Electronic Supplementary Information

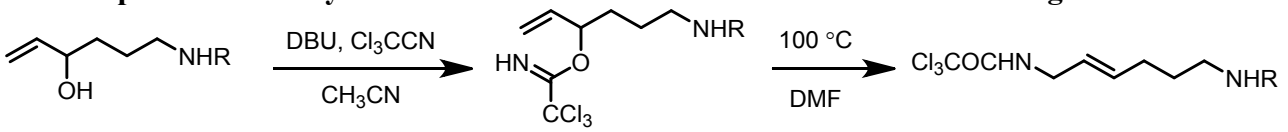
1. General Experimental

All non-aqueous reactions were performed under an atmosphere of N_2 unless otherwise stated. THF, CH_2Cl_2, toluene and MeCN were dried and purified by means of a Pure Solv MD solvent purification system (Innovative Technology Inc.). Anhydrous DMF and 1,4-dioxane were obtained in SureSeal bottles from Sigma-Aldrich. All other solvents used were of chromatography or analytical grade. Commercially available starting materials were obtained from Sigma-Aldrich, Fluka, Acros or Alfa-Aesar and were used without purification unless stated. TLC was carried out on aluminum backed silica gel (Merck silica gel 60 F_{254}) plates supplied by Merck. Flash chromatography was carried out using silica gel 60 (60–63 μm particles) supplied by Merck.

IR spectra were recorded on a PerkinElmer One FT-IR spectrometer. High-resolution mass spectra (HRMS) were recorded on a Bruker MaXis Impact spectrometer with electrospray ionisation (ESI) source. ^1H and ^{13}C NMR spectral data were collected on a Bruker Advance500 or DPX300 spectrometer. Chemical shifts (δ) are quoted in parts per million (ppm) and referenced to the residual solvent peak. Coupling constants are reported as measured within individual NMR signals.

2. Synthetic Details and Experimental Details

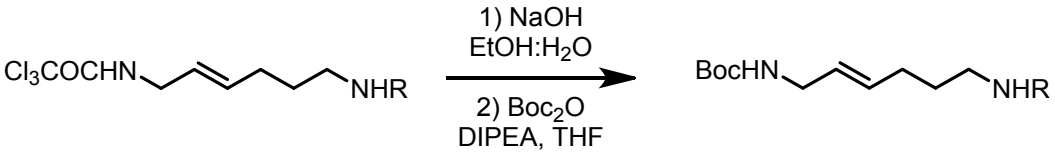
General procedure A: Synthesis trichloroacetimidates and Overman rearrangement.



To a cold solution (0 ºC) of the starting material in anhydrous CH_3CN (5.0 mL/mmol), DBU (0.60 equiv) and trichloroacetonitrile (2.0 equiv) were added. The mixture was allowed to warm up to rt and monitored by TLC. The solvent was evaporated and the crude product was filtered through a small pack of silica gel to remove polar impurities. The crude trichloroacetimidate was used in the next step without further purification.

A solution of the crude trichloroacetimidate in anhydrous DMF (5.0 mL/mmol) was stirred at 100 ºC and monitored by LC-MS until completion. The solvent was evaporated under reduced pressure to give the corresponding trichloroacetamide that was purified by chromatography on silica gel using the appropriate mixture of eluents.

General procedure B: Synthesis of dicarbamates.



To a solution of the trichloroacetamide starting material in a 9:1 mixture of EtOH:H_2O (5.0 mL/mmol), NaOH (0.60 equiv) was added. The mixture was heated at 60 ºC and monitored by TLC and LC-MS. The solvent was evaporated and the crude product was used in the next step without further purification.

To a solution of the crude amine in THF (5.0 mL/mmol), DIPEA (3.0 equiv) and Boc_2O (2.0 equiv) were added. The mixture was stirred at rt and monitored by TLC and LC-MS until completion. The solvent was evaporated under reduced pressure to give the corresponding dicarbamate, which was purified by chromatography on silica gel using the appropriate mixture of eluents.

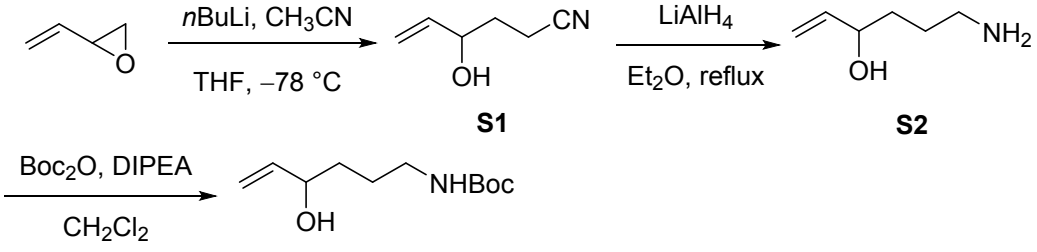
General procedure C: Iodine-mediated cyclization.

To a cold solution (0 ºC) of the starting material in anhydrous CH_3CN (10.0 mL/mmol), $NaHCO_3$ (0.60 equiv) and I_2 (2.0 equiv) were added. The mixture was allowed to warm up to rt and monitored by TLC and LC-MS. The reaction was quenched with saturated solution of $Na_2S_2O_3$ (2 mL/mmol), diluted with EtOAc (2 mL/mmol) and the layers were separated. The aqueous layer was extracted with EtOAc (3 x 5 mL/mmol) and the combined organic extracts were washed with saturated brine (5 mL/mmol), dried (Na_2SO_4) and concentrated under reduced pressure to give the product, that was purified by chromatography on silica gel using the appropriate mixture of eluents.

General procedure D: Alkene aminoarylation.

A mixture of the alkene starting material, $Pd(OAc)_2$ (5 mol %), BINAP (10 mol %), Cs_2CO_3 (2.3 equiv) and aryl halide (1.2 equiv) was dissolved in anhydrous 1,4-dioxane (0.2 M) at rt under N_2. The reaction was stirred at 100 ºC and monitored by LC-MS until completion. The crude product was filtered through a small pack of Cellite. The solvent was evaporated under reduced pressure to give the corresponding pyrrolidine, which was purified by chromatography on silica gel using the appropriate mixture of eluents

***tert*-Butyl (4-hydroxyhex-5-en-1-yl)carbamate, S3.**



To a solution of acetonitrile (3.75 mL, 71.76 mmol) in THF (350 mL) at –78 ºC, *n*-BuLi (2.5 M, 32 mL, 80 mmol) was added and then stirred for 1 hr. Butadienemonoxide (5.03 g, 71.76 mmol) was added dropwise and the reaction was stirred at –78 ºC for 1 hour and then at 0 ºC for 2 hours. The reaction mixture was quenched with a saturated solution of NH_4Cl (150 mL), diluted with EtOAc (150 mL), and the layers were separated. The aqueous layer was extracted with EtOAc (3 x 150 mL). The combined organic extracts were washed with a saturated aqueous solution of NaCl (150 mL), dried over Na_2SO_4 and concentrated under reduced pressure to give the crude alcohol **S1** which was used in the next step without further purification.

To a solution of crude **S1** in Et_2O (300 mL) at 0 ºC was carefully added $LiAlH_4$ (4.1 g, 107.6 mmol) and the resulting suspension was then stirred at reflux for 1 hour. The reaction mixture was quenched with 2.0 M solution of NaOH (70 mL) and the solid was filtered. The filtrate was concentrated under reduced pressure, to give the crude amine S**2** (3.07 g) which was used in the next step without further purification.

To a solution of crude **S2** (3.07 g) in CH_2Cl_2 (130 mL) at rt was added DIPEA (9.40 mL, 53.4 mmol) and Boc_2O (6.4 g, 29.4 mmol). The reaction was stirred overnight until complete as monitored by TLC. The solvent was evaporated and the crude residue was submitted to chromatographic purification (gradient elution: 10:90 → 50:50 Et_2O – CH_2Cl_2) to give the carbamate **S3** (4.2 g, 74%, over 3 steps), as a colourless oil.

R_f 0.20 (30% Et_2O – CH_2Cl_2);

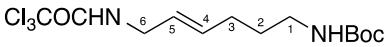
δ_H (500 MHz, $CDCl_3$, 27 °C) 5.87 (1 H, ddd, *J* 16.8 and 10.4 and 6.2 Hz, 5-H), 5.23 (1 H, d, *J* 17.2 Hz, 6-H_{trans}), 5.11 (1 H, t, *J* 10.4 Hz, 6-H_{cis}), 4.71 (1 H, s, NH), 4.13 (1 H, q, *J* 5.2 Hz, 4-H), 3.16 (2 H, m, 1-H_2), 2.16 (2 H, m, 3-H_2), 1.51-1.60 (2 H, m, 2-H_2), 1.47 (9 H, s, *t*-Bu);

δ_C (125 MHz, $CDCl_3$, 27 °C) 156.5 (C=O), 140.9 (C-5), 136.6 (C Ar), 128.5 (2 x CH Ar), 128.1 (3 x CH Ar), 114.9 (C-6), 72.8 (C-4), 66.7 (CH_2 Bn), 40.9 (C-1), 33.8 and 25.9 (C-2 and C-3);

IR (film): ν_{max} 3320, 1695, 1540, 1455, 1251, 1139, 922, 750, 698 cm^{-1};

HRMS (ES) m/z calcd for $C_{11}H_{21}NNaO_3$ $[M+Na]^+$ 238.141364, found 238.141832.

***tert*-Butyl [(4*E*)-6-(2,2,2-trichloroacetamido)hex-4-en-1-yl]carbamate, 1a.**



As general procedure A, using alcohol **S3**. After chromatographic purification (gradient elution: 20:80 → 80:20 EtOAc – hexane) alkene **1a** (2.03 g, 58%, over 2 steps) was obtained as a colourless oil.

R_f 0.40 (30% EtOAc – hexane).

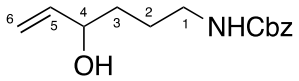
δ_H (500 MHz, $CDCl_3$, 27 °C) 6.85 (1 H, s, $NHCOCCl_3$), 5.70 (1 H, dt, *J* 15.0 and 6.7 Hz, 4-H), 5.51 (1 H, dt, *J* 15.3 and 6.3 Hz, 5-H), 4.55 (1 H, s, NHBoc), 3.92 (2 H, t, *J* 5.7 Hz, 6-H_2), 3.12 (2 H, q, *J* 6.3 Hz, 1-H_2), 2.10 (2 H, c, *J* 7.1 Hz, 3-H_2), 1.58 (2 H, quint, *J* 7.2 Hz, 2-H_2), 1.44 (9 H, s, *t*-Bu).

δ_C (125 MHz, $CDCl_3$, 27 °C) 163.8 (C=O), 161.8 (C=O), 134.3 (C-4), 124.6 (C-5), 91.9 (C *t*-Bu), 79.3 (CCl_3), 43.3 (C-6), 39.8 (C-1), 29.31 (C-3), 29.26 (C-2), 28.4 (*t*-Bu).

IR (film): ν_{max} 3325, 2936, 2874, 1695, 1540, 1518, 1455, 1251, 1137, 970, 822, 739, 697 cm^{-1}.

HRMS (ES) m/z calcd for $C_{13}H_{22}Cl_3N_2O_3$ $[M+H]^+$ 359.069052, found 359.068670.

Benzyl (4-hydroxyhex-5-en-1-yl)carbamate, S4.



To a solution of crude amine **S2** (0.95 g) in CH_2Cl_2 (40 mL) at rt was added DIPEA (2.9 mL, 16.6 mmol) and Cbz-Cl (2.36 mL, 16.6 mmol). The reaction was stirred overnight until complete as monitored by TLC. The solvent was evaporated and the crude residue was submitted to chromatographic purification (gradient elution: 10:90 → 30:70 Et_2O – CH_2Cl_2) to give the carbamate **S4** (1.8 g, 75%, over 3 steps), as colourless oil.

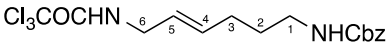
R_f 0.25 (20% Et_2O – CH_2Cl_2).

δ_H (500 MHz, $CDCl_3$, 27 °C) 7.32-7.47 (5 H, m, 5 x CH Ar), 5.88 (1 H, ddd, *J* 16.8 and 10.4 and 6.3 Hz, 5-H), 5.25 (1 H, d, *J* 17.2 Hz, 6-H_{trans}), 5.14 (1 H, t, *J* 10.7 Hz, 6-H_{cis}), 5.12 (2 H, s, CH_2 Bn), 4.91 (1 H, s, NH), 4.16 (1 H, br m, 4-H), 3.13-3.35 (2 H, m, 1-H_2), 1.77 (1 H, s, OH), 1.51-1.67 (4 H, m, 2-H_2 and 3-H_2).

δ_C (125 MHz, $CDCl_3$, 27 °C) 156.5 (C=O), 140.9 (C-5), 136.6 (C Ar), 128.5 (2 x CH Ar), 128.1 (3 x CH Ar), 114.9 (C-6), 72.8 (C-4), 66.7 (CH_2 Bn), 40.9 (C-1), 33.8 and 25.9 (C-2 and C-3).

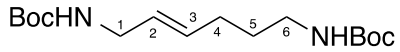
IR (film): ν_{max} 3320, 2942, 2874, 1699, 1540, 1455, 1258, 1139, 1026, 991, 922, 750, 698 cm^{-1}.
HRMS (ES) m/z calcd for $C_{14}H_{19}NNaO_3$ $[M+Na]^+$ 272.125714, found 272.126081.

Benzyl [(4*E*)-6-(2,2,2-trichloroacetamido)hex-4-en-1-yl]carbamate, S5.



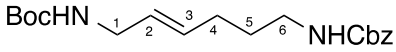
As general procedure A using alcohol **S4** (1.0 g, 4.01 mmol). After chromatographic purification (gradient elution: 20:80 → 80:20 EtOAc – hexane) **S5** (0.79 g, 51%, over 2 steps) was obtained as a colourless oil.
R_f 0.35 (50% EtOAc – hexane).
δ_H (500 MHz, $CDCl_3$, 27 °C) 7.31-7.48 (5 H, m, 5 x CH Ar), 6.86 (1 H, s, $NHCOCCl_3$), 5.72 (1 H, dt, *J* 15.2 and 7.1 Hz, 4-H), 5.51 (1 H, dt, *J* 15.3 and 6.5 Hz, 5-H), 5.12 (2 H, s, CH_2 Bn), 4.82 (1 H, s, NHCbz), 3.94 (2 H, t, *J* 5.6 Hz, 6-H_2), 3.23 (2 H, c, *J* 6.7 Hz, 1-H_2), 2.13 (2 H, c, *J* 6.9 Hz, 3-H_2), 1.63 (2 H, q, *J* 7.1 Hz, 2-H_2).
δ_C (125 MHz, $CDCl_3$, 27 °C) 161.7 ($COCCl_3$), 156.4 (C=O), 136.6 (C-4), 134.1 (C Ar), 128.6 (2 x CH Ar), 128.2 (2 x CH Ar), 128.1 (CH Ar), 124.8 (C-5), 66.7 (CH_2 Bn), 43.3 (C-6), 40.3 (C-1), 29.3 (C-3), 29.1 (C-2).
IR (film): ν_{max} 3329, 2935, 1695, 1518, 1455, 1252, 1138, 970, 823, 739, 697 cm^{-1}.
HRMS (ES) m/z calcd for $C_{16}H_{19}Cl_3N_2NaO_3$ $[M+Na]^+$ 415.035346, found 415.035999.

Di-*tert*-butyl [(2*E*)-hex-2-ene-1,6-diyl]dicarbamate, 1b.



As general procedure B using trichloracetamide **1a** (970 mg, 2.70 mmol). After chromatographic purification (gradient elution: 20:80 → 80:20 EtOAc – hexane) the dicarbamate **1b** (600 mg, 71%, over 2 steps) was obtained as a colourless oil.
R_f 0.30 (30% EtOAc – Hex).
δ_H (500 MHz, $CDCl_3$, 27 °C) 5.58 (1 H, dt, *J* 15.0 and 6.7 Hz, 3-H), 5.47 (1 H, dt, *J* 15.3 and 5.7 Hz, 2-H), 4.60 (2 H, s, NH), 3.68 (2 H, br m, 1-H_2), 3.11 (2 H, br m, 6-H_2), 2.06 (2 H, c, *J* 6.9 Hz, 4-H_2), 1.56 (2 H, q, *J* 7.1 Hz, 5-H_2), 1.45 (9 H, s, *t*-Bu).
δ_C (125 MHz, $CDCl_3$, 27 °C) 156.0 (C=O), 155.8 (C=O), 131.8 (C-3), 127.3 (C-2), 79.2 (C *t*-Bu), 79.1 (C *t*-Bu), 42.5 (C-1), 40.0 (C-6), 29.4 (C-4 and C-5), 28.4 (*t*-Bu).
IR (film): ν_{max} 3341, 2977, 2934, 1684, 1518, 1392, 1249, 1166, 1038, 918, 828, 679 cm^{-1}.
HRMS (ES) m/z calcd for $C_{16}H_{31}N_2O_4$ $[M+H]^+$ 315.227834, found 315.227599.

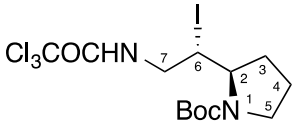
Benzyl *tert*-butyl [(2*E*)-hex-2-ene-1,6-diyl]dicarbamate, 1c.



As general procedure B, using trichloracetamide **S5** (310 mg, 0.787 mmol). After chromatographic purification (gradient elution: 20:80 → 80:20 EtOAc – hexane) the dicarbamate **1c** (170 mg, 62%, over 2 steps) was obtained as a colourless oil.
R_f 0.30 (40% EtOAc – Hex).
δ_H (500 MHz, $CDCl_3$, 27 °C) 7.31-7.42 (5 H, m, 5 x CH Ar), 5.59 (1 H, dt, *J* 14.9 and 6.5 Hz, 3-H), 5.48 (1 H, dt, *J* 14.6 and 5.8 Hz, 2-H), 5.12 (2 H, s, CH_2 Bn), 4.82 (1 H, s, NH), 4.60 (1 H, s, NH), 3.69 (2 H, br m, 1-H_2), 3.22 (2 H, c, *J* 6.5 Hz, 6-H_2), 2.08 (2 H, c, *J* 6.8 Hz, 4-H_2), 1.61 (2 H, q, *J* 7.1 Hz, 5-H_2), 1.47 (9 H, s, *t*-Bu).
δ_C (125 MHz, $CDCl_3$, 27 °C) 156.4 (C=O), 155.8 (C=O), 136.6 (C Ar), 131.5 (C-3), 128.5 (2 x CH Ar), 128.1 (3 x CH Ar), 127.4 (C-2), 79.3 (C *t*-Bu), 66.6 (CH_2 Bn), 42.5 (C-1), 40.5 (C-6), 29.3 (C-4 and C-5), 28.4 (*t*-Bu).

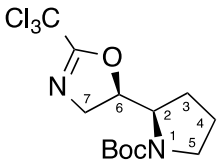
IR (film): ν_{max} 3350, 2977, 2933, 1686, 1518, 1366, 1167, 1041, 994, 828, 678 cm^{-1}.
HRMS (ES) m/z calcd for $C_{19}H_{29}N_2O_4$ $[M+H]^+$ 349.212184, found 349.211842.

***tert*-Butyl (2*S**)-2-[(1*R**)-1-iodo-2-(2,2,2-trichloroacetamido)ethyl]pyrrolidine-1-carboxylate, 4.**



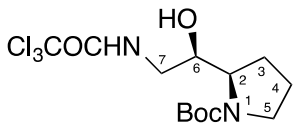
As general procedure C, using trichloracetamide **1a** (1.7 g, 4.73 mmol). After chromatographic purification (gradient elution: 20:80 → 80:20 EtOAc–hexane) the pyrrolidine **4** (1.74 g, 76%) was obtained as a colourless oil.
R_f 0.20 (25% EtOAc – hexane).
δ_H (500 MHz, $CDCl_3$, 27 °C) 8.45 (1 H, s, NH), 4.22-4.36 (1 H, m, 7-H_A), 3.99-4.14 (2 H, m, 7-H_B and 2-H), 3.41-3.49 (1 H, m, 5-H_A), 3.30-3.39 (2 H, m, 5-H_B and 6-H), 2.03-2.13 (2 H, m, 3-H_2), 1.85-1.98 (2 H, m, 4-H_2), 1.46 (9 H, s, *t*-Bu).
δ_C (125 MHz, $CDCl_3$, 27 °C) 162.2 (C=O), 156.1 (C=O), 93.1 (CCl_3), 80.8 (C *t*-Bu), 60.9 (C-2), 47.5 (C-6), 44.9 (C-5), 36.7 (C-7), 31.3 (C-3), 28.4 (*t*-Bu), 23.5 (C-4).
IR (film): ν_{max} 3318, 2974, 2928, 1687, 1519, 1480, 1452, 1392, 1271, 1251, 1167, 761 cm^{-1}.
HRMS (ES) m/z calcd for $C_{13}H_{21}Cl_3IN_2O_3$ $[M+H]^+$ 484.96624, found 484.96721.

***tert*-Butyl (2*R**)-2-[(5*S**)-2-(trichloromethyl)-4,5-dihydrooxazol-5-yl]pyrrolidine-1-carboxylate, 5.**



To a cold solution (0 ºC) of oxazine **12** (1.70 g, 3.50 mmol) in 35.0 mL of THF, LiHMDS 1.0 M solution in THF (7.0 mL, 7.00 mmol) was added. The solution was stirred and let it warm up to rt until completion (4 hr), followed by LC-MS. The volatiles were evaporated to give the crude reaction. After chromatographic purification (gradient elution: 20:80 → 80:20 EtOAc – hexane) the trichloroacetimidate **6** (1.1 g, 88%) was obtained as a colourless oil.
R_f 0.40 (50% EtOAc – hexane).
δ_H (500 MHz, $CDCl_3$, 27 °C) 5.29 (1 H, td, *J* 9.1 and 3.5 Hz, 6-H), 4.21 (1 H, br m, 2-H), 4.05 (2 H, t, *J* 11.0, 7-H_2), 3.50-3.74 (1 H, m, 5-H_A), 3.22-3.34 (1 H, m, 5-H_B), 1.92-2.19 (2 H, m, 3-H_2), 1.73-1.91 (2 H, m, 4-H_2), 1.47 (9 H, s, *t*-Bu).
δ_C (125 MHz, $CDCl_3$, 27 °C) 162.4 (C=N), 80.1 (C-6), 56.5 (C-2), 48.0 (C-5 and C-7), 29.7 (C-4), 28.4 (*t*-Bu), 27.5 (C-3).
IR (film): ν_{max} 2976, 2934, 1700, 1671, 1509, 1394, 1366, 1250, 1163, 1110, 820 cm^{-1}.
HRMS (ES) m/z calcd for $C_{13}H_{20}Cl_3N_2O_3$ $[M+H]^+$ 357.05395, found 357.05409.

***tert*-Butyl (2*R**)-2-[(1*S**)-1-hydroxy-2-(2,2,2-trichloroacetamido)ethyl]pyrrolidine-1-carboxylate, 6.**



To a solution of trichloroacetimidate **6** (1.1 g, 3.08 mmol) in a 6:4 mixture of Py:H_2O (30.1 mL), PTSA (0.292 mg, 1.54 mmol) was added at rt. The solution was stirred at 70 ºC until completion (12 hr), followed by LC-MS. The volatiles were evaporated to give the crude reaction. After

chromatographic purification (gradient elution: 20:80 → 80:20 EtOAc – CH_2Cl_2) the alcohol **7** (1.03 g, 90%) was obtained as a colourless oil.

R_f 0.30 (10% EtOAc – CH_2Cl_2).

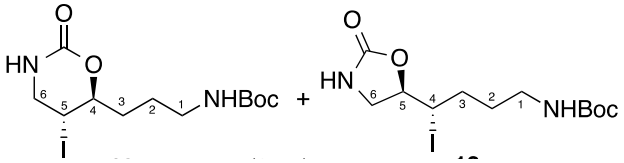
δ_H (500 MHz, $CDCl_3$, 27 °C) 7.64 (1 H, s, NH), 5.60 (1 H, s, OH), 3.83 (1 H, td, *J* 7.7 and 5.0 Hz, 2-H), 3.66 (1 H, td, *J* 7.1 and 4.5 Hz, 6-H), 3.41-3.55 (2 H, m, 5-H_A and 7-H_A), 3.37 (1 H, dt, *J* 13.4 and 6.0 Hz, 7-H_B), 3.26 (1 H, dt, *J* 11.0 and 6.8 Hz, 5-H_B), 2.02 (1 H, dt, *J* 12.8 and 7.5 Hz, 3-H_A), 1.89 (1 H, dquint, *J* 13.2 and 6.7 Hz, 4-H_A), 1.77 (1 H, dquint, *J* 13.3 and 6.7 Hz, 4-H_B), 1.72 (1 H, dt, *J* 11.9 and 6.2 Hz, 3-H_B), 1.45 (9 H, s, *t*-Bu).

δ_C (125 MHz, $CDCl_3$, 27 °C) 162.2 (CONH), 158.1 (C=O Boc), 92.7 (CCl_3), 81.0 (C *t*-Bu), 74.0 (C-6), 60.5 (C-2), 47.6 (C-5), 44.6 (C-7), 28.7 (C-4), 28.4 (*t*-Bu), 24.2 (C-3).

IR (film): ν_{max} 3326, 2976, 2935, 2885, 1701, 1672, 1509, 1397, 1366, 1250, 1164, 1129, 1110, 1033, 822, 666 cm^{-1}.

HRMS (ES) m/z calcd for $C_{13}H_{22}Cl_3N_2O_4$ $[M+H]^+$ 375.063967, found 375.064198.

***tert*-Butyl {3-[(5*R**,6*S**)-5-iodo-2-oxo-1,3-oxazinan-6-yl]propyl}carbamate, 7, and *tert*-Butyl {(4*S**)-4-iodo-4-[(5*R**)-2-oxooxazolidin-5-yl]butyl}carbamate, S6.**



As general procdure C, using dicarbamate **1b** (520 mg, 1.65 mmol). After chromatographic purification (gradient elution: 20:80 → 80:20 EtOAc – CH_2Cl_2) the oxazine **7** (411 mg, 66%) and the oxazolidinone **S6** (20 mg, 4%) were obtained as colourless oils.

Data for **7**:

R_f 0.30 (50% EtOAc – CH_2Cl_2).

δ_H (500 MHz, $CDCl_3$, 27 °C) 6.03 (1 H, s, NH), 4.62 (1 H, s, NHBoc), 4.41 (1 H, td, *J* 8.9 and 2.7 Hz, 4-H), 4.05 (1 H, td, *J* 9.4 and 5.4 Hz, 5-H), 3.72 (1 H, dt, *J* 11.8 and 4.5 Hz, 6-H_A), 3.66 (1 H, dd, *J* 11.9 and 9.9 Hz, 6-H_B), 3.18 (2 H, d, *J* 6.1 Hz, 1-H_2), 2.03-2.18 (1 H, m, 3-H_A), 1.71-1.90 (2 H, m, 3-H_B and 2-H_A), 1.58-1.70 (1 H, m, 2-H_B), 1.45 (9 H, s, *t*-Bu).

δ_C (125 MHz, $CDCl_3$, 27 °C) 156.1 (C=O), 153.2 (C=O), 81.8 (C-4), 79.3 (C *t*-Bu), 49.0 (C-6), 39.9 (C-1), 31.5 (C-3), 28.5 (*t*-Bu), 25.0 (C-2), 17.1 (C-5).

IR (film): ν_{max} 2980, 2950, 1701, 1671, 1509, 1394, 1366, 1251, 820 cm^{-1}.

HRMS (ES) m/z calcd for $C_{12}H_{22}IN_2O_4$ $[M+H]^+$ 385.061882, found 385.062247. The large coupling constant (9.4 Hz) was indicative of the *trans* configuration. The regioselectivity was confirmed by the chemical shifts of 5-H and C-5.

Data for **S6**:

R_f 0.40 (50% EtOAc – CH_2Cl_2).

δ_H (500 MHz, $CDCl_3$, 27 °C) 5.18 (1 H, s, NH), 4.66 (1 H, ap q, *J* 8.6 Hz, 5-H), 4.56 (1 H, s, NHBoc), 4.11 (1 H, td, *J* 9.4 and 2.9 Hz, 4-H), 3.79 (1 H, t, *J* 8.7 Hz, 6-H_A), 3.42 (1 H, t, *J* 8.1 Hz, 6-H_B), 3.18 (2 H, d, *J* 5.5 Hz, 1-H_2), 2.00-2.11 (1 H, m, 3-H_A), 1.76-1.86 (2 H, m, 3-H_B and 2-H_A), 1.58-1.65 (1 H, m, 2-H_B), 1.47 (9 H, s, *t*-Bu).

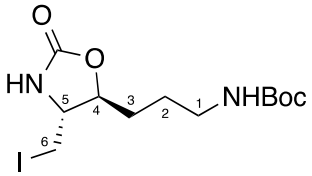
δ_C (125 MHz, $CDCl_3$, 27 °C) 158.4 (C=O), 156.0 (C=O), 79.0 (C-5), 47.3 (C-6), 36.7 (C-1), 32.6 (C-4), 29.7 (C-2 and C-3), 28.41 (*t*-Bu), 28.36 (C-1).

NOESY 2D (500 MHz, $CDCl_3$): between 5-H and 6-H_B, between 4-H and 6-H_A.

IR (film): ν_{max} 3318, 2974, 2928, 1687, 1519, 1480, 1452, 1392, 1271, 1251, 1167, 761 cm^{-1}.

HRMS (ES) m/z calcd for $C_{12}H_{22}IN_2O_4$ $[M+H]^+$ 385.061882, found 385.062211.

***tert*-Butyl {3-[(4*R**,5*S**)-4-(iodomethyl)-2-oxooxazolidin-5-yl]propyl}carbamate, 8.**



To a cold solution (0 ºC) of oxazine **7** (100 mg, 0.26 mmol) in 2.6 mL of THF, LiHMDS 1.0 M solution in THF (0.30 mL, 0.29 mmol) was added. The solution was stirred and warmed up to rt until completion (4 hr), followed by LC-MS. The volatiles were evaporated to give the crude reaction. After chromatographic purification (gradient elution: 20:80 → 80:20 EtOAc – CH_2Cl_2) the oxazolidinone **8** (85 mg, 85%) was obtained as a colourless oil.

R_f 0.30 (50% EtOAc – CH_2Cl_2).

δ_H (500 MHz, $CDCl_3$, 27 °C) 5.49 (1 H, s, NH), 4.58 (1 H, s, NHBoc), 4.28 (1 H, dt, *J* 6.3 and 4.5 Hz, 4-H), 3.66 (1 H, tdd, *J* 6.0 and 4.3 and 1.0 Hz, 5-H), 3.13-3.28 (4 H, m, 1-H_2 and 6-H_2), 1.74-1.84 (2 H, m, 3-H_2), 1.63-1.74 (2 H, m, 2-H_2), 1.45 (9 H, s, *t*-Bu).

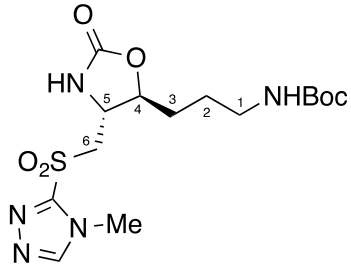
δ_C (125 MHz, $CDCl_3$, 27 °C) 157.4 (C=O), 156.1 (C=O), 81.9 (C-4), 79.4 (C *t*-Bu), 58.5 (C-5), 39.8 (C-1), 32.2 (C-3), 28.4 (*t*-Bu), 25.5 (C-2), 7.7 (C-6).

NOESY 2D (500 MHz, $CDCl_3$): between 4-H and 6-H, between 5-H and 6-H, between 4-H and 3-H_2, between 5-H and 3-H_2.

IR (film): ν_{max} 3342, 2922, 2851, 1746, 1689, 1522, 1455, 1392, 1366, 1251, 1167, 764 cm^{-1}.

HRMS (ES) m/z calcd for $C_{12}H_{22}IN_2O_4$ $[M+H]^+$ 385.061882, found 385.061898.

***tert*-Butyl {3-{(4*R**,5*S**)-4-[(4-methyl-4*H*-1,2,4-triazol-3-yl)sulfonylmethyl]-2-oxooxazolidin-5-yl]propyl}}carbamate, 9.**



To a solution of oxazolidinone **8** (34 mg, 0.088 mmol) in 1.0 mL of CH_2Cl_2, DBU (22 µL, 0.176 mmol) followed by 4-methyl-4H-1,2,4-triazole-3-thiol (20 mg, 0.176 mmol) were added. The solution was stirred at rt until completion (12 hr), followed by LC-MS. The volatiles were evaporated to give the crude reaction, which was filtered through a small pack of silica to give the crude sulfide.

To a solution of the crude sulfide in CH_2Cl_2 (1 mL), *m*-CPBA (77% purity, 78 mg, 0.352 mmol) was added. The solution was stirred at rt for 12 hr. The volatiles were evaporated to give the crude reaction. After chromatographic purification (gradient elution: 5:95 → 30:70 MeOH – CH_2Cl_2) the sulfone **9** (29 mg, 83% 2 steps) was obtained as a white foam.

R_f 0.30 (50% MeOH – CH_2Cl_2).

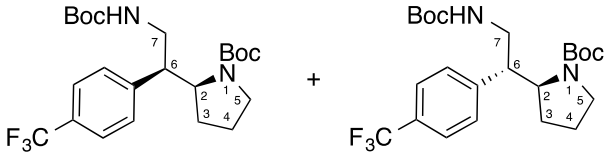
δ_H (500 MHz, CD_3OD, 27 °C) 8.59 (1 H, s, Ar), 4.47 (1 H, dt, *J* 8.3 and 4.5 Hz, 4-H), 4.04 (1 H, dt, *J* 6.3 and 4.8 Hz, 5-H), 3.99 (1 H, dd, *J* 14.8 and 5.2 Hz, 6-H_A), 3.89 (3 H, s, NMe), 3.83 (1 H, dd, *J* 14.8 and 6.5 Hz, 6-H_B), 2.99 (2 H, td, *J* 6.8 and 3.9 Hz, 1-H_2), 1.60-1.74 (2 H, m, 3-H_2), 1.42-1.56 (2 H, m, 2-H_2), 1.33 (9 H, s, *t*-Bu).

δ_C (125 MHz, CD_3OD, 27 °C) 160.4 (C=O), 149.3 (CH Ar), 82.1 (C-4), 79.9 (C *t*-Bu), 60.0 (C-5), 53.8 (C-6), 40.6 (C-1), 33.6 (NMe), 32.6 (C-3), 28.7 (*t*-Bu), 26.3 (C-2).

IR (film): ν_{max} 3327, 2976, 2929, 1748, 1695, 1513, 1366, 1246, 1164, 1136, 734 cm^{-1}.

HRMS (ES) m/z calcd for $C_{15}H_{25}N_5NaO_6S$ $[M+Na]^+$ 426.141775, found 426.142096.

***tert*-Butyl (2*S**)-2-{(1*S**)-2-(*tert*-butoxycarbonylamino)-1-[4-(trifluoromethyl)phenyl]ethyl}pyrrolidine-1-carboxylate, 10a, and *tert*-Butyl (2*S**)-2-{(1*R**)-2-*tert*-butoxycarbonylamino)-1-[4-(trifluoromethyl)phenyl]ethyl}pyrrolidine-1-carboxylate, 11a.**



As general procedure D, using alkene **1a** (55 mg, 0.158 mmol). The diastereomeric ratio as measured in the crude NMR was 4:1. After chromatographic purification (gradient elution: 10:90 → 30:70 EtOAc – hexane) **10a** (47 mg, 41%) and **11a** (12 mg, 10%) were obtained as colourless foams.

Data for **10a**:

R_f 0.30 (20% EtOAc – hexane).

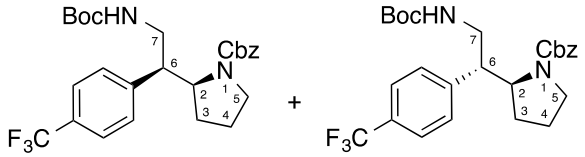
δ_H (500 MHz, $CDCl_3$, 27 °C) 7.58 (2 H, d, *J* 7.2 Hz, 2 x CH Ar), 7.32 (2 H, t, *J* 7.9 Hz, 2 x CH Ar), 5.60 (1 H, s, NH), 4.33 (1 H, br m, 2-H), 3.46-3.62 (1 H, m, 7-H_A), 3.25-3.44 (2 H, m, 7-H_B and 5-H_A), 3.15 (1 H, br m, 6-H), 2.98 (1 H, br m, 5-H_B), 1.81-2.00 (1 H, m, 3-H_A), 1.62-1.78 (1 H, m, 4-H_A), 1.46-1.53 (2 H, m, 3-H_B and 4-H_B), 1.50 (9 H, s, *t*-Bu), 1.46 (9 H, s, *t*-Bu).

δ_C (125 MHz, $CDCl_3$, 27 °C) 156.2 (C=O), 156.1 (C=O), 144.3 (C Ar), 129.2 (3 x C Ar), 125.2 (2 x CH Ar), 79.7 (C *t*-Bu), 77.3 (CF_3), 58.2 (C-2), 50.2 (C-5), 47.8 (C-6), 41.9 (C-7), 29.0 (C-3), 28.5 (6 x CH_3 *t*-Bu), 23.6 (C-4).

IR (film): ν_{max} 3351, 2977, 2933, 1688, 1615, 1519, 1391, 1366, 1323, 1250, 1159, 1119, 1067, 845, 770 cm^{-1}.

HRMS (ES) m/z calcd for $C_{23}H_{34}F_3N_2O_4$ $[M+H]^+$ 459.246519, found 459.246662.

Benzyl (2*S)-2-{(1*S**)-2-(*tert*-butoxycarbonylamino)-1-[4-(trifluoromethyl)phenyl]ethyl}pyrrolidine-1-carboxylate, 10b, and Benzyl (2*S**)-2-{(1*R**)-2-(*tert*-butoxycarbonylamino)-1-[4-(trifluoromethyl)phenyl]ethyl}pyrrolidine-1-carboxylate, 11b.**



As general procedure D, using alkene **1c** (80 mg, 0.254 mmol). The diastereomeric ratio as measured in the crude NMR was 5:1. After chromatographic purification (gradient elution: 10:90 → 30:70 EtOAc – hexane) **10b** (30 mg, 41%) and **11b** (7 mg, 8%) were obtained as colourless foams.

Data for **16b**:

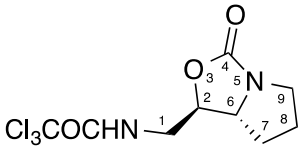
R_f 0.20 (20% EtOAc – hexane).

δ_H (500 MHz, $CDCl_3$, 27 °C) 7.56 (2 H, d, *J* 8.1 Hz, 2 x CH Ar), 7.30-7.42 (7 H, m, 7 x CH Ar), 5.63 (1 H, br m, NH), 5.18 (2 H, s, CH_2 Bn), 4.39 (1 H, ap t, *J* 9.8 Hz, 2-H), 4.06-4.21 (1 H, m, 7-H_A), 3.47-3.76 (2 H, m, 7-H_B and 5-H_A), 3.26-3.45 (2 H, br m, 5-H_B and 6-H), 2.00-2.11 (1 H, m, 3-H_A), 1.83-2.00 (2 H, m, 3-H_B and 4-H_B), 1.63-1.77 (1 H, m, 4-H_A), 1.44 (9 H, s, *t*-Bu).

δ_C (125 MHz, $CDCl_3$, 27 °C) 156.2 (C=O), 156.1 (C=O), 144.3 (C Ar), 129.2 (3 x C Ar), 125.2 (2 x CH Ar), 79.7 (C *t*-Bu), 77.3 (CF_3), 58.2 (C-2), 50.2 (C-5), 47.8 (C-6), 41.9 (C-7), 29.0 (C-3), 28.5 (6 x CH_3 *t*-Bu), 23.6 (C-4).

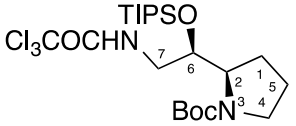
IR (film): ν_{max} 3325, 2927, 2854, 1691, 1500, 1412, 1324, 1251, 1162, 1118, 1068, 1018, 970, 844, 772 cm^{-1}.
HRMS (ES) m/z calcd for $C_{26}H_{32}F_3N_2O_4$ $[M+H]^+$ 493.230869, found 493.231182.

2,2,2-trichloro-*N*-((3-oxotetrahydro-1*H*,3*H*-pyrrolo[1,2-*c*]oxazol-1-yl)methyl)acetamide, 13.



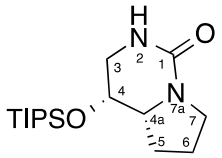
To a solution of *N*-Boc pyrrolidine **6** (32 mg, 0.085 mmol) in CH_2Cl_2 (1 mL), TFA (20 μL, 0.256 mmol) was added at rt and stirred overnight. The solvent was evaporated to give the crude amine as the TFA salt which was used without further purification in the next step.
To a cold solution (0 ºC) of the crude amine in CH_2Cl_2 (1 mL), CDI (28 μL, 0.17 mmol) was added. The solution was stirred at rt for 24 h. The volatiles were evaporated to give a crude material. After chromatographic purification (gradient elution: 20:80 → 80:20 EtOAc – CH_2Cl_2) the urea **13** (20 mg, 80%) was obtained as a colourless foam.
R_f 0.30 (50% EtOAc – CH_2Cl_2).
δ_H (500 MHz, $CDCl_3$, 27 °C) 7.23 (1 H, s, NH), 4.51 (1 H, dt, *J* 7.0 and 3.5 Hz, 2-H), 3.83 (1 H, ddd, *J* 14.4 and 6.6 and 3.3 Hz, 1-H_A), 3.54-3.73 (3 H, m, 1-H_B and 6-H and 9-H_A), 3.17 (1 H, ddd, *J* 11.4 and 9.1 and 4.4 Hz, 9-H_B), 2.03-2.20 (2 H, m, 8-H_A and 7-H_A), 1.85-2.00 (1 H, m, 8-H_B), 1.53 (1 H, dq, *J* 12.2 and 9.3 Hz, 7-H_B).
δ_C (125 MHz, $CDCl_3$, 27 °C) 162.9 (C=O), 160.3 (C-4), 92.1 (CCl_3), 78.1 (C-2), 62.1 (C-6), 45.7 (C-9), 44.5 (C-1), 30.5 (C-7), 25.6 (C-8).
NOESY 2D (500 MHz, $CDCl_3$): between 2-H and 7-H_B.
IR (film): ν_{max} 3325, 2926, 2853, 1748, 1706, 1541, 1250, 1040, 822, 767, 665 cm^{-1}.
HRMS (ES) m/z calcd for $C_{20}H_{24}N_3O_6S$ $[M+H]^+$ 300.991350, found 300.990504.

***tert*-Butyl (*R**)-2-((*S**)-2-(2,2,2-trichloroacetamido)-1-[(triisopropylsilyl)oxy)ethyl]pyrrolidine-1-carboxylate, S7.**



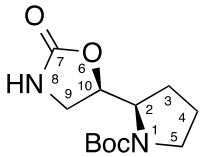
To a solution of alcohol **6** (30 mg, 0.080 mmol) in DMF (1.3 mL), ImH (17 mg, 0.24 mmol, 3.0 equiv), and TIPSOTf (44 μL, 0.16 mmol, 2.0 equiv) were added. The mixture was stirred at rt for 24 h. The reaction was quenched with H_2O (1 mL) and saturated $NaHCO_3$ solution (1 mL) and the layers were separated. The aqueous layer was extracted with EtOAc (3 × 5 mL), and the combined organic extracts were washed with saturated NaCl solution (5 mL), dried over Na_2SO_4 and concentrated under reduced pressure. Purification by column chromatography (gradient elution: 2:98 → 10:90 EtOAc – hexane) afforded **S7** (38 mg, 90%) as a colourless oil.
R_f 0.30 (10% EtOAc – hexane).
δ_H (500 MHz, $CDCl_3$, 27 °C) 8.75 (1 H, s, NH), 4.08 (1 H, br m, 6-H), 4.04 (1 H, br m, 2-H), 3.67-3.83 (1 H, m, 7-H_A), 3.47-3.60 (1 H, m, 5-H_A), 3.30 (1 H, d, *J* 6.2 Hz, 5-H_B), 2.87-3.02 (1 H, m, 7-H_B), 1.92-2.07 (3 H, m, 3-H_2 and 4-H_A), 1.71-1.83 (1 H, m, 4-H_B), 1.47 (9 H, s, *t*-Bu), 1.13-1.18 (3 H, m, 3 x CH *i*-Pr), 1.11 (9 H, d, *J* 5.1 Hz, *i*-Pr), 1.10 (9 H, d, *i*-Pr).
δ_C (125 MHz, $CDCl_3$, 27 °C) 162.3 ($COCCl_3$), 156.6 (C=O Boc), 92.9 (CCl_3), 80.0 (C *t*-Bu), 73.0 (C-6), 58.3 (C-2), 47.9 (C-5), 43.5 (C-7), 28.5 (*t*-Bu), 24.4 (C-4), 18.2 (*i*-Pr), 18.1 (*i*-Pr), 13.0 (*i*-Pr).
HRMS (ES) m/z calcd for $C_{22}H_{42}Cl_3N_2O_4Si$ $[M+H]^+$ 531.197394, found 531.198141.

(4a*R)-4-((triisopropylsilyl)oxy)hexahydropyrrolo[1,2-*c*]pyrimidin-1(2*H*)-one, 14.**



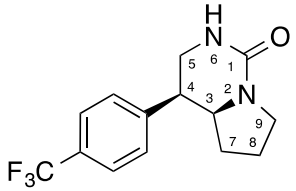
To a solution of *N*-Boc pyrrolidine **10** (25 mg, 0.047 mmol) in CH_2Cl_2 (1 mL), TFA (20 μL, 0.235 mmol) was added at rt and stirred overnight. The solvent was evaporated to give crude amine which was used withour further purification in next step.
To a cold solution (0 ºC) of the crude amine in THF (1 mL), LiHMDS (1.0 M solution in THF, 36 μL, 0.036 mmol) was added. The solution was stirred at rt for 4 hr. The volatiles were evaporated to give the crude material. After chromatographic purification (gradient elution: 5:95 → 30:70 MeOH – CH_2Cl_2) the urea **14** (12 mg, 82%) was obtained as a colourless foam.
R_f 0.30 (20% MeOH – CH_2Cl_2).
δ_H (500 MHz, $CDCl_3$, 27 °C) 4.45 (1 H, d, *J* 4.6 Hz, NH), 4.18 (1 H, q, *J* 2.0 Hz, 4-H), 3.59 (1 H, ddd, *J* 7.1 and 5.8 and 1.7 Hz, 4a-H), 3.52 (1 H, ddd, *J* 10.1 and 8.4 and 2.0 Hz, 7-H_A), 3.43 (1 H, td, *J* 10.3 and 6.7 Hz, 7-H_B), 3.37 (1 H, dt, *J* 12.6 and 1.1 Hz, 3-H_A), 3.31 (1 H, ddd, *J* 12.6 and 5.3 and 2.4 Hz, 3-H_B), 1.99-2.10 (2 H, m, 5-H_A), 1.87-1.98 (2 H, m, 5-H_B and 6-H_A), 1.70-1.83 (1 H, m, 6-H_B), 1.13-1.18 (3 H, m, 3 x CH *i*-Pr), 1.11 (9 H, d, *J* 5.1 Hz, *i*-Pr), 1.10 (9 H, d, *i*-Pr).
δ_C (125 MHz, $CDCl_3$, 27 °C) 155.6 (C=O), 62.9 (C-4), 61.0 (C-4a), 48.0 (C-3), 46.1 (C-7), 28.0 (C-5), 23.2 (C-6), 18.14 (*i*-Pr), 18.08 (*i*-Pr), 12.7 (*i*-Pr).
NOESY 2D (500 MHz, $CDCl_3$): between 4-H and 4a-H.
IR (film): ν_{max} 3292, 2944, 2890, 2866, 1646, 1515, 1460, 1247, 1131, 883, 665 cm^{-1}.
HRMS (ES) m/z calcd for $C_{32}H_{65}N_4O_4Si_2$ $[2M+H]^+$ 625.453457, found 625.453886.

***tert*-Butyl (2*R**)-2-[(5*S**)-2-oxoazolidin-5-yl]pyrrolidine-1-carboxylate, 15.**



To a cold solution (0 ºC) of alcohol **6** (470 mg, 1.25 mmol) in THF (13 mL), KO*t*-Bu (211 mg, 1.88 mmol) was added. The solution was stirred at rt until for 2 hr. The volatiles were evaporated to give the crude material. After chromatographic purification (gradient elution: 5:95 → 30:70 MeOH – CH_2Cl_2) the oxazolidinone **15** (280 mg, 89%) was obtained as a colourless oil.
R_f 0.30 (20% MeOH – CH_2Cl_2).
δ_H (500 MHz, $CDCl_3$, 50 °C) 5.06 (1 H, s, NH), 4.87 (1 H, td, *J* 8.6 and 3.6 Hz, 10-H), 4.05-4.25 (1 H, m, 2-H), 3.58 (2 H, br m, 9-H_A and 5-H_A), 3.53 (1 H, t, *J* 8.7 Hz, 9-H_B), 3.19-3.38 (1 H, m, 5-H_B), 1.95-2.15 (2 H, m, 3-H_A and 4-H_A), 1.75-1.94 (2 H, m, 3-H_B and 4-H_B), 1.47 (9 H, s, *t*-Bu).
δ_C (125 MHz, $CDCl_3$, 50 °C) 159.2 (C=O), 80.12 (C *t*-Bu), 80.07 (C-10), 58.2 (C-2), 47.8 (C-9), 42.3 (C-5), 28.4 (*t*-Bu), 27.7 (C-4), 23.8 (C-3).
NOESY 2D (500 MHz, $CDCl_3$): between 10-H and 2-H, between 10-H and 9-H_2, between 3-H_A and 9-H_B.
HRMS (ES) m/z calcd for $C_{12}H_{20}N_2NaO_4$ $[M+Na]^+$ 279.131528, found 279.131694.

(4*S,4a*S**)-4-[4-(Trifluoromethyl)phenyl]hexahydropyrrolo[1,2-*c*]pyrimidin-1(2*H*)-one, 16.**



To a solution of dicarbamate **16a** (27 mg, 0.06 mmol) in CH_2Cl_2 (1 mL), TFA (44 µL, 0.60 mmol) was added at rt and the resulting solution was stirred overnight. The solvent was evaporated to give crude amine that was used without further purification in next step.
To a cold solution (0 ºC) of the crude amine in CH_2Cl_2 (1 mL), CDI (15 µL, 0.09 mmol) and DIPEA (32 µL, 0.18 mmol) were added. The solution was stirred at rt for 1 d. The volatiles were evaporated to give the crude reaction. After chromatographic purification (gradient elution: 5:95 → 30:70 MeOH – CH_2Cl_2) pure urea **18** (13 mg, 80% over 2 steps) was obtained as a colourless foam.
R_f 0.30 (20% MeOH – CH_2Cl_2).
δ_H (500 MHz, $CDCl_3$, 27 °C) 7.60 (1 H, d, *J* 8.2 Hz, CH Ar), 7.35 (2 H, t, *J* 8.2 Hz, 2 x CH Ar), 4.84 (1 H, d, *J* 3.3 Hz, NH), 3.99 (1 H, ddd, *J* 10.1 and 5.8 and 4.3 Hz, 3-H), 3.82 (1 H, dd, *J* 12.1 and 4.9 Hz, 5-H_A), 3.58 (1 H, ddd, *J* 11.0 and 8.5 and 2.9 Hz, 9-H_A), 3.43 (1 H, ddd, *J* 12.3 and 4.7 and 1.3 Hz, 5-H_B), 3.38 (1 H, ddd, *J* 10.5 and 7.5 and 1.9 Hz, 9-H_B), 3.28 (1 H, t, *J* 4.4 Hz, 4-H), 1.94 (1 H, qd, *J* 5.8 and 2.3 Hz, 7-H_A), 1.68-1.81 (2 H, m, 8-H_2), 1.19-1.31 (1 H, m, 7-H_B).
δ_C (125 MHz, $CDCl_3$, 27 °C) 155.3 (C=O), 143.6 (1C, s, $C_{para\text{-}CF_3}$ Ar), 129.5 (1 C, d, $J_{Cipso\text{-}F}$ 32 Hz, $C_{ipso\text{-}F}$ Ar), 129.2 (2 C, s, 2 x $C_{meta\text{-}CF_3}$ Ar), 125.5 (2 C, d, $J_{Corto\text{-}F}$ 4 Hz, 2 x $C_{orto\text{-}F}$ Ar), 77.2 (CF_3), 55.8 (C-3), 46.1 (C-5), 45.3 (C-9), 39.4 (C-4), 29.8 (C-7), 22.7 (C-8).
IR (film): ν_{max} 3294, 2974, 2877, 1716, 1509, 1326, 1164, 1116, 1071 cm^{-1}.
NOESY 2D (500 MHz, $CDCl_3$): between 3-H and 4-H, between 3-H and 7-H_A.
HRMS (ES) m/z calcd for $C_{14}H_{16}F_3N_2O$ $[M+H]^+$ 285.120924, found 285.121232.

3. LLAMA Details

3.1. LLAMA system architecture

The LLAMA software comprises a web browser based user interface and a computational backend implemented in Python v2.6.6. Data is held in a MySQL[1] database. LLAMA is hosted by the School of Chemistry at the University of Leeds, and runs on a CentOS Linux server with a pair of Intel Xeon E5-2650 CPUs and 32 GB RAM.
Users interact with LLAMA via a web browser. The interface displays a variety of interactive plots that give insights into the properties of a user's molecules. The interface is implemented in PHP, HTML5 and JavaScript and uses the jQuery and D3.js[2] libraries to generate interactive user interface elements.
The computational backend is implemented in a highly multithreaded architecture to take advantage of the many available CPU cores. Three services are implemented in the backend; library decoration, bulk upload and novelty search. Inter-process communication between the frontend and backend, and between backend processes, is mostly mediated by ZeroMQ[3] messages.

3.2. Libraries

Collections of molecules and reactions are held in self-contained "libraries". Each library is owned by one user but can be shared with others to aid collaboration.
A default set of pharmaceutically relevant capping groups ("reactants") is added automatically to each library. Users may add additional reactants or remove ones that they do not require.
Libraries also contain carefully selected medicinally relevant organic reactions, implemented in SMARTS[4] format. Many of the reaction SMARTS in LLAMA were adapted from the examples provided by Hartenfeller *et al.*[5] Reactions can be enabled or disabled should the user wish to use

only a subset of the available chemistries. Users may add their own reactions in addition to the default set, if they wish.
Users must upload one or more "scaffolds" to be decorated with the capping groups using the reactions. Scaffolds and reactants can be drawn individually or uploaded in bulk.

3.3. Decoration engine

When a user decorates a library LLAMA attempts to react all of the library's scaffolds with all of the reactants, using all of the enabled reactions. The associated workflow is shown in Figure 1. Permutations of scaffolds and reactants that do not match a given reaction SMARTS are filtered out (Fig. S1, A) to reduce unnecessary computational burden. Valid permutations are sent to the reaction process pool (Fig S1, B). When a virtual reaction yields one or more product molecules their properties are calculated and written to the database. The products are then fed back into the start of the process to be decorated again (Fig S1, C). Each scaffold may be decorated with reactants a maximum of twice. When all of the possible permutations have been processed an email is sent to the user and a notification is displayed on the user interface. This process typically takes less than a minute for a single scaffold.
During decoration ZeroMQ messages are sent to the user interface to indicate when product molecules have been formed. These messages cause user interface elements such as the analysis plots to update, so the user can see the results of the decoration calculation as it progresses.

3.4. Bulk upload engine

To reduce the amount of time that users need to spend inputting molecules into LLAMA we have implemented a convenient bulk upload tool that supports Structure Data Format (.sdf) and ChemDraw (.cdx) files. When a user uploads a file containing molecules to LLAMA the backend's bulk upload engine. This engine uses a pool of worker processes to upload the molecules in the file in parallel. The upload engine uses ZeroMQ messages to inform the user of the status of the upload. The user interface displays a notification when this process is complete, and the backend sends a report to the user containing information about any molecules that failed to upload.

3.5. Novelty search engine

In the drug discovery process it can be helpful to know whether a scaffold is already known or novel. LLAMA gives an indication of the novelty of a scaffold by comparing it to a subset of all commercially-available compounds. It should be noted that this novelty assessment algorithm cannot guarantee that a given scaffold is truly novel. Rather, this information gives an insight into the likely novelty of the framework of a scaffold and the novelty of its substitution pattern. Highly novel scaffolds are generally based on scaffolds that are not found in commercially-available compounds.
The novelty search engine performs two searches for a given scaffold against a reference set comprising a random 2% sample of molecules in the ZINC 'all now' database.[6]
The first search finds the number of exact matches of the scaffold's Murcko framework[7] against those of the 2% ZINC subset. The canonical SMILES representation of the scaffold's Murcko framework is compared via a simple string comparison to the pre-computed canonical SMILES of the Murcko frameworks in the ZINC subset.
The second search finds the number of Murcko frameworks in the 2% ZINC reference set in which the scaffold's Murcko framework appears as a substructure.

3.6. Molecular properties

LLAMA computes a variety of medicinally-relevant molecular properties for all molecules. The following software packages and services are used to generate these properties.

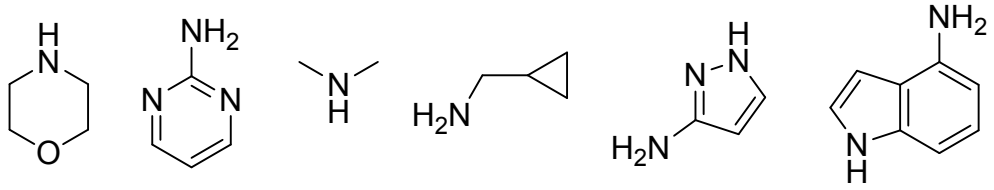
Property	Calculated with	Physico-chemical relevance
Canonical SMILES, RMM, Lipinski failure count[8], heavy atoms, number of aliphatic rings, number of aromatic rings, number of heteroatoms, number of carbon atoms, number of sp3 carbon atoms, number of chiral centres, principal moments of inertia.	Indigo toolkit[9]	Bioavailability
AlogP	VCCLab AlogPS service[10]	Bioavailability
Rotatable bonds	Implementation uses the definition of rotatable bonds in Veber *et al.*[11,a]	Oral bioavailability[10]
Topological polar surface area	The code in this module is derived from the code from the Chemistry Toolkit Rosetta Wiki that implements the algorithm described by Ertl et al.[12]	Penetration of cell membranes[10b, 13] and the blood-brain barrier[14]
Lowest energy 3D conformer and associated SD representation.	RDKit Q4 2012.12.1[15]	3D character
Plane of best fit	We are grateful to Greg Landrum for implementing a Python version of the RDKit "plane of best fit" module for use in LLAMA.[16] This code is freely available at https://github.com/rdkit/rdkit/blob/master/Contrib/PBF/pbf.py	3D character

a The number of rotatable bonds is the number of matches to the SMARTS string:
[!$([NH]!@C(=O))&!D1&!$(*#*)]-&!@[!$([NH]!@C(=O))&!D1&!$(*#*)]

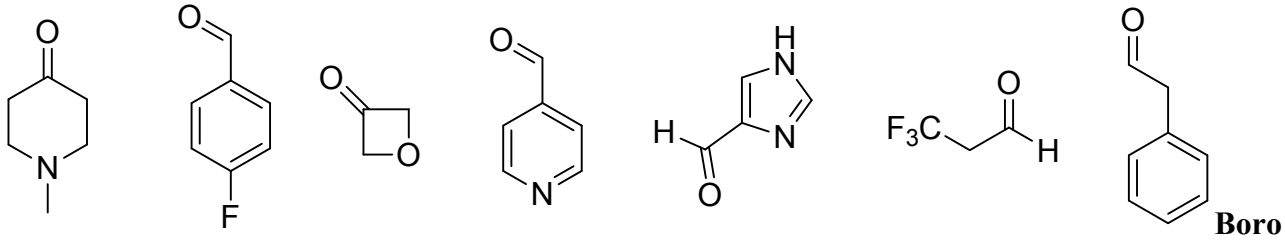
LLAMA uses the OpenBabel[17] package to convert molecules drawn by users into canonical SMILES format, and for a file type conversion in the PMI plot generation code.

3.7. Building Blocks used for diversification

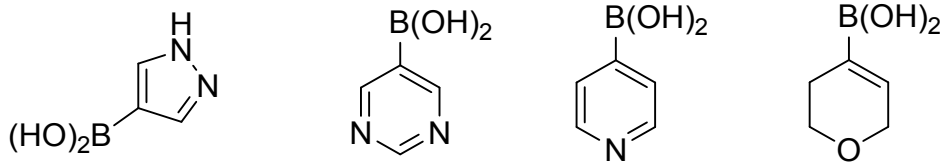
Amines



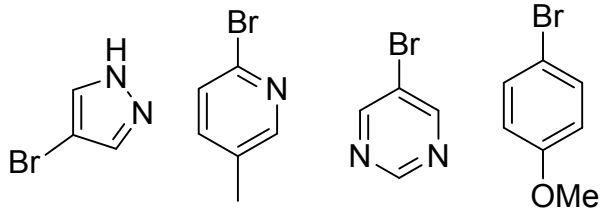
Aldehydes and Ketones



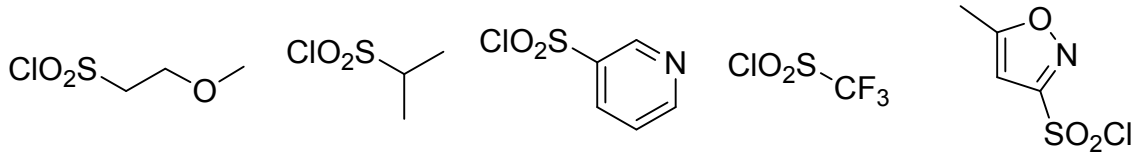
Boronic Acids



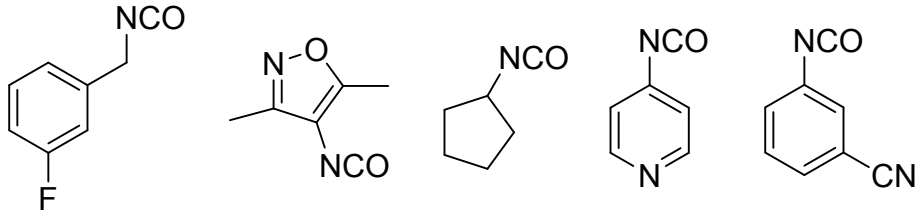
Aryl Bromides



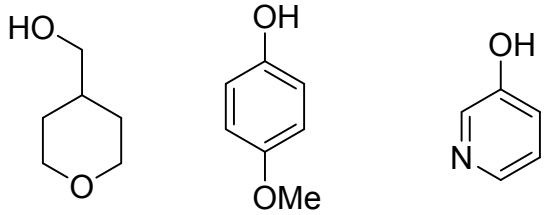
Sulfonyl Chlorides



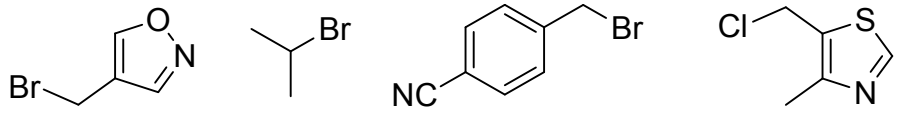
Isocyanates



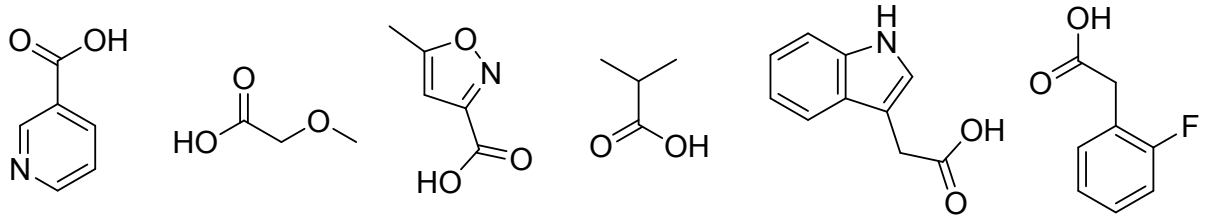
Alcohols



Alkyl Bromides

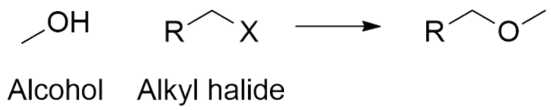


Carboxylic Acids

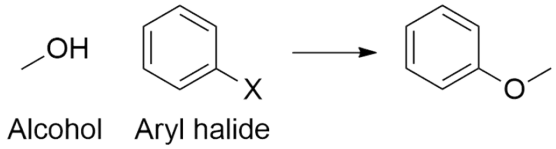


3.8. Diversification reactions

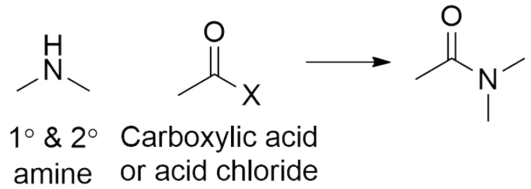
Alcohol Alkylation



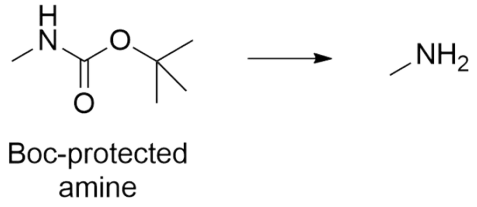
Alcohol Arylation



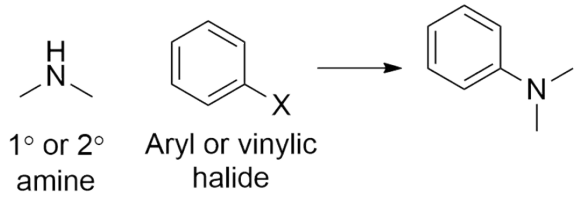
Amide formation



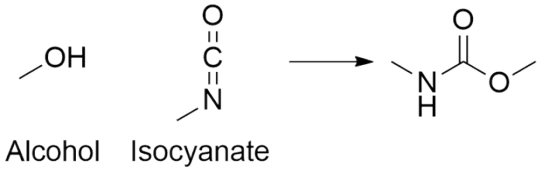
Boc deprotection



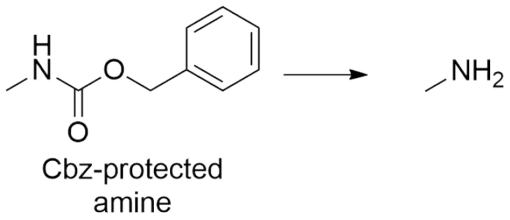
Buchwald-Hartwig Amination



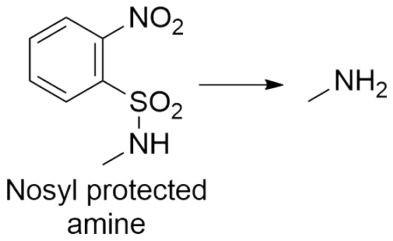
Carbmate formation



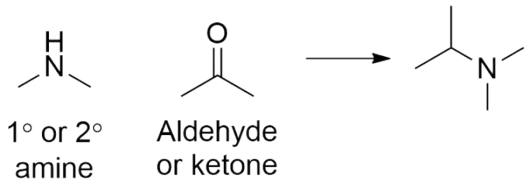
Cbz deprotection



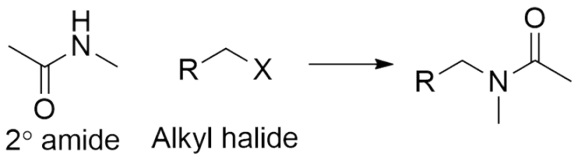
Denosylation



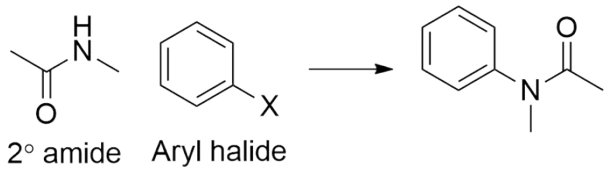
Reductive Amination



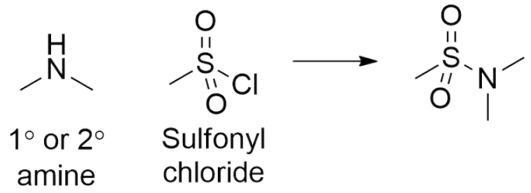
Secondary amide alkylation



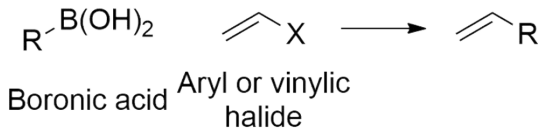
Secondary amide arylation



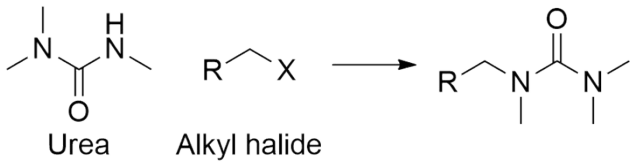
Sulfonamide formation



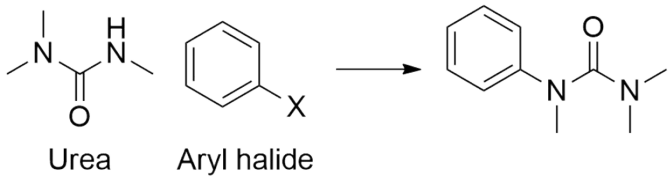
Suzuki coupling



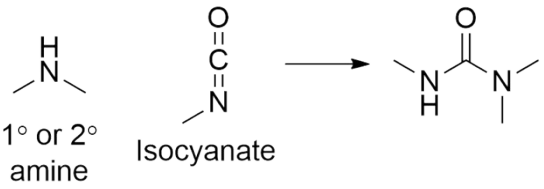
Area alkylation



Urea arylation



Urea Formation



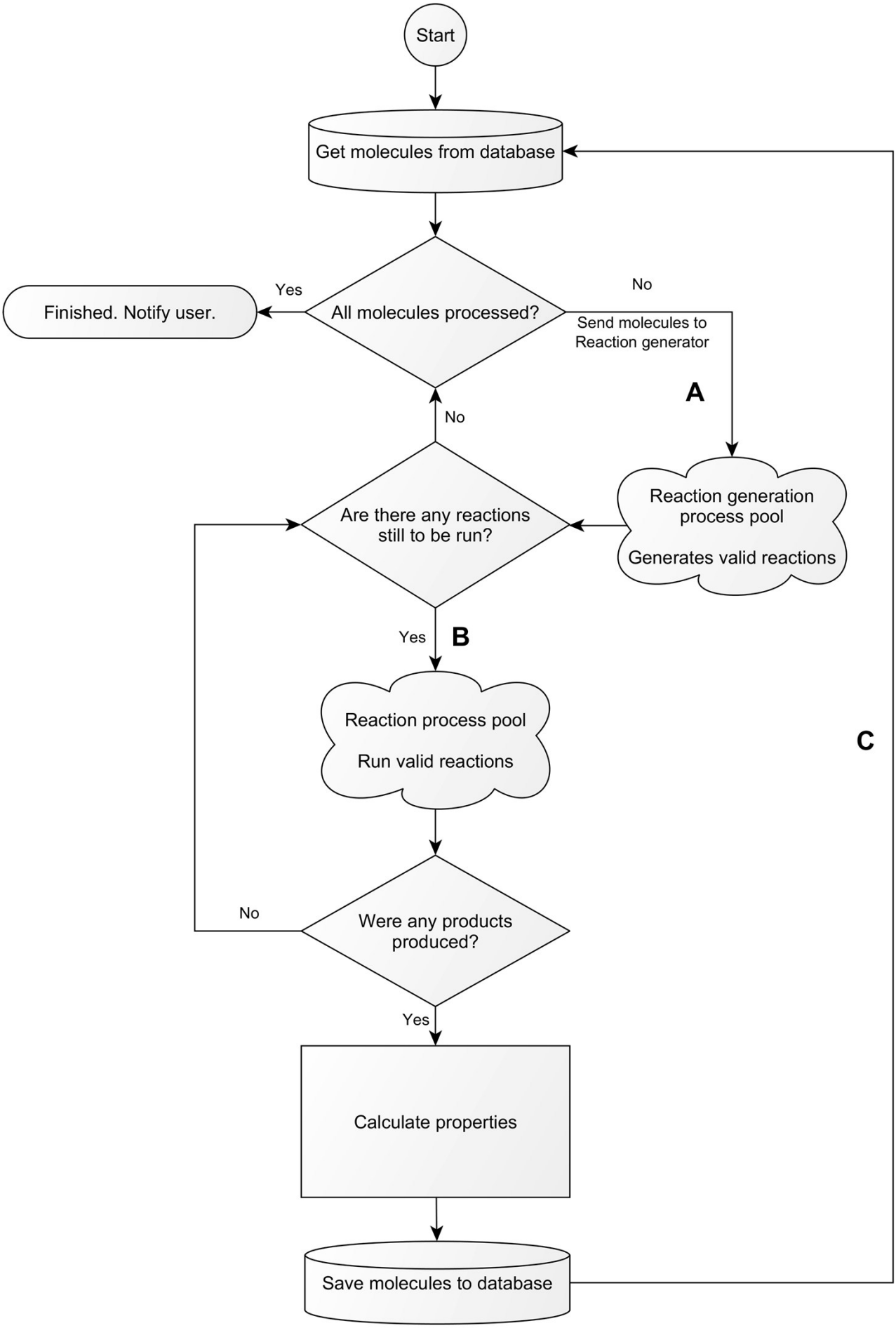
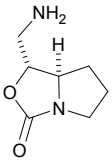


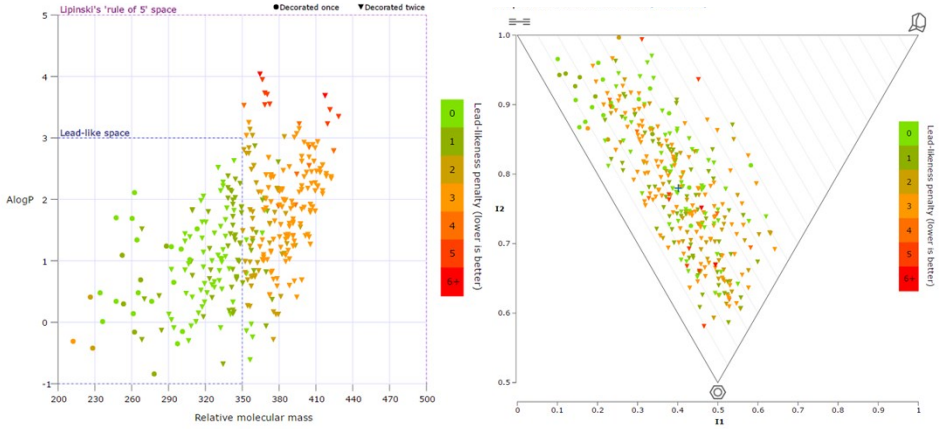
Figure S1. Library decoration workflow.

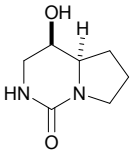
4. Scaffold Analysis

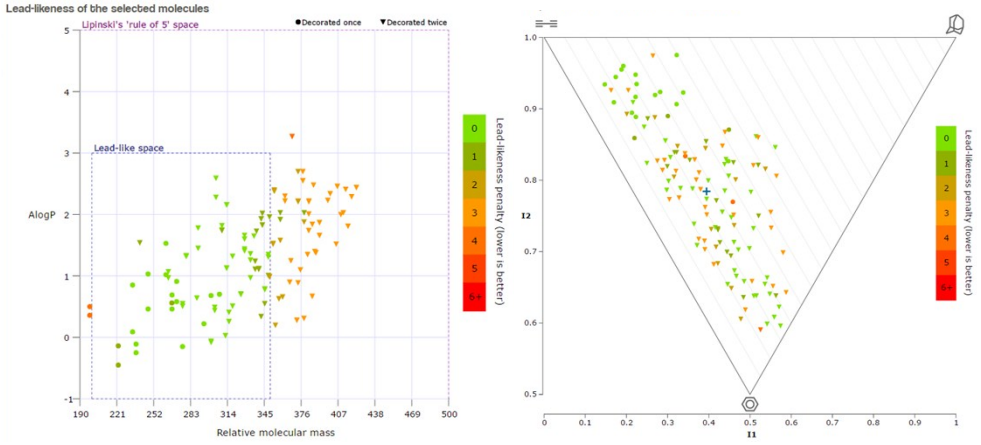
4.1. RMM vs. AlogP and PMI plots for individual scaffolds

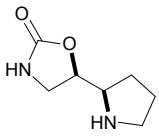
For each scaffold uploaded to LLAMA the RMM vs. AlogP plot and PMI plot is presented below.

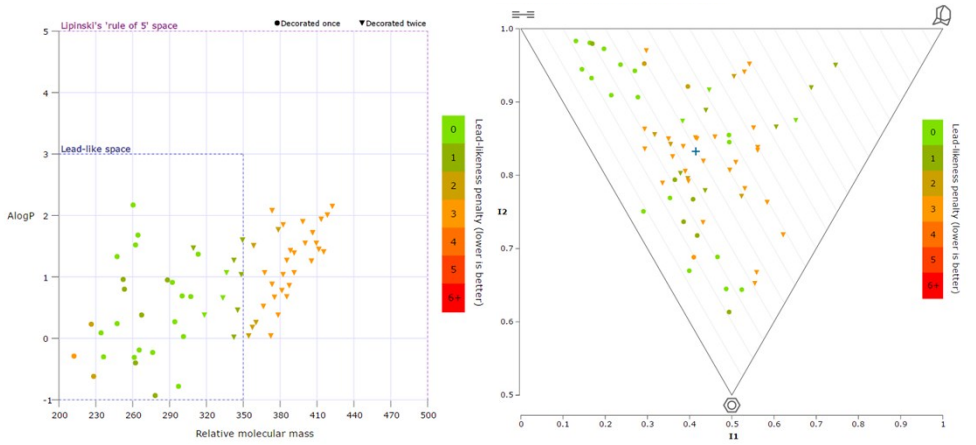


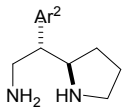


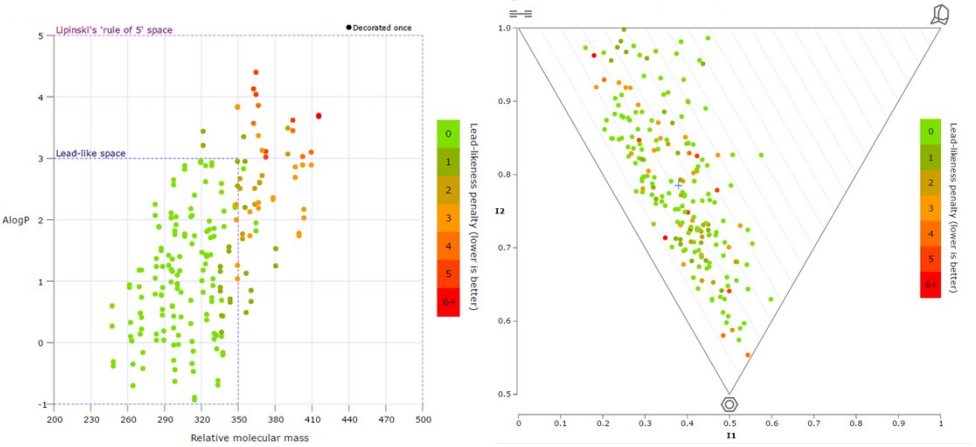


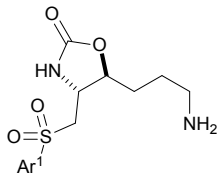


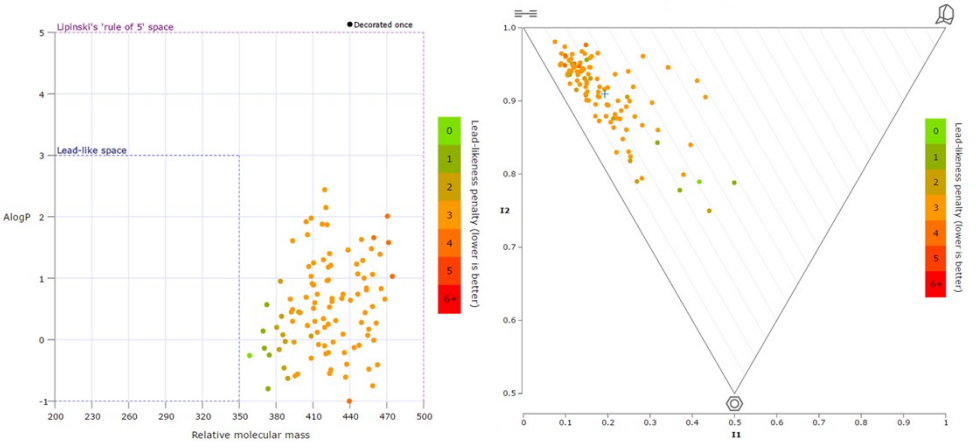


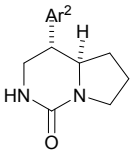

Lipinski's 'rule of 5' space
Decorated once
Decorated twice
Lead-like space
AlogP
Relative molecular mass
Lead-likeness penalty (lower is better)
I1
I2

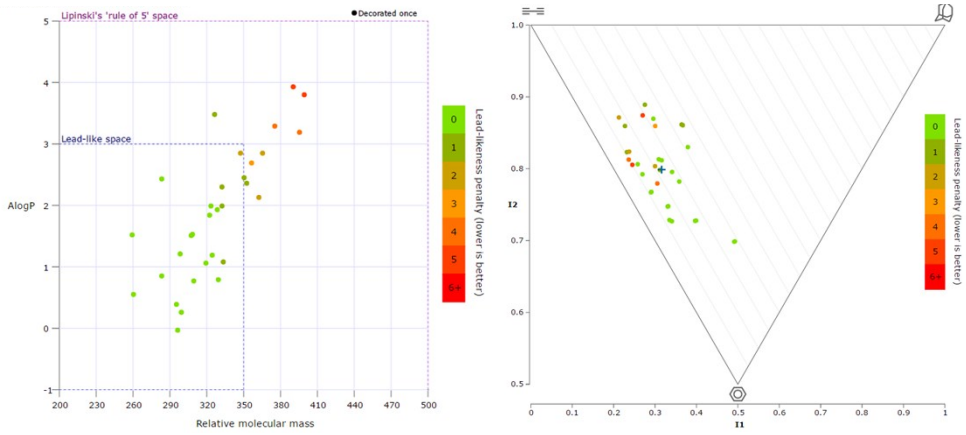

Ar2
NH2
HN


Lipinski's 'rule of 5' space
Decorated once
Lead-like space
AlogP
Relative molecular mass
Lead-likeness penalty (lower is better)
I1
I2

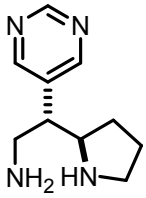

O
O
HN
O
O=S
Ar1
NH2

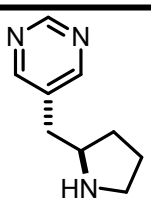
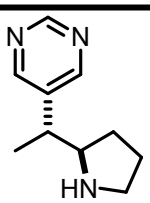

Lipinski's 'rule of 5' space
Decorated once
Lead-like space
AlogP
Relative molecular mass
Lead-likeness penalty (lower is better)
I1
I2

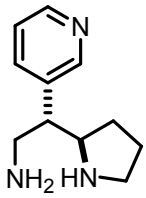

Ar2
H
HN
N
O

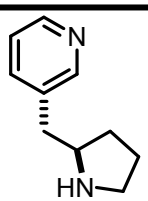
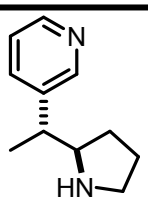


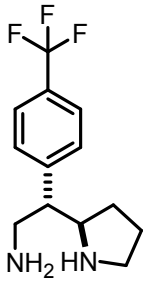
4.2. Novelty analysis

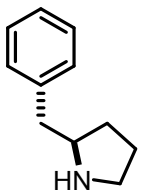
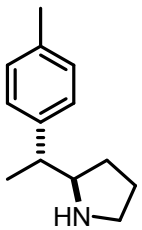


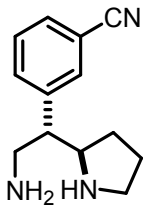
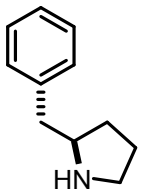
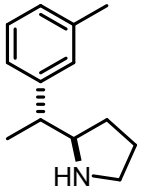
	Murcko	Murcko with α
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

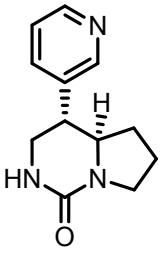
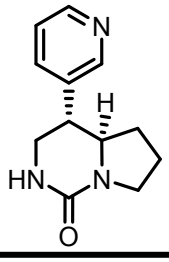
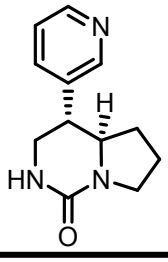


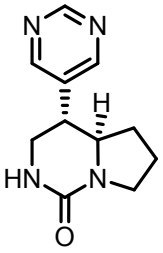
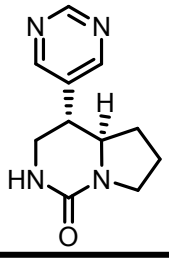
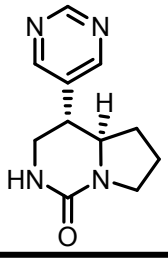
	Murcko	Murcko with α
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

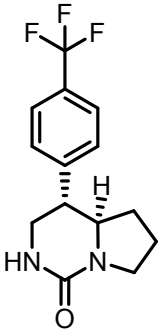
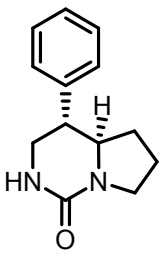
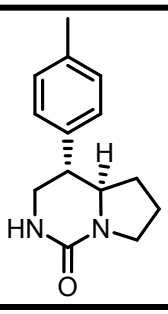


	Murcko	Murcko with α
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	226 (0.14%)	0

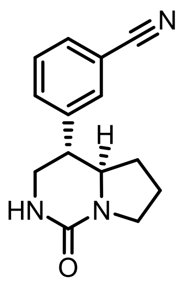
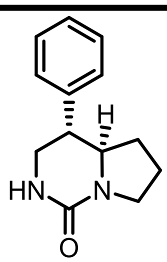
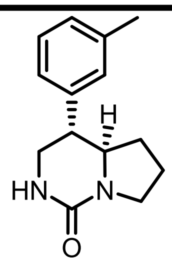
	**Murcko**	**Murcko with α**
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	226 (0.14%)	0

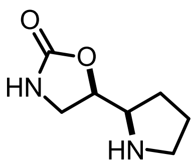
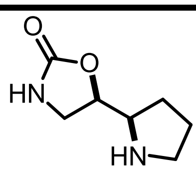
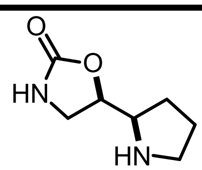
	**Murcko**	**Murcko with α**
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

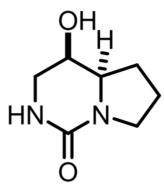
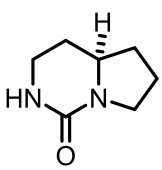
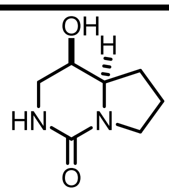
	**Murcko**	**Murcko with α**
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

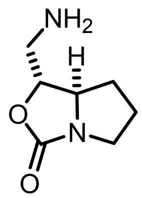
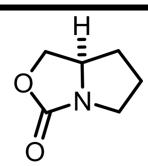
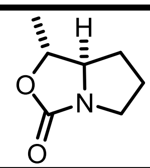
	**Murcko**	**Murcko with α**
		
No. of compounds with same framework in	0	0

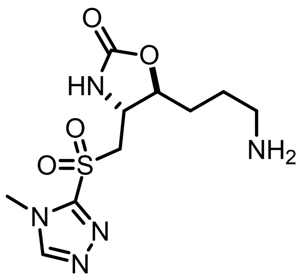
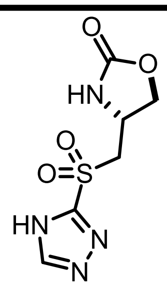
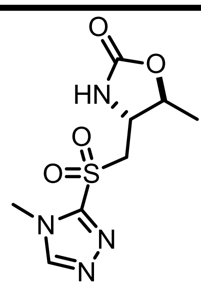
ZINC 2%		
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

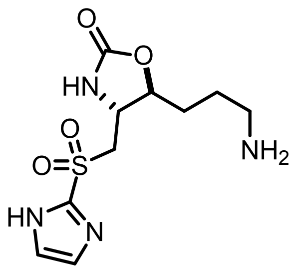
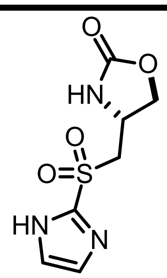
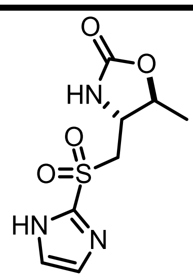
	Murcko	Murcko with α
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

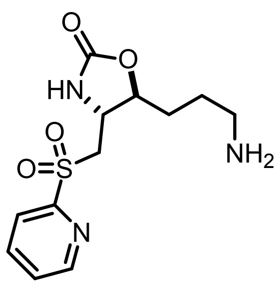
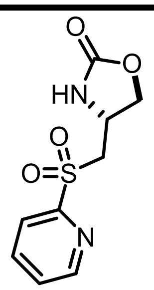
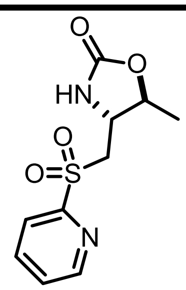
	Murcko	Murcko with α
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

	Murcko	Murcko with α
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

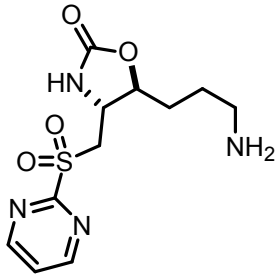
	**Murcko**	**Murcko with α**
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

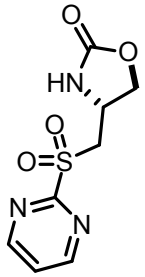
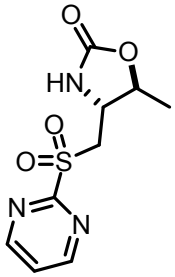
	**Murcko**	**Murcko with α**
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

	**Murcko**	**Murcko with α**
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

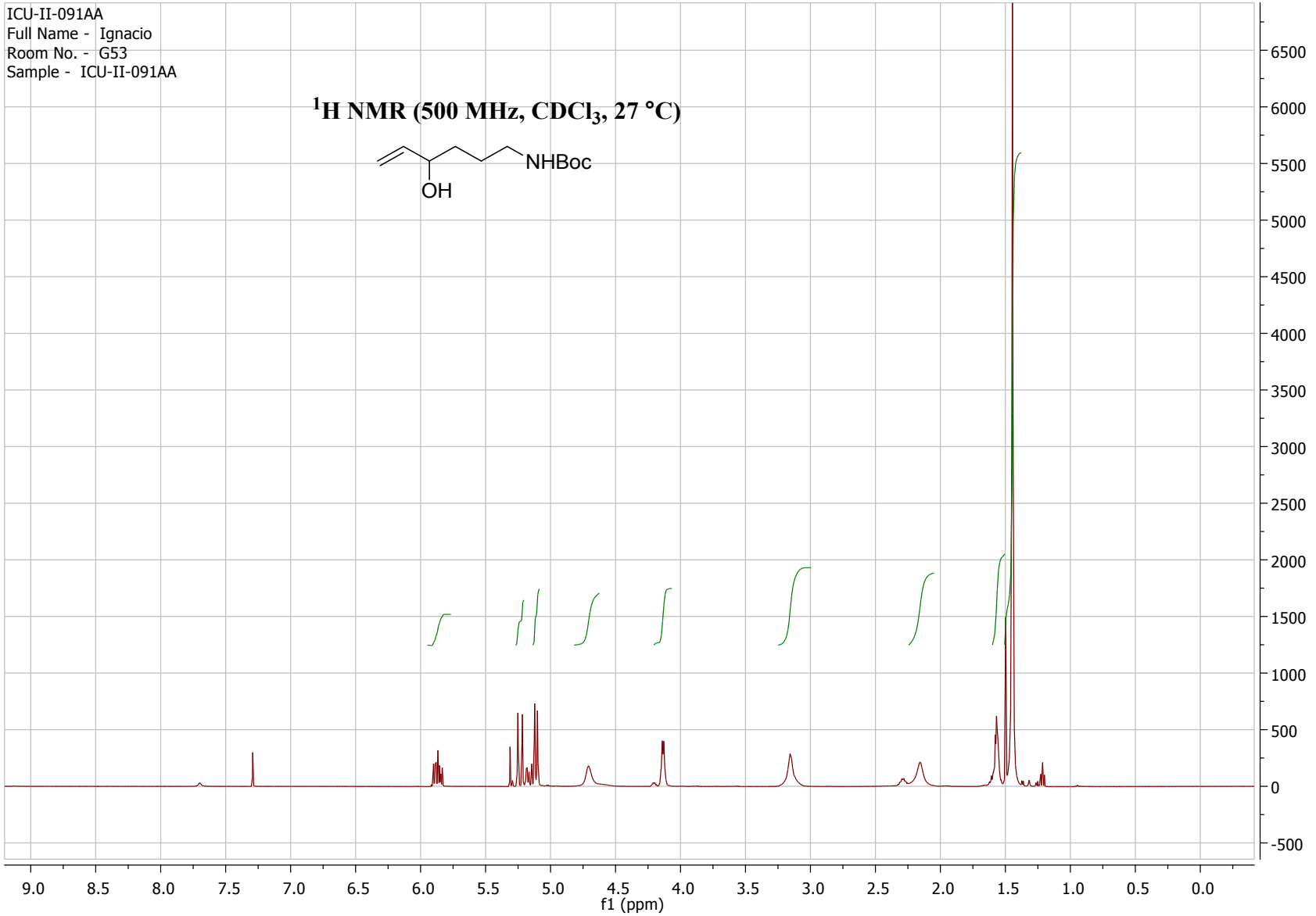
	**Murcko**	**Murcko with α**
		
No. of compounds with same framework in	0	0

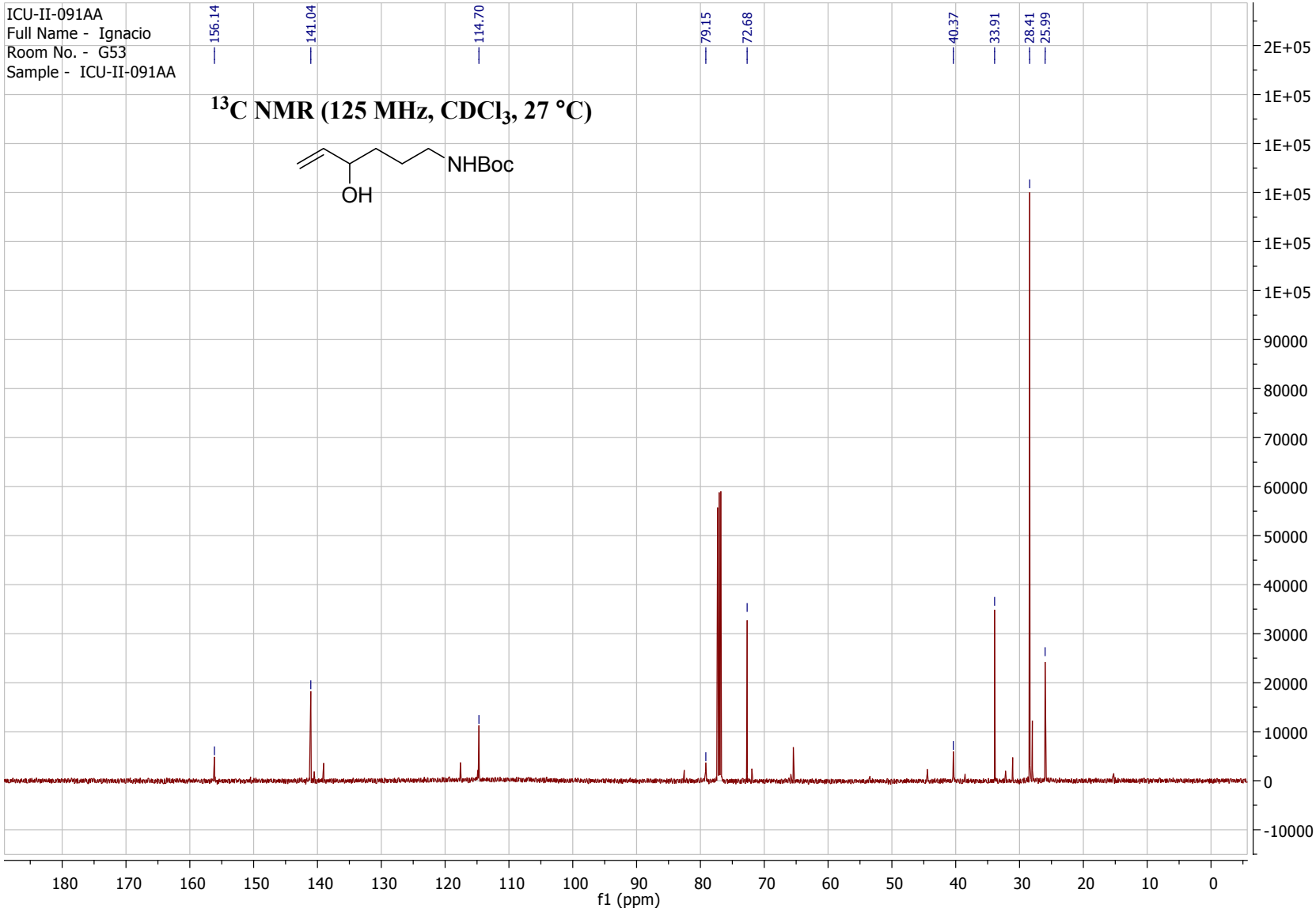
ZINC 2%		
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

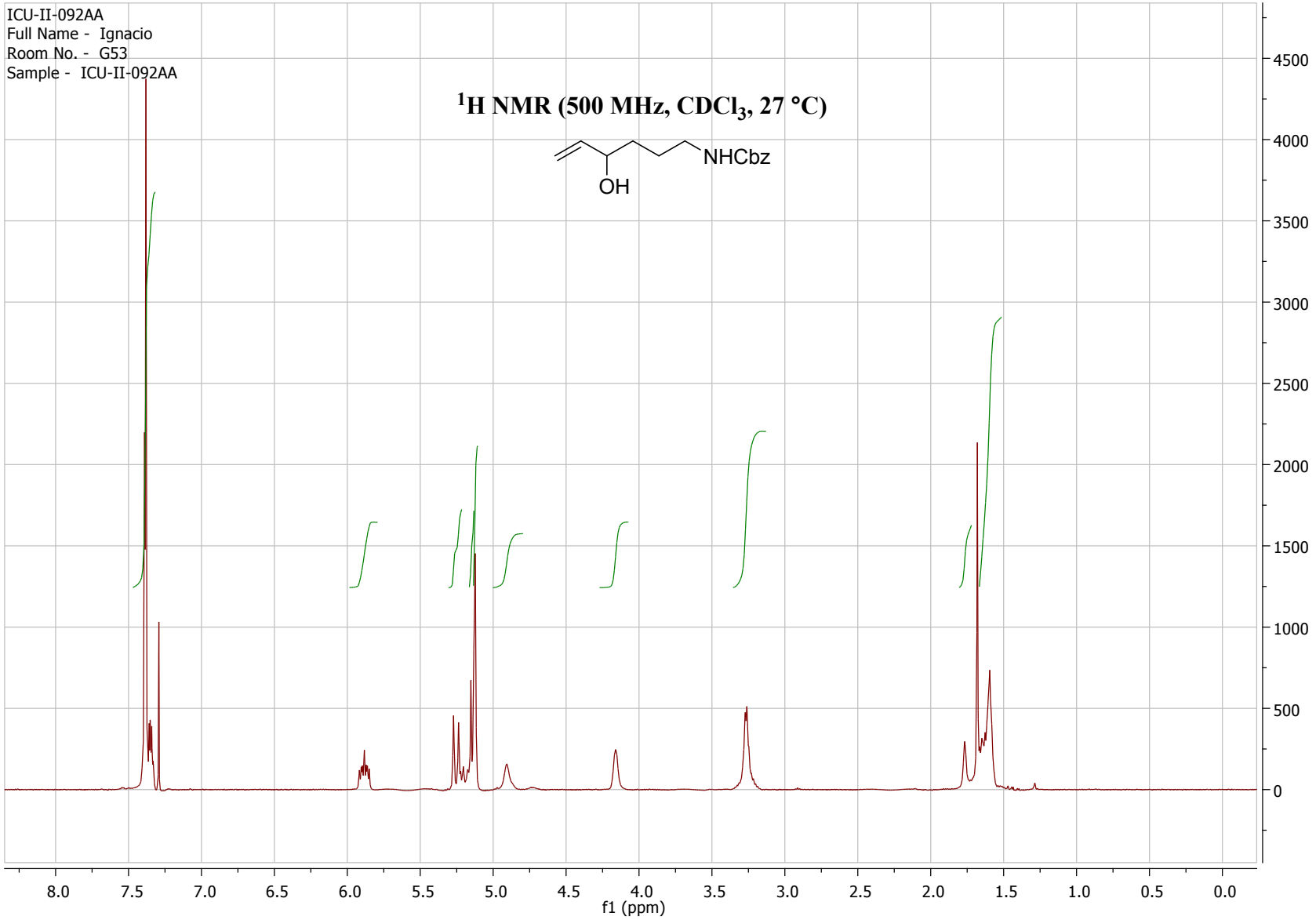


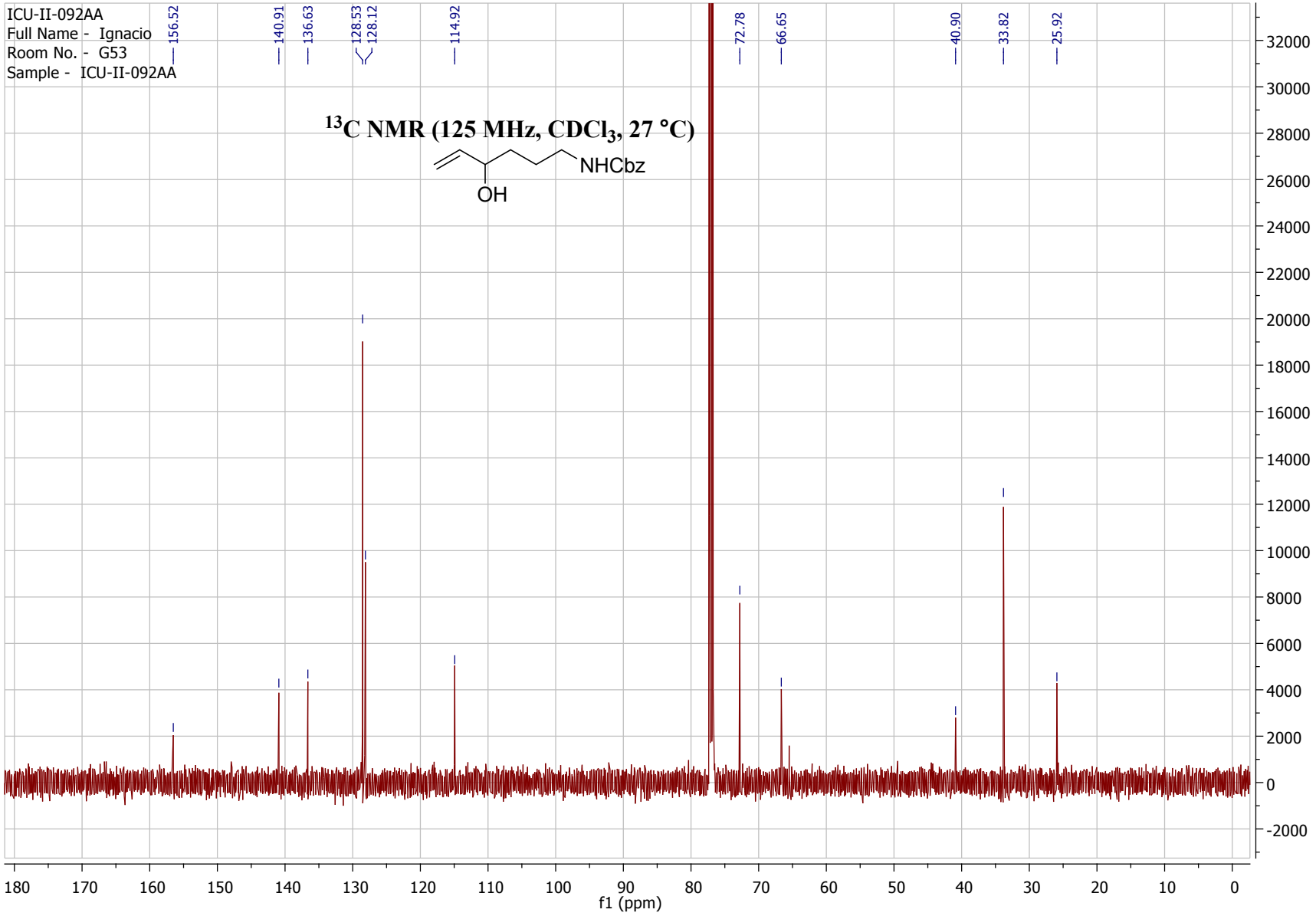
	Murcko	**Murcko with α**
		
No. of compounds with same framework in ZINC 2%	0	0
No. of compounds in ZINC 2% in which the framework is embedded as a substructure	0	0

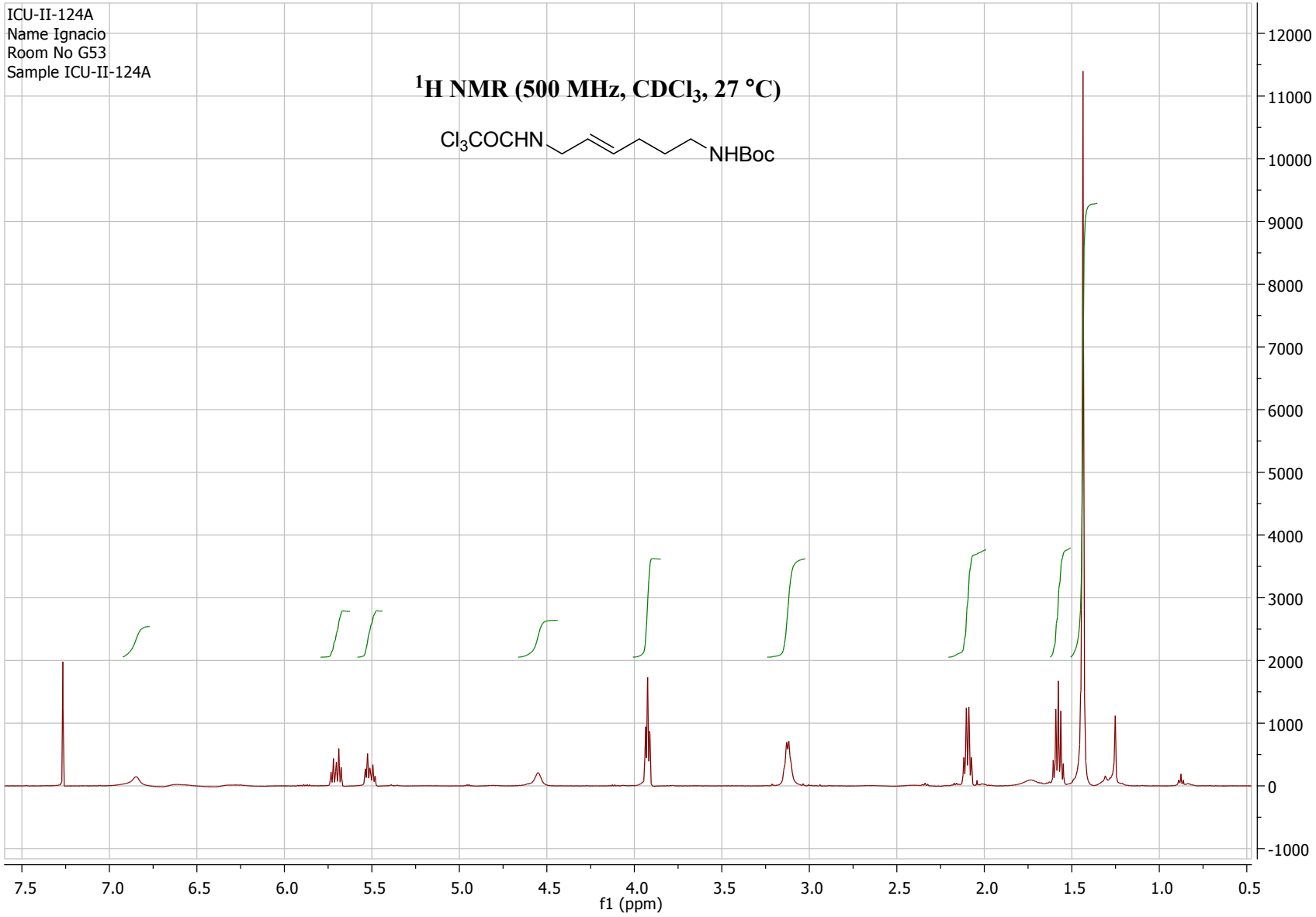
5. Spectra

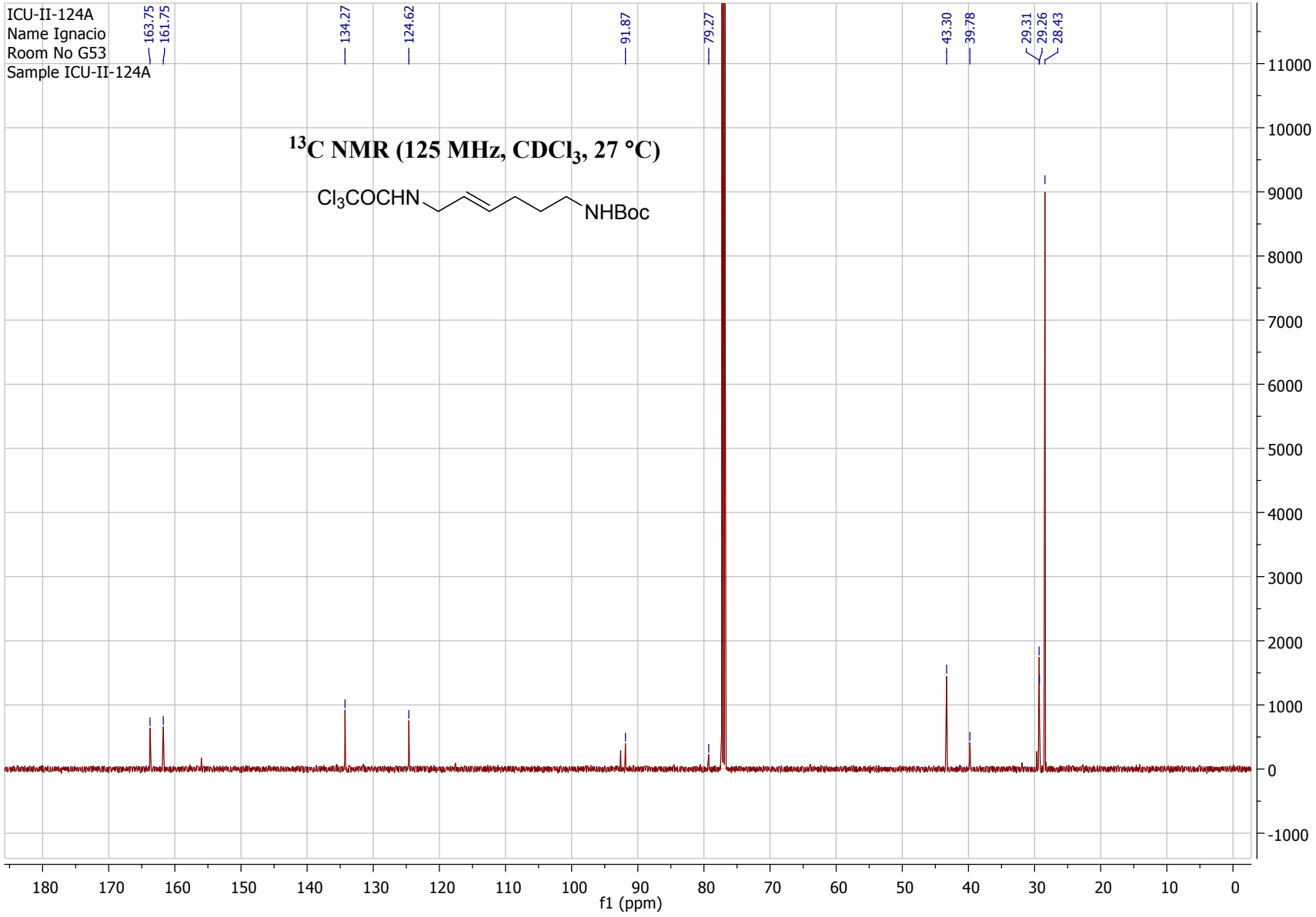

ICU-II-091AA
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-091AA
^{1}H NMR (500 MHz, $CDCl_3$, 27 °C)
NHBoc
OH
f1 (ppm)

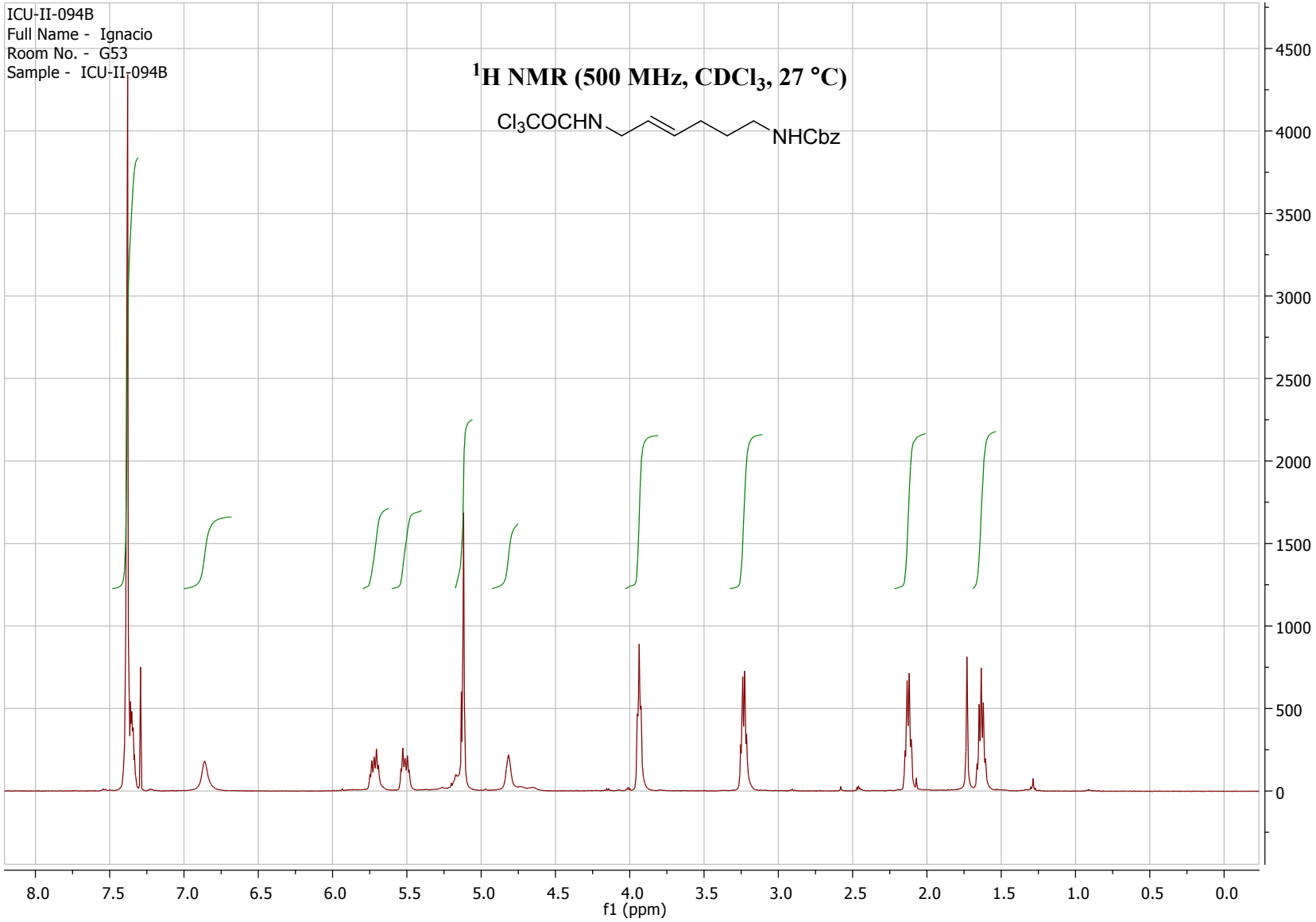

ICU-II-091AA
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-091AA
156.14
141.04
114.70
79.15
72.68
40.37
33.91
28.41
25.99
13C NMR (125 MHz, CDCl3, 27 °C)
NHBoc
OH
2E+05
1E+05
1E+05
1E+05
1E+05
1E+05
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
-10000
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

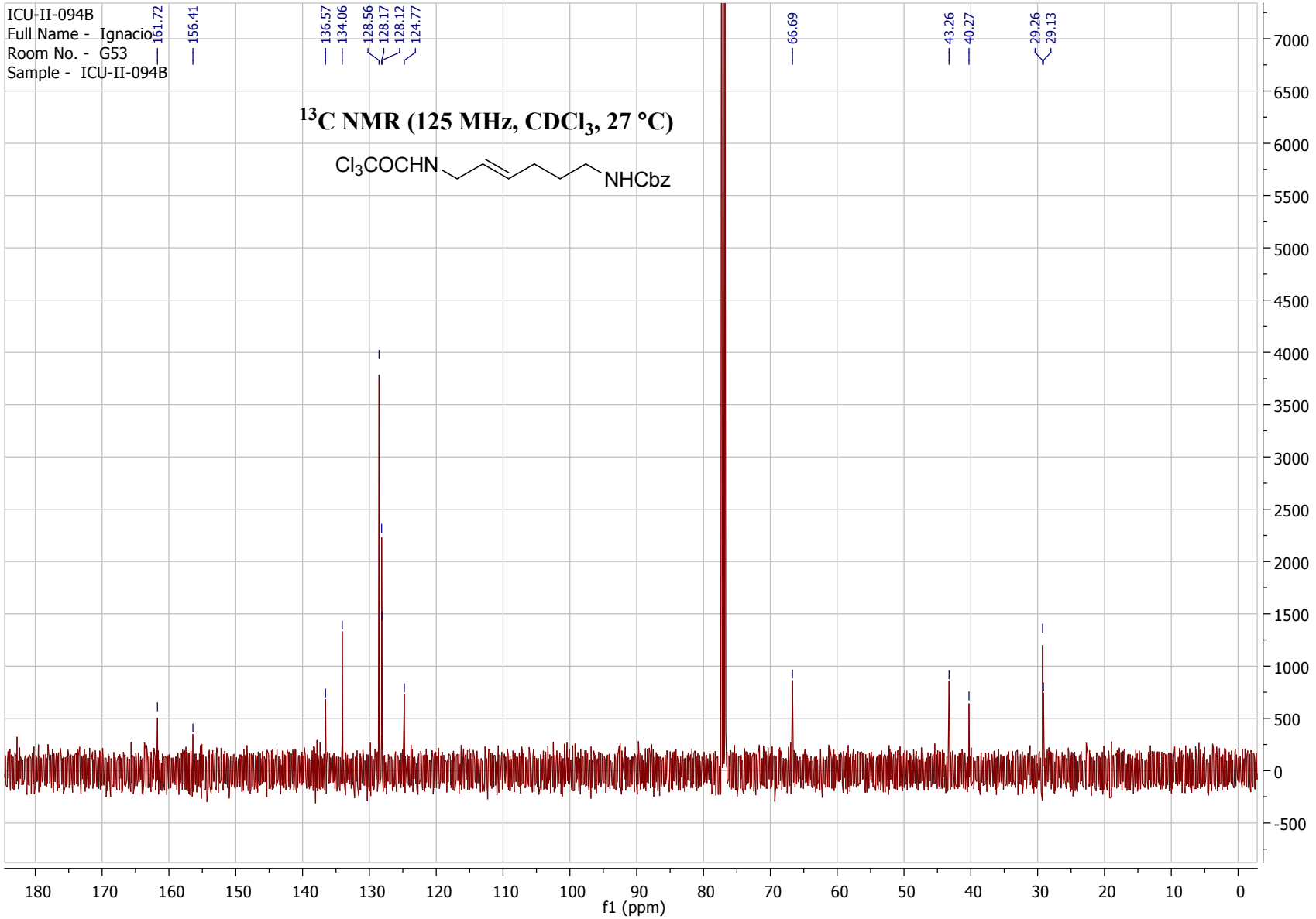

ICU-II-092AA
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-092AA
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
NHCbz
OH
4500
4000
3500
3000
2500
2000
1500
1000
500
0
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

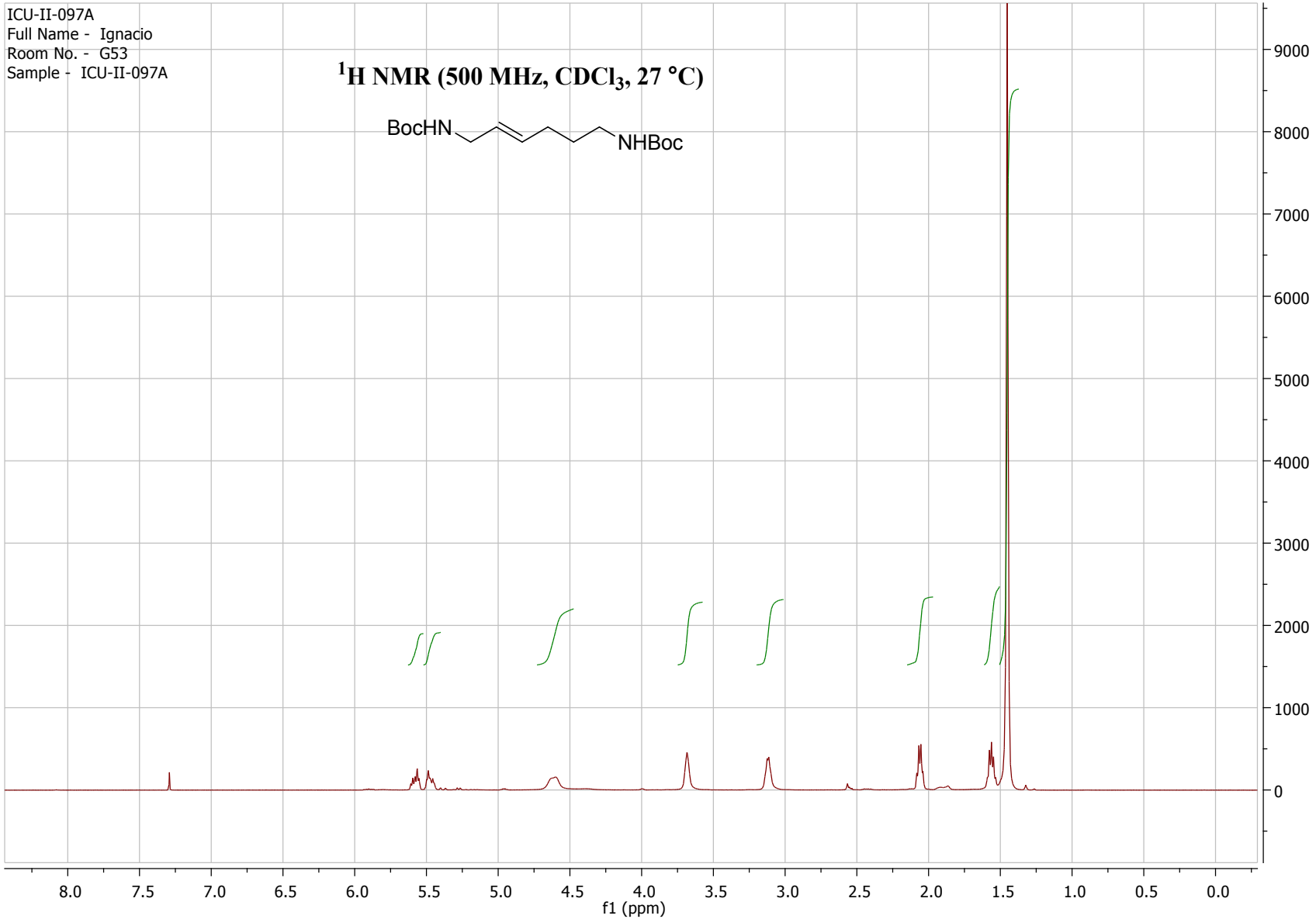

ICU-II-092AA
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-092AA
156.52
140.91
136.63
128.53
128.12
114.92
72.78
66.65
40.90
33.82
25.92
13C NMR (125 MHz, CDCl3, 27 °C)
NHCbz
OH
f1 (ppm)

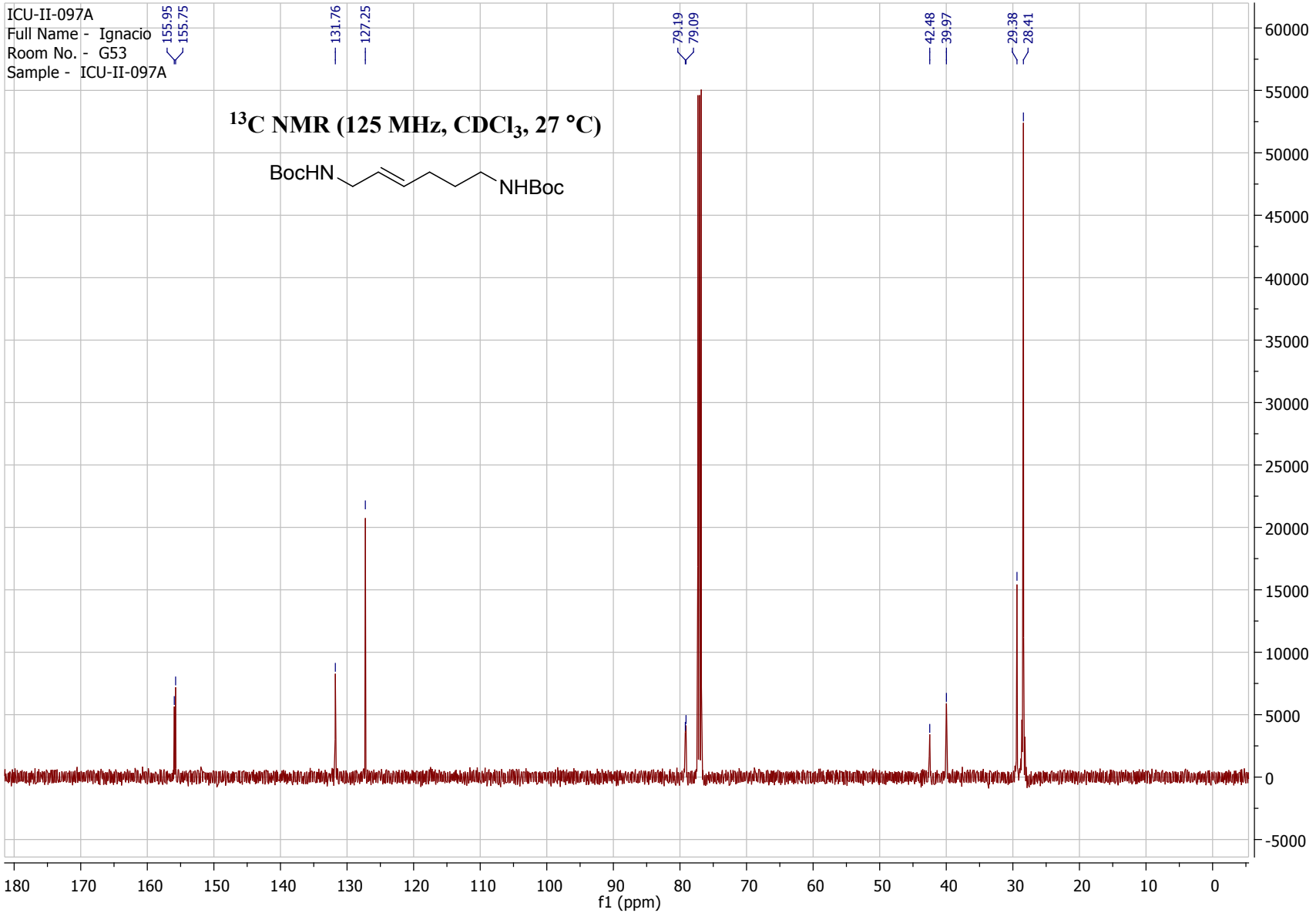

ICU-II-124A
Name Ignacio
Room No G53
Sample ICU-II-124A
^{1}H NMR (500 MHz, $CDCl_3$, 27 °C)
Cl_3COCHN
NHBoc
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)

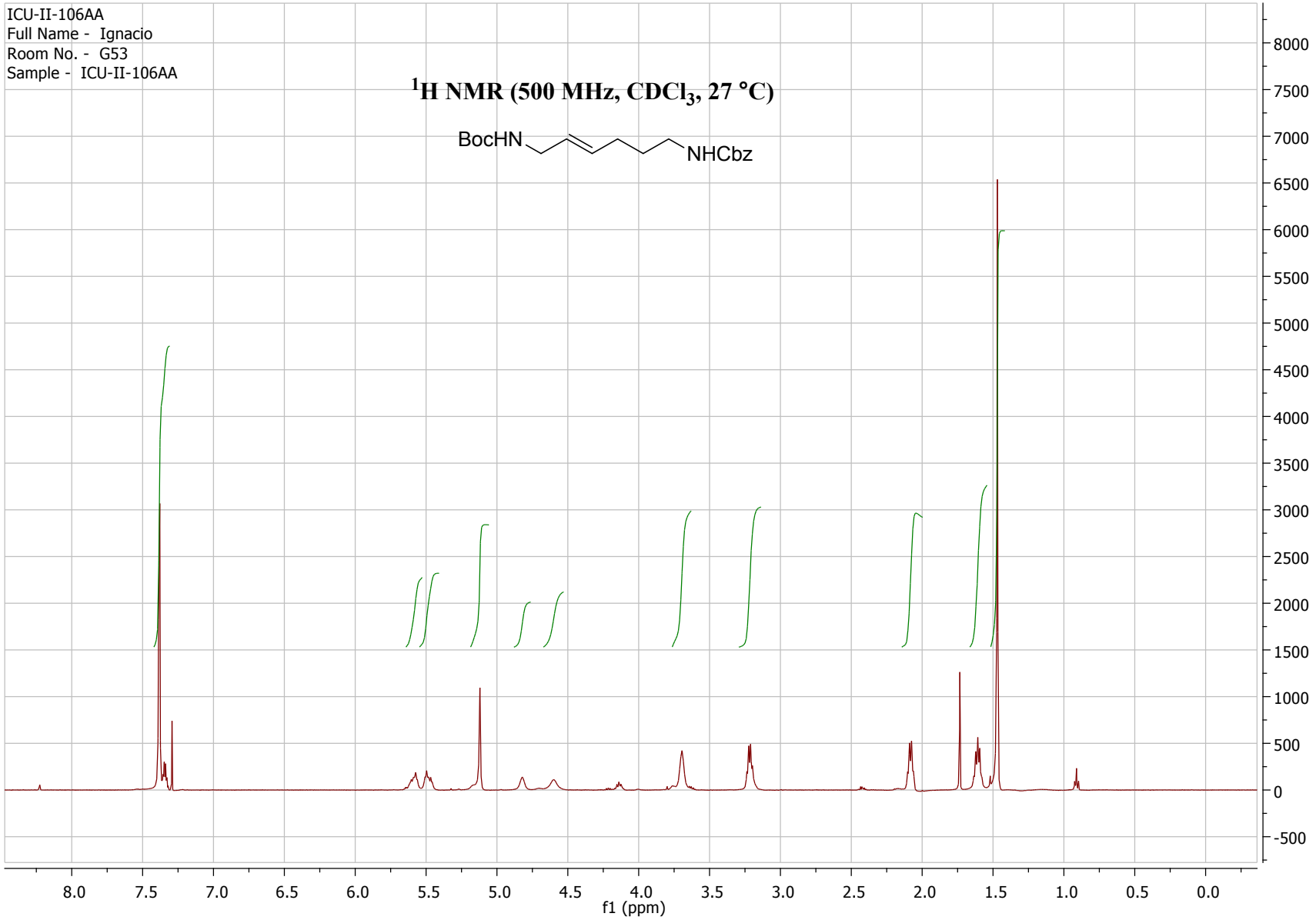

ICU-II-124A
Name Ignacio
Room No G53
Sample ICU-II-124A
163.75
161.75
134.27
124.62
91.87
79.27
43.30
39.78
29.31
29.26
28.43
13C NMR (125 MHz, CDCl3, 27 °C)
Cl3COCHN
NHBoc
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

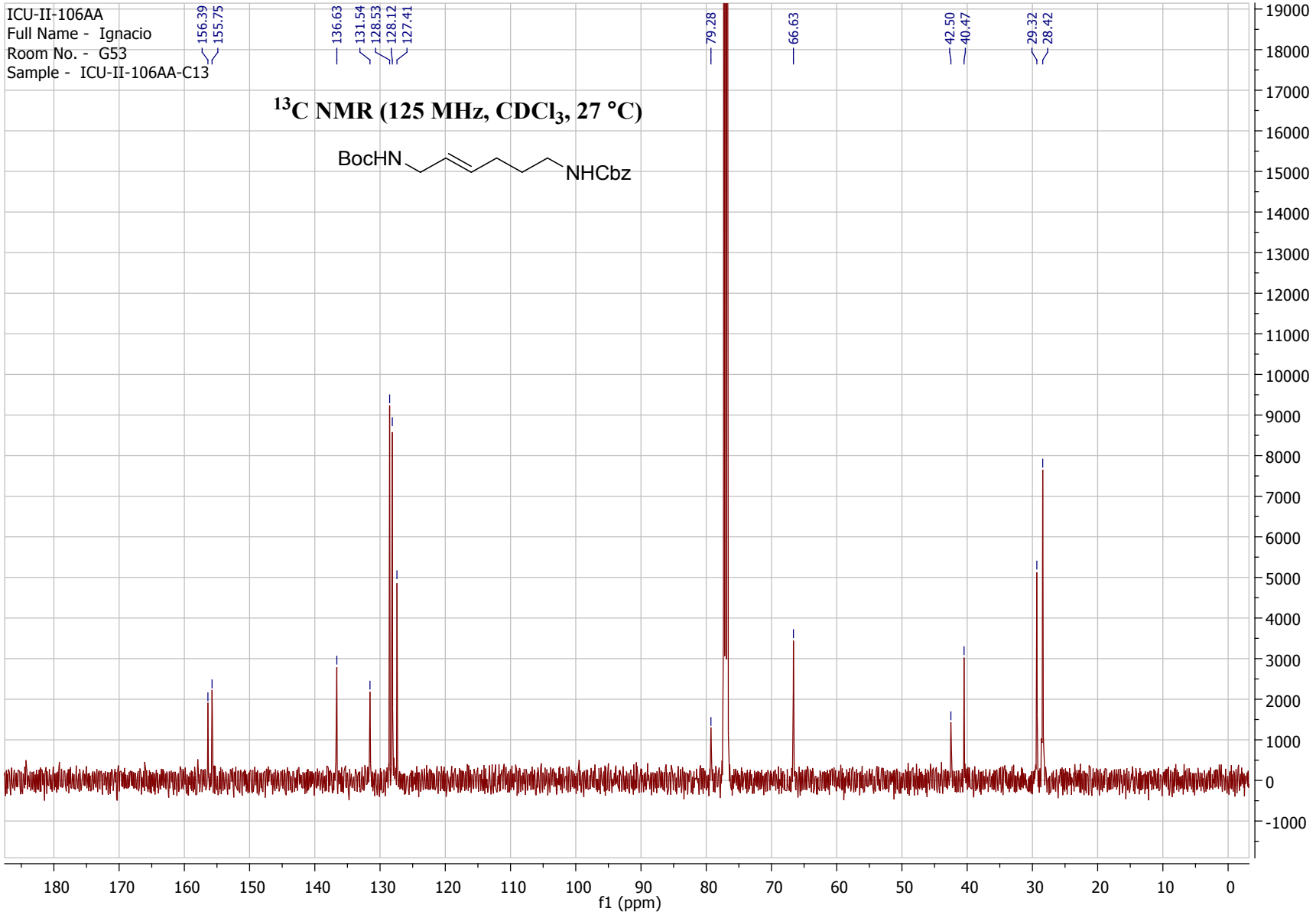

ICU-II-094B
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-094B
^{1}H NMR (500 MHz, $CDCl_3$, 27 °C)
Cl_3COCHN
NHCbz
4500
4000
3500
3000
2500
2000
1500
1000
500
0
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

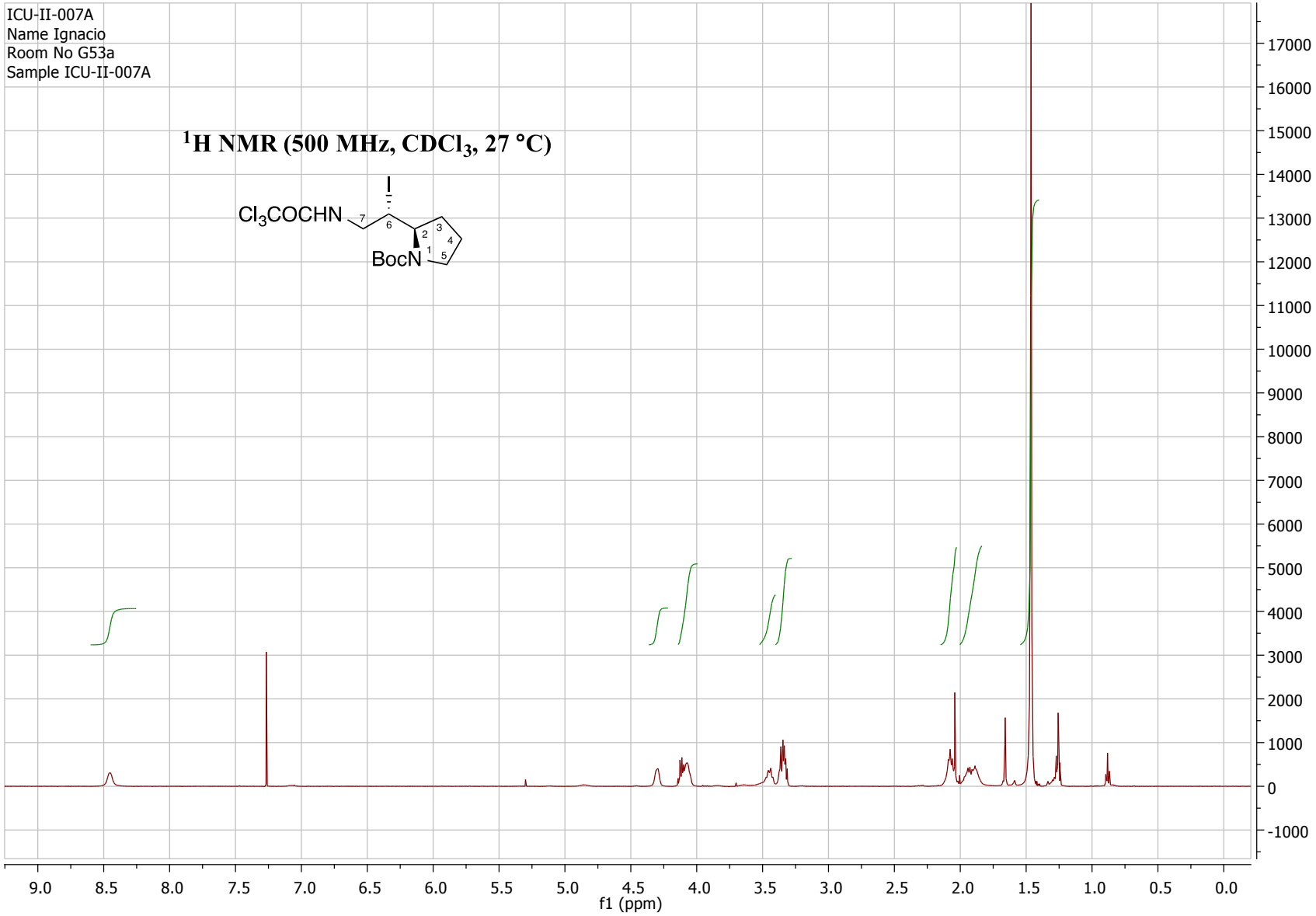

ICU-II-094B
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-094B
161.72
156.41
136.57
134.06
128.56
128.17
128.12
124.77
66.69
43.26
40.27
29.26
29.13
13C NMR (125 MHz, CDCl3, 27 °C)
Cl3COCHN
NHCbz
f1 (ppm)

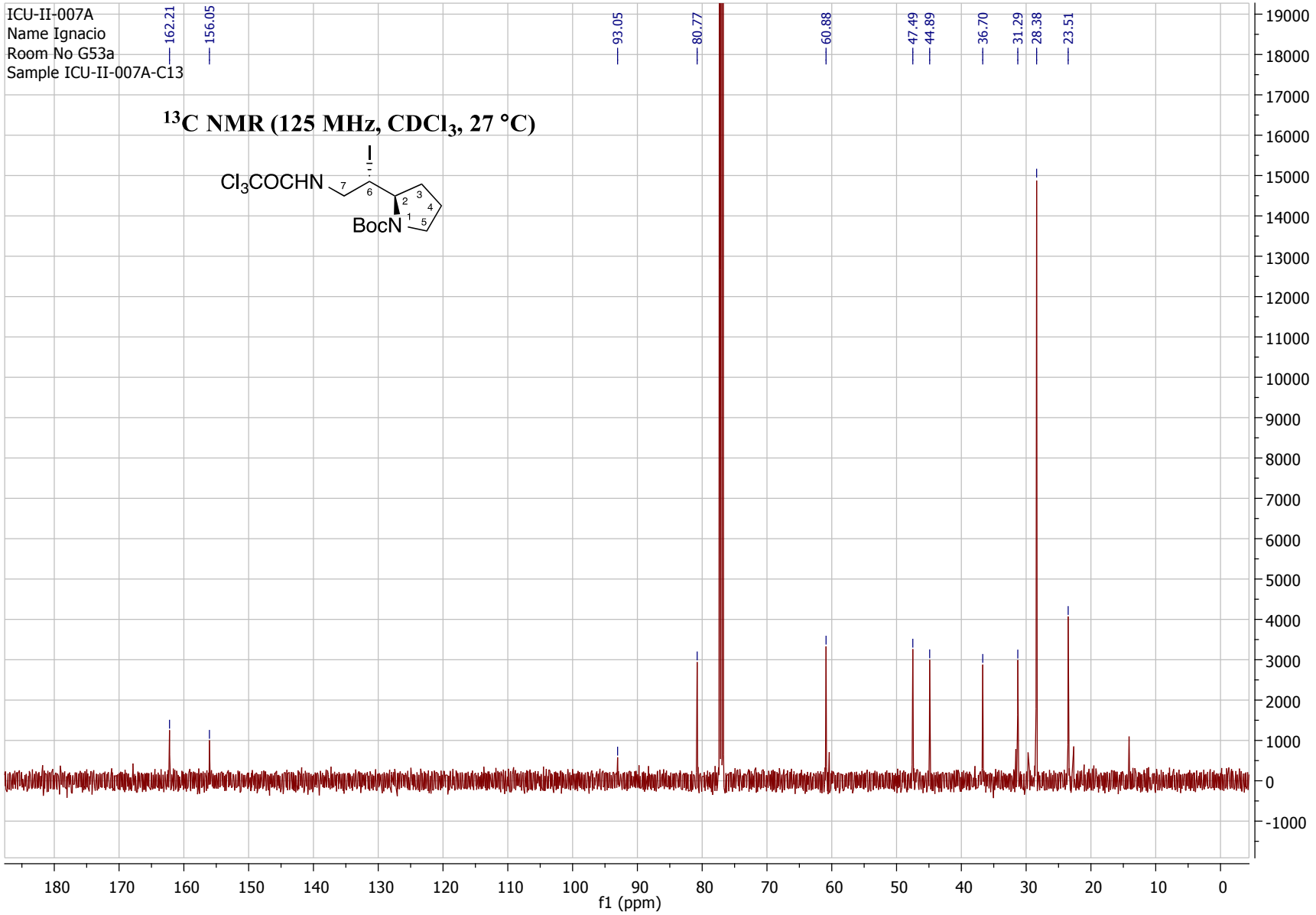

ICU-II-097A
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-097A
^{1}H NMR (500 MHz, $CDCl_3$, 27 °C)
BocHN
NHBoc
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

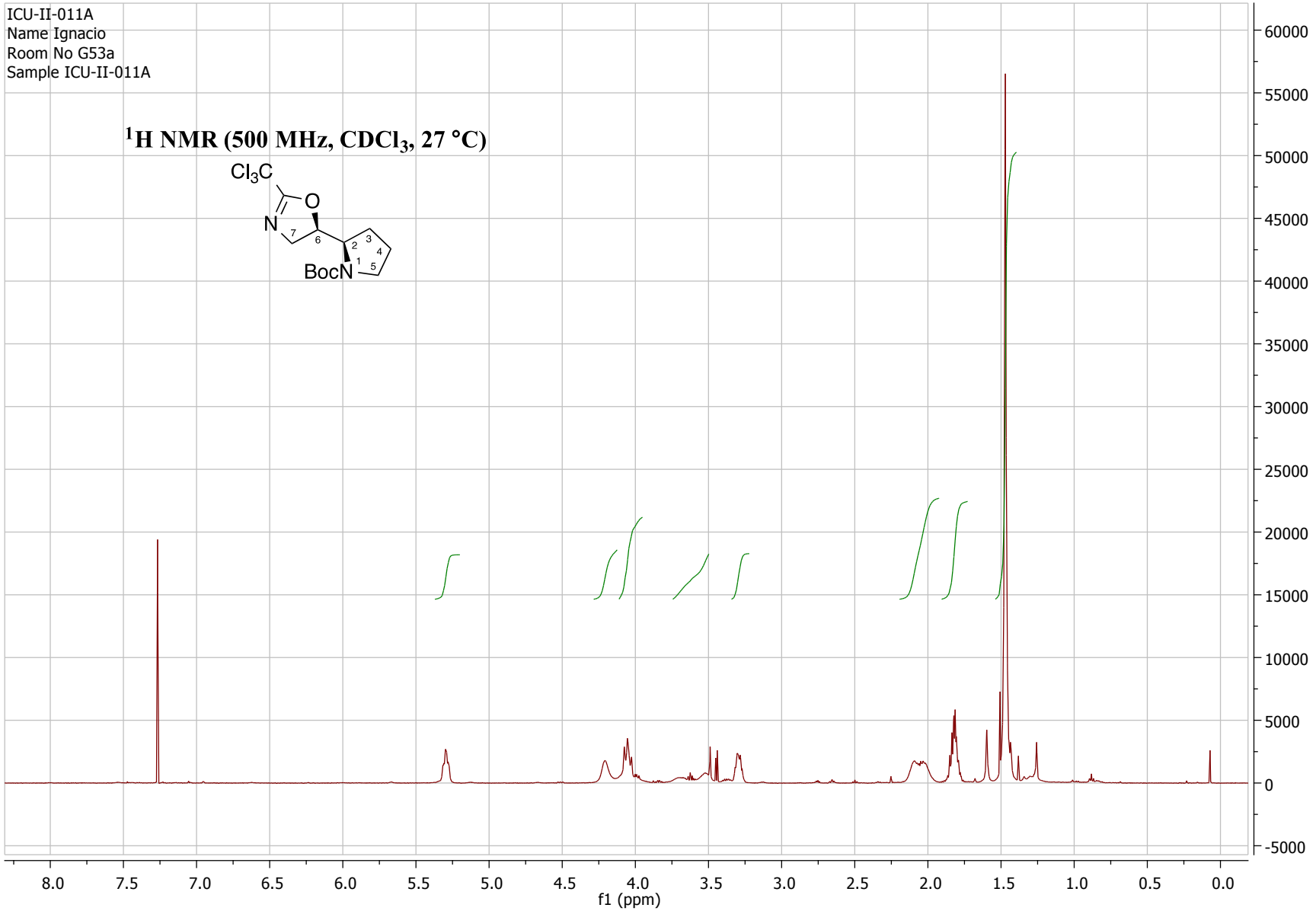

ICU-II-097A
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-097A
155.95
155.75
131.76
127.25
79.19
79.09
42.48
39.97
29.38
28.41
13C NMR (125 MHz, CDCl3, 27 °C)
BocHN
NHBoc
60000
55000
50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
-5000
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

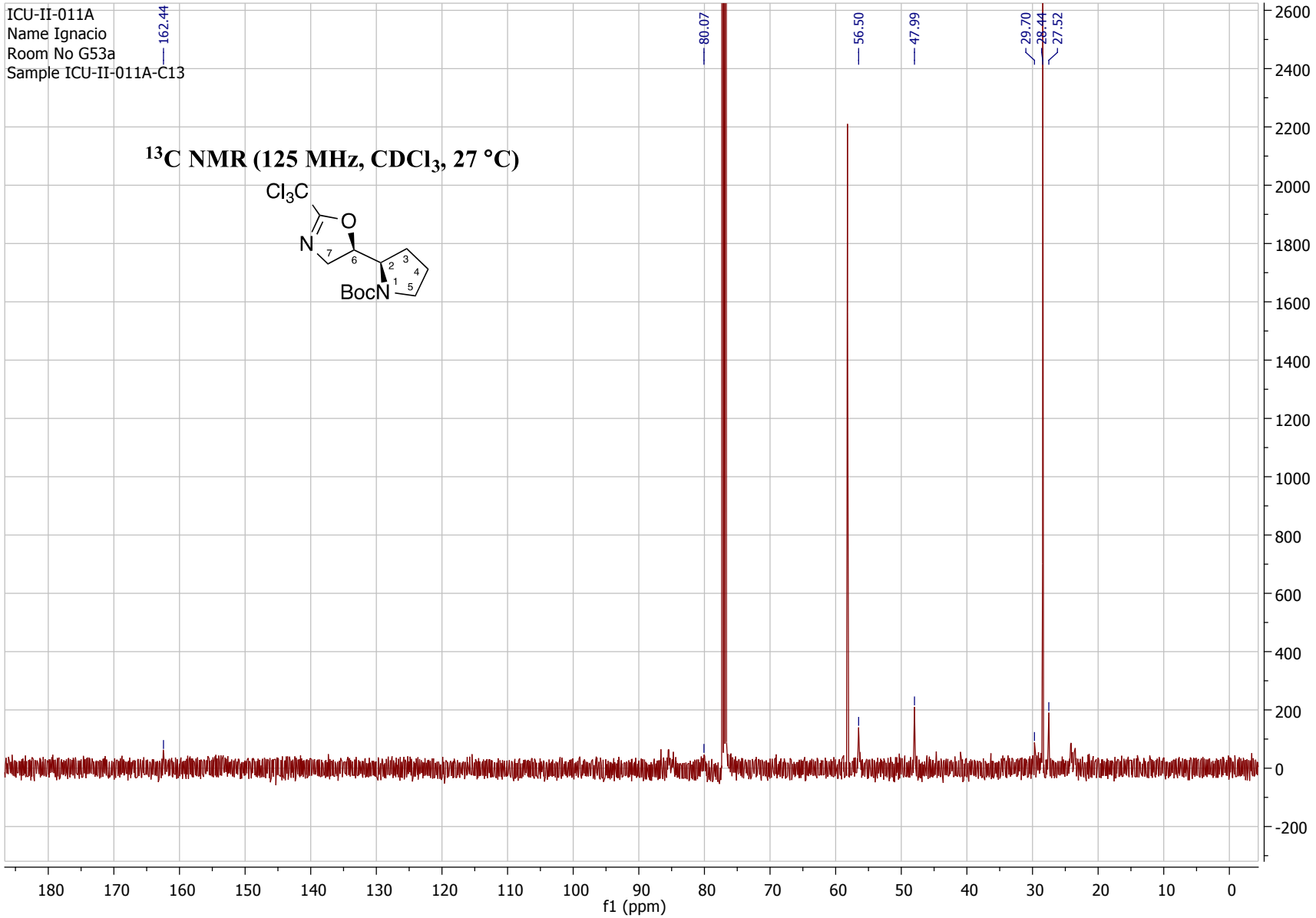

ICU-II-106AA
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-106AA
^{1}H NMR (500 MHz, $CDCl_3$, 27 °C)
BocHN
NHCbz
f1 (ppm)

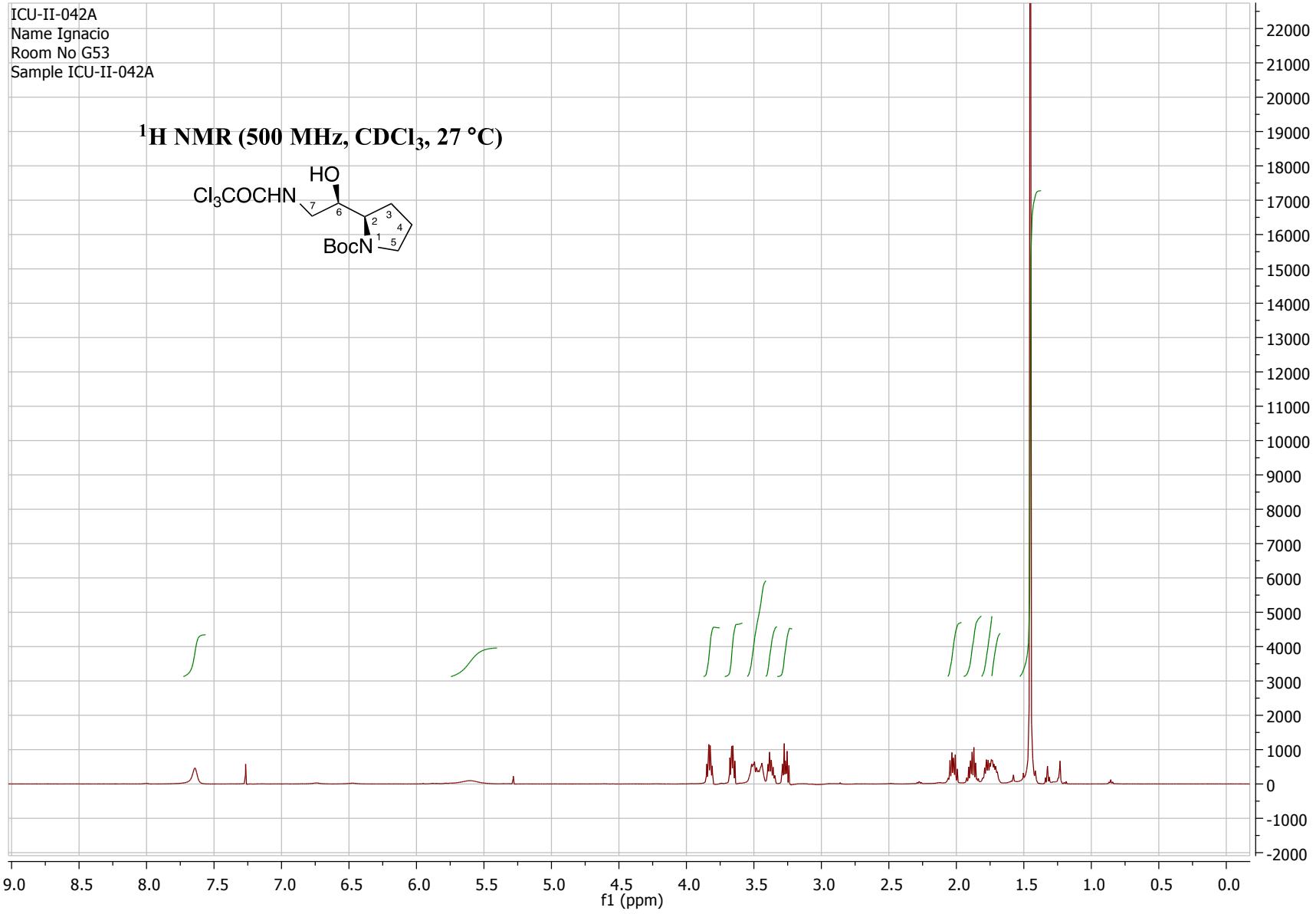

ICU-II-106AA
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-106AA-C13
156.39
155.75
136.63
131.54
128.53
128.12
127.41
79.28
66.63
42.50
40.47
29.32
28.42
13C NMR (125 MHz, CDCl3, 27 °C)
BocHN
NHCbz
f1 (ppm)

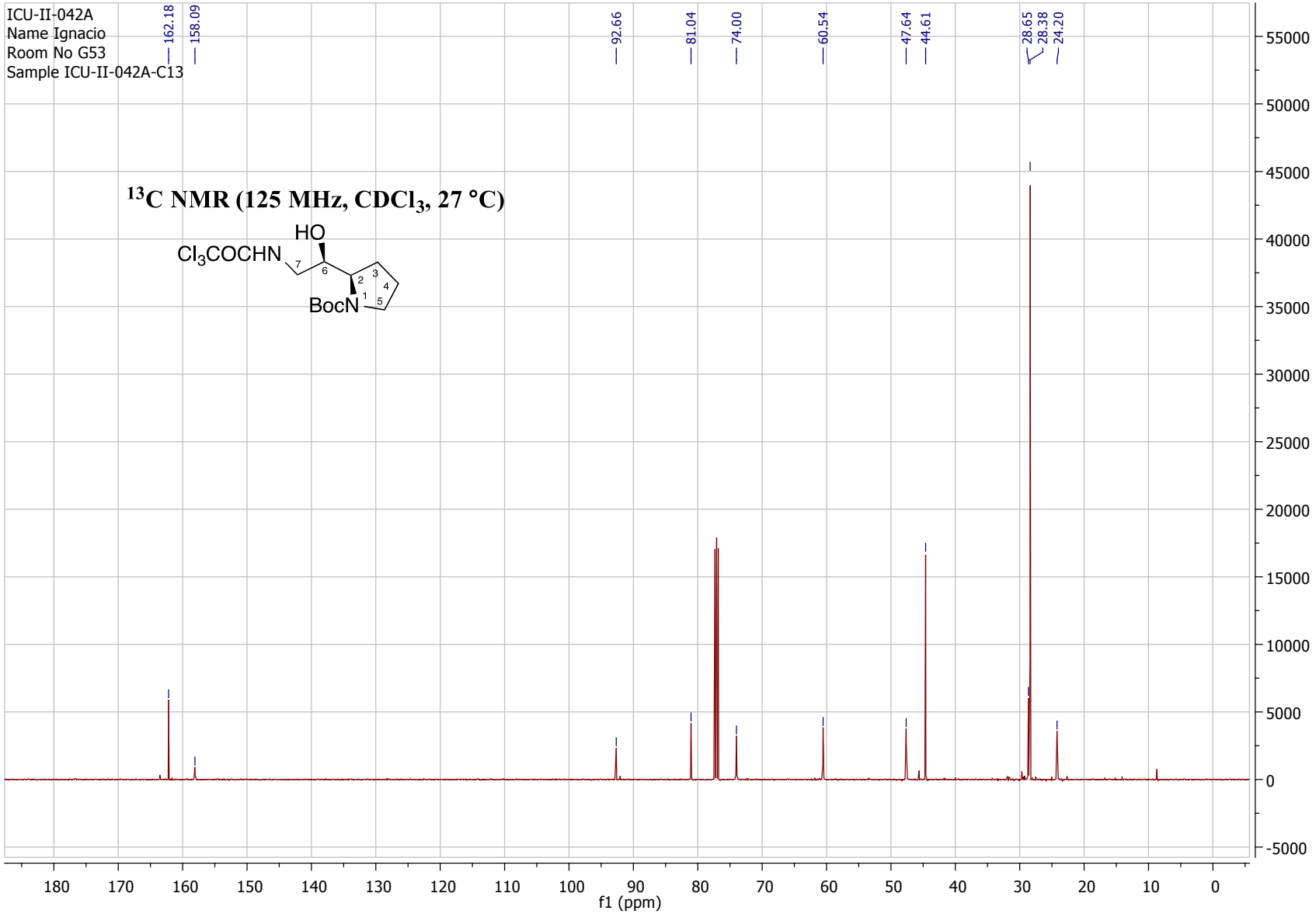

ICU-II-007A
Name Ignacio
Room No G53a
Sample ICU-II-007A
^{1}H NMR (500 MHz, $CDCl_3$, 27 °C)
Cl_3COCHN
BocN
f1 (ppm)

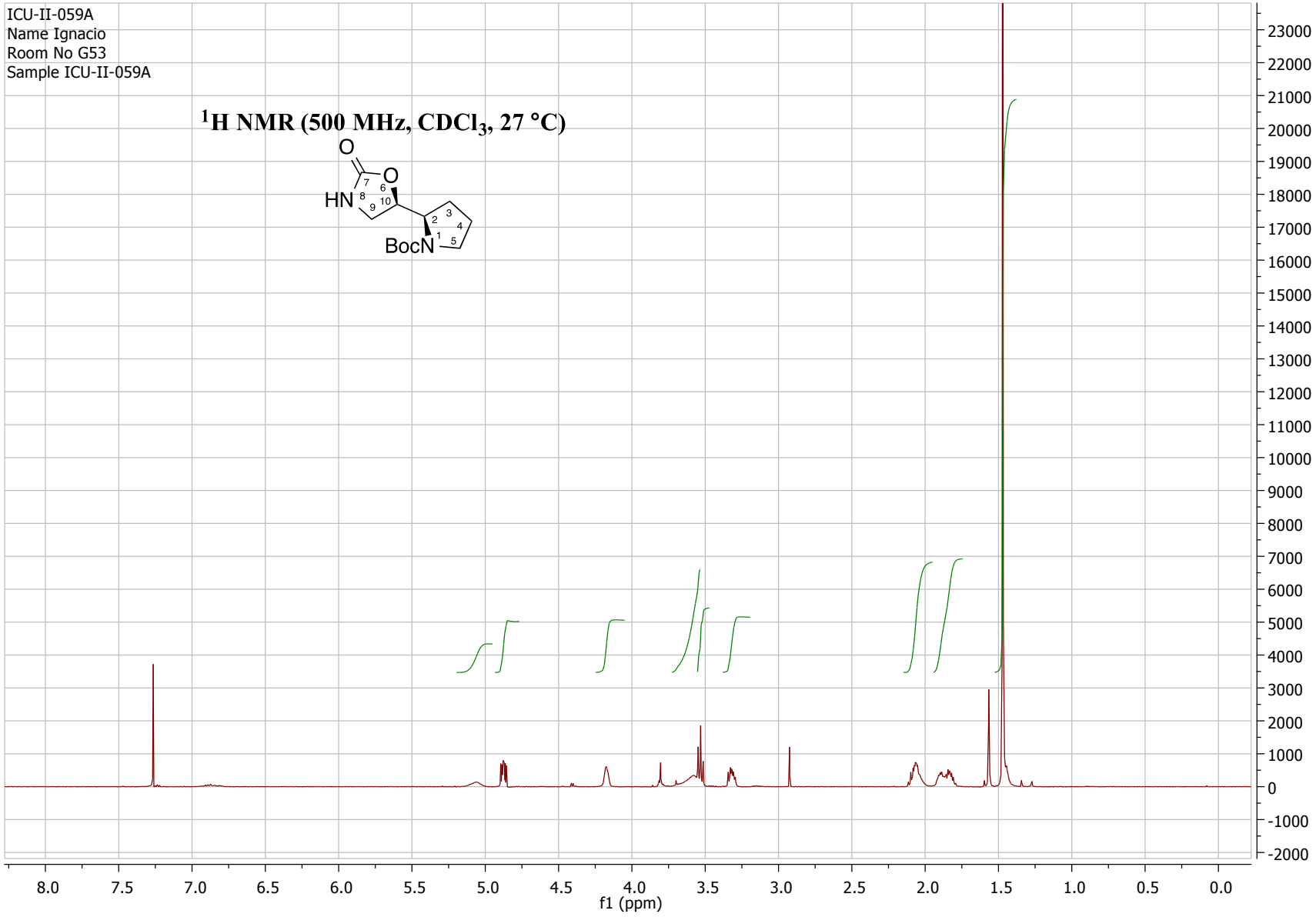

ICU-II-007A
Name Ignacio
Room No G53a
Sample ICU-II-007A-C13
162.21
156.05
93.05
80.77
60.88
47.49
44.89
36.70
31.29
28.38
23.51
13C NMR (125 MHz, CDCl3, 27 °C)
Cl3COCHN
BocN
f1 (ppm)

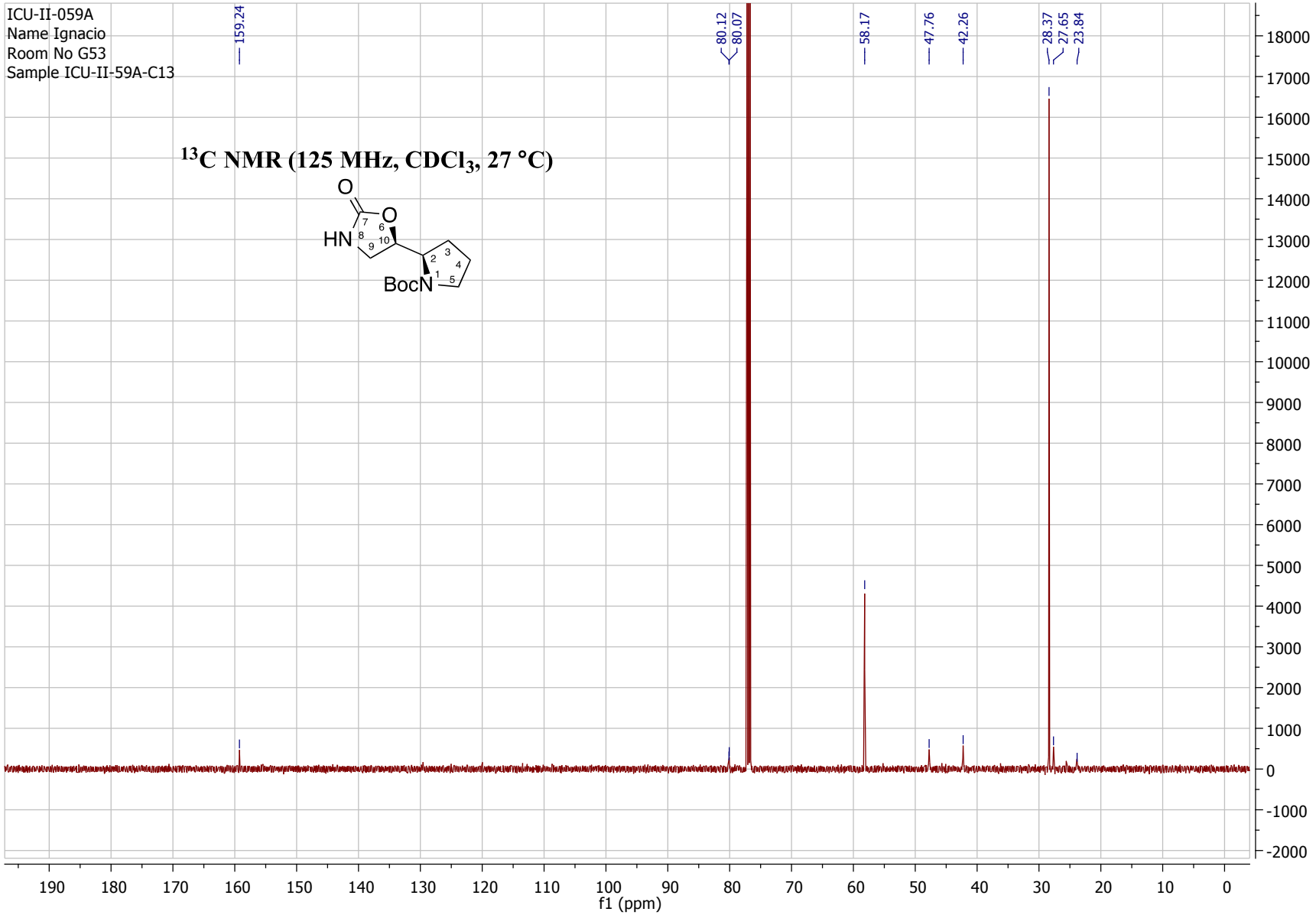

ICU-II-011A
Name Ignacio
Room No G53a
Sample ICU-II-011A
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
Cl$_3$C
O
N
7
6
2
3
4
1
5
BocN
60000
55000
50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
-5000
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)


ICU-II-011A
Name Ignacio
Room No G53a
Sample ICU-II-011A-C13
162.44
80.07
56.50
47.99
29.70
28.44
27.52
^{13}C NMR (125 MHz, CDCl$_3$, 27 °C)
Cl$_3$C
O
N
BocN
f1 (ppm)

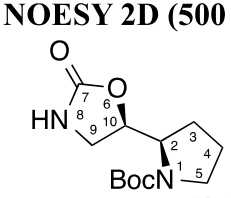

ICU-II-042A
Name Ignacio
Room No G53
Sample ICU-II-042A
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
HO
Cl$_3$COCHN
BocN
f1 (ppm)

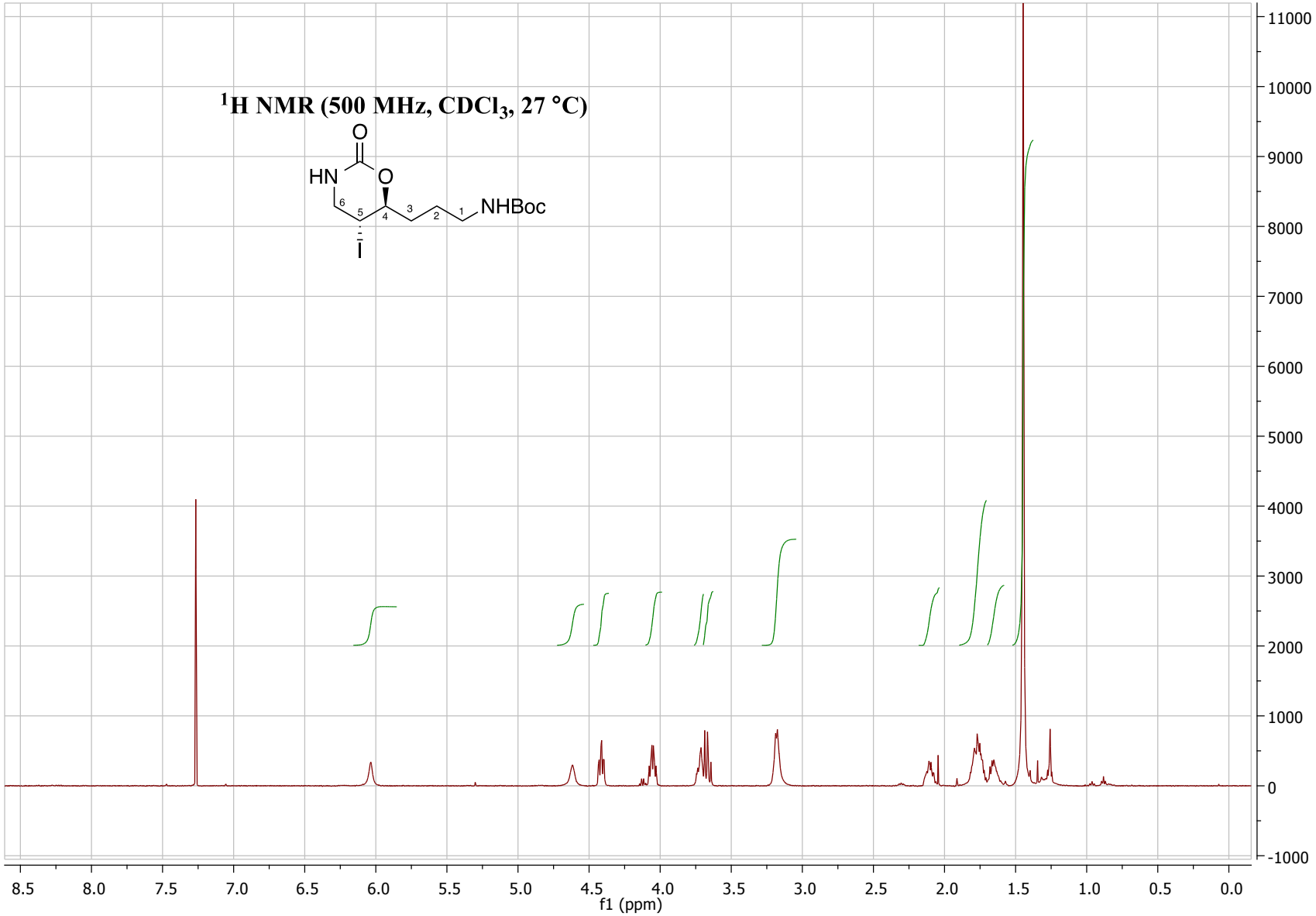

ICU-II-042A
Name Ignacio
Room No G53
Sample ICU-II-042A-C13
162.18
158.09
92.66
81.04
74.00
60.54
47.64
44.61
28.65
28.38
24.20
13C NMR (125 MHz, CDCl3, 27 °C)
HO
Cl3COCHN
BocN
f1 (ppm)

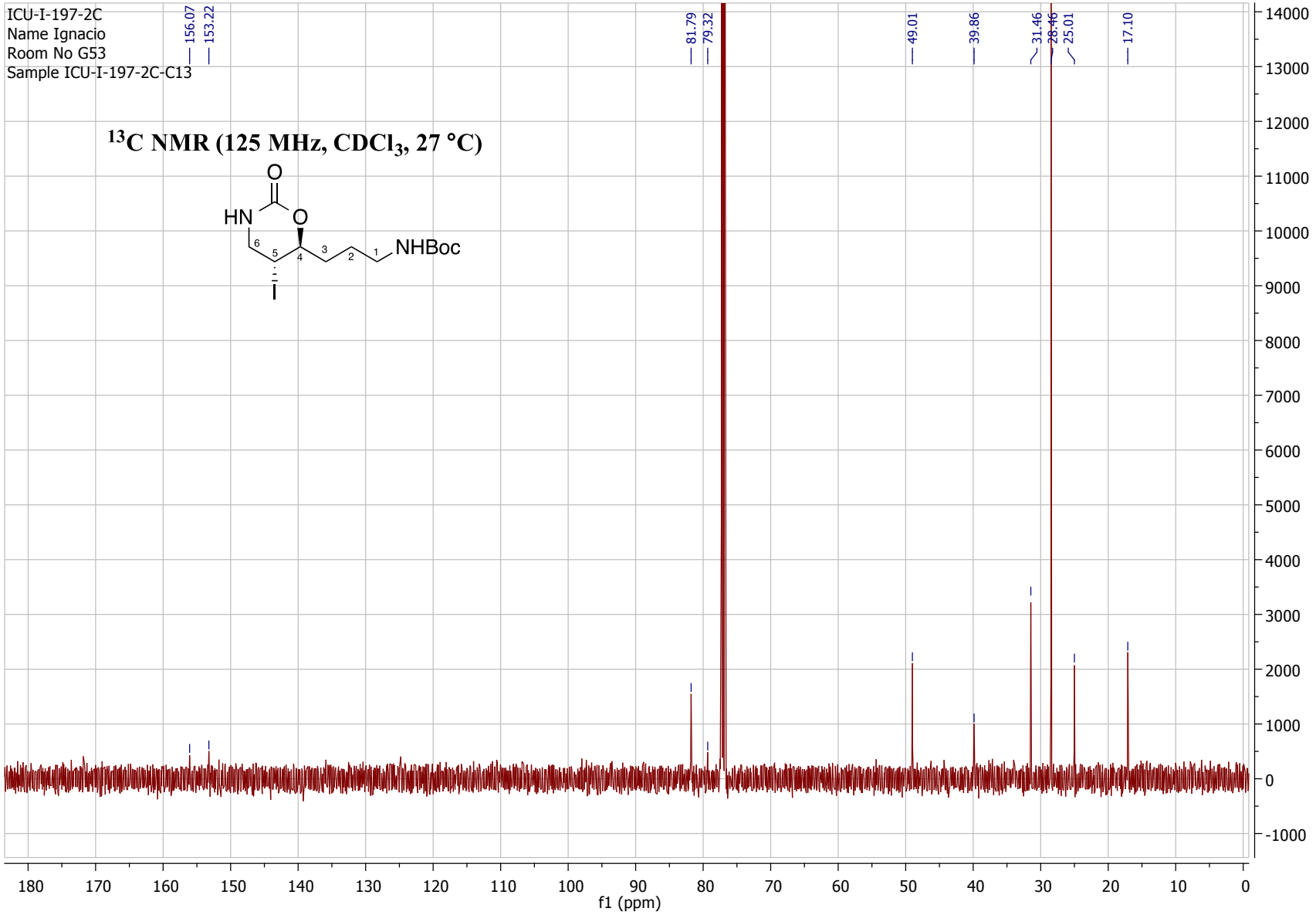

ICU-II-059A
Name Ignacio
Room No G53
Sample ICU-II-059A
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
O
O
HN
BocN
f1 (ppm)


ICU-II-059A
Name Ignacio
Room No G53
Sample ICU-II-59A-C13
159.24
80.12
80.07
58.17
47.76
42.26
28.37
27.65
23.84
13C NMR (125 MHz, CDCl3, 27 °C)
O
O
HN
BocN
f1 (ppm)

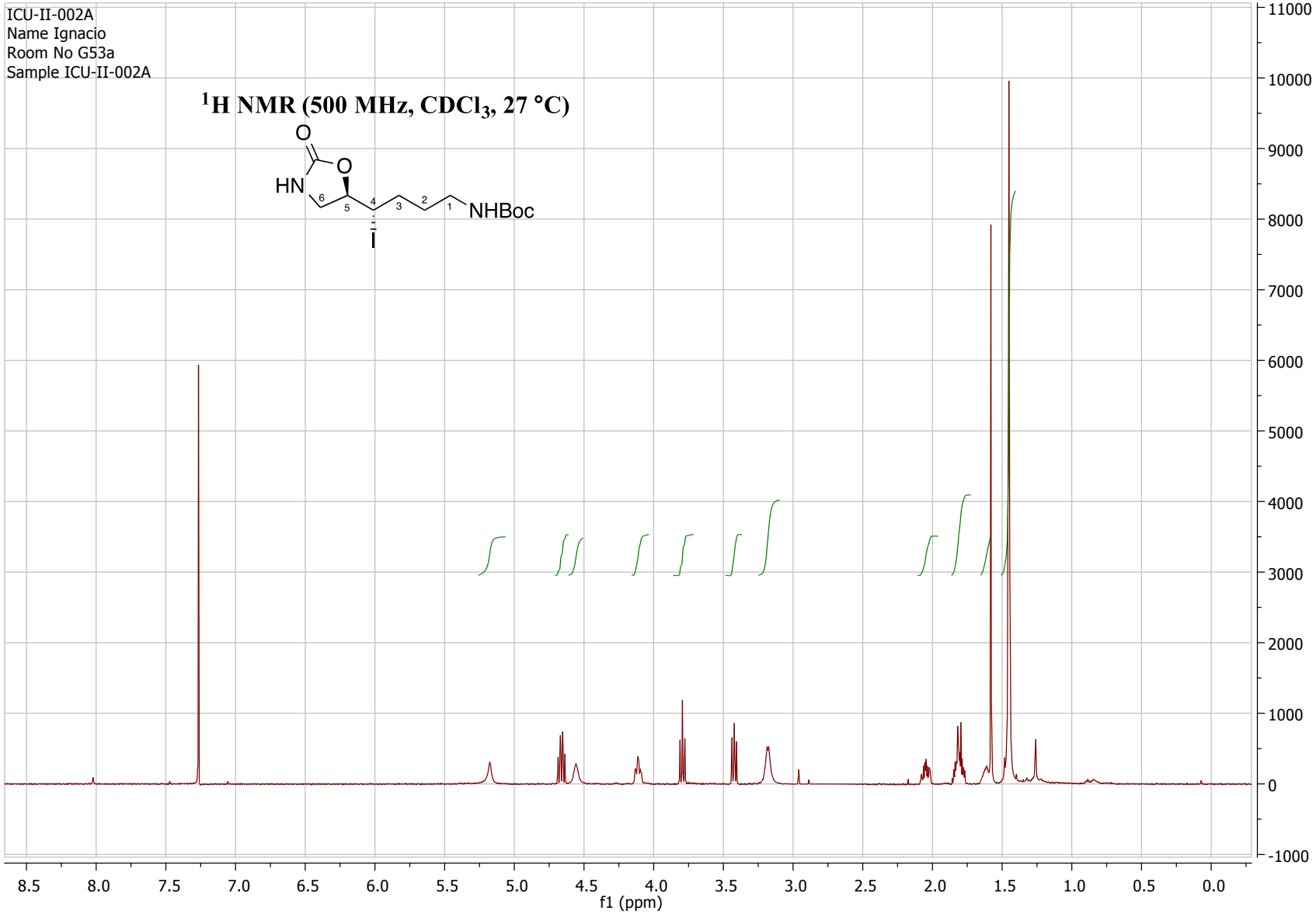
NOESY 2D (500 MHz, $CDCl_3$, 27 °C)

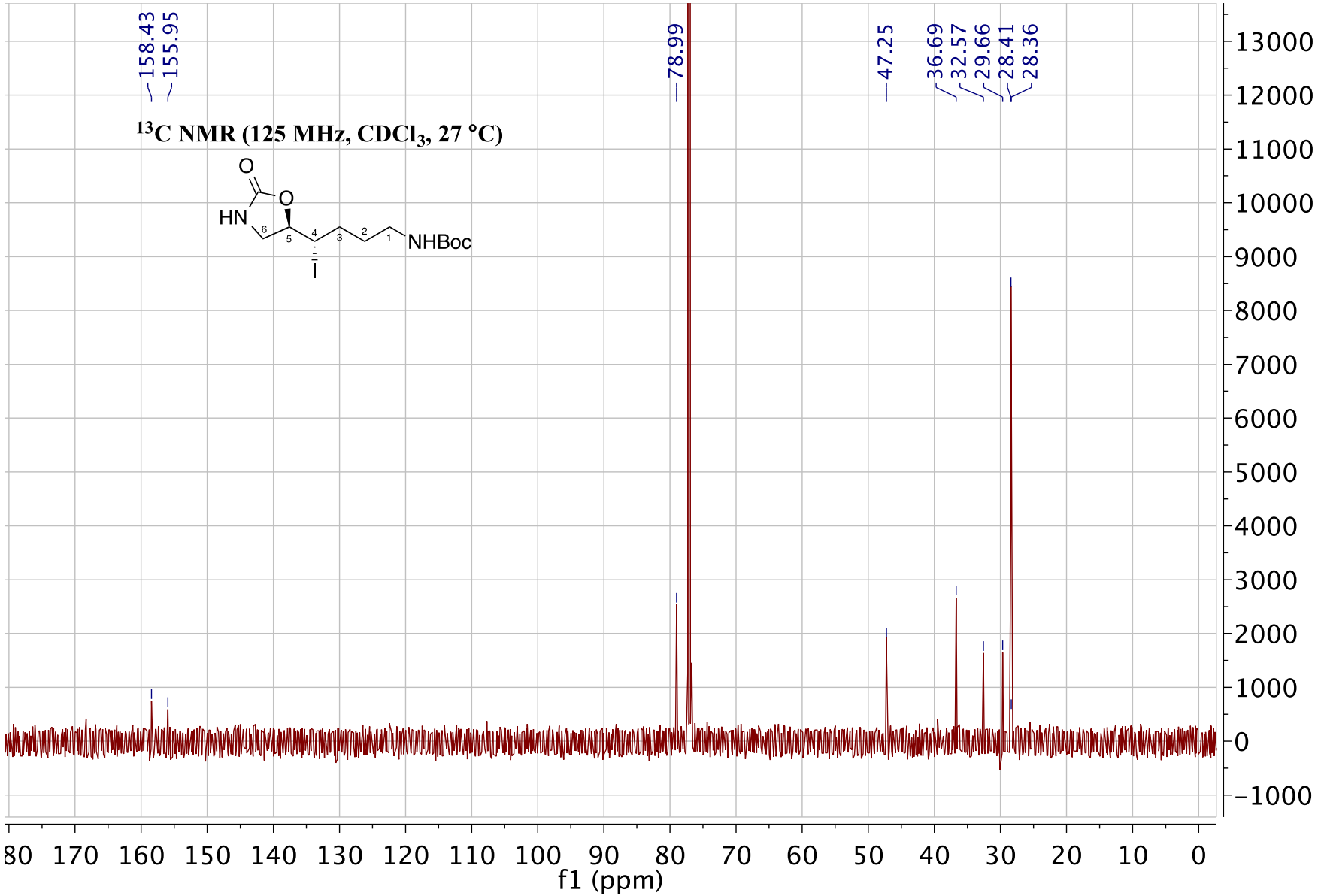



^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
HN
O
O
NHBoc
I
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

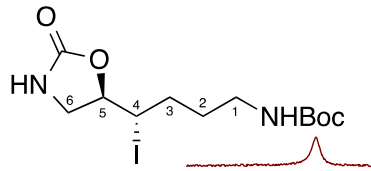

ICU-I-197-2C
Name Ignacio
Room No G53
Sample ICU-I-197-2C-C13
156.07
153.22
81.79
79.32
49.01
39.86
31.46
28.46
25.01
17.10
13C NMR (125 MHz, CDCl3, 27 °C)
HN
O
O
NHBoc
I
f1 (ppm)

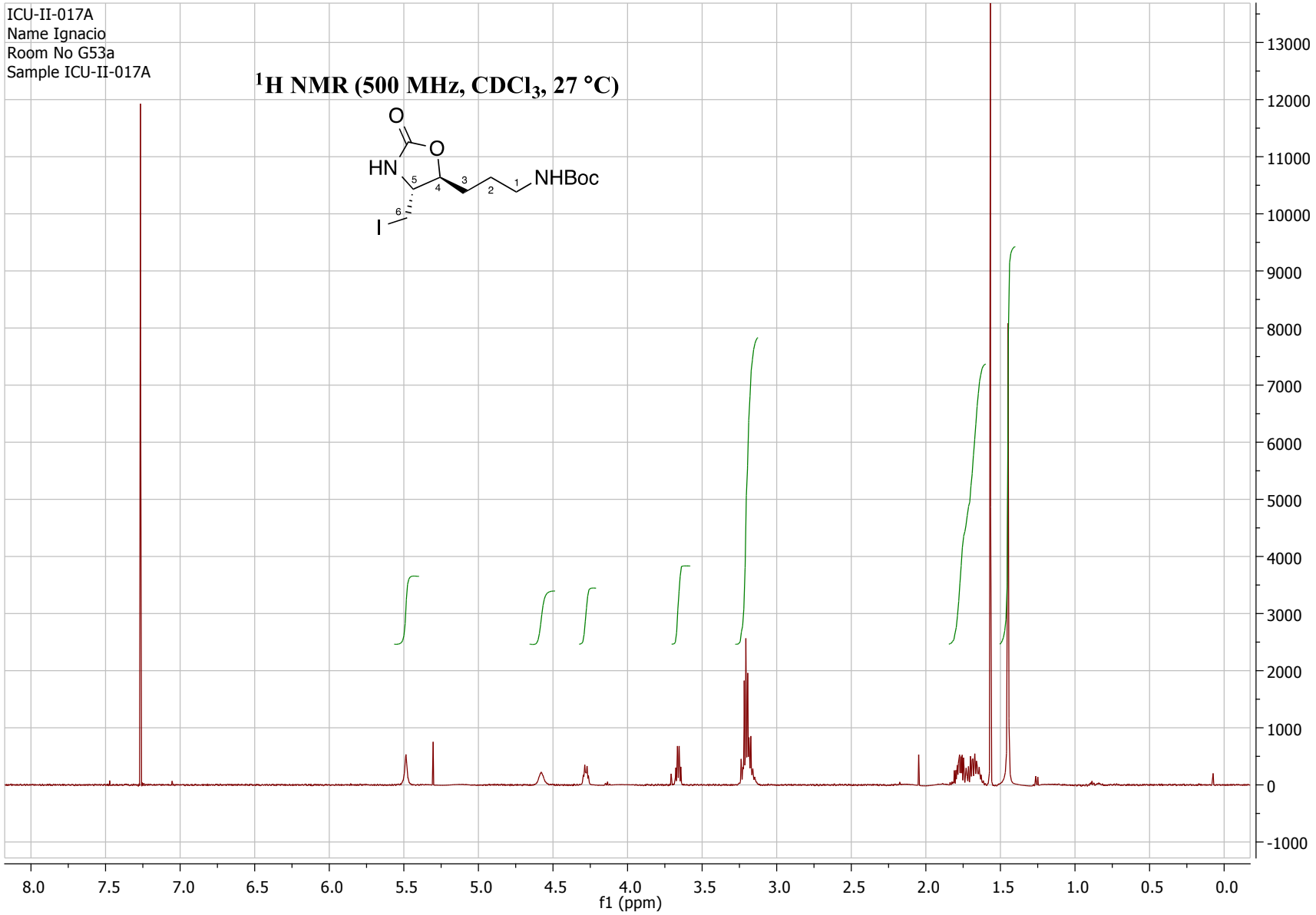
\

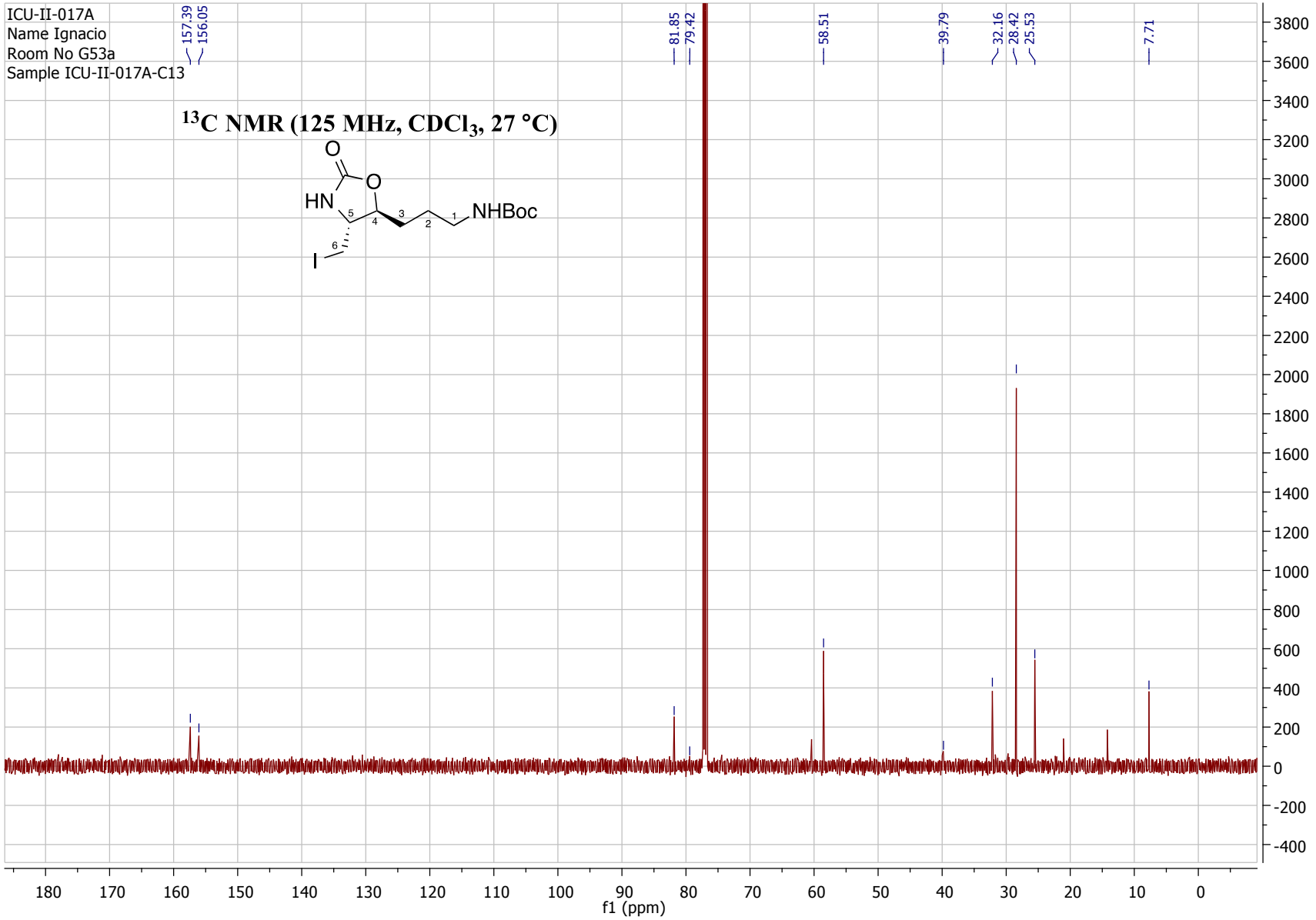

ICU-II-002A
Name Ignacio
Room No G53a
Sample ICU-II-002A
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
O
O
HN
6
5
4
3
2
1
NHBoc
I
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

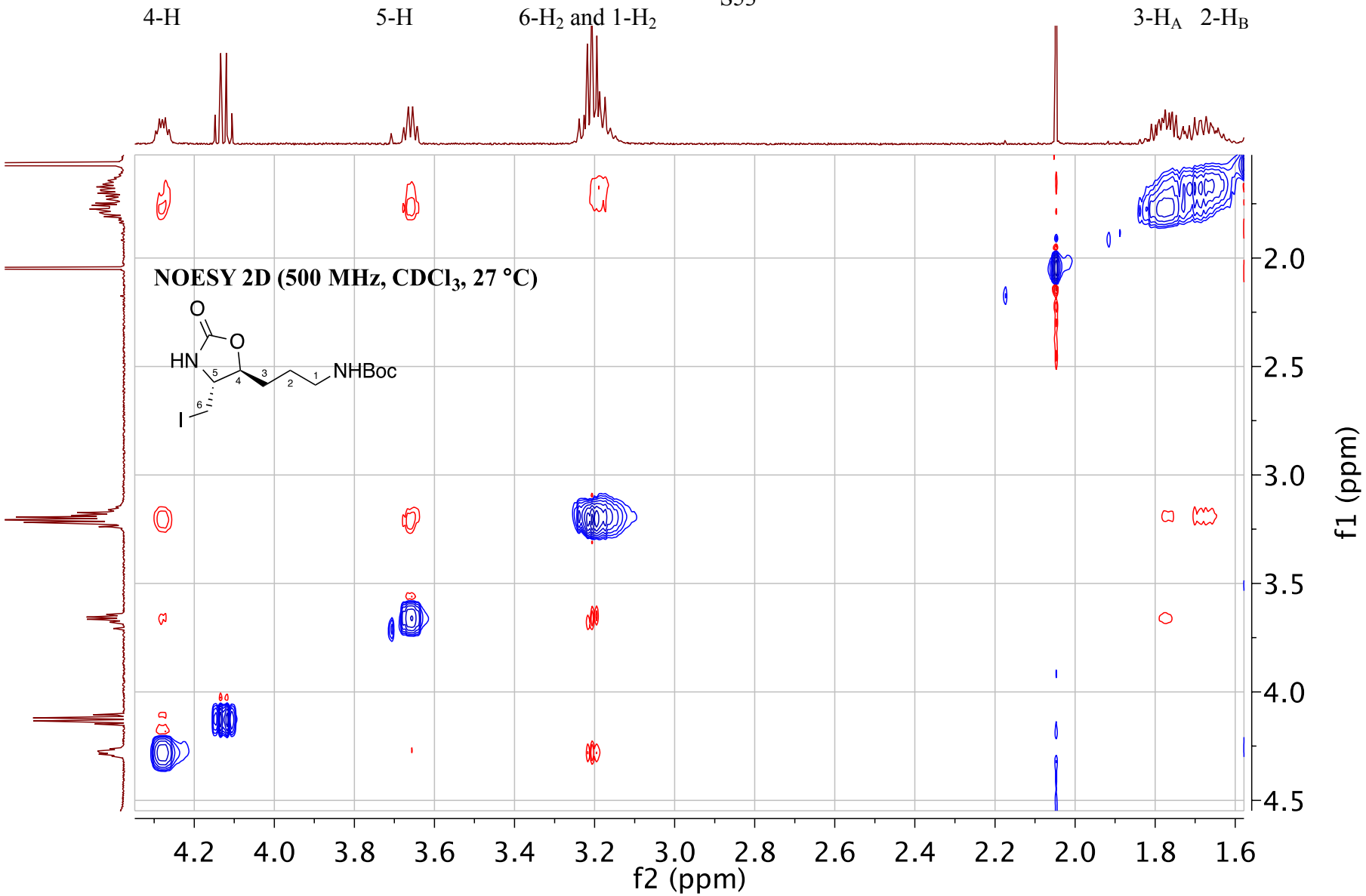

158.43
155.95
78.99
47.25
36.69
32.57
29.66
28.41
28.36
^{13}C NMR (125 MHz, CDCl$_3$, 27 °C)
O
O
HN
6
5
4
3
2
1
NHBoc
I
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
80 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

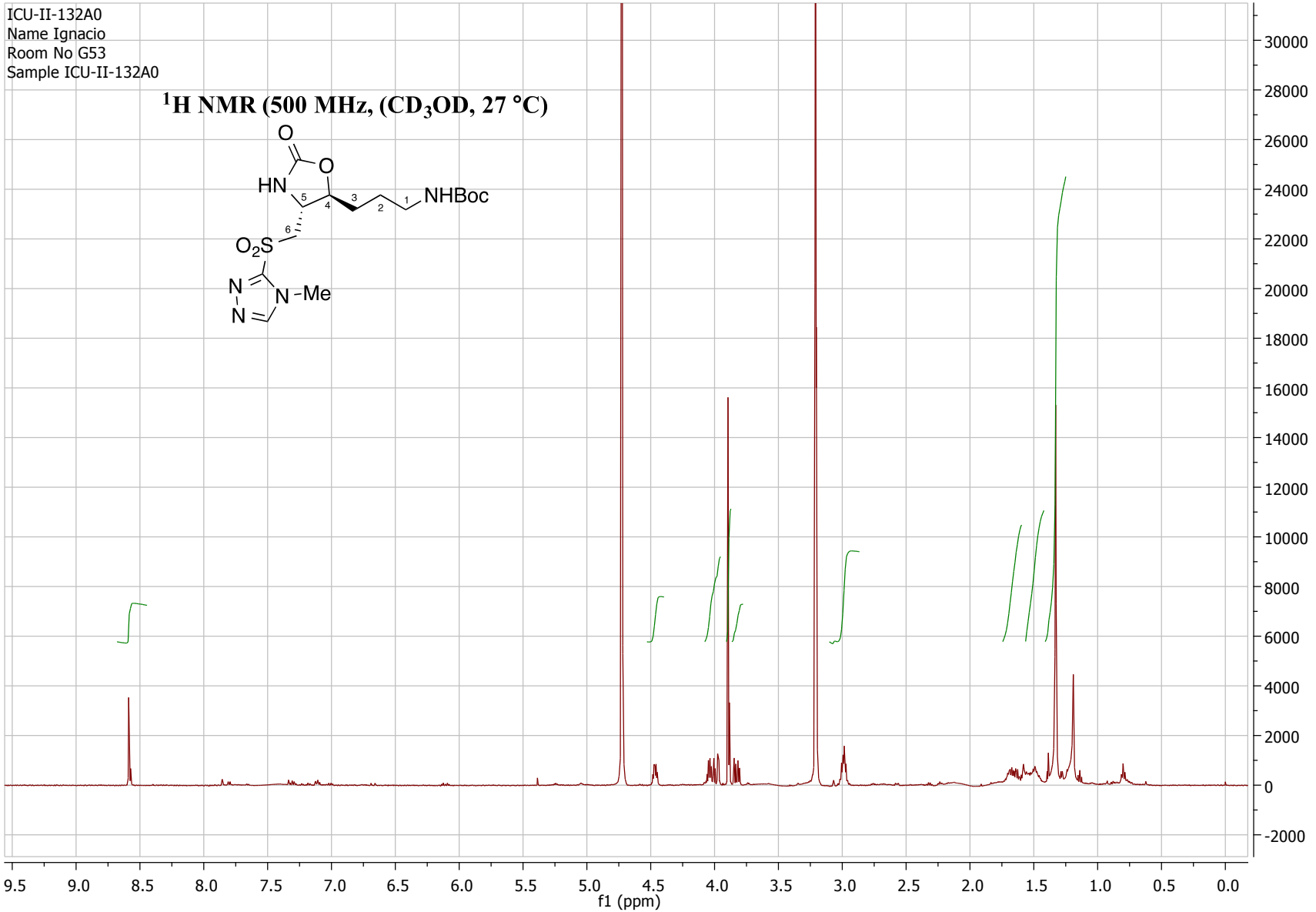
NOESY 2D (500 MHz, $CDCl_3$, 27 °C)

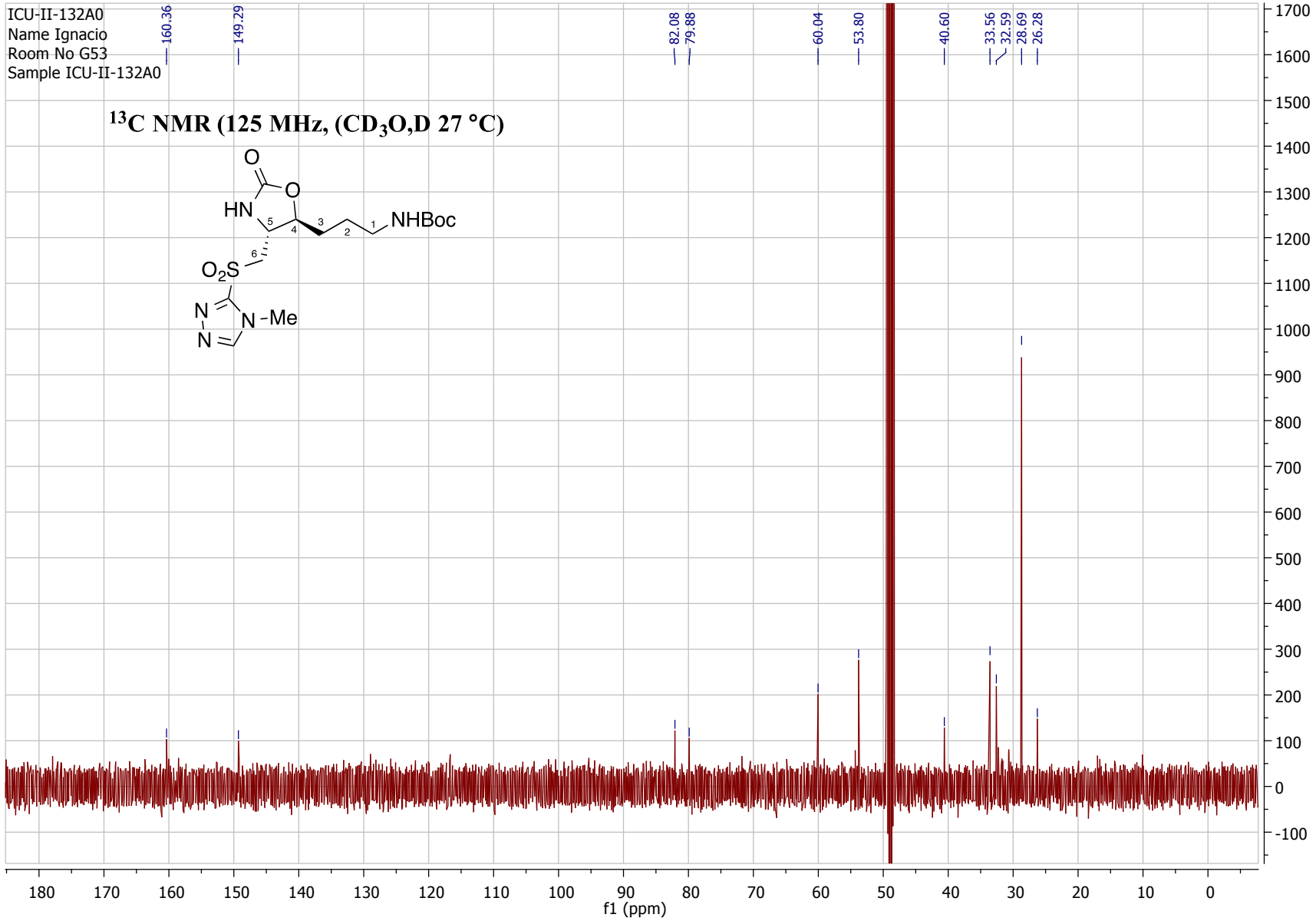


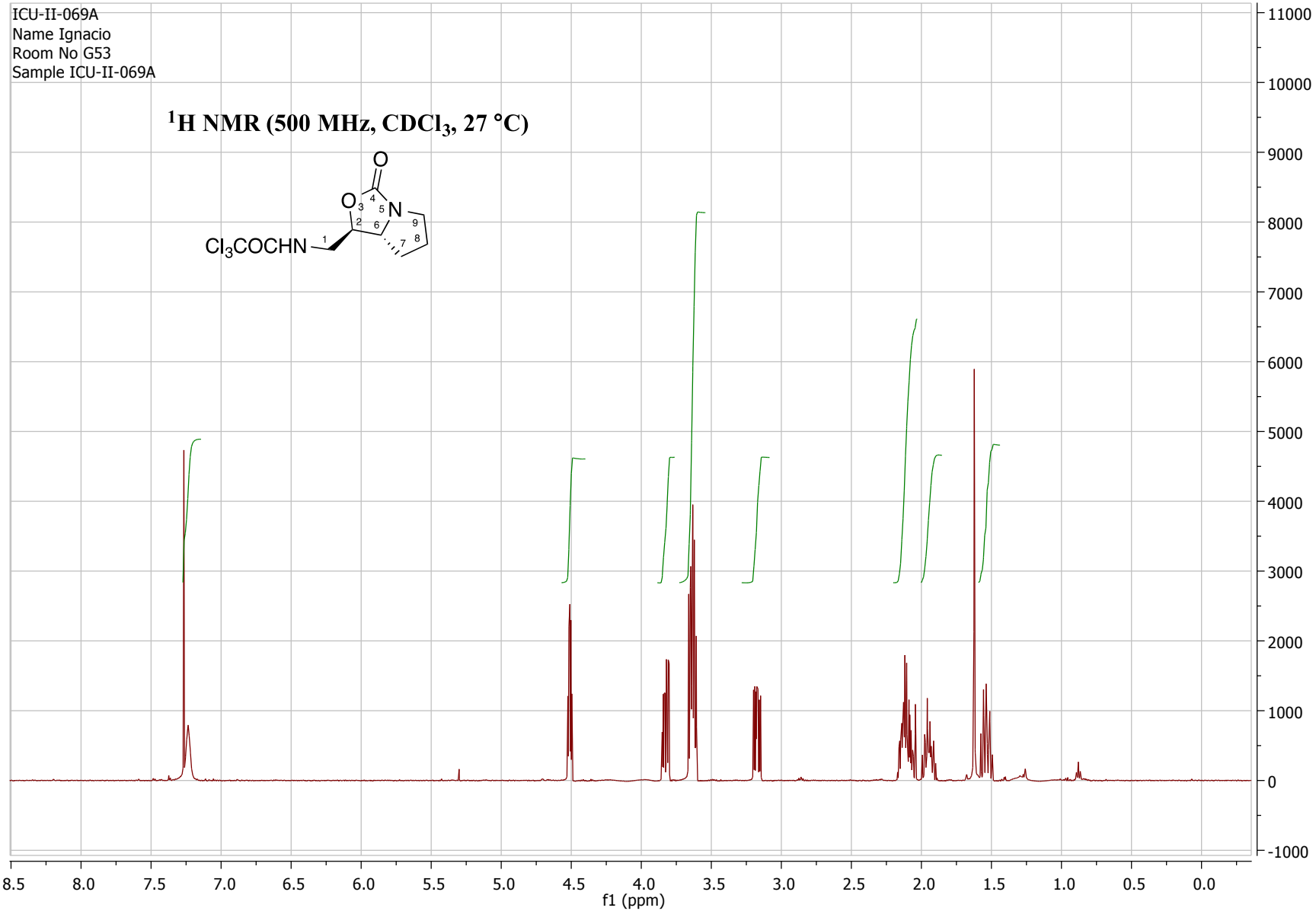

ICU-II-017A
Name Ignacio
Room No G53a
Sample ICU-II-017A
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
O
O
HN
5
4
3
2
1
NHBoc
6
I
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

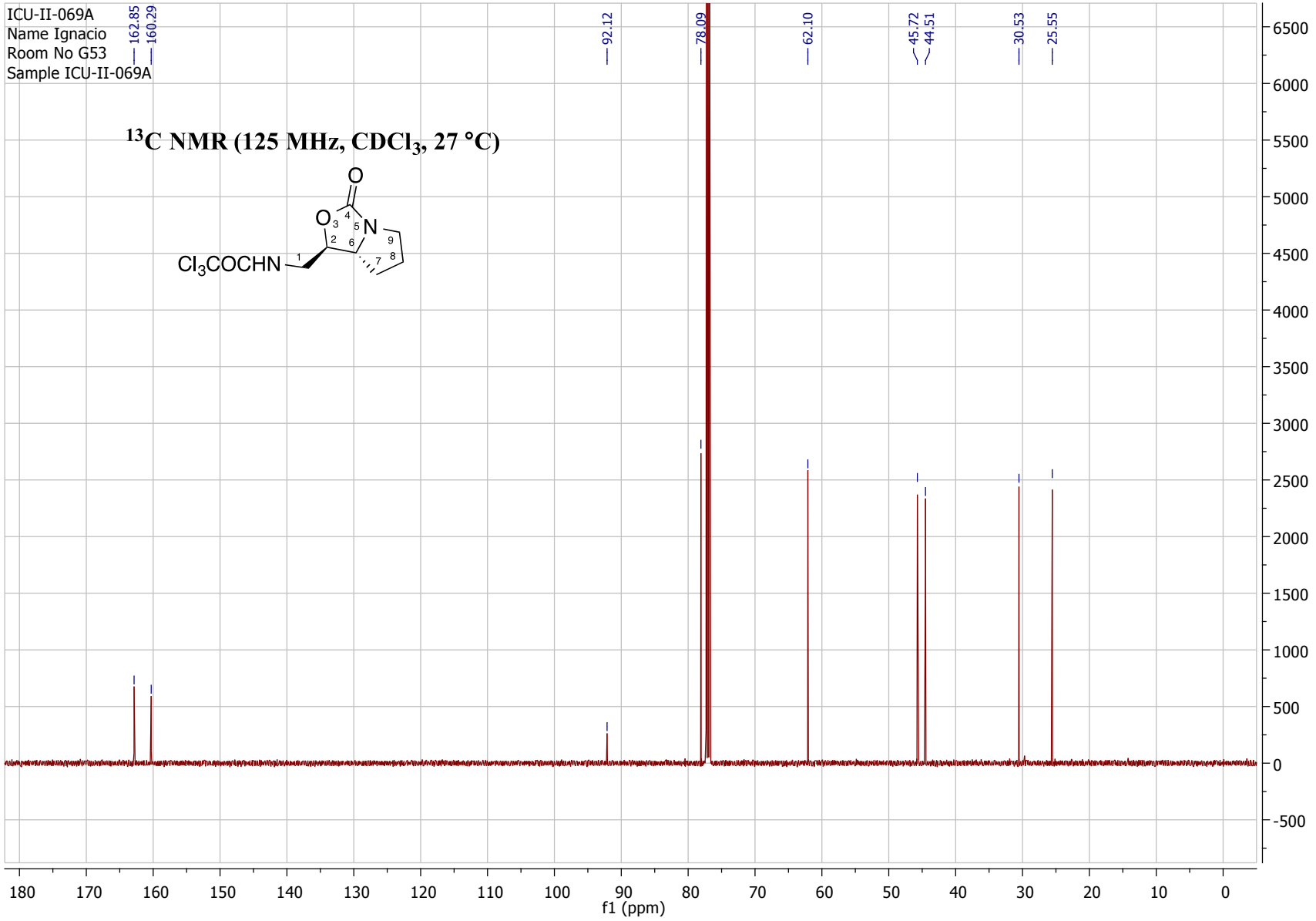

ICU-II-017A
Name Ignacio
Room No G53a
Sample ICU-II-017A-C13
157.39
156.05
81.85
79.42
58.51
39.79
32.16
28.42
25.53
7.71
13C NMR (125 MHz, CDCl3, 27 °C)
O
O
HN
NHBoc
I
f1 (ppm)


4-H
5-H
6-H2 and 1-H2
3-HA 2-HB
NOESY 2D (500 MHz, CDCl3, 27 °C)
O
O
HN
NHBoc
I
f2 (ppm)
f1 (ppm)

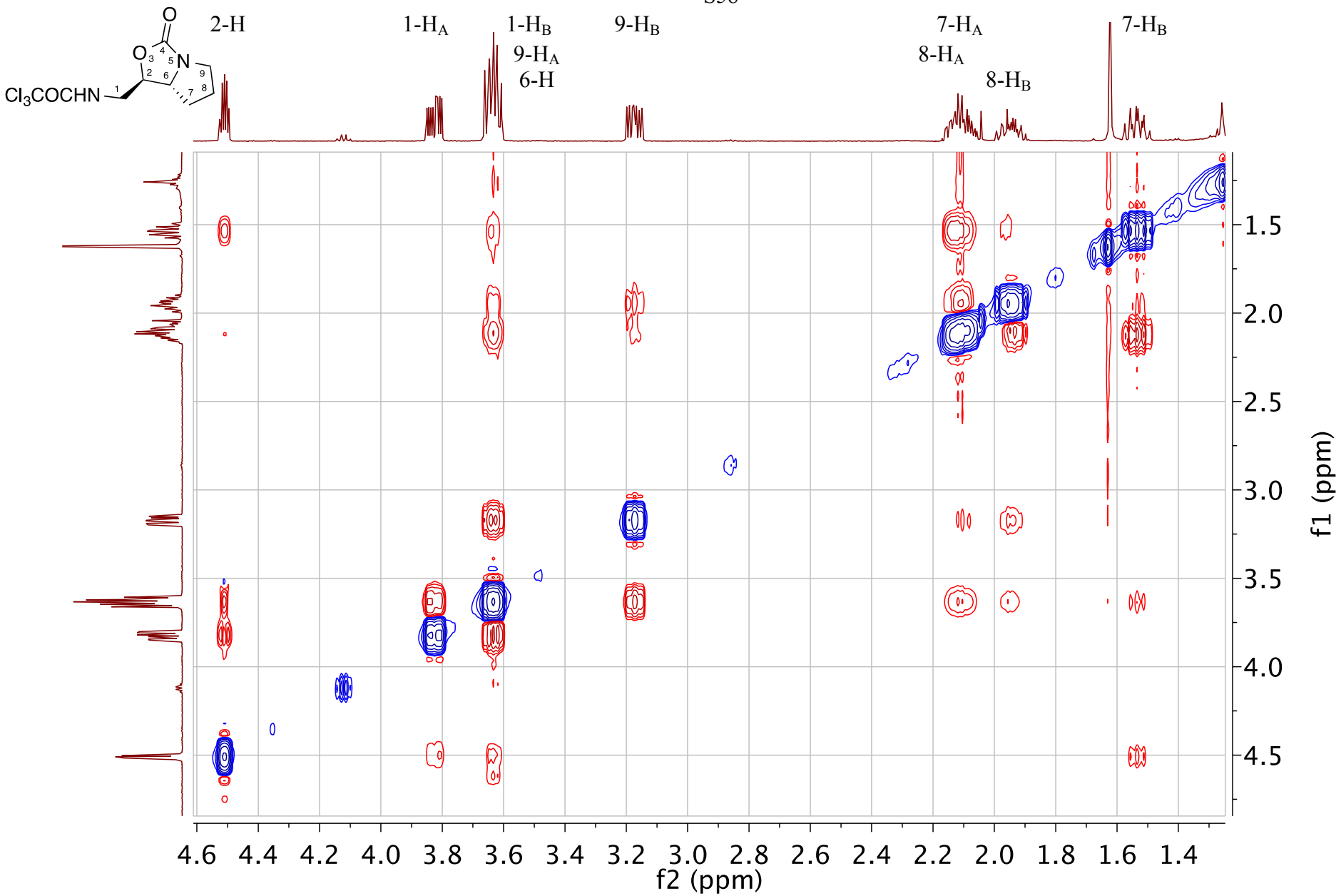

ICU-II-132A0
Name Ignacio
Room No G53
Sample ICU-II-132A0
^{1}H NMR (500 MHz, (CD_3OD, 27 °C)
O
O
HN
5
4
3
2
1
NHBoc
6
O_2S
N
N
N
Me
30000
28000
26000
24000
22000
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
-2000
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

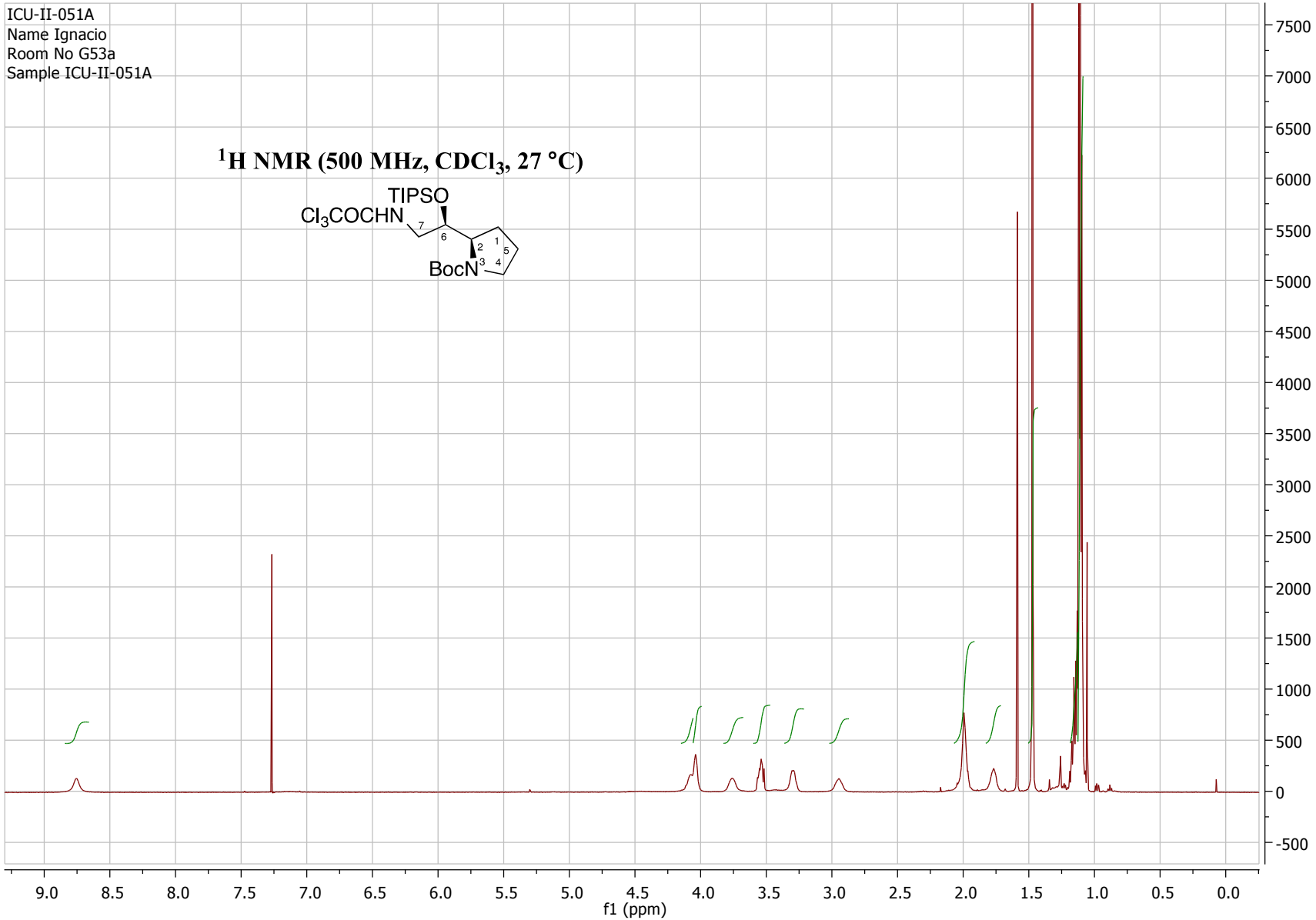

ICU-II-132A0
Name Ignacio
Room No G53
Sample ICU-II-132A0
160.36
149.29
82.08
79.88
60.04
53.80
40.60
33.56
32.59
28.69
26.28
13C NMR (125 MHz, (CD3O,D 27 °C)
O
O
HN
5
4
3
2
1
NHBoc
6
O2S
N
N
N
Me
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
-100
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

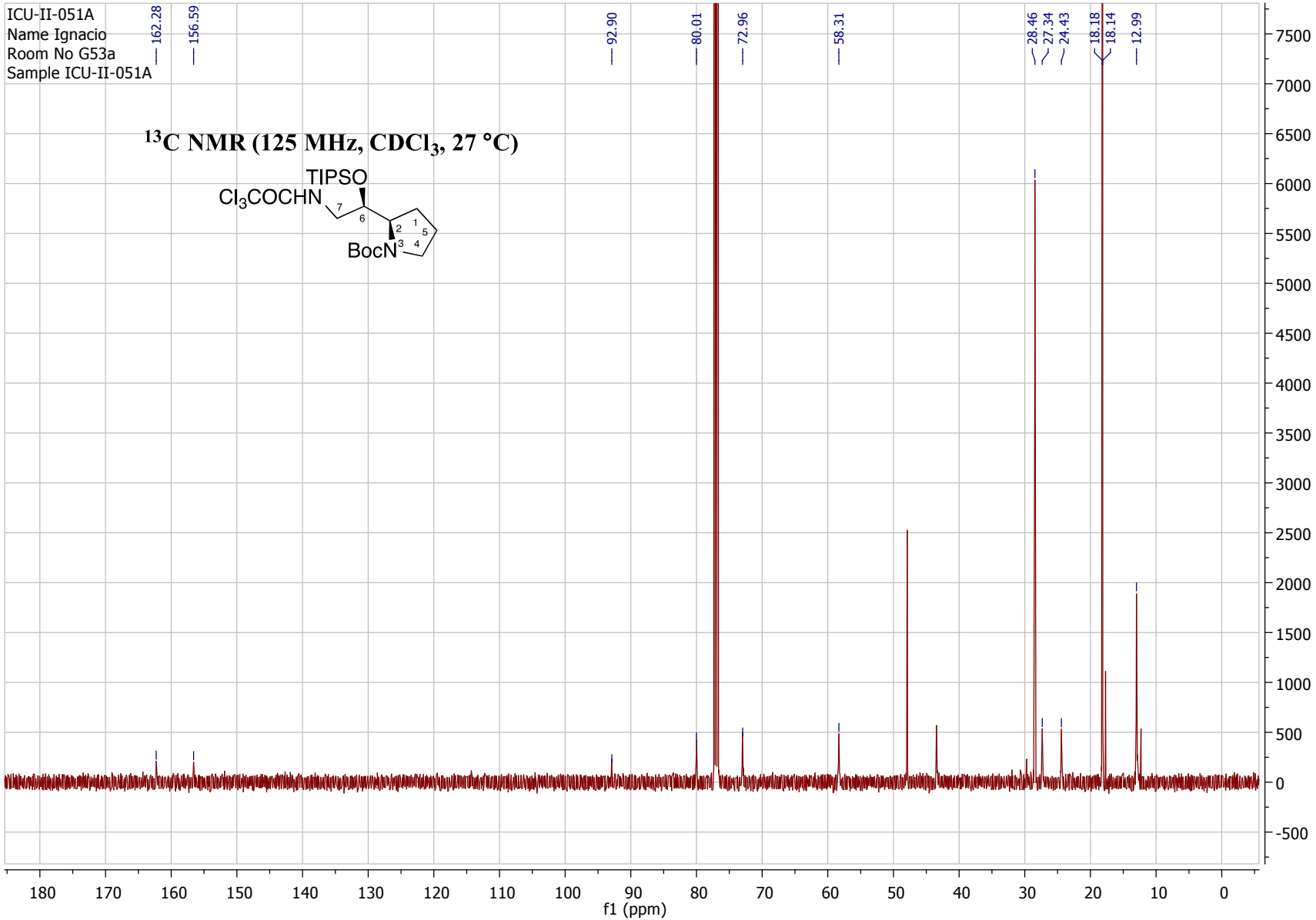

ICU-II-069A
Name Ignacio
Room No G53
Sample ICU-II-069A
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
Cl$_3$COCHN
f1 (ppm)

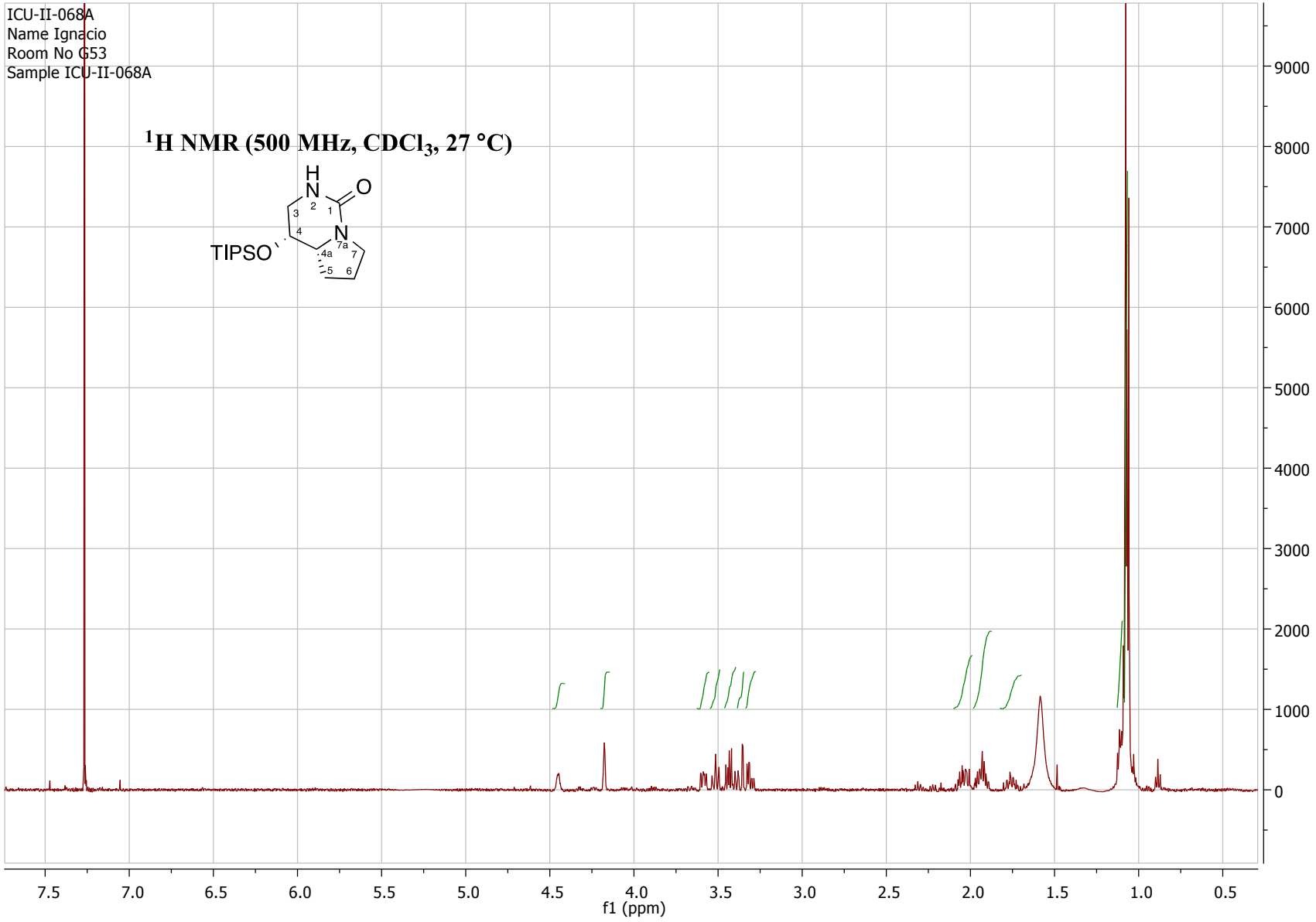

ICU-II-069A
Name Ignacio
Room No G53
Sample ICU-II-069A
162.85
160.29
92.12
78.09
62.10
45.72
44.51
30.53
25.55
^{13}C NMR (125 MHz, CDCl$_3$, 27 °C)
Cl$_3$COCHN
f1 (ppm)

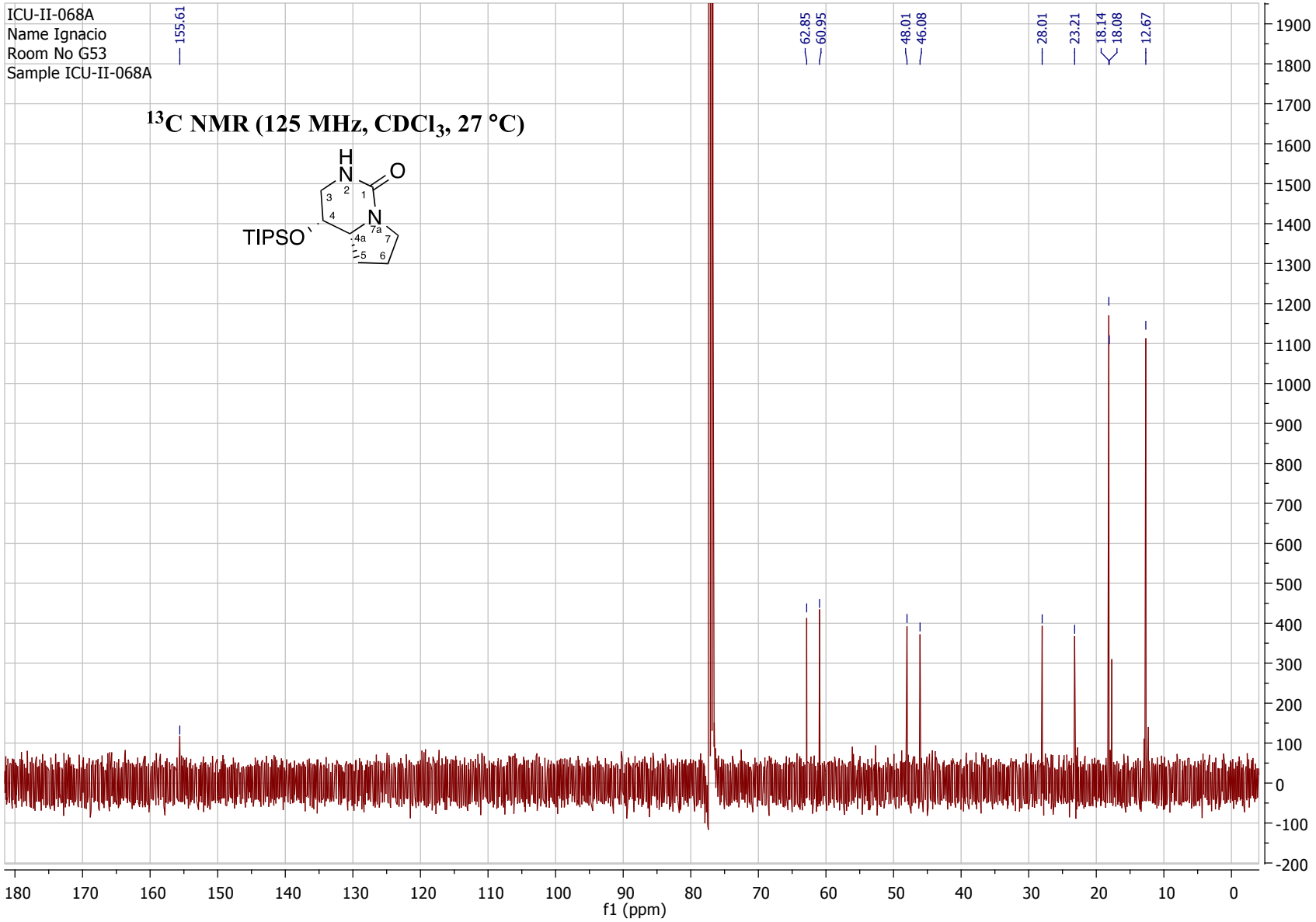
NOESY 2D (500 MHz, $CDCl_3$, 27 °C)

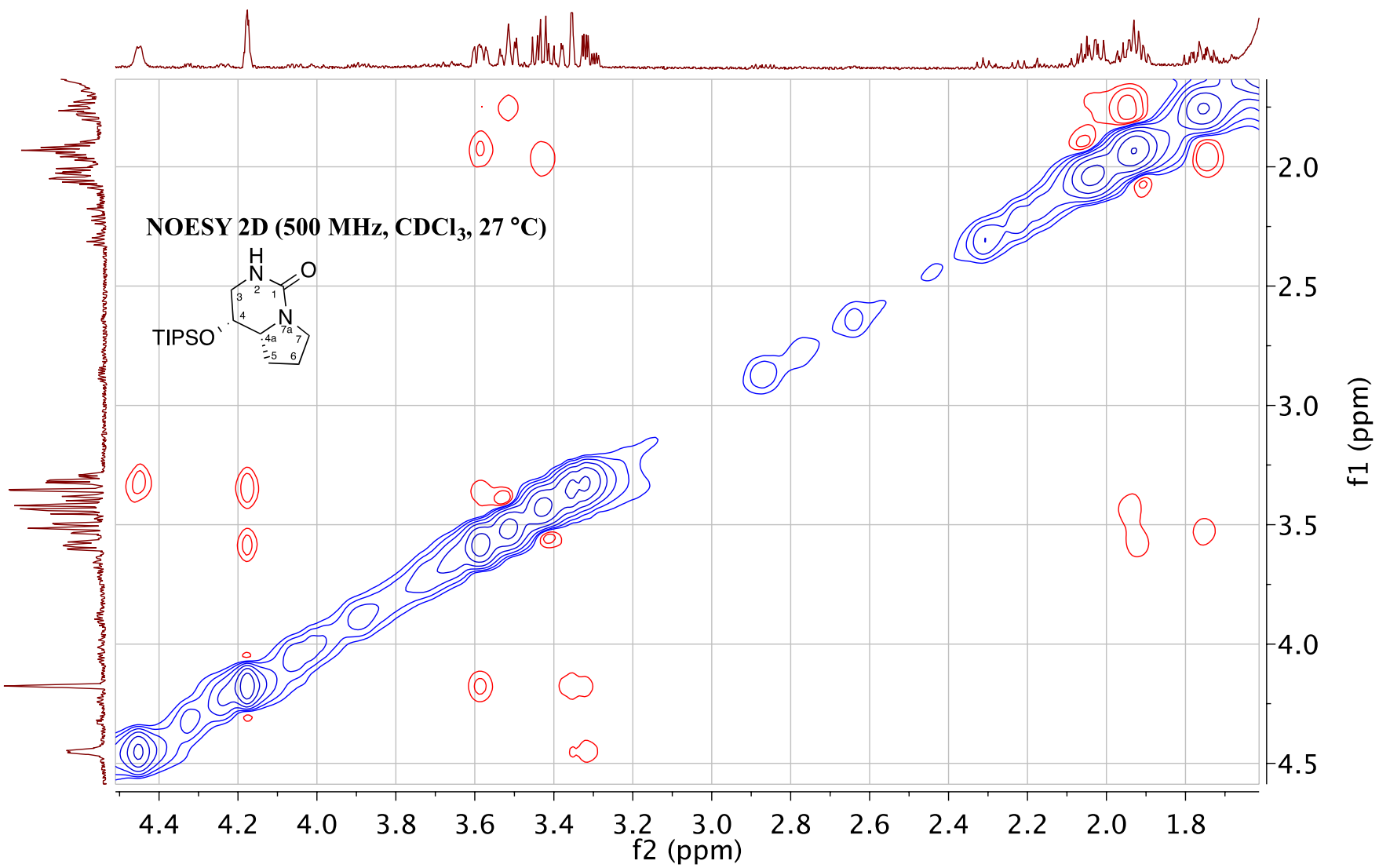


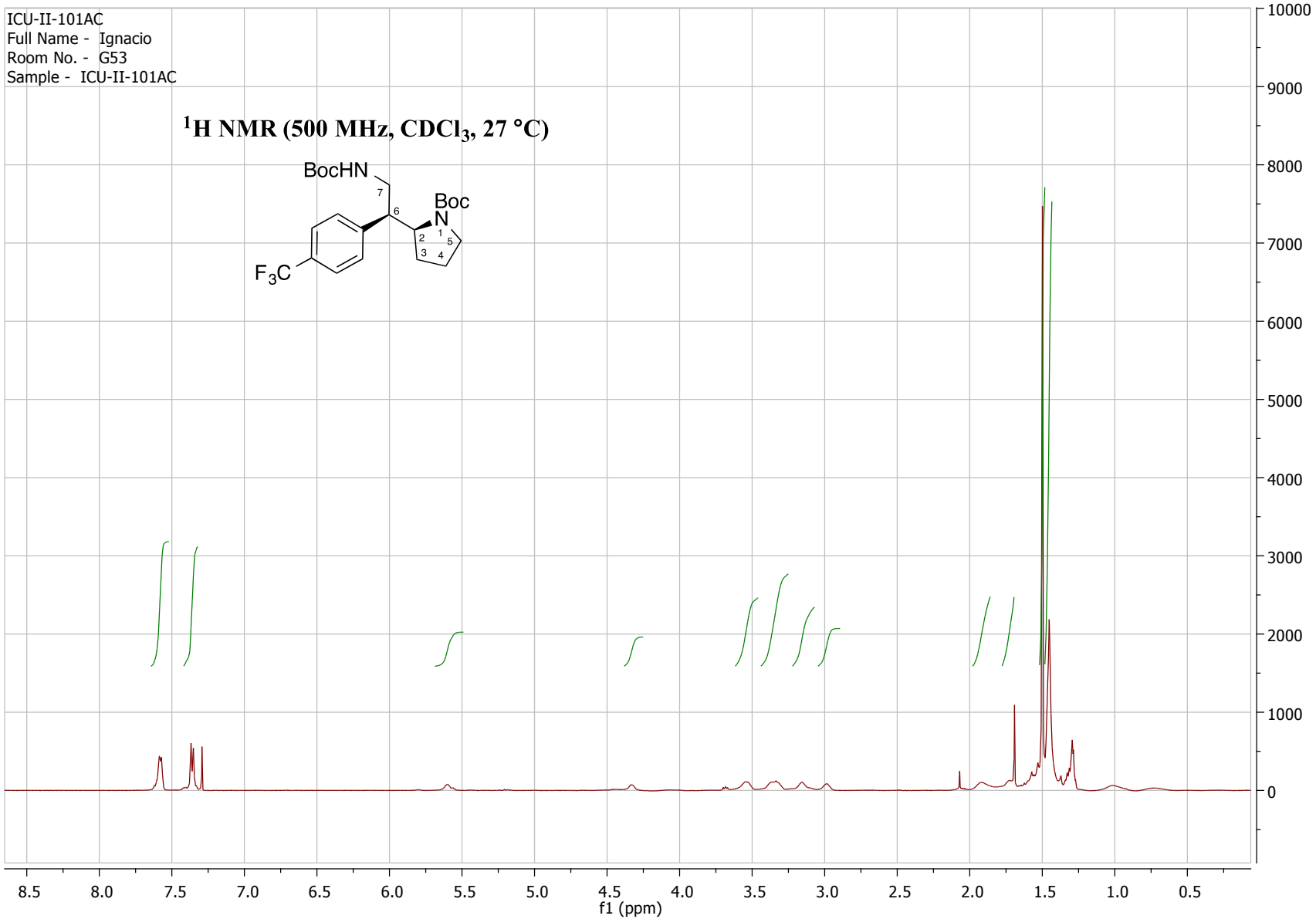

ICU-II-051A
Name Ignacio
Room No G53a
Sample ICU-II-051A
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
TIPSO
Cl$_3$COCHN
BocN
f1 (ppm)

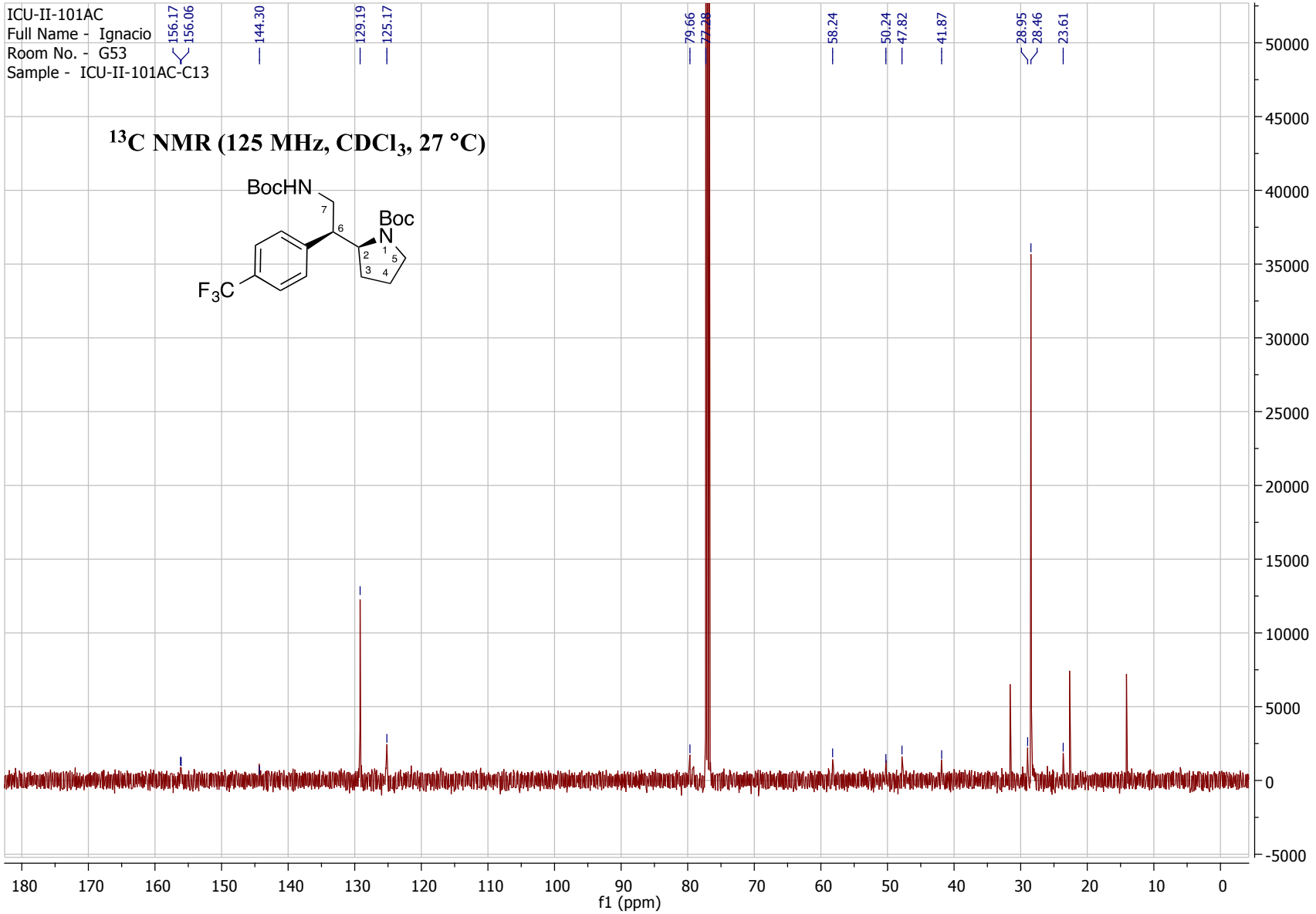

ICU-II-051A
Name Ignacio
Room No G53a
Sample ICU-II-051A
162.28
156.59
92.90
80.01
72.96
58.31
28.46
27.34
24.43
18.18
18.14
12.99
13C NMR (125 MHz, CDCl3, 27 °C)
TIPSO
Cl3COCHN
BocN
f1 (ppm)

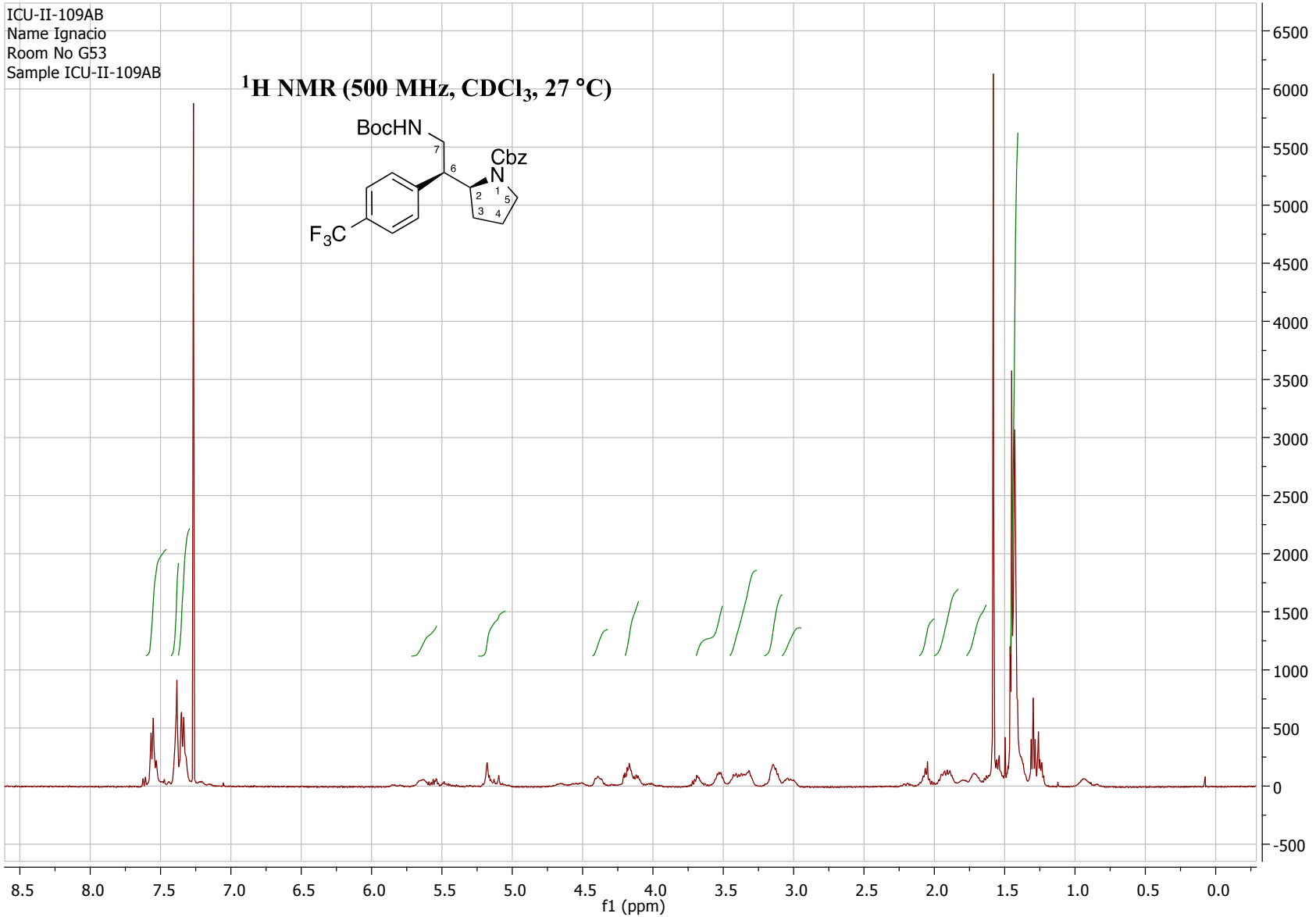

ICU-II-068A
Name Ignacio
Room No G53
Sample ICU-II-068A
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
TIPSO
H
N
O
N
7a
4a
f1 (ppm)

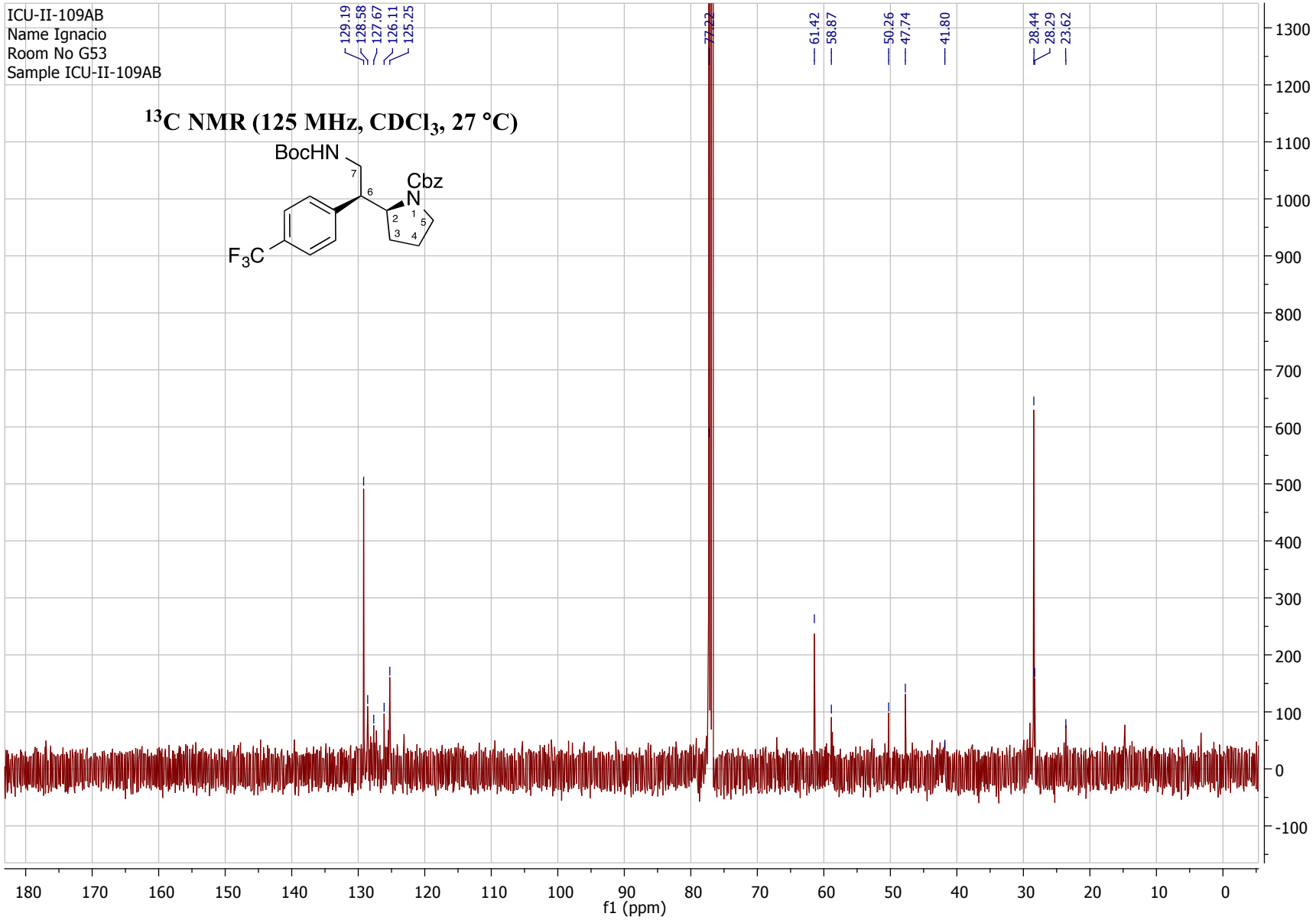

ICU-II-068A
Name Ignacio
Room No G53
Sample ICU-II-068A
155.61
62.85
60.95
48.01
46.08
28.01
23.21
18.14
18.08
12.67
13C NMR (125 MHz, CDCl3, 27 °C)
TIPSO
f1 (ppm)

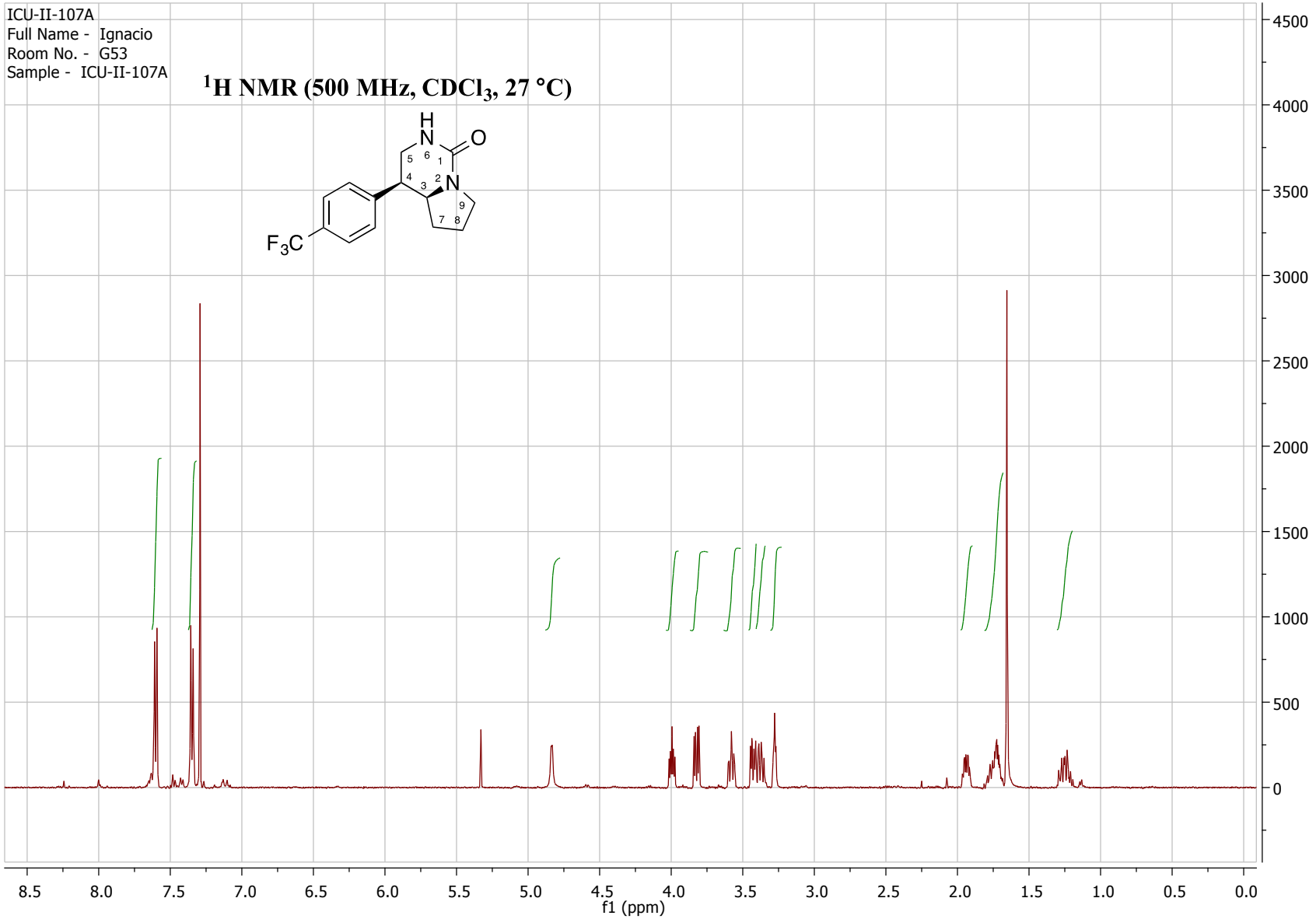

4-H
4a-H
7-H$_B$
3-H$_B$
7-H$_A$
3-H$_A$
5-H$_A$
5-H$_B$
6-H$_A$
6-H$_B$
NOESY 2D (500 MHz, CDCl$_3$, 27 °C)
TIPSO
f2 (ppm)
f1 (ppm)

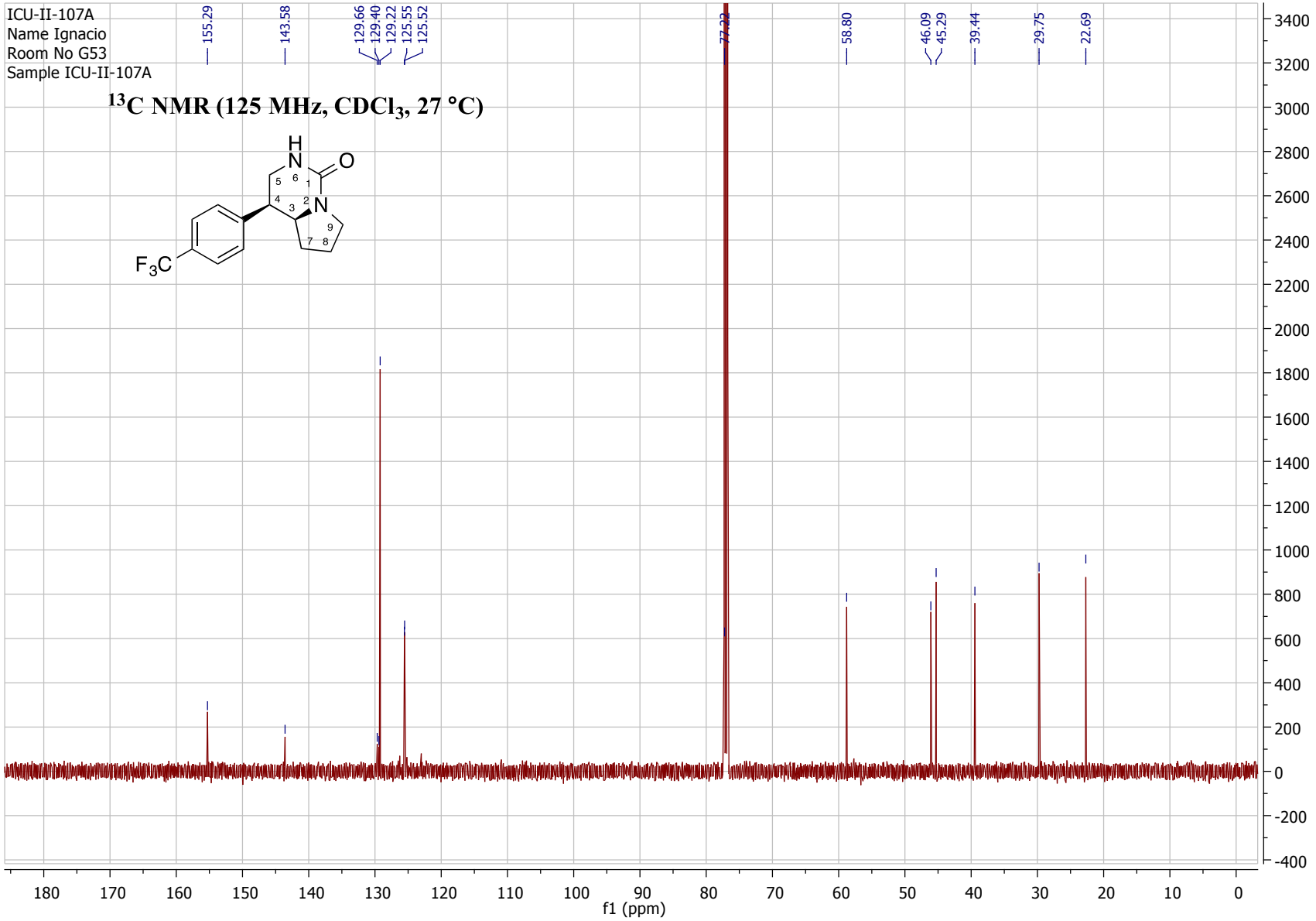

ICU-II-101AC
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-101AC
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
BocHN
Boc
N
F$_3$C
f1 (ppm)


ICU-II-101AC
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-101AC-C13
156.17
156.06
144.30
129.19
125.17
79.66
77.28
58.24
50.24
47.82
41.87
28.95
28.46
23.61
13C NMR (125 MHz, CDCl3, 27 °C)
BocHN
Boc
N
F3C
f1 (ppm)


ICU-II-109AB
Name Ignacio
Room No G53
Sample ICU-II-109AB
^{1}H NMR (500 MHz, $CDCl_3$, 27 °C)
BocHN
Cbz
N
F_3C
f1 (ppm)

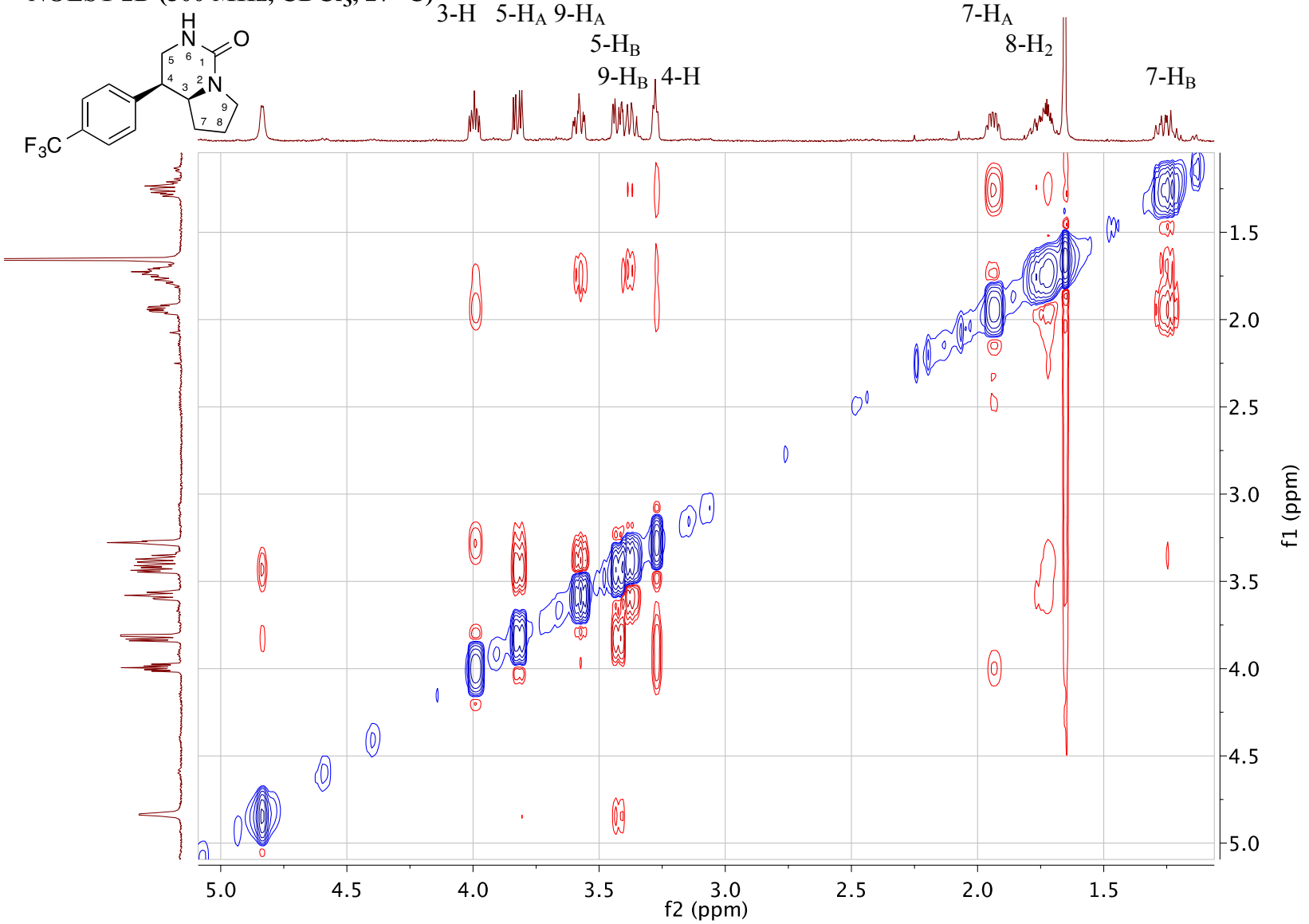

ICU-II-109AB
Name Ignacio
Room No G53
Sample ICU-II-109AB
129.19
128.58
127.67
126.11
125.25
77.22
61.42
58.87
50.26
47.74
41.80
28.44
28.29
23.62
13C NMR (125 MHz, CDCl3, 27 °C)
BocHN
Cbz
N
F3C
f1 (ppm)


ICU-II-107A
Full Name - Ignacio
Room No. - G53
Sample - ICU-II-107A
^{1}H NMR (500 MHz, CDCl$_3$, 27 °C)
H
N
O
N
F$_3$C
f1 (ppm)



^{13}C NMR (125 MHz, $CDCl_3$, 27 °C)

NOESY 2D (500 MHz, $CDCl_3$, 27 °C)



6. References

1. MySQL, version 5.1.73, Oracle Corporation, 500 Oracle Parkway, Redwood Shores, CA 94065, USA, http://www.mysql.com/ (accessed Jun 2015).
2. D3.js, M. Bostock, https://d3js.org/
3. ZeroMQ, version 3.2.1, iMatix Corp, Werkhuizenstraat 13-15, 1080 Brussels, Belgium, http://zeromq.org/ (accessed Jun 2015).
4. SMARTS – A language for describing molecular patterns, Daylight Chemical Information Systems, Inc., PO Box 7737, Laguna Niguel, CA 92677, USA, http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (accessed Apr 2015).
5. M. Hartenfeller, M. Eberle, P. Meier, C. Nieto-Oberhuber, K. H., Altmann, G. Schneider, E. Jacoby and S.Renner, *J. Chem. Inf. Model.*, 2011, **51**, 3093.
6. J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, *J. Chem. Inf. Model.*, 2012, **52**, 1757.
7. G. W. Bemis and M. A. Murcko, *J. Med. Chem.,* 1996, **39**, 2887.
8. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv. Rev.*, 2001, **46**, 3.
9. Indigo toolkit, version 1.2.1, EPAM Systems, Newtown, PA 18940, USA. http://lifescience.opensource.epam.com/indigo/ (accessed Apr 2015).
10. I. V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E.V. Radchenko, N.S. Zefirov, A. S. Makarenko, V. Y. Tanchuk and V. V. Prokopenko, *J. Comput. Aid. Mol. Des.*, 2005, **19**, 453.
11. D.F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple, *J. Med. Chem*., 2002, **45**, 2615.
12. (a) Chemistry Toolkit Rosetta Wiki: Calculate TPSA, http://ctr.wikia.com/wiki/Calculate_TPSA (accessed Oct 2015); (b) P. Ertl, B. Rohde and P. Selzer, J. Med. Chem., 2000, **43**, 3714.
13. K. Palm, P. Stenberg, K. Luthman and P. Artursson, *Pharma. Res.*, 1997, **14**, 568.
14. S. A. Hitchcock and L. D. Pennington, *J. Med. Chem.*, 2006, **49**, 7559.
15. RDKit: Open-source cheminformatics; http://www.rdkit.org (accessed Oct 2015).
16. N. C. Firth, N. Brown and J. Blagg, *J. Chem*. *Inf. Model*., 2012, **52**, 2516.
17. (a) The Open Babel Package, version 2.3.1 http://openbabel.org (accessed Oct 2015); (b) N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison., *J. Cheminform.* 2011, **3**, 33.