Modification of peptides using Palladium mediated C-H activation

David Turnbull

MSc by Research

University of York

Chemistry

February 2018

Abstract

Palladium mediated C-H bond activation has become a popular tool used by many organic chemists. Whilst there are many different types of C-H activation processes, the methods used for biologically relevant molecules is relatively low, with most catalysis on biological molecules being on cysteine residues (thiol functionalisation), and lysine residues (amine functionalisation).

The work presented in this thesis continues the use of tryptophan as the residue of choice for functionalisation. Tryptophan contains an indole ring lending itself to C-H bond activation, due to its intrinsic reactivity. A big question regarding peptide and protein modification is whether modification is feasible for more complex peptides, where selectivity (site of functionalisation) becomes a challenge. The aim of the project is to explore the scope of tryptophan arylation using Pd catalysis and aryl diazonium salts, to determine whether the arylation procedure caused any erosion of the chiral information at tryptophan through racemisation, and to what extent does the procedure work on peptides and what effect it has on the conformation.

The procedure used is typically used on L-tryptophan due to the biological usability, but does the procedure arylate both enantiomers? Using chiral HPLC, it can be seen that di-protected racemic tryptophan can be arylated and that two peaks will be observed next to each other. Overlaying the L-enantiomer onto the graph shows that the procedure works for both enantiomers meaning that it is not enantiomerically selective. The procedure worked on peptides and the conformational changes were monitored using ROESY NMR spectroscopy.

Contents

Table of Figures

Table of Tables

Table of Appendices

Declaration

I declare that this thesis is a presentation of original work and I am the sole author.

This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

All work is my own except for the following:

- All ESI-MS experiments were carried out by Dr Ed Bergstrom and Mr Karl Heaton.
- All NMR experiments on peptide starting materials and products were carried out by Dr Alex Heyam

Acknowledgements

This has been probably the most difficult year of my life. From moving away from home for the first time to the passing of friends and family alike. I would like to thank Ian Fairlamb for all the support he has given me, which has been a lot. I would like to thank my family and partner for keeping me somewhat sane during this tough time. There is an extra special thanks to the Fairlamb group, with Anders, Neil and Ben in particular helping me throughout this project. Charlotte Elkington, you have helped me even when I asked really silly questions and you are honestly the backbone that keeps the lab working, you are the real MVP. I must also thank Heather Fish, Karl Heaton and Alex Heyam for all the support with the techniques used.

I have learnt a lot about myself from this year, from never having done any NMR work to doing a project entirely dedicated to it. I would never have thought I could do this. There have indeed been times when I have wanted to quit, and with the help of the group, I got through it. I think that's the important lesson throughout all of this. Keep moving forward and you will get through this.

A special message for my uncle who passed away during this project. I must have got my fighting spirit from you, as you were always fighting to the end. There's not a day I don't wish you were here. I miss you and your practical jokes. I dedicate my life to living by the same principle you lived by.

In memory of Derek Knox, 2nd April 1958 – 9th July 2017

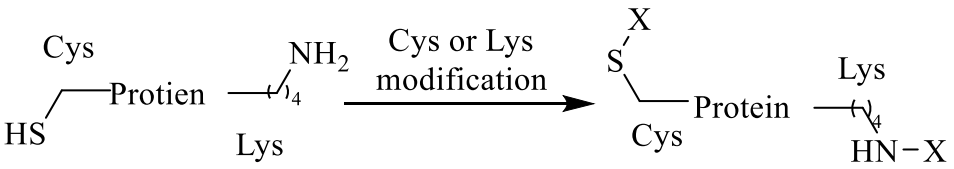
Chapter 1: Introduction

1.1. Protein Modification Challenges and Uses

Protein modification is found within nature and it increases the diversity of protein structures and function by up to two orders of magnitude.[1] Yet our ability to mimic nature in regard to synthetic protein modification has been limited due to the available synthetic tools. In addition to limited chemistry, there are several challenges with protein modification. Chemo- and regio- site-selectivity is difficult due to the presence of a sea of reactive functional groups; carboxylic acids, alcohols, amides, amines, and thiols, which makes modification around a single amino acid difficult. The reaction conditions are another challenge as proteins typically require biologically-ambient conditions (<37 ^{o}C, pH 6-8, aqueous solvent and salt buffer) to keep its structure and functionality. Any given reaction should ideally proceed to near total conversion to generate homogenous constructs.[2,3]

1.2. Classic Methods for Protein Modification

Protein bioconjugation is the formation of a stable covalent bond between two molecules, where one is a biological molecule. This involves using second-order reactions that selectively target a functional group present in the side chains of the proteogenic amino acids.[5] The two most commonly modified residues are cysteine and lysine.[2] The thiol group of cysteine can form mixed disulfides as well as alkylation with alkyl halides or reacting with α,β-unsaturated carbonyl compounds to yield thioethers via Michael addition.[5] Cysteine is relatively rare as an amino acid meaning it can more commonly be used for single-site modifications, this however can be on the proteins functional site resulting in loss of function.[2] Lysine is a popular target for modification due to the large amounts of methods to modify primary amines.[5] Lysine can react with activated esters, sulfonyl chlorides, isocyanates, or isothiocyanates to afford amides, sulfonamides, ureas, or thioureas respectively as well as with aldehydes to undergo reductive amination.[2]



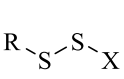
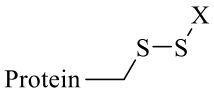
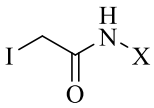
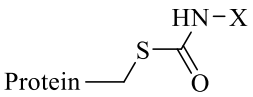
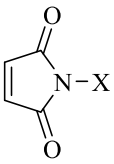
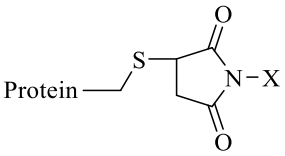
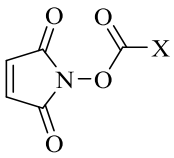
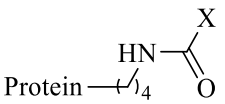
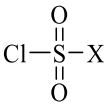
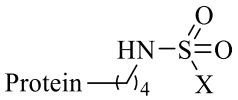
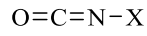
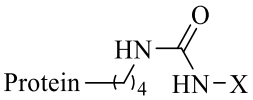
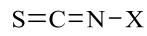
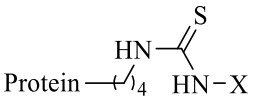
Entry	Residue	Reagent	Product
1	Cys		
2	Cys		
3	Cys		
4	Lys		
5	Lys		
6	Lys		
7	Lys		

Table 1: Modification of cysteine and lysine that have historically been used for protein modification.[2] Adapted with permission from Angew. Chem. Int. Ed Engl., 2009, 48, 6974–6998. Copyright of Wiley.

The remaining 18 proteogenic amino acids have not been exploited for residue-selective modification. Tyrosine has been modified through electrophilic aromatic substitution of the phenol moiety with aryl diazonium salts, iodine, or nitrous acid. Glutamate and aspartate residues have been targeted for bioconjugation to form carbodiimides by coupling with amines; although the possibility of cross linking proteins limits the utility of this method. Histidine residues have also been successfully modified using pyrocarbonates.[6] Through the use of these methods, conjugation of

small-molecule probes; biotin and fluorophores, to proteins has become routine and allows for a wide variety of products to be made, ranging from soluble polymers to microarray chips.

1.3. Modern Methods for protein modification

New methods have been developed for the modification of cysteine, lysine, tyrosine, and tryptophan especially through metal-mediated transformations.[7] Furthermore, the *N*-terminus has emerged as a popular target for protein modification (Table 2). Methods for the selective modification of amines and thiols of cysteine and lysine continue to be developed and optimized. McFarland and Francis reported a lysine-specific reductive alkylation that proceeds through an iridium-catalyzed transfer hydrogenation. Unlike the classic reaction based on sodium cyanoborohydride, the iridium-mediated process proceeds in high yield at pH 7.4 (Table 2, entry 1)[11] when compared to the original acidic conditions.

Davis and co-workers recently developed a two-step method for cysteine modification (Table 2 entry 2).[12] The first step is the transformation of cysteine into dehydroalanine by treatment with O-mesitylenesulfonylhydroxylamine under basic conditions which are non-ideal. The dehydroalanine residues then undergo a Michael addition with a thiol reagent to yield a thioether linkage. The Michael addition is not stereospecific resulting in a diastereomeric mixture of modified proteins.[12]

The *N*-terminus of a protein has unique pH-dependent reactivity and is thus an attractive target for single-site modification as shown by Dixon in 1984.[8] Its decreased pK_a value relative to amino groups on lysine side chains allows for selective acylation or alkylation. If there are many competing lysine side chains, it is difficult for this process to be selective. In 1964, Dixon performed a transamination at room temperature by using glyoxylate, catalytic base, and copper(I), which facilitated imine formation between the *N* terminus and the glyoxylate group (Table 2 entry 1 b).[9] This had previously been performed at 100 °C by Bonetti and Co-workers in 1956.[43]

Even with these improvements made by Dixon, the transamination reaction did not receive considerable attention until Francis and co-workers reported transamination that proceeds under ambient conditions without the need for metal or base additives (Table 2, entry 1 c).[10] The method involves condensation of the *N*-terminal amine with pyridoxal-5-phosphate and subsequent hydrolysis to result in a pyruvamide. The protein can then be further modified through the ketone.

Extensive screening revealed that the transamination reaction proceeded best when Ala, Gly, Asp, Glu, or Asn occupied the *N*-terminal position.

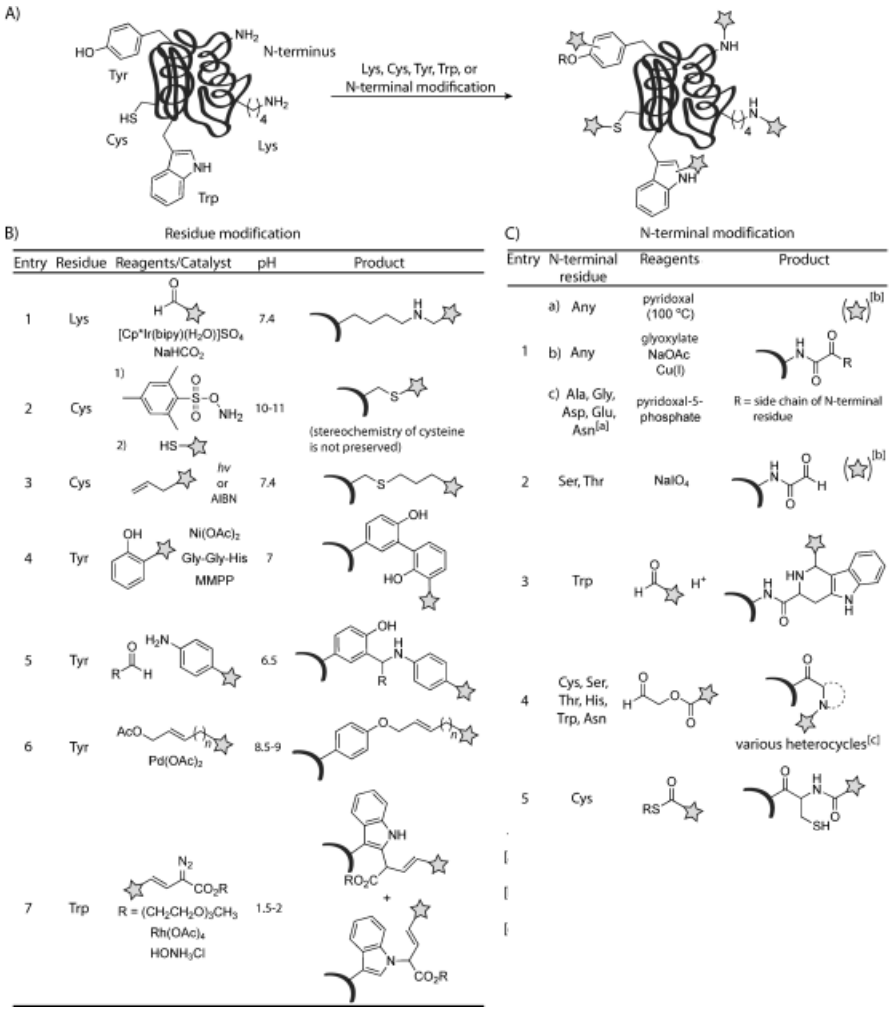


Table 2: Modern bioconjugate reactions for protein modifications involving particularly cysteine, tyrosine and tryptophan.[2]Adapted with permission from Angew. Chem. Int. Ed Engl., 2009, 48, 6974–6998. Copyright of Wiley.

There has been a lot of focus on the modification of tyrosine and tryptophan residues which often use transition-metal-mediated processes (see also section 1.4). These residues are relatively uncommon in proteins, whilst being present in the majority of proteins, presenting opportunities for single-site modifications. Kodadek and co-workers reported the first example of metal-mediated modification of tyrosine in 2005. This involves oxidative coupling of two phenol rings using a nickel(II) catalyst and co-oxidant to cross-link two peptides (Table 2, entry 4).[13] Additionally, Antos and Francis have developed a bioconjugation reaction for tryptophan, in which they used a rhodium carbenoid that was generated *in situ* from $[Rh_2(OAc)_4]$ and a diazo compound (Table 2, entry 7) which functionalised the C-2 position on the indole as well as the indole nitrogen atom.[14] However,

this reaction requires acidic conditions (pH≈2), which may affect the structure of some protein targets.

There are several other modification strategies that can be undertaken, as collated in Figure 1: 1) copper-catalysed azide alkyne cycloaddition (CuAAC); 2) Staudinger reaction; 3) strained-promoted azide alkyne cycloaddition (SPAAC); and, 4) inverse electron demand Diels–Alder (IEDDA) cycloaddition. Whilst each of these reactions allow for protein modification using mild, biologically-friendly, conditions, they each have separate issues. CuAAC requires Cu(I) salts which have a high toxicity as well as the issue that the terminal alkyne could possible react within a cell, particularly oxidative enzymes and cysteine residues in the active site of a protein.[16,19] There is an added limitation of lower stability in biological conditions, this is most likely due to the slightly acidic alkyne proton,[17] which can lead to alkyne homocoupling.[18] Limitations of SPAAC include side reactions between thiols and the alkyne, same as for CuAAC, but with the added issue that the large size of the strained alkyne is not as suitable for metabolic labelling.[20] The main limitation of the Staudinger reaction is that it is relatively slow compared to the other reactions in this section, and the synthesis of the components for the IEDDA reactions limits its application in the area.

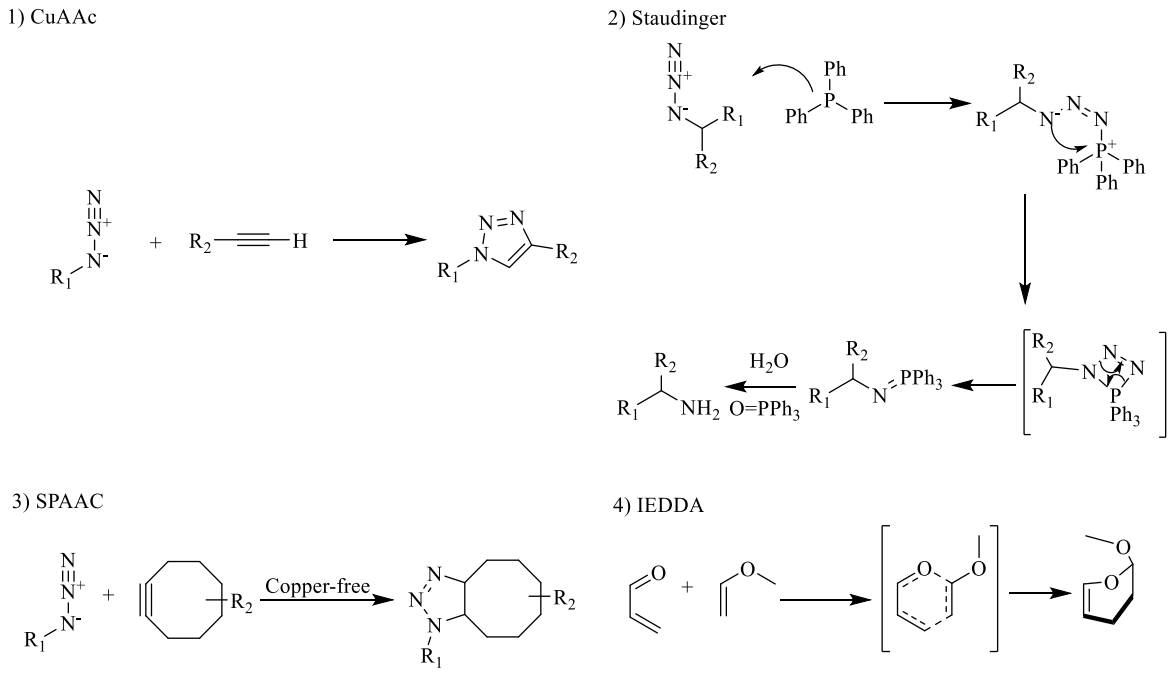


Figure 1: Different methods of adding a tag to a peptide. 1) CuAAC 2) Staudinger Reaction 3) SPAAC 4) IEDDA

1.4: Solid Phase Peptide Synthesis

Solid Phase Peptide Synthesis (SPPS) was pioneered by Robert Merrifield, receiving the Nobel prize in 1984, which is the method of choice for synthesising peptides. This is a particularly useful method for peptides which are difficult for bacteria to express, the incorporation of unnatural amino acids, peptide backbone modification and the synthesis of *D*-proteins.[34]

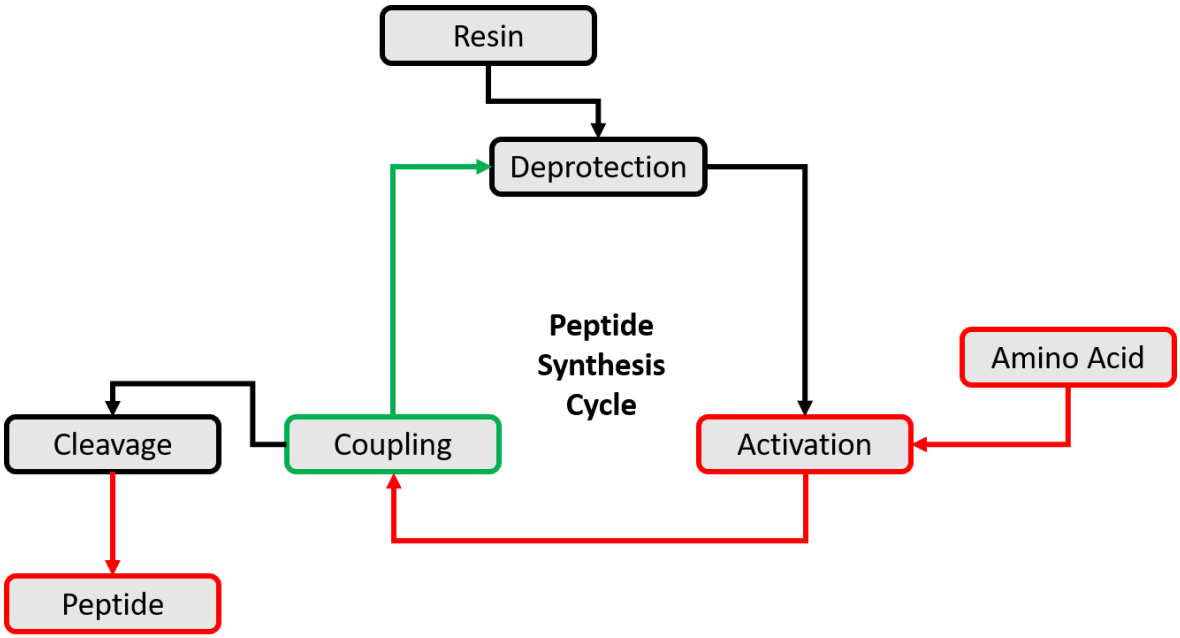


Figure 2: The process of SPPS

The solid phase support is usually one of three resins; polystyrene, polyamide or PEG-based. Each have their own benefits and issues. Polystyrene or the 'Merrifield resin' was the original support used and is still used today for its versatility, ease of use in automated peptide synthesis and minimum swelling in DCM. It is made up polystyrene cross-linked with 2% divinylbenzene, this produces a hydrophobic bead that is solvated by a nonpolar solvent such as DCM or DMF.[35] In recent years, new resins have been designed to be more chemically-inert or enhance swelling and rigidity. One example is cross-linked (50%) polystyrene, which allows for rapid reactions, better filtration of waste reagents and increased mechanistic stability.

Polyamide resin is also versatile, but it is known to swell much more than polystyrene making it unsuitable for automated synthesis and if the wells are too small.[36] The PEG-based resin is a modern

improvement to the previous resins and allows it to withstand repeated use of TFA during the deprotection stage.[36] It also allows for the C-terminus to be altered by introducing additional resins such as PAM and pMBHA resins. PAM produces a C-terminus carboxylic acid whilst pMBHA produces an amide.

1.4: Palladium Mediated Isolated Protein Modification

The use of palladium in modifying proteins is appealing due to how well studied palladium catalysis is, and its versatility as a catalytic metal. An early example, reported by Kodadek, was oxidative crosslinking which used palladium(II)/tetrakis-(4-methylpyridyl)-porphyrin complex (Pd^{II}TMPyP) with ammonium persulfate (APS) to undergo photo-redox upon irradiation with visible light. The reaction occurs due to the electron transferring from the palladium species to APS, acting as an electron acceptor, to form a radical which extracts an electron from tyrosine to form a tyrosine radical which can couple with electron-rich residues such as those found in protein active sites. This did however have flaws, cysteine and tryptophan were found to trap tyrosine radicals reducing crosslinking capabilities. The sustainability of this reaction in a cell would be low due to the toxicity of the palladium catalyst, as well as the fact that making radicals inside a cell would fuel unwanted side reactions and could be potentially fatal to the cell.[21]

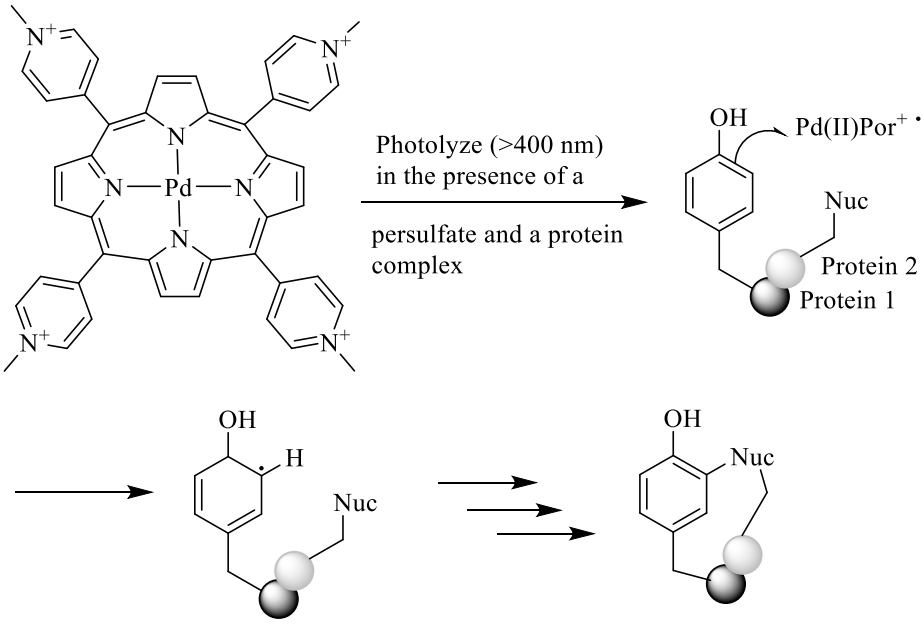


Figure 3: Reaction scheme for oxidative crosslinking reported by Kodadek [21]

Work conducted in 2006 by Tilley and Francis, based on the Tsuji-Trost reaction, allowed for site-specific tyrosine modification using an electrophilic π-allyl species and a palladium catalyst, prepared from allylic chloride and carbamate precursors then activated with palladium acetate in the presence of triphenylphosphine tris(sulfonate) (TPPTS) to aid water solubility. Conversions of 50-65% were seen in selective incorporation of lipophilic and dye molecules. [22]

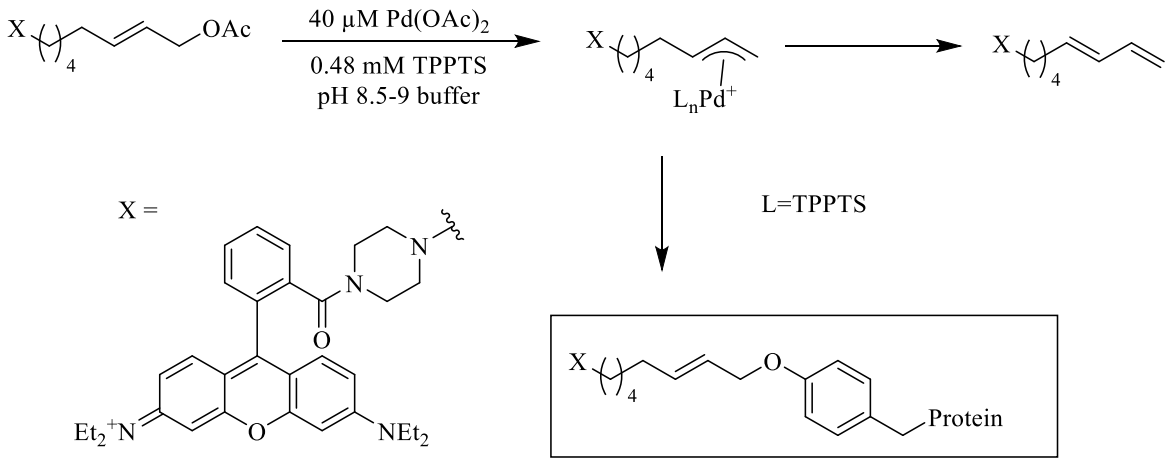


Figure 4: Site-selective alkylation using π-allylpalladium complexes [22]

There has been work done on modifying cysteine residues by several different groups worth noting. Buchwald and Pentelute reported cysteine-modificaiton with an aryl group in an unprotected peptide chain.[23] This was done using an organopalladium catalyst prepared using palladium chloride with an aryl halide and the 2-dicyclohexylphosphino-2,6-diisopropoxybiphenyl (RuPhos) ligand.[23] The reductive elimination gives a C-S bond which is efficient to form and allows the reaction to be done at low temperatures. Messaoudi and co-workers developed a third generation aminobiphenyl palladacycle precatalyst which in organic and aqueous solvents, allows cysteine to conjugate with either aryl or alkyne halides. This worked in large polypeptides and allowed tags to be attached in excellent yields (80-99%).[24] These two methods require that there is a solvent-exposed or single active residue to allow for the modification to occur in the correct way. When there are multiple active residues present, a heterogenous mixture is observed making these methods less favourable. In 2016, Davis found a way to bypass this issue by developing a method which binds the catalyst to a certain site, in this case aspartic acid, and allowing the nearby cysteine to be successfully modified. The catalyst for this method is prepared by using palladium acetate, aryl iodide and N,N-dimethyl-

2-amino-4,6-dihydroxypyrimidine (DMADHP) as a ligand. The reaction conditions are not as biologically friendly, 65 °C, but with the ability to selectively pick which residue to target, the benefits are certainly useful[25].

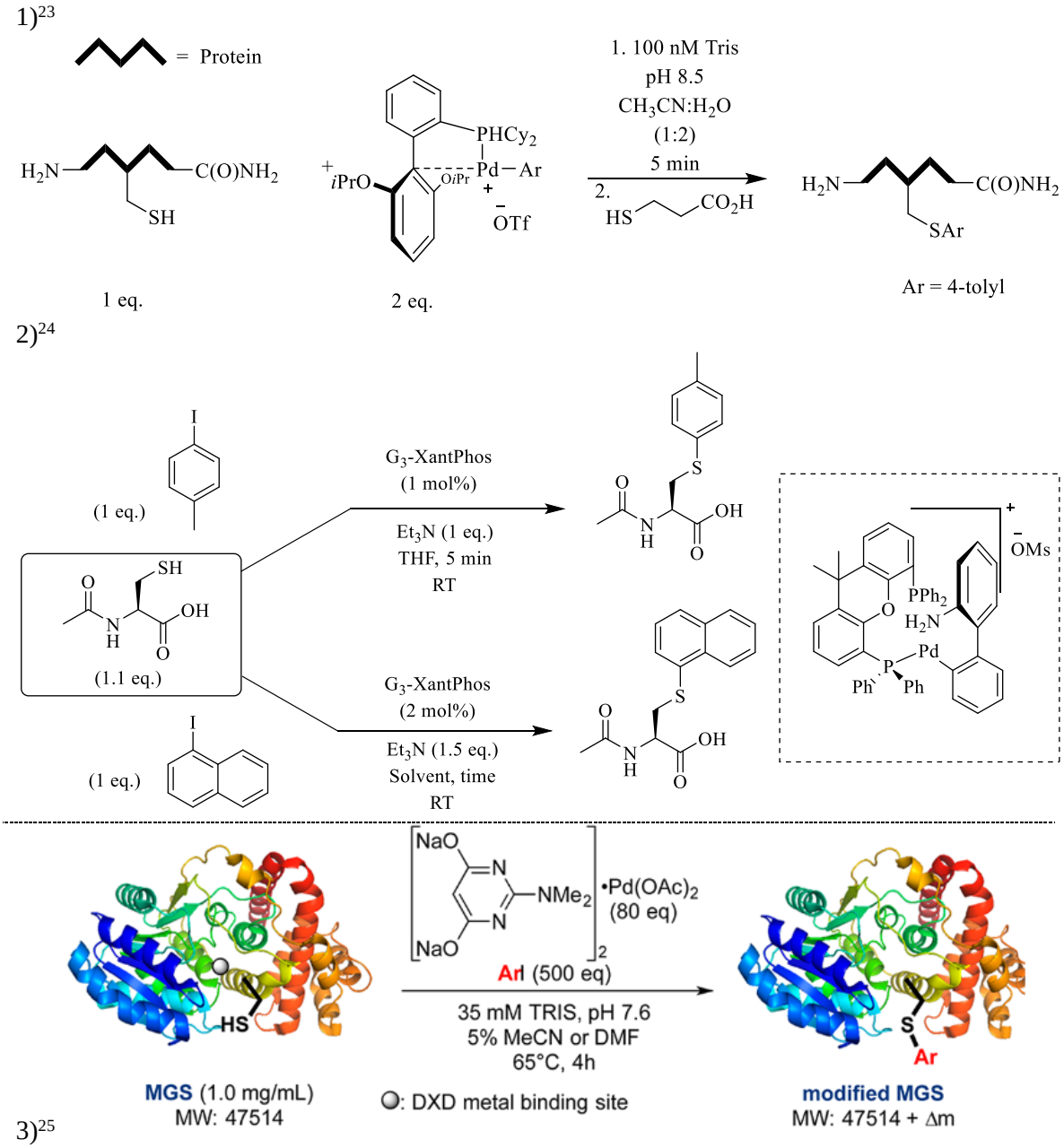
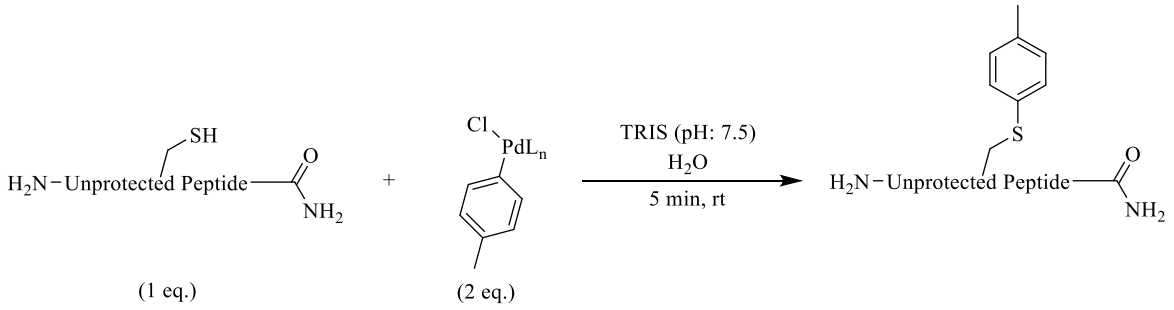


Figure 5: Palladium mediated modification method with cysteine 3) Reprinted with permission from J. Willwacher, R. Raj, S. Mohammed and B. G. Davis, J. Am. Chem. Soc., 2016, 138, 8678–8681. Copyright 2016 American Chemical Society.

In 2017, Buchwald and Pentelute managed to make a water stable palladium catalyst using a sulfonated biarylphosphine ligand and used this to modify an unprotected protein on a cysteine site. The catalyst activity was compared to other ligand, RuPhos and SPhos, and was found to get a 99% yield compared to the 20% from the other ligands. The reaction was completed in 5 minutes at room temperature in water with a TRIS additive.[44] The reaction scheme is below. (Figure 6)



Entry	Ligand	Yield (%)
1	L1	18
2	L2	14
3	L3	99

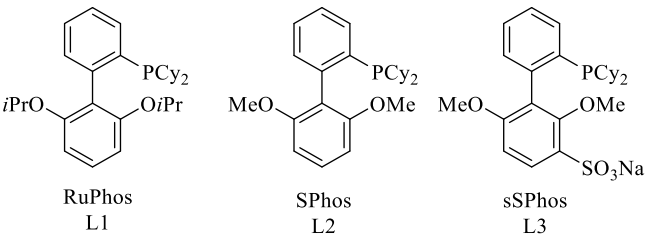


Figure 6: Pd mediated modification of unprotected peptides using a sSPhos ligand

1.5: Palladium mediated modified peptides within cell structures

Using the methods in the previous section on an isolated peptide/protein is one thing, to be able to do this inside a cell brings a whole host of additional problems but would allow for the selective modulation of cellular processes by modification of enzymes and proteins. The problems of using palladium species intracellularly include possible toxicity issues, the permeability/solubility of the palladium complex, a suitable bio-orthogonal handle and dealing with the chemically-rich environment inside the cell.[37,38] Despite these challenges, there have been studies that have shown selective modification of site specific proteins.

Davis in 2013 successfully demonstrated the first Suzuki-Miyaura coupling reaction on the membrane protein OmpC which required pre-modification via genetic incorporation of a para-iodophenol group for the selective modification to take place.[39] The reaction employed $Pd(OAc)_2$

and ADHP, which labelled the bacteria surface.[40] This work prompted other research groups to try different methods, *e.g.* Chen in 2014 reported NHC/palladium complexes that are efficient catalysts for Suzuki-Miyaura cross couplings and have allowed for two proteins to be labelled with aryl boronic acid derivatives on mammalian cell surfaces.[41]

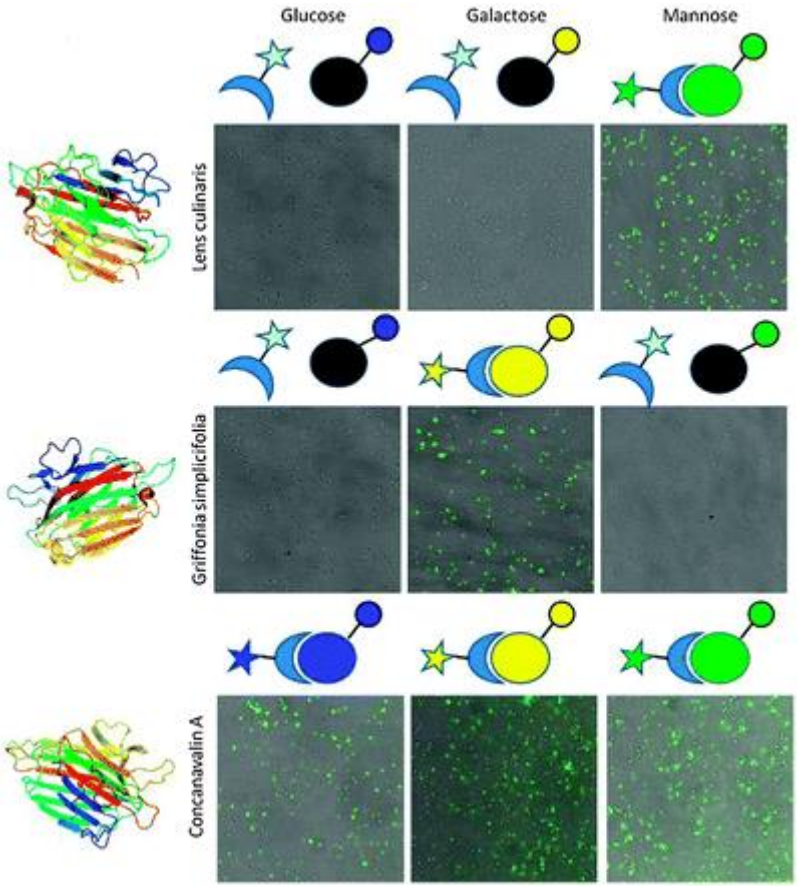


Figure 7: Interaction of fluorescein-lectin conjugate with E. coli with labelled boronic acids[40] Reproduced from Ref. 40 with permission from The Royal Society of Chemistry.

In 2011, Lin successfully conducted a Heck alkynylation using $Pd(OAc)_2$ and DM-ADHP, aryl iodides and alkyne-containing proteins. This was conducted using a modified amino acid with a propargyl group. The alkyne modified protein is found on the surface of the mammalian cell[41].This was further optimised to allow for specific labelling of the extracelluar domain of the cell, which is the epidermal growth factor for live mammalian cells.

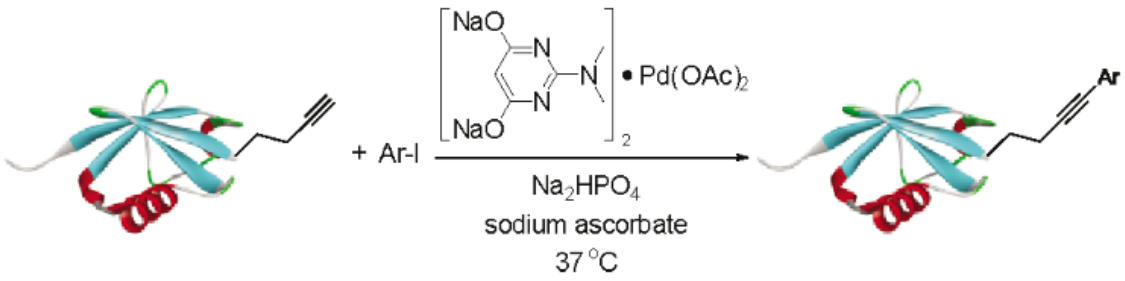


Figure 8: Heck alkynylation as reported by Lin in 2011.[41] Reprinted with permission from N. Li, R. K. V. Lim, S. Edwardraja and Q. Lin, J. Am. Chem. Soc., 2011, 133, 15316–15319. Copyright 2011 American Chemical Society.

In 2016, Hamachi reported a method for allosteric activation to modify protein function using palladium complexes with specifically inserted histidine residues. It works by replacing two specific amino acid in the protein sequence with histidine residues near the binding pocket in membrane receptor proteins. The histidine acts as a ligand with the $Pd(NO_3)_2$(bipy). When the palladium binds together to form the complex the protein adopts the required orientation for enhanced binding. The fact that the allosteric sites are artificially inserted and that a metal chelator residue in certain positions are an issue.[42]

1.6: Palladium Mediated Arylation of Tryptophan

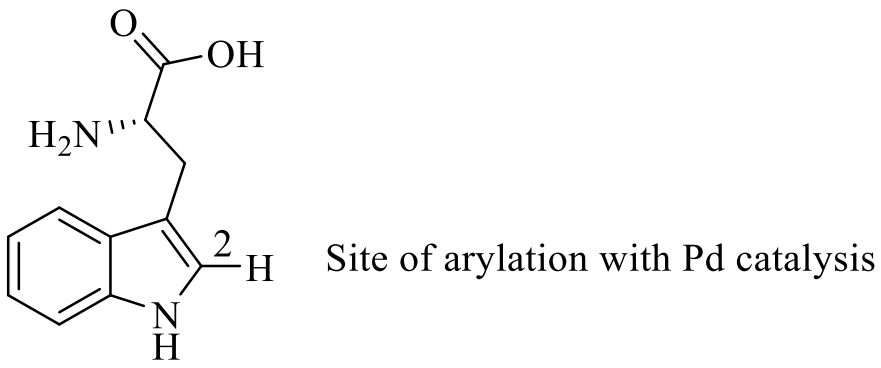
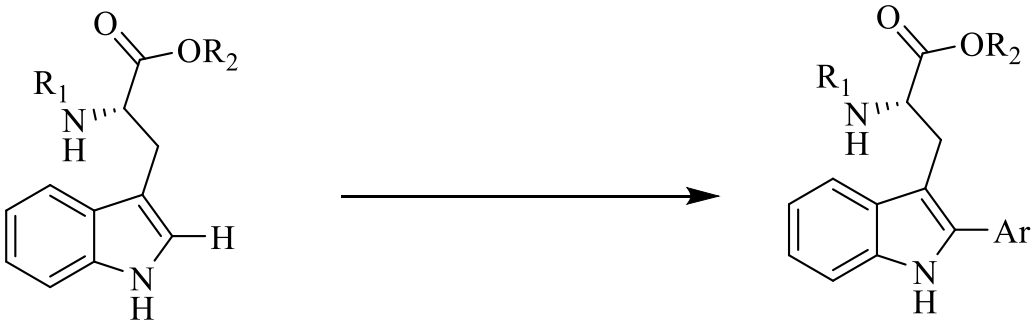


Figure 9: A tryptophan molecule showing the C2 hydrogen atom

Tryptophan is an amino acid which has displayed a high suitability for modification, especially via arylation. The structure of tryptophan is shown in Figure 9; with the indole side chain, intermolecular interactions, such as hydrogen bonds and aromatic π stacking to name a few, with biologically relevant molecules. The presence of tryptophan also endows fluorescence on the molecule with modification on the indole ring changing the wavelength of fluorescence of the molecule. With the use of a fluorescence detector, the modification of tryptophan can be determined whether to be successful or not by if the fluorescence has changed. By using tryptophan, which has a low abundance in natural proteins, estimated roughly at 1.1% in natural proteins and peptides. As well as tryptophan having a low abundance, about 90% of proteins contain tryptophan hence making this a viable way of modifying proteins. In recent years, numerous methodologies have been developed for selective C2 arylation of tryptophan and tryptophan derivatives whilst in 'free' form and in structures such as proteins and peptides.

There are several terms found in literature, these are 'C-H bond activation' and 'C-H bond functionalisation'. Often these terms have been interchangeable but for this review C-H activation refers to the formation of an organometallic species from activation of an initial C-H bond. C-H functionalisation refers to the formation of a functionalised C-R species from an initial C-H species which proceeds via an organometallic species. Whilst the idea of direct arylation is promising, it is not without its pitfalls, often the reactions do not have good green metrics, table 3 compares the work of the Fairlamb[26,27], Ackermann[28] and Albericio[29] groups in this area.

Table 3: Pd-Mediated C-H bond functionalisation of tryptophan



Entry	Ref.	R Groups	Cat.*	Arylating agent	Conditions	Yield / %
1	Fairlamb *et al*[26]	1.Ac-2.-OMe	$Pd(OAc)_2$, $Cu(OAc)_2$	PhB(OH)2 5 eq.	AcOH, 16 h, 40 °C, air	93
2	Fairlamb *et al*[27]	1.Ac-2.-OMe	$Pd(OAc)_2$	[MesPhI]OTf 2 eq.	EtOAc, 16 h, 25 °C	85 (91 conv.)
3	Ackermann *et al*[28]	1.Ac-Ala-2.-O-Et	$Pd(OAc)_2$	$[Ph_2I]OTs$ 1.5 eq.	AcOH, 17 h, 23 °C	99
4	Ackermann *et al*[28]	1.Ac-Ala-2.-O-Et	$Pd(OAc)_2$	$[Ph_2I]OTs$ 1.5 eq.	H_2O, 24 h, 23 °C	70
5	Ackermann *et al*[28]	1.Ac-Ala-2.-O-Et	$Pd(OAc)_2$	$[Ph_2I]OTs$ 1.5 eq.	H_2O, 24 h, 23 °C	60
6	Albericio/Lavilla *et al*[29]	1.Ac-2.-O-Me	$Pd(OAc)_2$	PhI 4 eq.	2-NO2Bz (1.5 eq.), AgBF4 (1 eq.), DMF, 5 min, 150 °C MW	89

*Catalyst loading was 5 mol% for $Pd(OAc)_2$ and 10 mol% for $Cu(OAc)_2$

Whilst there are issues, the advantages are that it removes the need for substrate pre-activation and has moderate to good yields. There are still changes being made to the procedures to make the reaction more environmentally friendly and still produce good yields. The type of catalyst typically used is $Pd(OAc)_2$ although alternatives such as $Pd(MeCN)_2(OTs)_2$ can be used to similar effect (in the correct solvent). When using $Pd(OAc)_2$, ethyl acetate is the solvent of choice, but it should be noted that the solvent does have an effect on the reaction rate as well as if an acid has been used in the system or not. The use of tosic and tetrafluoroboric acid reduces the catalyst induction period of the reaction which indicates that either the acids increase the rate of formation of the active catalyst

or the mechanism is different from the standard conditions. In a master's thesis from the Fairlamb group, an investigation into the role of acid additives was done.[30] When 5 mol% tosic acid was employed, a short induction period was found, which was then followed by a fast k_{obs} reaction. A similar trend was seen with tetrafluoroboric acid but with a longer induction time. This could be due to the acids having different pKa values, (TsOH -2.8, HBF_4 -0.4) which could mean the acid would need to form the corresponding anion to form the active catalytic compound. Whilst the induction period changes with the different acids, the k_{obs} value does not change when compared with the standard conditions which indicates that the active catalysts in all these reactions could be the same or have a similar structure.[30] The water content is also an important factor as reactions done using dry conditions had less yield than the reaction using 'wet' solvents.

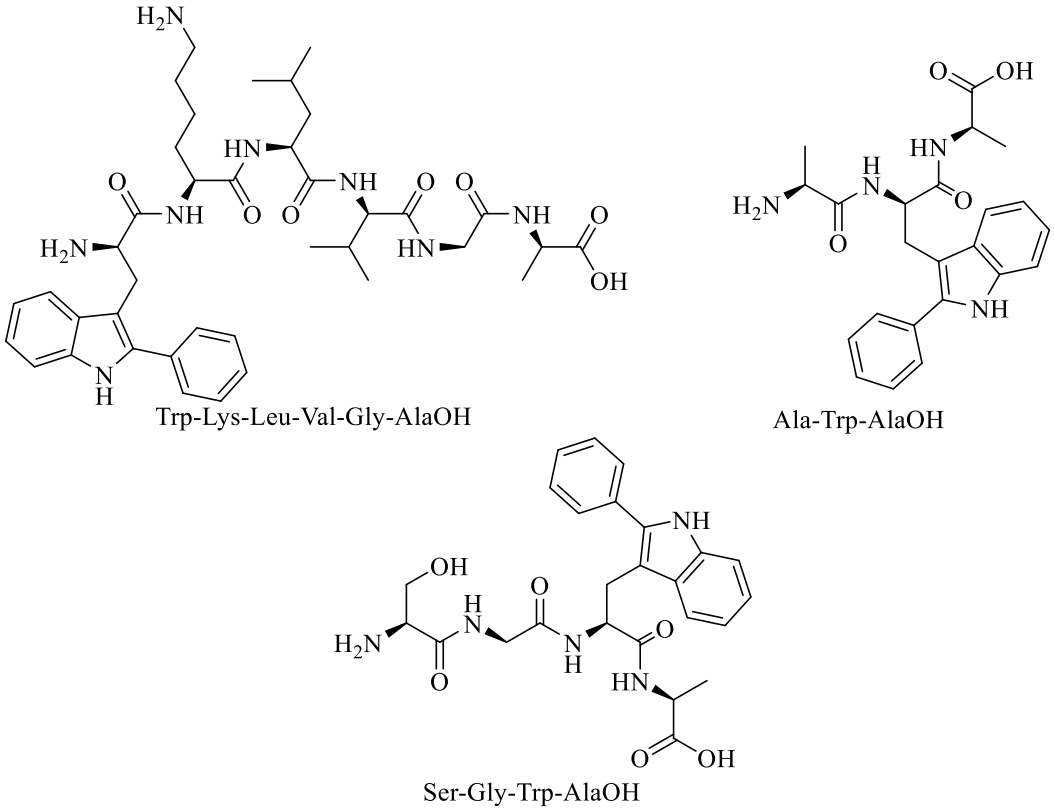


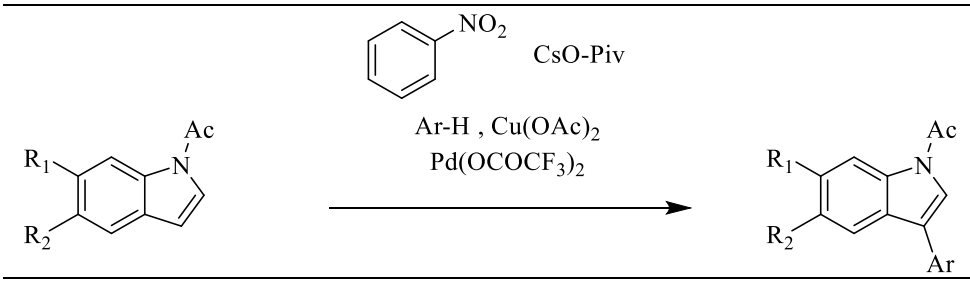
Figure 10: The peptides which were arylated by the Fairlamb group

Work has been done by the Fairlamb group which involves the arylation of peptides through palladium acetate and a copper acetate co-catalyst at ambient temperature in methanol. The peptide chains of different sequence and length, shown in Figure 10, from Trp-Lys-Leu-Val-Gly-Ala-*OH* to AcHN-Ala-Trp-Ala. The reaction of one peptide in particular, Ac-Ser-Gly-Trp-Ala-*OH* produced

the desired arylated product but also similar dihydroxylation byproducts, with a selectivity ratio of 1:1.4. The involvement of the copper(II) for the oxidation process was critical. The experiment was repeated using diaryliodonium salts and in acetic acid with no copper(II) acetate and the desired arylated product was isolated in a yield of 65%, meaning that using a diazonium salt negates the need for copper(II) in this reaction. This also brought about a byproduct of mesitylated product at a ratio of 10:1; this ratio is due the sterically hindered mesityl group. Use of more sterically hindered groups on the iodonium salt resulted in a lower reactivity in the system with a higher ratio of byproduct at 5:1. When using small peptides which contain tryptophan, a protection/deprotection strategy was chosen. The use of the *N*-Ac protecting group previously used on a singular tryptophan molecule proved difficult to synthesise on the peptide chains, so the use of *N*-Boc was evaluated. The *N*-Boc group allowed only a trace amount of desired product to be produced along with unreacted starting material and expected side products from the iodonium salt. The use of *N*-TFA, TFA protected *N* terminus, removed this issue and delivered the desired arylated product in a isolated yield of 82% (with 17% of the mesityl product). This procedure of using aryl iodonium salts without copper acetate was applied to the peptides with terminal alanine residues, which previously afforded dihydroxylation byproducts, now provides the desired arylated product, although the solvent for this reaction was changed from ethyl acetate to isopropanol due to the polar nature of the molecules.

The reactions of indoles *via* this methodology usually give the arylated product at the C-3 position. This is due to the nitrogen in the ring directing the aryl group onto the least hindered location. From work done by Stuart and Fagnou (2007), [31] it can be seen that the groups present on the indole group and on the aryl group effect the conversion and yield, as well as ratio of products made. The protecting groups always include an acetate group on the nitrogen but can be anything from a chlorine atom to an ester group on a carbon on the adjacent benzene ring. The products prepared include arylation on the C-3 position (major product), C-2 position and double arylation (undesired side product). Due to the fact tryptophan has an R group on the indole which allows for easier C-2 arylation, we do not need to concern ourselves with these issues but it should be noted that these protecting groups present in Stuart and Fagnou's paper could bring about positive benefits in similar work in regards to tryptophan.

Table 4:Catalytic Cross Coupling of unactivated arenes[31]



R_1	R_2	Ar	Yield(%)
H	H	Ph	87
OMe	H	Ph	84
H	OMe	Ph	74

Another interesting study on the subject of C-H bond activation of indoles was reported by Gaunt *et al* (2008)[32] which proposed different directing groups for indole arylation, using the directing group pyrimidine, with the only product formed the C-2 arylated product. This procedure by Gaunt however cannot be used in biologically systems due to possible formation of triflic acid from the copper catalyst which is dangerous in itself and then can racemise to increase toxicity. Again, this may not affect most of work being done on tryptophan at the current time, the main aim of Yang's 2016 paper showed how to arylate indoles at the C-7 position and if this was ever used on tryptophan, there could be a generation of C-7 modification of tryptophan.[33]

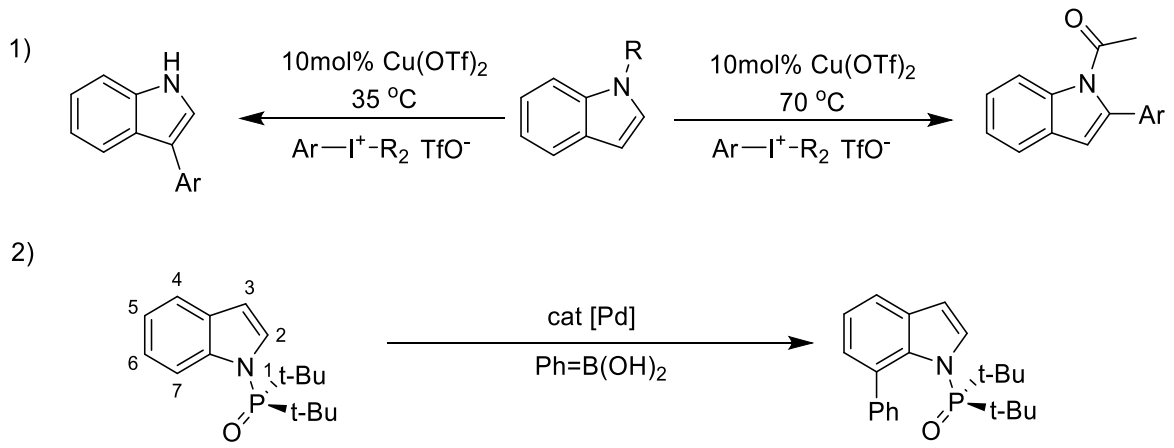


Figure 11: 1) C3 and C2 arylation of unactivated indoles[32] 2) C7 arylation of unactivated indoles[33]

When evaluating catalyst loadings, work reported by Lane and Sames in 2004 showed that the lower the catalyst loading for the $Pd(OAc)_2$, the higher the product yield is, at the expense of biphenyl

formation. This study was done using a high temperature, use of PPh_3, DMA as solvent and CsOAc as an additive. The palladium loadings were from 5mol% to almost trace amounts of catalyst. The loading with the best product formation was 0.5 mol %. The difference between 0.5 and 5 mol % loading in terms of product was about 10-15% while the bi-phenyl formation goes up to about 85%.

1.7: Project Aims

Research in the field of tryptophan arylation has been limited. So far it has been limited to smaller peptides. The aim for this study are therefore as follows:

1) To use the procedure with aryl-diazonium salts as used in the group (Thomas Sheridan's Thesis)[30], to assess if it is possible to modify peptides of different lengths with an aryl groups.
2) To monitor if there is a change in peptide conformation following arylation, this will be done using ROESY NMR spectroscopic analysis.
3) To develop a purification technique for the modified peptide that does not involve prep TLC or a column.
4) Determine whether the kinetics for the peptide arylations are the same as a simple tryptophan derivative.

The work herein attempts to address these aims, to gain an understanding of how this method affects longer-chain peptides.

Chapter 2: Results and Discussion

2.1: Arylation of Tryptophan and showing non-stereospecificity

To synthesise the starting material for the arylation, tryptophan is protected as a methyl ester to give compound 1 and then converted into an aceto-amide to form compound 2. The protection is done to ensure that the C-H activation does not interact with either of the amino groups. This was done in two high-yielding steps, as shown in Figure 12.

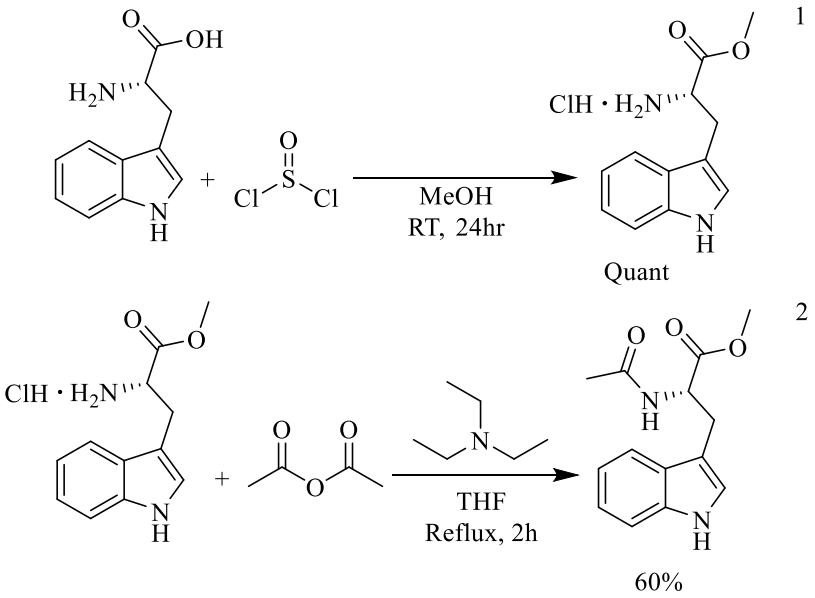


Figure 12: Protection procedure for tryptophan

Once the tryptophan was di-protected, the use of pre-synthesised aryl diazonium salts allowed for the C-2 selective arylation, in the presence of $Pd(OAc)_2$, in ethyl acetate. This was shown to give excellent conversion with all three of the aryl diazonium salts used, as shown in Figure 13.

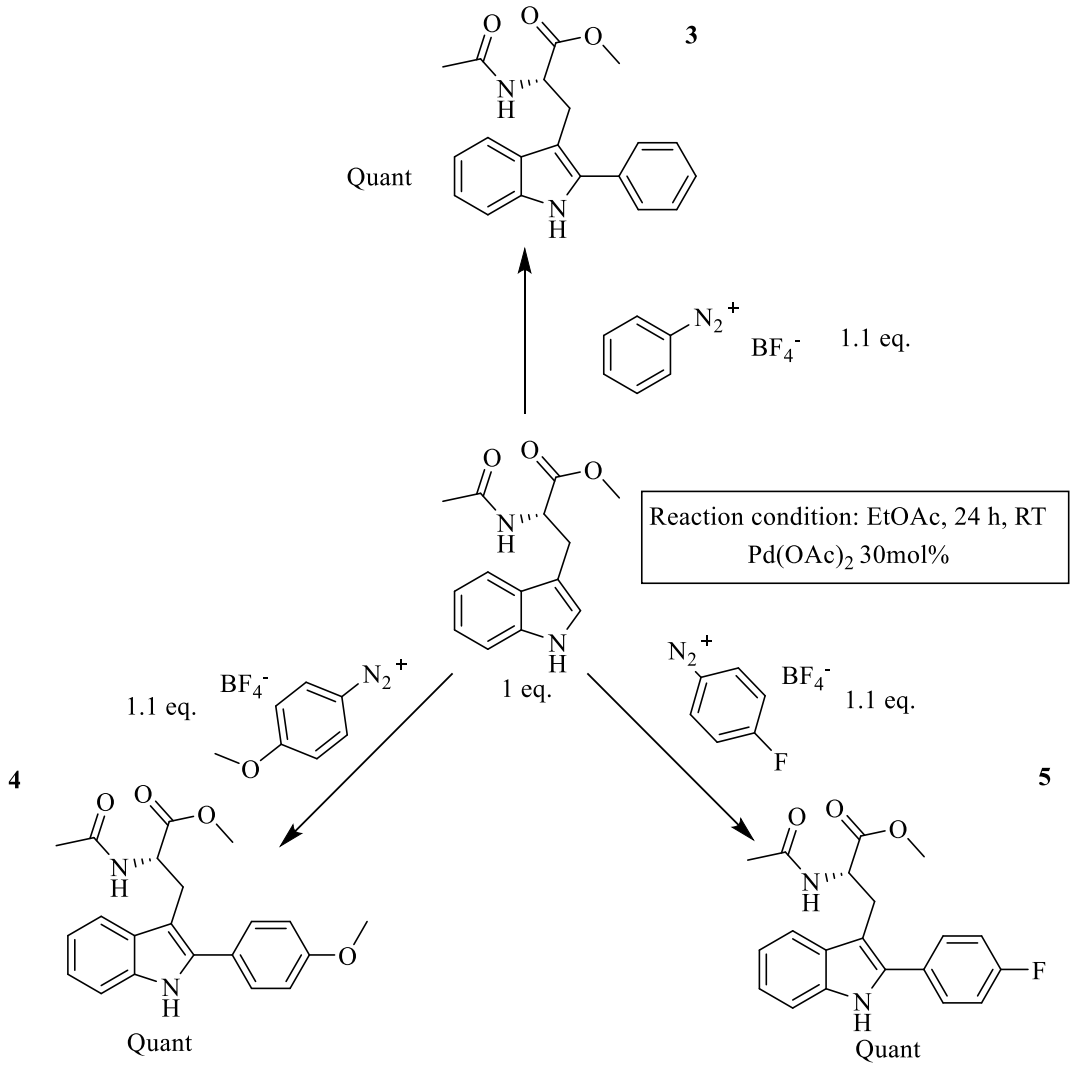


Figure 13: Different arylated products using different aryl diazonium salts

The synthetic methodology was tested using racemic and enantiomerically pure L-tryptophan to assess whether there was an erosion of enantioselectivity. This was tested via the use of chiral HPLC as it can show whether both enantiomers have been modified, and whether racemization has occurred for non-racemic starting material. From the data shown in Figures 14, 15 and 16, both enantiomers were modified for the different aryl groups. From the HPLC traces, it can be seen that both enantiomers of the tryptophan are arylated as the peaks are close together and the L-enantiomer is overlaid to show that retention times match.

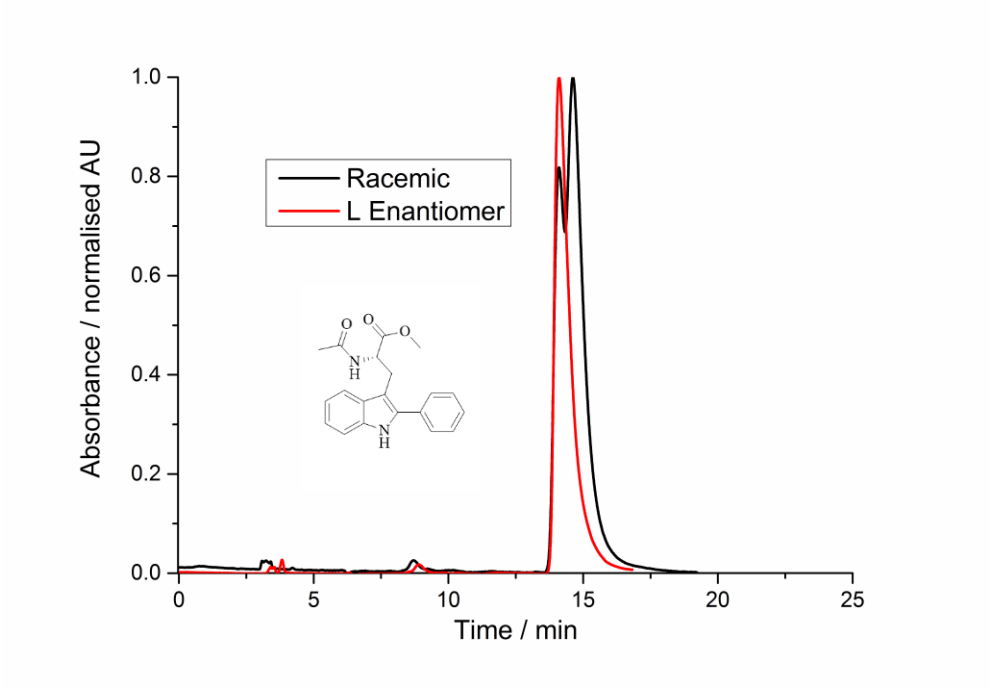


Figure 14: Chiral HPLC data for phenyl products **3**

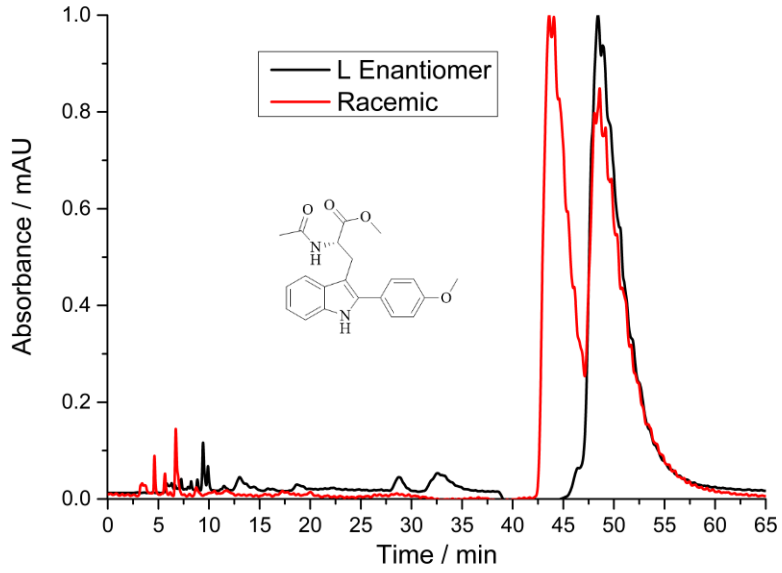


Figure 15: Chiral HPLC data for *para*-methoxy products **4**

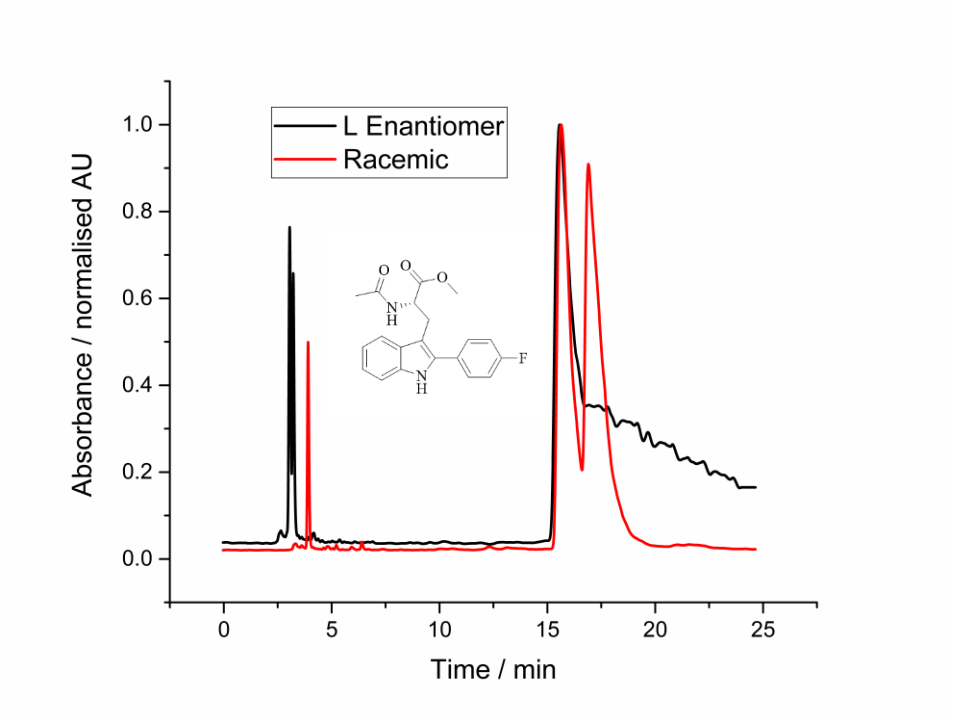


Figure 16: Chiral HPLC data for *para*-fluoro products **6**

2.2: Arylation of peptides

Using the same reagents shown in Figure 13, it was applied to tryptophan containing peptides, **6-9**, shown in Figure 17. These peptides were made using SPPS, a method discussed in the experimental section. The method was modified to run the reactions in methanol due to solubility issues and to increase the Pd catalyst loading to push the conversion higher. The peptides were chosen to contain different groups, specifically tyrosine, lysine and asparagine due to possible interactions with palladium, and to test the functional group tolerance. The peptides have a free carboxylic acid C-terminus to maintain comparability to other reported synthetic methodologies. Esterification to the methyl ester (reaction solvent) is a potential implication in the context of the current arylation technology.

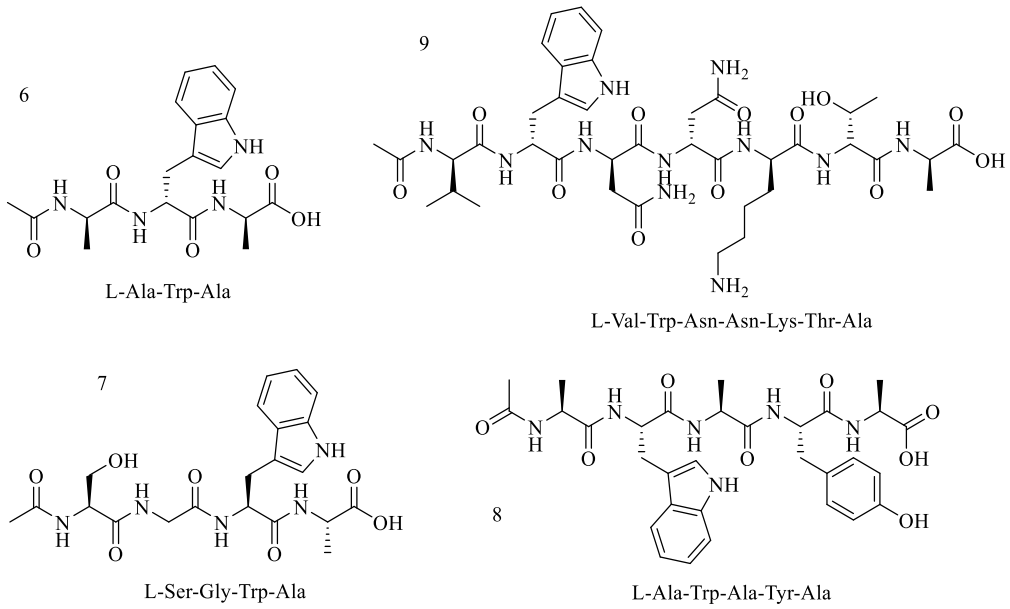


Figure 17: Peptides used in this work

2.3: NMR conformational study of the peptides

The conformation of a protein is an important factor to whether the protein will remain active or not. To assess this, the peptides modified previously were used with the phenyl diazonium salt to assess if there were any conformational changes. To discover the conformation, ROESY NMR spectroscopy was used to see what protons were within 5 angstroms of each other. To confirm the allocations of the atoms in the molecule, COSY, TOCSY, HSQC and HMBC NMR spectroscopy was used. COSY NMR spectroscopy shows interactions between adjacent protons, whilst TOCSY shows interactions with all the protons on consecutive atoms, as long as the proton chain is not broken the interactions will continue throughout the entire molecule. HSQC NMR spectroscopy shows which carbon peak corresponds to which proton peak by correlating a ^{1}H spectrum with a ^{13}C spectrum. HMBC NMR spectroscopy shows interactions between adjacent protons within a few bonds on carbon atoms.

The unmodified peptides were analysed first to have something to compare against. The tri-peptide, AcNH-Ala-Trp-Ala-OH was the first peptide looked at due to simplicity of the molecule. Using other NMR techniques such as COSY, TOCSY, HSQC, HMBC as well as the standard one-dimensional proton and carbon, each atom in the molecule is assigned and shown below.

AcNH-Ala-Trp-Ala-OH (6)

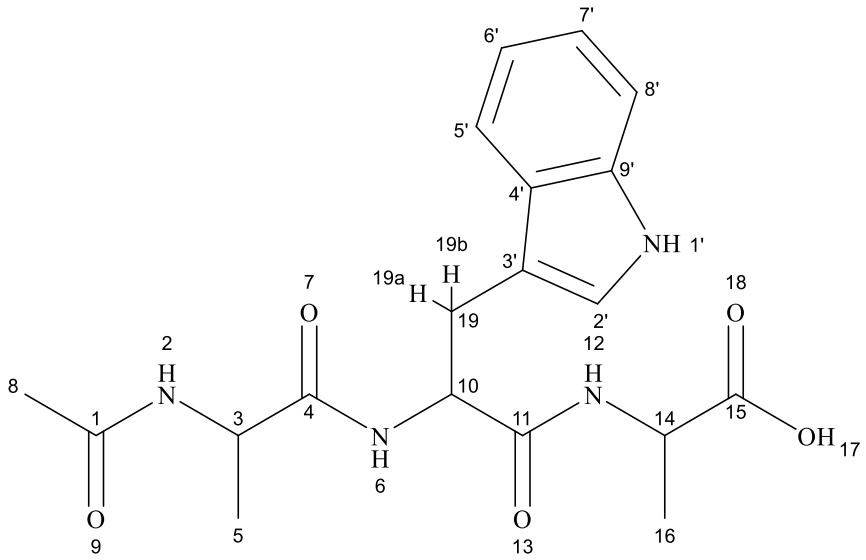


Figure 18: Molecule of *AcNH*-Ala-Trp-Ala-*OH* (**6**)

NMR spectrum data for **6**

^{1}H NMR (700 MHz, DMSO) δ 10.81 (s, 1H), 8.14 (d, *J* = 7.2 Hz, 1H), 7.98 (d, *J* = 7.2 Hz, 1H), 7.84 (d, *J* = 8.2 Hz, 1H), 7.58 (d, *J* = 7.9 Hz, 1H), 7.30 (d, *J* = 8.1 Hz, 1H), 7.13 (d, *J* = 2.1 Hz, 1H), 7.04 (t, *J* = 7.5 Hz, 1H), 6.96 (t, *J* = 7.3 Hz, 1H), 4.51 (td, *J* = 8.6, 4.4 Hz, 1H), 4.24 – 4.15 (m, 2H), 3.15 (dd, *J* = 14.8, 4.3 Hz, 1H), 2.95 (dd, *J* = 14.9, 9.0 Hz, 1H), 1.78 (s, 3H), 1.26 (d, *J* = 7.3 Hz, 3H), 1.10 (d, *J* = 7.1 Hz, 3H).

^{13}C NMR (176 MHz, DMSO) δ 174.21, 172.37, 171.36, 169.44, 136.23, 127.64, 123.90, 121.02, 118.67, 118.39, 111.44, 110.12, 53.13, 48.52, 47.79, 27.61, 22.69, 18.11, 17.39.

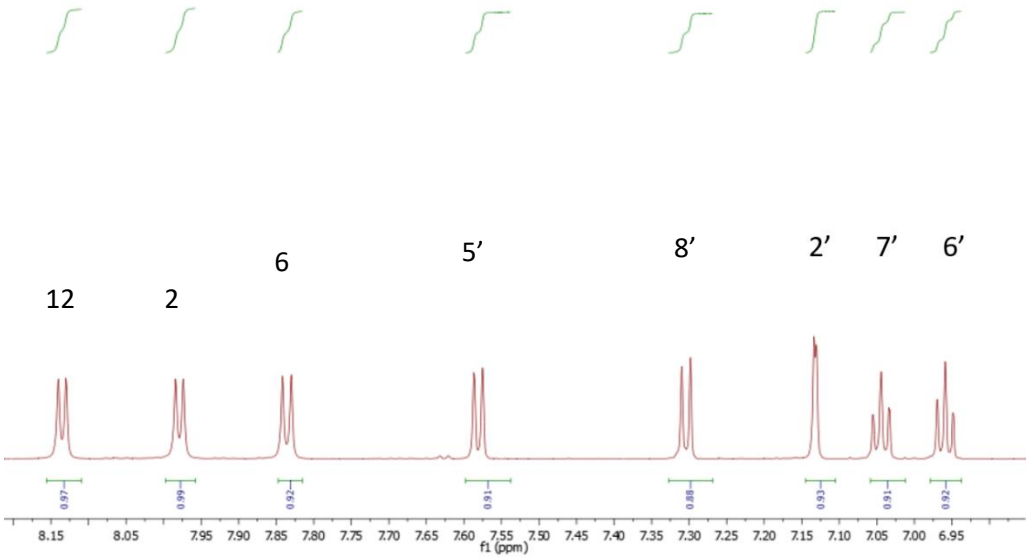


Figure 19: ^{1}H NMR spectrum of **6,** showing ^{1}H assignments between δ 6.95 – 8.15 ppm (DMSO, 700 MHz, 300 K)

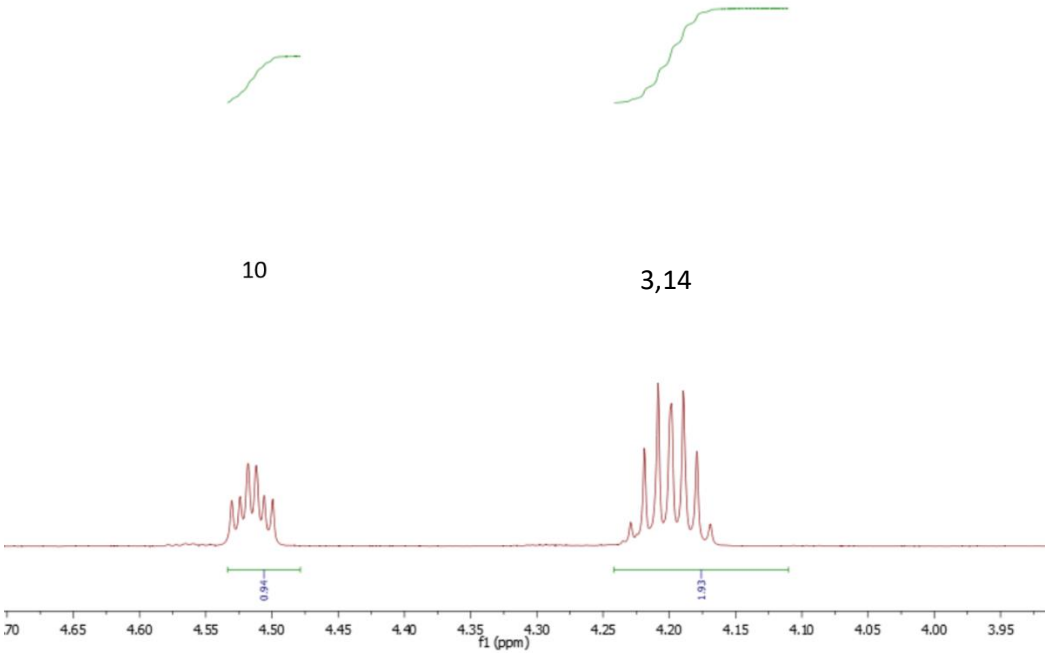


Figure 20: ^{1}H NMR spectrum of 6, ^{1}H assignments between δ 4.10 – 4.65 ppm (DMSO, 700 MHz, 300 K)

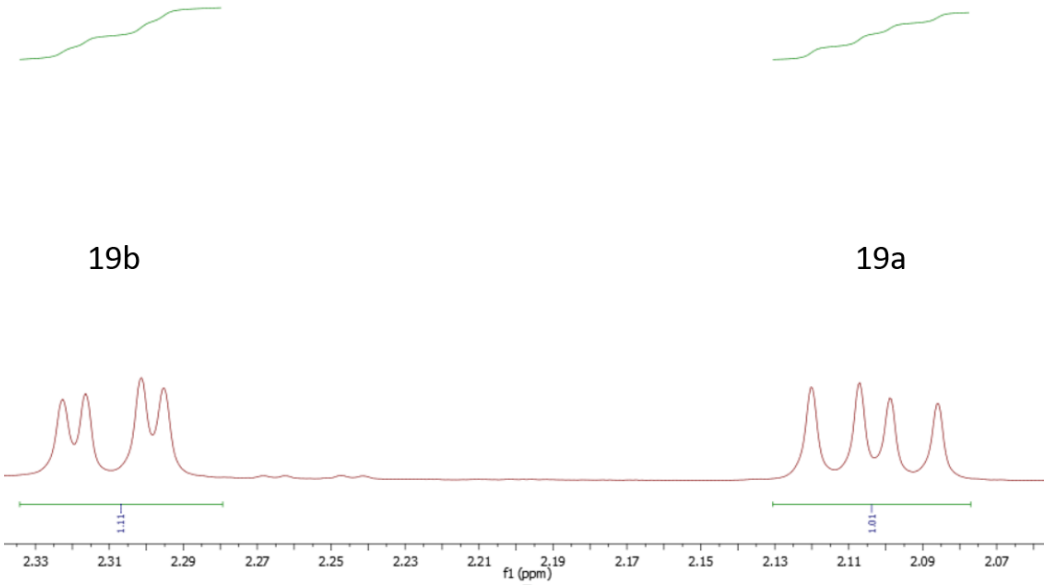


Figure 21: ^{1}H NMR spectrum of **6,** ^{1}H assignments between δ 2.07 – 2.33 ppm (DMSO, 700 MHz, 300 K)

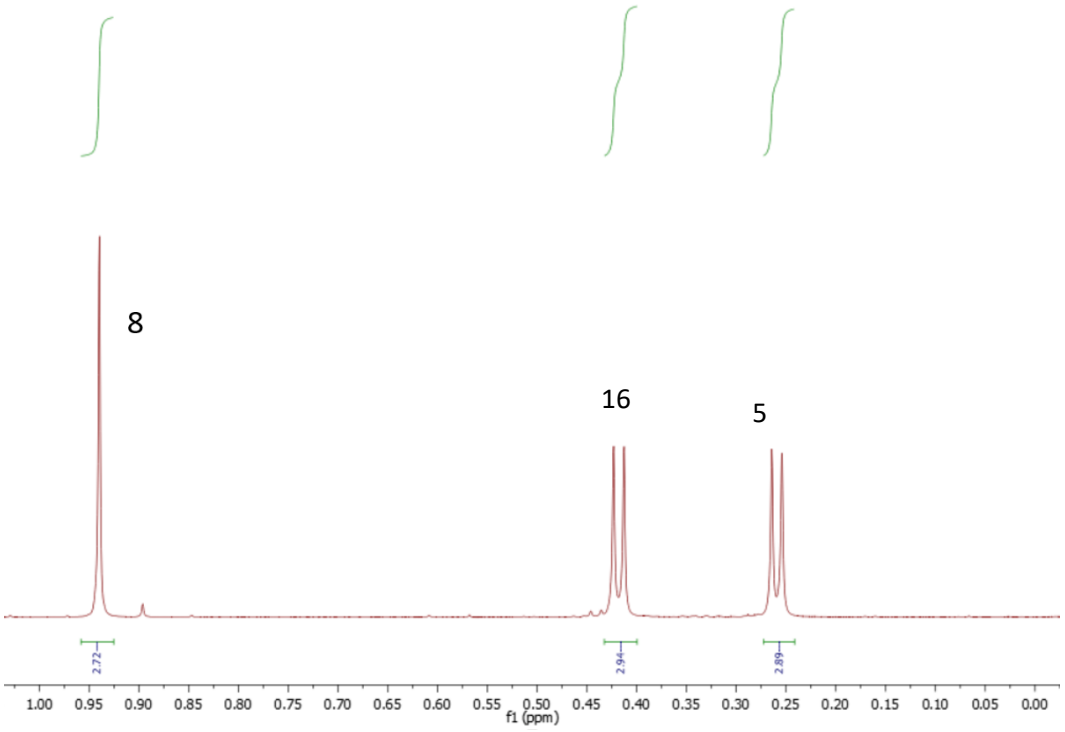


Figure 22: ^{1}H NMR spectrum of **6**, ^{1}H assignments between δ 0.25 – 0.95 ppm (DMSO, 700 MHz, 300 K)

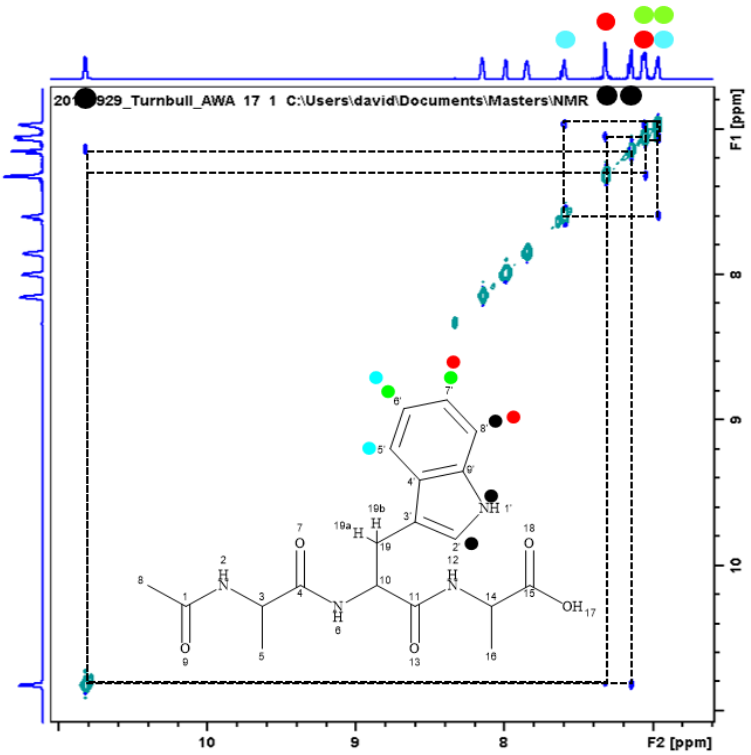


Figure 23: ROESY Spectrum of aromatic region of **6** – examination of the indole region (DMSO, 700 MHz, 300 K)

From the ROESY spectrum, the indole nitrogen proton is close in proximity to the C-2' proton, which is to be expected, but also close to the C-8' proton, which is interesting due to it being on the a few bonds away. It can also be seen that the remaining aromatic protons are close to the adjacent proton.

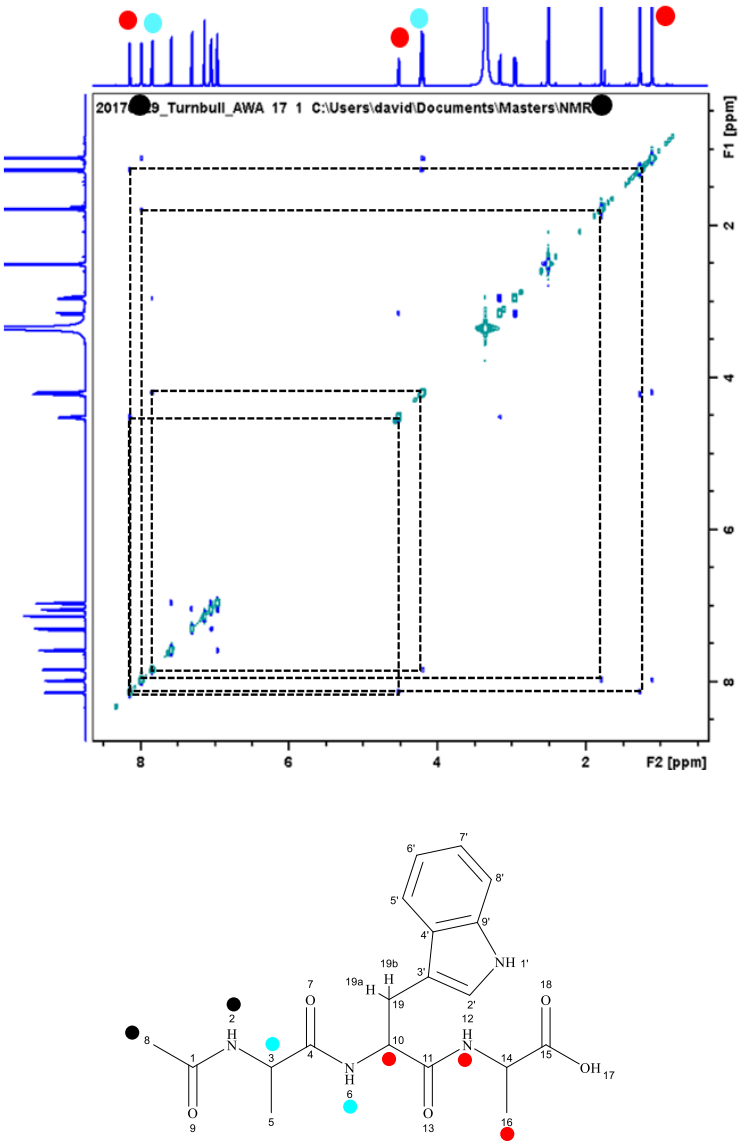


Figure 24: ROESY Spectrum of **6** – examination of the peptide chain region (DMSO, 700 MHz, 300 K)

Looking at the spectra above, proton 2 on the *N* terminus alanine is close to the methyl group on atom 8. This can be explained by rotation around C1 and C8 allowing the methyl group to be closer to the amide proton. The amide proton on the tryptophan residue, atom 6, is close to the protons on either atom 3 or 14. Atom 3 or 14 would be interesting as it is perceived to be further away from the proton on, it is more likely to be 3 due to the distance away from position 14. The amide proton on atom 12 can be seen to be close to atom 14 which is expected and atom 10 which is somewhat unexpected but could be due to the atom at 11 rotating.

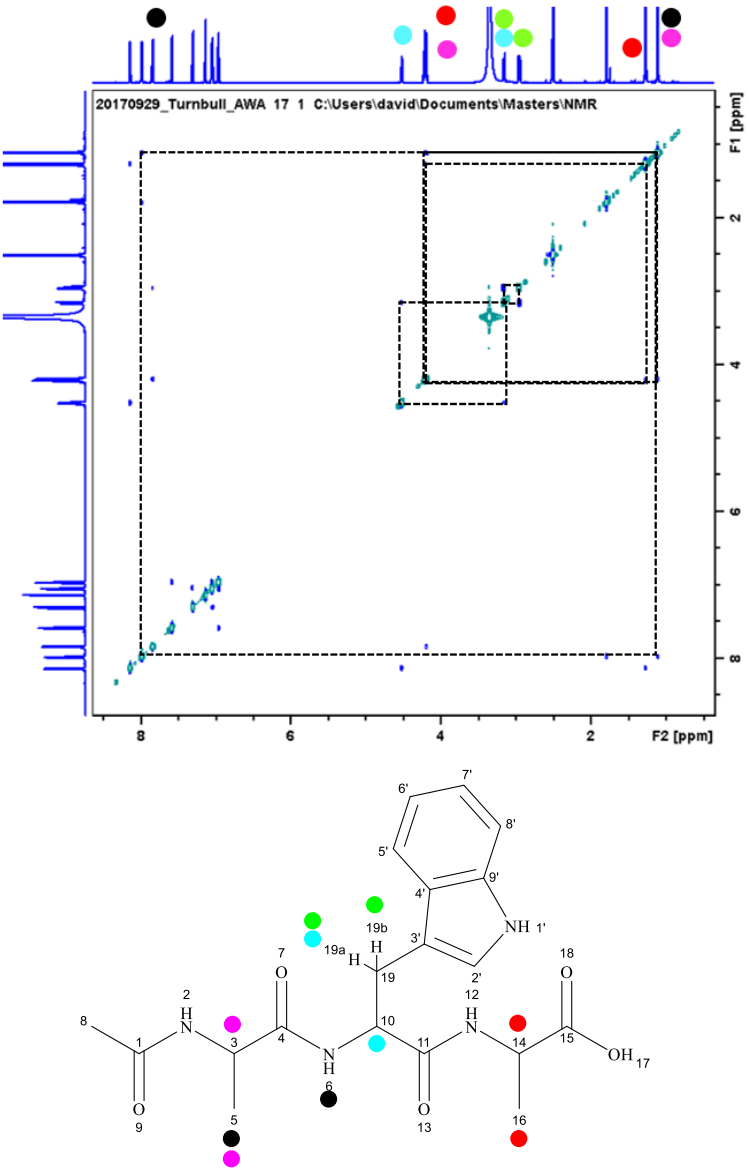


Figure 25: ROESY Spectrum of aliphatic region of **6** – examination of the peptide backbone (DMSO, 700 MHz, 300 K)

From the above spectra, the amide proton on atom 2 is close in proximity to atom 5, this requires that the molecule is somewhat contorted due to no interaction between the proton on atom 3, which is adjacent to the amide at position 2. The two diastereomeric protons on the β-carbon on the tryptophan residue are coupled together. The proton on atom 10, the α-carbon on the tryptophan residue is close in proximity to one of the diastereomeric protons but not the other. There is also coupling between the protons on atoms 3 to 5 and 14 to 16, this was to be expected for adjacent atoms.

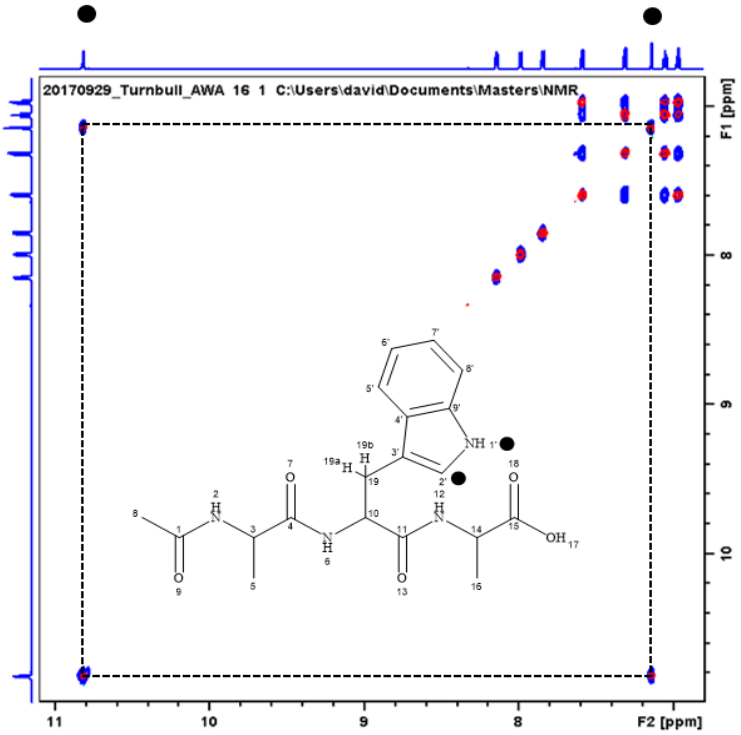


Figure 26: TOCSY (Blue) and COSY (Red) spectrum of aromatic region of **6** – examination of nitrogen ring of the indole (DMSO, 700 MHz, 300 K)

In the figure above, an interaction can be seen between the 1’ position proton and 2’ position proton. This is to be expected from adjacent protons.

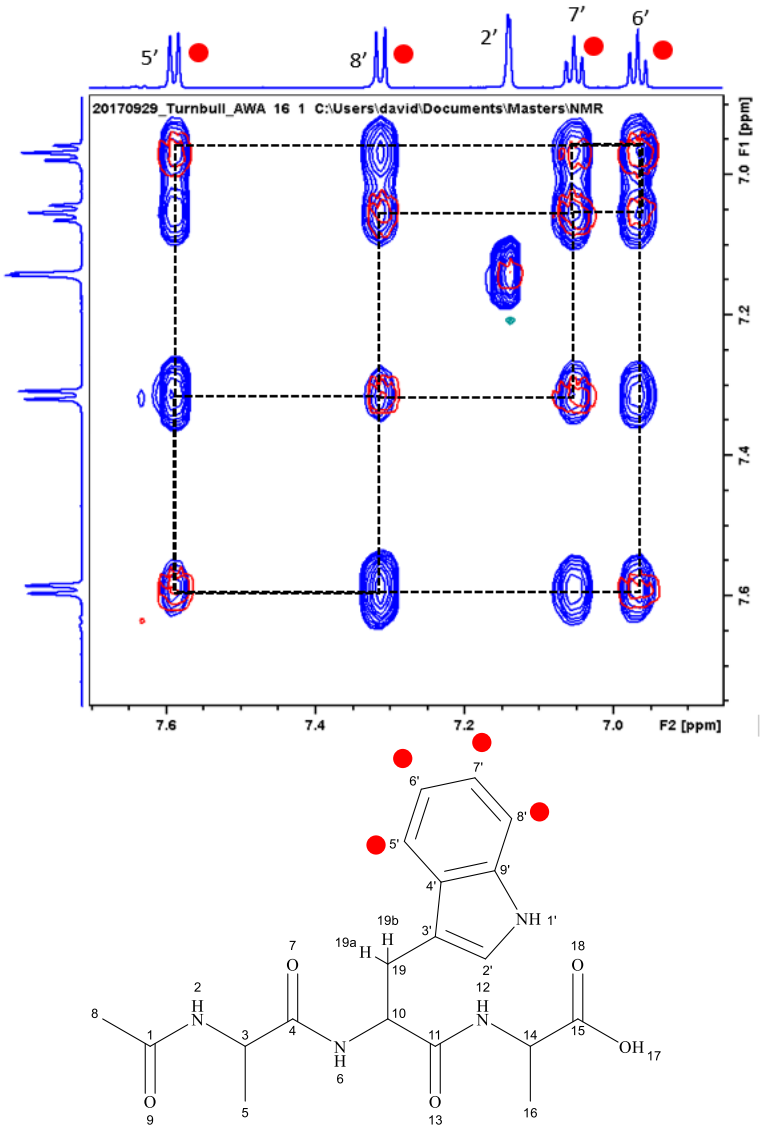


Figure 27: TOCSY (Blue) and COSY (Red) spectrum of aromatic region of **6** – examination of the remaining indole group (DMSO, 700 MHz, 300 K)

The NMR spectrum in Figure 27 contains the overlay of the COSY over the TOCSY. The TOCSY shows that the aromatic protons are all in series with no breaks in the proton chain. The COSY allows for the comparison of adjacent protons with the protons connected in the series.

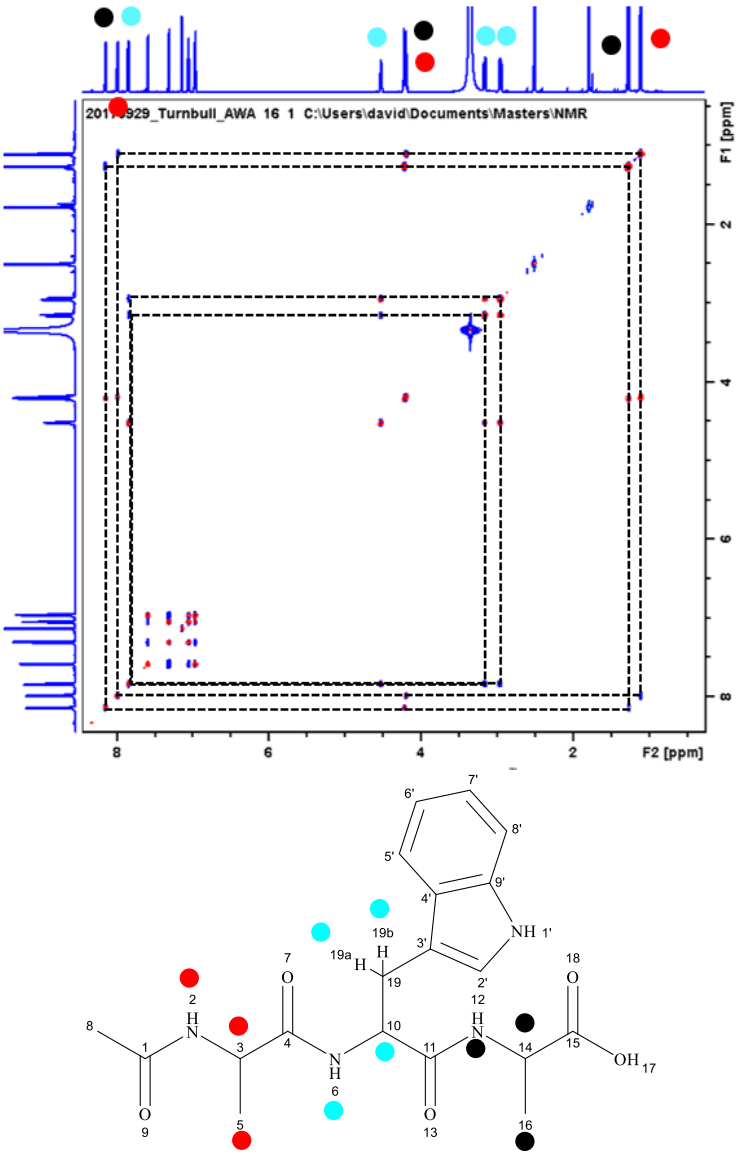


Figure 28: TOCSY (Blue) and COSY (Red) spectrum of **6** – examination of the peptide backbone (DMSO, 700 MHz, 300 K)

The use of the TOCSY and COSY shows the protons in series in certain parts of the molecule. It can be seen that the amide proton on position 6 correlates with the α-carbon proton at position 10 and then to the 2 diastereomeric protons on the β-carbon at position 19. Similarly, the series can be seen for the amide protons at position 2 and 12 and that the series continues to 5 and 16 respectively.

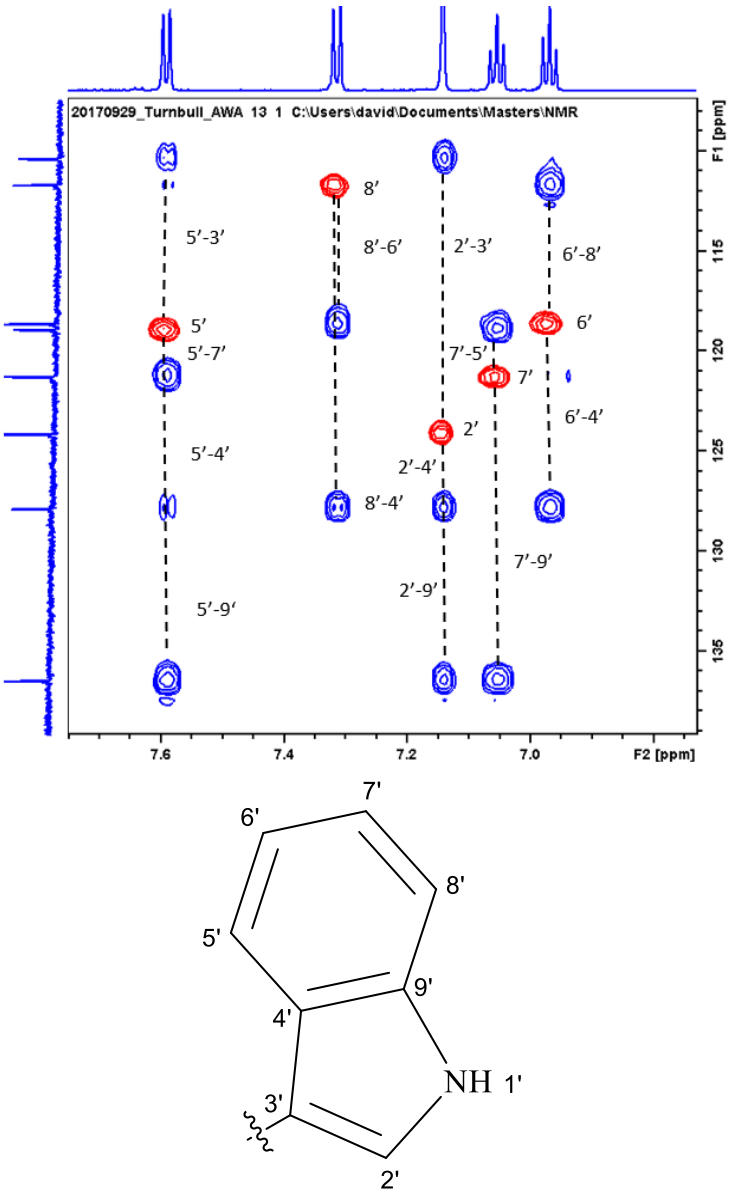


Figure 29: HSQC and HMBC spectrum of the aromatic region of 6 - examination of the indole ring (DMSO, for 1H -700 MHz, for ^{13}C - 176 MHz, 300 K)

By using the HMBC and ^{13}C HSQC, the links to nearby carbon atoms can be determined. The connections show interactions such as position 5’ on the aromatic ring interacts with positions 3’,7’,4’ and 9’ on the aromatic ring. The proton at position 2’ shows that the interactions can go past heteroatoms such as nitrogen.

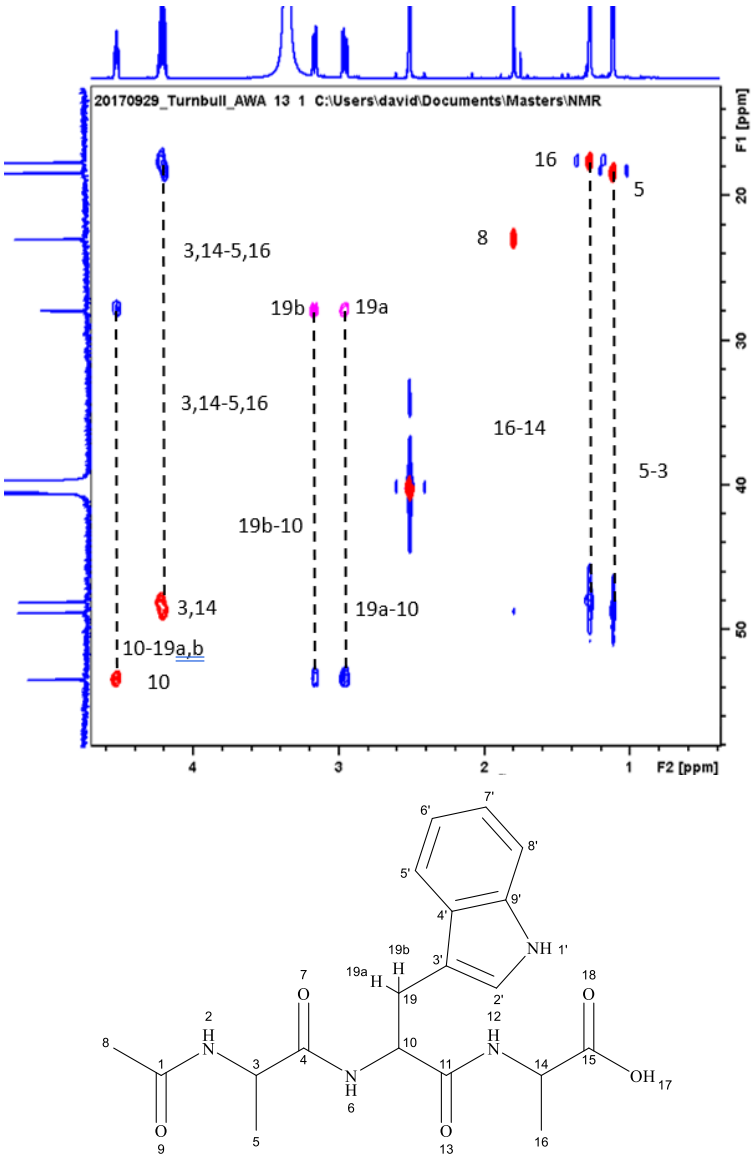


Figure 30: HSQC and HMBC spectrum of the aliphatic region of **6** - examination of the peptide backbone (DMSO, for ^{1}H -700 MHz, for ^{13}C - 176 MHz, 300 K)

The use of the HMQC and HSBC show distinct differences between diastereomeric protons and the α-carbon at position 10. Through this, it can be seen that each proton interacts with the α-carbon on the tryptophan residue.

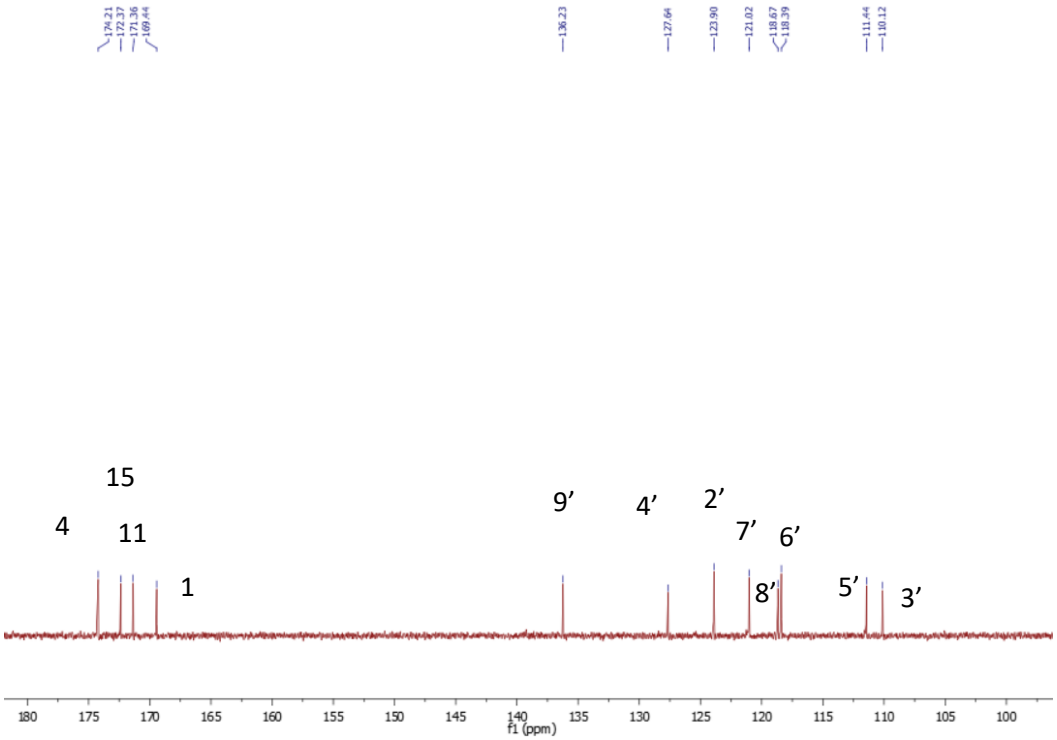


Figure 31: ^{13}C NMR spectrum of **6**, ^{13}C assignments between δ 110.0 – 180.0 ppm (DMSO, 176 MHz, 300K)

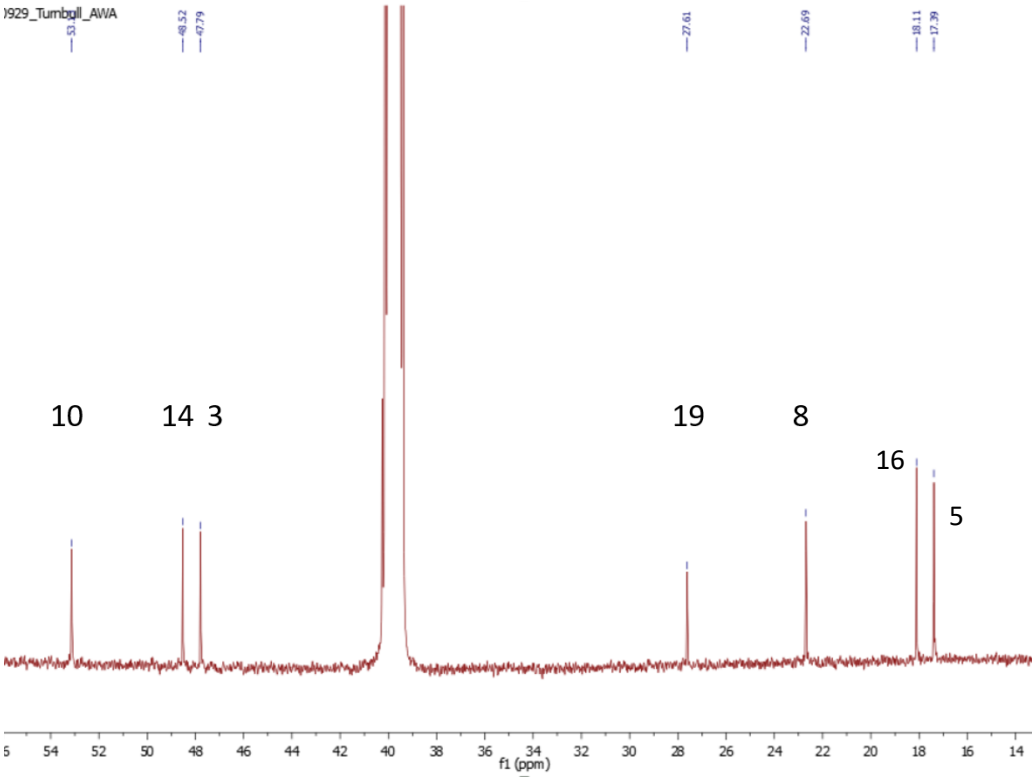


Figure 32: ^{13}C NMR spectrum of **6**, ^{13}C assignments between δ 15.0 – 54.0 ppm (DMSO, 176 MHz, 300K)

AcNH-Ala-Trp-Ala-Tyr-Ala-OH (8)

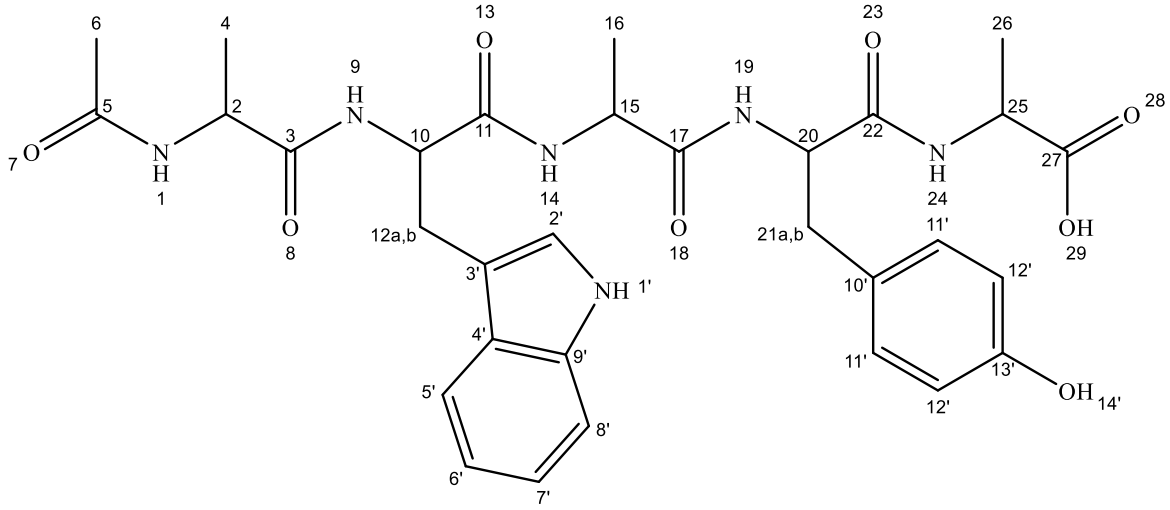


Figure 33: Reference molecule for Ala-Trp-Ala-Tyr-Ala (**8**)

NMR spectrum data for **8**

^{1}H NMR (700 MHz, DMSO) δ 10.80 (s, 1H), 9.12 (s, 1H), 8.13 (d, *J* = 7.1 Hz, 1H), 8.01 (d, *J* = 7.0 Hz, 1H), 7.89 (d, *J* = 7.4 Hz, 2H), 7.76 (d, *J* = 8.2 Hz, 1H), 7.55 (d, *J* = 7.9 Hz, 1H), 7.30 (d, *J* = 8.1 Hz, 1H), 7.12 (s, 1H), 7.03 (dd, *J* = 11.5, 8.0 Hz, 3H), 6.92 (t, *J* = 7.4 Hz, 1H), 6.62 (d, *J* = 8.3 Hz, 2H), 4.46 (td, *J* = 8.4, 4.5 Hz, 1H), 4.40 (td, *J* = 8.7, 4.5 Hz, 1H), 4.18 (dp, *J* = 14.1, 7.0 Hz, 3H), 3.12 (dd, *J* = 14.9, 4.0 Hz, 1H), 2.98 – 2.91 (m, 2H), 2.68 (dd, *J* = 14.0, 9.3 Hz, 1H), 1.91 (s, 1H), 1.78 (s, 3H), 1.27 (t, *J* = 9.0 Hz, 4H), 1.11 (t, *J* = 6.9 Hz, 7H).

^{13}C NMR (176 MHz, DMSO) δ 173.93, 172.46, 172.04, 171.79, 171.14, 170.71, 169.27, 155.73, 136.00, 130.17, 127.72, 127.26, 123.56, 120.75, 118.43, 118.14, 114.81, 111.12, 109.81, 53.84, 53.20, 48.45, 48.33, 47.45, 36.50, 27.13, 22.42, 21.05, 18.00, 17.79, 17.15.

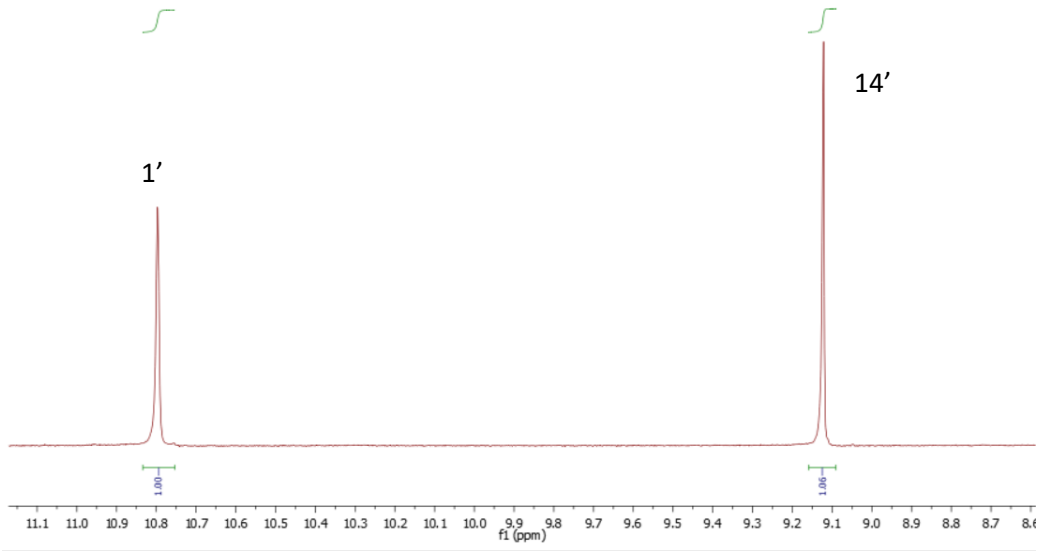


Figure 34: ^{1}H NMR spectrum of **8,** showing ^{1}H assignments between δ 9.1 – 10.8 ppm (DMSO, 700 MHz, 300 K)

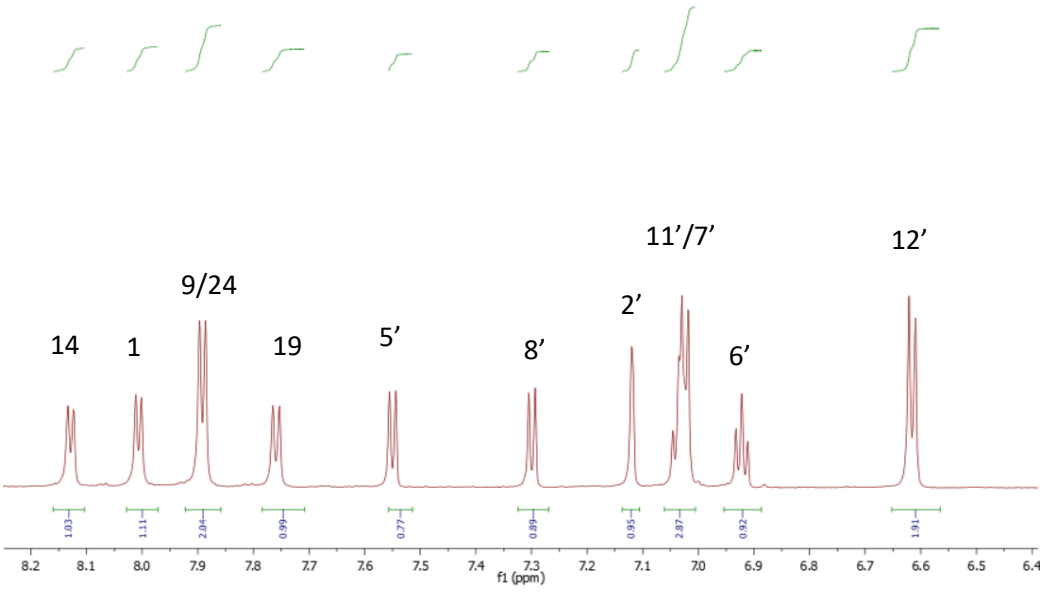


Figure 35: ^{1}H NMR spectrum of **8,** showing ^{1}H assignments between δ 6.6 – 8.15 ppm (DMSO, 700 MHz, 300 K)

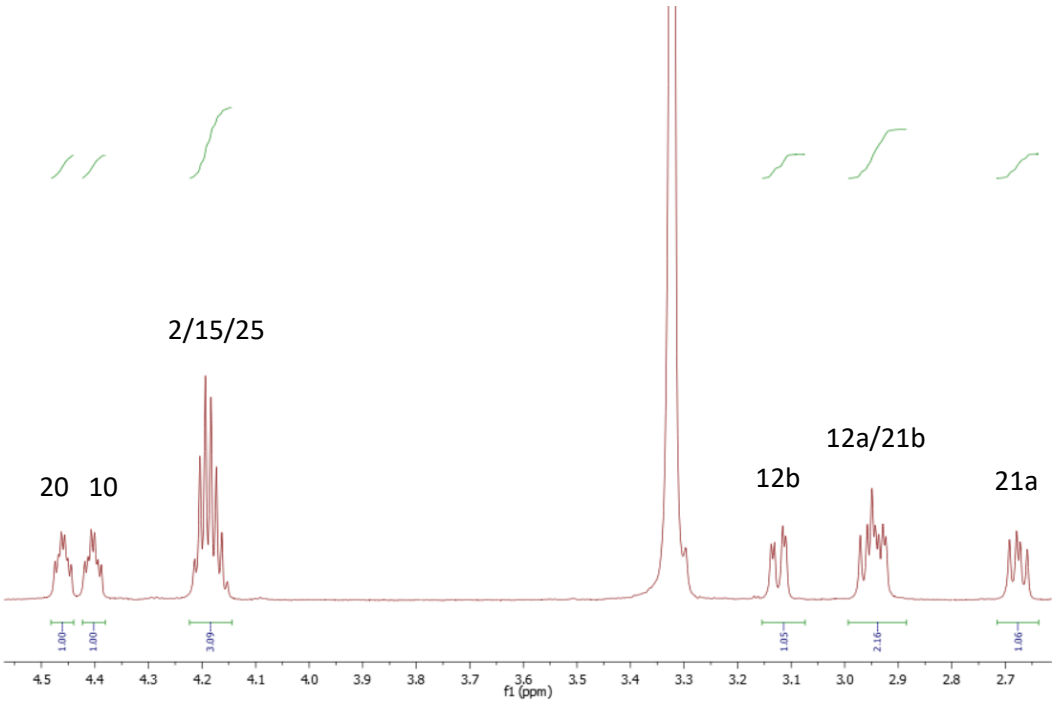


Figure 36: ^{1}H NMR spectrum of **8,** showing ^{1}H assignments between δ 2.6 – 4.5 ppm (DMSO, 700 MHz, 300 K)

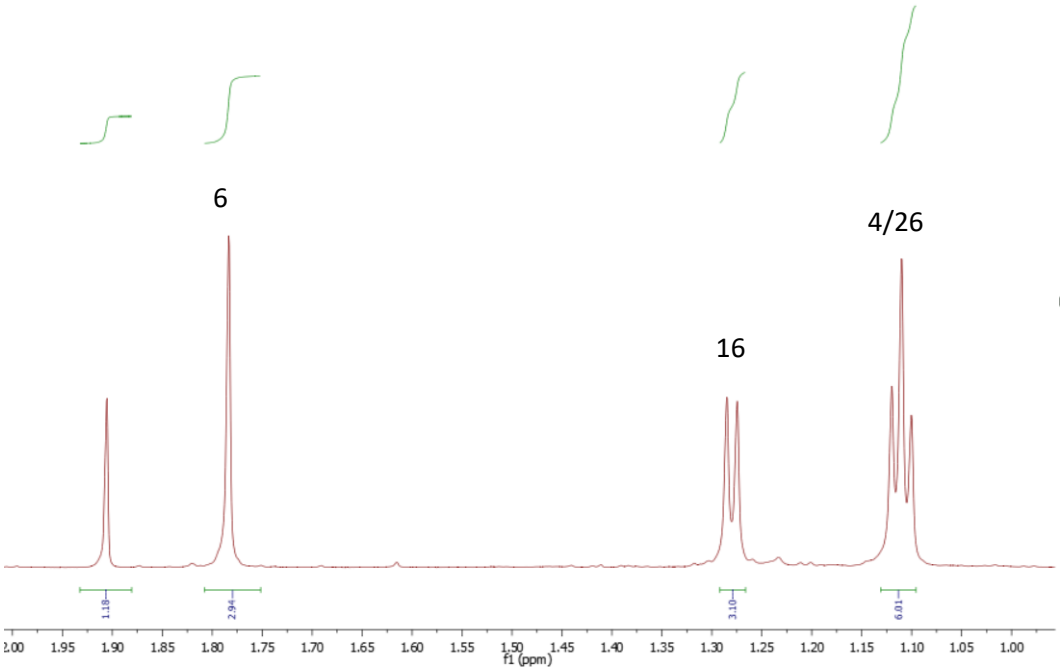


Figure 37: ^{1}H NMR spectrum of **8,** showing ^{1}H assignments between δ 1.00 – 1.95 ppm (DMSO, 700 MHz, 300 K)

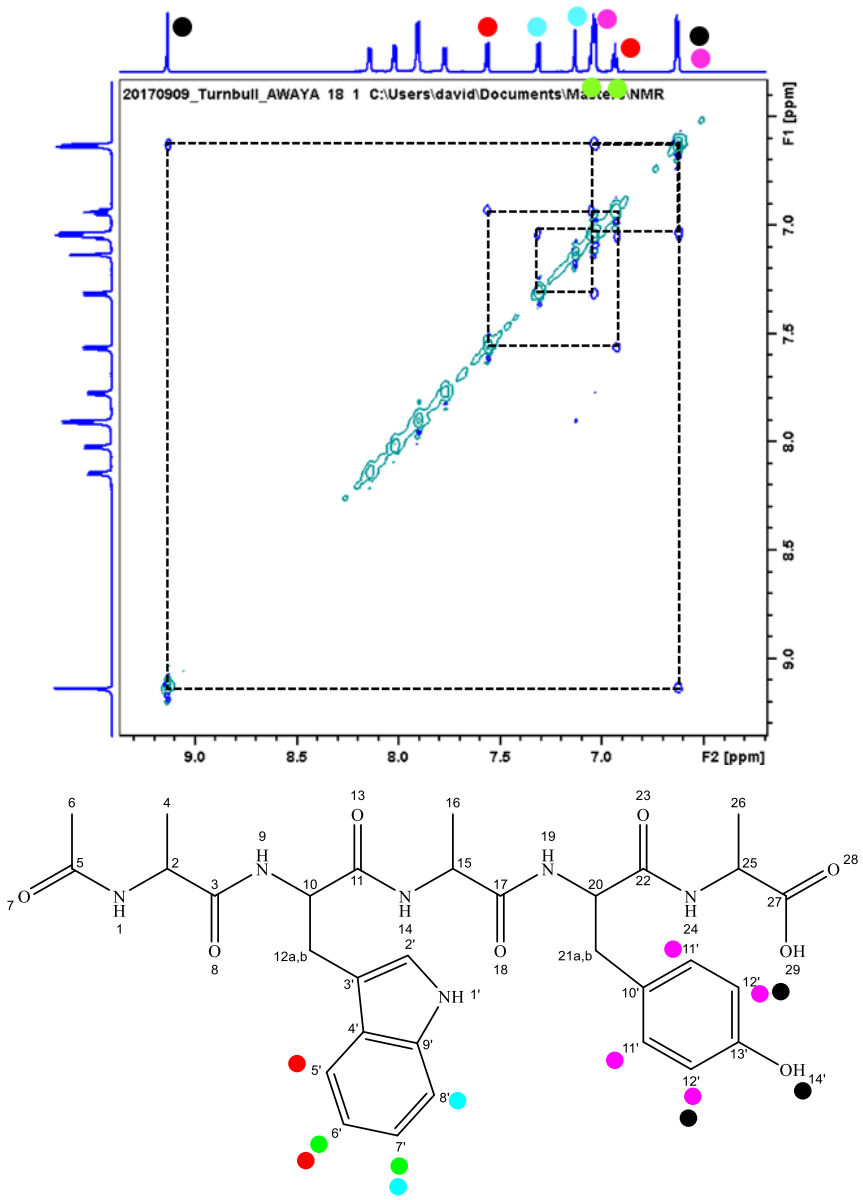


Figure 38: ROESY Spectrum of aromatic region of **8** – examination of the indole region (DMSO, 700 MHz, 300 K)

Looking at the ROESY spectrum in Figure 38, the interactions between the aromatic protons become clear. In the indole, there are interactions between position 6' and 7', to be expected. As are the interactions between 5' and 6', and 7' and 8'. In the tyrosine ring, interactions with the protons in positions 11' and 12' are seen, as well as an interaction between 12' and 14'. The latter interaction is more interesting due to the oxygen present in the alcohol group. The bent angle from the lone pairs on the oxygen could have meant that there were no interactions with the other protons or interactions with just one.

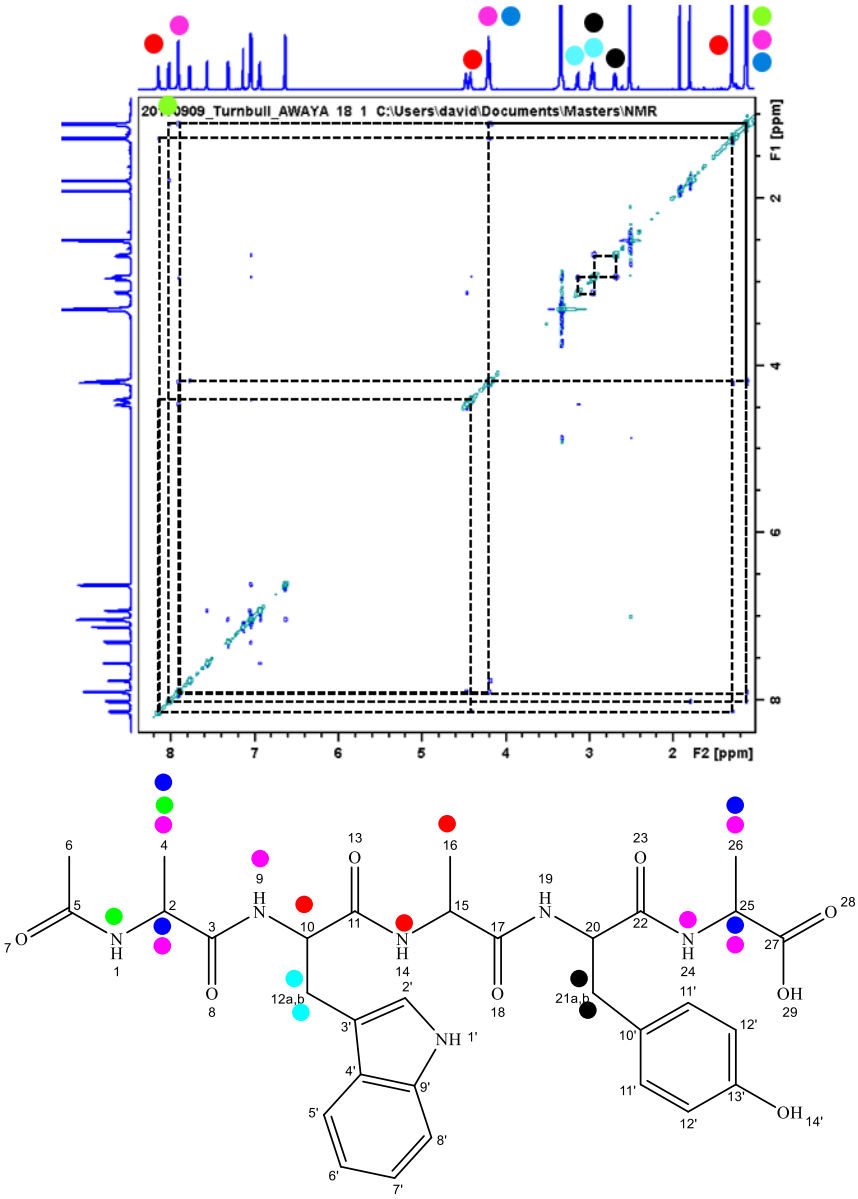


Figure 39: ROESY Spectrum of aromatic region of **8** – examination of the peptide chain region (DMSO, 700 MHz, 300 K)

The spectrum above shows interactions with the amide proton at position 14 with the methyl protons on the β-carbon at position 16, which is expected. There is an additional interaction between 14 and 10 which is slightly unusual due to the perceived distance but if the bond between 10 and 11 rotates, then 14 is near 10. The amide protons at positions 9 and 1 have interactions with positions 4 and 26 which are several bonds away, this is an indicator that the molecule is bending and allowing for unusual interactions. There are interactions between positions 2, 15,25 to 4, 16 and 26 respectively. These interactions are expected. There are 2 sets of diastereomeric protons in this molecule, one being on the tryptophan β-carbon and the other of the tyrosine β-carbon, they interact internally.

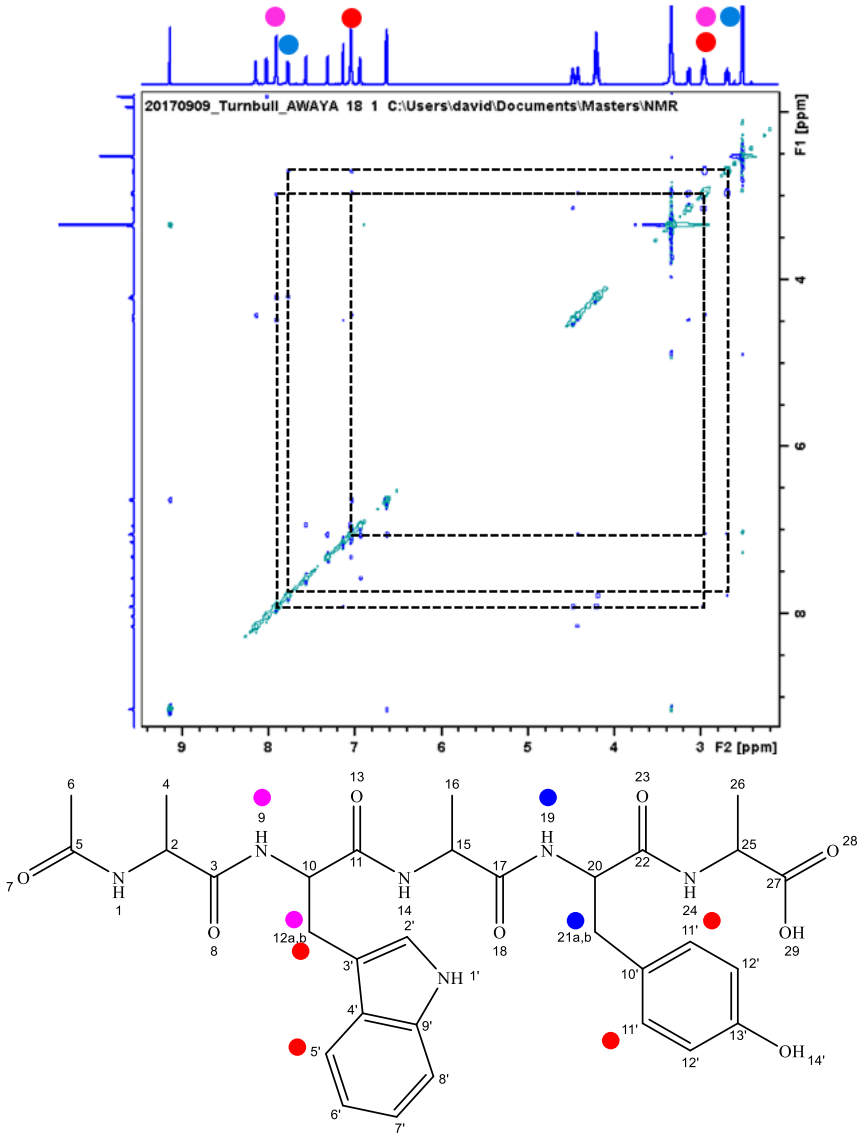


Figure 40: ROESY Spectrum of aromatic region of **8** – examination of the peptide backbone (DMSO, 700 MHz, 300 K)

From the spectrum above, interactions can be seen between either the 5'-position proton on the indole aromatic ring or the 11' position on the tyrosine aromatic ring with one of the diastereomeric protons. This is quite unusual due to the position of both of these protons in relation to the β-carbon proton on the tryptophan. Due to the sheer size of the indole ring, it would be predicted that the interaction is with the 11' position proton on the tyrosine indicating the molecule bending. There are interactions between the amide protons at position 9 and 19 with one of the diastereomeric protons on the nearby β-carbon protons, positions 12 and 21 respectively. Rotation of the protons around the nitrogen at position 9 explains the interactions with only one of the diastereomeric protons.

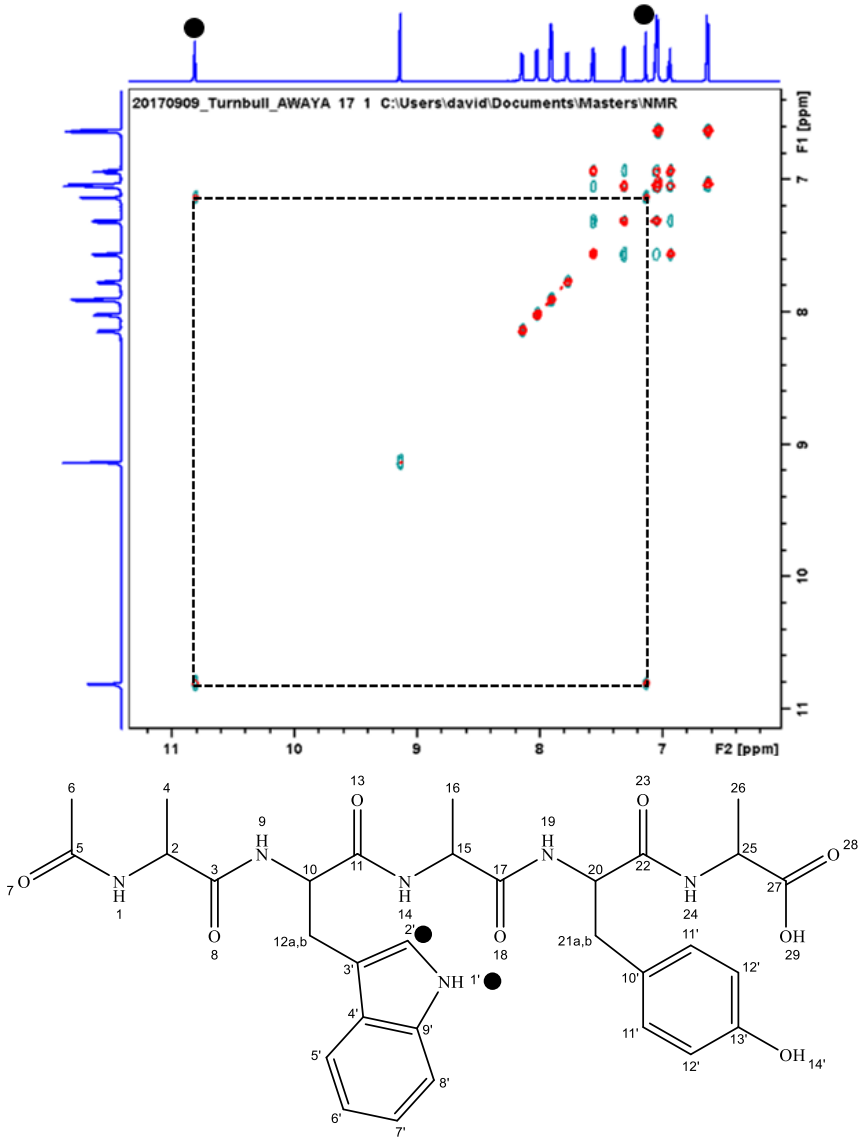


Figure 41: TOCSY (Blue) and COSY (Red) spectrum of **8** – examination of the indole ring (DMSO, 700 MHz, 300 K)

From the spectrum above, it can be seen that there is an interaction between protons 1' and 2' which is to be expected.

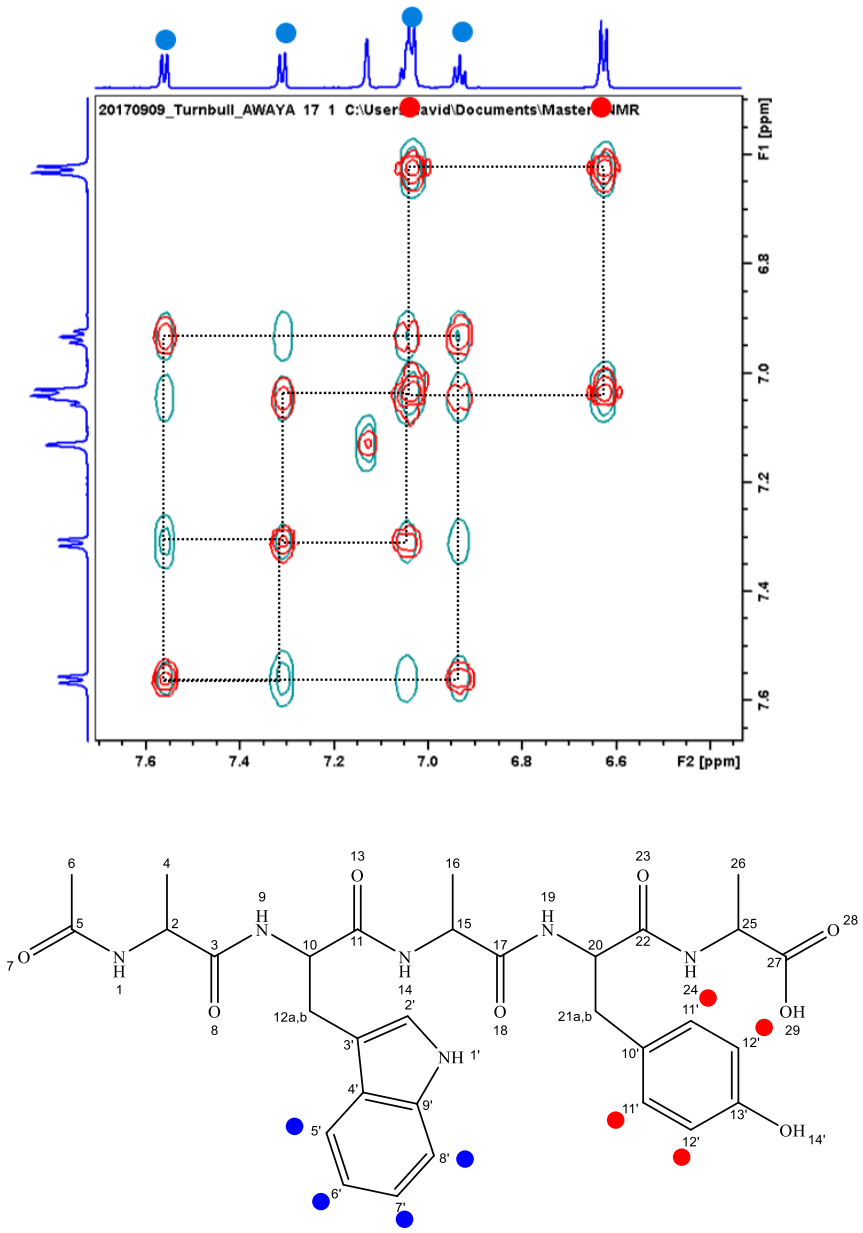


Figure 42: TOCSY (Blue) and COSY (Red) spectrum of **8** – examination of the aromatic rings (DMSO, 700 MHz, 300 K)

Through the TOCSY and COSY, it can be seen that there are 2 different aromatic rings present.

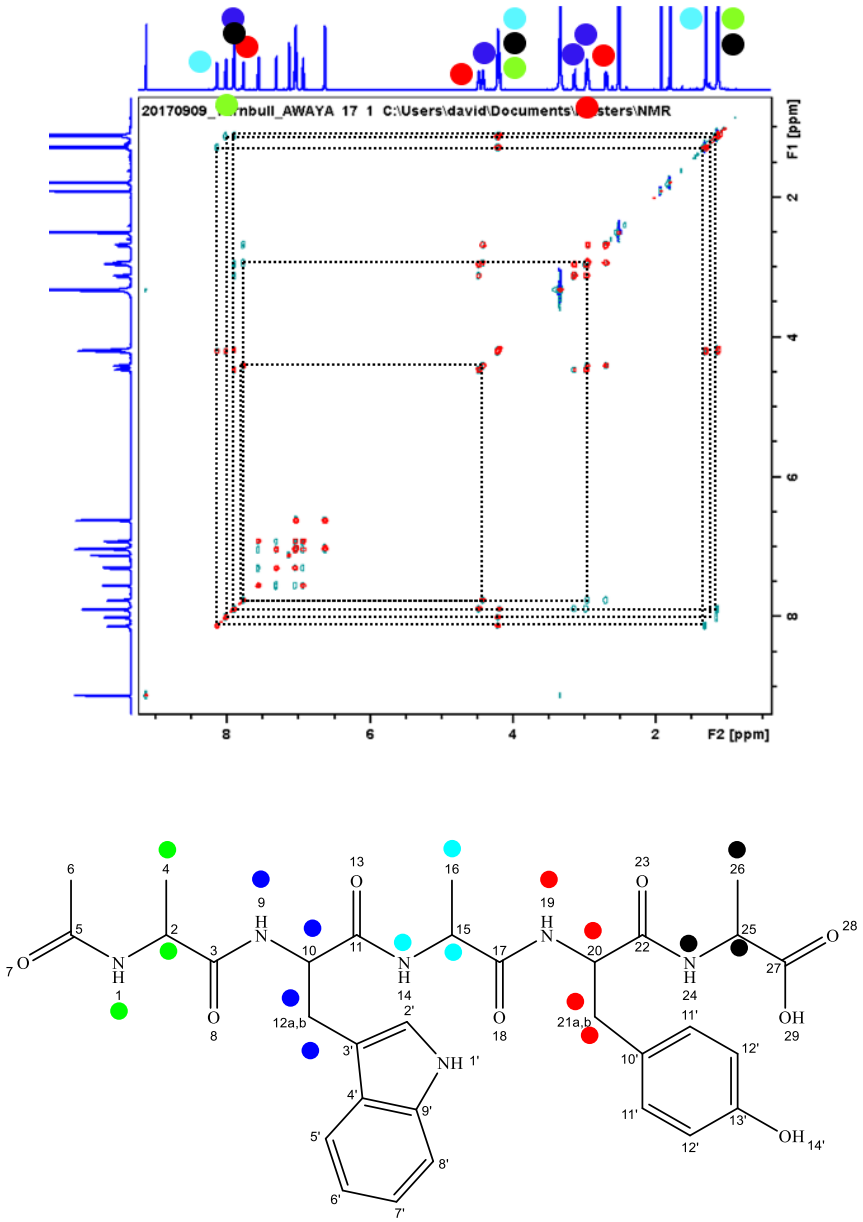


Figure 43: TOCSY (Blue) and COSY (Red) spectrum of **8** – examination of the peptide backbone (DMSO, 700 MHz, 300 K)

Through the spectrum above, it can be seen that the amide protons at 9 and 19 are in series with the diastereomeric protons on the β-carbons of the tryptophan and tyrosine residues. This indicates that the assignment of the protons is correct.

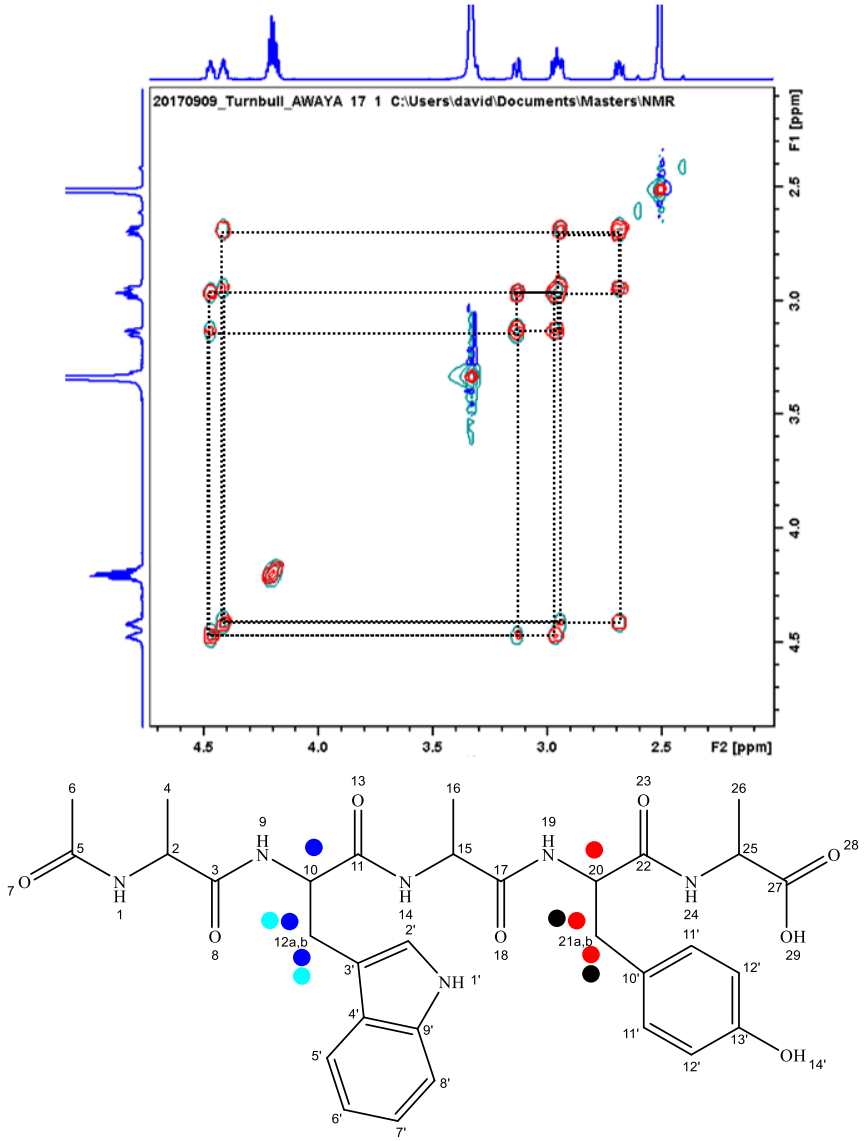
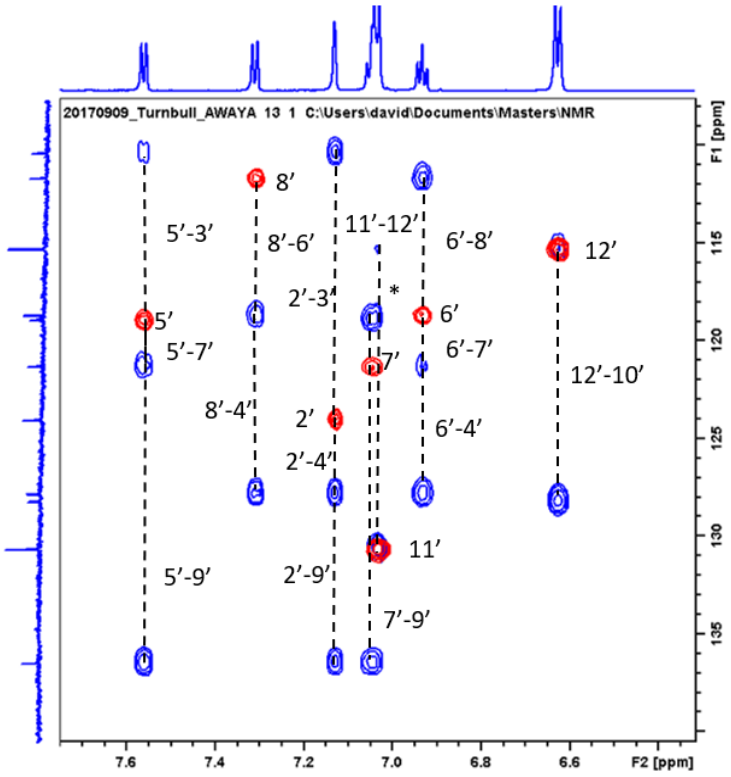


Figure 44: TOCSY (Blue) and COSY (Red) spectrum of **8** – examination of the diastereomeric protons (DMSO, 700 MHz, 300 K)



*7-5+6

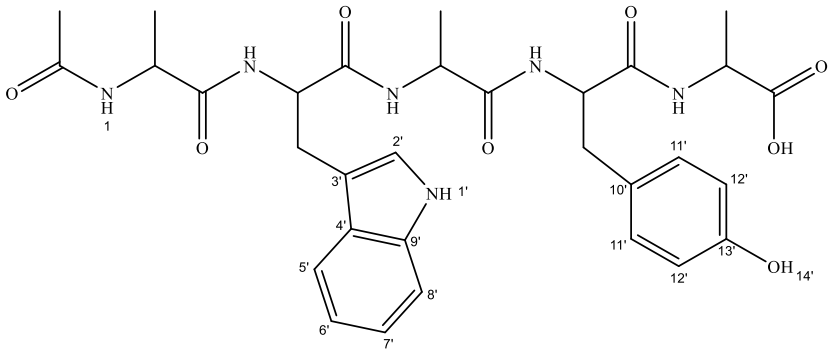


Figure 45: HSQC and HMBC spectrum of the aliphatic region of **8** - examination of the aromatic region (DMSO, for 1H -700 MHz, for ^{13}C - 176 MHz, 300 K)

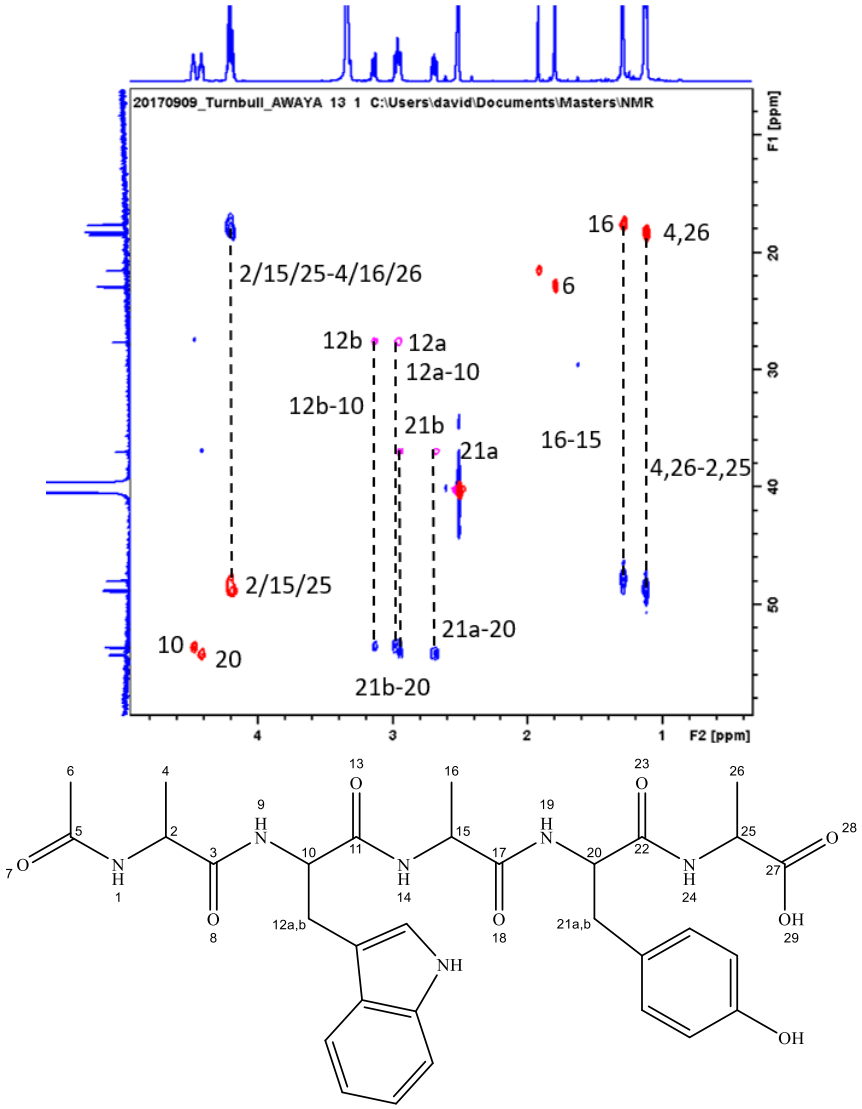


Figure 46: HSQC and HMBC spectrum of the aliphatic region of **8** - examination of the aliphatic region (DMSO, for 1H -700 MHz, for ^{13}C - 176 MHz, 300 K)

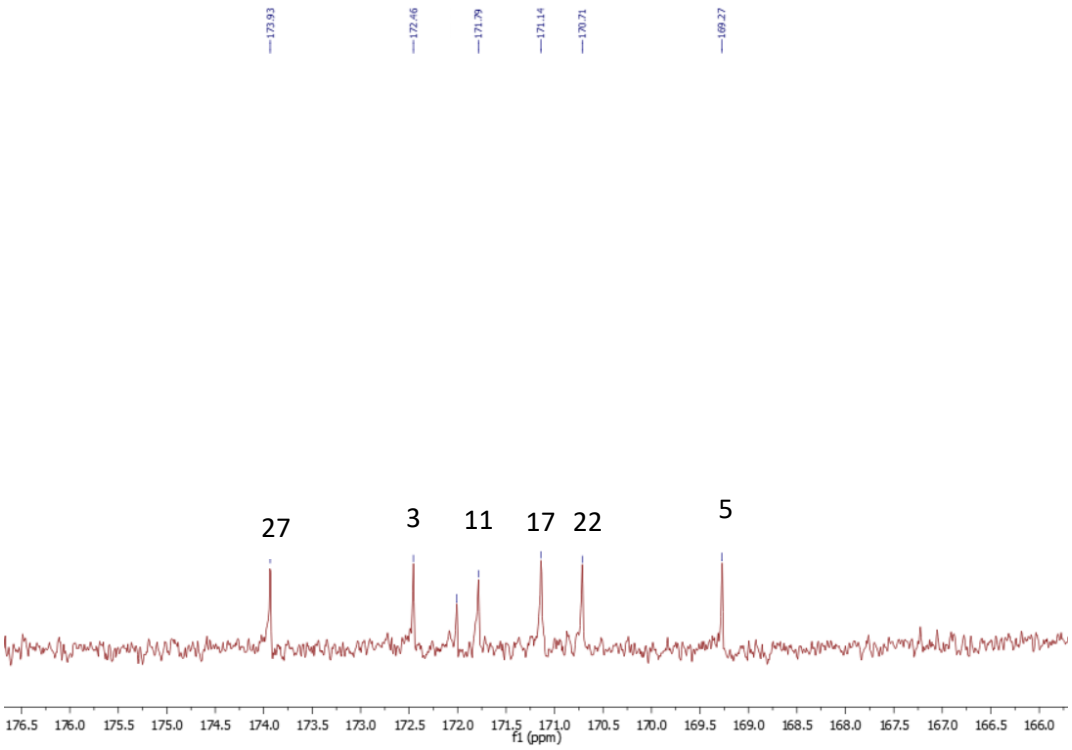


Figure 47: ^{13}C NMR spectrum of **8**, ^{13}C assignments between δ 169.5 – 174.0 ppm (DMSO, 176 MHz, 300K)

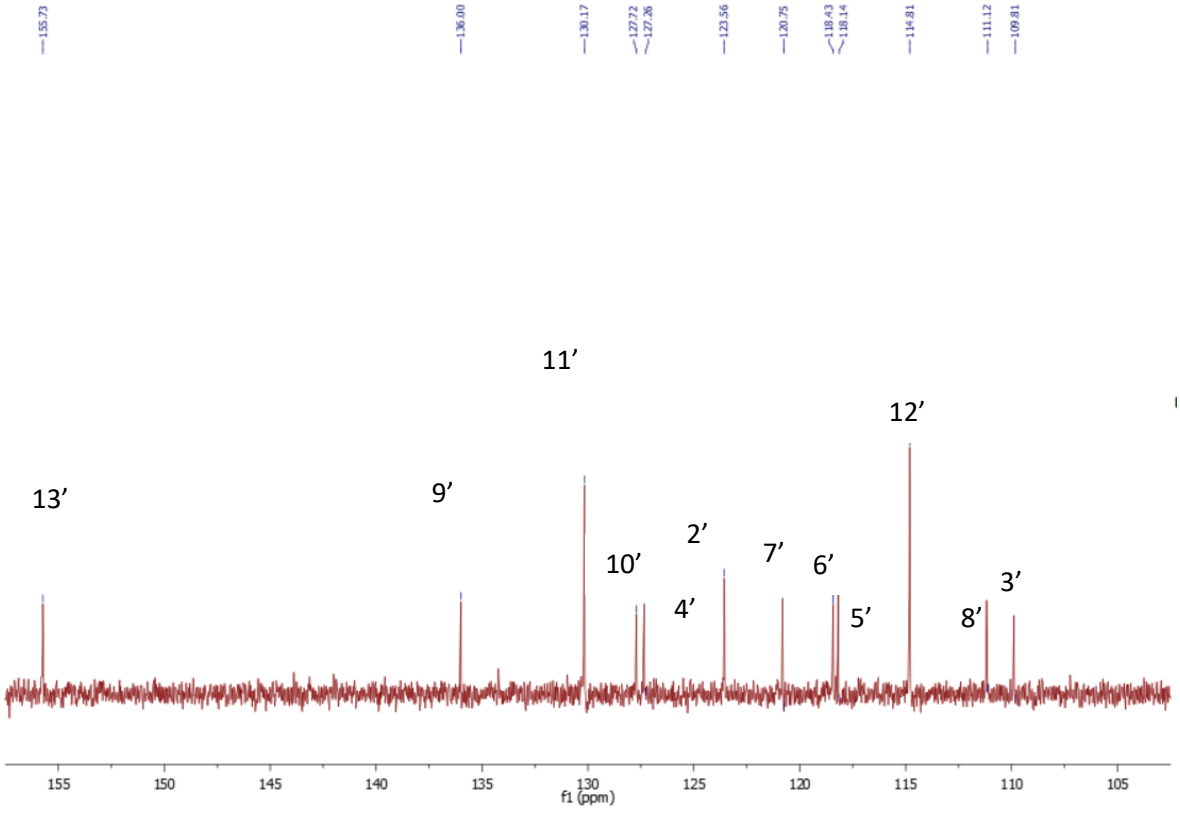


Figure 48: ^{13}C NMR spectrum of **8**, ^{13}C assignments between δ 110.0 – 157.0 ppm (DMSO, 176 MHz, 300K)

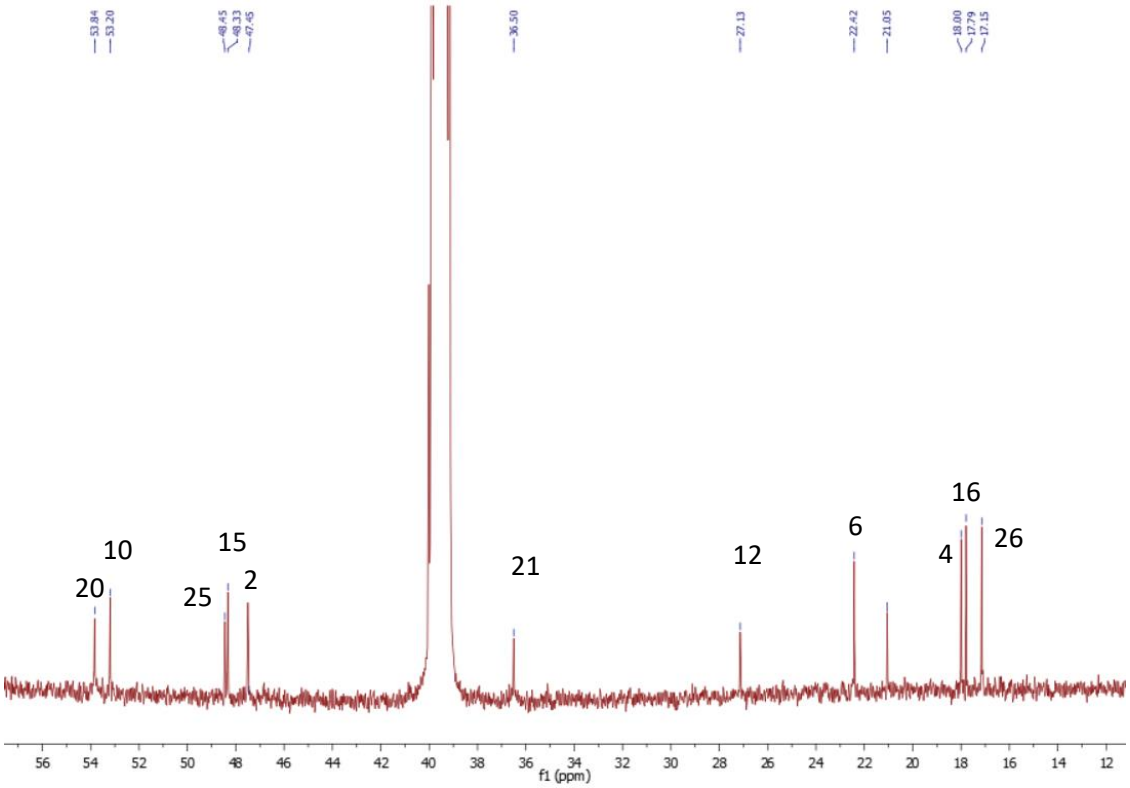


Figure 49: ^{13}C NMR spectrum of **8**, ^{13}C assignments between δ 16.0 – 55.0 ppm (DMSO, 176 MHz, 300K)

AcNH-Ser-Gly-Trp-Ala-OH (7)

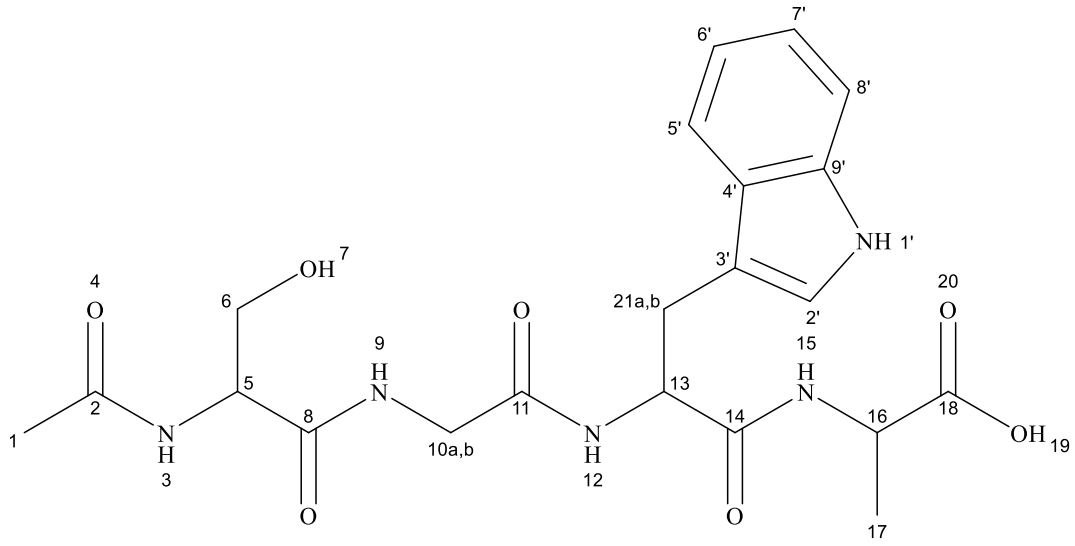


Figure 50: Reference molecule for Ser-Gly-Trp-Ala **7**

NMR spectrum data for **7**

^{1}H NMR (700 MHz, DMSO) δ 12.49 (s, 1H), 10.79 (s, 1H), 8.26 (d, *J* = 7.1 Hz, 1H), 8.14 (t, *J* = 5.5 Hz, 1H), 7.97 (dd, *J* = 12.5, 8.0 Hz, 2H), 7.61 (d, *J* = 7.9 Hz, 1H), 7.32 (d, *J* = 8.0 Hz, 1H), 7.14 (s, 1H), 7.06 (t, *J* = 7.5 Hz, 1H), 6.98 (t, *J* = 7.4 Hz, 1H), 4.98 (s, 1H), 4.56 (td, *J* = 9.0, 4.1 Hz, 1H), 4.21 (dd, *J* = 13.6, 6.7 Hz, 2H), 3.71 (dd, *J* = 16.8, 5.9 Hz, 1H), 3.62 – 3.53 (m, 3H), 3.15 (dd, *J* = 14.7, 3.7 Hz, 1H), 2.90 (dd, *J* = 14.7, 9.7 Hz, 1H), 1.86 (s, 3H), 1.30 (d, *J* = 7.3 Hz, 3H).

^{13}C NMR (176 MHz, DMSO) δ 173.95, 171.27, 170.73, 169.67, 168.48, 136.01, 127.22, 123.67, 120.72, 118.30, 118.16, 111.23, 110.00, 61.63, 55.40, 52.97, 47.55, 42.12, 27.67, 22.53, 17.05.

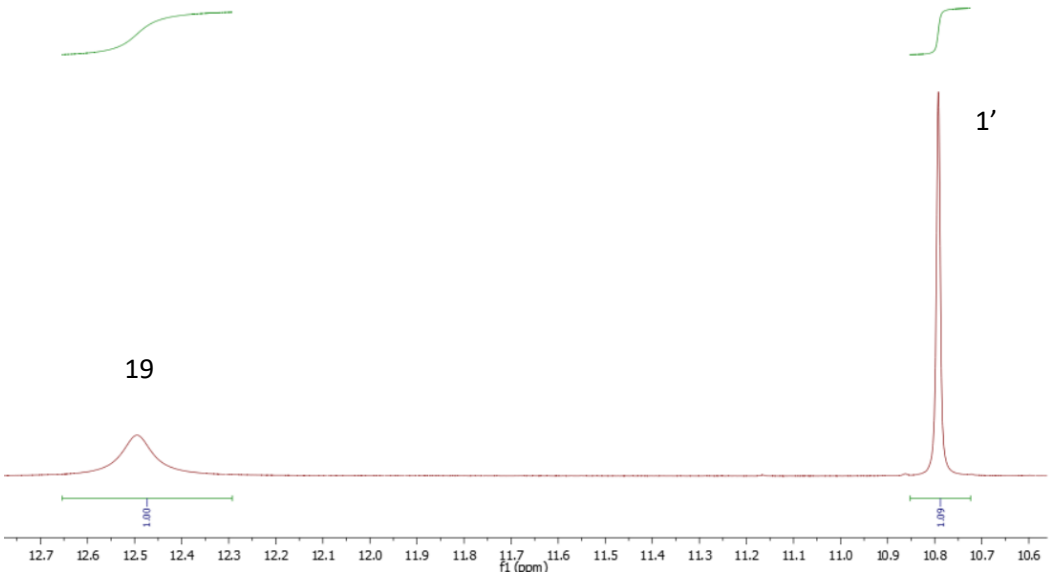


Figure 51: ^{1}H NMR spectrum of **7,** showing ^{1}H assignments between δ 10.7 – 12.7 ppm (DMSO, 700 MHz, 300 K)

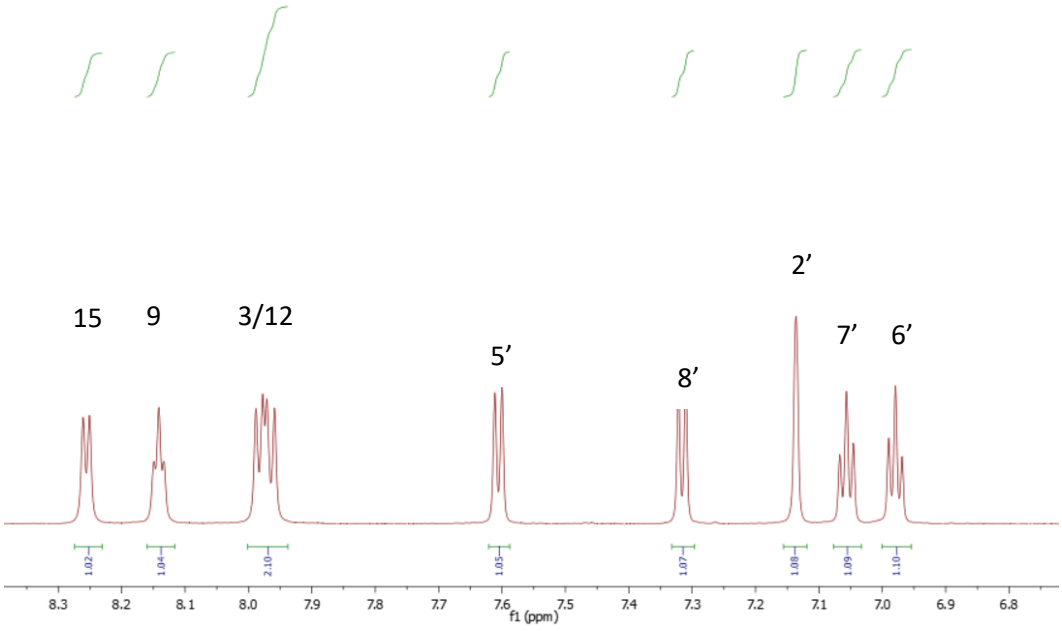


Figure 52: ^{1}H NMR spectrum of **7,** showing ^{1}H assignments between δ 6.9 – 8.3 ppm (DMSO, 700 MHz, 300 K)

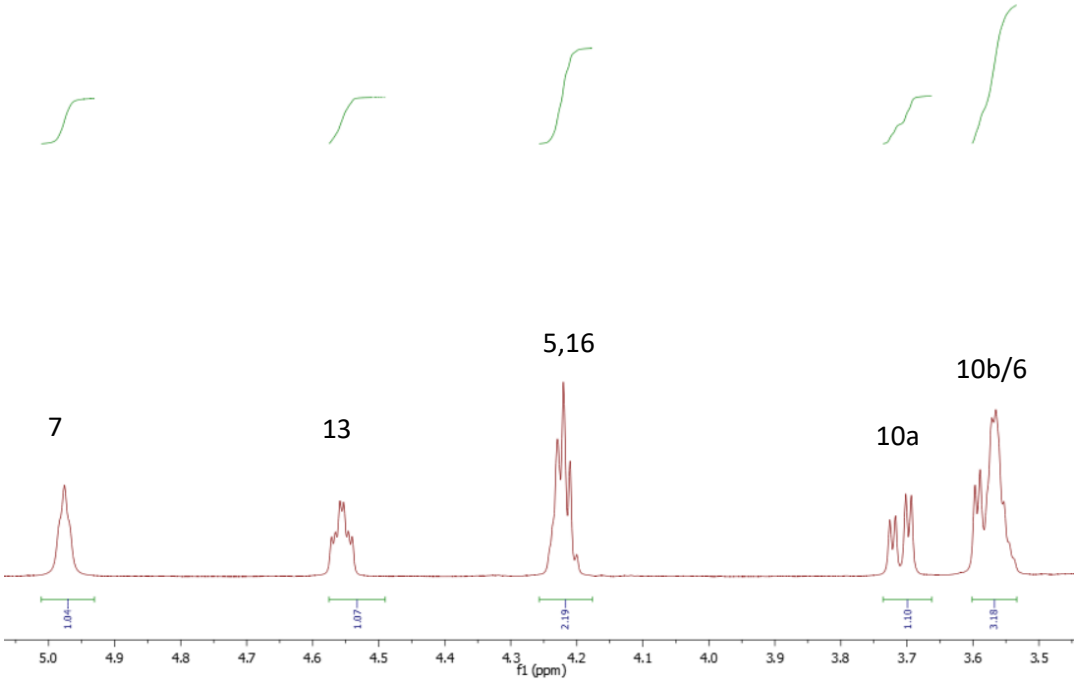


Figure 53: ^{1}H NMR spectrum of **7,** showing ^{1}H assignments between δ 3.5 – 5.0 ppm (DMSO, 700 MHz, 300 K)

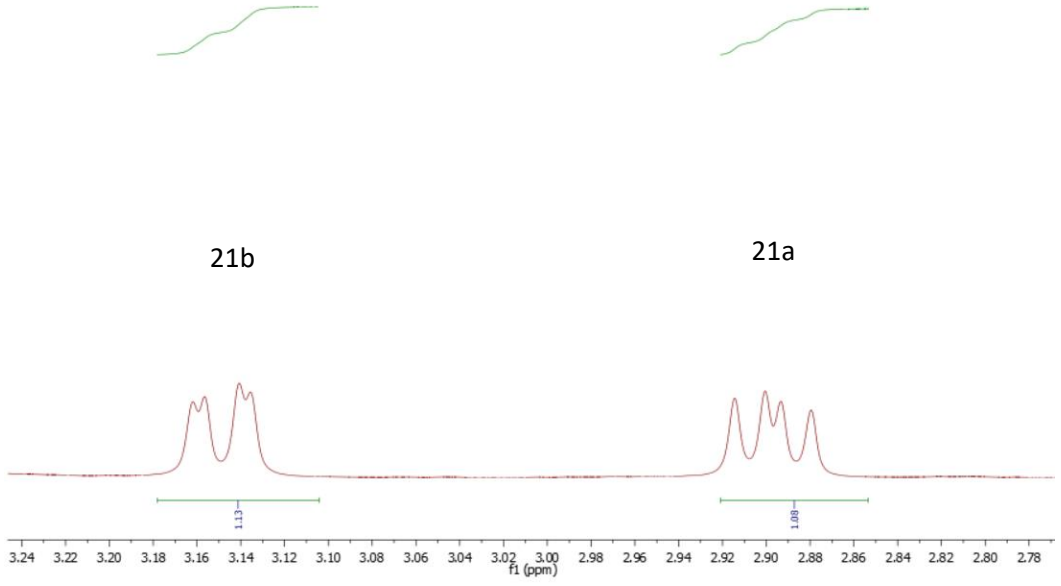


Figure 54: ^{1}H NMR spectrum of **7,** showing ^{1}H assignments between δ 2.86 – 3.16 ppm (DMSO, 700 MHz, 300 K)

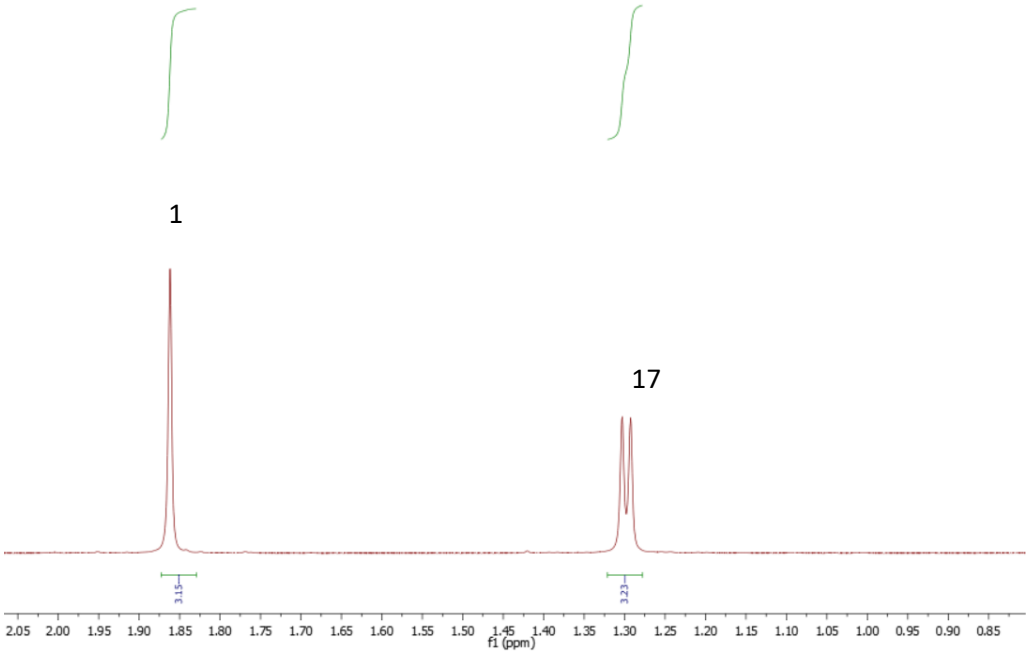


Figure 55: ^{1}H NMR spectrum of **7,** showing ^{1}H assignments between δ 1.25 – 1.9 ppm (DMSO, 700 MHz, 300 K)

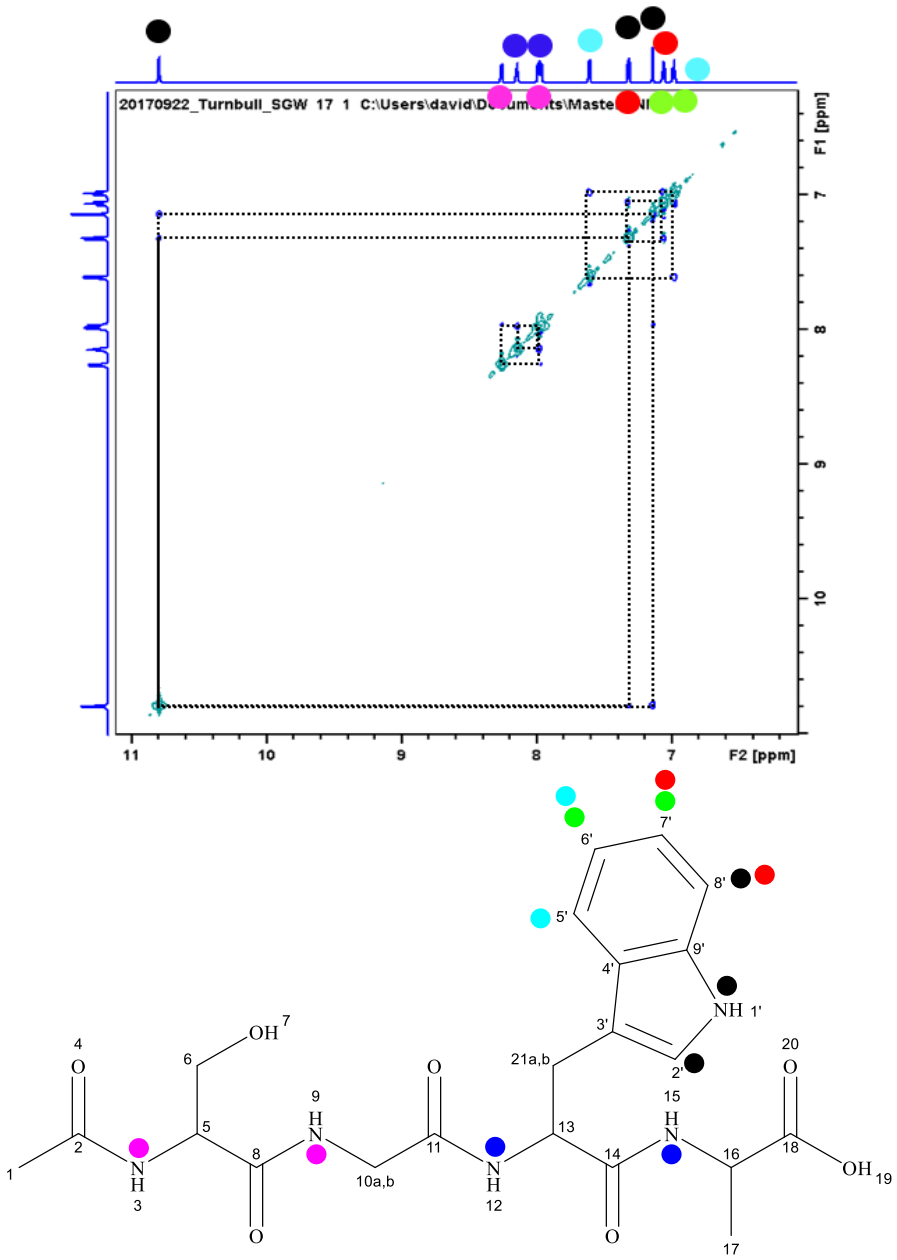


Figure 56: ROESY Spectrum of aromatic region of **7** – examination of the indole region (DMSO, 700 MHz, 300 K)

From the spectra in Figure 56, the usual interactions between the aromatic protons can be seen with additional interaction between the indole nitrogen proton and the position 8' proton on the aromatic ring, which is the same as Ala-Trp-Ala but not Ala-Trp-Ala-Tyr-Ala, this could be due to low steric bulk of the other amino acids in the peptide compared to the Tyrosine. Another unusual set of interactions involving the amide protons, the position 12 proton has an interaction with position 15 and position 9 has an interaction with position 3. From the molecule, it can be assumed that the molecule is bending which causes the molecule to fold back in on itself. If this is the case, it can be assumed that proton 9 interacts with 3 and 12 interacts with 15.

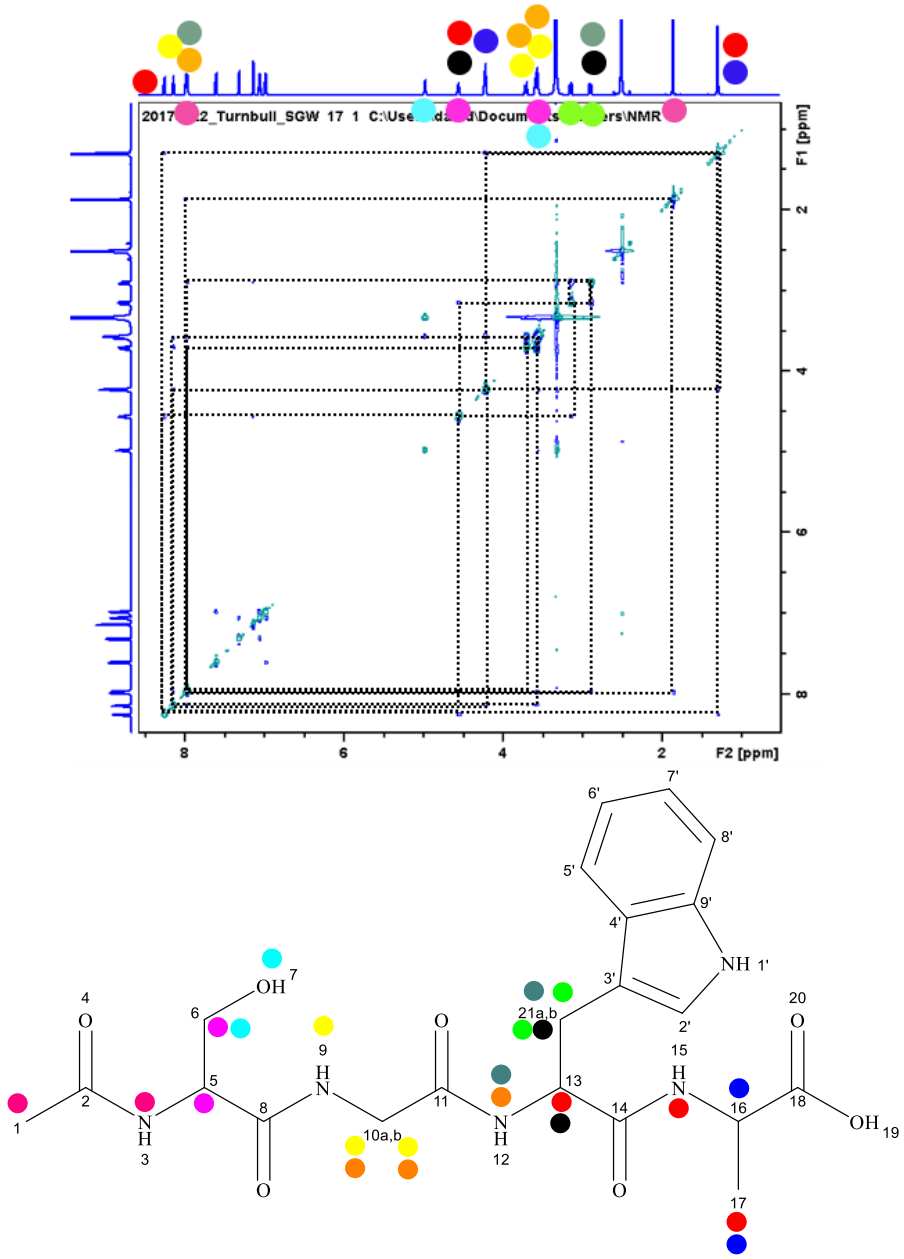


Figure 57: ROESY Spectrum of **7** – examination of the peptide backbone (DMSO, 700 MHz, 300 K

There are several interesting interactions in the spectrum above, the amide proton at position 9 interacts with the diastereomeric protons at position 10 but so does the proton at position 12 or 3, but with the closer proximity of 12, it is assumed to be proton 12 which is close to both the diastereomeric protons. The amide proton at position 3 is close to the methyl proton at position 1, which is to be expected. Proton 15 is close to the protons at position 17 which is expected. The amide proton at position 3 or 12 is close to one of the diastereomeric protons on atom 21. This is unusual but can be explained by the molecule curving on itself therefore it is likely to be the position 12 proton.

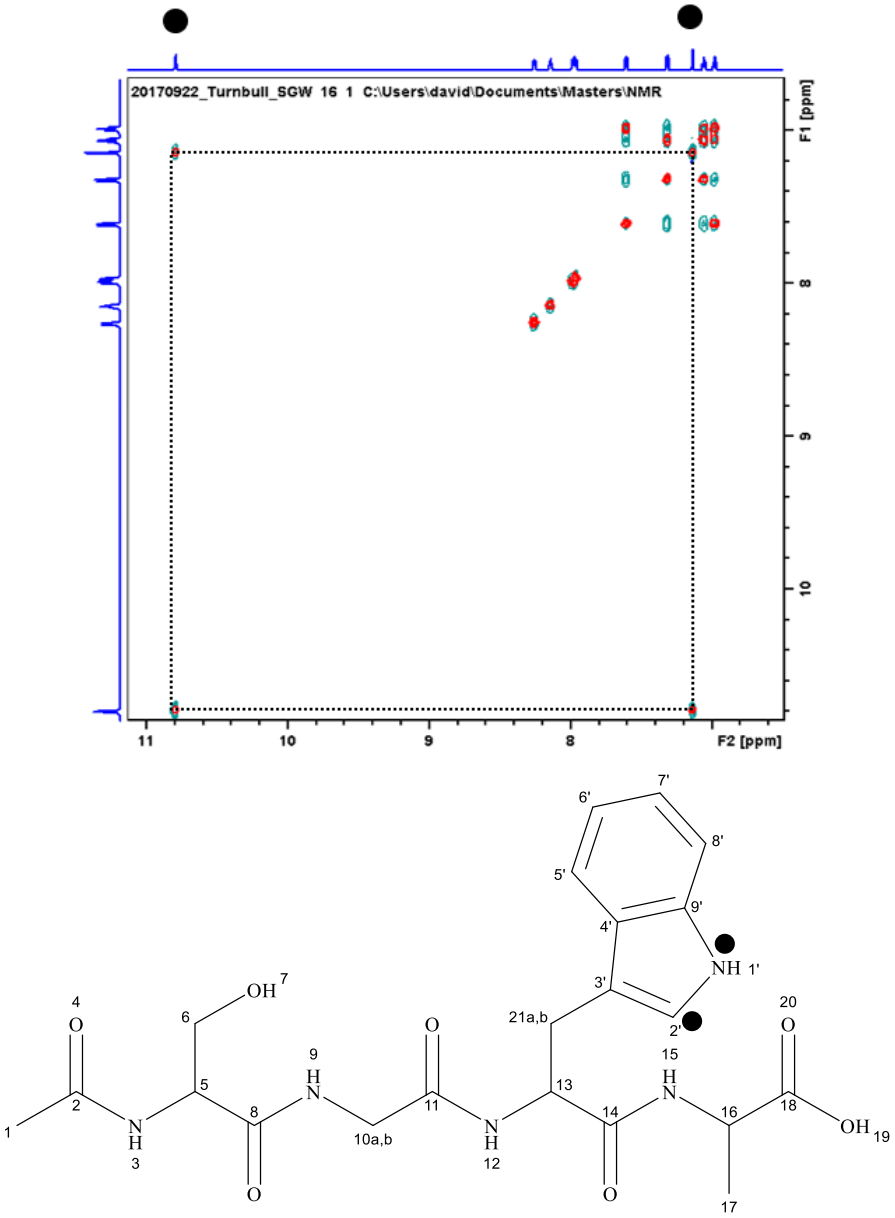


Figure 58: TOCSY (Blue) and COSY (Red) spectrum of **7** – examination of the indole ring on the nitrogen (DMSO, 700 MHz, 300 K)

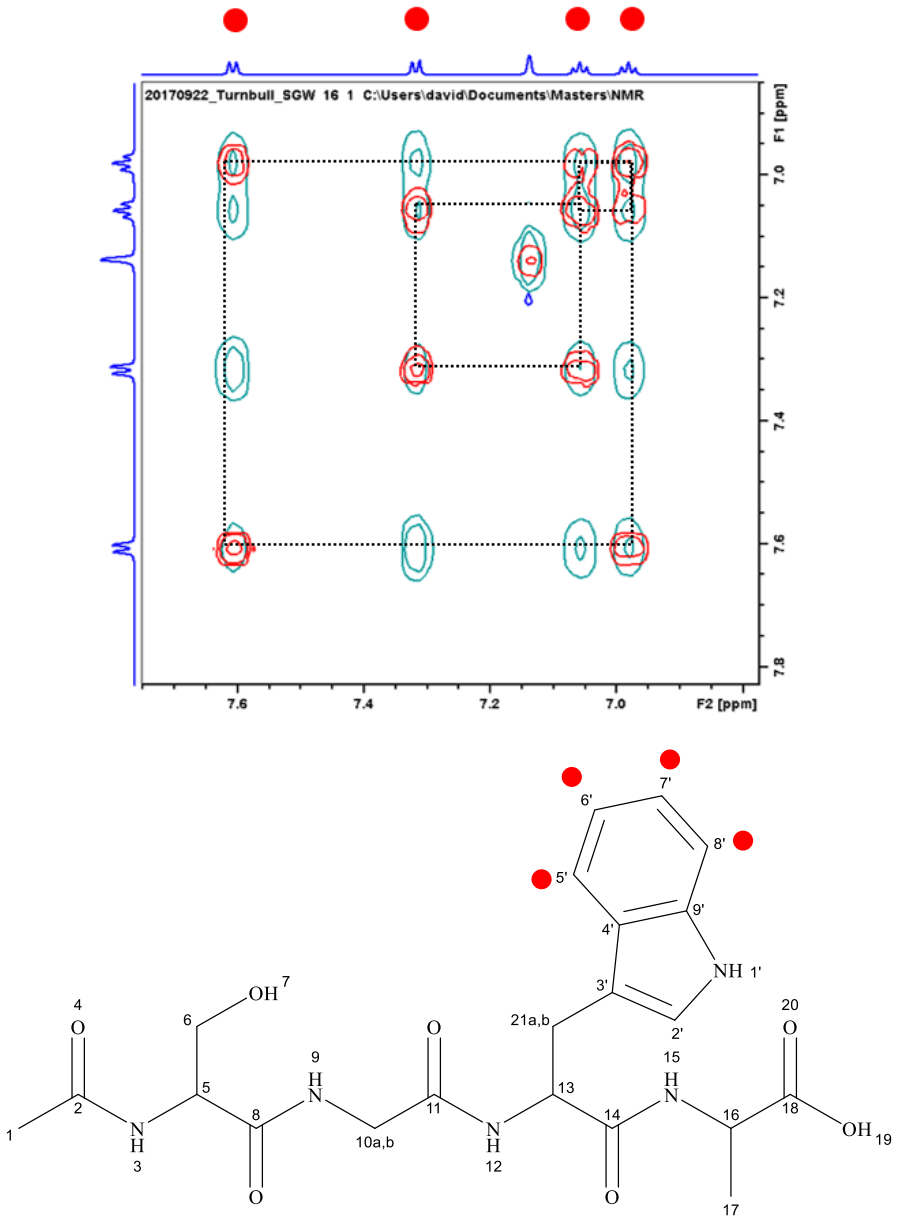


Figure 59: TOCSY (Blue) and COSY (Red) spectrum of **7** – examination of the indole ring (DMSO, 700 MHz, 300 K)

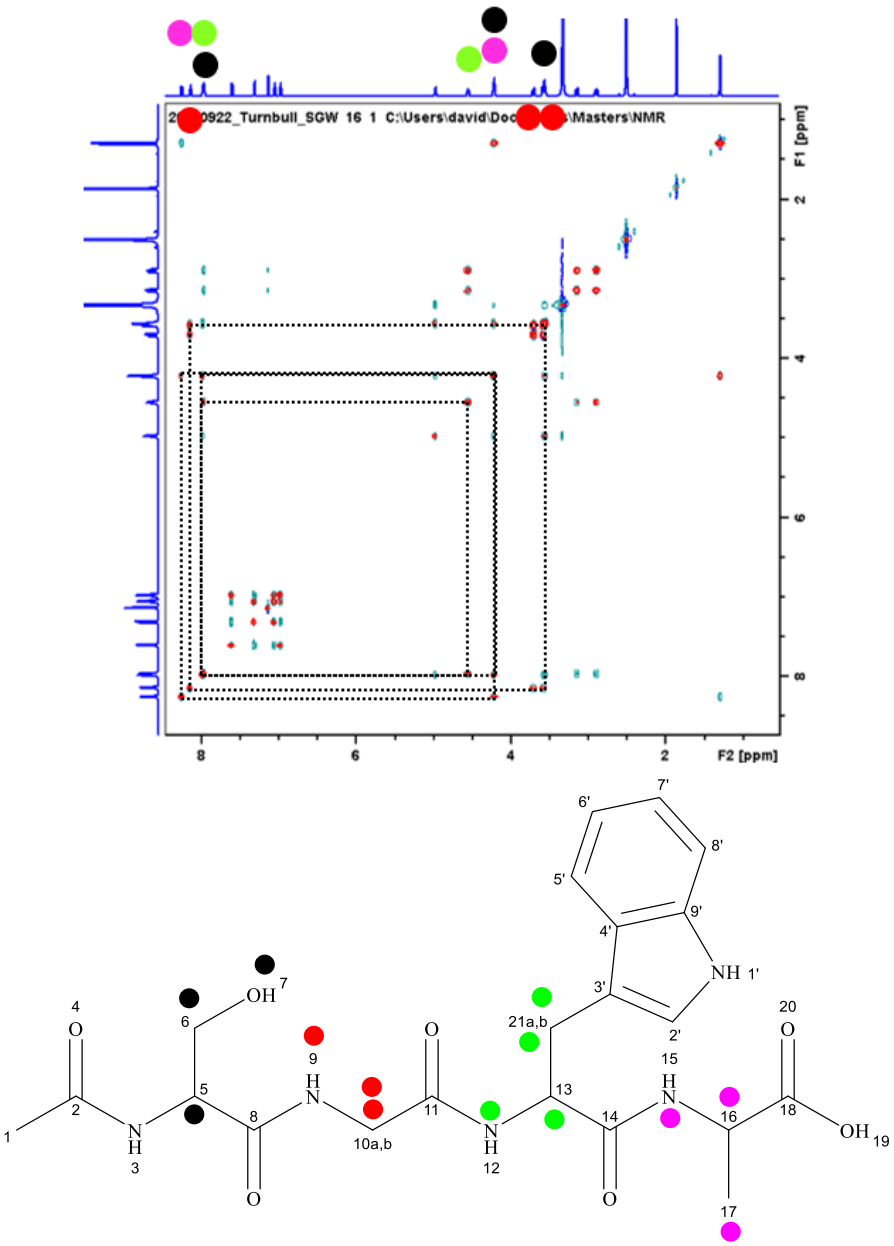


Figure 60: TOCSY (Blue) and COSY (Red) spectrum of **7** – examination of the peptide backbone (DMSO, 700 MHz, 300 K)

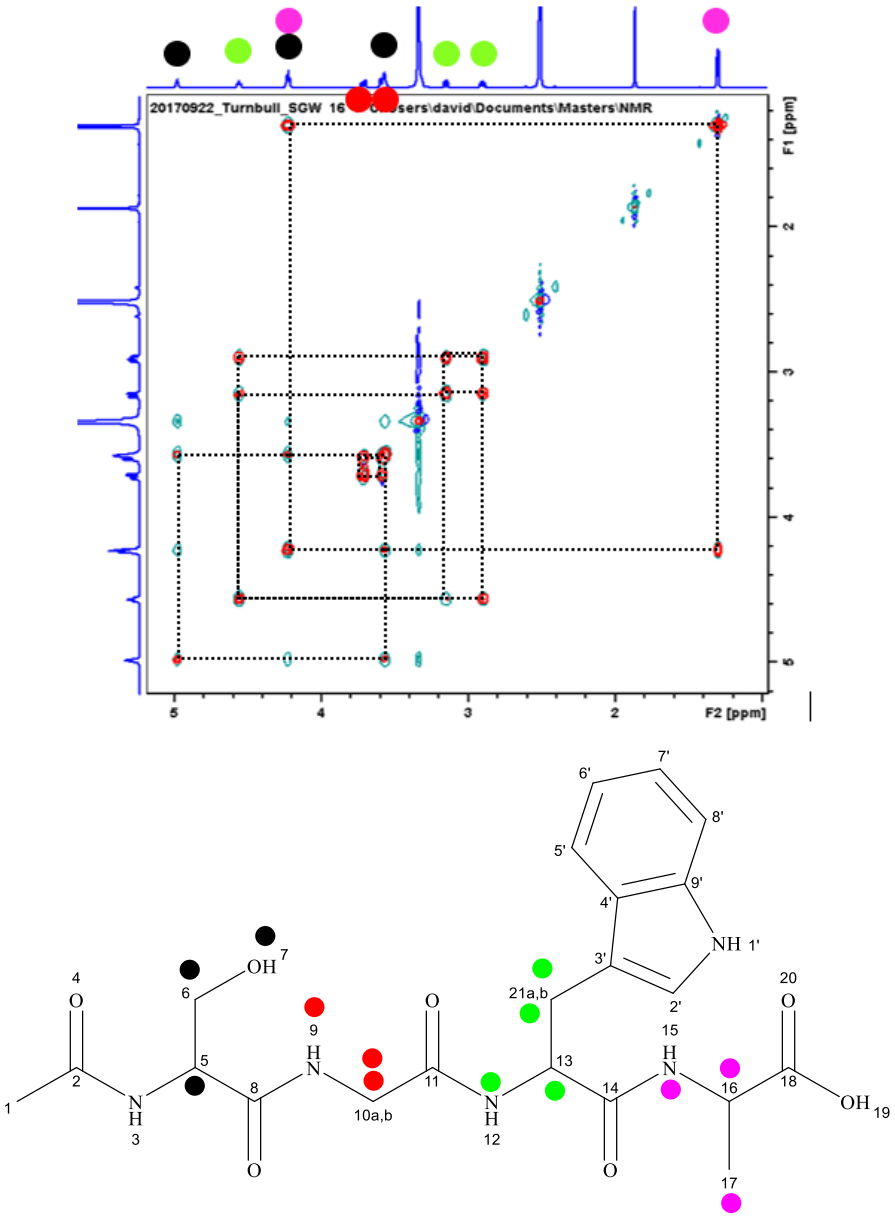


Figure 61: TOCSY (Blue) and COSY (Red) spectrum of **7** – examination of the aliphatic region (DMSO, 700 MHz, 300 K)

There are several positions that overlap therefore they need qualifying. Position 6 is overlapped with 10b. Through the spectrum above, it can be seen that proton 7 is interacting with position 6 and proton 6 interacts with position 5.

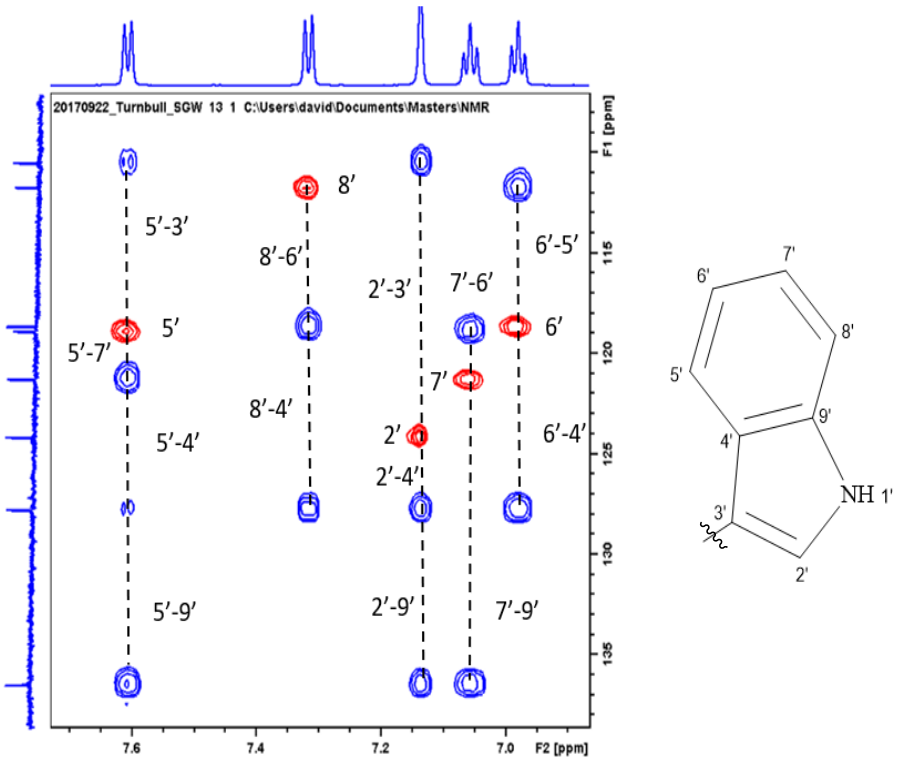


Figure 62: HSQC and HMBC spectrum of the aromatic region of **7** - examination of the indole ring (DMSO, for 1H -700 MHz, for ^{13}C - 176 MHz, 300 K)

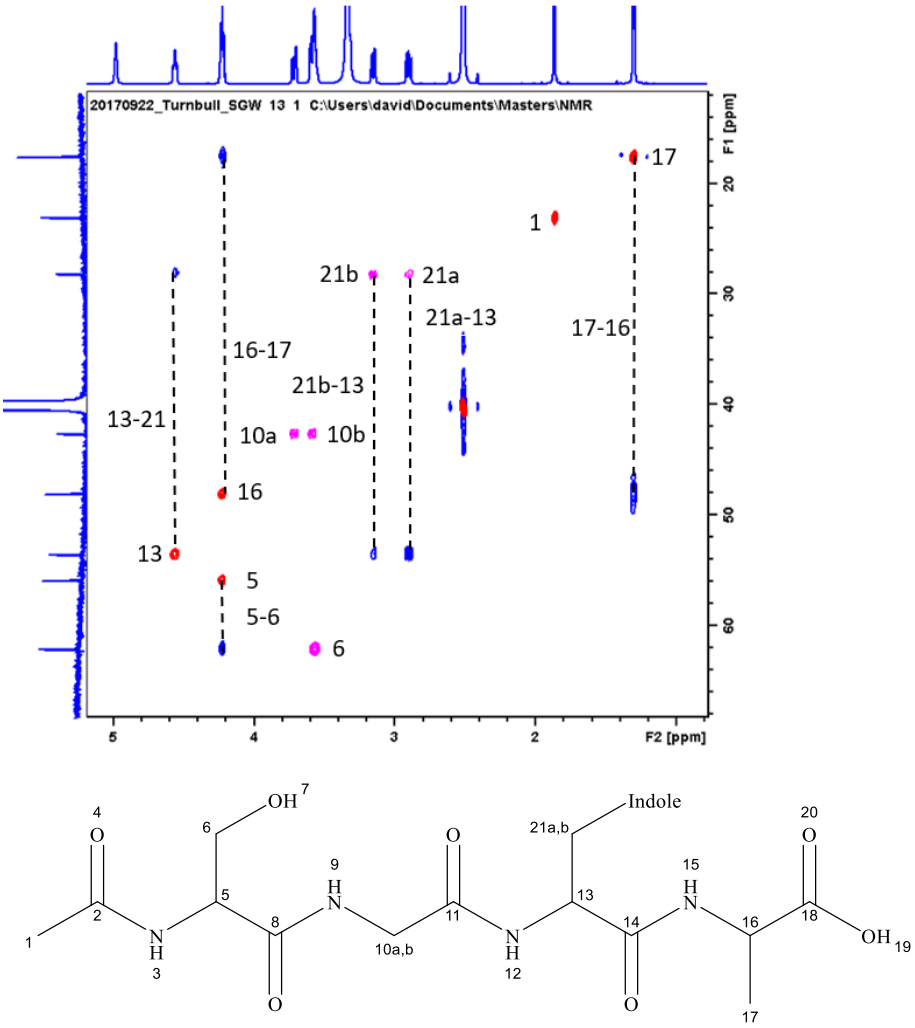


Figure 63: HSQC and HMBC spectrum of the aromatic region of **7** - examination of the aliphatic region (DMSO, for 1H -700 MHz, for ^{13}C - 176 MHz, 300 K)

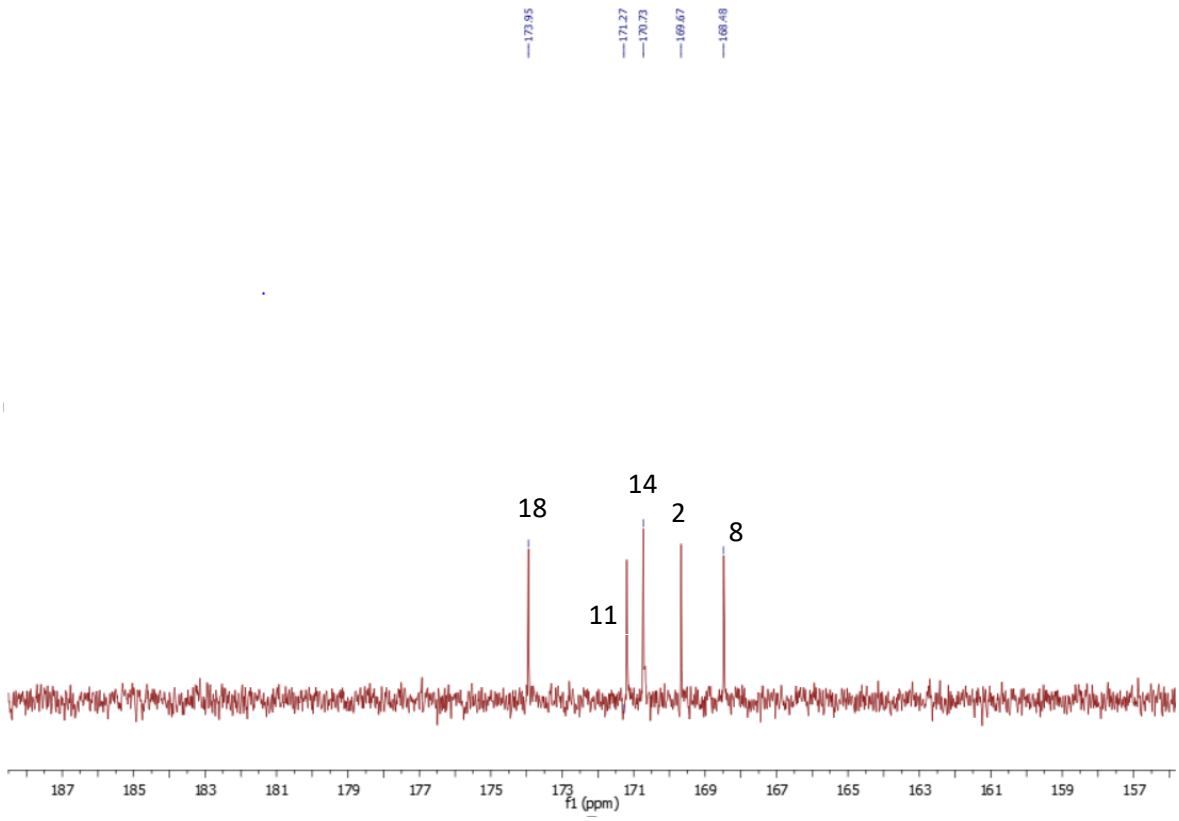


Figure 64: ^{13}C NMR spectrum of **7**, ^{13}C assignments between δ 156 - 188 ppm (DMSO, 176 MHz, 300K)

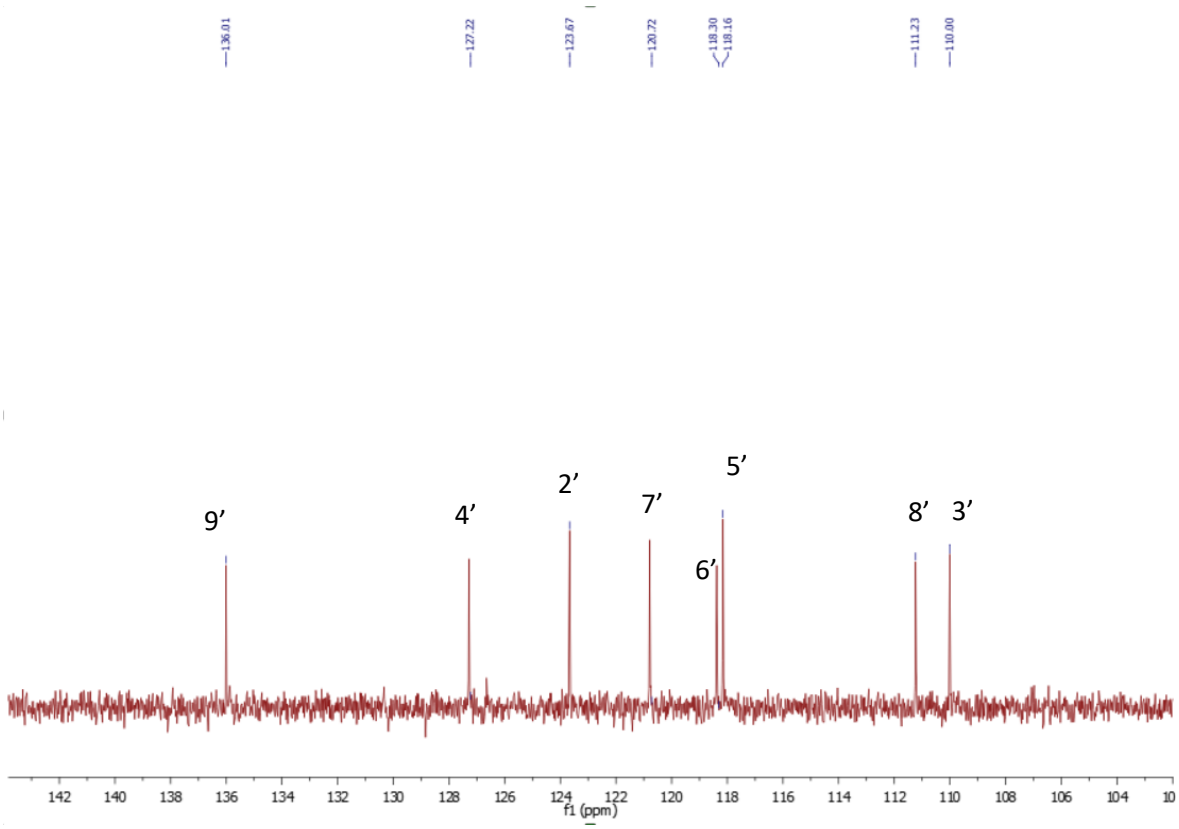


Figure 65: ^{13}C NMR spectrum of **7**, ^{13}C assignments between δ 110 - 138 ppm (DMSO, 176 MHz, 300K)

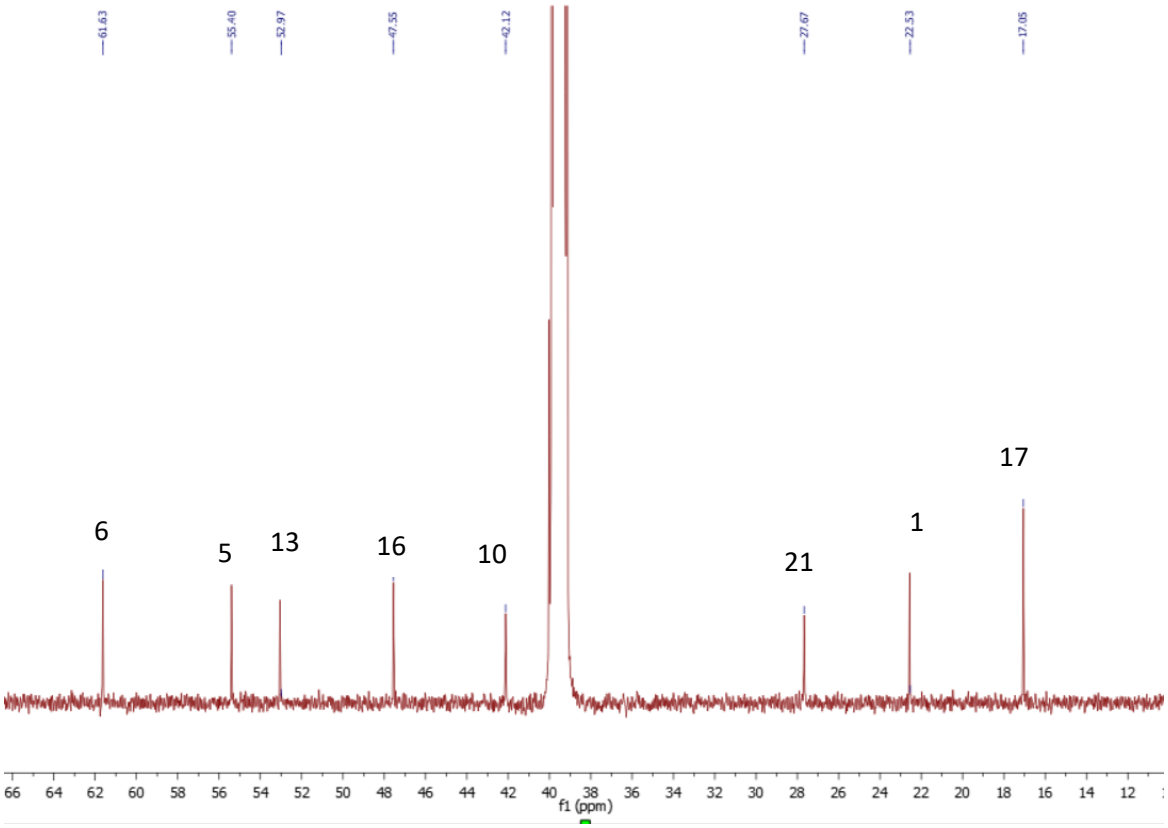


Figure 66: ^{13}C NMR spectrum of **7**, ^{13}C assignments between δ 15 – 65 ppm (DMSO, 176 MHz, 300K)

AcNH-Val-Trp-Asn-Asn-Lys-Thr-Ala-OH (9)

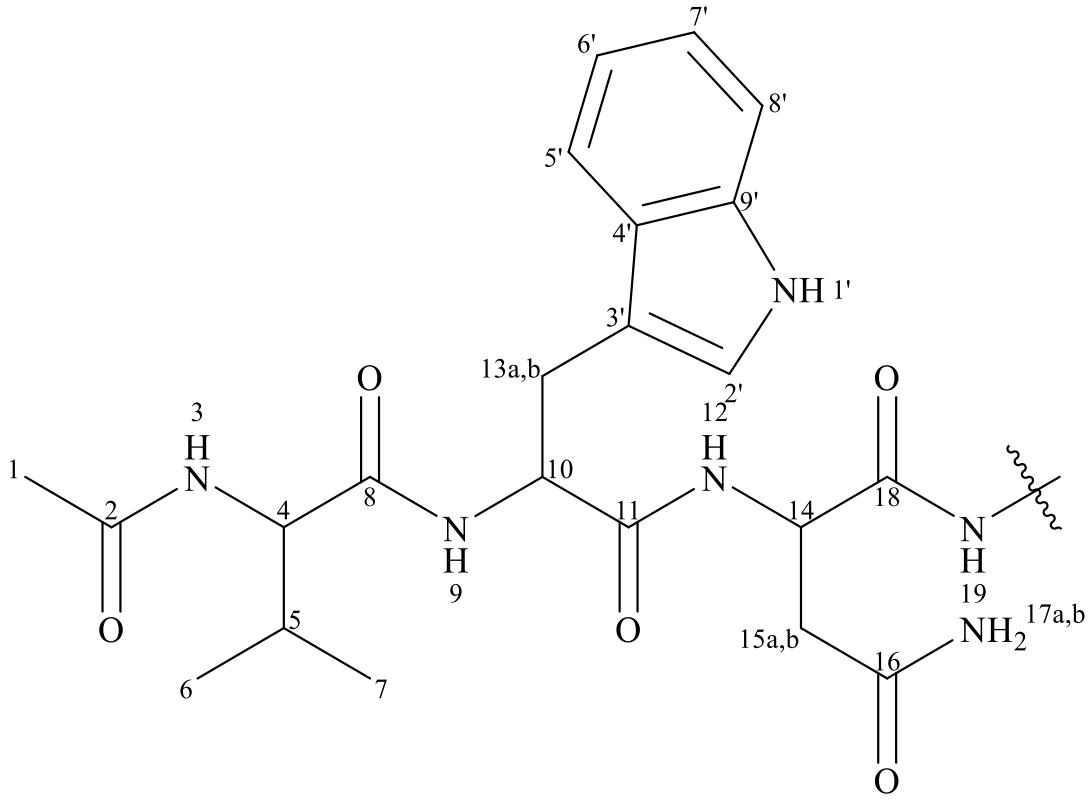


Figure 67: One half of the Val-Trp-Asn-Asn-Lys-Thr-Ala molecule *(9)*

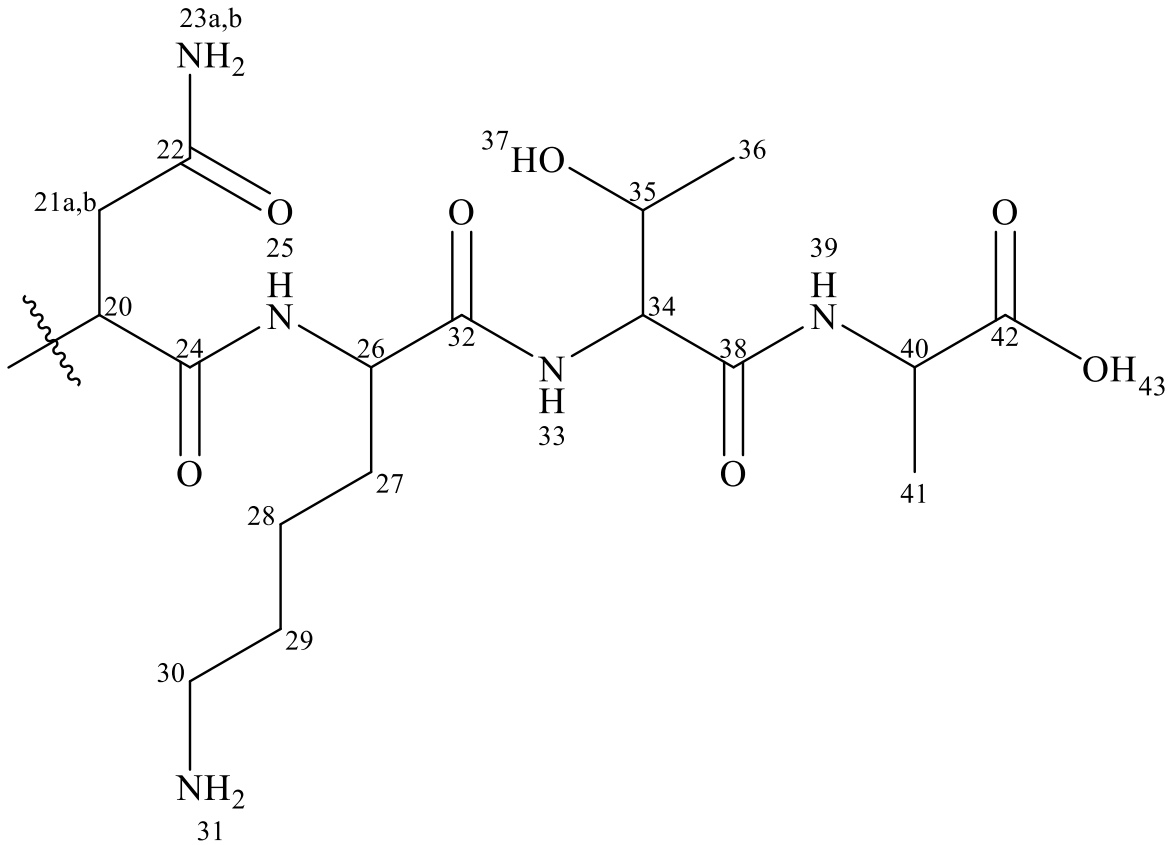


Figure 68: The other half of the Val-Trp-Asn-Asn-Lys-Thr-Ala molecule *(9)*

NMR spectrum data for **9**

^{1}H NMR (700 MHz, DMSO) δ 10.77 (s, 1H), 8.11 (t, *J* = 8.6 Hz, 2H), 7.98 (d, *J* = 7.4 Hz, 2H), 7.93 (d, *J* = 7.0 Hz, 1H), 7.82 (d, *J* = 8.5 Hz, 1H), 7.67 (d, *J* = 8.4 Hz, 1H), 7.62 – 7.55 (m, 4H), 7.45 (d, *J* = 14.3 Hz, 2H), 7.31 (d, *J* = 8.1 Hz, 1H), 7.11 (s, 1H), 7.04 (t, *J* = 7.5 Hz, 1H), 6.96 (dd, *J* = 14.1, 6.6 Hz, 3H), 4.76 (s, 1H), 4.53 (td, *J* = 13.7, 6.5 Hz, 3H), 4.26 (dd, *J* = 13.3, 8.7 Hz, 1H), 4.21 – 4.15 (m, 2H), 4.09 (t, *J* = 7.7 Hz, 1H), 3.98 – 3.92 (m, 1H), 3.12 (dd, *J* = 14.8, 4.2 Hz, 1H), 2.94 (dd, *J* = 14.9, 9.4 Hz, 1H), 2.76 (dt, *J* = 21.2, 10.6 Hz, 2H), 2.59 (dd, *J* = 15.6, 6.9 Hz, 1H), 2.46 – 2.37 (m, 1H), 1.91 (s, 2H), 1.88 (dd, *J* = 13.6, 6.8 Hz, 1H), 1.84 (s, 3H), 1.74 (d, *J* = 7.2 Hz, 1H), 1.63 – 1.56 (m, 1H), 1.51 (td, *J* = 14.9, 7.3 Hz, 3H), 1.38 – 1.29 (m, 3H), 1.27 (d, *J* = 7.3 Hz, 4H), 1.06 (d, *J* = 6.3 Hz, 4H), 0.76 (dd, *J* = 6.6, 3.1 Hz, 7H).

^{13}C NMR (176 MHz, DMSO) δ 173.70, 171.77, 171.59, 171.22, 171.10, 170.95, 170.64, 170.53, 169.40, 169.32, 135.76, 127.04, 123.28, 120.61, 118.14, 117.95, 110.98, 109.69, 66.34, 57.86, 57.80, 52.95, 52.33, 49.62, 49.46, 47.31, 38.52, 36.68, 36.37, 30.45, 29.89, 27.12, 26.20, 22.24, 21.80, 20.82, 19.48, 18.98, 17.88, 17.05.

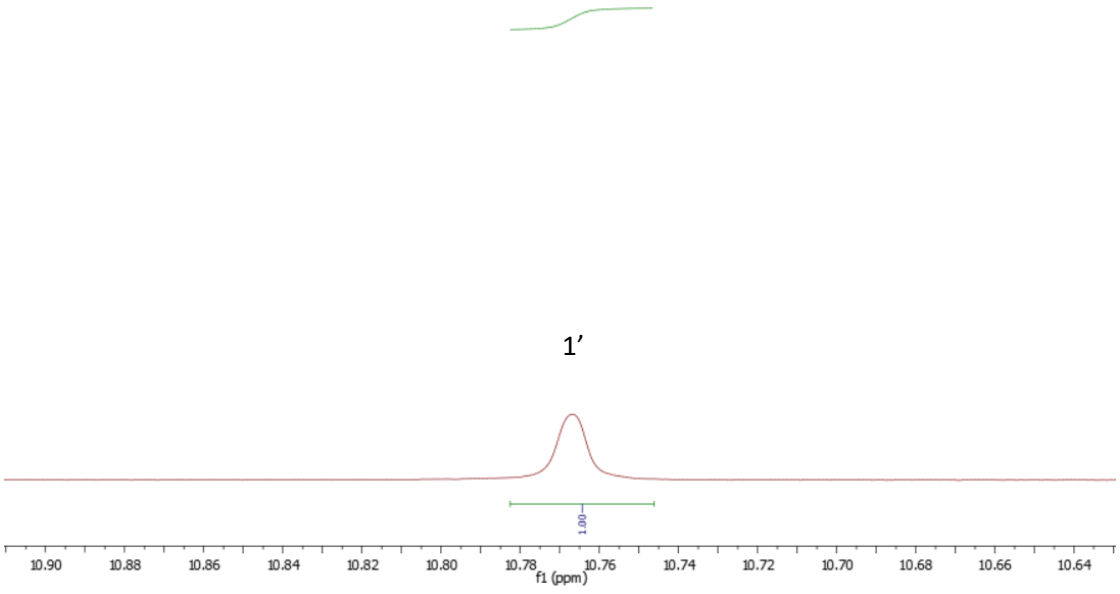


Figure 69: ^{1}H NMR spectrum of **9,** showing ^{1}H assignments between δ 10.70 – 10.80 ppm (DMSO, 700 MHz, 300 K)

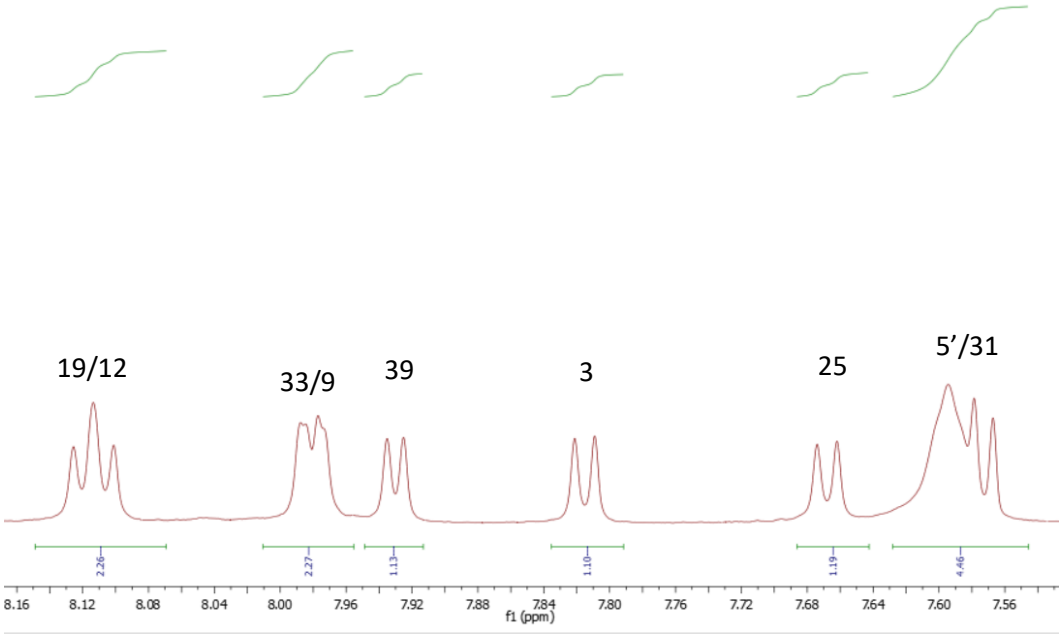


Figure 70: ^{1}H NMR spectrum of **9,** showing ^{1}H assignments between δ 7.56 – 8.16 ppm (DMSO, 700 MHz, 300 K)

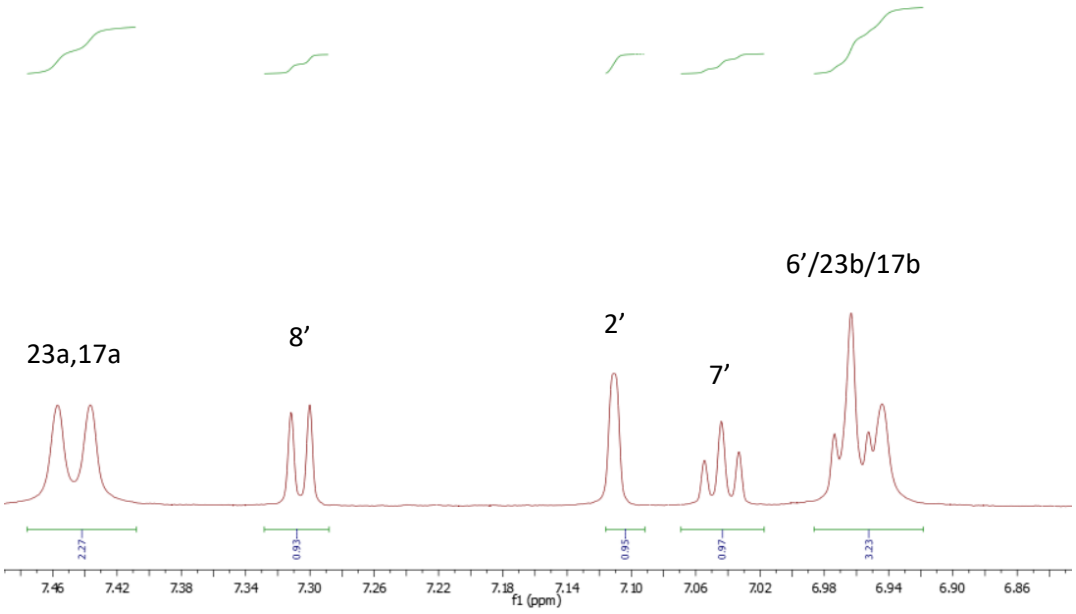


Figure 71: ^{1}H NMR spectrum of **9,** showing ^{1}H assignments between δ 6.90 – 7.46 ppm (DMSO, 700 MHz, 300 K)

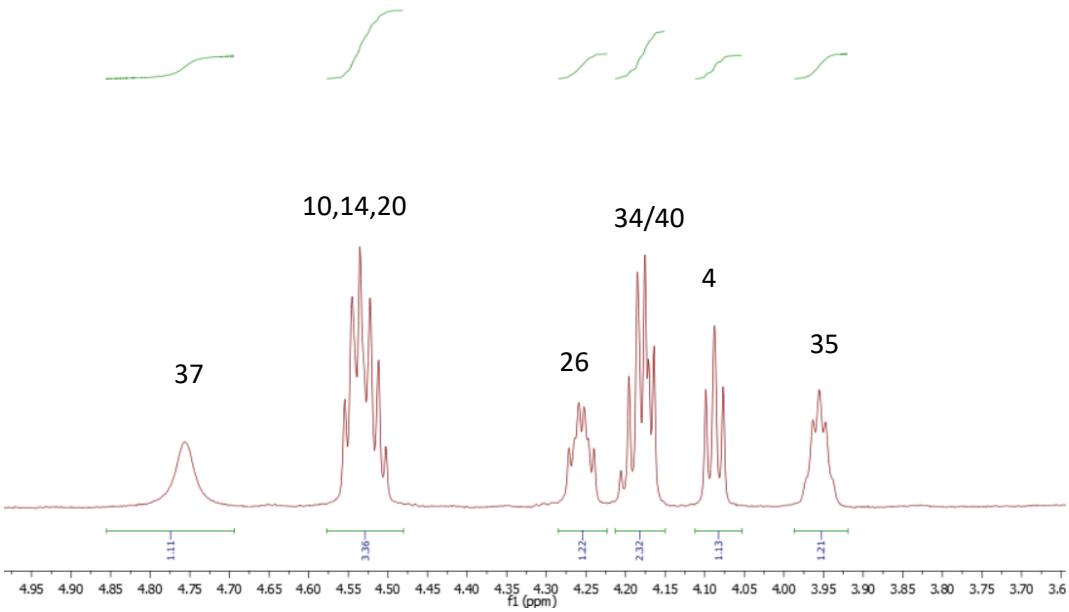


Figure 72: ^{1}H NMR spectrum of **9,** showing ^{1}H assignments between δ 3.90 – 4.80 ppm (DMSO, 700 MHz, 300 K)

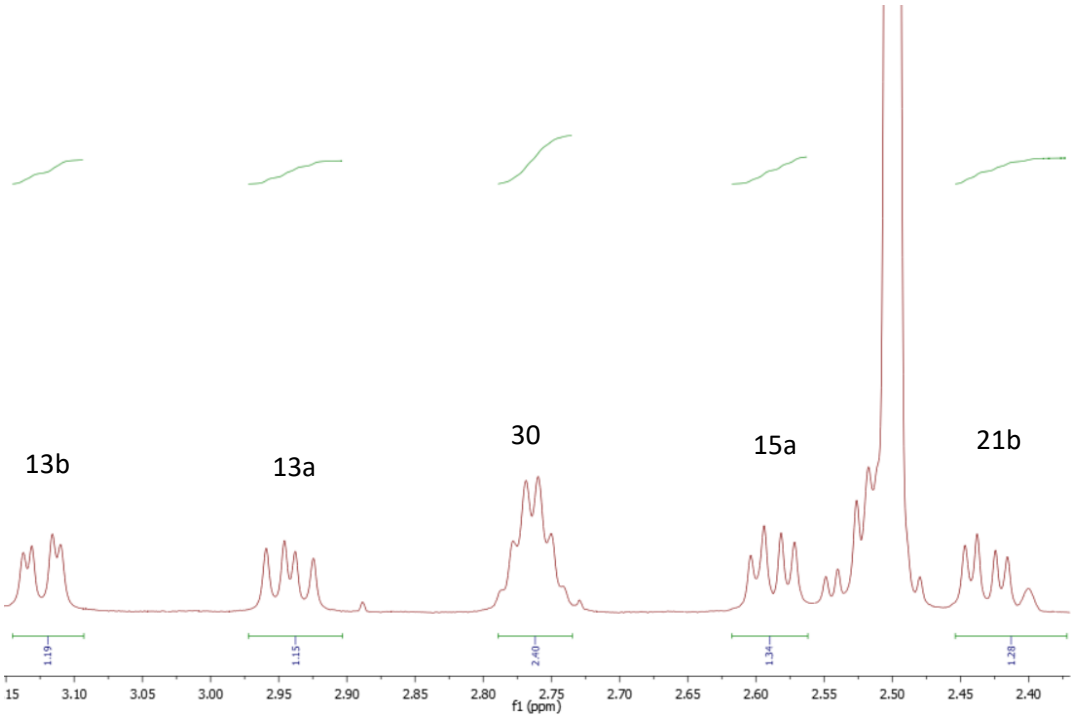


Figure 73: ^{1}H NMR spectrum of **9,** showing ^{1}H assignments between δ 2.4 – 3.15 ppm (DMSO, 700 MHz, 300 K)

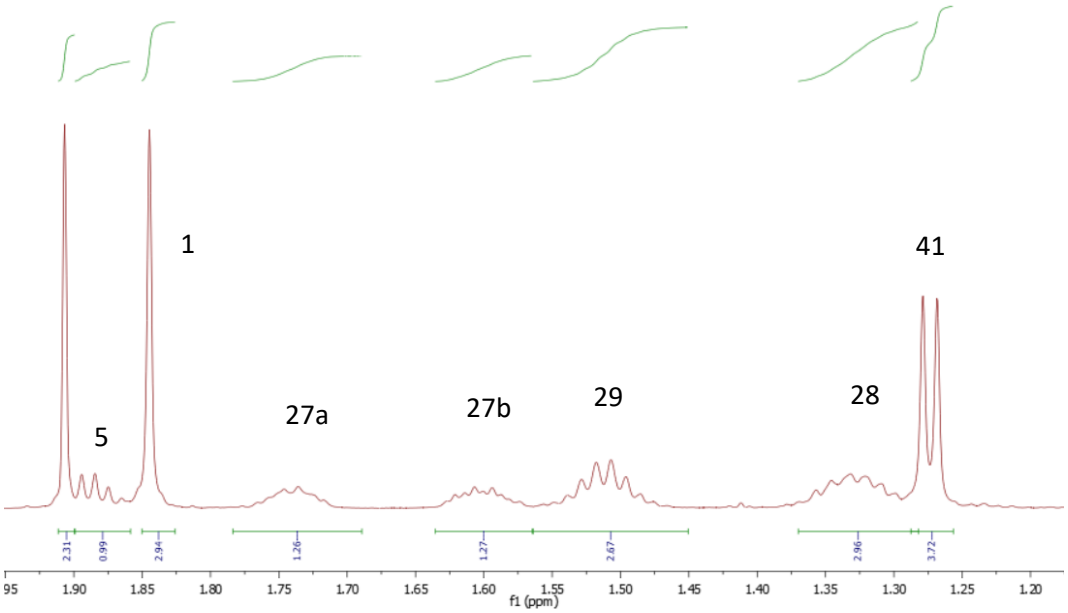


Figure 74: ^{1}H NMR spectrum of **9,** showing ^{1}H assignments between δ 1.25 – 1.90 ppm (DMSO, 700 MHz, 300 K)

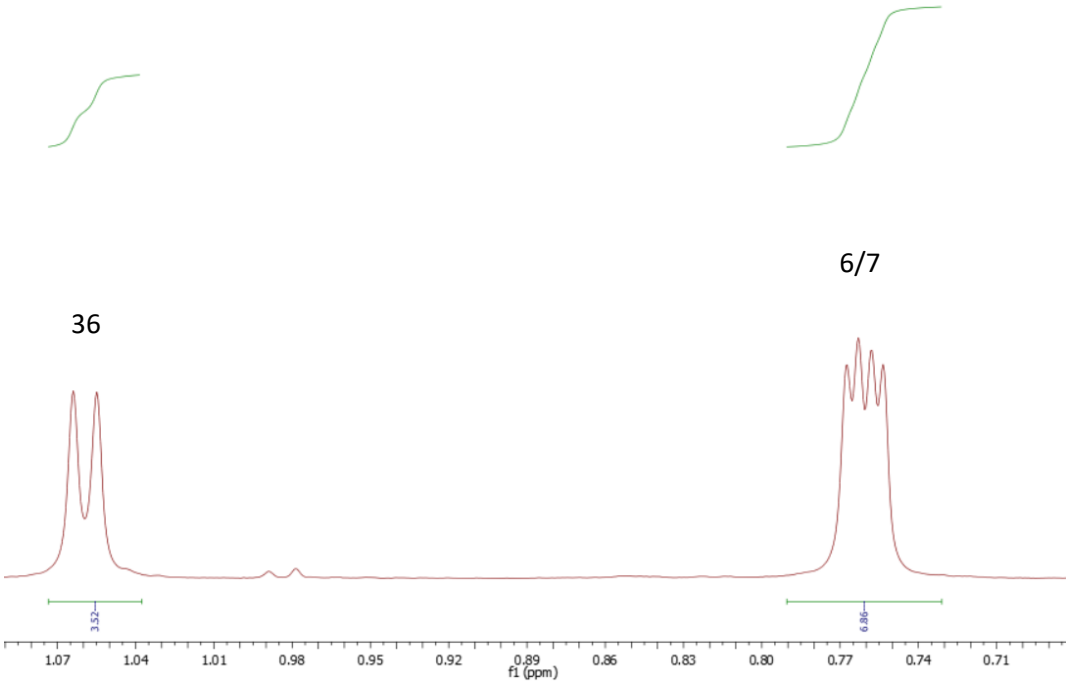


Figure 75: ^{1}H NMR spectrum of **9,** showing ^{1}H assignments between δ 0.72 – 1.08 ppm (DMSO, 700 MHz, 300 K)

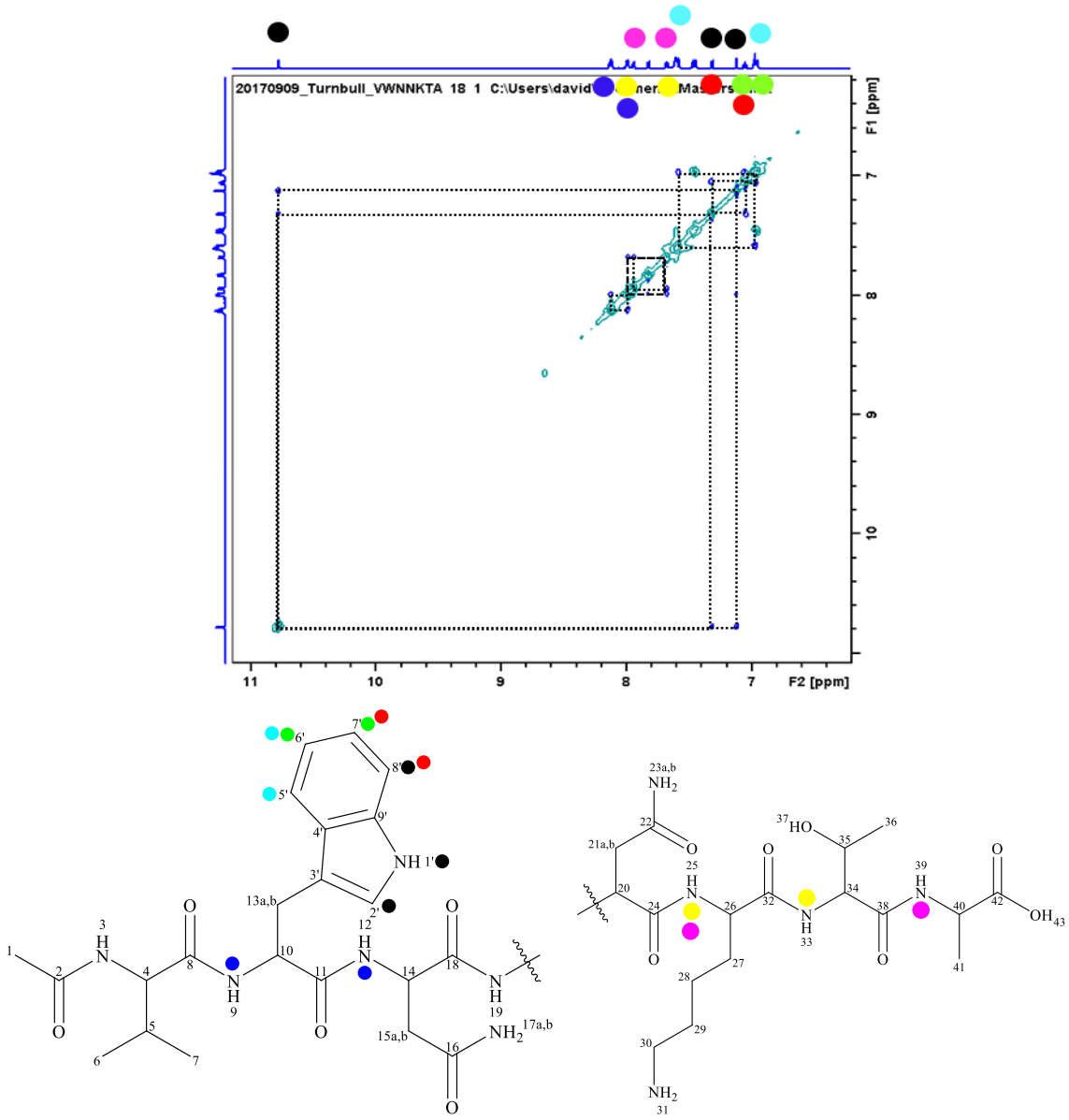


Figure 76: ROESY Spectrum of aromatic region of **9** – examination of the indole region (DMSO, 700 MHz, 300 K)

The spectra above show the usual interactions of the aromatic protons but with the additional interaction of the indole nitrogen proton and position 8' proton, similar to the interactions seen in Ala-Trp-Ala and Ser-Gly-Trp-Ala. Similar to Ser-Gly-Trp-Ala, there are interactions between several of the amide protons, mainly 25 to 39 and 25 to 33 or 9. These interactions are unusual due to the distance between the atoms but this indicates that there is a folding effect in molecule. There is an interaction between 9 or 33 to 19 or 12 , in regards to this the interaction must be between 9 and 12.

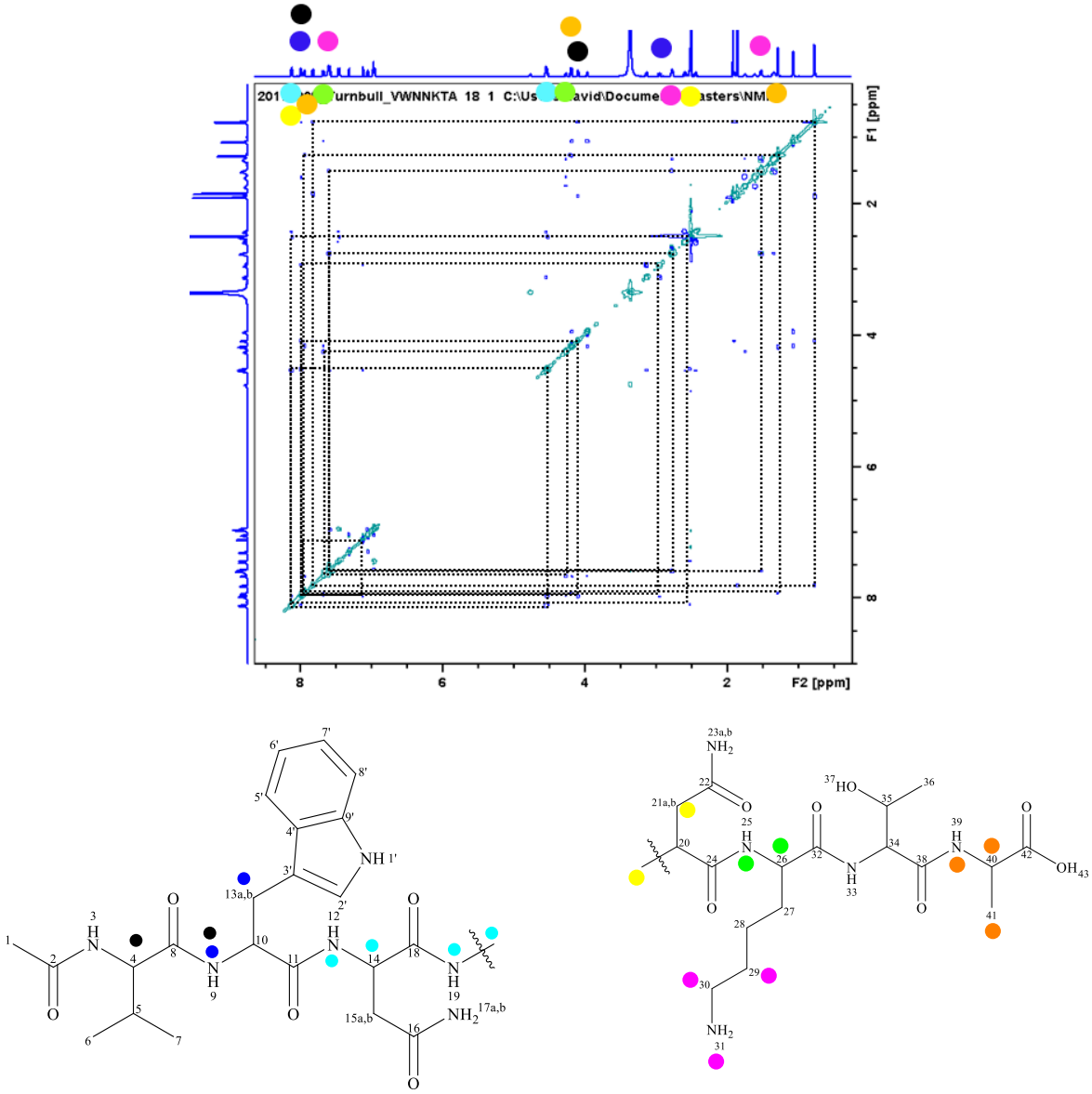


Figure 77: ROESY Spectrum of aromatic region of **9** – examination of the peptide chain (DMSO, 700 MHz, 300 K)

The interactions in this molecule are plentiful, there are interactions between position 19 and 20 and 12 and 14, this is to be expected from adjacent atoms. The amide proton at position 9 is interacting with the diastereomeric protons on the tryptophan residue at position 13 but only with one of the protons. The remaining interactions are between adjacent atoms which you would expect to see.

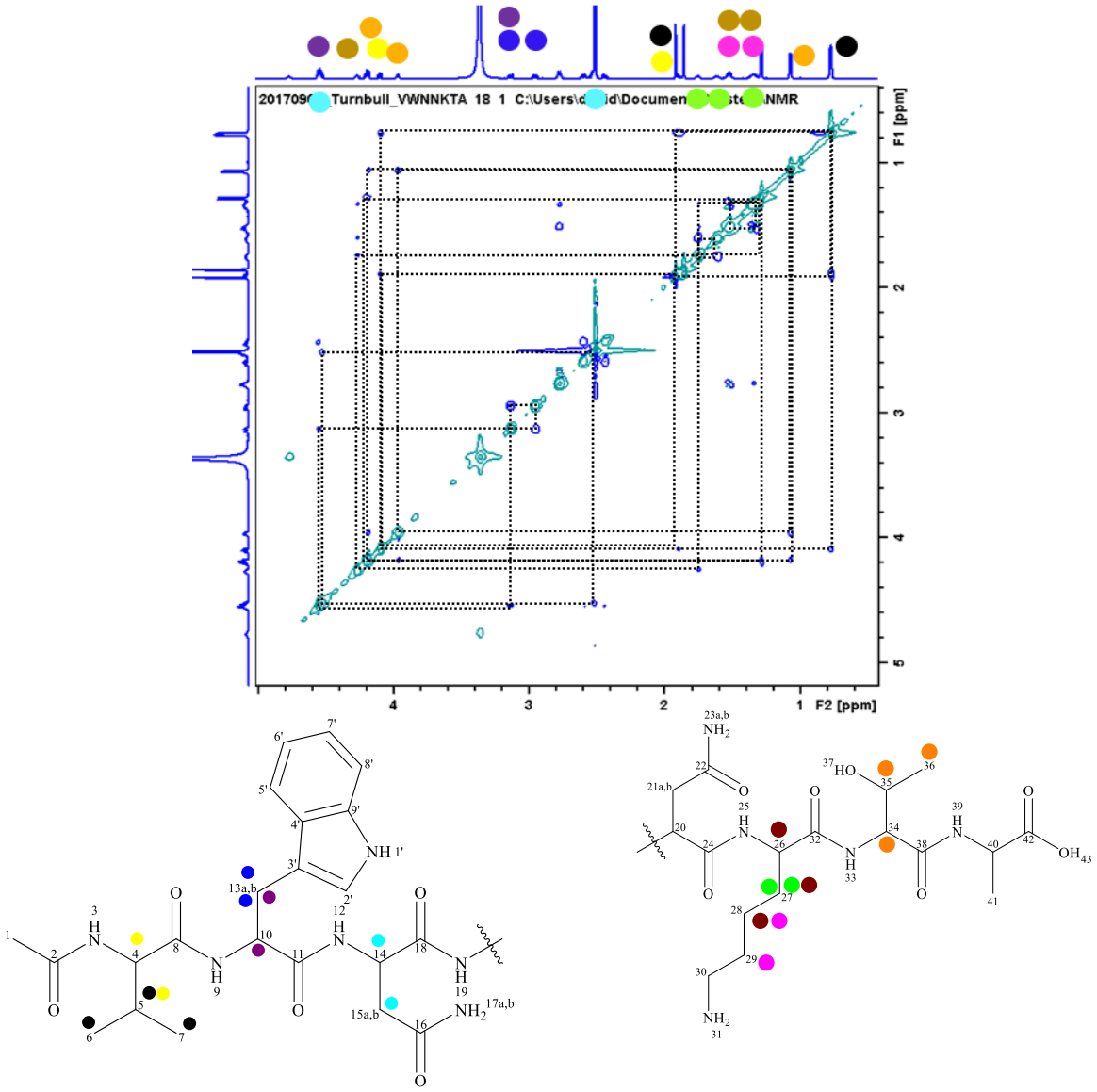


Figure 78: ROESY Spectrum of aromatic region of **9** – examination of the peptide backbone (DMSO, 700 MHz, 300 K)

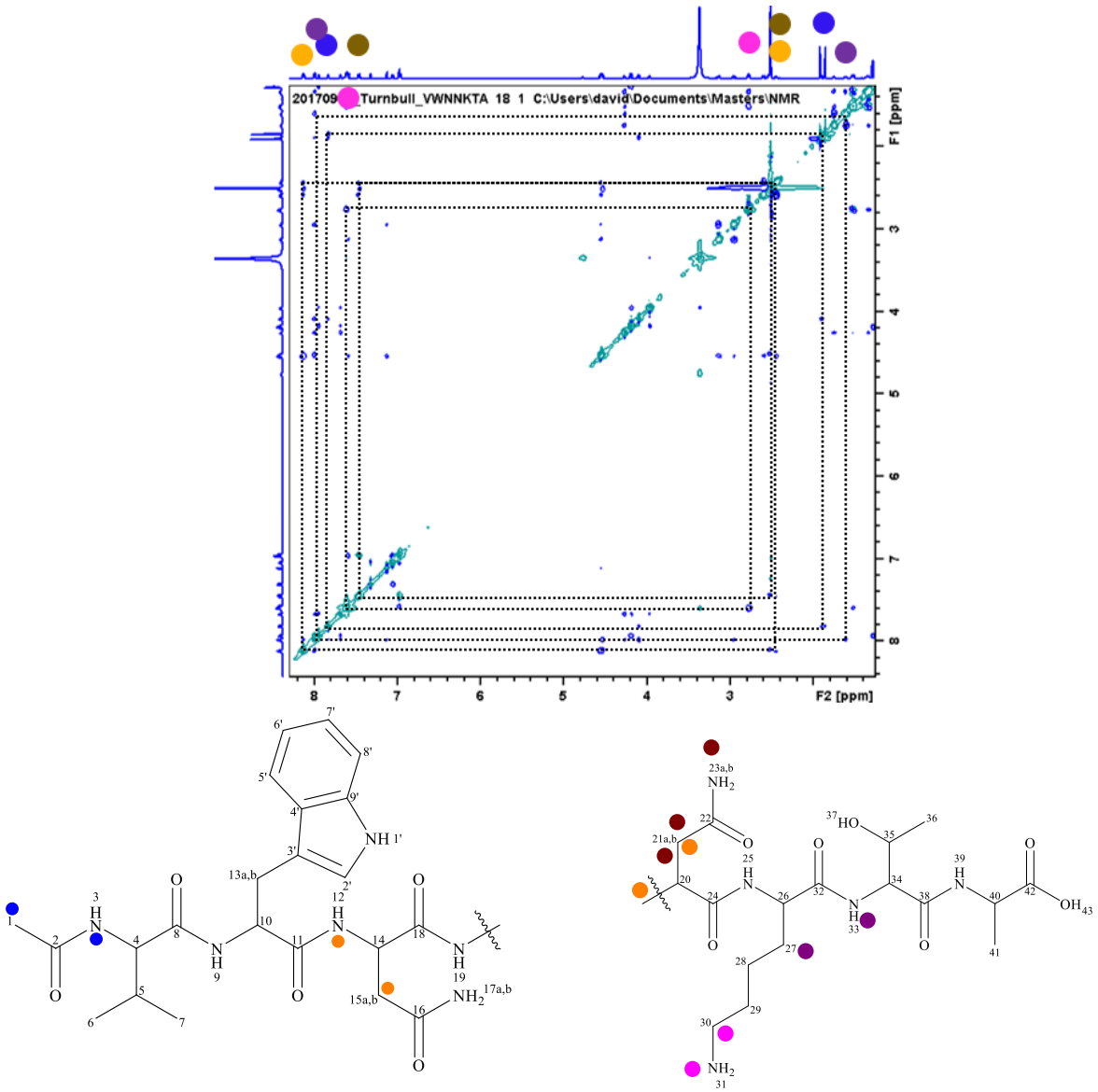


Figure 79: ROESY Spectrum of aromatic region of **9** – examination of the peptide backbone 2 (DMSO, 700 MHz, 300 K)

There is an interaction between the proton at position 10 and one of the diastereomeric protons on the tryptophan residue, 13. There is also an interaction between 14 and one of the diastereomeric protons on the asparagine residue, 15, and one between 20 and one of the diastereomeric protons on 21. The proton at position 5 links to the protons at position 6 and 7 as well as position 4. The diastereomeric protons at position 27 interact with themselves as well as with the proton on the α-carbon of the lysine at position 26. There is also interaction between 27, 28, 29 and 30 which indicates a contorted lysine residue. The amide protons at position 19 and 12 interact with both of the diastereomeric protons on position 15 and 21. There is an interaction with 21 and 23 which is unusual due to the positions. The proton at position 36 is close to the protons at 35 and 34, this is expected but the lack of interaction between position 37 with any other proton is interesting, with possible hydrogen bonding between the proton and the carboxyl group at 32.

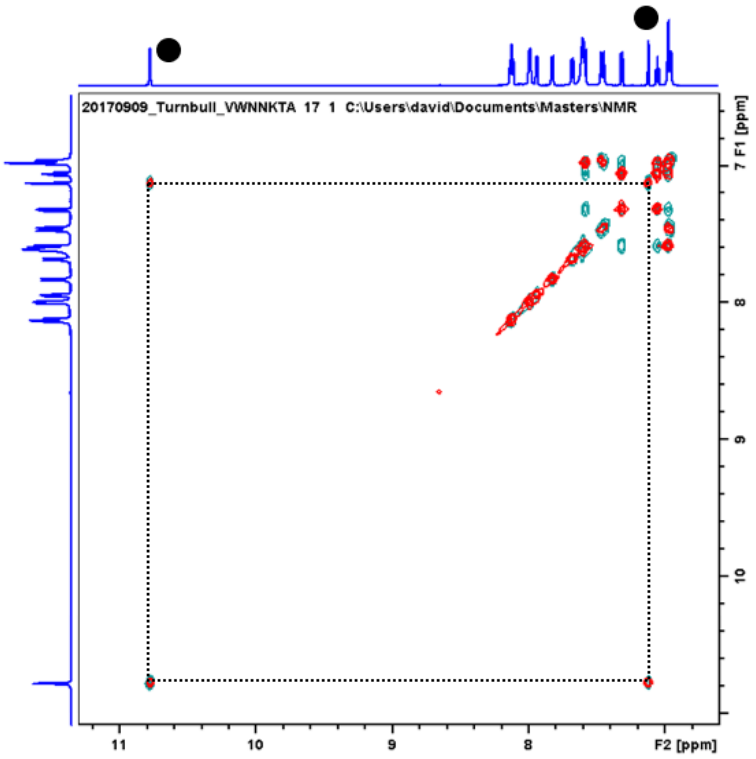


Figure 80: TOCSY (Blue) and COSY (Red) spectrum of **9** – examination of the indole ring on the nitrogen (DMSO, 700 MHz, 300 K)

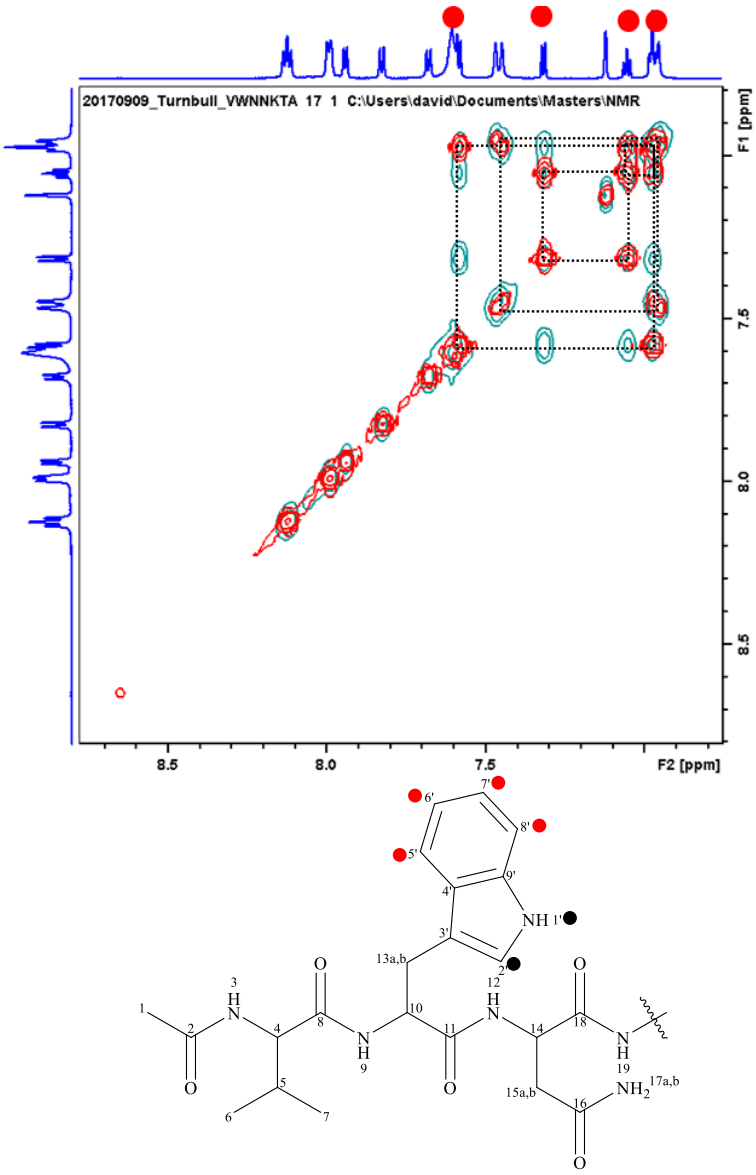


Figure 81: TOCSY (Blue) and COSY (Red) spectrum of **9** – examination of the indole group (DMSO, 700 MHz, 300 K)

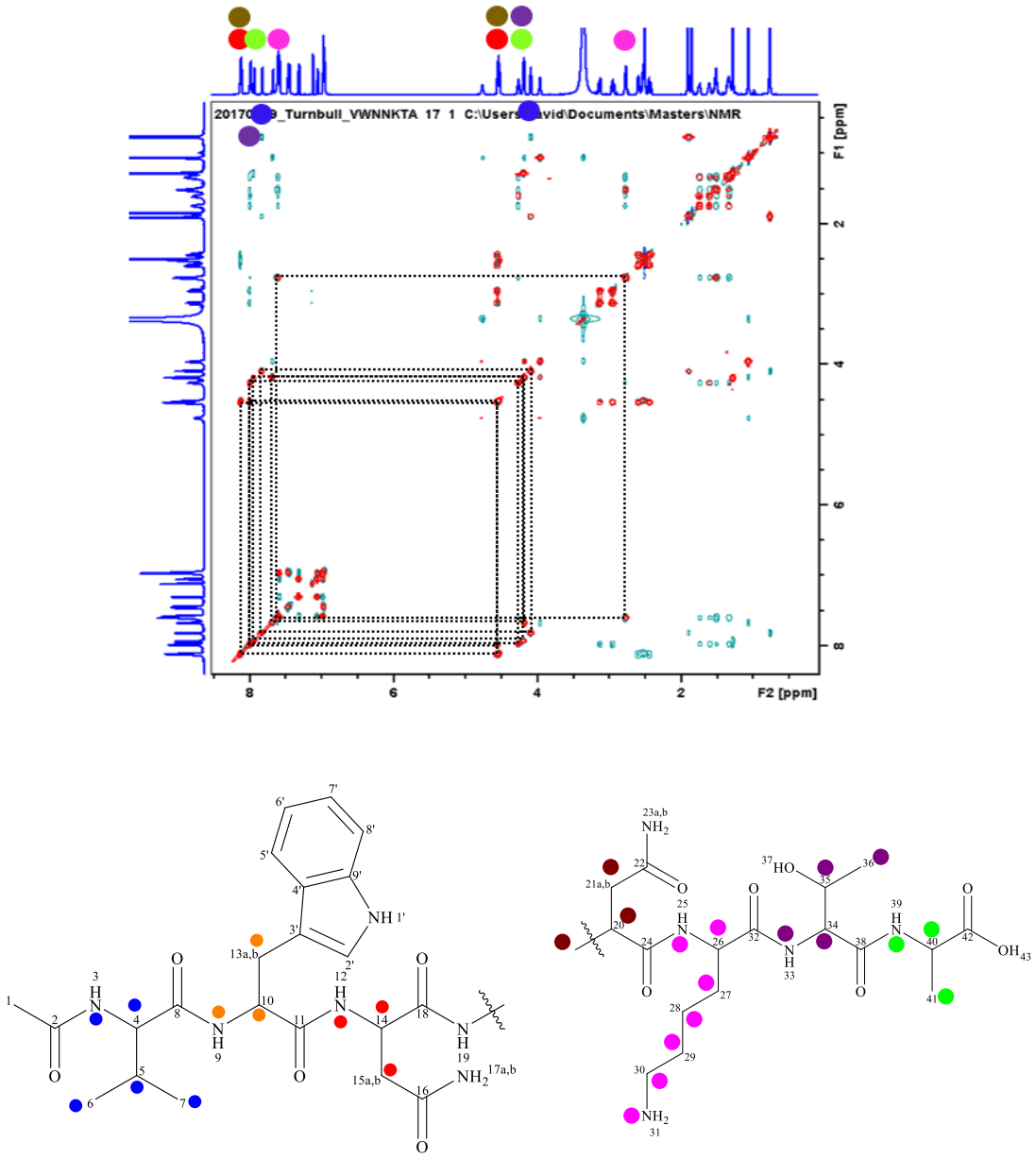


Figure 82: TOCSY (Blue) and COSY (Red) spectrum of **9** – examination of the peptide chain (DMSO, 700 MHz, 300 K)

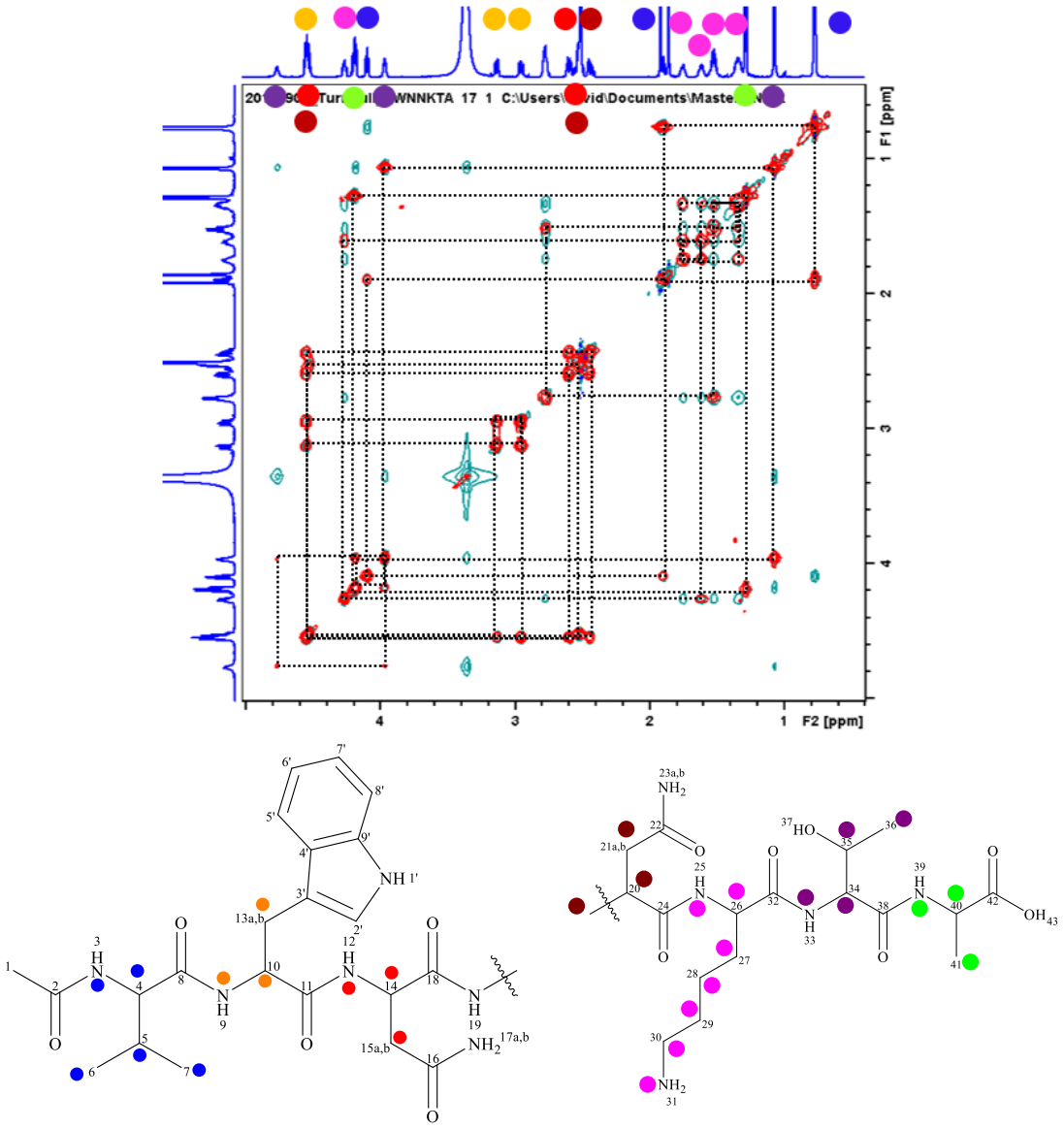


Figure 83: TOCSY (Blue) and COSY (Red) spectrum of **9** – examination of the aliphatic region (DMSO, 700 MHz, 300 K)

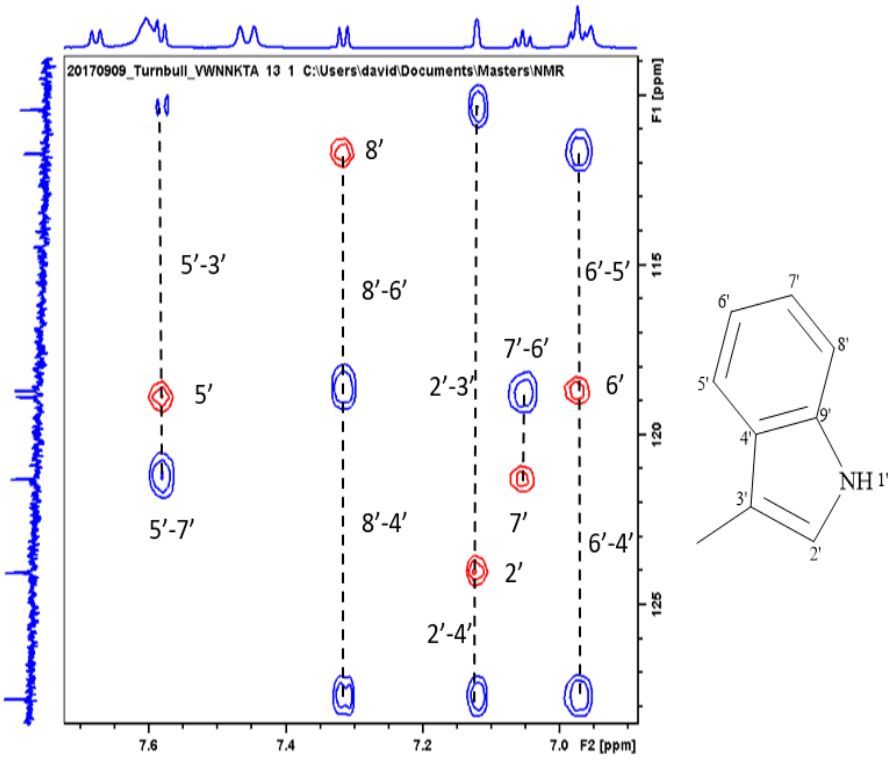
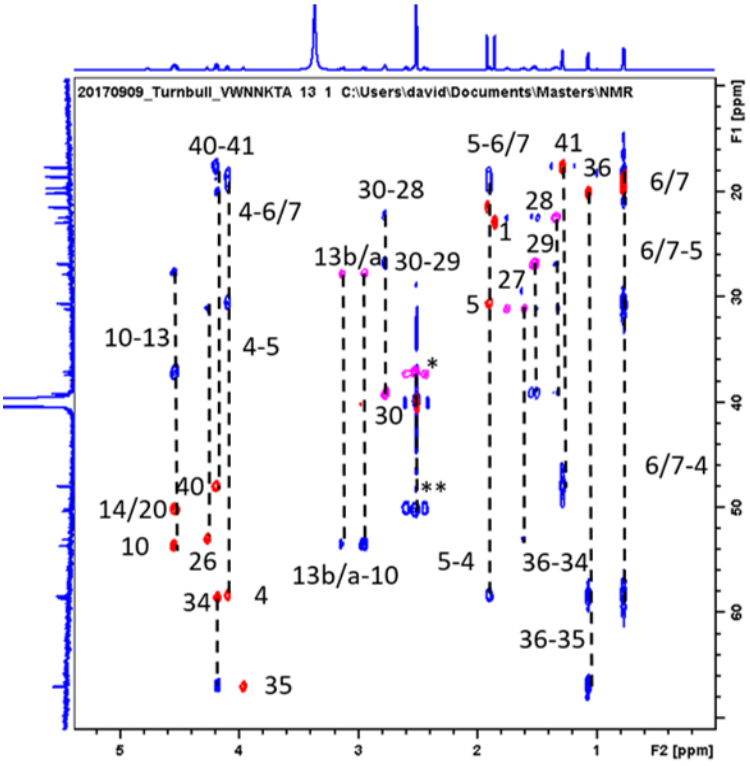


Figure 84: HSQC and HMBC spectrum of the aromatic region of **9** - examination of the indole group (DMSO, for 1H - 700 MHz, for ^{13}C - 176 MHz, 300 K)



*15b,a and 21b,a

**15b,a/21b,a-14/20

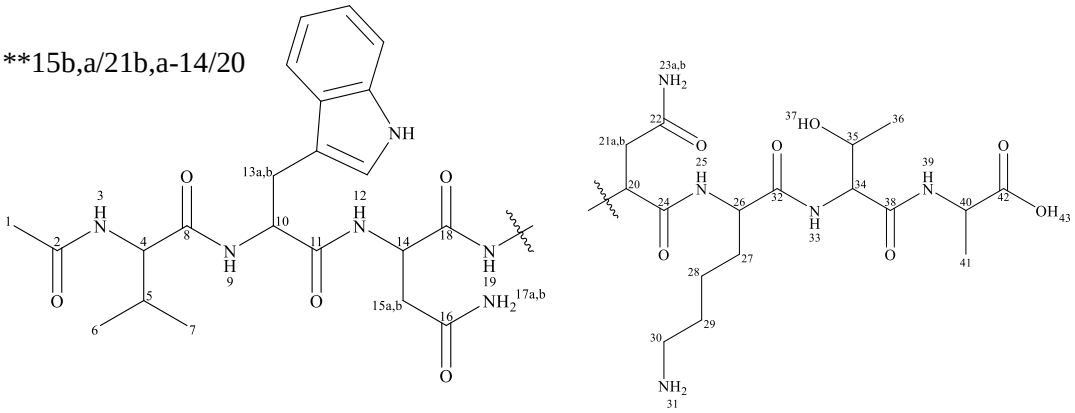


Figure 85: HSQC and HMBC spectrum of the aliphatic region of **9** - examination of the aliphatic region (DMSO, for 1H - 700 MHz, for `^{13}C - 176 MHz, 300 K)

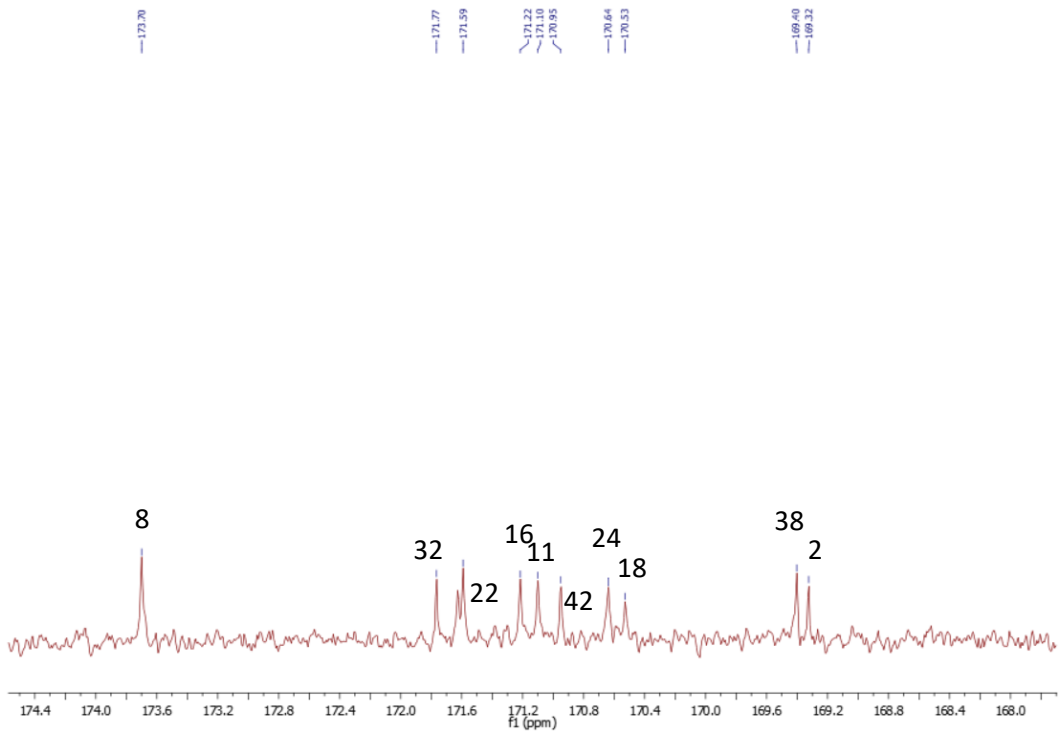


Figure 86: ^{13}C NMR spectrum of **9**, ^{13}C assignments between δ 168.8 – 174.0 ppm (DMSO, 176 MHz, 300K)

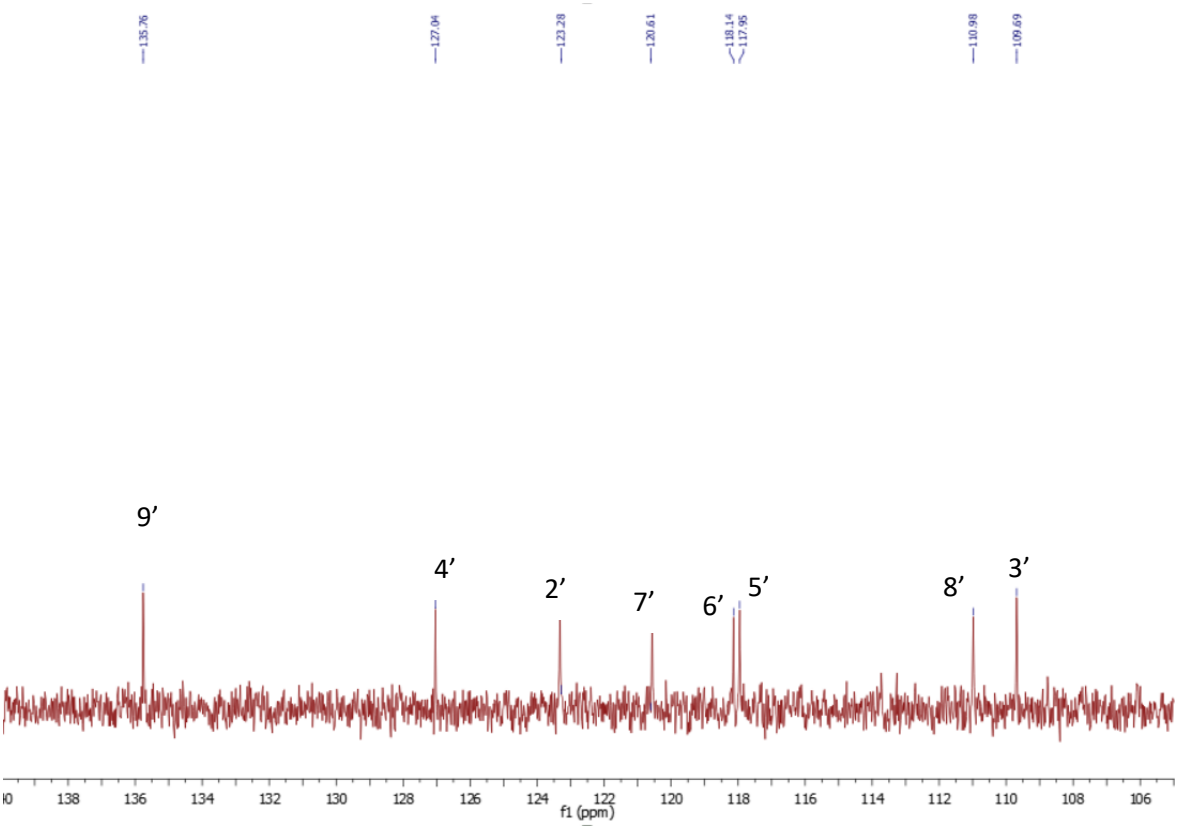


Figure 87: ^{13}C NMR spectrum of **9**, ^{13}C assignments between δ 105.0 – 138.0 ppm (DMSO, 176 MHz, 300K)

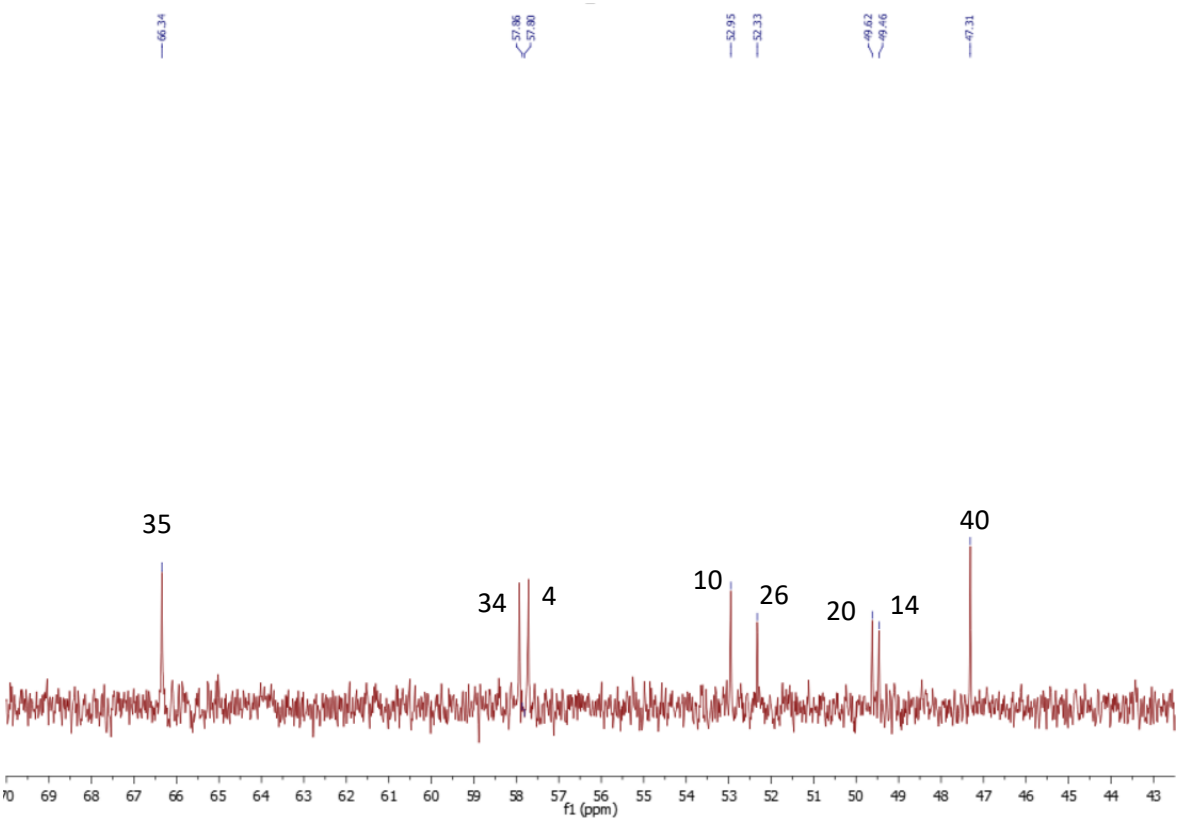


Figure 88: ^{13}C NMR spectrum of **9**, ^{13}C assignments between δ 45 - 67 ppm (DMSO, 176 MHz, 300K)

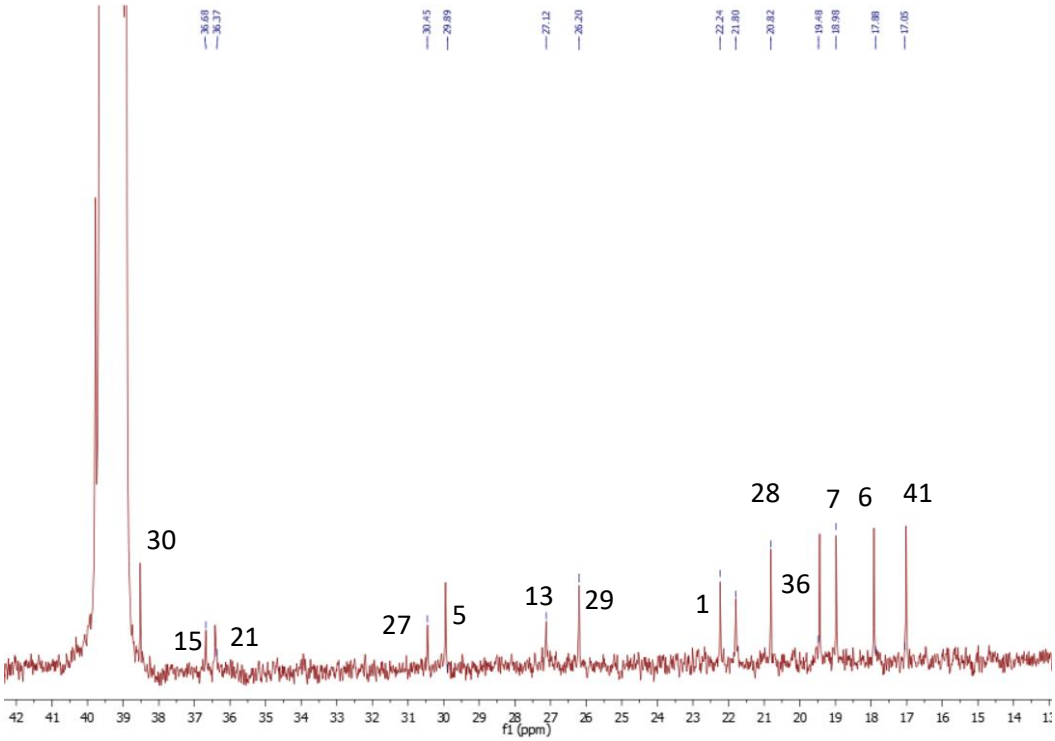


Figure 89: ^{13}C NMR spectrum of **9**, ^{13}C assignments between δ 17 -40 ppm (DMSO, 176 MHz, 300K)

AcNH-Ala-Trp(Ph)-Ala-OMe (10)

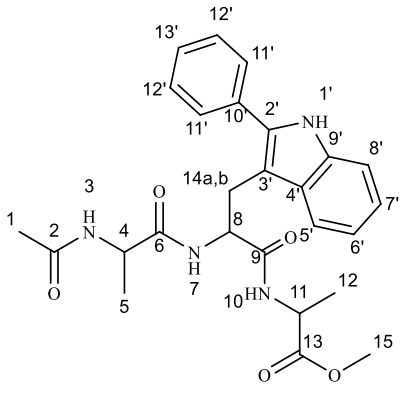


Figure 90: Reference molecule for Ala-Trp(Ph)-Ala (**10**)

NMR spectrum data for **10**

^{1}H NMR (700 MHz, DMSO) δ 11.16 (s, 1H), 8.02 – 7.98 (m, 2H), 7.88 (d, *J* = 8.4 Hz, 1H), 7.68 (d, *J* = 7.7 Hz, 3H), 7.48 (t, *J* = 7.6 Hz, 2H), 7.37 (t, *J* = 7.4 Hz, 1H), 7.33 (d, *J* = 8.0 Hz, 1H), 7.08 (t, *J* = 7.5 Hz, 1H), 6.98 (t, *J* = 7.5 Hz, 1H), 4.61 (dd, *J* = 14.9, 7.6 Hz, 1H), 4.23 – 4.17 (m, 1H), 4.10 (dt, *J* = 13.8, 7.0 Hz, 2H), 3.48 (s, 3H), 3.33 (dd, *J* = 14.6, 6.5 Hz, 1H), 3.10 (dd, *J* = 14.5, 7.8 Hz, 1H), 1.79 (s, 3H), 1.19 (d, *J* = 7.3 Hz, 4H), 1.01 (d, *J* = 7.1 Hz, 3H).

^{13}C NMR (176 MHz, DMSO) δ 172.53, 171.96, 170.90, 169.51, 135.93, 135.27, 132.75, 128.99, 128.59, 128.11, 127.29, 121.38, 119.29, 118.65, 110.94, 107.62, 53.72, 51.74, 48.59, 47.63, 27.45, 22.48, 17.51, 16.98.

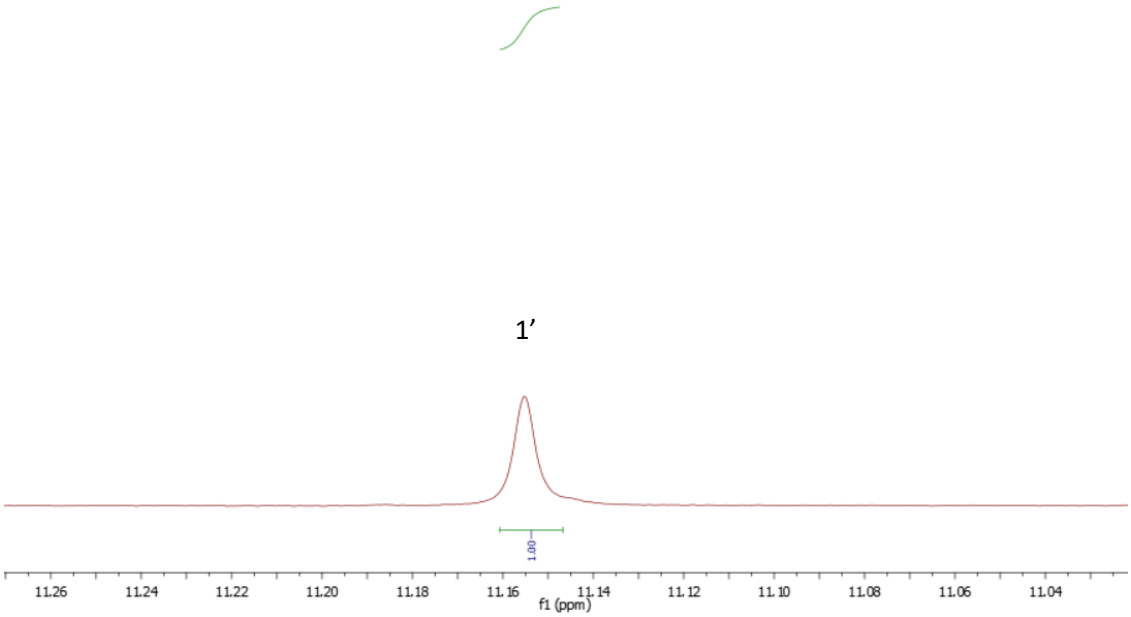


Figure 91: ^{1}H NMR spectrum of **10,** showing ^{1}H assignments between δ 11.1 – 11.2 ppm (DMSO, 700 MHz, 300 K)

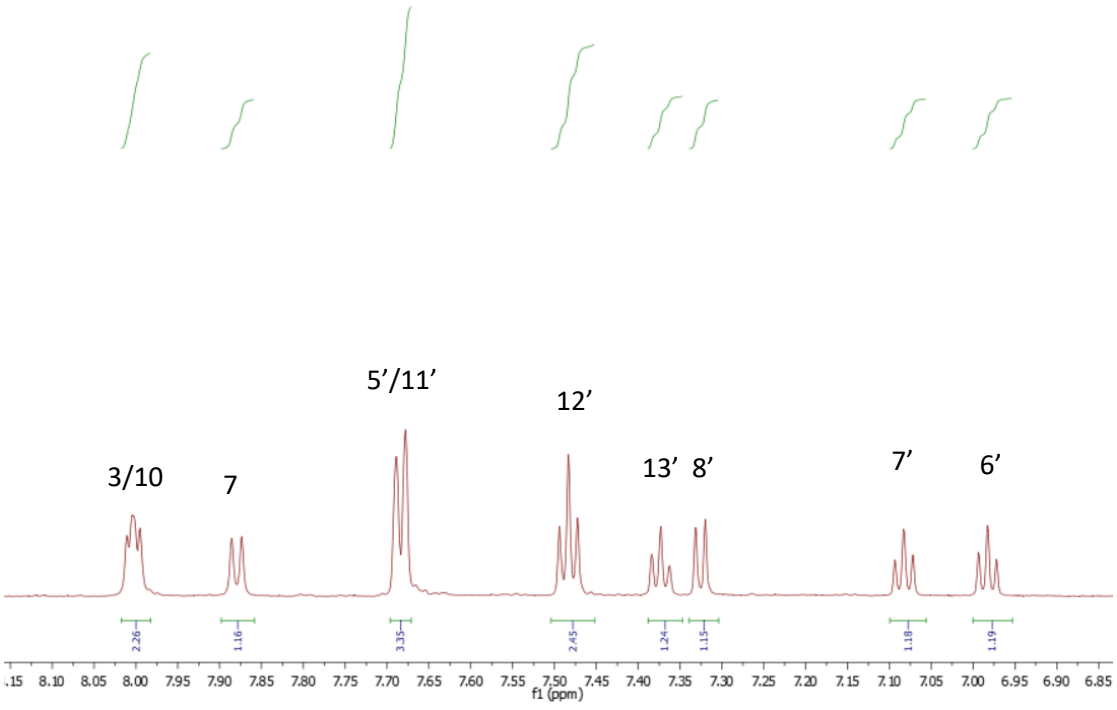


Figure 92: ^{1}H NMR spectrum of **10,** showing ^{1}H assignments between δ 6.95- 8.05 ppm (DMSO, 700 MHz, 300 K)

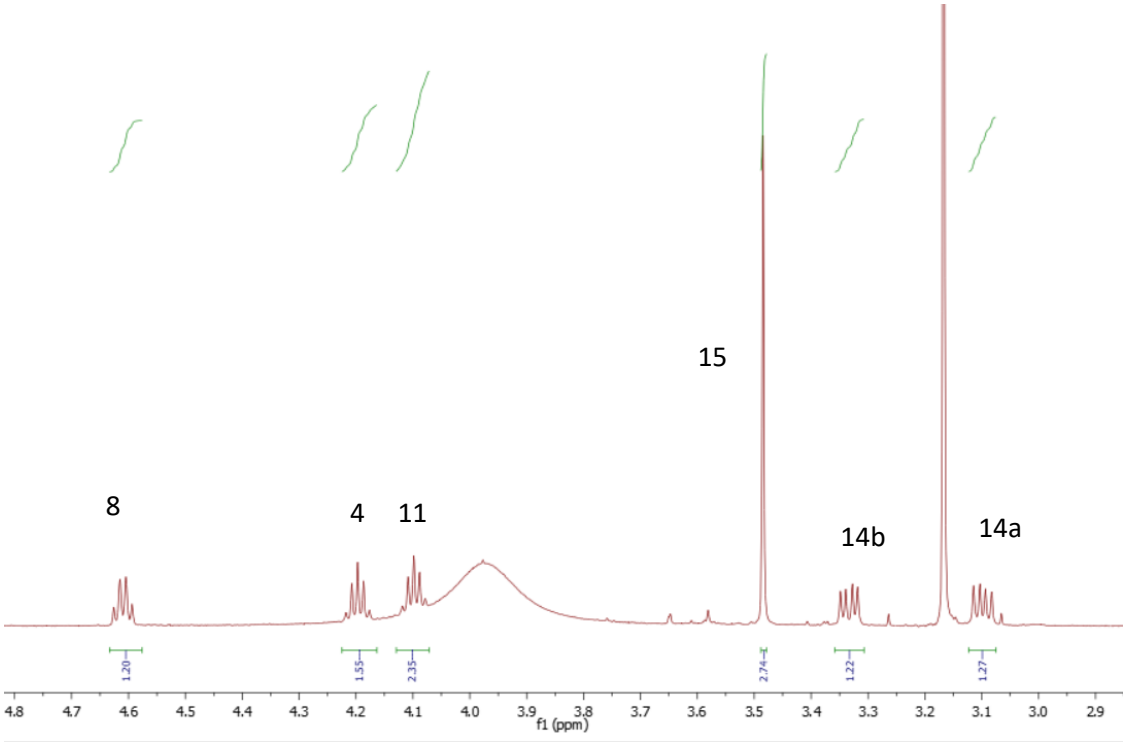


Figure 93: ^{1}H NMR spectrum of **10,** showing ^{1}H assignments between δ 3.0 – 4.7 ppm (DMSO, 700 MHz, 300 K)

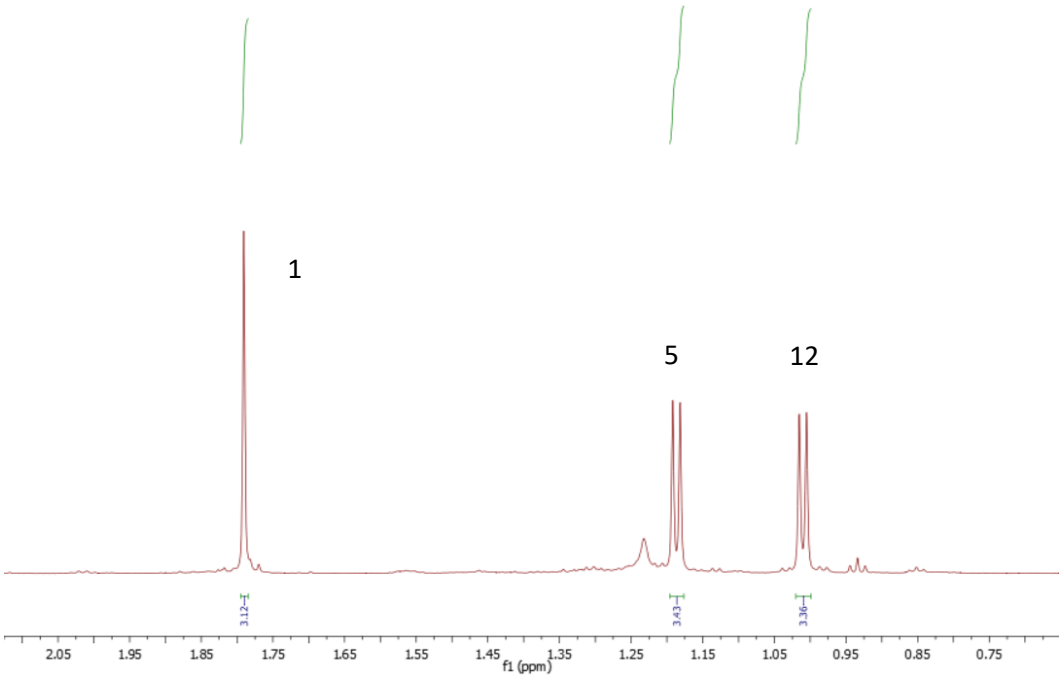


Figure 94: ^{1}H NMR spectrum of **10,** showing ^{1}H assignments between δ 0.95 – 1.85 ppm (DMSO, 700 MHz, 300 K)

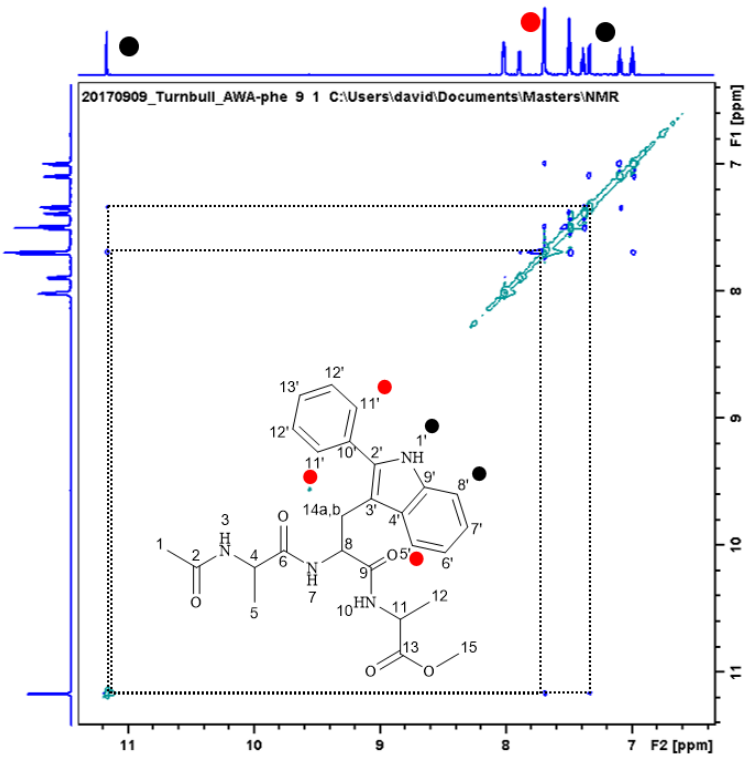


Figure 95: ROESY Spectrum of aromatic region of **10** – examination of the indole region (DMSO, 700 MHz, 300 K)

There are interactions between the indole nitrogen proton at position 1' and the newly inserted aromatic ring on position 11' or the proton at position 5', and there is also an interaction between position 1' and position 8'. This is similar to the unarylated version but could be due to the proton being slightly positioned away from the new aromatic ring and towards the 8 proton.

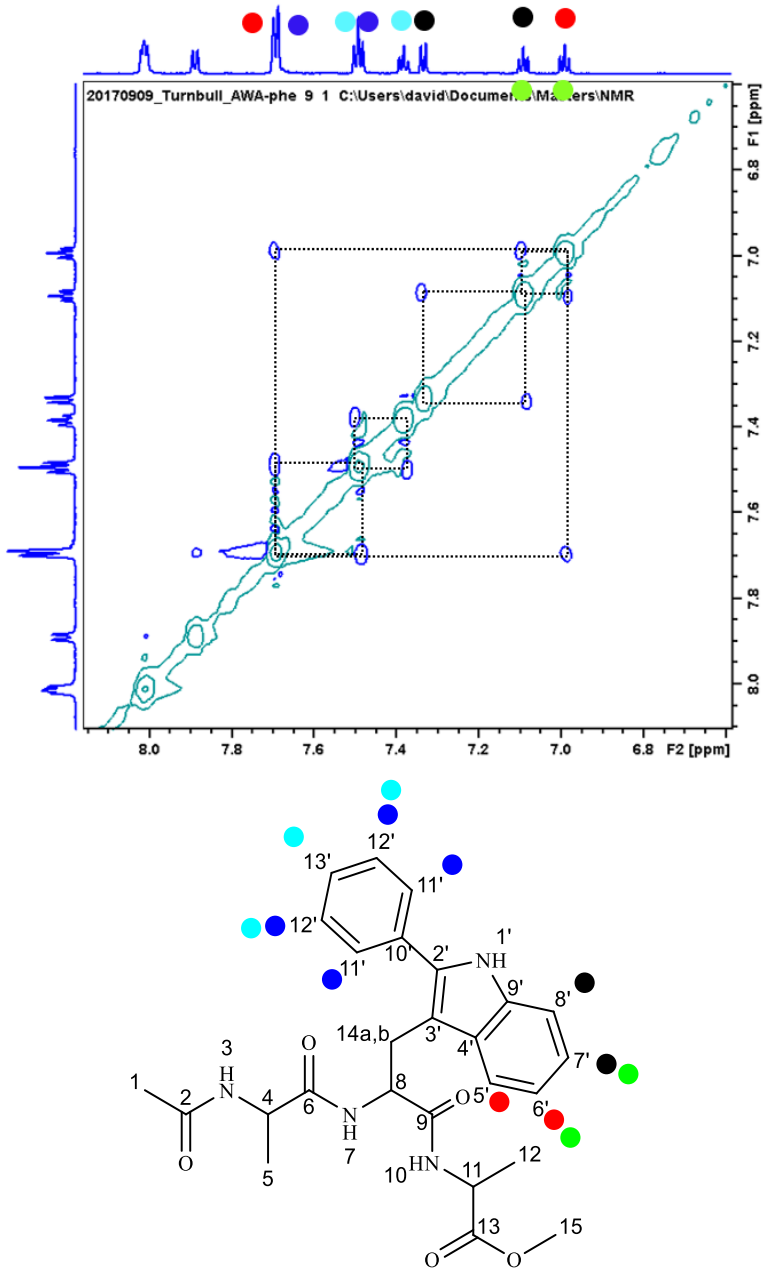


Figure 96: ROESY Spectrum of aromatic region of **10** – examination of the indole region part 2(DMSO, 700 MHz, 300 K)

The interactions between the aromatic protons show interaction between the indole protons at position 6’ and 7’, 8’ and 7’ and 5’ and 6’. The newly added aryl protons show interactions between themselves at position 11’ to 12’, and 12’ to 13’. All these interactions are to be expected.

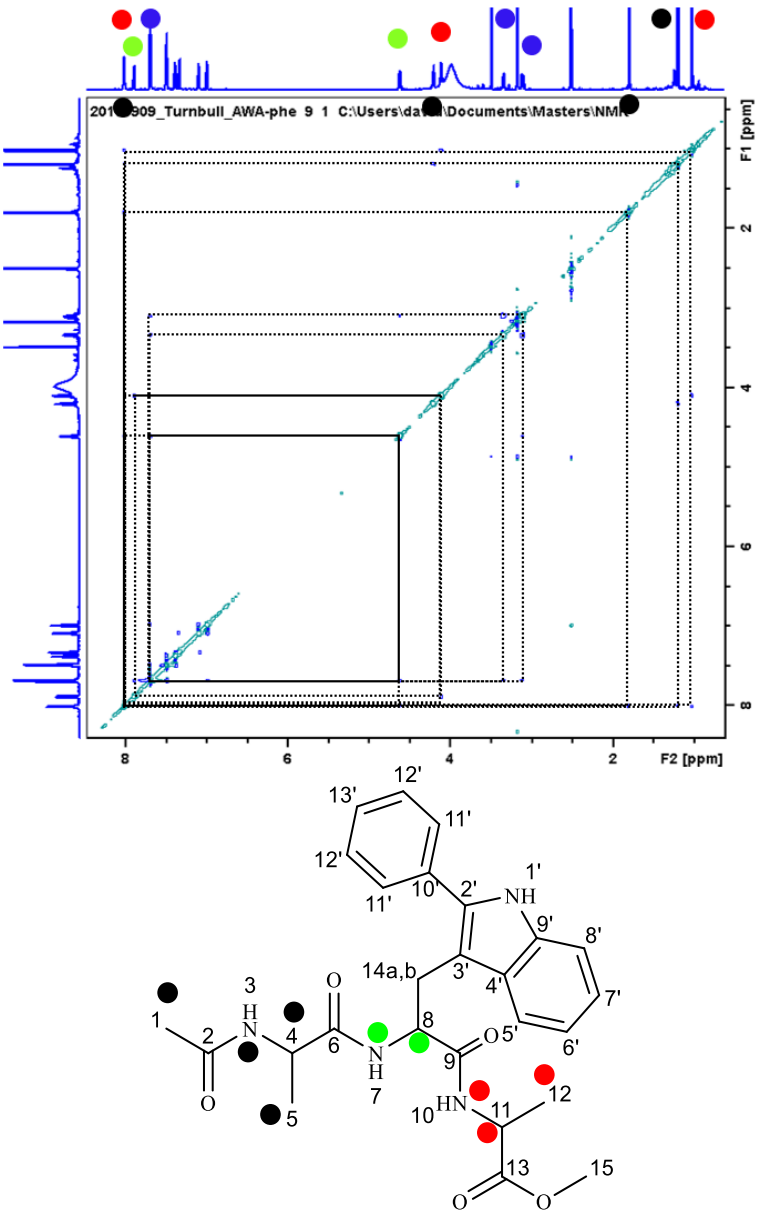


Figure 97: ROESY Spectrum of **10** – examination of the peptide chain (DMSO, 700 MHz, 300 K)

There are some interesting interactions in the spectra above, the proton at position 5 on the indole ring or 11 on the added aryl ring interacts with both of the diastereomeric protons at position 14 meaning that one of the protons is most likely positioned directly between the diastereomeric protons. There is also an interaction between 5 or 11 with the α-carbon proton at position 8. With these sets of interactions, it is likely that one interacts with the diastereomeric protons and the other with the α-carbon proton. The amide protons at position 3 and 10 have interactions with the α-carbon proton at position 4 and 11 respectively which is expected. The proton at position 3 also has an interaction at position 1. There is an interaction between the amide proton. The amide proton at position 3 or 10 interacts with the α-carbon proton at position 8, it is more likely position 10 than 3. The amide proton at position 7 interacts with the α-carbon proton at position 11.

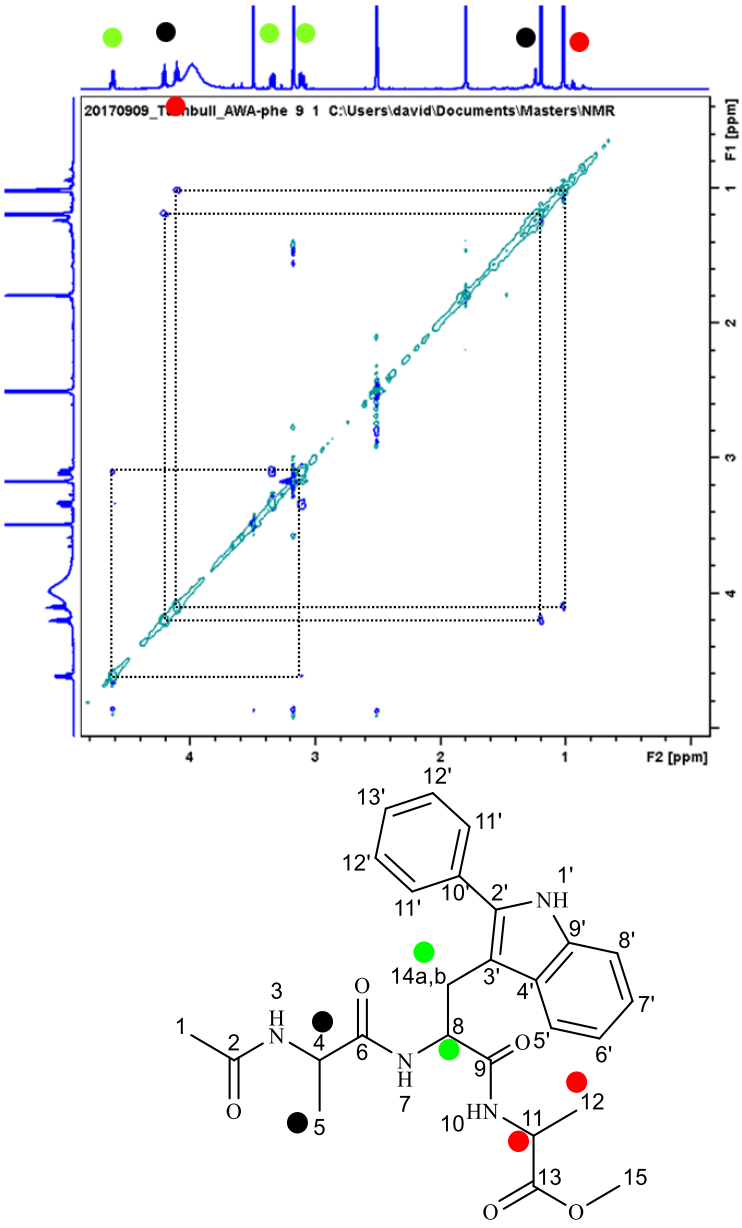


Figure 98: ROESY Spectrum of aliphatic region of **10** – examination of the peptide chain (DMSO, 700 MHz, 300 K)

From the spectra above, interactions can be seen between the tryptophan residue α-carbon proton at position 8 and one of the diastereomeric protons on the tryptophan residue β-carbon at position 14. There are interactions between the adjacent protons at positions 4 to 5 and 11 to 12 which is to be expected.

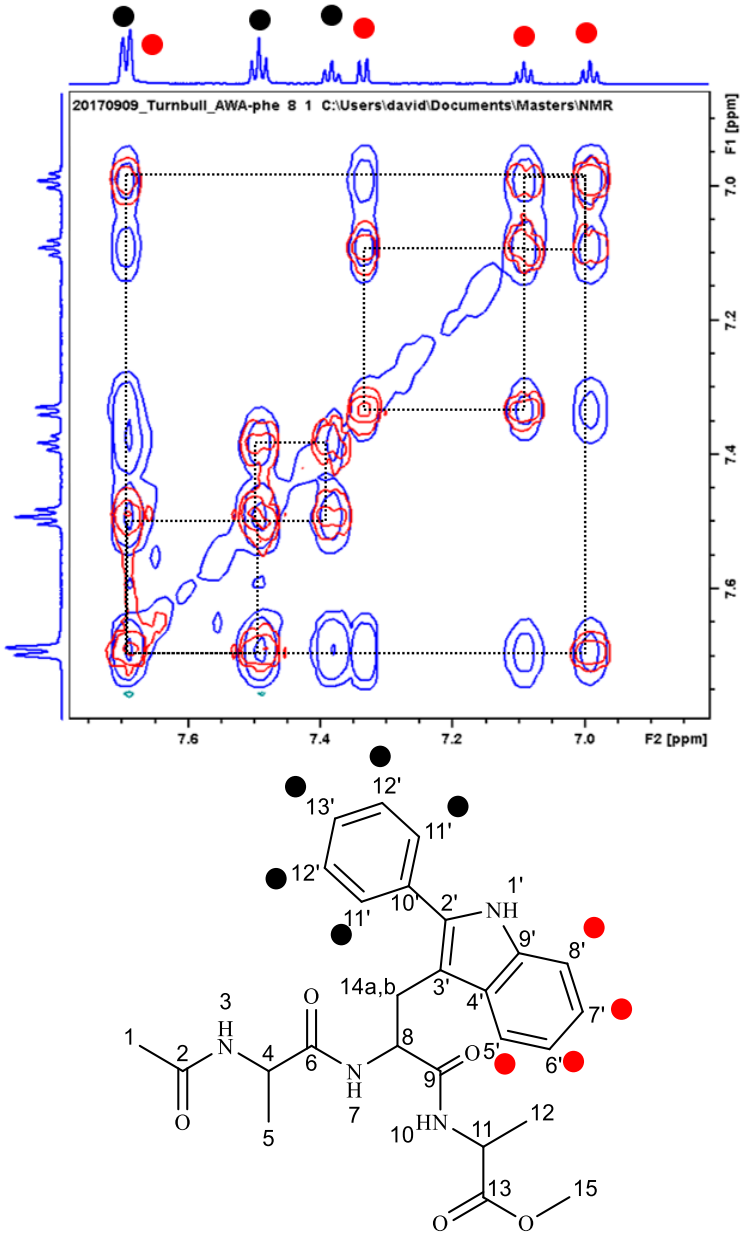


Figure 99: TOCSY (Blue) and COSY (Red) spectrum of **10** – examination of the indole ring (DMSO, 700 MHz, 300 K)

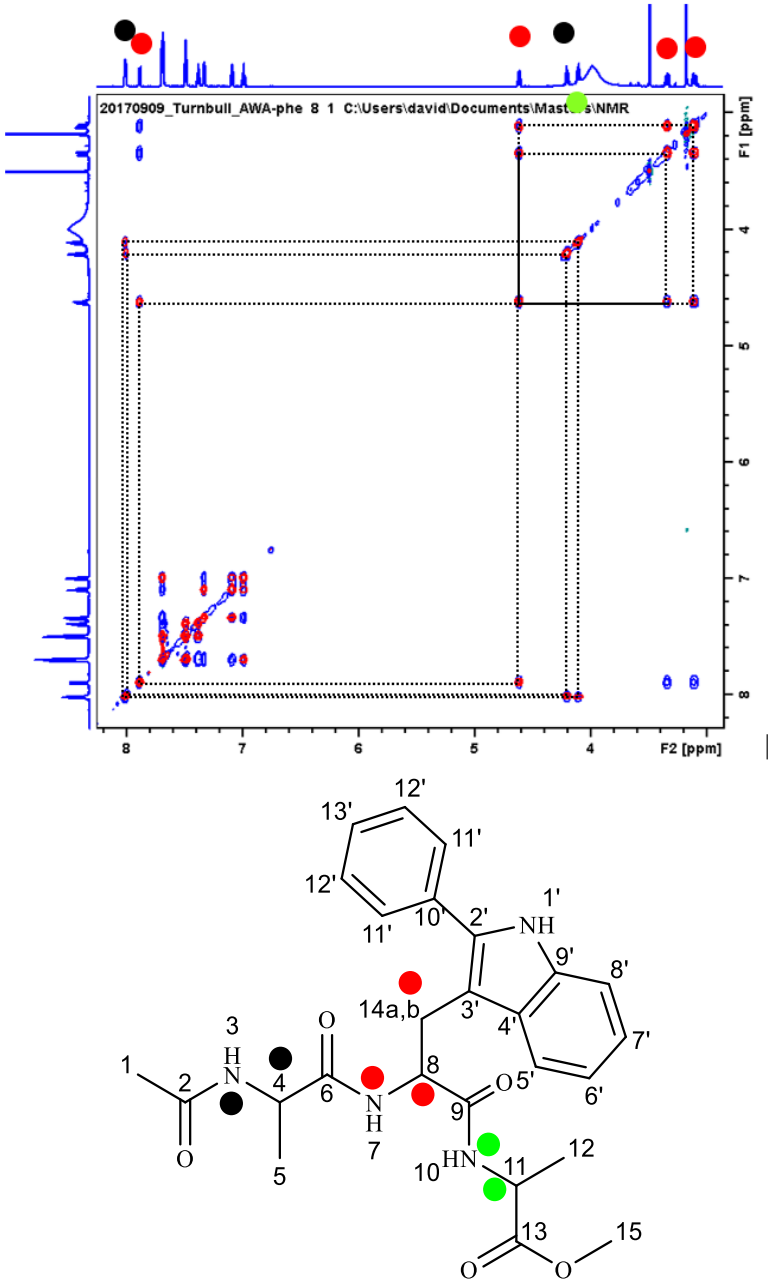


Figure 100: TOCSY (Blue) and COSY (Red) spectrum of **10** – examination of the peptide chain(DMSO, 700 MHz, 300 K)

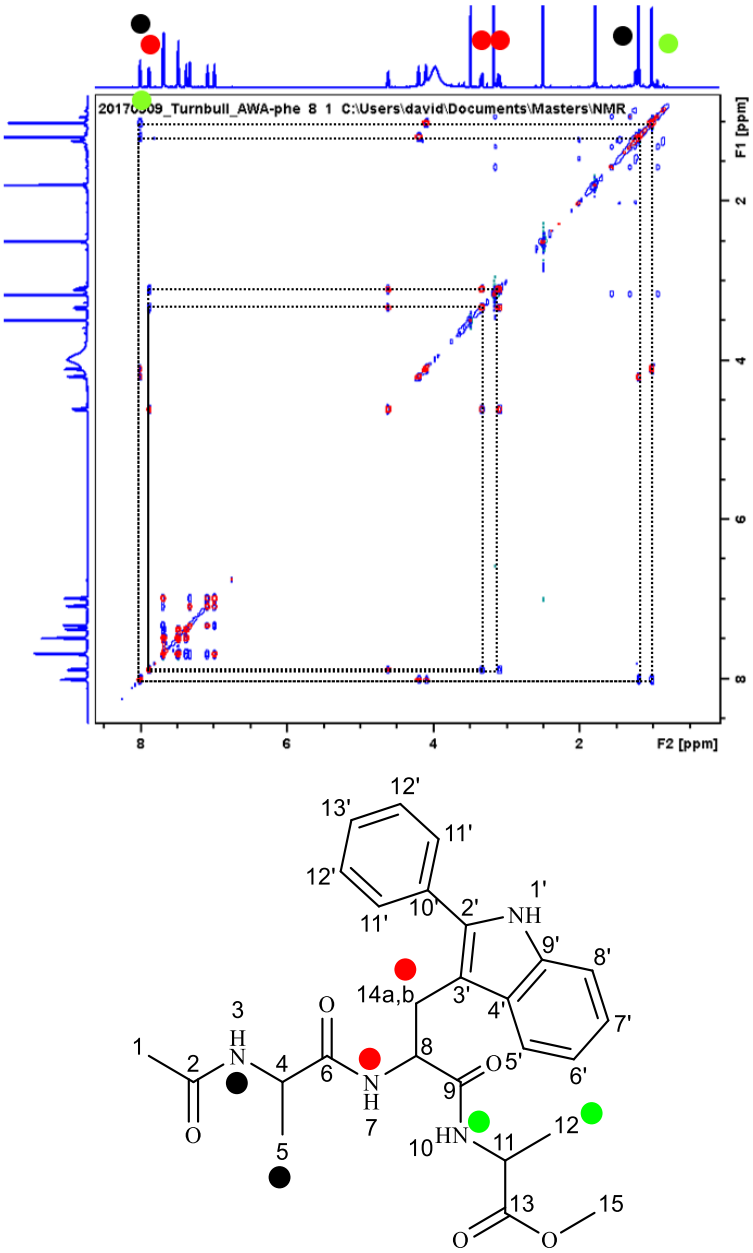


Figure 101: TOCSY (Blue) and COSY (Red) spectrum of **10** – examination of the peptide backbone (DMSO, 700 MHz, 300 K)

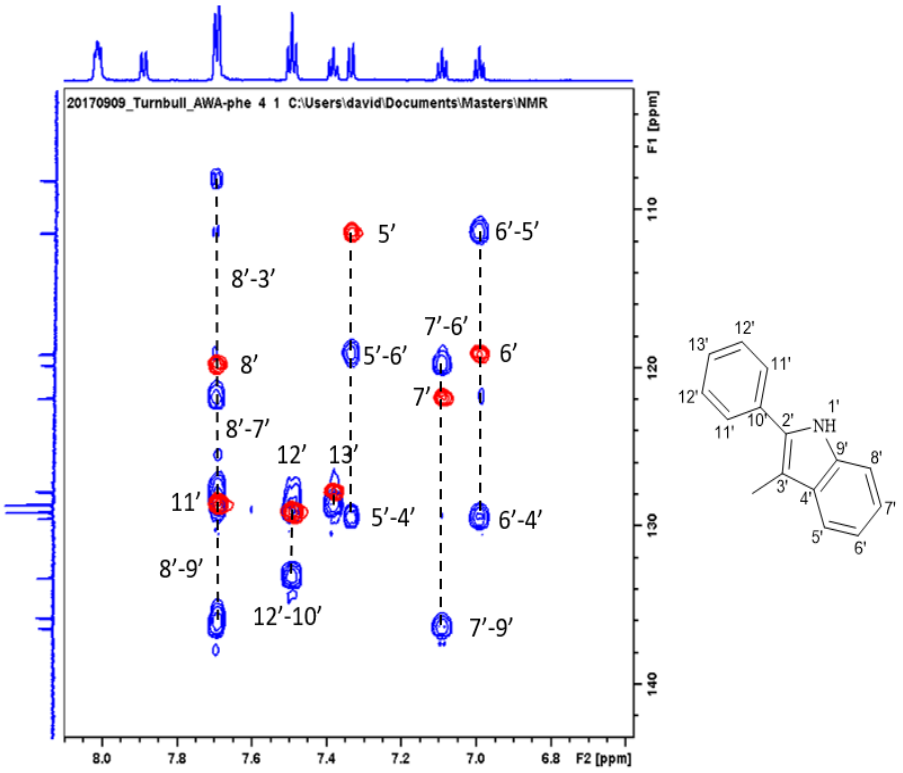


Figure 102: HSQC and HMBC spectrum of the aromatic region of **10** - examination of the indole ring (DMSO, for 1H - 700 MHz, for ^{13}C - 176 MHz, 300 K)

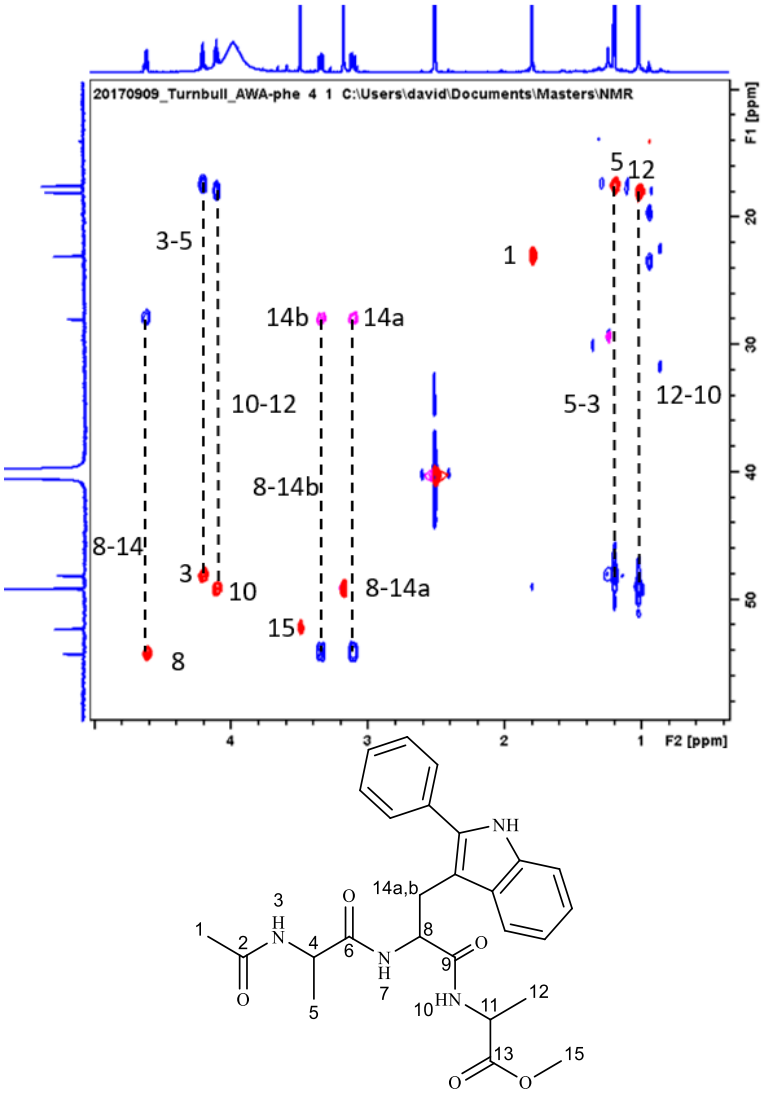


Figure 103: HSQC and HMBC spectrum of the aliphatic region of **10** - examination of the peptide backbone (DMSO, for 1H -700 MHz, for ^{13}C - 176 MHz, 300 K)

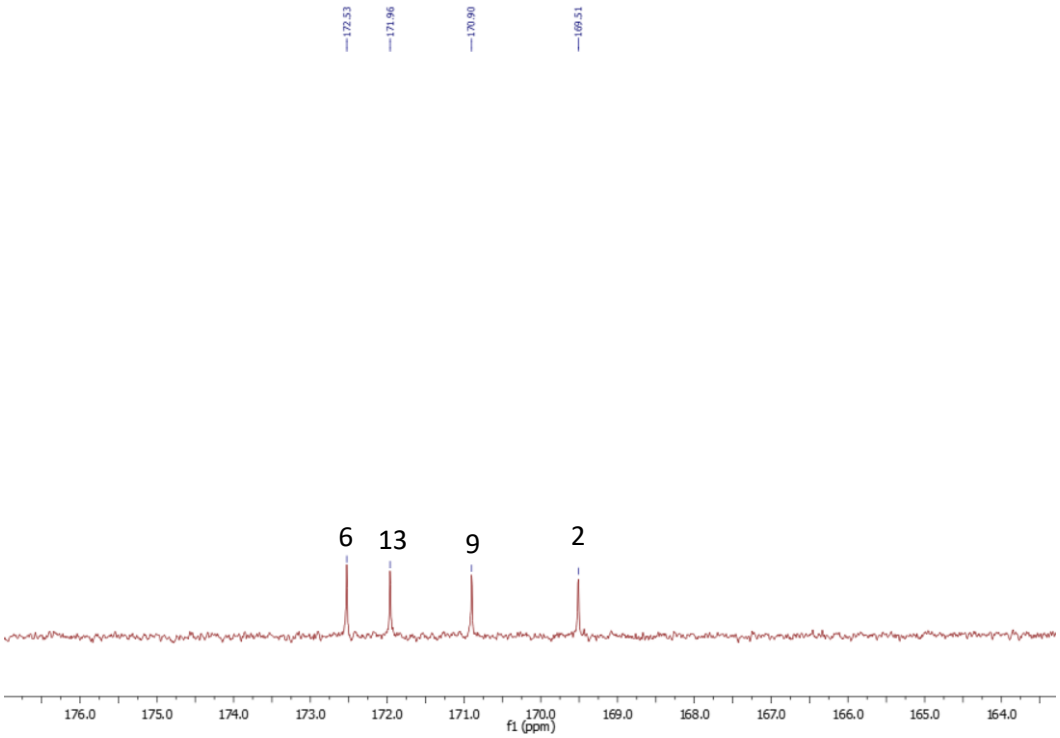


Figure 104: ^{13}C NMR spectrum of **10**, ^{13}C assignments between δ 169 - 173 ppm (DMSO, 176 MHz, 300K)

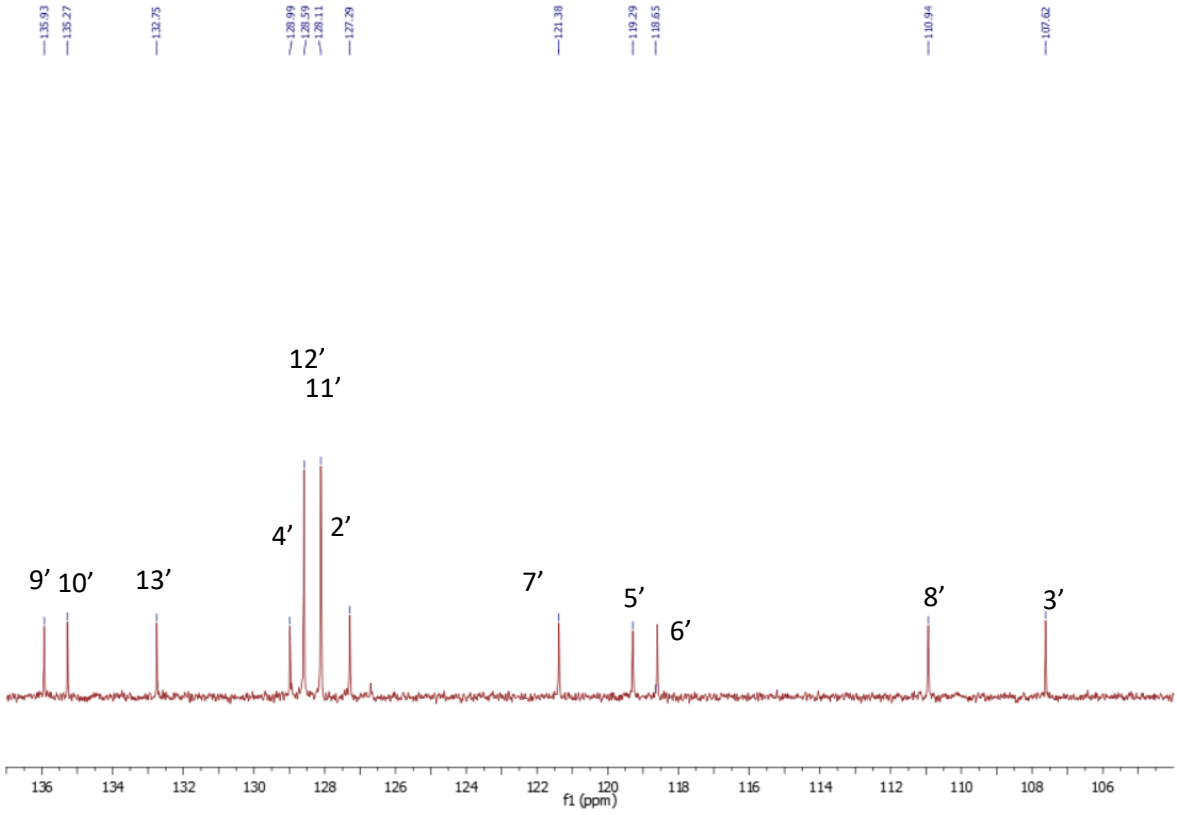


Figure 105: ^{13}C NMR spectrum of **10**, ^{13}C assignments between δ 107 - 137 ppm (DMSO, 176 MHz, 300K)

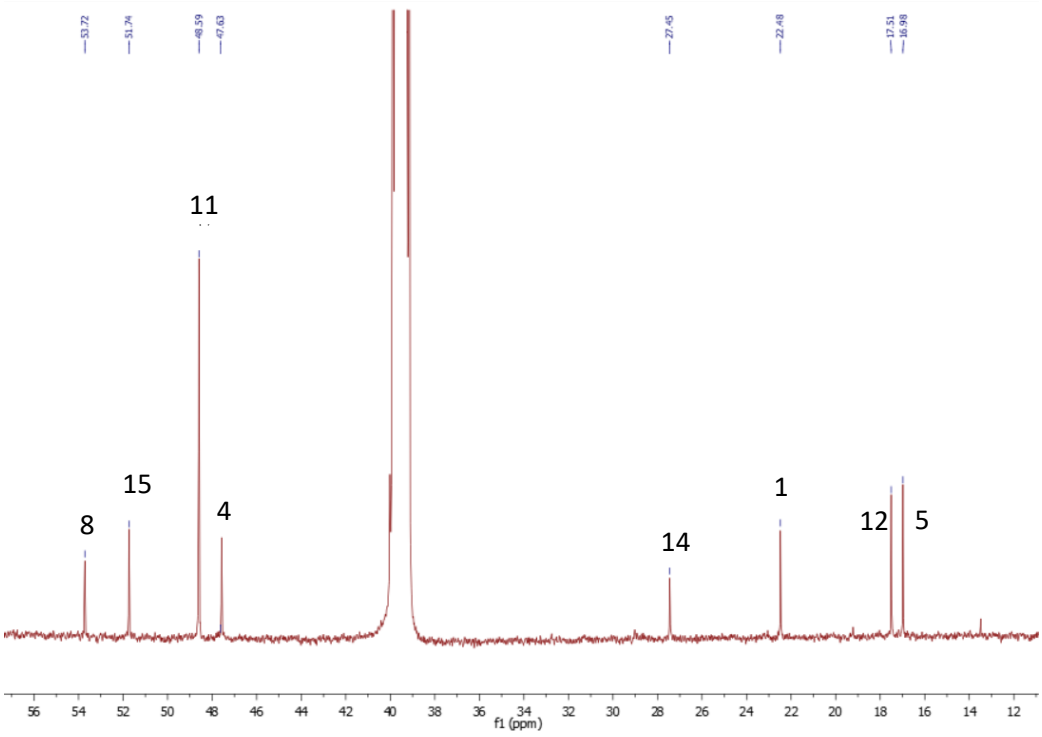


Figure 106: ^{13}C NMR spectrum of ^{13}C spectrum of **10**, ^{13}C assignments between δ 16 - 54 ppm (DMSO, 176 MHz, 300K)

AcNH-Ala-Trp-Ala-OH vs. AcNH-Ala-Trp(Ph)-Ala-OMe

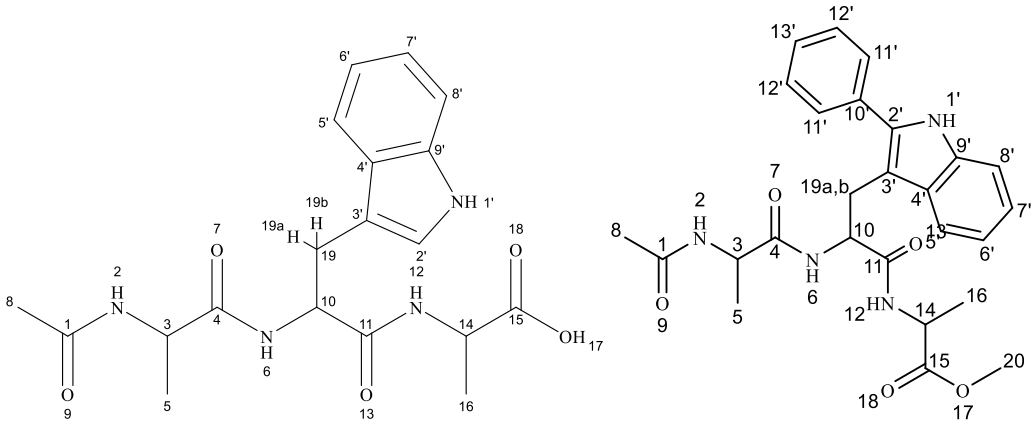


Figure 107: Ala-Trp-Ala (left) and Ala-Trp(Ph)-Ala (right)

Comparing the two molecules, the main differences would be in the aromatic region and low in the aliphatic region due to the methyl ester formation. When comparing the aromatic region, in Figure 23, the ROESY spectrum shows interactions between all adjacent protons including the 2' proton. Once arylated, the 2' proton peak is removed and included the extra aromatic protons 11', 12' 13' with an additional interaction between 1' and 11' which indicates that the aromatic ring is 90° from the indole ring, spectra shown in Figures 95 and 96. Using the TOCSY and COSY spectra in Figures 27 and 99, it can be seen that there is only one aromatic cycle in Ala-Trp-Ala whilst there are two in the arylated product. When comparing the chemical shifts for the aromatic protons, the indole proton in Ala-Trp-Ala is at 10.81 ppm whilst in the arylated product the indole proton shifts to 11.16 ppm meaning that the aromatic ring deshields the indole proton. The remaining aromatic protons did not change too much except for 5' which moved from 7.58 to 7.68.

Looking at the backbone of the peptide, for Ala-Trp-Ala the amide proton 2 had interactions with the protons at the acetyl protection group 8 and the adjacent proton on the α-carbon 3. The interactions of 2 changes to include the β-carbon protons as well as the other protons after arylation. The amide proton 12 has interactions with 10 and 16 but not with the adjacent proton on the α-carbon 13. The arylated product has interactions between 12, 13 and 16 but no longer with 10. The amide proton 6 showed interactions between positions 3 and 5 in Ala -Trp-Ala., whilst in the arylated product the proton at 6 only interacts with the adjacent proton 10. In both the arylated

and unarylated molecule, there are interactions between the α-carbon protons and the β-carbon protons in the peptide backbone which was expected.

The final area which had some difference was the low shift aliphatic region (<2 ppm). In the starting material, there were only two peaks of integral 3 showing the two methyl sidechains of the alanine residues. In the arylated product, there were three peaks with integrals of 3 which represents the two methyl sidechains of the alanine residues and the methyl ester formed in the product.

AcNH-Ala-Trp(Ph)-Ala-Tyr-Ala-OMe (12)

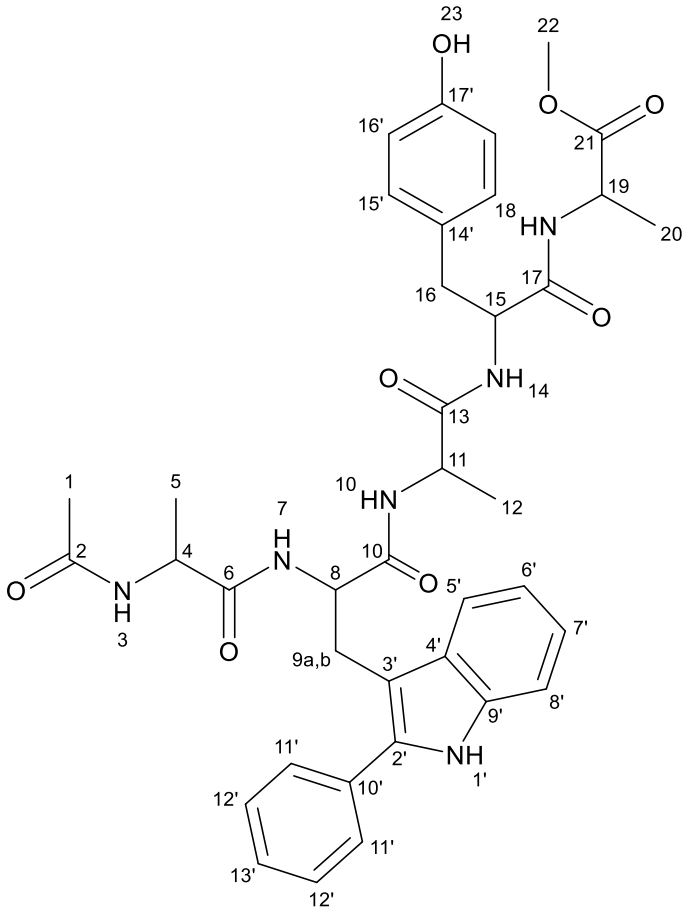


Figure 108: Reference molecule for Ala-Trp(Ph)-Ala-Tyr-Ala (**12**)

NMR spectrum data for **12**

^{1}H NMR (700 MHz, DMSO) δ 11.17 (s, 1H), 8.16 (d, *J* = 6.2 Hz, 2H), 8.00 (d, *J* = 23.1 Hz, 3H), 7.80 (d, *J* = 7.8 Hz, 1H), 7.68 – 7.63 (m, 5H), 7.46 (t, *J* = 7.5 Hz, 4H), 7.36 (t, *J* = 7.3 Hz, 2H), 7.32 (d, *J* = 8.0 Hz, 2H), 7.07 (t, *J* = 7.5 Hz, 2H), 6.99 (d, *J* = 8.0 Hz, 3H), 6.96 (t, *J* = 7.4 Hz, 2H), 6.62 (d, *J* = 7.9 Hz, 4H), 4.52 (dd, *J* = 13.9, 7.3 Hz, 2H), 4.33 (dd, *J* = 13.2, 8.6 Hz, 2H), 4.24 – 4.18 (m, 2H), 4.07 (dd, *J* = 14.5, 7.1 Hz, 4H), 3.59 (s, 5H), 2.90 (dd, *J* = 14.0, 4.3 Hz, 2H), 2.65 (dd, *J* = 13.9, 9.4 Hz, 2H), 2.58 (d, *J* = 15.5 Hz, 1H), 1.76 (s, 5H), 1.24 (d, *J* = 7.2 Hz, 8H), 1.10 – 1.06 (m, 7H), 1.01 (q, *J* = 7.5 Hz, 7H).

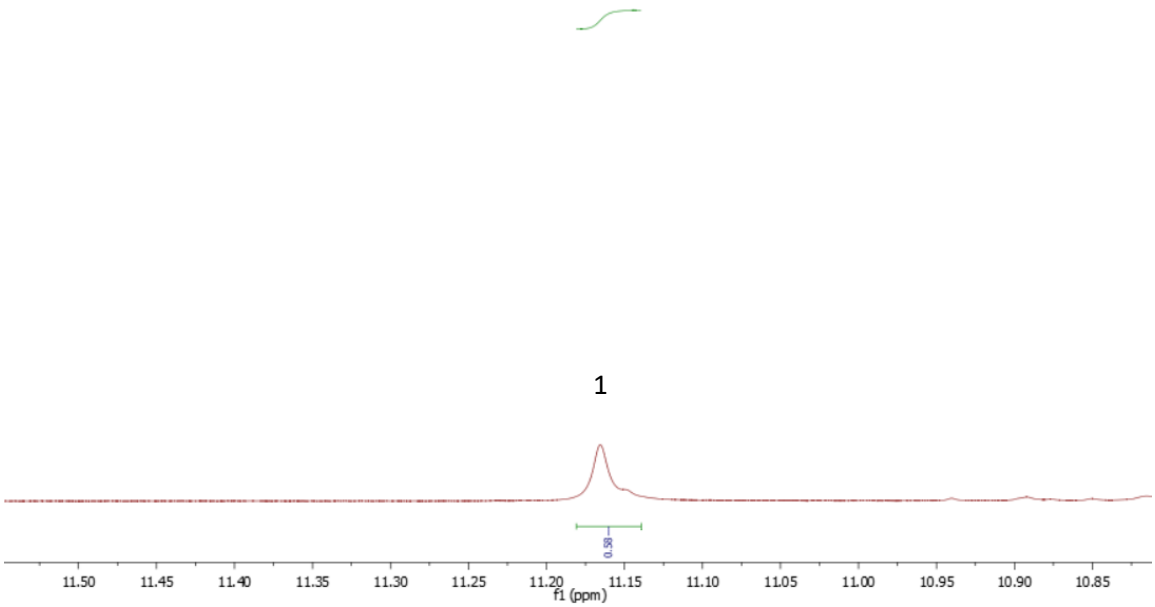


Figure 109: ^{1}H NMR spectrum of **12,** showing ^{1}H assignments between δ 11.1 – 11.2 ppm (DMSO, 700 MHz, 300 K)

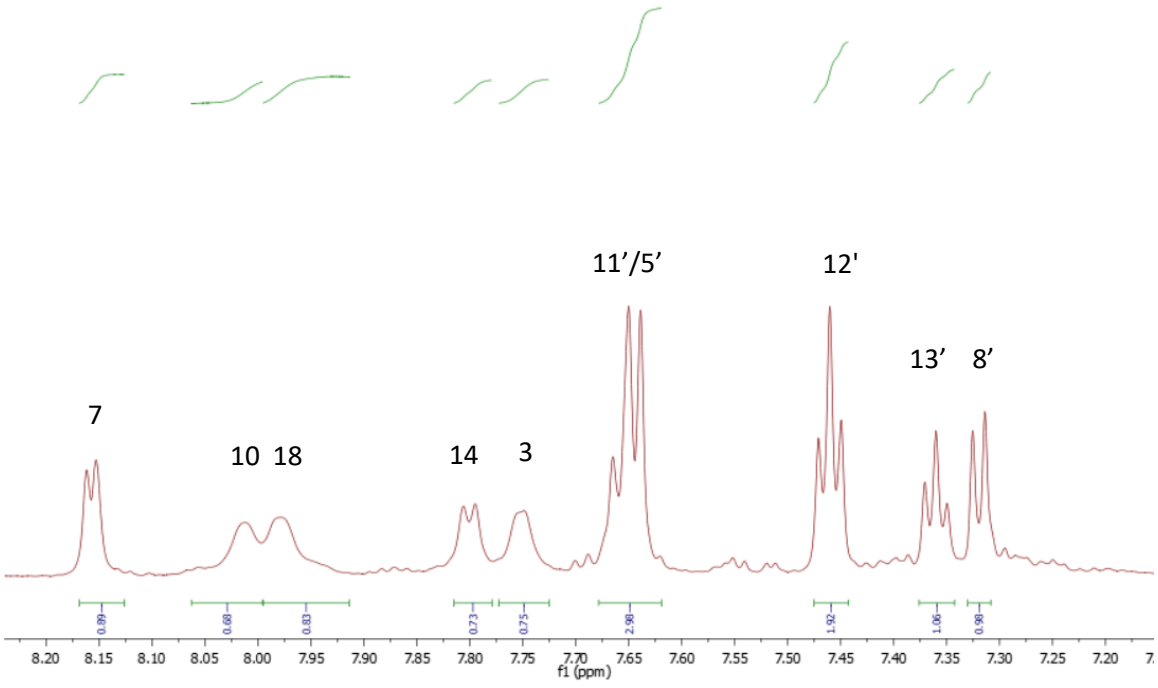


Figure 110^{1}H NMR spectrum of **12,** showing ^{1}H assignments between δ 7.3 – 8.2 ppm (DMSO, 700 MHz, 300 K)

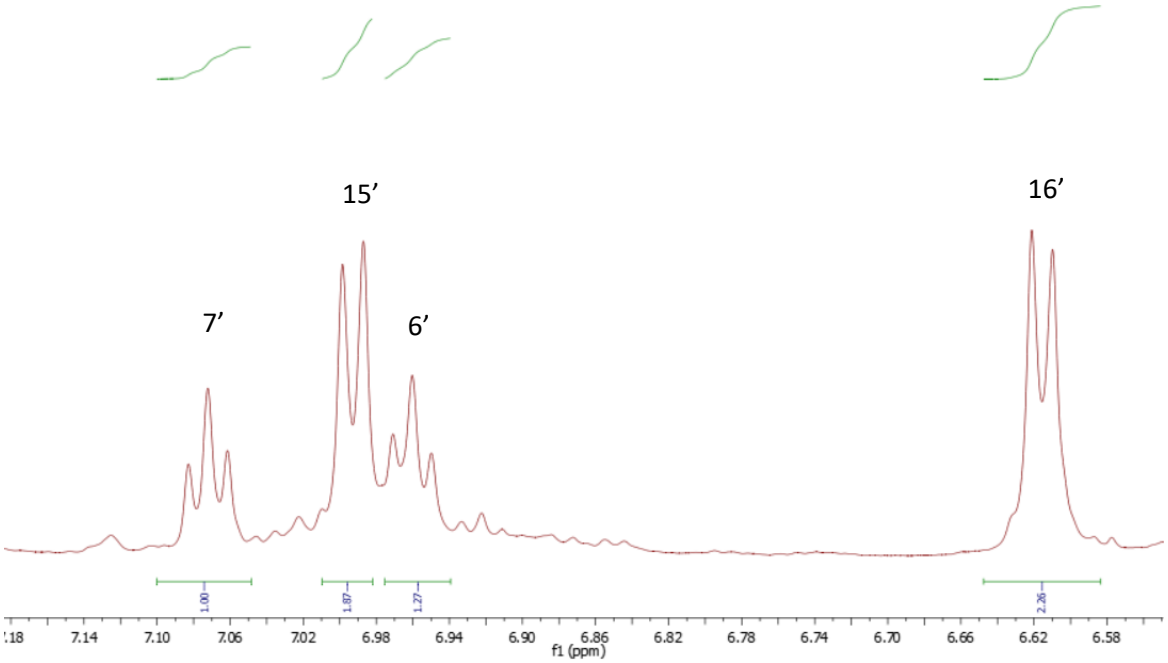


Figure 111: ^{1}H NMR spectrum of **12,** showing ^{1}H assignments between δ 6.58 – 7.1 ppm (DMSO, 700 MHz, 300 K)

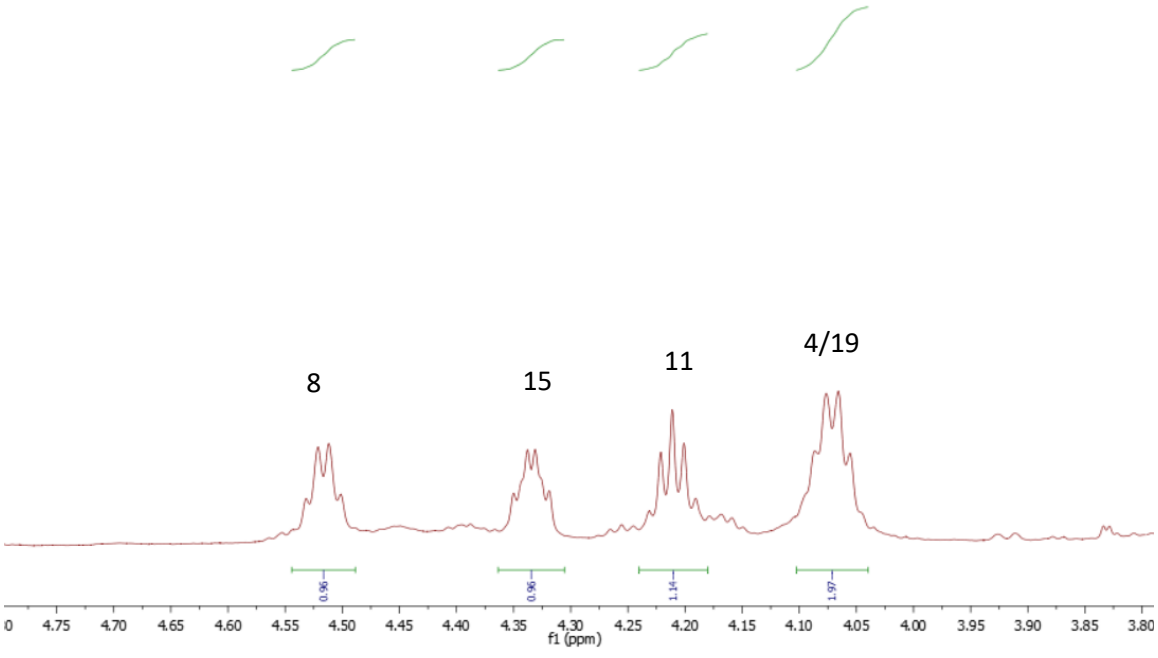


Figure 112: ^{1}H NMR spectrum of **12,** showing ^{1}H assignments between δ 4.0 – 4.6 ppm (DMSO, 700 MHz, 300 K)

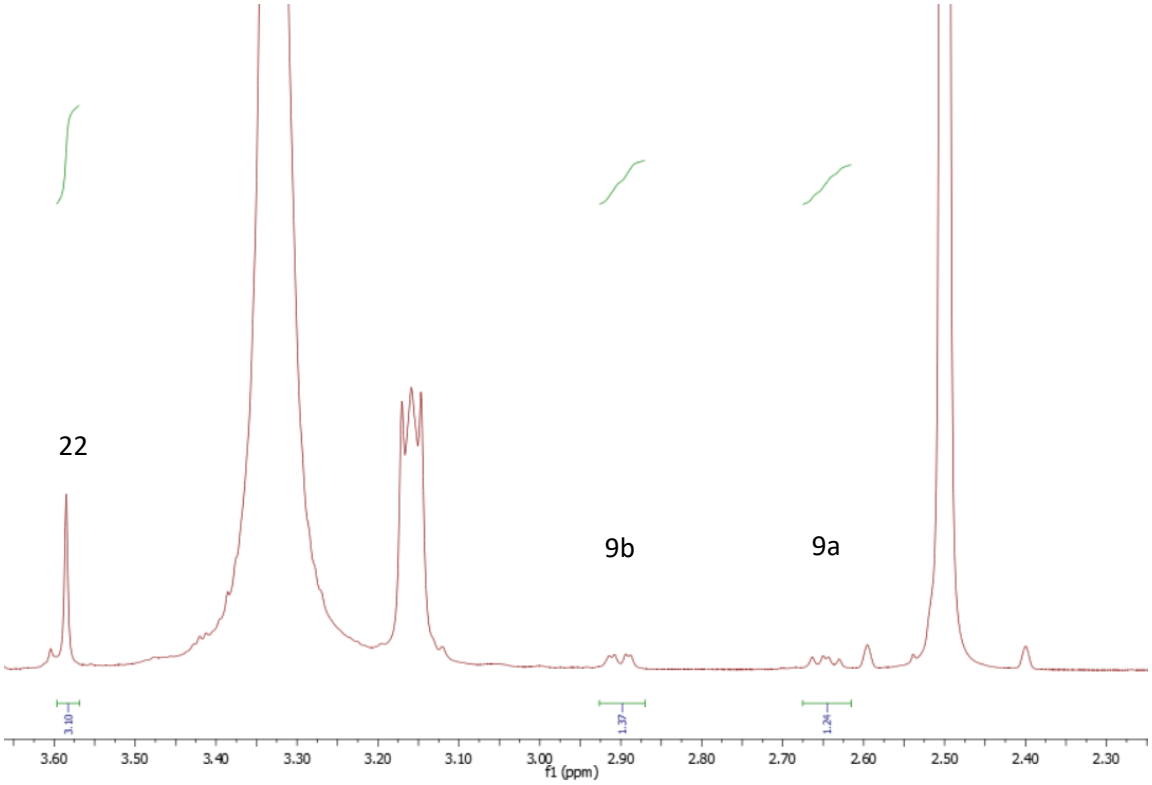


Figure 113: ^{1}H NMR spectrum of **12,** showing ^{1}H assignments between δ 2.6 – 3.6 ppm (DMSO, 700 MHz, 300 K)

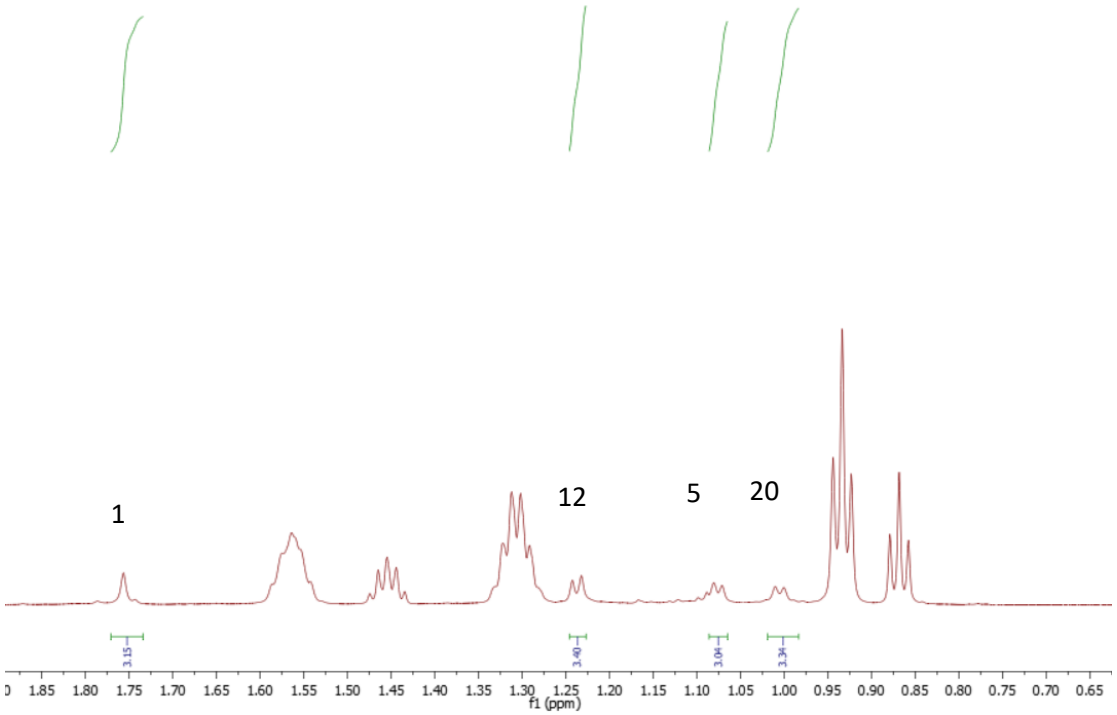


Figure 114: ^{1}H NMR spectrum of **12,** showing ^{1}H assignments between δ 1.0 – 1.75 ppm (DMSO, 700 MHz, 300 K)

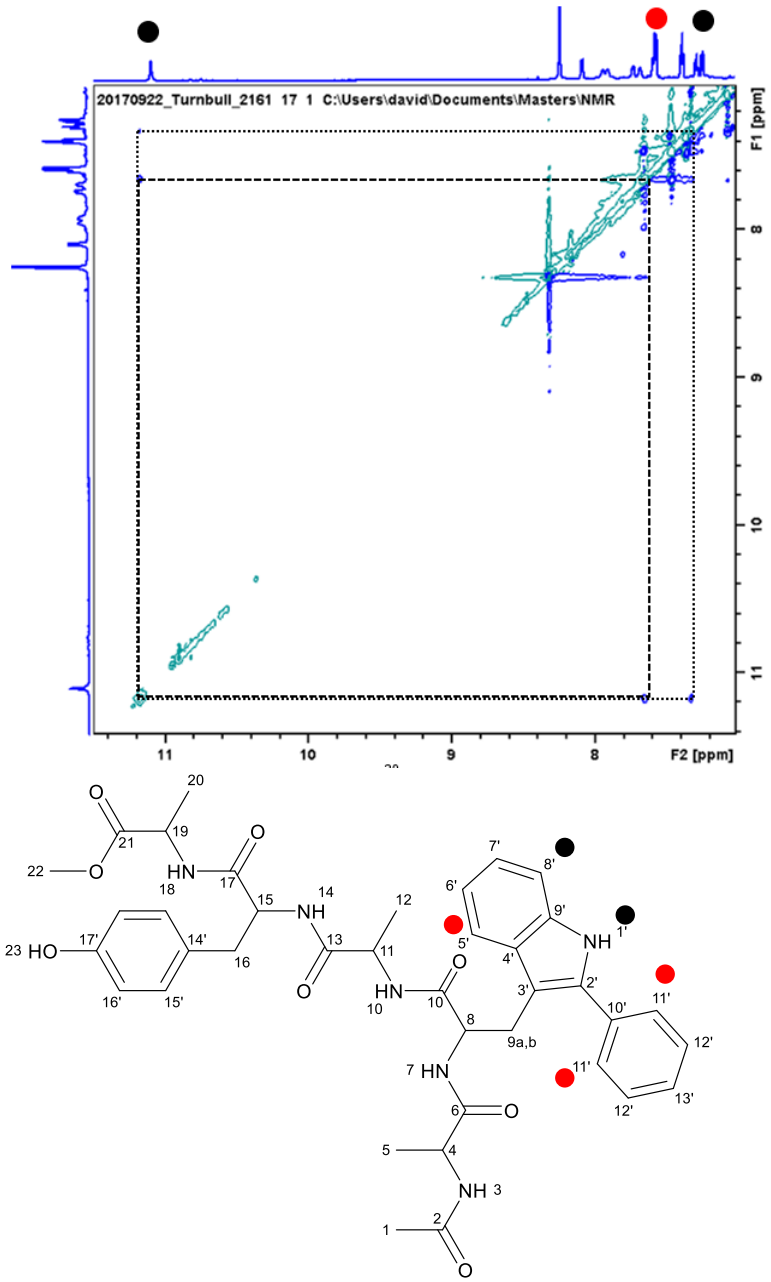


Figure 115: ROESY Spectrum of aromatic region of **12** – examination of the indole region (DMSO, 700 MHz, 300 K)

From the spectra above, interactions between the indole nitrogen proton at position 1' and the protons on the new aryl group at position 11' or the indole proton at position 5' and interactions with position 8'. This interaction wasn't seen in the starting material, this interaction could be due to repulsion from the new aryl group.

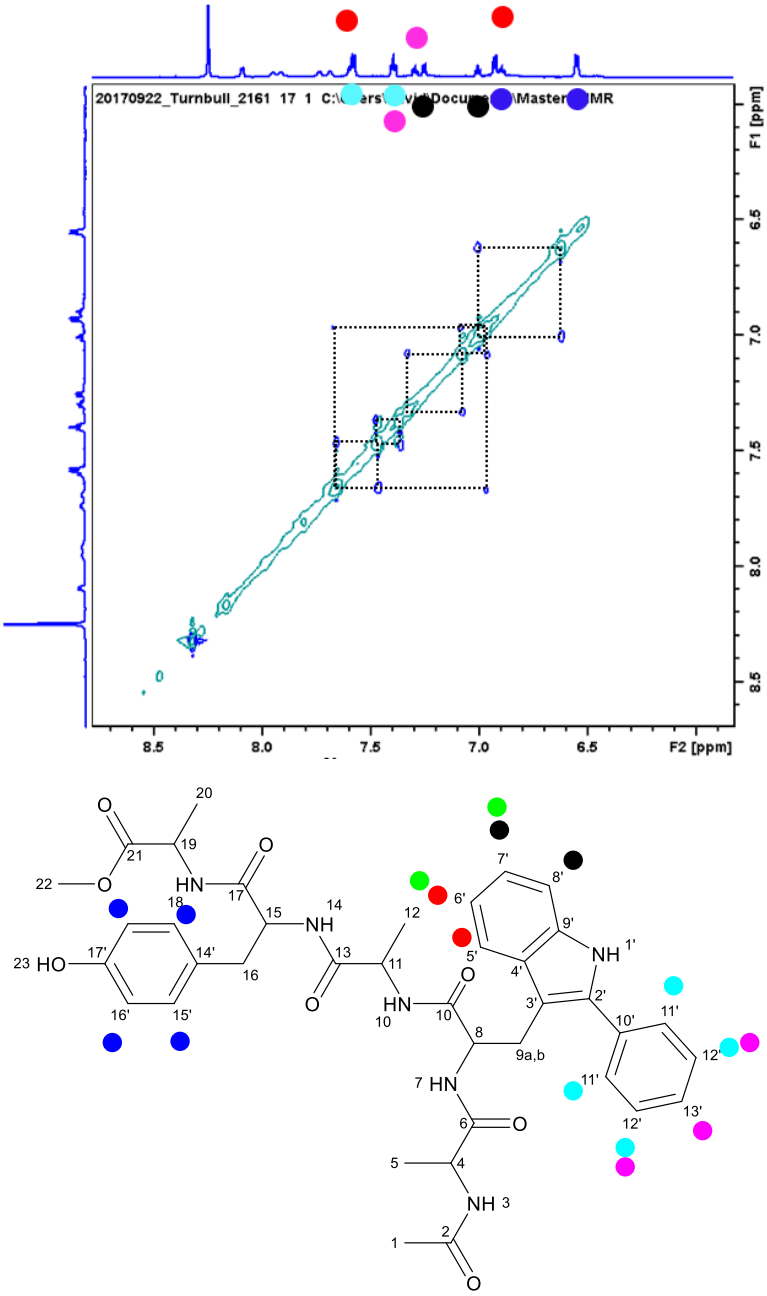


Figure 116: ROESY Spectrum of aromatic region of **12** – examination of the aromatic region (DMSO, 700 MHz, 300 K)

The new aryl protons can be seen interacting, 11' to 12' and 12' to 13' which is expected. The other aromatic protons all interact with each other in the expected manner.

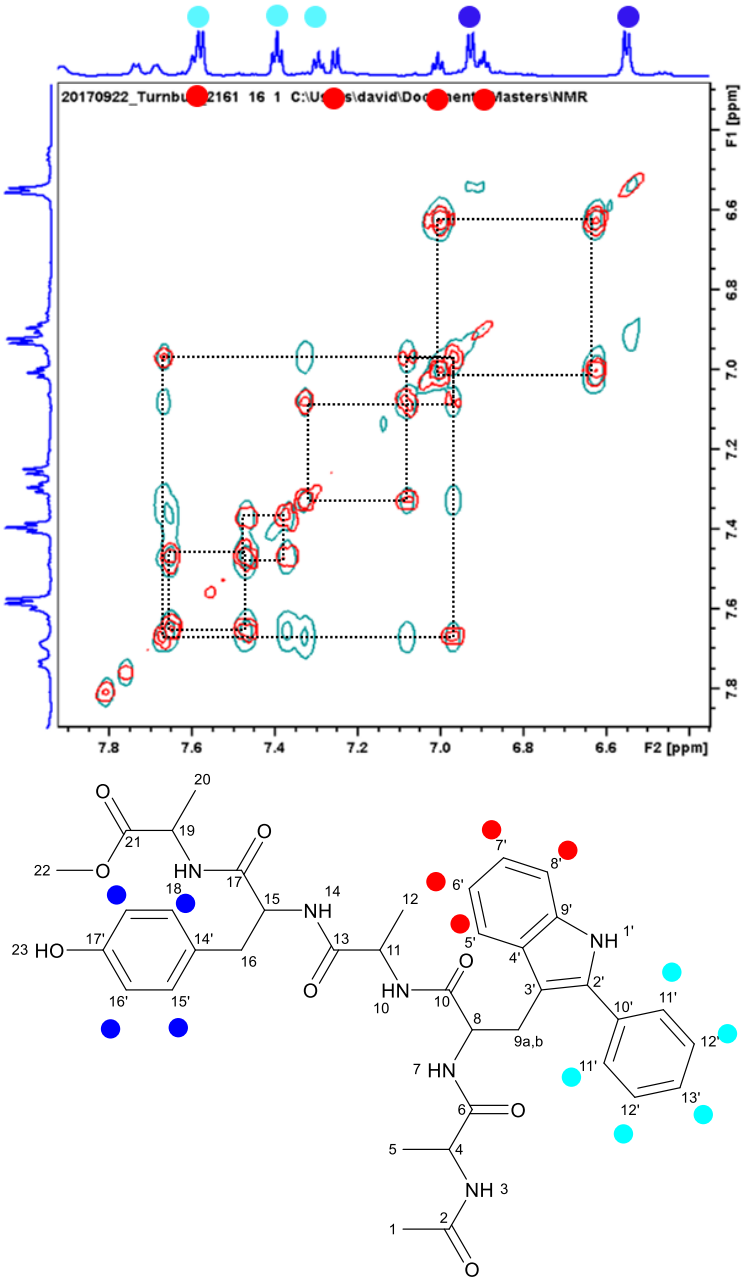


Figure 117: TOCSY (Blue) and COSY (Red) spectrum of **12** – examination of the aromatic area (DMSO, 700 MHz, 300 K)

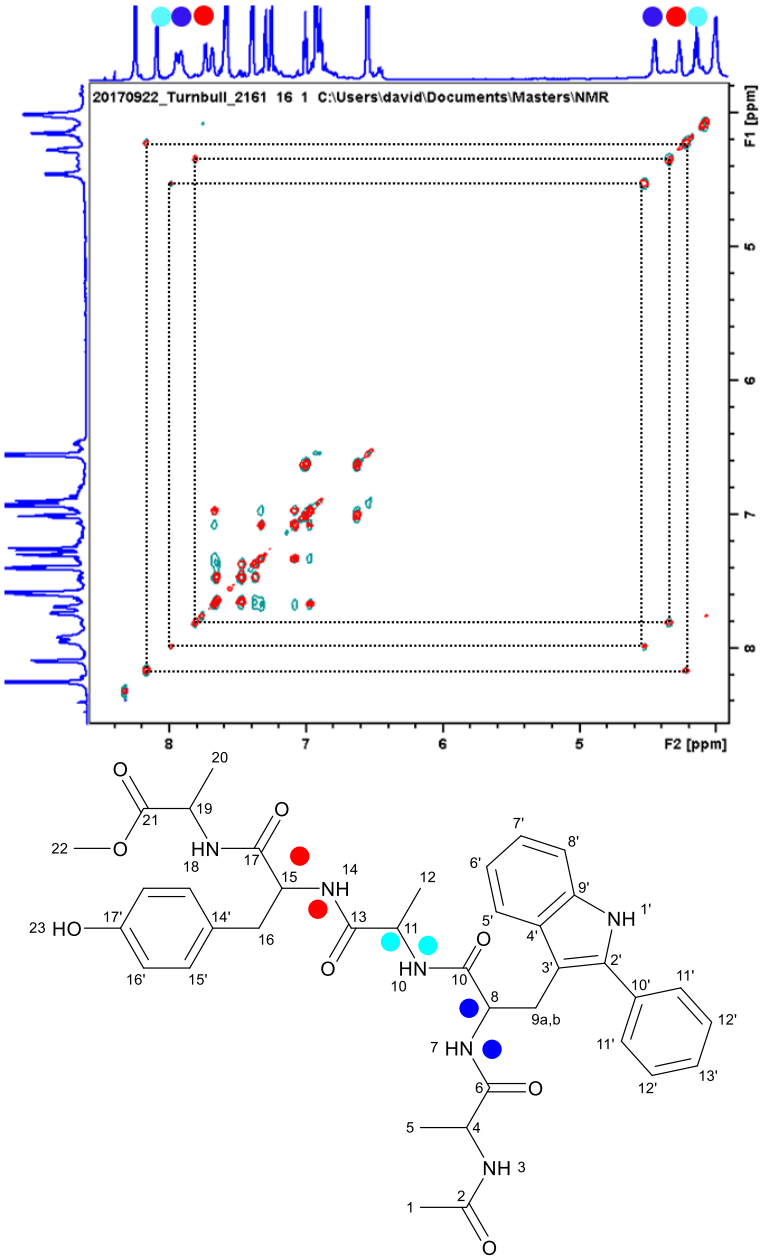


Figure 118: TOCSY (Blue) and COSY (Red) spectrum of **12** – examination of the peptide backbone (DMSO, 700 MHz, 300 K)

AcNH-Ser-Gly-Trp(Ph)-Ala-OMe (11)

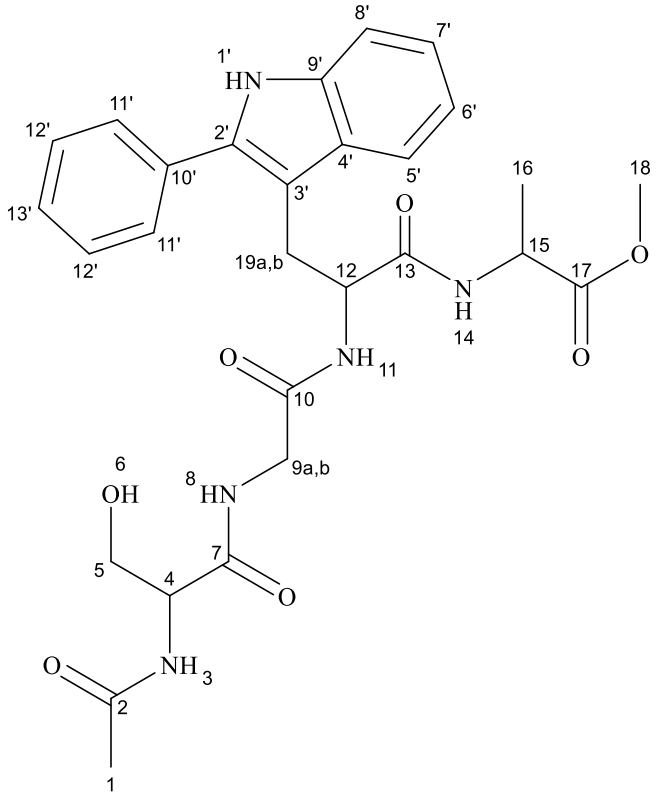


Figure 119: Reference Molecule for Ser-Gly-Trp(Ph)-Ala *(**11**)*

NMR spectrum data for **11**

^{1}H NMR (700 MHz, DMSO) δ 11.16 (s, 1H), 8.20 (d, *J* = 6.9 Hz, 1H), 8.16 (t, *J* = 4.9 Hz, 1H), 8.14 – 8.09 (m, 2H), 7.69 (d, *J* = 7.6 Hz, 3H), 7.49 (t, *J* = 7.5 Hz, 2H), 7.37 (t, *J* = 7.2 Hz, 1H), 7.32 (d, *J* = 8.0 Hz, 1H), 7.08 (t, *J* = 7.4 Hz, 1H), 6.99 (t, *J* = 7.4 Hz, 1H), 5.16 (s, 1H), 4.74 – 4.65 (m, 1H), 4.25 – 4.17 (m, 2H), 3.68 (dt, *J* = 15.0, 7.5 Hz, 1H), 3.55 (dd, *J* = 16.6, 10.3 Hz, 3H), 3.49 (d, *J* = 5.0 Hz, 1H), 3.47 (s, 3H), 3.02 (dd, *J* = 14.4, 7.7 Hz, 1H), 1.86 (s, 3H), 1.21 (d, *J* = 7.2 Hz, 3H).

^{13}C NMR (176 MHz, DMSO) δ 172.98, 171.30, 170.19, 168.76, 136.37, 135.70, 133.21, 129.47, 129.07, 128.57, 127.73, 121.71, 119.70, 119.13, 111.35, 108.05, 62.07, 55.97, 54.25, 52.25, 48.06, 42.61, 28.56, 23.03, 17.42.

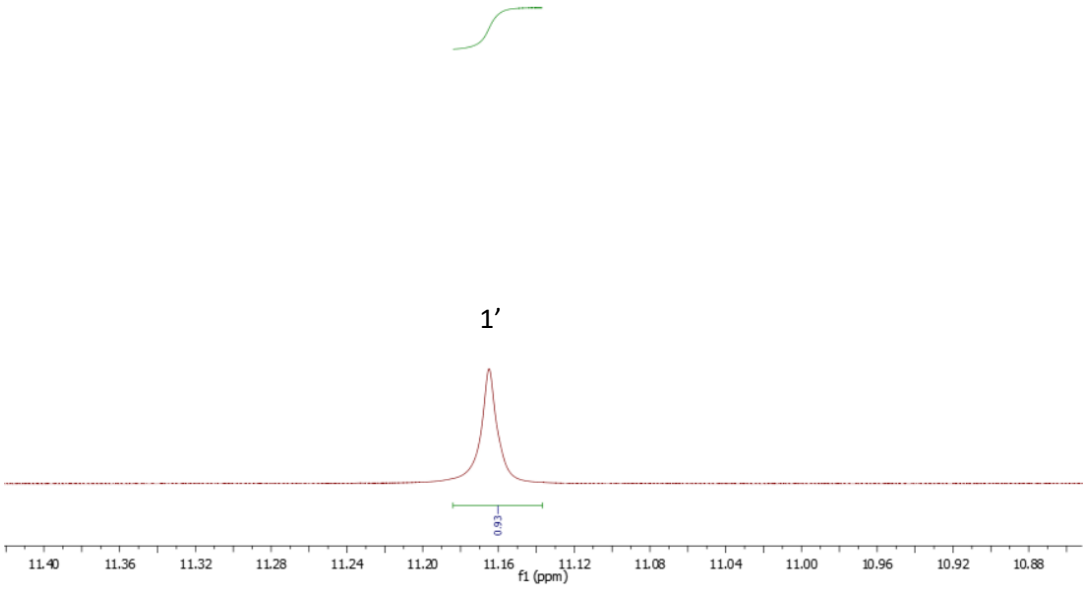


Figure 120: ^{1}H NMR spectrum of **11,** showing ^{1}H assignments between δ 11.12 – 11.2 ppm (DMSO, 700 MHz, 300 K)

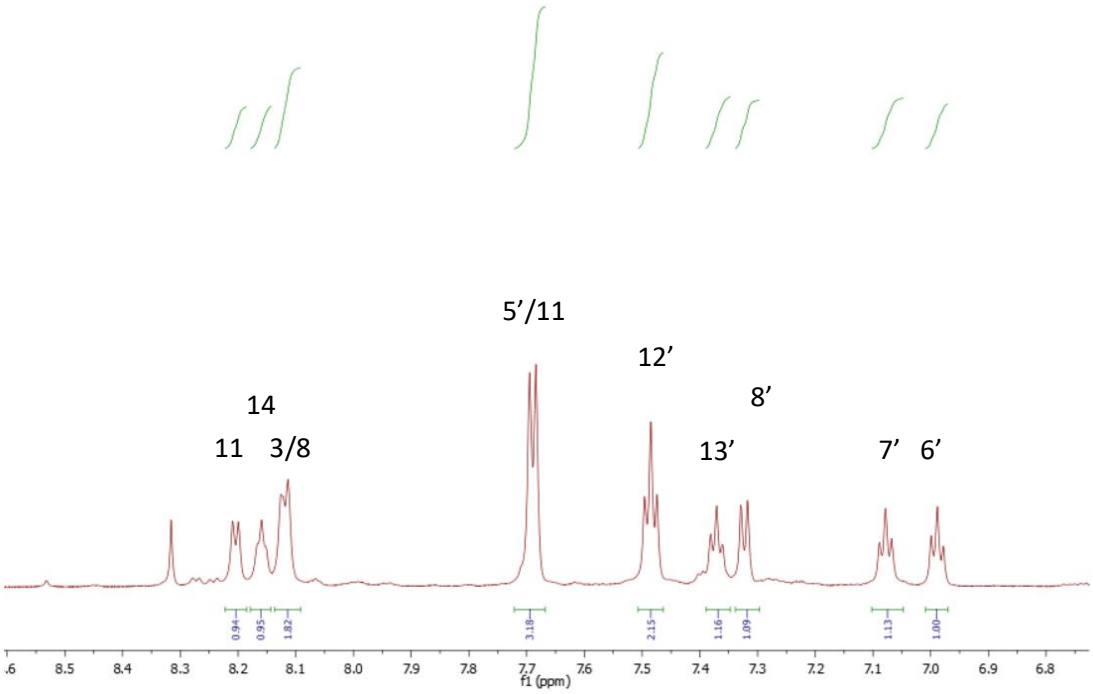


Figure 121: ^{1}H NMR spectrum of **11,** showing ^{1}H assignments between δ 6.9 – 8.2 ppm (DMSO, 700 MHz, 300 K)*)*

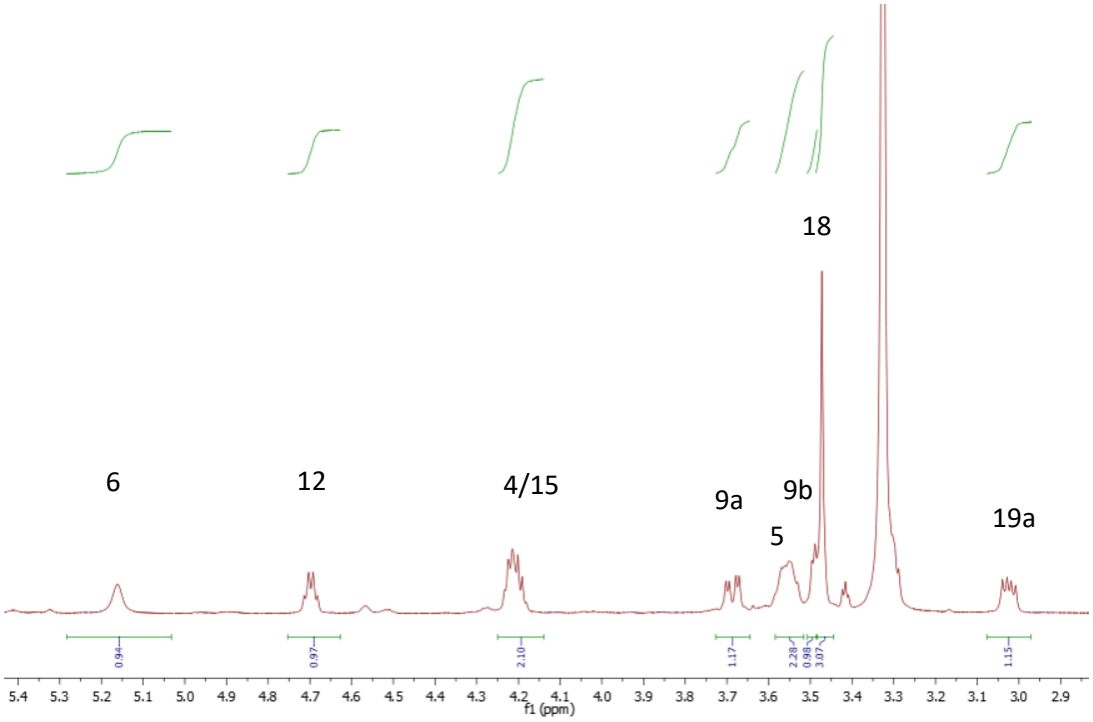


Figure 122: ^{1}H NMR spectrum of **11,** showing ^{1}H assignments between δ 2.9 – 5.2 ppm (DMSO, 700 MHz, 300 K)

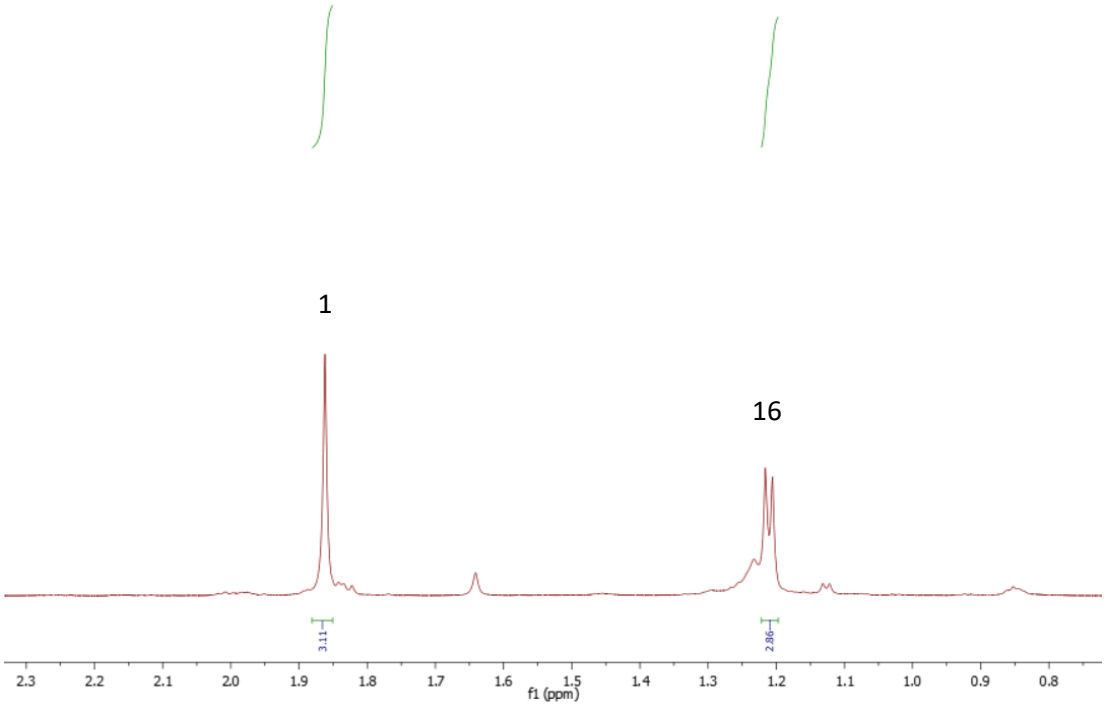


Figure 123: ^{1}H NMR spectrum of **11,** showing ^{1}H assignments between δ 1.1 – 1.9 ppm (DMSO, 700 MHz, 300 K)

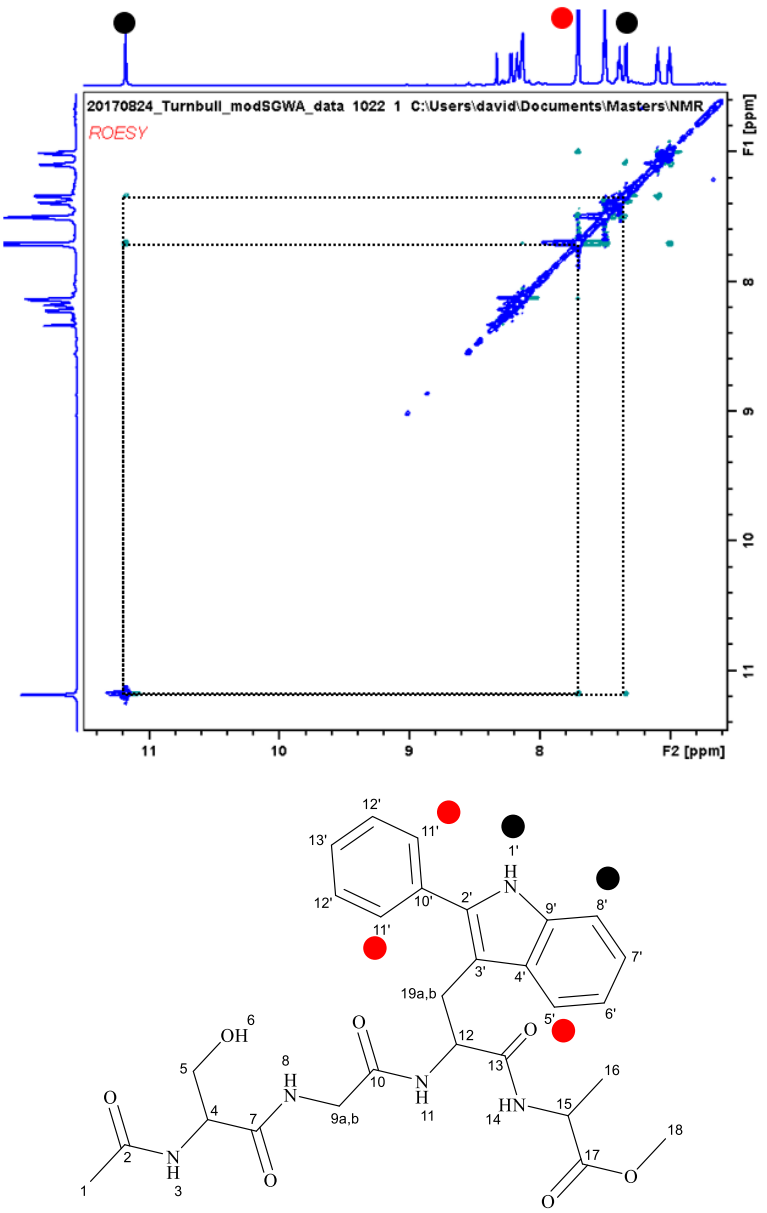


Figure 124: ROESY Spectrum of aromatic region of **11** – examination of the indole region with respect to the nitrogen (DMSO, 700 MHz, 300 K)

The spectrum above shows the interactions between the indole nitrogen proton at position 1' and position 8' on the indole ring and position 11' on the new aryl ring or the 5' position on the indole ring.

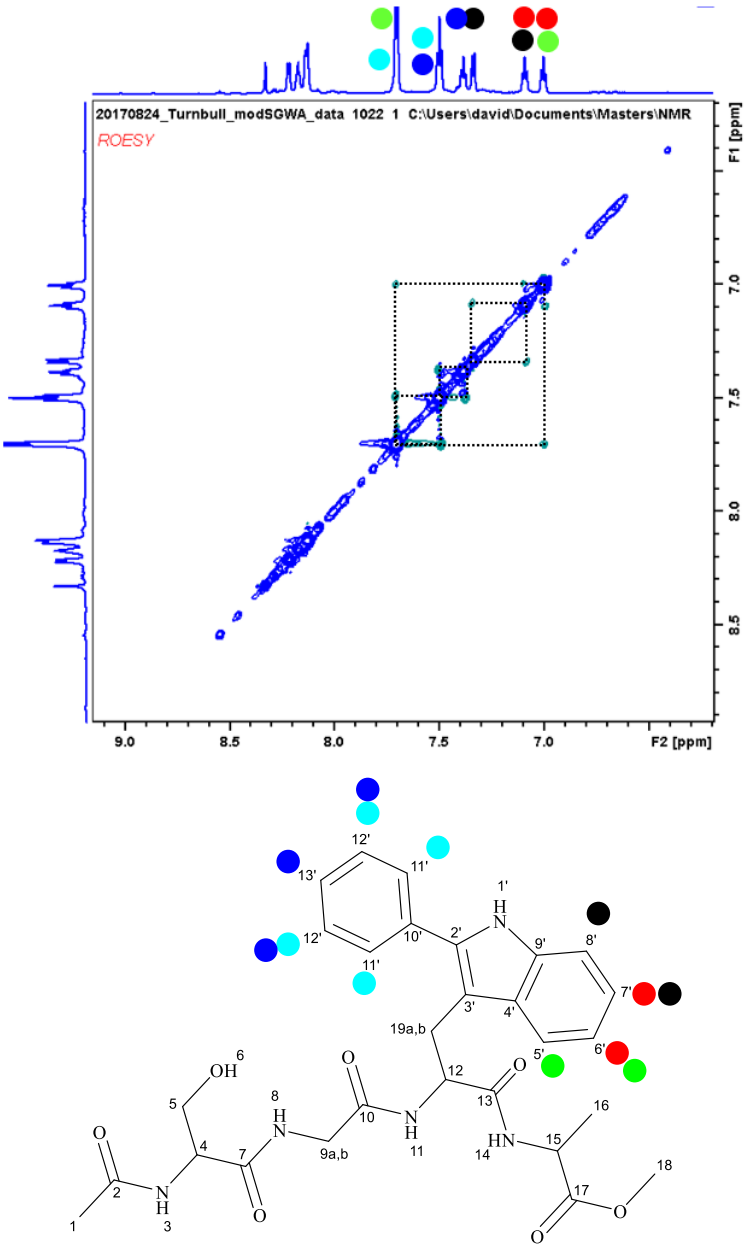


Figure 125: ROESY Spectrum of aromatic region of **11** – examination of the indole region (DMSO, 700 MHz, 300 K)

The spectra above indicates that the aromatic rings don't interact with each other, except for the indole proton and position 11' on the aryl ring. All the other interactions are as expected in the aromatic area.

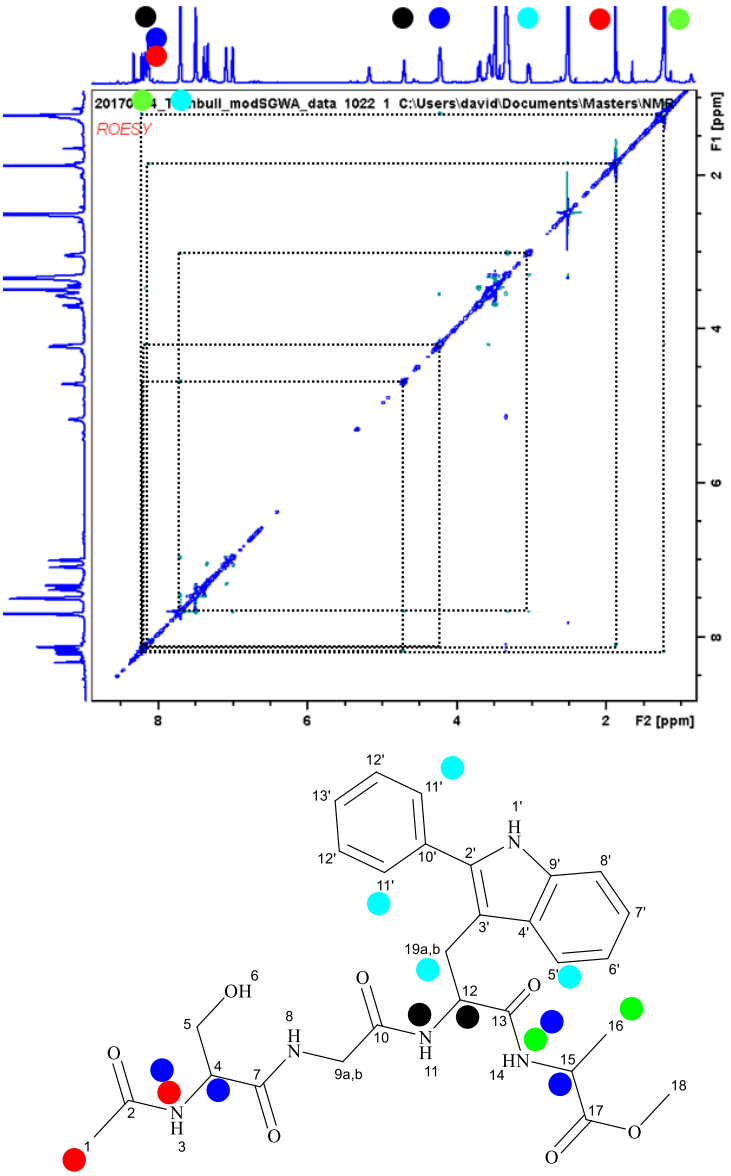


Figure 126: ROESY Spectrum of **11** – examination of the peptide chain (DMSO, 700 MHz, 300 K)

The interactions shown in the spectra above shows that the aromatic proton at either position 11’ or 5’ interact with the diastereomeric protons on the β-carbon of the tryptophan residues at position 19. There is an interaction between the amide protons at position 3 and 14 and position 1 and 16. The amide proton at position 11 interacts with the α-carbon proton of the tryptophan residue at position 12. The amide protons at positions 3 and 14 also interact with the protons at position 4 and 15 which is to be expected from adjacent atoms.

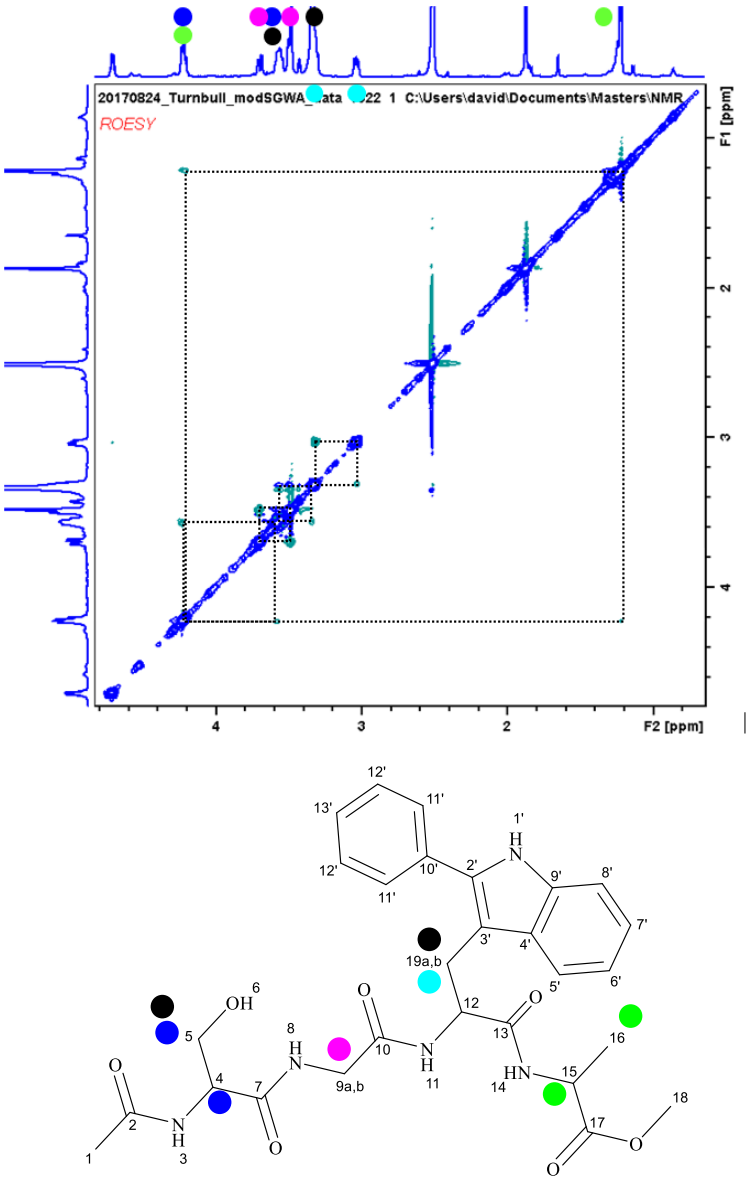


Figure 127: ROESY Spectrum of **11** – examination of the peptide backbone (DMSO, 700 MHz, 300 K)

There is an interaction between the proton 5 and proton 4 which is to be expected. There is also an interaction between positions 15 and 16 which, again, to be expected. There is an unusual interaction between position 5 and one of the diastereomeric protons at position 19. This might indicate that the molecule is bending and possibly folding in on itself.

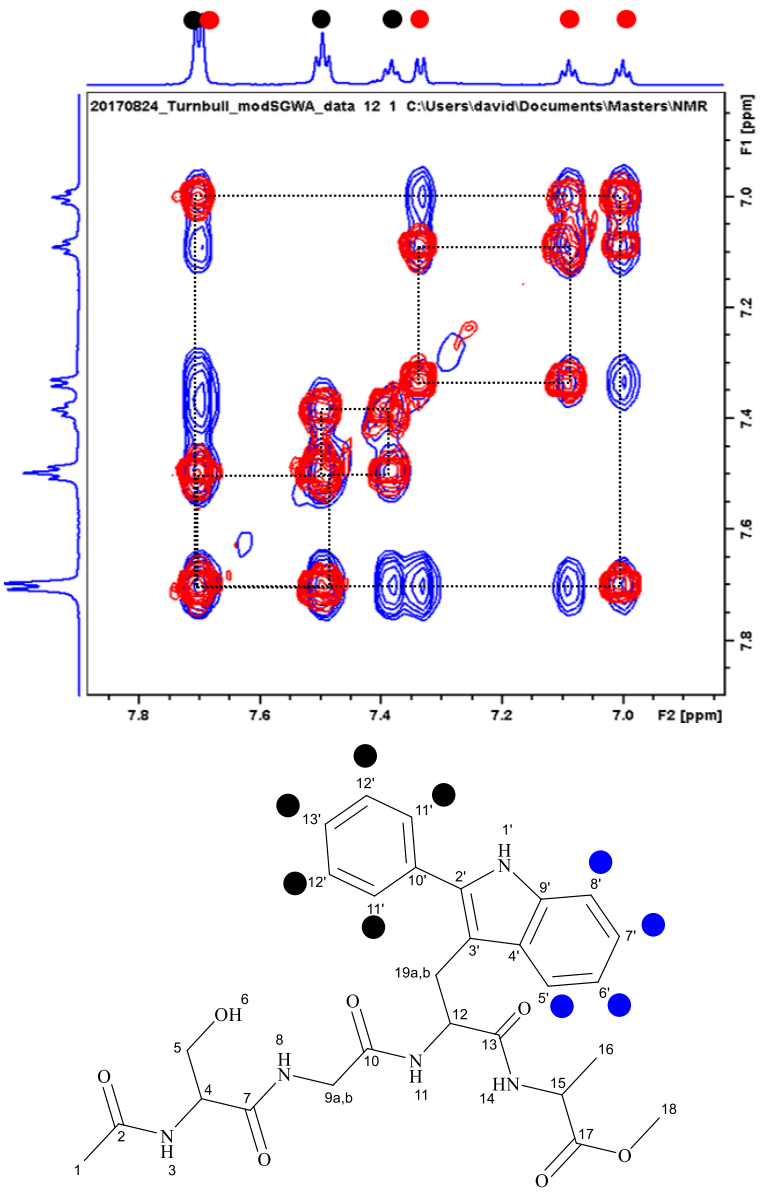


Figure 128: TOCSY (Blue) and COSY (Red) spectrum of **11** – examination of the aromatic rings(DMSO, 700 MHz, 300 K)

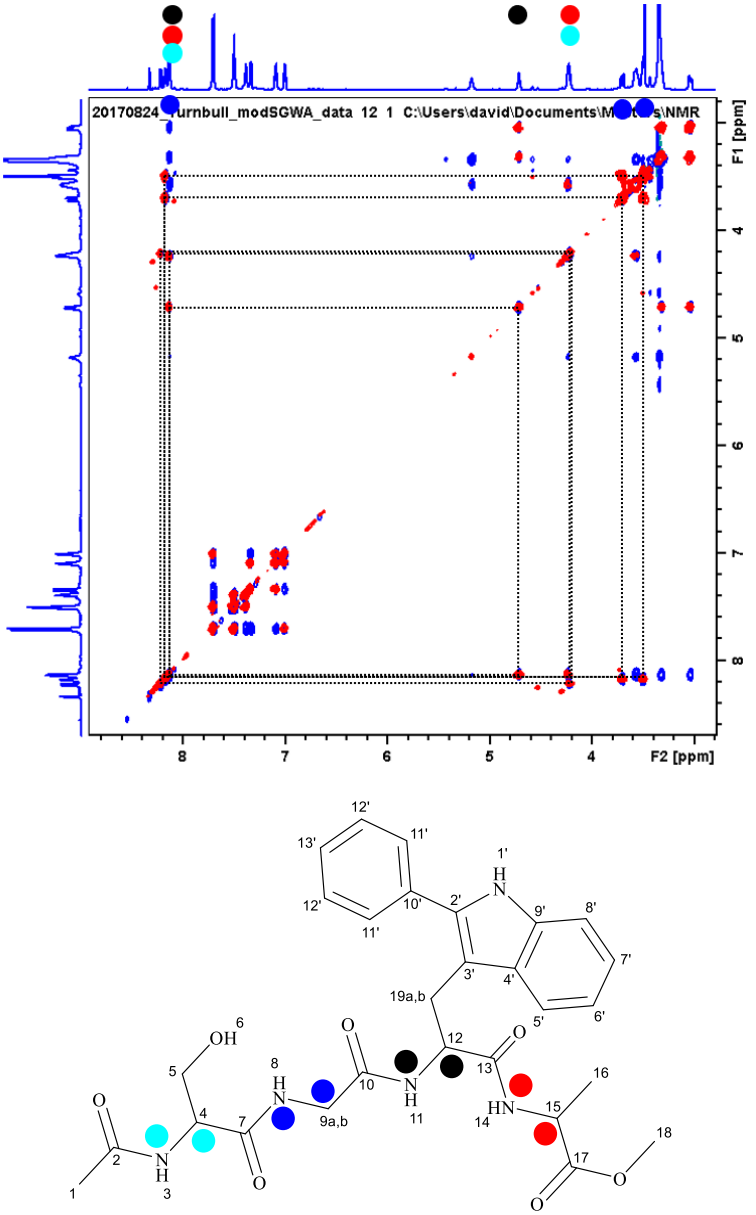


Figure 129: TOCSY (Blue) and COSY (Red) spectrum of **11** – examination of the peptide chain (DMSO, 700 MHz, 300 K)

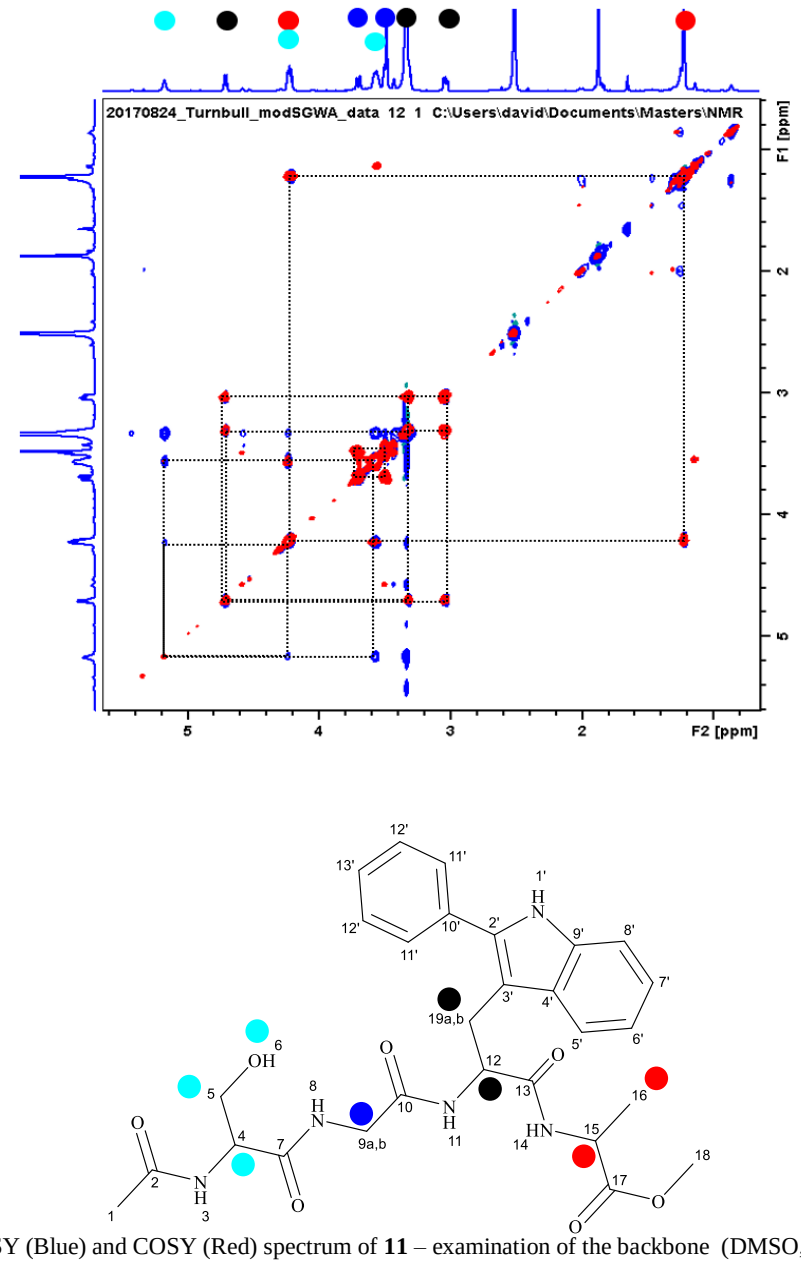


Figure 130: TOCSY (Blue) and COSY (Red) spectrum of **11** – examination of the backbone (DMSO, 700 MHz, 300 K)

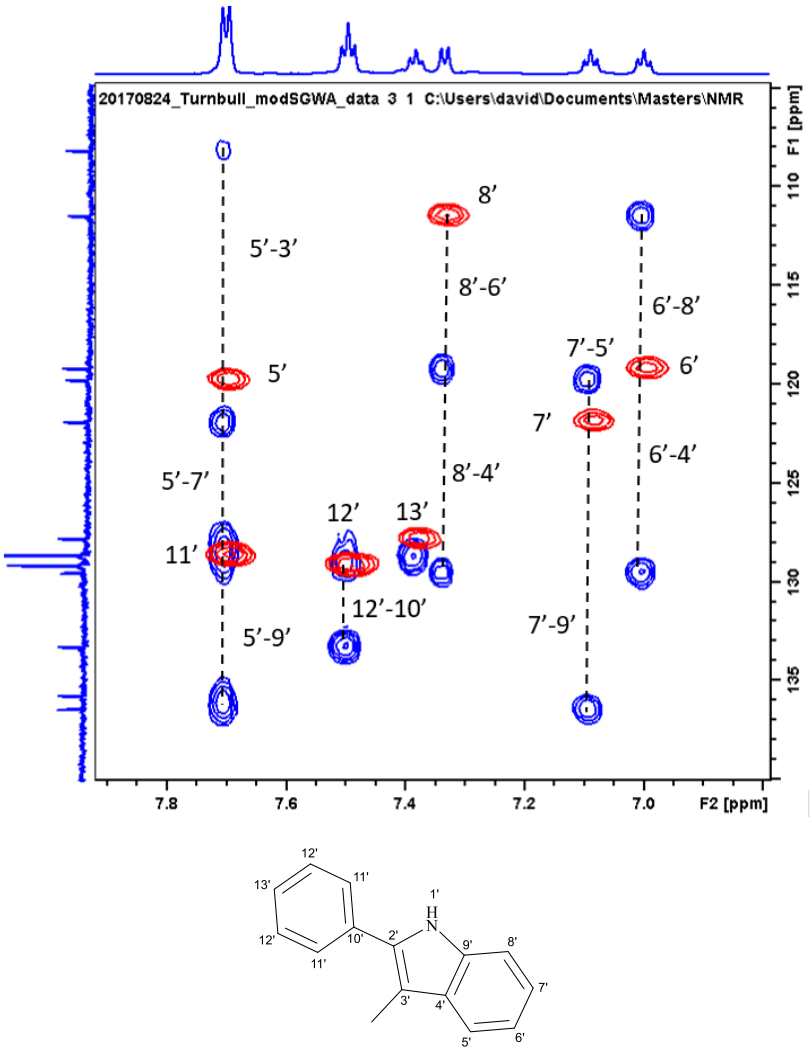


Figure 131: HSQC and HMBC spectrum of the aromatic region of **11** - examination of the indole group (DMSO, for 1H - 700 MHz, for ^{13}C - 176 MHz, 300 K)

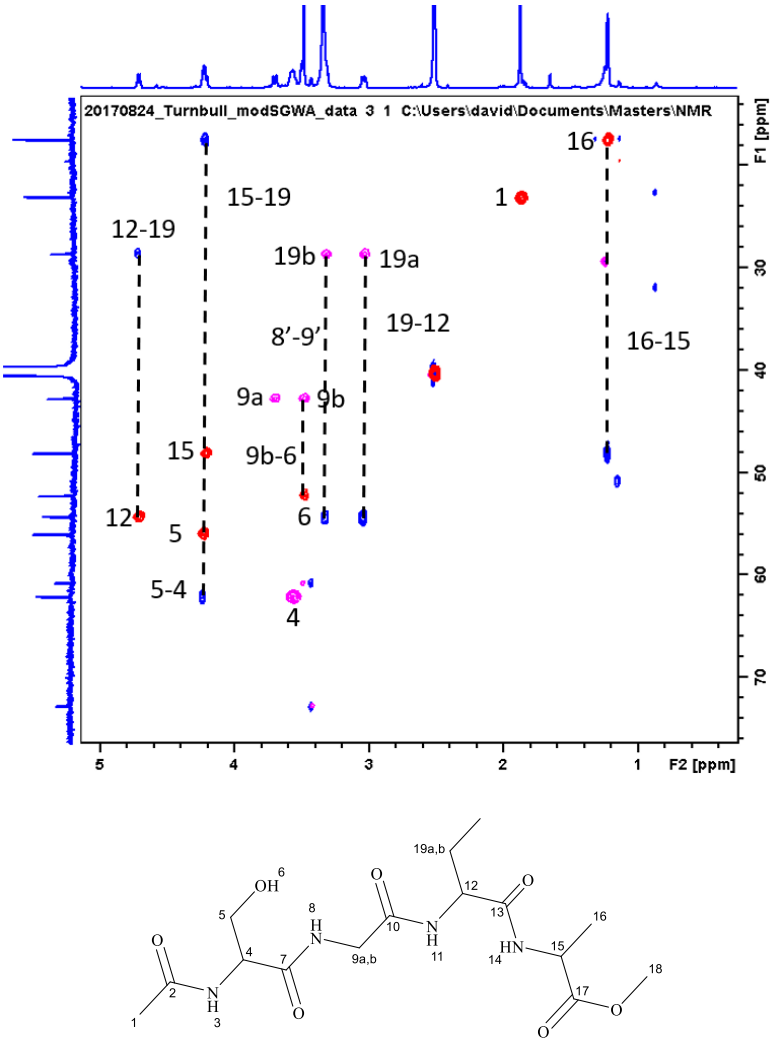


Figure 132: HSQC and HMBC spectrum of the aliphatic region of **11 -** examination of the peptide chain (DMSO, for 1H - 700 MHz, for ^{13}C - 176 MHz, 300 K)

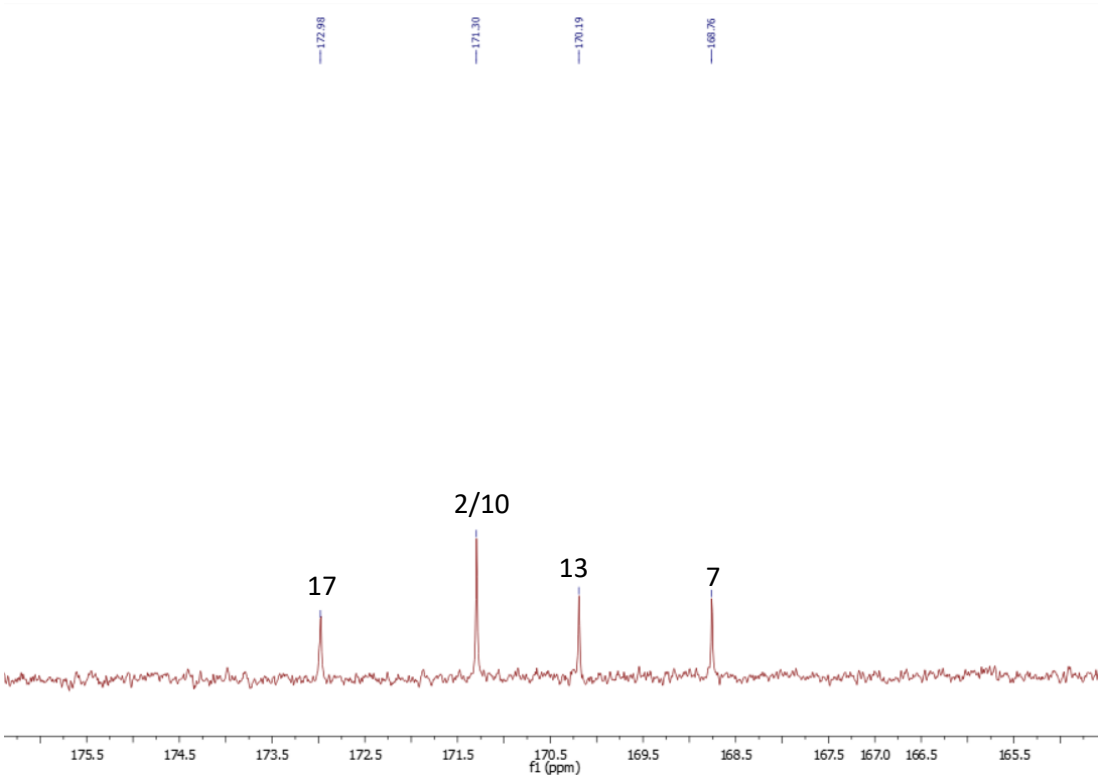


Figure 133: ^{13}C spectrum of **11**, ^{13}C assignments between δ 168.5 – 173.5 ppm (DMSO, 176 MHz, 300K)

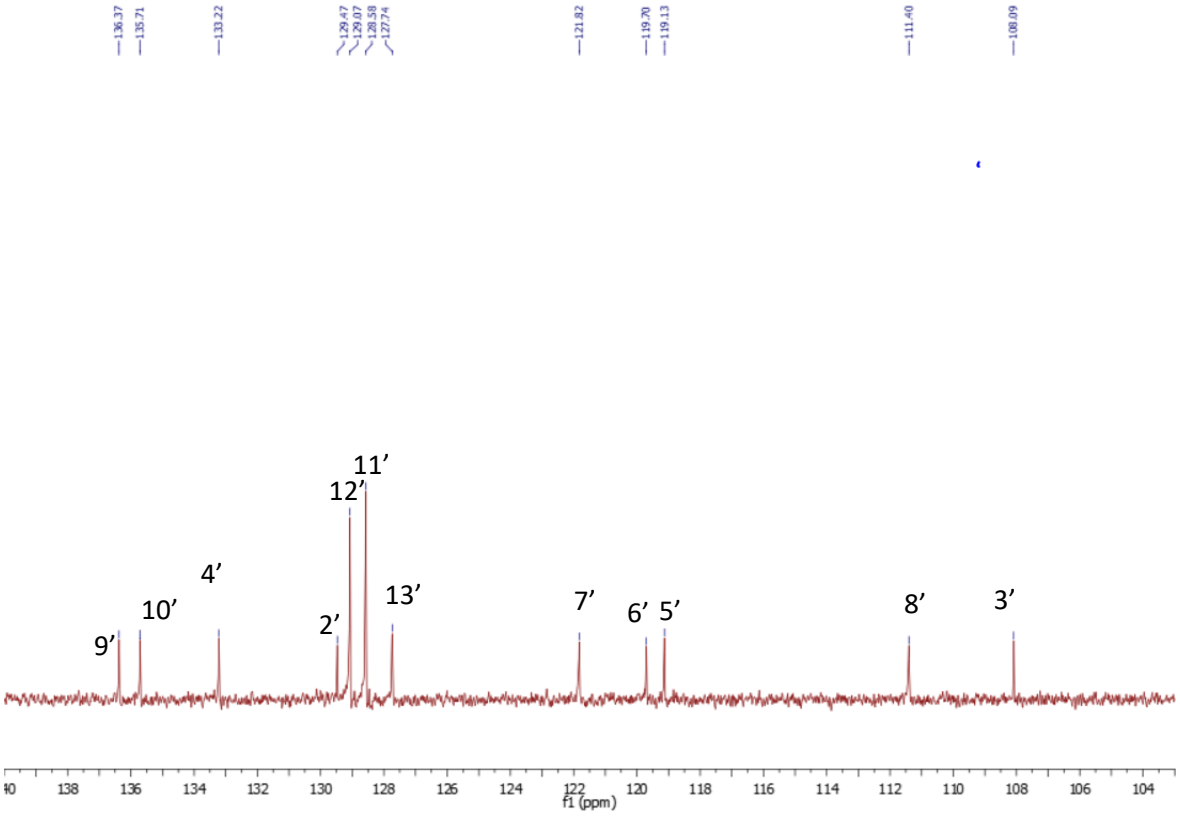


Figure 134: ^{13}C NMR spectrum of **11**, ^{13}C assignments between δ 108 – 136.5 ppm (DMSO, 176 MHz, 300K)

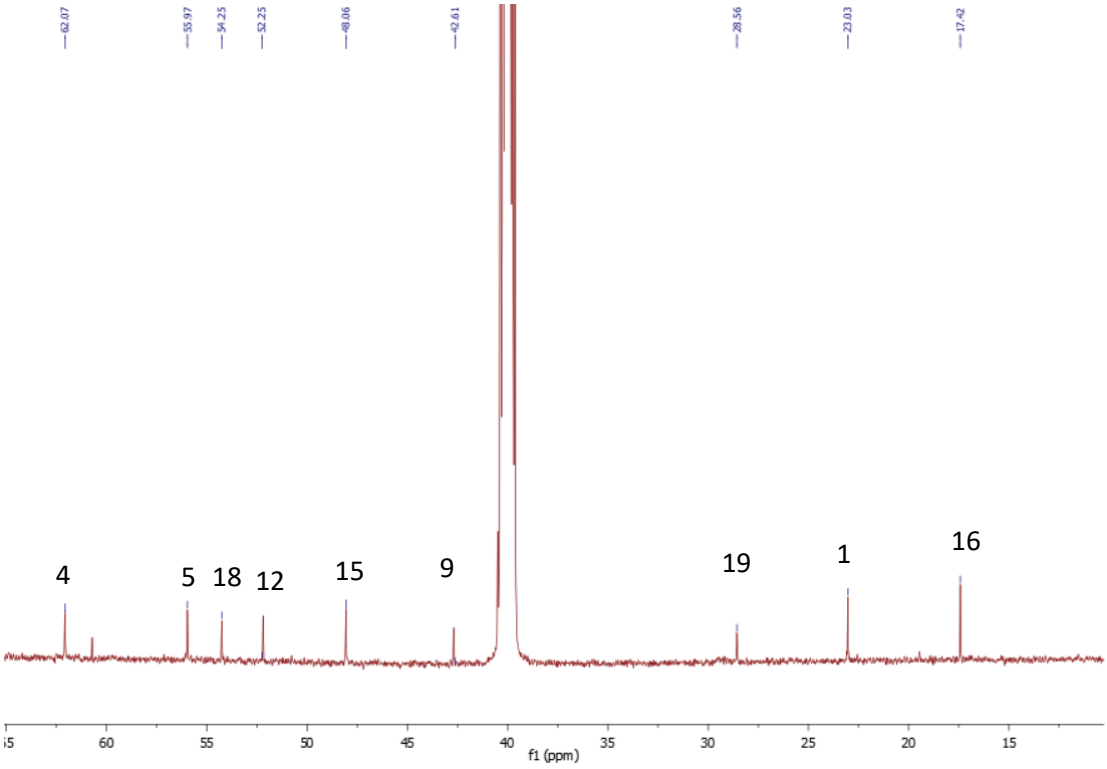


Figure 135: ^{13}C NMR spectrum of **11**, ^{13}C assignments between δ 15.0 – 65 ppm (DMSO, 176 MHz, 300K)

AcNH-Val-Trp(Ph)-Asn-Asn-Lys-Thr-Ala-OMe (13)

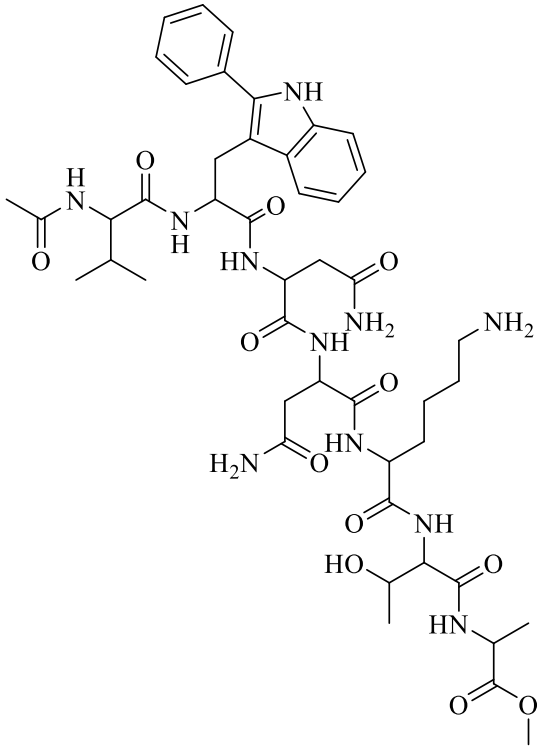


Figure 136: Reference molecule for Val-Trp(Ph)-Asn-Asn-Lys-Thr-Ala (13)

The spectrum from the NMR for this molecule is not of a high resolution due to the amount of the product isolated. Looking at the COSY, the C-2' proton can be seen meaning the arylation either didn't work or did not go to full conversion.

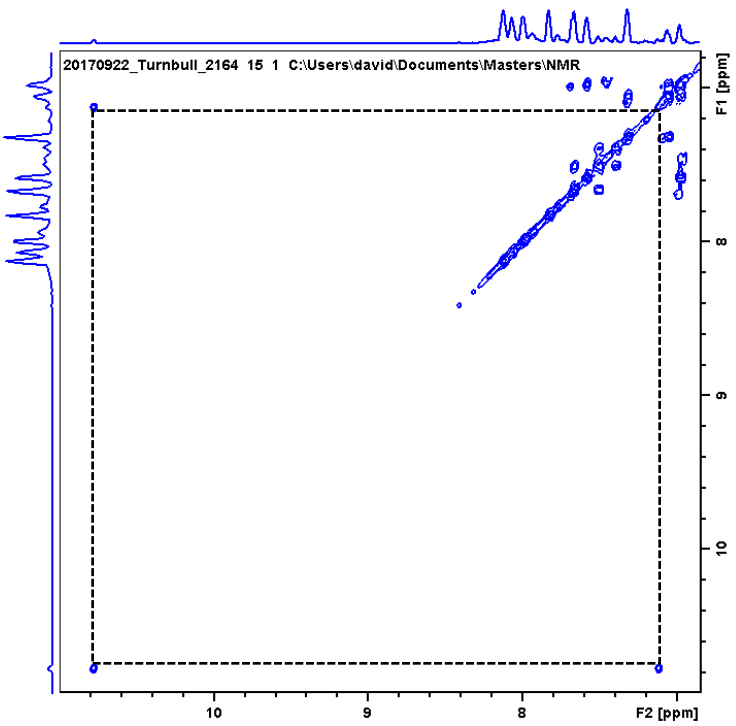


Figure 137: ROESY Spectrum of **13** showing C-2 proton present (DMSO, 700 MHz, 300 K)

The mass spectrum of **13** showed the product pseudo molecular ion. Using MS-MS, the alanine cleaves to form a fragment at m/z 861.4259. The threonine, lysine and an asparagine residue cleaves to form a fragment at m/z 518.2395. Finally, the asparagine cleaves to give the fragment at m/z at 404.1964, which is an indicator that the arylation went onto the tryptophan residue. Below is the fragmentation pattern for the peptide.

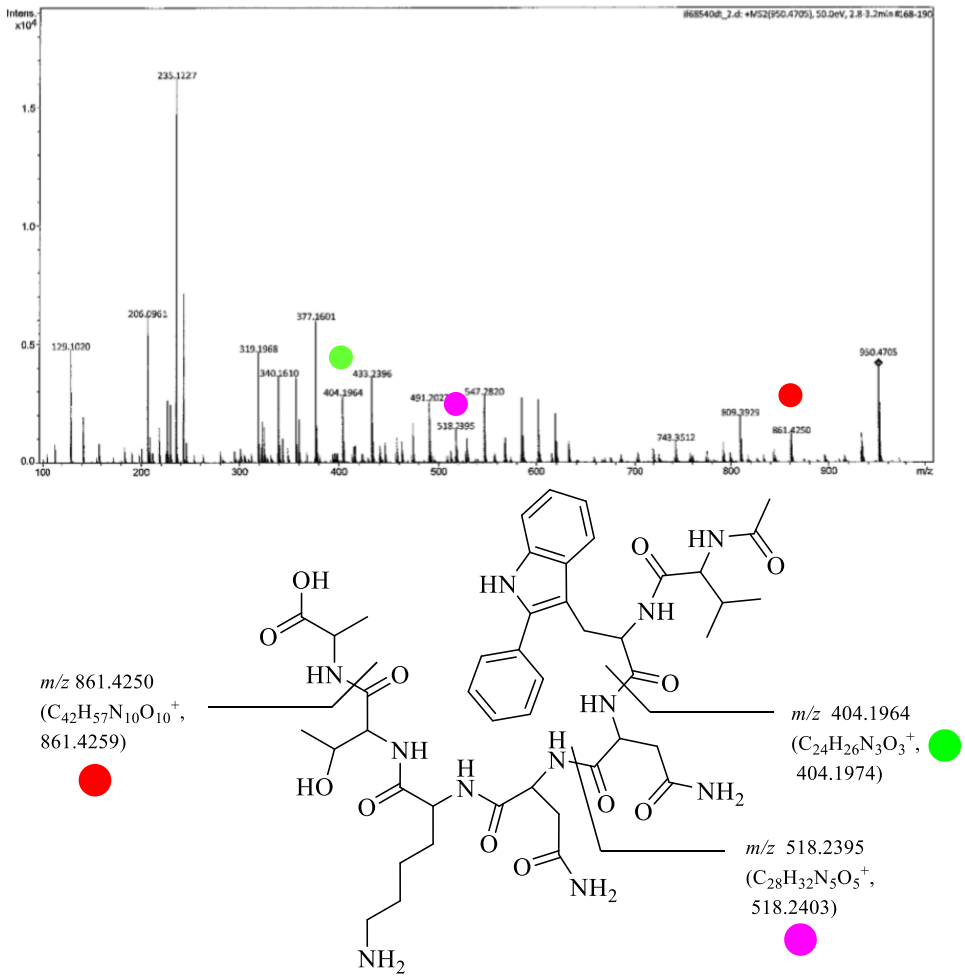


Figure 138: MSMS of product peak for **13** and fragmentation pattern

2.4: NMR Monitoring of the arylation reaction mediated by Pd

Using Ala-Trp-Ala reaction and Ser-Gly-Trp-Ala reaction, NMR monitoring was conducted to evaluate the kinetics. The monitoring looked at two particular peaks which are easily identified, one within the starting material, C-2, and one within the product, C-12.

For Ser-Gly-Trp-Ala, the graph and spectra below shows how the reaction progressed. The difference in height from the C-2 proton to the C-12 proton is due to the C-12 proton being an integral of 2 compared to the integral of 1 for the C2 proton. The initial time of the reaction was removed due to shimmy issues on the spectrometer.

The curve of the C-2 proton going down is different to the kinetic runs done by Thomas Sheridan in his MChem thesis on di-protected tryptophan. Whilst both results show sigmoidal kinetics, which has an induction period where no conversion is seen followed by a sharp increase in conversion and then plateau off, the reactions in this experiment are quicker to start than the simple tryptophan derivative. Without the starting data, from 0 to 30 minutes, it is impossible to say for certain what the induction period is.

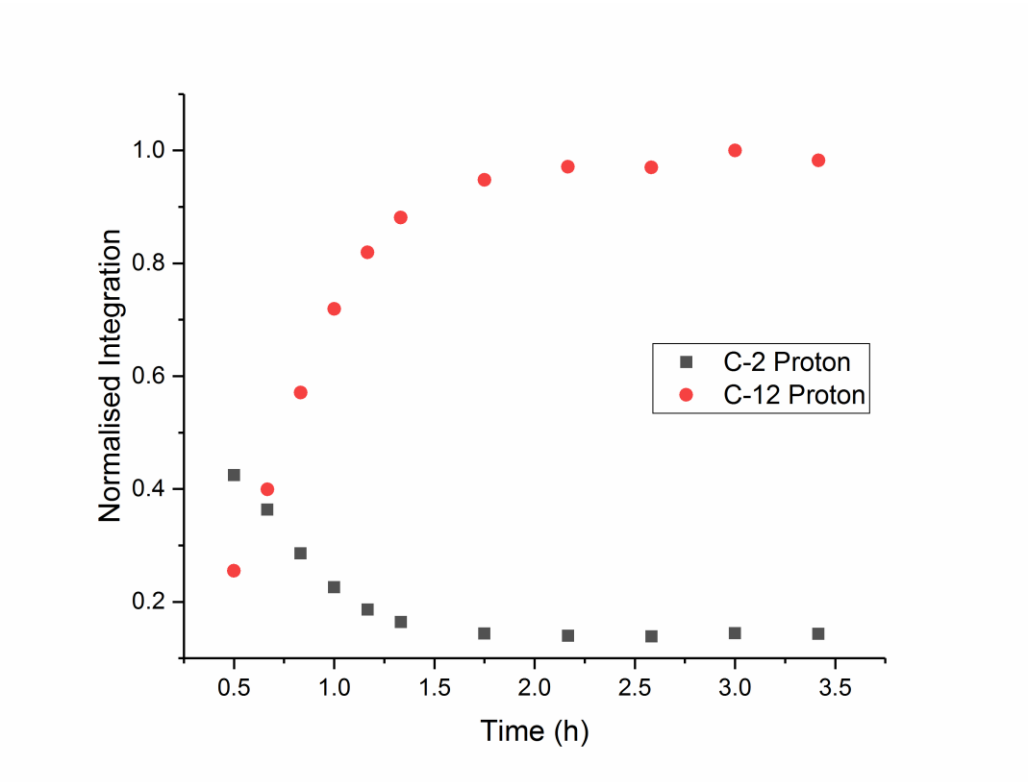


Figure 139: Kinetic plot for Ser-Gly-Trp-Ala reaction

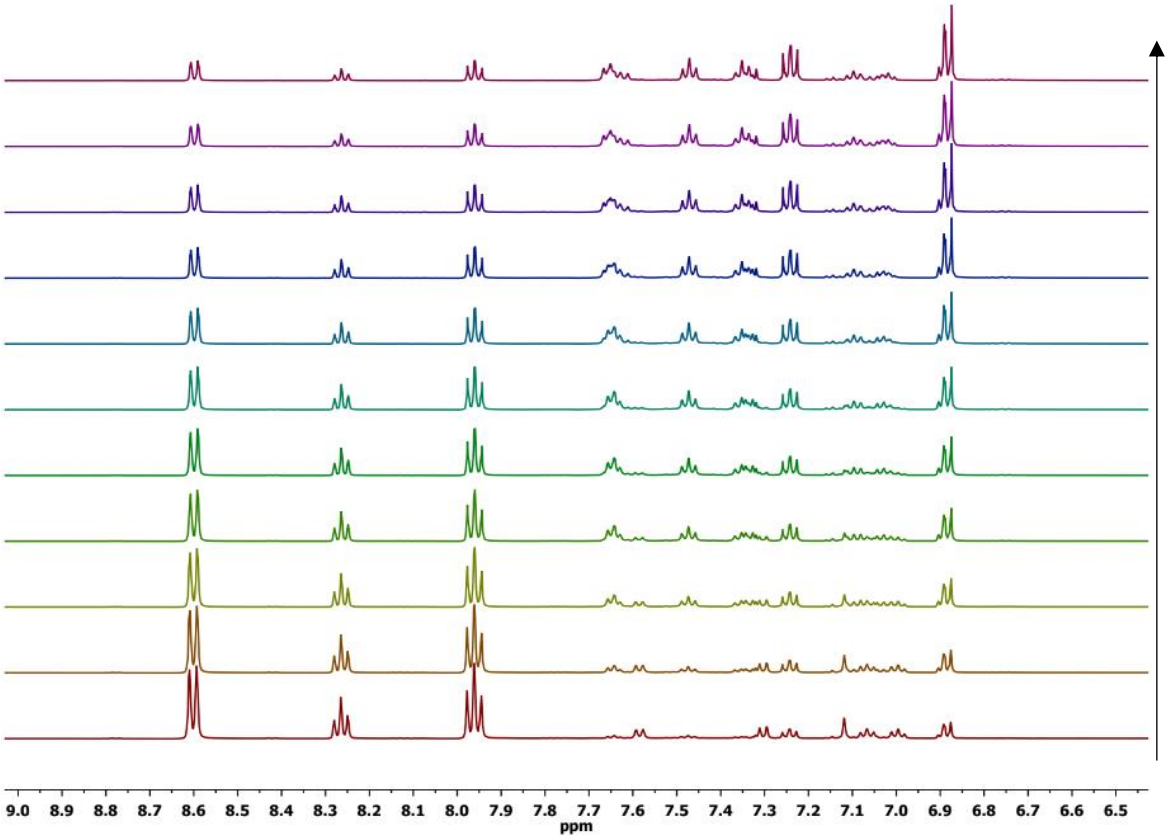


Figure 140: Stacked spectra from Ser-Gly-Trp-Ala **7** reaction - Aromatic region (MeOD-d4, 500 Mhz, 297K) - The arrow represents the flow of time with the bottom representing the starting scan.

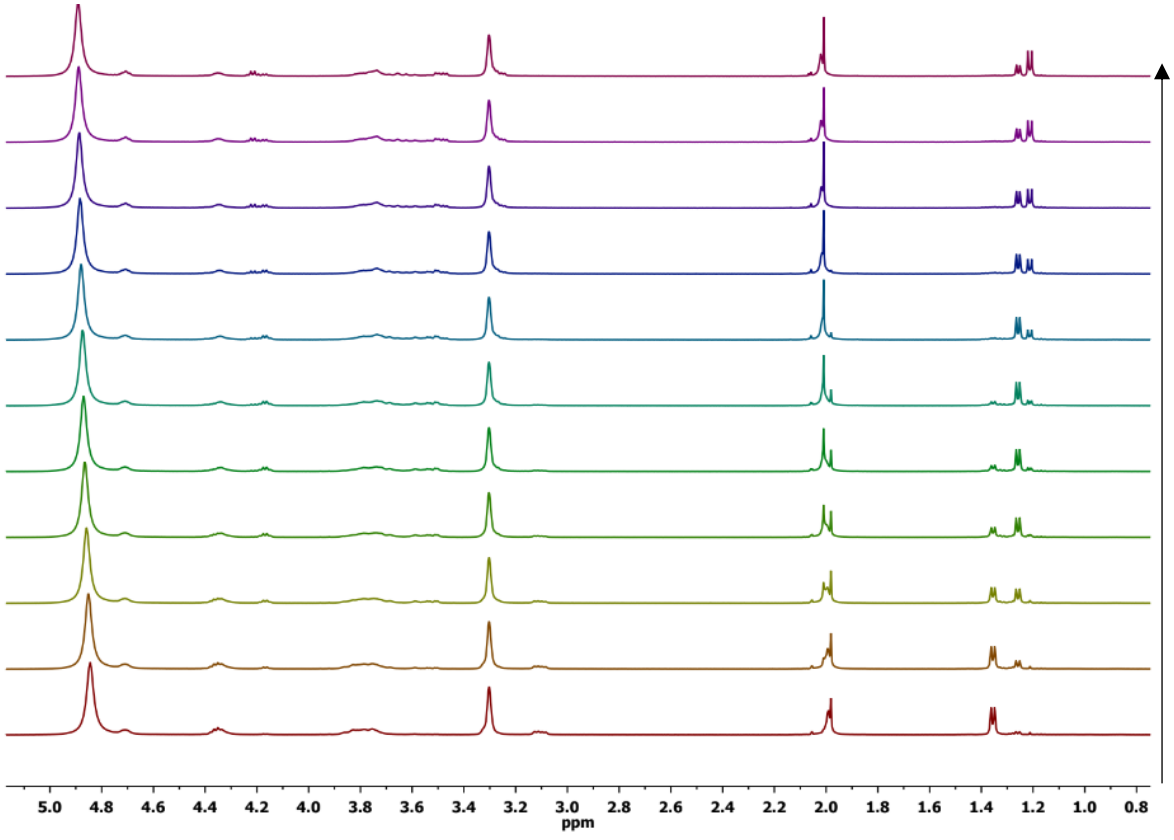


Figure 141: Stacked spectra from Ser-Gly-Trp-Ala **7** reaction - Aliphatic region (MeOD-d4, 500 MHz, 297 K) - The arrow represents the flow of time with the bottom representing the starting scan.

With this reaction, there are two possible pathways that could achieve the product. There are two reactions that take place, the arylation, which the procedure is designed for, and esterification which is a side product. It was not certain if these two reactions occur at the same time or not. From the data collected, it can be seen the arylated product is formed with no methyl ester formation, which should be in the area of δ 3.0 – 3.6 ppm, this suggests that the methyl ester occurs later than the arylation meaning that is could be possible to isolate the unprotected peptide before the protection occurs.

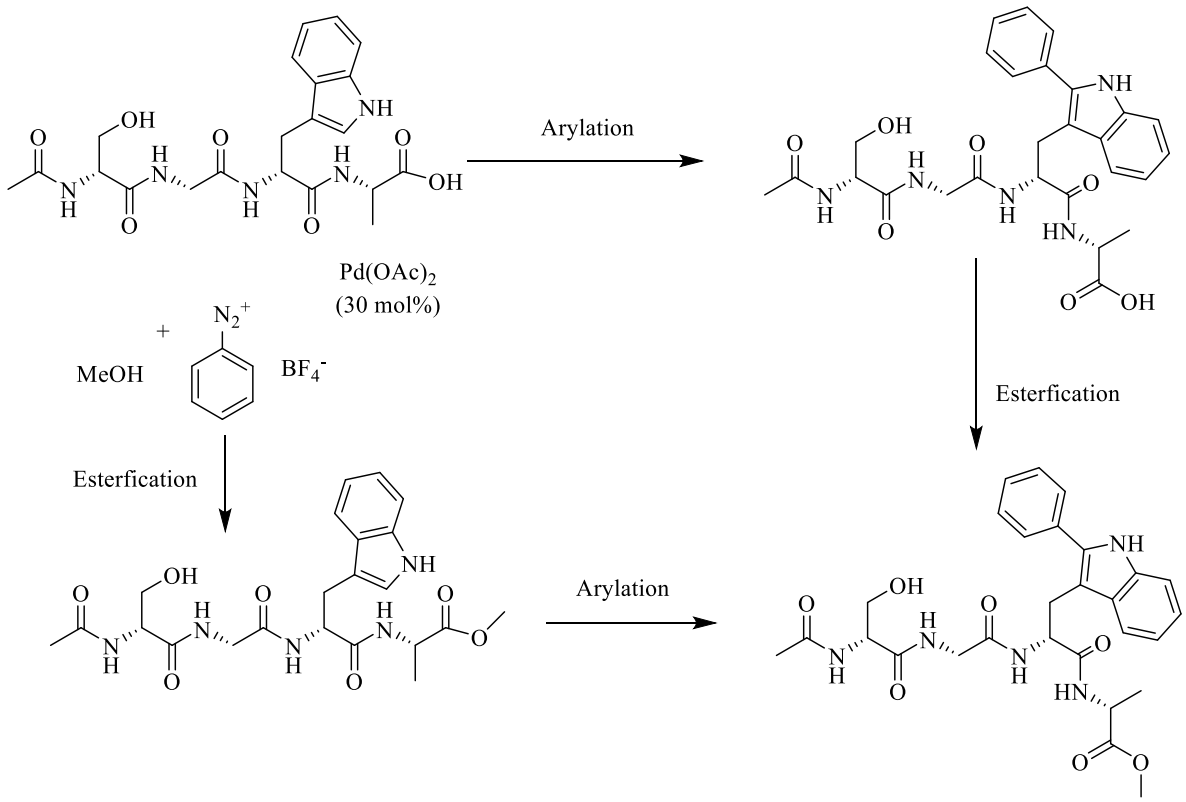
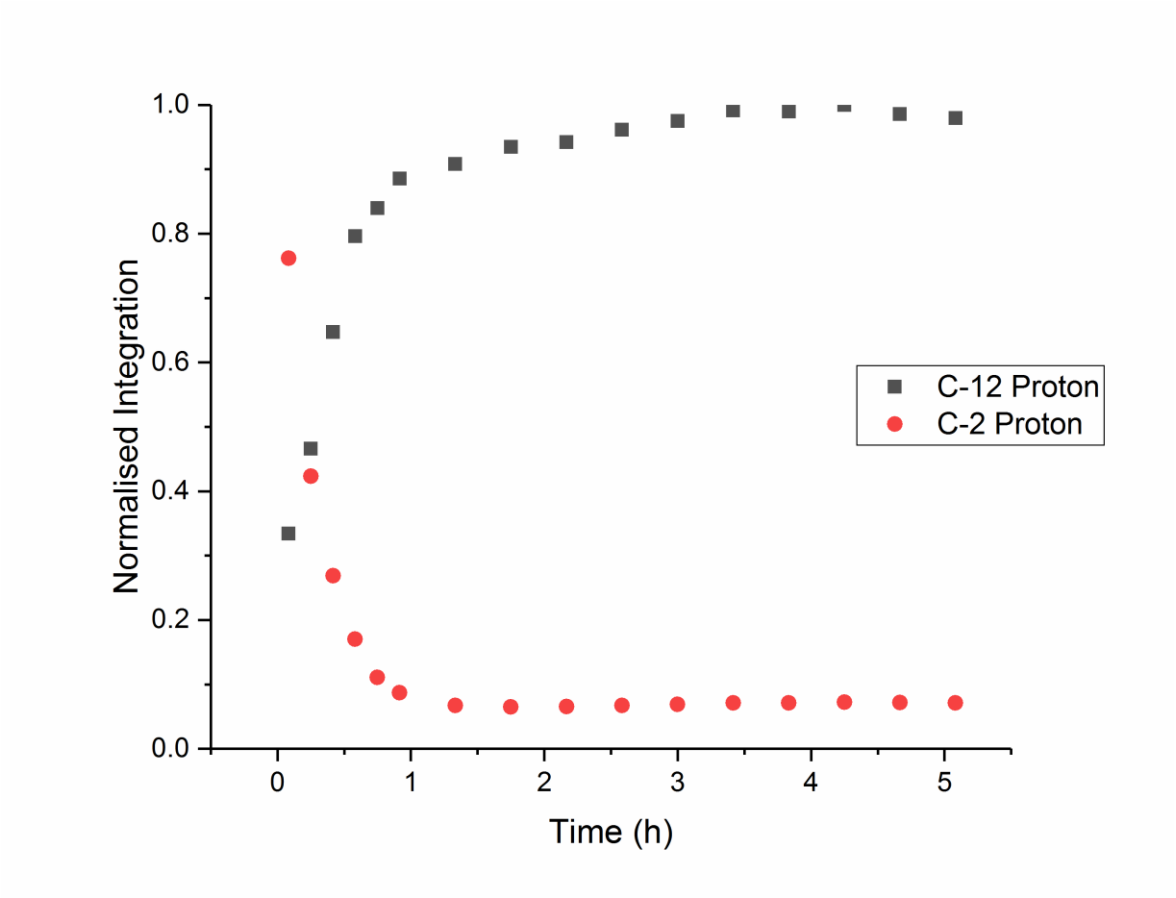


Figure 142: Ser-Gly-Trp-Ala (7) arylation reaction

For Ala-Trp-Ala, the graph and spectra below show how the reaction progressed. The data in this reaction is somewhat confusing, with the height of the C-12 proton being the same as the C-2 proton. This should not be the case as there are two C-12 protons and one C-2 proton.

The reaction itself progressed faster than Ser-Gly-Trp-Ala, with a sharper decrease seen. The initial few runs are also included therefore the lack of induction period can be seen. The speed of the reaction appears to be considerable faster than Ser-Gly-Trp-Ala, this could be due to the alcohol group on the serine group interacting with the aromatic ring or palladium species. From the reaction below, the reaction has achieved 50% completion by the third scan which is roughly 20

minutes into the experiment. *Figure 143: Kinetic plot for Ala-Trp-Ala reaction*



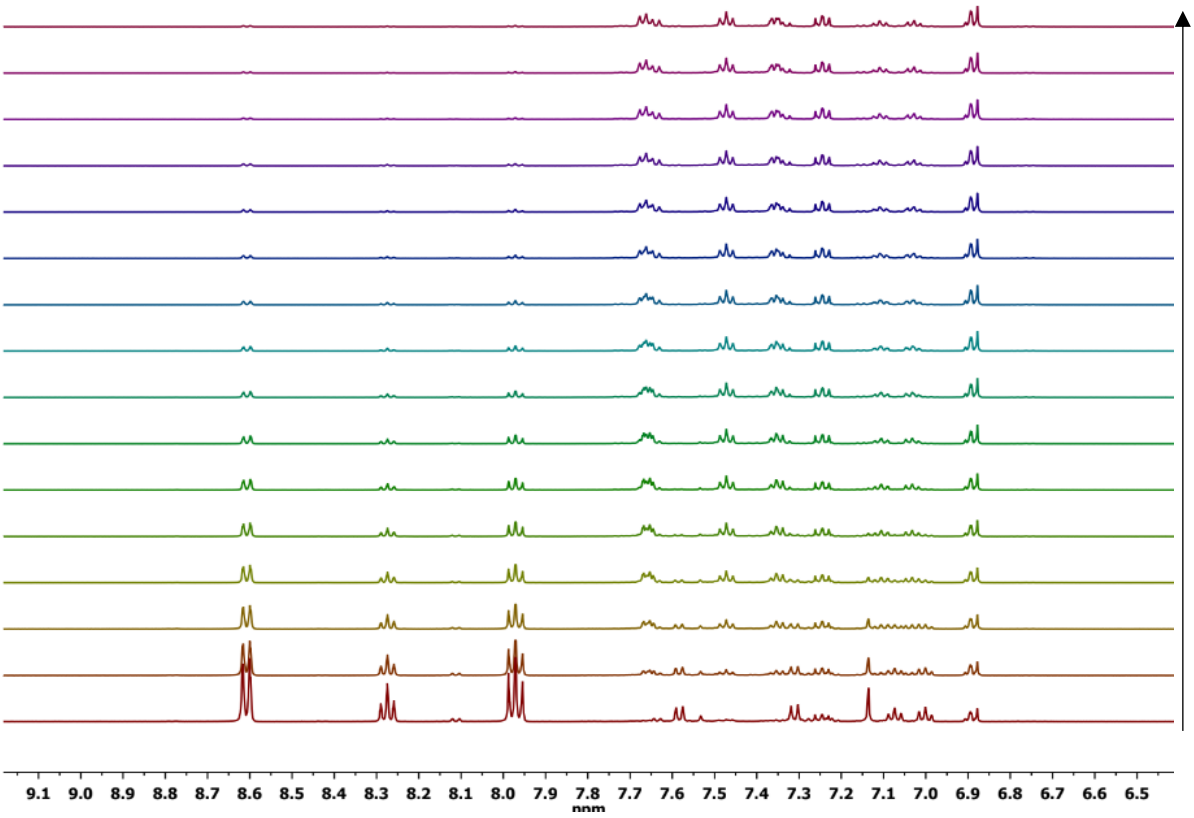


Figure 144: Stacked spectra from Ala-Trp-Ala **6** reaction - Aromatic region (MeOD-d4, 500 Mhz, 297K) The arrow represents the flow of time with the bottom representing the starting scan.

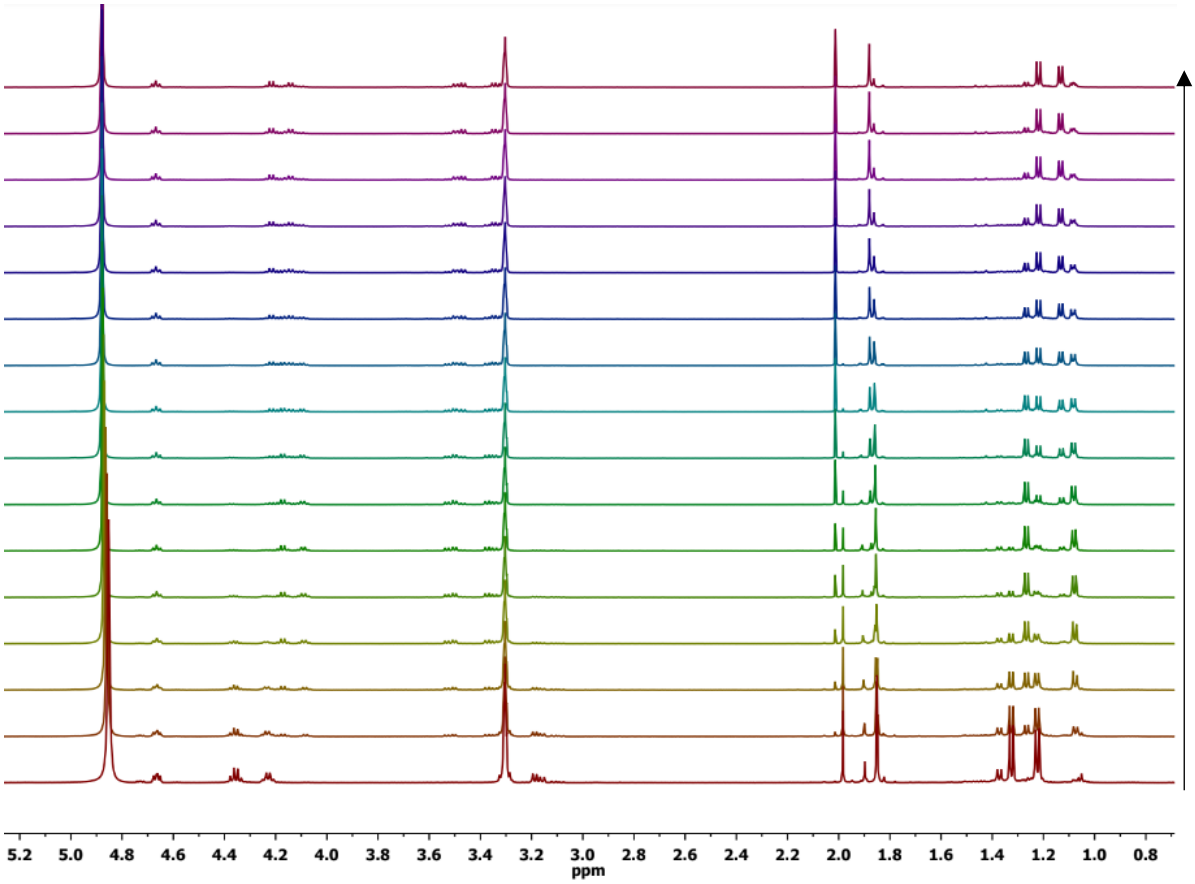


Figure 145: Stacked Spectra from Ala-Trp-Ala **6** reaction - Aliphatic region (MeOD-d4, 500 MHz, 297 K). The arrow represents the flow of time with the bottom representing the starting scan.

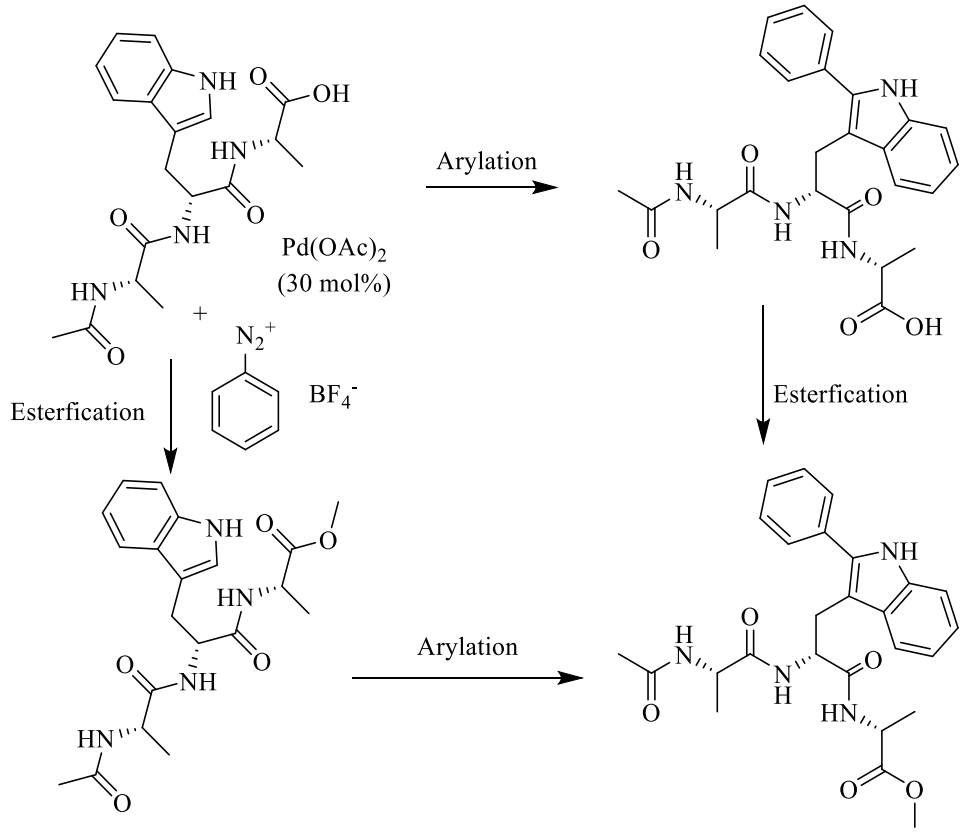


Figure 146: Ala-Trp-Ala (6) arylation reaction

Chapter 3: Conclusions

From the results presented in the previous chapter, the following conclusion can be made.

3.1: Arylation procedure – Assessment of impact on chiral centre

From arylating di-protected tryptophan with several different aryl groups (phenyl, mesityl and naphthalene), the chirality is not affected as seen in the chiral HPLC work. The racemic tryptophan produced racemic product and it could be seen that the $_{L}$-tryptophan product has the same retention time as the appropriate enantiomer in the racemic product.

This also shows that the procedure is compatible with $_{L}$ or $_{D}$ tryptophan, showing that racemistion does not occur.

3.2: Arylation of peptides

The procedure had been shown to work on peptides in previous papers from the Fairlamb group, and in this work this list of peptides was expanded. Using Ala-Trp-Ala-Tyr-Ala (8), in an attempt to use the second aromatic ring to possible interact with the palladium catalyst, and Val-Trp-Asn-Asn-Lys-Thr-Ala (9), which is part of the active site of an enzyme in a certain bacteria species. Ala-Trp-Ala-Tyr-Ala (8) worked to full completion with no arylation occurring on the tyrosine aromatic ring. For Val-Trp-Asn-Asn-Lys-Thr-Ala (9), arylation was shown to occur but full conversion from starting material to product was not seen meaning that a higher amount of aryl diazonium salt was needed. From doing MSMS, the arylated tryptophan product can be seen.

This is a significant result as it is possible for arylation to occur at the lysine amine group, which is a big issue as lysine is common in proteins and peptides. Since it did not, this shows the robustness of the procedure.

3.3: NMR conformation study of the peptides

By using ROESY NMR, it was possible to determine which protons are close in proximity to each other. The main point of this was to see if the conformation changes when arylated. It can be seen in the peptides that once arylated, there are definitely changes in the proton interactions. One change in interaction is between the amide proton on the glycine residue and tryptophan residue in Ser-Gly-Trp-Ala (7) (11) and the diastereomeric protons. In the unmodified peptide (7), the interaction can be seen whilst it cannot be seen in the modified peptide (11). Therefore, it can be assumed that if this methodology was used in a functional enzyme or a protein, the structure would change and this could possibly affect the activity of the molecule.

The newly added aryl ring interacts unusually in the molecule, allowing for interactions that wouldn't be associated with it normally, such as the interaction with the 11' protons or the 5' proton with the α and β-carbon on the tryptophan residue when there wasn't an interaction before. This is just an example of some of the changes in interactions when using the methodology.

3.4: NMR monitored reactions

By monitoring the reactions using NMR, the decrease in the C-2 indole proton can be seen and the increase of the product peaks can be seen. From this, we can see that the reaction with peptides has very little to no induction time, which the deprotected tryptophan does. The reaction is also seen to go quickly for both peptides but particularly the Ala-Trp-Ala. There are two different pathways for the product to be made, arylation followed by esterification or vice versa, through the work done here, it can be seen that the arylation occurs before the esterification meaning that it should be possible to isolate the monoprotected peptide.

Chapter 4: Future Work

It is clear that there is still much work to be done in this area with constant effort to expand the uses of tryptophan modification in peptide modification.

Firstly, it would be useful to produce a range of peptides which consist of three amino acids, alanine, tryptophan and another varying amino acid, shown in Figure 146. This should show if any amino acids, most likely cysteine, can interfere with the procedure.

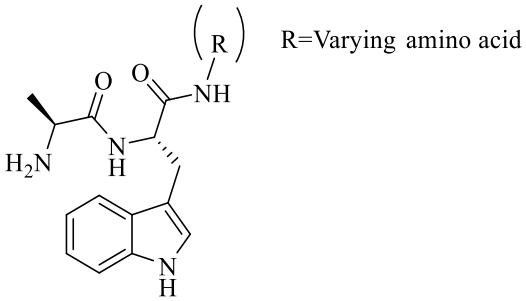


Figure 147: The structure of the peptide that finds interfering residues

Next, it is worth investigating if any amino acids have a promoting effect for the arylation reaction. Possible by using phenyl alanine on either side of the tryptophan residue, the arylation could be enhanced. If there are promoting effects being seen, competition reactions could be done with two tryptophan residues in the same molecule next to the promoting residues.

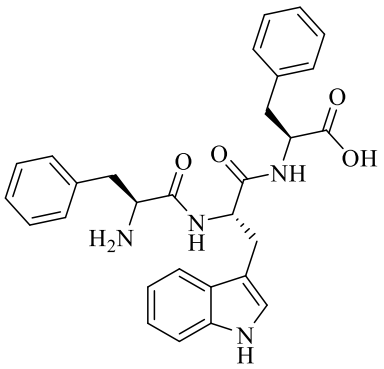


Figure 148: The proposed structure of a peptide that could promote the *rate of reaction*

It can be seen the methodology works for the benzene diazonium salt but would it be possible to use other aryl groups, such as a mesityl group, or would the size of alternative aryl groups be hindered by structure of the peptide.

Further work needs to be done to try and use water as a solvent for the reaction instead of methanol as to make it suitable for cell use. Once this is done, an attempt to modify a protein should be made.

Finally, computational studies should be done to determine the most likely conformation of the peptides and then compared to the data presented in this report.

Chapter 5: Experimental

Commercially available reagents were purchased from Sigma Aldrich, Fluorochem, Fisher Scientific, VWR, Alfa Aesar, Acros Organics or TCI and used as received unless otherwise noted. Room temperature (rt) refers to reactions where no thermostatic control was applied and was recorded as 16 – 22 °C. Petrol refers to the fraction of petroleum ether boiling in the range of 40 – 60 °C. Triethylamine were distilled from potassium hydroxide and stored under nitrogen. Air sensitive procedures were performed using standard Schlenk techniques and carried out in oven- or flame-dried glassware. Nitrogen gas was oxygen free and dried immediately prior to use by passing through a column of potassium hydroxide pellets and silica.

Thin layer chromatography (TLC) analysis was performed using Merck 5554 or Fluorochem aluminium backed silica plates and visualised using UV light (λmax = 254 nm). All flash column chromatography was performed using Fluorochem silica gel 60 (particle size 40 – 63 μm) and a solvent system as stated in the text. Retardation factors are quoted to two decimal places. Melting points were recorded using a Stuart digital SMP3 machine and are quoted to the nearest whole number. Where applicable, decomposition (dec) was noted. All NMR spectra were recorded on either a Jeol ECS400, Jeol ECX400 or 700 spectrometer and processed using MestReNova and TopSpin. Spectra were typically recorded at 298K, unless otherwise specified. Chemical shifts are reported in parts per million (ppm). Coupling constants are reported in Hz and quoted to ±0.5 Hz. Multiplicities are described as singlet (s), doublet (d), triplet (t), quartet (q), quintet (quin), sextet, (sext), heptet (hept), multiplet (m), apparent (app) and broad (br).

^{1}H NMR spectra were typically recorded at 400 MHz. All the NMR spectra for the peptides were done at 700 MHz. Chemical shifts are internally referenced to residual undeuterated solvent and given to two decimal places.

^{13}C NMR spectra were recorded at 101 MHz. Chemical shifts are internally referenced to residual undeuterated solvent and given to one decimal place.

^{19}F NMR spectra were recorded at 376 MHz and obtained with ^{1}H decoupling. Chemical shifts are externally referenced to $CFCl_3$ and given to one decimal place.

Mass spectrometry was performed using a Bruker Daltronics micrOTOF spectrometer using electrospray ionization (ESI). Mass to charge ratios (m/z) are reported in Daltons with percentage abundance in parentheses along with the corresponding fragment ion, where known. Where complex isotope patterns were observed, the most abundant ion was reported. High resolution mass spectra are reported with <5 ppm error (ESI).

IR spectrometry was performed using a Bruker Alpha FT-IR spectrometer. Signals are reported in wavenumbers (cm^{-1}) to the nearest whole number.

UV-visible spectroscopy was performed using a Jasco V-560 spectrometer. A baseline in the appropriate solvent was obtained prior to recording spectra.

General Procedure:

General Procedure 1: Arylation of Tryptophan Residue

To a microwave tube was added tryptophan 1 (50 mg, 0.192 mmol, 1 eq.), the appropriate aryldiazonium salt (0.192 mmol, 1 eq.), Pd(OAc)2 (2 mg, 9.6 μmol, 5 mol%) and EtOAc (5 mL). The reaction mixture was stirred at RT for 16 h. After 16 h the resulting brown reaction mixture was filtered through Celite then washed with sat. aq. $NaHCO_3$. The organic layer was collected and dried over $MgSO_4$, filtered and evaporated to give a brown solid. When purification was required, it was performed using dry-loaded flash column chromatography with a SiO_2 stationary phase and the solvent system specified for each compound.

General Procedure 2: Synthesis of Peptides via SPPS

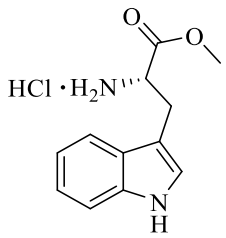
To a specialized syringe, alanine-2-chlorotrityl resin was added with DMF and rotated for 30 minutes to swell the resin up. It was then filtered into the manifold. 5 equivalents of DIPEA was added to a solution of 5 equivalents of amino acid and 5 equivalents of HCTU dissolved in minimum volume of DMF and then added to the resin. The mixture was then rotated for 60 minutes, and then the resin was filtered and washed with DMF (3 x 2 minutes

with rotation). Between each coupling, the fmoc of the latest amino acid was removed using 20% piperdine in DMF being added to the resin. The mixture was rotated for 2 minutes then filtered, this was done 5 times. The resin was then washed with DMF (5 x 2 minutes with rotation) ready for the next coupling. Once the peptide has been synthesized and the fmoc has been cleaved, the resin was washed with DCM (3 x 2 minutes with rotation), and methanol (3 x 2 minutes with rotation). The resin was dried on a vacuum manifold and further dried on a high vacuum line overnight. A cleavage solution, 38:1:1 TFA: H_2O: TIS, was added to the resin and rotated for 60 minutes. The reaction mixture was drained into -78 ^oC diethyl ether and centrifuged at 5000 rpm for 5 minutes at 4 ^oC. The supernatant was carefully decanted and subsequently resuspended, centrifuged and supernatant decanted three more times. The precipitated pellet was dissolved in water and acetonitrile and lyophilized.

General Procedure 3: Arylation of Peptides

To a microwave tube was added peptide (20 mg, 1 eq.), benzenediazonium tetrafluoroborate (1.1 eq.), $Pd(OAc)_2$ (20 mol%) and MeOH (2 mL), which was stirred at room temperature for 24 hours. The resulting brown reaction mixture was filtered through Celite with MeOH (5 mL) and then added dropwise to diethyl ether at -78 ^oC. The peptide crashes out and it then centrifuged 3 times at 5000 rpm for 5 minutes, draining off ether after each centrifuge, and then fill it back up with more ether for next centrifuge. Once done, drain off the ether, retaining as much peptide as possible, and leave to dry. Once dry, add water and acetonitrile to dissolve the peptide for freeze drying. Leave in the freeze dryer until dry.

Methyl (2R)-2-amino-3-(1H-indol-3-yl)propanoate hydrochloride (1)

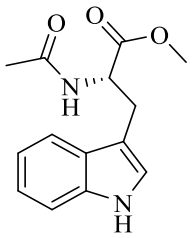


150 mL of MeOH was placed into a 500 mL round bottom flask which was cooled to −15 °C, addition of thionyl chloride (12.9 mL, 21.06 g, 177 mmol, 2.4 eq.) was then made dropwise. After complete addition, L-Tryptophan 73 (14.9994 g, 73.5 mmol, 1 eq.) was added in three portions, resulting in a white suspension. The mixture was then warmed to RT and stirred for 24 h, during which time a clear yellow solution was formed. 5 mL of deionised water was added to the reaction mixture and the solvent evaporated to afford the title compound as an off-white solid (18.11 g, 97%).

DRT 1 006

M.P. 212-213 °C (lit. 214 °C); ^{1}H NMR (400 MHz, CD_3OD, δ): 7.51 (d , J= 8.0 Hz, 1H), 7.36 (d, J=8 Hz, 1H), 7.18 (s, 1H), 7.11 (t , J= 16.0, 8.0 Hz, 1H), 7.03 (t , J= 16.0, 8.0 Hz, 1H), 4.29 (t , J= 12.0, 8.0 Hz, 1H), 3.76 (s, 3H), 3.43 (dd , J= 12.0, 8.0, 4.0 Hz, 1H), 3.34 (dd , J= 12.0, 8.0, 4.0 Hz, 1H); ^{13}C NMR (101 MHz, CD_3OD, δ), 170.8, 138.3, 128.1, 125.7, 122.9, 120.3, 118.8, 112.7, 107.4, 54.6, 53.7, 27.5; ESI–MS m/z (ion, rel. %): 219 ($[C_{12}H_{15}N_2O_2]^+$, 100); ESI–HRMS m/z: 219.1125 $[C_{12}H_{15}N_2O_2]^+$ ($C_{12}H_{15}N_2O_2$ requires 219.1128); Elemental anal.: C 56.08, H 5.90, N 10.78 ($C_{12}H_{15}ClN_2O_2$ requires C 56.58, H 5.94, N 11.00).

Methyl (2R)-2-acetamido-3-(1H-indol-3-yl)propanoate (2)

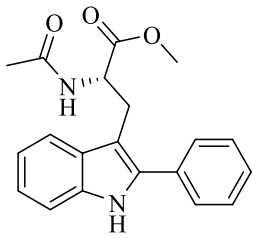


Mono-Protected tryptophan (2.9996 g, 13.7 mmol, 1 eq.) was added to a 2-necked round-bottom flask. To the flask, distilled Et_3N (2 mL, 1.45 g, 14.3 mmol, 1.05 eq.) and THF (150 mL) was added and stirred to give a white suspension before being cooled to 0 °C. Acetic anhydride (1.4 mL, 1.5 g, 15.1 mmol, 1.1 eq.) was then added in one portion. The reaction was then stirred for 2 hours under reflux to give a white suspension. After 2 hours, the mixture was left until it reached ambient temperature and then 150 mL of water was added. The mixture was then extracted with EtOAc (3 x 150 mL). The organic layers were combined and then washed sequentially with 1 M aq. HCl (100 mL), sat. aq. $NaHCO_3$ (100 mL) and brine (100 mL). The organic layer was collected, dried over $MgSO_4$, filtered and evaporated to afford a clear brown oil. Upon trituration with ether, an offish white powder was afforded (2.1798g, 60%).

DRT 1 013

M.P. 152 °C (lit. 155-156 °C); ^{1}H NMR (400 MHz, Chloroform-*d*) δ 8.26 (s, 1H), 7.57 – 7.48 (m, 5H), 7.36 (dt, *J* = 8.1, 0.9 Hz, 2H), 7.19 (ddd, *J* = 8.2, 7.2, 1.1 Hz, 2H), 7.12 (ddd, *J* = 8.0, 7.1, 1.1 Hz, 1H), 6.97 (d, *J* = 2.3 Hz, 1H), 6.03 (d, *J* = 7.6 Hz, 1H), 4.96 (dt, *J* = 7.9, 5.3 Hz, 1H), 3.70 (s, 3H), 3.41 – 3.24 (m, 3H), 1.96 (s, 3H); ^{13}C NMR (101 MHz, Choroform-*d*) δ 172.6, 169.9, 136.2, 127.9, 122.8, 122.4, 119.8, 118.6, 111.4, 110.1, 77.2, 53.2, 52.5, 27.7, 23.4; ESI–HRMS m/z: 283.1055 $[M^+Na]^+$ ($C_{14}H_{16}N_2O_3Na$ requires 283.1053); Elemental Anal.: C 64.6, H 6.3, N, 9.9 ($C_{14}H_{16}N_2O_3$ requires C 64.60, H 6.20, N 10.76)

Methyl (2R)-2-acetamido-3-(2-phenyl-1H-indol-3-yl)propanoate hydrochloride (3)

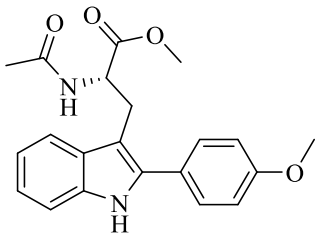


Synthesised using general procedure C with phenyl aryldiazonium salt (37 mg, 0.192 mmol, 1 eq.) to afford the title compound as an off-white solid (65 mg, quant.).

DRT 2 151

Rf 0.25 (EtOAc/petrol, 1:1, v/v); M.P. 82–83 °C (lit.[45] 85–86 °C); ^{1}H NMR (400 MHz, $CDCl_3$, δ): 8.43 (s, 1H), 7.59–7.52 (m, 3H), 7.45 (d, J = 15.1 Hz, 2H), 7.39–7.32 (m, 2H), 7.22–7.17 (m, 1H), 7.13 (t, J = 8.0 Hz, 1H), 5.81 (d, J = 8 Hz, 1H), 4.89– 4.76 (dt, J = 8.0, 5.0 Hz, 1H), 3.54 (d, J = 5.5 Hz, 2H), 3.29 (s, 3H), 1.64 (s, 3H). ^{13}C NMR (101 MHz, ($CDCl_3$, δ): 172.3, 169.8, 136.1, 135.8, 133.3, 129.5, 129.3, 128.4, 128.2, 122.6, 120.1, 119.0, 111.1, 106.8, 52.9, 52.1, 26.7, 23.0; ESI–HRMS m/z: 337.1547 $[M^+H]^+$ ($C_{20}H_{21}N_2O_3^+$ requires 337.1507)

Methyl (2S)-2-acetamido-3-[2-(4-methoxyphenyl)-1H-indol-3-yl]propanoate (4)



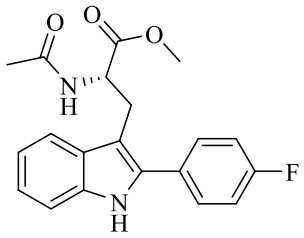
Synthesised using general procedure 1 (with a reaction time of 24 h) with Para-methyoxy aryldiazonium salt (43 mg, 0.192 mmol, 1 eq.) to afford the title compound as a brown solid (70 mg, quant.).

DRT 2 152

Rf 0.15 (petrol/EtOAc, 1:1, v/v); M.P. 200–203 °C; ^{1}H NMR (400 MHz, $CDCl_3$, δ): 8.41 (br s, 1H), 7.54 (d, J = 8.0 Hz, 1H), 7.50–7.40 (m, 2H), 7.32 (d, J = 8.0 Hz, 1H), 7.17 (ddd, J = 8.0, 7.0, 1.0 Hz, 1H), 7.12 (ddd, J = 8.5, 8.0, 1.0 Hz, 1H), 7.01– 6.91 (m, 2H), 5.85 (d, J = 8.0 Hz, 1H), 4.82 (dt, J =

8.0, 5.5 Hz, 1H), 3.83 (s, 3H), 3.49 (d, J = 5.5 Hz, 2H), 3.34 (s, 3H), 1.68 (s, 3H); ^{13}C NMR (101 MHz, ($CDCl_3$, δ): 172.4, 169.8, 159.5, 136.1, 135.7, 129.7, 129.5, 125.6, 122.3, 120.0, 118.7, 114.6, 111.0, 106.0, 55.5, 53.0, 52.2, 26.8, 23.0; ESI–HRMS m/z: 367.1646 $[M+H]^+$ ($C_{21}H_{22}N_2O_4^+$ requires 367.1613)

Methyl (2S)-3-[2-(4-fluorophenyl)-1H-indol-3-yl]-2-acetamidopropanoate (5)

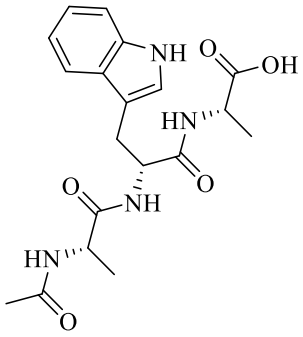


Synthesised using general procedure 1 with para-fluoro aryldiazonium salt (40 mg, 0.192 mmol, 1 eq.) to afford the title compound as a brown solid (68 mg, quant.).

DRT 2 153

Rf 0.23 (petrol/EtOAc, 1:1, v/v); ^{1}H NMR (400 MHz, $CDCl_3$, δ): 8.17 (br s, 1H), 7.59–7.50 (m, 3H), 7.36 (d, J = 8.0 Hz, 1H), 7.24–7.12 (m, 4H), 5.82 (d, J = 8.0 Hz, 1H), 4.84 (dt, J = 8.0, 5.5 Hz, 1H), 3.50 (m, 2H), 3.33 (s, 3H), 1.72 (s, 3H); ^{13}C NMR (101 MHz, ($CDCl_3$, δ): 172.3, 169.7, 162.6 (d, 1JC–F = 249.0 Hz), 135.8, 135.1, 130.2 (d, 3JC–F = 8.0 Hz), 129.5, 129.4 (d, 4JC–F = 3.5 Hz), 122.8, 120.3, 119.0, 116.3 (d, 2JC– F = 21.5 Hz), 111.1, 107.0, 52.9, 52.2, 26.8, 23.1; ^{19}F NMR (376 MHz, $CDCl_3$, δ): −112.8–−112.9 (m); ESI–HRMS m/z: 355.1449 $[M+H]^+$ ($C_{20}H_{20}FN_2O_3^+$ requires 355.1413)

***Ac*-AlaTrpAla-*OH* (6)**

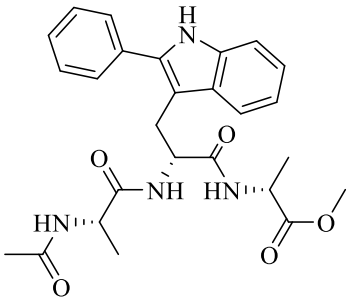


Using general procedure 2, with an acetyl protection at the *N*-terminus (30 mg).

DRT 2 AWA

^{1}H NMR (700 MHz, DMSO) δ 10.81 (s, 1H), 8.14 (d, *J* = 7.2 Hz, 1H), 7.98 (d, *J* = 7.2 Hz, 1H), 7.84 (d, *J* = 8.2 Hz, 1H), 7.58 (d, *J* = 7.9 Hz, 1H), 7.30 (d, *J* = 8.1 Hz, 1H), 7.13 (d, *J* = 2.1 Hz, 1H), 7.04 (t, *J* = 7.5 Hz, 1H), 6.96 (t, *J* = 7.3 Hz, 1H), 4.51 (td, *J* = 8.6, 4.4 Hz, 1H), 4.24 – 4.15 (m, 2H), 3.15 (dd, *J* = 14.8, 4.3 Hz, 1H), 2.95 (dd, *J* = 14.9, 9.0 Hz, 1H), 1.78 (s, 3H), 1.26 (d, *J* = 7.3 Hz, 3H), 1.10 (d, *J* = 7.1 Hz, 3H). ^{13}C NMR (176 MHz, DMSO) δ 174.21, 172.37, 171.36, 169.44, 136.23, 127.64, 123.90, 121.02, 118.67, 118.39, 111.44, 110.12, 53.13, 48.52, 47.79, 27.61, 22.69, 18.11, 17.39. ESI–HRMS m/z: 389.1819 $[M+H]^+$ ($C_{19}H_{25}N_4O_5^+$ requires 389.1780), 411.146 $[M+Na]^+$ ($C_{19}H_{24}N_4O_5Na^+$ requires 411.1644), 433.1458 $[M+2Na]^+$ ($C_{19}H_{23}N_4O_5Na_2^+$ requires 433.1464)

***Ac*-AlaTrp(Ph)Ala-O*Me* (10)**



Synthesized using general procedure 3, using 15.4 mg peptide **1** (0.04 mmol, 1 eq.), 9.2 mg benzenediazonium tetrafluoroborate (0.047. 1.2 eq.) and 2 mg $Pd(OAc)_2$ (0.0089 mmol, 20 mol%). (19 mg, quant.)

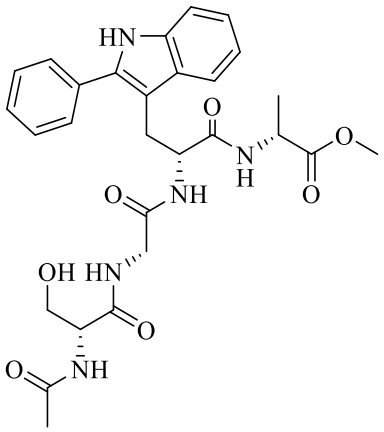
DRT 2 160

To a sample vial was added peptide **1** (5 mg, 0.013 mmol, 1 eq.), benzenediazonium tetrafluoroborate (4.9 mg, 0.028 mmol, 2 eq.), Pd(OAc)2 (0.75 mg, 0.0026mmol, 30 mol%) and d4-MeOD (0.5 mL), which was then monitored using NMR for 6 hours.

DRT 2 176

^{1}H NMR (700 MHz, DMSO) δ 11.16 (s, 1H), 8.02 – 7.98 (m, 2H), 7.88 (d, *J* = 8.4 Hz, 1H), 7.68 (d, *J* = 7.7 Hz, 3H), 7.48 (t, *J* = 7.6 Hz, 2H), 7.37 (t, *J* = 7.4 Hz, 1H), 7.33 (d, *J* = 8.0 Hz, 1H), 7.08 (t, *J* = 7.5 Hz, 1H), 6.98 (t, *J* = 7.5 Hz, 1H), 4.61 (dd, *J* = 14.9, 7.6 Hz, 1H), 4.23 – 4.17 (m, 1H), 4.10 (dt, *J* = 13.8, 7.0 Hz, 2H), 3.48 (s, 3H), 3.33 (dd, *J* = 14.6, 6.5 Hz, 1H), 3.10 (dd, *J* = 14.5, 7.8 Hz, 1H), 1.79 (s, 3H), 1.19 (d, *J* = 7.3 Hz, 4H), 1.01 (d, *J* = 7.1 Hz, 3H). ^{13}C NMR (176 MHz, DMSO) δ 172.53, 171.96, 170.90, 169.51, 135.93, 135.27, 132.75, 128.99, 128.59, 128.11, 127.29, 121.38, 119.29, 118.65, 110.94, 107.62, 53.72, 51.74, 48.59, 47.63, 27.45, 22.48, 17.51, 16.98. ESI–MS m/z: 479.11 $[M+H^+]$($C_{26}H_{31}N_4O_5^+$ requires 479.2250), 501.11 $[M+Na^+]$ (C26H30N4NaO$_5^+$ requires 501.2069)

***Ac*-SerGlyTrp(Ph)Ala-O*Me* (11)**



To a microwave tube was added peptide **6** (20 mg, 0.0433 mmol, 1 eq.), benzenediazonium tetrafluoroborate 2a (9.2 mg, 0.0476 mmol, 1.1 eq.), Pd(OAc)2 (2 mg, 0.0087 mmol, 20 mol%) and MeOH (2 mL), which was stirred at 37 °C for 8 h. The resulting brown reaction mixture was filtered through Celite with MeOH (5 mL) and the solvent removed under reduced pressure. This crude mixture was purified by preparative TLC (6% MeOH/CH2Cl2) to give an off white solid (10.8 mg, 45%).

Made by Josh Bray, part of the ACS Catalysis paper: Mild and Regioselective Pd(OAc)2-Catalyzed C–H Arylation of Tryptophans by [ArN2]X, Promoted by Tosic Acid

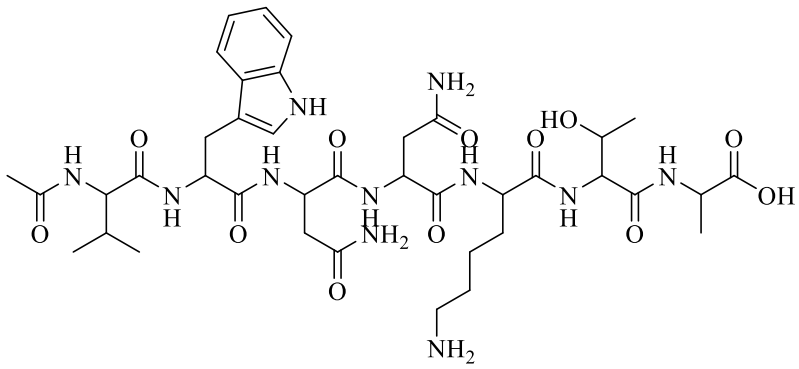
To a microwave tube was added peptide **2** (5 mg, 0.011 mmol, 1 eq.), benzenediazonium tetrafluoroborate (4.6 mg, 0.024 mmol, 2 eq.), Pd(OAc)2 (0.75 mg, 0.002 mmol, 30 mol%) and d4-MeOD (0.5 mL), which was then monitored using NMR for 4 hours.

DRT 2 177

^{1}H NMR (700 MHz, DMSO) δ 11.16 (s, 1H), 8.20 (d, *J* = 6.9 Hz, 1H), 8.16 (t, *J* = 4.9 Hz, 1H), 8.14 – 8.09 (m, 2H), 7.69 (d, *J* = 7.6 Hz, 3H), 7.49 (t, *J* = 7.5 Hz, 2H), 7.37 (t, *J* = 7.2 Hz, 1H), 7.32 (d, *J* = 8.0 Hz, 1H), 7.08 (t, *J* = 7.4 Hz, 1H), 6.99 (t, *J* = 7.4 Hz, 1H), 5.16 (s, 1H), 4.74 – 4.65 (m, 1H), 4.25 – 4.17 (m, 2H), 3.68 (dt, *J* = 15.0, 7.5 Hz, 1H), 3.55 (dd, *J* = 16.6, 10.3 Hz, 3H), 3.49 (d, *J* = 5.0 Hz, 1H), 3.47 (s, 3H), 3.02 (dd, *J* = 14.4, 7.7 Hz, 1H), 1.86 (s, 3H), 1.21 (d, *J* = 7.2 Hz, 3H). ^{13}C NMR (176 MHz, DMSO) δ 172.98, 171.30, 170.19, 168.76, 136.37, 135.70, 133.21, 129.47, 129.07, 128.57, 127.73, 121.71, 119.70, 119.13, 111.35, 108.05, 62.07, 55.97, 54.25, 52.25, 48.06, 42.61,

28.56, 23.03, 17.42. vmax/cm-1 (ATR) 3821.64, 1734.6, 1637.0, 1528.4, 1451.8, 1375.0, 1340.8, 1305.3, 1222.1, 1154.4, 1057.8, 745.0, 697.4, ESI–MS m/z: 552.2476 [M+H$^+$] ($C_{28}H_{34}N_5O_7^+$ requires 552.2414), 574.2289 [M+Na$^+$](C28H33N5NaO7$^+$ requires 574.2233); Rf = 0.13 (6% MeOH/CH2Cl2).

***Ac*-ValTrpAsnAsnLysThrAla-*OH* (9)**

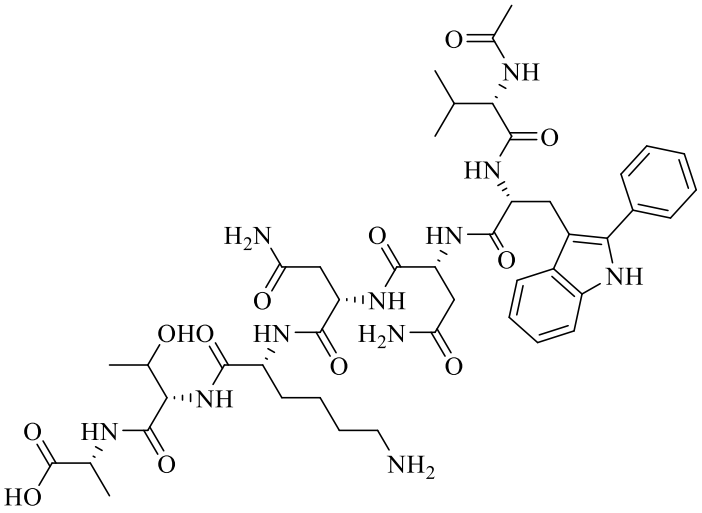


Using general procedure 2, with an acetyl protection at the *N*-terminus (60 mg).

DRT 2 VWNN

^{1}H NMR (700 MHz, DMSO) δ 10.77 (s, 1H), 8.11 (t, *J* = 8.6 Hz, 2H), 7.98 (d, *J* = 7.4 Hz, 2H), 7.93 (d, *J* = 7.0 Hz, 1H), 7.82 (d, *J* = 8.5 Hz, 1H), 7.67 (d, *J* = 8.4 Hz, 1H), 7.62 – 7.55 (m, 4H), 7.45 (d, *J* = 14.3 Hz, 2H), 7.31 (d, *J* = 8.1 Hz, 1H), 7.11 (s, 1H), 7.04 (t, *J* = 7.5 Hz, 1H), 6.96 (dd, *J* = 14.1, 6.6 Hz, 3H), 4.76 (s, 1H), 4.53 (td, *J* = 13.7, 6.5 Hz, 3H), 4.26 (dd, *J* = 13.3, 8.7 Hz, 1H), 4.21 – 4.15 (m, 2H), 4.09 (t, *J* = 7.7 Hz, 1H), 3.98 – 3.92 (m, 1H), 3.12 (dd, *J* = 14.8, 4.2 Hz, 1H), 2.94 (dd, *J* = 14.9, 9.4 Hz, 1H), 2.76 (dt, *J* = 21.2, 10.6 Hz, 2H), 2.59 (dd, *J* = 15.6, 6.9 Hz, 1H), 2.46 – 2.37 (m, 1H), 1.91 (s, 2H), 1.88 (dd, *J* = 13.6, 6.8 Hz, 1H), 1.84 (s, 3H), 1.74 (d, *J* = 7.2 Hz, 1H), 1.63 – 1.56 (m, 1H), 1.51 (td, *J* = 14.9, 7.3 Hz, 3H), 1.38 – 1.29 (m, 3H), 1.27 (d, *J* = 7.3 Hz, 4H), 1.06 (d, *J* = 6.3 Hz, 4H), 0.76 (dd, *J* = 6.6, 3.1 Hz, 7H). ^{13}C NMR (176 MHz, DMSO) δ 173.70, 171.77, 171.59, 171.22, 171.10, 170.95, 170.64, 170.53, 169.40, 169.32, 135.76, 127.04, 123.28, 120.61, 118.14, 117.95, 110.98, 109.69, 66.34, 57.86, 57.80, 52.95, 52.33, 49.62, 49.46, 47.31, 38.52, 36.68, 36.37, 30.45, 29.89, 27.12, 26.20, 22.24, 21.80, 20.82, 19.48, 18.98, 17.88, 17.05; ESI–MS m/z: 874.4417 [M+H$^+$] ($C_{39}H_{60}N_{11}O_{12}^+$ requires 874.4378), 896.4237 [M+Na$^+$] ($C_{39}H_{59}N_{11}O_{12}Na^+$ requires 896.4198), 918.4086 [M+2Na$^+$] ($C_{39}H_{58}N_{11}O_{12}Na_2^+$ requires 918.4017)

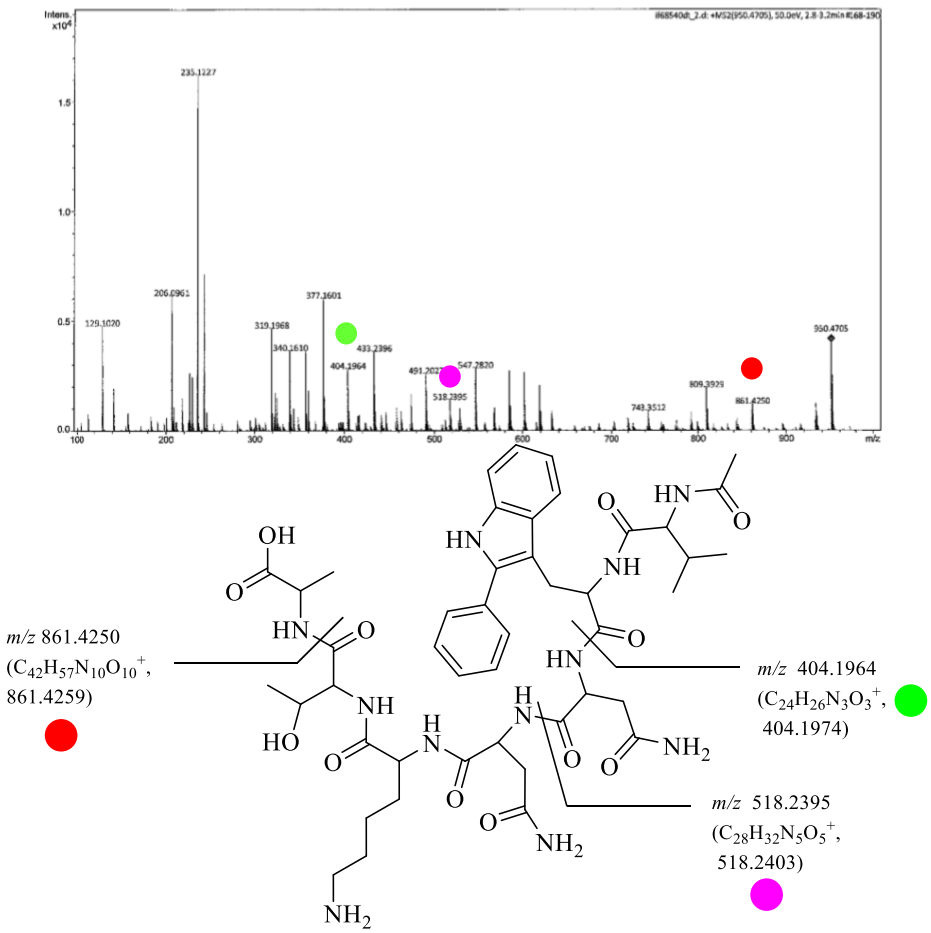
***Ac*-ValTrp(Ph)AsnAsnLysThrAla-O*Me* (13)**



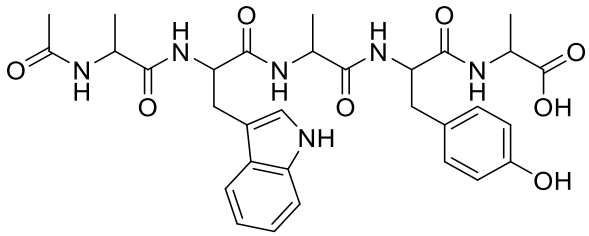
Using general procedure 3, 20 mg of peptide **3** (0.023 mmol, 1 eq.), 5.8 mg benzenediazonium tetrafluoroborate (0.025 mmol, 1.1 eq.) and 0.5 mg $Pd(OAc)_2$ (0.0026 mmol, 20 mol%).

DRT 2 164

The NMR data obtained showed that the C-2 proton was still present in the sample meaning that it did not go to completion or did not work. Through looking at the mass spectra and using MSMS the fragmentation pattern can be seen of the product peak and this shows that the product was indeed made with the pattern shown below.



***Ac*-AlaTrpAlaTyrAla-*OH* (8)**

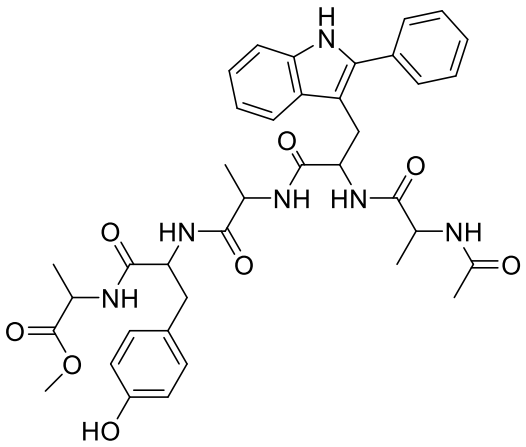


Using general procedure 2, with an acetyl protection at the *N*-terminus (25 mg).

DRT 2 AWAYA

^{1}H NMR (700 MHz, DMSO) δ 10.80 (s, 1H), 9.12 (s, 1H), 8.13 (d, *J* = 7.1 Hz, 1H), 8.01 (d, *J* = 7.0 Hz, 1H), 7.89 (d, *J* = 7.4 Hz, 2H), 7.76 (d, *J* = 8.2 Hz, 1H), 7.55 (d, *J* = 7.9 Hz, 1H), 7.30 (d, *J* = 8.1 Hz, 1H), 7.12 (s, 1H), 7.03 (dd, *J* = 11.5, 8.0 Hz, 3H), 6.92 (t, *J* = 7.4 Hz, 1H), 6.62 (d, *J* = 8.3 Hz, 2H), 4.46 (td, *J* = 8.4, 4.5 Hz, 1H), 4.40 (td, *J* = 8.7, 4.5 Hz, 1H), 4.18 (dp, *J* = 14.1, 7.0 Hz, 3H), 3.12 (dd, *J* = 14.9, 4.0 Hz, 1H), 2.98 – 2.91 (m, 2H), 2.68 (dd, *J* = 14.0, 9.3 Hz, 1H), 1.91 (s, 1H), 1.78 (s, 3H), 1.27 (t, *J* = 9.0 Hz, 4H), 1.11 (t, *J* = 6.9 Hz, 7H). ^{13}C NMR (176 MHz, DMSO) δ 173.93, 172.46, 172.04, 171.79, 171.14, 170.71, 169.27, 155.73, 136.00, 130.17, 127.72, 127.26, 123.56, 120.75, 118.43, 118.14, 114.81, 111.12, 109.81, 53.84, 53.20, 48.45, 48.33, 47.45, 36.50, 27.13, 22.42, 21.05, 18.00, 17.79, 17.15. ESI–HRMS m/z: 623.1851 [M+H]+ ($C_{31}H_{39}N_6O_6^+$ requires 623.2785)

***Ac*-AlaTrp(Ph)AlaTyrAla-O*Me* (12)**



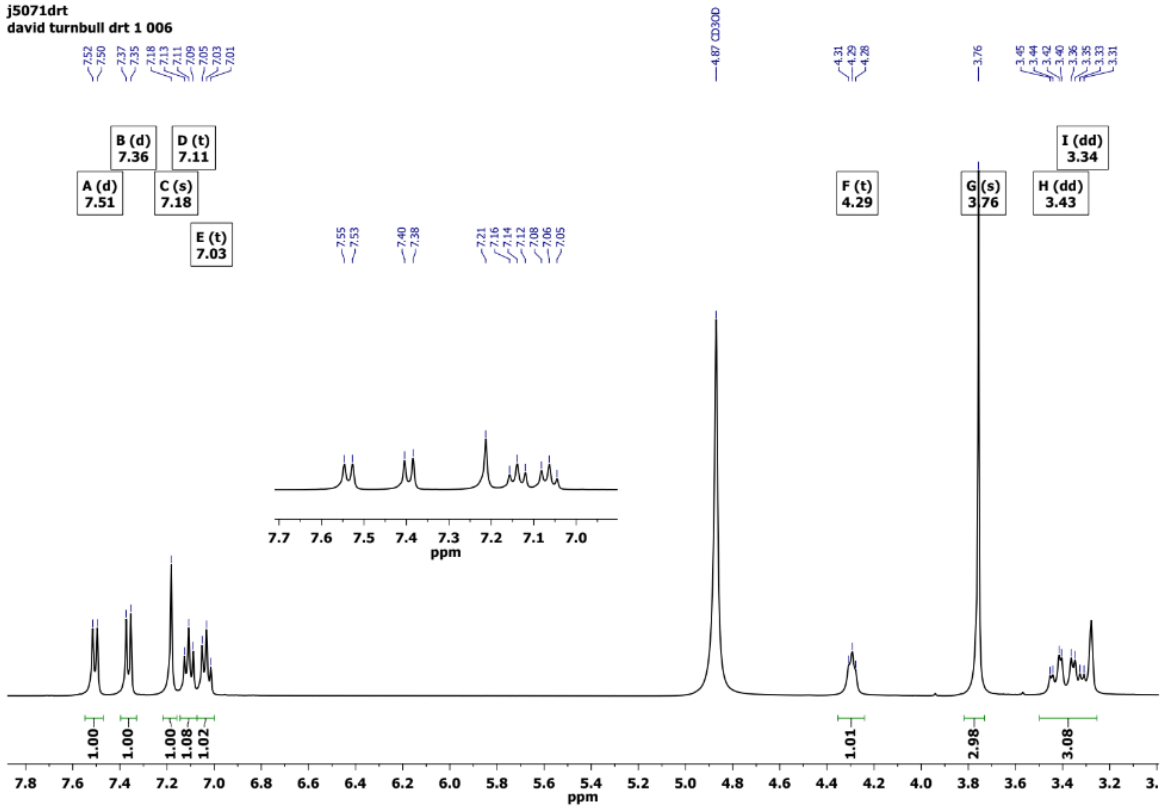
Using general procedure 2, using 15 mg of peptide **4** (0.024 mmol, 1 eq.), 5 mg of benzenediazonium tetrafluoroborate (0.026 mmol, 1.1 eq.) and 1 mg $Pd(OAc)_2$ (0.005 mmol, 20 mol%). (17.2 mg, quant.)

DRT 2 161

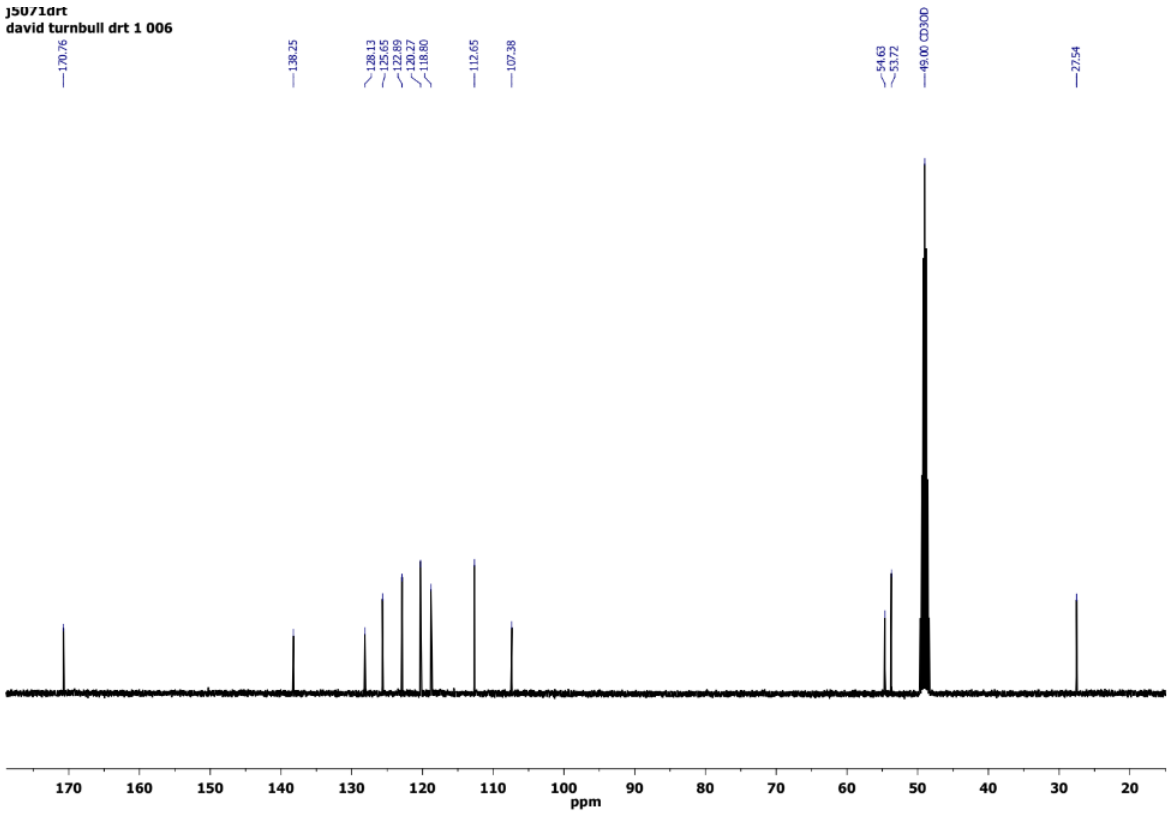
^{1}H NMR (700 MHz, DMSO) δ 11.17 (s, 1H), 8.16 (d, *J* = 6.2 Hz, 2H), 8.00 (d, *J* = 23.1 Hz, 3H), 7.80 (d, *J* = 7.8 Hz, 1H), 7.68 – 7.63 (m, 5H), 7.46 (t, *J* = 7.5 Hz, 4H), 7.36 (t, *J* = 7.3 Hz, 2H), 7.32 (d, *J* = 8.0 Hz, 2H), 7.07 (t, *J* = 7.5 Hz, 2H), 6.99 (d, *J* = 8.0 Hz, 3H), 6.96 (t, *J* = 7.4 Hz, 2H), 6.62 (d, *J* = 7.9 Hz, 4H), 4.52 (dd, *J* = 13.9, 7.3 Hz, 2H), 4.33 (dd, *J* = 13.2, 8.6 Hz, 2H), 4.24 – 4.18 (m, 2H), 4.07 (dd, *J* = 14.5, 7.1 Hz, 4H), 3.59 (s, 5H), 2.90 (dd, *J* = 14.0, 4.3 Hz, 2H), 2.65 (dd, *J* = 13.9, 9.4 Hz, 2H), 2.58 (d, *J* = 15.5 Hz, 1H), 1.76 (s, 5H), 1.24 (d, *J* = 7.2 Hz, 8H), 1.10 – 1.06 (m, 7H), 1.01 (q, *J* = 7.5 Hz, 7H). ESI–HRMS m/z: 735.3379 [M+Na]+ ($C_{38}H_{44}N_6O_8Na^+$ requires 735.3074)

Chapter 6: Appendices

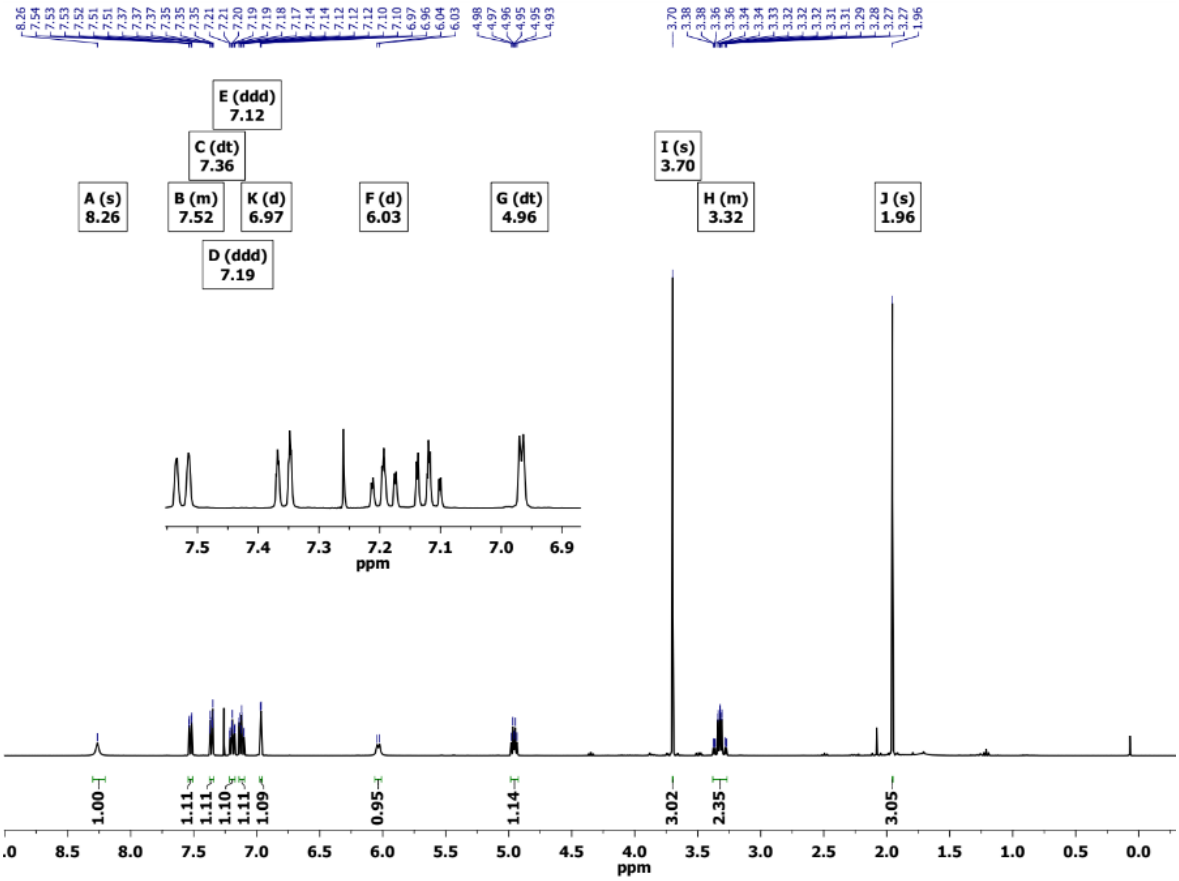
NMR Spectra



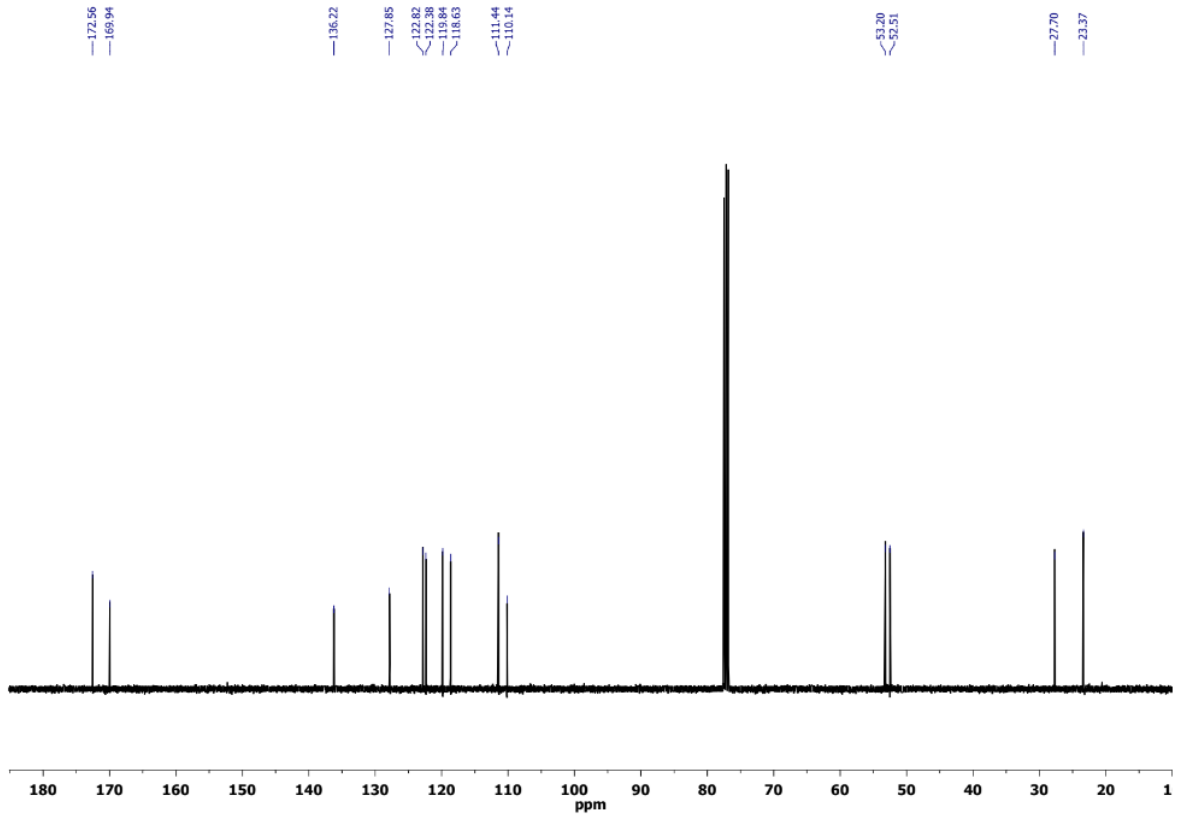
Appendix 1: ^{1}H NMR of compound 1



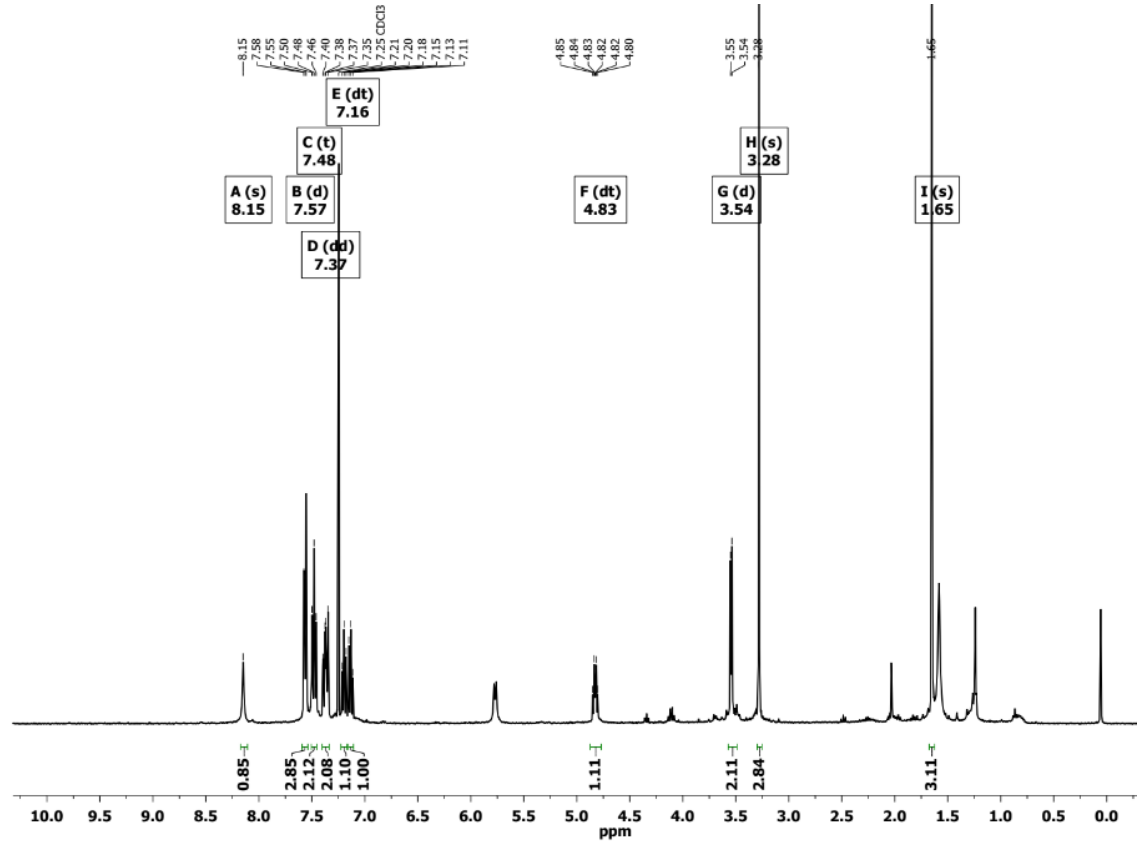
Appendix 2: ^{13}C NMR of compound 1



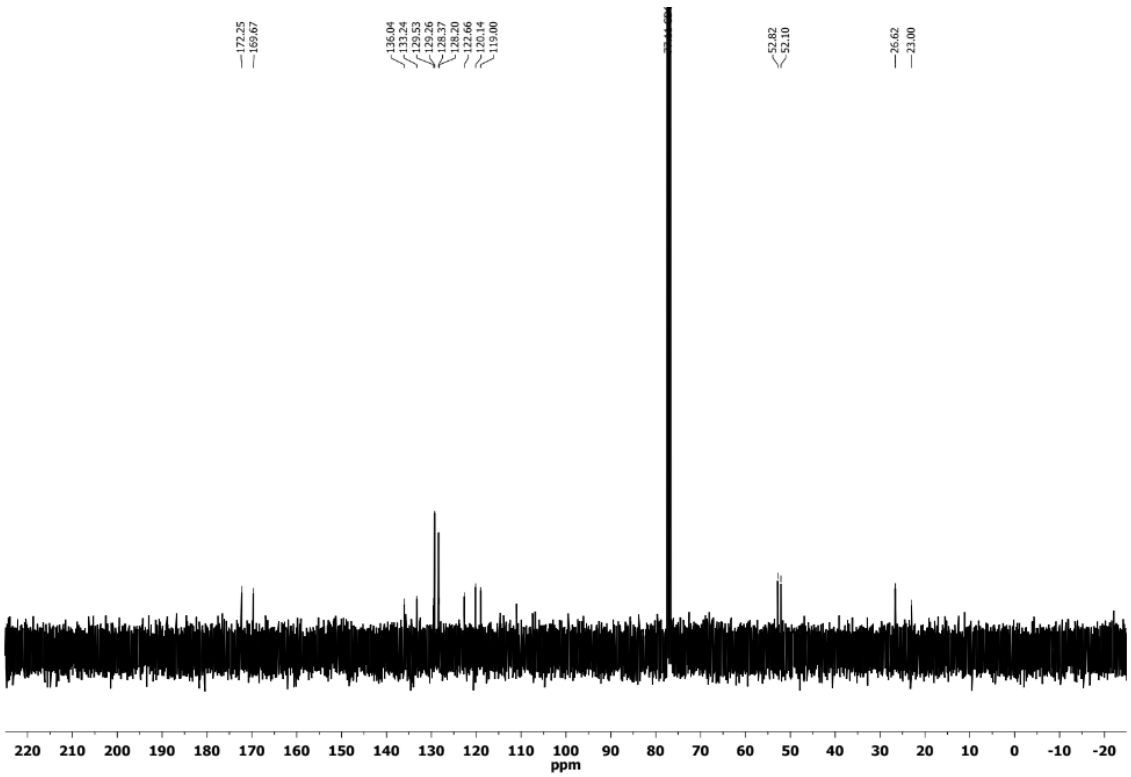
Appendix 3: ^{1}H NMR of compound 2



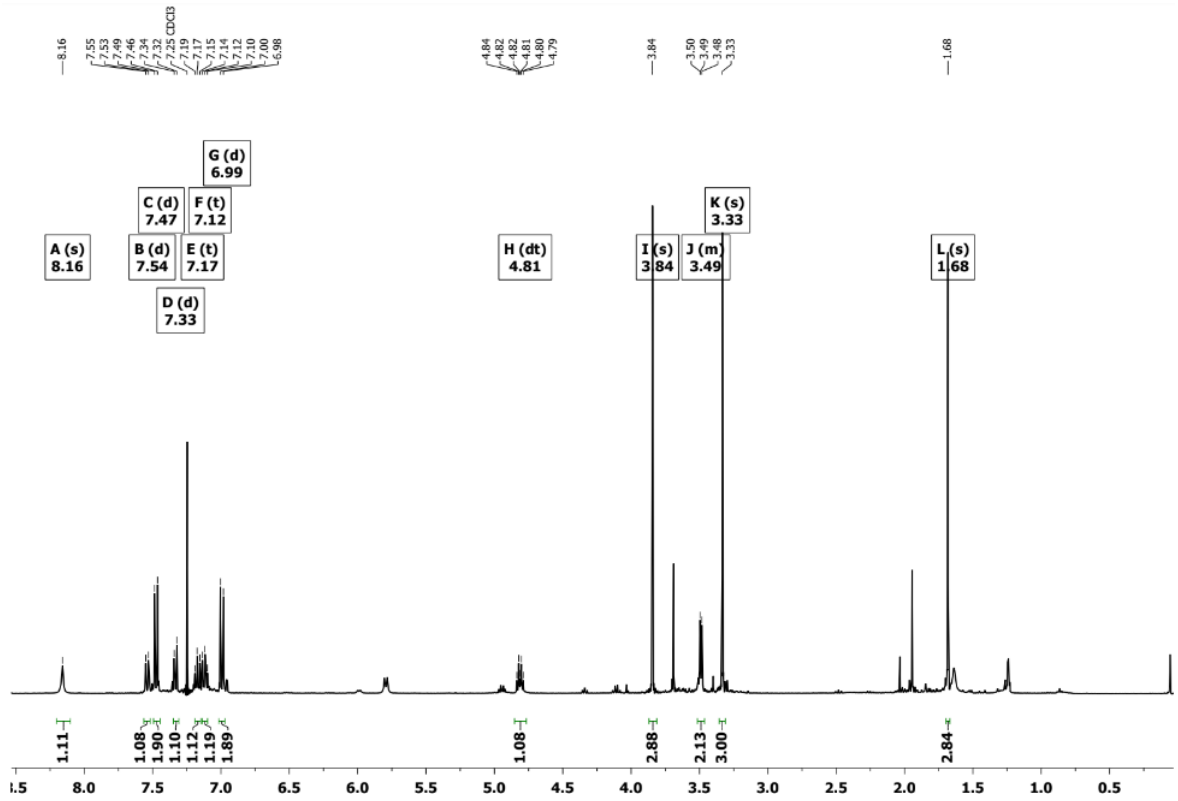
Appendix 4: 13C NMR of compound 2



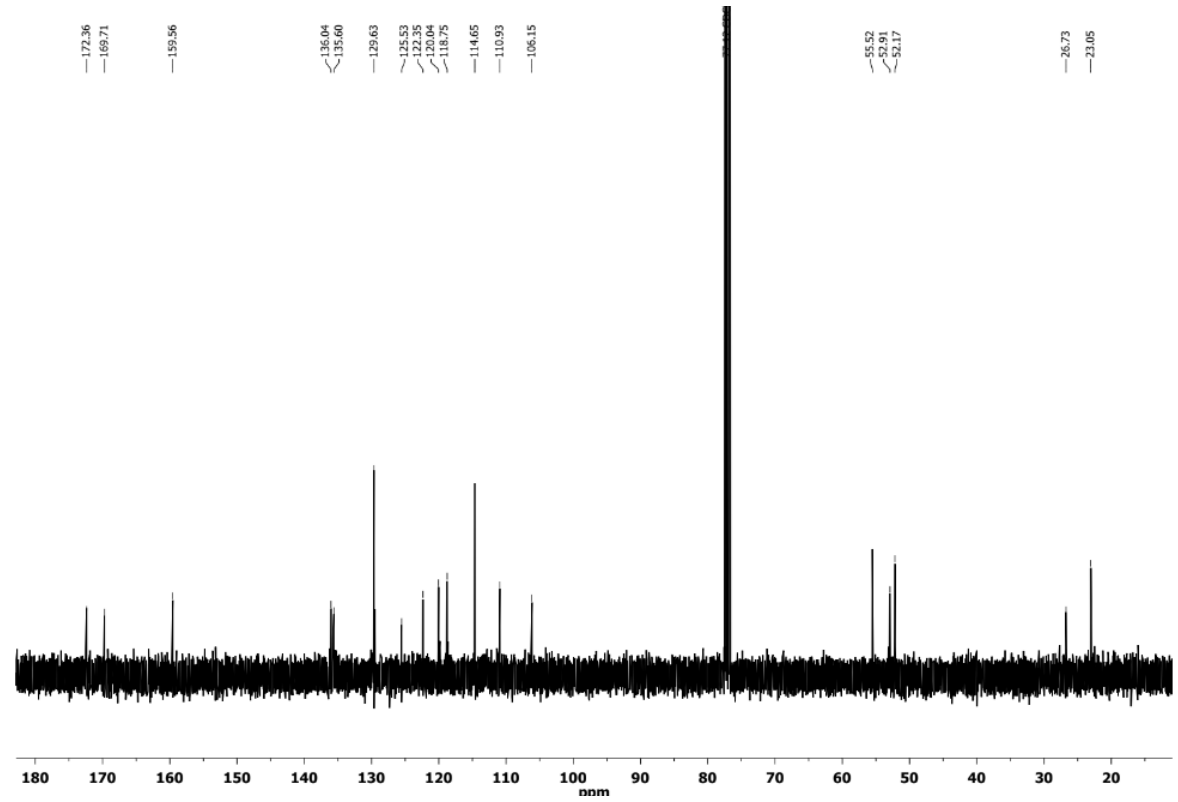
Appendix 5: ^{1}H spectrum of compound 3



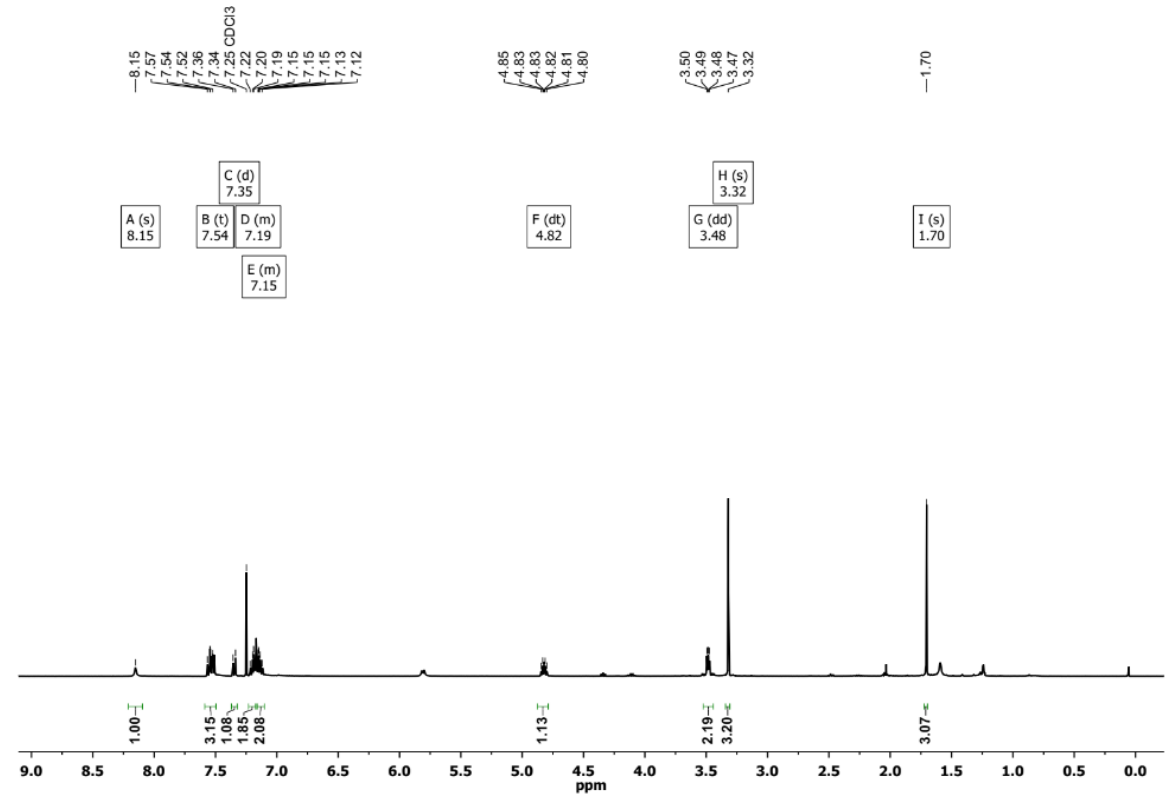
Appendix 6: ^{13}C spectrum of compound 3



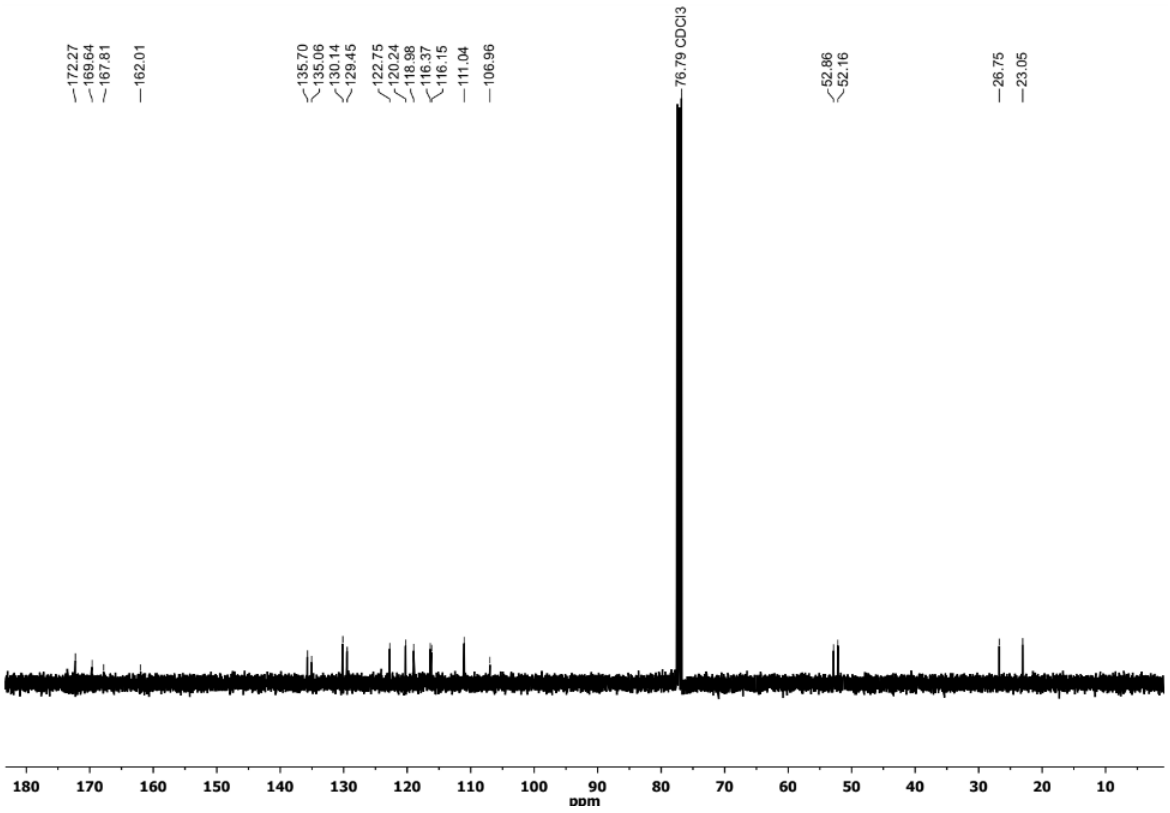
Appendix 7: ^{1}H spectrum of compound 4



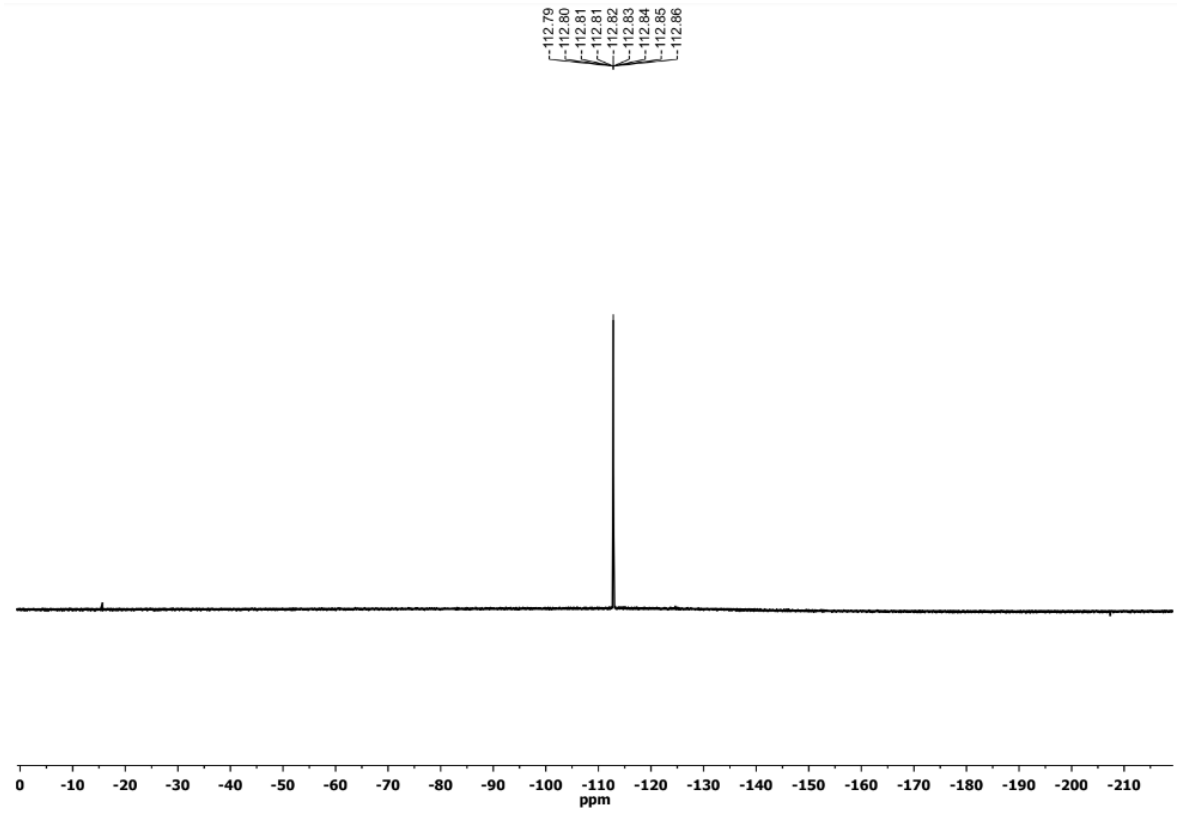
Appendix 8: ^{13}C spectrum of compound 4



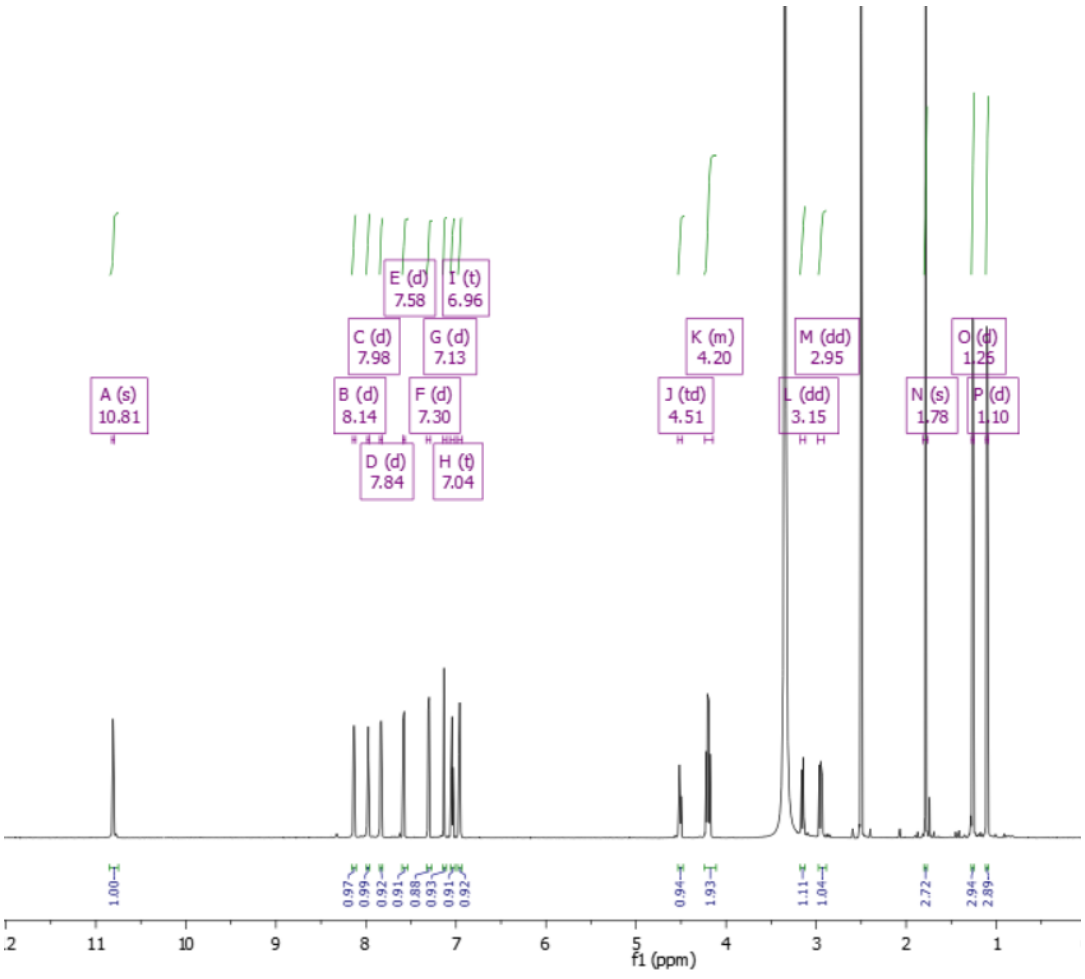
Appendix 9: ^{1}H spectrum of compound 5



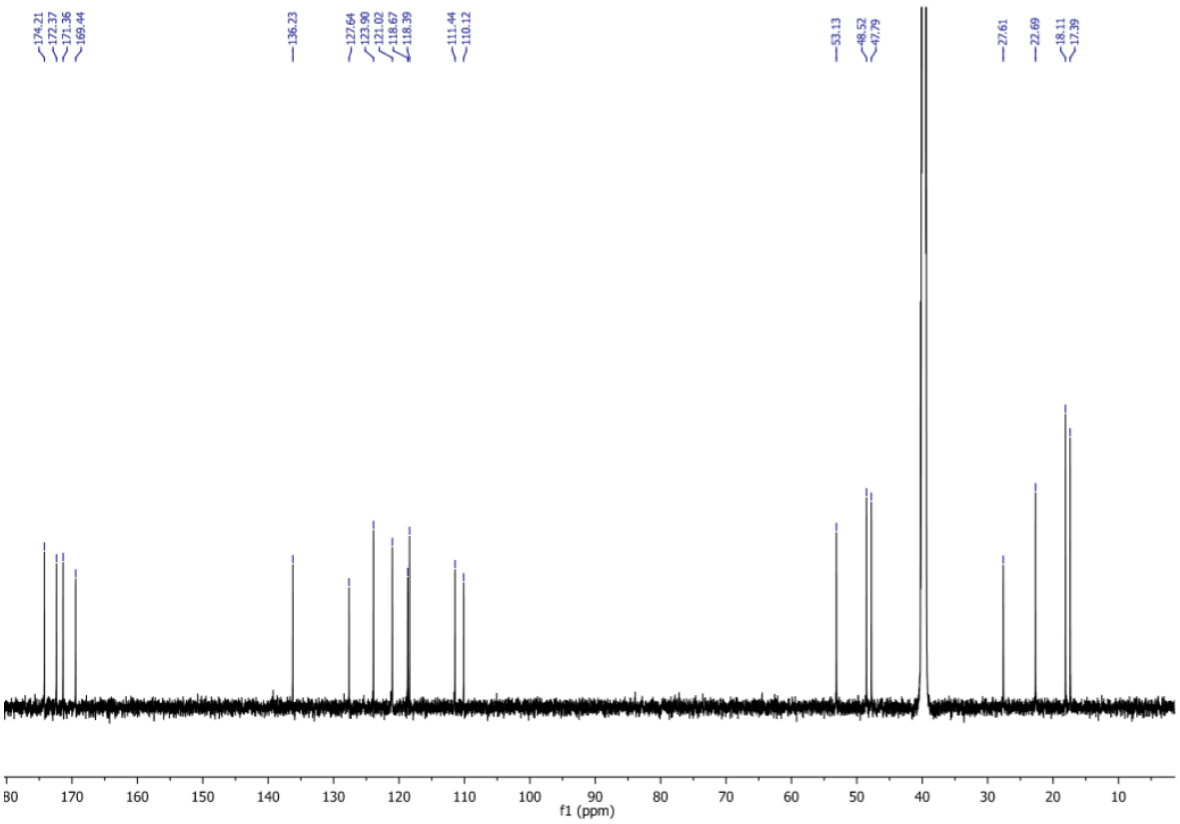
Appendix 10: ^{13}C spectrum of compound 5



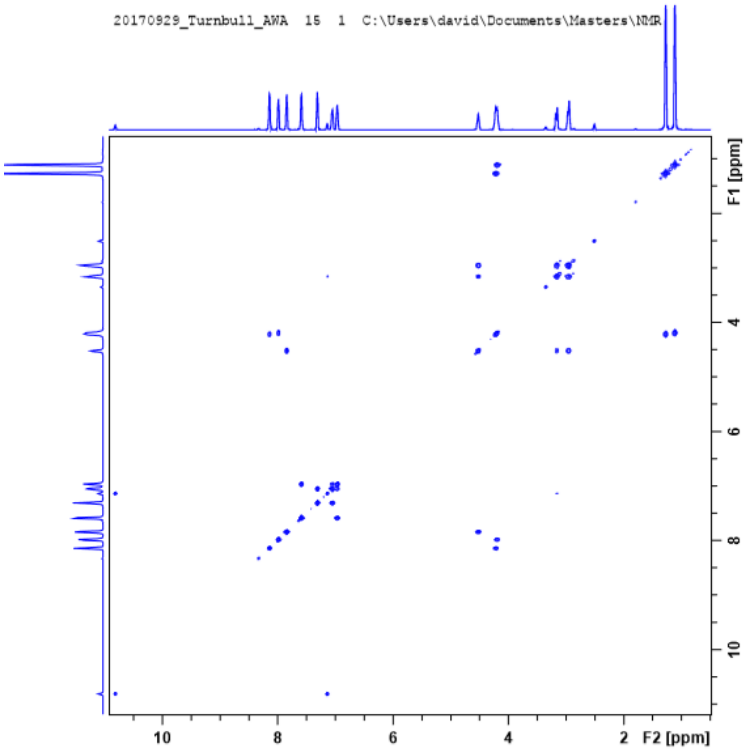
Appendix 11: ^{19}F spectrum of compound 5



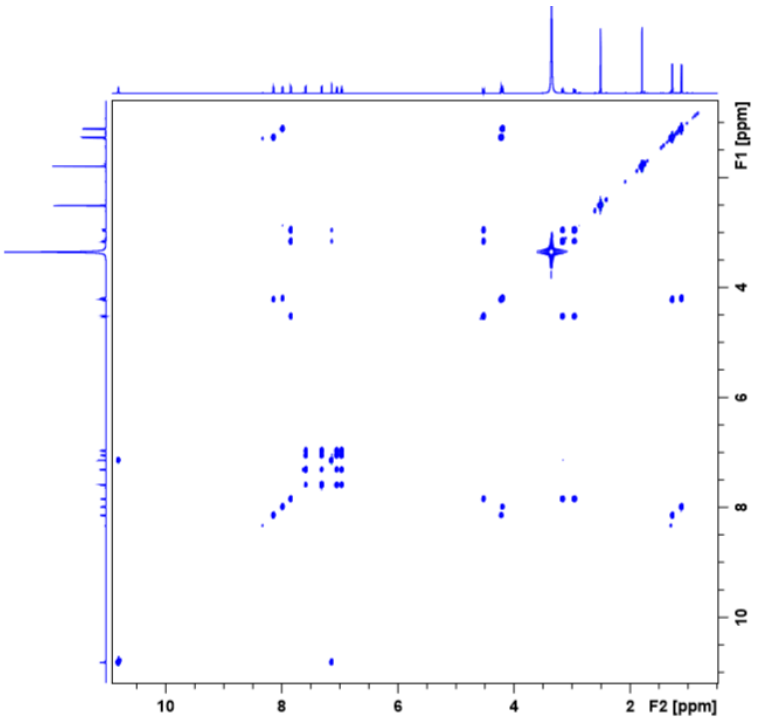
Appendix 12: ^{1}H spectrum of compound 6



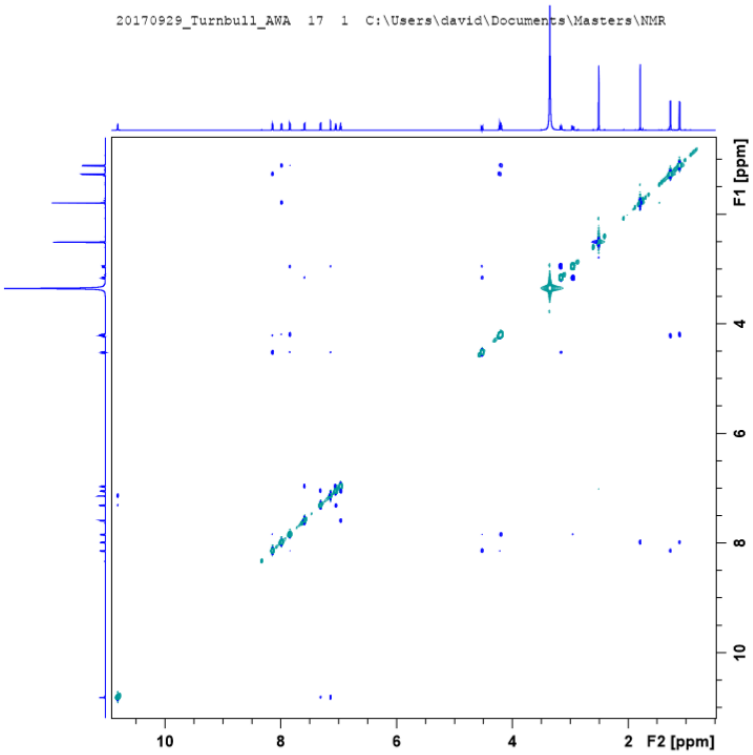
Appendix 13: ^{13}C spectrum of compound 6



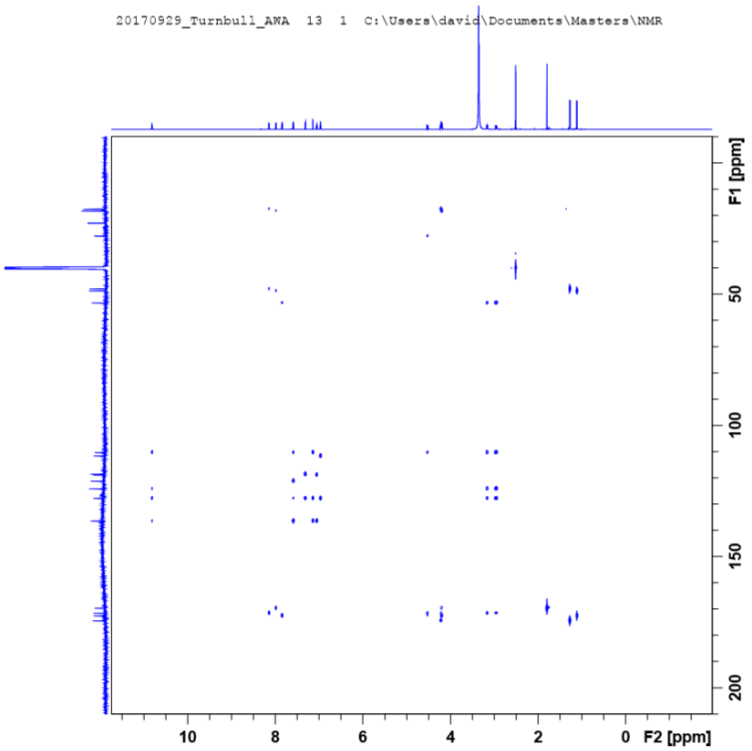
Appendix 14: COSY spectrum of compound 6



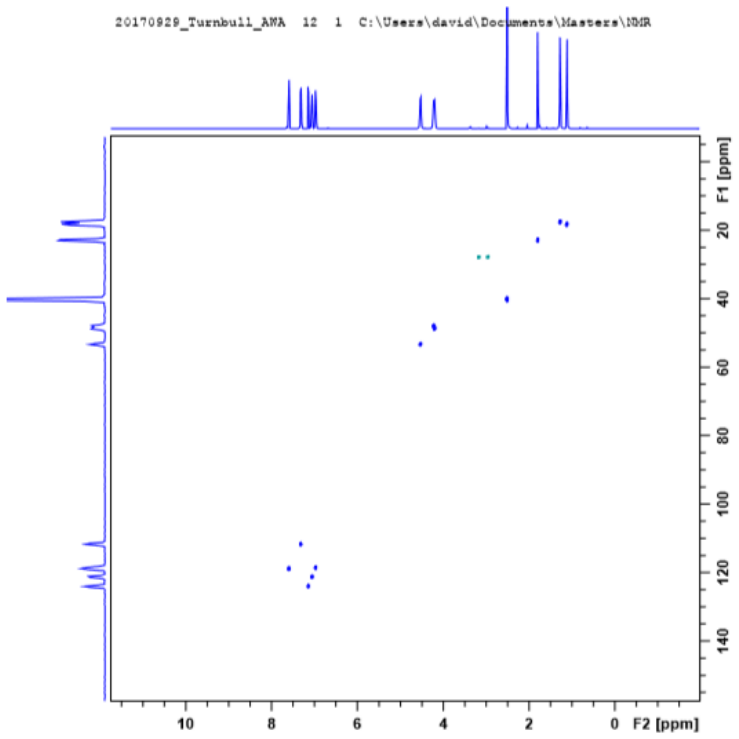
Appendix 15: TOCSY spectrum of compound 6



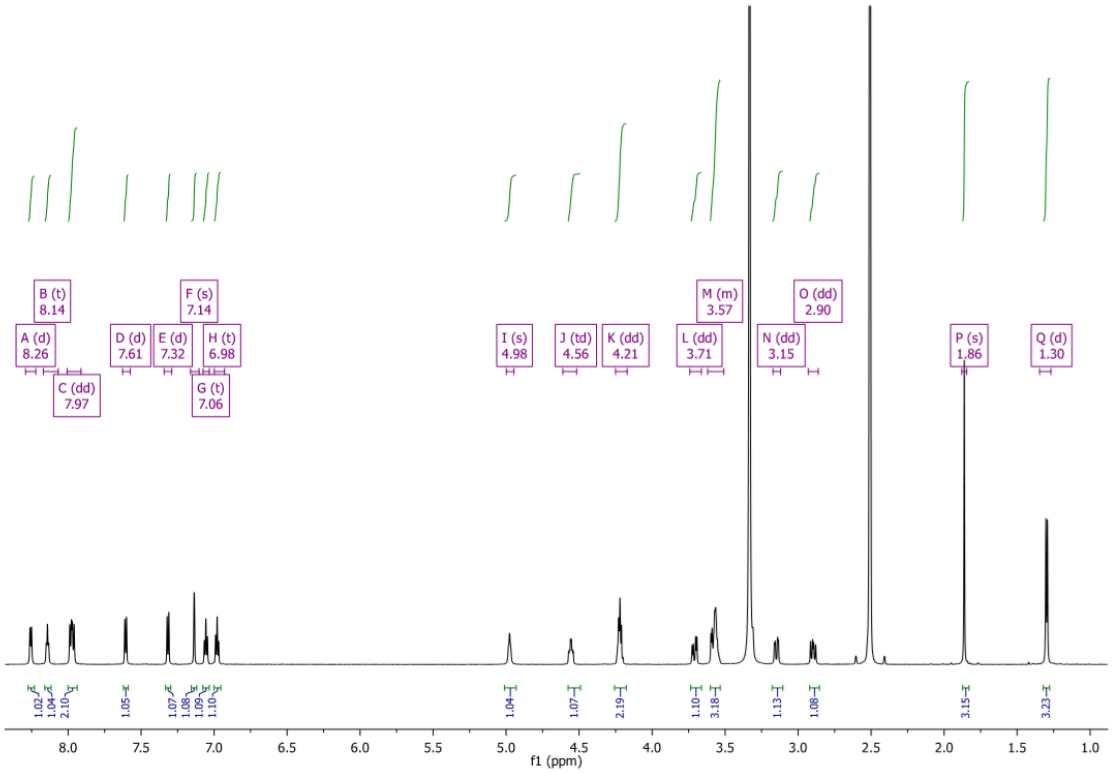
Appendix 16: ROESY spectrum of compound 6



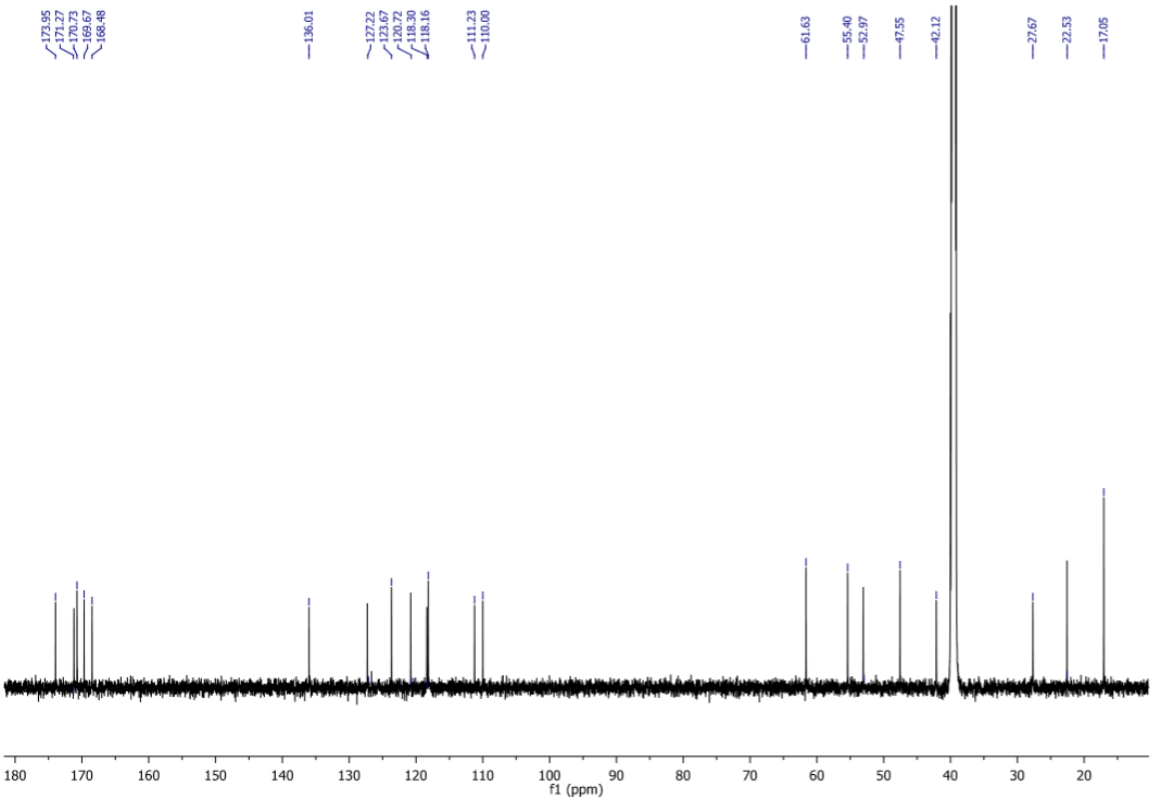
Appendix 17: HMBC spectrum of compound 6



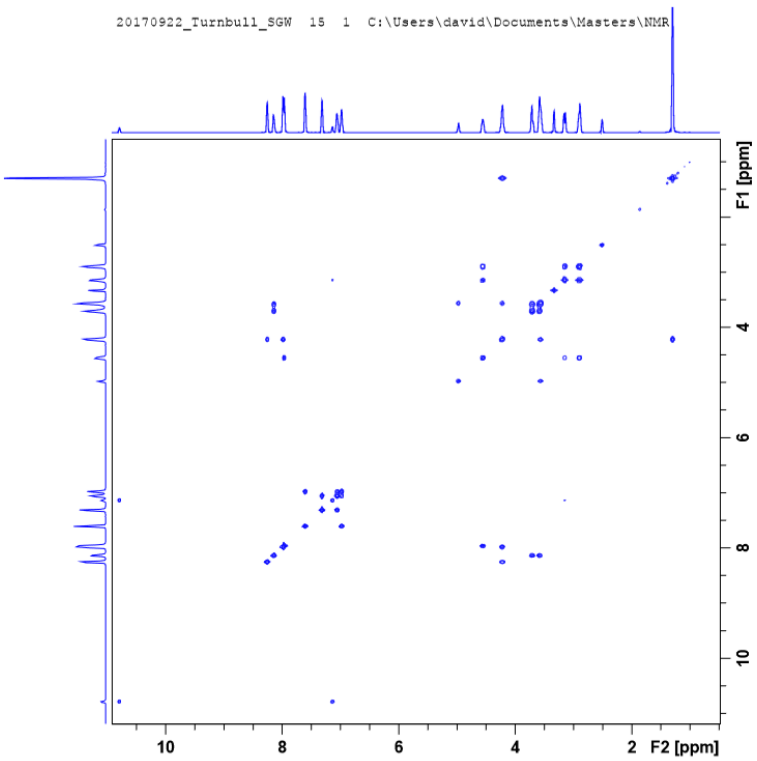
Appendix 18: HSQC spectrum of compound 6



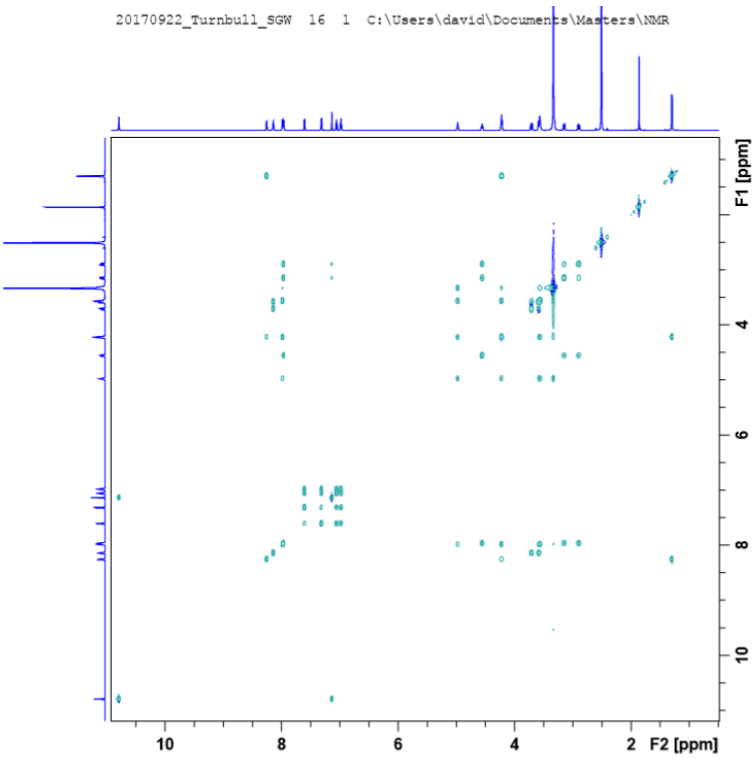
Appendix 19: ^{1}H spectrum of compound 7



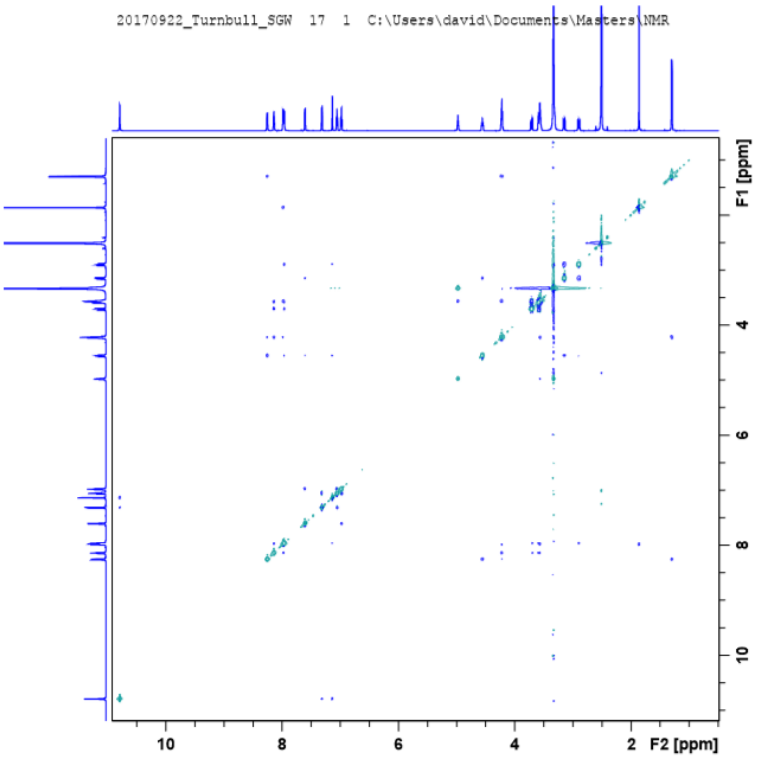
Appendix 20: ^{13}C spectrum of compound 7



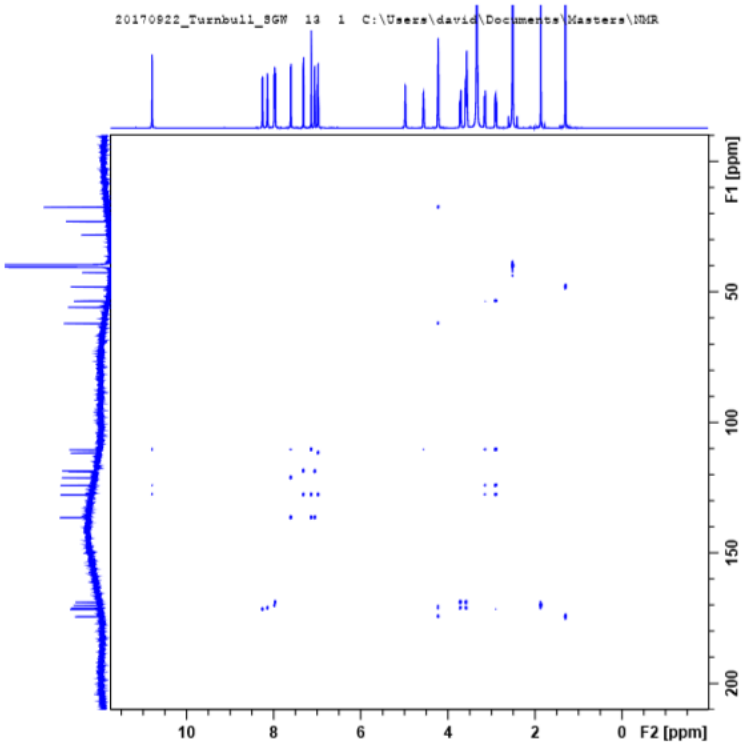
Appendix 21: COSY spectrum of compound 7



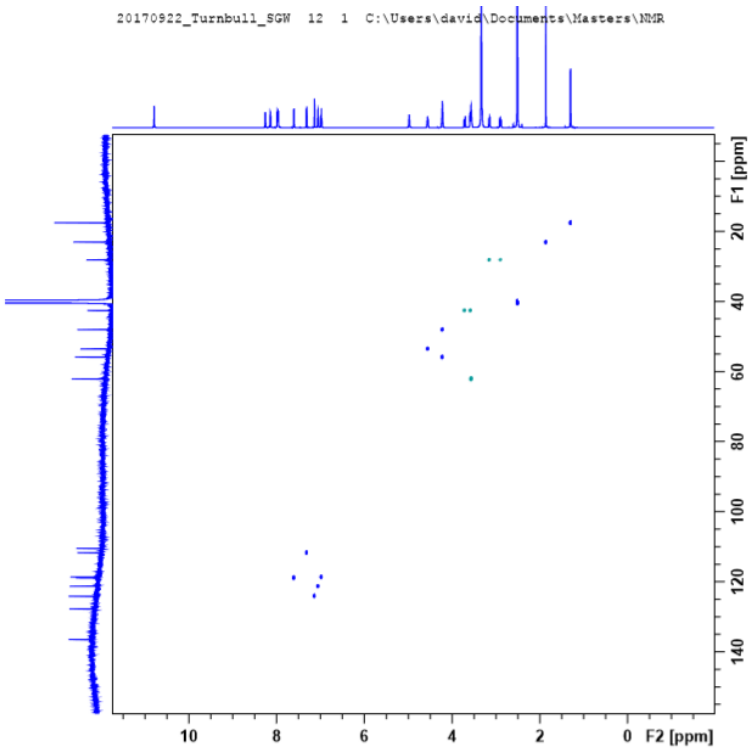
Appendix 22: TOCSY spectrum of compound 7



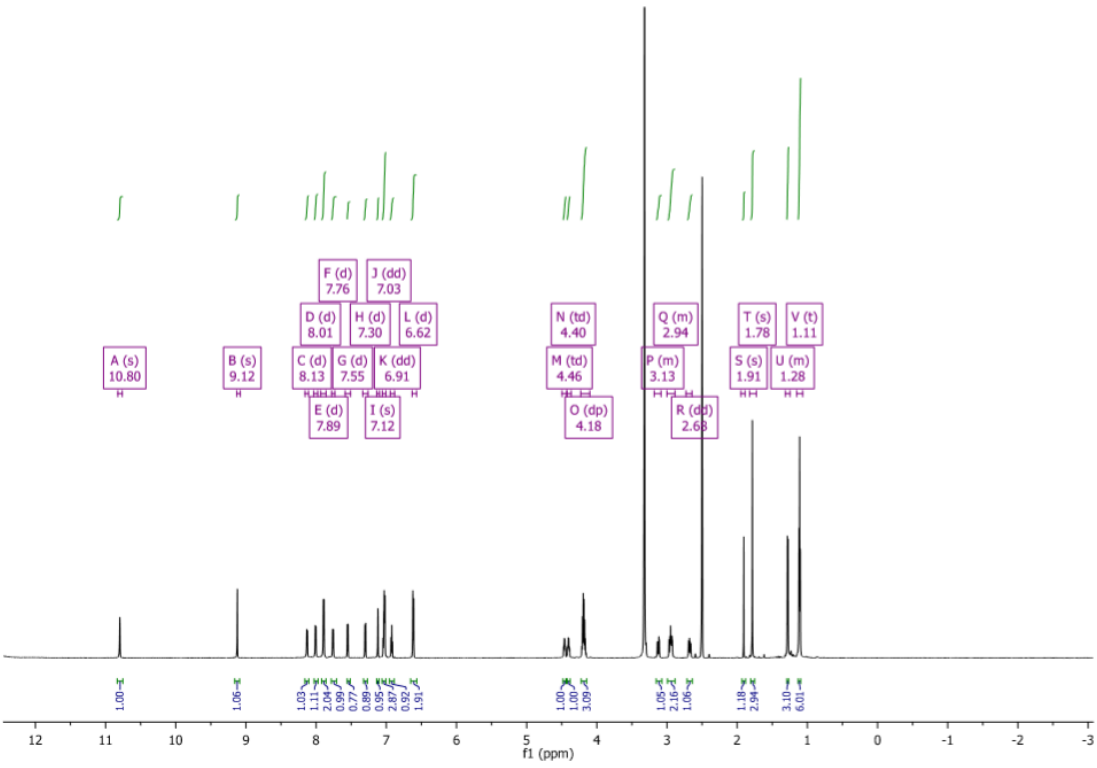
Appendix 23: ROESY spectrum of compound 7



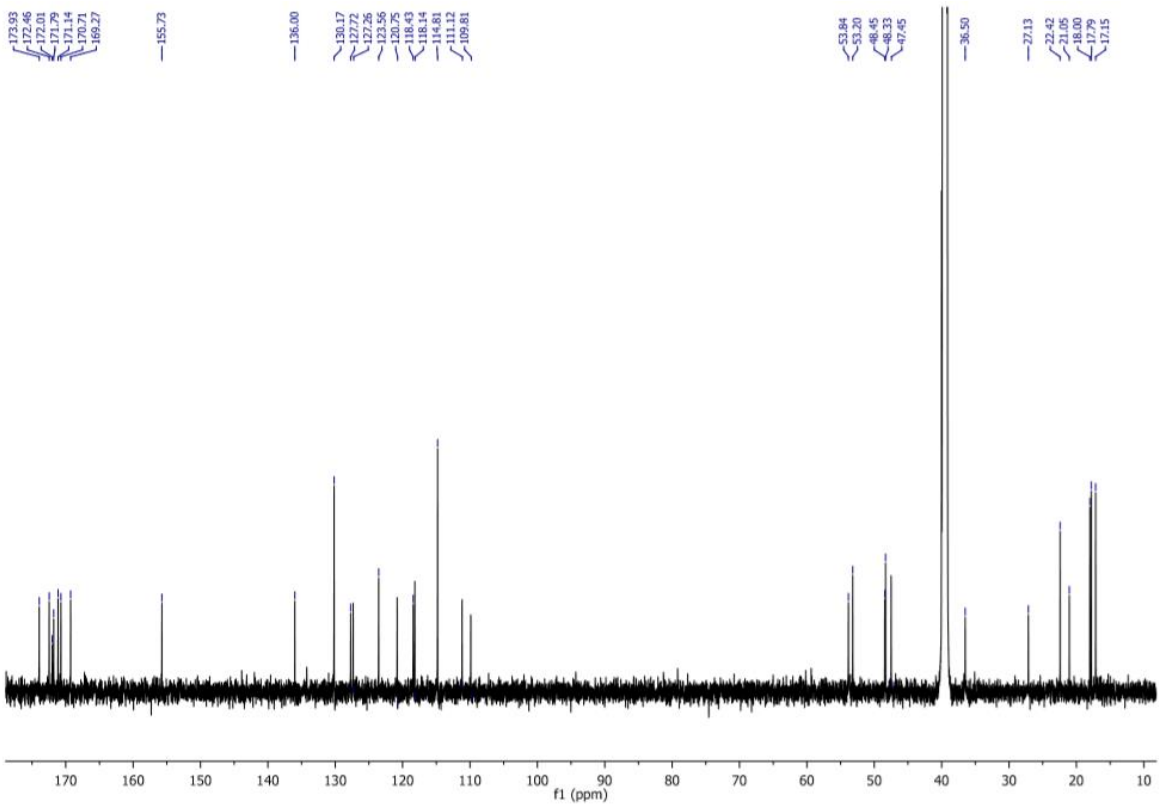
Appendix 24: HMBC spectrum of compound 7



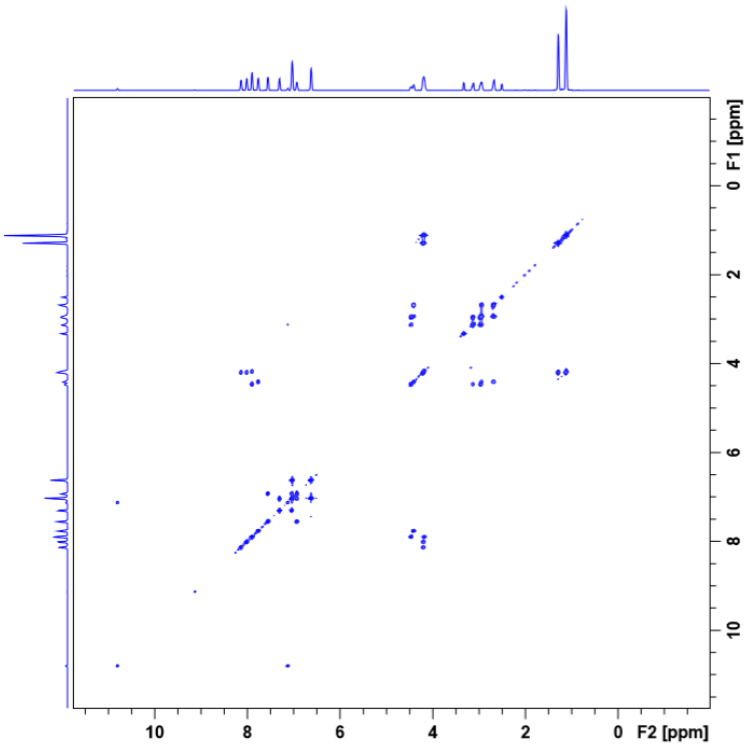
Appendix 25: HSQC spectrum of compound 7



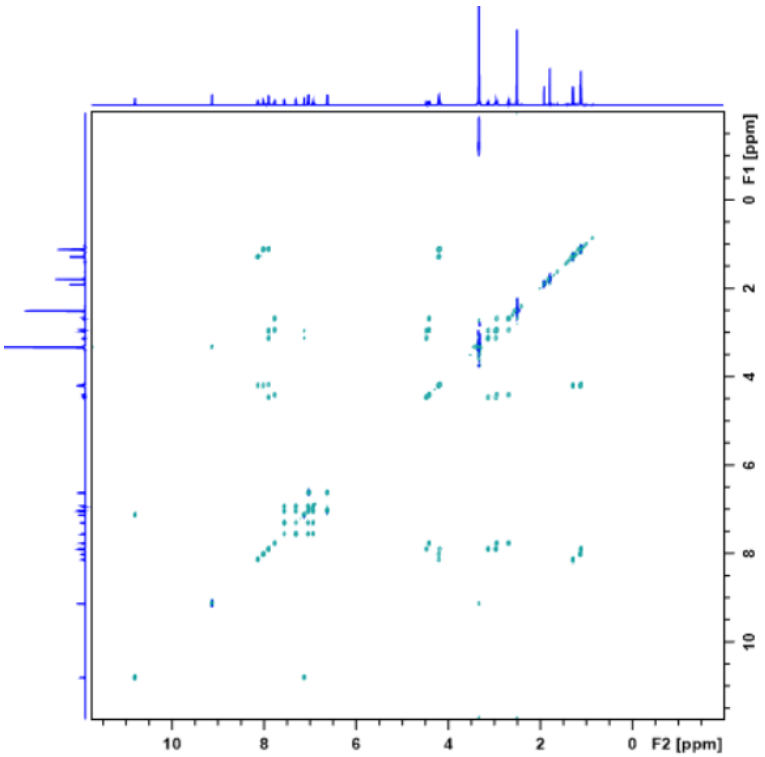
Appendix 26: ^{1}H spectrum of compound 8



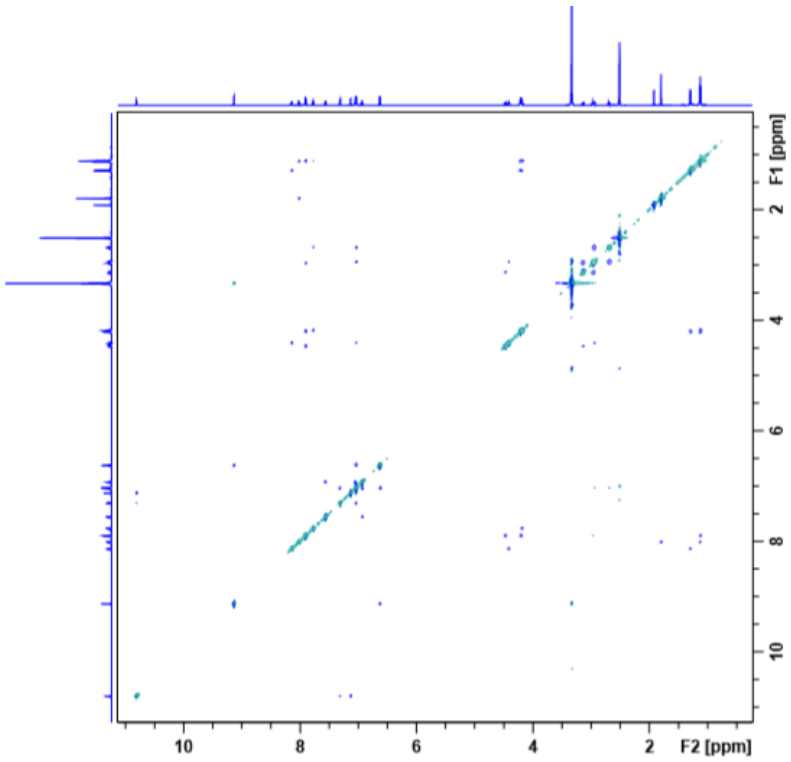
Appendix 27: ^{13}C spectrum of compound 8



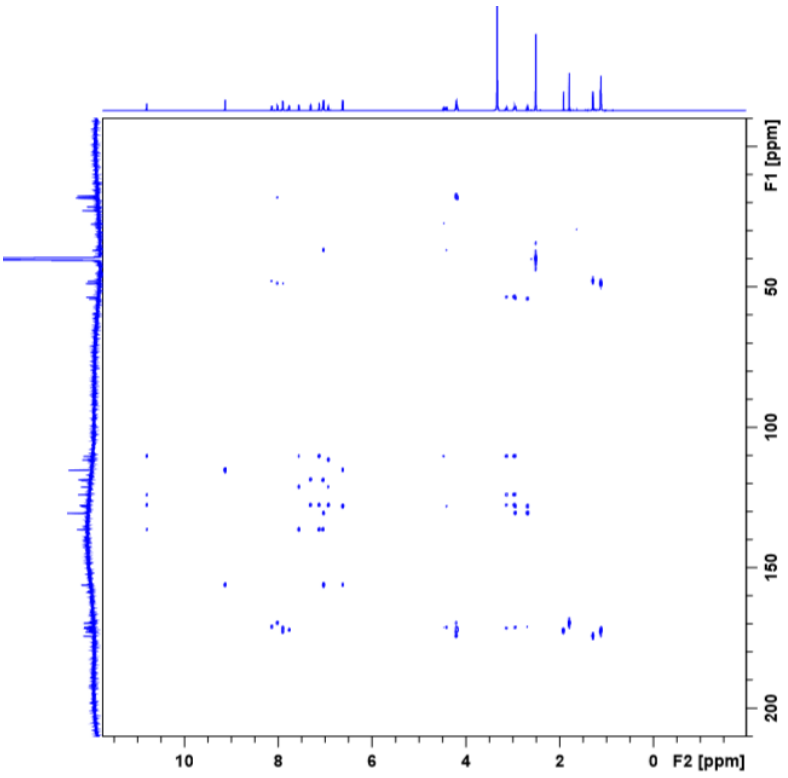
Appendix 28: COSY spectrum of compound 8



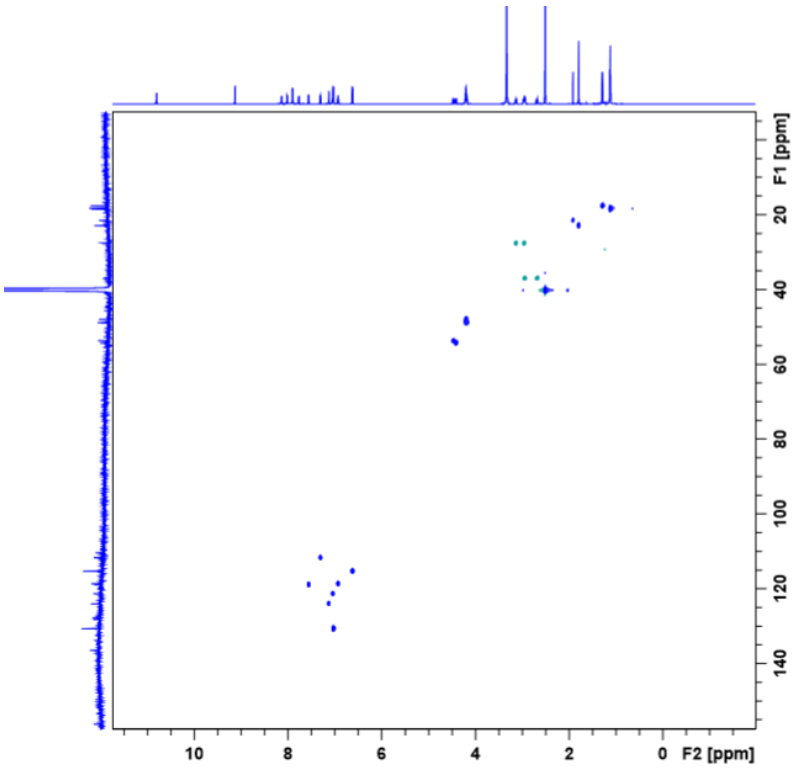
Appendix 29: TOCSY spectrum of compound 8



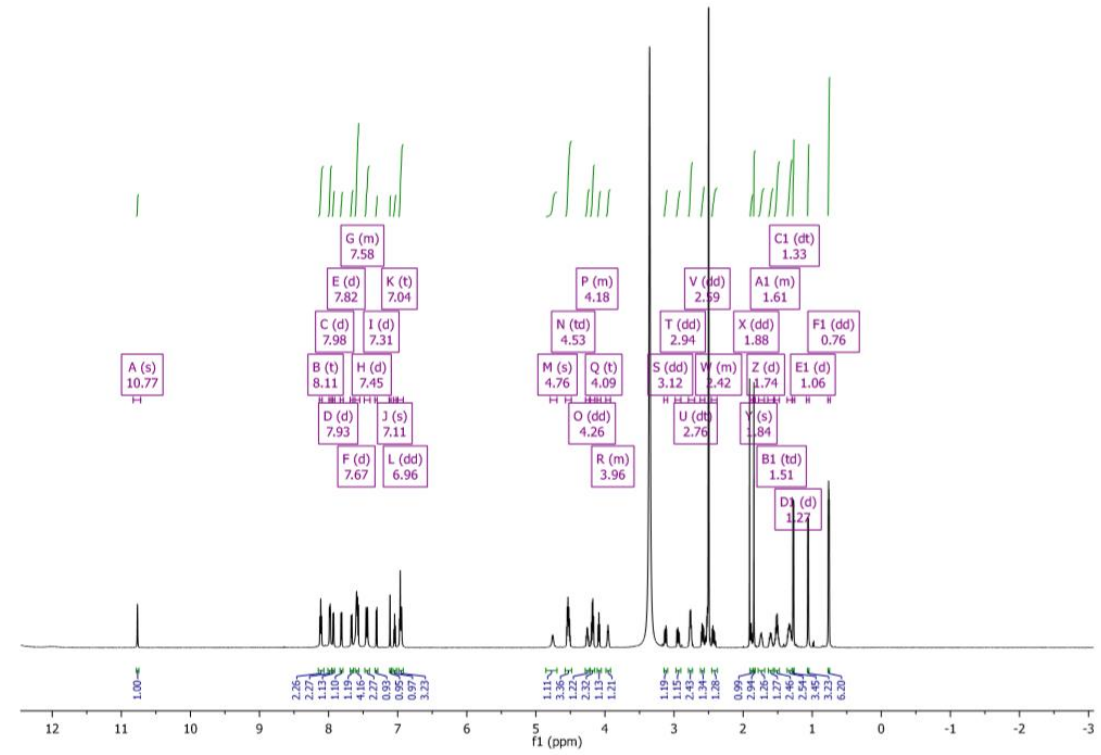
Appendix 30: ROESY spectrum of compound 8



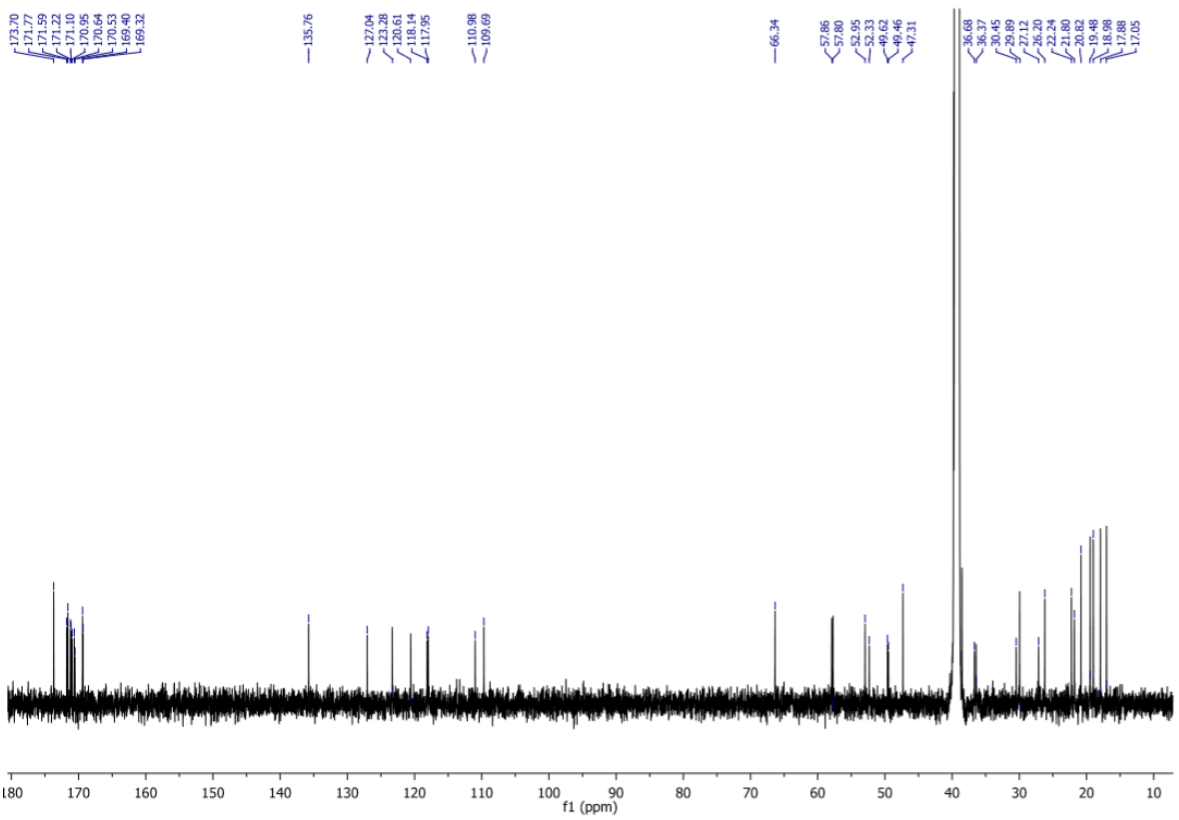
Appendix 31: HMBC spectrum of compound 8



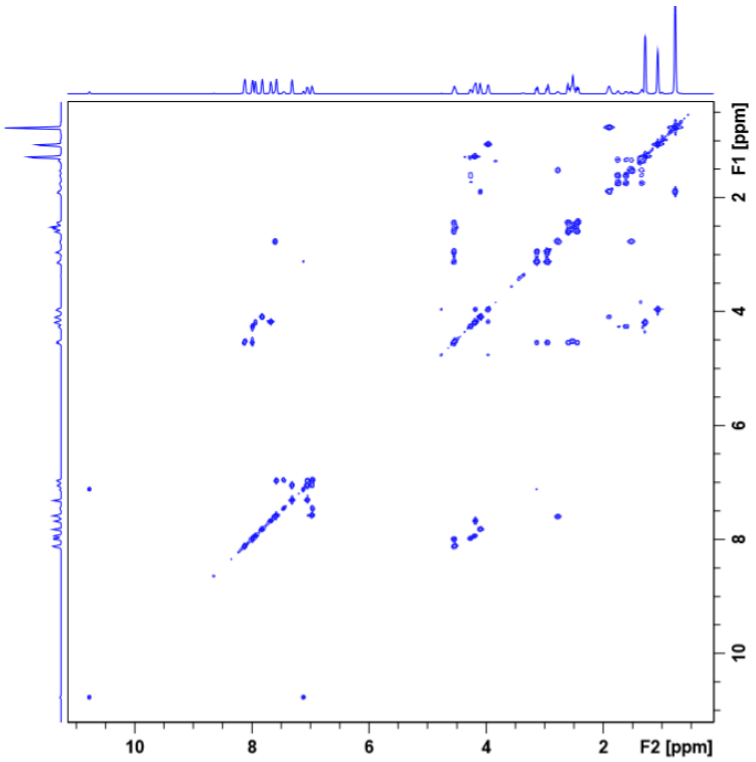
Appendix 32: HSQC spectrum of compound 8



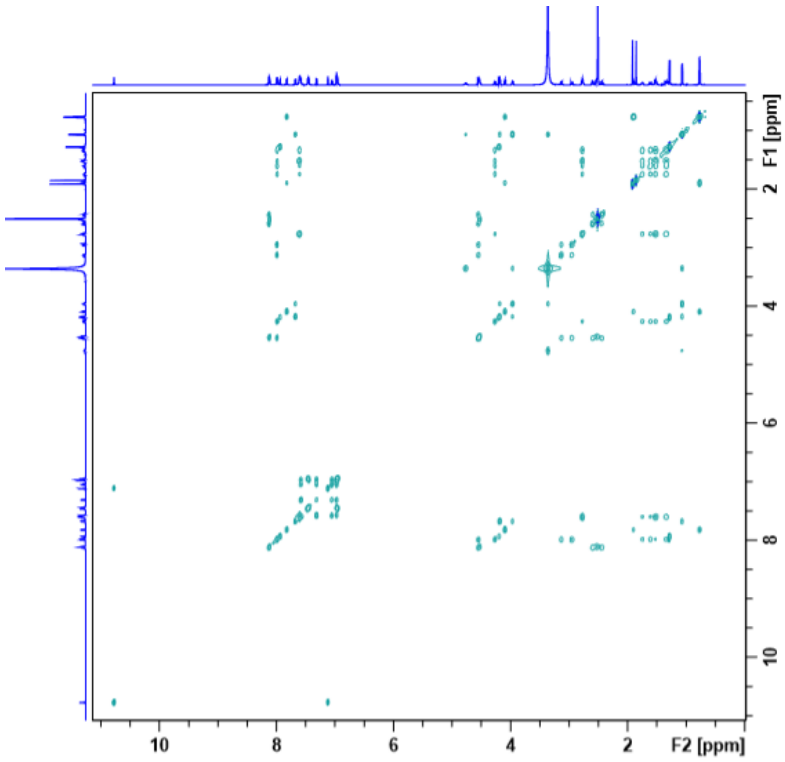
Appendix 33: ^{1}H spectrum of compound 9



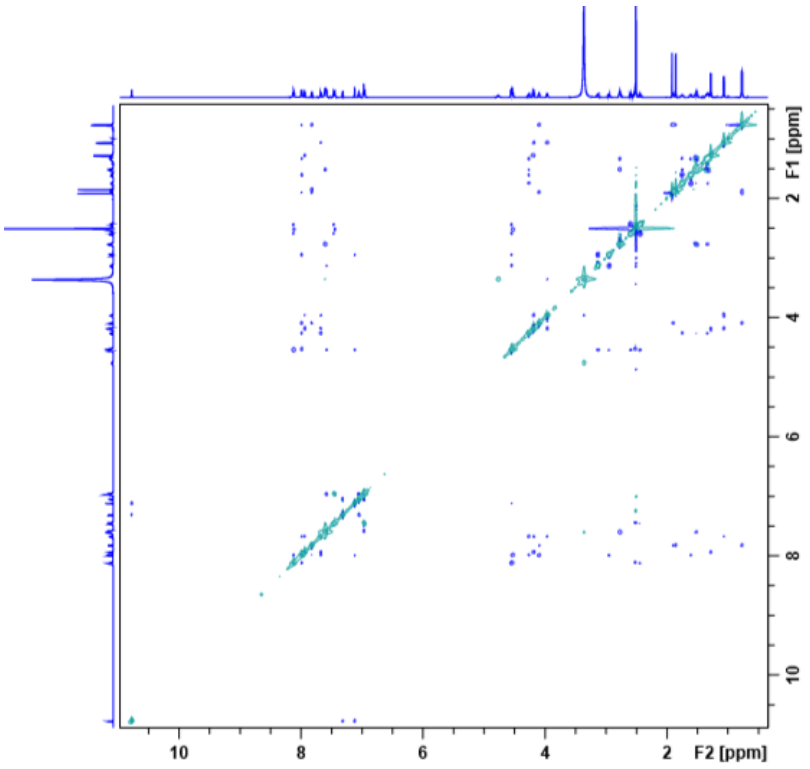
Appendix 34: ^{13}C spectrum of compound 9



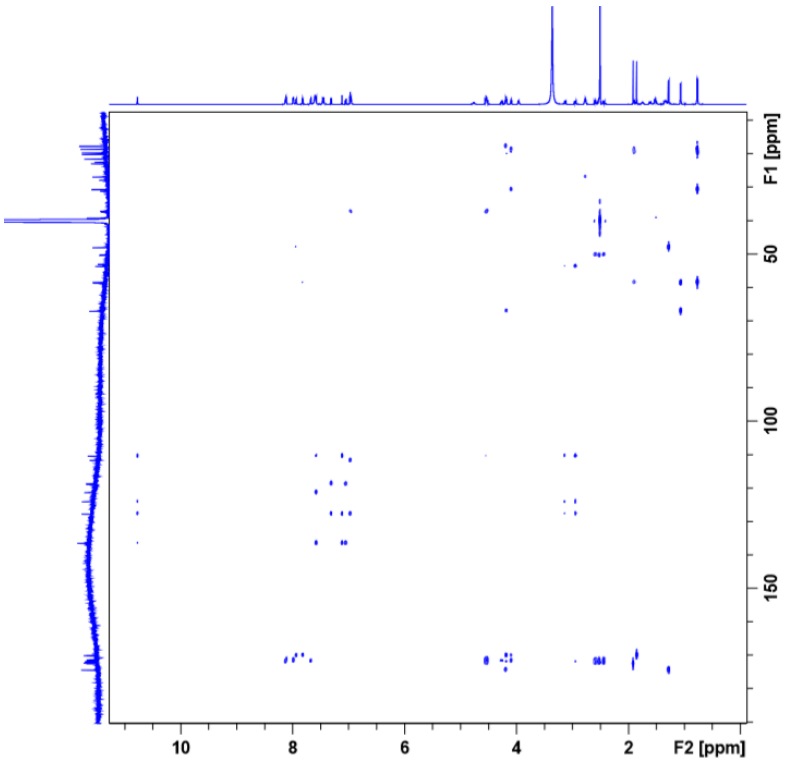
Appendix 35: COSY spectrum of compound 9



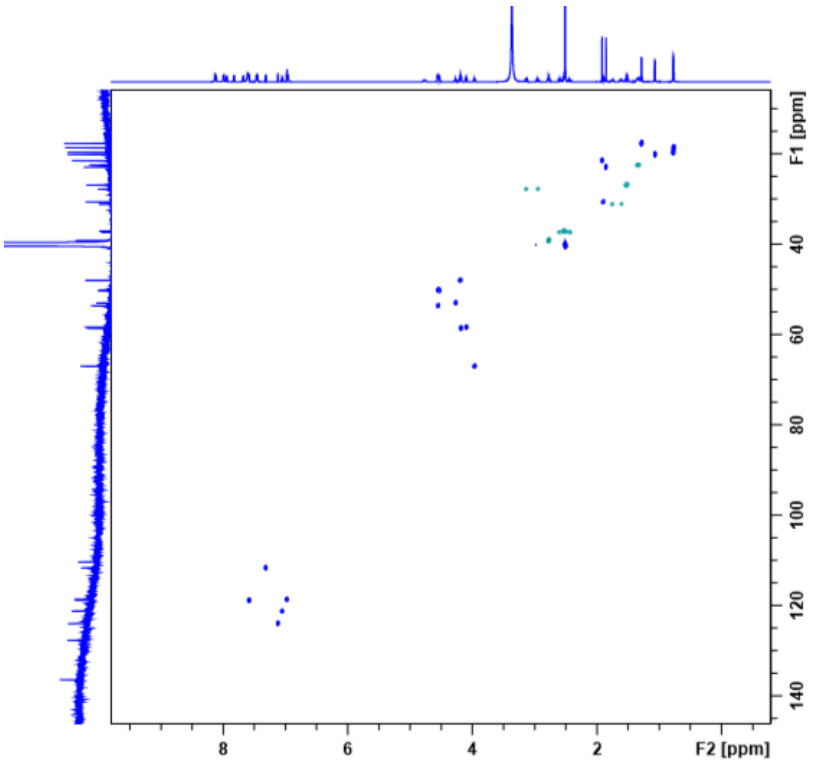
Appendix 36: TOCSY spectrum of compound 9



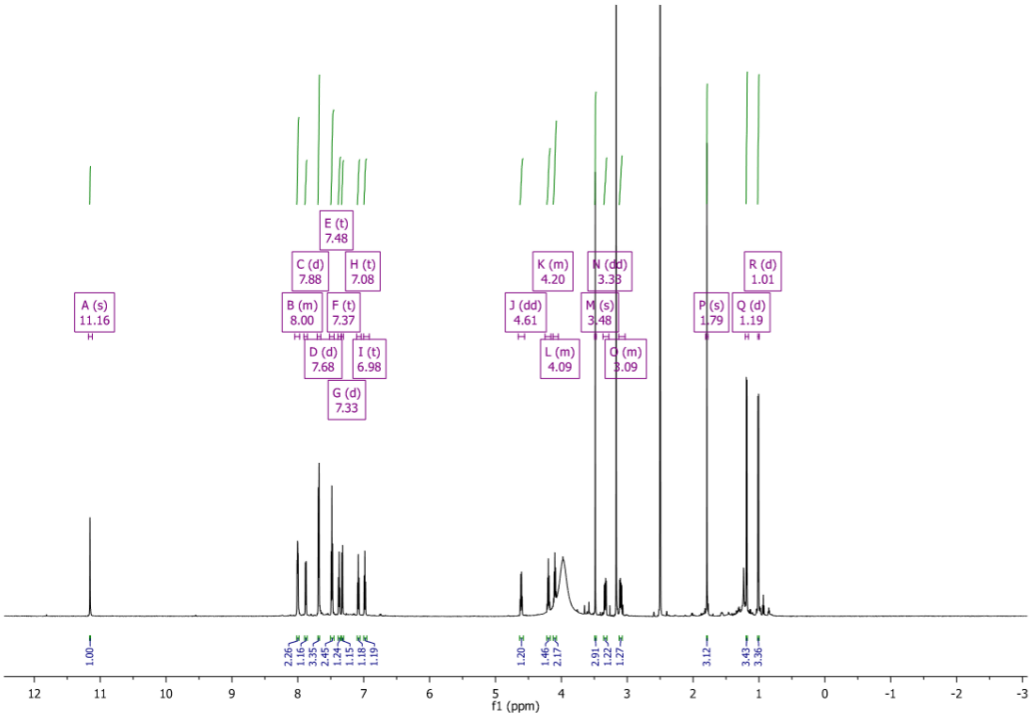
Appendix 37: ROESY spectrum of compound 9



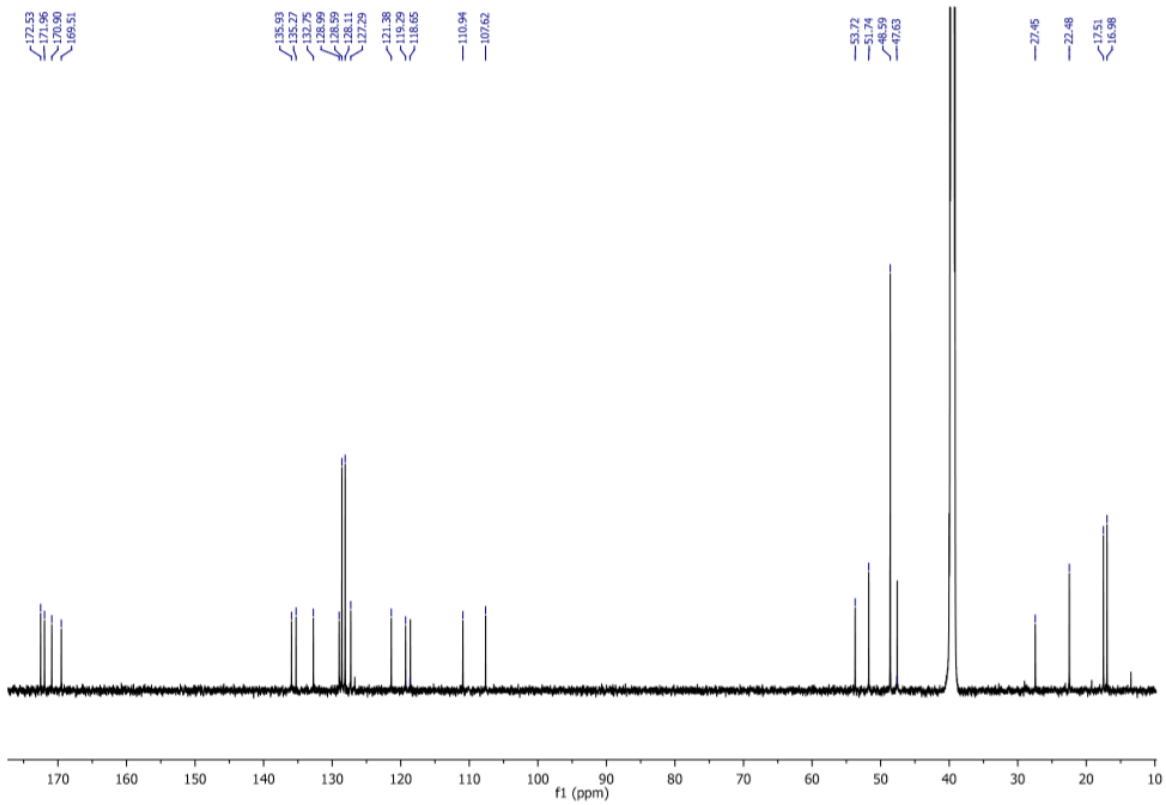
Appendix 38: HMBC spectrum of compound 9



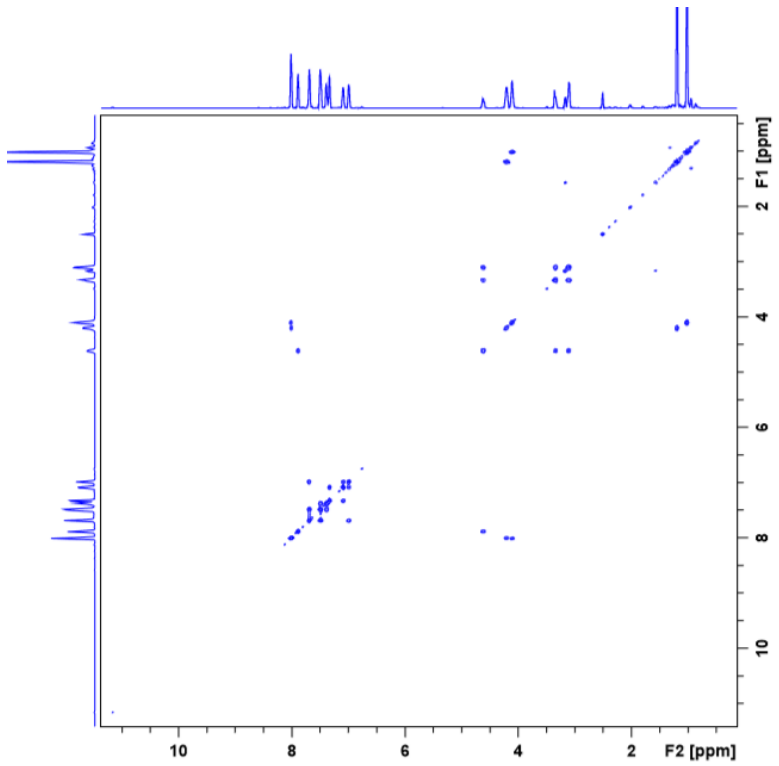
Appendix 39: HSQC spectrum of compound 9



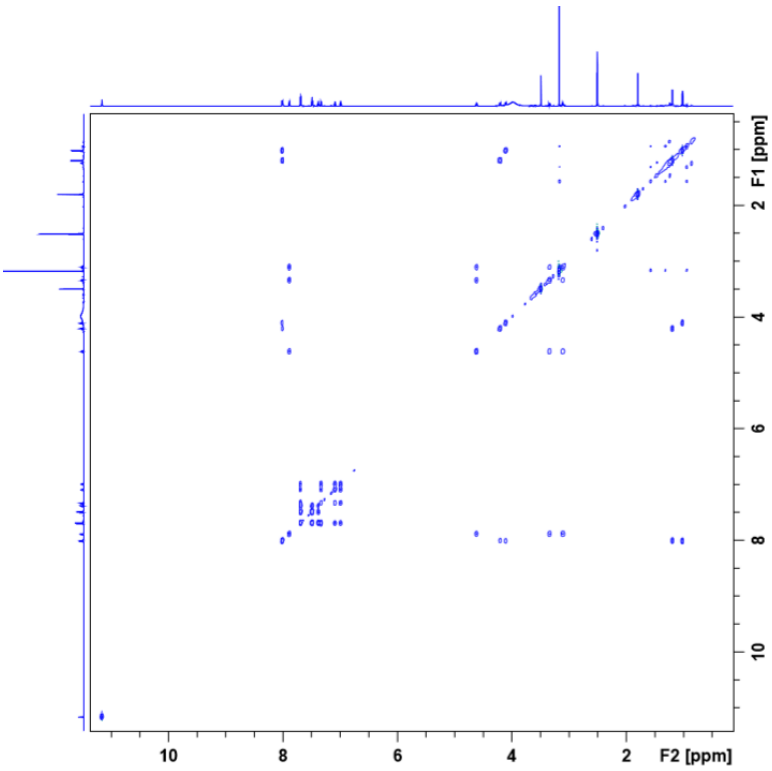
Appendix 40: ^{1}H spectrum of compound 11



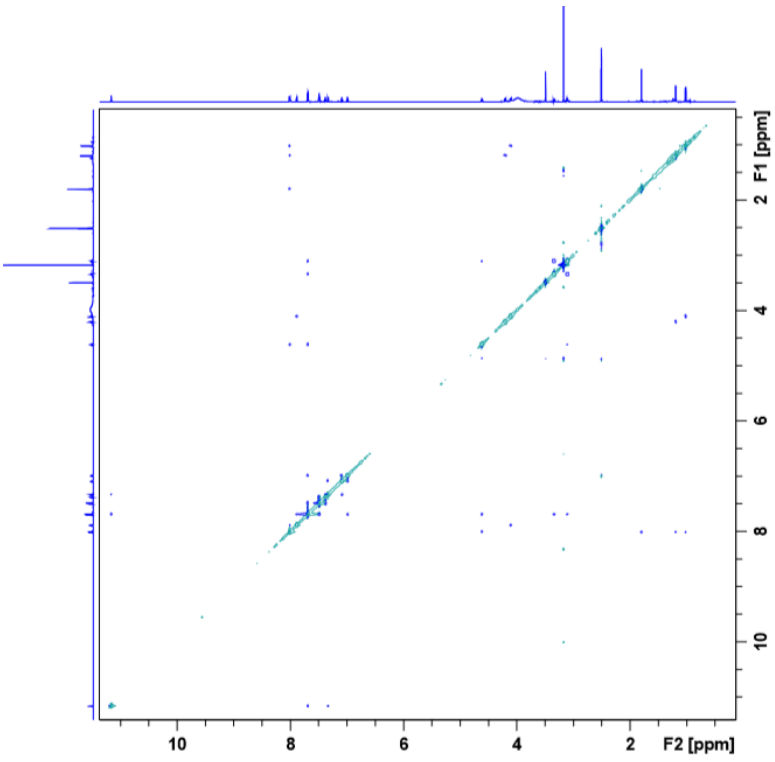
Appendix 41: ^{13}C spectrum of compound 11



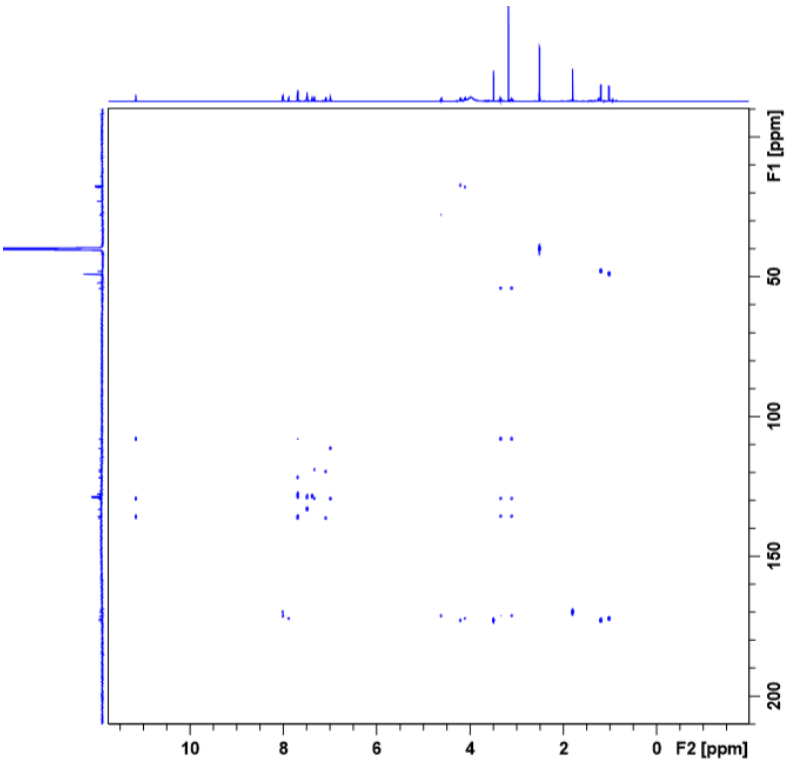
Appendix 42: COSY spectrum of compound 11



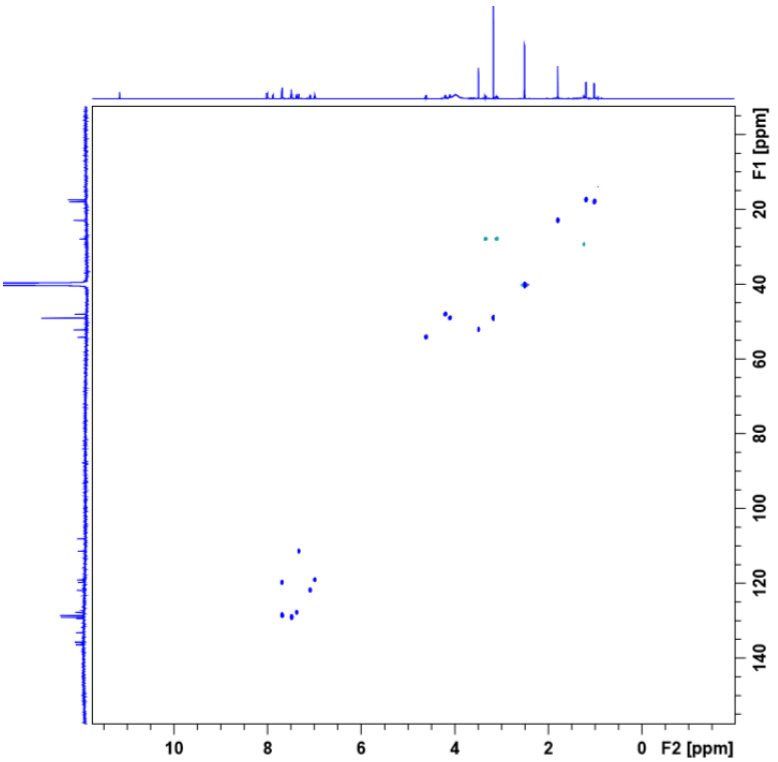
Appendix 43: TOCSY spectrum of compound 11



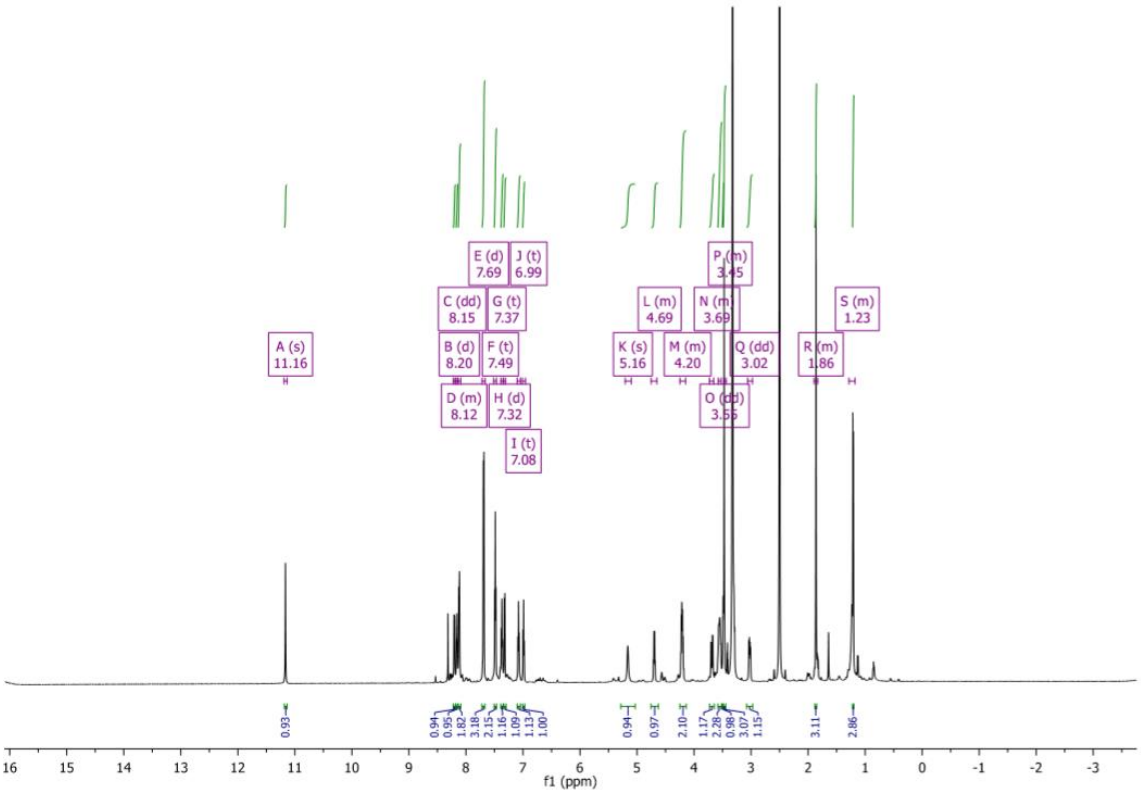
Appendix 44: ROESY spectrum of compound 11



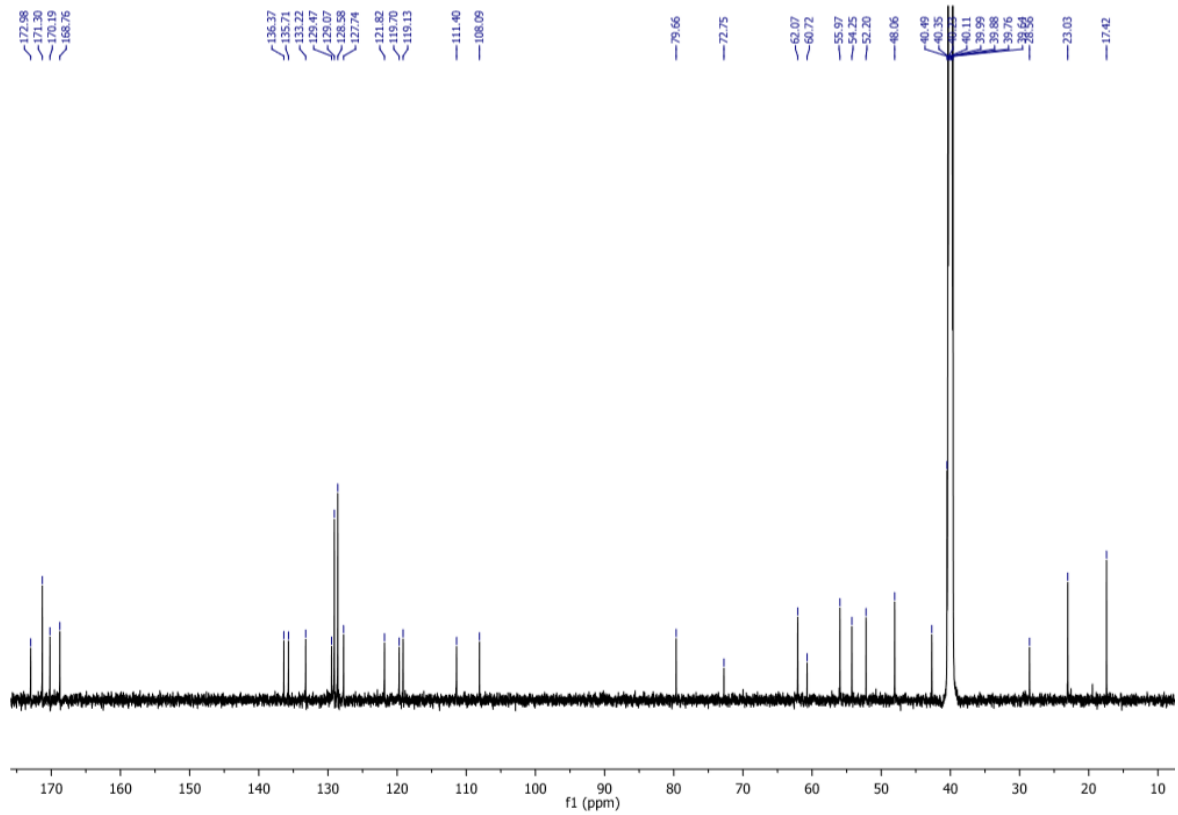
Appendix 45: HMBC spectrum of compound 11



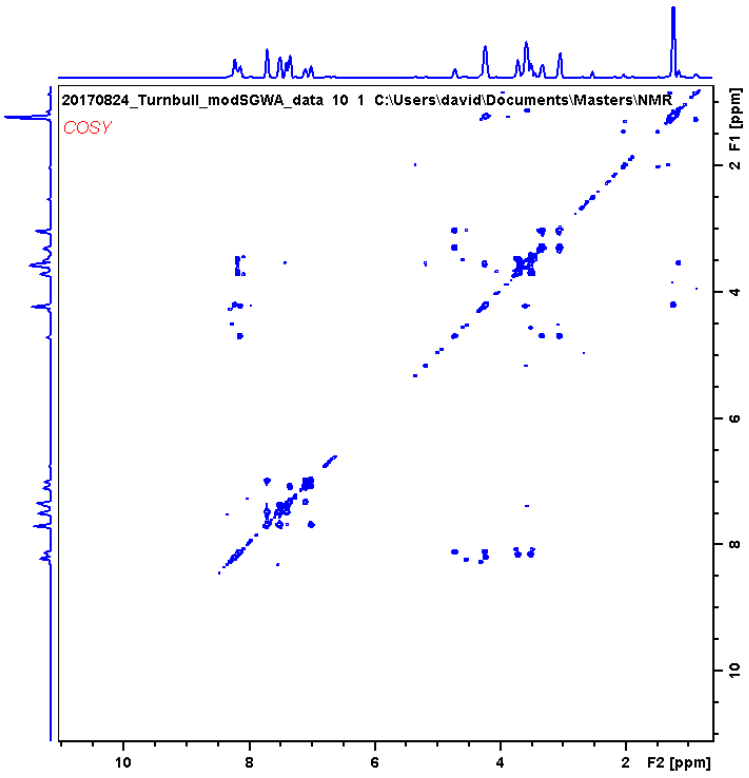
Appendix 46: HSQC spectrum of compound 11



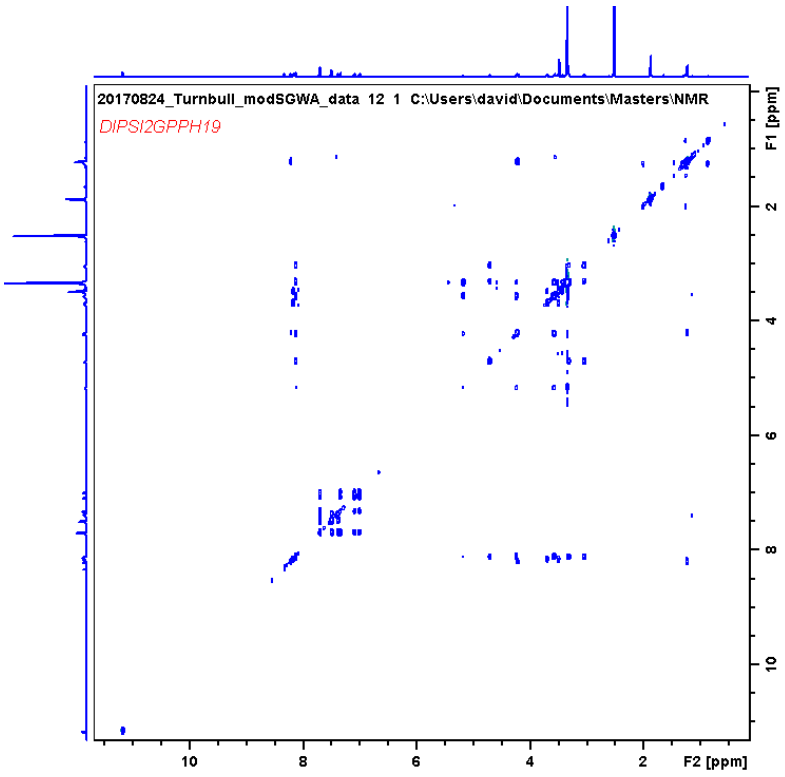
Appendix 47: ^{1}H spectrum of compound 11



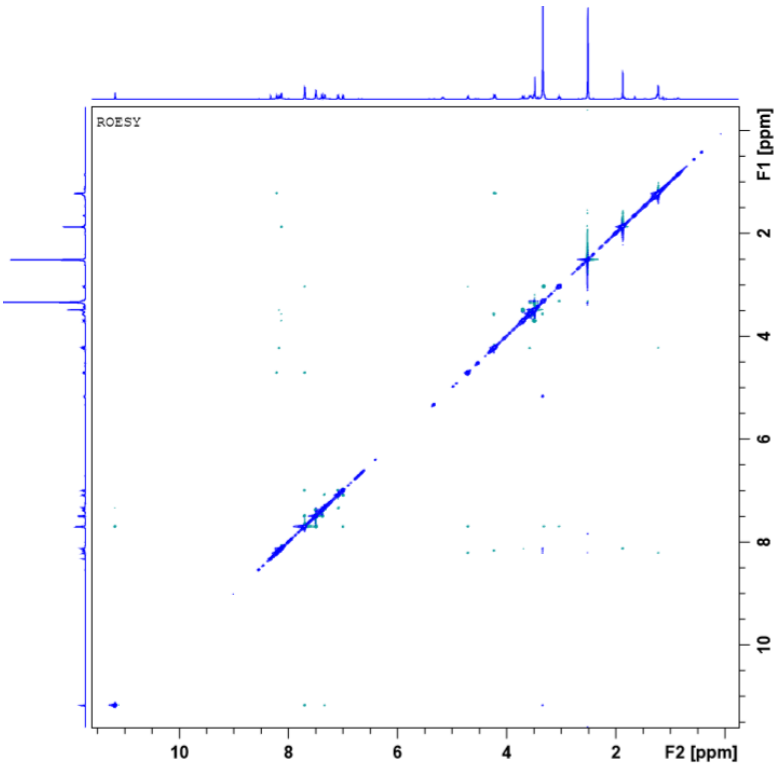
Appendix 48: ^{13}C spectrum of compound 11



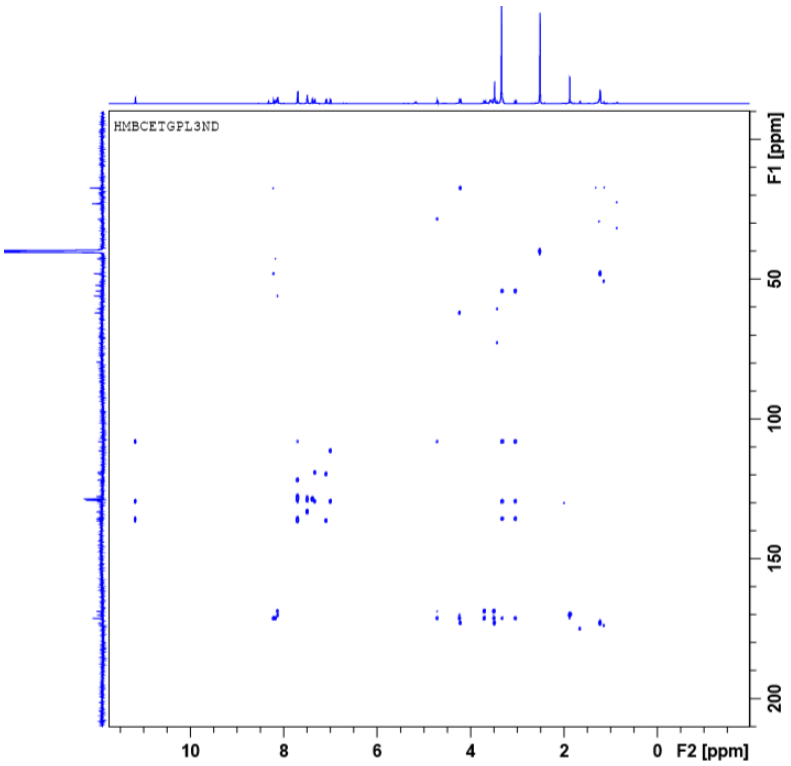
Appendix 49: COSY spectrum of compound 11



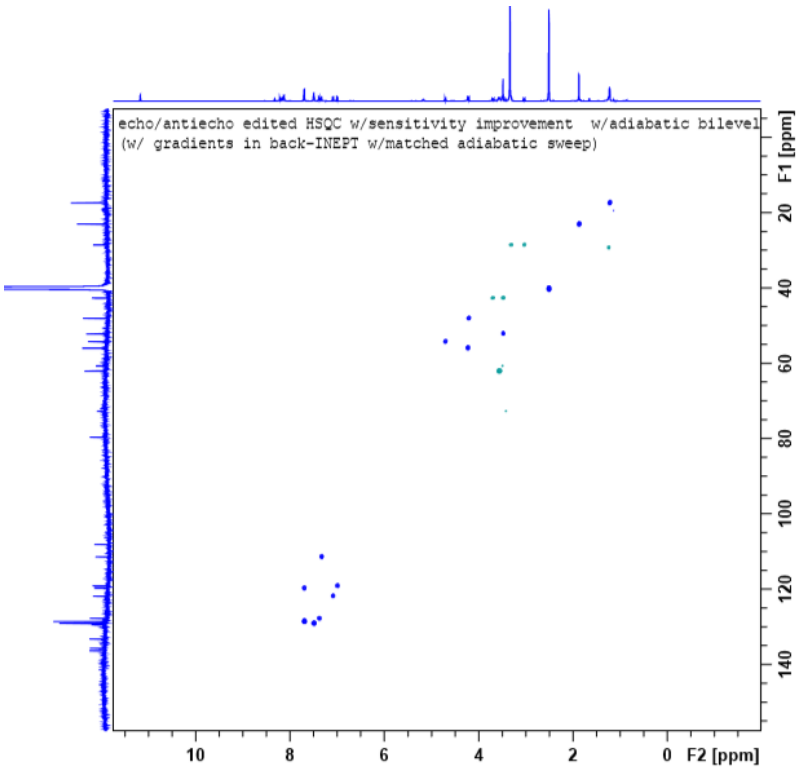
Appendix 50: TOCSY spectrum of compound 11



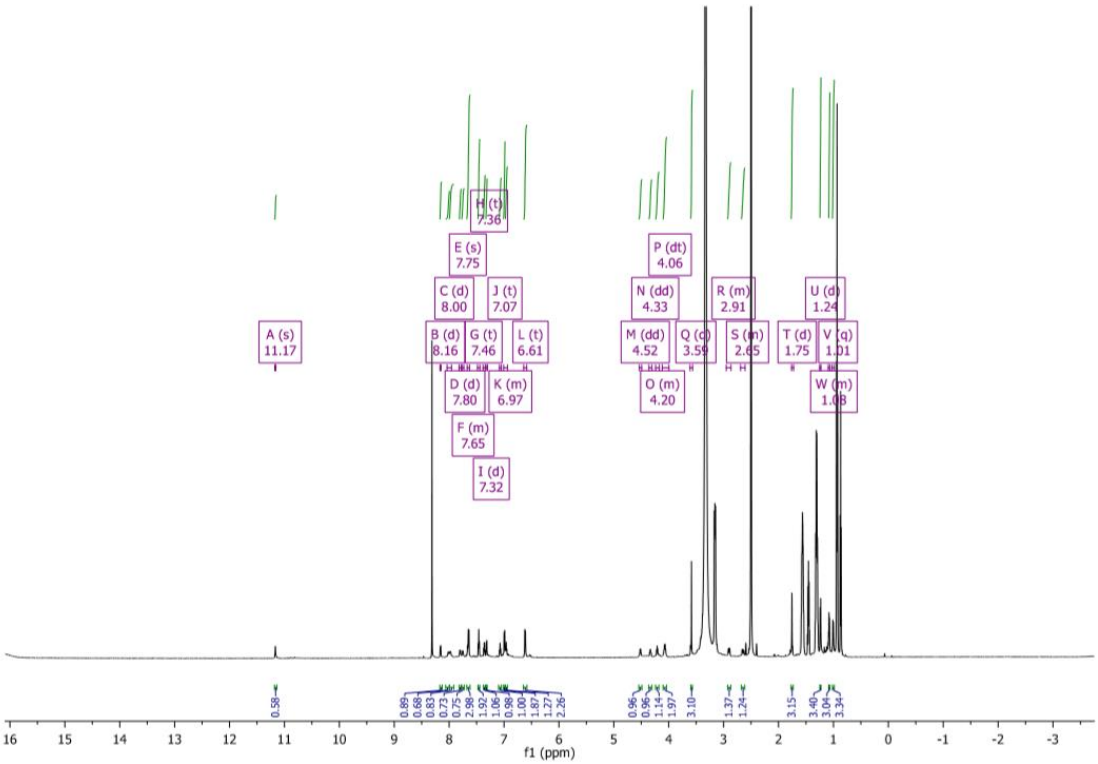
Appendix 51: ROESY spectrum of compound 11



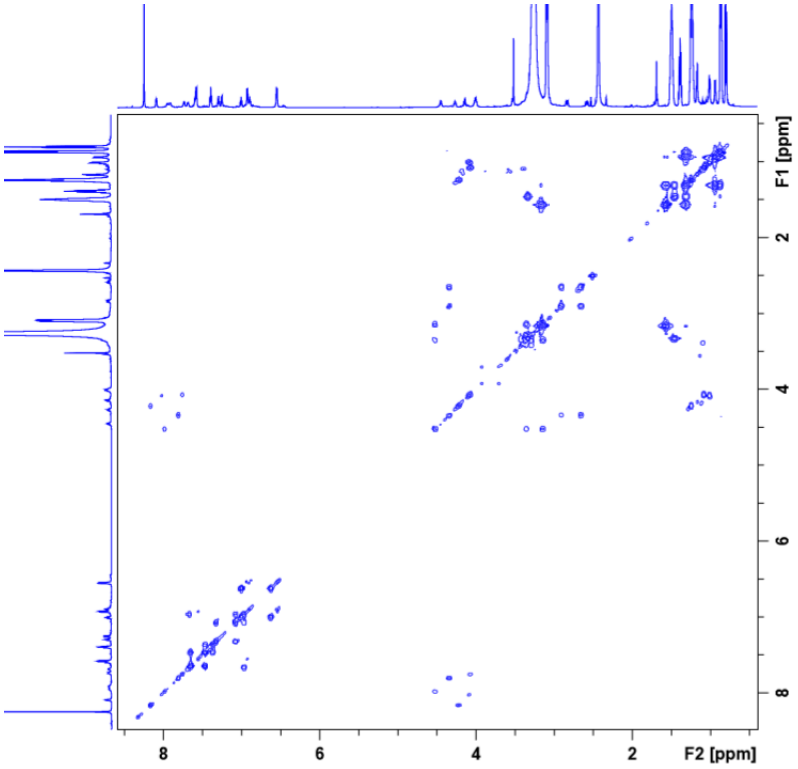
Appendix 52: HMBC spectrum of compound 11



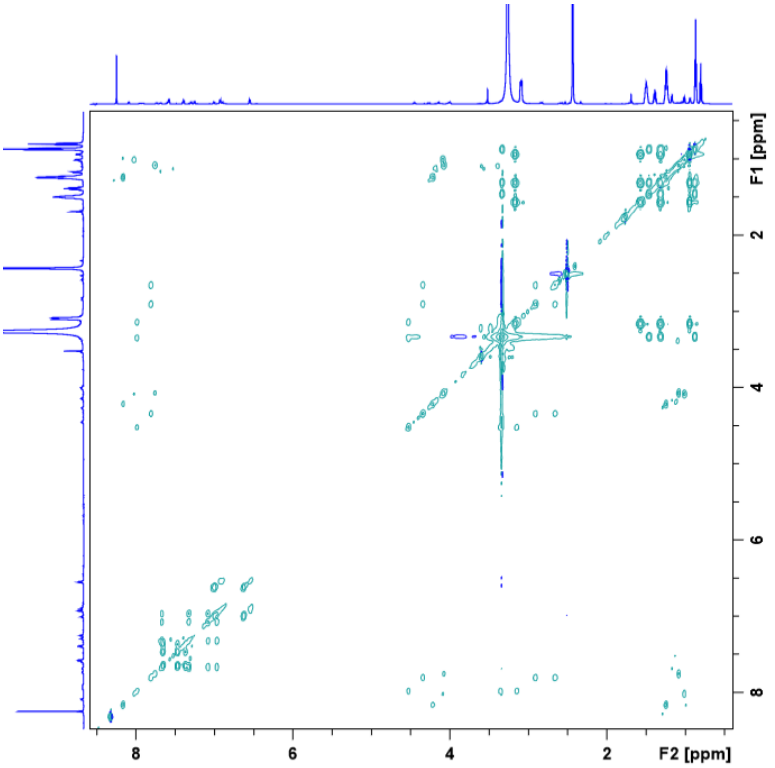
Appendix 53: HSQC spectrum of compound 11



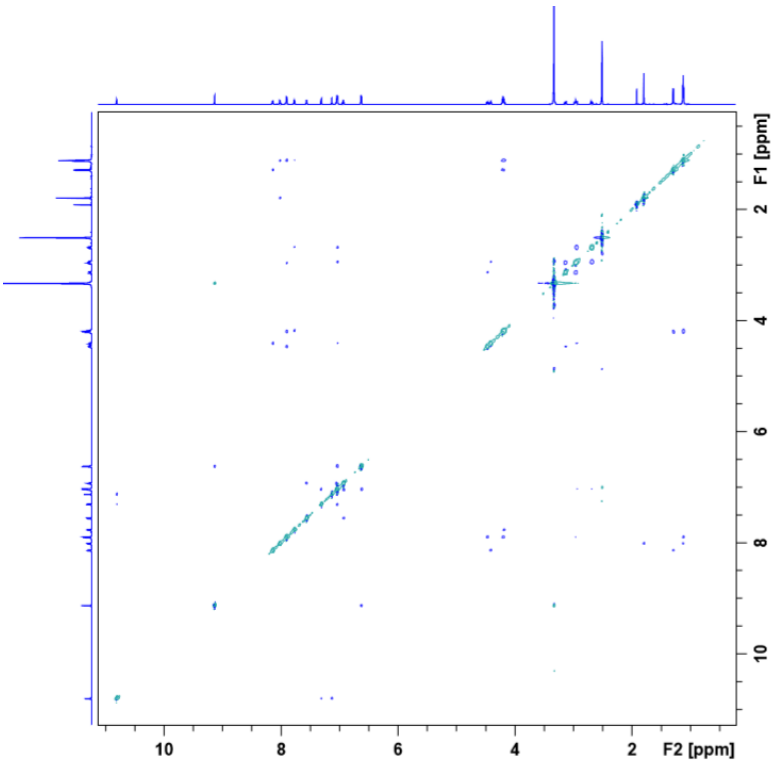
Appendix 54: ^{1}H spectrum of compound 12



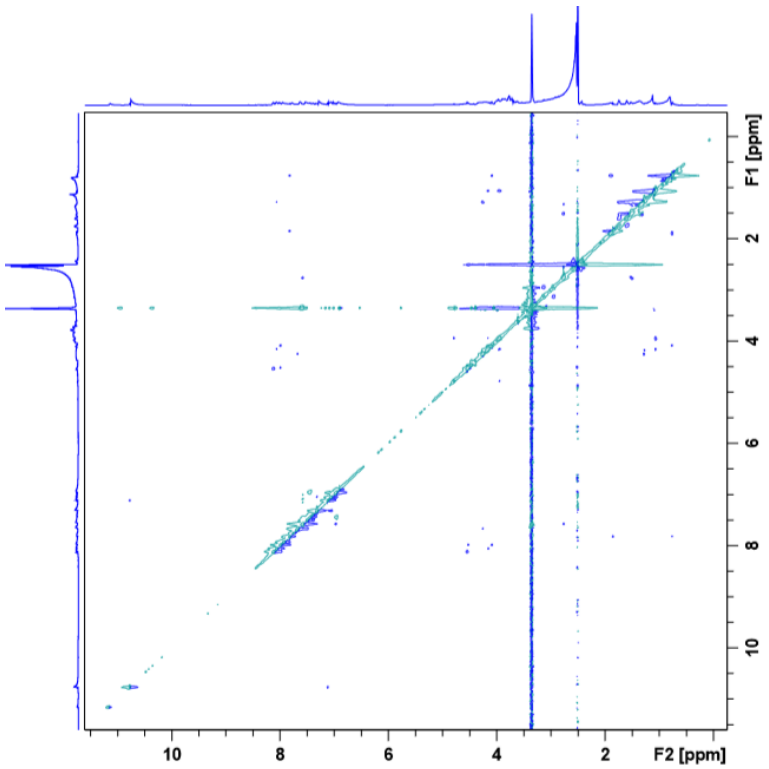
Appendix 55: COSY spectrum of compound 12



Appendix 56: TOCSY spectrum of compound 12



Appendix 57: ROESY spectrum of compound 12



Appendix 58: ROESY spectrum of compound 13

Abbreviations

Ac	Acetyl
ADHP	2-amino-4,6-dihydroxy-pyrimidine
Boc	Tert-Butyloxycarbonyl
COSY	Correlation spectroscopy
CsOAc	Caesium Acetate
DCM	Dichloromethane
DM-ADHP	N,N-dimethyl-2-amino-4,6dihydroxypyrimidine
DMF	Dimethyl formamide
DMSO	Dimethyl sulfoxide
Et	Ethyl
EtOAc	Ethyl Acetate
HBF_4	Tetrafluoroboric acid
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HPLC	High-performance liquid chromatography
HSQC	Heteronuclear single-quantum correlation spectroscopy
Hz	Hertz
Me	Methyl
MeOH	Methanol
MHz	Megahertz
Mmol	Millimolar
MSMS	Tandem mass spectrometry
m/z	Mass to charge ratio

NHC	N-heterocyclic carbenes
NMR	Nuclear magnetic resonance
OEt	Ethyl ester
OMe	Methyl ester
PAM	Polyacrylamide
PEG	Polyethylene glycol
Ph	Phenyl
p-MBHA	4-Methylbenzhydrylamine hydrochloride
ROESY	Rotating frame nuclear Overhauser effect spectroscopy
RT	Room Temperature
TFA	Trifluoroacetic acid
TLC	Thin-layer chromatography
TOCSY	Total correlation spectroscopy
TRIS	Tris(hydroxymethyl)aminomethane
TsOH	Tosic acid

References

[1] C. Walsh, *Posttranslational modification of proteins: expanding nature's inventory*, Roberts and Co. Publishers, Englewood, Colo., 2006.

[2] E. M. Sletten and C. R. Bertozzi, *Angew. Chem. Int. Ed Engl.*, 2009, **48**, 6974–6998.

[3] 3 J. M. Chalker, G. J. L. Bernardes, Y. A. Lin and B. G. Davis, *Chem. Asian J.*, 2009, **4**, 630–640.

[4] J. M. McFarland and M. B. Francis, *J. Am. Chem. Soc.*, 2005, **127**, 13490–13491

[5] A. N. Glazer, *Annu. Rev. Biochem.*, 1970, **39**, 101–130.

[6] G. Hermanson, (1996). *Bioconjugate techniques*. 1st ed. San Diego: Academic Press.

[7] J. M. Antos, M. B. Francis, *Curr. Opin. Chem. Biol*. 2006, **10**, 253.

[8] H. B. F. Dixon, *J. Protein Chem.* 1984, **3**, 99.

[9] H. B. F. Dixon, *Biochem. J.* 1964, **92**, 661.

[10] J. M. Gilmore, R. A. Scheck, A. P. Esser-Kahn, N. S. Joshi, M. B. Francis, Angew. Chem. 2006, 118, 5433;

[11] J. M. McFarland, M. B. Francis, *J. Am. Chem. Soc.* 2005, **127**, 13490.

[12] G. J. L. Bernardes, J. M. Chalker, J. C. Errey, B. G. Davis, *J. Am. Chem. Soc.* 2008, **130**, 5052

[13] T. Kodadek, I. Duroux-Richard, J. C. Bonnafous, *Trends Pharmacol. Sci.* 2005**, 26**, 210.

[14] J. M. Antos, M. B. Francis, *J. Am. Chem. Soc*. 2004, **126**, 10256.

[15] Brik, A., Maity, S., M., Jbara, M., *Angew. Chem. Int.*, 2017, **56**, 10644-10655

[16] R. Ekkebus, S.I. Van Kasteren, Y. Kulathu, A. Scholten, I. Berlin, P. P. Geurink, A. de Jong,S.Goerdayal, J. Neefjes, A. J. R. Heck, D. Komander, H.Ovaa, *J. Am. Chem. Soc.* 2013, **135**, 2867 –2870

[17] Y. Y. Yang, J. M. Ascano, H. C. Hang, *J. Am. Chem. Soc.* 2010, **132**, 3640–3641.

[18] C. D. Hein, X. M. Liu, D. Wang, *Pharm. Res.*, 2008, **25**, 2216–2230.

[19] A. L. Blobaum, *Drug Metab. Dispos*. 2006, **34**, 1–7.

[20] *J. C. van Hest, F. L. van Delft, ChemBioChem* 2011, **12**, *1309–1312*.

[21] K. Kim, D. A. Fancy, D. Carney, T. Kodadek, *J. Am. Chem. Soc.* 1999, **121**, 11896–11897.

[22] S. D. Tilley, M. B. Francis, *J. Am. Chem. Soc.*, 2006, **128**, 1080– 1081.

[23] E. V. Vinogradova, C. Zhang, A. M. Spokoyny, B. L. Pentelute, S. L. Buchwald, *Nature*, 2015, **526**,687–691.

[24] R. A. A. Al-Shuaeeb, S. Kolodych, O. Koniev, S. Delacroix, S. Erb, S. Nicolay, J. C. Cintrat, J. D. Brion, S. Cianferani, M. Alami, A. Wagner, S.Messaoudi, *Chem. Eur.J.*, 2016, **22**,11365– 11370.

[25] J. Willwacher, R. Raj, S. Mohammed and B. G. Davis, *J. Am. Chem. Soc.*, 2016, **138**, 8678–8681.

[26] A. J. Reay, T. J. Williams and I. J. S. Fairlamb, *Org. Biomol. Chem.*, 2015, **13**, 8298–8309.

[27] T. J. Williams, A. J. Reay, A. C. Whitwood and I. J. S. Fairlamb, *Chem. Commun.*, 2014, **50**, 3052–3054.

[28] Y. Zhu, M. Bauer and L. Ackermann, *Chem. Eur. J*., 2015, **21**, 9980–9983.

[29] J. Ruiz-Rodr Iguez, F. Albericio and R. Lavilla, *Chem. Eur. J*., 2010, **16**, 1124–1127.

[30] T. H. Sheridan, MChem Thesis, 2016

[31] D. R. Stuart and K. Fagnou, *Science*, 2007, **316**, 1172–1175.

[32] R. J. Phipps, N. P. Grimster and M. J. Gaunt, *J. Am. Chem. Soc.*, 2008, **130**, 8172–8174.

[33] Y. Yang, X. Qiu, Y. Zhao, Y. Mu and Z. Shi, *J. Am. Chem. Soc.*, 2016, **138**, 495–498.

[34] R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149–2154.

[35] P. Forns, F. Albericio. Encyclopedia of Reagents for Organic Synthesis, 2003

[36] R. S. Feinberg and R. B. Merrifield, *Tetrahedron*, 1974, **30**, 3209–3212.

[37] S. V. Chankeshwara, E. Indrigo and M. Bradley, *Curr. Opin. Chem. Biol.*, 2014, **21**, 128–135.

[38] R. M. Yusop, A. Unciti-Broceta, E. M. V. Johansson, R. M. Sánchez-Martín and M. Bradley, *Nat. Chem.*, 2011, **3**, 239–243.

[39] C. D. Spicer, T. Triemer and B. G. Davis, *J. Am. Chem. Soc.*, 2012, **134**, 800–803.

[40] C. D. Spicer and B. G. Davis, *Chem. Commun.* , 2013, **49**, 2747–2749.

[41] N. Li, R. K. V. Lim, S. Edwardraja and Q. Lin, *J. Am. Chem. Soc.*, 2011, **133**, 15316–15319.

[42] S. Kiyonaka, R. Kubota, Y. Michibata, M. Sakakura, H. Takahashi, T. Numata, R. Inoue, M. Yuzaki and I. Hamachi, *Nat. Chem.*, 2016, **8**, 958.

[43] C. Cennamo, B. Carafoli, E. P. Bonetti, *J. Amer. chem. Soc*. 1956. **78**, 3523.

[44] A. J. Rojas, B. L. Pentelute and S. L. Buchwald, *Org. Lett.*, 2017, **19**, 4263–4266.

[45] E., Angelini; C., Balsamini; F., Bartoccini; S., Lucarini; G.J., Piersanti. *Org. Chem.* 2008, **73**, 5654.