Electronic Supplementary Material (ESI) for Chemical Communications

Supporting Information

1 Materials and Methods

Products

Rink Amide AM resin (0.68 mmole/g) was obtained from Novabiochem. All amino acids were purchased from Novabiochem. L-amino acids were used throughout the synthesis. DMF peptide synthesis grade and *N*-methylpyrrolidone (NMP) were purchased from Biosolve. Dichloromethane (DCM) and *N,N*-di*iso*propylethylamine (DIPEA) were obtained from Aldrich. Trifluoroacetic acid (TFA) and coupling reagents were obtained from Iris Biotech GmbH. DMEQ-TAD was purchased from Wako Chemicals GmbH. All chemicals were used without further purification. All reagents used for automated peptide synthesis were peptide synthesis grade.

Peptide syntheses

Automated peptide syntheses were performed on a 24-reactor block SYRO Multiple Peptide Synthesizer equipped with a vortexing unit (Multisyntech, Witten, Germany). Peptides were synthesized by standard Fmoc/*t*Bu strategy using HBTU/DIPEA couplings.

Analyses

ESI-MS spectra were recorded using an LCQ ion trap mass spectrometer (Finnigan MAT). RP-HPLC analyses were performed on an Agilent 1100 Series instrument with a Phenomenex Luna C18 column (250 x 4.6 mm, 5 µm at 35 °C). A flow rate of 1 ml/min was used with the following solvent systems: 0.1% TFA in H_2O (A) and MeCN (B). The column was flushed for 3 min with 100% A, then a gradient from 0 to 100% B over 15 min was used, followed by 5 min of flushing with 100% B. LC-MS data were collected on an Agilent 1100 Series instrument with a Phenomenex Luna C18(2) column (250 x 4.6 mm, 5 µm at 35 °C) connected to an ESMSD type VL mass detector with a flow rate of 1 ml/min was used with the following solvent systems: 5 mM NH_4OAc in H_2O (A) and MeCN (B). The column was flushed with 100% A for 2 min, then a gradient from 0 to 100% B over 15 min was used, followed by 5 min of flushing with 100% B.

All NMR spectra were measured on an Avance II Bruker Spectrometer operating at a ^{1}H frequency of 700 MHz and equipped with a $^{1}H/^{13}C/^{15}N$ TXI-z probe. Measurements a -50°C were recorded with Avance II Bruker spectrometer operating at a ^{1}H frequency of 500 MHz. The dry samples (±1.5mg) were dissolved in 550µl DMF-d7 in order to obtain a concentration of 2.2mM. Two series of measurements were performed on the modified peptide sample: one series at 298K and another at elevated temperatures in order to resolve overlapping signals in the amide region. All spectra were referenced to the solvent signals at 8.03 (1) ppm and 163.15 (3) ppm for the ^{1}H and ^{13}C frequencies respectively. The spectra recorded on the samples included 1D ^{1}H, 1D ^{13}C-APT, 1D ^{13}C power gated, 2D ^{1}H-{^{1}H} COSY, ^{1}H-{^{1}H} TOCSY (100ms mixing time), ^{1}H-{^{1}H} NOESY (600ms mixing time).1H-{^{1}H} Off-Resonance ROESY (300ms mixing time), ^{1}H-{^{13}C} HSQC and ^{1}H-{^{13}C} HMBC with both 4 and 8Hz long range coupling constants. All spectra were processed using TOPSPIN 3.1pl2.

2 Protocol for deprotection of peptides

1mg of resin is treated with 300µl of TFA/TIS/H_2O (95:2.5:2.5), after 2 hours, the liquid is filtered of and the resin is washed 3 times with TFA. The filtrate is evaporated and redissolved in MeOH for HPLC and LC-MS analysis.

3 Synthesis of peptide 1:

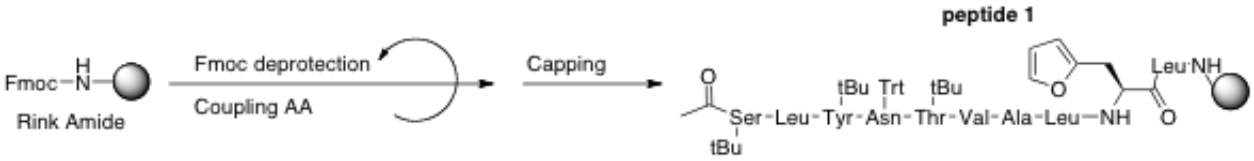

Figure 1: Synthesis of peptide 1

The resin is preswollen in DMF for 30 min and then filtered off. Peptide synthesis is performed on an automated peptide synthesizer using the following protocols for Fmoc deprotection and coupling.

Fmoc deprotection: A solution of 40% piperidine in DMF is added to the resin. The resin is shaken for 3 min and filtered off. Then a solution of 20% piperidine in DMF is added to the resin. The reaction mixture is shaken for 12 min. The resin is filtered off and washed with DMF (6 x 30 s)

Coupling: 5 equiv of a 0.5 M solution of amino acid in DMF, 5 equiv of a 0.5 M solution of HBTU in DMF and 10 equiv of a 2.0 M solution of DIPEA in NMP are added to the resin. The reaction mixture is shaken for 40 min. The resin is filtered off and washed with DMF (4 x 30 s). For difficult couplings, this step is repeated once.

Capping: 6 equiv of acetic anhydride and DIPEA in DMF are added to the resin. The reaction mixture is shaken for 30 min, filtered off and washed with DMF (3x30 s). The capping is repeated once to insure complete conversion.

The peptide is analysed according to the above protocol for deprotection of peptides

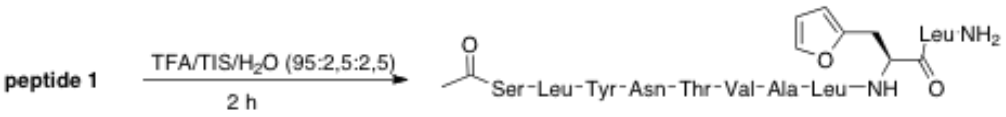

Figure 2: Deprotection of peptide 1 Exact Mass: 1057.54

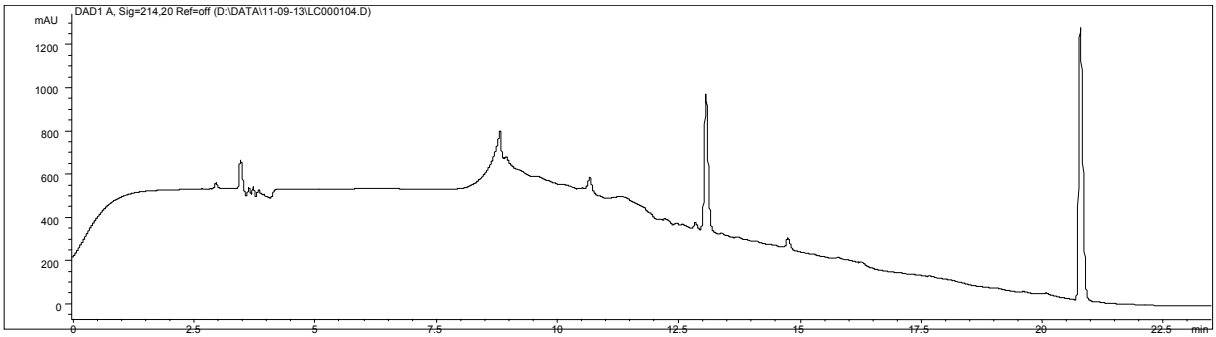

Figure 3: Crude HPLC chromatogram showing the main product (t_R= 13.1min) and trityl protecting group (t_R=20.9min)

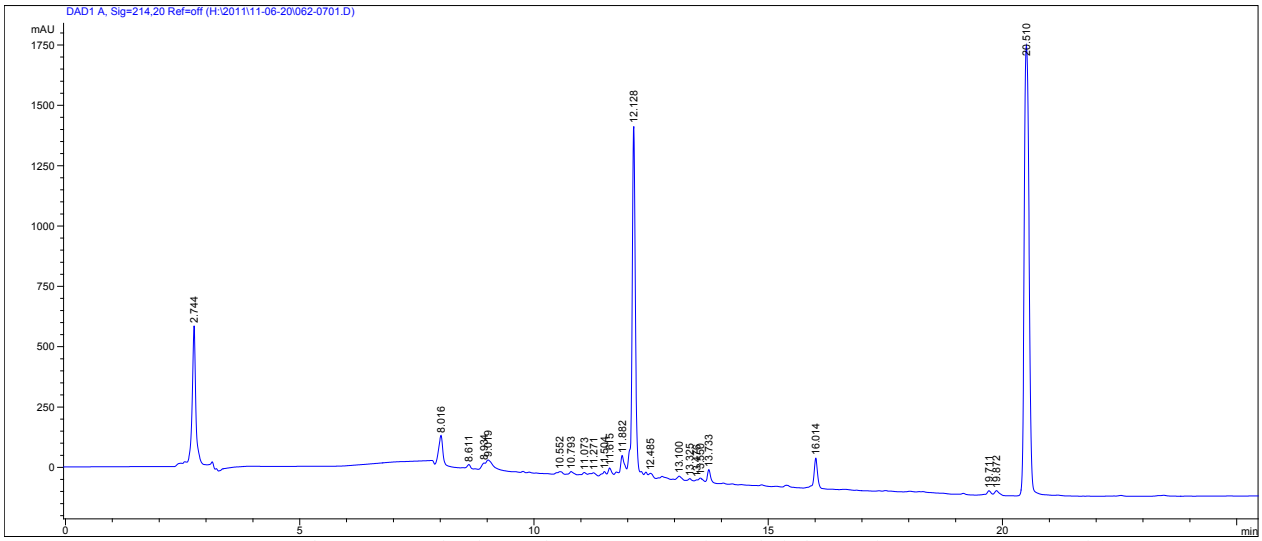

Figure 4: LC-MS chromatogram showing product (t_R= 12.128 min) and trityl protecting group (t_R= 20.510 min)

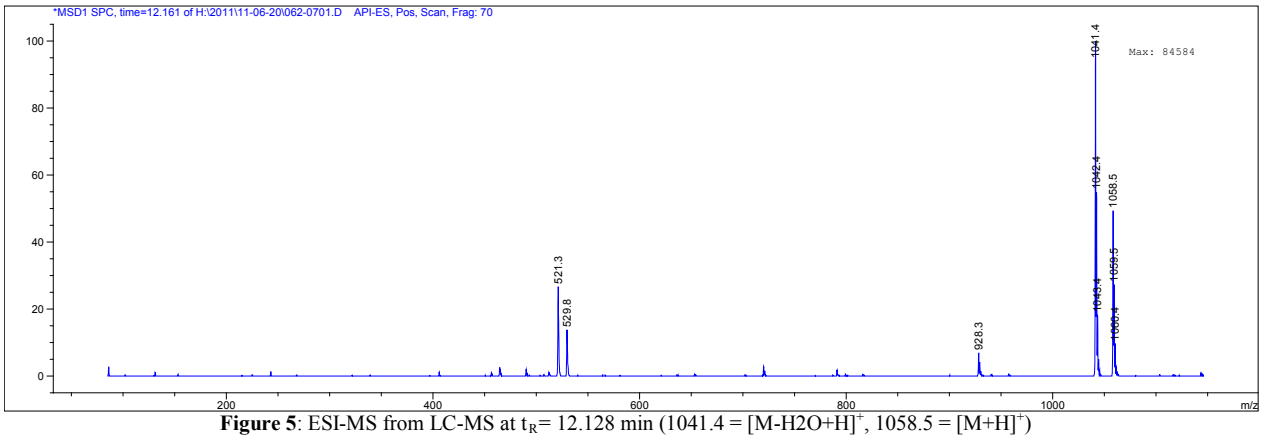

Figure 5: ESI-MS from LC-MS at t_R= 12.128 min (1041.4 = $[M\text{-}H2O+H]^+$, 1058.5 = $[M+H]^+$)

4 Diels-Alder reaction in toluene at room temperature.

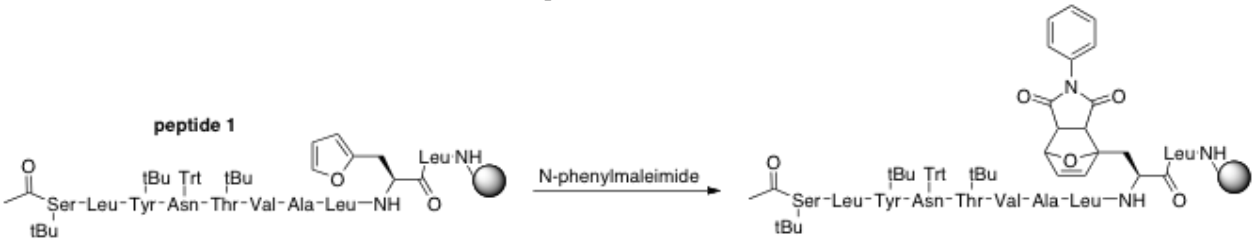

Figure 6: Diels-Alder reaction in toluene at room temperature.

N-Phenylmaleimide (6.6 mg, 38 µmole, 20 equiv) is dissolved in a minimal volume of toluene and added to the resin containing protected peptide **1** (4.5 mg, 1.9 µmole). After 48 hours the resin is washed with 3x DCM (3x30 s), 3x DMF (3x30 s), 3x MeOH (3x30 s), 3x DCM (3x30 s) and dried *in vacuo*.

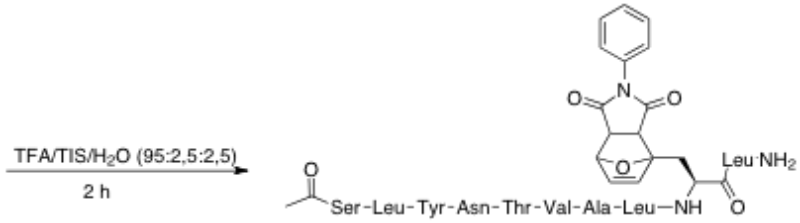

Figure 7: Deprotection of peptide after Diels-Alder reaction in toluene at room temperature.

A small sample of the peptide is cleaved off and analysed by LC-MS as mentioned in the protocol for deprotection of peptides (vide supra).

Exact Mass: 1230.59

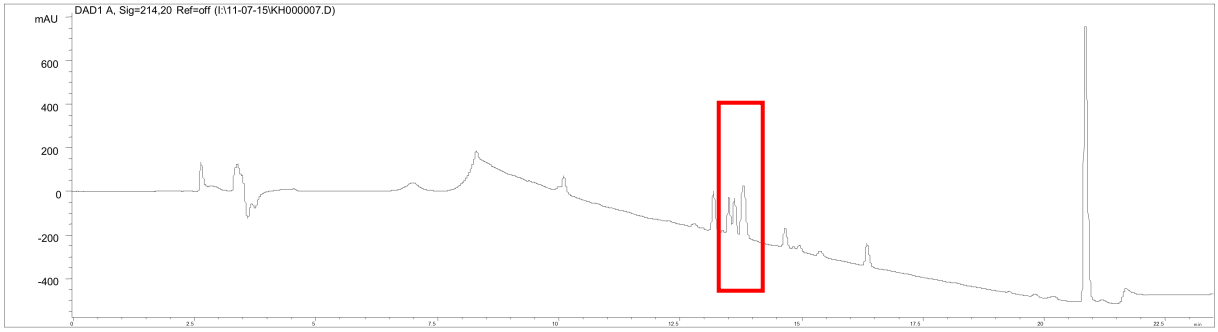

Figure 8: Crude HPLC chromatogram showing the starting product (t_R= 13,1min), Diels-Alder products indicated in red frame and trityl protecting group (t_R=20,9min)

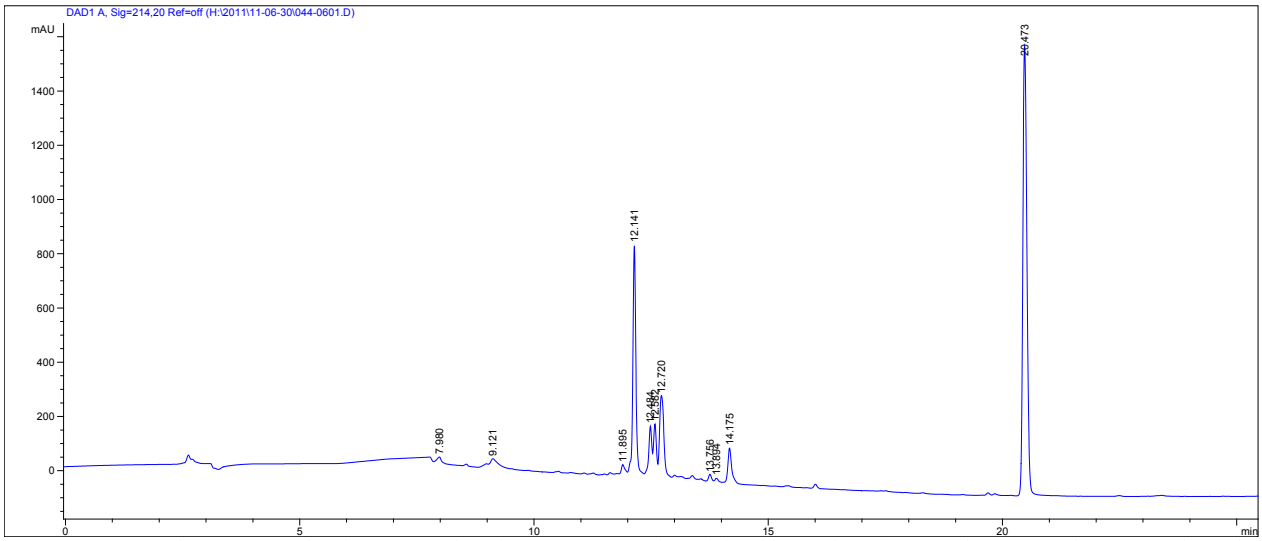

Figure 9: LC-MS chromatogram showing unreacted product (t_R= 12.141 min), Diels-Alder products (12.484 min; 12.582 min; 12.720 min) and trityl protecting group (t_R= 20.510 min)

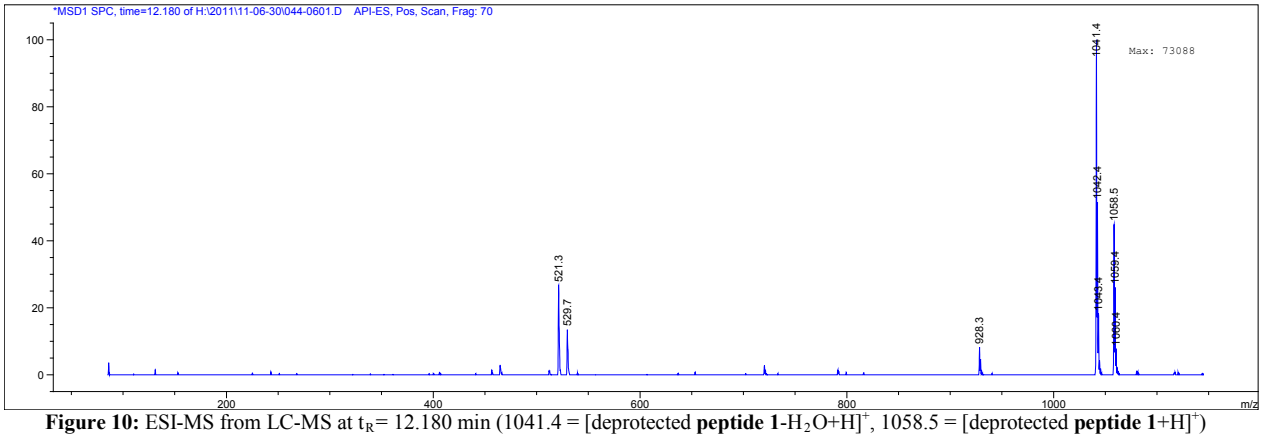

Figure 10: ESI-MS from LC-MS at t_R= 12.180 min (1041.4 = [deprotected **peptide 1**-H_2O+H]$^+$, 1058.5 = [deprotected **peptide 1**+H]$^+$)

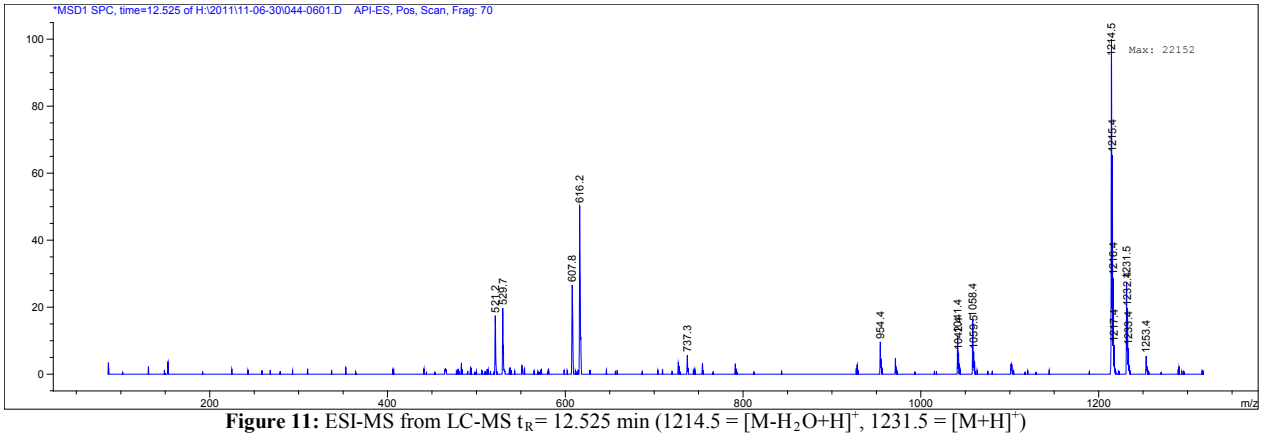

Figure 11: ESI-MS from LC-MS t_R= 12.525 min (1214.5 = [M-H_2O+H]$^+$, 1231.5 = [M+H]$^+$)

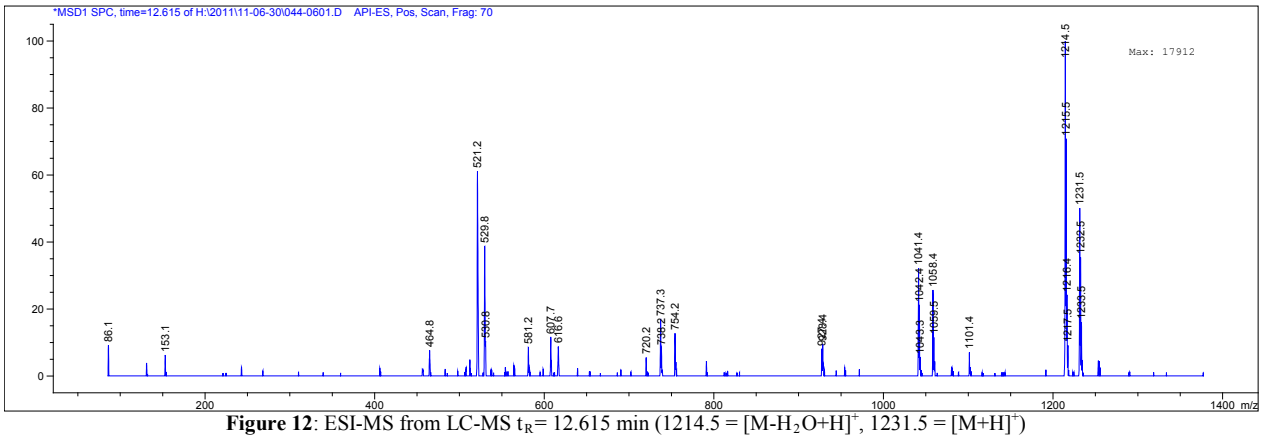

Figure 12: ESI-MS from LC-MS t_R= 12.615 min (1214.5 = $[M\text{-}H_2O+H]^+$, 1231.5 = $[M+H]^+$)

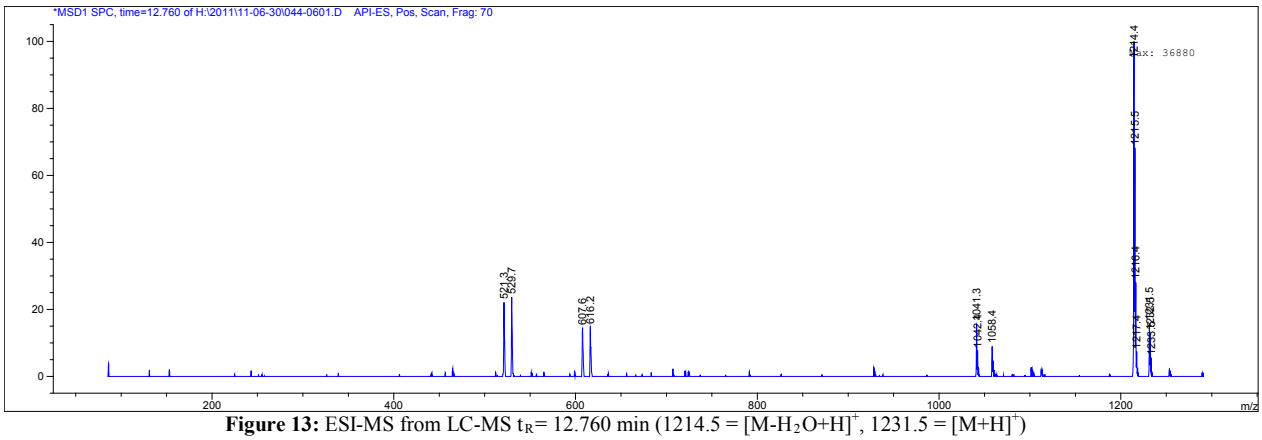

Figure 13: ESI-MS from LC-MS t_R= 12.760 min (1214.5 = $[M\text{-}H_2O+H]^+$, 1231.5 = $[M+H]^+$)

5 Diels-Alder reaction in toluene at 70°C.

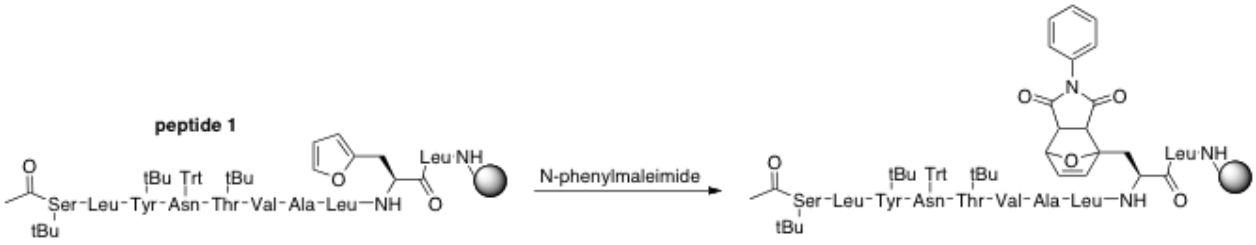

Figure 14: Diels-Alder reaction in toluene at 70°C.

N-Phenylmaleimide (7.7 mg, 44 µmole, 20 equiv) is dissolved in a minimal volume of toluene and added to the resin containing protected peptide **1** (5.3 mg, 2.2 µmole). The solution is heated for 24 hours at 70°C, after which the resin is washed with 3x DCM (3x30 s), 3x DMF (3x30 s), 3x MeOH (3x30 s), 3x DCM (3x30 s) and dried *in vacuo*.

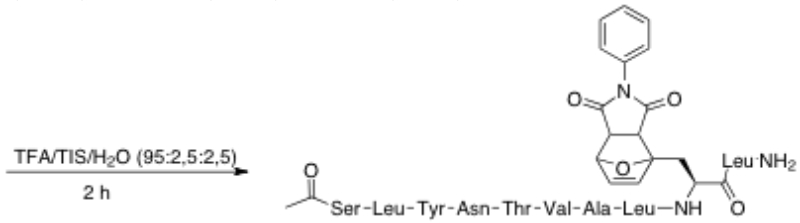

Figure 15: Deprotection of peptide after Diels-Alder reaction in toluene at 70°C.

A small sample of the peptide is cleaved off and analysed by LC-MS as mentioned in the protocol for deprotection of peptides (vide supra).

Exact Mass: 1230.59

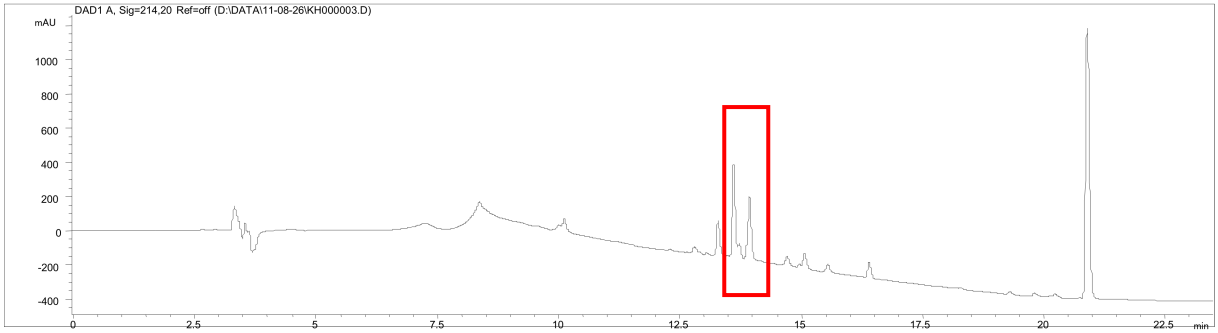

Figure 16: Crude HPLC chromatogram showing the starting product (t_R= 13.2 min), Diels-Alder products indicated in red frame and trityl protecting group (t_R=20.9min)

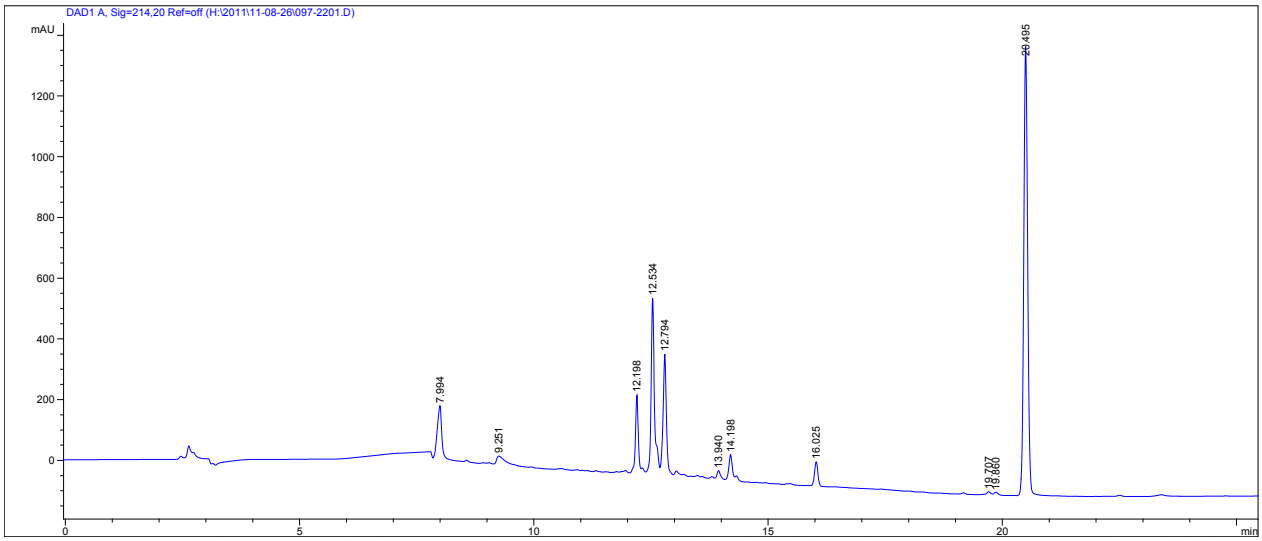

Figure 17: LC-MS chromatogram showing unreacted product (t_R= 12.198 min), Diels-Alder products (12.534 min; 12.794 min) and trityl protecting group (t_R= 20.495 min)

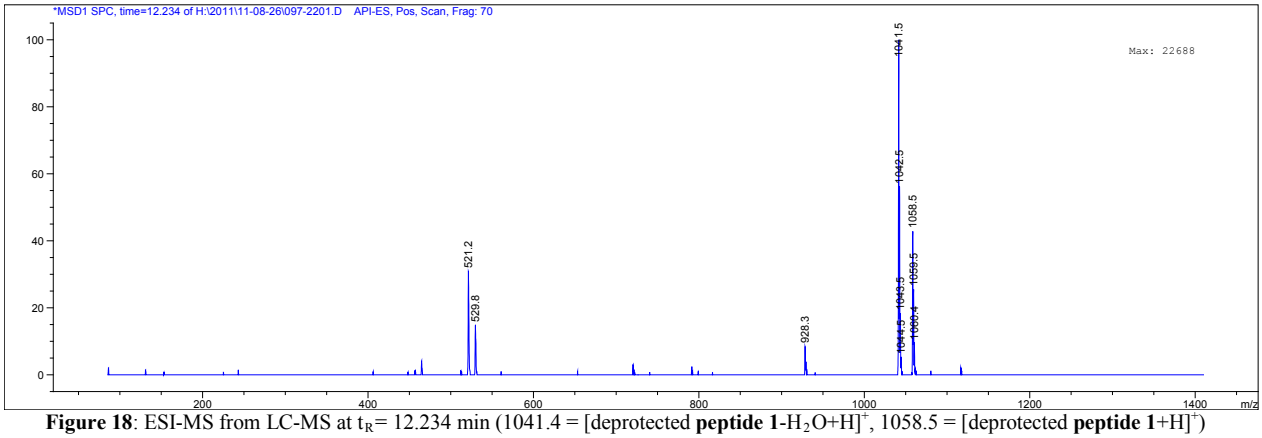

Figure 18: ESI-MS from LC-MS at t_R= 12.234 min (1041.4 = [deprotected **peptide 1**-H_2O+H]$^+$, 1058.5 = [deprotected **peptide 1**+H]$^+$)

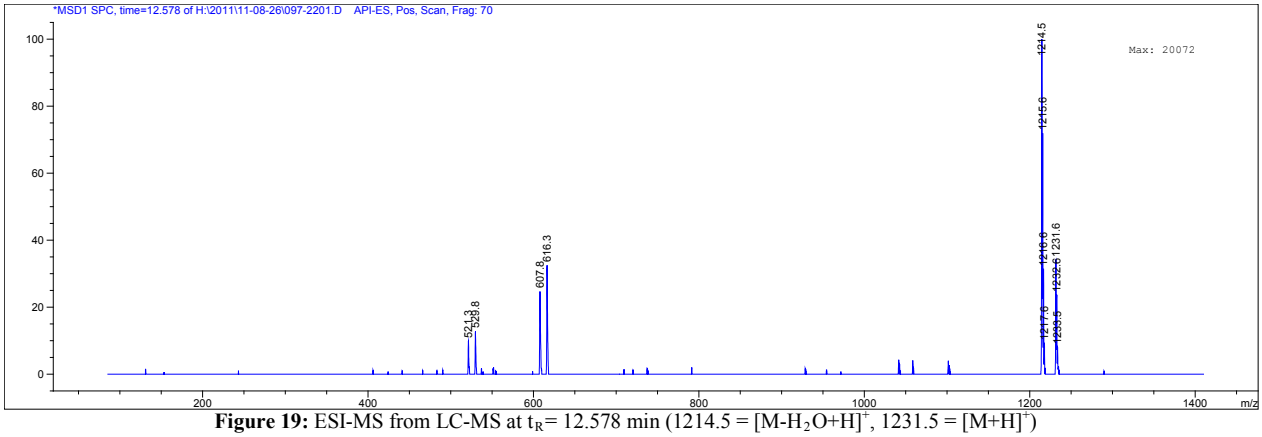

Figure 19: ESI-MS from LC-MS at t_R= 12.578 min (1214.5 = [M-H_2O+H]$^+$, 1231.5 = [M+H]$^+$)

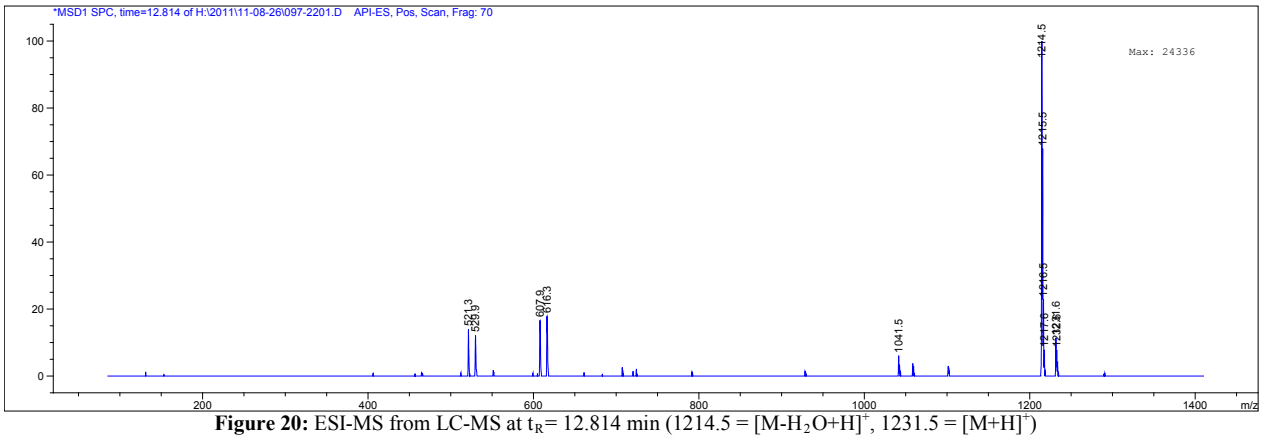

Figure 20: ESI-MS from LC-MS at t_R= 12.814 min (1214.5 = $[M\text{-}H_2O+H]^+$, 1231.5 = $[M+H]^+$)

6 Retro-Diels-Alder reaction from Diels-Alder product peptide 3

100 µl toluene is added to resin containing the protected peptide **3**. The solution is heated for 24 hours at 70°C, after which the resin is washed with 3x DCM (3x30 s), 3x DMF (3x30 s), 3x MeOH (3x30 s), 3x DCM (3x30 s) and dried *in vacuo*.

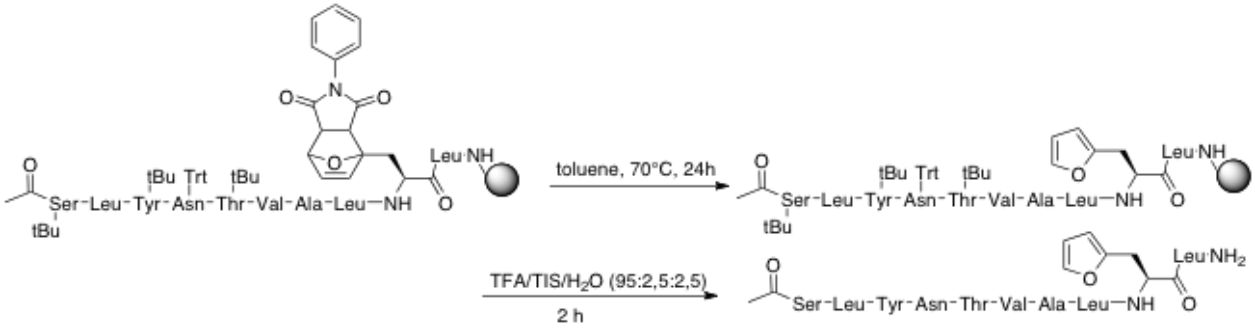

Figure 21: Retro-Diels-Alder and subsequent deprotection by TFA

A small sample of the peptide is cleaved off and analysed by HPLC as mentioned in the protocol for deprotection of peptides (vide supra).

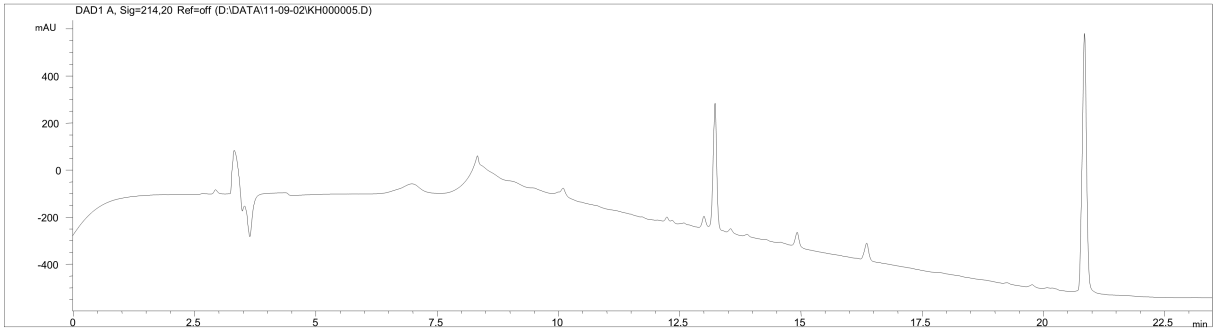

Figure 22: Crude HPLC chromatogram showing the retro-Diels-Alder product (t_R= 13,2 min) and trityl protecting group (t_R=20,9min)

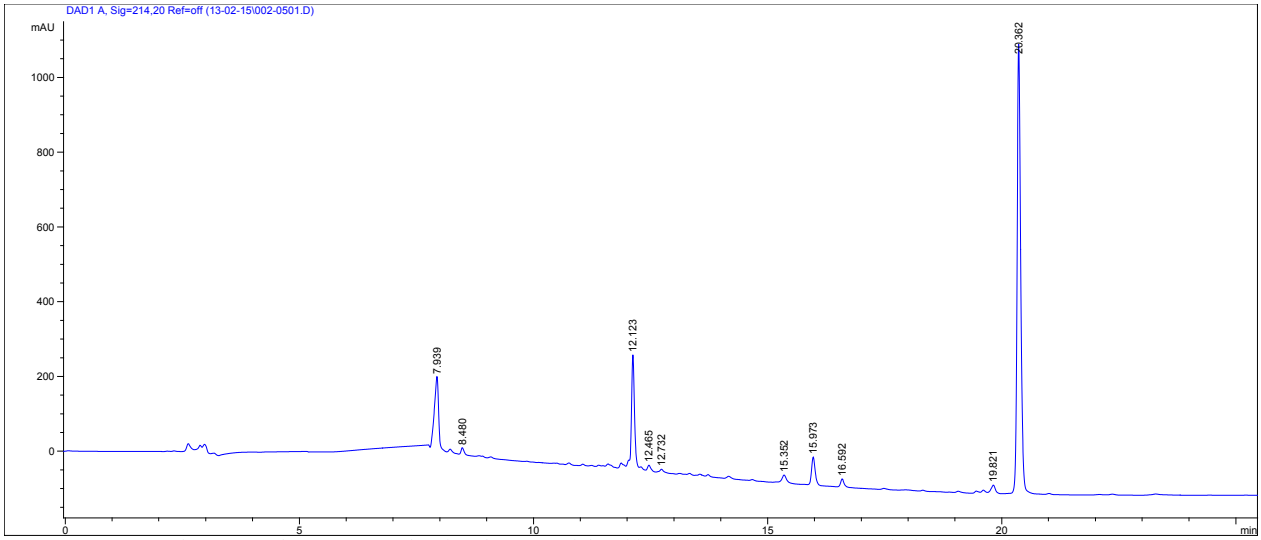

Figure 23: Crude LC-MS chromatogram showing the retro-Diels-Alder product (t_R= 12,1 min) and trityl protecting group (t_R=20,4min)

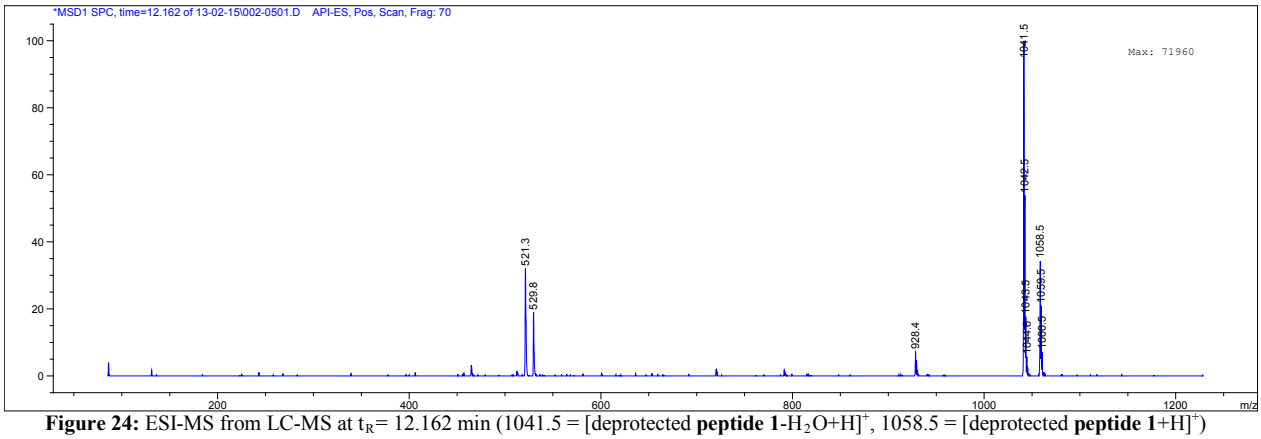

Figure 24: ESI-MS from LC-MS at t_R= 12.162 min (1041.5 = [deprotected **peptide 1**-H_2O+H]$^+$, 1058.5 = [deprotected **peptide 1**+H]$^+$)

7 Synthesis of peptide 4

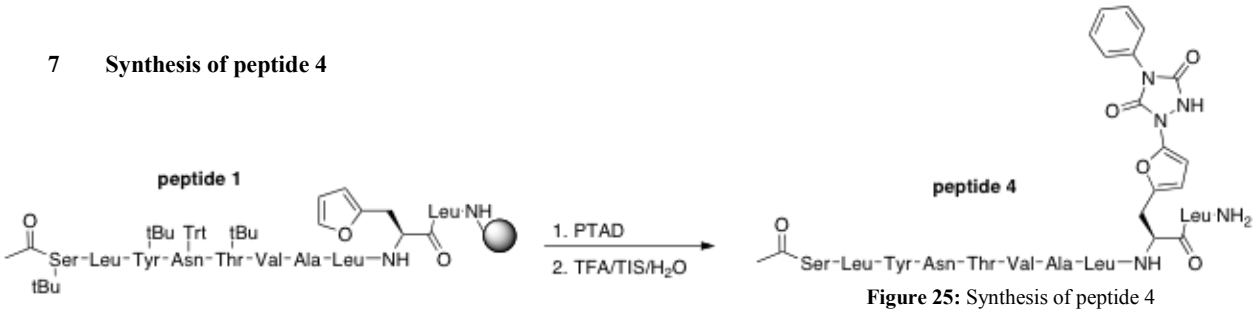

Figure 25: Synthesis of peptide 4

15 mg of peptidyl-resin **1** (5.7 µmole) is left to swell in 500µl of DCM, and filtered of after 20 minutes and washed 3 more times with DCM. Next, 3 mg PTAD (3 equiv, 17 µmole) is dissolved in 600 µl of DCM and added to 15 mg of peptidyl-resin **1** (5.7 µmole). The mixture is left for 15 minutes, filtered off, washed with 3x DCM (3x30 s), 3x DMF (3x30 s), 3x MeOH (3x30 s), 3x DCM (3x30 s) and dried *in vacuo*.
Completion of the reaction is confirmed by deprotecting and analysing a small sample (0.4mg) as mentioned in the protocol for deprotection of peptides (vide supra).
The resin is treated with 600 µl TFA/TIS/H_2O (95:2.5:2.5) for 2 hours. The liquid is filtered off and the resin is washed 3x with TFA, 2x with DCM. The filtrate is evaporated and dried *in vacuo*, yielding in 6.2mg of the crude product (4.7 µmole, yield 78%). LC-MS analysis revealed complete conversion to a product with the correct mass. After HPLC purification 1.5mg of peptide **4** (1.2 µmole) was obtained.

Exact Mass: 1232.58

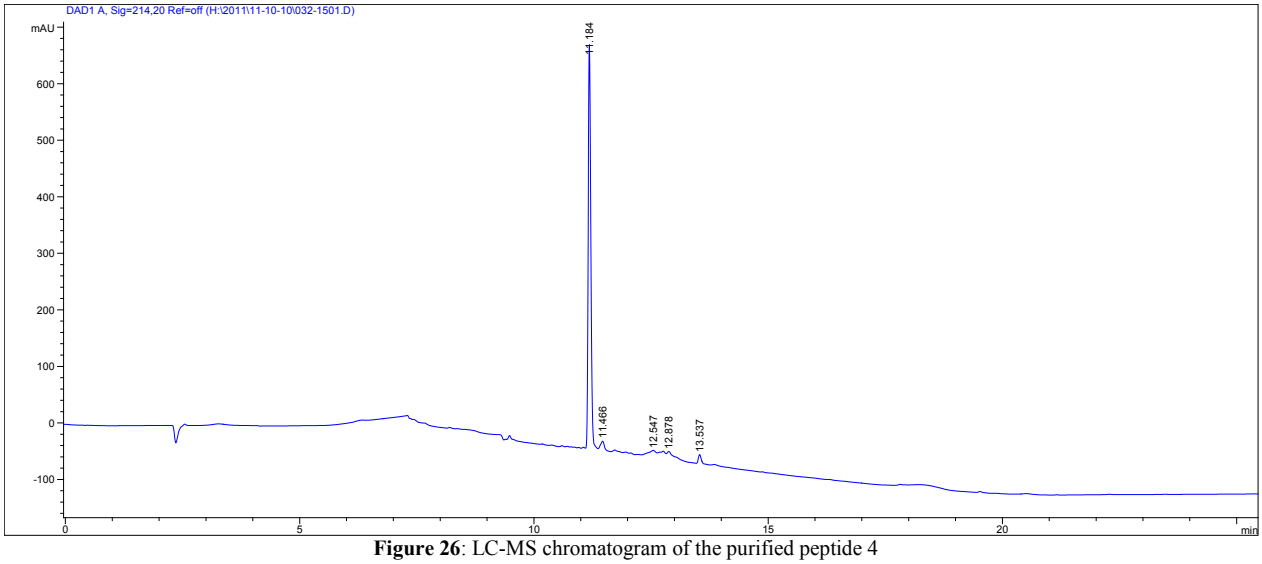

Figure 26: LC-MS chromatogram of the purified peptide 4

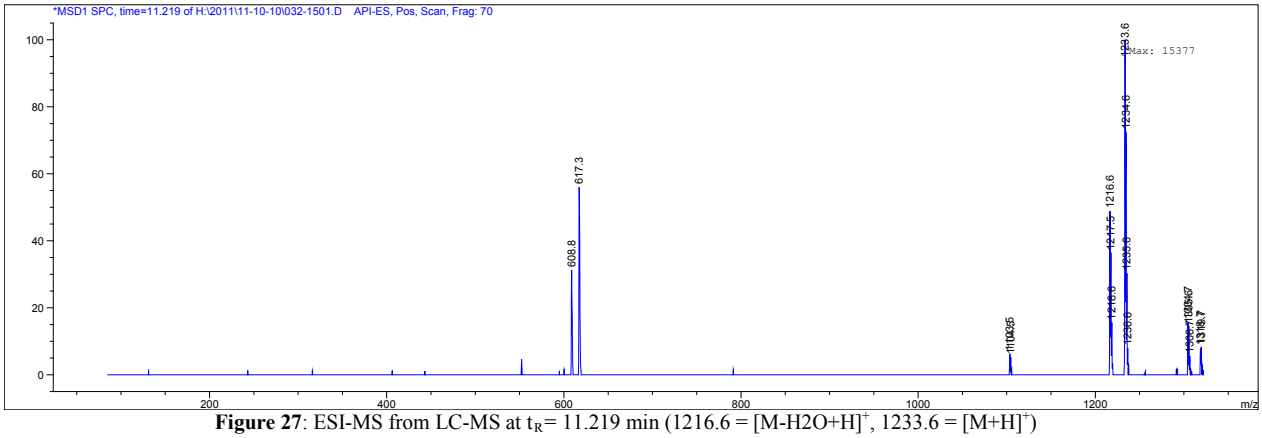

Figure 27: ESI-MS from LC-MS at t_R= 11.219 min (1216.6 = $[M-H2O+H]^+$, 1233.6 = $[M+H]^+$)

8 Electrophilic aromatic substitution of PTAD to acetylated furyl-alanine

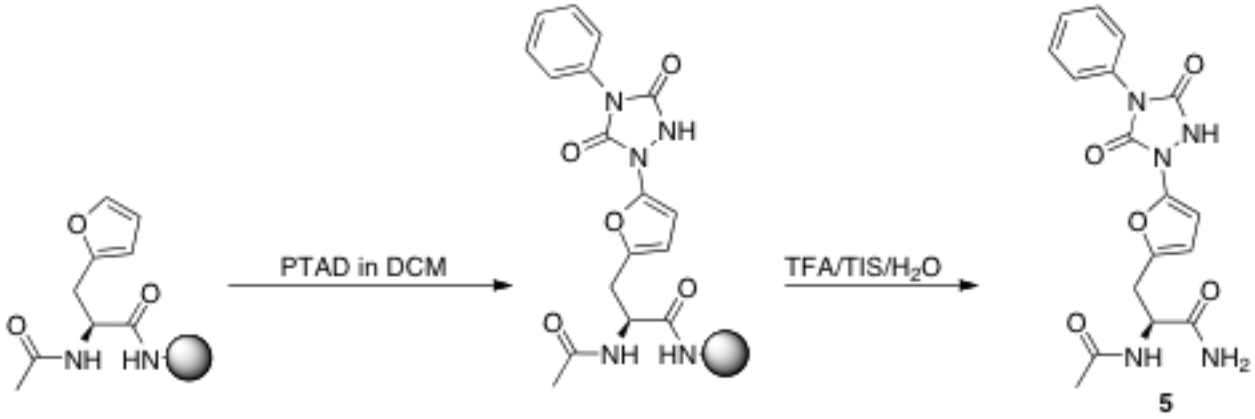

Figure 28: Cycloaddition of PTAD and subsequent cleavage of the resin

50 mg of resin (0.68 mmol/g, 34 µmol) was left to swell in 1 ml of DCM, filtered of after 20 minutes and washed 3 more times with DCM. Next, 17.8 mg PTAD (3 equiv, 102 µmole) is dissolved in 3 ml of DCM, and added to the resin. The mixture is left for 15 minutes, filtered off, washed with 3x DCM (3x30 s), 3x DMF (3x30 s), 3x MeOH (3x30 s), 3x DCM (3x30 s) and dried *in vacuo*.
The product is cleaved of the resin with 1 ml TFA/TIS/H_2O (95:2.5:2.5) for 1 hour. The resin is washed with 2x TFA, 2x DCM, the filtrate is evaporated, dissolved in MeOH, transferred to an eppendorf tube and dried in vacuo. 12.5 mg of the crude product was obtained. The reaction was repeated, and HPLC purified fractions were combined and dried by lyophilization to a white powder, yielding 2.1 mg of purified product **5**.
NMR analysis is described elsewhere in this document.

Exact Mass: 371.12

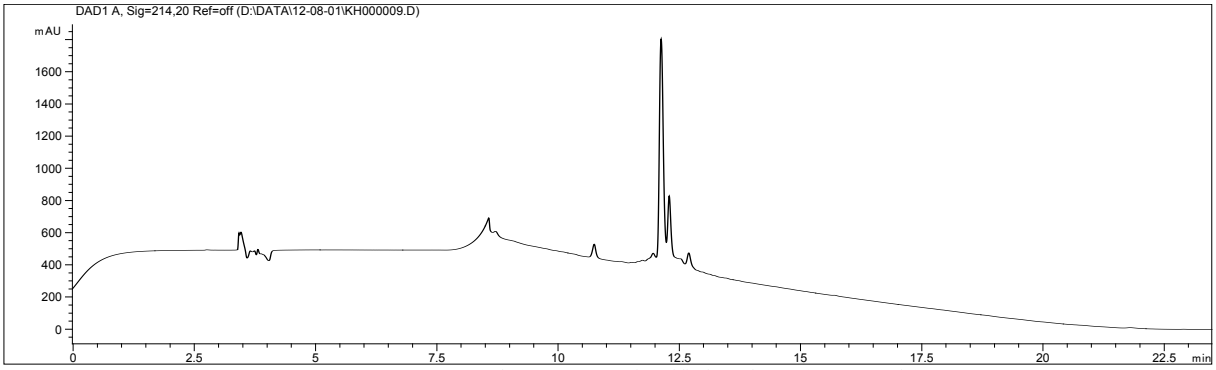

Figure 29: HPLC chromatogram of purified product 5 (t_R=12.1 min)

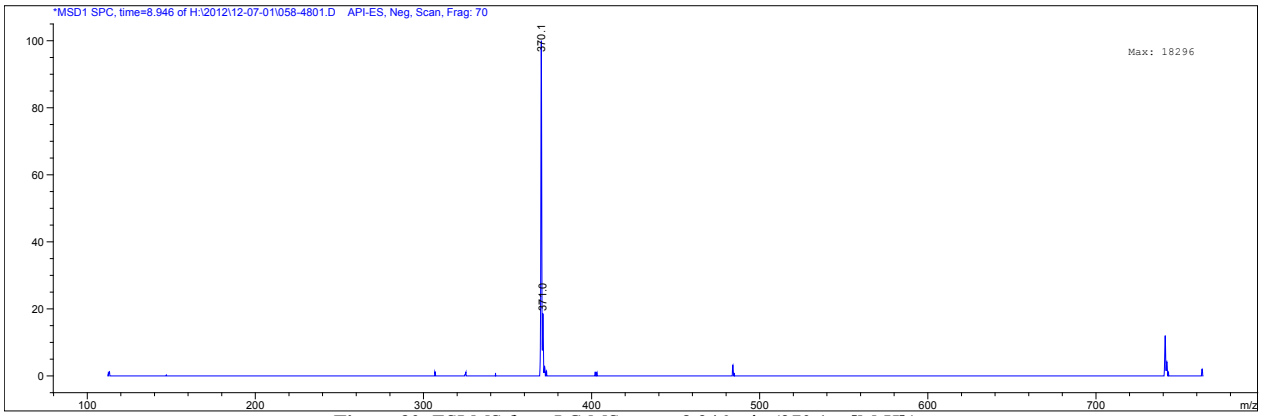

Figure 30: ESI-MS from LC-MS at t_R= 8.946 min (370.1 = $[M-H]^-$)

9 Synthesis of FRET-probe

9.1 introduction of DMEQ-TAD

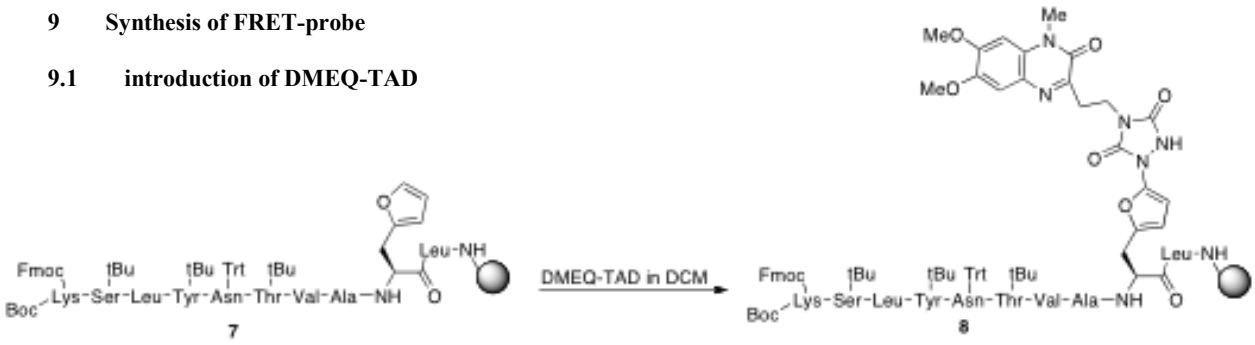

Figure 31: Cycloaddition of DMEQ-TAD

10 mg of resin (0.352 mmol/g, 3.52 µmol) was left to swell in 500 µl of DCM, filtered of after 20 minutes and washed 3 more times with DCM. Next, 3.65 mg DMEQ-TAD (3 equiv, 10.56 µmole) is dissolved in 600 µl of DCM and added to the resin. The mixture is left for 15 minutes in absence of light, filtered off, washed with 3x DCM (3x30 s), 3x DMF (3x30 s), 3x MeOH (3x30 s), 3x DCM (3x30 s) and dried *in vacuo*.
Completion of the reaction is confirmed by deprotecting and analysing a small sample (0.4mg) as mentioned in the protocol for deprotection of peptides (vide supra).

Exact Mass: 1710.80

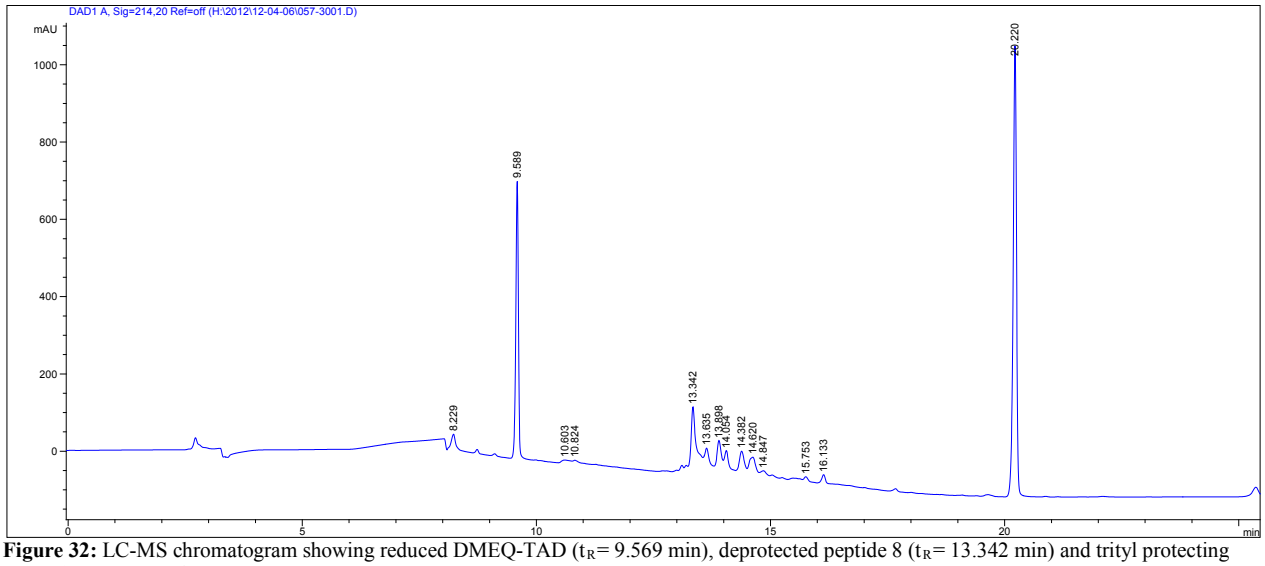

Figure 32: LC-MS chromatogram showing reduced DMEQ-TAD (t_R= 9.569 min), deprotected peptide 8 (t_R= 13.342 min) and trityl protecting group (t_R= 20.220 min)

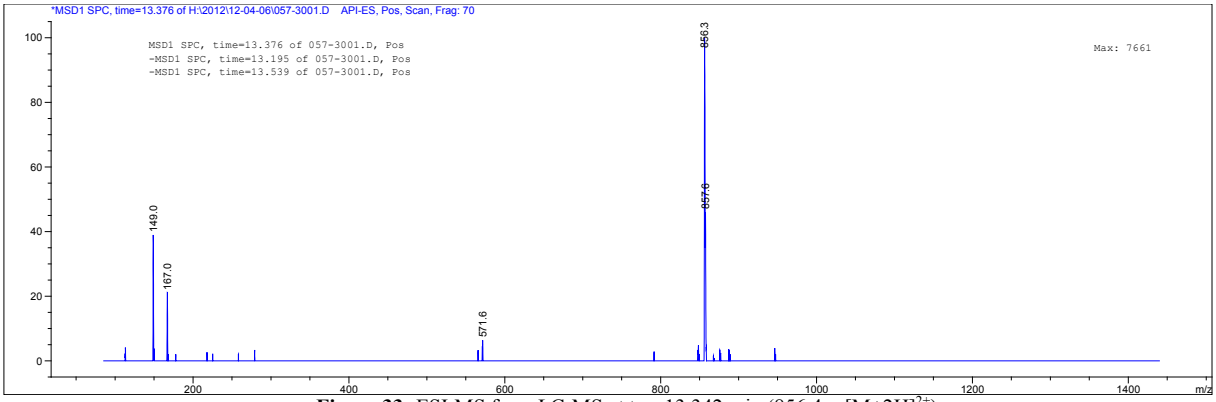

Figure 33: ESI-MS from LC-MS at t_R= 13.342 min (856.4 = $[M+2H]^{2+}$)

9.2 Fmoc deprotection peptide 8

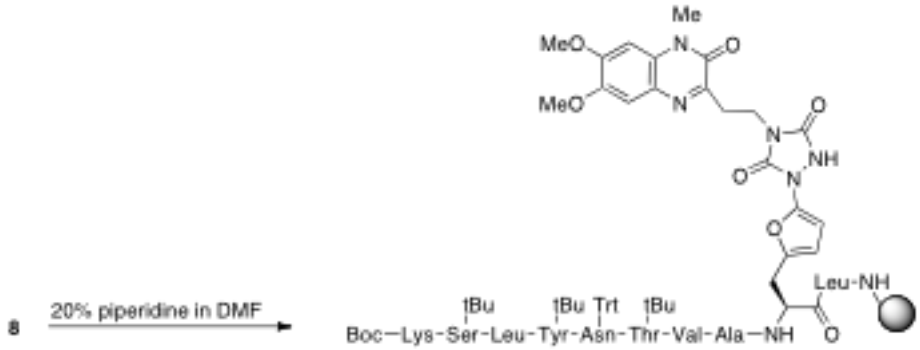

Figure 34: Fmoc deprotection peptide 8

The resin is left to swell in 200 µl DMF, and filtered off after 20 minutes. Next, 150 µl of a 20% piperidine in DMF solution is added and left to react for 2 minutes, and again filtered off. This step is repeated for 5 and 15 minutes.
The resin is washed with 3x DMF (3x30 s), 3x MeOH (3x30 s), 3x DCM (3x30 s) and dried *in vacuo*.
Completion of the reaction is confirmed by deprotecting and analysing a small sample (0.3mg) as mentioned in the protocol for deprotection of peptides (vide supra).

Exact Mass: 1488.74

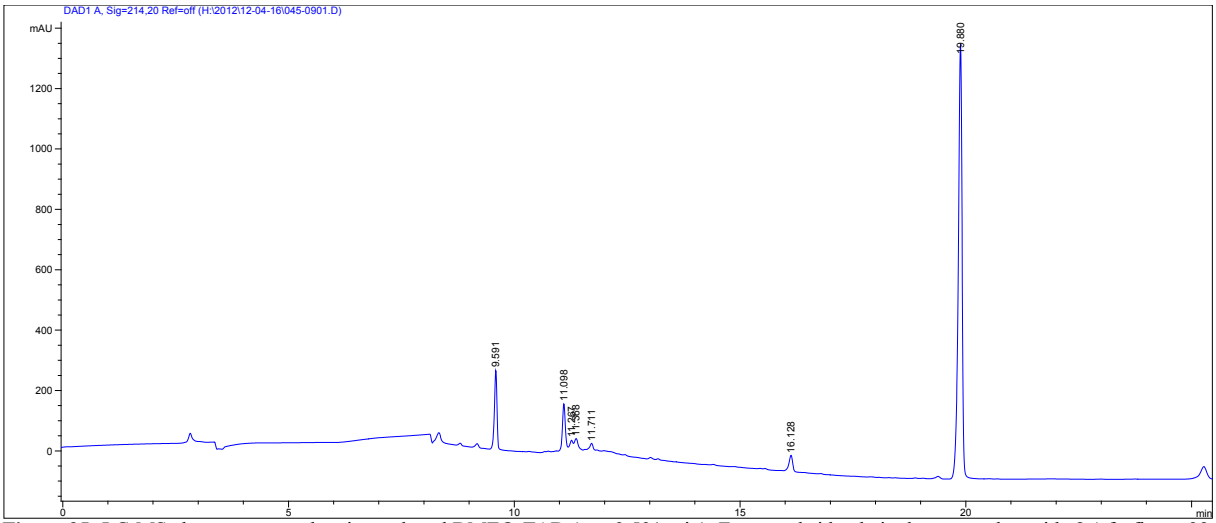

Figure 35: LC-MS chromatogram showing reduced DMEQ-TAD (t_R= 9.591 min), Fmoc and side-chain deprotected peptide 8 (cfr. figure 32; t_R= 11.096 min) and trityl protecting group (t_R= 19.880 min)

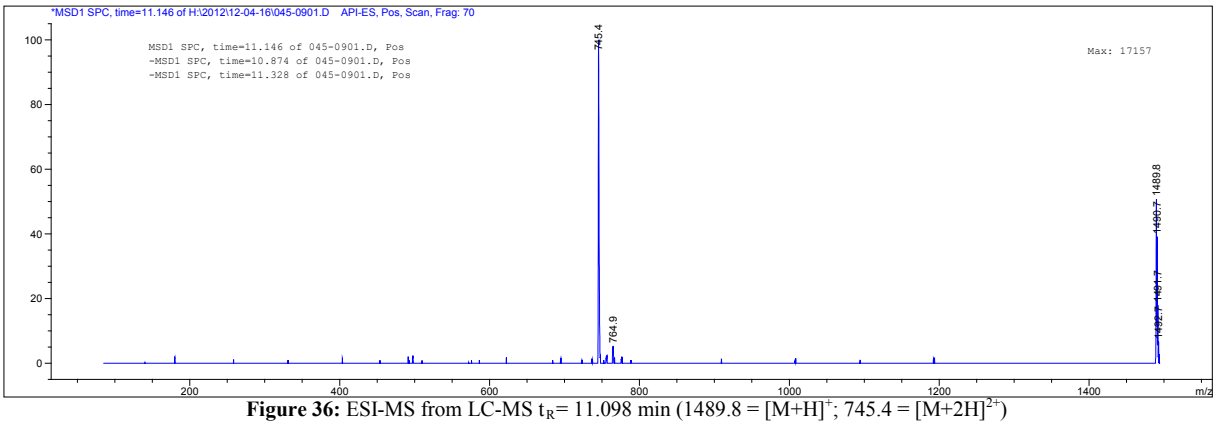

Figure 36: ESI-MS from LC-MS t_R= 11.098 min (1489.8 = $[M+H]^+$; 745.4 = $[M+2H]^{2+}$)

9.3 Dabsyl labeling

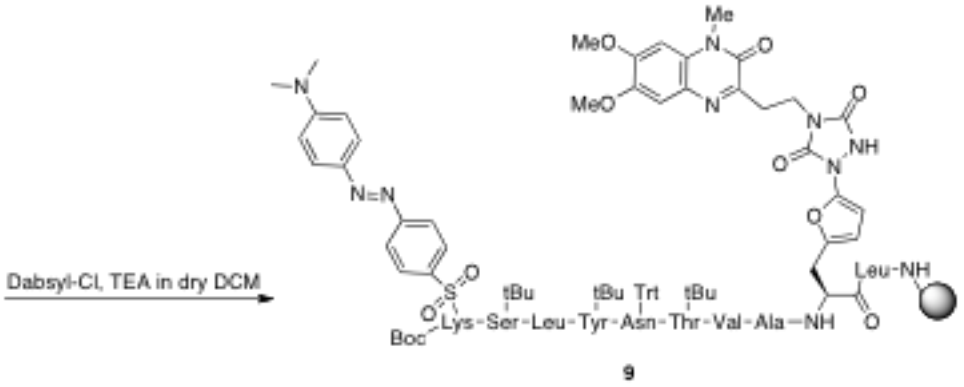

Figure 37: Dabsyl labeling at the free amine of lysine side chain

The resin (9 mg, 3.17 µmol) is left to swell in 200 µl dry DCM, and filtered off after 20 minutes. Next, Dabsyl-Cl (1.5 mg, 4.75 µmol) and TEA (0.58 µl, 4.1 µmol) in 135 µl dry DCM is added to the resin.
The resin is washed with 3x DCM (3x30 s), 3x DMF (3x30 s), 3x MeOH (3x30 s), 3x DCM (3x30 s) and dried *in vacuo*.

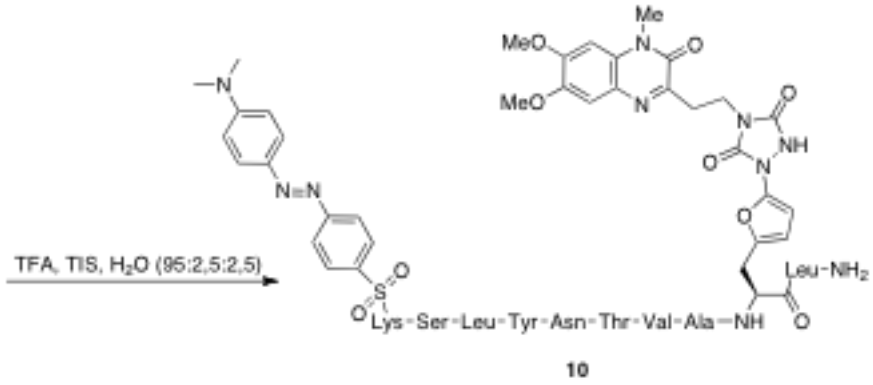

Figure 38: Deprotection of peptide 9 to peptide 10

Peptide **9** (6.5 mg) is treated with 1 ml of a TFA/TIS/H_2O solution and shaken for 1 hour. The resin is filtered of and washed twice with TFA, an twice with DCM. The filtrate is evaporated and the obtained peptide is dried *in vacuo*. The resin is treated again for 1 hour with TFA, washed and filtered off. The filtrate is combined with the first fraction, evaporated and again dried *in vacuo*. The resulting solid is redissolved in 600µl acetonitrile/H_2O (1:1) and purified by HPLC. The fraction containing peptide **10** was lyophilized, yielding 0.5 mg (281 nmol, 14% yield).

Exact Mass: 1775.81

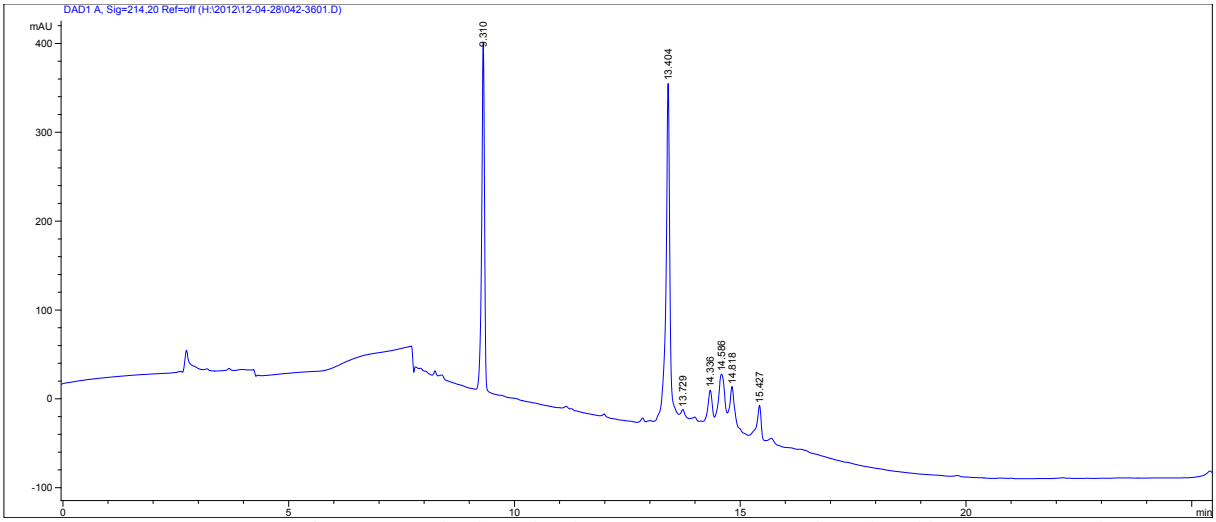

Figure 39: LC-MS chromatogram showing reduced DMEQ-TAD (t_R= 9.310 min) and peptide 10 (t_R= 13.404 min)

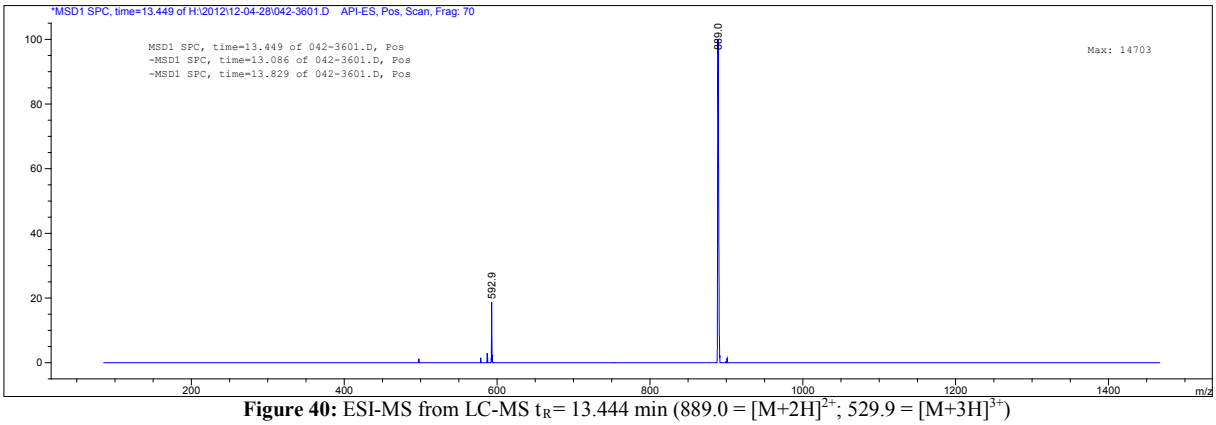

Figure 40: ESI-MS from LC-MS t_R= 13.444 min (889.0 = $[M+2H]^{2+}$; 529.9 = $[M+3H]^{3+}$)

9.4 FRET assay

25 µg of peptide **10** dissolved in 50 µl acetonitrile/water (1:1) is added to 2ml of a 0.02 M AmBiC buffer in a cuvette for fluorescence spectroscopy. Next, 2 µl of alpha-chymotrypsin (2.5 µg; 1.25 mg/ml) is added to initiate the enzymatic reaction. A second cuvette is prepared without enzyme as a negative control. Fluorescence is monitored for both samples simultaneously for 1 hour at 37°C, on a Varian Cary Eclipse Fluorescence Spectrophotometer.

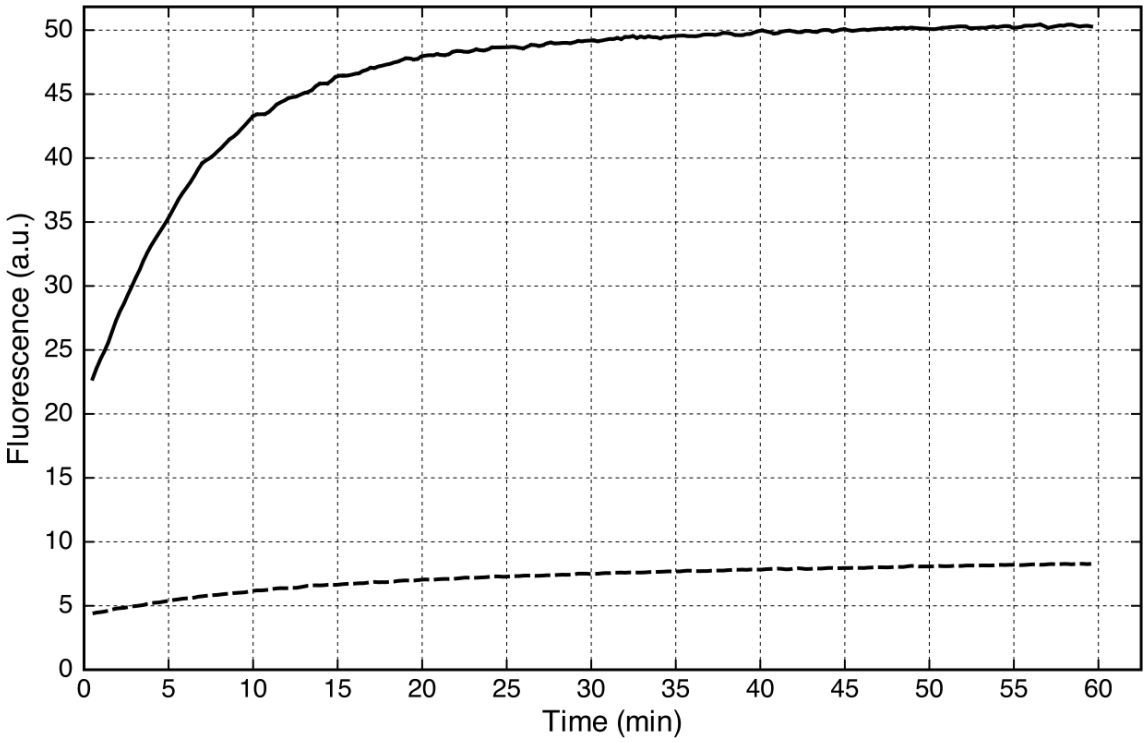

Figure 41: FRET-probe with alpha-chymotrypsin (solid) and control without enzyme (dashed).

10 Structure analysis of PTAD modified furylalanine

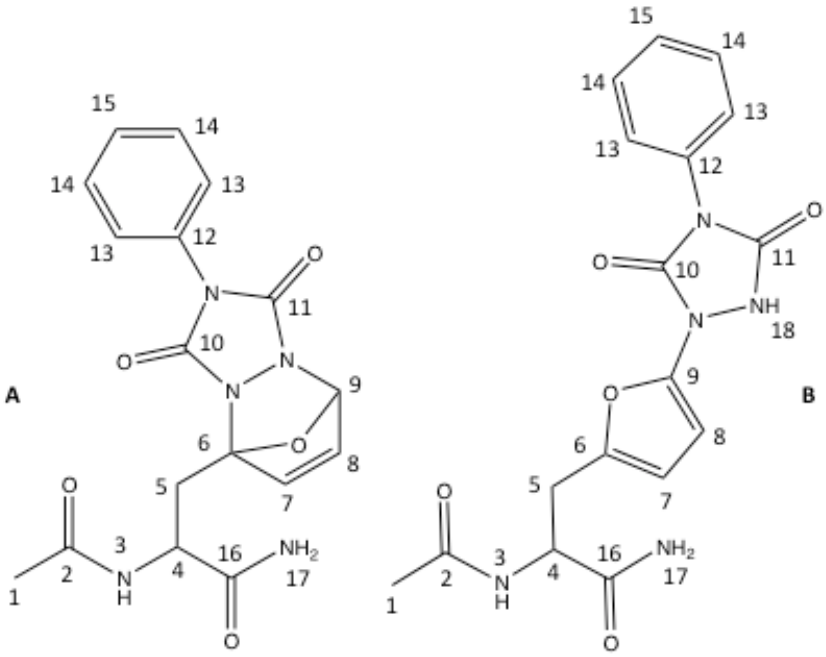

Figure 42: Two possible structures for the modified amino acid.

Since this modified building block is structurally quite diverse, the analysis of the spectra is quite straightforward. In the first section, the general assignment of the amino acid backbone is briefly discussed, followed by the elucidation of the structure of the modified side chain.

10.1 Information derived from 1D ^{1}H NMR

Considerable information can be obtained by simply analysing the 1D ^{1}H spectrum (figure 43, signals at 8.03; 2.92 and 2.74 correspond to DMF):

- In total all 17 different types of protons can be assigned to the respective NMR signals
- A broad signal can be seen around 11.5 ppm which mostly corresponds to exchangeable protons like phenolic OH-, carboxylic acid, amide or indole NH-like protons (the broad hump at 3.56ppm corresponds to some water still present in the sample.
- Some signals are located in regions that are typical for amino acid like structures: in the aliphatic region, a singlet corresponding to 3 protons at 1.92 ppm, two multiplets at 2.97 and 3.20 ppm each corresponding with 1 proton. In the Cα-region, a multiplet corresponding to one proton at 4.69 ppm and finally, in the amide region a doublet, integrating for a single proton with a chemical shift of 8.12 ppm.
- The remaining signals are either located in the aromatic region (7-8ppm) or in the alkene region (6.0-7.6 ppm), which is as expected. Several other signals of low intensity can be seen in the spectrum; these correspond to impurities still present in the sample but do not interfere with the analysis of the modified building block.

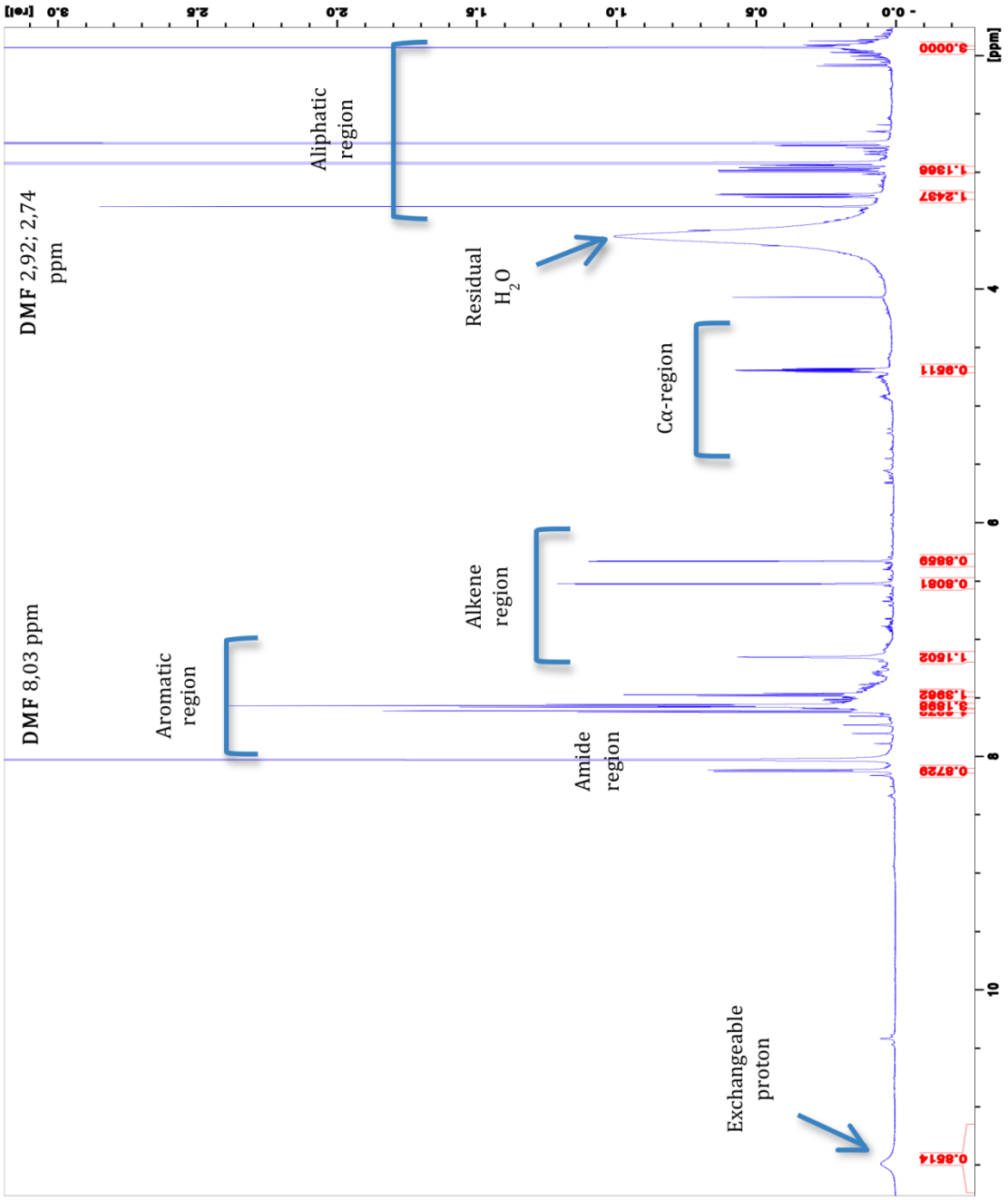

Figure 43: 1D ^{1}H spectrum (25°C, 700MHz)

10.2 Information derived from 2D NMR

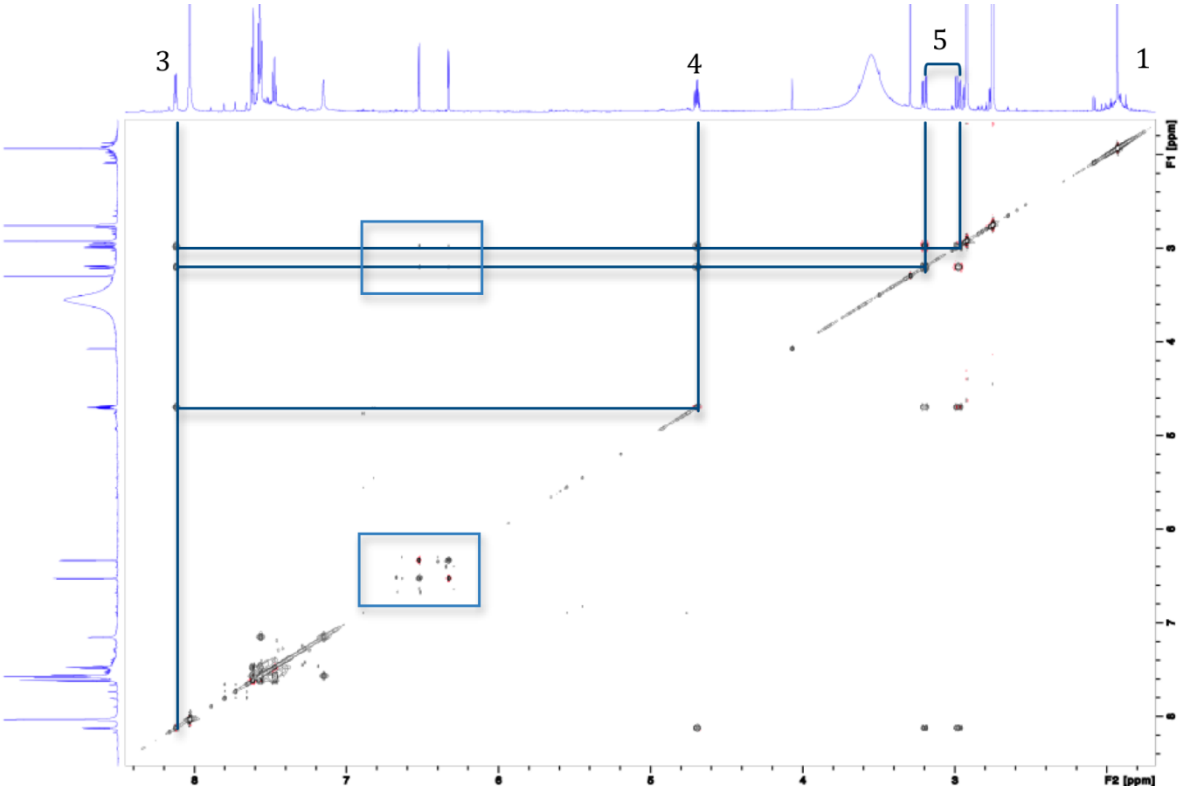

Figure 44: 2D TOCSY spectrum (25°C, 700MHz, 100ms mixing time)

From the TOCSY spectrum (figure 44), the singlet at 1.92 ppm integrating for three protons identifies it as CH_3 n°1 in the structure, since the carbonyl group next to it (n°2) is a quaternary carbon atom. This is also the expected chemical shift for such a $\underline{CH_3}$CONH-moiety. The two multiplets which each integrate for a single proton at 2.97 and 3.20 ppm respectively, show a clear correlation to the multiplet at 4.69 and doublet at 8.12 ppm, as does the multiplet at 4.69 ppm, indicating that each of these protons participate in a single, uninterrupted spin system with a pattern and chemical shifts (Hα (n°4), βCH_2 (n°5) and NH (n°3)) corresponding to the amino acid moiety of the compound.. Further proof is provided by the HSQC spectrum (fig.4). The carbon atom corresponding to CH_3 n°1 has a chemical shift of 22.64 ppm, typical for a methyl while the two multiplets at 2.97 and 3.20 ppm are located on the same CH_2 carbon atom C5 at 31.03ppm. H3 shows no correlation to a corresponding carbon atom in the HSQC-spectrum (figure 45), thus, together with the chemical shift, this confirms it to be a labile NH bound proton.

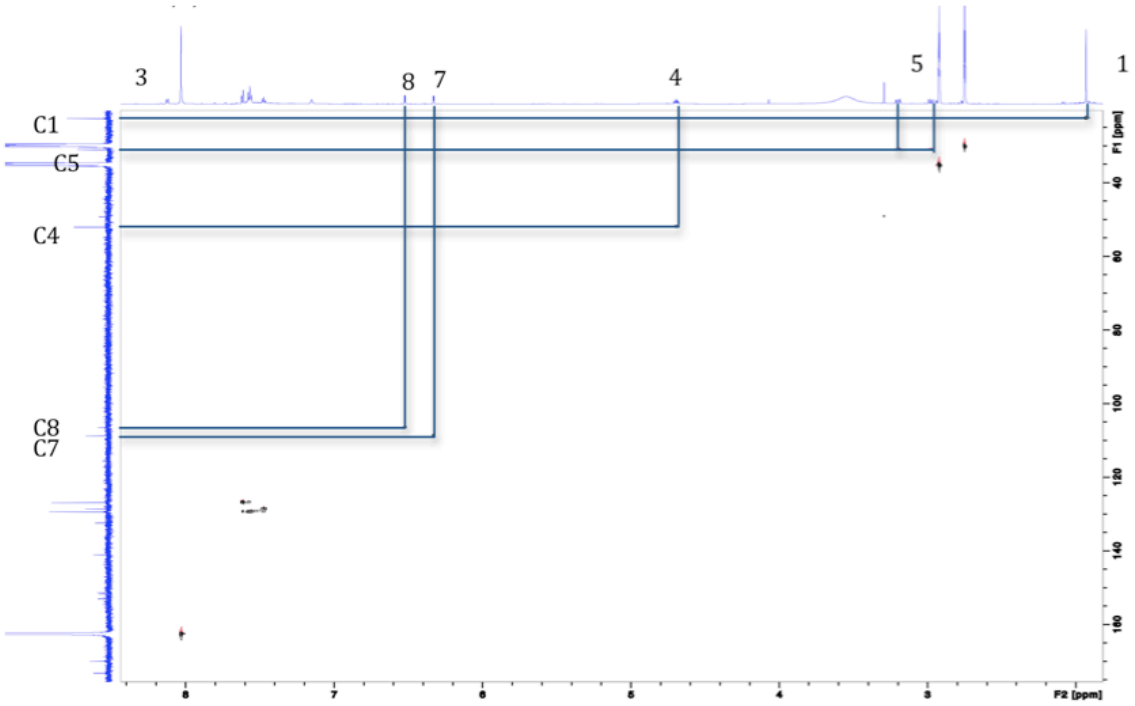

Figure 45: ^{1}H – ^{13}C HSQC spectrum (25°C, 700MHz)

The TOCSY pattern of the amino acid spin system originating from αH and the βCH_2 protons show weak long range correlations (unresolved in the 1D spectrum) to two doublets integrating for a single proton each at 6.33 and 6.52 ppm. They are mutually coupled and can be assigned to H7 and H8 in either structure A or structure B. Both protons form an otherwise isolated spin system since there are no further correlations to any other protons except for the ones already discussed. This fact, together with the presence of a labile proton (see also figure 42) already shows that the most likely structure is structure B and not A, since in the latter an additional scalar coupling with H9 is expected. A signal that could correspond with H9 in structure A in terms of its chemical shift is the somewhat broadened signal at 7.15 ppm. However, it is a singlet that has no directly attached carbon atom (figure 45) and shows and exchange peak in the NOESY spectrum (figure 46) with another proton at 7.57 ppm and with the residual water. This leads to the conclusion that both resonances originate from the primary amide group (H17a/b) expected in both structures.

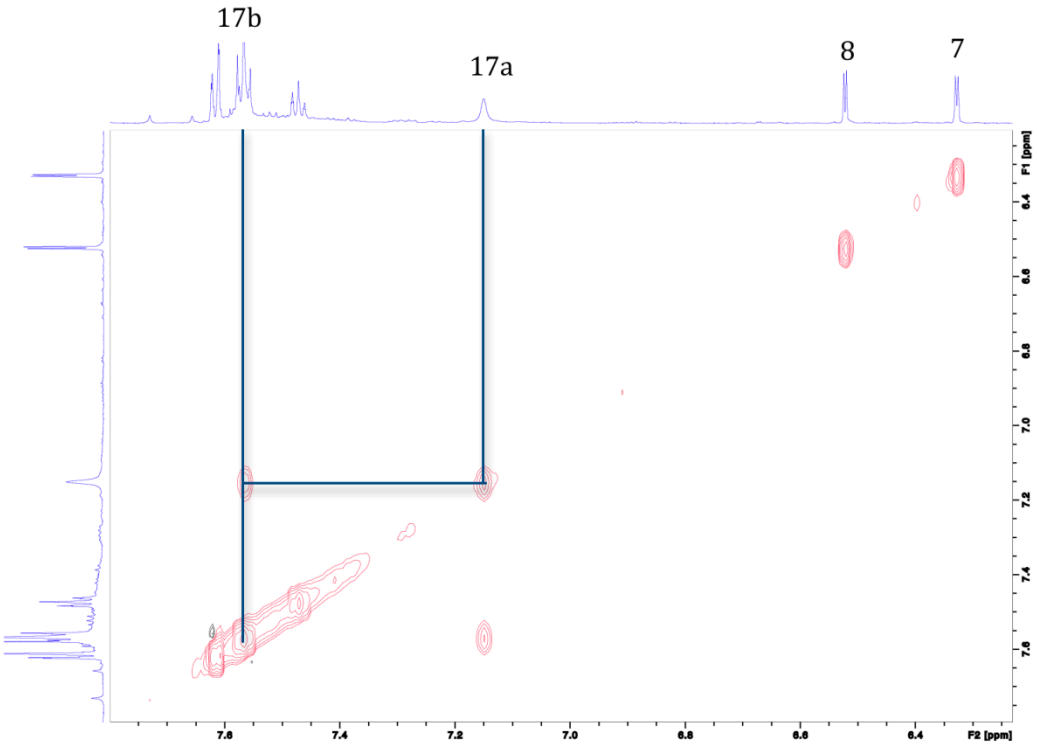

Figure 46: NOESY spectrum with the exchange peak (black) between the two amide protons (25°C, 700MHz)

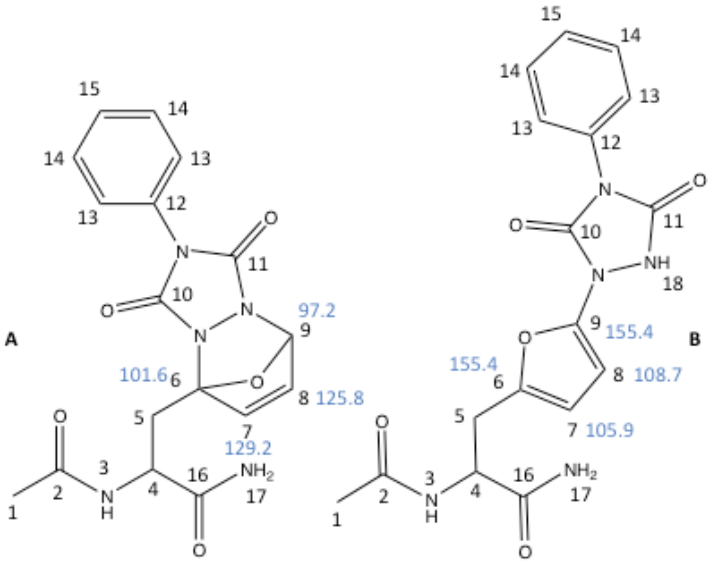

Figure 47: Predicted ^{13}C chemical shifts (obtained via 'ChemDraw Ultra v. 12')

A further argument for structure A comes from chemical shift prediction. In figure 47, the two possible structures for the amino acid are shown, together with their ^{13}C predicted chemical shifts. Note the blue coloured chemical shifts, indicating a prediction of sufficiently high confidence. The actual signals are found at 109.42 and 107.17 ppm, and are significantly different (~15-20 ppm) from those predicted for structure A, making it highly unlikely. In contrast, the small deviation with the predicted ^{13}C chemical shifts in structure B further supports the attribution of its structure. The protons associated with these carbons and thereby the

identity of both C–H units follows from the ROESY spectrum (figure 46), where the βCH_2 (n°5 at 2.97 and 3.20 ppm) and αH-proton (n°4 at 4.69 ppm) show nOe-contacts to H7 but not to H8.

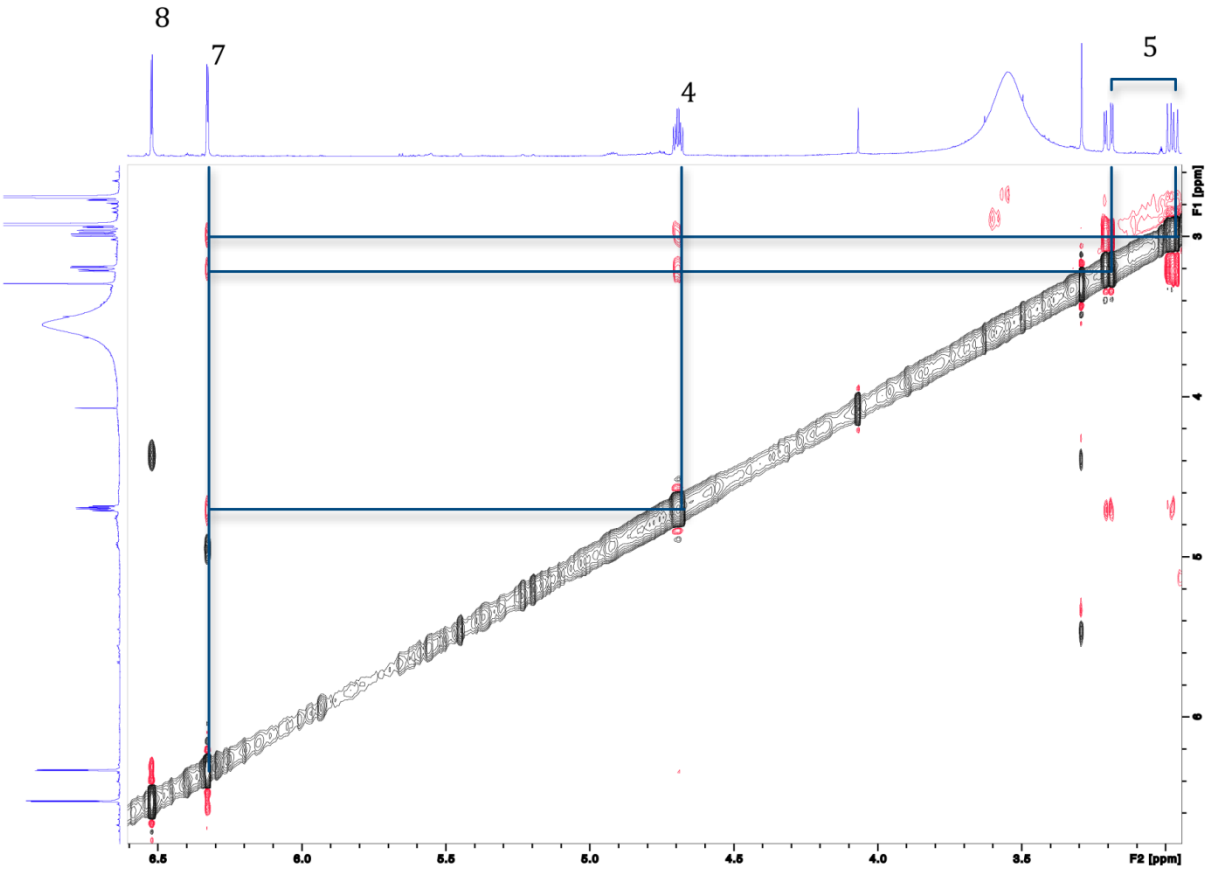

Figure 48: ^{1}H-^{1}H ROESY spectrum (25°C, 700MHz, 300ms mixing time)

Starting from H7 and H8, it can be expected that in the HMBC spectrum there are couplings to be seen which allow to identifying C6 and C9, where a distinction can be made between the two since the αH and βCH_2 protons are expected to show a correlation to C6 ($^{3}J_{CH}$ and $^{2}J_{CH}$ couplings) but not in the case of C9 ($^{6}J_{CH}$ and $^{5}J_{CH}$ couplings) because here no couplings are expected. As can be seen in figure 49 it is clear from the pattern of correlations that these couplings are indeed observed and involve quaternary carbons, in this way identifying C6 and C9.

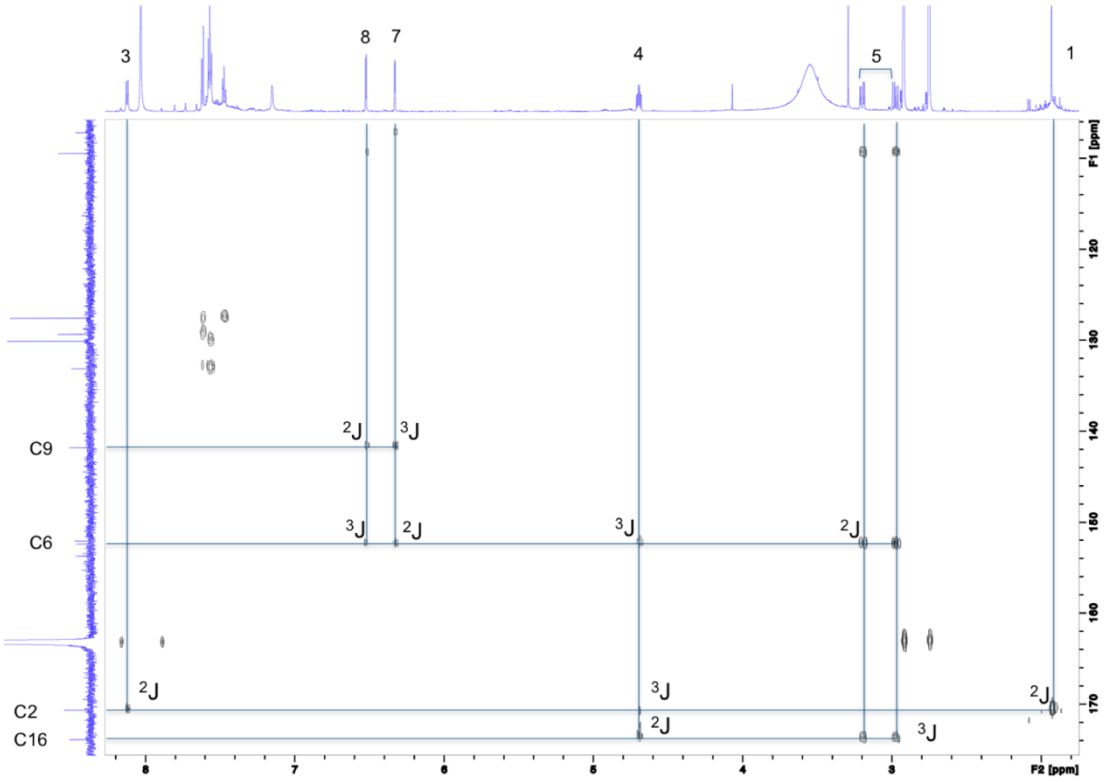

Figure 49: $^{1}H - ^{13}C$ HMBC spectrum (25°C, 700MHz, 8Hz long range coupling constant)

With the help of figure 49 two other quaternary amide type carbon atoms can be found: C2 and C16 at 170.65 and 174.11ppm respectively. C2 can be identified by a number of $^{n}J_{CH}$-couplings present in the HMBC spectrum: $^{2}J_{CH}$ coupling to amide proton 3, $^{3}J_{CH}$ coupling with Cα 4 and a $^{2}J_{CH}$ coupling with CH_3-group 1, while there are no couplings to be found to βCH_2 group n° 5. While this coupling is absent in the case of C2, it does show up with C16 as a ^{3}J-coupling, hence making the distinction between the two quaternary carbon atoms possible.

At this point only the ^{1}H and ^{13}C atoms from the aromatic ring and both quaternary carbon atoms C10 and C11 have to be assigned. The remaining aromatic proton signals can be found at 7.47, 7.57 and 7.62 ppm. From the fine structure and corresponding integral values alone, the signals can be assigned. H13 correspond to the doublet at 7.62ppm integrating for 2 protons. The apparent triplet like signal at 7.57 ppm which integrates for 3 protons is the result of a superposition with one of the primary amide signals as can be seen in figure 44. The 2 remaining protons at this position correspond to H14. Finally the other triplet at 7.47 ppm integrates for 1 proton and thus corresponds with H15. Each of the corresponding carbon atoms can be found via the HSQC. The final quaternary aromatic carbon C12 can be identified using the HMBC spectrum: it shows a ^{2}J and ^{3}J correlation respectively to H13 and H14 respectively.

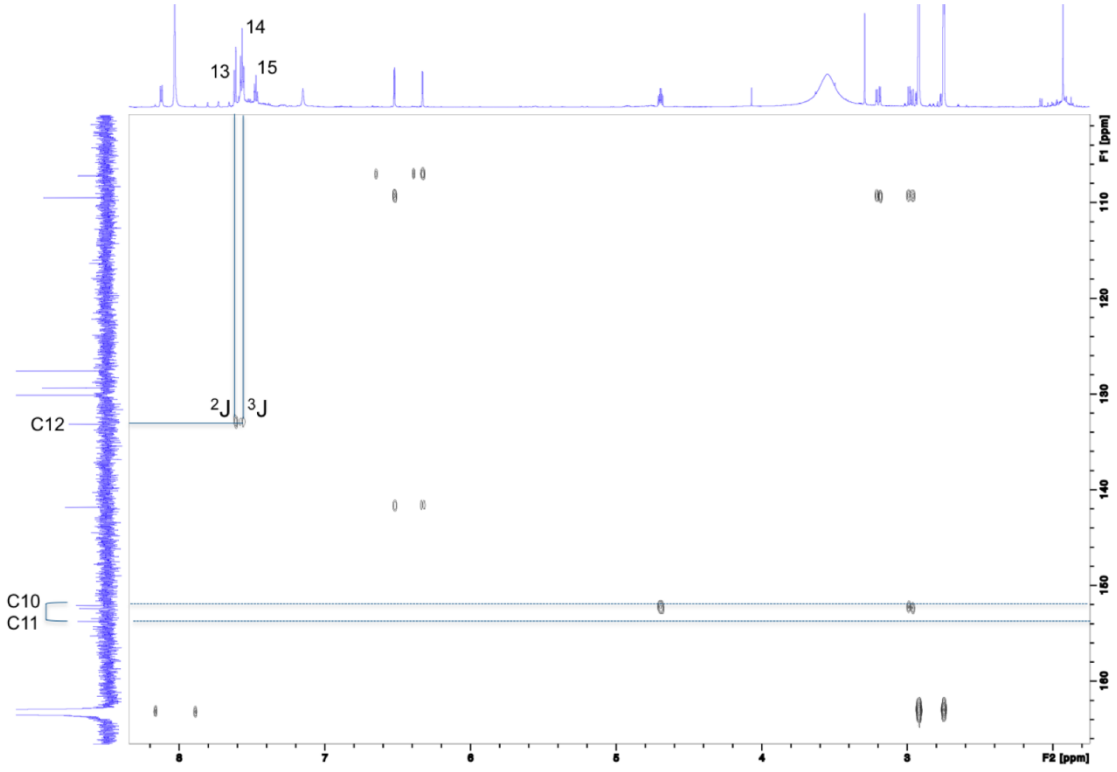

Figure 50: $^{1}H - ^{13}C$ HMBC spectrum (25°C, 700MHz, 4Hz long range coupling constant)

Last but not least, there are still two quaternary carbon atoms at 153.74 and 152.05 ppm which are unassigned in the spectrum. These show no correlations to any of the protons in the HSQC and HMBC (figure 48 dashed lines), meaning that these isolated quaternary carbon atoms. Again, chemical shift prediction places these carbons around 153 ppm, which is almost identical to the experimental chemical shift values.

From the analysis it is clear that the structure of the modified building block is structure B and not A, and the ^{1}H and ^{13}C chemical shifts are collected in the tables hereafter. Final confirmation comes from the analysis of H18, which is a labile proton unique for structure B. This resonance is by default assigned to the only remaining labile hydrogen, which appears at 11.61 ppm. Because of chemical exchange with residual water, this signal is too broad to generate correlations to neighbouring ^{1}H's (ROESY) or ^{13}C's (HSQC, HMBC). Therefore, additional measurements were done at low temperature to slow down the chemical exchange thus resuscitating these missing correlations as detailed in 10.5

10.3 ^{1}H Chemical shifts

Finestructure	Chemical shift δ (ppm)	# protons	Annotation
singlet	1.93	3	1
quintet	2.74	-	DMF
quintet	2.92	-	DMF
doublet of doublets	2.97	1	5
doublet of doublets	3.20	1	5
broad hump	3.55	-	Water
multiplet	4.69	1	4
doublet	6.33	1	7
doublet	6.52	1	8
broad singlet	7.15	1	17a/b
triplet	7.47	1	15
triplet	7.57	3	14; 17a/b
doublet	7.62	2	13
singlet	8.03	-	DMF
doublet	8.12	1	3
broad singlet	11.61	1	18
	Total	17	

Table 1: Overview of ^{1}H assignment modified furyl-alanine in DMF-d7

10.4 ^{13}C Chemical shifts

Atom type	Chemical shift δ (ppm)	Annotation	# carbon atoms
CH_3	23.23	1	1
Solvent	30.54	DMF	-
CH_2	31.72	5	1
Solvent	35.67	DMF	-
CH	52.77	4	1
CH	107.18	8	1
CH	109.43	7	1
CH(aromatic)	127.54	13	2
CH(aromatic)	129.35	15	1
CH(aromatic)	130.10	14	2
Cq(aromatic)	133.14	12	1
Cq	141.79	9	1
Cq	152.05	10/11	1
Cq	152.39	6	1
Cq	153.74	10/11	1
Solvent	163.15	DMF	-
Cq	170.65	2	1
Cq	174.11	15	1
		Total	17

Table 2: Overview of ^{13}C assignment of modified furyl-alanine in DMF-d7

10.5 Additional measurements at low temperature

The exchange broadening of H18, which prevents correlations to other spins to be detected was handled by cooling the sample to 223.1K (-50°C) after which a ROESY was measured with a mixing time of 150 ms at 500MHz.

Comparing the 1D spectrum to the one at room temperature in figure 51, the difference is quite clear: at -50°C the signal attributed to H18 has moved to 11.96 ppm and appears as sharp and therefore intense as the other non-exchangeable signals. It also integrates for one proton. Under these conditions, H18 is seen to generate a small but clear correlation to H8, which indeed is the only proton that is close in space according to structure B, finally confirming the identity of the lowest field resonance as H18.

With this additional information, the actual structure of the PTAD adduct of the furylalanine has definitely been identified as B.

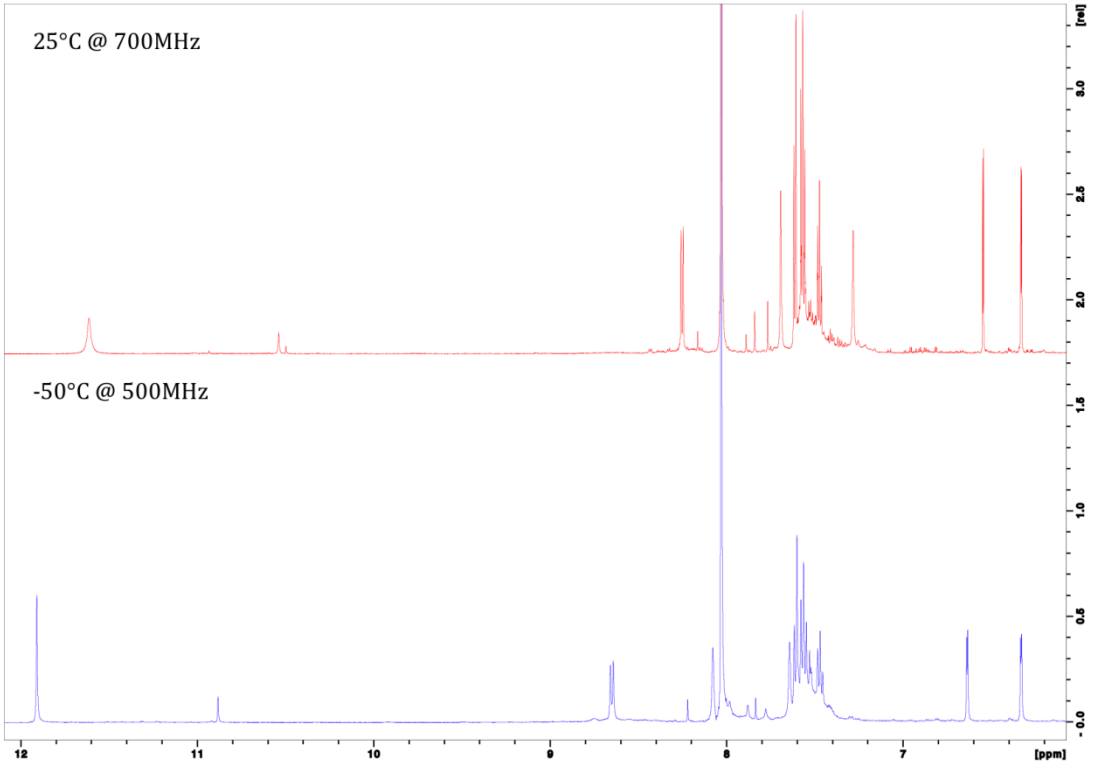

Figure 51: Comparison of two 1D 1H spectra at different temperatures (25°C, 700MHz; -50°C, 500MHz)

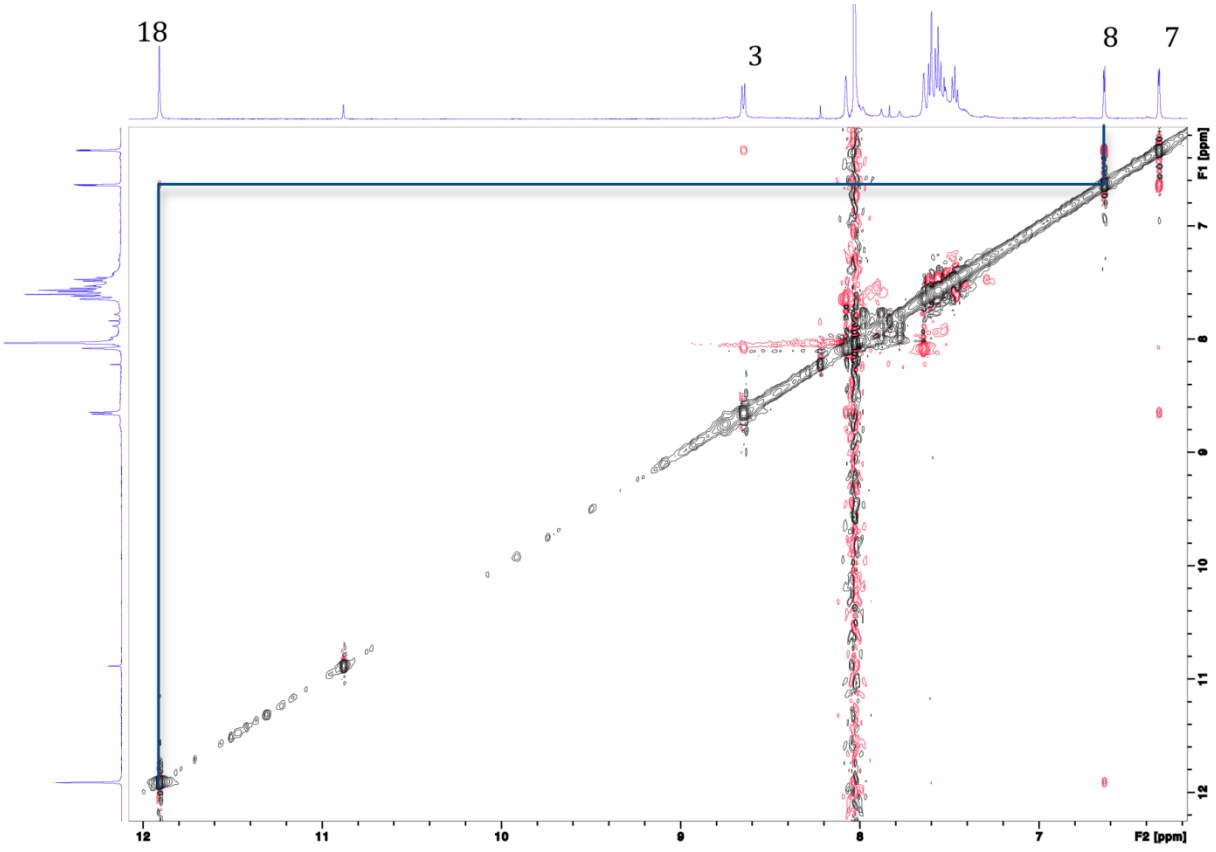

Figure 52: ROESY spectrum (-50°C.150ms mixing time, 500MHz)

10.6 Structure analysis of modified peptide 4.

The overall sequence of the peptide, together with the naming used in the structure analysis is depicted in figure 53. All NMR data was recorded at 25°C and in DMF-d7 as solvent. The following analysis allowed to confirm the incorporation of the modified amino acid at the proposed position in the sequence, and the integrity of the PTAD adduct following reaction.

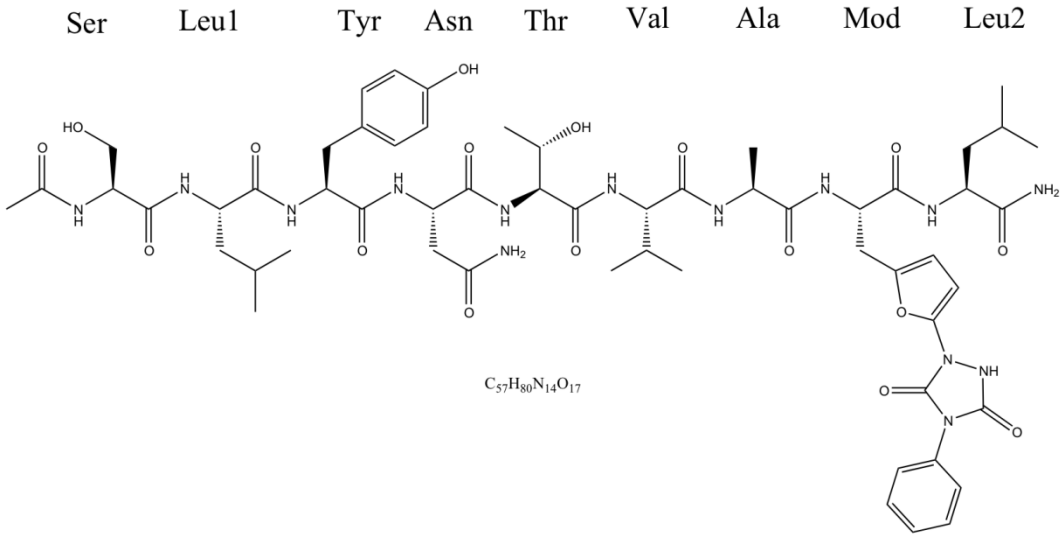

Figure 53: General structure of peptide 4

Except for the two leucines all the other amino acids are different, thus one can assume that the overlap of the different ^{1}H signals will be limited. By using the TOCSY spectrum (figure 54), the different individual amino acids are easily assigned. Since no subsets of signals can be found, we conclude there is only one compound or conformation present. The complete structure analysis will not be discussed here since it has limited added value, but the overall assignment of the signals in the ^{1}H and ^{13}C spectra can be found in the summarizing tables 3 and 4.

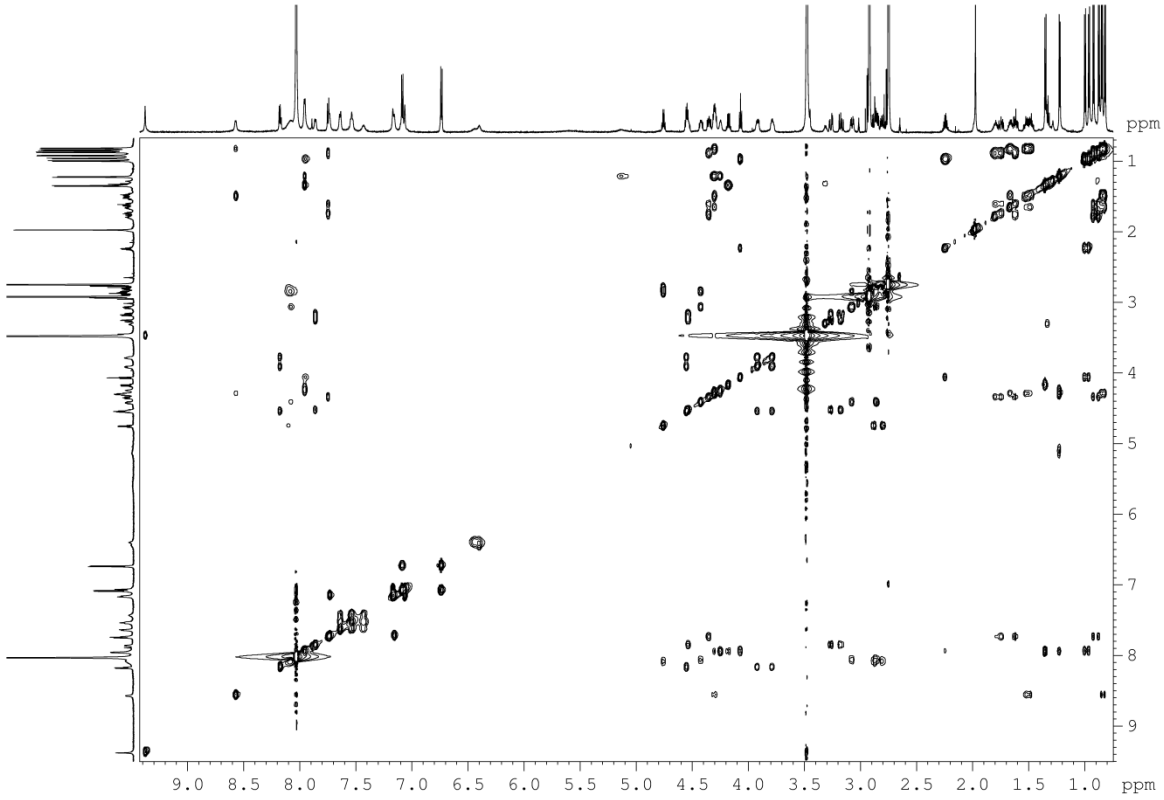

Figure 54: Overview TOCSY spectrum (25°C, 700MHz, tm 100ms)

In figure 55 the NH_i – NH_{i+1} sequential walk connecting the various amino acids identified from the TOCSY spectrum is shown and validates the peptide sequence. Possible ambiguities due to overlap were resolved through the α_iH – NH_{i+1} sequential walk, shown in figure 56. Here, an overlay of a NOESY (black) and TOCSY spectrum (red) at room temperature is shown.
The incorporation of the modified amino acid can be confirmed by a contact between the βCH_3 of the preceding alanine and the NH of the modified one and by a contact between the βCH_2 of the modified residue and the NH amide proton of following C-terminal leucine (Leu2). These and other distinct contacts provide conclusive proof for the incorporation of the modified amino acid between alanine and the second leucine in the sequence.

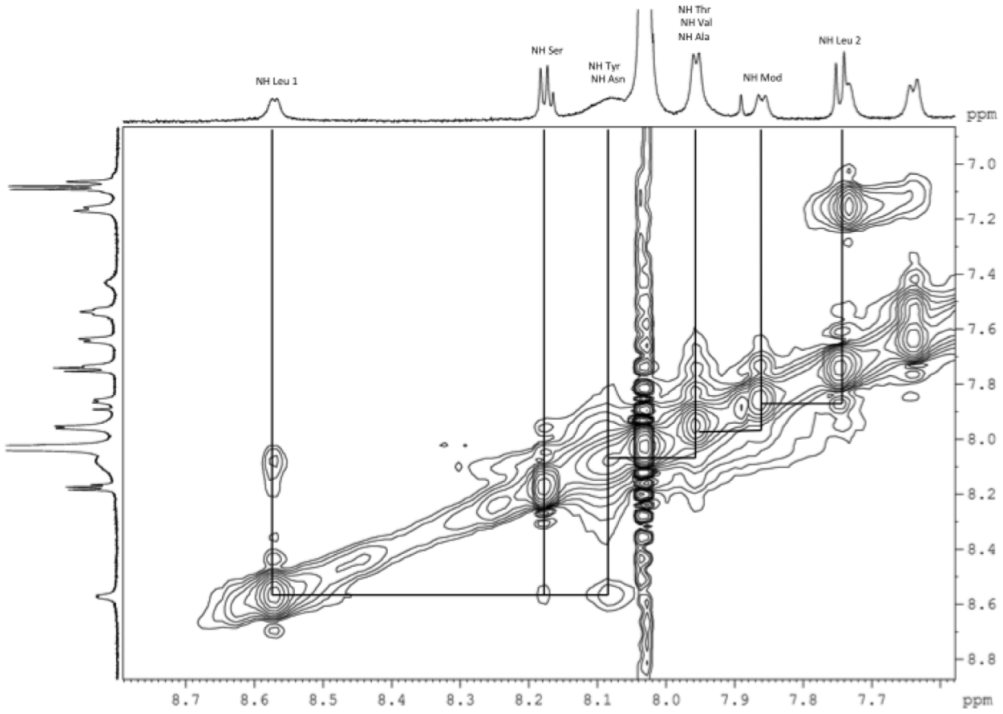

Figure 55: Zoom NOESY NH amide region (25°C, 700MHz, tm 600ms)

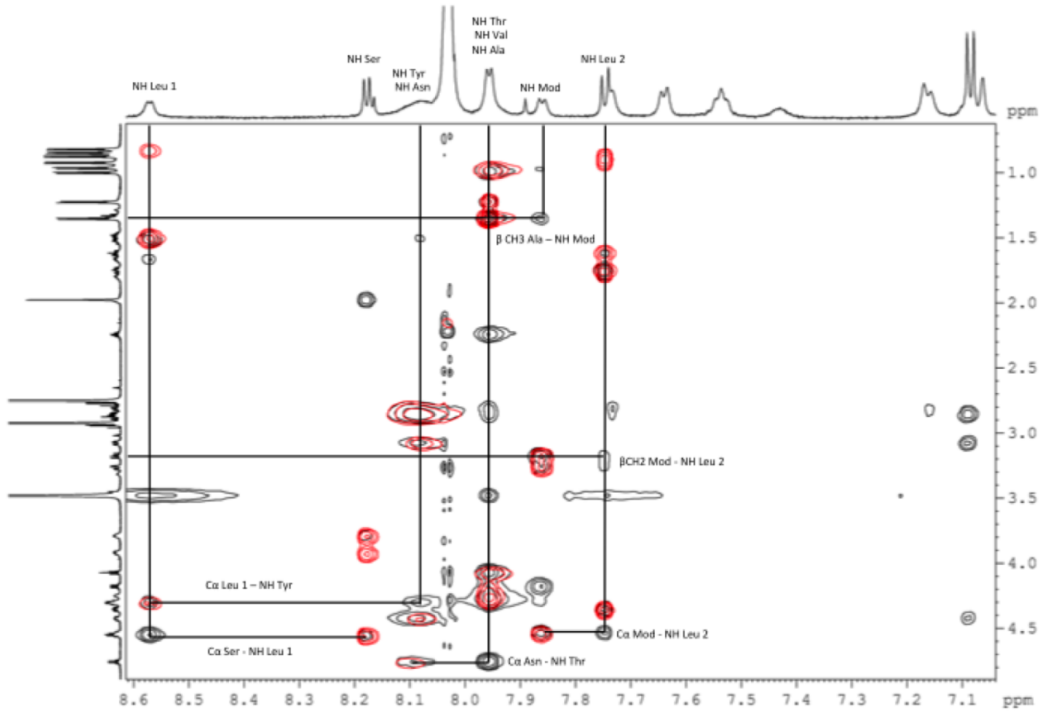

Figure 56: Overlay NOESY (black) – TOCSY (red) NH amide region (25°C, 700MHz,, t_m 100ms)

Integrity of the PTAD adduct.

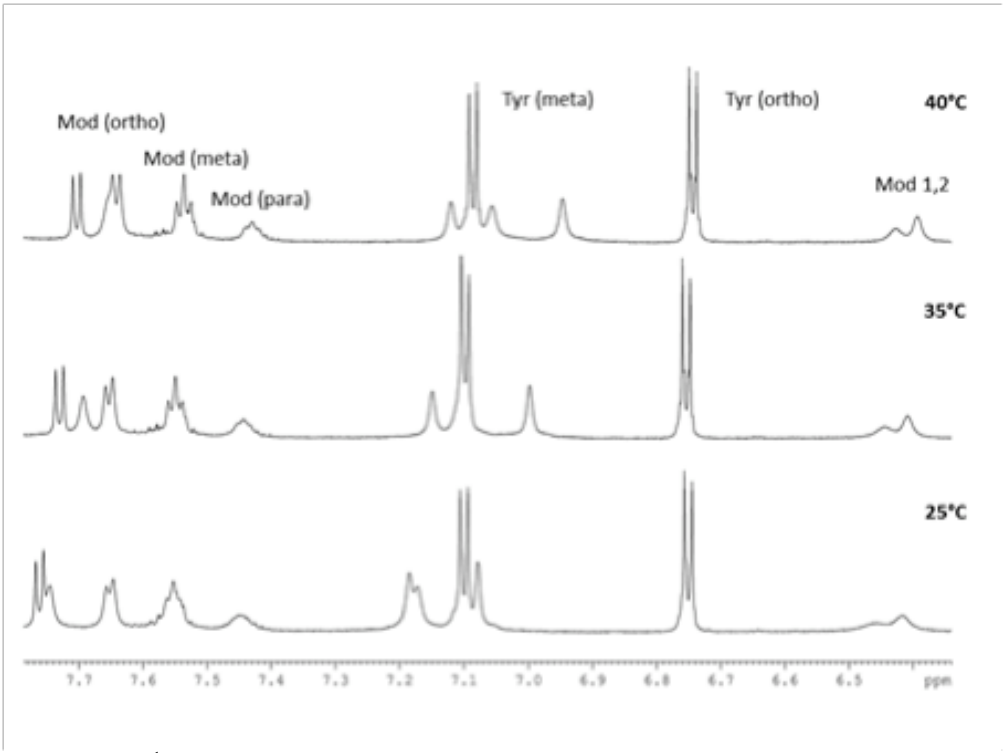

Figure 57: Overview of the 1D ^{1}H alkene/aromatic region in DMF-d_7 for peptide **4** showing the decrease in line width of furan signals (Mod1.2) and aromatic (Mod ortho, meta and para) signals of the introduced PTAD label with increasing temperature. Ortho and meta signals of tyrosine (Tyr ortho, meta) are intact. The moving signals are assigned as amide signals.

Assuming a Diels-Alder type furan-PTAD adduct, the broadening of some of the alkene and aromatic signals (Figure 57), was originally interpreted in terms of a diastereomeric mixture of the obtained adducts. However, given that the sharpening of these signals with increasing temperature suggests a conformational equilibrium, an alternative explanation could be the occurance of a pyramidal inversion process at the assumed ring fusion N atoms (main article, Figure 2). At this point of the analysis, all protons are assigned except for the two protons present on the furan-linker of the modified amino acid and the primary amide signals of Asn and the C-terminus. All these can typically occur between 6 and 8 ppm. The former are distinguished from the latter by noting the absence of correlations to an associated carbon for the amide hydrogens in the HSQC (not shown here). At 6.62ppm there are two broad and partially overlapping signals that also show a $^{1}J_{CH}$ correlation to two overlapping methyne type carbon resonances ~ 109ppm. Both the 1H and 13C chemical shifts are in line with those of H7 and H8 previously identified in structure B, and thus should correspond with protons mod 1 and mod 2 (figure 57).
Furthermore, these signals show a low intensity coupling pattern in the TOCSY spectrum with the αCH and βCH_2 protons of the modified amino acid, resulting from small but active long range couplings, similar to the ones seen in the spectra of structure B (figure 42).

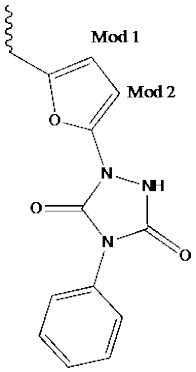

Figure 58: side chain modified amino acid

There is no additional correlation to be found to a third proton in the structure of the modified side chain, again providing additional proof that here, as well, the structure of the modified side chain corresponds to the one determined earlier and not to structure A. One last piece of evidence is the fact that a broad signal appears at 11.58ppm which corresponds to the labile NH proton of the side chain of the modified amino acid, which has almost the same chemical shift (11.61ppm).
The remaining unassigned resonances in the 6-8ppm region can be divided in two pairs based on mutual correlations in the COSY spectrum (figure 59) and do not show any correlation in the HSQC spectrum, as mentioned earlier. Their chemical shift is typical for a primary amide, and adding a drop of D_2O to the solution leads to their disappearance, confirming their exchangeable nature (not shown). The broad signals at 6.62ppm, assigned to mod1 and mod2 in the modified side chain, remain intact.

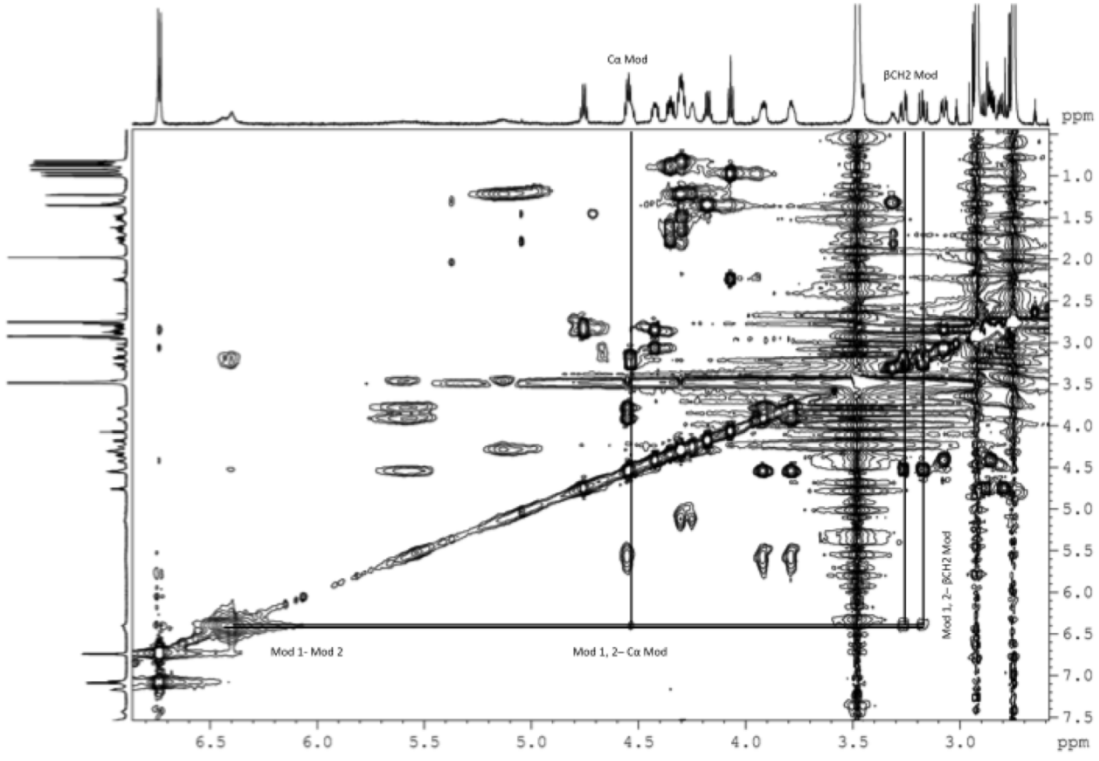

Figure 59: Zoom TOCSY coupling pattern modified amino acid (25°C, 700MHz, tm 100ms)

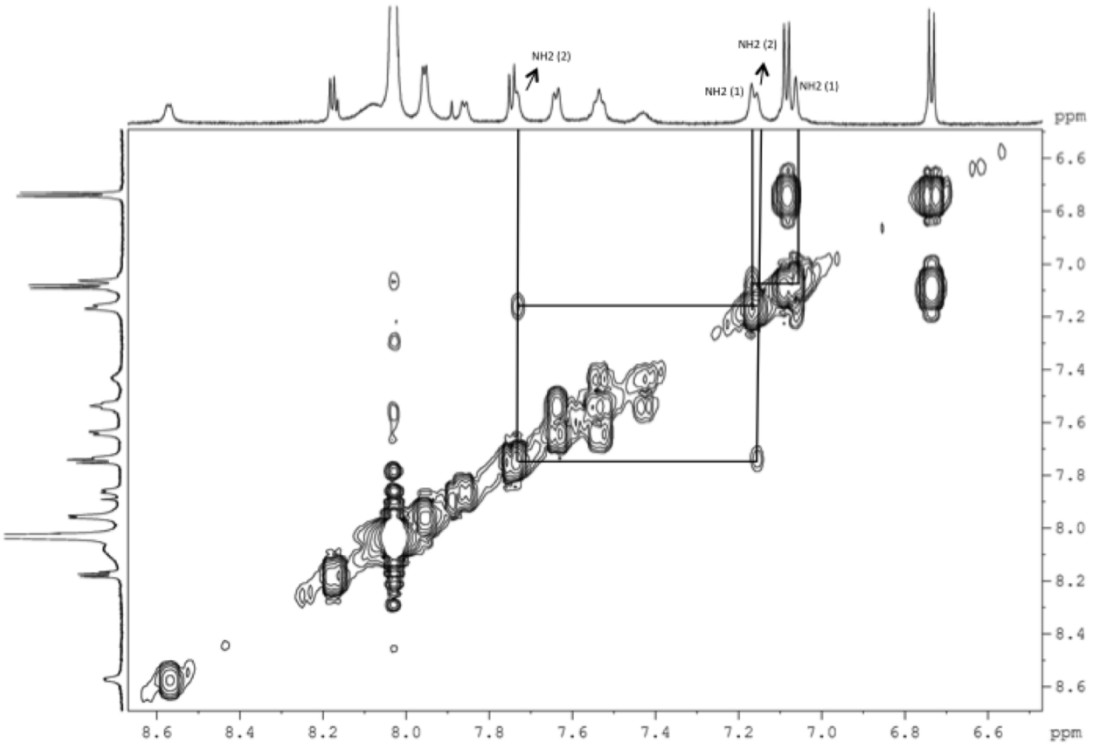

Figure 60: Zoom COSY coupling pattern two amide groups (25°C, 700MHz)

To conclude the analysis, an overview of the complete assignment is given in tables 3 and 4 for the ^{1}H and ^{13}C signals respectively. (DMF ^{1}H: 2.75; 2.92 and 8.03ppm – ^{13}C: 30.56; 35.72; 163.15ppm).

^{1}H Chemical shifts

Sequence	Finestructure	Chemical shift δ (ppm)	Integral	# protons	Annotation
C-terminus	singlet	1.98	3.06	3	CH3 (1)
Serine	doublet	8.17	0.97	1	NH Ser
	multiplet	4.54	1.99	2	Cα Ser + Cα Mod
	multiplet	3.79	1.22	1	β CH2 Ser
	multiplet	3.92	1.12	1	β CH2 Ser
	broad singlet	5.60	0.98	1	OH Ser
Leucine 1	doublet	8.57	1.16	1	NH Leu 1
	multiplet	4.30	2.29	2	Cα Leu 1 + β CH Thr
	multiplet	1.50	2.41	2	β CH2 Leu 1
	multiplet	1.66	1.28	1	γ CH Leu 1
	doublet	0.82	3.18	3	δ CH3 Leu 1
	doublet	0.85	3.43	3	δ CH3 Leu 1
Tyrosine	broad singlet	8.08	1.84	2	NH Tyr, NH Asn
	multiplet	4.44	0.94	1	Cα Tyr
	doublet of doublets	3.08	1.22	1	β CH2 Tyr (1)
	multiplet	2.83	4.92	3	β CH2 Asn + β CH2 Tyr (2)
	doublet	6.74	2.00	2	ortho aromatic CH Tyr
	doublet	7.08	2.07	2	meta aromatic CH Tyr
	singlet	9.38	0.93	1	OH Tyr
Asparagine	broad singlet	8.08	1.84	2	NH Tyr, NH Asn
	quartet	4.76	1.03	1	Cα Asn
	multiplet	2.83	4.92	3	β CH2 Asn + β CH2 Tyr
	singlet	7.06	0.92	1	NH2 (1)
	singlet	7.17	1.02	1	NH2 (1)
Threonine	doublet	7.95	2.87	3	NH Thr, NH Ala, NH Val
	triplet	4.25	0.93	1	Cα Thr
	multiplet	4.30	2.29	2	Cα Leu 1 + β CH Thr
	broad singlet	5.15	0.82	1	OH Thr
	doublet	1.23	2.93	3	γ CH3 Thr
Valine	doublet	7.95	2.87	3	NH Thr, NH Ala, NH Val
	triplet	4.07	0.95	1	Cα Val
	octoplet	2.24	1.09	1	β CH Val
	doublet	0.96	3.14	3	γ CH3 Val
	doublet	1.00	3.08	3	γ CH3 Val
Alanine	doublet	7.95	2.87	3	NH Thr, NH Ala, NH Val
	pentuplet	4.18	1.04	1	Cα Ala
	doublet	1.35	2.95	3	β CH3 Ala
Modified	doublet	7.86	1.08	1	NH Mod

	multiplet	4.54	1.99	2	Cα Ser + Cα Mod
	doublet of doublets	3.17	1.05	1	β CH2 Mod.
	doublet of doublets	3.27	1.01	1	β CH2 Mod.
	singlet	6.39	0.65	1	mod 1
	singlet	6.43	0.75	1	mod 2
	broad singlet	11.58	0.45	1	NH mod
	triplet	7.43	1.13	1	para aromatic mod
	triplet	7.54	2.28	2	meta aromatic mod
	doublet	7.64	1.79	2	ortho aromatic mod
Leucine 2	doublet	7.75	1.17	1	NH Leu 2
	multiplet	4.35	1.18	1	Cα Leu 2
	multiplet	1.75	1.20	1	β CH2 Leu 2
	multiplet	1.80	1.16	1	γ CH Leu 2
	doublet	0.88	3.24	3	δ CH3 Leu 2
	doublet	0.92	3.38	3	δ CH3 Leu 2
N-terminus	singlet	7.15	0.96	1	NH2 (2)
	singlet	7.73	0.71	1	NH2 (2)

Table 3: Overview ^{1}H assignment in DMF-d7 (*(1) region lies between two solvent peaks with some impurities, hence the deviation in integral value, *(2) two sets of amide protons could be identified NH_2 (1, 2) but not undoubtedly assigned)

As can be seen from table 4, there are still 10 carbon atoms missing. All of these carbon atoms are quaternary ones, which are known to have a long relaxation time (T_1) and thus can be hard to detect within a reasonable measuring time. Their assignment via a ^{13}C APT spectrum was not pursued considering the limited amount material available.

^{13}C Chemical shifts

Sequence	Atom type	Chemical shift δ (ppm)	Annotation	# carbon atoms
C-terminus	CH3	23.25	CH3 1	1
	Cq	171.45	Cq 2	1
Serine	CH	56.9	Cα Ser	1
	CH2	63.59	βCH2 Ser	1
Leucine 1	CH	54.28	Cα Leu 1	1
	CH2	40.94	βCH2 Leu 1	1
	CH	25.5	γ CH Leu 1	1
	CH3	22.09	δCH3 Leu 1	1
	CH3	23.63	δCH3 Leu 1	1
	Cq	174.41	Cq Leu 1	1
Tyrosine	CH	57.18	Cα Tyr	1
	CH2	37.26	βCH2 Tyr	1
	Cq	157.58	Cq Tyr 1	1
	CH	116.16	ortho CH Tyr (aromatic)	2
	CH	131.25	meta CH Tyr (aromatic)	2
	Cq	129	Cq Tyr 2	1
Asparagine	CH	52.27	Cα Asn	1
	CH2	37.86	βCH2 Asn	1
	Cq	173.6	Cq Asn/ Cq sidechain Asn	1
Threonine	CH	62.08	Cα Thr	1

	CH	67.71	Cβ Thr	1
	CH3	20.72	γCH3 Thr	1
Valine	CH	61.86	Cα Val	1
	CH	30.61	βCH Val	1
	CH3	19.78	γCH3 Val	1
	CH3	20.11	γCH3 Val	1
	Cq	173.83	Cq Val	1
Alanine	CH	51.66	Cα Ala	1
	CH3	17.64	βCH3 Ala	1
	Cq	174.55	Cq Ala	1
Modified	CH	54.28	Cα Mod	1
	CH2	30.94	βCH2 Mod	1
	CH	109.64	mod C1 + mod C2	2
	CH	127.33	ortho CH Mod (aromatic)	2
	CH	129.98	meta CH Mod (aromatic)	2
	CH	128.89	para CH Mod (aromatic)	1
	Cq	171.61	Cq Mod	1
Leucine 2	CH	53.05	Cα Leu 2	1
	CH2	41.88	βCH2 Leu 2	1
	CH	25.55	γ CH Leu 2	1
	CH3	21.87	δCH3 Leu 2	1
	CH3	24.07	δCH3 Leu 2	1
			Total	47

Table 4: Overview 13C assignment in DMF-d7