5 Dimensional Structural Characterization of Synthetic Peptides

Dr Osama Chahrour, Dr John Malone, Diego Cobice, Dr Dan Fletcher, Dr Alan Thompson, and Cliodhna McGurgan

Almac Sciences, 20 Seagoe Industrial Estate, Craigavon, Northern Ireland, BT63 5QD

Background

There is currently growing interest in the therapeutic benefits of synthetic peptides. These molecules, like small molecule pharmaceuticals, require comprehensive characterization to unequivocally confirm the structure as per the USP guidelines. Full identity elucidation requires integration of orthogonal techniques to overcome the limitations associated with each technique. Presented is an example of the characterization protocol developed at Almac for synthetic peptides. This process confirms peptide sequence by Nuclear Magnetic Resonance and Tandem Mass Spectrometry, peptide molecular formula by accurate mass determination, chiral purity by pre-column derivatization of the peptide hydrolysates followed by liquid chromatography with mass spectrometry detection as well as ratio determination of each amino acid in the sequence. This characterization protocol goes beyond the requirements of the USP and gives confidence in the identification of synthetic peptides. The protocol has already been used for peptides containing up to 34 amino acids . A case study is presented herein.

Peptide Sequencing by NMR

~25 mg of a 20-mer peptide was dissolved in DMSO-d_6. Spectra were obtained using a Bruker Avance NMR spectrometer operating at 500 MHz running under TopSpin 2.1. The sample was run at 28 and 50°C in the presence and absence of TFA. ^{1}H and ^{13}C spectra were assigned using DEPT-135, HSQC, HMBC, ROESY and TOCSY experiments. Individual peaks were picked out in the HSQC (finger-printing) and their multiplicity determined with the aid of the DEPT-135 spectrum. The correct number and multiplicity of carbons and protons was also confirmed.

COSY-DQF, Transverse-ROESY and HSQC-TOCSY data was to assign the protons to individual amino acid side chains and confirm they agreed with the expected residues. The ROESY and HMBC data was used to determine the relative positions and sequence of amino acids within the molecule

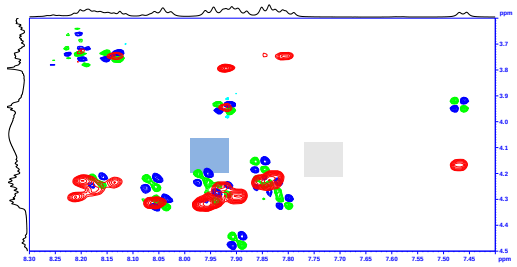
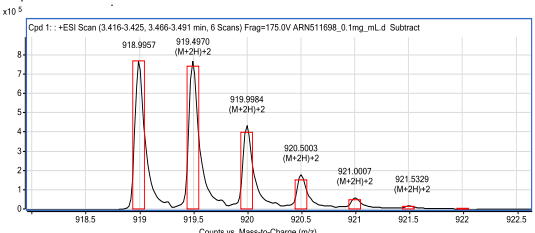


Figure 1: *Intra*-residue NH-H^{α} COSY correlations (green/blue). ROESY *Inter*-residue correlations (red) across the amide bonds from the α-H of residue *n* to the NH of the residue *n+1* were used to confirm the correct peptide sequence .

Accurate Mass Determination by LC-ESI-TOF

The accurate mass of the peptide was determined using an Agilent LC-QTOF to within 2ppm of theoretical. The Find By formula Algorithm of the MassHunter software was used to map the MS(1) data. It identified all major cluster ions and the algorithm confirmed the empirical formula .



Best	ID Source	Formula	Diff (ppm)
TRUE	FBF	$C_{79}H_{133}N_{23}O_{27}$	-1.60

Figure 2: Isotopic distribution matching between actual spectrum (black line) and FBF calculated theoretical distribution (red boxes) and table of mass confirmation.

Peptide Sequencing by LC-ESI-QTOF

The peptide precursor ions identified during the accurate mass determination were fragmented in the tandem mass spectrometer and the product (fragment) ions were subsequently mass analysed. The implemented low energy collision induced dissociation (CID) primarily produced y and b type ions. The data were processed using the Molecular Feature Extractor algorithm of MassHunter and matched with theoretical peptide fragment ions. The majority of y and b ions observed were within 10 ppm difference.

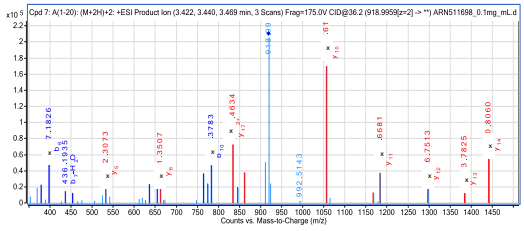


Figure 3: Isotopic distribution matching between actual spectrum (black line) and FBF calculated theoretical distribution (red boxes).

Precursor ion	m/z (prec.)	Number of charges Z
(M+H)+	1X36.9849	1

Product Ion	m/z	m/z (prod.)	Diff (ppm)	Sequence	Z
b6	3X7.1777	3X7.183	-13.3	R^1GGR^2GA	1
b10	78X.3853	78X.3791	7.8	$R^1GGR^2GAGR^3LQ$	1
y5	5X2.3074	5X2.3089	-2.9	$GSLR^6K$	1
y6	6X1.3497	6X1.3515	-2.8	R^5GSLR^6K	1
y7	7X4.431	7X4.4356	-5.9	LR^5GSLR^6K	1
y8	8X5.4695	8X5.4727	-3.8	ALR^5GSLR^6K	1
y10	105X.6084	105X.6095	-1.1	$R^4LALR^5GSLR^6K$	1
y11	118X.6594	118X.6681	-7.4	$QR^4LALR^5GSLR^6K$	1
b10-H2O	7X4.3609	7X4.3686	-10	$R^1GGR^2GAGR^3LQ$	1
y11-NH3	1X66.648	1X66.6416	5.5	$QR^3LALR^4GSLR^5K$	1

Figure 4: Example of peptide precursor ion and corresponding product ions.

Chirality Determination by HPLC-MS

The chiral purity of the amino acids in the peptide were determined by acid hydrolysis followed by amino acid separation. Racemisation was circumvented by hydrolysing peptides in DCl/ [D4] acetic acid (1:1). Amino acids were then derivatised with Marfey's reagent and analysed by LC-MS. This method allows the original chiral purity of each amino acid in a peptide to be determined using common achiral reversed-phase HPLC columns. Use of ESI prevented fragmentation during ionisation and allowed the protonated molecular ion to be monitored for each amino acid

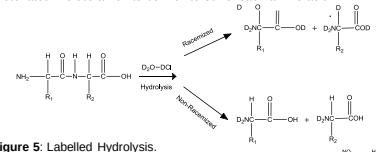


Figure 5: Labelled Hydrolysis.

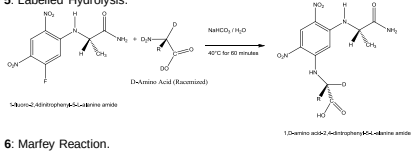


Figure 6: Marfey Reaction.

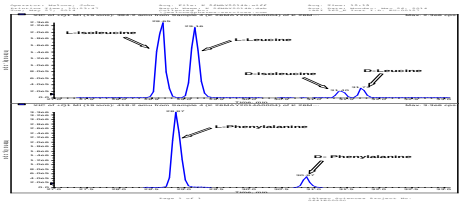


Figure 7: Example representative L and D amino acid standard chromatograms.

Amino Acid Analysis

Amino acid analysis was carried out by acid, oxidative and alkaline hydrolysis. The individual amino acids were then analysed with HPLC-UV/FLU and/or Biochrom amino acid analysers using classical ion-exchange liquid chromatography with post-column Ninhydrin derivatisation and photometric detection. Li-Citrate as well as Na-Citrate buffer system elution was used to cover all the amino acids. The presence and correct number of amino acids in the peptide sequence were confirmed.

Conclusion

A summary of the results is given in the table below:

R^1	G	G	R^2	G	A	G	R^3	L	Q	R^4	L	A	L	R^5	G	S	L	R^6	K
Sequence Confirmed by NMR																			
Sequence confirmed by CID Hybrid MS																			
Chirality Confirmed by LC-MS																			

The protocol for peptide characterization has been fully implemented by Almac has been used routinely to unequivocally confirm the structure of an number of peptides for our clients