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

Supporting Information

Direct Monitoring of Biocatalytic Deacetylation Reactions by 1H NMR Reveals Fine Details of Substrate Specificity

Silvia De Cesare, Catherine A. McKenna, Lorna Murray, Juraj Bella, Nicholas P. Mulholland and Dominic J. Campopiano*

Table of Contents

General Information 2

Protocols 2

Schemes, Figures and Tables 4

Substrate Synthesis 12

^{1}H and ^{13}C NMR spectra 15

ArgE catalysed hydrolysis of *N*-acetyl proteinogenic amino acids: ^{1}H NMR spectra 26

ArgE-catalysed hydrolysis of *N*-Ac-NCAAs 36

General Information

All reactions were performed in Eppendorfs, NMR tubes or round-bottom flasks, according with the final volume of the reaction or the kind of experiment performed, as specified in the Protocols section. The NCAAs used in this study, along with *N*-acetyl-L-ornithine (**1a**) and *N*-acetyl-L-methionine sulfone (**1c**) were a gift from Syngenta (Bracknell, UK). *N*-acetyl-L-proteinogenic amino acids, *N*-acetyl-L-methionine sulfoxide (**1b**), L-proteinogenic amino acids, cobalt chloride and all other chemicals were obtained from Sigma-Aldrich unless otherwise stated. Protein concentrations were calculated using a nanovue spectrophotometer (nanodrop, GE). Proton nuclear magnetic resonance (^{1}H NMR) and carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded on Ava-400 (400 MHz), or Ava 500 (500MHz) spectrometers. Proton and carbon chemical shifts are reported in parts per million downfield from tetramethylsilane and are referenced to residual proton in the NMR solvent (D_2O = δ 4.79). NMR data are represented as follows: chemical shift, multiplicity (, s = singlet, d = doublet, t = triplet, dd = doublet of doublets, ddd = doublet of doublets of doublets, ddt = doublet of doublets of triplet, q = quartet, m = multiplet), coupling constant in Hertz (Hz), integration. Mass spec analysis (MS) was carried out on a liquid chromatography-mass spectrometer Bruker Daltonics-micrOTOF (ESI-MS) in the University of Edinburgh Mass Spec facility, School of Chemistry. Proteins were purified on an ATKA Purifier (GE Healthcare) using immobilized metal affinity chromatography (HisTrap™ Fast Flow 1 mL, GE Healthcare) and size exclusion chromatography (HR S200 column, GE Healthcare). All curve fittings presented in this report were carried out using Microsoft Excel, GraphPad, or Origin Lab. NMR spectra were processed with the Mnova (Mestrelab Research).

Abbreviations used: Ac = acetyl, DMF = *N,N*-dimethylformamide, IMAC = immobilized metal affinity chromatography, NaPi = sodium phosphate buffer and r.t. = room temperature.

Protocols

Cloning. The *E. coli* ArgE synthetic gene was ordered from GenScript and cloned into a pETHisTEV vector using the NcoI and XhoI restriction site to obtain an N-terminus 6xHis TEV cleavable tagged construct. Recombinant ArgE was prepared as described below.

ArgE expression and purification. An overnight culture of *E. coli* BL21 (DE3) cells transformed with pETHisTEV-ArgE was used to inoculate 500 mL of LB broth (30 µg mL^{-1} kanamycin) and left to incubate at 37°C with shaking until an OD_{600} of ~0.6-0.8 was reached. The cells were then induced with 0.2 mM IPTG and 0.1 mM $ZnSO_4$ at 20 °C for 20 hours. Cells were harvested by centrifugation (4,000 rpm, 10 minutes) and re-suspended immediately in 10 mL of lysis/ buffer (50 mM Tris pH 8.0, 100 mM NaCl, 20 mM imidazole). This homogenous solution was then sonicated on ice (30s on/30s off) for 15 minutes. The supernatant was then clarified by centrifugation (12,000 rpm, 60 minutes, at 4 °C), and filtered (0.45 µm). Purification was carried out by Ni^{2+} affinity chromatography (HisTrap™ FF column, GE Healthcare) on a linear gradient of imidazole and size exclusion chromatography (HR S200 column, GE Healthcare). An extinction coefficient of 43360 M^{-1}cm^{-1}, calculated from the amino acid sequence with the ProtParam tool online software, was used to

calculate the protein concentration with the nanodrop (Abs at 280 nm). Elution buffer 50 mM Tris pH 8.0, 100 mM NaCl, 500 mM imidazole and storage buffer 50 mM Tris pH 8.0, 100 mM NaCl.

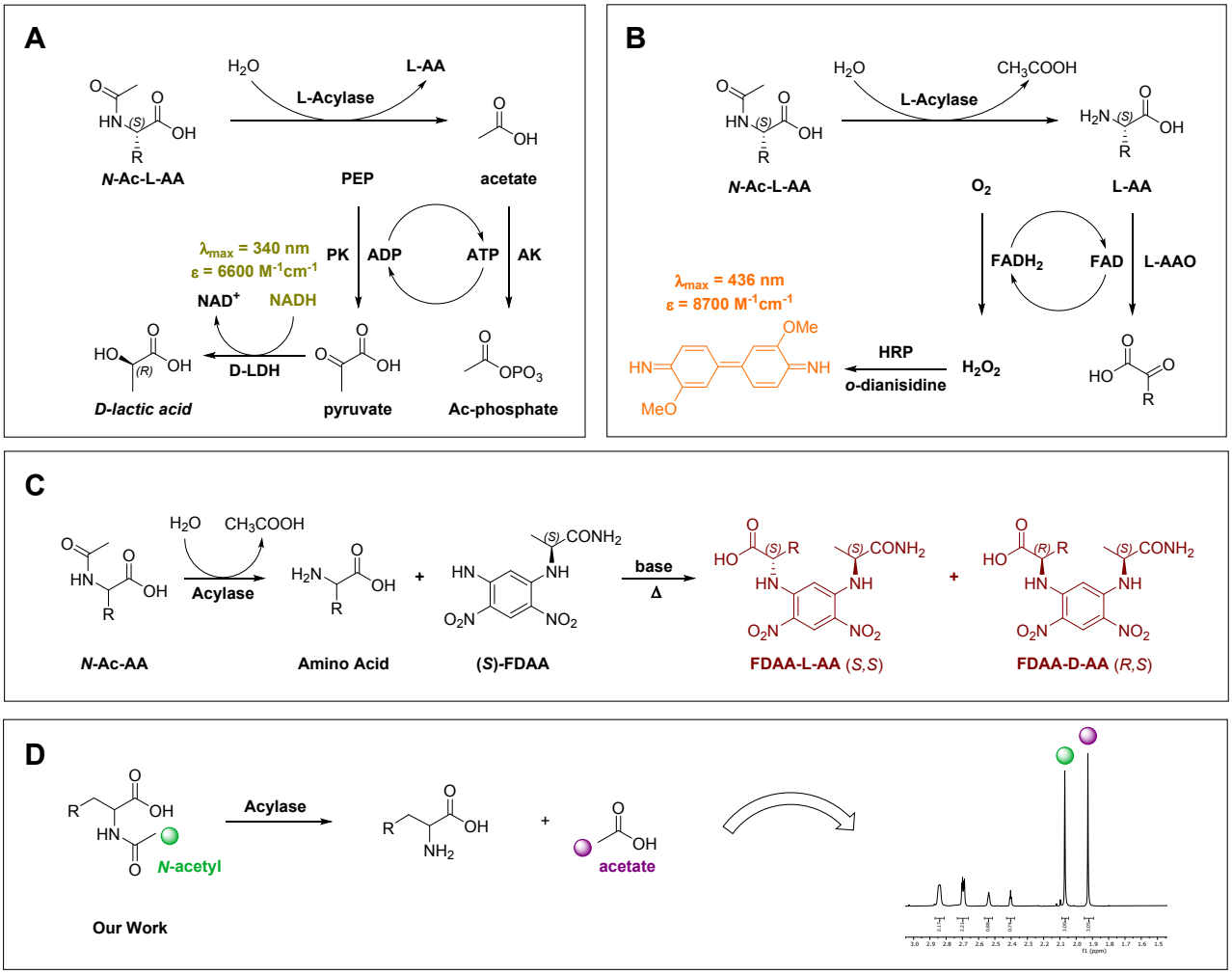
^{1}H NMR based assay. 500 μL reactions were set up in Eppendorf tubes using various *N*-acetyl-L-AAs (20 mM), *E. coli* ArgE 0.25 mg/mL and $CoCl_2$ 100 μM in 100 mM NaPi buffer pH 8.0. The reactions were incubated at 40 °C, 250 rpm for 24 h. 20 μL $HCl_{conc.}$ were added to the reaction mixture to precipitate the enzyme, the samples were centrifuged for 10 min at 13000 rpm at r.t. and the supernatant analyzed *via* ^{1}H NMR and LC-MS. ^{1}H NMR samples were prepared by adding 100 μL of D_2O to 400 μL of reaction mixtures. Samples for LC-MS characterization were prepared by dissolving 5-10 mg materials in 1 mL MS grade water or methanol, while for reaction monitoring 50 μL of reaction mixture were diluted in 50 μL of H_2O and AcCN 1:1 mixture (MS grade solvents) and filtered before injection. A single EspriSat experiment for water (16 scans, solvent peak suppression) was employed to record all NMR spectra, which were processed with the Mnova (Mestrelab Research). Reaction conversion (%) was calculated using the NMR peak integrals of methyl protons of the product acetate (I_P) and starting material amide (I_R) with the equation below. In the case of *N*-acetyl-Gln where the signals for the methyl protons could not be distinguished, we instead used the signals from the C-α proton (~4.2 ppm and ~3.75 ppm). Where the starting amide could not be detected we set this conversion as >99.0%.

$$Conv\ (\%) = \frac{(I_P)}{(I_P + I_R)} \cdot 100$$

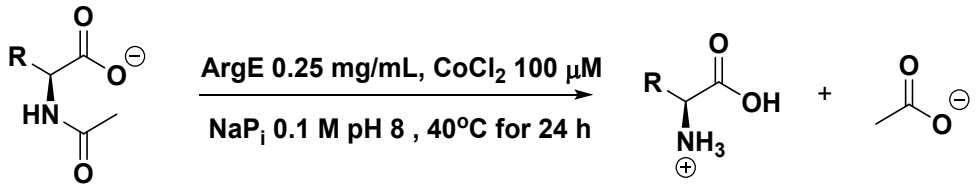
^{1}H NMR reaction monitoring. 500 μL reactions were set up in an NMR tube using *N*-acetyl-L-Val (30 mM), *E. coli* ArgE 0.50 mg/mL and $CoCl_2$ 100 μM in D_2O at either 25 °C or 40 °. The tubes were inserted in an AVA 400 (400 MHz) NMR, setting the probe at the desired temperature for the reaction. The ^{1}H NMR spectrum was recorded every 10 minutes for 2.5 h.

Coupled, colourimetric L-AAO assay. The kinetic analysis was carried out with a reaction mixture containing various concentrations of *N*-acyl- amino acid substrate (**1a-j**, typically 0-50 mM, see Fig. S5 for details), ArgE 0.1 mg/mL (2.0μM), $CoCl_2$ 100 μM, Horseradish Peroxidase (HRP, Sigma Aldrich) 10 U/mL, L-Amino Acid Oxidase (LAAO from *Crotalus Atrox,* Sigma Aldrich) 100-250 mU/mL and *o*-dianisidine 0.1 mg/mL in 100 mM NaPi buffer pH 8.0. The reactions were incubated at 40°C and the absorbance at 436 nm was measured at intervals of 30 s for 1 hour. The concentration of L-AAO was adjusted considering the enzyme activity towards the substrate used. Lab Origins was used to analyse and fit the data. All reactions were repeated in triplicates, data reported as the average and the associated standard deviation.

Schemes, Figures and Tables



Scheme S1. Examples of assay used to monitor biocatalytic deacetylation reactions. **A)** HT acetate assay; the release of free acetate into solution is monitored *via* disappearance of the NADH absorbance by coupling the hydrolytic enzyme with an AK, a PK and a D-LDH. **B)** The continuous L-AAO assay, the formation of free L-amino acid (AA) into solution is observed *via* oxidation of *o*-dianisidine in a three enzymatic cascade process. **C)** Derivatization of a AA racemic mixture *via* Marfey' reagent (FDAA) with the formation of a 50:50 mixture of two diasteremeric derivatives, FDAA-L-AA (*S,S*) and FDAA-D-AA (*R,S*). **D)** This work; direct monitoring of the deacetylation reaction *via* ^{1}H NMR.



Scheme S2. Reaction scheme for *E. coli* ArgE catalysed hydrolysis of *N*-acetylated canonical and NCAAs.

***N*-Ac-L-AA**	**Conv (%)**	***m/z* [M+H]$^+$**
Ala	93.9	90.0
Ile	>99.0	132.1
Leu	99.3	132.1
Pro	>53.7	116.0.
Val	85.4	118.1
Phe	2.0	166.1
Tyr	N.D.	N.D.
Trp	N.D.	N.D.
Asp	3.8	N.D.
Glu	N.D.	N.D.
Arg	96.4	124.1
His	>99.0	156.1
Lys	98.8	147.0
Ser*	89.8	106.1
Thr	27.4	147.1
Cys	77.9	120.1
Met	>99.0	122.1
Asn	95.4	150.1
Gln	60.6	133.1

Table S1 Reaction scheme of the ArgE-catalysed hydrolysis of *N*-acetyl-L-proteinogenic amino acids. The product conversion percentages for the ArgE-catalysed biotransformation were calculated from the ^{1}H NMR spectra and the *m/z* values from positive ion LC-TOF MS analysis. N.D. product not detected. See protocols above and primary data below. * All *N*-acyl substrates were L-enantiomers, apart from serine which was a racemic mixture of L- and D- (the % conversion is based on the L-form).

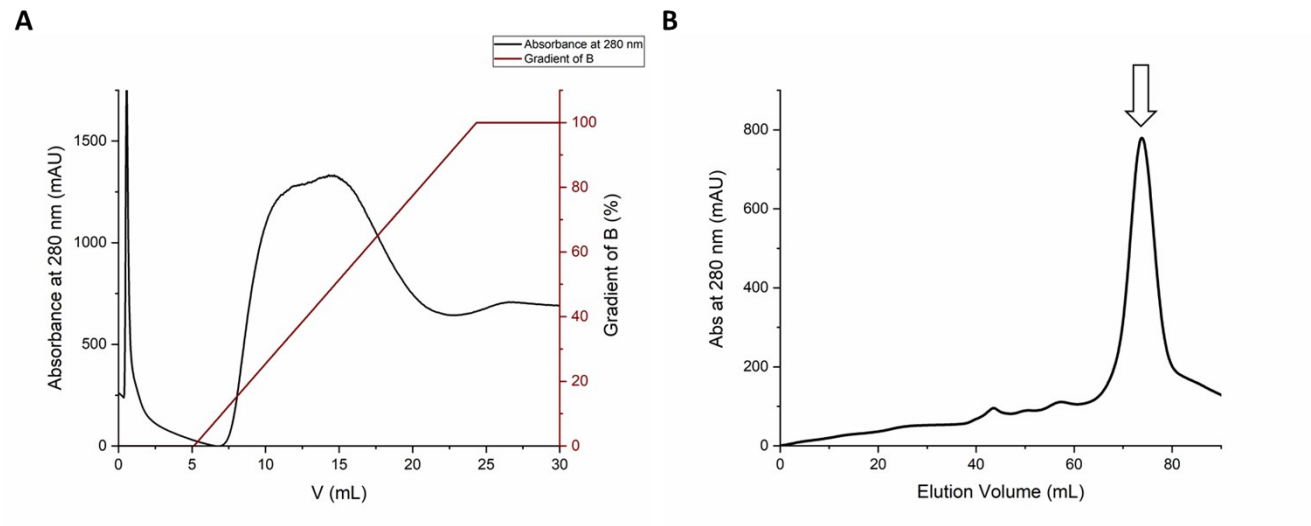


Figure S1. *E. coli* ArgE purification. **A)** IMAC chromatogram for His-tagged ArgE on a HisTrap FF (1 mL) column, elution of the target protein was carried out using a linear gradient of imidazole (red). **B)** Size exclusion chromatogram for un-tagged ArgE on a Superdex S200 column. The arrow points to the Acylase peak.

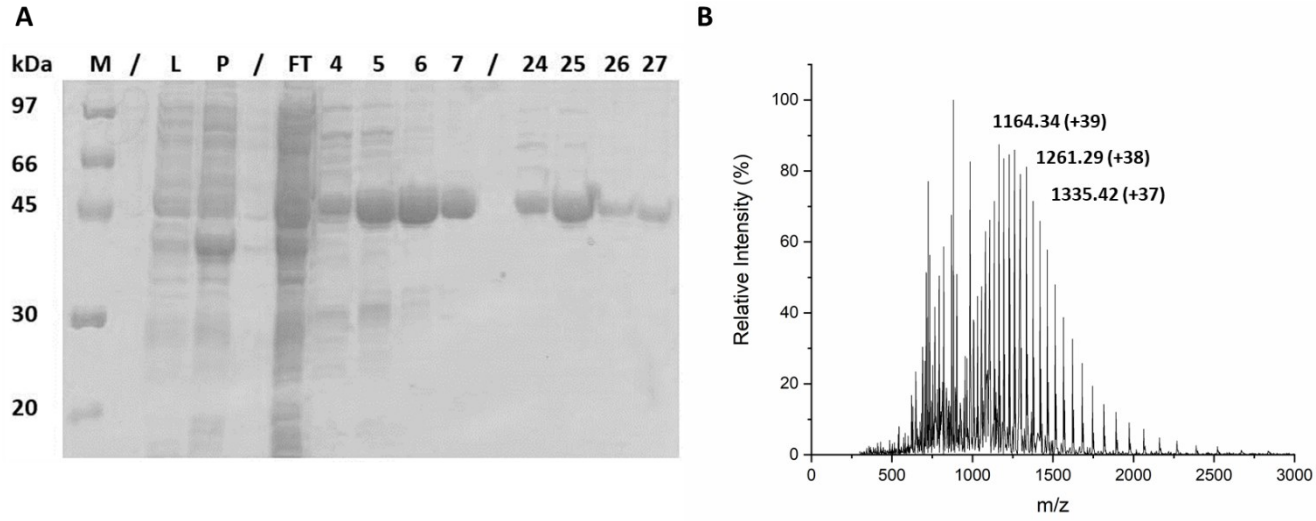


Figure S2. Characterization of purified ArgE. **A)** 12% SDS-PAGE gel of N-HisTag ArgE purification: lane 1 LMWM (M), lane 2 CFE (L), lane 3 insoluble fraction (P), lane 4 HisTrap flow through (FT), lane 5-8 HisTrap fractions (3 mL each, enzyme elution start at ~200 mM imidazole) and lane 9-12 S200 fractions (3 mL each). **B)** Denaturating LC ESI-MS of ArgE (20 μM) with a mass of 45369.00 ± 0.90 Da, in accordance with the MW calculated from the sequence (45369.50 Da). The values reported are m/z with the charges in brackets.

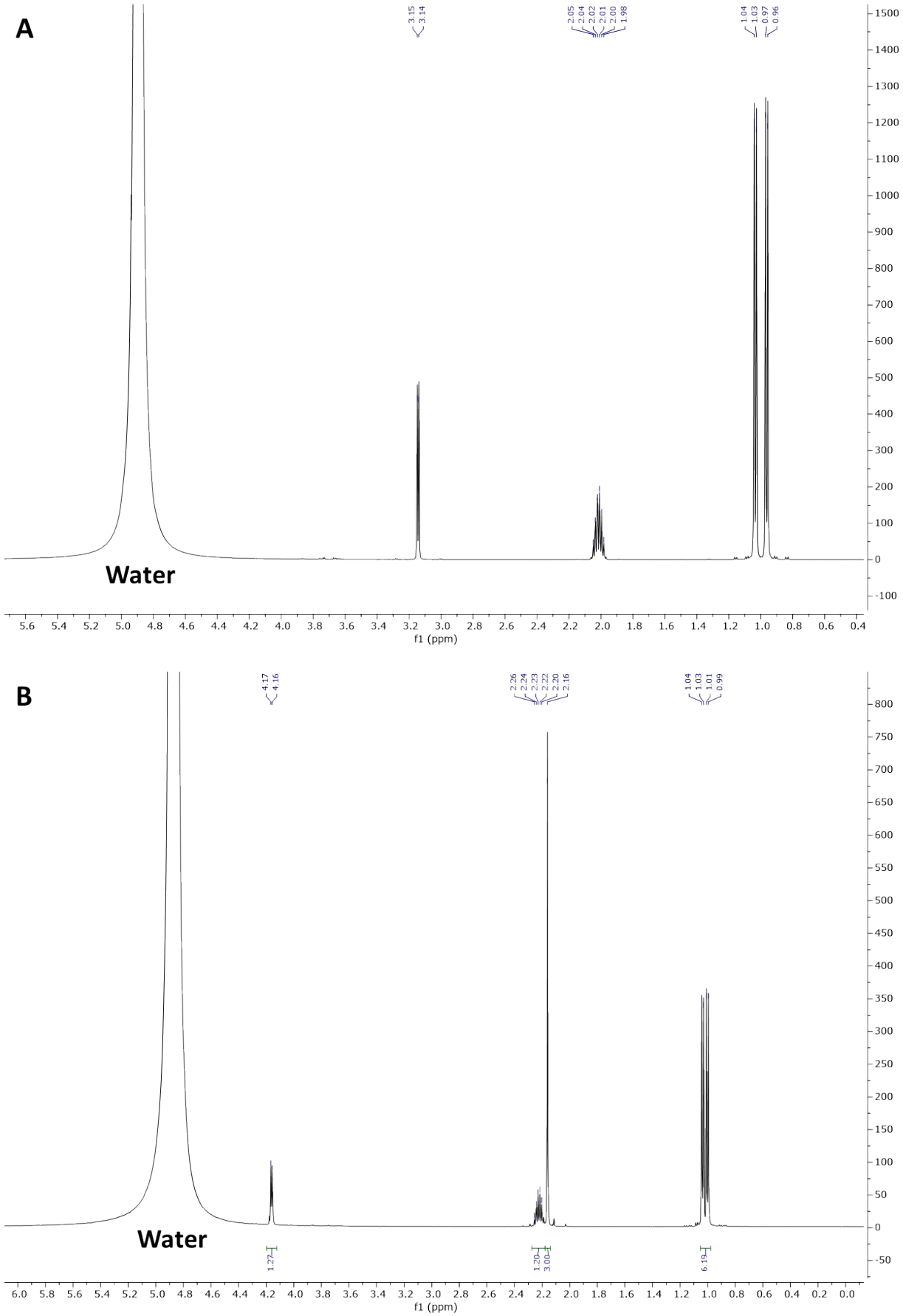


Figure S3. [1]H NMR spectra for **A)** L-Val and **B)** *N*-Ac-L-Val standards.

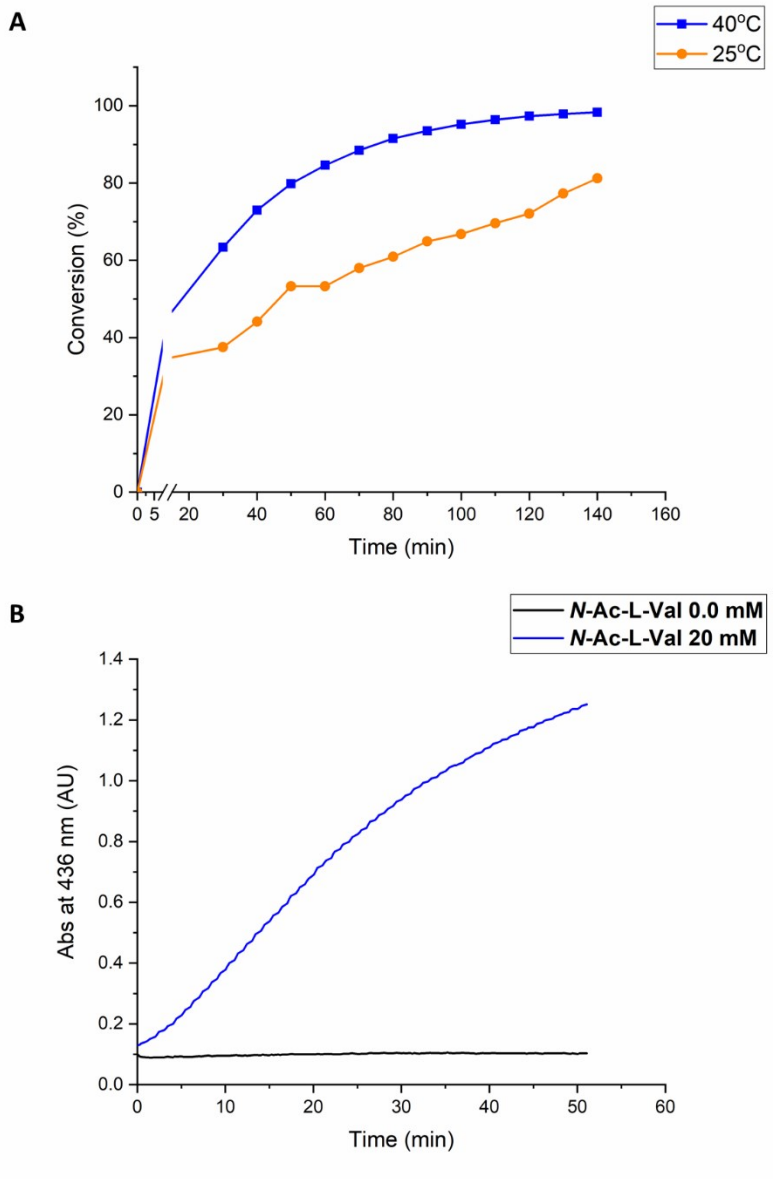
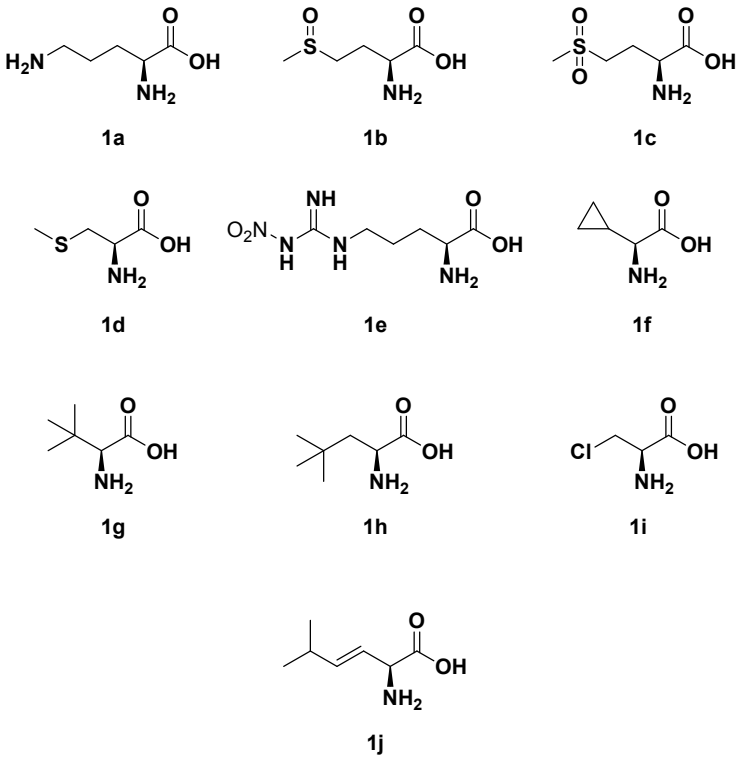


Figure S4. Monitoring ArgE catalysed biotransformations. **A)** Formation of L-Val from *N*-Ac-L-Val over time at 25°C (orange) and 40°C (blue). Conversion percentages were calculated from the integrals of the acetate signals (1.88 ppm) in the ^{1}H NMR spectra. The specific activity for L-Val formation was determined as 1.84 mM/min/mg at 40 °C. **B)** The ArgE-biocatalysed hydrolysis of *N*-Ac-L-Val at 40 °C (blue) monitored *via* the continuous HTP L-AAO coupled assay. Absorbance curves at 436 nm for the reaction (20 mM substrate in blue) and control (no substrate in black). The rate was linear from 10-25 min and the specific activity for o-dianisidine formation was 2.60 mM/min/mg.



Scheme S2. Structures of the NCAAs used (**1a-1j**).

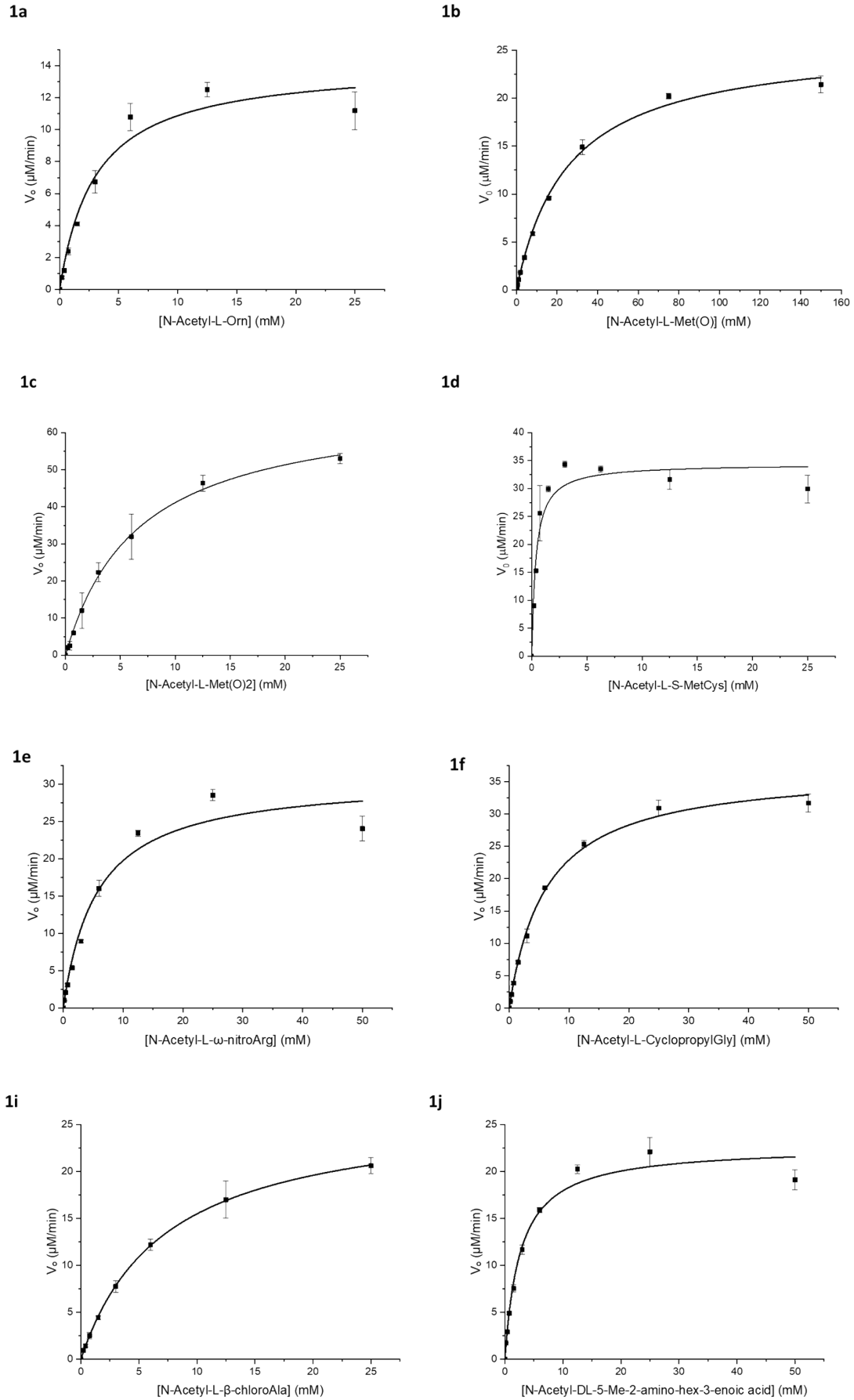


Figure S5. Monitoring ArgE *N*-Ac-L-NCAA substrate activity using the L-AAO coupled assay. The eight *N*-Ac-L-NCAA-substrates (**1a-1f**, **1i**, **1j**) identified using the ^{1}H NMR screen were analysed using the coupled

colourimetric L-AAO assay. Kinetic plots were fitted with Origin Lab (reactions were repeated in triplicates) and data presented in Table S2.

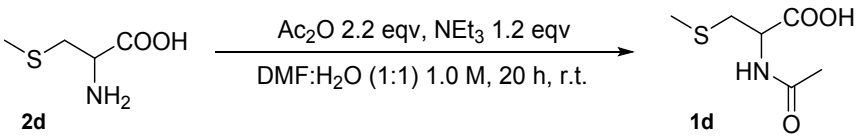
Substrate:	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}s^{-1}$)
1a	3.00 ± 0.30	0.106 ± 0.008	35.250 ± 0.008
1b	26.00 ± 3.00	1.970 ± 0.080	76.812 ± 0.005
1c	3.40 ± 0.90	0.360 ± 0.020	113.41 ± 0.010
1d	0.41 ± 0.01	0.264 ± 0.010	676.84 ± 0.01
1e	2.70 ± 0.10	0.170 ± 0.002	64.3415 ± 0.0005
1f	6.20 ± 0.50	0.280 ± 0.020	44.688 ± 0.004
1g	N.D.	N.D.	N.D.
1h	N.D.	N.D.	N.D.
1i	4.10 ± 0.30	0.152 ± 0.009	37.006 ± 0.005
1j	2.70 ± 0.10	0.170 ± 0.002	64.3415 ± 0.0005

Table S2 The kinetic parameters of *E. coli* ArgE for NCAAs substrates **1a-j**, calculated by employing the coupled colorimetric L-AAO assay.

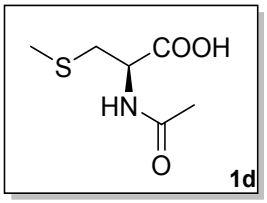
Substrate Synthesis

General acetylation of polar amino acids (using 2d as an example)

N-Ac-L-S-methyl cysteine, 1d



In a round bottom flask **2d** (297mg, 5.64 mmol) was dissolved in a 1:1 mixture of DMF and water (1.0 M, 5.64 mL). NEt_3 (944 µL, 1.2 eqv) and Ac_2O (1173 µL, 2.2 eqv) were added dropwise to the solution under stirring. After 20 h at r.t. the solvent was removed under vacuum and the residues re-dissolved in water; the pH of the solution was adjusted to 10 with $NaCO_{3\ sat}$ and the mixture washed with Et_2O (20 mLx3). The pH of the water phase was adjusted to 2 with HCl 2 M and the solvent removed under vacuum. The white solid was dissolved in the minimum amount of MeOH, the solution filtered and methanol removed *in vacuo* to give pure *N*-acetyl-*S*-methyl cysteine (**1d**).



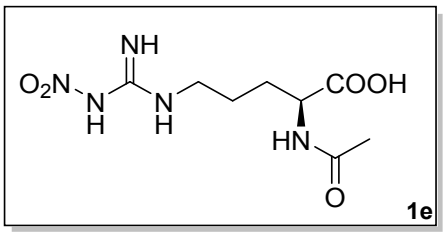
Yield: 49.6 %

^{1}H NMR: (500 MHz, D_2O) δ 4.63 (dd, *J* = 8, 4.5 Hz,1H), 3.08 (m, 1H), 2.94 (m, 1H), 2.16 (s, 3H) and 2.08 (s, 3H).

^{13}C NMR: (126 MHz, D_2O) δ 174.30, 174.13, 52.13, 34.59, 21.64, 14.80.

TOF-ESI MS: calculated 176.2 $[M - H]^-$, found 176

N-Ac-L-ω-nitro arginine, 1e



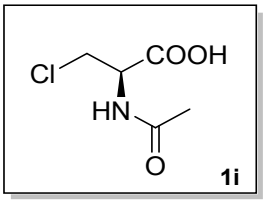
Yield: 41.4 %

^{1}H NMR: (500 MHz, D_2O) δ 4 4.35 (dd, *J* = 8.4, 4.8 Hz,1H), 3.29 (t, 2H), 2.02 (s, 3H) and 1.98-1.62 (bm, 4H).

^{13}C NMR: (126 MHz, D_2O) δ 178.83, 173.59, 158.85, 54.76, 40.75, 28.85, 23.91 and 21.88.

TOF-ESI MS: calculated 260.2 $[M - H]^-$, found 260.0.

***N*-Ac-L-β-chloro alanine, 1i**



Yield: 17.8 %

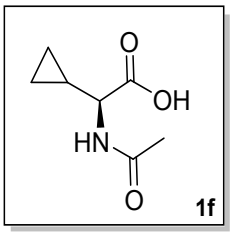
^{1}H NMR: (500 MHz, D_2O) δ 4.76 (m, 1H), 4.01 (dd, *J* =12.0, 5.2 Hz,1H), 3.90 (dd, *J* = 11.6, 3.6 Hz,1H) and 2.06 (s, 3H).

TOF-ESI MS: calculated 164.5 $[M - H]^-$, found 164.0.

Acetylation of hydrophobic amino acids

The reaction was carried out using the same protocol as **2e**. After 20 h at r.t. the solvent was removed under vacuum and the residues re-dissolved in water; the pH of the solution was adjusted to 10 with $NaCO_{3\ sat}$ and the mixture washed with Et_2O (20 mLx3). The organic phase discarded and the pH of the water phase was adjusted to 2 with HCl 2 M and the solution extracted with EtOAc (20 mLx3). The organic phase was dried under $MgSO_4$, filtered and the solvent removed under vacuum to give pure *N*-acetyl-cyclopropyl glycine (**1f**).

***N*-Ac-L-cyclopropyl glycine, 1f**



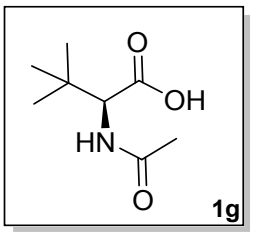
Yield: 58.8 %

^{1}H NMR: (500 MHz, D_2O) δ 3.52 (d, 1H), 2.00 (s, 3H), 1.11 (m, 1H), 0.62 (m, 2H), 0.49 (m, 1H), 0.32 (m, 1H).

^{13}C NMR: (126 MHz, D_2O) δ 178.92, 173.37, 59.44, 21.72, 12.78, 2.92 and 2.35.

TOF-ESI MS: calculated 156.2 $[M - H]^-$, found 156.0.

***N*-Ac-L-tert leucine, 1g**



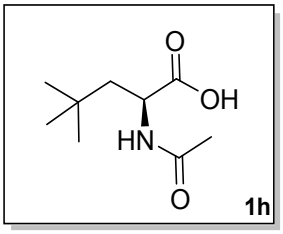
Yield: 53 %

^{1}H NMR: (500 MHz, D_2O) δ 4.11 (s, 1H), 2.04 (s, 3H) and 1.00 (s, 9H).

^{13}C NMR: (126 MHz, D_2O) δ 175.04, 174.26, 61.97, 32.99, 25.83 and 21.56.

TOF-ESI MS: calculated 172.2 $[M - H]^-$, found 172.0.

N-Ac-L-tertbutyl alanine, 1h



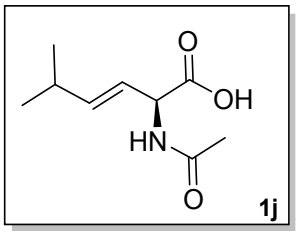
Yield: 79.3 %

^{1}H NMR: (500 MHz, D_2O) δ 4.49 (dd, *J* = 7.6, 6 Hz, 1H), 1.99 (s, 3H), 1.82 (dd, *J* = 11.6, 2.4 Hz, 1H), 1.61 (dd, *J* = 11.6, 7.6 Hz, 1H) and 1.00 (s, 9H).

^{13}C NMR: (126 MHz, D_2O) δ 180.78, 173.08, 53.16, 45.23, 29.91, 28.85 and 21.99.

TOF-ESI MS: calculated 186.2 $[M - H]^-$, found 186.0.

N-acetyl *trans*-2-amino-5-methylhex-3-enoic acid, 1j.



Yield: 56 %

^{1}H NMR: (500 MHz, D_2O) δ 5.91 (dd, *J* = 12.4, 5.2 Hz, 1H), 5.78 (dd, *J* = 12.4, 5.2 Hz, 1H), 5.54 (m, 2H), 4.64 (d, 1H), 2.57 (m, 2H), 2.07 (s, 3H), 2.06 (s, 3H), 1.00 (m, 12H).

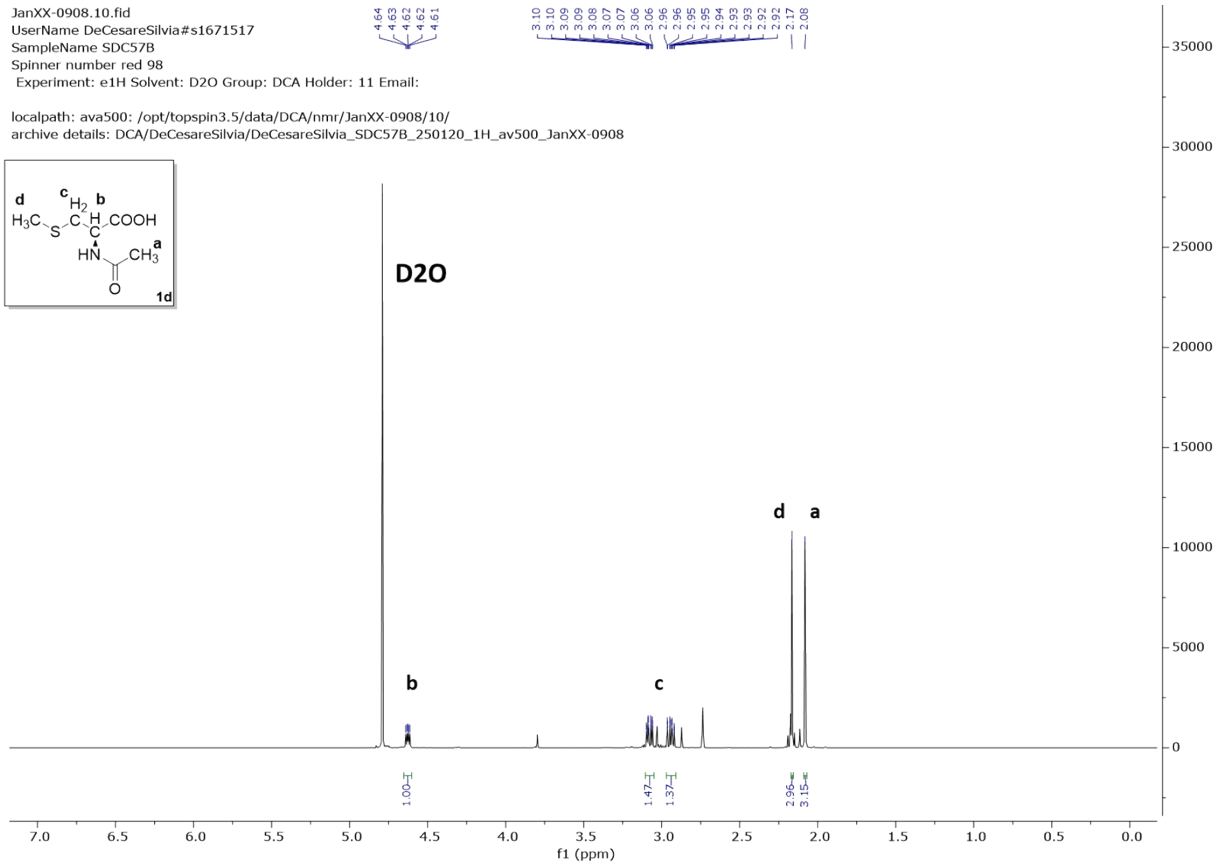
^{13}C NMR: (126 MHz, D_2O) δ 177.7 (both rotamers), 173.0 (major), 173.1 (minor), 144.2 (minor), 141.3 (major), 122.0 (major), 119.0 (minor), 57.5 (major), 55.2 (minor), 30.3 (minor), 30.2 (major), 22.0 (major), 21.6 (minor), 21.3 (major), 21.1 (minor), 21.0 (minor).

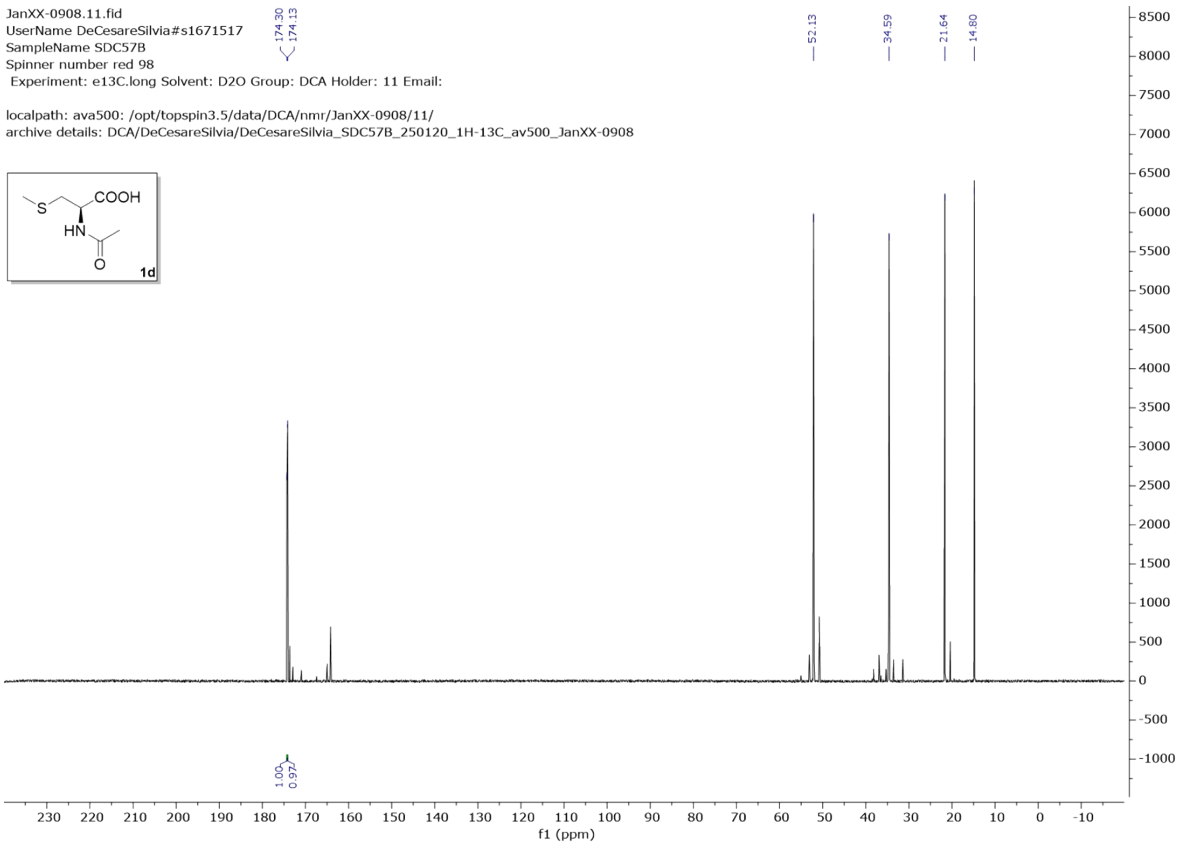
TOF-ESI MS: calculated 170.2 $[M - H]^-$, found 170.2.

^{1}H and ^{13}C NMR spectra

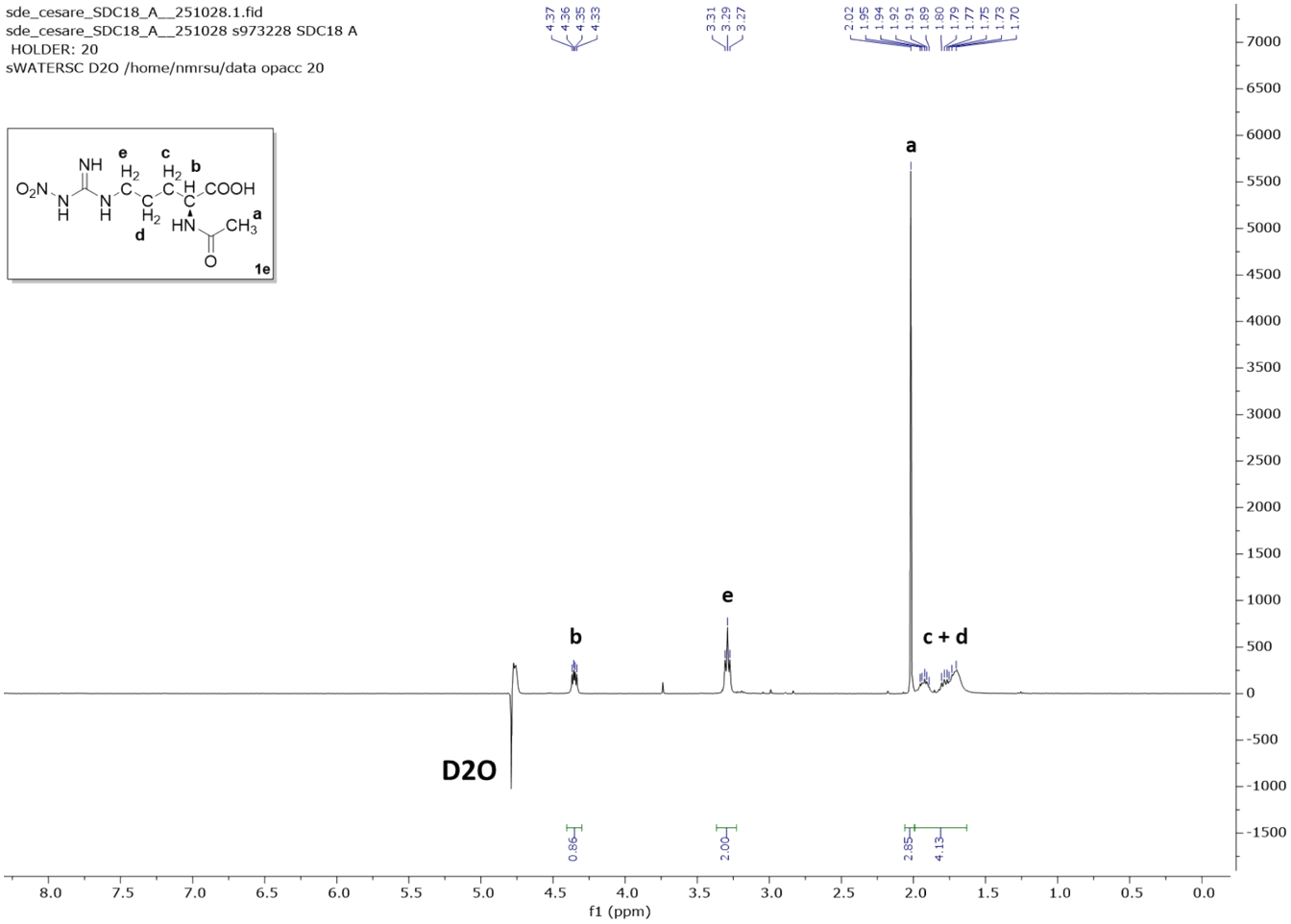
Substrate Characterization

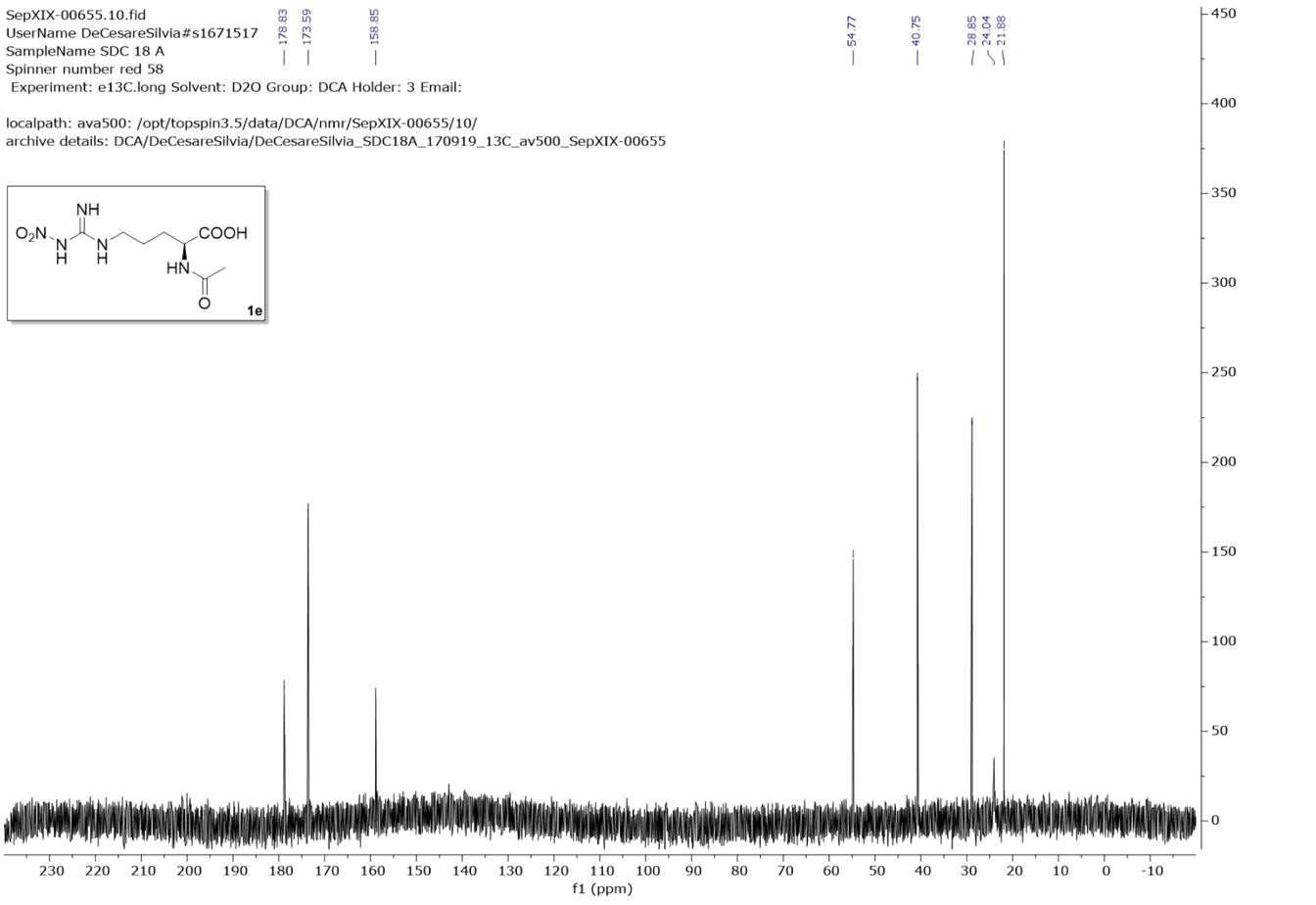
N-Ac-L-S-methyl cysteine, 1d



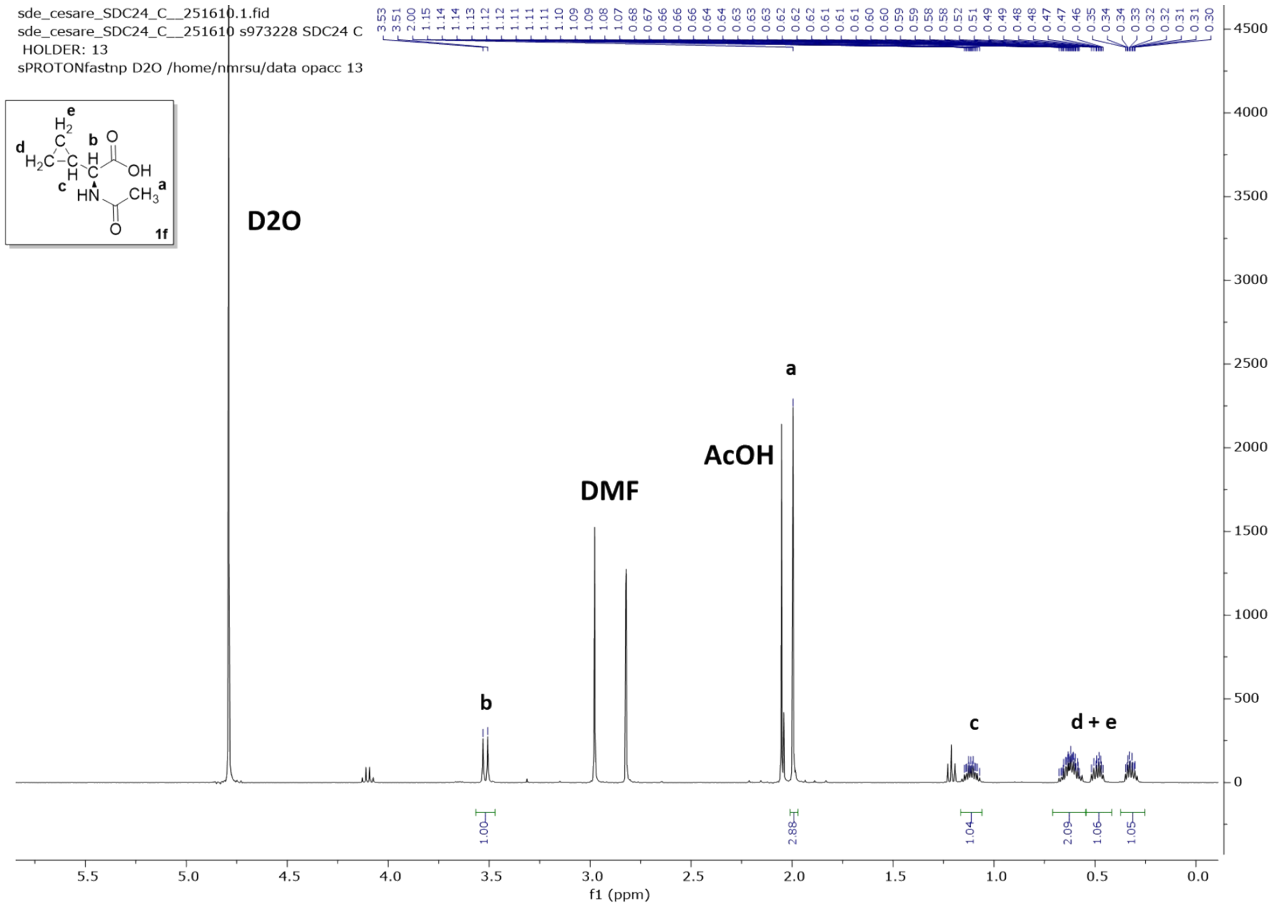


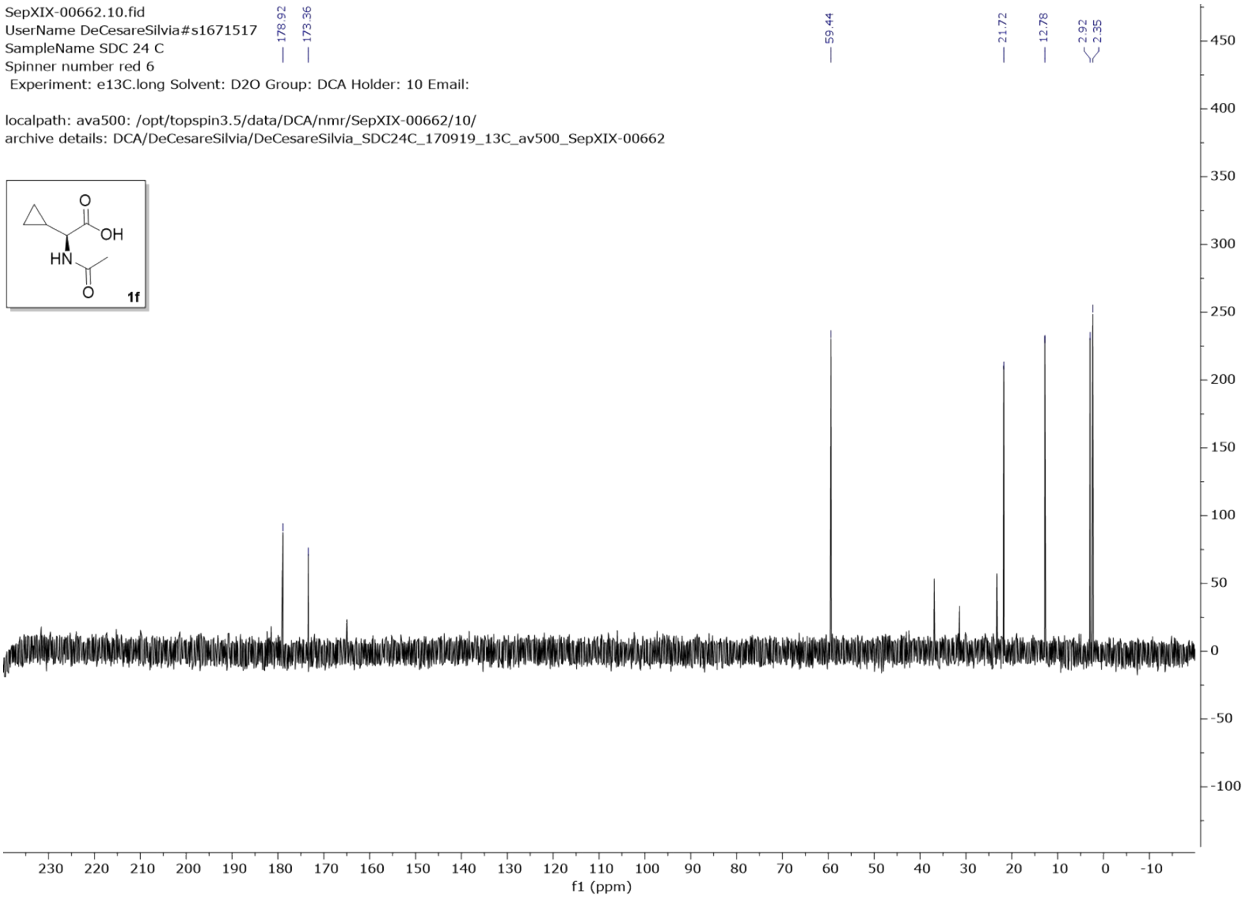
N-Ac-L-ω-nitro arginine, 1e



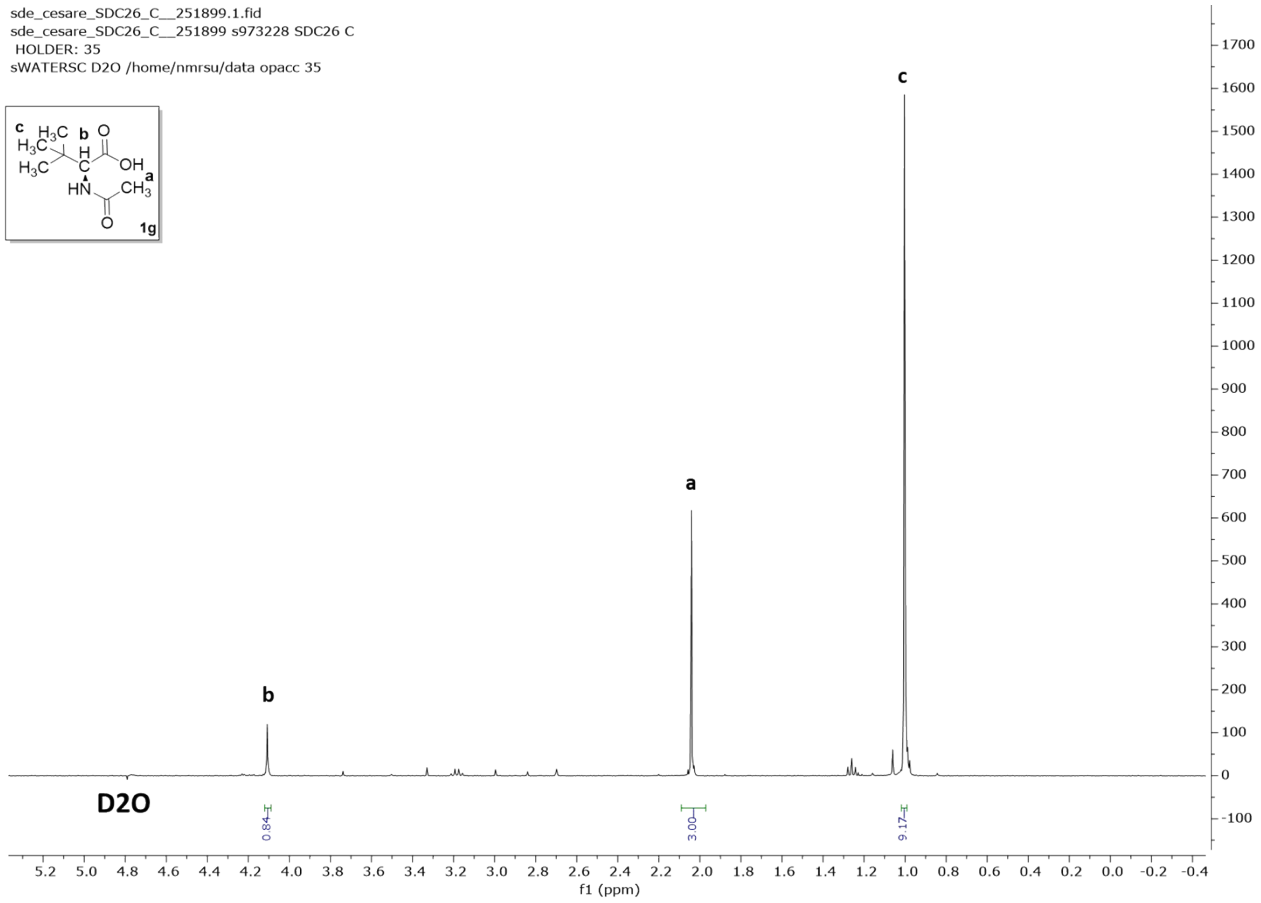


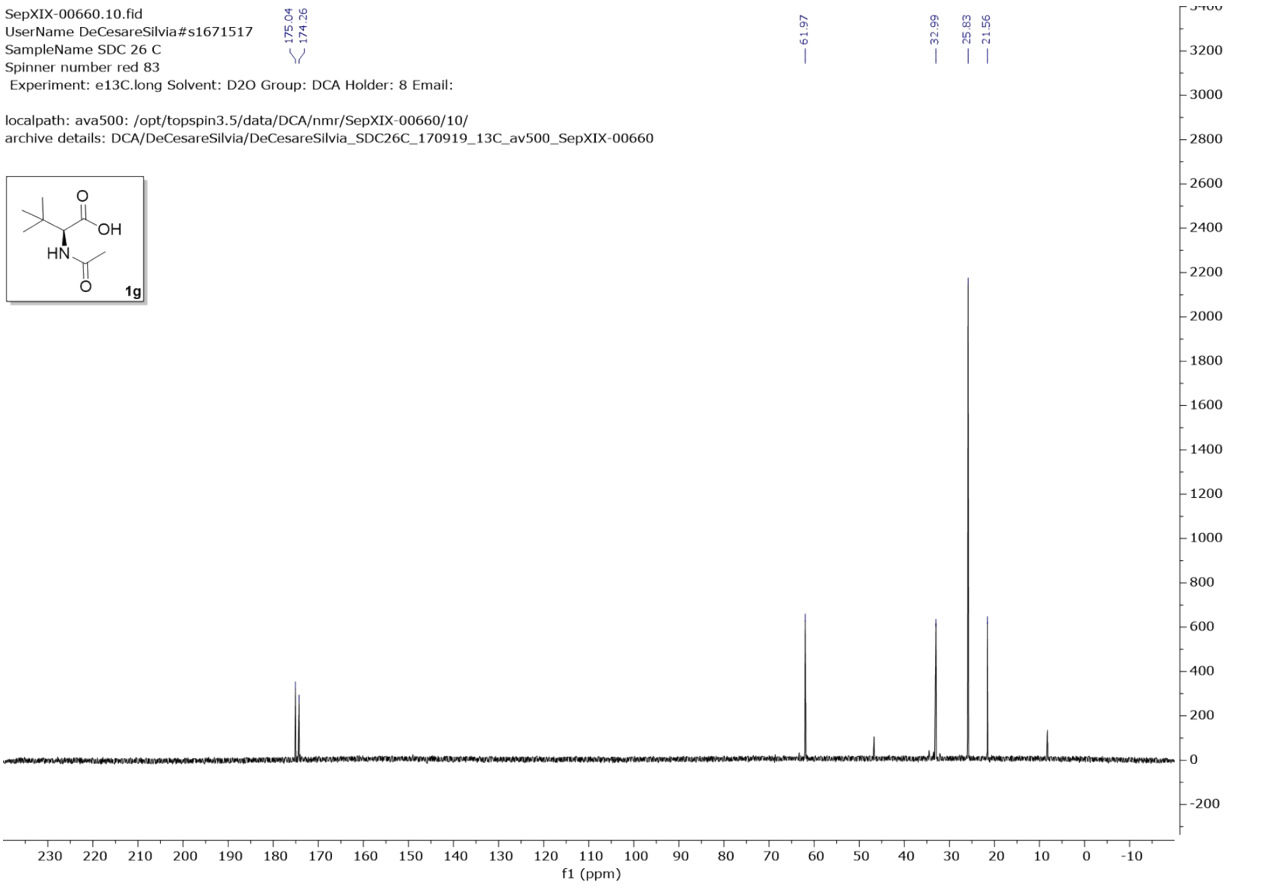
N-Ac-L-cyclopropyl glycine, 1f



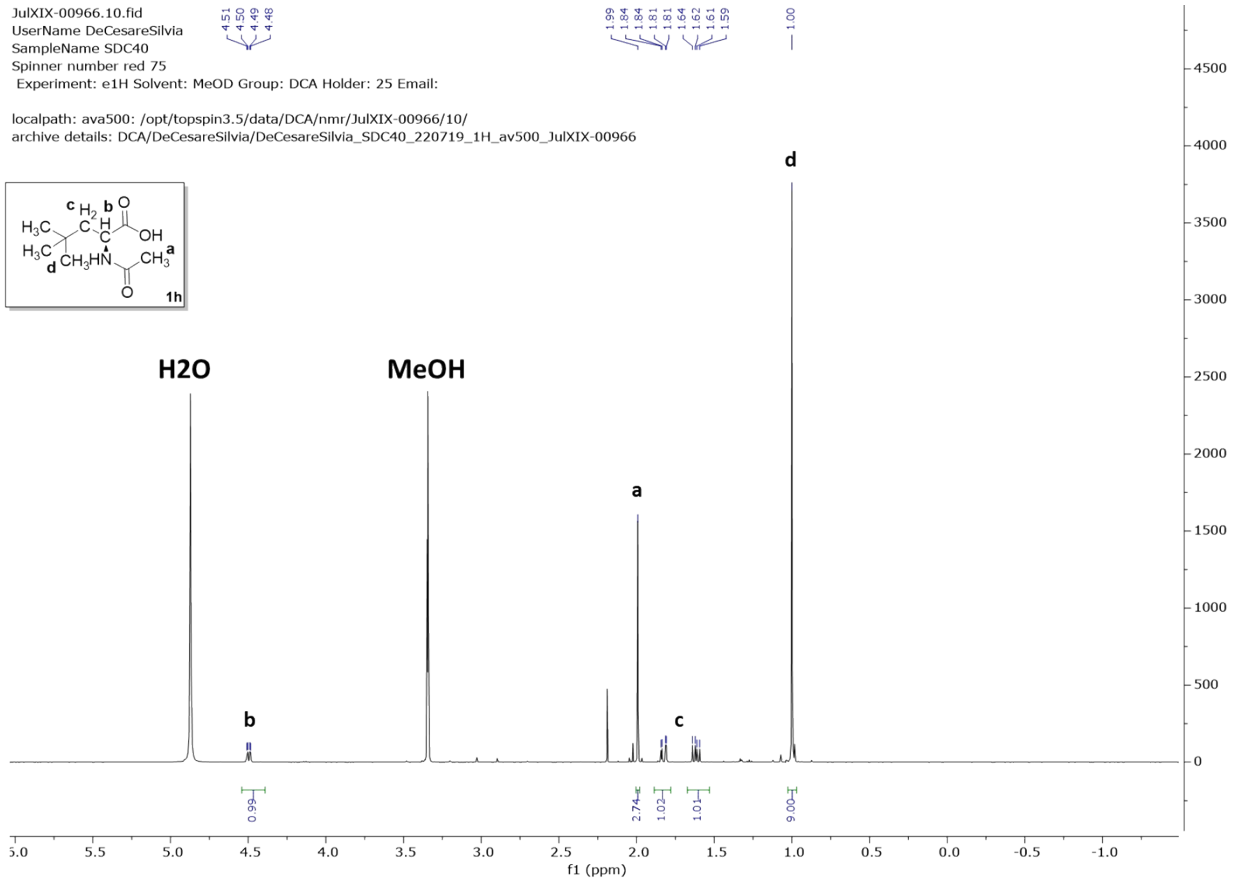


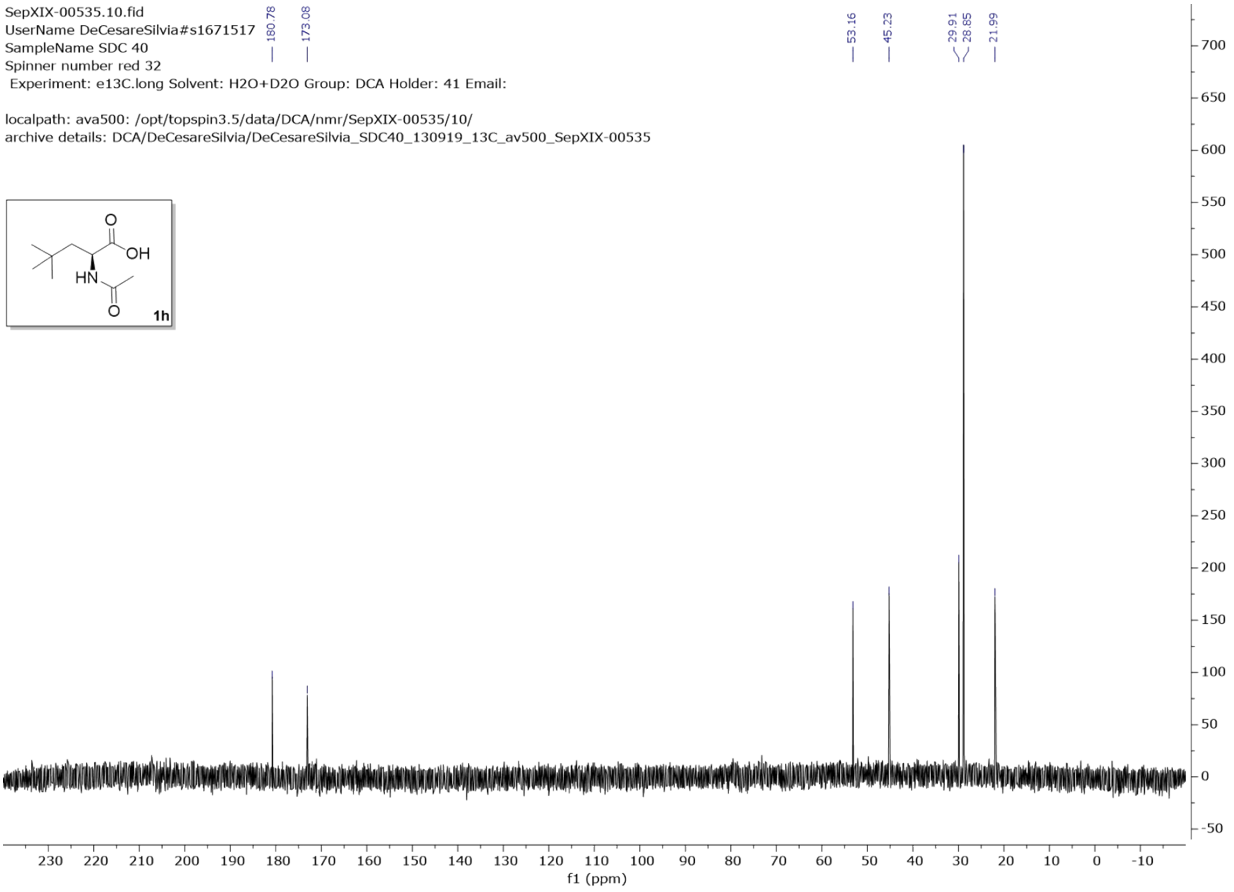
N-Ac-L-tert leucine, 1g



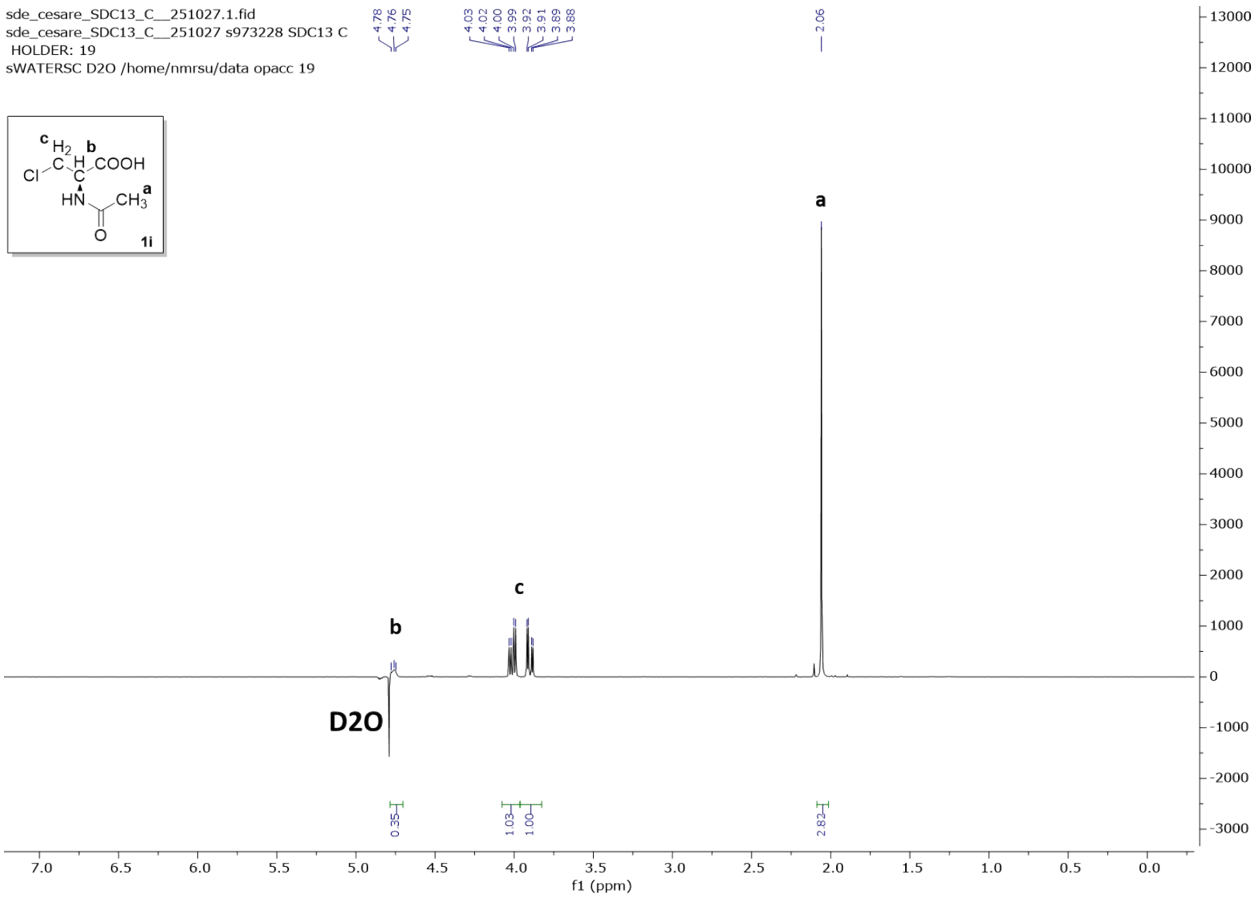


N-Ac-L-tertbutyl alanine, 1h



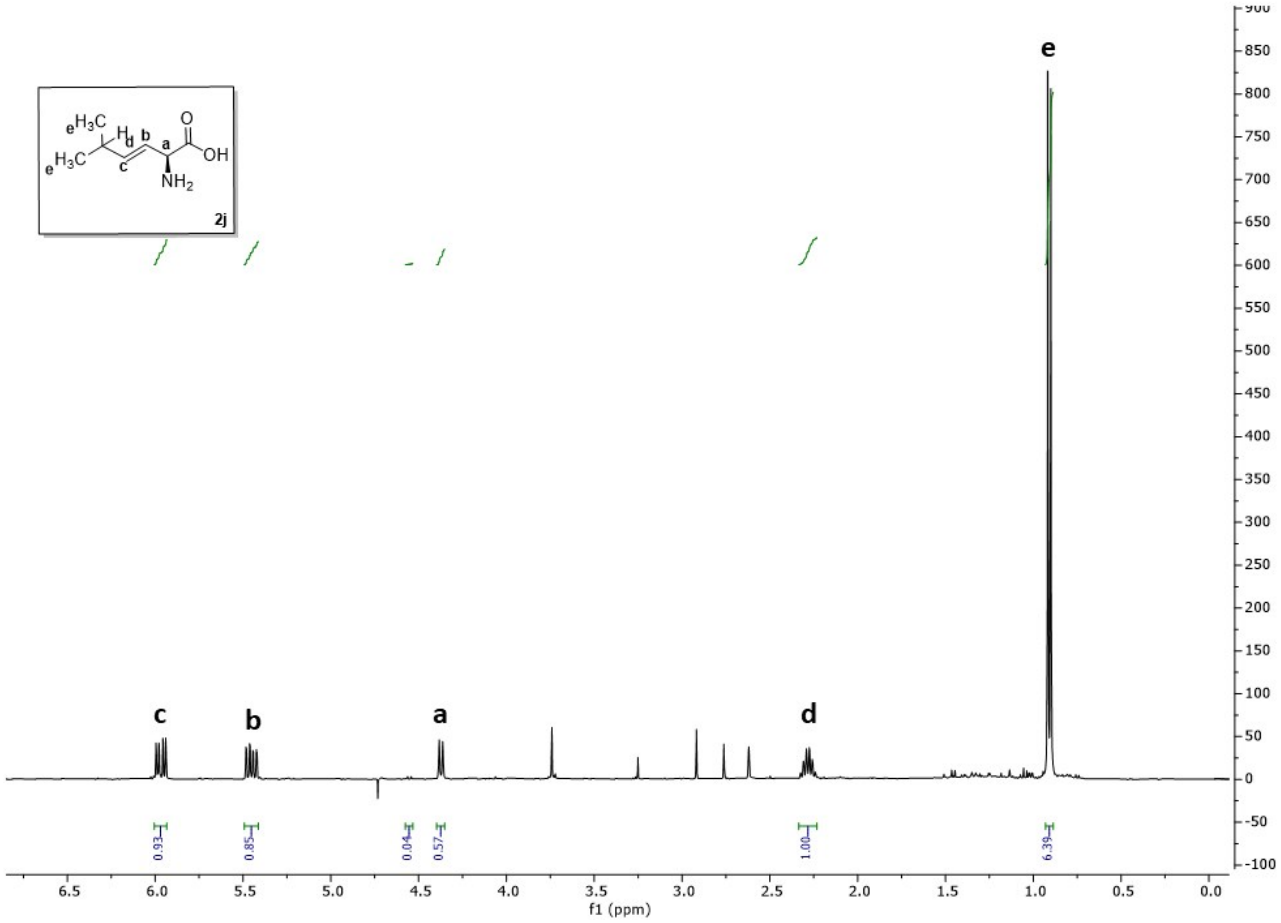


N-Ac-L-β-chloro alanine, 1i

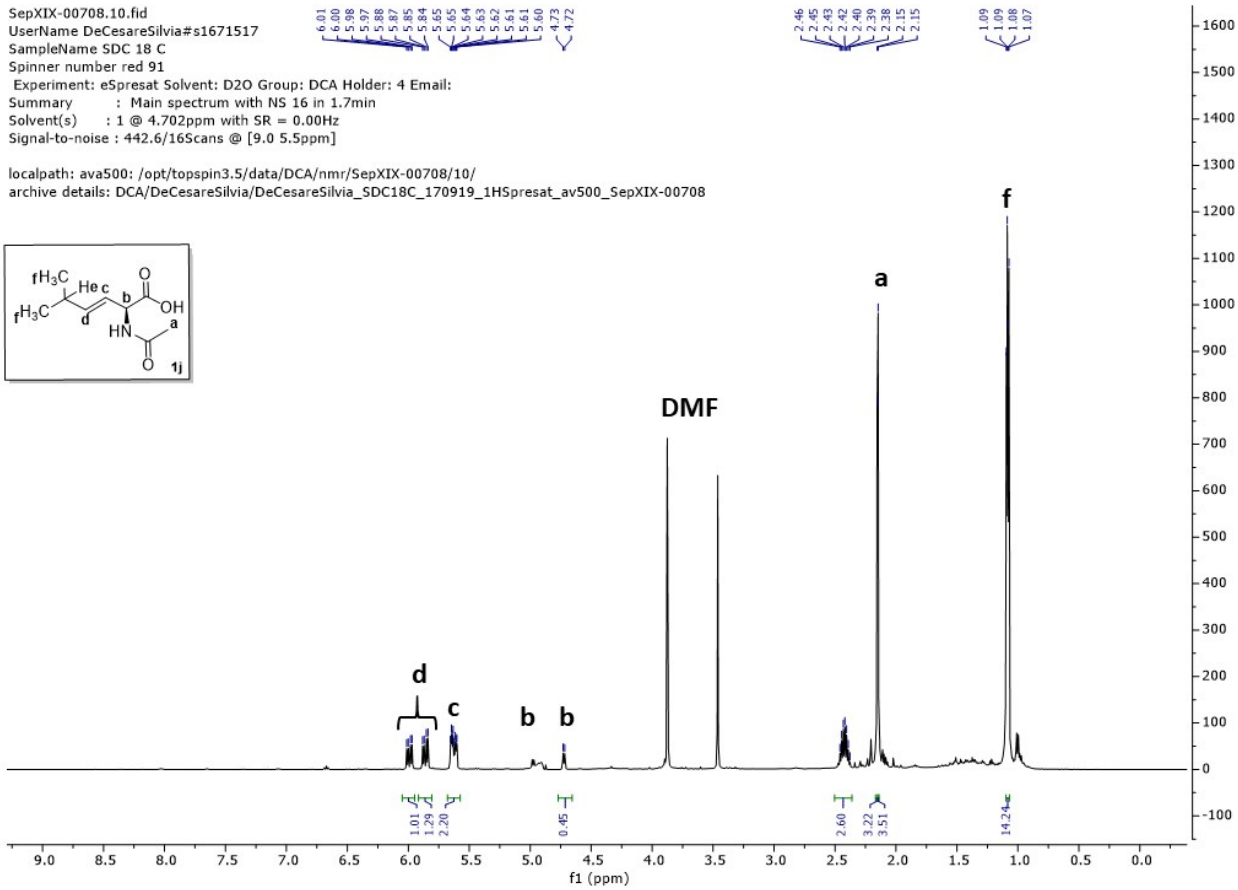


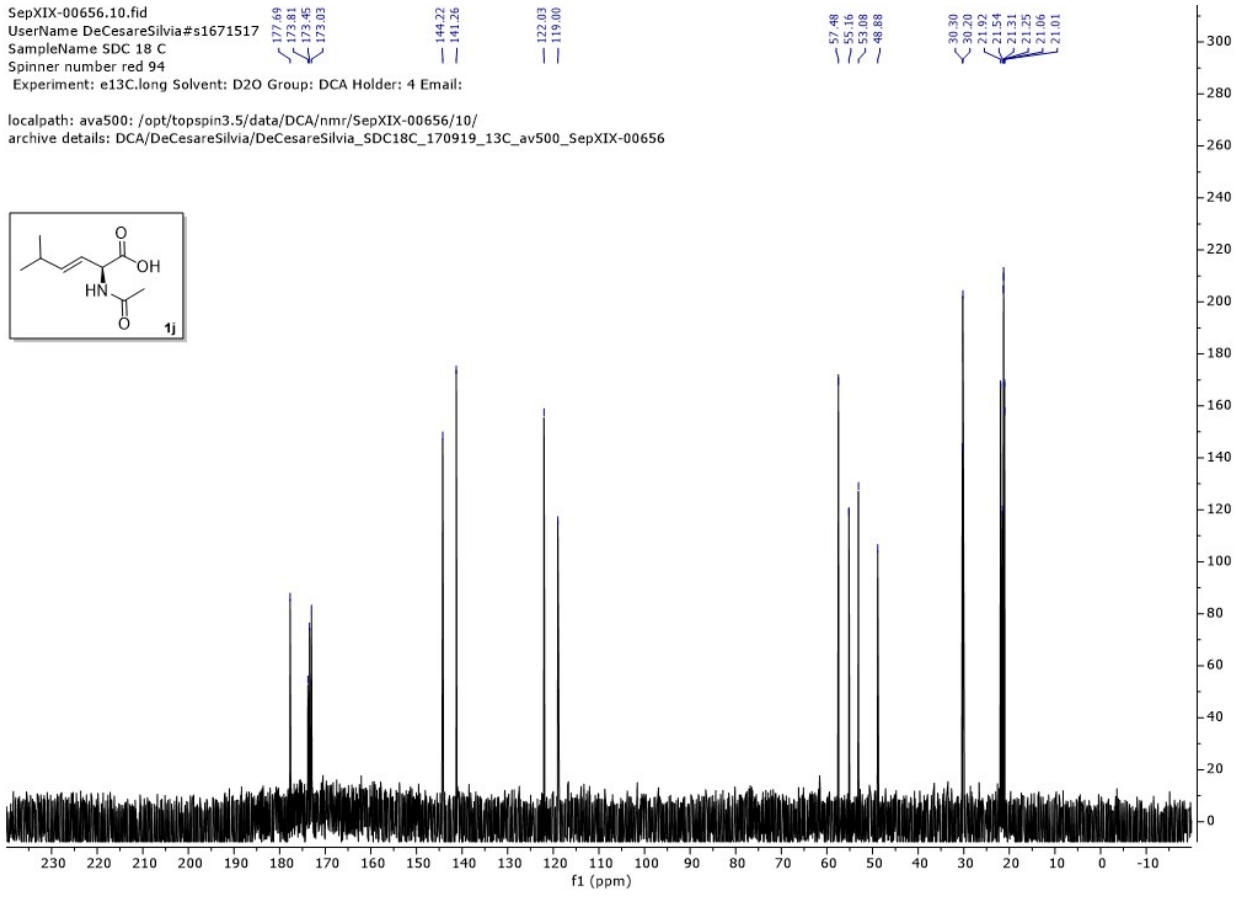
Unfortunately there was not enough substrate material to carry out a 13C NMR analysis of this substrate.

***trans*-2-amino-5-methylhex-3-enoic acid, 2j**

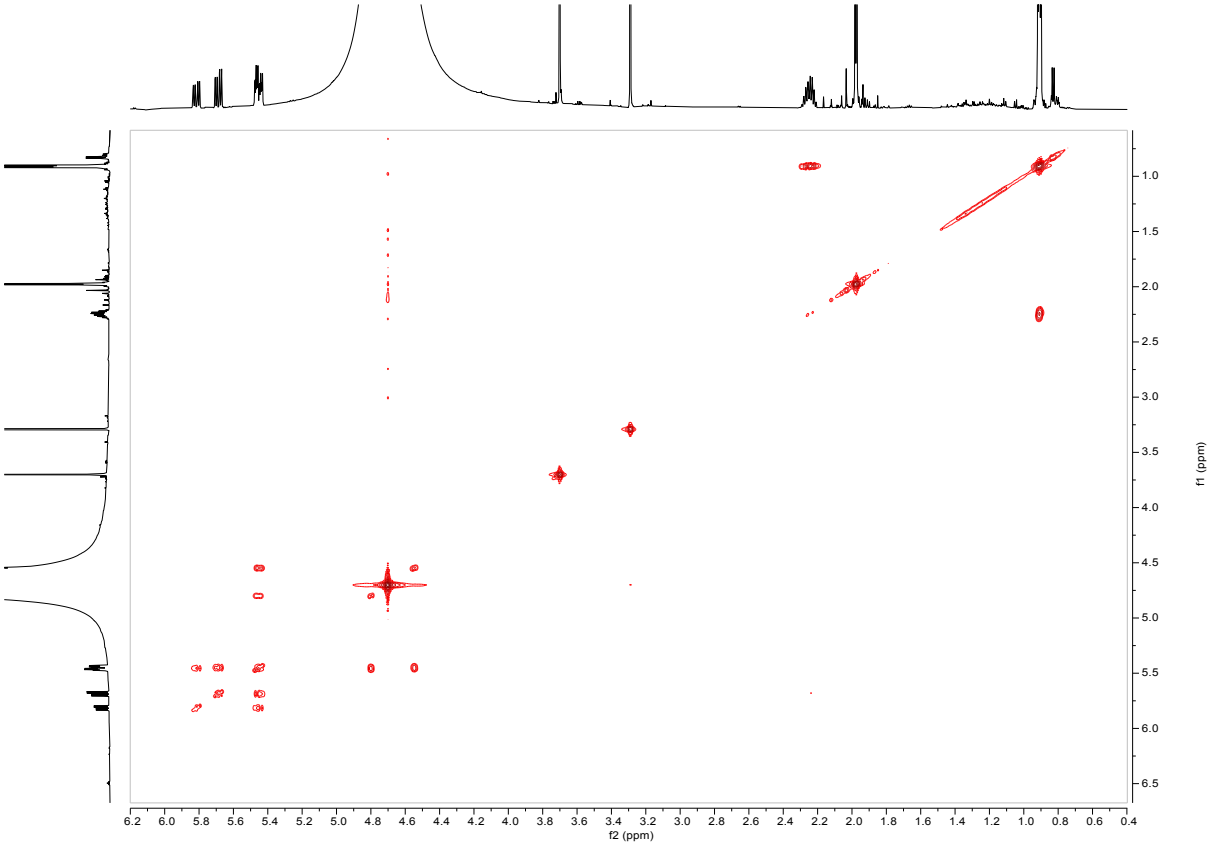


N-acetyl *trans*-2-amino-5-methylhex-3-enoic acid 1j (mixture of rotamers)

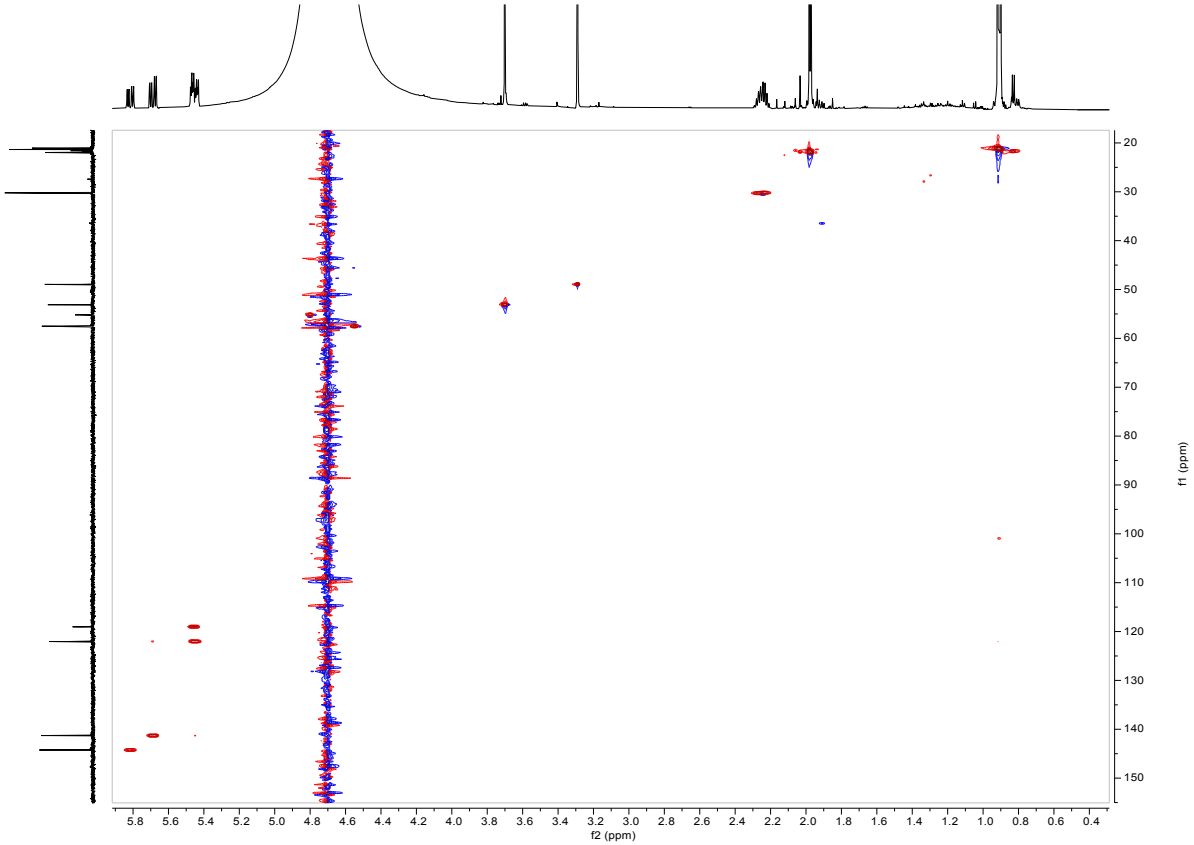




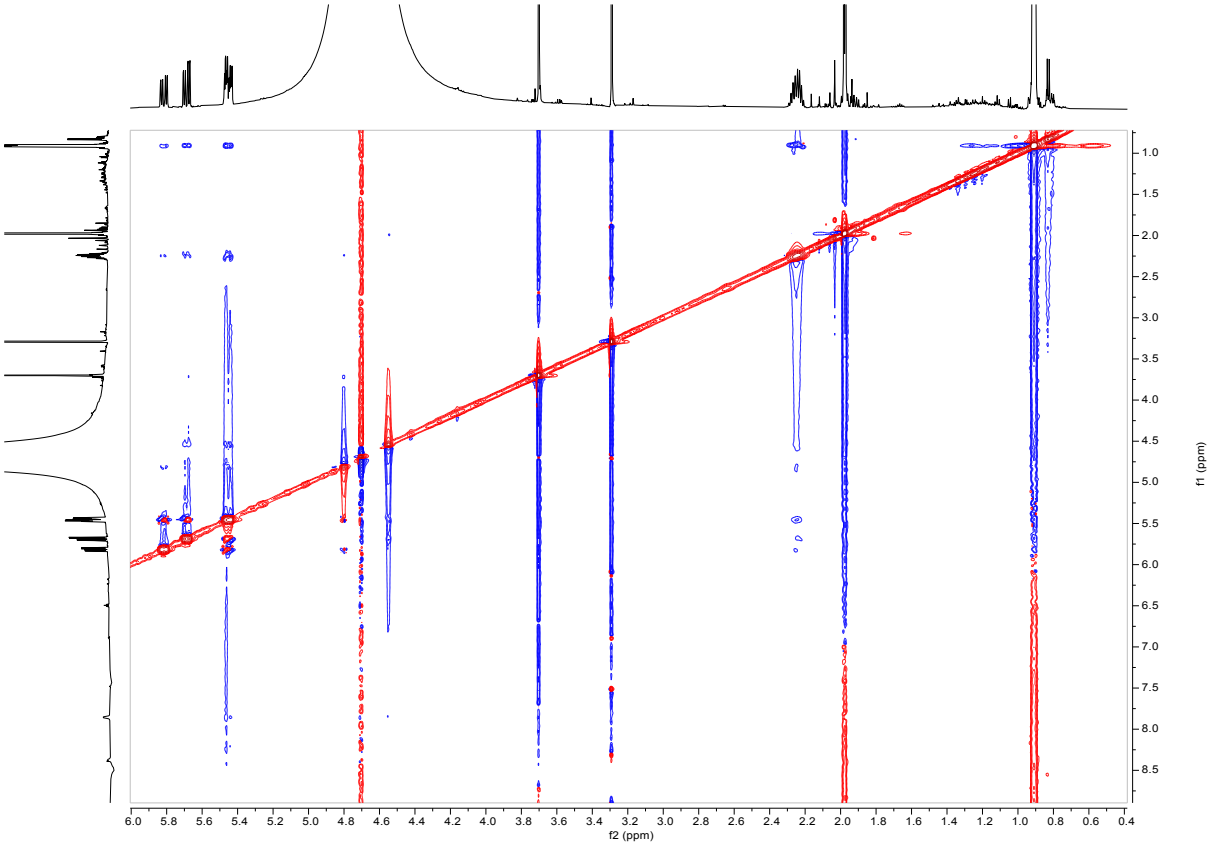
2j COSY NMR



2j HSQC NMR

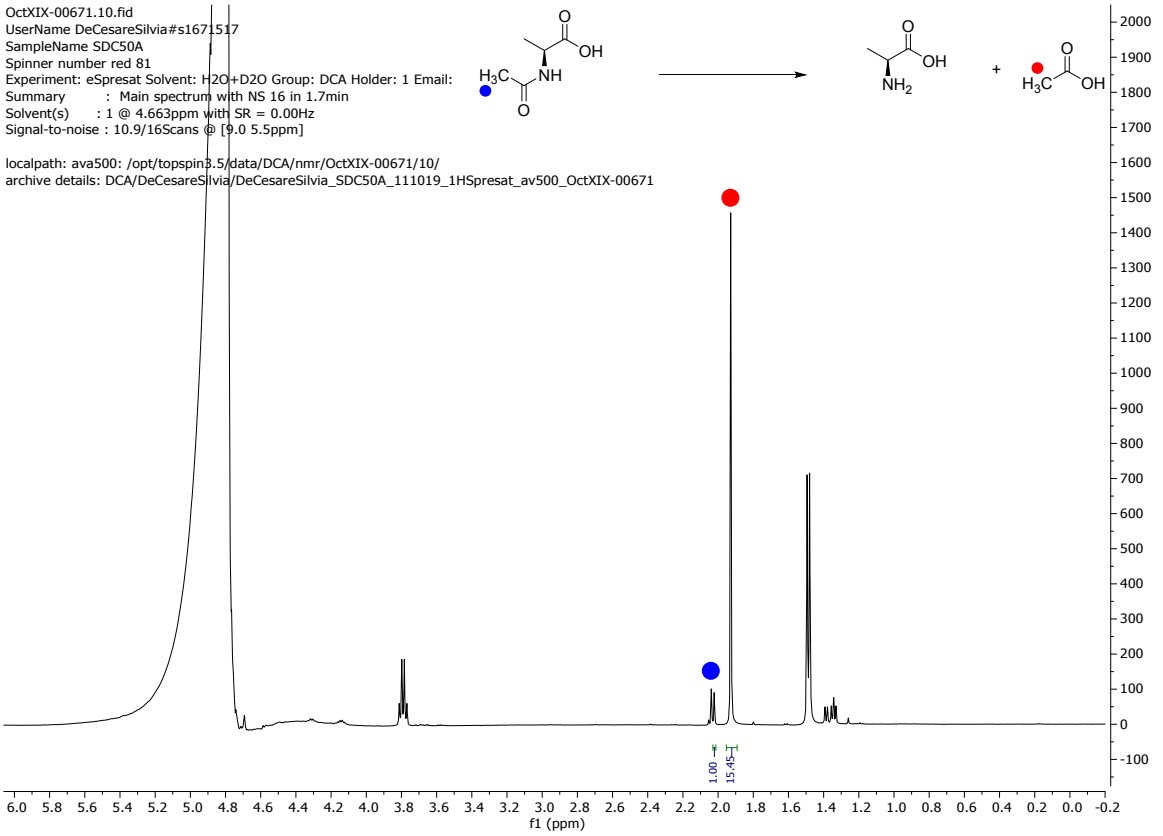


2j NOESY NMR

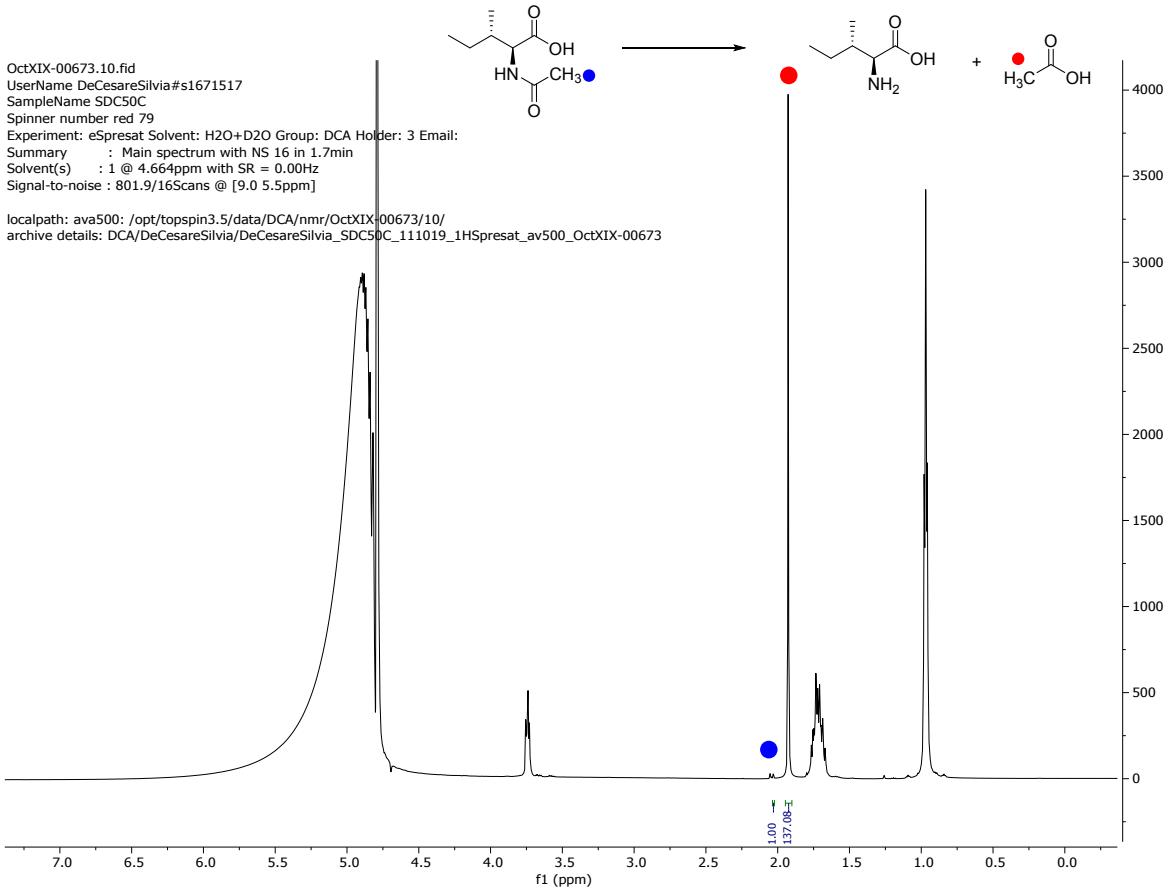


ArgE catalysed hydrolysis of *N*-acetyl proteinogenic amino acids: ^{1}H NMR spectra

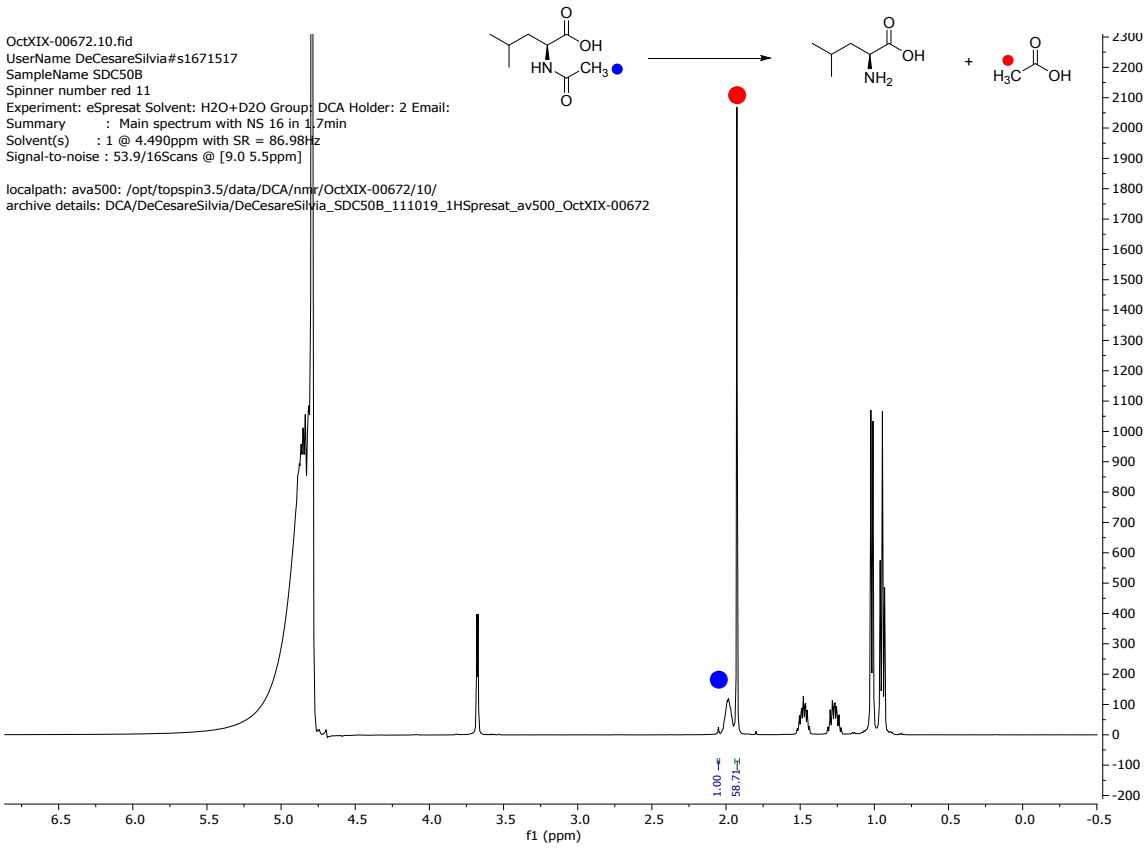
N-Ac-L-Ala hydrolysis



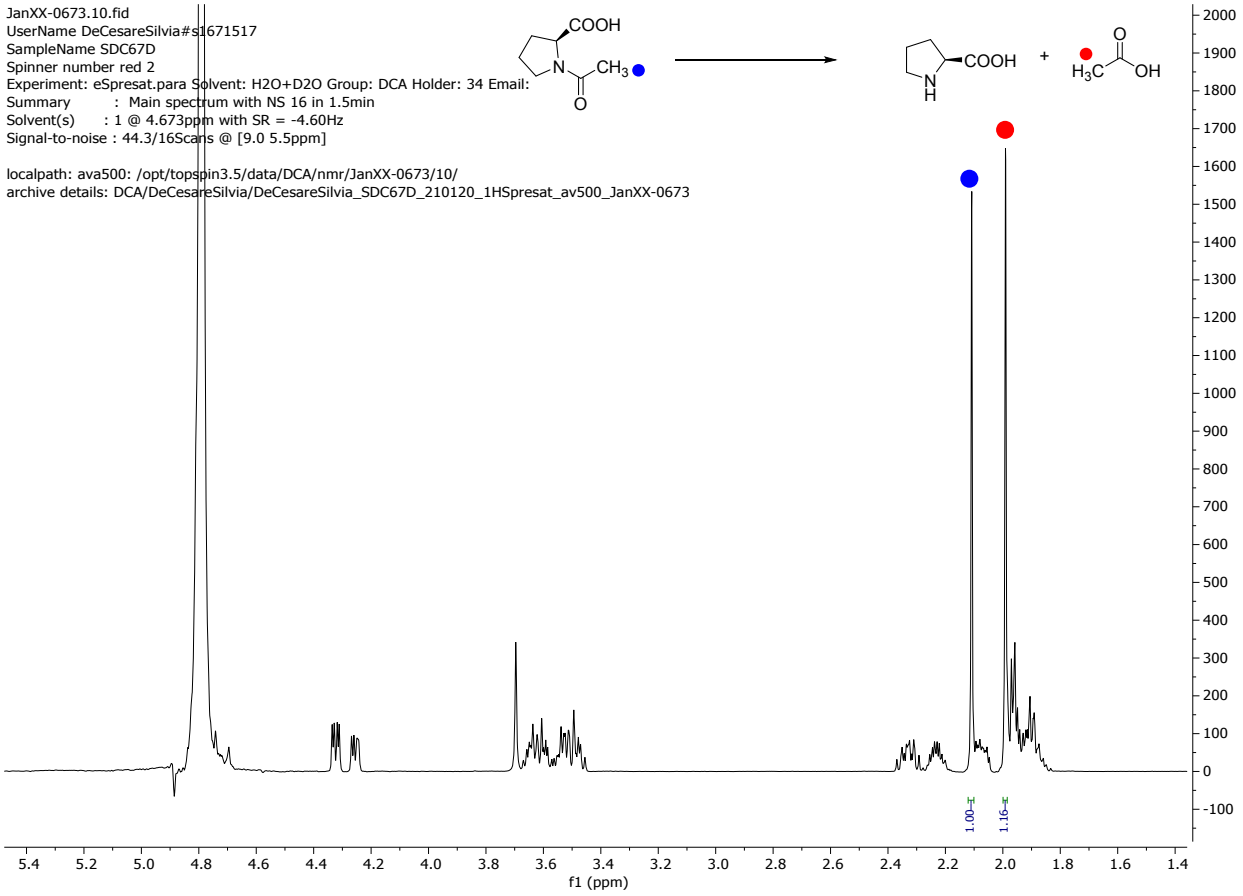
N-Ac-L-Ile hydrolysis



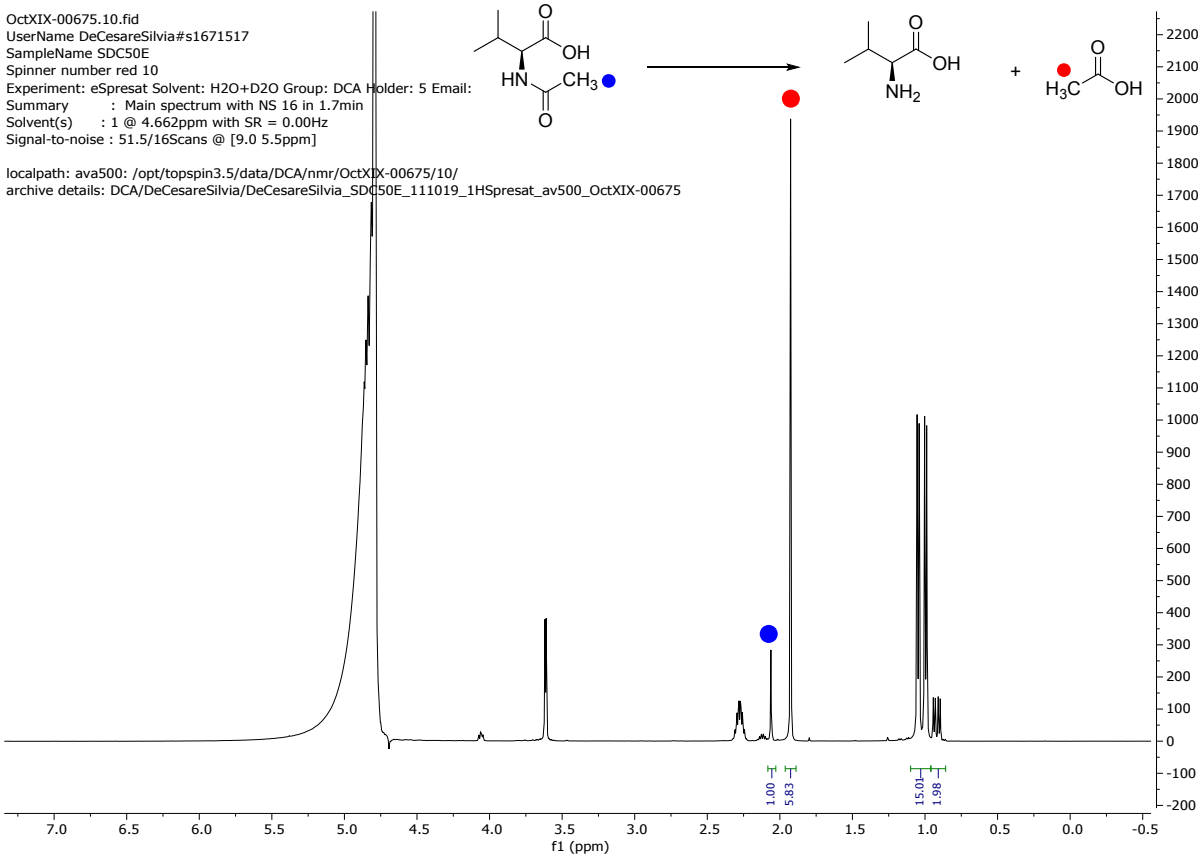
N-Ac-L-Leu hydrolysis



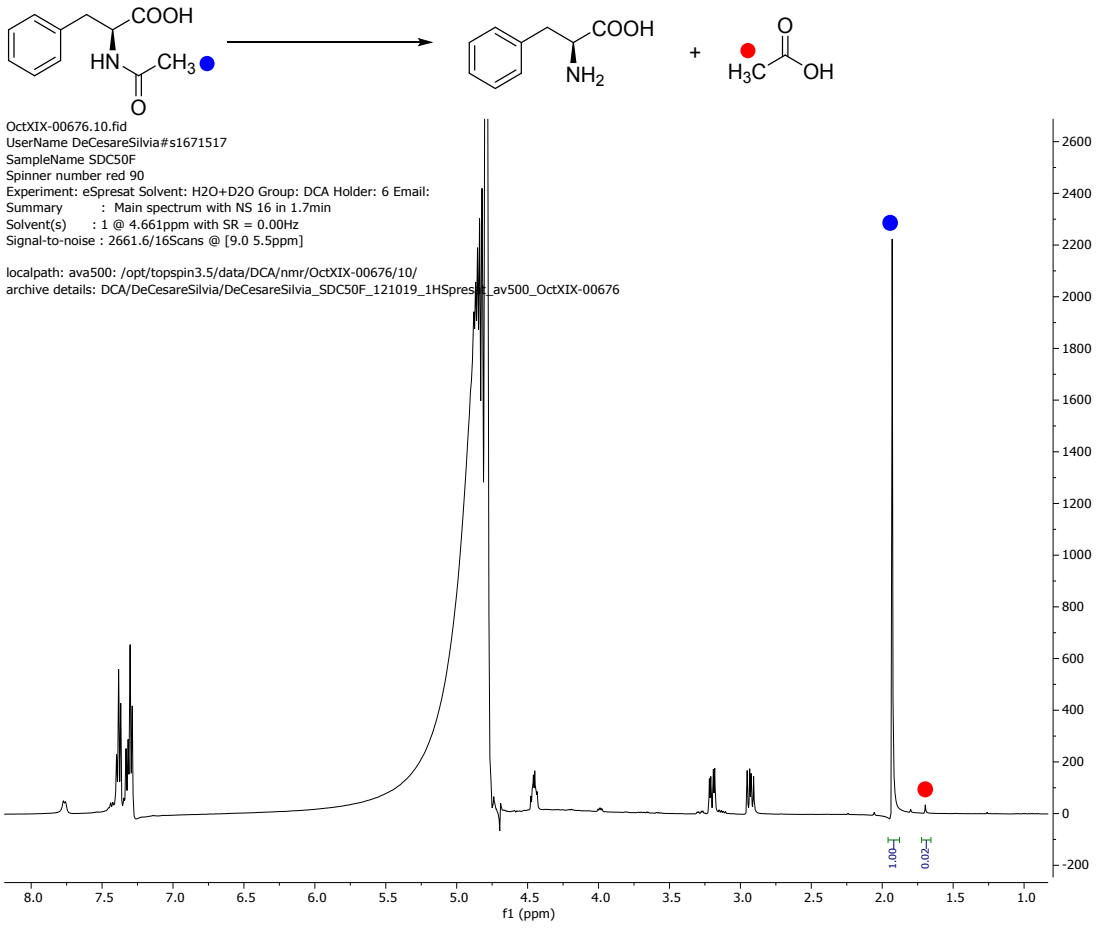
N-Ac-L-Pro hydrolysis



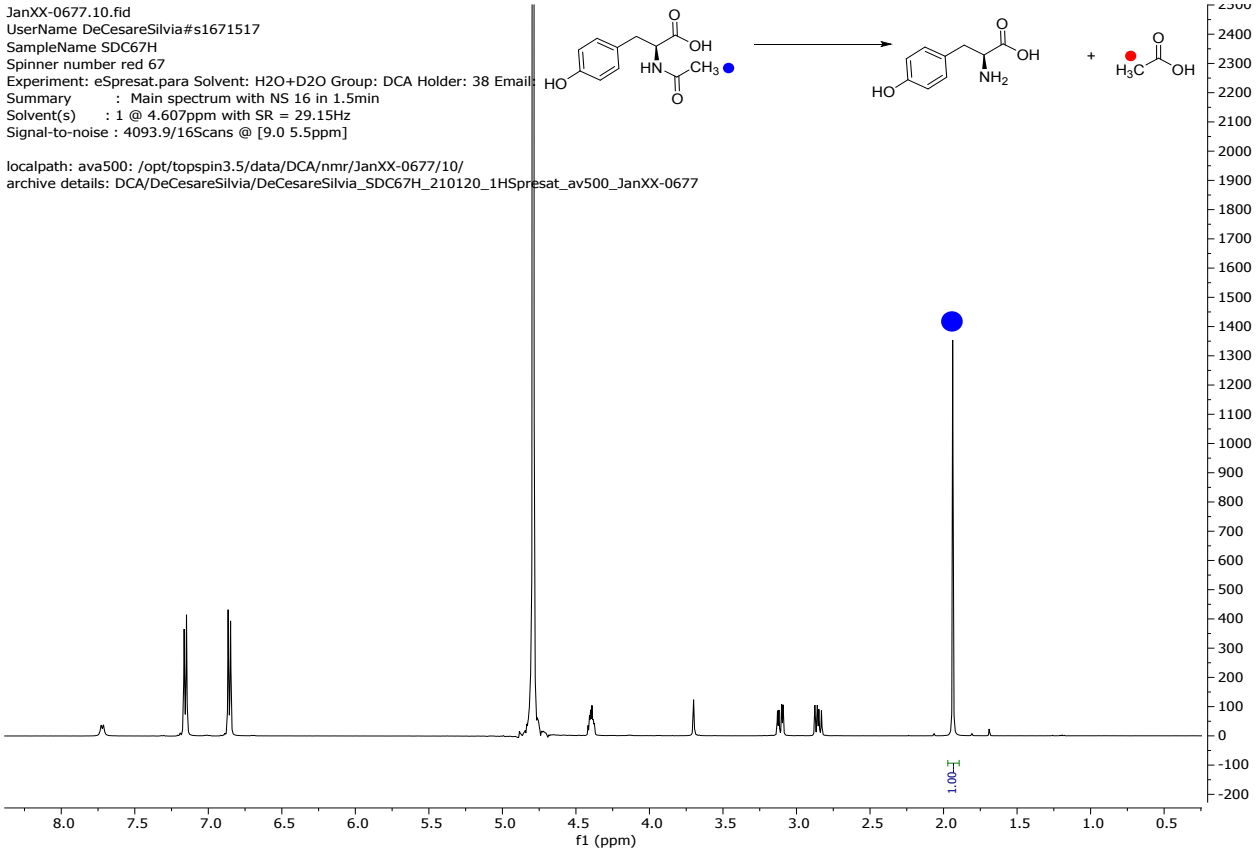
N-Ac-L-Val hydrolysis



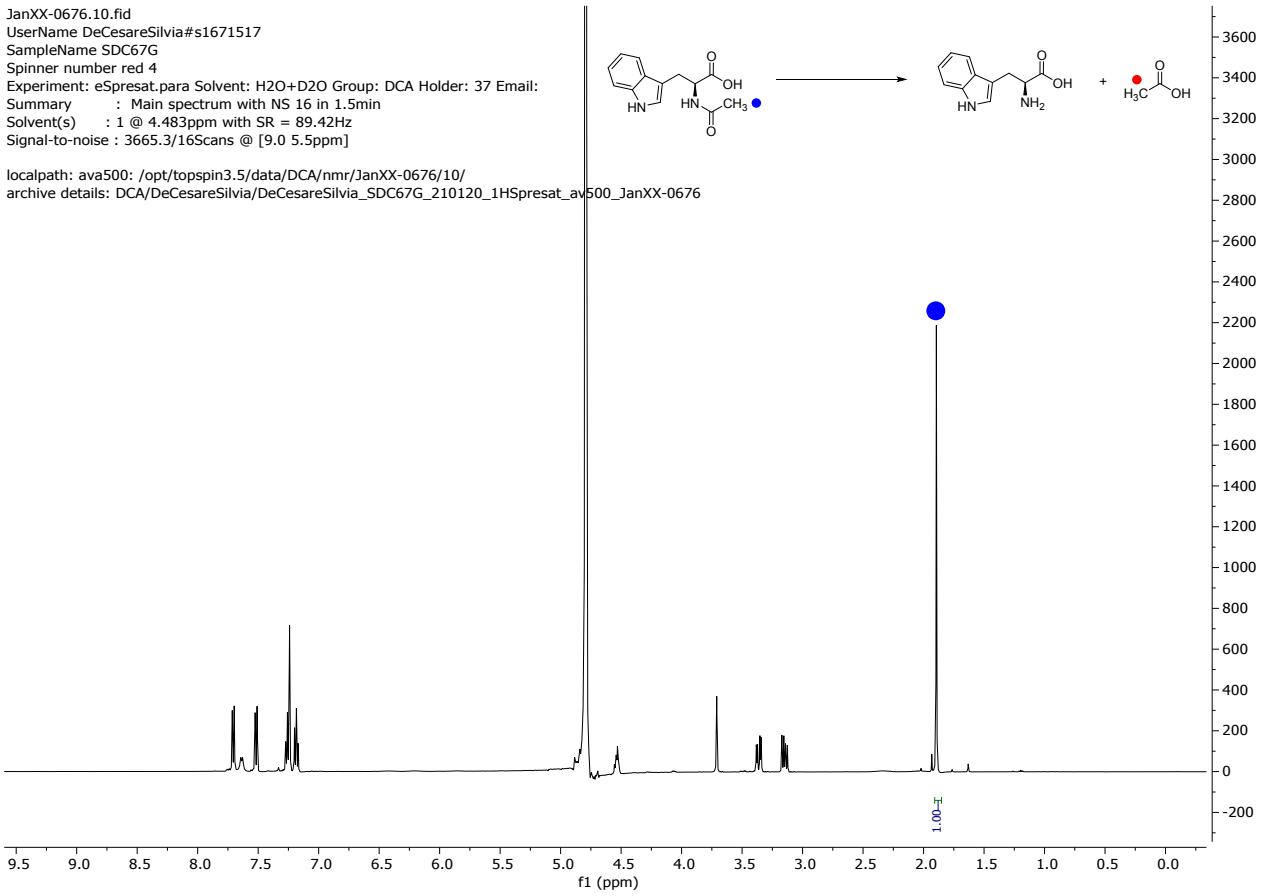
N-Ac-L-Phe hydrolysis



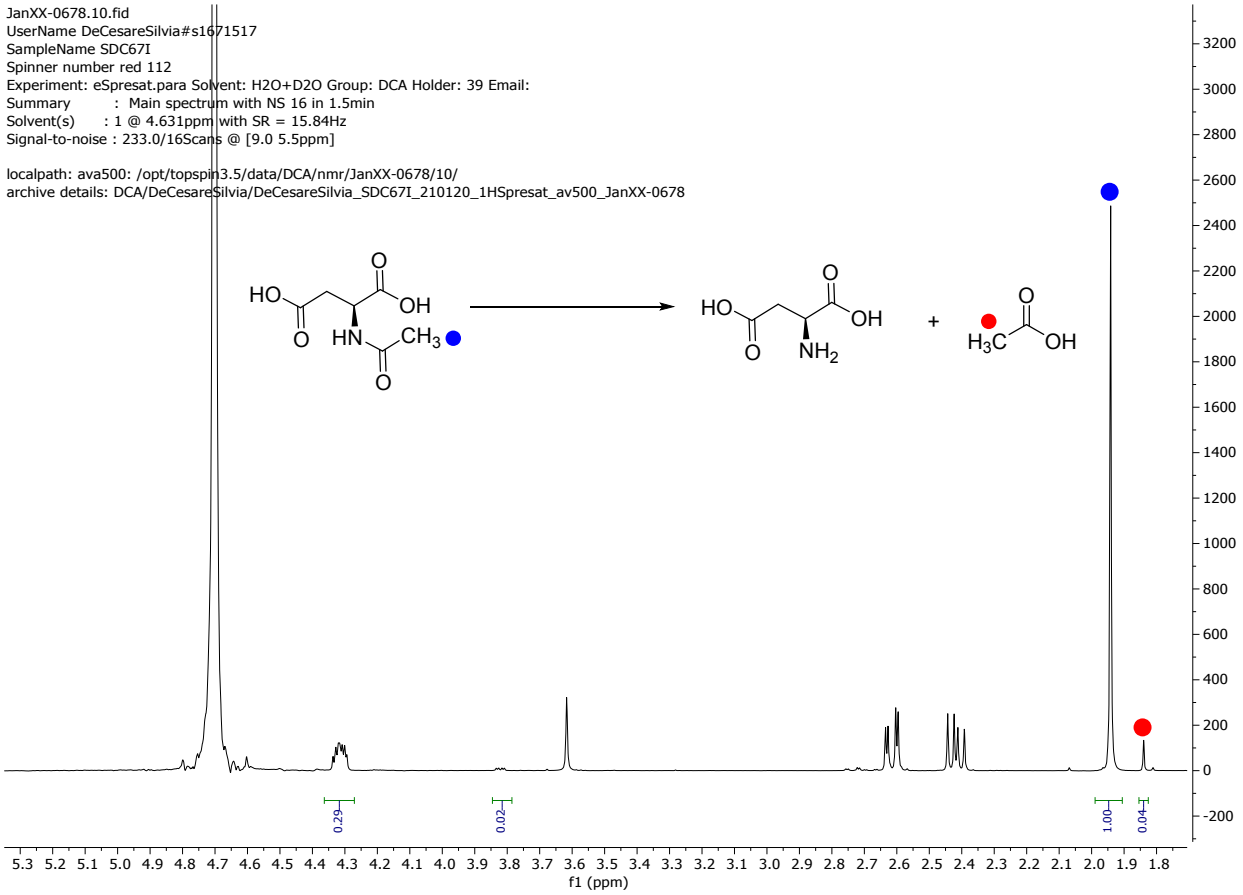
N-Ac-L-Tyr hydrolysis



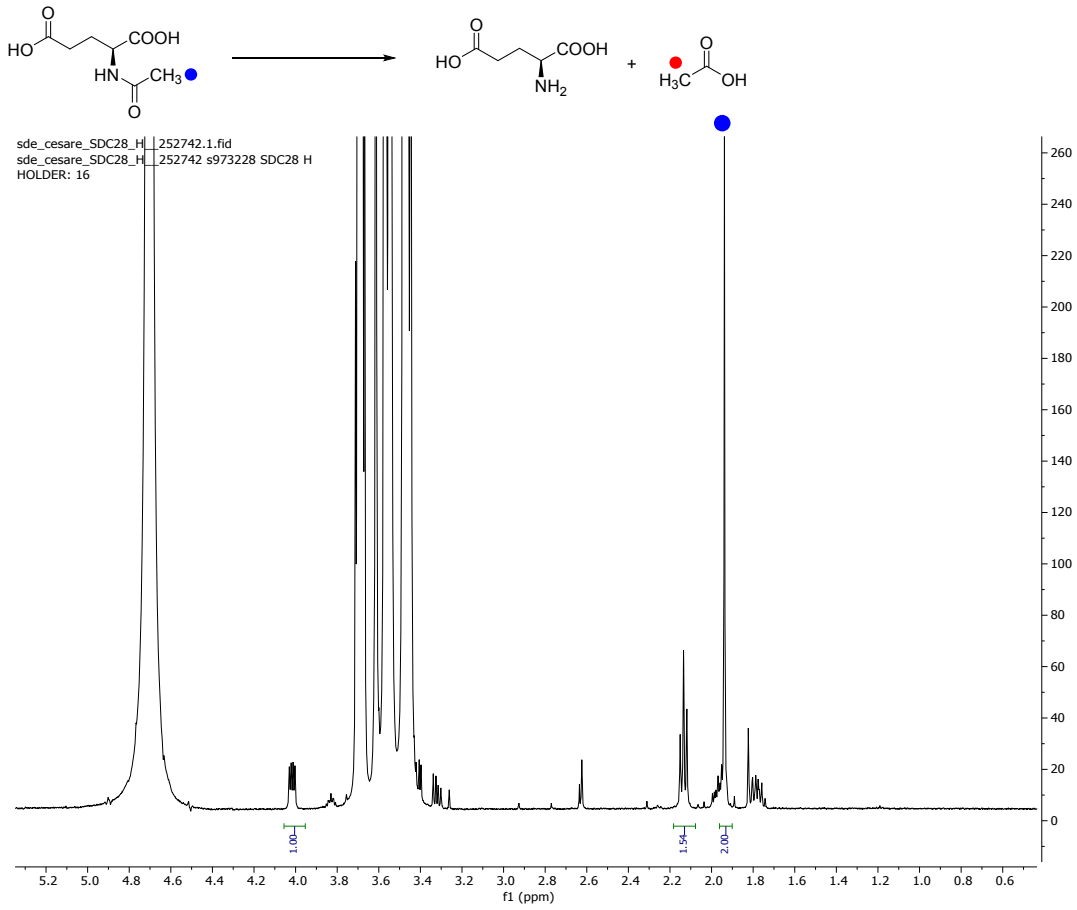
N-Ac-L-Trp hydrolysis



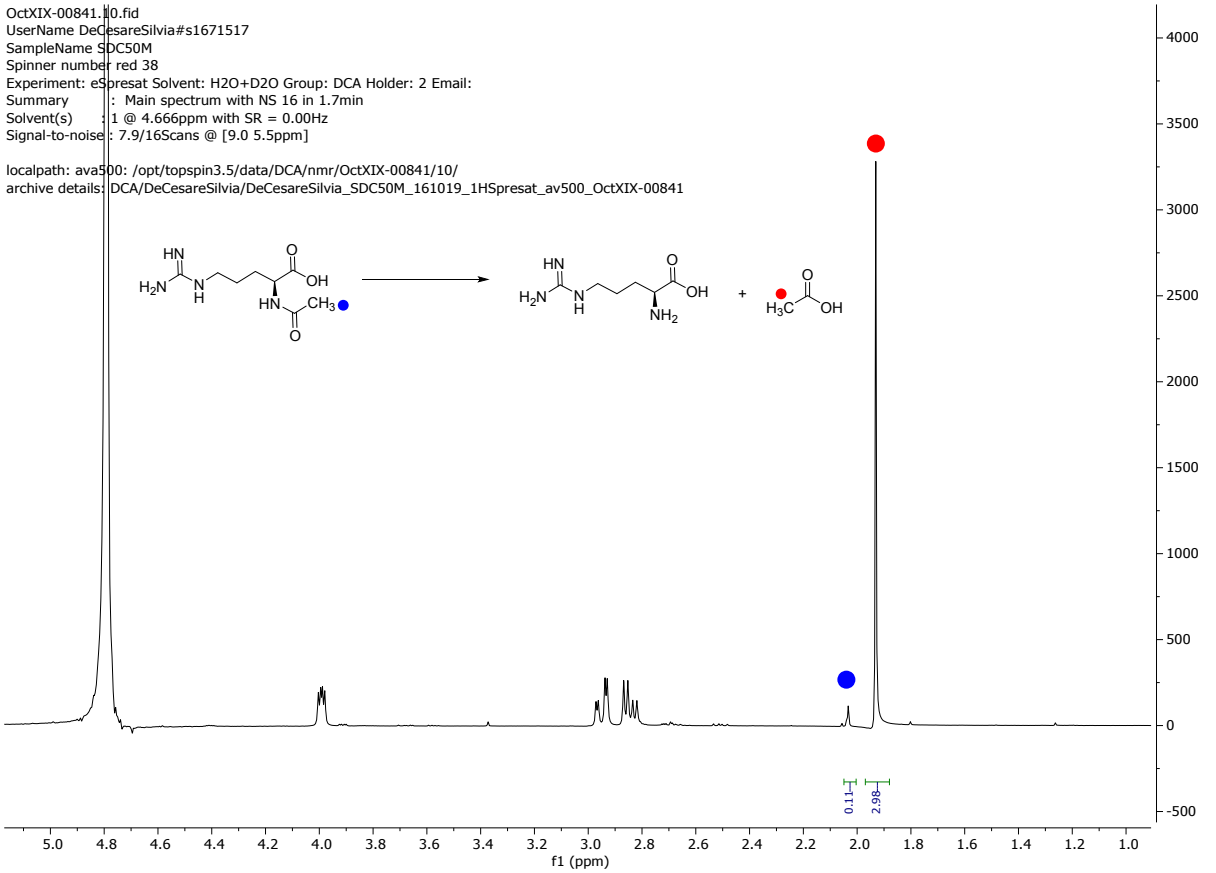
N-Ac-L-Asp hydrolysis



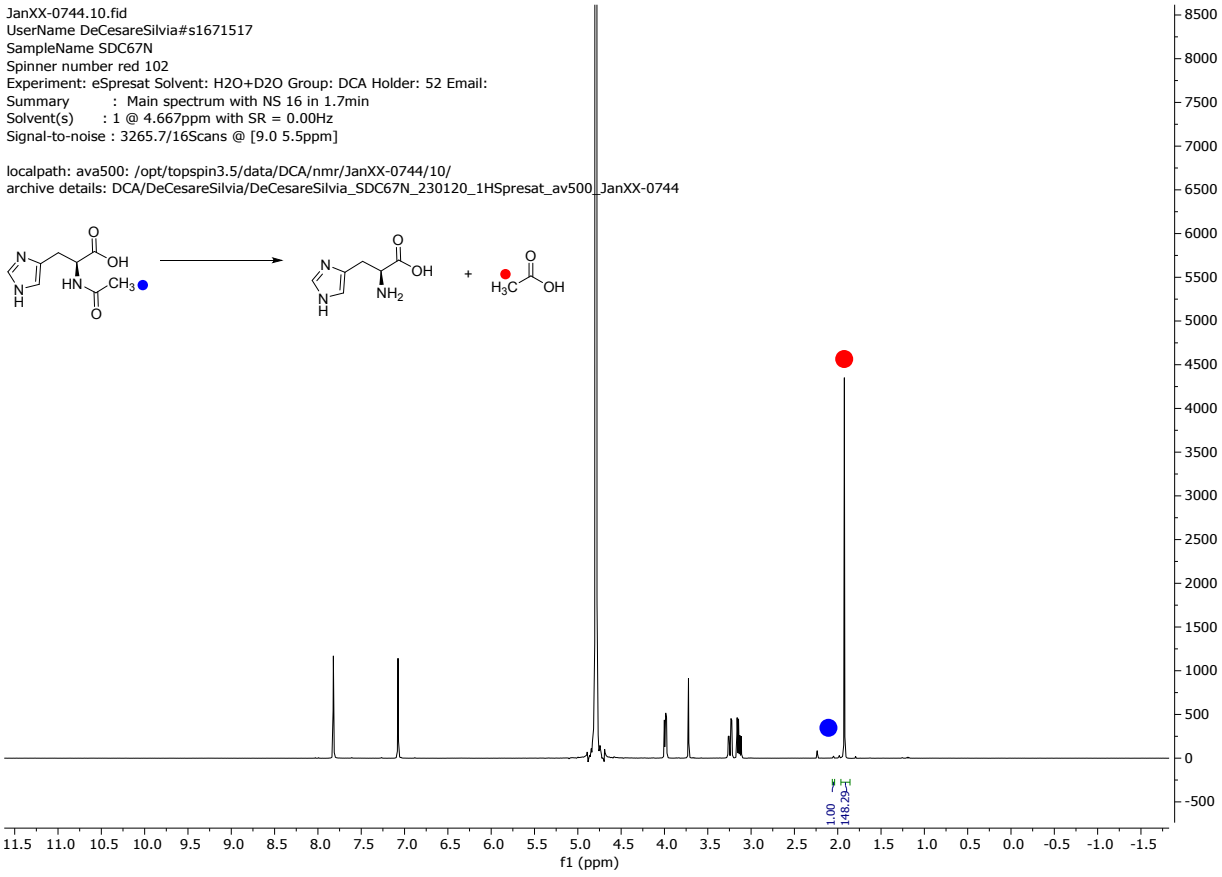
N-Ac-L-Glu hydrolysis



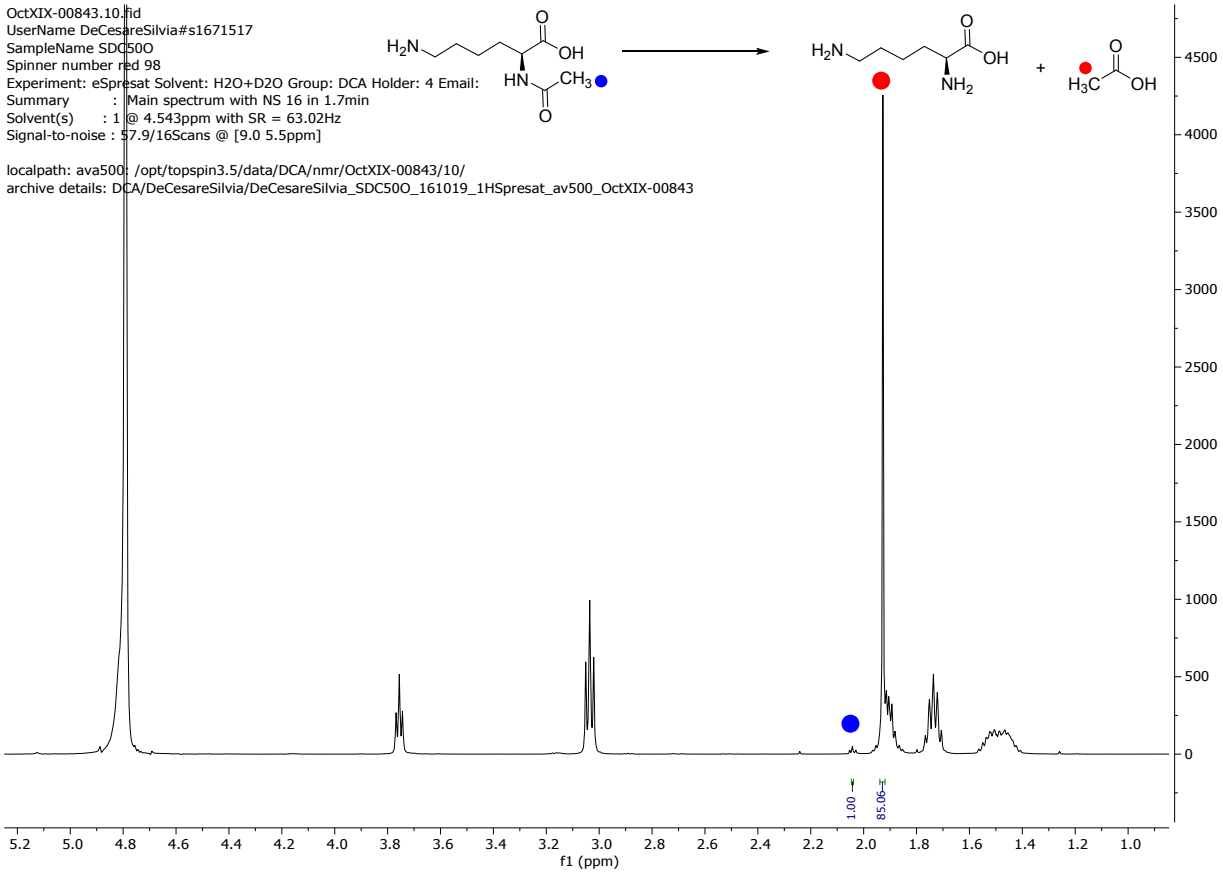
N-Ac-L-Arg hydrolysis



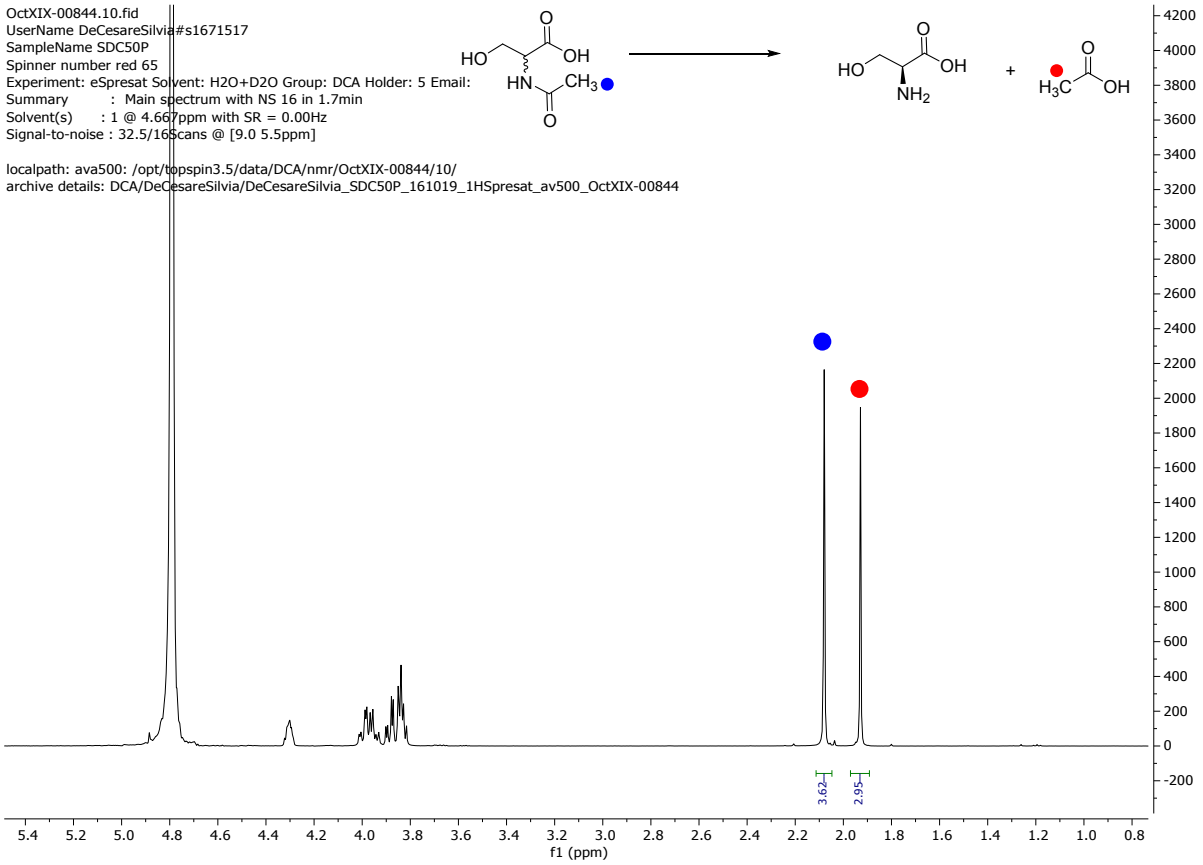
N-Ac-L-His hydrolysis



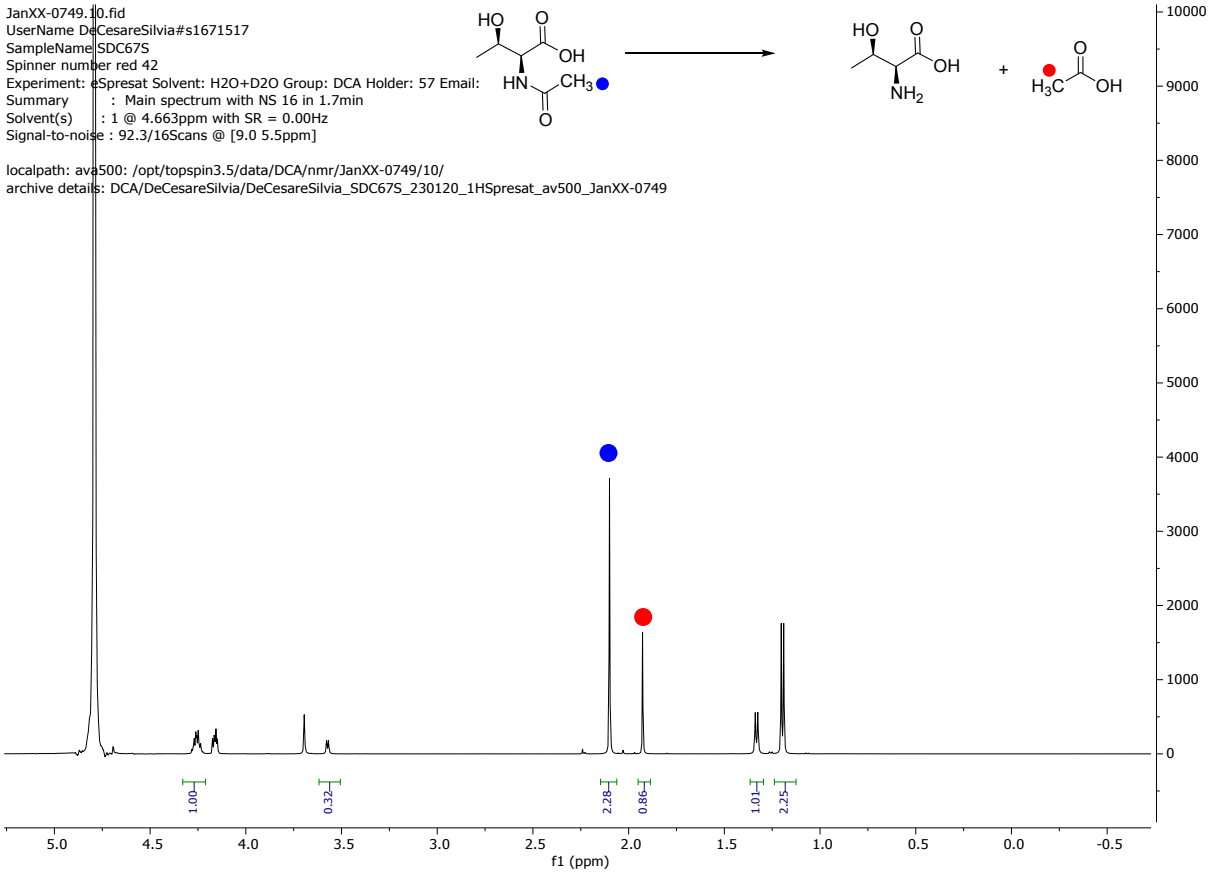
N-Ac-L-Lys hydrolysis



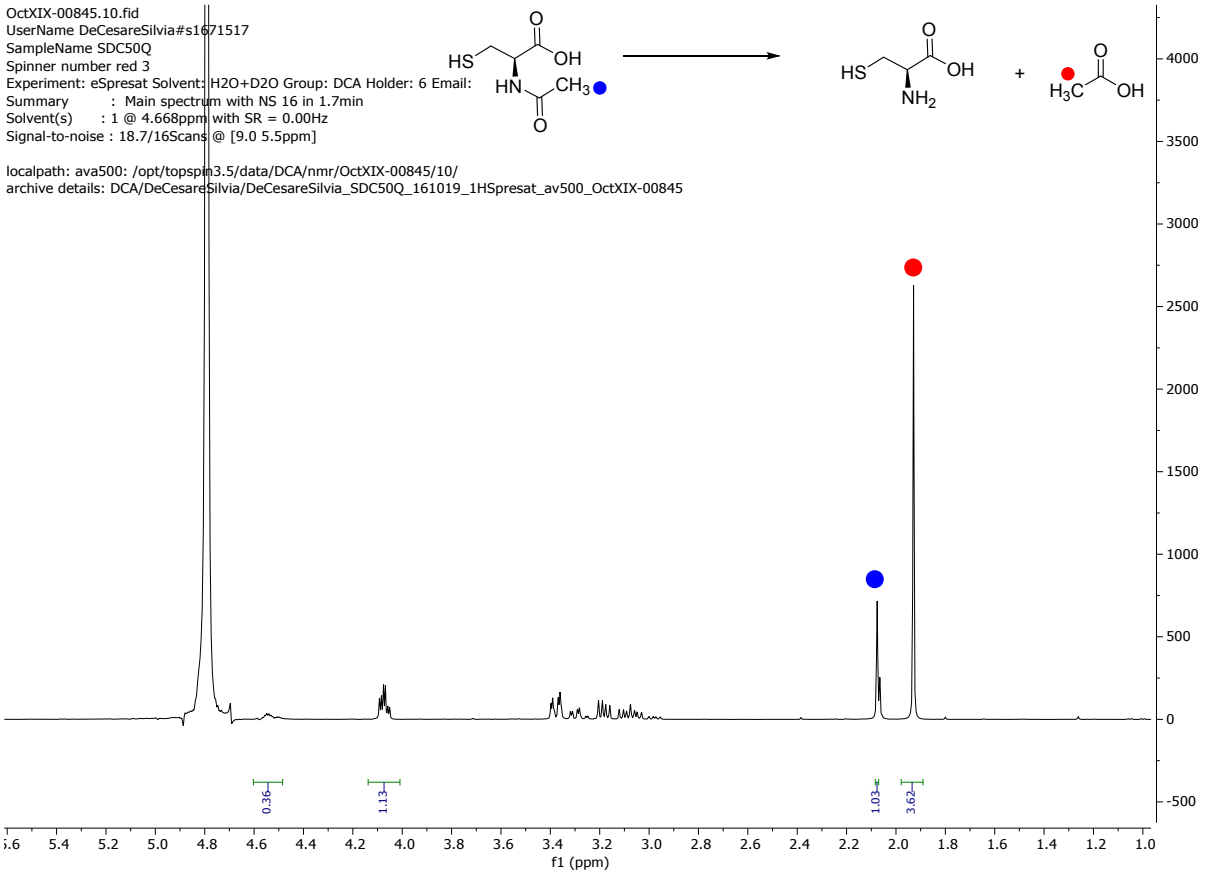
N-Ac-D/L-Ser hydrolysis



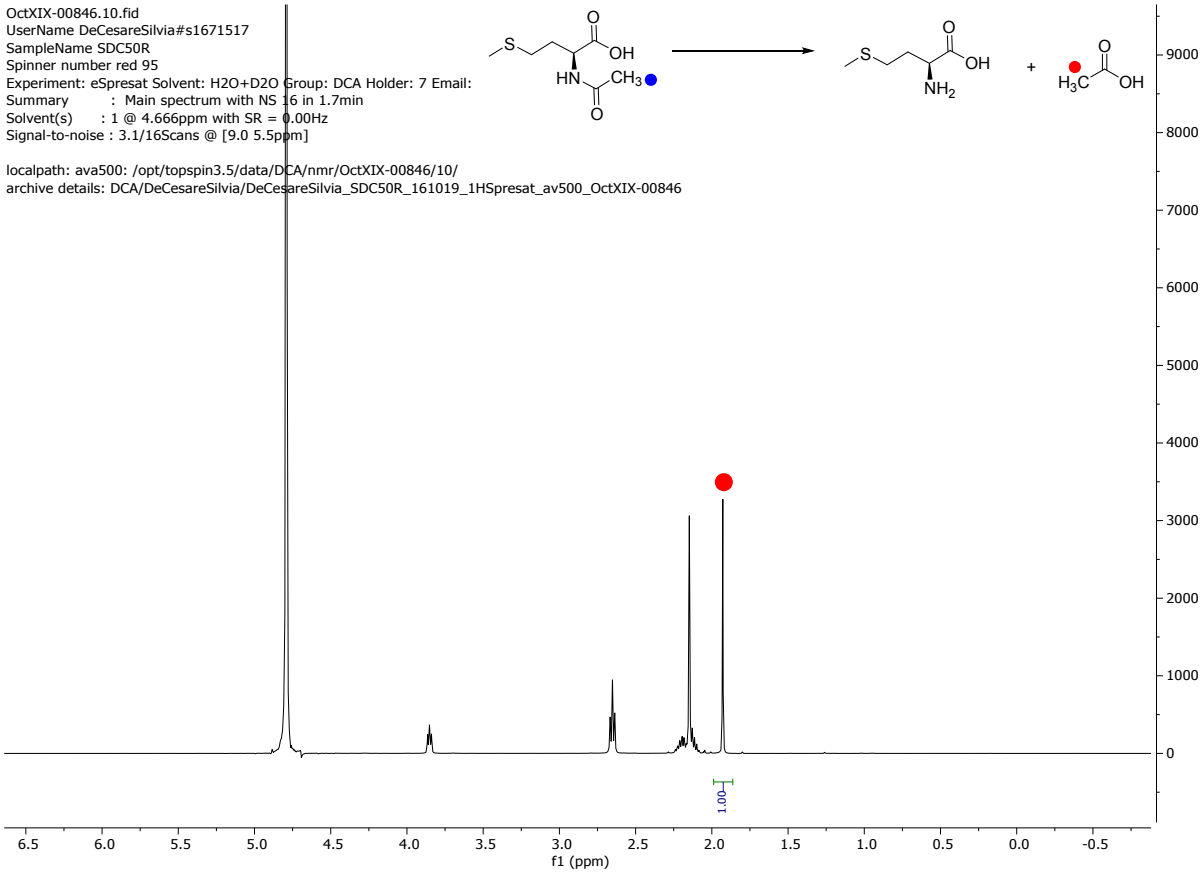
N-Ac-L-Thr hydrolysis



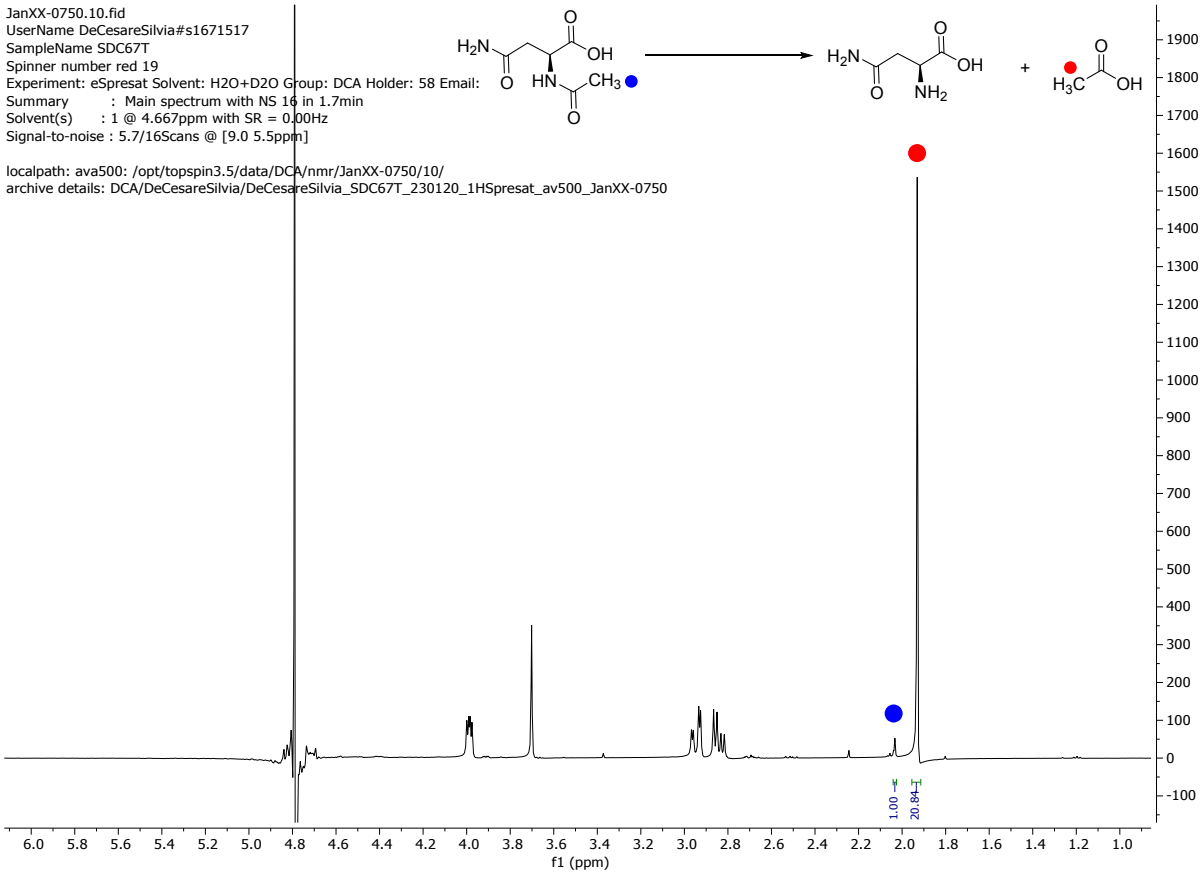
N-Ac-L-Cys hydrolysis



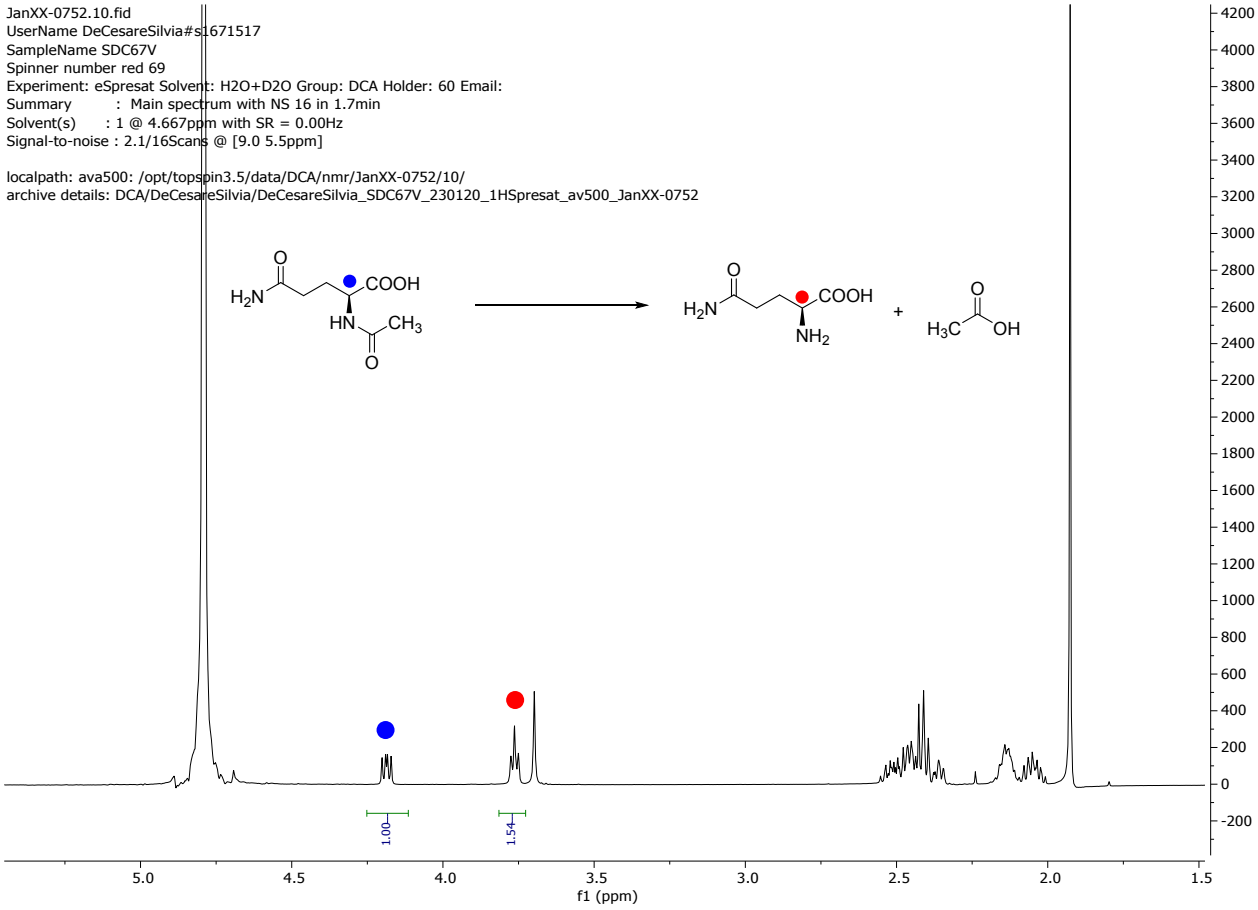
N-Ac-L-Met hydrolysis



N-Ac-L-Asn hydrolysis

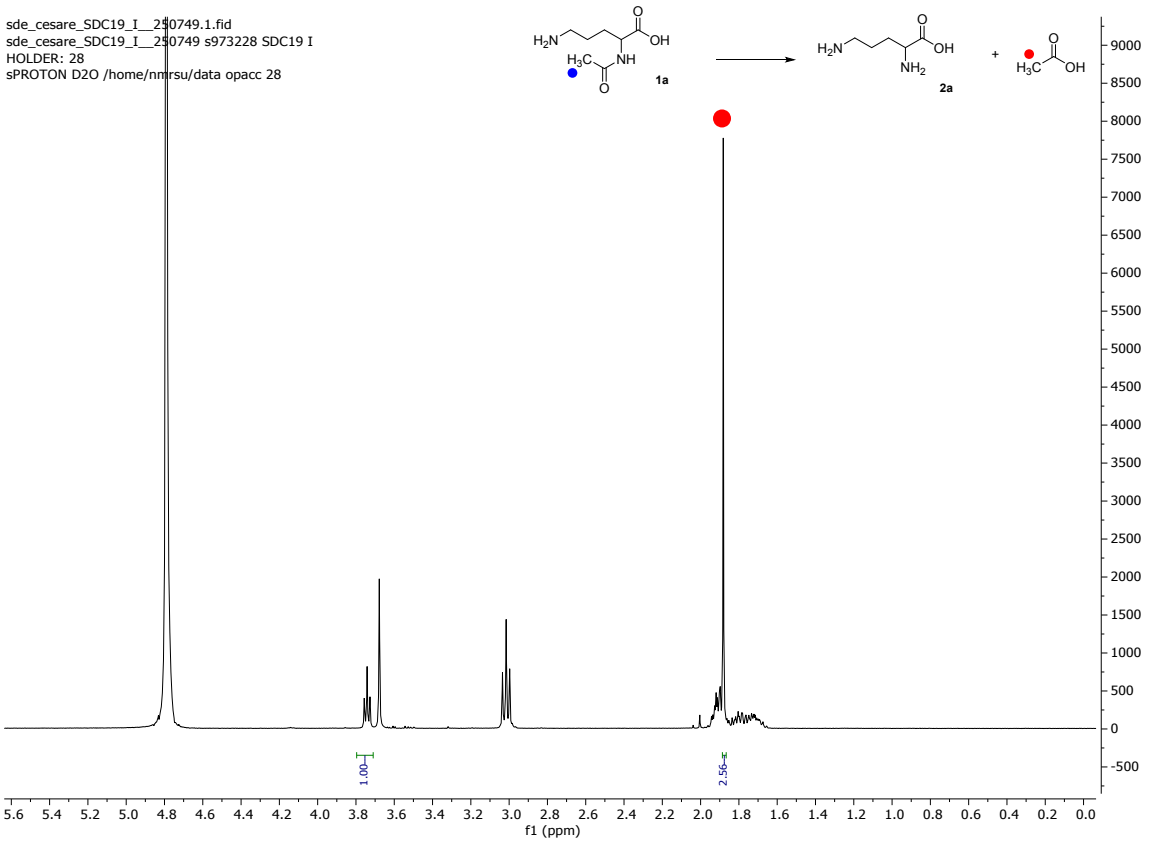


N-Ac-L-Gln hydrolysis

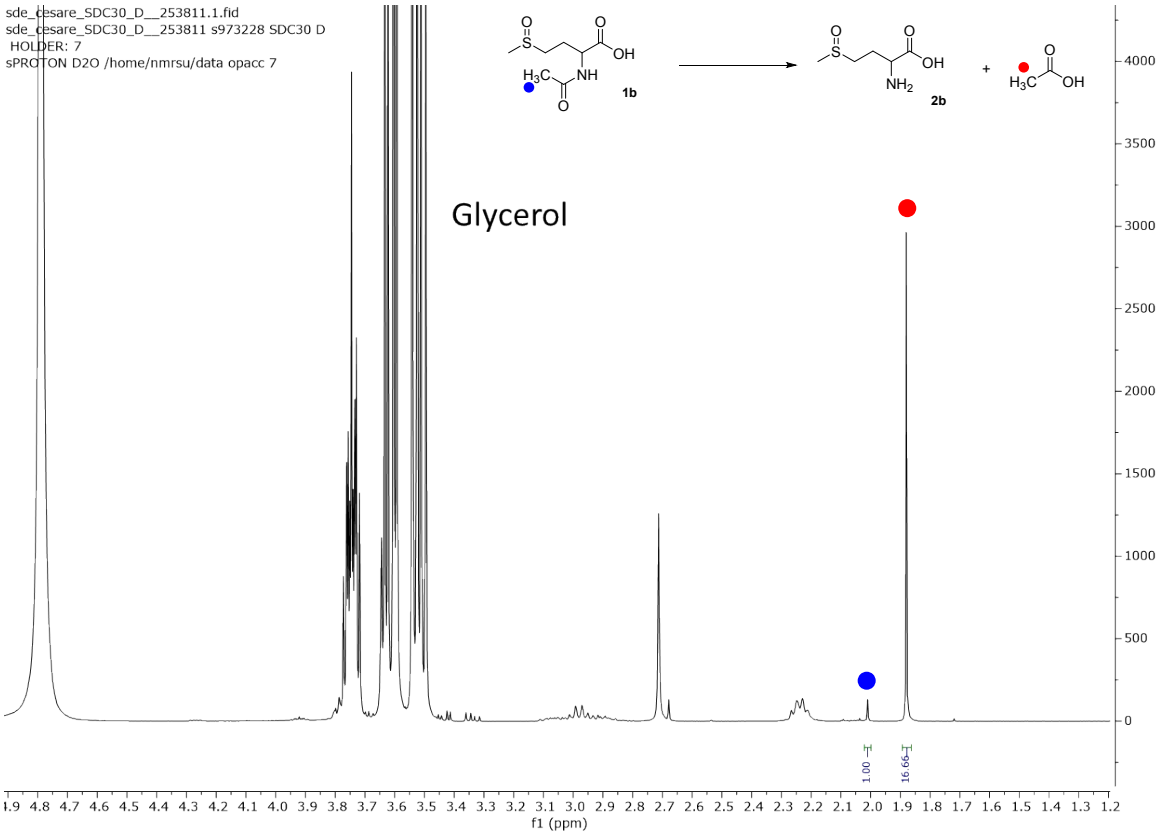


ArgE-catalysed hydrolysis of *N*-Ac-NCAAs

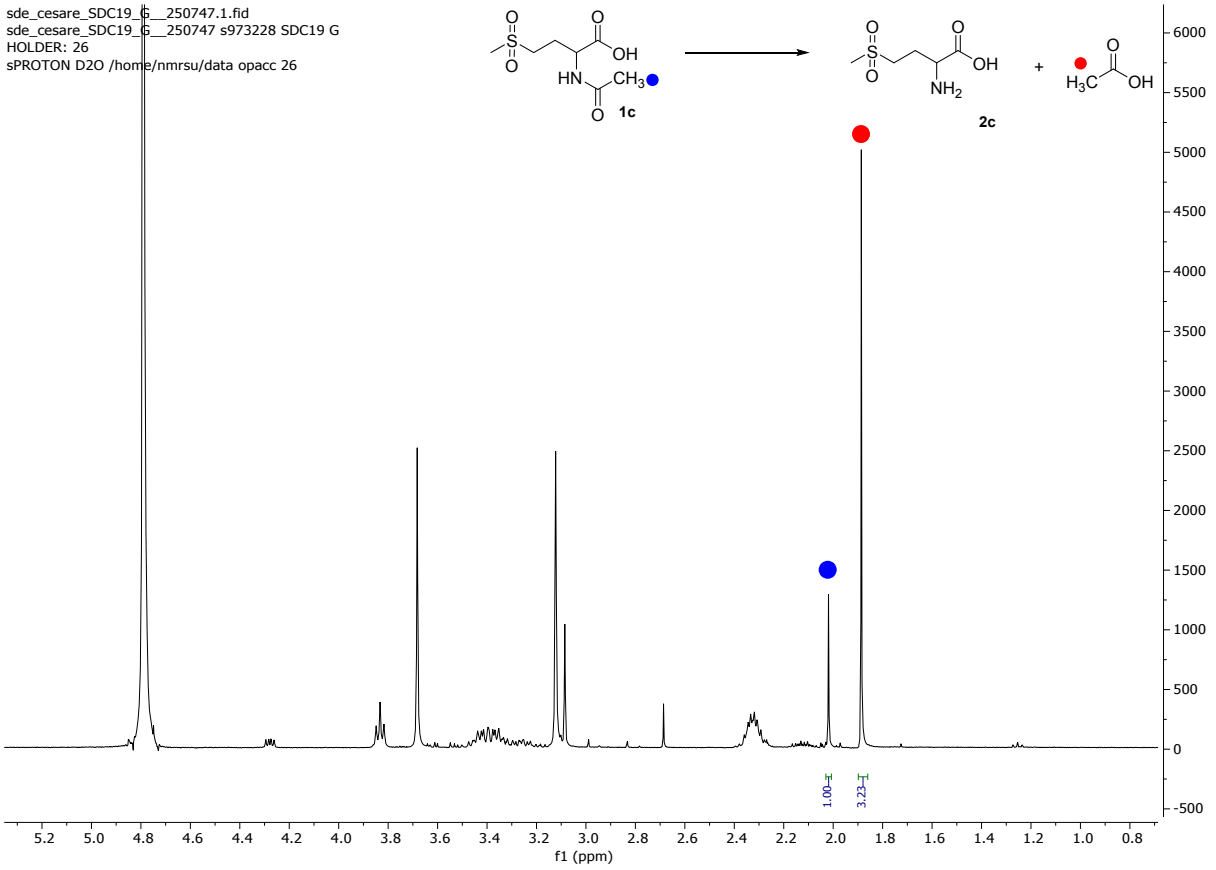
1a hydrolysis



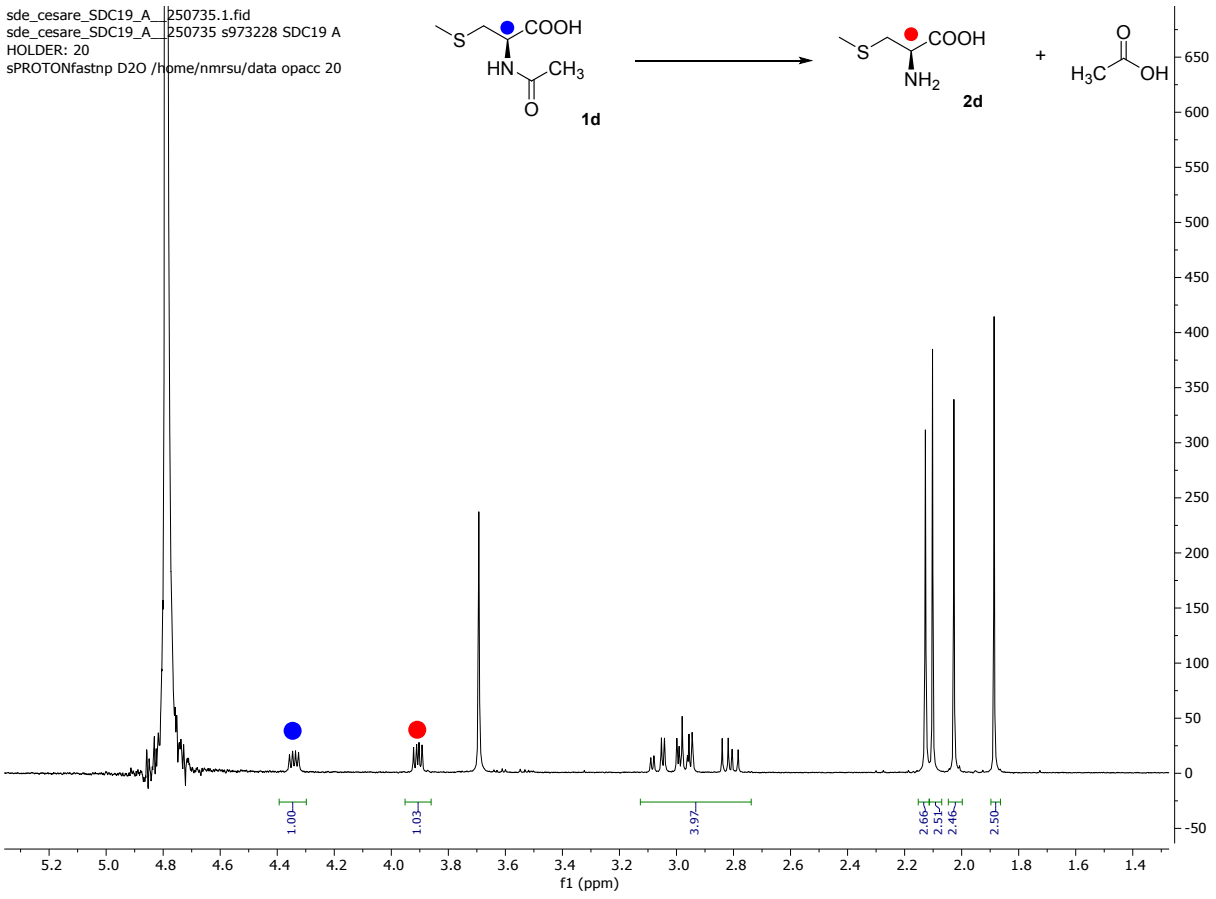
1b hydrolysis



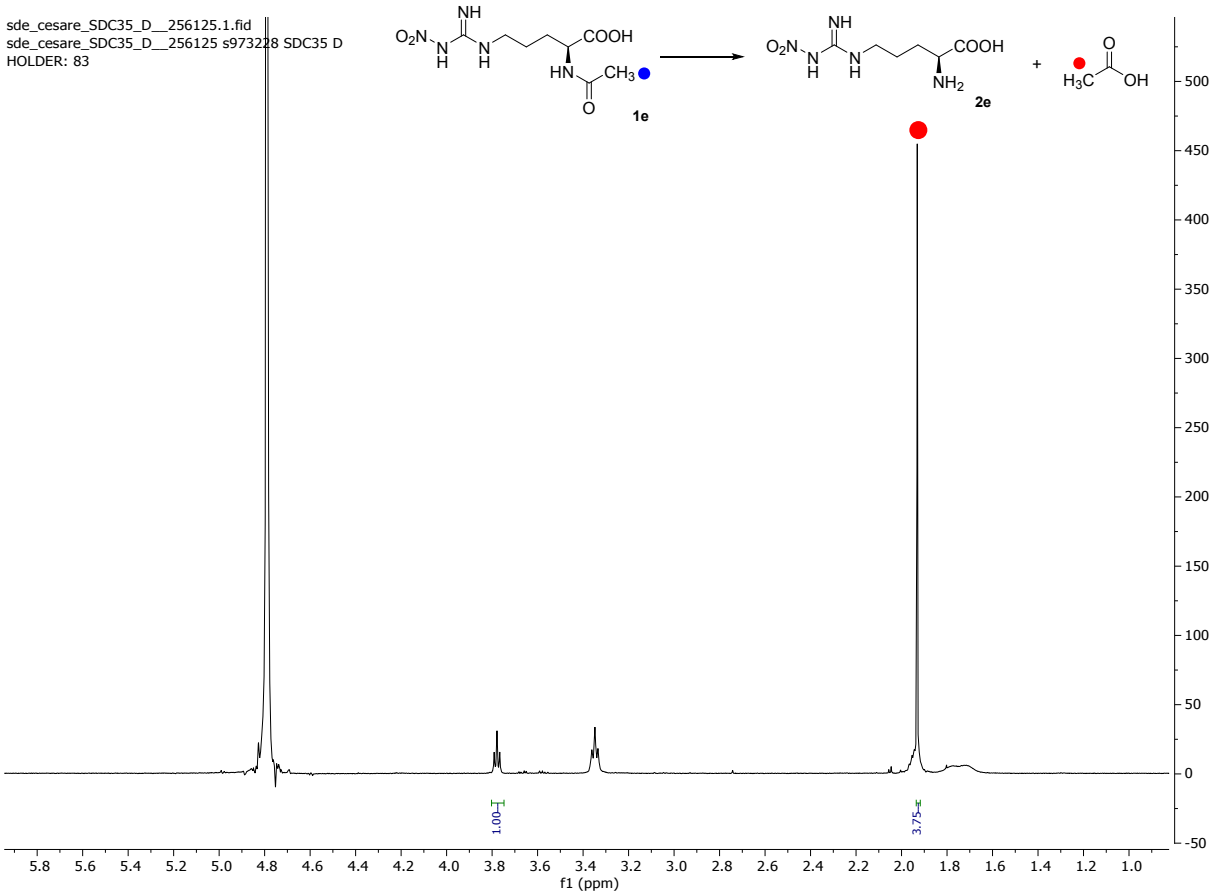
1c hydrolysis



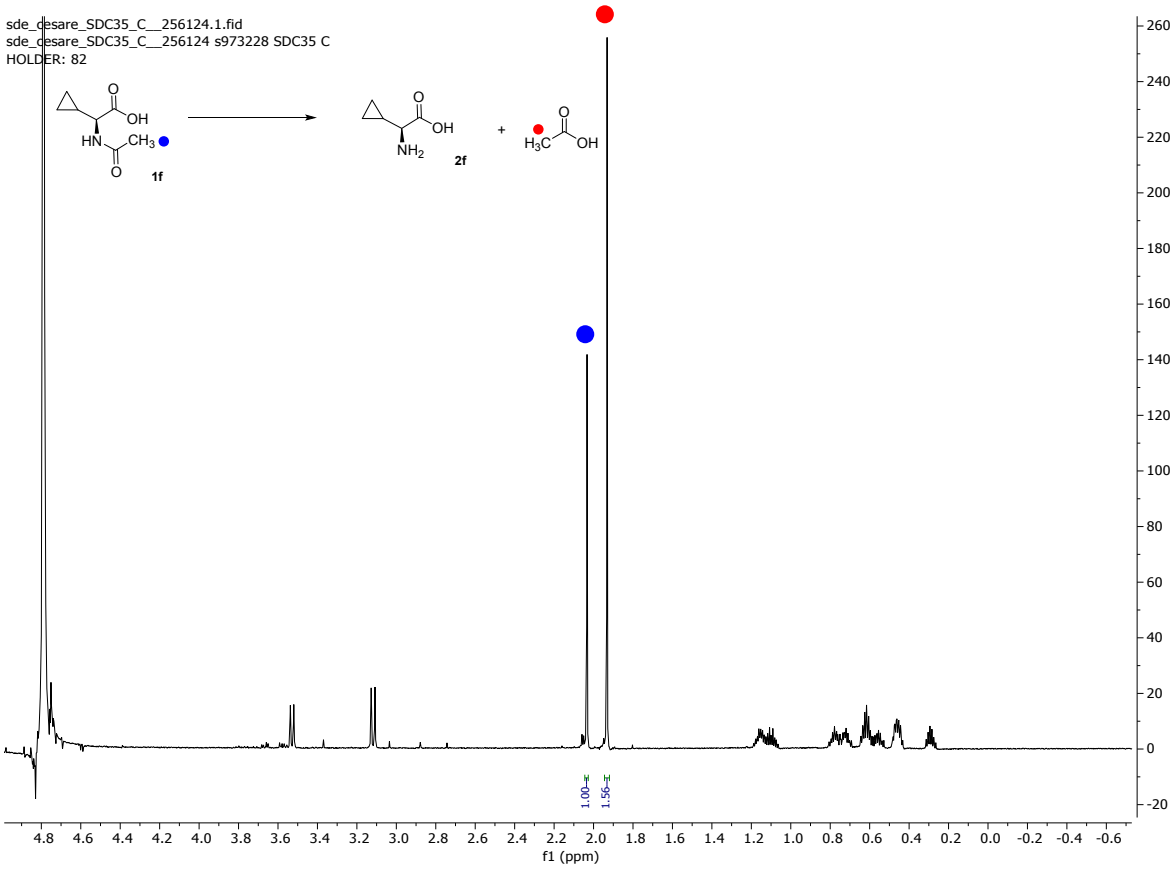
1d hydrolysis



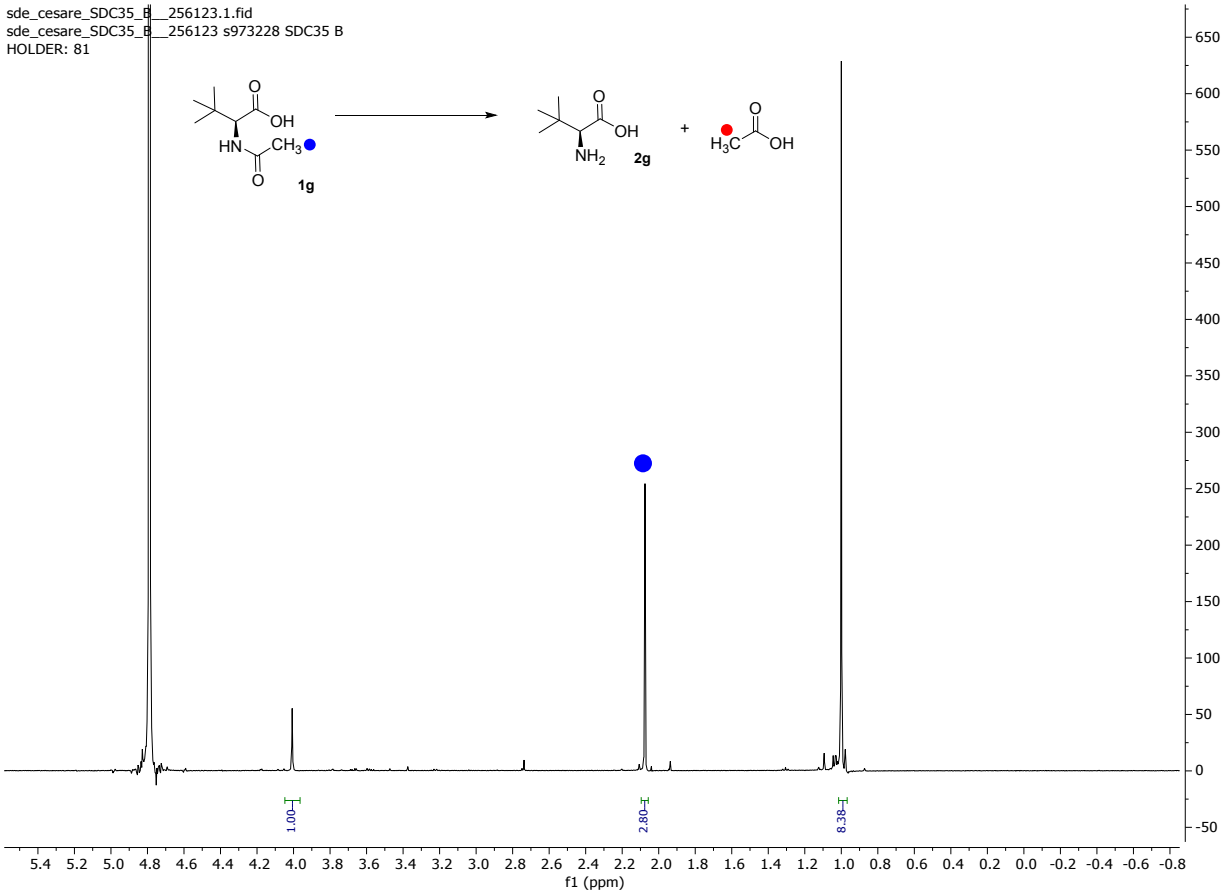
1e hydrolysis



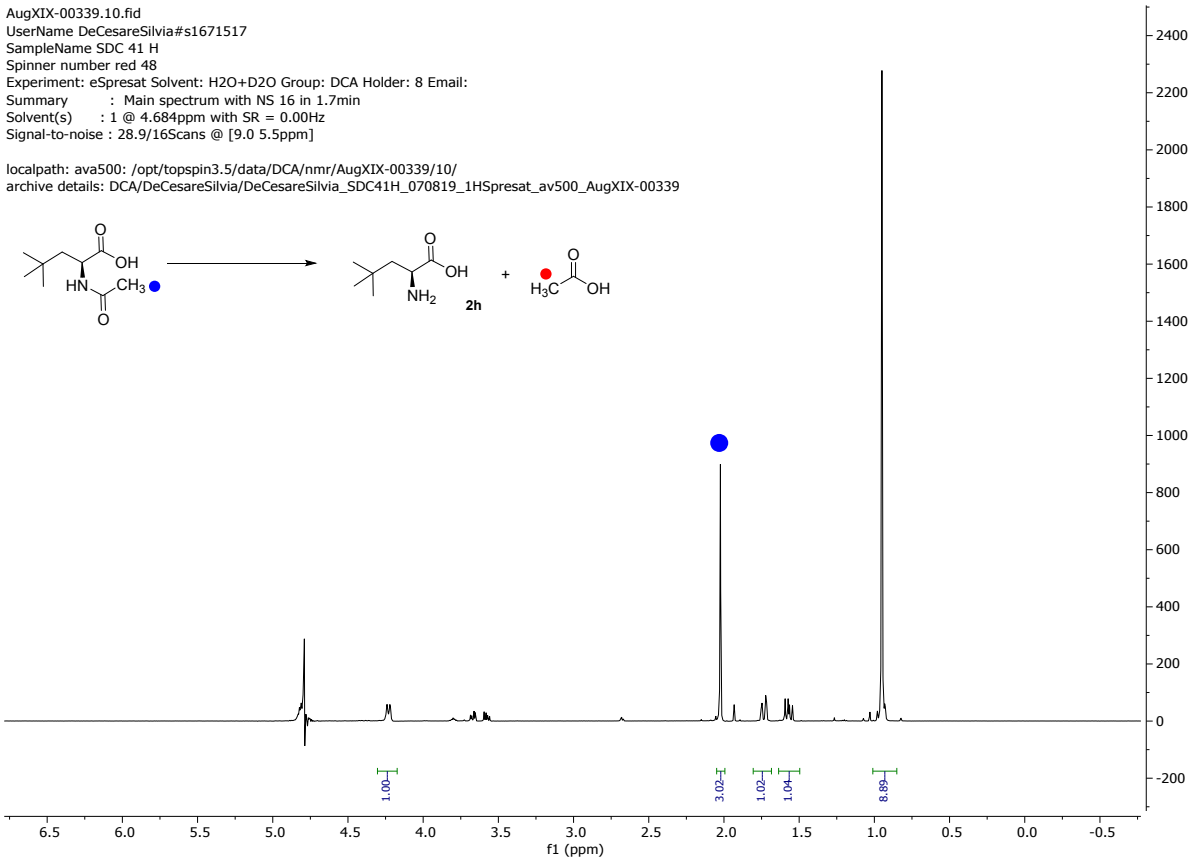
1f hydrolysis



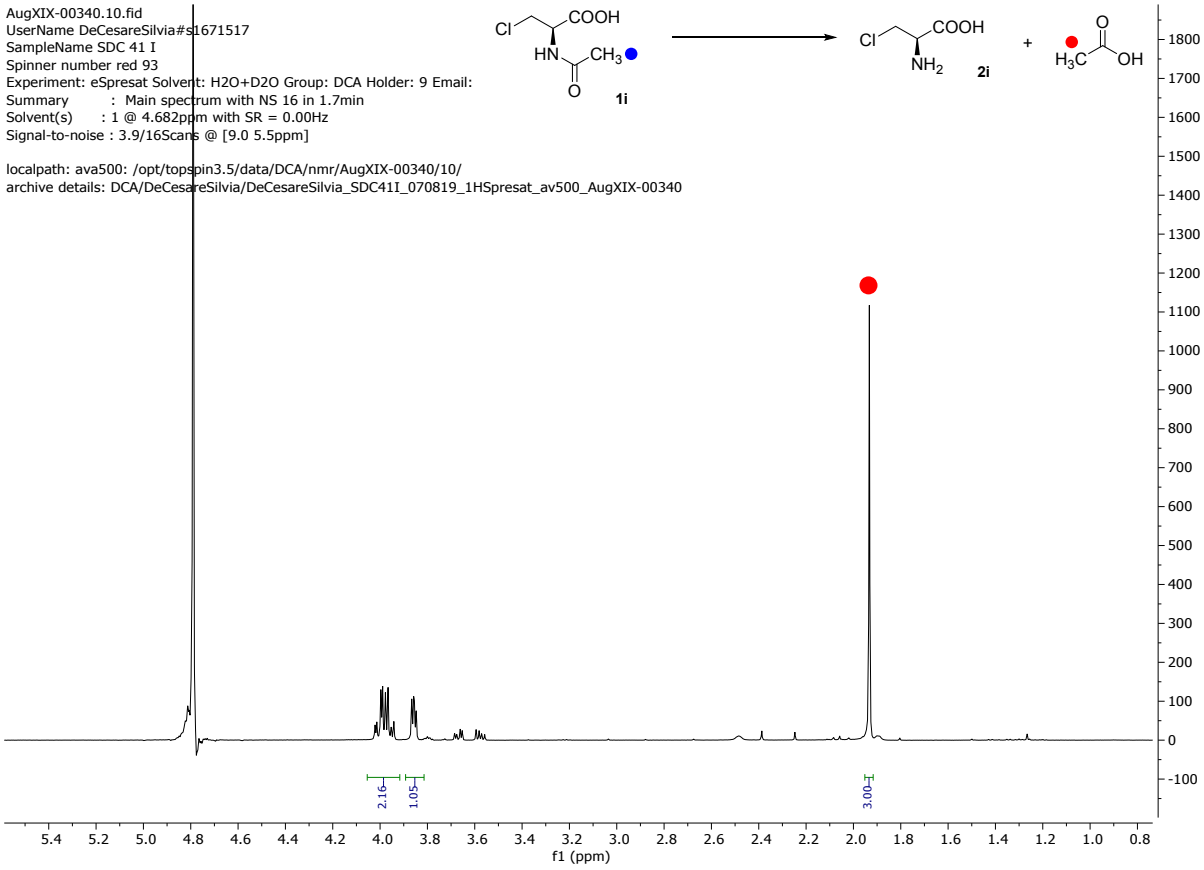
1g hydrolysis



1h hydrolysis



1i hydrolysis



1j hydrolysis