chem. Biol*. **14,** 269–278 (2007).

14. Zhang, K., Damaty, El, S. & Fasan, R. P450 Fingerprinting Method for Rapid Discovery of Terpene Hydroxylating P450 Catalysts with Diversified Regioselectivity. *J. Am. Chem. Soc.* **133,** 3242–3245 (2011).
15. Ashtekar, K. D. *et al.* A new tool to guide halofunctionalization reactions: the halenium affinity (HalA) scale. *J. Am. Chem. Soc.* **136,** 13355–13362 (2014).
16. Andorfer, M. C. *et al.* Understanding Flavin-Dependent Halogenase Reactivity via Substrate Activity Profiling. *ACS Catal.* **7,** 1897–1904 (2017).
17. Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases †,‖. *Biochemistry* **46,** 1284–1292 (2007).
18. Flecks, S. *et al.* New Insights into the Mechanism of Enzymatic Chlorination of Tryptophan. *Angew. Chem. Int. Ed.* **47,** 9533–9536 (2008).
19. Erlanson, D. A., McDowell, R. S. & O'Brien, T. Fragment-Based Drug Discovery. *J. Med. Chem.* **47,** 3463–3482 (2004).
20. Lewis, J. C. & Arnold, F. H. Catalysts on Demand: Selective Oxidations by Laboratory-Evolved Cytochrome P450 BM3. *CHIMIA* **63,** 309–312 (2009).
21. Matsumura, I. & Ellington, A. D. In vitro Evolution of Beta-glucuronidase into a Beta-galactosidase Proceeds Through Non-specific Intermediates. *J. Mol. Biol.* **305,** 331–339 (2001).
22. Pratihar, S. & Roy, S. Nucleophilicity and Site Selectivity of Commonly Used Arenes and Heteroarenes. *J. Org. Chem.* **75,** 4957–4963 (2010).
23. Löfdahl, C. G. & Svedmyr, N. Formoterol fumarate, a new β2-adrenoceptor agonist. *Allergy* **44,** 264–271 (1989).
24. Weise, N. J., Parmeggiani, F., Ahmed, S. T. & Turner, N. J. The Bacterial Ammonia Lyase EncP: A Tunable Biocatalyst for the Synthesis of Unnatural Amino Acids. *J. Am. Chem. Soc.* **137,** 12977–12983 (2015).
25. Dong, C. J. *et al.* Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* **309,** 2216–2219 (2005).
26. Tishchenko, O., Pham-Tran, N.-N., Kryachko, E. S. & Nguyen, M. T. Protonation of Gaseous Halogenated Phenols and Anisoles and Its Interpretation Using DFT-Based Local Reactivity Indices. *J. Phys. Chem. A* **105,** 8709–8717 (2001).
27. Schwabe, T. & Grimme, S. Theoretical Description of Substituent Effects in Electrophilic Aromatic Substitution Reactions. *Eur. J. Org. Chem.* **2008,** 5928–5935 (2008).
28. Aoki, S. *et al.* Regioselective Aromatic Substitution Reactions of Cyclometalated Ir(III) Complexes: Synthesis and Photochemical Properties of Substituted Ir(III) Complexes That Exhibit Blue, Green, and Red Color Luminescence Emission. *Inorg. Chem.* **50,** 806–818 (2011).
29. Abagyan, R., Totrov, M. & Kuznetsov, D. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry* **15,** 488–506 (1994).
30. Grant, J. A., Gallardo, M. A. & Pickup, B. T. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *Journal of Computational Chemistry* **17,** 1653–1666 (1996).

31. Negretti, S. *et al*. Directing Group-Controlled Regioselectivity in an Enzymatic C–H Bond Oxygenation. *J. Am. Chem. Soc*. **136,** 4901–4904 (2014).
32. Narayan, A. R. H. *et al*. Enzymatic hydroxylation of an unactivated methylene C-H bond guided by molecular dynamics simulations. *Nat. Chem*. **7,** 653–660 (2015).
33. Yeh, E., Garneau, S. & Walsh, C. T. Robust in vitro Activity of RebF and RebH, a Two-Component Reductase/Halogenase, Generating 7-Chlorotryptophan during Rebeccamycin Biosynthesis. *Proc. Natl. Acad. Sci*. **102,** 3960–3965 (2005).
34. Cabrita, L. D., Dai, W. & Bottomley, S. P. A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol* **6,** 12 (2006).

Chapter 6: Discovering new flavin-dependent halogenases through genome mining

6.1 Introduction

Genome mining is a powerful approach to discover new natural products, pathways, and biocatalysts.[1-3] As DNA sequencing techniques have become less expensive and more rapid, metagenomics endeavors (i.e. sequencing of whole genomes of microbial organisms within a sample) have allowed for the identification of millions of novel genes.[1,4,5] Bioinformatics tools such as BLAST have allowed researchers to easily search these enormous quantities of genomics data to identify homologues of a particular gene of interest. Engineering approaches have successfully led to novel FDHs with altered stability, substrate scope, and site selectivity; however, genome mining approaches could serve as an alternative method for FDH discovery and provide new starting points for further engineering efforts. Very recently, Micklefield and coworkers have used this approach to identify a thermostable flavin-dependent tryptophan halogenase (Trp-FDH).[6] Although FDHs have been characterized in this manner on a relatively small scale, the sequence diversity of FDHs is growing rapidly and thousands remain uncharacterized. In an effort to discover and characterize novel FDHs, we partnered with the Joint Genome Institute to synthesize and characterize a library of ~100 putative FDHs.

Authorship

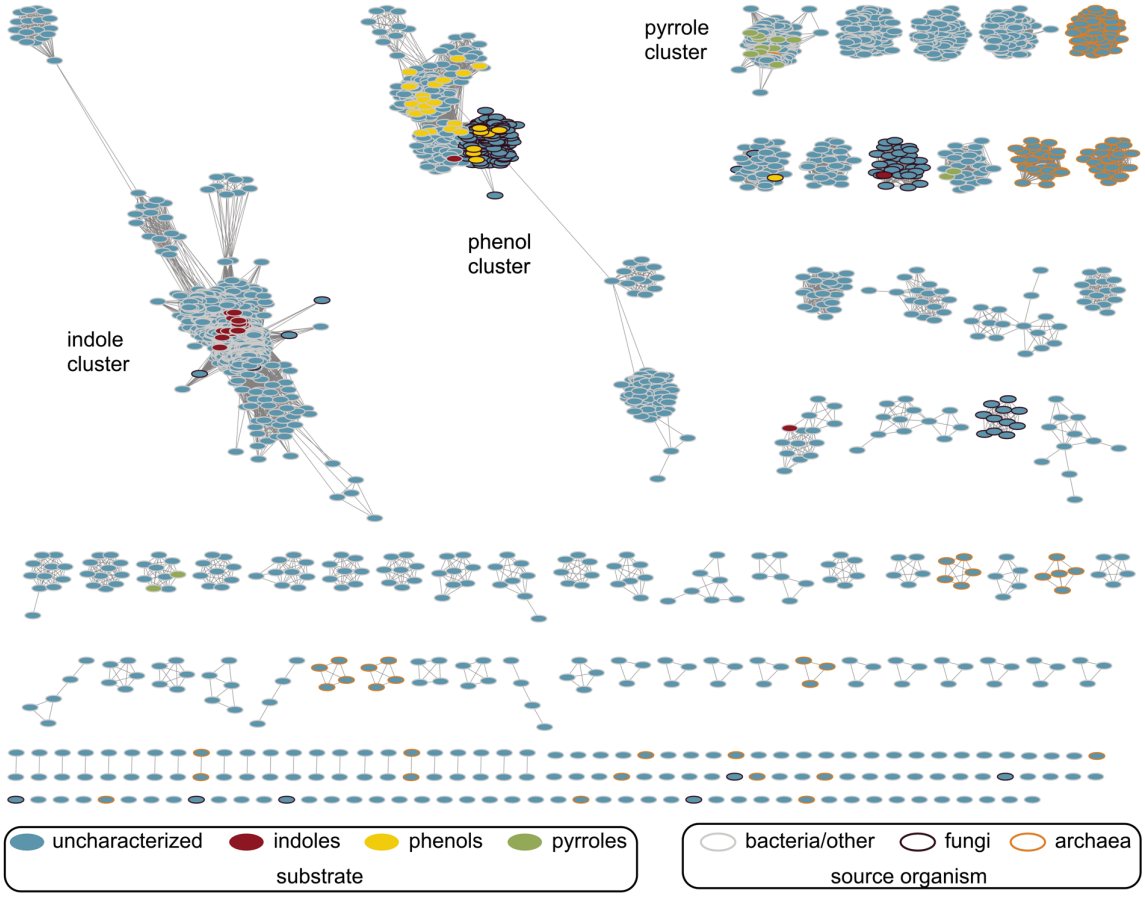
Prof. Jared Lewis performed the initial BLAST search and created the sequence similarity network shown in Fig. 6.1. Harrison Snodgrass, another graduate student in the Lewis lab, helped to verify the screen described in section 6.2.2 and is continuing this project. Putative FDH genes were synthesized by the Joint Genome Institute.

6.2 Results and Discussion

6.2.1 Halogenase sequence similarity network and analysis

To compile an initial set of putative flavin-dependent halogenase (FDH) genes, Prof. Lewis used BLAST (Basic Local Alignment Search Tool) to find gene sequences with similarity to the Trp-FDH RebH. Using RebH as the search sequence, we obtained approximately 4,000 sequences. A sequence similarity network (SSN) was then created for unique sequences from this list using the Enzyme Function Initiative's Enzyme Similarity Tool[7] (EFI-EST, Fig. 6.1). Approximately 60-70 genes in this network had been previously characterized as FDHs, and, when plotted on our SSN, they cluster together according to native substrate (Fig. 6.1, circles are color-coded according to substrate). For example, most known Trp-FDHs are located in a single cluster of the SSN, labeled as the indole cluster in Fig. 6.1. Similarly, FDHs thought to natively halogenate phenols, from both bacterial and fungal organisms, cluster together, as do pyrrole halogenases.

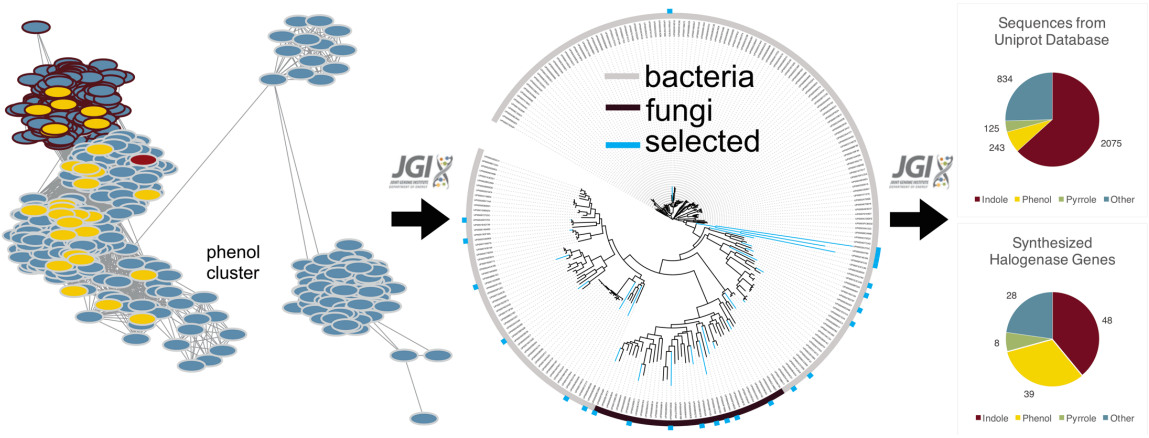
Figure 6.1: Sequence similarity network for putative flavin-dependent halogenases.



We next wanted to narrow the list of sequences to around 100 putative FDH genes. We split the full list of unique, putative FDH sequences according to cluster and/or source organism, which resulted in 6 categories – archaeal, bacterial phenol, fungal phenol, indole, bacterial "other," and pyrrole. Each category contained a different number of sequences, for example, the "indole" category contained 2,075, corresponding to ~63% of all sequences. We did not want, however, 63% of our final 100-member library to contain tryptophan FDHs, since they are the most well characterized in terms of both mechanism and substrate scope. To enrich our library with a higher diversity of novel FDHs, we determined what percentage of each category we wanted our final library to be (Fig. 6.2). Putative Trp-FDHs were underrepresented relative to total and putative phenol FDHs were overrepresented (Fig. 6.2, pie graphs). The Joint Genome

Institute (JGI) analyzed these six sequence lists, determined the maximally diverse set for each, and synthesized 123 putative FDH genes total. They cloned these into a pET28 vector, transformed them into BL21(DE3) containing GroEL/ES chaperones, and sent this library to us to analyze.

Figure 6.2: Overview of putative FDH sequence selection method.

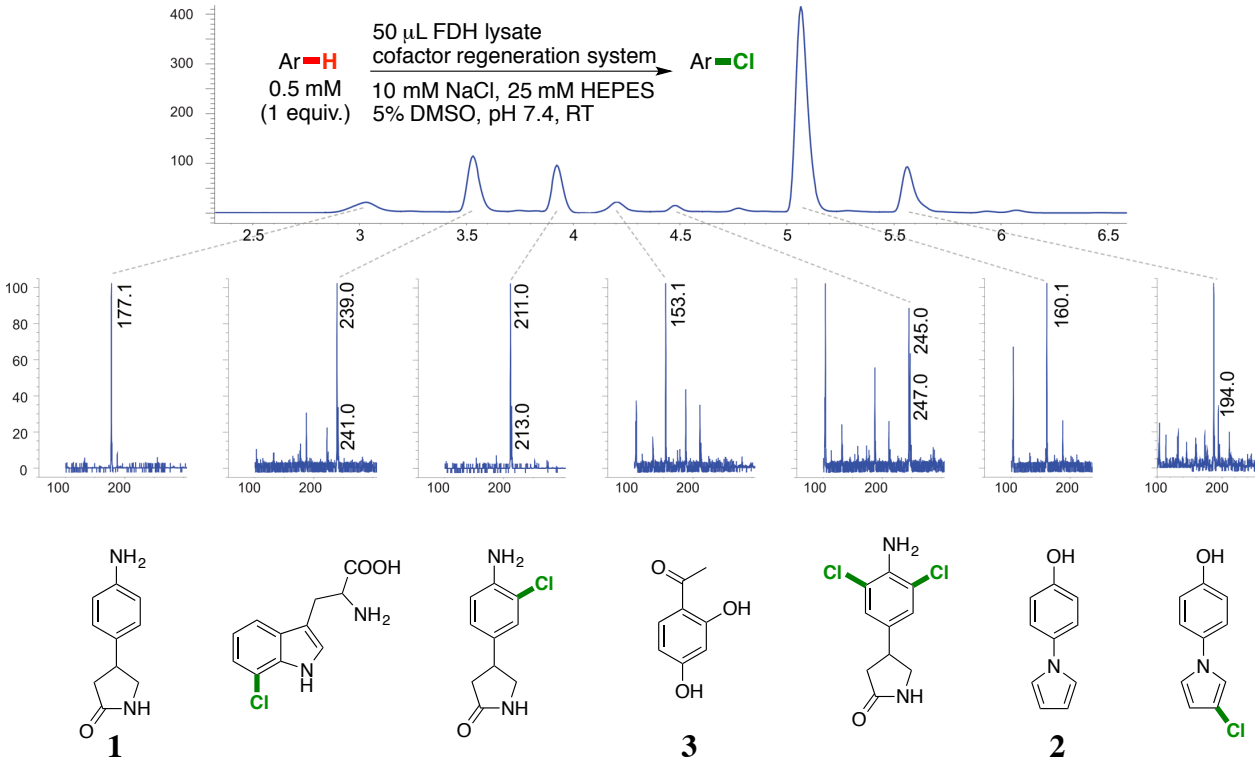


6.2.2 Initial results from halogenase library

Upon receiving a library of 123 putative FDHs from the JGI, we developed a UPLC method to assess chlorination activity. The halogenase genes were expressed and the cells were lysed in 96-well plates according to the standard procedure developed for RebH (chapter 2).[8] From our FDH substrate activity profiling efforts in collaboration with Novartis Institutes for Biomedical Research (NIBR), we had previously found that FDHs can have very different, complementary substrate scopes.[9] We do not know what the native substrates are for any of these putative FDHs. For these reasons, we did not want to limit our initial assessment to only one substrate and thus conducted analytical bioconversions with four substrates simultaneously (Fig. 6.3). L-tryptophan was included as one of these substrates to readily identify Trp-FDHs in our library. The aniline substrate **1** was halogenated in high conversion with all Trp-FDHs tested at

NIBR, even those that do not readily halogenate L-tryptophan,[9] and was therefore included as a general probe substrate in this screen. Substrate **2**, also halogenated by most FDHs during substrate activity profiling, was included as a general FDH probe because it contained two types of electron-rich aromatic rings – a phenol and pyrrole. Lastly, the dihydroxy-acetophenone **3** was included as another probe phenol compound. Control RebH-catalyzed chlorination reactions containing these four substrates showed good conversion of L-tryptophan, **1** and **2** by LCMS (Fig. 6.3).

Figure 6.3: LCMS method using in initial assessment of putative FDH library.[a]

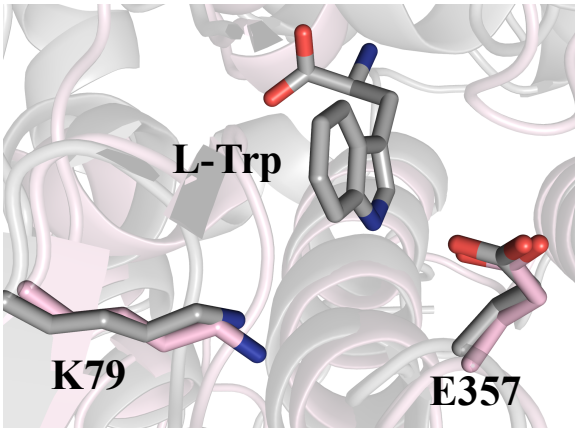


[a] 50 μL RebH lysate, 0.5 mM L-trp, 0.5 mM **1**, 0.5 mM **2**, 0.5 mM **3**, 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10 mM NaCl, 20 mM glucose, 100 μM NAD and FAD, 5% v/v DMSO/25 mM HEPES buffer pH 7.4, 25 °C. Analyzed by LCMS method 1 in the general procedures.

One hit was found from initial assessment of this library for chlorination of the four substrates shown in Fig. 6.3. Surprisingly, this hit, currently termed #65, readily chlorinated

substrates **1** and **2**, but not L-tryptophan. The identification of these probe substrates from substrate activity profiling for discovery of new FDHs was crucial to observing chlorination with #65. When I expressed and purified #65 on a larger scale (750 mL in shake flasks instead of 1 mL in plates), high yields of pure protein were obtained. Using pure protein, I retested L-tryptophan as a substrate for #65-catalyzed chlorination. Again, no conversion could be detected by LCMS. This is especially interesting to us because of the striking similarity between characterized Trp-FDHs and #65. Located within the "indole" cluster (Fig. 6.1), #65 has highest sequence identity to Trp-FDHs out of the characterized FDHs (33% identity to RebH). When a homology model is built using Phyre[2] server,[10] the active sites of RebH and #65 show good alignment (Fig. 6.4). The catalytically crucial lysine and glutamine residues found in all known Trp-FDHs are both present in #65.

Figure 6.4: Homology model of FDH #65 aligned with RebH crystal structure.[a]

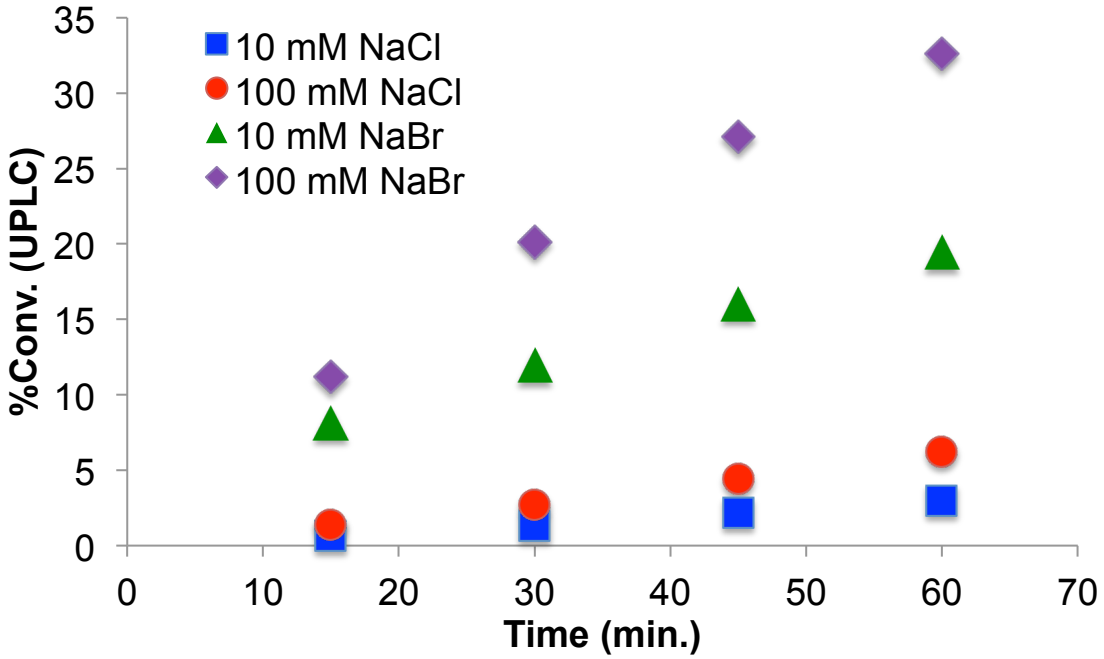


[a] RebH crystal structure (PDB ID: 2OA1)[11] and Phyre[2]-generated[10] #65 homology model are shown in grey and pink, respectively. Residue numbering corresponds to RebH.

Although #65 catalyzes chlorination, this does not necessarily mean that its native function is chlorination, given that several FDHs have been found to halogenate using more than one halide.[12-15] To test whether #65 catalyzes bromination or iodination, I conducted reactions

with varying concentrations of NaBr and NaI using substrate **2**. No iodination activity was observed; however, bromination of **2** occurred at significantly faster rates than chlorination (Fig. 6.5), perhaps (but not necessarily) indicating #65 acts as a brominase natively. #65 is from the bacterial species *Idiomarina atlantica*, which was first isolated from deep sea sediment in the North Atlantic Ocean in 2015.[16] Many brominated marine natural products have been discovered;[17] however, none are currently known to be from *I. atlantica*.

Figure 6.5: FDH #65 demonstrates higher bromination activity than chlorination.[a]



[a] 0.5 mM **2**, 25 μM #65, 2.5 μM MBP-RebF, 9 U mL^{-1} GDH, 10-100 mM NaCl or NaBr, 20 mM glucose, 100 μM NAD and FAD, 1.7% v/v DMSO/25 mM HEPES buffer pH 7.4, 25 °C. Time points collected from 15–60 minutes.

Upon analysis of other putative FDHs from our genome mining library, it appears that many of these enzymes are expressing insolubly. This could be a reason that only one hit was identified with the initial screen. Harrison Snodgrass, a graduate student in the Lewis lab, is currently investigating different expression conditions, such as expression in 24-well plates instead of 96 and lower expression temperatures. We hope to further characterize the novel

brominase #65 as well as use further expression and screening techniques to identity other FDHs in the remaining library.

6.3 Conclusions

The dramatically increased number of sequenced genomes has led to thousands of putative flavin-dependent halogenases found in current sequencing databases. In addition to directed evolution, we envisioned mining of these deposited but not yet characterized sequences as a tool for discovering new FDHs. We partnered with the Joint Genome Institute and obtained a maximally diverse library of 123 putative FDH genes. These were screened using four general probe substrates, and novel FDH #65 was identified. FDH #65 has higher bromination activity than chlorination activity, and the substrate specificity differs from that of previously characterized Trp-FDHs. The remaining 122 FDHs in this library are currently being tested for better expression.

6.4 Experimental

6.4.1 General experimental procedures

Materials:

Plasmids pET-28a/RebF and pET-28a/RebH in BL-21 DE3 E. coli were provided by the Walsh group of Harvard Medical School, Boston, MA. The pLIC-MBP plasmid was provided by the Bottomley group of Monash University, Clayton, Australia. The pGro7 plasmid encoding the groES and groEL chaperone set was purchased from Takara (Otsu, Shiga, Japan). BL21(DE3) E. coli cells were purchased from Invitrogen (Carlsbad, CA). Luria broth (LB) and Terrific broth (TB) media were purchased from Research Products International (Mt. Prospect, IL). Qiagen

Miniprep Kits were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. All genes were confirmed by sequencing by the Joint Genome Institute.

General Procedures:

Reactions were analyzed using an Agilent Technologies 1200 UHPLC or Agilent Technologies 6130 LC-MS. All mass spectra were collected at the University of Chicago Mass Spectrometry Service Center. Protein concentrations were measured using the Pierce® BCA Protein Assay Kit.

UPLC/LC-MS methods:

LCMS Method 1: Phenomenex Kinetex EVO C18 2.1 x 50 mm column, 2.6 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN; 0-5 min, B = 5-100%; 5-7 min, B = 100%. Absorbance at 254 and 280 nm was measured.

UPLC Method 2: Agilent Eclipse Plus C18 4.6 x 150 mm column, 3.5 μM particle size; solvent A = H_2O/0.1% TFA, solvent B = CH_3CN; 0-10 min, B = 15%; 10-17 min, B = 15-80%; 17-20 min, B = 80%. Absorbance at 280 nm was measured.

6.4.2 Specific experimental procedures

Library expression and lysis: Glycerol stocks for all putative FDH genes in pGro7-containing BL21(DE3) cells, in 96-well microtiter plate format, were stamped into 1 mL 96-well plates containing 300 μL LB with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol. Cells were

grown overnight (14-15 hours) at 37 °C, 200 rpm, and 100 µL of overnight culture were used to inoculate 1 mL TB (with 50 µg/mL kanamycin and 20 µg/mL chloramphenicol) in 2-mL 96-well plates. Following growth at 37 °C, 250 rpm, to an OD_{600} = 0.8-1, enzyme expression was induced with IPTG and arabinose (added using a multichannel pipette) to final concentrations of 10 µM and 0.2 mg/mL, respectively. Protein expression continued for ~20 h at 30 °C, 250 rpm, after which cells were pelleted by centrifugation and the supernatants were discarded.

Cell pellets were suspended in 100 µL HEPES buffer (25 mM, pH 7.4) containing 0.75 mg/mL lysozyme. After incubation at 37 °C, (250 rpm, for 30 min.) cells were flash frozen in liquid nitrogen and thawed in a ~30 °C water bath. DNaseI (10 µL of 1 mg/mL, 25 mM HEPES buffer, pH 7.4) was added, and the cells were incubated at 37 °C, 250 rpm, for 15 min. After centrifugation, 50 µL of supernatant containing cell lysates was transferred to a microtiter plate for screening.

Library screening: MBP-RebF (0.005 equiv., 2.5 µM final concentration) and glucose dehydrogenase (9 U/mL final concentration) were added as solutions (25 mM HEPES, pH 7.4) to the FDH lysate. A solution containing L-tryptophan (1 equiv., 0.5 mM final concentration), **1** (1 equiv., 0.5 mM final concentration), **2** (1 equiv., 0.5 mM final concentration), **3** (1 equiv., 0.5 mM final concentration), NAD (0.2 equiv., 100 µM final concentration), FAD (0.2 equiv., 100 µM final concentration), NaCl (20 equiv., 10 mM final concentration), and glucose (40 equiv., 20 mM final concentration) was added via multichannel pipette to simultaneously initiate the reactions (final reaction volume of 75 µL). The microtiter plates were sealed (ALPS 3000, Thermo Scientific) and were agitated at 650 rpm on an Eppendorf dry block incubator for 15-16 hours. Reactions were quenched with 75 µL of methanol. Precipitated protein was then pelleted

by centrifugation, and the supernatant was transferred to a 96-well 0.45 μm filter plate. The filter plate was loaded on top of a new 96-well microtiter plate and centrifuged at 2,453 g for 10 minutes. Reactions were analyzed by LCMS Method 1 in the general procedures.

Large-scale Expression of FDH #65: A primary culture of pET-28a/#65 & pGro7 BL21(DE3) *E. coli* were grown (7.5 mL LB containing 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin in 15 mL Falcon culture tubes) from the glycerol stock described above at 37 °C and 250 rpm overnight. This primary culture were used to inoculate 750 mL of TB containing 0.02 mg/mL chloramphenicol and 0.05 mg/mL kanamycin in a 2.8 L, non-beveled glass flask. The cultures were then grown at 37 °C and 250 rpm to OD_{600} = 0.8, at which point isopropyl β-D-1-thiogalactopyranoside (IPTG) and L(+)-arabinose were added to final concentrations of 100 μM and 2 mg/mL, respectively. The temperature was then reduced to 30 °C and the shake rate left at 250 rpm for an additional 20 hours, at which time the cells were pelleted by centrifugation (4,000g, 30 min, 4 °C). The pellet was each resuspended in 30 mL of HEPES (25 mM, pH 7.4), transferred to a 50 mL conical tube, and lysed by sonication.

FDH #65 Sonication Conditions: Sonication was performed on a Qsonica S-4000 Sonicator with a 0.5" horn using the following procedure: 5 x 1 min with 1 min rests, 20% duty cycle delivering 40-50 W. To keep the sample from over-heating, the conical tube was submerged was submerged in a circulating ice-water bath.

Purification of FDH #65: Following sonication, the lysate was then clarified by centrifugation (24,000g, 30 min, 4 °C). The clarified lysate was purified by Ni-NTA affinity chromatography.

The elution fractions were concentrated to 3 mL using an Amicon® 15 mL 30 kD cutoff centrifugal filter and exchanged three times into HEPES/glycerol buffer (25 mM, pH 7.5, 10% v/v glycerol).

Halogenation reactions using purified enzyme: Purified #65, (0.05 equiv., 25 μM final concentration), MBP-RebF (0.005 equiv., 2.5 μM final concentration) and glucose dehydrogenase (9 U/mL final concentration) were added as solutions (25 mM HEPES, pH 7.4) to eppendorf tubes containing 25 mM HEPES buffer. A solution containing L-tryptophan (1 equiv., 0.5 mM final concentration) or **2** (1 equiv., 0.5 mM final concentration), NAD (0.2 equiv., 100 μM final concentration), FAD (0.2 equiv., 100 μM final concentration), NaCl or NaBr (20 or 200 equiv., 10 or 100 mM final concentration), and glucose (40 equiv., 20 mM final concentration) was added to initiate the reactions (final reaction volume of 75 μL). The tubes were agitated at 650 rpm in an eppendorf dry block incubator. Reactions containing L-tryptophan proceeded overnight and were subsequently quenched with 75 μL of methanol, while reactions containing substrate **2** were quenched every 15 minutes for 60 minutes. Precipitated protein was then pelleted by centrifugation, and the supernatant was transferred to a 96-well 0.45 μm filter plate. The filter plate was loaded on top of a new 96-well microtiter plate and centrifuged at 2,453 g for 10 minutes. Reactions were analyzed by UPLC Method 2 in the general procedures.

6.5 References

1. Hess, M. *et al*. Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* **331,** 463–467 (2011).
2. Jensen, P. R., Chavarria, K. L., Fenical, W., Moore, B. S. & Ziemert, N. Challenges and triumphs to genomics-based natural product discovery. *J Ind Microbiol Biotechnol* **41,** 203–209 (2013).
3. Padhi, S. K. Modern Approaches to Discovering New Hydroxynitrile Lyases for Biocatalysis. *ChemBioChem* **18,** 152–160 (2016).
4. Roumpeka, D. D., Wallace, R. J. & Escalettes, F. A review of bioinformatics tools for bio-prospecting from metagenomic sequence data. *Front. Microbiol*. (2017). doi:10.3389/fgene.2017.00023/full#
5. Sunagawa, S. *et al*. Structure and function of the global ocean microbiome. *Science* **348,** 1261359–1261359 (2015).
6. Menon, B. R. K. *et al*. Structure and biocatalytic scope of thermophilic flavin-dependent halogenase and flavin reductase enzymes. *Org. Biomol. Chem*. **14,** 9354–9361 (2016).
7. Gerlt, J. A. *et al*. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* **1854,** 1019–1037 (2015).
8. Poor, C. B., Andorfer, M. C. & Lewis, J. C. Improving the Stability and Catalyst Lifetime of the Halogenase RebH By Directed Evolution. *ChemBioChem* **15,** 1286–1289 (2014).
9. Andorfer, M. C., Grob, J. E., Hajdin, C. E. & Chael, J. R. Understanding Flavin-Dependent Halogenase Reactivity via Substrate Activity Profiling. *ACS Catal*. **7,** 1897–1904 (2017).
10. Mezulis, S., Yates, C. M., Wass, M. N., Sternberg, M. J. E. & Kelley, L. A. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc*. **10,** 845–858 (2015).
11. Bitto, E. *et al*. The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins* **70,** 289–293 (2007).
12. Payne, J. T., Andorfer, M. C. & Lewis, J. C. Regioselective Arene Halogenation using the FAD-Dependent Halogenase RebH. *Angew. Chem. Int. Ed*. **52,** 5271–5274 (2013).
13. Smith, D. R. M. *et al*. An unusual flavin-dependent halogenase from the metagenome of the marine sponge Theonella swinhoei WA. *ACS Chem. Biol*. acschembio.6b01115–12 (2017). doi:10.1021/acschembio.6b01115
14. Weichold, V., Milbredt, D. & van Pée, K.-H. Specific Enzymatic Halogenation-From the Discovery of Halogenated Enzymes to Their Applications In Vitro and In Vivo. *Angew. Chem. Int. Ed. Engl*. **55,** 6374–6389 (2016).
15. Agarwal, V. *et al*. Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat. Chem. Biol*. **10,** 640–647 (2014).
16. Du, J. *et al*. Idiomarina atlantica sp. nov., a marine bacterium isolated from the deep sea sediment of the North Atlantic Ocean. *Antonie van Leeuwenhoek* **107,** 393–401 (2014).
17. Gribble, G. W. The natural production of organobromine compounds. *Environ Sci Pollut Res Int* **7,** 37–47 (2000).

Appendix I

NMR Spectra for Compounds from Chapter 2

Figure AI.1: ^{1}HNMR spectrum of **1**.

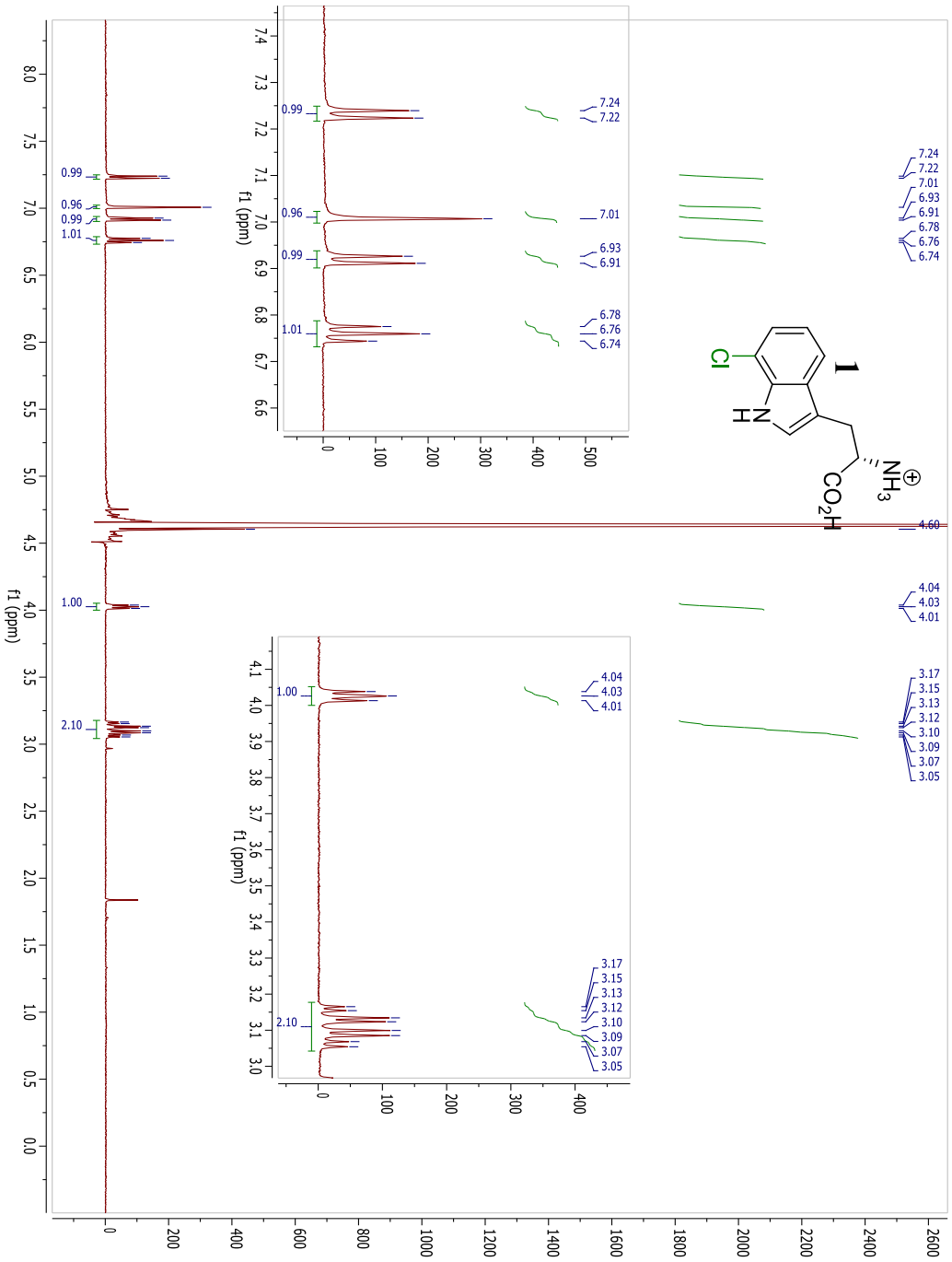


Figure AI.2: ^{13}CNMR spectrum of **1**.

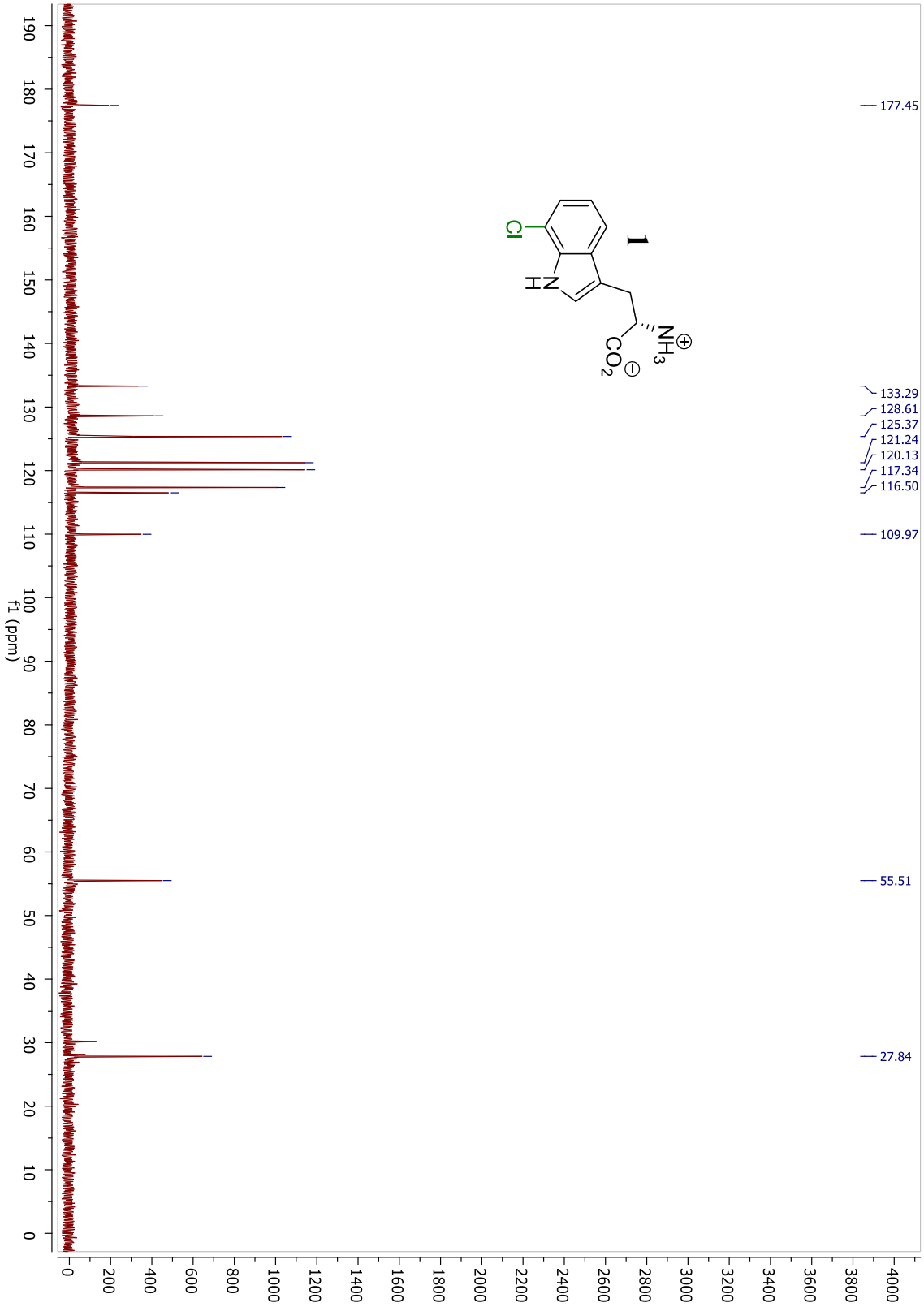
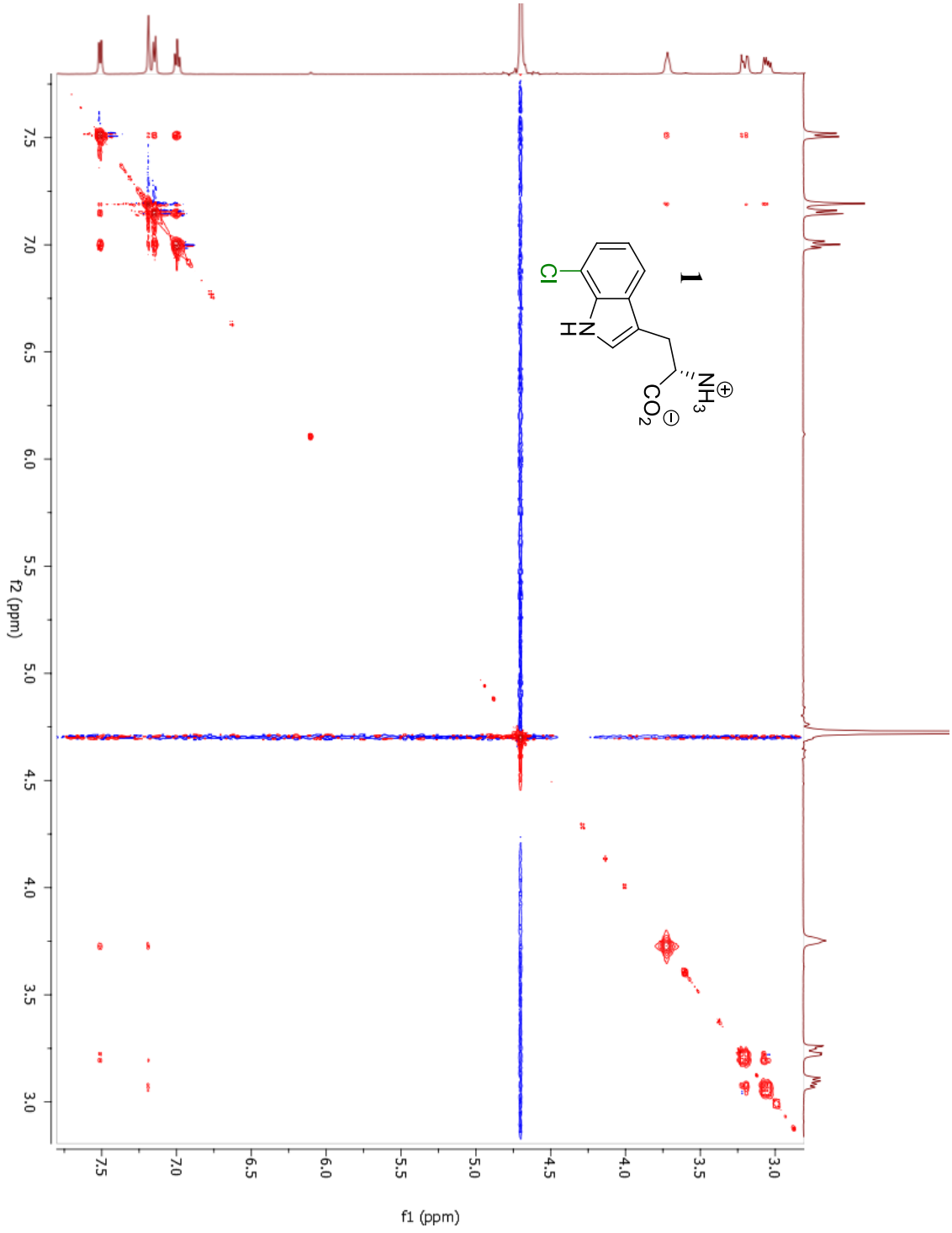


Figure AI.3: NOSEY spectrum of **1**.



NOESY of **1**: Resonances between peaks at 7.19 and 3.72 ppm, and between peaks at 7.51 and 3.72 ppm consistent with chlorination at the 7 position.

Figure AI.4: [1]HNMR spectrum of **2**.

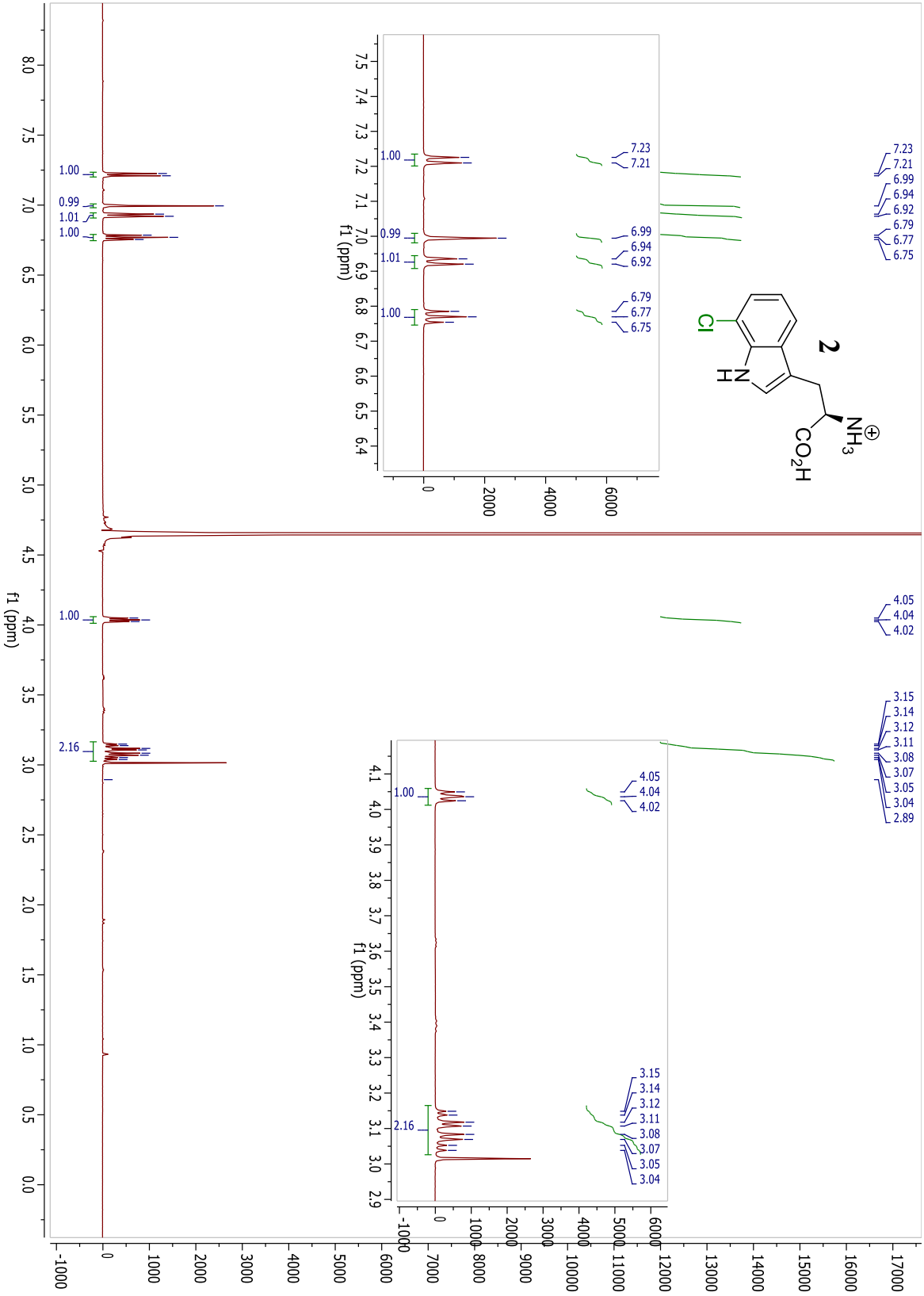


Figure AI.5: ^{13}CNMR spectrum of **2**.

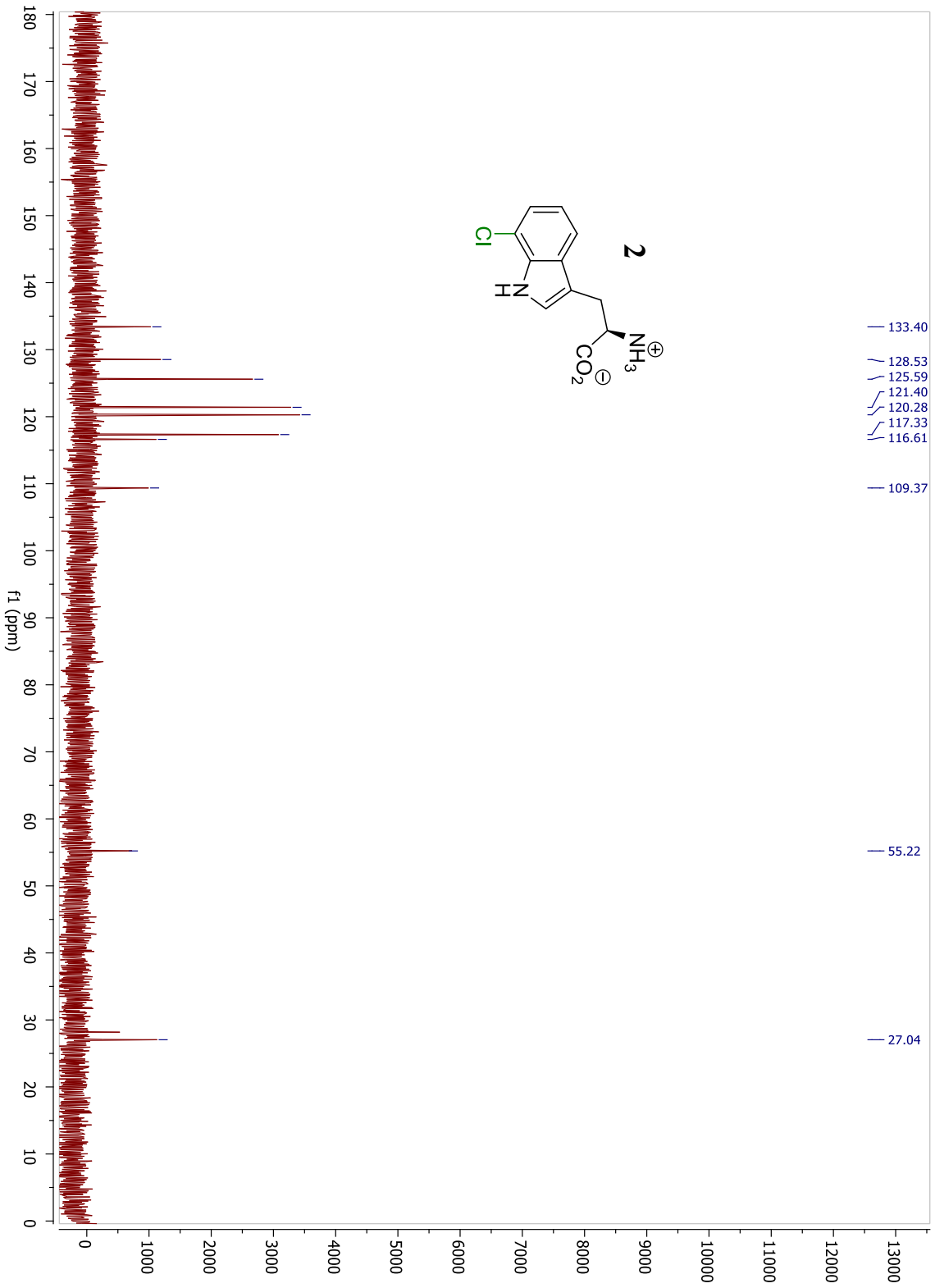
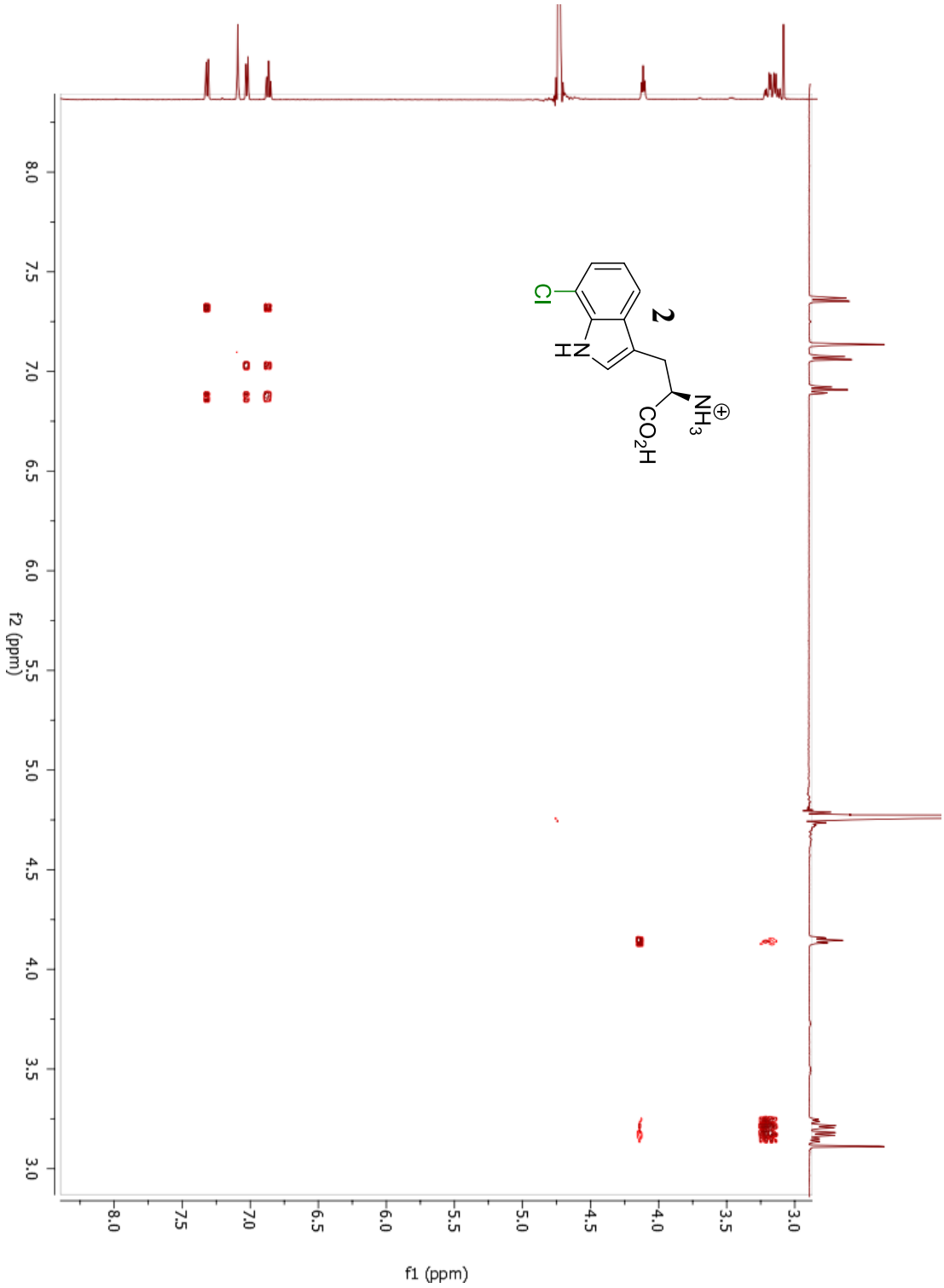
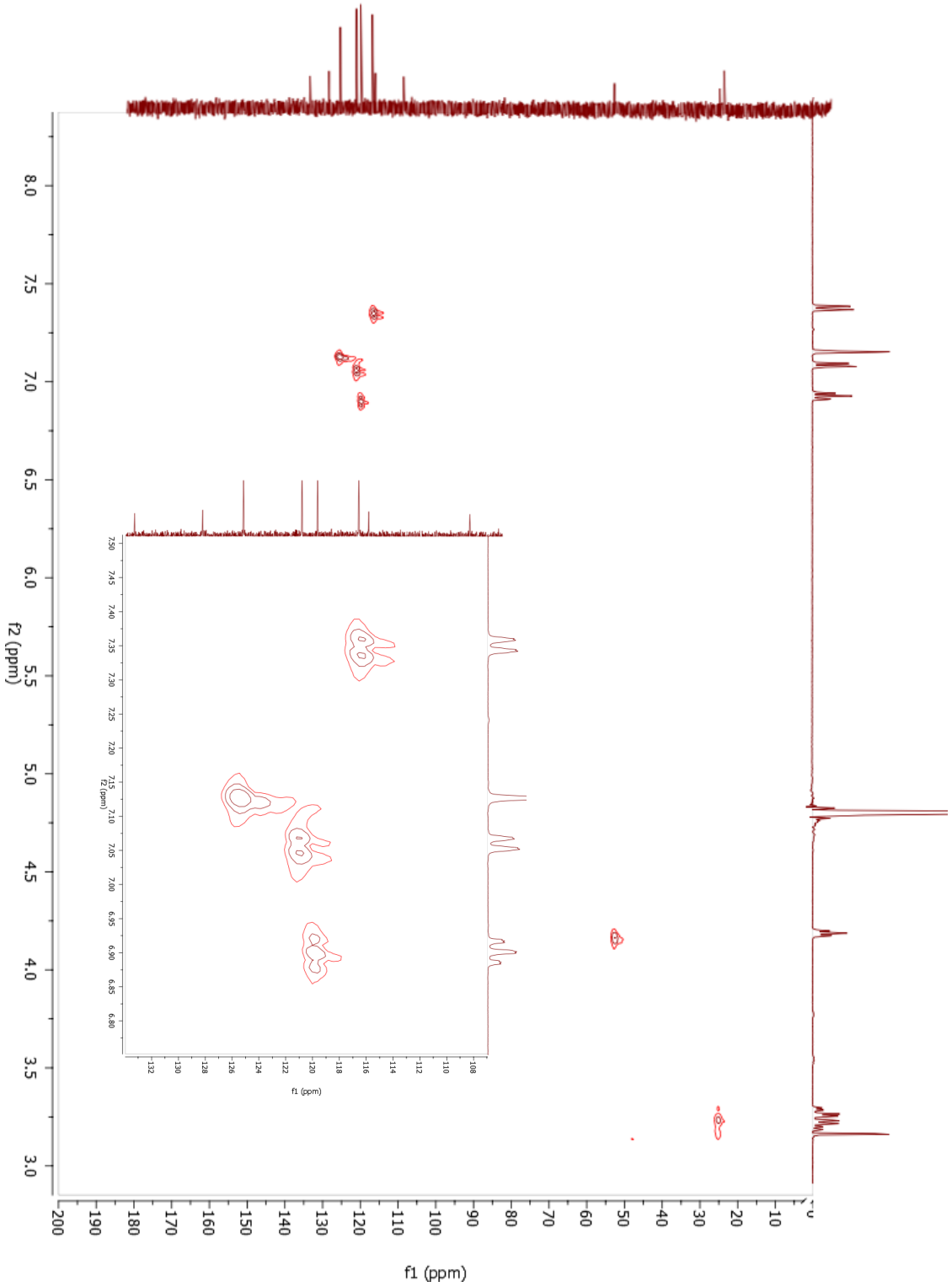


Figure AI.6: COSY spectrum of **2**.



COSY of **2**: Resonances between peaks at 7.22 and 6.77 ppm, and between peaks at 6.93 and 6.7 ppm support chlorination at the 7 position.

Figure AI.7: HMQC spectrum of **2**.



HMQC of **2**: Resonances support chlorination at the 7 position.

Figure AI.8: [1]HNMR spectrum of **3**.

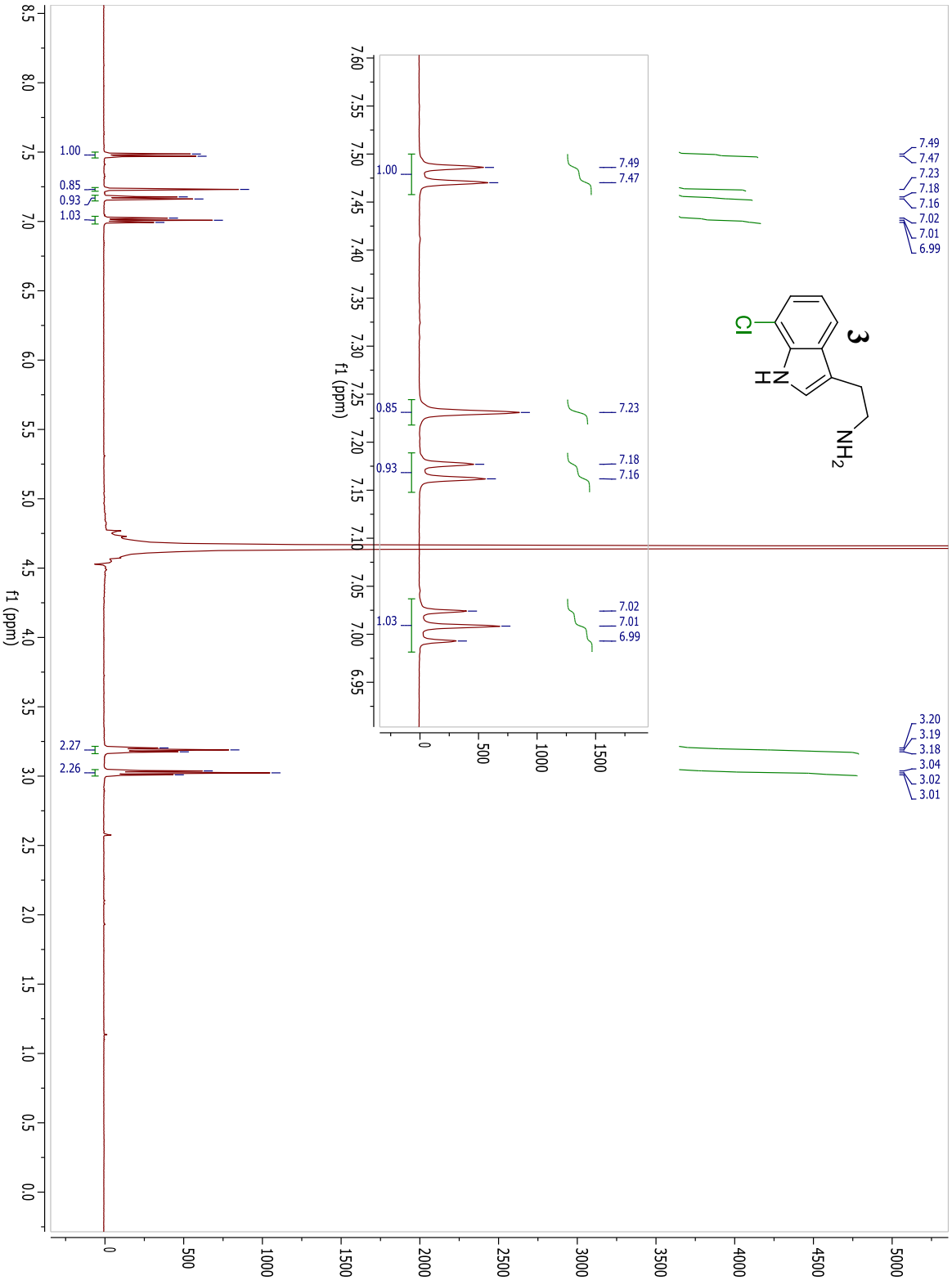


Figure AI.9: ^{13}CNMR spectrum of **3**.

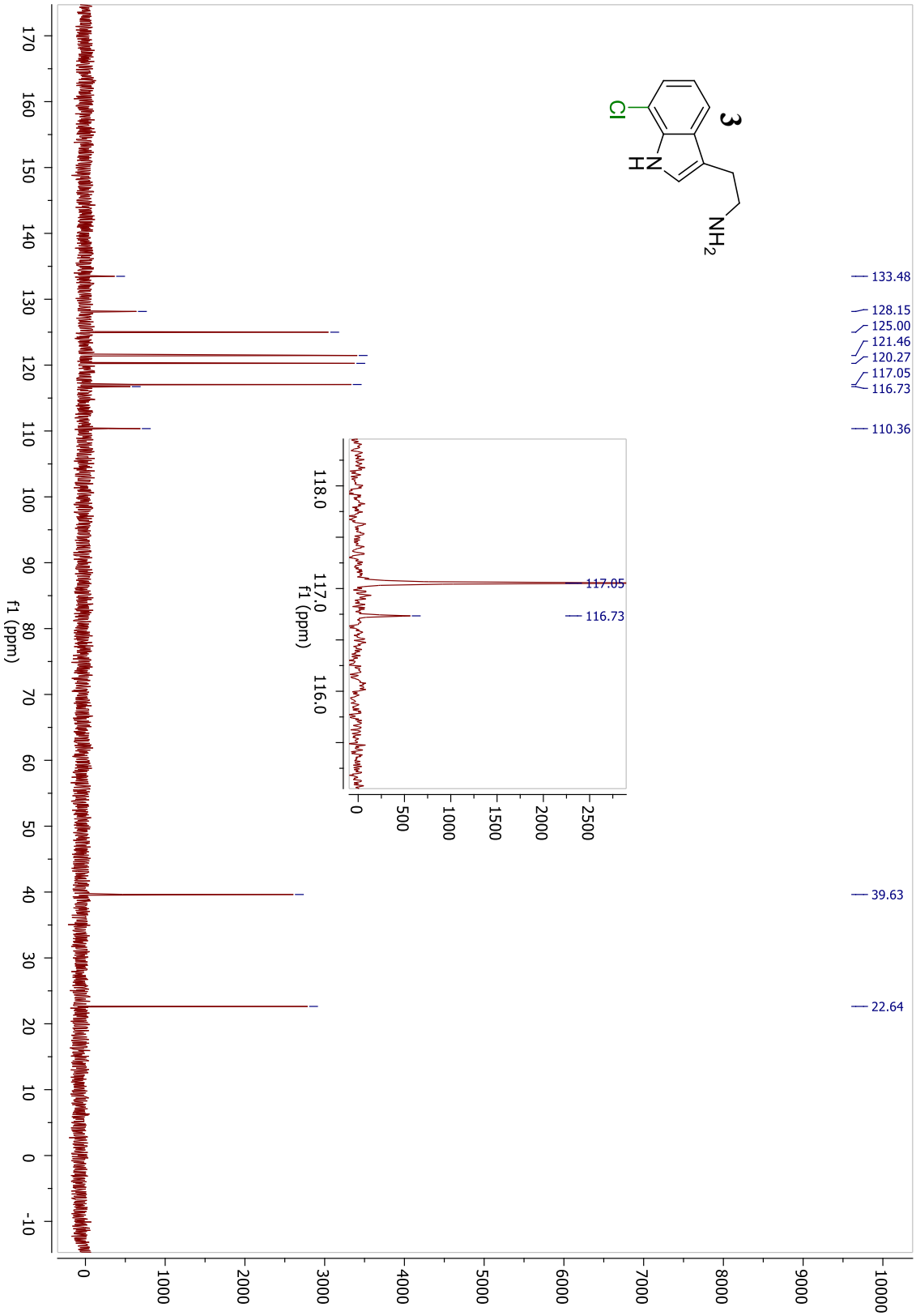
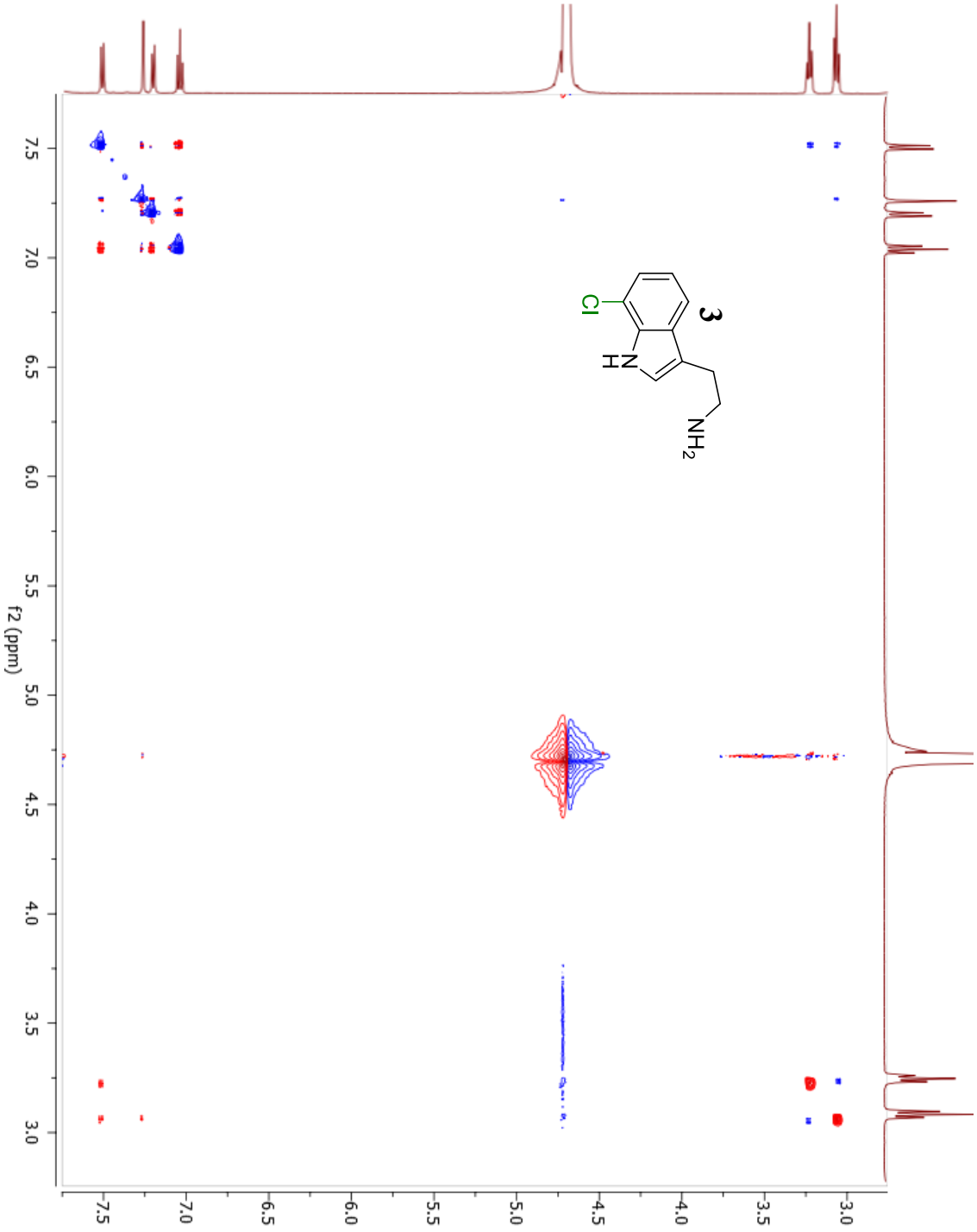


Figure AI.10: NOSEY spectrum of **3**.



NOESY of **3**: Resonances between peaks at 7.48 and 3.02 ppm, and between peaks at 7.23 and 3.02 ppm demonstrate chlorination at the 7 position.

Figure AI.11: [1]HNMR spectrum of **4**.

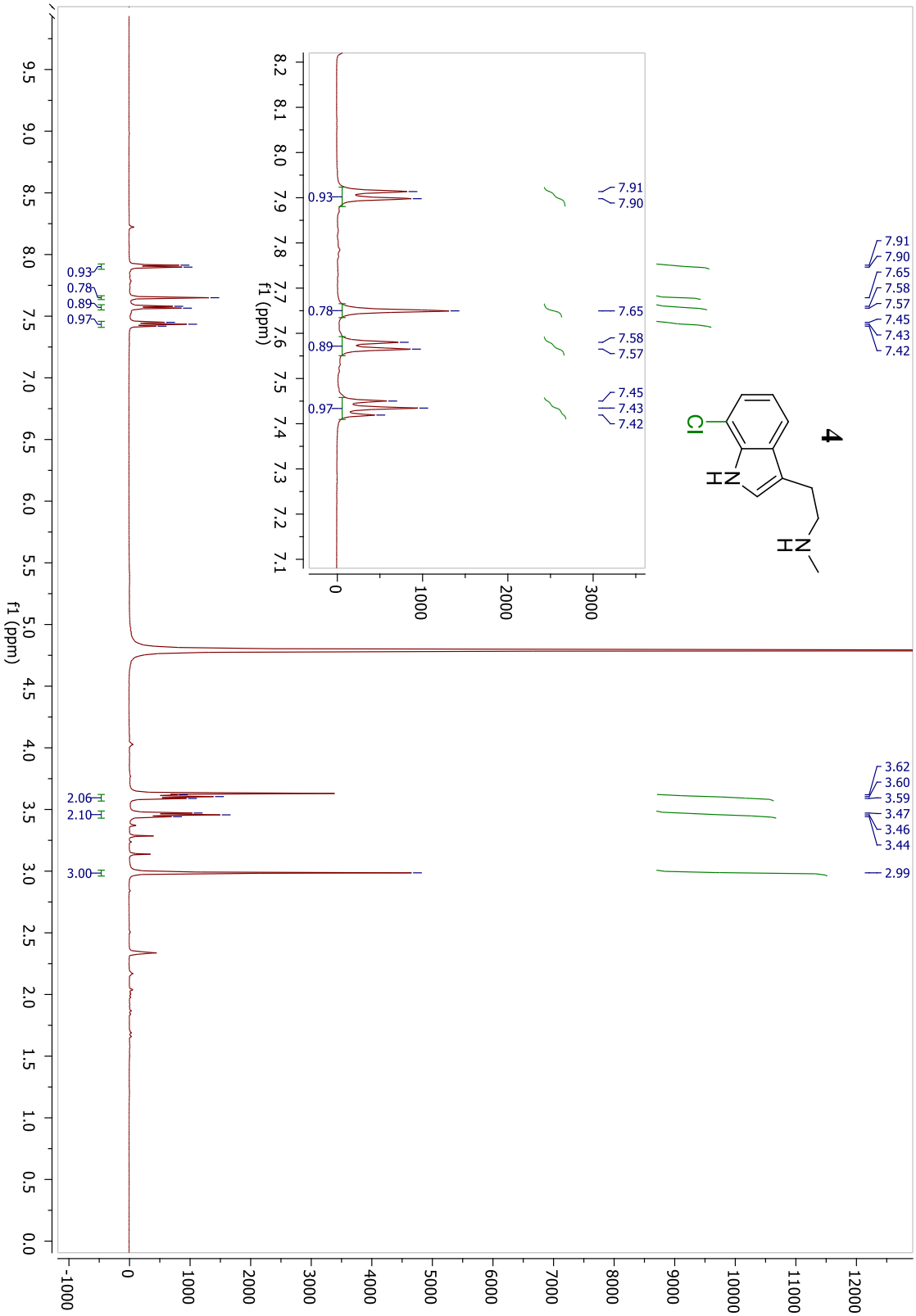


Figure AI.12: ^{13}CNMR spectrum of **4**.

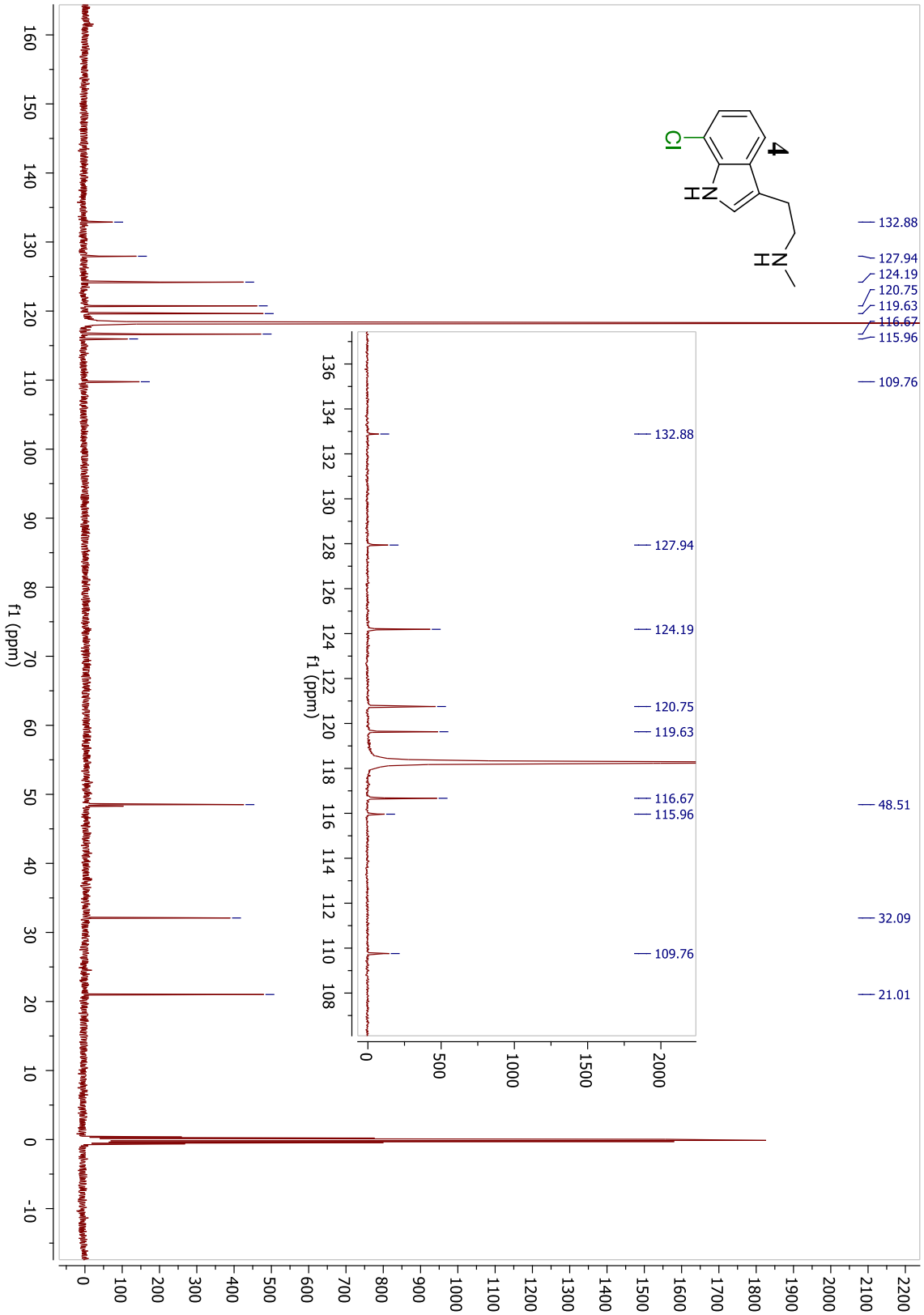
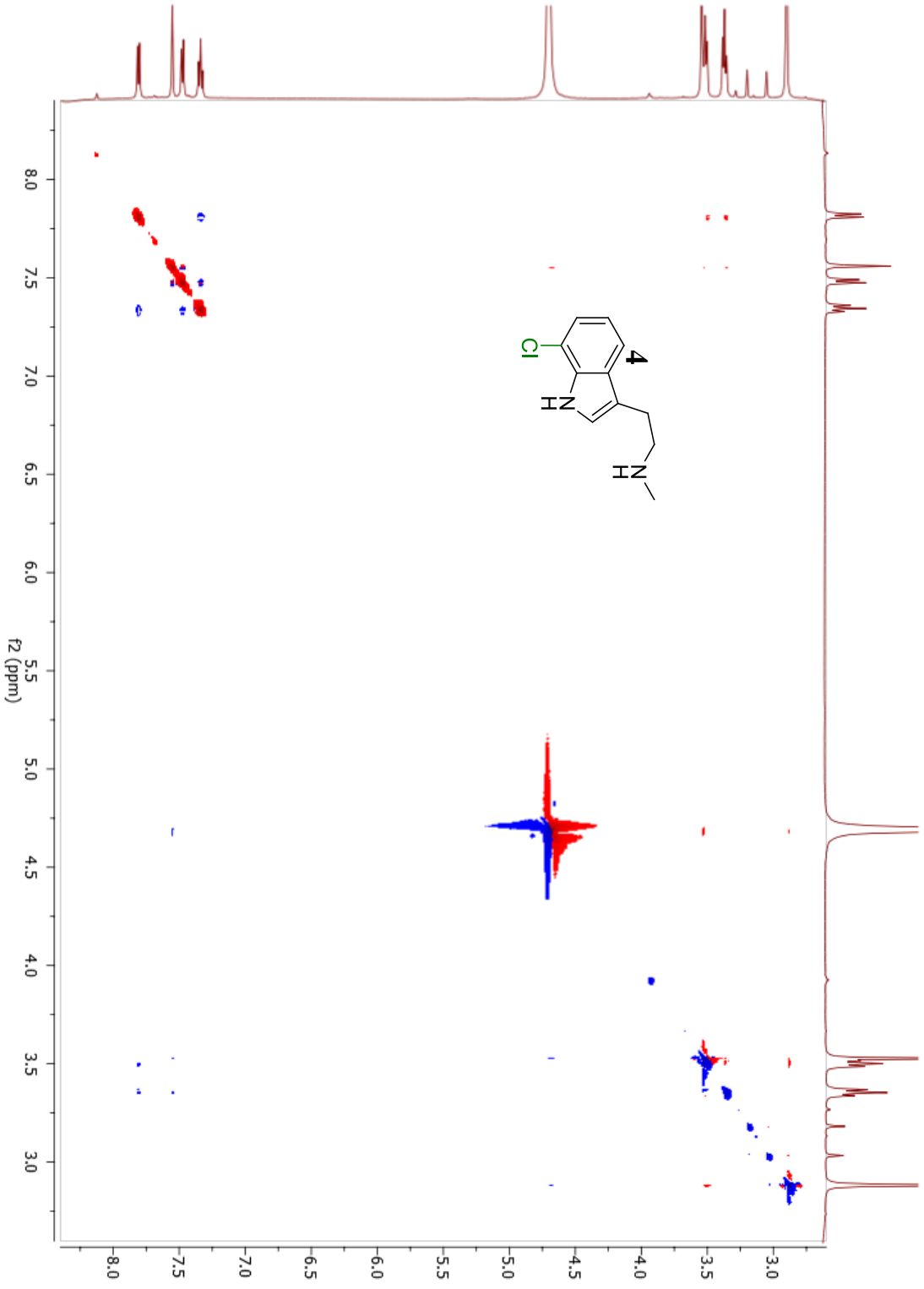


Figure AI.13: NOSEY spectrum of **4**.



NOESY of **4**: Resonances between peaks at 7.90 and 3.46 ppm, and between peaks at 7.65 and 3.46 ppm demonstrate chlorination at the 7 position.

Figure AI.14: [1]HNMR spectrum of **5**.

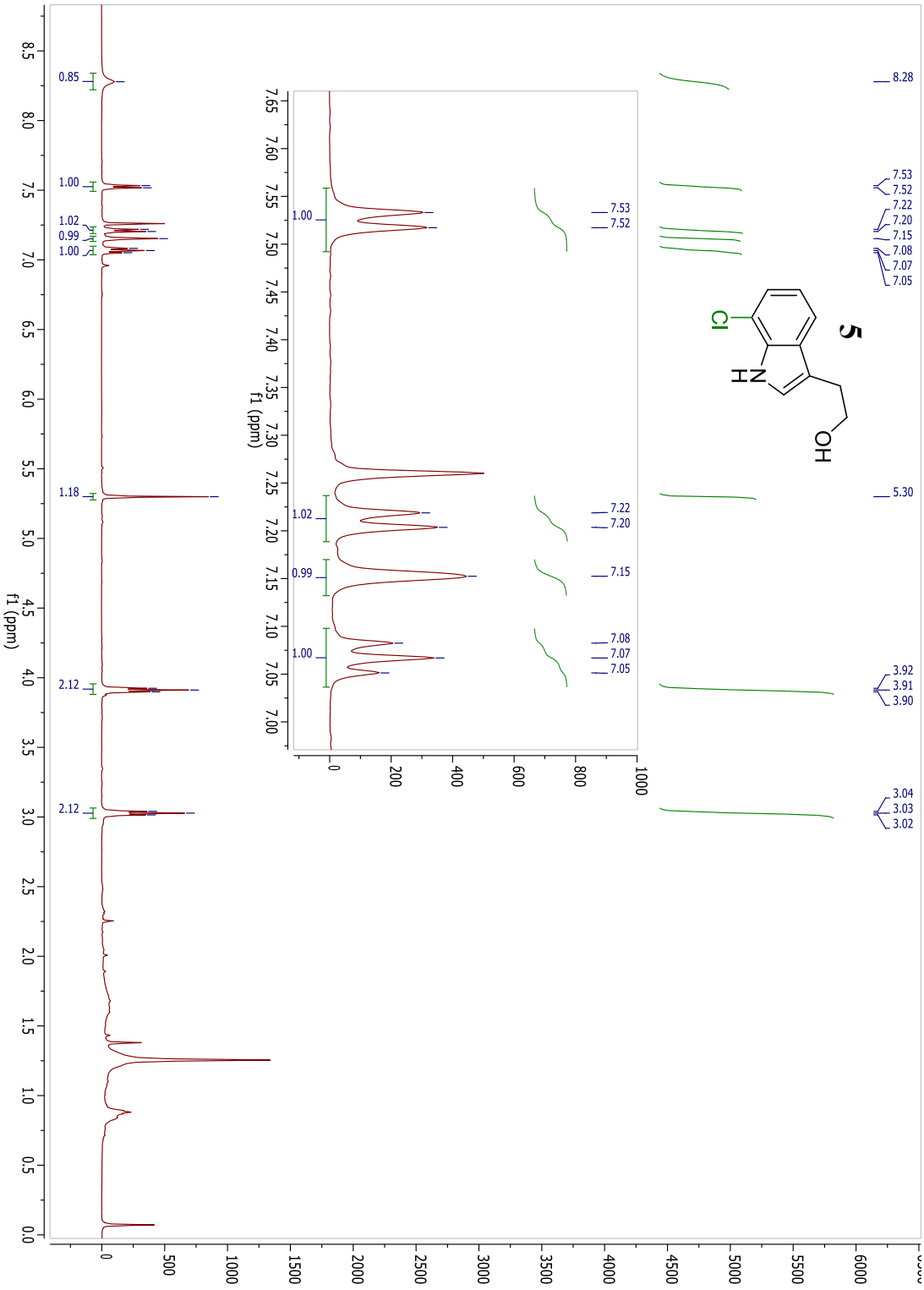


Figure AI.15: ^{13}CNMR spectrum of **5**.

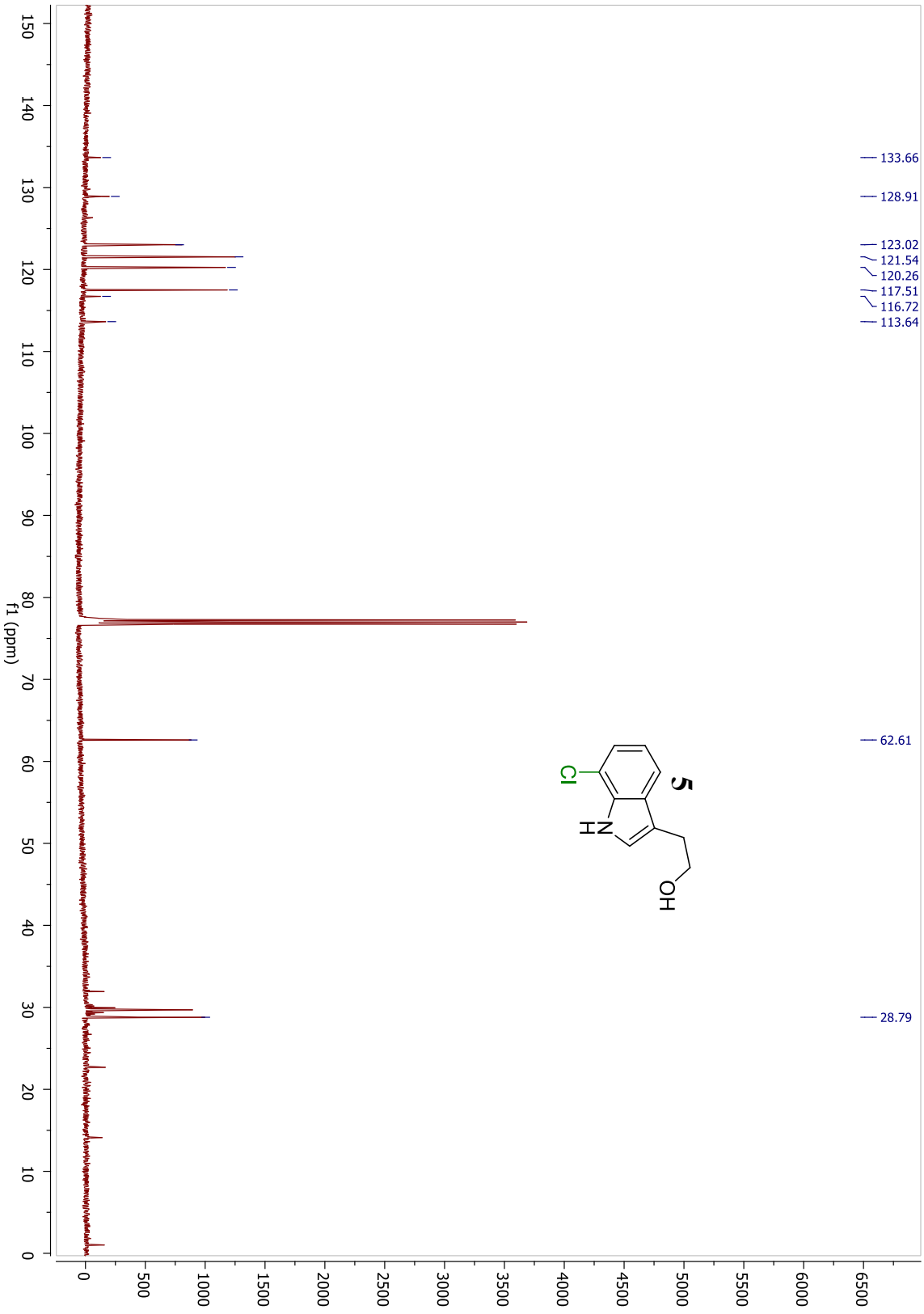
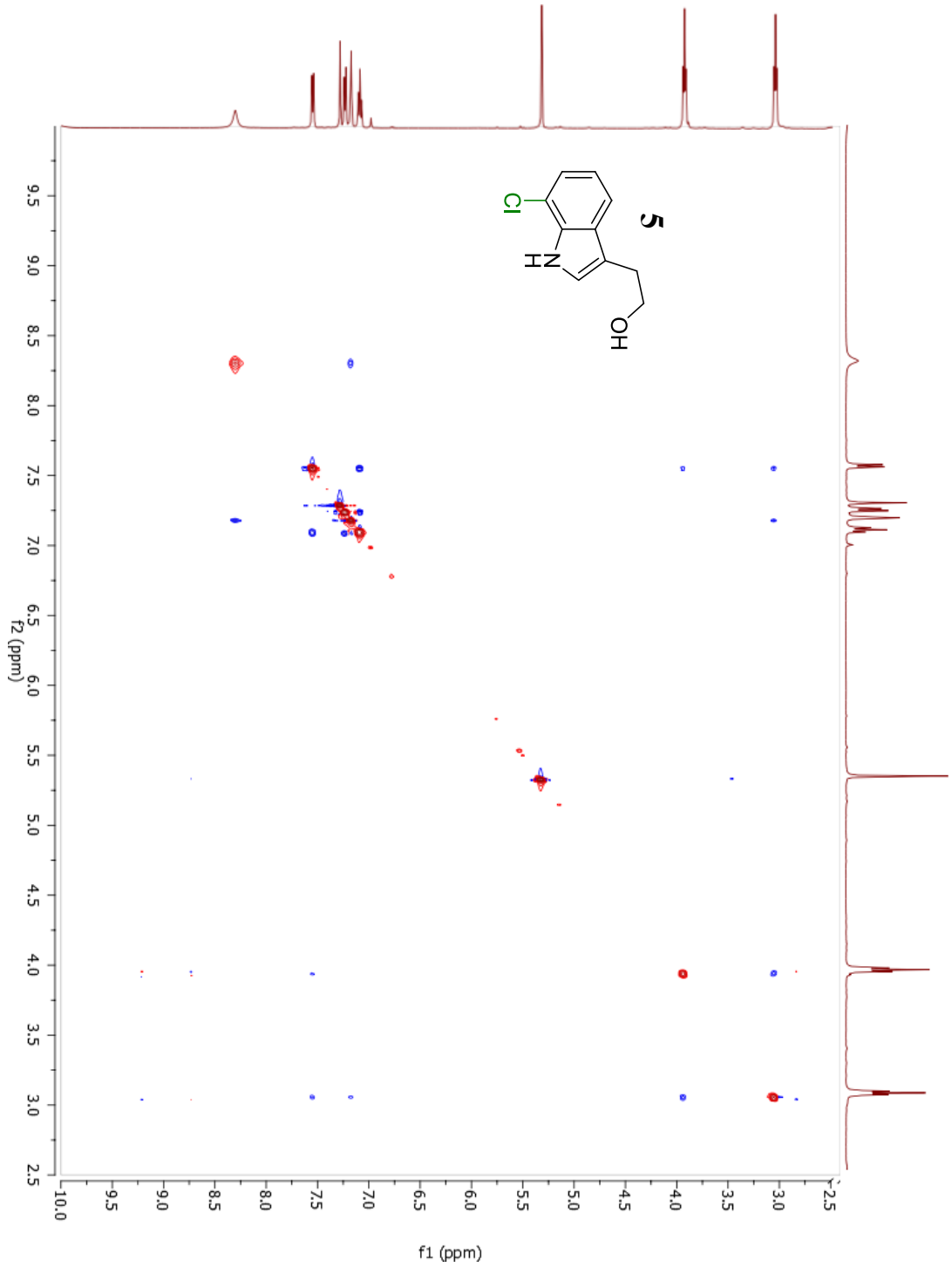


Figure AI.16: NOSEY spectrum of **5**.



NOESY of **5**: Resonances between peaks at 7.53 and 3.03 ppm, and between peaks at 7.15 and 3.03 ppm demonstrate chlorination at the 7 position.

Figure AI.17: [1]HNMR spectrum of **6**.

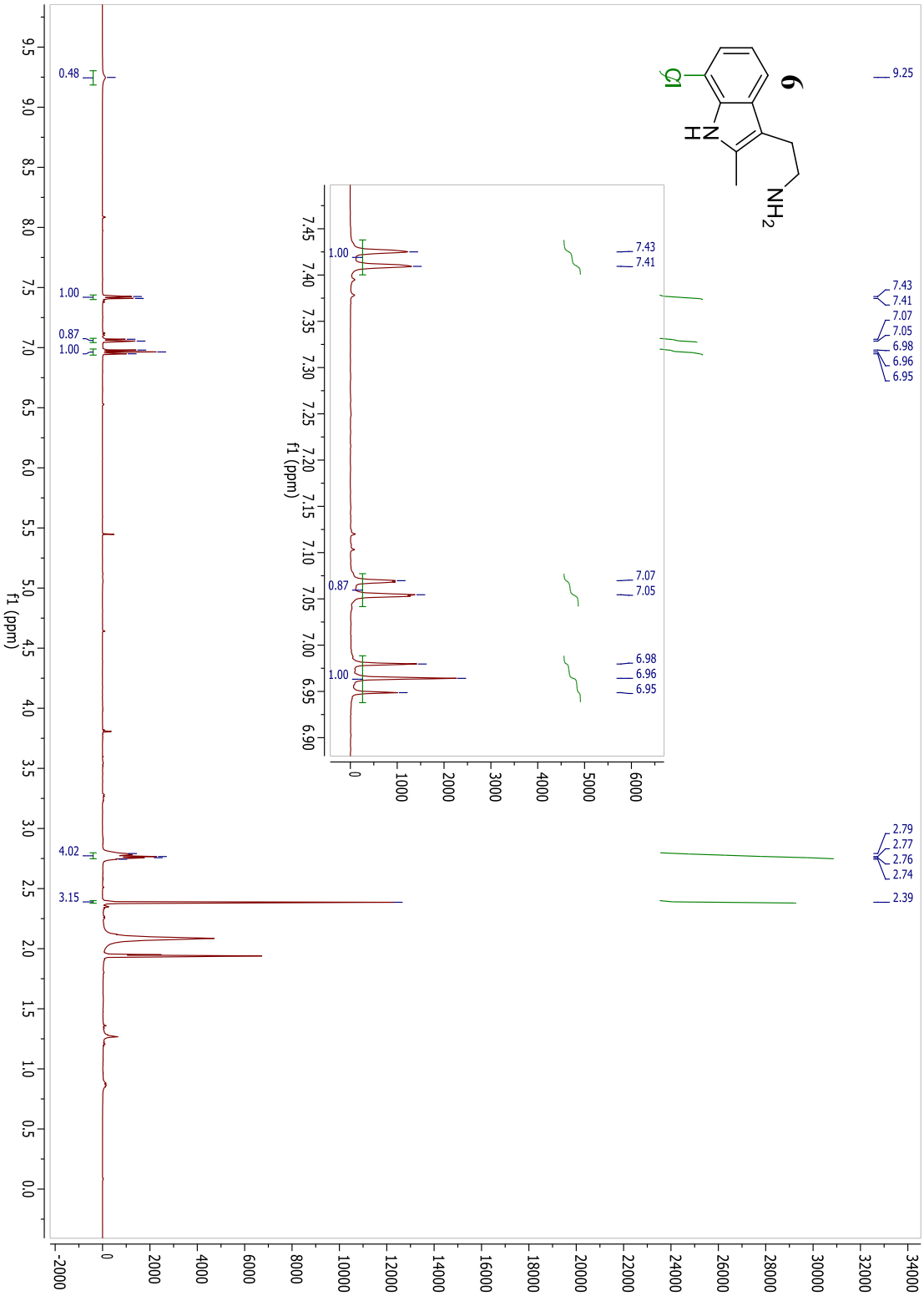


Figure AI.18: ^{13}CNMR spectrum of **6**.

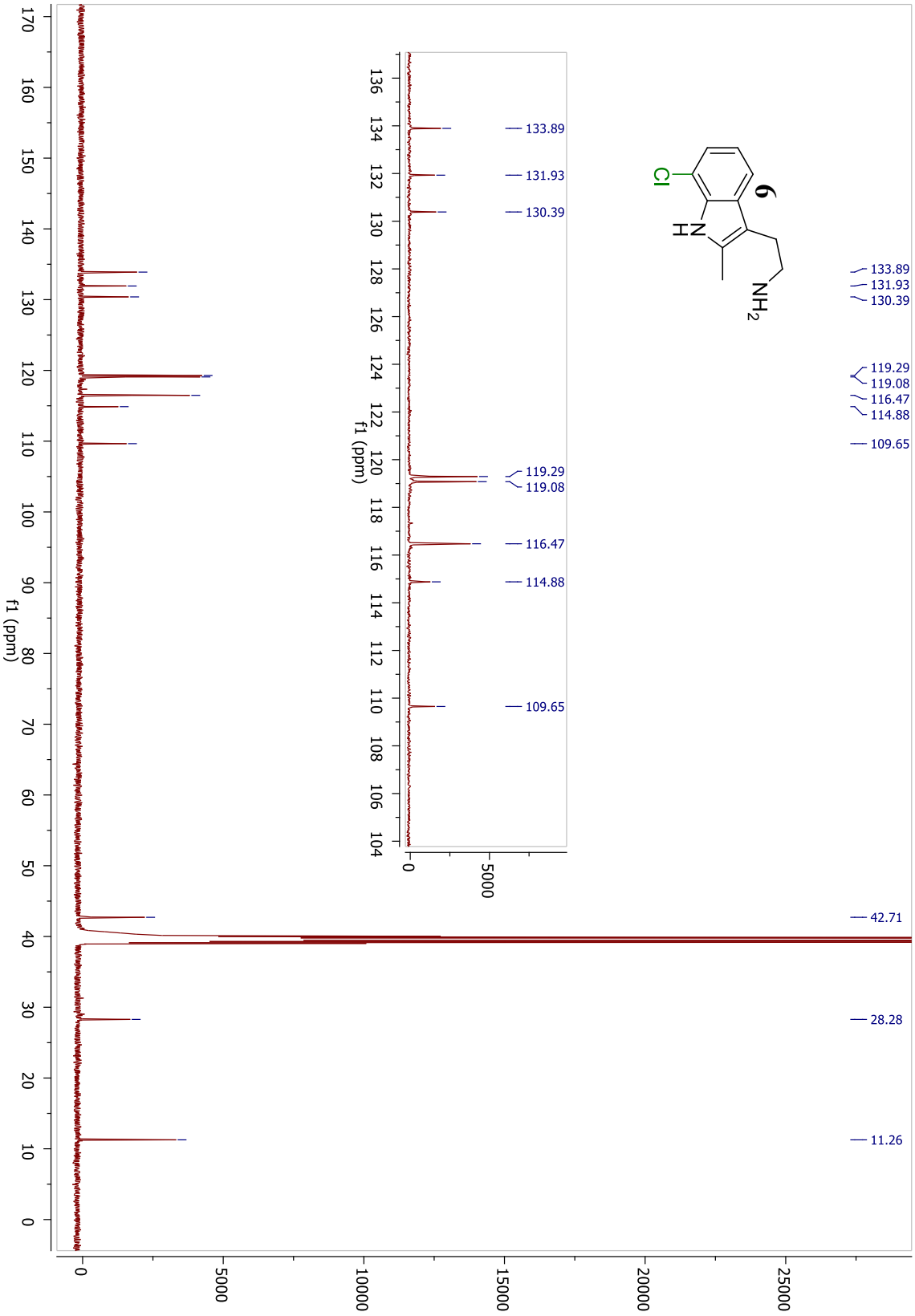
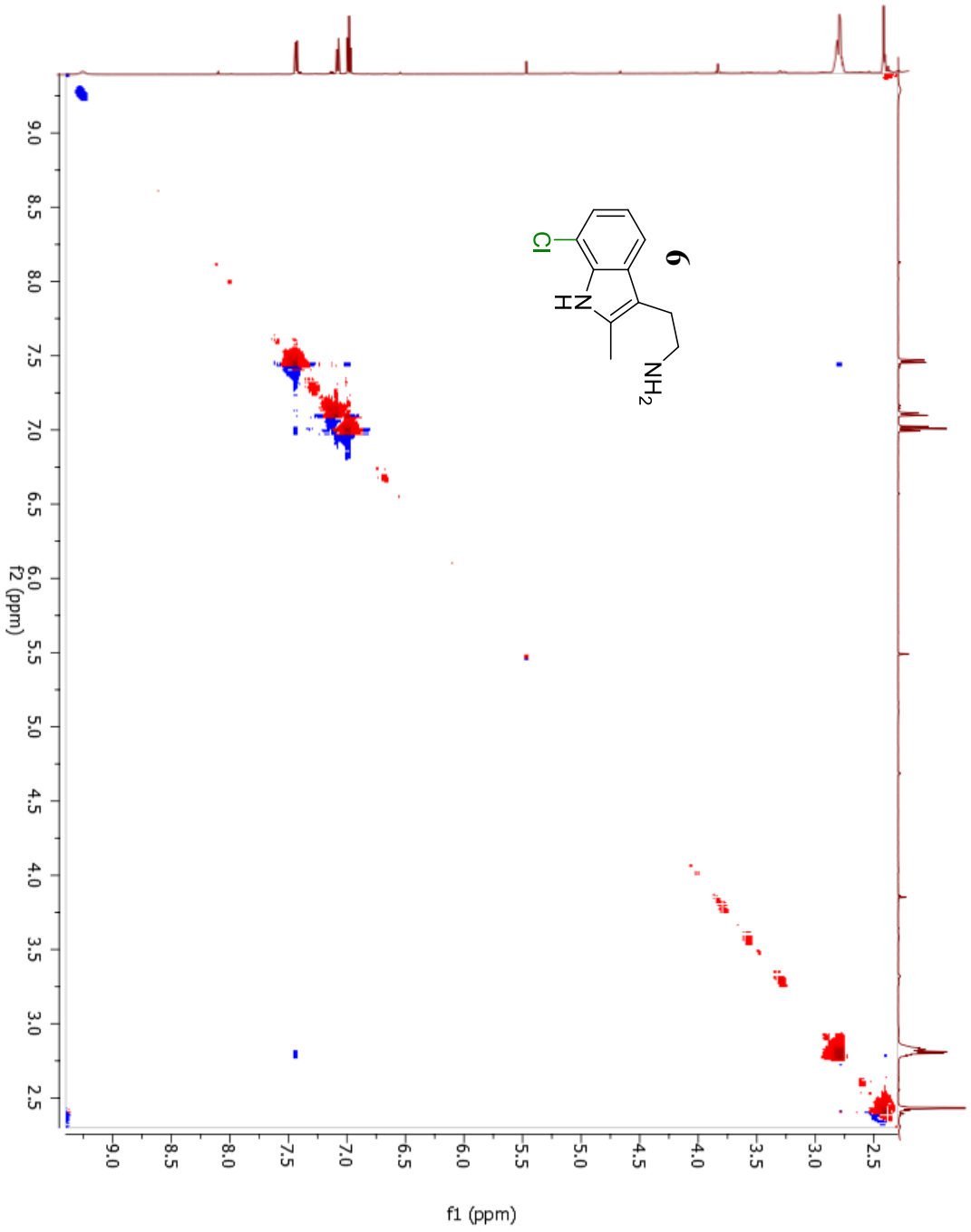
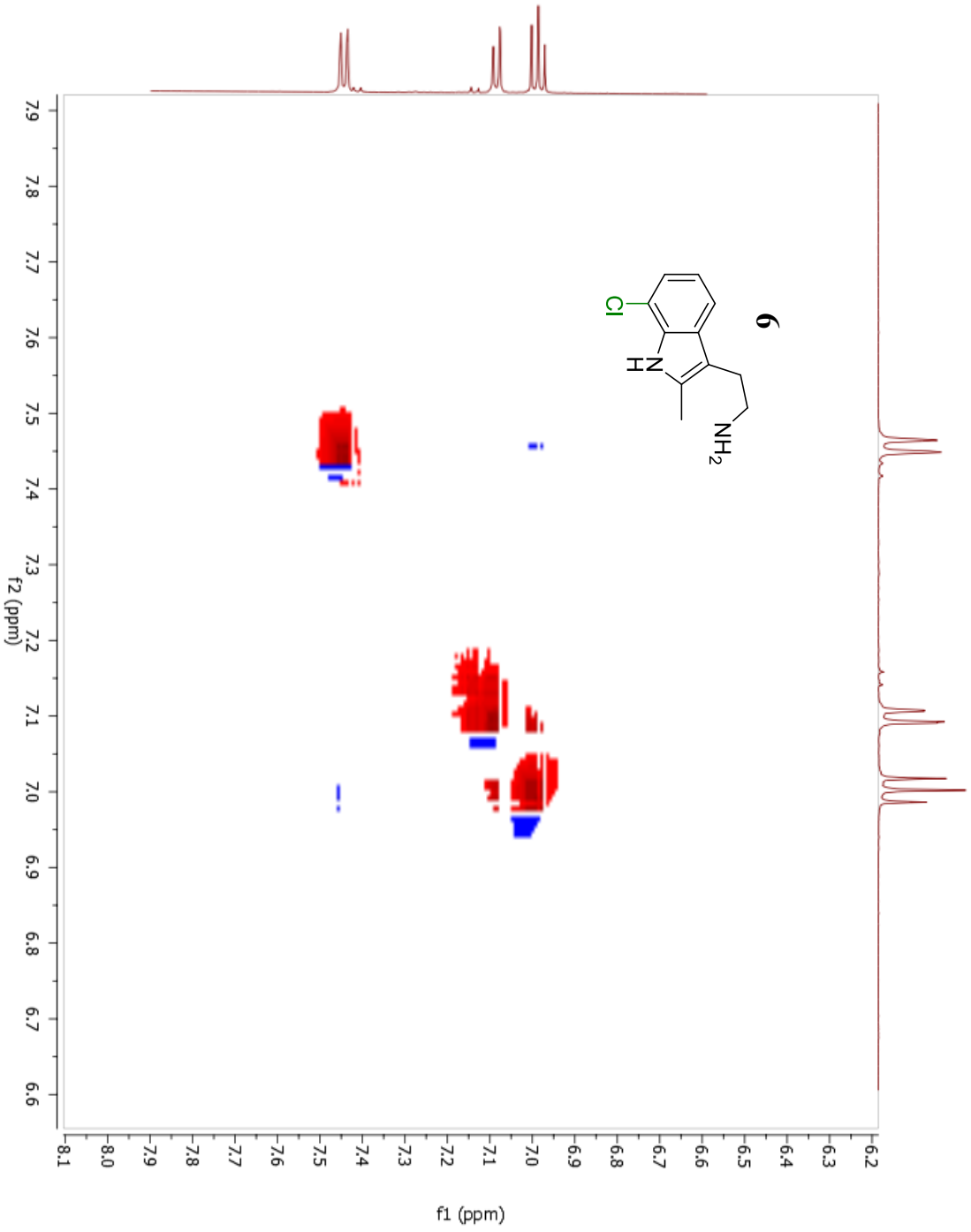


Figure AI.19: NOSEY spectrum of **6**.



NOESY of **6**: The resonance between peaks at 7.45 and 2.79 ppm demonstrate chlorination at the 7 position.

Figure AI.20: NOSEY spectrum of **6** (zoomed).



NOESY of **6**: Resonances between peaks at 7.45 and 7.00 ppm, and between peaks at 7.09 and 7.00 ppm support chlorination at the 7 position.

Figure AI.21: [1]HNMR spectrum of **7**.

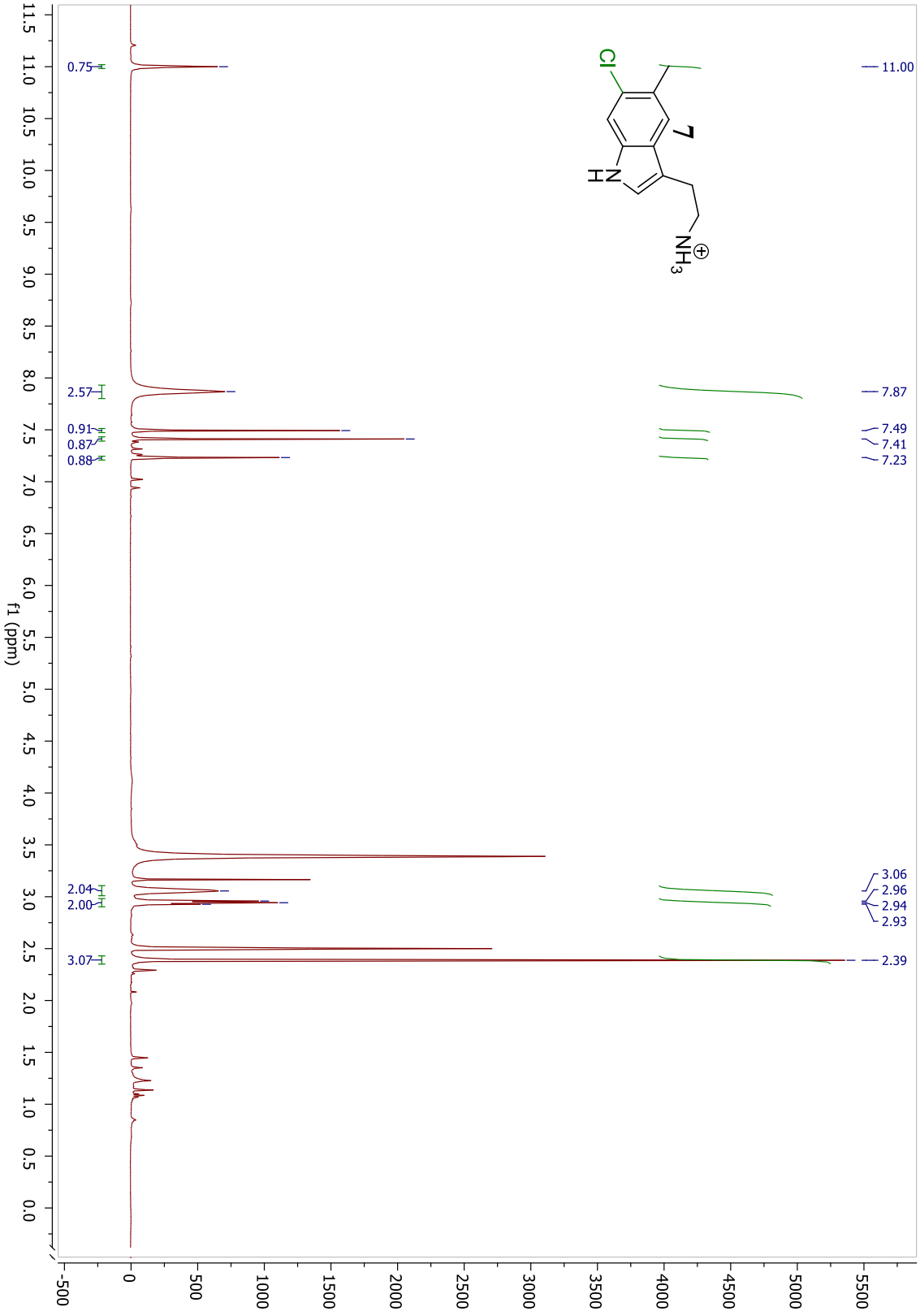


Figure AI.22: ^{13}CNMR spectrum of **7**.

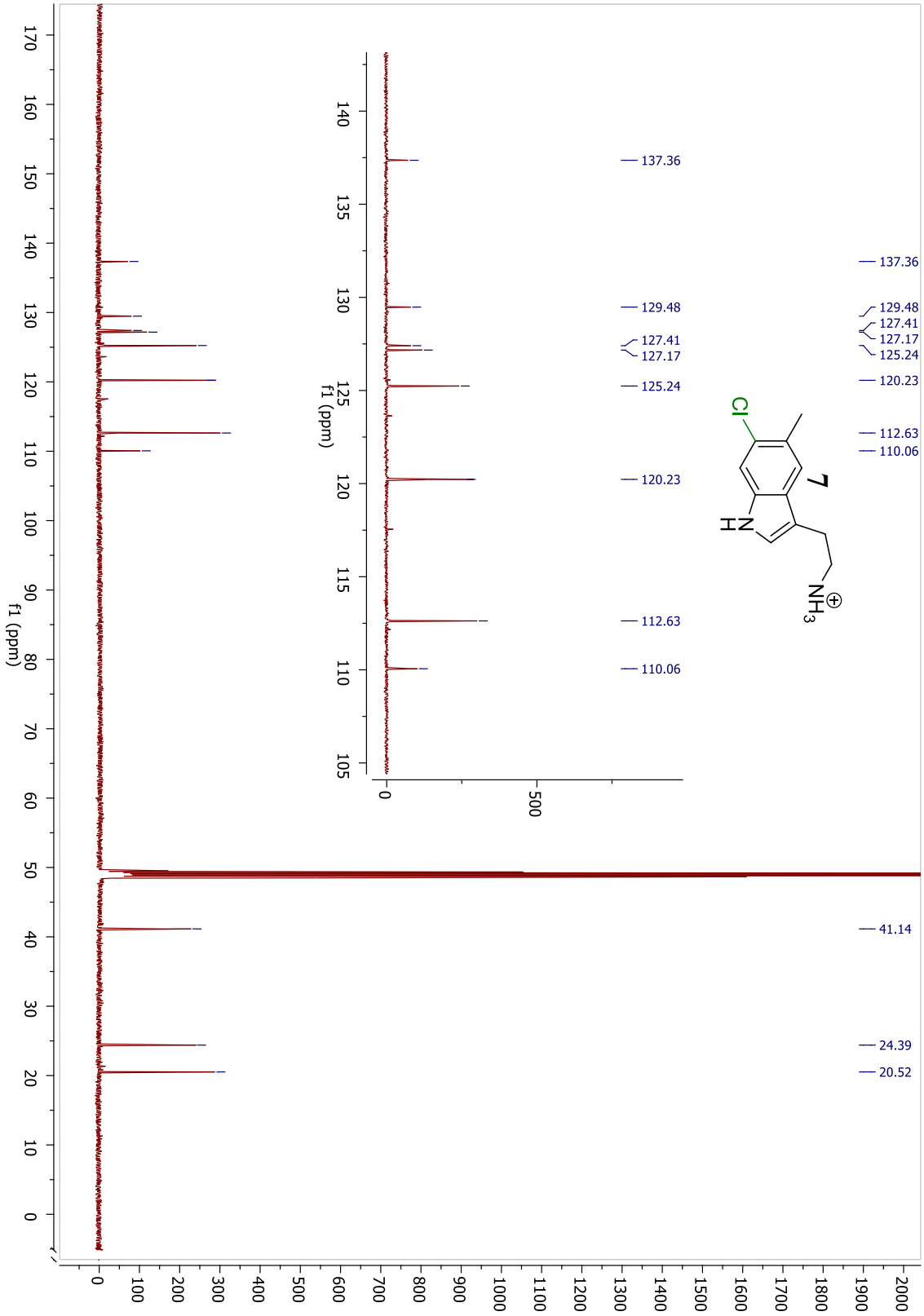
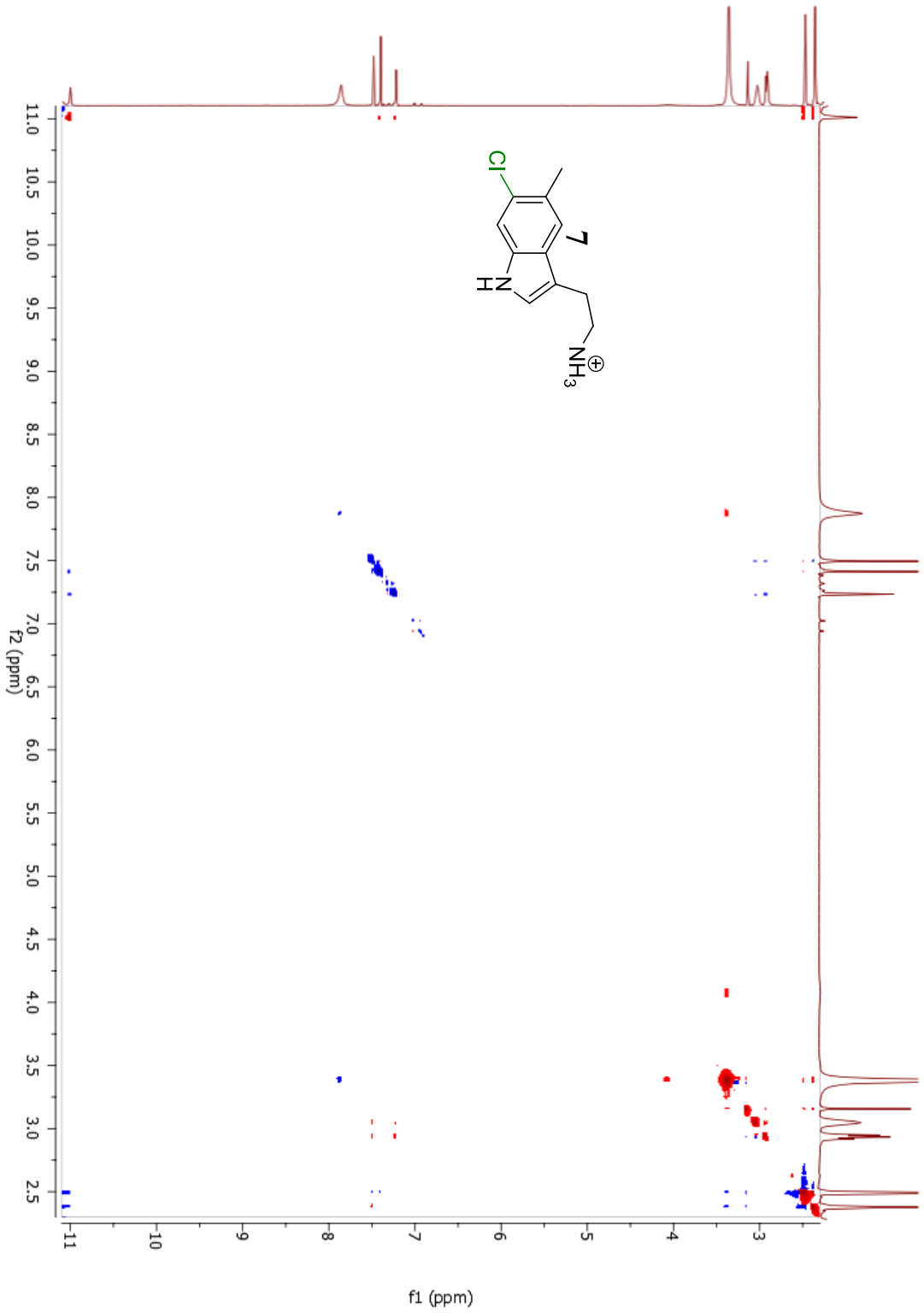
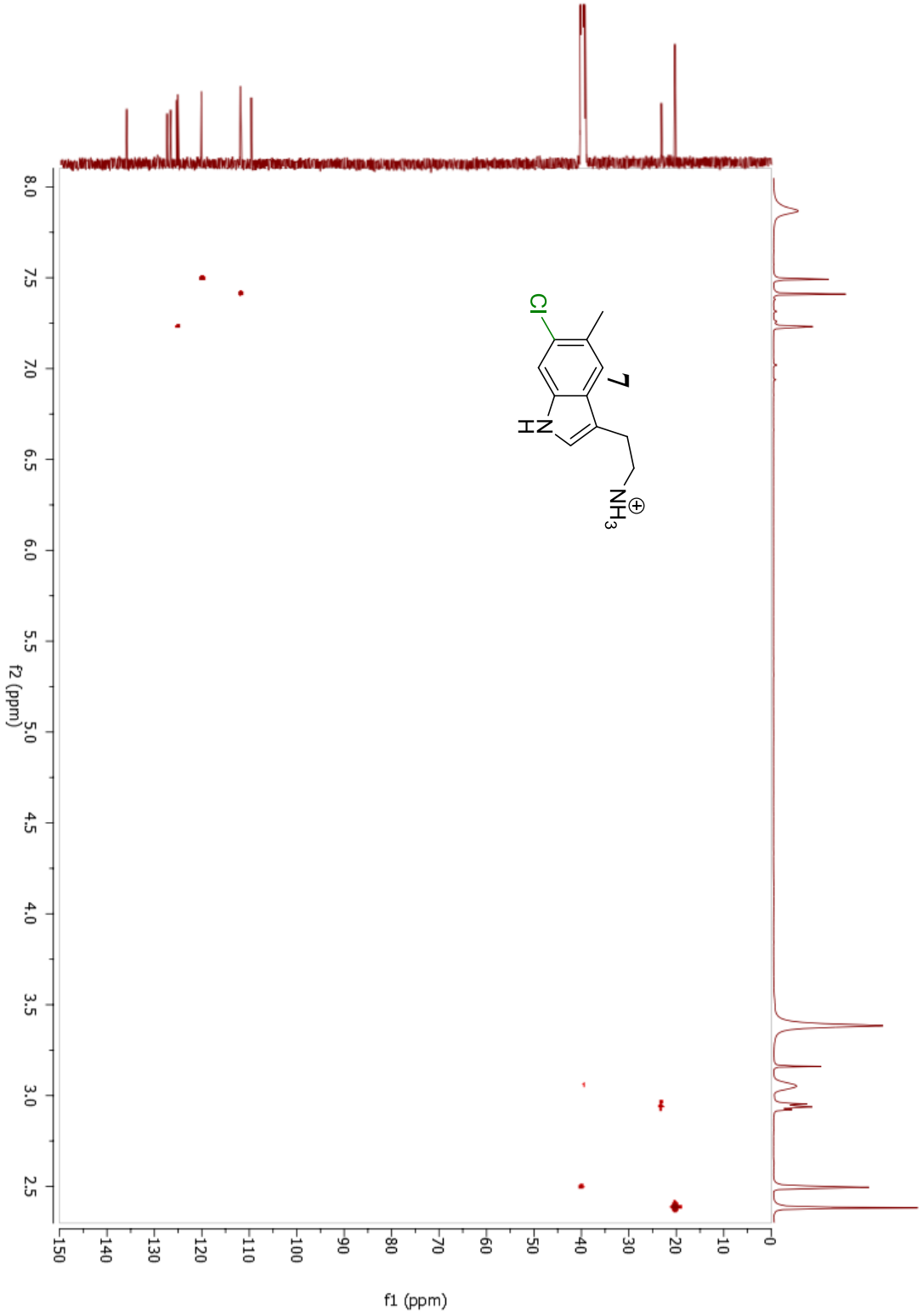


Figure AI.23: NOSEY spectrum of **7**.



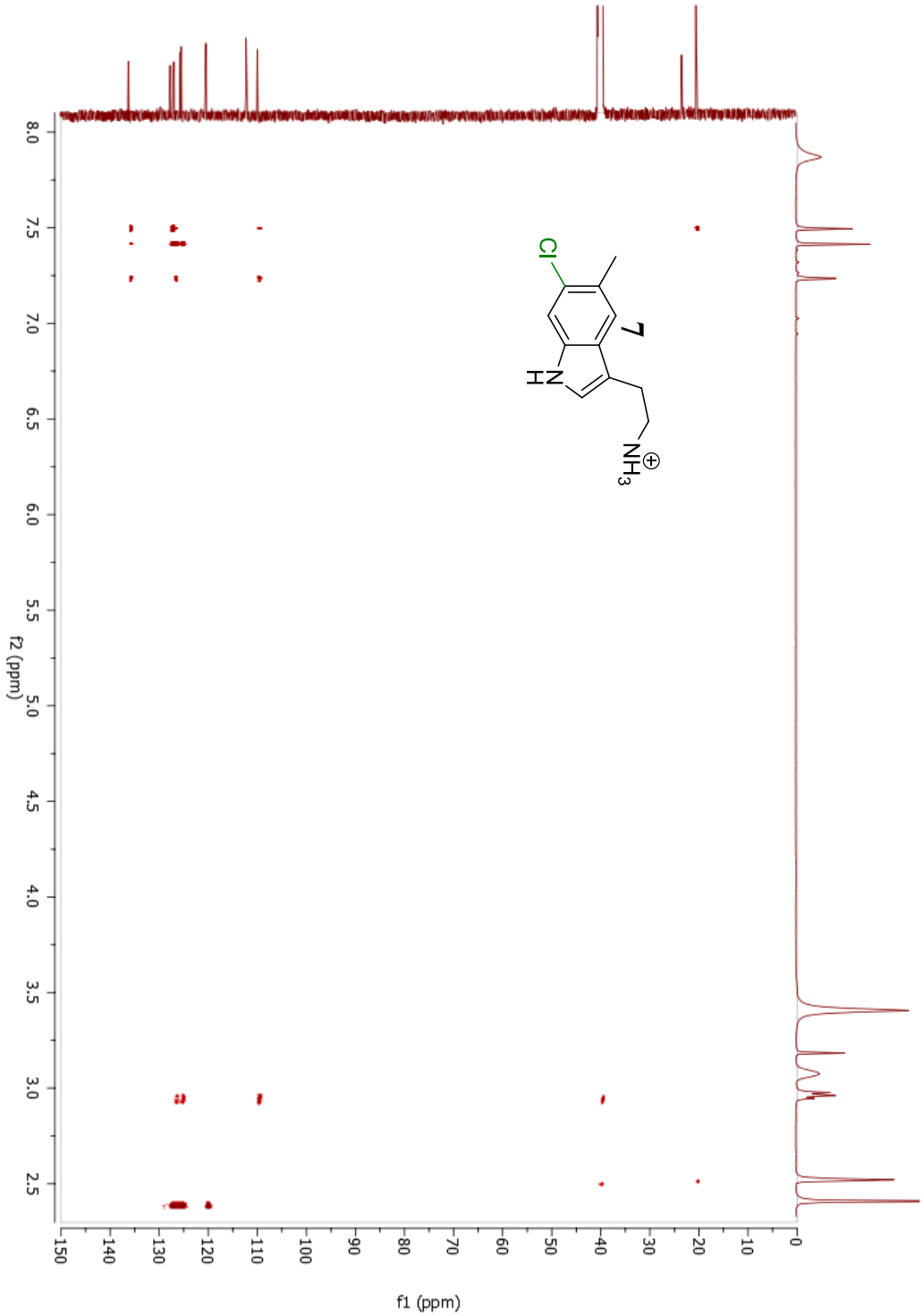
NOESY of **7**: Resonances between peaks at 7.50 and 2.95 ppm, between peaks at 7.24 and 2.95 ppm, and between peaks at 11.01 and 7.42 ppm support chlorination at

Figure AI.24: HMQC spectrum of **7**.



HMQC of **7**: The resonance between peaks at 2.40 and 20.54 ppm identifies the 5-methyl carbon, which is necessary to analyze the HMBC spectrum.

Figure AI.25: HMBC spectrum of **7**.



HMBC of **7**: The resonance between peaks at 7.50 and 20.54 ppm confirms the identity of the proton on C4. The HMBC of the starting material shows a clear resonance between the 5-methyl carbon and the proton at C6. This resonance does not appear in this product, thus supporting chlorination at the 6 position.

Figure AI.26: [1]HNMR spectrum of **8a** and **8b**.

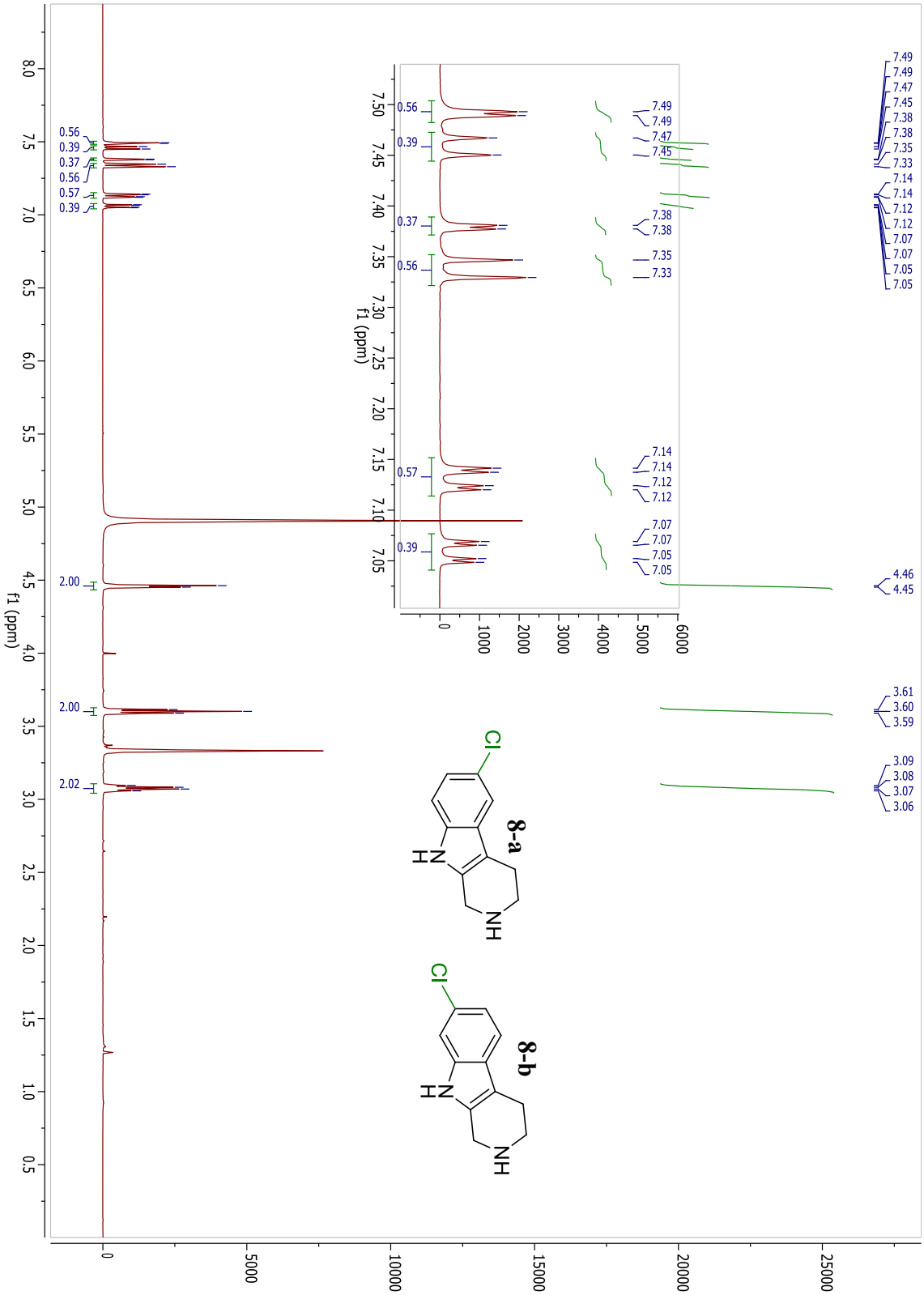


Figure AI.27: ^{13}CNMR spectrum of **8a** and **8b**.

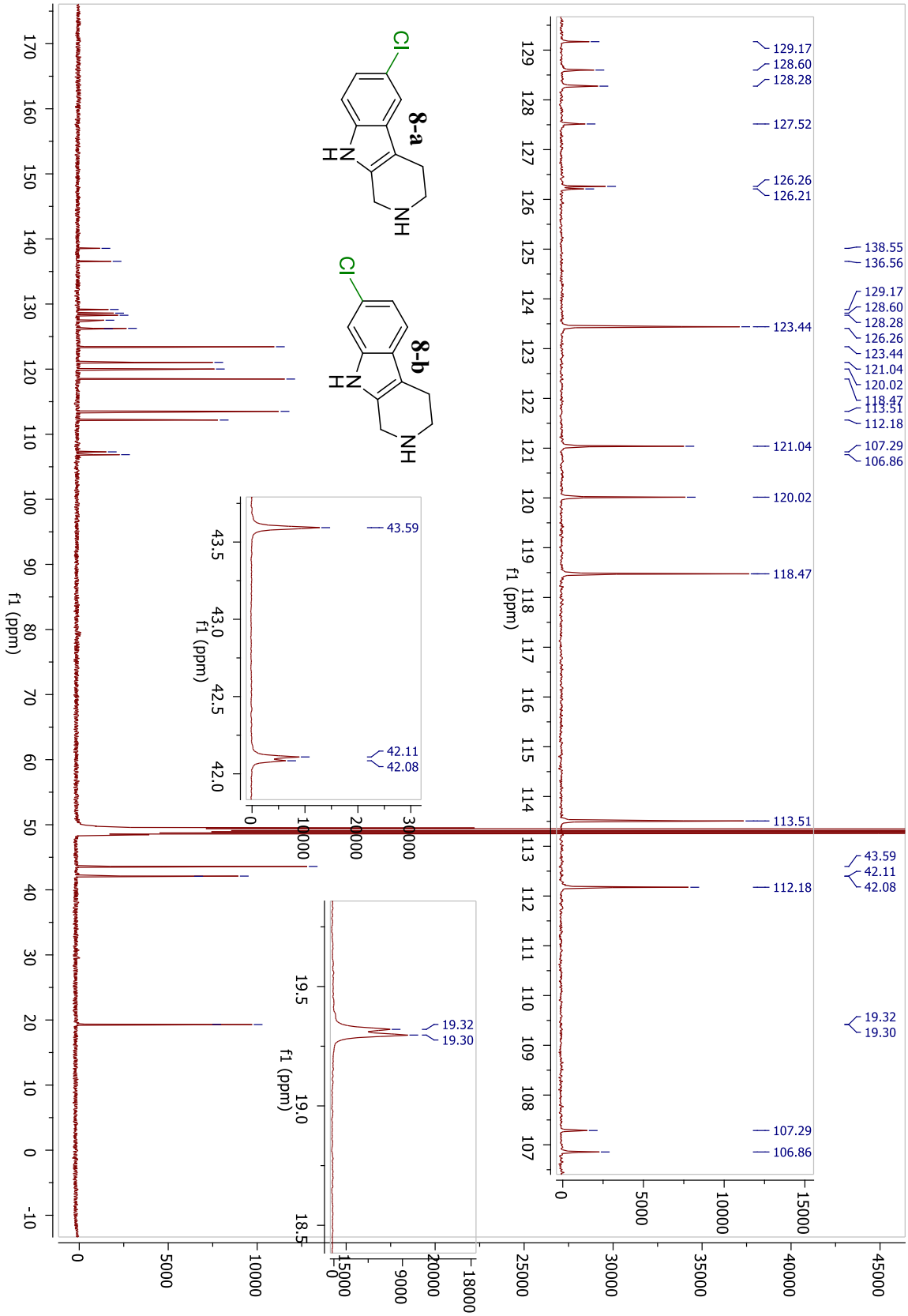


Figure AI.28: [1]HNMR spectrum of **9**.

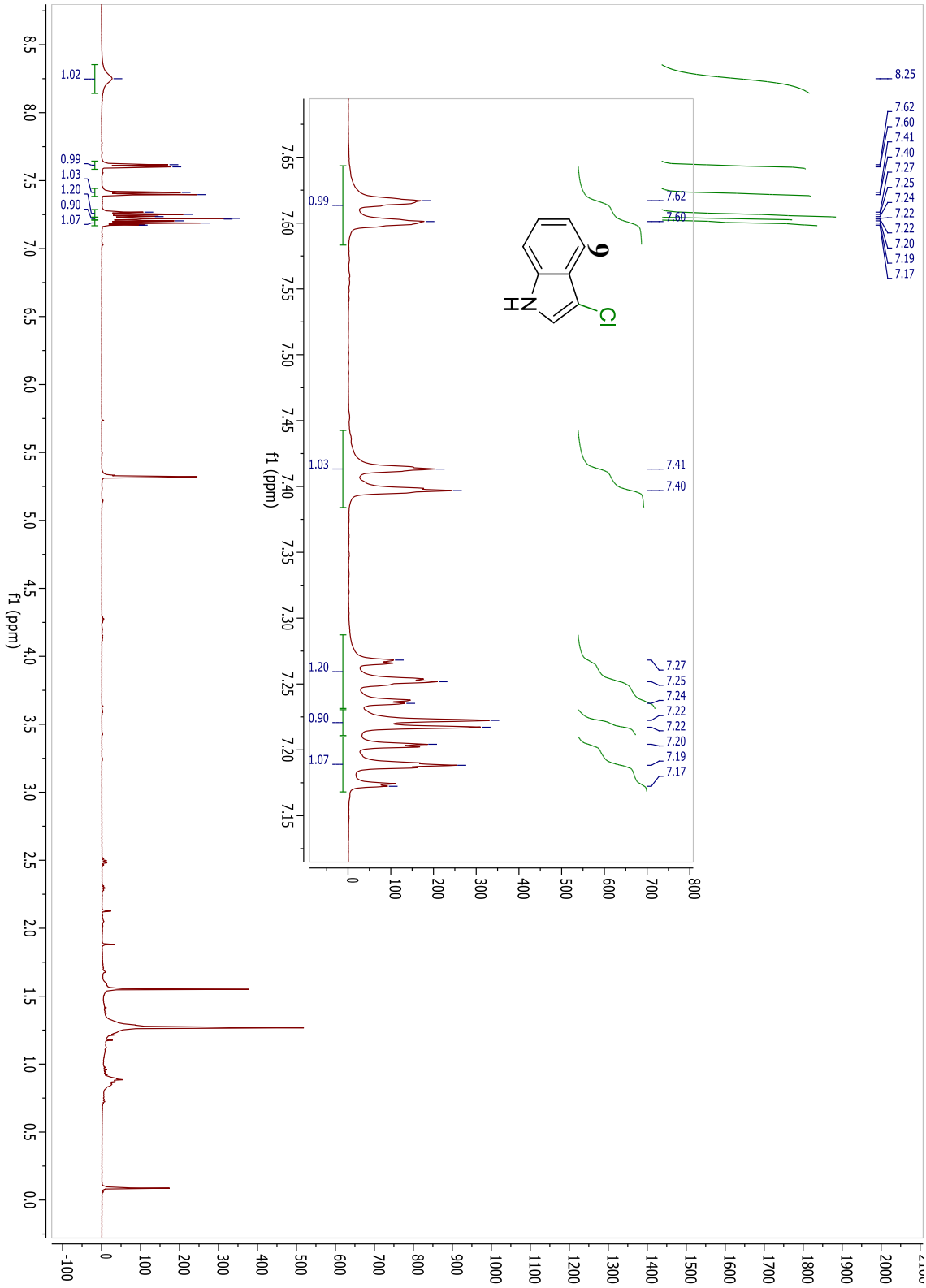


Figure AI.29: ^{13}CNMR spectrum of **9**.

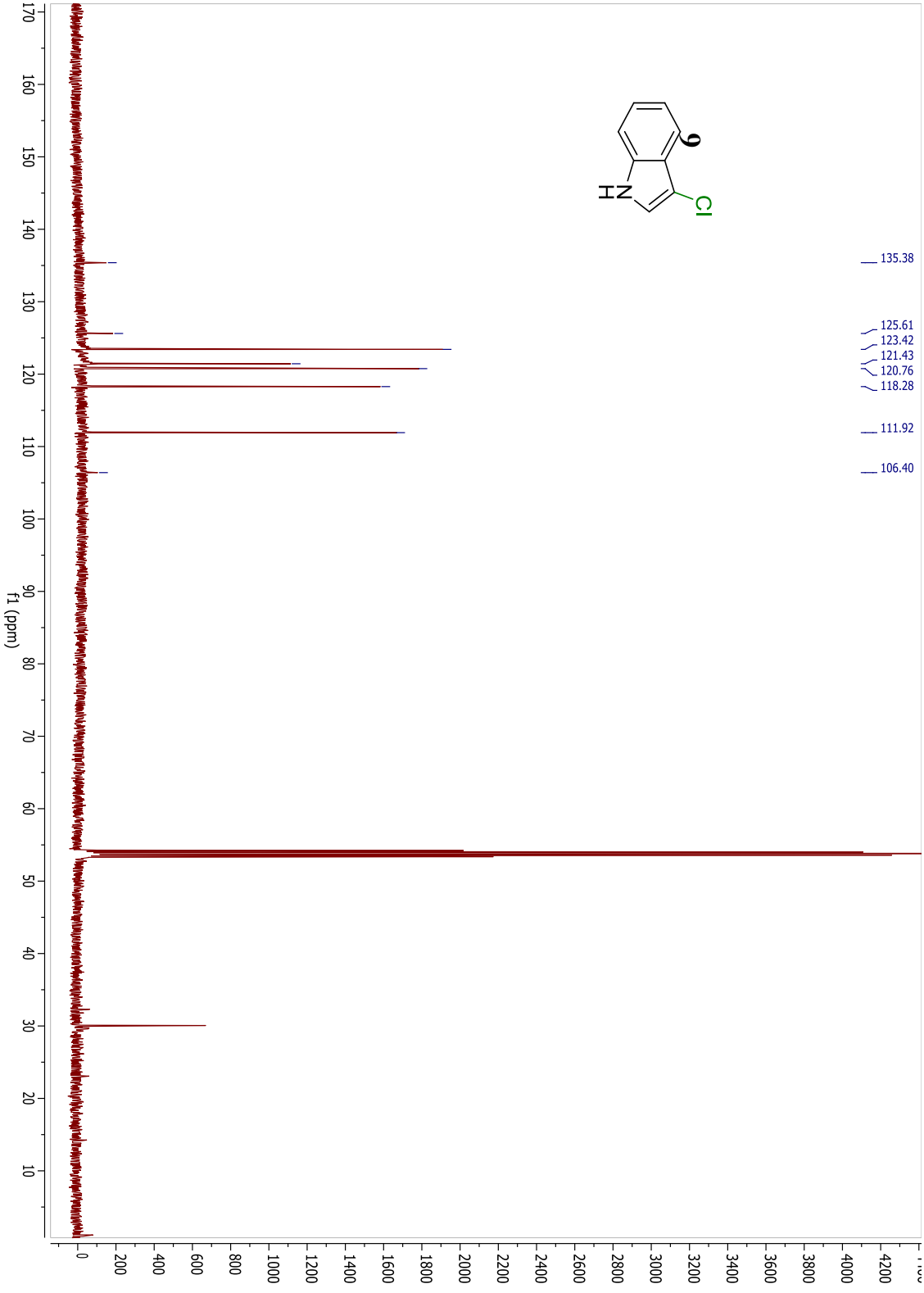


Figure AI.30: [1]HNMR spectrum of **10**.

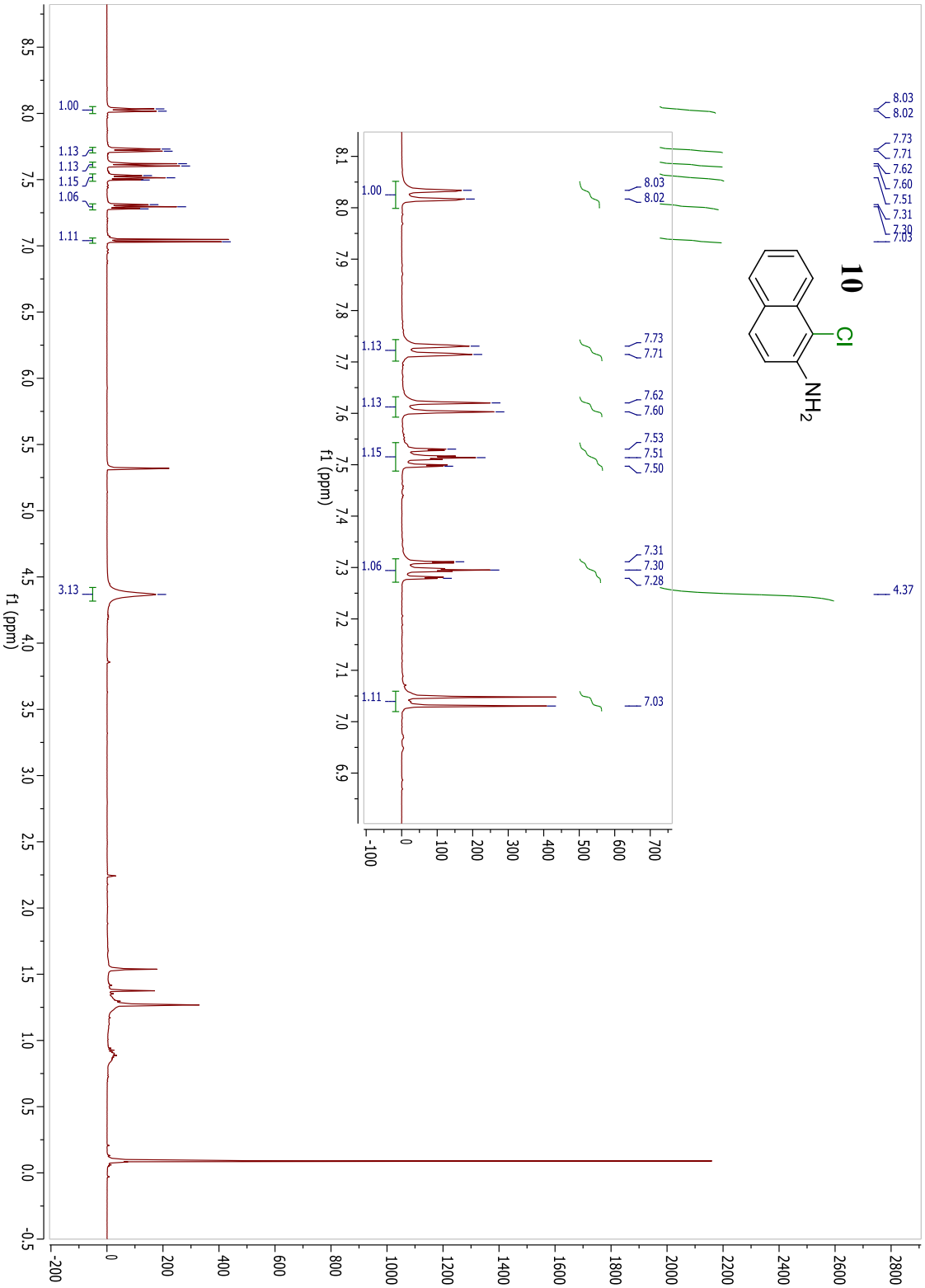


Figure AI.31: ^{13}CNMR spectrum of **10**.

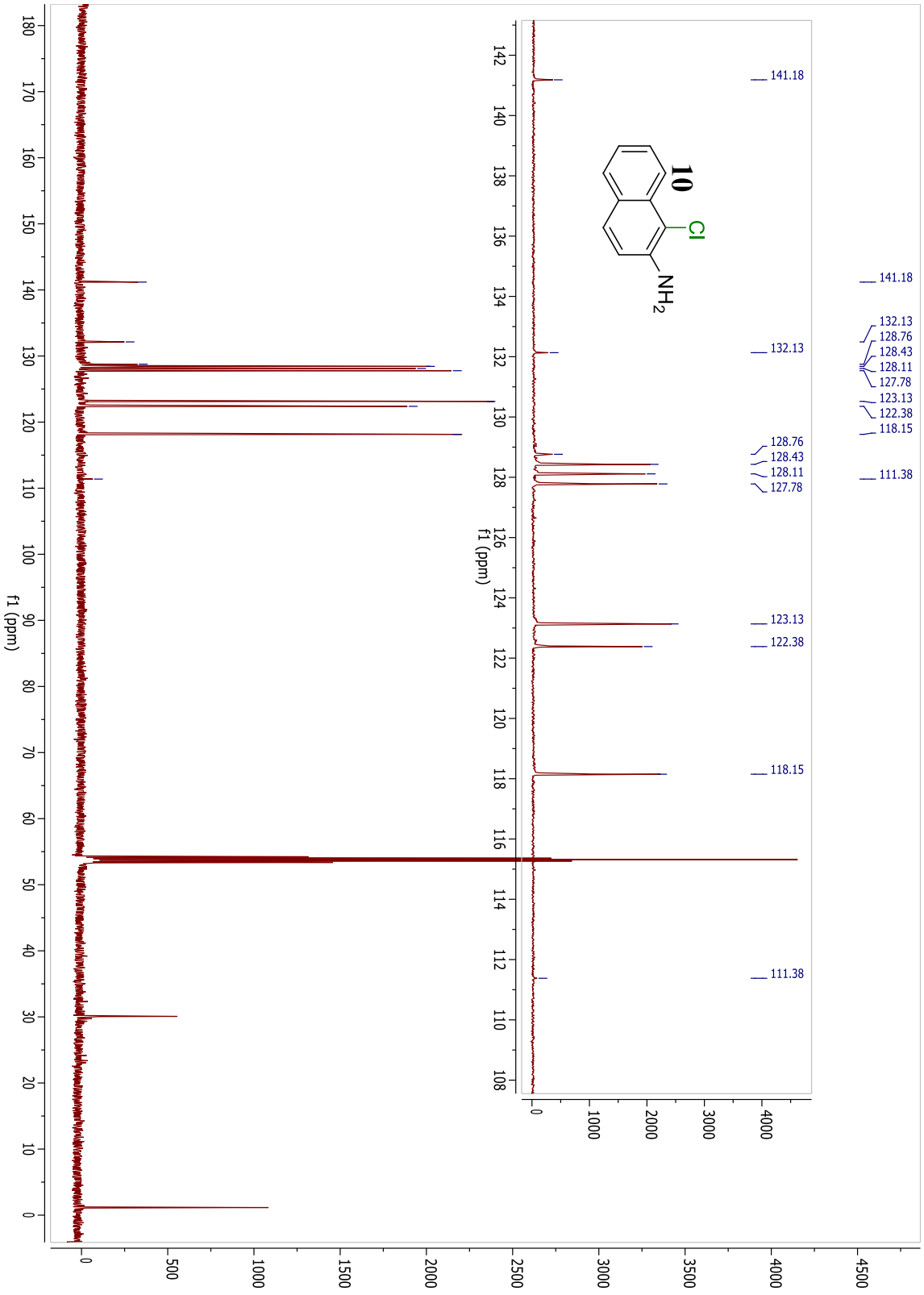


Figure AI.32: NOSEY spectrum of **10**.

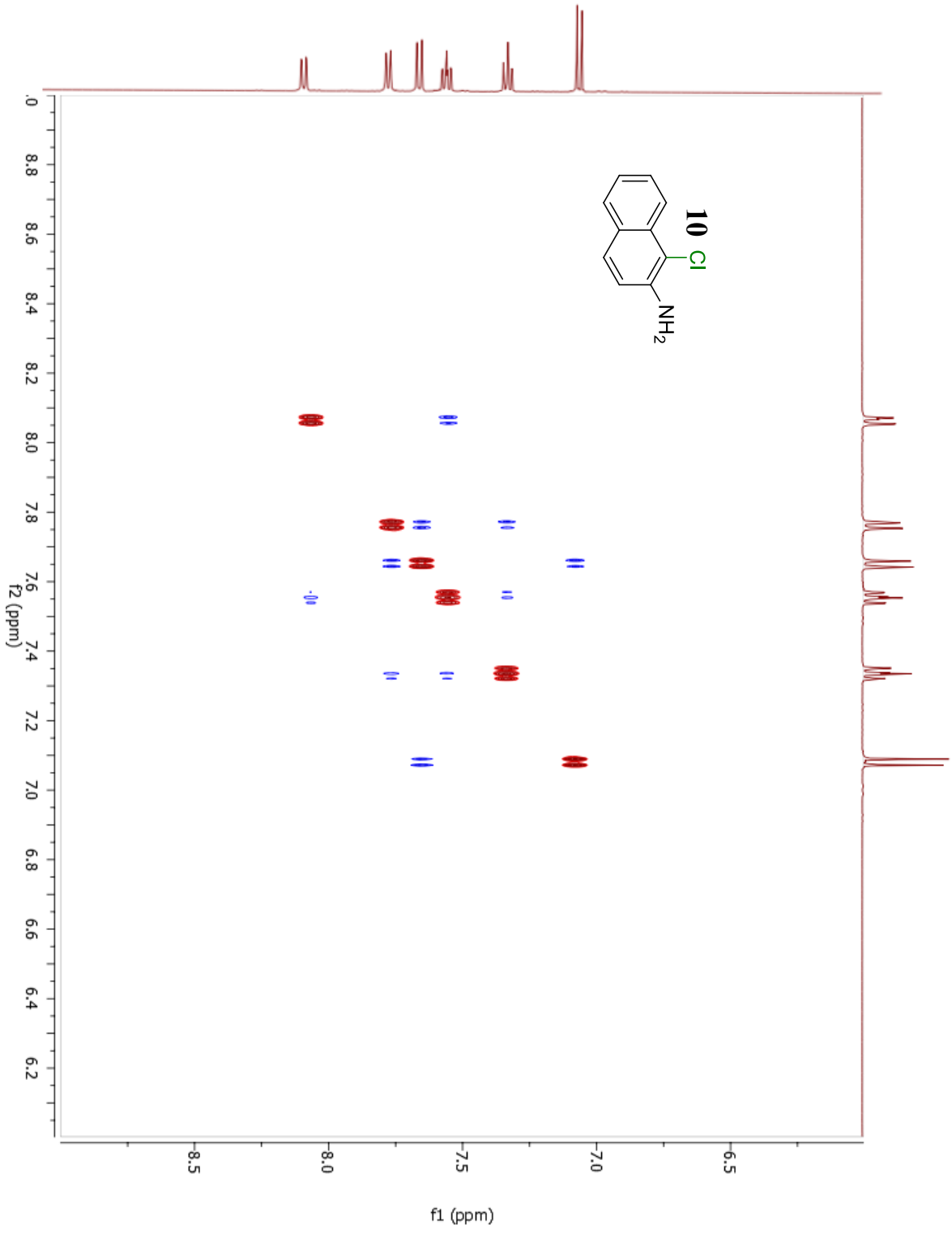


Figure AI.33: [1]HNMR spectrum of **11**.

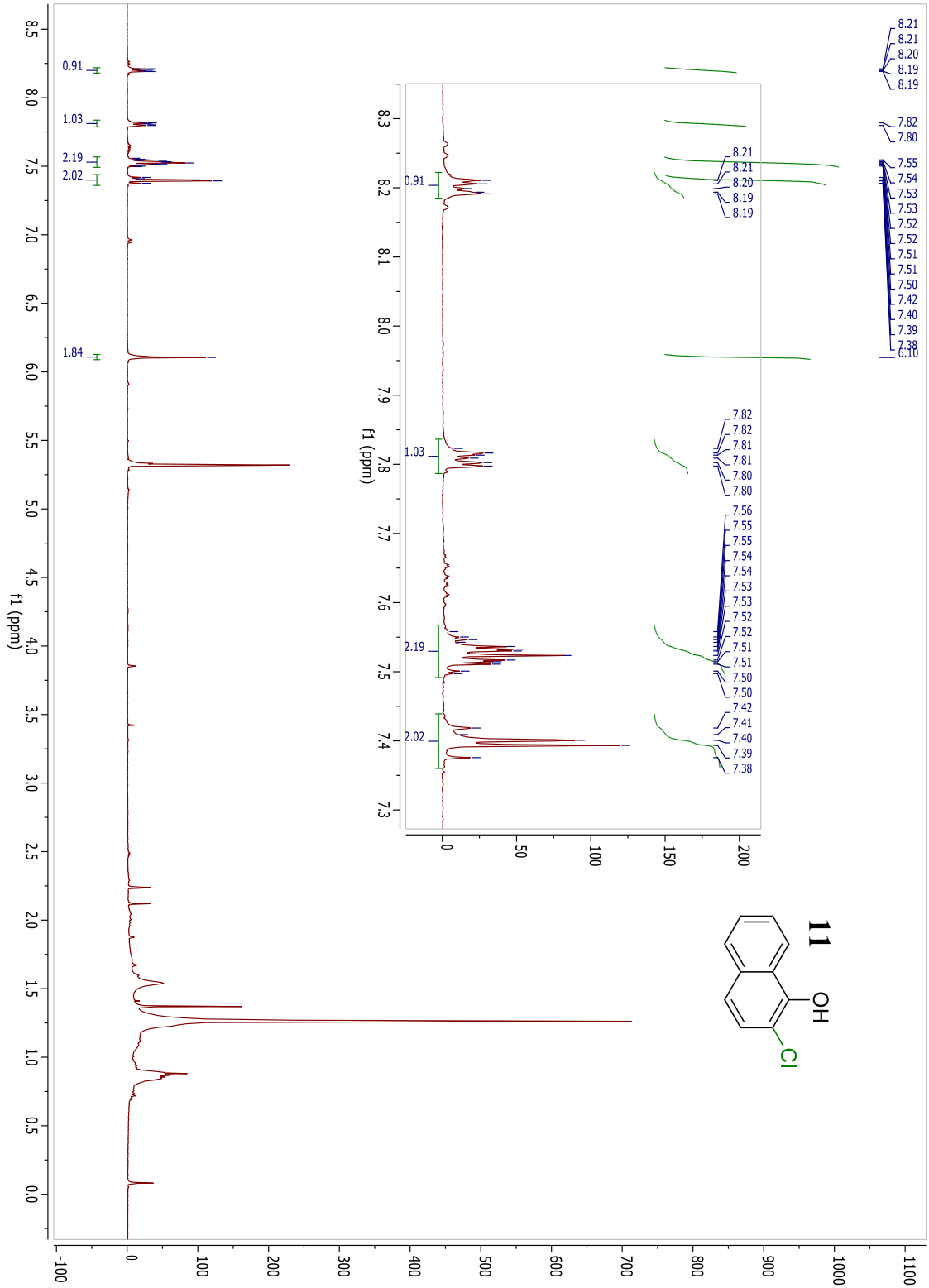


Figure AI.34: ^{13}CNMR spectrum of **11**.

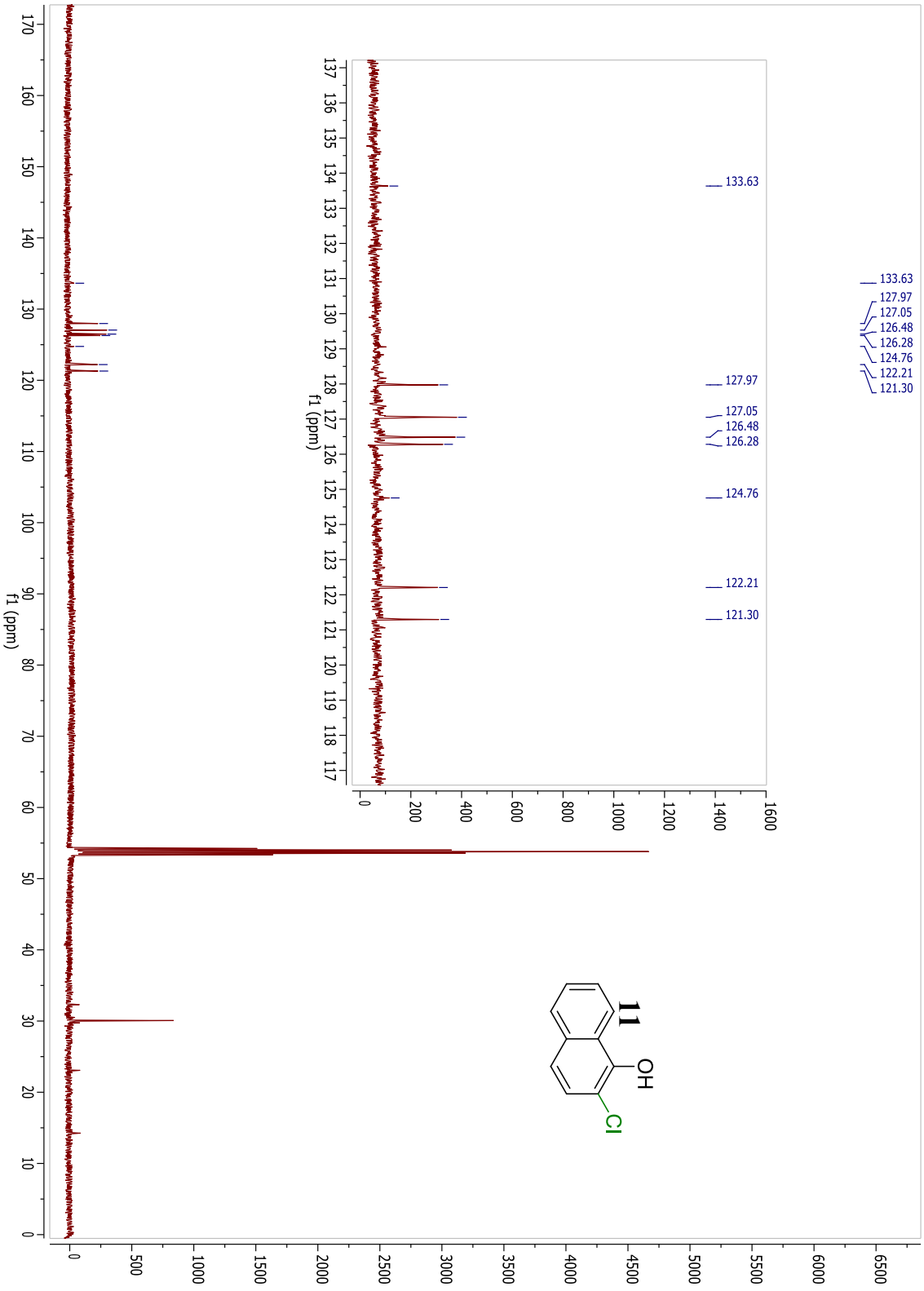


Figure AI.35: [1]HNMR spectrum of **12**.

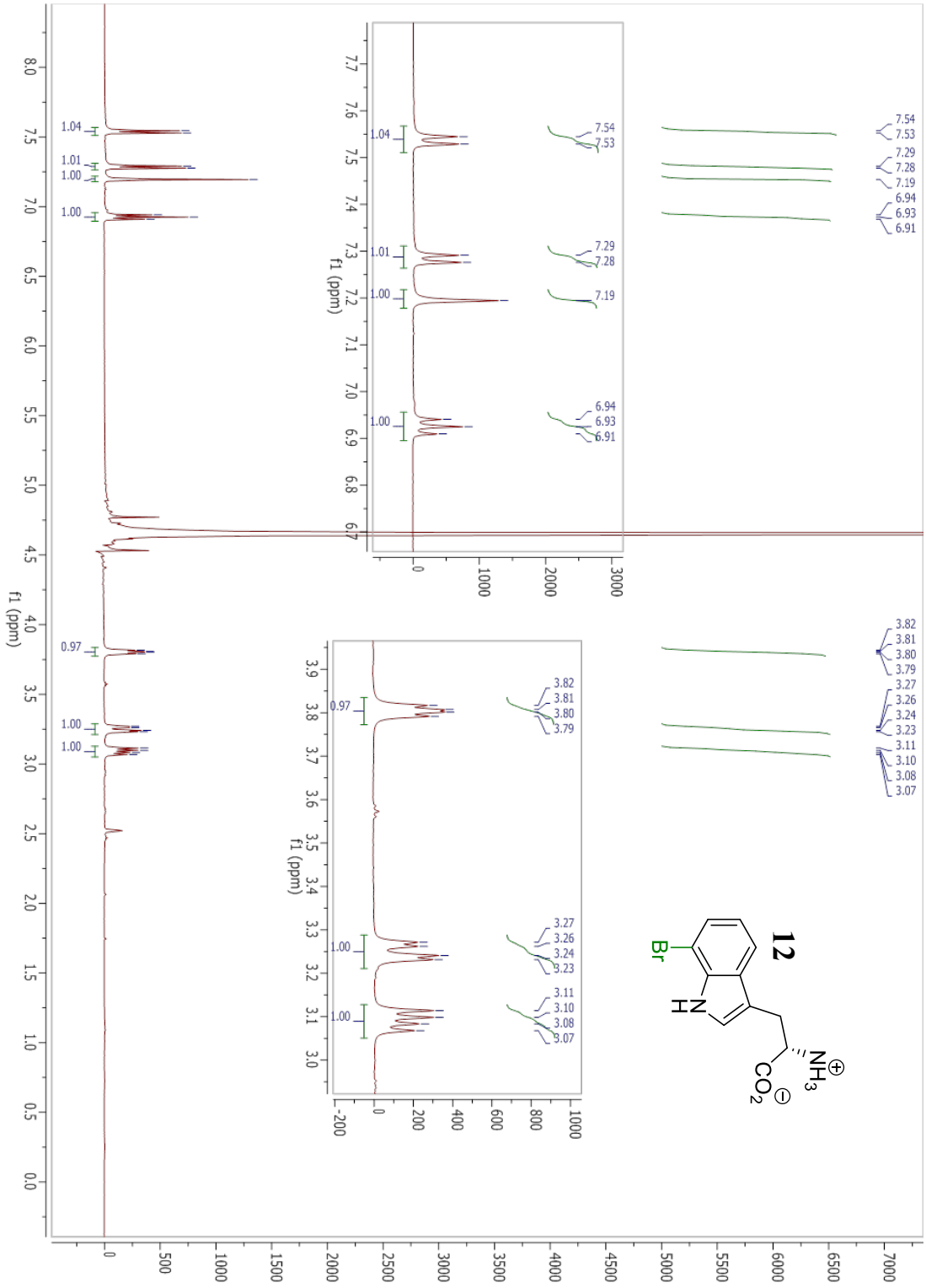


Figure AI.36: ^{13}CNMR spectrum of **12**.

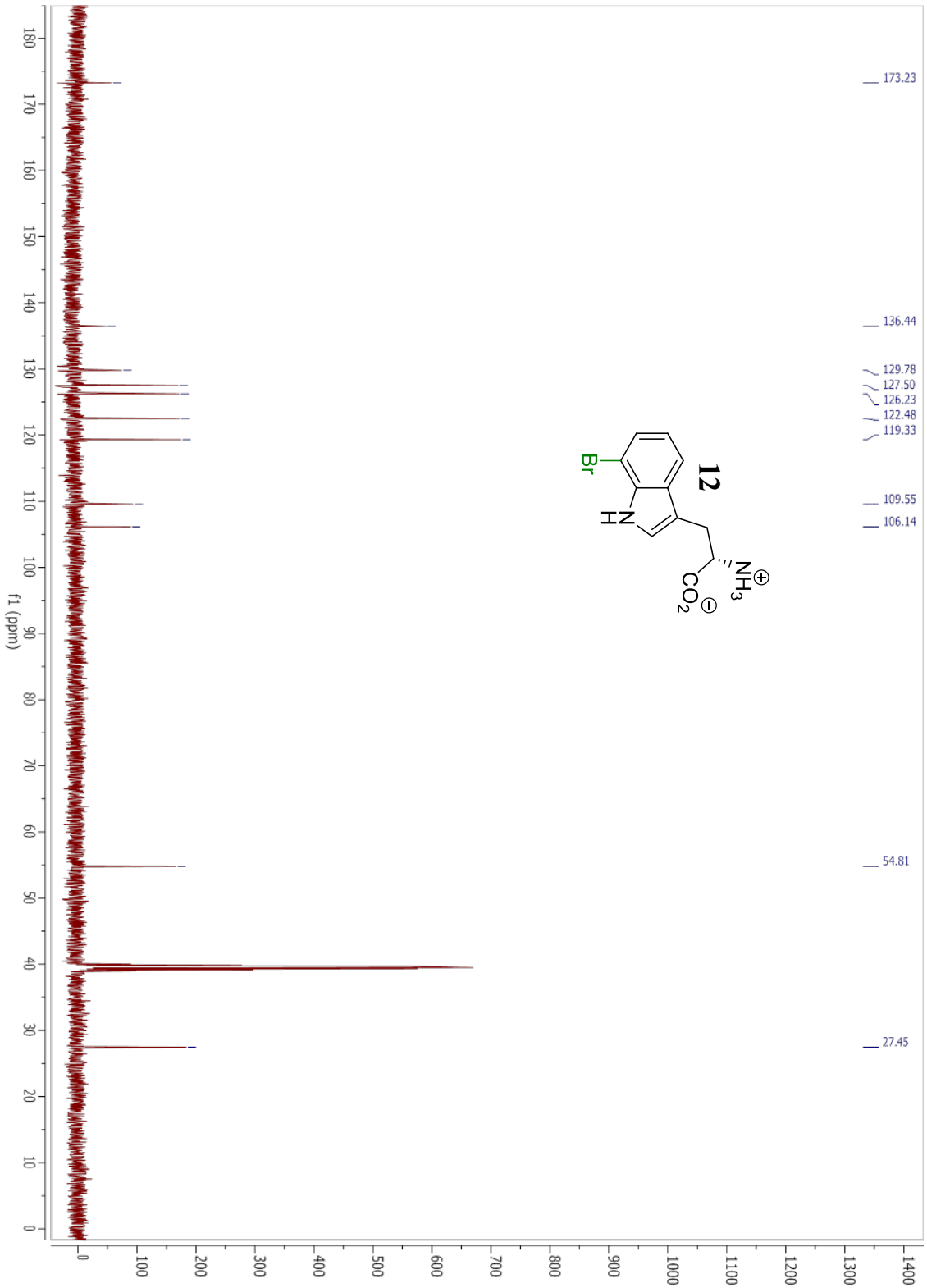
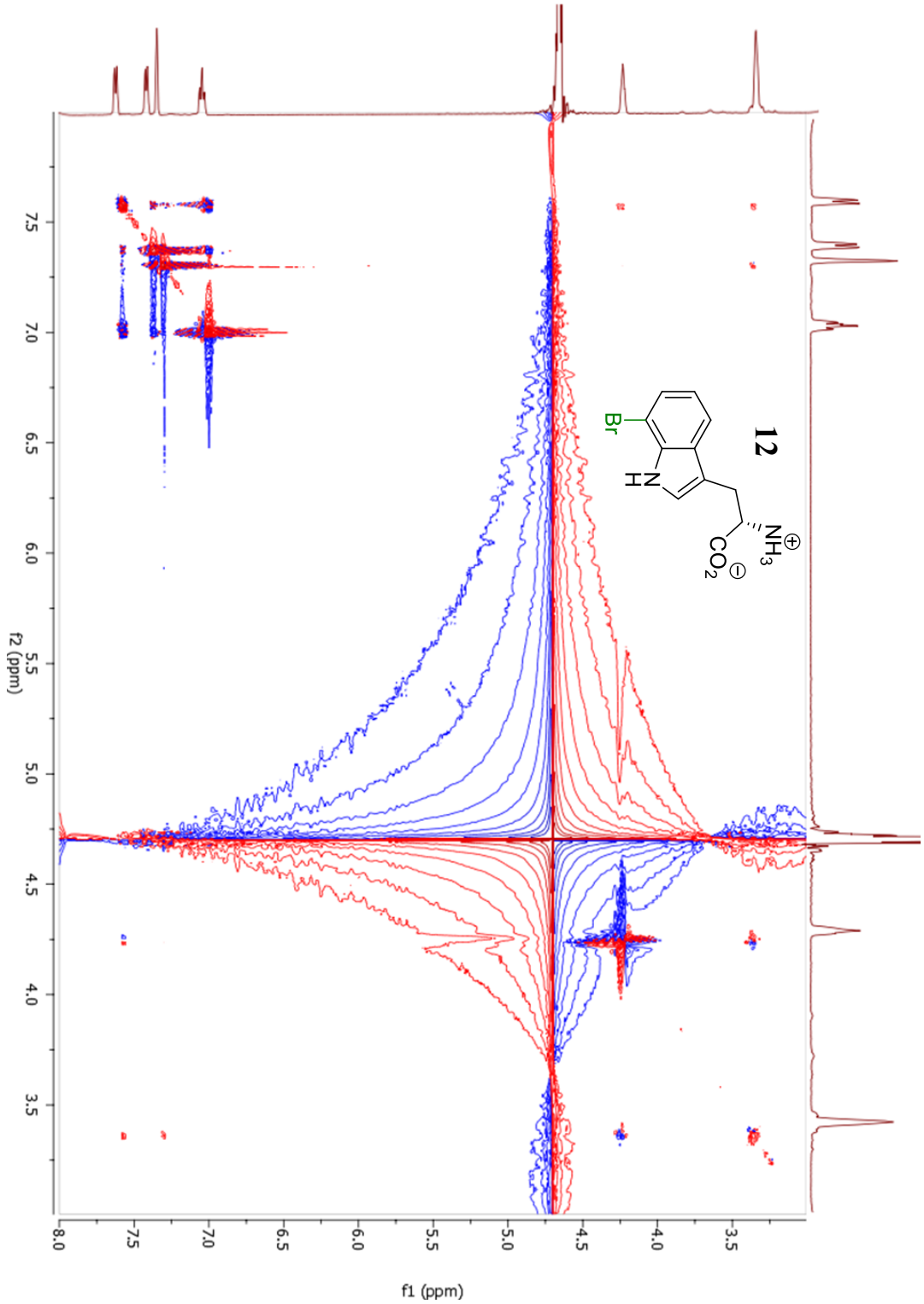


Figure AI.37: NOSEY spectrum of **12**.



NOESY of **12**: Resonances between peaks at 7.19 and 3.42 ppm, and between peaks at 7.54 and 3.42 ppm demonstrate chlorination at the 7 position.

Appendix II

NMR Spectra for Compounds from Chapter 3

Figure AII.1: ^{1}HNMR spectrum of **1a**.

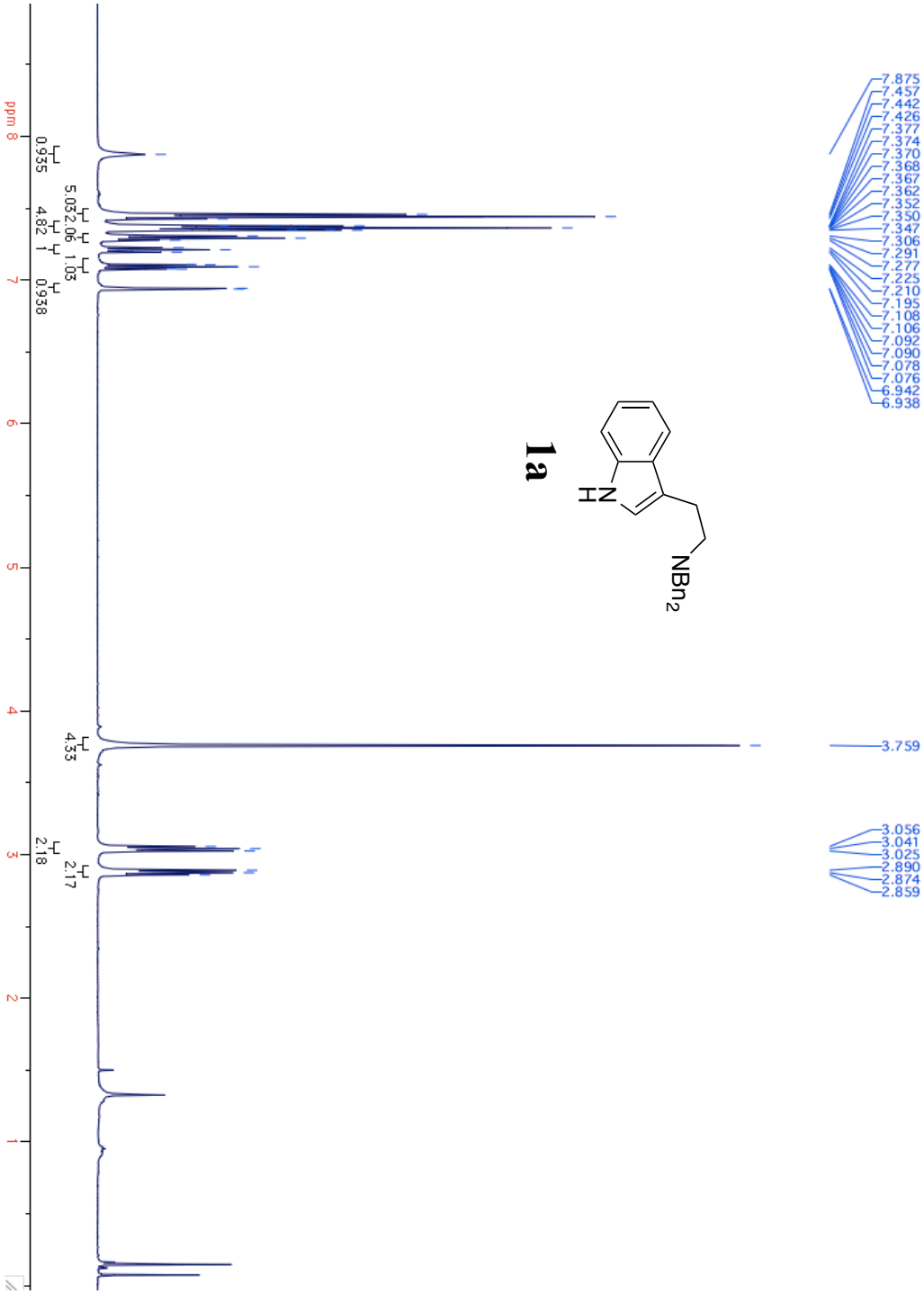


Figure AII.2: [1]HNMR spectrum of **1b**.

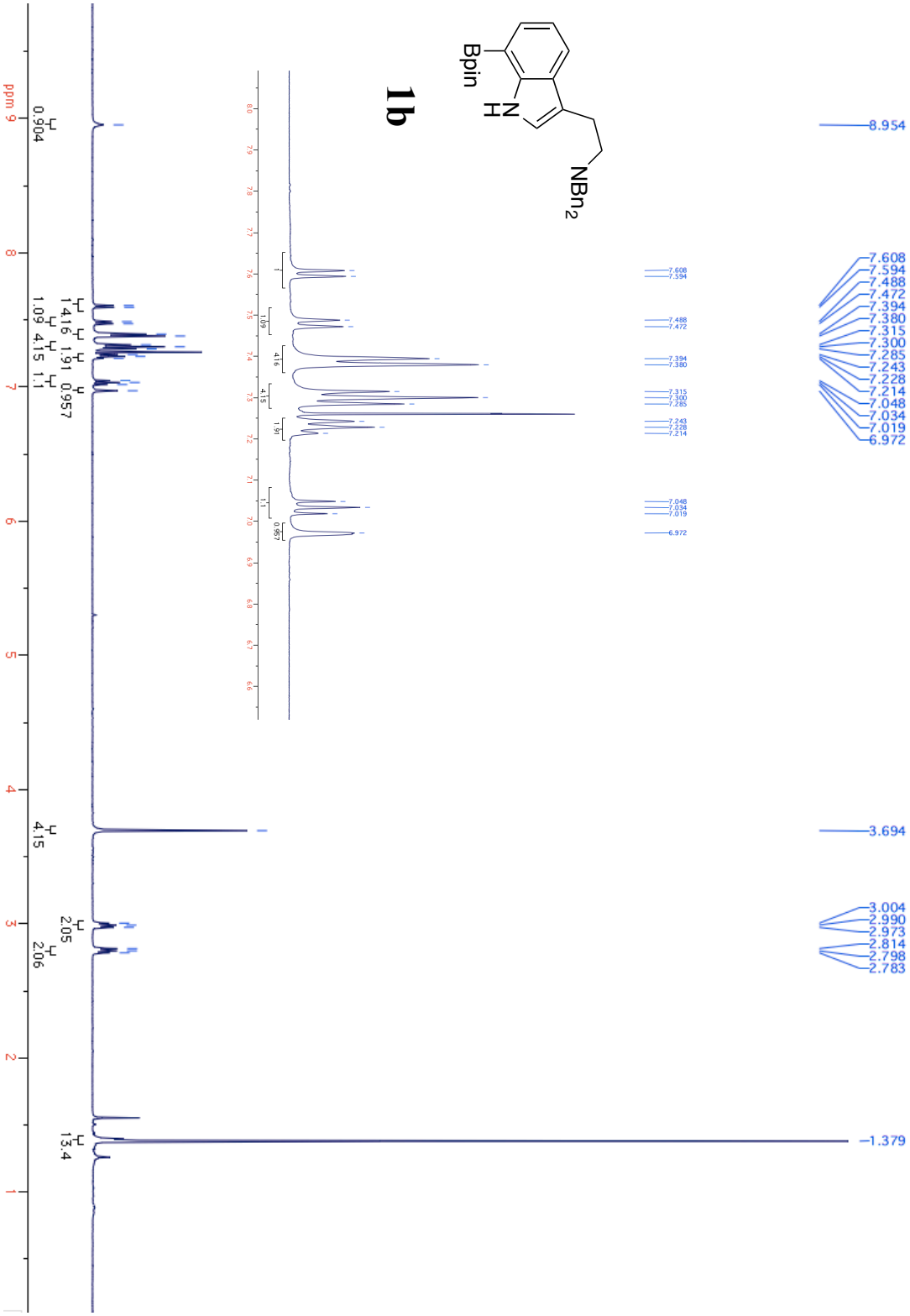


Figure AII.3: ^{13}CNMR spectrum of **1b**.

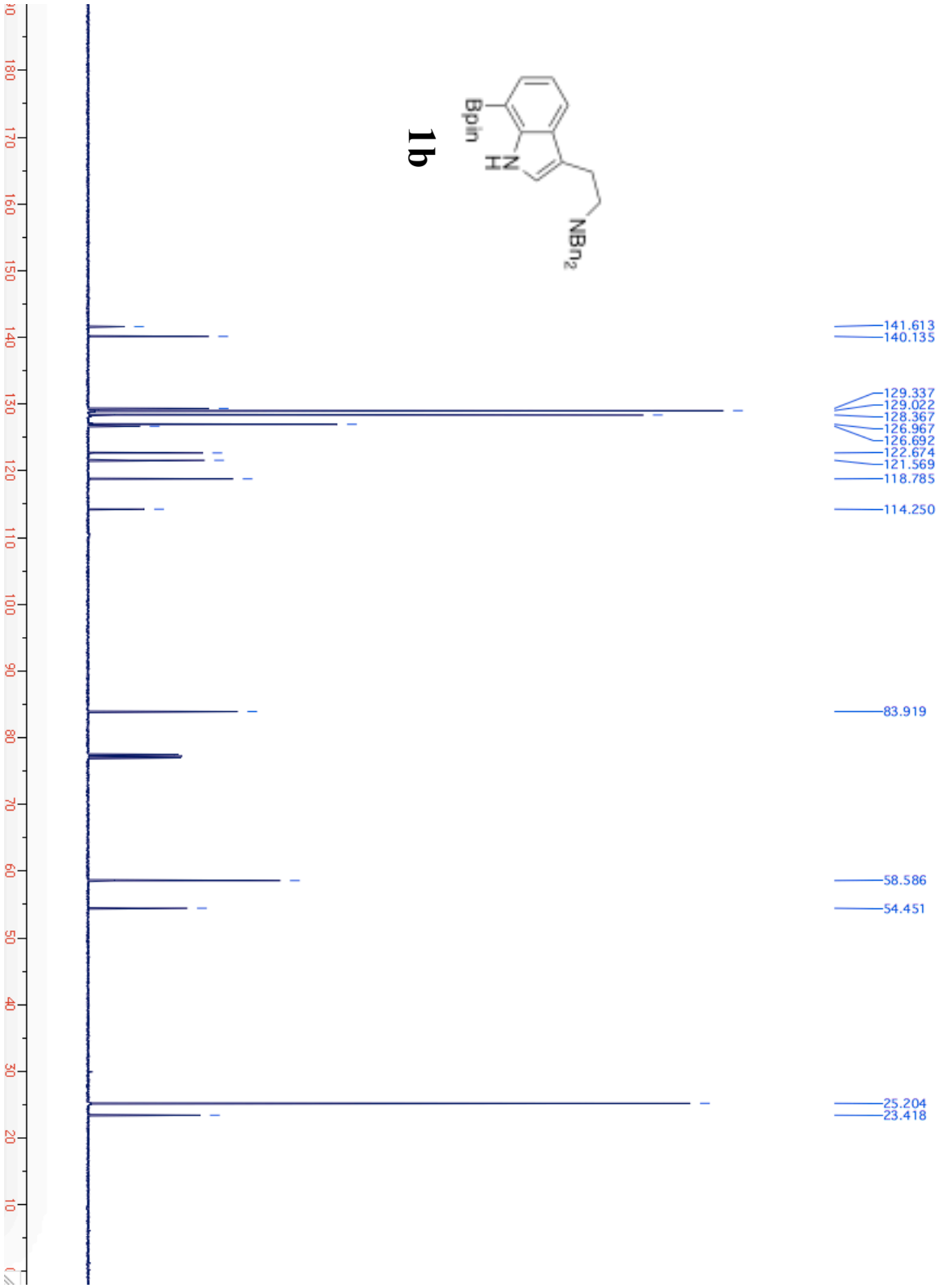


Figure AII.4: [1]HNMR spectrum of **1c**.

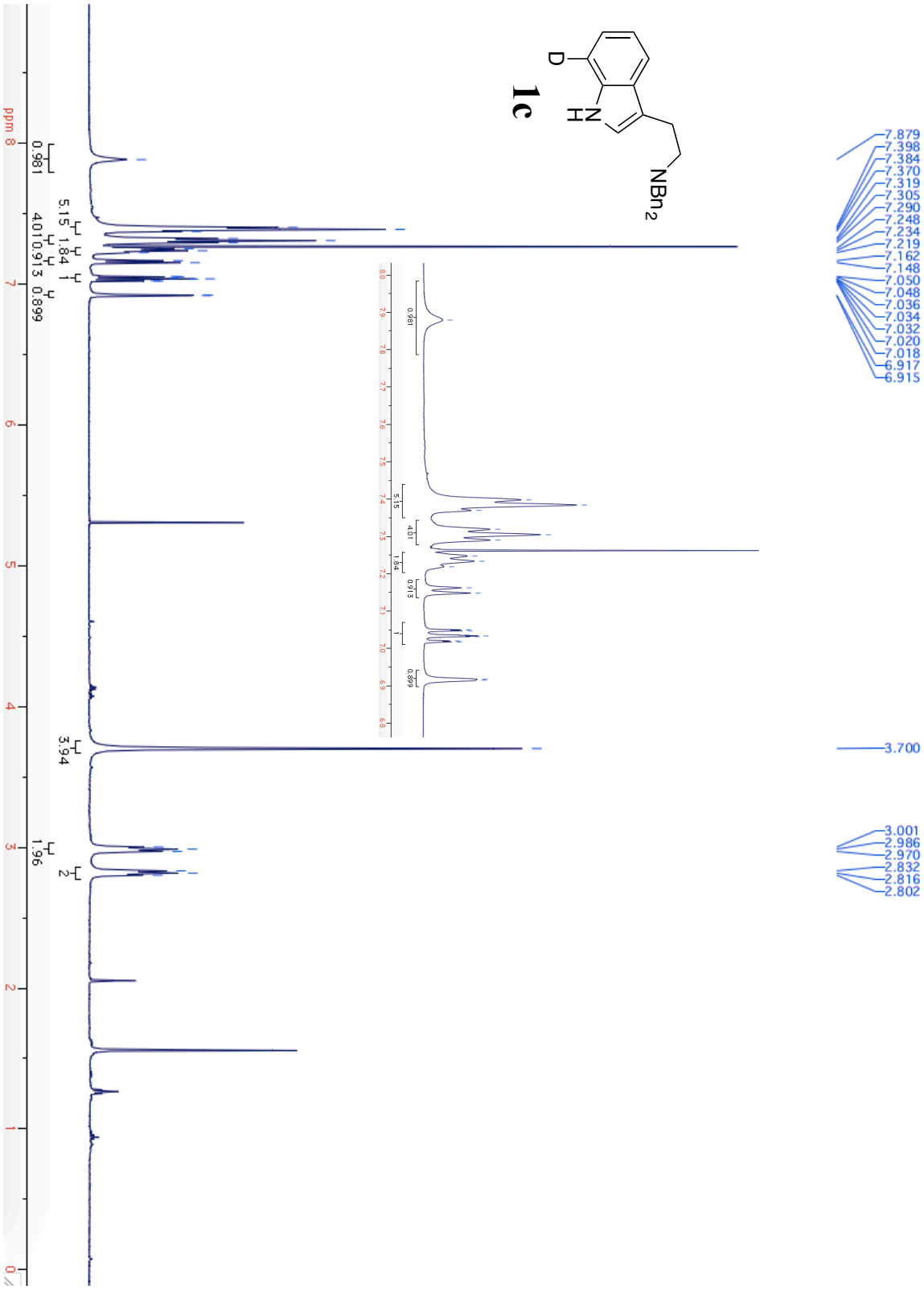


Figure AII.5: ^{13}CNMR spectrum of **1c**.

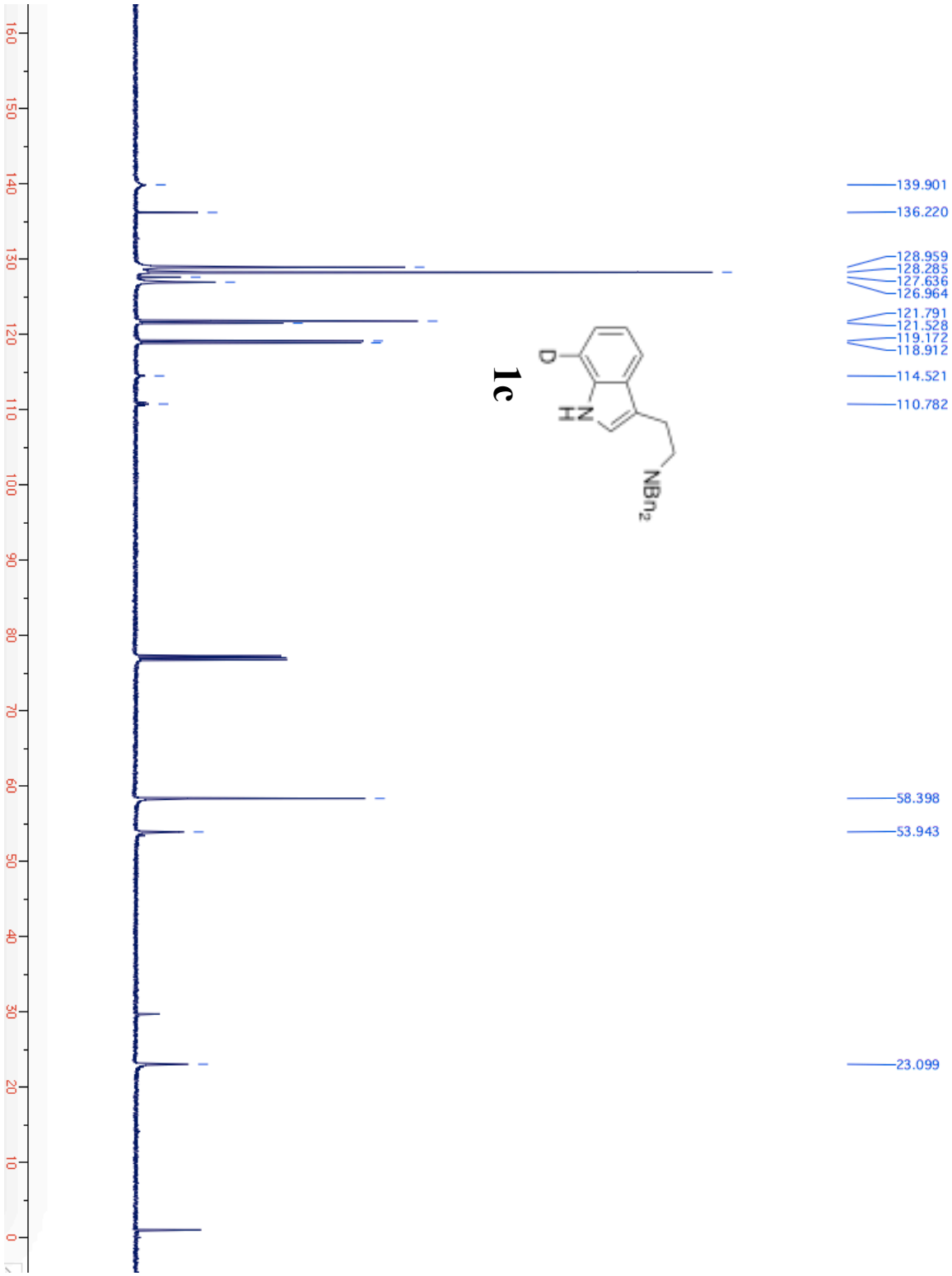


Figure AII.6: [1]HNMR spectrum of **Probe 1**.

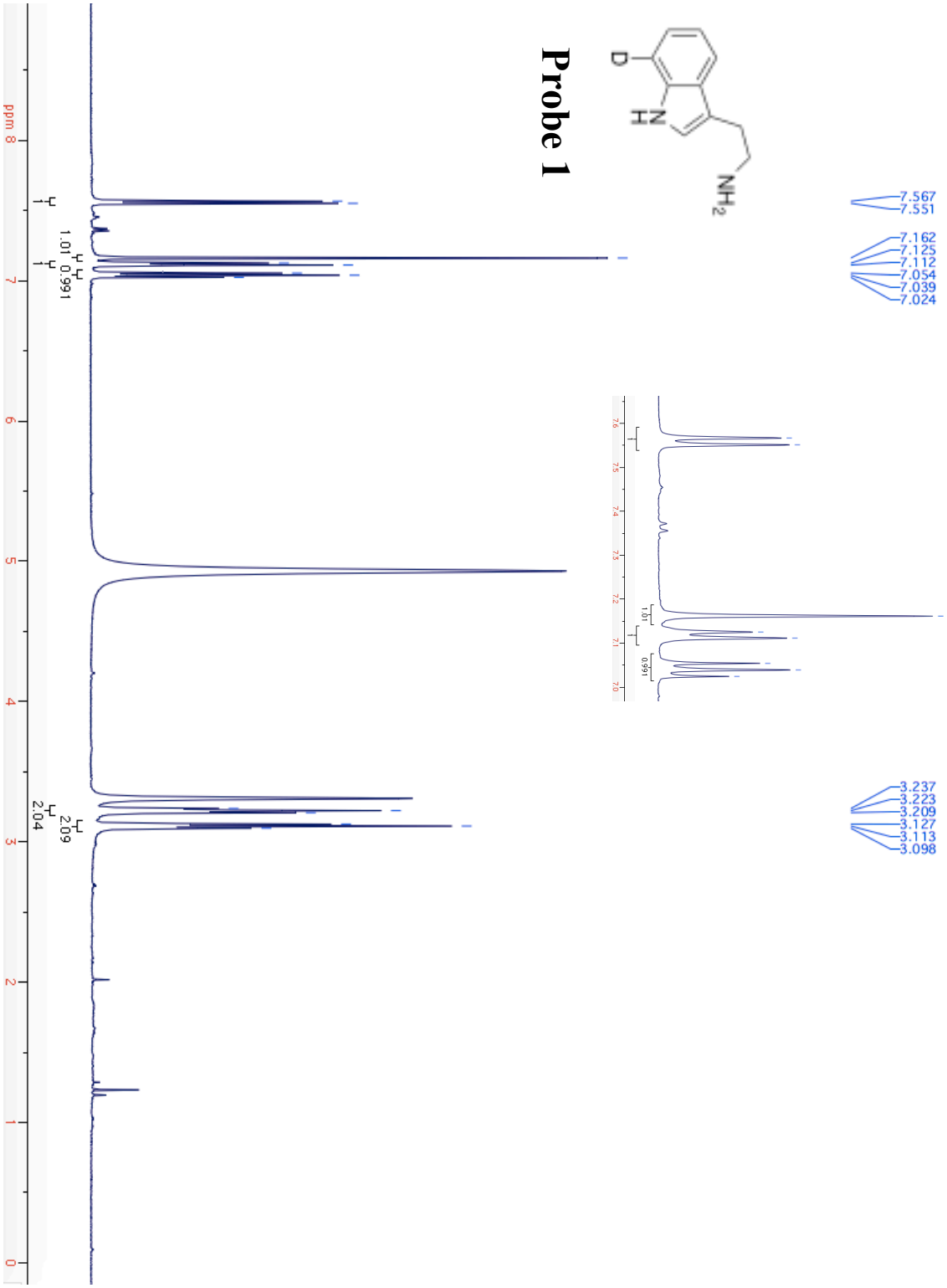


Figure AII.7: ^{13}CNMR spectrum of **Probe 1**.

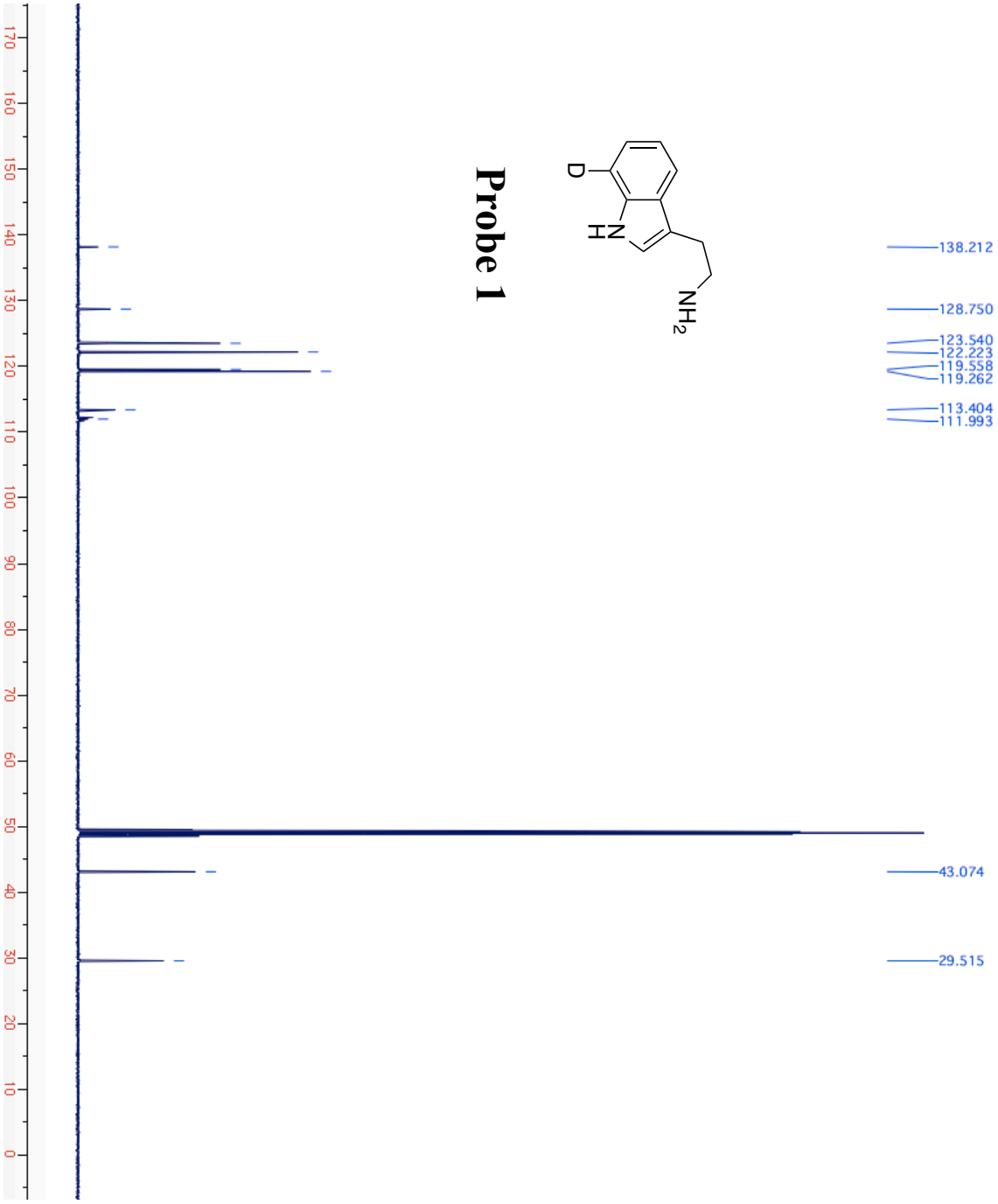


Figure AII.8: [1]HNMR spectrum of **2a**.

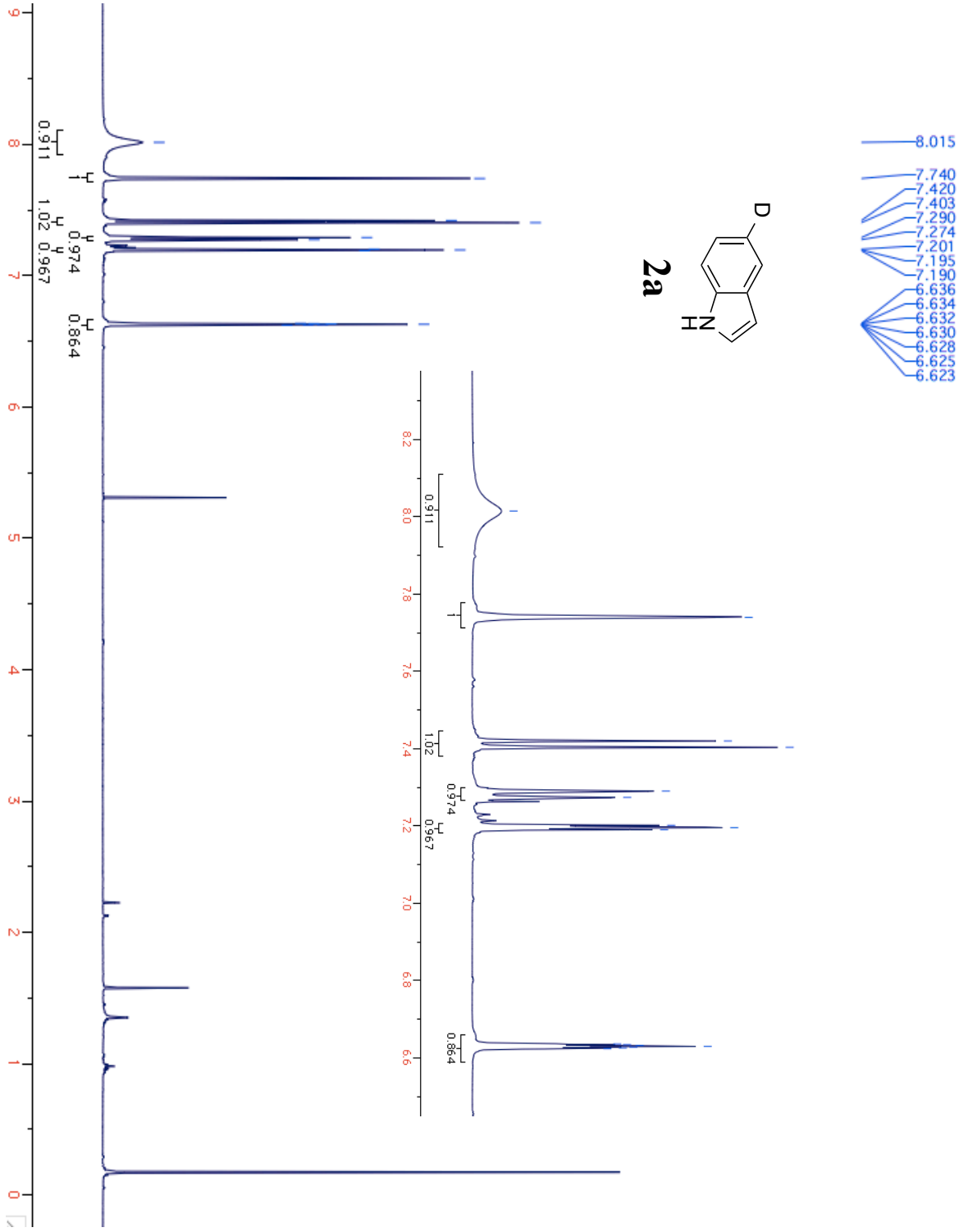


Figure AII.9: ^{13}CNMR spectrum of **2a**.

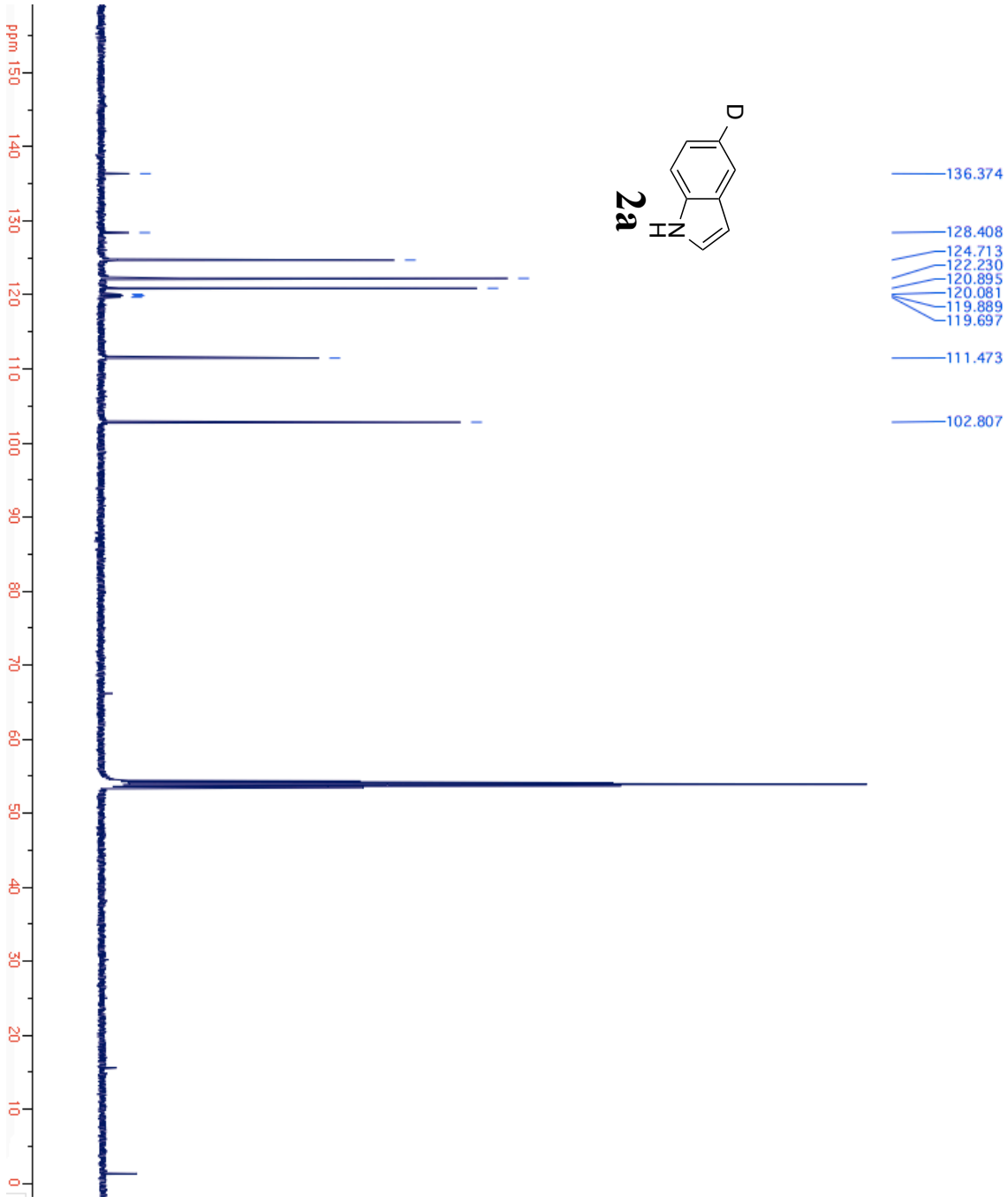


Figure AII.10: [1]HNMR spectrum of **2b**.

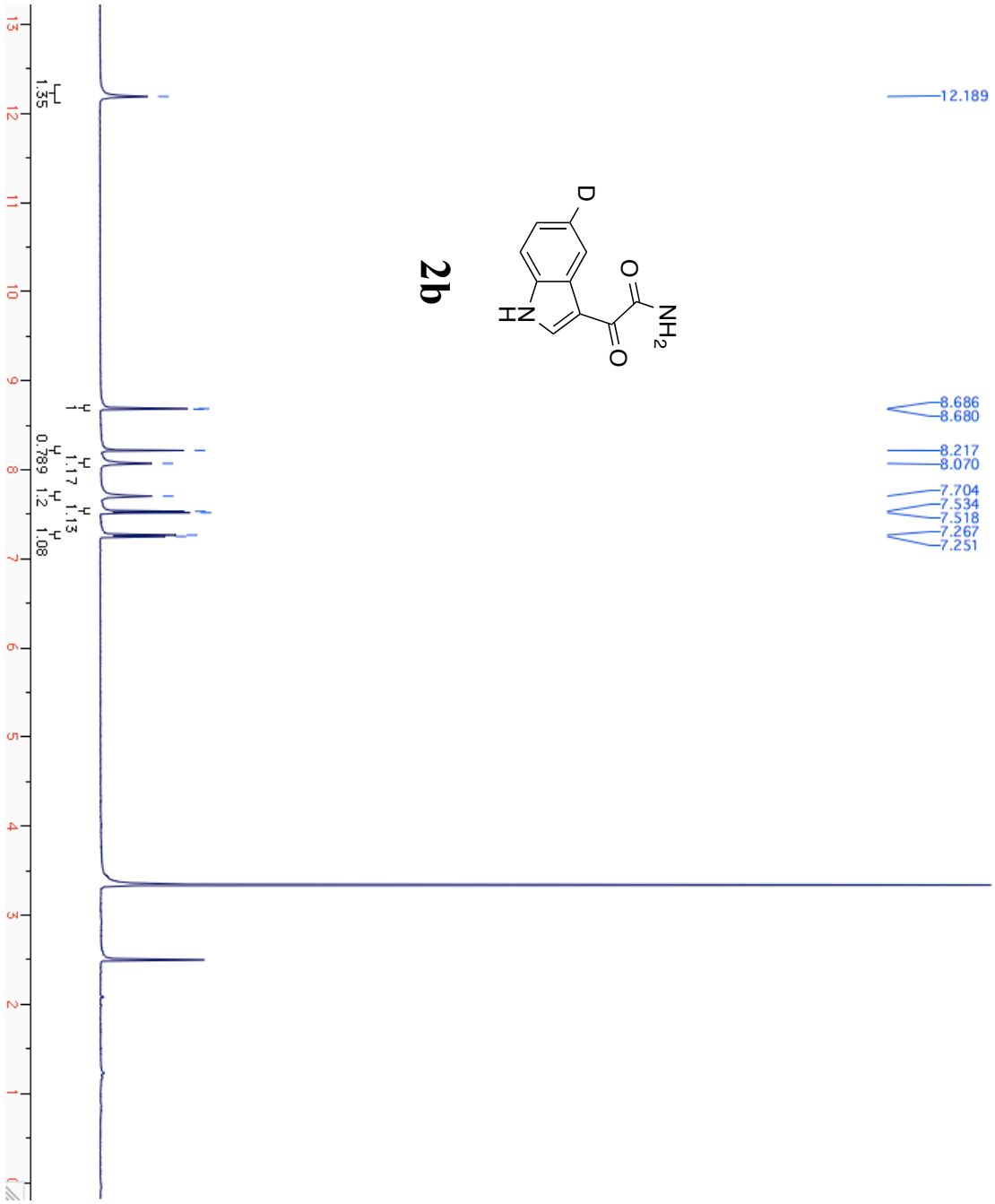


Figure AII.11: ^{13}CNMR spectrum of **2b**.

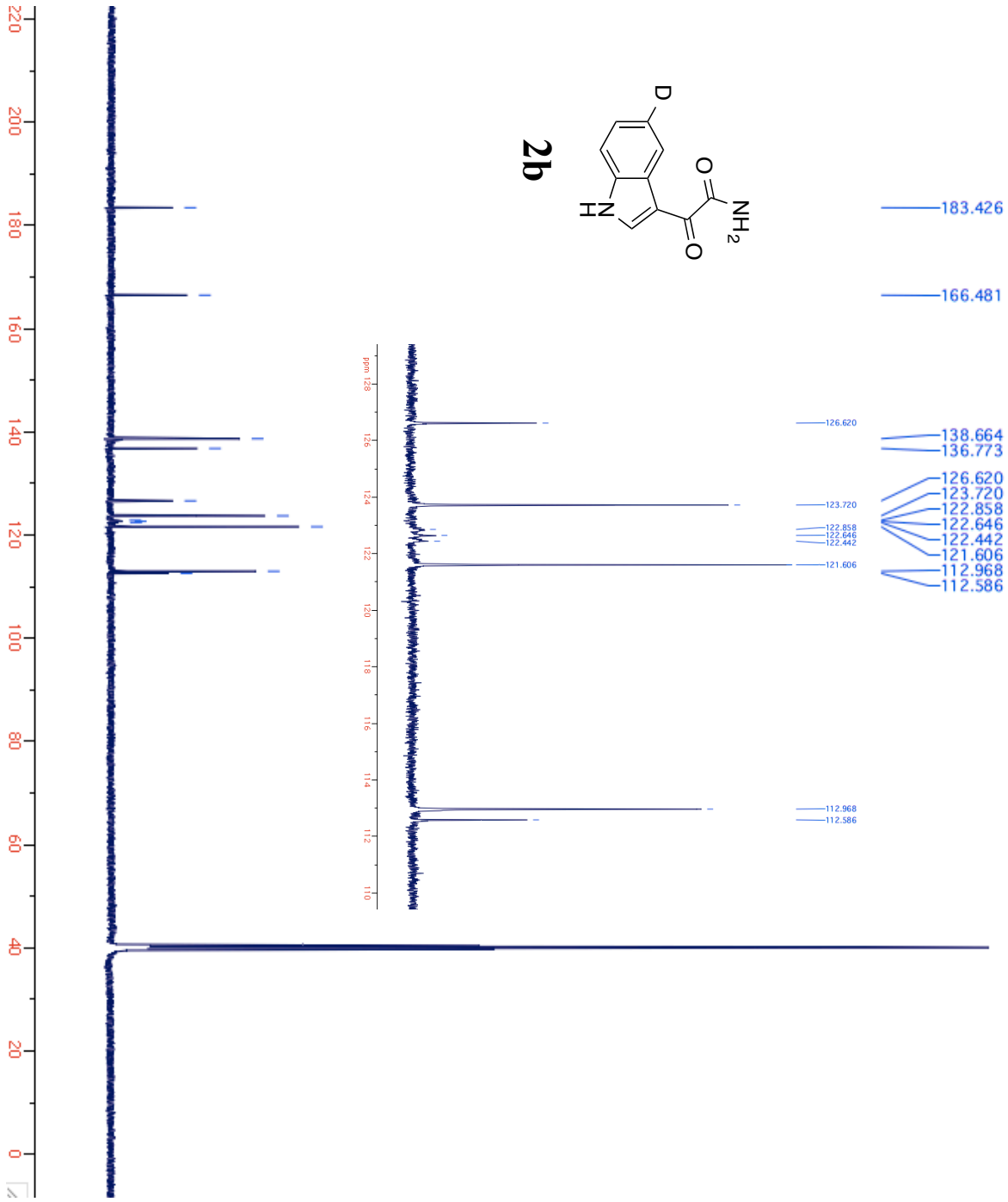


Figure AII.12: [1]HNMR spectrum of **Probe 2**.

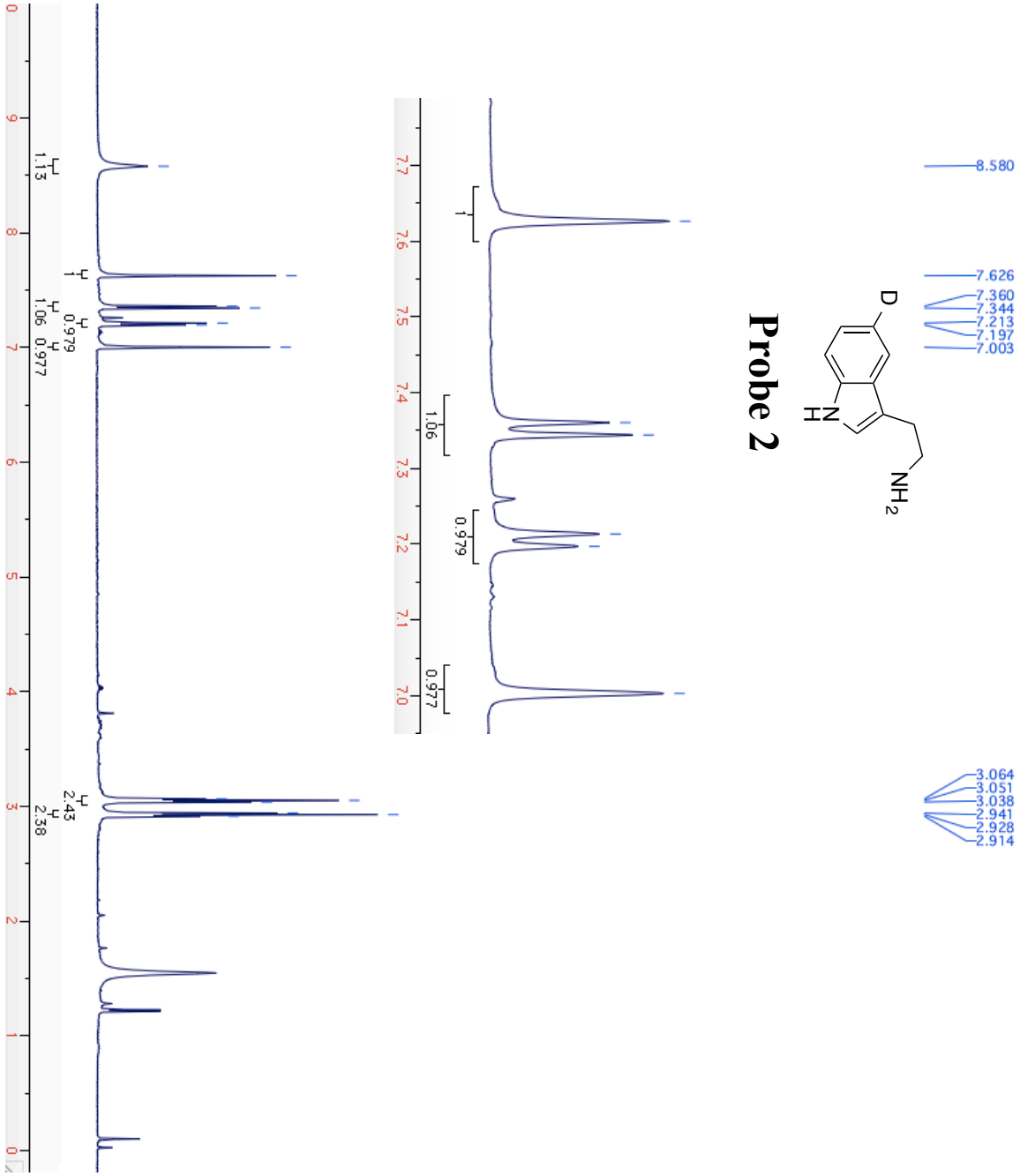
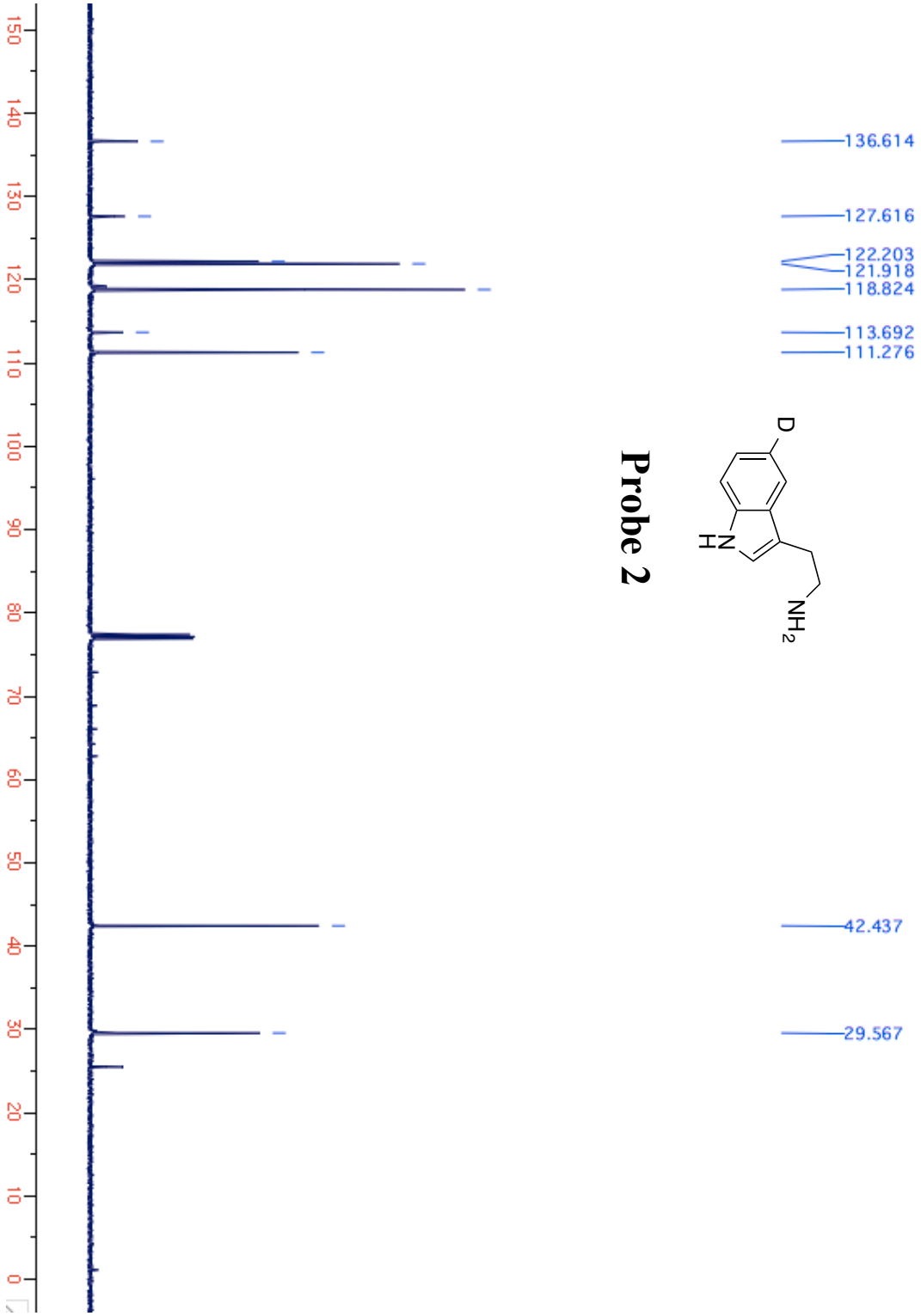


Figure AII.13: ^{13}CNMR spectrum of **Probe 2**.



Appendix III

NMR Spectra for Compounds from Chapter 4

Figure AIII.1: [1]HNMR spectrum of **3**.

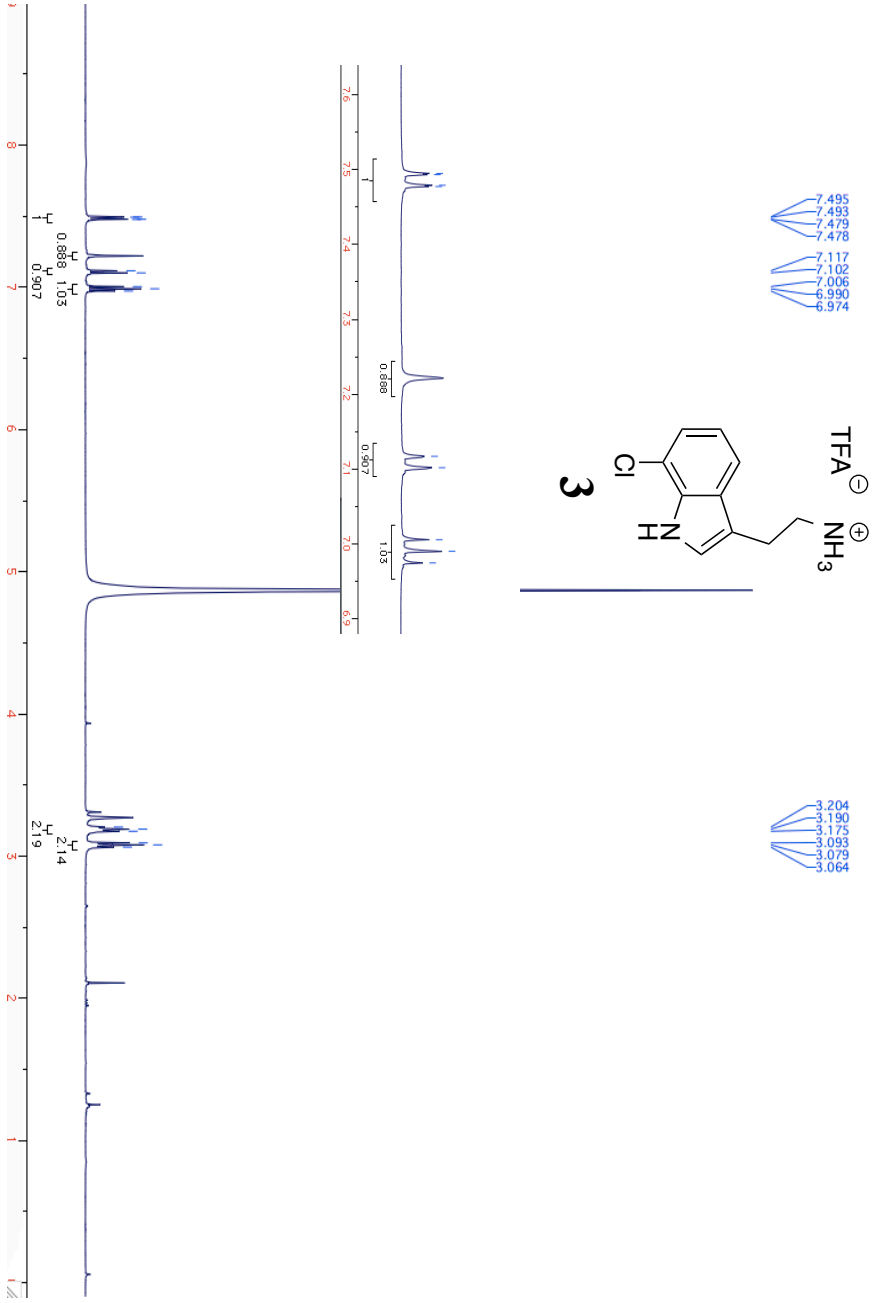


Figure AIII.2: [1]HNMR spectrum of **5**.

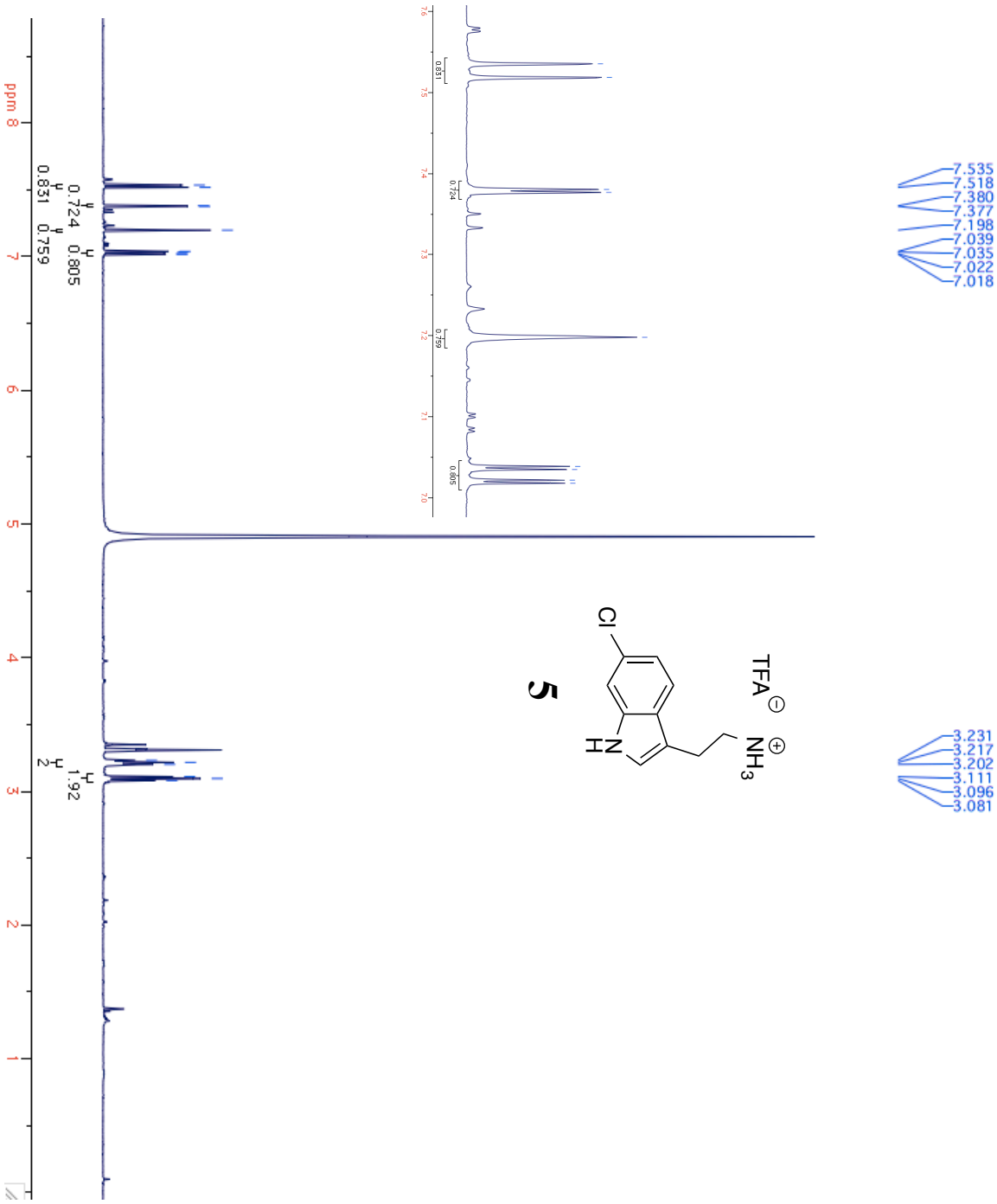


Figure AIII.3: [1]HNMR spectrum of **6**.

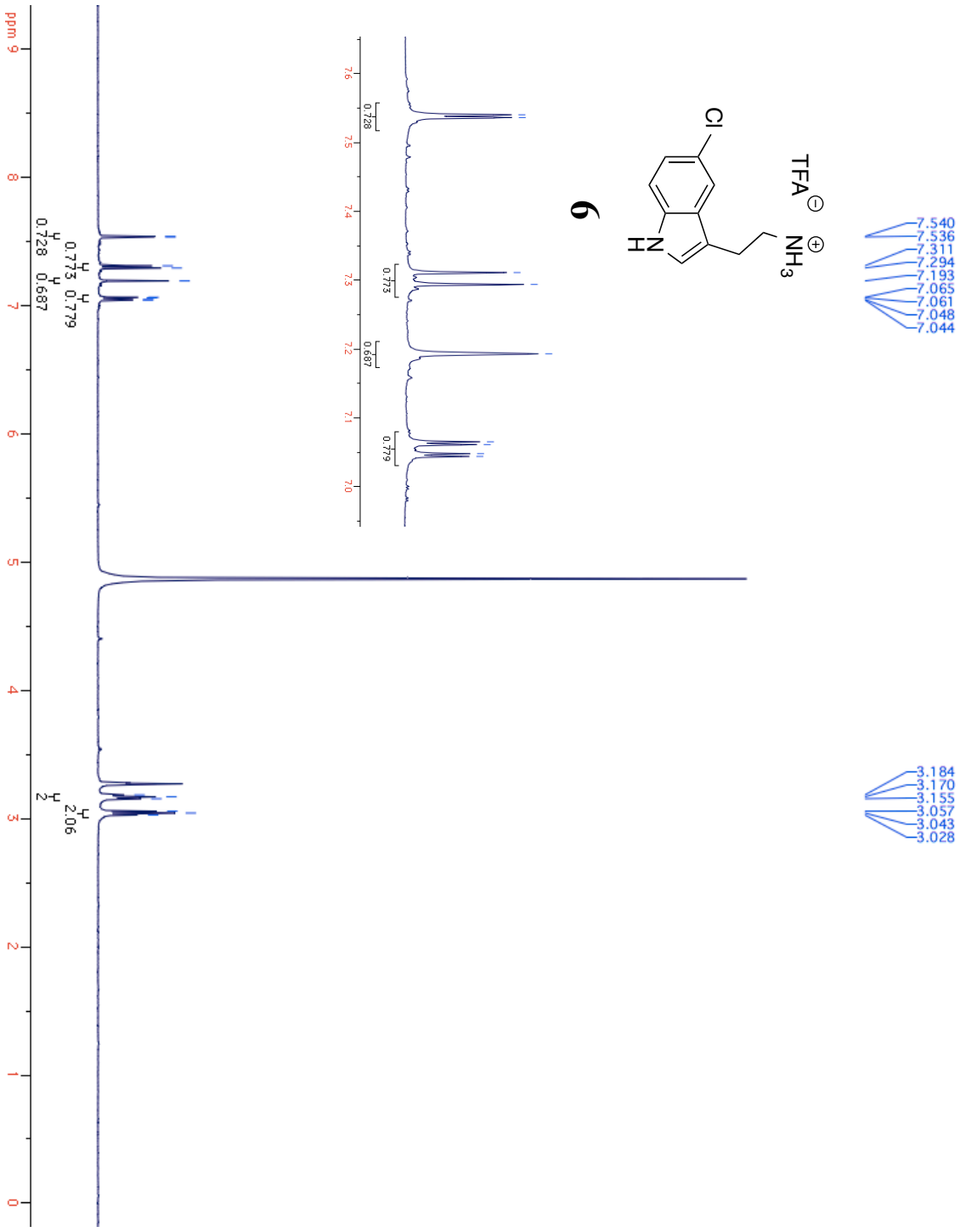


Figure AIII.4: [1]HNMR spectrum of **7**.

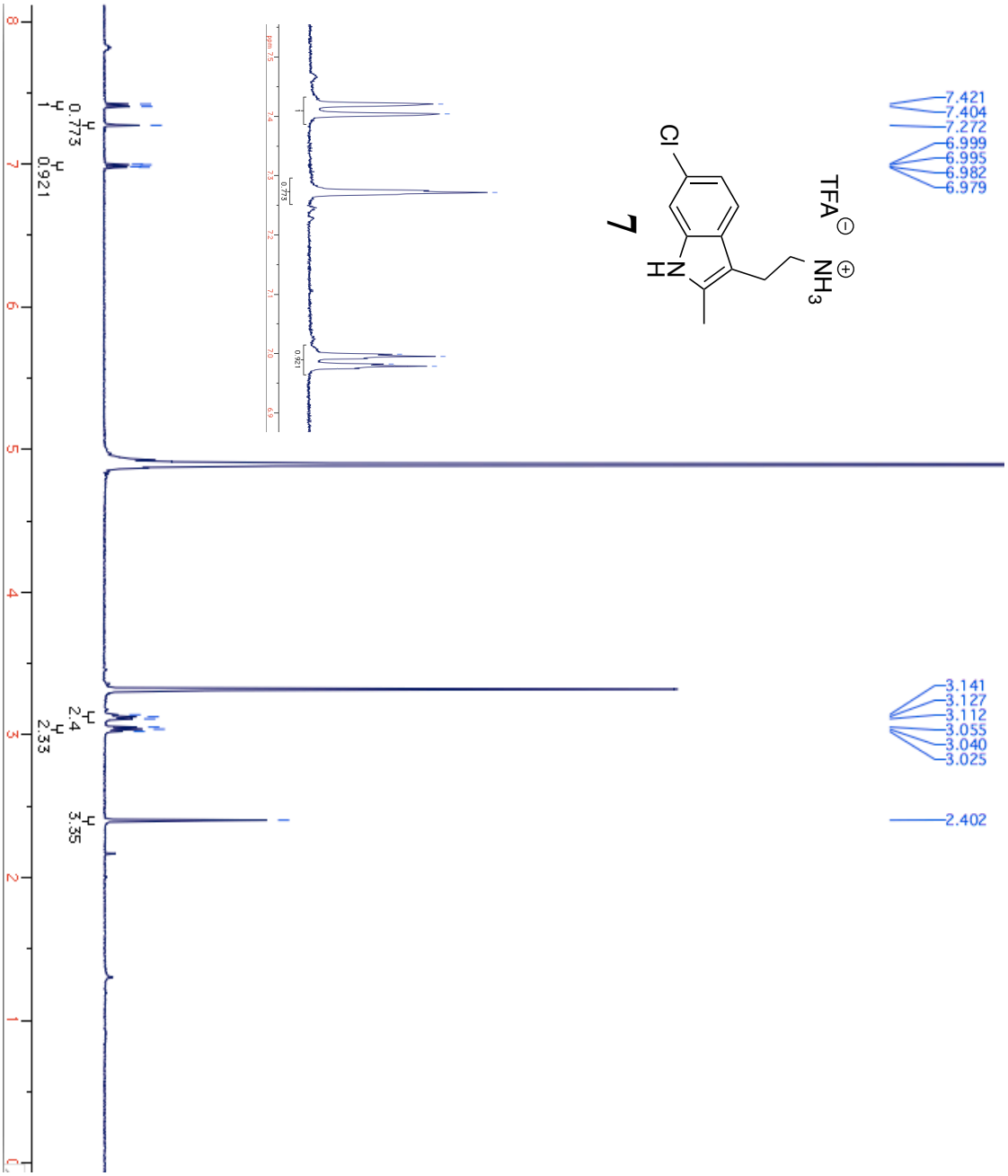


Figure AIII.5: [1]HNMR spectrum of **7** and **8**.

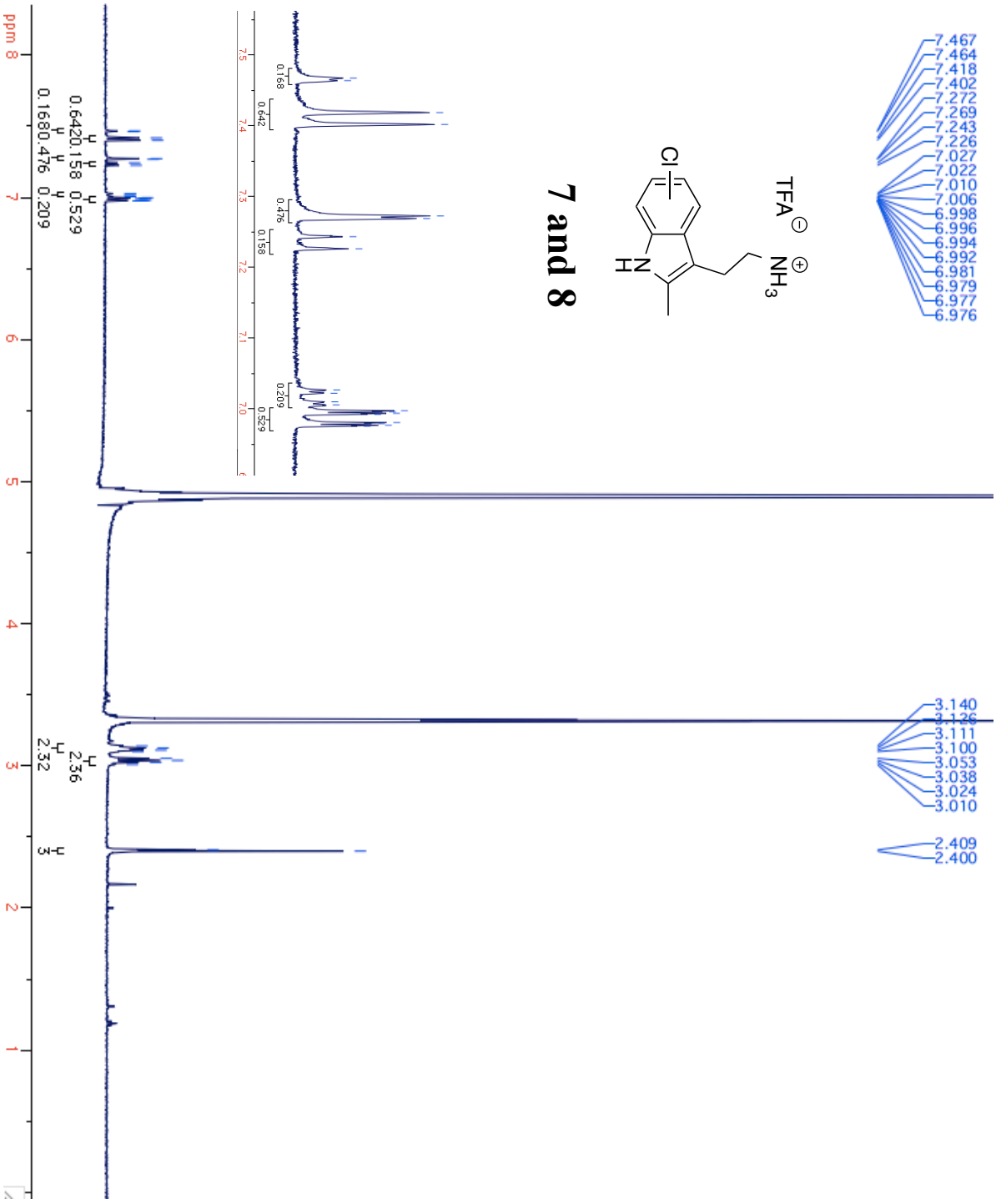


Figure AIII.6: [1]HNMR spectrum of **9**.

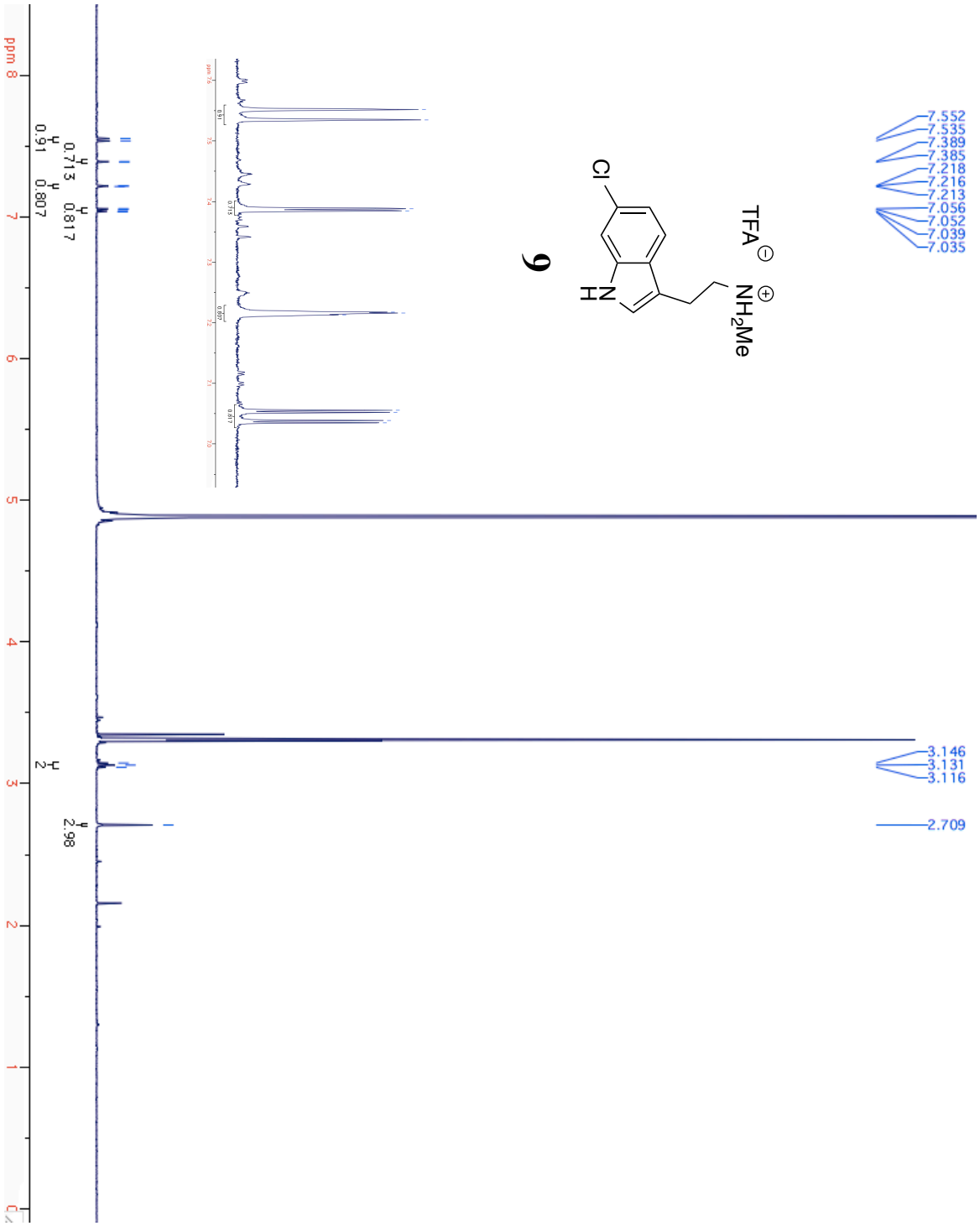


Figure AIII.7: [1]HNMR spectrum of **10**.

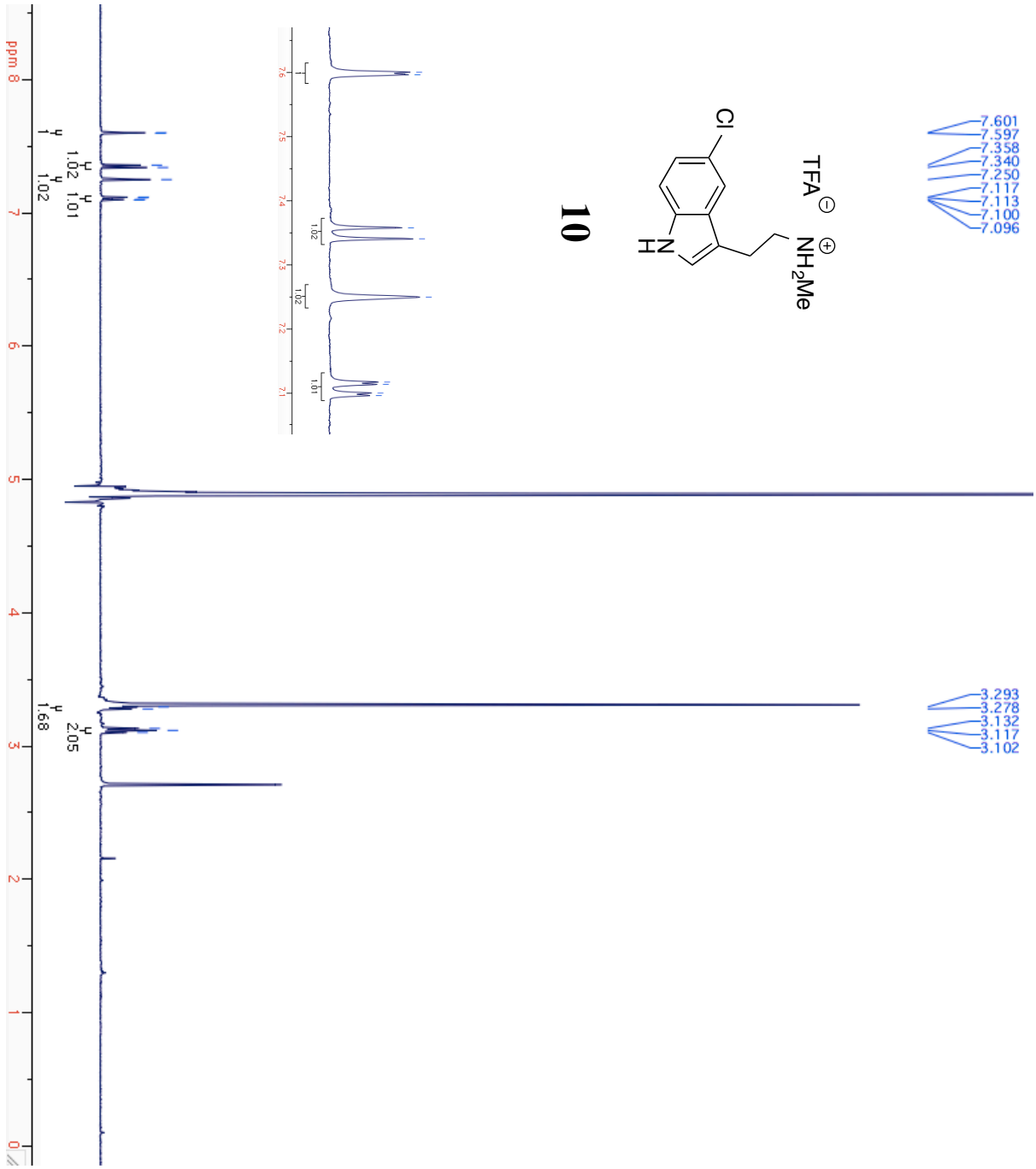


Figure AIII.8: [1]HNMR spectrum of **11**.

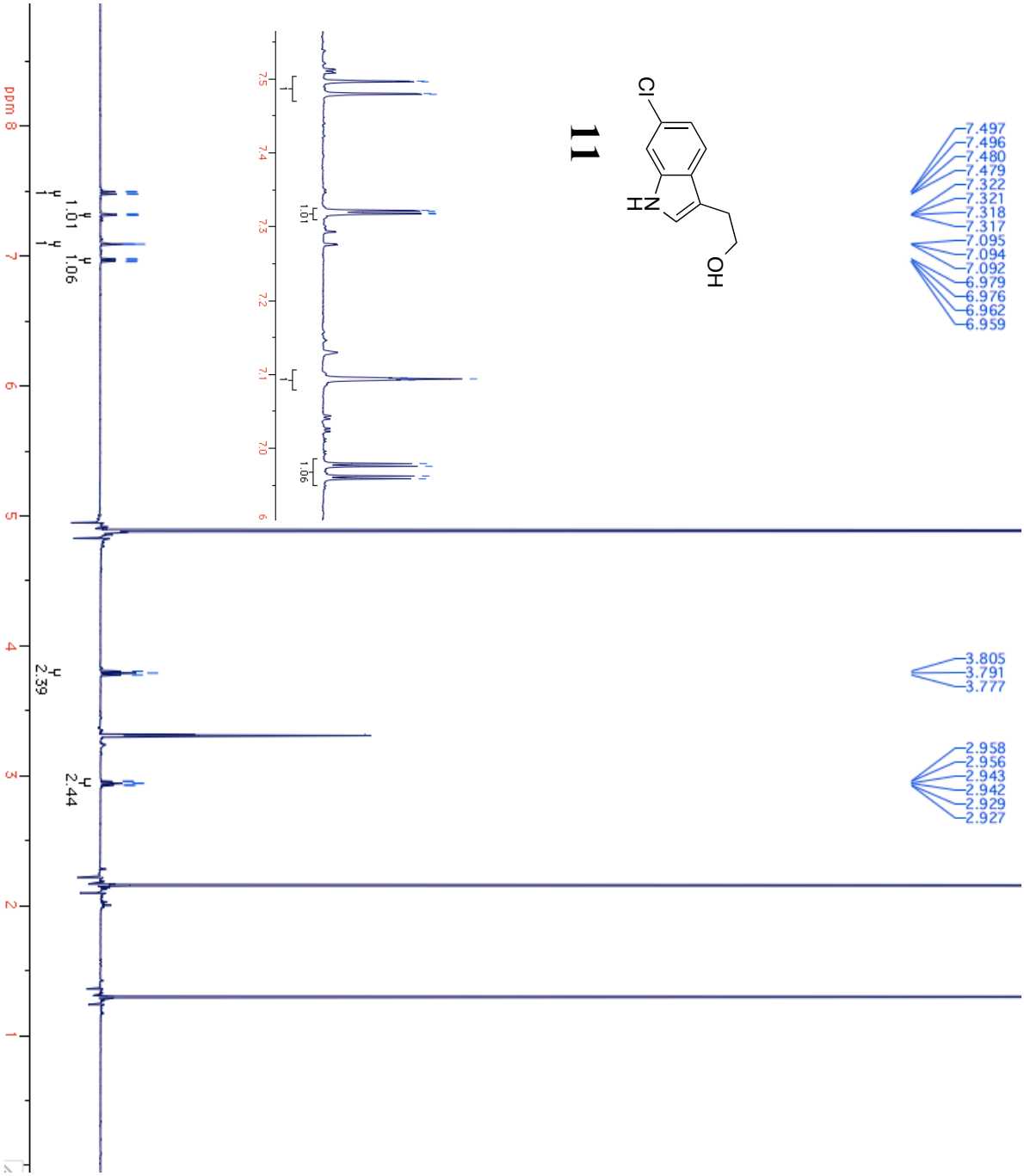


Figure AIII.9: [1]HNMR spectrum of **12**.

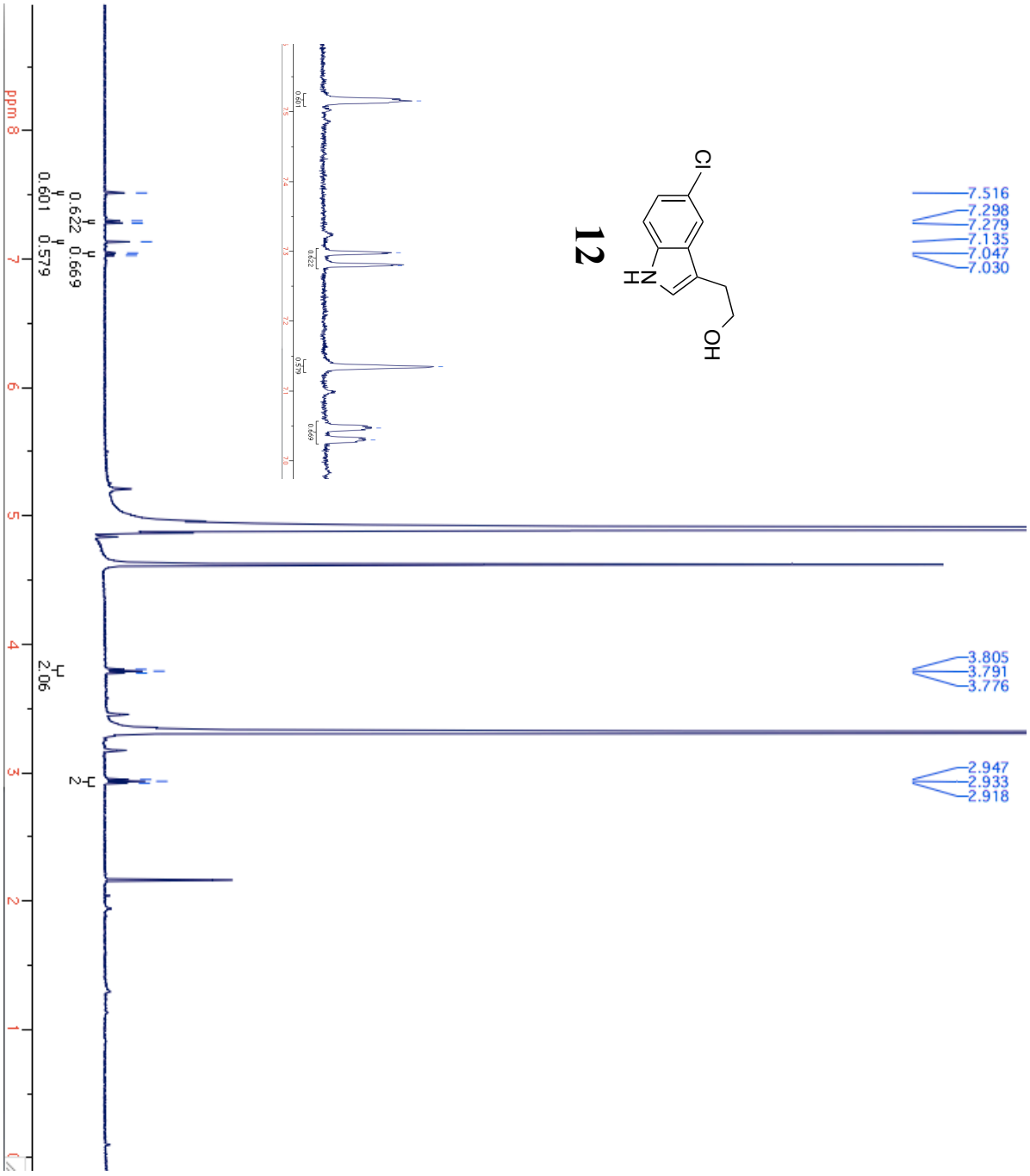
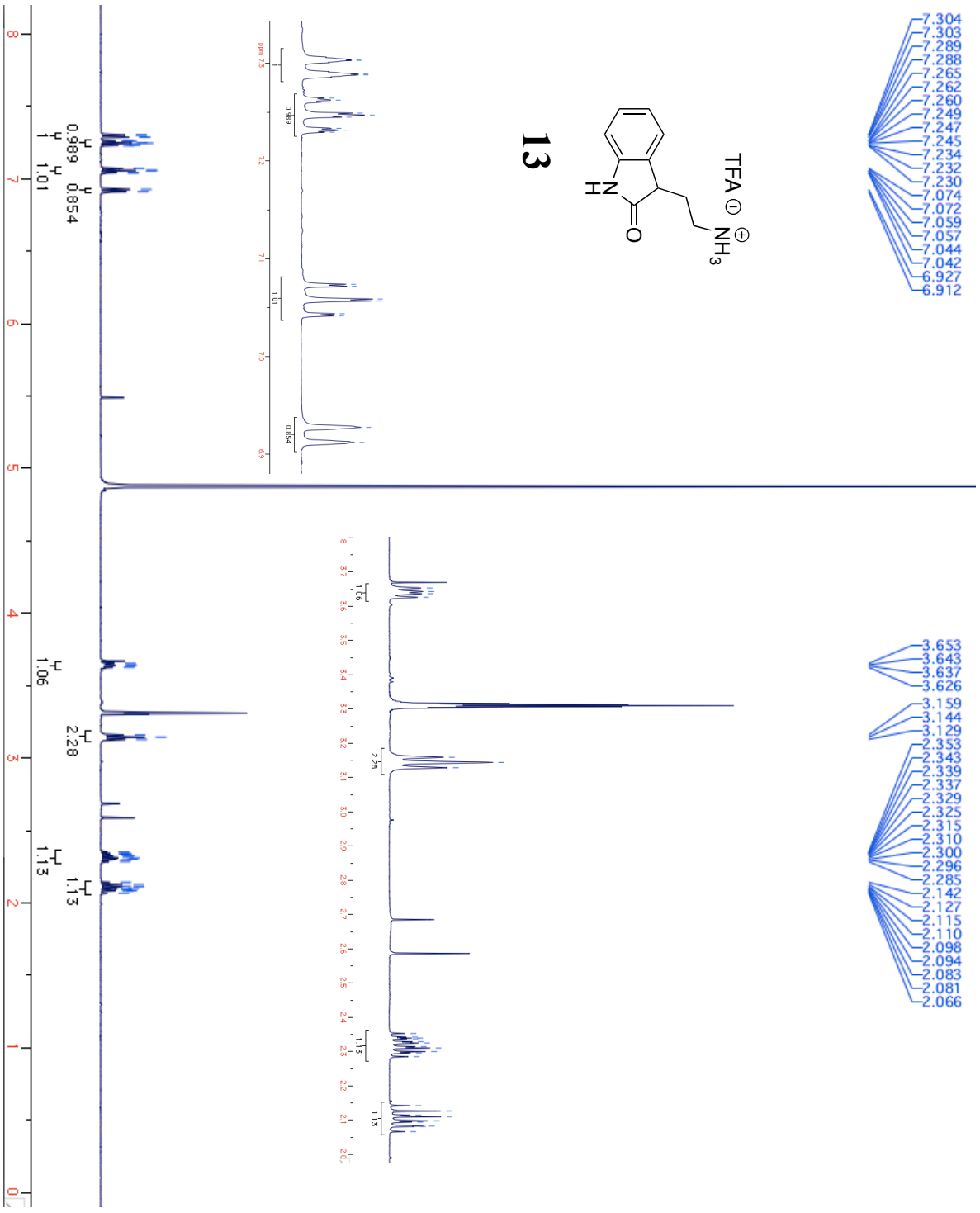


Figure AIII.10: [1]HNMR spectrum of **13**.



Appendix IV

NMR Spectra for Compounds from Chapter 5

Figure AIV.1: ^{1}HNMR spectrum of **29**.

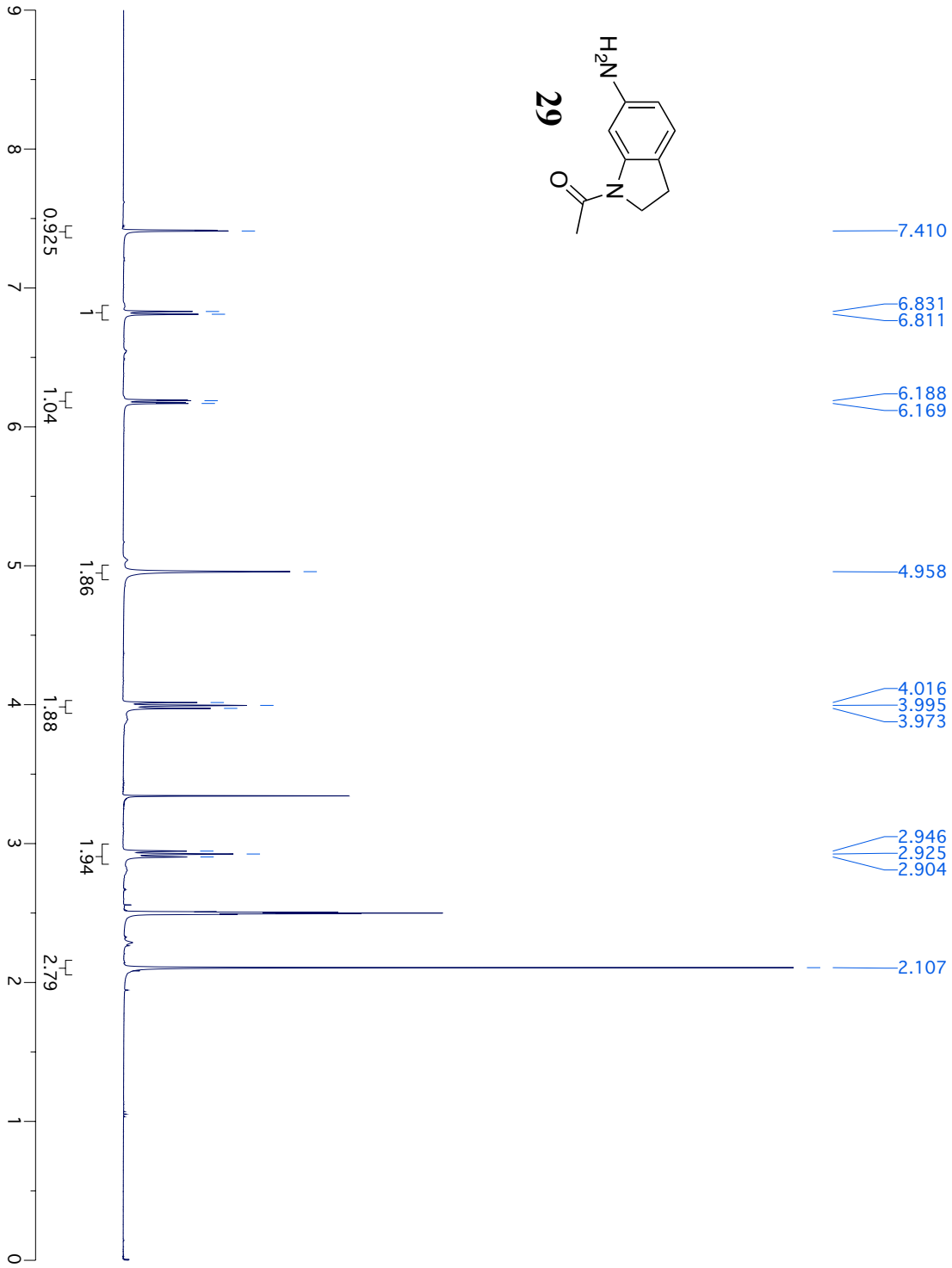


Figure AIV.2: [1]HNMR spectrum of **29a**.

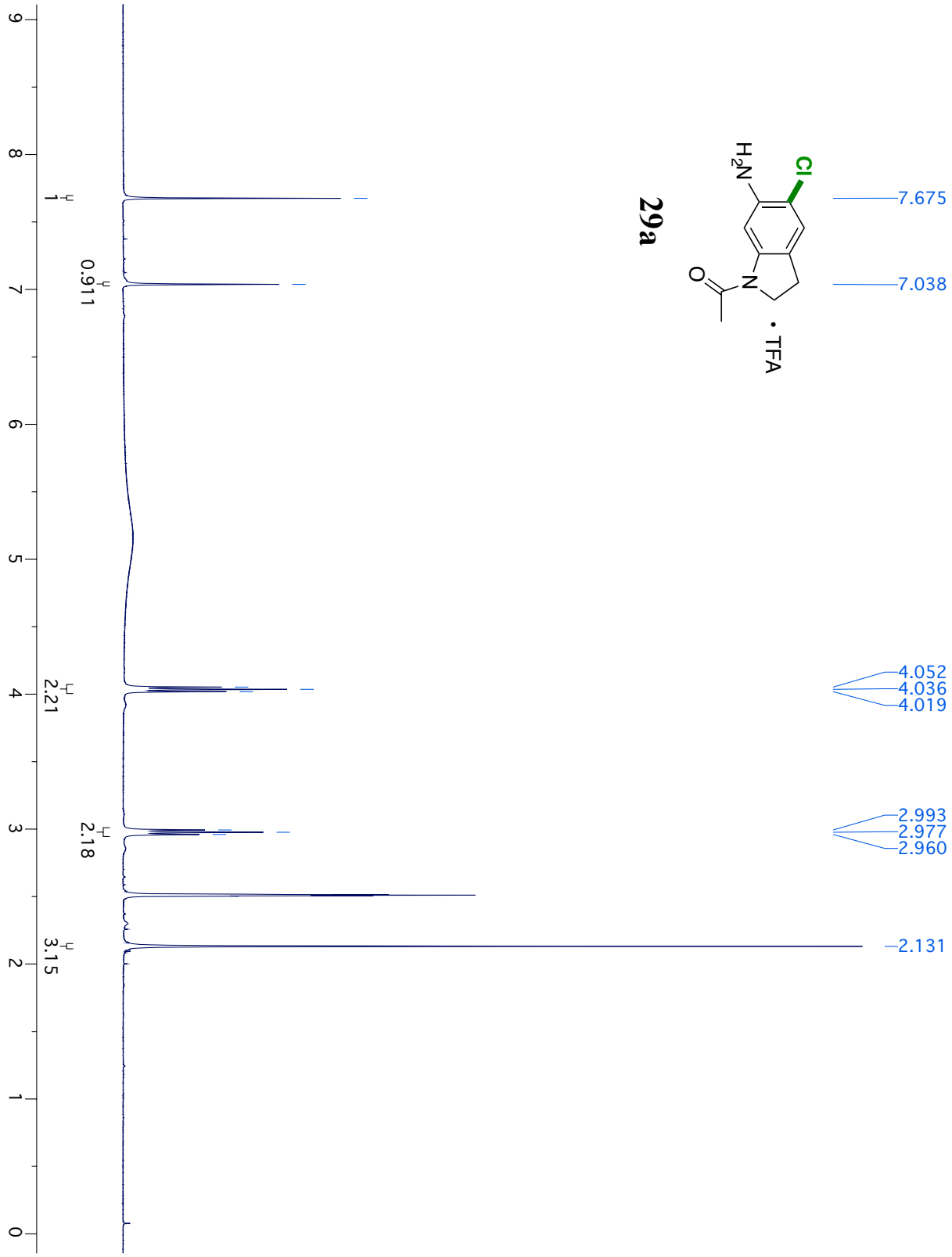


Figure AIV.3: NOSEY spectrum of **29a**.

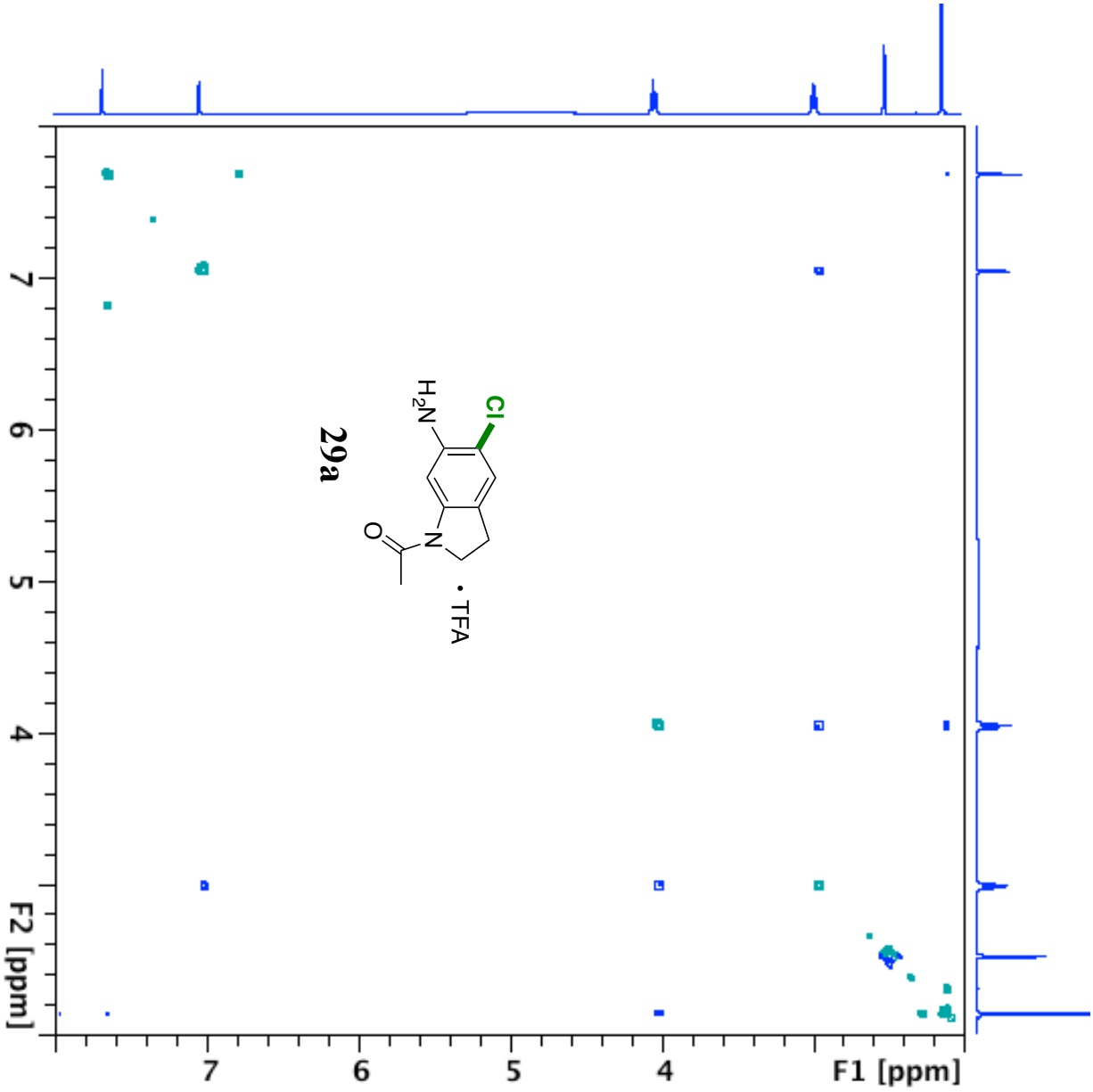


Figure AIV.4: [1]HNMR spectrum of **30**.

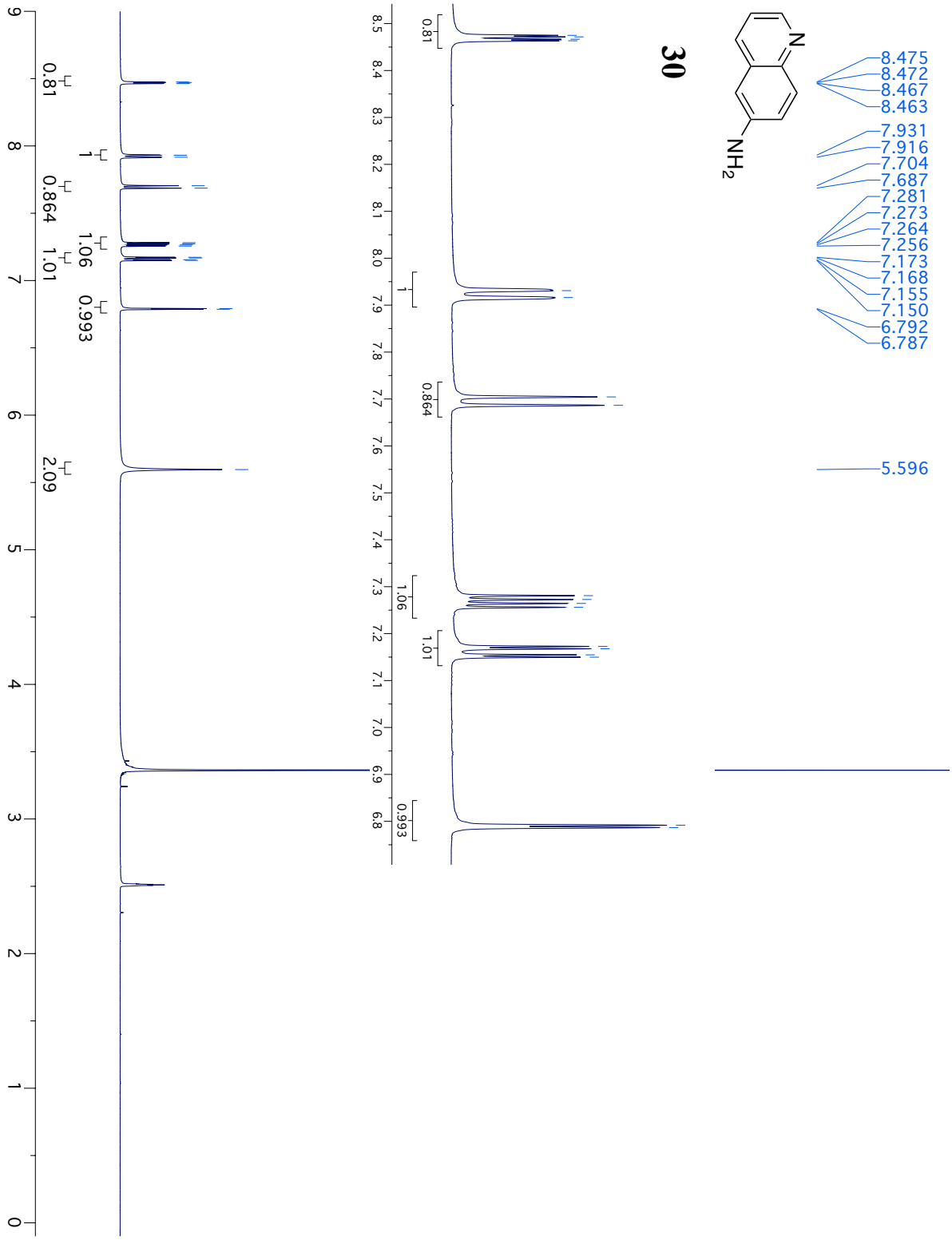


Figure AIV.5: [1]HNMR spectrum of **30a**.

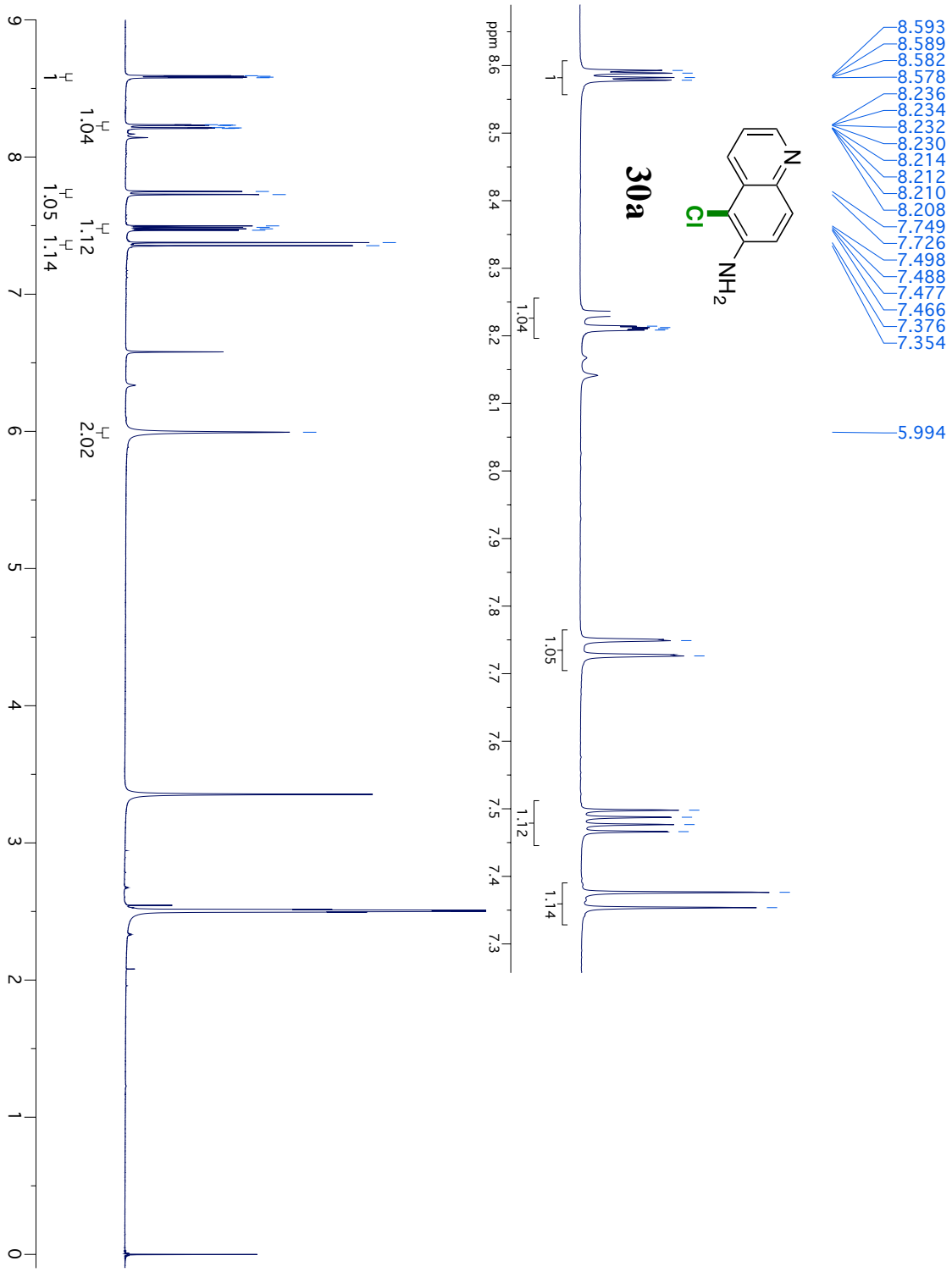


Figure AIV.6: [1]HNMR spectrum of **31**.

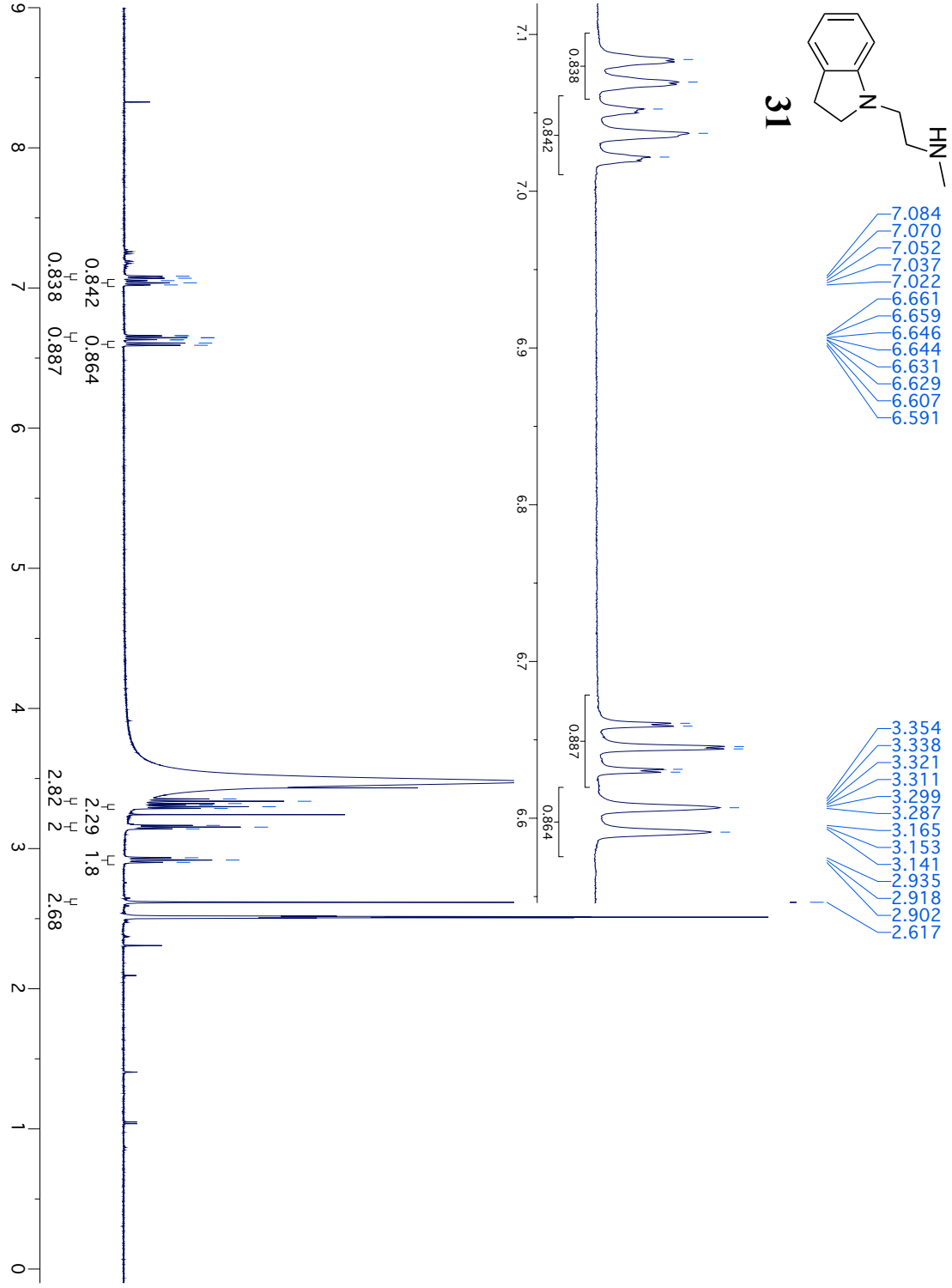


Figure AIV.7: [1]HNMR spectrum of **31a**.

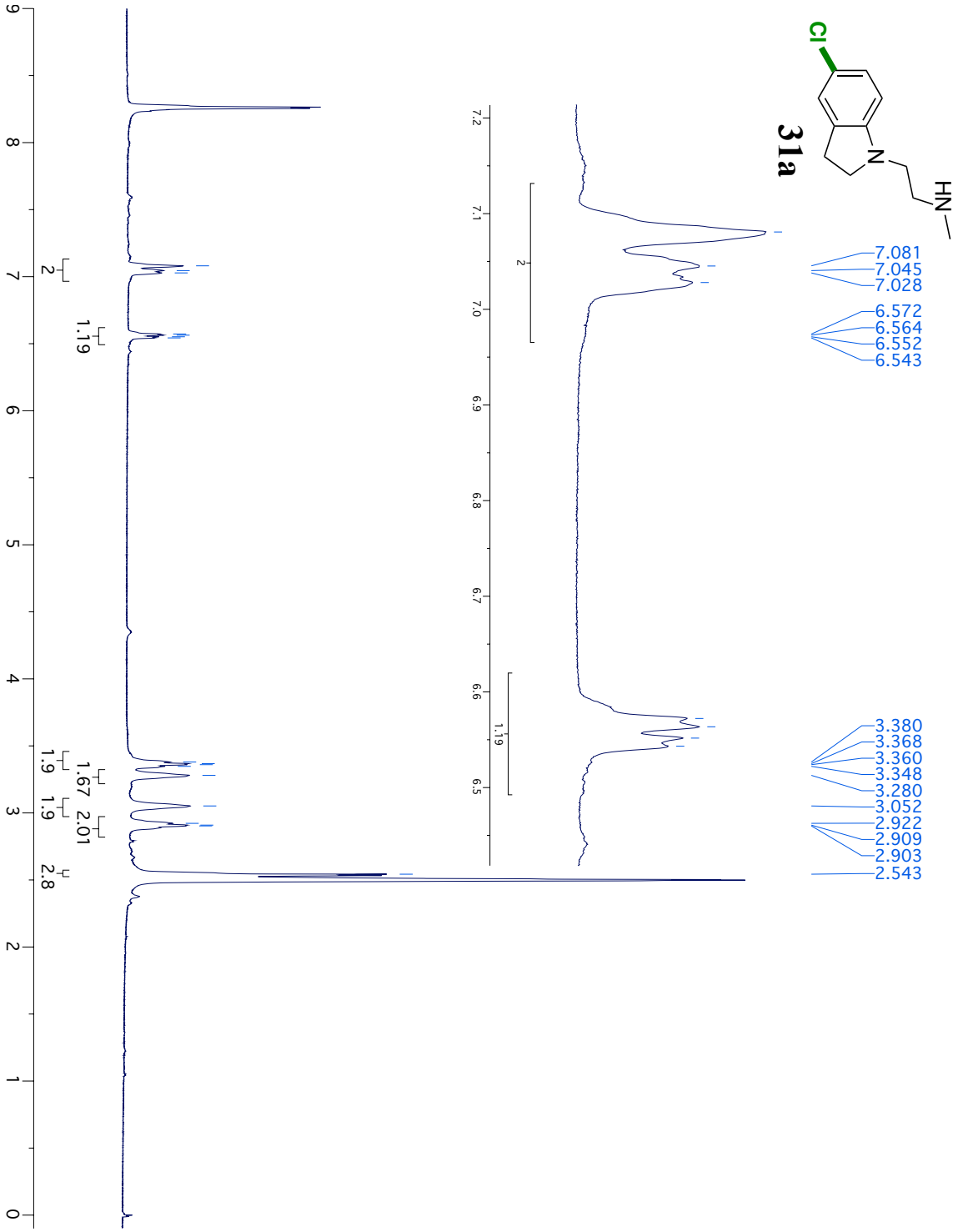


Figure AIV.8: NOSEY spectrum of **31a**.

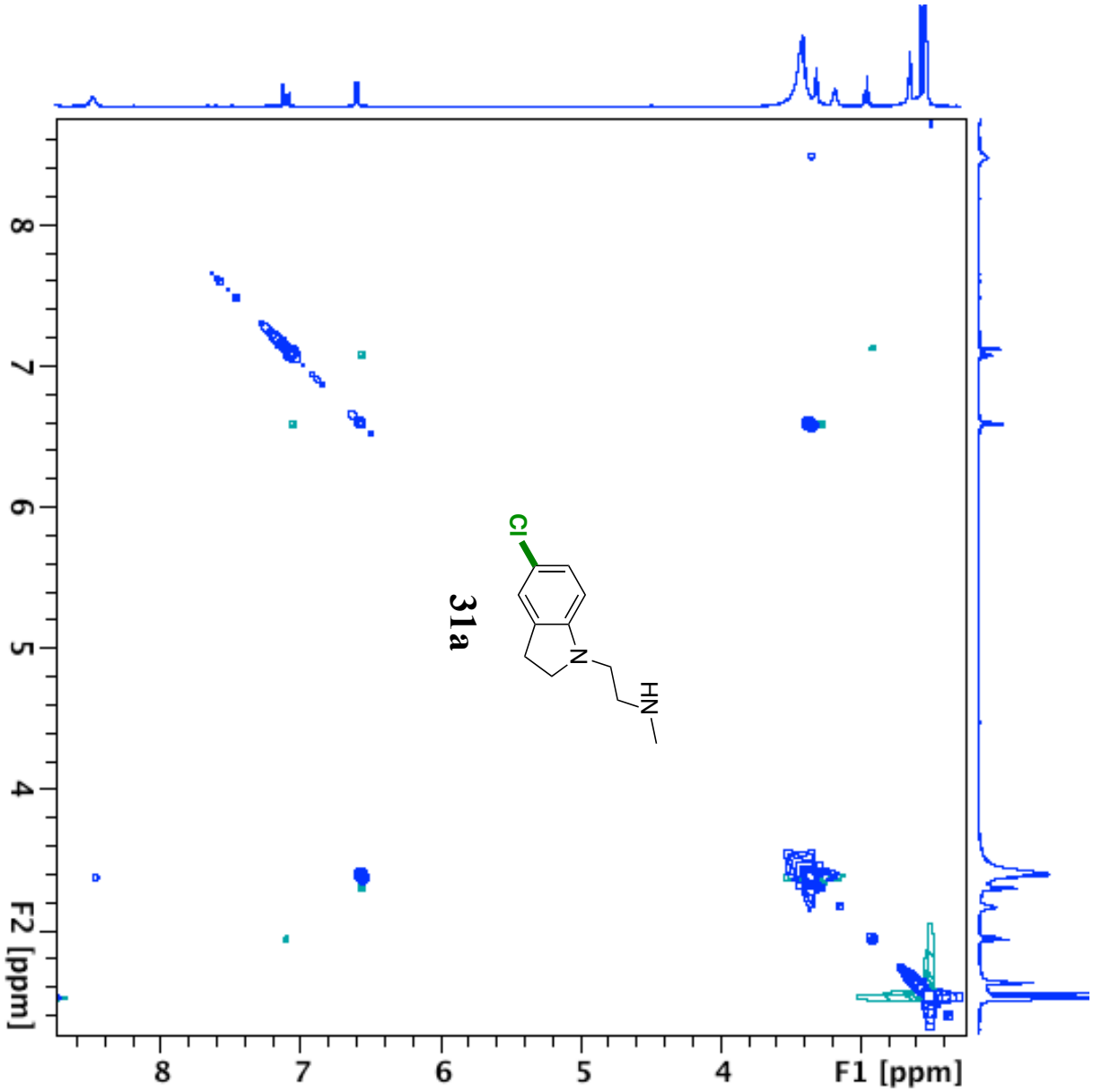


Figure AIV.9: [1]HNMR spectrum of **32**.

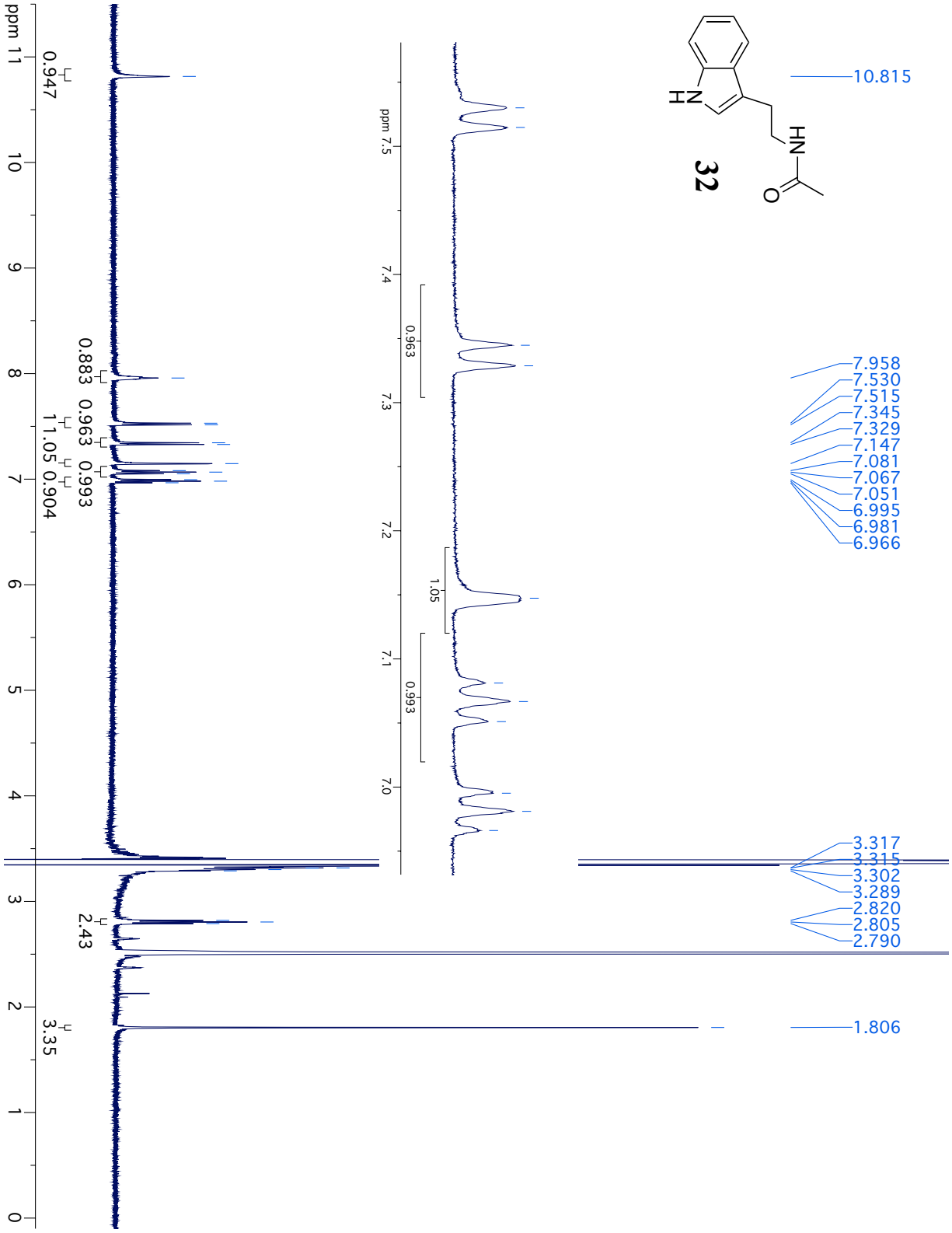


Figure AIV.10: [1]HNMR spectrum of **32a**.

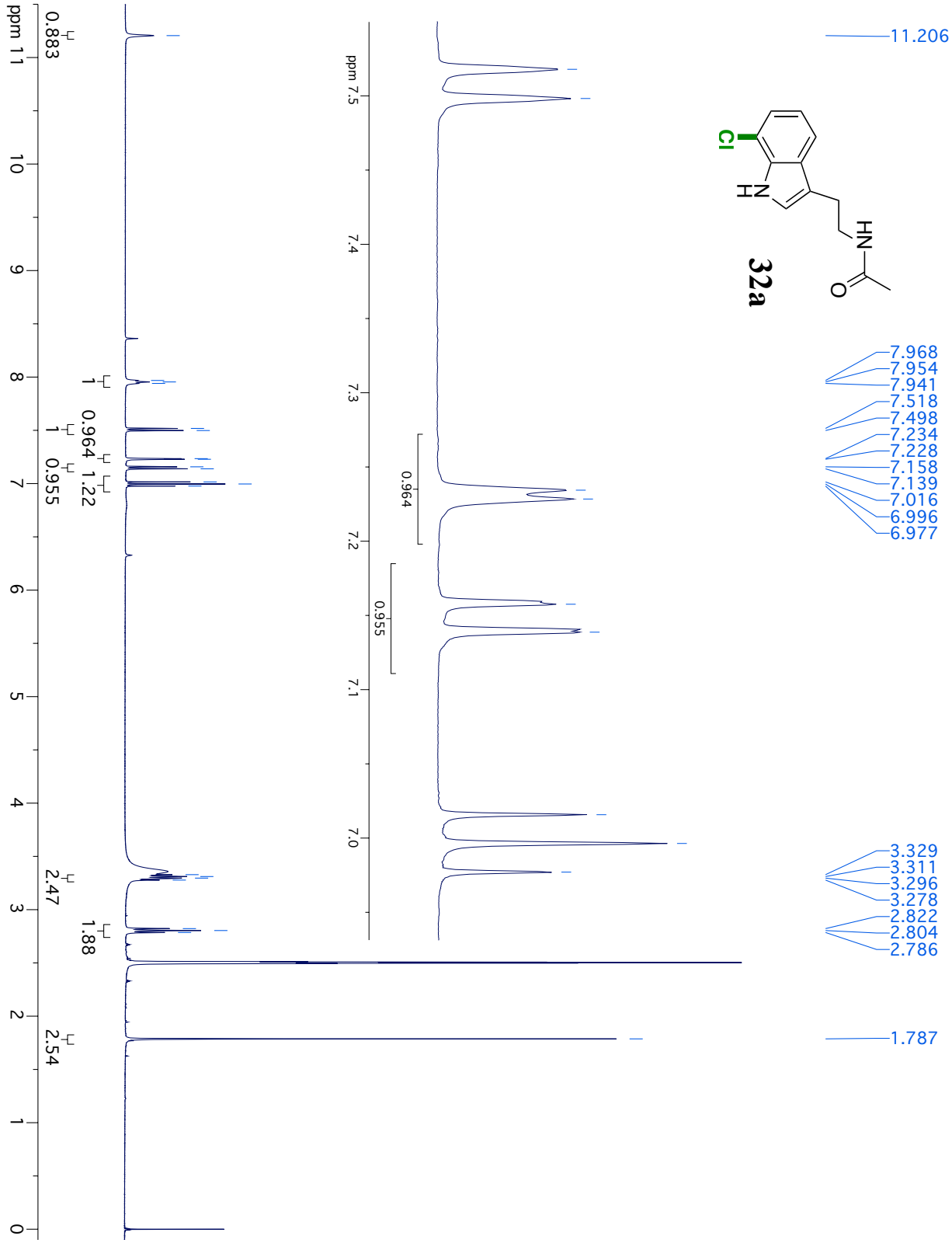


Figure AIV.11: NOSEY spectrum of **32a**.

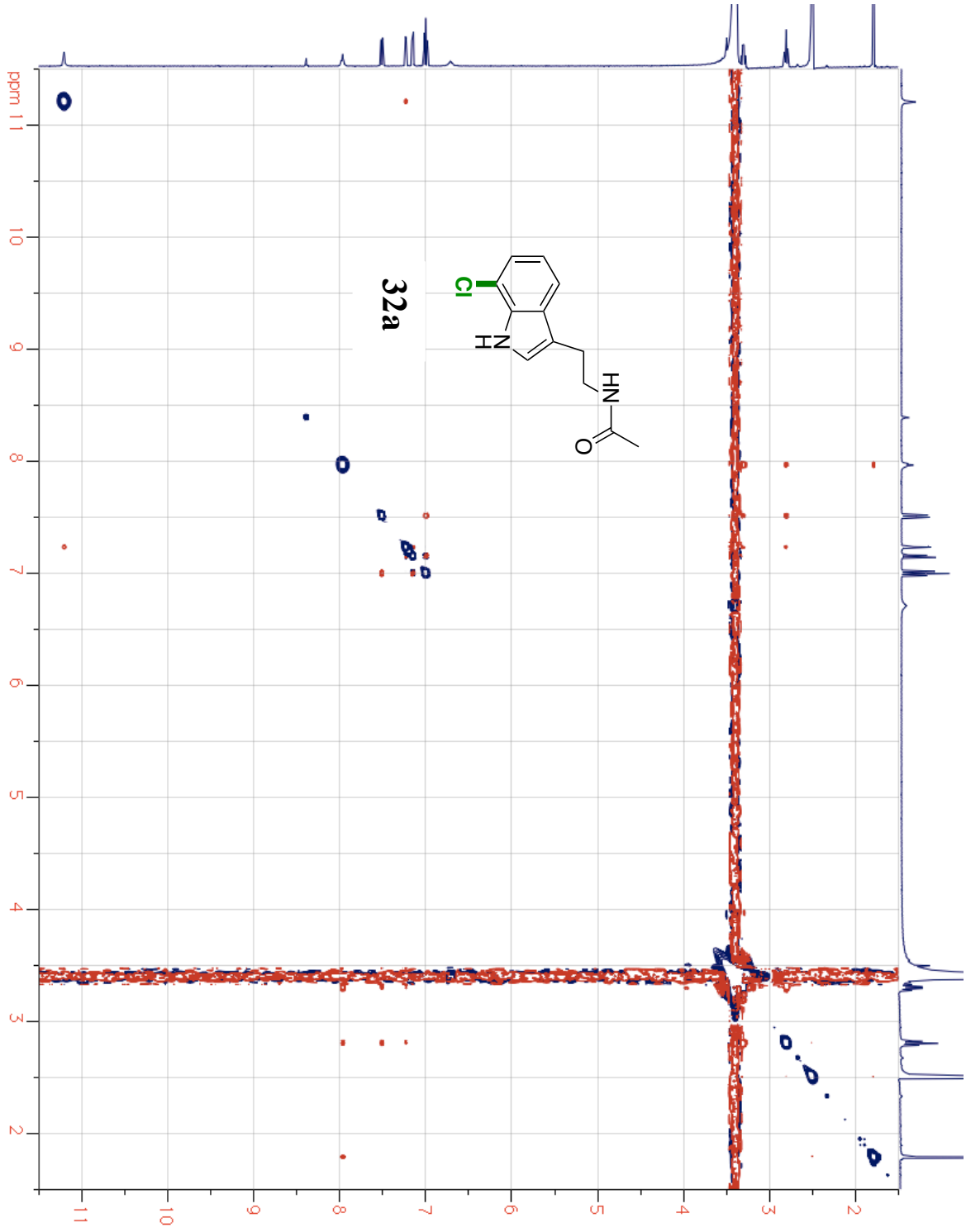


Figure AIV.12: [1]HNMR spectrum of **33**.

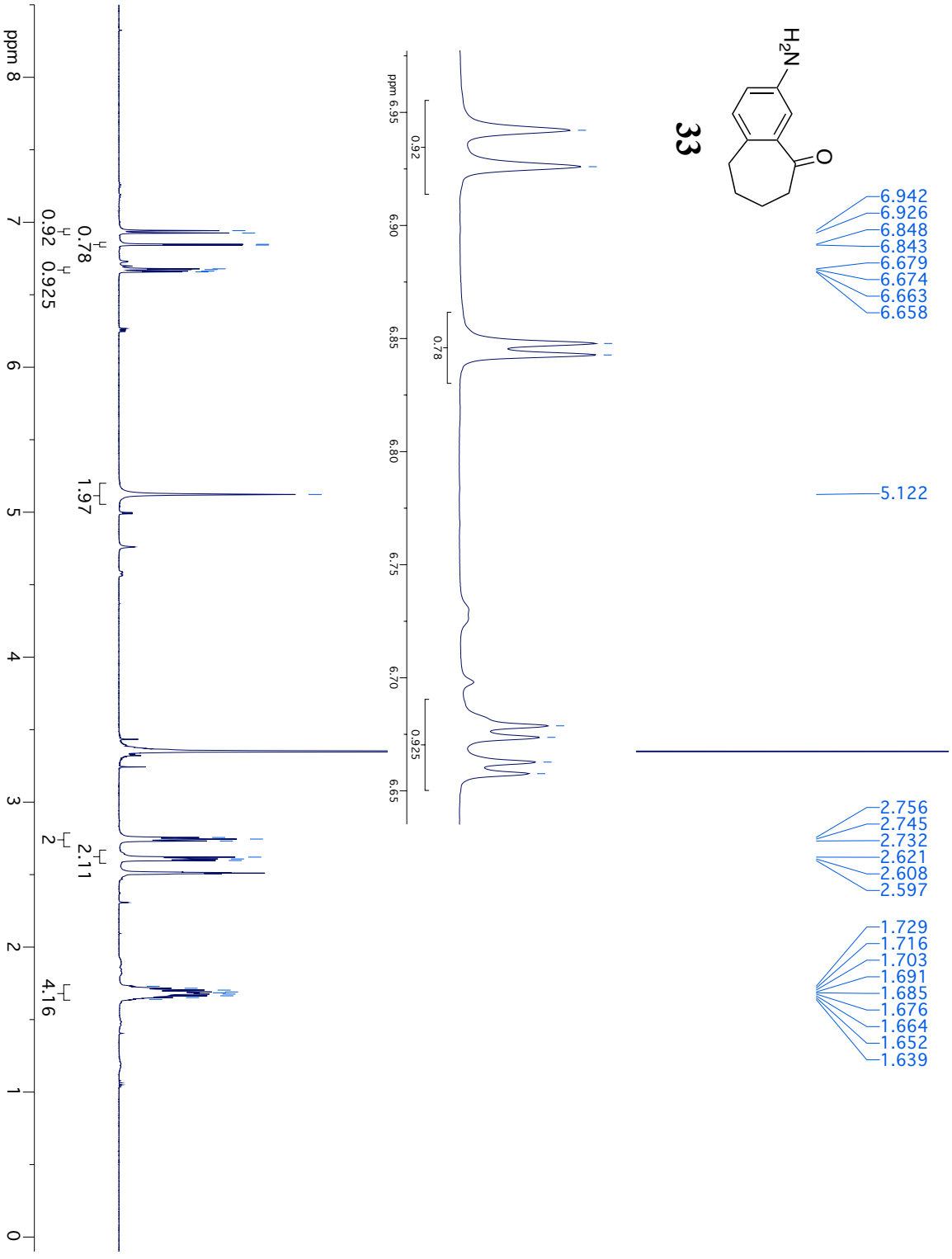


Figure AIV.13: [1]HNMR spectrum of **33a**.

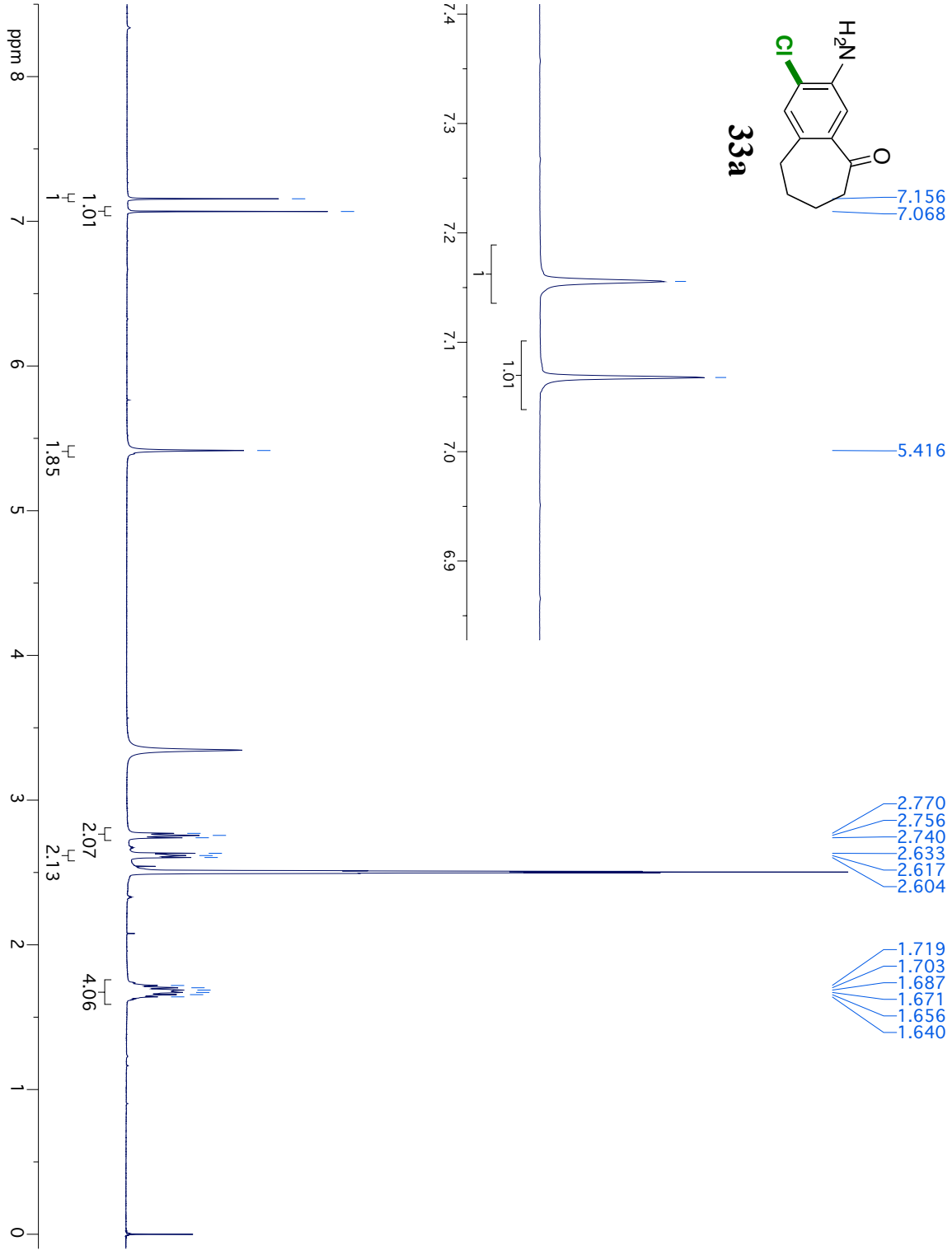


Figure AIV.14: NOSEY spectrum of **33a**.

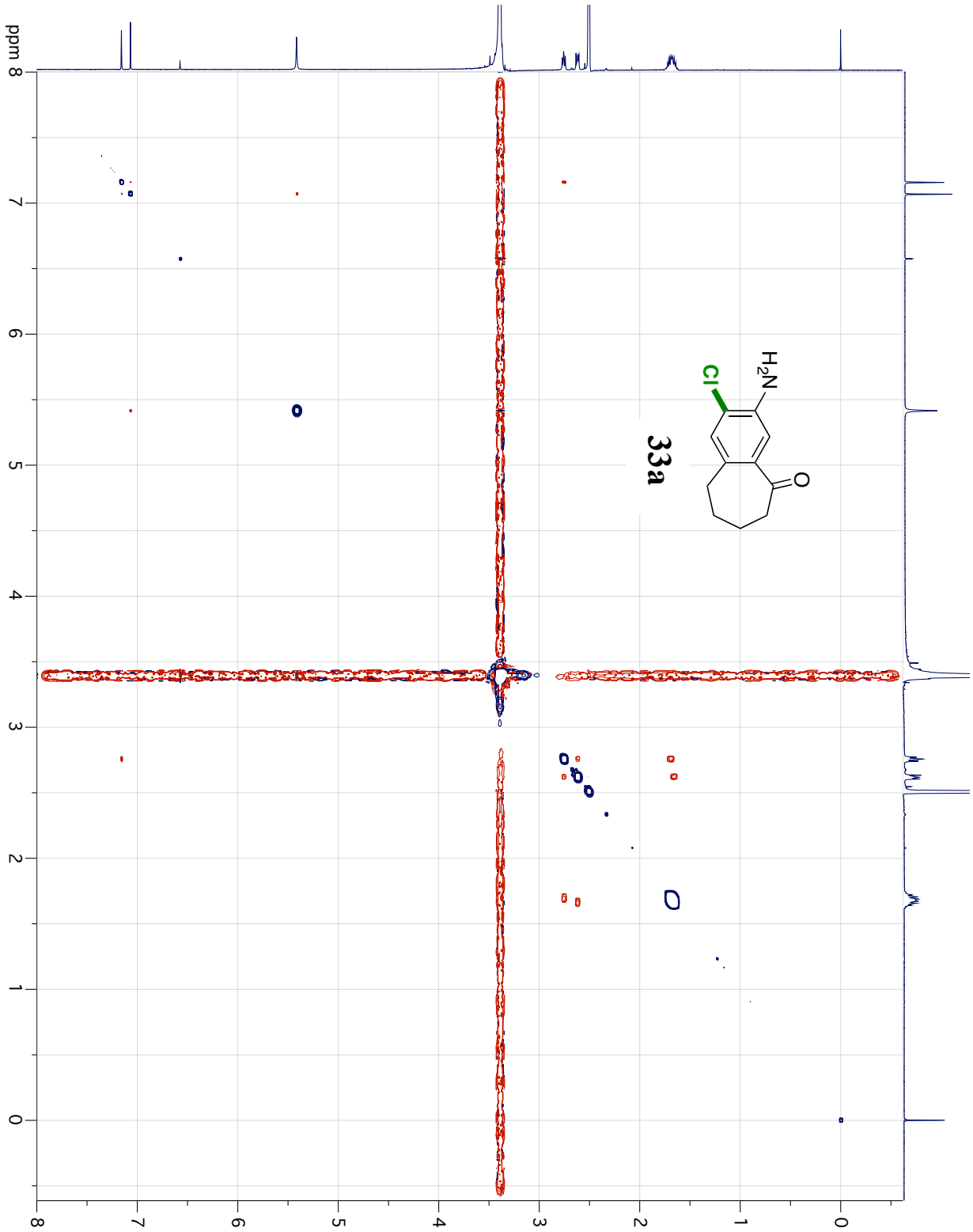


Figure AIV.15: [1]HNMR spectrum of **33b**.

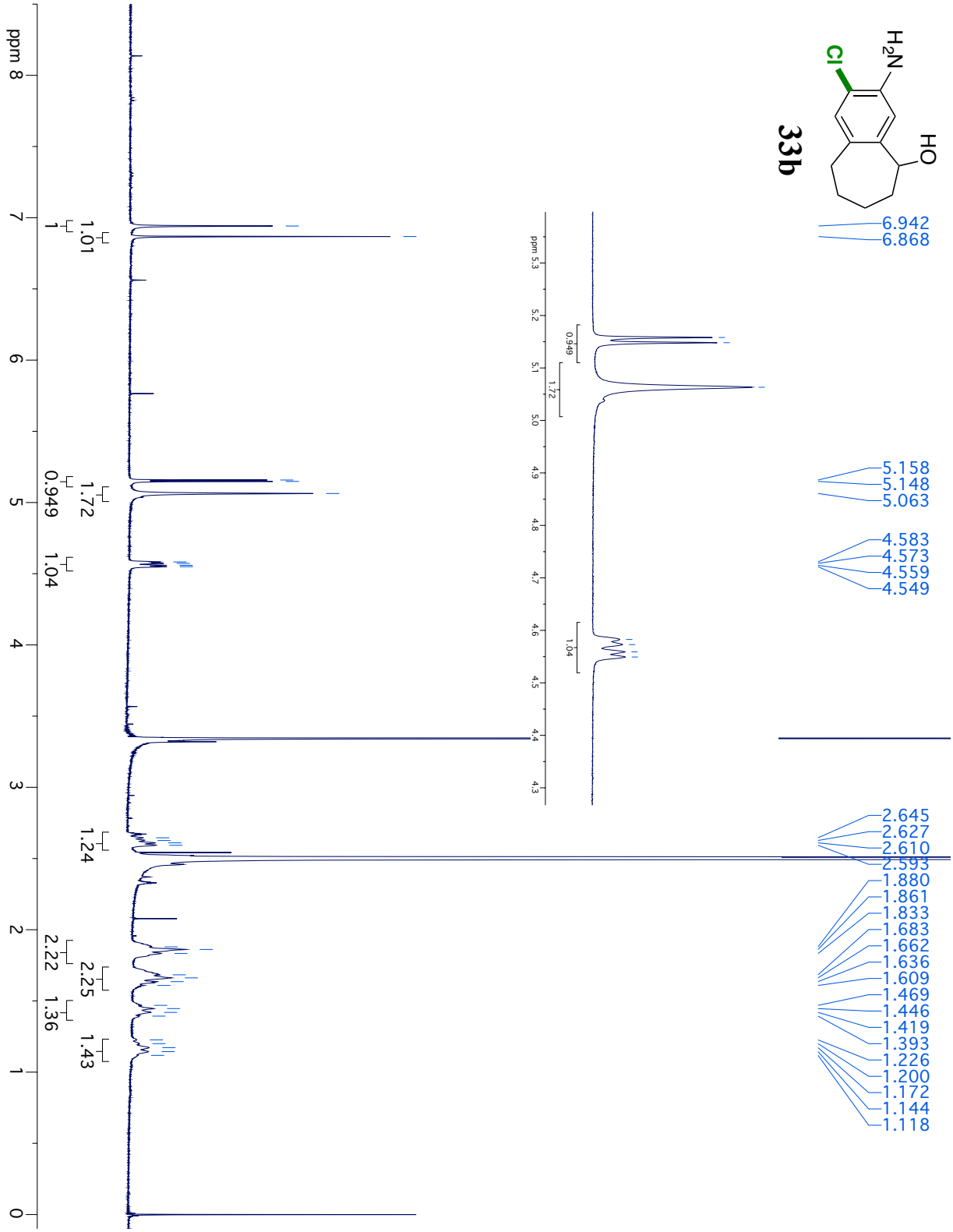


Figure AIV.16: [1]HNMR spectrum of **34**.

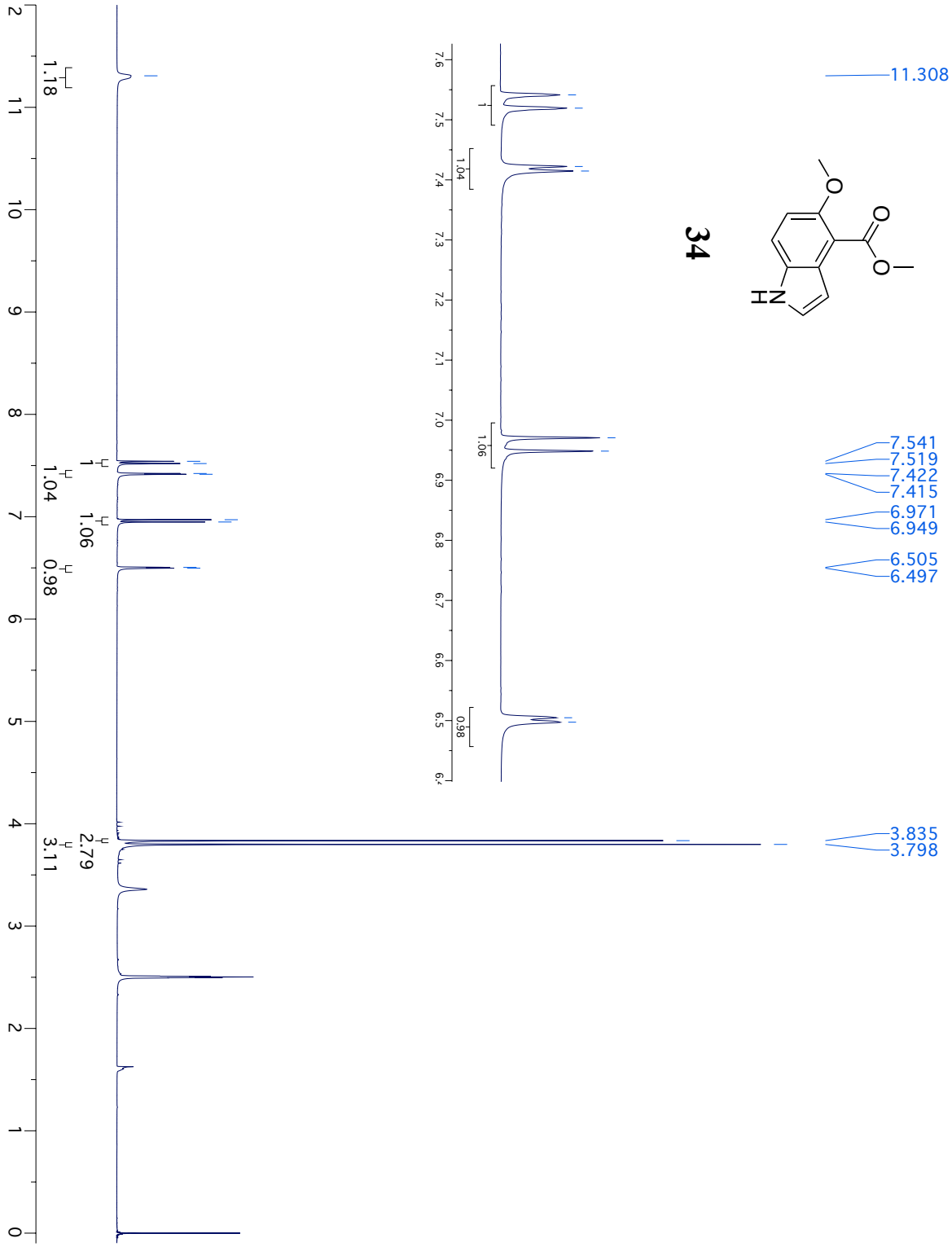


Figure AIV.17: ^{1}HNMR spectrum of **34a**.

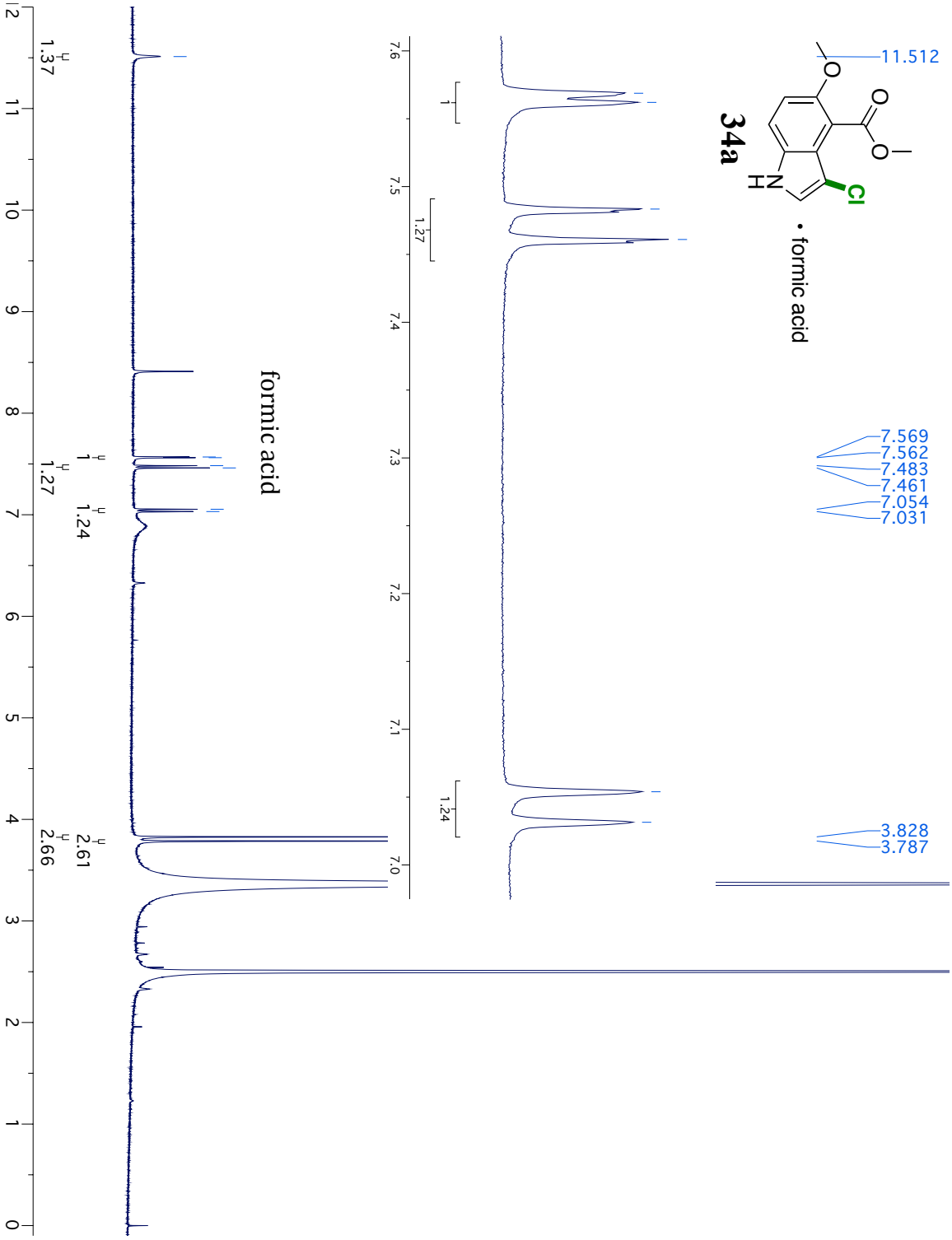


Figure AIV.18: NOSEY spectrum of **34a**.

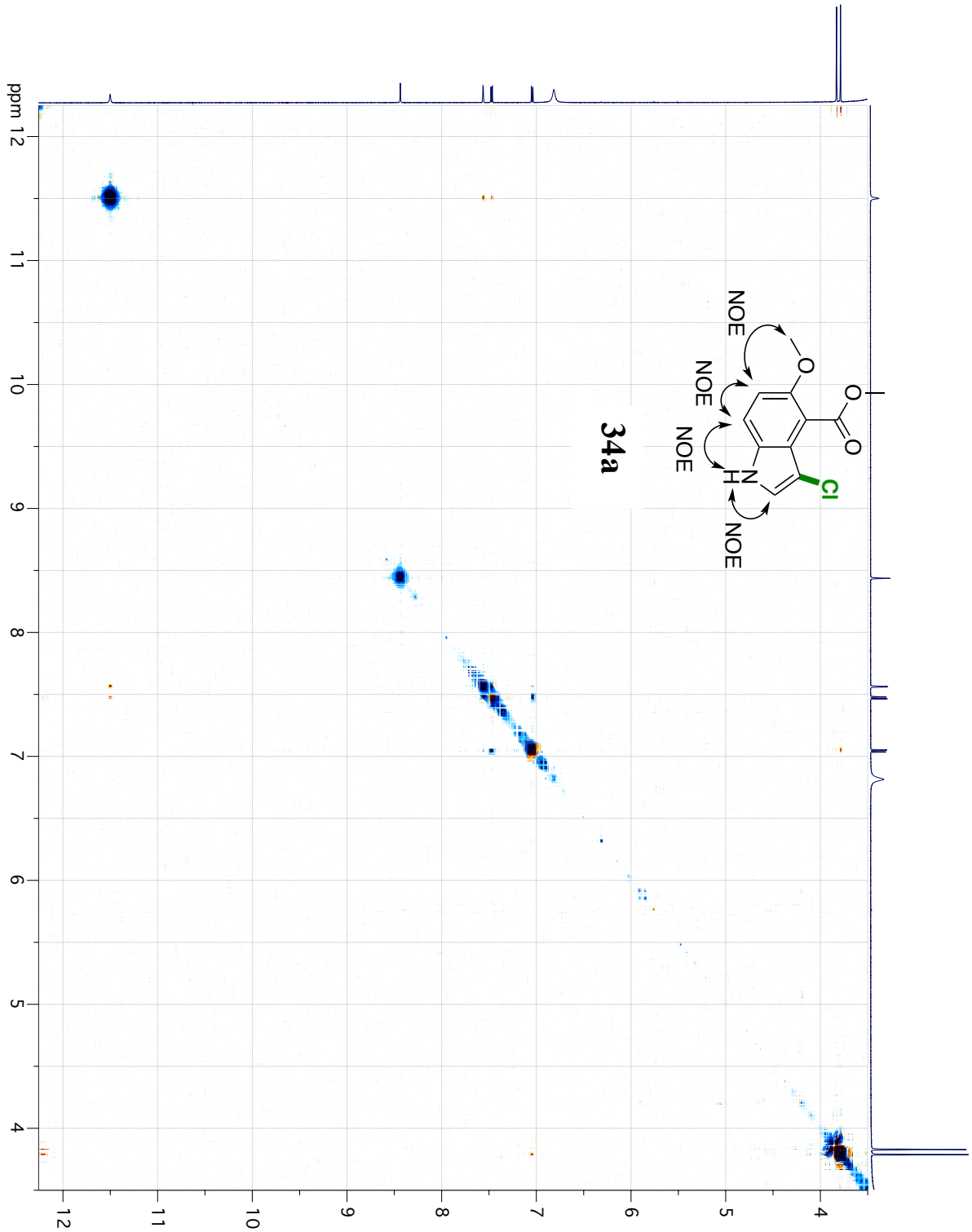


Figure AIV.19: [1]HNMR spectrum of **35**.

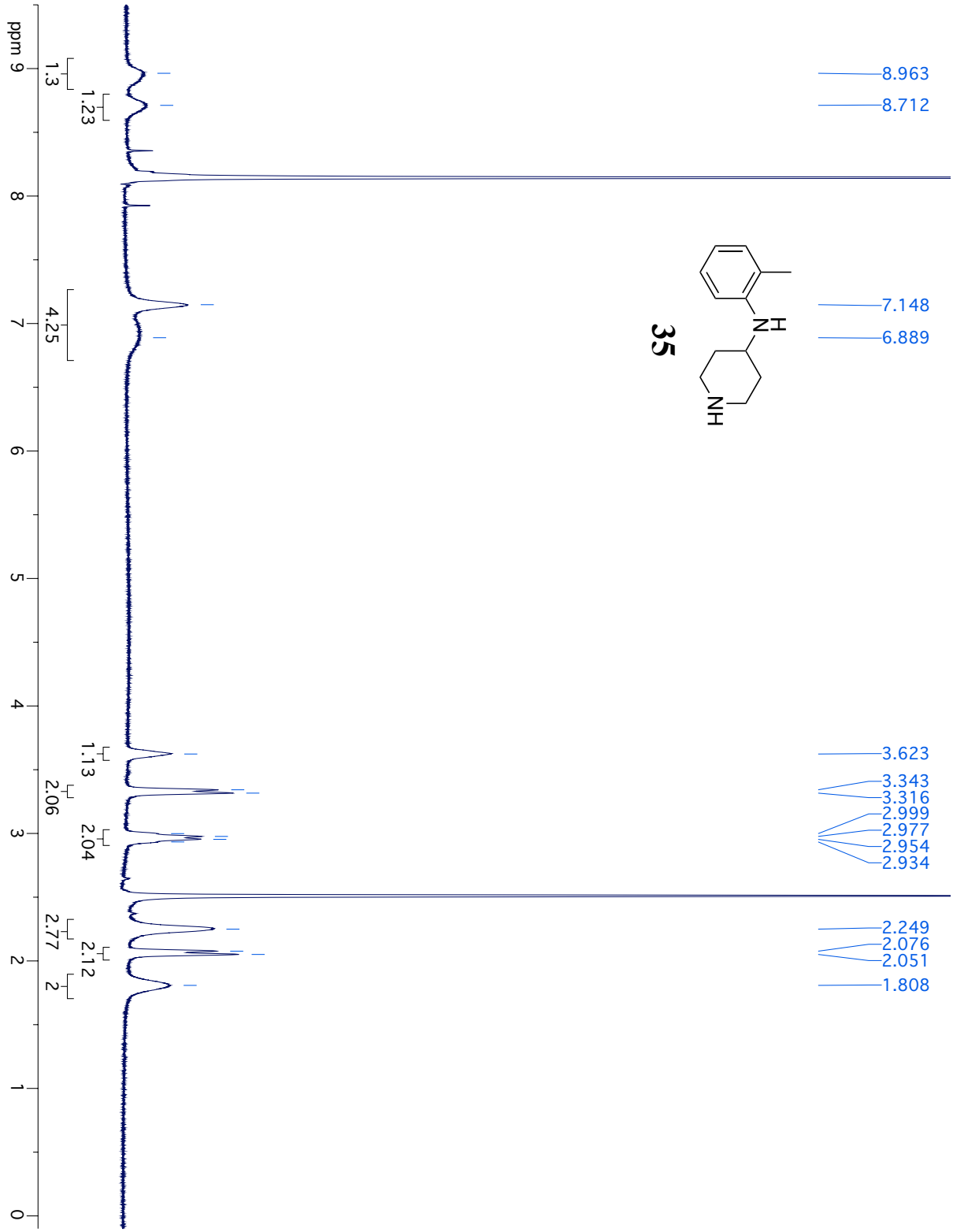


Figure AIV.20: [1]HNMR spectrum of **35a**.

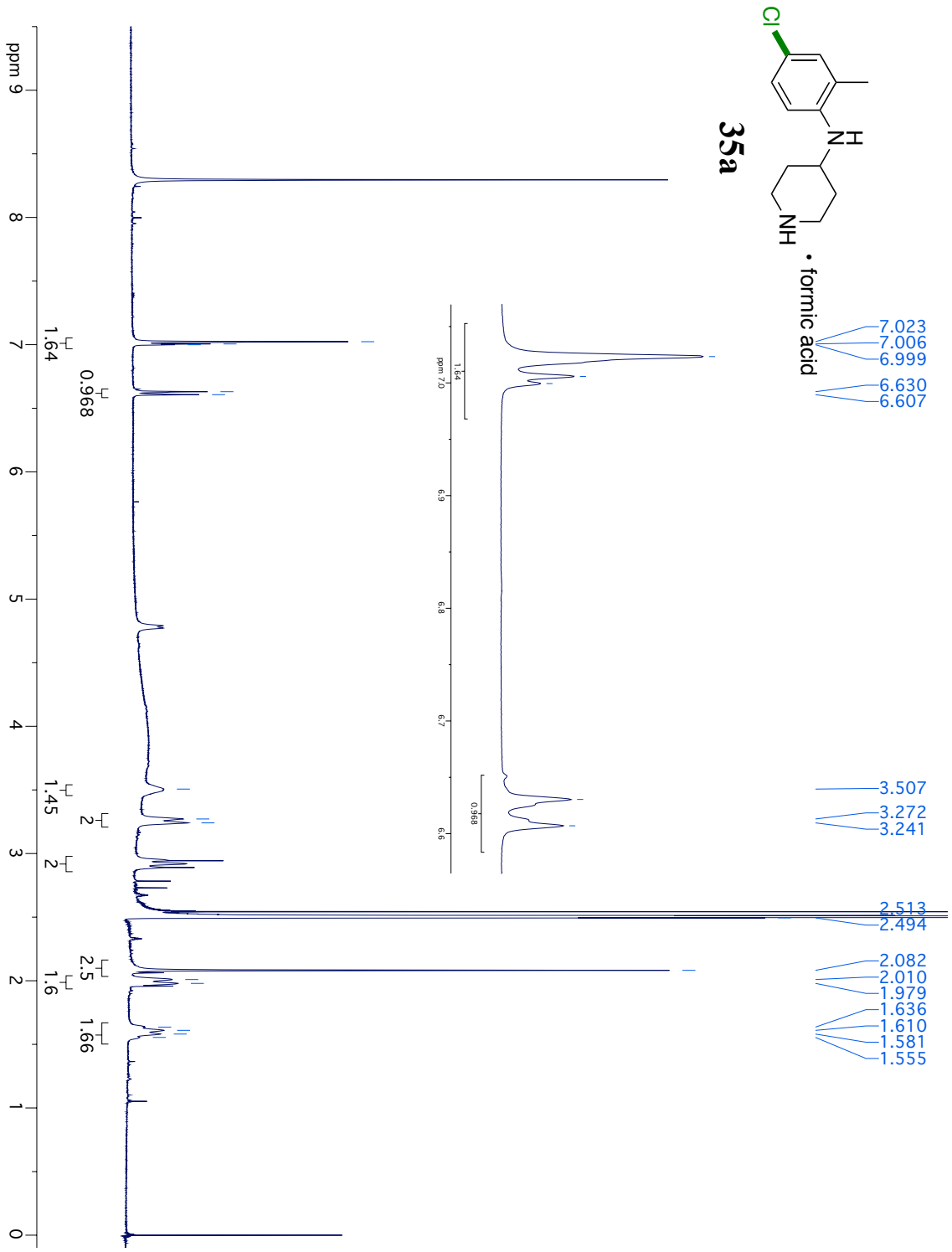


Figure AIV.21: NOSEY spectrum of **35a**.

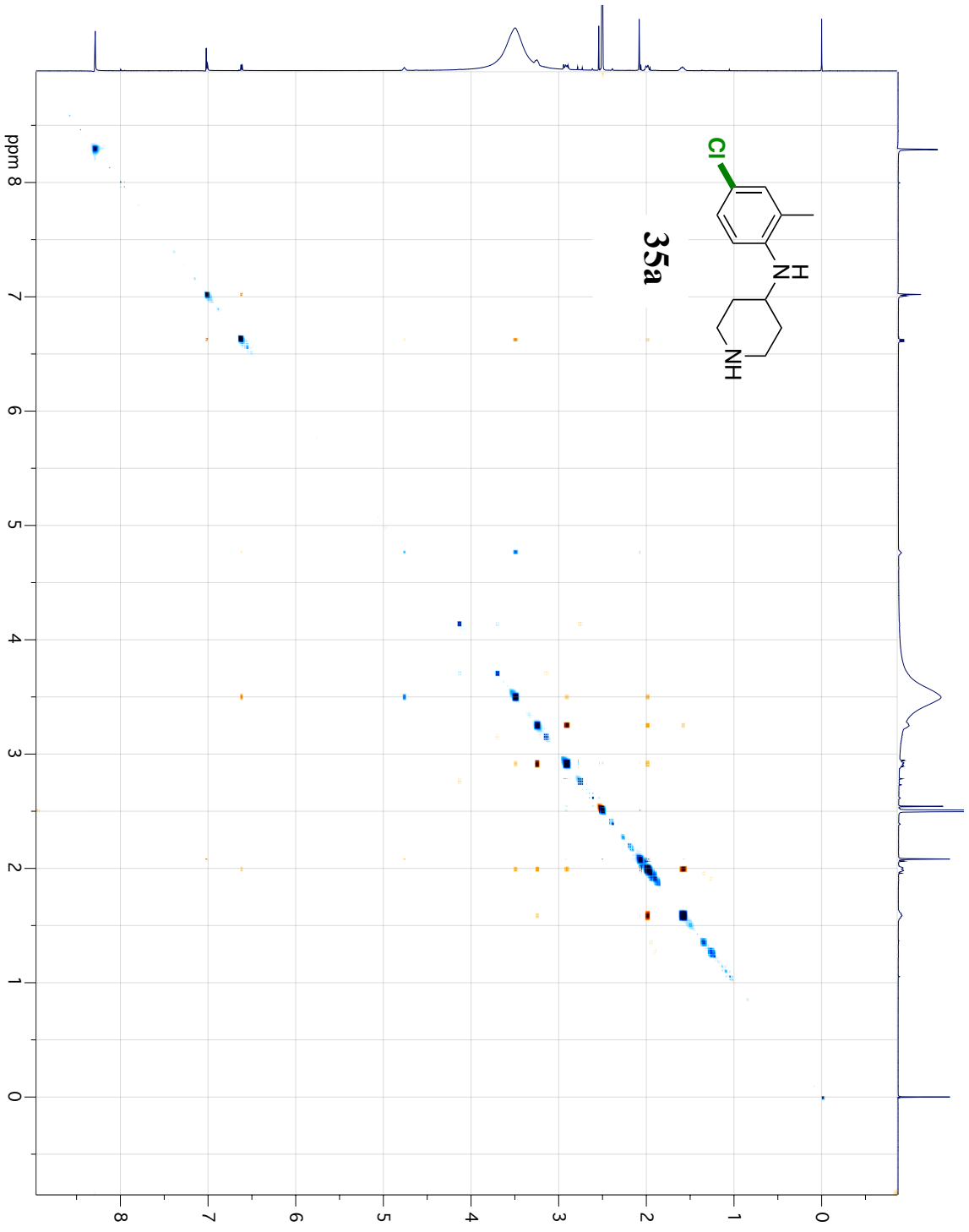


Figure AIV.22: Zoomed NOSEY spectrum of **35a**.

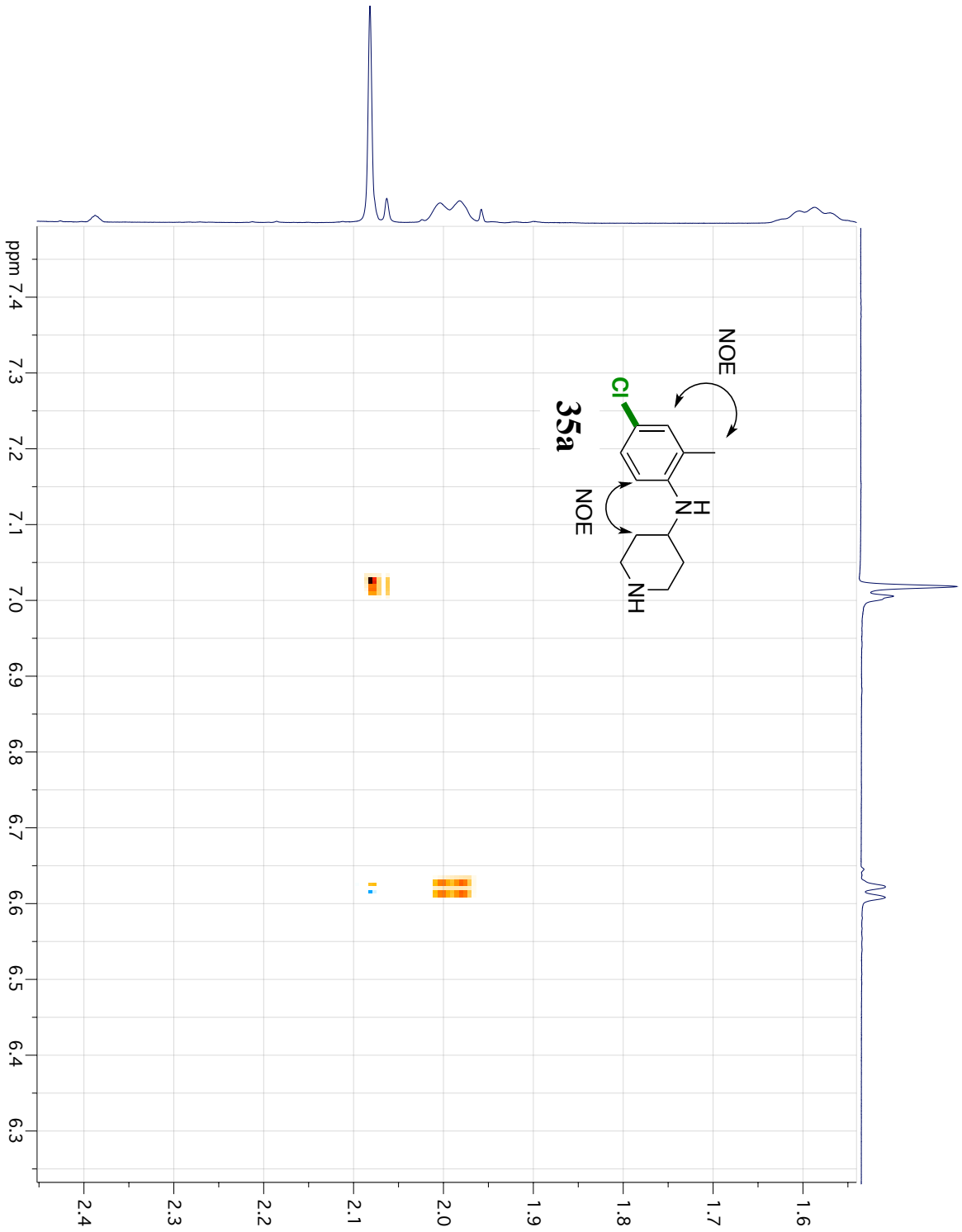


Figure AIV.23: [1]HNMR spectrum of **36**.

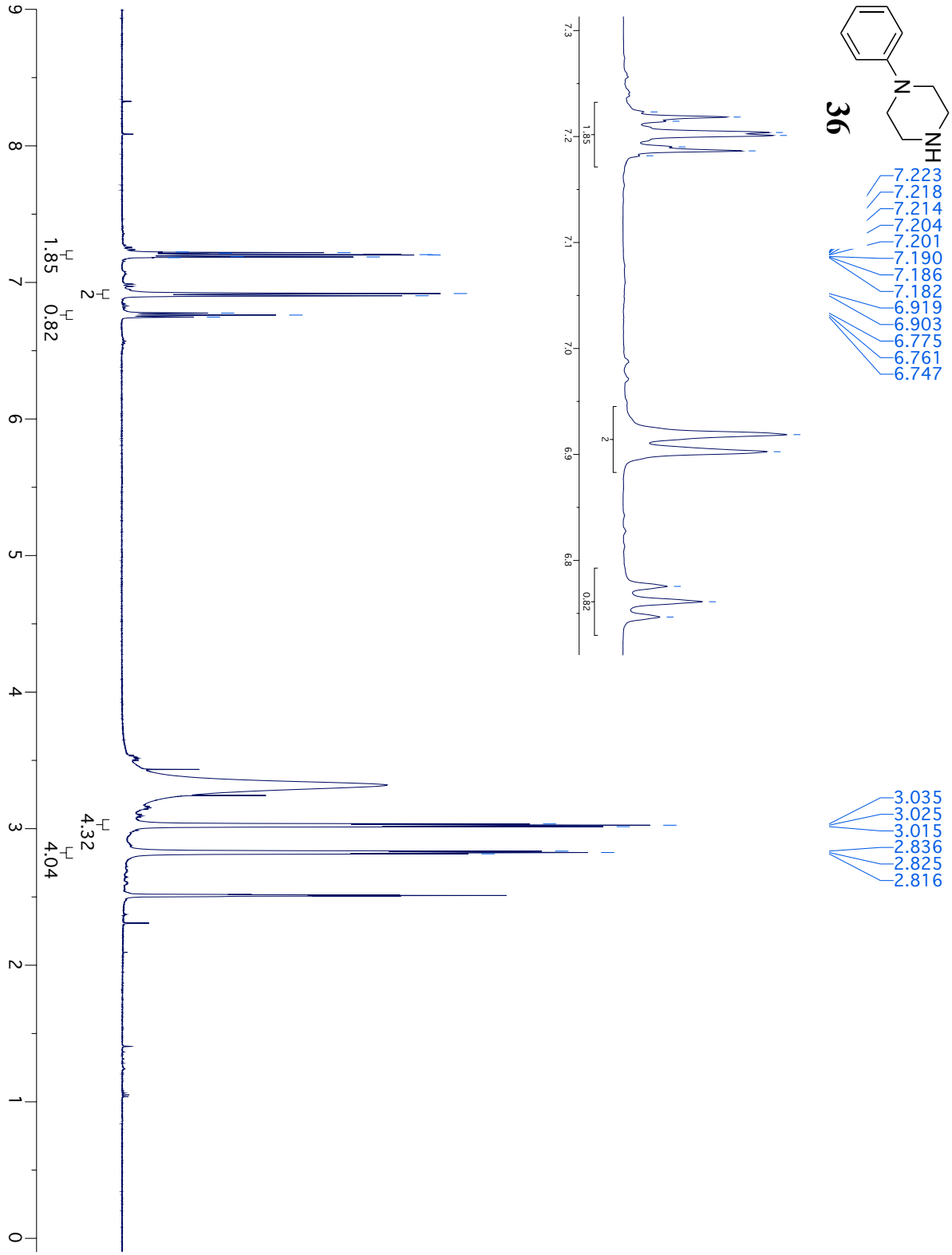


Figure AIV.24: [1]HNMR spectrum of **36a**.

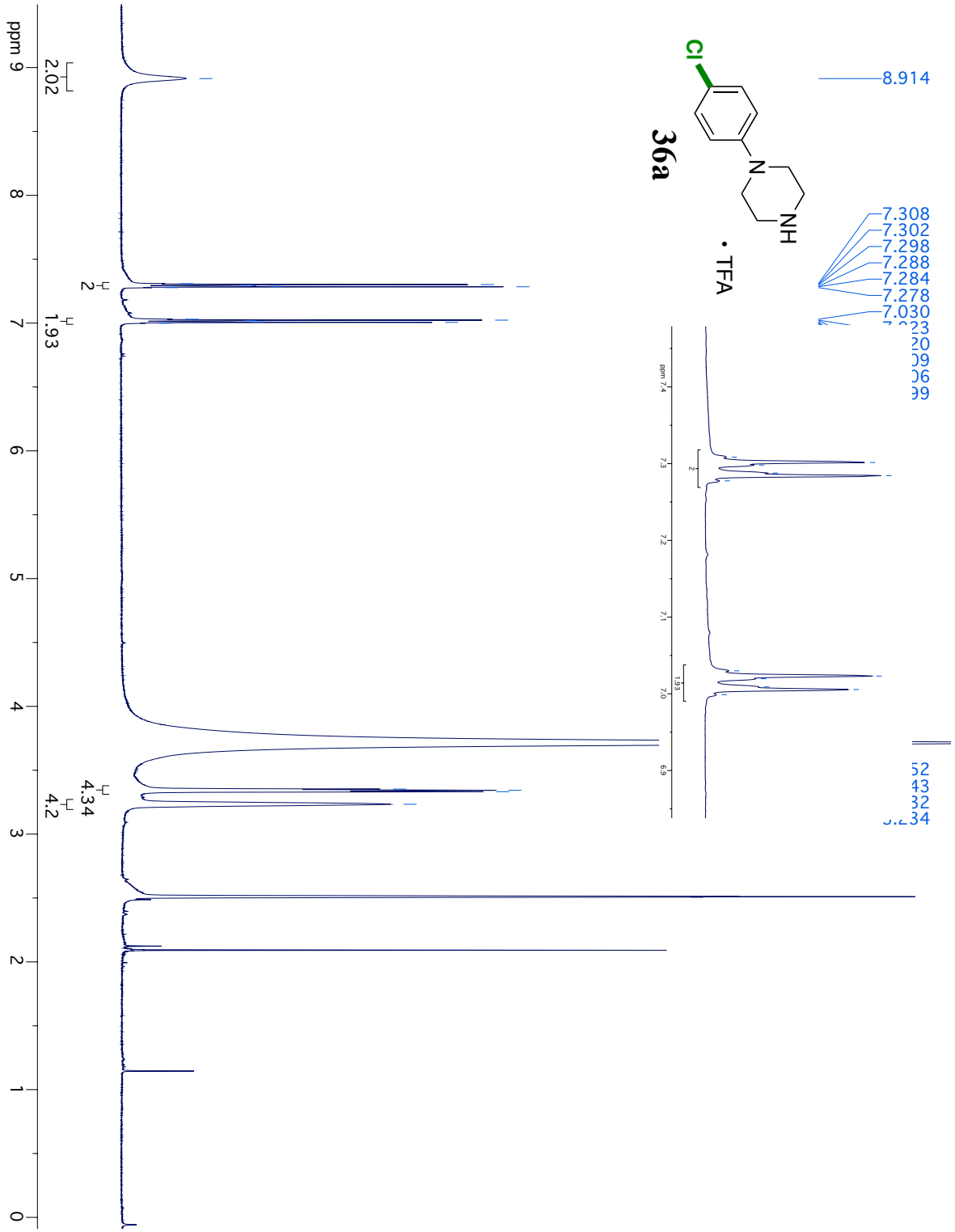


Figure AIV.25: [1]HNMR spectrum of **37**.

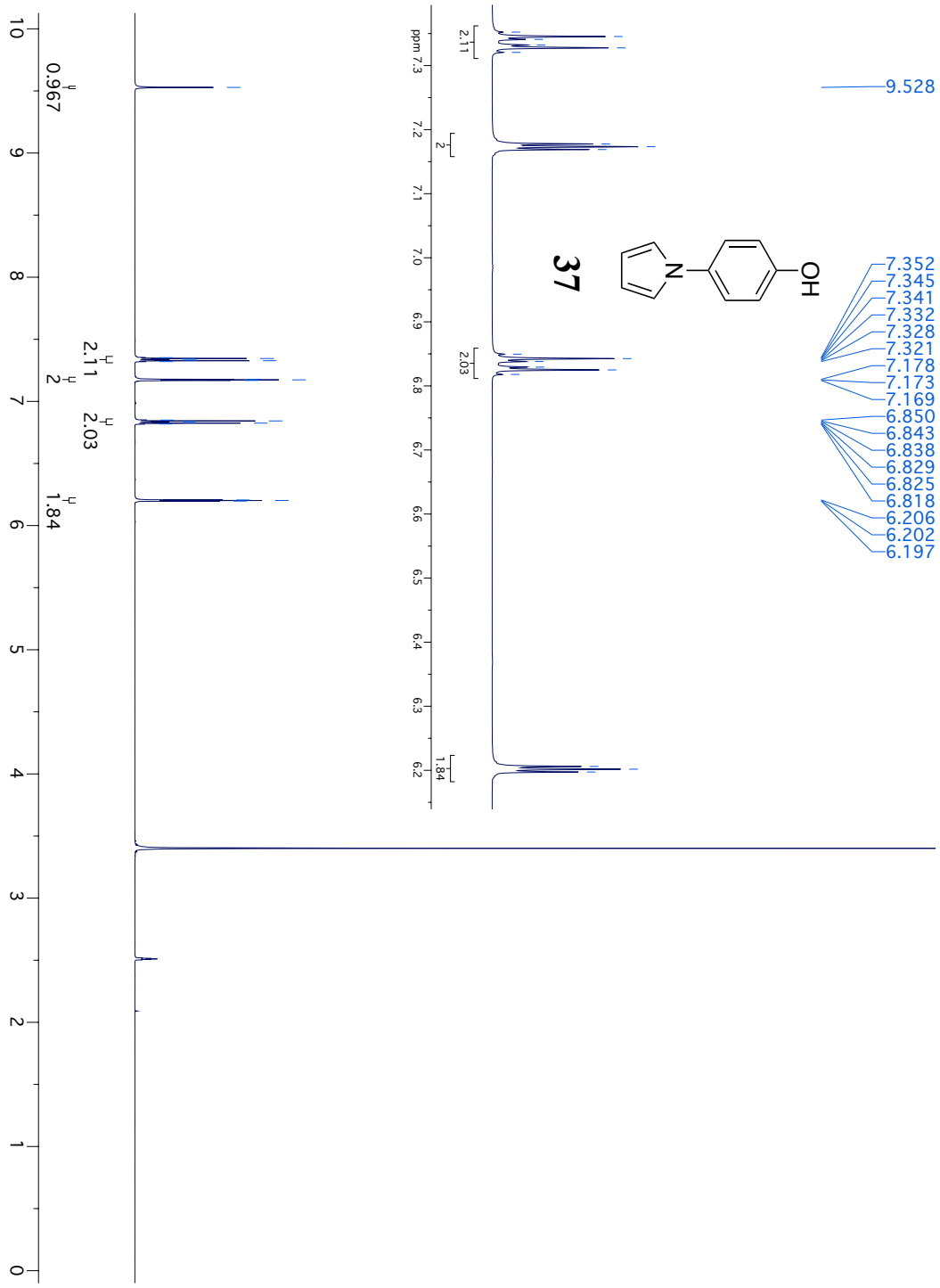


Figure AIV.26: NOSEY spectrum of **37**.

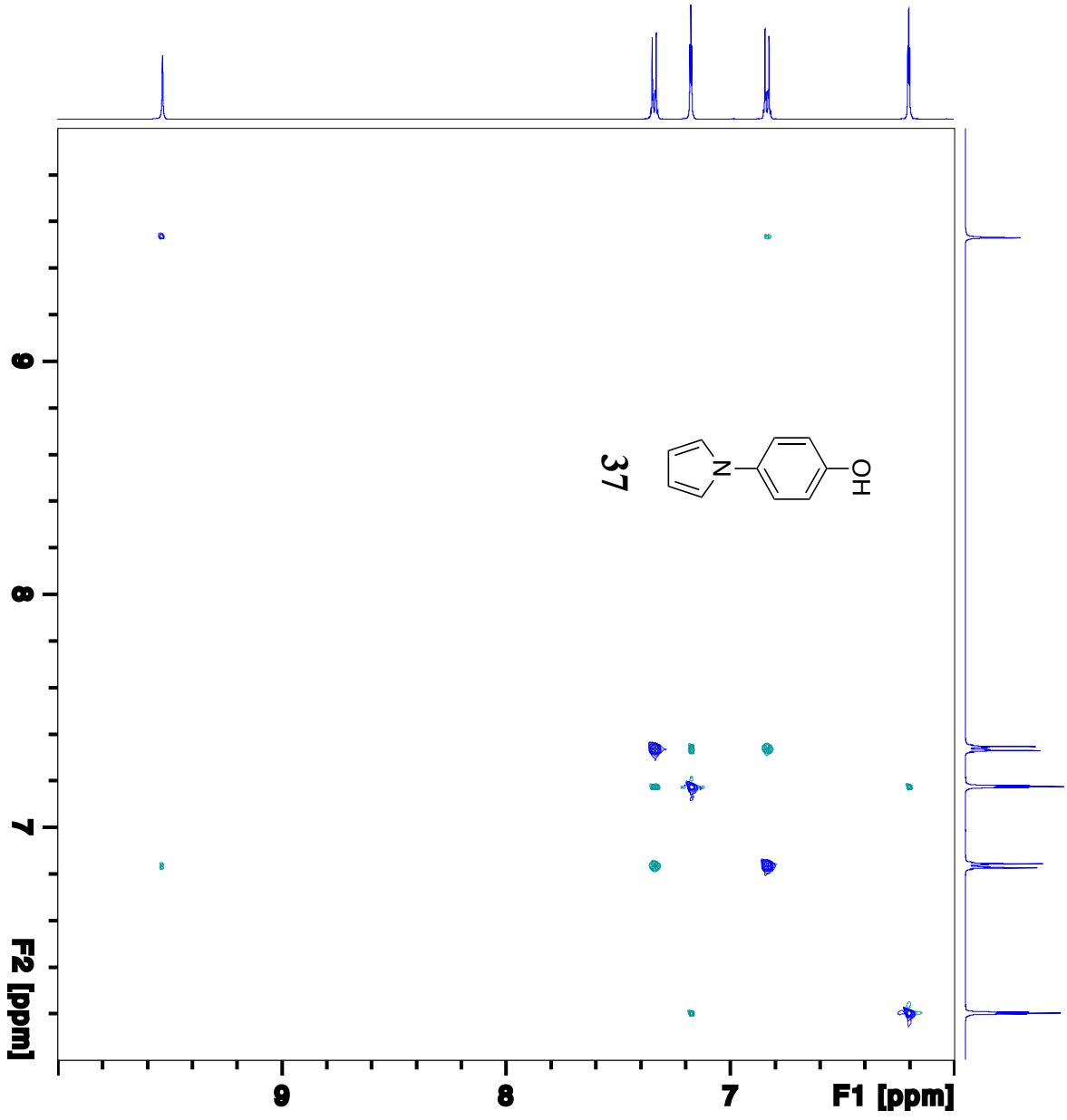


Figure AIV.27: [1]HNMR spectrum of **37a**.

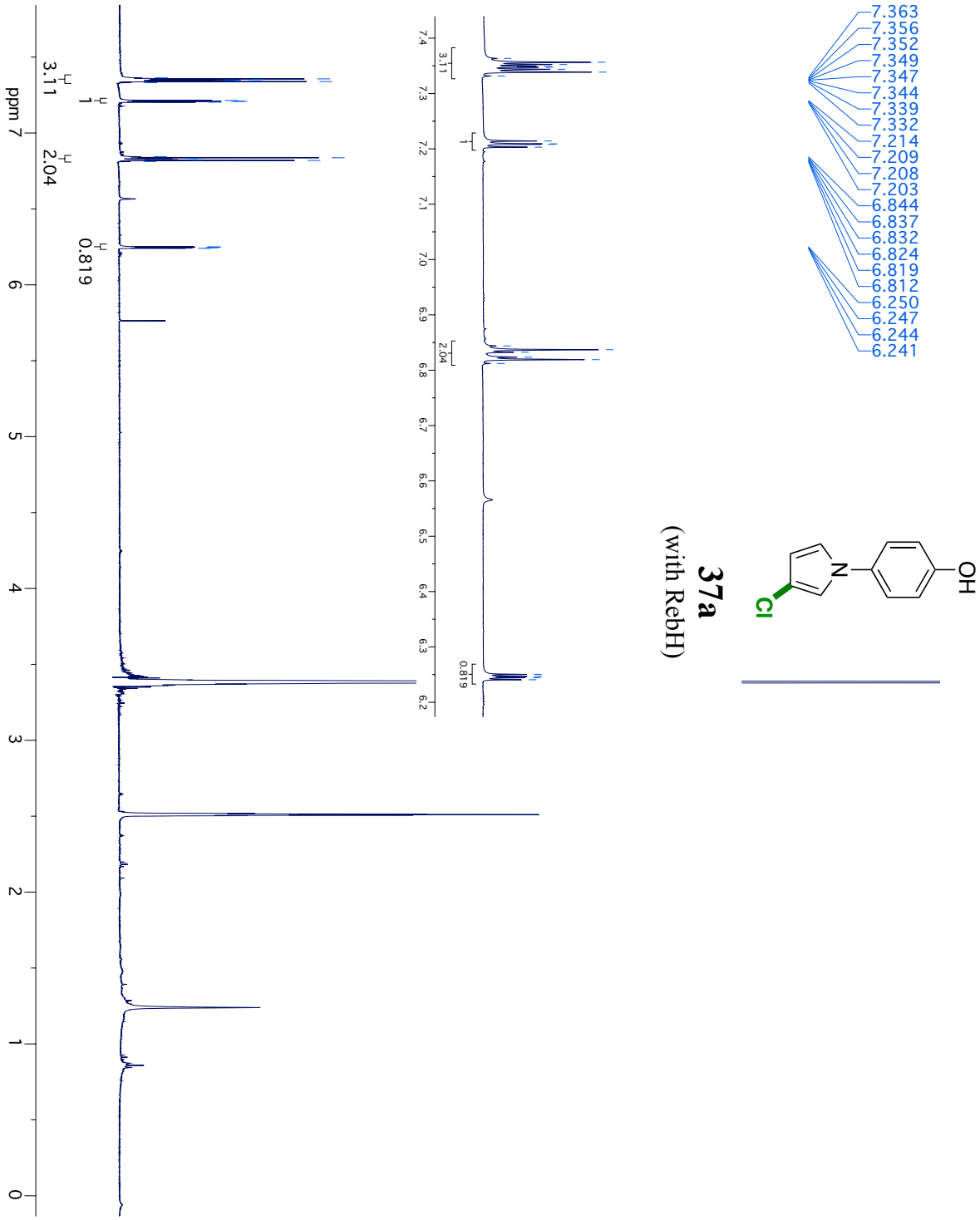


Figure AIV.28: [1]HNMR spectrum of **37b**.

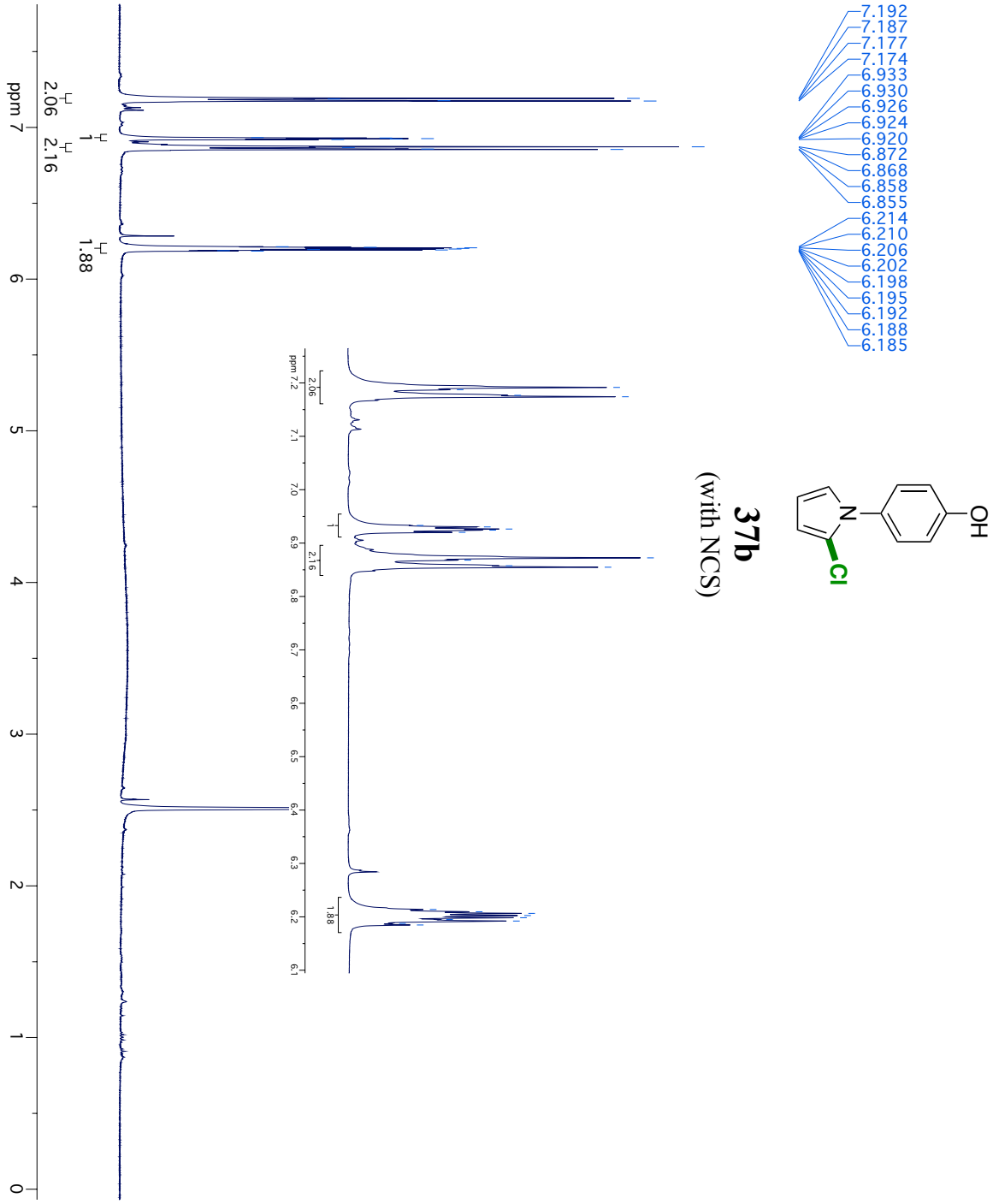


Figure AIV.29: [1]HNMR spectrum of **38**.

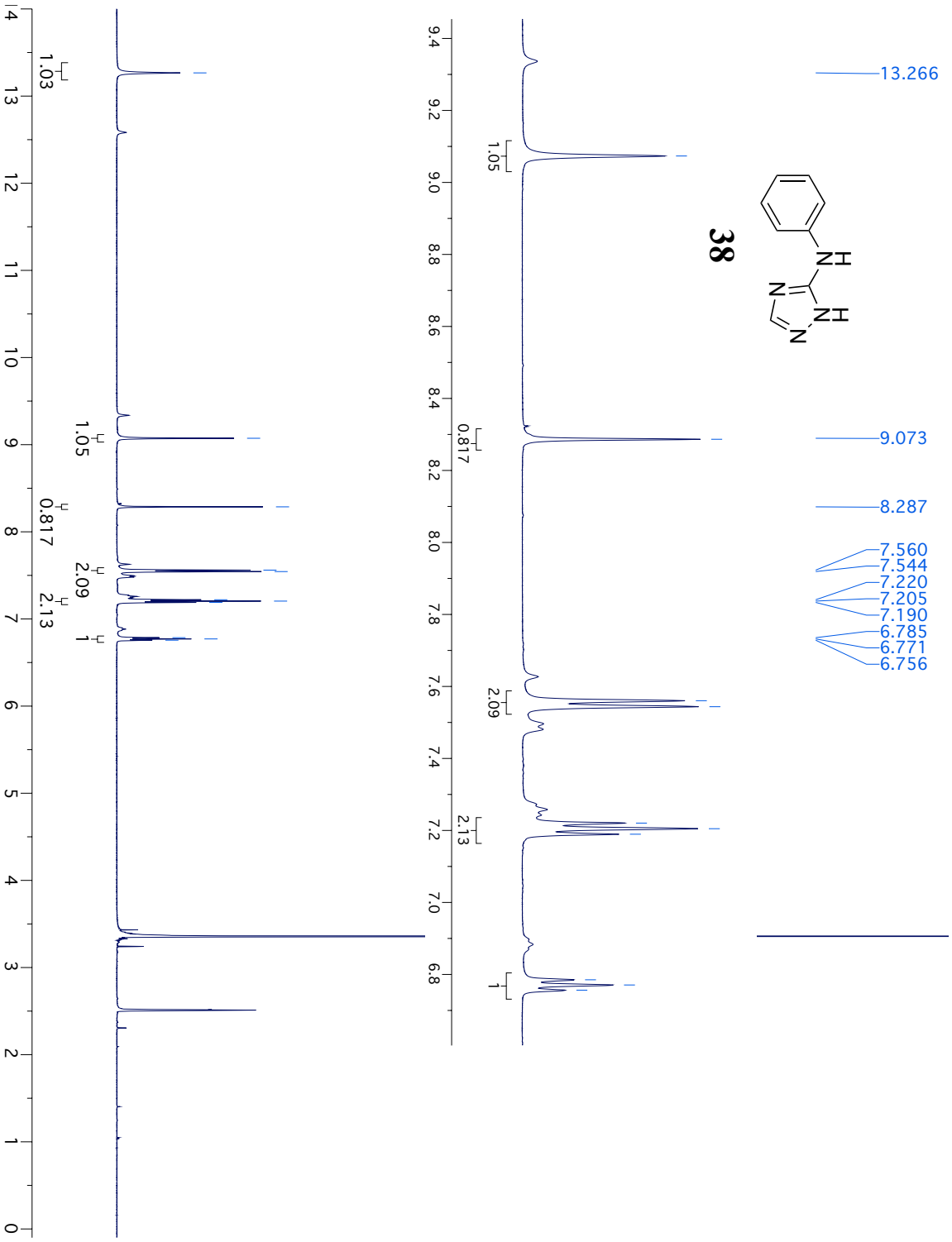


Figure AIV.30: [1]HNMR spectrum of **38a**.

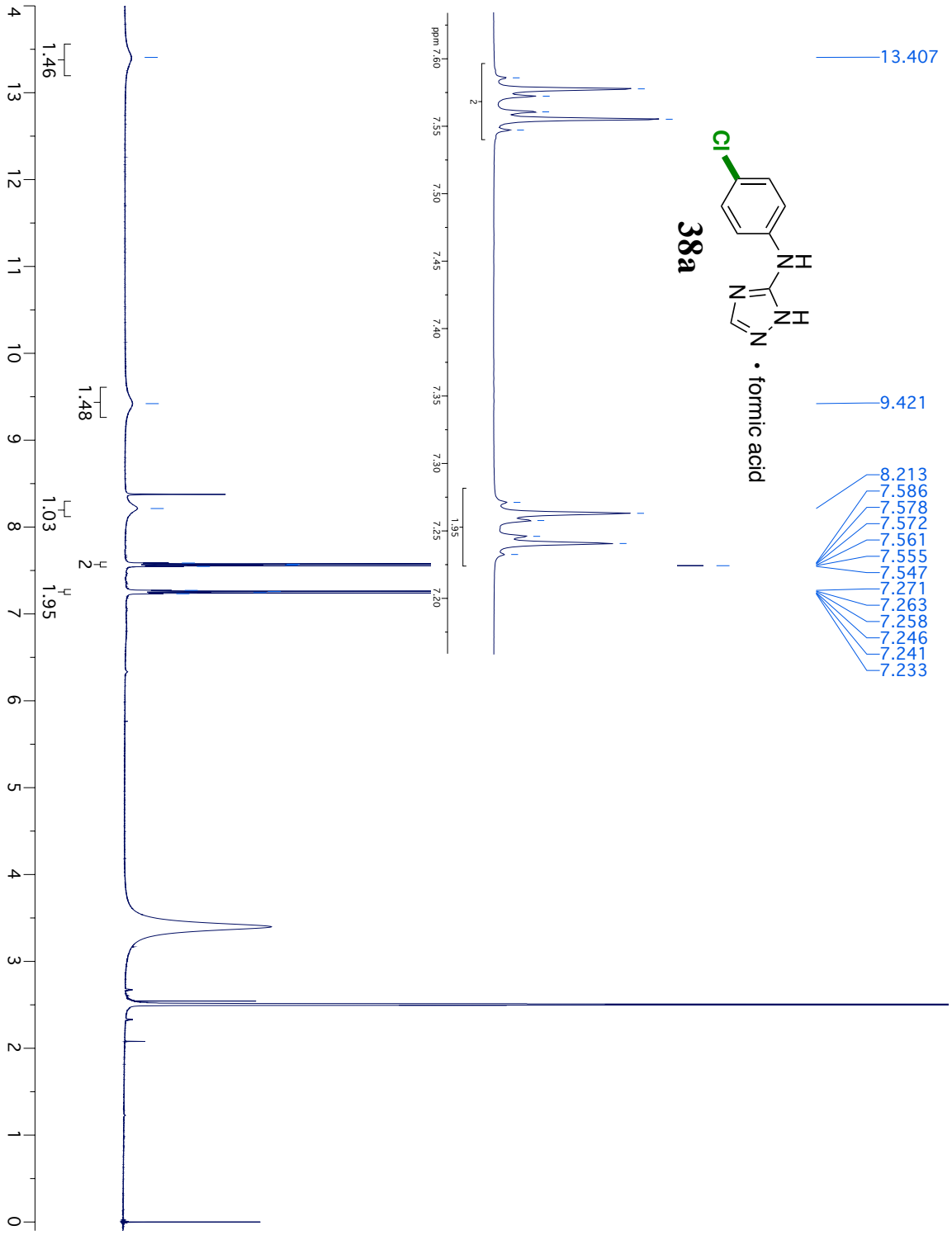


Figure AIV.31: [1]HNMR spectrum of **39**.

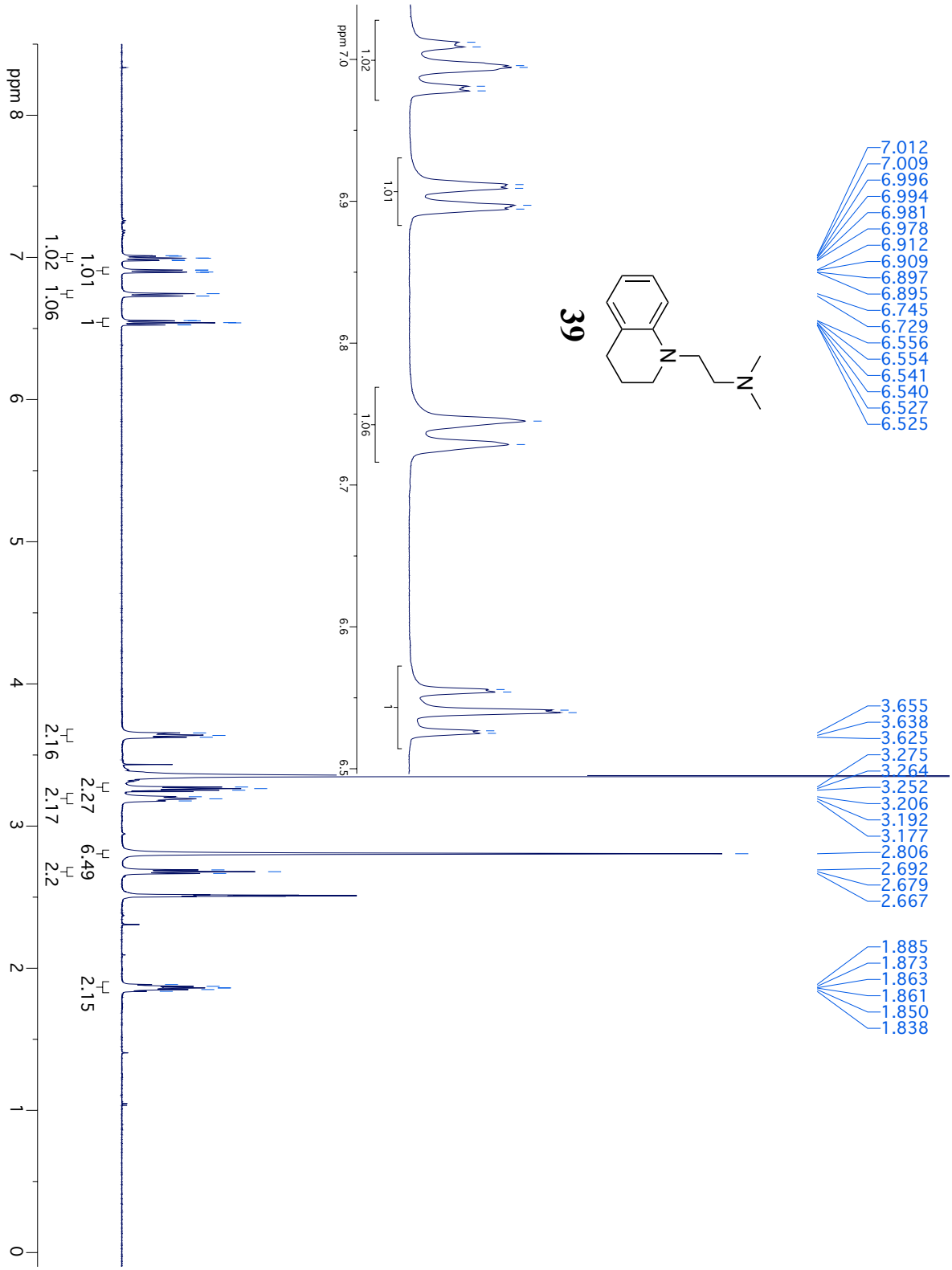


Figure AIV.32: [1]HNMR spectrum of **39a**.

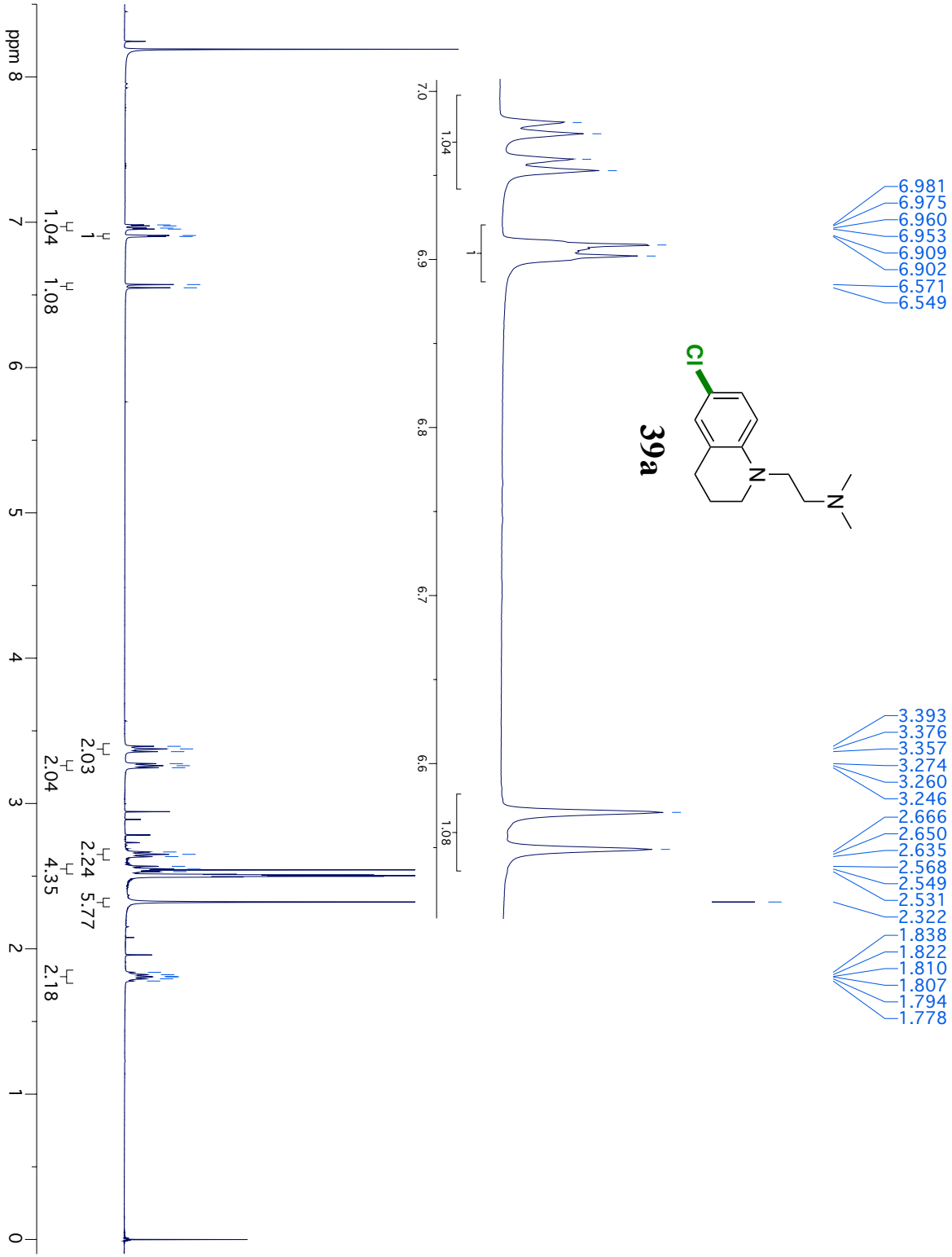


Figure AIV.33: [1]HNMR spectrum of **40**.

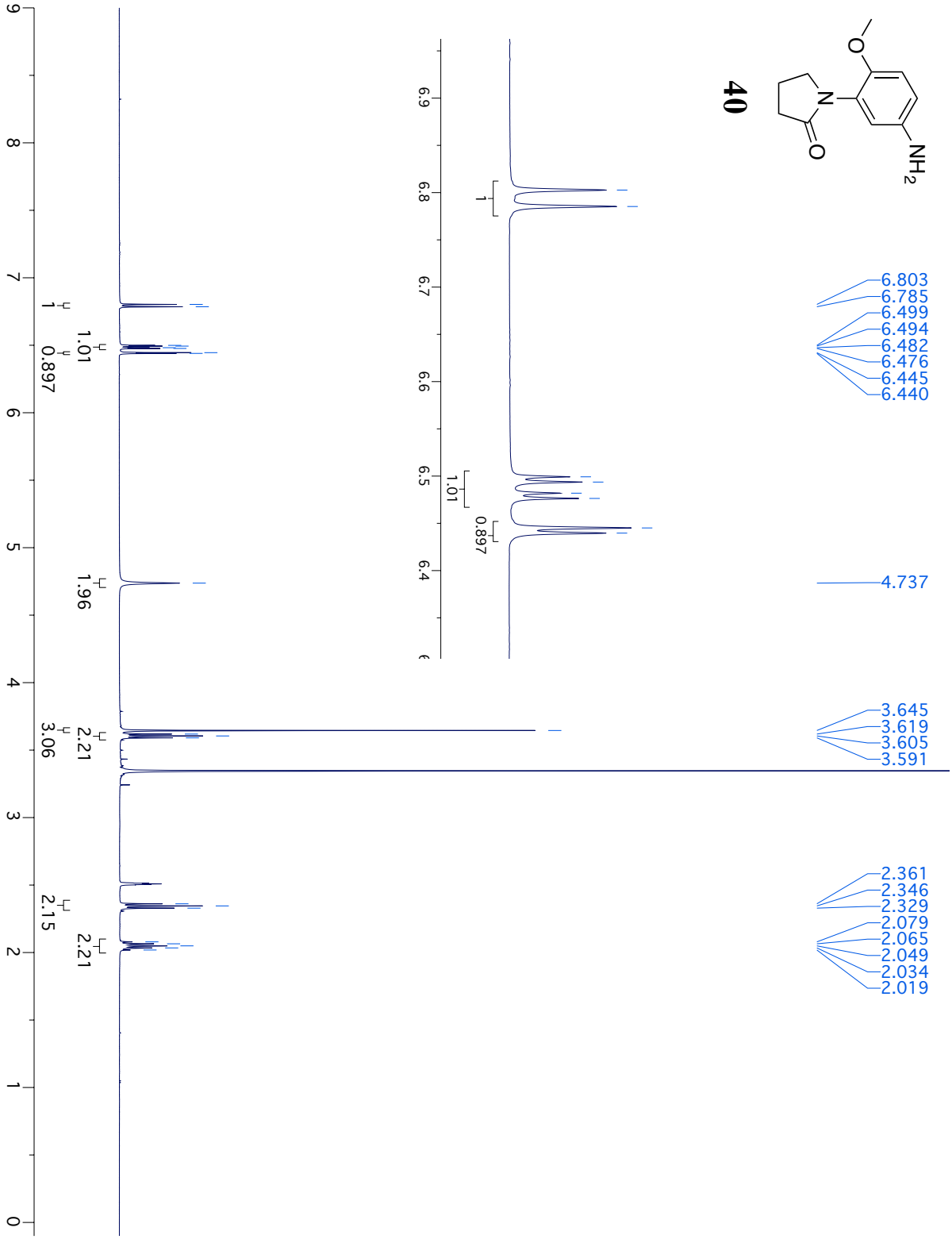


Figure AIV.34: [1]HNMR spectrum of **40a**.

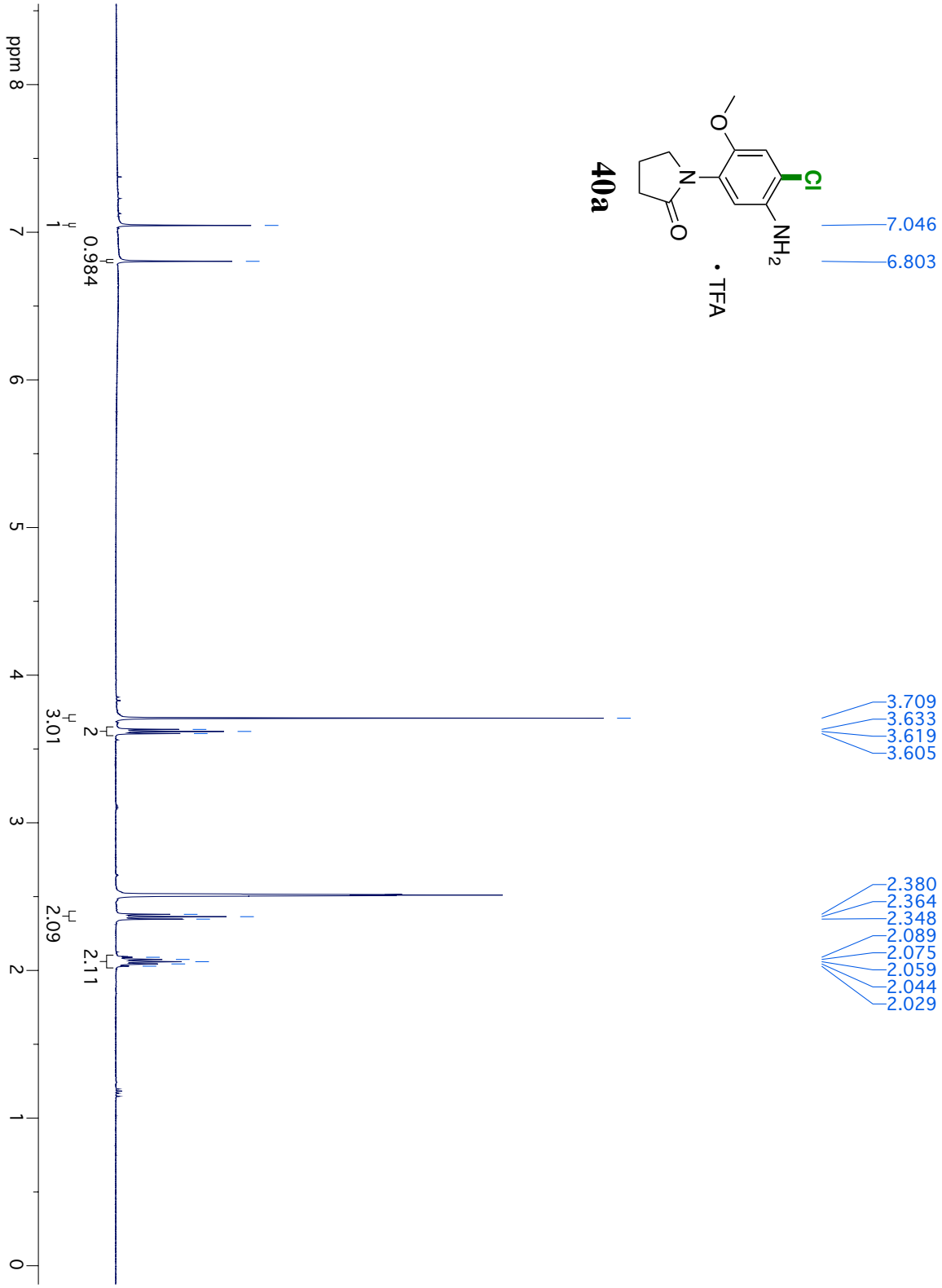


Figure AIV.35: NOSEY spectrum of **40a**.

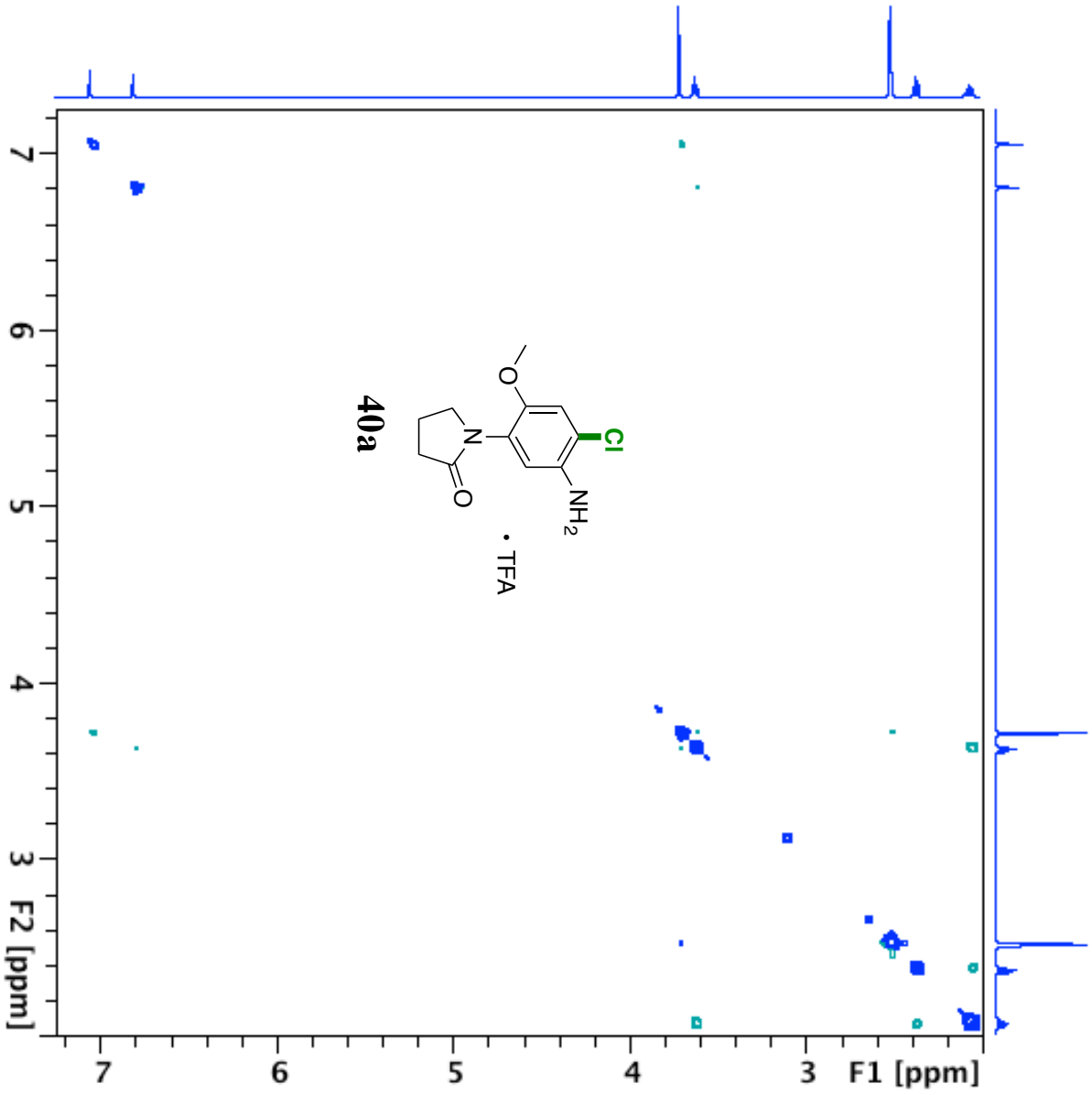


Figure AIV.36: [1]HNMR spectrum of **41**.

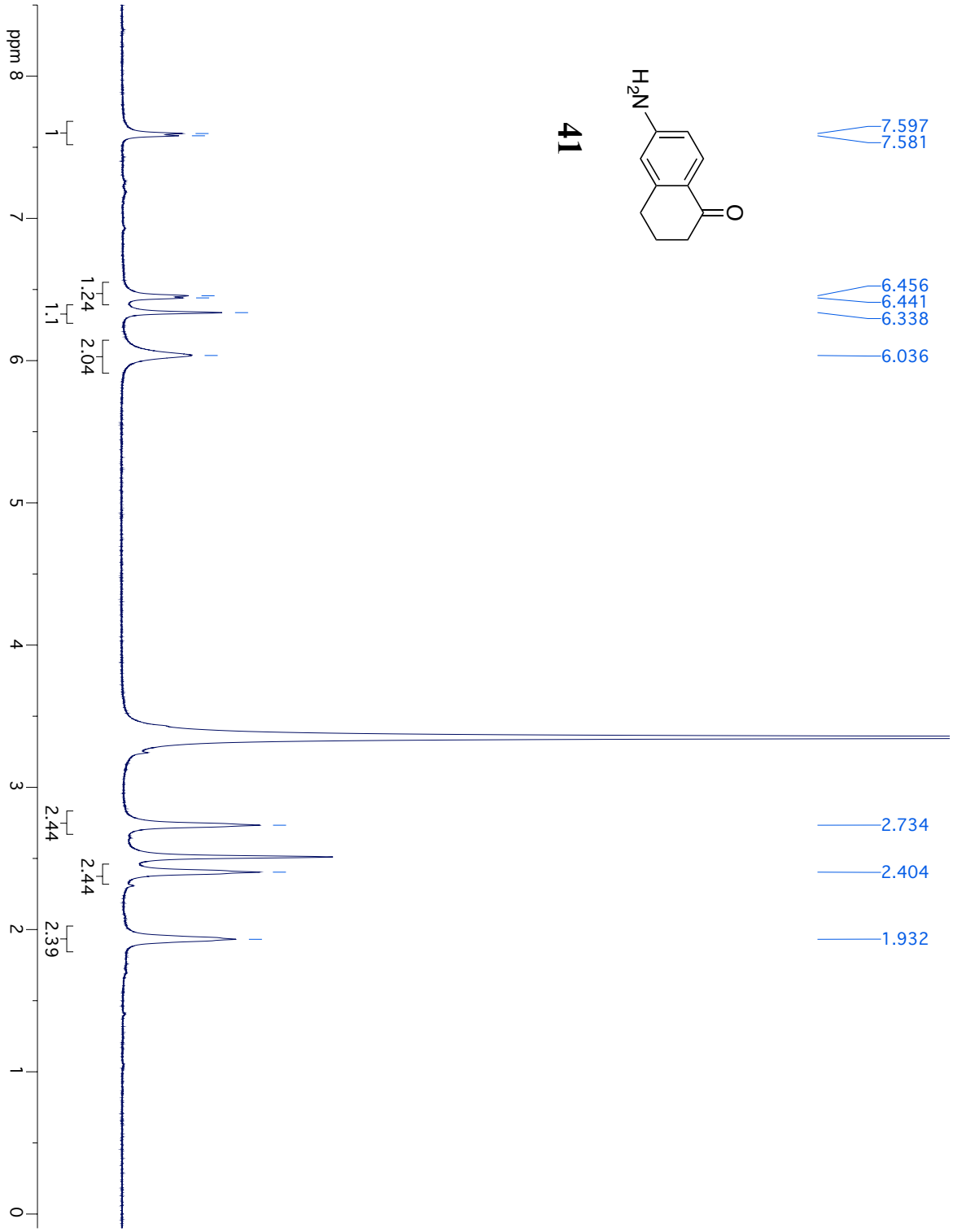


Figure AIV.37: [1]HNMR spectrum of **41a**.

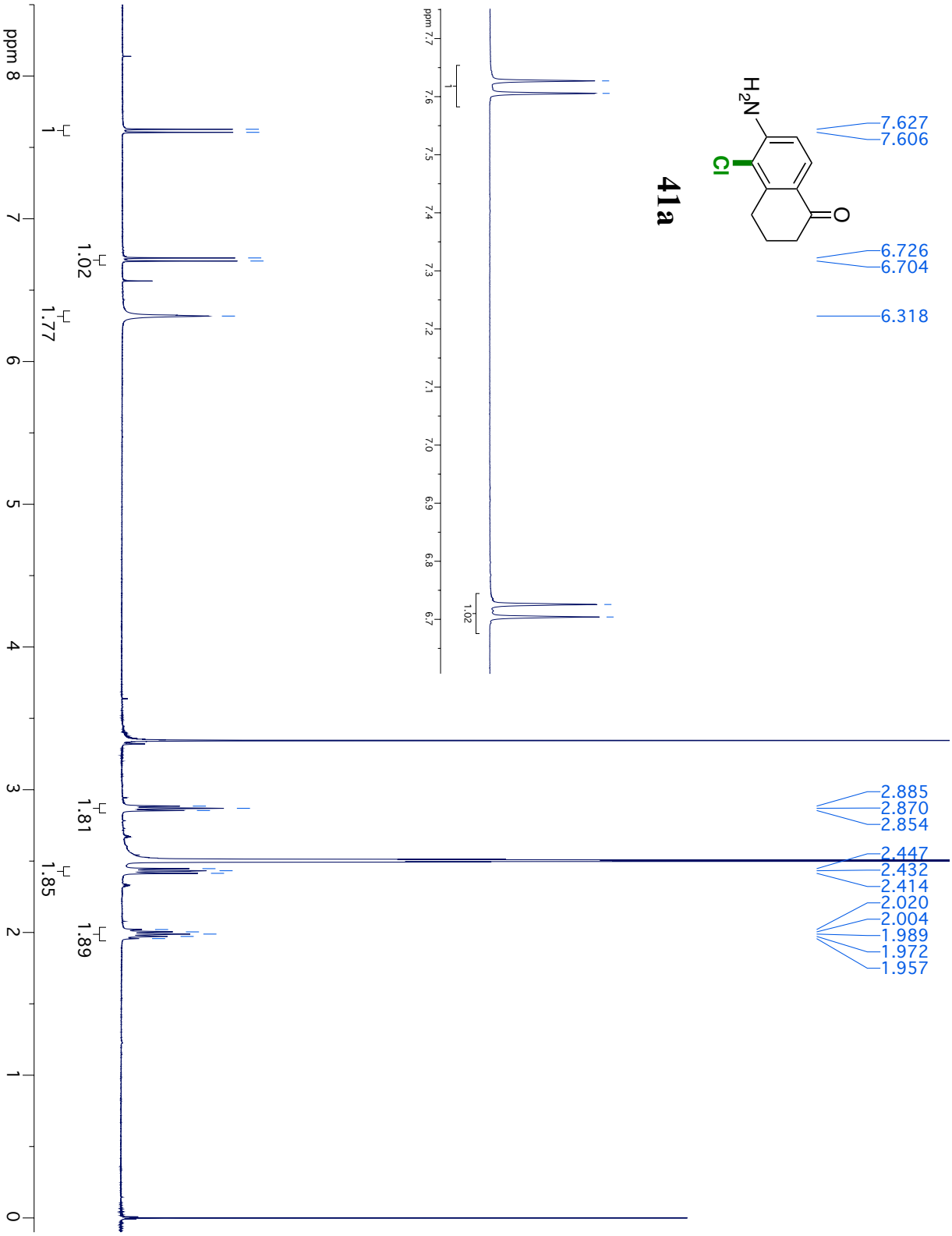


Figure AIV.38: [1]HNMR spectrum of **42**.

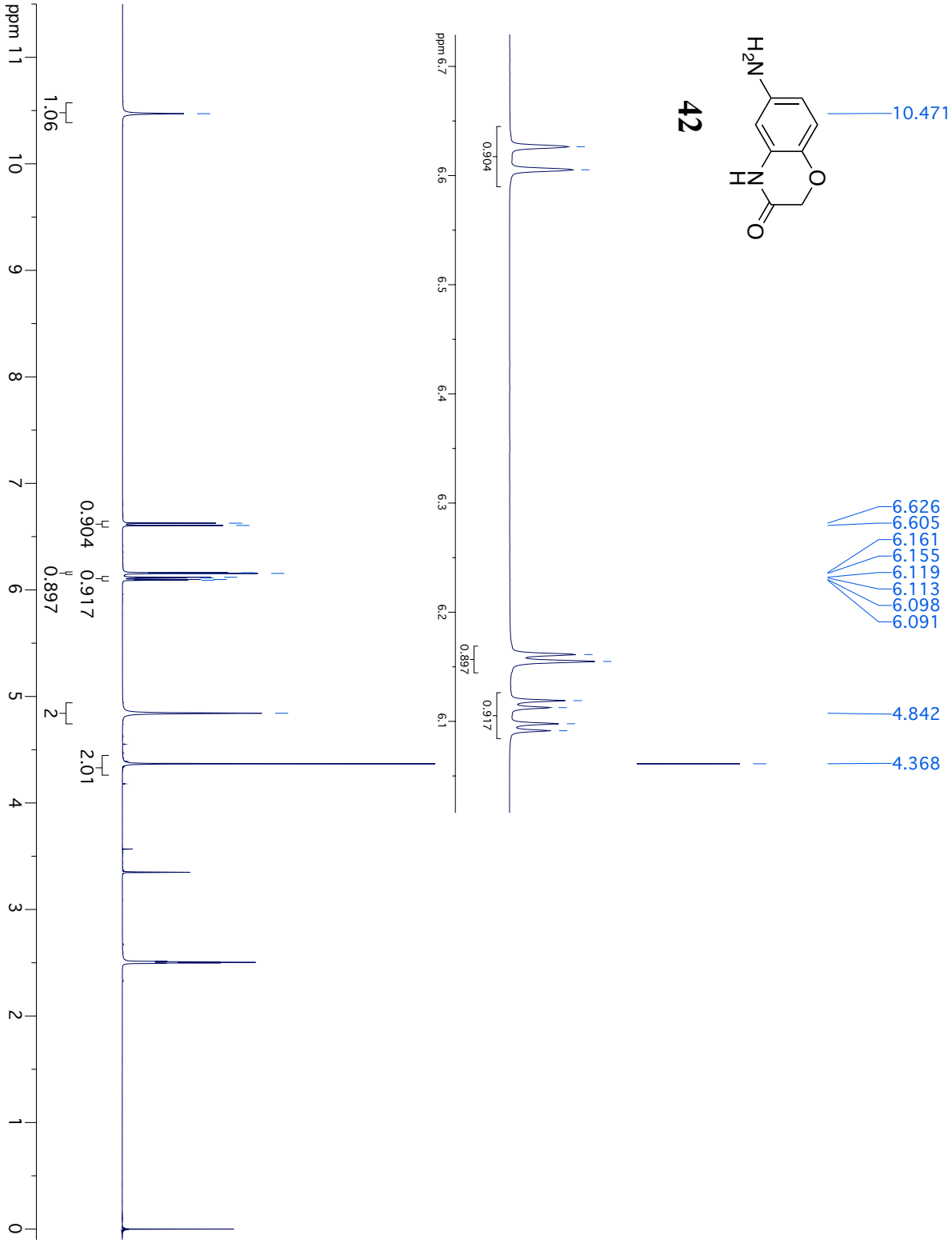


Figure AIV.39: [1]HNMR spectrum of **42a**.

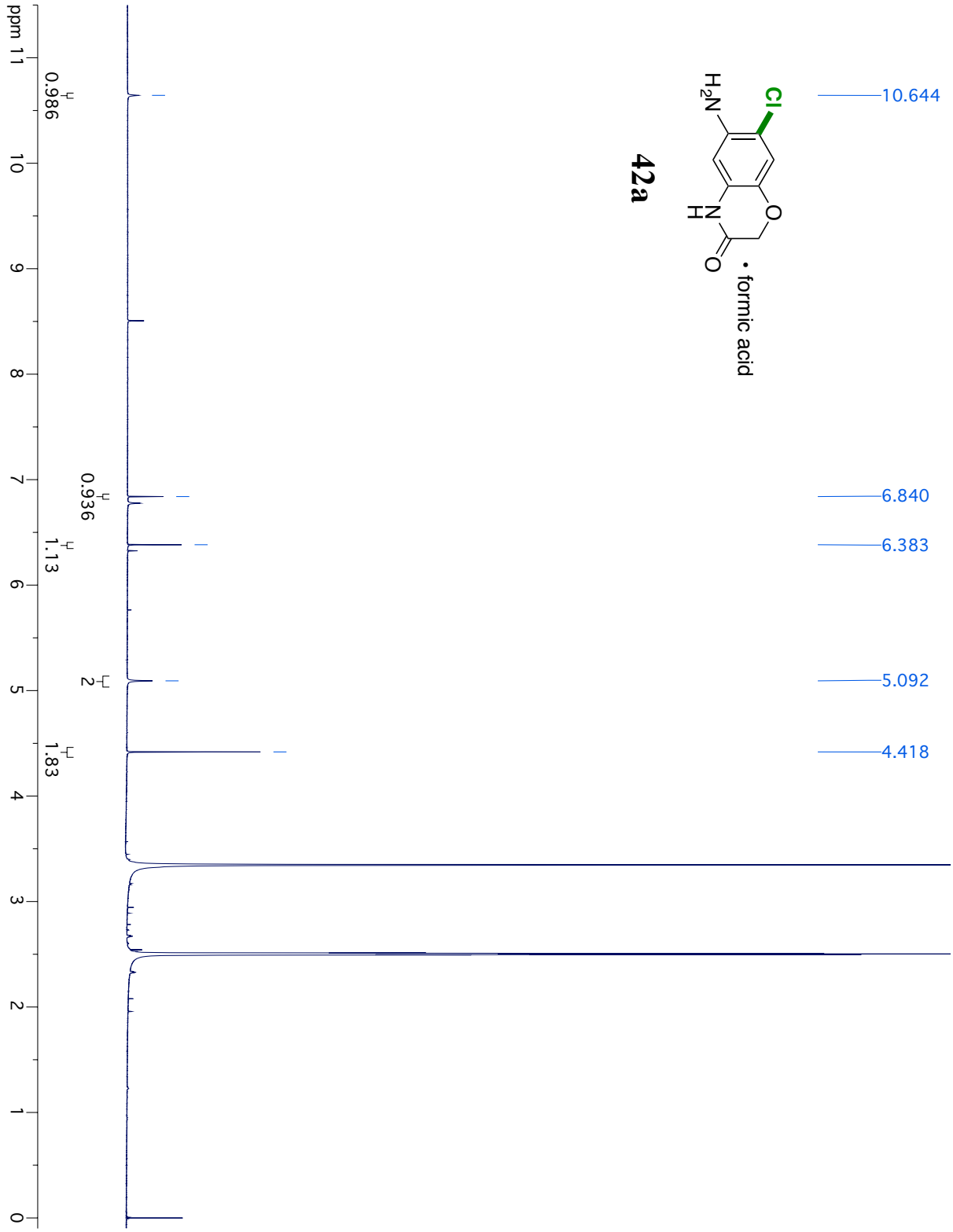


Figure AIV.40: NOSEY spectrum of **42a**.

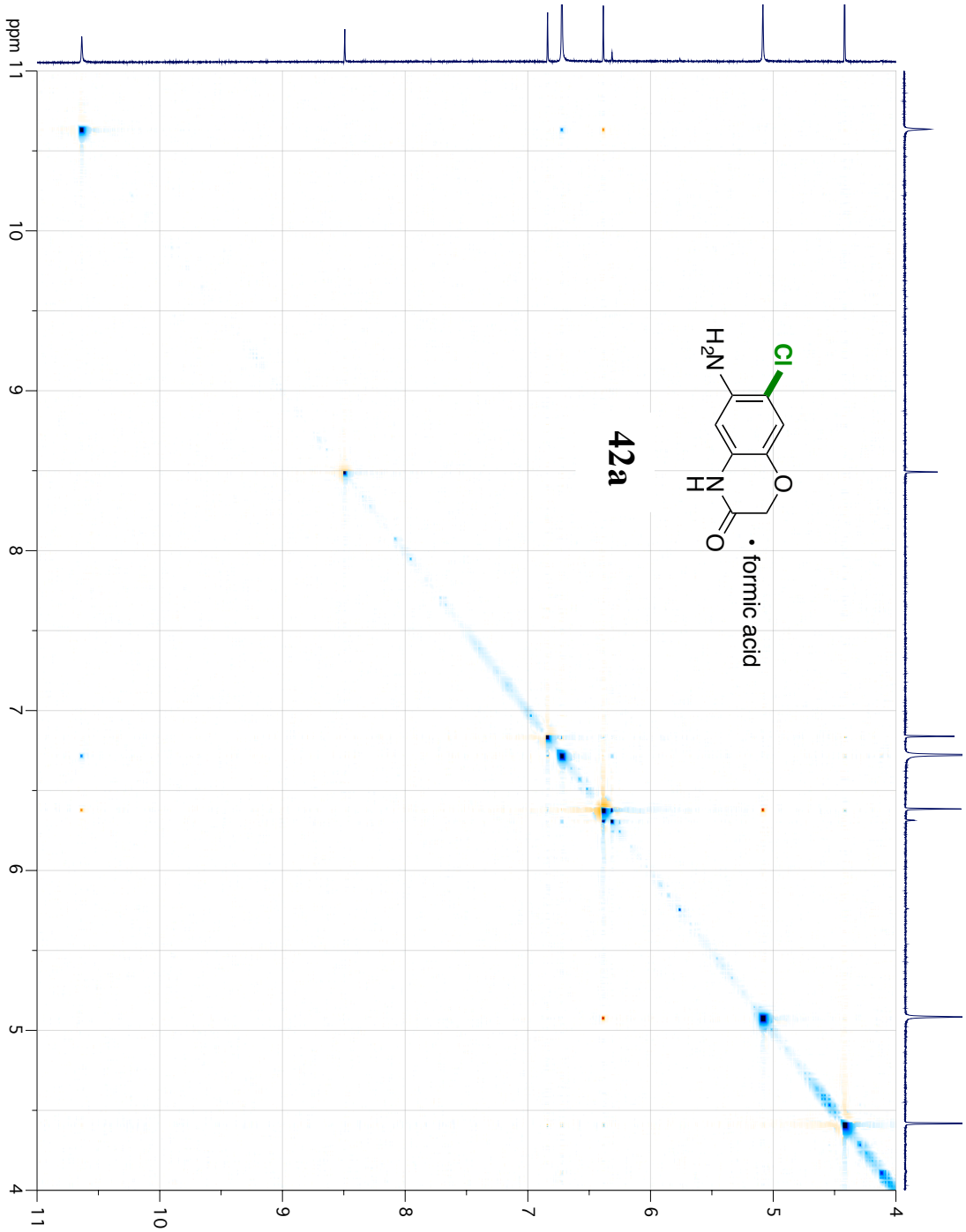


Figure AIV.41: [1]HNMR spectrum of **43**.

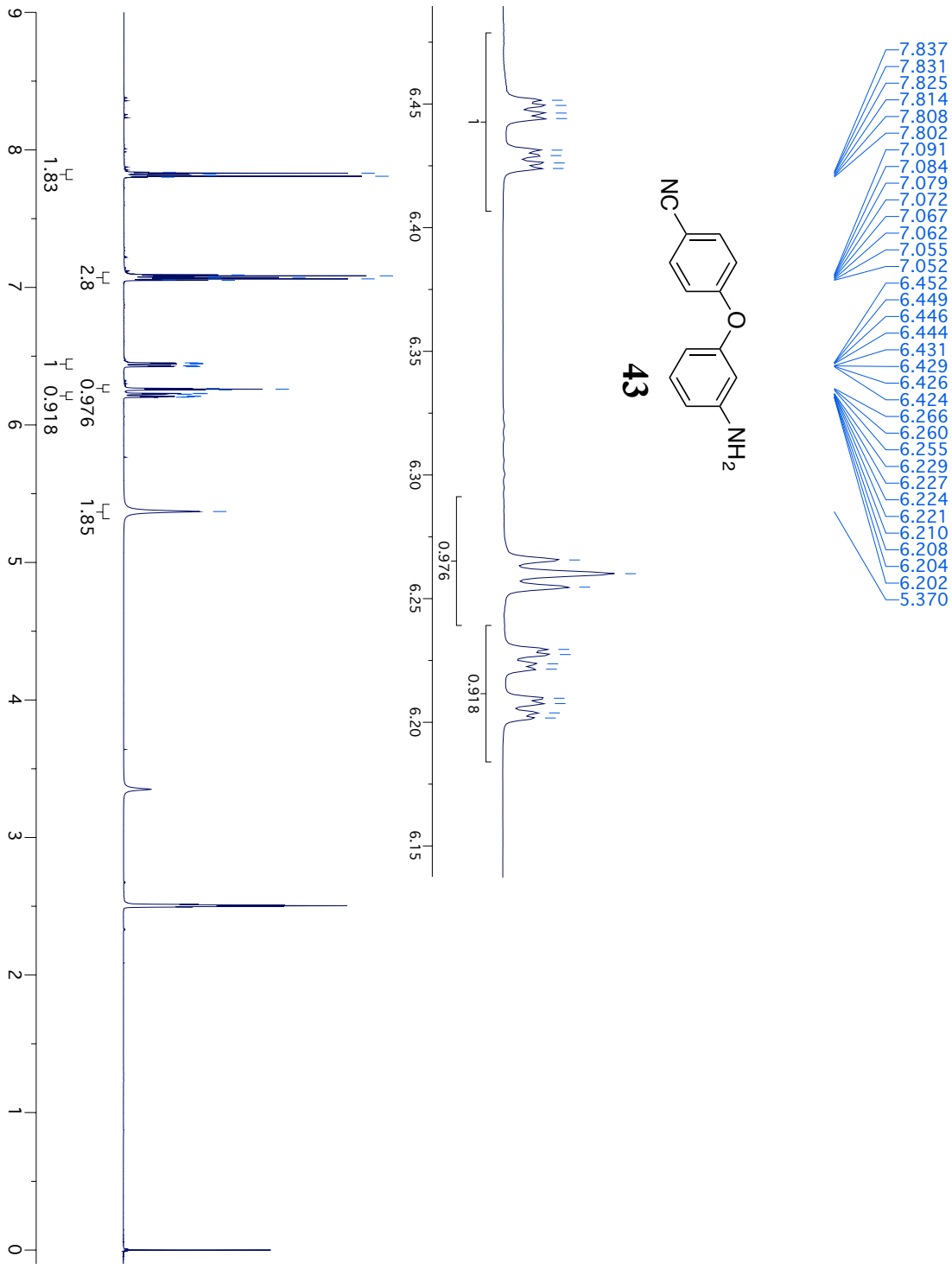


Figure AIV.42: [1]HNMR spectrum of **43a**.

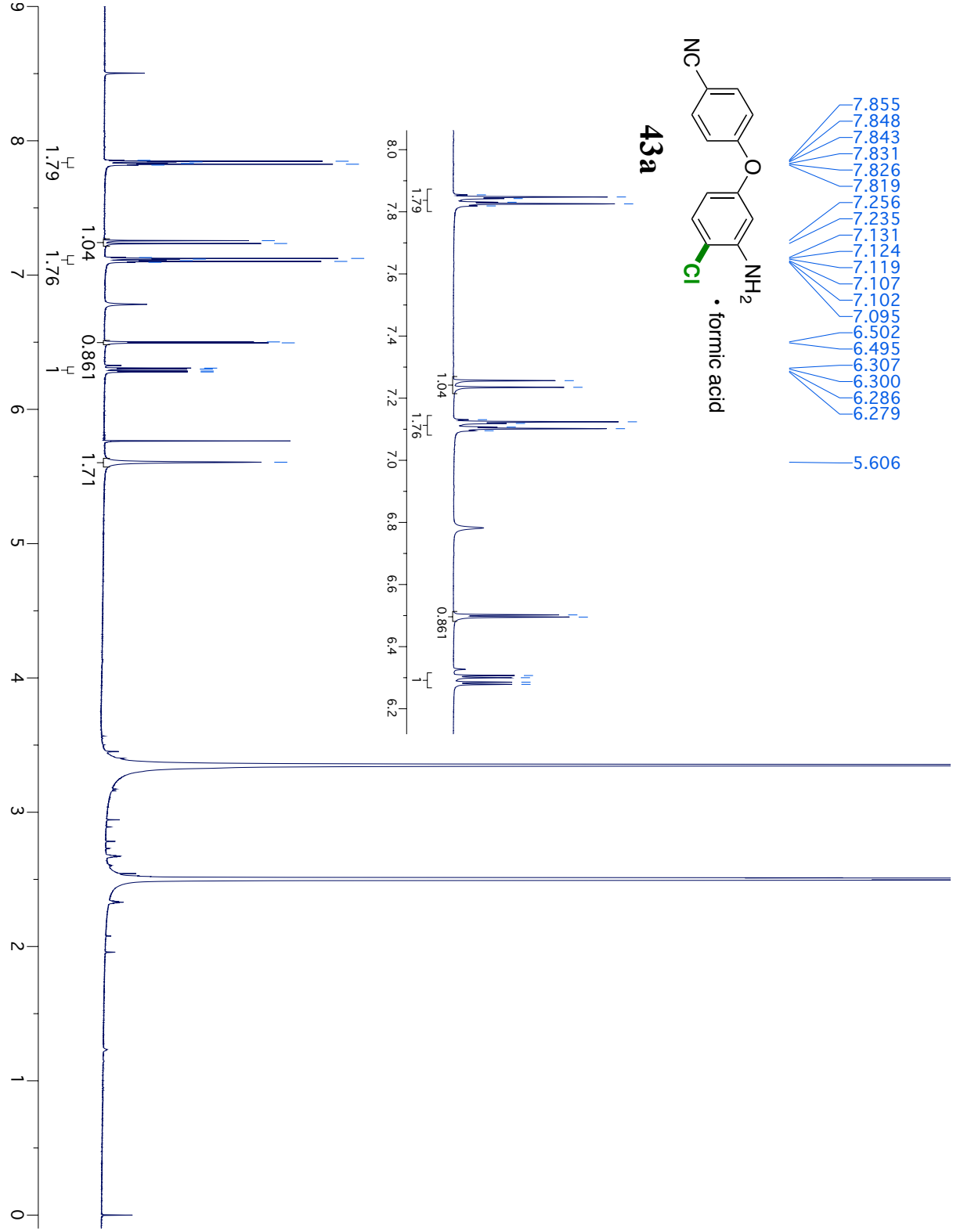


Figure AIV.43: NOSEY spectrum of **43a**.

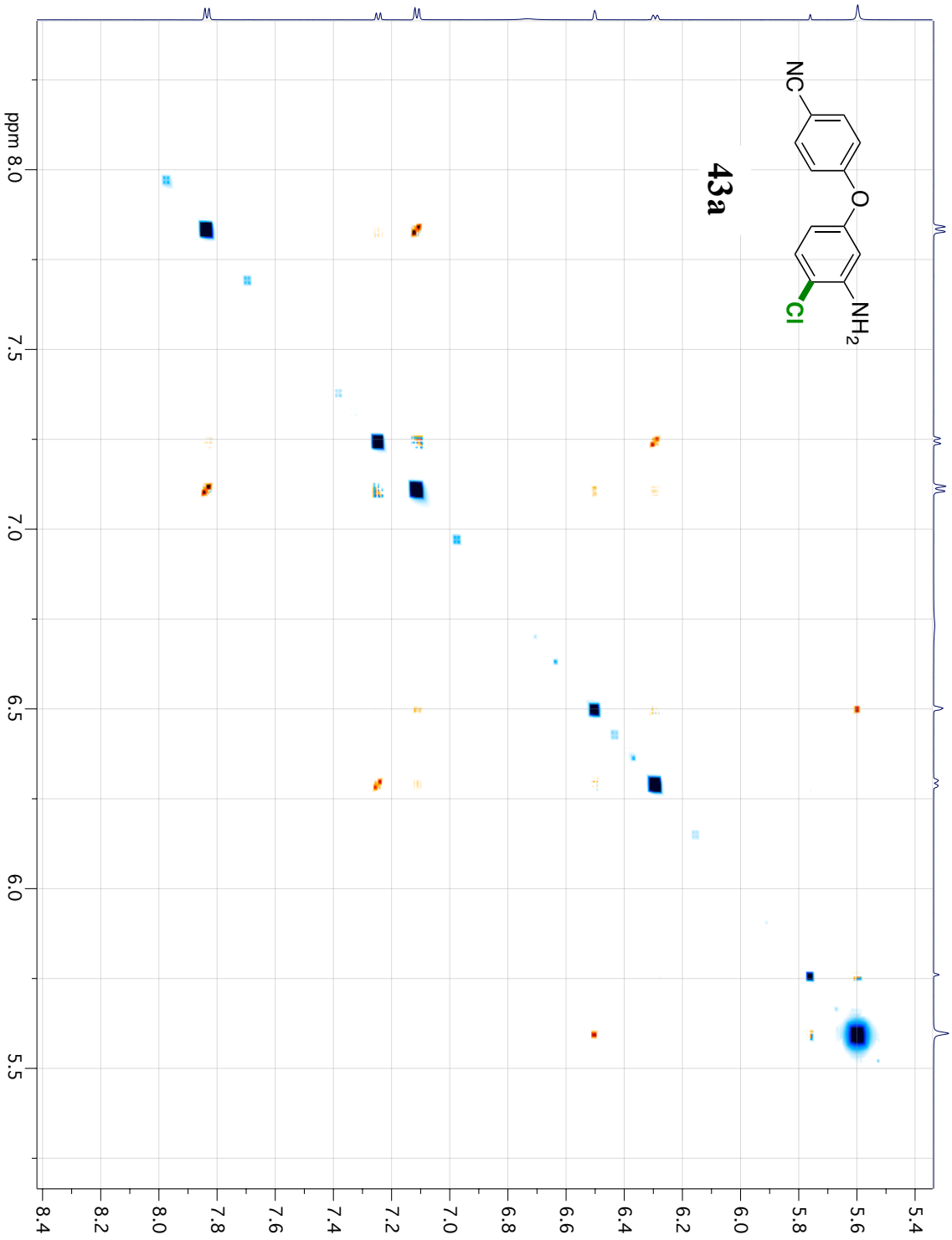


Figure AIV.44: [1]HNMR spectrum of **44**.

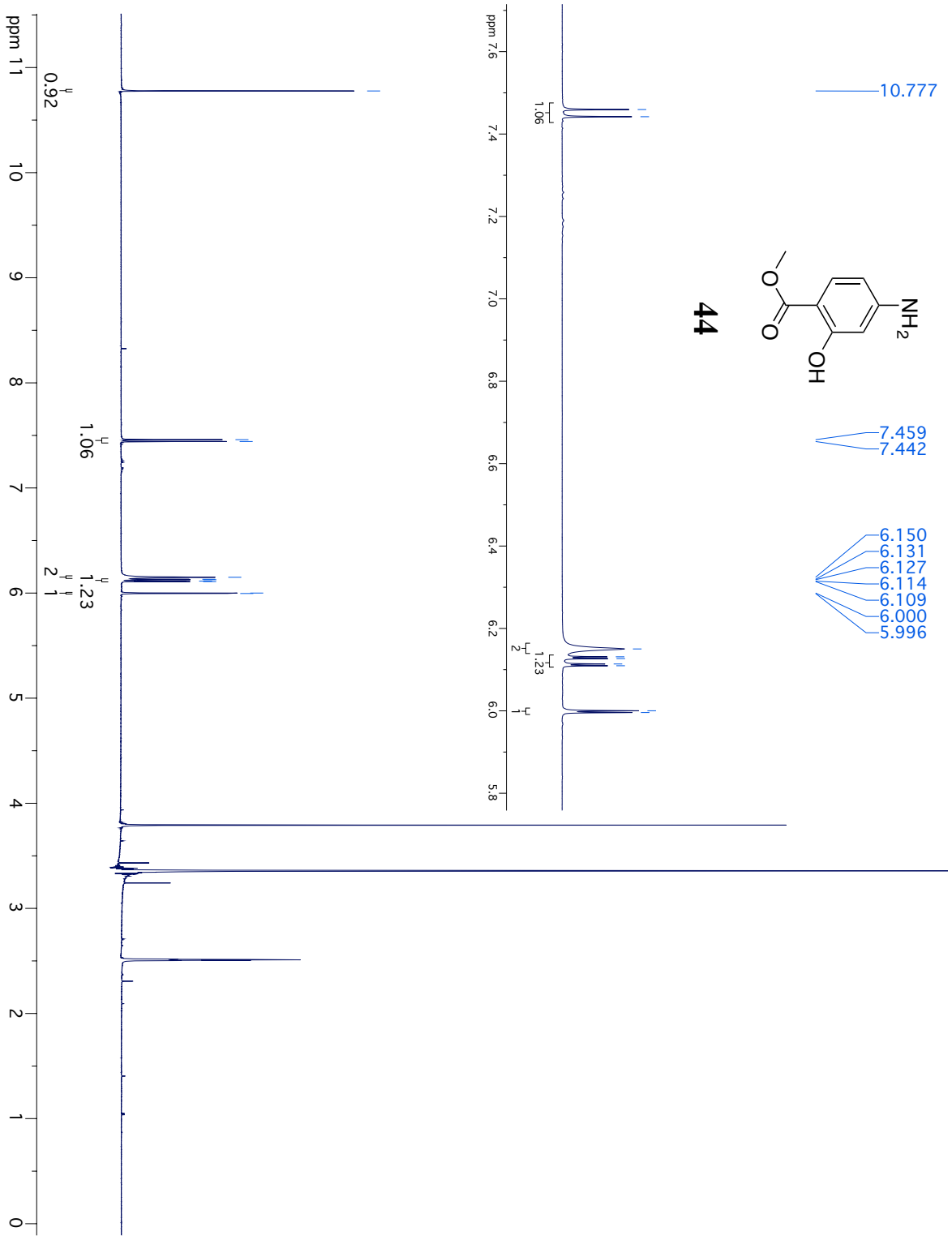


Figure AIV.45: [1]HNMR spectrum of **44a**.

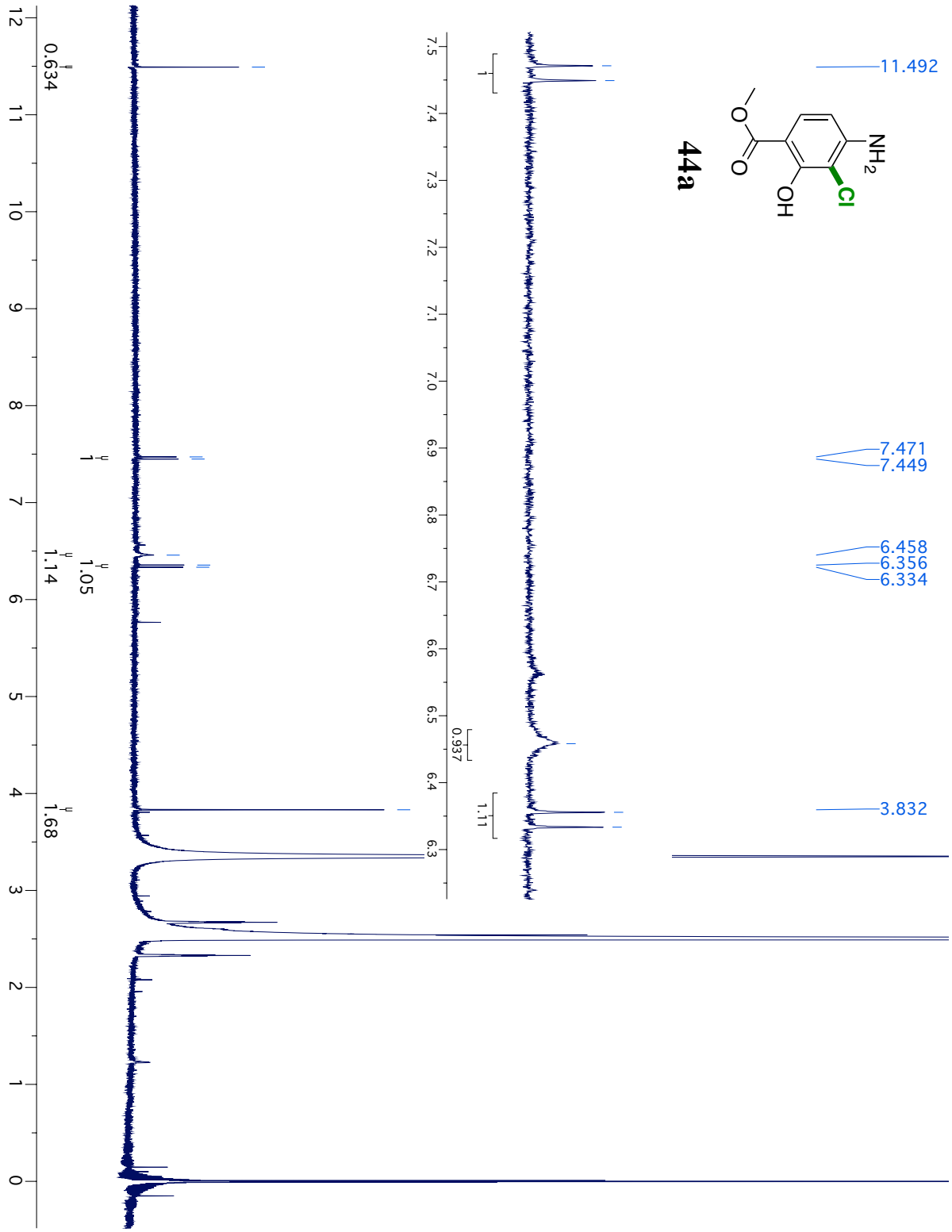


Figure AIV.46: [1]HNMR spectrum of **44b**.

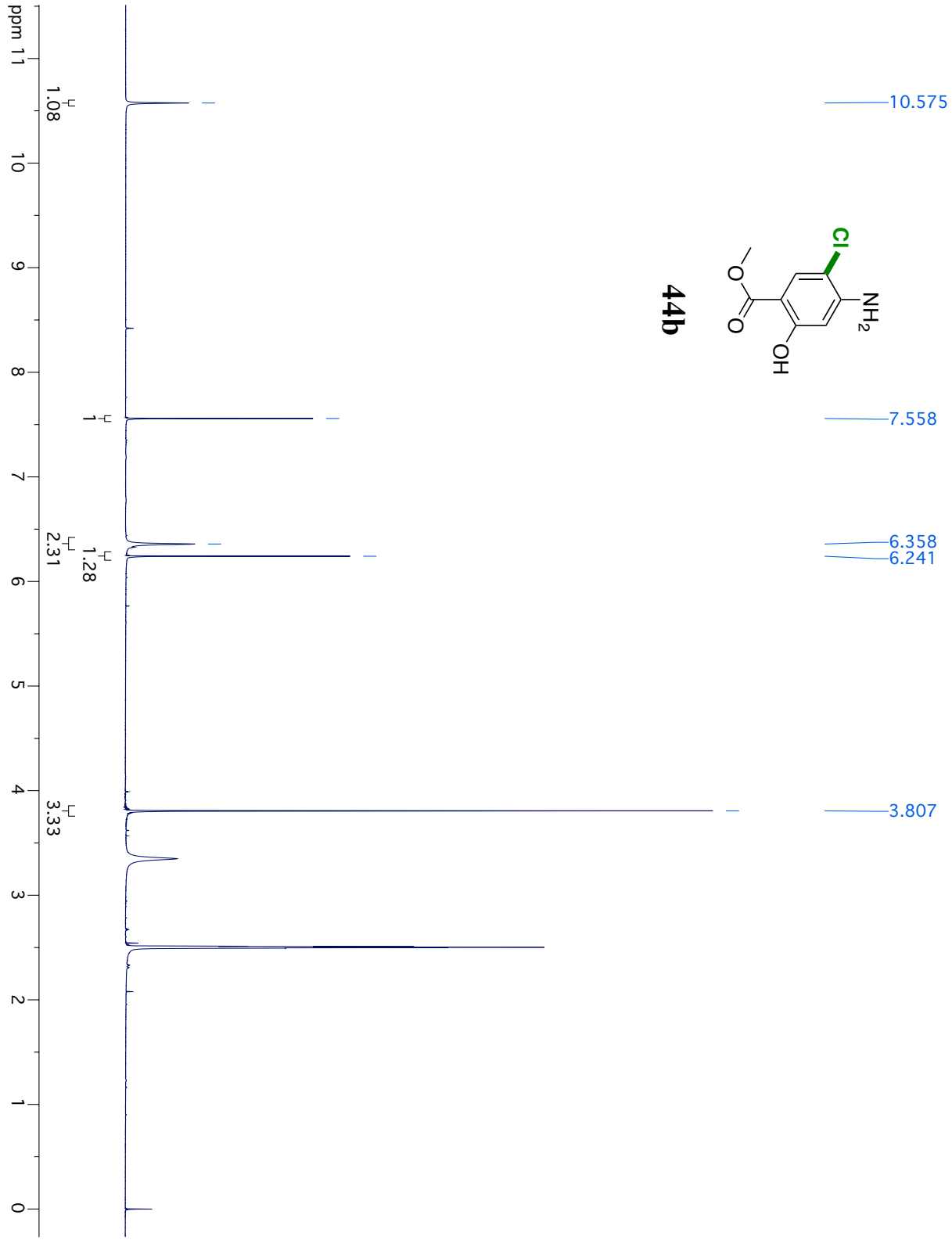


Figure AIV.47: NOSEY spectrum of **44b**.

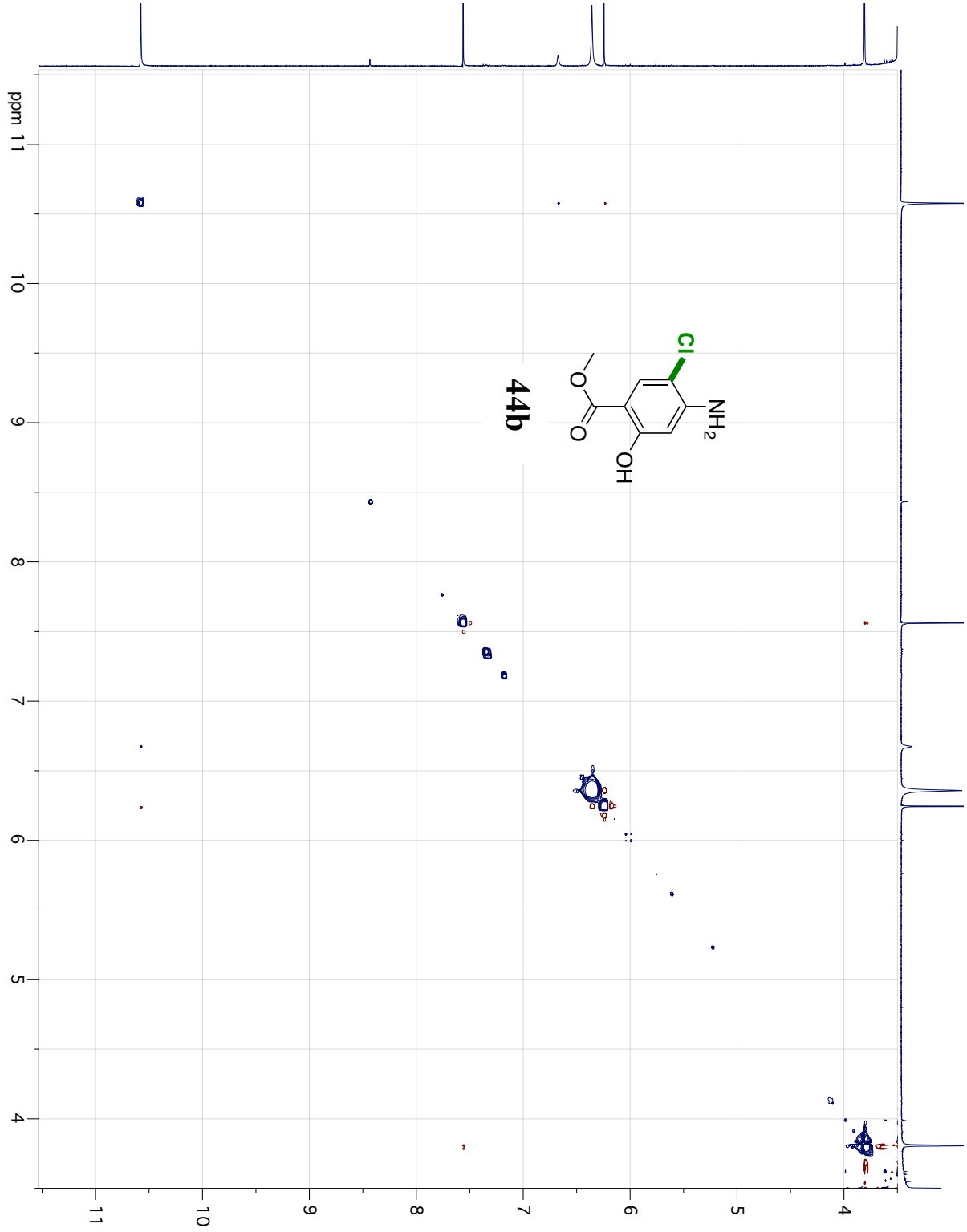


Figure AIV.48: [1]HNMR spectrum of **45**.

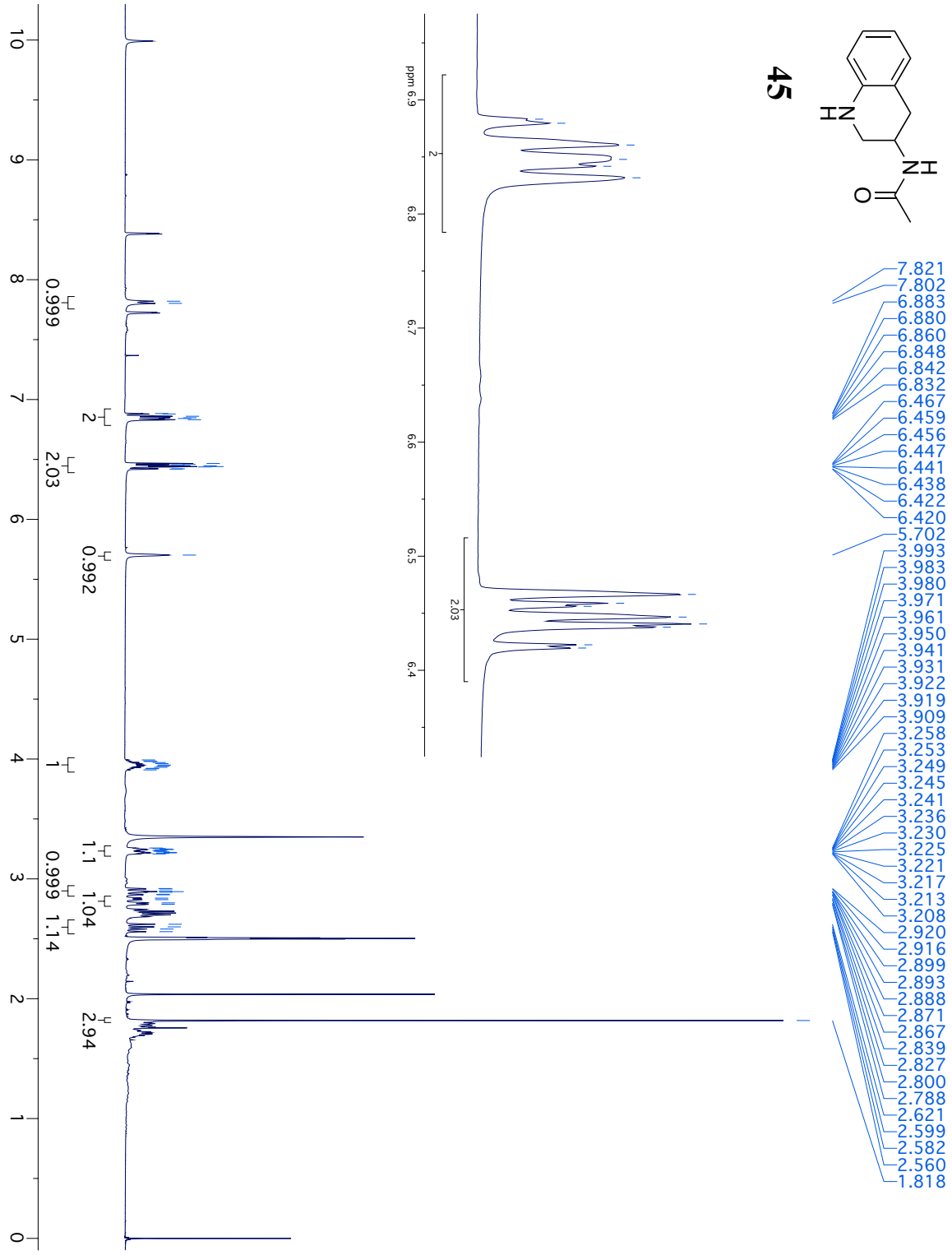


Figure AIV.49: [1]HNMR spectrum of **45a**.

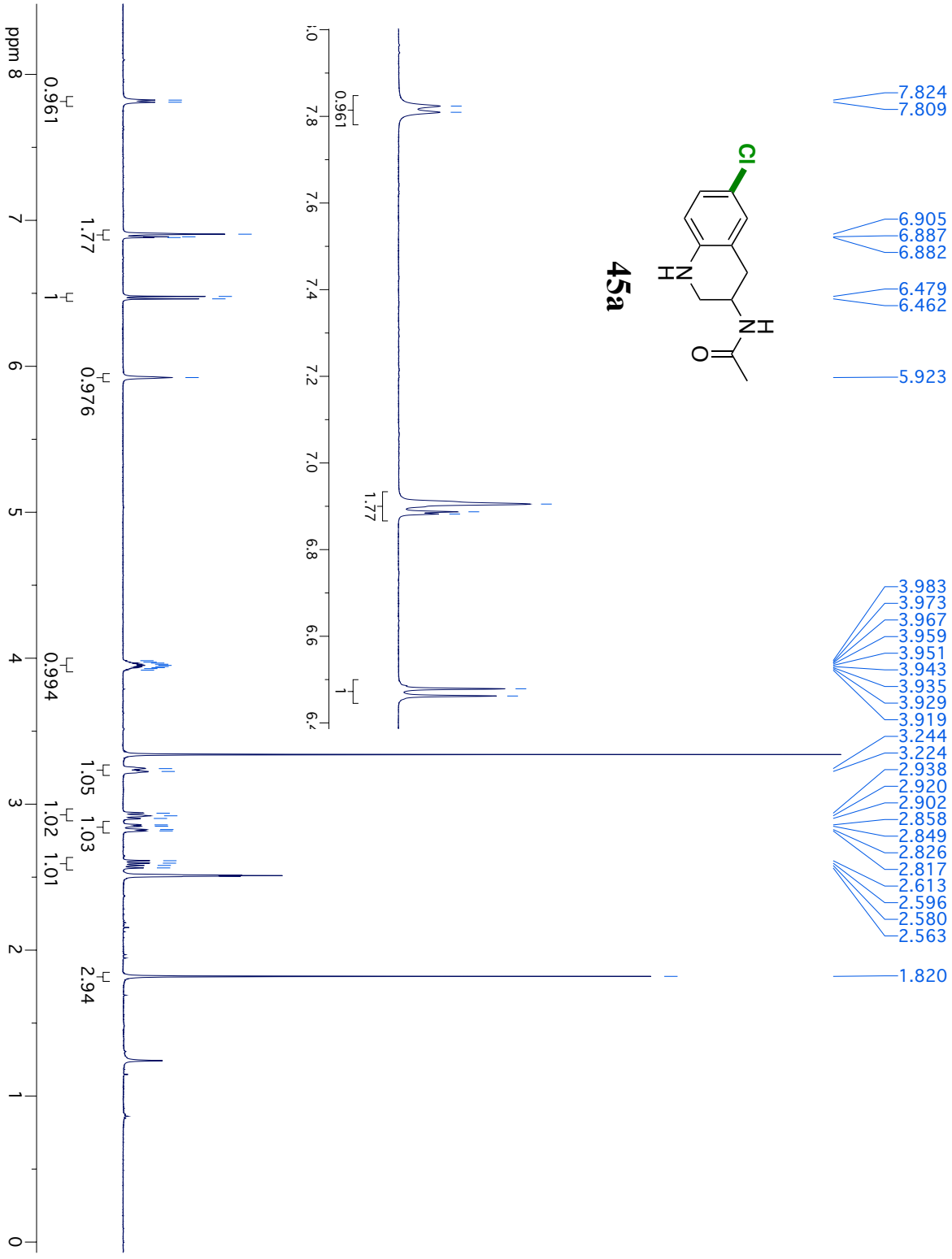


Figure AIV.50: NOSEY spectrum of **45a**.

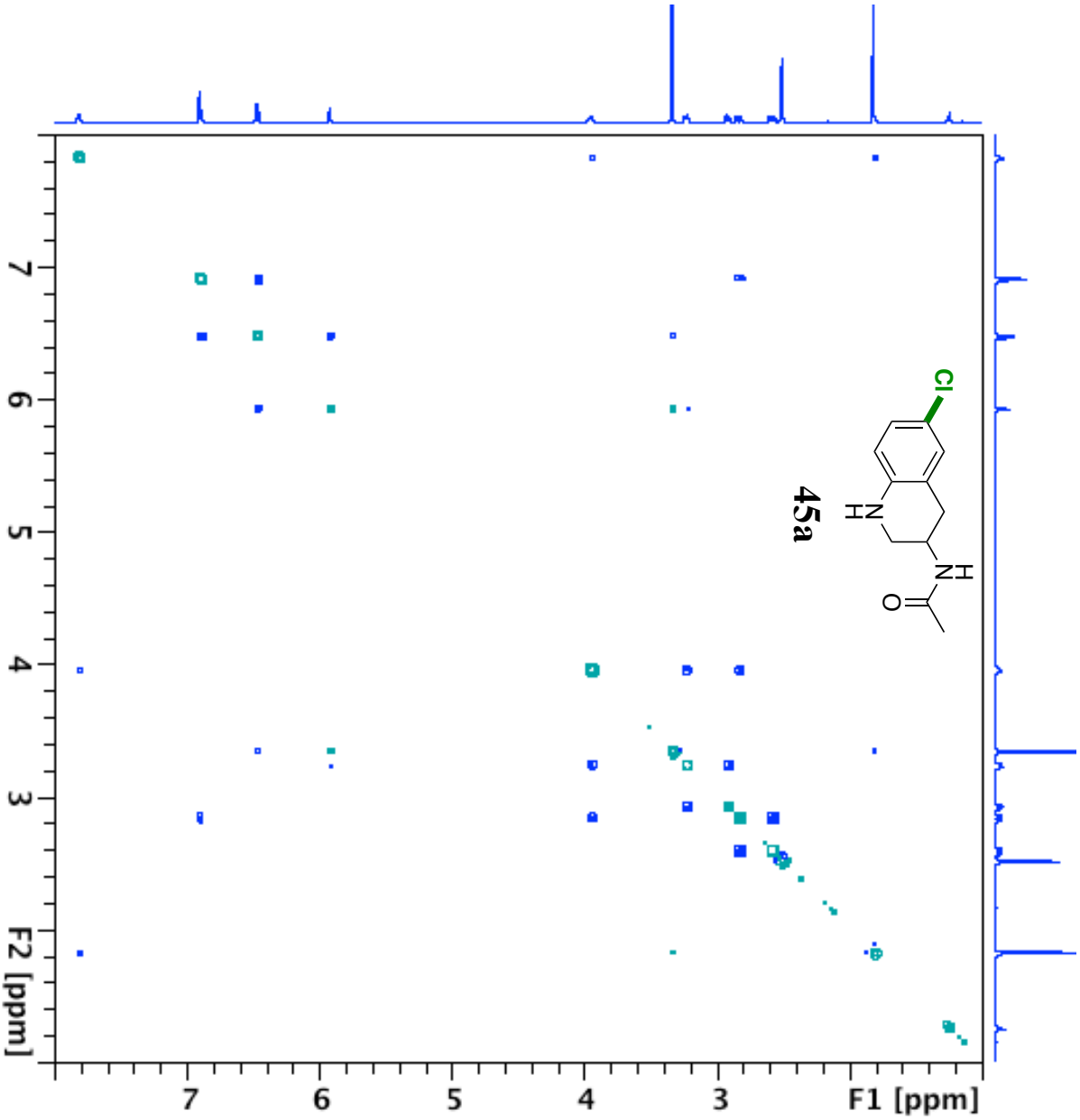


Figure AIV.51: [1]HNMR spectrum of **46**.

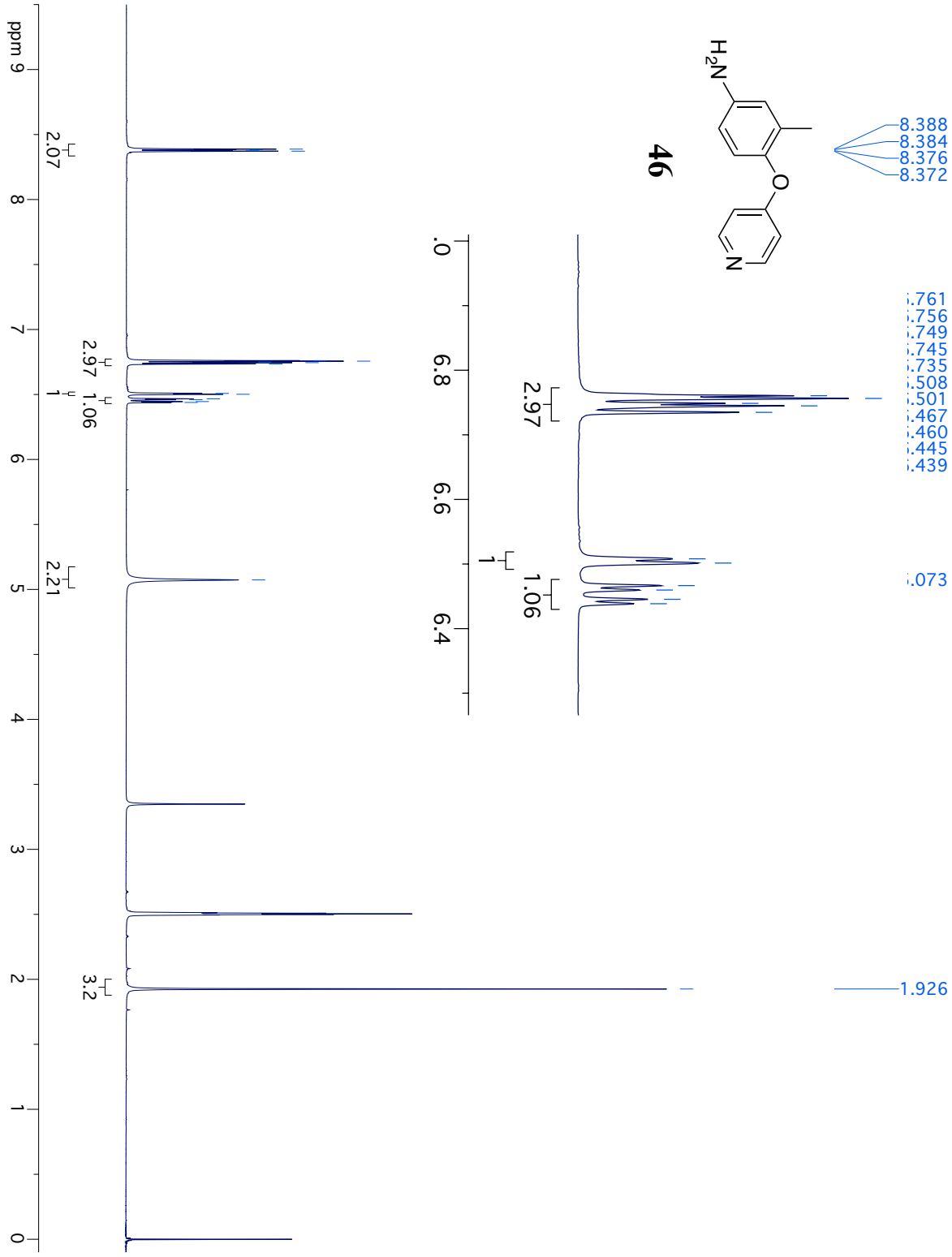


Figure AIV.52: [1]HNMR spectrum of **46a**.

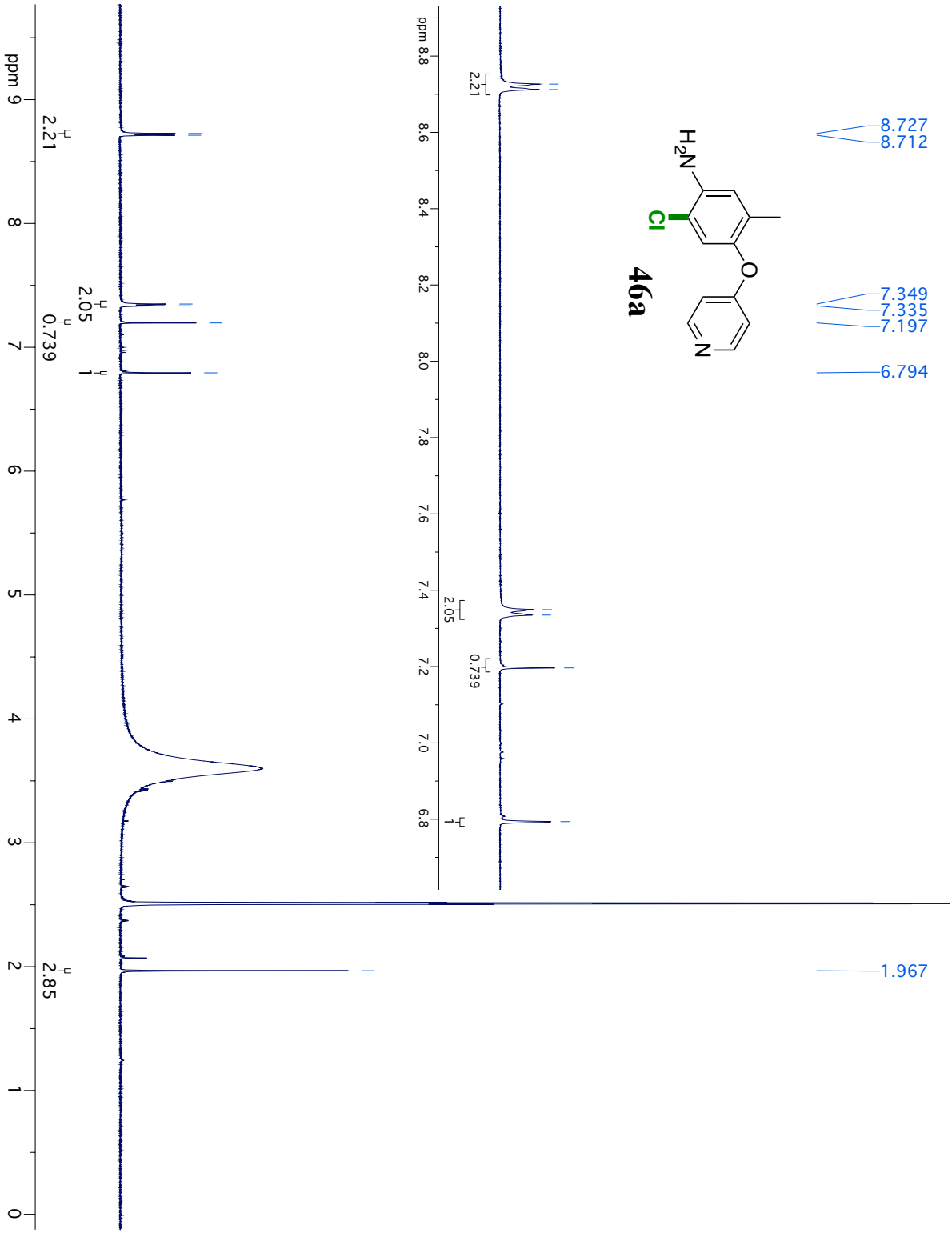


Figure AIV.53: NOSEY spectrum of **46a**.

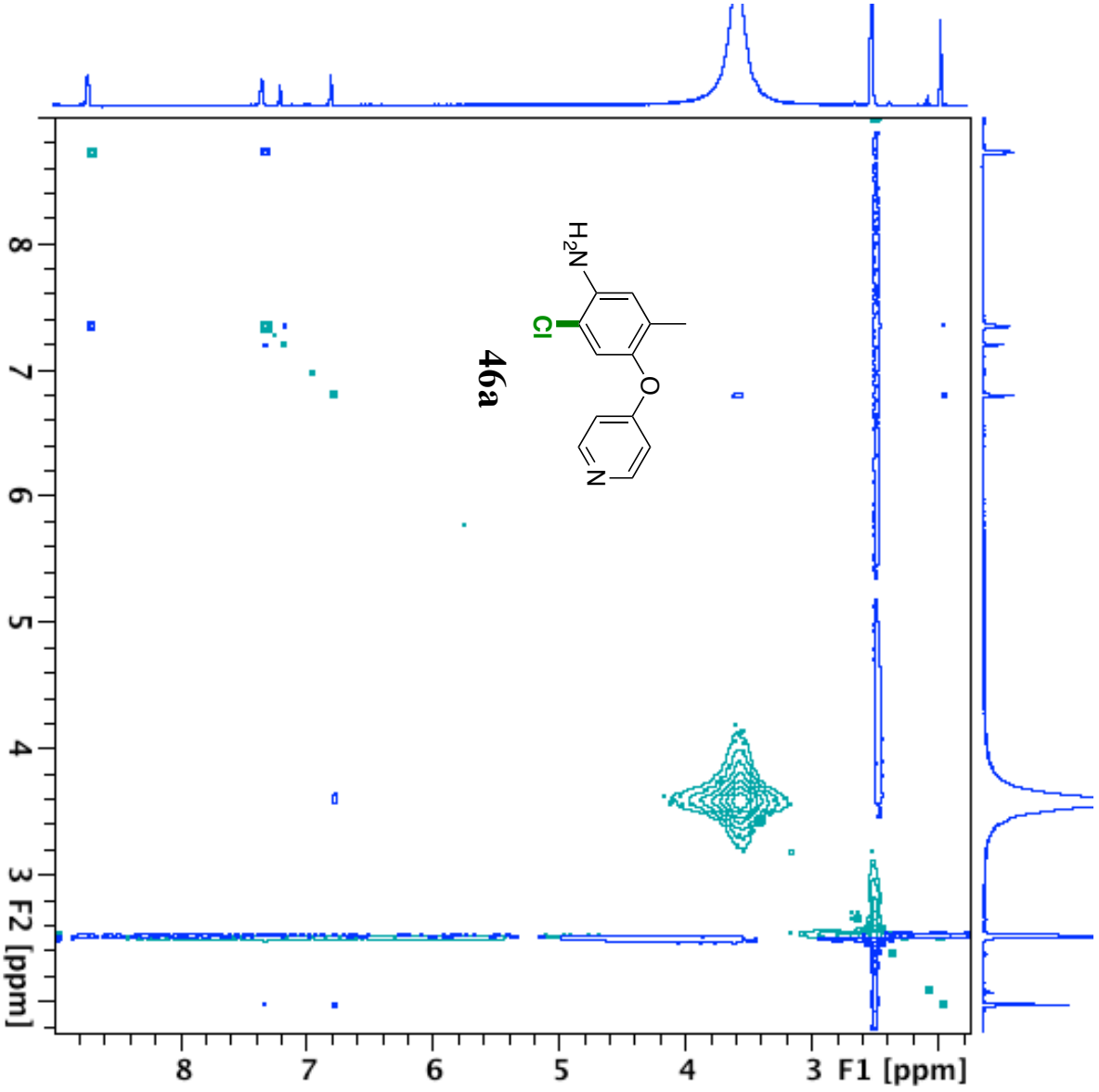


Figure AIV.54: [1]HNMR spectrum of **46b**.

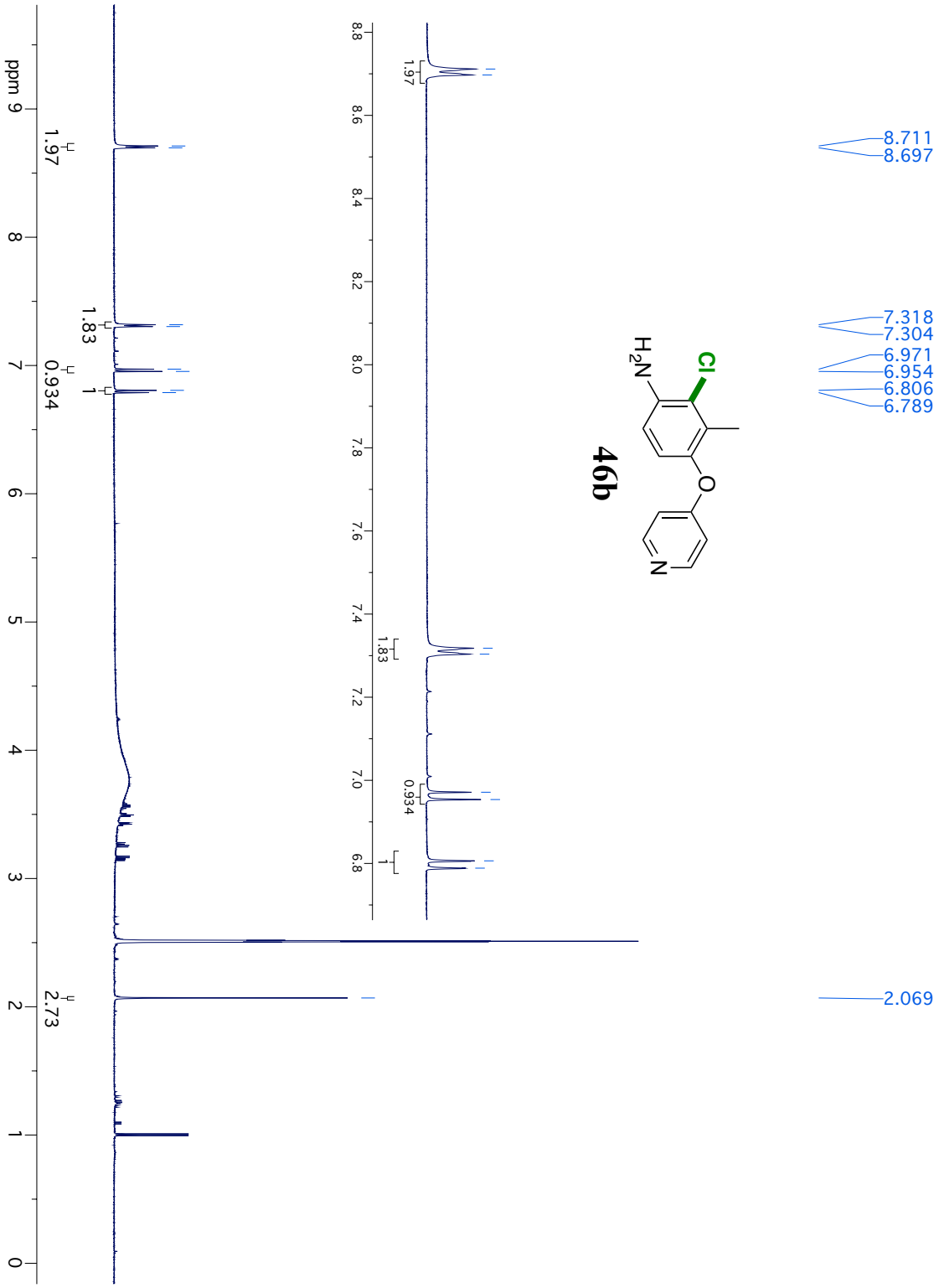


Figure AIV.55: [1]HNMR spectrum of **48**.

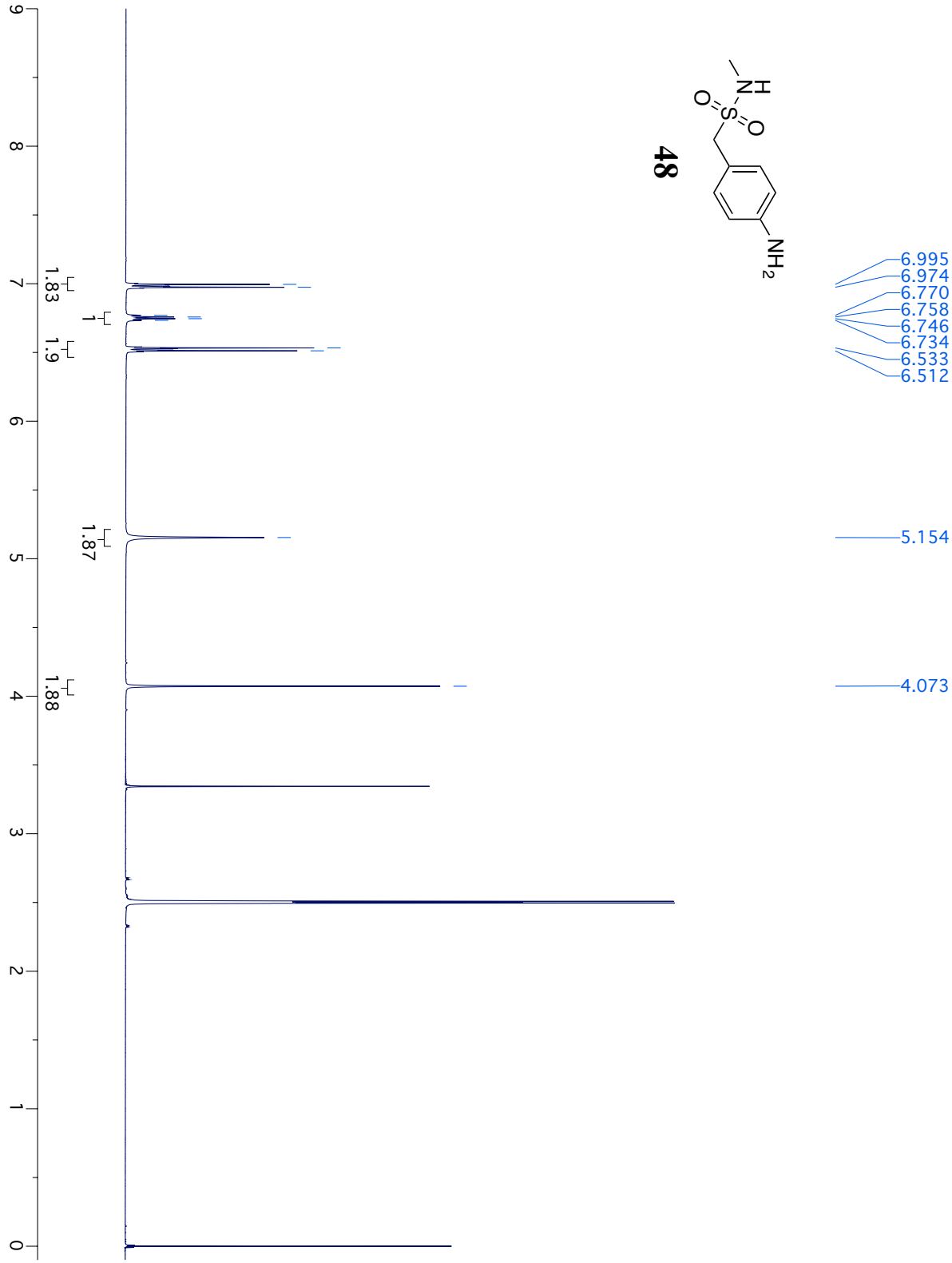


Figure AIV.56: [1]HNMR spectrum of **48a**.

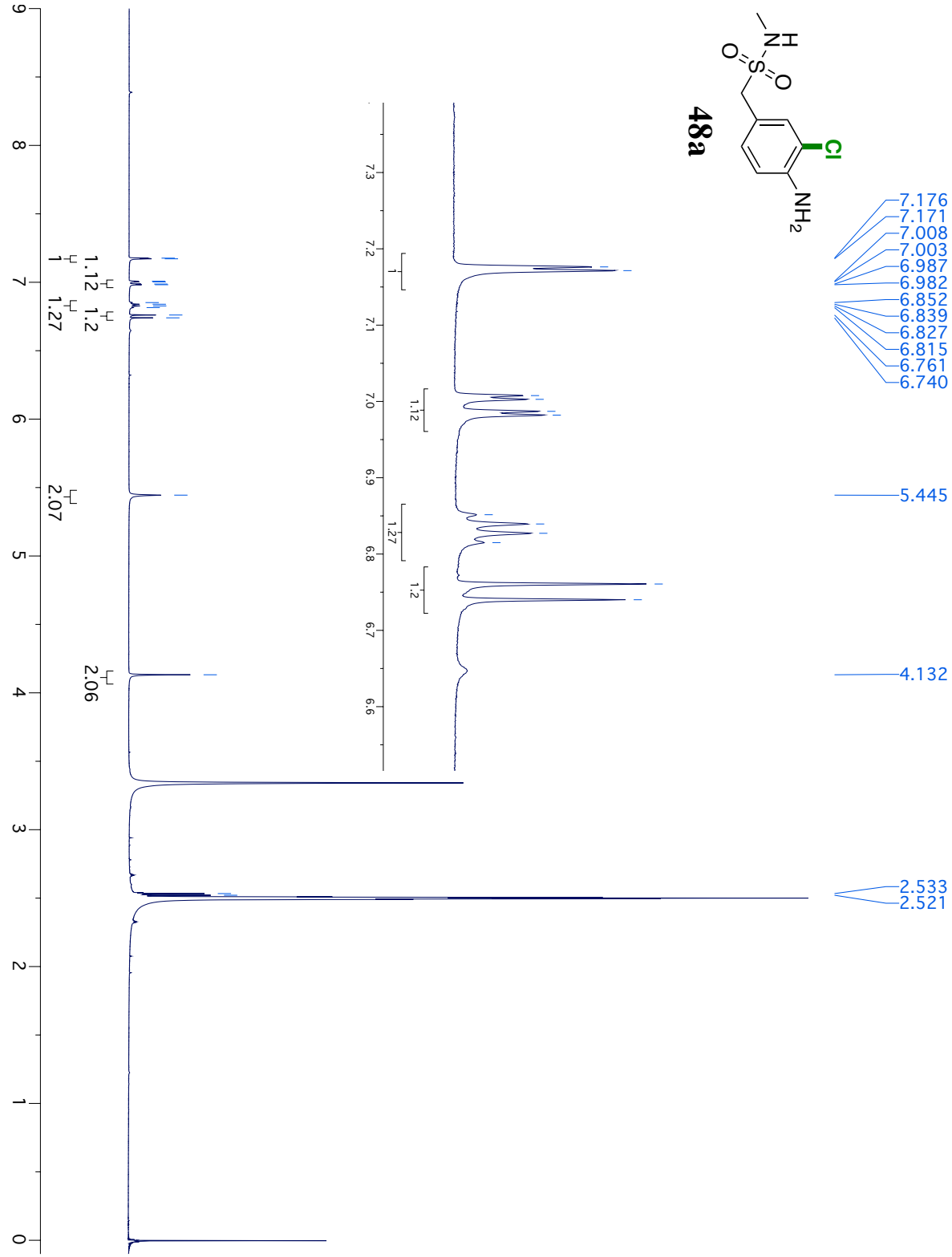


Figure AIV.57: NOSEY spectrum of **48a**.

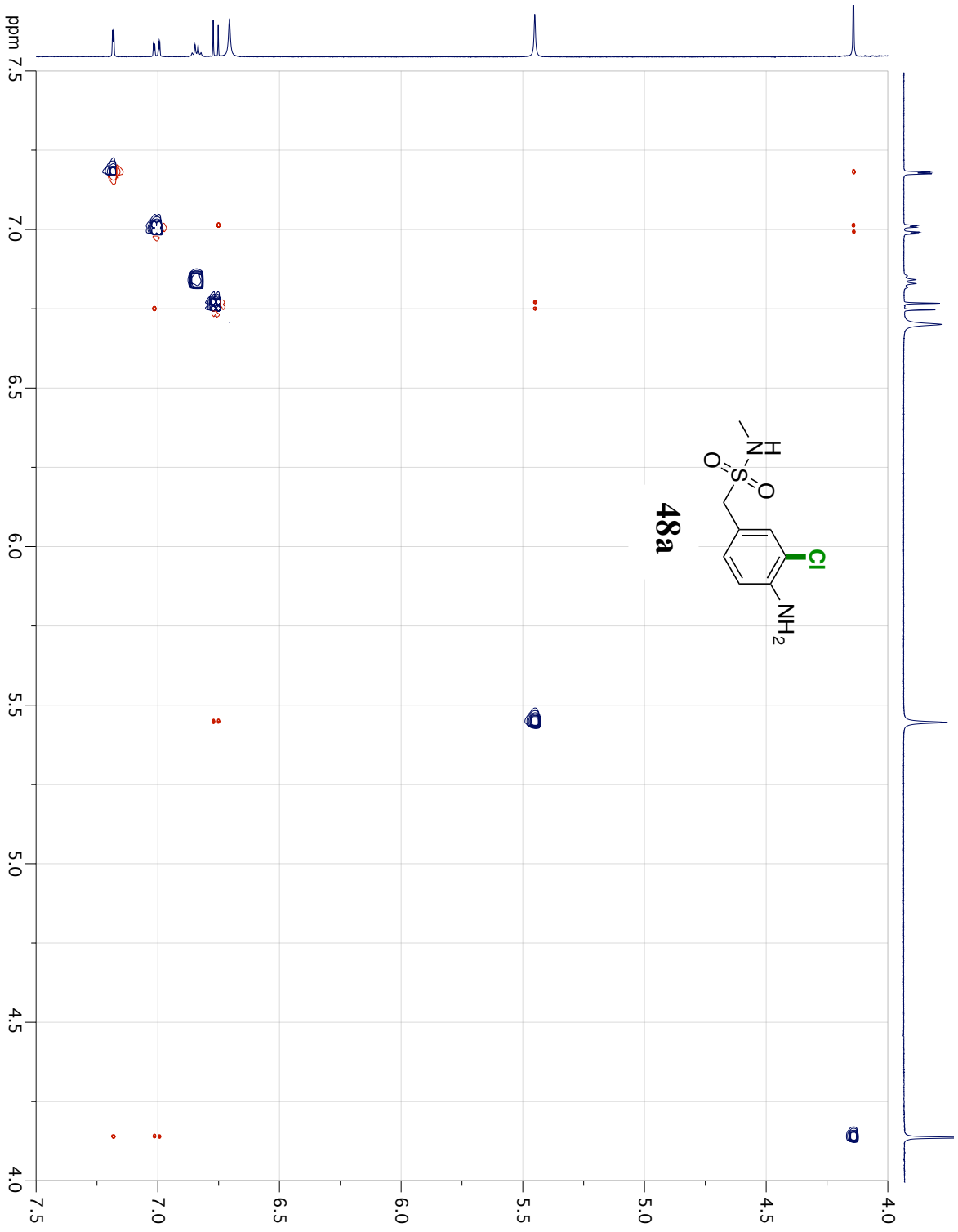


Figure AIV.58: [1]HNMR spectrum of **52**.

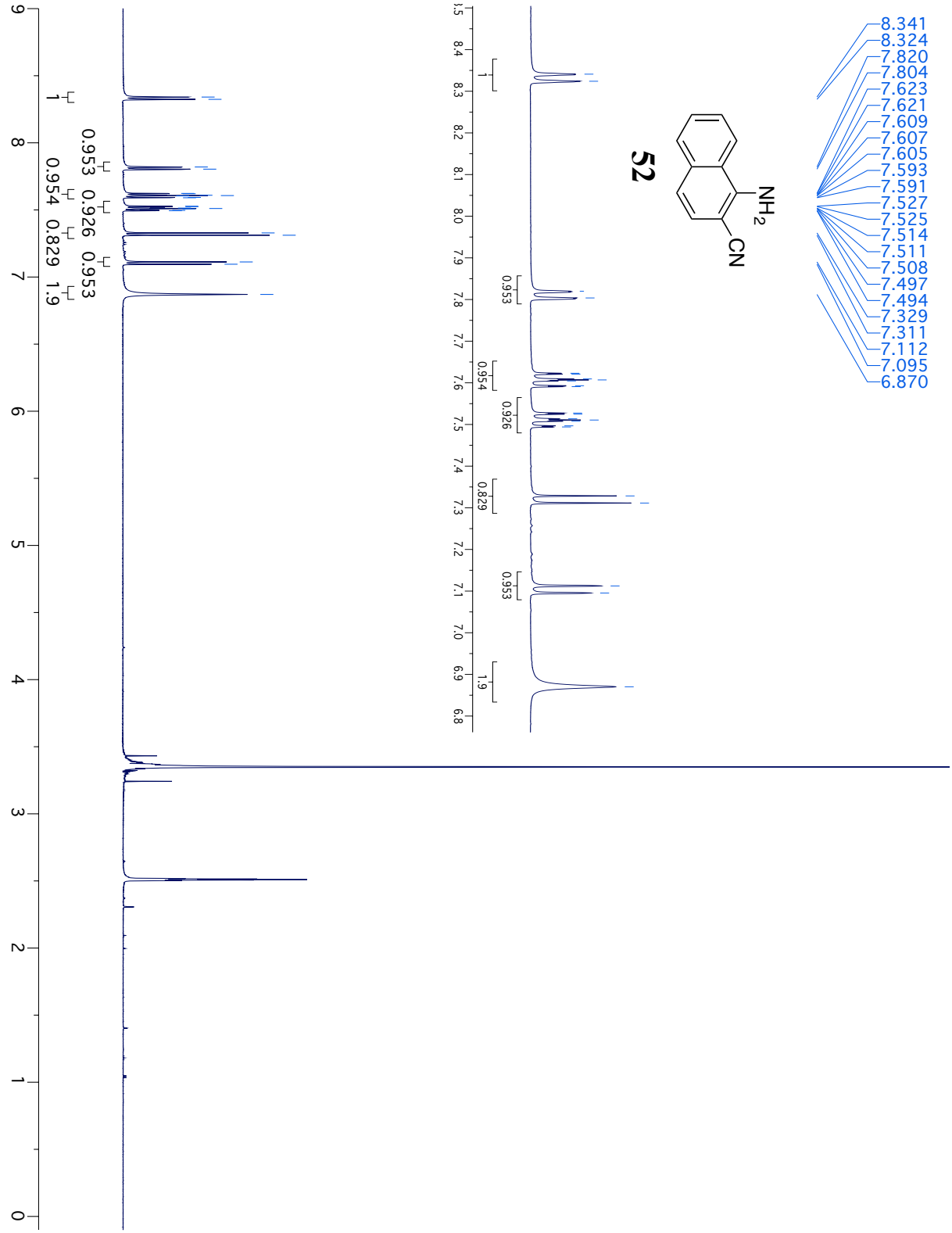


Figure AIV.59: [1]HNMR spectrum of **52a**.

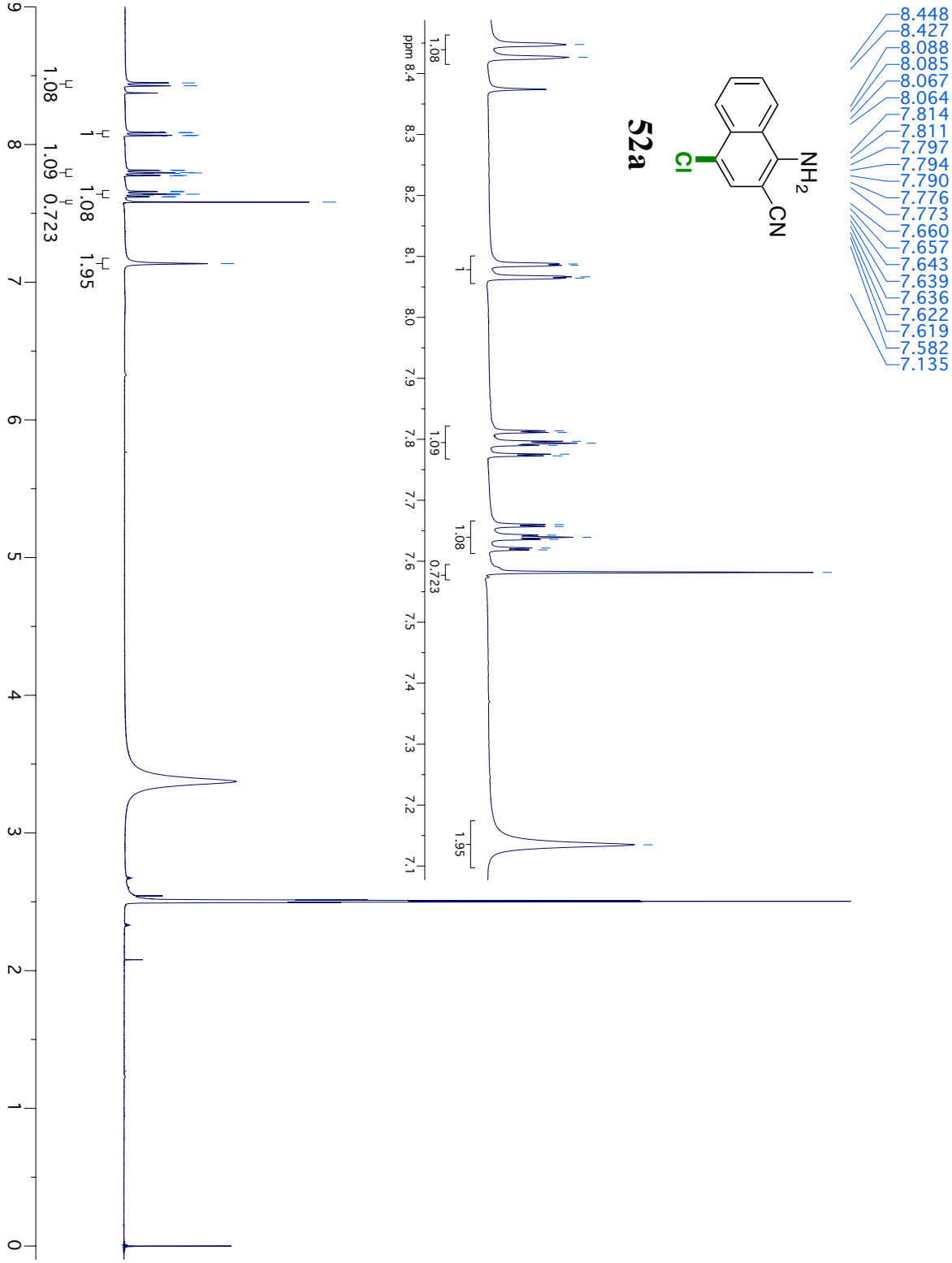


Figure AIV.60: NOSEY spectrum of **52a**.

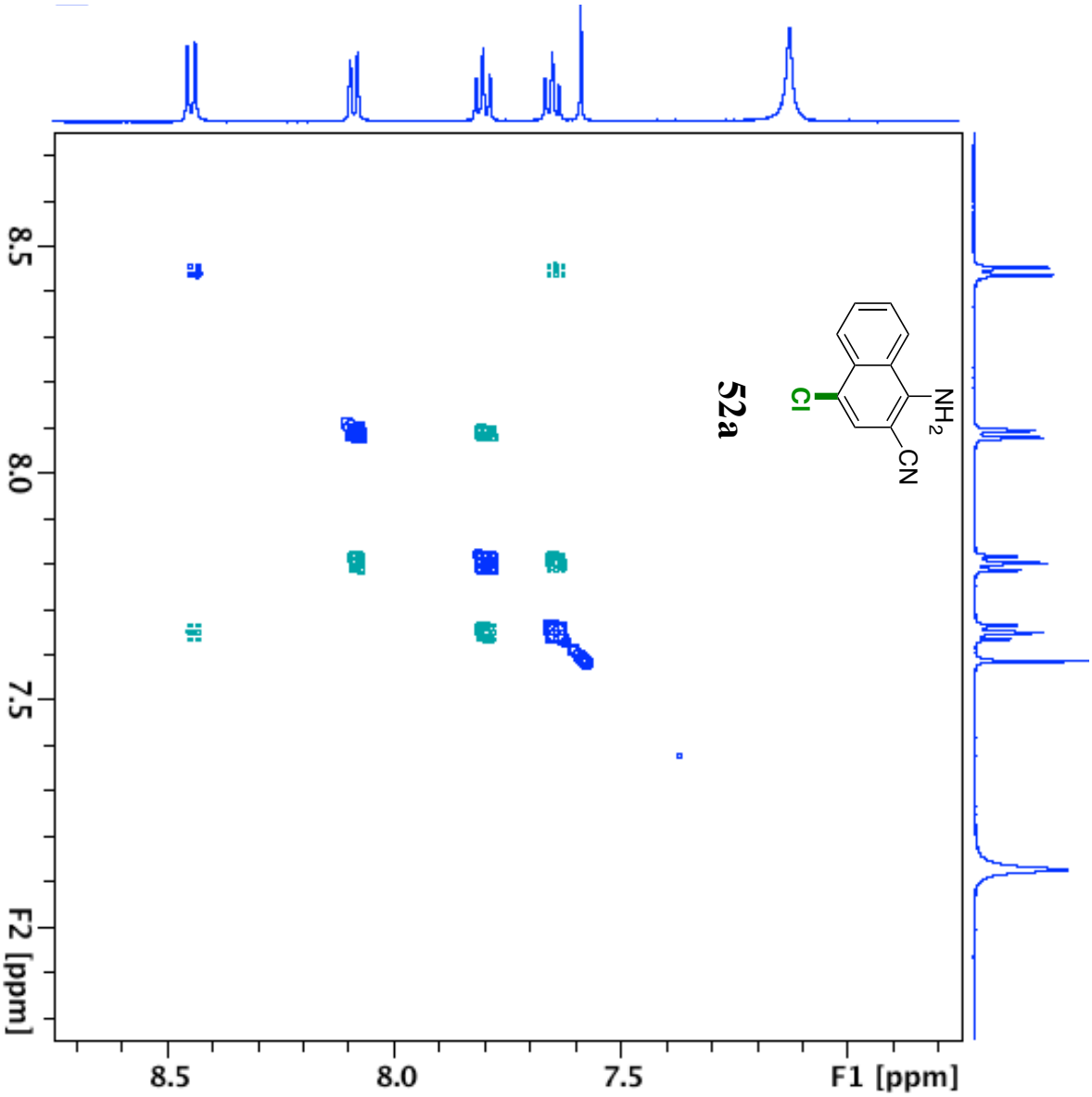


Figure AIV.61: [1]HNMR spectrum of **62**.

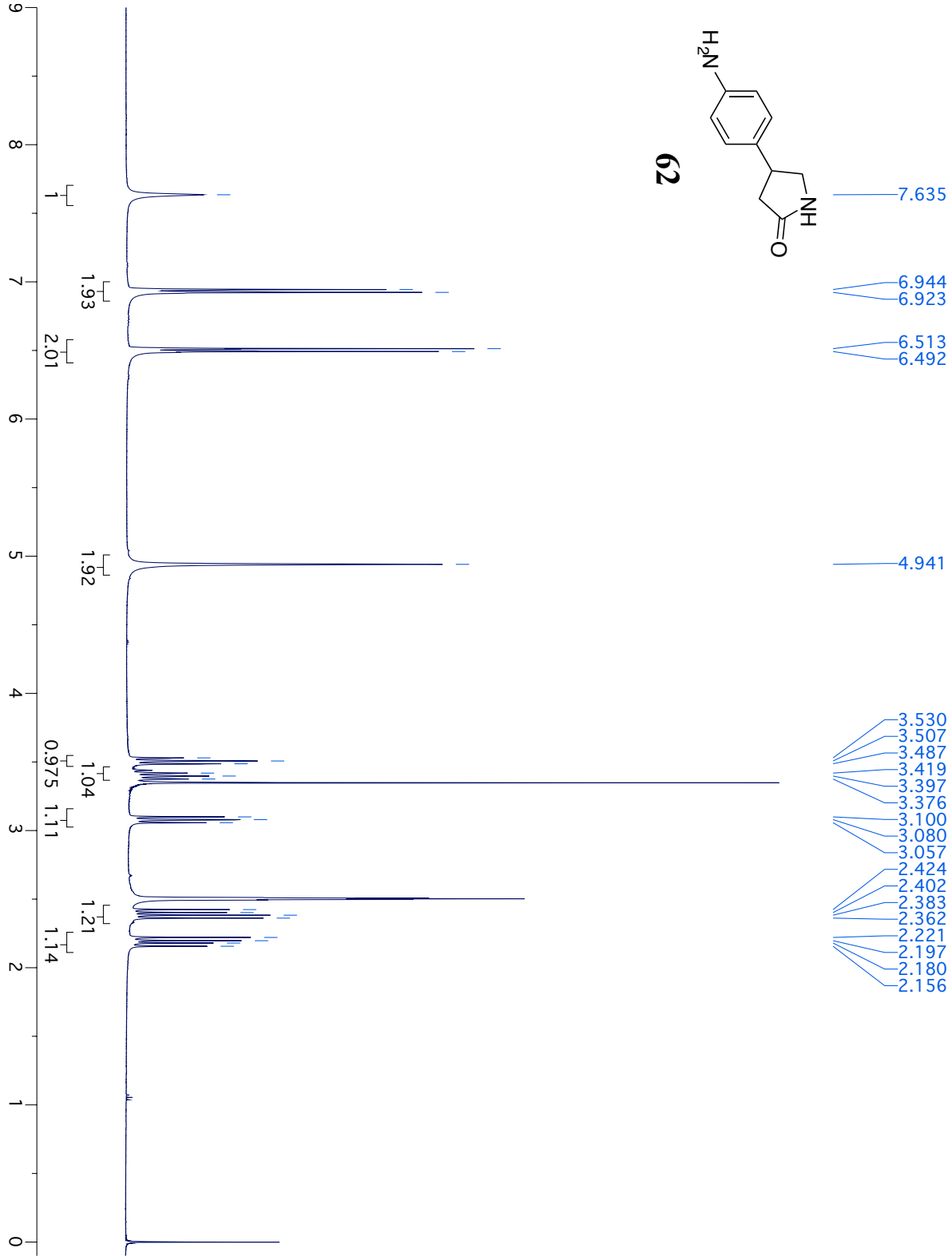


Figure AIV.62: [1]HNMR spectrum of **62a**.

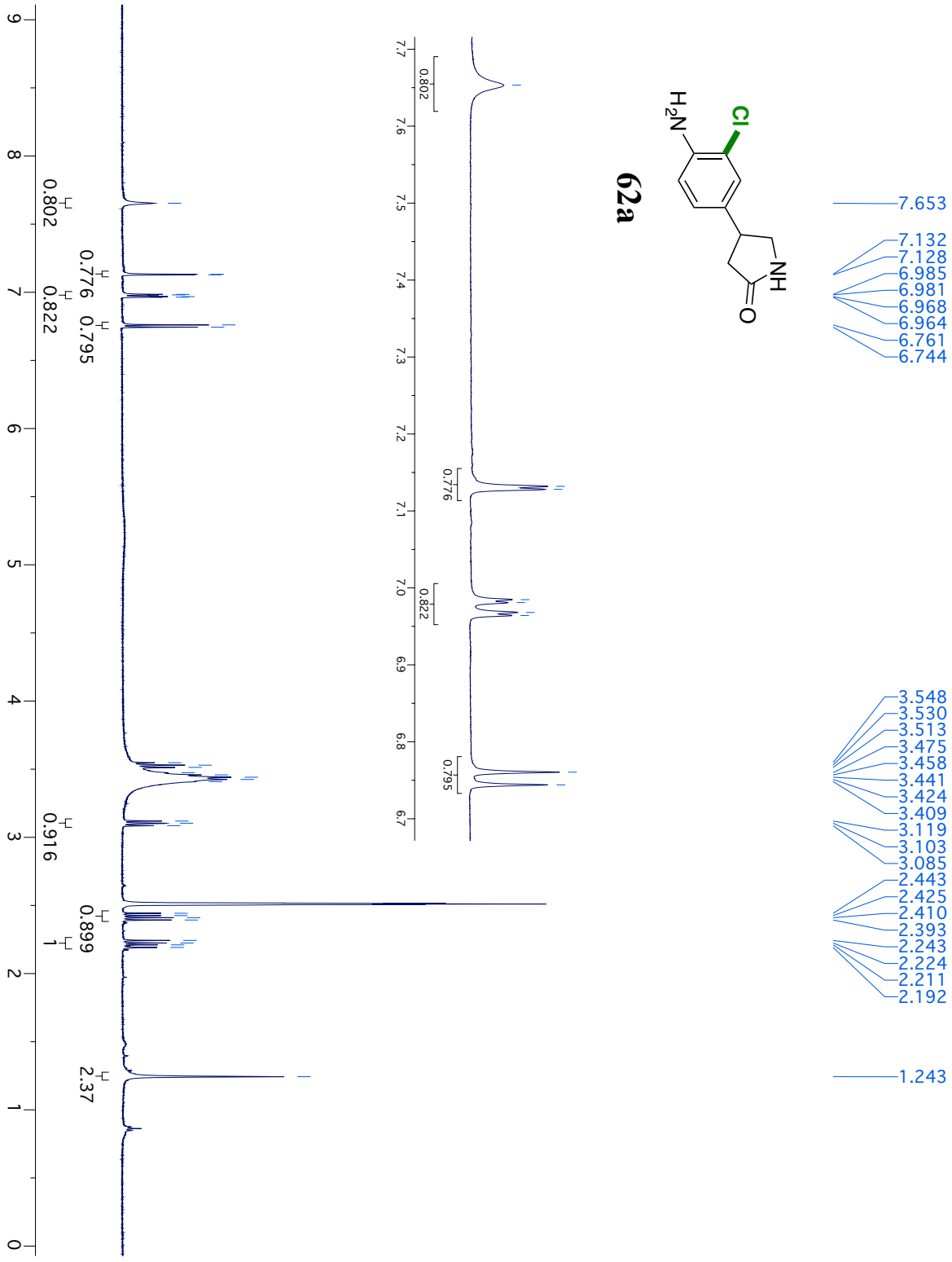


Figure AIV.63: NOSEY spectrum of **62a**.

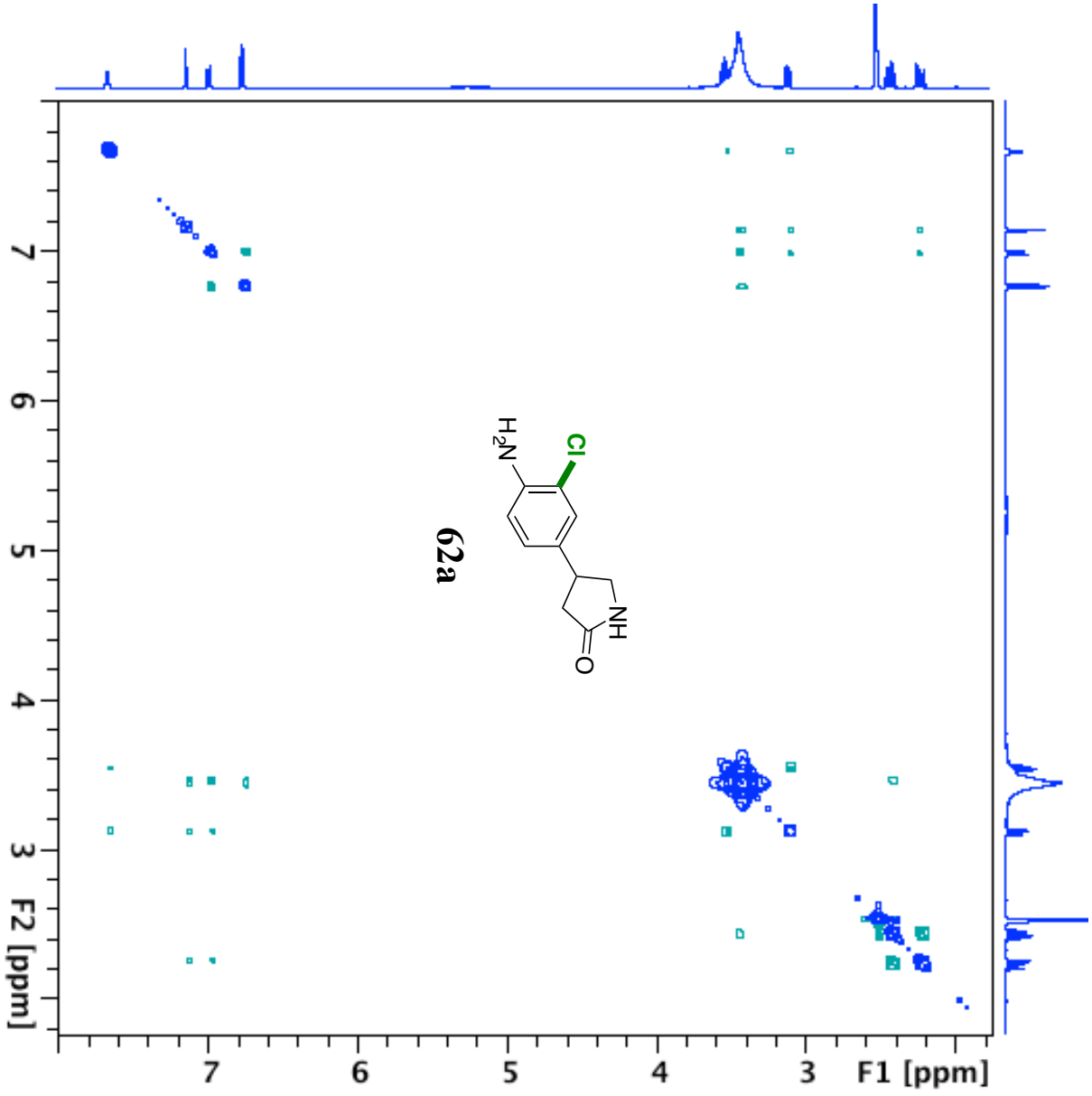


Figure AIV.64: [1]HNMR spectrum of **64**.

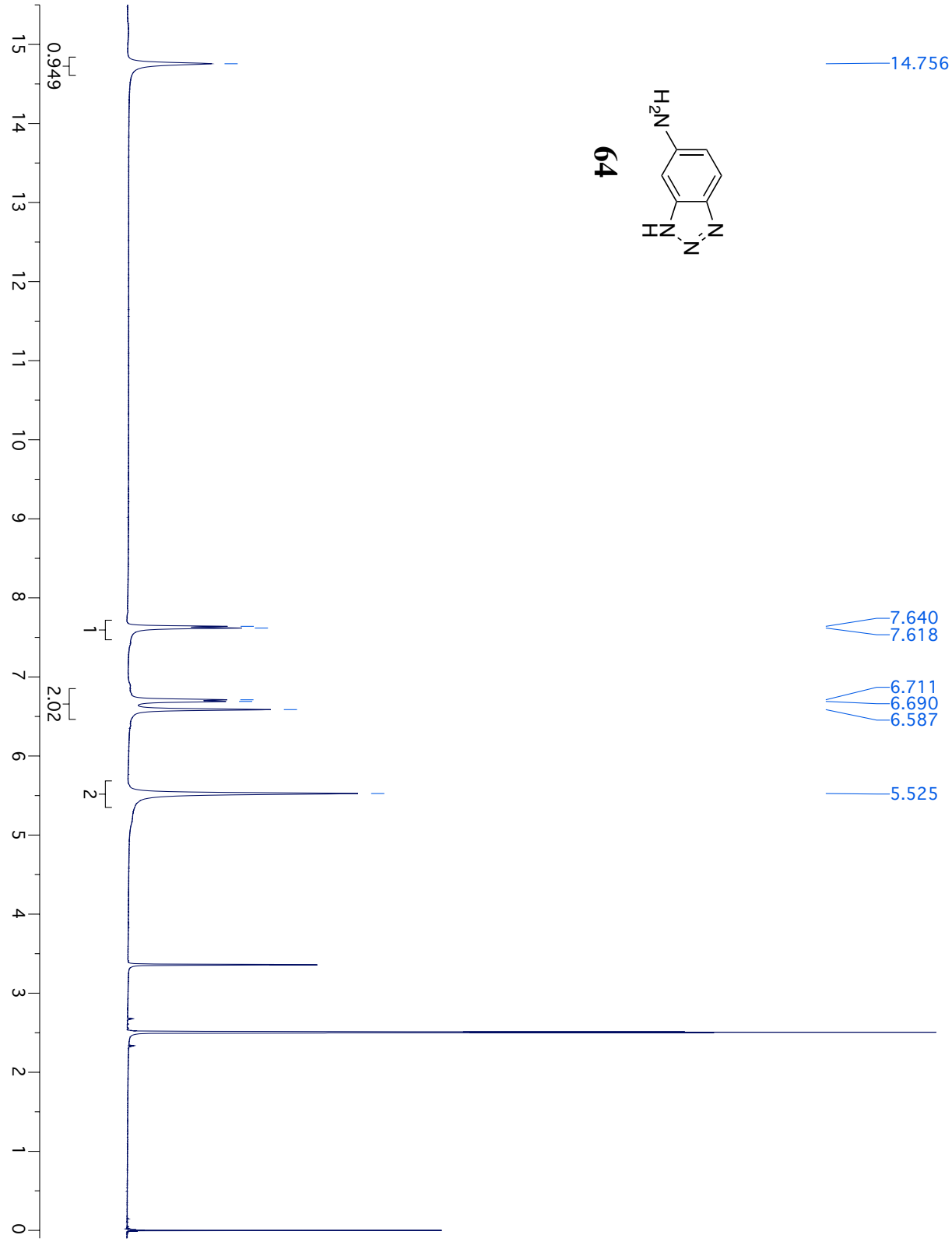


Figure AIV.65: [1]HNMR spectrum of **64a**.

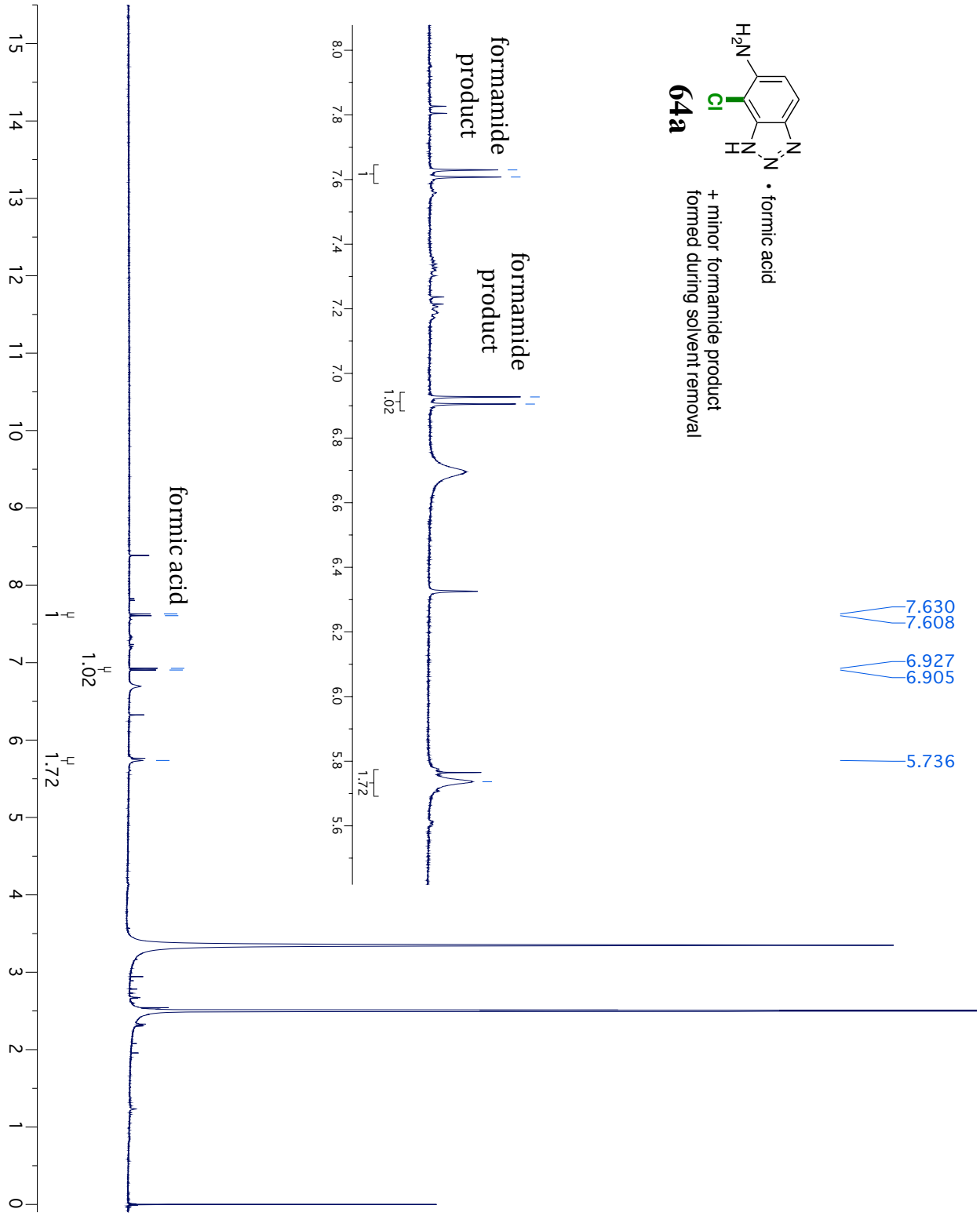


Figure AIV.66: [1]HNMR spectrum of **67**.

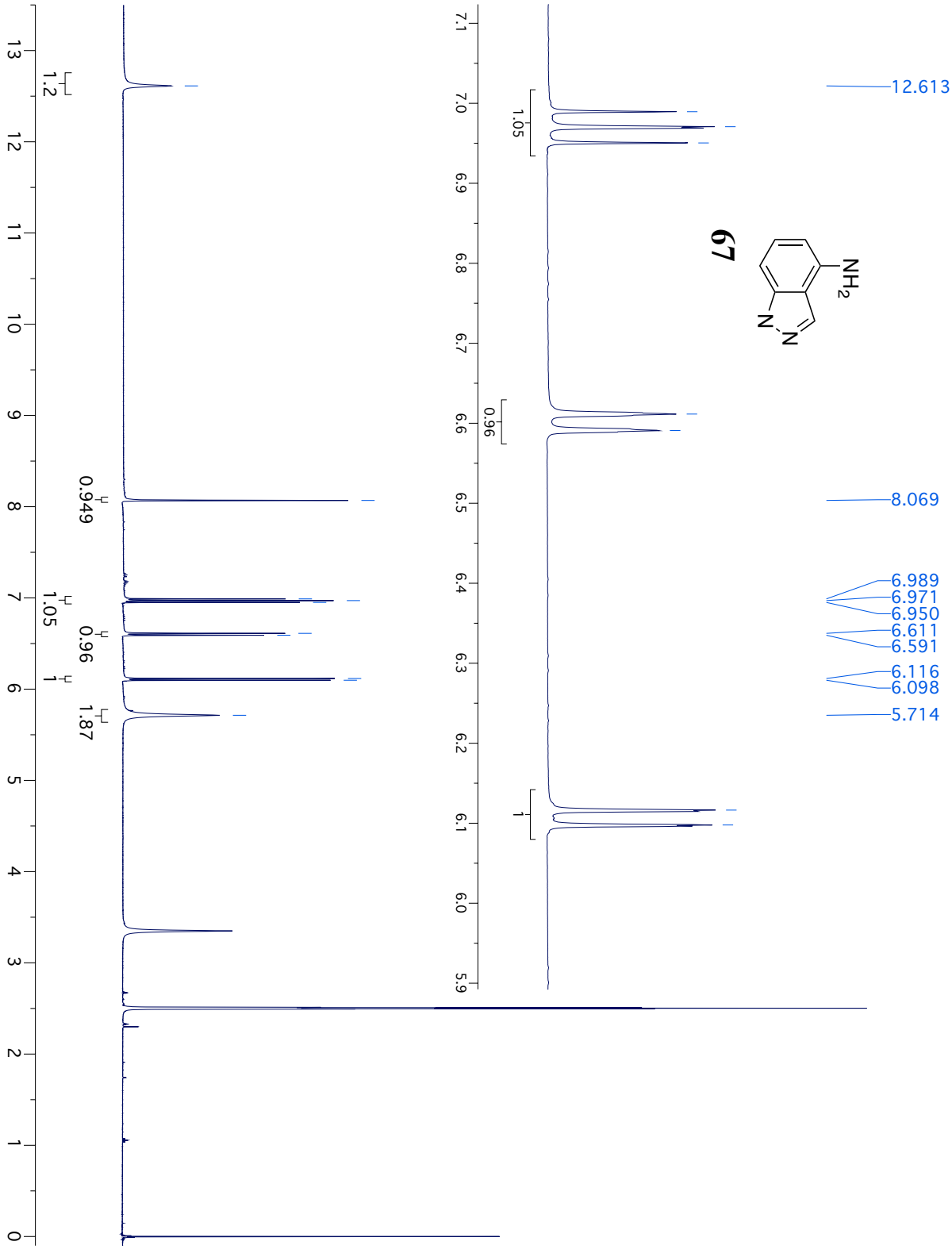


Figure AIV.67: NOSEY spectrum of **67**.

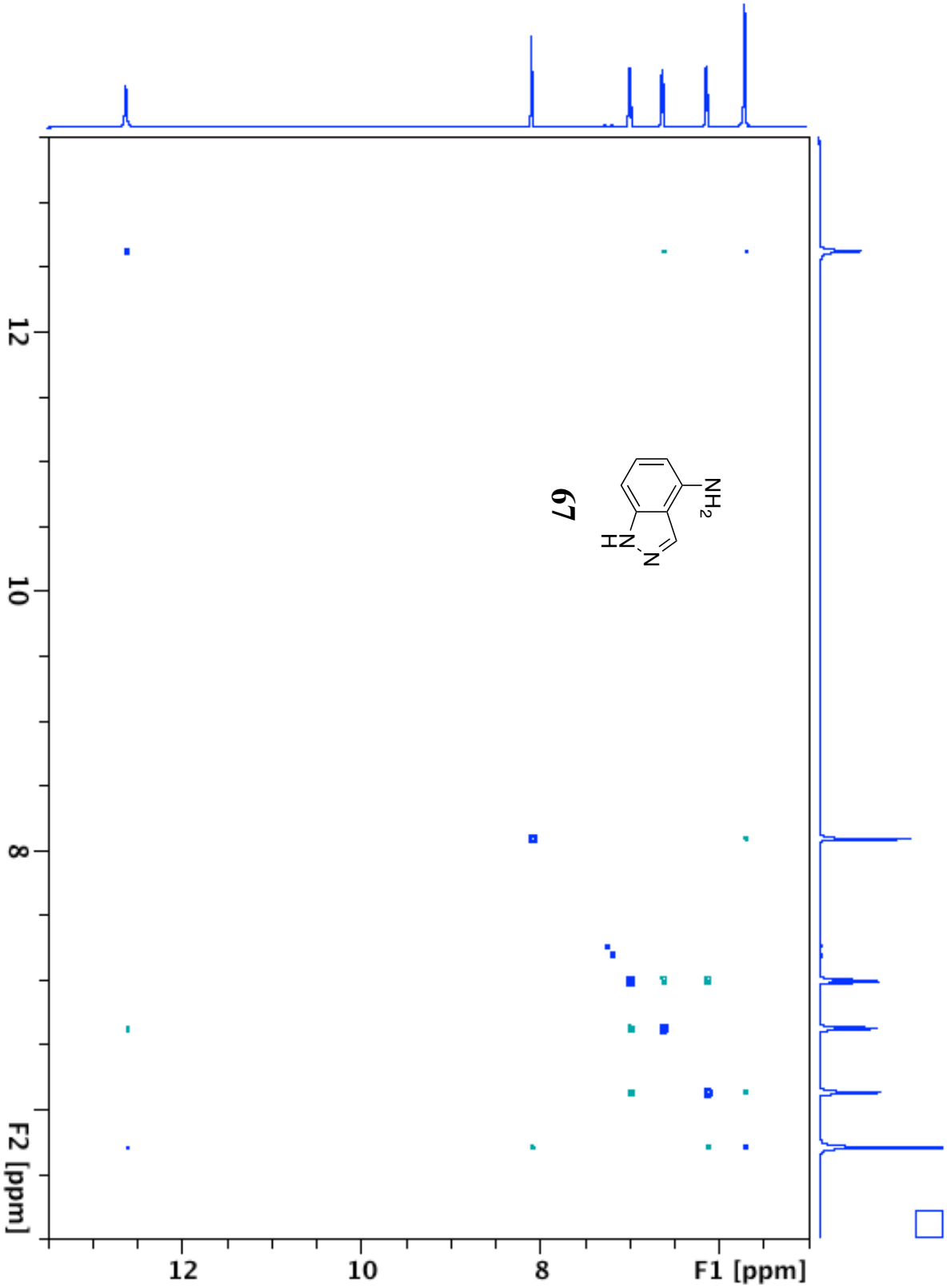


Figure AIV.68: [1]HNMR spectrum of **67a**.

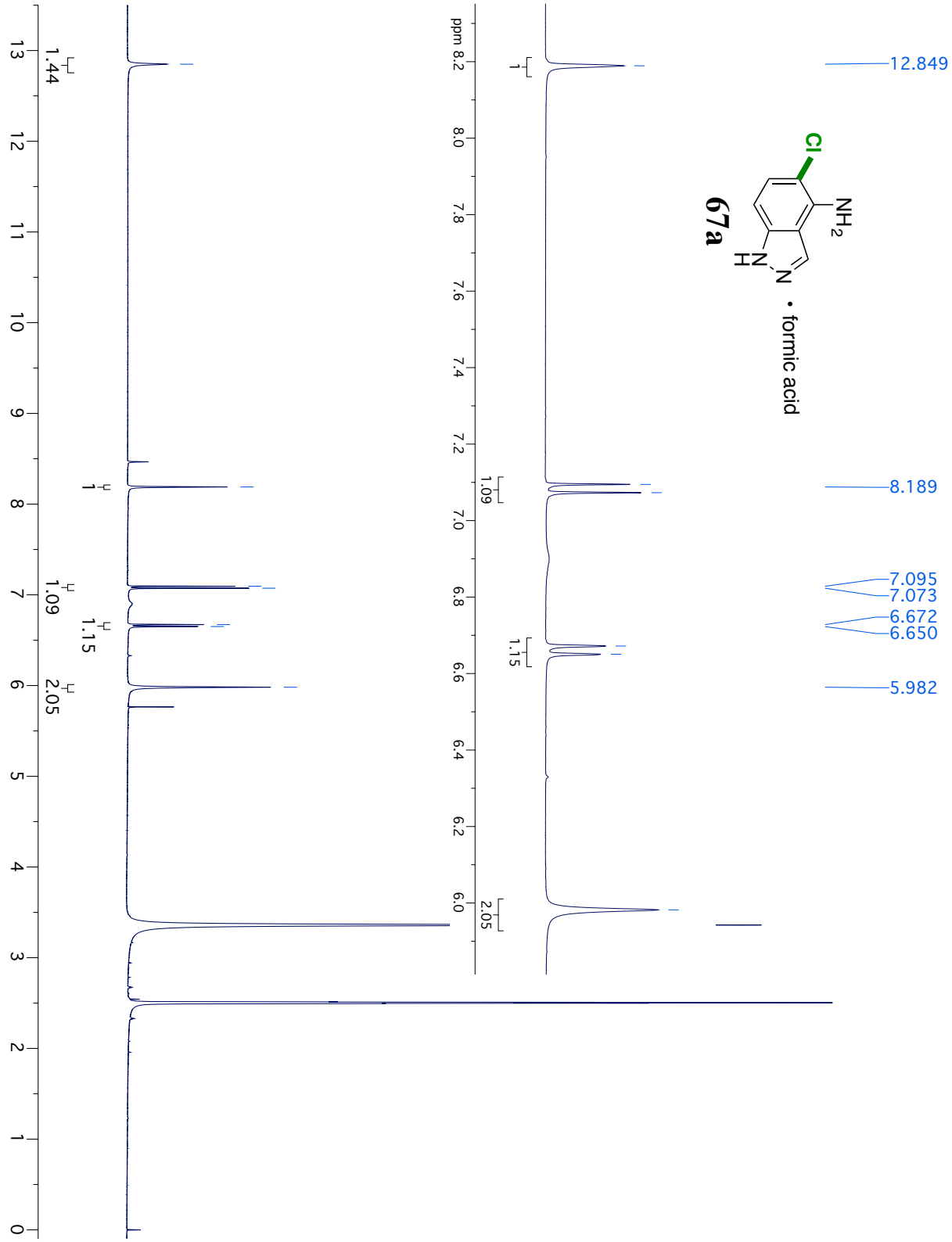


Figure AIV.69: [1]HNMR spectrum of **67b**.

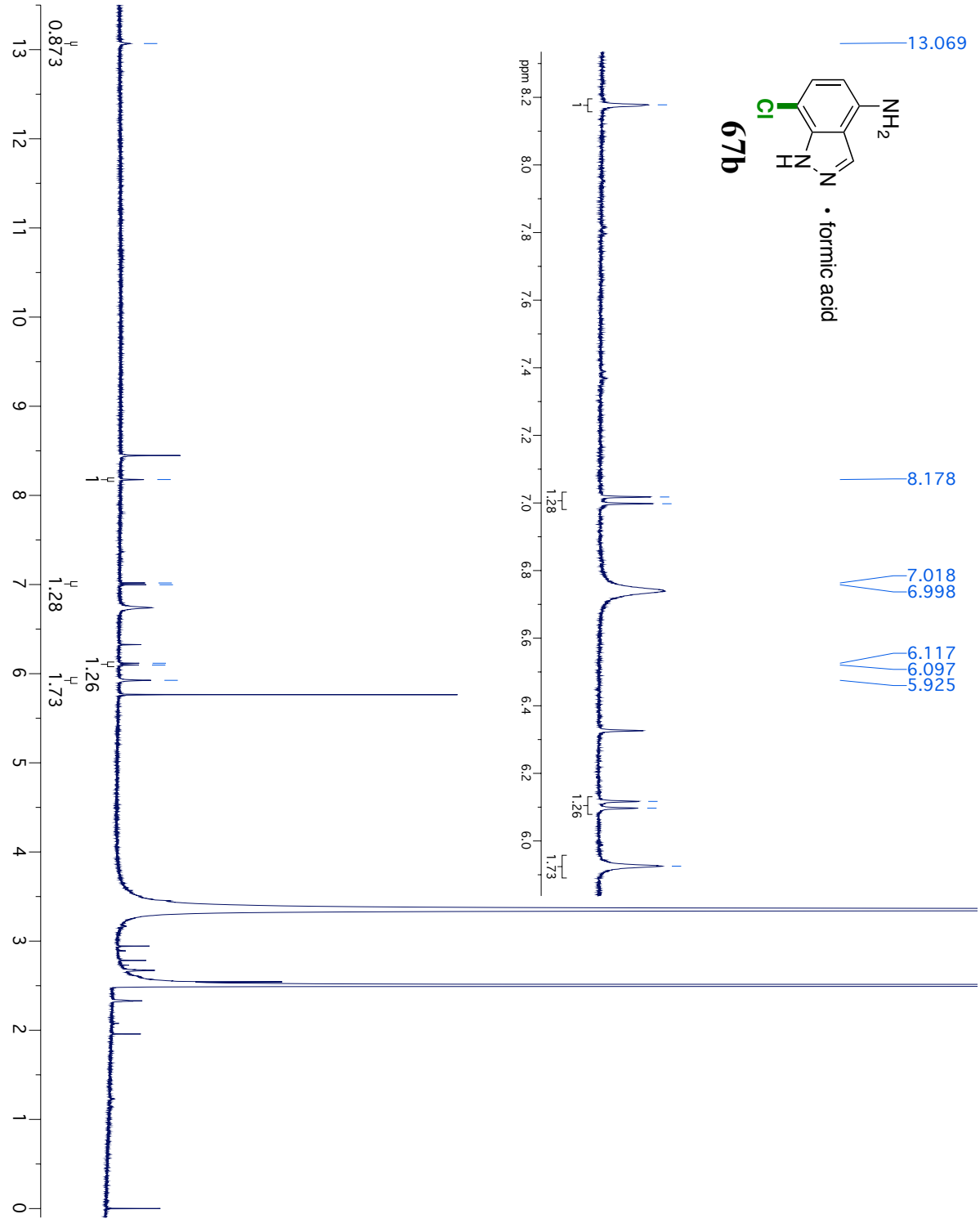


Figure AIV.70: [1]HNMR spectrum of **77**.

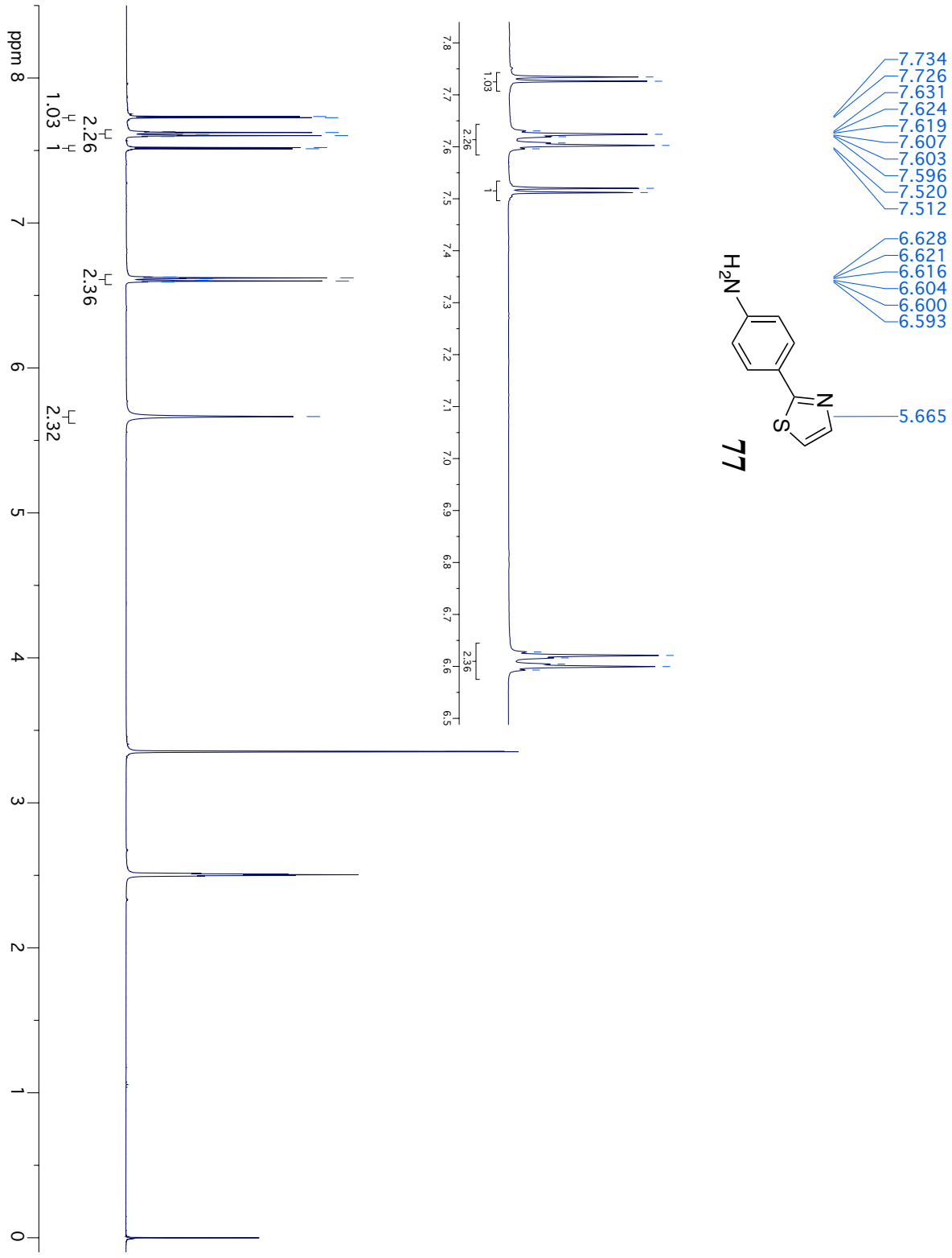


Figure AIV.71: [1]HNMR spectrum of **77a**.

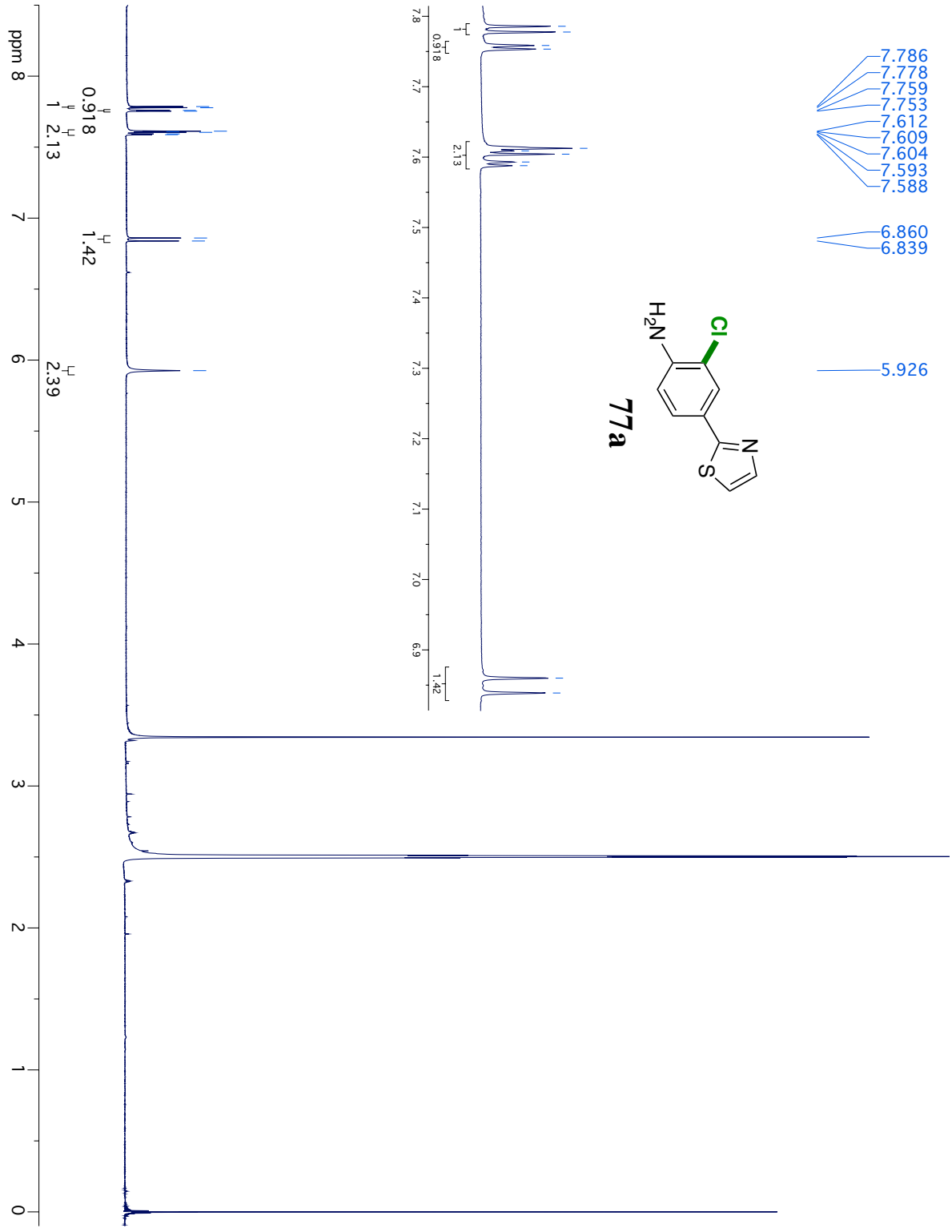


Figure AIV.72: NOSEY spectrum of **77a**.

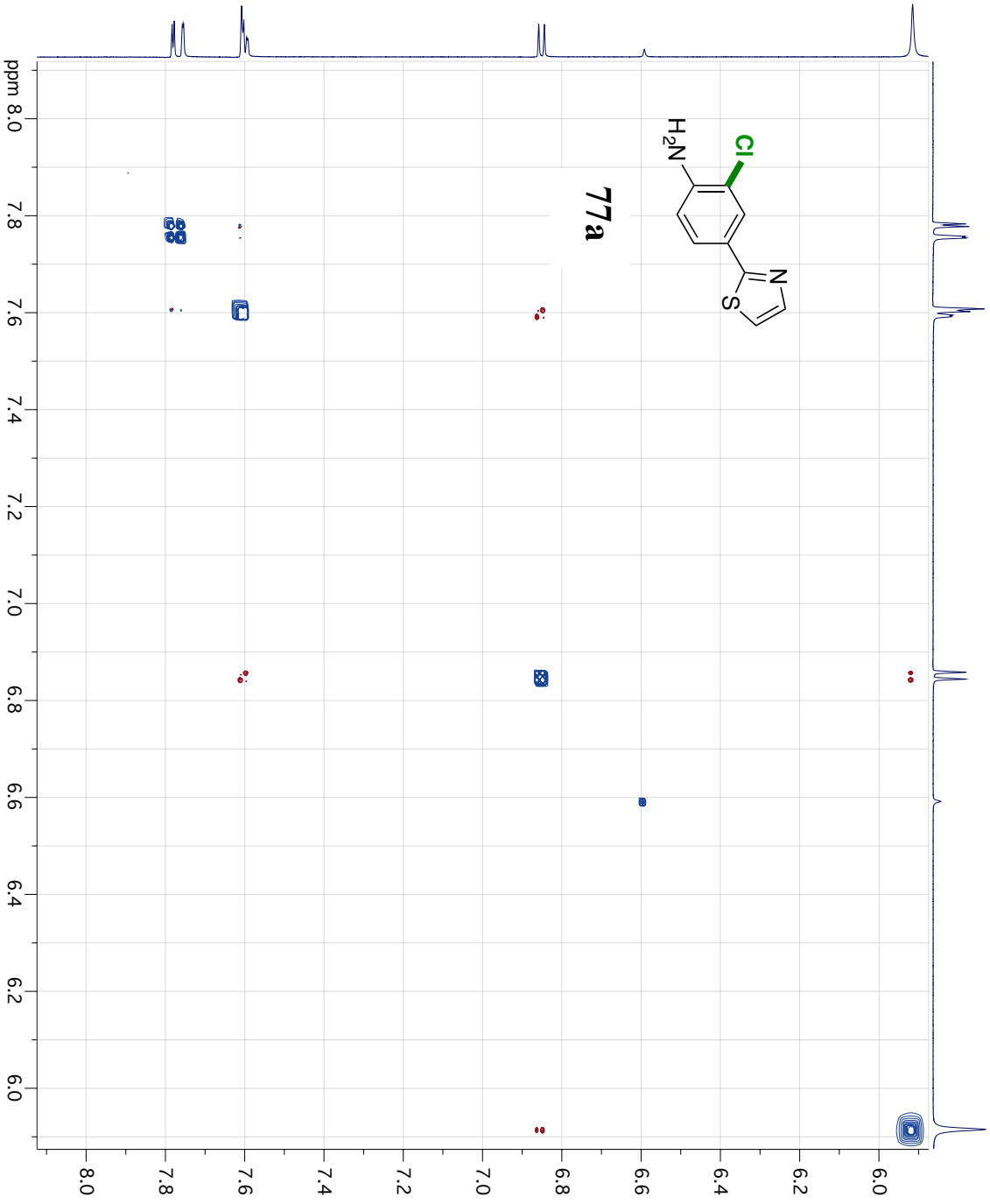


Figure AIV.73: [1]HNMR spectrum of **82**.

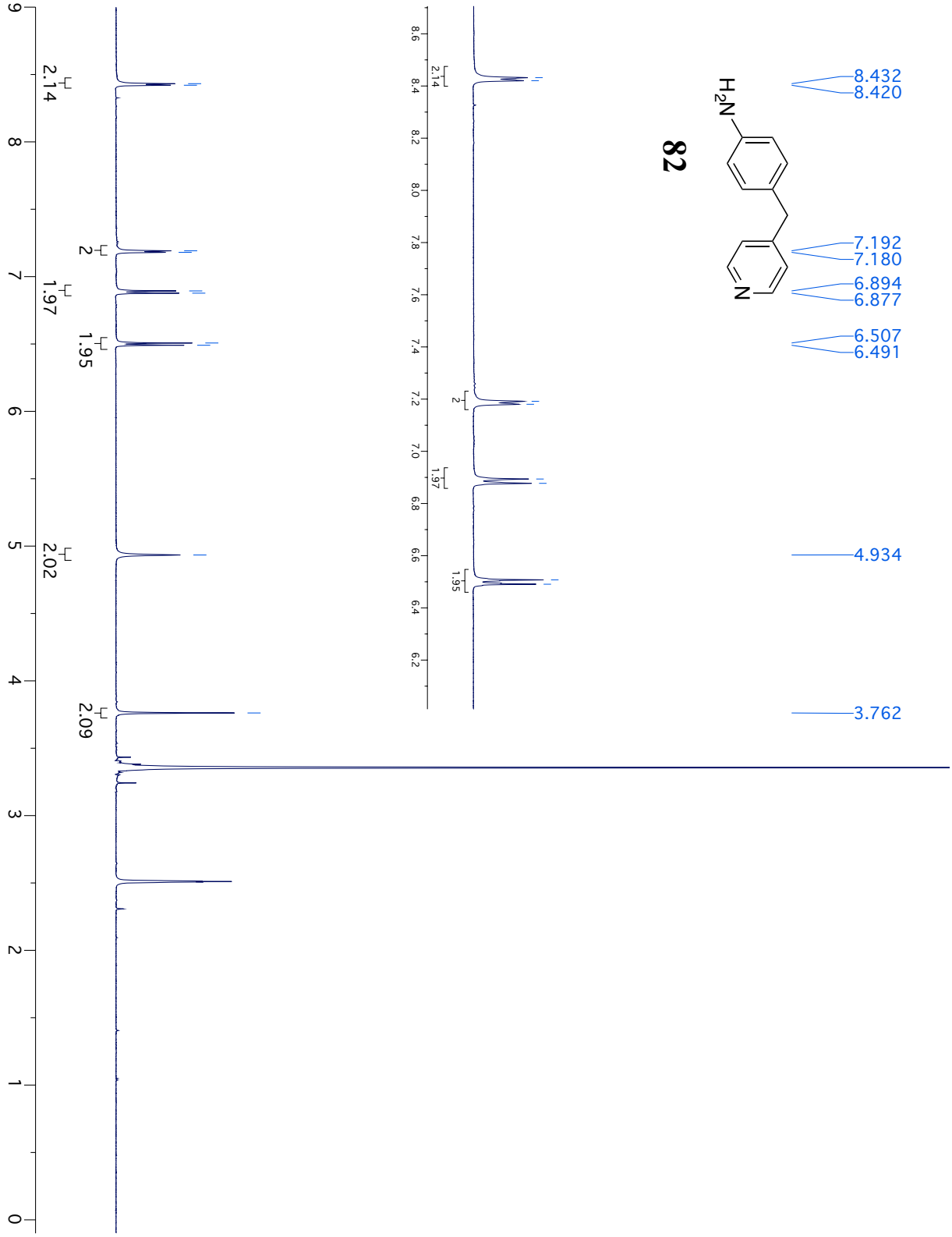


Figure AIV.74: [1]HNMR spectrum of **82a**.

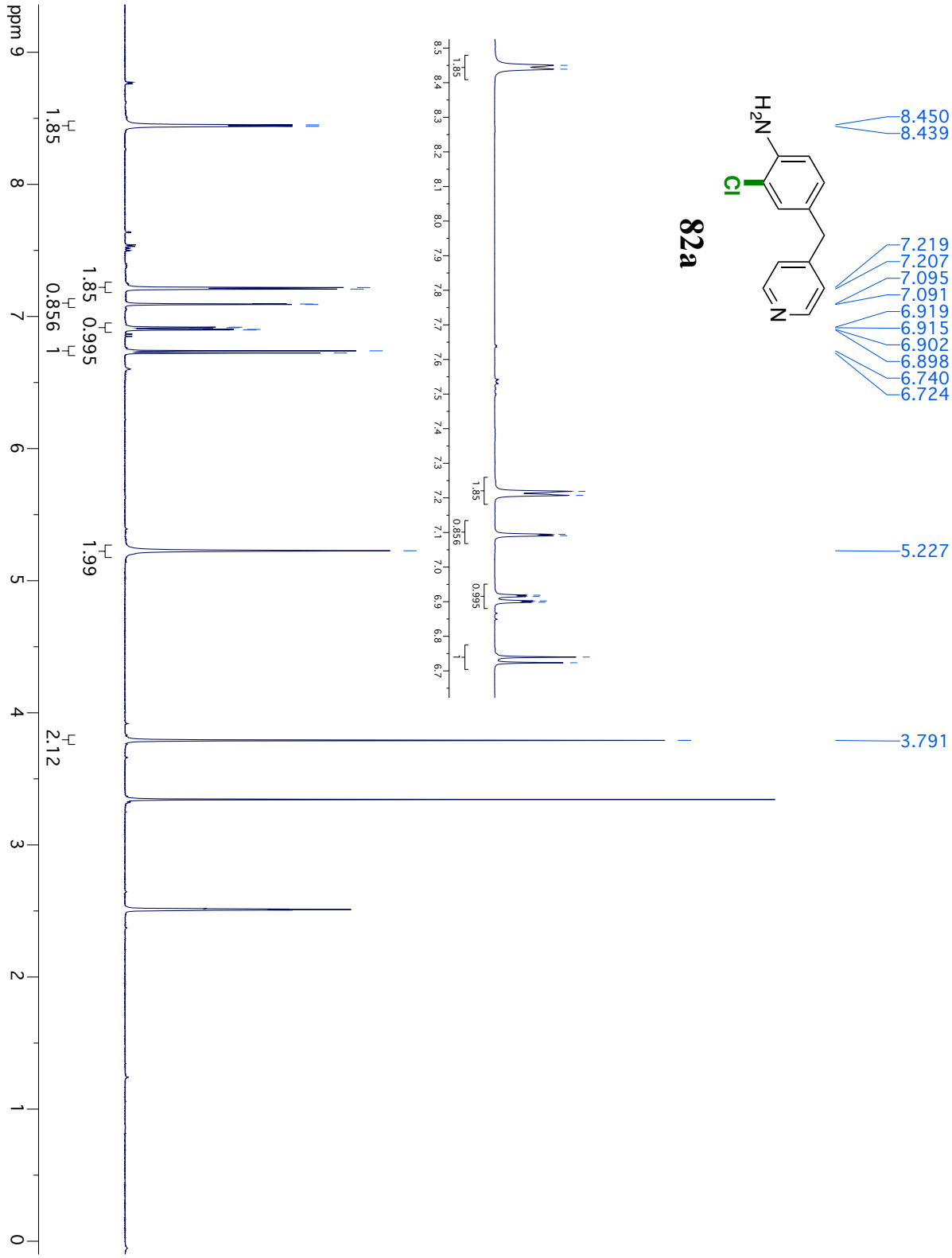


Figure AIV.75: NOSEY spectrum of **82a**.

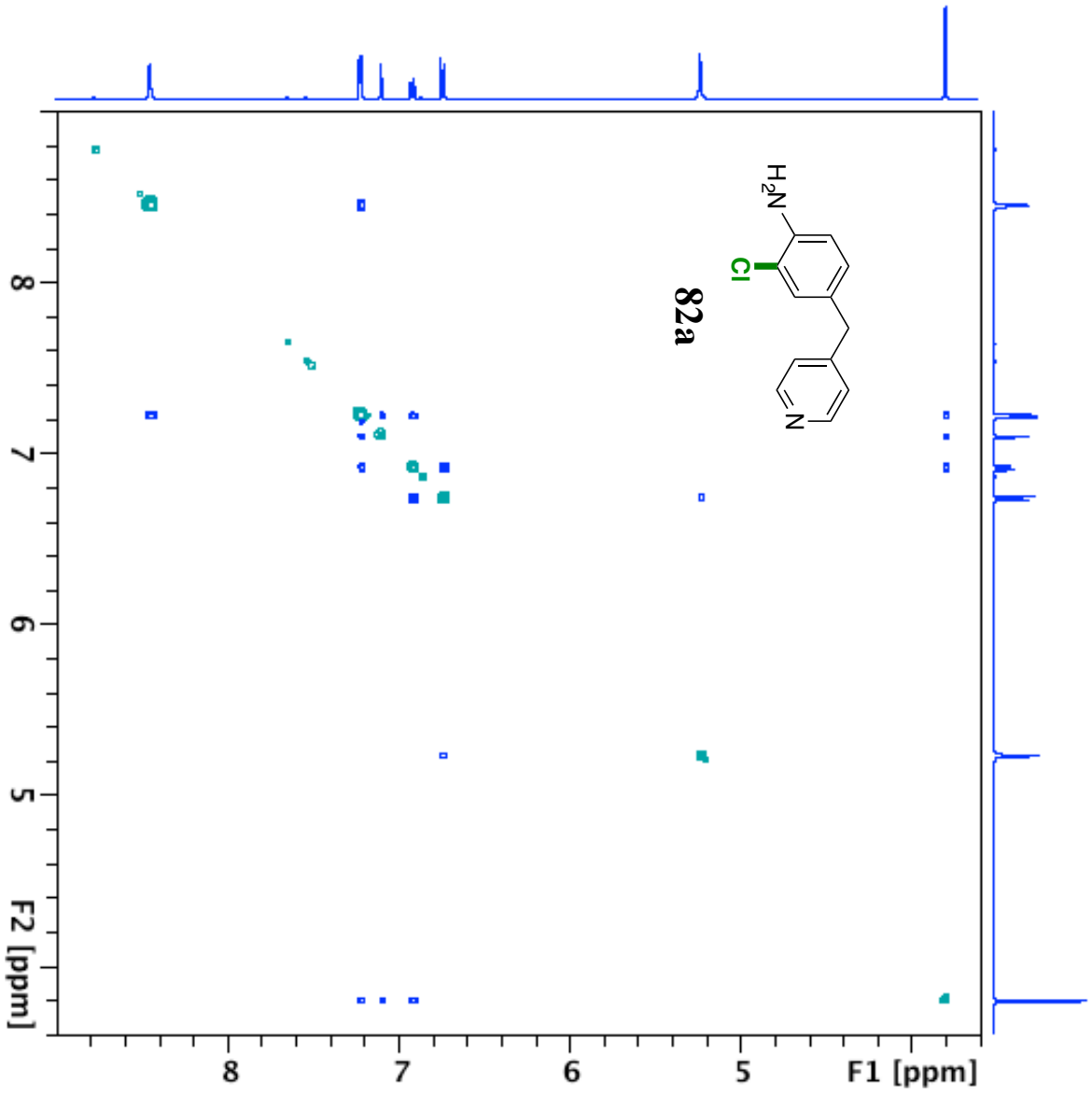


Figure AIV.76: [1]HNMR spectrum of **23a** and **23b**.

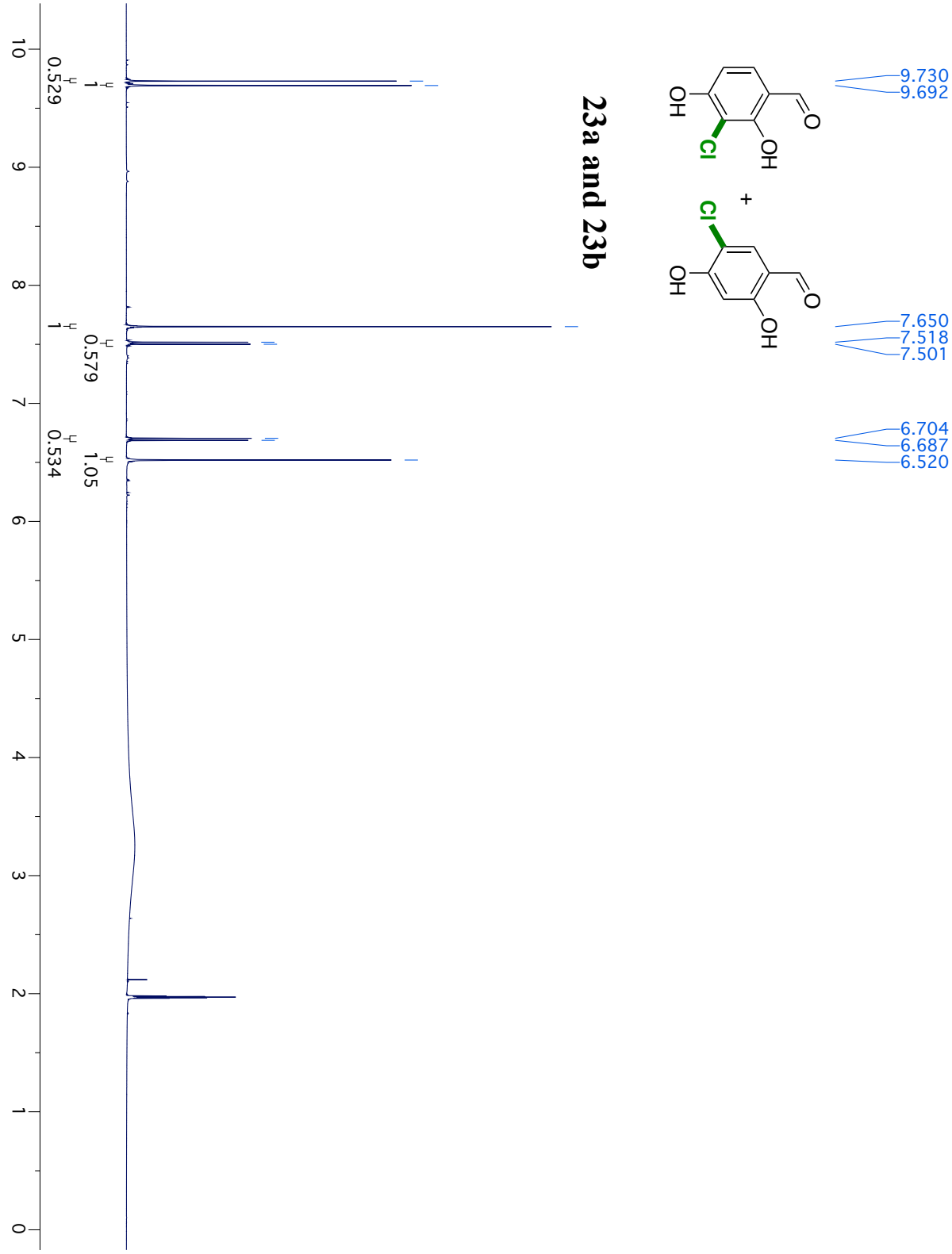


Figure AIV.77: [1]HNMR spectrum of **23c**.

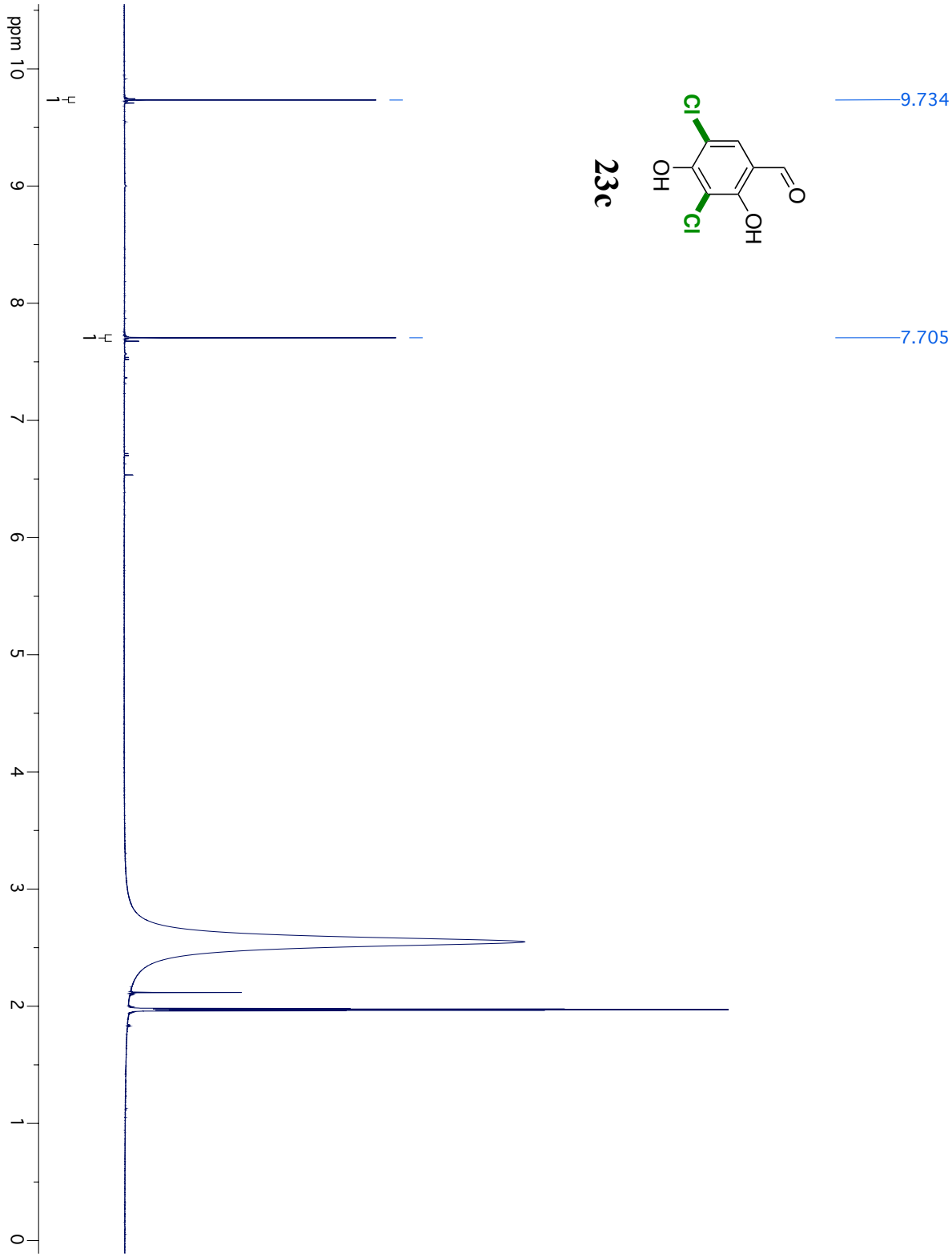


Figure AIV.78: [1]HNMR spectrum of **24a**.

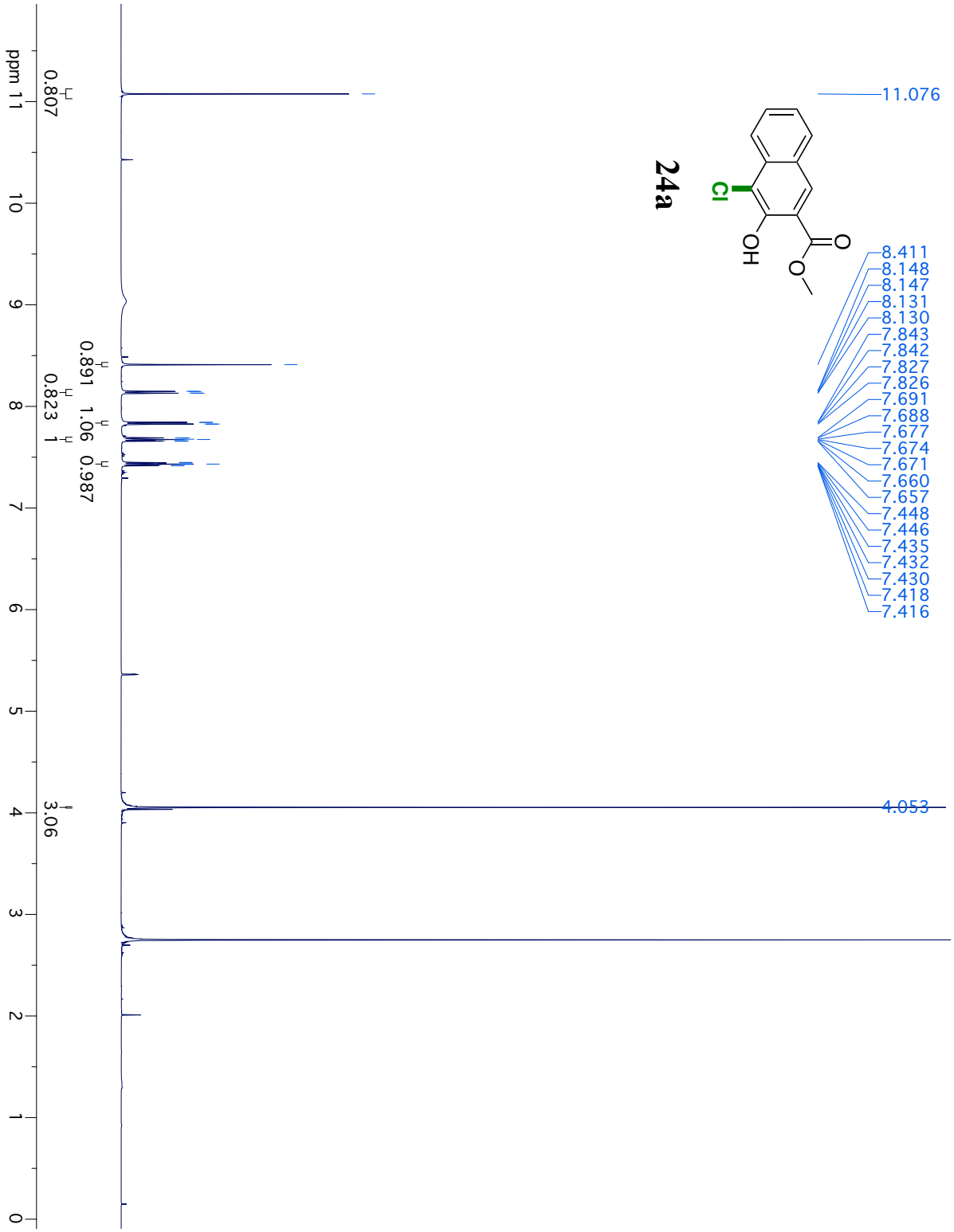


Figure AIV.79: [1]HNMR spectrum of **25a**.

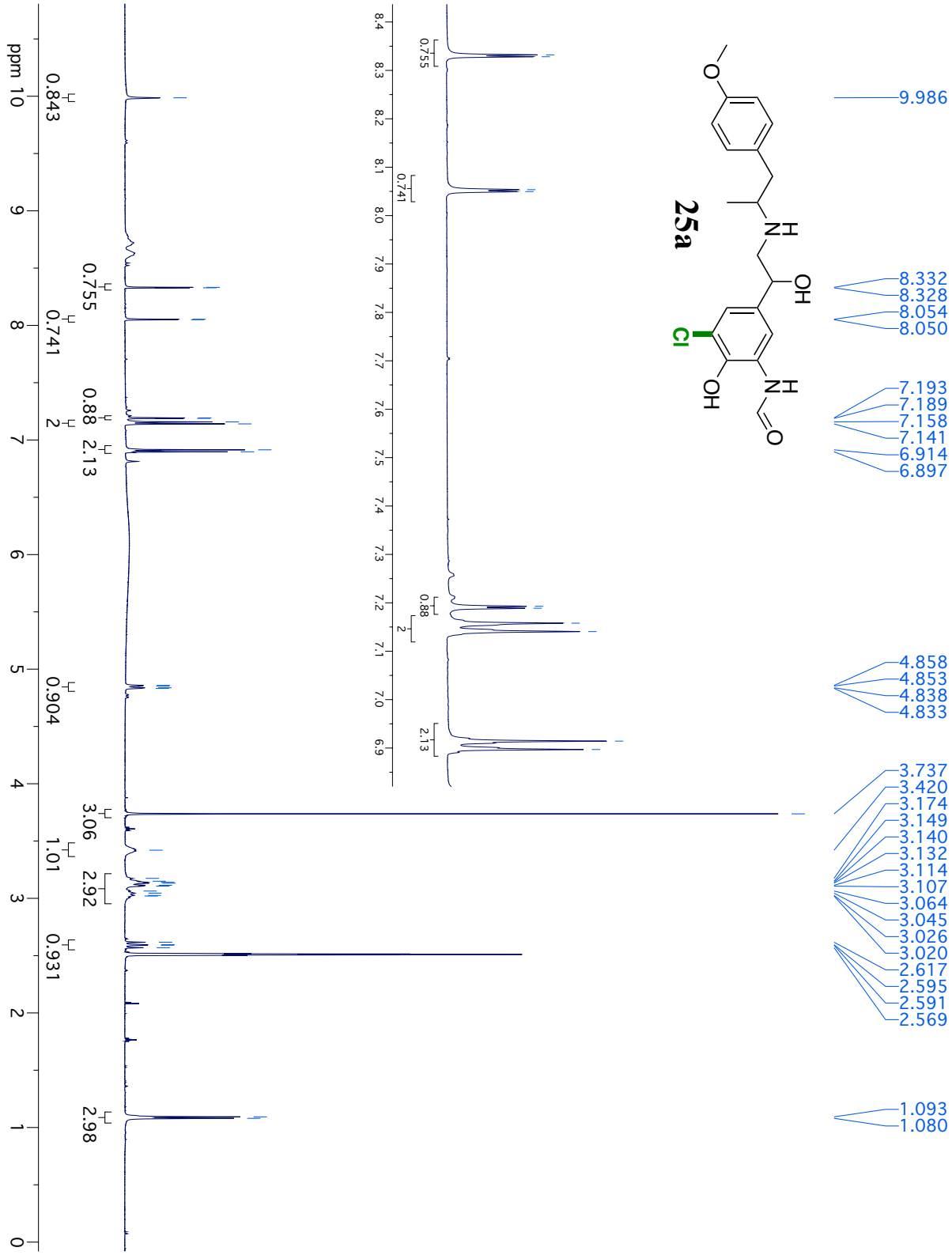


Figure AIV.80: ^{13}CNMR spectrum of **25a**.

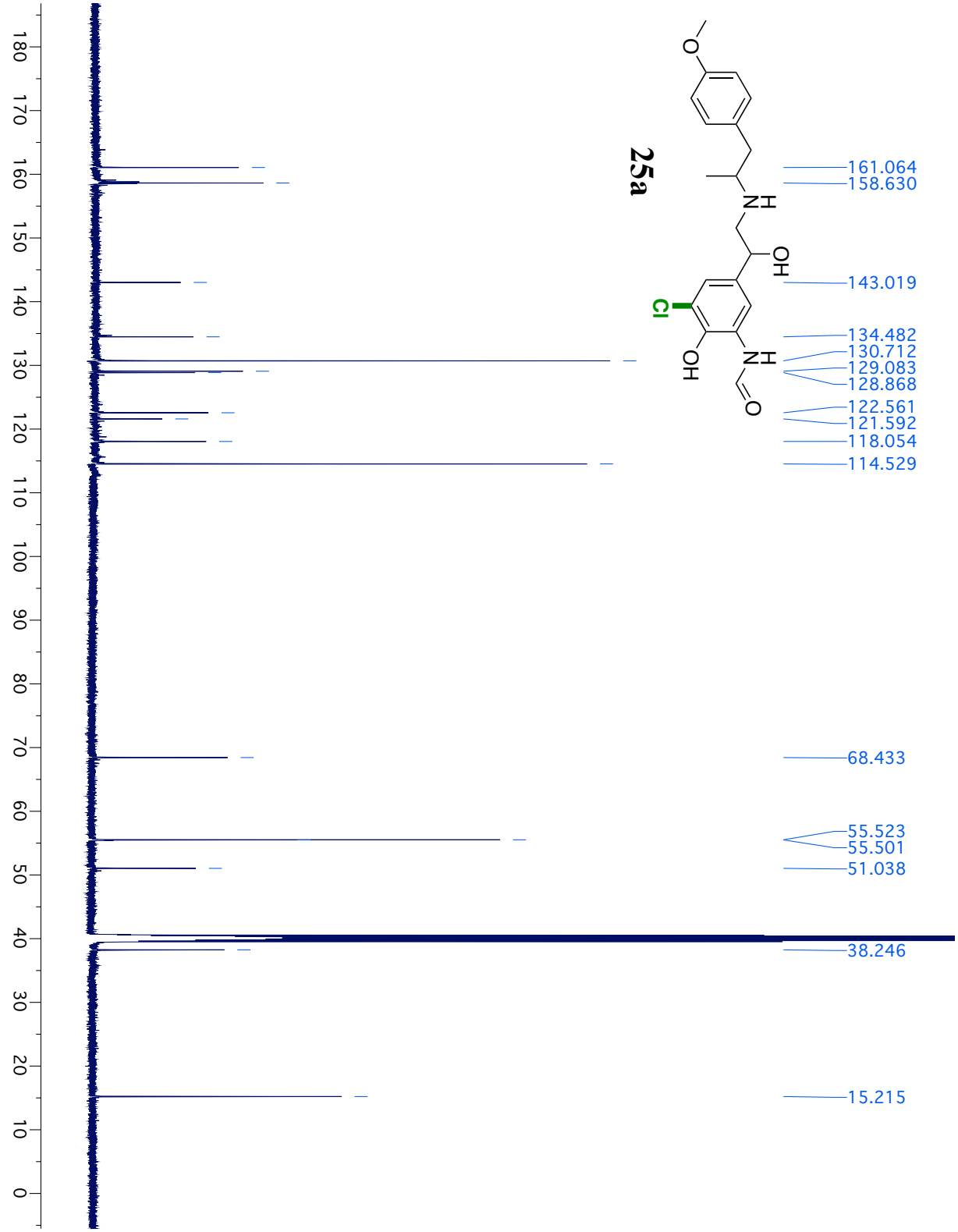


Figure AIV.81: ^{1}H-^{13}C HSQC spectrum of **25a**.

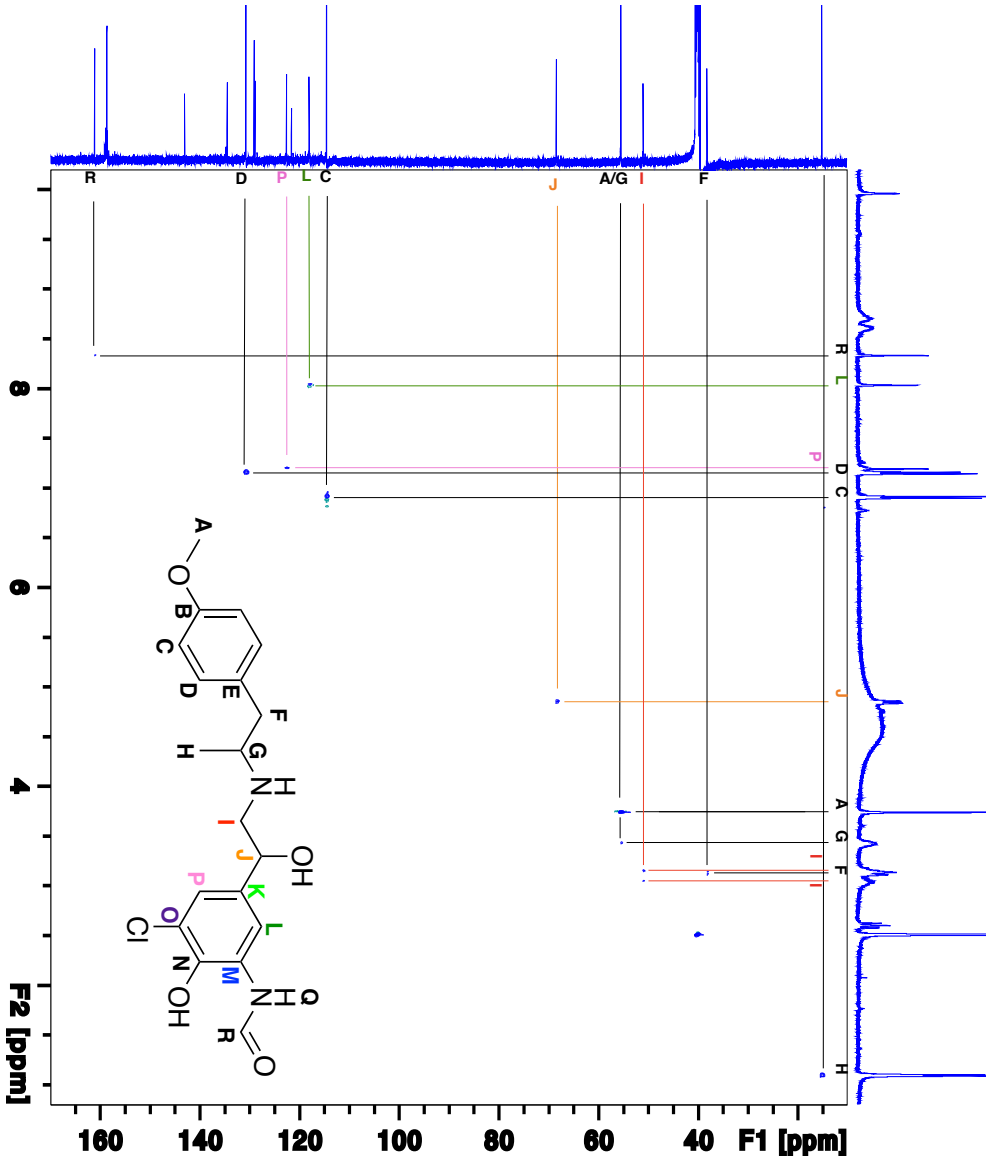


Figure AIV.82: ^{1}H-^{13}C HMBC spectrum of **25a**.

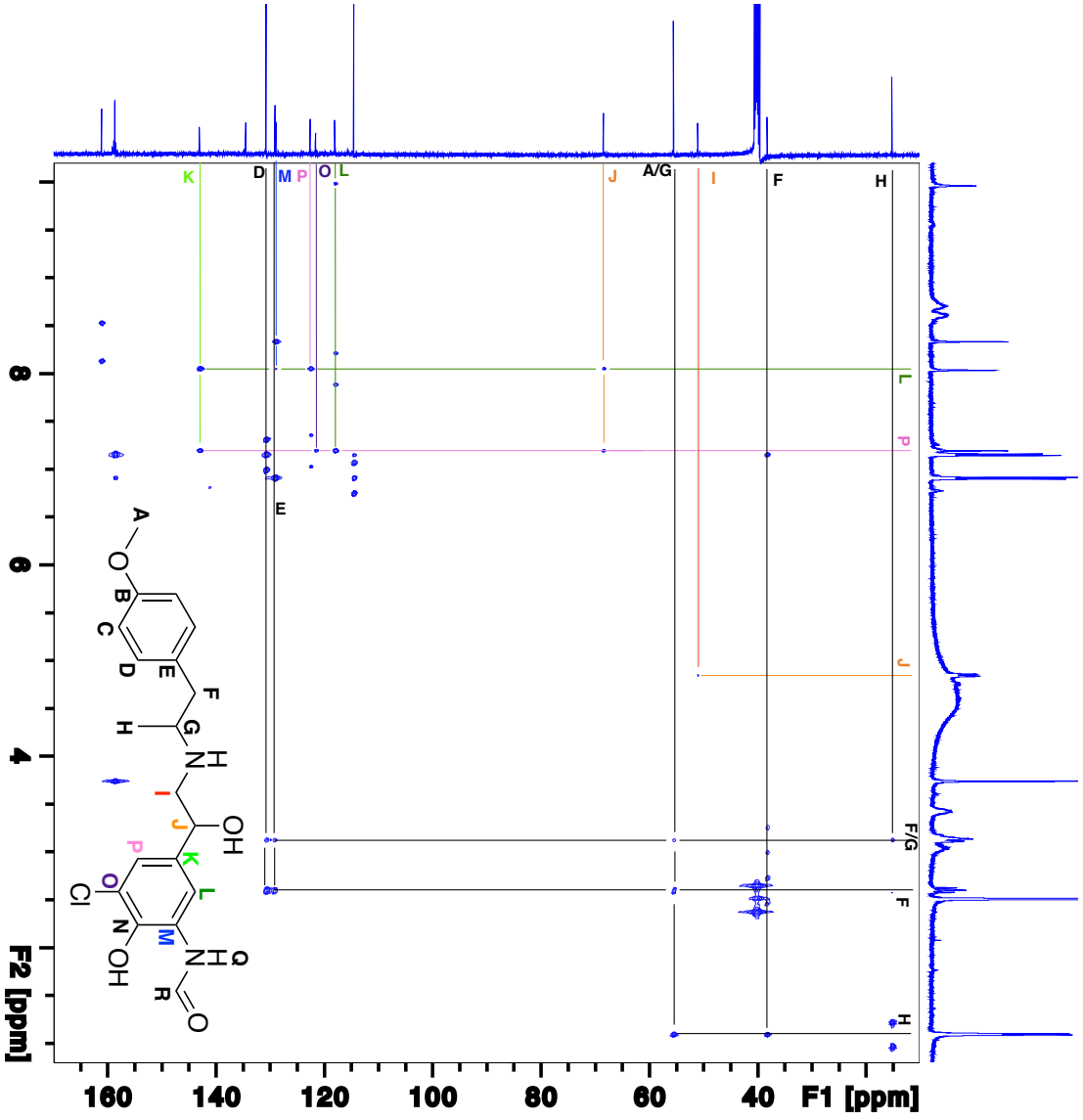


Figure AIV.83: [1]HNMR spectrum of **28a**.