Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

Supporting Information

A convenient synthetic route to (2*S*, 4*S*)-methylproline and its exploration for protein engineering of thioredoxin

Andrea Caporale,[a,b] Jennie O' Loughlin, [a] Yannick Ortin [a] and Marina Rubini[*] [a]

a. School of Chemistry, University College Dublin, Belfield, Dublin 4, Ireland. E-mail: marina.rubini@ucd.ie
b. Present address: Institute of Crystallography, C.N.R., Basovizza (TS), Italy

TABLE OF CONTENTS

General Methods

All compounds and reagents of reagent grade or better were purchased from commercial suppliers (Sigma-Aldrich, Across, Flourochem, Iris Biotech, Biosynth Carbosynth) and were used without further purification. When reaction conditions involved anhydrous solvents, glassware was either oven- or flame-dried. $NaHCO_3$ and brine (NaCl) refers to saturated aqueous solutions unless specified otherwise. Flash chromatography was performed with columns of silica gel Merck 60H, 230–240 mesh.

For enantiomer separation, the analytical HPLC was performed with chiral column Daicel Chiralcel® OJ-H of 4.6 mm ID x 250 mm L and 5 µm particle size coupled to a guard column of 4.6 mm ID x 50 mm L using linear gradients of solvent A (heptane) and solvent B (ethanol). The term “concentrated under reduced pressure” refers to the removal of solvents and other volatile materials using a rotary evaporator at water aspirator pressure (<20 torr) while maintaining the water-bath temperature below 40 °C. Residual solvent was removed from samples at high vacuum (<0.1 torr). The term “high vacuum” refers to vacuum achieved by a mechanical belt-drive oil pump.

NMR spectra were acquired with Varian VnmrS 400 MHz (^{1}H, 400 MHz; ^{13}C, 100.6 MHz), Varian VnmrS 500 MHz (^{1}H, 500 MHz; ^{13}C, 125 MHz) or Varian VnmrS 600 MHz (^{1}H, 600 MHz; ^{13}C, 150 MHz) or spectrometers. NMR spectra were obtained at room temperatures (25°C) on samples dissolved in $CDCl_3$ unless indicated otherwise. Data are reported as follows: chemical shifts (in ppm), integration, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, m = multiplet, br = broad), and coupling constants (J, in Hz). Chemical shifts are given in ppm relative to deuterated solvent used, $CDCl_3$ (^{1}H, 7.27 ppm; ^{13}C, 77.0 ppm) with trimethylsilane as the internal standard.

Mass spectrometry (MS) for proline compounds was performed with either a Waters micromass Quattro Micro LC-MS/MS instrument using electrospray ionization (ESI). The recombinant thioredoxin samples were analyzed by ESI-MS in the range 300 – 3000 Da mass spectrometer. The m/z data was deconvoluted using MaxEnt1 Software.

For tryptic digest analysis, the digested samples were injected on a nanoAcquity UPLC coupled to a Q-Exactive mass spectrometer (Thermo).

Experimental Section

Methyl-(2S)-N-tert-butoxycarbonyl-4-oxopyrrolidine, N-Boc-(2S)-4-ketoproline methyl ester (2): Following the protocol of Loosli et al.[1], trichloroisocyanuric acid (TCCA) (0.95 g,

4.05 mmol) was added to a solution of **1,** prepared according to the protocol of Priem et al.[2] (1.0 g, 4.05 mmol) in dry CH_2Cl_2 (10 mL), and the solution was stirred and maintained at 0 °C. TEMPO (12 mg, 0.04 mmol) was added to the reaction mixture. After the addition, the mixture was warmed to rt and stirred overnight. The mixture was filtered over a Celite pad and the filter cake was washed with CH_2Cl_2 (2 x 10 mL). The organic layer was washed with Na_2CO_3, then with 0.5 M HCl and brine and was dried over Na_2SO_4. The solvent was concentrated under reduced pressure providing **2.** The crude was purified with flash chromatography using EtOAc/cyclohexane 3:7. Yield 0.9 g, 90 %;

^{1}H and ^{13}C NMR spectra show double set of peaks due to the exo and endo conformers around the tertiary carbamate.

^{1}H NMR (400 MHz, $CDCl_3$): δ 4.64 - 4.79 (dd, 1H), 3.82 - 3.88 (m, 2H), 3.72 (s, 3H), 2.83 - 2.97 (m, 1H), 2.50 - 2.57 (dd, 1H), 1.39 - 1.46 (dd, 9H).

^{13}C NMR (100.6 MHz, $CDCl_3$): δ 28.10 (*C*H_3; BOC), 41.16 (β*C*), 52.60 (O*C*H_3), 56.19 (δ*C*), 62.11 (α*C*), 81.28 (O*C*(CH_3)$_3$), 153.47 (N*C*=O), 172.17 (*C*OOCH_3), 208.24 (*C*=O).

Isolated signals of the minor conformer: δ 28.10, 40.71, 52.60, 55.48, 81.28, 154.27,171.17, 207.17

(2*S*)-N-tert-butoxycarbonyl-4-methylene proline methyl ester (3): Following the protocol of Loosli et al.[1], KOtBu (5.0 ml, 1 M in THF) and $CH_3P(Ph)_3Br$ (1.79 g, 4.95 mmol) were suspended in dry THF (10 mL) and the reaction was stirred for 5 h at rt. The solution was cooled to –10°C and **2** (0.6 g, 2.47 mmol) was added in small portions. The reaction mixture was warmed to rt and stirred overnight. To quench the reaction, a solution of aqueous 2M $KHSO_4$ (10 mL) was added and THF was concentrated under reduced pressure. The residue was extracted with EtOAc (3 x 20 mL). The organic phase was washed with brine and dried over Na_2SO_4. The solvent was concentrated under reduced pressure to afford a brown oil. The crude was purified by flash chromatography (0-15% EtOAc in pentane) to yield the product **3**. Yield 200 g; 35%.

^{1}H and ^{13}C NMR spectra show double set of peaks due to the exo and endo conformers around the tertiary carbamate.

^{1}H NMR (400 MHz, $CDCl_3$): δ: 4.99-4.95 (m, 2H), 4.40-4.35 (dd, 1H), 4.09-4.05 (s, 2H), 3.70 (s, 3H), 3.00-2.87 (m, 1H), 2.65-2.55 (m, 1H), 1.42-1.39 (s, 9H). Isolated signals of the minor conformer: δ: 5.15-4.99, 4.51-4.46, 4.05-4.01, 1.75-1.43 (9H)

^{13}C NMR (100.6 MHz, $CDCl_3$), δ: 28.3 (*C*H_3; BOC), 35.9 (β*C*), 51.2 (δ*C*H_2), 52.3 (O*C*H_3), 59.2 (α*C*), 81.2 (*C*(CH_3)$_3$), 107.7 (=*C*H_2), 142.3 (*C*H=CH_2), 156.0 (N*C*=O), 173.4

(*C*OOCH$_3$)). Isolated signals of the minor conformer: δ 28.4, 36.7, 50.5, 52.2, 58.3, 59.4, 81.2, 107.7, 143.0, 156.0, 173.4.

(2*S*, 4*S*)-1-tert-Butoxycarbonyl-2-carboxymethyl, 4-methyl-pyrrolidine (4) and (2*S*, 4*R*)-1-tert-Butoxycarbonyl-2-carboxymethyl, 4-methyl-pyrrolidine (5): 3 (200 mg, 0.83 mmol) was dissolved in MeOH (10 ml) and Pd/C 5% (40 mg) was added to the solution. The mixture was stirred under hydrogen atmosphere overnight. The reaction mixture was filtrated through Celite pad. The solution recovered was evaporated under reduced pressure to afford an oil. The mixture of two diastereoisomers was analyzed by NMR as previously reported by Sun et al,[3] and by analytical HPLC with chiral column Daicel Chiralcel® OJ-H to calculate the d.r.

Methyl (2*S*)-N-(tert-butoxycarbonyl)-4-(dimethylaminomethylene)pyroglutamate (7): According the protocol reported by Coudert et al.,[4] the tert-butoxybis(dimethylamino)methane (Bredereck's reagent) (1.2 g, 4.96 mmol) was added to a mixture of **6** (1.0 g, 4.13 mmol) and toluene (10.0 mL), then the mixture was heated at 100 °C overnight. The solvent was removed under reduced pressure, and the solid residue was purified by flash chromatography [CH_2Cl_2/EtOAc (0-50% CH_2Cl_2)], afforded pale yellow powder (0.8 g, 65%). TLC: EtOAc/CH_2Cl_2 (8/2): *Rf* 0.50.

^{1}H NMR (400 MHz, $CDCl_3$) δ: 1.49 (s, 9H), 2.84-2.96 (m, 1H), 3.02 (s, 6H), 3.19-3.32 (m, 1H), 3.75 (s, 3H), 4.55 (dd, J = 3.8, 10.5 Hz, 1H), 7.10 (t, 1H)

^{13}C NMR (100.6 MHz, $CDCl_3$) δ: 26.49 (β*C*), 27.82 (*C*H$_3$; BOC), 41.76 (N(*C*H$_3$)$_2$), 52.35 (O*C*H$_3$), 55.70 (α*C*), 82.41 (NCH=*C*), 90.95 (*C*(CH$_3$)$_3$), 146.55 (N*C*H=C), 150.47 (N*C*=O), 169.34 (N*C*=O; BOC), 172.71 (*C*OOCH$_3$))

Methyl (2*S*,4*S*)-N-(tert-butoxycarbonyl)-4-methylpyroglutamate (8):

A stainless steel hydrogenation reaction high pressure reactor was charged with a mixture of **7** (0.35 g, 1.17 mmol) and 10% Pd/C (35 mg) in iPrOH (25.0 mL) and AcOEt (5.0 mL) was stirred under an atmosphere of H_2 at 200 psi for 72 h at rt. The reaction mixture was filtrated through Celite pad. Evaporation of the solvent afforded an oil, which was purified by flash chromatography [n-hexane:AcOEt, 3:2] to give the isomer **8** (240 mg, 80%) as a white solid:

^{1}H NMR (500 MHz, $CDCl_3$) δ: 1.25 (d, J = 6.8 Hz, 3H), 1.49 (s, 9H), 1.58-1.66 (m, 1H), 2.51-2.64 (m, 2H), 3.77 (s, 3H), 4.49 (dd, 1H)

^{13}C NMR (125 MHz, $CDCl_3$) δ: 16.2 (*C*H_3), 28.0 (3*C*H_3, Boc), 29.6 (β*C*), 37.5 (γ*C*), 52.3 (O-*C*H_3), 57.3 (α*C*), 83.6 (O-*C*-$(CH_3)_3$), 149.4 (N*C*=O, Boc),172.0 (γC=O), 175.3 (*C*OOCH_3)

The isomer **methyl (2*S*,4*R*)-N-(tert-butoxycarbonyl)-4-methylpyroglutamate** was afforded with a further elution with the same solvent system (3 mg, 1.3%) as a colorless oil:
^{1}H NMR (500 MHz, $CDCl_3$) δ: 4.57 (dd, J = 1.3, 9.6 Hz, 1H), 3.78 (s, 3H), 2.45-2.72 (m, 1H), 2.36-2.24 (m, 1H), 2.07-1.99 (m, 1H), 1.49 (s, 9H), 1.25 (d, 3H).
^{13}C NMR (125 MHz, $CDCl_3$): 16.50, 27.80, 29.48, 30.49, 52.54, 56.94, 83.47, 149.57, 173.17, 175.56

(2*S*, 4*S*)-1-tert-butoxycarbonyl-2-carboxymethyl, 4-methyl-pyrrolidine (4): Compound **8** (150 mg, 0.78 mmol) was dissolved in dry THF (10 mL) and the solution cooled at 0 °C. 1 mL of a BH_3Me_2S solution (1 M solution in THF, 2.03 mmol) was added and the solution stirred at 40 °C overnight. The solvent was evaporated under reduced pressure and the crude oily product was dissolved in ethyl acetate (20 ml). The organic layer was washed with water (2 × 10 ml) and saturated aqueous sodium chloride (10 ml), dried on Na_2SO_4), and filtered. The solvent was concentrated under reduced pressure. The mixture was purified by flash chromatography (eluent n-pentane:AcOEt 7: 1) to give 90 mg of **4** (60 % yield).
^{1}H and ^{13}C NMR spectra show double set of peaks due to the exo and endo conformers around the tertiary carbamate.
^{1}H NMR (600 MHz, $CDCl_3$): δ: 1.05 (d, J = 6.6 Hz, 3H), 1.39 (s, 9H), 1.54 (m, 1H), 2.21 (m, 1H), 2.39 (m, 1H), 2.97 (m, 1H), 3.73 (m, 1H), 3.71 (s, 3H), 4.18 (dd, AJ = 9.2 Hz, BJ = 7.2Hz, 1H). Isolated signals of the minor conformer: δ: 1.04, 1.44, 1.51, 2.24, 2.38, 2.96, 3.65, 3.725, 4.25
^{13}C NMR (150 MHz, $CDCl_3$) δ: 16.88 (*C*H_3), 28.26 (3*C*H_3, Boc), 32.60 (γ*C*), 38.97 (β*C*), 51.87 (*C*H_3), 53.28 (δ*C*), 59.73 (α*C*), 79.87 (O-*C*-$(CH_3)_3$), 153.51 (N*C*=O, Boc), 173.81 (*C*OOCH_3). Isolated signals of the minor conformer: δ 17.00, 28.44, 33.53, 38.03, 52.05, 53.74, 59.23, 79.81, 154.23, 173.56
The by-product was isolated and analysed by mass spectroscopy and NMR (Fig. S10). The mass found was consistent with (2*S*, 4*S*)-methylprolinol.
^{1}H NMR (400 MHz, $CDCl_3$) δ: 0.90-1.00 (d, 3H), 1.00-1.05 (m, 1H); 1.45 (s, 9H), 2.00-2.18 (m, 2H), 2.65-2.80 (t, 1H), 3.50-3.60 (t, 1H), 3.60-3.75 (m, 2H), 3.90 (m, 1H), 5.30 (d, 1H)

(2*S*, 4*S*)-4-methyl-pyrrolidin-2-carboxylic acid, (2*S*, 4*S*)-4-methyl-proline (9):

To a solution of compound **4** (70 mg, 0.29 mmol) in THF (2.5 mL) LiOH (25.0 mg, 0.58 mmol in 1 mL of H_2O) was added and the mixture was stirred at rt overnight. The solvent was evaporated under reduced pressure and the crude residue was dissolved in water (2 mL). HCl 0.5 N was added dropwise until pH=6 was reached and a precipitation of a white solid was observed that was extracted with CH_2Cl_2 (5 mL x 3 times). The organic layers were dried over Na_2SO_4, then the solvent was removed under reduced pressure. The residue was dissolved in a solution of CH_2Cl_2/TFA (5 ml). After stirring for 3 h at rt, the solvent was evaporated under reduced pressure and the crude product was triturated with Et_2O to obtain a white powder of **9** TFA salt. Yield: 57 mg; 81.9%.

^{1}H NMR (500 MHz, D_2O) δ: 0.96 (d, J=7, 3H), 1.60 (m, AJ=13 Hz, BJ= 20 Hz, 1H), 2.35 (m, AJ= 7 Hz, 1H), 2.48 (m, AJ = 13 H, 1H), 2.83 (dd, AJ= 10 Hz, BJ= 12 Hz 1H), 3.39 (dd, AJ = 10 Hz, BJ = 7 Hz, 1H), 4.24 (dd, AJ = 10 Hz, BJ = 8.5 Hz, 1H).

^{13}C NMR (125 MHz, D_2O) δ: 15.7 (*C*H_3), 32.9 (γ*C*), 36.2 (β*C*), 51.8 (δ*C*), 60.3 (α*C*), 173.0 (*C*OOH)

Incorporation of 4-methylproline isomers

Incorporation of (2*S*, 4*S*)- and (2*S*, 4*R*)-methylproline into Trx1P

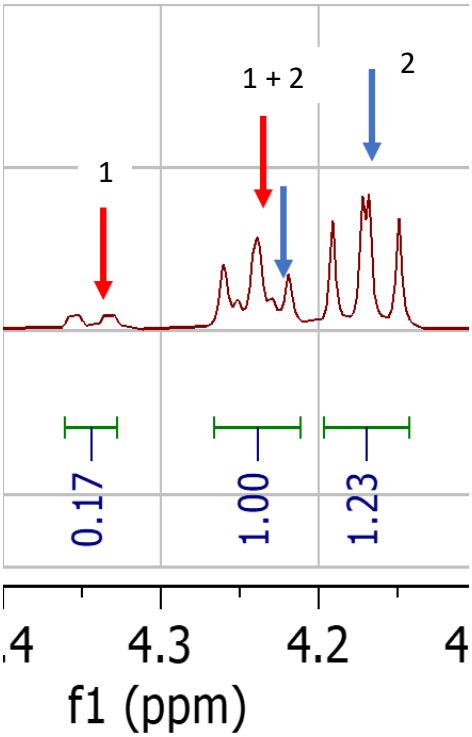
A single CAG18515 *E. coli* cell colony co-transformed with pGDR11-Trx1P and pTARA-ProRS(C443G) was grown in 50 mL LB, with ampicillin and chloramphenicol, overnight at 37 ºC. The overnight cell culture was centrifuged at 4500 rpm for 15 min. The cell pellet was collected and resuspended in 1 L M9 MM (0.3 mM Pro, ampicillin 100 µg/mL and chloramphenicol 34 µg/mL). Arabinose (0.1 % w/v) was also added to the medium, to induce expression of ProRS(C443G). The culture was incubated at 37 ºC, until a plateau of growth was reached. After centrifugation at 4000 rpm for 15 min, the cell pellet was washed twice with 0.9 % NaCl (50 mL). The cell pellet was then resuspended in 1 L M9 MM (0.6 M NaCl, ampicillin 100 µg/mL and chloramphenicol 34 µg/mL, no Pro). The cell culture was incubated for 20 min at 37 °C. The non-natural Pro analogue (2.5 mM final concentration) was then added to the culture and it was incubated for further 20 min. IPTG (1 mM final concentration) was added and expression was allowed to proceed for 3 h at 37 ºC. Cells were harvested by centrifugation (4000 rpm) and the pellet was suspended in 30 mL 20 mM Tris/HCl pH 7.5 and disrupted by sonication. The lysate was centrifuged (10000 rpm, 40 min, 4 °C) and the supernatant was applied to a Thermo Scientific HisPur

Ni-NTA Spin Columns (3 mL). The protein was eluted by increasing the concentration of imidazole. Fractions containing thioredoxin were pooled, concentrated to 2 mL by ultra-filtration (10 kDa membrane, Amicon) and dialyzed against 20 mM Tris/HCI pH 7.5 buffer and stored at -20°C.

Figures

Supplementary Figure 1

a) NMR of diastereomeric mixture



(2*S*;4*S*) Methylproline with Endo and Exo Pucker
(2*S*;4*R*) Methylproline with Endo and Exo Pucker

4.15 ppm (2*S*;4*S*) Methylproline (Endo) with area 1.23;
4.25 ppm (2*S*;4*S*) Methylproline (Exo) + (2S;4R) Methylproline (Endo) with total area 1.00;
4.35 ppm (2*S*;4*R*) Methylproline (Exo) with area 0.17;

$$\begin{cases} 0.17 \div x = y \div 1.23 \\ x + y = 1 \end{cases} \rightarrow x = 0.7 \text{ ppm}$$

$$\frac{Integral\ 1}{Integral\ 1 + Integral\ 2} = \frac{0.17 + 0.7}{0.17 + 1 + 1.23} \times 100 = 36$$

b) Analytical HPLC

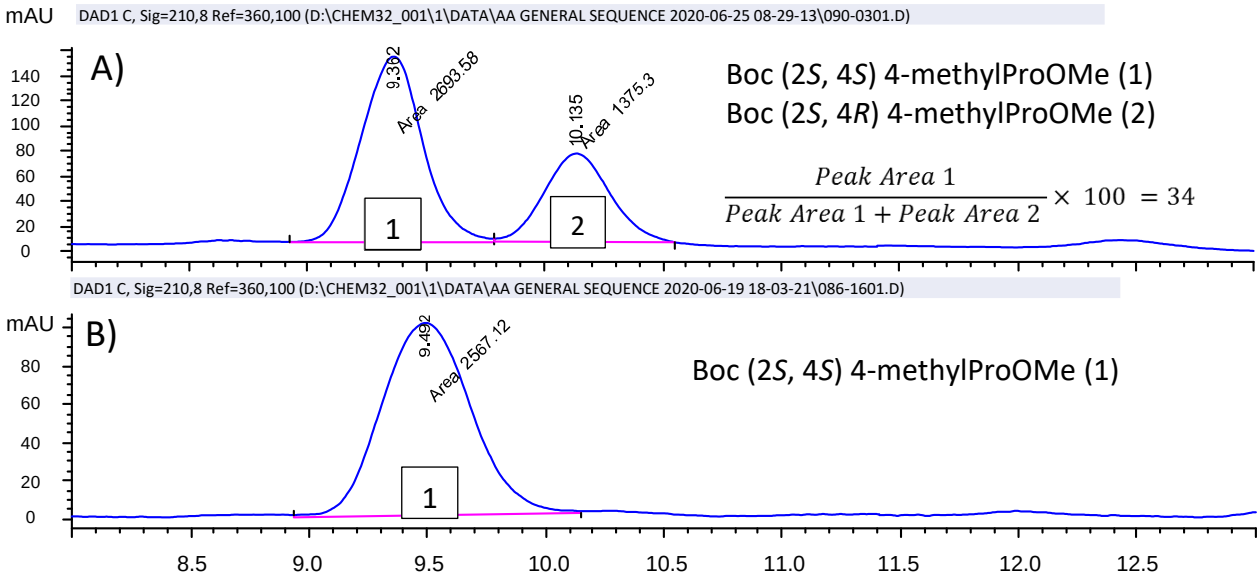


Figure S1: Determination of 4-Mep diastereomers by ^{1}H-NMR and analytical HPLC chiral column a) ^{1}H NMR analysis of Boc (2*S*, 4*R*,*S*)-methylproline methyl ester diasteroisomers

mixture. The analysis of spin system of proton alpha of proline is reported. b) HPLC analysis of the Boc (2*S*, 4*R*,*S*)-methylproline methyl ester diasteroisomers mixture using chiral column Daicel Chiralcel® OJ-H). The mixture of the two 4-Mep diastereoisomers obtained from the synthetic route starting from 4-hydroxyproline is shown. B) Boc (2*S*; 4*S*)-4-methylProOMe obtained from the synthetic route starting from pyroglutamic acid. Only one diastereisomer is detected.

Supplementary Figure 2

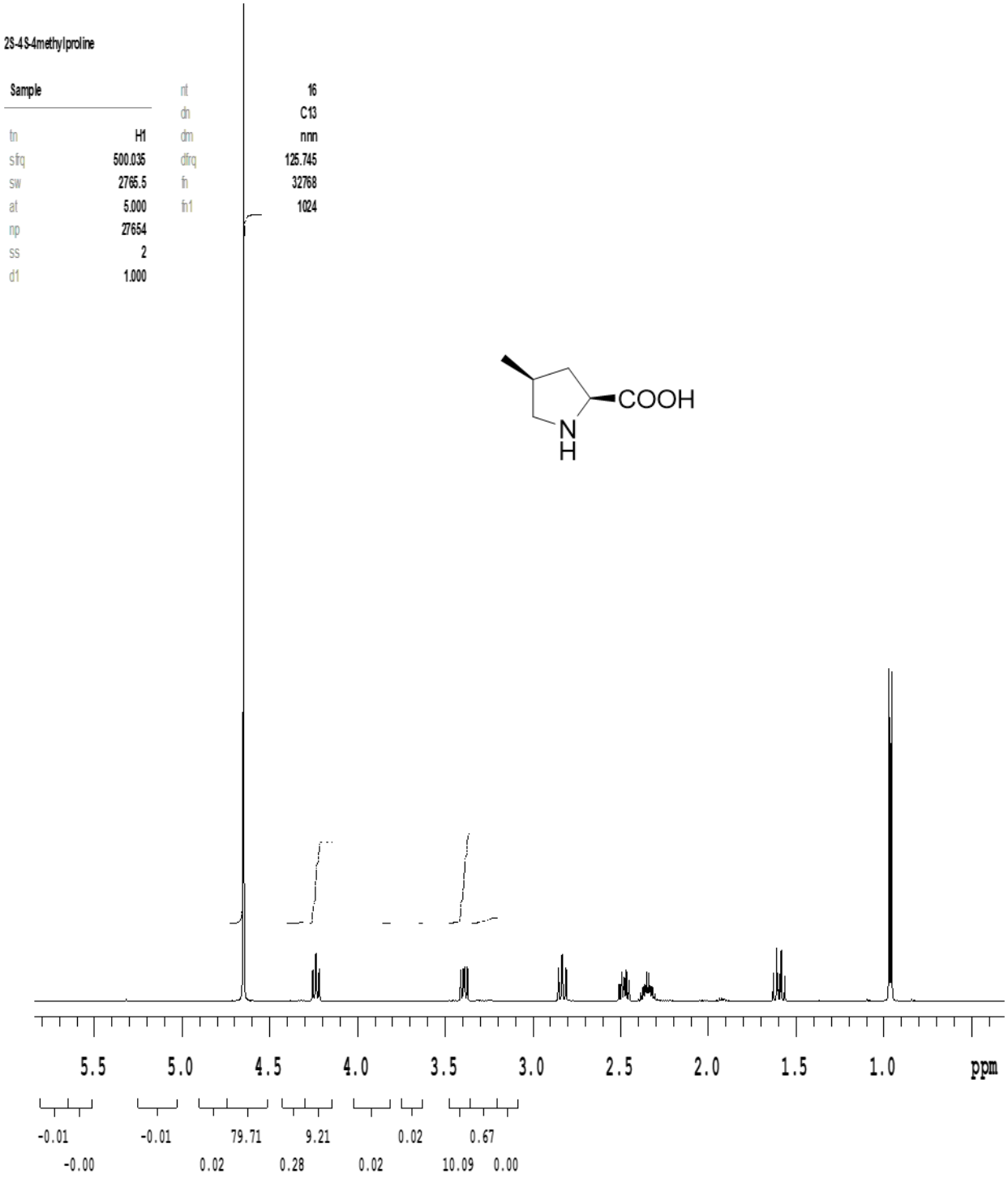


Figure S2: ^{1}H-NMR of (2*S*, 4*S*)-4-methyl-proline **(9)**

Supplementary Figure 3

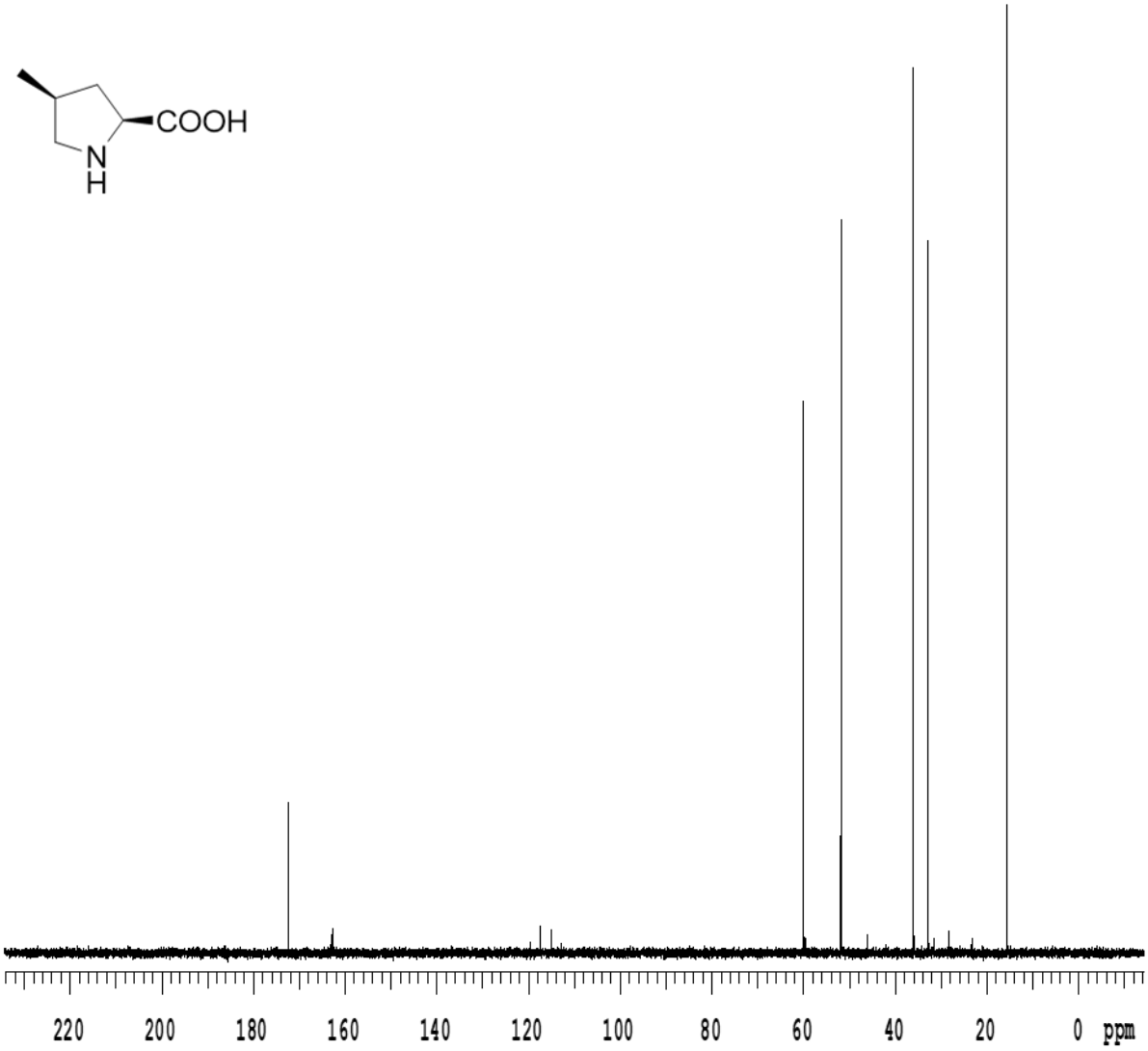


Figure S3: 13C-NMR of (2*S*, 4*S*)-4-methyl-proline (**9**)

Supplementary Figure 4

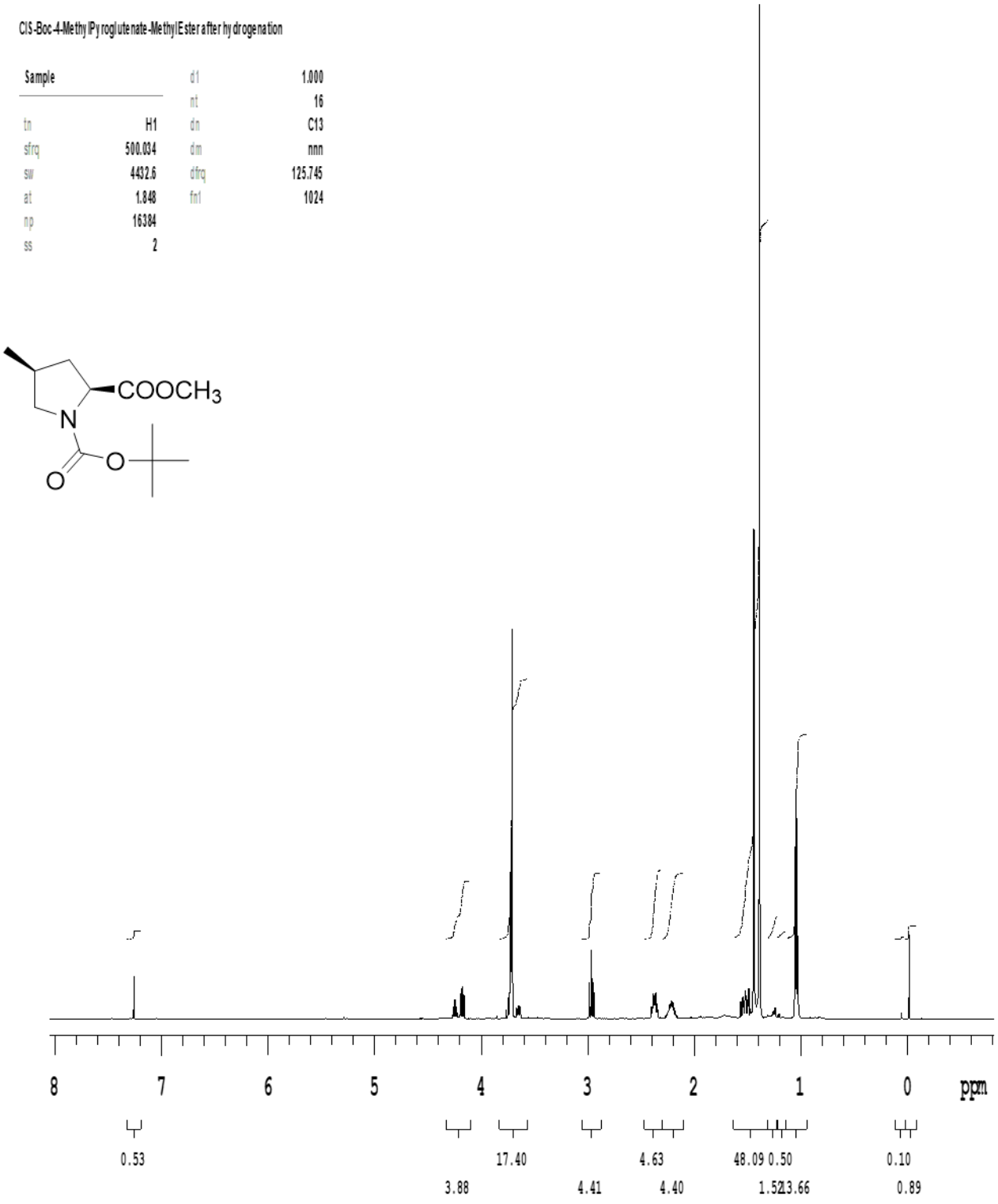


Figure S4: 1H-NMR of (2*S*, 4*S*)-1-tert-butoxycarbonyl-2-carboxymethyl, 4-methyl-pyrrolidine (**4**)

Supplementary Figure 5

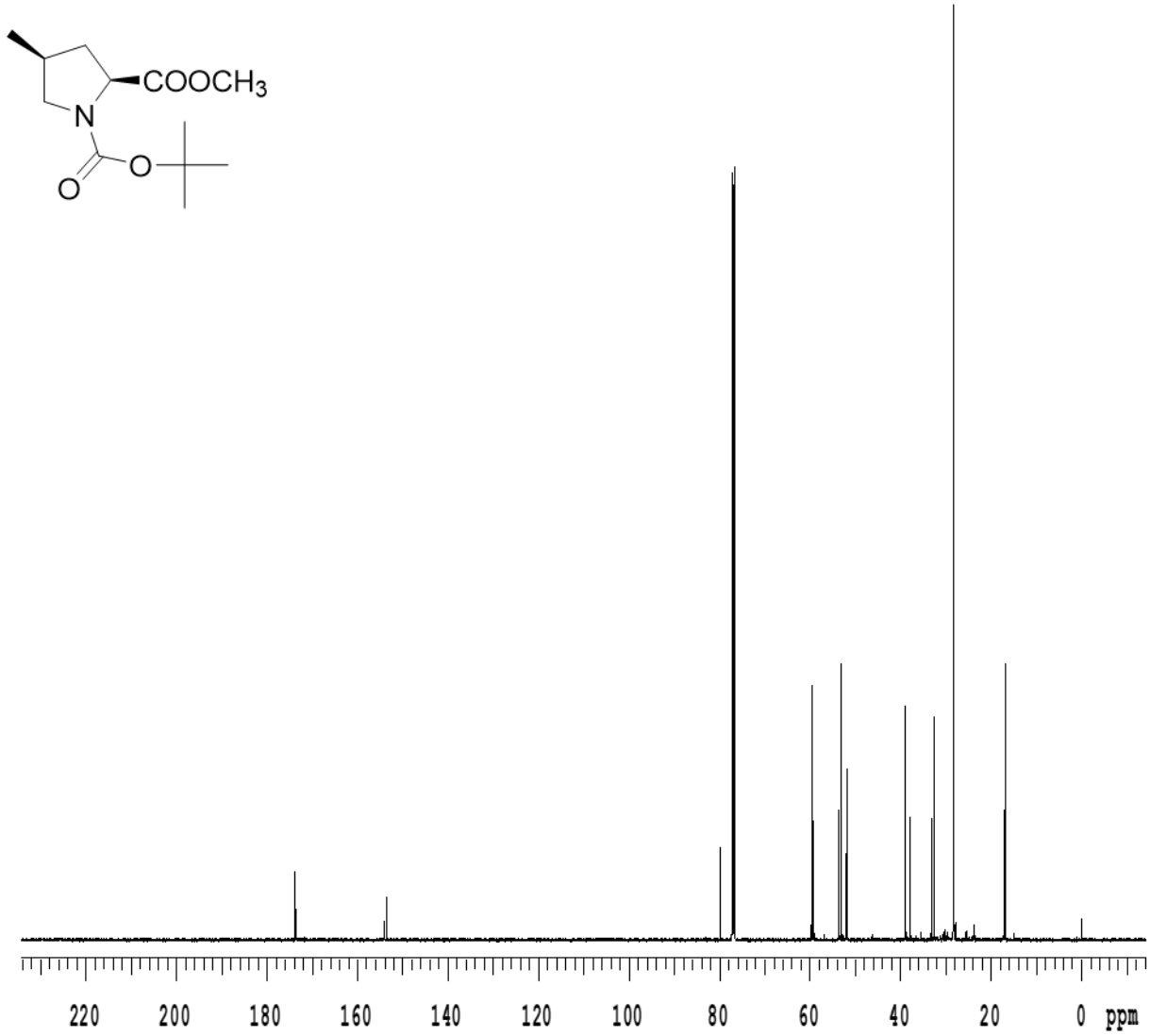


Figure S5: 13C-NMR of (2*S*, 4*S*)-1-tert-butoxycarbonyl-2-carboxymethyl, 4-methyl-pyrrolidine (**4**)

Supplementary Figure 6

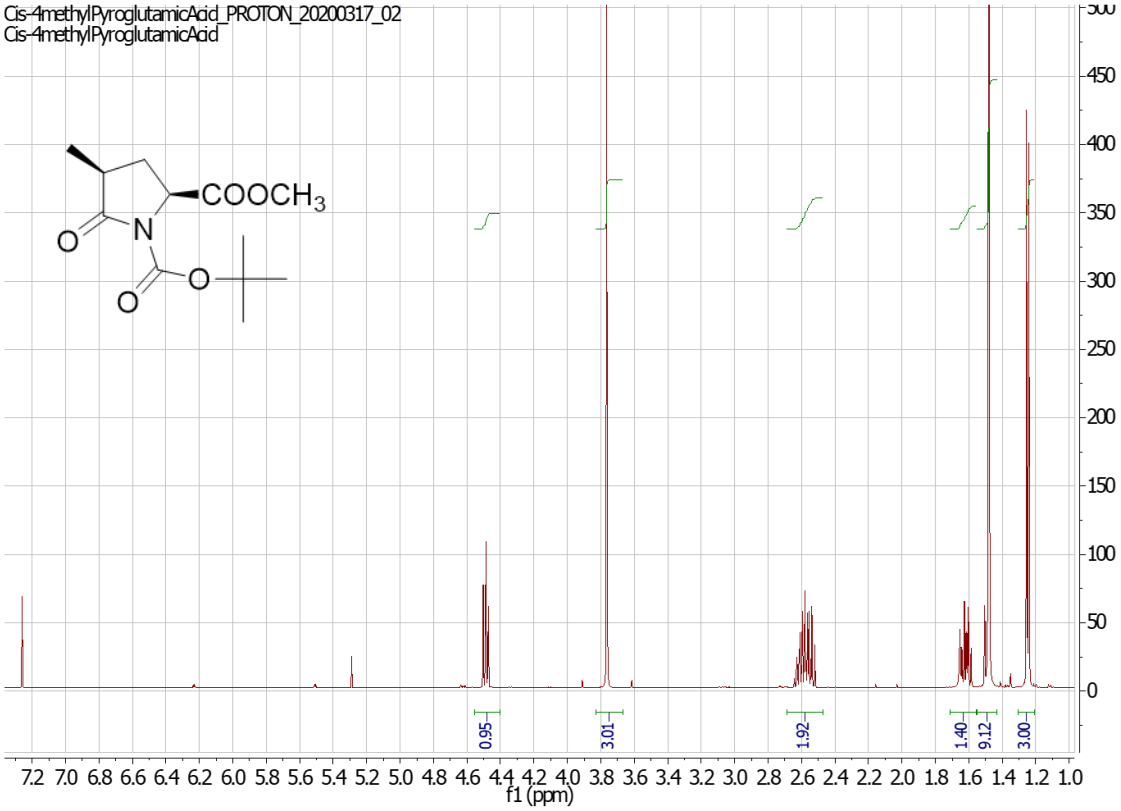


Figure S6: 1H-NMR of Methyl (2*S*,4*S*)-N-(tert-butoxycarbonyl)-4-methylpyroglutamate (**8**)

Supplementary Figure 7

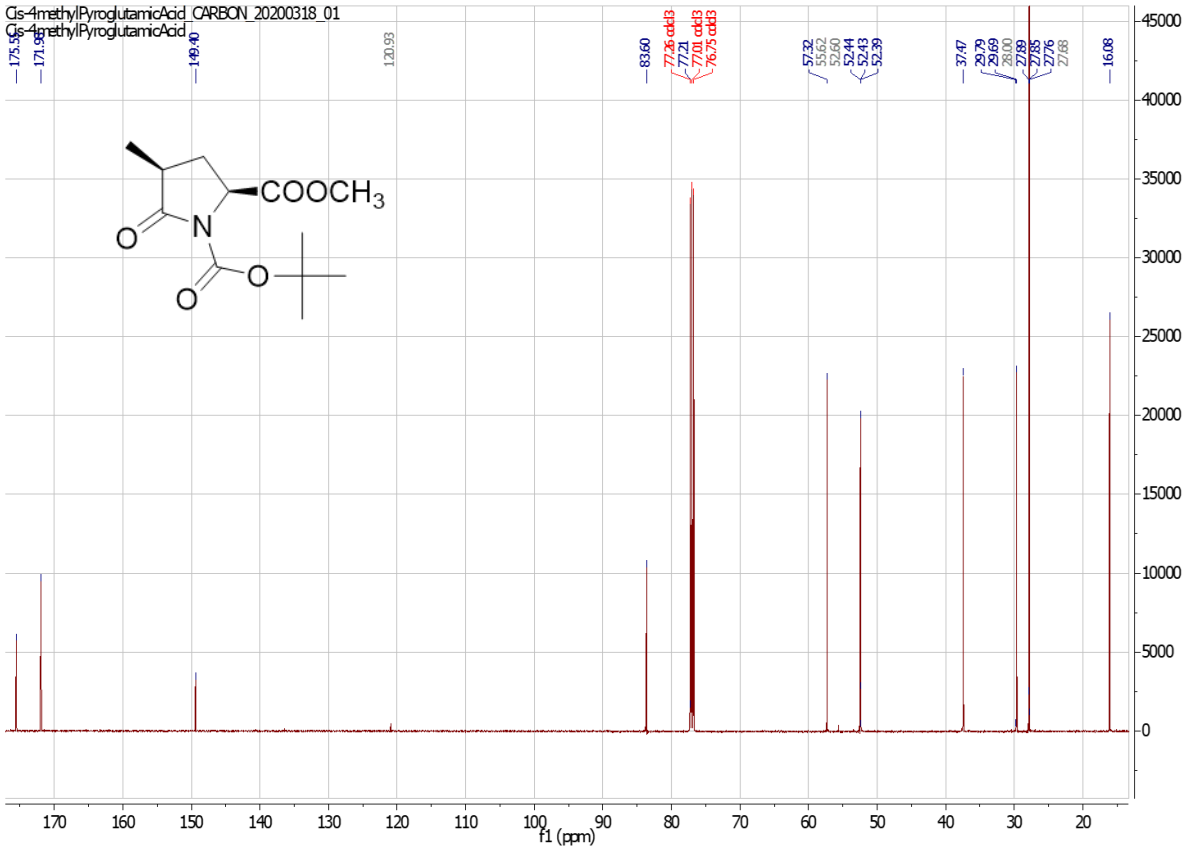


Figure S7: 13C-NMR of Methyl (2*S*,4*S*)-N-(tert-butoxycarbonyl)-4-methylpyroglutamate (**8**)

Supplementary Figure 8

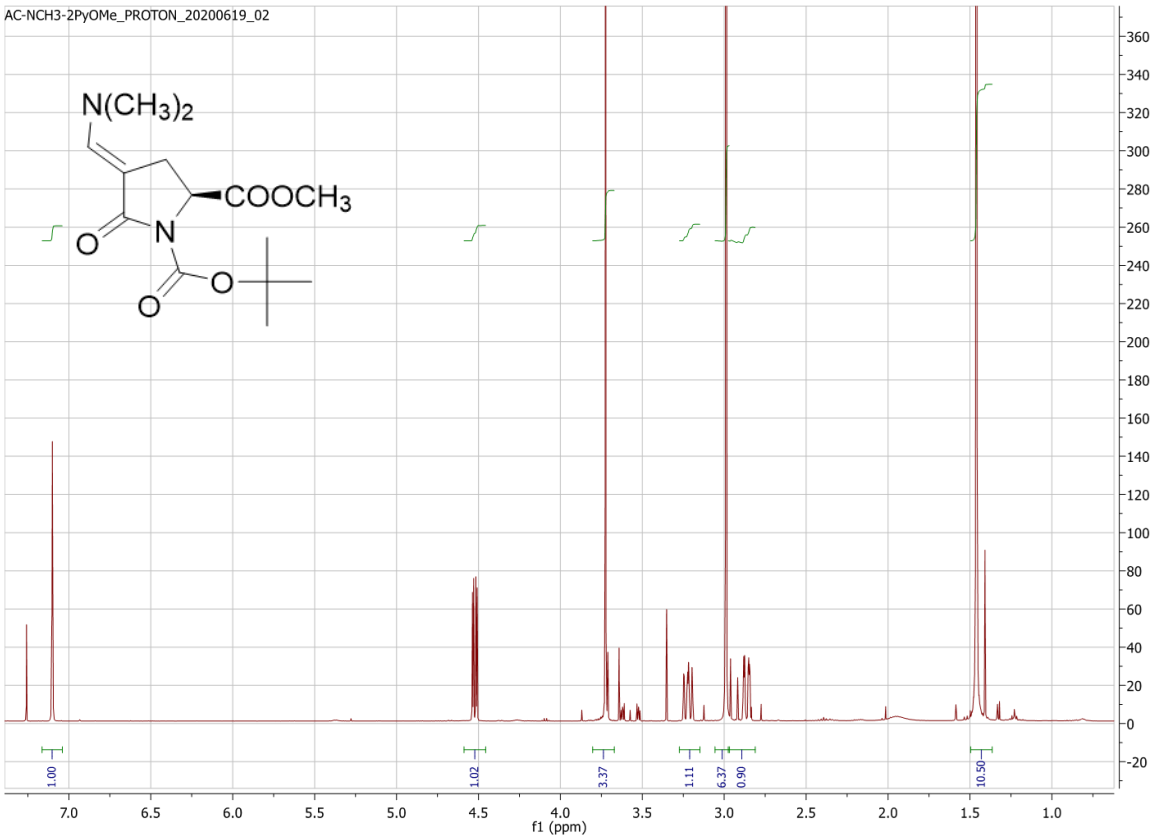


Figure S8: 1H-NMR of Methyl (2*S*)-N-(tert-butoxycarbonyl)-4-(dimethylaminomethylene)pyroglutamate (**7**)

Supplementary Figure 9

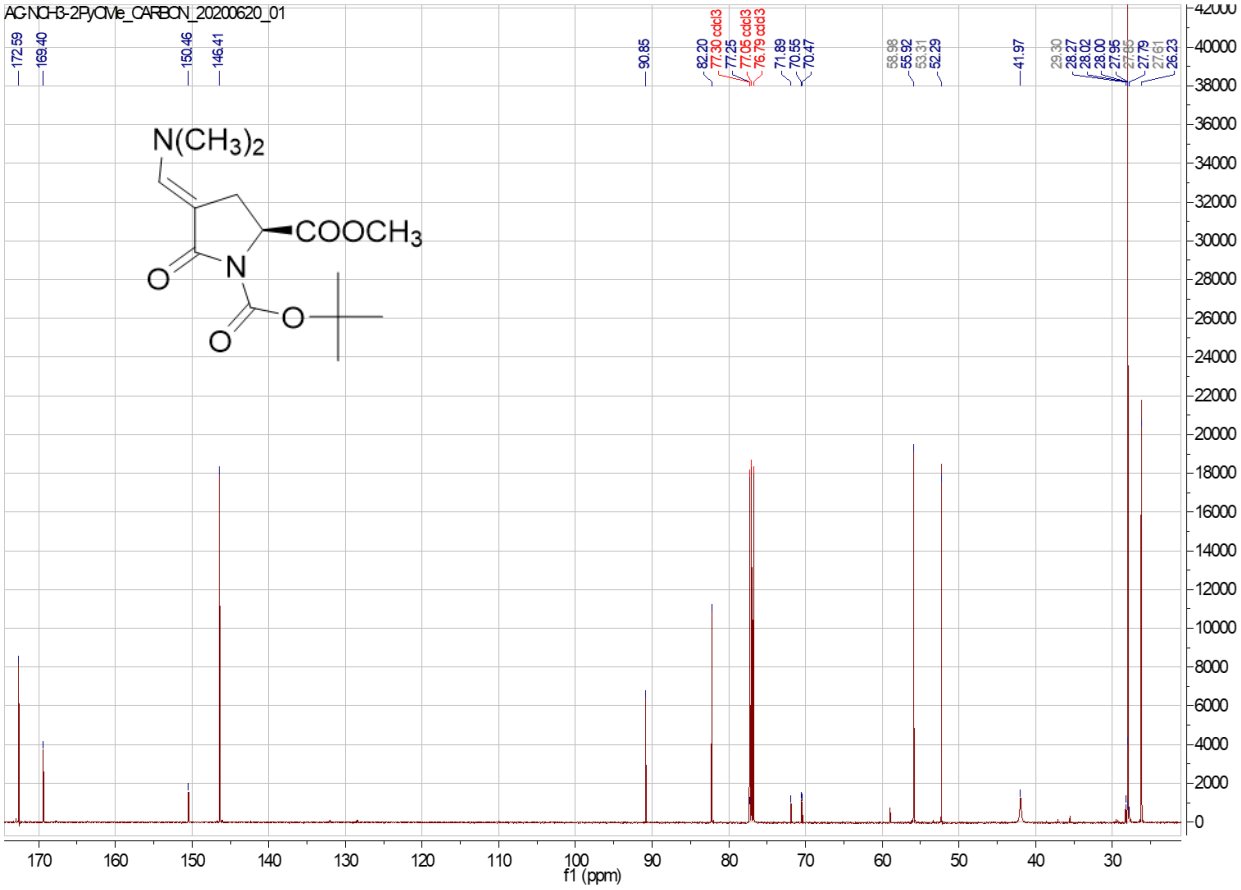
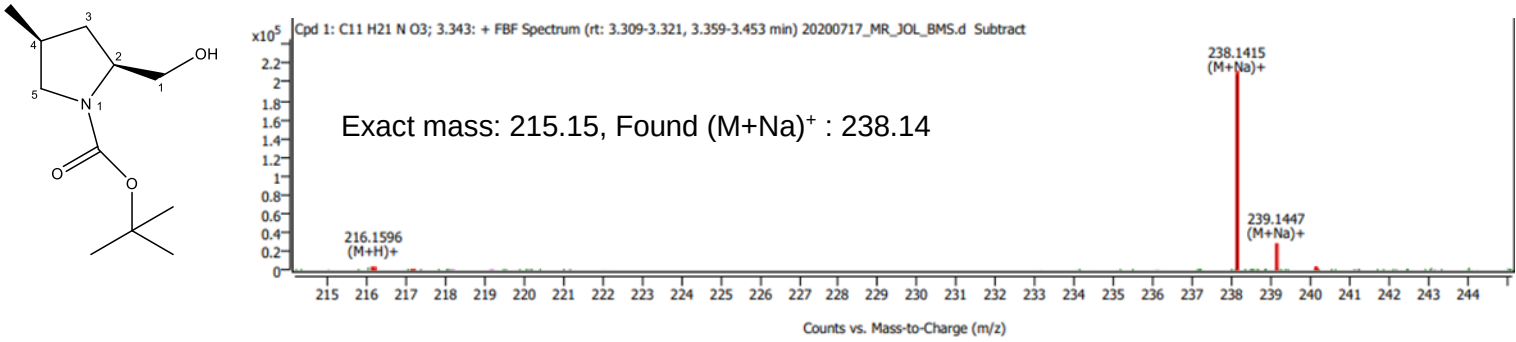


Figure S9: 13C-NMR of Methyl (2*S*)-N-(tert-butoxycarbonyl)-4-(dimethylaminomethylene)pyroglutamate (**7**)

Supplementary Figure 10

a)



b)

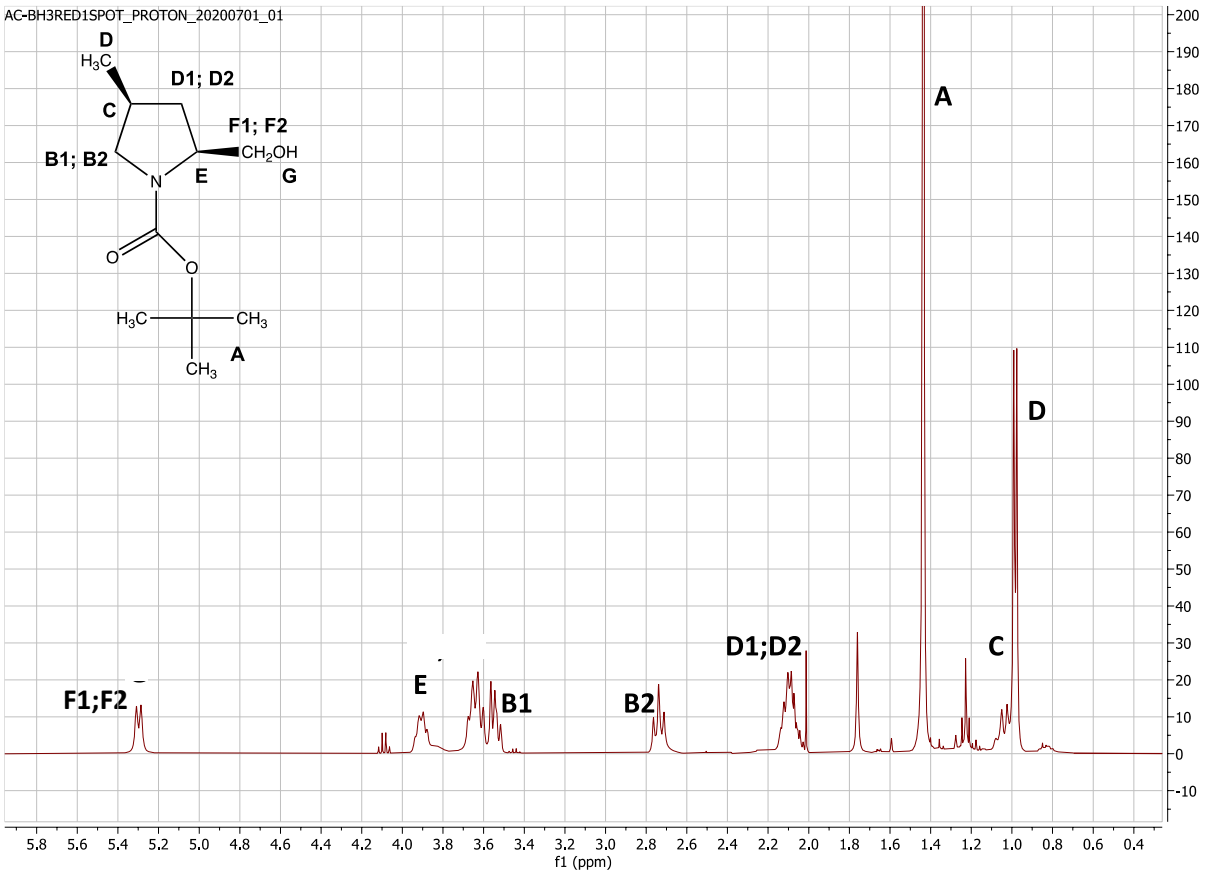


Figure S10: Characterization of by-product isolated during reduction of 4*S*-methylpyroglutamate **8** with BH_3Me_2S a) Mass spectrometry analysis b) ^{1}H-NMR spectrometry analysis

REFERENCE

1. S. Loosli, C. Foletti, M. Papmeyer and H. Wennemers, *SynLett*, 2019, **30**, 508-510.
2. C. Priem and A. Geyer, *Org Lett*, 2018, **20**, 162-165.
3. K. Sun, C. Tao, B. Long, X. Zeng, Z. Wu and R. Zhang, *RSC Advances*, 2019, **9**, 32017-32020.
4. E. Coudert, F. Acher and R. Azerad, *Synthesis-Stuttgart*, 1997, 863-&.