SYNTHESIS & COMPARATIVE BIOPHYSICAL STUDIES OF PEPTIDES DERIVED FROM 3 AND 4-SUBSTITUTED PROLINES

THESIS SUBMITTED TO
THE UNIVERSITY OF PUNE

FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
CHEMISTRY

BY

MANASWINI NANDA

RESEARCH SUPERVISOR
DR. KRISHNA N. GANESH

DIVISION OF ORGANIC CHEMISTRY
NATIONAL CHEMICAL LABORATORY
PUNE 411008

APRIL 2011

Candidate's Declaration

I hereby declare that the thesis entitled **"Synthesis & Comparative Biophysical Studies of Peptides Derived from 3 and 4-Substituted Prolines"** submitted for the degree of Doctor of Philosophy in Chemistry to the University of Pune has not been submitted by me to any other University or Institution. This work was carried out at the National Chemical Laboratory, Pune, India.

Manaswini Nanda

Research Fellow
Organic Chemistry Division
National Chemical Laboratory
Pune - 411008

April 2011

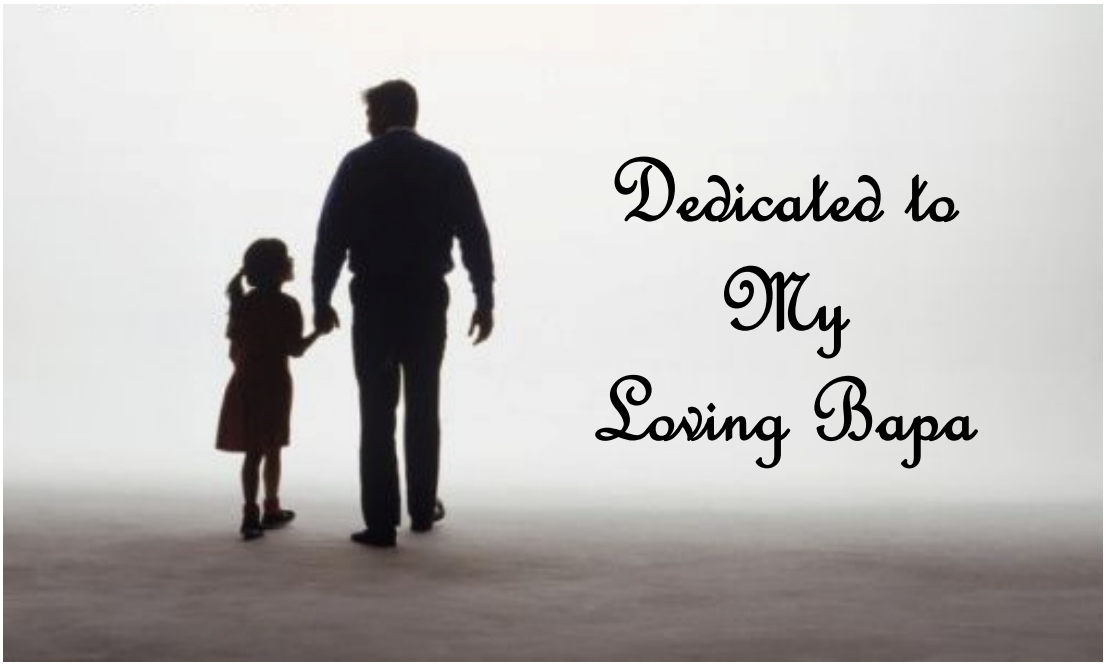
CERTIFICATE

This is to certify that the work presented in the thesis entitled "**Synthesis & Comparative Biophysical Studies of Peptides Derived from 3 and 4-Substituted Prolines**" submitted by ***Manaswini Nanda*** was carried out by the candidate at the National Chemical Laboratory, Pune, under my supervision. Such materials as obtained from other sources have been duly acknowledged in the thesis.

Prof. K. N. Ganesh, FNA, FNASc
(Research Supervisor)

Director, IISER, Pune
J. C. Bose Fellow
National Chemical Laboratory
Pune-411008

April 2011


Dedicated to
My
Loving Bapa

Acknowledgement

First of all I want to express my heartfelt sincere profound thanks to my beloved father (Bapa) and mother (Maa) in bringing me up to this stage to become a good academician in the intellectual society.

It gives me a great pleasure to express my deep sense of gratitude to my research supervisor Prof. K. N. Ganesh for all his advice, guidance, support and encouragement in chemistry and its relevance in biological systems. His tireless enthusiasm was always a source of inspiration. Additionally, I revere his lessons on independent thinking, perfection softs kills and many more in shaping up the researcher in me. I am also thankful to for his valuable suggestions and advice at the time of difficulty on personal matters.

I am grateful to the Director, Dr. Sourav Pal, NCL, Pune for all infrastructural and administrative support and facilities that have been provided during my research period. I am also thankful to Dr. Ganesh Pandey, Head, Division of Organic Chemistry, NCL, Pune for rendering all laboratory facilities to carry out my research work smoothly. The financial support from CSIR in the form of research-fellowship for 5 years is also acknowledged.

I acknowledge my sincere thanks to the faculty professors of IISER-Pune Dr. H. N. Gopi, Dr. A. A. Natu, Dr. Partha Hazra, Dr. Girish Ratnaparkhi, Dr. S. G. Srivastav for their support in peptide synthesis, HPLC, fluroscence analysis, cell studies and DNA related work respectively along with many fruitful discussions that helped me design my experiments.

My special vote of thanks would be to Dr. V. A. Kumar, Dr. Moneesha, Dr. H. V. Thulasiram, Dr. M. J. Kulkarni and Dr. G. J. Sanjayan along with their research groups for their help, encouragement and co-operation during my research work. I am thankful to Mrs Meenakshi Mane & Sunita Kunte for HPLC analysis, Mrs Shantakumari for LCMS work, Mrs Anita Gunjal for providing synthesised DNA, Mrs V.Puranik for X-ray analysis, Dr. M. J. Kulkarni for MALDI-TOF experiments. The kind support from NMR group is greatly acknowledged and thanks to Dr Rajamahanan and Hilda Davis for their extra efforts. I also thank Mr K D. Deshpande, Mr Mane, and Mr Javed for solving my computer related problems.

I must acknowledge the help from students of IISER pune, specially from Senthil (Cell Studies), Hina (enzyme assay), Umer, Madhan, Sandip (peptide synthesis), Sachin (CD), Abhigyan (Fluorescence & ITC), Anupam (X-ray), Rahul, Maroti and Sekhar. The support from Mayura and Pooja relating to MALDI and NMR work carried at IISER is also highly appreciated. I am thankful to Naina & Mahesh

for administrative support, Chengaiah & Vrushali for proving purchase support on behalf of IISER-pune in the NCL-IISER collaborative research work.

The research becomes boring if I would not have wonderful labmates around. I am fortunate to have a excellent blend of talented and unique labmates who can rise in any occasion in their own different way, in addition to being cooperative and helpful. The association of educative seniors like Dr. Pravin, Dr. Nagendra, Dr. Dinesh, Dr. Khirud, Dr. Sunil, Dr. Gourishankar, Dr. Raman, Dr. Umashankara, Dr. Patwa, Dr. Gitali, Dr. Sachin, Dr. Nasrin, Dr. Madhuri, Ashwani, Sreedhar, Roopa and Pradnya helped me a lot during my research work. My special thanks to Dr. Umashankara and Roopa who extended their moral support for solving my experimental problems in novel way. I enjoyed working with my labmates like Mahesh, Tanpreet, Deepak, Vijay, Nitin and Satish in lab 209. The experiences in lab 226 will also remain an unforgotable memory in my research life.

Further I thank my labmates for bearing my presence and my roommates (Priya, Meera, Shreaddha, Ramya, Sridevi & Manisha) for bearing my absence. My friends were always there for me whenever I required some serious relaxation with frustrating results at lab:, who caused the much needed "load-sheding " to kickback power to move forward afresh. During my stay in New hostel and GJ hostel of NCL, Pune for the last 5 years life was cheerful, enjoyable and pleasant with friends like Shreaddha, Suman, Bibhuti, Divya, Roopa, Bibhasda, Manisha, Venu, Payal, Ruby, Priyanka, Gowri, Manasi, Lalitha, Sampa *et al*.. My early research days at NCL made colourful friendship with another charged group of oriya friends and aquintances as Sabita, Seetaram, Debasish, Rosy, Gokarneshwar, Shashi, Bhuban, Dibyadarshani, Sivaram, Pitambar, Tiny, Ramakanta, Mandakini, Chakadola, Subash S., Subash M., Raju, Jitu, Pushpanjali who filled in the gaps of occupational failures & lonliness with numerous joyous moments of happiness. I am indebted to Sivaram, Gowri and Tanpreet for their kind help in literature referencing and arrangement of the supporting data in this thesis.

The thesis remains incomplete without mentioning the motivation blessings and love of my father-in-law and family. Their love support and credence have strengthened me for all my ventures in different ambits. At this point I would like to mention about the special person in my life Dr. Satyartha Prakash (Bapi) as this thesis would be incomplete if his encouragement, support and above all his patience would not have lingered up till now. He has been and would be my beloved husband and soul-mate as we move forward together for eternity.

Finally I pray to almighty Lord Sri Jagannath, whose blessings (kṛupa) has made me able to complete the research work and submit this thesis for PhD degree.

Manaswini Nanda

Contents

Abstract v

Abbreviations xiii

List of Figures xvi

List of Tables xxi

List of Schemes xxii

Prologue xxiii

Section -I

1. Synthesis and Biophysical Studies on DNA Binding and Cell Penetrating ability of Chimeric 4-Amino & Guanidino-proline Collageneous Peptides

1. 1 Introduction…….... 1

1. 2 DNA – A structural overview…….... 1

1. 3 Plasmid DNA……..... 4

1. 4 DNA and molecular recognition……..... 5

1.4.1 Non-specific outside-edge interactions……..... 7

1.4.1a Condensation type interactions……..... 7

1.4.1b Electrostatic Interactions……..... 7

1.4.2 DNA binding drugs……..…… 8

1.4.2a Intercalating molecules……..... 9

1.4.2b Groove binding molecules……..... 10

1.4.2c Intercation *vs.* Groove binding molecules…….. 12

1.4.3 DNA-Protein interactions……. 13

1. 5 Gene / DNA delivery……..... 16

1.6 DNA delivering peptides/ Peptide mediated DNA delivery…… 17

1.7 Physical and structural techniques applied to study of nucleic acids (DNA) …….. 21

1.7.1 Spectroscopic Techniques……..……..... 21

1.7.1a Ultraviolet Absorption……..... 21

1.7.1b Circular Dichroism……..... 22

1.7.1c Fluorescence……..……... 25

1.7.2 Electrophorectic mobility shift assay……..... 26

1.7.3 Microcalorimetry……..……..... 28

1.7.4 Mass Spectrometry: MALDI-TOF……..... 29

1.7.5 Transfection…….…….... 30

1. 8 Aim and rationale of the present work…….... 32

1. 9 Results…….... 34

1.9.1 Synthesis of monomers…….... 34

1.9.1a Design of monomers…….... 34

1.9.1b Synthesis of (2*S*,4*R*)- and (2*S*,4*S*)-4-aminoproline monomers..... 35

1.9.2 Solid phase synthesis…….... 37

1.9.2a Synthesis of cationic peptides…….... 40

1.9.2b Synthesis of cationic fluorescent peptides……..... 42

1.9.2c Synthesis of oligonucleotides…….... 46

1.9.3 Biophysical Studies…….... 47

1.9.3a Conformational analysis of peptides by CD spectroscopy..... 47

(i) Concentration dependent CD spectroscopy..... 48

(ii) CD - thermal denaturation studies…… 49

(iii) CD study of salt effect on peptides..... 50

1.9.3b UV-T_m studies on Interaction of peptides with *ds*DNA..... 51

1.9.3c CD study of *ds*DNA binding to cationic peptides..... 56

1.9.3d CD study of plasmid DNA binding to cationic peptides..... 59

1.9.3e Fluorescence study - EtBr displacement in *ds*DNA..... 61

1.9.3f Fluorescence study - EtBr exclusion in *ds*DNA..... 63

1.9.3g Fluorescence study - EtBr assays with plasmid DNA..... 64

1.9.3h Native PAGE…….... 65

1.9.3i Agarose Gel Retardation Assay…….... 67

1.9.3j Restriction enzyme digestion of plasmid DNA-peptide complexes. 69

1.9.3k Cell Studies…….…….. 70

(i) Cell transfection with cationic peptides..... 71

(ii) Cell permeation of fluorescent cationic peptides..... 73

(iii) Cell transfection with fluorescent peptides..... 75

1.9.3l Isothermal titration calorimetry…….... 77

1.10 Discussion…….…….... 79

1.11 Conclusions…….…….... 87

1.12 Experimental…….. 88

1.12.1 Synthesis of compounds/monomers…….... 88

1.12.2 Solid Phase synthesis…….... 98

1.12.3 Purification & characterization……..... 104
1.12.4 Biophysical techniques……..... 106
1.12.5 Biological protocols……..... 111
1.13 References…… 117
1.14 Appendix A……..... 125

Section -II

2. NMR Studies of *cis:trans* Isomerization in Prolyl Peptide bond with respect to Substitution at 4-position of Proline in Ala-Pro based Dipeptides

2.1 Introduction……..... 159
2.2 Prolyl-Peptidyl bond isomerization……..... 161
2.2.1 Pyrrolidine ring conformation in proline and substituted proline..... 163
2.2.2 *Gauche* effect on the ring-pucker preferences……..... 164
2.3 Biological Relevance of Prolyl-Peptidyl bond isomerization…….. 166
2.3.1 Some cited applications with examples……..... 167
2.3.2 Implications to the triple helical structure of collagen…… 167
2.4 Aim and rationale of the present work……..... 168
2.5 Results…… 170
2.5.1 Solution phase synthesis of Ala-Pro model dipeptides……..... 170
2.5.2 Crystal structure……..... 178
2.5.3 Two-dimensional NMR study of Ala-Pro model dipeptides…….. 180
2.5.3a COSY: Assignment of dipeptide backbone……. 180
2.5.3b NOESY: Characterization of *cis* & *trans* isomers..... 187
2.5.4 pH dependent NMR analysis of dipeptides……..... 194
2.5.5 pK_a determination of Ala-Pro model dipeptides……..... 199
2.5.6 Conformational review by CD spectroscopy…….. 200
2.6 Discussion……..... 203
2.7 Conclusions……..... 206
2.8 Experimental…….. 207
2.8.1 Synthesis of dipeptides……..... 207
2.8.2 HPLC purification ……..…… 223
2.8.3 pK_a Determination……..... 223
2.8.4 pH depentendent NMR spectroscopy……..... 224
2.8.5 pH depentendent CD spectroscopy……..... 224

2.8.6 X-ray crystal structure determination……..... 225

2. 9 References……. 228

2.10 Appendix B……..... 232

Section -III

3. 3-Hydroxy and 3-Aminoproline in Chimeric Collagen Peptides: Synthesis & CD Studies

3.1 Introduction……..... 299

3.2 Positional preferences of amino acids in collagen……..... 301

3.3 Three dimensional structure & crystal packing of collagen…… 302

3.4 Role of hydroxyprolines in collagen structure……..... 303

3.4.1 Collagen structures with 4(*R/S*)-hydroxyproline…….. 303

3.4.2 Collagen structures with *trans*-3-hydroxyproline…….. 304

3.5 Characterization of triple-helical structures……..... 305

3.6 Collagen mimetics……..…… 306

3.7 Collagen applications: Ageing, disease & biomaterials…….. 307

3.8 Aim and rationale of the present work…….. 308

3.9 Results……..…… 311

3.9.1 Synthesis of monomers……..... 311

3.9.1a Synthesis of (2*S*,3*S*) and (2*S*,3*R*)- aminoproline monomers..... 311

3.9.1b Synthesis of N^{α}-Fmoc (2*S*,3*S*) and (2*S*,3*R*)-hydroxyproline monomers 313

3.9.2 Solid phase peptide synthesis……..... 314

3.9.3 Conformational study by CD spectroscopy…….. 322

3.9.3a Concentration dependent CD spectroscopy..... 322

3.9.3b Thermal denaturation of peptides..... 324

3.9.3c Behaviour of peptides in ethylene glycol..... 326

3.10 Discussion ……..... 328

3.11 Conclusions……..……..... 329

3.12 Experimental…….. 330

3.12.1 Synthesis of monomers/compounds……..... 330

3.12.2 Purification & characterization……..... 340

3.12.3 CD spectroscopy & thermal denaturation studies……..... 341

3.13 References…… 342

3.14 Appendix C……..... 347

Abstract

The thesis entitled "**Synthesis & Comparative Biophysical Studies of Peptides Derived from 3 and 4-Substituted Prolines**'' is comprised of studies towards the design and synthesis of peptides based on 4-substituted prolines along with their behavior under variable physical envionments which may have importance in understanding their biological functions. The work also describes DNA recognition, cell penetrating properties and/or stability of some of the derived collageneous peptides. The thesis is divided into three parts:

Section I

Synthesis and Biophysical Studies on DNA Binding and Cell Penetrating ability of chimeric 4-Amino & Guanidino-proline Collageneous Peptides

This section involves the study of interaction of cationic peptides with DNA. The role of 4-substituted aminoprolines and their guanidium derivatives in molecular recognition of peptides leading to DNA complexation is elaborated. Improvements in pharmacokinetic properties for better cellular uptake have been targeted through designed cationic collagen synthesis. The effects on hybridization properties of DNA-peptide complexes have been investigated through different biophysical and biological studies.

Recent literature and advancements in the area of cell penetrating peptides emphasize on the utility of proline based cationic peptides and rationale for undertaking the research work.

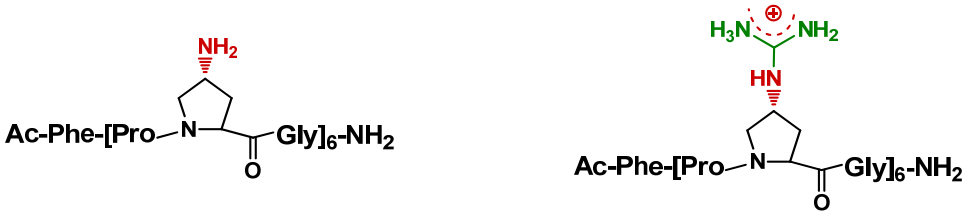


Fig. 1. (**A**)Amino and (**B**)Guanidino chimeric collagen peptides

Since the cationic peptides are known to be efficient DNA transporting agents,[1] the cationic guanidino collagen peptide analouges are examined for their DNA binding and cell penetrating abilities. Guanidinium modifications to the 4-amino side chain of proline incorporates extra positive charge for better recognition as observed in arginine[2] or lysine[3] containing peptides.

The 4-Amino proline collagen peptide sequence were synthesized by solid phase peptide synthesis and all 4-NH_2 prolyl function was converted into corresponding guanidium function on resin using the reagent *N*,*N*'-bis-Boc-1*H*-pyrazole-1-carboxamidine. The purity of these sequences were checked by RP-HPLC and the sequences were characterized by MALDI-TOF mass spectrometry.

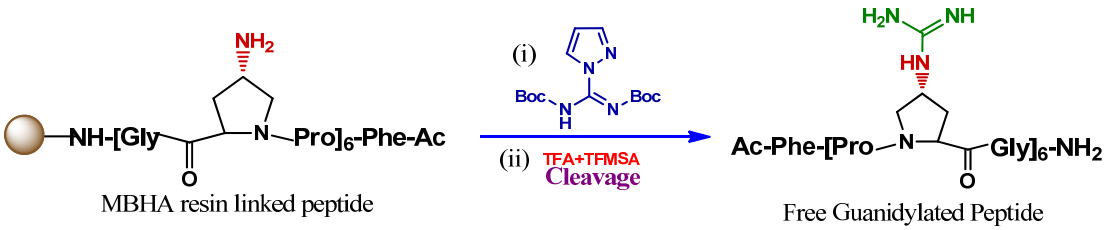


Fig.2 Solid Phase Synthesis & Cleavage of **P4** (RG)

UV-T_m values suggest that the cationic peptides provide improved thermal stability to small duplex oligonucleotides, while CD studies highlight the possible role of peptide conformation in their interaction towards DNA strands. Moreover the peptide binding to binding cooperatively distorts the DNA double helix into condensed forms. Competitive binding of these peptides with respect to Ethidium ions in displacement and exclusion assay were observed through fluorescence studies. The cationic guanidino peptides effectively neutralised the charge on DNA molecules to completely retard them in electrophoretic mobility shift assays (agarose gel electrophoresis). These peptides play a protective shield for the DNA and shield them from enzymatic digestion.

The non-specific electrostatic interaction not only enable cationic peptides to bind to DNA but also fecilitate their entry into cells. This was verified from the transfection studies when complexation of GFP reporter vector with synthesized cationic peptides resulted in enhanced fluorescence or GFP expression by replacing commercially available transfection reagent (*enhancer*). Furthermore texas red

conjugated 4-guanidinoproline peptides trasverse the cell membrane to localise themselves in various cell organelles are a direct proof of their prospective applications in gene therapy.

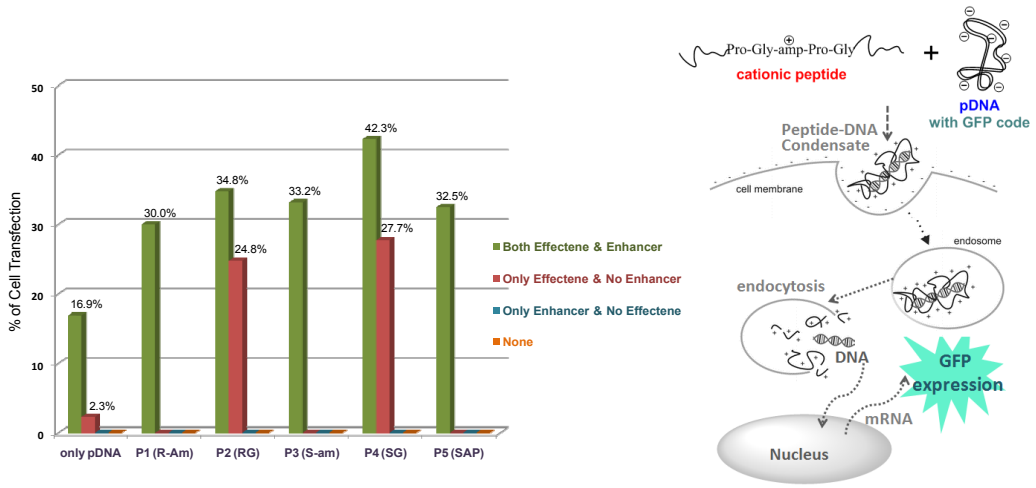


Fig.3 (**A**) Representative diagram of cell penetration of peptide mediated cell transfection, (**B**) Comparative transfection profile (with respect to only *p*DNA) for transfecting reagents in presence of various synthesised cationic peptides

Thus it was shown that 4-guanidinoproline peptides with either stereochemistry bind to DNA, but with varying affinities and probably with different sequence specificities.

Section II

NMR Studies of *cis:trans* Isomerization in Prolyl Peptide bond with respect to Substitution at 4-position of Proline in Ala-Pro based Dipeptides

This section starts with introduction to occurance and properties of 4-substituted prolines and their peptides, with a focus on the likely importance of 4-substitution on prolyl-peptide bond isomerization. It discusses the NMR investigation done to explore the effect of stereochemistry and pH on overall *cis-trans* orientation selectivity of those molecules, backed by crystal structure, CD and pK_a studies. Ala-Pro based dipeptides were chosen as best model for undertaking the research work .

Proline derivatives having 5-membered pyrrolidine ring with strong conformational biases are potent tools to elucidate the roles of secondary structure

like turn geometry and *cis:trans* isomerization in protein structure and function.[4] Thus stereochemically well-defined and 4-substituted prolines can play an important role in modulating the conformation of derived peptides thereby leading to a detailed understanding of structure-function relationship.

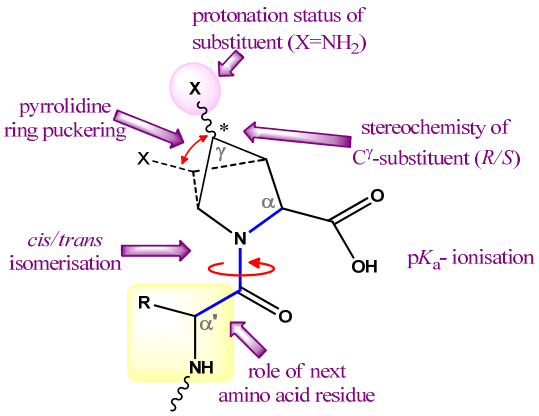
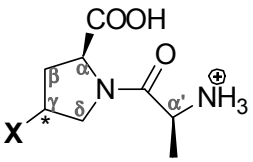


Fig. 4 Factors involving *cis-trans* isomerization in 4-substituted proline

C*is-trans* isomerization in 4-substituted prolines, depends on the steric and electronic effects exerted by the 4-substitutent. Thus, pyrrolidine ring may prefer any one of the ring-puckers which may influence the variation in *cis:trans* isomerization.[5] In order to understand the relative effects of protonation/nonprotonation, stereo-chemical and electronic influences of the 4-substituent on the *cis:trans* equilibrium of peptidyl-prolyl bond and the conformation of the pyrrolidine ring the effect of 4-substituents in proline model compounds of type H_2N-Ala-Pro-OH (4-substituted-proline-alanine dipeptides) under different pH conditions were examined *via* NMR study. For this purpose *hydroxyl*, *amino* and *azido* groups were chosen (Fig **5**).



Dipeptide 19. X = H = *Pro*

23. X = 4*R*-OH = *Hyp*	**28.** X = 4*S*-OH = *hyp*
31. X = 4*R*-N_3 = *Azp*	**34.** X = 4*S*-N_3 = *azp*
36. X = 4*R*-NH_2 = *Amp*	**38.** X = 4*S*-NH_2 = *amp*

Fig. 5 Ala-Pro based dipeptides with substitution at 4-position

The section II of the thesis describes the synthesis and comparative study of the peptidyl-prolyl bond isomerization in the above Ala-Pro model compounds. The syntheses of all these dipeptides were done by the solution phase peptide coupling

method using mostly EDCI [1-Ethyl-3-(3-dimethyl aminopropyl) carbodiimide hydrochloride] and HOBt (N^1-Hydroxybenzotriazole) as coupling reagents.

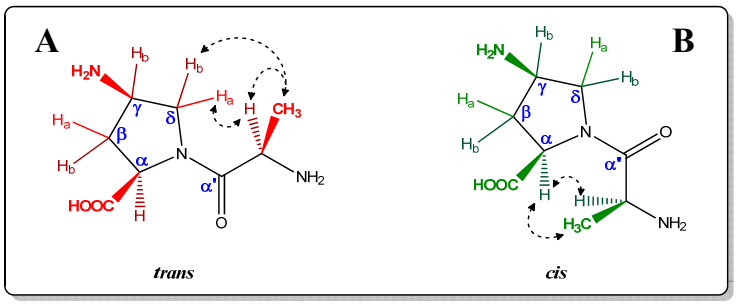
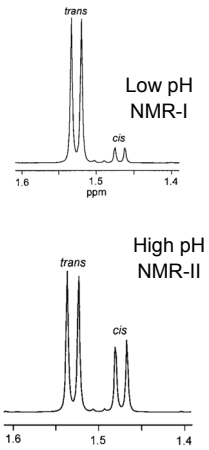


Fig. 6 Expected NOE interaction in case of Ala-amp dipeptide **38**, as (**A**) *trans* & (**B**) *cis* isomers

The dipeptides were investigated by ^{1}H NMR under five different pH conditions. The assignments of different protons were done by 1D ^{1}H decoupling and 2D ^{1}H-COSY, NOESY experiments. The marker alanyl CH_3 protons appeared as doublet and the major isomer was identified as *trans* and the minor as *cis* by NOESY. The ratio of *trans:cis* isomers as a function of pH were estimated from the relative ratio of *trans:cis* as seen for dipeptide **19** shown in Fig **6**. The results tabulated in Table **1** indicate that the *cis* isomer increases in content at alkaline pH.

Table 1: Variation of *trans* isomer content as a function of pH in various dipeptides.

	Dipeptide	% of *trans* Isomer				
		pH 2.2	pH 4.4	pH 7.0	pH 9.5	pH 12.0
19	Ala-Pro	*87*	*63*	*61*	*50*	*49*
23	Ala-Hyp	*89*	*73*	*72*	*58*	*52*
28	Ala-hyp	*78*	*62*	*55*	*52*	*51*
31	Ala-Azp	*87*	*69*	*66*	*53*	*53*
34	Ala-azp	*72*	*42*	*39*	*36*	*36*
36	Ala-Amp	*80*	*59*	*55*	*53*	*52*
38	Ala-amp	*68*	*60*	*59*	*59*	*54*
41	Ac-Ala-Amp	*88*	*88*	*88*	*88*	*88*
44	Ac-Ala-amp	*83*	*81*	*81*	*90*	*91*



The 4*R*-substituted dipeptides showed a trend more or less as that of (unsubstituted) dipeptide Ala-Pro **19**, while the 4*S*-substituted dipeptides showed a slightly different result with relatively higher predominance of the *cis*-isomers.

Overall the c*is:trans* ratio is dependent on pH where *trans* conformation is favoured at acididc pH.

From all the results obtained it was infered that near the pK_a value of the acid/amino group, the dipeptide exists mostly in zwitterionic form; there is little difference in the energy of the two isomers (*cis/trans*) due to formation of intramolecular salt bridge. Although the free amino group at N-terminus of dipeptide has a major role in determining the *cis: trans* ratio, the effect of substituents at 4-position cannot be ruled out.

Section III

3-Hydroxy and 3-Aminoproline in Chimeric Collagen Peptides: Synthesis & CD Studies

This section presents a brief introduction of collagen and the recent literature trends in research of this area. The effects of the different structural modifications on the properties of collagen-triplex stability have been overviewed to draw directions for the present work, targeted towards exploring new pathways of chimeric sequence design and investigation. The work in this section of thesis is mainly devoted to collagen sequence modification by introduction of hydroxyl or amino-functionality (as non-natural amino acid) at 3-position of proline ring at X/Y positions in the (X-Y-Gly)$_n$ backbone and study of its conformational properties.

In recent years, the biomedical importance of collagen has led to increasing interest in the structure of this protein and in establishing relationships between sequence, structure, and function.

During the biosynthesis of natural collagen, post-translational modifications occur, where in addition to prolyl 4-hydroxylation, a small numbers of proline residues are modified to 3(*S*)-hydroxyproline [3(*S*)Hyp][6] in many types of vertebrate collagens. Whether 3(*S*)Hyp residue in the Xaa position stabilizes or destabilizes the desired collagen helix is still controversial as the role of 3(*S*)Hyp residues in the Xaa position is not very well understood.[7,8] In host-guest peptides, it was found that the stability of the triple helix is decreased when Pro in the Xaa position is replaced by 3(*S*)Hyp[7.7]

The 4(*R/S*)-aminoprolines have already been tested as a surrogate of 4-hydroxyproline in both Xaa and Yaa position in collagen sequence of $(X\text{-}Y\text{-}Gly)_6$ type which show better triple-helical propencity.[9] In case of 3(*R/S)*-aminoprolines the 3-NH_2 group, like hydroxyl is a potent hydrogen bond donor. Moreover, its higher basicity compared to -OH causes its protonation at physiological pH (Fig **7**). Overall amino group is ionisable, electronegative with potential to form hydrogen bond.

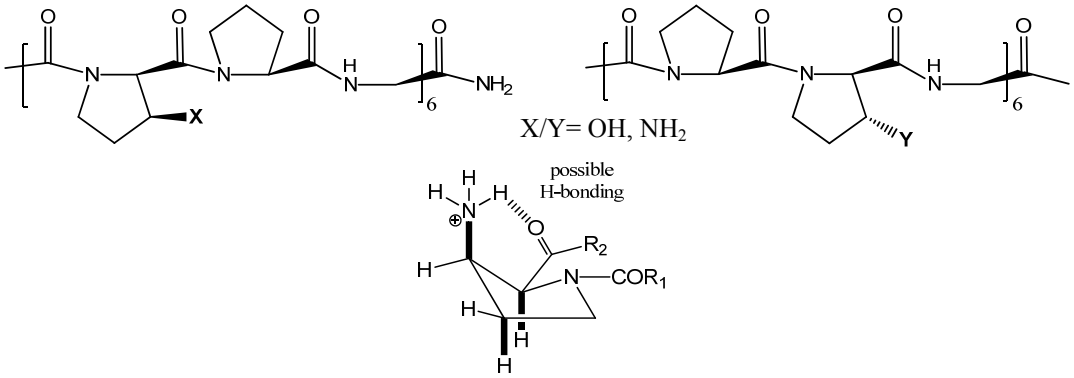


Fig.7 Collagen sequences with 3-substituted prolines and possible hydrogen bonding in 3*S*-aminoproline

(2*S*,3*S*)-N^{α}-(*t*-Boc)-3NH-Fmoc-aminoproline (**52**), (2*S*,3*R*)-N^{α}-(*t*-Boc)-3NH-Fmoc-aminoproline (**56**), (2*S*,3*S*)-N^{α}-Fmoc-3-Hydroxyproline (**57**) and (2*S*,3*R*)-N^{α}-Fmoc-3-Hydroxyproline (**58**) monomers were synthesized starting from commercially available *trans*-3-hydroxy-L-Proline. These monomers were then incorporated into both X and/or Y position of the collagen peptide $(X\text{-}Y\text{-}Gly)_n$ by solid phase peptide synthesis and the compatibility of 3*S* and 3*R* hydroxyl and aminoprolines with respect to the conformation of pyrrolidine ring were examined via CD studies.

Table 2: Thermal transition temperature of peptides in both aq. and ethylene glycol systems

Peptide Sequence		T_m (°C)	
		pH 7	EG:W
Ac-Phe-(Pro-3*(S)*Hyp-Gly)$_6$-NH_2	**P11**	58.3 °C	35.9 °C
Ac-Phe-(Pro-3*(S)*Amp-Gly)$_6$-NH_2	**P13**	-nd-	28.3 °C

CD spectra measured at various concentrations, temperatures and solvent conditions (Table **2**) suggest 3(*S*)Hyp in X position of collagen sequence $(Gly\text{-}X\text{-}Y)_n$ forming a stable coiled-coil kind of structure contrary to previous literature report,

while the 3(*S*)Amp in equivalent sequence behaves differently without any conformational transitions probably due to intrastrand H-bonding which does not allow it to form any coiled-coil structure.

References

1) (a) Chmielewski, J.; Geisler, I.; *Chem Biol Drug Des* **2009**; *73,* 39-45. (b) Fillon, Y. A.; Anderson, J. P.; Chmielewski, J. *J. Am. Chem. Soc.* **2005**, *127*, 11798-11803.

2) (a) Pozo-Rodríguez, A. D.; Pujals**,** S.; Delgado**,** D.; Solinís**,** M.A.; Gascón**,** A.R.; Giralt**,** E.; Pedraz**,** J. L. *Journal of Controlled Release* **2009**, *133*, 52–59. (b) Nakase, I.; Tadokoro, A.; T. Takeuchi, T.; Jones, A.T.; Futaki, S. *Biochem. Soc. Trans*. **2007**, *35*, 784-787.

3) (a) Yang-Gyun, K.; Hyun-Ju, P.; Kyeong, K. K.; Alexander, R.; Ky, L. *Nucleic Acids Research*, **2006**, *34*, 4937-4942. (b) Haley, J.; Kabiru, P.; Geng, Y.; *Mol. BioSyst.*, **2010**, *6*, 249–255.

4) Krista, M.T.; Naduthambi, D.; Tririya, G.; Zondlo*;* N. J. *Org.Lett.,* **2005**, *7*, 2397-2400.

5) (a) Taylor, C. M.; Hardré, R.; Edwards, P. J. B. *J. Org. Chem.* **2005,** *70*, 1306-1315. (b) Koskinen, A. M. P.; Helaja, J.; Kumpulainen, E. T. T.; Koivisto, J.; Mansikkamäki, H.; Rissanen, K. *J. Org. Chem.* **2005,** *70*, 6447-6453.

6) (a) Wolff III, J. S.; Logan, M. A.; Ogle, J. D. *Fed Proc.* **1966**, *25*, 862–863. (b) Wolff III, J. S.; Ogle, J. D.; Logan, M. A. J. Biol. Chem. **1966**, *241*, 1300-1307.

7) Hodges, J. A. Raines, R. T. *J. Am. Chem. Soc*. **2005**, *127*, 15923-15932.

8) Schumacher, M. A.; Mizuno, K. Bachinger, H. P. *J. Biol. Chem.* **2006**, *281*, 27566-27574. (c) Jenkins, C. L.; Bretscher, L. E.; Guzei. I. A.; Raines, R.T. *J. Am. Chem. Soc.* **2003**, *125*, 6422-6427.

9) (a) Babu, I.R.; Ganesh, K. N. *J. Am. Chem. Soc.*, **2001**, *123*, 2079-2080. (b) Umashankara, M.; Babu, I. R.; Ganesh, K. N. *Chem. Commun*, **2003**, 2606-2607.

Abbreviations & Symbols

A	Adenine
Ac	Acetyl
Ac_2O	Acetic anhydride
Ala	Alanine
Amp	(2*S*,4*R*)-aminoproline
amp	(2*S*,4*S*)-aminoproline
anhy.	Anhydrous
aq.	Aqueous
Arg	Arginine
Azp	(2*S*,4*R*)-azidoproline
azp	(2*S*,4*S*)-azidoproline
BSA	Bovine Serum Albumin
$(Boc)_2O$	Boc anhydride
bp	Base pairs
C	Cytosine
^oC	Degree Celsius
Calc.	Calculated
Cbz	Benzyloxycarbonyl
CD	Circular Dichroism / dichroic
COSY	Correlation Spectroscopy
$CuSO_4$	Copper Sulfate
D-	Dextro-
DCC	Dicyclohexylcarbodiimide
DCM	Dichloromethane
ΔG	Change in Gibb's free energy
ΔH	Change in enthalpy
ΔS	Change in entropy
Da	Dalton
DAPI	4', 6-diamidino-2-phenylindole
DIAD	Diisopropyl azodicarboxylate
DIPCDI	Diisopropyl carbodiimide
DIPEA/DIEA	N,N-Diisopropylethylamine
DMAP	N,N-Dimethyl-4-aminopyridine
DMF	N,N-dimethylformamide
DMSO	N,N-Dimethyl sulfoxide
DNA	2'-deoxyribonucleic acid
ds	Double stranded
E	*entgegen*
EB / EtBr	Ethidium bromide
EDTA	Ethylene diamine tetraacetic acid
eq.	equivalents
EG	Ethyleneglycol
ESI-MS	Electro Spray Ionisation Mass Spectrometry
Et	Ethyl
EtOAc	Ethyl acetate
FBS	Fetal bovine serum
FI	Fluorescence Intensity
Flp	(2*S*,4*R*)-fluoroproline

flp	(2*S*,4*S*)-fluoroproline
Fmoc	9-Fluorenylmethoxycarbonyl
g	Gram
G	Guanine
Gly	Glycine
hrs	Hours
Hz	Hertz
HBTU	2-(1H-Benzotriazole-1-yl)-1,1,3,3 tetramethyl-uronium-hexafluoro-phosphate
HMBC	Heteronuclear Multiple Bond Coherence
HOBt	N-Hydroxybenzotriazole
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence
Hyp	(2*S*,4*R*)-hydroxyproline (*trans*-4-hydroxy-L-proline)
hyp	(2*S*,4*S*)-hydroxyproline
3-Hyp	(2*S*,3S)-hydroxyproline (*trans*-3-hydroxy-L-proline)
3-hyp	(2*S*,4R)-hydroxyproline
IR	Infra Red
in situ	In the reaction mixture
ITC	Isothermal Titration Calorimetry
in vivo	Within the living
K	Kelvin / Kilo / Binding Constant
$K_{Z/E}$	equilibrium constant for *Z*/*E* isomerization
L-	Levo-
LC-MS	Liquid Chromatography-Mass Spectrometry
Leu	Leucine
Lys	Lysine
M	Molar
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation-Time of Flight
MBHA	4-Methyl benzhydryl amine
mg	milligram
MHz	Megahertz
min	minutes
µ	Micron
µg	Microgram
µl	Microliter
µm	Micrometer
µM	Micromolar
ma	major isomer
mA	miliampere
mi	minor isomer
MeCN	Acetonitrile
ml	milliliter
mM	millimolar
mmol	millimoles
m.p.	melting point
MS	Mass spectrometry
MsCl	Mesyl Chloride
MW	Molecular weight
N	Normal
nm	Nanometer
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Magnetic Overhauser Effect

N/P	Nitrogen/Phosphate ratio
ν	nu (frequency)
Obsv.	Observed
PBS	Phosphate-buffer-saline
PCR	Polymerase chain reaction
pDNA	Plasmid DNA
Pro	Proline
ppm	Parts per million
PNA	Peptide Nucleic Acid
p-TSA	*para*-toluenesulfonic acid
R	Rectus
R_f	Retention factor
RNA	Ribonucleic Acid
$R_{p/n}$	Ratio of positive to negative band intensities
RP	Reverse Phase (-HPLC)
rpm	Rotations per minute
rt	Room temperature
S	Sinister
S2	Schneider 2 *Drosophila melanogaster* cells
SPPS	Solid Phase Peptide Synthesis
ss	Single strand
T	Thymine
TAE	Tris Acetate Edetate
TAT	Transcellular activating transcription factor
t-Boc	*tert*-butyloxycarbonyl
TBE	Tris Borate Edetate
TEA/Et_3N	Triethylamine
TE	Tris Edetate
TEAA	Triethyl ammonium acetate
TFA	Trifluoroacetic acid
TFMSA	Trifluoromethane sulfonic acid
THF	Tetrahydrofuran
TIS	Triisopropylsilane
T_m	Melting temperature
TPP	Triphenylphosphine (PPh_3)
t_R	Retention time
Trp	Tryptophan
TsCl	Tosyl Chloride
UV-Vis	Ultraviolet-Visible
V	Volts
Val	Valine
v/v	volume to volume ratio
w/o	with and/or without
w/w	weight to weight ratio
Z	zusammen

List of Figures

FIGURE NO.	SECTION-I	Page
FIGURE 1	(A) Double helical structure of DNA (B) The phosphodiester linkage and the sugar phosphate backbone of DNA	2
FIGURE 2	Watson-Crick base pairing for A:T and G:C base pairing	2
FIGURE 3	Molecular models of B, A and Z form DNA	3
FIGURE 4	A short section of B-DNA with the principal modes of binding of ligands as intercalators, minor-groove binders, and major-groove binders.	5
FIGURE 5	Schematic representation of hydrogen bonding donor, acceptor sites available for molecular recognition in major and minor grooves of DNA.	6
FIGURE 6	Biogenic polyamines (A) Putrescine, (B) Cadavarine, (C) Spermine (D) Spermidine	8
FIGURE 7	Representative intercalating molecules	9
FIGURE 8	Structures of some DNA minor groove binding ligands	10
FIGURE 9	Schematic representation of the specific molecular interactions that occur between DNA and (A) Netropsin and (B) Hairpin polyamide	11
FIGURE 10	Amino acid specificity postulated in case of DNA and protein interactions	13
FIGURE 11	A few typical helix-turn-helix DNA-binding proteins: tryptophan repressor, lambda repressor fragment, lambda Cro fragment and CAP	15
FIGURE 12	DNA binding by a zinc finger protein	15
FIGURE 13	Leucine zipper region binding to a specific DNA sequence in the major groove of DNA.	15
FIGURE 14	Gene delivery using cationic co-polymers, peptides and lipids	19
FIGURE 15	(A) Typical UV absorption curves for equimolar base concentrations of mononucleotides, of single-stranded (random coil) oligonucleotide and of double-helical DNA; (B) An oligonucleotide melting curve and the derivative	22
FIGURE 16	Pathway of Linearly and Circularly polarized light	23
FIGURE 17	Circular dichroism spectra for three different conformations of DNA, a poly(dA)poly(dT) B-type duplex, a guanine quadruplex (tetraplex) and a single-stranded random coil	24
FIGURE 18	(A) A simplified Jablonski diagram of a fluorescence event (B) Generalized representation of the absorbance and emission spectra of a fluorophore	25
FIGURE 19	Mechanism of intercalated EtBr displacement from *ds*DNA on binding to protein	26
FIGURE 20	Relative position of different DNA forms of a plasmid on agarose gel, before and after enzyme digestion	27
FIGURE 21	Method of polyacrylamide gel electrophoresis	28
FIGURE 22	Mechanism of ionization in MALDI-TOF spectroscopy	29
FIGURE 23	*Aequorea Victoria* wtGFP chromophore formation	30
FIGURE 24	Expression of green fluorescent protein having its origin from transfected pRmHa3-GFP plasmid	31
FIGURE 25	(A) Guanidylated chimeric collagen peptide (B) Possible DNA-peptide interaction	32
FIGURE 26	Few representative structures of resin used in solid phase peptide synthesis	38
FIGURE 27	General protocol for SPPS via (I) Boc-chemistry (II) Fmoc chemistry	39
FIGURE 28	Peptides used in the present study. P1-P4 peptides are end capped, while peptide P5 is with free *N*-termini	42

FIGURE 29	Regioisomers of Texas Red dye	43
FIGURE 30	Fluorescent peptides (P8-P10) used in the present study. *N*-terminal free Peptides P6 and P7 are precursor for peptide P8 and P9 respectively.	44
FIGURE 31	(A) UV-absorption spectra of *TR-X*-labelled peptides (B) their corresponding fluorescence emission spectra at 615 nm	45
FIGURE 32	CD spectra at 10 °C at different concentrations of peptides (A) P2, (B) P4 and (C) The plot of $R_{p/n}$ values for P2.	48
FIGURE 33	Boltzman/Sigmoidal fit curve for molar ellipticity at 221.5 nm, from variable temperature CD spectra in different solvent conditions of pH 7.2, 12.0 and ethylene glycol–water mixture (3:1) for peptides (A)P2, (B)P4 at 0.03 mM concentration	50
FIGURE 34	CD spectra at various salt (NaCl) concentrations for peptide (A) P2 and (B) P4	51
FIGURE 35	UV-T_m curves of (A) AT rich *ds*DNA D1D2, (B) CG rich *ds*DNA D3D4	52
FIGURE 36	UV-melting profiles of (A) AT rich *ds*DNA D1D2 + P5, (C) CG rich *ds*DNA D3D4 + P5; first derivative UV-T_m of curves (B) AT rich *ds*DNA D1D2 + P5, (D) CG rich *ds*DNA D3D4 + P5	53
FIGURE 37	UV-melting profiles of (A) AT rich *ds*DNA D1D2 + P1, (C) CG rich *ds*DNA D1D2 + P2, (E) AT rich *ds*DNA D1D2 + P3, (G) CG rich *ds*DNA D1D2 + P4; Corresponding first derivative UV-T_m of curves in (B), (D), (F) and (H) respectively	54
FIGURE 38	UV-melting profiles of peptides (A) P1, P2, P3, P4 and P5, (B) Corresponding first derivative UV-T_m of curves	56
FIGURE 39	CD spectra of AT rich (D1D2) and CG rich (D3D4) duplex DNA (1 μM)	57
FIGURE 40	CD spectra of peptide P5 (100 μM) along with (A) AT rich *ds*DNA D1D2 (1 μM) and *ds*DNA+P5 complex, (B) CG rich *ds*DNA D3D4 (1 μM) and *ds*DNA+P5 complex at different ratios (1:50, 1:100)	57
FIGURE 41	CD spectra of ATrich *ds*DNA (1 μM) along with (A) peptide P1 (100 μM) and *p*DNA+P1 complex; (B) peptide P3 (100 μM) and *p*DNA+P3 complex; (C) peptide P2 (100 μM) and *p*DNA+P2 complex; (D) peptide P4 (100 μM) and *p*DNA+P4 complex at different ratios (1:50, 1:100)	58
FIGURE 42	(A) CD spectra of *p*DNA (5 μg/ml), peptide P5 (250 μg/ml) and *p*DNA-P5 (1:50 w/w) complex (B) zoomed view of the same CD spectra	59
FIGURE 43	CD spectra of *p*DNA (5 μg/ml) along with (A) peptide P1 (250 μg/ml) and *p*DNA+P1 (1:50 w/w) complex, (B) peptide P3 (250 μg/ml) and *p*DNA+P3 (1:50 w/w) complex	60
FIGURE 44	CD spectra of *p*DNA (5 μg/ml) along with (A) peptide P2 (250 μg/ml) and *p*DNA+P2 (1:50 w/w) complex, (B) peptide P4 (250 μg/ml) and *p*DNA+P4 (1:50 w/w) complex	61
FIGURE 45	Ethidium bromide displacement assay with peptides (A) P1, P3 and P5; (B) P2, P4 and P5 with AT rich *ds*DNA	62
FIGURE 46	Ethidium bromide displacement assay with peptides (A) P1, P3 and P5; (B) P2, P4 and P5 with CG rich *ds*DNA	62
FIGURE 47	Ethidium bromide exclusion assay with peptides (A) P1, P3 and P5; (B) P2, P4 and P5 with AT rich *ds*DNA	64
FIGURE 48	Ethidium bromide exclusion assay with peptides (A) P1, P3 and P5; (B) P2, P4 and P5 with CG rich *ds*DNA	64
FIGURE 49	(A) Ethidium bromide displacement assay and (B) Ethidium bromide exclusion assay of *p*DNA with peptides P1-P5	65
FIGURE 50	Polyacrylamide gel electrophoresis of AT rich *ds*DNA+peptide complexes	66
FIGURE 51	Polyacrylamide gel electrophoresis of CG rich *ds*DNA+peptide complexes	66
FIGURE 52	Time dependent binding of Peptides to *p*DNA (pRmHa3-GFP) on 1% Agarose Gel	67
FIGURE 53	Binding efficiency of peptides to *p*DNA at different ratios by agarose gel (1%) electrophoresis	68

FIGURE 54	DNA restriction site recognised by enzyme BamH1	69
FIGURE 55	Agarose gel electrophoresis after restriction enzyme (BamH1) digestion of pDNA (pRmHa3-GFP) and its complexes with peptides P2 (RG) and P4 (SG)	70
FIGURE 56	Live S2 cell line & various transfection levels (A) image of cells before transfection (B) high green fluorescence in cells after transfection (C) moderate fluorescence (D) weak fluorescence	71
FIGURE 57	Percentage cell transfection of pRmHa3-GFP on S2 cells along with different peptides using Qiagen Transfection kit reagents	72
FIGURE 58	Comparative transfection profile (with respect to only *p*DNA) for transfecting reagents in presence of various peptides	72
FIGURE 59	Pictures of fixed S2 cell line by light microscopy (A) DIC image (B)Green fluorescent cells image (C) Overlapping image of DIC and green fluorescent image	73
FIGURE 60	Confocal fluorescence microscopy analysis of S2 cells on treatment with fluorescent peptides. (I) DAPI stained image, (II) Fluorescent image of peptide staining, (III) Superimposed DAPI stained and fluorescent image of cells. (A-C) peptide P8, (D-F) peptide P9, (G-H) peptide P10	74
FIGURE 61	Distribution of fluorescent peptides inside S2 cells. Confocal image after incubation with 1 μM of (A) Peptide P8 (B) Peptide P9 (C) Peptide P10	75
FIGURE 62	Fluorescence microscopy analysis of S2 cells on transfection with range of peptides (I) DAPI stained image, (II) Green fluorescence image for GFP expression, (III) Red fluorescence image of peptide staining, (IV) Superimposed DAPI stained and fluorescent image of cells. (A-D) *p*DNA transfection images (E-H) transfection images with peptide P8, (I-L) transfection images with peptide P9, (M-P) transfection images with peptide P10	76
FIGURE 63	Isothermal titration of *p*DNA (pRmHa3-GFP) with the P4 (SG) peptide	78
FIGURE 64	Model illustrating (A) Cationic peptide complexation with GFP coding DNA, (B) Mechanistic pathway of peptide-DNA complex during transfection of eukaryotic cells, (C) Expression of fluorescent peptide and GFP in the cells.	86

SECTION-II

FIGURE NO.		Page
FIGURE 1	(A) Illustration of peptide plane and the torsion angles therein, and (B) Ramachandran plot for variety of peptide structures	159
FIGURE 2	(A) Resonance phenomenon in amide bond, (B) *cis-trans* isomerization in non-proline peptide fragments, (C) *cis-trans* conformations in proline fragments	160
FIGURE 3	Main chain torsion angles and endocyclic torsion angles in imino acid proline (Pro)	161
FIGURE 4	*cis -trans* (*E-Z*) isomers of the peptide bond in proline and substituted prolines	161
FIGURE 5	Ring-puckers of the pyrrolidine ring in proline and substituted prolines.	163
FIGURE 6	Ring conformations of 4-substituted proline residues.	164
FIGURE 7	Newman and saw-horse projections depicting the *gauche* effect and the resulting pucker preferences in the prolines with electronegative (A) 4*R*-substitutent, (B) 4*S*-substituent	165
FIGURE 8	Dicisive structural determinants in Xaa-Pro prolyl-peptide isomerization process	169
FIGURE 9	Methyl doublets of alanine in the ^{1}H NMR of thiopeptide at two different pH	169
FIGURE 10	^{1}H-NMR spectra of dipeptide 19 (500 MHz) in D_2O (pD $\leq$ 2); Inset: appearance of compound 19a along with dipeptide 19.	172
FIGURE 11	^{1}H-NMR spectra of compound 19a (500 MHz) in D_2O	173
FIGURE 12	^{1}H-NMR spectra of dipeptide 21 (500 MHz) in $CDCl_3$	174
FIGURE 13	ORTEP diagram of crystal structures of N-Boc protected dipeptides 18, 22, 26, 32 and 33	178

obtained by single crystal X-ray diffraction.

FIGURE 14	Main chain torsion angles and endocyclic torsion angles in Ala-Pro based model dipeptides;	179
FIGURE 15	Correlations expected with respect to 2D ^{1}H-^{1}H COSY NMR spectra of Dipeptide 19 at pD 2.2.	180
FIGURE 16	2D ^{1}H-^{1}H COSY NMR spectra of pyrrolidine ring protons in compound 19 (500 MHz) in D_2O, pD 2.2	181
FIGURE 17	Expanded view of selected portion from 2D ^{1}H-^{1}H COSY NMR spectra of pyrrolidine ring protons in compound 19 (500 MHz) in D_2O at pD 2.	182
FIGURE 18	2D ^{1}H-^{1}H COSY NMR spectra of pyrrolidine ring protons in compound 19 (500 MHz) in D_2O at pD 12.0	183
FIGURE 19	2D ^{1}H-^{1}H COSY NMR spectra of pyrrolidine ring protons in compound 36 (500 MHz) in D_2O at pD 12.0	184
FIGURE 20	Selected view of 2D ^{1}H-^{1}H COSY NMR spectra of 38 at pD 12.0	185
FIGURE 21	2D ^{1}H-^{1}H COSY NMR spectra of pyrrolidine ring protons in compound 38 (500 MHz) in D_2O at pD 12.0	186
FIGURE 22	Expected NOE interaction in case of Ala-Pro dipeptide 19, as (A) *trans* & (B) *cis* isomers	187
FIGURE 23	2D ^{1}H-^{1}H NOESY NMR spectra of pyrrolidine ring protons in compound 19 (500 MHz) in D_2O at pD 12.0	188
FIGURE 24	2D ^{1}H-^{1}H NOESY NMR spectra of pyrrolidine ring protons in compound 36 (500 MHz) in D_2O at pD 12.0	189
FIGURE 25	2D ^{1}H-^{1}H NOESY NMR spectra of pyrrolidine ring protons in compound 38 (500 MHz) in D_2O at pD 12.0	189
FIGURE 26	Expected NOE interaction in case of Ala-amp dipeptide 38, as (A) *trans* & (B) *cis* isomers	190
FIGURE 27	Stacked plot of 1D ^{1}H-NMR spectra of dipeptide 19 (500 MHz) in D_2O at various pD/pH	194
FIGURE 28	Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide 23 (500 MHz) in D_2O at pD 2.2 and 12.0	195
FIGURE 29	Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide 28 (500 MHz) in D_2O at pD 2.2 and 12.0	195
FIGURE 30	Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide 31 (500 MHz) in D_2O at pD 2.2 and 12.0	195
FIGURE 31	Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide 34 (500 MHz) in D_2O at pD 2.2 and 12.0	196
FIGURE 32	Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide 36 (500 MHz) in D_2O at pD 2.2 and 12.0	196
FIGURE 33	Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide 38 (500 MHz) in D_2O at pD 2.2 and 12.0	196
FIGURE 34	Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide 41 and 44 (500 MHz) in D_2O at pD 2.2 and 12.0	197
FIGURE 35	Variation plot of *cis*-isomers with pH for the dipeptides; (A) 4-Hydoxy dipeptides, (B) 4-Azido dipeptides, (C) 4-amino dipeptides	198
FIGURE 36	pK_a titration with 0.5 M solution of NaOH for dipeptides (A) Ala-Amp 36, (B) Ala-amp 38, (C) Ac-NH-Ala-Amp 41, (D) Ac-NH-Ala-amp 44 and (E) Ala-Pro 19	199
FIGURE 37	CD spectra of dipeptides; (A) Comparative plot of all dipeptides at their original pH values ($\leq$ 2) and (B) Ala-Pro 19, (5 mM, H_2O) at various pH values	201
FIGURE 38	CD spectra of dipeptides (5 mM, H_2O) at various pH values; (C) Ala-Hyp 23, (D) Ala-hyp 28, (E) Ala-Azp 31 (F) and Ala-azp 34	201
FIGURE 39	CD spectra of dipeptides (5 mM, H_2O) at various pH values; (C) Ala-Hyp 23, (D) Ala-hyp 28, (E) Ala-Azp 31 (F) Ala-azp 34, (G) Ala-Amp 36 (H) Ala-amp 38, (I) Ac-NH-Ala-Amp 41	202

and (J) Ac-NH-Ala-amp 44

FIGURE 40 Proline dipeptides in the (A) *trans* and (B) *cis* conformations at non-ionised & ionised state respectively. 204

FIGURE NO.	SECTION-III	Page

FIGURE 1 Classification of collagen as a structure 299

FIGURE 2 Extracellular events in the biosynthesis of fibrillar collagens 300

FIGURE 3 (A) Schematic view of Rich & Crick model with one hydrogen bond per trimer repeat (Gly-Pro-Hyp) repeat as an example (B) Crystal structure of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ 302

FIGURE 4 Collageneous peptides with (A) Hydrogen bonding inbetween strands of $(Gly\text{-}Hyp\text{-}Pro)_n$ (B) Ring puckering preferences in 4(*R/S*)-Hyp 304

FIGURE 5 (A) CD spectrum at 10 °C and (B) triple-helix ⇔ coil transition curve for $(Pro\text{-}Hyp\text{-}Gly)_{10}$ in water. 306

FIGURE 6 Various proline surrogates that have been incorporated in to collagen sequences. (A) Thiozilidine (Thz), (B) *trans*-3-Methylproline (MePro), (C) *trans*-4-Fluoroproline (Flp), (D) *cis*-4-Fluroproline (flp), (E) *trans*-3-Hydroxyproline, (F) L-piperidine-2-carboxylic acid (Pipec), (G) Azitidine (Aze). 306

FIGURE 7 Putative interstrand hydrogen bonds in triple-helical $(3\text{-}Hyp\text{-}4\text{-}Hyp\text{-}Gly)_n$ 309

FIGURE 8 Schematic view of both α and β peptide linkages in 3-aminoproline. 309

FIGURE 9 Expected pyrrolidine conformation for 3-aminoproline peptides 310

FIGURE 10 End capped peptides used in the present study. P11-P12 peptides on Rink Amide resin *via* Fmoc-chemistry, P13 and P14 peptides synthesized *via t*-Boc chemistry on MBHA resin 318

FIGURE 11 HPLC Profiles of peptides(A) P11, (B) P12, (C) P13 and (D) P14 319-320

FIGURE 12 MALDI-TOF spectra of peptides (A) P11, (B) P12, (C) P13 and (D) P14 320-321

FIGURE 13 CD spectra at 10 °C at concentrations from 0.05 mM-0.25 mM in steps of 0.05 mM (20 mM phosphate buffer pH 7.2, 10 mM NaCl) of peptides (A) P11, λ_{max} = 225 nm, λ_{min} = 201 nm, isobestic point at 221 nm; (C) P13, , λ_{max} = 223.5 nm, λ_{min} = 201 nm, isobestic point at 219 nm; The plot of $R_{p/n}$ values deduced from these spectra against the concentration of (B) P11 and (D) P13. 323

FIGURE 14 CD spectra at 10 °C at concentrations at 0.2 mM & 0.5 mM (20 mM phosphate buffer pH 7.2, 10 mM NaCl) of peptides (A) P12, λ_{max} = -nd-, λ_{min} = ~200 nm, (B) P14, λ_{max} = -nd-, λ_{min} = 205 nm 324

FIGURE 15 CD spectra at (A) 5-75 °C, (B) Boltzman/Sigmoidal fit curve for molar ellipticity at 224 nm (C) T_m curve of 0.2 mM solutions of peptide P11 in 20 mM phosphate buffer pH 7.2, 10 mM NaCl 325

FIGURE 16 CD spectra at (A) 5-75 °C, (B) Boltzman/Sigmoidal fit curve for molar ellipticity at 224 nm (C) T_m curve of 0.2 mM solutions of peptide P13 in 20 mM phosphate buffer pH 7.2, 10 mM NaCl 325

FIGURE 17 CD spectra at (A) 10 °C, (B) 5-90 °C, (C) Boltzman/Sigmoidal fit curve for molar ellipticity at 224 nm (D) T_m curve of 0.2 mM solutions of peptide P11 in 3:1 (v/v) EG:Water mixture. 326

FIGURE 18 CD spectra at (A) 10 °C, (B) 5-90 °C, (C) Boltzman/Sigmoidal fit curve for molar ellipticity at 223 nm (D) T_m curve of 0.2 mM solutions of peptide P13 in 3:1 (v/v) EG:Water mixture 327

List of Tables

TABLE NO.	SECTION-I	Page
Table 1	Average helix parameters for the major DNA conformations	4
Table 2	Summary of DNA transfection methods.	17
Table 3	Common peptides used in DNA delivery	20
Table 4	HPLC retention time and MALDI-TOF mass spectral analysis of the synthesized peptides	45
Table 5	HPLC retention time and MALDI-TOF mass spectral analysis of the DNA oligonucleotides	46
Table 6	Thermal transition temperature of peptides in both aq. and ethylene glycol systems	50
Table 7	UV-T_m data for interaction of peptide P5 (SAP) with *ds*DNA.	52
Table 8	UV-T_m data for interaction of cationic peptides with *ds*DNA (AT-rich) sequence	55
Table 9	Difference in UV-T_m values between cationic peptides -*ds*DNA (AT-rich) complex and only *ds*DNA.	55
Table 10	UV-T_m data for interaction of peptide P5 (SAP) with *ds*DNA.	63
Table 11	ITC thermodynamic parameters describing the interaction of peptide with DNA	78
Table 12	Details of restriction enzyme digestion protocol	111
Table 13	Details of fluorescent peptide permeation and transfection protocol	115

TABLE NO.	SECTION-II	Page
Table 1	Percententage of *cis* isomers found in various peptides and with respect to their neighbouring amino acids.	162
Table 2	Main-chain torsion angles of Ala-Pro based model dipeptides	179
Table 3.A	Chemical shift and splitting pattern of ^{1}H-NMR signals for dipeptide 19 at acidic & alkaline pH	191
Table 3.B	Chemical shift and splitting pattern of ^{1}H-NMR signals for dipeptide 36 from acidic to alkaline pH range.	192
Table 3.C	Chemical shift and splitting pattern of ^{1}H-NMR signals for dipeptide 38 from acidic to alkaline pH range.	193
Table 4	General trend of *trans* isomers at various pH	197
Table 5	pK_a values obtained for titrated dipeptides	200
Table 6	ORTEP diagram and Crystal data with structure refinement for compound 18	293
Table 7	ORTEP diagram and Crystal data with structure refinement for compound 22	294
Table 8	ORTEP diagram and Crystal data with structure refinement for compound 26	295
Table 9	ORTEP diagram and Crystal data with structure refinement for compound 32	296
Table 10	ORTEP diagram and Crystal data with structure refinement for compound 33	297

TABLE NO.	SECTION-III	Page
Table 1	List of examples of collagen peptides with high resolution structures solved by X-Ray crystallography	303
Table 2	HPLC retention time and MALDI-TOF mass spectral analysis of the synthesized peptides	319
Table 3	Thermal transition temperature of peptides in both aq. and ethylene glycol systems	327

List of Schemes

SCHEME NO.	SECTION-I	Page
Scheme 1.1	(A) Non-viable synthetic route with use of N$^{\alpha}$-Fmoc and N$^{\alpha}$-Boc protected monomers, (B) Proposed synthetic route for synthesis of required peptides with N$^{\alpha}$-Boc and N$^{\alpha}$-Fmoc protected monomers	34
Scheme 1.2	Synthesis of (2*S*,4*R*)-N^{1}-(t-butoxycarbonyl)-4-(9-fluorenylmethyloxycarbonylamino) proline	36
Scheme 1.3	Synthesis of (2*S*,4*S*)-N^{1}-(t-butoxycarbonyl)-4-(9-fluorenylmethyloxycarbonylamino) proline	36
Scheme 1.4	Solid phase synthesis of peptides using *t*-Boc chemistry	41
Scheme 1.5	Solid phase synthesis of fluorescent peptides and subsequent guanidylation	43

SCHEME NO.	SECTION-II	Page
Scheme 2.1	Synthesis of (*S*)-N^{1}-(t-butoxycarbonyl)-alanine 14	171
Scheme 2.2	Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-proline dipeptide methylester 17	171
Scheme 2.3	Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-hydroxyproline dipeptide methylester 23	173
Scheme 2.4	Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-hydroxyproline dipeptide methylester 28	175
Scheme 2.5	Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-azidoproline dipeptide methylester 31	175
Scheme 2.6	Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-azidoproline dipeptide methylester 34	176
Scheme 2.7	Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-aminoproline dipeptide methylester 36	176
Scheme 2.8	Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-aminoproline dipeptide methylester 38	176
Scheme 2.9	Synthesis of (*S*)-N-acetyl-alanine-(2*S*,4*R*)-4-aminoproline dipeptide 41	177
Scheme 2.10	Synthesis of (*S*)-N-acetyl-alanine-(2*S*,4*S*)-4-aminoproline dipeptide 44	177

SCHEME NO.	SECTION-III	Page
Scheme 3.1	Synthesis of (2*S*,3*S*)-*N*$^{\alpha}$-(t-butoxycarbonyl)-3-(9-fluorenylmethyloxycarbonylamino) proline	312
Scheme 3.2	Synthesis of (2*S*,3*R*)-*N*$^{\alpha}$-(t-butoxycarbonyl)-3-(9-fluorenylmethyloxycarbonylamino) proline	313
Scheme 3.3	(A) Synthesis of (2*S*,3*S*)-*N*$^{\alpha}$-(9-fluorenylmethyloxycarbonylamino)-3-hydroxyproline, (B) Synthesis of (2*S*,3*R*)-*N*$^{\alpha}$-(9-fluorenylmethyloxycarbonylamino)-3-hydroxyproline	314
Scheme 3.4	Schematic representation of solid phase synthesis; (A) *t*-Boc strategy used for synthesis of P13 and P14 peptides, (B) Fmoc- strategy used for peptide P11 and P12	316
Scheme 3.5	(A) Reaction of ninhydrin (trioxohydrindene hydrate) with the amino group of a bound residue generates the Schiff's base. (B) Hydrolysis generates the aldehyde and another amine. (C) Amine reacts with a second molecule of ninhydrin to give an equilibrium mixture of the anion depicted and its tetraoxo form with a maximum of absorbance at 570 nm	317
Scheme 3.6	Proposed mechanism for the reaction between free amine group of peptide on solid phase and p-chloranil to produce charge-transfer complex	317

Prologue

Proline derivatives having 5-membered pyrrolidine ring with strong conformational biases are potent tools to elucidate the roles of secondary structure like turn geometry and *cis:trans* isomerization which indirectly affect the protein structure and function. Hence, stereochemically well-defined 3- and 4-substituted prolines can play an important role in modulating the conformation of derived peptides thereby leading to a detailed understanding of structure-function relationship.

Chimeric amino collagen peptides $[Pro(X)\text{-}Amp(Y)\text{-}Gly]_6$ & $[amp(X)\text{-}Pro(Y)\text{-}Gly]_6$ show pH dependent triple-helices & higher stability compared to natural collagen peptides, where Amp and amp stand for 4(*R/S*)-Aminoproline. Replacement of amino groups by guanidino function ($pK_a \approx 13$) these peptides should ensure cationic substituent even at neutral pH. In this context, **Section I** examines syntheses and biophysical studies of 4-guanidino proline collagen like peptides.

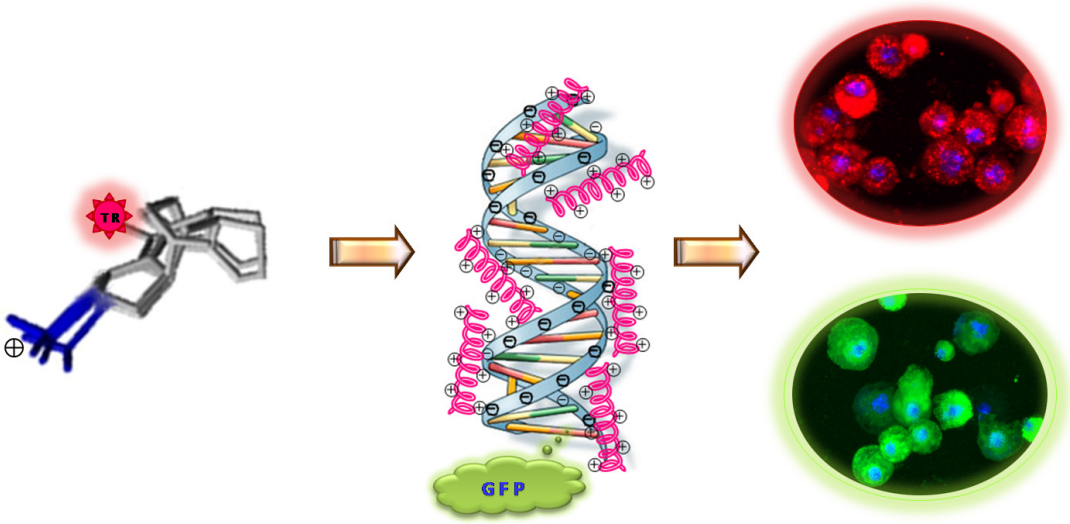
The pH dependent 4-amino collagen peptides also insist that in case of 4-aminoproline there is a possibility of different types of ring pucker and *trans:cis* ratio in its protonated and non-protonated forms. This may result in the formation of pH regulated helical forms, because the protonated and free amino group have different electronegativity. Study of pepdidyl-prolyl bond isomerization is also intimately linked to the understanding of the folding and stability of the collagen. In order to understand the relative effects of protonation/non-protonation, stereo-chemical and electronic influences, **Section II** focuses on effect of various sustituents at 4-position of proline which is a part of the dipeptide system showing *cis:trans* isomerisation.

Further synthesis of conformationally constrained peptides are emerging as useful means for developing different mimics of existing secondary structure of peptides and/or proteins. **Section III** investigates the impact of amino group in the C3/C^{β}-position of pyrolidine ring as a part of the collagen sequence, since 4(*R/S*)-aminoprolines have proved to be better surrogates of 4-hydroxyproline in both Xaa and Yaa position of collagen sequence $[(Gly\text{-}X\text{-}Y)_6]$ type by showing higher triple-helical propencity.

The work presented in this thesis explores these aspects and thus, the work in the thesis entitled "**Synthesis & Comparative Biophysical Studies of Peptides Derived from 3 and 4-Substituted Prolines**'' is divided into three sections having interrelated rationale.

Section I

Synthesis and Biophysical Studies on DNA Binding and Cell Penetrating Ability of Chimeric 4-Amino & Guanidino-proline Collageneous Peptides



This study involves characterization of interaction of cationic peptides with DNA. The role of 4-substituted aminoprolines and their guanidium derivatives in molecular recognition of peptides leading to DNA complexation is elaborated in this section. Improvements in pharmacokinetic properties for better cellular uptake have been targeted through designed cationic collagen synthesis. The effects on hybridization properties of DNA-peptide complexes have been investigated through different biophysical and biological studies. It also portrays recent literature and advancements in the area of cell penetrating peptides emphasizing on utility of proline based cationic peptides. Important properties of the CPP namely, the DNA-condensing issues related to its mechanisms and cellular uptake etc., have been discussed in the light of its structural features.

1.1 Introduction

Intracellular drug delivery is the focal point of present day's drug development. Efficient possibilities to penetrate cell membrane and deliver molecules to intracellular targets create vast number of new therapeutic opportunities including gene therapy. However, cell membranes are impermeable to most molecules not actively imported by living cells, including practically all macromolecules such as peptides, proteins, DNA and even some small molecules whose physiochemical properties prevent passive membrane diffusion. Delivery agents with cell penetrating ability present a new dimension to drug research and nucleic acid therapeutics. Study and development of such systems can help increasing the efficiency of many treatment protocols with intracellular delivery of various drugs, including DNA.

The present work focuses DNA binding and cell penetrating ability of peptides leading to their use in gene therapy *i.e.* DNA incorporation into the cell genome. The following sections give an introduction to structure of DNA and its molecular recognition properties.

1.2 DNA – A structural overview

Deoxyribonucleic acid (DNA) is among the largest molecules in the cell. Being the blueprint of genetic information, understanding its structure and activity assumes extreme importance. DNA in living cells can interact with broad classes of drugs, carcinogens, mutagens, dyes and large molecules such as proteins and peptides. Owing to DNA's central role in biological replication and protein synthesis, modulation of its structure or activity by such interactions greatly affects the cell metabolism. Moreover many such interactions lead to condensation of long strands of DNA into compact, ordered structures, which is an important biological process for gene protection, storage and replication. Thus, understanding the process of DNA-ligand interactions is of vital importance. Such studies not only throw light on the mechanism of action of many natural products that attack DNA, but also aid as probes to elucidate its structure-function relationship in addition to revealing common principles underlying gene therapy and development of non-viral gene delivery vectors.

The primary structure consists of two hydrophilic sugar–phosphate backbones that lie on the outside of the molecule, and the purines and pyrimidines lie on the inside of the molecule (Figure **1**).

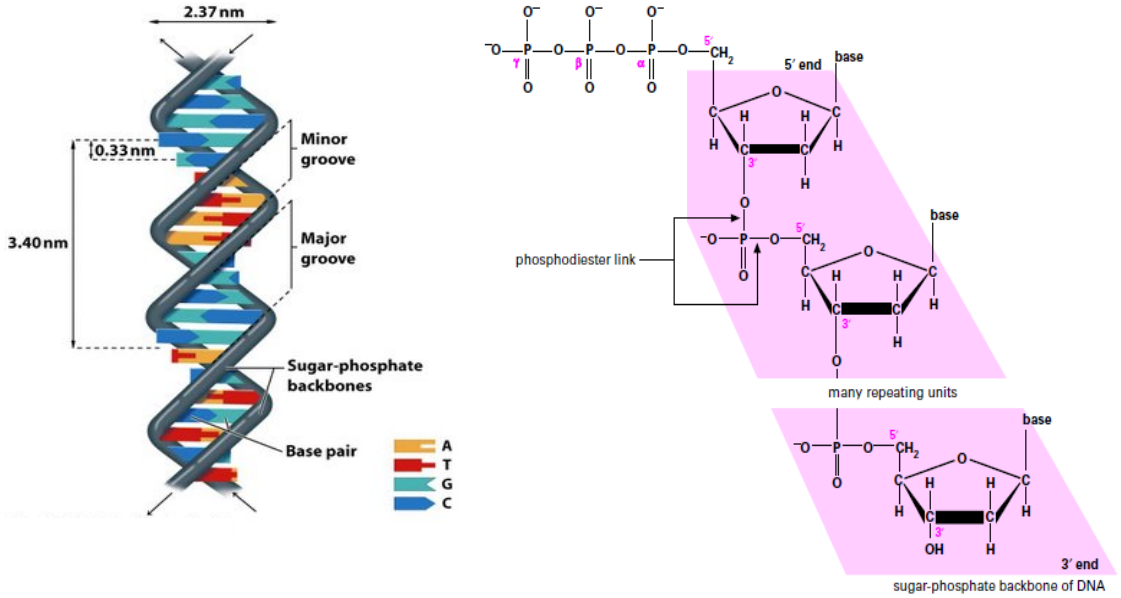


FIGURE 1: (**A**) Double helical structure of DNA (**B**) The phosphodiester linkage and the sugar phosphate backbone of DNA

DNA is a polynucleotide made up of four different nucleotide units A, G (purines) and C, T (pyrimidines). Accurate DNA replication is most necessary for cells survival and at a molecular level, this is achieved through Watson-Crick hydrogen bonding (Figure **2**).[1] This mutual recognition of bases by use of hydrogen bonds primarily ensures the fidelity of DNA transcription and translation. There are two hydrogen bonds in an A:T base pair and three in C:G base pair involved in Watson-Crick pairing. This results in an isomorphous geometry whereby the four base combinations A:T, T:A, C:G and G:C can all be built into same geometrical framework. Further, it leads to an antiparallel double helical structure for DNA in which the two strands held by hydrogen bonds, and are twisted around each other. The topological consequence of this is the formation of helical major and minor grooves in B form of DNA - the most prevalent conformation of DNA in cells. Most specific interactions of DNA with other molecules such as proteins, drugs, water and metal ions take place in these grooves.

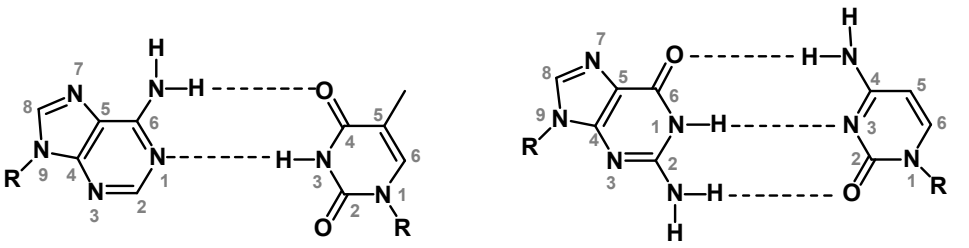


FIGURE 2: Watson-Crick base pairing for A:T and G:C base pairing

Depending on the base sequences[2] and environment, DNA duplex can exist in three major conformations: A, B and Z-DNA (Figure **3**).[3] A-DNA and B-DNA are the major regular DNA secondary structures with right-handed double helices and Watson-Crick base pairs. At low humidity and high salt, the favored form is the highly crystalline A-DNA while at high humidity and low salt, the dominant structure is B-DNA.

A-DNA has 11 residues per turn; the bases are tilted by 20° to enhance stacking and lay 4.5 Å away from the helix axis (Figure **3.a**). This gives a fairly stiff right handed helix which shows little sequence-dependent variation in structure. The furanose ring has a *C*3´-*endo* pucker and the glycoside is in the *anti*-conformation. The major groove is deep and narrow while the minor grove is broad and shallow.

B-DNA has ten bases per turn with little tilting of the bases (Figure **3.b**). The sugars have the *C*2´-*endo* pucker and all the glycosides have the *anti*-conformation. The B-form structure is sufficiently flexible to permit a conformational response in the backbone to particular local base sequences. The wide major groove and narrow minor groove are both of moderate depth and are well solvated by water molecules.

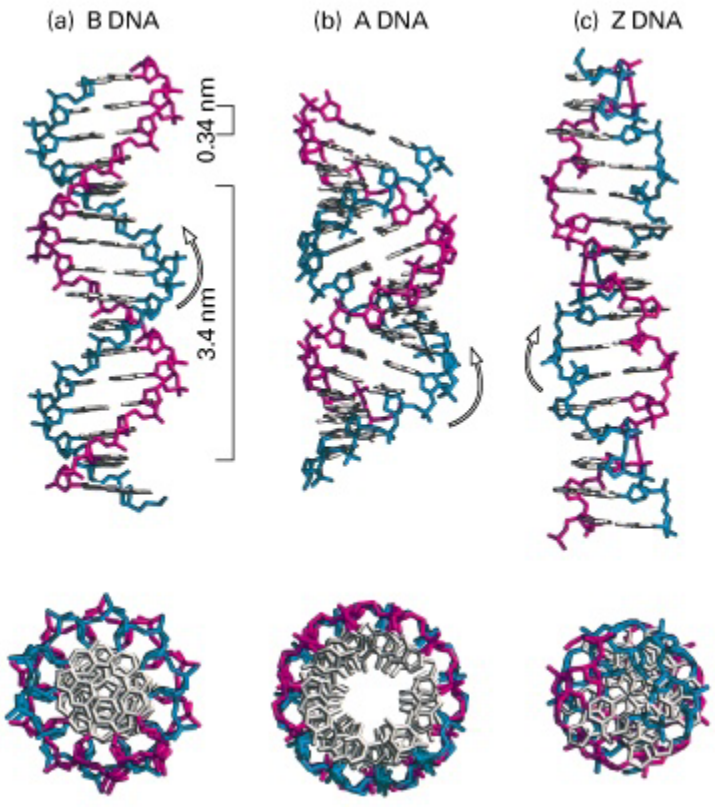


FIGURE 3: Molecular models of B, A and Z form DNA (**Top**), views along the helical axis (**Bottom**); The arrows indicate direction of twist along the helix

Z-DNA is a left-handed double-helical structure stabilized by high concentrations of $MgCl_2$, NaCl or ethanol. It is most favoured for alternating G-C sequences. The left-handedness of this antiparallel duplex is the result of a switch in the glycosidic torsion angle from the regular *anti*- to the unusual *syn*- conformation for one nucleoside in each base-pair, which is normally the purine. This *anti-syn* feature alternates regularly along the DNA backbone, causing the phosphorus atoms to follow a zig-zag course (Figure **3.c**), while the sugar pucker alternates from *C2´-endo* for the *anti*-residues to *C3´-endo* for the *syn*-nucleosides. The grooves are shallow and it is difficult to distinguish the major and minor gooves.

Table 1: Average helix parameters for the major DNA conformations

	A form	*B form*	*Z form*
Helical sense	Right handed	Right handed	Left handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11	10.5	12
Helix rise per base pair	2.6 Å	3.4 Å	3.7 Å
Base tilt normal to the helix axis	20°	6°	7°
Sugar pucker conformation	C-3′ endo	C-2′ endo	C-2′ endo for pyrimidines; C-3′ endo for purines
Glycosyl bond conformation	Anti	Anti	Anti for pyrimidines; syn for purines

The above mentioned Table **1** summarizes the major helix parameters observed for the three different forms of DNA.

1.3 Plasmid DNA

Plasmid is extranuclear, covalenty closed circular, self replicating, double stranded form of DNA which can exist independently in the cells. Plasmids contains genes ranging between three to hundreds. Genes present in plasmid do not play any role in normal life cycle or survival of a cell. Functions of these genes are additional to the main function of chromosomal genes. Plasmids are mainly seen in bacteria and few eukaryotic organism like yeast cells. The term *plasmid* was first introduced by the American molecular biologist Joshua Lederberg in 1952.

Plasmids are important in certain bacteria since plasmids code for proteins, especially enzymes, which can confer resistance to antibiotics. Certain plasmids are able to insert themselves into the chromosomes particularly in regions where there is

a common sequence of nucleotides. Hence, they are used in recombinant DNA technology and research as means for transferring genes between cells or as cloning vectors. In this section plasmid DNA has been abbreviated as *p*DNA at many places for the sake of simplicity.

1.4 DNA and molecular recognition

Most biological processes are reliant upon molecular recognition and the reversible interactions of one set of molecules with another. In particular DNA at many places is dependent on numerous protein–DNA interactions as well as, in a more subtle way, interactions with ions and water. There are three principal binding modes in which low molecular weight ligands can interact with double-stranded DNA *viz.* **Outside-edge binding**, **Intercalation** and **Groove binding** (Figure **4**) based on three sites of molecular recognition presented by DNA. The phosphate backbone and the two grooves, namely major and minor grooves are the locations where the structure-specific recognitions mostly happen.[4]

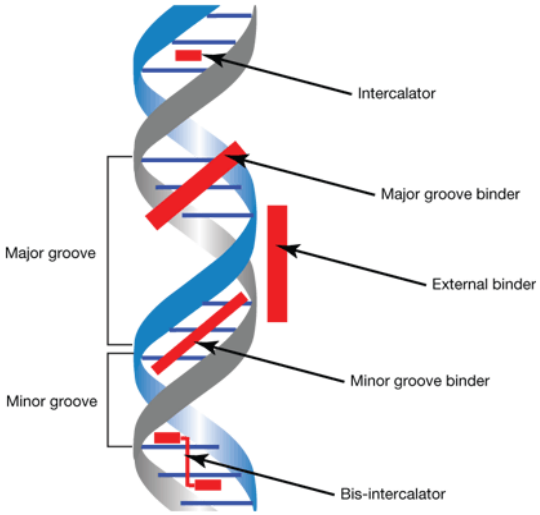


FIGURE 4: A short section of B-DNA with the principal modes of binding of ligands as intercalators, minor-groove binders, and major-groove binders.

The recognition of the phosphate backbone in DNA is predominantly by electrostatic interactions, while, hydrogen-bonding interactions dominate the major/minor groove recognition.

Moreover the self-recognition of bases through hydrogen bonding is also one of the most straightforward principles of molecular recognition in complex systems like DNA. Hydrogen bonds are formed between the major *amino-keto* tautomers of the bases. The *N-H* groups of the bases are good hydrogen-bond donors, while the sp^2-hybridized electron pairs on the oxygens of the base, the *C=O* groups and on the

ring nitrogens are much better hydrogen bond acceptors than the oxygens of either the phosphate or the pentose. The donor-acceptor hydrogen bonds so formed are largely ionic in character.

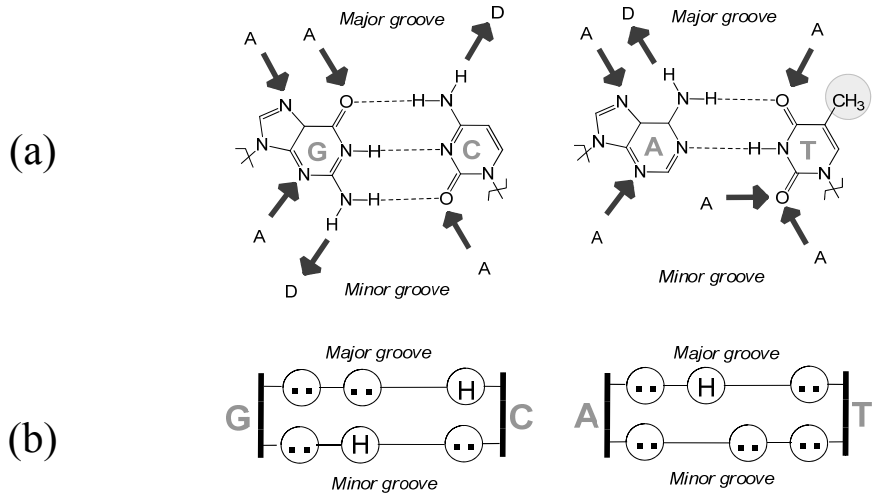


FIGURE 5: Schematic representation of hydrogen bonding donor, acceptor sites available for molecular recognition in major and minor grooves of DNA.

The most effective sequence discrimination by DNA binding agents is possible on the floor of the major groove where four recognition sites are present for A:T base pairs and three sites for G:C base pairs (Figure **5.a**). These are adenine-N7, thymine-O4 (as hydrogen bond acceptors) and the adenine amino group (N6) (as hydrogen bond donor) the thymine methyl group (van der Waals interaction) for A:T. Similarly, the recognition sites available at G:C base pairs are two hydrogen acceptors (gaunine-N7 and guanine O6) and one hydrogen donor (the cytosine amino group at C4). Thus the DNA sequence can be unambiguously read digitally from the major groove as a "two dimensional grid" of hydrogen donor/acceptor and van der Waals sites. Considerably less sequence discrimination is possible in the major groove of the DNA (Figure **5.b**) with the G:C base pair presenting two hydrogen-bond acceptor sites (gaunine-N3 and cytosine-O2) and one hydrogen bond donor (gaunine-C2-NH_2) site, while, the A:T base pair presents only two hydrogen accepting sites.

Generally it has been seen that the compounds that have the potential to be clinically useful are normally either intercalators or groove-binders. However, outside-edge, electrostatic interactions are also important, not least because the association of positively charged counterions with the DNA polyanion has a large effect on DNA conformation and stability.

1.4.1 Non-specific outside-edge interactions

1.4.1a Condensation type interactions

Nucleic acids are highly charged polyelectrolytes whose anionic phosphate groups strongly affect their structure and interactions with other molecules like charged drugs and proteins. The initial interaction of simple counter-ions as those of the alkali metals with nucleic acids is referred to as condensation because the cations generally form a cloud around the charge density of the nucleic acid and are not bound at specific sites.

The nucleic acid polyelectrolytes, in which the charges are closely spaced, possess enormous amount of repulsive electrostatic interactions and to achieve stability these molecules must uptake counterions from solution. This polyelectrolyte association with ions called ***counterion condensation***, which contributes unfavorably to the entropy because of the loss in conformational freedom of both the polymer and the counterions. The unfavorable entropy term is more than outweighed by the numerous coulombic attractive forces in the folded polymer (such as the DNA double helix). Counterions condense until the charge density is reduced to the stable level of approximately one charge per 7Å.

1.4.1b Electrostatic interactions

This involves ligand binding (*e.g.* Na^+, Mg^{2+} or polyamines) to the outside of the helix through non-specific, primarily **electrostatic interactions** with the sugar–phosphate backbone. The negative charges present on DNA sugar-phosphate backbone or the exterior of the helix provide a means of interaction for ligands with positive charges resulting from coulombic attractive forces between them. Several biogenic polyamines such as *putrescine, cadavarine, spermine* and *spermidine* (Figure **6**) are among the well-characterized molecules that bind DNA by electrostatic interactions. These amines are ubiquitously present in plant and animal cells and are required for the normal growth of all cells.[5] They are thought to play multiple roles in cellular function like DNA packaging into chromatin, through neutralization of negative charges on the phosphates, analogous to histones. They also have an important role in cell proliferation as direct binding of polyamines to DNA results modulation of protein–DNA interactions.[6] In addition, polyamines have some regulatory role in the synthesis of biological macromolecules and enhance the activity

of several DNA-binding proteins such as DNA polymerase α.[7]

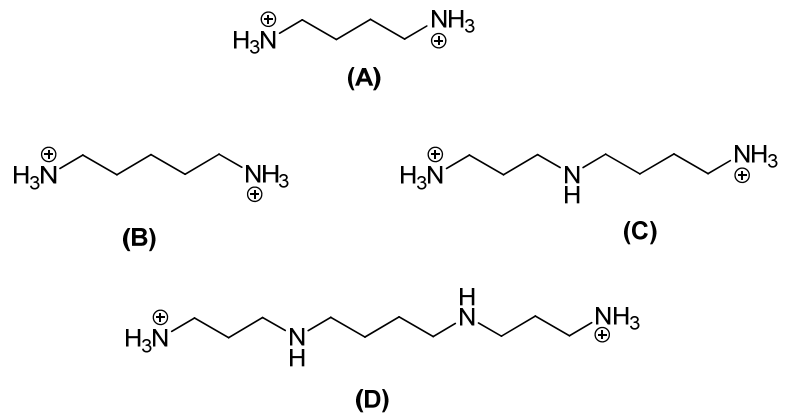


FIGURE 6: Biogenic polyamines (**A**) Putrescine, (**B**) Cadavarine, (**C**) Spermine (**D**) Spermidine

The structural effects of polyamine-DNA interaction to some extent can be explained in terms of non-specific electrostatic binding based on the counter-ion condensation theory.[8] A significant degree of polyamine binding interactions with base and/sequence specificity has been shown from X-ray structures and by molecular modeling.[9] Base/sequence specificity of *spermine* binding was observed in the crystal structure of the self-complementary octamer d(GTGTACAC):*spermine* complex.[10] *Spermine* was bound in the deep groove of the DNA and was located on the molecular dyad axis. Besides hydrogen bonding, the complex showed significant number of hydrophobic interactions between the methylene groups of the *spermine* and the base atoms on the floor of the major groove.

The co-crystal of d(CGCGCG) and *spermine* showed two typical modes of *spermine* binding, also found in solution.[11] Some of the *spermine* molecules were wound along one side of the groove effectively neutralizing the negative charges on the phosphate groups. *Spermine* molecules have also bridged the opening to the major groove by hydrogen bonding to phosphate groups across the grooves on opposite edges. These specific interactions of *spermine* molecules might constitute the major reason for their very effective stabilization of left-handed Z-DNA where, phosphate groups come very close together.

1.4.2 DNA binding drugs

A majority of antitumor /anticancer drugs present in clinical use are known to exert their cytotoxic action by inhibiting the biological function of DNA. Ligands bind to DNA by a combination of several attractive forces such as electrostatic,

hydrophobic and hydrogen bonding. On a structural basis, drugs that bind to DNA are classified into either intercalating or groove binding.

1.4.2a Intercalating molecules

This class of molecules posses prominent heterocyclic, planar aromatic moiety and form complexes by inserting themselves (intercalating) between the base pairs in DNA (e.g. *acridines* and *ethidium ions*; Figure **7**).[12] Structural and biological aspects of intercalating drugs, especially those active as anticancer agents have been extensively surveyed.[13]

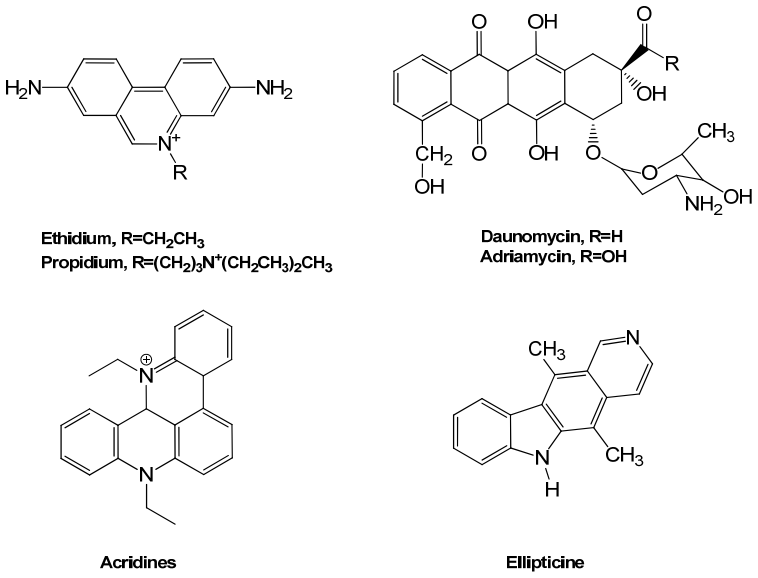


FIGURE 7: Representative intercalating molecules

Various naturally occurring anti-tumor agents like *actinomycin D*, *daunomycin*, *adriamycin*, and drugs such as *chloroquine* also bind DNA in intercalating mode. Intercalating agents come in many guises and interact with DNA with or without sequence selectivity.[14] In general, intercalating drugs have a polarizable, electron deficient chromophore that is typically composed of atleast three fused six-membered rings with a formal positive charge – located either on the central chromophore or on an attached substituent group. The substituents on the aromatic ring of the intercalating agent may protrude into one of the DNA-grooves, interacting sequence specifically with the base pairs. Intercalation also leads to an unwinding of the helix, resulting in a lengthening of the DNA-double helix. Solution studies have indicated that at saturating concentrations of the intercalating species, binding occurs at most between base pairs at alternate sites on DNA - the so called neighbor

exclusion principle.[12] In addition to inducing conformational changes, binding of the intercalators affects other physical properties of DNA such as stabilization towards thermal-denaturation and reduction of its buoyant density in CsCl gradient (hydrodynamic property). Changes in the spectroscopic and chemical properties of the bound drug have also been observed.[14]

1.4.2b Groove binding molecules

The major and minor grooves of DNA differ significantly in electrostatic potential, hydrogen bonding characteristics, steric effects and hydration. Specificity in binding to the nucleic acids results from molecular contacts between the bound ligand and the edges of base pairs on the floor of the groove. Many proteins exhibit binding specificity with nucleic acids primarily via major groove interactions while smaller molecules in general bind in the minor groove of the DNA duplex.

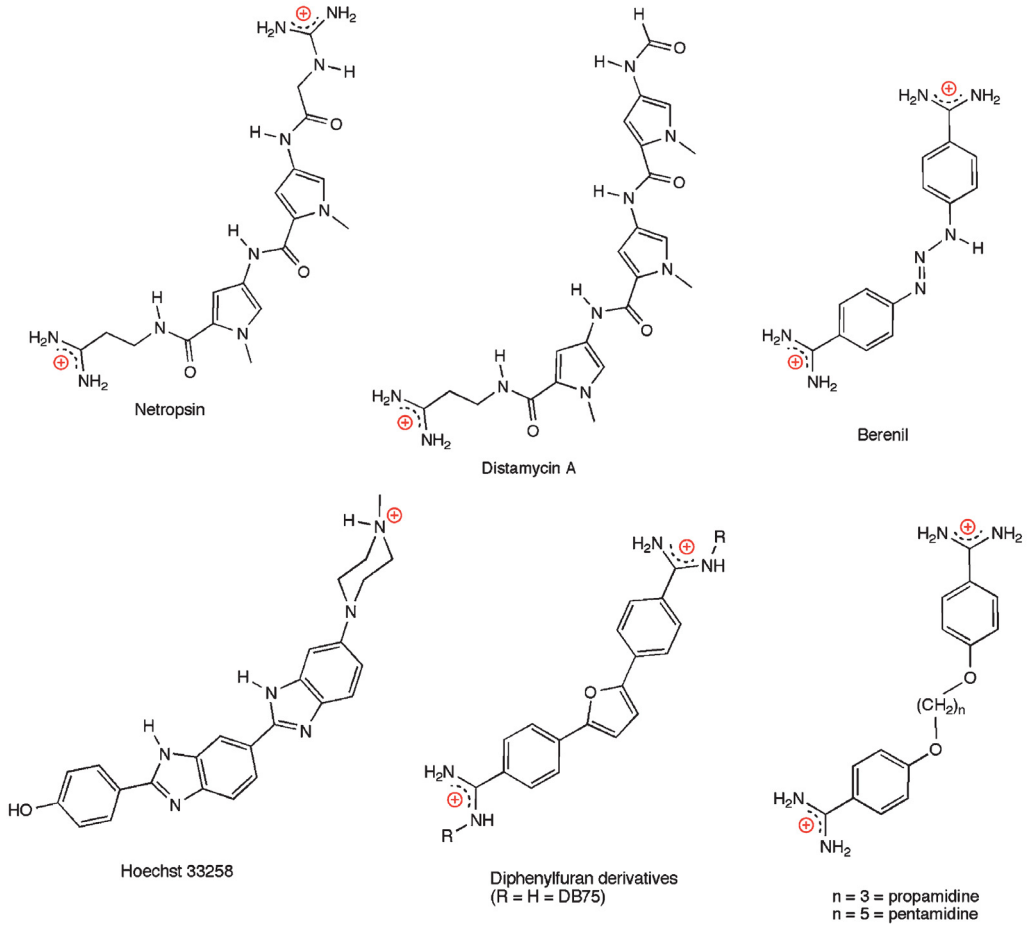


FIGURE 8: Structures of some DNA minor groove binding ligands

As described in the earlier (Figure **5**) it is possible to distinguish A:T from T:A and C:G from G:C in the major groove since the ordered array of hydrogen bonding sites and hydrophobic centers differ among the four pairs. The negative electrostatic potential due to phosphate charges is greater in A:T minor groove than in G:C rich regions and this provides an additional important source for A:T specific minor groove recognition. Figure **8** provides a list of minor groove binding ligands.

DNA-minor groove binders have been reviewed by Lown *et al.*[15]. Much of the intimate structural details of minor groove binding have come from the crystal structure studies of DNA-netropsin complex (Figure **9.A**). Netropsin is an anti-tumor, antibiotic that binds in the minor groove of DNA at the AATT center of the symmetric duplex d(CGCGAATTCGCG). Dickerson *et al.*[16] have demonstrated by X-ray crystallography that netropsin with two aromatic *N*-methylpyrrole (Py rings), forms a

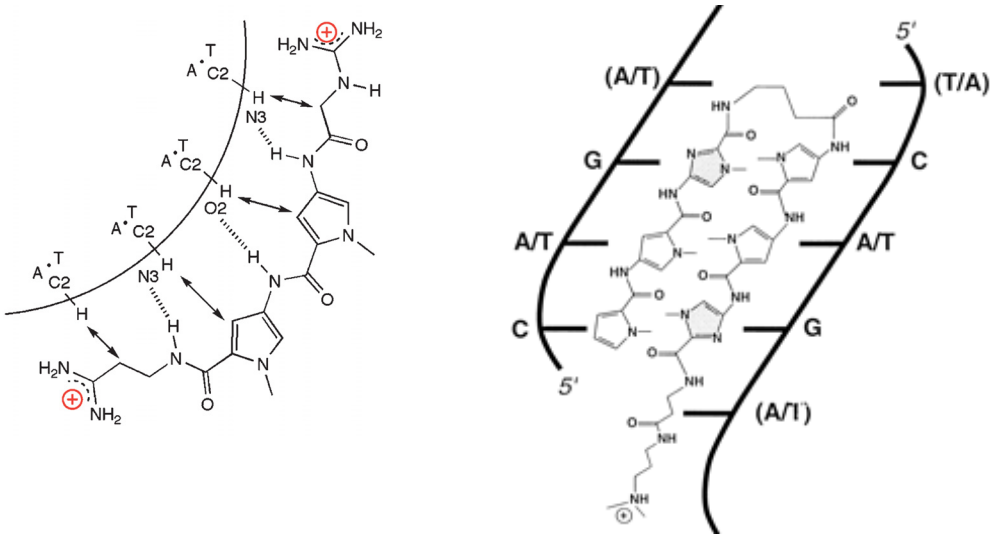


FIGURE 9: Schematic representation of the specific molecular interactions that occur between DNA and (**A**) Netropsin and (**B**) Hairpin polyamide; Dotted lines indicate hydrogen bonds. Double-headed arrows represent nonbonded van der Waals contacts. The curved line is the floor of the minor groove

1:1 complex in the minor groove with an adenine and thymine-rich DNA fragment. Other DNA-minor groove binders include synthetic Hoechst 33258, SN 6999. By NMR spectroscopy Wemmer *et al.*[17] have found that the natural product distamycin (PyPyPy) can bind to DNA in a 2:1 as well as in a 1:1 stoichiometry, depending on the concentration. Dickerson *et al.*[16] and Lown *et al.*[18] have also suggested that replacing *N*-methylpyrrole in netropsin by an *N*-methylimidazole (Im) unit should allow the recognition of a G:C bp. This led Dervan *et al.*[19] to synthesize polyamides containing ImPyPy sequences which allowed sequence specific recognition of the

DNA duplex. This new class of compounds called "lexitropsins" can differentiate between A:T and G:G base pairs.

Polyamides containing imidazole and pyrrole amino acids can be combined in antiparallel, side-by-side, dimeric complexes in "hairpin-linked lexitropsins" for sequence-specific recognition in the minor groove of DNA (Figure **9.B**). Based on footprint titration and affinity cleavage experiments Dervan *et a*[19,20] have formulated a set of pairing rules for the sequence specific recognition of DNA by these aromatic polyamides. A pyrrole opposite an imidazole (Py/Im pairing) targets a C:G bp whereas an Im/Py pair recognizes a G:C bp. A Py/Py pair is partially degenerate and binds both A:T and T:A pairs in preference to G:C/C:G . For Im/Py/Py the anti-parallel 2:1 binding motif with the polyamide oriented N→C with respect to the 5'→3' direction of the DNA helix in the minor groove was confirmed by NMR methods.[21] More recently, a high resolution X-ray crystal structure of the Im/Py pair bound to DNA revealed the hydrogen bond between the N3 of imidazole and the exocyclic amino group of guanine and a total of three hydrogen bonds between the Im/Py pair and the edge of the G:C base pair.[22] In their later work, to break the degeneracy of A:T/T:A recognition by Py/Py pair Dervan *et al*.[23] have introduced a new aromatic amino acid N-methyl-3-hydroxypyrrole (Hp). A Hp/Py pair recognizes a T.A bp whereas a Py/Hp pair targets the A:T bp. Both pairs however, disfavor binding to a G:C or C:G pair. These polyamides bind DNA with very high affinity and sequence specificity and have very high potential as transcription inhibitors. In a first case study, it has been shown that the DNA-binding activity of *the 5S RNA gene-specific transcription factor* TFIIA was inhibited by an eight-ring hairpin polyamide that bound within the recognition site of the nine-zinc-finger protein in the minor groove.[24] As a result, the transcription of 5S RNA genes by RNA polymerase III was suppressed *in vitro* and in cultured *Xenopus* kidney cells.[25]

1.4.2c Intercalation *vs.* Groove binding molecules

There are some interesting compounds that possess structural features that could, in principle, involve binding by either intercalation or groove binding. In reality, one is favoured over the other. The factors that dictate such choice of binding mode are free energy, torsional freedom of ligands, solvent release from grooves *etc.*. SN6999, DAPI, Chloroquine and Diphenylpyridine are a few specific examples of this type.

Therefore at this point it can be said that for all drug–DNA interactions, intercalation and groove binding represent two potential low-energy wells in a continuous free energy surface. The binding mode with lower energy will depend on DNA sequence and conformation as well as the molecular features of the bound ligand.

1.4.3 DNA-Protein interactions

Nucleic acids are rarely found alone in the cells and occur as complexes with proteins. These proteins/enzymes mediate all nucleic acid transformations such as replication, gene regulation, splicing or degradation. Protein-nucleic acid interactions are far more complex than protein-ligand or nucleic acid small molecule interactions.[26] Mostly the DNA-binding motifs provide structural frameworks from which specific amino acid side chains extend to contact specific base pairs in the DNA (Figure **10**). Charged residues, in particular the guanidinium group of the arginine side-chain, are known to be essential to several DNA-protein interactions. Following is a short review of the known principles of protein–nucleic acid interactions and explanation for how these molecules recognize and distinguish specific binding sites from all other potential binding sites in the cell.

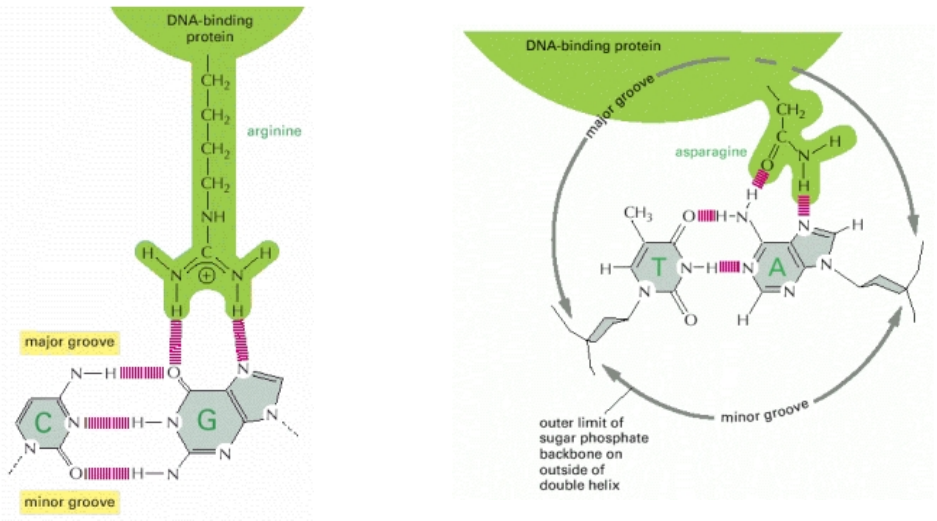


FIGURE 10: Amino acid specificity postulated in case of DNA and protein interactions

The main electrostatic component of DNA-protein binding comes from interaction of negatively charged phosphate backbone with cationic side chains of lysine and arginine. These interactions however, are predominantly non-specific and are incapable of producing sequence specificity on their own. Much of the specificity in DNA-protein interactions is derived from hydrogen bonding of protein side chains

with specific base pairs in the major/minor grooves of DNA. The topological interaction of proteins with DNA occurs via insertion of an α-helix into major groove of DNA duplex and a β-motif interaction with DNA in the minor groove. In such complexes, the side chains of amino acids recognize the edges of base pairs to form a specific recognition system.[27] Much of the present knowledge of sequence-specific interactions has grown from extensive studies of interaction of regulatory proteins with operator DNA, DNA-polymerase complexes, DNA restriction enzymes and recent work on eukaryotic transcription factors.[28]

Figure **11**-**13** show representative helical motifs and their docking into the grooves of DNA. The DNA binding proteins Cro,[29] λ,[30] Trp repressor,[31] CAP all have quite different overall structure, but share a common structural feature in their DNA recognition; they contain a pair of α-helices termed helix-turn-helix motif (Figure **11**).[32] One of these pairs is the recognition helix, which resides in the major groove of DNA, making sequence-specific contacts. Two other structural motifs have been identified as major determinants of DNA-protein interactions. In the zinc-finger motif[32c,33] (Figure **12**) which is most commonly found in several eukaryotic transcription factors, the recognition element is held in place by zinc through interaction of β-sheet/α-helix substructure *via* Cys/His side chains, forming Zn-S and Zn-N coordinating bonds. In linear leucine zipper proteins[32c] (Figure **13**), the interacting recognition elements are constructed *via* two protein chains containing Leu at every seventh residue. This forces a stable zipper like hydrophobic locking of Leu side chains, thus holding the recognition units together in space, for a dimeric complexation with DNA.

Restriction enzymes present a class of DNA-binding proteins, which show extremely high sequence-specificity, and cleave DNA hydrolytically with great precision. These, therefore, have evolved as major tools in genetic engineering and recombinant DNA technology. For example the enzyme *Eco*R1 recognizes the symmetric hexameric sequence GAATTC and cleaves it specifically between G and A to produce sticky ends. The DNA binding element of this enzyme itself shares no structural homology to other established motifs such as helix-turn-helix, zinc-finger, etc. This enzyme employs a four α-helix bundle with two extended side chains that project into the major groove of DNA to contact the substrate purines and pyrimidines. The enzyme binds as a dimer symmetrically with each monomer enwrapping DNA and anchoring the scissile bonds at active sites. The binding of the

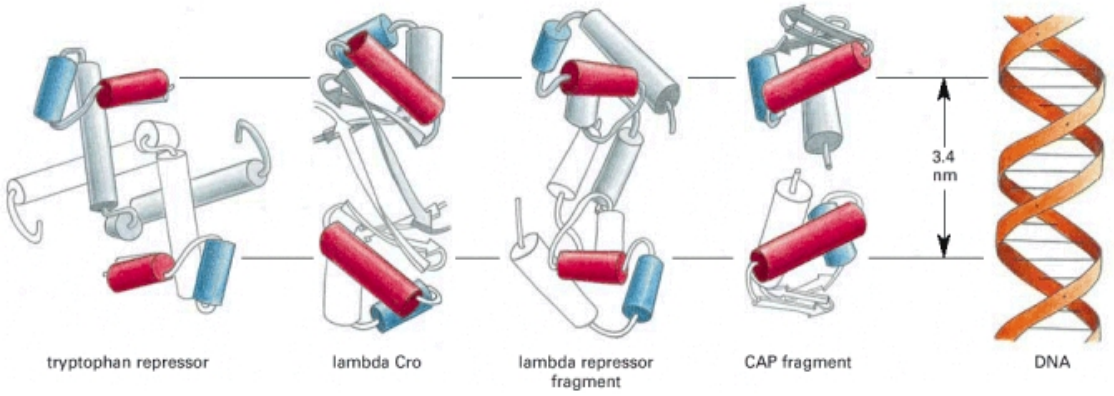


Figure 11: A few typical helix-turn-helix DNA-binding proteins; All of the proteins bind DNA as dimers in which the two copies of the recognition helix (red cylinder) are separated by exactly one turn of the DNA helix (3.4 nm). The other helix of the helix-turn-helix motif is colored blue. The lambda repressor and Cro proteins control bacteriophage lambda gene expression, and the tryptophan repressor and the catabolite activator protein (CAP) control the expression of sets of E. coli genes.

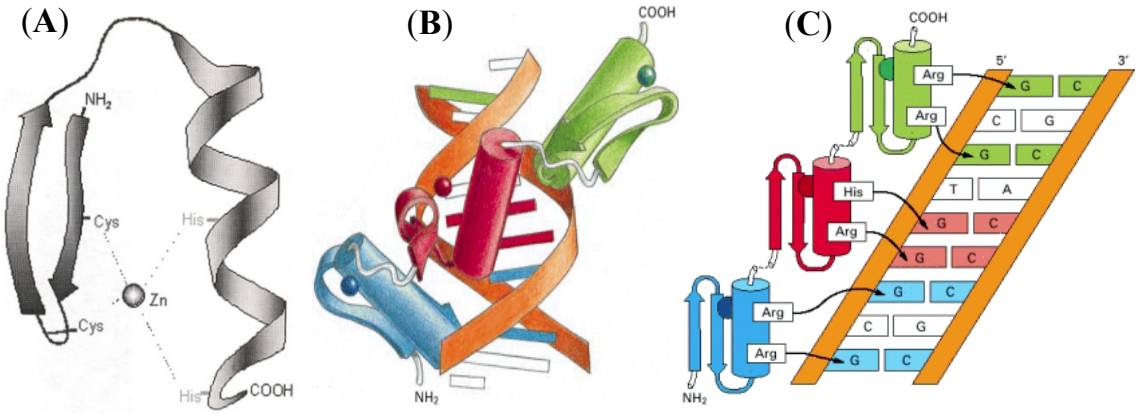


Figure 12: DNA binding by a zinc finger protein; (**A**) Zinc fingers of the Cys-Cys-His-His type (**B**) The structure of a fragment of a mouse gene regulatory protein bound to a specific DNA site. This protein recognizes DNA using three zinc fingers arranged as direct repeats. (**C**) The three fingers have similar amino acid sequences and contact the DNA in similar ways. In all these figures zinc atom in each finger is represented by a small sphere.[34]

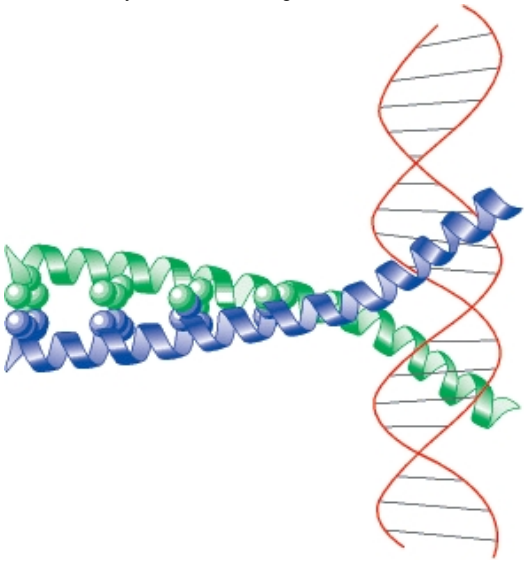


Figure 13: Two α-helical DNA-binding domains (bottom) dimerize through their α-helical leucine zipper region (top) to form an inverted Y-shaped structure. Each arm of the Y is formed by a single α helix, one from each monomer that mediates binding to a specific DNA sequence in the major groove of DNA. Each α helix binds to one-half of a symmetric DNA structure.

enzymes induces unwinding and bending of the DNA-axis resulting in a widening of the major groove by 3.5 Å at the cleavage site.[35] Such structural changes are lacking in non-recognizable sequences of DNA. This discrimination in specificity is a result of specific recognition mechanism involving hydrogen bonding with purine, hydrogen bonding and van der Waals interaction with pyrimidines and sequence dependent phosphate interactions which lead to structure distortion.

Protein-DNA interactions offer both specific and non-specific recognitions. In sequence-specific recognition there is direct readout of the DNA sequence by hydrogen bonding between the protein and the DNA, while in a non-specific complex, there are fewer direct interactions with the bases, and the buried surface area is often smaller. Consequently, a non-specific complex has smaller favourable contributions from these effects along with smaller entropic penalties from conformational effects. Thus, target recognition is here associated with structural changes in the protein that alter both the surface charge distribution and the shape.

1.5 Gene delivery

Gene delivery is the process of introducing foreign DNA into host cells and is one of the necessary steps for gene therapy or the genetic modification to deal with genetic disorders (diseases or mutations).[36] The success of such therapies rely heavily on efficient intracellular delivery of exogenous DNA for blocking undesired gene expression, but naked DNA is unable to efficiently cross cellular barriers by passive diffusion because of its large size, strong negative charge, hydrophilicity and susceptibility to nuclease attack. And so this generally requires delivery vehicles that are capable of carrying/protecting genetic materials along with target-specifically delivering to the desired tissues or subsets of cells.[37] There are many different methods of gene delivery developed for various types of cells and tissues, from bacterial to mammalian. Generally, the methods can be divided into two broad categories: **viral** and **non-viral**.

Virus mediated gene delivery utilizes ability of a virus to inject its DNA inside a host cell. A gene that is intended for delivery is thus packaged into a viral particle, while non-viral methods can include varied range of mechanical, physical or chemical processes. But over the past decades, significant endeavors have been devoted to develop non-viral gene delivery vehicles[38] as alternatives to their viral counterparts, whose applications are restricted due to the potential safety issues, immunogenicity, possible random genomic integration and complex processes of preparation. Table **2**

lists the different non-viral methods used for enhancing DNA penetration of the plasma membrane.

Table 2. Summary of DNA transfection methods.

Approach	Method
Physical	Microinjection[39] Hydrostatic pressure[40] Gene gun[41] Electroporation (high[42]/low voltage[43]) Sonication[44]
Chemical[45]	DEAE-dextran Calcium phosphate Artificial lipids Proteins Dendrimers Other polymers (including controlled-release polymers)[46] Peptides Nanoparticles

Among the existing non-viral gene delivery systems, peptide and polymer based delivery vehicles (*e.g.* liposomes & cationic polymers) have received extensive attention. However, lack of control during synthesis of cationic polymers often results in polymer–DNA formulations with non-uniform physico-chemical characteristics. Cellular toxicity and difficulties in selectively modifying the polymer with ligands for targeted delivery pose additional problems. Although liposomes are commonly employed for *in vitro* DNA delivery, lipoplexes are unstable in the presence of serum or blood components and tend to change in size, surface charge and lipid composition, restricting their *in vivo* application.

The use of low molecular weight (LMW) peptides as alternate DNA delivery systems is advantageous in many ways as they are relatively stable, easy to synthesize and functionalize, less toxic or immunogenic compared to other non-viral and viral vectors and only weak activators of the complement system.[47]

1.6 Peptide mediated DNA delivery

Peptides are emerging as attractive alternatives to cationic polymers and lipids for nonviral DNA delivery. Their remarkable properties such as efficient condensation of DNA, translocation across the cellular membrane, pH-sensitive membrane disruption, and efficient targeting of attached cargoes to the nucleus make them lucrative for exploitation as DNA delivery vehicles.

Nevertheless the binding and specific release of DNA from the delivery vehicles is majorly guided by its affinity for each other and also the strength of molecular interactions. These interactions occur through nonspecific binding based on DNA and peptide composition.

DNA delivery formulations can be prepared by conjugating DNA to the peptide electrostatically or using covalent linkages where the single formulation has all the desired properties. Of all the peptides involved in DNA delivery known, can be classified into four different classes depending on their functions as:

- **DNA-condensing cationic peptides** (lysine and arginine-rich) which form complexes with DNA via electrostatic interaction,
- **Cell-penetrating peptides (CPPs)** which efficiently translocate across the cell membrane along with the cargo,
- **Membrane perturbing peptides (MTPs)** which primarily mediate DNA entry as a function of pH,
- **Nuclear Localization Sequences (NLSs)** which has specific sequence peptides that targets DNA to the nucleus for subsequent gene expression.

In case of DNA-condensing cationic peptides low molecular weight oligolysines usually containing ~18 consecutive lysine residues efficiently transfer DNA by forming small (<50 nm)-sized peptide–DNA complexes[48] and are more stable in the presence of serum and salt which closely resemble the *in vivo* conditions. Interestingly arginine-based peptides show better condensation, while lysine-based ones are weaker activators of the complement system.[49] But since these complexes are formed through electrostatic interaction between peptide and DNA, they are seen to be unstable during *in vitro* and *in vivo* gene delivery because the poor peptide–DNA binding leads to easy DNA release.

Cell penetrating peptides have inherent property of membrane translocation and enhance DNA delivery across the cell membrane.[50] Such CPPs include the naturally derived peptides (e.g. TAT and penetratin), chimeric peptides (e.g. transportan) and synthetic ones (e.g. oligoarginine).[51] The chemical nature of these CPPs varies from highly basic to amphipathic and some are even totally hydrophobic.[52] In this case also arginine plays more crucial role than lysine with respect to residue number and backbone structure. It is also suggested to have a sort of molecular mechanism between the guanidinium group and a hydrogen-bond acceptor moiety (proteoglycans – GAGs and heparan sulfate) present on the plasma membrane.[53] A

number of CPPs have been utilized for efficient delivery of ODNs, peptide nucleic acids (PNA), morpholinos, plasmid DNA as well as siRNA.[54] One of the major issues in using CPP–DNA complexes formed through electrostatic interaction is that the arginine moieties which impart efficient translocation ability are likely to be masked in the complex formation. Thus a careful control on the number and position of arginine residues in the peptide architecture and charge ratios used in their design become important for overall mechanism of the CPPs.

Membrane perturbing peptides have the ability to escape from the endosome and release the DNA once they are inside the cell at endosomal pH. These are rich in histidines and are again subdivided into two categories. (i) Fusogenic peptides, mostly adopt amphipathic α-helical structures and are components of viral proteins (*e.g.* HA-2, the N-terminal end of influenza virus hemagglutinin), their derivatives (GALA, KALA, etc.) or synthetically designed (JTS).[55] (ii) Endosmolytic Peptides, assists in both early and late endosomal release of DNA cargoes like H5WYG peptide which is active even in the presence of serum.[56] In addition the LAH series of peptides are multifunctional. These peptides can adopt a transmembrane orientation at neutral pH. When the pH is lowered the histidines become protonated and the peptide flips into an in-plane orientation at the membrane surface thereby destabilizing anionic lipids of the membrane. The position and number of histidine residues (4–5 at the core) is crucial in obtaining the best transfection efficiency.[57]

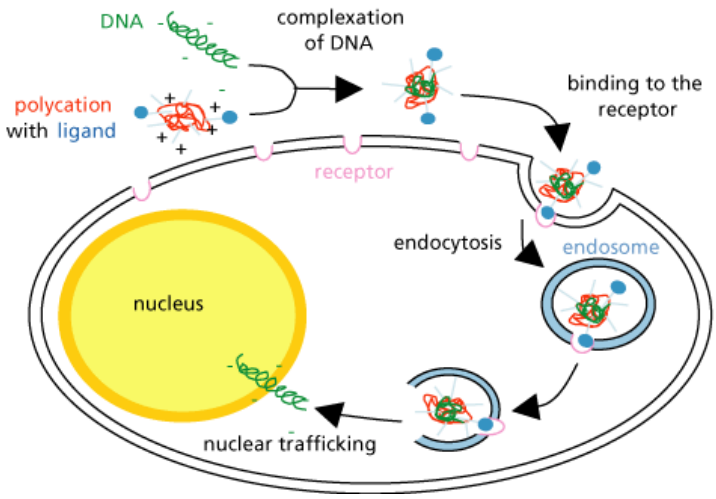


FIGURE 14: Gene delivery using cationic co-polymers, peptides and lipids

A generalized mechanism of gene delivery using cationic polymers, peptides and lipids is shown in Figure **14**.

Table 3. Common peptides used in DNA delivery

Peptide name	Sequence
DNA condensing peptides	
Linear lysine-based cationic peptides	Y K A K_n W K AlkC W K_n (n = 3 to 36) $(K)_n$ W C (s – s) C W $(K)_n$ CWK_{17} C (modification from CWK_{18})
Branched cationic oligopeptide	$(K K K)_2$ K G G C
Cell penetrating peptides (CPP)	
Penetratin	R Q I K I W F Q N R R M K W K K
Transportan Transportan-10	G W T L N S A G Y L L G K I N L K A L A A L A K I S I L A G Y L L G K I N L K A L A A L A K _ **K** I L
VP22	D A A T A T R G R S A A S R P T E R P R A P A R S A S R P R R P V D
HIV TAT	C G R K K R R Q R R R P P Q C
HIV-1 Rev	T R Q A R R N R R R R W R E R Q R
FHV coat	R R R R N R T R R N R R R V R
$(Arg)_n$, n=(7-16) Stearylated R8, Linear & Branched oligoarginine	$(R)_n$, Sterylated $(R)_8 (R)_n$-G-C
Haemagglutinin (HA-2) derived amphipathic peptides	
HA	G L F E A I A G F I E N G W E G M I D G
K5 E5 E5CA E5WYG H5WYG	G L F **K** A I A **K** F I **K G** G W **K** G **L** I **K** G G L F K A I A **E** F I **E** G G W **E** G L I **E** G G L F K A I A E F I E G G W E G L I E G **C A** G L F K A I A E F I E G G W E G L I E G **W Y G** G L F **H** A I A **H** F I **H** G G W **H** G L I **H** G W Y G
GALA KALA	W E A A L A E A L A E A L A E A L A E H L A E A L A E A L E A L A A W E A **K** L A **K** A L A **K** A L A **K H** L A **K A** L A **K** A L _ **K** A **C** E A
Cationic amphipathic histidine rich peptides	
LAH4 LAH4-L1 LAH6-L1 LAH6-L1-80	K K A L L A _ L A L H H L A H L A L H L A L A L _ _ _ K K A K K A L L A _ **H** A L H **L** L A **L** L A L H L A **H** A L _ _ _ K K A K K A L L A **L** H A L H **H** L A L L A **H** H L A H A L _ _ _ K K A K K **H** L L A _ H A L H **L** L A L L A **L** H L A H A L **A H L** K K A
JTS ppTG-1 ppTG-20	G L F E A L L E L L E S L W E L L L E A G L F **K** A L L **K** L L **K** S L **T K** L L L **K** A G L F **R** A L L **R** L L **R** S L T **A** L L L **R** A
Nuclear localisation sequences (NLS)	
SV40 large T Ag Ext SV40	P K **K*** K R K V E D P Y C S S D D E A T A D S Q H S T P P K **K*** K R K V E D P Y C
Mu	M R R A H H R R R R A S H R R M R G G
MPG	G A L F L G F L G A A G S T M G A W S Q P K S K R K V
HTLV	M P K T R R R P R R S Q R K R P P T W A H F P G F G Q G S L C
Mellitin	G I G A V L K V L T T G L P A L I S W I K R K R Q Q

Peptides with nuclear localization sequence transport the DNA across the nuclear membrane through the nuclear pore complex (NPC).[58] The classical NLS sequences used in DNA delivery have stretches of highly basic amino acids: either one cluster (monopartite) or two clusters of basic residues separated by 10–12 neutral residues (bipartite). Conjugation of DNA (plasmid, linearized plasmids and oligonucloetides) to NLS moieties can be achieved using both electrostatic and covalent methods. However, covalent method of NLS–DNA conjugation allows greater control on the peptide–DNA formulation. The number of NLS sequences required for efficient nuclear entry depends on the conformation, size and length of

the DNA to be transported. Covalently attached NLSs with reduced positive charge are more efficient in transport.[59]

Although peptides have the potential to develop new paradigms in cellular gene delivery, application of peptides for in vivo delivery of therapeutic DNA is still in its infancy and various areas of concern like CPP entry with and without cargo is a gray area, single and composite multifunctional peptides, *in vivo* efficacy and physico-chemical properties of peptide–DNA complexes need to be critically addressed before significant solutions emerge.

1.7 Physical and structural techniques applied to study of nucleic acids (DNA)

The equilibrium binding of small molecules & peptides to nucleic acids has been an active area of research for half a century. The previous sections focused on basic concepts and major issues by use of selected examples. Here many of the biophysical and structural methods used in studying the overall ligand and peptide binding to nucleic acids are outlined.

1.7.1 Spectroscopic Techniques

1.7.1a Ultraviolet Absorption

The light absorption characteristics of nucleic acids[60] result from the combination of the strong ultraviolet (UV) absorption of the purine and pyrimidine bases in the 240–280 nm range modulated by the stereochemistry and conformational influences of a ribose-phosphate backbone that is essentially transparent to light of that wavelength. Oligonucleotides exhibit a strong UV absorption maximum λ_{max} at approximately 260 nm and a molar extinction coefficient ε of the order of 10^4 dm^3 $mol^{-1}cm^{-1}$. This absorption arises almost entirely from complex electronic transitions in the purine and pyrimidine components. The intensity and exact position of λ_{max} are functions not only of the base composition of the nucleic acid but also of the state of the base-pairing interactions present, the salt concentration of the solution and its pH. In practical terms this allows estimatation of the molar quantity of an oligonucleotide, which is for example chemically synthesised, by measurement of the number of absorbance units at 260 nm (A_{260}, the optical density) of its solution in a UV spectrometer.

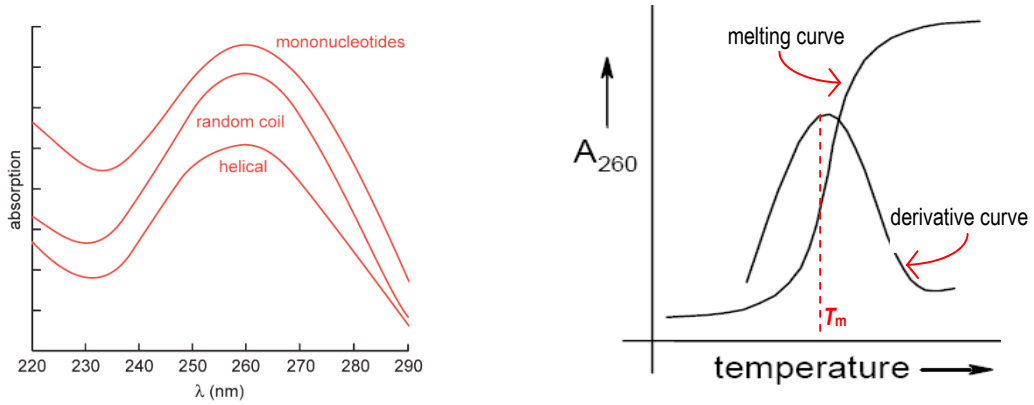


FIGURE 15: (**A**) Typical UV absorption curves for equimolar base concentrations of mononucleotides, of single-stranded (random coil) oligonucleotide and of double-helical DNA; (**B**) An oligonucleotide melting curve and the derivative

Ultraviolet absorption is a sensitive and convenient way to monitor the thermodynamics of DNA. The most common method for determining nucleic acid thermodynamics is thermal denaturation, as the unfolding transitions have analogies to phase transitions and the experimental observable is the intrinsic hyperchromicity of the nucleobases. It arises because the intrinsic electric dipole transition moments tend to cancel when bases stack or pair, giving rise to a smaller extinction coefficient in the folded state.[61] When the UV absorption of a nucleic acid sample is measured as a function of temperature, the resulting plot is known as a melting curve (Figure **15.B**). The midpoint in the melting curve showing the increase in absorbance with increasing temperature is known as the melting temperature T_m and this is dependent on various conditions of pH, temperature, or ionic strength that disrupt hydrogen bonds leading the strands to be no longer held together. That is, the double helix is denatured and the strands separate as individual random coils. UV-melting is a co-operative phenomenon and the observed melting curves become progressively steeper with increasing length of the oligonucleotides (Figure **15.A**).

1.7.1b Circular Dichroism

Circular dichroism is a spectroscopic method which depends on the fact that certain molecules interact differently with right and left circularly polarized light. Circularly polarized light is chiral and hence to discriminate between the two chiral forms of light, a molecule must be chiral, which includes the vast majority of biological molecules, in particular determining the helical content of proteins and conformational analysis of nucleic acids and interaction with other ligands. This method can discern the subtle differences between non-superimposable mirror

image molecules (enantiomers) and is highly sensitive to the three-dimensional features of molecules or its conformation. This expectation has been amply-verified and forms the basis for the many applications of CD in biochemistry.

CD spectra are particularly valuable in determining the following aspects:

- Whether the individual strands themselves are self-complexed
- Whether the individual strands alter their conformations during duplex or triplex formation and
- How the conformations of hybrid complexes, with mixed DNA-peptide strands, are related to those having all DNA strands

Linearly polarized light can be shown to consist of two circularly polarized components of equal intensity (Figure **16**). A CD signal results from the differential absorption of left and right circularly polarised light. Circularly polarised light has chirality and therefore it will exhibit chiral discrimination. A chiral molecule will absorb left and right circularly polarised light to differing extents and this difference gives rise to the phenomenon of CD.

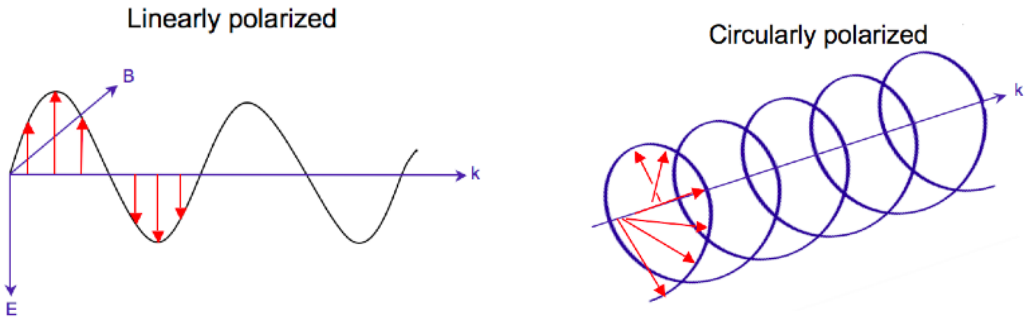


FIGURE 16: Pathway of Linearly and Circularly polarized light

According to the Beer-Lambert law, if the absorbing species has a molar concentration c and the sample thickness or path length is l, the absorbance is given by

$$A = \varepsilon cl$$

where, ε is defined as the molar decadic absorption coefficient or molar extinction coefficient. Then, molar circular dichroism, $\Delta\varepsilon$,

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R = (A_L - A_R)/\, cl$$

where ε_R and ε_L are molar extinction coefficients for right circularly polarized and left circularly polarized light respectively.

Since $\Delta\varepsilon$ is a function of wavelength, a molar circular dichroism value ($\Delta\varepsilon$) must specify the wavelength at which it is valid. An alternative measure of CD is molar ellipticity. The ellipticity is an angular measure which is related to ΔA as follows:

$$\Theta = 32.98\ \Delta A$$

where Θ is in degrees. To eliminate the effects of path length and concentration, the molar ellipticity is defined as

$$[\Theta] = 100\ \Theta / lc = 3298\ \Delta A$$

The units for molar ellipticity are degrees $cm^2/dmol$. Typically, data are presented as plots of [Θ] *versus* wavelength.

Circular dichroism (CD)-the chiral spectroscopy has been widely applied to the study of peptides[62] and nucleic acids[63]. Single stranded DNAs are structurally less defined than duplex DNAs and their CD signal is smaller, while the CD spectra of naturally occurring RNA show them to be mainly well structured with extensive duplex regions (stem-bulb).

In an isolated nucleotide, the chirality originates from the ribose sugar and influences the base absorption. The CD signal of a nucleic acid increases with length due to the co-operativity of chiral interactions between contiguous bases. This effect occurs both as a result of sequence effects arising from nearest-neighbour interactions as well as from overall gross secondary structure.

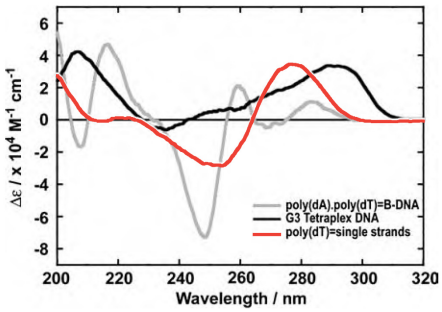


FIGURE 17: Circular dichroism spectra for three different conformations of DNA, a poly(dA)poly(dT) B-type duplex, a guanine quadruplex (tetraplex) and a single-stranded random coil

In the normal spectral region (down to 180 nm), the chiral ribose-phosphate backbone has no transitions and in the CD spectrum of nucleic acids, one detects the CD induced into base transitions as a result of their coupling with the backbone transitions.[64,65] The magnitude of $\Delta\varepsilon_{max}$ is of the order of 2 mol^{-1} dm^3 cm^{-1} at 270 nm and the purine bases have a negative signal whereas the pyrimidines have a positive signal. In the case of CD spectrum of DNA polynucleotides with stacked bases, the

magnitude is larger at 270 nm and significantly larger at 200 nm than that of individual bases (Figure **17**). The spectrum is dominated by the induced CD transitions of the bases from their coupling among each other, due to their stacking on top of each other in a chiral (helical) fashion.

1.7.1c Fluorescence

Fluorescence is defined as the emission of radiation as a molecule returns to its ground state from an excited electronic state.[66]

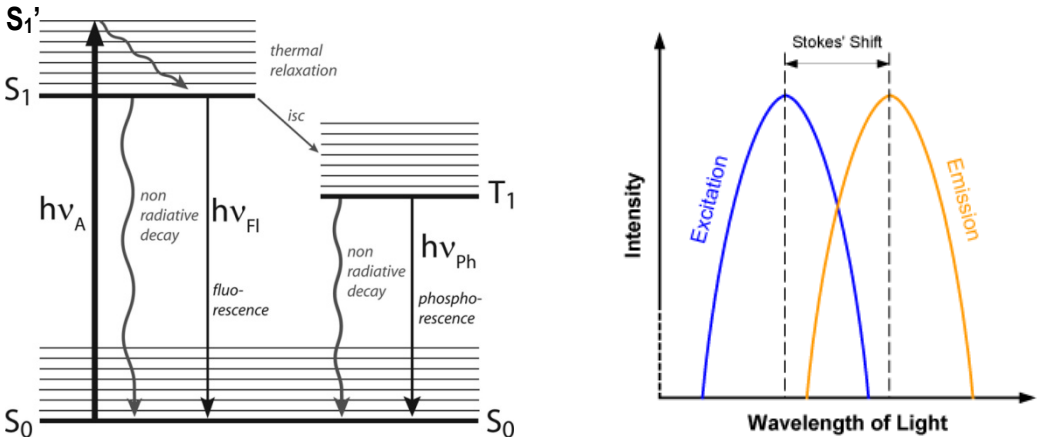


FIGURE 18: (**A**) A simplified Jablonski diagram of a fluorescence event. The fluorescent molecule begins in its ground energy state, S_0, and is converted to an excited singlet state, S_1', by absorbing energy ($h\nu_A$) in a specific wavelength. The molecule has a transition to the relaxed singlet state, S_1 or T_1, by releasing some of the absorbed energy. Finally, the molecule returns to its ground energy state by releasing the remaining energy ($h\nu_{Fl}$ - fluorescence) or ($h\nu_{Ph}$ - phosphorescence). The duration of a single fluorescence event is a few nanoseconds.
(**B**) Generalized representation of the absorbance and emission spectra of a fluorophore

Fluorescence spectroscopy is nevertheless invaluable for examining nucleic acid–ligand interactions. Many DNA-binding ligands have little or no fluorescence in aqueous solution. However, upon binding to a nucleic acid the ligand is in a hydrophobic environment and the solvent can no longer quench the intrinsic ligand fluorescence. Therefore fluorescence emission is a direct probe of the concentration of bound ligand.

Ethidium bromide (EB) interacts with DNA by intercalation between the base pairs. It is a weakly fluorescent molecule, but when intercalated to DNA exhibits a strong fluorescence with emission maxima at 595 nm upon excitation at 475 nm. A competent DNA ligand can displace the intercalated ethidium bromide thus leading to a decrease in the fluorescence intensity of the complex.[67] An attractive feature of the

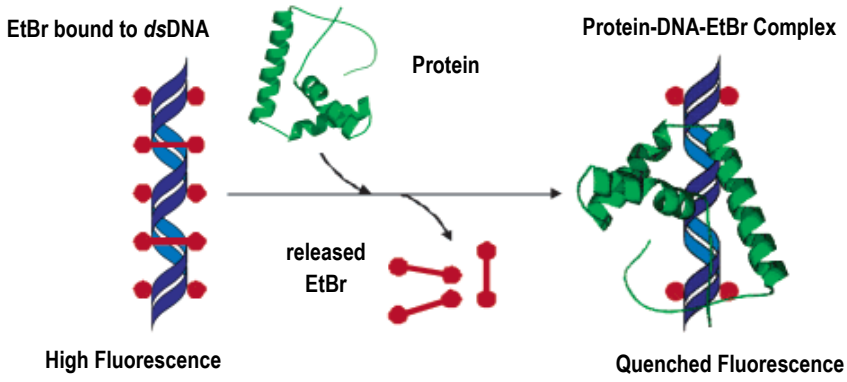


FIGURE 19: Mechanism of intercalated EtBr displacement from *ds*DNA on binding to protein

ethidium binding is that its fluorescence can increase 50 fold upon intercalation with DNA making it a very useful probe for monitoring the interactions of DNA with small molecules.[68] When DNA-ethidium complexes are challenged by other DNA binding agents (e.g. groove binders) it is displaced from the complex leading to a fall in fluorescence intensity (Figure **19**). In such experiments, a measure of the decrease in ethidium cation fluorescence intensity is useful to compute the relative binding strengths of the added ligands to DNA.

1.7.2 Electrophoretic mobility shift assay (EMSA)

One of the important technique for studying gene regulation and determining protein:DNA interactions is the electrophoretic mobility shift assay (EMSA). This can be used qualitatively to identify sequence-specific DNA-binding proteins

The advantages of studying protein:DNA interactions by an electrophoretic assay are:

- Ability to resolve complexes of different stoichiometry or conformation.[69]
- Crude samples directly from nuclear or whole cell extract, *in vitro* transcription product all can be analysed at the same time.

In free solution, the movement of DNA in an electric field is independent of shape and molecular mass and dependent only on charge. However when the DNA is exposed to an electric field in a gel matrix, the movement is dependent on size and shape as well as charge. Gels commonly used for nucleic acid electrophoresis are made of agarose and polyacrylamide. Both types of material consist of 3D networks of cross-linked polymer strands, which contain pores whose size varies according to the concentration of polymer used. The mobility of DNA in such gels is dependent

mainly on size and shape, since the charge per unit length of DNA is effectively constant. Shorter molecules move faster and migrate farther than longer ones because shorter molecules migrate more easily through the pores of the gel. This phenomenon is called sieving.[70]

Agarose gel electrophoresis is generally adequate for resolving nucleic acid fragments in the size range of 100 nucleotides to around 10-15 kb. Below this range, fragments are both difficult to separate and hard to visualize because of diffusion within the gel matrix.

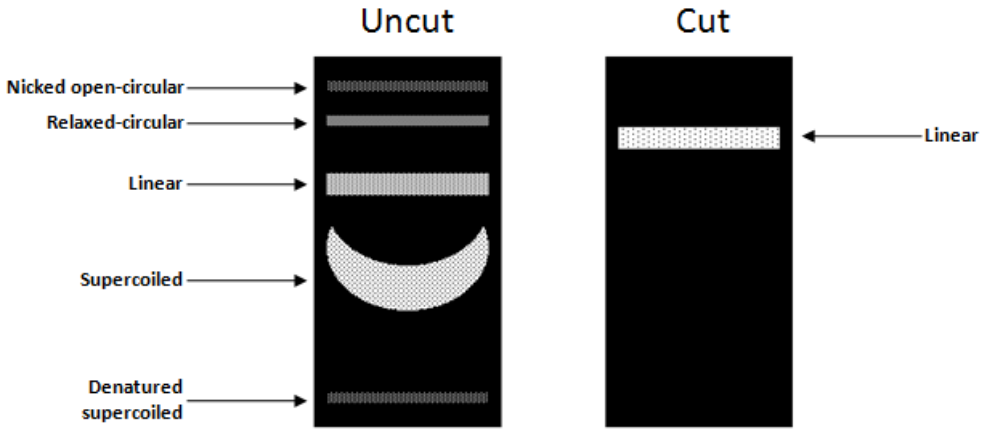


FIGURE 20: Relative position of different DNA forms of a plasmid on agarose gel, before and after enzyme digestion

Plasmid DNA appears as five conformations (*Nicked Open-Circular, Relaxed Circular, Linear, Supercoiled, Supercoiled Denatured*), which (for a given size) run at different speeds in a gel during electrophoresis. *Supercoiled* DNA migrates faster through a gel than linear or open-circular DNA because of its tight conformation. The electrophoresis pattern of plasmid DNA after digestion with restriction enzymes gives only one band for the short linear fragment (Figure **20**). The most common dye used to make DNA bands visible for agarose gel electrophoresis is ethidium bromide (EtBr).

Native polyacrylamide gel electrophoresis (PAGE) can separate fragments as small as 10 bp and upto 1 kb with a resolution of as little as 1 bp. It can also accommodate much larger quantities of DNA (up to 10 μg) and allow recovery of very pure DNA. Figure **21** describes a typical method of PAGE.

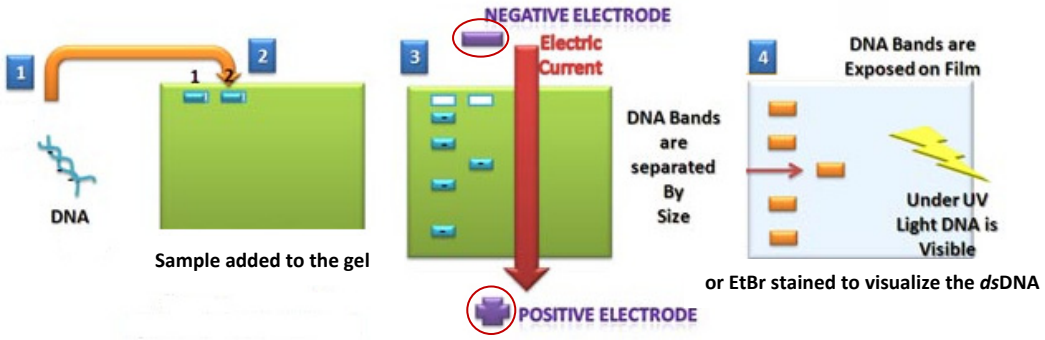


FIGURE 21: Method of polyacrylamide gel electrophoresis

There are two types of polyacrylamide gels:

Denaturing polyacrylamide gels - used for the separation and purification of single-stranded fragments of DNA. These gels are polymerized in the presence of an agent (urea and/or, less frequently, formamide) that suppresses base pairing in nucleic acids. Denatured DNA migrates through these gels at a rate that is almost completely independent of its base composition and sequence.

Nondenaturing (native) polyacrylamide gels - used for the separation and purification of fragments of *ds*DNA. As a general rule, *ds*DNAs migrate through nondenaturing polyacrylamide gels at rates that are inversely proportional to the $\log_{10}$ of their size but are affected by their base composition and sequence, so that duplex DNAs of exactly the same size can differ in mobility by up to 10 %. Nondenaturing or native polyacrylamide gels are used chiefly to prepare highly purified fragments of DNA and to detect protein-DNA complexes.

While in denaturing-PAGE the electrophoretic mobility of DNA/proteins depends primarily on their molecular mass, in native PAGE the mobility depends on both the DNA/protein's charge and its hydrodynamic size. Thus, mobility on the gel varies with the nature of their conformation (higher mobility for more compact conformations, lower for larger structures like oligomers.[71]

1.7.3 Microcalorimetry

Calorimetry is a technique in which the heat of a reaction is measured. One of the commonly used methods used for measuring the enthalpy change associated with either a binding interaction or a heat-induced conformational transition for nucleic acids and proteins is isothermal titration calorimetry (ITC).[72]

Isothermal titration calorimetry is a powerful and flexible technique that is used to determine directly and independently the change in enthalpy (ΔH) for almost any bimolecular binding interaction at a fixed and constant temperature. Depending on the magnitude of the binding constant and solubility of reactants, ITC can also be used to obtain binding isotherms, from which the equilibrium-binding constant (K_b = K_a) and binding stoichiometry (n) can be determined. This can therefore yield an almost complete thermodynamic profile for a binding interaction in a single experiment.

1.7.4 Mass spectrometry: MALDI-TOF

The determination of intrinsic molecular masses of nucleic acids using mass spectrometry is widely accepted as one of the most accurate methods to detect nucleic acids.[73] **Matrix-assisted laser desorption/ionization time-of-flight** (MALDI-TOF) is one key technology of second-generation called 'soft ionization' method. Some of the current clinical applications of mass spectrometry are (a) DNA sequencing[74], (b) genotyping[75] and detection of genetic variations[76], microsatellites[77], short tandem repeats[78], small insertions/deletions and (c) gene expression.

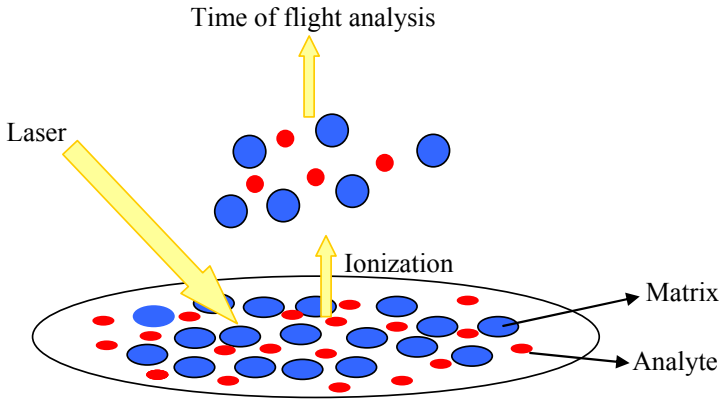


FIGURE 22: Mechanism of ionization in MALDI-TOF spectroscopy

Matrix-assisted laser desorption mass spectrometry was first introduced for proteins by Karas and Hillenkamp in 1988.[79] The nucleic acid and protein samples are embedded in a crystalline matrix of a small molecule (*e.g* α-Cyano-4-hydroxycinnamic acid). This target is excited by a pulse from an ultraviolet laser beam that, in high vacuum, results in intact molecules of the sample becoming desorbed into the gas phase and ionized by the UV radiation to give (mostly) singly charged ions. The matrix assists in the desorption and ionization of the analyte and

molecular weight of up to 500 kDa can be analyzed. The basic concept of TOF Mass analyzer (time-of-flight) is that the ions are separated based on the time taken by the ion to drift down the flight tube to the detector. Lighter ions have higher velocities than heavier ions and reach the detector first.

1.7.5 Transfection

Transfection is the process of deliberately introducing nucleic acids into cells. The term is used notably for non-viral methods[80] in eukaryotic cells. Genetic material (such as supercoiled plasmid DNA or siRNA constructs), or even proteins such as antibodies, may be transfected.

Transfection can be of two types,

Transient transfection- the processes where the transfected gene is expressed only for a short period and is usually not integrated into the nuclear genome, the foreign DNA eventually gets diluted through mitosis or degraded.

Stable transfection- the processes where the transfected gene actually remains in the genome of the cell and its daughter cells. It mostly has a marker gene (Geneticin or G418) which gives the cell some selectable advantage, such as resistance towards a certain toxin. Clones of stably transfected cells are selected by these selectable markers on the vectors.

A Reporter gene is used for determination of the percentage of cells in a transfection experiment that receive and express the foreign DNA sequence. The reporter gene can be present on the same vector as the gene of interest or can be on

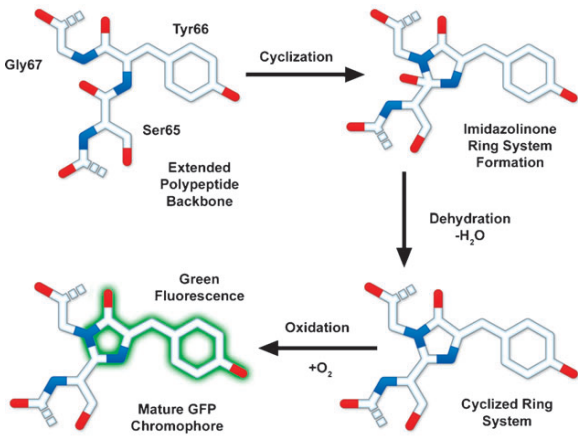


FIGURE 23: *Aequorea Victoria* wtGFP chromophore formation[81]

a separate plasmid. The reporter gene can also be used to create a fusion protein with the gene of interest for protein localization studies. A convenient reporter for monitoring transfection efficiency is the green fluorescent protein (GFP). When excited by blue or UV light, the protein emits bright green fluorescence light through cyclization of a tripeptide chromophore embedded within the complete amino acid sequence as illustrated in Figure **22**.[81]

Genes encoding green fluorescent proteins have been cloned from various coelenterates such as the jellyfish *Aequorea victoria* and the sea pansy *Renilla reniformis*. To facilitate their use as reporters, several GFP variants have been developed by introducing amino acid substitutions into the chromophore, which result in a shift in the emission wavelength as well as an increase in fluorescence intensity. Additional mutations have been introduced to create preferred gene codons in order to increase expression efficiency in different cells. Expression of GFP is typically detected by fluorescence microscopy or flow cytometry. Figure **24** gives the pathway of a GFP tagged reporter gene (pRmHa3-GFP) inside the cell after being transfected till the expression of GFP.

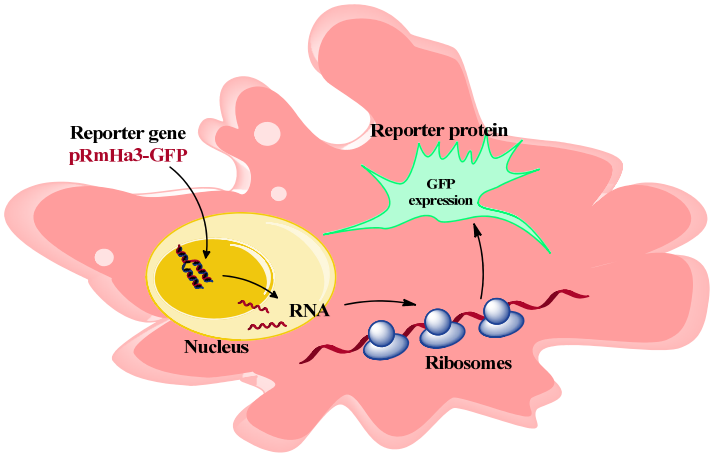


FIGURE 24: Expression of green fluorescent protein having its origin from transfected pRmHa3GFP plasmid.

The use of various chemical, lipid or physical methods, makes 'transfection' a powerful tool to study gene function and protein expression in the context of a cell. Currently, the most widely used method for transfecting cells is with cationic lipids (commercially available) that result in very high transfection efficiencies with low cytotoxicity.[82] Moreover development of reporter gene systems and selection methods for stable maintenance and expression of transferred DNA have greatly expanded the applications for transfection.

Effectene® is a common transfection reagent, produced and sold by Qiagen, used in molecular and cellular biology. It is used to introduce, that is transfect, siRNA or plasmid DNA into *in vitro* cell cultures by lipofection/transfection. Effectene® treatment alters the cellular plasma membrane, allowing nucleic acids to cross into the cytoplasm. It was invented by Dr. Yongliang Chu at *Life Technologies, Inc.*

1.8 Aim and rationale of the present work

Proline based peptides when appropriately functionalized with positive charge bearing groups become good CPP, with inherent ability to translocate through cell membranes. Many peptides have been tested towards this goal with different chain length and charge densities.[83]

Literature reports suggest a variety of natural and synthetic materials have been employed for DNA delivery, which can be categorized as either hydrophobic, hydrophilic or ampphiphilic polymers. Collagen is one such example which has helped in the transport of plasmid DNA and siRNA into cells.[84] The complexation of DNA and collagen has also been studied theoretically.[85]

The possibility of aminoproline based collagen peptides being efficient gene delivery agents are explored here. Aminocollagens present well defined molecular scaffolds. The amino groups in these peptides remain protonated even in physiological pH, which puts them in the list of cationic peptides and since the cationic peptides are known to be efficient DNA transporting agents[86], the cationic guanidino collagen peptide analouges are examined for their DNA binding and cell penetrating abilities. Guanidinium modifications to the 4-amino side chain of proline incorporates extra positive charge for better recognition as observed in arginine[87] or lysine[88] containing peptides (Figure **25**).

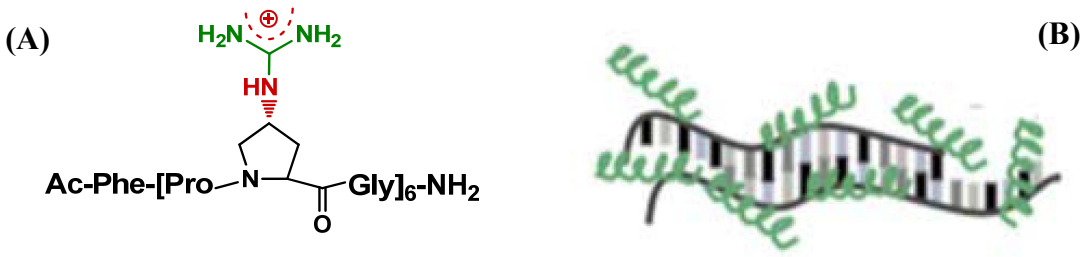


FIGURE 25: (**A**) Guanidylated chimeric collagen peptide (**B**) Possible DNA-peptide interaction

In this design a continuous collagen sequence of (Xaa-Yaa-Gly) type is maintained to ensure a well-folded structure with modifications to the side chain of

proline to incorporate the charged guanidinium groups, which have been shown to be superior to amino groups as the cationic moiety needed for cellular uptake.[89]

In this case chimeric collagen peptides were made that have amino and guanidinium groups in every third position. The influence of primary as well as the secondary structures, determined by CD spectroscopy, on their cellular uptake were investigated. The CD and UV-T_m studies regarding the self-assembly properties of this family of peptides may highlight the possible role of aggregated species in the interaction process. In order to understand the nature of chimeric collagen–DNA complex structure and the molecular and environmental determinants of their association processes, biological as well as biophysical experiments were planned.

Sweet arrow peptide (SAP) a well known CPP[90] was choosen as a control peptide to verify the DNA binding properties under all working conditions along with other synthesised chimeric collagen peptides.

The specific objectives of this section are

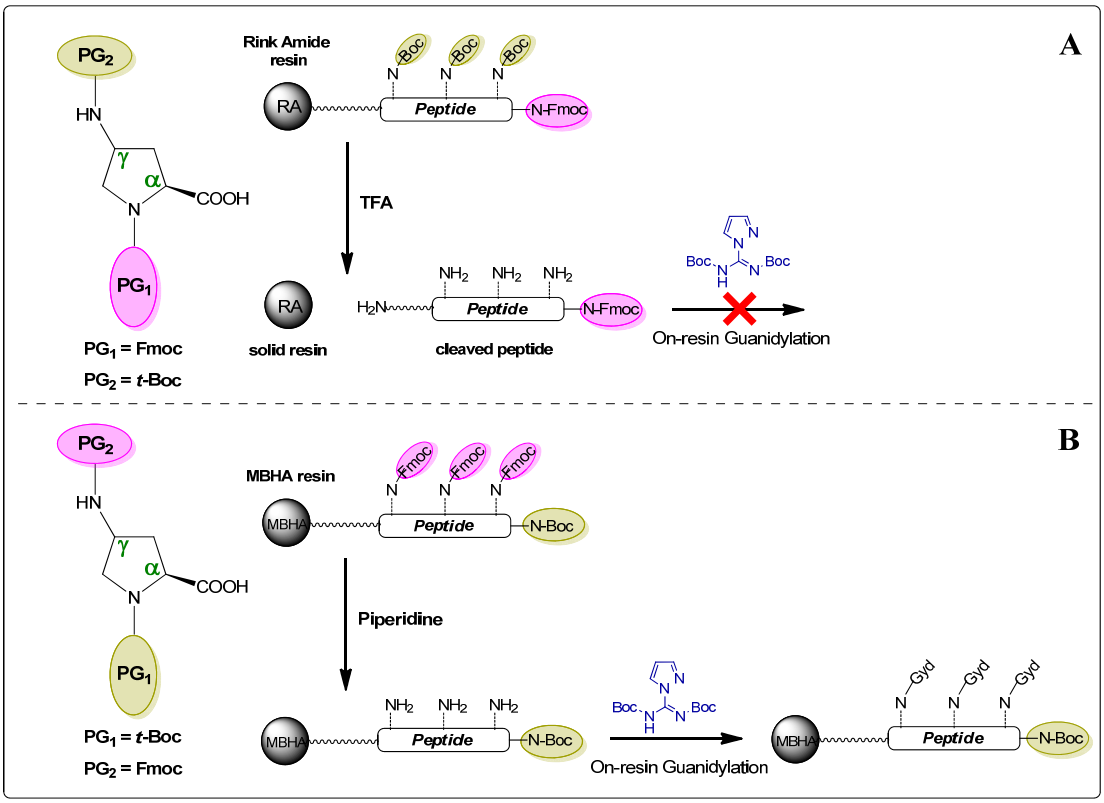
1) Synthesis of (2*S*,4*R*)-N$^{\alpha}$-(*t*-Boc)-4NH-Fmoc-aminoproline (**6**) and (2*S*,4*S*)-N$^{\alpha}$-(*t*-Boc)-4NH-Fmoc-aminoproline (**12**) monomers.
2) Incorporation of these monomers in collagen chimeric peptides using 4Amp, and 4amp at Y and X position respectively through solid phase synthesis followed by on-resin conversion of amino to guanidium functionality.
3) Solid phase synthesis of Sweet Arrow Peptide (SAP) to be used as a control for biological as well as biophysical experiments.
4) Cleavage of these peptides from the solid support, purification and characterization.
5) Investigation of conformation of these peptides using temperature dependent CD spectropolarimetry.
6) Study of their interaction with oligopeptides through UV-T_m, CD spectropolarimetry, fluorescence quenching of EtBr and non-denaturing PAGE.
7) Study of their interaction with plasmid DNA through agarose gel retardation, CD spectropolarimetry, fluorescence quenching of EtBr, isothermal titration calorimetry.
8) Test of transfecting ability of these peptides along with plasmid DNA into S2 cells.
9) Labelling of red colored flurophore to N-terminus of amino-collagen peptides and check their cell permeation ability into S2 cells after subsequent guanidylation.

1.9 Results

1.9.1 Synthesis of monomers

1.9.1a Design of monomers

4-Aminoproline is a versatile building block with two stereocenters, resulting in a stereochemical diversity of four isomers. This amino acid possesses two amino groups and these groups when linked to the α-carboxyl group can provide both α- and γ-peptides.



SCHEME 1.1: (**A**) Non-viable synthetic route with use of N^{α}-Fmoc and N^{γ}-Boc protected monomers, (**B**) Proposed synthetic route for synthesis of required peptides with N^{α}-Boc and N^{γ}-Fmoc protected monomers

In the present case, use of 4-aminoproline in the synthesis of collagen like structures is described, wherein they form α-peptide bonds with glycine and proline. In order to introduce this amino acid into peptide sequence, orthogonal protection (PG_1 and PG_2) of the two amino functionalities is a prerequisite. PG_1 is a primary/temporary protecting group that is removed in each step of growing peptide chain, while PG_2 is a secondary/semi-permanent protection which is retained through

the peptide synthesis and removed at the end as per design. Hence two easily detachable protecting groups *t*-Boc and Fmoc were choosen to design the monomers. The regiospecificity of these protecting groups in the monomers depends on the rationale illustrated in above scheme **1.1**.

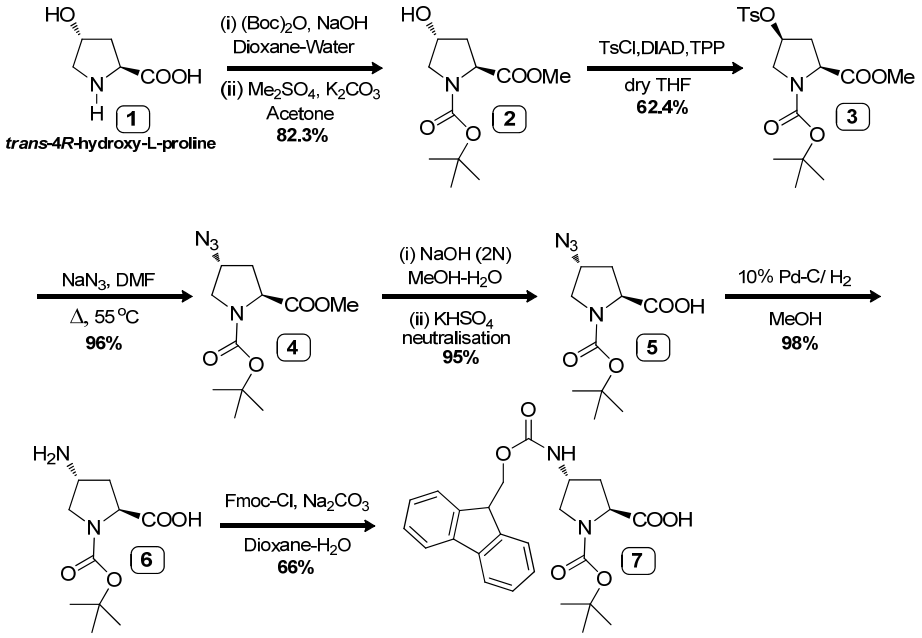
To achieve the synthesis of peptides (**P1**-**P4** & **P6**-**P9**) on solid phase using t-Boc chemistry, monomers **7** and **12** were chosen with a combination of N^{α}-Boc and N^{γ}-Fmoc protection. Although the solid phase peptide synthesis using Fmoc strategy offers several advantages, it would not be appropriate for the present synthetic route. As in case of Fmoc-strategy, using N^{α}-Fmoc and N^{γ}-Boc protected aminoproline, the peptide would need to be deprotected of all the N^{γ}-Boc groups (side chain protection) with 50% TFA/DCM at least prior to on-resin guanidylation, thus cleaving the whole peptide from the resin simultaneously.

1.9.1b Synthesis of fully protected (2*S*,4*R*)- and (2*S*,4*S*)-4-aminoproline monomers

Synthesis of orthogonally protected (2*S*,4*R*) and (2*S*,4*S*) 4-aminoproline monomers (**7** and **12**) were achieved in six steps from the naturally occurring trans-4-hydroxyproline **1** (Scheme **1.2** & **1.3**). The free imino group of 4*R*-hydroxyproline **1** was treated with bis-*t*-butoxycarbonyl anhydride in presence of NaOH as base in water-dioxane mixture to give N^{α}-*t*-Boc protected 4*R*-hydroxyproline, the carboxyl group of which was protected as methyl ester **2** using dimethyl sulfate and potassium carbonate in anhydrous acetone. The integrity of compound **2** was verified with characteristic signals in the ^{1}H NMR spectrum at δ 3.73 ($-OCH_3$) and δ 1.41 & 1.46 (*t*-Boc). Protection of pyrrolidine ring nitrogen resulted in the tertiary amide **2** which showed two sets of peaks - corresponding to a major and a minor isomer - in its NMR spectra arising due to presence of the *cis*-*trans* isomers for rotation about the >N-CO bond.

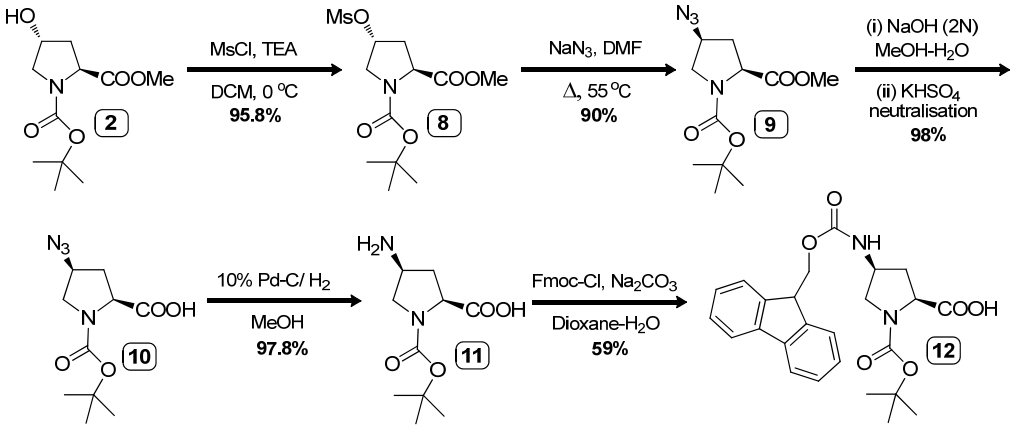
The 4*R*-OH group of compound **2** was converted into 4*S*-O-tosylate **3** using Mitsunobu reaction with DIAD/PPh_3 and methyl-p-toluenesulfonate, resulting in an inversion of configuration at γ-C (^{1}H NMR δ 7.33 – 7.79, aromatic protons; δ 4.30-4.47, C^{γ} – proton). The reaction of tosylate **3** under S_N2 condition with NaN_3 in DMF resulted in a second inversion of stereochemistry at γ-C, yielding (2*S*, 4*R*)-N^{α}-Boc-azidoprolinemethylester **4**. A peak at 2105 cm^{-1} characteristic of azide was seen in IR spectrum of **4**. The methyl ester group of compound **4** was hydrolyzed with aq. NaOH

(2N) in methanol to give free acid **5** to be confirmed by disappearance of –OC<u>H</u>$_3$ δ 3.71 from the ^{1}H NMR spectrum. This azide group of this resulting acid was selectively reduced to the corresponding amine **6** using 10% Pd-C under hydrogenation conditions. The amine was then protected with Fmoc, by reaction with 9-fluorenylmethylchloroformate to yield the fully protected 4*R*-aminoproline (Amp) **7**.



SCHEME 1.2: Synthesis of (2*S*,4*R*)-N^1-(t-butoxycarbonyl)-4-(9-fluorenylmethyloxycarbonylamino) proline

The appearance of signals in the aromatic region typical of Fmoc group [δ 7.31-7.35 (m, 2H), 7.39-7.43 (t, 2H), 7.64 (bs, 1H), 7.68-7.7(m, 2H), 7.87-7.89 (d, 2H); Ar<u>H</u> & C<u>H</u> of Fmoc], peaks at 453 [M+H], 475 [M+Na] and 491 [M+K] in the LCMS confirmed the structural integrity of the monomer **7** (Scheme **1.2**).



SCHEME 1.3: Synthesis of (2*S*,4*S*)-N^1-(t-butoxycarbonyl)-4-(9-fluorenylmethyloxycarbonylamino) proline

The 4-OH group of **2** was also converted to the corresponding mesyl derivative **8** by treatment with methanesulfonylchloride, in DCM in presence of Et_3N (^{1}H NMR: δ 3.04, s, 3H, –SO_3CH_3). The treatment of mesylate compound **8** with NaN_3 in DMF at 55-60 ^oC resulted in an S_N2 displacement of 4-O-mesyl group to yield the azide **9**. A peak at 2106 cm^{-1} corresponding to azide functional group seen in IR spectrum confirmed the conversion of mesyl to azide. The methyl ester of compound **9** was hydrolysed to acid **10** with aq. NaOH (2N) in methanol which was verified from its ^{1}H NMR spectra where the δ 3.71 & peaks for –OCH_3 vanished. Hydrogenolysis of azide to amine was achieved by catalytic hydrogenation of **10** with H_2-Pd/C to yield compound **11** to be confirmed from disappearance of azide peak in IR at 2106 cm^{-1}. The γ-amino acid **11** was reacted with fluorenylmethylchloroformate in dioxane:water with Na_2CO_3 as base to yield the fully protected monomer **12**. Signals in the aromatic region, typical of Fmoc group [δ 8.5-8.55 (t, 2H), 8.61-8.65 (t, 2H), 8.75-8.77 (d, 1H), 8.97-8.99 (d, 2H), 9.2-9.23 (d, 2H), ArH & CH of Fmoc] and LC-MS peak at 475 [M+Na] confirmed the formation of orthogonally protected (2*S*,4*S*) 4-aminoproline monomer (Scheme **1.3**).

1.9.2 Solid phase synthesis

General principles

Peptide synthesis can be carried out either by solution phase or by solid phase methods.[91] In the solution phase method, the *C*-terminal amino acid with a carboxyl protection (amine free) is coupled with the second N^{α}-protected protected (carboxyl free) amino acid. Coupling reactions are performed by an *in situ* activation of the carboxylic acid using carbodiimide coupling reagents such as dicyclohexylcarbodiimide (DCC).[92] In these reactions, HOBt is often used as both a catalyst and racemization-suppressing agent.[93] Alternatively, preactivated amino acid esters with pentafluorophenyl (pfp)[94] or 3-hydroxy-4-oxo-3,4-dihydro-1,2,3-benzotriazole (DhBt)[95] ester may be used. The solution phase method however, requires a tedious separation step after each coupling reaction, thus preventing the usage of large excess of the carboxylic acid component.

The solid phase synthesis devised by Merrifield,[96] involves the combination of reagents with functional groups that are located on the surface and on the inside of beaded polymers.[97] The beads are immersed in solvent containing the reagents, which approach the solvated sites by diffusion. Success is contingent on the sites

being accessible to the reagents. In contrast to solution phase method, the solid phase method offers several advantages.[91] In this method, the *C*-terminal amino acid is linked to an insoluble support that also acts as a permanent protection for the carboxylic acid (Figure **26**). Then the next N^{α}-protected amino acid is coupled to the resin bound amino acid either by using an active pfp or DHBt ester, or by *in situ* activation with carbodiimide reagents. The excess amino acid is washed out and the deprotection and coupling reactions are repeated until the desired peptide sequence is achieved. The need to purify intermediates at every step is obviated. Finally, the resin bound peptide and the side chain protecting groups are cleaved in one step.

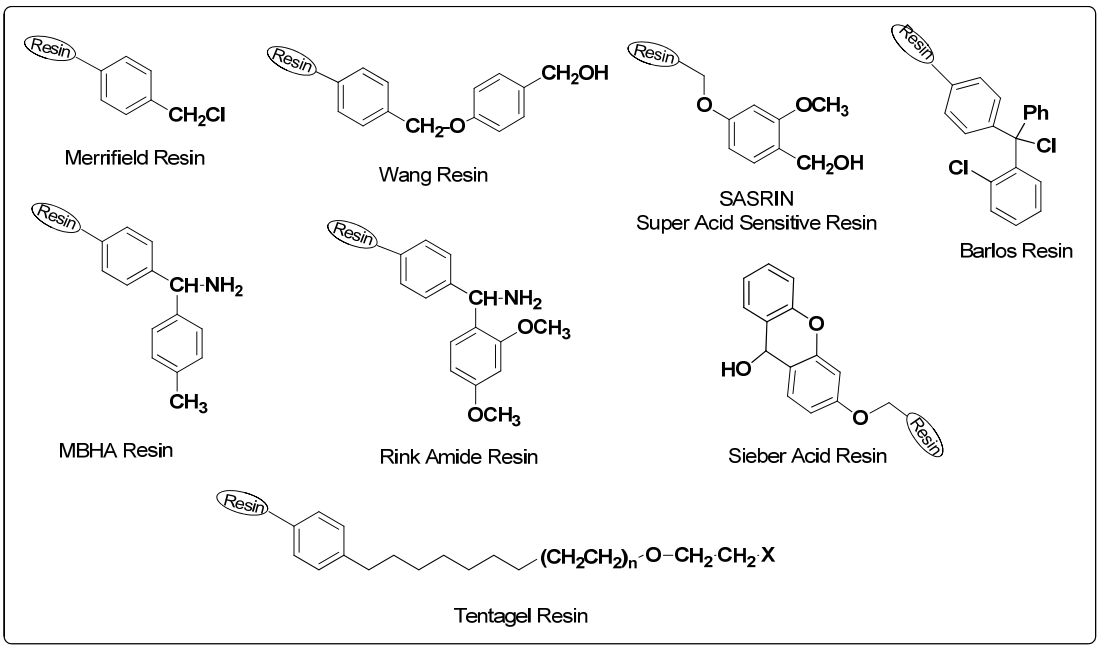


FIGURE 26: Few representative structures of resin used in solid phase peptide synthesis

There are several features that are characteristic of solid-phase synthesis. There is no loss of material accompanying the synthetic steps because the peptide is never taken out of the vessel. Secondary products arising from the reagents and protectors do not affect the purity of the peptide because they are quantitatively removed after each reaction. Reactions can be encouraged to go to completion by the use of excess of reagents. The question of the solubility of the peptide does not arise as it does for the synthesis carried out in solution. No purification of the intermediates is effected during the synthesis, and operations are repetitive and have been automated.

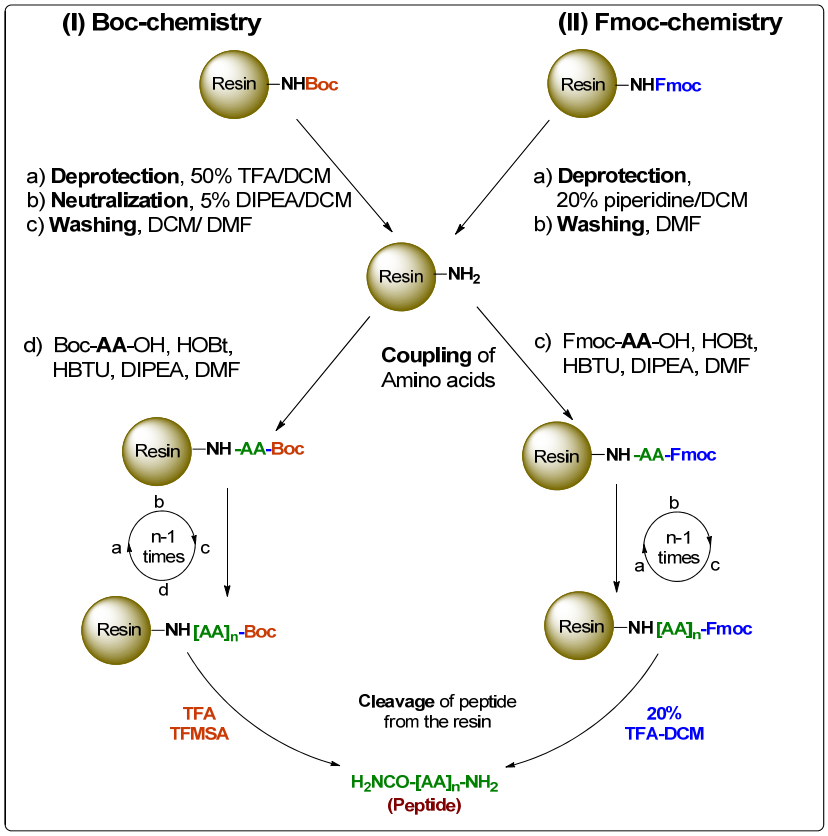


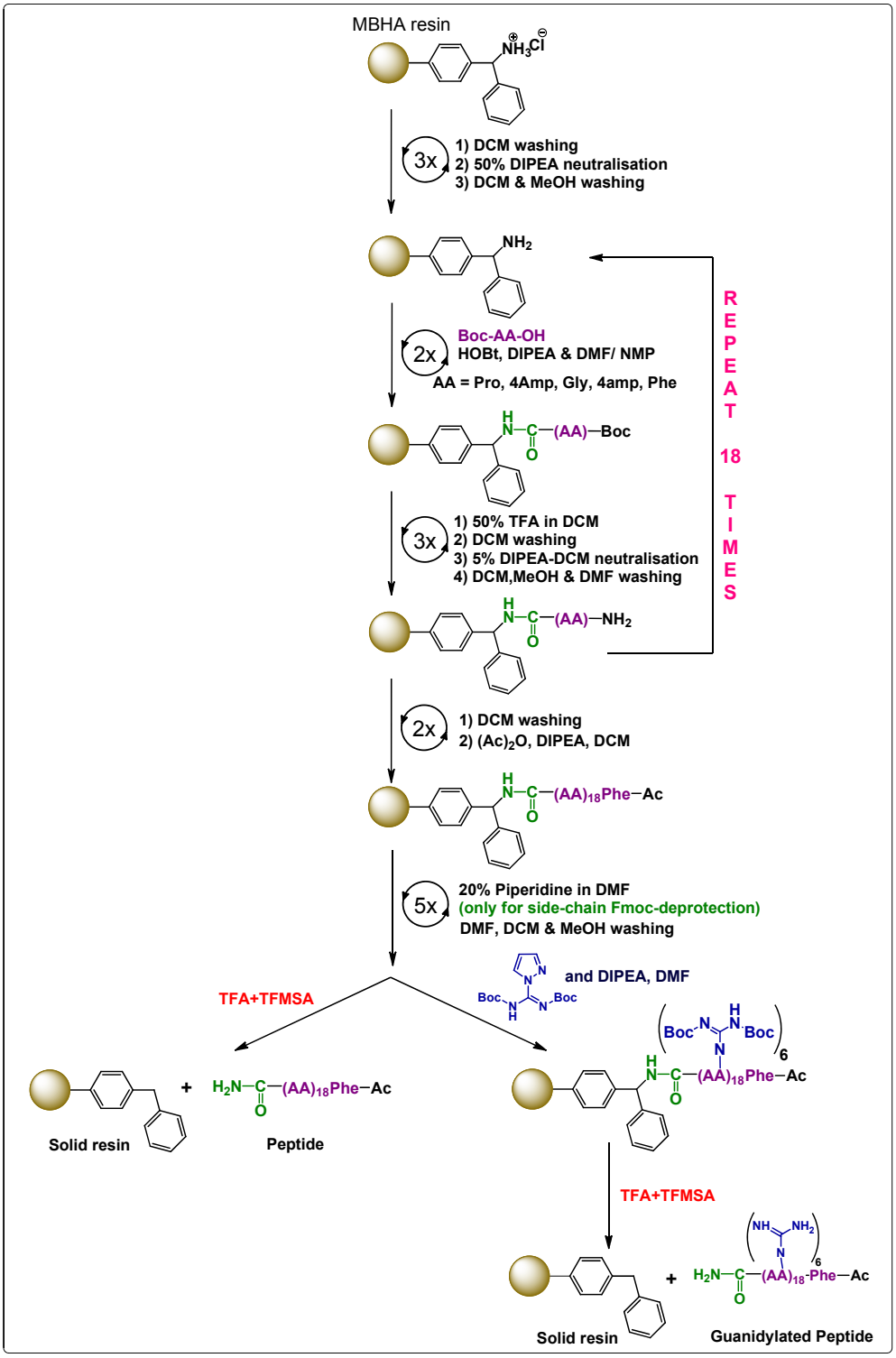
FIGURE 27: General protocol for SPPS via (**I**) Boc-chemistry (**II**) Fmoc chemistry

Two major protocols are followed for the routine synthesis of peptides by solid phase peptide synthesis –***Fmoc* strategy** (base-labile alpha-amino protecting group) and ***Boc* strategy** (acid-labile protecting group) (Figure **27**). First protocol uses the *t*-butoxycarbony (*t*-Boc) group as N^{α}-protection that is removed by mild acidic conditions such as 50% TFA in DCM. The reactive side chains are protected with groups that are stable to *t*-Boc deprotection conditions and can be removed under strongly acidic conditions using HF in dimethylsufide or TFMSA in TFA. The alternative protocol is to use fluorenylmethyloxycarbonyl (Fmoc) group as N^{α}-protection which is extremely stable to acidic conditions but can be cleaved efficiently with a secondary base such as piperidine. Coupled with acid (50% TFA) cleavable side chain protection this method offers a milder strategy for the SPPS. Fmoc chemistry is also known for generating peptides of higher quality and in greater yield than Boc-chemistry. In both chemistries, the linker group that connects or binds the peptide to the resin is chosen such that the side chains and the linker are cleaved in one step. But the main limitation of solid phase peptide synthesis is that it cannot be applied for synthesizing peptides routinely in large scale.

1.9.2a Synthesis of cationic peptides

In the present work, peptides (**P1**-**P4**) were synthesized by manual solid phase synthesis on readily available MBHA (4-methyl–Benzhydryl amine) resin using standard *t*-Boc protocol from the *C*-terminus to the *N*-terminus using orthogonally protected N^{γ}-Fmoc and N^{α}-*t*-Boc monomeric units (Scheme **1.4**), which upon cleavage directly yielded the peptide-C-terminal amide. Monomers having two amino groups were orthogonally protected, Boc-group at main site of elongation and Fmoc-group at sidechain. The loading value of the resin (0.8 g/mole) was unaltered for this synthesis.

The hydrochloride salt of MBHA resin was neutralised with 50% DIPEA-DCM and the monomers were coupled as free acids using *in situ* activation procedure with 3 eq. of aminoacid, HBTU as a coupling reagent and DIPEA, HOBt as catalyst and recemization-suppressant respectively. Subsequently the resin bound *Boc*-group was cleaved with 50% TFA/DCM before coupling the next amino acid. The deprotection and coupling reactions were monitored using qualitative Ninhydrin (Kaiser)[98] test for Gly and Chloranil test[99] for iminoacids. In case of a positive color test after the first coupling which indicates incomplete reaction, the coupling was repeated. To avoid deletion of sequences, a capping step with Ac_2O/ DIPEA in DCM was performed. The terminal amino group of the final peptide were also capped with Ac_2O and the side chain fmoc protection removed with 20% piperidine/DMF. Peptides **P1** and **P3** were cleaved from the resin at this stage using TFA and TFMSA while for peptides **P2** and **P4** prior to cleavage all 4-NH_2 prolyl function were converted into corresponding guanidium function on resin using the reagent *N*,*N*'-bis-Boc-1*H*-pyrazole-1-carboxamidine in DMF.



SCHEME 1.4: Solid phase synthesis of peptides using *t*-Boc chemistry

The control peptide **P5** (SAP) was synthesised on an automated synthesizer pursuing Fmoc-chemistry and using commercially available *Fmoc*-protected amino acids. This peptide was cleaved from its resin (Rink Amide) with 20% TFA in DCM and 0.1% TIS (Triisopropylsilane) used as a scavenger.

The following Figure **29** shows the synthesised peptides (**P1**-**P5**) used for current biophysical and biological studies.

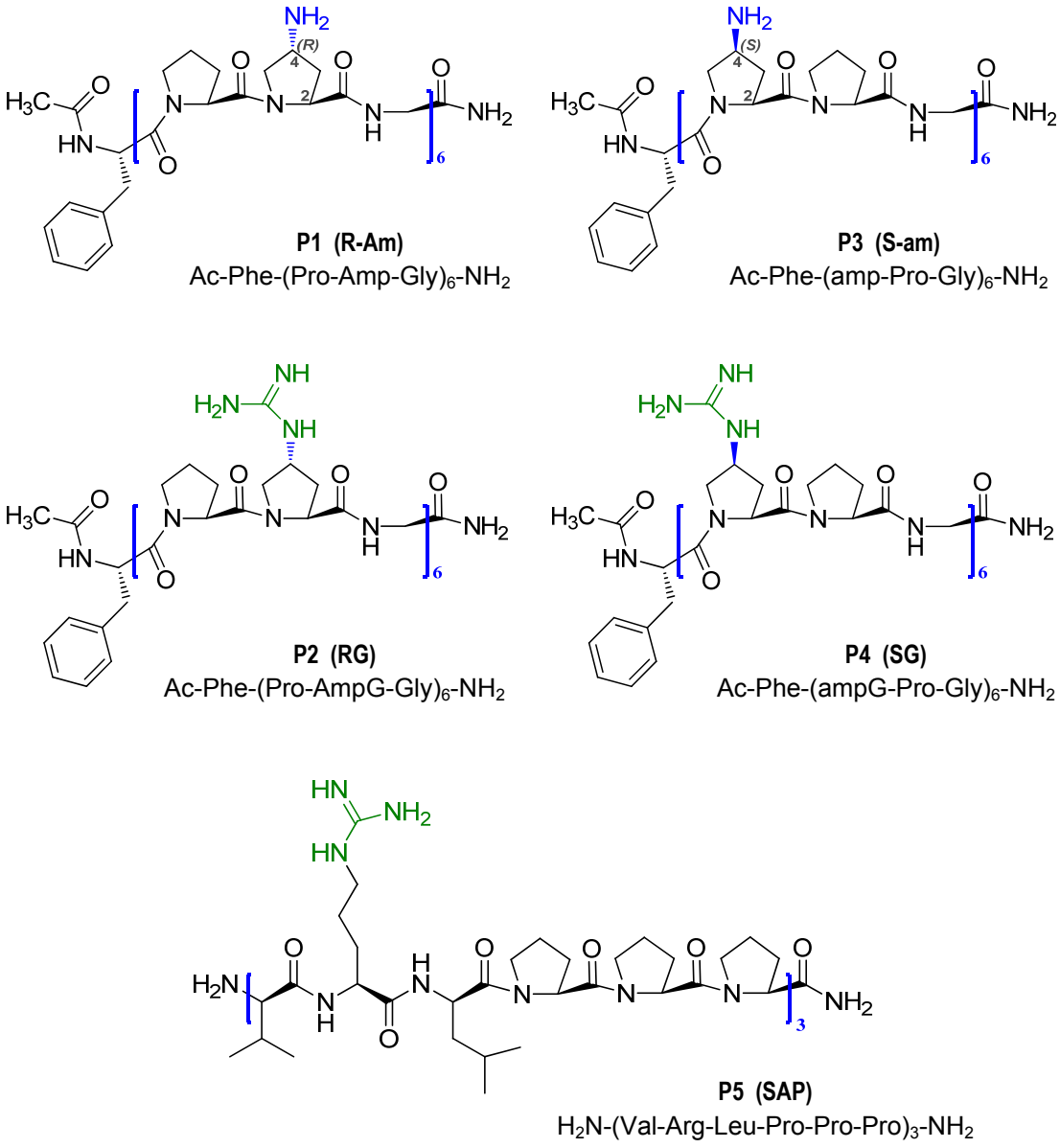
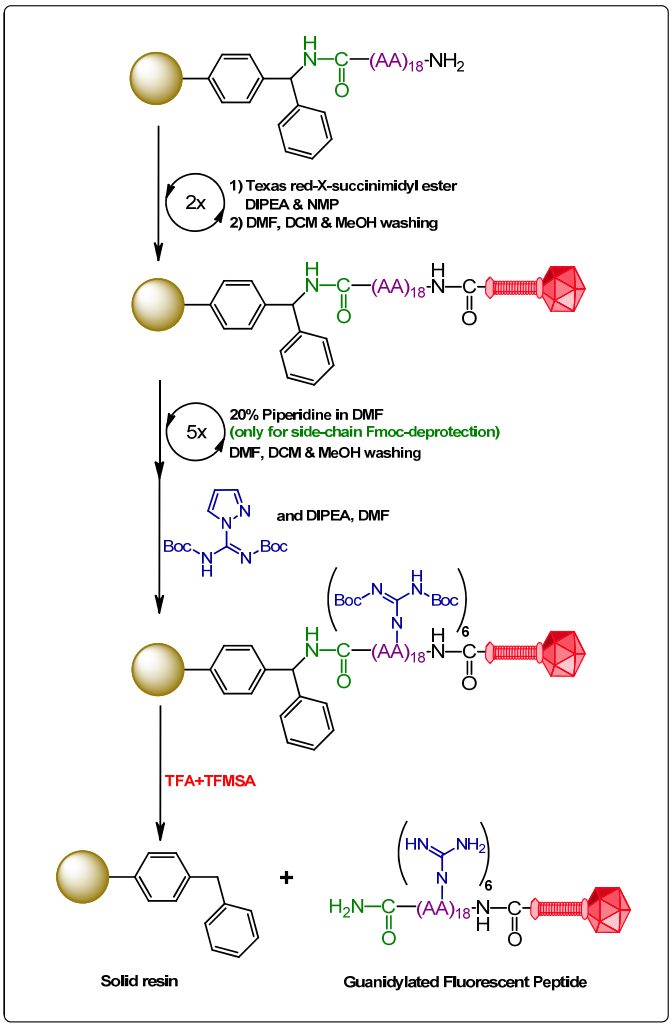


FIGURE 28: Peptides used in the present study. **P1**-**P4** peptides are end capped, while peptide **P5** is with free *N*-termini

1.9.2b Synthesis of fluorescent cationic peptides

Two peptides **P6** and **P7** were also synthesized with identical sequence of amino acids as that of peptides **P1** and **P3** respectively, except N-acetyl ion of phenylalanine group at the N-terminal. Small portions of it were cleaved from resin to verify the structural integrity at this juncture of synthesis. These peptides were then tagged with fluorescent Texas red group by coupling them with Texas red-X-

succinimidyl ester in presence of catalytic amount of DIPEA while they were still anchored to the resin (Scheme **1.5**).



SCHEME 1.5: Solid phase synthesis of fluorescent peptides and subsequent guanidylation

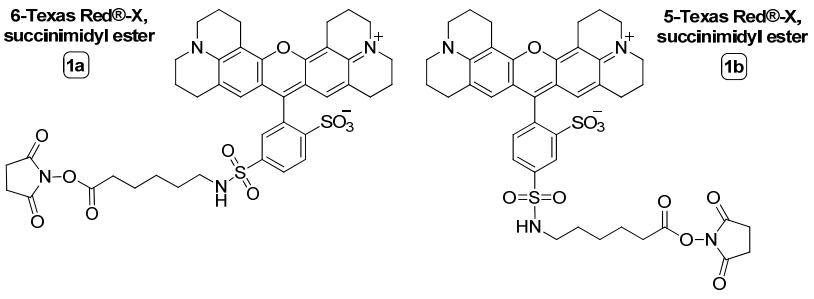
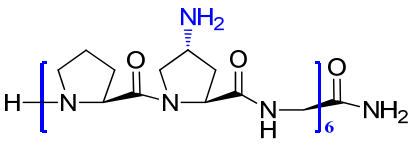


FIGURE 29: Regioisomers of Texas Red dye

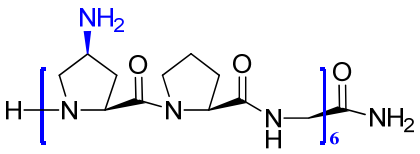
Texas red[100] is a red fluorescent dye which is used to create bright red-fluorescent bio-conjugates having application as tracer agents. It refers to two fluorescent dyes (**1a** and **1b**) with λ_{ex} at 595 nm and λ_{em} at 615 nm, respectively (Figure **28**). The dyes are membrane impermeant and are widely used as molecular

labels.[101] This reactive dye contains an additional seven-atom aminohexanoyl spacer ("X") between the fluorophore and the succinimidyl ester group. This spacer helps to separate the fluorophore from its point of attachment, potentially reducing the interaction of the fluorophore with the biomolecule to which it is conjugated.

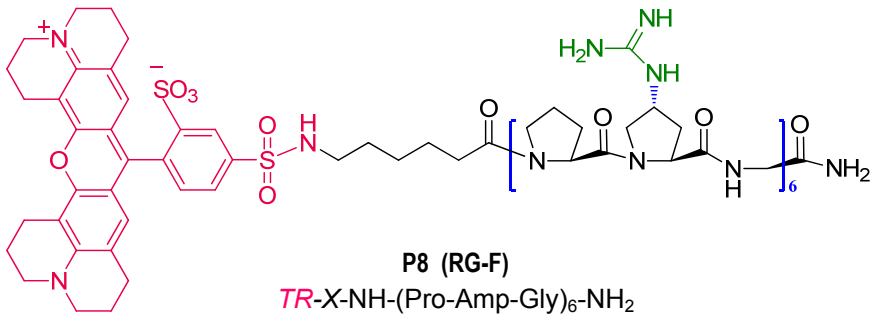
Figure **30** gives the list of fluorescent peptides synthesised along with their precursors.



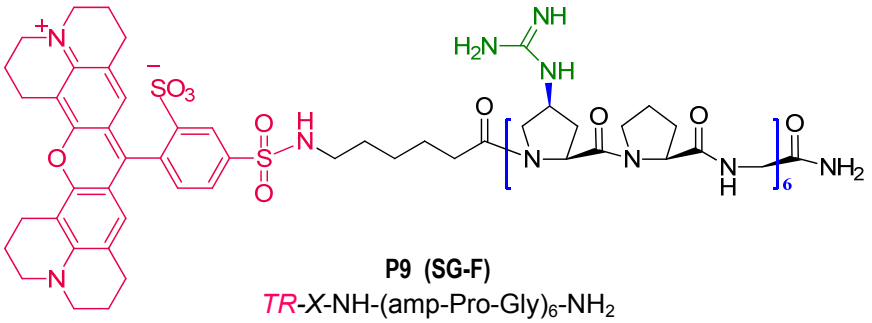
P6 (RA)
HN-(Pro-Amp-Gly)$_6$-NH$_2$



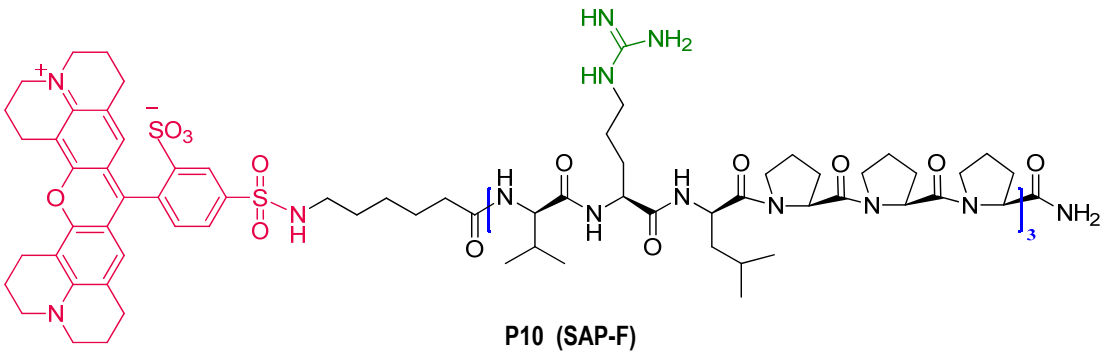
P7 (Sa)
HN-(amp-Pro-Gly)$_6$-NH$_2$



P8 (RG-F)
TR-*X*-NH-(Pro-Amp-Gly)$_6$-NH$_2$



P9 (SG-F)
TR-*X*-NH-(amp-Pro-Gly)$_6$-NH$_2$



P10 (SAP-F)
TR-*X*-NH-(Val-Arg-Leu-Pro-Pro-Pro)$_3$-NH$_2$

FIGURE 30: Fluorescent peptides (**P8-P10**) used in the present study. *N*-terminal free Peptides **P6** and **P7** are precursor for peptide **P8** and **P9** respectively.

The integrity of peptides was confirmed by MALDI-TOF mass spectrometric analysis. The HPLC retention time of the synthesized peptides is shown in Table **4**. The observed molecular weight has been mentioned in the same table along with the calculated molecular weight and the molecular formula of all the peptides. The HPLC profile and MALDI-TOF data for confirmation of the synthesized control as well as modified peptides are shown later in Appendix A.

TABLE 4: HPLC retention time and MALDI-TOF mass spectral analysis of the synthesized peptides

Peptide	Sequence Code	HPLC t_R # (min)	Mol. Formula	$M_{(Calcd)}$*/ $M_{(Obsd)}$
P1	R-Am	15.02	$C_{83}H_{122}N_{26}O_{20}$	1804.0190 / 1804.0044
P2	RG	15.99	$C_{89}H_{134}N_{38}O_{20}$	2056.2589 / 2056.8967
P3	S-am	14.81	$C_{83}H_{122}N_{26}O_{20}$	1804.0190 / 1805.0488
P4	SG	15.68	$C_{89}H_{134}N_{38}O_{20}$	2056.2589 / 2056.8789
P5	SAP	21.56	$C_{96}H_{162}N_{28}O_{18}$	1996.4903 / 1996.7594
P6	RA	9.61	$C_{72}H_{111}N_{25}O_{18}$	1614.8084 / 1614.3934
P7	Sa	10.86	$C_{72}H_{111}N_{25}O_{18}$	1614.8084 / 1614.5562
P8	RG-F	21.27	$C_{124}H_{171}N_{41}O_{26}S_2$	2716.0736 / 2717.8298
P9	SG-F	20.92	$C_{124}H_{171}N_{41}O_{26}S_2$	2716.0736 / 2716.3008
P10	SAP-F	24.39	$C_{133}H_{201}N_{31}O_{25}S_2$	2698.3417 / 2699.1646

Solvent gradient used in HPLC are not uniform for all the peptides. Details are mentioned in Appendix A. * All the molecular weight has been calculated by Chem BioDraw 12.0

The synthesized fluorescent peptides were analyzed for their excitation and fluorescence emission spectra (Figure **31**). Texas Red shows excitation spectrum at around 595 nm and emission spectra at around 615 nm.

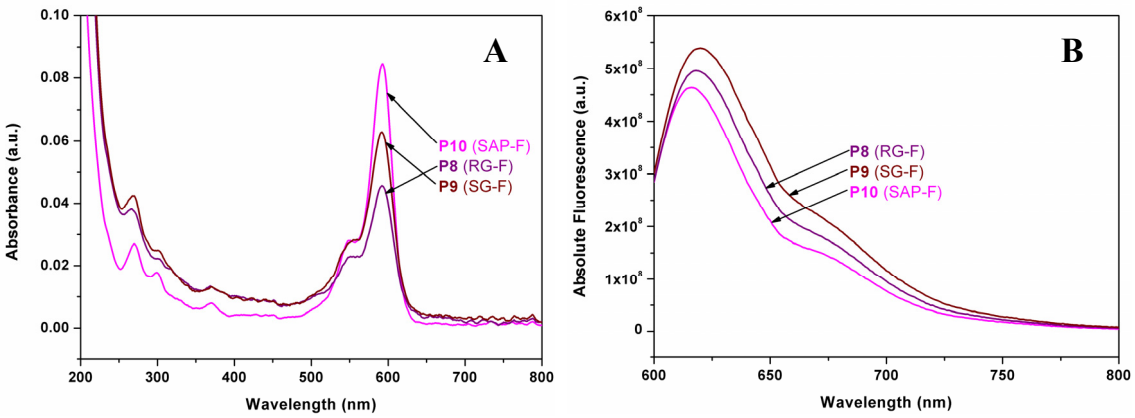


FIGURE 31: (**A**) UV-absorption spectra of *TR-X*-labelled peptides (**B**) their corresponding fluorescence emission spectra at 615 nm

All the fluorescent peptides showed maximum absorbance at 592 nm because of texas red moiety tagged to it (Figure **31.A**). On excitation of fluorescent peptide samples at 592 nm, the corresponding fluorescence emission spectra shows band near to 616-618 nm in accordance to the presence of texas red (Figure **31.B**).

The phenylalanine group attached at the end of the non-fluorescent peptide sequences served the purpose of estimating their concentration in the stock-solutions at 260 nm, with known extinction coefficient (ε_{260} = 2.00 x 10^3 M^{-1} cm^{-1}), whereas for the synthesized fluorescent peptides the absorbance from fluorophore texas red (ε_{595} = 8 x 10^4 M^{-1} cm^{-1}) was used as a measure for their concentrations.

1.9.2c Synthesis of oligonucleotides

A random series of both pyrimidine (A & T) and purine (C & G) rich sequences, along with their complementary strands were chosen for the biophysical studies of peptides, comprising of 18 bases each. These DNA oligonucleotides **D1**-**D4** (Table **5**) were synthesized on Applied Biosystems ABI 3900 High Throughput instrument using standard phosphoramidite chemistry.[102] The details of oligomer synthesis, HPLC profile and MALDI-TOF characterization are mentioned later in Appendix A.

TABLE 5: HPLC retention time and MALDI-TOF mass spectral analysis of the DNA oligonucleotides

Oligo	Seq. Code	Sequences (5' to 3')	HPLC t_R (min)	$M_{(Calcd)}$*/$M_{(Obsd)}$	Remark
D1	DNA 1	CGTTATAATATTATAAGC	10.82	5496.7/ 5500.3	AT-rich
D2	DNA 1a	GCTTATAATATTATAACG	10.50	5496.7/ 5497.1	
D3	DNA 2	ATGGCGCCGCGGCGCCTA	9.04	5501.6/ 5501.0	CG-rich
D4	DNA 2a	TAGGCGCCGCGGCGCCAT	9.03	5501.6/ 5503.0	

∗ All the molecular weight has been calculated by Oligo Calc tool available on internet[103]

In designing the duplex one set of sequences was taken with 22% purine/ GC-content (**D1:D2**) while another with 78% GC content (**D3:D4**) for investigating the sequence specificity of binding in synthesized peptides.

1.9.3 Biophysical Studies

The physical principles underlying all processes of living systems constitute the subject of biophysics. It includes the understanding of interactions of various physical influences on physiological functions. Biophysics is an interdisciplinary science which applies the techniques from the physical sciences in order to understand the biological structure and their functions. These techniques are useful in studying the structure and properties of nucleic acids, proteins, peptides and their analogues.[104]

As discussed in the previously, the amino functionality of chimeric collagen sequence has been rationally functionalized to guanidinium group using on-resin modification strategy, with different stereochemistry. This section focuses on the biophysical studies which has been carried out on these designed collagen analogues in order to investigate the effect of amino and guanidinium groups on its binding to oligomeric and plasmid DNA with respect to SAP (**P5**) considered as control peptide. The techniques used in the study mainly includes circular dichroism (CD), temperature dependent UV-spectroscopy (UV-T_m), fluorescence and gel electrophoresis. Moreover the secondary structure or conformations of these peptides were also studied by CD spectroscopy.

These studies give information on the binding selectivity, specificity and thermal stability of the modified peptide and DNA chimeras. The following sections elucidate the systematic analysis of the application of biophysical techniques on the novel chimeric collagen synthesized.

1.9.3a Conformational analysis of Guanidino-peptides by CD spectroscopy

Proteins and peptides have a number of groups (chromophores) that have characteristic absorption bands in specific areas of the visible and ultraviolet region. The most numerous and characteristic chromophores in proteins and peptides are the amide bonds (peptide bond) linking the amino acid residues. In the far-ultraviolet region amides have two well-characterized low-energy electronic transitions: the $n \rightarrow \pi^*$ transition, weak in absorption (ε_{max} ~100) but often strong in CD, at about 220 nm, and the $\pi \rightarrow \pi^*$ transition near 190 nm, which is strong in both absorption (ε_{max} ~7000) and CD. These two transitions dominate the CD and absorption spectra of peptides and proteins in the far-ultraviolet region.

The synthesized peptides are chimeric collageneous peptides, with **P2** and **P4** being directly modified sequences of **P1** and **P3** respectively and which are known to show characteristic CD spectra typical of collagen triple helices.[105] In solution, collagen-like triple-helical and polyproline II like structures exhibit fingerprint CD spectra.[106] These spectra show characteristic presence of a large negative band around 200 nm, a crossover at around 213 nm, and a small positive band around 215-227 nm. In order to investigate the conformational behavior of peptides **P2** and **P4**, CD spectra were measured at various concentrations, temperatures and solvent conditions.

(i) Concentration dependent CD spectroscopy

Figure **32(A-B)** shows the CD spectra of guanidino-peptides (2*S*,4*R*) **P2** and (2*R*,4*S*) **P4** at pH 7.2 at various molar concentrations. In case of peptide **P2** the CD spectrum has positive maxima at 221 nm and intense negative maxima at around 200 nm. In the entire concentration range peptide **P2** shows similar positive and negative maxima with the magnitude of positive and negative bands varing as a function of concentration. Importantly, all the spectral traces pass through an isobestic point at 215 nm which also happens to be the cross-over point from

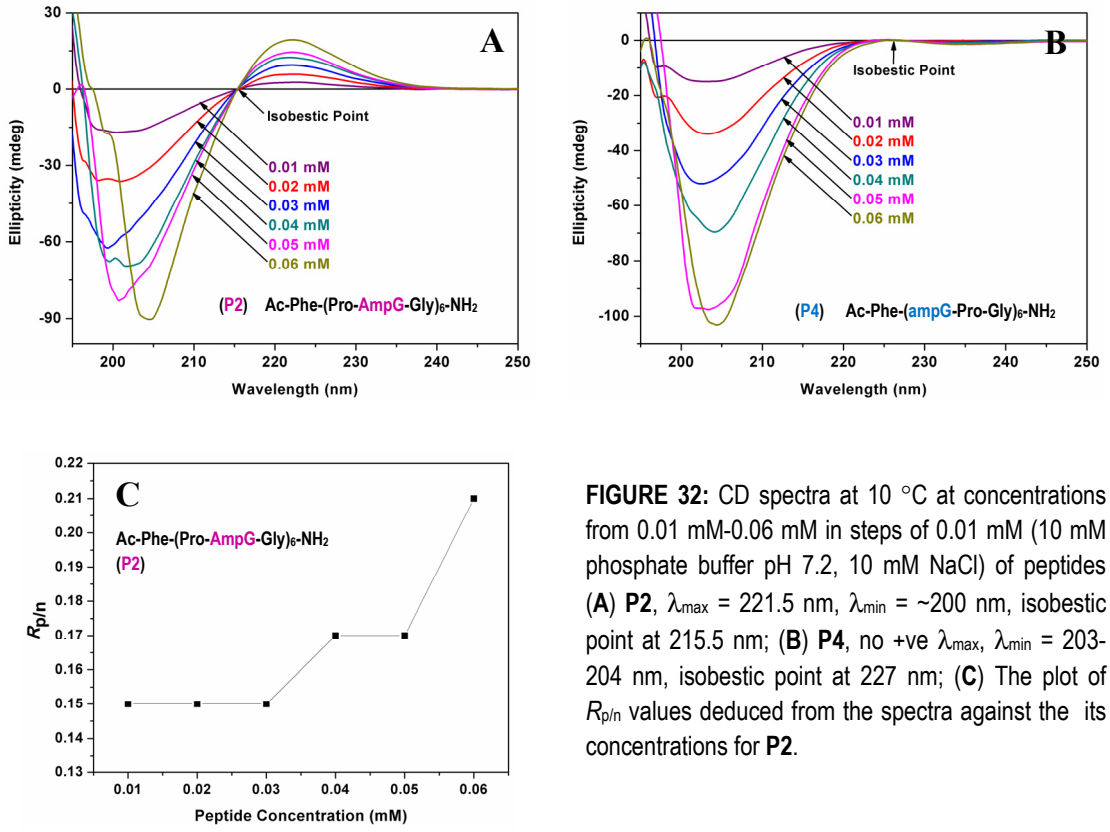


FIGURE 32: CD spectra at 10 °C at concentrations from 0.01 mM-0.06 mM in steps of 0.01 mM (10 mM phosphate buffer pH 7.2, 10 mM NaCl) of peptides (**A**) **P2**, λ_{max} = 221.5 nm, λ_{min} = ~200 nm, isobestic point at 215.5 nm; (**B**) **P4**, no +ve λ_{max}, λ_{min} = 203-204 nm, isobestic point at 227 nm; (**C**) The plot of $R_{p/n}$ values deduced from the spectra against the its concentrations for **P2**.

positive to negative region. Peptide **P4** on the other hand has a strong negative band near about 203-204 nm but does not show any maxima in the positive range. In Figure **32.B** an isobestic point is also detected at 227 nm through which all the spectral lines pass irrespective of the concentration. Spectra could not be satisfactorily measured below 190 nm (vacuum CD region) due to high noise.

This does not comply with the trend observed in case of triple helical peptide **P1**, where $R_{p/n}$ value increases exponentially to saturate at the critical triple-helical concentration and remains constant thereafter.[105a] Hence, 0.03 mM was chosen for performing all the further studies assuming the peptide **P2** to be in fully associated form at this concentration. Peptide **P4** was also taken at the same concentration for ease of comparison of data.

(ii) CD - thermal denaturation studies

In order to verify the conformation of peptides to be either PPII or triple helical CD-thermal denaturation study of peptides **P2** and **P4** were carried out. Upon heating, triple-helical structures generally undergo a cooperative triple-helix to single strand coil transition. This change of conformation is reflected in several CD properties such as the degree of the molar ellipticity typically around λ_{max} in the CD spectra. The thermal denaturation curves so obtained show a decrease in molar ellipticity with increase in temperature for the peptides.

The conformational transition of peptides was therefore monitored by observing the changes in ellipticity at 221.5 nm with temperature. Further, thermal denaturation study was carried out for the peptides **P2** & **P4** at both neutral and basic conditions along with that in ethylene glycol-water (3:1) mixture. The temperature dependent CD at basic pH 12 was done since the pK_a of guanidine groups in peptides **P1** & **P4** are near about 13, at which a greater proportion of protonated states can be expected. Polyols and sugars are known to offer protection against thermal denaturation to most proteins,[107] including collagen triple-helix.[108] Ethylene glycol stabilizes the helical structure and therefore can be very useful in amplifying and detecting very weak triple-helical propensities.[109]

Figure **33** shows the CD-thermal denaturation curves of peptides **P2** & **P4** at different conditions. The peptides (0.03 mM) were taken in respective buffers at different pH and the molar ellipticity was monitored at 222 nm, from 4 ^{o}C to 85 ^{o}C.

The T_m values were obtained from the first derivative of the sigmoidal fitted molar ellipticity values.

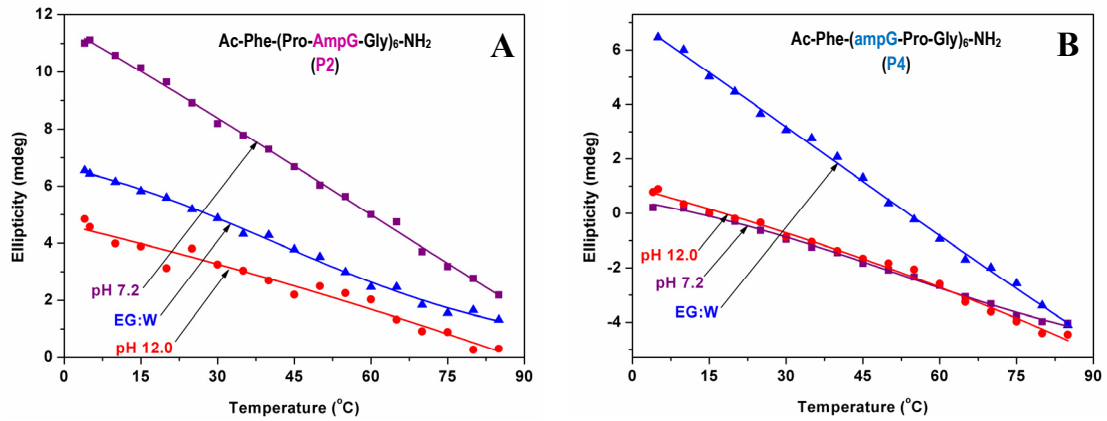


FIGURE 33: Boltzman/Sigmoidal fit curve for molar ellipticity at 222 nm, from variable temperature CD spectra in different solvent conditions of pH 7.2, 12.0 (10 mM phosphate buffer & borate buffer with 10 mM NaCl respectively) and ethylene glycol –water mixture (3:1) for peptides (**A**) **P2**, (**B**) **P4** at 0.03 mM concentration

At pH 12.0, both the peptides (**P2** & **P4**) show a linear decrease in ellipticity with increasing temperature At pH 7.2 and in EG:W system, the T_m values could be deduced from their first derivative curves although the rate of change of molar ellipticity with temperature was not very sharp. The thermal denaturation profile of these two peptides is summarized in Table **6**.

Table 6: Thermal transition temperature of peptides in both aq. and ethylene glycol systems

Peptide Sequence		T_m (°C)		
		pH 7.2	pH 12.0	EG:W
Ac-Phe-(Pro-AmpG-Gly)$_6$-NH$_2$	**P2**	58 °C	-nd-	41.6 °C
Ac-Phe-(ampG-Pro-Gly)$_6$-NH$_2$	**P4**	50.6 °C	-nd-	49 °C

(iii) CD study of salt effect on peptides

The interaction of protein molecules with solvent or with other molecules is determined primarily by its surface, with most favorable interactions provided by the charged or polar residues. Additives such as salts affect the solubility and stability of secondary and tertiary structure of proteins. The increase in ionic strength due to presence of salt decreases the electrostatic free energy of the protein and enhances its structural stability. At high concentrations, salts decrease the solubility of proteins and denature the folded structure by disrupting the structure of water.[110]

In order to evaluate the role of salt in conformations of **P2** and **P4** CD spectra of both guanidino peptides were carried out at pH 7.2 with varying concentrations of NaCl.

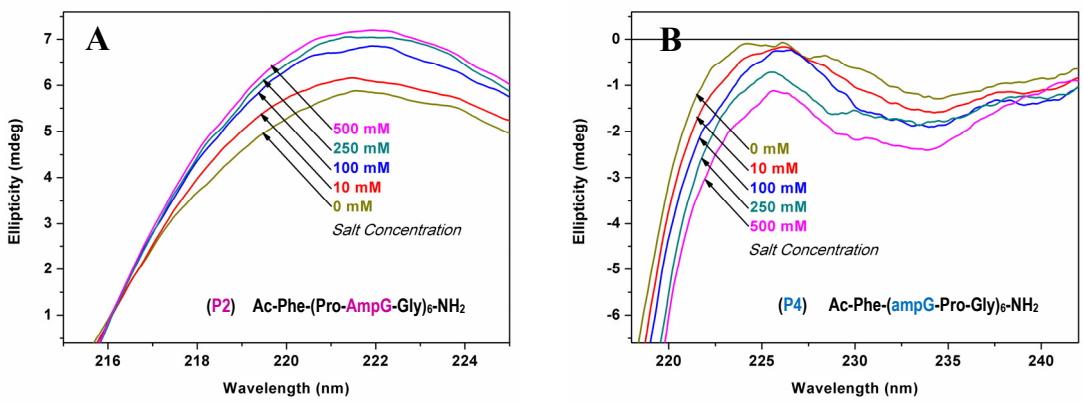


FIGURE 34: CD spectra at various salt (NaCl) concentrations for peptide (**A**) **P2** and (**B**) **P4** in phosphate buffer pH 7.2 (10 mM)

Figure **34** shows the plot of ellipticity of peptide **P2** and **P4** against concentration of NaCl. Concentration of salt had similar effects on both peptides. In case of **P2**, the increase in salt concentration increased the intensity of positive and negative peaks. The negative peak intensities increased uniformly for peptide **P4** over the entire salt concentration range.

1.9.3b UV-T_m studies on Interaction of peptides with *ds*DNA

DNA duplexes and triplexes undergo a denaturing transition upon heating to a higher temperature. This transition typically involves a change in several spectroscopic properties of the sample arising from the difference in the spectral behavior of the complex and the single strands since purines and pyrimidines in DNA absorb UV-radiation around 260 nm, any denaturation of the duplex-DNA is accompanied by *hyperchromatic* change in the absorbance, followed by monitoring the change in absorbance at 260 nm (A_{260}) with temperature. A plot of A_{260} *vs.* temperature shows a sigmoidal curve resulting from the co-operative nature of the duplex ⇔ single-strand transition with mid-point of the curve (T_m) being an indicator of the strength of complex formation. Peptide as an external agent when bound to duplex-DNA modulates the stability of the double-helical structure resulting in an increase of T_m. Such changes in T_m can be used to estimate the strength of peptide-DNA interaction.

UV-T_m experiments were carried out with an AT rich (**D1:D2**) and a CG rich (**D3:D4**) 18-mer duplex-DNAs to determine their melting temperatures (Figure **35**). The AT rich (**D1:D2**) duplex (1 μM) denatured at 36 °C in phosphate buffer (10 mM, pH 7.2, 10 mM salt), while CG rich (**D3:D4**) duplex (1 μM) melted at 73.4 °C.

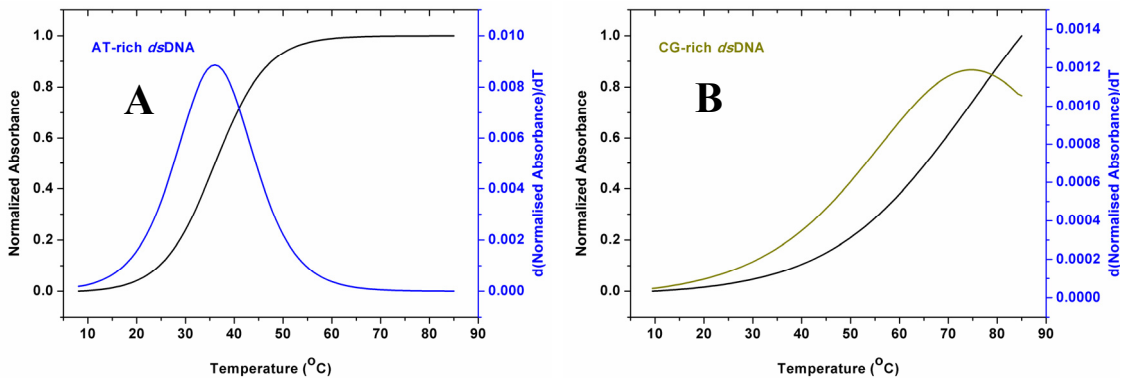


FIGURE 35: UV-T_m curves of (**A**) AT rich *ds*DNA **D1:D2**, (**B**) CG rich *ds*DNA **D3:D4**; Conditions: 1 μM DNA-duplex in phosphate buffer (10 mM with 10 mM salt) at pH 7.2. Experiments were performed by monitoring the absorbance at 260 nm with temperature. Heating rate = 0.5 °C min^{-1}. Black lines are the melting profiles and colored lines indicate the first derivative curves from which the T_m values are derived.

These peptides were mixed with SAP (**P5**) peptide at different ratios (1:1, 1:50 and 1:100) and UV-T_m studies were done with similar DNA and buffer conditions as mentioned above. As the salts are known to screen charges interaction of polycationic substances with DNA, was carried out in presence of very little salt (10 mM). The T_m-values obtained from the first derivative curves of the temperature-dependent absorbance data are shown in Table **7**.

Table 7: UV-T_m data for interaction of peptide **P5** (SAP) with *ds*DNA.

	T_m (°C)				
	only *ds*DNA	**1:1**	**1:50**	**1:100**	**only Peptide**
AT-rich Seqn	36.0	36.0	36.0	36.0	ND
CG-rich Seqn	73.4	>85	>85	>85	

Irrespective of the peptide concentration, the peptide **P5** had no effect on the strength of the AT rich *ds*DNA as can be inferred from the Figure **36(A-B)**. In case of CG rich *ds*DNA-SAP complexes, there seemed to be increase in melting temperatures (beyond 85 °C) with respect to only *ds*DNA (Figure **36(C-D)**). But these values could not be determined as they were out of scale.

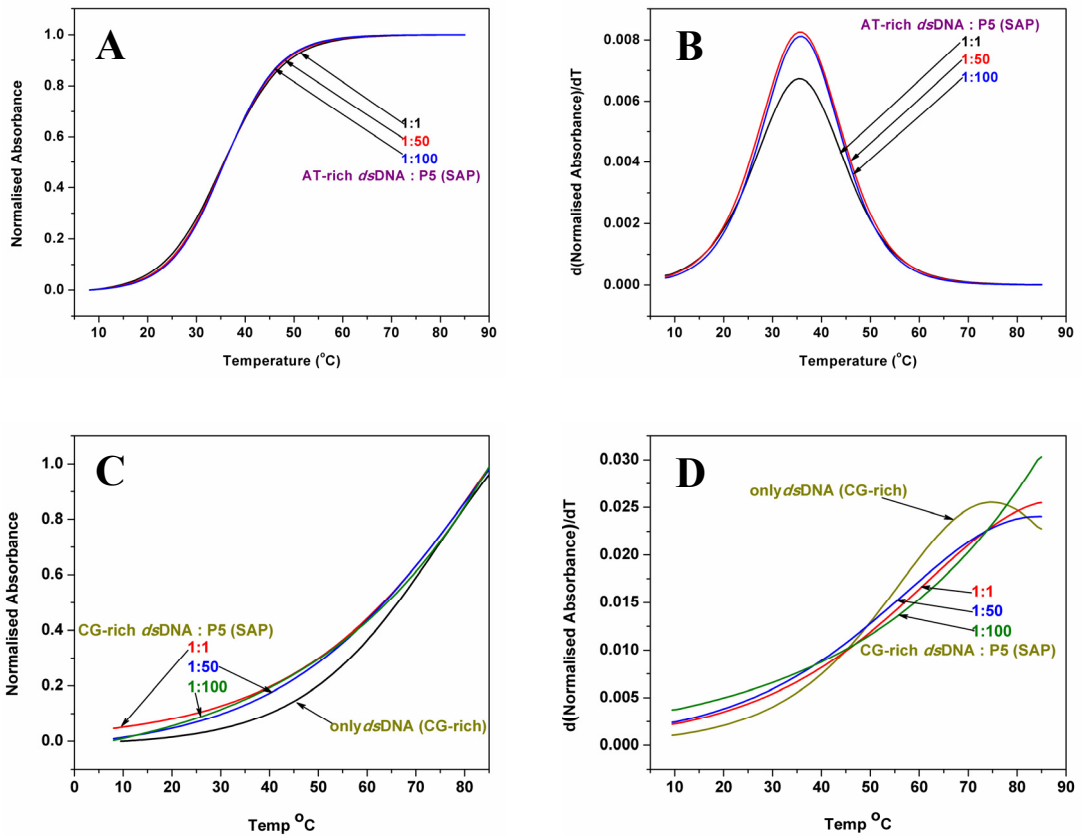


FIGURE 36: UV-melting profiles of (**A**) AT rich *ds*DNA **D1:D2** + **P5**, (**C**) CG rich *ds*DNA **D3:D4** + **P5**; first derivative UV-T_m of curves (**B**) AT rich *ds*DNA **D1:D2** + **P5**, (**D**) CG rich *ds*DNA **D3:D4** + **P5**; Conditions: 1 μM DNA-duplex and (1 μM, 50 μM, 100 μM) peptide in phosphate buffer (10 mM with 10 mM salt) at pH 7.2. Experiments were performed by monitoring the absorbance at 260 nm with temperature. Heating rate = 0.5 °C min^{-1}.

Similarly for other synthesized peptides (**P1**, **P2**, **P3** and **P4**) the UV-melting studies with *ds*DNA were carried out. Figures **37(A-H)** show the UV-melting profiles of the complex [*ds*DNA **D1:D2** (1 μM) + peptide (1 μM, 50 μM, 100 μM)] all at pH 7.2 (10 mM phosphate buffer with 10 mM NaCl) obtained by monitoring the A_{260} with temperature. Results of CG-rich *ds*DNA complexing with the collageneous peptides have been excluded as all of them had out of range UV-T_m profiles.

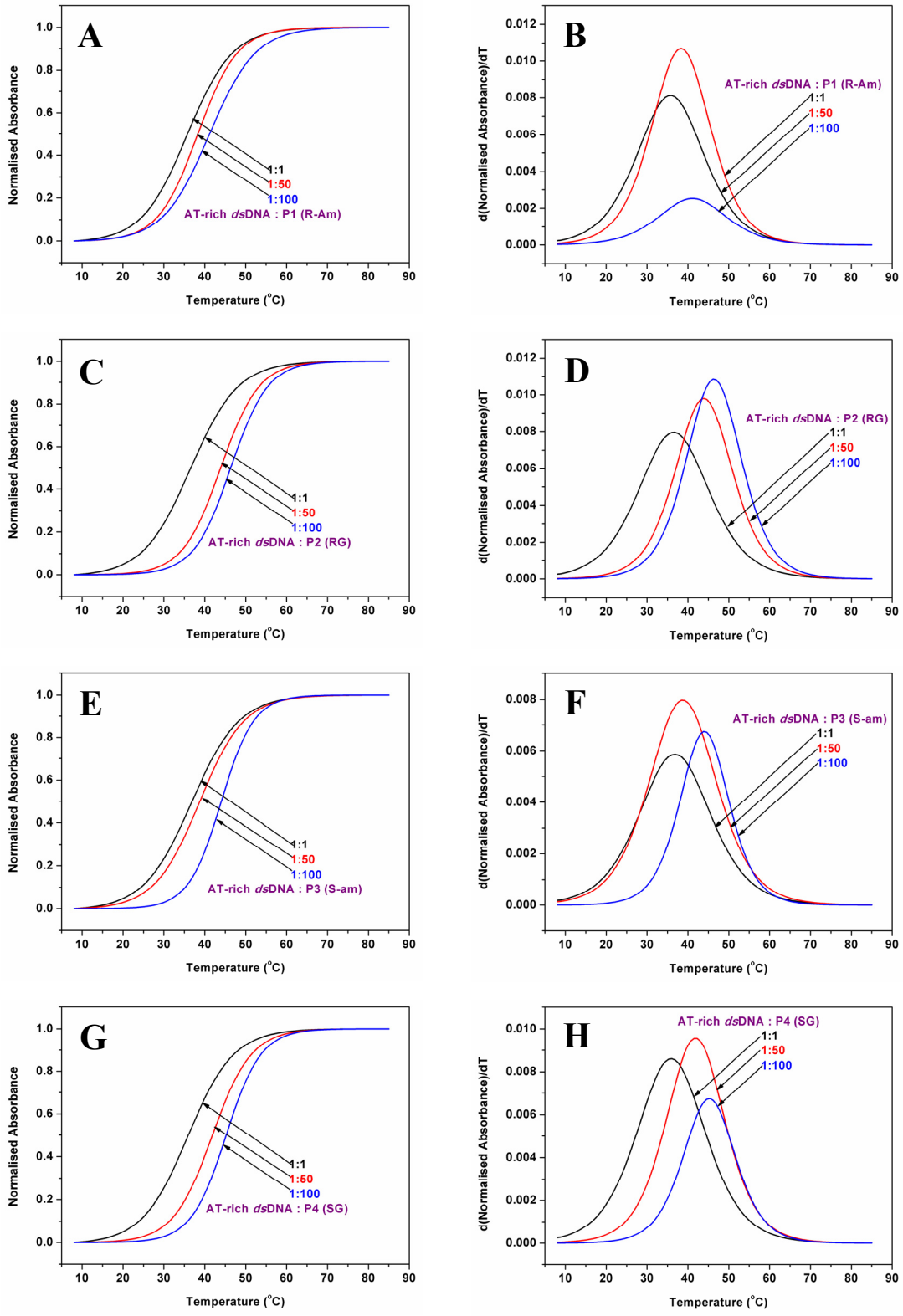


FIGURE 37: UV-melting profiles of (**A**) AT rich *ds*DNA **D1:D2** + **P1**, (**C**) CG rich *ds*DNA **D1:D2** + **P2**, (**E**) AT rich *ds*DNA **D1:D2** + **P3**, (**G**) CG rich *ds*DNA **D1:D2** + **P4**; Corresponding first derivative UV-T_m of curves in (**B**), (**D**), (**F**) and (**H**) respectively; Conditions: 1 μM DNA-duplex and (1 μM, 50 μM, 100 μM) peptide in phosphate buffer (10 mM with 10 mM salt) at pH 7.2. Experiments were performed by monitoring the absorbance at 260 nm with temperature. Heating rate = 0.5 °C min^{-1}.

Table **8** lists the T_m values obtained from the above plots for variable concentration peptide-*ds*DNA complexes. And the differences in melting temperatures of only *ds*DNA and peptide-*ds*DNA complexes are compiled in Table **9**.

Table 8: UV-T_m data for interaction of cationic peptides with *ds*DNA (AT-rich) sequence

Peptide	T_m (°C)				
	only *ds*DNA	1:1	1:50	1:100	only Peptide
(P1) R-Am	36.0	36.0	38.4	41.5	ND
(P2) RG		36.0	43.8	46.1	ND
(P3) S-am		36.8	39.1	43.8	ND
(P4) SG		36.0	42.2	45.4	ND

The complex of AT rich duplex-DNA **D1:D2** (1 μM) with all peptides of same concentrations at pH 7.2 in 10 mM phosphate shows T_m of ~36 °C, while there is increase in the duplex strength in presence of high concentrations (50 & 100 fold) peptides by at least 2.4 °C.

Table 9: Difference in UV-T_m values between cationic peptides -*ds*DNA (AT-rich) complex and only *ds*DNA.

Peptide	ΔT_m (°C)		
	1:1	1:50	1:100
(P1) R-Am	0	2.4	5.5
(P2) RG	0	7.8	10.1
(P3) S-am	0.8	3.1	7.8
(P4) SG	0	6.2	9.4

The T_m of the duplex is significantly altered by guanidino-peptides, **P2** (RG) with DNA to peptide ratio (1:100 fold) where it gains 10.1 °C, followed by **P4** with 9.4 °C of ΔT_m. Among the amino-collagen peptides **P1** (R-Am) with ΔT_m of 2.4 °C at (1:50) and 5.5 °C at (1:100) it is relatively less affected.

In order to distinguish the change in T_m values to be a result of peptide-DNA complexation and not from self melting of peptides, the thermal denaturation of peptides were monitored by A_{260} under identical conditions (Figure **38**). The presence of phenylalanine residue at the N-terminus of each of the above peptides enabled this procedure.

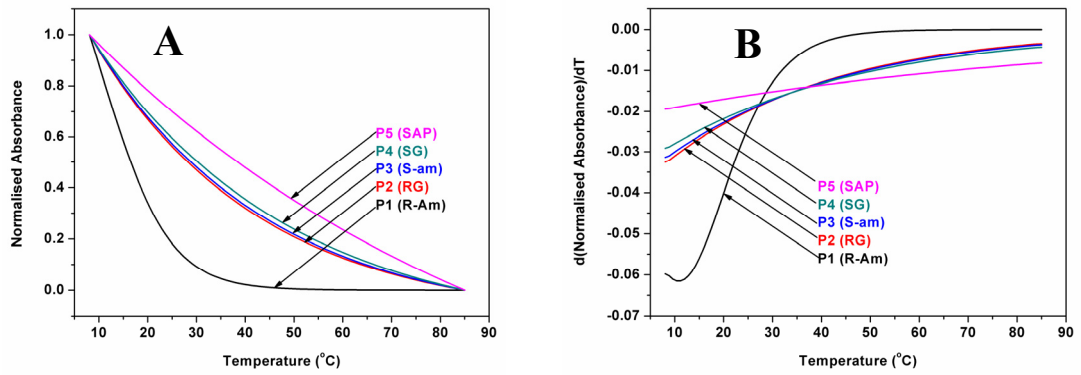


FIGURE 38: UV-melting profiles of peptides (**A**) **P1**, **P2**, **P3**, **P4** and **P5**, (**B**) Corresponding first derivative UV-T_m of curves; Conditions: 100 μM peptide in phosphate buffer (10 mM with 10 mM salt) at pH 7.2. Experiments were performed by monitoring the absorbance at 260 nm with temperature. Heating rate = 0.5 °C min^{-1}.

The changes in absorbance at 260 nm of peptides were random and did not present any sigmoidal transitions, which clearly indicated that the increased melting temperatures are due to complexation between *ds*DNA and cationic peptides.

1.9.3c CD study of *ds*DNA binding to cationic peptides

The chromophores responsible for the intrinsic CD of nucleic acids are the purine and pyrimidine bases.[111] The CD spectra indicates the conformational state of nucleic acids and this technique has been used extensively for following conformational transitions in DNA, such as denaturation and transitions from B→A and B→Z DNA.[112] The CD spectrum of native DNA generally shows a typical B-form conformation, composed of four major peaks in the UV-Vis region: a negative 210 nm peak, a positive 220 nm peak, a negative 245 nm peak and a positive ~280 nm peak.[113] Transition of the DNA conformation from the normal B-form to the less compact C-form, is characterized by a decrease in the intensity of the positive 280 nm peak commonly observed in condensed DNA systems.[114] Hence changes in CD spectra around 280 nm can be monitored for understanding peptide-DNA interactions.

The purified peptide samples were individually studied by CD spectroscopy under different conditions for ascertaining their conformation. To follow the conformational changes of peptide after complex formation with DNA, CD spectroscopy was performed on same composition of samples that showed increment in UV-T_m studies described earlier in this section.

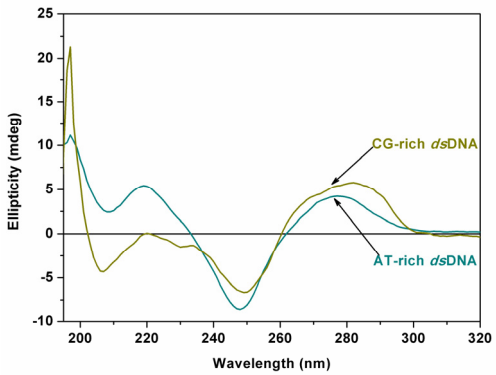


FIGURE 39: CD spectra of AT rich (**D1:D2**) and CG rich (**D3:D4**) duplex DNA (1 µM); 10 mM phosphate buffer (10 mM NaCl) pH 7.2

Initially the CD spectra of the two different type of duplex DNA were taken in phosphate buffer at neutral pH (Figure **39**). The spectrum obtained in each case was in accordance with literature reports.[115] The CG rich *ds*DNA has a sharp positive peak at 197 nm, a hump in the negative range just touching the zero line, a negative band at 249 nm and a broad positive band around 280 nm. In the CD spectrum of AT rich *ds*DNA similar pattern was observed with a few marginal changes. In this case also broad positive peaks at 277 nm, at 219 nm and at 197 nm (lower intensity) along with a negative band at 248 nm were observed.

Figure **40** illustrates the CD spectra of resulting complexes along with the individual CD of peptide and *ds*DNA for the sake of comparison. It is seen that the addition of peptide **P5** (SAP) changes the CD spectrum of DNA duplexes.

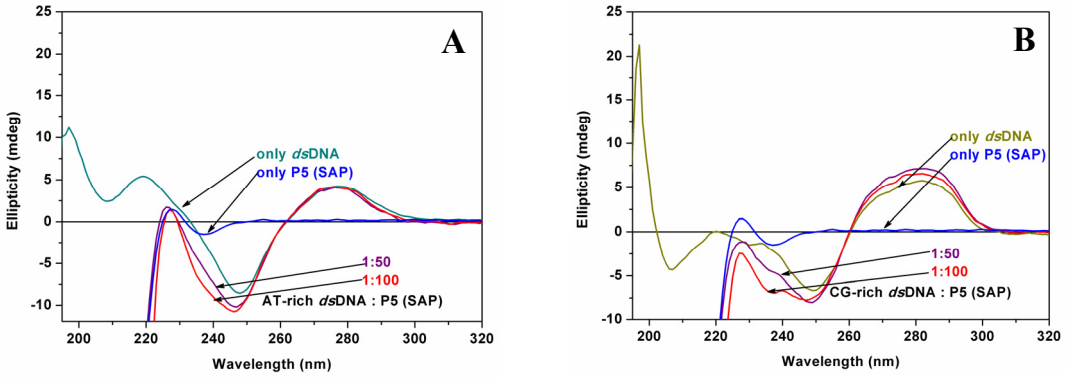


FIGURE 40: CD spectra of peptide **P5** (100 µM) along with (**A**) AT rich *ds*DNA **D1:D2** (1 µM) and *ds*DNA+**P5** complex, (**B**) CG rich *ds*DNA **D3:D4** (1 µM) and *ds*DNA+**P5** complex at different ratios (1:50, 1:100) in 10 mM phosphate buffer (10 mM NaCl) pH 7.2

The interaction of peptide **P5** (SAP) with AT rich *ds*DNA in high ratios (1:50, 1:100) does not induce any conformational changes to be observed in the CD spectrum of the *ds*DNA+**P5** mixture (Figure **40.A**). The spectra of DNA and peptide seemed to be additive in the mixture in case of AT rich *ds*DNA, while for CG rich *ds*DNA+**P5** the intensity of positive peak at 282 nm increased with higher ratio of *ds*DNA to peptide in the mixture (Figure **40.B**).

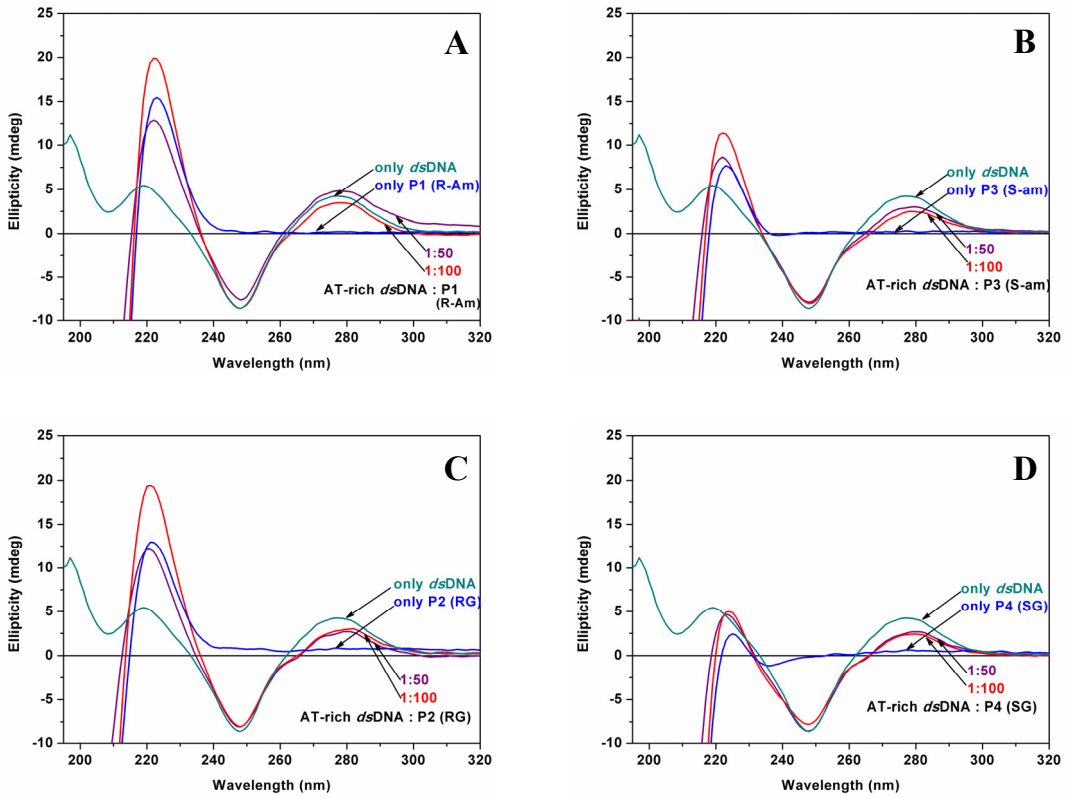


FIGURE 41: CD spectra of ATrich *ds*DNA (1 μM) along with (**A**) peptide **P1** (100 μM) and *p*DNA+**P1** complex; (**B**) peptide **P3** (100 μM) and *p*DNA+**P3** complex; (**C**) peptide **P2** (100 μM) and *p*DNA+**P2** complex; (**D**) peptide **P4** (100 μM) and *p*DNA+**P4** complex at different ratios (1:50, 1:100) in 10 mM phosphate buffer (10 mM NaCl) pH 7.2

Figure **41** shows CD spectra of the similar complexes of other synthesized peptides (**P1**, **P2**, **P3** and **P4**) with AT rich *ds*DNA, and in all the cases the change in intensity around 277 nm positive peak was observed. However, guanidino peptides **P2** (RG) and **P4** (SG) showed more prominent changes than the amino peptides **P1** and **P3**. Interestingly the results of CD spectroscopy are in agreement with that of UV-T_m for same set of *ds*DNA+peptide samples, where the guanidino peptides had induced maximum rise in the UV-T_m values.

1.9.3d CD study of plasmid DNA binding to cationic peptides

Figure **42-43** show the CD spectra of plasmid DNA (*p*DNA) taken with various cationic peptides **P1**-**P5** along with only *p*DNA and only peptide CD spectra separately. All the spectra are corrected for buffer contributions. The CD spectrum of *p*DNA (pRmHa3-GFP) taken in 10 µM EDTA buffer shows a moderate positive band at 271 nm and a weak negative band at 246 nm. A weak positive band of almost equal intensity at 220 nm followed by a strong positive peak at 196 nm is also observed. It is evident that the addition of peptides induces a significant conformational transition in the *p*DNA as seen from the changes in both intensity and position of the bands in the CD spectra.

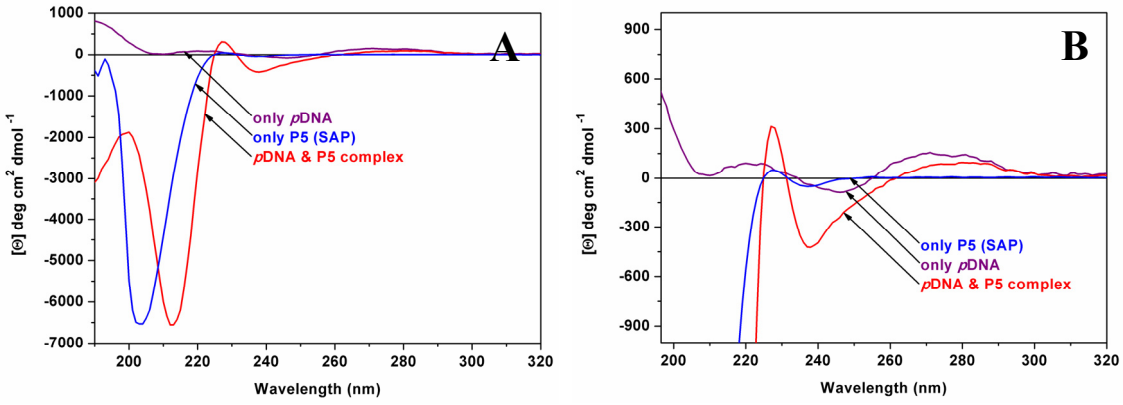


FIGURE 42: (**A**) CD spectra of *p*DNA (5 µg/ml), peptide **P5** (250 µg/ml) and *p*DNA-**P5** (1:50 w/w) complex in 10 µM EDTA using 10 mm path length cell, (**B**) zoomed view of the same CD spectra

Upon complexation of *p*DNA with peptide **P5** (SAP), the following changes are observed in its CD pattern (Figure **42**). The intensity of the positive band in the short wavelength region at 196 nm, changes to negative peak at 200 nm. A positive band that is seen in the *p*DNA at 220 nm amplifies six times accompanied by a red-shift to 227 nm. The weak negative band 246 nm becomes a major trough with blue shift to 238 nm. A drop in the intensity of the 271 nm band is also observed. Further, when compared with CD spectra of **P5**, the *p*DNA+SAP spectra seems to be additive apart from the deep negative peak which moves from 203 nm to 212 nm. The other peaks at 228 nm and 238 nm get amplified but retain their original positions as earlier.

The amino-peptides **P1** and **P3** had much more pronounced effect on the CD pattern of *p*DNA when mixed together at 1:50 ratio. Figure **43.A** shows the

complexation of **P1** (R-Am) with *p*DNA. In this case also the positive band in the short wavelength region at 196 nm, changes to negative band. The positive band at 220 nm intensifies more than 25 times with red-shift to 227 nm. A weak negative band present in 246 nm region becomes more shallow and connects to the 271 nm band which shifts phase from negative to positive. On comparison with CD spectrum of the peptide **P1** taken alone under similar conditions, the CD pattern seems to have totally changed. The only peaks to be correlated are 223 nm positive peak in peptide shifts to 227 nm with 10 fold amplification.

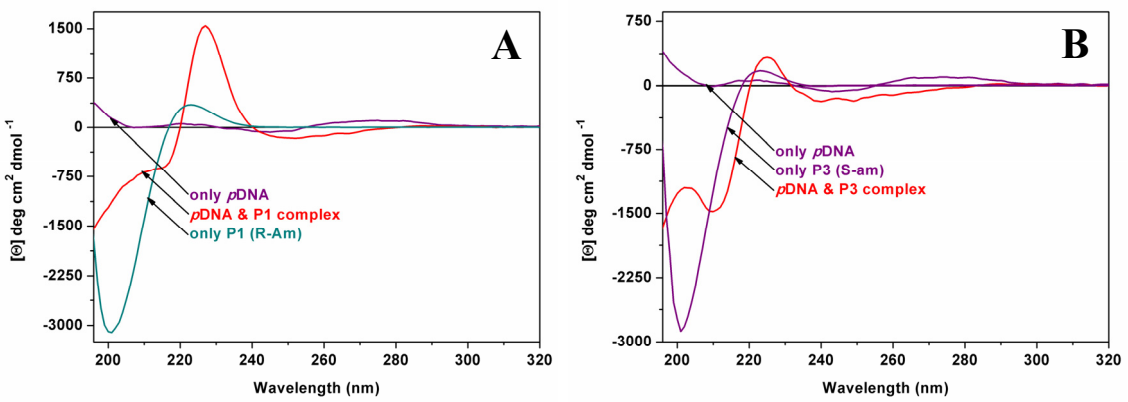


FIGURE 43: CD spectra of *p*DNA (5 µg/ml) along with (**A**) peptide **P1** (250 µg/ml) and *p*DNA+**P1** (1:50 w/w) complex, (**B**) peptide **P3** (250 µg/ml) and *p*DNA+**P3** (1:50 w/w) complex in 10 µM EDTA using 10 mm path length cell

Similarly for peptide **P3** (S-am) the complexation with *p*DNA shows changes in the same wavelength regions of the CD spectrum (Figure **43.B**). The positive band in the short wavelength region at 196 nm, now changes to a hump at 200 nm in negative ellipticity. A positive band that is seen in the *p*DNA at 220 nm and 223 nm for only **P3** undergoes red-shift to 225 nm with marginal amplification. The positive band at 271 nm vanishes and the negative band of 246 nm amplifies along with broadening.

The CD spectra of guanidino peptides **P2** and **P4** have more severe changes on being combined with *p*DNA. Peptide **P2** (RG) on complexation with *p*DNA results the following CD pattern (Figure **44.A**). A positive band at 196 nm is slightly reduced and less broad than in only *p*DNA. Another intensified positive band is seen at 223 nm from the merging of 220 nm (only *p*DNA) and 221 nm (only **P2**) peaks. The 271 nm positive band disappears, while the 246 nm negative band stretches out. The conformational signature band for peptide **P2** at 200 nm retains itself though with reduced intensity and red-shift of peak position to 200 nm.

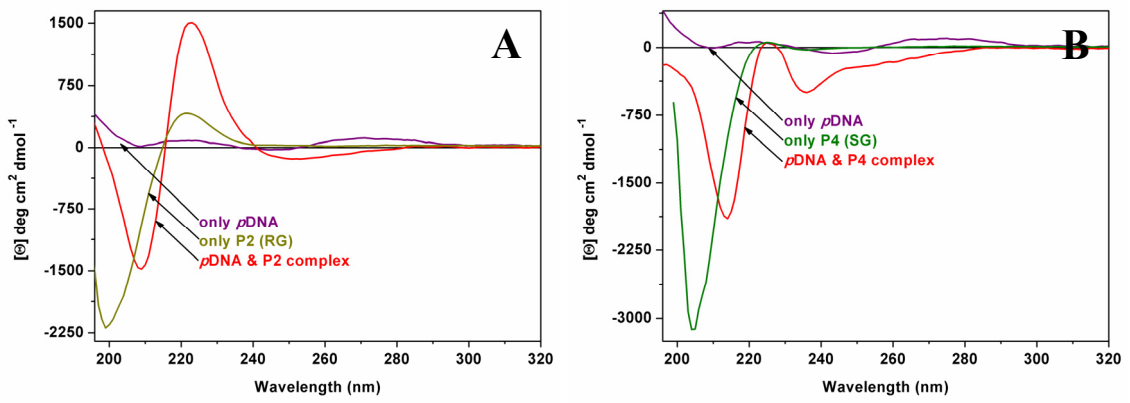


FIGURE 44: CD spectra of *p*DNA (5 µg/ml) along with (**A**) peptide **P2** (250 µg/ml) and *p*DNA+**P2** (1:50 w/w) complex, (**B**) peptide **P4** (250 µg/ml) and *p*DNA+**P4** (1:50 w/w) complex in 10 µM EDTA using 10 mm path length cell

The *p*DNA on complexation with peptide **P3** (SG), gives a new CD pattern different from the *p*DNA or **P3** taken individually (Figure **44.B**). The intensity of the positive band at 190 nm reduces to negative. A positive band that is found in the *p*DNA at 220 nm makes a red-shift to match the intensity of **P3** at 225 nm. The 246 nm negative band intensifies along with blue-shift to a large negative band at 236 nm and no positive band at 271 nm is observed anymore. The deep negative peak of **P3** at 204 nm shifts to 214 nm with loss of intensity.

1.9.3e Fluorescence study - EtBr displacement assay with *ds*DNA

To examine whether the differential stabilization observed in DNA-duplex by cationic-peptides is due to their different binding affinities towards DNA, the displacement of ethidium bromide upon binding of these peptides to DNA was followed by fluorescence assay. Experiment was performed by incremental addition of *ds*DNA stock solution (1 µM) to ethidium bromide (250 µM) in water (10 mM NaCl solution) and recording the fluorescence emission intensity at 610 nm till the saturation point is attained. The fluorescence intensity steadily increases and reaches a saturation value. The resulting ethidium-DNA complex was individually titrated with aqueous solutions of peptides **P1**, **P2**, **P3**, **P4** and **P5**, and the fluorescence intensity changes were recorded.

Figure **45** shows the fluorescence intensity of AT rich *ds*DNA (**D1:D2**) as a percentage of fraction of initial value (I/I_0) plotted against the concentration of the peptide separately for amino peptides (**P1** & **P3**) and guanidino peptides (**P2** & **P4**) while comparing them with control peptide **P5** (SAP).

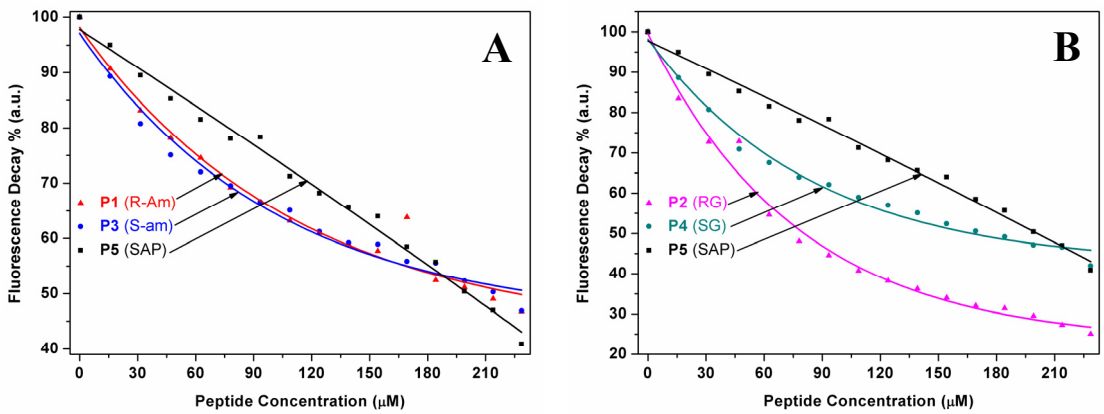


FIGURE 45: Ethidium bromide displacement assay with peptides (**A**) **P1**, **P3** and **P5**; (**B**) **P2**, **P4** and **P5** with AT rich *ds*DNA; 1 μM *p*DNA in 800 μl of 10 μM NaCl saturated with 0.59 μg EtBr, $\lambda_{excitation}$=490 nm and $\lambda_{emission}$= 610 nm

Up to a concentration of 228 μM, the amino-peptides **P1** and **P3**, the fluorescence intensity of the ethidium:*ds*DNA decreases to 47% gradually in an exponential pattern. The guanidine peptides on the other hand have more rapid intensity changes, with **P2** quenching the fluorescence to near about 25% followed by **P4**, which quenches till 42%. Interestingly for peptide **P5** (SAP) the fluorescence decay is more linear with overall reduction to 41% of the original intensity.

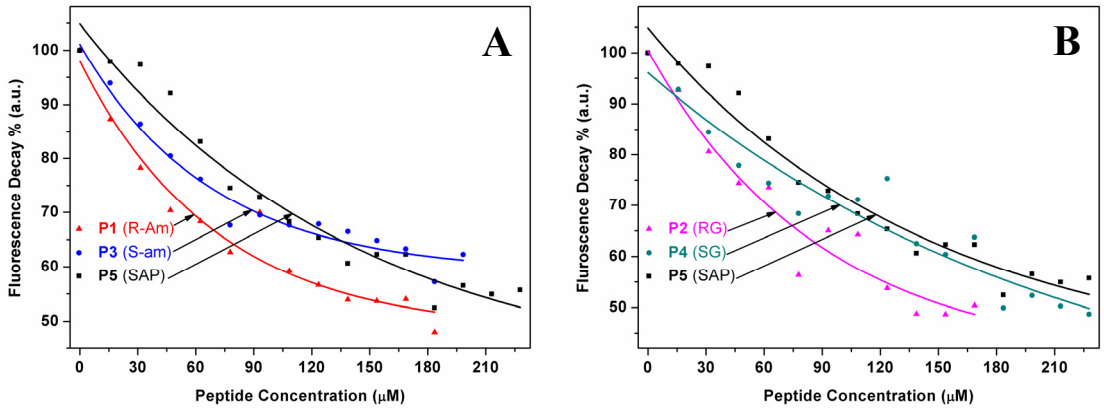


FIGURE 46: Ethidium bromide displacement assay with peptides (**A**) **P1**, **P3** and **P5**; (**B**) **P2**, **P4** and **P5** with CG rich *ds*DNA; 1 μM *p*DNA in 800 μl of 10 μM NaCl saturated with 0.88 μg EtBr, $\lambda_{excitation}$= 490 nm and $\lambda_{emission}$= 610 nm

In a similar way the fluorescence intensity of CG rich *ds*DNA (**D3:D4**) has been presented in Figure **46**. The binding of **P5** peptide (SAP) to CG rich *ds*DNA reduces 50% of the fluorescence by displacing the ethidium ion. In this case the rate of change of fluorescence intensity is not linear as observed with AT rich DNA

duplex. All the other cationic peptides (**P1**, **P2**, **P3** and **P4**) have about similar fluorescence quenching ability with CG rich DNA duplex *i.e* around 50% overall.

Table **10** summarizes the different percentage of fluorescence decay in both AT rich and CG rich DNA duplexes as a result of condensation with cationic peptides at ~154 μM concentration beyond which no significant change is observed.

Table 10: UV-T_m data for interaction of peptide **P5** (SAP) with *ds*DNA.

	Fluorescence decay (%)				
***ds*DNA**	**P1 (R-Am)**	**P2 (RG)**	**P3 (S-am)**	**P4 (SG)**	**P5 (SAP)**
AT-rich	58	34	59	52	64
CG-rich	54	49	65	60	62

1.9.3f Fluorescence study - EtBr exclusion assay with *ds*DNA

Another most common way of analyzing the formation of cationic peptide-DNA complex is the EtBr exclusion assay[116], where the *ds*DNA is pre-incubated with peptide and then the complex is challenged with competition for binding with EtBr titration. In all the experiments AT rich and CG rich *ds*DNA (1 μM) in presence of peptide (100 μM), at pH 7.0 (10 mM NaCl) were incubated for 6 hrs (4 ^{o}C) individually and titrated with aliquots of EtBr solution (100 μM). The fluorescence spectra were of these complexes were recorded as a function of EtBr concentration by $\lambda_{excitation}$ at 490 nm and $\lambda_{emission}$ 601 nm respectively. The results are expressed as percentage of fluorescence with EtBr intercalation taken as 100%.

Figure **47** shows the fluorescence data for AT rich *ds*DNA–EtBr intercalation in the presence of various amino and guanidino peptides. In this case, for all peptide-*ds*DNA mixtures, similar features are exhibited with an initial increase of fluorescence as a result of increasing EtBr intercalation, but reaches saturation fluorescence at ~10 μM of EtBr. Further it is observed that **P5** (SAP) offers 80% of EtBr intercalation, **P1** (R-Am) intercalates only 45% EtBr in the *ds*DNA-peptide complexation, and the other three peptides **P2**, **P3** and **P4** allow about 60-70%.

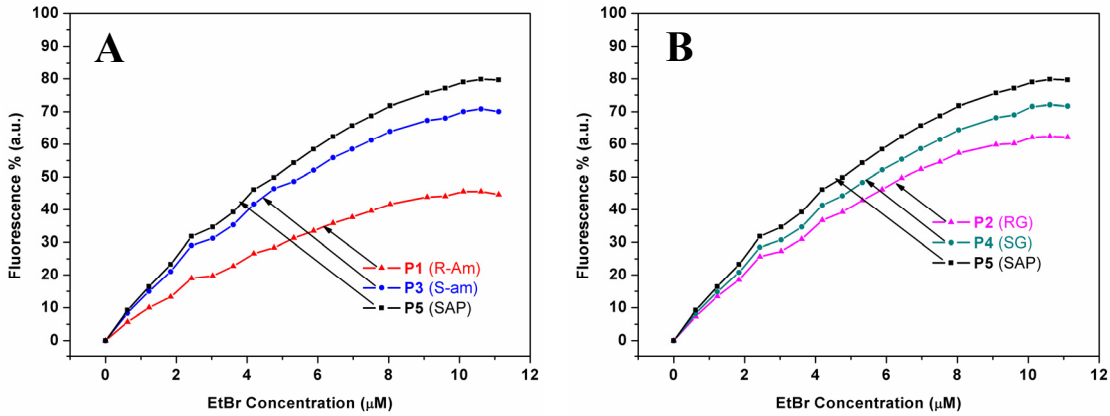


FIGURE 47: Ethidium bromide exclusion assay with peptides (**A**) **P1**, **P3** and **P5**; (**B**) **P2**, **P4** and **P5** with AT rich *ds*DNA; 1 μM *p*DNA in 800 μl of 10 μM NaCl with 100 μM (100X) peptides, $\lambda_{excitation}$ = 490 nm and $\lambda_{emission}$ = 601 nm

When the same set of experiments was repeated with CG rich *ds*DNA, a slightly different pattern of results was noticed. The fluorescence increases initially with increase of EtBr concentration but decreases beyond a particular concentration (4.6 μM) of EtBr as can be seen in Figure **48**.

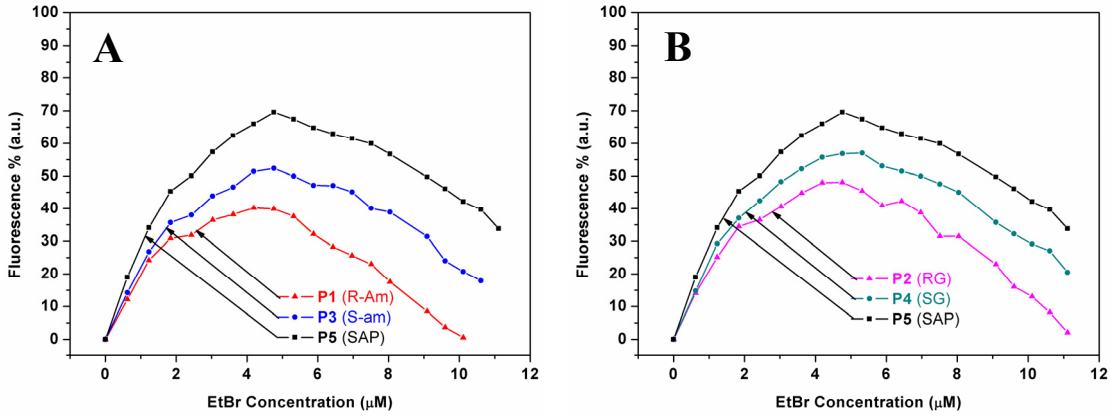


FIGURE 48: Ethidium bromide exclusion assay with peptides (**A**) **P1**, **P3** and **P5**; (**B**) **P2**, **P4** and **P5** with CG rich *ds*DNA; 1 μM *p*DNA in 800 μl of 10 μM NaCl with 100 μM (100X) peptides, $\lambda_{excitation}$= 490 nm and $\lambda_{emission}$= 600 nm

1.9.3g Fluorescence study - EtBr displacement & exclusion assay with plasmid DNA

The binding affinities of cationic peptides towards *p*DNA (pRma3H-GFP) were also checked through fluorescence studies (displacement and exclusion) with EtBr as a probe.

The displacement assay was carried out by mixing the minimum quantity of EtBr (35 µl, 250 µM) required to achive a fluorescence saturation with 3 µg of pRmHa3-GFP in 800 µl with buffer (10 µM EDTA). And this mixture was then titrated with aliquots of each peptide solution. The resulting fluorescence is expressed as a function of peptide to *p*DNA (w/w) ratio in Figure **49.A**. However not much difference was observed various *p*DNA-peptide complexes, except that there is sharp increase in fluorescence intensity initially which later reduces to almost same value with increasing peptide amount.

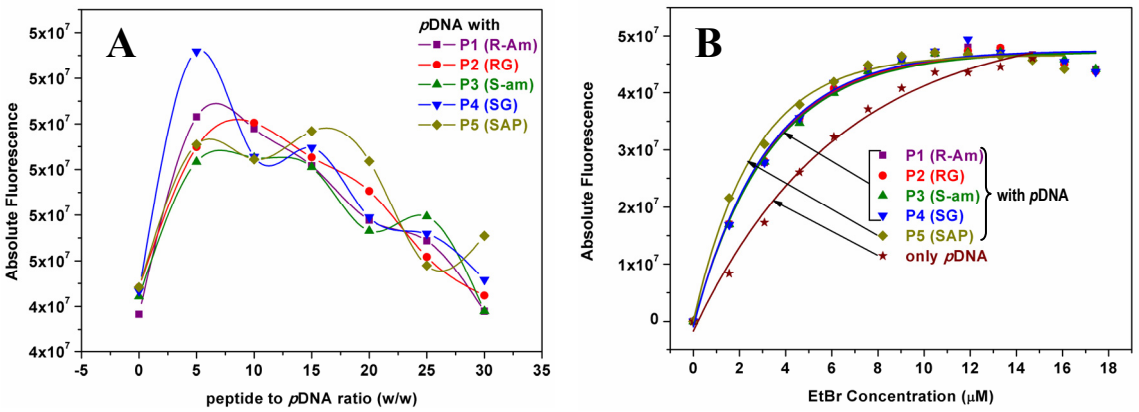


FIGURE 49: (**A**) Ethidium bromide displacement assay [3 µg *p*DNA in 800 µl of 10 µM EDTA saturated with 3.45 µg EtBr], (**B**) Ethidium bromide exclusion assay of *p*DNA with peptides **P1**-**P5** [3 µg *p*DNA in 800 µl of 10 µM EDTA with 150 µg (50X) peptides]; $\lambda_{excitation}$= 490 nm and $\lambda_{emission}$= 632 nm.

For exclusion assay, 3 µg of pRmHa3-GFP and 150 µg of various peptides were diluted to 800 µl with buffer (10 µM EDTA) individually and incubated for 30 min at 0 °C temperature. Each solution was then titrated with aliquots of EtBr (5 µl, 250 µM) and the fluorescence measured. Figure **49.B** shows the results obtained from this assay, where more fluorescence is observed for cationic peptide-*p*DNA complexes than only *p*DNA at any particular concentration of EtBr. Although beyond 7.5 µM of EtBr, the fluorescence is seen to saturate in case of peptide-*p*DNA complexes unlike free *p*DNA. Among all the cationic peptides, peptide **P5** (SAP) complex was slightly higher fluorescence intensity.

1.9.3h Native PAGE

Both covalent binding and electrostatic association of larger molecules to DNA are known to induce changes in gel migration patterns of the resulting structurally distorted DNA, which are expected to move slowly on a gel.[71]

The binding experiments were carried out by adding various peptides to the DNA duplex (AT or CG rich) followed by annealing and the complexation was monitored by non-denaturing gel electrophoresis at 10 °C and the results are depicted in Figure **50** and **51**.

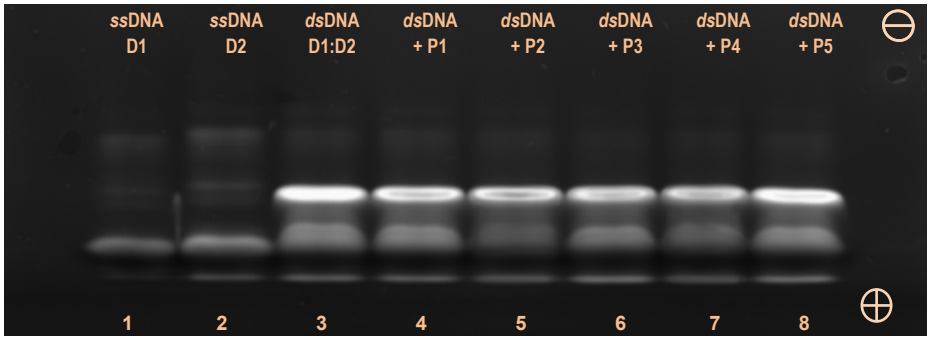


FIGURE 50: Polyacrylamide gel electrophoresis of AT rich *ds*DNA+peptide complexes (20% acrylamide, 1X-TBE buffer, 2 hrs, 30 A); Lane 1: DNA **D1** single strand, Lane 2: DNA **D2** single strand, Lane 3: *ds*DNA **D1:D2** duplex, Lane 4: *ds*DNA + **P1**, Lane 5: *ds*DNA + **P2**, Lane 6: *ds*DNA + **P3**, Lane 7: *ds*DNA + **P4**, Lane 8: *ds*DNA + **P5**.

The amino and guanidino peptide chimera **P1-P5** were found to cause only minor changes in the native PAGE. In Figure **50** the AT rich DNA duplex in lane 3 was more brightly luminescent than its peptide complexed counterparts. Peptide **P2** and **P4** in lane 5 and 7 respectively made the DNA bands look more dull in intensity. These bands also showed more prominent dark patches at the center of these *ds*DNA bands.

Similarly for cationic peptides binding to the CG rich duplexes were tested on native PAGE (Figure **51**). In this case also very faint change in the intensity of *ds*DNA was observed. Almost all the peptide complexed *ds*DNA seem to be less intense (though marginally) and have smiley kind of band pattern.

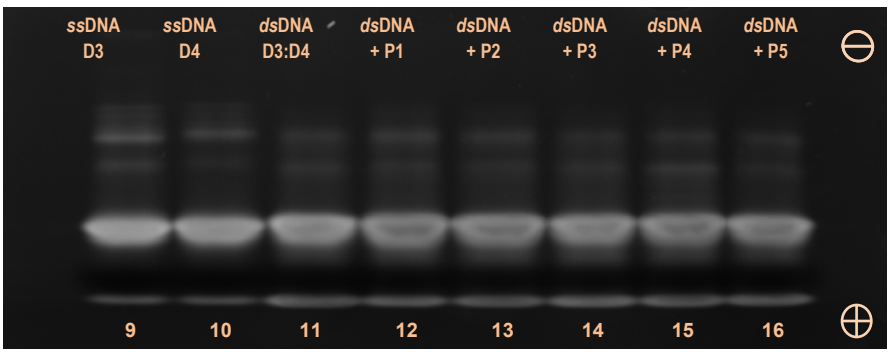


FIGURE 51: Polyacrylamide gel electrophoresis of CG rich *ds*DNA+peptide complexes (20% acrylamide, 1X-TBE buffer, 2 hrs, 30 A); Lane 1: DNA **D3** single strand, Lane 2: DNA **D4** single strand, Lane 3: *ds*DNA **D3:D4** duplex, Lane 4: *ds*DNA + **P1**, Lane 5: *ds*DNA + **P2**, Lane 6: *ds*DNA + **P3**, Lane 7: *ds*DNA + **P4**, Lane 8: *ds*DNA + **P5**.

Another important aspect to be noted from this experiment is that the peptides do not cause any dissociation of duplex into individual strands.

1.9.3i Agarose gel retardation assay

In the preliminary experiments, the amount of peptide required to prevent penetration of plasmid DNA into the agarose mesh and migration of the complex toward the cathode during gel electrophoresis was determined. The results of the gel-shift experiments with various peptides at constant w/w ratio (1:50) indicate that, for complexes prepared by peptide-DNA incubation period of 3 min to 15 min time are found to be enough for most peptides. This time dependent complexation of peptides to *p*DNA is shown in Figure **52** where lane 2 corresponds to free *p*DNA. In lanes 3–5 (*p*DNA with Peptide **P5**) the intensity of the bands indicates that almost half *p*DNA is free, still lane 3 (3 min incubation) shows a slight higher intensity at the wells signifying better complexation. Lanes 6-11 (with Peptide **P2** and **P4**) *p*DNA is totally retained in the wells and is not affected by period of incubation, whereas *p*DNA in lane 12-14 does not show any complexation with peptide **P1** even after 30 min incubation. Peptide **P3** though is partially able to retain the *p*DNA in well but varies with incubation time.

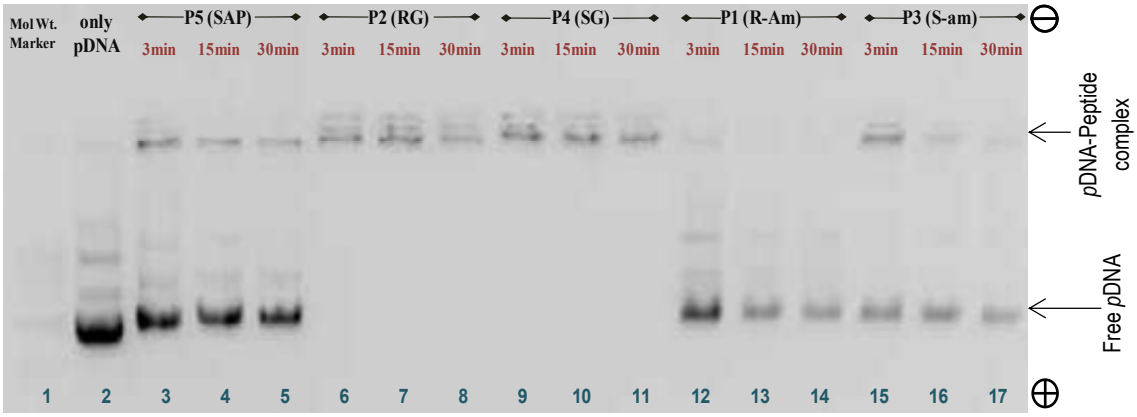


FIGURE 52: Time dependent binding of Peptides to *p*DNA (pRmHa3-GFP) on 1% Agarose Gel; Negative Image, 1X-TAE Buffer, 2hrs, 90Volts, *p*DNA to Peptide ratio (1:50 w/w) in EDTA buffer (10 µM); Lane 2: only *p*DNA, Lane 3-5: *p*DNA + **P5**, Lane 6-8: *p*DNA + **P2**, Lane 9-11: *p*DNA + **P4**, Lane 12-14: *p*DNA + **P1**, Lane 15-17: *p*DNA + **P3**.

Hence in subsequent experiments an average incubation time of 5 min was maintained for all the peptides complexing with *p*DNA.

Another binding assay was conducted to estimate concentation dependent complexation particularly with peptides **P2** and **P4** which were seen to retard the *p*DNA completely in the well. Therefore individual peptides and *p*DNA were

separately combined at different w/w ratios, and the resulting complexes subjected to electrophoresis on agarose gel with same conditions as previous (Figure **53**). The gel results illustrate the capacity of peptide to condense *p*DNA depending on their binding ratios.

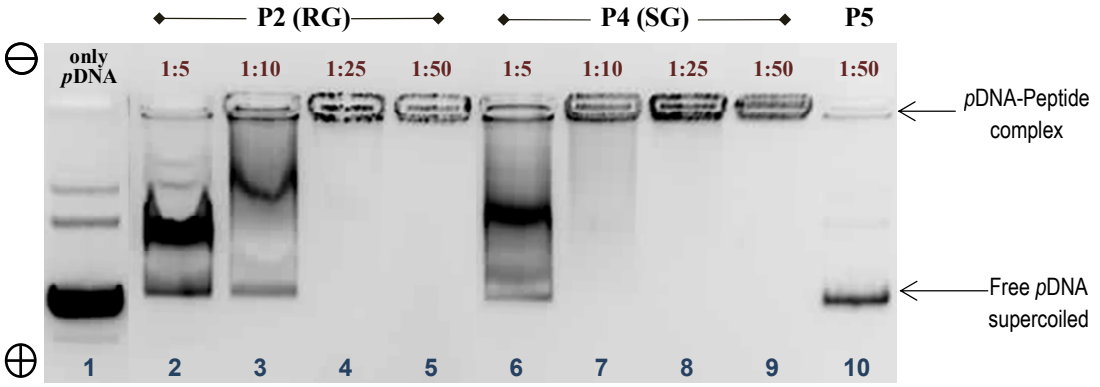


FIGURE 53: Binding efficiency of peptides to *p*DNA at different ratios (w/w in 10 µM EDTA buffer), as studied by agarose gel (1%) electrophoresis (Negative Image, 1X-TAE Buffer, 2hrs, 90Volts); Lane 1: only *p*DNA, Lane 2-5: *p*DNA + **P2** (1:5, 1:10, 1:25, 1:50 w/w), Lane 6-9: *p*DNA + **P4** (1:5, 1:10, 1:25, 1:50 w/w), Lane 10: *p*DNA + **P5** (1:50 w/w).

Lane 1 represents *p*DNA which separates into all its components and the free supercoiled DNA moves fastest in the electric field. In lanes 2–5 peptide **P2** and lanes 6–9 peptide **P4** is used along with *p*DNA which show retardation as a function of their binding (w/w) ratios. Lane 10 shows the control peptide **P5** (SAP) taken for comparision. It can be seen from Figure **53** that at 1:25 and 1:50 ratios most *p*DNA is retained in the well, where the complexes had been placed for both peptide **P2** and **P4**. At 1:10 ratio for peptide **P2** almost half the DNA is retained in the well and half diffuses out to form a smudged band, while peptide **P4** at same retains more DNA and only a slight amount diffuses out. A similar pattern is also observed both peptides at 1:5 ratio, more percentage of DNA diffuses out. Interestingly in cases of peptide complexed *p*DNA, the DNA that diffuses or leaks out from the wells runs less distance than free the *p*DNA.

Thus the minimum binding ratio of *p*DNA to peptides was determined to be 1:25 for both peptide **P2** and peptide **P4** as at this ratio almost all DNA is retarded in the well. This 1:25 ratio has been further maintained in all other biological studies involving the **P2** and **P4** peptides, while for other peptides 1:50 ratio was adhered, since 1:50 ratio is reported in literature as best binding ratio for control peptide **P5** (SAP).

1.9.3j Restriction enzyme digestion of *p*DNA-peptide complexes

One of the prerequisites for an efficient gene delivery system is that the DNA is still intact upon reaching the target in the cell, and has not been degraded by nucleases. Complexation with polycations has shown to protect DNA against degradation.[117] Therefore, a vital step in the development of any new delivery system based on polycations is demonstration of stability of the DNA against nuclease attack. The most common method used in such situations, is incubation of complexes in the presence of nucleases or endonucleases followed by determination of integrity of DNA by agarose gel electrophoresis.

In this case a restriction enzyme digestion assay has been used to verify the stability of peptide-*p*DNA complexes. A restriction enzyme (endo-nuclease) is one which recognizes and cuts DNA only at a particular sequence of nucleotides.[118] It cuts the double-stranded DNA by making two breaks, one through each of the phosphate backbones of the double helix, without damaging the bases of the DNA. BamH1 is one such resctriction enzyme derived from *Bacillus amyloliquefaciens* that recognizes and cuts the DNA following way (Figure **54**). It has the recognition site (G'GATCC), and leaves a sticky end.[119]

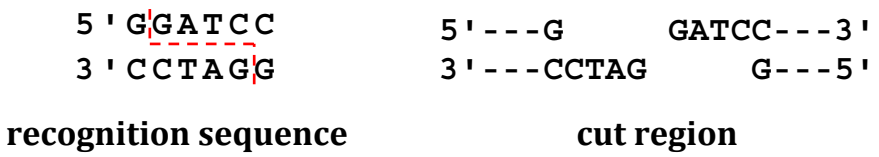


FIGURE 54: DNA restriction site recognised by enzyme BamH1

Plasmid DNA (pRmHa3-GFP) and the peptides **P2** (RG) and **P4** (SG) individuallly were incubated with restriction enzyme (BamH1) and products analyzed over agarose gel in comparison with reaction of free *p*DNA (Figure **55**).

The results of the agarose gel indicate that *p*DNA was fragmented into two bands by the restriction enzyme BamH1 (lane 2 of Figure **55**). The untreated *p*DNA shows three bands, the major band corresponding to the free supercoiled plasmid DNA in lane 1.

The enzyme digested *p*DNA has two bands corresponding to linear DNA which differ in length and the smaller one moving faster in the gel and the longer one being comparatively retarded. The same *p*DNA when complexed with guanidino

peptides show a difference in their digestion pattern. The *p*DNA-**P4** mixture in lane 6 shows no linear DNA, and most of the DNA is retartded in the well where it had been placed and seems to be not digested by the restriction enzyme. On the otherhand *p*DNA-**P2** mixture in lane 4 has reduced amount of linear DNA in comparison to simple *p*DNA's response to restriction enzyme as per the intensity of bands. In this case also a diffusing patch of DNA close to the well is seen.

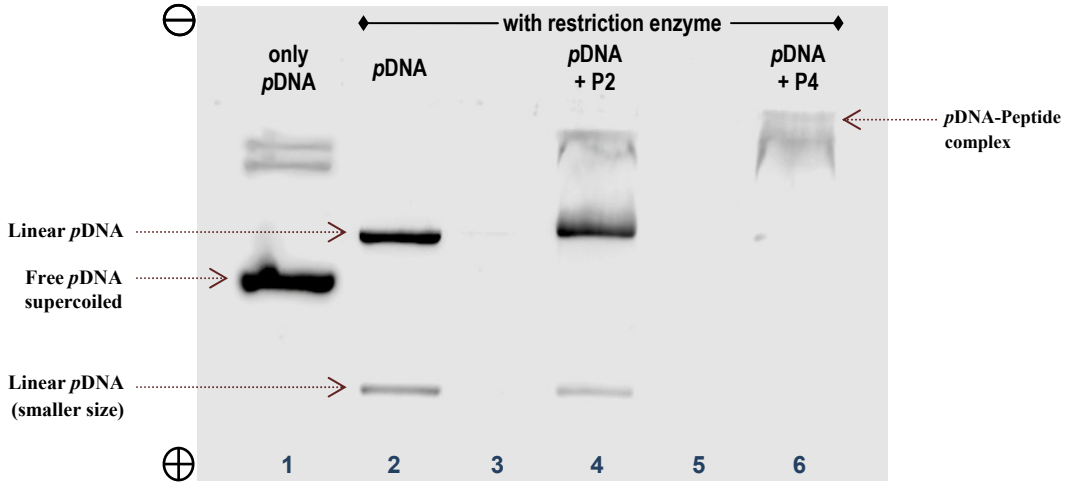


FIGURE 55: Agarose gel electrophoresis after restriction enzyme (BamH1) digestion of pDNA (pRmHa3-GFP) and its complexes with peptides **P2** (RG) and **P4** (SG) at 1:25 w/w ratio; (1% agarose gel, negative image, 1X-TAE buffer, 45 mins, 90 Volts); Lane 1: only *p*DNA untreated by enzyme, Lane 2: *p*DNA, Lane 4: *p*DNA with **P2**, Lane 6: *p*DNA with **P4** treated by restriction enzyme. Lane 3 and 5 are blank.

1.9.3k Cell Studies

To gain insight into the cell-uptake properties of these cataionic peptides, the intracellular distribution of the five peptides (**P1**, **P2**, **P3**, **P4** and **P5**) in *Drosophilia melanogaster*; Schneider 2 (S2) cell line was investigated.

The S2 cell line was derived from a primary culture of late stage (20-24 hours old) *Drosophila melanogaster* embryos by Schneider in 1972.[120] These cells grow at room temperature without CO_2 as a loose, semi-adherent monolayer in tissue culture flasks and in suspension in spinners and shake flasks. S2 cells have been increasingly utilized over the past few years for the expression of heterologous proteins[121] and particularly useful as it is easy to grow and maintain them in the lab. Moreover they are well suited to high-resolution light microscopic assays.[122]

(i) Cell transfection with cationic peptides

In the first step, the transfection efficiency of plasmid DNA (pRmHa3-GFP) encoding green fluorescent protein was determined using the Qiagen transfection reagents kit in *Drosophilia* (S2) cells. Later the transfection efficiencies were determined in the presence of each synthesised peptides (**P1**, **P2**, **P3**, **P4** and **P5**). The Qiagen transfection kit contains two major reagents, '*effectene*' and '*enhancer*', apart from the '*EC buffer*'. The replacement of either/both of these reagents with individual synthesised peptide provided vital information on the possible role of peptides.

Figure **56** shows a representative picture of live S2 cell lines before and after transfection. Figure **56.A** represents the DIC (Differential Interference Contrast) image before transfection while the subsequent images (Figure **56(B-D)**) show various degree of GFP expression after successful transfection with pRmHa3-GFP.

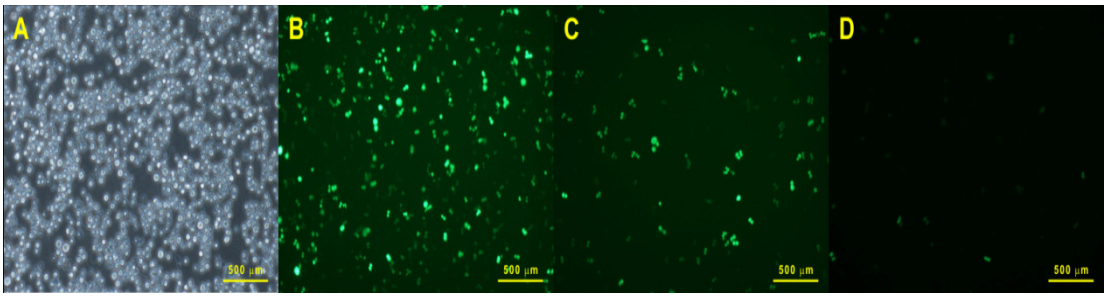


FIGURE 56: Live S2 cell line & various transfection levels (**A**) image of cells before transfection (**B**) high green fluorescence in cells after transfection (**C**) moderate fluorescence (**D**) weak fluorescence

Around 500 to 700 cells were spotted in each field of the cluster-plate well. Counting the number of green fluorescent cells among these gave an approximate value of transfection percentage. All the values from separate fields are averaged and plotted into bar graphs (Figure **57**). The outcome showed that the control cells without cationic peptide showed 17% transfection while the percentage of transfections in the case of cells having cationic peptide **P4**, followed by peptide **P2** were 42% and 35% respectively.

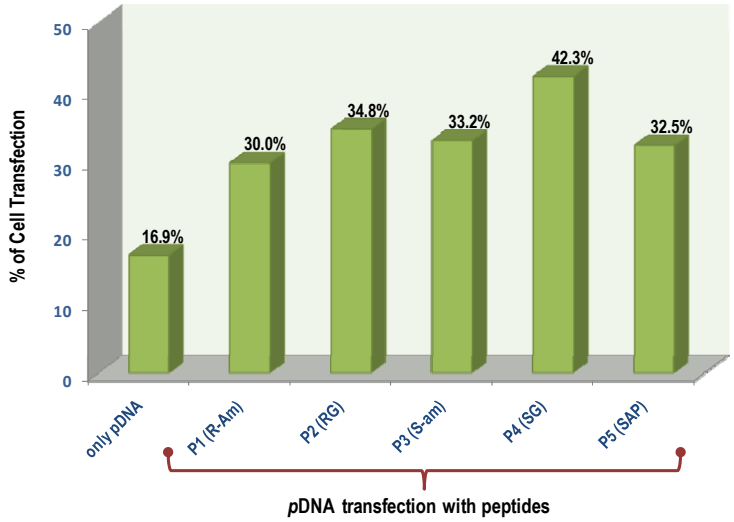


FIGURE 57: Percentage cell transfection of pRmHa3-GFP on S2 cells along with different peptides using Qiagen tansfection kit reagents

The transfection experiments with *p*DNA (pRmHa3-GFP) when carried without commercial reagents show no GFP expression, while replacements of each of the two reagents (Effectene & Enhancer) one by one with individual peptides results are interesting. The peptides **P2** and **P4** are able to replace the '*enhancer*' reagent of Qiagen transfection kit effectively.

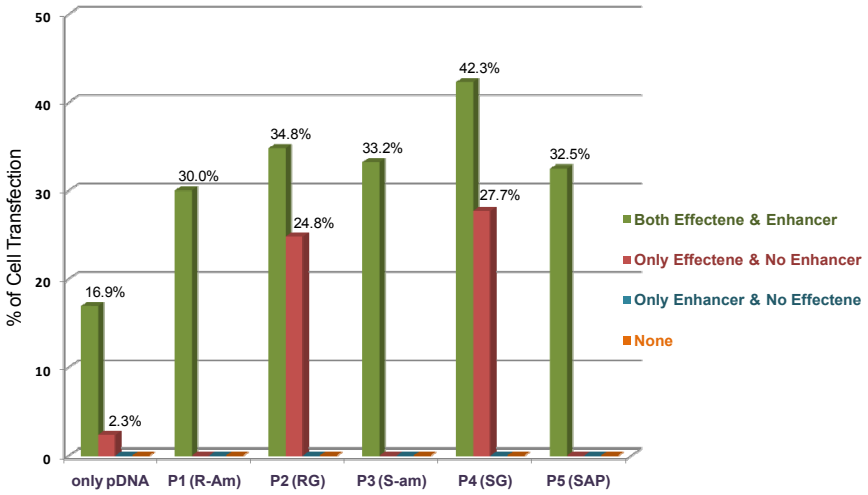


FIGURE 58: Comparative transfection profile (with respect to only *p*DNA) for transfecting reagents in presence of various peptides

The bar graph in above Figure **58** shows the result from various combinations of transfection reagents along with the synthesized peptides in transfecting pRmHa3-GFP into S2 cells. The first series represented by green bars are for experiments where both *effectene* and *enhancer* (Qiagen transfection kit) reagents have been used. This series also shows the maximum percentage of transfection. The next series of red bars correspond to experiments done with only *effectene* and no *enhancer*. It is to be noted here that while only *p*DNA transfection in this series is

2.3%, the pepetides **P2** and **P4** surrogate the absence of enhancer by increasing the transfection efficiency 10-fold (24.8% for peptide **P2** and 27.7% for peptide **P4**), much higher than the control transfection efficiency of 16.9% with all reagents. But the replacement of *effectene* with the synthesized cationic peptides in the transfection process does not provide any increment in efficiency as observed from blue bars of subsequent series. The series denoted by orange bars where none of the reagents have been used shows no transfection at all denoted by orange bars. Considering the whole results and comparing the different series, the necessity of *effectene* in transfection process is proven although the synthesized cationic peptides (**P2** and **P4**) promisingly replace the *enhancer*.

After transfection a portion of the cells were fixed on cover slips and Figure **59** shows typical illustrative pictures of these cells ascertaining the green fluorescence due to GFP expression in the cytoplasm of S2 cells.

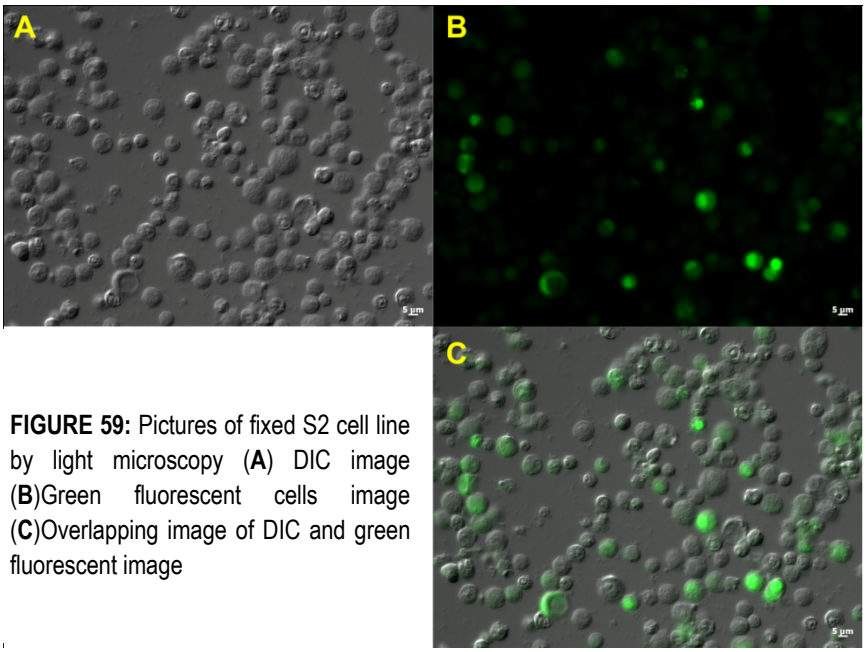


FIGURE 59: Pictures of fixed S2 cell line by light microscopy (**A**) DIC image (**B**)Green fluorescent cells image (**C**)Overlapping image of DIC and green fluorescent image

(ii) Cell permeation of fluorescent cationic peptides

Texas red covalently attached at the *N*-terminus of peptides was added to cell cultures without any liposome vehicle (Qiagen transfection reagents) and the internalization was observed by confocal fluorescence microscopy.

As shown in Figure **60(A-I)**, all the fluorescent peptides (**P8**, **P9** and **P10**), irrespective of the position of modification, traversed the cell membrane effectively under identical conditions. More remarkably, the fluorescent guanidino peptides **P8**

and **P9** appear to localize specifically in certain areas of cytoplasm although the relative uptake efficiency remains same.

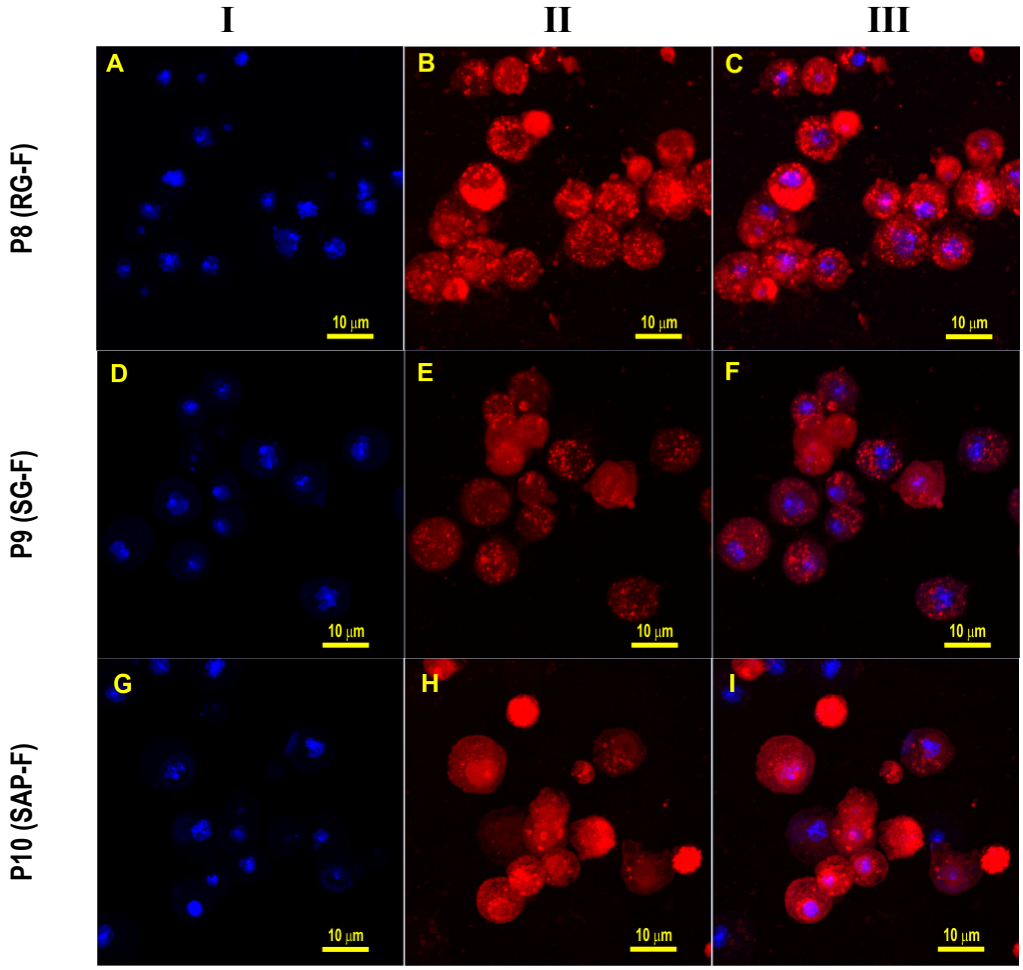


FIGURE 60: Confocal fluorescence microscopy analysis of S2 cells on treatment with fluorescent peptides. (**I**) DAPI stained image, (**II**) Fluorescent image of peptide staining, (**III**) Superimposed DAPI stained and fluorescent image of cells. (**A**-**C**) peptide **P8**, (**D**-**F**) peptide **P9**, (**G**-**H**) peptide **P10**

The fluorescent peptides were found to be taken up by the cells and no significant toxicity was observed since the size distribution of the S2 cells remained essentially unchanged in relation to untreated cells. A more convincing evidence of this uptake selectivity came from superimposition of fluorescent images (enhanced red signal) (Figure **60.II**) and DAPI stained images (Figure **60.I**) of individual cells treated with fluorescent peptides, which clearly showed that the fluorescence intensity was singular in the nucleus of the cells. The nucleus shows blue stain when treated with DAPI but the overlapped image (Figure **60.III**) does not show any difference in the intensity of fluorescence signal which confirms the distribution of fluorescent peptides inside the cells.

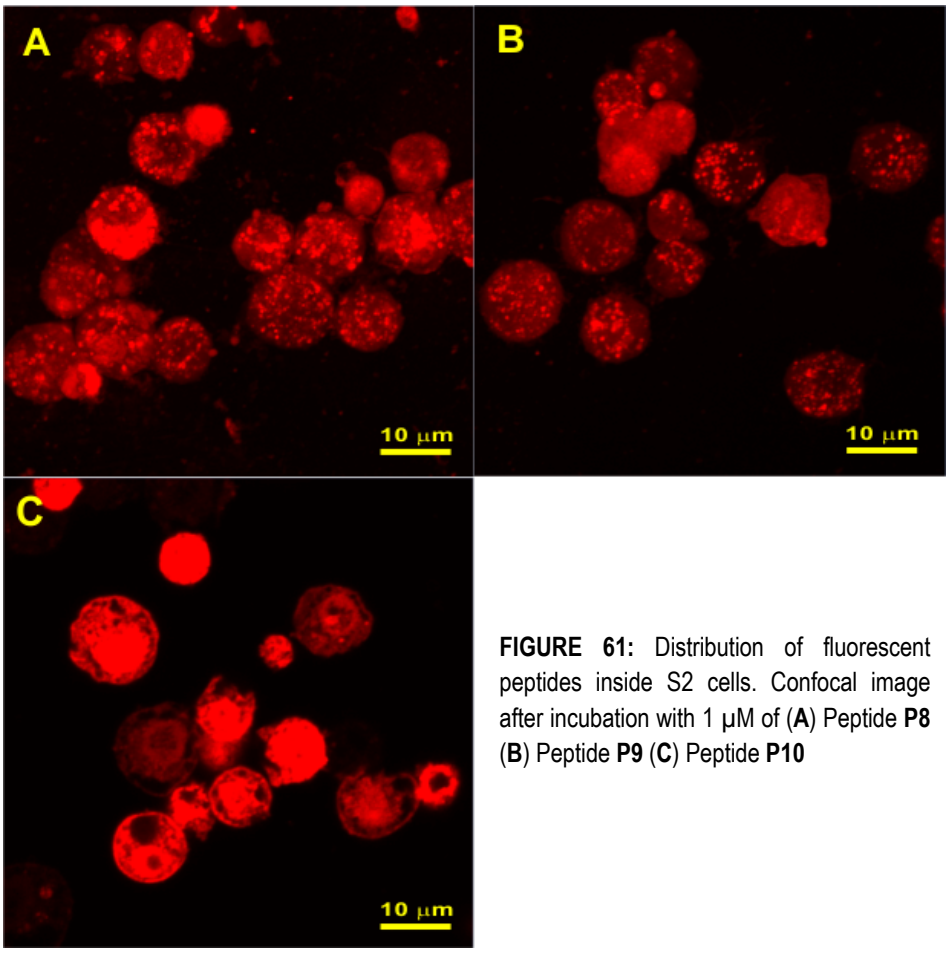


FIGURE 61: Distribution of fluorescent peptides inside S2 cells. Confocal image after incubation with 1 µM of (**A**) Peptide **P8** (**B**) Peptide **P9** (**C**) Peptide **P10**

The fluorescent SAP (**P10**) behaves slightly different from other two cationic peptides (**P8** & **P9**) can be observed by comparing the maximum intensity projection images in Figure **61** of all the three peptides. The peptide **P8** (Figure **61.A**) and **P9** (Figure **61.B**) appear to localize themselves into more organelles as viewed from the intense red punctates, whereas in case of fluorescent SAP peptide **P10** (Figure **61.C**) such spots are lacking, with the peptide spread all over the cytoplasm.

(iii) Cell transfection with fluorescent peptides

Similarly experiments were carried out with fluorescent peptides (**P8**, **P9** and **P10**). Qiagen's transfection reagents were considered to be one of the variables. The experiments could not be done at 1:25 ratio of plasmid DNA to fluorescent peptides as done for non-fluorescent peptides **P2** and **P4** due to interference of excess fluorescence from peptides.

A considerably lower amount (1 µM) of each fluorescent peptide was used for transfection of pRmHa3-GFP on S2 cells, as a result of which the effective

replacement of vehicular liposome (Qiagen's transfection reagents) did not show any positive result; rather a qualitative reflection of fluorescent peptides' transport and presence inside the cell, during transfection process, was confirmed.

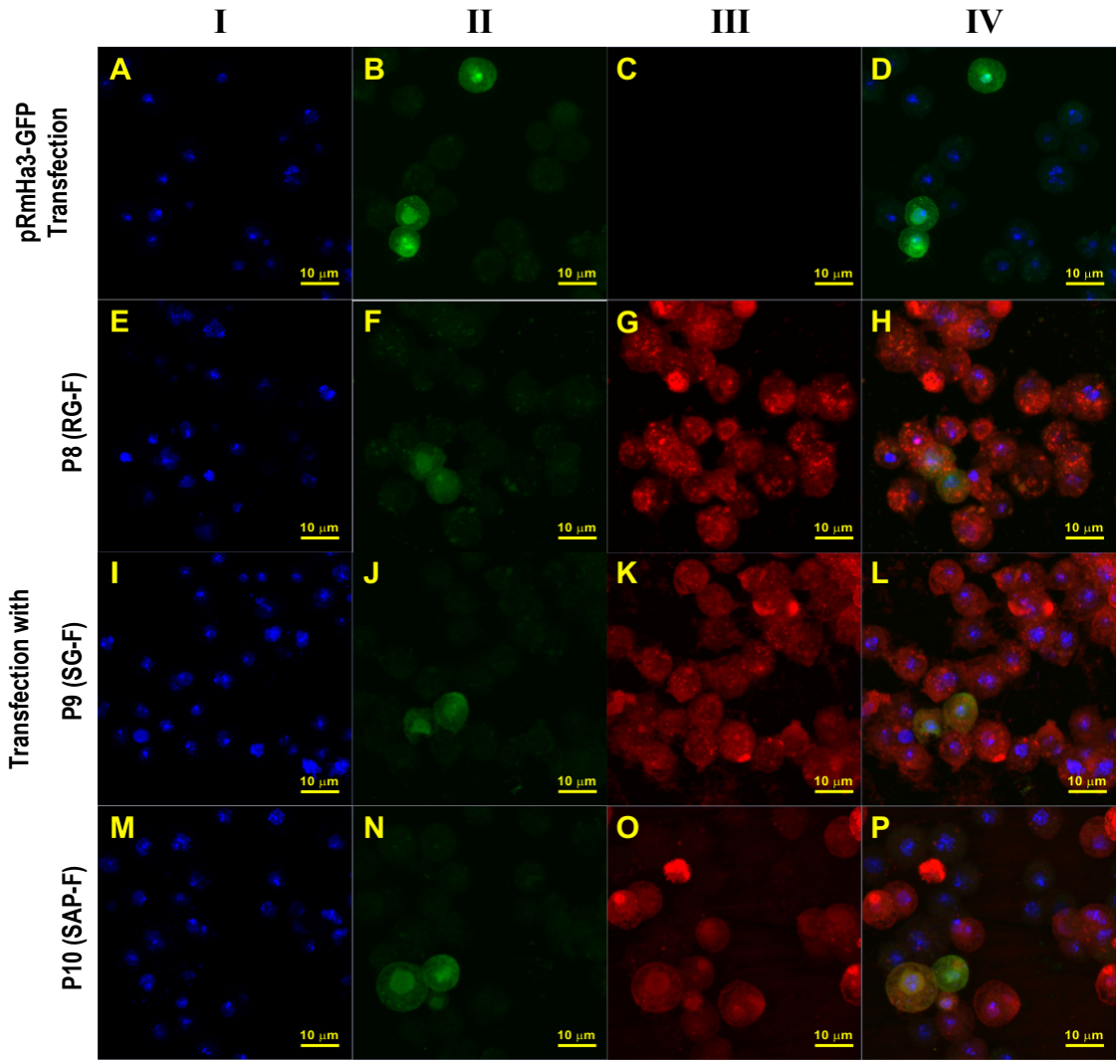


FIGURE 62: Fluorescence microscopy analysis of S2 cells on transfection with range of peptides (**I**) DAPI stained image, (**II**) Green fluorescence image for GFP expression, (**III**) Red fluorescence image of peptide staining, (**IV**) Superimposed DAPI stained and fluorescent image of cells. (**A-D**) *p*DNA transfection images (**E-H**) transfection images with peptide **P8**, (**I-L**) transfection images with peptide **P9**, (**M-P**) transfection images with peptide **P10**

The above Figure **62** illustrates the series of confocal images obtained after transfection of plasmid DNA into S2 cells along with the synthesized fluorescent peptides using the commercially available transfection reagents. The GFP expression after successful transfection in S2 cells is evident from the green fluorescent cells in Figure **62.B**. DAPI staining exclusively in the nucleus is seen in Figure **62.A**. A superimposed image of DAPI stain and GFP in Figure **62.D** shows uniform

expression of green fluorescent protein (GFP) throughout the cell. Enhanced intensity of bluish-green signal in circular patches inside the cell also confirms the same.

In case of peptide **P8** (Figure **62(E-H)**), peptide **P9** (Figure **61(I-L)**) and peptide **P10** (Figure **62(M-P)**) the cells fluoresce red apart from the green fluorescence due to GFP expression. Comparing 2nd and 3rd column in Figure **62**(**II** & **III**) it is also observed that the fluorescent red cells are more distinct than the few transfected green cells. An overlapping image corresponding to each peptide in the fourth column of Figure **62**(**IV**) suggests the prolonged co-existence of these peptides inside the cell even after transfection process.

1.9.3I Isothermal titration calorimetry

Having established that chimeric collageneous guanidino peptides (SG and RG) bind better to the DNA than amino collagens, ITC was used to investigate the equilibrium thermodynamics of peptide-DNA complex formation. For this the peptide **P4** (SG) showing best results so far was chosen.

In this experiment, the heat changes are directly measured upon the addition of small volumes of concentrated **P4** (SG) to the reaction cell containing the *p*DNA solution (Figure **63**). Integration of each peak after the titrant addition yields the calorimetric binding enthalpy (ΔH) as a function of the peptide concentration. In control experiments, the enthalpy of the peptide **P4** dilution was determined[123] and subtracted from the total change in enthalpy of the formation of the **P4**:*p*DNA complex. In contrast, *p*DNA dilution results in negligible heat changes under the conditions tested. Titrations were performed at ~20 °C, at which both the peptide and the *p*DNA undergo only limited conformational changes.[124]

A non-linear, least squares minimization software program (Origin 7.0 from Microcal Inc.) was used to fit the data and generate the titration curve using *one site* binding model for peptide:*p*DNA complex. The stochiometry of peptide to *p*DNA binding (N) was thus found to be between 12 to 13 peptide residues per *p*DNA strand. The free-energy change (ΔG) during complex formation was obtained using standard thermodynamic relationships from above data.

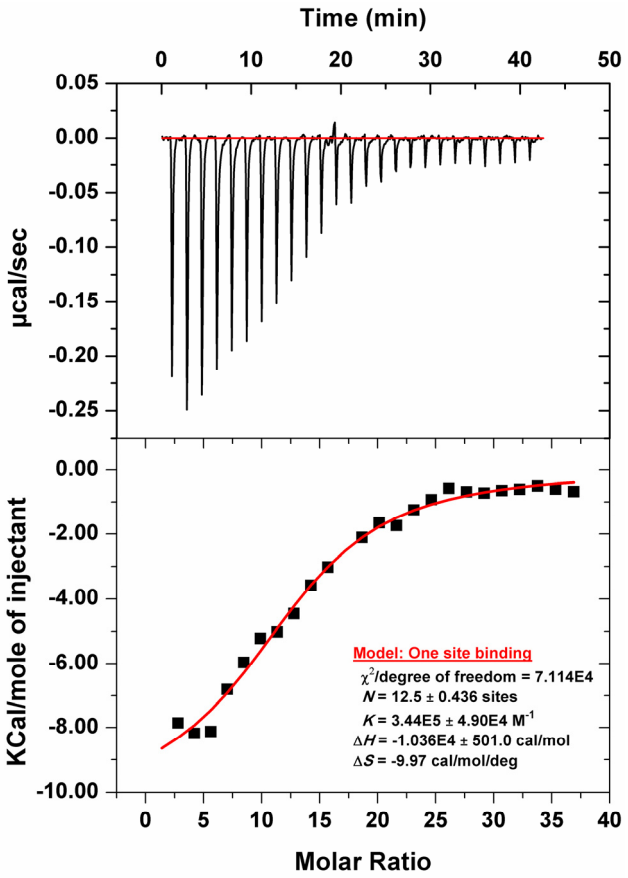


FIGURE 63: Isothermal titration of *p*DNA (pRmHa3-GFP) with the **P4** (SG) peptide; Titration of 1.3 µM DNA with 0.37 mM **P4** (upper panel). The peptide was added in 25 aliquots of 1 µl each, in 10 µM EDTA at 20 °C. Integration of these peaks and subtraction of the heats of dilution produces the binding isotherm shown in the lower panel. The points are fitted to a model for a single type and independent binding sites.

The thermodynamic parameters thus obtained at 273 K are listed above in Table **11**. The value of ΔG is negative under the conditions tested, favouring association, which is entropy-driven. However the reaction is exothermic. The ITC data confirm in a quantitative manner that there are more than 10 binding sites on the *p*DNA at room temperature and millimolar concentrations of the peptide.

Table 11: ITC thermodynamic parameters describing the interaction of peptide with DNA

Temp (K)	Concentration (mM)	ΔH (kCal/mol)	$T\Delta S$ (kCal/mol)	ΔG (kCal/mol)	K_a (μM^{-1})	N
273	0.37	-10.36	-2.72	-7.64	3.44×10^5	12.5

From these experimentally determined parameters, the free energy of binding (ΔG) and the entropy change (ΔS) are obtained using the standard thermodynamic relationship $\Delta G = -RT \ln K = \Delta H - T\Delta S$.

1.10 Discussion

In this section the focus has been on how the chemical nature, structural features and DNA complexation strategies of different amino and guanidino collagen peptides can be utilized for efficient DNA delivery. Several approaches to understand the role of physico-chemical properties of the DNA–peptide complex in relation to DNA delivery were attempted. Ligands bind to DNA with varying affinities and sequence specificities. As a consequence of change in microenvironment, the binding to DNA is always accompanied by change in the spectroscopic properties of either the ligand or DNA.

UV-T_m studies

A significant difference between the in UV T_m data of AT rich duplex-DNA **D1:D2** (36 ^oC) and CG rich duplex-DNA **D3:D4** (73.4 ^oC) obtained from their self-melting can be correlated to their strength of hydrogen bonding among complementary strands.

A binding agent such as a peptide when complexed to *ds*DNA modulates the stability of the double-helical structure resulting in an increase of T_m.[125] Such changes in T_m are related to the strength of peptide-DNA interactions.

The control peptide **P5** (SAP) when complexed with various ratios with the two different types of *ds*DNA, showed absolutely no change in case of AT rich duplex, whereas with CG rich duplex, the Tm of the complex was more than 80 ^oC, beyond the temperature range of UV-spectrophotometer (Table **8**). The amino and guanidino peptides (**P1**, **P2**, **P3** and **P4**) differ in the way in which they interact with AT rich *ds*DNA. Upon binding, the amino-peptides **P1** & **P3** stabilize the AT rich dsDNA **D1:D2** by 5-8 °C (Table **9**). Under identical conditions and ratios, the guanidino derivatives **P2** & **P4** produce larger change in the T_m about 9 to 10 °C due to the higher cataionic nature of the guanidino peptides. Even among these the (*R*)- and (*S*)- enantiomers the observed difference in the T_m suggest that in addition to electrostatic interaction from cationic side chains, the stereochemistry also matters. Further all the peptides more than 1:1 stochiometry were needed to see changes in T_m when (1:50 or 1:100) DNA:peptide.

Overall, the UV T_m data implies that the cationic peptides provide thermal stability to the small duplex oligonucleotides.

CD studies

Conformational analysis of guanidino-peptides

The guanidino peptides **P2** [-(Pro-AmpG-Gly)$_n$-] and **P4** [-(ampG-Pro-Gly)$_n$-] are direct modification of chimeric amino collagen sequences of **P1** [-(Pro-Amp-Gly)$_n$-] and **P3** [-(amp-Pro-Gly)$_n$-] respectively. **P1** and **P3** have been reported[105] to form well defined triple helices. In the present work the effect of incorporation of extra positive charge by means of guanidium groups onto these peptide sequences has been studied through CD spectroscopy. According to the results described in Section **1.9.3a**, there is a change in the CD pattern of the peptides on conversion of **P3** to **P4**, while transformation of **P1** to **P2** does not have much effect on overall pattern of the CD.

The peptide **P2** (RG) has a CD spectra as that of its precursor **P1** (R-Am). Positive maxima at ~221 nm and a intense negative maxima around 200 nm with an isobestic point at 215 nm are seen in the CD spectra of **P2**. The intensity of positive and negative bands change as a function of concentration, where the change of $R_{p/n}$ value as a function of peptide concentration suggests a coiled-coil type of structure.[109c] Nevertheless the triplex formation does not hold good any further as the change in $R_{p/n}$ values for peptide P4 does not match with reported literature,[109c] where it should increase sharply till the critical triple helical concentration is attended and thereafter remain constant. In case of **P4** (SG) a PPII kind[126] of CD spectra is observed with a strong negative band around 203-204 nm and no positive band. This CD pattern is unaffected by concentration of peptide, which rules out existence of any higher order structure structure unlike its enantiomer **P2** (RG).

Further thermal denaturation of these peptides (self melting) under different pH and solvent conditions supports the above hypothesis. At neutral pH (7.2) **P2** dissociates around 58 ^{o}C while **P4** ~50 ^{o}C. But at pH 12.0 no T_m value could be deduced for either of the peptides as they showed a linear decrease in ellipticity with increasing temperature, indicating that these peptides do not associate into any coiled-coil conformation at basic condition. The pK_a value is near about 13, and so guanidino peptides at basic pH 12.0 were expected to have a a greater proportion of both protonated & non-protonated species. The guanidino peptides probably loosen their conformational stability on switching from highly protonated (pH 7.2) to lesser protonated version at pH 12.0.

A decrease in the T_m values was observed on change of solvent system from aq. buffer (pH 7.2) to ethylene glycol for both **P2** and **P4**. This indicates the ethylene glycol molecules interfere in proper orientation of peptide strands in the solution as they contain hydrophilic carbon chains that prevent the water from entering into the main chain domain of peptides, thereby enhancing the hydrogen bond strength between amide hydrogen and main chain carbonyl oxygen atoms.[127] But the additional interstrand electrostatic repulsions arising from the cationic guanidino groups in these peptides remain unscreened in EG:W mixture. This repulsive interaction is aggravated by the absence of salt which also results in the decreased thermal stability. Since EG:W enhances the hydrogen bonding stability and decrease the electrostatic stability compared to H_2O, the formation of less stable PPII structure in EG:H_2O system indicates that electrostatic interactions play larger role for stabilizing the helical structure of 4-guanidinoproline containing collagen peptides.

Moreover a large difference in T_m values is found in case of **P2** (58 ^oC at pH 7.2; 41 ^oC in EG:W; ΔT_m = 7 ^oC) in comparison to **P4** with close T_m (50 ^oC at pH 7.2; 49 ^oC in EG:W; ΔT_m = 2 ^oC) values. This may due to the fact that in **P2** (RG) has a coiled-coil structure and the interaction of guanidium groups (in Y position) with bulk solvent and other molecules leads to opening up of the helix resulting in large difference in T_m values. **P4** on the otherhand has a PPII-like helix and the 4*S*-guanidinoproline residue present in the X position is more exposed and amenable to interactions with bulk solvent and other molecules in both aq. and EG:W systems does not induce much change in their denaturation temperatures.

Thus it can be said that **P2** (RG) forms more ordered structure in aqueous buffer with highly protonated side groups rather than **P4** (SG) which has similar conformational order in both aq. and ethylene glycol systems and the screening of positive charge have very little effect on its conformation.

Furthermore the presence of high concentration of salt had similar effect on both peptides. The intensities of positive and negative peaks increased with higher salt concentrations suggesting enhanced structural stability by increasing ionic strength & decreasing electrostatic free energy.[110,128] This stabilization by NaCl arises from the screening of interstrand electrostatic repulsion caused by cationic guanidino groups.

The CD data at physiological pH confirm that side chain extensions to proline do not affect the fold of the backbone, and that the polyproline scaffold maintains its

structure even with cationic groups along the same face of the helix although aggregation to triplex is lost.

Conformation of *ds*DNA-peptide complexes

P5 (SAP) which did not show any change in T_m values has no effective change in 280 nm region on complexation with AT rich *ds*DNA, but with CG-rich *ds*DNA an increase in this region indicates DNA complexation, which relaxes the DNA structure. For **P2**, **P3** and **P4** peptides, complexation with AT rich *ds*DNA brings about structural changes in the DNA which results in decreased intensity of the positive band ~280 nm, suggesting formation of slightly condensed DNA.[129] Only peptide **P1** induces a mixed structural change in the DNA conformation. At 1:50 ratio of (*ds*DNA : peptide) binding of DNA increases the ellipticity ~ 280 nm, and subsequently with a higher ratio of peptide (1:100), DNA undergoes slight compaction. Thus, a higher amount of aminopeptide **P1** (R-Am) is needed to bring about a condensed DNA conformation.

Conformation of plasmid DNA-peptide complexes

Interestingly each peptide showed different pattern of CD spectra on complexation with *p*DNA. For aminopeptides **P1** and **P3** there were drastic changes in the CD pattern which did not match with CD spectra of either *p*DNA alone or with peptides under similar conditions. This observation suggests that the original conformation of both pDNA and peptide are compromised in the complexation process. The **P5** (SAP) and guanidylated peptides on the other hand retain the peptide conformation to some extent as can be seen from the deep negative bands in 220 nm -190 nm region. The positive band of *p*DNA in ~280 nm region is seen to reduce in all the peptide cases, while the **P4** (SG) peptide has more significant change in the intensity of this band suggesting maximum condensation of the *p*DNA structure.[129]

Thus, it seems that the cationic guanidino peptide–DNA binding can cooperatively distort and loosen the DNA double helix into the less compact C form, which facilitates DNA condensation.[130] It must be pointed out that due to overlap in CD bands of DNA and peptides, a rigorous analysis of conformational changes in DNA-peptide complexes by CD is not straightforward and has limitations.

Fluorescence Studies

EtBr displacement assay with *ds*DNA

The displacement of intercalated ethidium by the addition of non-intercalative ligand happens as a consequence of structural transitions induced by the ligand. Significantly, neither of the peptides could completely displace the ethidium from the both types (AT rich & CG rich) of *ds*DNA suggesting EtBr to be a stronger ligand than any of the synthesized cationic peptides.

The binding of ethidium to *ds*DNA in itself is sequence dependent with GC rich sequences more preferred than the AT rich ones. Thus, ethidium displacement assay of peptides with AT rich *ds*DNA, more fluorescence decay is observed in comparission to CG rich duplexes.

EtBr exclusion assay with *ds*DNA

The protection offered by synthesized peptides to *ds*DNA from ethidium intercalation is not 100%. This is indicated by the lesser change in the fluorescence percentage with respect to only *ds*DNA – ethidium intercalation. Such observation is a result of the weaker binding affinity between the *ds*DNA and peptides.

It is also interesting to see that in the ethidium exclusion curves with CG rich *ds*DNA shows initial increase and then gradual fall in the fluorescence intensity. This result can be attributed to repulsion between cationic peptides and positively charged ethidium ions leading to collapse of DNA structure with release of bound ethidium.

EtBr displacement & exclusion assay with *p*DNA

Plasmid DNA being a large molecule intercalates ethidium very strongly and does not have much difference with respect to individual peptides' exclusion or displacement ability.

Interestingly in the displacement assay the fluorescence intensity increases sharply with the addition of first aliquot of peptide and falls thereafter with subsequent increase in peptide concentration. It can be suggested that there is a relaxation in the conformational state of *p*DNA due to peptide complexation at a lower concentration, followed by condensation with increasing concentration. Hence at relaxed state the *p*DNA allows extra ethidium present in the solution to intercalate which are later displaced with competitive binding from peptides and DNA condensation. Such behavior has also been observed for the interaction of several cationic lipids with

DNA and has been ascribed to initial charge neutralization followed by DNA collapse or "condensation", which results in a sudden release of bound ethidium.[131] Among all the synthesized cationic peptides **P4** (SG) presents noticeable change with intial fluorescence intensity indicating maximum conformational change being in agreement with previously discussed CD data.

The exclusion assay shows that the presence of these peptides on the *p*DNA surface is enough to efficiently block the access of EtBr to intercalation sites on the *p*DNA but only after a particular concentration. Overall the peptide complexed *p*DNA showed higher fluorescence intensity with respect to free *p*DNA at any particular concentration of EtBr. This observation is in line with the suggested structural effect observed during plasmid binding.

Electrophoretic mobility shift assay

The native PAGE experiment results do not offer any meaningful conclusion, except that the guanidino peptides when bound to either AT rich or CG rich *ds*DNA have less intense bands representing the duplex. This may be due to the fact that these gels were stained with EtBr, and the peptide complexed *ds*DNA is comparatively shielded from the ethidium ions. This behaviour has already been discussed under fluorescence studies.

The agarose gel electrophoresis gives slightly more meaningful results, where the guanidino peptides restrict the movement of *p*DNA on complexation with cationic peptides. This indicates charge neutralisation, which hinder the movement of negatively charged *p*DNA towards cathode. And since guanidino-peptides **P2** and **P4** are highly cationic they retard the DNA completely in the well where it had been placed. Further peptide **P4** (SG) is more effective in such neutralisation as the guanidium groups are more exposed in simple PPII conformation than a coiled coil structure of **P2** (RG).

Restriction Enzyme Digestion

The observations from this assay suggest that cationic guanidino peptides offer *p*DNA protection against restriction enzyme digestion. The intense lower band in free *p*DNA corresponds to supercoiled (SC) *p*DNA, which has been reported in the literature to be the most bioactive form.[132] While the enzyme treated free *p*DNA presented two linear bands. This is related to the BamH1 mechanism of action, which

cuts SC–DNA into L-DNA (linear DNA)[133] by cutting at two sites of the DNA. The absence or reduced intensity of L-bands in the formulations containing peptides indicates that the peptide maintained the protective capacity.

P2-*p*DNA complexes are partially cut by restriction enzyme, while **P4**-*p*DNA complexes are fully shielded from the enzyme action, apart from being retarded in agarose gel. This is an important advantage of non-viral systems in gene therapy for their capacity to protect DNA from components of the medium, above all from digestion by DNases. However, the protective role of free cationic peptides depends not only on their positive charge, but also on the primary structure of the peptide.

Cell studies

Positively charged amino acids not only enable cationic peptides to bind DNA, but are also important for electrostatic interactions with the anionic groups of the cell surface. Incorporation of synthesized amino and guanidino-peptides significantly increased the number of cells expressing green fluorescent protein (GFP), with **P4** inducing maximum transfection as seen from the GFP expression. Further analysis suggested that the guanidino peptides efficiently replace the commercially available *enhancer* that is used as DNA condensing agent, rather use of **P2** and **P4** in place of *enhancer* increase the transfection efficiency by 1.5 times.

Fluorescent counterpart of these guanidino-peptides (RG-F; **P8** & SG-F; **P9**) also permeate the cell and localise themselves into certain organelles unlike **P10** (SAP-F) which distributes itself uniformly into the cell cytoplasm. This observation is linked with peptide release mechanism inside the cell.

Calorimetric binding study

Investigation of thermodynamic parameters of guanidino peptide **P4** (SG), which has been showing best results in efficiently complexing DNA, shows entropy driven association, and confirm in a quantitative manner that there are more than 12 binding sites on the DNA at room temperature and at millimolar concentrations of the peptide. But this associative binding due to non-specific interactions does not have very high binding constant (3.4×10^5 Kcal/mol) to be challenged with intercalators (like EtBr) and groove binders. Such entropy-enthalpy compensation has been described for other biological systems discussed in literature.[134]

General discussion

In comparison to polypeptides that bind to specific DNA sequences,[135] the results shown here indicate that very high molar ratios of guanidino peptides are required to saturate the binding sites of DNA. The high density of peptides on the DNA double helix seems clearly too large to account for a tight and specific interaction of all peptides or to correlate with a model where the helix is well-inserted within the major groove of the DNA. Instead, one could imagine that several layers of peptides reside around the DNA, possibly by interacting with each other through their cationic and hydrophobic surfaces. Alternatively, a high density of peptides can be created by end-on association of the peptides with the DNA double helix. In this case, one charged peptide terminus is in contact with the DNA, whereas the other one is oriented away from the double helix. In both models, the DNA is covered by peptides exposing a positively charged surface. These charges are available for interactions with neighboring DNA strands, thereby condensing the elongated DNA double helix into large globular aggregates. The charged surface permits the complex to stay in suspension and helps association with negatively charged biomolecules, for example, at the cellular outer layers. Notably, DNA condensation has been demonstrated to be a critical feature for DNA stability and enhanced transfection efficiency.[136] The above-mentioned models provide an explanation of how the high peptide density can be accommodated.

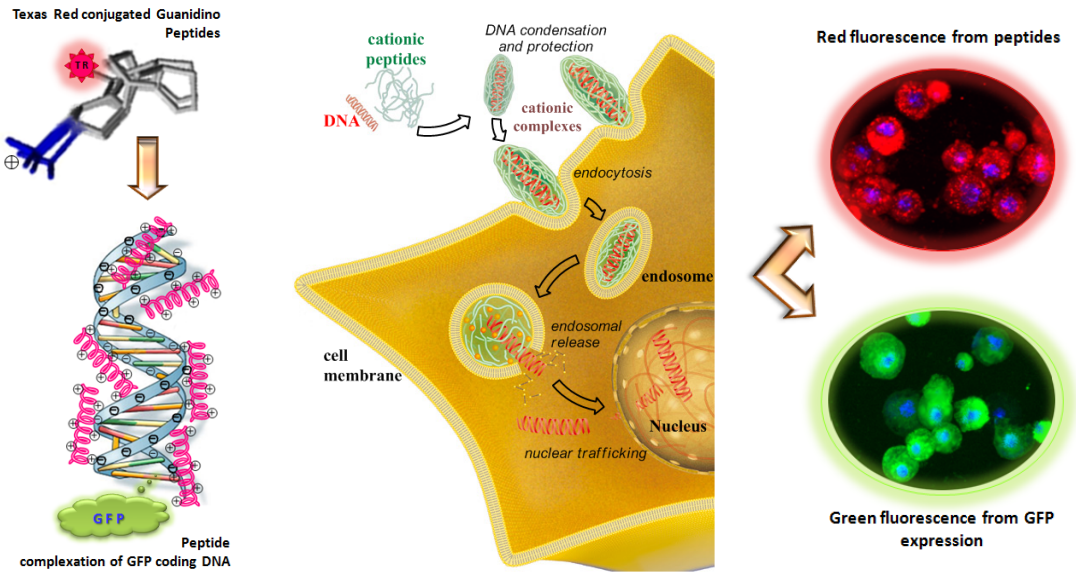


FIGURE 64: Model illustrating (**A**) Cationic peptide complexation with GFP coding DNA, (**B**) Mechanistic pathway of peptide-DNA complex during transfection of eukaryotic cells, (**C**) Expression of fluorescent peptide and GFP in the cells.

A probable mechanistic model and pathway of these peptides is suggested in Figure **64**. The guanidino peptides bind to DNA by nonspecific interactions, generating large complexes. After cell binding, the peptide-*p*DNA complexes enter the cell via an endosomal pathway.[137] Other processes, which are thus far much less well-understood, help in the transfer of DNA to the nucleus, where the genetic information is expressed.

1.11 Conclusions

In conclusion it is demonstrated that the stereoregular, enantiomeric, 4-aminoproline and 4-guanidinoproline peptides with both (4*R/S*) stereochemistry can bind to DNA. The binding of these peptides exhibits considerable amount of sequence specificity. These peptides were also shown to induce conformational transitions in both *ds*DNA and the *p*DNA and condense DNA. These peptides have future as use of artificial DNA transfection-enhancer agents or in the design of novel DNA binding agent.

These preliminary, but positive results are encouraging and further studies are necessary to establish the exact sequences to which these peptides bind, and elaborate them for specific applications of gene expression.

1.12 Experimental

1.12.1 Synthesis of compounds/monomers

General remarks: All reagents and chemicals were of laboratory or analytical grade obtained from commercial sources and were used without further purification except in cases mentioned. Thin layer chromatography (TLC) was carried out on pre-coated silica gel 60 F_{254} plates (E. Merck). TLCs were visualized under UV light, iodine and/or ninhydrin spray followed by heating up to 110 ^{o}C with heat gun. Silica gel 60-120 and 100-200 mesh (Merck) was used for routine column chromatography with ethyl acetate/petroleum ether or dichloromethane/methanol mixture as elution solvent depending upon the compound polarity and chemical nature whereas for flash chromatography 230-400 mesh silica gel was used. All solvents were distilled under an inert atmosphere with appropriate desiccant. Reactions in aqueous medium and workup processes were done using double distilled water. Unless otherwise noted, all reactions were carried out at room temperature.

^{1}H NMR spectra were routinely recorded at 200MHz on a Bruker AC-200 instrument controlled by an Aspect 2000 computer. ^{13}C NMR and ^{13}C-DEPT spectra (at 50MHz) were recorded on the same instrument. The spectra were analyzed using ACD spec-viewer software from ACD labs. For some compounds, NMR spectra were also recorded on 400MHz JEOL spectrometer; and data processed Cambridge Soft's MestReNova software. All chemical shifts are referenced with respect to TMS as internal standard and are expressed in δ-scale (ppm). Mass spectra were obtained by ESI-MS technique on AP-QSTAR spectrometer. Melting points of the samples were determined in open capillary tubes using Büchi Melting Point M-560 apparatus and are uncorrected. IR spectra were recorded on an Infrared Fourier Transform Spectrophotometer using chloroform, nujol or neat sample. The optical rotation values were measured on JASCO P2000 polarimeter.

Resins for solid phase peptide synthesis and Fmoc-protected amino acids (Fmoc-Pro, Fmoc-Gly, fmoc-Phe, Fmoc-Val, Fmoc-Arg-pbf and Fmoc-Leu) were bought from Novabiochem and were used without further purification.

The fluorescent molecular probe Texas Red®-succinimidyl ester was obtained from Invitrogen whereas the transfection reagents were procured from QIAGEN. The plasmid DNA (pRmHa3-GFP) was received as a gift from A. J. Courey's Lab; ULCA.

The apparatus used in the processes like centrifuge tubes, microtips, slides and other glasswares were autoclaved for 30 min before use.

1.12.1a Buffers & solutions for biophysical studies

Salts and reagents used in buffer preparation such as NaCl, Na_2BO_4, NaH_2PO_4, Na_2HPO_4 and Tris base etc. were obtained from Sigma-Aldrich and were of molecular biology reagent grade. Double distilled water was demineralized using Millipore MilliQ deionizer and was used for the preparation of buffers. pH was adjusted using NaOH and HCl in case of phosphate and borate buffer except in case of acetate buffer where acetic acid was used.

Loading dye-6X for agarose gel

Sucrose (4 g) and bromophenol blue (25 mg) in deionised H_2O (10 ml) final volume was stored at 4°C, and appropriate amount was added to DNA sample at 1:5 ratio, e.g. 5 µl to 25 µl.

Loading dye-6X for PAGE

Glycerol (3 ml, 30%), Bromophenol Blue (25 mg) were dissolved in deionized H_2O (10 ml)

Acrylamide:bisacrylamide (29:1)

A mixture of acrylamide (29.0 g) and *N,N'*-methylenebisacrylamide (1.0 g) was dissolved in minimum amount of deionised H_2O (15 ml) and diluted to 100 ml with deionised H_2O.

Ammonium persulfate (10 % w/v)

$(NH_4)_2S_2O_8$ (ammonium persulfate, 228 mg) was dissolved in deionised H_2O (1 ml) was used as a catalyst for the copolymerization of acrylamide and bisacrylamide gels.

Borate buffer (pH = 12.0, 10 mM NaCl)

Borax ($Na_2B_4O_7.10H_2O$ or sodium tetraborate, 762 mg) and Boric Acid (H_3BO_3, 123 mg) was mixed in deionised H_2O (200 ml) contains of NaCl (117 mg). The pH was raised to 12 with NaOH solution.

Ethidium bromide solution

Sigma BioReagent EtBr (3,8-Diamino-5-ethyl-6-phenylphenanthridinium bromide, 15.76 mg) was dissolved in of deionized H_2O (200 µl) to obtain a stock solution (1 mM). It was diluted to appropriate concentration as per the requirement whenever necessary. It was stored in cool and dark place.

PBS buffer – 1X (pH=7.4)

Phosphate-buffered saline was prepared by NaCl (800 mg), KCl (20 mg), Na_2HPO_4 (144 mg), KH_2PO_4 (24 mg) in deionized H_2O (80 ml). The pH adjusted to 7.4 with aq. HCl and the volume made up to 100 ml with deionised water. The buffer was then sterilized by autoclave and stored at RT.

Phosphate buffer (pH = 7.2, 10mM NaCl)

Na_2HPO_4 (110 mg), $NaH_2PO_4 \cdot 2H_2O$ (35.3 mg), NaCl (58.5 mg) was dissolved in minimum quantity of water and the total volume was made 100 ml. The pH of the solution was adjusted 7.2 with aq. NaOH solution (in deionised water), and stored at 4 ^oC.

TAE buffer – 50X (pH=7.4)

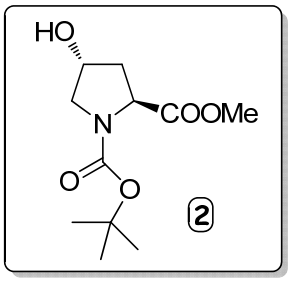
Tris-acetate-edetate (TAE) buffer was prepared from tris(hydroxymethyl)amino methane (Tris base, 242 mg] was dissolved in deionised water (80 ml) with glacial acetic acid (5.7 ml) and EDTA (10 ml 0.5 M) was added to it. The pH was adjusted to 7.4 with aq. HCl and volume was made up to 1 liter with deionised water. This buffer was stored at RT.

TBE buffer – 5X (pH=8.0)

For preparation of Tris-borate-edetate (TAE) buffer, the mixture of tris(hydroxymethyl)amino methane (tris-base, 27.0 g), boric acid (H_3BO_3, 13.75 g) and EDTA (1.46 g) were dissolved in minimum quantity of water and the total volume was made 500 ml. The pH was adjusted to 8.0 and stored at 4 ^oC. 100 ml of 5X-TBE diluted upto 500 ml with deionised water to 500 ml to obtain 1X-TBE.

1.12.1b Experimental procedures and spectral data

(2*S*, 4*R*)-N^1-(t-butoxycarbonyl)-4-hydroxyproline methylester (2)



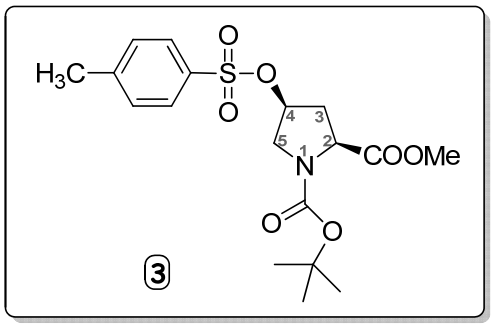
A solution of trans-4*R*-hydroxyproline (1 g, 7.6 mmol) in a solution of aq. NaOH (2N, 10 ml) and dioxane (10 ml) was cooled to 0 ^oC and Boc-anhydride (2 ml, 8.4 mmol) was added dropwise using addition funnel. The reaction mixture was stirred at 0 ^oC for 1 h, and dioxane was removed under reduced pressure. The residue was extracted with diethyl ether (5 x 25 ml) to remove unreacted $(Boc)_2O$ The aqueous layer was then vigorously stirred by ethyl acetate under ice-cold condition and acidified with saturated $KHSO_4$ solution to pH 3. The ethyl acetate layer was separated and the aqueous layers were further extracted into ethyl acetate (3 x 20 ml). The organic layer were pooled and washed with water followed by brine, dried over anhydrous Na_2SO_4 and concentrated under vacuum. Removal of ethyl acetate under reduced pressure, gave a white solid which was recrystalized with EtOAc/Petroleum-ether.

The white solid compound (1 g, 4.3 mmol) obtained in the above step was mixed with anhydrous K_2CO_3 (1.8 g, 13 mmol) and dissolved in anhydrous acetone (25 ml). The reaction mixture was stirred for 30 min at room temperature and dimethylsulphate (4.9 ml, 5.18 mmol) was added. Stirring was continued under reflux conditions for 5 h, when TLC confirmed maximum product formation. The solvent was evaporated under reduced pressure and resulting solid was dissolved in ethyl acetate (50 ml), washed with water (2 X 30 ml), followed by saturated brine solution and dried over anhydrous Na_2SO_4. The ethyl acetate extract was concentrated under reduced pressure to a pale yellow solid which was then purified by silica gel chromatography, eluting with ethylacetate:pet ether (1:1) to yield compound **2** as white crystalline solid. Yield 1.54 g, 82.3% over the two steps.

Mol. Formula	:	$C_{11}H_{19}NO_5$
mp	:	91.8 – 93.1 ^oC
Mol. Weight	:	245.27
ESI-MS *m/z*	:	246.3 $[M+1]^{+\bullet}$
^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.41 & 1.46 (2s, 9H), 2.0-2.13 (m, 1H), 2.2-2.33 (m, 1H), 3.58-3.62 (m, 2H), 3.73 (s, 3H), 4.35-4.49 (m, 2H)
^{13}C NMR (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) 28.1, 38.3 (mi), 38.9 (ma), 52.0 (ma), 52.1 (mi) 54.5, 57.5 (mi), 57.9 (ma), 69.0 (ma), 69.7 (mi), 80.4, 154.0 (ma), 54.6 (mi), 173.5 (mi), 173.7 (ma)

^{13}C-DEPT (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) *Positive Peaks:* 28.1, 32.3 (mi), 52.0 (ma), 52.2 (mi), 54.5, 57.5 (mi), 57.9 (ma),69.0 (ma), 69.7 (mi) *Negative Peaks:* 38.3 (mi), 38.9 (ma), 54.5

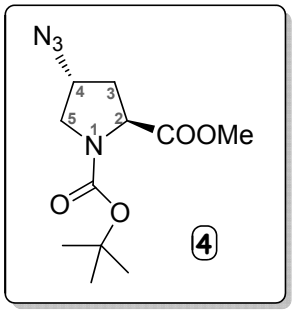
(2*S*, 4*S*)-N^1-(t-butoxycarbonyl)-4-(p-toluenesulfonyloxy)proline methylester (3)



A mixture of compound **2,** (1 g, 4 mmol), PPh_3, (1.2 g, 4.5 mmol) and *p*-toluenemethylsulfonate (6.8 ml, 4.5 mmol) was dissolved in dry THF (15 ml) cooled to 0 ^oC on ice bath, under Argon. After stirring for 30 min at ice-cold temperature, Diisopropylazodicarboxylate (8.8 ml, 4.47 mmol) was added slowly with syringe. The reaction mixture was brought to RT and stirred for another 8 hrs after which THF was removed under vacuum. The resulting orange colored thick oil was dissolved in diethyl ether (60 ml). The resulting solution was kept overnight at room temperature and the white crystals were filtered and the residue was washed with petroleum ether followed by diethyl ether. The filtrate was concentrated under reduced pressure and column purified with silica gel, which gave compound **3** as white crystalline powder. Yield 1.56 g (62.4%).

Mol. Formula	:	$C_{18}H_{25}NO_7S$
mp	:	75.2 – 78.3 ^oC
Mol. Weight	:	399.46
ESI-MS *m/z*	:	422.57 $[M+Na]^{+\bullet}$
$[\alpha]^{25}_D$	:	-21.77 (C=0.58, $CHCl_3$)
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3320, 3019, 2982, 1695, 1598, 1404, 1368, 1307, 1215, 1176, 1133, 1052, 959, 901, 757, 668, 620, 555, 485
^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.40 &1.44 (2s, 9H), 2.33-2.45 (m, 2H), 2.45 (s, 3H), 3.62 (m,2H) 3.69 (s, 3H), 4.30-4.47 (m, 1H), 5.03-5.05 (m, 1H), 7.33-7.37 (dd, 2H), 7.70-7.79 (dd, 2H)
^{13}C NMR (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) 21.6 (ma), 21.9 (mi), 28.1, 35.9 (ma), 35.9 (mi), 51.5 (ma), 52.2 (mi), 56.9 (mi), 57.3 (ma), 69.8, 78.9, 80.5, 127.7, 129.9, 133.4, 145.1, 153.3 (ma), 153.7 (mi), 171.6 (mi),171.9 (ma)
^{13}C-DEPT (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) *Positive Peaks:* 21.6 (ma), 21.9 (mi), 28.1, 52.2, 56.9 (mi), 57.3 (ma), 69.8, 77.7, 78.9, 127.7, 129.9 *Negative Peaks:* 35.97 (mi), 36.9(ma), 51.5 (ma), 52.0 (mi)

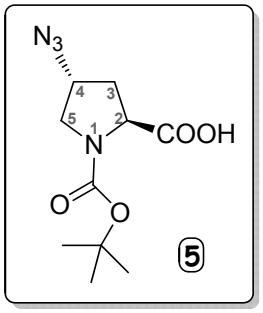
(2*S*,4*R*)N¹-(t-butyloxycabonyl)-4-azidoproline methylester (4)



A solution of compound **3,** (1 g, 2.5 mmol) and NaN_3, (1.3 g, 20 mmol) in dry DMF (10 ml) was stirred at 55-60 ^oC for 12 hrs under Argon. When TLC indicated complete conversion of the starting material, DMF was removed under vacuum and the residue was dissolved in water. The aqueous layer was extracted with ethyl acetate (3 x 25 ml). The combined organic layer was washed with water followed by brine, dried over anhydrous Na_2SO_4 and concentrated under vacuum. The crude product obtained was purified by silica gel chromatography (40% ethyl acetate/hexane elute) to afford compound **4** as colorless thick oil. Yield 0.65 g (96%).

Mol. Formula	:	$C_{11}H_{18}N_4O_4$
Mol. Weight	:	270.29
ESI-MS *m/z*	:	293.23 $[M+Na]^{+\bullet}$
$[\alpha]_D^{25}$	:	-31 (C = 1, EtOH)
IR (thin film /neat)	:	**ν** (cm^{-1}) 3307, 2982, 2105, 1714, 1394, 1256, 1179, 1110, 1025, 976, 920, 896, 855, 767, 622, 567, 475
^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm)1.38 &1.43 (2s,9H), 2.07-2.34 (m,2H), 3.40-3.57 (m, 2H), 3.71 (s, 3H), 4.07-4.19 (m, 1H), 4.26-4.42 (m, 1H), 4.88-5.0 (t,1H)
^{13}C NMR (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) 28.1, 35.2, 36.1, 51.1, 52, 52.2, 58.6, 59.1, 80, 153.8, 256.3, 172.8
^{13}C-DEPT (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) *Positive Peaks:* 28.1, 52, 52.2, 58.6, 59.1 *Negative Peaks:* 35.2, 36.1, 51.1

(2*S*,4*R*)N¹-(t-butyloxycabonyl)-4-azidoproline (5)

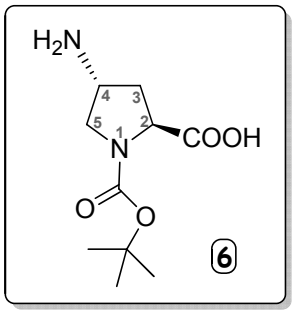


The methyl ester **4** (1 g, 3.7 mmol) was hydrolysed using aqueous NaOH (2N) in water-MeOH (1:1, 10 ml) mixture. The reaction mixture was stirred in RT for 2 hrs and methanol was removed under reduced pressure. The aqueous layer was acidified with saturated $KHSO_4$ solution, extracted with ethyl acetate and dried over Na_2SO_4. The free acid **5** was obtained as a dark brown liquid upon concentration. Yield 0.9 g (95%)

Mol. Formula	:	$C_{10}H_{16}N_4O_4$
Mol. Weight	:	256.26
ESI-MS *m/z*	:	279.14 $[M+Na]^{+\bullet}$
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3320, 3019, 2980, 2107, 1694, 1477, 1408, 1369, 1310,

		1261, 1216, 1162, 1135, 1067, 929, 851, 767, 669, 621, 485
^{1}H NMR (200 MHz, MeOH-d_4)	:	δ_H (ppm) 1.43, 1.47 & 1.49 (3s, 9H), 2.13-2.48 (m, 2H), 3.46-3.61 (m, 2H), 4.1-4.43 (m, 2H)
^{13}C NMR (50 MHz, MeOH-d_4)	:	δ_C (ppm) 28.5 (ma), 28.7 (mi), 37.3 (ma), 36.6 (mi), 52.5 (ma), 52.8 (mi) 58.9 (mi), 59.4 (ma), 60 (mi), 60.5 (ma), 81.8 (mi), 82.1 (ma), 155.7, 176.2
^{13}C-DEPT (50 MHz, MeOH-d_4)	:	δ_C (ppm) *Positive Peaks:* 28.5 (ma), 28.7 (mi), 58.9 (mi), 59.4 (ma), 60 (mi), 60.5 (ma) *Negative Peaks:* 37.3 (ma), 36.6 (mi), 52.5 (ma), 52.8 (mi)

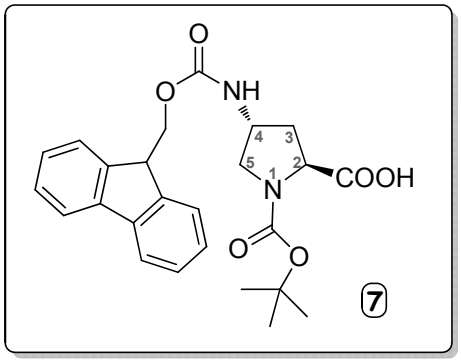
(2*S*,4*R*)N^1-(t-butyloxycabonyl)-4-aminoproline (6)



Compound **5** (1 g, 3.9 mmol) was taken in methanol (10 ml) and to this was added 10% Pd-C (0.15 g). The mixture was subjected to hydrogenation in a Parr-Hydrogenation apparatus at a pressure of 40 psi for 4 hrs. Pd-C suspension was removed by filtration through Celite and the filtrate was evaporated under reduced pressure to give reddish brown liquid. The (2*S*,4*R*)N^1-(*t*-butoxycarbonyl)-4-aminoproline **6** thus obtained (0.88 g, 98% yield) was used without further purification.

Mol. Formula	:	$C_{10}H_{18}N_2O_4$
Mol. Weight	:	230.26
ESI-MS *m/z*	:	231.19 $[M+1]^{+\bullet}$,253.19 $[M+Na]^{+\bullet}$
IR (thin film on $CHCl_3$)	:	ν (cm^{-1}) 3421, 3019, 2975, 2654, 2178, 2120, 1690, 1602, 1408, 1365, 1299, 1257, 1215, 1174, 1137, 1120, 1067, 983, 956, 904, 851, 759, 670, 620, 576, 488
^{1}H NMR (400 MHz, MeOH-d_4)	:	δ_H (ppm) 1.45 & 1.47 (2s, 9H), 2.28-2.34 (m, 2H), 3.42-3.49 (m, 1H), 3.83-3.9 (m, 1H), 4.15-4.29 (m, 1H)

(2*S*,4*R*)-N^1-(t-butoxycarbonyl)-4-(9-fluorenylmethyloxycarbonylamino) proline (7)

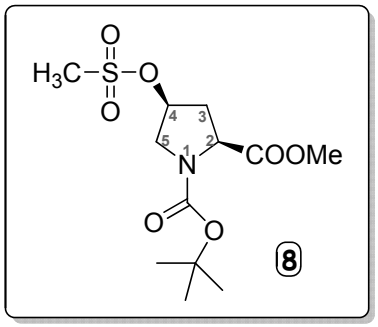


An aqueous solution (100 ml) of compound **6** (1 g, 4.3 mmol) was cooled to 0 °C and 10% aqueous Na_2CO_3 (20 ml) and dioxane (100 ml) were added. Fmoc-Cl (1.7 g, 6.5 mmol) dissolved in dioxane (10 ml) was added dropwise during 1 hr. The resulting mixture was stirred for 15 hrs at room temperature,

maintaining the pH of solution around 8.0. The reaction mixture was concentrated on a rotary evaporator to remove dioxane and to the resulting slurry 50 ml of water was added. The mixture was extracted with diethyl ether (5 x 50 ml) to remove unreacted Fmoc-Cl. The aqueous layer was then vigorously stirred by ethyl acetate under ice-cold condition, and acidified with saturated $KHSO_4$ solution to pH 3. The ethyl acetate layer was separated and the aqueous layer was further extracted into fresh ethyl acetate (3 x 50 ml), the organic layer were pooled and washed with water followed by brine, dried over anhydrous Na_2SO_4 and concentrated under vacuum. The crude material was purified by silica gel chromatography (80% ethyl acetate/hexane elute) afford compound **7** as white solid. Yield 1.28 g (66%).

Mol. Formula	:	$C_{25}H_{28}N_2O_6$
mp	:	Decomposes around 183 – 189°C
Mol. Weight	:	452.50
ESI-MS *m/z*	:	453.30 $[M+1]^{+\bullet}$, 475.57 $[M+Na]^{+\bullet}$, 491.33 $[M+K]^{+\bullet}$
$[\alpha]_D^{25}$	:	-11 (c = 0.2, EtOH)
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3307, 3020, 2923, 2400, 1667, 1546, 1412, 1311, 1216, 1157, 1041, 928, 758, 669, 625
^{1}H NMR (400 MHz,DMSO-d_6)	:	**δ_H** (ppm) 1.36 & 1.39 (2s, 9H), 2.05-2.15 (bm, 2H), 3.16-3.23 (bs,1H), 4.07 (bs, 1H), 4.2-4.23 (m, 2H), 4.34-4.36 (m, 2H), 7.31-7.35 (m, 2H), 7.39-7.43 (t, 2H), 7.64 (bs, 1H), 7.68-7.7(m, 2H), 7.87-7.89 (d, 2H), 12.57 (s, 1H)
^{13}C NMR (125 MHz,DMSO-d_6)	:	**δ_C** (ppm) 25 (ma), 25.2 (mi), 33.4 (mi), 34.5 (ma), 48.5, 54.1 (ma), 54.4 (mi) 61.9 (mi), 62.2 (ma), 71.8, 88.8, 140.3, 146.5, 148.9, 149.6, 166.1, 169.9, 181.4 (ma), 181.9 (mi), 184.8, 207.2 (mi), 207.8 (ma)
^{13}C-DEPT (125 MHz,DMSO-d_6)	:	**δ_C** (ppm) *Positive Peaks:* 25 (ma), 25.2 (mi), 48.5, 61.9 (mi), 62.2 (ma), 88.8, 140.3, 146.5, 148.9, 149.6, 166.1, 169.9 *Negative Peaks:* 33.4 (mi), 34.5 (ma), 54.1 (ma), 54.4 (mi), 71.8

(2*S*,4*R*)N-(t-butyloxycabonyl)-4-O-mesyl proline methylester (8)

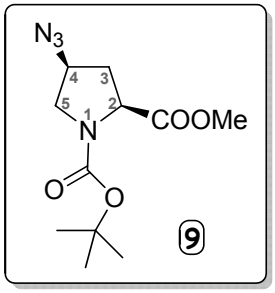


The solution of methylester **2** (1 g, 4.1 mmol) and triethylamine 1.7 ml (12.2 mmol) in 20 ml of dry dichloromethane was cooled to 0 °C on ice bath under Argon. While stirring (4.7 ml, 6.1 mmol) methanesulfonyl chloride was added in one shot. Stirring was continued over a period of 3 hrs at 0 °C. The reaction mixture was then washed with water, followed by saturated brine solution. The organic layer was dried over anhydrous

Na_2SO_4 and concentrated under vacuum. The crude material was purified by silica gel chromatography (30% ethyl acetate/hexane) which afforded mesylated compound **8** as a white solid. Yield 1.26 g (95.8%)

Mol. Formula	:	$C_{12}H_{21}NO_7S$
mp	:	83 – 86.4 ^oC
Mol. Weight	:	323.1
ESI-MS *m/z*	:	346.30 $[M+Na]^{+\bullet}$
$[\alpha]_D^{25}$	:	-41.22 (C=0.55, $CHCl_3$)
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3319, 3020, 2400, 1744, 1703, 1599, 1405, 1306, 1264, 1215, 1174, 1135, 1068, 905, 758, 669, 620, 485
^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.42 & 1.47 (2s, 9H), 2.17-2.30 (m, 1H), 2.5-2.71 (m, 1H), 3.04 (s,3H), 3.71-3.87 (m, 5H), 4.32-4.48 (m, 1H), 5.24-5.3 (m, 1H)
^{13}C NMR (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) 27.7 (mi), 28.1 (ma), 36.2 (mi), 37.4 (ma), 38.6, 52.1, 52.2, 57.0 (mi), 57.4 (ma), 77.9 (ma), 78.2 (mi), 80.8, 153.3 (ma), 153.9 (mi), 172.7
^{13}C-DEPT (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) *Positive Peaks:* 27.7 (mi), 28.1 (ma), 57.0 (mi), 57.4 (ma), 77.9 (ma), 78.2 (mi) *Negative Peaks:* 32.2 (mi), 37.4 (ma), 38.6, 52.1, 52.2

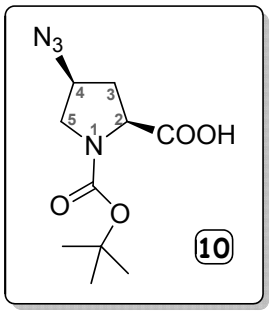
(2*S*,4*S*)N¹-(t-butyloxycabonyl)-4-azidoproline methylester (9)



(*2S,4S*)-azido compound **9** (0.75 g, 90%) was synthesized from (1 g, 3.1 mmol) the mesylate **8** accordingly as mentioned earlier for compound **4**.

Mol. Formula	:	$C_{11}H_{18}N_4O_4$
Mol. Weight	:	270.29
ESI-MS *m/z*	:	293.23 $[M+Na]^{+\bullet}$
IR (thin film / neat)	:	**ν** (cm^{-1}) 3502, 2978, 2509, 2106, 1755, 1704, 1478, 1398, 1367, 1260, 1161, 1120, 1055, 992, 951, 893, 865, 842, 770, 624, 567
^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.37 & 1.43 (s, 9H), 2.07-2.108 (m, 1H), 2.33-2.52 (m, 1H), 3.41-3.49 (m, 1H), 3.64-3.66 (m, 1H), 3.71 (s, 3H), 4.06-4.13 (d, 1H), 4.25-4.42 (m, 1H)
^{13}C NMR (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) 27.8, 34.7, 35.6, 50.4, 51.9, 57.0, 57.3, 57.9, 58.9, 80.1, 153.0, 153.6, 171.6, 171.9
^{13}C-DEPT (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) *Positive Peaks:* 27.8, 51.9, 57.0, 57.3, 57.9, 58.9 *Negative Peaks:* 34.7, 35.6, 50.4, 50.9

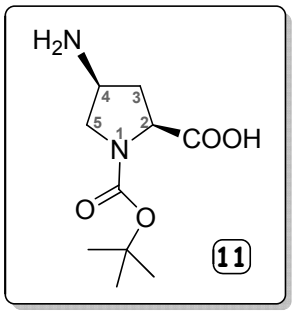
(2*S*,4*S*)N¹-(t-butyloxycabonyl)-4-azidoproline (10)



This acid was hydrolysed from its corresponding methyl ester **9** by aq. NaOH (2N) in water/MeOH. The product was obtained as brown liquid in 98% yield following previously mentioned procedure in case of compound **5**.

Mol. Formula	:	$C_{10}H_{16}N_4O_4$
Mol. Weight	:	256.26
ESI-MS *m/z*	:	257.12 $[M+1]^{+\bullet}$, 279.12 $[M+Na]^{+\bullet}$
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3319, 3019, 2981, 2401, 2106, 1695, 1477, 1404, 1369, 1307, 1260, 1215, 1162, 1132, 1068, 928, 902, 757, 669, 622, 485
^{1}H NMR (200 MHz, MeOH-d_4)	:	**δ_H** (ppm) 1.43 & 1.47 (2s, 9H), 2.1-2.19 (m, 1H), 2.45-2.65 (m, 1H), 3.34-3.42 (dd, 1H, J=12, 3 Hz), 3.64-3.73 (dd, 1H, J=12, 6 Hz), 4.28-4.33 (m, 2H)
^{13}C NMR (50 MHz, MeOH-d_4)	:	**δ_C** (ppm) 28.2, 28.7 (mi), 28.5 (ma), 35.6, 36 (mi), 36.8 (ma), 51.7, 52.6 (mi), 52 (ma), 58.8, 59.2 (mi), 59.7 (ma), 60.8, 60.6 (mi) 60.7 (ma), 81.7 (mi), 81.8 (ma), 155.7 (ma), 155.9 (mi), 175.2(mi), 175.4(ma)
^{13}C-DEPT (50 MHz, MeOH-d_4)	:	**δ_C** (ppm) *Positive Peaks:* 28.2, 28.7(mi), 28.5(ma), 58.8, 59.2(mi), 59.7(ma), 60.8, 60.6(mi) 60.7(ma) *Negative Peaks:* 35.6, 36(mi), 36.8(ma), 51.7, 52.6(mi), 52(ma)

(2*S*,4*S*)N¹-(t-butyloxycabonyl)-4-aminoproline (11)

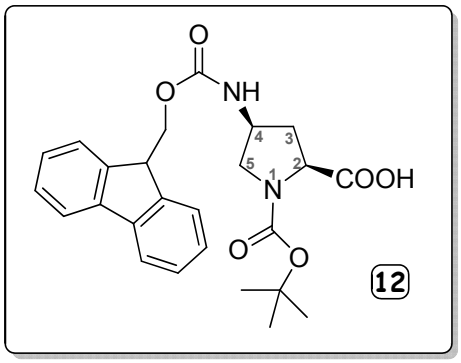


(*2S,4S*)-amino compound **11** (0.87 g, 97.8%) was synthesized from the analogous azide **10** accordingly by Pd-C/H_2 reduction as mentioned earlier for compound **6**.

Mol. Formula	:	$C_{10}H_{18}N_2O_4$
mp	:	Decomposes around 150°C
Mol. Weight	:	230.26
ESI-MS *m/z*	:	231.25 $[M+1]^{+\bullet}$, 253.24 $[M+Na]^{+\bullet}$
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3321, 2983, 2105, 1679, 1585, 1415, 1307, 1260, 1216, 1164, 1133, 1067, 908, 853, 759, 669, 485

^{1}H NMR 400 MHz ($MeOH-d_4 + D_2O$)	:	**δ_H** (ppm) 0.96 & 0.99 (2s, 9H), 1.79-1.86 (m, 1H), 2.35-2.45 (m, 1H), 3.68-3.8 (m, 1H), 3.85-3.92 (m, 1H), 4.3-4.37 (m, 1H), 4.7 (bs, 1H)

(2*S*,4*S*)N^1-(t-butoxycarbonyl)-4-(fluorenylmethyloxycarbonylamino) proline (12)



The compound (2*S*,4*S*)-**12** was prepared from the free amine **11** using the similar procedure as the one described for the compound (2*S*,4R)-**7** as foamy white solid with Yield 1.14 g (59%).

Mol. Formula	:	$C_{25}H_{28}N_2O_6$
mp	:	148.6 – 153.6^{o}C
Mol. Weight	:	452.50
ESI-MS *m/z*	:	475.49 $[M+Na]^{+\bullet}$
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3319, 309, 2978, 2098, 1685, 1522, 1407, 1305, 1261, 1215, 1163, 1136, 1068, 928, 757, 669, 621, 485
^{1}H NMR (400 MHz,DMSO-d_6)	:	**δ_H** (ppm) 1.06 & 1.12 (2s, 9H), 1.57-1.68 (m, 1H), 2.4-2.46 (m, 1H), 3.17-3.23 (m, 1H), 3.91-4.02 (m, 1H), 4.37-4.41 (m, 1H), 4.46-4.53 (m, 1H), 4.61-4.65 (t, 1H), 4.76-4.78 (d, 2H), 8.5-8.55 (t, 2H), 8.61-8.65 (t, 2H), 8.75-8.77 (d, 1H), 8.97-8.99 (d, 2H), 9.2-9.23 (d, 2H), 15.11 (s, 1H)
^{13}C NMR (125 MHz,DMSO-d_6)	:	**δ_C** (ppm) 25(ma), 25.2(mi), 33.1(mi), 34.3(ma), 48.51, 50.9(ma), 51.7(mi), 52.9(ma), 53.7(mi), 61.5(mi), 61.9(ma), 71.9, 88.8, 140.2, 146.5, 148.9, 149.6, 166, 169.8, 181.7, 184.7, 207(mi), 207.6(ma)
^{13}C-DEPT (125 MHz,DMSO-d_6)	:	**δ_C** (ppm) *Positive Peaks:* 25(ma), 25.2(mi), 48.51, 50.9(ma), 51.7(mi), 61.5(mi), 61.9(ma), 88.8, 140.2, 146.5, 148.9, 149.6, 166, 169.8 *Negative Peaks:*33.1(mi), 34.3(ma), 52.9(ma), 53.7(mi), 71.9

1.12.2 Solid phase synthesis

1.12.2a Synthesis of peptides

Majority of peptides were synthesized manually using a sintered (G2) reaction vessel equipped with a Teflon stopcock. All the materials used were of peptide synthesis grade (Novabiochem) and was used without further purification. Analytical grade DMF was purchased from (Merck, India) and was distilled over P_2O_5 under

vacuum at 45 ^{o}C, stored over 4Å molecular sieves for 2 days before using for peptide synthesis. Similarly DCM was distilled over CaH_2 and stored over 4Å molecular sieves. Coupling reactions were carried out using *in situ* active ester method, by HBTU as coupling reagent, HOBt as recemization suppresser and DIPEA as catalyst.

(i) Resin functionalization

Rink Amide (2',4'-Dimethoxyphenyl-Fmoc-aminomethyl)-phenoxy) resin (100 mg, Novabiochem, 100-200 mesh) was taken in afore mentioned sintered reaction vessel vessel rinsed with dry DCM (5 x 5 ml) and filtered under N_2-pressure. The resulting resin was swollen by allowing it to stand in dry DCM for at least 2 hrs. The same procedure was again repeated with dry DMF and the Deprotection of 'Fmoc-group' attached to the resin was done with 20% piperidine in DMF (3 x 3 ml) before proceeding for first amino acid coupling.

In case of commercially available MBHA (4-methyl benzhydrylamine) resin (Novabiochem, 100-200 mesh) the hydrochloride salt neutralization was performed prior to amino acid coupling and the following steps were done for synthesis (100 mg scale).

- The resin was washed and swollen in dry DCM for atleast 1 hr.
- The DCM was drained and 50% solution of DIPEA in DCM was added (3 x 3 ml; 10 min each) followed by thorough washing with DCM.
- Further washing and swelling with dry DMF for 1 hr.
- Immediate coupling of 1st amino acid desired in C-terminus of peptide.

(ii) Synthesis protocol

Following are the various steps performed on the pre-swollen resin in each step of either strategy based on their functionality.

(A) Fmoc–strategy *was used for synthesis on Rink Amide resin for base-labile protecting groups*

- DMF wash (4 x 5 ml)
- 20% piperidine in DMF (2 X 5 ml; 20 min each) for deeprotection of Fmoc group
- DMF, DCM, MeOH wash (3 x 3 ml each)

- Kaiser's/Chloranil test
- DCM, DMF wash (3 x 3 ml each)
- Coupling reaction (3 eq. each of aminoacid, DIPEA, HOBt and HBTU in 10 eq. volume of DMF and NMP mixture)
- Kaiser's/Chloranil test

(B) Boc–strategy *was used for synthesis on MBHA resin for acid-labile protecting groups*

- DCM wash (4 x 5 ml)
- 50% TFA in DCM (2 X 5 ml; 10 min each) for deprotection of Boc group
- 5% DIPEA in DCM (3 X 5 ml; 5 min each) for neutralization of TFA salt
- DCM, MeOH wash (3 x 3 ml each)
- Kaiser's/Chloranil test
- DCM, DMF wash (3 x 3 ml each)
- Coupling reaction (3 eq. each of aminoacid, DIPEA, HOBt and HBTU in 10eq. volume of DMF and NMP mixture)
- Kaiser's/Chloranil test

These cycles ware repeated for every amino acid. The coupling and deprotection reactions were monitored by a combination of Kaiser's (ninhydrin) test and chloronil test. In case of negative test after coupling the re-coupling was performed with same aminoacid followed by capping of the unreacted amino groups using Ac_2O, pyridine & DCM (1:1:1), in case coupling does not go to completion even after re-coupling.

Kaiser's test[138,139]

Kaiser's was used to monitor the *t*-Boc/Fmoc deprotection and coupling reactions of glycine (or basically primary amines) in the solid phase peptide synthesis using three solutions.

Solution A: Ninhydrin (5.0 g) dissolved in ethanol (100 ml)
Solution B: Phenol (80.0 mg) dissolved in ethanol (20 ml)
Solution C: KCN (2 ml, 0.001 M aqueous solution) added to 98 ml pyridine

- Few beads of resin to be tested were taken in a test tube and washed 3 times with ethanol.
- 3-4 drops of each of the three solutions described above were added to it
- The test tube was heated to 120 °C for 4-6 min

The successful deprotection was indicated by blue resin beads while colourless beads indicate the completion of coupling step.

Chloranil test[139,140]

A few beads of resin were taken in a glass test tube (5 ml capacity) and were washed with methanol followed by toluene. To this three drops of saturated chloranil solution in toluene and 200 µl of acetone were added. The mixture was shaken for 2-3 minutes. Blue or green color is observed on the resin beads if free amines are present.

(iii) Automated peptide synthesis[141]

Peptide **P5** [SAP or $(VRLPPP)_3$] was synthesized by FastMoc™-strategy on Applied Biosystems 433A Peptide Synthesizer connected to a windows PC with SynthAssist® software version 3.1 to control the synthesiser. Low loading Rink Amide resin (LV = 0.37 mmol/gm) and peptide synthesis grade commercially available Fmoc-protected amino acids (Novabiochem) were used for this synthesis.

- Resin (0.1mmol) was loaded to the 3 ml capacity reaction vessel and placed on the vortexer.
- Equivalent amount of amino acids filled into sealed cartridges were aligned on the cartridge guideway of machine as per sequence in the C to N terminus order.
- Piperidine, NMP, MeOH, DIPEA-NMP (2 M), HBTU/HOBt-NMP (0.45 M) and Ac_2O(0.5 M)-DIPEA(0.125 M)-NMP solutions were added to their respective bottles.
- The synthesis cycles were programmed on software with each cycle of peptide coupling running for about 50 mins.
- The resin was quantitatively transferred from the reaction vessel after completion of all cycles and the terminal Fmoc-group was manually deprotected and acetylated (capped).

- Afterwards the resin was also manually cleaved to isolate the peptide in excellent yield.

(iv) On-resin guanidylation of peptides

On-resin guanidylation was carried out on peptide **P1** [Ac-Phe-(Pro-Amp-Gly)$_6$-NH_2] and peptide **P3** [Ac-Phe-(amp-Pro-Gly)$_6$-NH_2], that were manually synthesized by Boc-strategy (previously described) using (2*S*,4*R*)N^1-(t-butoxycarbonyl)-4-(9-fluorenylmethyloxycarbonylamino)proline and (2*S*,4*S*)N^1-(t-butoxycarbonyl)-4-(9-fluorenylmethyloxycarbonylamino)proline as monomers respectively on MBHA resin of 0.87 mmol/gm loading value.

- The Fmoc-protection on side chains of the peptides were deprotected using 20% piperidine/DMF treatment.
- *N*,*N*'-bis-Boc-1*H*-pyrazole-1-carboxamidine (5 eq. for each amino group on peptide chain) with equivalent amount of DIPEA in minimum volume of DMF were added to the on-resin peptides and 4 hrs of reaction time was allowed. This step was repeated thrice till Kaiser's test confirmed complete conversion.
- The peptides were later cleaved from their corresponding resin to obtain peptide **P2** [Ac-Phe-(Pro-AmpG-Gly)$_6$-NH_2] and peptide **P4** [Ac-Phe-(ampG-Pro-Gly)$_6$-NH_2] in good yields.

(v) Peptide cleavage from resin

Finally after the last coupling the resin was washed with DMF (5 x 5 ml), DCM (5 x 5 ml), then with toluene (5 x 5 ml) and methanol (5 x 5 ml) then flushed with N_2-gas for 3 mins. The resin along with the sintered flask was dried in vacuum desiccator over KOH and P_2O_5 for 6 hrs.

Two different protocols of cleavage were followed for peptides synthesizes via Fmoc- & Boc-strategy.

(A) Rink amide resin - Peptide cleavage

Dry peptide-resin (100 mg) of was treated with 10 ml of TFA:TIS (95: 5) in DCM solution. The resulting mixture was stirred gently using magnetic stirrer for 1 hr; the mixture was filtered through a sintered funnel and the resin was further washed with TFA:TIS (95: 5) in DCM (3 x 5 ml). The filtrate was collected and evaporated

under reduced pressure. The resulting crude peptide was precipitated with anhydrous diethyl ether and centrifuged (5 times at 5000 rpm). The off-white pellet obtained through centrifugation was dried with N_2-gas and stored at 4 °C.

(B) MBHA resin - Peptide cleavage

The resin MBHA (10 mg) with peptides attached to it was stirred with TFA (200 µl) and deionized H_2O (10 µl) in an ice bath for 10 min. TFMSA (16 µl) was added slowly with vigorous shaking to dissipate the heat generated. The reaction mixture was stirred for 1.5 hrs at room temperature. The resin was removed by filtration under reduced pressure and washed with TFA (3 x 1 ml). The combined filtrate was concentrated, the product was precipitated with cold dry ether and the peptide was isolated by centrifugation (5 times at 5000 rpm) as white powder.

Peptides **P1** and **P3** which had Fmoc-protection in side groups were removed by 20% piperidine/DMF treatment and dried properly before cleavage.

1.12.2b Synthesis of fluorescent labeled peptides

In order to label the peptides **P2**, **P4** and **P5** with fluorescent group, peptides were synthesized as per the earlier mentioned process, except the final N-acetylation step where the N-terminal was being capped with Ac_2O. Texas Red® was attached to peptides as a fluorophore.

Details are described stepwise below for each kind of peptide separately. Precautions were taken to protect the reaction mixture, resin and peptide from light in all the cases.

(i) Peptide P10 (SAP-F)

- The terminal Fmoc-group of peptide **P5** [SAP or Fmoc-NH-$(VRLPPP)_3$] synthesized through automated peptide synthesizer was deprotected with 20% piperidine/DMF.
- The resin (50 mg) was washed with NMP (3 x 3 ml) soaked in NMP (0.5 ml) having Texas Red®-X-succinimidyl ester (1.2 eq.) and DIPEA (1.2 eq.) for 15 hrs.
- The above step was repeated 3 times to ensure higher fluorophore labeling.

- The Texas Red®-labeled peptide **P10** [SAP-F or *TR-X*-NH-(VRLPPP)$_3$] was cleaved from its resin following standard cleavage protocol as described earlier for Rink Amide resins.

(ii) Peptides P8 (RG-F) and P9 (SG-F)

- The Boc-group of non-acetylated peptides **P6** [Boc-NH-(Pro-Amp-Gly)$_6$-NH$_2$] and **P7** [Boc-NH-(amp-Pro-Gly)$_6$-NH$_2$] were deprotected using 50% TFA followed by 5% DIPEA.
- The resin (50 mg) was washed with NMP (3 x 3 ml) soaked in NMP (0.5 ml) having Texas Red®-X-succinimidyl ester (1.2 eq.) and DIPEA (1.2 eq.) for 15 hrs.
- The above step was repeated 3 times to ensure higher fluorophore labeling.
- Resin was DMF washed and *N*,*N*′-bis-Boc-1*H*-pyrazole-1-carboxamidine (5 eq. for each NH$_2$ group on peptide chain) with DIPEA (5 eq.) in DMF (0.5 ml) were added to the on-resin peptides and 4 hrs of reaction time was allowed. This step was also repeated thrice to make sure complete conversion.
- Finally fluorescent peptides **P8** [RG-F or *TR-X*-NH-(Pro-AmpG-Gly)$_6$-NH$_2$] & **P9** [SG-F or *TR-X*-NH-(ampG-Pro-Gly)$_6$-NH$_2$] were isolated in moderate yields after cleavage.

1.12.2c Synthesis of oligonucleotides

DNA oligonucleotides **1-4** (Table **5**) were synthesized on Applied Biosystems ABI 3900 High Throughput DNA Synthesizer at 40 nmol scale using standard β-cyanoethyl phosphoramidite chemistry.[102] The oligomers were synthesized in the 3′ to 5′ direction on polystyrene solid support, followed by ammonia treatment for 8 hr at 60 ^{o}C. The oligonucleotides were desalted by gel filtration and their purity as ascertained by RP-HPLC on a C-18 column was more than 98% and they were used without further purification in the biophysical studies of peptides.

1.12.3 Purification & characterization

1.12.3a Gel filtration chromatography

The crude peptides obtained after ether precipitation were dissolved in de-ionized water (~0.5 ml) and loaded onto a gel filtration G10 Sephadex column with a

void volume of 1 ml. The presence of the peptide was detected by measuring the absorbance at 220 nm. The fractions containing the peptides were pooled together and freeze-dried. The purity of the cleaved crude peptide was determined by analytical RP-HPLC on a C18 column.

The same procedure was repeated with crude oligonucleotides but in G-25 Sephadex columns and the OD being checked at 260 nm.

1.12.3b High performance liquid chromatography (HPLC):

Peptides (**P1** to **P10**) were purified by reverse phase-HPLC on Waters 600 equipped with 2998-Photodiode array detector (PDA). Semi-preparative RP-C18 columns (250 x 10 mm, 10 µm) of Allteck-Alltima make were used for peptides. A gradient of 0-100% B at a flow rate of 3 ml/min was used to elute the peptide and the eluant was monitored at 220 nm unless otherwise mentioned for fluorophore labeled peptide (595 nm incase of Texas Red®-tagged peptides). The peak corresponding to the peptide was collected and the fractions were freezed. Subsequently these peptides were centri-evaporated to dryness on Eppendof-Concentrator plus speed vac. The purity of the final peptides were further analyzed on Merck LiChrospher 100 RP-18 (250 x 4 mm, 5 µm) column by using a gradient flow of 0 to 100% B in 30 min at a flow rate of 1.5 ml/min. The spectra acquisition, analysis and processing was done on Waters Empower-2154 software. The absorbance of the eluant was monitored at its corresponding wavelength and the purity was obtained from the integrator output. The solvent system comprised of MeCN:Water (5:95) with 0.1% TFA for solution A and for solution B MeCN:Water (50:50), 0.1% TFA. The purities of the hence purified peptides were found to be more than 95%.

The purity of **oligonucleotides** (DNA **D1** to **D4**) were ascertained on Varian ProStar HPLC attached to Dynamax absorbance detector model UV-DII using GRACE LiChrospher 100 RP-18 (125 x 4 mm, 5 µm). A gradient elution method 0% A to 100% B in 20 min was used with flow rate 2 ml/min and the eluent was monitored at 260 nm. [A= 5% MeCN in TEAA (0.1 M, pH 7); B=30% MeCN in TEAA (0.1 M, pH 7)]

1.12.3c MALDI-TOF Characterization:

MALDI-TOF mass spectra were obtained on either Voyager-Elite instrument (PerSeptive Biosystems Inc., Farmingham, MA) equipped with delayed extraction or

on Voyager-De-STR (Applied Biosystems) instrument. Sinapinic acid and α-cyano-4-hydroxycinnamic acid (CHCA) both were used as matrix for peptides of which CHCA was found to give satisfactory results. And in case of oligonucleotides, 2,5-dihydroxybenzoic acid (DHB) was well suited for the negative ion MALDI-MS. A saturated matrix solution was prepared with typical dilution solvent (50:50:0.1 Water:MeCN:TFA) and spotted on the metal plate along with the oligomers. The metal plate was loaded to the instrument and the analyte ions are then accelerated by an applied high voltage (15-25 kV) in reflector mode, separated in a field-free flight tube and detected as an electrical signal at the end of the flight tube.

HPLC purified peptides and oligonucleotides were characterized through this method and were observed to give good signal to noise ratio, mostly producing higher molecular ion signals.

1.12.4 Biophysical techniques

1.12.4a Estimation of *p*DNA concentration & purity

The plasmid DNA concentration (1626 ng/µl) and purity were determined using triplicate spectrophotometric readings at 260 nm (for DNA concentration) and 280 nm (protein contamination) on Thermo Fisher Scientific's NanoDrop 2000c spectrophotometer instrument. The ratio of absorbance at 260/280 (~1.8) was considered satisfactory for our further spectroscopic experiments as the ratio of absorbance at 260 and 280 nm is used to assess the purity of DNA and RNA. If the ratio is appreciably lower in any case, it may indicate the presence of protein or other contaminants that absorb strongly at or near 280 nm.

1.12.4b Circular dichroism (CD) spectroscopy

CD spectrometry study was carried out on JASCO J-715 spectropolarimeter using cylindrical, jacketed quartz cell (10 mm path length), which was connected to Julabo-UC-25 water circulator. Spectra were recorded with a spectral resolution of 0.05 nm, band width 1 nm at a scan speed of 100 nm/min and a response time 1 sec. All the spectra were corrected for respective buffer condition and are typically averaged over 6-12 scans.

For **CD thermal denaturation studies** samples were annealed in water bath at 90 °C and slowly cooled to room temperature over a period of 6 hrs prior to spectroscopic analysis. These samples were then incubated at 4 °C for 12 hr

followed by additional half an hour incubation in the instrument at the initial measurement temperature. The temperature was varied in steps of 5 °C and the spectra were recorded at each step. An equilibration period of 5 min was allowed at each temperature. Data processing and curve fitting was performed using MicroCal Origin 8.0 software. Ellipticity at specified wavelength for each temperature was plotted, normalized data was fitted to a sigmoid curve and the T_m (melting temperature) values are derived from the first derivative curve of the fit.

1.12.4c UV-T_m measurements

UV-melting experiments were carried out on Varian Cary 300 UV-spectrophotometer equipped with a Peltier temperature programmer and Julabo water circulator. The sample for T_m measurement was prepared by mixing calculated amount of stock oligonucleotide and peptide solutions together in 1 ml of 10 mM sodium phosphate buffer with 10 mM NaCl (pH 7.2). The samples 1 ml were transferred to rectangular quartz cell (1 cc, 10 mm), sealed with Teflon stopper after degassing with N_2-gas for 15 min, and equilibrated at the starting temperature for at least 15 min. N_2-gas was purged through the cuvette chamber below 15 °C to prevent condensation of moisture on the cuvette walls. The OD at 260 nm was recorded in steps from 10-85 °C with temperature increment of 0.5 °C/min.

Complementary DNA oligonucleotides were mixed together in stoichiometric amounts (1:1) for forming the duplex in 0.01 M sodium phosphate buffer, pH 7.2 to achieve a final strand concentration of either 1 or 2 μM of each strand. Both for the AT rich and CG-rich DNAs *parallel* complexes were constituted by annealing the samples. They were heated at 90 °C for 5 mins followed by slow cooling to room temperature over a time span of at least 8 hrs, allowed to remain at room temperature for an hour and then refrigerated at 4 °C overnight prior to running the melting experiments. Each melting experiment was repeated thrice to confirm the consistency. The normalized absorbance at 260 nm was plotted as a function of the temperature. Weak background absorbance from the buffer was directly subtracted from the measurements. The T_m was determined from the first derivative plots of normalized absorbance with respect to temperature and is accurate to ±0.5 °C. The data were processed using MicroCal Origin 8.0 version software. The concentration of DNA oligonucleotides were calculated with the help of extinction coefficients, A = 15400, T = 8700, C = 7400 and G = 11500 litre/(mol.cm).[142] Similarly the UV-T_m

values were determined for peptide-*ds*DNA complexes in various (1:1, 1:50, 1:100) molar ratios.

1.12.4d Fluorescence spectroscopy

Experiments were performed on a Horiba Jobin Yvon Flurolog 3 (Fl3-22) spectrometer equipped with a jacketed cell holder, which was connected to a temperature controller. The fluorescence results were articulated as percentage of the maximum fluorescence signal when EtBr was bound to the DNA w/o the presence of peptides with due rectification for buffer.

Ethidium bromide-DNA assay

The measurements were performed in a rectangular quartz cell (1cc, 10 mm) at 15 °C with λ_{excit} 490 nm; λ_{emiss} 610 nm and an excitation slit width of 2 nm and an emission slit width of 4 nm. Fluorescence spectra were obtained at a scan speed of 50 nm min^{-1} and are an average of 5 scans.

(i) With oligonucleotides

For **exclusion assay** duplex DNA (both AT and CG rich parallel complexes, 1 μM of each strand) was prepared in with NaCl buffer (850 μl, 10 mM). Two sets with varying amounts of peptides (1 eq. and 100 eq.) were added to same buffer with DNA, mixed in a microcentrifuge, annealed and incubated (as described earlier for UV-T_m experiments). Immediately prior to each analysis, aliquots of EtBr solution (5 μl, 100 μM) was added, the sample was mixed on a bench top vortexer and the fluorescence was measured. The solution of EtBr was titrated till a saturation value was observed in fluorescence intensity.

In case of **displacement assay**, duplex DNA (1 μM, 850 μl in 10 mM NaCl buffer) was added with EtBr (9 μl for CG- & 6 μl for AT-rich oligos, 250 μM)* to achieve maximum fluorescence intensity. Thereafter the above *ds*DNA-EtBr solution was titrated with different peptide solutions (6.37 mM) individually in aliquot of 2 μl and change in fluorescence intensity was measured.

* The amount of EtBr (250 μM) required for maximum fluorescence intensity was determined through its titration with *ds*DNA (1 μM, 850 μl in 10 mM NaCl buffer) for both CG- & AT-rich sequences individually.

(ii) With plasmid DNA

Similarly with plasmid DNA (pRmHa3-GFP) the EtBr assay was carried out through both displacement and exclusion method.

Akin to above described **displacement assay**, pRmHa3-GFP (3 µg) was dissolved in EDTA (1 mM, 800 µl) was titrated with EtBr (250 µM) solution to know the minimum amount of EtBr required to observe the maximum fluorescence intensity. Later this *p*DNA-EtBr mixture was titrated with various peptides in increasing concentrations and the change in fluorescence was noted. While in case of **exclusion assay**, 50 eq. of each peptide were mixed with pRmHa3-GFP (3 µg) dissolved in EDTA (1 mM, 800 µl). The peptide-*p*DNA mixture was then titrated with EtBr (250 µM) solution in aliquots of 5 µl each time.

1.12.4e Electrophoretic mobility shift assay

(i) Agarose gel electrophoresis

The agarose gel electrophoresis was used for both assessment of plasmid DNA purity and its retardation when complexed with various peptides.

A mixture of pRmHa3-GFP plasmid DNA (1 µg) and cataionic peptides (**P1**, **P2**, **P3**, **P4** & **P5**; 50 µg each) EDTA buffer (1 mM, 15 µl) diluted to a final concentration of 0.1 µg DNA/µl. This mixture was incubated at 0 ^{o}C over 3 different time intervals of 3 min, 15 min and 30 min each. Samples (10 µl) of the above mixture along with dye (2 µl, 40% w/v sucrose & 0.25% Bromophenol blue in deionised H_2O) were loaded on a agarose gel (1% with 50 ng/ml EtBr for visualization). A DNA molecular weight ladder (50 KDa) was also loaded alongside the samples. The electrophoresis was carried out at 90 V for 120 min in TAE buffer. The gel was periodically checked on a UV-transilluminator for determining sufficient run time. The bands were observed with a transilluminator and images were captured using a Syngene G:BOX digital camera.

*p*DNA (1 µg) with increasing amounts of cationic peptides at ratios of 1:1, 1:2, 1:5, 1:10, 1: 25, 1:50 (w/w) were electrophoresed on 1% agarose gel using conditions as mentioned above, to determine the minimum concentration of peptide for complexation with DNA.

(ii) Native polyacrylamide gel electrophoresis (PAGE)

Acrylamide:bisacrylamide (29:1, 6.6 ml), deionised H_2O (1.27 ml) and 5X-TBE (2.0 ml) were degassed by under vacuum (30 min). Ammonium persulfate solution (10 %, 70 µl) and TEMED (4.6 µl) were later added to it as polymerization activators. This solution was used to cast the 20% nondenaturing/native polyacrylamide gel.

Peptides were individually mixed with oligonucleotides (*ds*DNA) in 1:1 ratio (each strand 0.4 mM) in water. The samples were lyophilized to dryness and re-suspended in sodium phosphate buffer (10 mM, 4 µl, pH 7.4) containing NaCl (10 mM). The samples were annealed by heating to 85 °C for 5 min followed by slow cooling to RT and refrigeration at 4 °C overnight. To this, 40 % sucrose in 5X-TBE buffer (4 µl, pH 8.0) was added and sample was loaded on the gel. Bromophenol blue (BPB) was used as the tracer dye and separately loaded in an adjacent well. Gel electrophoresis was performed on 20 % nondenaturing polyacrylamide gel (acrylamide:*N*,*N*'-methylenebisacrylamide, 29:1) with 1X-TBE as tank buffer at constant power supply of 200 V and 10 mA, until the BPB migrated to three-fourth of the gel length. During electrophoresis, the temperature was maintained at 10 °C. The spots were visualized through UV-transilluminator by illumination the gel placed on a fluorescent silica gel plate, GF_{254} using UV-light or UV-transilluminator after treating the gel with 0.1% EtBr solution for 5 mins.

1.12.4f Isothermal titration calorimetry

Isothermal titration calorimetric studies were done on MicroCal iTC_{200} instrument. Experiments were carried out at 20 °C. The sample cell was filled with plasmid DNA (2 µg) in 1 mM EDTA (300 µl). The peptides (0.37 mM, 50 µl) diluted in the same buffer were loaded into the syringe. The injection volumes were 1 µl each, injection time 6 s, a 150 s delay between each injection and stirring speed 1000 rpm.

The integrated peaks (pulses) of the heat production/absorption upon each injection were plotted against the molar ratio. The interaction of buffer in the process was corrected by substracting the values from obtained from titration of peptide on to blank buffer using the same set of conditions. With the built-in software, MicroCal Origin, the binding isotherms were fitted to a two-state/one-state binding model, giving numerical values of both the enthalpy of binding (ΔH) and the binding constant (K).

1.12.5 Biological Protocols

1.12.5a Plasmid DNA isolation & purification

E.coli cells transformed with pRmHa3-GFP plasmid DNA was allowed to grow overnight in Luria-Bertani (LB) media with 50 µg/ml of ampicillin as selection agent. DNA was isolated using standard Qiagen Maxiprep protocol. Purified *p*DNA was re-suspended in deionised water with a concentration around 1.5 µg/µl. The purity of isolated plasmid was determined by measurement of the A260:A280 ratio on Thermo Fisher Scientific's NanoDrop 2000c spectrophotometer instrument.

1.12.5b Restriction enzyme digestion

Three separate batches of restriction enzyme digest of plasmid DNA (pRmHa3-GFP) w/o cationic peptides **P2** and **P4** were set up. The details are as per the following table.

Table 12: Details of restriction enzyme digestion protocol

Materials	**Batch-I (control)**	**Batch-II**	**Batch-III**
Sterile water	16.3µl	14.9 µl	14.9 µl
Enzyme Buffer(1X)	2 µl	2 µl	2 µl
BSA(1X)	0.2 µl	0.2 µl	0.2 µl
*p*DNA (0.5 µg/µl)	0.5 µg/ µl	0.5 µg/ µl	0.5 µg/ µl
Peptide (25X)	---	1.31 µl (**P2**/RG)	1.36 µl (**P4**/SG)
Enzyme (BamH1)	1 µl	1 µl	1 µl
TOTAL VOLUME	**20 µl**	**20 µl**	**20 µl**

The three samples were prepared in eppendorfs 200 µl centrifuge tubes vortexed and centrifuged for 2 seconds. The reaction mixtures were then incubated at 37 ^{o}C for 3 hrs, after which they were transferred to 65 ^{o}C for 10 min for inactivating the enzyme. The digested products along with the uncut plasmid were analyzed on 1% agarose gel using TAE buffer.

1.12.5c Cell culture & transfection

Drosophila melanogaster Schneider 2 (S2) cells were grown at 25 ^{o}C under normal atmosphere (100% air without CO_2). Schneider's cells medium (Sigma) was supplemented with 10% heat-inactivated FBS (Invitrogen).

Cells were maintained in 25-cm^2 T-flasks (Corning Inc). They were allowed to grow to a density of 1 to 5 × 10^6 cells/ml, with the culture being split into fresh medium at a 1:4 or 1:5 dilutions every 3 days.

The following protocol was undertaken day wise for optimised transfection using QIAGEN's transfection kit.

DAY 1: Plating

- Under sterile conditions the S2 cells were resuspended and counted in an aliquot of 10 µl inside Neubauer chamber (haemocytometer).
- Cells were seeded at a concentration of 2 × 10^4 cells/ml of fresh Schneider's cells medium supplemented with 10% FBS, 350 µl cells/well was seeded in a 24-well culture cluster.
- Plates were sealed with Parafilm and were grown under above mentioned conditions for 24 hrs prior to transfection.

DAY 2: Transfection

- The plate was checked under microscope to observe around 80% confluence of cells.
- In a 500 µl centrifuge tube 0.2 µg plasmid DNA (pRmHa3-GFP) was diluted with the DNA-condensation buffer (Buffer EC) to a total volume of 60 µl. Enhancer (1.6 µl) was added to it and mixed by vortexing for 1 s. This mixture was incubated at room temperature for 2–5 min then spun down for a few seconds to remove drops from the top of the tube.
- 5 µl Effectene® (transfection reagent – QIAGEN) was added to the DNA-Enhancer mixture. DNA-effectene mixture was vortexed for 10 s followed by incubation at RT for 5-10 mins to allow transfection complex formation.
- Mean while the growth medium was gently aspirated from the each well of the plate and 350 µl of fresh growth medium was added to the cells.

- 350 µl of growth medium was also added to the tube containing the transfection complexes. Mixed well and immediately added the transfection complexes drop-wise onto the cells in the 24-well cluster plate. The plate was gently swirled to ensure uniform distribution of the transfection complexes.
- The plate was sealed with Parafilm and incubated for 18 hrs.

DAY 3: Post-transfection I

- Growth medium from the wells was replaced smoothly.
- 2.5 µl of 100 mM $CuSO_4$ solution was added to all the wells swirled carefully, again incubated at 25 ^{o}C for 20 hrs with normal air flow but covered with aluminium foil to protect it from light.

DAY 4: Post-transfection II

- The cluster plate was analysed under Zeiss- Microscope with camera at 40X optical zoom. GFP expression was confirmed by illuminating at 480 nm monochrome light using standard filter.
- Green fluorescent cells in addition to non-fluorescent cells were counted manually by focusing in different areas of the well. Average transfection efficiency was thus calculated from the above numbers.

DAY 4: Cell fixation & mounting

- The transfected cells were washed twice with 1X-PBS buffer by placing on rocker (5 min each time) and later spinning them at a maximum speed of 1000 rpm.
- Thin corning glass coverslips treated with poly-L-lysine were incubated at RT till dry.
- 100 µl cell suspensions in 1X-PBS were gently spread over the poly-L-lysine coating and kept aside in dark undisturbed for 15 min.
- The excess cell-suspension was gently aspirated out and the adhered cells were treated with 50 µl of fixing solution (3% paraformaldehyde in PBS; MeOH free) for around 15 to 20 min.
- Thereafter the coverslips were washed with excess of 1X-PBS thrice to remove excess fixing solution and loosely bound cells.

- The rehydrated cover slips were then mounted to glass-slides with 20 µl of mounting media (DAPI-Fluoromount-G^{TM}). It was gently sponged with a tissue to remove excess mounting medium before sealing and allowed to air dry for 5 min before examination.
- The slides were incubated at 4 ^{o}C in dark to preserve fluorescence till visualisation under microscope.

(i) Transfection with peptides

Analogous to above process, S2 cell transfection with pRmHa3-GFP was carried out in the presence of the synthesized peptides (**P1**, **P2**, **P3**, **P4** and **P5**) individually. Apart from control transfection with transfection kit (QIAGEN), three other sets of experiments were designed for each peptide *viz*

- S2 cells + pRmHa3GFP + Enhancer + Effectene® (Control Experiment)
- S2 cells + pRmHa3GFP + Enhancer + Peptide
- S2 cells + pRmHa3GFP + Peptide + Effectene®
- S2 cells + pRmHa3GFP + Peptide

Effectene® and Enhancer are the reagents available in the transfection kit procured from QIAGEN. These reagents were replaced with 25 eq. (**P2** & **P4**) and 50 eq. (**P1**, **P3** & **P5**) w/w peptides dissolved in EC buffer (QIAGEN) were added to S2 cell culture in 24-well plate as per afore mentioned transfection protocol.

The transfection efficiencies in each case were determined by manual counting of both fluorescent and non-fluorescent cells with the help of Fluorescence light microscopy. The data was compared to know the role and efficacy of peptides as substitute for transfection reagents.

(ii) Cell permeation & transfection with fluorescent peptides

Texas Red® labelled fluorescent peptides (**P8**, **P9** & **P10**) were tested for their cell permeation ability w/o the plasmid DNA (pRmHa3-GFP).

- In a 12-well culture plate, S2 cells (800 µl cells/well) were seeded at a concentration of 2 × 10^4 cells/ml in fresh Schneider's cells medium supplemented with 10% FBS, and grown at 25 ^{o}C for 24 hrs under normal atmosphere.

Table 13: Details of fluorescent peptide permeation and transfection protocol

Materials	Well-1	Well-2	Well-3	Well-4
pRmHa3-GFP	---	0.3 µg	0.3 µg	0.3 µg
Fluorescent Peptide*	5 µl	5 µl	---	5 µl
Transfection reagents#	x	x	√	√
Buffer EC	70 µl	70 µl	70 µl	70 µl
Media to Complex	400 µl	400 µl	400 µl	400 µl

* All fluorescent peptides had concentration around 2.0 µM
# QIAGEN Transfection Reagents: Enhancer (2.4 µl) and Effectene® (6 µl)

- In 500 µl centrifuge tubes the above mentioned materials were mixed in specific amounts with Buffer EC (QIAGEN) to a total volume of 78 µl disjointedly. These were then added with 400 µl of S2-cell media (Sigma) and transferred to respective wells with fresh media (800 µl). This setup was repeated for each fluorescent peptide.
- After 24 hrs incubation at 25 ^{o}C, all the cells were given $CuSO_4$ (5 µl, 100mM) induction irrespective of presence of *p*DNA to maintain homogeneity and further incubated for another 24 hrs.
- The cells were then examined for fluorescence. A portion of these cells were fixed on cover slips and studied in detail by confocal microscopy.

1.12.5d Fluorescence light microscopy

Fluorescence was viewed by wide field epifluorescence microscopy on a Imager.Z1 inverted microscope (Objective: 40X/0.75; Carl Zeiss, Germany) with filter sets designed for GFP (480 nm) or Texas Red® (590 nm). Digital images were taken using a Cooled digital CCD camera (AxioCam HRm/HRc) and saved as 8-bit raw TIFF images on a hp-workstation with monitor. For quantitative comparisons, images were acquired with the same exposure settings and were rescaled according to a standard offset and gain using AxioVision.

1.12.5e Confocal microscopy

Fixed cells were scanned using the 488 nm laser line and light emitted between 500 and 600 nm was collected for confocal detection of GFP expression in transfected cells. And for detection of Texas Red labelled peptides in the S2 cells, excitation wavelength of 590 nm was used, whose fluorescence was followed in the range of 610 to 650 nm.

Images were collected on a LSM 710 Observer.Z1 laser scanning confocal microscope equipped with a HAL 100 objective (Carl Zeiss, Germany). The same settings for image acquisition and processing (Z-stacking & maximum intensity projection) have been applied for all samples to allow comparison of the fluorescence intensities among different samples.

1.13 References

1) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.

2) Arnott, S.; Chandrashekaran, R.; Hall, I. H.; Puigjaner, L. C.; Walker, J. K.; Wang, M. *Cold Spring Harbor Symp. Quant. Biol.* **1982**, *47*, 53-65.

3) Saenger, W. *Principles of Nucleic Acid Structure*, Sprigner Verlag: New York, **1984**.

4) (a) Zimmerman, S. B. *Annu. Rev. Biochem.* **1982**, *51*, 395-427. (b) Rich, A.; Nordheim, A.; Wang, A. H. -J. *Annu. Rev. Biochem.* **1984**, *53*, 791-846. (c) Schleif, R. *Science* **1988**, *241*, 1182-1187.

5) (a) Tabor, C. W.; Tabor, H. *Annu. Rev. Biochem*. **1984**, *53*, 749-790. (b) Marton, L. J.; Pegg, A. E. *Annu. Rev. Pharmacol. Toxicol.* **1995**, *33*, 55-91.

6) Westhof, E.; Beveridge, D.L. *Water Sci. Rev.*, **1990**, *5*, 24-136.

7) Fisher, P. A.; Korn, D. *J. Biol. Chem.* **1979**, *254*, 11033-11039.

8) Manning, G. S. Q. *Rev. Biophys.* **1978**, *11*, 179-246.

9) Dickerson, R. E.; Drew, H. R. *Nucleic Acids Res.* **1981**, *12*, 2381.

10) Jain, S.; Zon, G.; Sundaralingam, M. *Biochemistry* **1989**, *28*, 2360-2364.

11) Gessner, R. V.; Frederick, C. A.; Quigley, G. J.; Rich, A.; Wang. A. H-J. *J. Biol. Chem.* **1989**, *264*, 7921-7935.

12) Waring, M. J; Wakelin, L. P. G. *Comprehensive Med. Chem.* Ed. Sammes, P. G. Pergamon Press, Oxford, **1990**.

13) Neidle, S.; Waring, M. J. Eds. *Molecular Aspects of Anti-Cancer Drug Action*, Macmillan Press: London, **1977**.

14) Gale, E. F.; Cundliffe, E.; Reynolds, P. E.; Richmond, M. H.: Waring M. J. in *The Molecular Basis of Antibiotic Action* 2nd edn. Wiley, London, **1981**.

15) Reddy, B. S. P.; Sondhi, S. M.; Lown, J. W. *Pharmacol. & Therapeutics* **1999**, *84*, 1-111.

16) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R.E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380.

17) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723-5727.

18) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408-7416.

19) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8789.

20) (a) Dervan, P. B.; Burli, R. *Current Opin. Chem. Biol.* **1999**, *3*, 688-693. (b) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569-578.

21) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstranger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586-7590.

22) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature Struct. Biol.* **1998**, *5*, 104-109.

23) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468-471.

24) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205.

25) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfield, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12890-12895.

26) Saenger, W.; Heinemann, U. *Protein-Nucleic Acid Interactions* Macmillan Press, London, **1989.**

27) (a) Harrison, S. C. *Nature* **1991**, *252*, 809-817. (b) Helene, C.; Lancelot, G. *Prog. Biophys. Mol. Biol.* **1982**, *39*, 1-68.

28) (a) Steitz, T. A. *Quart. Rev. Biophys.* **1990,** *23*, 205-280. (b) Ollis, D. L.; White, S. W. *Chem. Rev.* **1987**, *87*, 981-985.

29) Brennan, R. G.; Roderick, S. L.; Takeda, Y.; Matthews, B. W. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8165-8169.

30) Sauer, R. T.; Jordan, S. R.; Pabo, C. O. *Adv. Protein Chem.* **1990**, *40*, 1-61.

31) Luisi, B. F.; Sigler, P. B. *Biochim. Biophys. Acta* **1990**, *1048*, 113-126.

32) (a) Creighton, T. E. *Proteins* 2nd ed. W.H. Freeman & Co, New York, **1993**, pp. 355-360. (b) Brennan, R. G.; Matthews, B. W. *J. Biol. Chem.* **1989**, *264*, 1903-1906. (c) Struhl, K. *Trends Biochem. Sci.* **1989**, *14*, 137-140.

33) (a) Berg, J. M. *Ann. Rev. Biophys. Biophys. chem.* **1990**, *19*, 405-421. (b) Pavletich, N. K.; Pabo, C. O. *Science* **1991**, *252*, 809-817.

34) Pavletich, N.; Pabo, C. *Science* **1991**,*252,* 810–817.

35) (a) Dickerson, R. E.; Drew, H. R. J. Mol. Biol. **1981**. *149*, 781-786. (b) McClarine, J. A.; Frederick, C. A.; Wang, B. C.; Green, P.; Boyer, H. W.; Grable, J.; Rosenberg, J. M. *Science* **1986**, *30*, 9792-9798.

36) Edelstein, M.L. ; Mohammad, R.; Abedi, M. R.; Wixon, J. *J. Gene Med*. **2007**, *9*, 833–842.

37) Kim, D. H.; Rossi, J.J. *Nat Rev Genet.* **2007**, *8*,173-184.

38) (a) Glover, D. J.; Lipps, H. J.; Jans, D. A. *Nat Rev Genet*. **2005**, 6, 299-310. (b) Rosi, N. L.; Mirkin, C. A. *Chem Rev.* **2005**, *105*, 1547-62.

39) Wolff, J. A.; Malone, R. W.; Williams, P. *Science* **1990**, *247*, 1465-1468.

40) (a) Liu, F.; Song, Y.; Liu, D. *Gene Ther*. **1999**, *6*, 1258-1266. (b) Zhang. G, Budker, V. Wolff, J. A. *Hum Gene Ther*. **1999**, *10*, 1735-1737.

41) (a)Yang, N. S.; Burkholder, J.; Roberts, B.; Martinell, B.; McCabe, D. *Proc Natl Acad. Sci. USA* **1990**, *87*, 9568-9572.(b) Yang, N.S.; Sun, W. H. *Nat Med*. **1995**, *1*, 481-483.

42) Neumann, E.; Schaefer-Ridder, M.; Wang, Y. Hofschneider, P.H. *EMBO J***. 1982**, *1*, 841-845.

43) Heller, L. C.; Ugen, K.; Heller, R. *Expert Opin Drug Deliv*. **2005**, *2*, 255-268.

44) Brisken, L. A.; Francis, A. F.; Cumberland, S. E.; Crossman, D. C.; Newman, C. M. *Gene Ther*. **2000***, 7*, 2023-2027.

45) (a) Neu, M.; Fischer, D.; Kissel, T. *J Gene Med*. **2005**, *7,* 992-1009. (b) Liu, D.; Ren, T. Gao, X. *Curr. Med. Chem.* **2003**, *10,* 1307-1315. (c) Huang, L.; Hung, M.C.; Wagner, E. *Nonviral Vectors for Gene Therapy* .San Diego, California, Academic Press, **1999**. (d) Mahato, R. I.; Kim, S. W. *Pharmaceutical Perspectives of Nucleic Acid- Based Therapeutics.* London, United Kingdom, Taylor & Francis, **2002**.

46) Saul, J. M.; Linnes, M. P.; Ratner, B. D.; Giachelli, C. M.; Pun, S. H. *Biomaterials* **2007**, *28*, 4705-4716.

47) Fabre, J. W.; Collins, L. *Curr. Gene Ther*. **2006**, *6*, 459-480.

48) McKenzie, D. L.;Collard, W. T.; Rice, K. G. *J. Peptide Res*. **1999**, *54*, 311-318.

49) Plank, C.; Tang, M. X.; Wolfe, A. R. Hum. Gene Ther. **1999**, *10*, 319-332.

50) Schwarze, S.R.; Dowdy, S. F., *Trends Pharmacol. Sci.* **2000**, *21*, 45-48.

51) Trehin, R.; Merkle, H.P.; *Eur. J. Pharmaceut. Biopharmaceut.* **2004,** *58*, 209-223.

52) Futaki, S. *Biopolymers* **2006**, *84*, 241-249.

53) Wender, P. A.; Wender, P. A.; Mitchell, D. J; Pattabiraman, K.; Pelkey, E. T.; Steinman, L.; Rothbard, J. B. *Proc. Natl. Acad. Sci*. **2000**, *97*, 13003-13008.

54) (a) Gait, M.J. *Cell. Mol. Life Sci.* **2003**, *60*, 844-853, (b) Koppelhus, U. Nielsen, P. E. *Adv. Drug Deliv*. *Rev*. **2003,** *55*, 267-280, (c) Eguchi, A.; Akuta, T. ; Okuyama, H.; Senda, T.; Yokoi, H. ; Inokuchi, H.; Fujita, S.; Hayakawa, T.; Takeda, K.; Hasegawa, M.; Nakanishi, M. *J. Biol. Chem.* **2001**, *276*, 26204-26210.

55) Wagner, E. *Adv. Drug Deliv*. *Rev.* **1999,** *38*, 279-289.

56) Midoux, P.; Kichler, A.; Boutin, V.; Jean-Claude, M.; Monsigny, M. *Bioconjugate Chem*. **1998**, *9*, 260-267.

57) (a) Kichler, A. Leborgne, C.; Marz, J.; Danos, O.; Bechinger, B. *Proc. Natl. Acad. Sci*. **2003** *100*, 1564-1568 (b) Mason, A. J.; Leborgne, C.; Moulay, G. Martinez, A. Danos, O.; Bechinger, B.; Kichler, A. *J. Control. Rel.* **2007,** *118*, 95-104.

58) Escriou, V.; Carrière, M.; Scherman, D.; Wils, P. *Adv. Drug Deliv. Rev*. **2003**, *55*, 295-306.

59) Ludtke, J. J.; Zhang, G.; Sebestyen, M. G.; Wolff, J. A. *J. Cell Sci.* **1999**, *112*, 2033–2041.

60) Bloomfield, V. A.; Crothers, D.; Tinoco, I. *Physical Chemistry of the Nucleic Acids.* Harper and Row: New York, **1974**.

61) Freeman, W. H. *The behavior of biological macromolecules* (Cantor and Schimmel, San Francisco, **1980**, pp. 624.

62) Martin, S. R.; Schilstra, M. J. Biophysical Tools for Biologists, Volume One: In Vitro Techniques, *Methods in cell biology*, Elsevier, **2008,** pp. 263-93.

63) (a) Eriksson, M.; Nordén, B. *Meth. Enzymol.* **2001**,*340*, 68–99. (b) Nordén, B.; Kurucsev, T. *J. Mol. Recog.* **1994**, *7*, 141-156.

64) Callis, P. R. *Ann. Rev. Phys. Chem*. **1983**, *34*, 329-357.

65) Ho, P. S.; Zhou, G.; Clark, L. B. *Biopolymers* **1990**, *30*, 151-163.

66) Brand, L.; Johnson, M. L. *Meth. Enzymol*. **1997**, *278*, 1-628.

67) Cain, B. F.; Baguley, B. C.; Denny, W. A. *J. Med. Chem*. **1978**, *21*, 658-668.

68) Gershan, H.; Chiralando, R.; Guttman, S. B.; Minsky, A. *Biochemistry* **1993**, *32*, 7143.

69) (a) Hendrickson, W. *BioTechniques* **1985**, *3*, 346-354. (b) Revzin, A. *BioTechniques* **1989**, *7*, 346-354.

70) Sambrook, J.; Russel, D. W. *Molecular Cloning: A Laboratory Manual* 3rd Ed. Cold Spring Harbor Laboratory Press.: Cold Spring Harbor, New York, **2001**.

71) (a) Liebler, E. K.; Diederichsen, U.; *Org. Lett.* **2004**, *6*, 2893–2896. (b) Lerman, L. S.; Frisch, H. L. *Biopolymers* **1982**, *21*, 995–997. (c) Leng, M. *Biophys. Chem.* **1990**, *35*, 155-163.

72) (a) Haq, I.; Jenkins, T. C.; Chowdhry, B. Z; Ren, J; Chaires, J. B. *Meth. Enzymol.* **2000**, *323*, 373-405. (b) Haq, I.; Jenkins, T. C.; Chowdhry, B. Z. *Meth. Enzymol.***2001**, *340*, 109-149. (c) Jelesarov, I.; Bosshard, H.R. *J. Mol. Recog.***1999**, *12*, 3-18. (d) Cooper, A. *Curr. Opin. Chem. Biol.***1999,** *3*,557-563. (e) Wiseman, T.; Williston, S.; Brandts, J.F.; Lin, L. N. *Anal. Biochem.***1989**, *179*, 131-137.

73) Crain, P.F, ; McCloskey, J. A. *Curr. Opin. Biotechnol.* **1998**, *9*, 25-34.

74) Fu, D. J.; Tang, K.; Braun, A.; Reuter, D.; Iverson, B. L.; Darnhofer-Demar, B.; Little, D.P.; O'Donnell, M. J.; Cantor, C. R.; Koster, H. *Nat. Biotech.*, **1998**, *16*, 381-384

75) Ross, P.; Hall, L.; Smirnov, I.; Haff, L. *Nat. Biotech.* **1998**, *16*, 1347-1351.

76) Braun, A.; Little, D.P.; Koster, H. *Clin. Chem.*, **1997**, *43*, 1151-1158.

77) Braun, A.; Little, D.P.; Reuter, D.; Muller-Mysok, B.; Koster, H. *Genomics*, **1997**, *46*, 18-23.

78) Ross, P.; Belgrader, P. *Anal. Chem.*, **1997**, *69*, 3966-3972.

79) Little, D. J.; Cornish, T. J.; O'Donnell, M. J.; Braun, A.; Cotter, R. J.; Koster, H. *Anal. Chem.*, **1997**, 69, 4540-4546.

80) Promega's *Protocols and Applications Guide-Transfection,* http://www.promega.com/paguide/chap12.htm (accessed January 6, 2011)

81) Sample, V.; Newman, R. H.; Zhang, J. *Chem. Soc. Rev*., **2009**, *38*, 2852–2864.

82) (a) Felgner, P.L.; Gadek, T. R.; Holm, M.; Roman, R.; Chan, H. W.; Wenz, M.; Northrop, J. P. ; Ringold, G. M.; Danielsen, M. *Proc. Natl. Acad. Sci. USA* **1987,** *84*, 7413-7417. (b) Capaccioli, S.; Di Pasquale, G.; Mini, E.; Mazzei, T.; Quattrone, A. *Biochem. Biophys. Res. Commun.* **1993,** *197*, 818-825. (c) Felgner, J.; Bennett, F.; Felgner, P. L. *Methods* **1993**, *5*, 67-75. (d) Haensler, J.; Szoka, F.C. *Bioconj. Chem.* **1993,** *4*, 372–379. (e) Lee, J.T.; Jaenisch, R. *Nucleic Acids Res.***1996**, *24*, 5054–5055. (f) Lamb, B.T.; Gearhart, J.D. *Curr. Opin. Genet. Dev.* **1995,** *5*, 342–348.

83) Smith, B. A.; Daniels, D. S.; Coplin, A. E.; Jordan, G. E.; McGregor, L. M.; Schepartz, A. *J. Am. Chem. Soc.* **2008**, *130*, 2948.

84) (a) Sano, A.; Maeda M.; Nagahara, S.; Ochiya, T.; Honma, K.; Itoh, H.; Miyata, T.; Fujioka, K. *Advanced Drug Delivery Reviews* **2003**, *55,* 1651-1677 (b) Takeshita, F.;

Minakuchi, Y.; Nagahara, S.; Honma, K.; Sasaki, H.; Hirai, K.; Teratani, T.; Namatame, N.; Yamamoto, Y.; Hanai, K.; Kato, T.; Sano, A.; Ochiya, T. *Proc. Nat. Acad. Sci. USA* , **2005**, *102*, 12177–12182.

85) (a) Mrevlishvili, G. M.; Svintradze, D. V. *International Journal of Biological Macromolecules* **2005**, *36*, 324-326, (b) Creager, M. S. ; Jenkins, J. E. ; Thagard-Yeaman, L. A.; Brooks, A. E. ; Jones, J. A.; Lewis, R. V.; Holland, G. P. ; Yarger, J. L. *Biomacromolecules* **2008,** *9*, 21-28.

86) (a) Chmielewski, J.; Geisler, I.; *Chem Biol Drug Des* **2009**; *73,* 39-45. (b) Fillon, Y. A.; Anderson, J. P.; Chmielewski, J. *J. Am. Chem. Soc.* **2005**, *127*, 11798-11803.

87) (a) Pozo-Rodríguez, A. D.; Pujals**,** S.; Delgado**,** D.; Solinís**,** M.A.; Gascón**,** A.R.; Giralt**,** E.; Pedraz**,** J. L. *J. Contr. Rel.* **2009**, *133*, 52–59 (b) Nakase, I.; Tadokoro, A.; T. Takeuchi, T.; Jones, A.T.; Futaki, S. *Biochem. Soc. Trans*. **2007**, *35*, 784-787.

88) (a) Yang-Gyun, K.; Hyun-Ju, P.; Kyeong, K. K.; Alexander, R.; Ky, L. *Nucleic Acids Research*, **2006**, *34*, 4937-4942, (b) Haley, J.; Kabiru, P.; Geng, Y.; *Mol. BioSyst.*, **2010**, *6*, 249–255.

89) Mitchell, D. J.; Kim, D. T.; Steinman, L.; Fathman, C. G.; Rothbard, J. B. *J. Pept. Res.* **2000**, *56*, 318-325.

90) (a) Pujals, S.; Giralt, E. *Advanced drug delivery reviews* **2008**, *60*, 473-84. (b) Fernandez-Carneado, J.; Kogan, M. J.; Castel, S.; Giralt, E. *Angew. Chem., Int. Ed.* **2004**, *43*, 1811-1814.

91) (a) Bodansky, M.; Bodansky, A. *The Practice of Peptide Synthesis*, Springer-Verlog, Berlin, **1984**. (b) Stewart, J. M.; Young, J. D. *Solid Phase Peptide Synthesis*, W. H. Freeman & Co, New York, **1969**.

92) Rich, D. H.; Singh, J. *The carbodiimide method*. in *The Peptides* Vol 1, Gross, E.; Meienhofer, J. eds., Acad. Press New York **1979**, pp.241-314.

93) (a) Köing, W.; Geiger, R. *Chem. Ber.* **1970**, *103*, 788-798. (b) Köing, W.; Geiger, R. *Chem. Ber.* **1973**, *106*, 3626-3635.

94) (a) Kisfaludy, L.; Löw, M.; Nyéki, O.; Szirtes, T.; Schön, I. *Ann. Chem.* **1973**, 1421-1429. (b) Kisfaludy, L.; Schön, I. *Synthesis* **1983**, 325-327.

95) (a) Köing, W.; Geiger, R. *Chem. Ber.* **1970**, *103*, 2024-2033. (b) Köing, W.; Geiger, R. *Chem. Ber.* **1970**, *103*, 2034-2040

96) Merrifield, R. B. *J. Am. Chem. Soc.***1963**, *85*, 2149-2154.

97) Benoiton, N. L. *Chemistry of Peptide Synthesis*, CRC Press. **2005**

98) Kaiser, E.; Colescott, R. L.; Bossinger, C.D.; Cook, P. I. *Anal. Biochem.* **1970**, *34*, 595-598.

99) (a) Dörwald, F. Z., Organic synthesis on solid phase: supports, linkers, reactions. In 2nd ed.; Wiley-VCH, **2002**; pp. 7. (b) Christensen, T.; *Acta Chemica Scandinavica B* **1979**, 763-766.

100) Orndorff, W. R.; Hemmer, A. J. *J. Am. Chem. Soc.***1927**, *49*, 1272-1280.

101) (a) Adamczyk, M.; Fishpaugh, J. R.; Heuser, K. J. *Bioconjugate Chem.* **1997**, *8*, 253-255. (b) Adamczyk, M.; Chan, C. M.; Fino, J. R.; Mattingly, P. G. *J. Org. Chem*. **2000**, *65*, 596-601.

102) (a) Gait, J. M.; *Oligonucleotide synthesis: A practical approach.* IRL Press Oxford, United Kingdom 217, **1984**. (b) Agrawal, S.; *Protocols for oligonucleotides and analogs: Synthesis and properties*; *Methods in Molecular Biology*. (ed): vol 20: Totowa, NJ. Humana Press, Inc., **1993**.

103) (a) Kibbe W. A.; *Nucleic Acids Res.* **2007**, *35.* (b) Oligo Calculator version 3.26, http://www.basic.northwestern.edu/ biotools/oligocalc.html (accessed June 7, 2010)

104) Cantor, C. R.; Warshaw , M. W.; Shapiro., H. *Biopolymers*, **1970**, *9*, 1059-1070.

105) (a) Babu, I. R.; Ganesh, K. N. *J. Am. Chem. Soc.*, **2001**, *123*, 2079-2080. (b) Umashankara, M.; Babu, I. R.; Ganesh, K. N. *Chem. Commun*, **2003**, 2606-2607.

106) (a) Kobayashi, Y.; Isemura, T. in *Progress in Polymer Science* **1972**, *3*, Okamura, S.; Takayanagi, M., Eds., Kodansha Limited, Tokyo & John Wiley & Sons, New York, pp 315-380. (b) Inouye, K.; Sakakibara, S.; Prockop, D. J. *Biochim. Biophys. Acta* **1976**, *420*, 133-141. (c) Weber, R. W.; Nitschmann, H. *Helv. Chim. Acta* **1978**, *61*, 701-708. (d) Inouye, K.; Kobayashi, Y.; Kyogoku, Y.; Kishida, Y.; Sakakibara, S.; Prockop, D. J. *Arch. Biochem. Biophys.* **1982**, *219*, 198-203. (e) Engel, J.; Chen, H. T.; Prockop, D. J.; Klump, H. *Biopolymers* **1977**, *16*, 601-622.

107) Cheung, M. S.; Garcia, A. E.; Onuchic, J. N. *Proc. Nat. Acad. Sci. USA* **2002**, *99* 685-690

108) (a) Gerisma, S. Y.; Stuur, E. R. *Int. J. Pept. Protein Res*. **1972**, *4*, 377-383. (b) Back, J. F.; Oakenfull, D.; Smith, M. B. *Biochemistry* **1979**, *18*, 5191-5196. (c) Arakawa, T.; Timasheff, S. N. *Biochemistry* **1982**, *21*, 6536-6544.

109) (a) Harrap, B. S. *Int. J. Pept. Protein Res*. **1969**, *1*, 2527-2532. (b) Gekko, K.; Koga, S. *J. Biochem*. **1983**, *94*, 199-205. (c) Feng, Y.; Melacini, G.; Taulane, J. P.; Goodman, M. *J. Am. Chem.Soc*. **1996**, *118*, 10351-10358.

110) Creighton, T. E. *Proteins*, W. H. Feeman & Co., New York, **1993,** pp. 262-263.

111) (a) Callis, P. R. *Annu. Rev. Phys. Chem.* **1983**, *34*, 329-357. (b) Ho, P. S.; Zhou, G.; Clark, L. B. *Biopolymers* **1990**, *30*, 151-163. (c) Theiste, D.; Callis, P. R.; Woody, R. W. *J. Am. Chem. Soc.* **1991**, 113, 3260-3267.

112) (a) Gray, D. M.; Hung, S. H.; Johnson, K. H. *Methd. In Enzymol.* **1995**, *246*, 19-34 (b) Gray, D. B.; Ratliff, R. L.; Vaughan, M. R. *Methods in Enzymol.* **1992**, *211*, 389-406. (c) Ivanov, V. I.; Krylov, D. Yu. Methd. Enzymol. **1992**, *211*, 111-127.

113) Baase, W. A.; Johnson, W. C. J. *Nucleic Acids Res.* **1979***, 6*, 797-814.

114) Haley, J.; Kabiru, P.; Geng, Y. *Mol. BioSyst.* **2010,** *6*, 249-255.

115) Gray, D. M.; Morgan, A. R.; Ratliff, R. L. *Nucleic acids research* **1978,** *5*, 3679-95.

116) Damian, M. S.; Hedman, H. K.; Elmroth, S. K. C.; Diederichsen, U. *Eur. J. Org. Chem*. **2010,** *2010*, 6161-6170.

117) (a) Dash, P.R.; Toncheva, V.; Schacht, E.; Seymour, L.W. *Journal of controlled release*, **1997**, *48*, 269-276. (b) Chiou, H. C.; Tangco, M. V. ; Levine, S. M.; Robertson, D.; Kormis, K. Wu, C. H.; Wu, G. Y. *Nucleic Acids Res*. **1994**, *22*, 5439-46.

118) Kessler, C.; Manta, V. *Gene* **1990**, *92*, 1–248.

119) Lodish, H. F. *Molecular cell biology*. 5th ed.; New York : W. H. Freeman and Co., **2003**, *973* s. b ill. ISBN 0-7167-4366-3

120) Schneider, I. *J. Embryol. Exp. Morphol.* **1972**, *27*, 353-365.

121) Santos, M. G.; Jorge, S. A. C.; Brillet, K.; Pereira, C. A.; *Cytotechnology* **2007**, *54*,15-24.

122) Buster, D. W.; Nye, J.; Klebba, J. E.; Rogers, G. C. *J. Vis. Exp.* **2010**, *40.*1982.

123) (a) Seelig, J. *Biochim. Biophys. Acta*, **1997**, *1331*, 103-116. (b) Berger, C.; Jelesarov, I.; Bosshard, H. R., *Biochemistry* **1996,** *35*, 14984-4991.

124) (a) Dragan, A. I.; Klass, J.; Read, C.; Churchill, M. E.; Crane- Robinson, C.; Privalov, P. L. *J. Mol. Biol.* **2003,** *331*,795-813. (b) Lundback, T.; Hansson, H., Knapp, S., Adenstein, R., Hard, T. *J. Mol. Biol.* **1998**, *276*, 775-786.

125) (a) Poklar, N.; Pitch, D. S.; Lippard, S. J.; Redding, E. A.; Dun ham, S. U.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA* **1996**, 93, 7606–7611. (b) Olmsted, M. C.; Bond, J. P.; Anderson, C. F.; Record, M. T. *Bio. phys. J.* **1995**, 68, 634–647.

126) (a) Woody, R. W. *Adv. Biophys. Chem.* **1992**, *2*, 37-79. (b) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II structure in a sequence of seven alanine residues, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190-9195. (c) Rucker, A. L., Creamer, T. P. *Protein Sci.* **2002**, *11*, 980-985. (d) Chellgren, B. W.; Creamer, T. P.; *Biochemistry* **2004**, *43*, 5864-5869.

127) Ikura, T.; Urakubo, Y.; Nobutoshi, I. *Chem. Phys.* **2004**, *307*, 111-119.

128) (a) Von Hippel, P. H.; Wong, K. *Biochemistry* **1962**, *1*, 664-674. (b) Von Hippel, P. H.; Wong, K. *Biochemistry* **1963**, *2*, 1387-1398. (c) Komsa-penkova, R.; Koynova, R.; Kostov, G.; Techov, B. G. *Biochim. Biophys. Acta* **1996**, *1297*, 171-181. (d) Zanaboni, G.; Rossi, A.; Onana, A. M. T.; Tenni, R. *Matrix Biol*. **2000**, *19*, 511-520

129) Böttcher, C.; Endisch, C.; Fuhrhop, J. H.; Catterall C.; Eaton, M. *J. Am. Chem. Soc.,* **1998**, *120*, 12–17.

130) Haley, J.; Kabiru, P.; Geng, Y. *Molecular bioSystems* **2010**, *6*, 249-55.

131) Felgner, P. L.; Barenholz, Y.; Behr, J. P; Cheng, S. H.; Cullis, P.; Huang, L.; Jessee, J. A.; Seymour, L.; Szoka, F.; Thierry, A. R.; Wagner, E.; Wu, G. *Hum. Gene Ther.* **1997**, *8*, 511-512.

132) (a) Middaugh, C. R.; Evans, R. K.; Montgomery, D.L.; Casimiro, D.R. *J. Pharm. Sci.* **1998**, *87* 130–146. (b) Remaut, K.; Sanders, N. N.; Fayazpour, F.; Demeester, J.; De Smedt, S. C. *J. Control. Release* **2006**, *115*, 335–343.

133) Sanders, N. N.; De Smedt, S. C.; Demeester, J.; McGrath, B. M.; Walsh, G. (Eds.), *Therapeutic Enzymes*, Taylor & Fracis Group: Boca Raton, **2006**, pp. 97–116.

134) Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nollmann, M. *Biophys. Chem.* **2001**, *93*, 215-230.

135) (a) Dragan, A. I.; Klass, J.; Read, C.; Churchill, M. E.; Crane- Robinson, C.; Privalov, P. L. *J. Mol. Biol.* **2003**, *331*, 795-813. (b) Berger, C.; Jelesarov, I.; and Bosshard, H. R. *Biochemistry* **1996**, *35*, 14984-14991.

136) Bloomfield, V. A. *Curr. Opin. Struct. Biol.* **1996**, 6, 334-341.

137) Kichler, A.; Leborgne, C.; März, J.; Danos, O.; Bechinger, B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 1564-1568.

138) Sarin, V. K.; Kent, S. B.; Tam, J. P.; Merrifield, R. B. *Anal Biochem.* **1981,** *117*, 14757.

139) Dorwald, F. Z., Organic synthesis on solid phase: supports, linkers, reactions. In 2nd ed.; Wiley-VCH, **2002**; pp. 7.

140) Christensen, T. *Acta Chemica Scandinavica B* **1979**, 763-766.

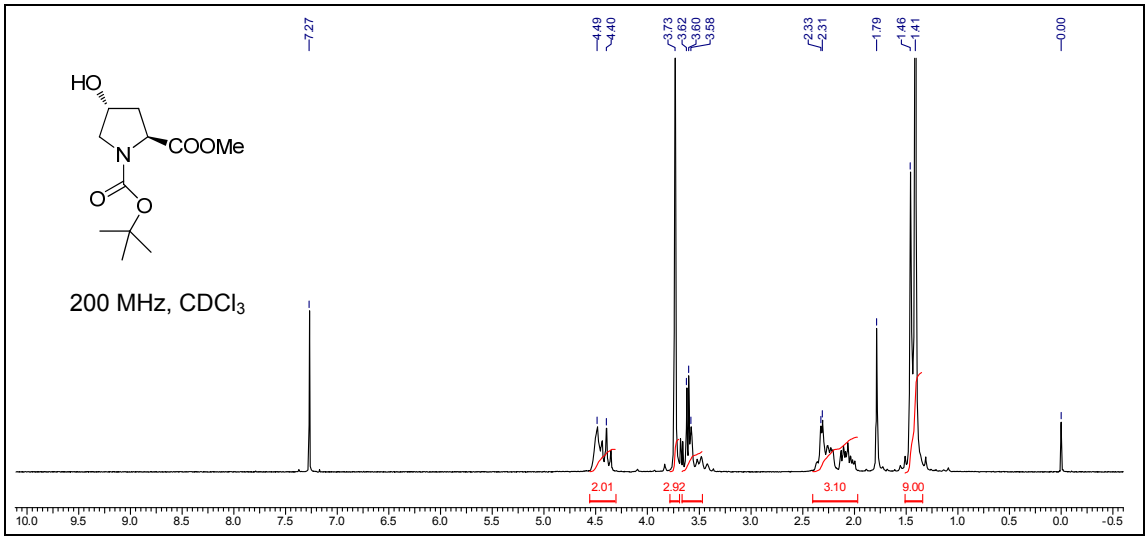
141) *Applied Biosystems* ABI 433A Peptide Synthesis User's Manual, **2001**

142) Cavaluzzi, M.J.; Borer, P. N.; *Nucleic Acids Research* **2004,** *32*, e13.

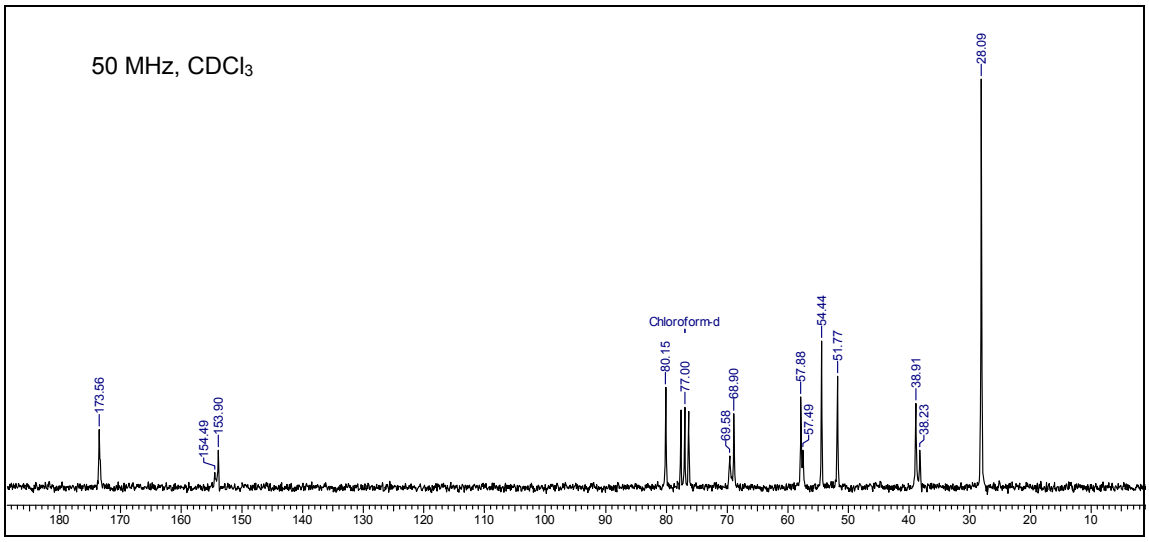
1.13 Appendix A: Characterization data of synthesized compounds

Designation	Description	Page
Compound **2**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR	126
Compound **3**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	127-128
Compound **4**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	128-130
Compound **5**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	131-132
Compound **6**	^{1}H NMR, LC-MS, FT-IR	133
Compound **7**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	134-135
Compound **8**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	136-137
Compound **9**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	138-139
Compound **10**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	140-141
Compound **11**	^{1}H NMR, LC-MS, FT-IR	142
Compound **12**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	143-144
Peptide **P1**	HPLC & MALDI-TOF spectrum (R-Am)	145
Peptide **P2**	HPLC & MALDI-TOF spectrum (RG)	146
Peptide **P3**	HPLC & MALDI-TOF spectrum (S-am)	147
Peptide **P4**	HPLC & MALDI-TOF spectrum (SG)	148
Peptide **P5**	HPLC & MALDI-TOF spectrum (SAP)	149
Peptide **P6**	HPLC & MALDI-TOF spectrum (RA)	150
Peptide **P7**	HPLC & MALDI-TOF spectrum (SA)	151
Peptide **P8**	HPLC & MALDI-TOF spectrum (RG-F)	152
Peptide **P9**	HPLC & MALDI-TOF spectrum (SG-F)	153
Peptide **P10**	HPLC & MALDI-TOF spectrum (SAP-F)	154
Oligo (**1** to **4**)	HPLC of Oligonucleotides (DNA D1, D2, D3 and D4)	155
Oligo (**1**, **2**)	MALDI-TOF spectrum (DNA D1 and D2)	156
Oligo (**3**, **4**)	MALDI-TOF spectrum (DNA D3 and D4)	157

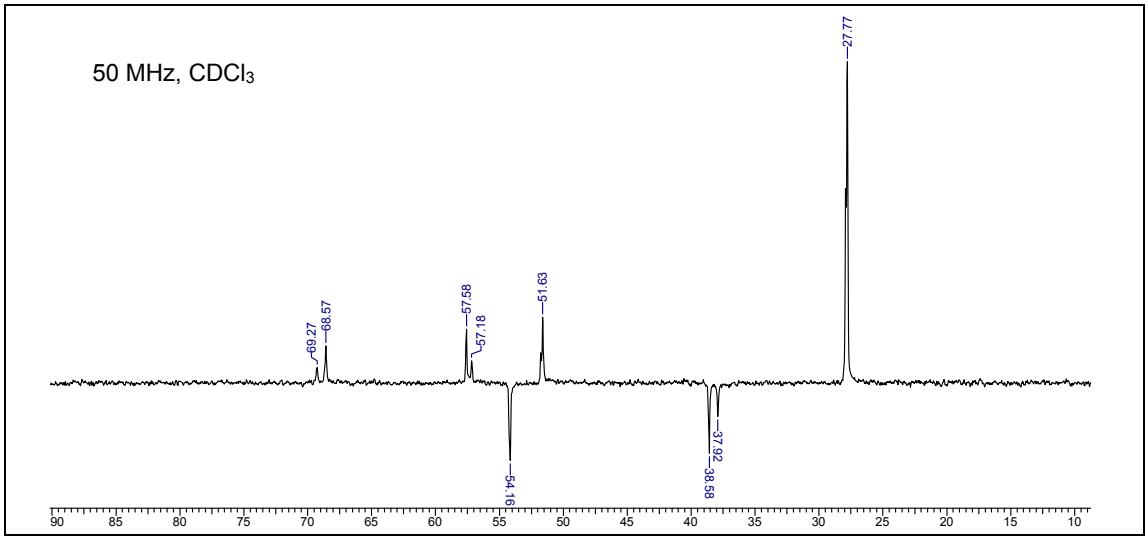
^{1}H NMR of compound 2



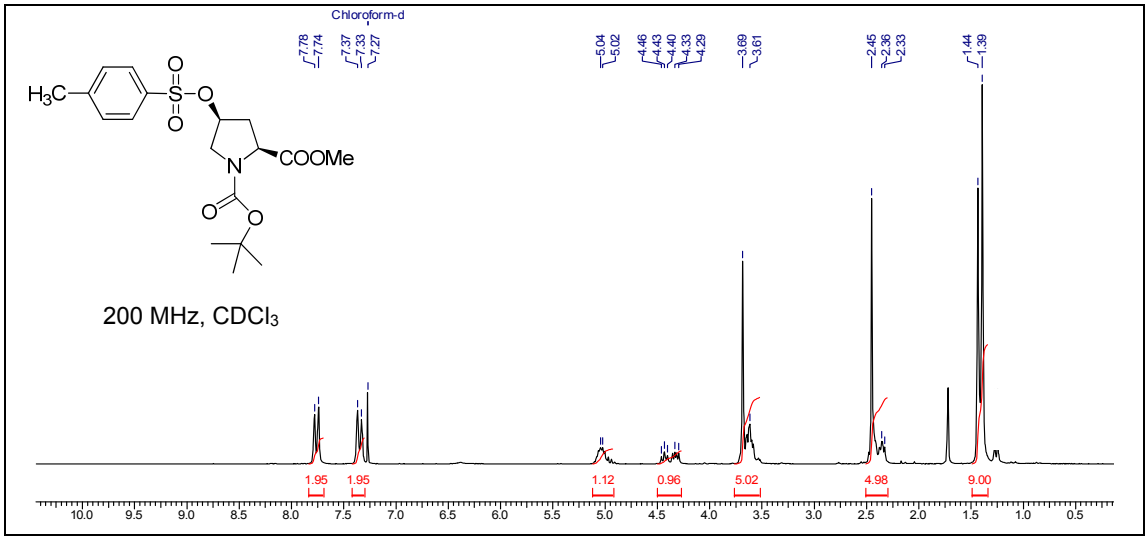
^{13}C NMR of compound 2



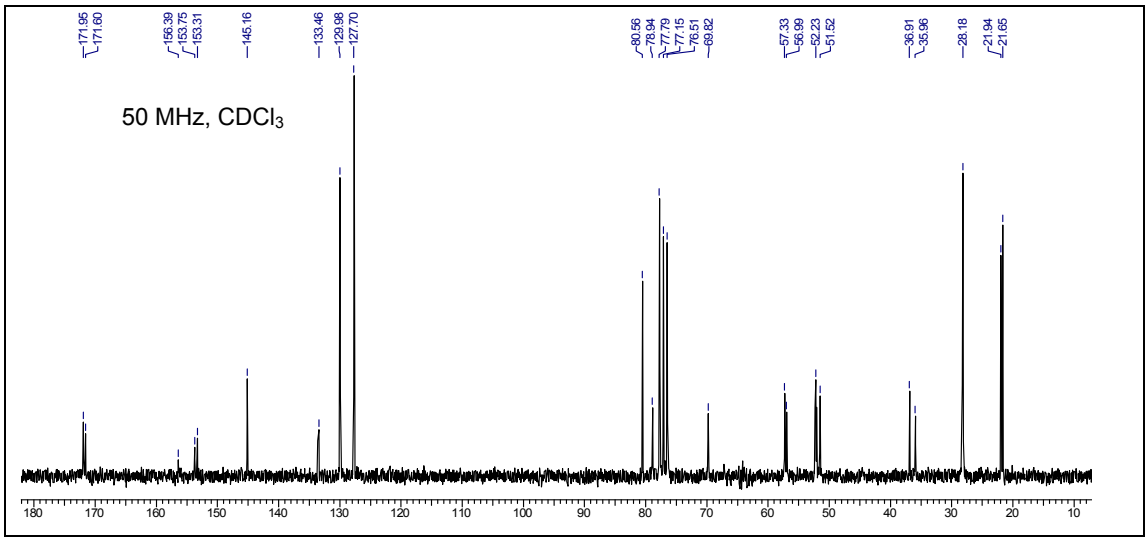
DEPT-^{13}C NMR of compound 2



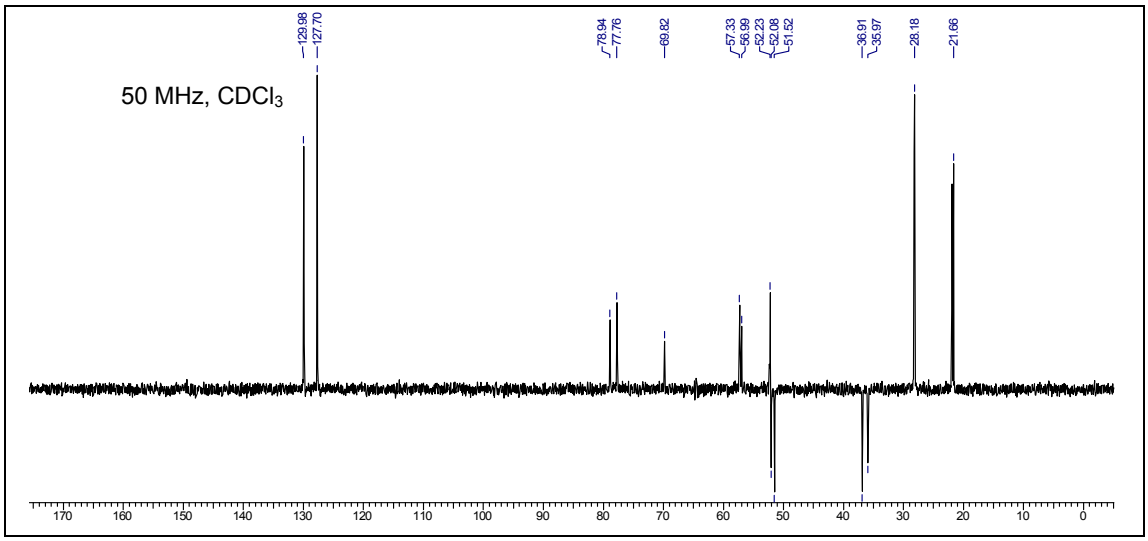
^{1}H NMR of compound 3



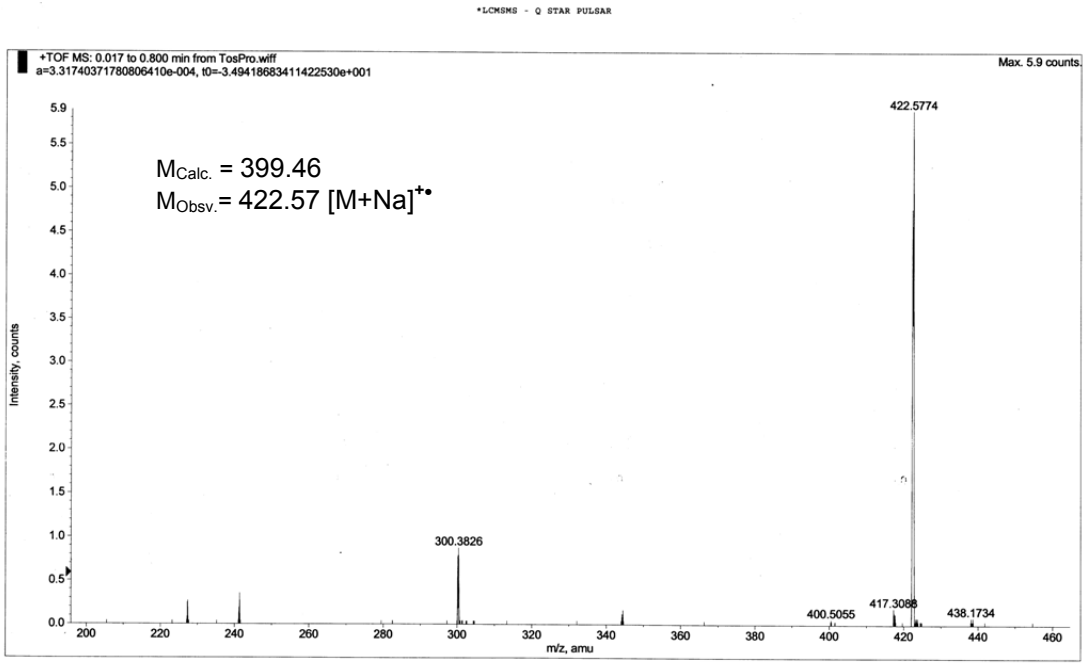
^{13}C NMR of compound 3



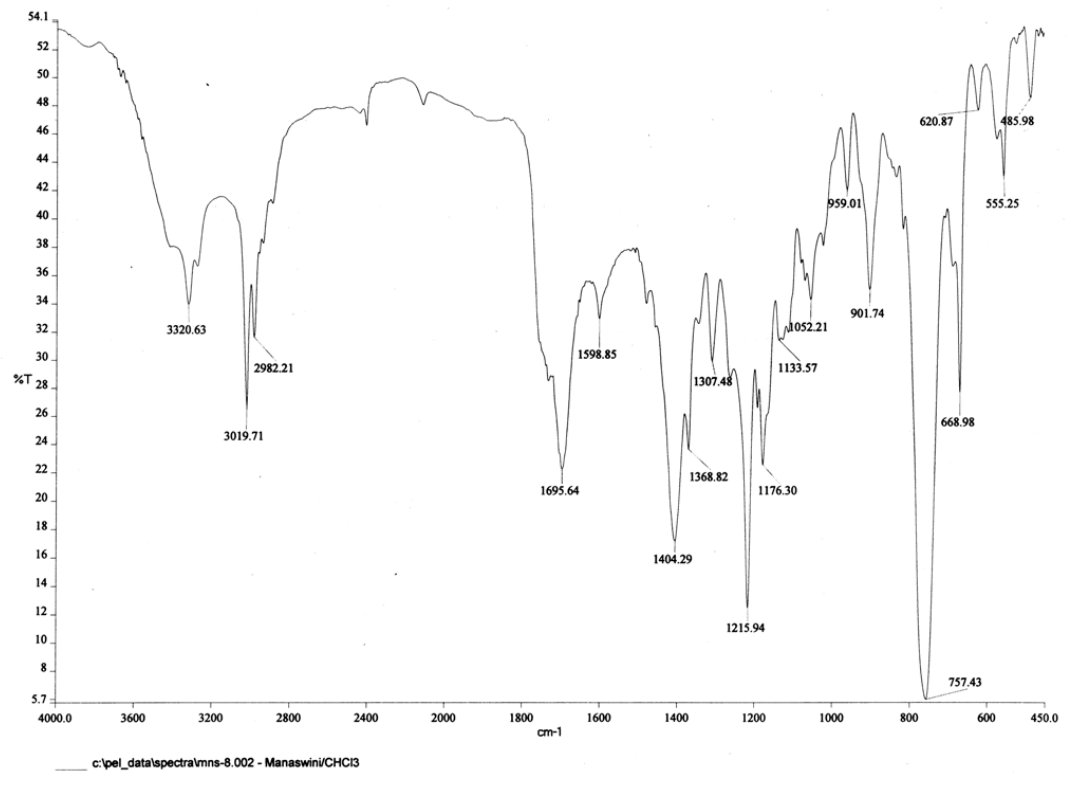
DEPT-^{13}C NMR of compound 3



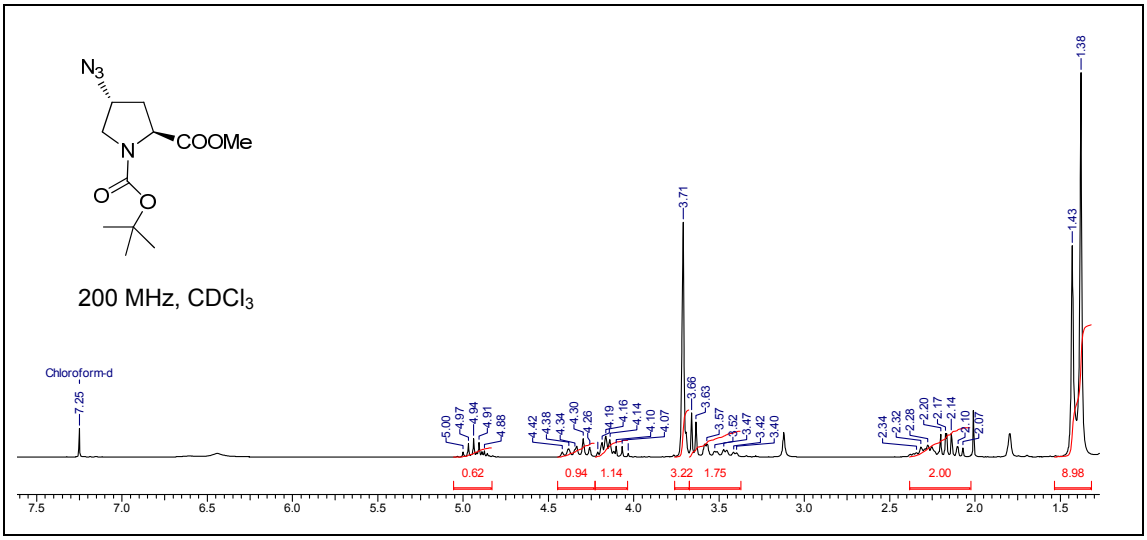
LC-MS of compound 3



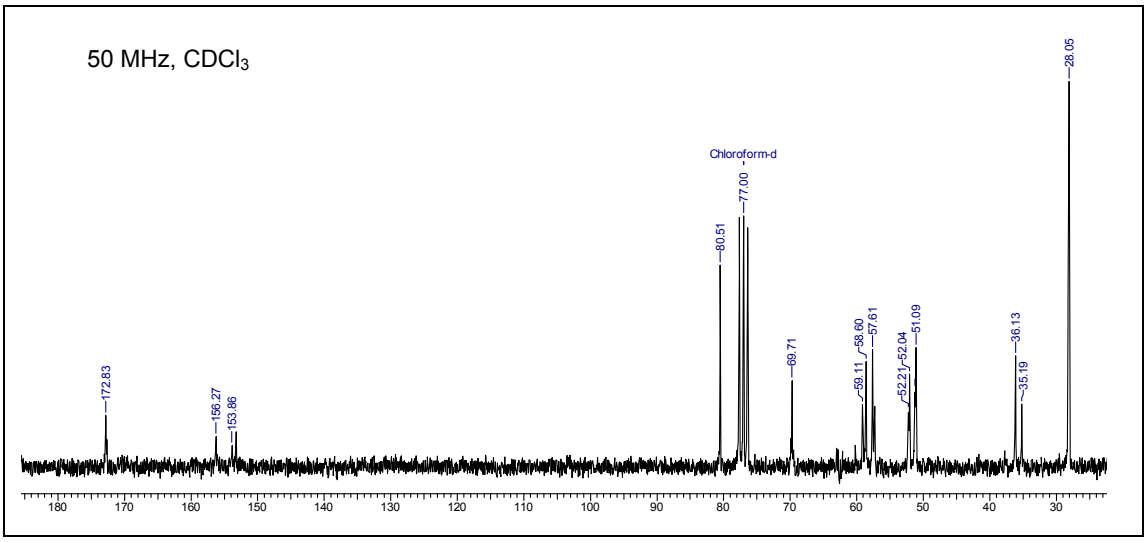
IR Spectra of compound 3



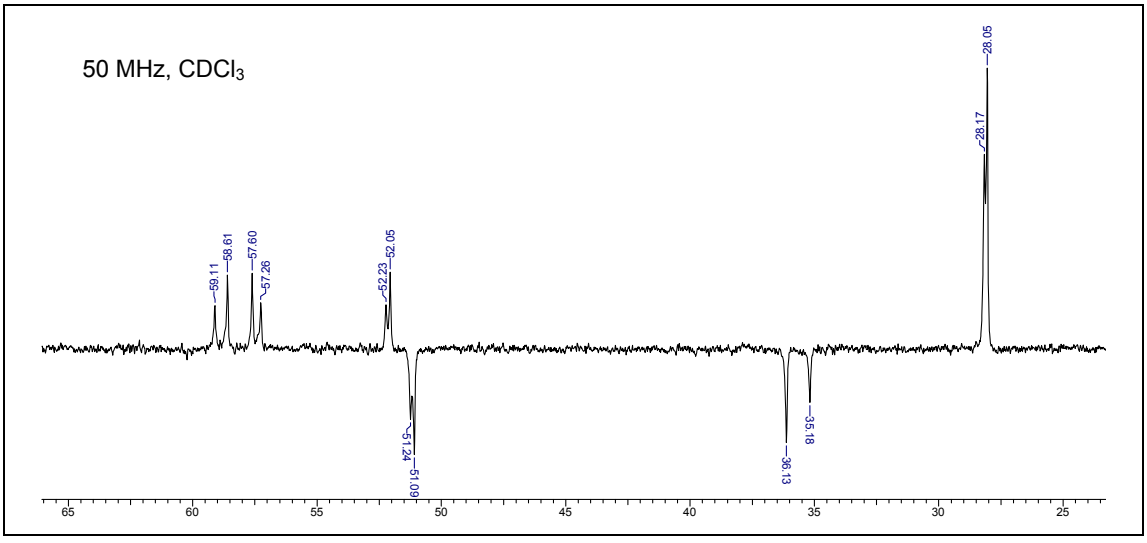
^{1}H NMR of compound 4



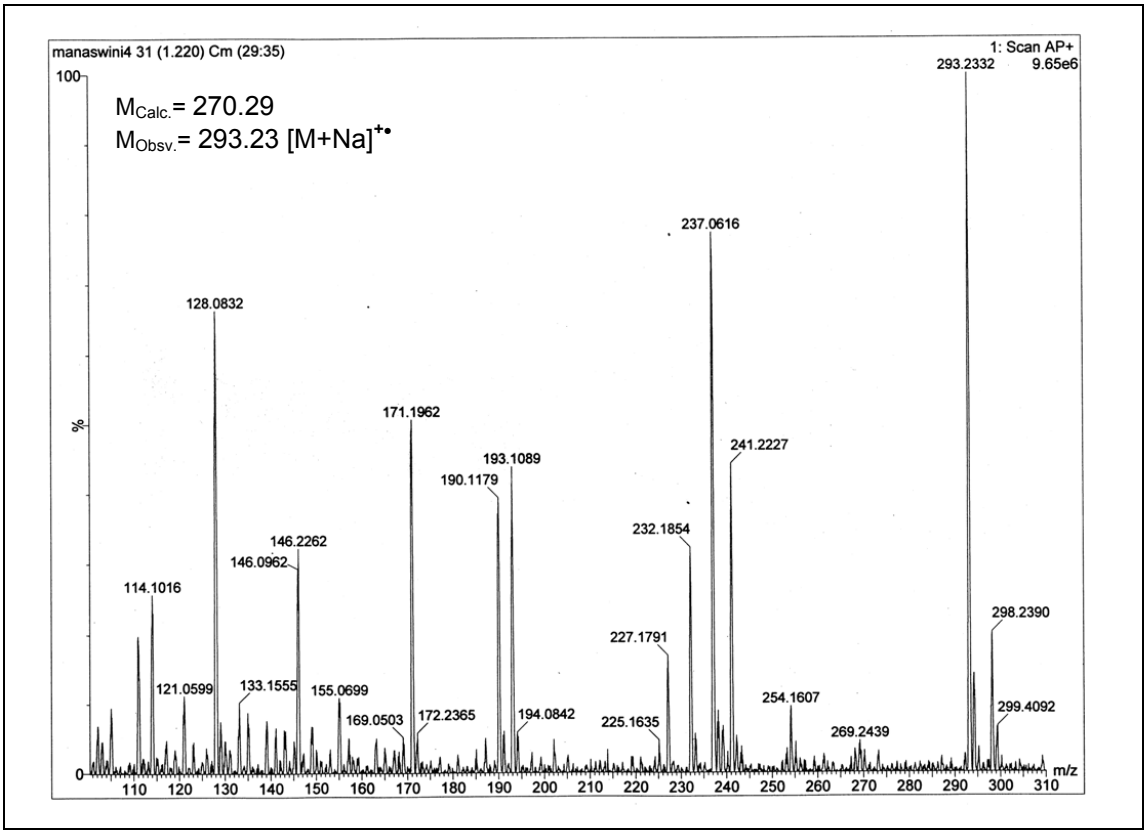
^{13}C NMR of compound 4



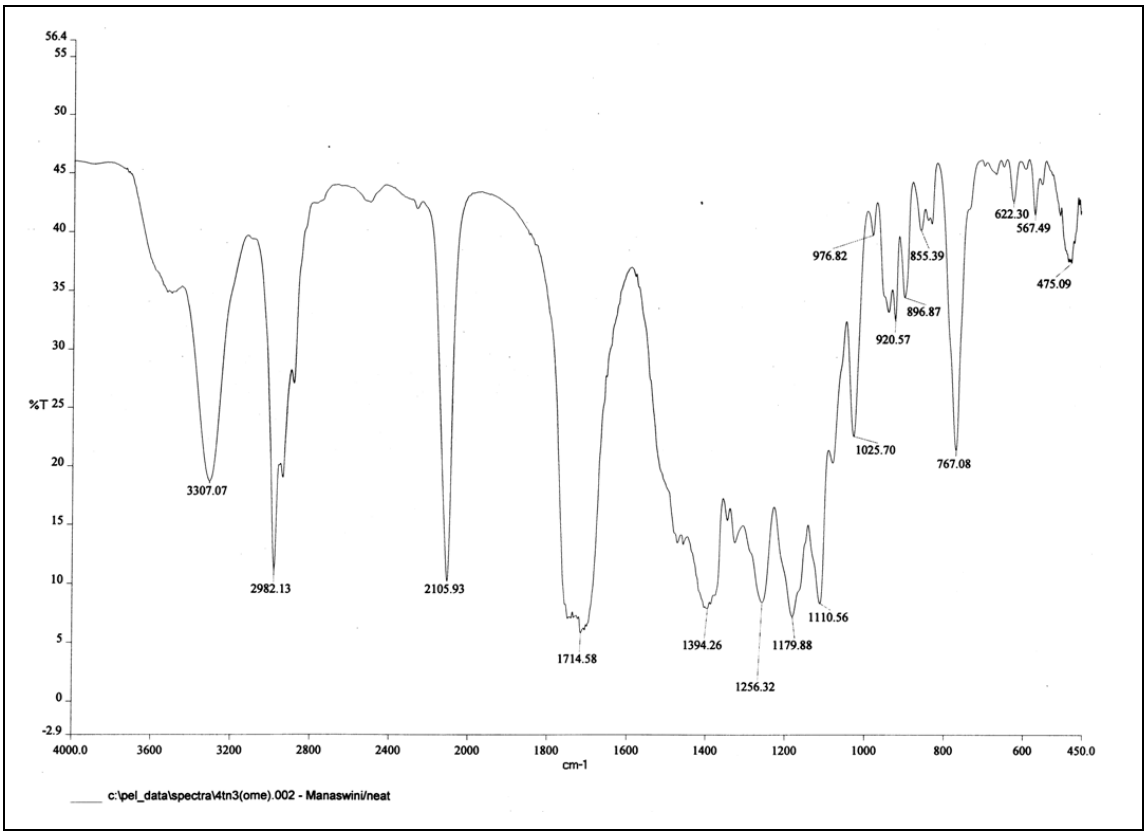
DEPT-^{13}C NMR of compound 4



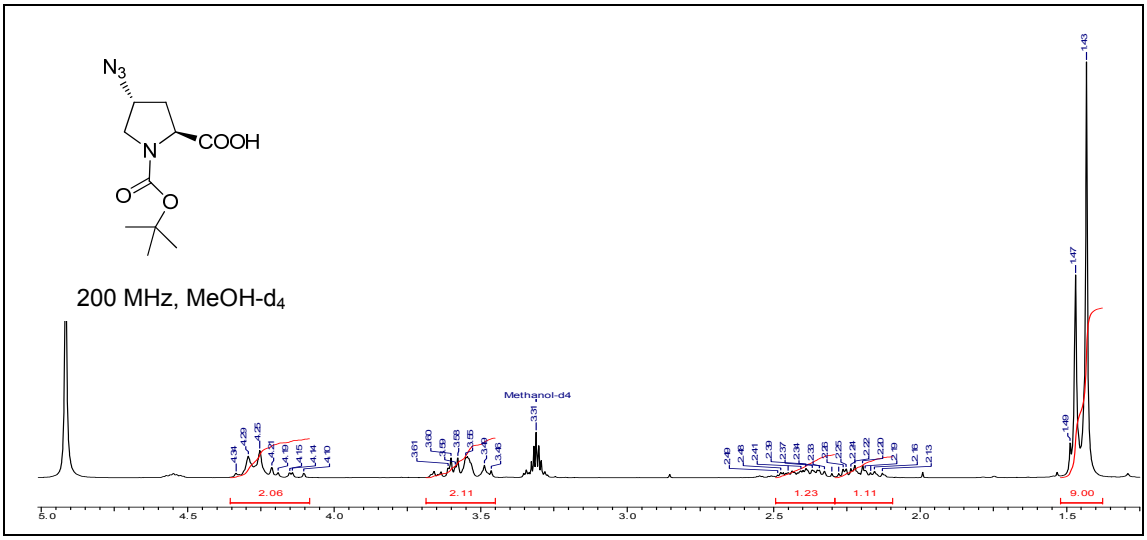
LC-MS of compound 4



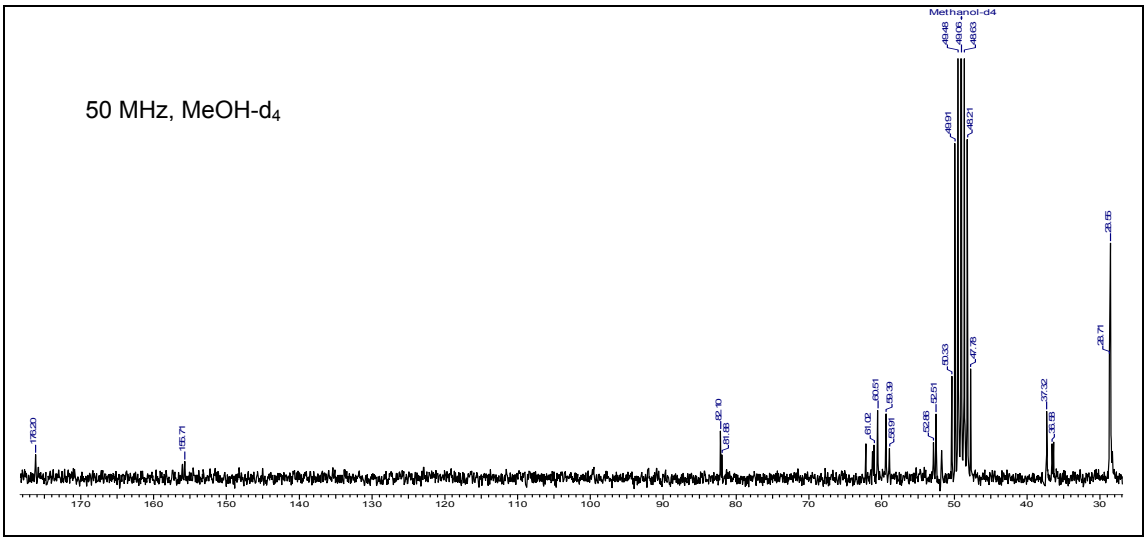
IR Spectra of compound 4



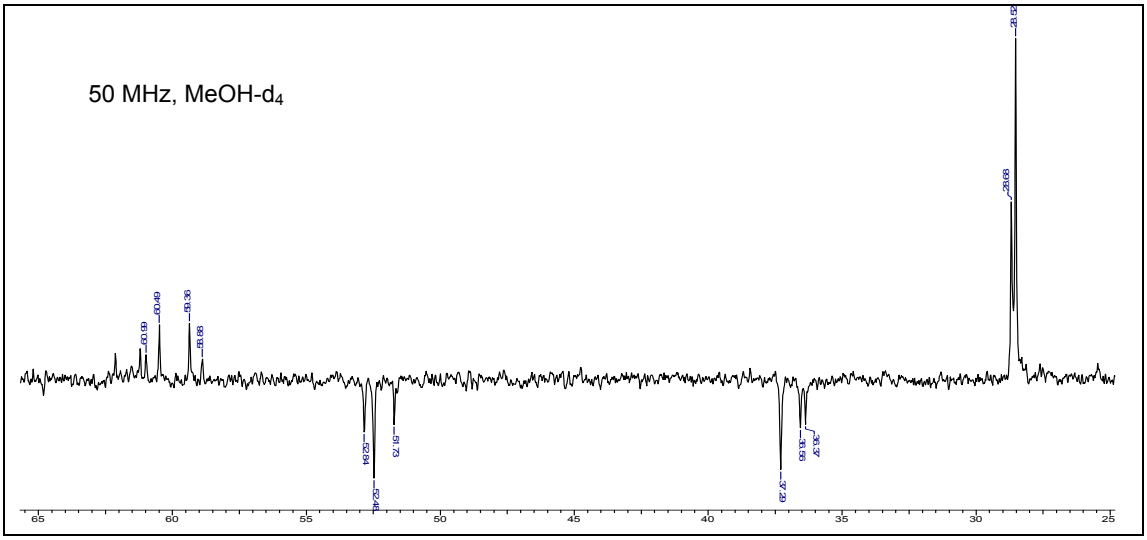
^{1}H NMR of compound 5



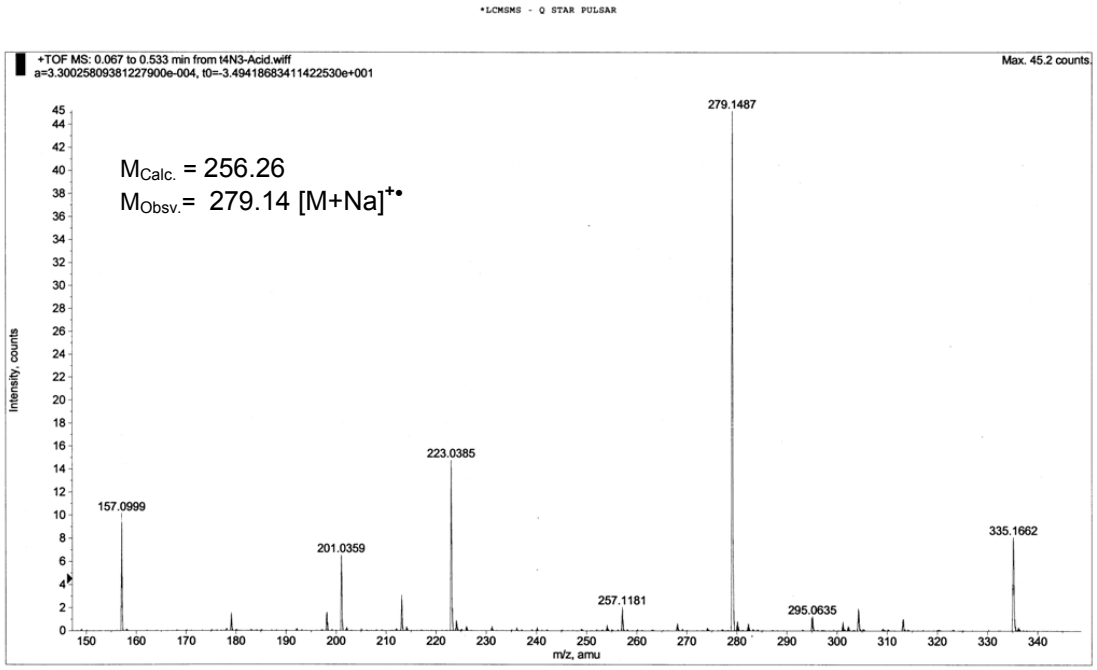
^{13}C NMR of compound 5



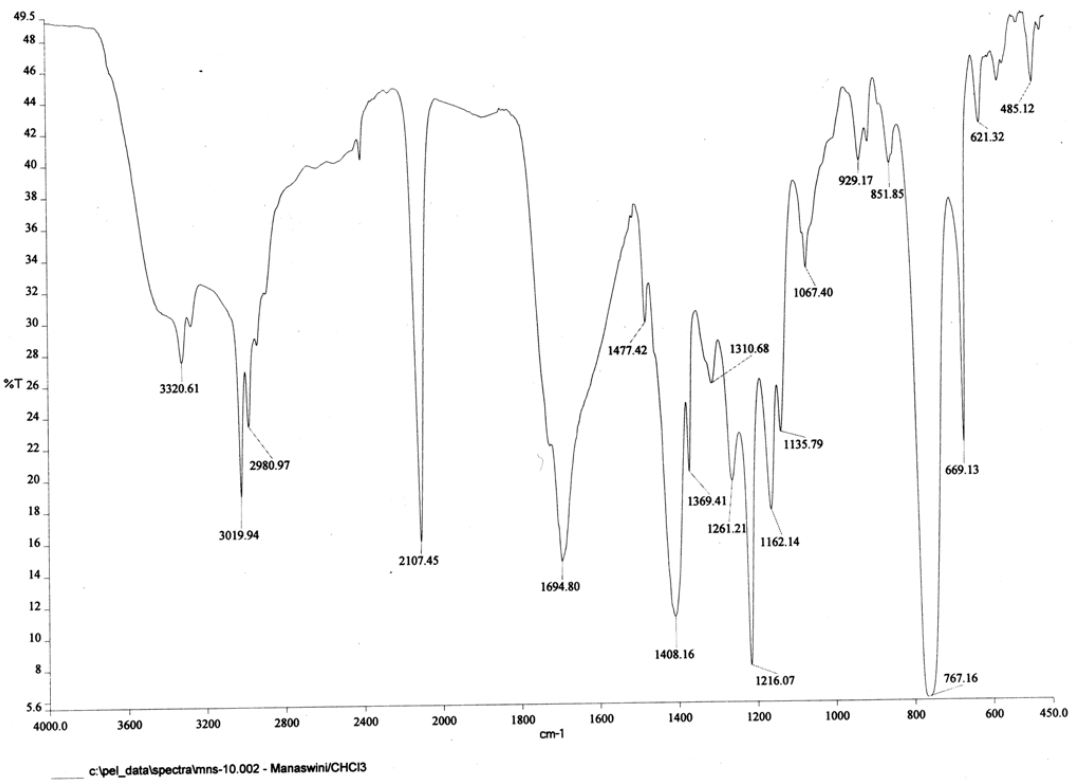
DEPT-^{13}C NMR of compound 5



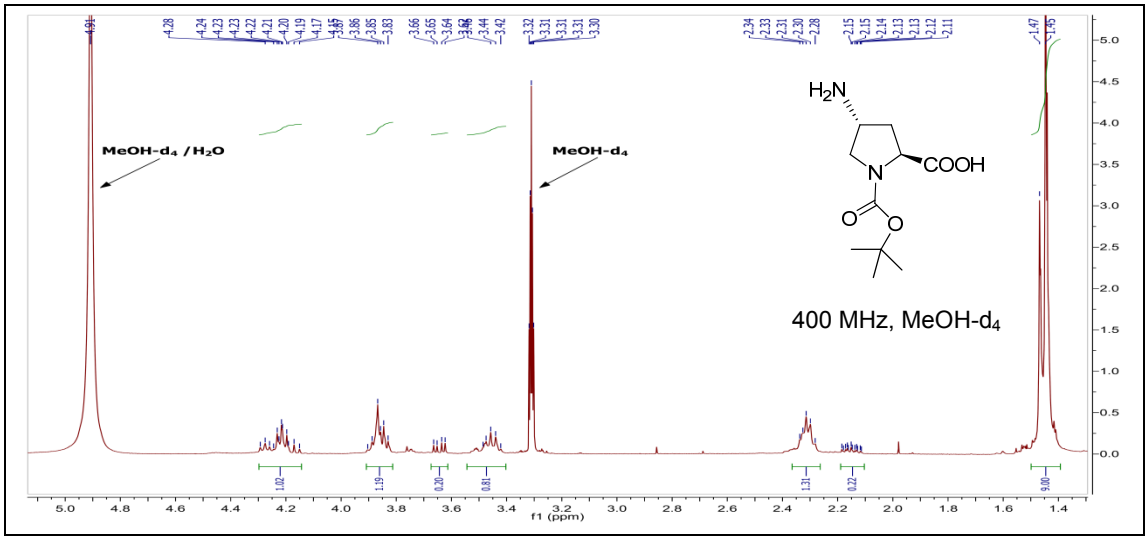
LC-MS of compound 5



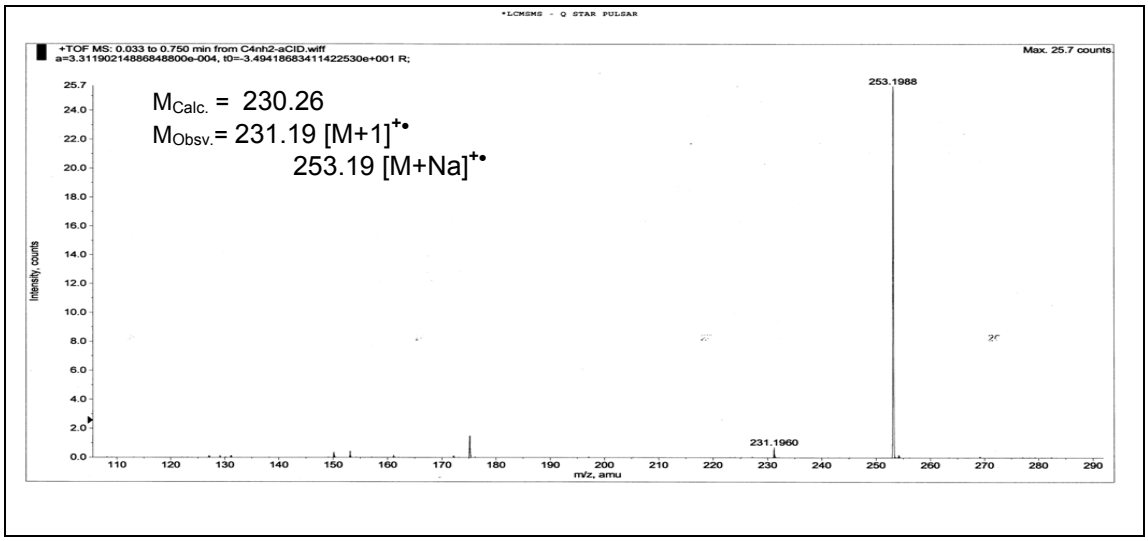
IR Spectra of compound 5



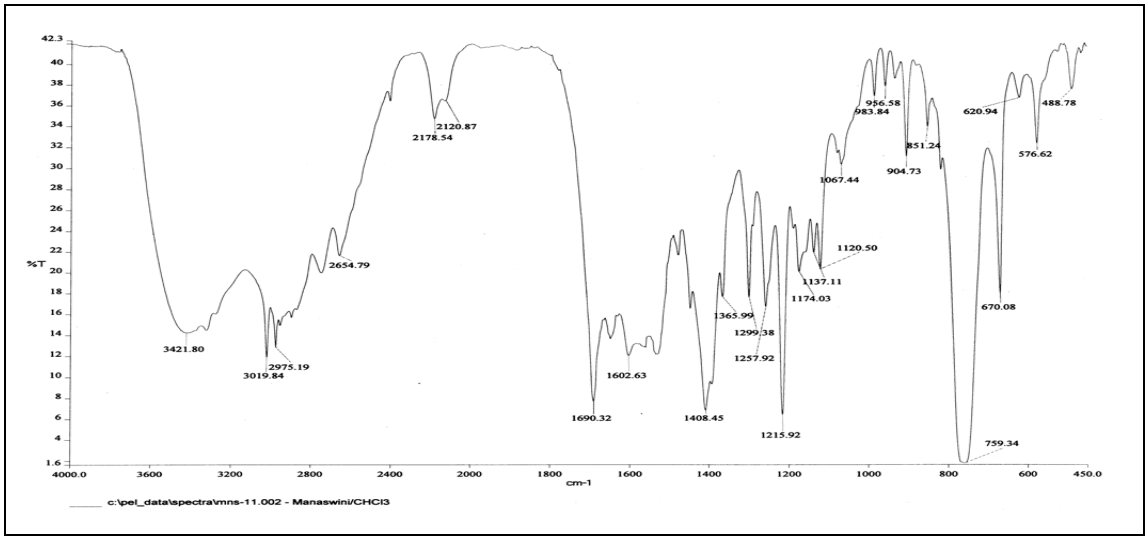
^{1}H NMR of compound 6



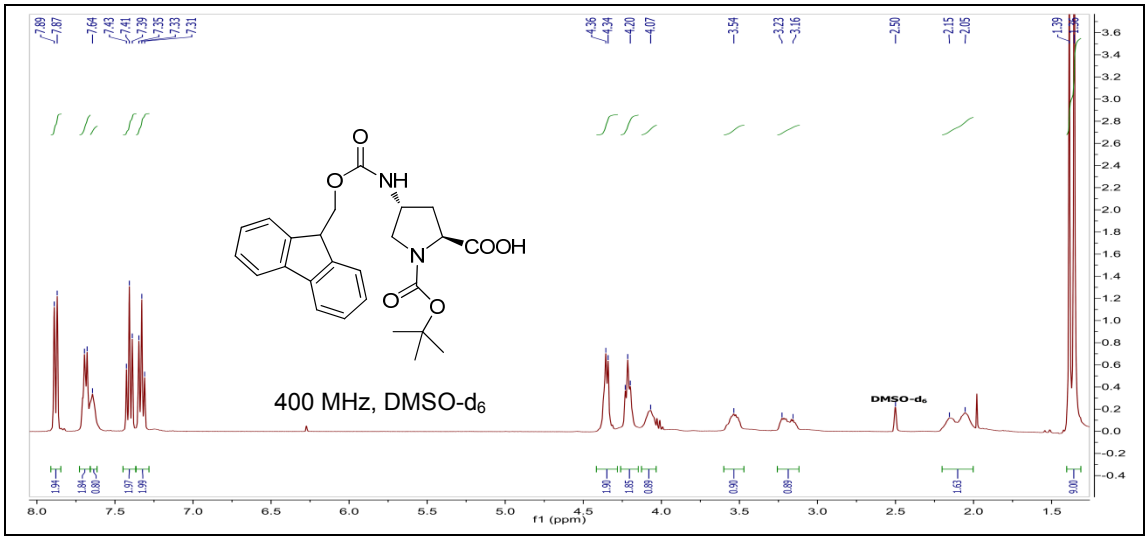
LC-MS of compound 6



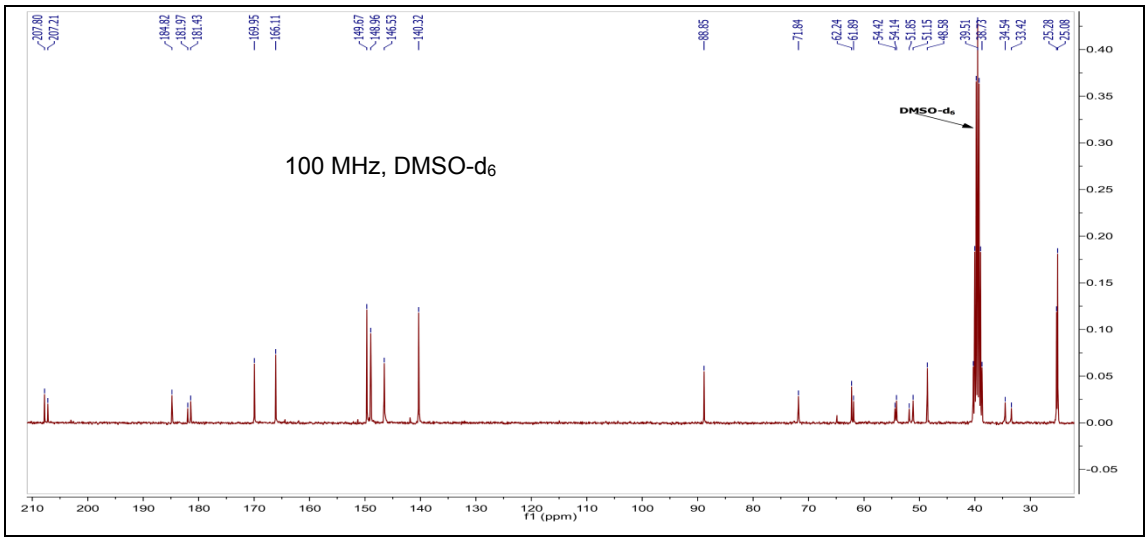
IR Spectra of compound 6



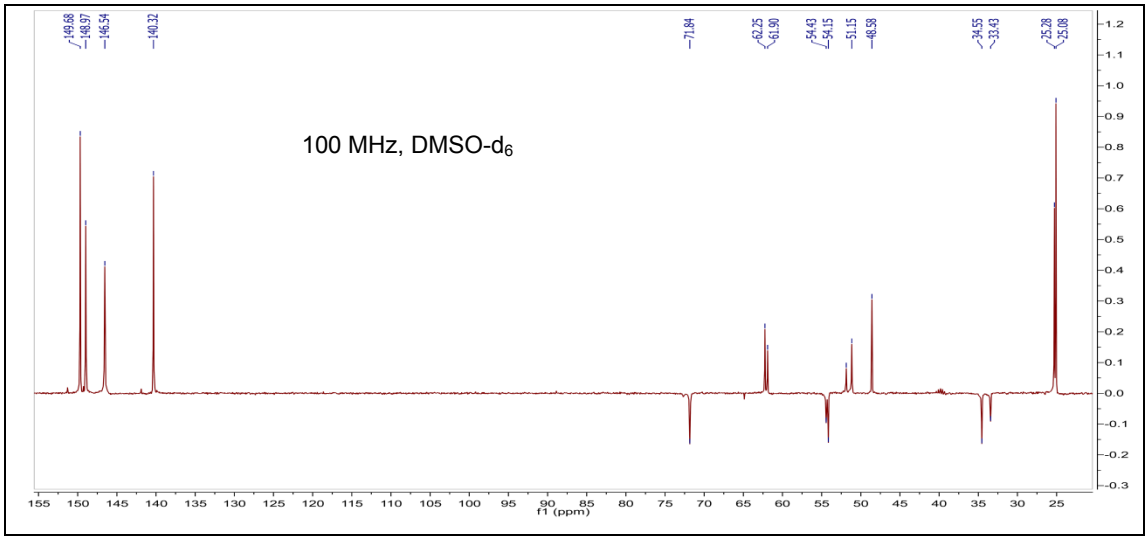
^{1}H NMR of compound 7



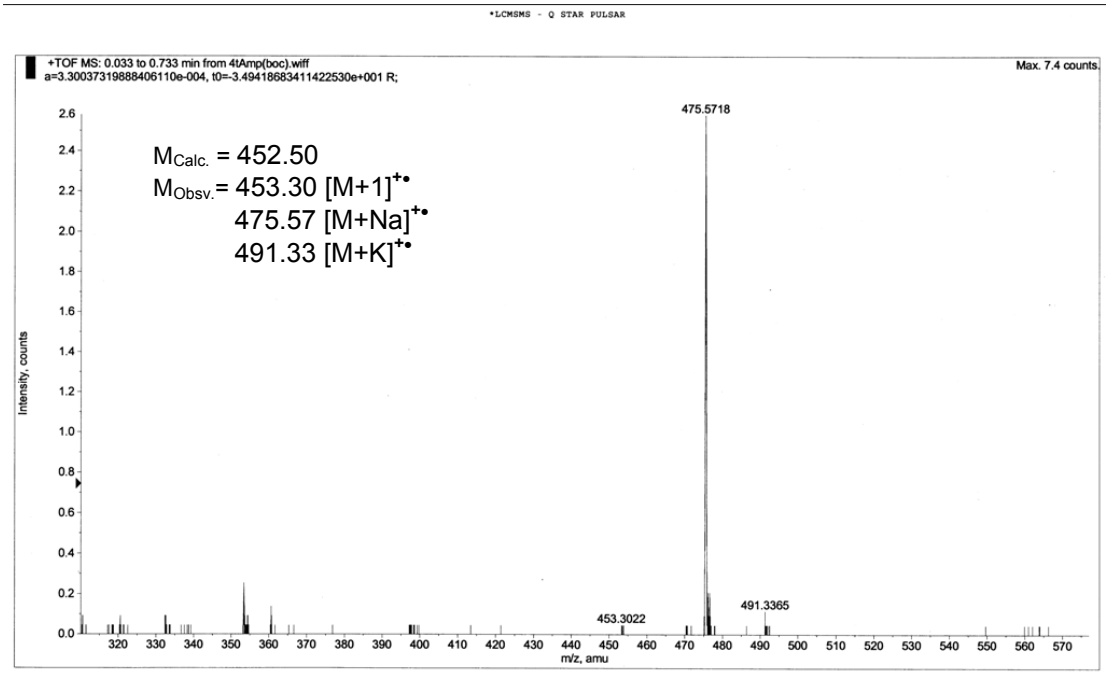
^{13}C NMR of compound 7



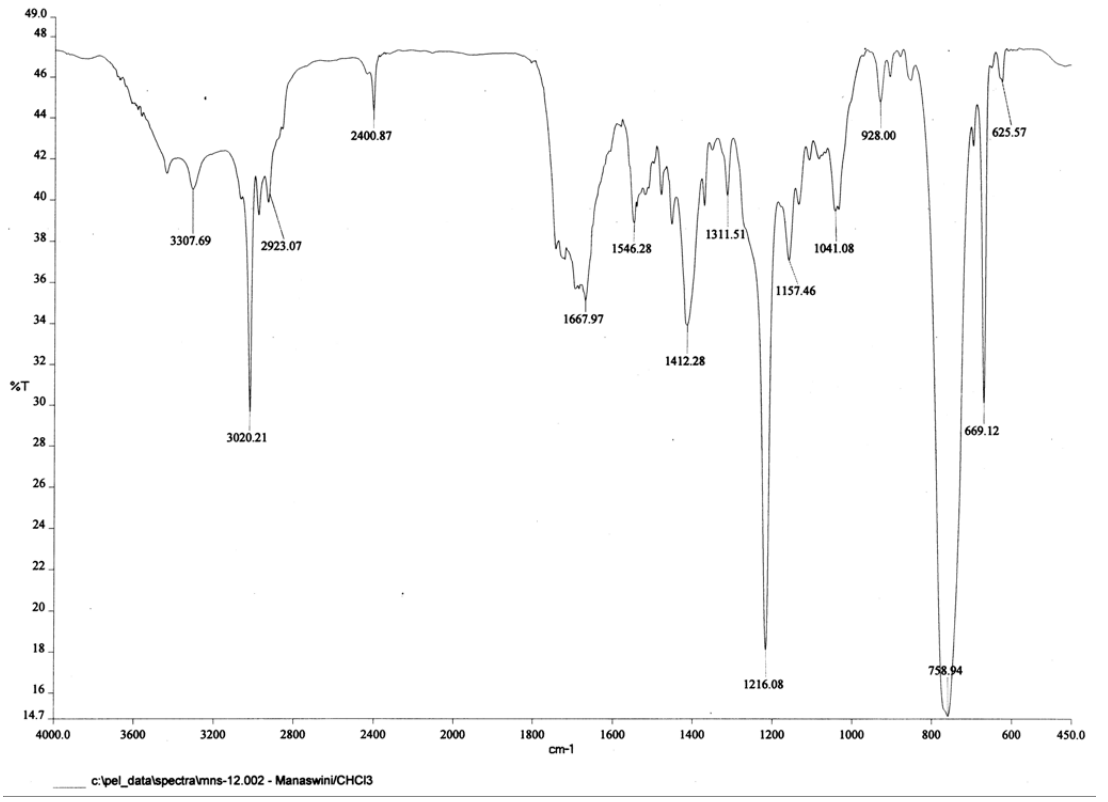
DEPT-^{13}C NMR of compound 7



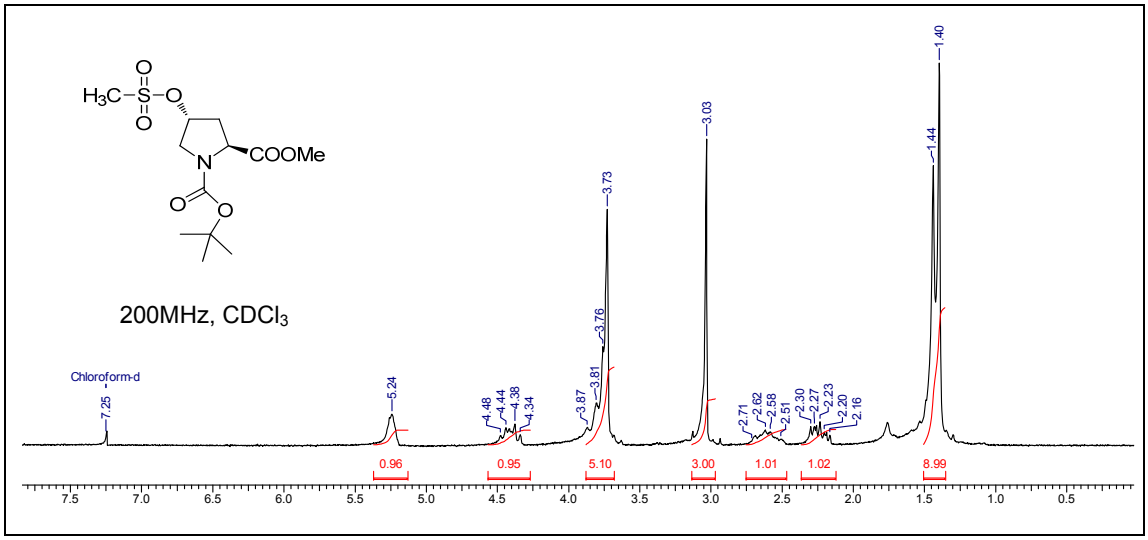
LC-MS of compound 7



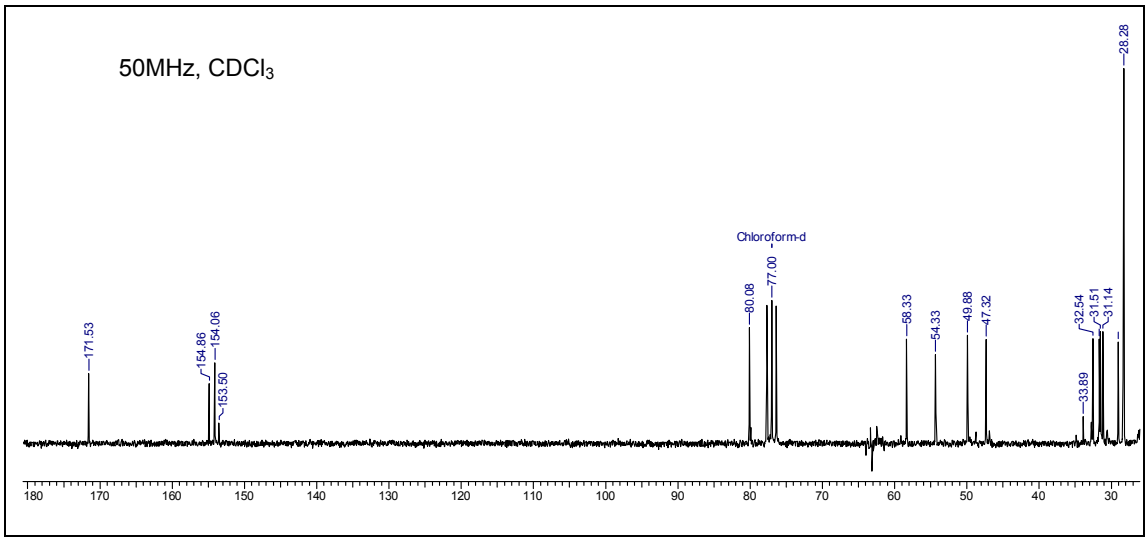
IR Spectra of compound 7



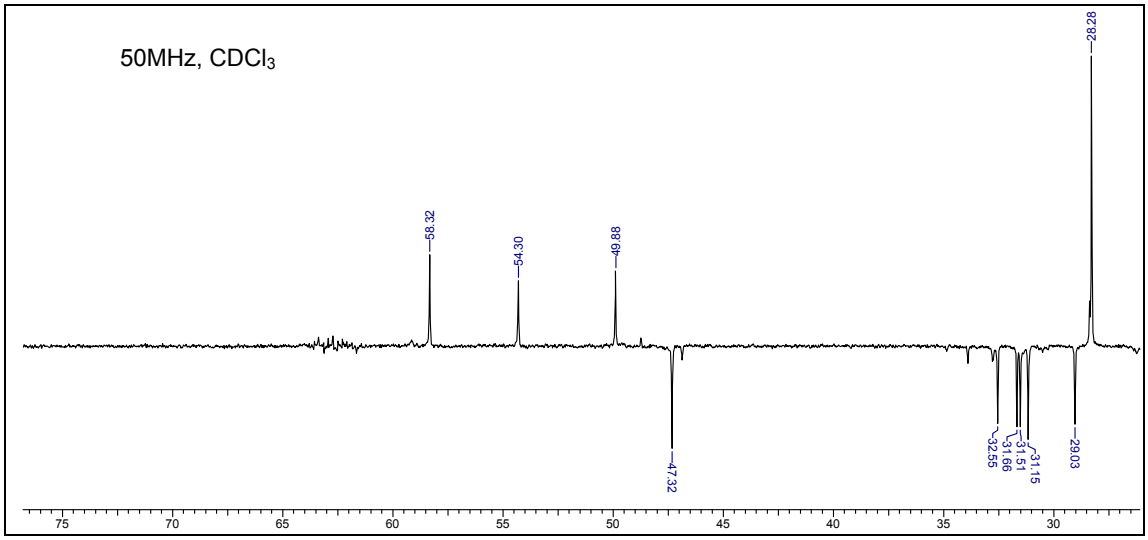
^{1}H NMR of compound 8



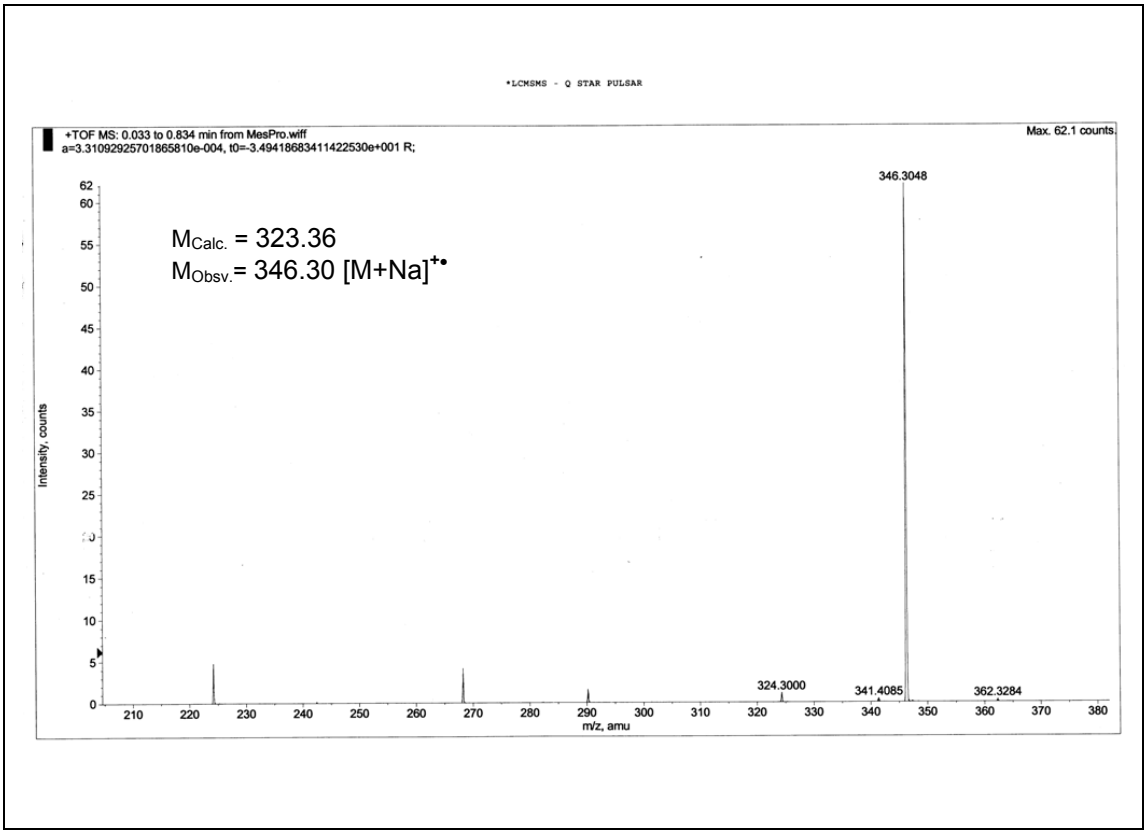
^{13}C NMR of compound 8



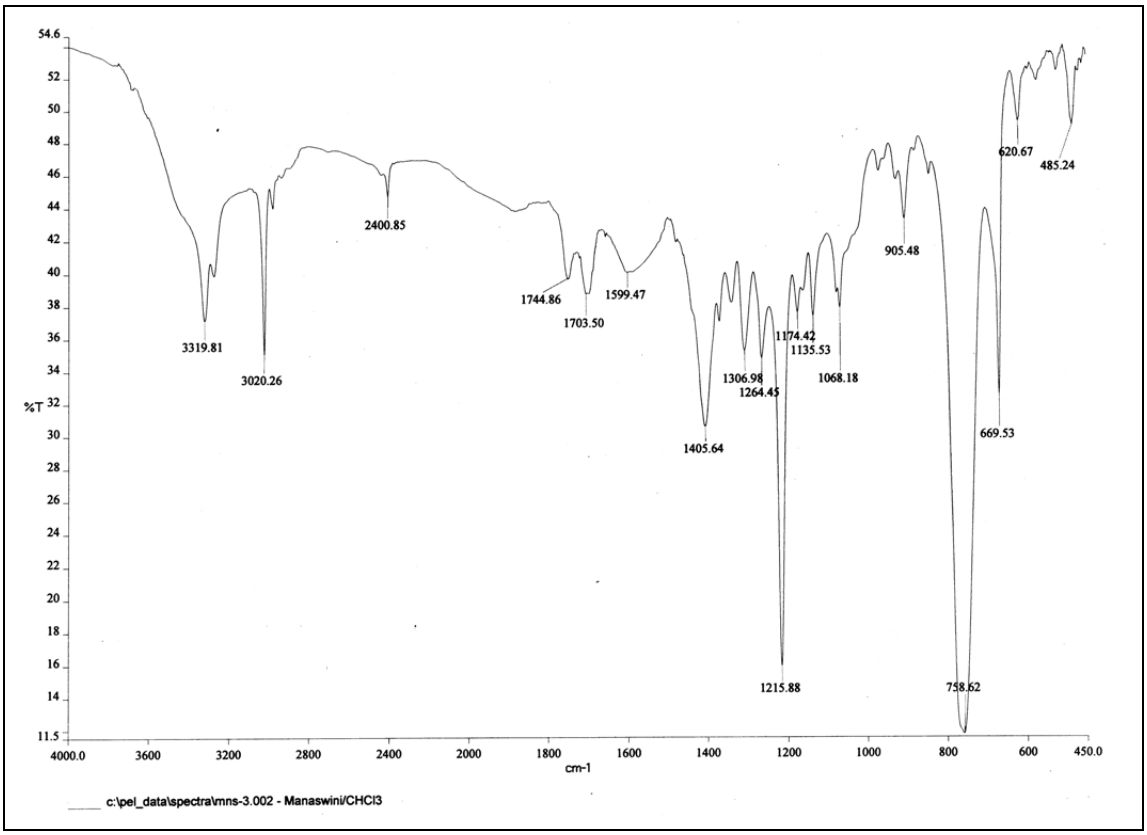
DEPT-^{13}C NMR of compound 8



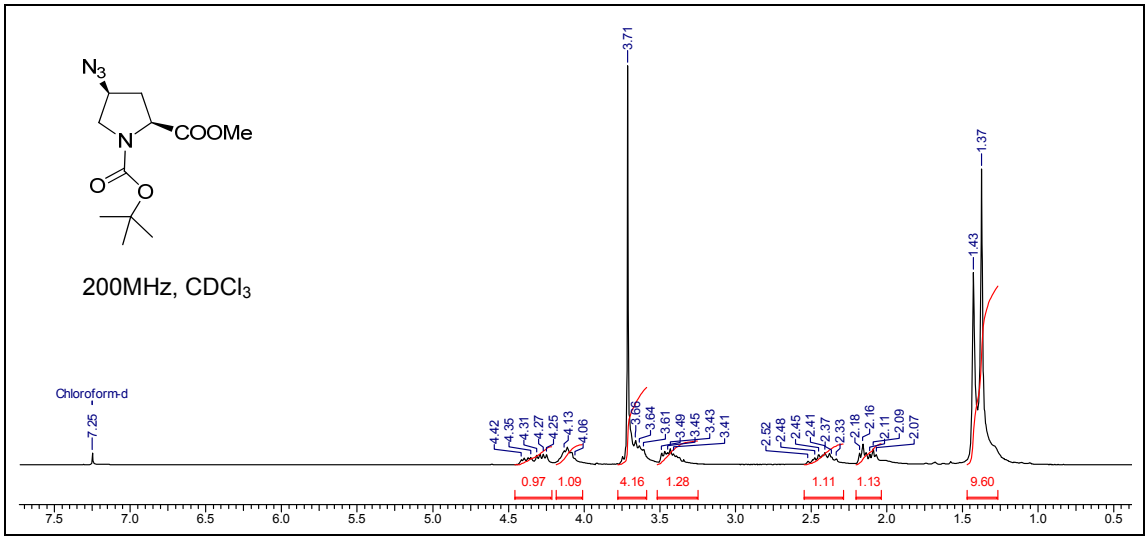
LC-MS of compound 8



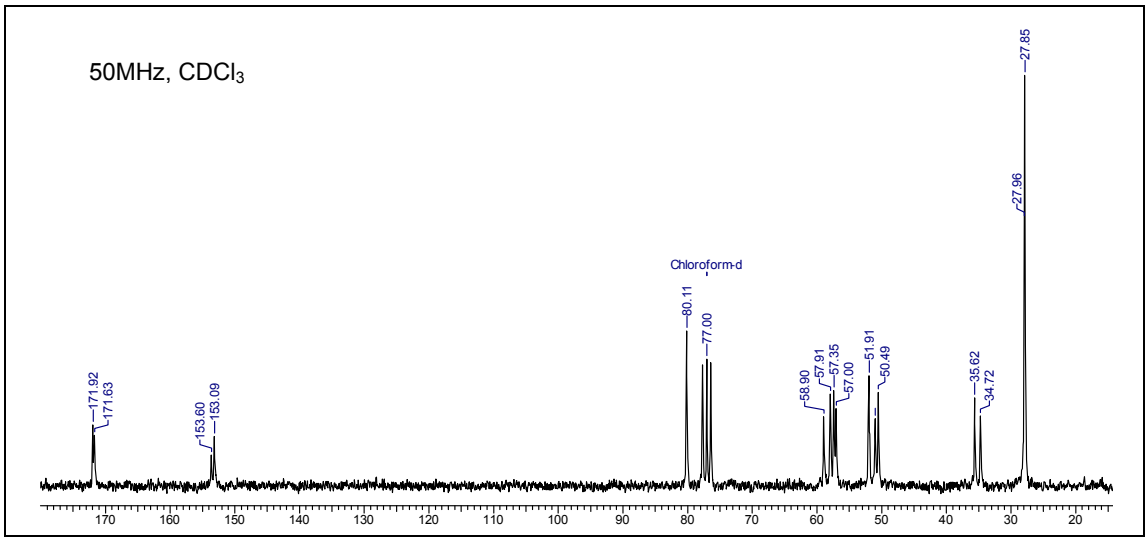
IR Spectra of compound 8



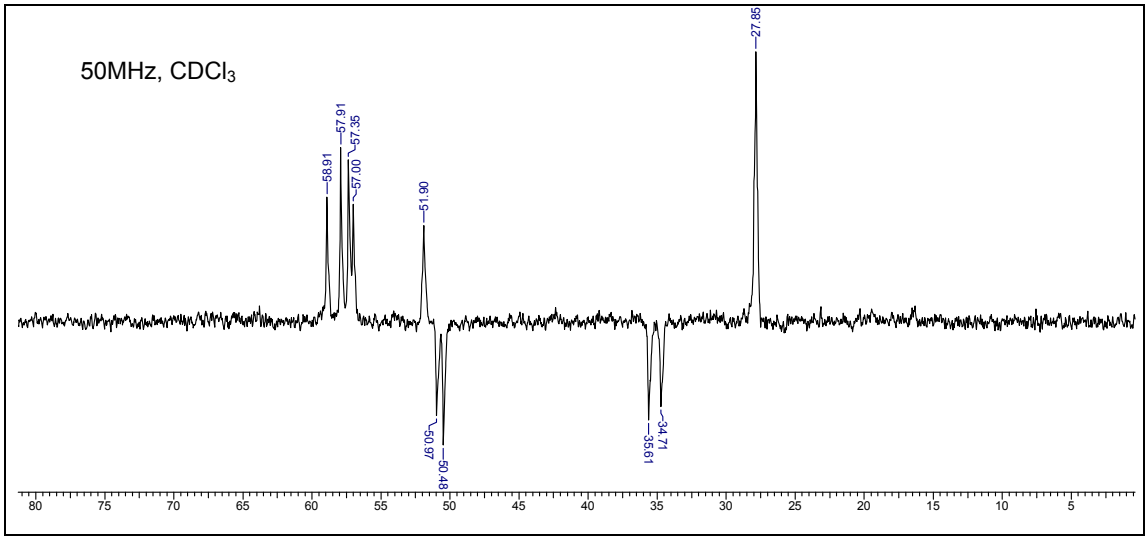
^{1}H NMR of compound 9



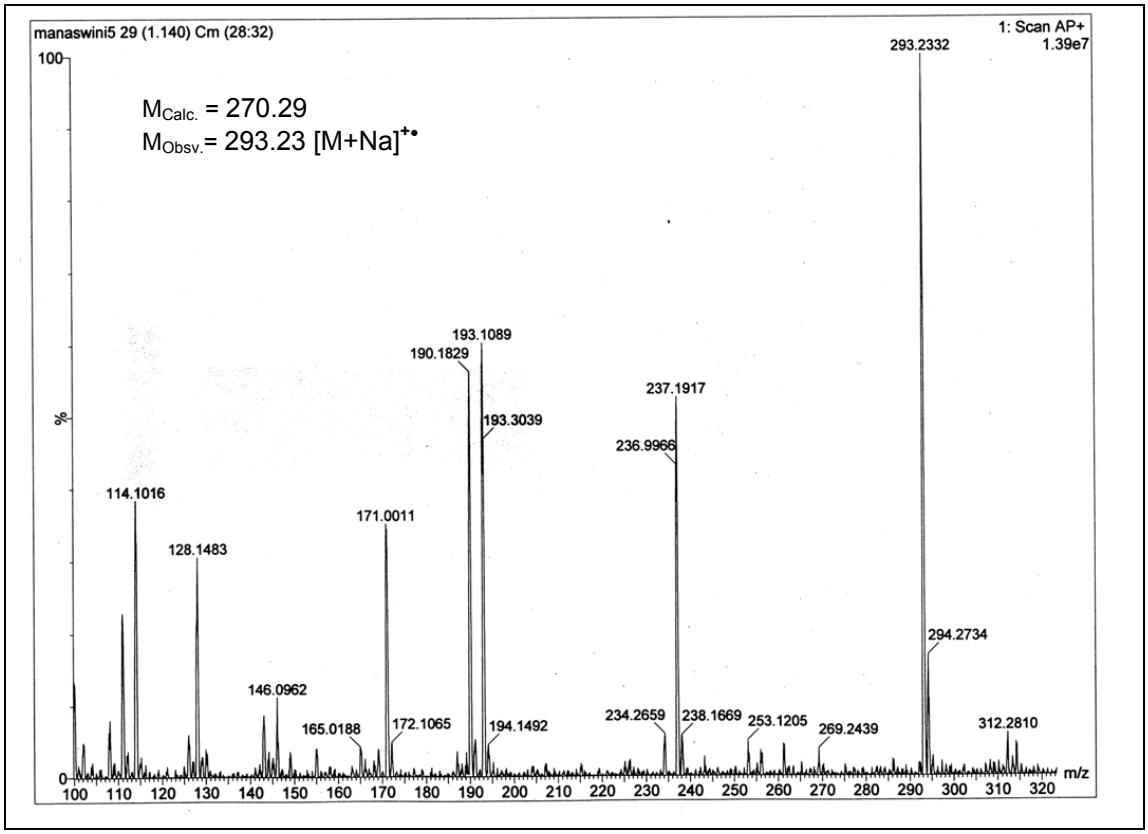
^{13}C NMR of compound 9



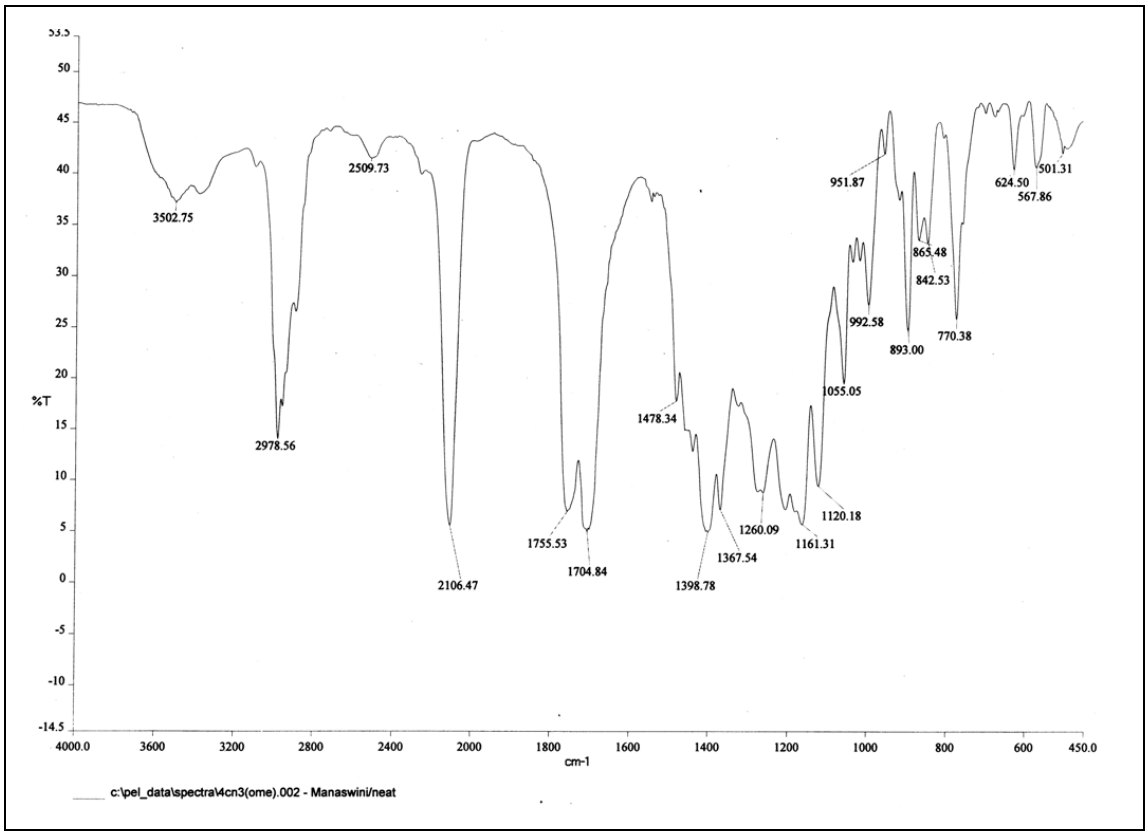
DEPT-^{13}C NMR of compound 9



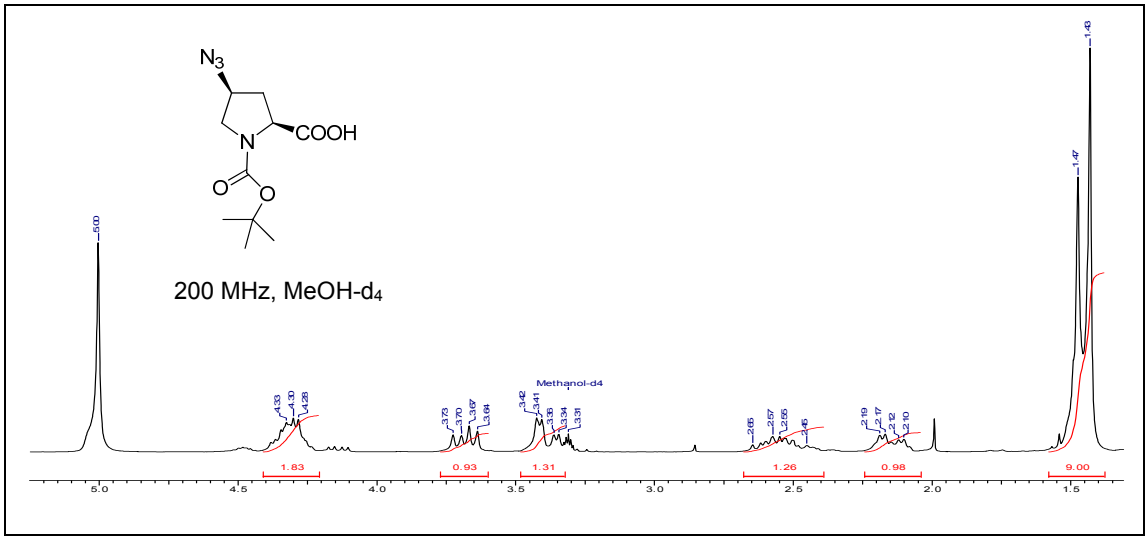
LC-MS of compound 9



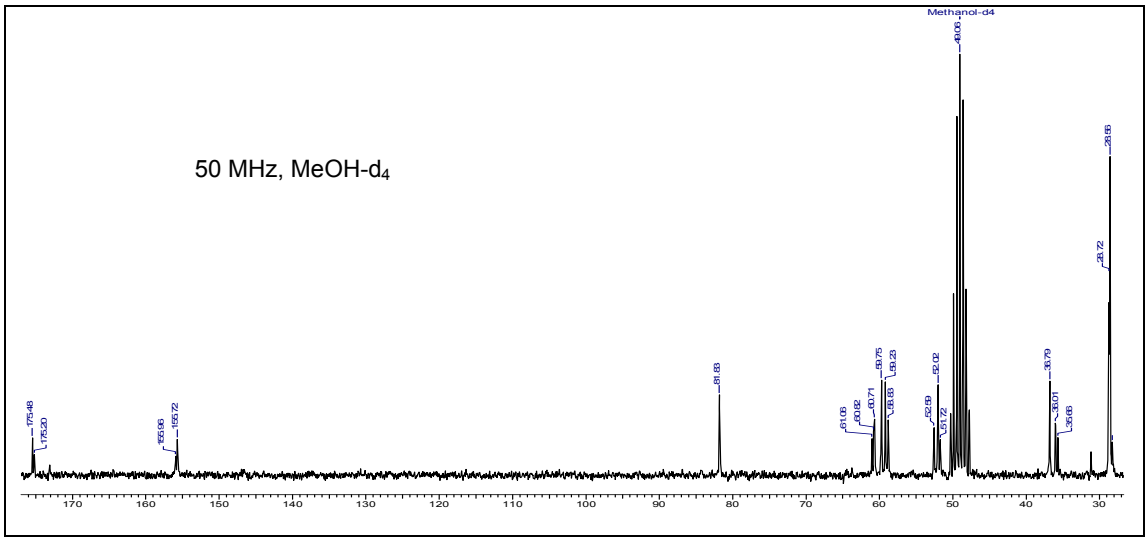
IR Spectra of compound 9



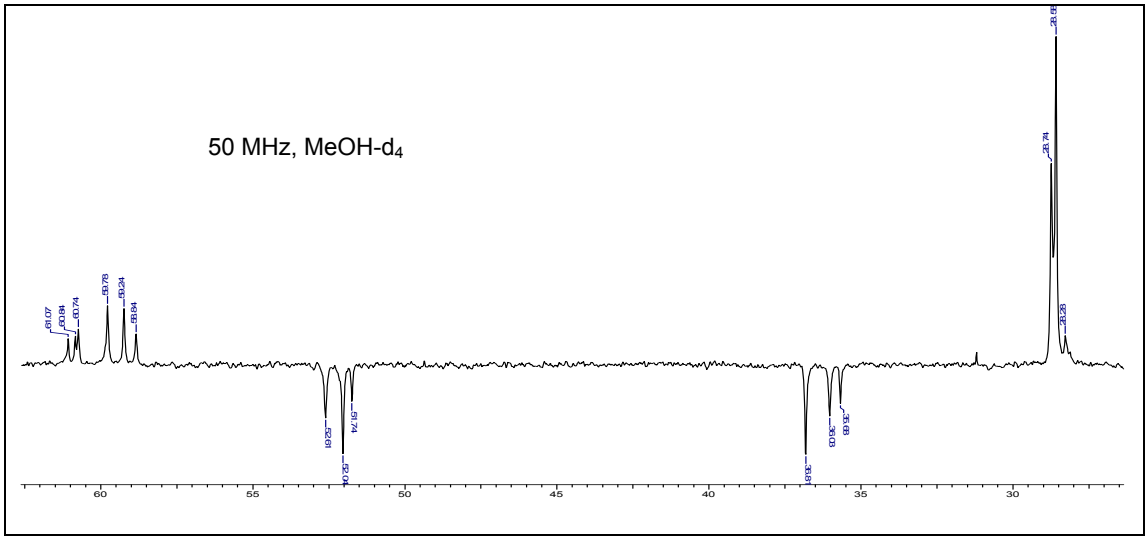
^{1}H NMR of compound 10



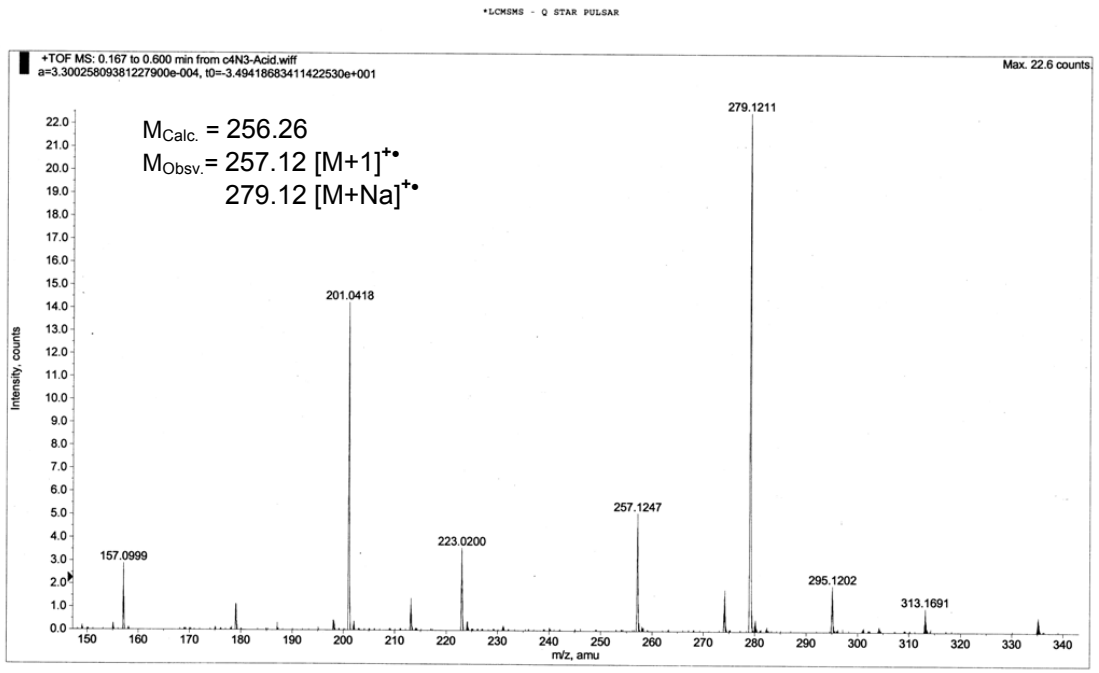
^{13}C NMR of compound 10



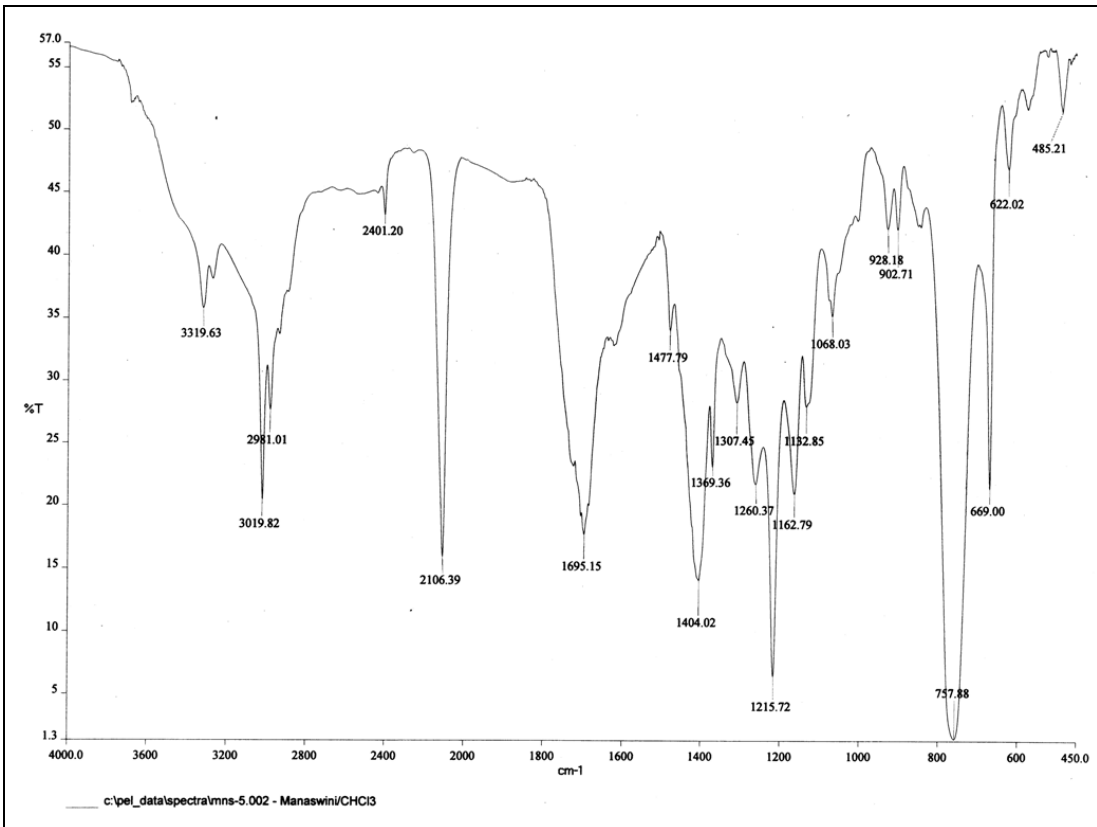
DEPT-^{13}C NMR of compound 10



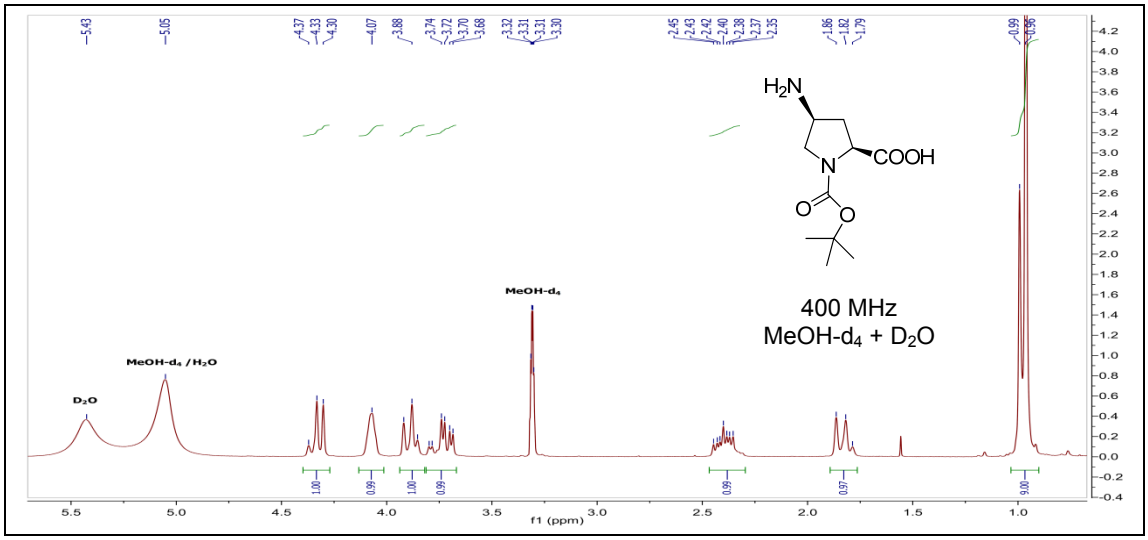
LC-MS of compound 10



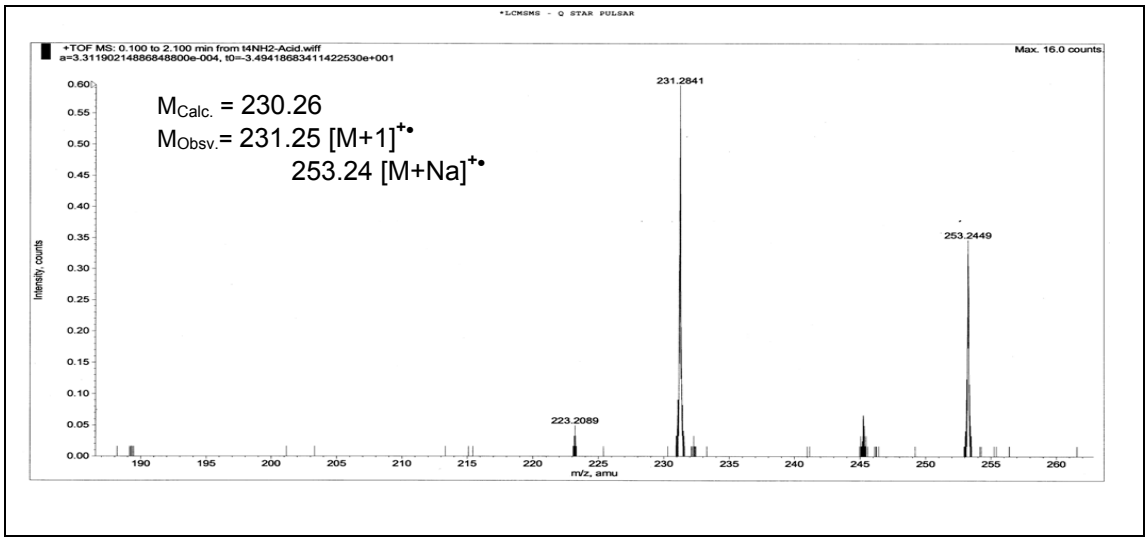
IR Spectra of compound 10



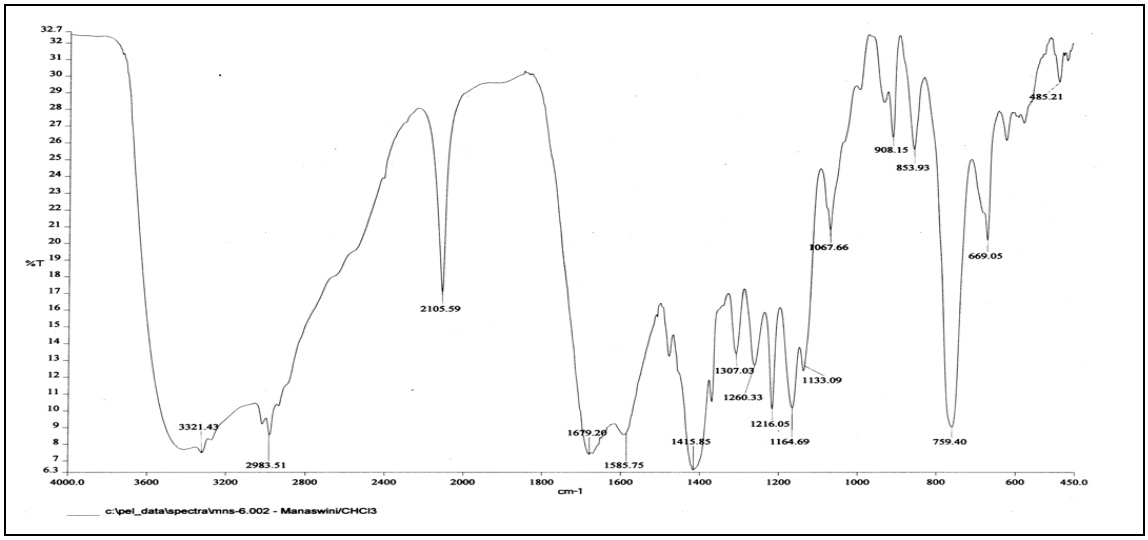
^{1}H NMR of compound 11



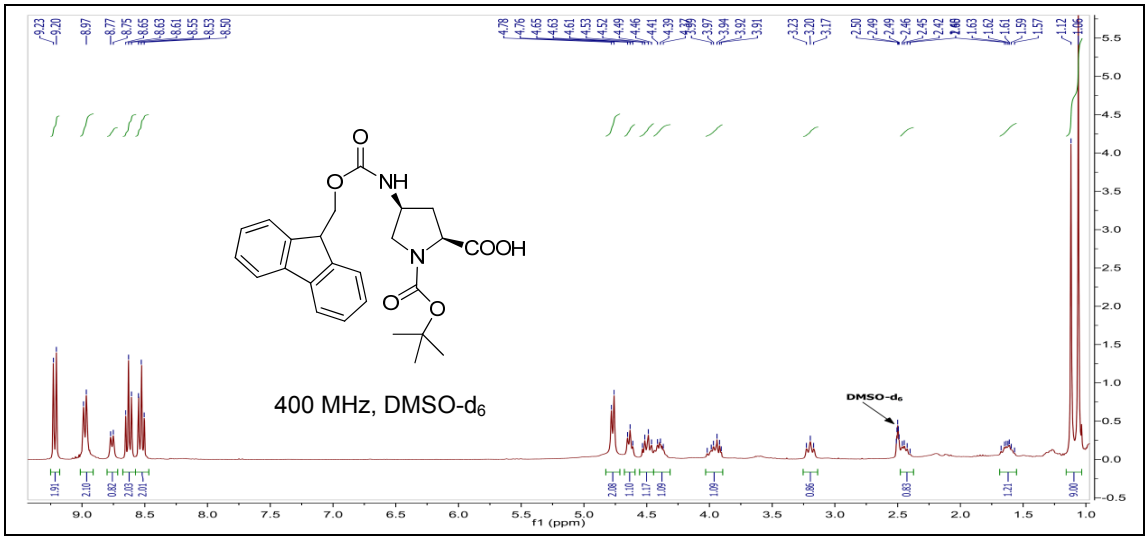
LC-MS of compound 11



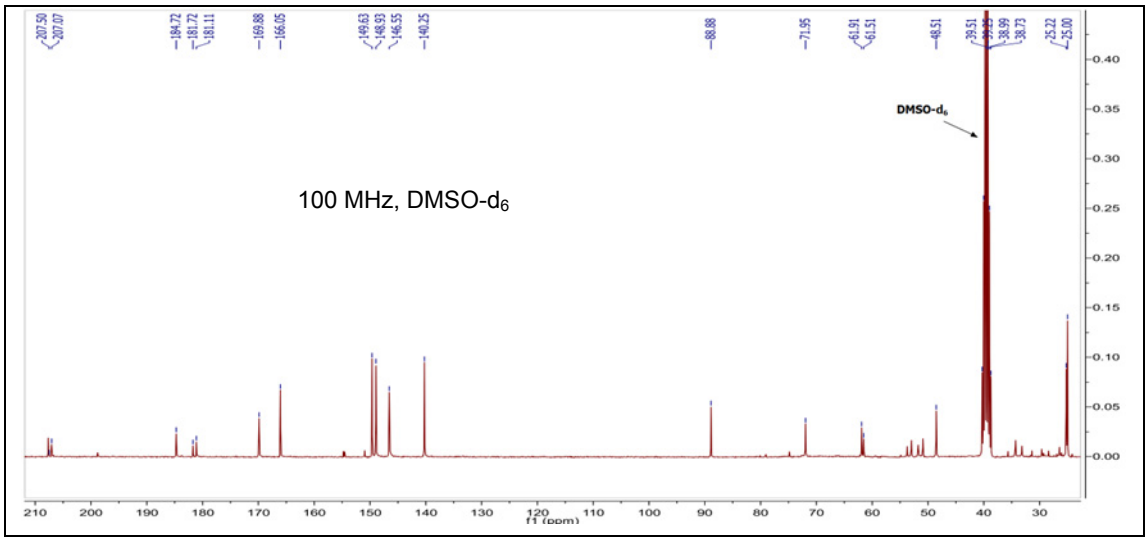
IR Spectra of compound 11



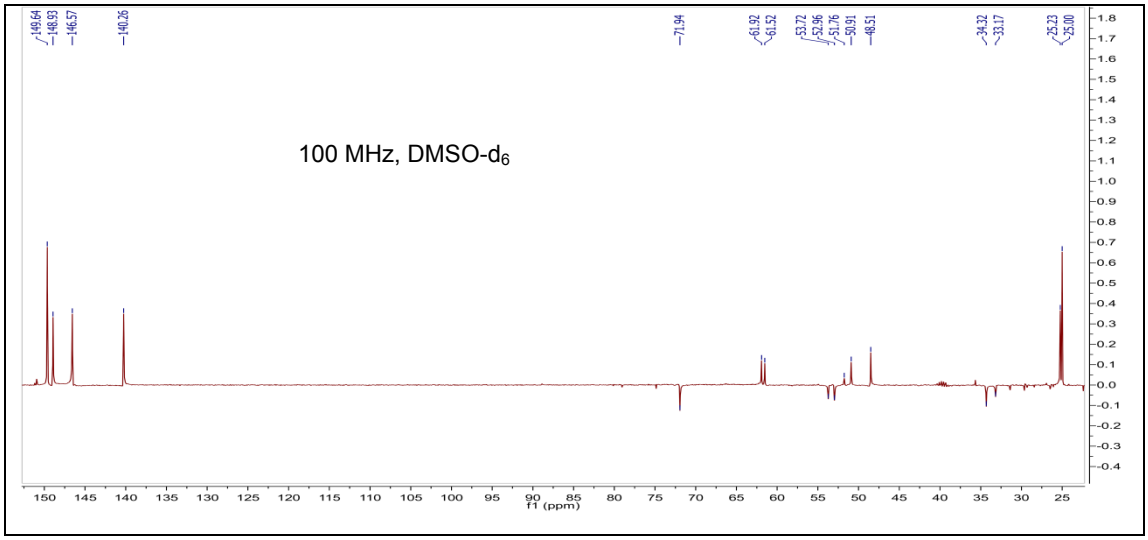
^{1}H NMR of compound 12



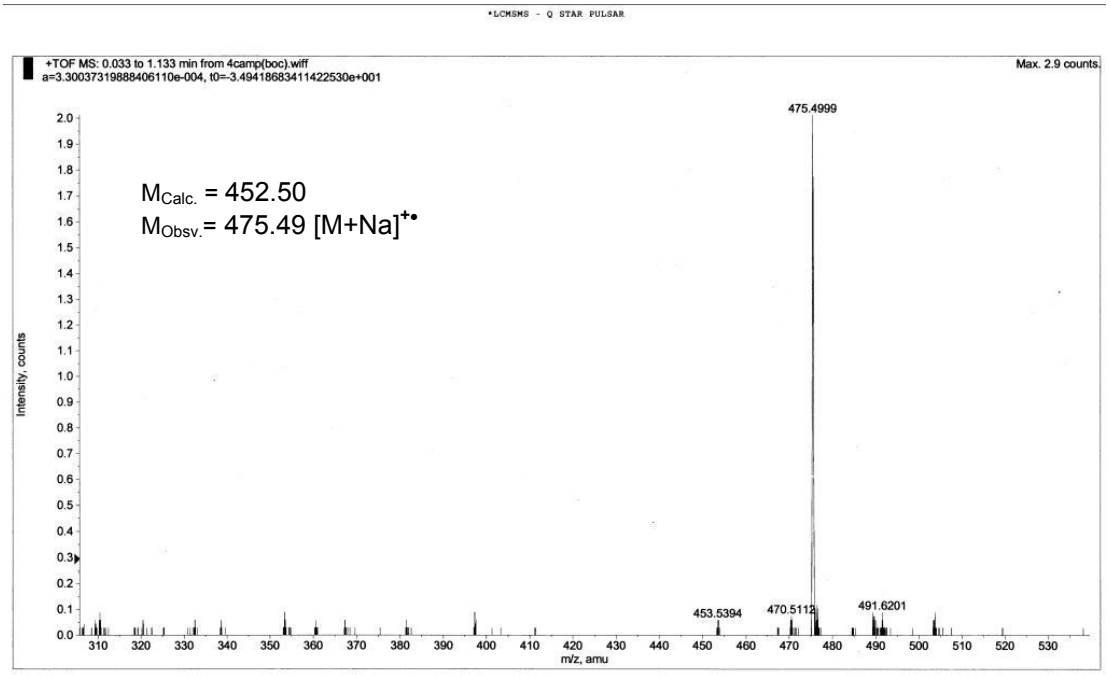
^{13}C NMR of compound 12



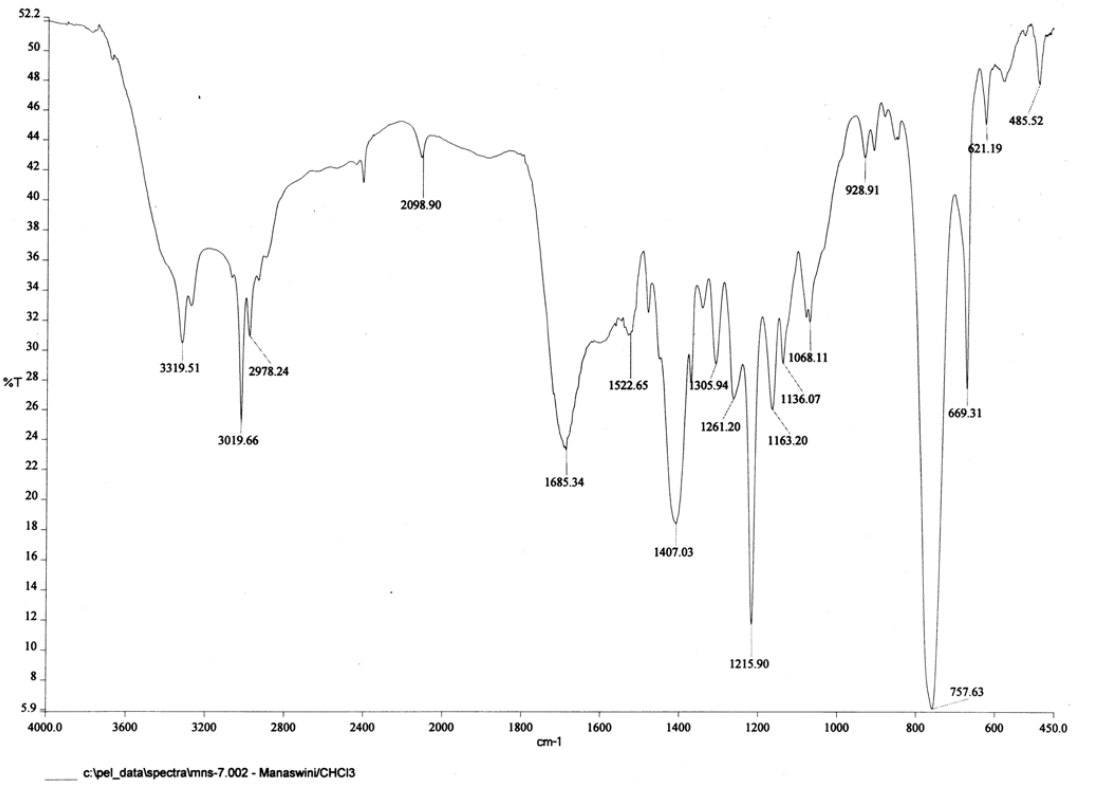
DEPT-^{13}C NMR of compound 12



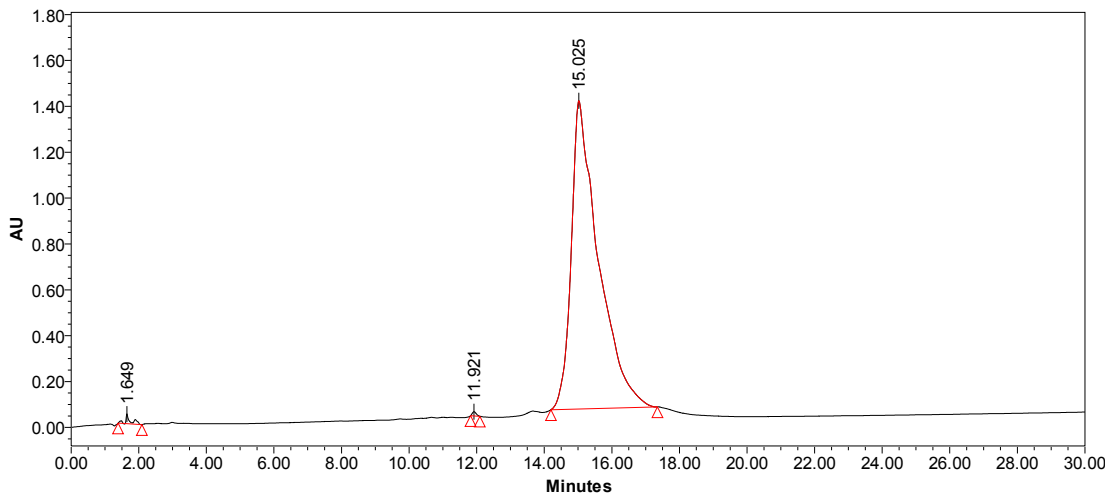
LC-MS of compound 12



IR Spectra of compound 12



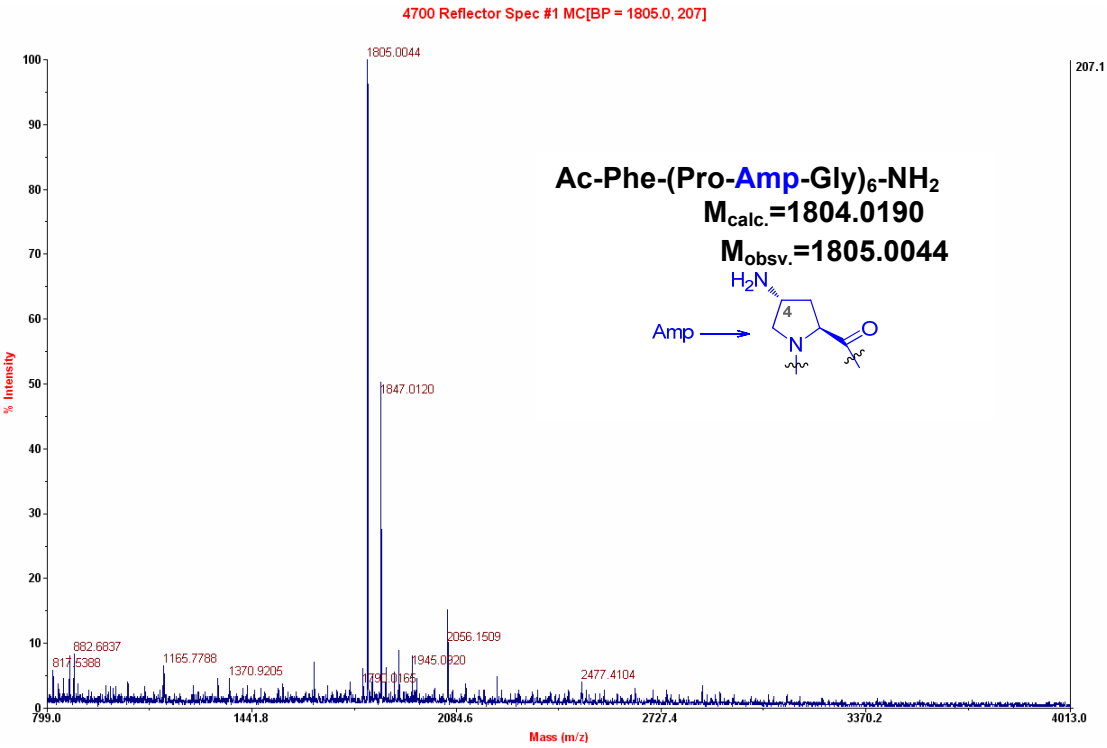
HPLC of Peptide P1 (R-Am)



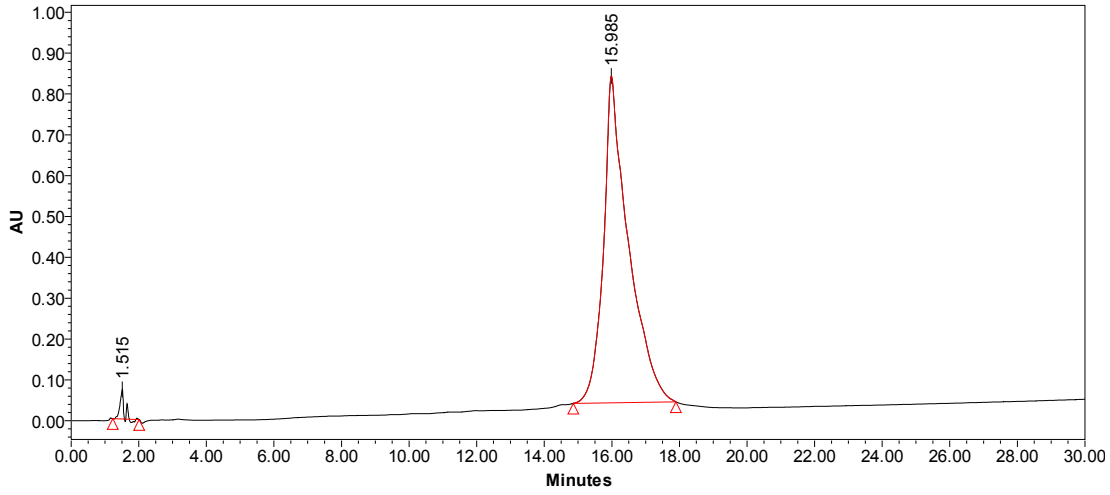
Name	Retention Time	% Area	Height
	1.649	0.56	41554
	11.921	0.19	18251
R-Am	**15.025**	**99.25**	**1344057**

30 min Gradient: 5% to 50% MeCN/H_2O with 0.1% TFA, flow rate = 1.5 ml/min; λ = 220 nm

MALDI-TOF of Peptide P1 (R-Am)



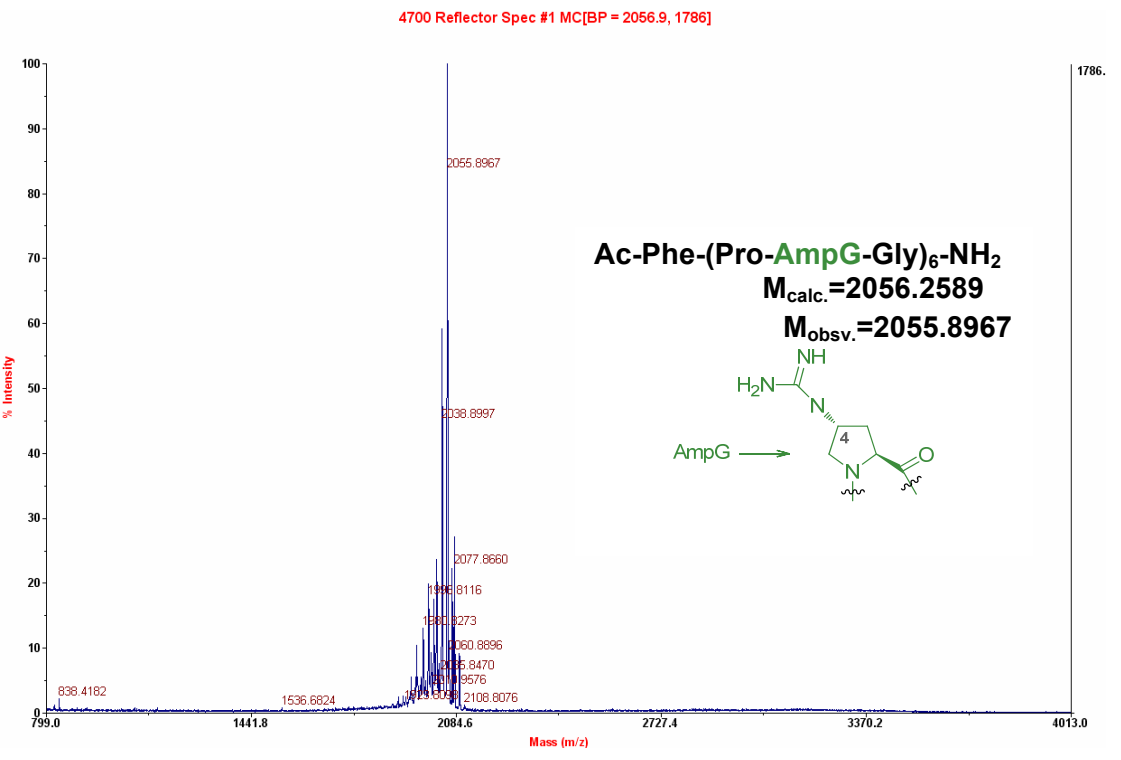
HPLC of Peptide P2 (RG)



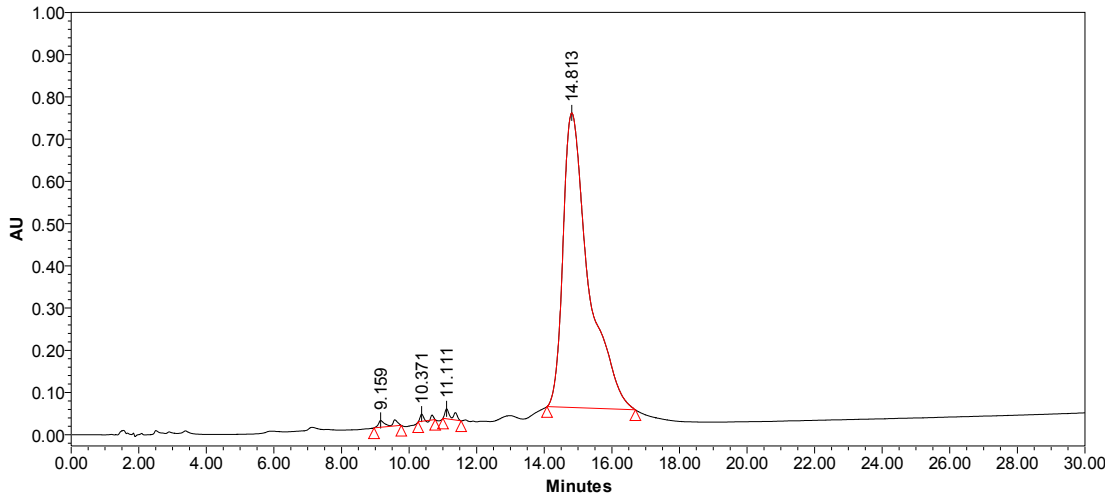
Name	Retention Time	% Area	Height
	1.515	1.61	71983
RG	15.985	98.39	799890

30 min Gradient: 5% to 50% MeCN/H_2O with 0.1% TFA, flow rate = 1.5 ml/min; λ = 220 nm

MALDI-TOF of Peptide P2 (RG)



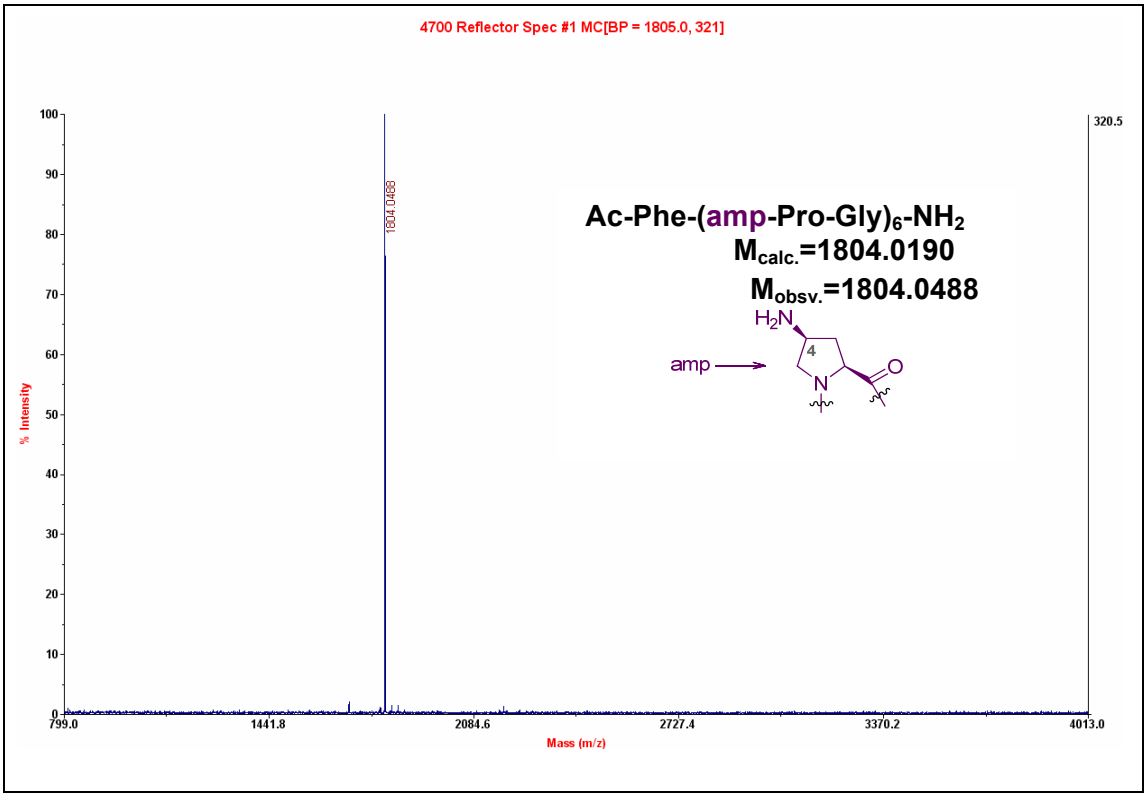
HPLC of Peptide P3 (S-am)



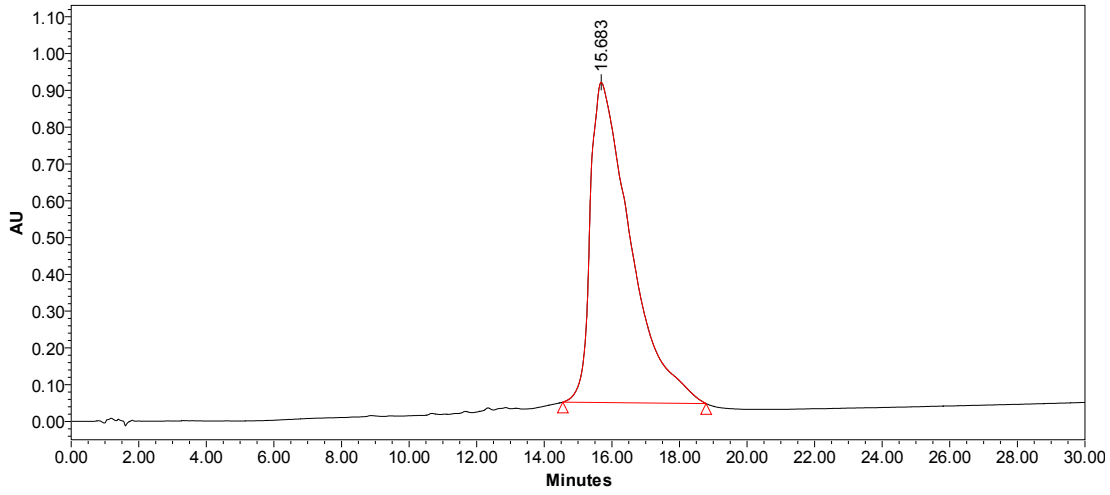
Name	Retention Time	% Area	Height
	9.159	0.89	15946
	10.371	0.53	17195
	11.111	0.87	23555
S-am	**14.813**	**97.70**	**697704**

30 min Gradient: 5% to 50% MeCN/H_2O with 0.1% TFA, flow rate = 1.5 ml/min; λ = 220 nm

MALDI-TOF of Peptide P3 (S-am)



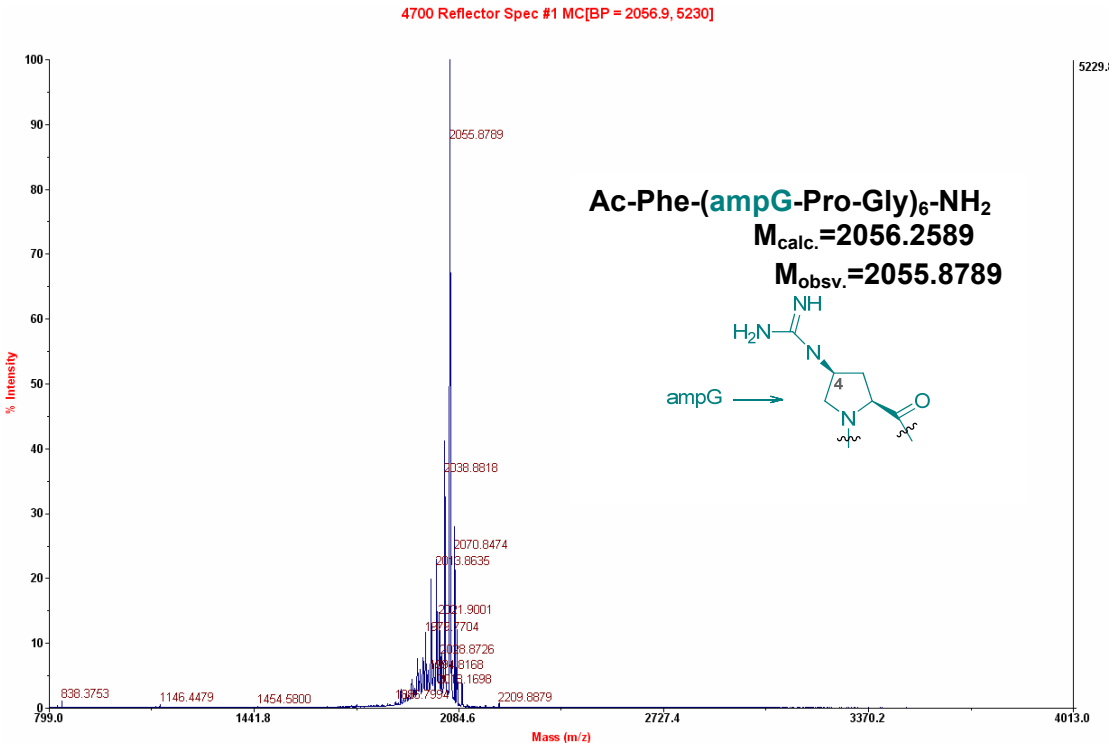
HPLC of Peptide P4 (SG)



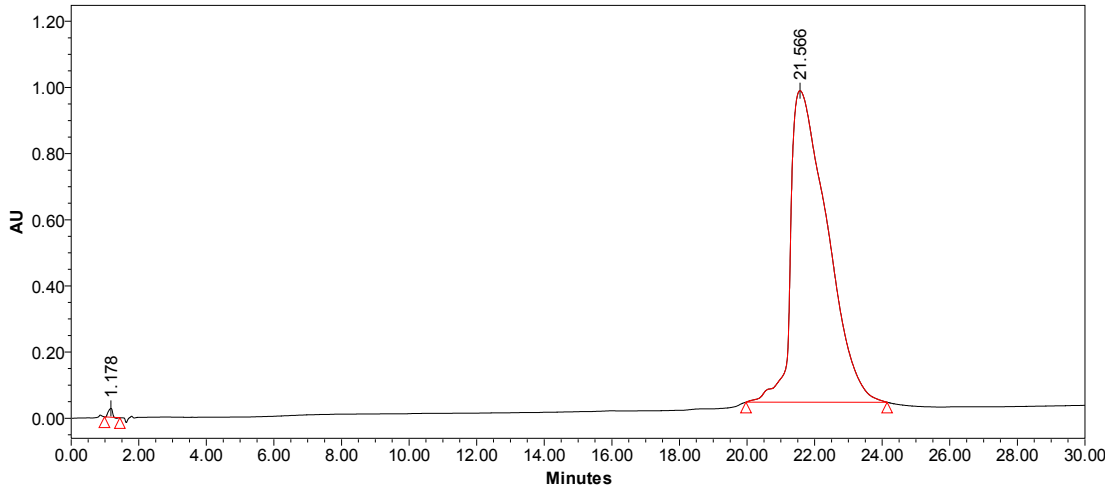
Name	Retention Time	% Area	Height
SG	15.683	100.00	870320

30 min Gradient: 5% to 50% MeCN/H_2O with 0.1% TFA, flow rate = 1.5 ml/min; λ = 220 nm

MALDI-TOF of Peptide P4 (SG)



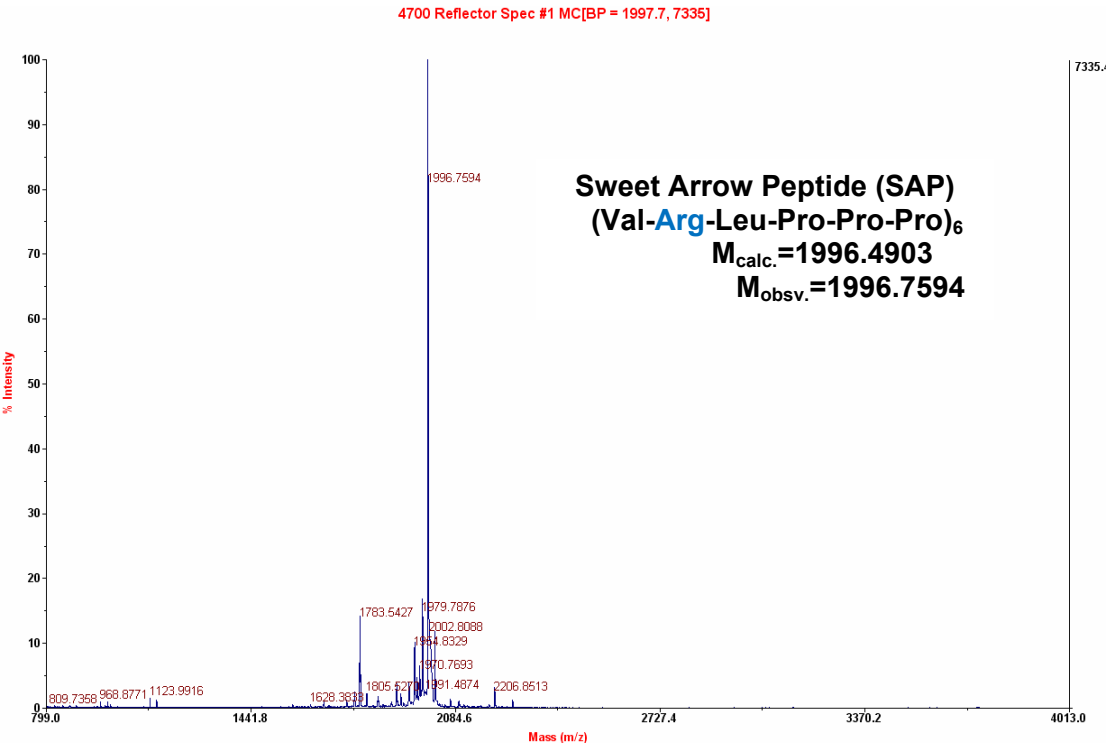
HPLC of Peptide P5 (SAP)



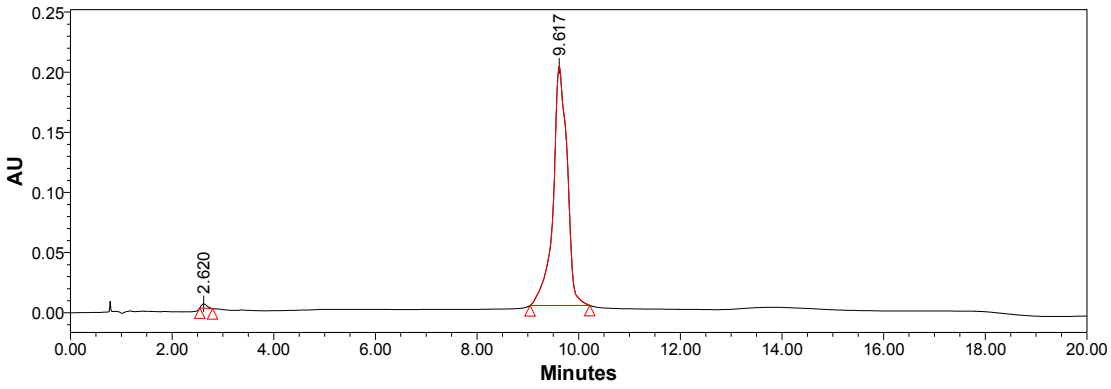
Name	Retention Time	% Area	Height
	1.178	0.33	26963
SAP	21.566	99.67	941885

30 min Gradient: 5% to 50% MeCN/H_2O with 0.1% TFA, flow rate = 1.5 ml/min; λ = 220 nm

MALDI-TOF of Peptide P5 (SAP)



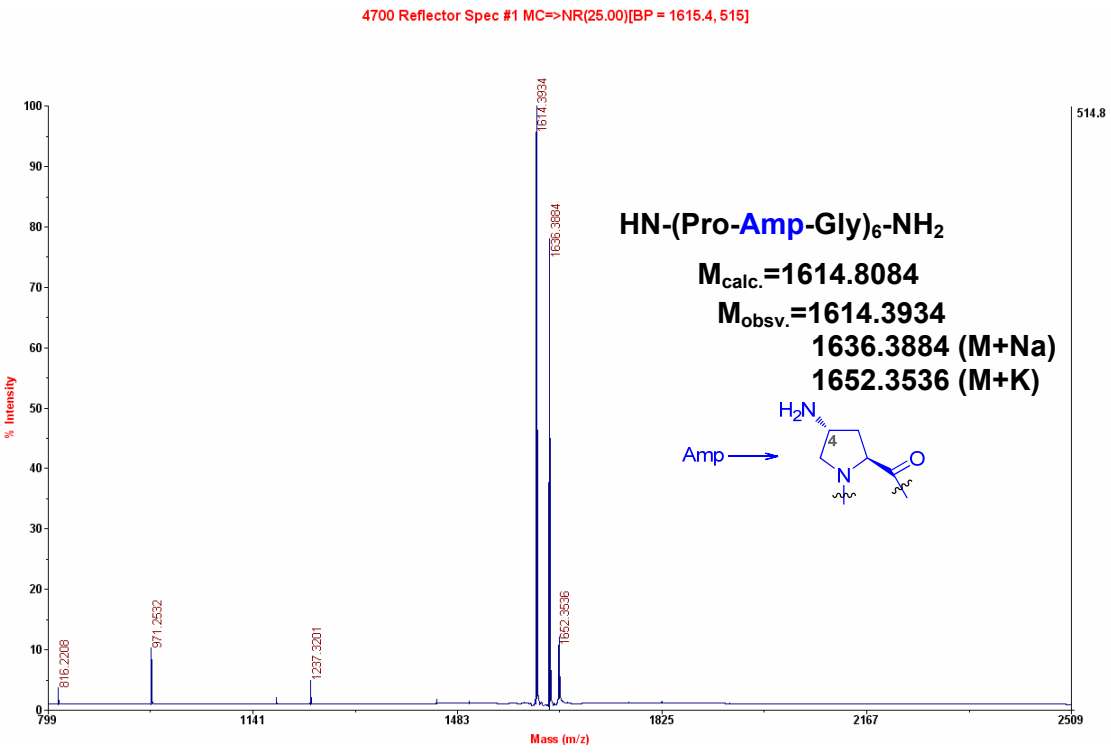
HPLC of Peptide P6 (RA)



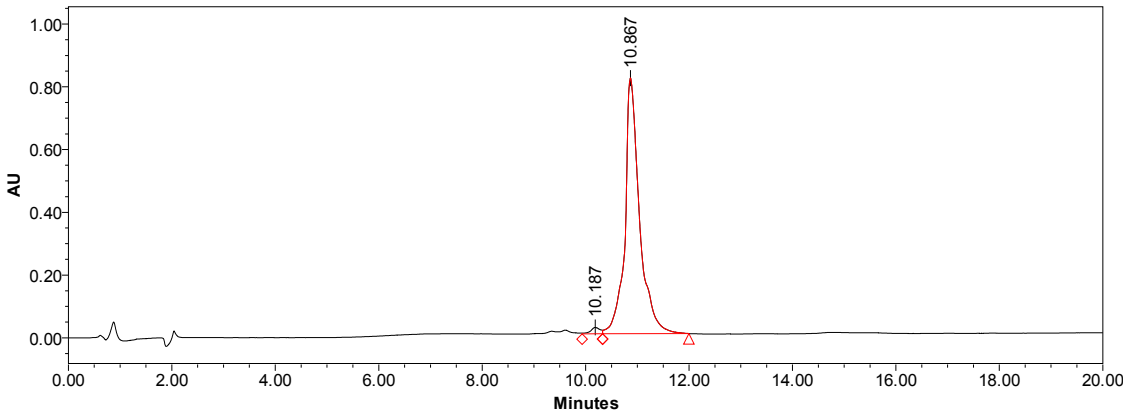
Name	Retention Time	% Area	Height
	2.620	0.68	3489
RA	9.617	99.32	199042

20 min Gradient: 5% to 35% MeCN/H_2O with 0.1% TFA, flow rate = 1.5 ml/min; λ = 220 nm

MALDI-TOF of Peptide P6 (RA)



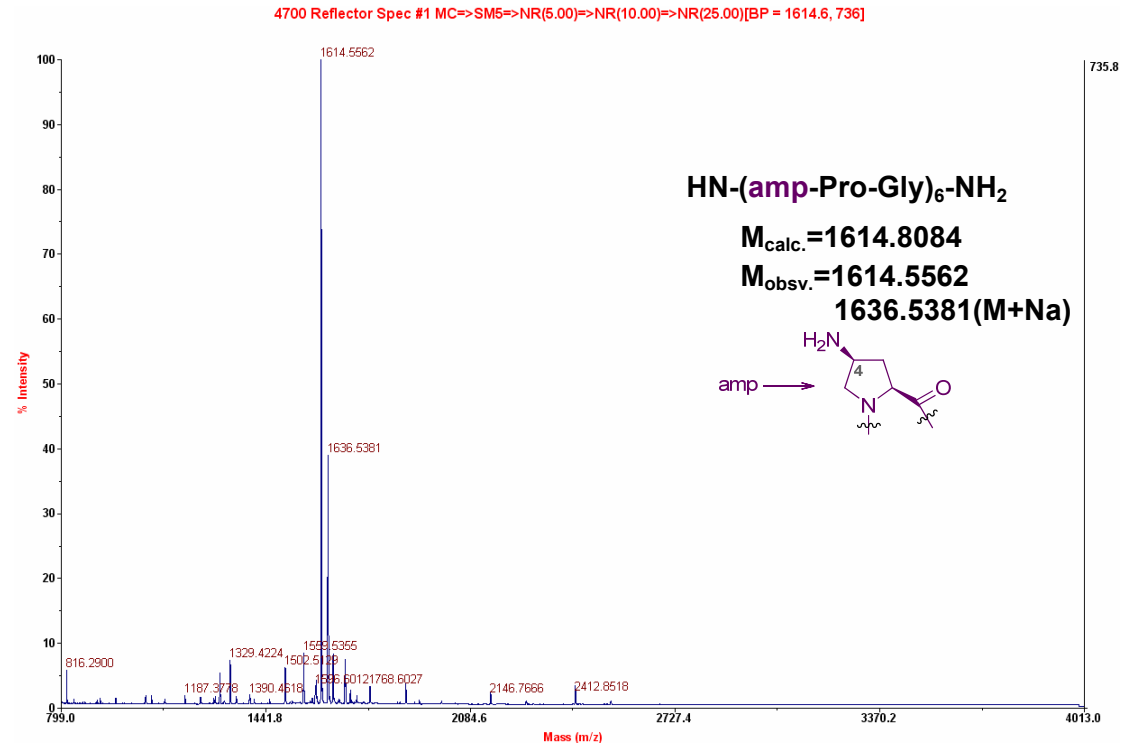
HPLC of Peptide P7



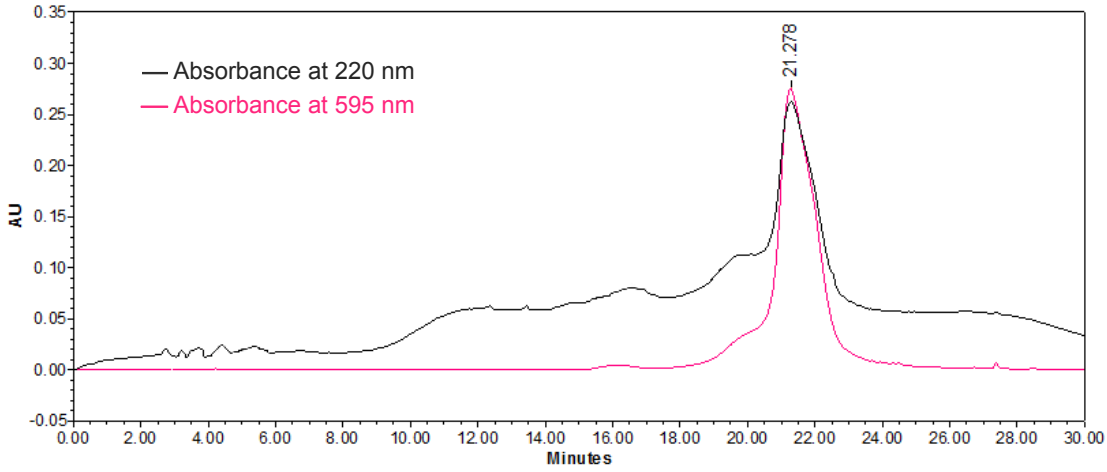
Name	Retention Time	% Area	Height
	10.187	1.60	20600
SA	**10.867**	**98.40**	**815011**

20 min Gradient: 5% to 35% MeCN/H_2O with 0.1% TFA, flow rate = 1.5 ml/min; λ = 220 nm

MALDI-TOF of Peptide P7 (SA)

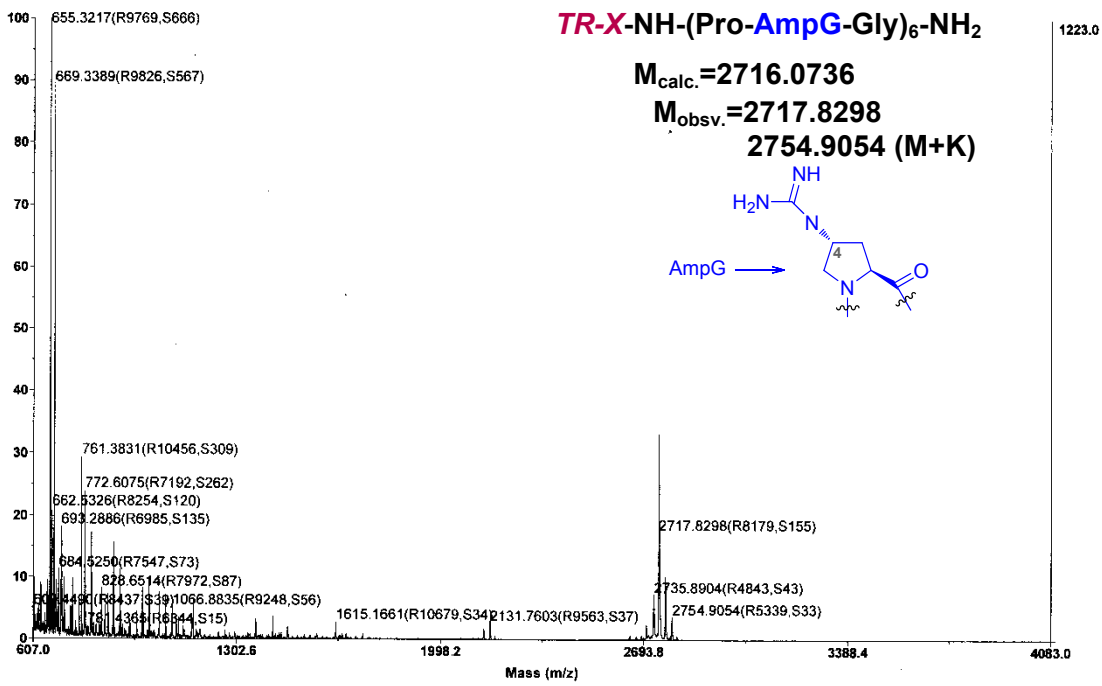


HPLC of Peptide P8 (RG-F)

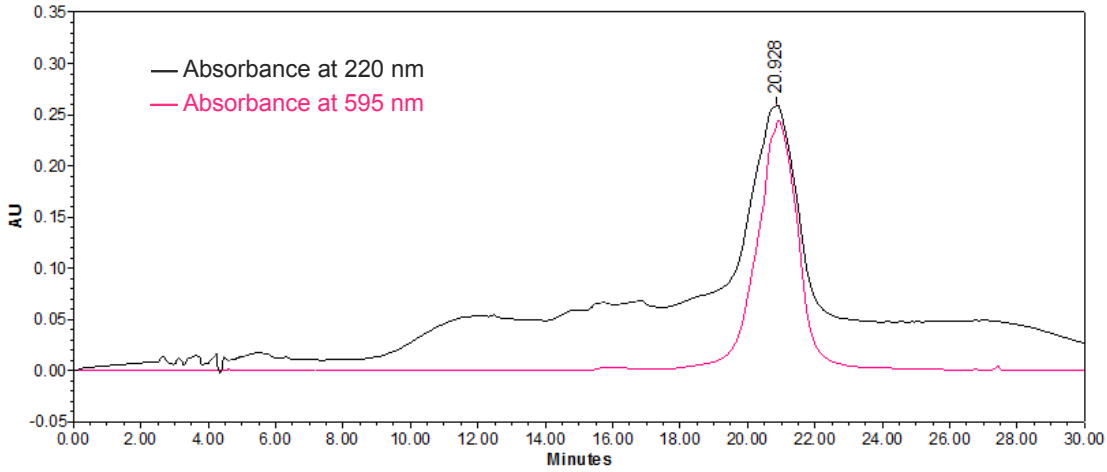


30 min Gradient: 5% to 50% MeCN/H_2O with 0.1% TFA in 1st 20 min then 50% to 5% next 10 min, flow rate = 1.0 ml/min; λ = 220 nm & 595 nm

MALDI-TOF of Peptide P8 (RG-F)

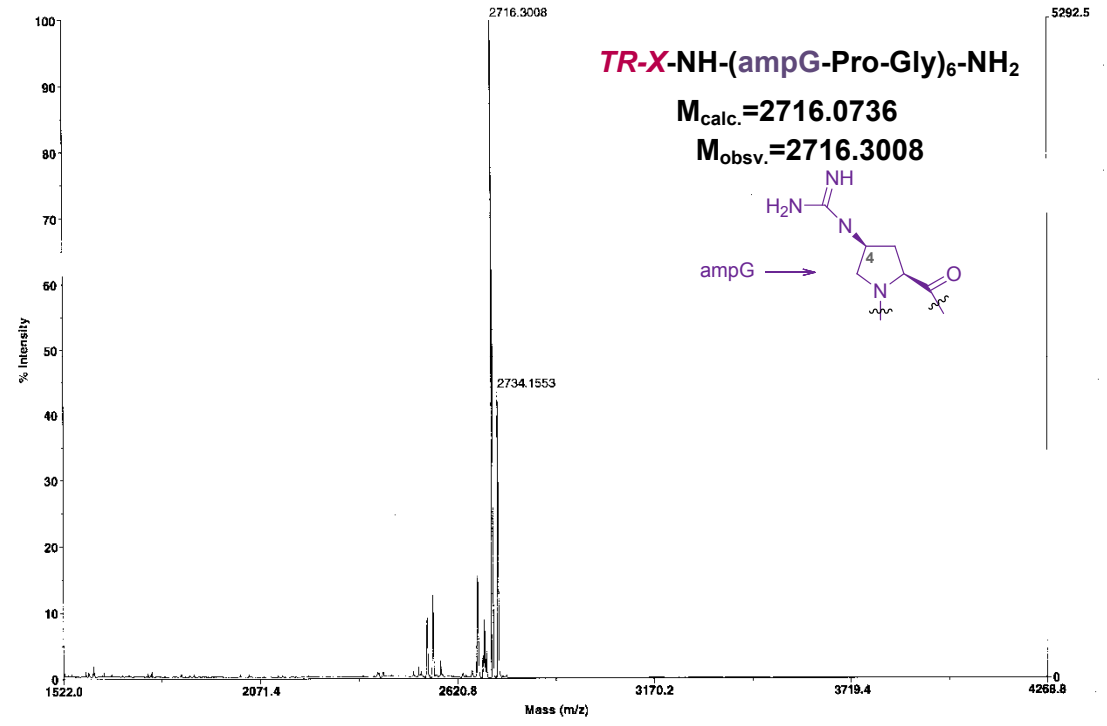


HPLC of Peptide P9 (SG-F)

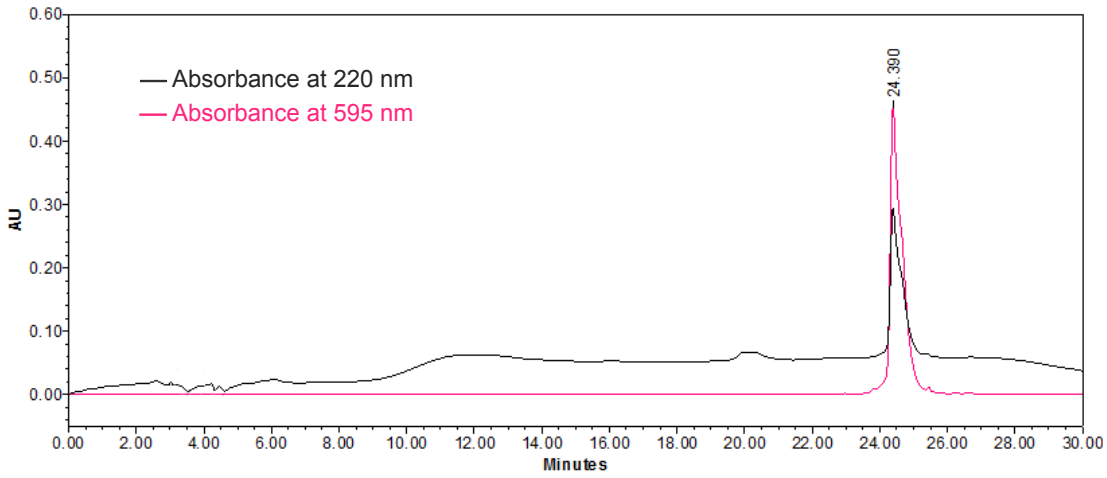


30 min Gradient: 5% to 50% MeCN/H_2O with 0.1% TFA in 1st 20 min then 50% to 5% next 10 min, flow rate = 1.0 ml/min; λ = 220 nm & 595 nm

MALDI-TOF of Peptide P9 (SG-F)

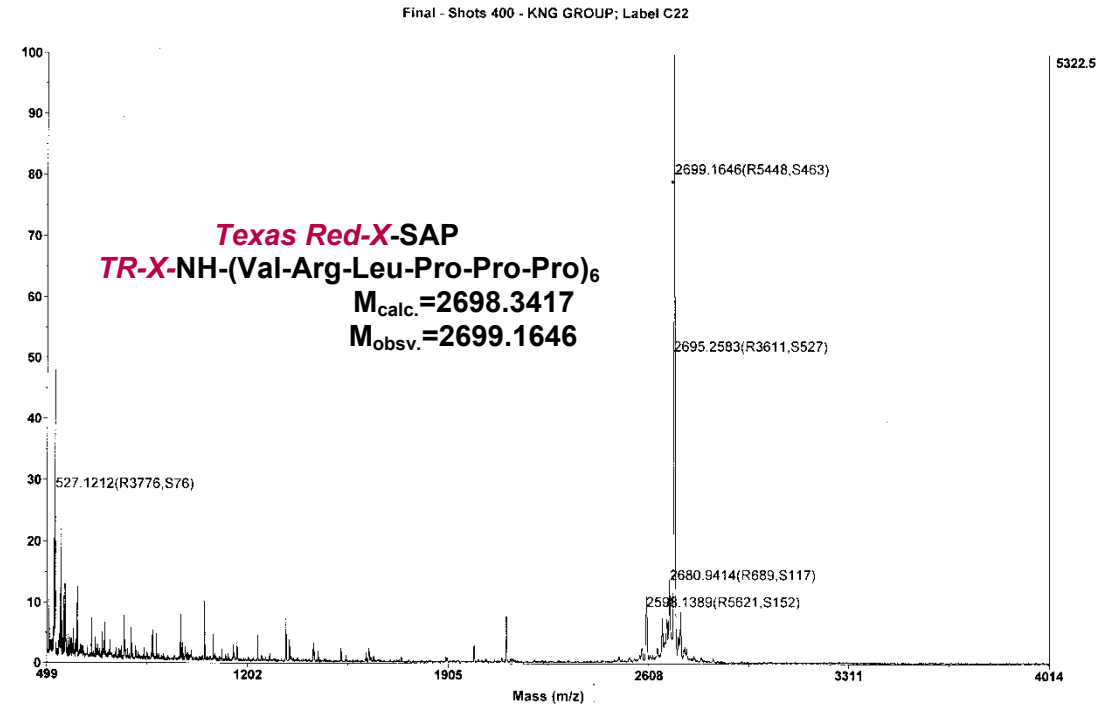


HPLC of Peptide P10 (SAP-F)

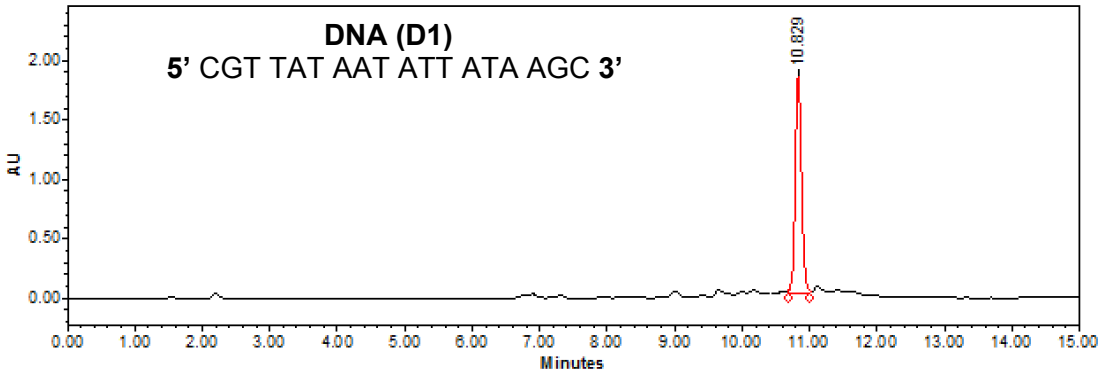


30 min Gradient: 5% to 50% MeCN/H_2O with 0.1% TFA in 1st 20 min then 50% to 5% next 10 min, flow rate = 1.0 ml/min; λ = 220 nm & 595 nm

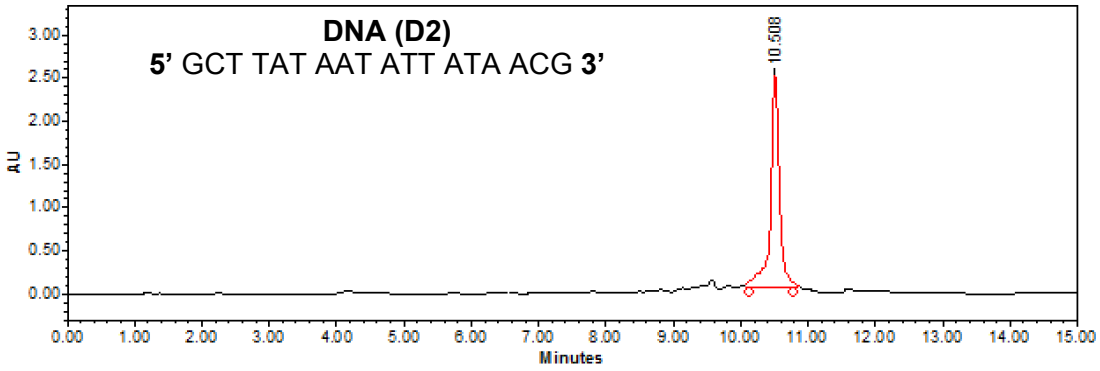
MALDI-TOF of Peptide P10 (SAP-F)



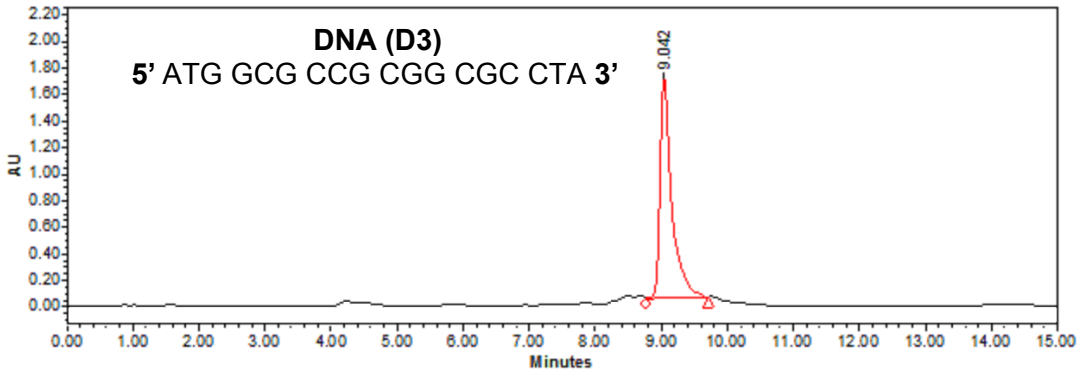
HPLC of Oligonucleotide D1



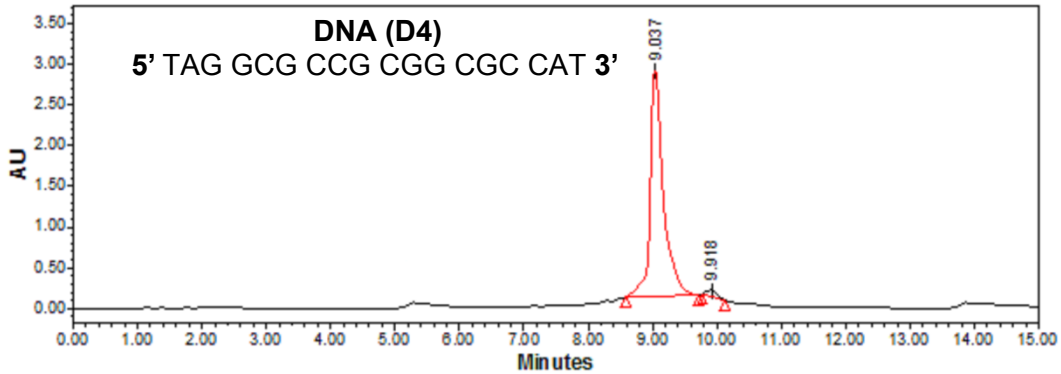
HPLC of Oligonucleotide D2



HPLC of Oligonucleotide D3

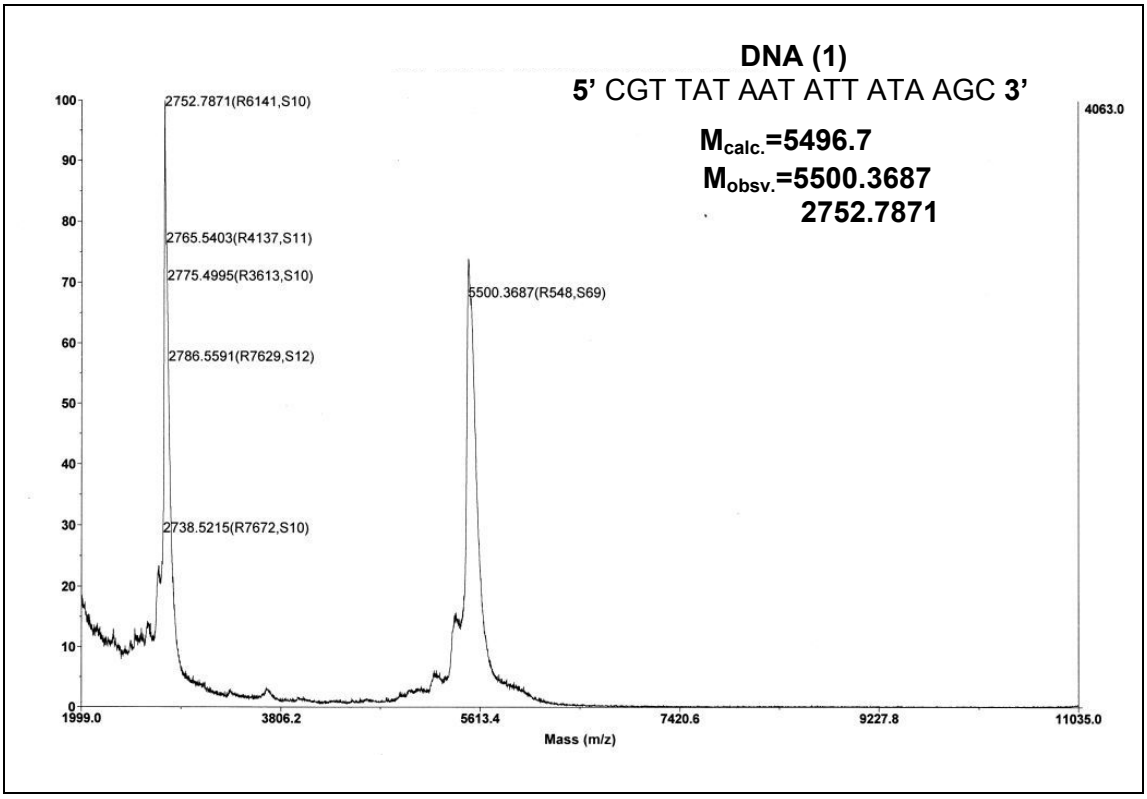


HPLC of Oligonucleotide D4

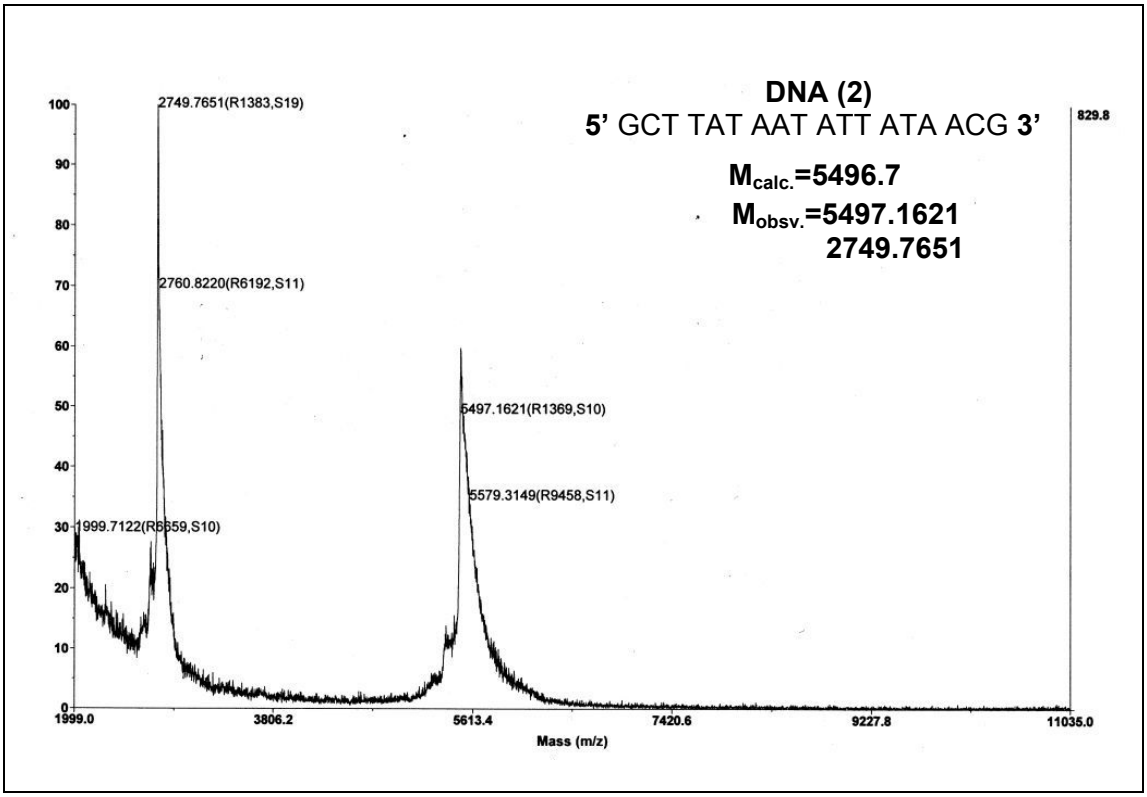


15 min Gradient: 5% to 30% MeCN/TEAA, flow rate = 2 ml/min; λ = 260 nm

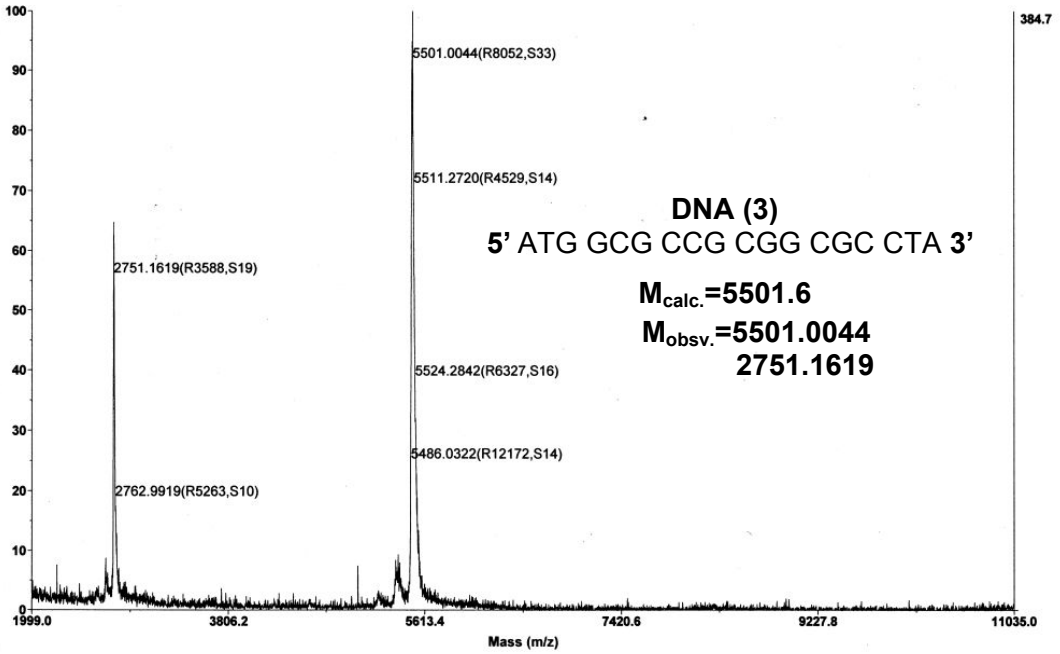
MALDI-TOF of Oligonucleotide D1



MALDI-TOF of Oligonucleotide D2



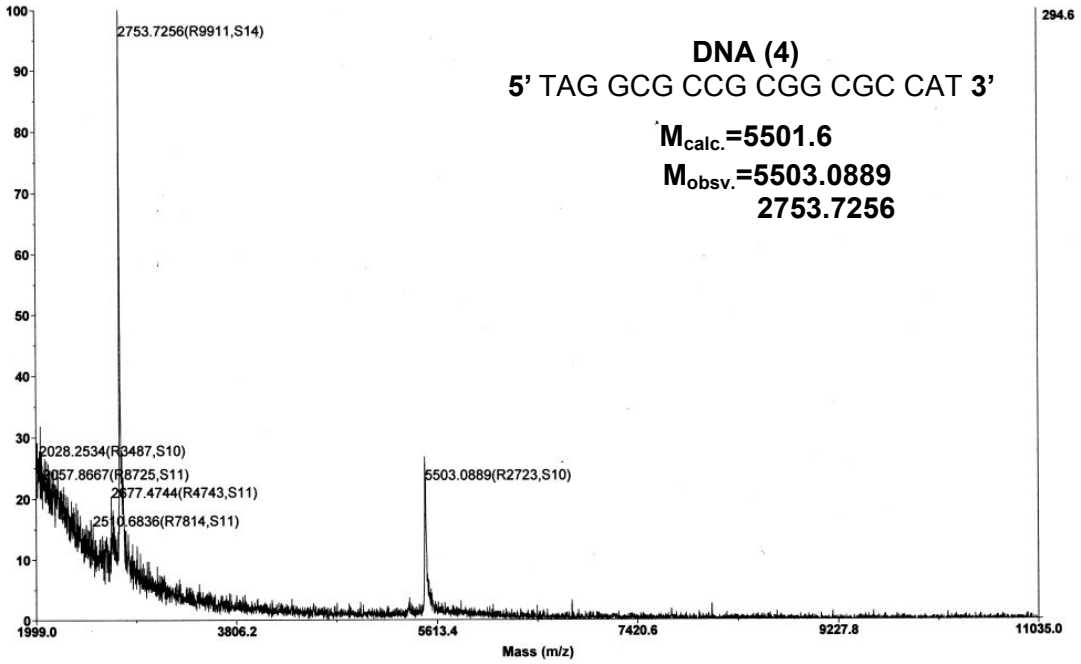
MALDI-TOF of Oligonucleotide D3



DNA (3)
5' ATG GCG CCG CGG CGC CTA **3'**

$M_{calc.}$=5501.6
$M_{obsv.}$=5501.0044
2751.1619

MALDI-TOF of Oligonucleotide D4

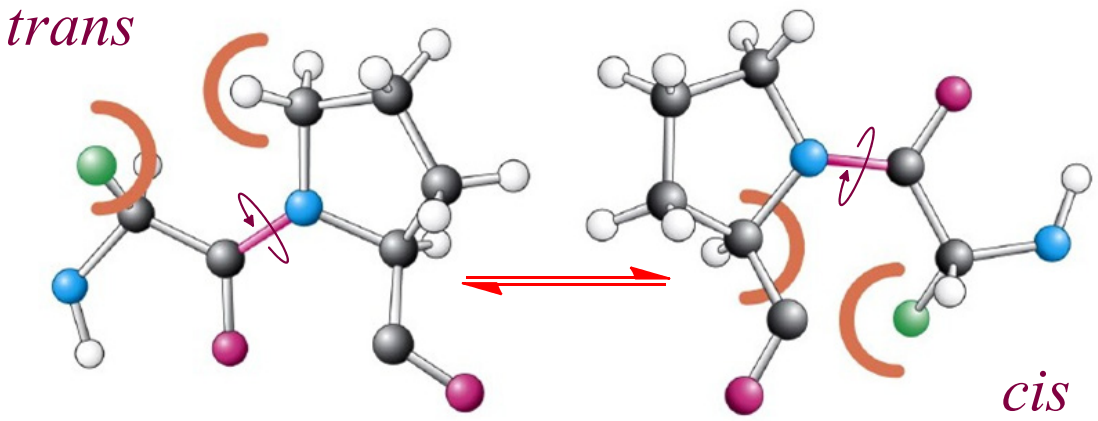


DNA (4)
5' TAG GCG CCG CGG CGC CAT **3'**

$M_{calc.}$=5501.6
$M_{obsv.}$=5503.0889
2753.7256

Section II

NMR Studies of *cis:trans* Isomerisation in Prolyl Peptide Bond with respect to Substitution at 4-position of Proline in Ala-Pro based Dipeptides



This section starts with introduction to occurance and properties of 4-substituted prolines and their peptides, with a focus on the likely importance of 4-substitution on prolyl-peptide bond isomerization. It majorly discusses the NMR investigation done to explore the effect of stereochemistry and pH on overall *cis-trans* orientation selectivity of those molecules, backed by crystal structure, CD and pK_a studies. Ala-Pro based dipeptides were chosen as best model for undertaking the research work .

2.1 Introduction

Proteins are long chain polypeptides that adopt many different conformations; yet, the sequence of their amino acid residues directs the folding to a particular native state conformation.[1] Polypeptide conformations can be described in terms of 3 main chain torsion angles; (i) the torsion angle about C^{α} -N σ-bond Φ, (ii) the angle about the σ-bond between carbonyl group and C^{α} Ψ, and (iii) the angle about the amide bond ω (Figure **1.A**). The σ-bonds (except in the imino acid proline) are relatively flexible and the preferred values for Φ and Ψ angles depend primarily on the nature of the α-substituent. Allowed values for Φ and Ψ are graphically revealed when Φ is plotted *versus* Ψ in a Ramachandran plot (Figure **1.B**), introduced by G. N. Ramachandran.[2] The torsional angles of each residue in a peptide define the

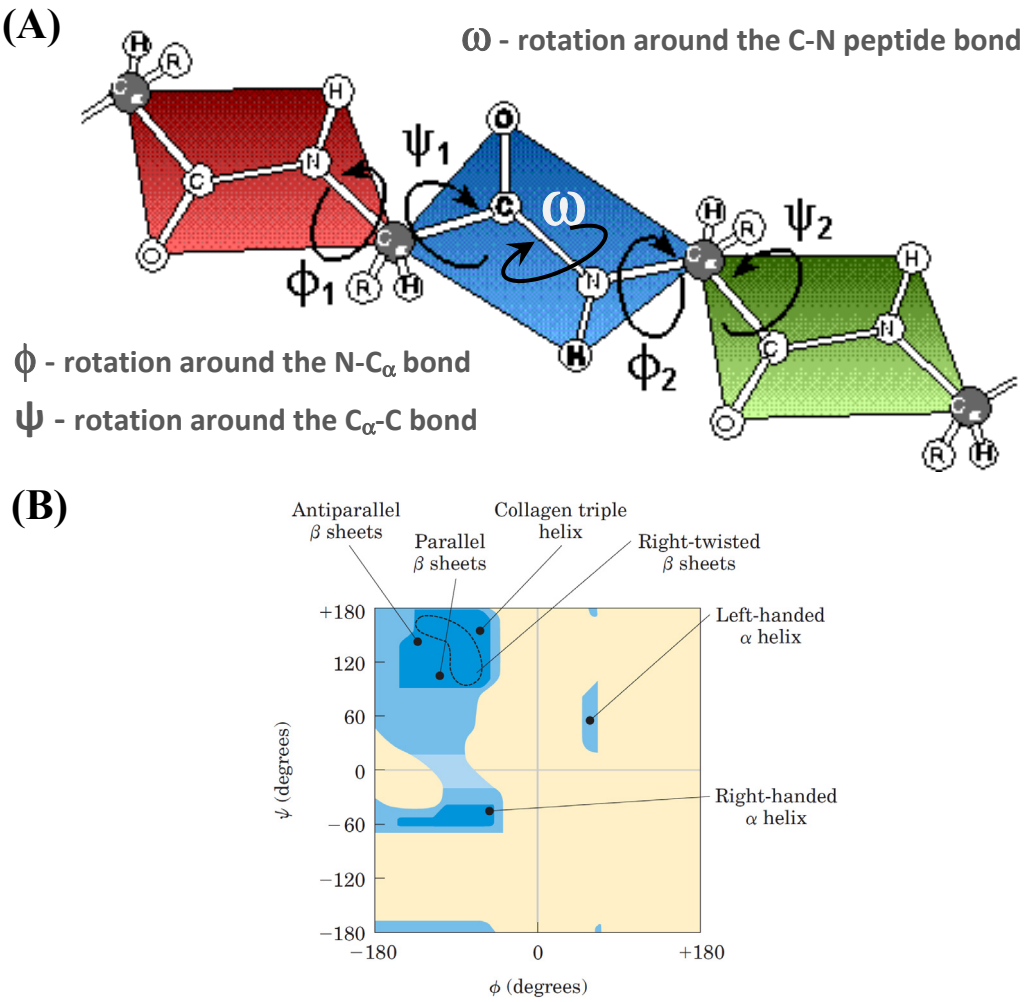


FIGURE 1: (**A**) Illustration of peptide plane and the torsion angles therein, and (**B**) Ramachandran plot for variety of peptide structures

geometry of its attachment to its two adjacent residues by positioning its planar peptide bond relative to the two adjacent planar peptide bonds; thereby the torsional angles determine the conformation of the residues and the peptide as a whole.

On the other hand, α-nitrogen atom possessing a lone pair of electrons participates in resonance with the carbonyl double bond. This resonance phenomenon imparts a partial double bond character to the N-CO bond because of the π–orbital overlapp, leading to planar amide bond (Figure **2.A**). Further, due to the higher energy associated with N-CO partial double bond, the amide bond is rigid and the peptide bond exists in two distinct conformations namely *cis* and *trans* (Figure **2(B-C)**).

The free energy barrier between the energetically favored conformations of a single residue in polypeptide chain, by rotation of Φ and Ψ, are only of the order of 0.5-1.5 kcal mol^{-1} and these rotations occur at rates of the order of 10^{12} s^{-1}. The only

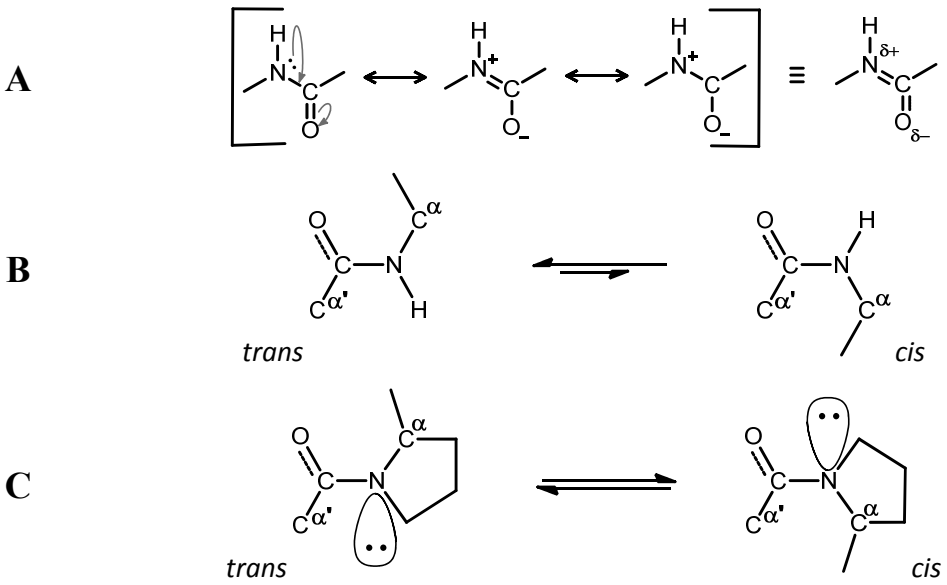


FIGURE 2: (**A**) Resonance phenomenon in amide bond, (**B**) *cis-trans* isomerization in non-proline peptide fragments, (**C**) *cis-trans* conformations in proline fragments; Longer backward arrow indicates that the *trans* isomer is greatly favored over the *cis* isomer while equal length signifies similar preference.

conformational transition in disordered polypeptide chains known to have an intrinsically high free-energy barrier is rotation about the peptide bond, interconverting *cis* (*E*) and *trans* (*Z*) form. This rotation requires disruption of the partially double bonded peptide bond but, the rate of interconversion is not known except for proline.

2.2 Peptidyl-prolyl bond isomerization

Proline is a cyclic imino acid and the bridging of the α-carbon atom to the main chain amide nitrogen by 3 methylene bridge imposes added constraint on the main chain torsion angles Φ and Ψ. Further, pyrrolidine ring conformation of Pro is defined in terms of four endocyclic torsion angles χ_1, χ_2, χ_3 and χ_4 (Figure **3**).

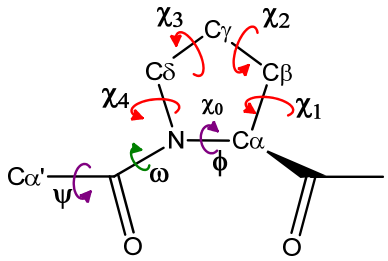
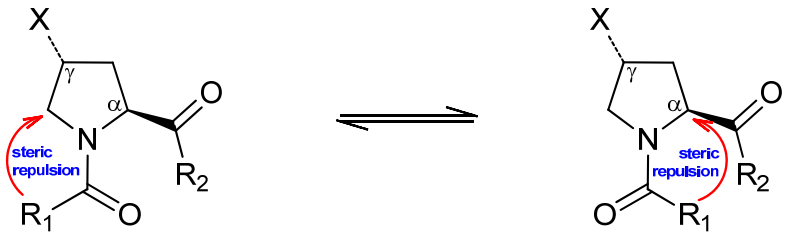


FIGURE 3: Main chain torsion angles and endocyclic torsion angles in imino acid proline (Pro)

The peptide bond in case of proline residues is known to interconvert at a measurable rate, but this rate of *E-Z* interconversion is very slow, with a $t_{1/2}$ of 20 min at 0 °C and an activation energy barrier of 20.4 kcal mol^{-1}, and is very temperature dependent. Thus at room temperature, for *cis-trans* isomerization of the peptide bonds of most amino acids, the *trans* (*Z*) form is favored over the *cis* (*E*) form, as in the latter case the C^{α} atom and the side chains are in close proximity to the C^{α} atom and the side chains of the preceding residue. Generally, the ratio of *Z* to *E* conformation is of the order of 1000:1 whereas in Pro - a cyclic imino acid, the steric repulsion with C^{δ} atoms decreases (Figure **4**) the energy difference between the two forms (0.5 kcal mol^{-1}). Consequently, with Pro the *Z* form is only slightly favored over the *E* form, generally by a ratio of 66:33.[3] The kinetics of the *E-Z* isomerization is intrinsically slow, and has been identified as the rate-limiting step in several protein-folding pathways.[4]



trans (*Z*) *cis* (*E*)

FIGURE 4: *cis -trans* (*E-Z*) isomers of the peptide bond in proline and substituted prolines

Thus a considerable proportion (about 4–5%) of Xaa-Pro peptide bonds (also called peptidyl-prolyl isomerization, where Xaa is any amino acid) adopt the *cis* conformation, while only 0.03–0.05% Xaa-nonPro bonds occur in the *cis* form.[5] In recent years, there are increasing number of known protein structures determined which exhibit conformational heterogeneity of one or more prolyl peptide bonds.[6] According to a set of published results aromatic residues in the Xaa position gives the highest fraction of *cis* isomer (>20%), while the *cis* content covered the range from 6.0% (Xaa = Pro) to 13.7% (Xaa = Gly) for all other peptides.[4] Table **1** summarizes the above frequency of *cis* isomers observed in various peptides with respect to their neighbouring amino acids. The overall frequency of *cis*-Xaa-Pro listed in the PDB is also about 5-6%, a value that compares well with the typical percentage observed in model peptides.[5] In a recent report *N*-acetyl-L-proline methyl ester and its corresponding methyl amides were studied for their thermodynamics of the *cis* $\rightarrow$ *trans* isomerization for prolyl amide bonds. Increasing the steric bulk of the *N*-terminal residue from Ac- to Ac-Gly- favors the *trans* conformation. Incorporation of a Phe residue *N*-terminal to Pro, however, shifts the equilibrium in favor of the *cis* conformation, *via* a stabilizing aromatic-proline interaction.[7]

Table 1: Percententage of *cis* isomers found in various peptides and with respect to their neighboring amino acids.[4,6]

Peptide bond	% of *Cis* form
Xaa-nonPro	0.03 to 0.05%
Xaa-Pro	4 to 5%
	6.0% (Xaa = Pro)
	13.7% (Xaa = Gly)
	> 20% (Xaa = Ar)

Xaa= any amino acid except those specified

It has also been tried to determine whether the nature of the amino acid preceding proline controls the probability of *cis* prolyl bonds in native proteins. Systematic studies on the thermodynamics and kinetics of the prolyl isomerisation in the pentapeptide series Ac-Ala-Xaa-Pro-Ala-Lys-NH_2 were performed with proteinogenic amino acids substituted in the position preceding proline. It was found that unionisable amino acids like Ala/Trp have much higher kinetics than amino acids with deprotanable side chains as His/Tyr.[8]

Furthermore, a variety of methods have been developed to control the *cis/trans* ratio of prolyl peptide bond. These include buttressing β, γ, and δ positions of proline with a functional group,[9] replacing the peptide bond with an alkene isostere[10] with inclusion of the amide in a ring system that is fused to the pyrrolidine ring[9e] and introducing bridge in the pyrrolidine ring.[10b] Among these, substitutions on C-4 of proline residues seem to have a large effect on the *trans/cis* ratio.[11]

2.2.1 Pyrrolidine ring conformation in proline and substituted proline

Proline is a cyclic imino acid and the bridging of the α-carbon atom to the main chain amide nitrogen by 3 methylene bridge imposes further constraint on the main chain torsion angles Φ and Ψ. Proline and substituted-proline rings exhibit two types of ring-pucker, and in analogy to *ribo* and *deoxyribo* sugars, they are named as *N* (γ-*exo*) and *S* (γ-*endo*) puckers (Figure **5**). These ring-puckers are also expressed in terms of *endocyclic* torsion angles (Figure **3**). In proline the two puckers are almost equally preferred and the energy barrier to inter-conversion is very low.[12]

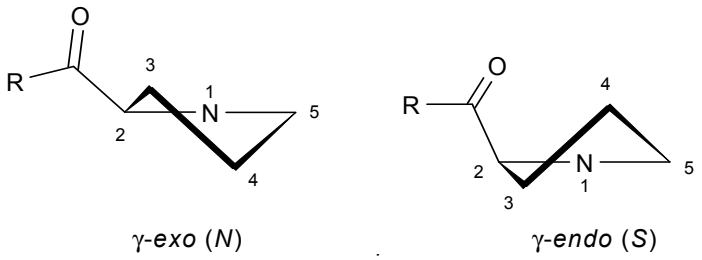


FIGURE 5: Ring-puckers of the pyrrolidine ring in proline and substituted prolines.

The pH dependence of *exo-endo* interconversion of the pyrrolidine ring of L-proline has also been studied in detail.[13] The *endo-exo* equilibrium of proline shifts towards more *endo* in basic solution, because of the steric effect of the nitrogen lone pair. At increased pH the preferred rotamer has the lone pair in the *cis* position to the carboxylic group and its smaller volume requirements, favors more *endo* population.

Extensive study of both the thermodynamics and kinetics of *E-Z* isomerization of 4-substituted proline model compounds has shown that the nature and stereochemistry of electron withdrawing substituent at 4-position has a marked effect on the proline ring conformation and peptidyl-prolyl bond isomerization.[11a,14] For example, the electronegative substituents in the 4*R*-stereochemistry of the proline

increase the stability of *trans* isomer, whereas electronegative substituents in the 4*S*-stereochemistry decrease that stability. In Ac-Xaa-OCH_3 model compounds, it was observed that an electron-withdrawing 4-substituent exerts a significant inductive (-I) effect on pyrrolidine ring nitrogen. This -I effect, transmitted through σ-bonds, increases pyramidalization of the ring nitrogen, decreases the bond order of N-CO bond, and thereby altering the *E-Z* isomerization of the peptidyl-prolyl bond. Thus, in 4-substituted prolines, depending on the steric and electronic effects exerted by the 4-substitutent, pyrrolidine ring may prefer any one of the ring-puckers. For example *trans*-4-hydroxy-L-proline (Hyp) shows almost exclusively γ-*exo* pucker, as found in both crystal structure and in solution.[12] This pucker preference has been attributed to the phenomenon of *gauche* effect between hydroxyl group and amide-N-atom.[11a,15] Further it has also been demonstrated that conformational stability arises from stereoelectronic effects that fix the pucker of the pyrrolidine ring and thereby preorganize the peptide backbone as the steric effect of a 4-methyl group can confer stability similar to that from a 4-fluoro group in the opposite configuration (Figure **6**).[16]

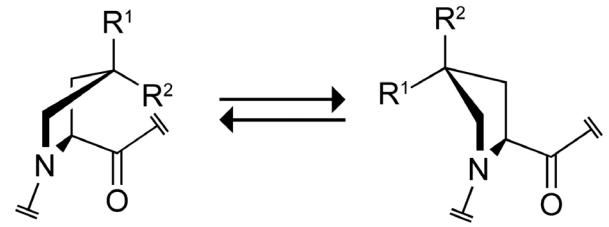


FIGURE 6: Ring conformations of 4-substituted proline residues. The C^{γ}-endo conformation is favored strongly by steric effects when R^1 = Me, R^2 = H (mep) or stereoelectronic effects when R^1 = H and R^2 = F (flp). Similarly, the C^{γ}-exo conformation is favored strongly by steric effects when R^1 = H, R^2 = Me (Mep) or stereoelectronic effects when R^1 = OH, R^2 = H (Hyp) or R^1 = F, R^2 = H (Flp)

2.2.2 *Gauche* effect on the ring-pucker preferences

Gauche effect may be described as "the preference of two electronegative atoms X and Y in vicinally substituted ethanes to remain *gauche* with respect each other rather than *anti*".[17] This effect is surprising since, due to a combination of dipole repulsion and steric effect between the electronegative atoms the vicinal substituents are expected to remain *anti* to each other. Many molecules containing N, O, P, S, F or Cl show a preference for *gauche* conformation. Though the origin of the *gauche* effect is not very clear and *ab initio* quantum chemical calculations underestimate the *gauche* effect still σ-Hyper-conjugation[18] and bent-bonds[19] have been proposed to

explain the phenomenon of the *gauche* effect.

The structure and reactivity of an organic molecule also relies on the stereochemistry of its bonded and non-bonded electron pairs.[20] Stereolectronic (*gauche*) effects, arise from the mixing of an electron pair with the antibonding σ* of an adjacent polar bond (C-X, where X = N or O) and are important in stabilizing the conformation of nucleic acids and carbohydrates.[21] For example, the multiple gauche effects (X-C-C-X) arising from 2' oxygen distinguishes stable conformations of RNA.RNA and RNA.DNA duplexes where anomeric effect (X-C-X) is shown to enhance the stability of the α-isomer of the glycosides.[22]

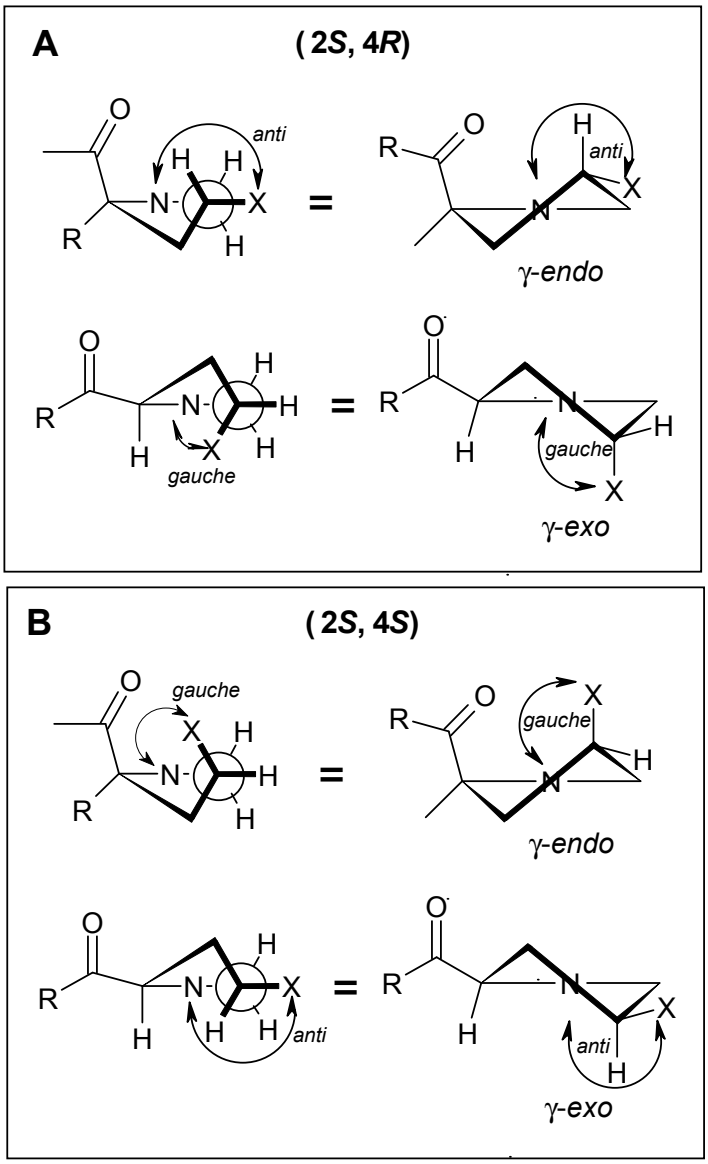


FIGURE 7: Newman and saw-horse projections depicting the *gauche* effect and the resulting pucker preferences in the prolines with electronegative (**A**) 4*R*-substitutent, (**B**) 4*S*-substituent

In 4-substituted prolines, the steric repulsion between the amide-ring nitrogen and the 4-substituent should result in a pseudo-equatorial positioning of the 4-substitutent i.e., *anti* with respect to the ring amide-nitrogen. For example, 4*R*-hydroxyproline and 4*R*-fluoroproline may be expected to exhibit γ-*endo* ring-pucker (Figure **7.A**). However, X-ray crystal structure, ^{1}H-^{1}H and ^{19}F-^{1}H coupling constant analyses confirm the pseudo-axial positioning of fluorine and the resulting γ-*exo* ring-pucker.[20,23] This suggests that the *gauche* effect may be a dominating factor in determining the ring-pucker preference for proline with 4-electronegative substituent. Similarly, in 4*S*-fluoroproline with an opposite stereochemistry at C-4 the *gauche* effect leads to γ-*endo* ring-pucker (Figure **3.5B**).[20,23b]

It has also been proposed that restricted rotation about the N-C^{α} bond leads to a favorable C=O·····C=O electrostatic interaction,[24] which leads to stability almost as strong (80%) as backbone hydrogen bond in a protein[25]. This stabilization involves n→π* interaction *via* electron donation from the oxygen lone pair of the (*i – 1*) amide C=O to the antibonding orbital of the C=O belonging to the proline (*i*) residue. Moreover in evidence for the significance of this interaction, quantum mechanical rather than electrostatic effect was seen to dominate with an estimated contribution of 0.7 kcal mol^{-1} to the stability of the *trans* conformation at 298K.[26]

2.3 Biological relevance of prolyl-peptidyl bond isomerization

Proline is arguably the most important amino acid, *vis-à-vis* the determination of protein structure and function. *Cis-trans* isomerization of the peptide bond plays an important role in many protein-folding pathways[4] including collagen,[27] and in most amino acids, the *trans* (*Z*) form is favored over the *cis* (*E*) form. In contrast, the *trans* isomer of a prolyl peptide bond is only slightly favored over the *cis* isomer and so a considerable fraction of prolyl peptide bonds adopt the *cis* form. The interconversion of *cis* and *trans* isomers about prolyl peptide bonds gives rise to biological effects being concerned with isomer-specific recognition in peptide/protein interactions.[28] Moreover in proline *cis*/*trans* isomerization is also known to play a critical role in splicing, cell signalling, trans-membrane active transport and direct assistance to other catalytic processes apart from protein folding. Prediction of proline *cis*/*trans* isomerization in proteins would have many important applications towards the understanding of protein structure and function.

2.3.1 Some cited applications with examples

The importance of proline *cis*/*trans* isomerization as rate-limiting step in *in vitro* protein folding has been well characterized.[8,29] This process becomes a necessary step to achieve the final stable conformational state. For example, it has been suggested to dominate the folding of the alpha subunit of trp synthase in *E. coli*.[30] And a remarkable example of the nonreceptor protein kinase interleukin-2 tyrosine kinase (*itk*), where *cis-trans* isomerism about the Asn286-Pro287 amide bond is responsible for the coexistence of two different *itk* conformations that display distinct biological activities was suggested by NMR spectroscopic data.[31]

The isomerisation process of Xaa-Pro peptide bonds can be catalyzed and accelerated by the so-called peptidyl prolyl *cis*/*trans* isomerise (PPIase) enzymes,[32] which are found to be involved in cell signalling and cell replication, and be implicated in the induction of severe diseases such as cancer, AIDS, Alzheimer's disease and other neurodegenerative disorders.[33] In addition, proline isomerization functions as molecular switch due to its potential ability to control protein activity within the confines of the intrinsic conformational exchange.[6]

The intestinal H^+/peptide sympoter (PepT1) is an essential ekaryotic membrane protein that differentiates between peptide bond conformers of substrates and selectively binds to the *trans* conformer of a peptide derivative while transporting.[28b] This is also observed in case many cell penetrating peptides.[34] It has also been observed by Lummis *et al.* that *cis–trans* isomerization at a proline opens the pore of a neurotransmitter-gated ion channel.[35] Further the report on incorporation of 4-fluoroprolines into *barstar* using a bacterial expression system is a demonstration of the importance of substituted prolines as tools for protein design and engineering.[36] Another study indicates that prolyl *cis-trans* isomerization can act as a novel molecular timer to help control the amplitude and duration of a cellular process as an important regulatory mechanism in human physiology and pathology. Thus prolyl *cis-trans* isomerization becomes a new target for therapeutic interventions.[37]

2.3.2 Implications to the triple-helical structure of collagen

Proline is a major constituent of collagen with approximately 20% of residues being either Pro or Hyp.[38] The primary structure of collagen is composed of Gly-X-Y repeats where X and Y positions are frequently occupied by proline (Pro) and

hydroxyproline (Hyp). Three chains of collagen come together in parallel fashion with one residue shift register to form a right-handed super-helix - the collagen triple-helix.[39] Each of these chains exists in a left handed polyproline type II helix, where all the peptide bonds assume *trans* conformation.[40] The ability of collagen chains to attain an extended structure has been attributed to the presence of Hyp residues.[41] Several studies have suggested the role of 4-OH in stabilizing the collagen triple-helix by virtue of hydrogen bonding to the amide carbonyls of adjacent chains through water bridges,[39a,42] later challenged by electron withdrawing effect of 4-substituent in the proline ring.[43]

The enzyme *peptidyl-prolyl-cis-trans isomerase* is also known to accelerate the proper assembly of this proline rich protein - collagen.[27b] Study of peptidyl-prolyl bond isomerization is intimately linked to the understanding of the folding and stability of collagen. This knowledge has resulted in the design of novel amino acids for use in collagen mimetics.[43,44]

Since collagen is an imino acid rich protein, both the kinetics and thermodynamics of the *cis-trans* isomerization plays an important role in controlling the overall structure and stability of the folded structure, namely the triple-helix. As demonstrated in earlier work from this laboratory,[45,46] both the electronegativity and the stereochemistry of the C4/C$^{\gamma}$ substituent of proline have a significant effect on T_m of collagen triple-helix. Larger T_m values of (Pro-Yaa-Gly)$_7$ triple-helices correlate with larger $K_{trans/cis}$ values of Ac-Yaa-OMe mimics.[46] It has been found that a C4/C$^{\gamma}$ substituent can enhance the conformational stability by favoring the *trans* isomer, thereby pre-organizing the individual strands to resemble more closely the strands in the triple-helix.

2.4 Aim and rationale of the present work

As discussed in the previous sections it is summarized that the prolyl-peptidyl isomerisation is interplay of several factors. The decisive structural determinants (Figure **8**) revolving around the Xaa-Pro peptide structure are: (i) the state of ionisation of side chain to the proline ring or protonation status of a substituent (like NH_2), (ii) stereochemistry of the substituent mainly at C$^{\gamma}$-position, (iii) the arising pyrrolidine pucker, (iv) pK_a / ionisation of polar groups, (v) role of neighbouring amino acid and its steric properties. Overall the Xaa-Pro peptide bond's conformational flexibility is ruled by all the above parameters.

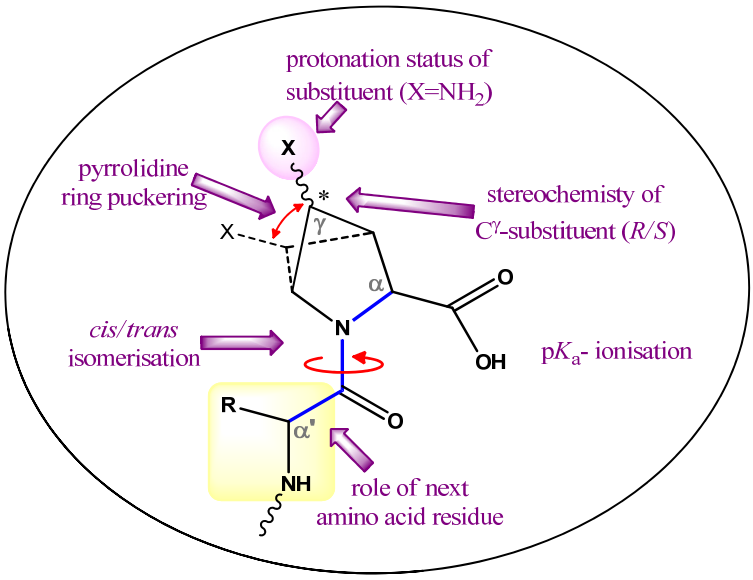


FIGURE 8: Decisive structural determinants in Xaa-Pro prolyl-peptide isomerization process

Chimeric amino collagen peptides [Pro(X)-Amp(Y)-Gly]$_6$ & [amp(X)-Pro(Y)-Gly]$_6$ show pH dependent triple-helices.[47] In case of 4-aminoproline, it is possible that the protonated and non-protonated forms have different types of ring pucker and *trans* : *cis* ratio. This may result in the formation of pH regulated helical forms, because the protonated and free amino group will have different electronegativity.

Ala-Pro thiopeptides are known to show pH dependent *trans* : *cis* isomerisation[48] as thioamides undergo slower rotation of the C–N bond than the corresponding amides (typical $t_{1/2}$ measured in hours instead of milliseconds at RT)[49]. It is also known in literature[50] that the amides of secondary amines show the presence of both *trans* and *cis* isomers easily detectable by ^{1}H / ^{13}C NMR. Analysis by ^{1}H-NMR of the thiopeptide[48] have allowed determination of the *trans:cis* ratio, following the methyl doublet of the alanyl residue as a clear marker (Figure **9**).

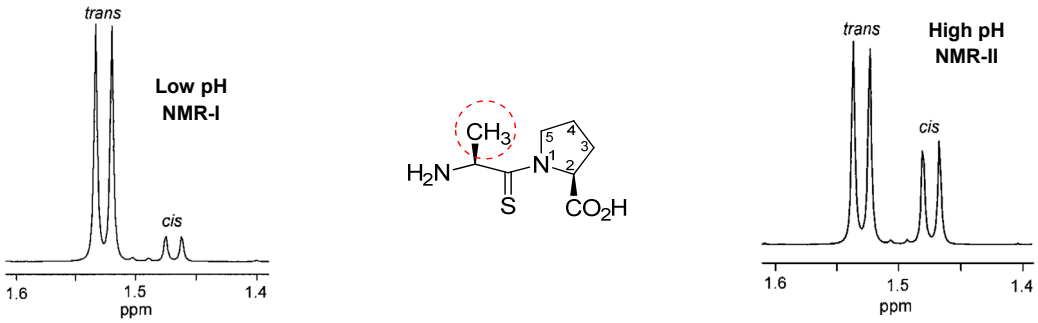


FIGURE 9: Methyl doublets of alanine in the ^{1}H NMR of thiopeptide at two different pH.

In order to understand the relative effects of protonation/non-protonation, stereo-chemical and electronic influences of the various 4-substituent on the *trans* : *cis* equilibrium of peptidyl-prolyl bond and the conformation of the pyrrolidine ring it was decided to first examine the effect of 4-substituents in proline model compounds of type H_2N-Ala-Pro-OH (4-substituted-proline-alanine dipeptides) under different pH conditions using NMR spectroscopy. For this purpose *hydroxyl*, *amino* and *azido* groups were chosen as 4-substituents.

The objective of the present work is to measure the variation of *trans* : *cis* isomer ratio in Ala-Pro dipeptide as a function of pH. This section deals with the chemical synthesis and ^{1}H-NMR analysis of Ala-Pro and its substituted derivatives in order to estimate the *trans* : *cis* ratio, by monitoring the intensity of the doublet due to methyl group of the alanyl residue.

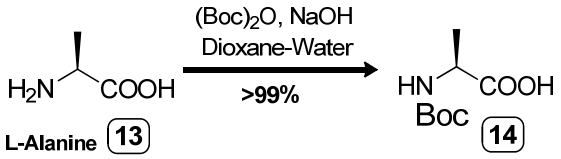
The specific objectives of this section are

1) Synthesis of unsubstituted dipeptide Ala-Pro (**19**) and its 4-substituted proline analouges Ala-Hyp (**23**), Ala-hyp (**28**), Ala-Azp (**31**), Ala-azp (**34**), Ala-Amp (**36**), Ala-amp (**38**), $N^{\alpha\prime}$-acetyl-Ala-Amp (**41**) and $N^{\alpha\prime}$-acetyl-Ala-amp (**44**) by solution phase peptide coupling reactions.
2) Assignment of dipeptide structures and characterisation of *cis* & *trans* isomers through 2D-NMR.
3) Analysis of variation of *cis* & *trans* isomers at various pH.
4) pK_a determination of carboxylic acid of proline, amino group of alanine and amino groups in 4-substituted aminoproline dipeptides. Elucidation of its correlation with percentage of *cis* & *trans* isomers.
5) Study of change in conformation of the above mentioned Ala-Pro model dipeptides as a function of pH by means of CD spectroscopy.

2.5 Results

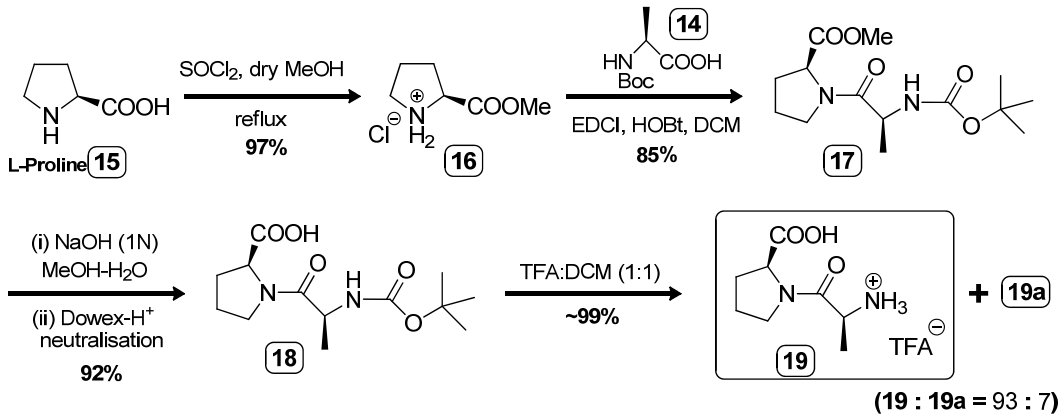
2.5.1 Solution phase synthesis of Ala-Pro model dipeptides

Synthesis of Ala-Pro (**19**) and its 4-substituted counterparts for the present study were done according to the Scheme **2.2** – **2.10**. These schemes elaborate the synthetic route for obtaining specifically Ala-Hyp **23**, Ala-hyp **28**, Ala-Azp **31**, Ala-azp **34**, Ala-Amp **36**, Ala-amp **38**, $N^{\alpha\prime}$-acetyl-Ala-Amp **41** and $N^{\alpha\prime}$-acetyl-Ala-amp **44** for the purpose of the study.



Scheme 2.1: Synthesis of (*S*)-N[1]-(*t*-butoxycarbonyl)-alanine **14**

The naturally occurring amino acid *L*-proline **15** was first converted to its methyl ester **16** using $SOCl_2$-methanol. Boc-Alanine **14** prepared from *L*-alanine (Scheme **2.1**) was coupled to **16** in the presence of EDCI and HOBt in DCM as coupling reagents to obtain the protected Boc-Ala-Pro-OMe dipeptide **17** in 85 % yield. Dipeptide formation was confirmed by the upfield shift of signals corresponding to α'-C$\underline{H}_3$ of alanine from δ 1.42 & 1.45 (d, 3H) to (δ 1.23 & 1.29, 2d, 3H) in the peptide **17**. In case of dipeptides arising from proline *E-Z* isomers resulting from the rotation of peptidyl-prolyl bond gives rise to two sets of signals corresponding to major and minor isomers (designated as *ma* and *mi* respectively) in the NMR spectra. Correlation of ^1H and ^{13}C in HETCOR spectroscopy confirmed the carbon-hydrogen framework of the molecule. LC-MS peaks at 301.14 $[M+1]^+$, 323.09 $[M+Na]^+$ and 339.09 $[M+K]^+$ along with IR band at (1650 cm^{-1} for >N-C=O vibration) also supported the structural integrity of this dipeptide. The methyl ester of **17** was then hydrolysed with aq. NaOH in methanol followed by its neutralisation with Dowex-H^+ to get the free acid **18,** validated by the disappearance of -OC$\underline{H}_3$ signals (δ 3.66) from ^1H-NMR spectrum. This compound crystallized in methanol-water system which upon X-ray analysis (Appendix-B) re-affirmed the structural integrity of the molecule. The deprotection of N-Boc group using 50% TFA-DCM solution resulted in desired dipeptide **19** (9.6 t_R, 92.7%) yield accompanied by another product (11.9 t_R, 6.3%) identified as the lactam (**19a**) checked by reverse-phase HPLC.



Scheme 2.2: Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-proline dipeptide methylester **17**

The ^{1}H-NMR (400 MHz) spectrum of compound **19** in D_2O at pH ≤ 2 shows two doublets due to α'-CH_3 of alanine at δ 1.43 and 1.38 for major (91%) and minor (9%) isomers respectively. The corresponding signals for Ala-α'H appeared as multiplets at δ 4.26 and 3.99. The signals due to each of the proline ring protons also appear as double multiplets which are assigned with the help of 2D NMR as explained later. The assignment of signals given according to Figure **10** are as follows: δ 4.58-4.55 (ma) & 4.42-4.39 (mi) (2m, 1H, rel. area 92:8 resp., Pro-αH), 4.26 (ma) & 3.99 (mi) (2q, J= 7.2 & 6.9 Hz, resp. 1H, rel. area 93:7 resp., Ala-α'H), 3.64-3.51 (ma) & 3.42-3.30 (mi) (2m, 2H, rel. area 96:4 resp., Pro-δH_2), 2.32-2.21 (m, 1H, Pro-βH_a), 2.01-1.90 (m, 3H, Pro-βH_a & Pro-γH_2), 1.43 (ma) & 1.38 (mi) (2d, J= 7.3 & 6.9 Hz, resp. 3H, rel. area 91:9 resp., Ala-C<u>H</u>$_3$).

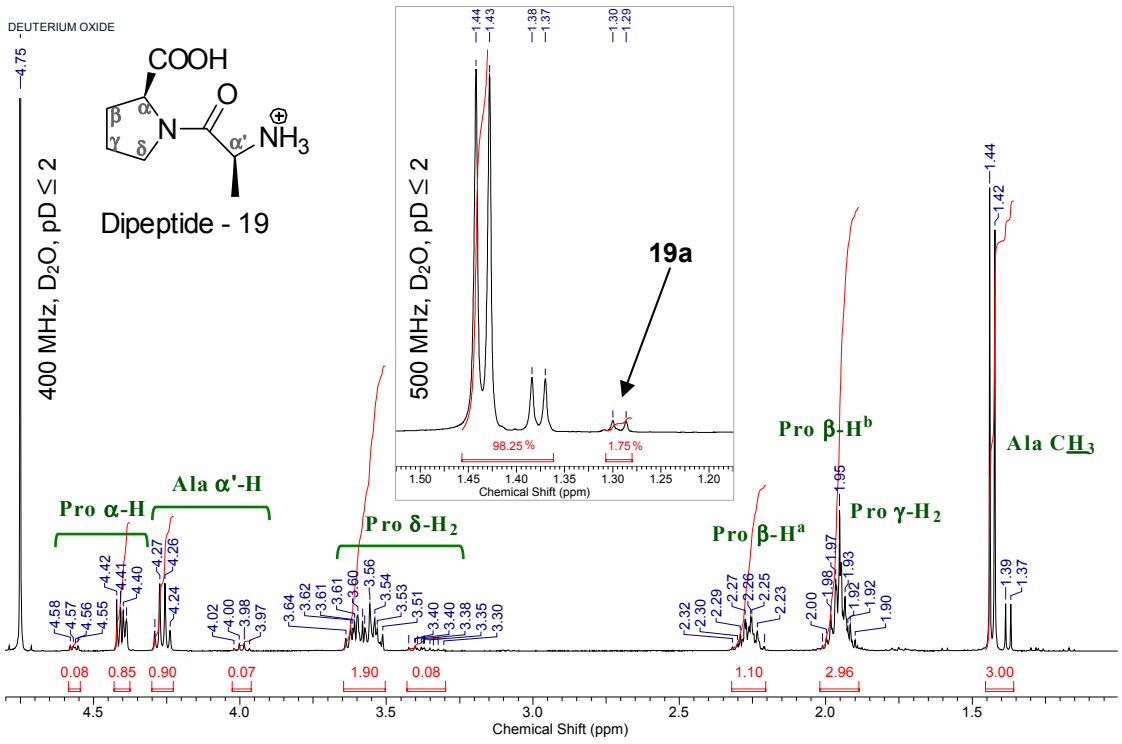


Figure 10: ^{1}H-NMR spectra of dipeptide **19** (500 MHz) in D_2O (pD ≤ 2); **Inset:** appearance of compound **19a** along with dipeptide **19**.

The minor product seen in HPLC at t_R=11.9 was identified as the 6-membered lactam (**19a**) arising from intra-molecular cyclization between Ala-NH_2 and Pro-COOH residues. This product was separated by preparative RP-HPLC and the presence of 6-membered lactam fuctionalisation was suggested by signal in IR at 1655 cm^{-1} (Figure **11 inset**). The structure of **19a** as the lactam was further confirmed by ^{1}H-NMR (Figure **11**) and LC-MS. δ 4.29-4.19 (m, 2H, Pro-αH & Ala-

α'H), 3.49-3.38 (m, 2H, Pro-δH_2), 2.26-2.20 (m, 1H, Pro-βH^a), 2.0-1.94 (m, 1H, Pro-βH^b), 1.92-1.82 (m, 2H, Pro-γH_2) and 1.31 (d, J= 7.3 Hz, 3H, Ala-CH_3). LC-MS: 169 $[M+1]^{+\bullet}$ and 192 $[M+Na]^{+\bullet}$

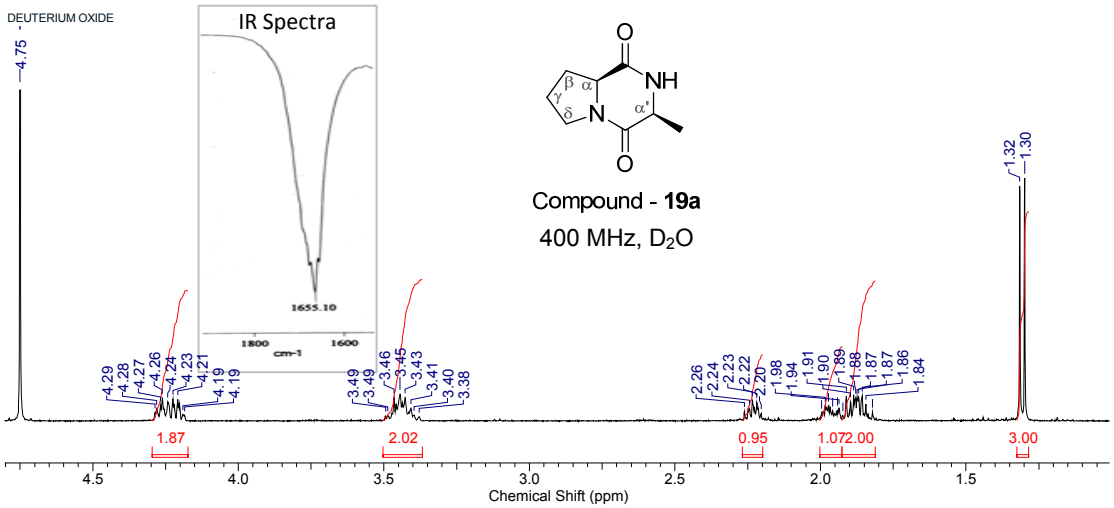
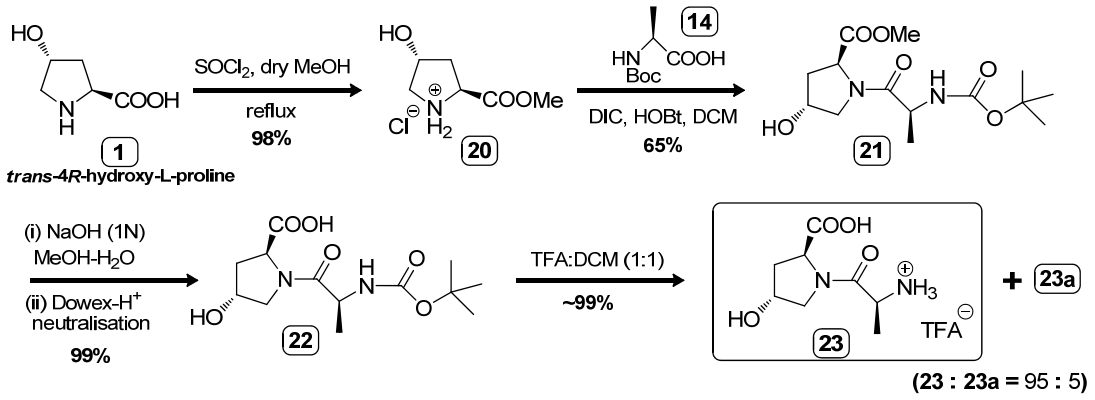


Figure 11: ^{1}H-NMR spectra of compound **19a** (500 MHz) in D_2O

Although most of the required dipeptide **19** was isolated by semi-preparative HPLC, a slight contamination (<1%) of **19a** was detected in the ^{1}H-NMR spectrum as a third set of doublet at δ 1.30 for the three methyl protons of alanine (Figure **10 inset**) and since this peak was very small to interfere with other peaks, it was ignored in further NMR study of compound **19**.



Scheme 2.3: Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-hydroxyproline dipeptide methylester **23**

Scheme **2.3** shows the synthesis of the dipeptide Ala-Hyp (**23**) starting from the naturally occurring *trans*-4-hydroxy-*L*-proline. This was first converted to its

methyl ester **20** using $SOCl_2$-MeOH and coupled to Boc-Ala **14** using DIC and HOBt in DCM to obtain the protected dipeptide **21**. The structure of this compound was characterized through its ^1H and ^{13}C-NMR spectra. The 1D ^1H-NMR spectra (Figure **12**) showed a pattern of signals for α'-CH_3 of alanine, similar to that seen for Ala-Pro (**17**) with major and minor isomers in 94:6 ratio. Compound **21** is similar to compound **17** except a hydroxyl substitution at γ-carbon of the pyrrolidine ring resulting in the downfield shift of γ-proton while chemical shift of other protons remain unchanged from that of dipeptide **17**.

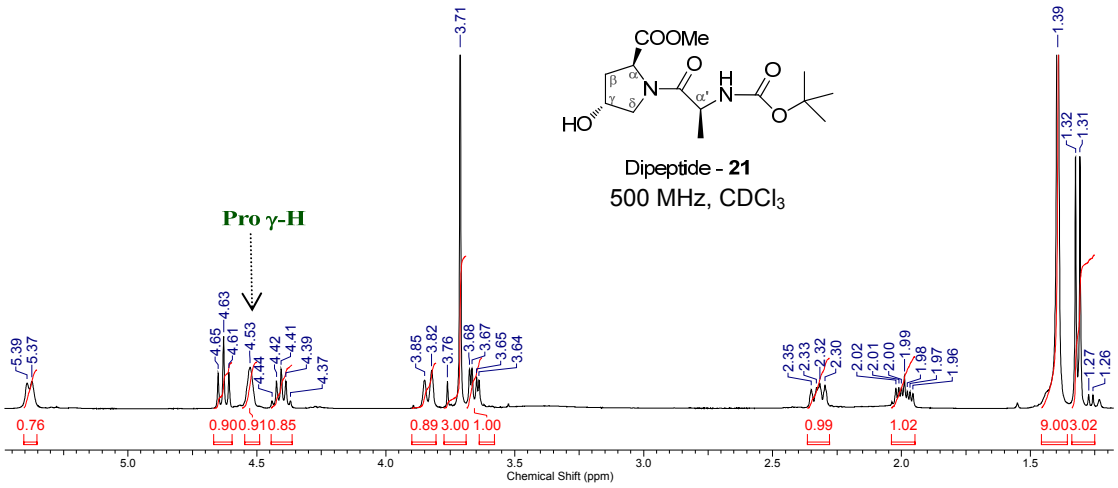
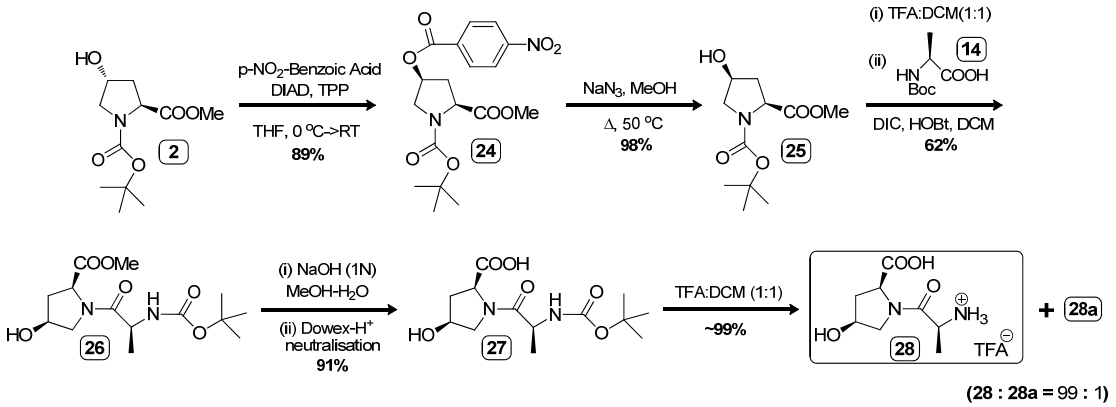


Figure 12: ^1H-NMR spectra of dipeptide **21** (500 MHz) in $CDCl_3$

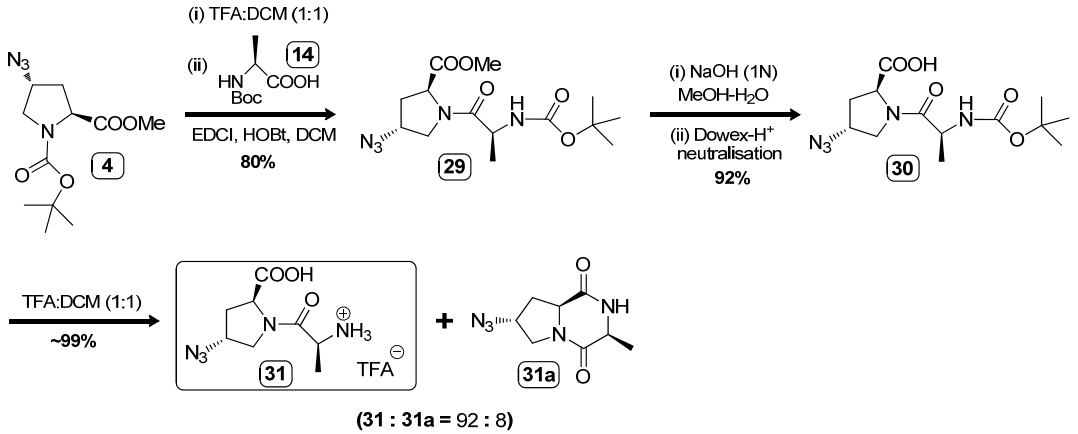
In the subsequent steps **21** was hydrolysed to **22** and Boc-deprotected to get the desired dipeptide **23** as TFA salt. HPLC of the product suggested presence of **23** as major constituent and the lactam **23a** as minor constituent in relative ratios of 95 : 5.

Similarly the diastereomeric dipeptide (Ala-hyp) **28** was synthesized from *trans*-4(*S*)-hydroxy-L-proline as per scheme **2.4.** The 4*S*-proline methylester **25** was obtained from 4*R*-Proline derivative (Scheme-**1.1**/Section-I) in two steps involving the inversion of the stereocenter at γ-carbon. The 4*R*-hydroxyl group in **2** was reacted with p-nitrobenzoic acid under Mitsunobu conditions to get the 4*S*-O-benzoate ester **24**. Hydrolysis of this ester with sodium azide in methanol yielded 4*S*-OH-derivative **25**, which was Boc-deprotected followed by coupling it with compound **14** to obtain the protected dipeptide **26** in 62% yield. This was transformed to the desired dipeptide **28** as shown in the scheme.



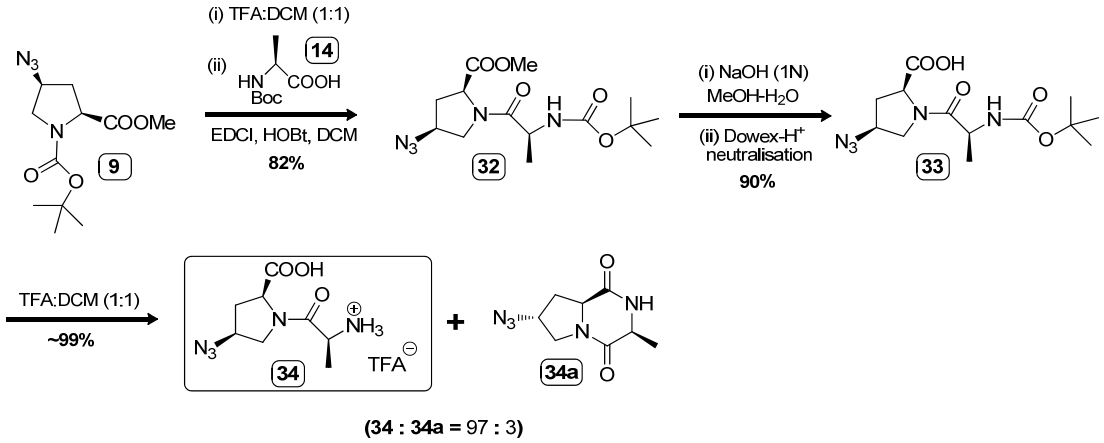
Scheme 2.4: Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-hydroxyproline dipeptide methylester **28**

The diastereomers Ala-4*R*-azidoproline **31** and Ala-4*S*-azidoproline **34** were synthesised in analogous method from their precursors **4** and **9** (Section-I) respectively by Schemes **2.5** and **2.6**.



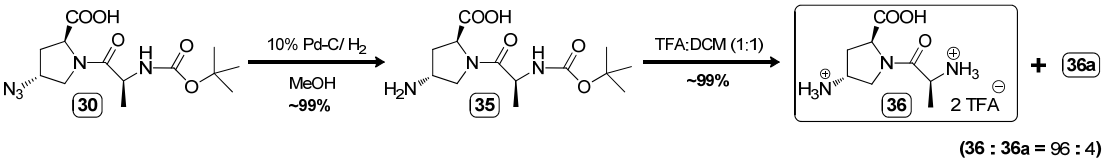
Scheme 2.5: Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-azidoproline dipeptide methylester **31**

All final compounds and intermediates in schemes **2.4**, **2.5** & **2.6** were characterised by ^{1}H and ^{13}C NMR as mentioned before. The relative ratios of the dipeptide and the lactams were estimated from corresponding RP-HPLC data. The spectral data are shown in Appendix B.

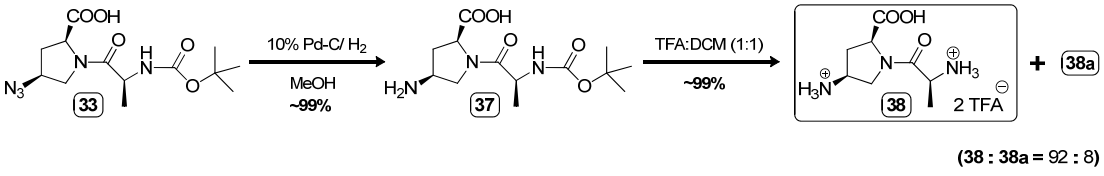


Scheme 2.6: Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-azidoproline dipeptide methylester **34**

Schemes **2.7** and **2.8** give the synthetic route for dipeptides Ala-Amp (**36**) and Ala-amp (**38**). The synthesis of these 4(*R/S*)-aminoproline substituted Ala-Pro dipeptides were achieved by hydrogenation of 4(*R/S*)-azidoproline dipeptides (**30** and **33**) using 10% Pd-C in methanol. The reduction of azide to amine was confirmed from the IR spectra with disappearance of the azide signature band around 2107 cm^{-1}.

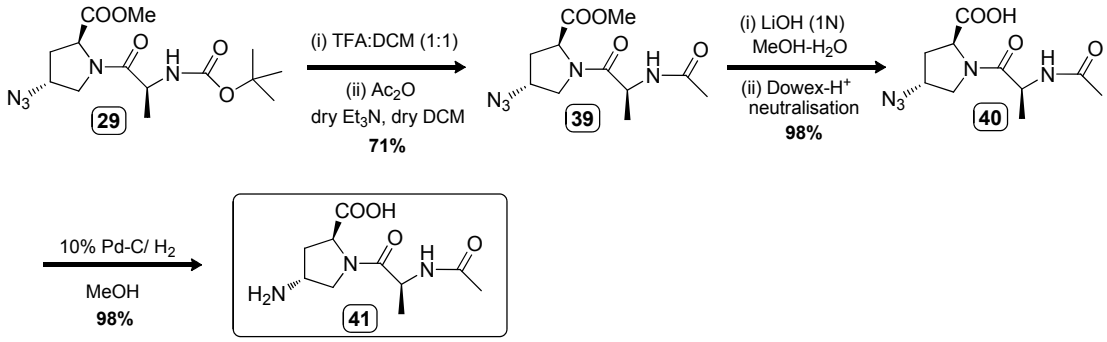


Scheme 2.7: Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-aminoproline dipeptide methylester **36**



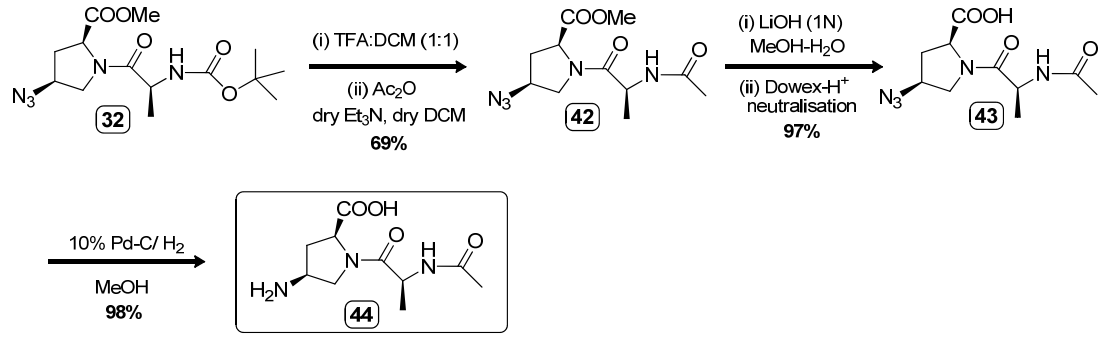
Scheme 2.8: Synthesis of (*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-aminoproline dipeptide methylester **38**

In order to investigate the effect of charged amino group of alanyl residue in the isomerization process of the dipeptides, N-capped dipeptides, (N^{α}'-acetyl)-Ala-Amp **41** and (N^{α}'-acetyl)-Ala-amp **44** were also synthesized (Scheme **2.9** and **2.10**). Since direct acetylation of **36** and **38** would lead to diacetylated products, the acetylation was done on the 4(*R/S*)-azidoproline dipeptides (**29** and **32**), followed by reduction to 4(*R/S*)-aminoproline dipeptides (**35** and **37**) respectively.



Scheme 2.9: Synthesis of (*S*)-N-acetyl-alanine-(2*S*,4*R*)-4-aminoproline dipeptide **41**

The treatment of the Boc-Ala-4*R*-azidoproline methylester **29** with 50% TFA-DCM followed by reaction with acetic anhydride in presence of triethyl amine resulted in $N^{\alpha\prime}$ -acetylation to give compound **39**. The structure of product was confirmed by presence of two sets of signals corresponding to acetyl group (δ 1.98 & 2.00, 2s, 3H) rotamers. Selective hydrolysis of –OMe group using aq. LiOH in methanol gave compound **40** which upon reduction with Pd-C/H_2 yielded compound **41**. Similarly compound **44** was synthesized in 98% yield starting from the protected dipeptide **32**.

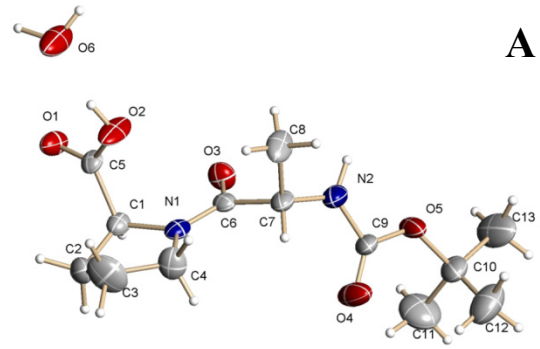


Scheme 2.10: Synthesis of (*S*)-N-acetyl-alanine-(2*S*,4*S*)-4-aminoproline dipeptide **44**

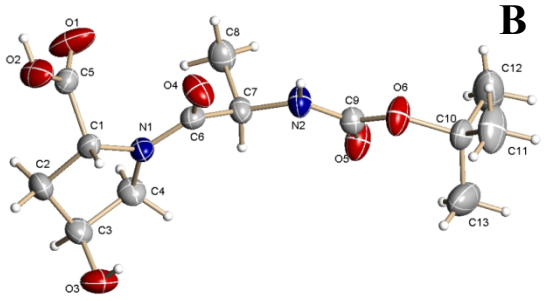
The structure of all intermediate and final compounds (Schemes **2.9** & **2.10**) are supported by ^{1}H, ^{13}C and confirmed by LC-MS data. The purity of the final compounds was established by RP-HPLC data (Appendix B).

2.5.2 Crystal structure

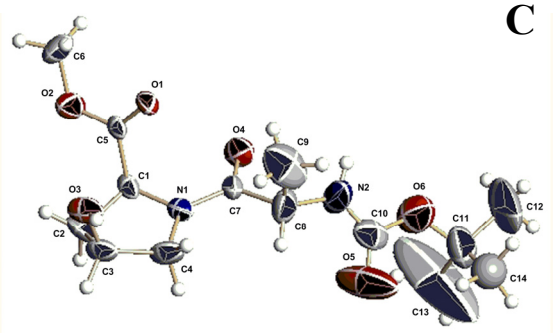
The dipeptides **18**, **22**, **26**, **32** and **33** formed good crystals enabling their structure analysis by single crystal X-ray diffraction. The crystal structures correspond to the major rotameric isomer and confirm the stereochemical disposition of various proline substituents.



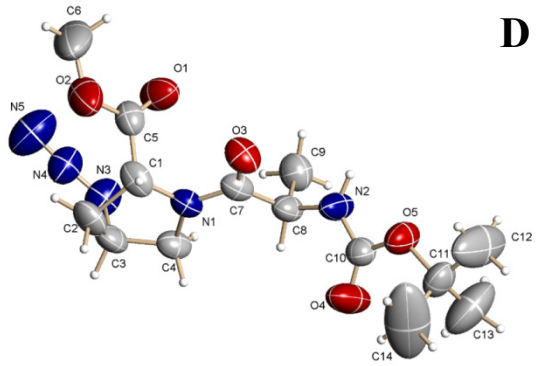
Dipeptide **18** (Boc-Ala-Pro-OH)



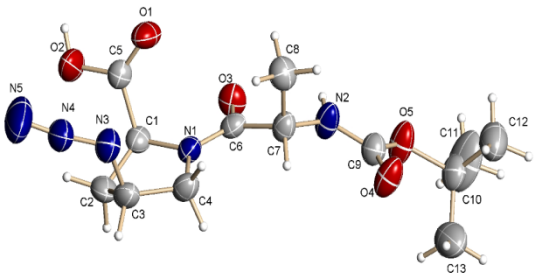
Dipeptide **22** (Boc-Ala-Hyp-OH)



Dipeptide **26** (Boc-Ala-hyp-OMe)



Dipeptide **32** (Boc-Ala-azp-OMe)



Dipeptide **33** (Boc-Ala-azp-OH)

FIGURE 13: ORTEP diagram of crystal structures of N-Boc protected dipeptides **18**, **22**, **26**, **32** and **33** obtained by single crystal X-ray diffraction.

In order to estimate the effect of the 4-substituent on the main-chain torsion angles and the ring-pucker adopted by the Ala-Pro model compounds, the crystals structures were analyzed for individual ϕ, ψ, ω and χ values through their ORTEP

diagrams (Figure **13**) with 3D-modeled structure (Figure **14**). And the main-chain torsions and the ring-pucker types derived from these structures are listed in Table **2**.

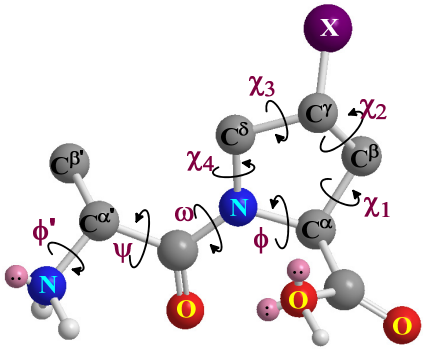


FIGURE 14: Main chain torsion angles and endocyclic torsion angles in Ala-Pro based model dipeptides; only terminal (polar) hydrogen atoms are shown and "X" signifies any substituent.

Table 2: Main-chain torsion angles of Ala-Pro based model dipeptides

Dipeptides	Main chain torsion angles (degrees)							
	ω	ϕ	ψ	ϕ'	χ_1	χ_2	χ_3	χ_4
Boc-Ala-Pro-OH **(18)**	170.3	-71.8	153.7	-95.7	29.6	-34.2	24.5	-5.8
Boc-Ala-Hyp-OH **(22)**	-165.9	58.5	-153.4	148.6	26.3	-38.6	35.4	-19.8
Boc-Ala-hyp-OMe **(26)**	169.4	-71.6	160.4	-116.0	33.6	-37.9	29.6	-9.1
Boc-Ala-azp-OMe **(32)**	-172.3	71.1	-157.8	109.3	-29.8	36.1	-27.7	9.0
Boc-Ala-azp-OH **(33)**	168.3	-74.2	158.7	-143.4	33.1	-37.3	26.3	-5.4

Although most of the crystals were expected to be frozen in their most stable and minimum energy conformation, the stereoelectronic effect from 4-substitution gives them variation in spatial arrangement. For all the model compounds the ω torsion-angle corresponding to the amide bond are approximately near -180°, i.e., an almost planar amide bond suggesting a *trans* conformation. The other two main-chain torsion angles ψ and ϕ obtained from the crystal structures of Boc-Ala-hyp-OMe **26** (ψ = 160.4°; ϕ = -71.6°), Boc-Ala-azp-OMe **32** (ψ = -157.8°; ϕ = 71.1°) and Boc-Ala-Azp-OH **33** (ψ = 158.7°; ϕ = -74.2°) are very close to each other and with unsubstituted proline ring in dipeptide Boc-Ala-Pro-OH **18** (ψ = 153.7°; ϕ = -71.8°). However the 4*R*-hydroxy substituted **22** showed deviation towards (ψ = -153.4°; ϕ = 58.5°) along with (χ_4 = -19.8°) the torsion angle about C^{δ}-N atoms of proline ring.

2.5.3 Two-dimensional NMR study of Ala-Pro model dipeptides

Evaluation of interconvertible (*cis–trans* / proline puckers) isomeric forms required unambiguous assignments of chemical shift of the all protons. Therefore the primary assignment of common framework of all target Ala-Pro derivatives in dipeptide **19** was done by a combination of ^{1}H-NMR, 2D COSY (^{1}H-^{1}H) and 2D NOESY experiments.

2.5.3a COSY : Assignment of dipeptide backbone and pyrrolidine ring protons in Ala-Pro dipeptide & its derivatives

In the ^{1}H-NMR spectra scanned at 500 MHz in D_2O (pD 2.2), the methyl group (Ala-C$\underline{H}_3$, δ 1.47 (*ma*) & 1.41 (*mi*), 2d, 3H) of alanine is assigned by its characteristic chemical shift (ppm). Next is the Pro-αH which is easily assignable as the most downfield shifted signal (δ 4.40-4.43 (*ma*) & 4.45-4.48 (*mi*), 2dd, 1H) and its identification triggers the assignment of all J-coupled protons in ring by ^{1}H-^{1}H COSY spectral analysis. Both for Ala-C$\underline{H}_3$ and Pro-αH peaks, two sets of isomers corresponding to major and minor isomers (designated as *ma* and *mi* respectively) are observed arising due to rotation around the prolyl-peptidyl bond.

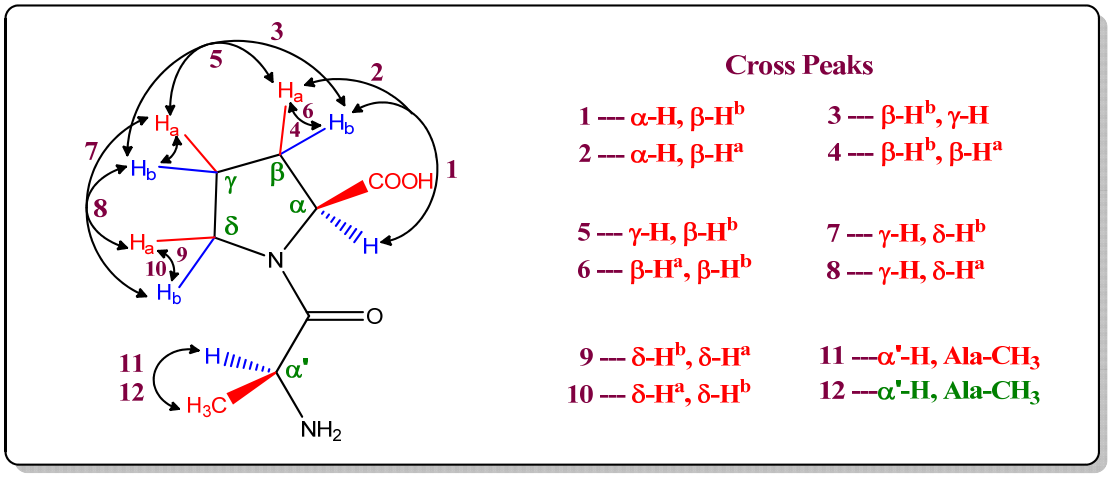


FIGURE 15: Correlations expected with respect to 2D ^{1}H-^{1}H COSY NMR spectra of Dipeptide **19** at pD 2.2.

Figure **15** illustrates the typical 2D NMR assignment of vicinal connectivity by the use of ^{1}H-^{1}H COSY has been given in Figure **16**. The peak for Pro-αH at δ 4.40-4.43 (*ma*) shows two cross peaks (1-2) corresponding to its coupling with neighbours Pro-βH_2, having two non-equivalent protons (βH^a and βH^b). Among these, the relative chemical shift values assists in identifying the upfield components (cross peaks 3 and

5) as due to vicinal β– and γ–protons while the downfield signal at (cross peaks 4 and 6) correspond to germinal β–protons, thus accounting for the observed set of four cross peaks 3 ($\beta H^b\gamma H$), 4 ($\beta H^b\beta H^a$), 5 ($\beta H^a\gamma H$) and 6 ($\beta H^a\beta H^b$). A zoomed area of these four cross peaks is shown in Figure **17** which gives a clear view of the coupling protons. The assignment of $\beta H^a\beta H^b$ paves way for identifying the downfield signals at δ 3.58 & 3.63 as due to δH_2 *via* $\gamma H\delta H^b$ and $\gamma H\delta H^a$ cross peaks 7 and 8. These further connect to two the other cross peaks 9 ($\delta H^b\delta H^a$) and 10 ($\delta H^a\delta H^b$) for geminal protons. This completes the assignment of all prolyl ring protons. The assignment of alanine residue protons Ala-C<u>H</u>$_3$ and Ala-α'H starts from the easily identifiable Ala-C<u>H</u>$_3$ at δ 1.47 (*ma*) & 1.41 (*mi*). These protons show two cross peaks 11 & 12 at 4.3 & 3.98 respectively due to α'HAla-C<u>H</u>$_3$ separately for its major (*ma*) and minor (*mi*) isomers.

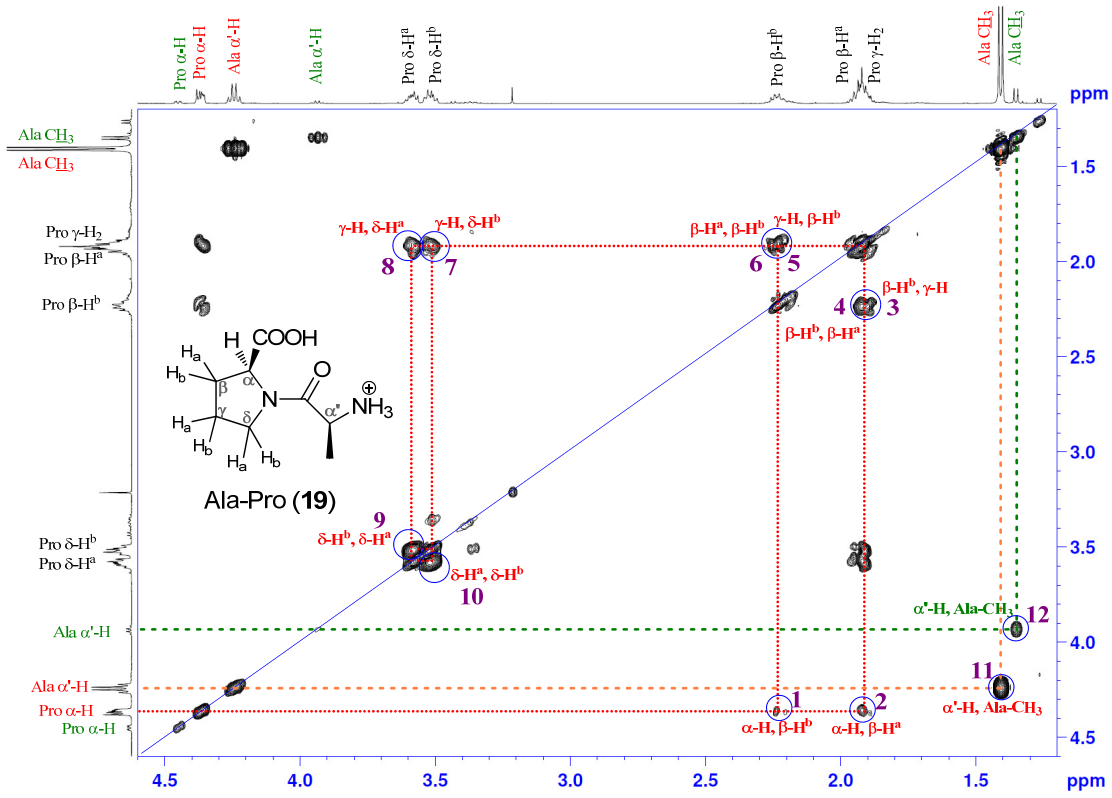


FIGURE 16: 2D ^{1}H-^{1}H COSY NMR spectra of pyrrolidine ring protons in dipeptide **19** (500 MHz) in D_2O, pD 2.2; The numbers 1-12 correspond to assigned cross peaks; Color code: Red – major isomer (*ma*), Green – minor isomer (*mi*)

The signals due to the major (*ma*) and minor (*mi*) isomers are designated by two different colors; (*ma*) in red and (*mi*) in green.

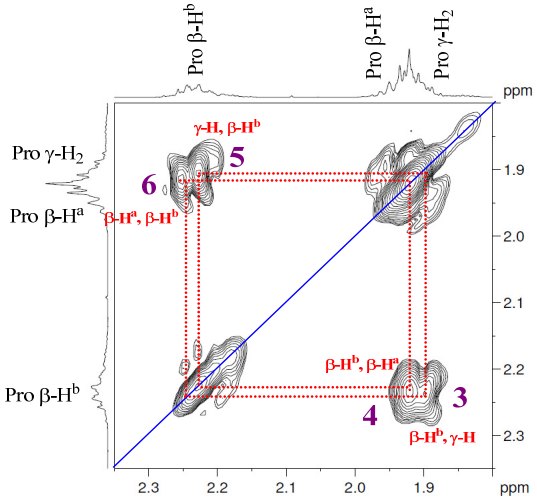


FIGURE 17: Expanded view of selected portion from 2D ^{1}H-^{1}H COSY NMR spectra of pyrrolidine ring protons in compound **19** (500 MHz) in D_2O at pD 2.

Interestingly in the ^{1}H-NMR spectra of the same dipeptide **19** recorded in D_2O at pD 12, the intensities of the C$\underline{H}_3$ signal due to the two isomers was equal and are now represented as *i1* and *i2*. An upfield shift was observed for the methyl group of alanine (Ala-C$\underline{H}_3$, δ 1.18 (*i1*) & 1.11 (*i2*), 2d, 3H) pD 12 compared to that at pD 2. The vicinal connectivity observed from the ^{1}H-^{1}H COSY spectrum for both isomers illustrated below in Figure **18** the assignments for isomer *i1* is designated in red color while that of *i2* in green color. And the protons that could not be differentiated into two separate isomers are designated in black color.

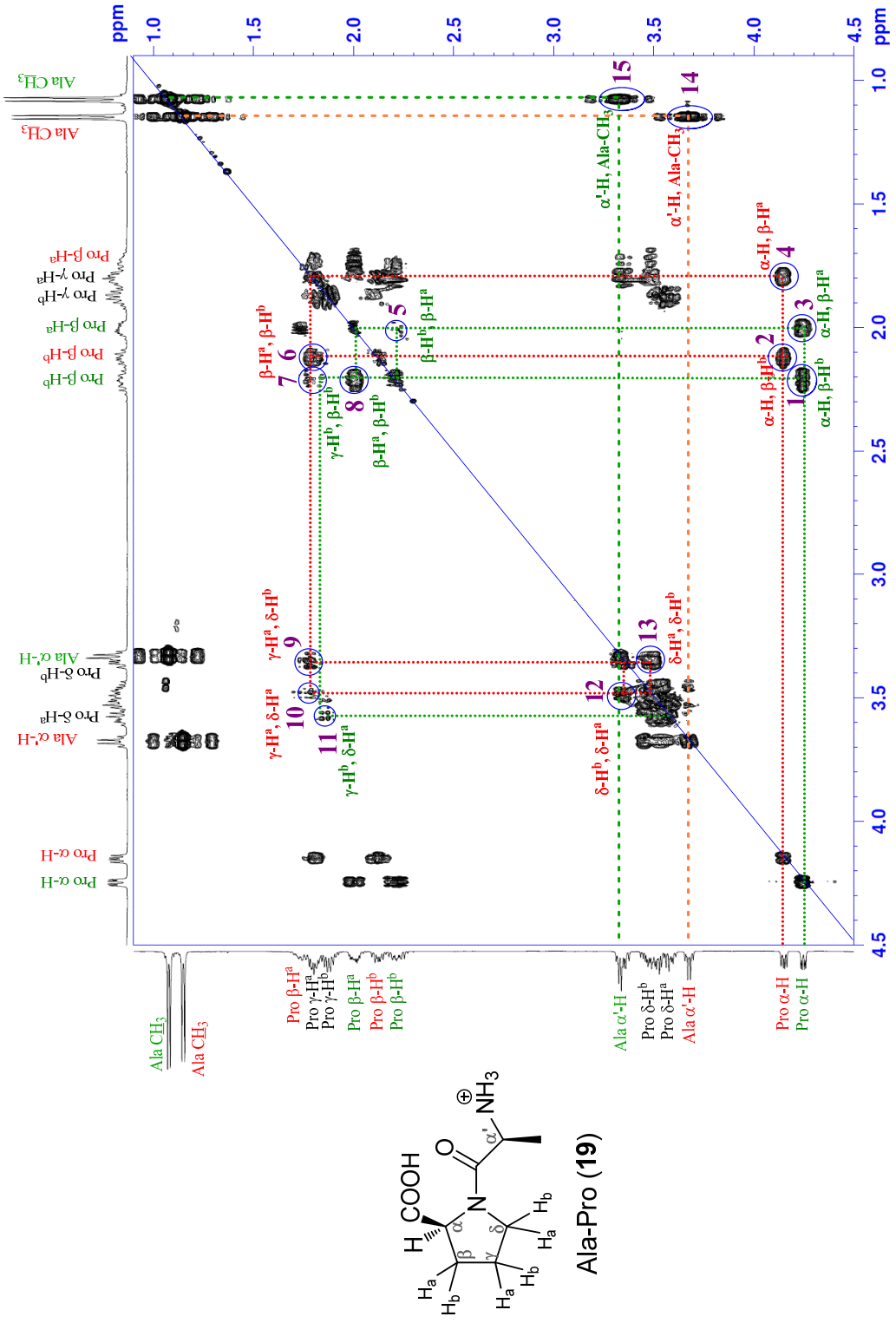


FIGURE 18: 2D ^{1}H-^{1}H COSY NMR spectra of pyrrolidine ring protons in compound **19** (500 MHz) in D_2O at pD 12.0; The numbers 1-15 correspond to assigned cross peaks; Color code: Red – major isomer (*i1*), Green – minor isomer (*i2*)

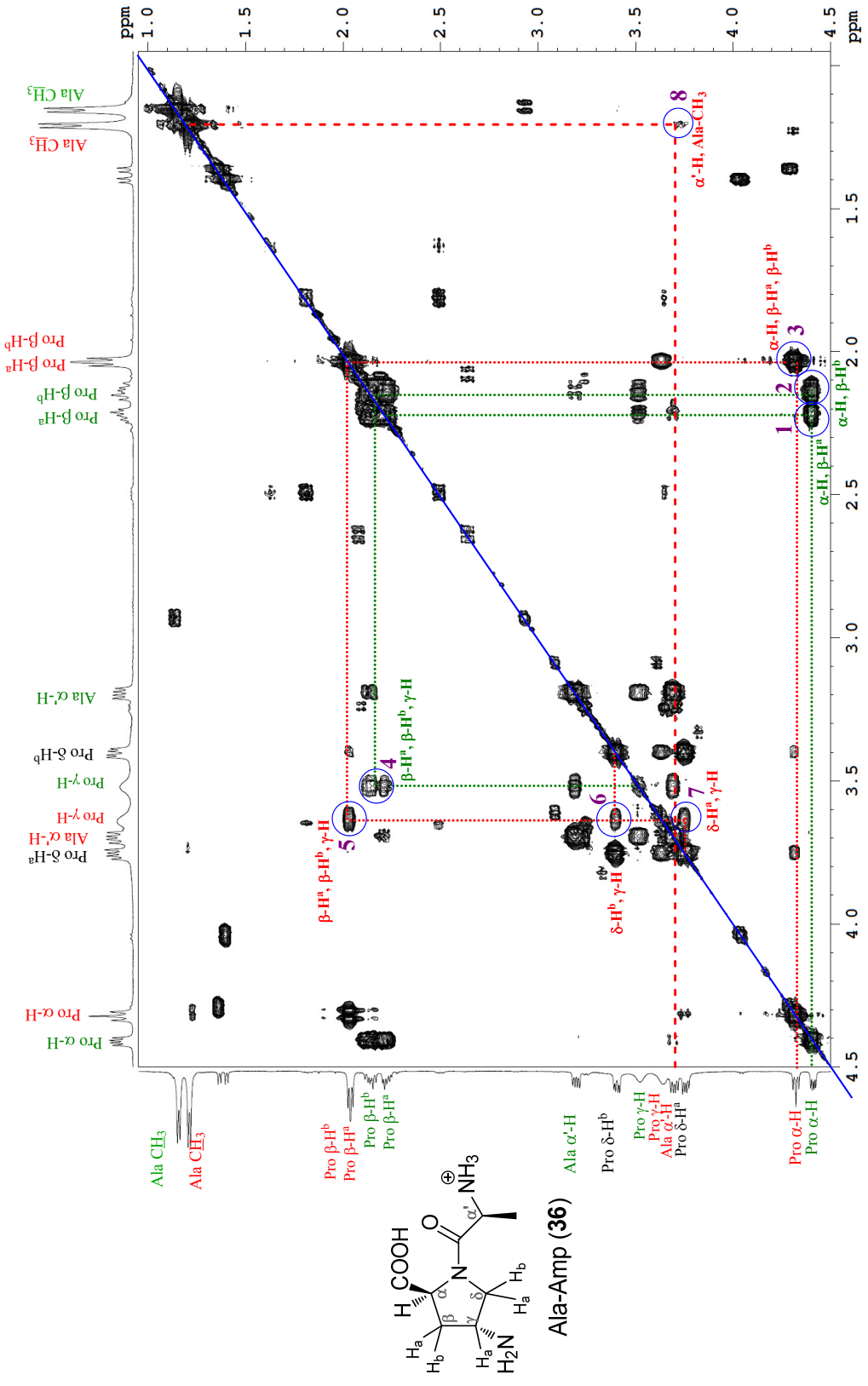


FIGURE 19: 2D ^{1}H-^{1}H COSY NMR spectra of pyrrolidine ring protons in compound **36** (500 MHz) in D_2O at pD 12.0; The numbers 1-8 correspond to assigned cross peaks; Color code: Red–major isomer (*ma*), Green–minor isomer (*mi*),

The substituted dipeptides **36** and **38** were analysed with similar analogy with ^{1}H-^{1}H COSY and the connectivity of protons as per the Figures **19** and **20** respectively.

Among the substituted dipeptides, Ala-amp **38** showed much better splitting of peaks into major (*i1*) & minor (*i2*) isomers. The cross peaks 1-16 in Figure **21** represent the intra-residue J-coupling of the protons in the dipeptide. The dashed lines correspond to ^{1}H-^{1}H couplings in alanine residue while the tiny dotted lines stand for proline residues. Figure **20** gives an expanded view of the selected portion from 2D 1H-NMR spectrum of dipeptide **38**. Here also the red lines stand for major *i1* isomer, while green lines stand for minor *i2* isomer.

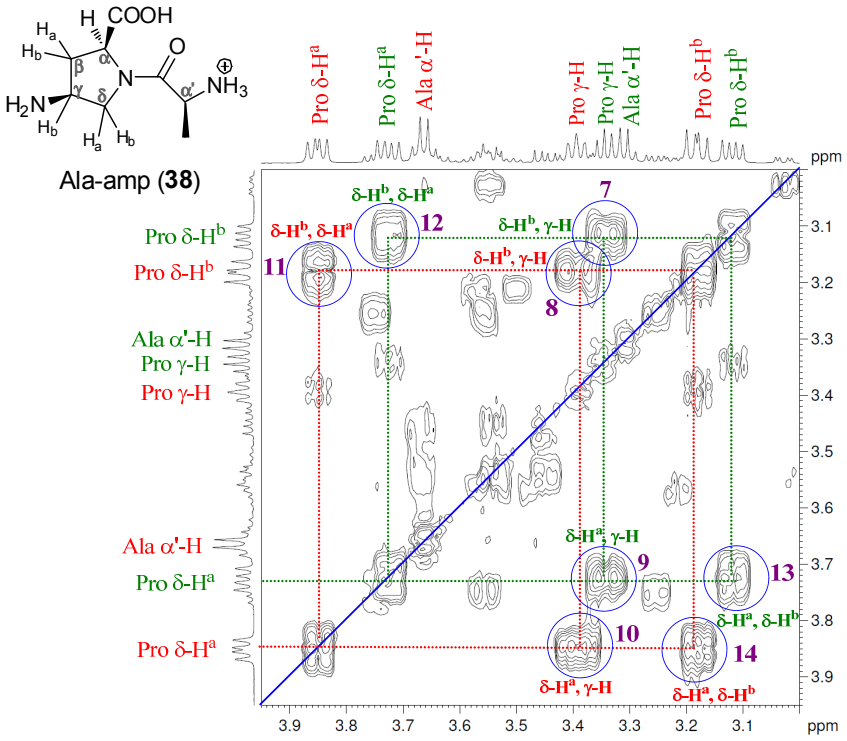


FIGURE 20: Selected view of 2D ^{1}H-^{1}H COSY NMR spectra of **38** at pD 12.0

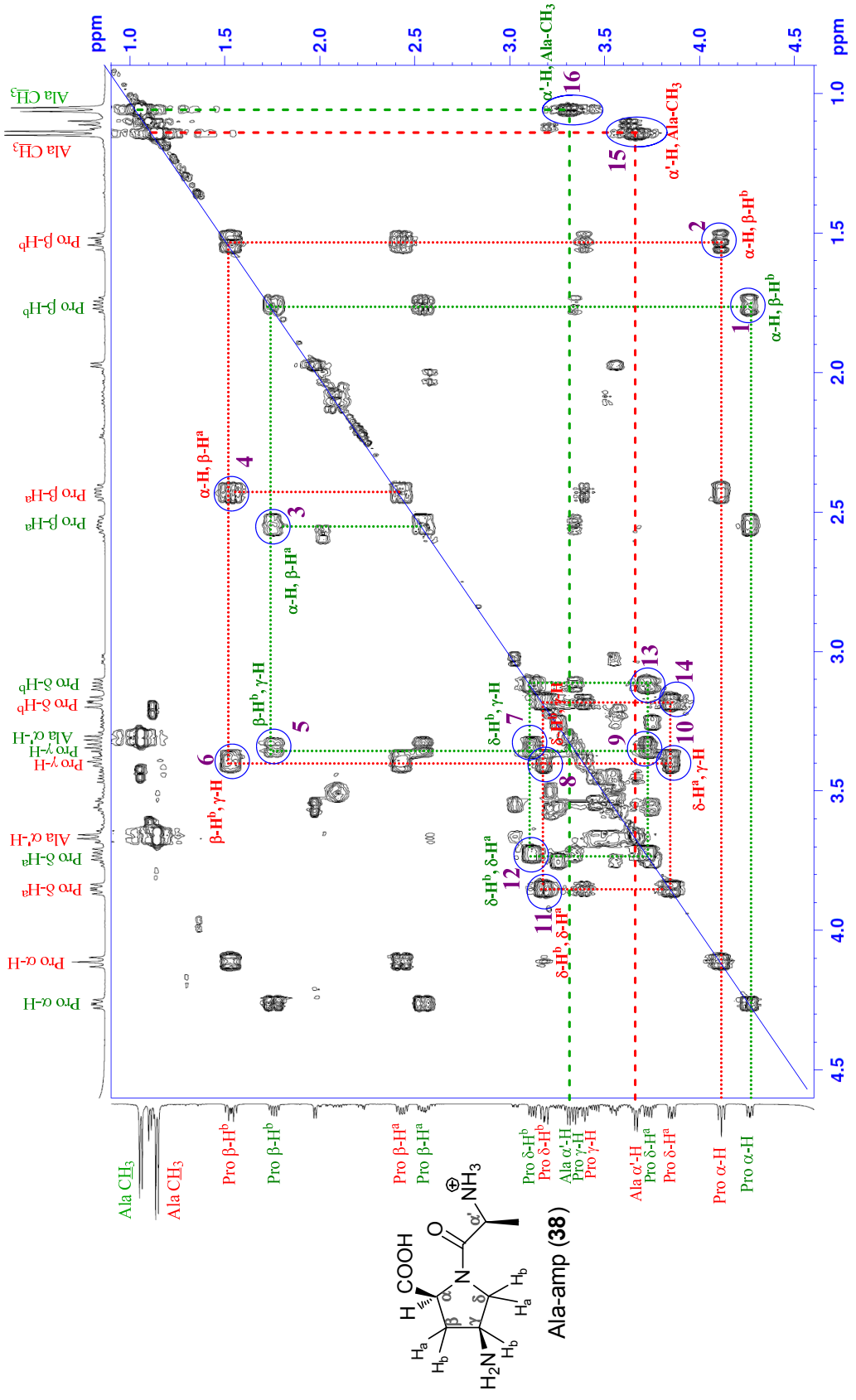


FIGURE 21: 2D ^{1}H-^{1}H COSY NMR spectra of pyrrolidine ring protons in compound **38** (500 MHz) in D_2O at pD 12.0; The numbers 1-16 correspond to assigned cross peaks; Color code: Red–major isomer (*i1*), Green–minor isomer (*i2*).

At pD 12 and room temperature, separate sets of cross peaks were seen to connect the protons of each isomer separately, which proves that at basic pH equal population of the two isomers co-exist and have enough $t_{1/2}$ to be distinctly detected by NMR at room temperature.

2.5.3b NOESY : Characterization of *cis* & *trans* isomers

The intraresidue assignment of various protons in proline through ^{1}H-^{1}H COSY and alanine residues of the dipeptide should lead to the identification of *cis* and *trans* isomers (rotamers) through inter-residue ^{1}H-^{1}H NOESY spectra. Figure **22** shows the structure of the *cis* and *trans* isomers of the dipeptide **19**.

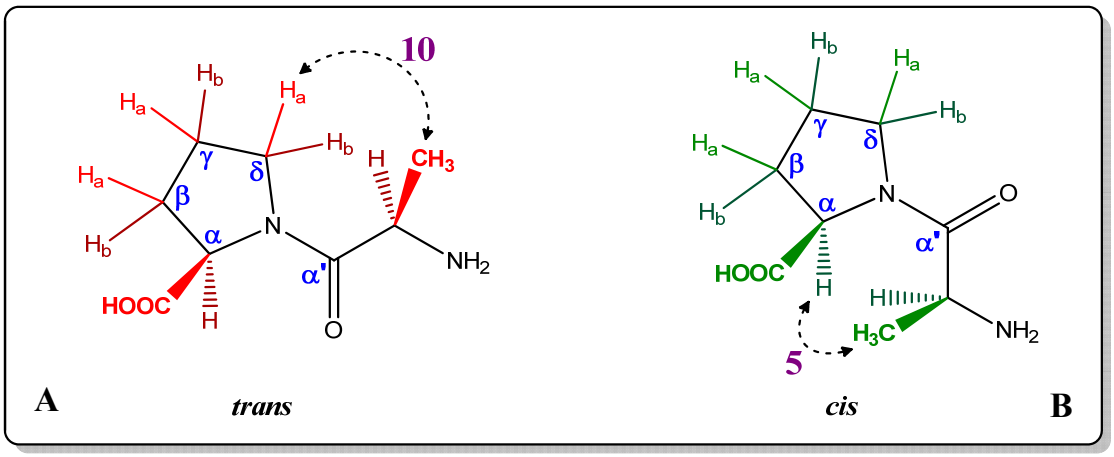


FIGURE 22: Expected NOE interaction in case of Ala-Pro dipeptide **19**, as (**A**) *trans* & (**B**) *cis* isomers; the numbers over the dotted arrows indicate the cross peak no. assigned to the interaction in the 2D ^{1}H-^{1}H NOESY spectrum.

It can be seen that in *trans* isomer (Figure **22.A**), Ala-C<u>H</u>$_3$ is close to one of the δ-hydrogens (H_a or H_b) of the proline while in *cis* isomer (Figure **22.B**), the Ala-C<u>H</u>$_3$ is nearer to the α-H of proline. Hence in the ^{1}H-^{1}H NOESY spectrum the corresponding inter-residue connection can be employed for assignment which becomes the basis for assignment of *cis* & *trans* isomers.

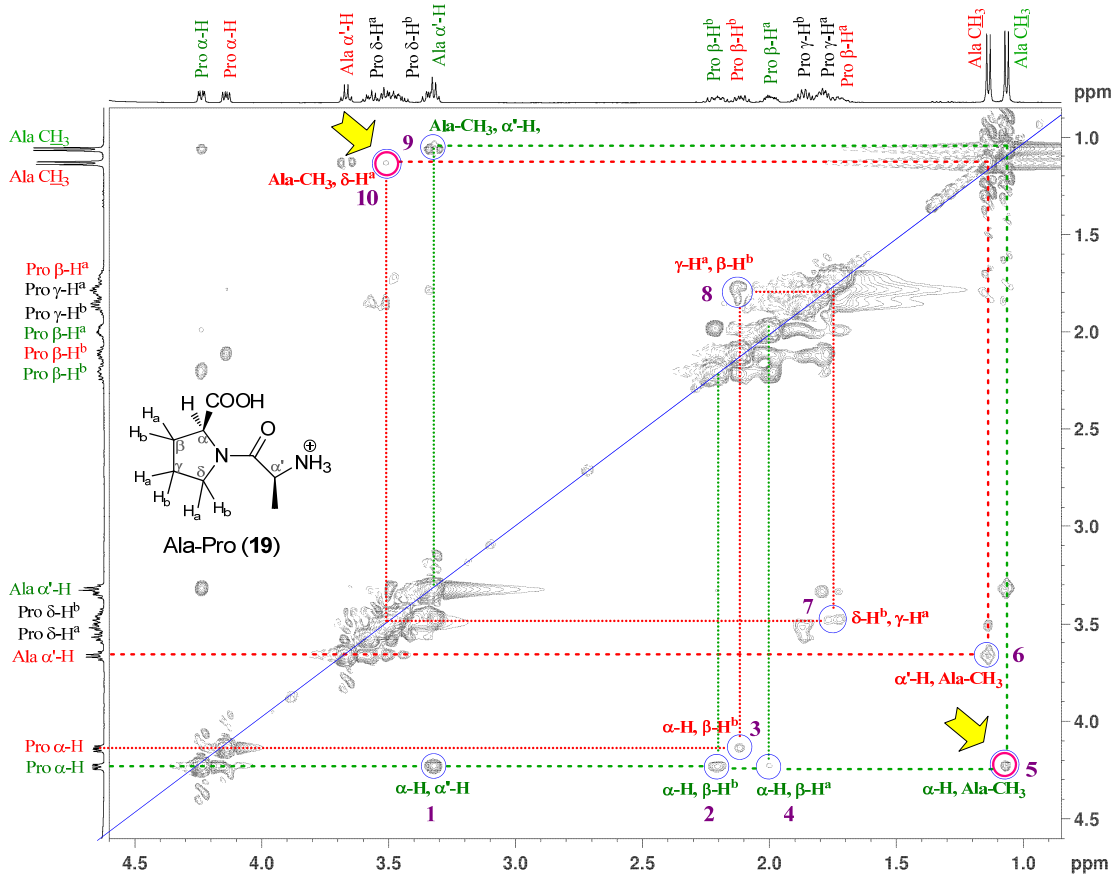


FIGURE 23: 2D ^{1}H-^{1}H NOESY NMR spectra of pyrrolidine ring protons in compound **19** (500 MHz) in D_2O at pD 12.0; The numbers 1-10 correspond to assigned cross peaks; Color code: Green – *cis* isomer, Red – *trans* isomer.

The identity of isomers as to which is *cis* and which is *trans* was determined from 2D NOESY spectra of compound **19** (Figure **23**), in which the two observed cross peaks 5 (Pro-αH, Ala-CH_3) and 10 (Ala-CH_3, δ-H^a) indicate the expected spatial correlation in each of the two isomers. Isomer *i1* shows proximity of alanine methyl protons with one of the δ-protons of proline, while the isomer *i2* shows expected connectivity for *cis* isomer between α-H of proline with methyl protons of alanine (cross peak 5). Hence by virtue of observation of these two cross peaks, the upfield component of Ala-CH_3 at δ 1.11 was assigned to be due to *cis* isomer, while the downfield component at δ 1.18 arises from the *trans* isomer.

Once the isomers *i1* and *i2* were assigned to be *trans* and *cis* isomers respectively, this analogy was extrapolated to assign the major (*ma*) and minor (*mi*) isomers in ^{1}H-NMR of dipeptide at other pHs' as the *trans* and *cis* isomers respectively in that order.

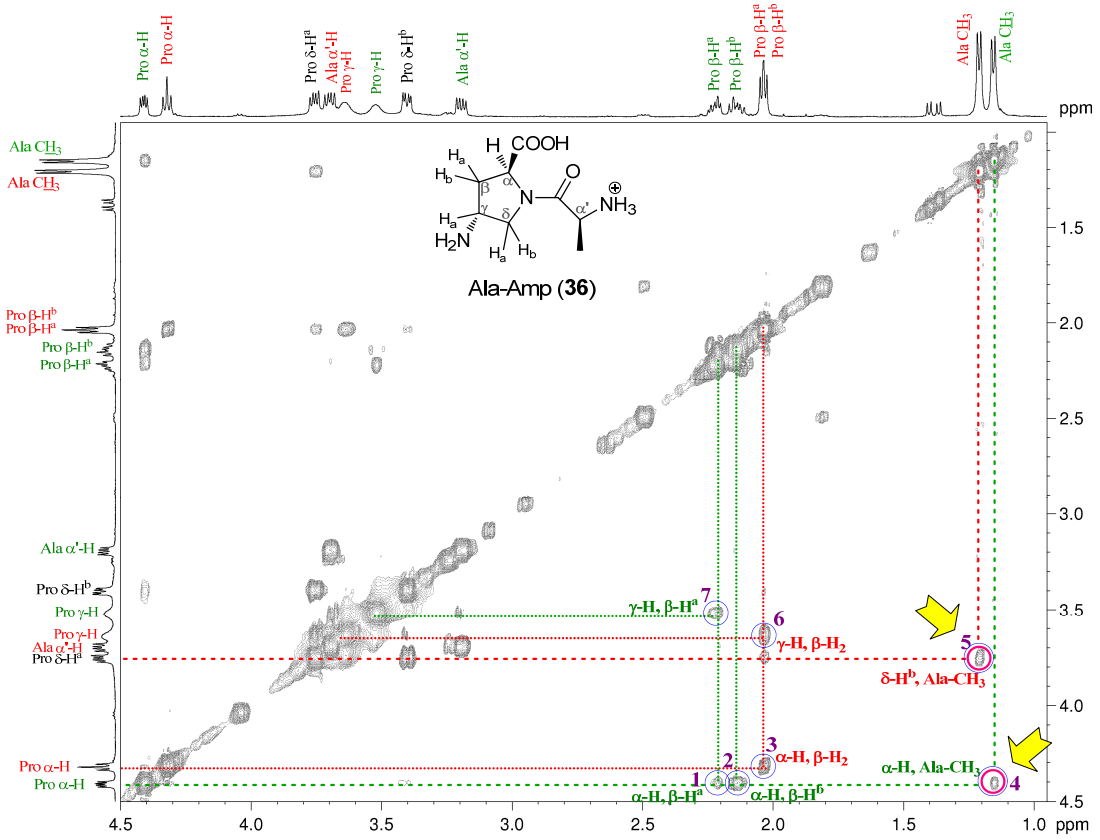


FIGURE 24: 2D ^{1}H-^{1}H NOESY NMR spectra of pyrrolidine ring protons in compound **36** (500 MHz) in D_2O at pD 12.0; The numbers 1-7 correspond to assigned cross peaks; Color code: Green – *cis* isomer, Red – *trans* isomer.

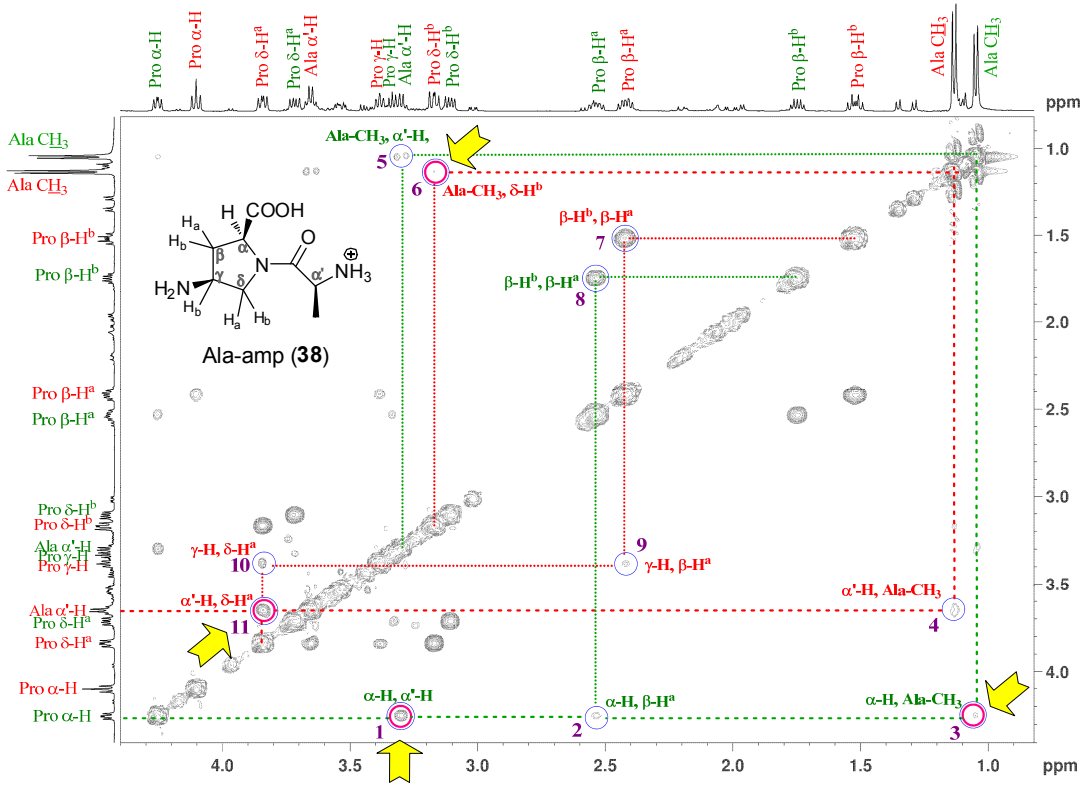


FIGURE 25: 2D ^{1}H-^{1}H NOESY NMR spectra of pyrrolidine ring protons in compound **38** (500 MHz) in D_2O at pD 12.0. The numbers 1-11 correspond to assigned cross peaks; Color code: Green– *cis* isomer, Red– *trans* isomer.

In the dipeptide Ala-Amp **36**, the 2D ^{1}H-^{1}H NOESY spectrum (Figure **24**) shows only two inter-residue cross peaks 4 (α-H, Ala-CH_3) for one isomer and 5 (δ^b-H, Ala-CH_3) for the other isomer. This indicates that in this instance, the alanyl methyl group orients itself to towards C^{α} of proline in one isomer and in other isomer it rotates itself towards the C^{δ}. Thus, going by the previous hypothesis as applied to dipeptide **19**, the two isomers were assigned to be *trans* and *cis* respectively.

In the case of Ala-amp **38**, the number of inter-residue cross peaks observed in NOESY spectrum was much higher. Five cross peaks 1 (α-H, α'-H), 3 (α-H, Ala-CH_3), 4 (α'-H, Ala-CH_3,), 6 (Ala-CH_3, δ-H^b) and 7 (α'-H, δ-H^a) accounting for the through space ^{1}H-^{1}H couplings indicated in Figure **25.** This helped in assignment of *trans* and *cis* isomers of the dipeptide **38**. Thus NOESY interactions (Figure **26**) giving rise to cross peak 4, 6 and 7 come from *trans* isomer, while cross peaks 1 and 3 are from *cis* isomer.

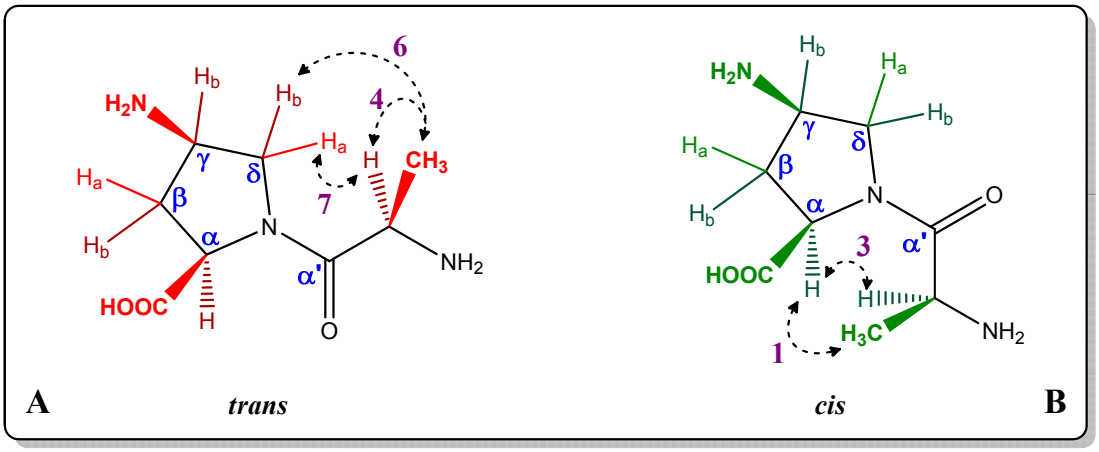


FIGURE 26: Expected NOE interaction in case of Ala-amp dipeptide **38**, as (**A**) *trans* & (**B**) *cis* isomers; The numbers over the dotted arrows indicate the cross peak no. assigned to the interaction in the 2D ^{1}H-^{1}H NOESY spectrum.

The assignment of vicinal J values for proline ring protons should in principle allow determination of the proline ring pucker (*exo*, *endo*, *twist etc.*) for each of the *trans* and *cis* isomers). However due to severe overlap of different multiplets and lack of resolution did not permit such an analysis in the present work. Nevertheless some components of coupling constants could be extracted from the different spectra and Table **3(A-C)** indicates the chemical shift and J values for dipeptides **19**, **36** and **38** at different pHs (2.2, 4.2, 7.0, 9.5 and 12.0).

Table 3.A: Chemical shift and splitting pattern of ^{1}H-NMR signals for dipeptide **19** at different pHs

Assignment		^{1}H Chemical shifts				
		pH 2.2	**pH 4.2**	**pH 7.0**	**pH 9.5**	**pH 12**
Pro-αH	*trans*	4.40-4.43 (dd, J=8.7, 5.2 Hz)	4.23 (dd, J=8.5, 6.1 Hz)	4.20-4.22 (dd, J=8.4, 6.0 Hz)	4.17-4.20 (dd, J=8.6, 5.2 Hz)	4.17-4.20 (dd, J=8.6, 5.1 Hz)
	cis	4.45-4.48 (dd, J=8.5, 2.5 Hz)	4.27-4.29 (m)	4.23-4.28 (m)	4.27-4.29 (dd, J=8.7, 3.2 Hz)	4.27-4.29 (dd, J=8.7, 3.2 Hz)
Pro-βH$_a$	*trans*	1.94-2.02 (m)	1.72-1.79 (m)	1.73-1.80 (m)	1.72-1.79 (m)	1.74-1.79 (m)
	cis		2.06-2.13 (m)	2.08-2.13 (m)	2.02-2.08 (m)	2.01-2.08 (m)
Pro-βH$_b$	*trans*	2.25-2.33 (m)	2.19-2.31 (m)	2.17-2.30 (m)	2.13-2.20 (m)	2.13-2.20 (m)
	cis				2.21-2.30 (m)	2.21-2.30 (m)
Pro-γH$_a$		1.94-2.02 (m)	1.83-1.99 (m)	1.83-1.99 (m)	1.81-1.86 (m)	1.81-1.86 (m)
Pro-γH$_b$		1.94-2.02 (m)	1.83-1.99 (m)	1.83-1.99 (m)	1.88-1.96 (m)	1.88-1.97 (m)
Pro-δH$_a$	*trans*	3.62-3.66 (m)	3.59-3.64 (m)	3.58-3.63 (m)	3.47-3.64 (m)	3.47-3.65 (m)
	cis	3.34-3.43 (m)	3.34-3.43 (m)	3.34-3.43 (m)		
Pro-δH$_b$	*trans*	3.55-3.60 (m)	3.52-3.57 (m)	3.51-3.57 (m)	3.35-3.38 (m)	3.39-3.41 (m)
	cis	3.45-3.50 (m)	3.45-3.50 (m)	3.44-3.50 (m)		
Ala-α'H	*trans*	4.30 (q, J=7.0 Hz)	4.27-4.29 (m)	4.23-4.28 (m)	3.78 (q, J=6.7 Hz)	3.71 (q, J=6.8 Hz)
	cis	3.98 (q, J=6.9 Hz)	3.91 (q, J=6.9 Hz)	3.91 (q, J=6.9 Hz)	3.38-3.43 (m)	3.35-3.37 (m)
Ala-C$\underline{H}_3$	*trans*	1.47 (d, J=7.0Hz)	1.48 (d, J=7.0Hz)	1.47 (d, J=7.0Hz)	1.22 (d, J=6.8 Hz)	1.18 (d, J=6.8 Hz)
	cis	1.41 (d, J=6.9 Hz)	1.41 (d, J=6.9 Hz)	1.40 (d, J=6.9 Hz)	1.13 (d, J=6.7 Hz)	1.11 (d, J=6.7 Hz)

Table 3.B: Chemical shift and splitting pattern of ^{1}H-NMR signals for dipeptide **36** at different pHs.

Assignment		^{1}H Chemical shifts				
		pH 2.2	**pH 4.2**	**pH 7.0**	**pH 9.5**	**pH 12**
Pro-αH	*trans*	4.64-4.67 (m)	4.44 (d, J=7.9 Hz)	4.42 (d, J=7.8 Hz)	4.27-4.32 (m)	4.29 (t, J=7.5 Hz)
	cis		4.48-4.50 (dd, J=8.7, 3.6 Hz)	4.48 (d, J=5.4 Hz)	4.38-4.41 (dd, J=8.4, 4.5 Hz)	4.37-4.40 (dd, J=8.6, 4.6 Hz)
Pro-βH_a	*trans*	2.43-2.48 (m)	2.43-2.47 (m)	2.40-2.44 (m)	2.00-2.07 (br m)	2.01 (t, J=6.5 Hz)
	cis		2.56-2.60 (m)	2.55-2.58 (br m)	2.20-2.24 (m)	2.17-2.22 (m)
Pro-βH_b	*trans*	2.50-2.55 (m)	2.27-2.32 (m)	2.24-2.29 (m)	2.00-2.07 (br m)	2.01 (t, J=6.5 Hz)
	cis		2.49-2.53 (m)	2.47-2.51 (m)	2.11-2.17 (m)	2.09-2.15 (m)
Pro-γH_a	*trans*	4.12-4.16 (m)	4.07-4.12 (br m)	3.97-4.07 (m)	3.52-3.65 (m)	3.57-3.61 (m)
	cis		4.00-4.03 (m)			3.47-3.51 (m)
Pro-δH_a	*trans*	3.98-4.02 (dd, J=12, 5.9 Hz)	3.90-3.98 (m)	3.89-3.92 (m)	3.71-3.75 (m)	3.65-3.69 (dd, J=11.9, 6.5 Hz)
	cis					
Pro-δH_b	*trans*	3.88-3.92 (m)	3.80-3.83 (dd, J=12, 2.4 Hz)	3.78 (br d, J=10.8 Hz)	3.47-3.51 (m)	3.37-3.40 (dd, J=10.7, 4.3 Hz)
	cis	3.83-3.86 (dd, J=11.9, 3.2 Hz)	3.90-3.98 (m)		3.40 (d, J=7.3 Hz)	
Ala-α'H	*trans*	4.33 (q, J=7 Hz)	4.3 (q, J=7 Hz)	4.28 (d, J=5.3 Hz)	3.67-3.70 (m)	3.71-3.75 (dd, J=10.7, 5.7 Hz)
	cis	4.07 (q, J=6.9 Hz)	3.55-3.63 (m)	3.55-3.61 (m)	3.18-3.21 (dd, J= 11.9, 5.6 Hz)	3.15-3.19 (dd, J=11.9, 5.8 Hz)
Ala-C$\underline{H}_3$	*trans*	1.47 (d, J=7 Hz)	1.49 (d, J=7 Hz)	1.45 (br s)	1.24 (br s)	1.18 (d, J=6.7 Hz)
	cis	1.44 (d, J=7 Hz)	1.45 (d, J=6.9 Hz)	1.44 (d, J=5.8 Hz)	1.16 (d, J=5.1 Hz)	1.13 (d, J=6.7 Hz)

Table 3.C: Chemical shift and splitting pattern of ^{1}H-NMR signals for dipeptide **38** at different pHs.

Assignment		^{1}H Chemical shifts				
		pH 2.2	**pH 4.2**	**pH 7.0**	**pH 9.5**	**pH 12**
Pro-αH	*trans*	4.39-4.43 (m)	4.29-4.32 (dd, J=9.5, 3.9 Hz)	4.27-4.30 (dd, J=9.4, 4.2 Hz)	4.16 (t, J=8.1 Hz)	4.15 (t, J=8.3 Hz)
	cis	4.5 (t, J=7.8 Hz)	4.36-4.39 (dd, J=9.4, 2.5 Hz)	4.35-4.38 (dd, J=9.4, 2.6 Hz)	4.35-4.38 (dd, J=8.9, 4.7 Hz)	4.28-4.31 (dd, J=8.0, 6.2 Hz)
Pro-βH$_a$	*trans*	2.69-2.82 (m)	2.60-2.66 (dq, J=15.1, 9.3, 6 Hz)	2.58-2.64 (m)	2.44-2.51 (m)	2.42-2.49 (m)
	cis		2.69-2.80 (m)	2.72-2.78 (m)	2.57-2.65 (m)	2.55-2.64 (m)
Pro-βH$_b$	*trans*	2.07-2.12 (dt, J=13.7, 5.2 Hz)	2.05-2.09 (dt, J=14.3, 3.6 Hz)	2.00-2.04 (dt, J=14.3, 3.8 Hz)	1.59-1.65 (dt, J=13, 7.7 Hz)	1.54-1.59 (dt, J=12.7, 8.3 Hz)
	cis	2.29 (d, J=14.7 Hz)	2.28 (br d, J=14.7 Hz)	2.24 (br d, J=14.6 Hz)	1.82-1.88 (m)	1.77-1.82 (dt, J=12.9, 6.4 Hz)
Pro-γH$_b$	*trans*	4.05-4.12 (m)	4.04-4.08 (m)	3.98-4.01 (m)	3.48-3.53 (m)	3.43 (t, J=7.4 Hz)
	cis					3.38 (t, J=6.4 Hz)
Pro-δH$_a$	*trans*	3.99-4.03 (m)	3.87-3.91 (m)	3.86-3.97 (m)	3.87-3.91 (dd, J=10.6, 6.6 Hz)	3.87-3.90 (dd, J=12.1, 6.6 Hz)
	cis	3.90-3.94 (m)	3.97-4.00 (m)		3.75-3.79 (dd, J=12, 6.6 Hz)	3.74-3.78 (dd, J=10.3, 6.7 Hz)
Pro-δH$_b$	*trans*	3.72-3.83 (m)	3.77-3.83 (m)	3.70-3.81 (m)	3.26-3.30 (dd, J=10.5, 7.3 Hz)	3.20-3.23 (dd, J=10.3, 7.8 Hz)
	cis				3.22-3.24 (dd, J=12.1, 5.7 Hz)	3.13-3.17 (dd, J=12.0, 6.2 Hz)
Ala-α'H	*trans*	4.28 (q, J=6.9 Hz)	4.25 (q, J=7 Hz)	4.20 (q, J=6.7 Hz)	3.7-3.79 (dd, J=12, 6.6 Hz)	3.32-3.35 (m)
	cis	4.05 (br s)	4.01-4.03 (m)	3.98-4.01 (m)	3.42-3.47 (m)	3.70 (q, J=6.9 Hz)
Ala-C$\underline{H}_3$	*trans*	1.48 (d, J=7 Hz)	1.48 (d, J=6.9 Hz)	1.45 (d, J=7 Hz)	1.21 (d, J=6.9 Hz)	1.18 (d, J=6.8 Hz)
	cis	1.44 (d, J=7 Hz)	1.44 (d, J=6.9 Hz)	1.41 (d, J=6.9 Hz)	1.12 (d, J=6.7 Hz)	1.09 (d, J=6.8 Hz)

2.5.4 pH dependent NMR analysis of dipeptides

As described in the litereture for Ala-Pro thiopeptides[48], the relative ratios of *cis* & *trans* rotamers depend on pH and can be determined from change in relative ratios of the doublets of α'-CH_3 of alanine as a function of pH. Hence pH dependent ^{1}H NMR of different peptides (**19**, **23**, **28**, **31**, **34**, **36**, **38**, **41** and **44**) were recorded at pHs 2.2, 4.2, 7.0, 9.5, 12.0. Chemical shift assignments were done following methods described in previous section. Figure **27** shows a typical stack plot of ^{1}H NMR of dipeptide **19** at pHs 2, 4, 7, 9.5, 12 and Table **4** gives the ratios of *trans* isomers in each case determined from change in Ala-CH_3 intensities.

Dipeptides (**19** & its 4-substituted derivatives) were investigated by 1D ^{1}H NMR at five different pH conditions. The dipeptides were dissolved in D_2O and the pD value was altered in ascending order with quantitative addition of TFA-d_4 and NaOD. The pD was measured directly in the NMR tube with the help of a specialised electrode introduced into NMR tube and attached to a pH meter.

The specific observations obtained from the stacked ^{1}H-NMR plot of dipeptide **19** (Figure **27**) taken at 500 MHz are:

➔ At acidic pH (2.2, 4.2), *trans* > *cis* (90:10) while at pH 12.0, *trans* ≈ *cis* (50:50)

➔ With increase in pH, going towards alkaline region

- Most of the peaks shift upfield, although to different extent.
- Methyl protons of alanine and alpha proton of proline are the most affected.
- The multiplets due to different non equivalent geminal protons are much better resolved
- The ratio of *cis* : *trans* isomer changes gradually.

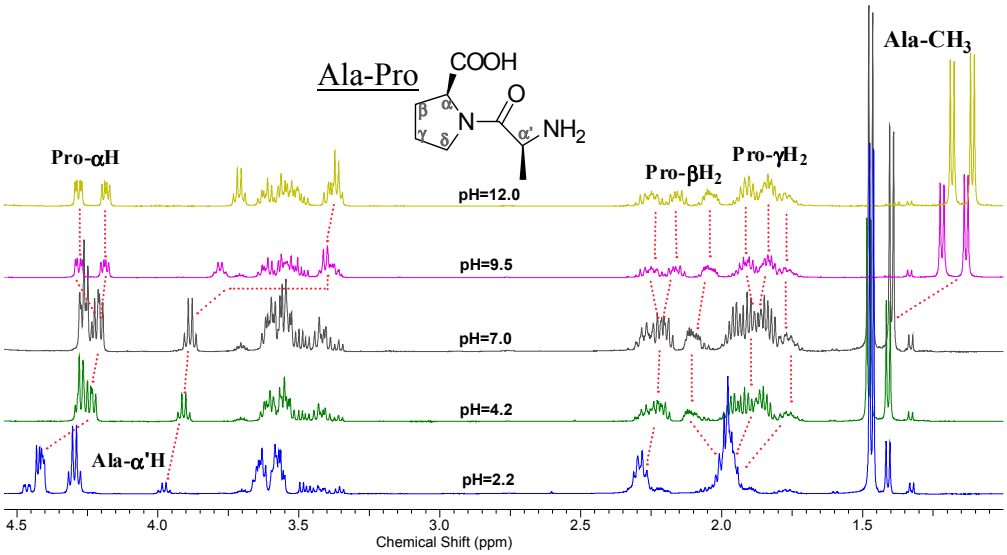


FIGURE 27: Stacked plot of 1D ^{1}H-NMR spectra of dipeptide **19** (500 MHz) in D_2O at various pD/pH.

The other 4-substituted proline dipeptides showed a similar trend in the observed ^{1}H-NMR spectra at different pHs. This is illustrated in ^{1}H-NMR spectra (Figure **28** – **33**) of peptides (**23**, **28**, **31**, **34**, **36** and **38**) at the two extreme pH values 2 and 12. The NMR spectra at other pHs are shown in Appendix B.

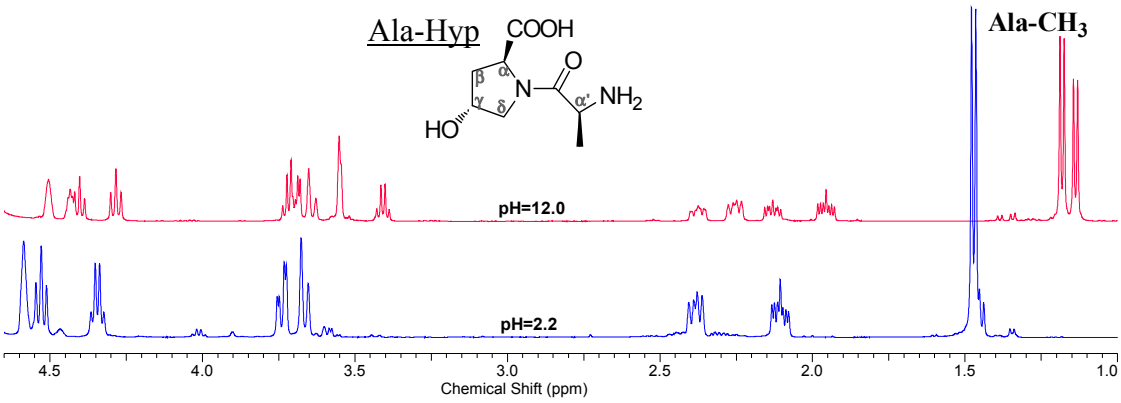


FIGURE 28: Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide **23** (500 MHz) in D_2O at pD 2.2 and 12.0.

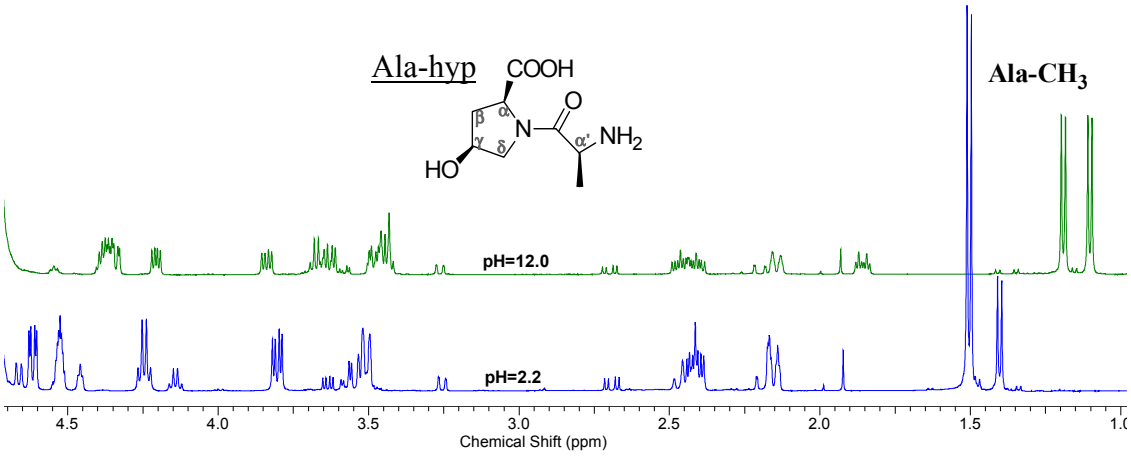


FIGURE 29: Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide **28** (500 MHz) in D_2O at pD 2.2 and 12.0.

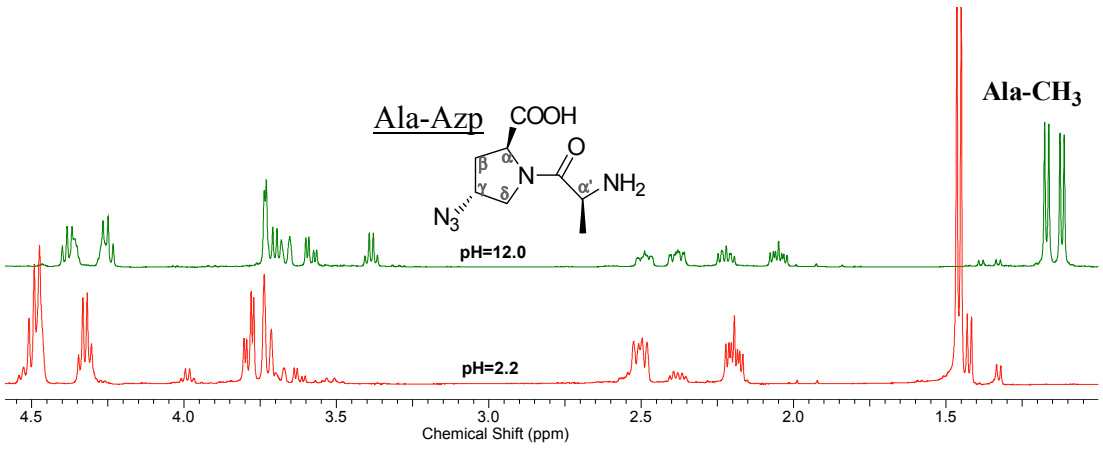


FIGURE 30: Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide **31** (500 MHz) in D_2O at pD 2.2 and 12.0.

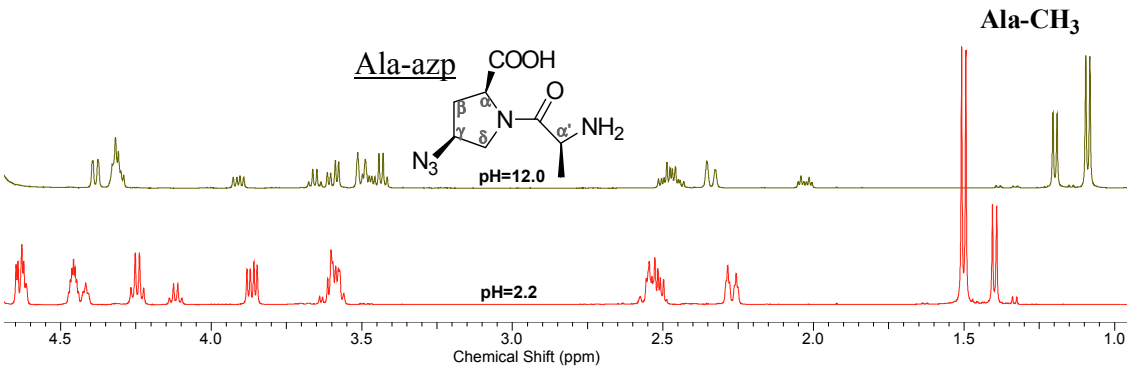


FIGURE 31: Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide **34** (500 MHz) in D_2O at pD 2.2 and 12.0.

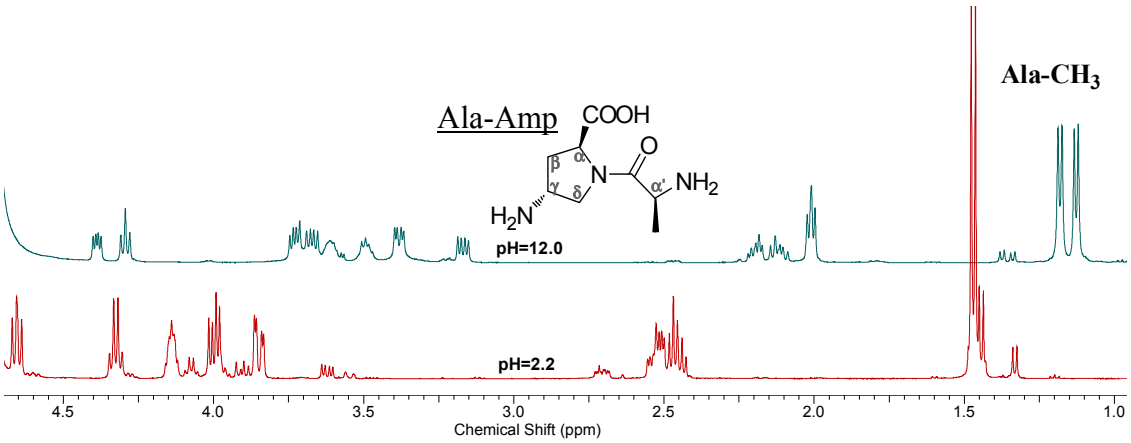


FIGURE 32: Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide **36** (500 MHz) in D_2O at pD 2.2 and 12.0.

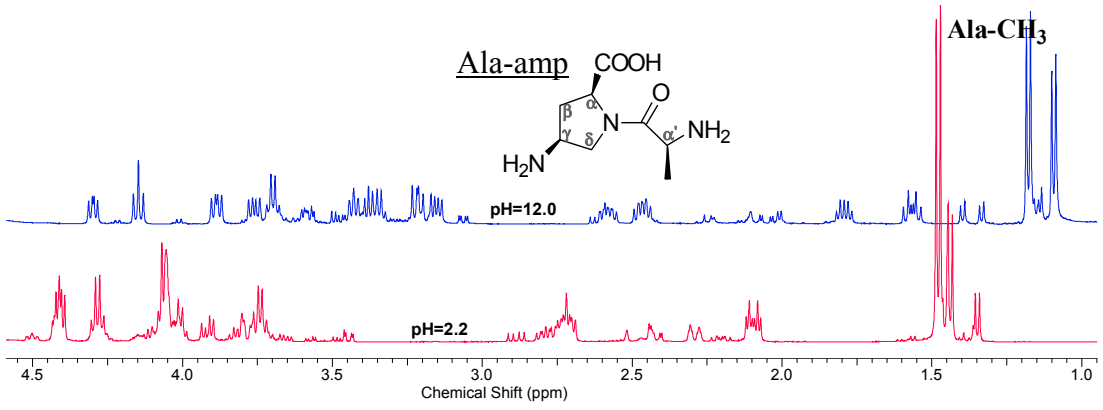


FIGURE 33: Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide **38** (500 MHz) in D_2O at pD 2.2 and 12.0.

The methyl doublet of the alanyl residue is a clear marker, since the α'–H of alanine is the most affected atom in peptide bond flipping. Hence the *trans* : *cis* ratio could be easily quantified by analysis of ^{1}H-NMR following the intensity of doublet of CH_3 of alanine residue.

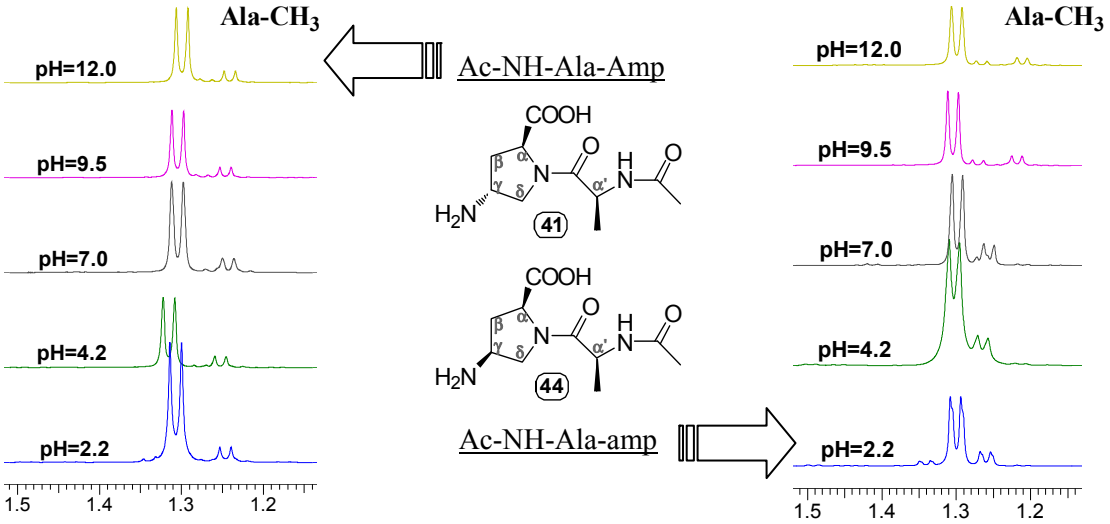


FIGURE 34: Stacked plot of 1D ^{1}H-NMR spectra of Dipeptide **41** and **44** (500 MHz) in D_2O at various pD.

The dipeptides (**19**, **23**, **28**, **31**, **34**, **36** and **38**) contain free NH_2 groups at N-terminus, which can ionize as a function of pH. This may also influence the *trans* : *cis* isomer ratio. To examine this contribution, similar data was recorded for the Ala-N-capped peptides **41** and **44** (Figure **34**). Interestingly it is seen that the *trans* : *cis* ratios do not vary as a function of pH, with *trans* isomer always dominating (~90%).

Thus the *trans* isomer content of all the peptides was determined from the peak areas of the *cis* and *trans* signals through doublets of α'-CH_3 of alanine in ^{1}H-NMR spectra obtained at all pHs (2.2, 4.2, 7.0, 9.5 and 12.0) in deuterated-aqueous solution (Table **4**). The data were verified by evaluation of ^{13}C-NMR spectra, having the signals of proline ring carbons well resolved for *cis* and *trans* isomers (Apendix B).

Table 4: Variation of *trans* isomer content as a function of pH in various dipeptides.

Dipeptide		% of *trans* Isomer				
		pH 2.2	pH 4.4	pH 7.0	pH 9.5	pH 12.0
19	Ala-Pro	*87*	*63*	*61*	*50*	*49*
23	Ala-Hyp	*89*	*73*	*72*	*58*	*52*
28	Ala-hyp	*78*	*62*	*55*	*52*	*51*
31	Ala-Azp	*87*	*69*	*66*	*53*	*53*
34	Ala-azp	*72*	*42*	*39*	*36*	*36*
36	Ala-Amp	*80*	*59*	*55*	*53*	*52*
38	Ala-amp	*68*	*60*	*59*	*59*	*54*
41	Ac-Ala-Amp	*88*	*88*	*88*	*88*	*88*
44	Ac-Ala-amp	*83*	*81*	*81*	*90*	*91*

The 4*R*-substituted dipeptides showed a trend more or less as that of (unsubstituted) dipeptide Ala-Pro **19**, while the 4*S*-substituted dipeptides showed a slightly different result with relatively higher predominance of the *cis*-isomers. A comparative plot of 4*R*- *versus* 4*S*-diastereomers along with control dipeptide Ala-Pro is given in Figure **35**.

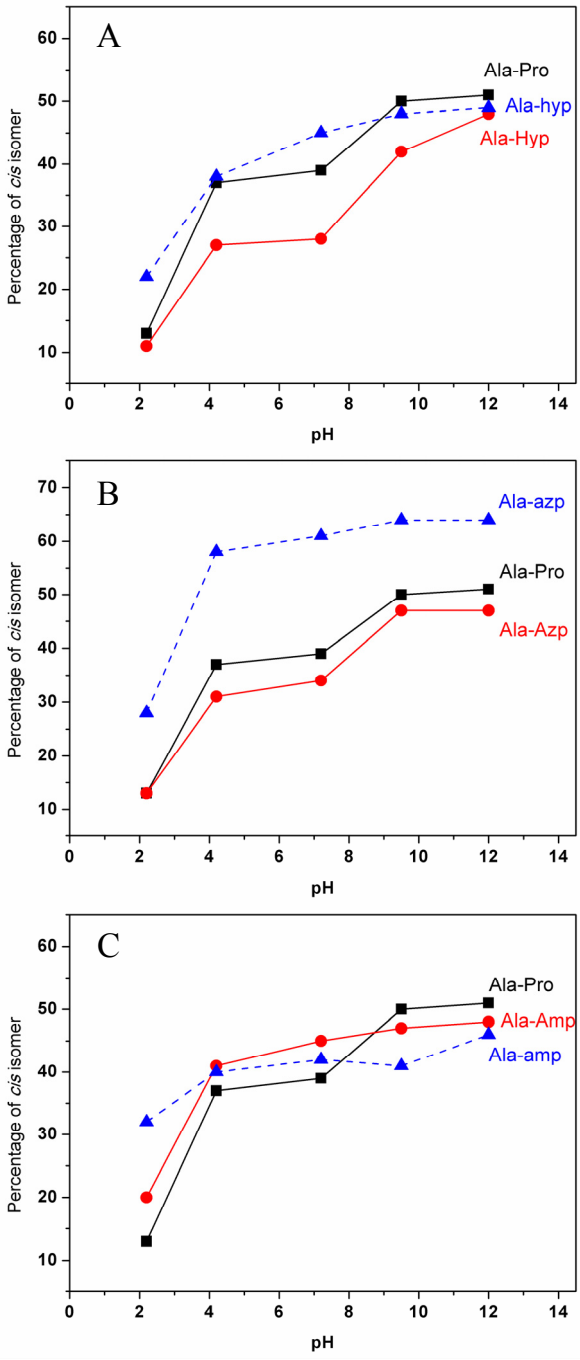


FIGURE 35: Variation plot of *cis*-isomers with pH for the dipeptides; (**A**) 4-Hydoxy dipeptides, (**B**) 4-Azido dipeptides, (**C**) 4-amino dipeptides; In the plot black squares (■) represent Ala-Pro dipeptide, **19**; red circles (●) represent 4*R*-isomer of dipeptides and blue triangles (▲) represent 4*S*- isomers.

2.5.5 pK_a determination of Ala-Pro model dipeptides

In order to examine whether the pH dependence of *cis-trans* isomer ratio in 4(*R/S*)-aminoproline dipeptide on the protonation status of the 4-amino substituent, the pK_a of various ionizable groups were determined. This can be conveniently determined by acid/base titration.[51] Typically, the sample of interest is first brought to either high alkaline or low acidic pH and then titrated with acid or base by titrating with fixed volume aliquots. After the addition of each aliquot, the pH was measured. The pH of the sample solution changes rapidly at the pK_a of the functional group. In a pH titration experiment the number of pH transitions are equal to the number of ionizable groups with different pK_a values.

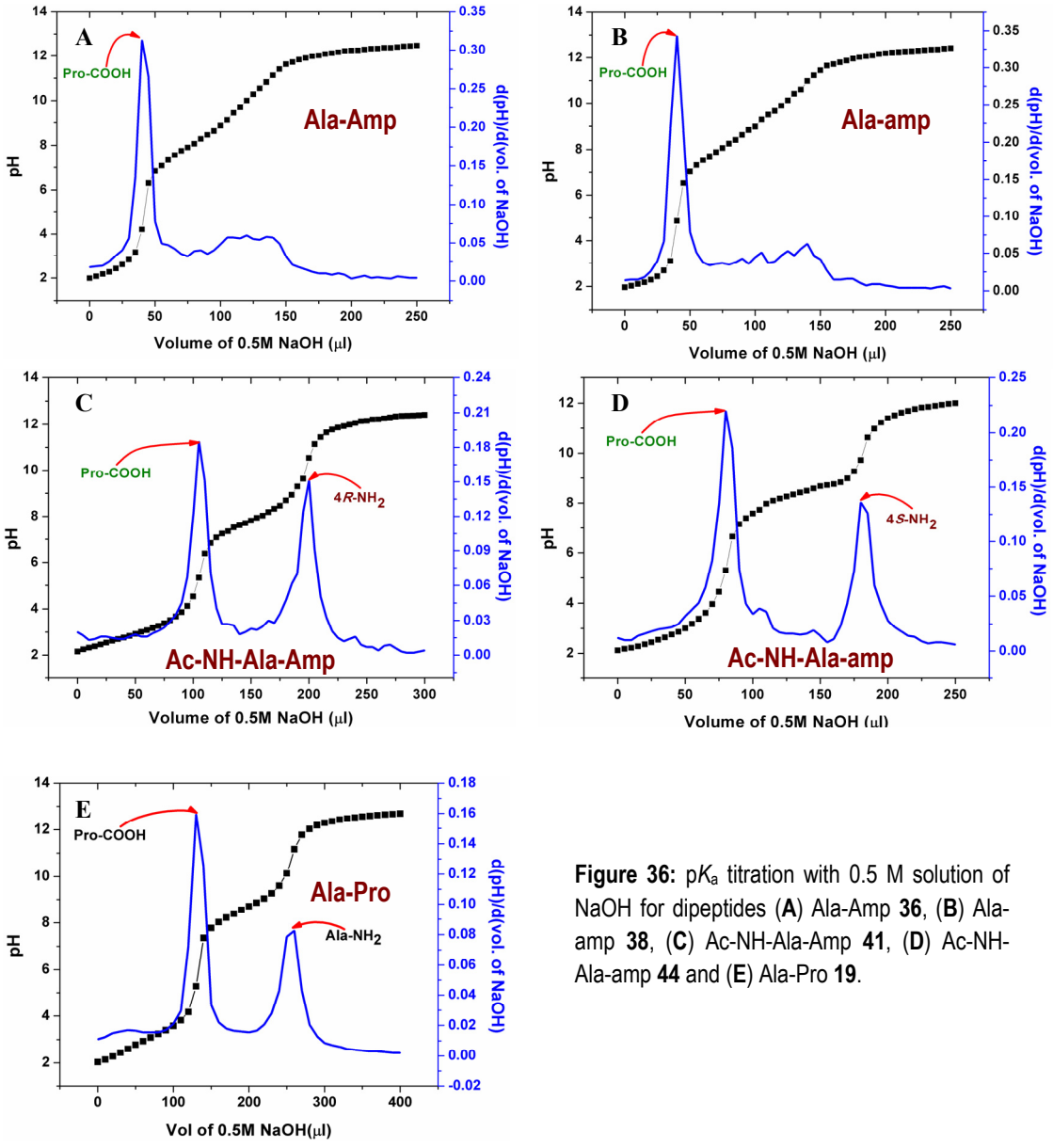


Figure 36: pK_a titration with 0.5 M solution of NaOH for dipeptides (**A**) Ala-Amp **36**, (**B**) Ala-amp **38**, (**C**) Ac-NH-Ala-Amp **41**, (**D**) Ac-NH-Ala-amp **44** and (**E**) Ala-Pro **19**.

The pH titration plot of 4(*R/S*)-amino substituted Ala-Pro dipeptides **36** and **38** (0.1 M) with NaOH (0.5 M) using direct pH meter reading shows only one sharp pH transition around 4.2 to 4.9 for the ionization of α-carboxyl group of proline. The presence of two free amino groups in the molecule with perhaps closer pKa value lends to broad (unresolved) transition not enough to be detected manually. N-terminal acetylated (α'-NH_2 of Ala) dipeptides were used to determine the pK_a of 4(*R/S*)-NH_2 groups.

In case of $N^{\alpha'}$-acetyl-Ala-Amp (**41**), one peak at pK_a of 5.3 corresponding to the α-carboxylic acid group, and a second transition at pK_a of 10.5 for 4*R*-amino group were observed (Figure **36**). For the $N^{\alpha'}$-acetyl-Ala-amp, peaks at 5.3 (for α-COOH) and 9.9 (for 4*S*-NH_2) were seen. Titration of unsubstituted dipeptide Ala-Pro **19**, gave the pK_a values for carboxylic acid of proline and amino group of alanine as 5.3 and 11.2 respectively. Table **5** gives a comparative outline of all the pK_a values determined at dipeptide level. However, it may be noted that the pK_a of 4-amino group may be slightly different when it is part of a larger peptide chain.

Table 5: pK_a values obtained for titrated dipeptides

Figure	Dipeptide		pK_a	
			COOH	NH_2
A	Ala-Amp	(**36**)	4.20	ND
B	Ala-amp	(**38**)	4.87	ND
C	Ac-NH-Ala-Amp	(**41**)	5.33	10.54 / 10.5*
D	Ac-NH-Ala-amp	(**44**)	5.30	9.72 / 9.3*
E	Ala-Pro	(**19**)	5.28	11.17

*pK_a values measured in N^{α}-acetyl-4(*R/S*)-aminoprolines individually

It is seen that the 4*S*-amino are more acidic ($\Delta pK_a \approx 0.8$) compared to the 4*R*-amino group and found close to the values obtained for their corresponding N^a-acetyl-(*R/S*) -aminoprolines.

2.5.6 Conformational review by CD spectroscopy

The conformation of dipeptides was analysed by CD spectroscopy. The CD spectra of dipeptide (5 mM) in water were recorded from 190 to 260 nm wavelengths. The dipeptides were seen to have varied CD pattern. On measuring the pH of these solutions, it was found that they had pH less than equal to 2.0. Subsequently CD spectra of each of the dipeptide were recorded at increasing pH values, at which its NMR spectra were recorded. Figure **37(A-B)** shows a comparative plot of all

dipeptides in water (at their original pH values) and a variable pH CD-spectra of unsubstituted dipeptide Ala-Pro **19**. The Ala-Pro dipeptide shows a random coil pattern CD in high acidic region indicated by a strong negative band around 200 –

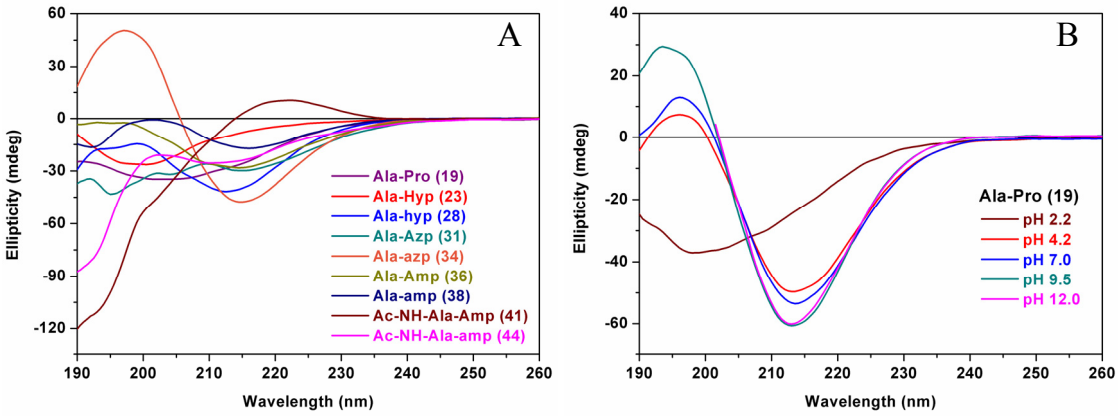


FIGURE 37: CD spectra of dipeptides; (**A**) Comparative plot of all dipeptides at their original pH values (≤ 2) and (**B**) Ala-Pro **19**, (5 mM, H_2O) at various pH values

– 190 nm, but suddenly changes to a more helical pattern (shift of negative band towards 215 nm) around pH 4.2 and maintains the same on proceeding to higher pH.

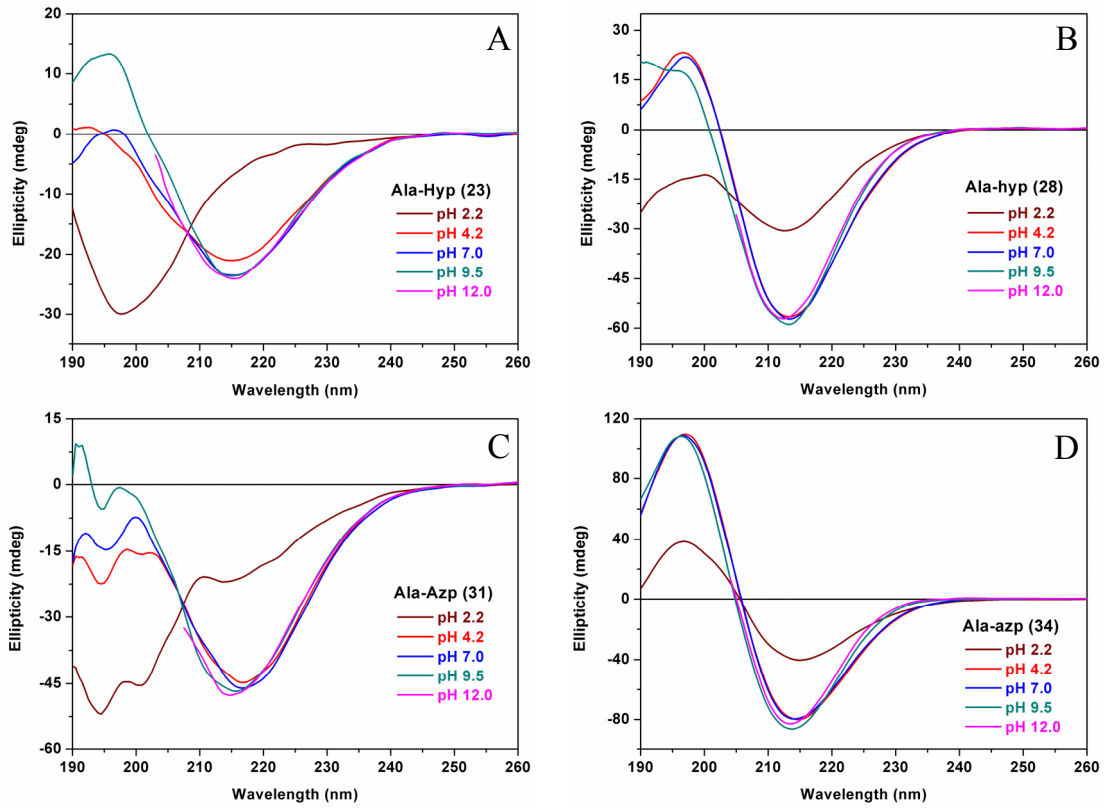


FIGURE 38: CD spectra of dipeptides (5 mM, H_2O) at various pH values; (**A**) Ala-Hyp **23**, (**B**) Ala-hyp **28**, (**C**) Ala-Azp **31** (**D**) and Ala-azp **34**

Similar pattern of CD spectra with variation of pH was observed in case of 4*R*-substituted dipeptides **23** and **31**, while 4*S*-substituted dipeptides had slightly dissimilar pattern, where instead of band shift, the intensity of +ve and –ve bands increased after reaching 4.2 pH value. The Figure **38(A-D)** illustrates the spectral plots of these peptides at pH 2.2, 4.2, 7.0, 9.5 and 12.0.

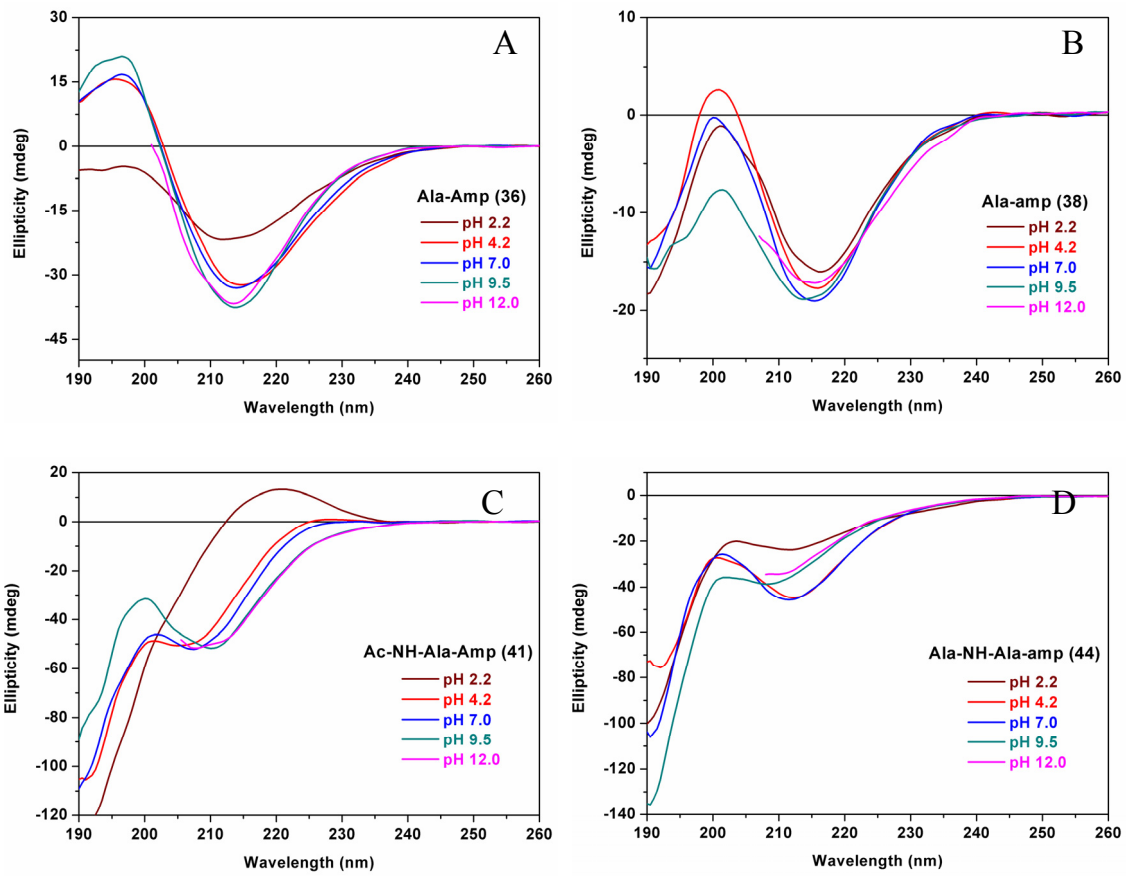


FIGURE 39: CD spectra of dipeptides (5 mM, H_2O) at various pH values; (**A**) Ala-Amp **36** (**B**) Ala-amp **38,** (**C**) Ac-NH-Ala-Amp **41** and (**D**) Ac-NH-Ala-amp **44**.

The same experiment was also repeated with 4-amino substituted dipeptides (Figure **39(A-D)**). In case of Ala-Amp **36**, the change in the CD spectra as a function of pH was very much similar to the previously examined dipeptides **31** and **34**. But Ala-amp **38** dipeptide with 4*S*-amino group showed very negligible change in its CD spectra. Moreover the end-capped Ac-NH-Ala-Amp **41** and Ac-NH-Ala-amp **44** dipeptides had a completely different pattern of spectra which had no effect of variation in pH either.

2.6 Discussion

The study of relation between structure and function of biomolecules is very important in understanding the biological functions of living systems. Proline is unique in the realm of amino acids in its ability to adopt completely distinct *cis* and *trans* conformations in peptides, which allows it to act as a backbone switch that is controlled by prolyl *cis-trans* isomerization. Further the conformation of the ring can be modulated as a function of 4-substituent, pH, temperature *etc.*

The nature and stereochemistry of -I substituents on the pyrrolidine ring influences the conformational behavior such as *E-Z* isomerization equilibrium and ring-pucker of proline residues.[11-16] Such behavior arises from the inductive effect exerted by the 4-substituent on the peptidyl-prolyl bond. The -I group in 4-position decreases the bond order of the peptide bond, *i.e.*, weakens the partial double bond character.[52] This weakening of the double bond character is believed to increase the *E-Z* isomerization rates and increase the preference to *Z* isomer.[11a] The preference to *Z* isomer also arises from a decreased C^{γ}-C^{δ} bond length, in effect pulling away the C^{δ} atom of the proline from the C^{α} of the preceding residue thus decreasing the steric repulsion. This structural manifestation of the inductive effect increases the stability of the *Z* (*trans*) isomer.

The molecular resolution of interconverting conformers' analysis may be done by X-ray crystallography and NMR techniques. The X-ray diffraction study provides only the structure of a stable conformer while NMR has the potential to provide information on a mixture of conformers, including the kinetics and thermodynamics of their interconversion.[53]

Numerous attempts to understand this complex blend of steric, electronic & conformational factors ruling *cis/trans* equilibrium are found in literature. Taylor *et al.*[54] studied the impact of pyrrolidine hydroxylation on the conformation of proline containing dipeptides by VT-NMR and reported that the preferred conformation is one in which there are a maximum number of *gauche* effects. The pyrrolidine conformation that has the strongest influence on the position of the *cis/trans* equilibrium for the preceding X-Pro peptide bond is C^{γ}-*exo*, for the electronegative OH-group in 4*R*-position. In another study by Koskinen *et al.*[55] a bulky *t*-Bu group substitution at C4/C^{γ}-position induced opposite puckering effects for the pyrrolidine ring (L-Proline) *i.e.* 4*S*-*exo* and 4*R*-*endo* in both solution and solid state. A theoretical

work by Alema *et al.*[56] suggested intramolecular H-bonds between the backbone carbonyl groups and the amino side group at C4/C$^{\gamma}$ position. This was further supported with their work related to influence of side chain protonation on the conformation and its effect on peptide bond isomerisation.

The present work is focussed to examine the pH–dependent effect of various substituents like *hydroxyl*, *azido* and *amino* groups at C4/C$^{\gamma}$-position of the pyrrolidine ring along with their relative steric (4*R*/*S*)-disposition on *cis-trans* isomerisation.

The present results on the ratio of *trans : cis* rotamers at varying pH estimated by NMR for unsubstituted dipeptide Ala-Pro **19**, is in agreement with Bailey's report on Ala-Ψ[CS-N]-Pro.[48] In the acidic pH range the percentage of *cis/trans* isomer was found to be almost same (10:90), while it becomes equimolar (50:50) at alkaline pH, which was around 35:65 for the Ala-Pro thiopeptide. This suggests a slightly faster rotation of peptide bond in comparison to the thiopeptides.

At the pK_a value of the acid/amino group, the dipeptide exists mostly in zwitterionic form (Figure **40**); there is little difference in the energy of the two isomers (*cis/trans*) due to formation of intramolecular salt bridge.

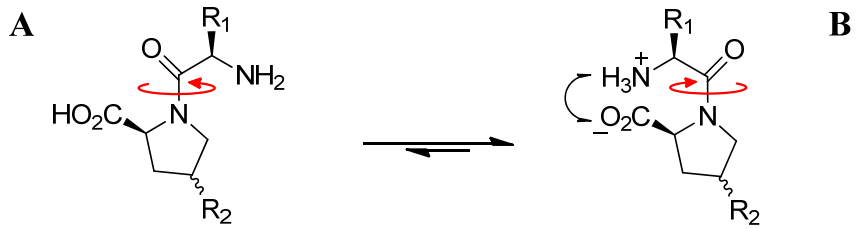


FIGURE 40: Proline dipeptides in the (**A**) *trans* and (**B**) *cis* conformations at non-ionised & ionised state respectively.

However in case of proline peptides the contribution from C4/C$^{\gamma}$-position substitution on the pyrrolidine ring may influence the conformation of the five-membered ring, which in turn modulate the backbone dihedral angles and hence the relative ratio of the *cis* and *trans* species.

Not surprisingly, significant effect of pH on chemical shift was seen for both C-terminal proline and N-terminal alanine residues. Since the pK_a of proline-carboxylic acid is 5.3 and alanine's free amine group is 11.2, in the middle of the pH range, the degrees of protonation of these groups vary with the change in pH (*i.e.*, the percentage of the deprotonated carboxylic acid increases with increasing pH). And shielding and upfield shifts for α-proton of proline in response to the population

of carboxylate ions are observed probably due to increase in average electron density around it. Same way happens for the N-terminal alanine residues α'-protons which are downfield when present in protonated state as NH_3^+ ions.

The initial identification of *cis* and *trans* isomers was guided by NOESY, the reason being NOESY spectra are capable for providing information on short proton–proton distances in the range from the van der Waals distance of 2.0 Å to approximately 4.4 Å. Wüthrich *et al.* provided a detailed description of proton–proton distances in standard tight turns of type I and II for L-Pro–L-X sequences using NOESY.[57] Thus from the NOESY data the proximity of alanine methyl protons towards the δ-protons of proline is observed at all instances (pH ~ 2) for major isomers of dipeptides. In another set the same alanine methyl protons (although low in population) show closeness with the α-protons of proline, so by logic the major isomer is assigned to be the *trans* isomer and the minor to be *cis*.

Among the substituted dipeptides the hydroxyl group at C^γ of proline has high importance as it is seen in the repeating triads of collagen sequence. Ala-Hyp **23**, with the hydroxyl group at 4*R*-position shows a sudden change in the percentage of *cis/trans* equilibrium within pH 7 to 9.5, where the *trans* isomer reduces from 72 to 58%. In comparission, the 4*S*-diastereomer shows a more gradual change increases the % of *cis* isomer in the entire pH range.

It is to be noted that all the 4*R*-stereoisomers present a similar *cis-trans* equilibrium profile to that of unsubstituted dipeptide Ala-Pro **19**, whereas the 4*S* –dipeptides differ from 4*R*-isomers significantly.

More interestingly, the non-ionisable azido group in its 4*S*-conformation prefers *cis*-peptide bond at all pH values examined except at pH 2.2. This behaviour is probably due to the inherent *gauche* effect involved. Wennemers *et al.* have recently shown that for Ala-(4*S*)azp derivatives, the azido gauche effect induces a C4/C^γ-*endo* conformation of the pyrrolidine ring which prevents the stabilizing $n\rightarrow\pi^*$ interaction of the *trans* conformer.[58] Consequently, the equilibrium shifts towards *cis* peptide and higher *cis* conformer ratio is observed for Ala-azp **34** (61%, 4*S*-), compared to Ala-Azp **31** (44%, 4*R*-) derivatives at pH 7 in D_2O.

Further the 4-amino substituted dipeptides Ala-Amp **36** and Ala-amp **38**, have almost similar *cis/trans* isomer preference except at the strong acidic pH where the Ala-Amp (4*R*-) predominates on the *trans* isomer with 80% population while the Ala-amp (4*S*-) has only 68%.

4(*R/S*)-substituent thus plays a major role in determining the pucker of the pyrrolidine ring of proline and hence the conformational stability of peptides and when this substituent is an ionisable amino group, it has the added property of change in stereoelectronic effects as a function of pH. The effect of such a combination on the secondary structure of aminoproline based polypeptides is reported very recently in solvent dependent switch of β-structure to PPII helix in case of 4*S*-aminoproline based polypeptides.[59]

The pH dependent ionisation of amino group also acts as a switch to control the *cis-trans* ratios in the dipeptide. But the free-amino group at the N-terminal of alanine residue seems to play a major role in dynamic *cis-trans* equilibrium. The dipeptides **41** and **44** with capped N-terminal showed meagre or no change in *cis/trans* ratios with respect to the change in pH. Similar effect of N-terminus on the stability of PPII structure is also seen in a recent paper.[60]

Crystal structures of certain dipeptide intermediate like Boc-NH-Ala-Pro-OH, **18** seemed to have a slight *endo*-puckering at C^{γ} position. Other dipeptides (Boc-Ala-hyp-OMe **26**, Boc-Ala-azp-OMe **32** and Boc-Ala-azp-OH **33**) with 4*S*-stereochemistry showed C^{γ}-*endo* pucker to various degrees, while Boc-NH-Ala-Hyp-OH, **22** had C^{γ}-*exo* conformation. This envisages that the unsubstituted proline ring which prefers both *endo* and *exo*-puckering to equal extent in solution can be freezed to one of the states when studied in crystal. The conformation of proline ring in substituted dipeptides had similar results in accordance to earlier literature reports.

Although the dipeptides do not show any evidence for ordered structures in the CD study, the change of CD pattern with pH indicates some conformational differences in ionisation state of amino function.

2.7 Conclusions

It is often difficult to reduce different and complex processes to a simple concept. *Cis-trans* isomerization causes reversible modification of the molecule geometry which becomes the basis of changes in its physical, chemical, and biological properties. This isomerization is also a simple way, originally found by nature, of increasing molecular diversity.

The present work elucidated the importance of pH in probing the dipeptide *cis/trans* equilibrium using Ala-Pro based substituted model dipeptides.

This process has high potential to be exploited in drug design and understanding structure-activity-relationships which can lead to development of new ligands and inhibitors.

2.8 Experimental

2.8.1 Synthesis of dipeptides

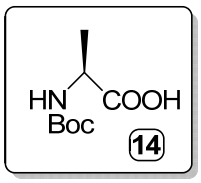
General Remarks: All reagents and chemicals were of laboratory or analytical grade obtained from commercial sources and were used without further purification except in cases mentioned. Thin layer chromatography (TLC) was carried out on pre-coated silica gel 60 F_{254} plates (E. Merck). TLCs were visualized under UV light, iodine and/or ninhydrin spray, phosphomolybdic acid charring solution (only for N-acetylated compounds) followed by heating up to 110 ^{o}C with heat gun. Silica gel 60-120 and 100-200 mesh (Merck) was used for routine column chromatography with ethyl acetate/petroleum ether or dichloromethane/methanol mixture as elution solvent depending upon the compound polarity and chemical nature, whereas for flash chromatography 230-400 mesh silica gel was used. All solvents were distilled under an inert atmosphere with appropriate desiccant. Reactions in aqueous medium and workup processes were done using double distilled water. Unless otherwise noted, all reactions were carried out at room temperature. The dipeptide syntheses mentioned in this section are done by solution phase using standard EDCI /HOBt coupling method.

^{1}H NMR spectra were routinely recorded at 200 MHz on a Bruker AC-200 instrument controlled by an Aspect 2000 computer. ^{13}C NMR and ^{13}C-DEPT spectra (at 50 MHz) were recorded on the same instrument. The spectra were analyzed using ACD NMR processor (Version 12) software from ACD labs. For final dipeptides, NMR spectra were also recorded on 400 MHz JEOL and 500 MHz Bruker spectrometers. All chemical shifts are referenced with respect to TMS as internal standard and are expressed in δ-scale (ppm). Mass spectra were obtained by ESI-MS technique on AP-QSTAR spectrometer. Melting points of the samples were determined in open capillary tubes using Büchi Melting Point M-560 apparatus and are uncorrected. IR spectra were recorded on an Infrared Fourier Transform Spectrophotometer using chloroform, nujol or neat sample. The optical rotation values were measured on JASCO P2000 polarimeter. $CDCl_3$, MeOH-d_4 and D_2O

(98.9% atom) ware obtained from Aldrich, while solutions of NaOH and TFA were prepared in D_2O for adjustment of pH in NMR samples.

Procedures and Spectral Data:

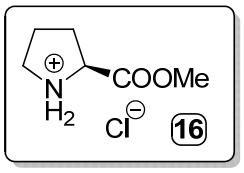
(*S*)-N^1-(t-butoxycarbonyl)-alanine (14)



A solution of amino acid **Alanine** (1 g, 11.2 mmol) in a solution of aq. 2N NaOH (10 ml) and dioxan (10 ml) was cooled to 0 ^oC and Boc anhydride (2.8 ml, 12.3 mmol) was added drop wise using addition funnel. The reaction mixture was stirred at 0 ^oC for 1 h. Dioxane was removed under reduced pressure; aqueous layer was covered with EtOAc (25 ml). The reaction mixture was neutralized by adding, solid $KHSO_4$ portion wise with vigorous stirring until pH 2. The organic layer was separated, aqueous layer was extracted with ethyl acetate (3 x 25 ml), the combined organic layer was washed with water followed by saturated brine solution and dried over anhydrous Na_2SO_4. Removal of ethyl acetate under reduced pressure, afforded a white solid (2.1 g, >99% yield) which was recrystallized with EtOAc/Petroleum-ether.

Mol. Formula	:	$C_8H_{15}NO_4$
mp	:	60.3 – 62.3 ^oC
Mol. Weight	:	189.20
ESI-MS *m/z*	:	212.21 $[M+Na]^{+\bullet}$
$[\alpha]_D^{25}$	:	+2.96 (C=1.56, $CHCl_3$)
^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.42 (s, 3H), 1.45 (s, 9H), 4.26-4.40 (m, 1H), 5.02-5.05 (d, J=5.6 Hz, 1H)
^{13}C NMR (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) 18.4, 28.3, 49.1, 80.3, 155.5, 177.7
^{13}C-DEPT	:	**δ_C** (ppm) 18.4, 28.3, 49.1

L-Proline methyl ester hydrochloride salt (16)



A suspension of L-Proline **15** (1 g, 8.6 mmol) in anhydrous methanol (40 ml) was stirred at 0 ^oC. Thionyl chloride (0.8 ml, 10.7 mmol) was added dropwise to the reaction mixture. The stirring was continued at 0 ^oC during the addition of thionyl chloride. After completing the addition it was allowed to come to room temperature. The reaction mixture was refluxed for another 6 hrs. Methanol was removed under vacuum with KOH trap till a white solid was obtained. The precipitate was washed

with EtOAc, followed by diethyl ether. The residue dried under vacuum over phosphorus pentoxide yielded methyl ester hydrochloride **16** as a white solid 1.37 g (97% yield). This solid was used for the next reaction without further purification.

General procedure for synthesis of N-Boc-dipeptide methylesters 17, 21, 26, 29 and 32

N^1-(t-butoxycarbonyl)-alanine **14** (1 g, 5.3 mmol), HOBt (1.1 g, 8 mmol) and EDCI (1.5 g, 8 mmol) / DIC (1.3 ml, 8 mmol) were taken in dry DCM (80 ml) under inert atmosphere (argon), and the reaction mixture was stirred for 30 min at -10 ^oC (ice-salt). Free amine of compounds (**16**, **2**, **25**, **4** and **9**; 5.1 mmol) dissolved in dry DCM (10 ml) were added separately to the above mixture. The reaction mixture was stirred overnight at RT.

After completion of reaction as monitored by TLC, water (80 ml) was added to the reaction mixture and it was extracted with DCM (3 x 50 ml). The organic layer was washed successively with 10% $NaHCO_3$ (2 x 50 ml), 10% citric acid (50 ml) and 10% $NaHCO_3$ (150 ml) and finally with water (3 x 100 ml). The organic layer was dried over anhydrous Na_2SO_4, filtered and concentrated on rotary-evaporator to obtain crude dipeptide. Column chromatography with ethyl acetate/pet-ether yielded the respective dipeptides (**17**, **21**, **26**, **29** and **32**) in good to moderate yields (85 to 62%).

DIC was used exclusively for synthesis of dipeptide **21** and **26**. Free amine of compounds (**16**, **2**, **25**, **4** and **9**) was generated as follows. TFA in dry DCM (50%, 3 ml) was added to the above compounds under ice-cold condition and stirred for 2 hr. Subsequent neutralization of TFA salt with DIPEA in inert atmosphere resulted in the free amine. For compounds **16** & **20**, which were already hydrochlorides, only DIPEA neutralization was carried out.

General procedure for synthesis of N-acetylated dipeptides 39 and 42

The azide (**29**/ **32**, 1 g, 3 mmol) compound (**29** and **32**) was stirred for 2 hr in mixture of DCM and TFA (1:1, 8 ml). The solvent was removed under reduced pressure, and the resulting slurry was dissolved in diethyl ether (10 ml). Removal of ether yielded white solid of the corresponding trifluoroacetate salts. The salt obtained was stirred in a solution of acetic anhydride, Et_3N and anhydrous DCM (1:1:1, 20 ml) for 8 hrs after which the solvent was removed under reduced pressure. The resulting

crude product was purified by column chromatography (EtOAc/pet-ether, 7:6). Yield (~70%).

General method of hydrolysis used for synthesis of dipeptide-acids 18, 22, 27, 30, 33, 40 and 43

The methyl esters (**17**, **21**, **26**, **29** and **32**; 3.5 mmol) were generally hydrolyzed using aq. NaOH (1N) in water-MeOH (1:1, 20 ml) mixture. This reaction mixture was stirred in RT for 2 hrs and methanol was removed under reduced pressure. The remaining aq. solution was washed with diethyl ether (3 x 15 ml) and neutralized with preactivated Dowex H^+ (8X, 100-200 mesh) resin till the pH of the solution turned 7.0. The resin was removed by filtration and the filtrate was concentrated to obtain the Boc-protected acid (**18**, **22**, **27**, **30** and **33**) in excellent yields (mostly > 95%). Only esters of **39** and **42** were hydrolysed by aq. LiOH (1N) instead of NaOH to obtain **40** and **43** respectively in good yields (~90%).

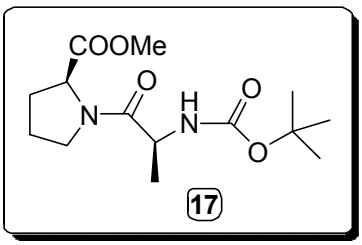
General procedure for synthesis of dipeptide-amines 35, 37, 41 and 44

Individual compounds (**30**, **33**, **40** and **43**; 5 mmol) taken in methanol, was treated with 10% Pd-C (150 mg) and the mixture was subjected to hydrogenation in a Parr-Hydrogenation shaker at a pressure of 40 psi for 3 hr. Pd-C suspension was removed by filtration through Celite and the filtrate was evaporated under reduced pressure to give the reduced products (**35**, **37**, **41** and **44**). The compounds thus obtained were used for next reaction without further purification.

General procedure for synthesis of dipeptide-amine TFA salts 19, 23, 28, 31, 34, 36 and 38

Trifluroacetic acid in dry DCM (50%, 3 ml) was added individually to cooled (0 ^oC) solution of compounds (**18**, **22**, **27**, **30**, **33**, **35** and **37**; 3 mmol) in dry DCM (1 ml) in inert atmosphere (argon) drop wise with the syringe and the reaction mixture stirred at same ice-cold condition for 2.5 hr (reaction was monitored by TLC, 50% n-butanol + 20% H_2O + 15% pyridine + 15% acetic acid). After the completion of reaction the solvent was evaporated on rotary evaporator under vacuum. The final desired dipeptides were obtained as TFA salts in excellent yields. (≥ 99%)

(*S*)-N-(t-butoxycarbonyl)-alanine-proline dipeptide methylester (17)



Mol. Formula : $C_{14}H_{24}N_2O_5$
Mol. Weight : 300.35
ESI-MS *m/z* : 301.14 $[M+1]^{+\bullet}$, 323.09 $[M+Na]^{+\bullet}$ 339.09 $[M+K]^{+\bullet}$
$[\alpha]^{25}_D$: -79.23 (C=1.22, $CHCl_3$)

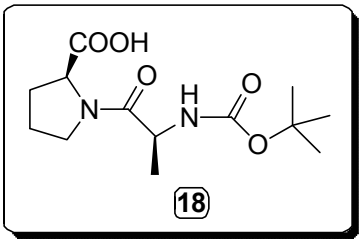
IR : ν (cm^{-1}) 3423, 2980, 2881, 2453, 1749, 1707, 1650, 1450, 1391, 1367, 1170, 1095, 1047, 924, 865, 755, 665, 493
thin film on $CHCl_3$

^{1}H NMR : δ_H (ppm) 1.23 & 1.29 (2d, J=6.8 Hz, 3H), 1.36 (s, 9H), 1.89-2.06 (m, 3H), 2.12-2.22 (m, 1H), 3.52-3.68 (m, 2H), 3.66 & 3.71 (2s, 3H), 4.38-4.43 (dd, J=7.5, 7.0 Hz, 1H), 4.46-4.49 (dd, J=8.4, 4.4 Hz, 1H), 5.33 (d, J=8 Hz, 1H)
(400 MHz, $CDCl_3$)

^{13}C NMR : δ_C (ppm) 18.1, 24.8, 28.2, 28.8, 46.6, 47.6, 52.1, 58.5, 79.4, 155.1, 171.6, 172.3
(100 MHz, $CDCl_3$)

^{13}C-DEPT : δ_C (ppm) *Positive Peaks:* 18.1, 28.2, 47.6, 52.1, 58.5
(100 MHz, $CDCl_3$) *Negative Peaks:* 24.8, 28.8, 46.6

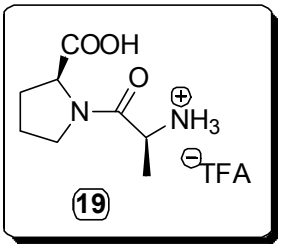
(*S*)-N-(t-butoxycarbonyl)-alanine-proline dipeptide (18)



Mol. Formula : $C_{13}H_{22}N_2O_5$
mp : 107 – 111.1 °C
Mol. Weight : 286.32
ESI-MS *m/z* : 287.18 $[M+1]^{+\bullet}$, 309.10 $[M+Na]^{+\bullet}$, 325.08 $[M+K]^{+\bullet}$

^{1}H NMR : δ_H (ppm) 1.29 (d, J=7.3 Hz, 3H), 1.43 (s, 9H), 1.99-2.09 (m, 3H), 2.21-2.31 (m, 1H), 3.62-3.67 (m, 1H), 3.75-3.81 (m, 1H), 4.35-4.40 (q, J=7.2 Hz, 1H), 4.44-4.47 (dd, J=8.9, 3.9 Hz, 1H)
(400 MHz, MeOH-d_4)

(*S*)-alanine-(2*S*)-proline dipeptide TFA salt (19)



Mol. Formula : $C_8H_{15}N_2O_3^+$
Mol. Weight : 187.22
ESI-MS *m/z* : 187.19 $[M]^{+\bullet}$
$[\alpha]^{25}_D$: -67.26 (C=0.36, H_2O, pH=1.79)

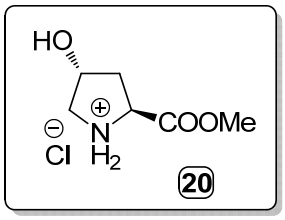
^{1}H NMR : δ_H (ppm) 1.11 & 1.18 (2d, J=6.8 Hz, 3H), 1.74-1.79 & 2.01-2.08 (2m, 1H), 2.13-2.20 & 2.21-2.30 (2m, 1H), 1.81-1.86 & 1.88-1.97 (2m, 2H), 3.39-3.41 & 3.47-3.65 (2m, 2H), 3.35-3.37 & 3.69-3.73 (m & q, J=6.8 Hz; 1H), 4.17-4.20 & 4.27-4.29 (dd, J=8.6, 5.1 Hz
(500 MHz, D_2O)
pD=12

and dd, J=8.7, 3.2 Hz; 1H)

^{13}C NMR (100 MHz, D_2O) pD=12 : δ_C (ppm) 18.6 (mi), 18.8 (ma), 22.3 (mi), 29.3 (mi), 31.4 (ma), 47.1 (mi), 47.2, 47.4 (mi), 47.7 (ma), 61.9 (mi), 62.0 (ma), 175.9 & 177.0 (equal), 179.6 (ma), 179.7 (mi)

^{13}C-DEPT (125 MHz, D_2O) pD=12 : δ_C (ppm) *Positive Peaks:* 18.6 (mi), 18.8 (ma), 47.4 (mi), 47.7 (ma), 61.9 (mi), 62.0 (ma); *Negative Peaks:* 22.3 (mi), 29.3 (mi), 31.4 (ma), 47.1 (mi), 47.2 (ma)

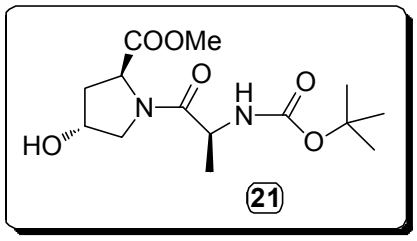
(2*S*,4*R*)-4-Hydroxyproline methyl ester hydrochloride salt (20)



20

(2*S*,4*R*)-4-hydroxyproline **1** (1 g, 7.6 mmol) was esterfied in similar process as for compound **16** in anhydrous methanol (40 ml) and Thionyl chloride (0.7 ml, 9.5 mmol) to give 4-hydroxyproline methyl ester hydrochloride **20** as white solid.

(*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-Hydroxyproline dipeptide methylester (21)



21

Mol. Formula : $C_{14}H_{24}N_2O_6$

mp : 112.4 -114.9 °C

Mol. Weight : 316.35

ESI-MS *m/z* : 317.16 $[M+1]^{+\bullet}$,339.07 $[M+Na]^{+\bullet}$, 355.05 $[M+K]^{+\bullet}$

$[\alpha]_D^{25}$: -78.94 (C=0.33, $CHCl_3$)

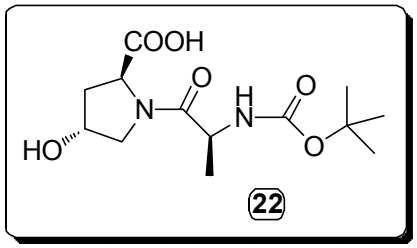
IR thin film on $CHCl_3$: ν (cm^{-1}) 3412, 3018, 1640, 1218, 772, 666, 510

^{1}H NMR (400 MHz, $CDCl_3$) : δ_H (ppm) 1.27 & 1.32 (2d, J=6.8 Hz, 3H), 1.39 (s, 9H), 1.96-2.02 (ddd, J=13.3, 8.7, 4.6, 1H), 2.30-2.35 (dd, J=13.4, 8.2 Hz, 1H), 3.64-3.68 (dd, J=10.9, 3.9 Hz, 1H), 3.71 & 3.76 (2s, 3H), 3.83 (d, J=10.8 Hz, 1H), 4.41 (quin, J=7.2 Hz, 1H), 4.53 (bs, 1H), 4.63 (t, J=8.3 Hz, 1H), 5.38 (d, J=7.5 Hz, 1H)

^{13}C NMR (50 MHz, $CDCl_3$) : δ_C (ppm) 17.7, 28.2, 37.3, 47.7, 52.2, 54.9, 57.6, 70.1, 79.9, 155.5, 172.0, 172.5

^{13}C-DEPT (50 MHz, $CDCl_3$) : δ_C (ppm) *Positive Peaks:* 17.7, 28.2, 47.7, 52.2, 57.6, 70.1; *Negative Peaks:* 37.3, 54.9

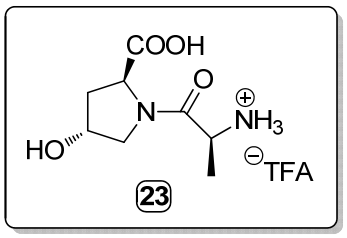
(*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-Hydroxyproline dipeptide (22)



Mol. Formula : $C_{13}H_{22}N_2O_6$
Mol. Weight : 302.32
ESI-MS *m/z* : 303.27 $[M+1]^{+\bullet}$, 325.27 $[M+Na]^{+\bullet}$, 341.23 $[M+K]^{+\bullet}$

^{1}H NMR (400 MHz, MeOH-d_4) : δ_H (ppm) 1.29 (d, J=6.9 Hz, 3H), 1.43 (s, 9H), 2.04-2.10 (ddd, J=13.2, 8.1, 4.8 Hz, 1H), 2.24-2.30 (m, 1H), 3.59 & 3.74 (2d, J=4.1 Hz & 3.7 Hz; 2H), 4.39 (q, J=7 Hz, 1H), 4.48-4.54 (m, 2H)

(*S*)-alanine-(2*S*,4*R*)-Hydroxyproline dipeptide TFA salt (23)



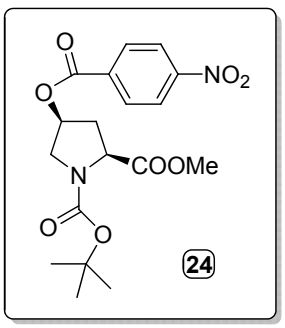
Mol. Formula : $C_8H_{15}N_2O_4^+$
Mol. Weight : 203.22
ESI-MS *m/z* : 203.33 $[M]^{+\bullet}$
$[\alpha]_D^{25}$: -42.78 (C=0.36, H_2O, pH=1.74)

^{1}H NMR (500 MHz, D_2O) pD=12 : δ_H (ppm) 1.14 & 1.18 (2d, J=6.8, 1H), 1.93-1.98 & 2.10-2.16 (ddd, J=13.7, 9.1, 4.6 Hz & m; 1H), 2.23-2.28 & 2.35-2.40 (dd, J=13.6, 8 Hz & m; 1H), 3.41 & 3.70-3.74 (q, J=6.8 Hz & m, 1H), 3.55 & 3.63-3.69 (br s & m, 2H), 4.28 & 4.43 (t, J=8.5 Hz & br s 1H), 4.40 & 4.50 (t, J=8.5 Hz & br s, 1H)

^{13}C NMR (100 MHz, D_2O) pD=12 : δ_C (ppm) 18.75 (mi), 18.78 (ma), 37.1(ma), 39.0(mi), 47.4(mi), 47.5 (ma), 54.3 (mi), 54.7 (ma), 60.50 (mi), 60.53(ma), 68.1(mi), 70.0 (ma), 176.4 (ma), 177.9 (mi), 179.22(mi), 179.27 (ma)

^{13}C-DEPT (125 MHz, D_2O) pD=12 : δ_C (ppm) *Positive Peaks:* 18.75 (mi), 18.78 (ma), 47.4(mi), 47.5 (ma), 60.50 (mi), 60.53(ma), 68.1(mi), 70.0 (ma); *Negative Peaks:* 37.1(ma), 39.0(mi), 54.3 (mi), 54.7 (ma)

(2*S*,4*S*)-N^1-(t-butoxycarbonyl)-4-(p-nitrobenzoyloxy) proline methyl ester (24)

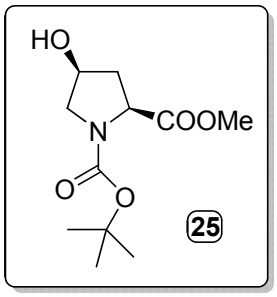


The compound **2,** (ref: *Page 91*) (1 g, 4 mmol) along with PPh_3, (2.3 g, 8.8 mmol) and *p*-nitrobenzoic acid (1.3 g, 8 mmol) were dissolved in dry THF (100 ml) cooled to 0 ^oC on ice bath, under Argon. After stirring for 30 min at ice-cold temperature, DIAD (1.7 ml, 8.8 mmol) was added slowly with syringe. The reaction mixture brought to RT and stirred for

another 24 hrs. THF was removed under vacuum. The resulting orange colored thick oil was dissolved in 150 ml diethyl ether, by triturating with spatula; the resulting solution was kept overnight at room temperature. The white crystals were filtered and the residue was washed with petroleum ether followed by diethyl ether. The filtrate was concentrated under reduced pressure and purified by silica gel column chromatography, which offered the compound **24** as a white crystalline substance. Yield 1.4 g (89%).

Mol. Formula	:	$C_{18}H_{22}N_2O_8$
MP	:	94.4 – 96.4 ^oC
Mol. Weight	:	394.38
ESI-MS *m/z*	:	393.07 $[M-1]^{+\bullet}$,417.00 $[M+Na]^{+\bullet}$
^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.44 & 1.48 (2s, 9H), 2.41-2.48 (m, 1H), 2.53-2.68 (m, 1H), 3.69 & 3.70 (2s, 3H), 3.74-3.84 (m, 2H), 4.46- 4.64 (m, 1H), 5.55-5.58 (br m, 1H), 8.13-8.17 (m, 2H), 8.26-8.31 (m, 2H)
^{13}C NMR (100 MHz, $CDCl_3$)	:	**δ_C** (ppm) 28.2(ma), 28.3(mi), 35.6(mi), 36.4(ma), 52.1(ma), 52.5(mi), 57.4(mi), 57.7(ma), 73.4, 74.5, 80.6, 123.5, 130.8, 135.0, 150.7, 153.6(ma), 154.0(mi), 163.9(ma), 164.0(mi), 172.3(ma), 172.1(mi)
^{13}C-DEPT (100 MHz, $CDCl_3$)	:	**δ_C** (ppm) *Positive Peaks:* 28.2(ma), 28.3(mi), 57.4(mi), 57.7(ma), 73.4, 74.5; *Negative Peaks:* 35.6(mi), 36.4(ma), 52.1(ma), 52.5(mi)

(2*S*,4*S*)N^1-(t-butyloxycabonyl)-4-hydroxyproline methylester (25)

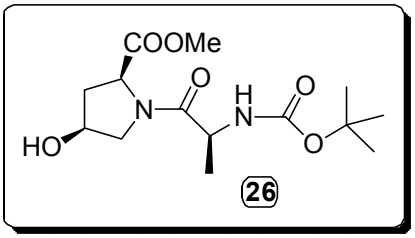


A solution of **24** (1 g, 2.5 mmol) and sodium azide (0.65 g, 10 mmol) in dry MeOH (60 ml) was stirred overnight at 40 ^oC under nitrogen. The solvent was removed on a rotary evaporator, and the compound was purified by flash chromatography (ethyl acetate:hexane, 1:3 to 1:1) to give 0.96 g (98%) of a thick yellowish oil.

Mol. Formula	:	$C_{11}H_{19}NO_5$
MP	:	75.3 – 79.1 ^oC
Mol. Weight	:	245.27
ESI-MS *m/z*	:	246.18 $[M+1]^{+\bullet}$, 268.18 $[M+Na]^{+\bullet}$, 284.20 $[M+K]^{+\bullet}$
^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.40 & 1.44 (2s, 9H), 2.30-2.11 (m, 1H), 2.22-2.41 (m, 1H), 3.36-3.56 (m, 2H), 3.57-3.70 (m, 1H), 3.76 & 3.77 (2s, 3H), 4.24-4.38 (ddd, J=16.3, 9.6, 1.6 Hz, 2H)
^{13}C NMR (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) 28.1 (ma), 28.2 (mi), 37.6 (mi), 38.4 (ma), 52.3 (ma), 52.6 (mi), 55.0 (ma), 55.8 (mi), 57.5 (mi), 57.7 (ma), 69.9 (ma), 71.0 (mi), 80.3, 153.6 (ma), 154.3 (mi), 175.1 (ma), 175.4 (mi)

^{13}C-DEPT (50 MHz, $CDCl_3$) : **δ_C** (ppm) *Positive Peaks:* 28.2 (ma), 28.3 (mi), 52.4 (ma), 52.7 (mi), 57.6 (mi), 57.8 (ma), 70.0 (ma), 71.1 (mi); *Negative Peaks:* 37.7 (mi), 38.5 (ma), 55.1 (ma), 55.7 (mi)

(*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-Hydroxyproline dipeptide methylester (26)



26

Mol. Formula : $C_{14}H_{24}N_2O_6$
mp : 112.5 – 117.1 °C
Mol. Weight : 316.35
ESI-MS *m/z* : 317.45 $[M+1]^{+\bullet}$, 339.53 $[M+Na]^{+\bullet}$, 355.64 $[M+K]^{+\bullet}$
$[\alpha]_D^{25}$: -38.45 (C=0.49, $CHCl_3$)

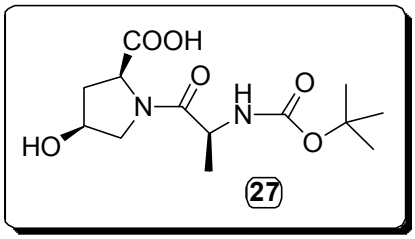
IR thin film on $CHCl_3$: **ν** (cm^{-1}) 3615, 3432, 3019, 2980, 2931, 2400, 1705, 1651, 1502, 1454, 1369, 1216, 1170, 1090, 1049, 929, 861, 770, 668, 509

^{1}H NMR (200 MHz, $CDCl_3$) : **δ_H** (ppm) 1.36 (d, J=6.9 Hz, 3H), 1.42()s, 9H), 2.05-2.14 (m,1H), 2.26-2.40 (m, 1H), 3.50 (d, J=9.7 Hz, 1H), 3.67-3.73 (m, 2H), 3.79 (s, 3H), 4.33-4.50 (m, 2H), 4.54-4.60 (dd, J=9.7, 1.8Hz, 1H), 5.24 (d, J=8.1 Hz, 1H)

^{13}C NMR (50 MHz, $CDCl_3$) : **δ_C** (ppm) 18.1, 28.2, 36.6, 47.6, 52.8, 55.6, 57.5, 71.2, 79.7, 155.2, 172.4, 174.5

^{13}C-DEPT (50 MHz, $CDCl_3$) : **δ_C** (ppm) *Positive Peaks:* 18.1, 28.2, 47.6, 52.8, 57.5, 71.2; *Negative Peaks:* 36.6, 55.6

(*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-Hydroxyproline dipeptide (27)

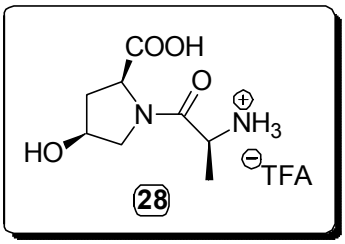


27

Mol. Formula : $C_{13}H_{22}N_2O_6$
Mol. Weight : 302.32
ESI-MS *m/z* : 303.43 $[M+1]^{+\bullet}$, 325.39 $[M+Na]^{+\bullet}$ 341.40 $[M+K]^{+\bullet}$

^{1}H NMR (400 MHz, MeOH-d_4) : **δ_H** (ppm) 1.28 & 1.31 (d, J=7.3 Hz, 3H), 1.43 (s, 9H), 1.99-2.04 (dt, J=13.3, 2.7 Hz, 1H), 2.34-2.41 (ddd, J=13.6, 9.0, 4.8 Hz, 1H), 3.56-3.64 (m, 1H), 3.89-3.93 (dd, J=10.8, 4.8 Hz, 1H), 4.33-4.35 (m, 1H), 4.38-4.42 (m, 2H)

(*S*)-alanine-(2*S*,4*S*)-Hydroxyproline dipeptide TFA salt (28)



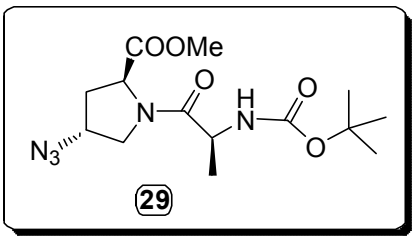
Mol. Formula : $C_8H_{15}N_2O_4^+$
Mol. Weight : 203.22
ESI-MS *m/z* : 203.33 $[M]^{+\bullet}$
$[\alpha]_D^{25}$: -40.96 (C=0.36, H_2O, pH=1.72)

^{1}H NMR (500 MHz, D_2O, pD12) : δ_H (ppm) 1.10 & 1.19 (2d, J=6.8 Hz, 3H), 1.83-1.88 & 2.13-2.22 (dt, J=13.3, 4.8 Hz & m; 1H) and 2.38-2.49 (ddd, J=13.7, 1.3, 17.8, 1.6 Hz; 1H), 3.42-3.45 & 3.66-3.69 (m, 1H), 3.46-3.50 (m, 1H) and 3.60-3.65 & 3.82-3.86 (m & dd, J=11.1, 5.3 Hz; 1H), 4.19-4.22 & 4.36- 4.40 (dd, J=9.1, 4.9 Hz & m; 1H), 4.33-4.38 (m, 1H)

^{13}C NMR (100 MHz, D_2O) pD=12 : δ_C (ppm) 18.6(ma), 18.8(mi), 36.4(mi), 38.7(ma), 47.4(mi), 47.5(ma), 54.1(mi), 54.8(ma), 60.61(ma), 60.64(mi), 68.1 (ma), 69.9(mi), 176.4 & 176.6 (equal), 179.1 & 179.4 (equal)

^{13}C-DEPT (125 MHz, D_2O) pD=12 : δ_C (ppm) *Positive Peaks:* 18.6(ma), 18.8(mi), 47.4(mi), 47.5(ma), 60.61(ma), 60.64(mi), 68.1 (ma), 69.9(mi); *Negative Peaks:* 36.4(mi), 38.7(ma), 54.1(mi), 54.8(ma)

(*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-4-azidoproline dipeptide methyl ester (29)



Mol. Formula : $C_{14}H_{23}N_5O_5$
Mol. Weight : 341.36
ESI-MS *m/z* : 364.04 $[M+Na]^{+\bullet}$
$[\alpha]_D^{25}$: -37.02 (C=0.97, $CHCl_3$)

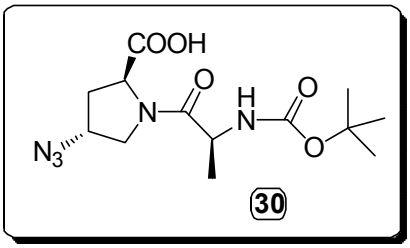
IR thin film on $CHCl_3$: ν (cm^{-1}) 3430, 3016, 2981, 2933, 2459, 2107, 1750, 1707, 1654, 1500, 1437, 1392, 1368, 1322, 1216, 1174, 1092, 1054, 1028, 949, 927, 858, 757, 667, 623, 488

^{1}H NMR (500 MHz, $CDCl_3$) : δ_H (ppm) 1.33 & 1.37 (2d, J=6.8, 3H), 1.42 (s, 9H), 2.15-2.21 (m, 1H), 2.31-2.38 (m, 1H), 3.74 (s, 3H), 3.76-3.81 (m, 2H), 4.28-4.34 (m, 1H), 4.43 (quin, J=7.2 Hz, 1H) 4.59 (t, J=7.5 Hz, 1H), 5.29 (d, J=8 Hz, 1H)

^{13}C NMR (50 MHz, $CDCl_3$) : δ_C (ppm) 18.1, 28.2, 34.0, 47.6, 51.2, 52.5, 57.1, 59.3, 79.7, 155.2, 170.9, 171.9

^{13}C-DEPT (50 MHz, $CDCl_3$) : δ_C (ppm) *Positive Peaks:* 18.1, 28.2, 47.6, 52.5, 57.1, 59.3; *Negative Peaks:* 34.0, 51.2

(*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-4-azidoproline dipeptide (30)

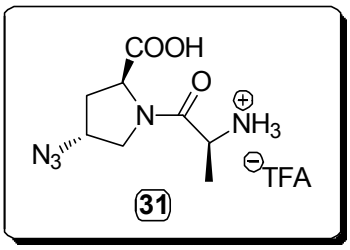


Mol. Formula : $C_{13}H_{21}N_5O_5$
Mol. Weight : 327.34
ESI-MS *m/z* : 328.21 $[M+1]^{+\bullet}$, 350.08 $[M+Na]^{+\bullet}$

IR : ν (cm^{-1}) 3431, 3019, 2981, 2934, 2400, 2107, 1704, 1652, 1502, 1455, 1392, 1369, 1323, 1216, 1165, 1053, 1028, 928, 852, 762, 668, 624, 508
thin film on $CHCl_3$

^{1}H NMR : δ_H (ppm) 1.28 (d, J=6.9 Hz, 3H), 1.43 (s, 9H), 2.18-2.24 (ddd, J=13.3, 7.3, 5.5 Hz, 1H), 2.34-2.40 (m, 1H), 3.68-3.90 (m, 2H), 4.27-4.40 (m, 2H), 4.46 (t, J=7.8 Hz, 1H)
(200 MHz, $CDCl_3$)

(*S*)-alanine-(2*S*,4*R*)-4-azidoproline dipeptide TFA salt (31)



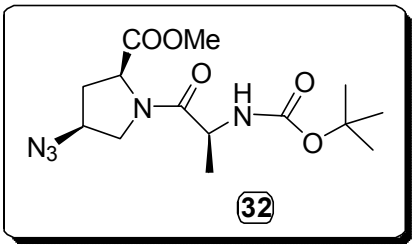
Mol. Formula : $C_8H_{14}N_5O_3^+$
Mol. Weight : 228.23
ESI-MS *m/z* : 228.30 $[M+1]^{+\bullet}$
$[\alpha]_D^{25}$: -28.66 (C=0.022, H_2O, pH=3.06)

^{1}H NMR : δ_H (ppm) 1.12 & 1.17 (2d, J=6.8 Hz, 3H), 2.02-2.08 and 2.19- 2.25 (2 ddd, J= 13.8, 8.6, 5.2 Hz and 13.5, 7.5, 5.6 Hz; 1H), 2.36-2.41 and 2.47-2.51 (ddd, J= 13.6, 8.1, 2.6 Hz & m; 1H), 3.39 & 3.57-3.60 (q, J=6.8 Hz & dd, J=12.8, 5Hz; 1H), 3.65-3.71 (m, 1H), 3.73 (d, J=3.4 Hz, 1H) 4.23-4.28 (m, 1H), 4.34-4.40 (m, 1H)
(500 MHz, D_2O)
pD=12

^{13}C NMR : δ_C (ppm) 18.7, 34.5 & 36.5 (equal), 47.4(mi), 47.5(ma), 51.5(mi), 52.0(ma), 58.2(mi), 60.1(ma), 60.3(ma), 60.4(mi), 176.4(ma), 177.8(mi), 178.5(ma), 178.6(mi)
(100 MHz, D_2O)
pD=12

^{13}C-DEPT : δ_C (ppm) *Positive Peaks:* 18.7, 47.4(mi), 47.5(ma), 58.2(mi), 60.1(ma), 60.3(ma), 60.4(mi); *Negative Peaks:* 34.5 & 36.5 (equal), 51.5(mi), 52.0(ma)
(100 MHz, D_2O)
pD=12

(*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-4-azidoproline dipeptide methylester (32)



Mol. Formula : $C_{14}H_{23}N_5O_5$
mp : 93.4 – 94.9 °C
Mol. Weight : 341.36
ESI-MS *m/z* : 342.10 $[M+1]^{+\bullet}$, 364.07 $[M+Na]^{+\bullet}$, 380.04 $[M+K]^{+\bullet}$
$[\alpha]_D^{25}$: -30.27 (C=0.5, $CHCl_3$)

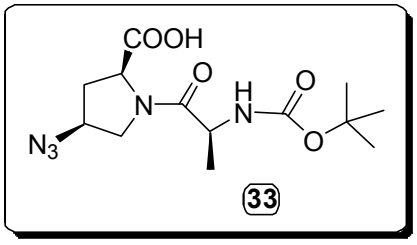
IR : ν (cm^{-1}) 3432, 3019, 2982, 2401, 2108, 1750, 1701, 1659, 1503, 1455, 1535, 1393, 1368, 1216, 1167, 1053, 1027, 929,
thin film on $CHCl_3$

^{1}H NMR : δ_H (ppm) 1.38 (d, J=6.9, 3H), 1.42 (s, 9H), 2.15-2.26 (m, 1H), 2.41-2.55 (ddd, J=13.6, 8.9, 5.9 Hz, 1H), 3.48-3.56 (dd, J=10.7, 4.3 Hz, 1H), 3.75 & 3.79 (s, 3H), 3.97-4.05 (dd, J=10.7, 6.3 Hz, 1H), 4.23-4.49 (m, 2H), 4.65-4.71 (dd, J=8.8, 4.4 Hz, 1H), 5.25 (d, J=7.7 Hz, 1H)
(200 MHz, $CDCl_3$)

^{13}C NMR : δ_C (ppm) 18.1, 28.2, 34.0, 47.6, 51.2, 52.5, 57.1, 59.3, 79.5, 155.2, 170.9, 171.9
(50 MHz, $CDCl_3$)

^{13}C-DEPT : δ_C (ppm) *Positive Peaks:* 18.1, 28.2, 47.6, 52.5, 57.1, 59.3; *Negative Peaks:* 34.0, 51.2
(50 MHz, $CDCl_3$)

(*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-4-azidoproline dipeptide (33)

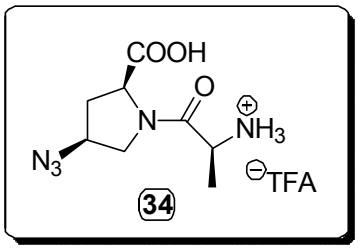


Mol. Formula : $C_{13}H_{21}N_5O_5$
Mol. Weight : 327.34
ESI-MS *m/z* : 328.26 $[M+1]^{+\bullet}$, 350.27 $[M+Na]^{+\bullet}$

IR : ν (cm^{-1}) 3554, 3370, 2924, 2854, 2724, 2116, 1741, 1675, 1634, 1462, 1377, 1316, 1189, 1163, 1051, 973, 849, 722, 666
thin film on nujol

^{1}H NMR : δ_H (ppm) 1.26 & 1.32 (2d, J=6.9 Hz, 3H) , 1.43 (s, 9H), 2.14-2.19 (m, 1H), 2.52-2.59 (m, 1H), 3.54-3.57 (dd, J=10.8, 3.4 Hz, 1H), 4.06-4.10 (dd, J=11, 6 Hz, 1H), 4.33 (q, J=6.9 Hz, 1H), 4.39-4.44 (tt, J=5.8, 3.8 Hz, 1H), 4.58-4.61 (dd, J=9.4, 3.9 Hz, 1H)
(400 MHz, MeOH-d_4)

(*S*)-alanine-(2*S*,4*S*)-4-azidoproline dipeptide TFA salt (34)



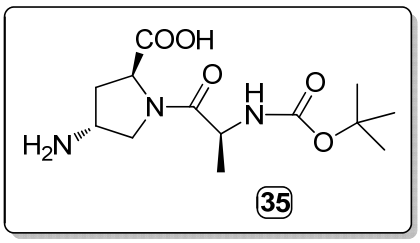
Mol. Formula : $C_8H_{14}N_5O_3^+$
Mol. Weight : 228.23
ESI-MS *m/z* : 228.25 $[M]^{+\bullet}$
$[\alpha]_D^{25}$: -16.54 (C=0.36, H_2O, pH=1.97)

^{1}H NMR (500 MHz, D_2O) pD=12 : **δ_H** (ppm) 1.09 & 1.20 (2d, J=6.8Hz, 3H), 2.01-2.05 & 2.32- 2.35 (dt, J=13.5, 4.8 Hz & dd, J=13.9, 1.1 Hz; 1H) and 2.43- 2.51 (m, 1H), 3.44 & 3.66 (2q, J=6.8 Hz; 1H), 3.47-3.51 (m, 1H) and 3.58-3.61 & 3.89-3.93 (dd, J=13.3, 5.6 Hz & dd, J=11.3, 6.2 Hz; 1H), 4.29-4.32 (m, 1H), 4.32-4.39 (m, 1H)

^{13}C NMR (100 MHz, D_2O) pD=12 : **δ_C** (ppm) 18.5(ma), 18.9 (mi), 34.0(mi), 36.0(ma), 47.4(mi), 47.7(ma), 51.4(ma), 51.7(mi), 57.9(ma), 59.1(mi), 60.2, 176.3(mi), 177.7(ma), 178.0(mi), 178.05(ma)

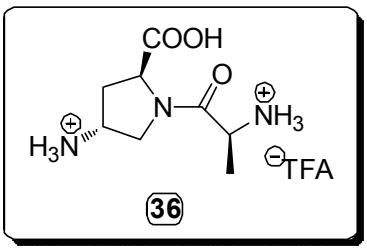
^{13}C-DEPT (125 MHz, D_2O) pD=12 : **δ_C** (ppm) *Positive Peaks:* 18.5(ma), 18.9 (mi), 47.4(mi), 47.7(ma), 57.9(ma), 59.1(mi), 60.2; *Negative Peaks:* 34.0(mi), 36.0(ma), 51.4(ma), 51.7(mi)

(*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*R*)-4-aminoproline dipeptide (35)



Mol. Formula : $C_{13}H_{23}N_3O_5$
Mol. Weight : 301.34
ESI-MS *m/z* : 302.42 $[M+1]^{+\bullet}$,324.47 $[M+Na]^{+\bullet}$
IR thin film on nujol : **ν** (cm^{-1}) 3307, 2924, 2854, 2726, 1445, 1461, 1377, 1305, 1169, 1066, 1022, 857, 722

(*S*)-alanine-(2*S*,4*R*)-4-aminoproline dipeptide TFA salt (36)



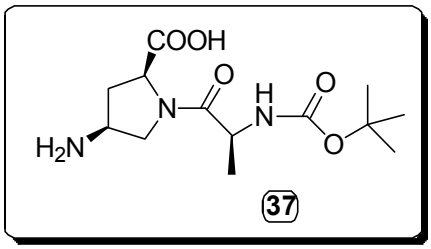
Mol. Formula : $C_8H_{17}N_3O_3^{2+}$
Mol. Weight : 203.23
ESI-MS *m/z* : 202.16 $[M]^{+\bullet}$
$[\alpha]_D^{25}$: -35.03 (C=0.36, H_2O, pH=1.74)

^{1}H NMR (500 MHz, D_2O) pD=12 : **δ_H** (ppm) 1.13-1.18 (2d, J=6.7 Hz, 3H), 2.01 (t, J=6.5 Hz, 1H) and 2.09-2.15 & 2.17-2.22 (2m, 1H), 3.47-3.51 & 3.57-3.61 (2m, 1H), 3.37-3.40 & 3.65-3.69 (dd, J=10.7, 4.3 Hz & dd, J=11.9, 6.5 Hz; 2H), 3.15-3.19 & 3.71-3.75 (dd, J=11.9, 5.8 Hz & dd, J=10.7, 5.7 Hz, 1H), 4.29 & 4.37-4.40 (t, J=7.5 Hz & dd, J=8.6, 4.6 Hz; 1H)

^{13}C NMR (100 MHz, D_2O) pD=12 : **δ_C** (ppm) 18.7 (ma),18.8 (mi), 37.1 (ma), 39.1 (mi), 47.3 & 47.4 (mi), 47.8 (ma), 49.8, 53.8 (mi), 54.3 (ma), 60.7 (ma), 61.1 (mi), 176.3 (ma), 177.5 (mi), 179.2 (mi), 179.3 (ma)

^{13}C-DEPT (125 MHz, D_2O) pD=12 : **δ_C** (ppm) *Positive Peaks:* 18.7 (ma),18.8 (mi), 47.8, 49.8, 60.7 (ma), 61.1 (mi); *Negative Peaks:* 37.1 (ma), 39.1 (mi), 53.8 (mi), 54.3 (ma)

(*S*)-N-(t-butoxycarbonyl)-alanine-(2*S*,4*S*)-4-aminoproline dipeptide (37)



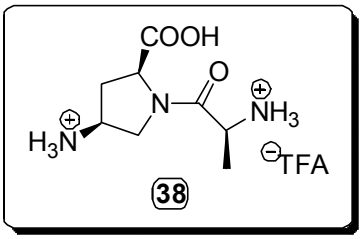
Mol. Formula : $C_{13}H_{23}N_3O_5$

Mol. Weight : 301.34

ESI-MS *m/z* : 302.47 $[M+1]^{+\bullet}$,324.51 $[M+Na]^{+\bullet}$

IR thin film on nujol : **ν** (cm^{-1}) 3353, 2925, 2855, 2726, 1702, 1683, 1654, 1609, 1527, 1460, 1376, 1290, 1247, 1172, 1058, 972, 861, 722, 668, 488

(*S*)-alanine-(2*S*,4*S*)-4-aminoproline dipeptide TFA salt (38)



Mol. Formula : $C_8H_{17}N_3O_3^{2+}$

Mol. Weight : 203.24

ESI-MS *m/z* : 202.26 $[M+1]^{+\bullet}$

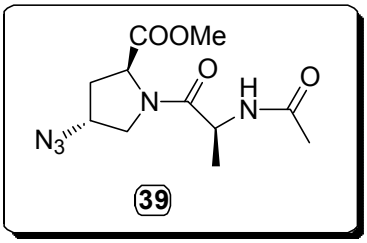
$[\alpha]_D^{25}$: -23.04 (C=0.36, H_2O, pH=1.83)

^{1}H NMR (500 MHz, D_2O) pD=12 : **δ_H** (ppm) 1.09 & 1.18 (2d, J=6.8 Hz, 3H), 1.54-1.59 & 1.77-1.82 (dt, J=12.7, 8.3 Hz & dt, J=12.9, 6.4 Hz; 1H), 2.42-2.49 & 2.55-2.64 (2m, 1H), 3.38 & 3.43 (t, J=6.4 Hz & t, J=7.4 Hz, 1H), 3.13-3.17 & 3.20-3.23 (dd, J=12, 6.2 Hz & dd, J=10.3, 7.8 Hz; 1H), 3.74-3.78 & 3.87-3.90 (dd, J=12.1, 6.6 Hz & dd, J=10.3, 6.7 Hz; 1H), 3.32-3.35 & 3.70 (m & q, J=6.9 Hz; 1H), 4.15 & 4.28-4.31 (t, J=8.3 Hz & dd, J=8, 6.2 Hz; 1H)

^{13}C NMR (100 MHz, D_2O) pD=12 : **δ_C** (ppm) 18.6(mi), 18.8(ma), 37.3(ma), 39.3(mi), 47.3(mi), 47.4(ma), 48.4(mi), 50.0, 54.1(ma), 54.2(mi), 61.0(mi), 61.2 (ma), 176.1(ma), 177.6(mi), 179.3(mi), 179.6(ma)

^{13}C-DEPT (125 MHz, D_2O) pD=12 : **δ_C** (ppm) *Positive Peaks:* 18.6(mi), 18.8(ma), , 47.3(mi), 47.4(ma), 48.4(mi), 50.0, 61.0(mi), 61.2 (ma);
Negative Peaks: 37.3(ma), 39.3(mi) 54.1(ma), 54.2(mi)

(*S*)-N-acetyl-alanine-(2*S*,4*R*)-4-azidoproline dipeptide methyl ester (39)



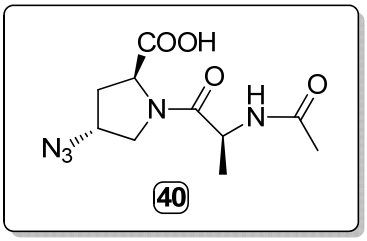
Mol. Formula : $C_{11}H_{17}N_5O_4$
Mol. Weight : 283.28
ESI-MS *m/z* : 284.20 $[M+1]^{+\bullet}$, 306.28 $[M+Na]^{+\bullet}$

^{1}H NMR (200 MHz, $CDCl_3$) : **δ_H** (ppm) 1.37 (d, J=6.8 Hz, 3H), 1.98 & 2.0 (2s, 3H), 2.18-2.25 & 2.33-2.48 (dd, J=7.6, 5.4 Hz & m; 3H), 3.76 (s, 3H), 3.79 (d, J=4.3 Hz, 2H), 4.30-4.38 (m, 1H), 4.59 (t, J=7.8 Hz, 1H), 4.73 (quin, J=7 Hz, 1H), 6.51 (d, J= 7.2 Hz, 1H)

^{13}C NMR (50 MHz, $CDCl_3$) : **δ_C** (ppm) 17.8, 23.0, 34.3, 46.6, 51.7, 52.5, 57.4, 59.6, 169.5, 171.4, 171.6

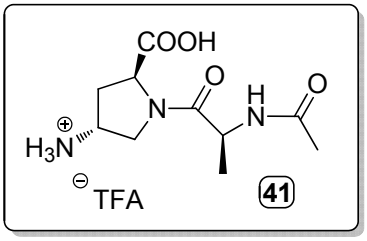
^{13}C-DEPT (50 MHz, $CDCl_3$) : **δ_C** (ppm) *Positive Peaks:* 17.8, 23.0, 46.6, 52.5, 57.4, 59.6; *Negative Peaks:* 34.3, 51.7

(*S*)-N-acetyl-alanine-(2*S*,4*R*)-4-aminoproline dipeptide (40)



Mol. Formula : $C_{10}H_{15}N_5O_4$
Mol. Weight : 269.26
ESI-MS *m/z* : 270.25 $[M+1]^{+\bullet}$, 292.23 $[M+Na]^{+\bullet}$

(*S*)-N-acetyl-alanine-(2*S*,4*R*)-4-aminoproline dipeptide (41)



Mol. Formula : $C_{10}H_{18}N_3O_4^+$
Mol. Weight : 244.27
ESI-MS *m/z* : 244.27 $[M]^{+\bullet}$
$[\alpha]_D^{25}$: -57.18 (C=0.36, H_2O, pH=1.86)

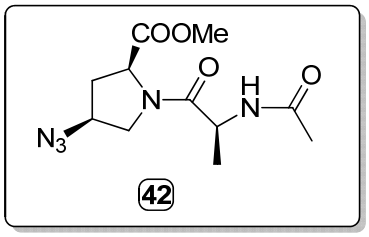
^{1}H NMR (500 MHz, D_2O) pD=12 : **δ_H** (ppm) 1.24 & 1.30 ()2d, J=7 Hz, 3H), 1.93 & 1.95 (2s, 3H), 2.01-2.04 & 2.08-2.25 (dd, J=7.1, 6.1 Hz & m; 1H), 3.48-3.51 & 3.53 (dd, J=10.7, 4.5 Hz & d, J=8 Hz; 1H), 3.62 (quin, J=5.4 Hz, 1H), 3.69-3.73 & 3.75-3.78 (dd, J=12, 7 Hz & dd, J=10.7, 5.7 Hz; 1H), 4.31 & 4.36-4.38 (t, J=7.4 Hz & dd, J=8.7, 3.4 Hz; 1H), 4.43 & 4.5 (2q, J=7 Hz, 1H)

^{13}C NMR (100 MHz, D_2O) : **δ_C** (ppm) 15.2(ma), 17.0(mi), 21.4(ma), 21.8(mi), 37.1(ma), 39.3(mi), 47.5(ma), 47.1 & 47.7(mi), 49.9, 53.8 (mi), 54.4 (ma),

pD=12 60.8(ma), 61.3(mi), 172.7(ma), 173.7(mi), 178.4(mi), 179.1(ma)

^{13}C-DEPT : **δ_C** (ppm) *Positive Peaks:* 15.2(ma), 17.0(mi), 21.4(ma), 21.8(mi), 47.5(ma), 47.1 & 47.7(mi), 49.9, 60.8(ma), 61.3(mi); *Negative Peaks:* 37.1(ma), 39.3(mi), 53.8 (mi), 54.4 (ma)
(125 MHz, D_2O)
pD=12

(*S*)-N-acetyl-alanine-(2*S*,4*S*)-4-azidoproline dipeptide methyl ester (42)



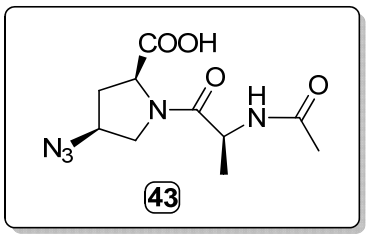
Mol. Formula : $C_{11}H_{17}N_5O_4$
Mol. Weight : 283.28
ESI-MS *m/z* : 284.39 $[M+1]^{+\bullet}$, 306.41 $[M+Na]^{+\bullet}$

^{1}H NMR : **δ_H** (ppm) 1.41 (d, J=6.9 Hz, 3H), 1.99 & 2.01 (2s, 3H), 2.17-2.28 (m, 2H), 2.42-2.56 (m, 2H), 3.51-3.59 (dd, J=10.9, 4.2, Hz, 1H), 3..76 & 3.77 (2s, 3H), 3.8 (d, J=4.4 Hz, 1H), 4.01-4.10 (dd, J=10.9, 6.1 Hz, 1H), 4.25- 4.35 (m,1H), 4.66-4.71 (m, 1H), 6.50 (bs, 1H)
(200 MHz, $CDCl_3$)

^{13}C NMR : **δ_C** (ppm) 17.4, 22.6, 33.9, 46.4, 51.3, 52.4, 57.0, 59.1, 170.0, 170.6, 171.9
(50 MHz, $CDCl_3$)

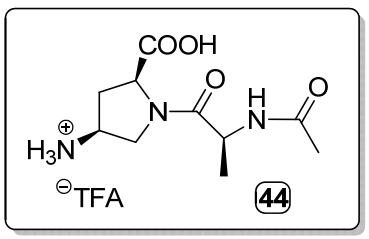
^{13}C-DEPT : **δ_C** (ppm) *Positive Peaks:* 17.4, 22.6, 46.4, , 52.4, 57.0, 59.1; *Negative Peaks:* 33.9, 51.3
(50 MHz, $CDCl_3$)

(*S*)-N-acetyl-alanine-(2*S*,4*S*)-4-aminoproline dipeptide (43)



Mol. Formula : $C_{10}H_{15}N_5O_4$
Mol. Weight : 269.26
ESI-MS *m/z* : 270.23 $[M+1]^{+\bullet}$

(*S*)-N-acetyl-alanine-(2*S*,4*S*)-4-aminoproline dipeptide (44)



Mol. Formula : $C_{10}H_{18}N_3O_4^+$
Mol. Weight : 244.27
ESI-MS *m/z* : 244.20 $[M]^{+\bullet}$
$[\alpha]_D^{25}$: 47.10 (C=0.354, H_2O, pH=1.77)

^{1}H NMR : **δ_H** (ppm) 1.21-1.30 (2d, J=7 Hz, 3H), 1.56-1.61 & 1.81-1.86 (dt, J=12.7, 8.2 Hz & m, 1H), 1.94 (s, 3H), 2.45-2.51 & 2.56-2.62 (m, 1H), 3.17-3.21 & 3.23-3.27 (m & dd, J= 10.2, 7.8 Hz; 1H), 3.39-3.42
(500 MHz, D_2O)
pD=12

& 3.44-3.50 (2m, 1H), 3.99-4.02 & 4.04-4.08 (dd, J=10.3, 6.7 Hz & m; 1H), 4.16 & 4.28-4.31 (t, J=8.2 Hz & dd, J=9.1, 5.1 Hz; 1H), 4.38-4.42 & 4.49-4.53 (m & q, J= 7.1 Hz; 1H)

^{13}C NMR (100 MHz, D_2O, pD=12) : **δ_C** (ppm) 15.4, 21.4, 37.2, 47.5, 50.1, 54.3, 61.3, 172.6, 173.7, 179.2

^{13}C-DEPT (125 MHz, D_2O, pD=12) : **δ_C** (ppm) *Positive Peaks:* 15.4(ma), 17.0(mi), 21.4(ma), 21.8(mi), 47.5(ma), 47.2 & 48.4(mi), 50.1, 61.1(mi), 61.3(ma); *Negative Peaks:* 37.2(ma), 39.5(mi), 54.3(ma), 54.5(mi)

2.8.2 HPLC purification

Dipeptides (**19**, **23**, **28**, **31**, **34**, **36**, **38**, **41** and **44**) were purified by reverse phase-HPLC on Waters 600 equipped with 2998-Photodiode array detector (PDA). Semi-preparative RP-C18 columns (250 x 10 mm, 10 µm) of Allteck-Alltima make were used for peptides. A combination of gradient and isocratic methods were used to elute the peptide at a flow rate of 3 ml/min and the eluant was monitored at 220 nm. The peaks corresponding to the dipeptides were collected and the fractions were lyophilized. The spectra acquisition, analysis and processing was done on Waters Empower-2154 software. The absorbance of the eluant was monitored at its corresponding wavelength and the purity was obtained from the integrator output.

The following solvent system composition was used for different peptides.

Method A: Gradient; 5% to 80% MeCN/H_2O with 0.1% TFA in 20 minutes for dipeptides **19**, **31** and **34**

Method B: Isocratic; 3% MeCN/H_2O with 0.1% TFA for 12 min for dipeptides **23**, **28**, **41** and **44**

Method C: Isocratic: only H_2O with 0.1% TFA for 12 min for dipeptides **36** and **38**

The purities of the HPLC purified peptides were found to be more than 90%.

2.8.3 pK_a Determiation

The pH of dipeptides **19**, **36**, **38**, **41** and **44** (50mM, 1 ml) in deionized water was adjusted to 2.0 using conc. HCl. This solution was titrated with 5 µl aliquots of 0.5 M aq. NaOH. After each addition of NaOH solution, pH was recorded after the reading reached a stable value (1 min). The pK_a values were derived from the first derivative of the plot of pH *vs* volume of NaOH.

2.8.4 pH dependent NMR spectroscopy

NMR experiments at different pH were performed on a Bruker 500 MHz instrument controlled by Bruker Topspin NMR software. The ratio of the *trans* : *cis* isomers of compounds (**19**, **23**, **28**, **31**, **34**, **36**, **38**, **41** and **44**) were directly obtained from the peak integration of ^{1}H signals (methyl doublet of the alanyl residue) in case of both minor (*cis*) and major (*trans*) isomers. A typical experiment was performed by first bringing the sample to the lowest pH followed by equilibration (15 – 20 min). The pH was increased in steps to 2.0, 4.0, 7.0, 9.5 and 12.0. Spectra were collected at 0.1 Hz resolution and were an average of 256 scans. In order to achieve a reliable integration, a post acquisition delay of 5 seconds was applied and the number of scans was kept constant throughout the experiment. Necessary phase corrections were applied and the peak integration was taken for each set of well-separated signals. The percentage of integration values for either of the isomers for different pH were plotted into graphs using MicroCal Origin 8.0 software.

2D-^{1}H-NMR of representative dipeptides were also recorded for assigning the position of both *trans* & *cis* peaks.

Samples were prepared in D_2O and various pH values were adjusted in the NMR tube (5 mm) by adding small aliquots of NaOD /D_2O with the help of Sigma-Hanna pH electrode (Ag/AgCl, 3 x 130 mm, nmr tube application specific). In order to remove any volatile impurities present, all the HPLC purified samples were freeze dried, desiccated and the resulting solids were redissolved in D_2O.

2.8.5 pH dependent CD spectroscopy

Circular dichroism (CD) spectrometry was carried out on JASCO J-715 spectropolarimeter using cylindrical, jacketed quartz cell (1 mm path length). Spectra were recorded with a spectral resolution of 0.05 nm, band width 1 nm at a scan speed of 200 nm/min with a response time 1 sec. All the spectra were collected at room temperature and corrected for respective buffer condition and are typically averaged over 5 scans. Data processing was performed using MicroCal Origin 8.0 software.

The dipeptide samples (5 mM) were prepared in deionized water without any extra buffer and the various pH values (2, 4, 7, 9.5 and 12) were achieved using aq.

TFA and aq. NaOH, apart from one set having its original pH. Separate samples of different pH values were prepared to avoid excessive variation in concentration.

2.8.6 X-ray crystal structure determination

X-ray diffraction data for all the crystallized compounds were collected at *T* = 296 K, on SMART APEX CCD Single Crystal X-ray diffractometer using Mo-Kα radiation (λ = 0.7107 Å) to a maximum θ range of 25.00°. Crystal to detector distance was 6.05 cm, 512 x 512 pixels / frame and other conditions used are oscillation / frame (-0.3º), maximum detector swing angle (–30.0º), beam center (260.2, 252.5) and in plane spot width (1.24). SAINT integration and SADABS correction were also applied. The structures were solved by direct methods using SHELXTL. All the data were corrected for Lorentzian, polarisation and absorption effects. SHELX-97 (ShelxTL)[61] was used for structure solution and full matrix least squares refinement on F^2. Hydrogen atoms were included in the refinement as per the riding model. The refinements were carried out using SHELXL-97.

Compound 18

Single crystals of the compound were found to grow best in solution mixture of methanol and water by slow evaporation. Colorless needle like crystal of approximate size 0.13 x 0.05 x 0.03 mm, was used for data collection. Multirun data acquisition, total scans (3), total frames (1818), exposure / frame (15.0 sec), θ range (1.80 to 24.99º) and completeness to θ of 24.99º (99.9%) were registered. The compound has molecular formula $C_{13}H_{22}N_2O_5 .H_2O$ and *M* = 304.34. Crystals belong to Orthorhombic system with $P2_12_12_1$ space group with unit cell dimensions as *a* = 5.8687(16), *b* = 13.458(4), *c* = 20.747(6) Å. Other parameters like volume 1638.6(8) $Å^3$, *Z* = 4, D_c = 1.234 g/cc, absorption coefficient *μ* (Mo–$K_α$) = 0.097 mm^{-1} were also recorded. 11542 reflections measured of which 2874 are unique. The final R indices [I>2σ(I)] are R1 (0.1030) and wR2 (0.2414).

Compound 22

Single crystals of the compound were found to grow best in solution of methanol by slow evaporation. Colorless needle-like crystal of approximate size 0.19 x 0.14 x 0.01 mm, was used for data collection. Quadrant data acquisition, total scans (4), total frames (2424), exposure / frame (25.0 sec), θ range (2.48 to 25.0º) and completeness to θ of 25.0º (99.9%) were registered. The compound has

molecular formula $C_{13}H_{22}N_2O_6$ and *M* = 302.33. Crystals belong to Monoclinic system with $P2_1$ space group with unit cell dimensions as *a* = 5.5597(4), *b* = 10.5695(8), *c* = 13.198(11) Å. Other parameters like volume 767.12(10) Å^3, *Z* = 2, D_c = 1.309 g/cc, absorption coefficient μ (Mo–K_α) = 0.104 mm^{-1} were also recorded. 7482 reflections measured of which 2701 are unique. The final R indices [I>2σ(I)] are R1 (0.0474) and wR2 (0.1005).

Compound 26

Single crystals of the compound were found to grow best in solution mixture of petroleum ether and ethyl acetate by slow evaporation. Colorless crystal of approximate size 0.45 x 0.32 x 0.25 mm, was used for data collection. Multi-scan data acquisition, total scans (3), total frames (1818), exposure / frame (10.0 sec), θ range (3.17 to 28.28º) and completeness to θ of 28.28º (100%) were registered. The compound has molecular formula $C_{14}H_{24}N_2O_6$ and *M* = 316.16. Crystals belong to Monoclinic system with C2 space group with unit cell dimensions as *a* = 12.526(15), *b* = 5.694(7), *c* = 26.17(3) Å. Other parameters like volume 1835(4) Å^3, *Z* = 4, D_c = 1.145 g/cc, absorption coefficient μ (Mo–K_α) = 0.089 mm^{-1} were also recorded. 3769 reflections measured of which 2793 are unique. The final R indices [I>2σ(I)] are R1 (0.1030) and wR2 (0.2414).

Compound 32

Single crystals of the compound were found to grow best in solution mixture of chloroform and petroleum ether by slow evaporation. Colorless rectangular crystal of approximate size 0.43 x 0.33 x 0.03 mm, was used for data collection. Hemisphere data acquisition, total scans (3), total frames (1271), exposure / frame (10.0 sec), θ range (2.32 to 28.15º) and completeness to θ of 28.15º (92.1%) were registered. The compound has molecular formula $C_{14}H_{23}N_5O_5$, *M* = 341.37. Crystals belong to Monoclinic system with $P2_1$ space group with unit cell dimensions as *a* = 5.6061(5), *b* = 11.078(1), *c* = 14.469(1) Å. Other parameters like volume 896.90(14) Å^3, *Z* = 2, D_c = 1.264 g/cc, absorption coefficient μ (Mo–K_α) = 0.097 mm^{-1} were also recorded. 5349 reflections measured of which 2937 are unique. The final R indices [I>2σ(I)] are R1 (0.0708) and wR2 (0.1809).

Compound 33

Single crystals of the compound were found to grow best in solution mixture of methanol and dichloromethane by slow evaporation. Colorless rectangular crystal of approximate size 0.42 x 0.18 x 0.02 mm, was used for data collection. Quadrant data acquisition, total scans (4), total frames (2424), exposure / frame (5.0 sec), θ range (1.39 to 25.0º) and completeness to θ of 25.0º (100.0%) were registered. The compound has molecular formula $C_{13}H_{21}N_5O_5$ and *M* = 327.35. Crystals belong to Monoclinic system with $P2_1$ space group with unit cell dimensions as *a* = 5.5489(4), *b* = 10.3689(8), *c* = 14.6022(11) Å. Other parameters like volume 840.10(11) $Å^3$, *Z* = 2, D_c = 1.294 g/cc, absorption coefficient μ (Mo–K_α) = 0.101 mm^{-1} were also recorded. 8163 reflections measured of which 2959 are unique. The final R indices [I>2σ(I)] are R1 (0.0417) and wR2 (0.0956).

2.9 References

1) Anfinsen, C. B. *Science* **1973**, *181*, 223-230.

2) Creighton, T. E. *Proteins: Structures and Molecular properties,* 2nd ed.; W. H. Freeman & Co., **1993**, 173-174.

3) Vila, J. A.; Baldoni, H. A.; Ripoll D. R.; Ghosh, A.; Scheraga, H. A. *Biophys. J.* **2004**, *86* 731–742

4) Brandts, J. F.; Halvorson, H. R.; Brennan, M. *Biochemistry* **1975**, *14*, 4853-4963.

5) (a) Weiss, M. S.; Jabs, A.; Hilgenfeld, R. *Nat. Struct. Biol.* **1998**, *5*, 676. (b) Jabs, A; Weiss, M. S.; Hilgenfeld, R. *J. Mol. Biol.* **1999**, *286***,** 291-304. (c) Pall, D.; Chakraabarti, P. *J. Mol. Biol.* **1999**, *294***,** 271-288.

6) Andreotti, A. H. *Biochemistry,* **2003**, *42,* 9515-9524.

7) Taylor, C. M.; Hardre, R.; Edwards, P. J. B.; Park, J. H. *Org. Lett.* **2003,** *5*, 4413-4416.

8) Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G. *J. Mol. Biol*. **1998,** *279*, 449-460.

9) (a) Dealaney, N. G.; Madison, V. *J. Am. Chem. Soc*. **1982,** *104*, 6635-6641. (b) Beausoleil, E.; Sharma, R. M.; Michnick, S. W.; Lubell, W. D. *J. Org. Chem*. **1998**, *63*, 6572-6578. (c) Magaard, V. W.; Sanchez, R. M.; Bean, J. W.; Moore, M. L. *Tetrahedron Lett.* **1993,** *34,* 381-384. (d) Aa, S.S. A.; Lester, C. C.; Peng, J. L.; Li, Y.J.; Rothwarf, D. M.; Welker, E.; Tannhauser, T. W.; Zhang, L. S.; Tam, J. P.; Scheraga, H. A. *J. Am. Chem. Soc.* **1999,** *121,* 11558-11566. (e) Halab, L.; Gosselin, F.; Lubell, W. D. *Biopolymers* **2000**, *55*,101-122. (f) Arnold, U.; Hinderaker, M. P.; Koeditz, J.; Golbik, R.; Ulbrich-Hoffman, R.; Raines, R. T. *J. Am. Chem. Soc*. **2003**, *125*, 7500-7501.

10) (a) Andres, C. J.; Macdonald, T. L.; Ocain, T. D.; Longhi, D. *J. Org. Chem.* **1993**, *58*, 6609-6613. (b) Welch, J. T.; Lin, J. *Tetrahedran* **1996**, *52*, 291-304. (c) Hart, S. A.; Sabt, M.; Etzkorn, F. A. *J. Org. Chem*. **1998**, *63*, 7580-7581. (d) Lin, J.; Toscano, P. J.; Welch, J. T. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14020-14024. (e) Otaka, A.; Katagiri, F.; Kinoshita, T.; Odagaki, Y.; Oishi, S.; Tammamura, H.; Hamanaka, N.; Fujii, N. *J. Org. Chem*. **2002**, *67*, 6152-6161. (f) Wang, X. J.; Hart, S. A.; Xu, B.; Mason, M. D.; Goodell, J. R.; Etzkorn, F. A. *J. Org. Chem.* **2003**, *68*, 2343-2349.

11) (a) Eberhardt, E. S.; Panasik, N. J.; Raines, R. T. *J. Am. Chem. Soc*. **1996**, *118*, 12261-12266. (b) Bretscher, L. E.; Taylor, K. M.; Raines, R. T. *In Peptides for the New Millennium: Proceedings of the Sixteenth American Peptide Symposium*; Dordrecht, The Netherlands, **2000**, 335-356. (c) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. *J. Am. Chem. Soc*. **2001**, *123*, 777-778. (d) Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L. *Angew. Chem. Int. Ed.* **2001**, *40*, 923-925.

12) (a) Haasnoot, C. A. G.; De Leeuw, A. A. M.; De Leeuw, H. P. M.; Altona, C. *Biopolymers* **1981**, *20*, 1211-1245. (b) Flores-Ortega, A.; Casanovas, J.; Assfeld, X.; Alemán, C. *J. Org. Chem.* **2009**, *74*, 3101–3108. (c) Shoulders, M. D.; Raines, R. T. *Annu. Rev. Biochem.* **2009**, *78*, 929–958.

13) Jankowski, K.; Soler, F.; Ellenberger, M. *J. Mol. Struct.***1978,** *48*, 63-68.

14) Eberhardt, E. S.; Loh, S. N.; P; Raines, R. T. *Tetrahedron Lett.* **1993**, *33*, 3055-3056.

15) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. *J. Am. Chem. Soc.* **2001**, *123*, 777-778.

16) Shoulders, M. D.; Hodges, J. A.; Raines, R. T., *J. Am. Chem. Soc.* **2006,** *128*, 8112-8113.

17) (a) Wolfe, S. *Acc. Chem. Res.* **1972**, *5*, 102. (b) Brunck, T. K.; Weinhold, F. *J. Am. Chem. Soc.* **1979**, *101*, 1700-1790. (c) Senderowitz, H.; Aped, P.; Fuchs, B. *J. Comput. Chem.* **1993**, *14*, 944. (d) Houk, K. N.; Eksterwicz, J. E.; Wu, Y. D.; Fuglesang, C. D.; Mitchell, D. B. *J. Am. Chem. Soc.* **1993**, *115*, 4170.

18) Parker, D.; Senanayake, K.; Vepsailainen, J.; Williams, S.; Batsanov A. S.; Howard, J. A. K. *J. Chem. Soc.*, *Perkin Trans. 2,* **1997**, *8*, 1445-1452.

19) Wiberg, K. B. *Acc. Chem. Res.* **1996**, *29*, 229-234.

20) Deslongchamps, P. *Stereoelectronic Effects in Organic Chemistry*; Pergamon Press: New York, **1983**.

21) Gorenstein, D. G. *Chem. Rev*. **1987**, *87*, 1047-1077.

22) Plavec, J.; Thibaudeau, C.; Chattopadhyaya, J. *J. Am. Chem. Soc*. **1994**, *166*, 6558-6560.

23) (a) Panasik, N. Jr.; Eberhardt, E. S.; Edison, A. S.; Powell, D. R.; Raines, R. T. *Int. J. Pept. Protein Res.* **1994**, *44*, 262-269. (b) Gerig, J. T.; McLeod, R. S. *J. Am. Chem. Soc.* **1973**, *95*, 5725-5729.

24) Zimmerman, S. S.; Scheraga, H. A. Macromolecules **1976**, *9*, 408-416.

25) (a) Maccallum, P. H.; Poet, R.; Milner-White, E. J. *J. Mol. Biol*. **1995**, *248*, 361-373. (b) Maccallum, P. H.; Poet, R.; Milner-White, E. J. *J. Mol. Biol*. **1995**, *248*, 374-384.

26) (a) DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markely, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 2497-2505. (b) Hinderaker, M. P.; Raines, R. T. *Protein Sci*. **2003**, *12*, 1188-1194.

27) (a) Bachinger, H. P.; Bruckner, P.; Timpl, R.; Prockop, D. J.; Engel, J. *Eur. J. Biochem.* **1980**, *106*, 619-632.(b) Bachinger, H. P. *J. Biol. Chem*. **1987**, *262*, 17144-17148. (c) Liu, X.; Siegel, D. L.; Fan, P.; Brodsky, B.; Baum, J. *Biochemistry* **1996**, *35*, 4306-4313.

28) (a) Merker, M. P.; Armitage, I. M.; Audi, S. H.; Kakalis, L. T.; Linehan, J. H.; Maehl, J. R.; Roerig, D. L. ; Dawson, C. A. Am. *J. Physiol.-Lung Cell Mol. Physiol*.**1996**. *14*, L251-L259. (b) Brandsch, M.; Thunecke, F.; Schutkowski, M.; Fischer, G.; Neubert, K. *J. Biol. Chem*. **1998**. *273*, 3861-3864. (c) Vanhove, M.; Raquet, X.; Palzkill, T.; Pain, R. H.; Frere, J. M. *Proteins: Struct. Funct. Genet.* **1996**, *25*, 104-111.

29) (a) Eckert, B.; Martin, A.; Balbach, J.; Schmid, F. X. *Nat. Struct. Mol. Biol.* **2005**, *12*,619-623. (b) Wedemeyer, W.J.; Welker, E.; Scheraga, H. A. *Biochemistry* **2002**, *41,* 14637-14644.

30) Wu, Y.; Matthews, C. R. *J. Mol. Biol.* **2002**, 322, 7-13.

31) Brazin, K. N.; Mallis, R. J.; Fulton, D. B.; Andreotti, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1899.

32) Schmid, F. X.; Mayr, L. M.; Mücke, M. Schönbrunner, E. R. *Advan. Protein Chem.* **1993**, *44***,** 25-66.

33) Dugave, C.; Demange, L. *Chem. Rev.* **2003**, *103*, 2475-2532.

34) (a) Farrera-Sinfreu, J.; Giralt, E.; Castel, S.; Albericio, F.; Royo, M. *J. Am. Chem. Soc.* **2005**, *127*, 9459–9468. (b) Fillon, Y. A.; Anderson, J. P.; Chmielewski, J. *J. Am. Chem. Soc.* **2005**, *127*, 11798–11803.

35) Lummis, S. C. R.; Beene, D. L.; Lee, L. W.; Lester, H. A.; Broadhurst, R. W.; Dougherty, D. A. *Nature (London)*, **2005**, *438*, 248-252.

36) Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L. *Angew. Chem. Int. Ed*. **2001**, *40*, 923-925.

37) Lu, K. P.; Finn, G.;Lee, T. H.; Nicholson, L. K. *Nat. Chem. Biol*. **2007**, *3*, 619-629.

38) (a) Traub, W.; Piez, K. *Adv. Protein Chem.* **1971**, *25*, 243-341. (b) Ramachandran, G. N.; Ramakrishnan, C. In *Biochemistry of Collagen*; Ramachandran, G. N., Reddi, A. H., Eds.; Plenum Press: NY **1976**, 45-84.

39) (a) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. *Science* **1994**, *266*, 75-81. (b) Venugopal, M. G.; Ramshaw, J. A. M.; Braswell, E.; Zhu, D.; Brodsky, B. *Biochemistry* **1994**, *33*, 7948-7956.

40) Ramachandran G. N.; *Aspects of Protein Structure*; Ramachandran G. N., E.; Academic Press: NY, **1963**, 39.

41) (a) Sakakibara, S.; Inouye, K.; Shudo, K.; Kishida, Y.; Kobayashi, Y.; Prockop, D. J. *Biochim. Biophys. Acta* **1973**, *303*, 198-202. (b) Inouye, K.; Kobayashi, Y.; Kyogoku, Y.; Kishida, Y.; Sakakibara, S.; Prockop, D. J. *Arch. Biochem. Biophys*. **1982**, *219*, 198-203.

42) (a) Ramachandran, G. N.; Bansal, M.; Bhatnagar, R. S. *Biochim. Bophys. Acta* **1973**, *322*, 166-171. (b) Bella, J.; Brodsky, B.; Berman, H. M. *Structure* **1995**, *3*, 893-906.

43) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. *Nature* **1998**, *392*, 666-667.

44) Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T. *Chem. Biol.* **1999**, *6*, 63-70.

45) Babu, I. R. PhD thesis, University of Pune, India, December **2001**.

46) Umashankara, M. PhD thesis, University of Pune, India, October **2006**.

47) (a) Babu, I. R.; Ganesh, K. N. *J. Am. Chem. Soc.* **2001**, *123*, 2079-2080, (b) M. Umashankara, M.; Babu, I.R.; Ganesh, K. N. *Chem. Commun*. **2003**, 2606-2607.

48) Bailey, P. D.; Boyd, C. A. R.; Collier, I. D.; Kellett, G. L.; Meredith, D.; Morgan, K. M.; Pettecrew, R.; Price, R. A. *Org. Biomol. Chem.***2005**, *3*, 4038-4039.

49) Schutkowski, M.; Landgraf, G.; Born, I.; Neubert, K.; Fischer, G. *Eur. J. Biochem.***1997**, *245*, 381.

50) (a) Grathwohl C.; Wuthrich K.; *Biopolymers*, **1976**, *15***,** 2025-2041, (b) Gerothanassis I.P.; *Biopolymers*, **2000**, *53*, 72-83.

51) Glastone, S. A. *Text Book of Physical Cheistry* 2nd ed. Macmillan India Ltd.,**1946**, 1002-1005.

52) (a) Panasik, N., Jr.; Eberhardt, E. S.; Edison, A. S.; Powell, D. R.; Raines, R. T. *Int. J. Pept. Protein Res.* **1994**, *44*, 262-269. (b) Gerig, J. T.; McLeod, R. S. *J. Am. Chem. Soc.* **1973**, *95*, 5725-5729.

53) (a) Barchi, J. J. Jr.; Karki, R. G.; Nicklaus, M. C.; Siddiqui, M. A.; George, C.; Mikhailopulo, I. A.; Marquez, V. E. *J. Am. Chem. Soc.* **2008**, *130*, 9048-9057; (b) van den Hoogen, Y. T.; Hilgersom, C. M.; Brozda, D.; Lesiak, K.; Torrence, P. F.; Altona, C. *Eur. J. Biochem.* **1989**, *182*, 629-637.

54) Taylor, C. M.; Hardré, R.; Edwards, P. J. B. *J. Org. Chem.* **2005,** *70*, 1306-1315.

55) Koskinen, A. M. P.; Helaja, J.; Kumpulainen, E. T. T.; Koivisto, J.; Mansikkamäki, H.; Rissanen, K. *J. Org. Chem.* **2005,** *70*, 6447-6453.

56) Alema, C.; Flores-ortega, A.; Casanovas, J.; Nussinov, R. *J. Phy. Chem. B* **2008,** 112, 14045-14055.

57) (a) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, **1986**. (b) Wüthrich, K.; Billeter, M.; Braun, W. *J. Mol. Biol.* **1984**, *180*, 715–740.

58) Sonntag, L.S.; Schweizer, S.; Ochsenfeld, C.; Wennemers, H. *J. Am. Chem. Soc.* **2006,** *128*, 14697-14703

59) Sonar, M. V.; Ganesh, K. N. *Org. Lett.* **2011,** 12, (23), 5390-5393.

60) Kuemin, M.; Schweizer, S.; Ochsenfeld, C.; Wennemers, H. *J. Am. Chem. Soc.* **2009**, *131*, 15474-15482.

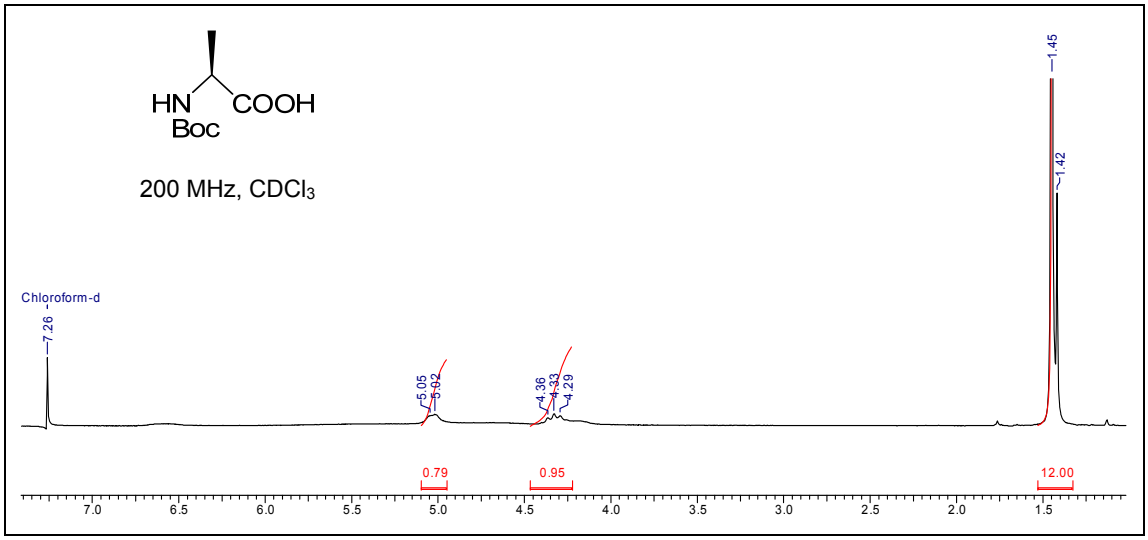
61) Sheldrick, G. M. *Acta Crystallogr.* **2008**, *A64*, 112-122.

2.10 Appendix B: Characterization data of synthesized compounds and dipeptides

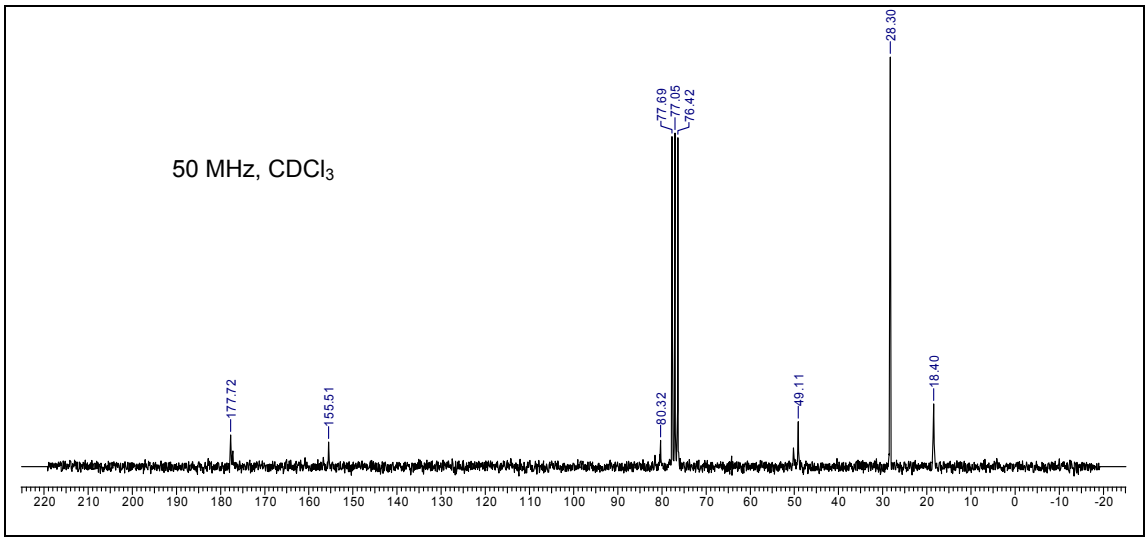
Designation	Description	Page
Compound **14**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR	234
Compound **17**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, ^{1}H-^{13}C HSQC LC-MS, FT-IR	235-237
Compound **18**	^{1}H NMR, LC-MS Crystal structure	238 293
Compound **19**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, ^{1}H-^{1}H COSY & NOESY, Variable pH ^{1}H NMR, LC-MS	239-241
	HPLC profile	288
Compound **21**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, ^{1}H-^{13}C HSQC, ^{1}H-^{1}H COSY, LC-MS, FT-IR	242-244
Compound **22**	^{1}H NMR, LC-MS Crystal structure	245 294
Compound **23**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, Variable pH ^{1}H NMR, ^{1}H-^{1}H NOESY	246-248
	HPLC profile	289
Compound **24**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	249-250
Compound **25**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	251-252
Compound **26**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR Crystal structure	253-254 295
Compound **27**	^{1}H NMR, LC-MS	255
Compound **28**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS Variable pH ^{1}H NMR, ^{1}H-^{1}H NOESY	256-258
	HPLC profile	289
Compound **29**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	259-260
Compound **30**	^{1}H NMR, LC-MS, FT-IR	261
Compound **31**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS Variable pH ^{1}H NMR, ^{1}H-^{1}H NOESY	262-264
	HPLC profile	290
Compound **32**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR Crystal structure	265-266 296
Compound **33**	^{1}H NMR, LC-MS, FT-IR Crystal structure	267 297

Designation	Description	Page
Compound **34**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	268-270
	Variable pH ^{1}H NMR, ^{1}H-^{1}H NOESY	
	HPLC profile	290
Compound **35**	LC-MS, FT-IR	271
Compound **36**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, ^{1}H-^{1}H COSY & NOESY, Variable pH ^{1}H NMR, LC-MS	272-274
	HPLC profile	291
Compound **37**	LC-MS, FT-IR	275
Compound **38**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, ^{1}H-^{1}H COSY & NOESY, Variable pH ^{1}H NMR, LC-MS	276-278
	HPLC profile	291
Compound **39**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	279-280
Compound **40**	LC-MS	285
Compound **41**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, Variable pH ^{1}H NMR, LC-MS	281-282
	HPLC profile	292
Compound **42**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	283-284
Compound **43**	LC-MS	285
Compound **44**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, Variable pH ^{1}H NMR, LC-MS	286-287
	HPLC profile	292

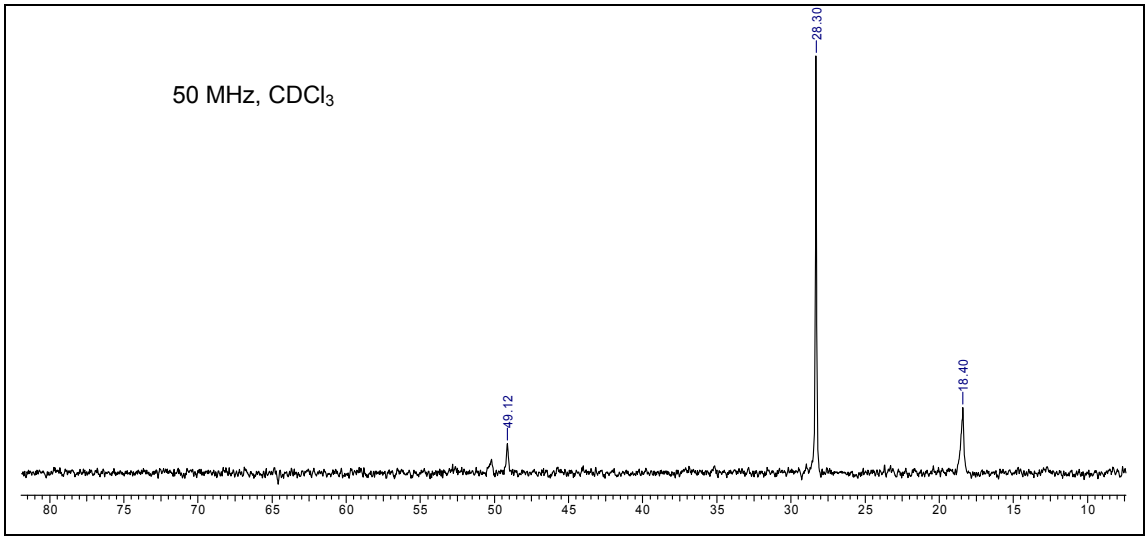
^{1}H NMR of compound 14



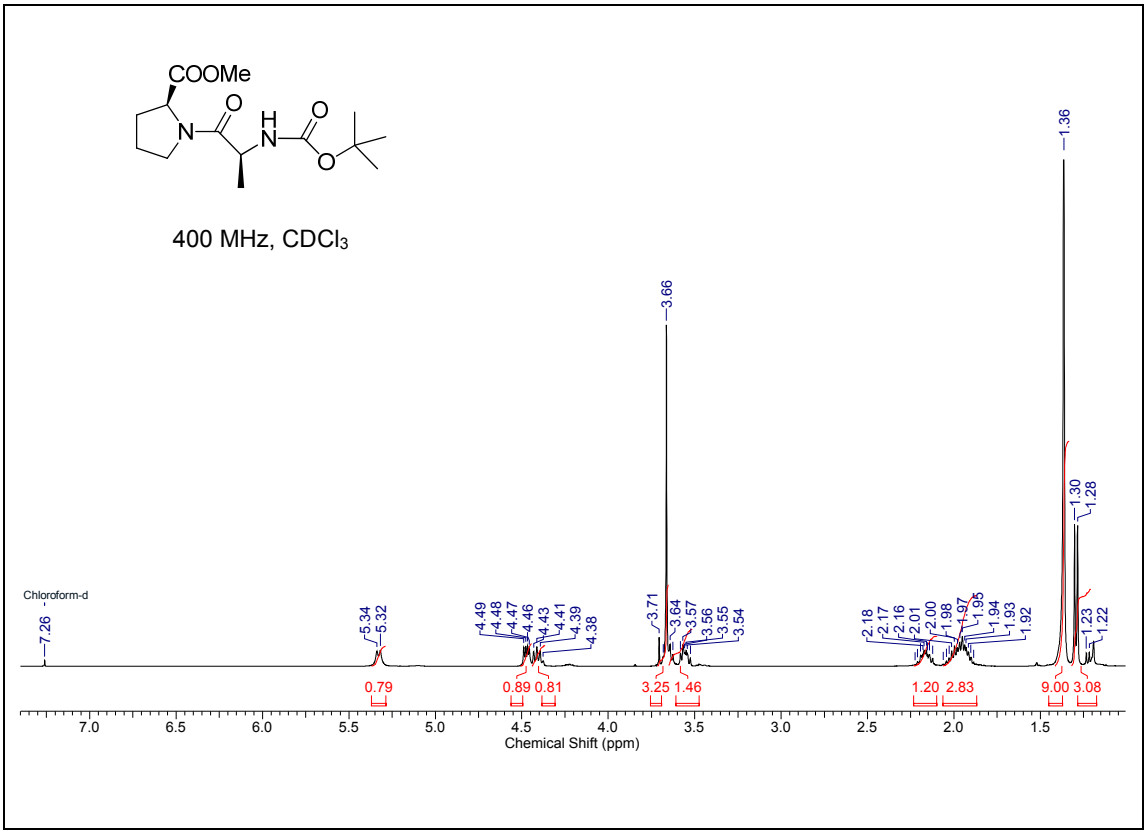
^{13}C NMR of compound 14



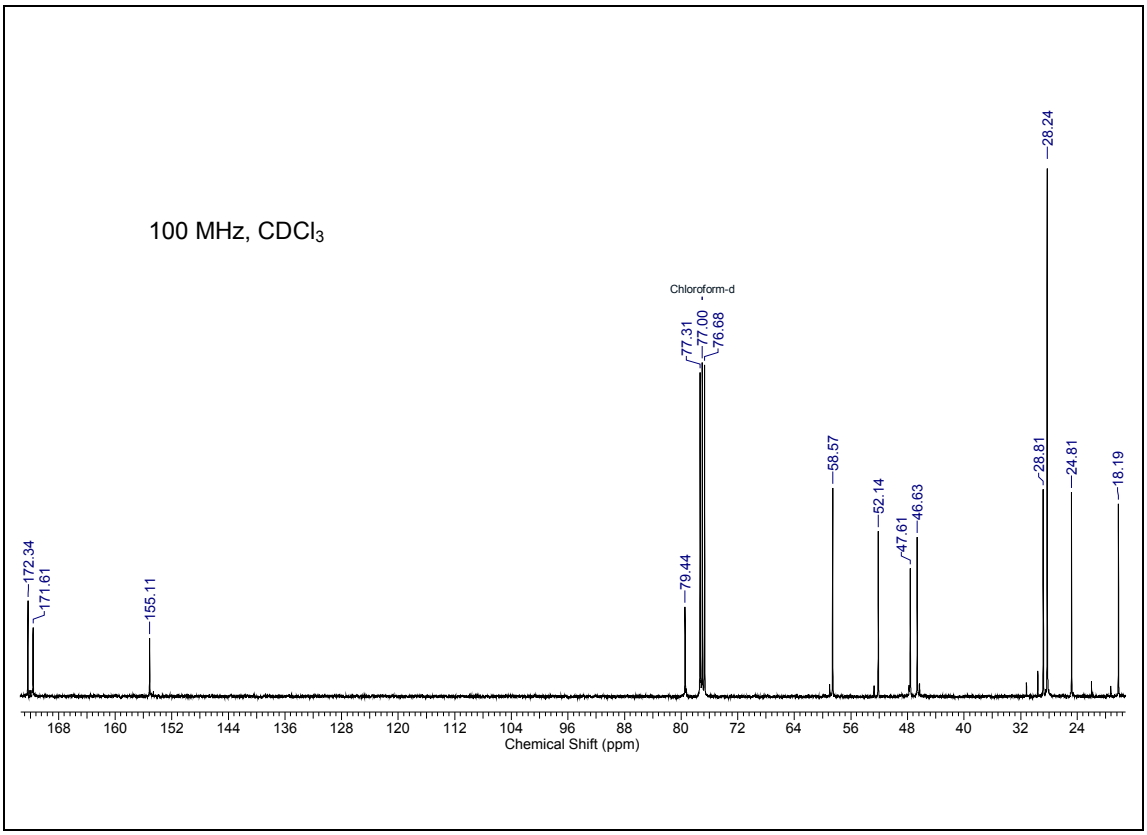
DEPT-^{13}C NMR of compound 14



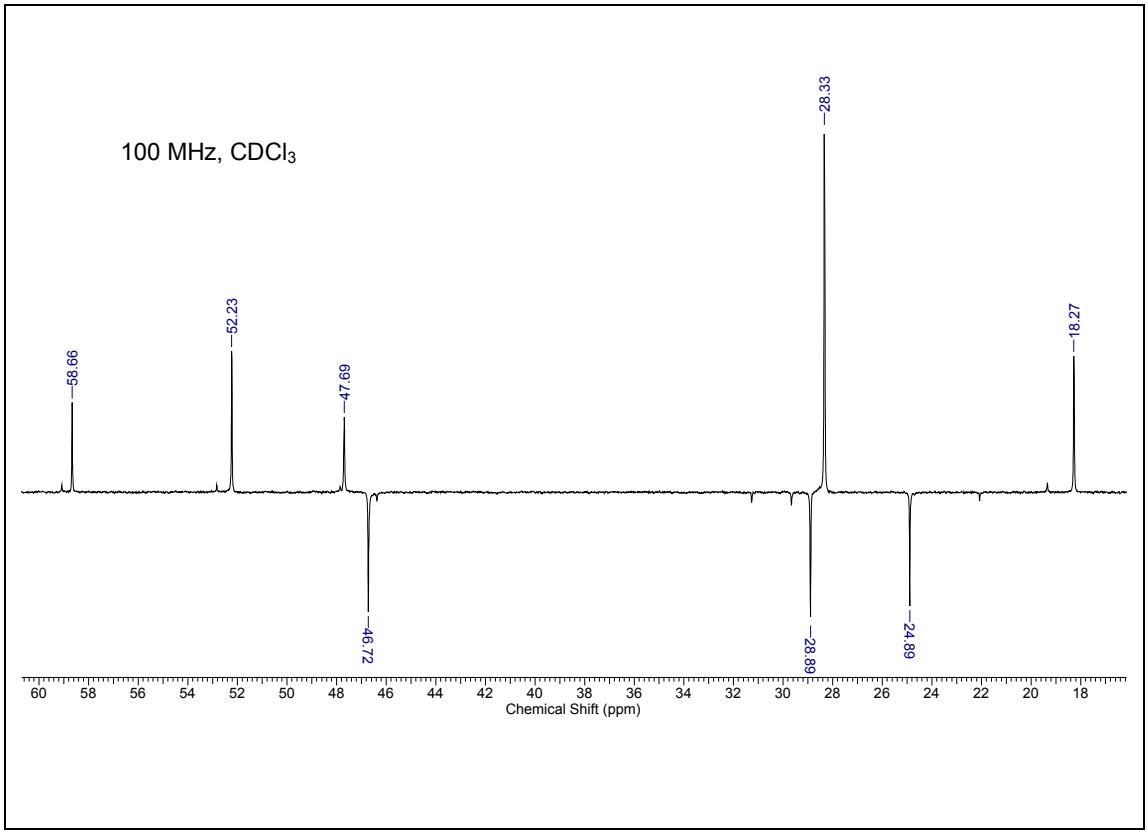
^{1}H NMR of compound 17



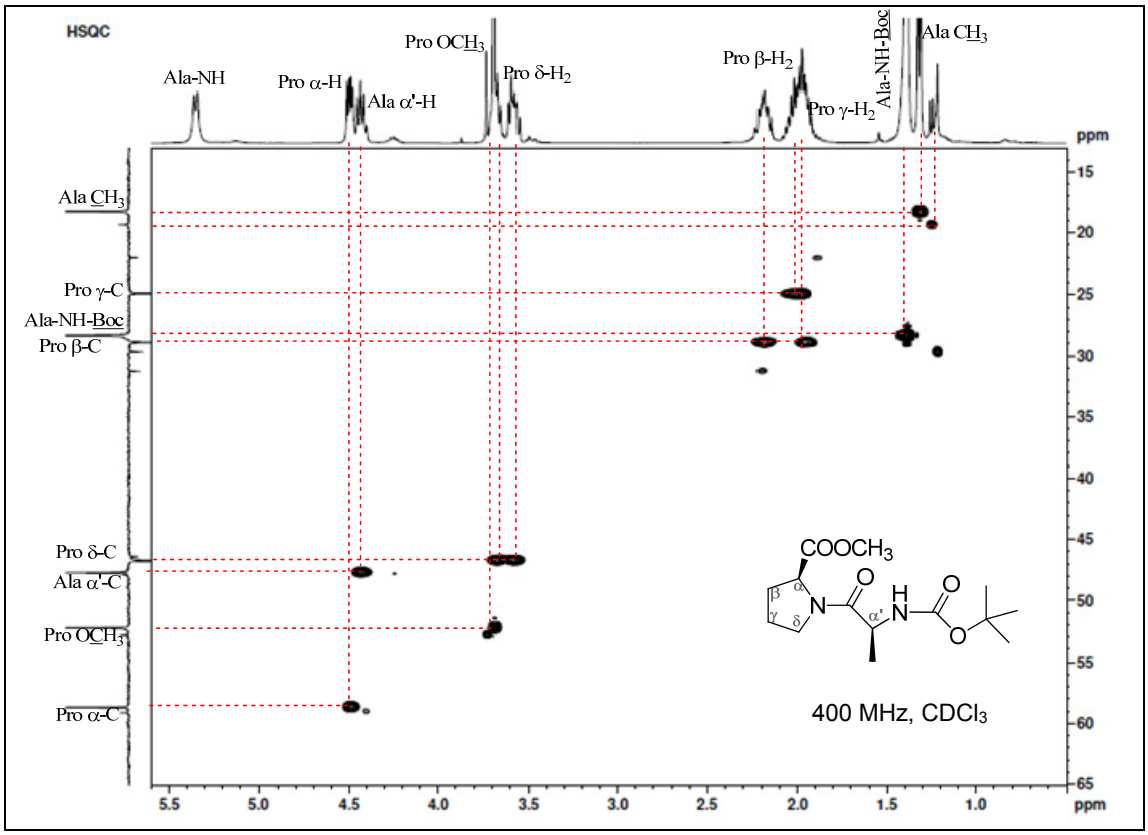
^{13}C NMR of compound 17



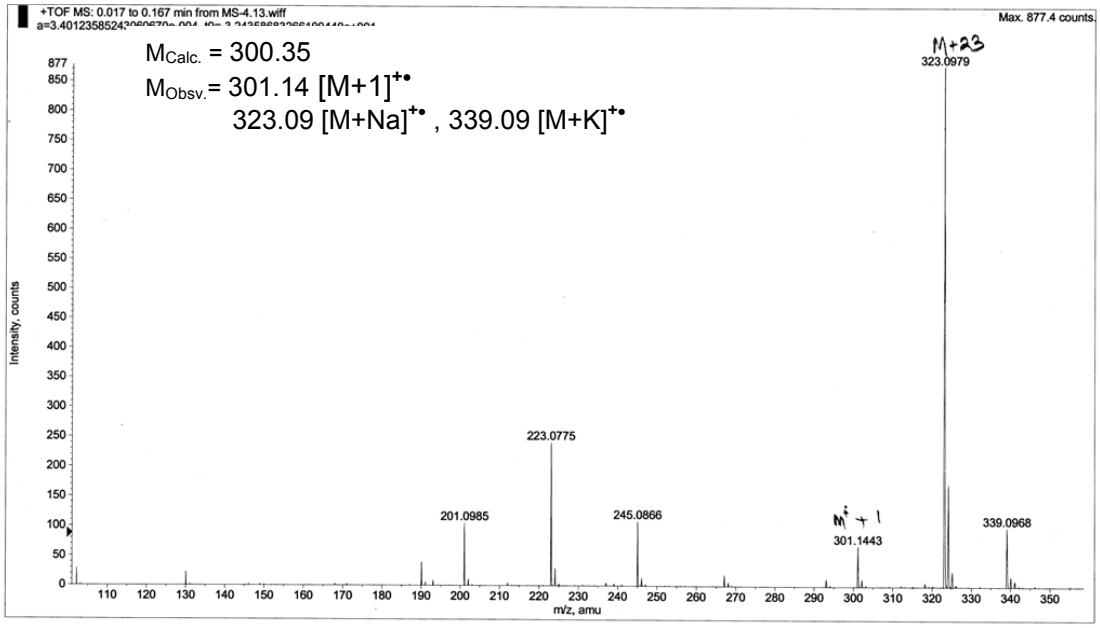
DEPT-^{13}C NMR of compound 17



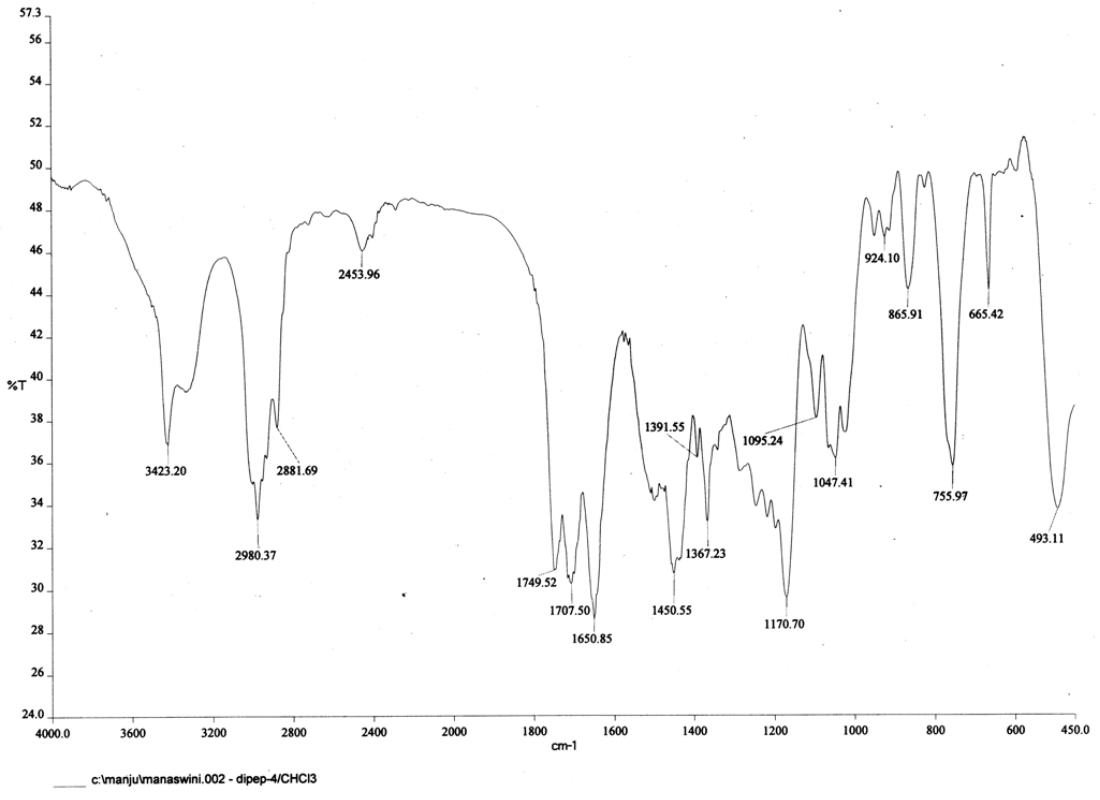
^{1}H-^{13}C HSQC of compound 17



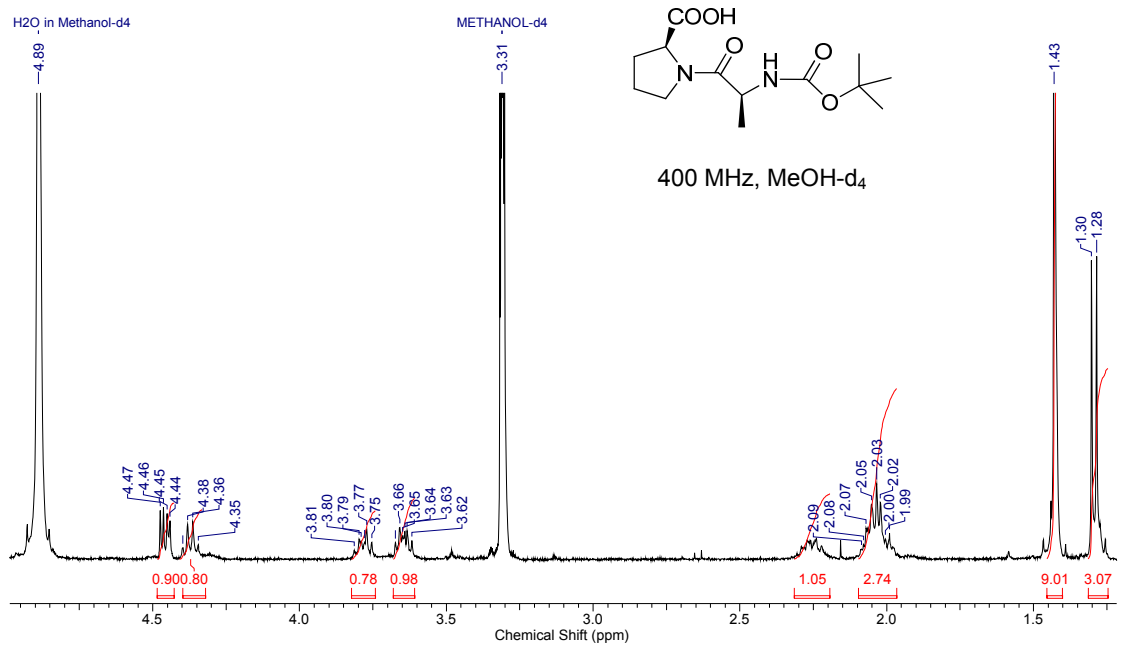
ESI-MS of compound 17



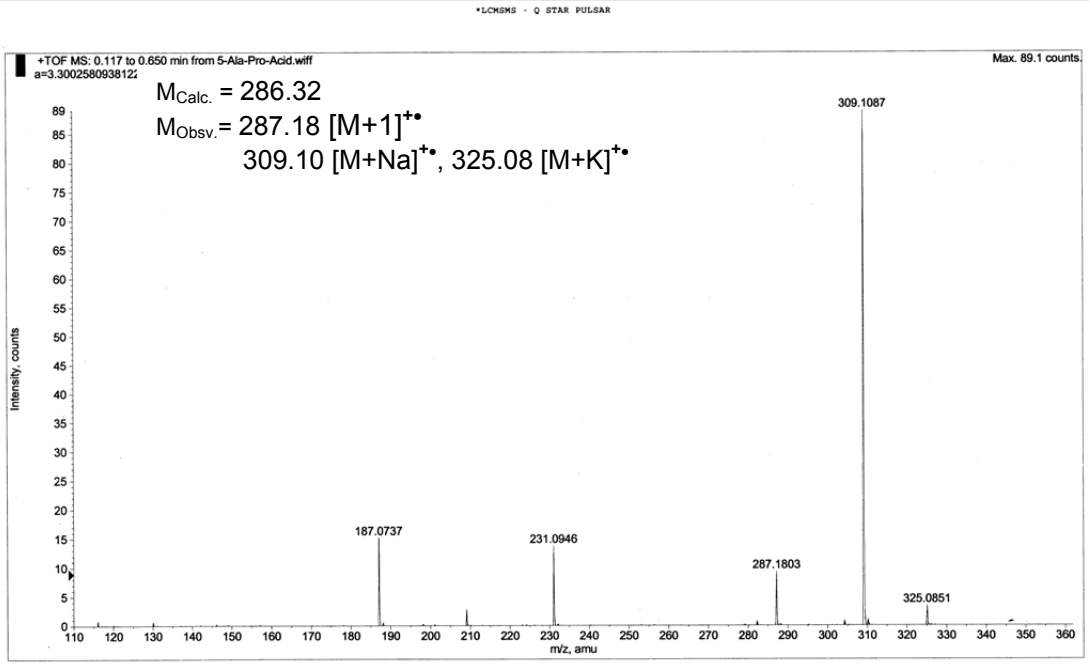
IR Spectra of compound 17



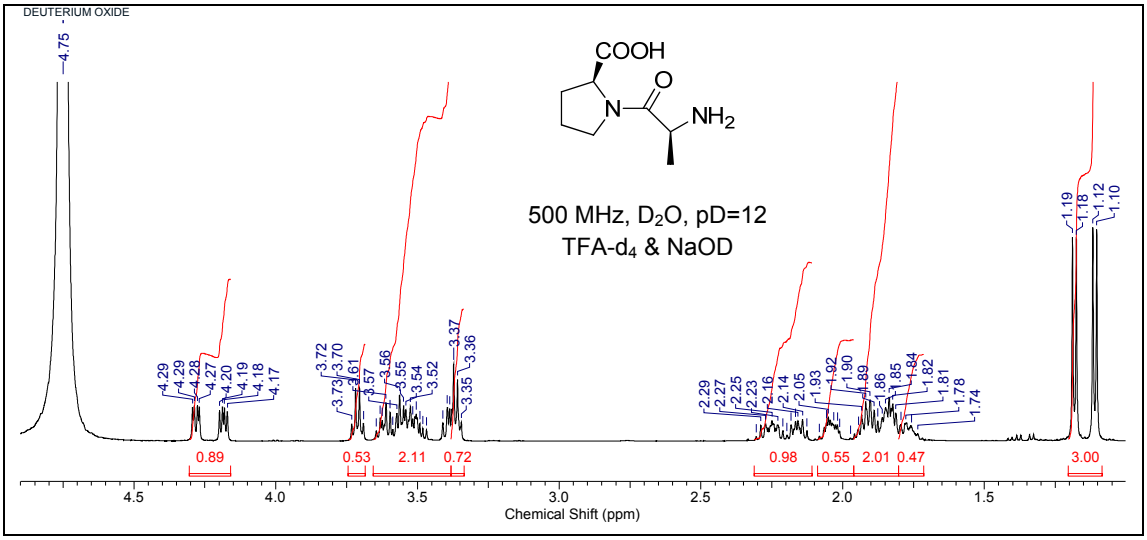
^{1}H NMR of compound 18



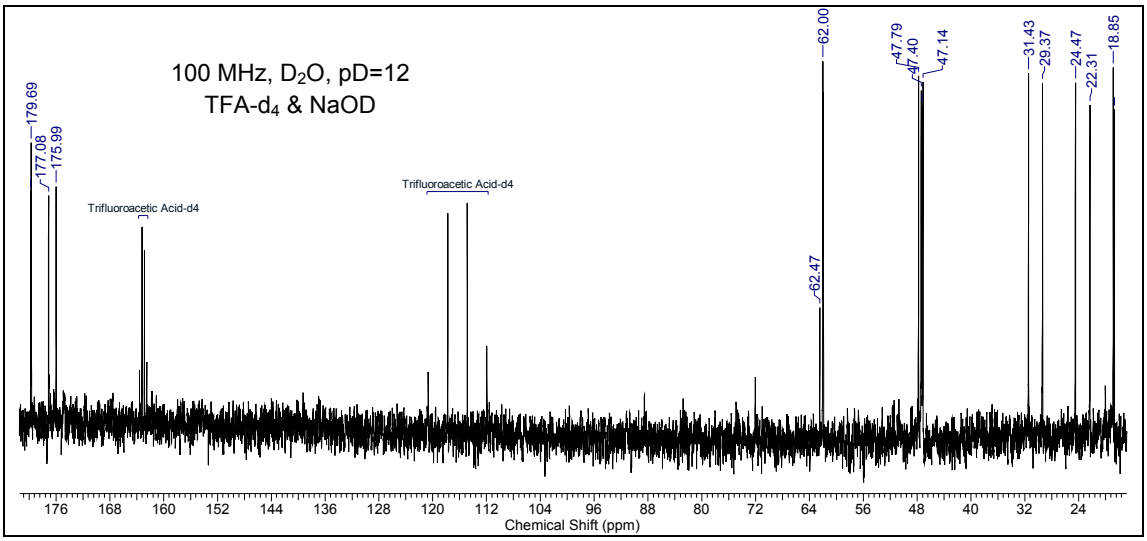
ESI-MS of compound 18



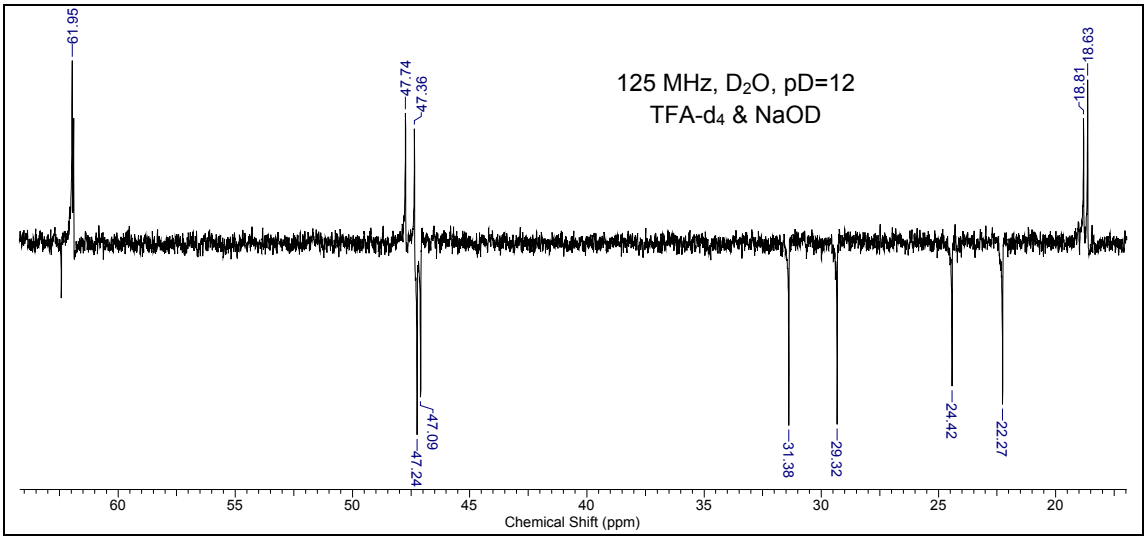
^{1}H NMR of compound 19



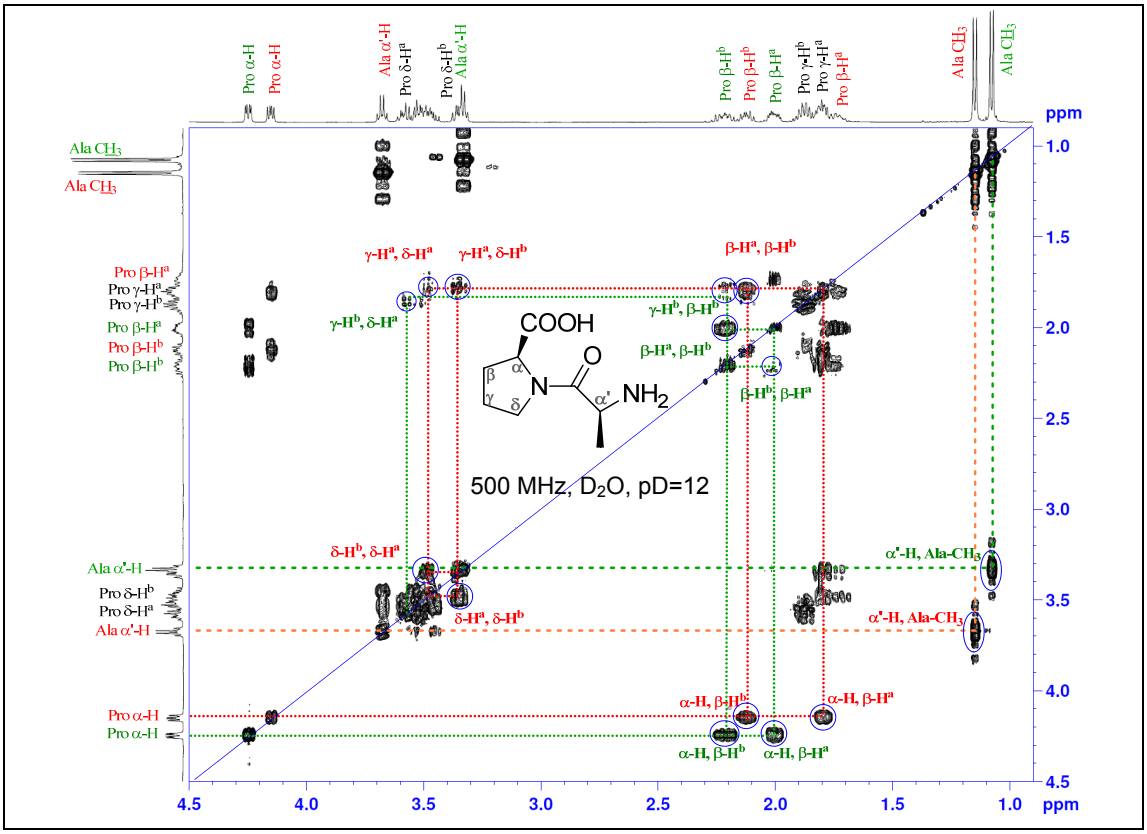
^{13}C NMR of compound 19



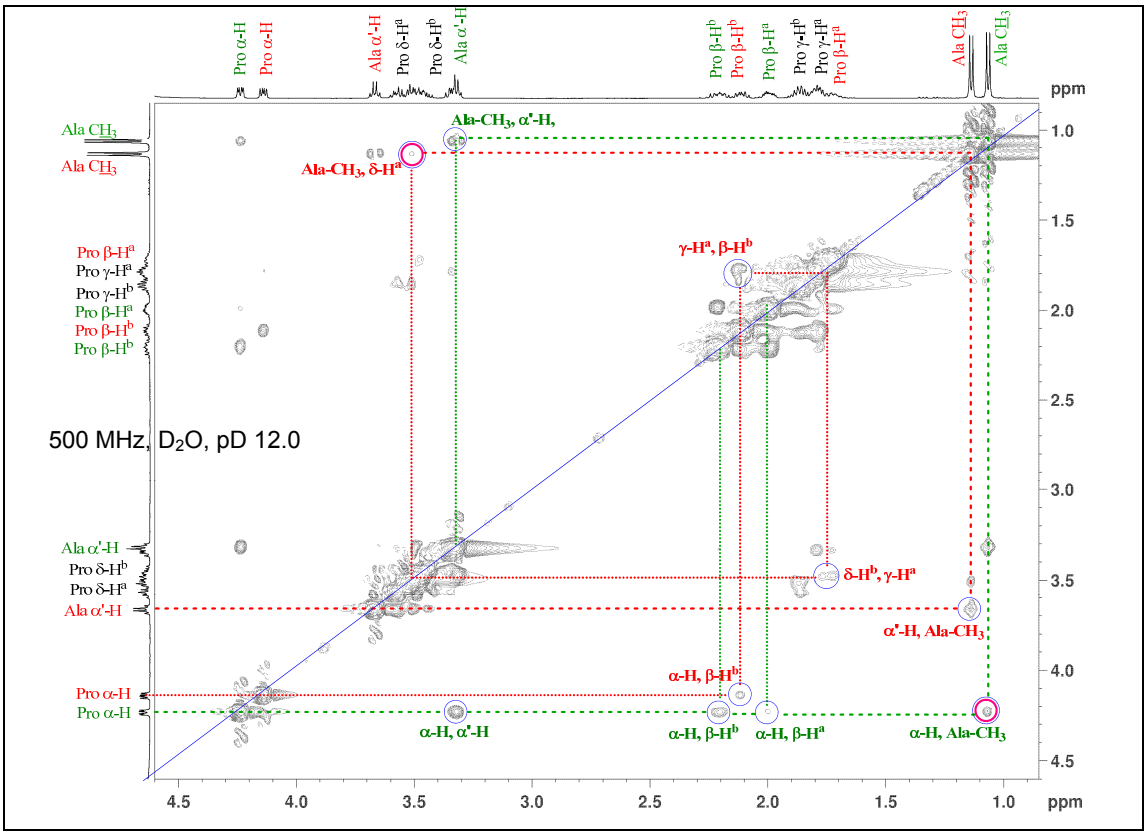
DEPT-^{13}C NMR of compound 19



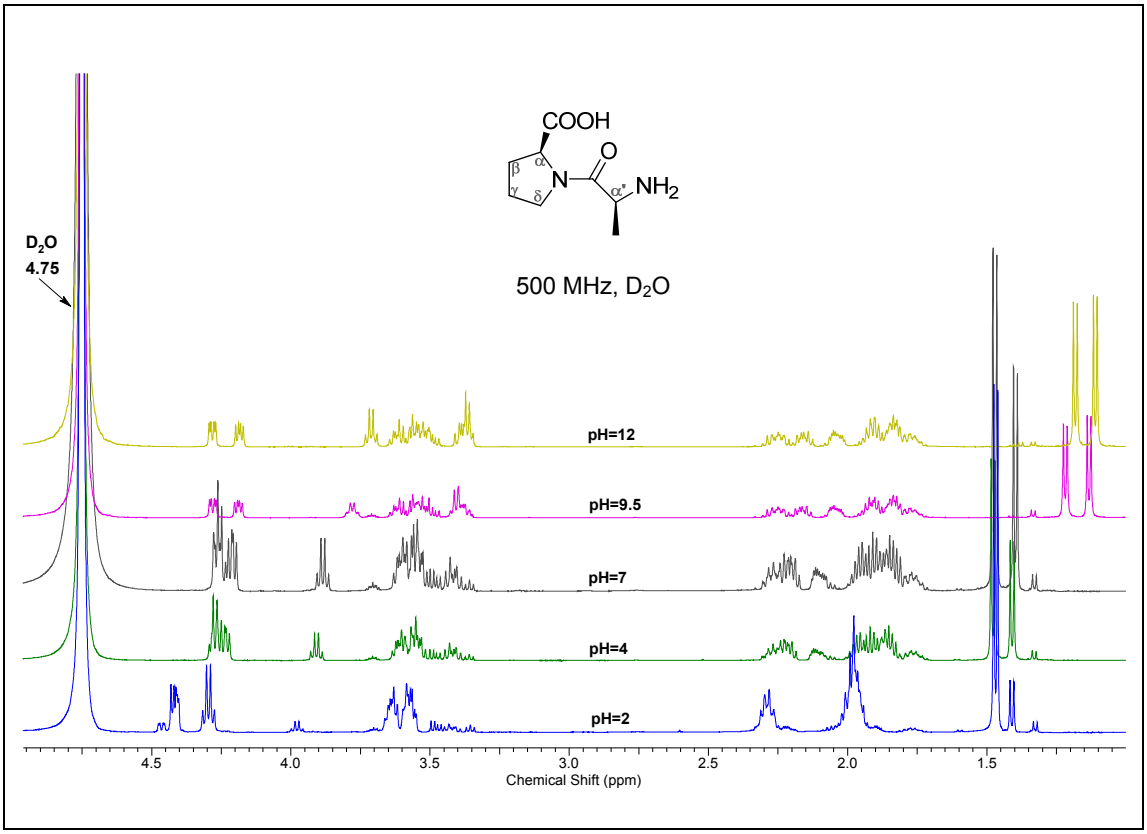
^{1}H-^{1}H COSY of compound 19



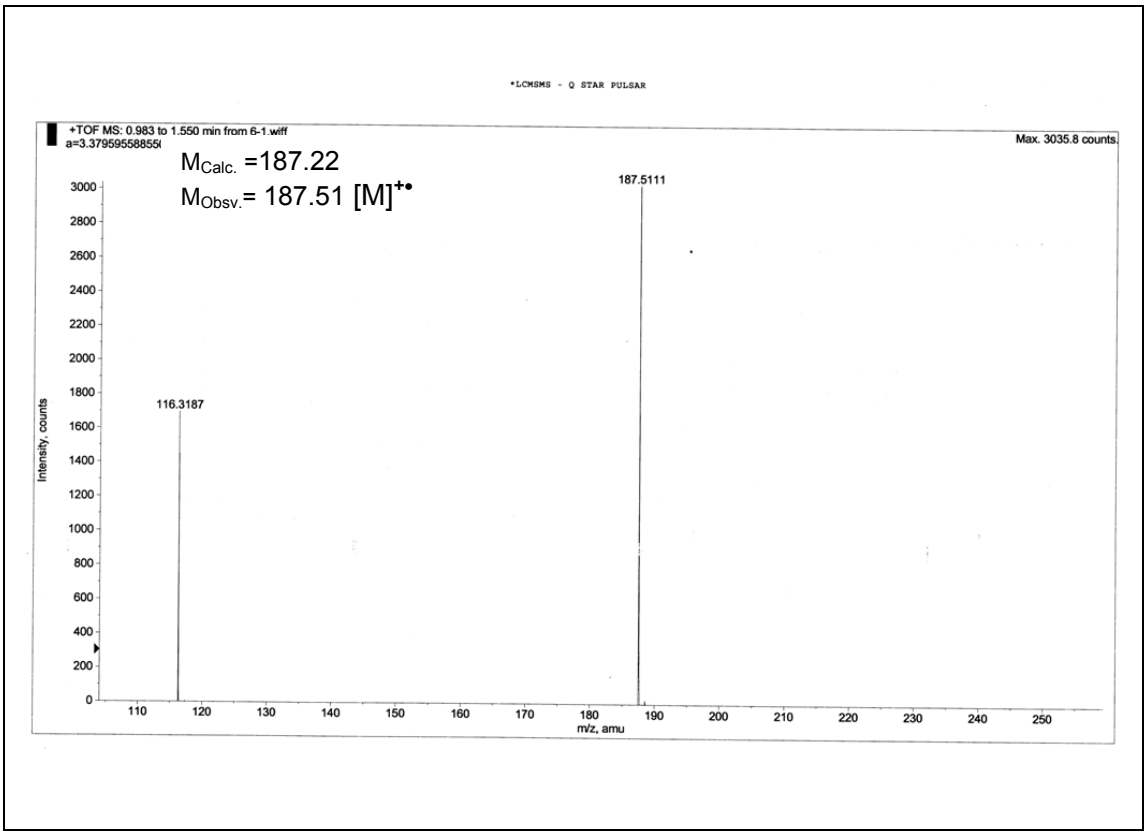
^{1}H-^{1}H NOESY of compound 19



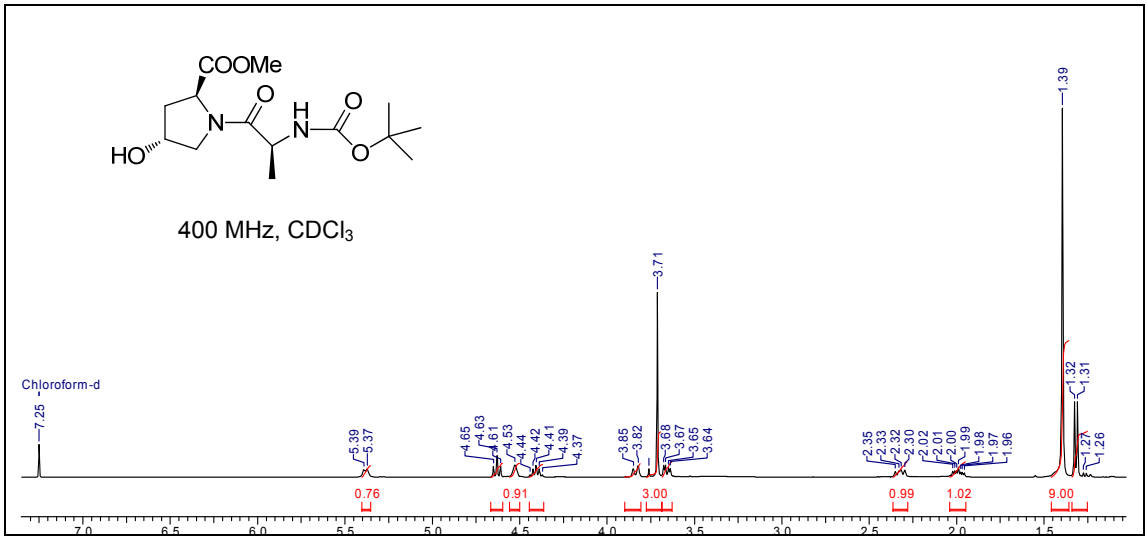
Variable pH ^{1}H NMR of compound 19



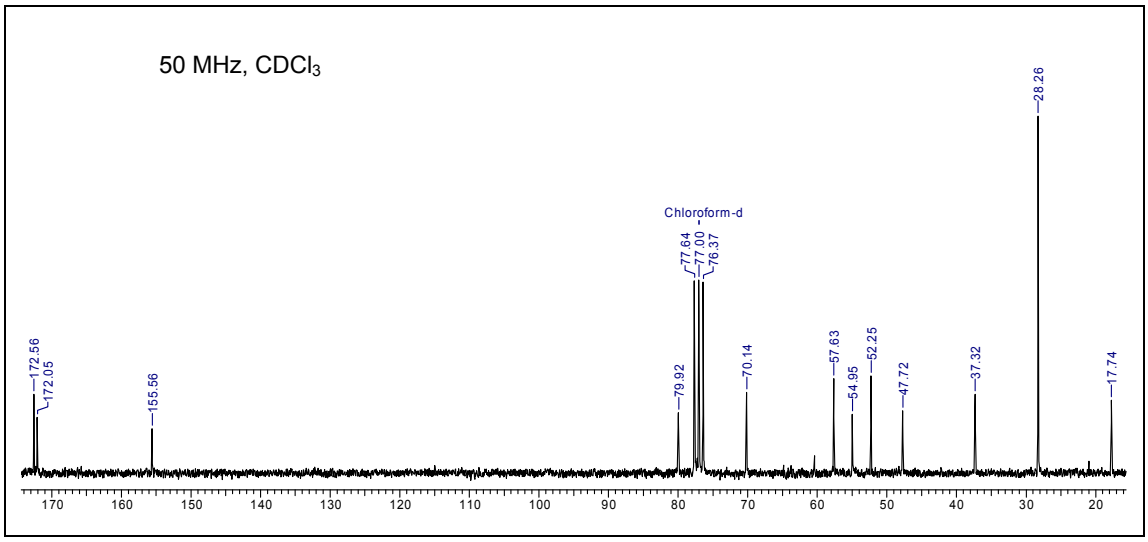
ESI-MS of compound 19



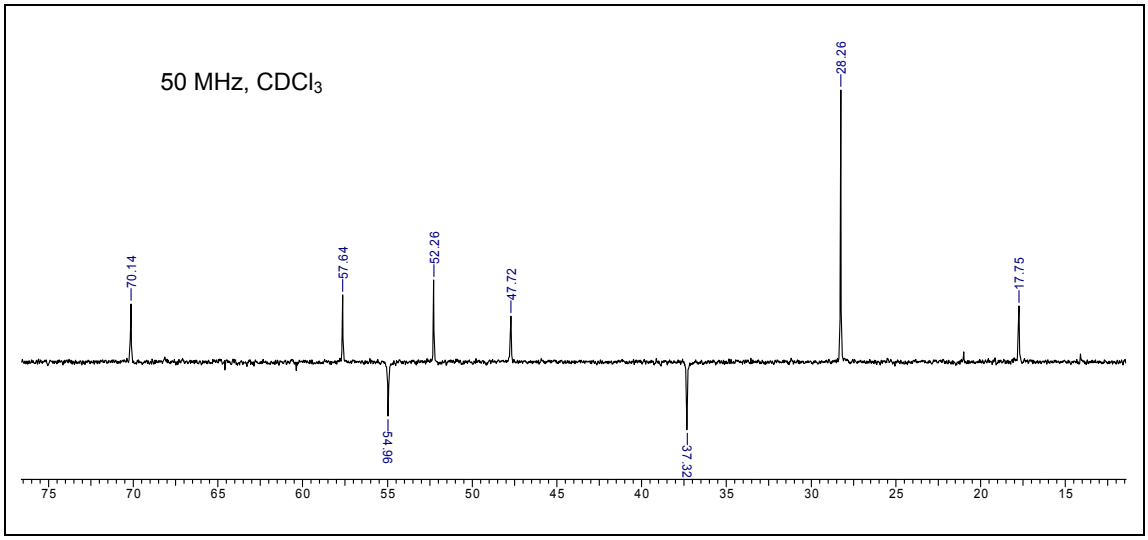
^{1}H NMR of compound 21



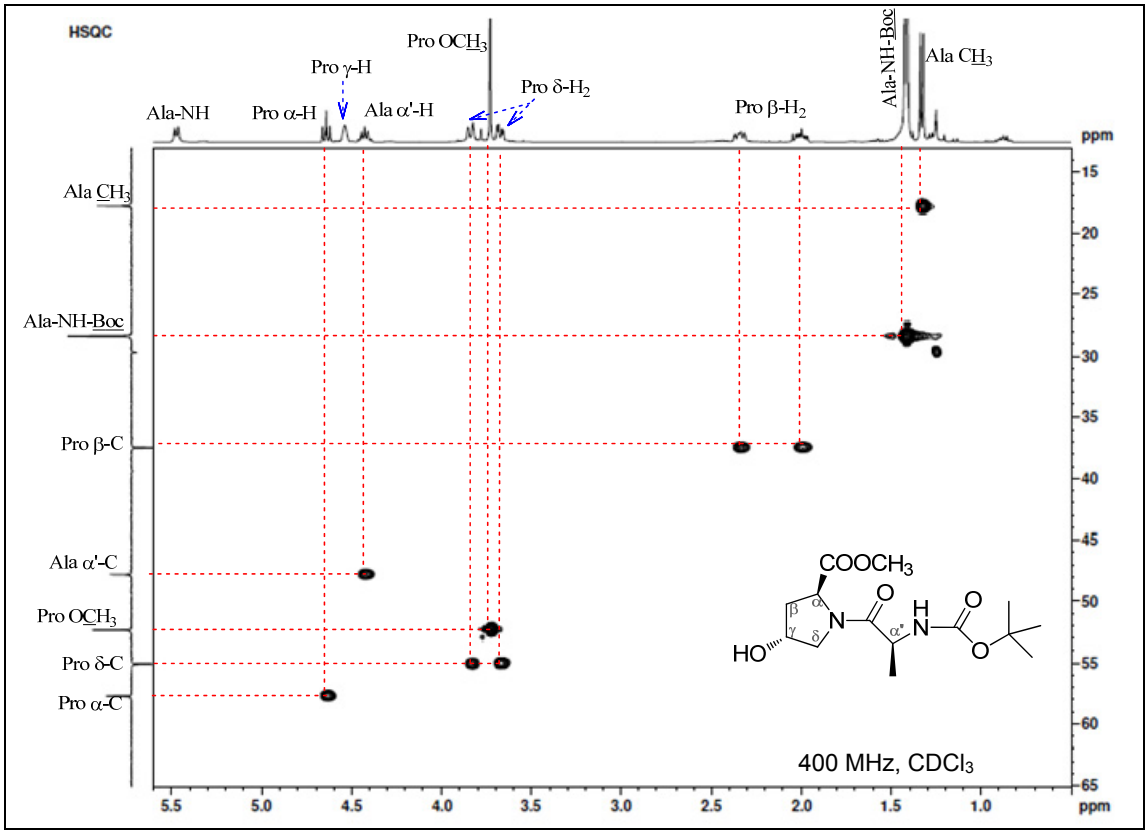
^{13}C NMR of compound 21



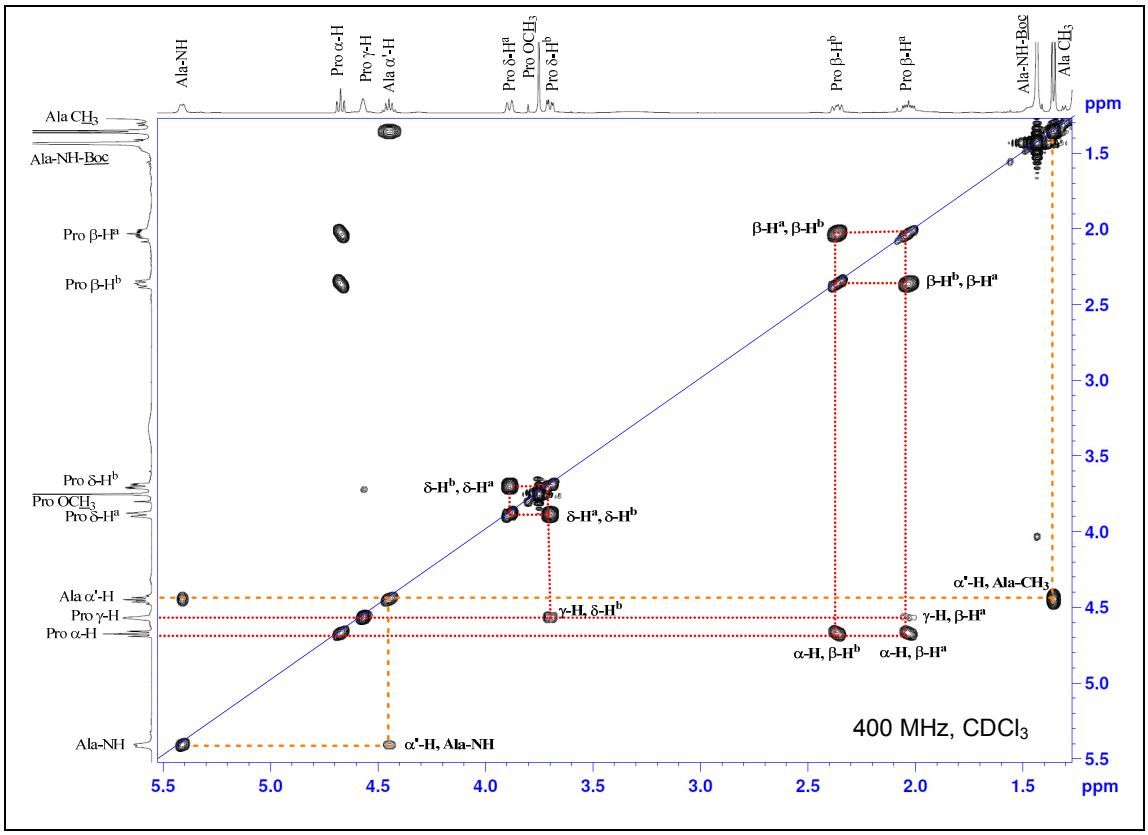
DEPT-^{13}C NMR of compound 21



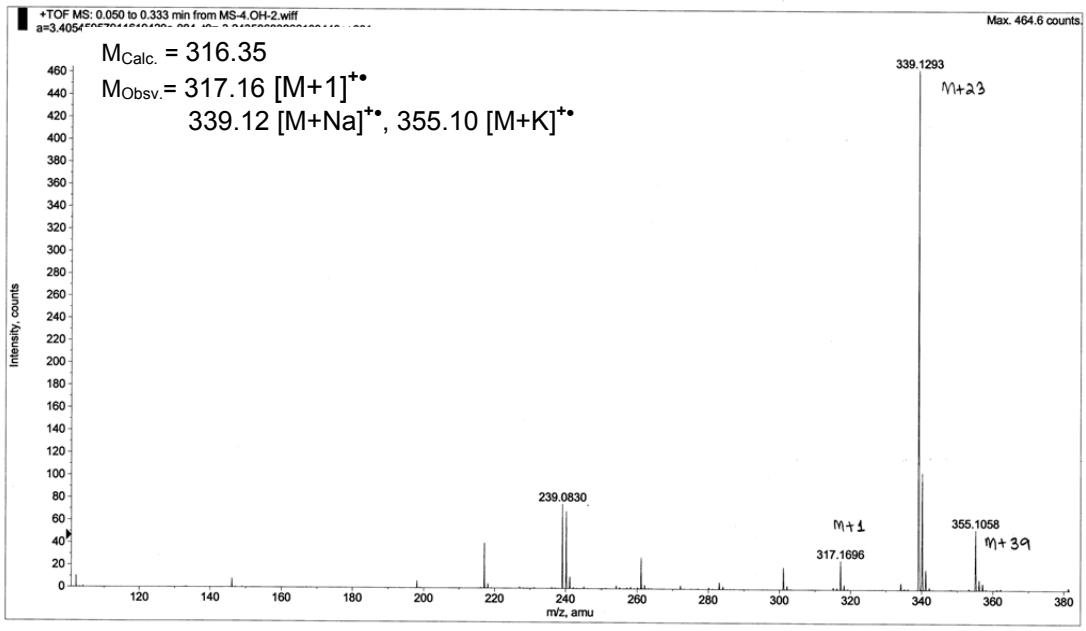
^{1}H-^{13}C HSQC of compound 21



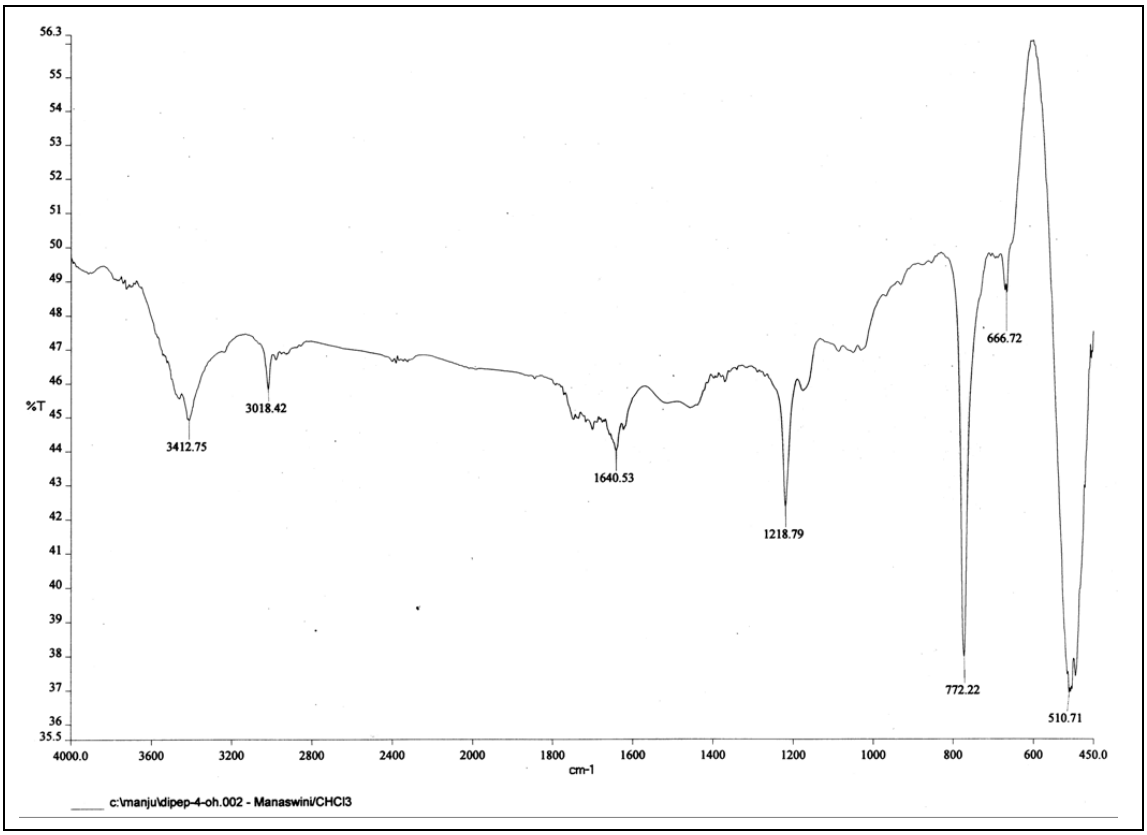
^{1}H-^{1}H COSY of compound 21



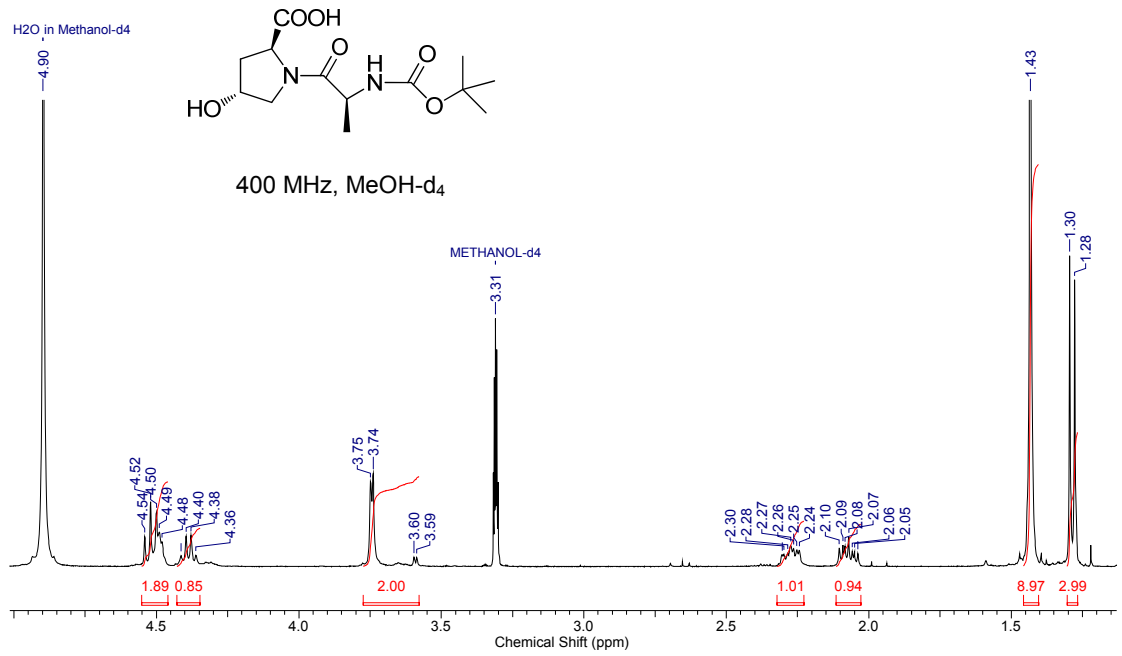
ESI-MS of compound 21



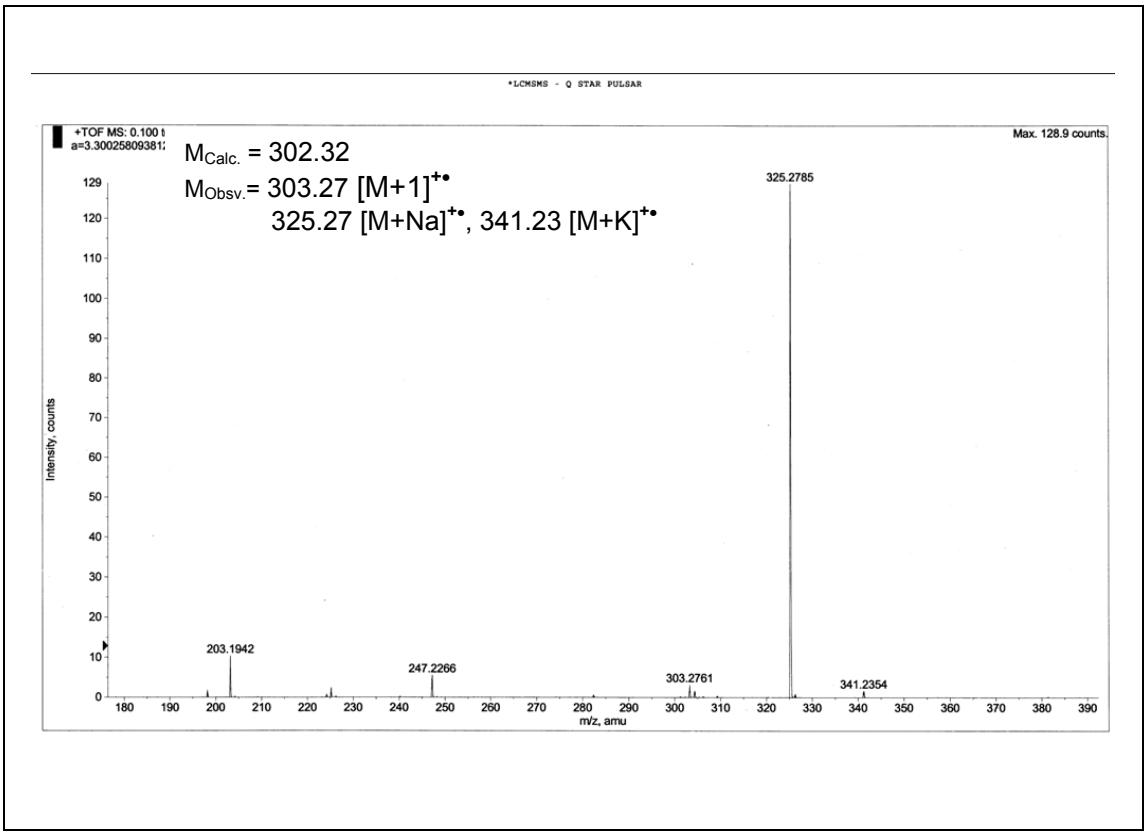
IR Spectra of compound 21



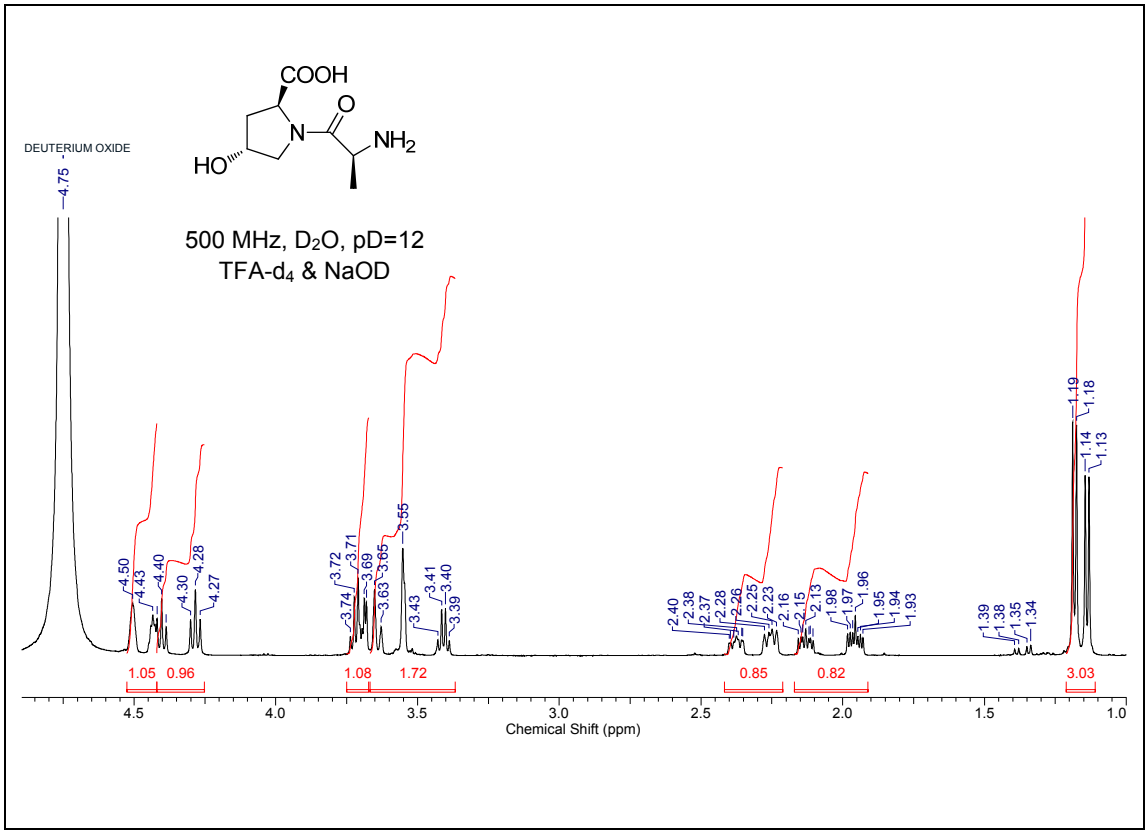
^{1}H NMR of compound 22



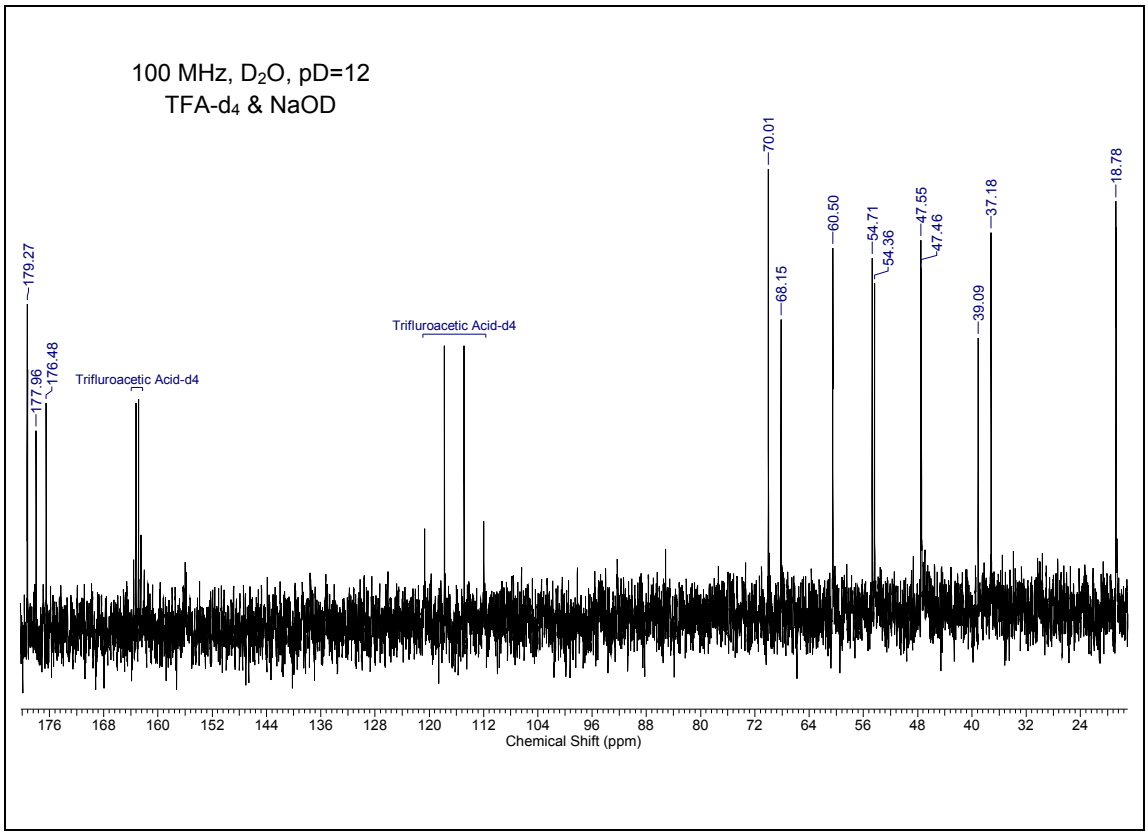
ESI-MS of compound 22



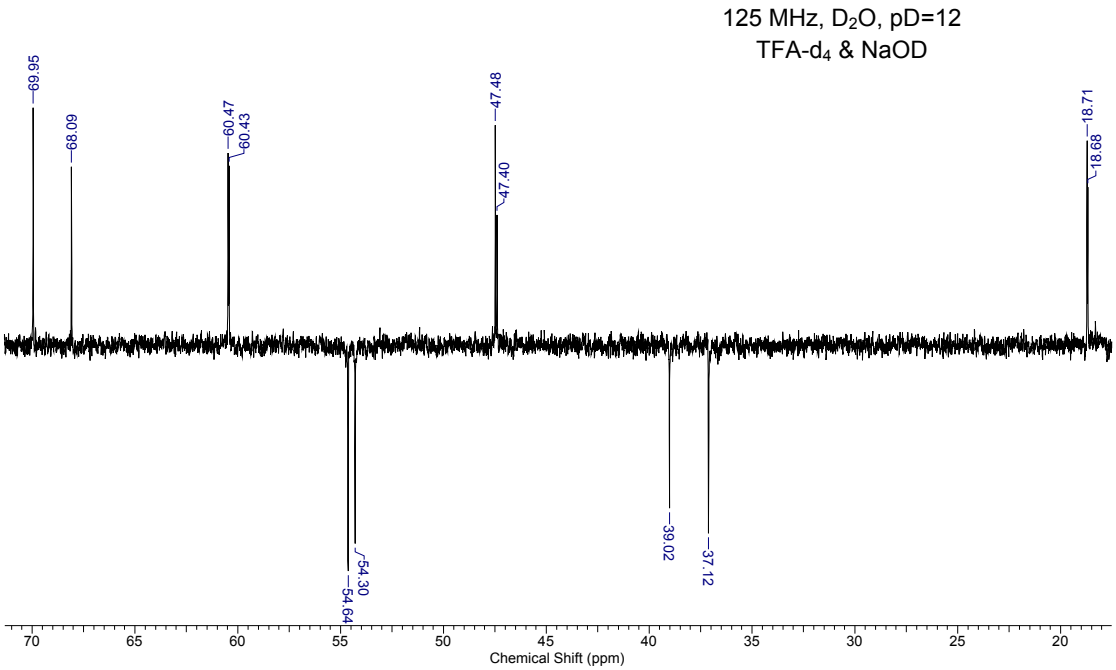
^{1}H NMR of compound 23



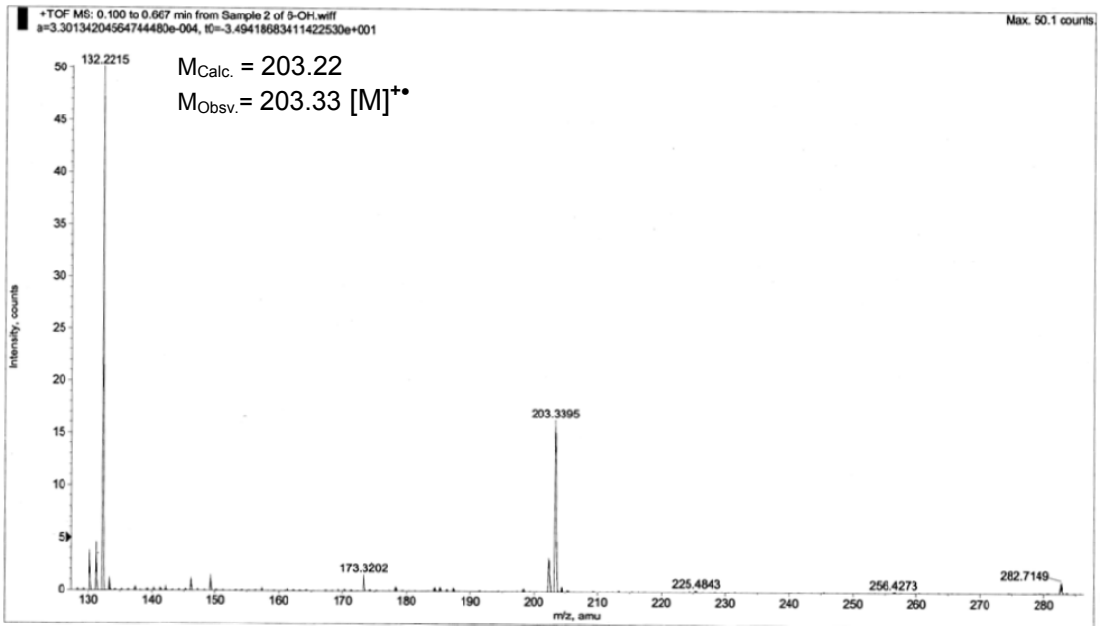
^{13}C NMR of compound 23



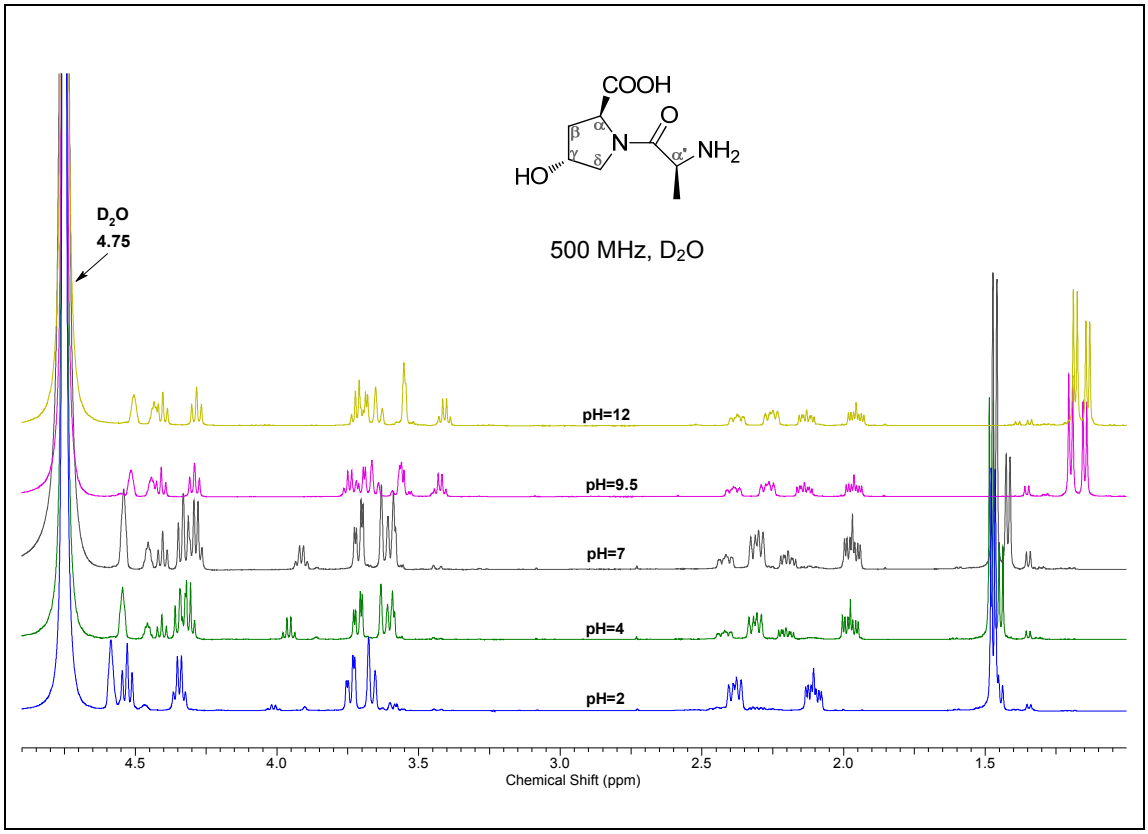
DEPT-^{13}C NMR of compound 23



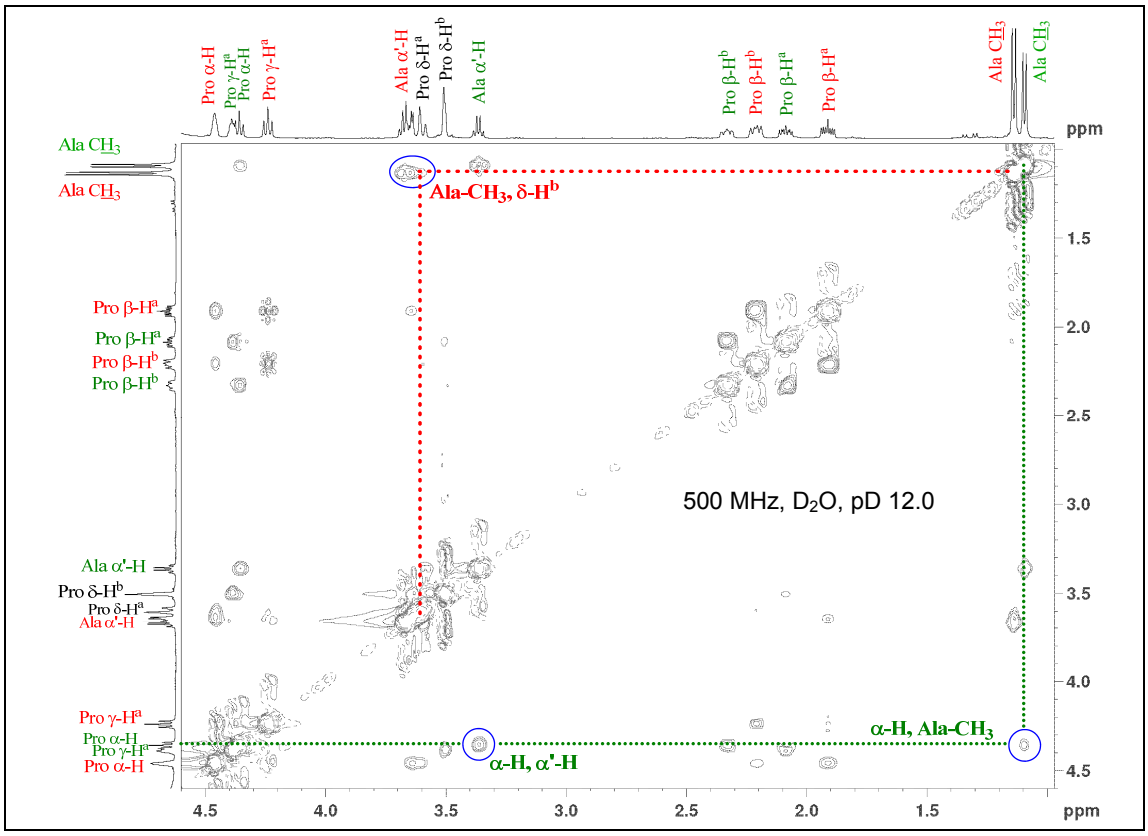
ESI-MS of compound 23



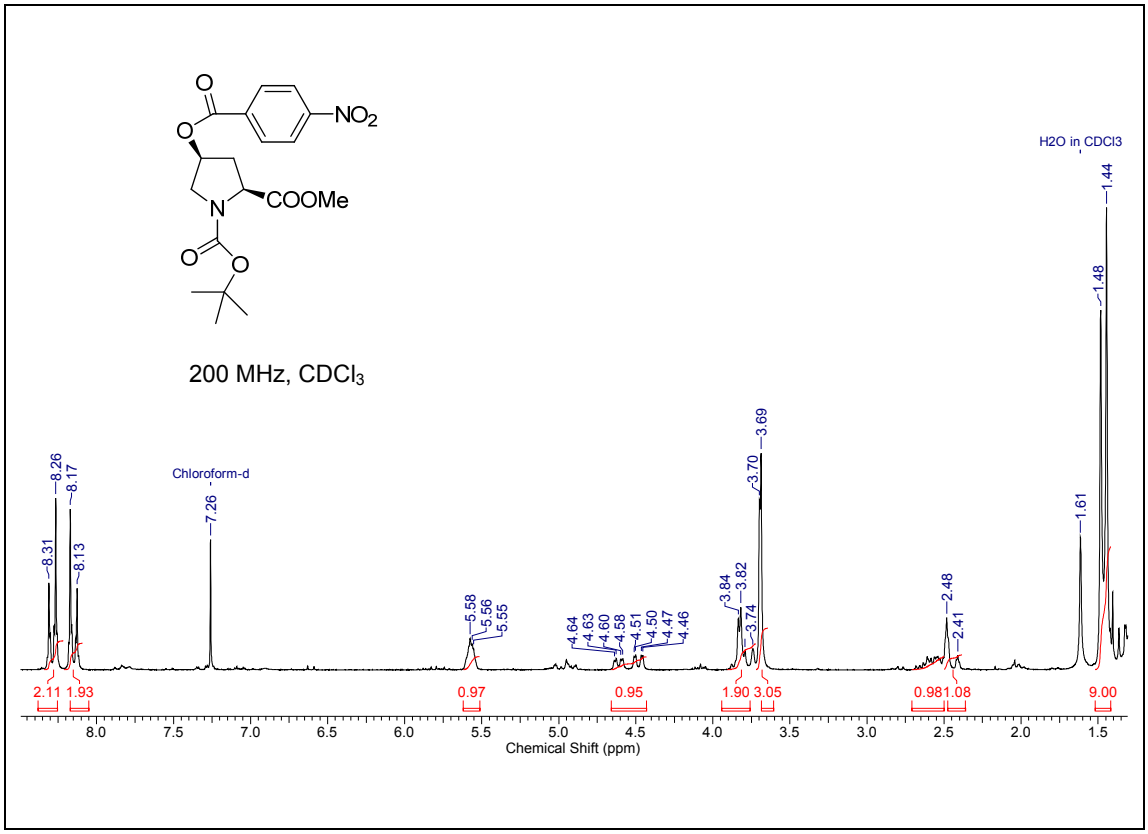
Variable pH ^{1}H NMR of compound 23



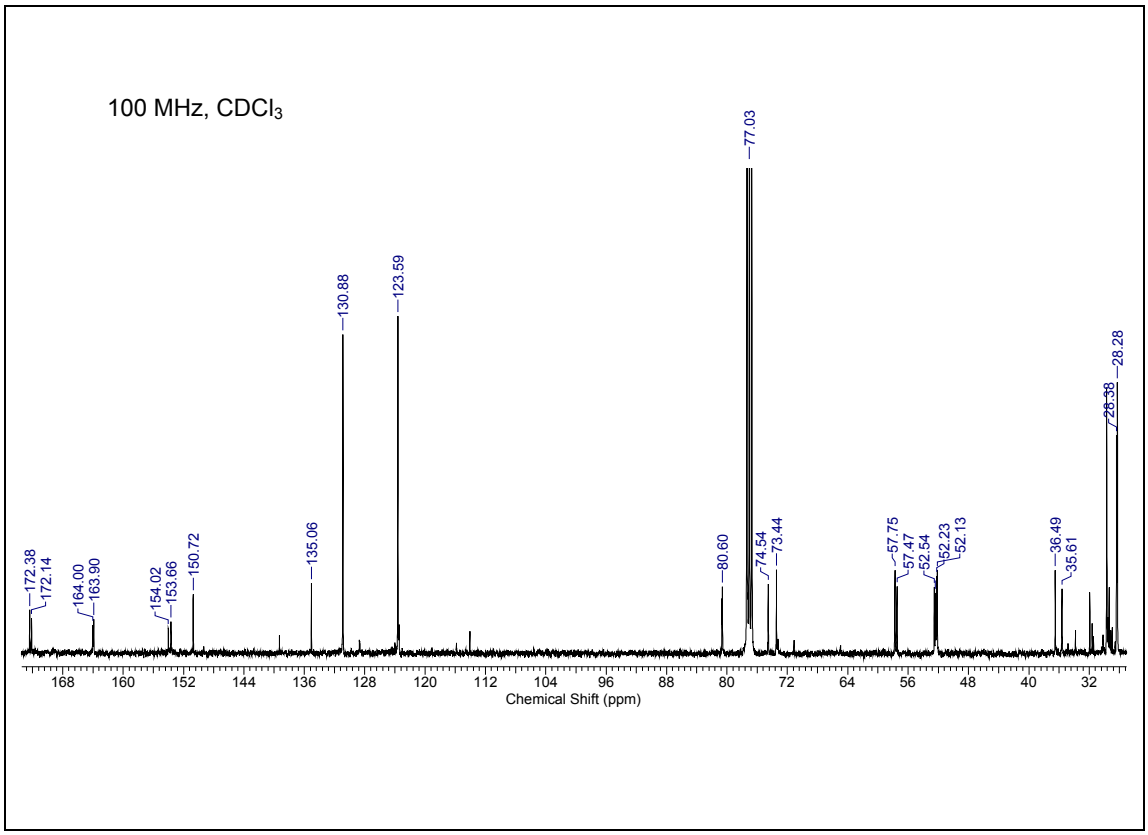
^{1}H-^{1}H NOESY of compound 23



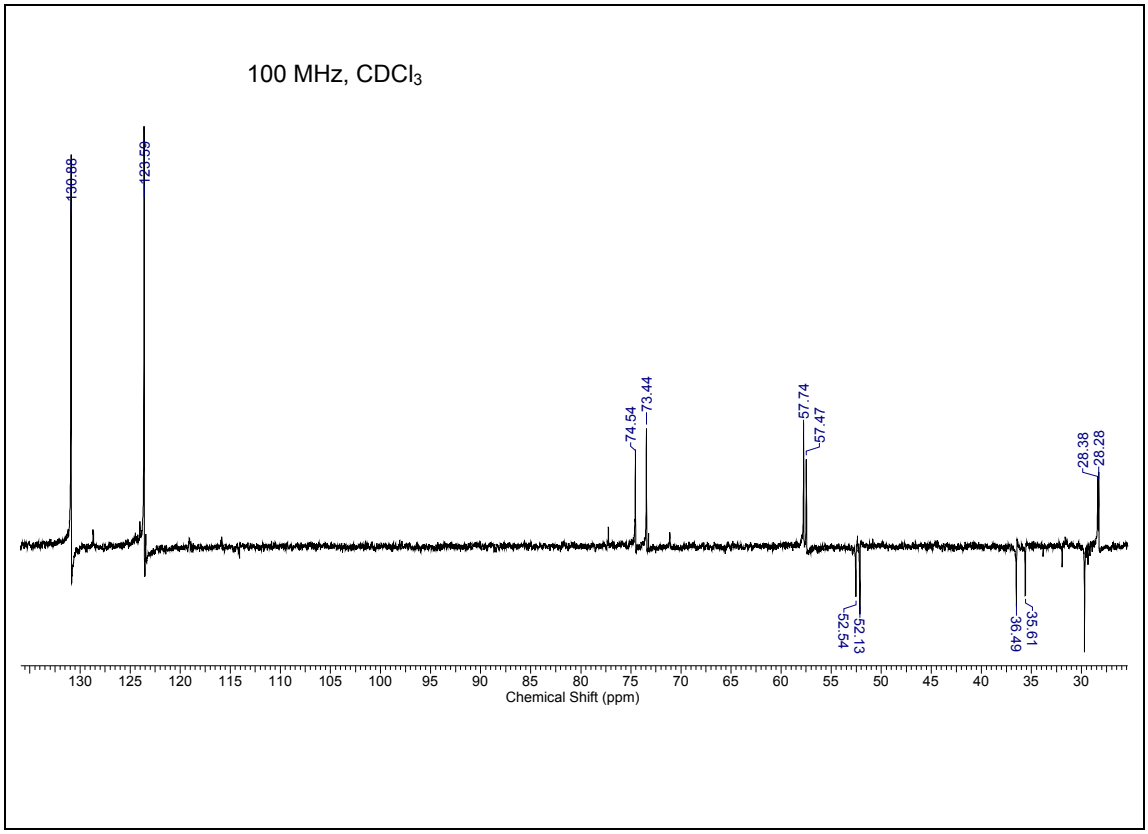
^{1}H NMR of compound 24



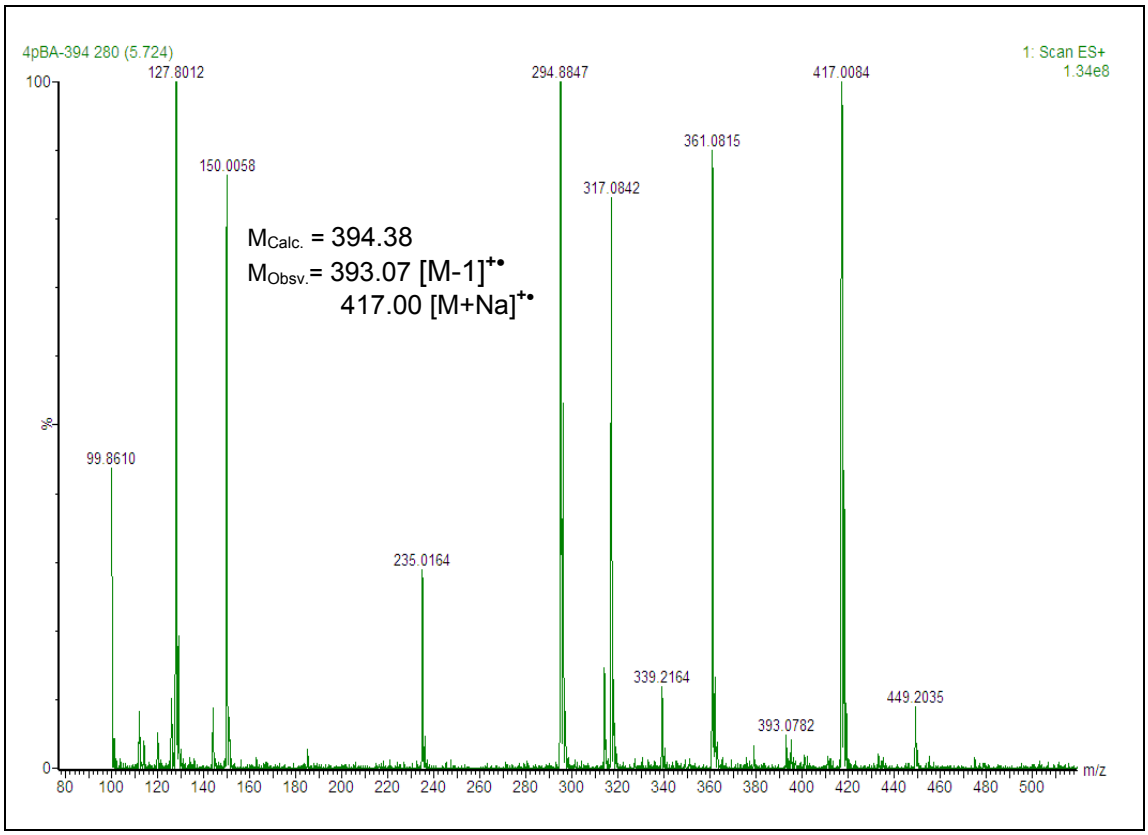
^{13}C NMR of compound 24



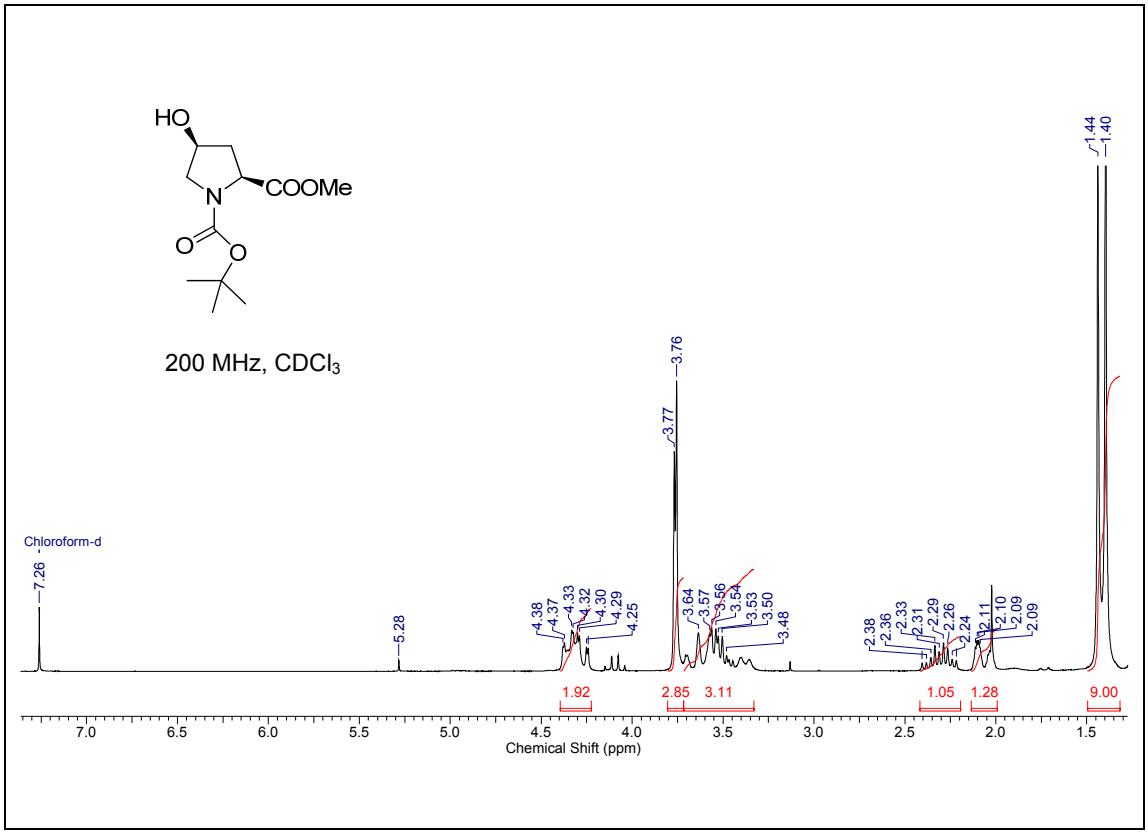
DEPT-^{13}C NMR of compound 24



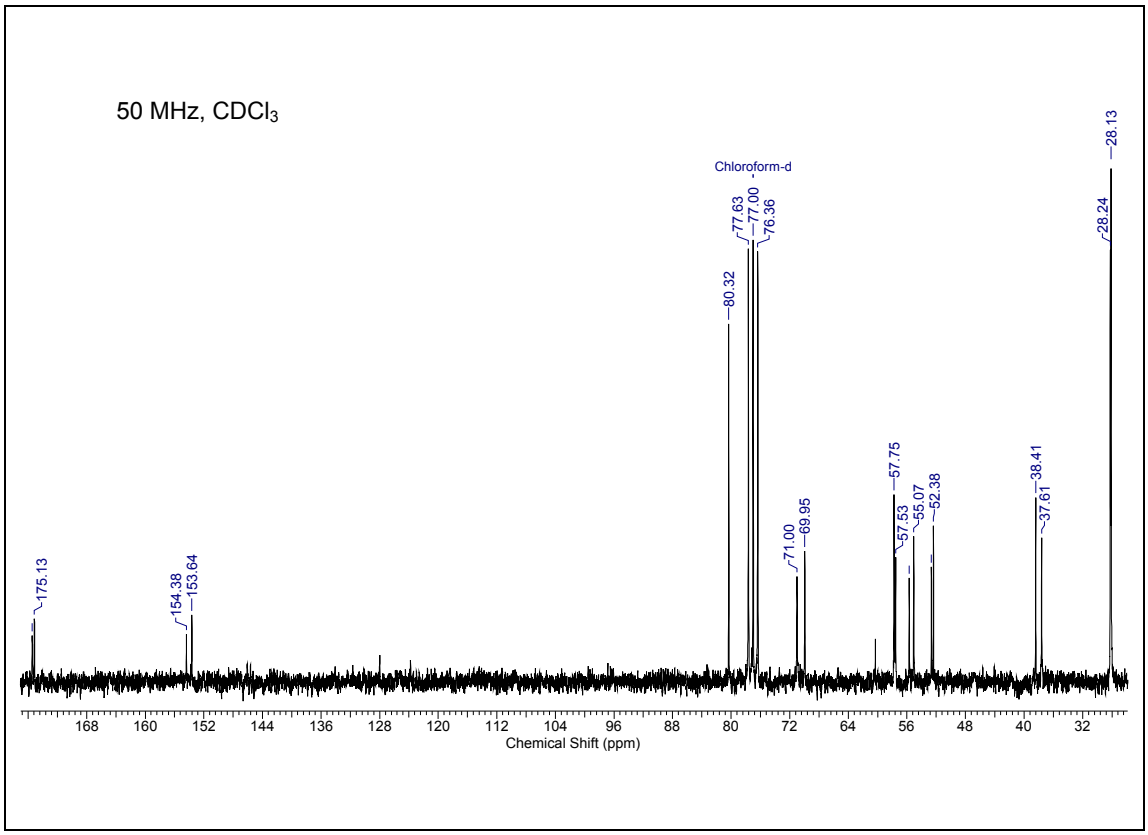
ESI-MS of compound 24



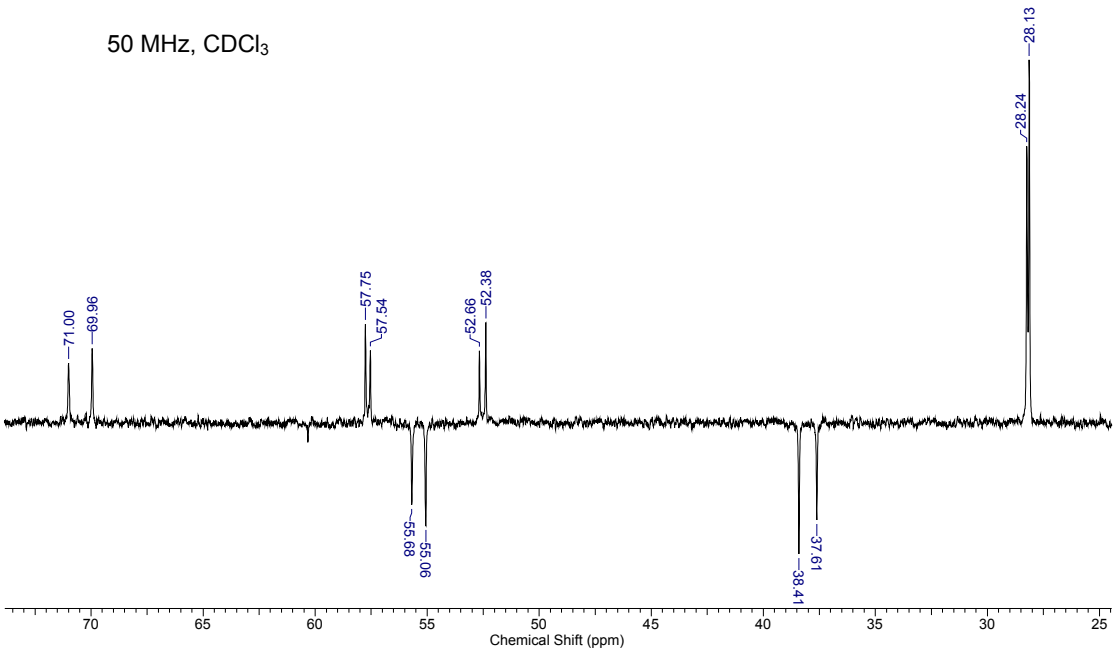
^{1}H NMR of compound 25



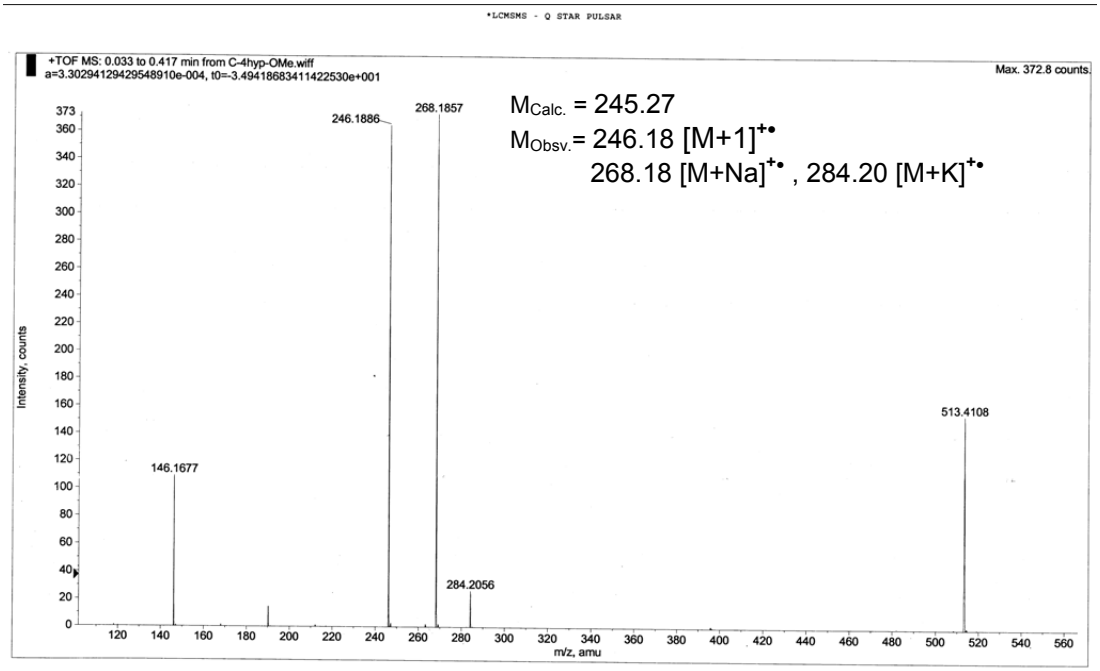
^{13}C NMR of compound 25



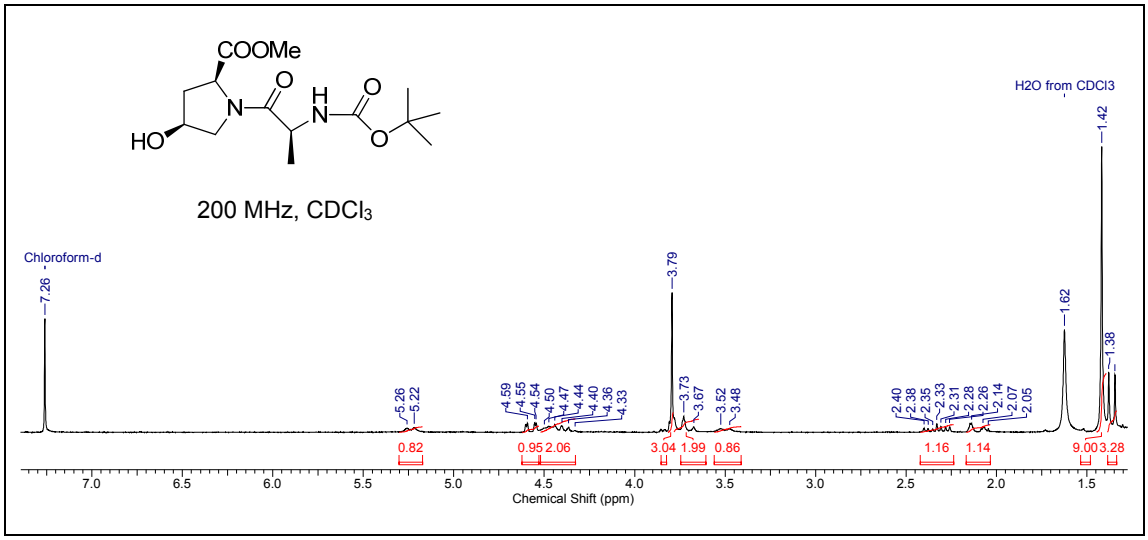
DEPT-^{13}C NMR of compound 25



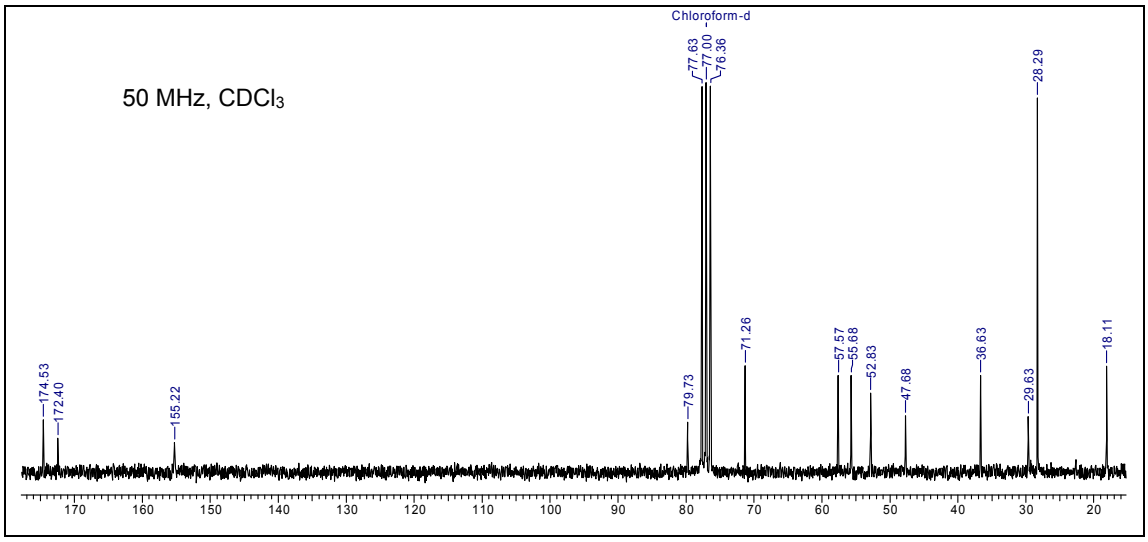
ESI-MS of compound 25



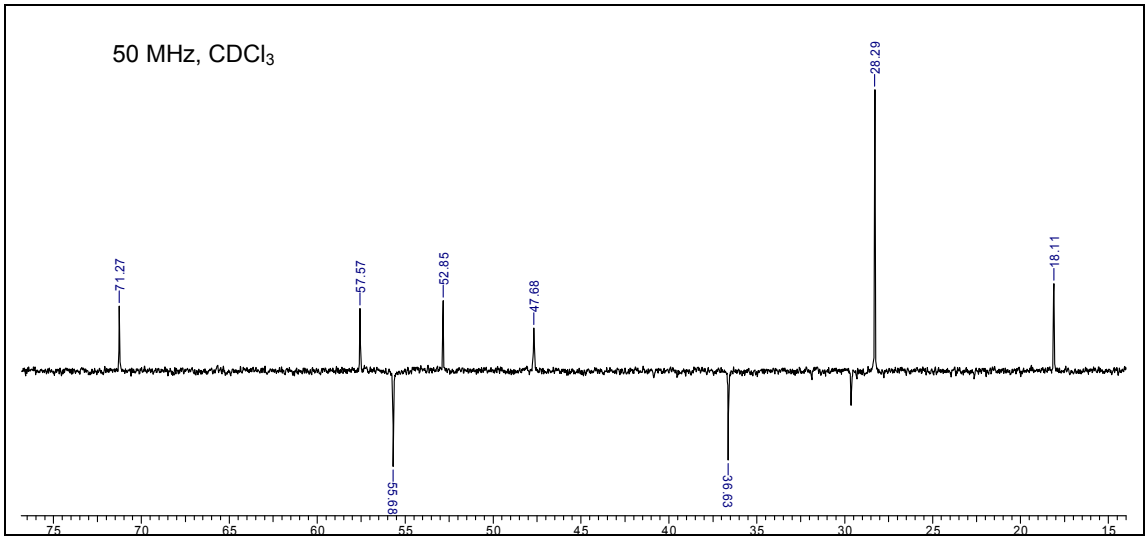
^{1}H NMR of compound 26



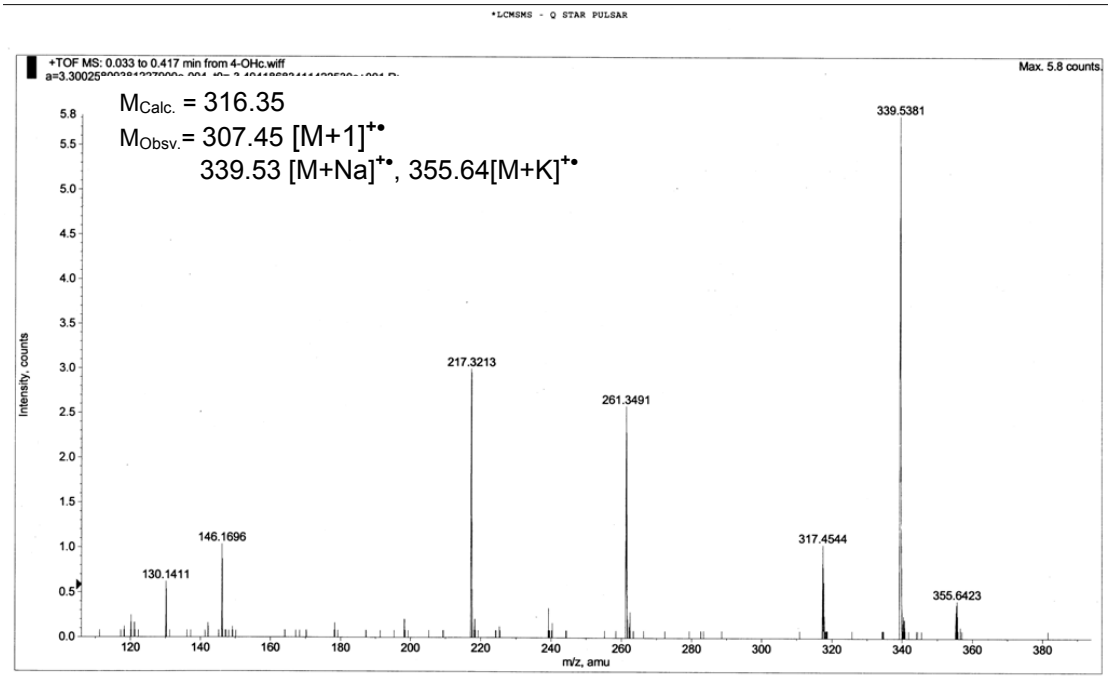
^{13}C NMR of compound 26



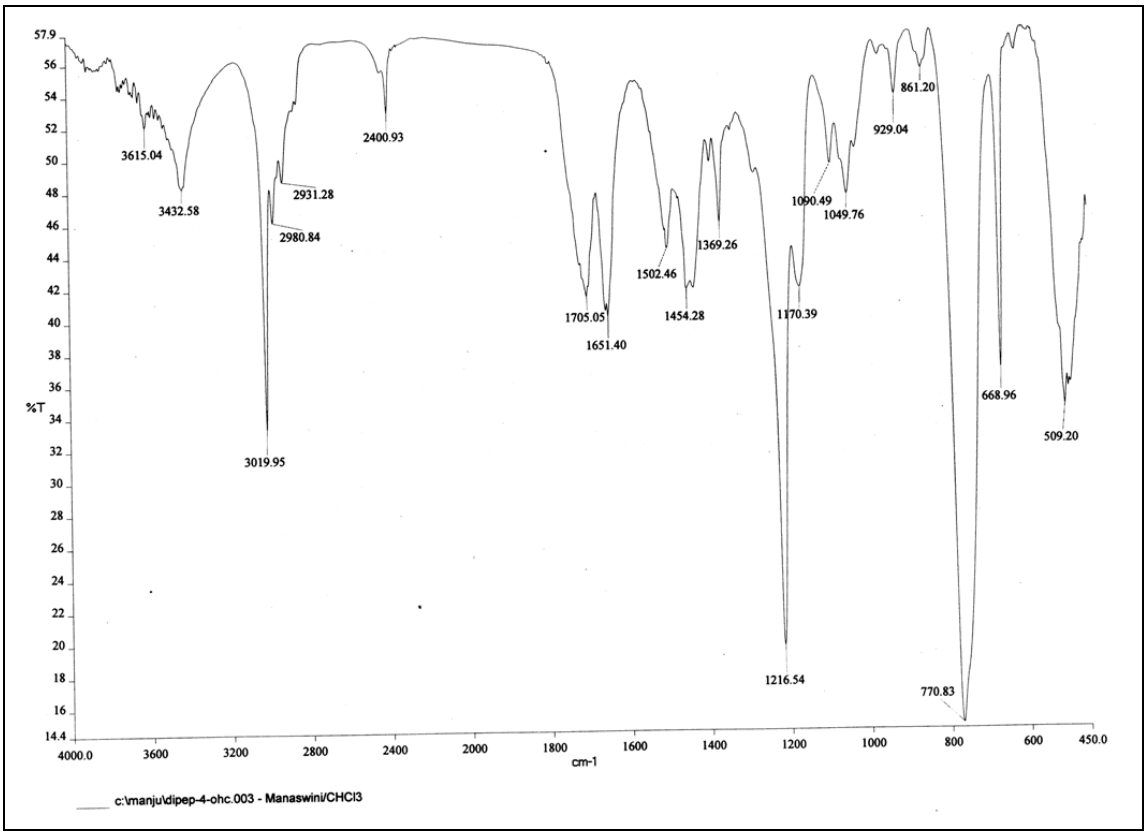
DEPT-^{13}C NMR of compound 26



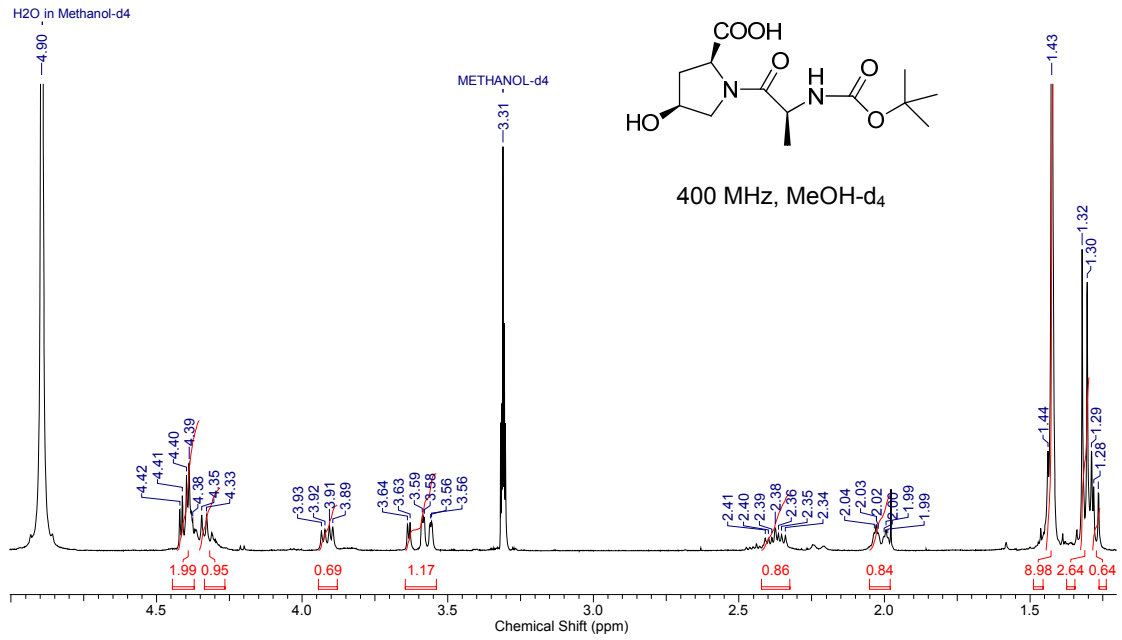
ESI-MS of compound 26



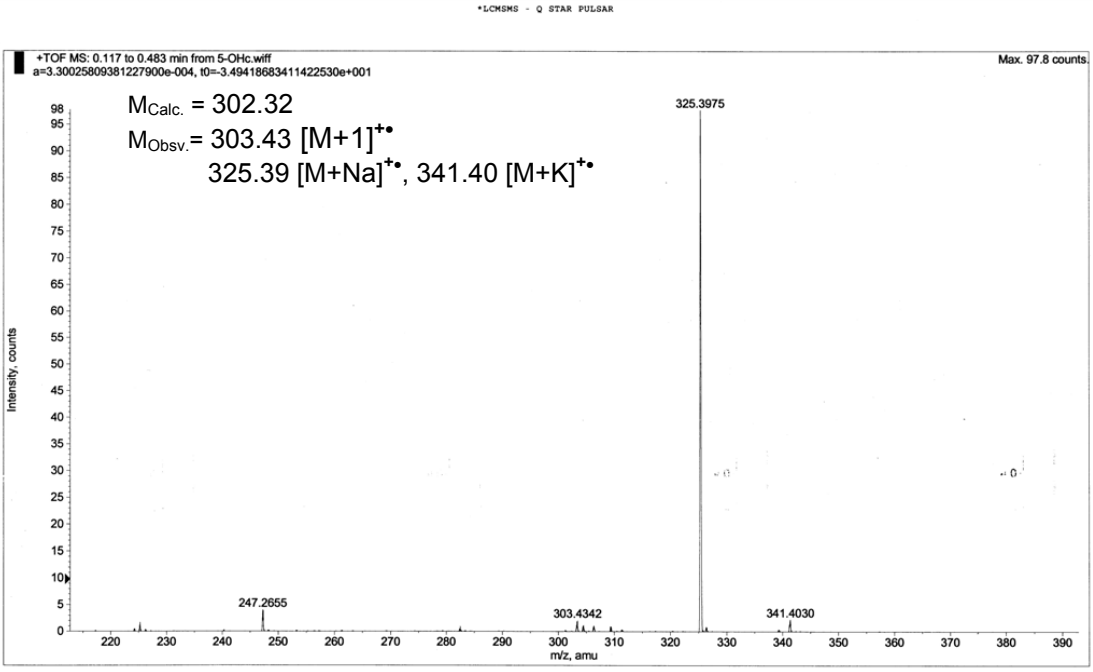
IR Spectra of compound 26



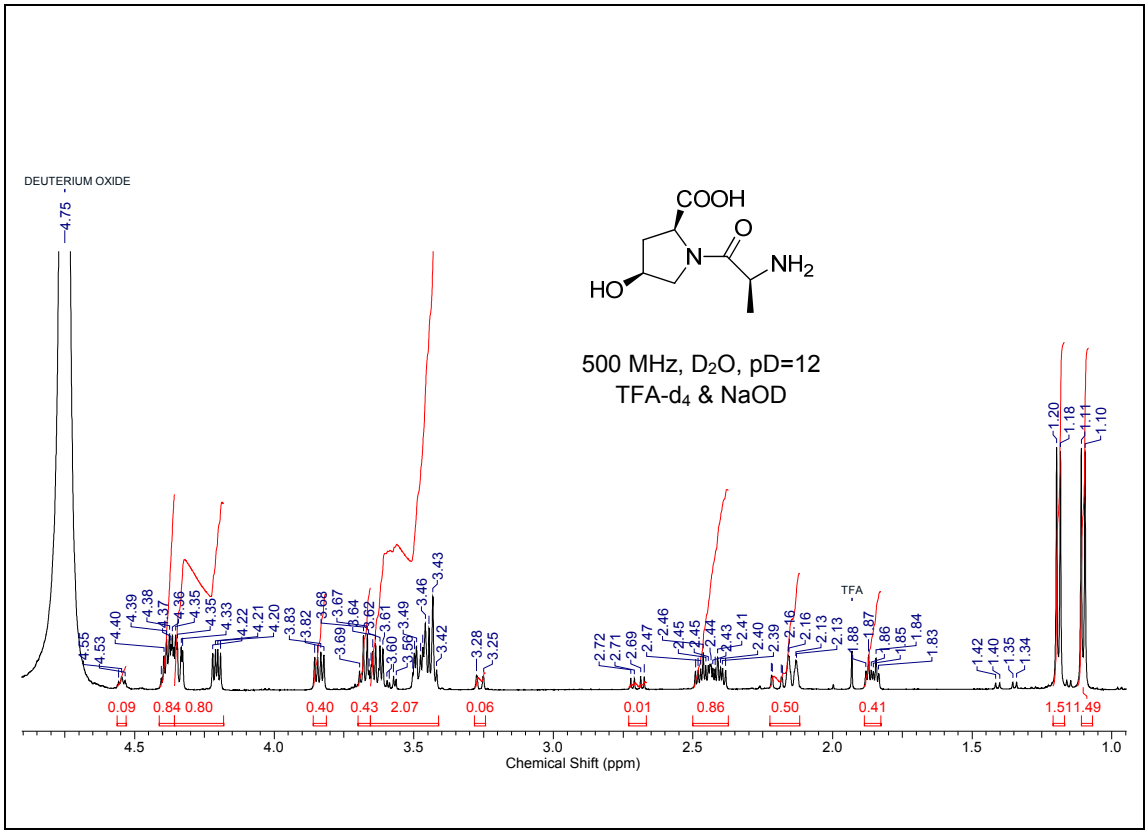
^{1}H NMR of compound 27



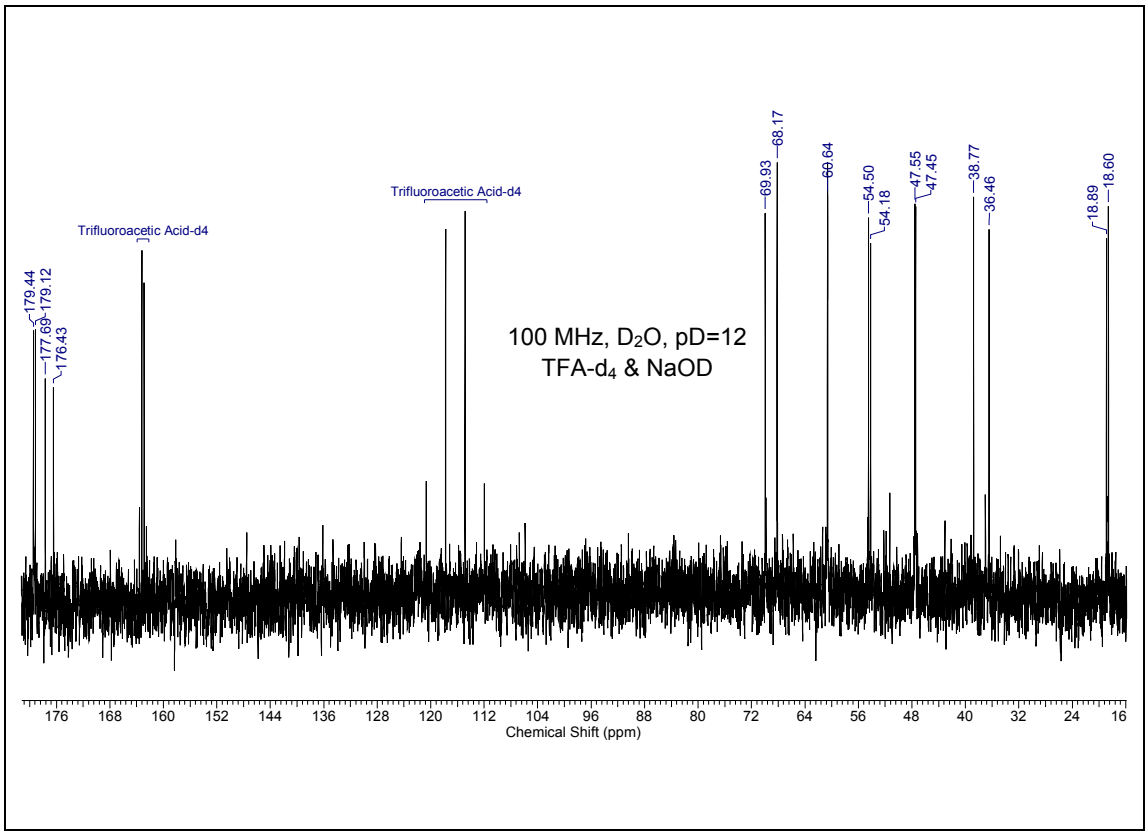
ESI-MS of compound 27



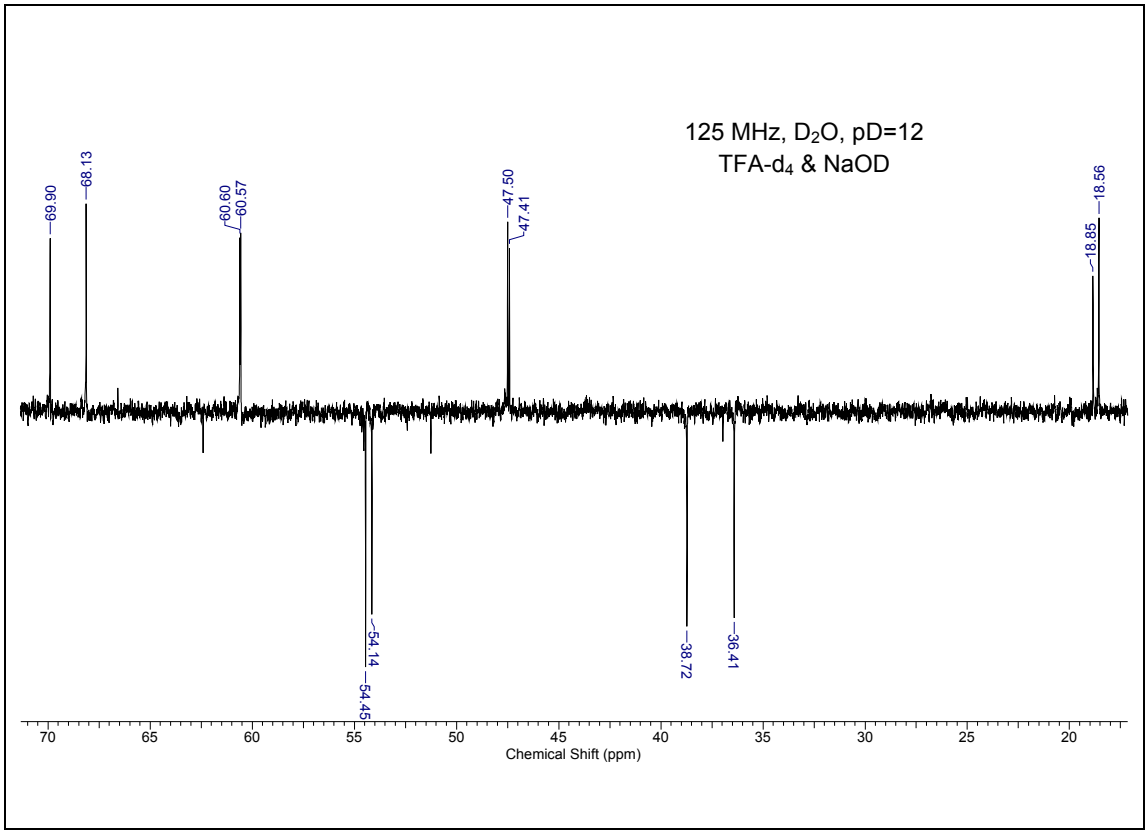
^{1}H NMR of compound 28



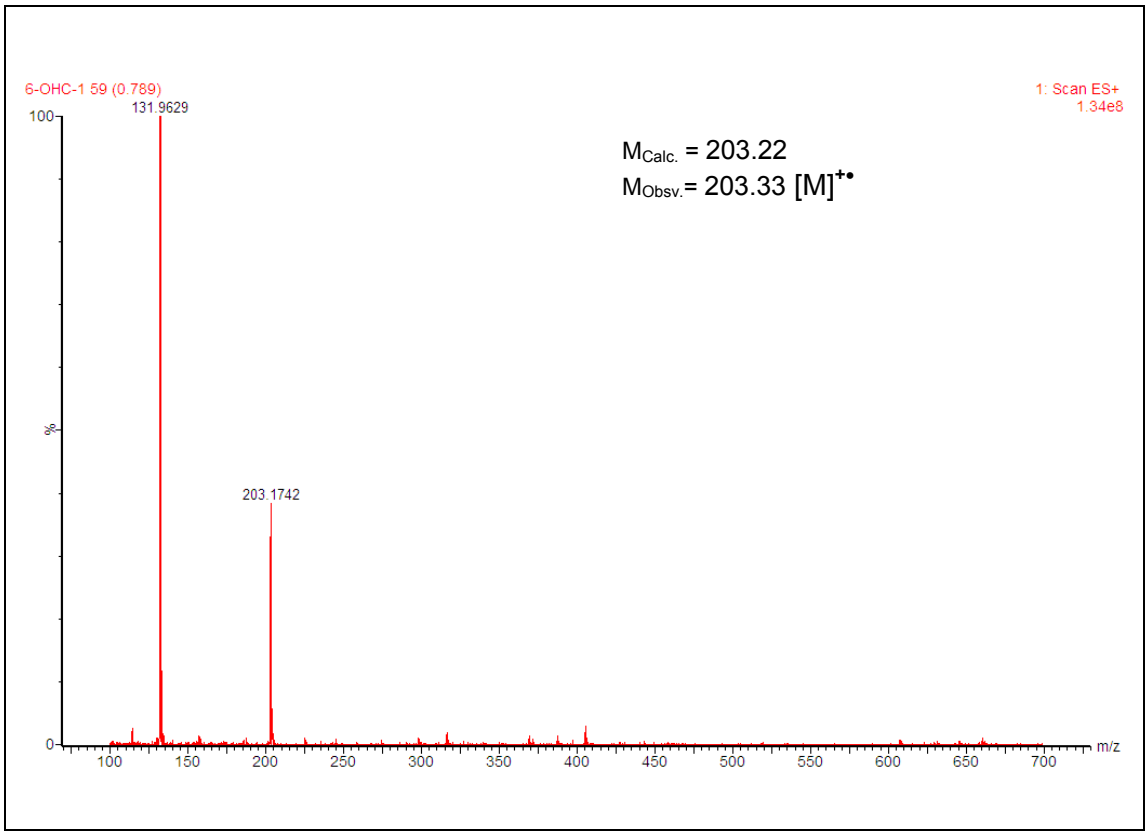
^{13}C NMR of compound 28



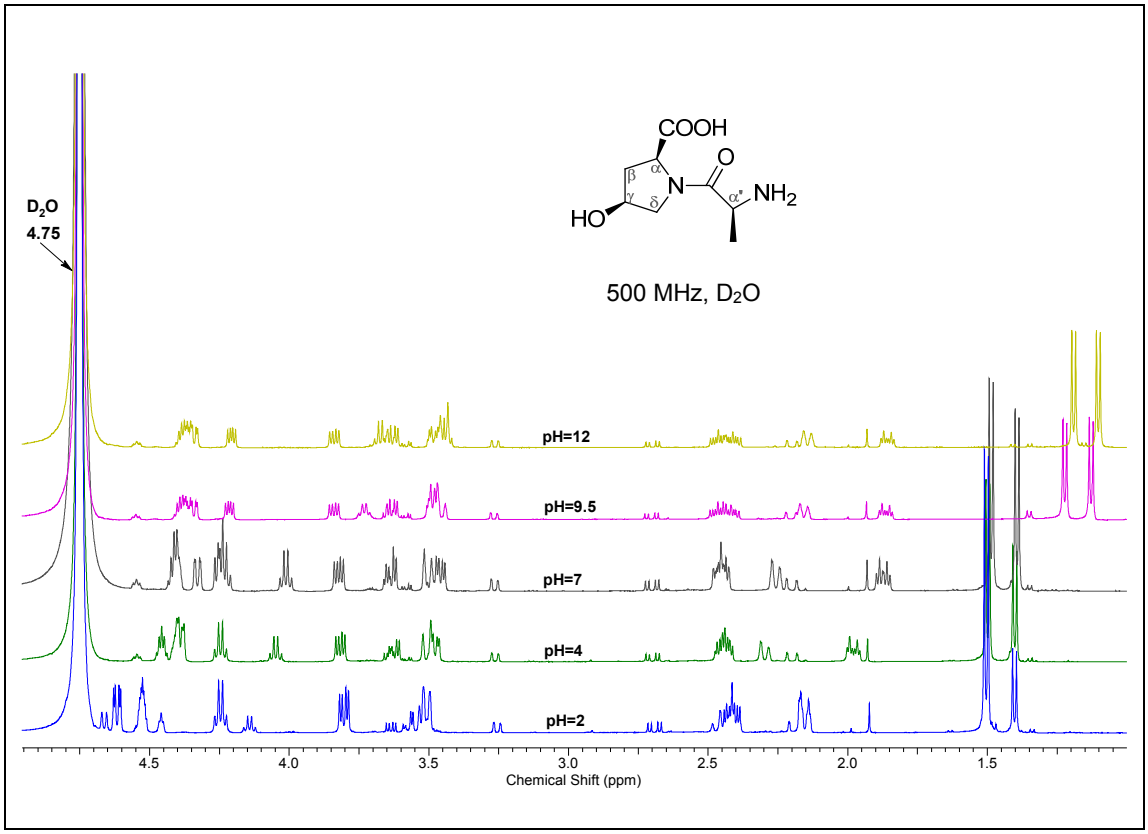
DEPT-^{13}C NMR of compound 28



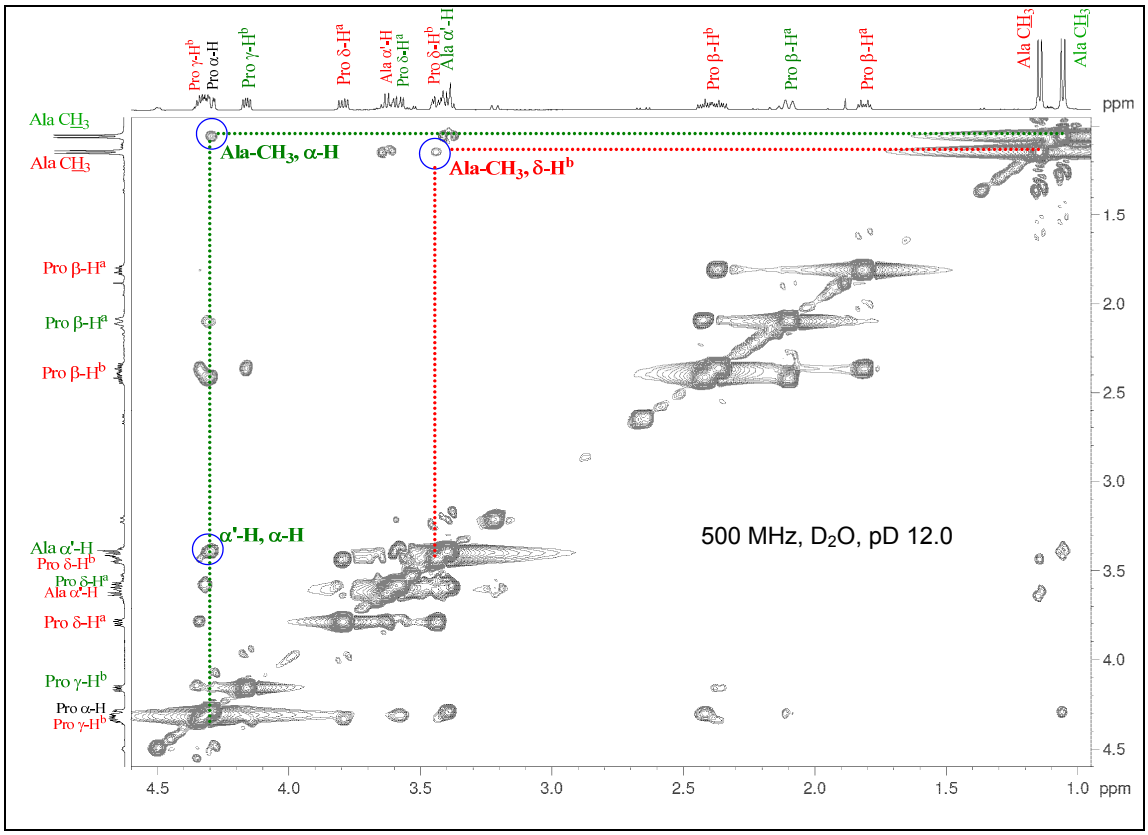
ESI-MS of compound 28



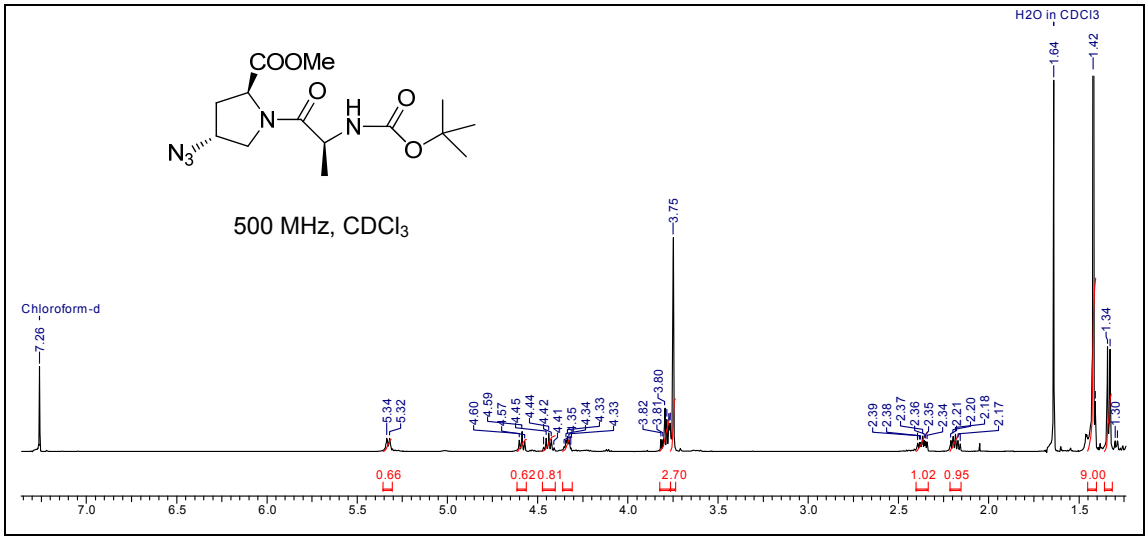
Variable pH ^{1}H NMR of compound 28



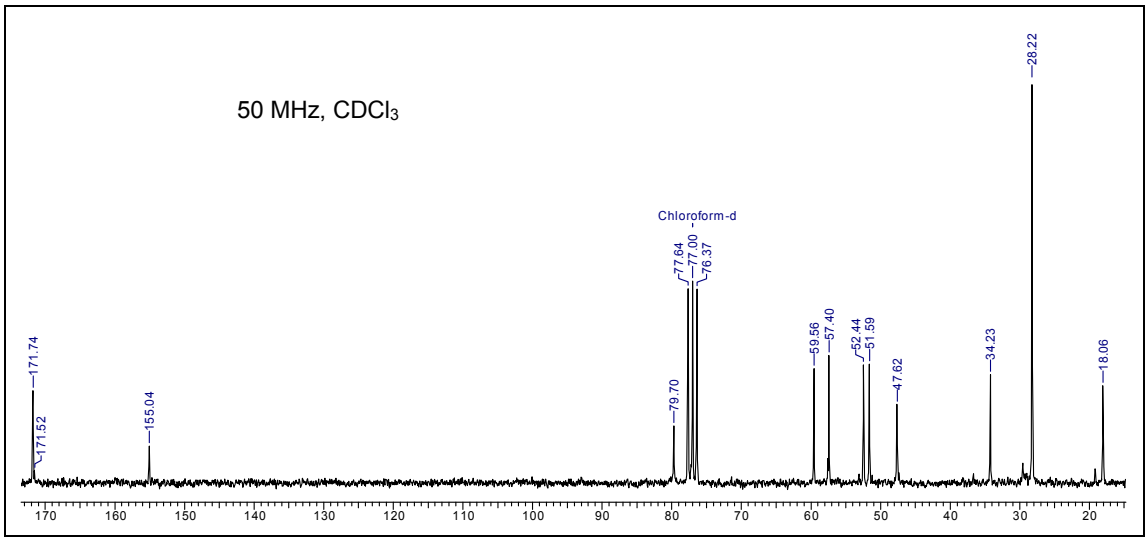
^{1}H-^{1}H NOESY of compound 28



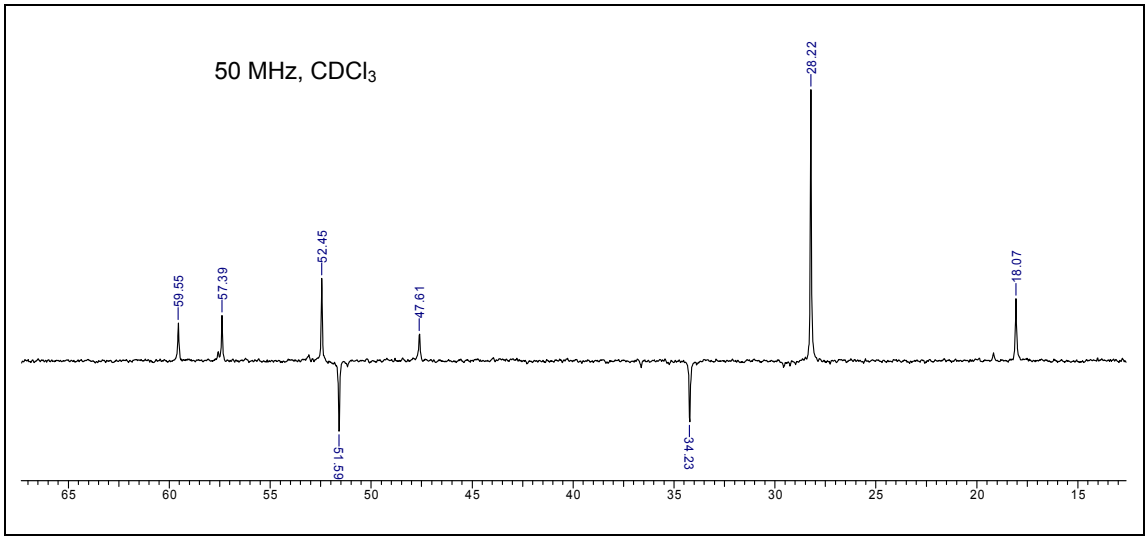
^{1}H NMR of compound 29



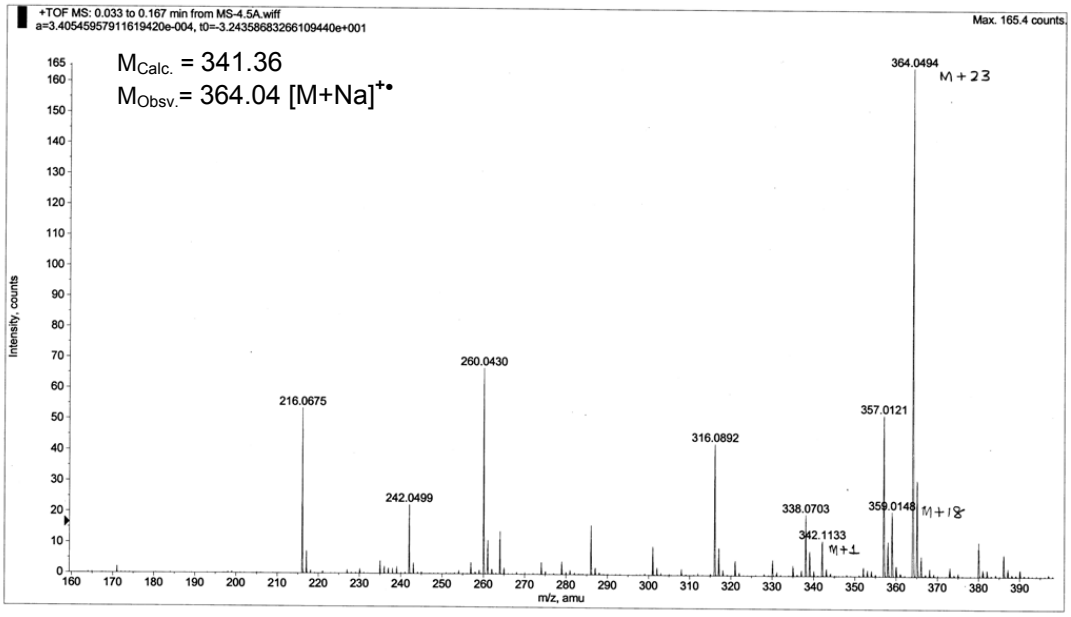
^{13}C NMR of compound 29



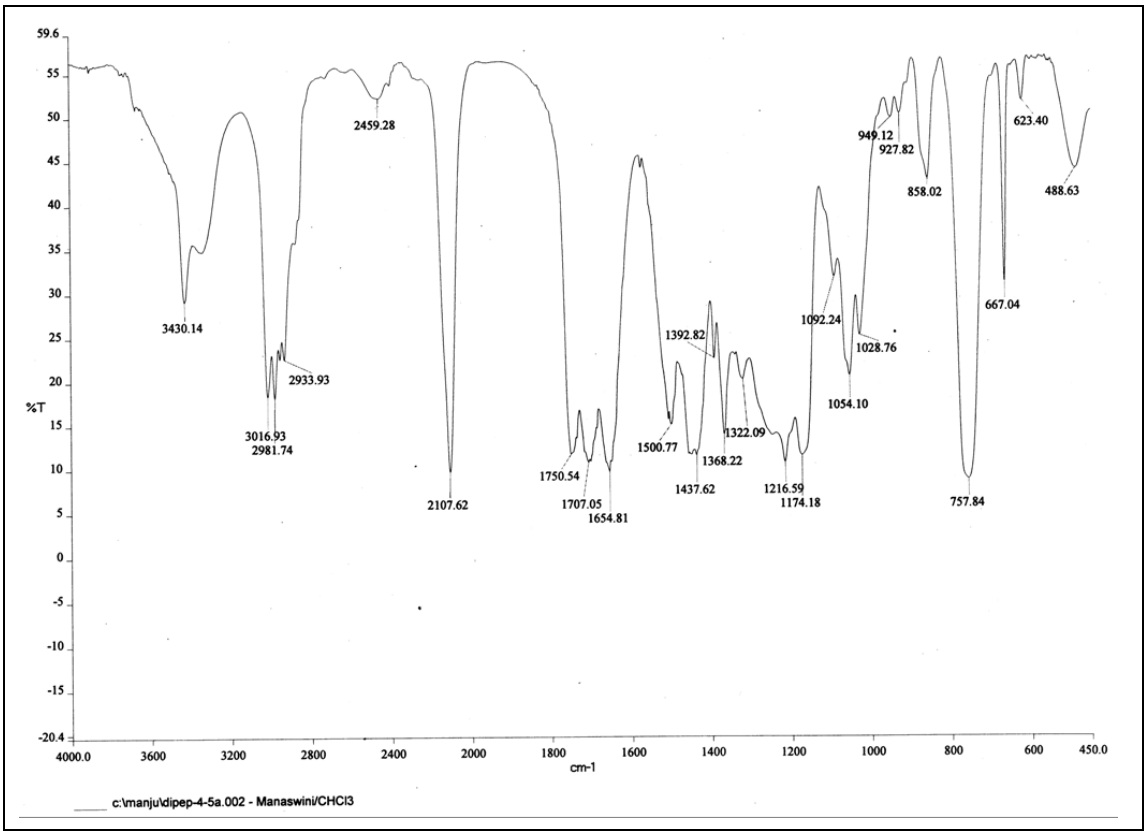
DEPT-^{13}C NMR of compound 29



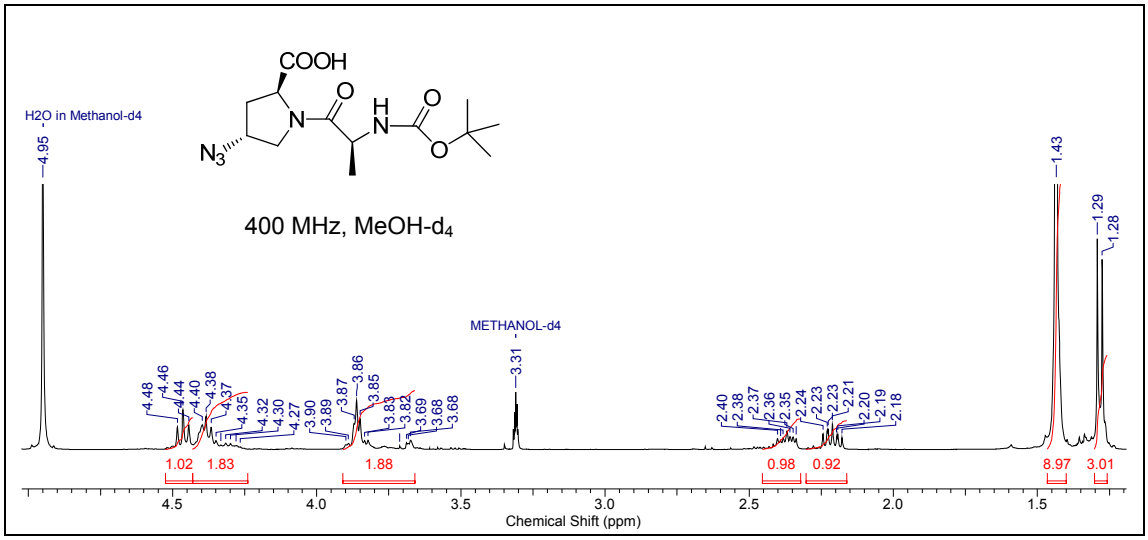
ESI-MS of compound 29



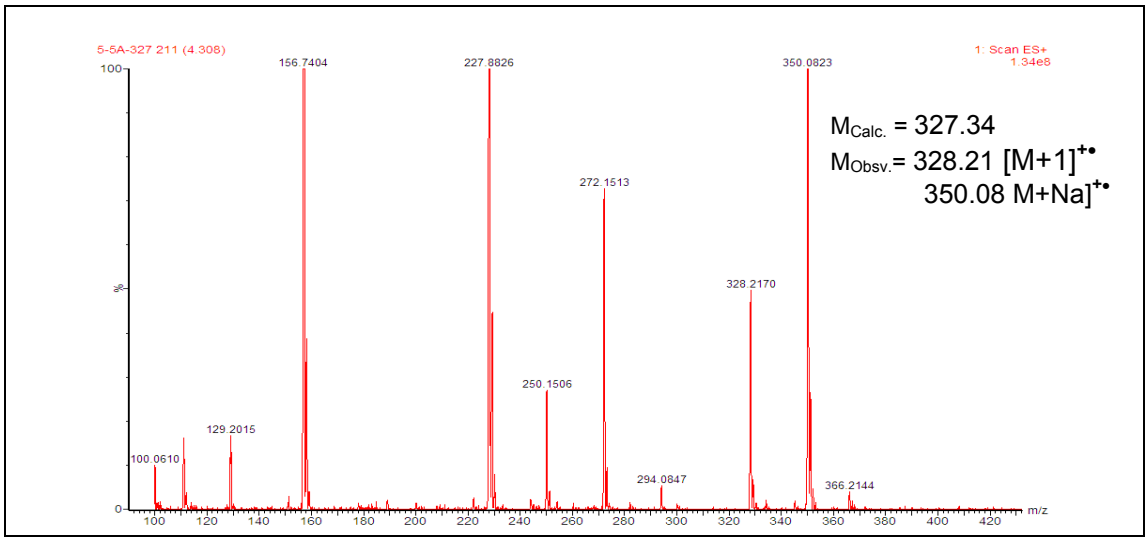
IR Spectra of compound 29



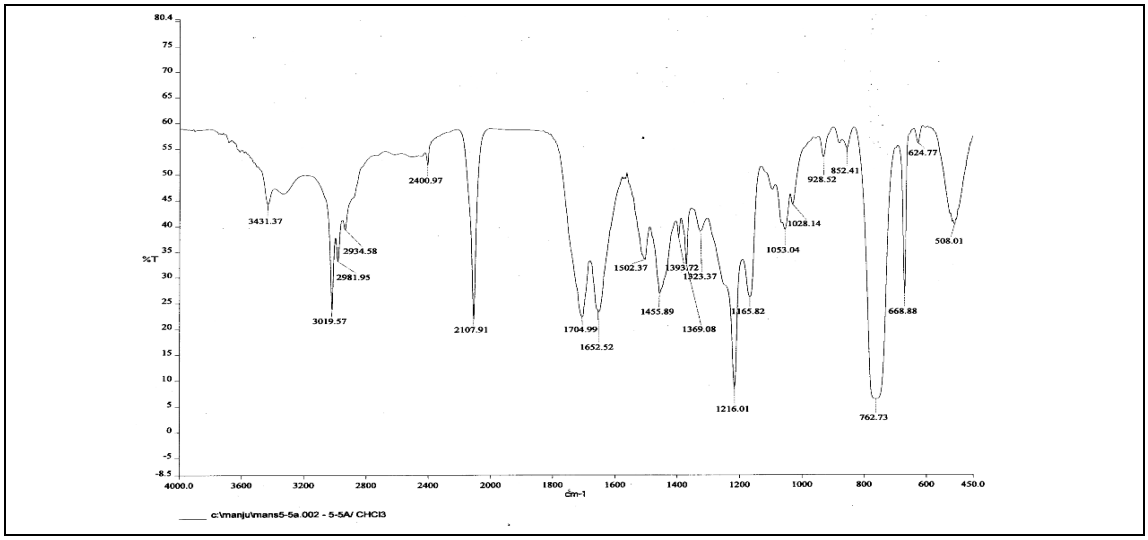
^{1}H NMR of compound 30



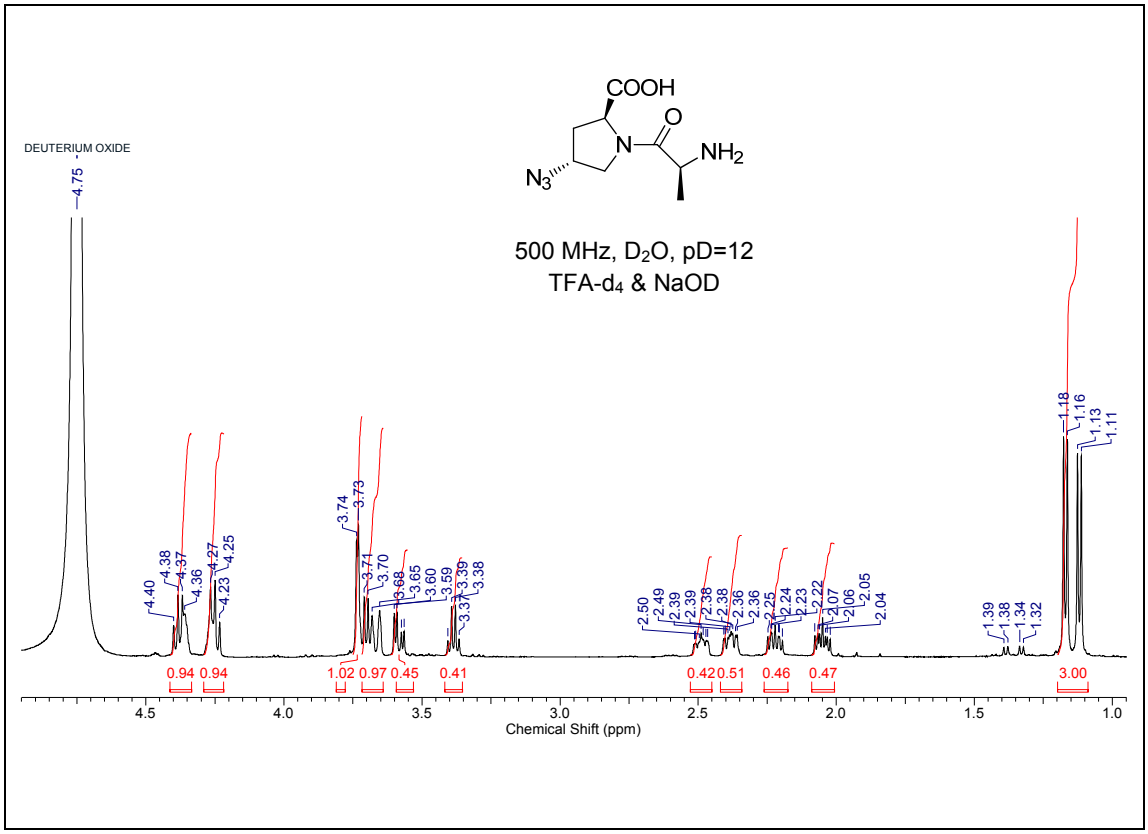
ESI-MS of compound 30



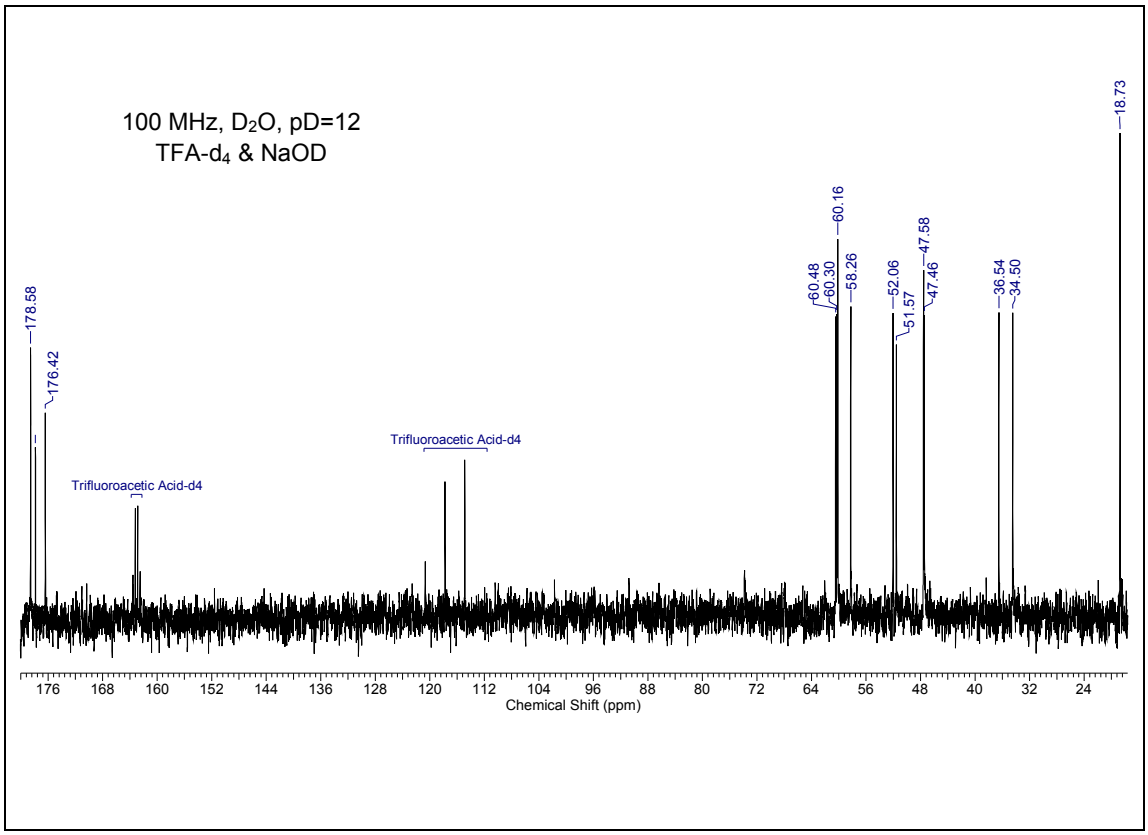
IR Spectra of compound 30



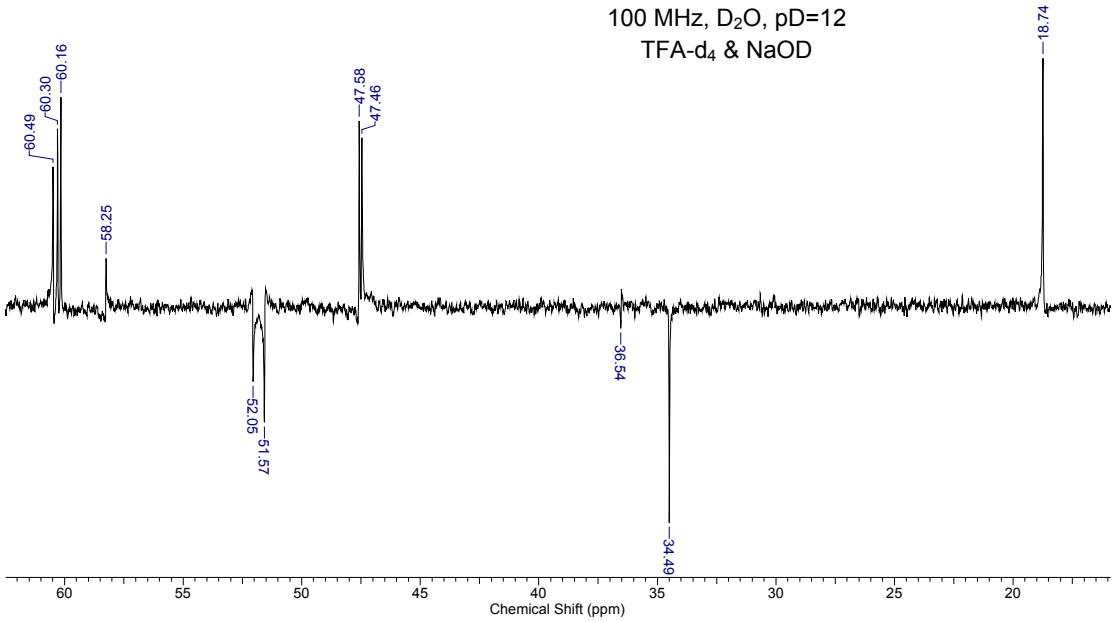
^{1}H NMR of compound 31



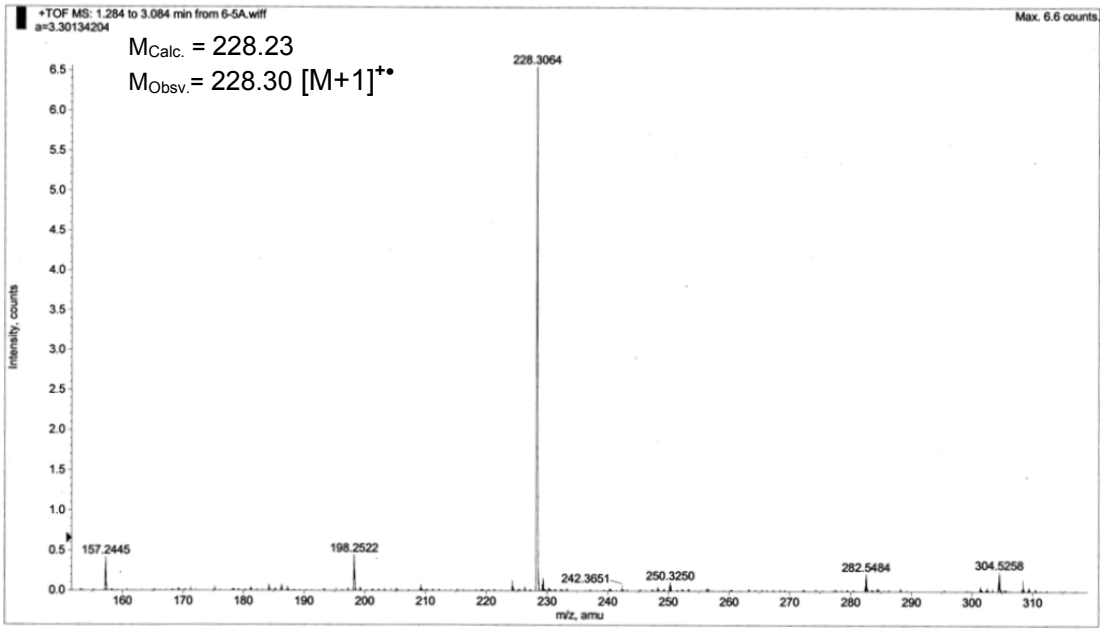
^{13}C NMR of compound 31



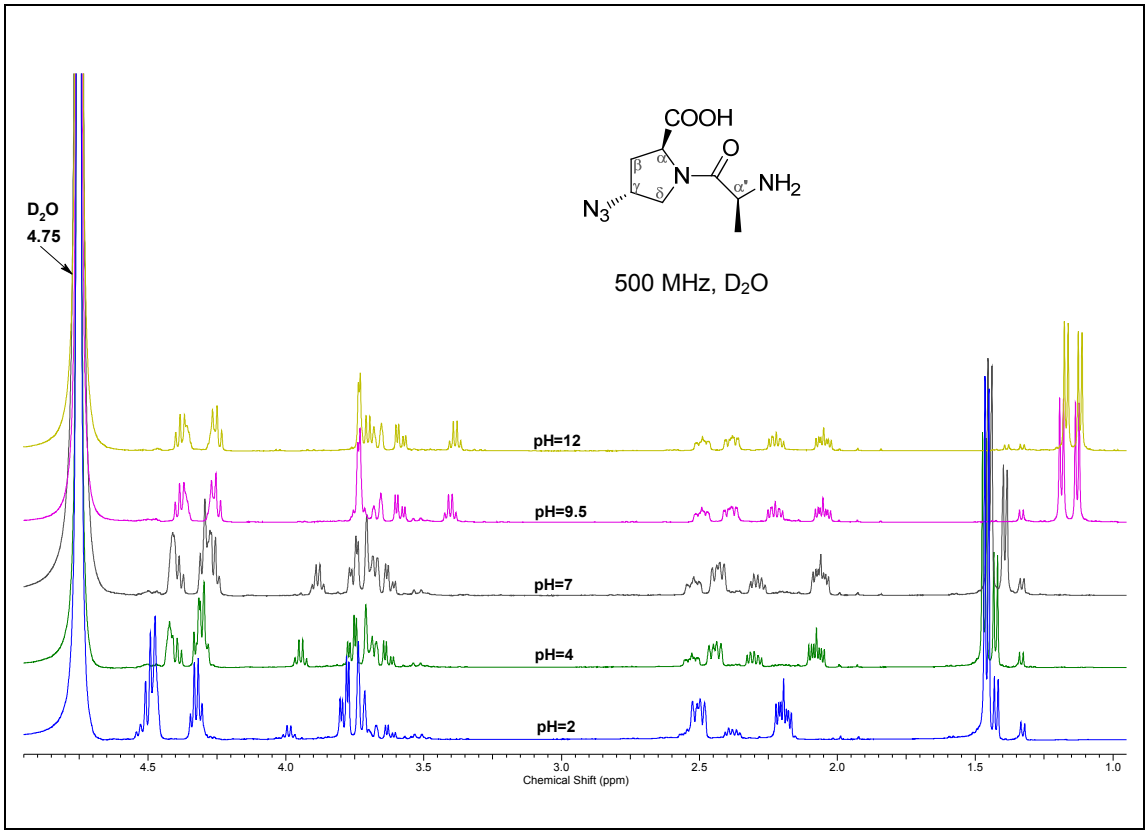
DEPT-^{13}C NMR of compound 31



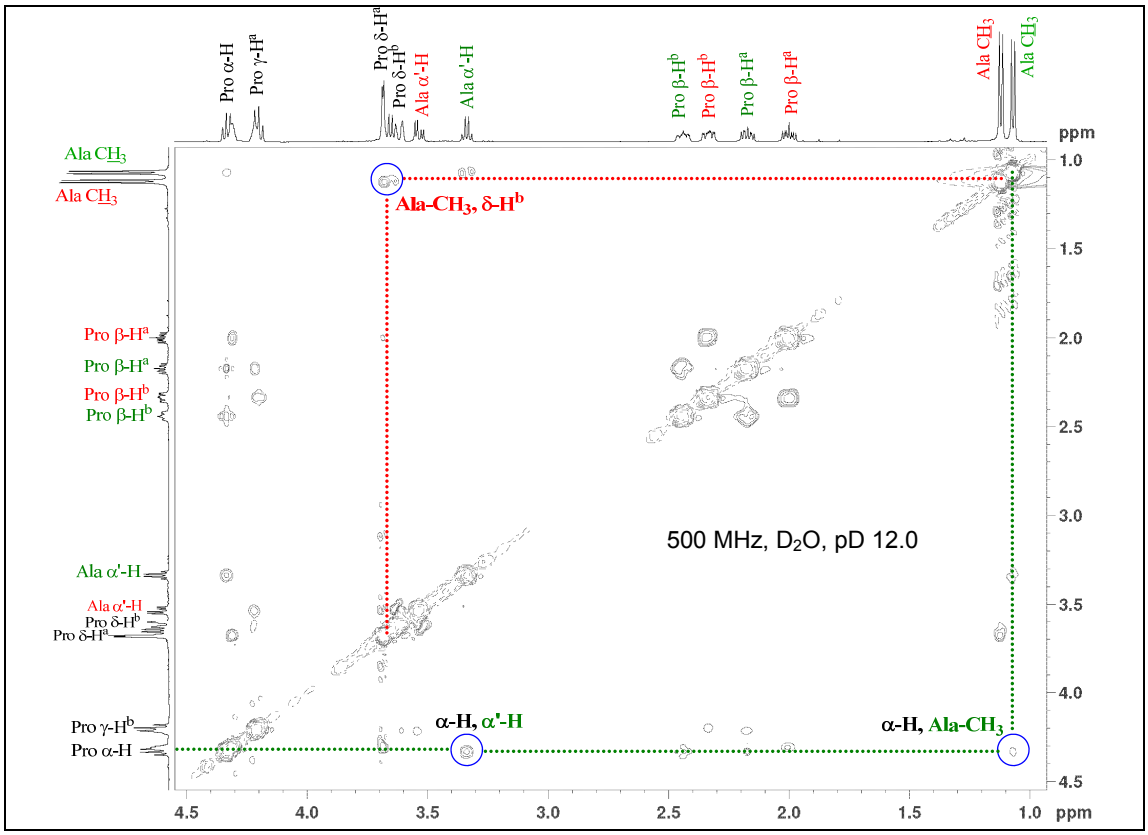
ESI-MS of compound 31



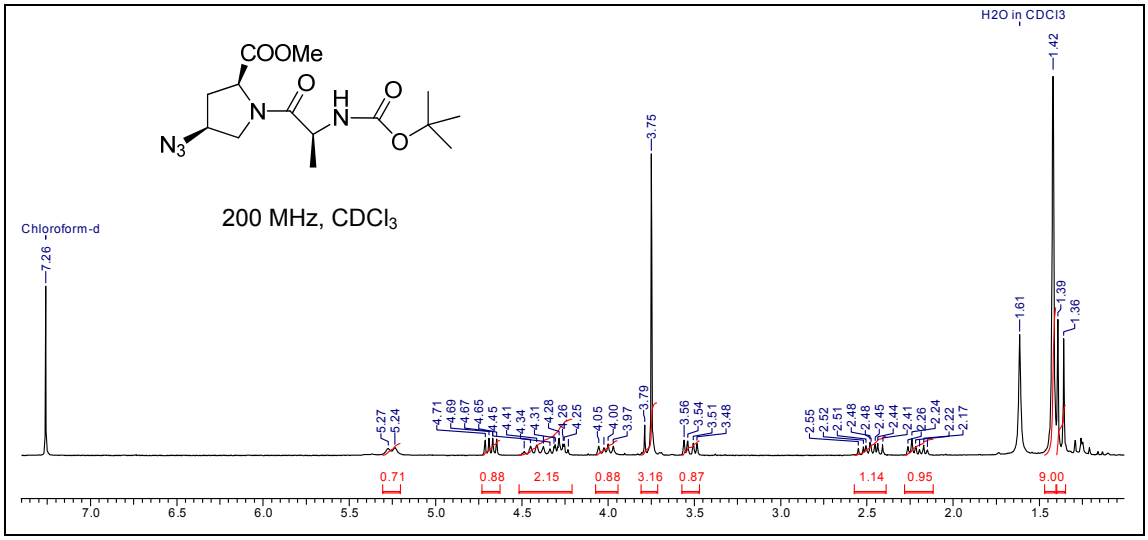
Variable pH ^{1}H NMR of compound 31



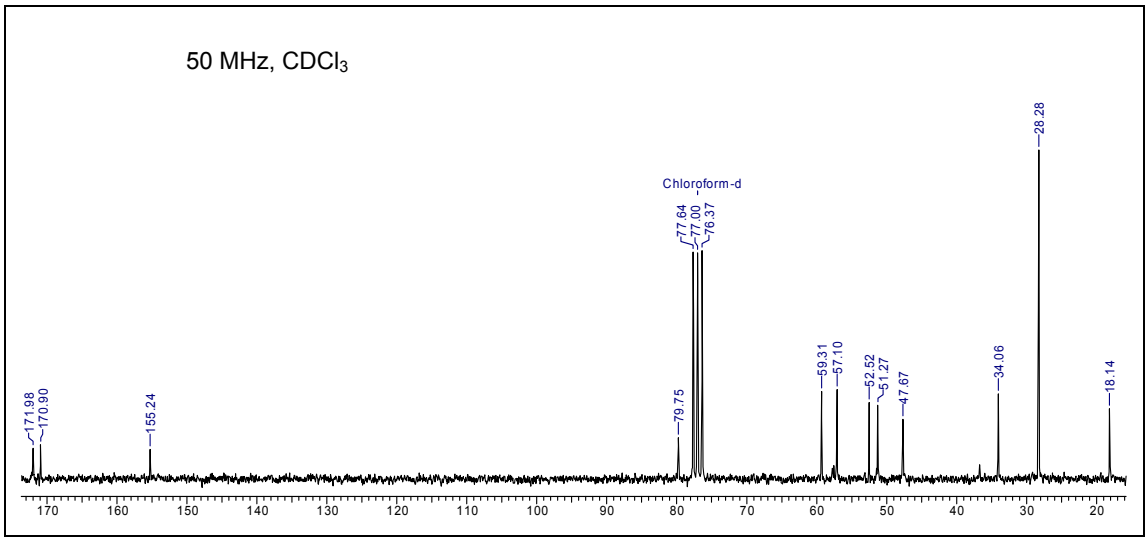
^{1}H-^{1}H NOESY of compound 31



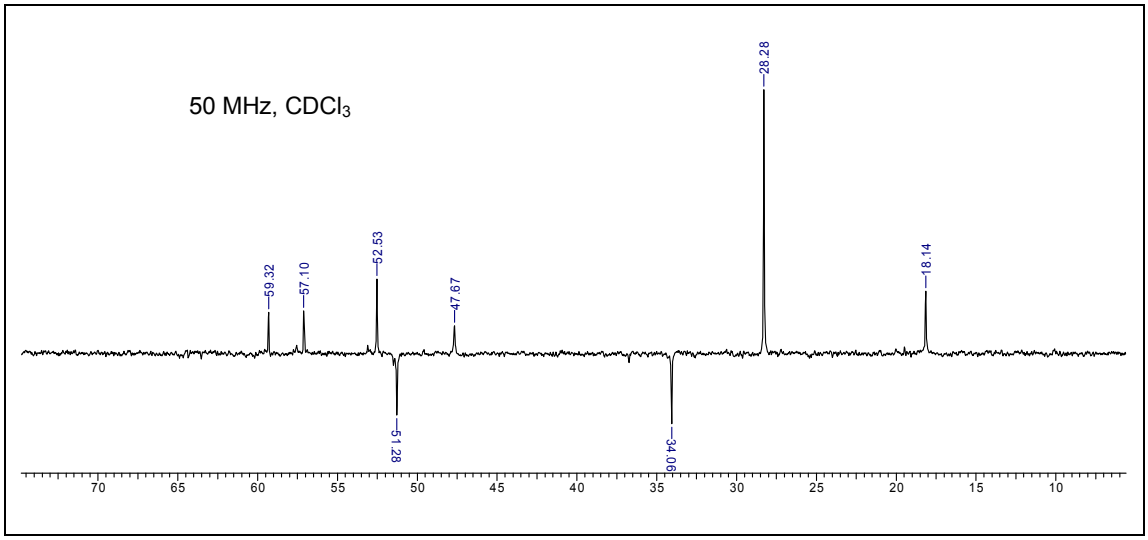
^{1}H NMR of compound 32



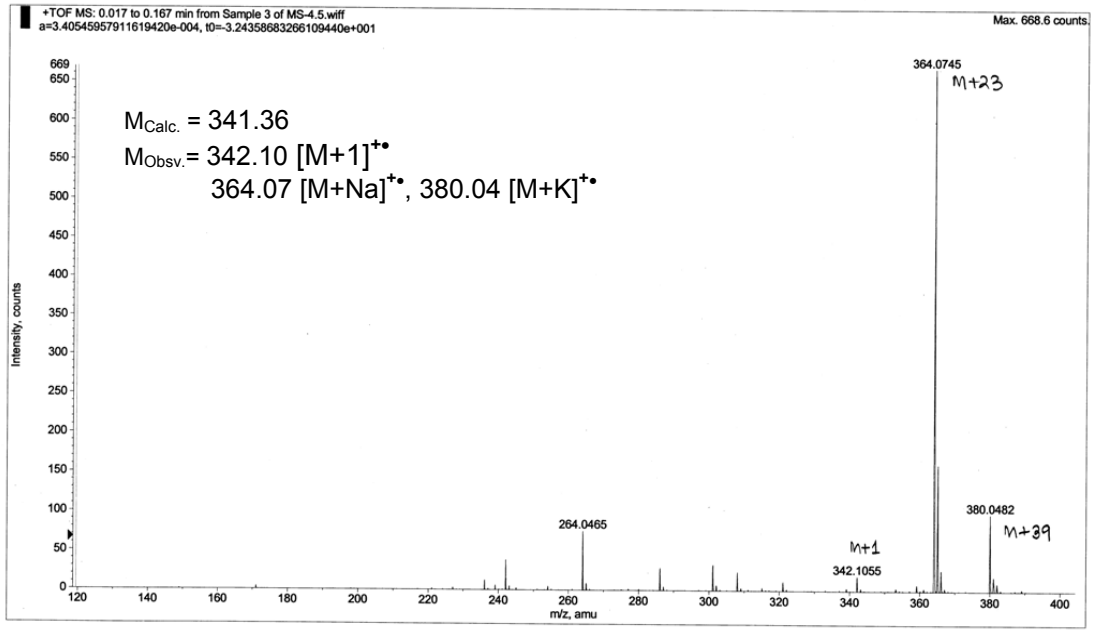
^{13}C NMR of compound 32



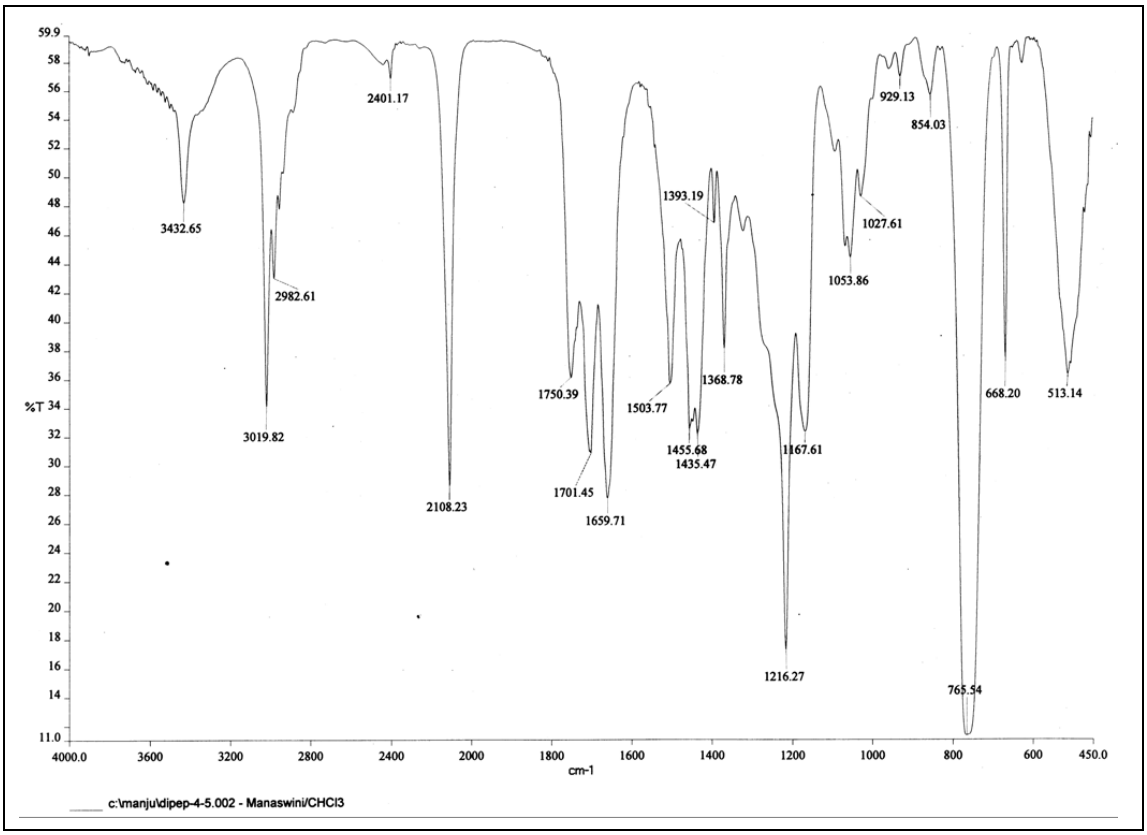
DEPT-^{13}C NMR of compound 32



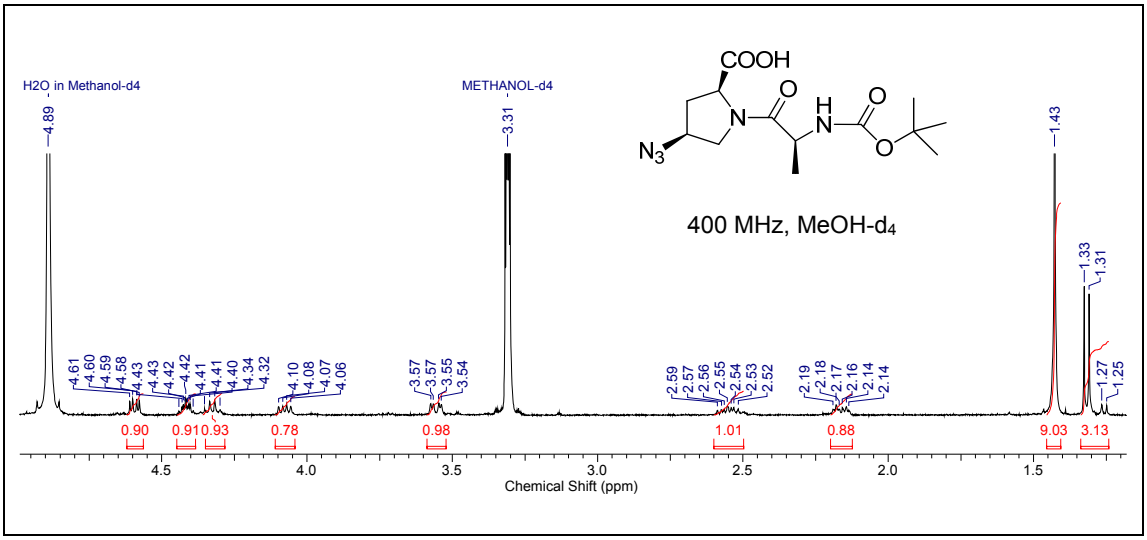
ESI-MS of compound 32



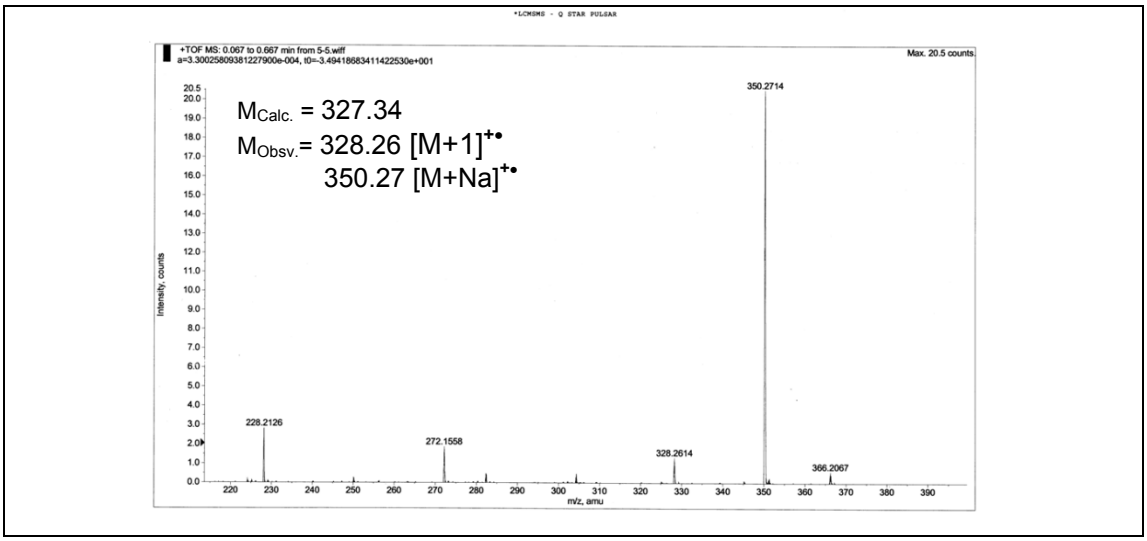
IR Spectra of compound 32



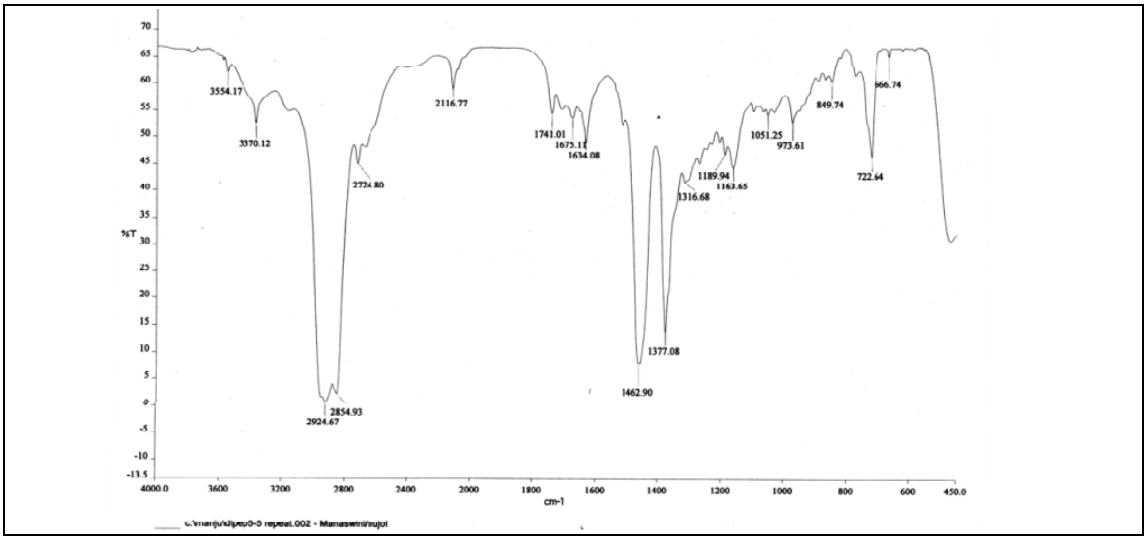
^{1}H NMR of compound 33



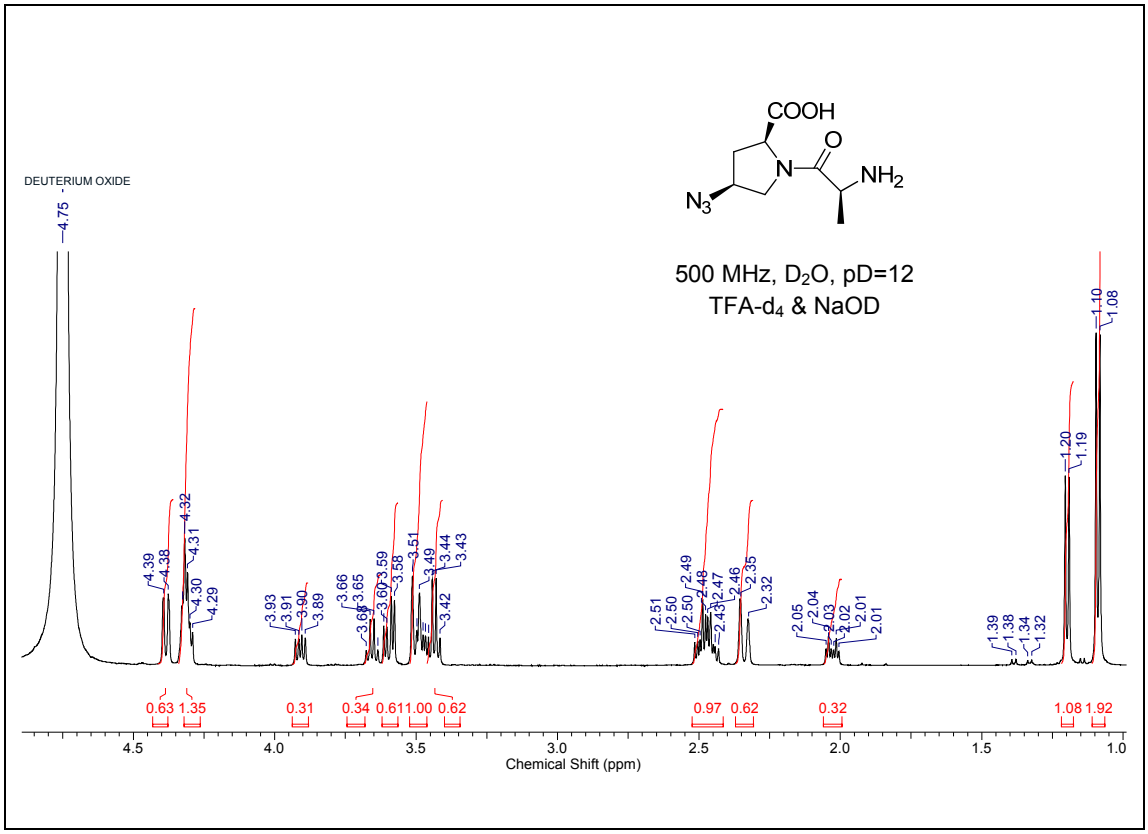
ESI-MS of compound 33



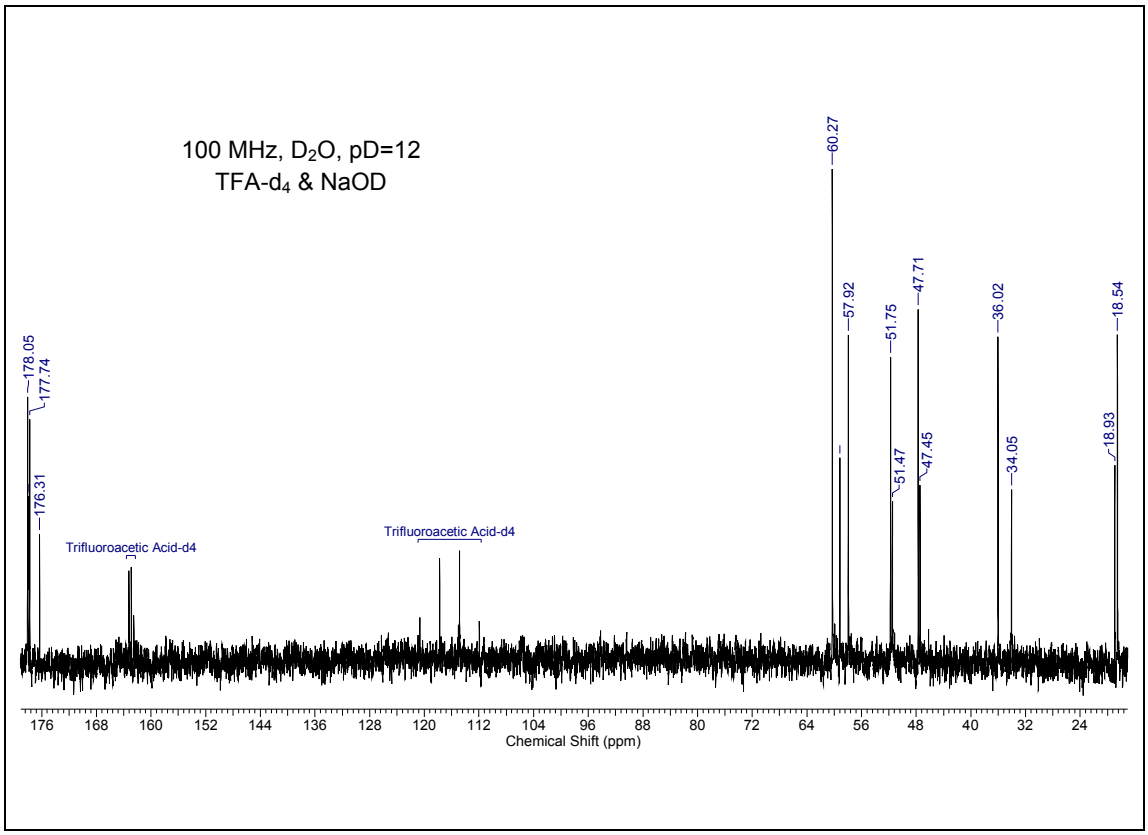
IR Spectra of compound 33



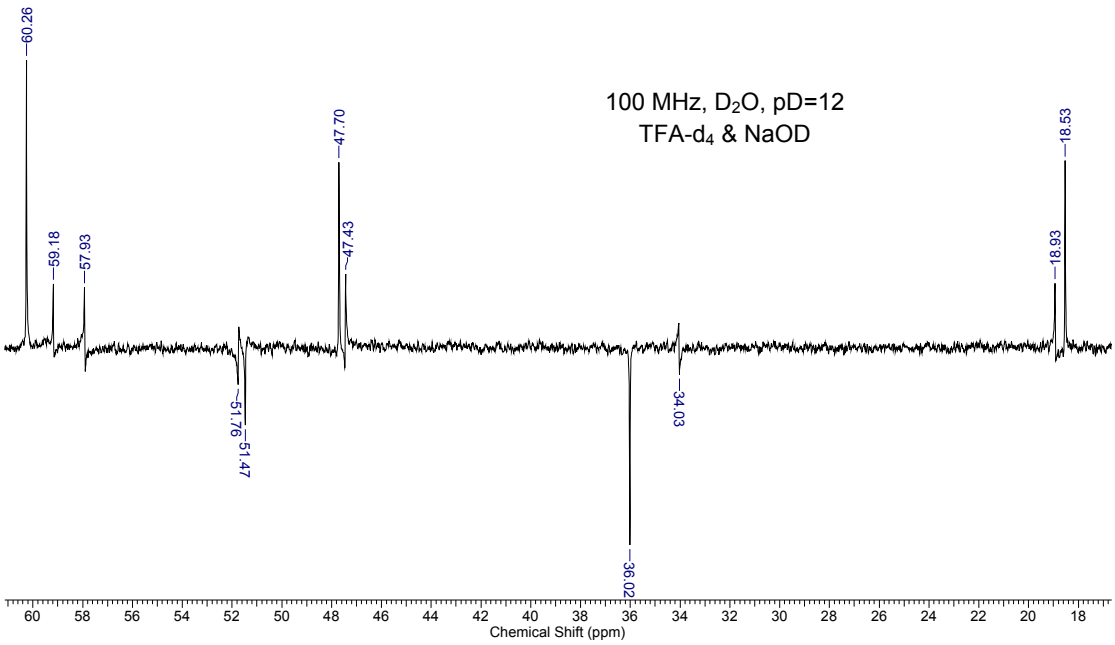
^{1}H NMR of compound 34



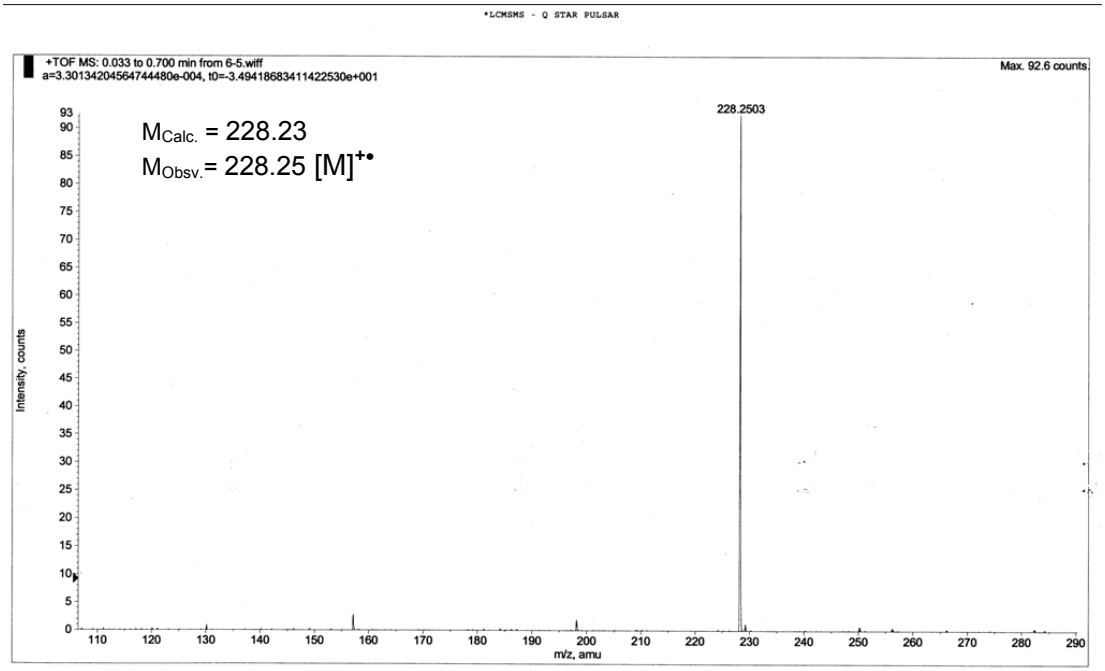
^{13}C NMR of compound 34



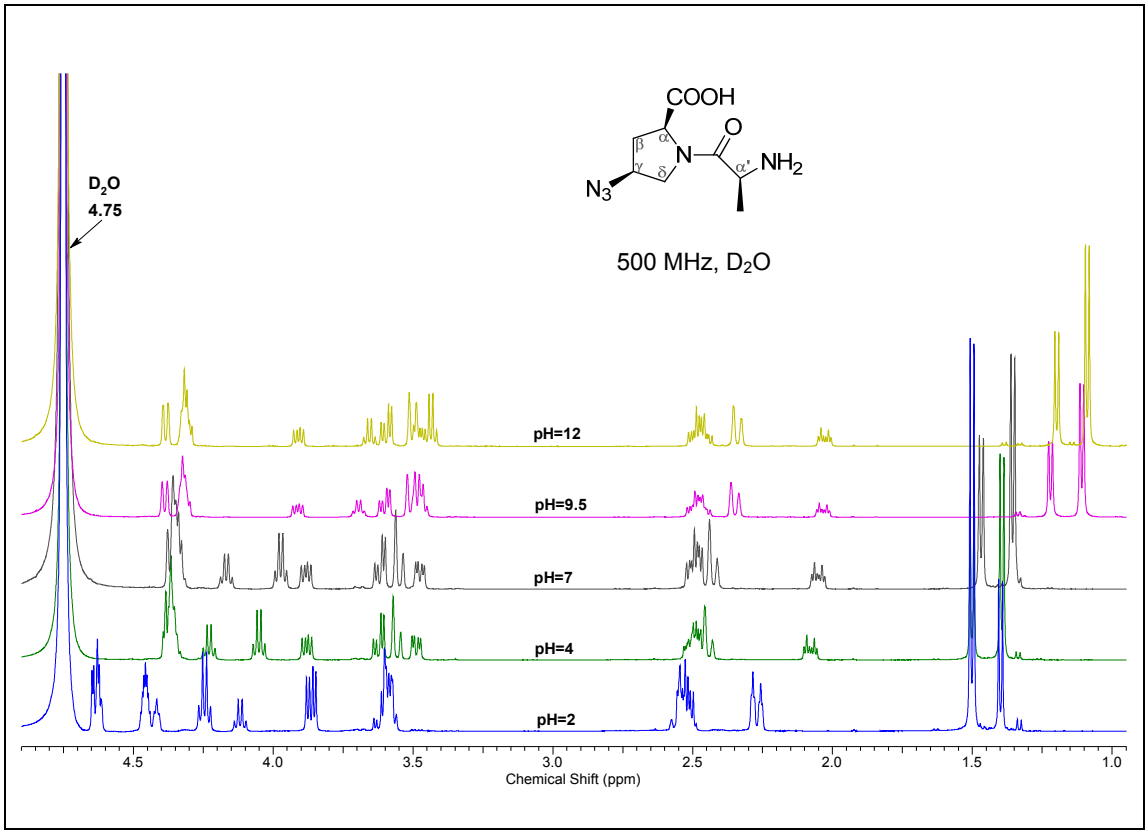
DEPT-^{13}C NMR of compound 34



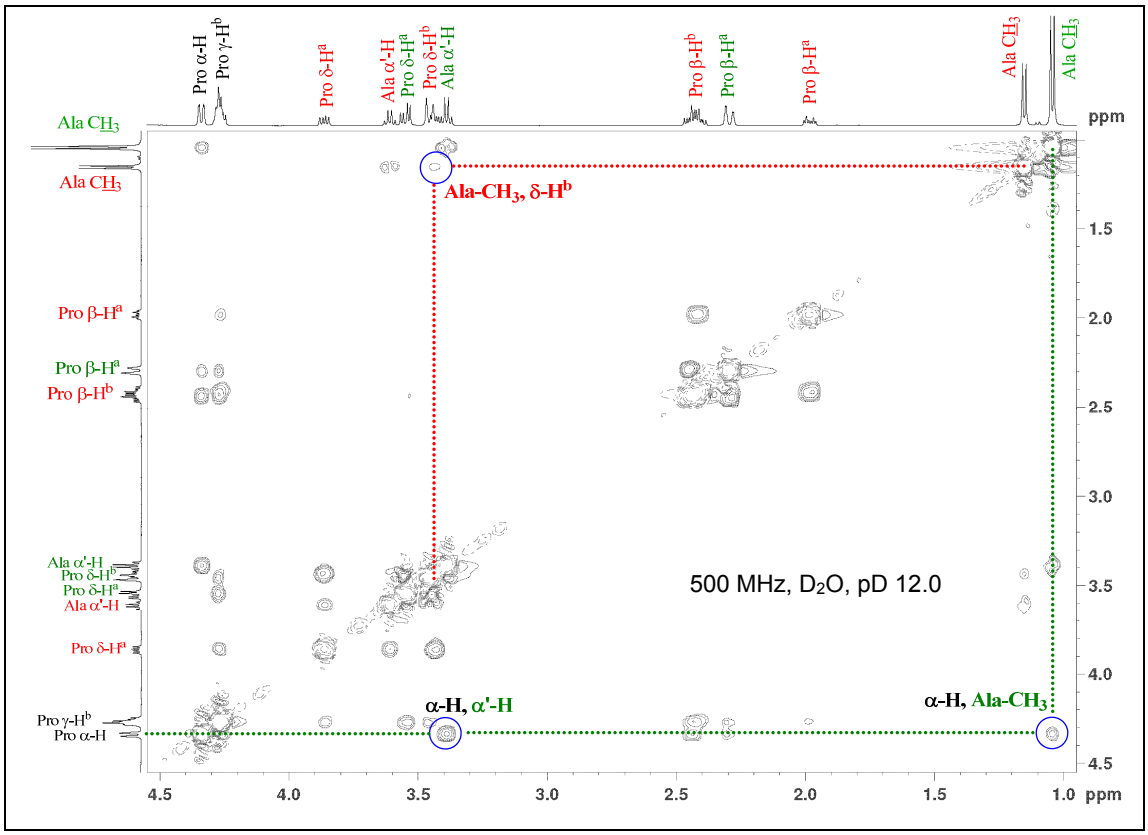
ESI-MS of compound 34



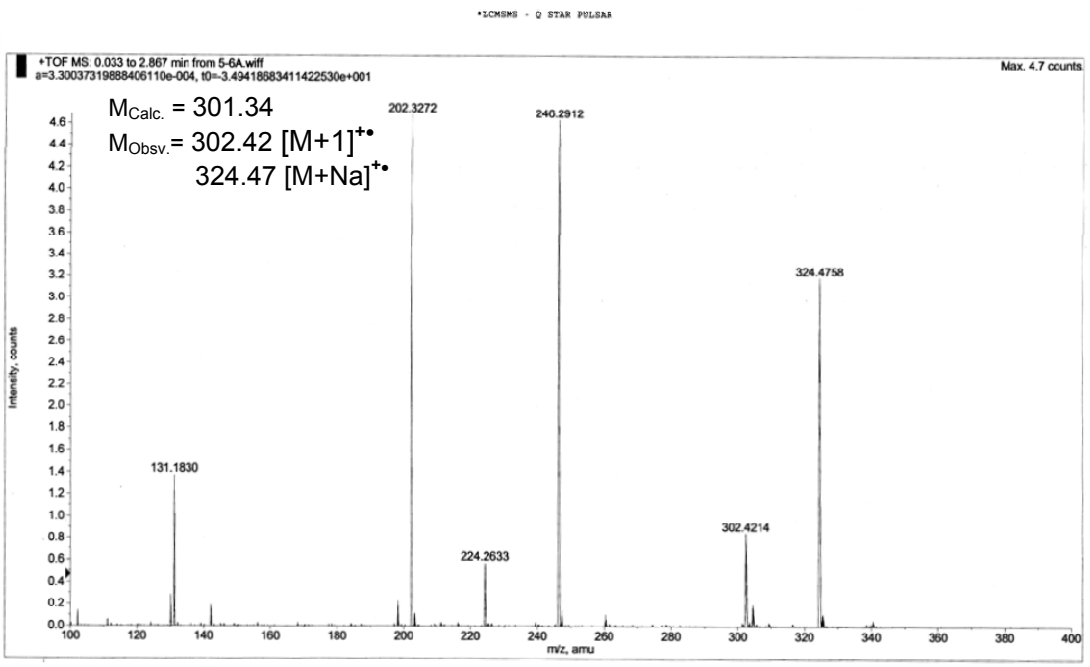
Variable pH ^{1}H NMR of compound 34



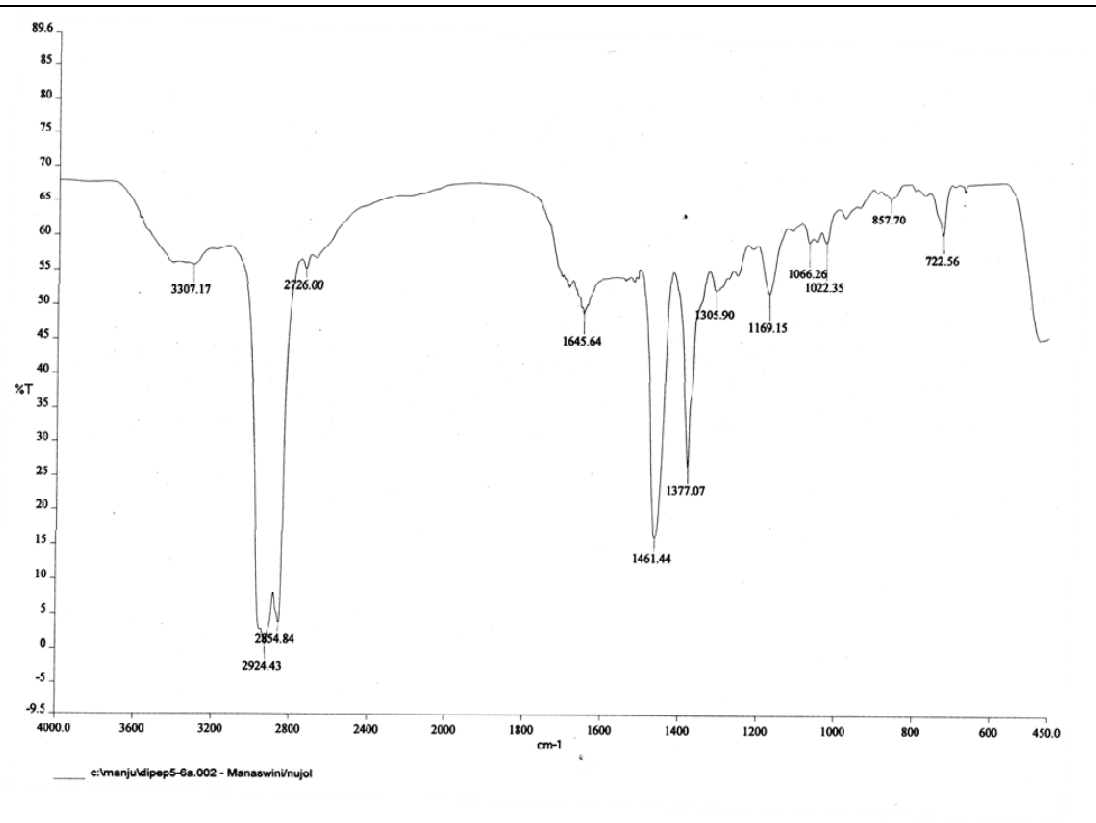
^{1}H-^{1}H NOESY of compound 34



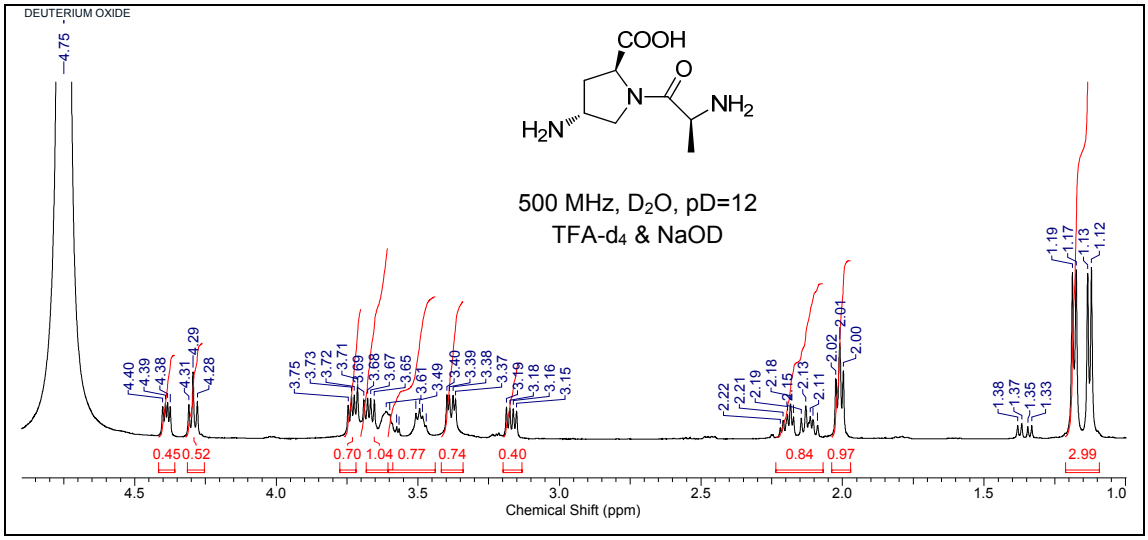
ESI-MS of compound 35



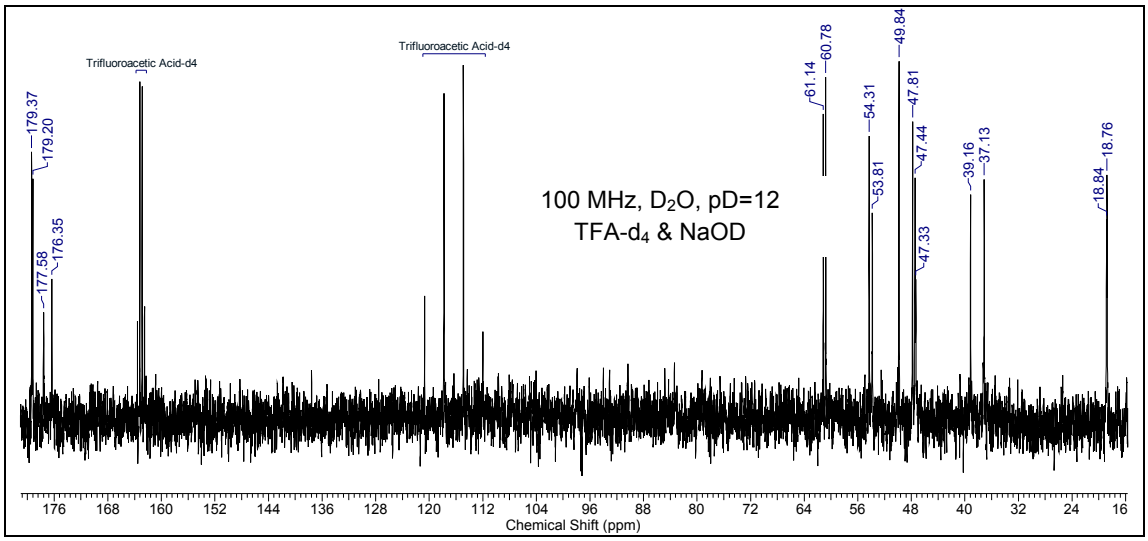
IR Spectra of compound 35



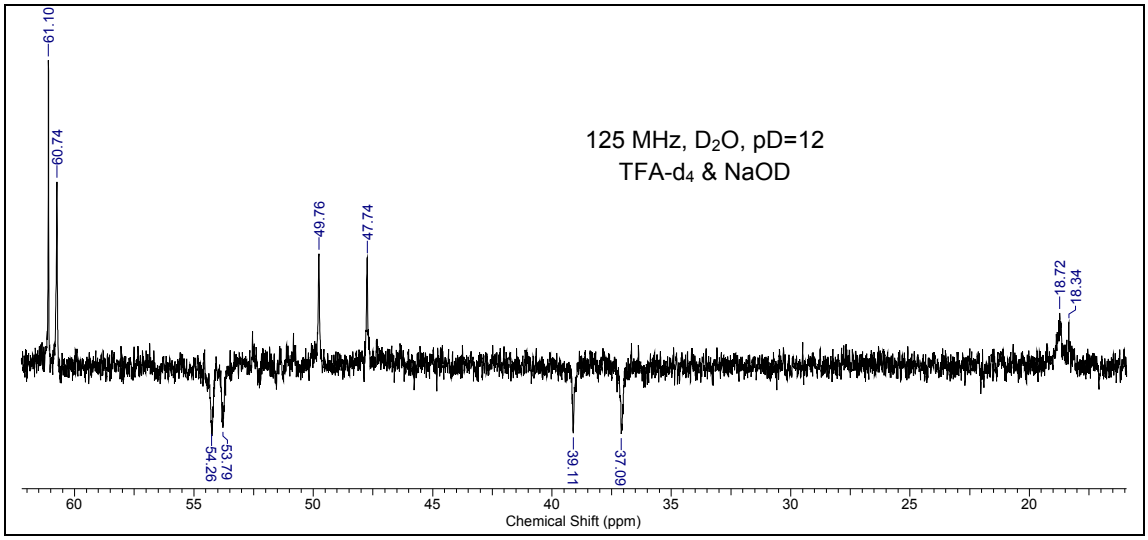
^{1}H NMR of compound 36



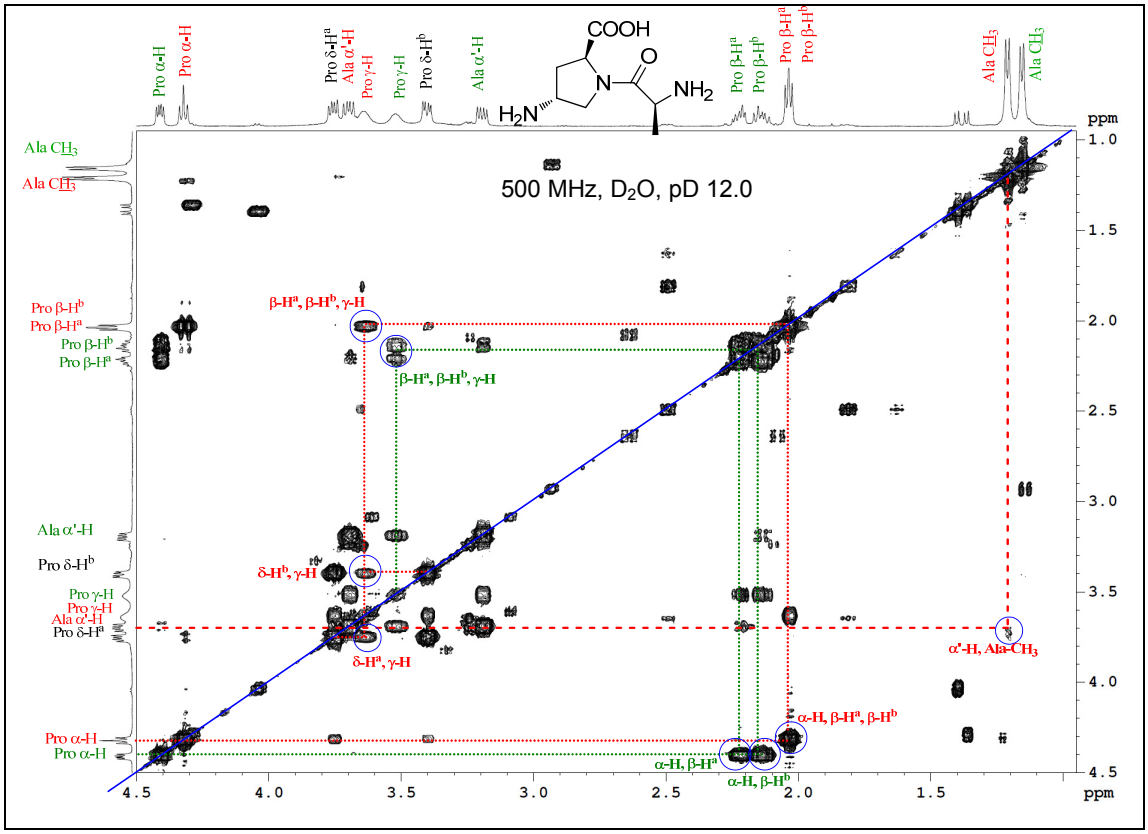
^{13}C NMR of compound 36



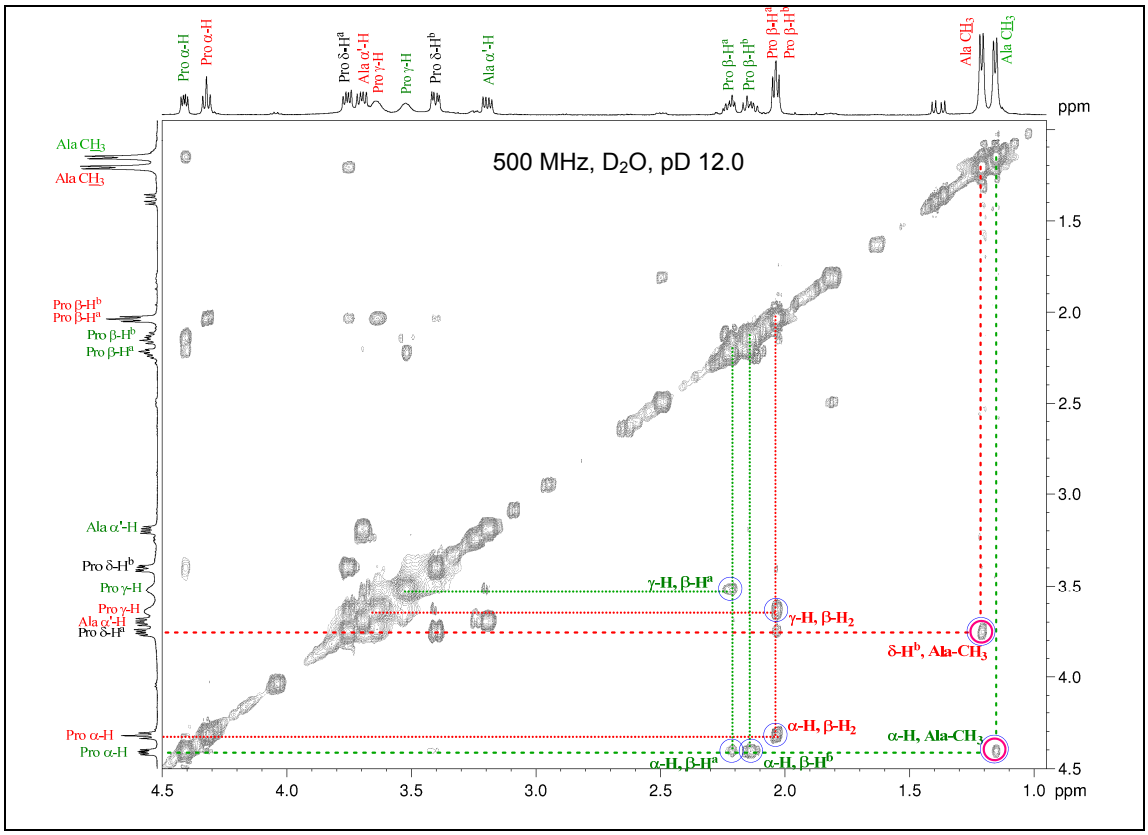
DEPT-^{13}C NMR of compound 36



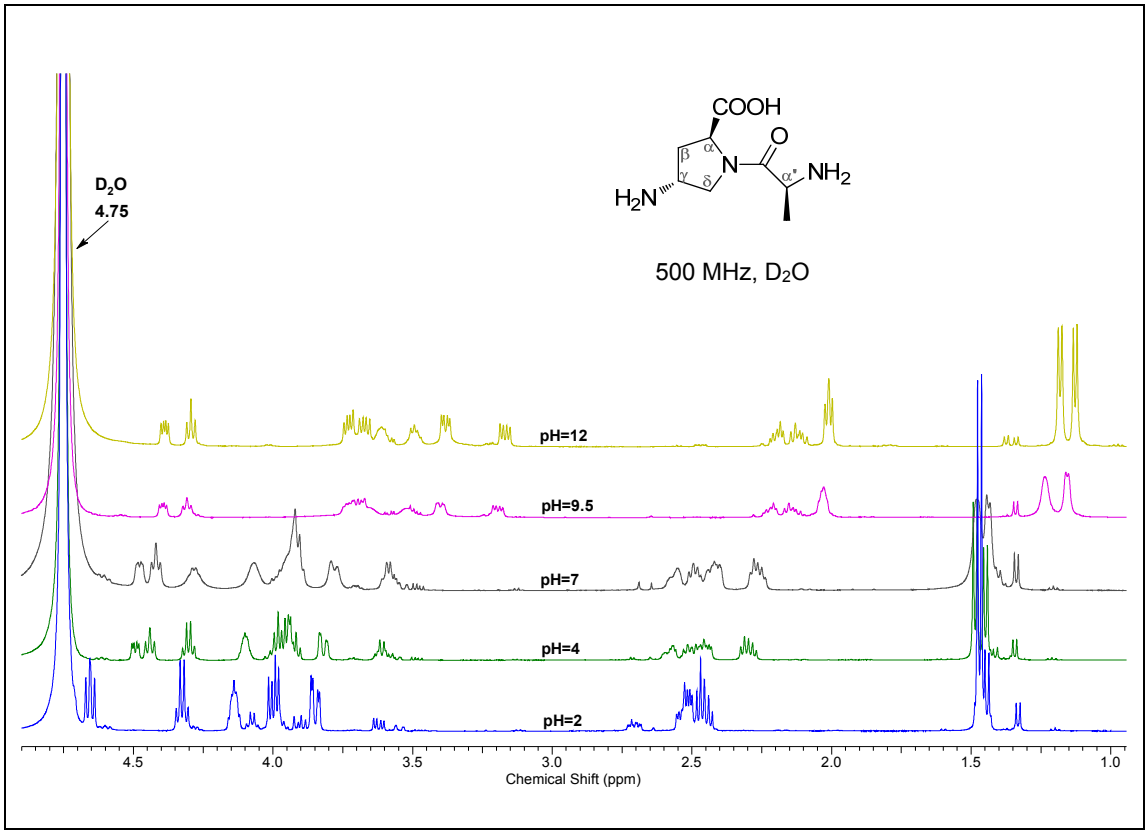
^{1}H-^{1}H COSY of compound 36



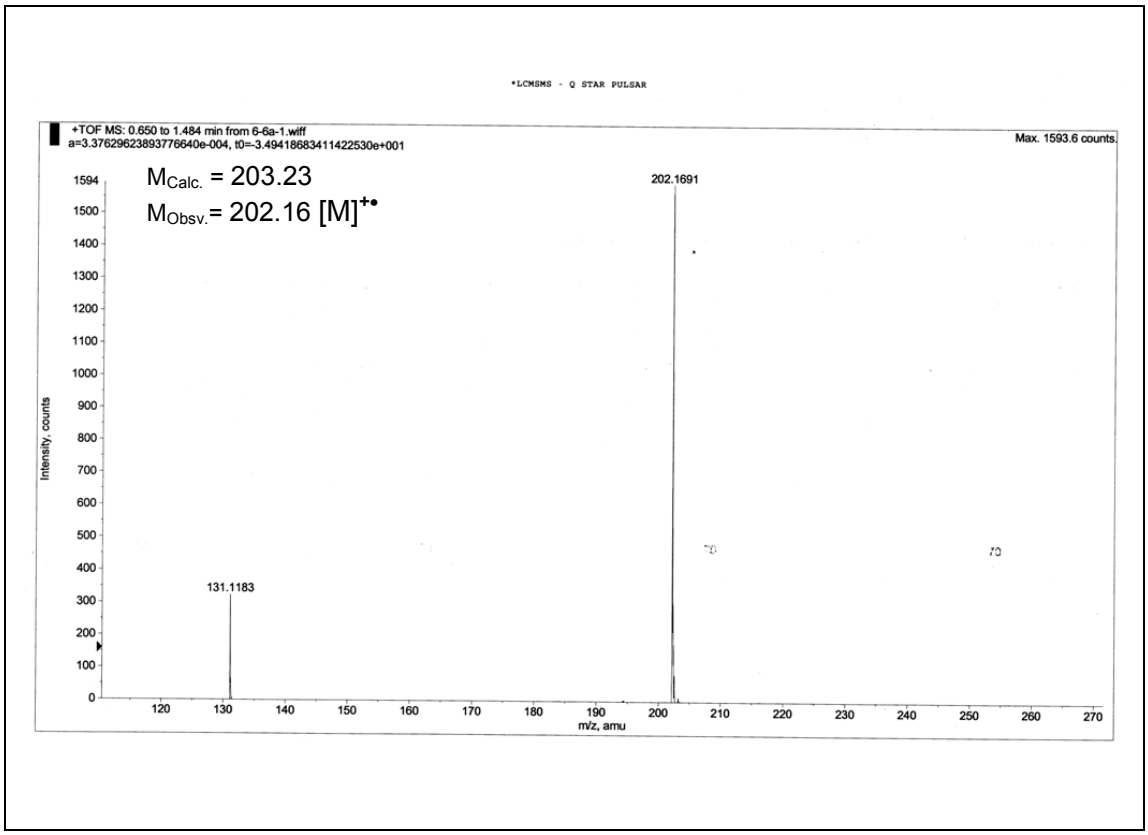
^{1}H-^{1}H NOESY of compound 36



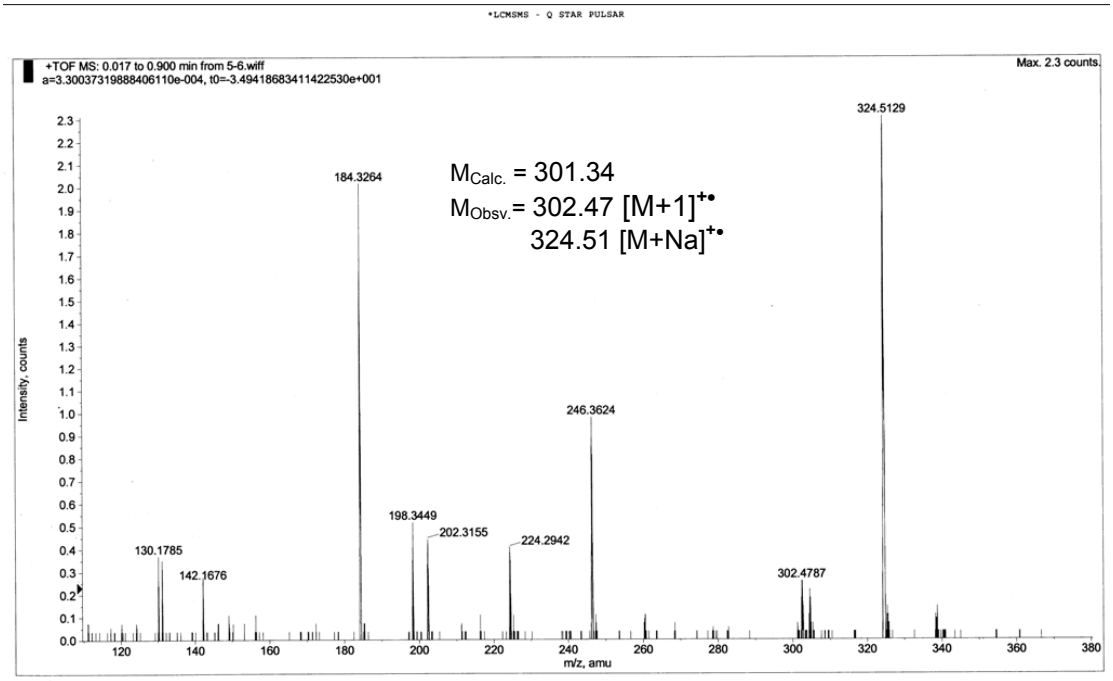
Variable pH ^{1}H NMR of compound 36



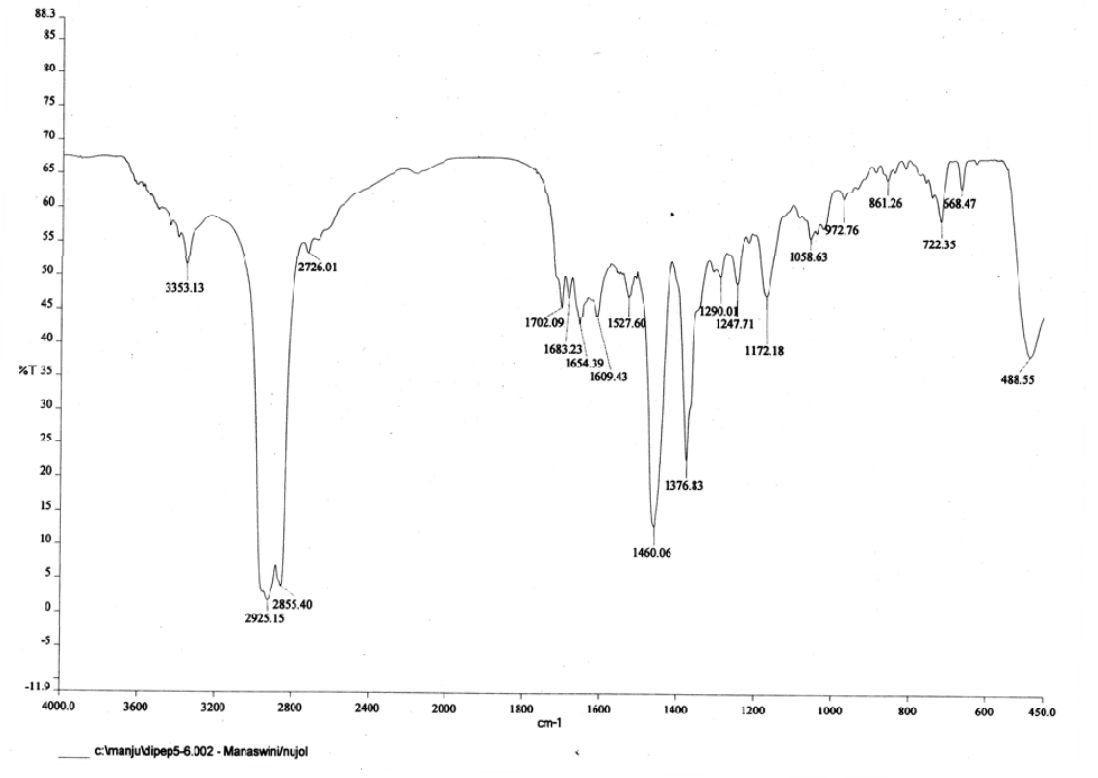
ESI-MS of compound 36



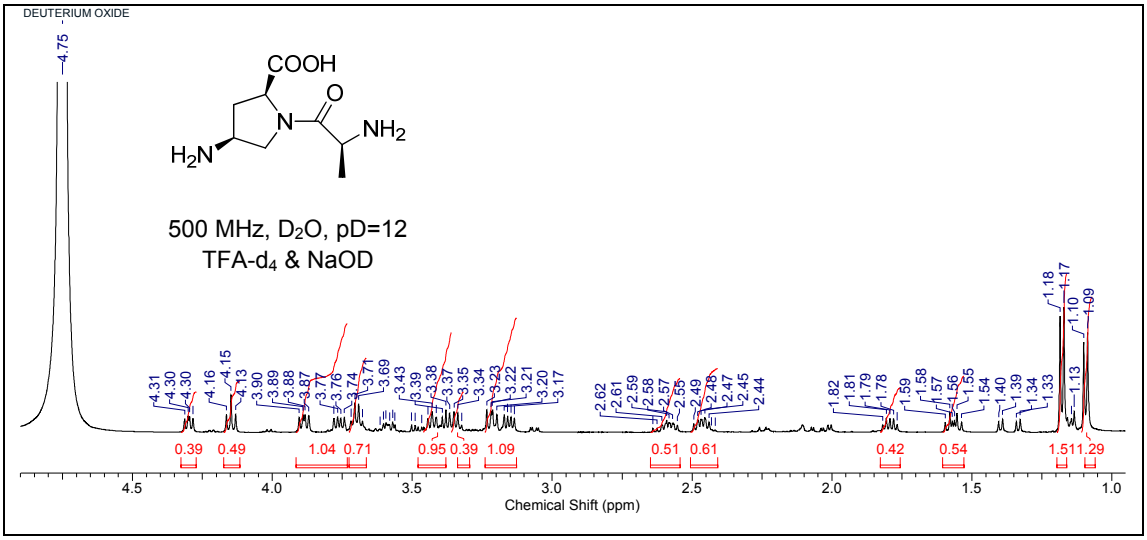
ESI-MS of compound 37



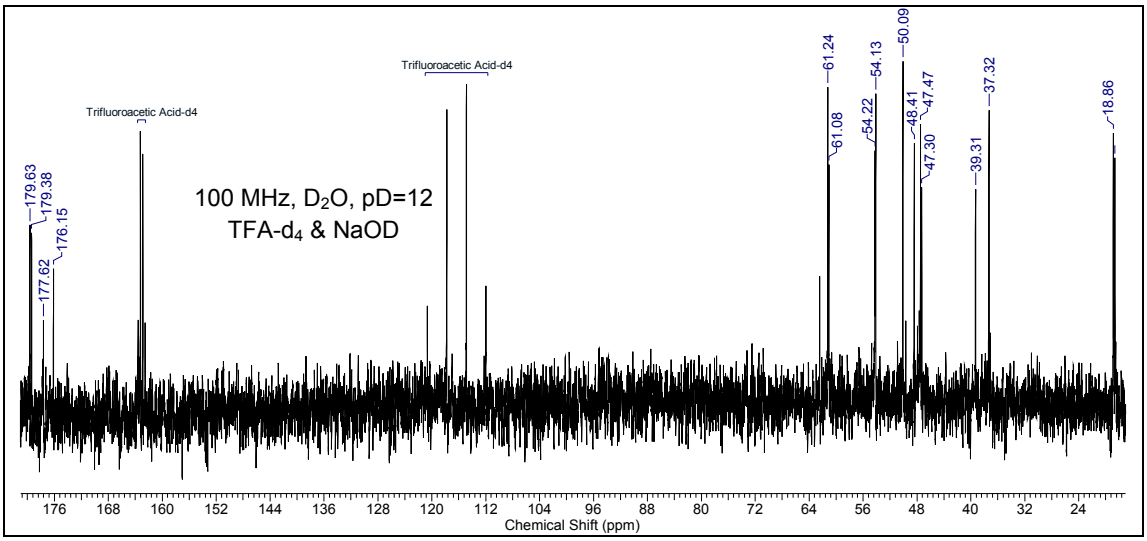
IR Spectra of compound 37



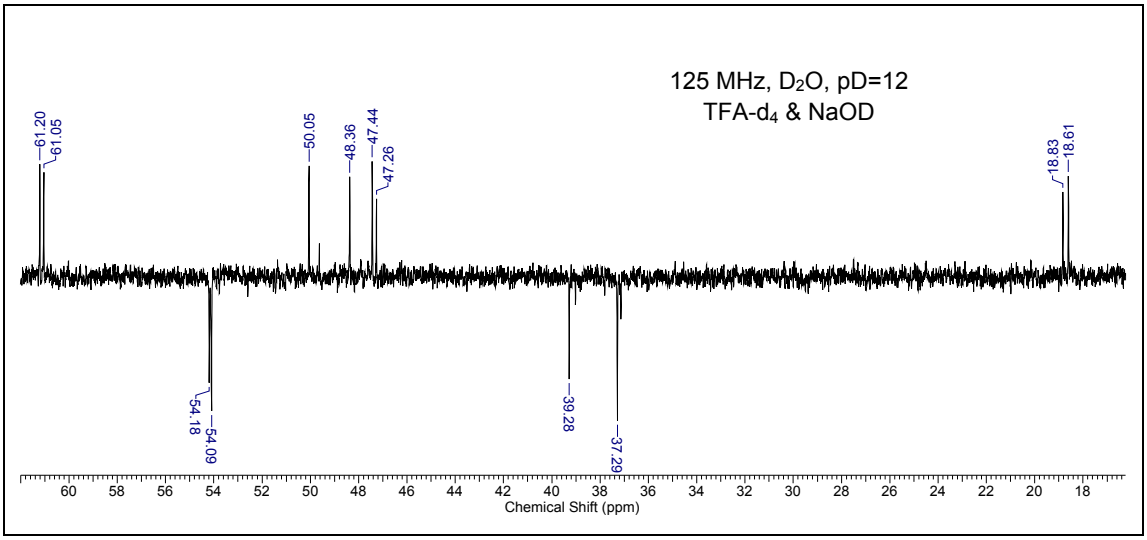
^{1}H NMR of compound 38



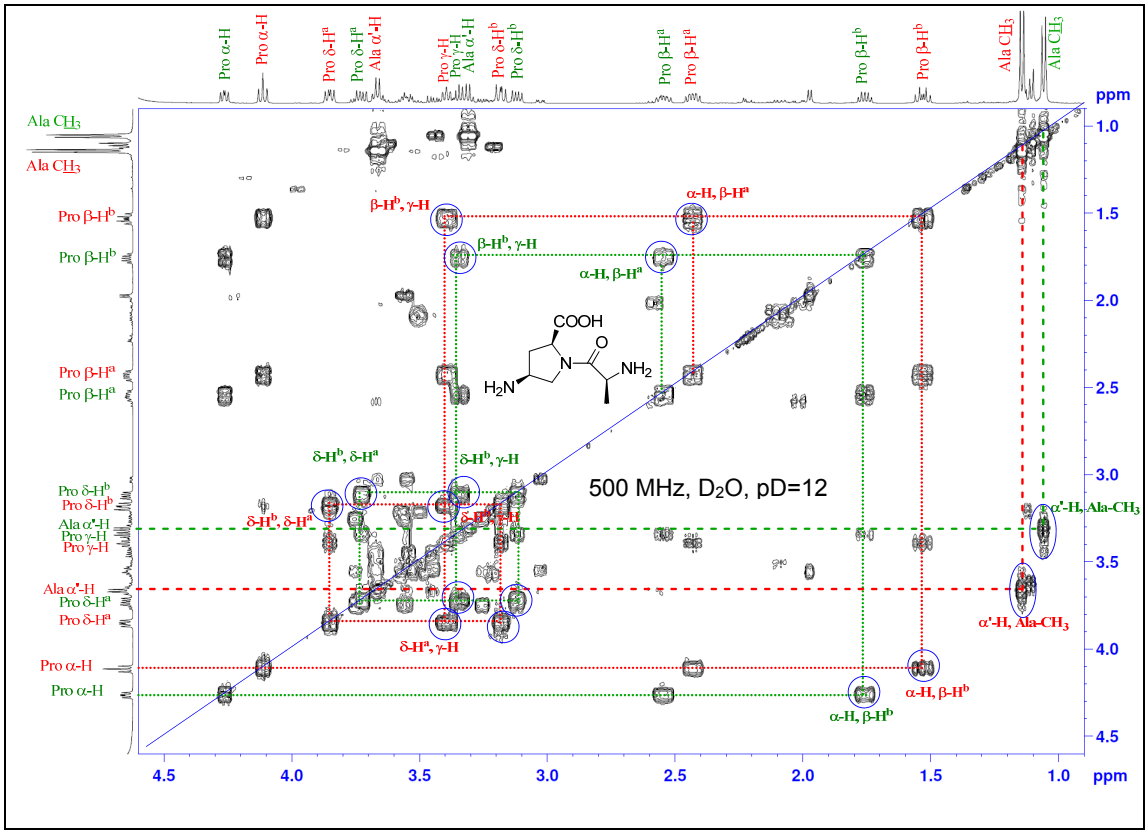
^{13}C NMR of compound 38



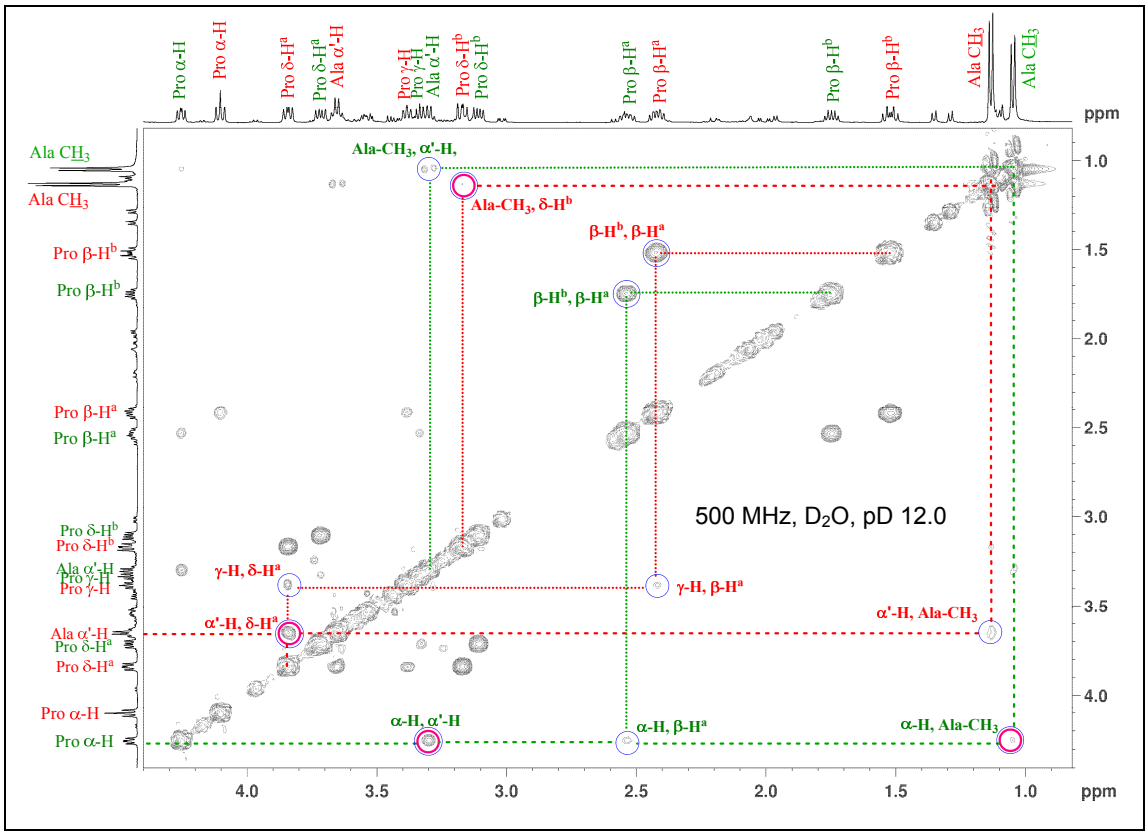
DEPT-^{13}C NMR of compound 38



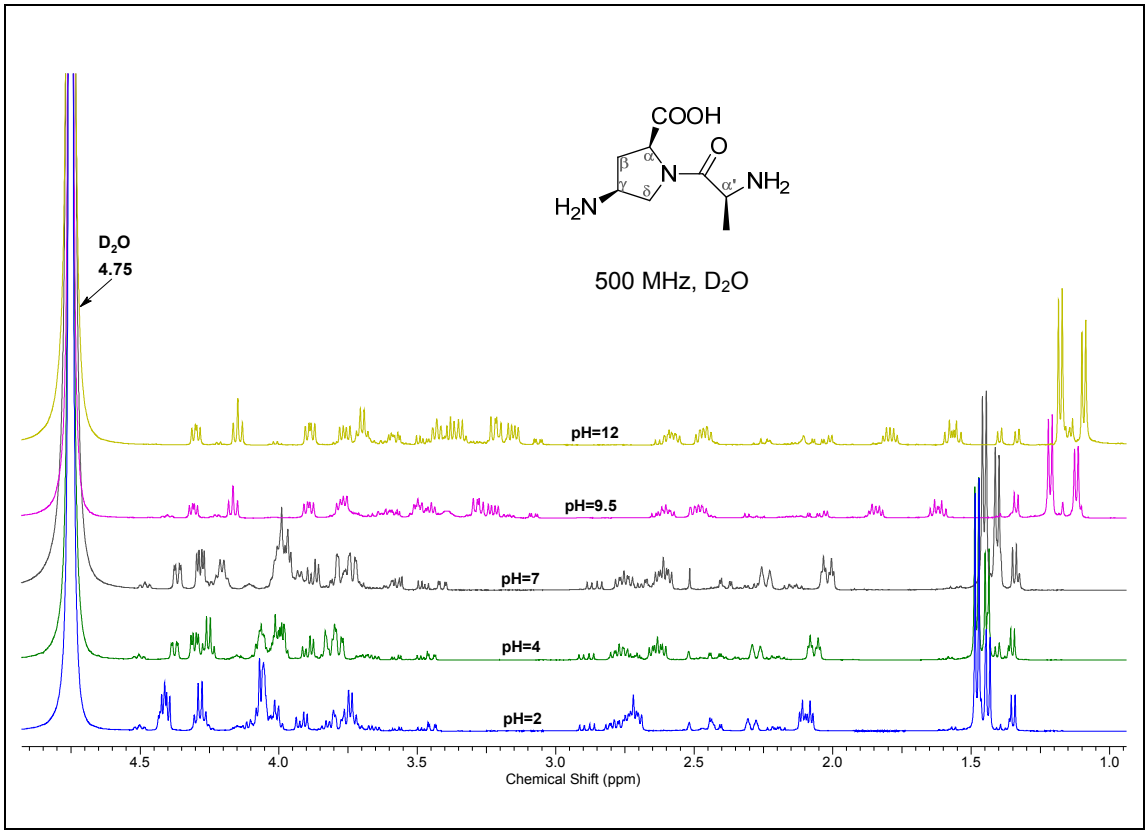
^{1}H-^{1}H COSY of compound 38



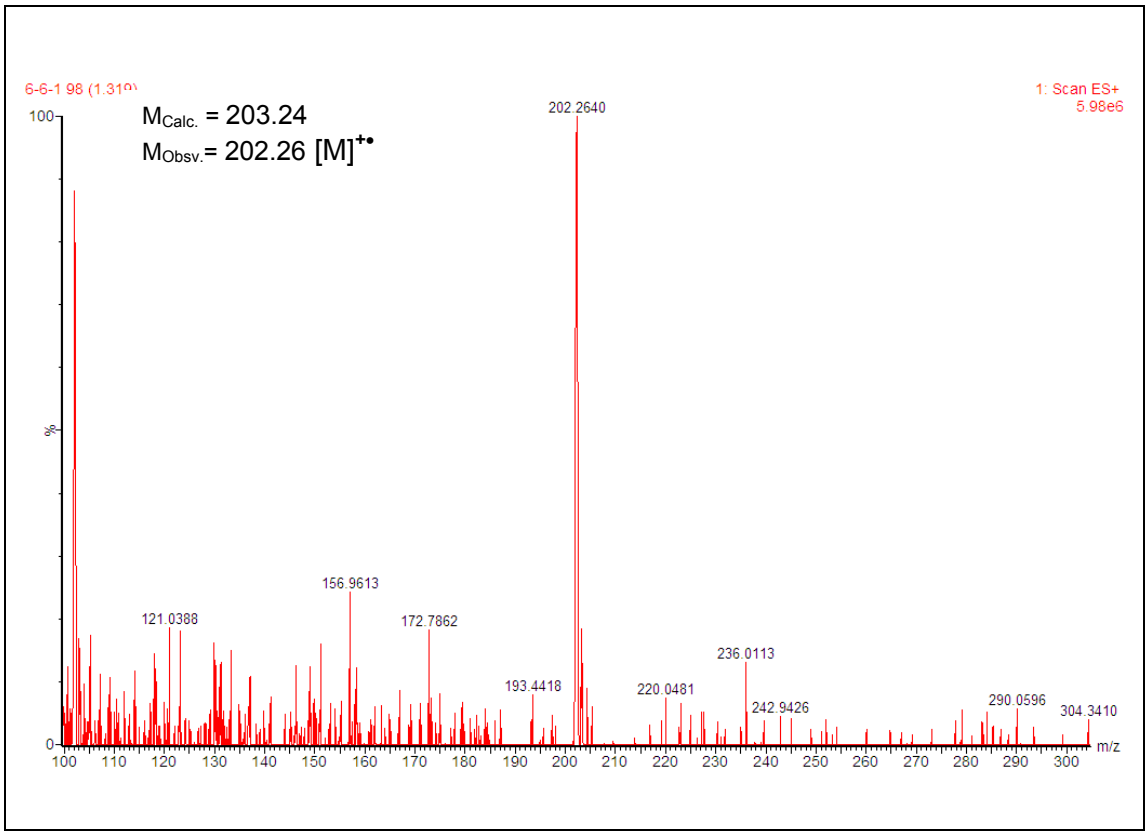
^{1}H-^{1}H NOESY of compound 38



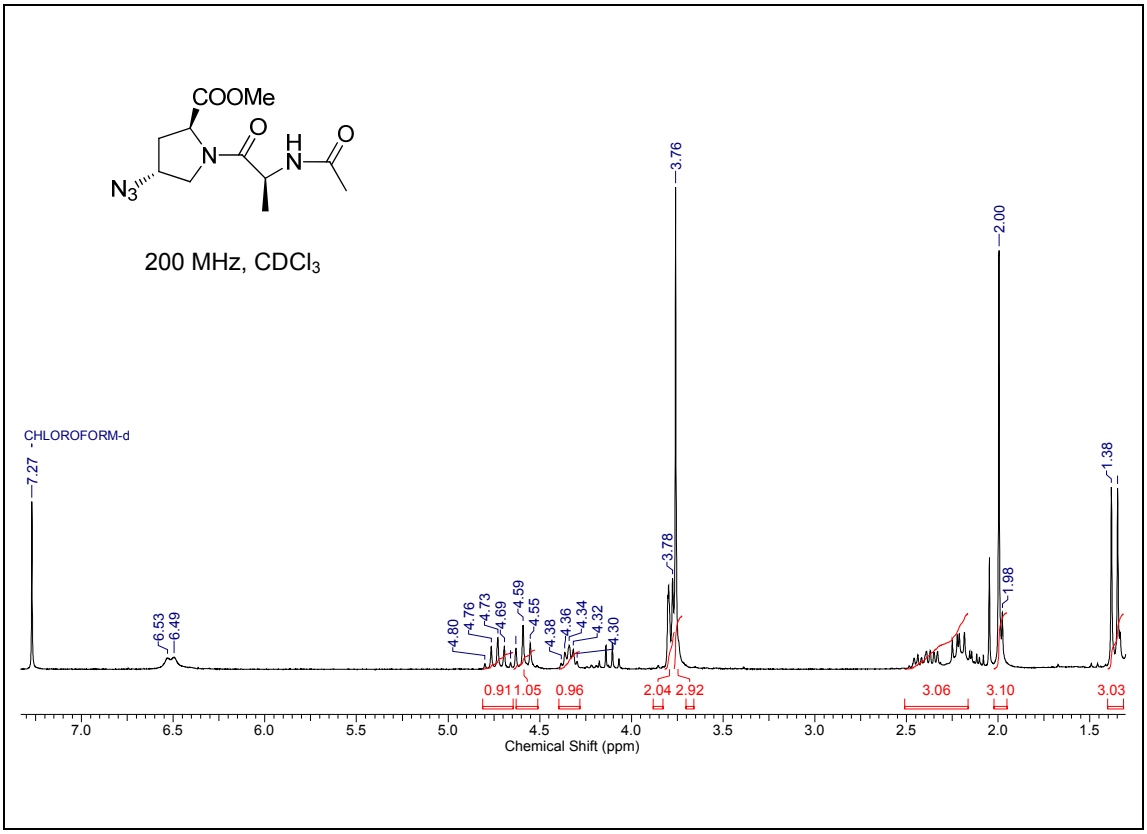
Variable pH ^{1}H NMR of compound 38



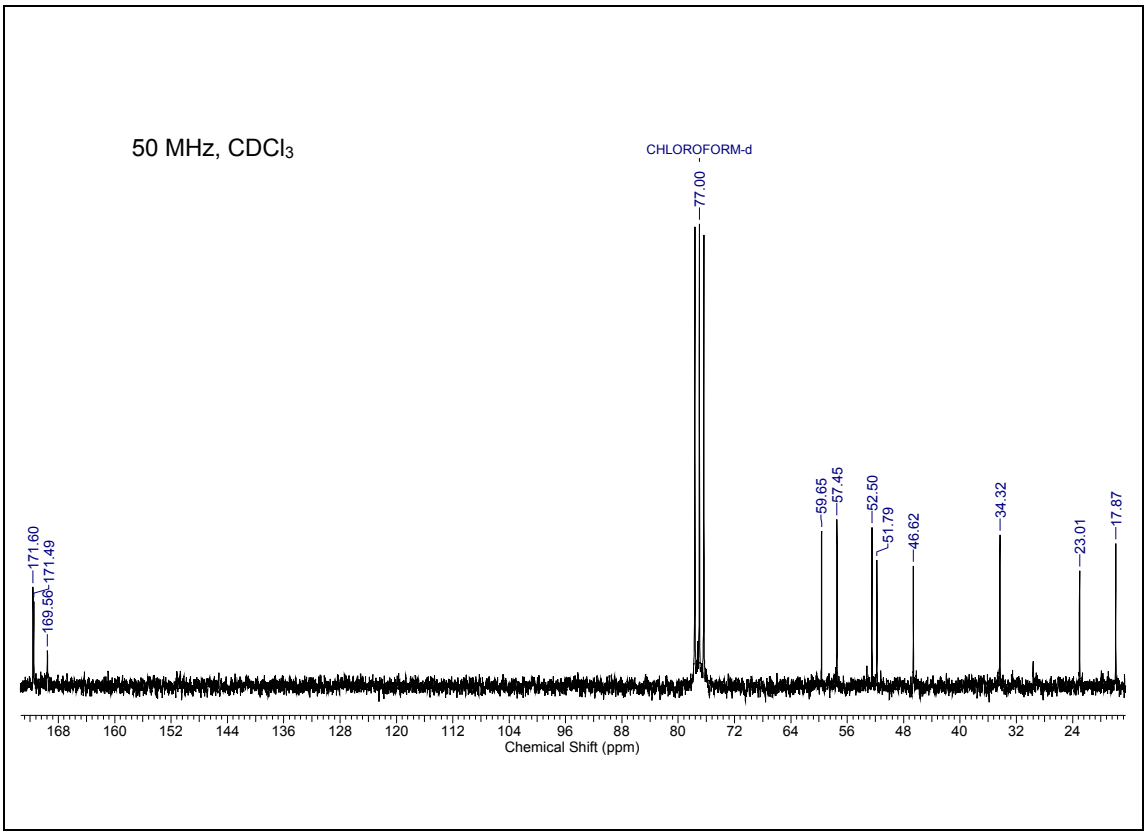
ESI-MS of compound 38



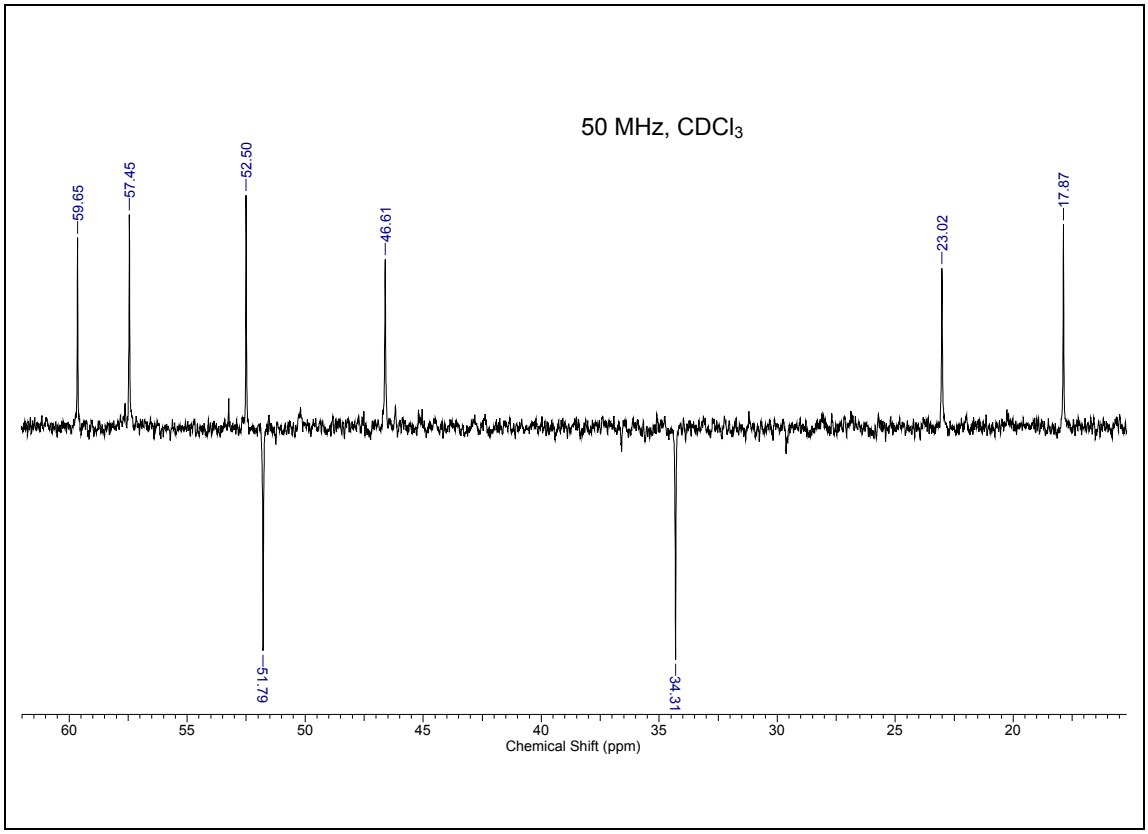
^{1}H NMR of compound 39



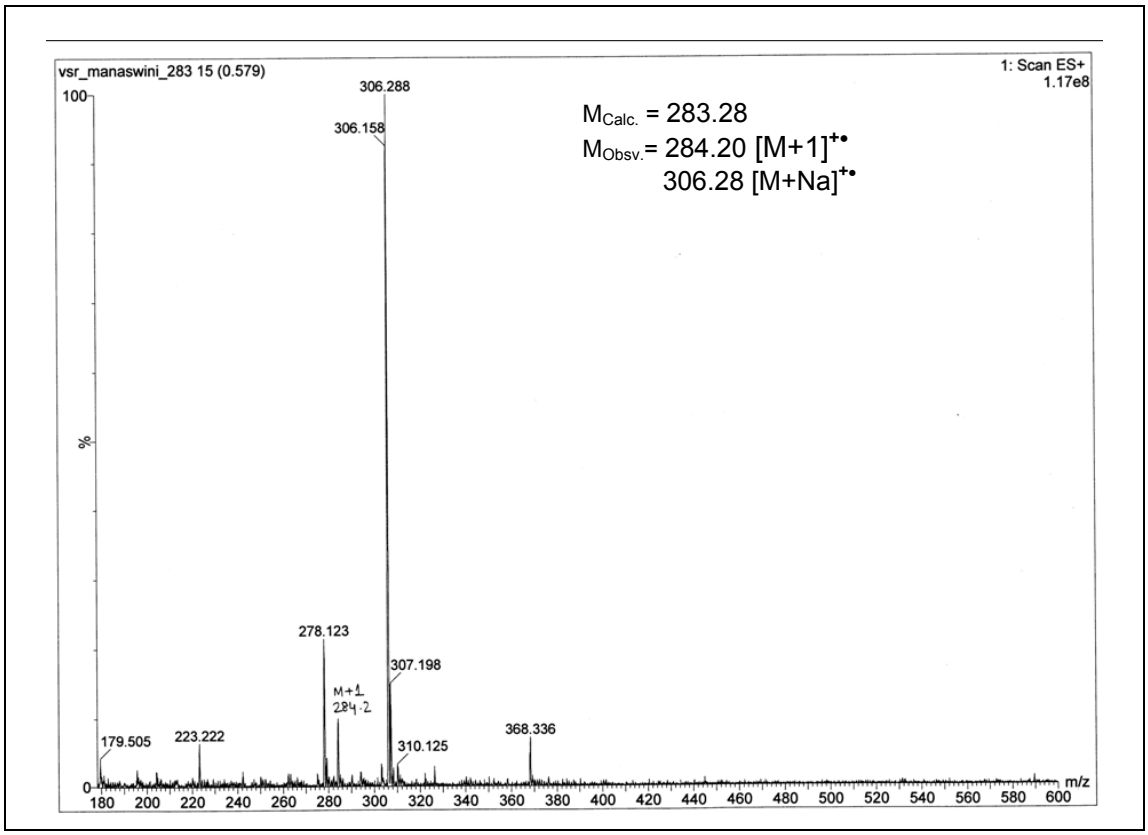
^{13}C NMR of compound 39



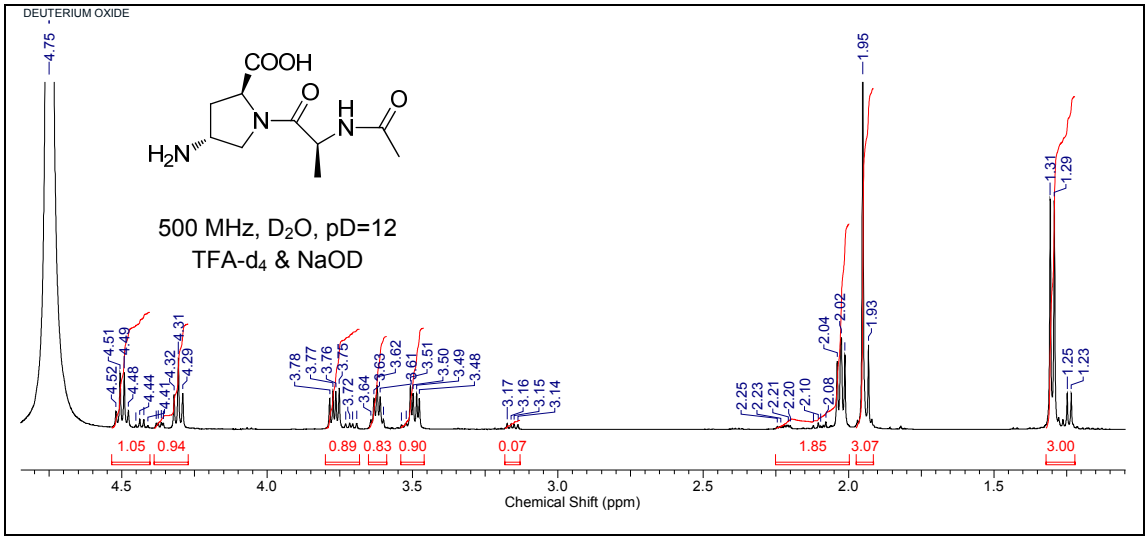
DEPT-^{13}C NMR of compound 39



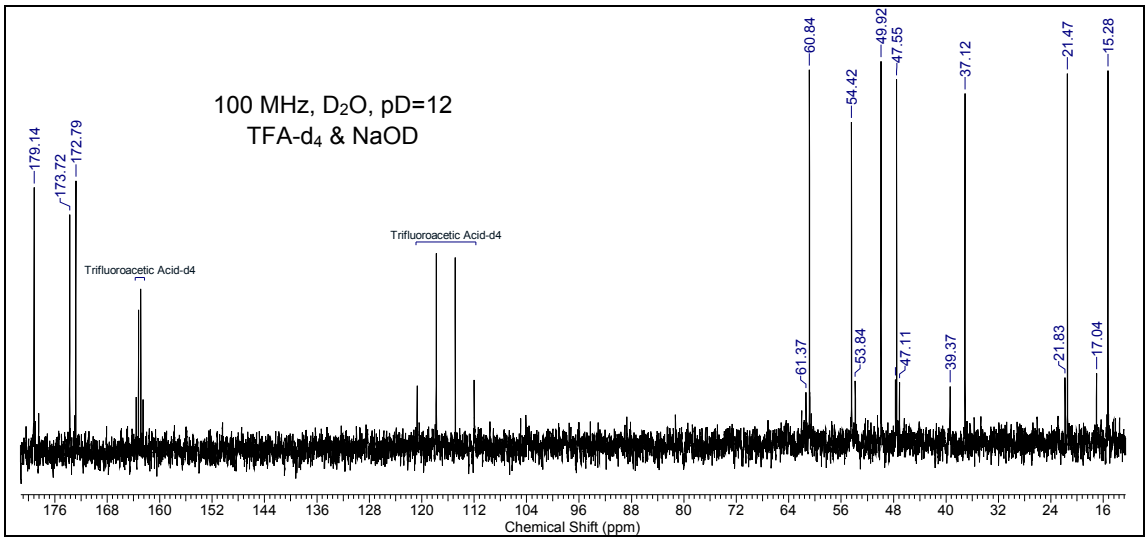
ESI-MS of compound 39



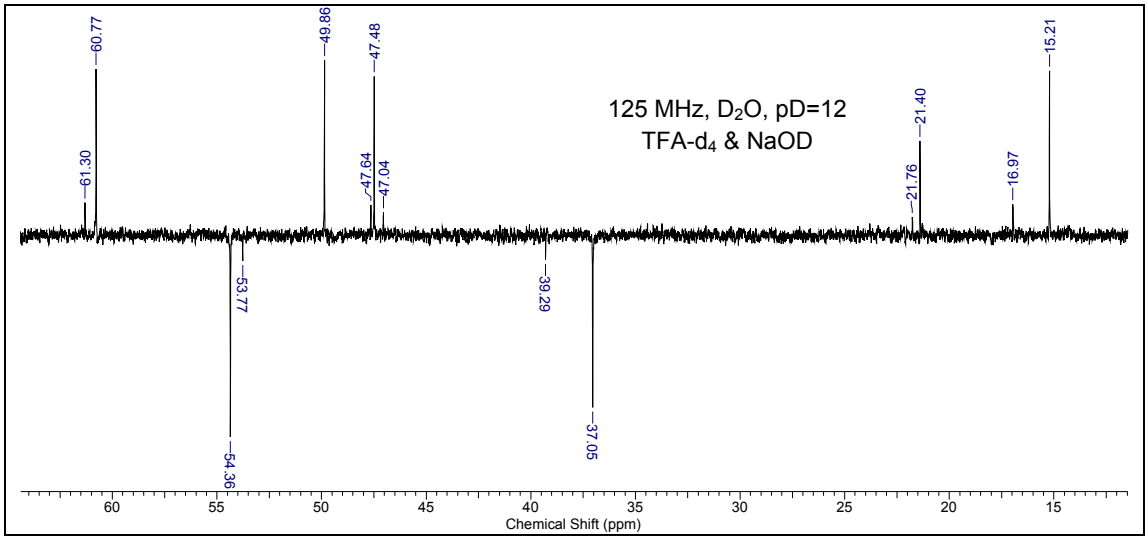
^{1}H NMR of compound 41



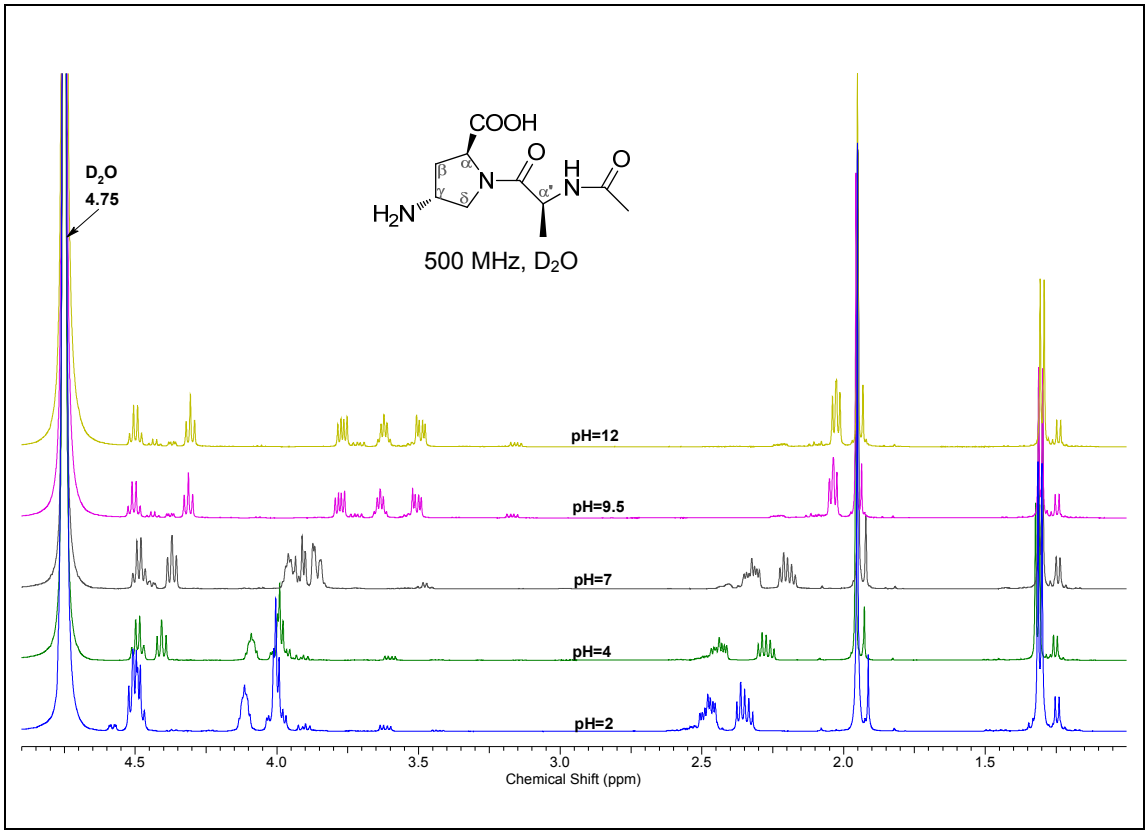
^{13}C NMR of compound 41



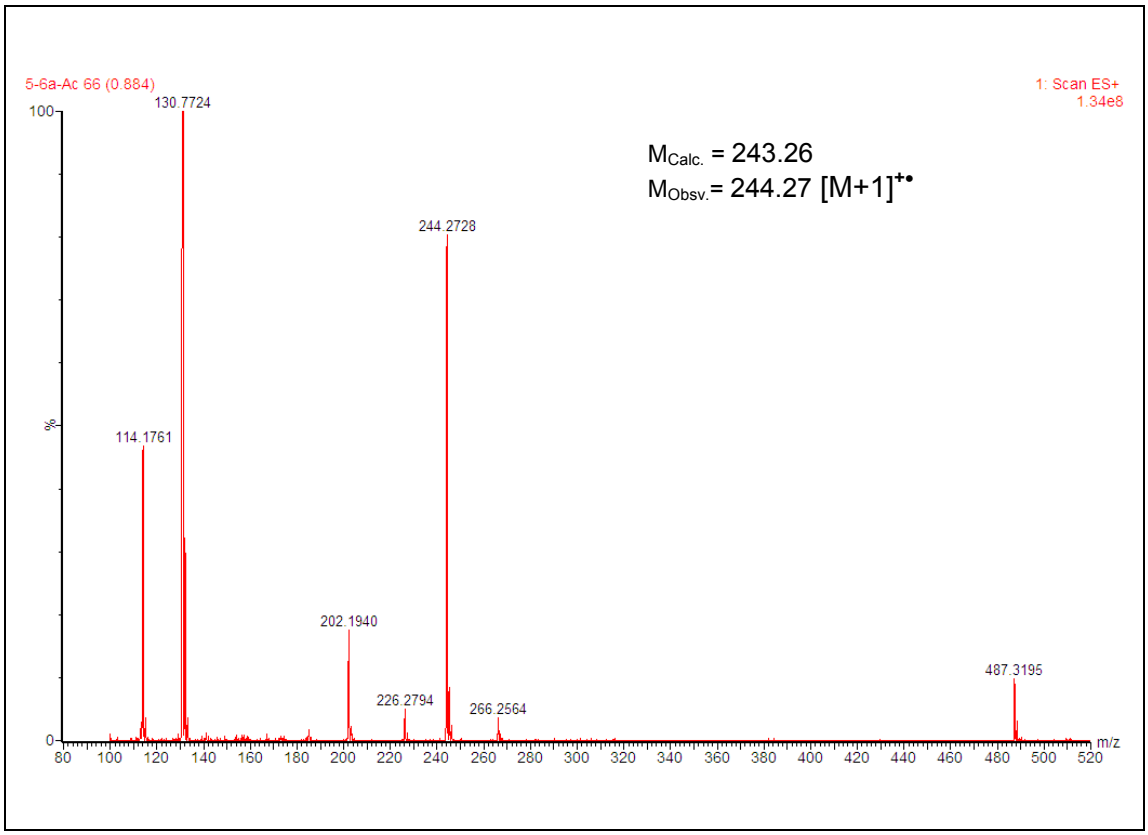
DEPT-^{13}C NMR of compound 41



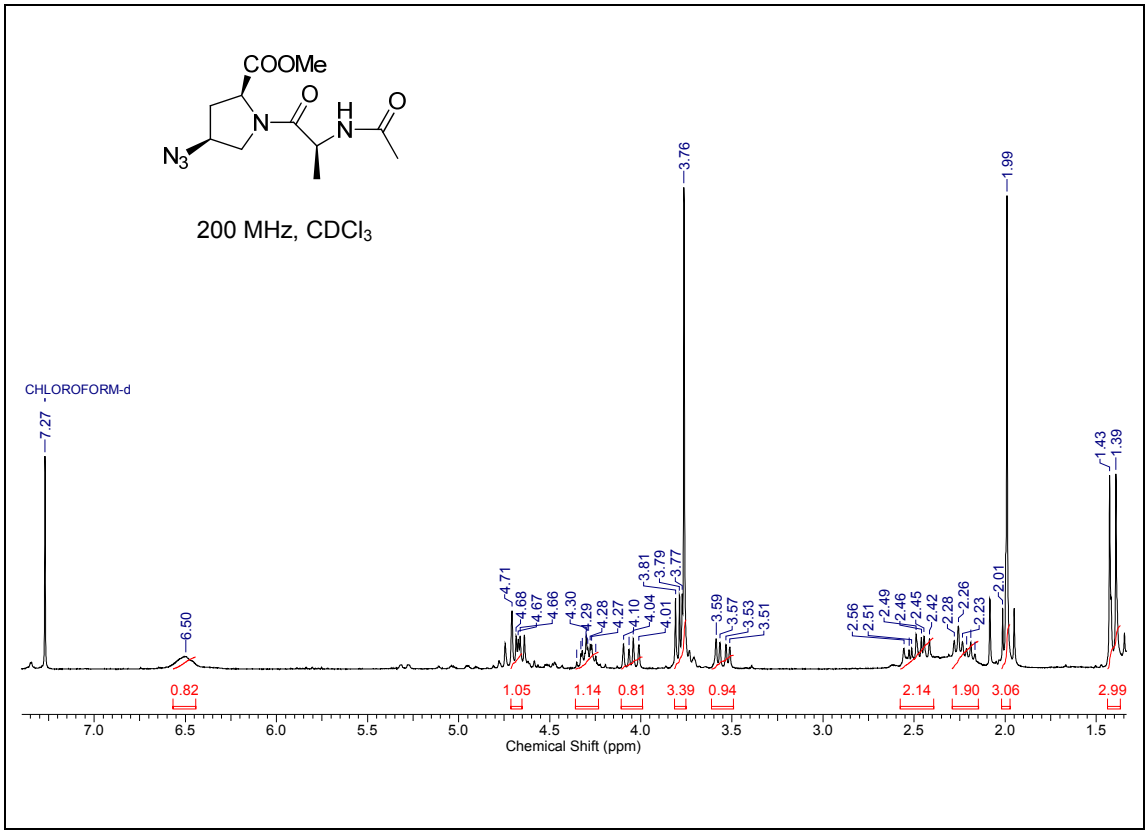
Variable pH ^{1}H NMR of compound 41



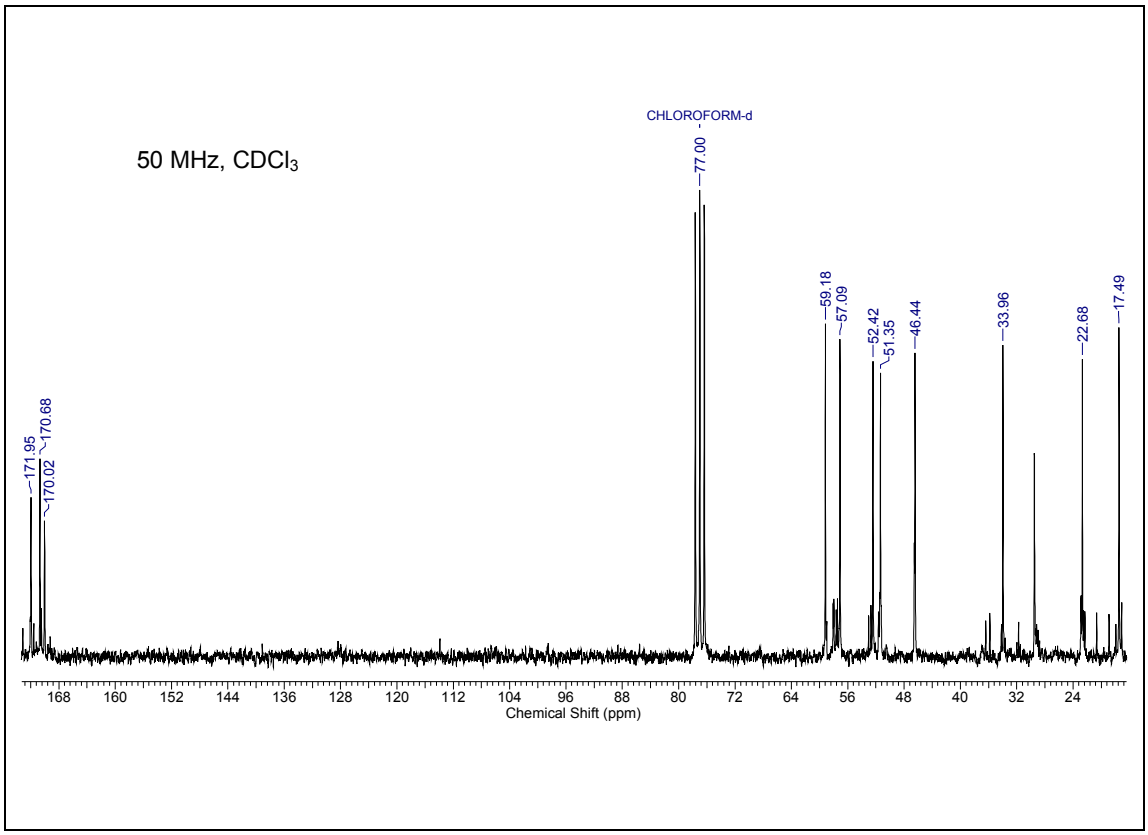
ESI-MS of compound 41



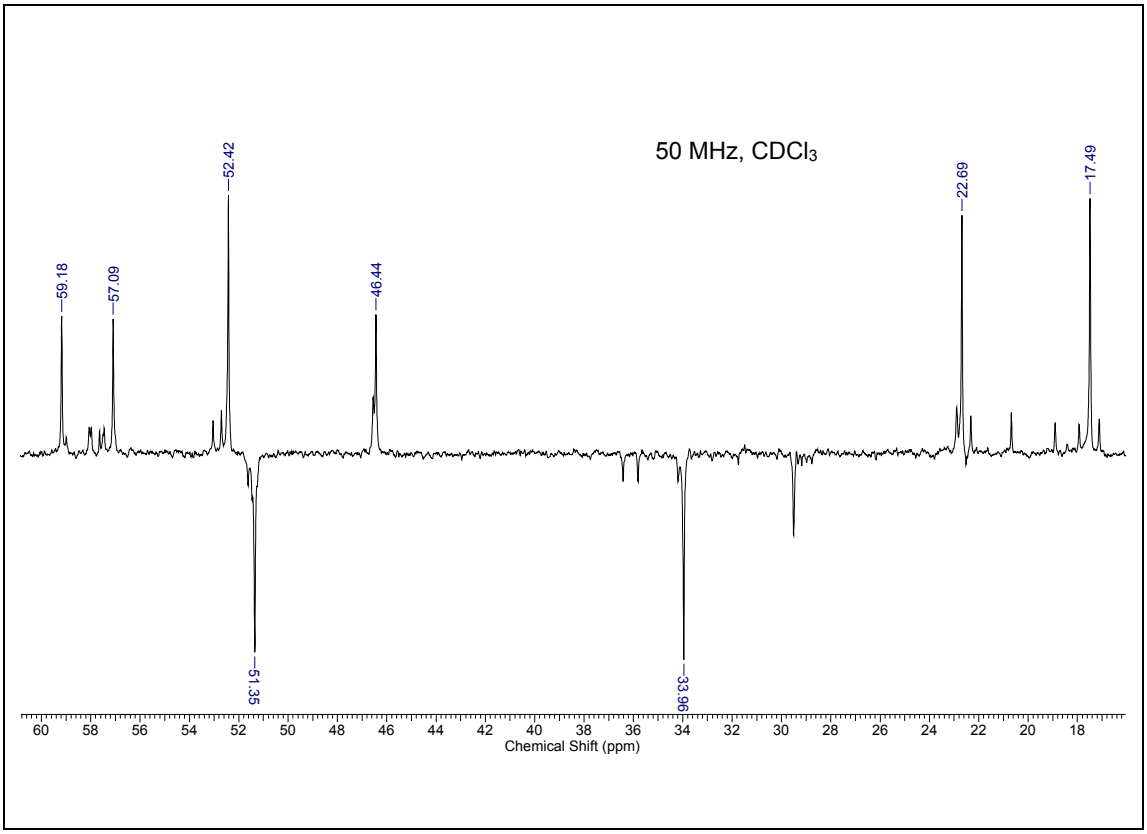
^{1}H NMR of compound 42



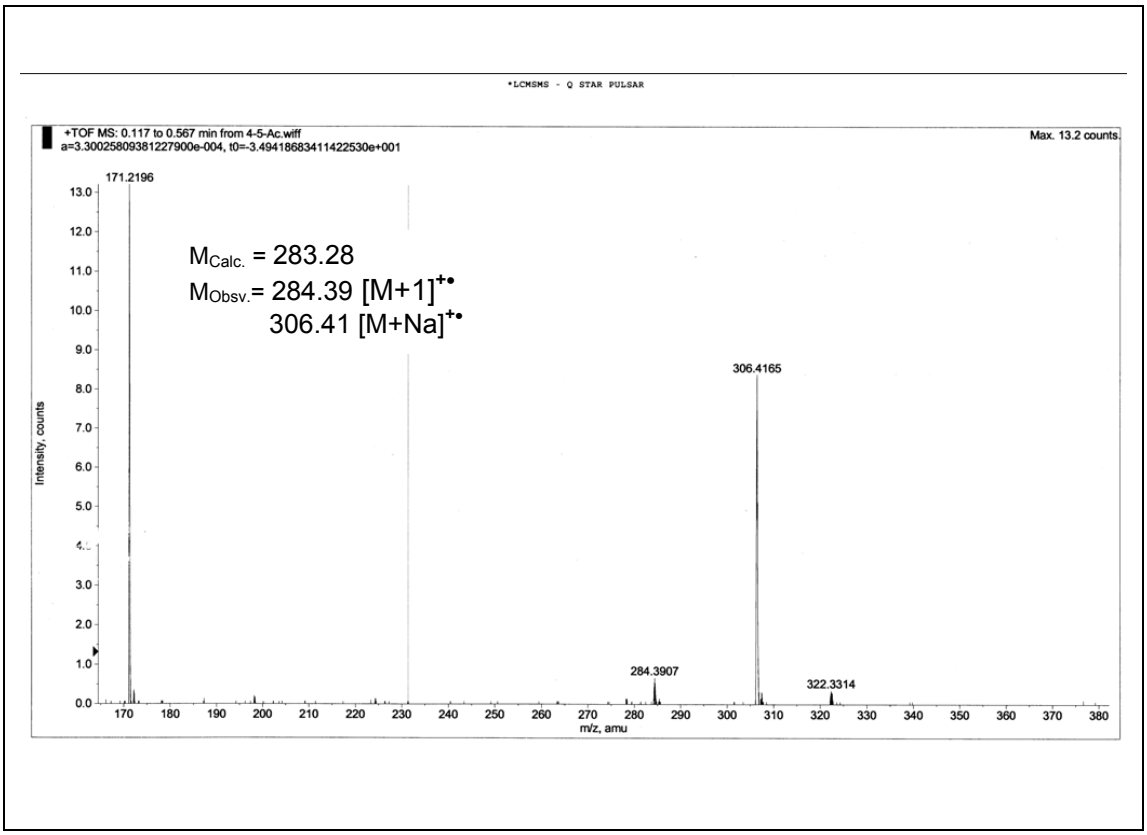
^{13}C NMR of compound 42



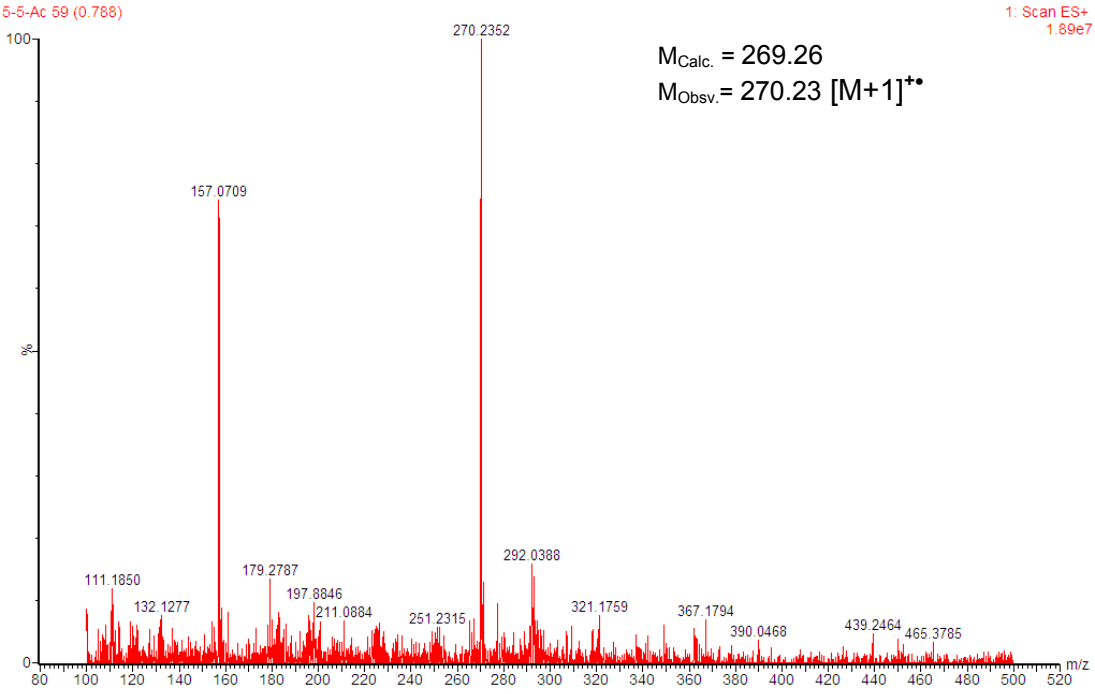
DEPT-^{13}C NMR of compound 42



ESI-MS of compound 42



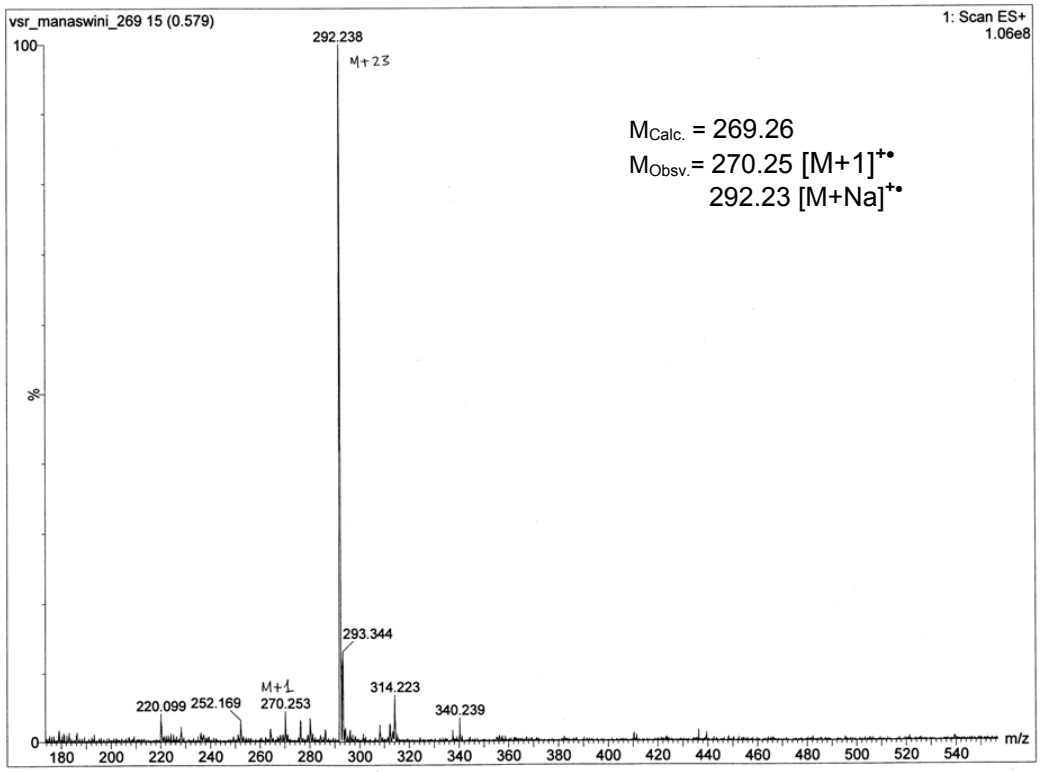
ESI-MS of compound 43



$M_{Calc.}$ = 269.26

$M_{Obsv.}$ = 270.23 $[M+1]^{+\bullet}$

ESI-MS of compound 40

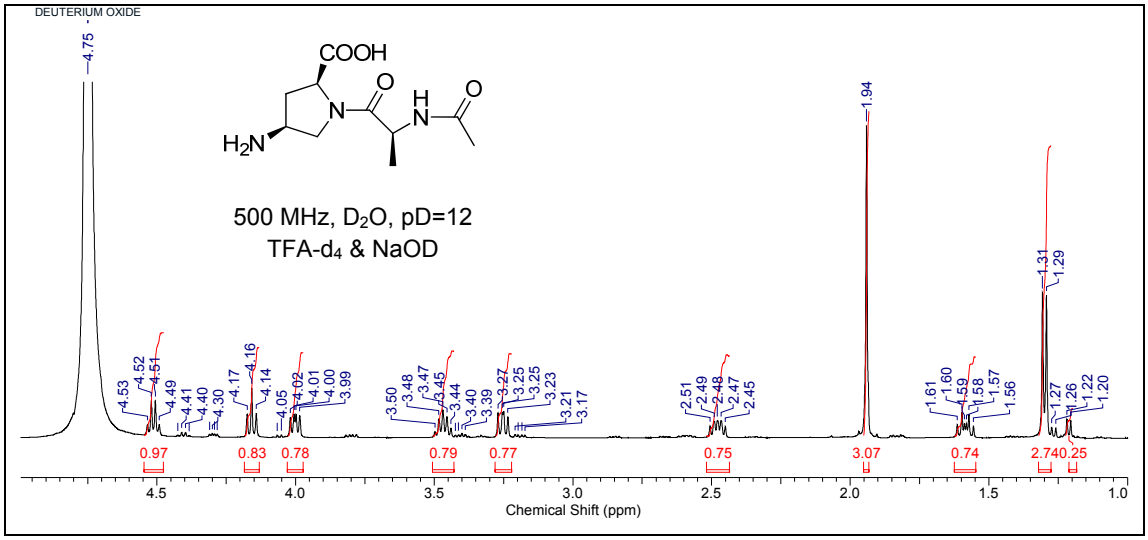


$M_{Calc.}$ = 269.26

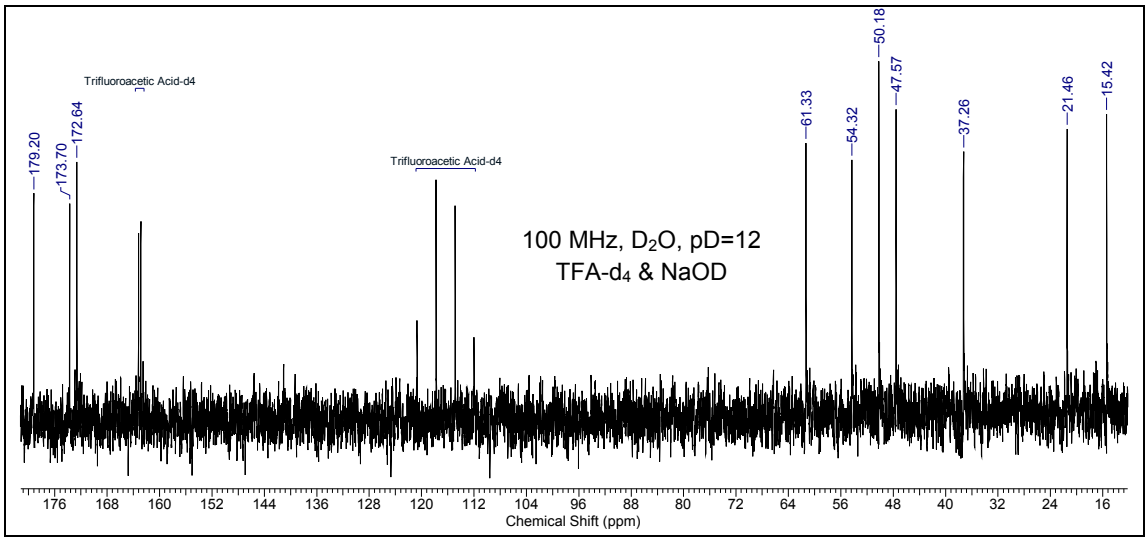
$M_{Obsv.}$ = 270.25 $[M+1]^{+\bullet}$

292.23 $[M+Na]^{+\bullet}$

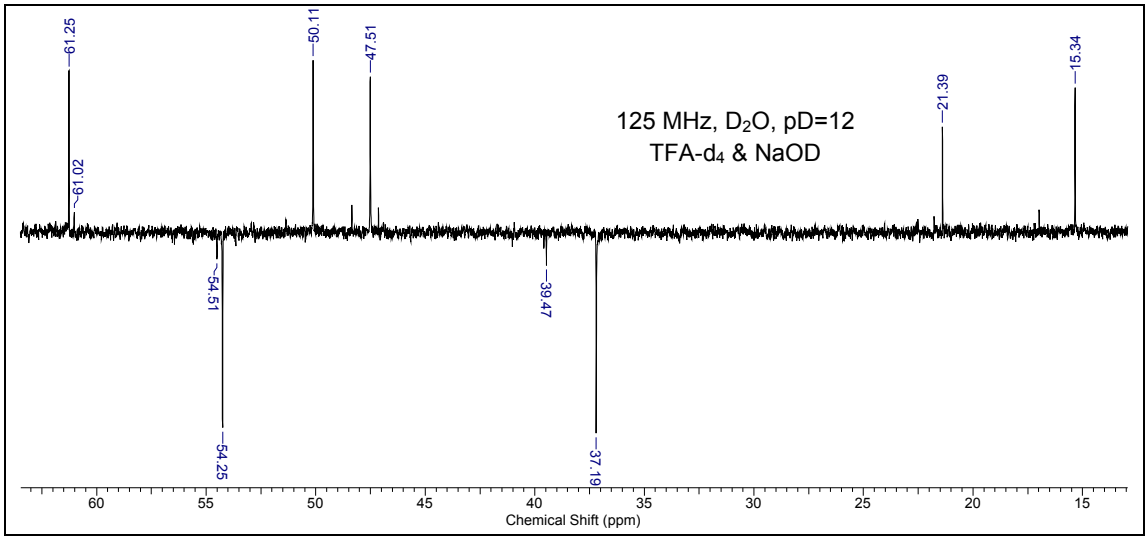
^{1}H NMR of compound 44



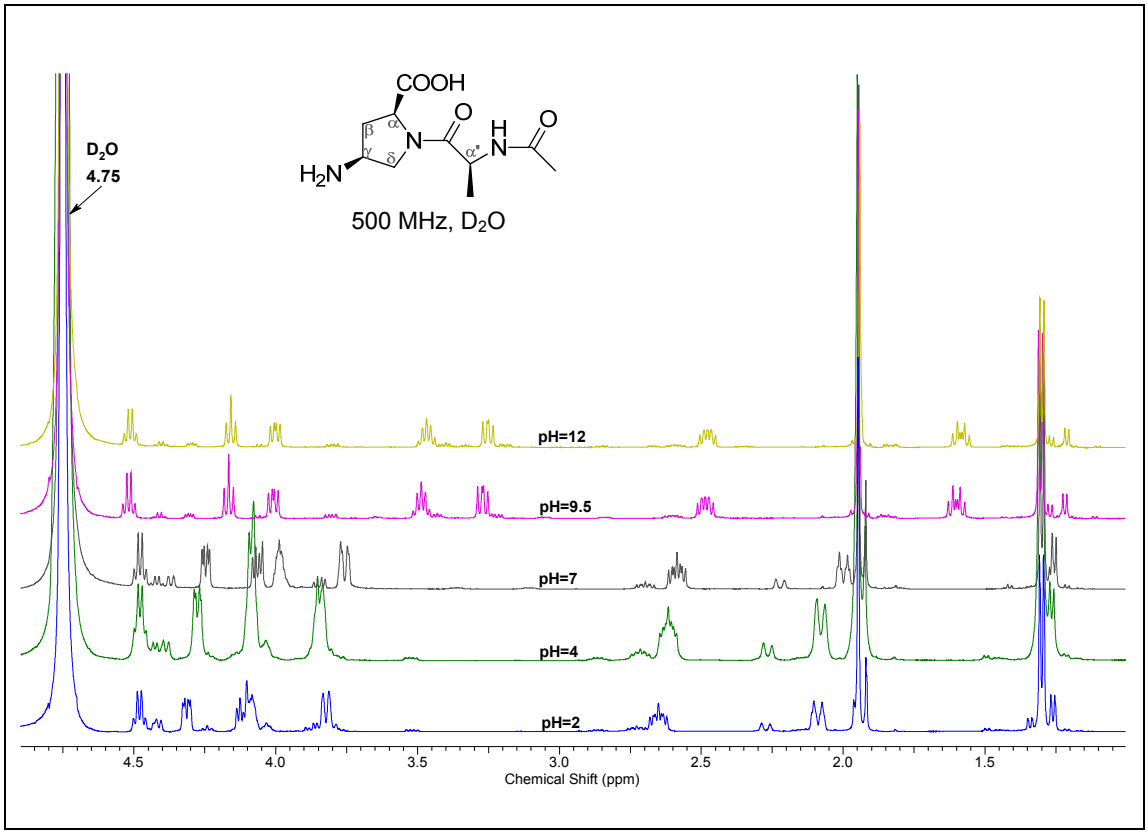
^{13}C NMR of compound 44



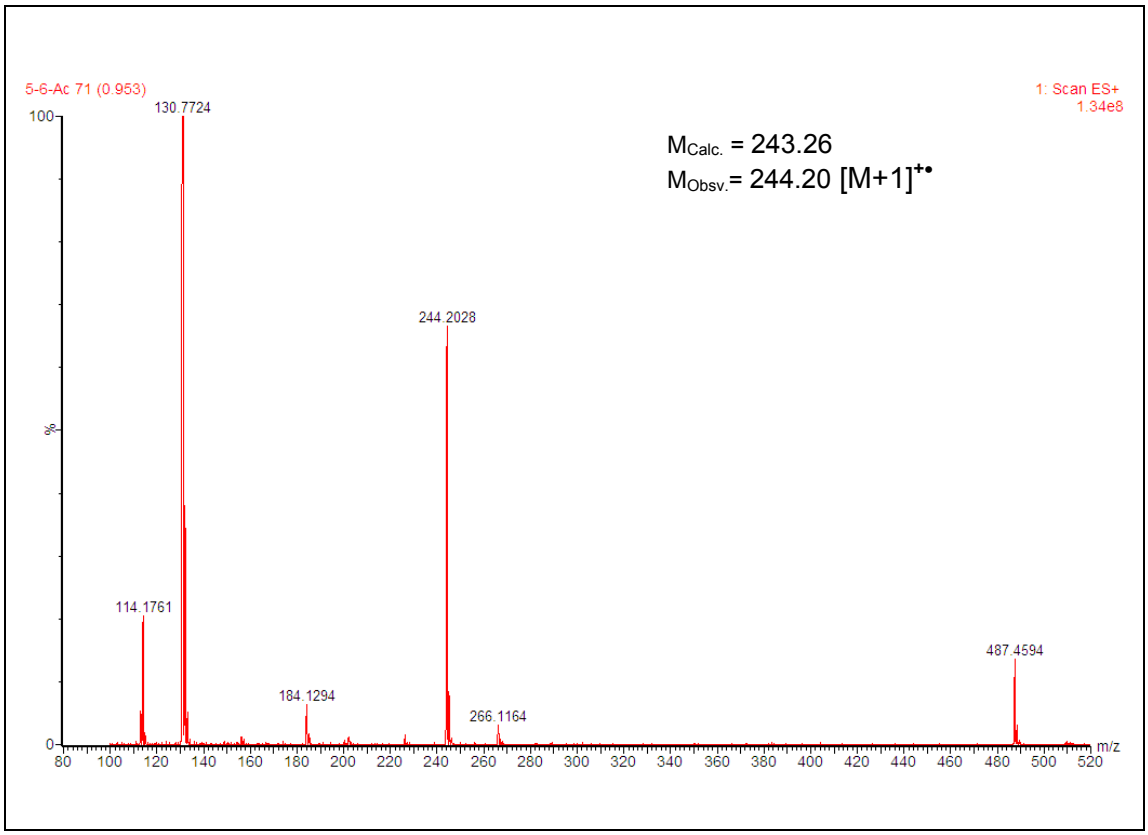
DEPT-^{13}C NMR of compound 44



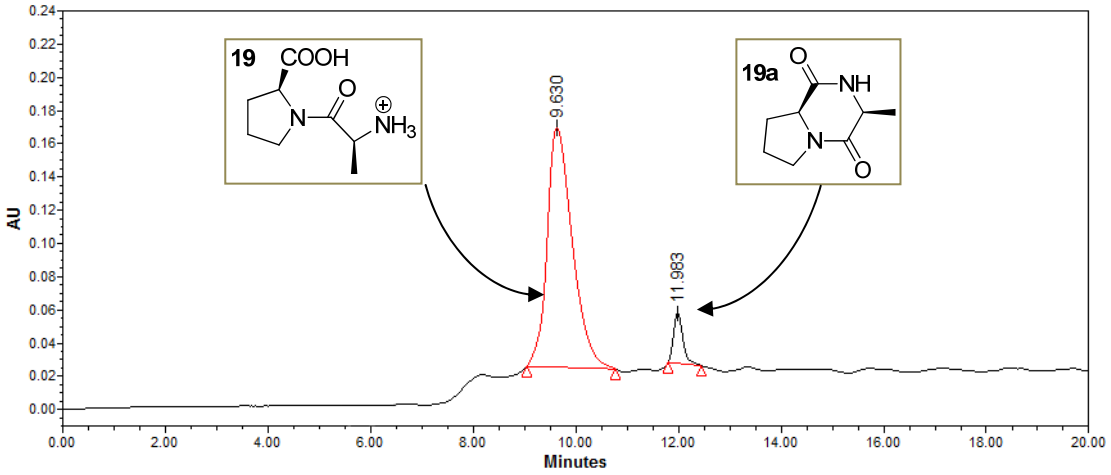
Variable pH ^{1}H NMR of compound 44



ESI-MS of compound 44



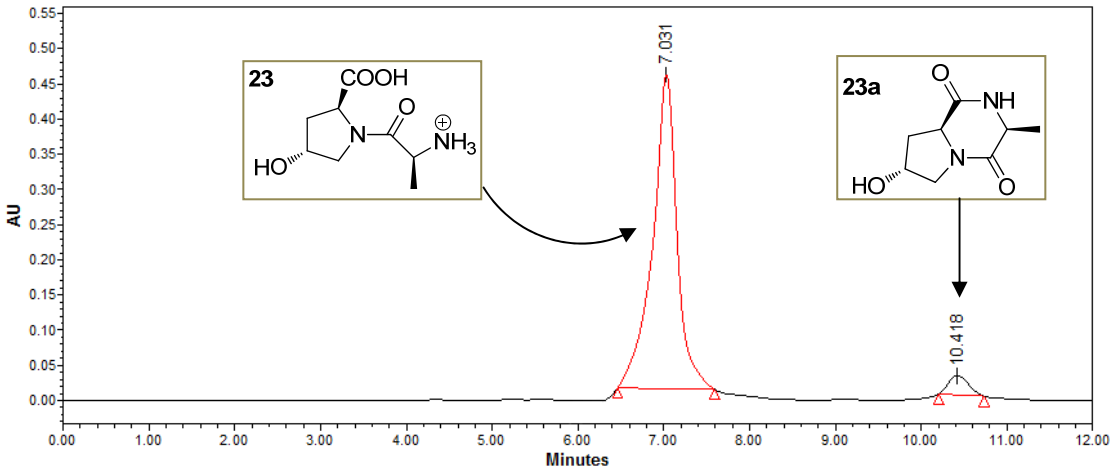
HPLC of Compound 19



Compound	t_R (min)	% Area	Height
19	**9.630**	**92.77**	**144368**
19a	11.983	7.23	30171

20 min Gradient: 5% to 80% MeCN/H_2O with 0.1% TFA, flow rate = 3 ml/min; λ = 220 nm

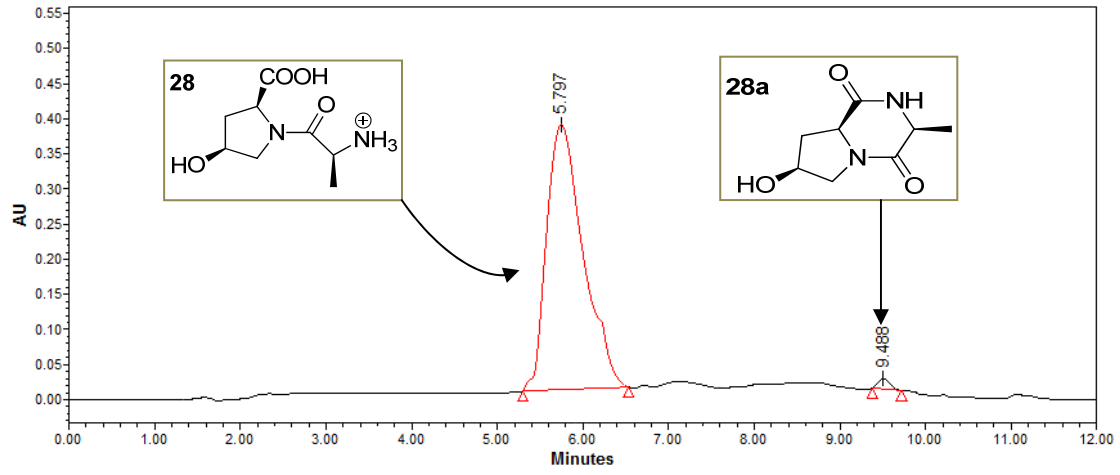
HPLC of Compound 23



Compound	t_R (min)	% Area	Height
23	**7.031**	**95.40**	**451026**
23a	10.418	4.60	27330

12 min Isocratic: 3% MeCN/H_2O with 0.1% TFA, flow rate = 3 ml/min; λ = 220 nm

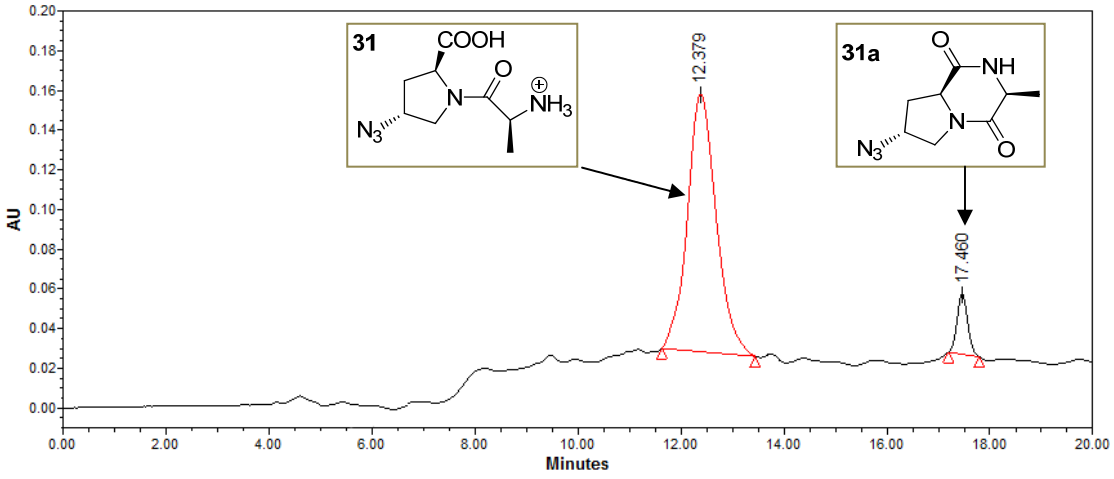
HPLC of Compound 28



Compound	t_R (min)	% Area	Height
28	**5.797**	**98.77**	**746555**
28a	9.488	1.23	30068

12 min Isocratic: 3% MeCN/H_2O with 0.1% TFA, flow rate = 3 ml/min; λ = 220 nm

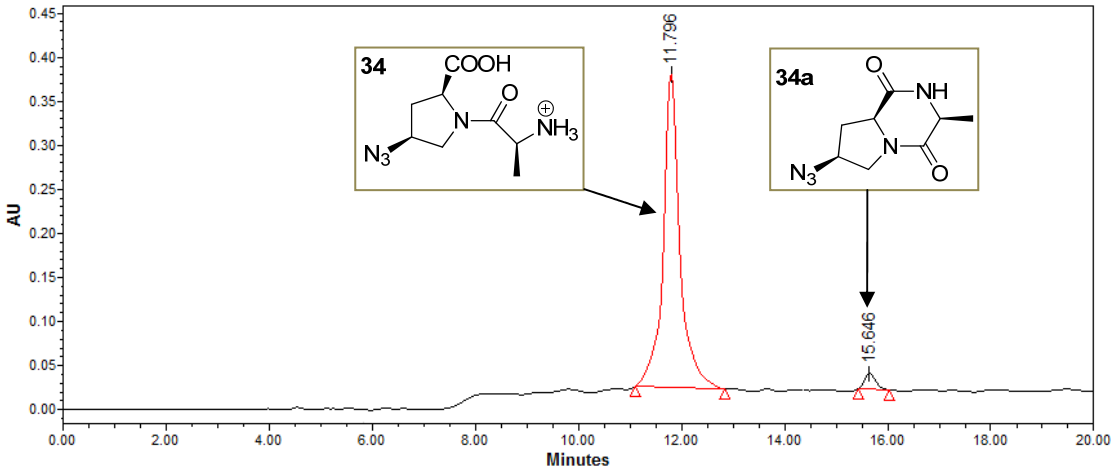
HPLC of Compound 31



Compound	t_R (min)	% Area	Height
31	**12.379**	**91.93**	**129433**
31a	17.460	8.07	30444

20 min Gradient: 5% to 80% MeCN/H_2O with 0.1% TFA, flow rate = 3 ml/min; λ = 220 nm

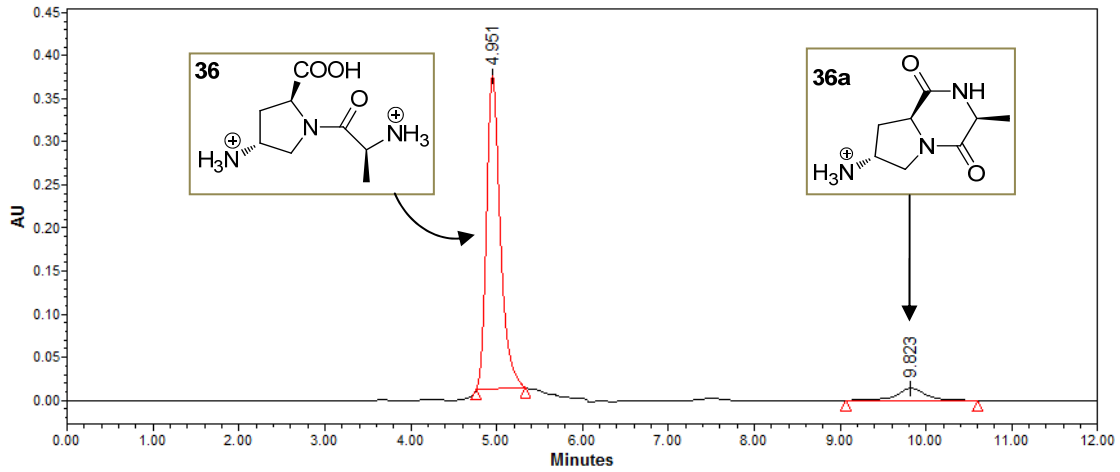
HPLC of Compound 34



Compound	t_R (min)	% Area	Height
34	**11.796**	**96.85**	**355212**
34a	15.646	3.15	17651

20 min Gradient: 5% to 80% MeCN/H_2O with 0.1% TFA, flow rate = 3 ml/min; λ = 220 nm

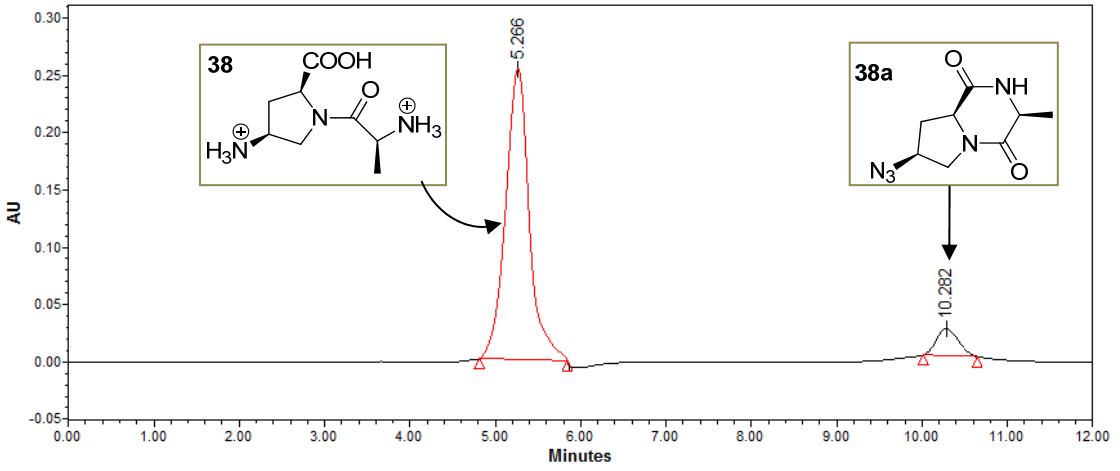
HPLC of Compound 36



Compound	t_R (min)	% Area	Height
36	**4.951**	**96.07**	**350610**
36a	9.823	3.93	10243

12 min Isocratic: H_2O with 0.1% TFA, flow rate = 3 ml/min; λ = 220 nm

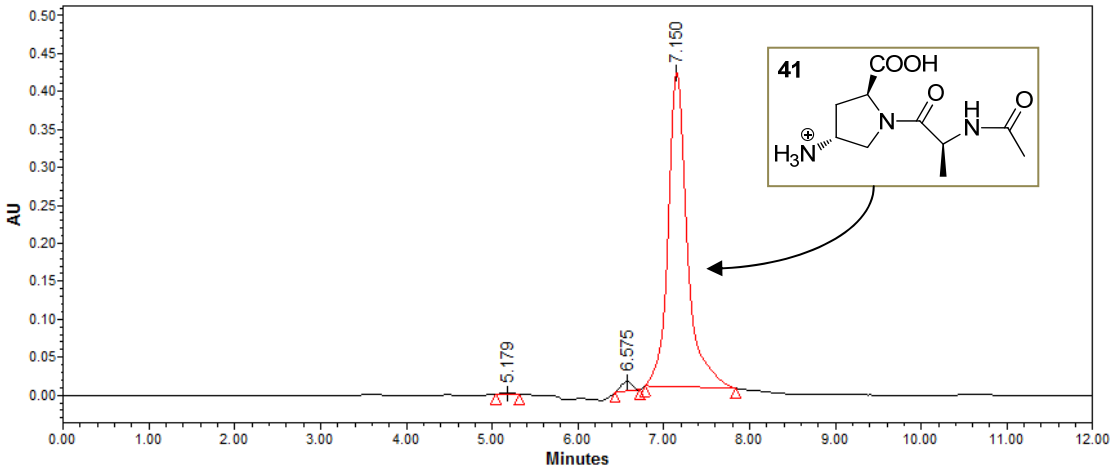
HPLC of Compound 38



Compound	t_R (min)	% Area	Height
38	**5.266**	**92.14**	**253218**
38a	10.282	7.86	23389

12 min Isocratic: H_2O with 0.1% TFA, flow rate = 3 ml/min; λ = 220 nm

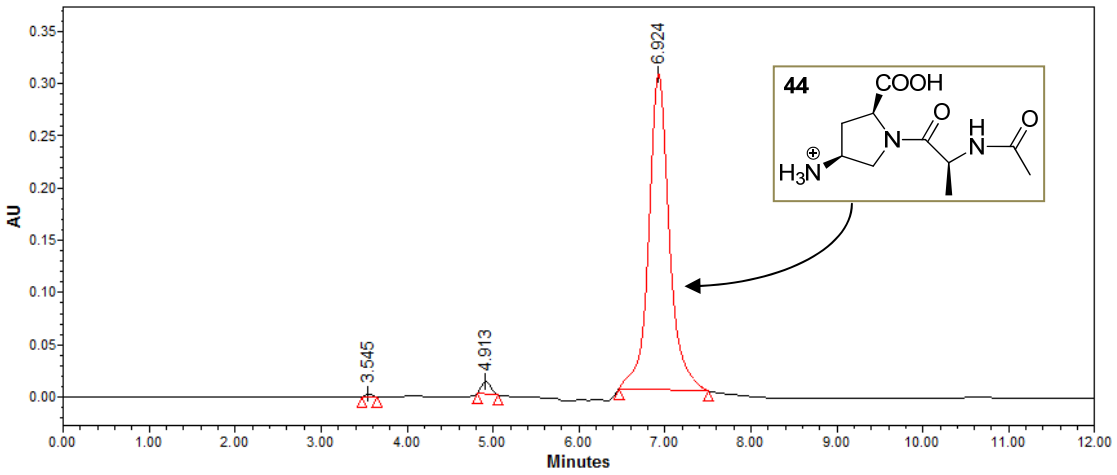
HPLC of Compound 41



Compound	t_R (min)	% Area	Height
	5.179	0.31	2077
	6.575	1.66	12704
41	**7.150**	**98.04**	**413259**

12 min Isocratic: 3% MeCN/H_2O with 0.1% TFA, flow rate = 3 ml/min; λ = 220 nm

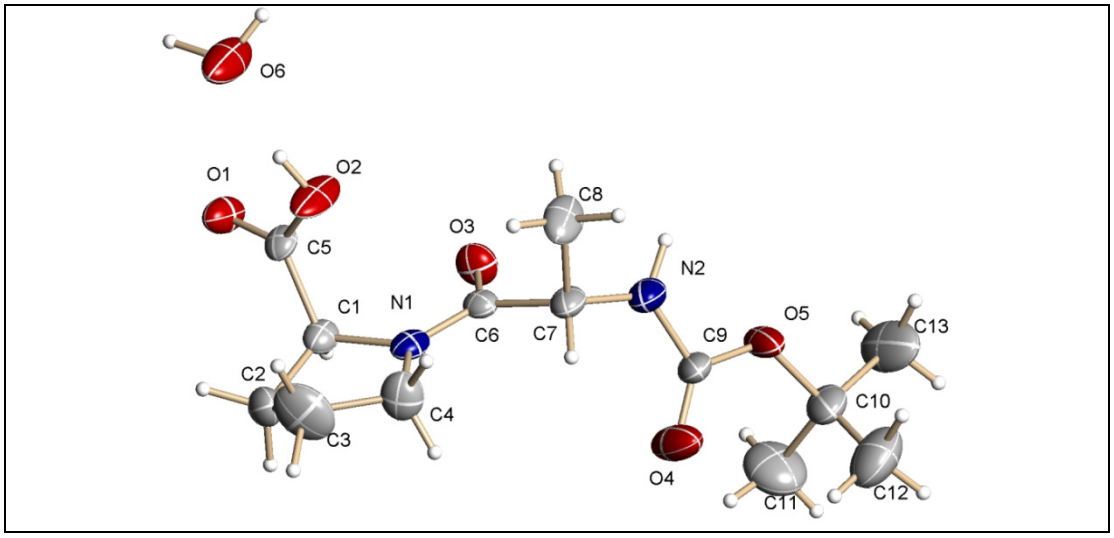
HPLC of Compound 44



Compound	t_R	% Area	Height
	3.545	0.34	3115
	4.913	1.71	11950
44	**6.924**	**97.95**	**302088**

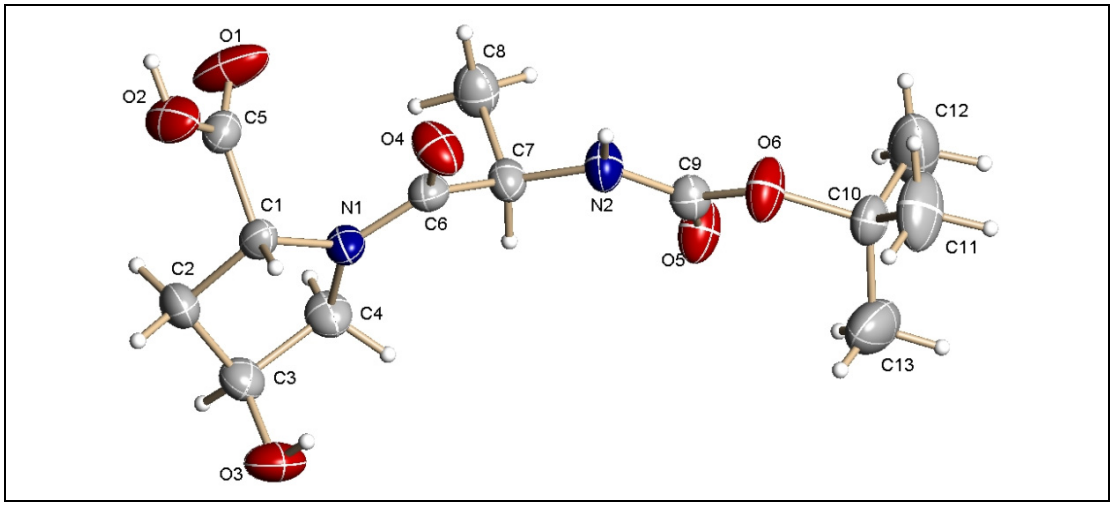
12 min Isocrc: 3% MeCN/H_2O with 0.1% TFA, flow rate = 3 ml/min; λ = 220 nm

Table 6: ORTEP diagram and Crystal data with structure refinement for compound **18**:



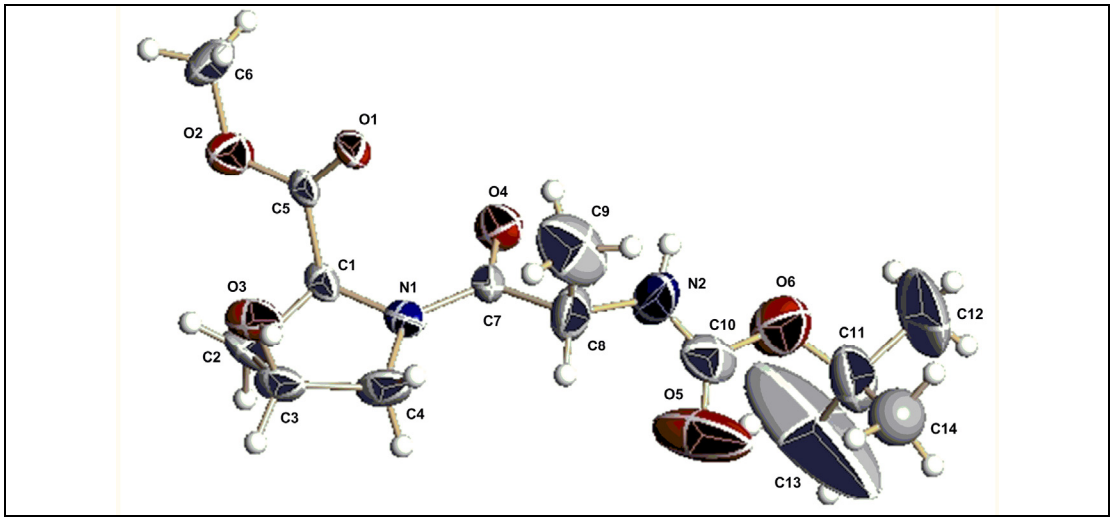
Empirical formula	:	$C_{13}\ H_{22}\ N_2\ O_5\ .H_2O$
Formula weight	:	304.34
Temperature	:	273(2) K
Wavelength	:	0.71073 Å
Crystal system, space group	:	Orthorhombic, $P2_12_12_1$
Unit cell dimensions	:	a = 5.868(16) Å alpha = 90°
		b = 13.458(4) Å beta = 90°
		c = 20.747(6) Å gamma = 90°
Volume	:	1638.6(8) $Å^{-3}$
Z, Calculated density	:	4, 1.234 g/cc
Absorption coefficient	:	0.097 mm^{-1}
F(000)	:	656
Crystal size	:	0.13 x 0.05 x 0.03 mm
θ range for data collection	:	1.80 to 24.99°
Limiting indices	:	-6<=h<=6, -16<=k<=15, -24<=l<=24
Reflections collected / unique	:	11542 / 2874 [R(int) = 0.0910]
Completeness to θ= 24.99°	:	99.9%
Max. and min. transmission	:	0.9969 and 0.9875
Refinement method	:	Full-matrix least-squares on F^2
Data / restraints / parameters	:	2874 / 0 / 203
Goodness-of-fit on F^2	:	1.193
Final R indices [I>2sigma(I)]	:	R1 = 0.1030, wR2 = 0.2414
R indices (all data)	:	R1 = 0.1173, wR2 = 0.2499
Absolute structure parameter	:	1(4)
Largest diff. peak and hole	:	0.440 and -0.358 $e.Å^{-3}$

Table 7: ORTEP diagram and Crystal data with structure refinement for compound **22**:



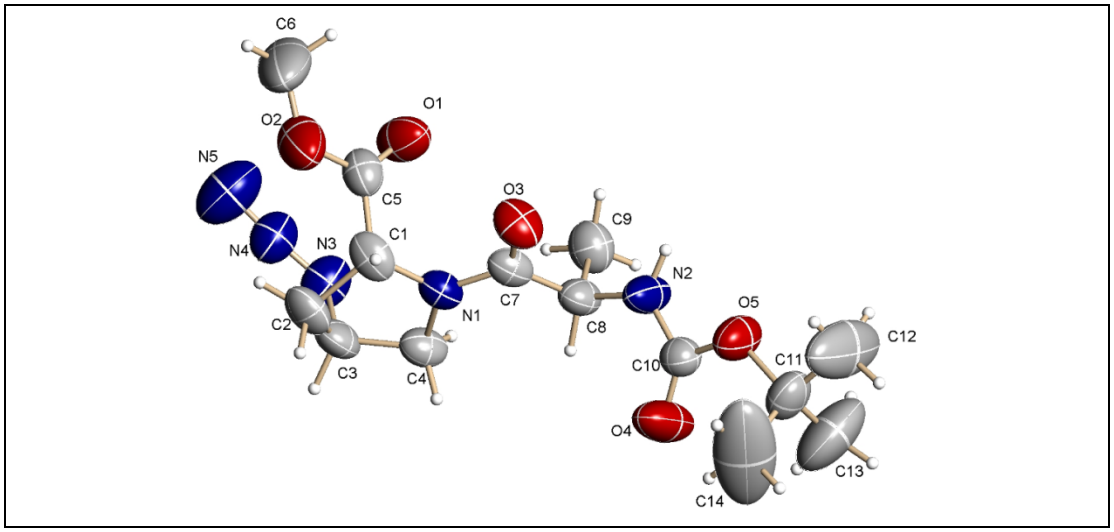
Empirical formula	:	$C_{13} H_{22} N_2 O_6$
Formula weight	:	302.33
Temperature	:	296(2) K
Wavelength	:	0.71073 Å
Crystal system, space group	:	Monoclinic, $P2_1$
Unit cell dimensions	:	a = 5.5597(4) Å alpha = 90°
		b = 10.5695(8) Å beta = 98.464(1)°
		c = 13.198(1) Å gamma = 90°
Volume	:	767.12(10) $Å^3$
Z, Calculated density	:	2, 1.309 g/cc
Absorption coefficient	:	0.104 mm^{-1}
F(000)	:	324
Crystal size	:	0.19 x 0.14 x 0.01 mm
θ range for data collection	:	2.48 to 25.00°
Limiting indices	:	-6<=h<=6, -12<=k<=12, -15<=l<=15
Reflections collected / unique	:	7482 / 2701 [R(int) = 0.0286]
Completeness to θ= 25.00°	:	99.9%
Max. and min. transmission	:	0.9989 and 0.9806
Refinement method	:	Full-matrix least-squares on F^2
Data / restraints / parameters	:	2701 / 1 / 196
Goodness-of-fit on F^2	:	1.174
Final R indices [I>2sigma(I)]	:	R1 = 0.0474, wR2 = 0.1005
R indices (all data)	:	R1 = 0.0526, wR2 = 0.1030
Absolute structure parameter	:	1.7(13)
Largest diff. peak and hole	:	0.175 and -0.137 e. $Å^{-3}$

Table 8: ORTEP diagram and Crystal data with structure refinement for compound **26**:



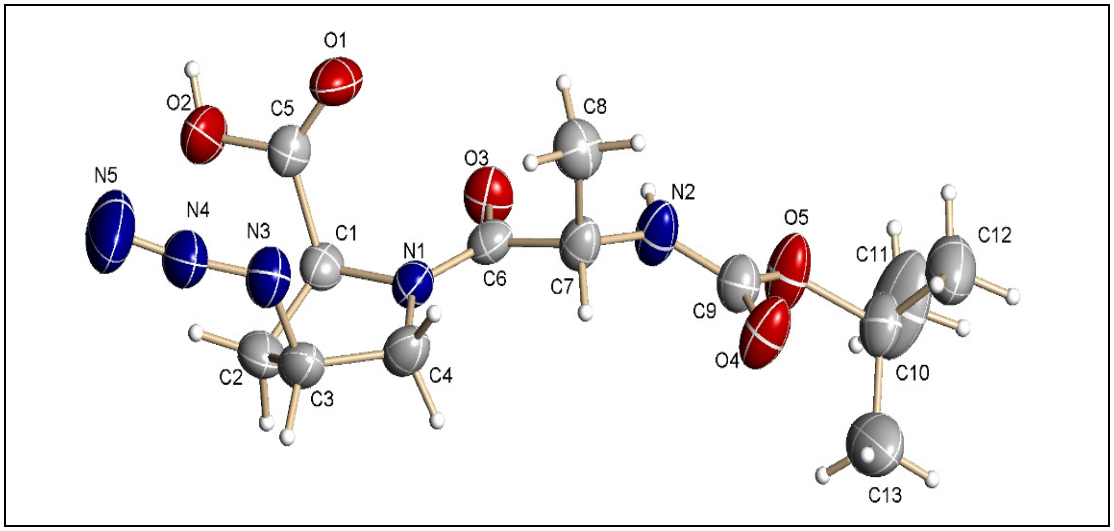
Empirical formula	:	$C_{14}\ H_{24}\ N_2\ O_6$
Formula weight	:	316.35
Temperature	:	200K
Wavelength	:	0.71073 Å
Crystal system, space group	:	Monoclinic, C2
Unit cell dimensions	:	a = 12.526(15) Å, α = 90.00° b = 5.694(7) Å, β = 100.53(2)° c = 26.17(3) Å, γ= 90.00°
Volume	:	1835(4) $Å^3$
Z, Calculated density	:	2, 1.145 gcm^{-3}
Absorption coefficient	:	0.089 mm^{-1}
F(000)	:	680
Crystal size	:	0.45 × 0.32 × 0.25 mm
θ range for data collection	:	3.17 to 28.28°
Limiting indices	:	-16<=h<=12, -7<=k<=7, -29<=l<=34
Reflections collected / unique	:	3769 / 2793
Completeness to θ= 28.28°	:	100%
Max. and min. transmission	:	0.978 and 0.966
Refinement method	:	Full-matrix least-squares on F^2
Data / restraints / parameters	:	2793 / 1 / 200
Goodness-of-fit on F^2	:	1.926
Final R indices [I>2sigma(I)]	:	R1 = 0.1775, wR2 = 0.4808
R indices (all data)	:	R1 = 0.1937, wR2 = 0.4911
Absolute structure parameter	:	2.3
Largest diff. peak and hole	:	0.680 and -0.600 e $Å^3$

Table 9: ORTEP diagram and Crystal data with structure refinement for compound **32**:



Empirical formula	:	C_{14} H_{23} N_5 O_5
Formula weight	:	341.37
Temperature	:	569(2) K
Wavelength	:	0.71073 Å
Crystal system, space group	:	Monoclinic, $P2_1$
Unit cell dimensions	:	a = 5.6061(5) Å alpha = 90°
		b = 11.078(1) Å beta = 93.560(2)°
		c = 14.469(1) Å gamma = 90°
Volume	:	896.90(14) $Å^3$
Z, Calculated density	:	2, 1.264 g/cc
Absorption coefficient	:	0.097 mm^{-1}
F(000)	:	364
Crystal size	:	0.43 x 0.33 x 0.03 mm
θ range for data collection	:	2.32 to 28.15°
Limiting indices	:	-7<=h<=7, -14<=k<=10, -18<=l<=18
Reflections collected / unique	:	5349 / 2937 [R(int) = 0.0197]
Completeness to θ= 28.15°	:	92.1%
Max. and min. transmission	:	0.9975 and 0.9594
Refinement method	:	Full-matrix least-squares on F^2
Data / restraints / parameters	:	2937 / 1 / 222
Goodness-of-fit on F^2	:	1.169
Final R indices [I>2sigma(I)]	:	R1 = 0.0708, wR2 = 0.1809
R indices (all data)	:	R1 = 0.0850, wR2 = 0.1921
Absolute structure parameter	:	0(2)
Largest diff. peak and hole	:	0.241 and -0.159 $e.Å^{-3}$

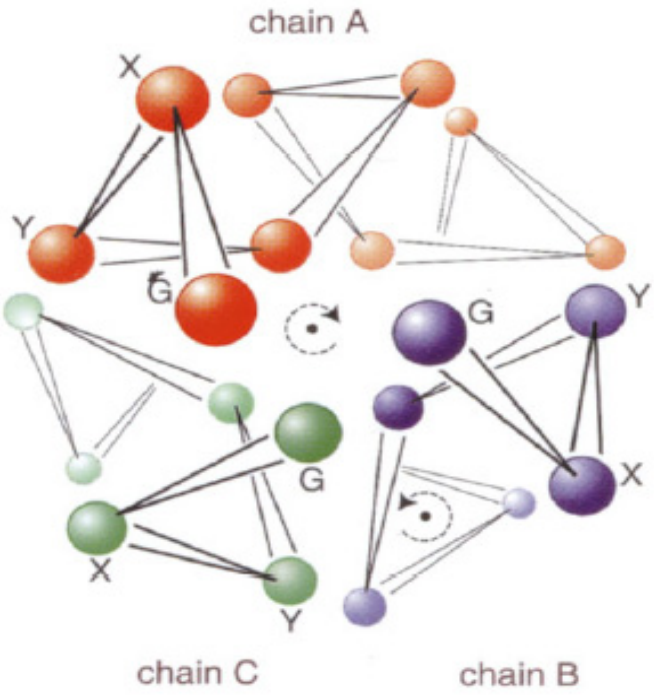
Table 10: ORTEP diagram and Crystal data with structure refinement for compound **33**:



Empirical formula	:	$C_{13}\ H_{21}\ N_5\ O_5$
Formula weight	:	327.35
Temperature	:	566(2) K
Wavelength	:	0.71073 Å
Crystal system, space group	:	Monoclinic, $P2_1$
Unit cell dimensions	:	a = 5.5489(4) Å alpha = 90°
		b = 10.3689(8) Å beta = 90.660(1)°
		c = 14.6022(1) Å gamma = 90°
Volume	:	840.10(11) $Å^3$
Z, Calculated density	:	2, 1.294 g/cc
Absorption coefficient	:	0.101 mm^{-1}
F(000)	:	348
Crystal size	:	0.42 x 0.18 x 0.02 mm
θ range for data collection	:	1.39 to 25.00°
Limiting indices	:	-6<=h<=6, -12<=k<=12, -17<=l<=17
Reflections collected / unique	:	8163 / 2959 [R(int) = 0.0275]
Completeness to θ= 25.00°	:	100.0%
Max. and min. transmission	:	0.9980 and 0.9587
Refinement method	:	Full-matrix least-squares on F^2
Data / restraints / parameters	:	2959 / 1 / 213
Goodness-of-fit on F^2	:	1.063
Final R indices [I>2sigma(I)]	:	R1 = 0.0417, wR2 = 0.0956
R indices (all data)	:	R1 = 0.0498, wR2 = 0.1002
Absolute structure parameter	:	0.2(12)
Largest diff. peak and hole	:	0.126 and -0.116 $e.Å^{-3}$

Section III

3-Hydroxy & Aminoproline in Chimeric Collagen Peptides: Synthesis & CD Studies



A brief introduction to the world of collagen followed by the recent literature trends in research is presented in this section. The effects of the different structural modifications on the properties of collagen-triplex stability have been overviewed to draw directions for the present work, targeted towards exploring new pathways of chimeric sequence design and investigation. Collagen sequence modifications by introduction of hydroxyl and amino-functionality (as non-natural amino acid) at 3-position of proline ring at X/Y positions in the $(X\text{-}Y\text{-}Gly)_n$ backbone have been described in this section.

3.1 Introduction

Collagen is a major structural protein in vertebrates. This fibrous protein is present in all connective tissues such as skin, bone, cartilage, tendons, basement membrane and blood vessels, and imparts stability and integrity to these tissues, performing somewhat the same function as cellulose molecules in plants. Figure **1** shows a typical collagen fiber originating from human tendon (a tough band of fibrous connective tissue that usually connecting muscle to bone) split into microfibril to triple-helix to the polypeptide chains. Collagenous sequences are also found in invertebrates[1] and in lower organisms like bacteria and phages.[2] Depending on the amino acid composition, collagen in vertebrates is divided into at least 19 different types.[3] Among these, collagen types I-III are the most abundant and form fibrils of similar structure.

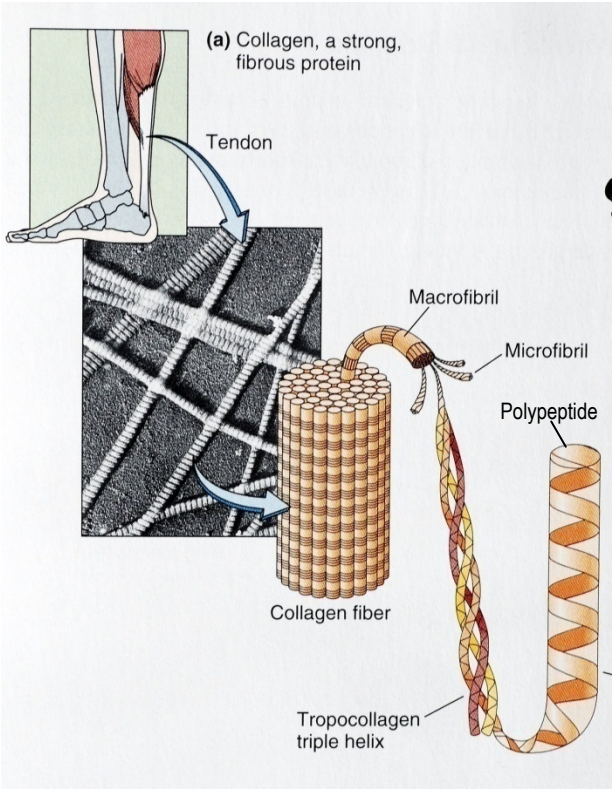


FIGURE 1: Classification of collagen as a structure

The molecular hallmarks of collagen proteins are the multiple repetitions of Gly-X-Y amino acid sequences and the unique triple-helical structure built by three polypeptide chains. The primary structure of collagen is composed of approximately 300 repeats of the trimer Gly-X-Y, where X and Y can be any amino acid. The most

commonly found amino acids in X and Y positions are proline (Pro) and *trans*-4-hydroxyproline (Hyp) and account for nearly 20% of the total amino acid content of natural collagen. The other commonly found amino acids are Ala, Lys, Arg, Leu, Val, Ser and Thr.[4] The secondary structure of collagen chain is an extended left-handed helix called - polyproline type-II conformation.[5] In the tertiary structure of collagen, three such polyproline II helices are intertwined in a parallel fashion with one residue shift to form a right handed super-helix called - "collagen triple-helix".[6]

Depending up on the amino acid composition and the supramolecular assembly, collagen super-family is classified into different subfamilies. Up till now 42 different polypeptide chains have been identified, which are encoded by 41 specific genes and compose 27 unique collagen types.[3] Among these, Collagen I-III are the most abundant and form fibrils of similar structure.

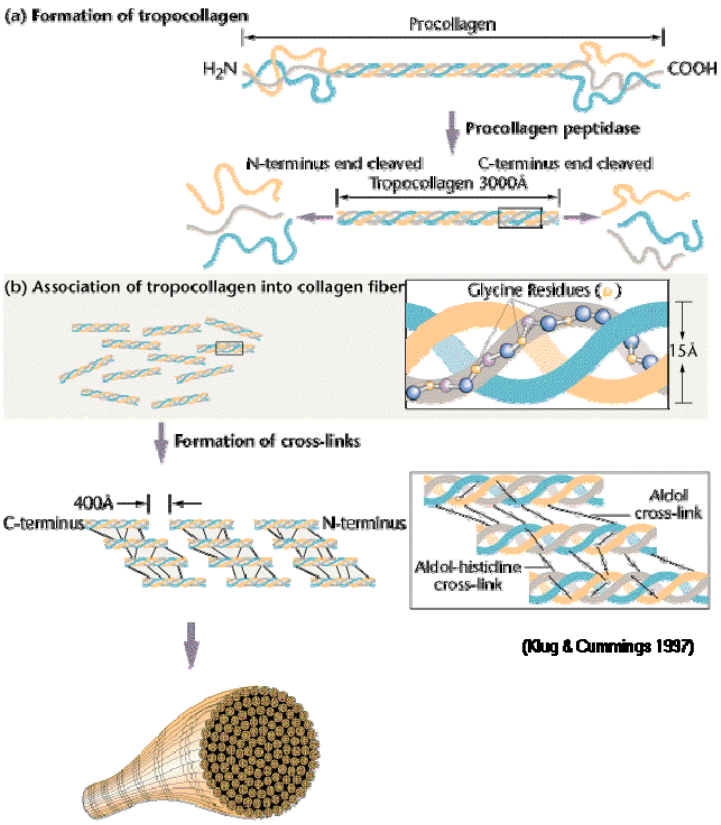


FIGURE 2: Extracellular events in the biosynthesis of fibrillar collagens

The biosynthesis of collagen is well studied (Figure **2**).[7] Collagen is synthesized in endoplasmic reticulum (ER) as procollagen, which is the precursor protein that bears propeptide domains at either end of the triple helical domain. This process also involves correct alignment of triple-helix without mis-staggering and

variety of post-transnational modifications, in particular, the prolyl-4 hydroxylation of Y-position amino acids through the action of enzymes *Prolyl 4-hydroxyase* and *lysyl hydoxyase*. The polypeptide chains are then secreted into extra cellular matrix, where they fold into a triple-helix.[8]

3.2 Positional preferences of amino acids in collagen

In principle, there are 400 different possible Gly-X-Y triplets for composing peptide with 20 amino acids, but only a limited number are actually found in collagen sequences.[9] Several attempts have been made to measure the stability of collagen model peptides with various amino acids and the triple helical propensities.[10] In one approach, a set of host triple-helical peptides with the sequence $(Gly\text{-}Pro\text{-}Hyp)_3$-Gly-X-Y-$(Gly\text{-}Pro\text{-}Hyp)_4$, where GXY is the guest triplet have been used to elucidate positional preference of various amino acids commonly found in natural collagen. All 20 amino acids were tested individually at both X or Y position with the help of T_m measurements and thermodynamic analyses of the host sequence. The results have revealed a striking relationship between the nature of the amino acid and the triple-helix stability.[11]

Pro was found to be the most stabilizing residue (T_m = 47.3 °C) in X position and also offered the lowest enthalpic contribution ($\Delta H°$ = 435 kJ mol^{-1}) in the entire series. The host peptides containing charged amino acids (Glu, Lys, Arg and Asp) were among the most stable peptides but these residues possessed higher enthalpic contribution than Pro, which reflected side-chain interactions with available backbone carbonyl or solvent. Trp was found to be the most destabilizing amino acid in X position (T_m = 31.9 °C). Among the Y position replacements Hyp/Arg were most stabilizing amino acids with nearly equal T_m values of 47.3 °C and 47.2 °C respectively. But, Arg peptide showed a higher $\Delta H°$ of denaturation than Hyp (610 kJ mol^{-1} and 435 kJ mol^{-1} respectively), suggesting a different mechanism for stabilization. The aromatic amino acids Tyr, Phe and Trp as usual were most destabilizing (T_m values 30.2 °C, 28.3 °C and 26.1 °C respectively).

Thermodynamic analysis has shown that both Hyp in Y position and Pro in X position have the same enthalpy contribution ($\Delta H°$ = 435 kJ mol^{-1}). It is also interesting to see that the most destabilizing amino acid Trp has the highest enthalpy value (593 kJ mol^{-1} in X position and 670 kJ mol^{-1} in Y position). Analysis of amino acid sequences of fibril-forming collagens in SWISS-PROT database showed that the

frequency of occurrence of each of the amino acids in fibril-forming collagens correlate well with triple-helix stability.

3.3 Three dimensional structure & crystal packing of collagen

In 1954, Ramachandran and Kartha provided the first[12] correct model for the structure of collagen as three parallel left-handed polypeptide chains, held together by interchain hydrogen bonds, which was later modified by themselves to be a coiled-coil structure.[13] Soon after Rich & Crick[14] tailored this model to propose another model where each of the strands in triple-helix has left-handed coil (PPII) which interwind one hydrogen bond per every trimer repeat to form a right-handed triple-helix conformation (Figure **3.A**).

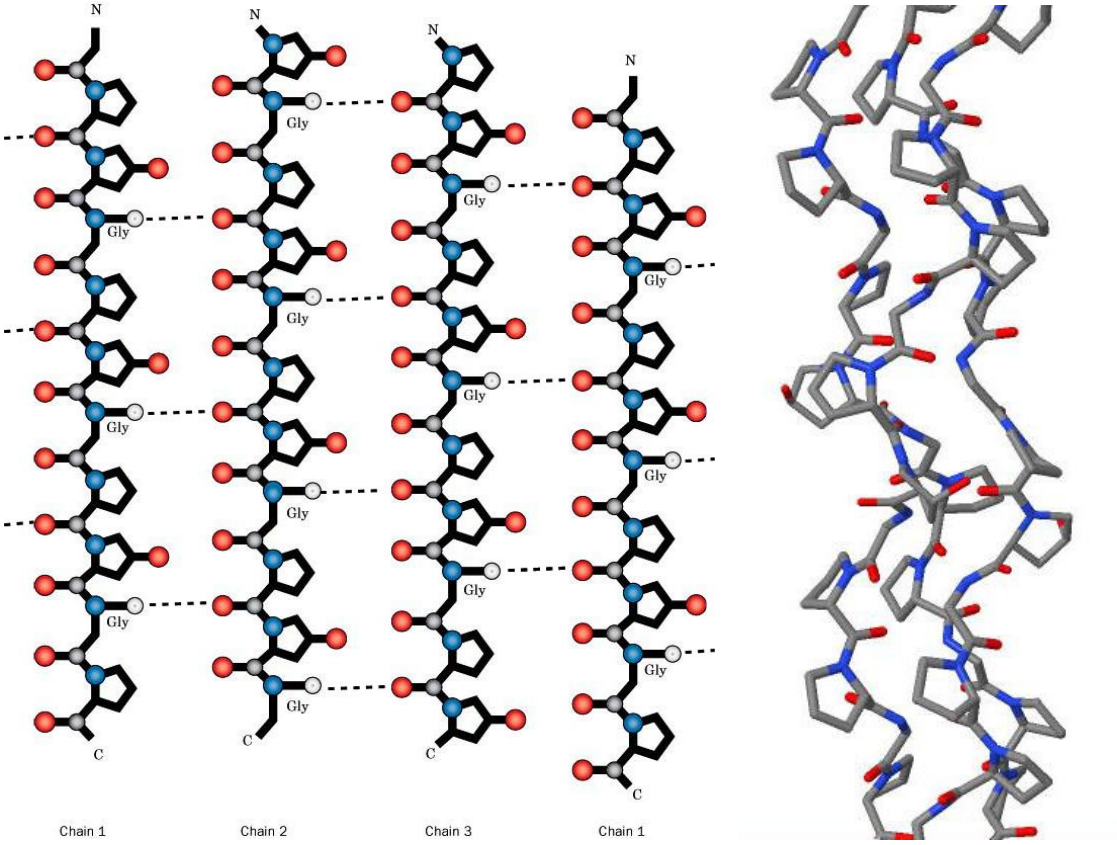


FIGURE 3: (**A**) Schematic view of Rich & Crick model with one hydrogen bond per trimer repeat (Gly-Pro-Hyp) repeat (**B**) Crystal structure of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ suggested by Okuyama *et al.*(PDB: 1V4F)[15]

But high resolution structures of triple-helix peptides obtained by X-ray crystallography provided confirmation and new insights into the molecular conformation. The first crystal structure was obtained by Okuyama *et al.* (1981)[15] for

$(Pro\text{-}Pro\text{-}Gly)_{10}$, followed by structures for peptides with varying sequences from a number of laboratories (Table **1**)[16]. These high-resolution structures of collagen- like peptides confirm the basic triple-helical model and show details of hydration, hydrogen bonding, and helical parameters.

Table 1: List of Examples of Collagen Peptides with High Resolution Structures Solved by X-Ray Crystallography[a]

	Peptide Sequence	PDB ID	Properties
Imino acid Polytripeptides	$(Pro\text{-}Pro\text{-}Gly)_n$ where $n = 9,10$	1A3I, 1K6F, 2CUO	7/2 symmetry,
	$(Pro\text{-}Hyp\text{-}Gly)_{10}$	1V7H	NH(Gly)...CO(X)
	$(Gly\text{-}Hyp\text{-}Hyp)_9$	1YM8	
Peptides with collagen sequences:			
LOG2	$(Pro\text{-}Hyp\text{-}Gly)_4$-$(Leu\text{-}Hyp\text{-}Gly)_2$-$(Pro\text{-}Hyp\text{-}Gly)_4$[b]		NH(Gly)...CO(X), NH(X)..W..CO(Gly),
Integrin binding peptide	$(Gly\text{-}Pro\text{-}Hyp)_2$ -Gly-Phe-Hyp-Gly-Glu-Arg-$(Gly\text{-}Pro\text{-}Hyp)_3$	1DZI, 1Q7D	7/2 symmetry in ends, more relaxed in central imino
T3-785 peptide	$(Pro\text{-}Hyp\text{-}Gly)_3$-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala-Gly-$(Pro\text{-}Hyp\text{-}Gly)_4$	1BKV	acid poor zone
Peptides with Imperfect Gly-X-Y repeat:			
Peptide with natural interruption	$(Pro–Hyp–Gly)_3$– Pro–Hyp–Gly–Pro–Gly–$(Pro–Hyp–Gly)_5$	1EI8	Altered H bonding at interruption site, good helix
Peptide with Gly→Ala mutation	$(Pro\text{-}Hyp\text{-}Gly)_4$-Pro-Hyp-Ala-$(Pro\text{-}Hyp\text{-}Gly)_5$	1CAG, 1CGD	at both ends, but out of register

[a] For a complete listing of all solved structures as of 2006, see Okuyama *et al*., 2006.[16b]
[b] Okuyama et al, 2007.[16c]

3.4 Role of hydroxyprolines in collagen structure

The extremely high content of hydroxyproline in collagen (about 10% in natural collagen) suggests an important role for this amino acid in collagen structure. Proline hydroxylation occurs as a major post–transnational event. During biosynthesis, proline residues are incorporated both in the X and the Y positions. However, only those in the Y positions are hydroxylated in vertebrate collagen to produce the specific diastereoisomer 4*R*-hydroxyproline (Hyp). By contrast, prolines both in X and Y positions may be hydroxylated to 3-hydroxyproline in some basement membrane collagens.[17]

3.4.1 Collagen structures with 4(*R/S*)-hydroxyproline

Several studies on collagen model peptides have suggested that peptide conformation is significantly affected by the 4*R*-hydroxyproline (Hyp) content. The replacement of Pro in either X or Y position by Hyp was found to have significant effect on the triple-helicity. For example, $(Pro.Hyp.Gly)_{10}$ had a T_m = 58 °C while $(Pro.Pro.Gly)_{10}$ had a T_m = 24 °C.[18] On the other hand, polypeptides, having repeating sequence Hyp-Pro-Gly, with 4*R*-Hyp in X position, do not associate in

triple-helix.[19] Furthermore it has been shown that triple-helix folding is also inhibited by the presence of the diastereoisomer 4*S*-hydroxyproline (4*S*-Hyp)[20] in either X or Y position. Figure **4** shows the postulated hydrogen bonding and puckering for 4(*R/S*)-hydroxyproline residues as a part of collagen sequence.

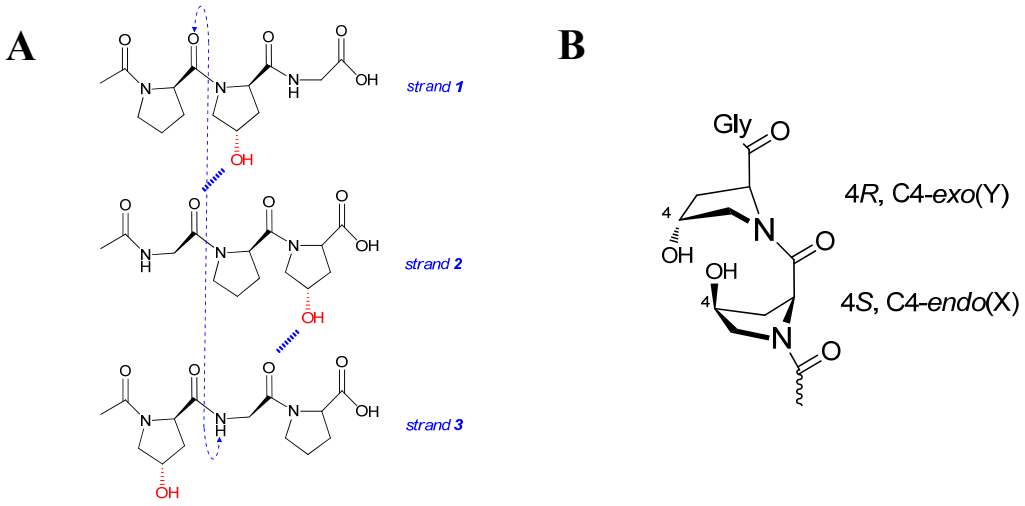


FIGURE 4: Collageneous peptides with (**A**) Hydrogen bonding inbetween strands of $(Gly\text{-}Hyp\text{-}Pro)_n$ (**B**) Ring puckering preferences in 4(*R/S*)-Hyp

To explain the stabilizing role of 4*R*-Hyp in Y position several hypotheses have been put forward since early 1950. Later, such enhanced triple-helical stability has been attributed to the extensive hydration induced by the presence of Hyp.[21] But more recently this was explained to arise from the strong inductive effects of the electronegative group, which imposes C_4-*exo* pucker on the pyrrolidine ring to attain the required *trans* conformation for prolyl-peptide bond through ring NH-C_5-C_4-X gauche effect and thus preorganize all three torsion angles.[22] Furthermore recent data have also shown that, when compared to Pro, 4*S*-Hyp favors the *cis* state of the peptide bond, which indirectly destabilizes the triple-helix.[23]

3.4.2 Collagen structures with *trans*-3-hydroxyproline

3(*S*)Hyp is clearly an integral part of the collagen triple helix in fibrillar and basement membrane collagens and is also present at low levels in other collagen molecules such as type X collagen.[24] During the post-translational modifications process in addition to prolyl 4-hydroxylation, a small numbers of proline residues are modified to 3(*S*)-hydroxyproline [3(*S*)Hyp][25] in many types of vertebrate collagens, such as types I, II, III, IV, V and X. Among these, Type I collagen, which consists of two α1 chains and one α2 chain, has a single 3(*S*)Hyp residue per chain.[26] The

proline residue at position 986 in the α1 chain is modified to 3(*S*)Hyp[27] by the protein complex P3H1/CRTAP/ cyclophilin B.[28] Invertebrate collagens also contain 3(*S*)Hyp, for example interstitial and cuticle collagens of annelids[29], crab sub-cuticular collagen, lobster sub-cuticular membrane collagen, squid skin collagen, abalone muscle collagen,[30] octopus skin collagen[31], octopus arm collagen[32] and jellyfish mesogloea collagen[33].

However a few studies have determined the position of 3(*S*)Hyp by amino acid sequencing, and the 3(*S*)Hyp was found in (Xaa position) a -Gly-3Hyp-4Hyp-Gly- sequence in all instances.[34] Although 4(*R*)Hyp residues in the Yaa positions have been shown to be critical for the formation of a stable triple helix, the role of 3(*S*)Hyp residues in the Xaa position is not very well understood.[35]

3.5 Characterization of triple-helical structures

Collagen triple-helix upon heating undergoes a helix ⇔ coil transition resulting in change of various spectroscopic and physical properties. These include change in molar extinction coefficient, molar ellipticity, specific rotation and integrations of NMR signals of associated and dissociated forms.[36,37,38] The thermal denaturation of associated triple-helical peptides obtained by monitoring one of these properties shows a sharp cooperative transition. The midpoint of such a transition (T_m) is used as a measure of triple-helical strengths of collagenous peptides.

A variety of experimental techniques have been used to detect and characterize triple-helical structures and these include spectroscopic methods such infrared-vibrational spectroscopy,[39] Raman spectroscopy[40] and two-dimensional NMR spectroscopy.[41] X-ray diffraction and electron microscopy have been used to study the sizes and shapes of collagen fibrils as well as synthetic collagen-like polymers.[42,43] More recently, single crystal X-ray crystallography of collagen like peptides of defined length has been used to determine triple-helical structures at atomic resolution. In solution, sedimentation-equilibrium and light scattering measurements have been utilized to study molecular weight-dependent properties such as formation and denaturation of triple-helices.[18,44]

Circular dichroism (CD) spectroscopy has been the most commonly employed tool to the study triple-helical conformations of collagen and collagen-like peptides in solution.[45] The CD spectrum of collagen triple-helix is unique with a small positive peak at around 225 nm, a crossover at 213 nm and a very large negative

peak at around 200 nm (Figure **5.A**). Polyproline II and polyproline II-like structures also exhibit CD spectra similar that of collagen triple-helices. Upon heating to higher temperature, natural collagen undergoes a triple-helix ⇔ coil transition resulting in decreased molar ellipticity. The thermal-denaturation curves of associated triple-helical peptides, obtained by monitoring the change in molar ellipticity at 222-227 nm, show a sharp, cooperative sigmoid transition (Figure **5.B**).

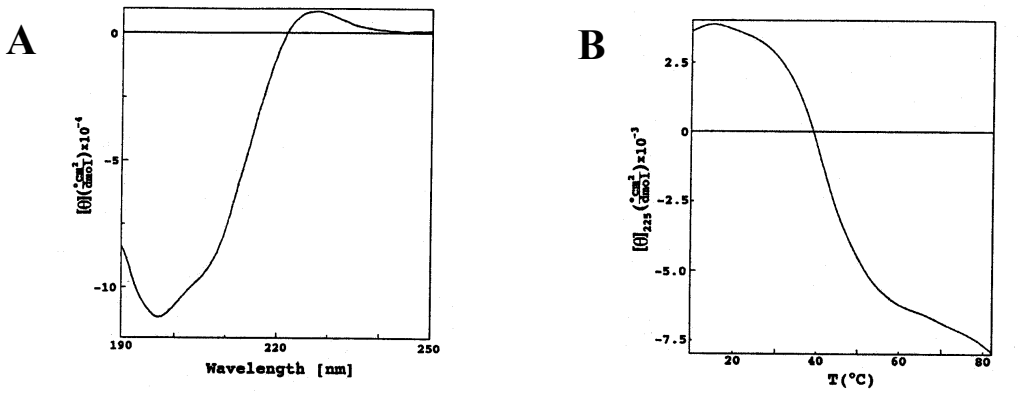


FIGURE 5: (**A**) CD spectrum at 10 °C and (**B**) triple-helix ⇔ coil transition curve for $(Pro\text{-}Hyp\text{-}Gly)_{10}$ in water.[45b]

3.6 Collagen mimetics

The development of collagen research has expanded to elucidate the interaction involved in stabilization and folding of collagen-like structures and to design alternatives to natural collagen-based biomaterials. The primary peptide sequence can be manipulated by incorporating different residues and sequences and the effect of such alternation can be investigated. By implementing a *de novo* approach, it is possible to design structures to induce triple helicity and enhance collagen stability.

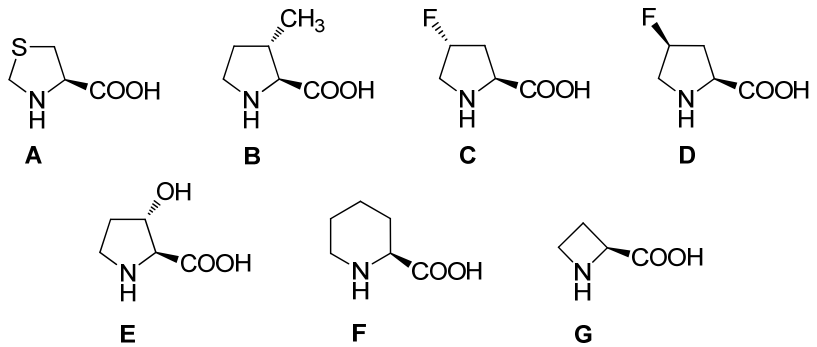


FIGURE 6: Various proline surrogates that have been incorporated in to collagen sequences (**A**) Thiozilidine (Thz), (**B**) *trans*-3-Methylproline (MePro), (**C**) *trans*-4-Fluoroproline (Flp), (**D**) *cis*-4-Fluroproline (flp), (**E**) *trans*-3-Hydroxyproline, (**F**) L-piperidine-2-carboxylic acid (Pipec), (**G**) Azitidine (Aze).

In attempts to understand and modify the triple-helical strength of collagen, several unnatural amino acids have been incorporated into the X and Y positions of collagen sequence (Figure **6**). Synthesis of sequential polypeptides poly-(Ala.Gly.Thz), poly-(Gly.Pipec.Ala) and poly-(Gly.Aze.Ala)† have been reported.[46] Peptides $(Gly.Pro.MePro)_n$, $(Pro.Gly.Flp)_nPro$† (n = 1-4) were synthesized and tested for the *proline-hydroxylase* inhibition activity.[47] However, the conformational properties of these peptides are not known. Later work on poly-(Gly.Pro.Aze), poly-(Gly.Aze.Pro), and poly-(Gly.Aze.Aze) has shown that poly-(Gly.Pro.Aze) exhibits triple-helical character according to CD vibrational spectroscopy and molecular modeling studies.[48] Goodman *et al.* have incorporated a peptoid residue, *N*-isobutylglycine (Nleu) into the collagenous peptide sequences to obtain stable triplexes. In addition to electrostatic interactions with charged amino acids,[49] Brodsky *et al.*[50] have also studied intermolecular hydrophobic interactions of non polar aminoacid residues for triple-helical stabilization.

A template assembly of peptides enhances secondary structure formation with high conformational stability which is a fundamental requirement to induce proper biological response or activity. Hence incorporation of template into the design of peptidomimetics can direct and reinforce the intramolecular folding of peptides.[51] In case of collagen, a template favors intramolecular folding as opposed to single chain intermolecular folding, and stabilizes the triple helicity by reducing any entropy loss in triple helical formation.[52] Heidemann *et al.*[53] have used 1,2,3-propane tricarboxylic acid and a Lys-dimer, Fields *et al.* have employed two consecutively connected Lys residues with three amino groups and Goodman *et al*.[54] have used Kemp's triacid or KTA as templates to anchor individual strands of the collagen triple-helix.

3.7 Collagen applications: Ageing, diseases & biomaterials

The biological diversity of the function of collagenous tissues is primarily due to several genetically distinct classes of collagens that are tissue specific. During aging, several chemical changes occur in the collagenous framework (mostly increase in its cross-linking) and these changes reflect in the physical properties of the fibers. The major change is an increase in the rigidity of the tissue, with the fibers ultimately becoming brittle.[55] Collagen synthesis decreases steadily with maturation,

† Thz = L-thiazolidine-4-carboxylic acid; Aze = L-azetidine-2-carboxylic acid; Pipec = L-piperidine-2-carboxylic acid; MePro = *trans*-3-methyl-L-proline; Flp = *trans*-4-fluoro-L-proline

and with subsequent aging drops 10-fold in the majority of tissues. Many common diseases, such as diabetes, arthritis, and cancer, involve abnormal degradation, reactivity, or cross-linking of collagen. There are also rare hereditary disorders that originate from specific mutations in collagen genes. The clinical manifestation of these genetic diseases depends on the type, location, and function of the collagen that has the mutation: a mutation in Type I collagen, the major fibril forming collagen in bone, leads to *Osteogenesis Imperfecta* (OI), a disease characterized by fragile bones,[56] whereas a mutation in Type IV collagen in basement membranes, the filtration barrier in the kidney glomerulus, leads to *Alport Syndrome* with progressive kidney failure.[57] Most of the mutations are missense mutations which replace one Gly residue in the Gly-X-Y repeating pattern by a larger residue.

Collagen and collagen-based materials have extensive applications in biomedical devices and tissue engineering.[58] They are used, for example, in the fabrication of cardiac and aortal valves[58b] and vascular grafts,[59] and have been proposed for use in ophthalmic implants,[60] abdominal repair fabrics,[61] and ligaments.[62] Chemical stabilization of collagen is often used to reduce the potential of immune and inflammatory reactions, and to stabilize it against proteolytic degradation. Such collagen and collagen-based materials seem to have an inherent biocompatibility that is difficult to mimic in synthetic polymers.

In recent years, the biomedical importance of collagen has led to increasing interest in the structure of this protein and in establishing relationships between sequence, structure, and function.

3.8 Aim and rationale for the present work

The work in this section of thesis is mainly devoted to the synthesis and structural characterization of *new class of synthetic collagens with 3-aminoproline* as *a Hyp and Pro surrogate along with a comparision with 3-hydroxyproline collagen analouges*.

Whether 3(*S*)Hyp residue in the Xaa position stabilizes or destabilizes the desired collagen helix is still controversial. In host-guest peptides, it was found that the stability of the triple helix is decreased when Pro in the Xaa position is replaced by either 3(*S*)Hyp[35b] or 3(*S*)fluoroproline.[35c] It was explained that, the inductive effect of its 3-hydroxyl group slightly diminishes the strength of the interstrand 3-HypC=O••••H–NGly hydrogen bond (Figure **7**). The destabilization is large when 3-

Hyp is in the unnatural Yaa position where, the pyrrolidine ring pucker leads to inappropriate mainchain dihedral angles and interstrand steric clashes.[35c]

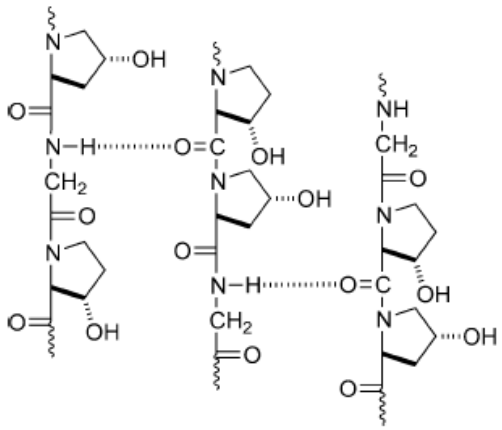


FIGURE 7: Putative interstrand hydrogen bonds in triple-helical[35c] $(3\text{-Hyp-4-Hyp-Gly})_n$.

An amino group attached to the C3/C^{β} of proline ring makes a versatile building block with two stereocenters, resulting in a stereochemical diversity of four isomers. This amino acid possesses two amino groups and these groups when linked to the α-carboxyl group can provide both α- and β-peptides (Figure **8.A**). The aminoproline units linked using the 3-amino group *i.e.*, β-peptide are reported in many biologically important foldamers.[63,64]

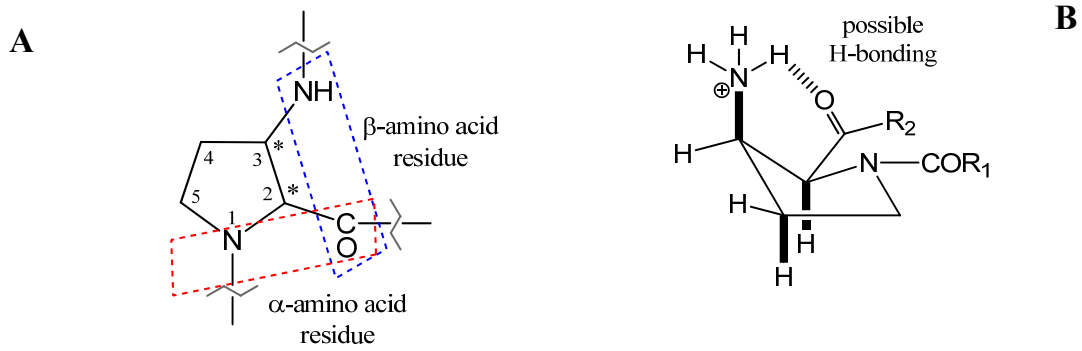


FIGURE 8: Schematic view of both α and β peptide linkages in 3-aminoproline. The two stereocentres on the pyrrolidine ring are indicated by an asterisk.

In view of the above a 3-aminoproline substituted triple-helix was conceived, where it would form α-peptide bonds. In case of 3*R*-aminoproline (3-Amp) and 3*S*-aminoproline (3-amp) the 3-NH_2 group, like hydroxyl is a potent hydrogen bond donor. Moreover, its higher basicity compared to -OH causes its protonation at physiological pH (Figure **8.B**). Overall amino group is ionisable, electronegative with potential to form hydrogen bond.

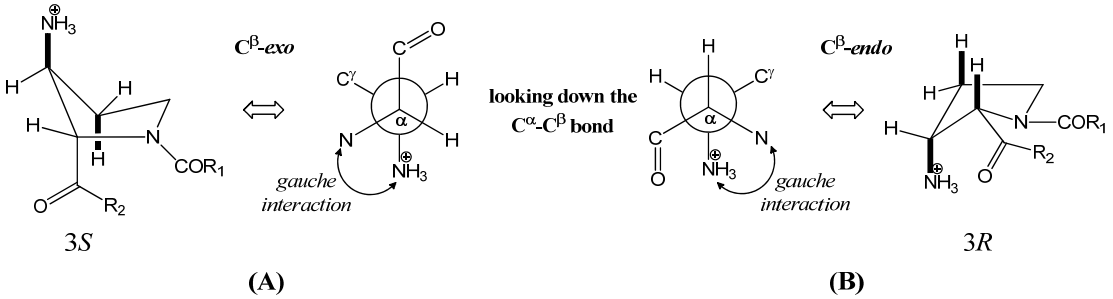


FIGURE 9: Expected pyrrolidine conformation for 3-aminoproline peptides (**A**) 3*S*-aminoproline and (**B**) 3*R*-aminoproline

As per Jenkins *et al.*'s reported X-ray data for *N*-($^{13}C_2$-acetyl)-3(*S*)-hydroxyproline methyl ester, this molecule showed the conformation of the pyrrolidine ring with the C^{α}, C^{γ}, and N atoms located within a plane and C^{δ} deviating slightly.[35c] Assuming a pyrrolidine conformation in which carbon bearing the substituent, C^{β} is the "flap atom" of the five membered ring, analysis of the potential stereoelectronic effects predicts that the 3*S*-Hyp would contain a pyrrolidine ring in a C^{β}-*exo* conformation while the 3*R*-hyp residue have C^{β}-*endo* conformation (Figure **9**).

Moreover the 4(*R/S*)-aminoprolines have already been tested as a surrogate of 4-hydroxyproline in both Xaa and Yaa position in collagen sequence of $(Gly\text{-}X\text{-}Y)_6$ type which show better triple-helical propencity.[65] Thus it seemed worth investigating the impact of amino group in the C3/C^{β}-position of pyrolidine ring as a part of the collagen sequence.

The objective of the present work is to examine the compatibility of 3*S* and 3*R* hydroxyl and aminoprolines wrt to their conformation of pyrrolidine ring adopted at both X and Y position of the collagen peptide $(X\text{-}Y\text{-}Gly)_n$**.** This section deals with the chemical synthesis and biophysical studies of the collagen peptides $(X\text{-}Y\text{-}Gly)_n$ having one modified amino acid.

The specific objectives of this section are

1) Synthesis of (2*S*,3*S*)-N^{α}-(*t*-Boc)-3NH-Fmoc-aminoproline (**52**), (2*S*,3*R*)-N^{α}-(*t*-Boc)-3NH-Fmoc-aminoproline (**56**), (2*S*,3*S*)-N^{α}-Fmoc-3-Hydroxyproline (**57**) and (2*S*,3*R*)-N^{α}-Fmoc-3-Hydroxyproline (**58**) monomers.
2) Incorporation of these monomers in collagen chimeric peptides using 3Amp/3Hyp, and 3amp/3hyp at Y and X position respectively through solid phase synthesis.

3) Cleavage of peptides from the solid support, purification, and characterization of these peptides.

4) Investigation of triple-helix forming ability using concentration, temperature and solvent dependent CD spectropolarimetry.

The *N*-terminal-acetylated and *C*-terminal-amidated peptides were designed to enable the delineation of the direct contribution of 3-NH_2 group towards the triple-helical stability. In order to facilitate the easy determination of concentration of the peptide solutions, a chromophore pheylalanine (Phe) was incorporated into the sequence.[66] Since, all peptides contained Phe any effect of this residue on the triple-helical stability would be identical for all the peptides.

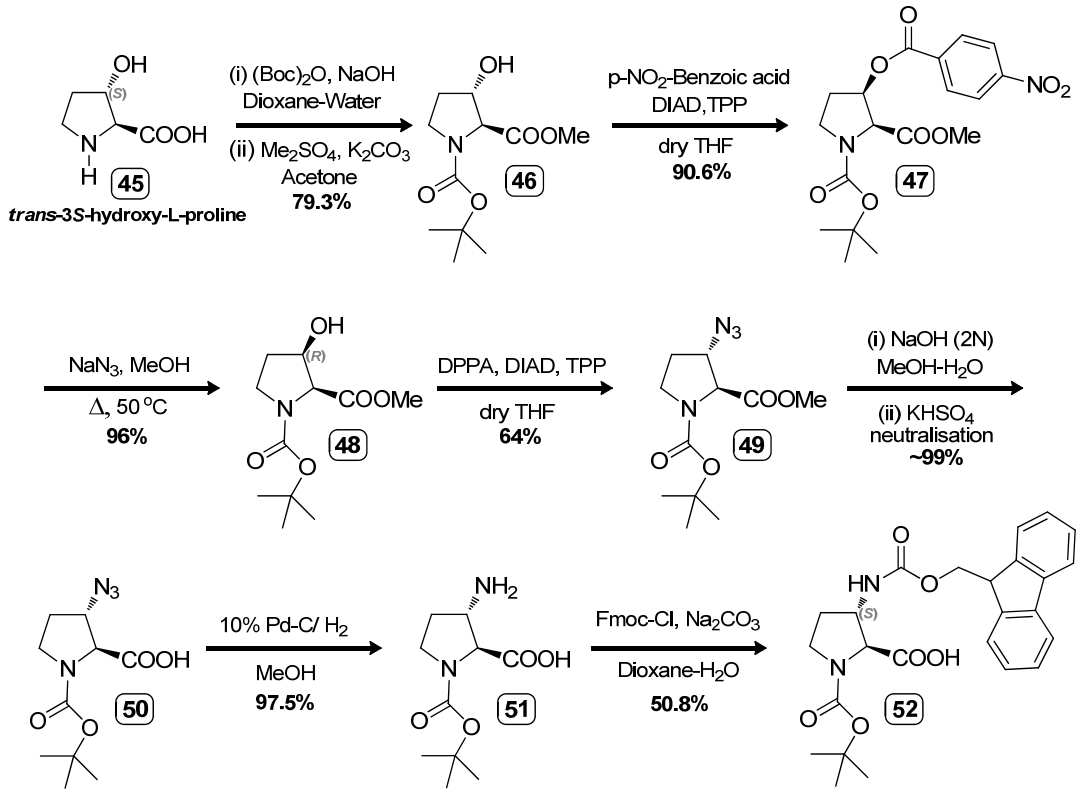
3.9 Results

3.9.1 Synthesis of Monomers

3.9.1a Synthesis of fully protected (2*S*,3*S*) and (2*S*,3*R*)-aminoproline monomers

Synthesis of orthogonally protected (2*S*,3*S*) and (2*S*,3*R*)-aminoproline monomer (**52** and **56**) were achieved in seven and five steps respectively from the commercially available *trans*-3*S*-hydroxy-L-proline **45** (Scheme **3.1** & **3.2**).

The two-step protection of the α-amino group with bis-*t*-butoxycarbonyl anhydride and esterification of the carboxylic acid with DMS in anhydrous acetone and K_2CO_3 afforded protected *trans*-3*S*-hydroxyproline **46** as evidenced by the appearance of signals in the ^{1}H NMR spectrum at δ 3.73 (-OCH_3) and δ 1.40 & 1.46 (*t*-Boc). Inversion of configuration of 3*S*-hydroxy group in **46** using p-nitrobenzoic acid under Mitsunobu conditions gave *p*-nitrobenzoate **47** in 90% isolated yield. The compound **47** was verified from appearance of aromatic peaks δ 8.13-8.17 (d, 2H) and 8.27-8.32 (d, 2H) in ^{1}H NMR. Use of *p*-nitrobenzoic acid under Mitsunobu conditions[67] reported in literature to increase the reactivity of the benzoate derivative. Inversion of configuration proceeds in nearly the same yield as with benzoic acid, and the ester could be cleaved by treatment with sodium azide in methanol in a quantitative yield.[68] Thus compound **47** was selectively hydrolyzed with sodium azide

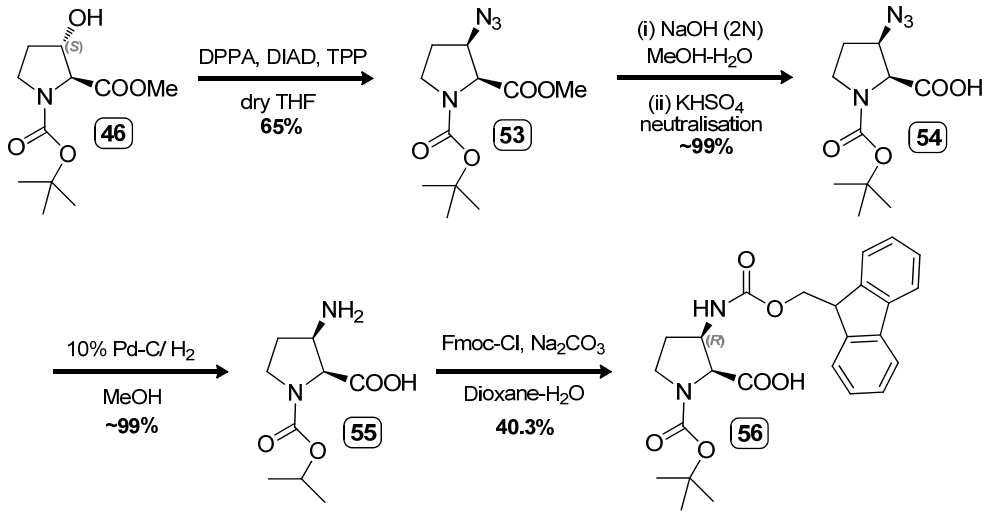


SCHEME 3.1: Synthesis of (2*S*,3*S*)-N^{α}-(t-butoxycarbonyl)-3-(9-fluorenylmethyloxycarbonylamino) proline

in methanol to afford protected *cis*-3*R*-hydroxyproline **48** in an excellent yield. The hydroxyl group was then converted to azide with DPPA with re-inversion of stereocenter to give 64% yield of compound **49**. A peak at 2105 cm^{-1} characteristic of azide was seen in the IR spectrum of **49**. The α-methyl ester group of compound **49** was hydrolyzed with aq. NaOH (1N) in methanol to give free acid **50**, confirmed by the disappearance of –OC<u>H</u>$_3$ at δ 3.76 from the ^{1}H NMR spectrum. This azide group of this resulting acid was selectively reduced to the corresponding amine **51** using 10% Pd-C under hydrogenation conditions. The amine was then protected with Fmoc, by reacting with 9-fluorenylmethylchloroformate to yield the fully protected 3*S*-aminoproline (3*S*-Amp) **52** (Scheme **3.1**). The appearance of signals in the aromatic region typical of Fmoc group [δ 7.29-7.33 (t, 2H), 7.38-7.41 (t, 2H), 7.56-7.58 (d, 2H), 7.74-7.76 (d, 2H); Ar<u>H</u> of Fmoc], peaks at 453 [M+H] and 475 [M+Na] in the LCMS confirmed the structural integrity of the monomer **52**.

Similarly for synthesis of N^{α}-Boc,N^{γ}-Fmoc-(2*S*,3*R*)-aminoproline (3*R*-amp) monomer, protected *trans*-3*S*-hydroxyproline **46**. Conversion of 3*S*-OH to 3*R*-N_3 of compound **46** using a Mitsunobu procedure with DIAD/PPh_3/DPPA yielded 65% of azide **53** with characteristic IR peak at 2111 cm^{-1}. Subsequent hydrolysis with NaOH

in aqueous methanol gave acid **54** with no peak at δ 3.77 for –OC<u>H</u>$_3$ in ^{1}H NMR spectrum. The hydrogenolysis of azide to amine was achieved by catalytic

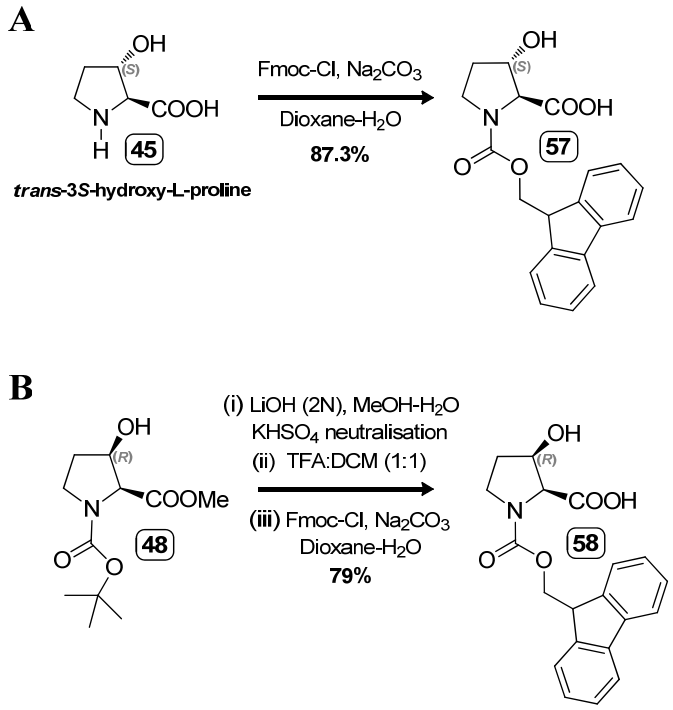


SCHEME 3.2: Synthesis of (2*S*,3*R*)-*N*$^{\alpha}$-(t-butoxycarbonyl)-3-(9-fluorenylmethyloxycarbonylamino) proline

hydrogenation of **54** with 10% Pd/C-H_2 to yield compound **55**, confirmed from the loss of IR-band characteristic of azido group at 2111 cm^{-1}. The β-amino acid **55** was reacted with fluorenylmethylchloroformate in dioxane:water with Na_2CO_3 as base to yield the fully protected monomer **56** (Scheme **3.2**). ^{1}H NMR signals in the aromatic region, typical of Fmoc group [δ 7.32-7.35 (m, 2H), 7.40-7.44 (m, 2H), 7.57-7.63 (m, 2H), 7.75-7.80 (m, 2H); Ar<u>H</u> of Fmoc] and LC-MS peaks at 453 [M+H] and 475 [M+Na] confirmed the formation of orthogonally protected (2*S*,3*R*) 3-aminoproline monomer.

3.9.1b Synthesis of Nα-Fmoc protected (2*S*,3*S*) and (2*S*,3*R*)-hydroxyproline monomers

Direct Fmoc-protection of pyrrolidine ring nitrogen of *trans*-3*S*-hydroxy-L-proline **45** using Fmoc-Cl in dioxane:water with Na_2CO_3 as base resulted the synthesis of *N*$^{\alpha}$-Fmoc-(2*S*,3*S*)-aminoproline monomer **57** (Scheme **3.3 A**). The ^{1}H NMR signals (δ 7.26-7.74, 4m, 8H; Ar<u>H</u> of Fmoc) together with LC-MS peaks 354.24 [M+1] and 376.17 [M+Na] confirm the structural integrity of the monomer.



SCHEME 3.3: (**A**) Synthesis of (2*S*,3*S*)-N^{α}-(9-fluorenylmethyloxycarbonylamino)-3-hydroxyproline, (**B**) Synthesis of (2*S*,3*R*)-N^{α}-(9-fluorenylmethyloxycarbonylamino)-3-hydroxyproline

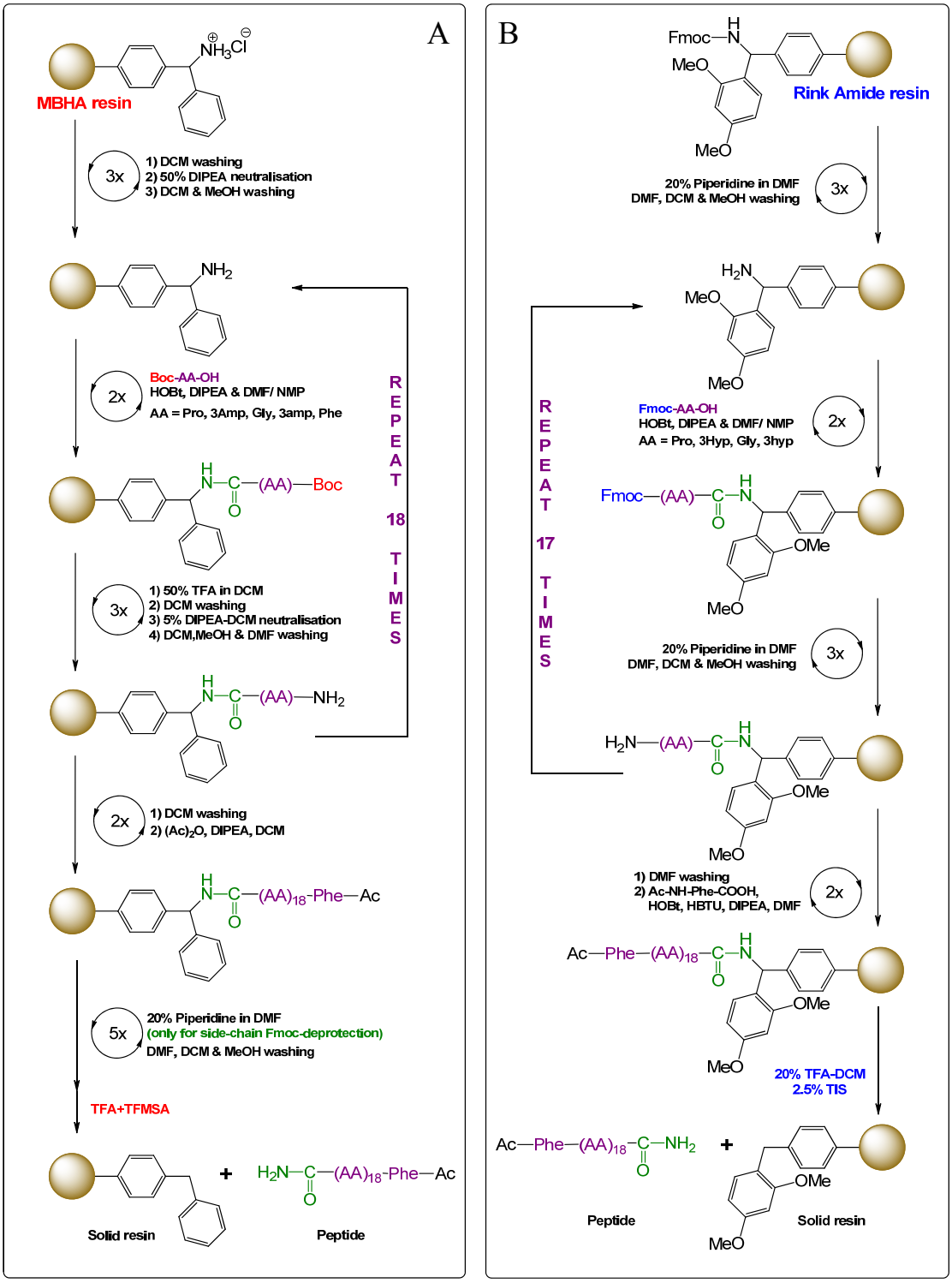
The synthesis of N^{α}-Fmoc-(2*S*,3*R*)-aminoproline monomer **58** (Scheme **3.3 B**) was done from protected *cis*-3*R*-hydroxyproline **48**. The methyl ester was hydrolysed with aqueous NaOH (2N) followed by removal of N^{α}-Boc group with 50% TFA in DCM, which was then reacted with fluorenylmethylchloroformate in dioxane:water with Na_2CO_3 as base to yield **58** with similar characteristic peaks in LC-MS (354 & 376) and ^{1}H NMR spectrum (δ 7.28-7.79, 4t, 8H; ArH of Fmoc).

3.9.2 Solid Phase Peptide Synthesis

The synthesis of (2*S*,4*R*)-aminoproline (4-Amp) and (2*S*,4*S*)-aminoproline (4-amp) monomers with N^{α}-*t*-Boc and N^{γ}-Fmoc protection and its use in the solid-phase synthesis of peptides using *t*-Boc chemistry has been described in Section I. Following similar procedure, the synthesized (2*S*,3*S*)-aminoproline (3*S*-Amp) and (2*S*,3*R*)-aminoproline (3*R*-amp) with N^{α}-*t*-Boc and N^{β}-Fmoc protection were incorporated into X and Y position of collagen sequence respectively using the *t*-Boc chemistry. The synthesis of peptides **P13** and **P14** (Scheme **3.4 A**) was carried out on MBHA resin, which upon cleavage directly yielded the peptide-*C*-terminal amide. Coupling reactions were carried out using 3 eq. of amino acid with HBTU as coupling reagent and HOBt as racemization-suppressant along with catalytic amount of

DIPEA. The *t*-Boc deprotection was carried out using 1:1 mixture of DCM:TFA followed by 5% DIPEA-DCM neutralization. Couplings were monitored using a combination of Kaiser test[69] (for Gly) and chloranil[70] tests (for imino acids Pro and 3(*R/S*)-aminoprolines). In order to avoid the deletion sequences resulting from the incomplete coupling reactions, whenever necessary the couplings were repeated and/or a capping step with Ac_2O/DIPEA/DCM was performed. A capping step was included to truncate failed sequences from growing, which indirectly helps in simplifying the purification procedure.

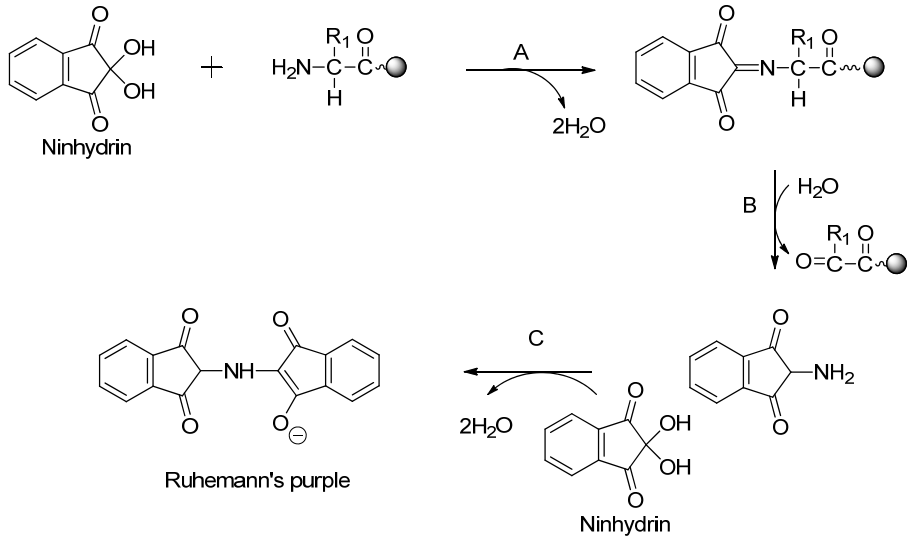
The synthesis peptides **P11** and **P12** (Scheme **3.4 B**) was carried out on a resin with Rink-amide linker. N^{α}-(Fmoc)-3-hydroxyproline isomers 3*S*-Hyp and 3*R*-hyp (monomers) were successfully incorporated into peptides by solid-phase method, without the protection of 3-hydroxyl group. The resin bound Fmoc group was cleaved with 20% piperidine/DMF and the monomers were coupled as free acids using an *in situ* activation procedure with HBTU as coupling reagent and HOBt/DIPEA as catalysts. The test for deprotection and coupling reactions were monitored analogous to previous Boc-strategy. In order to cap the terminal amino group in final peptide N-acetylated phenylalanine was coupled as the last amino acid of the sequence. No capping reaction with Ac_2O was carried to avoid O-acetylation on free hydroxyl groups. The peptide was cleaved from the resin using 20% TFA in DCM and 2.5% TIS.



SCHEME 3.4: Schematic representation of solid phase synthesis; (**A**) *t*-Boc strategy used for synthesis of **P13** [Ac-Phe-(Pro-3(*S*)Amp-Gly)$_6$-NH$_2$] and **P14** [Ac-Phe-(3(*R*)amp-Pro-Gly)$_6$-NH$_2$] peptides; (**B**) Fmoc- strategy used for peptide **P11** [Ac-Phe-(Pro-3(*S*)Hyp-Gly)$_6$-NH$_2$]and **P12** [Ac-Phe-(3(*R*)hyp-Pro-Gly)$_6$-NH$_2$].

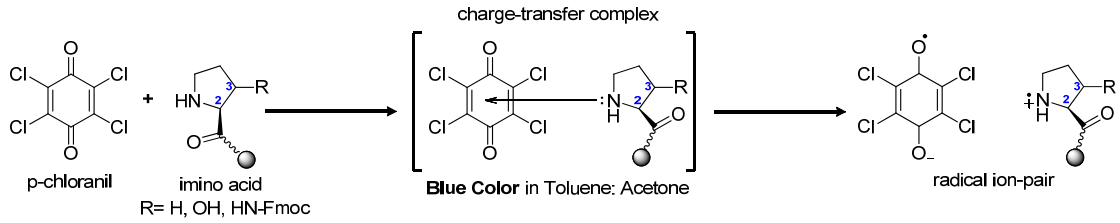
Kaiser's test is the most widely used qualitative test for the presence or absence of free amino groups (deprotection/coupling). It is used in monitoring the *t*-

Boc deprotection and amide bond (peptide bond) formation steps in solid phase peptide synthesis. The *t*-Boc-deprotection step leads to a positive Kaiser's test, where in the resin beads as well as the solution are blue in color (Rheumann's purple). On the other hand, upon completion of the coupling reaction, the Kaiser's test is negative, the resin beads remaining colorless (Scheme **3.5**).



SCHEME 3.5: (**A**) Reaction of ninhydrin (trioxohydrindene hydrate) with the amino group of a bound residue generates the Schiff's base. (**B**) Hydrolysis generates the aldehyde and another amine. (**C**) Amine reacts with a second molecule of ninhydrin to give an equilibrium mixture of the anion depicted and its tetraoxo form with a maximum of absorbance at 570 nm

Alternatively, Chloranil[70] or De Clercq[71] tests (Scheme **3.6**) are also used which detects secondary amine. The probable mechanism of the process is through charge-transfer complex between chloranil molecule and free amine of the imino acids.



SCHEME 3.6: Proposed mechanism for the reaction between free amine group of peptide on solid phase and p-chloranil to produce charge-transfer complex

The following peptides (Figure **10**), were synthesized for the present study by incorporating the modified aminoacids (2*S*,3*S*)-aminoproline **52** and (2*S*,3*R*)-

aminoproline **56**, at **Y** and **X** positions of the collagen model peptide $(X\text{-}Y\text{-}Gly)_n$. The peptides Ac-Phe(Pro-3*(S)*Amp-Gly)$_6$-NH_2 (**P13**) and Ac-Phe(Pro-3*(S)*Hyp-Gly)$_6$-NH_2 (**P11**) contain 3*S*-aminoproline (3*S*-Amp) and 3*S*-hydroxyproline (3*S*-Hyp) at **Y** position respectively, while peptides Ac-Phe(3*(R)*amp-Pro-Gly)$_6$-NH_2 (**P14**) and Ac-Phe(3*(R)*hyp-Pro-Gly)$_6$-NH_2 (**P12**) contain 3*R*-aminoproline (3*R*-amp) and 3*R*-hydroxyproline (3*R*-hyp) at **X** position.

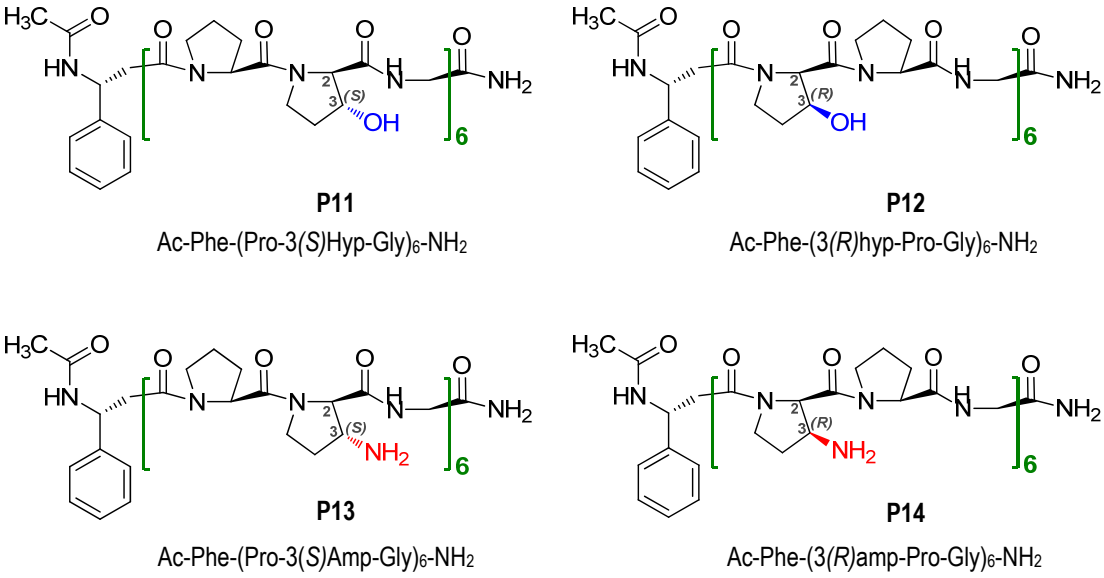


FIGURE 10: End capped peptides used in the present study. **P11-P12** peptides on Rink Amide resin *via* Fmoc-chemistry, **P13** and **P14** peptides synthesized *via t*-Boc chemistry on MBHA resin

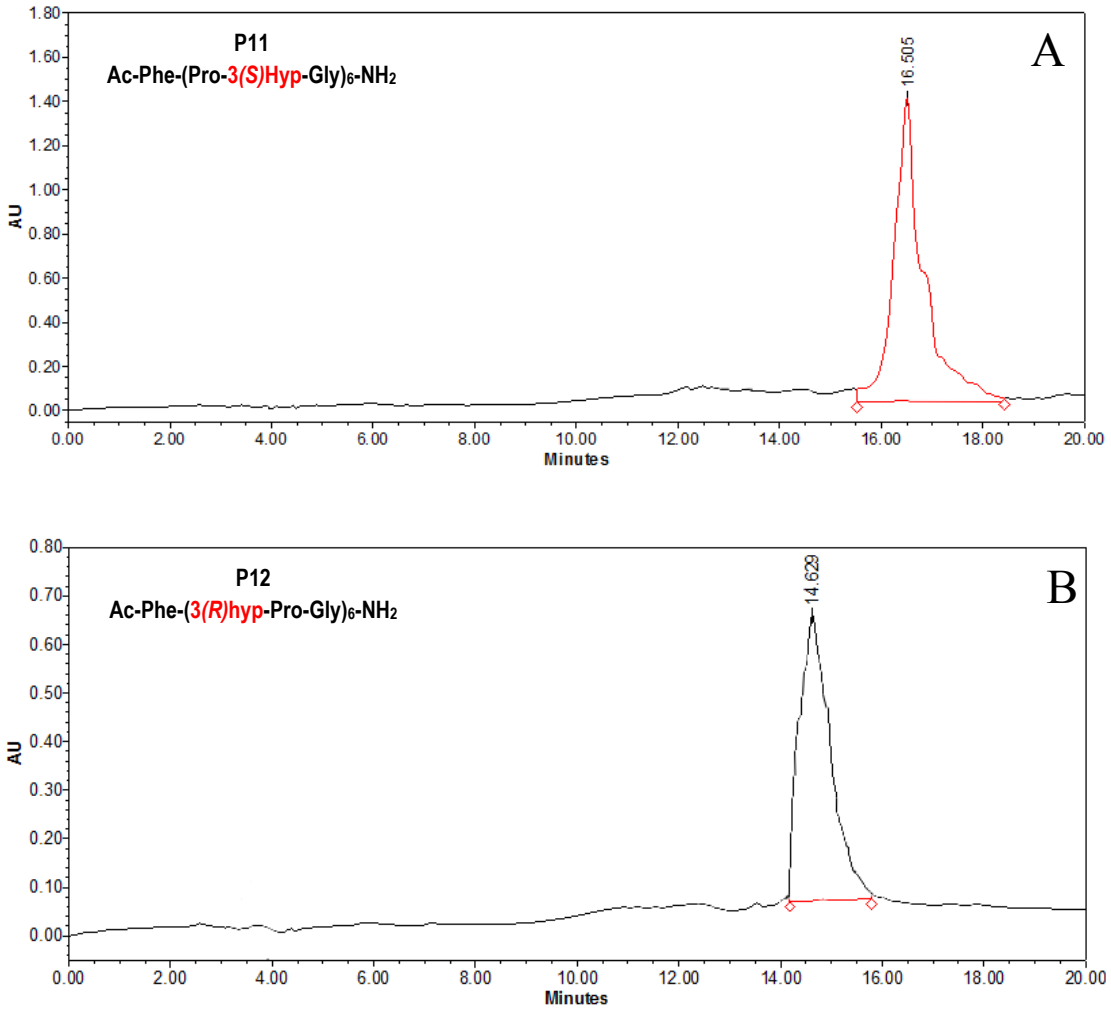
The structural integrity of peptides was confirmed by MALDI-TOF mass spectrometric analysis. The HPLC retention time of the synthesized peptides is shown in Table **2**. The observed molecular weight has been mentioned in the same table along with the calculated molecular weight and the molecular formula of all the peptides. The HPLC profile and MALDI-TOF data for confirmation of these synthesized peptides are shown later in Figure **11** and **12**.

Table 2: HPLC retention time and MALDI-TOF mass spectral analysis of the synthesized peptides

Peptide	Sequence Code	HPLC t_R (min)	Mol. Formula	$M_{(Calcd)}$*/ $M_{(Obsd)}$
P11	3*S*-Hyp	16.50	$C_{83}H_{116}N_{20}O_{26}$	1809.9275 /1832.0874 $[M+Na]^+$ 1848.0579 $[M+K]^+$
P12	3*R*-hyp	14.69	$C_{83}H_{116}N_{20}O_{26}$	1809.9275 /1831.5874 $[M+Na]^+$ 1847.5479 $[M+K]^+$
P13	3*S*-Amp	16.21	$C_{83}H_{122}N_{26}O_{20}$	1804.0190 /1827.6000 $[M+Na]^+$ 1827.6086 $[M+K]^+$
P14	3*R*-amp	17.02	$C_{83}H_{122}N_{26}O_{20}$	1804.0190 /1801.7133

* All the molecular weight has been calculated by Chem BioDraw 12.0

The N-terminal acetylated and C-terminal amidated peptides (**P11**-**P14**) were purified on semi-preparative RP-18 HPLC column using water acetenitrile gradient to which 0.1% TFA added as ionizer. Figure **11** gives the HPLC profile for the above peptides.



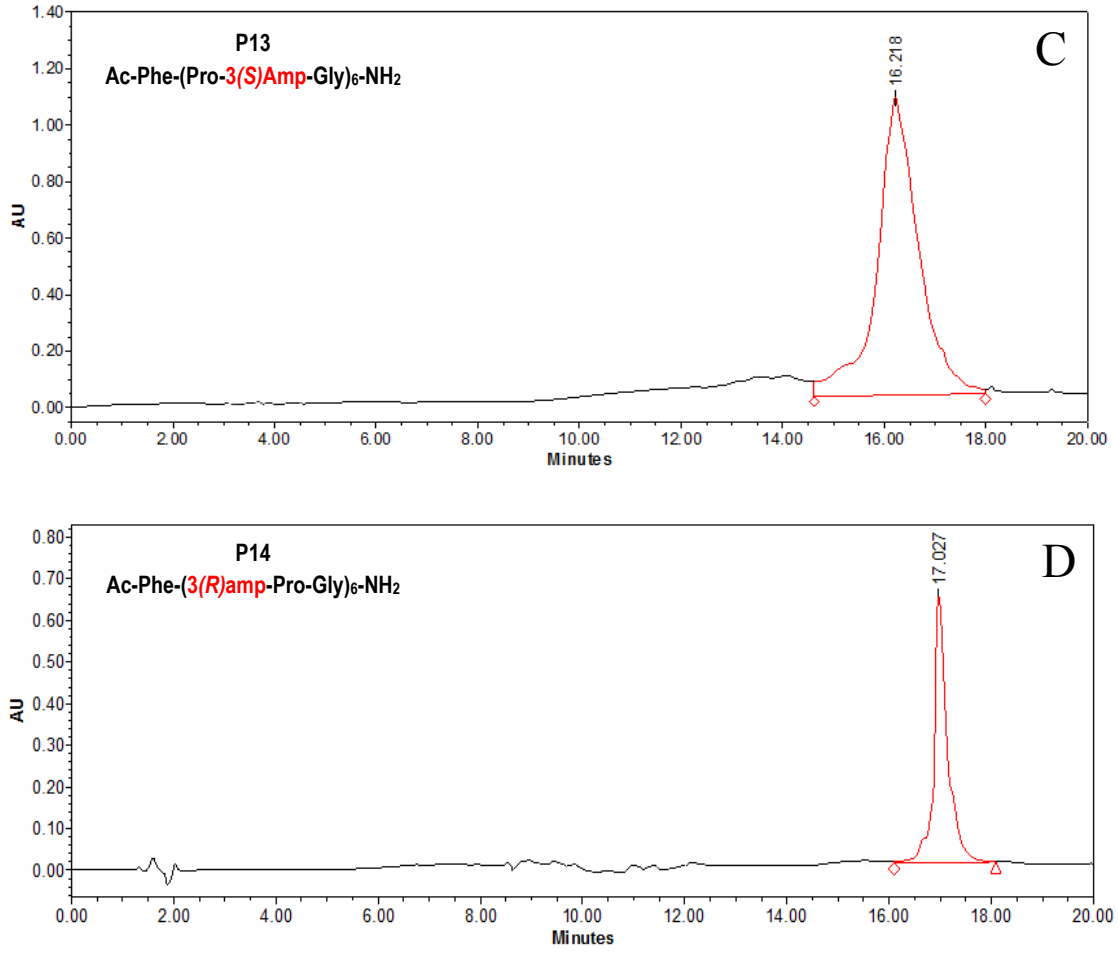
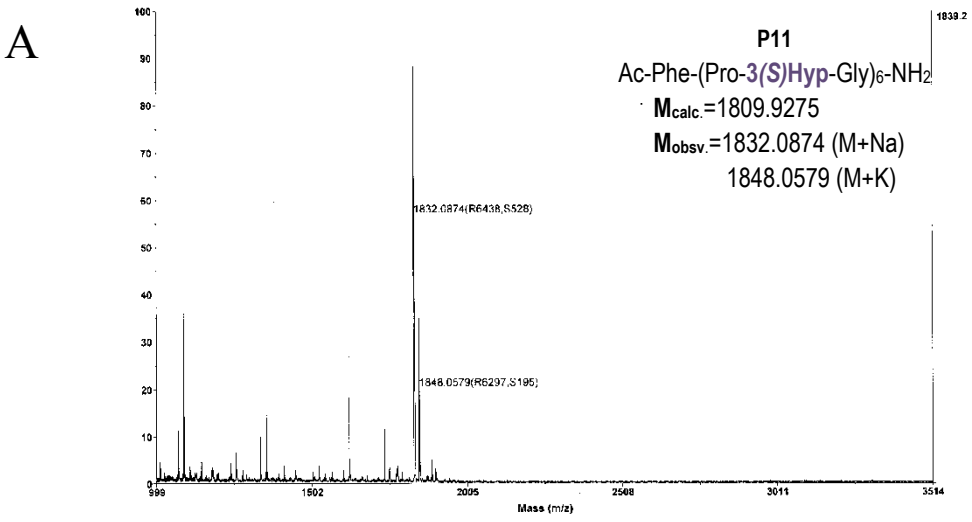
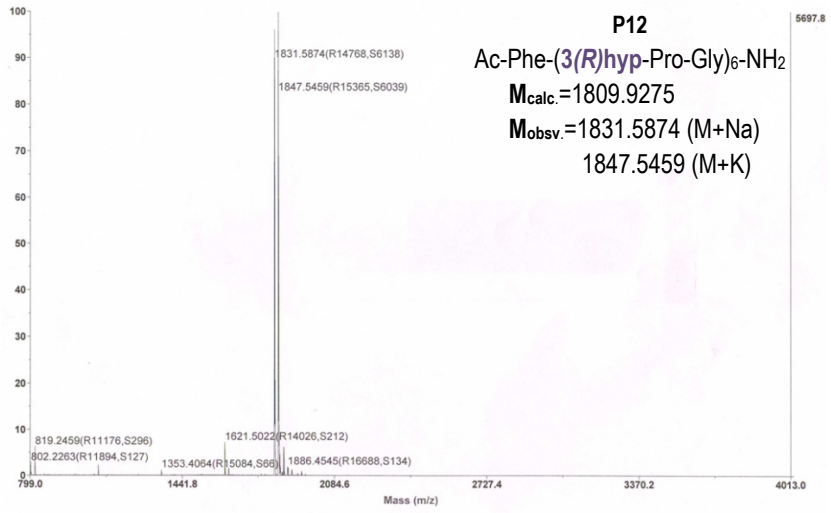


FIGURE 11: HPLC Profiles of peptides(**A**) **P11**, (**B**) **P12**, (**C**) **P13** and (**D**) **P14**; The conditions used in HPLC are 20 min Gradient: 5% to 80% MeCN/H_2O with 0.1% TFA,, flow rate = 3 ml/min; λ = 260 nm

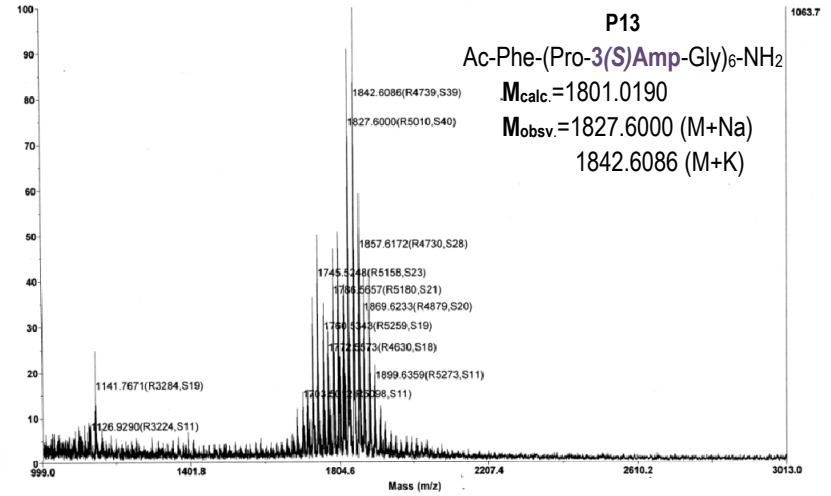
The MALDI-TOF spectra recorded for the peptides reported in this section, CHCA (α-cyano-4-hydroxycinnamic acid) was used as matrix and the molecular mass obtained for them were within the calculated range are shown in Figure **12**.



B



C



D

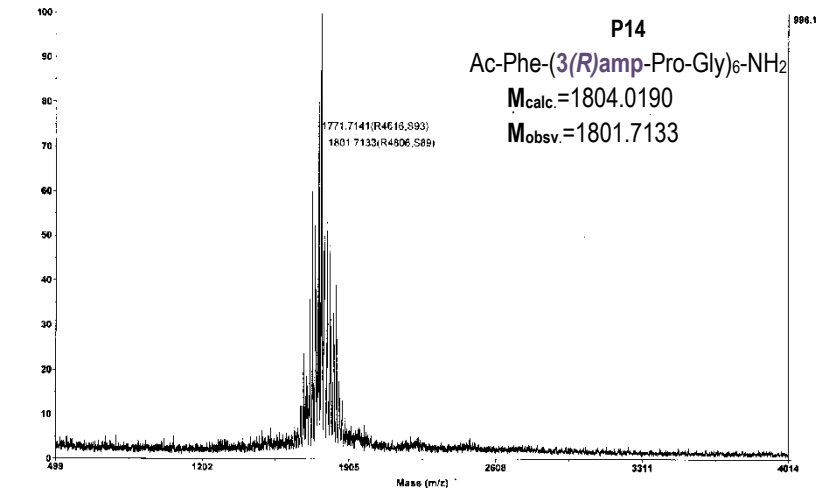


FIGURE 12: MALDI-TOF spectra of peptides (**A**) **P11**, (**B**) **P12**, (**C**) **P13** and (**D**) **P14**

3.9.3 Conformational study by CD spectroscopy

Circular dichroism (CD) spectroscopy is one of the most commonly employed tool in the characterization of triple-helical structures. In solution, collagen-like triple-helical and polyproline II like structures exhibit fingerprint CD spectra.[19,72] These spectra are characteristic by the presence of a large negative band around 200 nm, a crossover at around 213 nm, and a small positive band around 215-227 nm. In order to investigate the conformational behavior of peptides **P11**-**P14**, CD spectra were measured at various concentrations, temperatures and solvent conditions.

3.9.3a Concentration dependent CD spectroscopy

Formation of aggregated structures such as collagen triple-helix (a ternary association) is a concentration dependent phenomenon. Goodman *et al.*,[73] have shown that higher concentrations of peptide increase the percentage of triple-helical complex in solution. The triple-helical content for collagenous peptides is maximal when the concentration is greater than the critical triple-helical concentration.

The magnitude of the ratio of positive to negative band intensity in the CD spectra of collagen peptides ($R_{p/n}$) has been proposed to quantitate the triple-helical strength.[73] In the present study, this parameter has been used to determine the concentration dependent triple-helix formation of 3(*S*)Hyp containing peptide **P11** at pH 7.2 in presence of 0.02 M NaCl. Figure **13.A** shows CD spectra of 3(*S*)Hyp peptide **P11** taken at 10 °C in the concentration range of 0.05-0.25 mM. These spectra, in the entire concentration range show identical positive and negative maxima at 225 nm and 201 nm respectively. Importantly, all the spectral traces pass through an isobestic point at 221 nm. Figure **13.B** shows a plot of $R_{p/n}$ values derived from these spectra against concentration of **P11** at pH 7.2. Despite low $R_{p/n}$ values, this increases stepwise from 0.05 mM through 0.10 mM to reach a near saturation at 0.2 mM and remains nearly constant thereafter. A critical coiled-coil aggregation concentration of ~0.2 mM is derived from this plot. Hence, all further studies were performed at a concentration of 0.2 mM peptides assuming peptide **P11** is in fully associated form at this concentration.

Figure **13.C** shows the CD spectra with varying concentrations (0.05 mM-0.25 mM) of peptide **P13** –(Pro-3(*S*)Amp-Gly)$_6$– with 3(*S*)Amp at Y position of collagen sequence. In the entire concentration range the peptide **P13** displays the CD spectra

with identical positive band at 223.5 nm and negative band at 201 nm. Like **P11** this peptide also shows isobestic point at 219 nm.

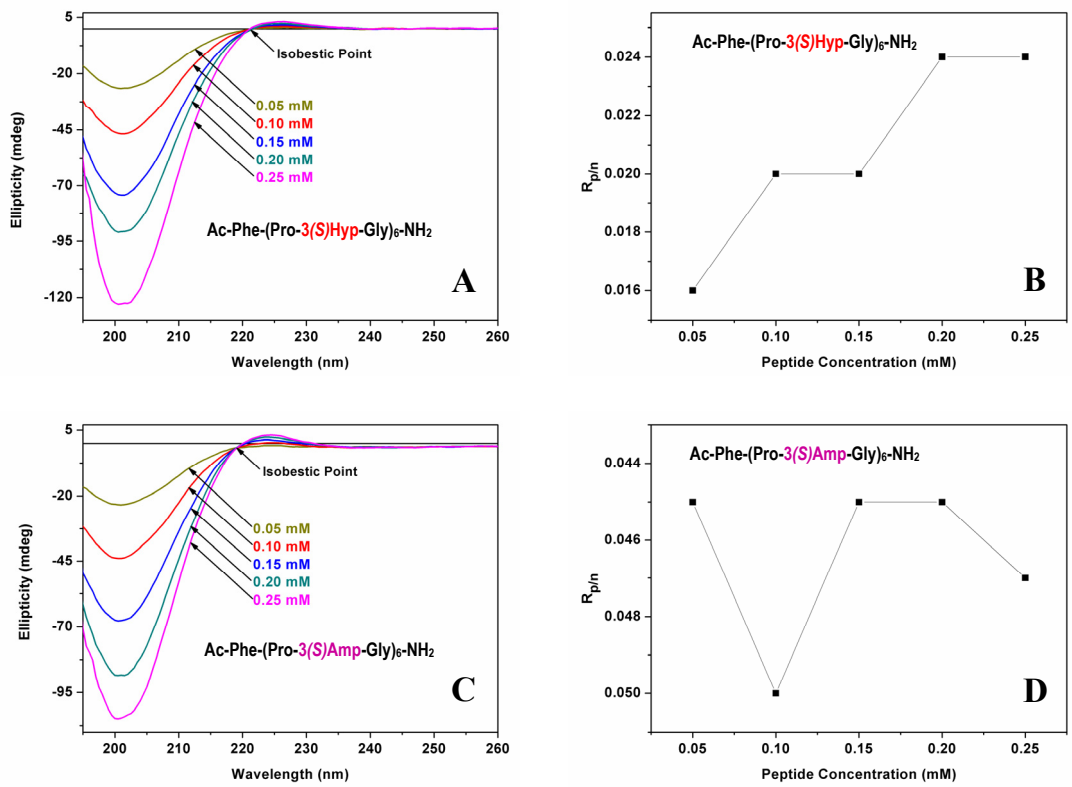


FIGURE 13: CD spectra at 10 °C at concentrations from 0.05 mM-0.25 mM in steps of 0.05 mM (20 mM phosphate buffer pH 7.2, 10 mM NaCl) of peptides (**A**) **P11**, λ_{max} = 225 nm, λ_{min} = 201 nm, isobestic point at 221 nm; (**C**) **P13**, λ_{max} = 223.5 nm, λ_{min} = 201 nm, isobestic point at 219 nm; The plot of $R_{p/n}$ values deduced from these spectra against the concentration of (**B**) **P11** and (**D**) **P13**.

In this case also an attempt was made to determine the minimum aggregation concentration for the peptide **P13** by plotting $R_{p/n}$ *versus* the concentration (Figure **13.D**). But it was found that unlike **P11** –(Pro-3(*S*)Hyp-Gly)$_6$– the $R_{p/n}$ values do not increase progressively with concentration for **P13**.

Peptides **P12** –(3(*R*)hyp-Pro-Gly)$_6$– and **P14** –(3(*R*)amp-Pro-Gly)$_6$– however had a slightly different CD spectra. Figure **14 (A-B)** shows the CD spectra taken at concentrations 0.2 mM and 0.5 mM, which do not have any positive band in 215-227 nm range although there is a large negative band in 200-205 nm region. Since these peptides did not show any +ve band even at high concentrations of 0.5 mM, the concentration dependent plot of $R_{p/n}$ could not be deduced and so the formation of any aggregated triple helical structure was also ruled out.

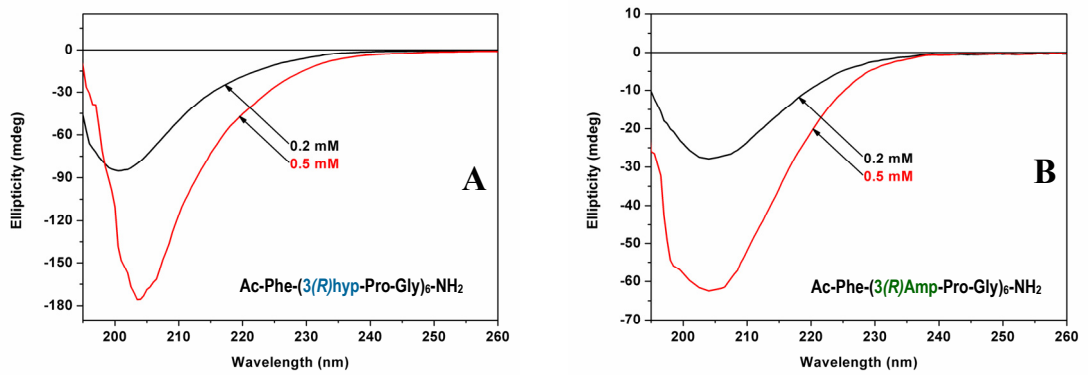


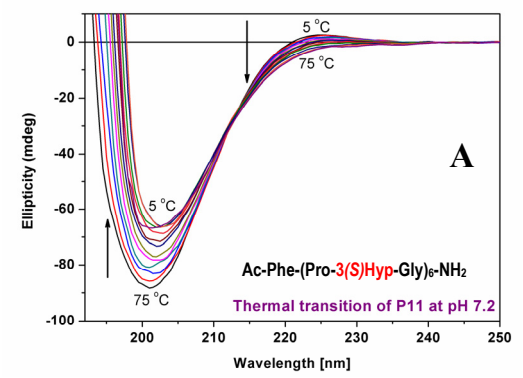
FIGURE 14: CD spectra at 10 °C at concentrations at 0.2 mM & 0.5 mM (20 mM phosphate buffer pH 7.2, 10 mM NaCl) of peptides (**A**) **P12**, λ_{max} = -nd-, λ_{min} = ~200 nm, (**B**) **P14**, λ_{max} = -nd-, λ_{min} = 205 nm

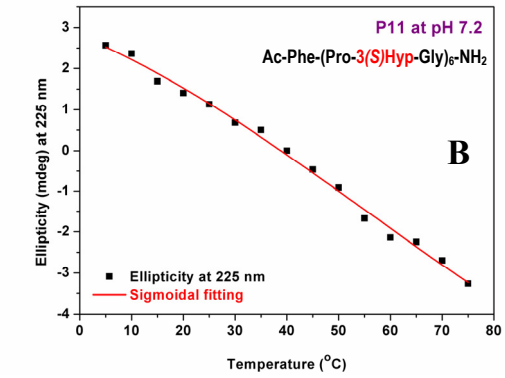
3.9.3b Thermal denaturation of peptides

It was shown in the previous section that peptides of 3(*S*)Hyp **P11** and 3(*S*)Amp **P13** show collagen-like CD spectra while the –(3(*R*)hyp-Pro-Gly)$_6$– **P12** and –(3(*R*)amp-Pro-Gly)$_6$– **P14** peptides do not. Conversely, appearance of collagen-like CD spectrum is neither a sufficient evidence for the presence of triple-helical structures, nor gives any information about the triple-helical strength. In order to determine the relative triple-helical strengths of peptides **P11** and **P13**, CD-thermal denaturation studies were carried out in phosphate buffer with 0.02 M NaCl salt and pH12.0.

Upon heating, triple-helical structures undergo a cooperative triple-helix ⇔ coil transition. This change of conformation is reflected in several spectroscopic properties including a decrease in the molar ellipticity - typically around 225 nm. Therefore, the triple-helix ⇔ coil transition was monitored by observing the change in ellipticity at 225 nm with temperature. The thermal denaturation curves so obtained show a decrease in molar ellipticity at 225 nm with increase in temperature for all peptides. Normalized ellipticity data at 225 nm plotted against temperature is shown in Figures **15** & **16**.

At pH 7.2 peptide **P11** shows cooperative melting transition curves (Figure **15.A**) with increasing temperature. This indicates that peptide –(Pro-3(*S*)Hyp-Gly)$_6$– **P11** is associated in a coiled-coil structure at physiological pH. A 53.8 ^{o}C T_m value is obtained from the integration of thermal denaturation curves.





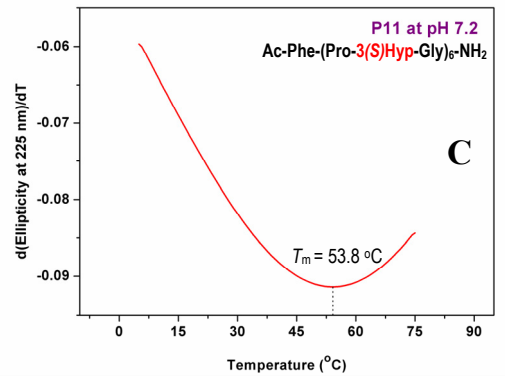
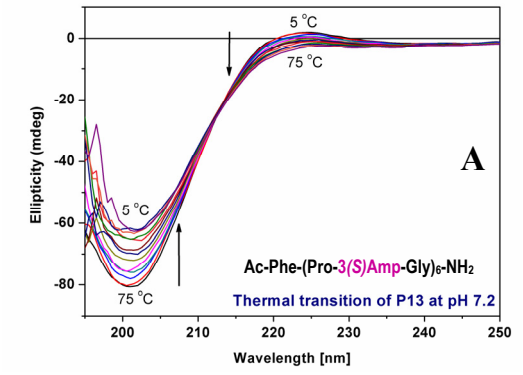
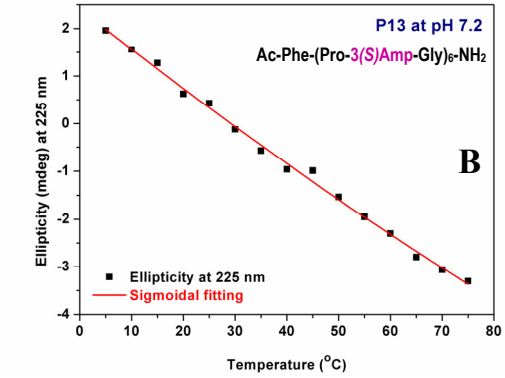


FIGURE 15: CD spectra at (**A**) 5-75 °C, (**B**) Boltzman/Sigmoidal fit curve for molar ellipticity at 224 nm (**C**) First derivative curve of 0.2 mM solutions of peptide **P11** in 20 mM phosphate buffer pH 7.2, 10 mM NaCl

Similar thermal denaturation of peptide **P13** with 3(*S*)Amp at Y position brings about only a linear decrease in ellipticity with increasing temperature that is a result of single chain melting as can be seen from Figure **16 (B-C)**.





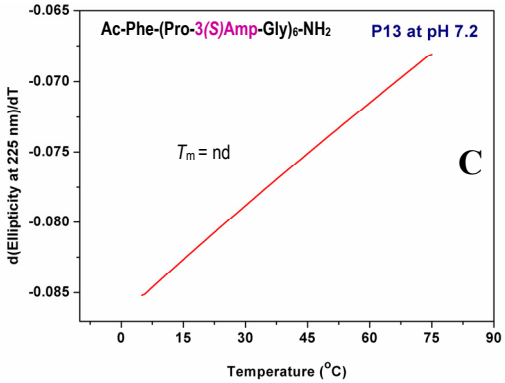


FIGURE 16: CD spectra at (**A**) 5-75 °C, (**B**) Boltzman/Sigmoidal fit curve for molar ellipticity at 224 nm (**C**) First derivative curve of 0.2 mM solutions of peptide **P13** in 20 mM phosphate buffer pH 7.2, 10 mM NaCl

3.9.3c Behaviour of peptides in ethylene glycol

Polyols and sugars are known to offer protection to most proteins,[74] including collagen triple-helix against thermal denaturation.[75] Ethylene glycol (EG) stabilizes helical structure and therefore can be very useful to amplify and detect very weak triple-helical propensities.[73,76]

Figure **17.A** shows the CD spectra of 0.2 mM solution the **P11** taken in a 3:1 v/v mixture of EG:Water. In comparison to aqueous solution the spectrum of peptide –(Pro-3(*S*)Hyp-Gly)$_6$– **P11** is characterized by the appearance of significant negative band at 201 nm (Θ = 89.36 mdeg) and a good increase in the molar ellipticity of the positive band (224 nm, Θ = 5.67 mdeg) with $R_{p/n}$ value of 0.06. The thermal transition (Figure **17 (B-D)**) of this peptide EG:Water system showed opposite behavior than the aqueous solutions. The derivative plot of change in ellipticity with respect to change in temperature for peptide **P11** gave a comparatively lower melting temperature of 35.9 °C (T_m).

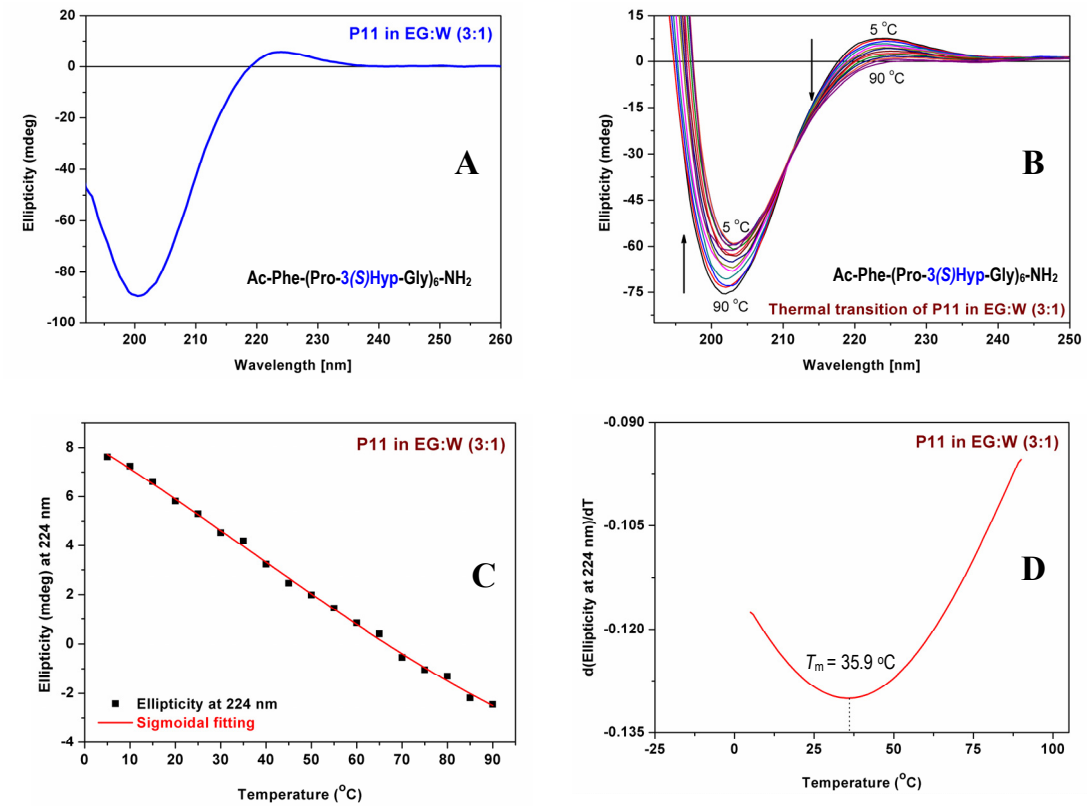


FIGURE 17: CD spectra at (**A**) 10 °C, (**B**) 5-90 °C, (**C**) Boltzman/Sigmoidal fit curve for molar ellipticity at 224 nm (**D**) First derivative curve of 0.2 mM solutions of peptide **P11** in 3:1 (v/v) EG:Water mixture.

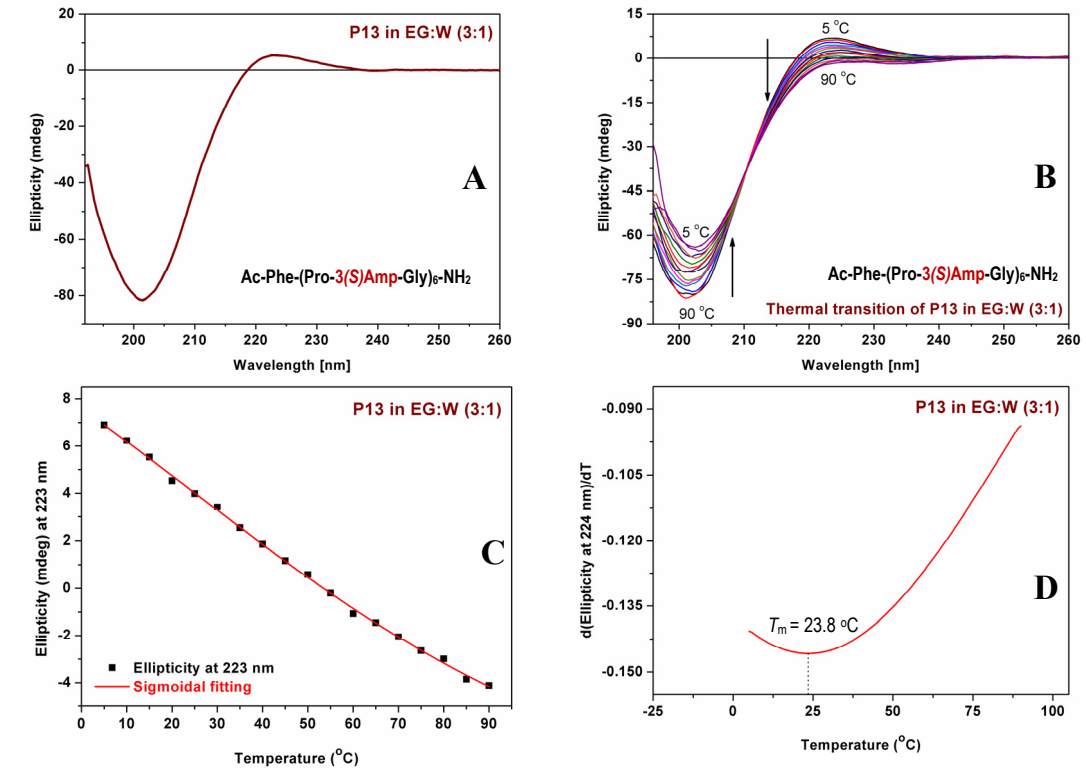


FIGURE 18: CD spectra at (**A**) 10 °C, (**B**) 5-90 °C, (**C**) Boltzman/Sigmoidal fit curve for molar ellipticity at 223 nm (**D**) First derivative curve of 0.2 mM solutions of peptide **P13** in 3:1 (v/v) EG:Water mixture.

On the otherhand peptide –(Pro-3(*S*)Amp-Gly)$_6$– **P13** shows better stability in mixture of EG:Water solvent system in context to aqueous phosphate buffer. This peptide too had a good increase in the molar ellipticity of the positive band (223 nm, Θ = 5.30 mdeg) and the same large negative band at 201 nm (Θ = 81.46 mdeg) giving same 0.06 $R_{p/n}$ value. The T_m obtained from the derivative plot is less than 25 °C (23.8 °C). These results are depicted in the above Figure **18 (A-D)**.

Summarizing the melting temperatures (T_m values) obtained from the various thermal transitions carried for peptides **P11** and **P13**, the following Table **3** was deduced.

Table 3: Thermal transition temperature of peptides in both aq. and ethylene glycol systems

Peptide Sequence		T_m (°C)	
		pH 7	EG:W
Ac-Phe-(Pro-3(*S*)Hyp-Gly)$_6$-NH$_2$	**P11**	58.3 °C	35.9 °C
Ac-Phe-(Pro-3(*S*)Amp-Gly)$_6$-NH$_2$	**P13**	-nd-	28.3 °C

3.10 Discussion

The chimeric collagen 3(*S*)Hyp peptide **P11** and 3(*S*)Amp peptide **P13** at pH 7.2 forms polyproline II like structure as indicated by a positive maximum to strong negative band at 225 /201 nm and 223 /201 nm in the CD-spectra respectively. The peak positions are nearly independent of the concentration and most importantly, all the spectral traces pass through the same isobestic point at 221 nm for **P11** and 219 nm for **P13** (Figure **13(A-B)**). These CD spectral properties indicate that within the concentration range measured i.e., 0.05-0.25 mM peptides exists in the same conformational state. A progressive increase in the $R_{p/n}$ values of **P11** with concentration also suggests a two state equilibrium, though very low when compared to minimum acceptable value for a triple helix.[73] The observed sigmoid transition in the variable temperature CD measurements provides additional evidence for the two state super-coil⇔coil transitions. While this doesn't hold true for **P13** with random variation of $R_{p/n}$ which also shows linear transition in temperature dependent CD spectra. Thus **P13** can be said to be in only one kind of conformation.

However this result of 3(*S*)Hyp in X position of collagen sequence $(Gly\text{-}X\text{-}Y)_n$ forming a stable supercoil/triplex kind of structure does not relate to previous reports in literature by different groups Raines *et al.*[35c] and Mizuno *et al.*.[77] But the evidence of supercoil structure from CD can be attributed to design of sequence. The **P11** has six 3(*S*)Hyp in X position of the sequence Ac–Phe–$(Pro\text{-}3(S)Hyp\text{-}Gly)_6$–$NH_2$ and no 4(*R*)-Hyp in its proximity, while the sequences analysed in above reports are of host-guest type *viz.* H-$(Gly\text{-}3(S)Hyp\text{-}4(R)Hyp)_9$-OH, $(Pro\text{-}4\text{-}Hyp\text{-}Gly)_3$-Pro-3(*S*)Hyp-Gly-$(Pro\text{-}4\text{-}Hyp\text{-}Gly)_3$ *etc*. can have interstrand clashes.

On the contrary **P13** with 3(*S*)Amp in equivalent sequence behaves differently without any conformational transitions probably due to intrastrand H-bonding which does not allow it to form any supercoils.

Again having substitution at Y position of chimeric collagen sequences $(Gly\text{-}X\text{-}Y)_n$ with 3(*R*)hyp in peptide **P12** and 3(*R*)amp as a part of peptide **P14** at pH 7.2 offered polyproline kind of CD spectra with no positive band.[78,79] This result is linked to its pyrrolidine ring pucker leading to inappropriate mainchain dihedral angles.

Further the 3(*S*)Hyp peptide **P11** in EG:W solvent system showed enhanced $R_{p/n}$ value but a lower T_m (35.9 ^oC) suggested decrease in the stability of supercoiled structure. Whereas **P13** peptide with 3(*S*)Amp displayed some change in

conformation evidenced by a broad melting transition at 23.8 ^{o}C (T_m) in CD spectra recorded in EG:W.

Suggested explanation for totally opposite behavior of **P11** and **P13** peptides in ethylene glycol can be because of intramolecular peptide (amide) hydrogen bonding coming to picture in the absence of accessible competing water which generally lengthens[80] the peptide hydrogen bonds by loosening the structure. Polyols like ethylene glycol offer stability to peptides/proteins depending on H-bonds compared to aqueous buffer condition.[75] These polyols generally contain hydrophilic carbon chains that prevent the water from entering into the main chain domain of proteins, thereby enhancing the hydrogen bond strength between amide hydrogen and main chain carbonyl oxygen atoms.

3.11 Conclusions

In conclusion, it is shown that the 3*S*-hydroxyproline residues in X position of collagen peptides $(Gly\text{-}X\text{-}Y)_6$ leads to stabilization of the derived supercoiled structure in aq. buffer while the 3*S*-aminoproline has weak triple helicity to be detected only in ethylene glycol.

More research needs to be done to verify these results in terms of effect of pH, salt concentration, peptide length, end-group capping *etc.* to obtain better insight into this aspect.

Even then results have a direct bearing on the current interest in collagen structure and mimetics. Since, 3-OH group in natural collagen can potentially be transformed chemically to 3-NH_2, the properties of this analogue may have significant applications in the design of new collagen based biomaterials.[81]

3.12 Experimental

3.12.1 Synthesis of compounds/monomers

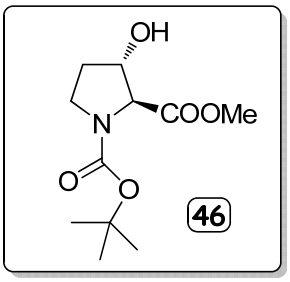
General remarks: All reagents and chemicals were of laboratory or analytical grade obtained from commercial sources and were used without further purification except in cases mentioned. Thin layer chromatography (TLC) was carried out on pre-coated silica gel 60 F_{254} plates (E. Merck). TLCs were visualized under UV light, iodine and/or ninhydrin spray, phosphomolybdic acid charring solution (only for N-acetylated compounds) followed by heating up to 110 ^{o}C with heat gun. Silica gel 60-120 and 100-200 mesh (Merck) was used for routine column chromatography with ethyl acetate/petroleum ether or dichloromethane/methanol mixture as elution solvent depending upon the compound polarity and chemical nature whereas for flash chromatography 230-400 mesh silica gel was used. All solvents were distilled under an inert atmosphere with appropriate desiccant. Reactions in aqueous medium and workup processes were done using double distilled water. Unless otherwise noted, all reactions were carried out at room temperature.

^{1}H NMR spectra were routinely recorded at 200MHz on a Bruker AC-200 instrument controlled by an Aspect 2000 computer. ^{13}C NMR and ^{13}C-DEPT spectra (at 50MHz) were recorded on the same instrument. The spectra were analyzed using ACD NMR processor software from ACD labs. For some compounds, NMR spectra were also recorded on 400MHz JEOL and 500MHz Bruker spectrometer. All chemical shifts are referenced with respect to TMS as internal standard and are expressed in δ-scale (ppm). Mass spectra were obtained by ESI-MS technique on AP-QSTAR spectrometer. Melting points of the samples were determined in open capillary tubes using Büchi Melting Point M-560 apparatus and are uncorrected. IR spectra were recorded on an Infrared Fourier Transform Spectrophotometer using chloroform, nujol or neat sample. The optical rotation values were measured on JASCO P2000 polarimeter.

Salts and reagents used in buffer preparation such as NaCl, NaH_2PO_4, Na_2HPO_4 and ethylene glycol etc. were obtained from Sigma-Aldrich and were of molecular biology reagent grade. Double distilled water was demineralized using Millipore MilliQ deionizer and was used for the preparation of buffers. In case of phosphate buffer, pH was adjusted using NaOH and HCl.

Procedures and spectral data:

(2*S*,3*S*)-N¹-(t-butoxycarbonyl)-3-hydroxyproline methylester (46)



A solution of trans-3*S*-hydroxyproline **45** (1 g, 7.6 mmol) in a solution of aq. NaOH (2N, 10 ml) and dioxane (10 ml) was cooled to 0 °C and Boc-anhydride (2 ml, 8.4 mmol) was added dropwise using addition funnel. The reaction mixture was stirred at 0 °C for 1 h, and dioxane was removed under reduced pressure. The residue was extracted with diethyl ether (5 x 25 ml) to remove unreacted $(Boc)_2O$. The aqueous layer was then vigorously stirred with ethyl acetate under ice-cold condition and acidified with saturated $KHSO_4$ solution to pH 3. The ethyl acetate layer was separated and the aqueous layer was further extracted into ethyl acetate (3 x 20 ml). The organic layers were pooled and washed with water followed by brine, dried over anhydrous Na_2SO_4 and concentrated under vacuum. Removal of ethyl acetate under reduced pressure, gave a white solid which was recrystalized with EtOAc/Petroleum ether.

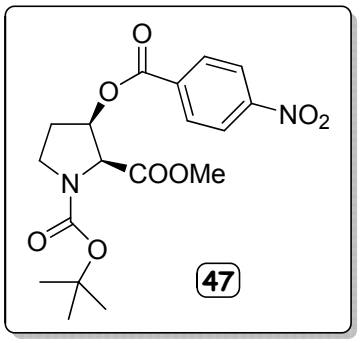
The white solid compound (1 g, 4.3 mmol) obtained in the above step was mixed with anhydrous K_2CO_3 (1.8 g, 13 mmol) and dissolved in anhydrous acetone (25 ml). The reaction mixture was stirred for 30 min at room temperature and dimethylsulphate (4.9 ml, 5.2 mmol) was added. The stirring was continued under reflux conditions for 5 h, when TLC confirmed maximum product formation. The solvent was evaporated under reduced pressure and the resulting solid was dissolved in ethyl acetate (50 ml), washed with water (2 X 30 ml), followed by saturated brine solution and dried over anhydrous Na_2SO_4. The ethyl acetate extract was concentrated under reduced pressure to a pale yellow solid which was then purified by silica gel chromatography, eluting with ethylacetate:pet-ether (1:1) to yield compound **46** as pale yellow liquid. Yield 1.48 g, 79.3% over the two steps.

Mol. Formula	:	$C_{11}H_{19}NO_5$
Mol. Weight	:	245.27
ESI-MS *m/z*	:	246.06 $[M+1]^{+\bullet}$, 268.03 $[M+Na]^{+\bullet}$
$[\alpha]_D^{25}$	:	-21.89 (C=1.05, $CHCl_3$)
^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.40 & 1.46 (2s, 9H), 2.01-2.16 (m, 2H), 3.53-3.66 (m, 2H), 3.73 (s, 3H), 4.16-4.29 (dd, 1H), 4.40-4.44 (m, 1H)
^{13}C NMR (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) 28.1(ma), 28.2(mi), 32.0(ma), 32.3(mi), 44.1(ma), 44.5(mi), 52.1(ma), 52.2(mi), 67.6(mi), 67.8(ma),73.8(mi),

74.9(ma), 154.0(ma), 154.6(mi), 171.4(mi), 171.8(ma)

^{13}C-DEPT : **δ_C** (ppm) *Positive Peaks:* 28.1(ma), 28.2(mi), 52.1(ma), 52.2(mi), 67.6(mi), 67.8(ma),73.8(mi), 74.9(ma) *Negative Peaks:* 32.0(ma), 32.3(mi), 44.1(ma), 44.5(mi)

(2*S*,3*R*)-N^1-(t-butoxycarbonyl)-3-(p-nitrobenzoyloxy) proline methylester (47)[82]



A mixture of compound **46,** (1g, 4.07 mmol), PPh_3, (2.35 g, 8.95 mmol) and *p*-nitrobenzoic acid (1.36 g, 8.14 mmol) were dissolved in dry THF (60 ml) cooled to 0 ^oC on ice bath, under Argon. After stirring for 30 min at ice-cold temperature Diisopropylazodicarboxylate (1.8 ml, 9 mmol) was added slowly with syringe. The reaction mixture was brought to RT and stirred for another 24 hrs after which THF was removed under vacuum. The resulting orange colored thick oil was dissolved in diethyl ether (50 ml). The resulting solution was kept overnight at room temperatureand the white crystals were filtered and the residue was washed with petroleum ether followed by diethyl ether. The filtrate was concentrated under reduced pressure and column purified with silica gel, which gave compound **47** as greasy pale yellow liquid. Yield 1.44 g (90.6%).

Mol. Formula : $C_{18}H_{22}N_2O_8$

Mol. Weight : 394.38

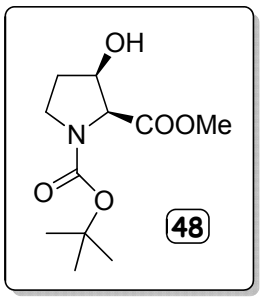
ESI-MS *m/z* : 417.07 $[M+Na]^{+\bullet}$

^{1}H NMR (200 MHz, $CDCl_3$) : **δ_H** (ppm) 1.42 & 1.48 (2s, 9H), 2.22-2.36 (m, 2H), 3.36 (s, 3H), 3.57-3.81 (m, 2H), 4.63-4.73 (dd, 1H, J=10.5), 5.69-5.78 (q, 1H), 8.13-8.17 (d, 2H), 8.27-8.32 (d, 2H)

^{13}C NMR (50 MHz, $CDCl_3$) : **δ_C** (ppm) 28.0(ma), 28.2(mi), 29.5(ma), 30.3(mi), 44.1(mi), 43.6(ma), 52.0, 61.6(ma), 61.2(mi), 73.7(mi), 74.5(ma), 80.5, 123.3(mi), 123.6(ma), 128.5(mi), 130.6(ma), 134.5, 150.6, 153.3(ma), 153.9(mi), 163.4, 169.1(mi), 169.4(ma)

^{13}C-DEPT : **δ_C** (ppm) *Positive Peaks:* 28.0(ma), 28.2(mi), 52.0, 61.6(ma), 61.2(mi), 73.7(mi), 74.5(ma), 123.3(mi), 123.6(ma), 128.5(mi), 130.6(ma) *Negative Peaks:* 29.5(ma), 30.3(mi), 44.1(mi), 43.6(ma)

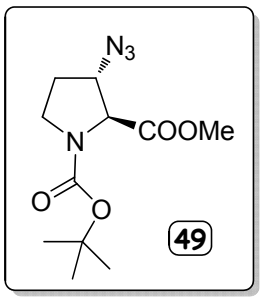
(2*S*,3*R*)N¹-(t-butyloxycabonyl)-3-hydroxyproline (48)[83]



A solution of **47** (1 g, 2.53 mmol) and sodium azide (60 mg, 9.36 mmol) in dry MeOH (70 ml) was stirred for 20 hrs at 40 °C under nitrogen. The solvent was removed on a rotary evaporator, and the compound was purified by column chromatography (ethyl acetate:hexane, 1:3 to 1:1) to give 0.54 g (96%) of a thick yellowish oil.

Mol. Formula	:	$C_{11}H_{19}NO_5$
Mol. Weight	:	245.27
ESI-MS *m/z*	:	246.24 $[M+1]^{+\bullet}$, 268.18 $[M+Na]^{+\bullet}$
$[\alpha]_D^{25}$	:	-28.33 (C=0.54, $CHCl_3$)
^{1}H NMR (200 MHz, $CDCl_3$)	:	δ_H (ppm) 1.34 & 1.39 (2s, 9H), 1.92-2.05 (m, 2H), 3.30-3.63 (m, 3H), 3.7 (s, 3H), 4.26-4.36 (dd, 1H), 4.54 (bm)
^{13}C NMR (50 MHz, $CDCl_3$)	:	δ_C (ppm) 28.0(ma), 28.2(mi), 32.0(ma), 32.4(mi), 43.7(ma), 44.1(mi), 51.8(ma), 52.0(mi), 63.2(mi), 63.9(ma), 71.1(mi), 72.0(ma), 80.1, 153.8(ma), 154.3(mi), 170.8(mi), 170.9 (ma)
^{13}C-DEPT	:	δ_C (ppm) *Positive Peaks:* 28.0(ma), 28.2(mi), 51.8(ma), 52.0(mi), 63.2(mi), 63.9(ma), 71.1(mi), 72.0(ma) *Negative Peaks:* 32.0(ma), 32.4(mi), 43.7(ma), 44.1(mi)

(2*S*,3*S*)N¹-(t-butyloxycabonyl)-3-azidoproline methylester (49)[84]

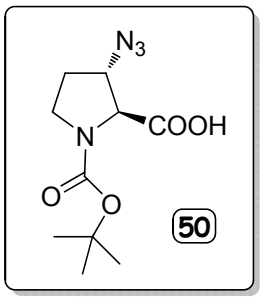


The compound **48,** (1 g, 4 mmol) and PPh_3, (1.3 g, 5 mmol) were dissolved in dry THF (40 ml) cooled to 0 °C on ice bath, under argon. After 30 mins of stirring at ice-cold temperature DIAD (1 ml, 5.1 mmol) and DPPA (1.2 ml, 5.3 mmol) were added slowly with syringe one after the other. The reaction mixture was brought to RT and stirred for another 8 hrs after which THF was removed under vacuum. The resulting orange colored thick oil was column purified with silica gel, which offered compound **49** as yellow liquid. Yield 0.7 g (64%).

Mol. Formula	:	$C_{11}H_{18}N_4O_4$
Mol. Weight	:	270.29
ESI-MS *m/z*	:	293.27 $[M+Na]^{+\bullet}$
$[\alpha]_D^{25}$	:	+35.03 (C=0.61, $CHCl_3$)
IR (thin film /neat)	:	ν (cm^{-1}) 3356, 2978, 2105, 1752, 1704, 1479, 1437, 1398, 1367, 1257, 1206, 1166, 1123, 1003, 901, 772, 559

^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.41 & 1.46 (2s, 9H), 1.93-2.04 (m, 1H), 2.09-2.25 (m, 1H), 3.47-3.65 (m, 2H), 3.76 (s, 3H), 4.12-4.18 (m, 1H), 4.19-4.35 (dd, 1H)
^{13}C NMR (100 MHz, $CDCl_3$)	:	**δ_C** (ppm) 27.7(ma), 27.8(mi), 29.0(ma), 29.7(mi), 43.8(ma), 44.1(mi), 52.0(ma), 52.1(mi), 63.0(mi), 63.9 & 64.3 (ma), 79.9, 152.9(ma), 153.6(mi), 170.1(mi), 170.5(ma)
^{13}C-DEPT	:	**δ_C** (ppm) *Positive Peaks:* 27.7(ma), 27.8(mi), 52.0(ma), 52.1(mi), 63.0(mi), 63.9 & 64.3 (ma) *Negative Peaks:* 29.0(ma), 29.7(mi), 43.8(ma), 44.1(mi)

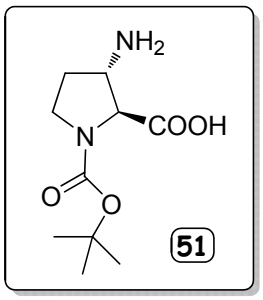
(2*S*,3*S*)N^1-(t-butyloxycabonyl)-3-azidoproline (50)



The methyl ester **49** (1 g, 3.7 mmol) was hydrolysed using aqueous LiOH (2N) in water-MeOH (1:1, 10 ml) mixture. This reaction mixture was stirred in RT for 2 hrs and methanol was removed under reduced pressure. The aqueous layer was washed with diethyl ether (3 x 10 ml) and then acidified with saturated $KHSO_4$ solution under ice-cold conditions. The desired compound **50** was extracted from the acidic solution with ethyl acetate and dried over Na_2SO_4. Free acid **50** was obtained as a gum upon concentration. Yield 0.94 g (~99%)

Mol. Formula	:	$C_{10}H_{16}N_4O_4$
Mol. Weight	:	256.26
ESI-MS *m/z*	:	257.18 $[M+1]^{+\bullet}$, 279.14 $[M+Na]^{+\bullet}$
^{1}H NMR (200 MHz, MeOH-d_4)	:	**δ_H** (ppm) 1.43 & 1.47 (2s, 9H), 1.96-2.06 (m, 1H), 2.09-2.28 (m, 1H), 3.43-3.64 (m, 2H), 4.13-4.22 (dd, 1H), 4.28-4.32 (m, 1H)
^{13}C NMR (100 MHz, MeOH-d_4)	:	**δ_C** (ppm) 28.6(ma), 28.8(mi), 30.4(ma), 31.1(mi), 45.4(ma), 45.8 (mi), 65.3(mi), 66.0(ma), 66.3(mi), 66.7(ma), 81.8(mi), 81.9(ma), 155.9(ma), 156.2(mi), 173.6(mi), 174.0(ma)
^{13}C-DEPT	:	**δ_C** (ppm) *Positive Peaks:* 28.6(ma), 28.8(mi), 65.3(mi), 66.0(ma), 66.3(mi), 66.7(ma) *Negative Peaks:* 30.4(ma), 31.1(mi), 45.4(ma), 45.8 (mi)

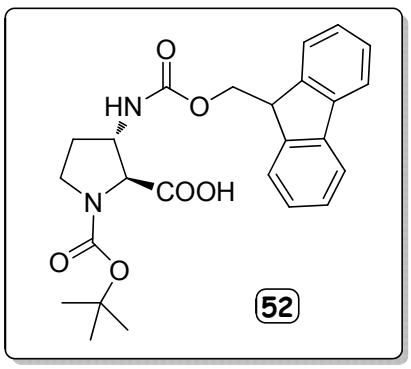
(2*S*,3*R*)N^1-(t-butyloxycabonyl)-3-aminoproline (51)



Compound **50** (1 g, 3.9 mmol) was taken in methanol (10 ml) to which was added 10% Pd-C (150 mg). The mixture was subjected to hydrogenation in a Parr-Hydrogenation apparatus at a pressure of 40 psi for 4 hrs. Pd-C suspension was removed by filtration through Celite and the filtrate was evaporated under reduced pressure to give white solid. The (2*S*,3*R*)N^1-(*t*-butoxycarbonyl)-3-aminoproline **51** (0.87 g, 97.5%) thus obtained was used without further purification.

Mol. Formula	:	$C_{10}H_{18}N_2O_4$
mp	:	Decomposes around 170 – 174^oC
Mol. Weight	:	230.26
ESI-MS *m/z*	:	231.32 $[M+1]^{+\bullet}$,253.37 $[M+Na]^{+\bullet}$, 269.28$[M+K]^{+\bullet}$
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3207, 3020, 2979, 1703, 1586, 1530, 1477, 1415, 1385, 1294, 1215, 1169, 1131, 932, 769, 669, 599, 551
^{1}H NMR (400 MHz, MeOH-d_4)	:	**δ_H** (ppm) 1.45 & 1.48 (2s, 9H), 1.97 (bs, 1H), 2.26-2.35 (m, 1H), 3.50-3.58 (m, 1H), 3.61-3.69 (m, 1H), 3.83-3.84 (d, 1H), 4.12-4.20 (d, 1H)

(2*S*,3*R*)-N^1-(t-butoxycarbonyl)-3-(fluorenylmethyloxycarbonylamino) proline (52)

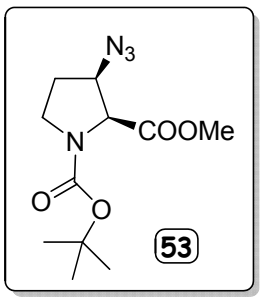


The aqueous solution (100 ml) of compound **51** (1 g, 4.3 mmol) was cooled to 0 ^oC and 10% aqueous Na_2CO_3 (20 ml) in dioxane (100 ml) was added. Fmoc-Cl (1.7 g, 6.5 mmol), dissolved in dioxane (10 ml) was added dropwise about 1 hr. The resulting mixture was stirred for 15 hrs at room temperature, maintaining the pH of solution around 8.0. The reaction mixture was concentrated on a rotary evaporator to remove dioxane and to the resulting slurry of water (50 ml) was added. The mixture was extracted with diethyl ether (5 x 50 ml) to remove unreacted Fmoc-Cl. The aqueous layer was then vigorously stirred by ethyl acetate under ice-cold condition and acidified with saturated $KHSO_4$ solution to pH 3. The ethyl acetate layer was separated and the aquous layer was further extracted into fresh ethyl acetate (3 x 50 ml), the organic layer were pooled and washed with water followed by brine, dried over anhydrous Na_2SO_4 and concentrated under vacuum. The crude

material was purified by silica gel chromatography (60% ethyl acetate/hexane elute) afford compound **52** as white solid. Yield 0.99 g (50.8%).

Mol. Formula	:	$C_{25}H_{28}N_2O_6$
mp	:	89.1 – 96.1 ^oC
Mol. Weight	:	452.50
ESI-MS *m/z*	:	453.40 $[M+1]^{+\bullet}$, 475.30 $[M+Na]^{+\bullet}$
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3437, 3019, 2400, 1720, 1700, 1512, 1407, 1216, 1163, 928, 770, 669, 491
^{1}H NMR (400 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.42 & 1.48 (2s, 9H), 1.80-1.84 (m, 1H), 1.88-1.92 (m, 1H), 3.51 (t, 1H), 3.57 (t, 1H), 4.19-4.20 (m, 1H), 4.28-4.36 (d, 1H), 4.42-4.44 (d, 2H), 5.29-5.36 (dd, 1H), 7.29-7.33 (t, 2H), 7.38-7.41 (t, 2H), 7.56-7.58 (d, 2H), 7.74-7.76 (d, 2H)
^{13}C NMR (100 MHz, $CDCl_3$)	:	**δ_C** (ppm) 28.1(ma), 28.3(mi), 29.6(ma), 29.9(mi), 44.2(mi), 44.6(ma),47.0, 54.2(mi), 55.5(ma), 64.5(mi), 65.2(ma), 66.7(mi), 66.9 (ma), 80.9(ma), 81.4(mi), 119.9, 124.9, 127.0, 127.7, 141.2, 143.6, 153.9, 155.9, 174.2
^{13}C-DEPT	:	**δ_C** (ppm) *Positive Peaks:* 28.1(ma), 28.3(mi), 47.0, 54.2(mi), 55.5(ma), 64.5(mi), 65.2(ma), 119.9, 124.9, 127.0, 127.7 *Negative Peaks:* 29.6(ma), 29.9(mi), 44.2(mi), 44.6(ma), 66.7(mi), 66.9 (ma)

(2*S*,3*R*)N^1-(t-butyloxycabonyl)-3-azidoproline methylester (53)[84]

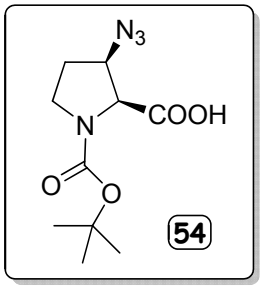


Compound **52** (1 g, 4.1 mmol) under similar reaction conditions (as mentioned for compound **54**) gave compound **53** as yellowish liquid. Yield 0.71 g (65%).

Mol. Formula	:	$C_{11}H_{18}N_4O_4$
Mol. Weight	:	270.29
ESI-MS *m/z*	:	270.98 $[M+1]^{+\bullet}$, 293.09 $[M+Na]^{+\bullet}$
$[\alpha]_D^{25}$	:	+26.86 (C=0.73, $CHCl_3$)
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3019, 2983, 2111, 1746, 1701, 1400, 1256, 1216, 1168, 1135, 1091, 1004, 757, 667
^{1}H NMR (200 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.39 & 1.44(2s, 9H),2.01-2.22(m, 2H),3.36-3.52 (m, 1H), 3.56-3.71(m, 1H), 3.77(s, 3H), 4.25-4.38 (m, 1H), 4.40-4.51(dd, 1H)
^{13}C NMR (50 MHz, $CDCl_3$)	:	**δ_C** (ppm) 27.8(ma), 28.0(mi), 29.0(mi), 29.7(ma), 43.8(ma), 44.2(mi), 51.8(ma), 51.9(mi), 60.7(ma), 61.3(mi), 61.5(ma), 61.7(mi), 80.2, 153.1(ma), 153.7(mi), 169.6(mi), 169.8(ma)

^{13}C-DEPT : **δ_C** (ppm) *Positive Peaks:* 27.8(ma), 28.0(mi), 51.8(ma), 51.9(mi), 60.7(ma), 61.3(mi), 61.5(ma), 61.7(mi) *Negative Peaks:* 29.0(mi), 29.7(ma), 43.8(ma), 44.2(mi)

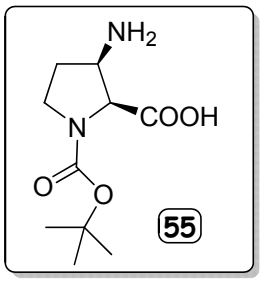
(2*S*,3*R*)N^1-(t-butyloxycabonyl)-3-azidoproline (54)



The compound **53** (1 g, 3.7 mmol) was hydrolysed using analogous conditions as stated for compound **50**. The free acid **54** was obtained as colourless liquid upon concentration. Yield 0.94 g (~99%)

Mol. Formula : $C_{10}H_{16}N_4O_4$
Mol. Weight : 256.26
ESI-MS *m/z* : 279.20 $[M+Na]^{+\bullet}$
^{1}H NMR (200 MHz, (MeOH-d_4 + D_2O)) : **δ_H** (ppm) 1.42 & 1.46 (2s, 9H), 1.91-2.27 (m, 3H), 3.43-3.79 (m, 2H), 4.20-4.56 (m, 1H)
^{13}C NMR (100 MHz, (MeOH-d_4 + D_2O)) : **δ_C** (ppm) 28.6(ma), 28.7(mi), 30.6(ma), 31.2(mi), 45.4(ma), 45.9(mi), 62.5(mi), 63.2 & 63.7(ma), 81.7(mi), 82.0(ma), 155.5(mi), 155.7(ma), 172.9(mi), 173.2(ma)
^{13}C-DEPT : **δ_C** (ppm) *Positive Peaks:* 28.6(ma), 28.7(mi), 62.5(mi), 63.2 & 63.7(ma) *Negative Peaks:* 30.6(ma), 31.2(mi), 45.4(ma), 45.9(mi)

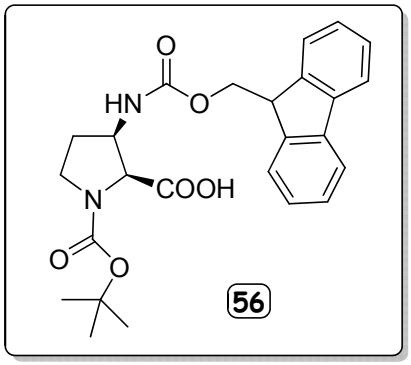
(2*S*,3*R*)N^1-(t-butyloxycabonyl)-3-aminoproline (55)



(2*S*,3*R*)-amino compound **55** (0.9 g, >99%) was synthesized from the corresponding azide **54** accordingly by reduction with Pd-C/H_2 as cited earlier for compound **51**.

Mol. Formula : $C_{10}H_{18}N_2O_4$
mp : Decomposes in the range 191 – 231 ^oC
Mol. Weight : 230.26
ESI-MS *m/z* : 231.29 $[M+1]^{+\bullet}$, 253.36 $[M+Na]^{+\bullet}$
IR (thin film on $CHCl_3$) : ν (cm^{-1}) 3019, 1696, 1625, 1406, 1215, 770, 669
^{1}H NMR (200 MHz, (MeOH-d_4 + D_2O)) : **δ_H** (ppm) 1.43 & 1.46 (2s, 9H), 1.96-2.05 (m, 1H), 2.29 (bs, 1H), 3.37-3.44 (m, 1H), 3.61-3.66 (m, 1H), 3.84-3.93 (m, 1H), 4.28-4.32 (m, 1H)

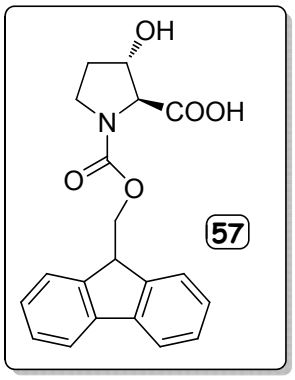
(2*S*,3*R*)-N¹-(t-butoxycarbonyl)-3-(9-fluorenylmethyloxycarbonylamino) proline (56)



The compound (2*S*,3*R*)-**56** was prepared from the free amine **55** using the similar procedure as the one described for the compound (2*S*,3*S*)-**52.** The compound **56** was isolated as a foamy white solid. Yield 0.79 g (40.3%)

Mol. Formula	:	$C_{25}H_{28}N_2O_6$
mp	:	53.3 – 57.4°C
Mol. Weight	:	452.50
ESI-MS *m/z*	:	453.33 $[M+1]^{+\bullet}$, 475.30 $[M+Na]^{+\bullet}$
$[\alpha]_D^{25}$	:	+42.77 (C= 0.29, EtOH)
IR (thin film on $CHCl_3$)	:	**ν** (cm^{-1}) 3431, 3019, 2980, 2930, 1725, 1691, 1407, 1368, 1216, 1168, 1096, 1039, 929, 856, 758, 668
^{1}H NMR (400 MHz, $CDCl_3$)	:	**δ_H** (ppm) 1.43 & 1.49 (2s, 9H), 1.98-2.07 (m, 2H), 3.47-3.57 (m, 1H), 3.63-3.75 (m, 1H), 4.07-4.23 (m, 2H), 4.30-4.41 (m, 2H), 4.49-4.62 (m, 2H), 7.32-7.35 (m, 2H), 7.40-7.44 (m, 2H), 7.57-7.63 (m, 2H), 7.75-7.80 (m, 2H)
^{13}C NMR (100 MHz, $CDCl_3$)	:	**δ_C** (ppm) 28.2(ma), 28.3(mi), 29.6(ma), 30.0(mi), 44.4(mi), 43.6 & 45.1(ma), 46.9, 56.8, 60.7(mi), 60.9(ma), 68.0(ma), 68.2(mi), 80.4(mi), 80.6(ma), 119.9(ma), 120.0(mi), 124.7(mi), 124.8(ma), 127.1 (ma), 127.2 (mi), 127.7 (mi), 127.8 (ma), 141.2 (ma), 141.3 (mi), 143.5, 153.6, 158.3, 174.9
^{13}C-DEPT	:	**δ_C** (ppm) *Positive Peaks:* 28.2(ma), 28.3(mi), 46.9, 56.8, 60.7(mi), 60.9(ma), 119.9(ma), 120.0(mi), 124.7(mi), 124.8(ma), 127.1(ma), 127.2(mi), 127.7(mi), 127.8(ma) *Negative Peaks:* 29.6(ma), 30.0(mi), 44.4(mi), 43.6 & 45.1(ma), 68.0(ma), 68.2(mi)

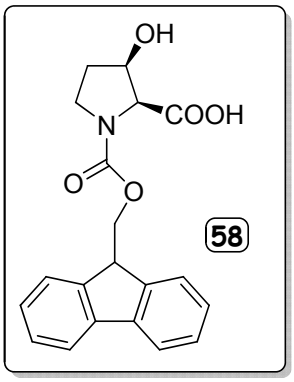
(2*S*,3*S*)-N^1(9-fluorenylmethyloxycarbonylamino)-3-hydroxy proline (57)



Commercial grade *trans*-3-Hydroxy-L-Proline **45** (1 g, 7.6 mmol) was dissolved in 100 ml of water-dioxane (1:1) was added with 10% Na_2CO_3 (10 ml) and cooled to 0 ^oC. pH was maintained around 10.0 for few minutes then (2.95 g, 11.4 mmol) of Fmoc-Cl, dissolved in dioxane (10 ml) was added dropwise for about 1 hr. The resulting mixture was stirred for 15 hrs at room temperature, maintaining the pH of solution at 8.0. The reaction mixture was concentrated on a rotary evaporator to remove dioxane and the resulting slurry was treated with 50 ml of water. The mixture was extracted with diethyl ether (5 x 50 ml) to remove unreacted Fmoc-Cl. The aqueous layer was then vigorously stirred along with ethyl acetate under ice-cold condition, and acidified with saturated $KHSO_4$ solution to pH 3. The ethyl acetate layer was separated and the aqueous layer was further extracted into fresh ethyl acetate (3 x 50 ml), the organic layers were pooled and washed with water followed by brine, dried over anhydrous Na_2SO_4 and concentrated under vacuum. The crude material was recrystallised from ethyl acetate to get compound **57** as crystalline white solid (2.35 g, 87.3% yield).

Mol. Formula	:	$C_{21}H_{21}NO_5$
mp	:	Decomposition with sublimation around 204.5 – 207.8 ^oC
Mol. Weight	:	353.37
ESI-MS *m/z*	:	354.24 $[M+1]^{+\bullet}$, 376.17 $[M+Na]^{+\bullet}$
$[\alpha]_D^{25}$	:	-16.7 (C=0.61, EtOH)
^{1}H NMR 500 MHz ($MeOH\text{-}d_4$ + $CDCl_3$)	:	**δ_H** (ppm) 1.90-2.0 (m, 1H), 2.05-2.14 (m, 1H), 3.65-3.72 (m, 2H), 4.15-4.25 (m, 2H), 4.30-4.35 (m, 2H), 4.47 (bs, 1H), 7.26-7.29 (m, 2H), 7.33-7.38 (m, 2H), 7.54-7.6 (m, 2H), 7.70-7.74 (m, 2H)
^{13}C NMR 100 MHz ($MeOH\text{-}d_4$ + $CDCl_3$)	:	**δ_C** (ppm) 30.8(mi), 31.7(ma), 44.0(mi), 44.3(ma), 46.4, 67.4(ma), 67.1(mi), 67.6, 73.0(ma), 74.0(mi), 119.2, 124.4, 126.4, 127.0, 140.6, 143.1(mi), 143.2(ma), 154.8(ma), 155.0(mi), 171.9(ma), 172.1(mi)
^{13}C-DEPT	:	**δ_C** (ppm) *Positive Peaks:* 46.4, 67.6, 73.0(ma), 74.0(mi), 119.2, 124.4, 126.4, 127.0 *Negative Peaks:* 30.8(mi), 31.7(ma), 44.0(mi), 44.3(ma), 67.4(ma), 67.1(mi)

(2*S*,3*R*)-N^1(9-fluorenylmethyloxycarbonylamino)-3-hydroxy proline (58)



The hydrolysis of methylester of **48** (1 g, 4.1 mmol) with aqueous NaOH (2N) was followed by removal of N^1-Boc group of compound **48** by addition of 50% TFA in DCM (3 ml) in ice-cold conditions. The mixture was stirred for at least 2 hrs, when TLC indicated complete deprotection. The reaction solvent was removed under reduced pressure with KOH trap followed by high vacuum. This TFA-salt of compound **48** was protected with Fmoc using the same reaction conditions as stated for compound **57**. The purification of the crude product afforded compound **58** as a white solid. Yield 1.1 g, 79%.

Mol. Formula	:	$C_{21}H_{21}NO_5$
mp	:	Decomposition around 63.8 – 81.3 ^oC
Mol. Weight	:	353.37
ESI-MS *m/z*	:	354.17 $[M+1]^{+\bullet}$, 376.24 $[M+Na]^{+\bullet}$
^{1}H NMR 500 MHz ($MeOH\text{-}d_4$ + $CDCl_3$)	:	δ_H (ppm) 1.97-2.04 (m, 1H), 2.05-2.12 (m, 1H), 3.46-3.55 (m, 1H), 3.62-3.68 (m, 1H), 4.23-4.27 (m, 1H), 4.33-4.38 (m, 2H), 4.39-4.44 (m, 1H), 4.56-4.64 (m, 1H), 7.28-7.31 (t, 2H), 7.36-7.37 (t, 2H), 7.60-7.64 (t, 2H), 7.76-7.79 (t, 2H)
^{13}C NMR 500 MHz ($MeOH\text{-}d_4$ + $CDCl_3$)	:	δ_C (ppm) 32.6(ma), 33.5(mi), 45.4(mi), 45.6(ma), 48.2, 64.8(ma), 65.1(mi), 68.6(mi), 69.1(ma), 71.8(mi), 72.6(ma), 120.8, 126.0(ma), 126.1(mi), 128.1, 128.7, 142.3(mi), 142.4(ma), 144.8(mi), 145.1(ma), 156.5(mi), 172.8(mi), 173.2(ma)
^{13}C-DEPT	:	δ_C (ppm) *Positive Peaks:* 48.2, 64.8(ma), 65.1(mi), 120.8, 126.0(ma), 126.1(mi), 128.1, 128.7 *Negative Peaks:* 32.6(ma), 33.5(mi), 45.4(mi), 45.6(ma), 68.6(mi), 69.1(ma)

3.12.2 Purification & characterization

3.12.2a High performance liquid chromatography (HPLC)

The peptides (**P11** - **P14**) were purified by reverse phase on Waters 600 HPLC instrument equipped with 2998-Photodiode array detector (PDA). Semi-preparative RP-C18 columns (250 x 10 mm, 10 µm) of Allteck-Alltima make were used for peptides. A gradient of 0-100% B at a flow rate of 3 ml/min was used to elute the peptide and the eluant was monitored at 220 nm. The peaks corresponding to the peptide were collected and the fractions were lyophilized. The purity of the final peptides were further analyzed on Merck LiChrospher 100 RP-18 (250 x 4 mm, 5

μm) column by using a gradient flow of 0 to 100% B in 20 min at a flow rate of 1.0 ml/min. The spectral acquisition, analysis and processing was done on Waters Empower-2154 software. The absorbance of the eluant was monitored at its corresponding wavelength and the purity was obtained from the integrator output. The solvent system comprised of MeCN:Water (5:95) with 0.1% TFA for solution A and for solution B MeCN:Water (50:50), 0.1% TFA. The purities of the hence purified peptides were found to be more than 95%.

3.12.2b MALDI-TOF Characterization

MALDI-TOF mass spectra of peptides purified by HPLC were obtained on Voyager-De-STR (Applied Biosystems) instrument. α-Cyano-4-hydroxycinnamic acid (CHCA) was used as matrix in a dilution solvent (50:50:0.1 Water:MeCN:TFA).

3.12.3 Circular dichroism spectroscopy & thermal denaturation studies

CD spectra were recorded on JASCO J-715 spectropolarimeter using cylindrical, jacketed quartz cell (1 mm path length) with an attached Julabo-UC-25 water circulator. And spectra were recorded with a spectral resolution of 0.05 nm, band width of 1 nm at a scan speed of 200 nm/min and a response time of 1 sec. All the spectra were corrected for respective buffer condition and are typically averaged over 6-12 scans.

For **CD thermal denaturation studies**, the samples were annealed in water bath at 90 °C and slowly cooled to room temperature over a period of 6 hrs prior to spectroscopic analysis. These samples were then incubated at 4 °C for 12 hr followed by additional half an hour incubation in the instrument at the initial measurement temperature. The temperature was varied in steps of 5 °C and the spectra were recorded at each step. An equilibration period of 5 min was allowed at each temperature. Data processing and curve fitting was performed using MicroCal Origin 8.0 software. Ellipticity at specified wavelength for each temperature was plotted, normalized data was fitted to a sigmoid curve and the T_m (melting temperature) values are derived from the first derivative curve of the fit.

3.13 References

1) Engel, J. *Science* **1997**, *277*, 1785-1786.

2) Engel, J.; Bachinger, H. P. in *Trends in Collagen*; eds. Chandrasekaran, G.; Yathindra, N.; Ramasami, T. *Proceed. Indian Acad. Sciences* **1999**, *111*, 81-86.

3) (a) Linsenmayar, T. F. in *Cell Biology of Extracellular Matrix* 2nd ed. Hay, E. D., Ed., Plenum Press, New York, **1991**,7-44. (b) Prockop, D. J.; Kivirikko, K. I. *Ann. Rev. Biochem.* **1995**, *64*, 403-434.

4) Woodhead-Galloway, J. in *Collagen: The Anatomy of a Protein* Edward Arnold (Publishers) Ltd. London, 10-19.

5) (a) Creighton, T. E. *Proteins*, W. H. Feeman & Co., New York, **1993,** 45-60. (b) Ramachandran, G. N.; Ramakrishnan, C. in *Biochemistry of Collagen* Ramachandran, G. N.; Reddi, A. H., Ed. Academic Press, London, **1976**, 45-84. (c) Ramachandran, G. N.; Ramakrishnan, C.; Venkatachalam, C. M. in *Conformations of Collagen* Ramachandran, G., Ed., Academic Press, London, **1967**, 429-438.

6) Ramachandran, G. N., in *Aspects of Protein Structure* Ramachandran, G. N., Ed., Academic Press, New York, **1963**, 39-55.

7) (a) Prockop, D. J.; Kivrikko, K. I.; Tuderman, L.; Guzman, N. A. *New Eng. J. Med.* **1979,** *301***,** 13-23. (b) Prockop, D. J. *New Eng. J. Med.* **1992**, *326*, 540-546. (c) Brodsky, B.; Shah, N. K. *FASEB J.* **1995**, *9*, 153-1546.

8) Koide T.; Nagata, K. *Top. Curr. Chem.* **2005**, *247*, 85-114.

9) Dolz, R.; Heidemann, E. *Bioplymers* **1986**, *25*, 1069-1080.

10) Heidemann, E.; Roth, W. *Adv. Ploym. Sci.* **1982**, *43*, 143-203.

11) Persikov, A. V.; Ramshaw, J. A.; Brodsky, B. *J. Biol. Chem.* **2005**, *280*, 19343-19349.

12) Ramachandran, G. N.; Kartha, G. *Nature* **1954**, *174*, 269-270.

13) (a) Ramachandran, G. N.; Kartha, G. *Nature* **1955**, *176*, 593-595. (b) Ramachandran, G. N.; Kartha, G. *Proc. Indian Acad. Sci*. **1955**, *XLII*, 215-234.

14) (a) Rich, A.; Crick, F. H. C. *Nature* **1955**, *176*, 915-916. (b) Rich, A.; Crick, F. H. C. *J. Mol. Biol.* **1961**, *3*, 483-506.

15) Okuyama, K.; Arnott, S.; Takayanagi, M.; Kakudo, M. *J. Mo.l Biol.* **1981**, *152*, 427–443.

16) (a) Brodsky, B.; Persikov, A. V. *Adv Protein Chem* **2005**, *70*, 301-339. (b) Okuyama, K.; Wu, G.; Jiravanichanun, N.; Hongo, C.; Noguchi, K. *Biopolymers* **2006**, *84*, 421-432. (c) Okuyama, K.; Narita, H.; Kawaguchi, T.; Noguchi, K.; Tanaka, Y.; Nishino, N. *Biopolymers* **2007**, *86*, 212–221.

17) Brg, R. A.; Prockop, D. J. *Biopchem. Biophys. Res. Commun*. **1973**, *52*, 115-120.

18) Sakakibara, S.; Inouye, K.; Shudo, K. S.; Kishida, Y.; Kobayashi, Y.; Prockop, D. J. *Biophys. Biochem. Acta* **1973**, *303*, 198-202

19) Inouye, K.; Kobayashi, Y.; Kyogoku, Y.; Kishida, Y.; Sakakibara, S.; Prockop, D. J. *Arch. Biochem. Biophys*, **1982**, *219*, 198-203.

20) Inouye, K.; Sakakibara, K.; Prockop, D. J. *Biochim. Biophys. Acta*, **1976**, *420*, 133-141.

21) Ramachandran, G. N.; Bansal, M.; Batnagar, R. S. *Biochim. Biophys.Acta* **1973**, *322*, 166-171.

22) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. *Nature* **1998**, *392*, 666-667.

23) Bann, J. G.; Bachinger, H. P. *J. Biol. Chem*. **2000**, *275*, 24466-24469.

24) Bos, K. J.; Rucklidge, G. J.; Dunbar, B.; Robins, S. P. *Matrix Biol.* **1999**, *18*, 149-153.

25) (a) Wolff III, J. S.; Logan, M. A.; Ogle, J. D. *Fed Proc.* **1966**, *25*, 862–863. (b) Wolff III, J. S.; Ogle, J. D.; Logan, M. A. J. Biol. Chem. **1966**, *241*, 1300-1307

26) (a) Click, E. M.; Bornstein, P. *Biochemistry* **1970**, *9*, 4699-4706. (b) Lane, J. M.; Miller, E. J. *Biochemistry,* **1969**, *8*, 2134-2139. (c) Miller, E. J.; Martin, G. R.; Piez, K. A. Powers, M. J. *J. Biol. Chem*. **1967**, *242*, 5481-5489.

27) Fietzek, P. P.; Rexrodt, F. W.; Wendt, P.; Stark, M.; Kuhn, K. *Eur. J. Biochem*. **1972**, *30*, 163–168.

28) Vranka, J. A.; Sakai, L.Y.; Bachinger, H. P. *J. Biol. Chem.* **2004**, *279*, 23615–23621.

29) Gaill, F.; Mann, K.; Wiedemann, H.; Engel, J. Timpl, R. *J. Mol. Biol.***1995**, *246*, 284-294.

30) Kimura, S.; Matsuura, F. *J. Biochem*. **1974**, *75*, 1231-1240.

31) Kimura S.; Kamimura, T; Takema, Y.; Kubota, M., *Biochim. Biophys. Acta* **1982**, *669,* 251-257.

32) Takema, Y. Kimura, S. *Biochim Biophys Acta* **1982**, *706*, 123-128.

33) Miura, S.; Kimura, S. *J. Biol. Chem*.**1985**, *260*, 15352-15356.

34) Gryder, R. M.; Lamon, M.; Adams, E. *J. Biol. Chem.* **1975**, *250*, 2470- 2474.

35) Schumacher, M. A.; Mizuno, K. Bachinger, H. P. *J. Biol. Chem.* **2006**, *281*, 27566-27574. (b) Hodges, J. A. Raines, R. T. *J. Am. Chem. Soc*. **2005**, *127*, 15923-15932. (c) Jenkins, C. L.; Bretscher, L. E.; Guzei. I. A.; Raines, R.T. *J. Am. Chem. Soc.* **2003**,*125*, 6422-6427.

36) Li, M.; Fan, P.; Brodsky, B.; Baum, J. *Biochemistry* **1993**, *32*, 7377-7387.

37) Venugopal, M. G.; Ramshaw, J. A. M.; Braswell, E.; Zhu, D.; Brodsky, B. *Biochemistry* **1994**, *33*, 7948-7956.

38) (a) Gratzer, W. B.; Rhodes, W.; Fasman, G. D. *Biopolymers* **1963**, *1*, 319-330. (b) Wood, G. C. *Biochem. Biophys. Res. Commun.* **1963**, *13*, 95-98.

39) (a) Lazarev, Yu. A.; Lazareva, A. V.; Shibnev, V. A.; Esipova, N. G. *Biopolymers* **1978**, *17*, 1197-1214. (b) Khromova, T. B.; Lazarev, Y. *A. Biophysics* **1986**, *31*, 227-232.

40) (a) Saito, H.; Tabeta, R.; Shoji, A.; Ozaki, T.; Ando, I.; Miyata, T. *Biopolymers* **1984**, *23*, 2279-2297. (b) Renugopalakrishnan, V.; Colette, T. W.; Carreira, L. A.; Bhatnagar, R. S. *Macromolecules* **1985**, *18*, 1786-1788.

41) Brodsky, B.; Li, M. H.; Long, C. G.; Apigo, J.; Baum, J. *Biopolymers* **1992**, *32*, 447-451.

42) Bruns, R. R.; Gross, J. *Biochemistry* **1973**, *12*, 808-815.

43) (a) Olsen, B. R.; Berg, R. A.; Sakakibara, S.; Kishida, Y.; Prockop, J. D. *J. Mol. Biol.* **1971**, *57*, 589-595. (b) Miller, A.; Wray, J. S. *Nature* **1971**, *230*, 437-439.

44) (a) Shaw, B. R.; Schurr, J. M. *Biopolymers* **1975**, *14*, 1951-1985. (b) Weber, R. W. Nitschmann, H. *Helv. Chem. Acta.* **1973**, *61*, 701-708.

45) (a) Fields, C. G.; Lovdhal, C. M.; Miles, A. J.; Matthias Hagen, V. L.; Fields, G. B. *Biopolymers* **1993**, *33*, 1695-1707. (b) Fields, C. G.; Grab, B.; Lauer, J. L.; Fields, G. B. *Anal. Biochem.* **1995**, *231*, 57-64.

46) Fairwather, R.; Jones, J. H. *J. Chem. Soc. Perkin* I **1972**, 2475-2481.

47) Hutton, J. J.; Marglin, A.; Witkop, B.; Kurtz, J.; Berger, A.; Undenfreind, S. *Arch. Biochim. Biophys.* **1968**, *125*, 779-785.

48) Goodman, M., and Kwak, J. *Proc. Indian Acad. Sci. (Chem. Sci.)* **1999***, 111*, 35-49.

49) (a) Li, S. T.; Golub, E.; Katz, E. P. *J. Mol. Biol.* **1975**, *98*, 791. (b) Katz, E. P.; David, C. W. *Biopolymers* **1990**, *29*, 791. (c) Zagari, A.; Nemethy, G.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 967.

50) Venugopal, M. G.; Ramshaw, J. W.; Brestscher, L. E.; Raines, R. T. *Nature (London)* **1998**, *392*, 666.

51) (a) Tuchscherer, G. Dorner, B.; Sila, U; Kambar, B.; Mutter, M. *Tetrahedron* **1993**, *49*, 3559-3574. (b) Tuchscherer, G.; Mutter, M. *Chem. Ind.* **1997**, 597-601.

52) Goodman, M.; Bhumralkar, M.; Jefferson, E. A.; Kwak, J.; Locardi, E. *Biopolymers* **1998**, *47*, 127.

53) (a) Roth, W.; Heidemann, E. *Biopolymers* **1980**, *19*, 1909-1917. (b) Greiche, Y.; Heidemann, E. *Biopolymers*, **1979**, *18*, 2359-2361. (c) Thakur, S.; Vadolas, D.; Germann, H. P.; Heidemann, E. *Biopolymers* **1986**, *25*, 1081-1086.

54) Goodman, M.; Feng, Y.; Melacini, G.; Taulance, J. P. *J. Am. Chem. Soc.* **1996**, *118*, 5156-5157.

55) (a) Torp, S.; Arridge, R. S. C.; Armeniades, C. D.; Baer, E. in *Structure of Fibrous Biopolymers* Atkins, E. D. T.; Keller, A. eds., Butterworth, London, **1975**, *26*,197-221. (b) Uitto, J. *Dermatol. Clin.* **1986**, *4*, 433-436.

56) Byers, P. H.; Cole, W. G. In Connective Tissue and Its Heritable Disorders; Royce, P. M.; Steinmann, B., Eds.; Wiley-Liss: New York, **2002**; 385-430.

57) Hudson, B. G.; Tryggvason, K.; Sundaramoorthy, M.; Neilson, E. G. *N. Engl. J. Med.* **2003**, *348*, 2543–2556.

58) (a) Ramshaw, J. A.; Werkmeister, J. A.; Glattauer, V. *Biotechnol. Genet. Eng. Rev.* **1996**, *13*, 335-382. (b) Tabata, Y. *Drug Discovery Today* **2001**, *6*, 484-487.

59) Ramshaw, J. A.; Peters, D. E.; Werkmeister, J. A.; Ketharanathan, V. *J. Biomed. Mater. Res.* **1989**, *23*, 649-660.

60) (a) Thompson, K. P.; Hanna, K.; Waring III G. O.; Gipson, I.; Liu, Y.; Gailitis, R. P, Johnson-Wint, B.; Green, K. *Refract. Corneal. Surg.* **1991**, *7*, 240-248. (b) Kirkham, S. M.; Dangel, M. E. *Ophthalmic Sur.* **1991**, *22*, 455-461.

61) Cavallaro, J. F.; Kemp, P. D.; Kraus, K. H. *Biotechnol. Bioeng.* **1994**, *43*, 781-791.

62) Dunn, M. G.; Avasarala, P. N.; Zawadsky, J. P. *J. Biomed. Mater. Res.* **1993**, *27*, 1545-1552.

63) Gelman, M. A.; Gellman, S. H. Using Constrained β-Amino Acid Residues to Control β-Peptide Shape and Function. In *Enantioselective Synthesis of β-Amino Acids*, John Wiley & Sons, **2005**, 527-591.

64) Vasudev, P. G., Chatterjee, S., Shamala, N., and Balaram, P. *Chem. Rev.* **2011**, *111*, 657-87.

65) (a) Babu, I. R.; Ganesh, K. N. *J. Am. Chem. Soc.*, **2001**, *123*, 2079-2080, (b) Umashankara, M.; Babu, I. R.; Ganesh, K. N. *Chem. Commun*, **2003**, 2606-2607,

66) Shah, N. K.; Ramshaw, J. A. M.; Kirkpatrick, A.; Shah, C.; Brodsky, B. *Biochemistry* **1996**, *35*, 10262-10268.

67) (a) Martin, S. F.; Dodge, J. A. *Tetrahedron Lett*. **1991**, *32*, 3017-3020. (b) Dodge, J. A.; Trujillo, J. I.; Presnell, M. *J. Org. Chem*. **1994**, *59*, 234-236.

68) Gómez-Vidal, J. A.; Forrester, M. T.; Silverman, R. B. *Org. Lett.* **2001**, *3*, 2477.

69) Kaiser, E.; Colescott, R. L.; Bossinger, C.D.; Cook, P. I. *Anal. Biochem.* **1970**, *34*, 595-598.

70) Christensen, T. *Acta. Chem. Scand*. **1979**, *34*, 594-598.

71) Madder, A.; Farcy, N.; Hosten, N. G. C.; De Muynck, H.; De Clereq, P.; Barry, J.; Davis, A.P: *Eur. J. Org. Chem*. **1999**, *11*, 2787-2791.

72) (a) Kobayashi, Y.; Isemura, T. in *Progress in Polymer Science*, Okamura, S.; Takayanagi, M., Eds., Kodansha Limited, Tokyo & John Wiley & Sons, New York, **1972**, *3*, 315-380. (b) Inouye, K.; Sakakibara, S.; Prockop, D. J. *Biochim. Biophys. Acta* **1976**, *420*, 133-141. (c) Weber, R. W.; Nitschmann, H. *Helv. Chim. Acta* **1978**, *61*, 701-708. (d) Engel, J.; Chen, H. T.; Prockop, D. J.; Klump, H. *Biopolymers* **1977**, *16*, 601-622.

73) Feng, Y.; Melacini, G.; Taulane, J. P.; Goodman, M. *J. Am. Chem. Soc.* **1996**, *118*, 10351-10358.

74) (a) Gerlsma, S. Y.; Stuur, E. R. *Int. J. Pept. Protein Res.* **1972**, *4*, 377-383. (b) Back, J. F.; Oakenfull, D.; Smith, M. B. *Biochemistry*, **1979**, *18*, 5191-5196. (c) Arakawa, T.; Timasheff, S. N. *Biochemistry*, **1982**, 6536-6544.

75) (a) Harrap, B. S. *Int. J. Pept. Protein Res.* **1969**, *1*, 245-2532. (b) Gekko, K.; Koga, S. *J. Biochem.* **1983**, *94*, 199-205.

76) (a) Brown III, F. R.; Carver, J. P.; Blout, E. R. *J. Mol. Biol.* **1969**, *39*, 307-313. (b) Brown III, F. R.,; Di Corato, A.; Lorenzi, G. P.; Blout, E. R. *J. Mol. Biol.* **1972**, *63*, 85-99. (c) Vuilleumier, S.; Mutter, M. *Biopolymers* **1993**, *33*, 389-400. (d) Choma, C. T.; Lear, J. D.; Nelson, M. J.; Dutton, P. L.; Roberson, D. E.; DeGrado, W. F. *J. Am. Chem. Soc.* **1994**, *116*, 856-865.

77) (a) Mizuno, K.; Hayashi, T.; Peyton, D. H.; Bachinger, H. P. *J. Biol. Chem.* **2004,** *279*, 282-287. (b) Mizuno, K.; Peyton, D. H.; Hayashi, T.; Engel, J.; Bachinger, H. P. *The FEBS journal* **2008,** 275, 5830-5840.

78) Woody, R. W. *Adv. Biophys. Chem.* **1992**, *2*, 37–79.

79) (a) Shi, Z., Olson, C. A., Rose, G. D., Baldwin, R. L., and Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190-9195, (b) Rucker, A. L., and Creamer, T. P. *Protein Sci.* **2002**, *11*, 980-985. (c) Chellgren, B. W., and Creamer, T. P. *Biochemistry* **2004**, *43*, 5864-5869.

80) Fernandez, F.; Scherega, H. A. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 113

81) Werkmeister, J. A.; Ramshaw, J. A. M. (Ed.) *Collagen Biomaterials*, Elsevier Science, Barking, Essex, **1992**.

82) (a) Martin, S. F.; Dodge, J. A. *Tetrahedron Lett*. **1991**, *32*, 3017-3020. (b) Dodge, J. A.; Trujillo, J. I.; Presnell, M. *J. Org. Chem*. **1994**, *59*,234-236.

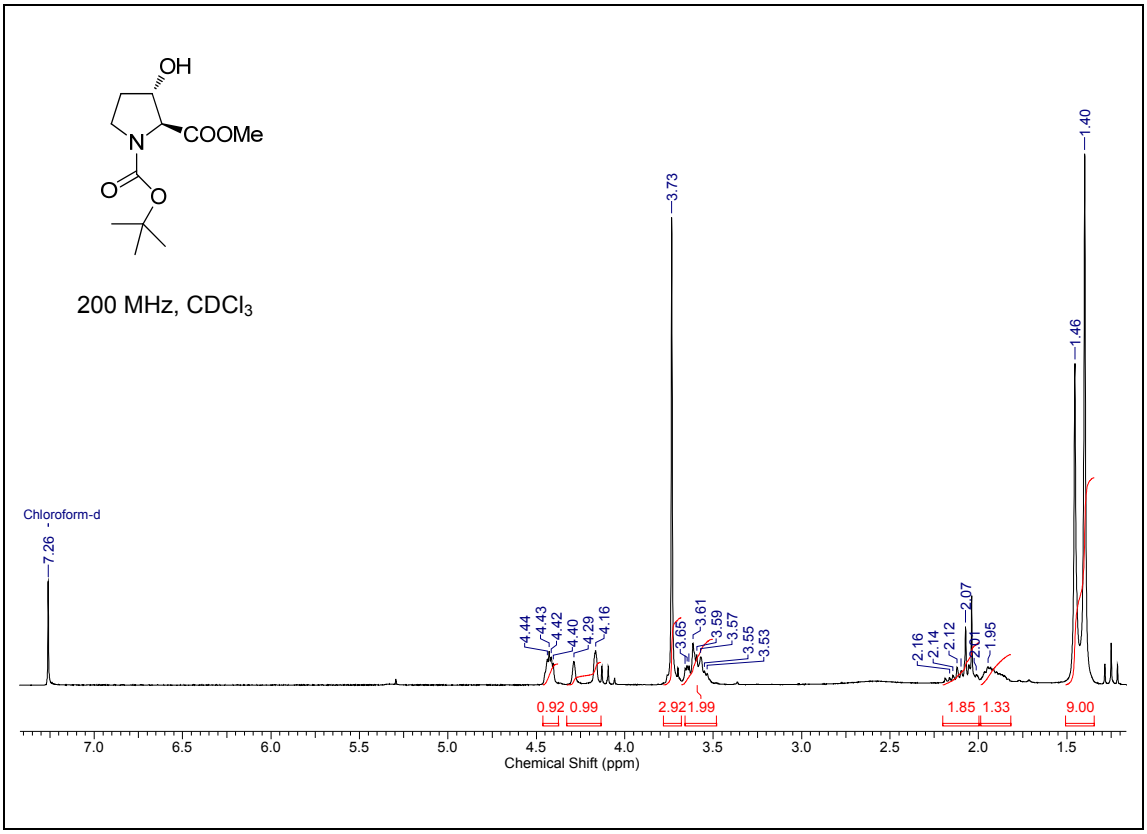
83) Gómez-Vidal, J. A.; Forrester, M. T.; Silverman, R. B. *Org. Lett.* **2001**, *3*, 2477

84) Gómez-Vidal, J. A.; Silverman, R. B. *Organic letters* **2001**, *3*, 2481-2484.

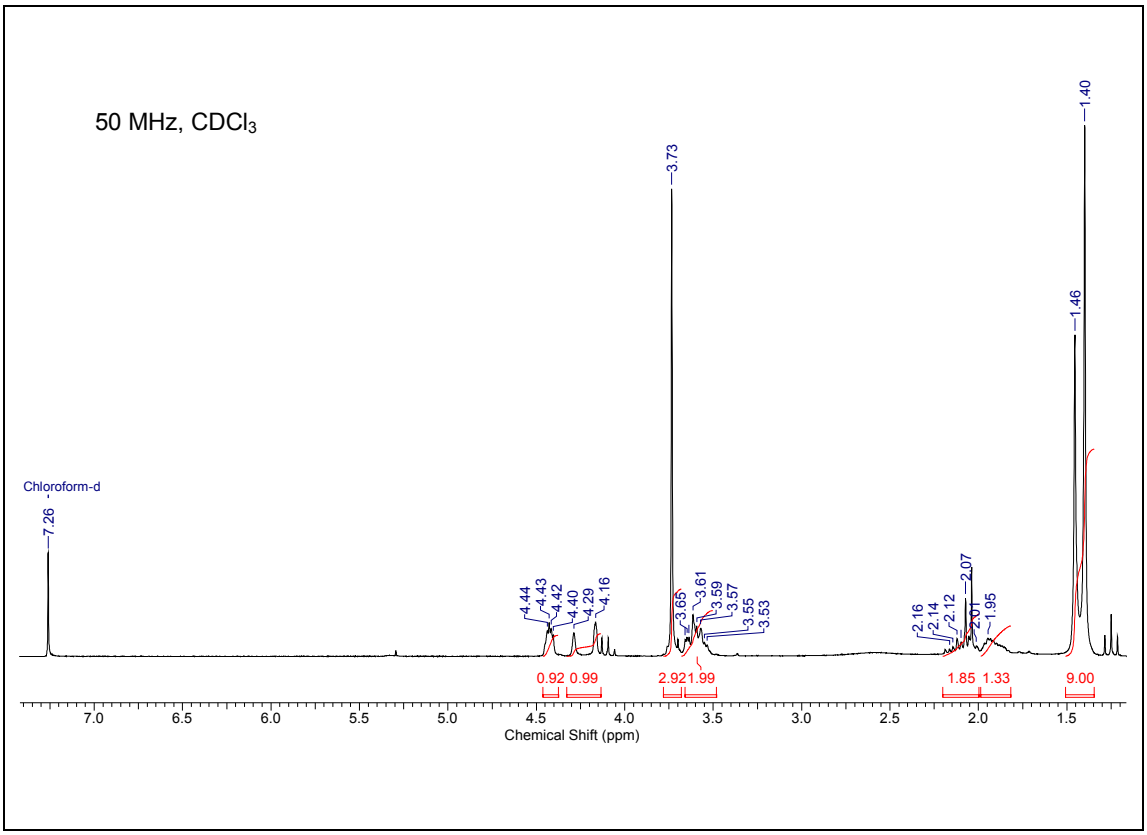
3.14 Appendix C: Characterization data of synthesized compounds

Designation	Description	Page
Compound **46**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	348-349
Compound **47**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	350-351
Compound **48**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	352-353
Compound **49**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	354-355
Compound **50**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	356-357
Compound **51**	^{1}H NMR, LC-MS, FT-IR	358
Compound **52**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	359-360
Compound **53**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	361-362
Compound **54**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	363-364
Compound **55**	^{1}H NMR, LC-MS, FT-IR	365
Compound **56**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS, FT-IR	366-367
Compound **57**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	368-369
Compound **58**	^{1}H NMR, ^{13}C NMR, DEPT-^{13}C NMR, LC-MS	370-371

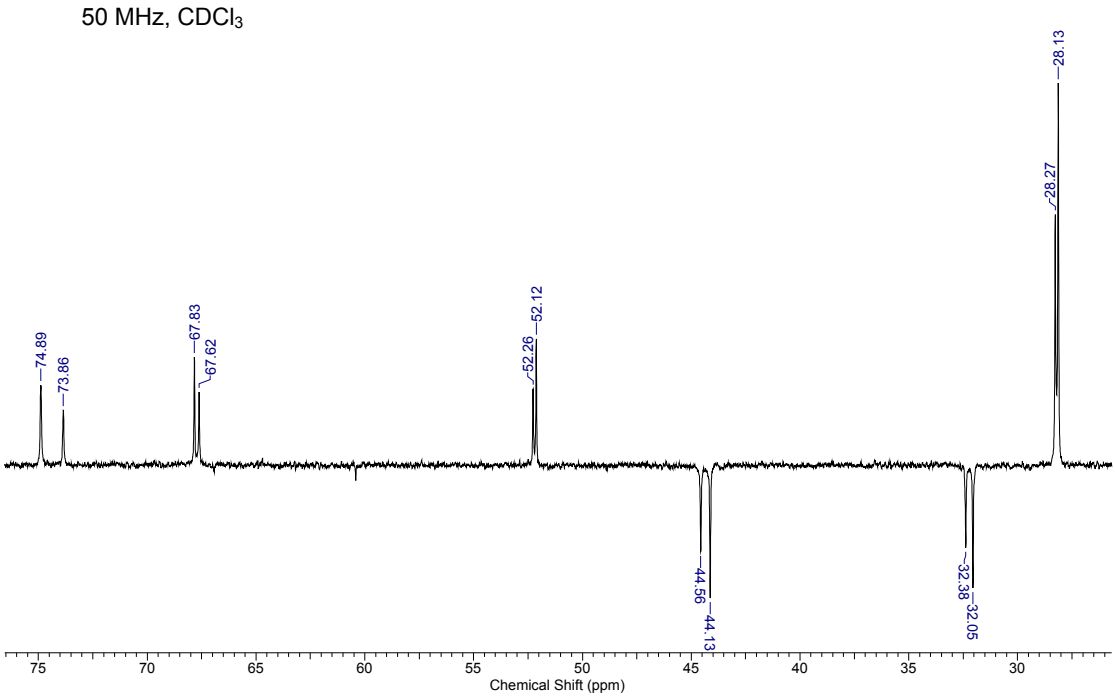
^{1}H NMR of compound 46



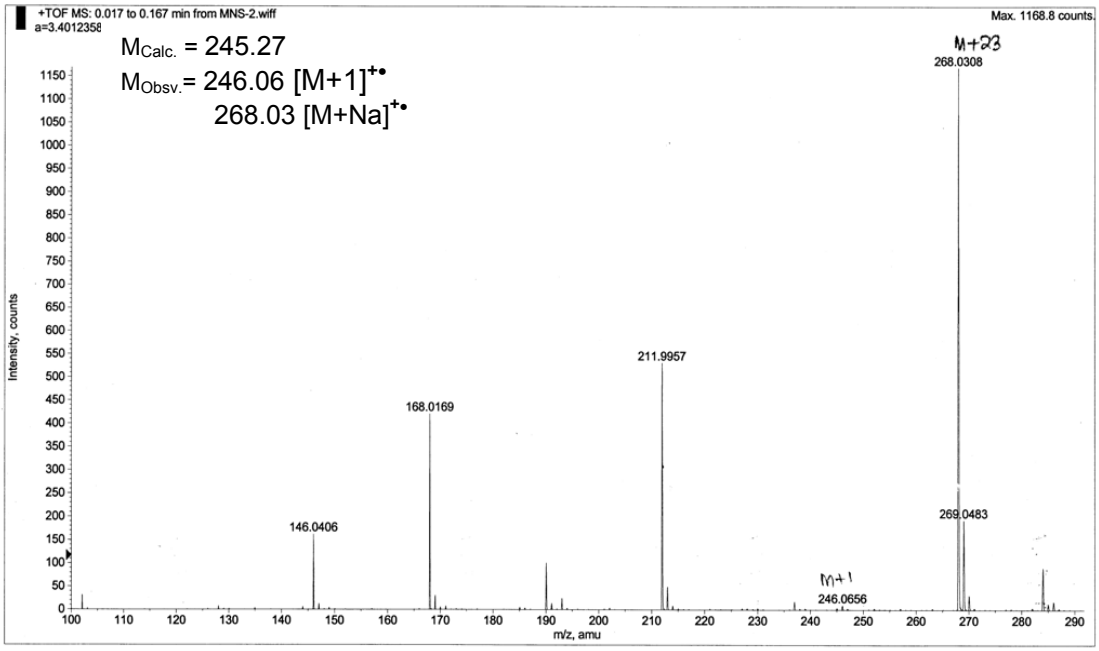
^{13}C NMR of compound 46



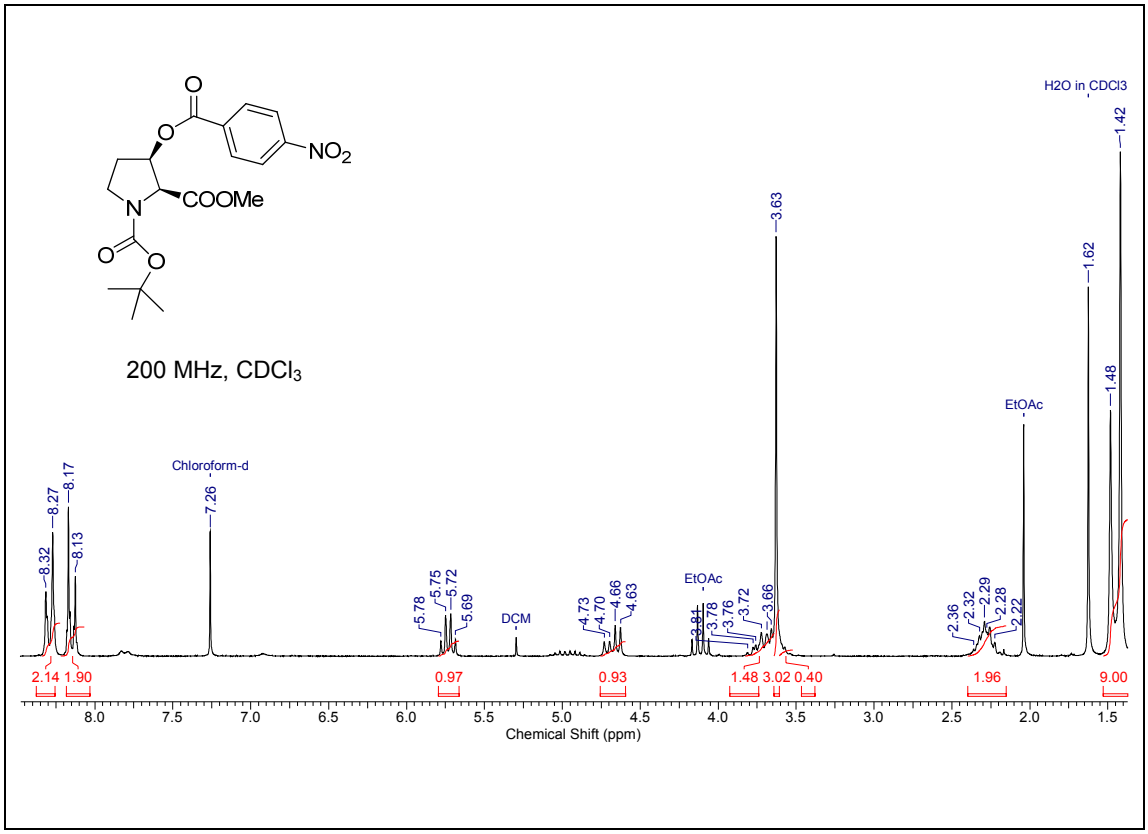
DEPT-^{13}C NMR of compound 46



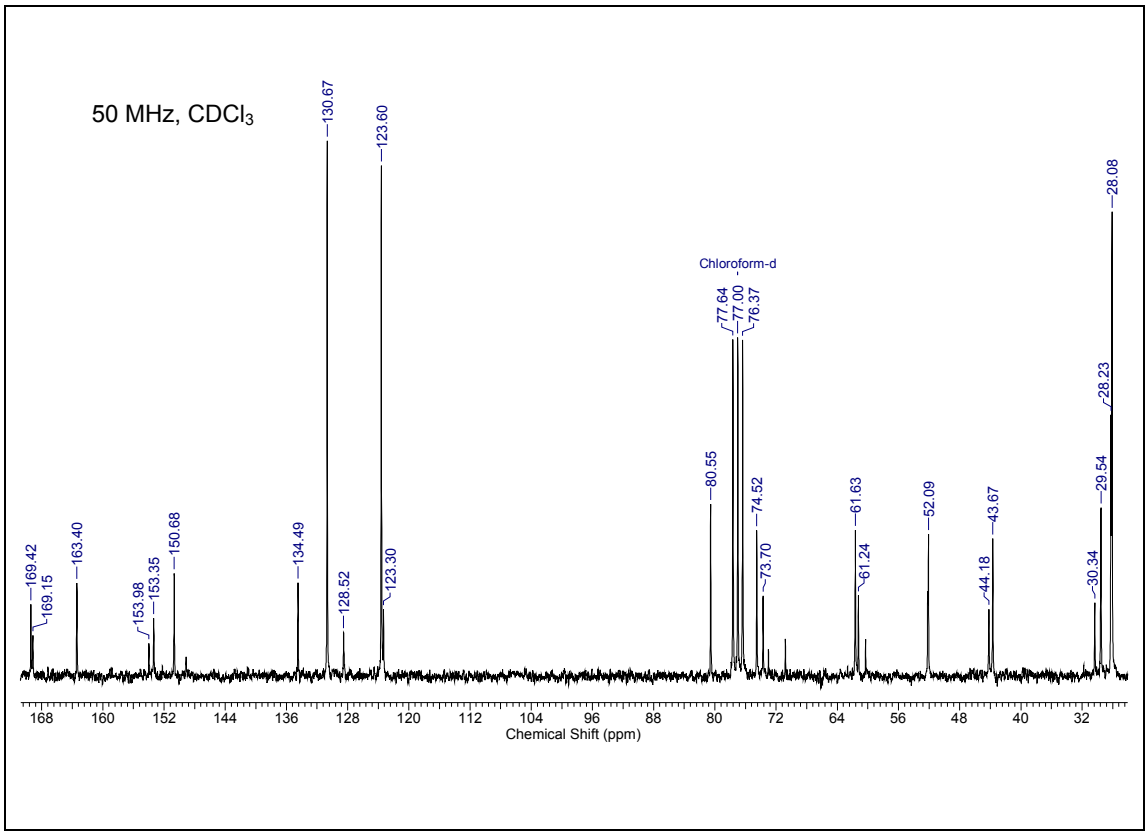
ESI-MS of compound 46



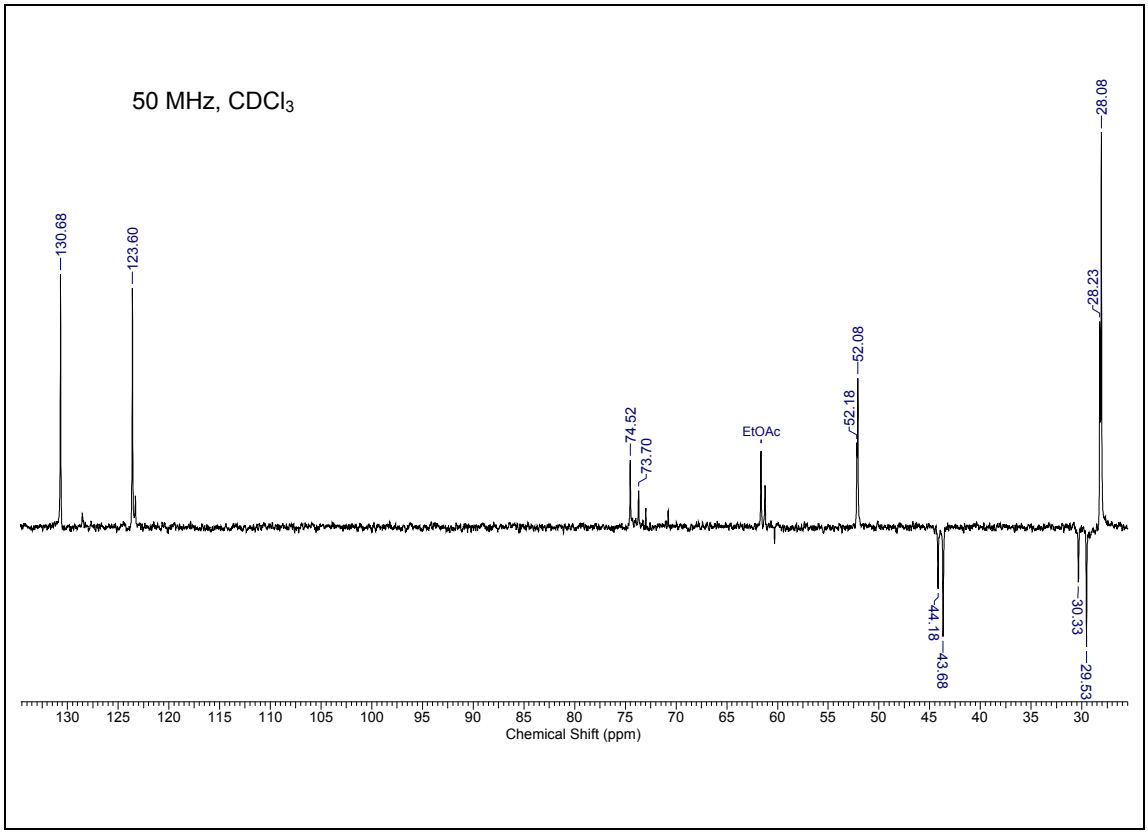
^{1}H NMR of compound 47



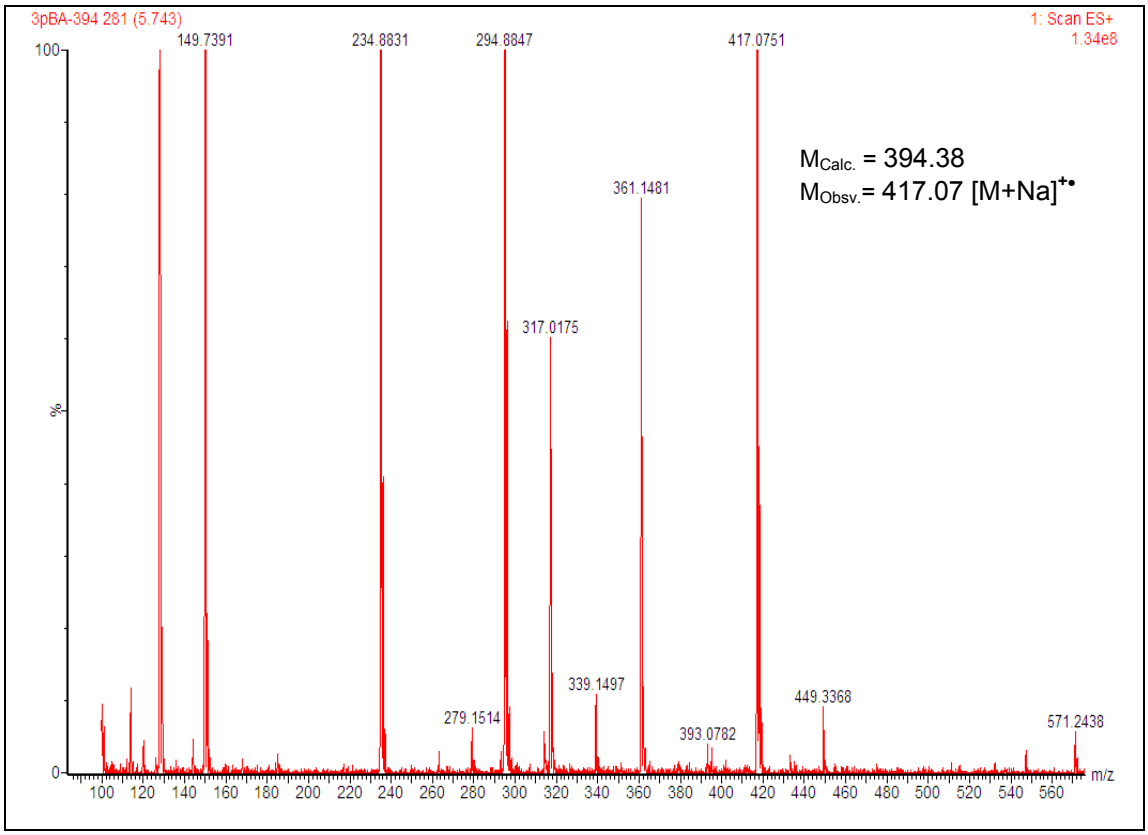
^{13}C NMR of compound 47



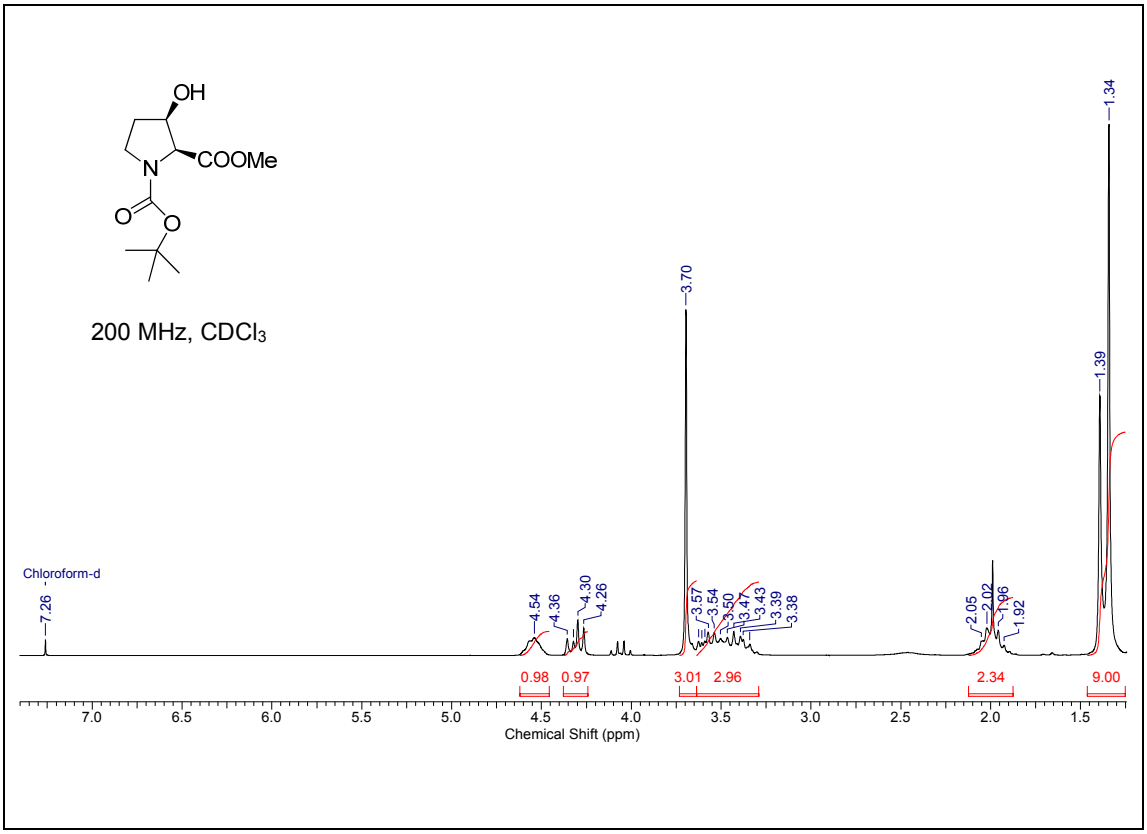
DEPT-^{13}C NMR of compound 47



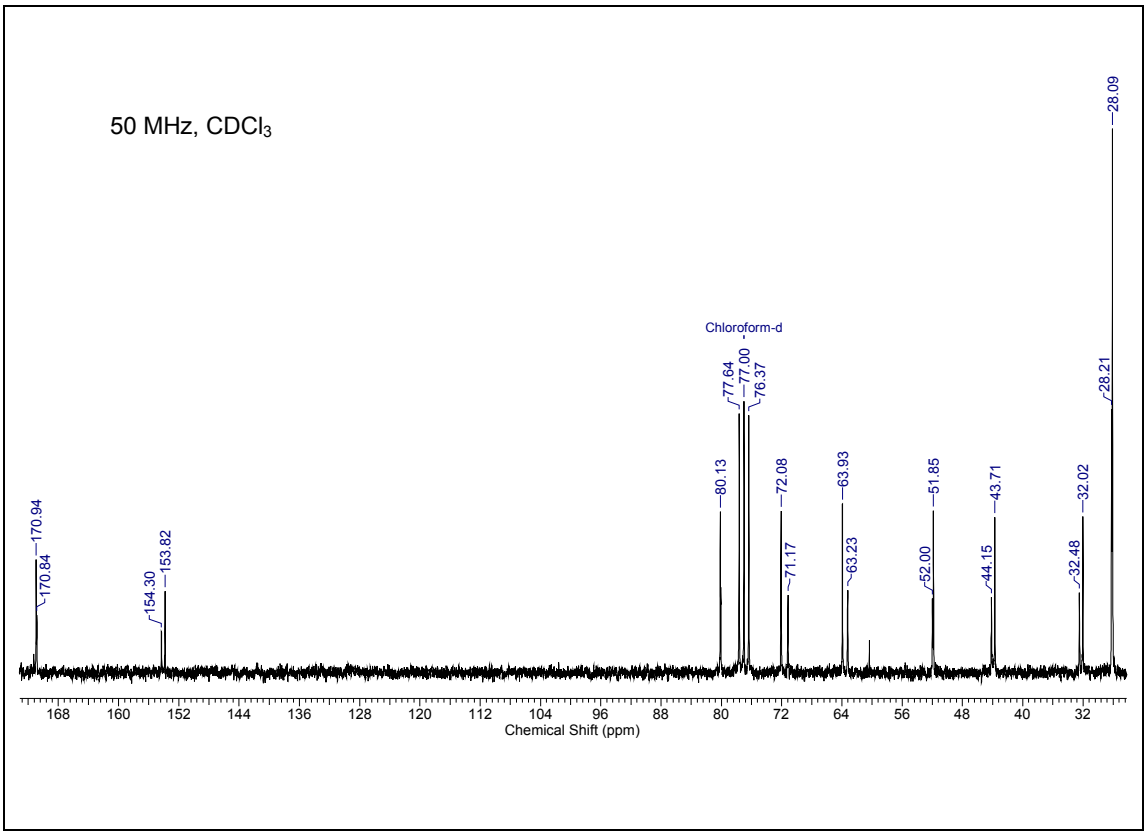
ESI-MS of compound 47



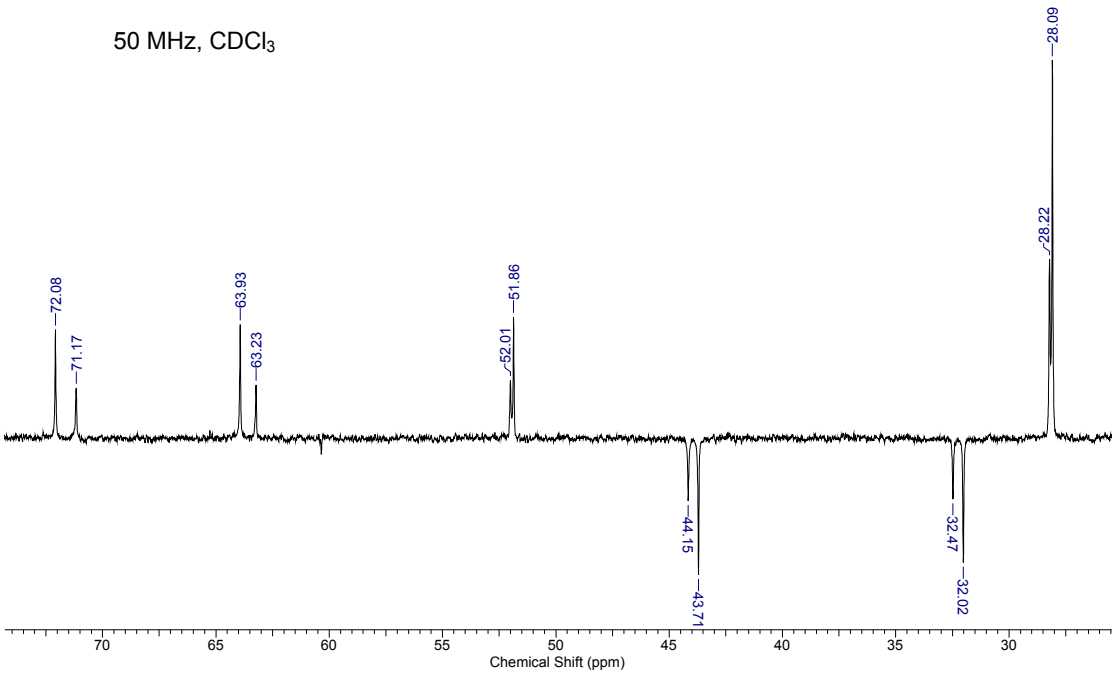
^{1}H NMR of compound 48



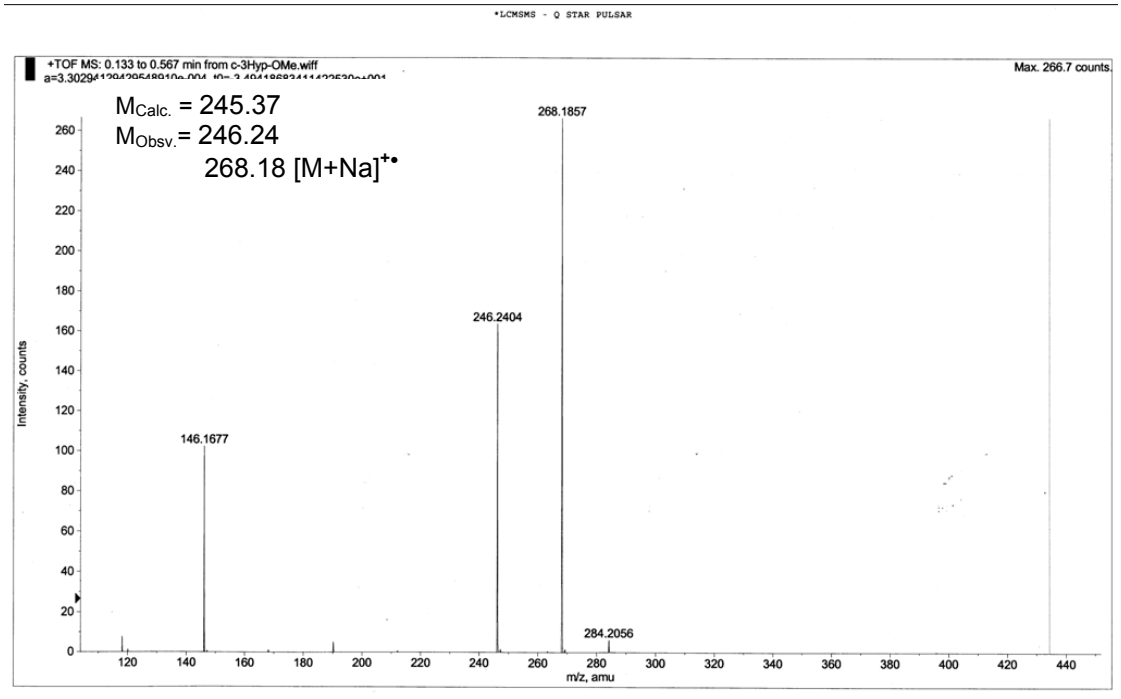
^{13}C NMR of compound 48



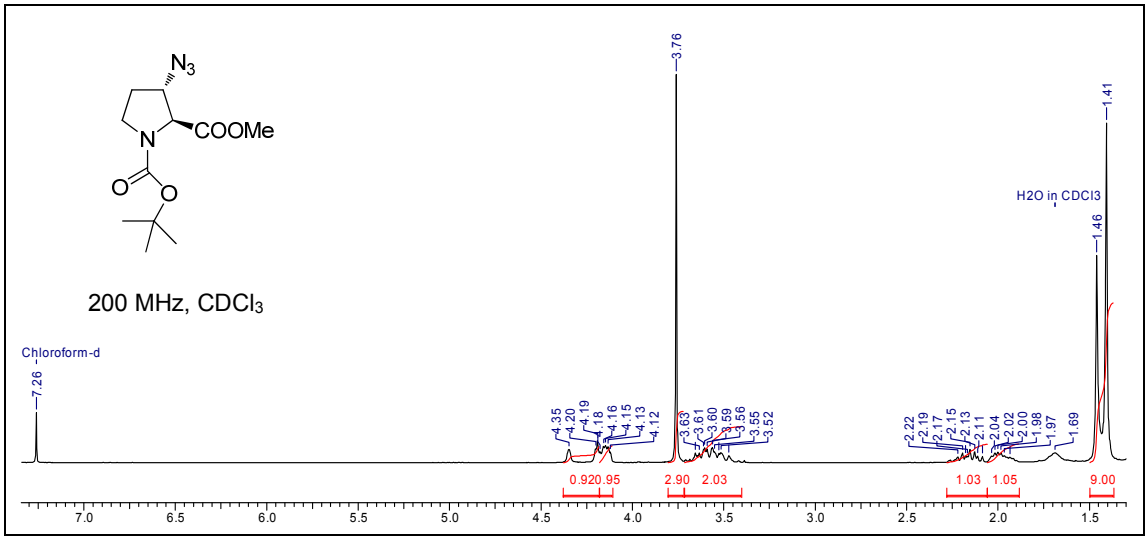
DEPT-^{13}C NMR of compound 48



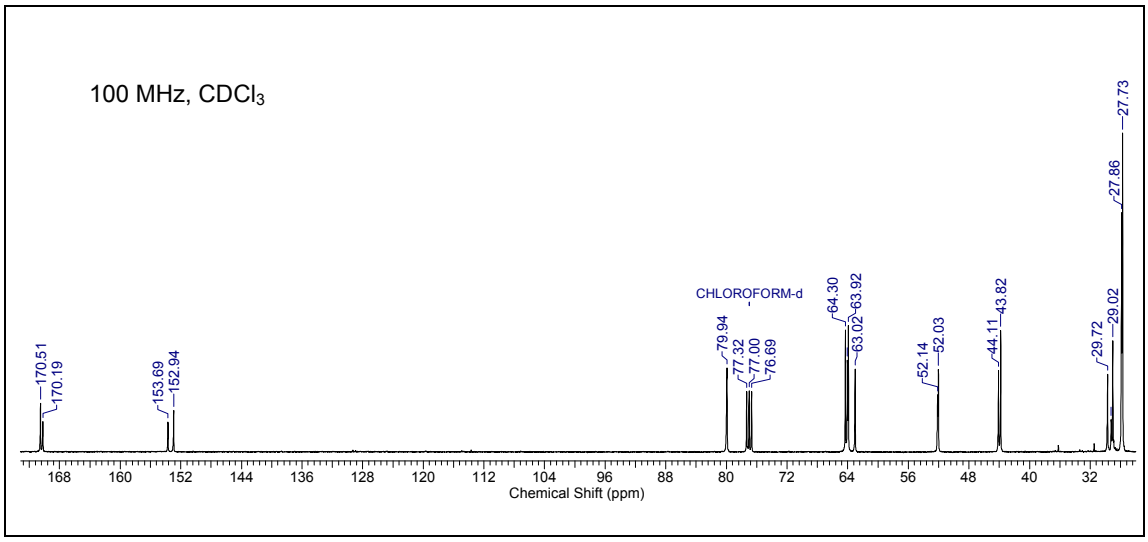
ESI-MS of compound 48



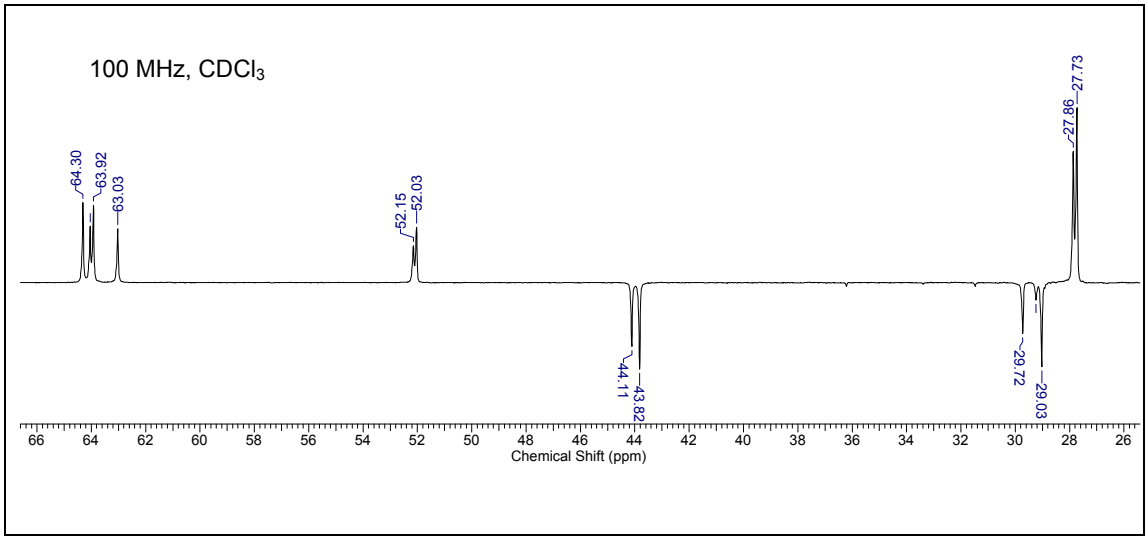
^{1}H NMR of compound 49



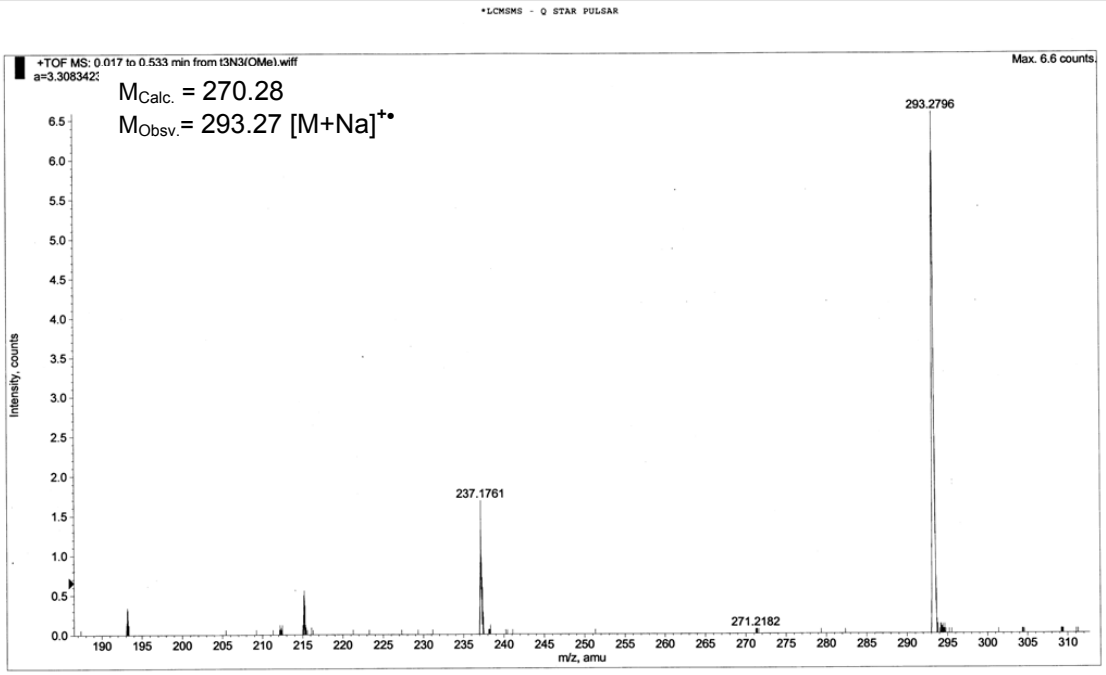
^{13}C NMR of compound 49



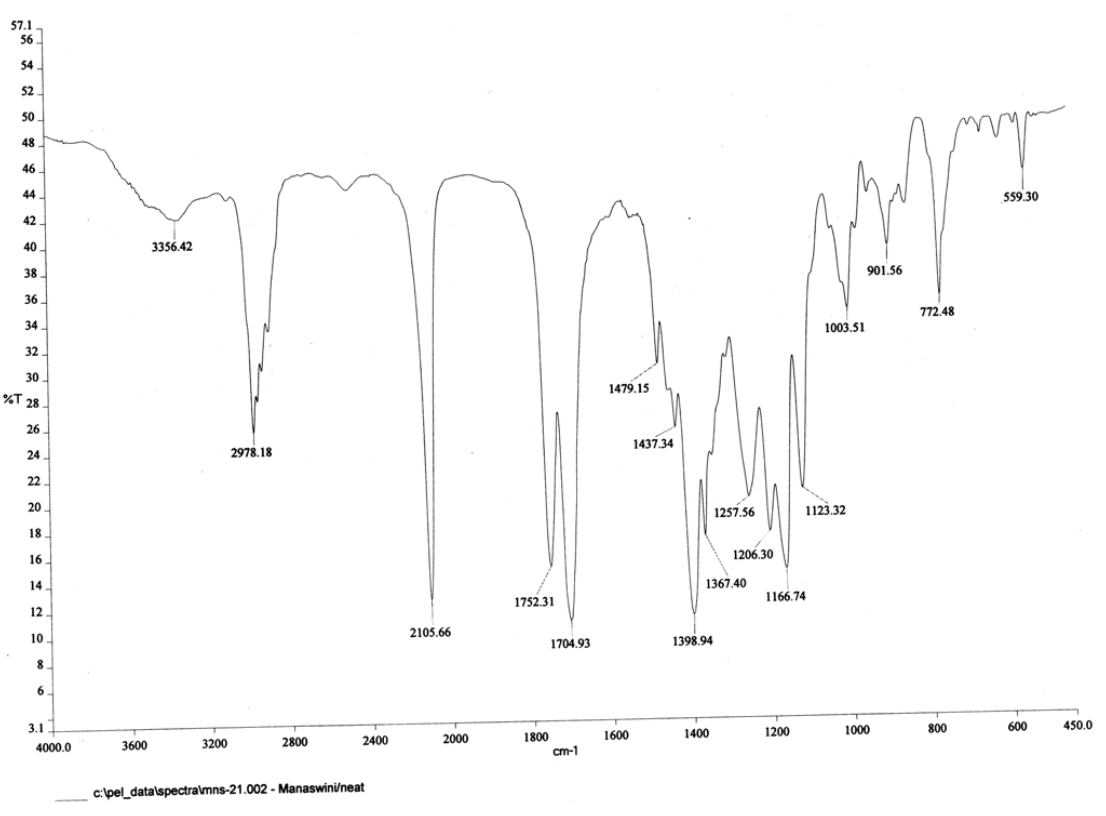
DEPT-^{13}C NMR of compound 49



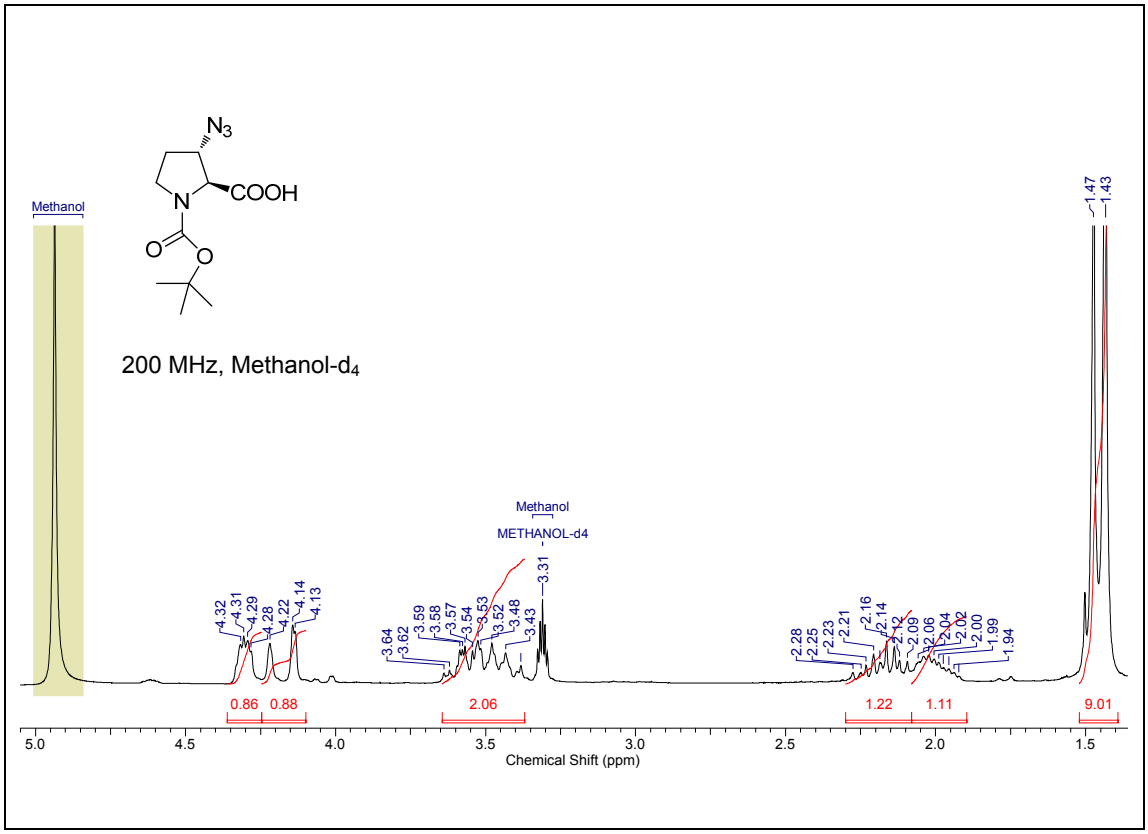
ESI-MS of compound 49



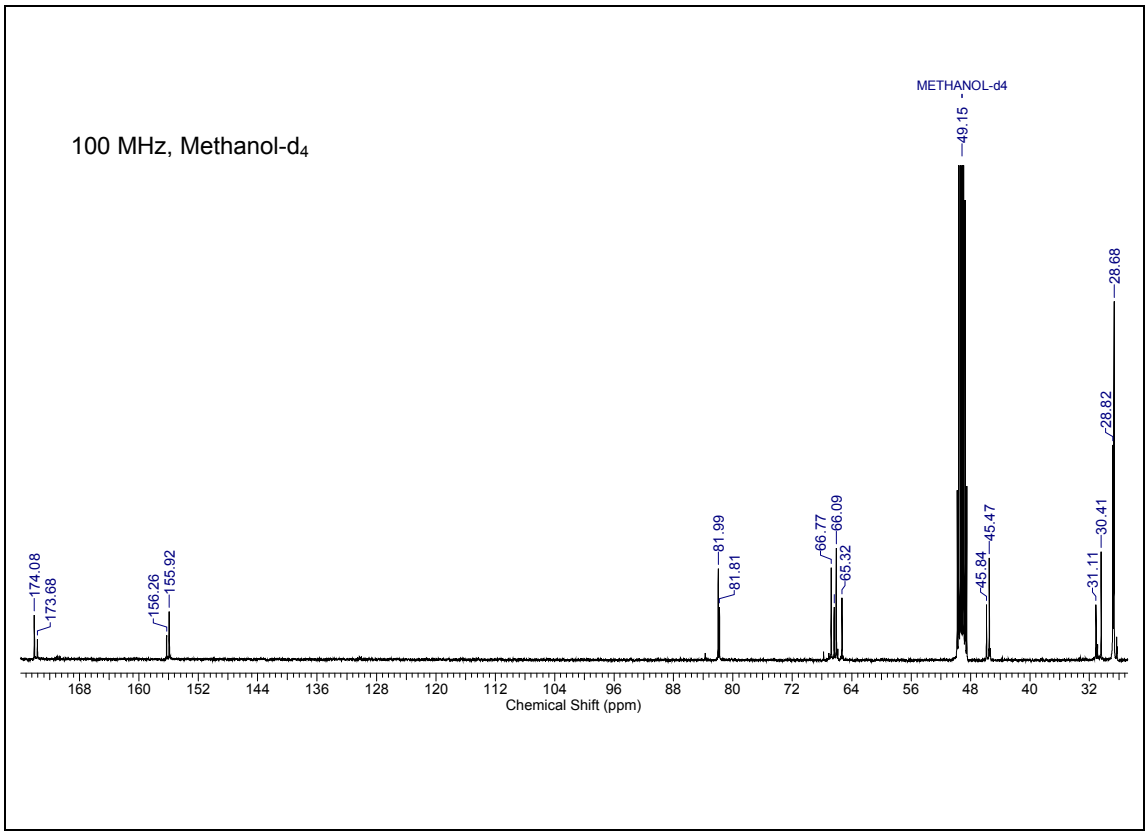
IR Spectra of compound 49



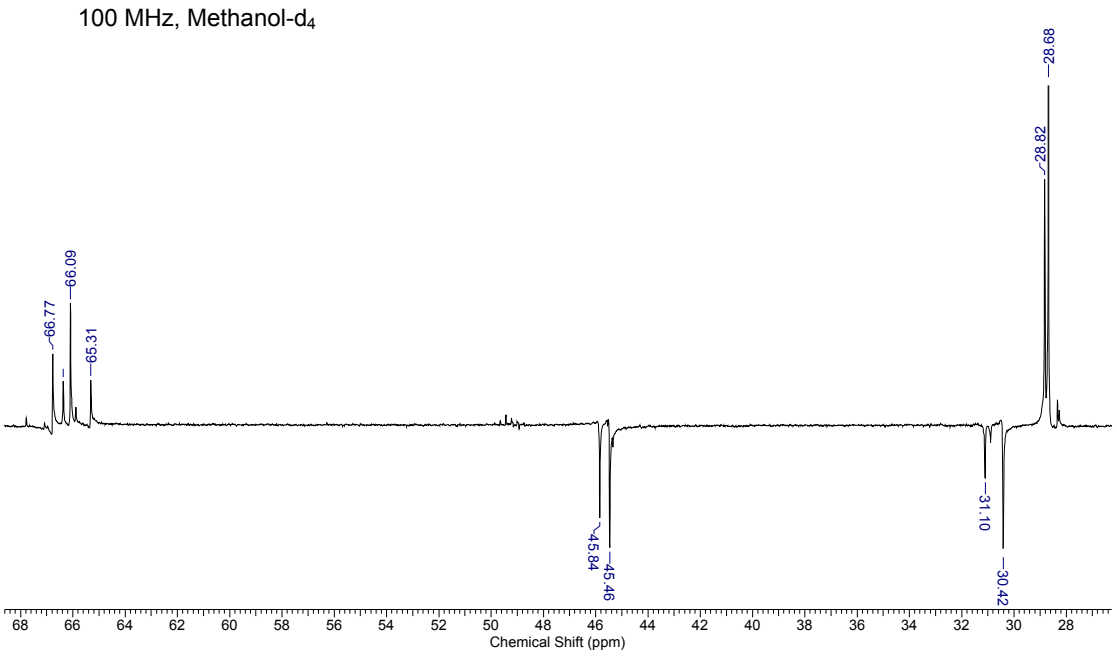
^{1}H NMR of compound 50



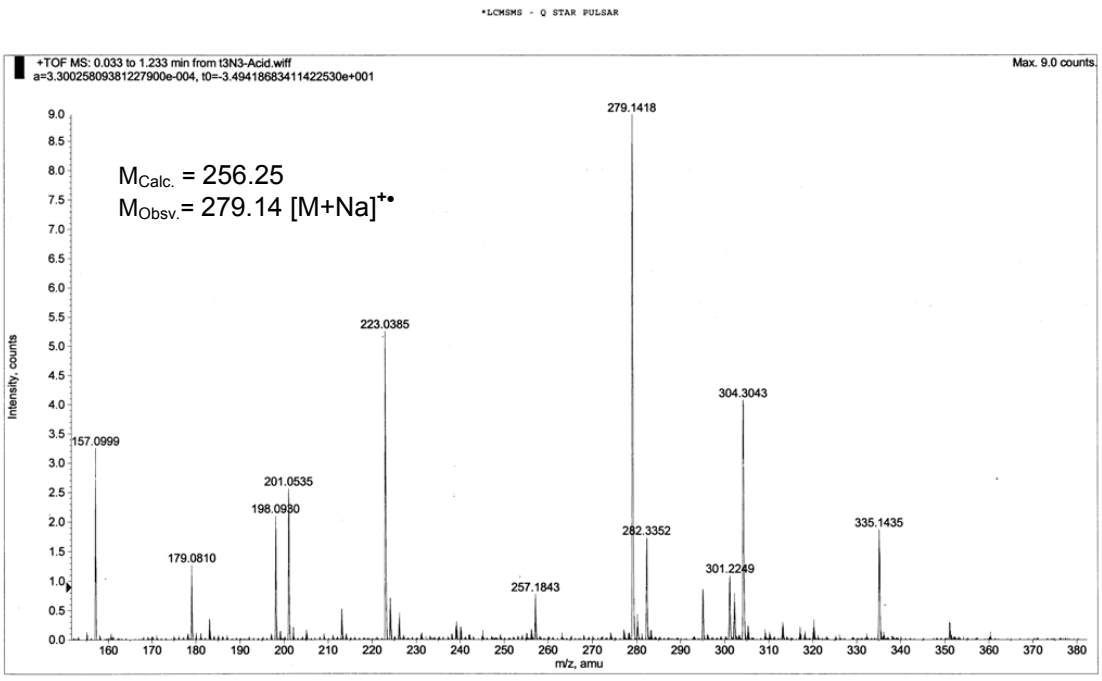
^{13}C NMR of compound 50



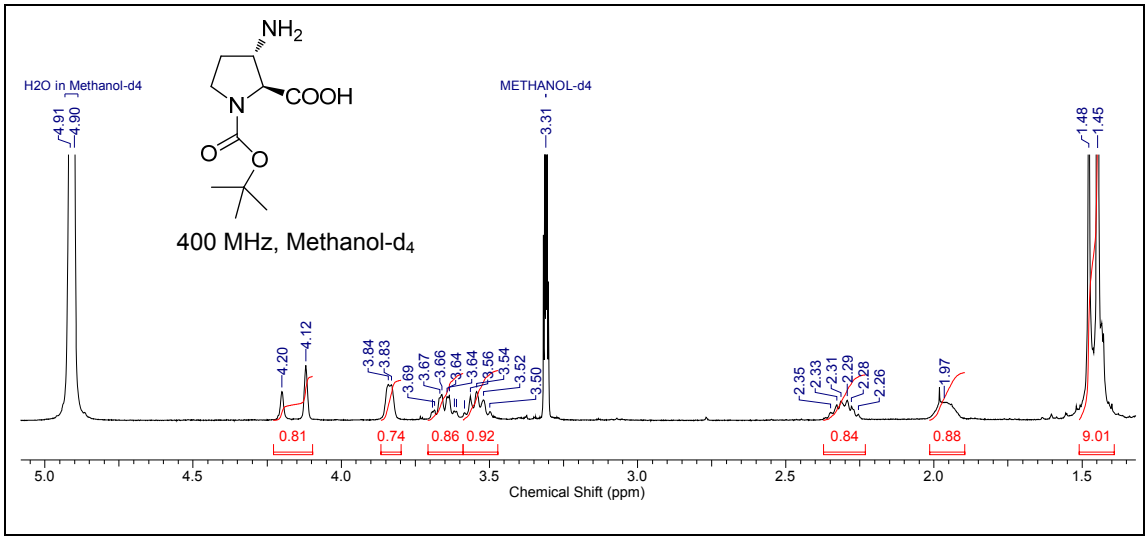
DEPT-^{13}C NMR of compound 50



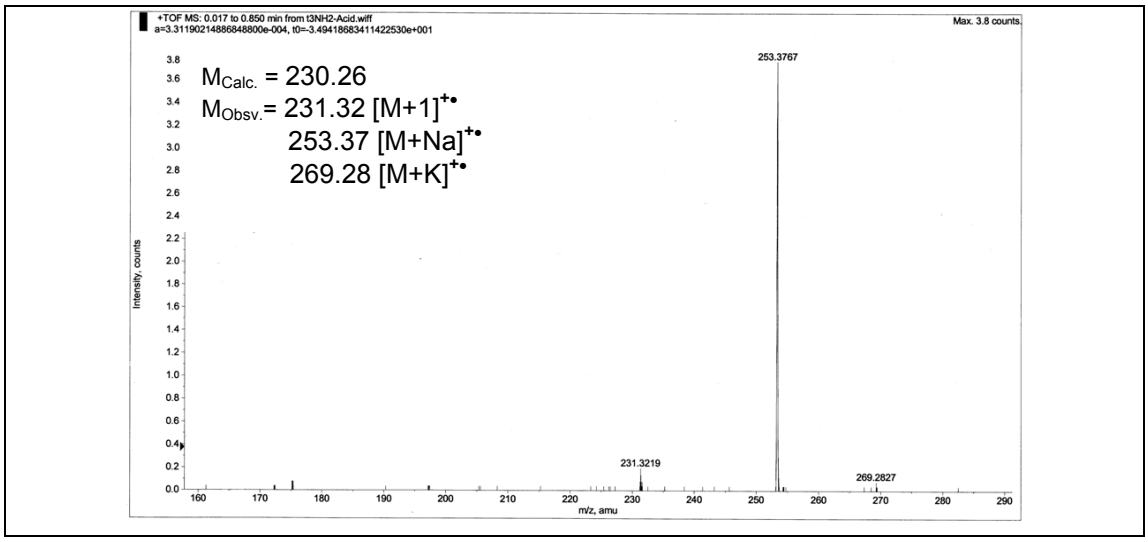
ESI-MS of compound 50



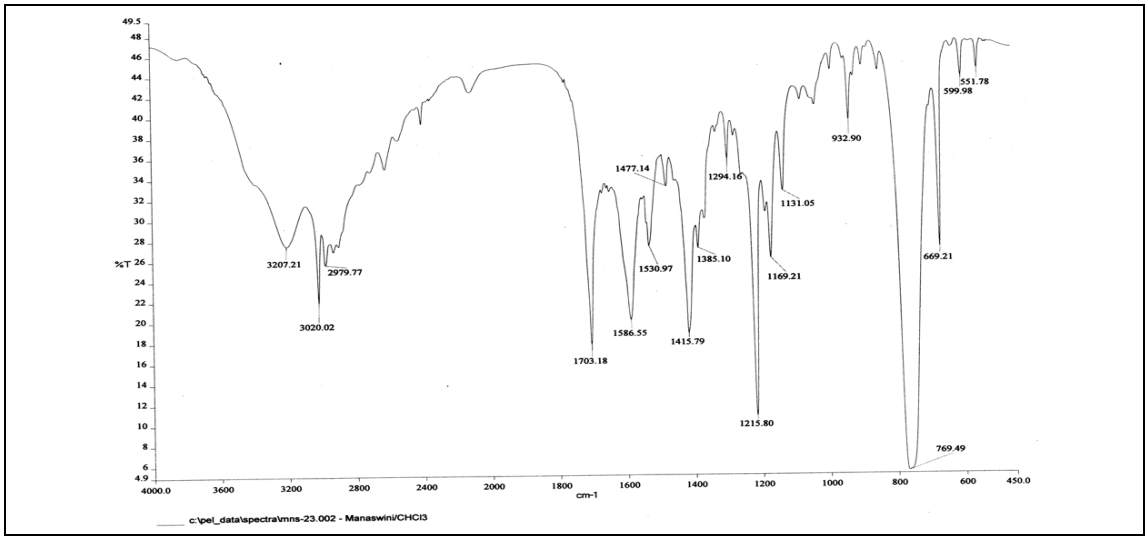
^{1}H NMR of compound 51



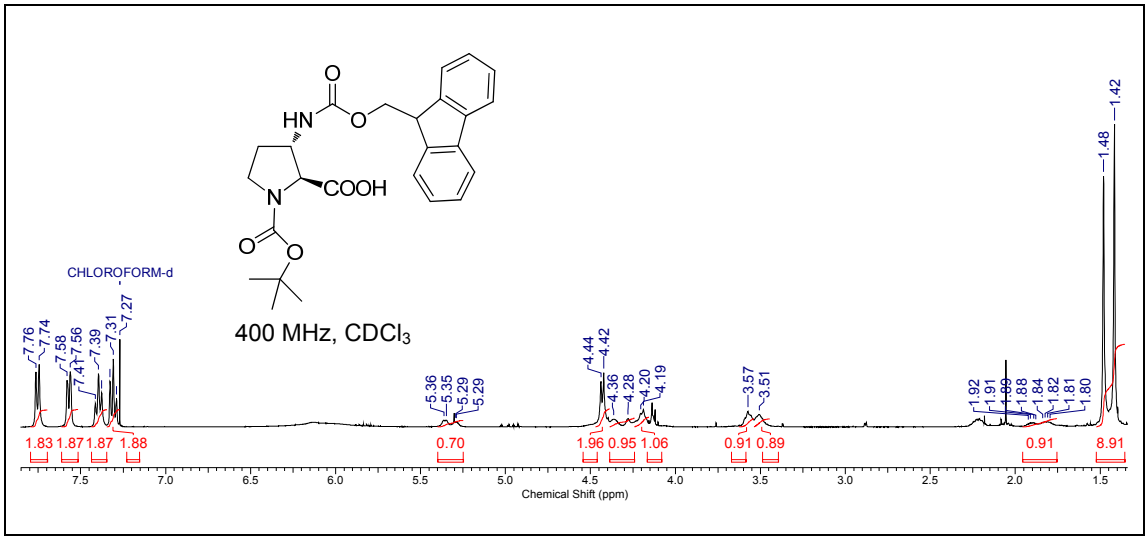
ESI-MS of compound 51



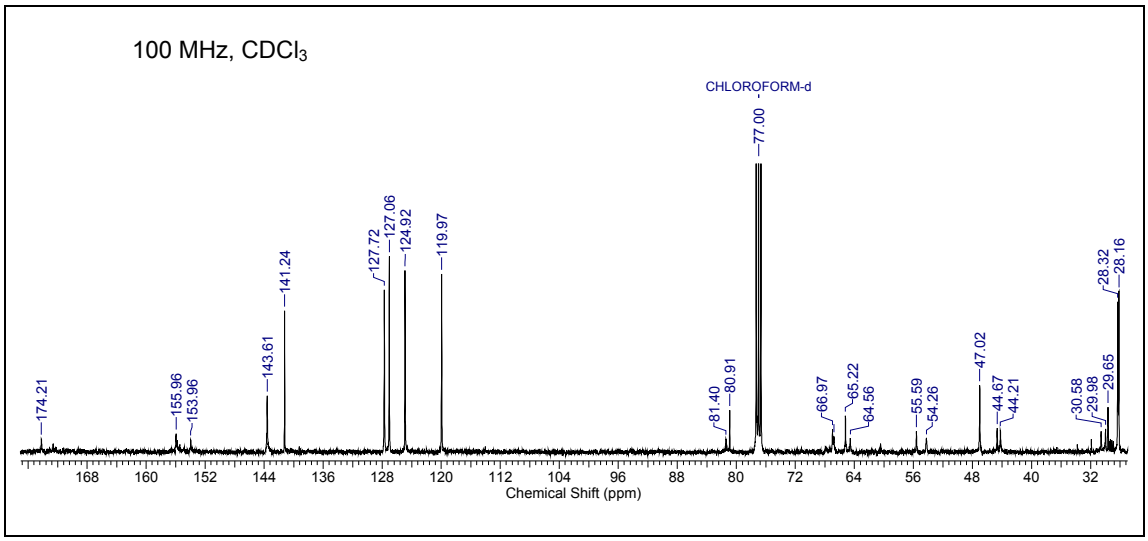
IR Spectra of compound 51



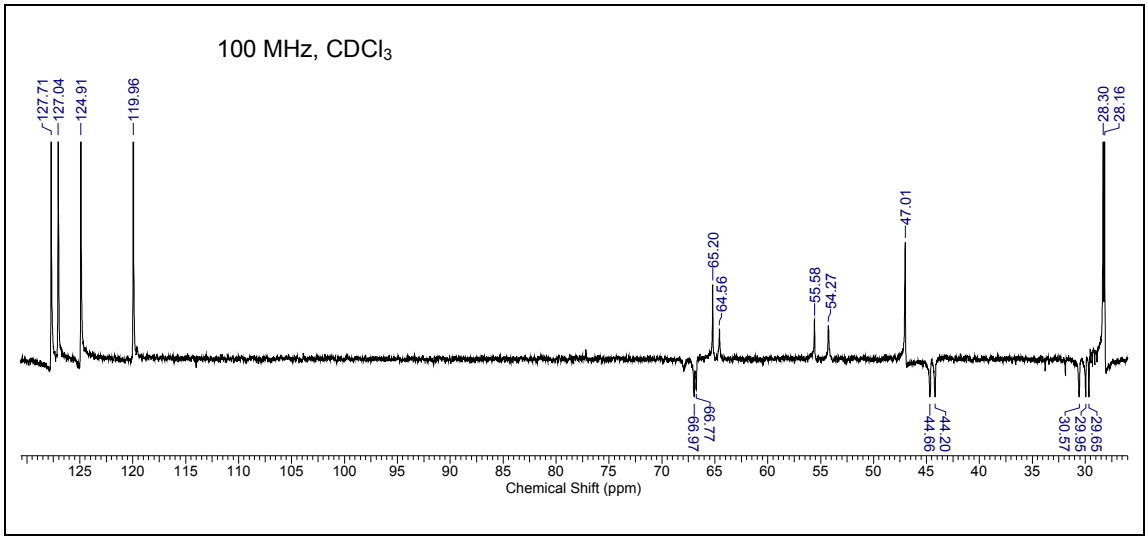
^{1}H NMR of compound 52



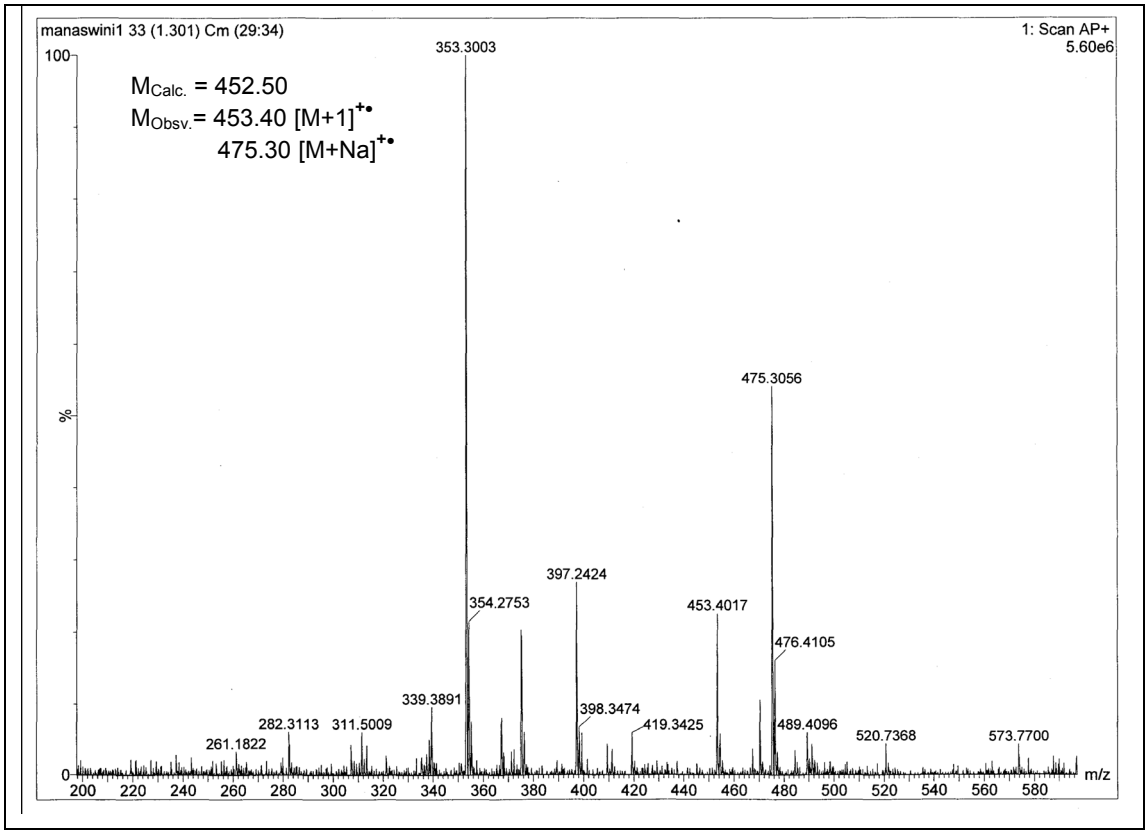
^{13}C NMR of compound 52



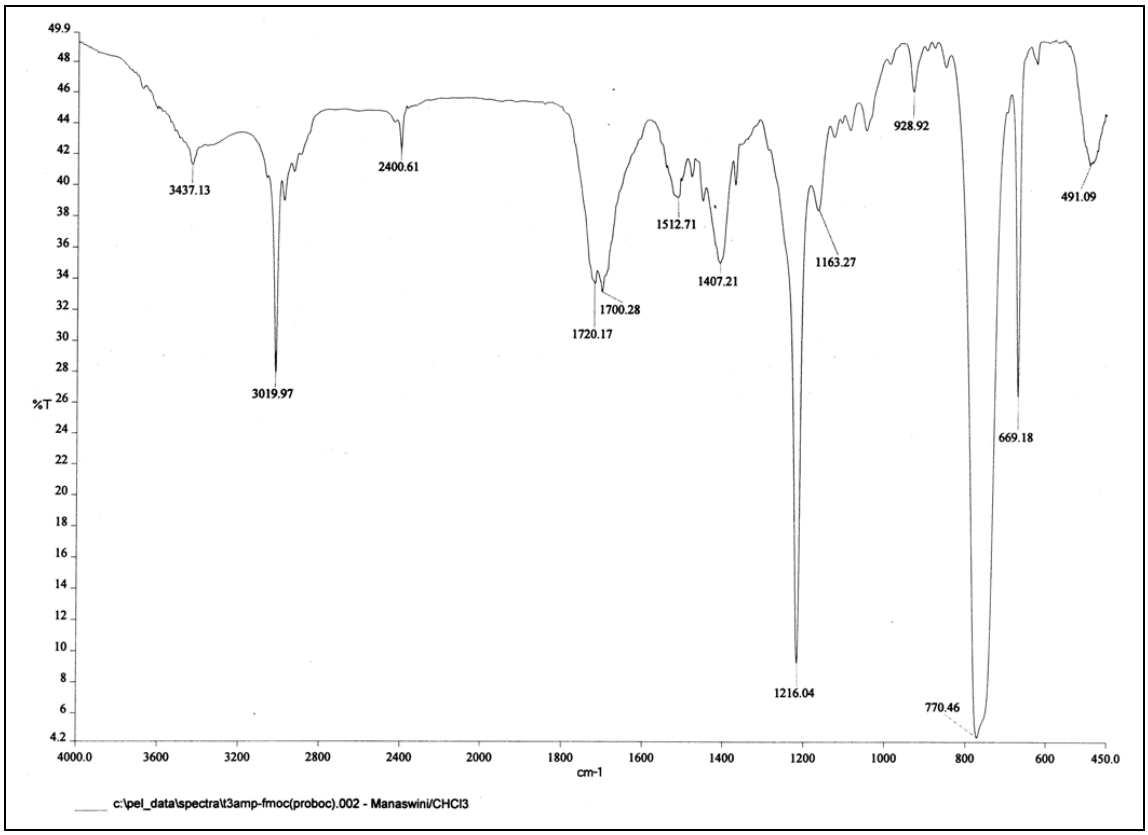
DEPT-^{13}C NMR of compound 52



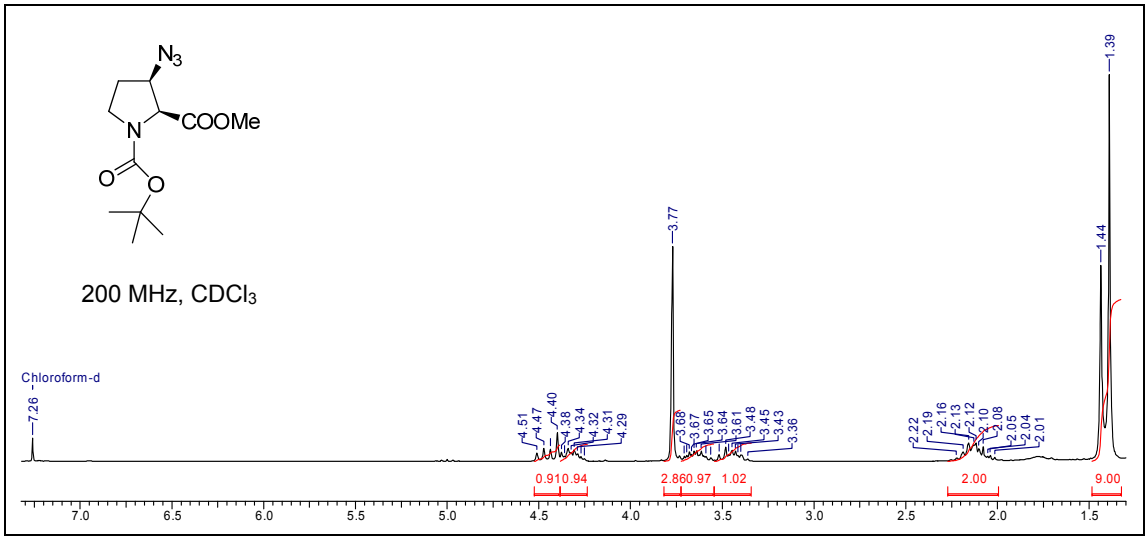
ESI-MS of compound 52



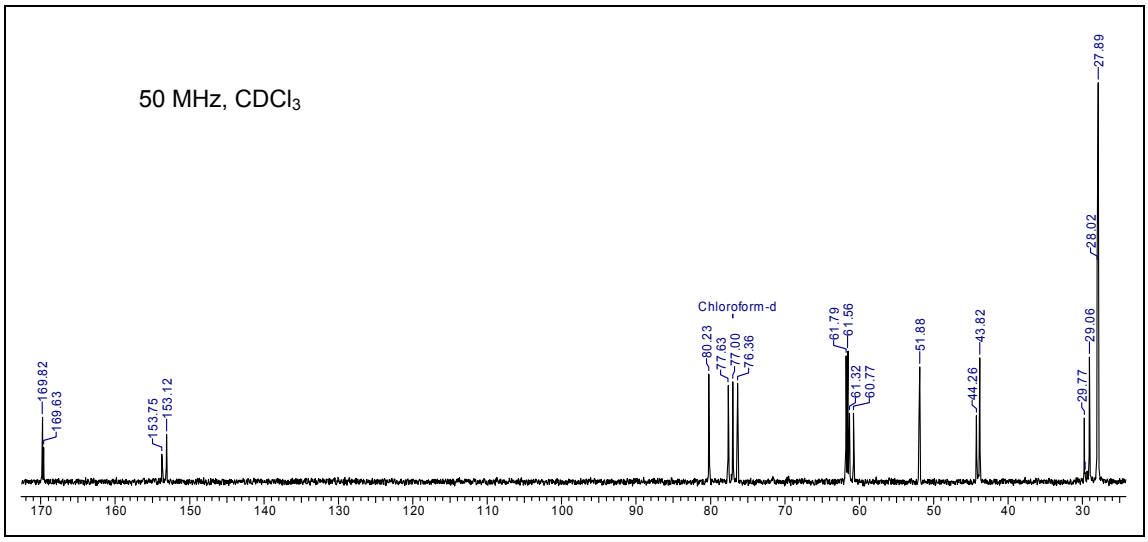
IR Spectra of compound 52



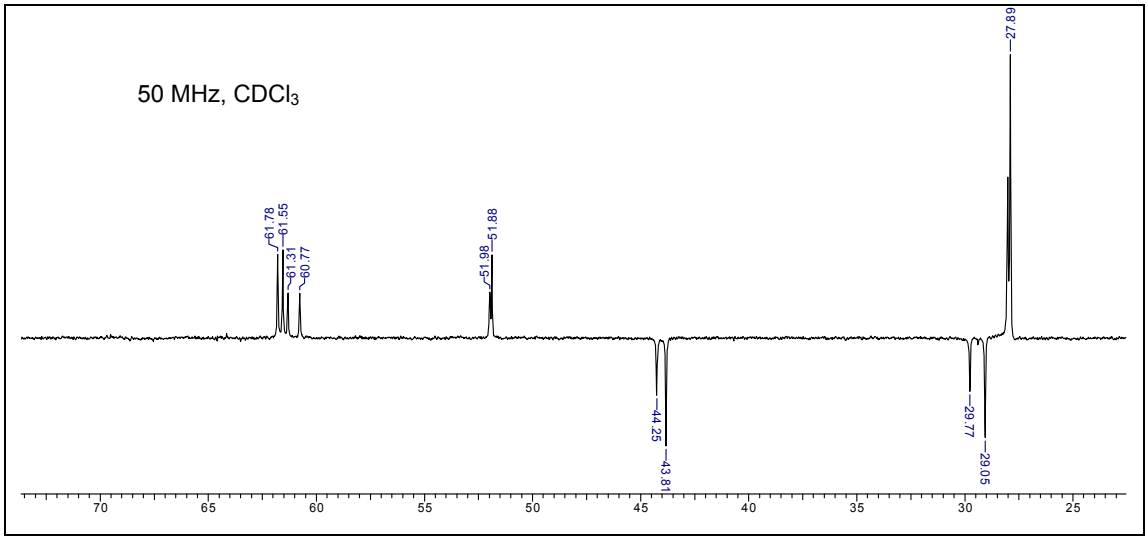
^{1}H NMR of compound 53



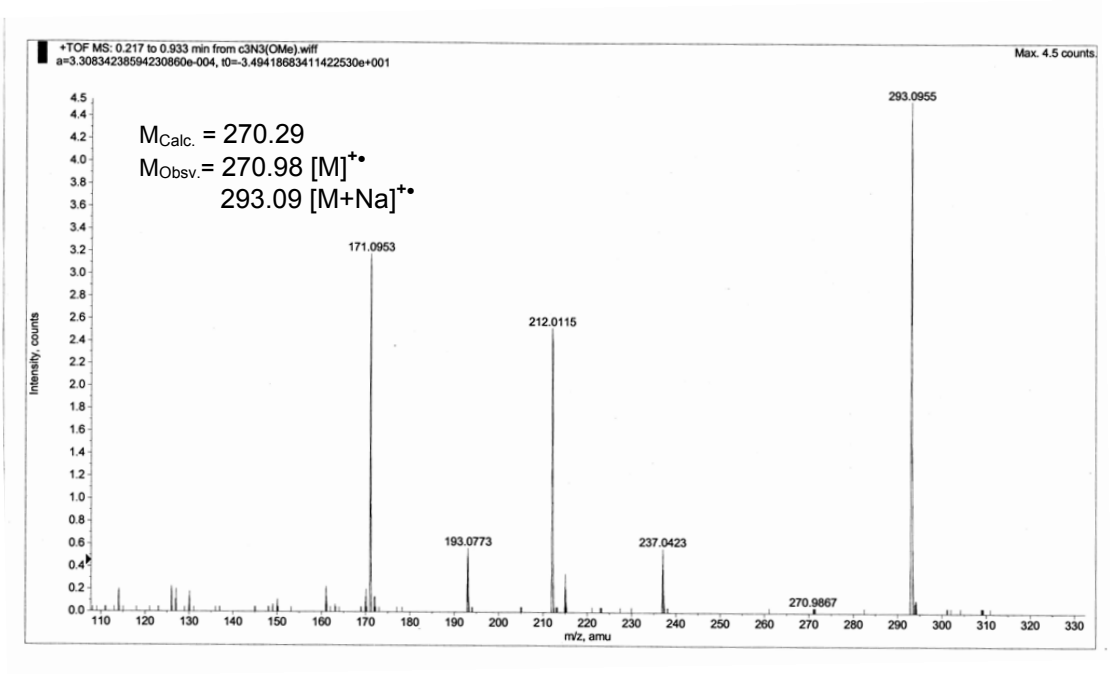
^{13}C NMR of compound 53



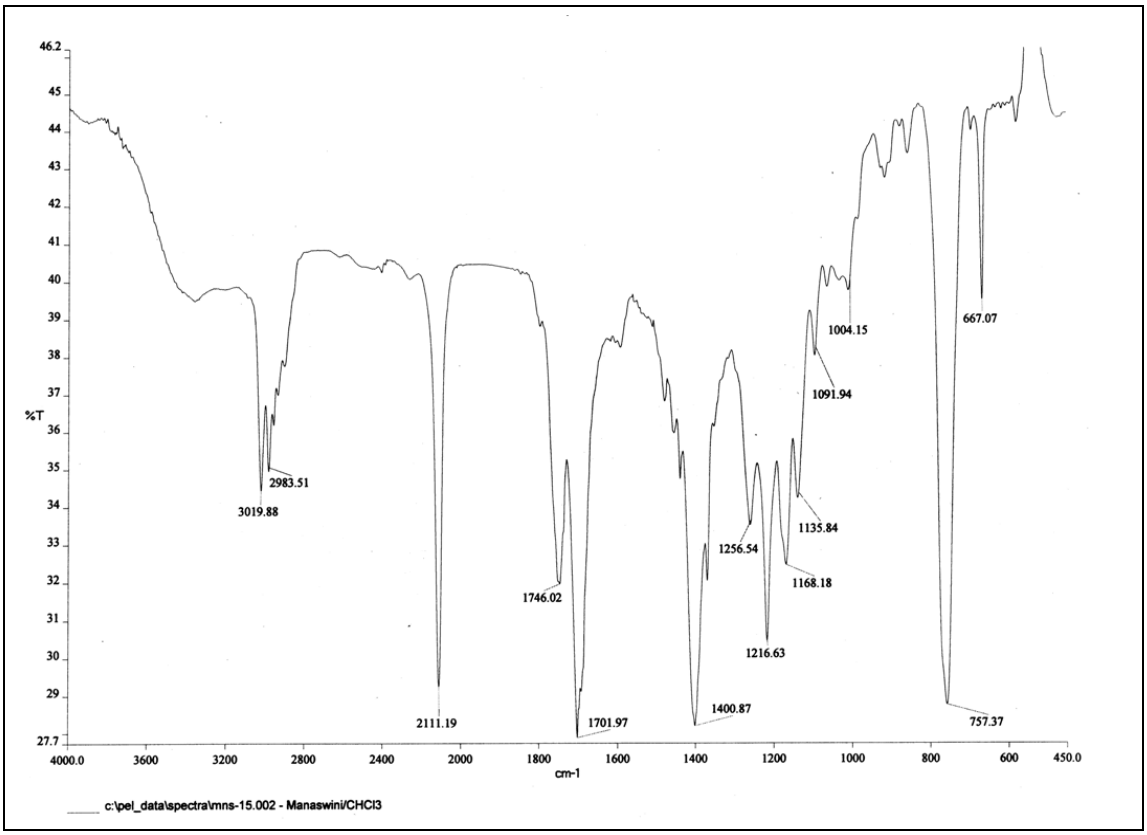
DEPT-^{13}C NMR of compound 53



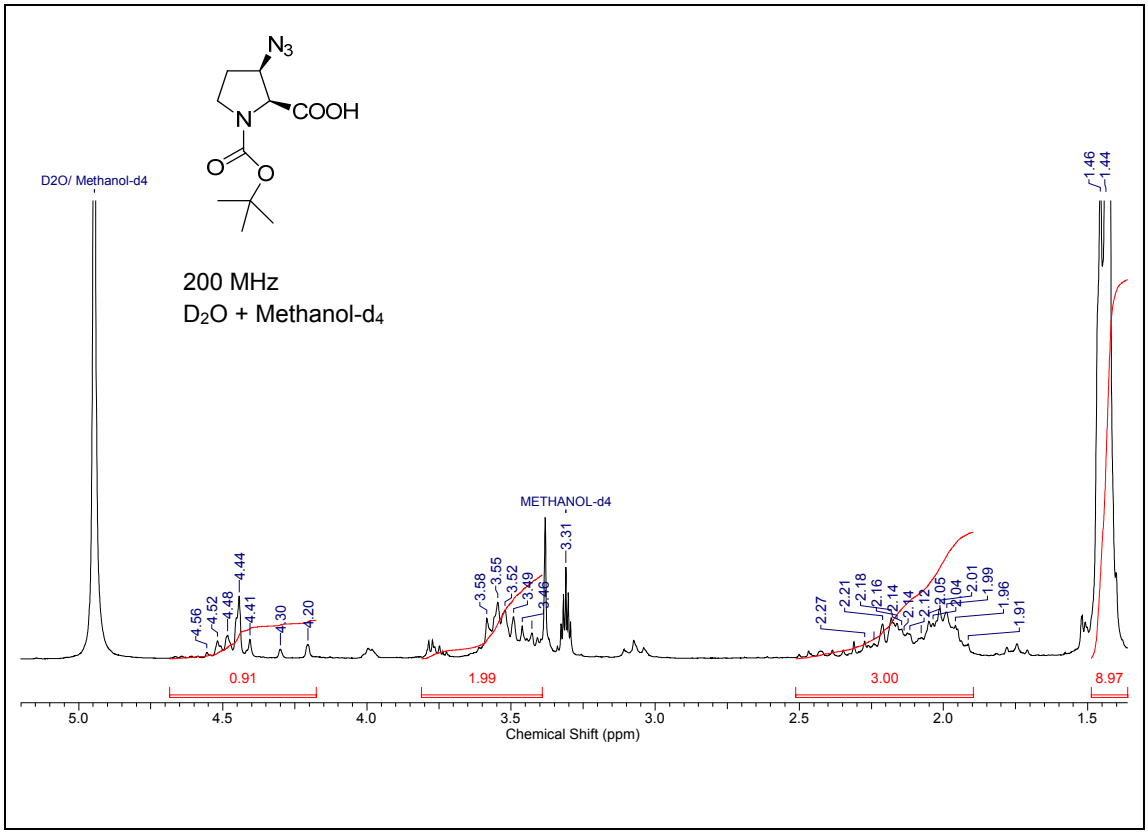
ESI-MS of compound 53



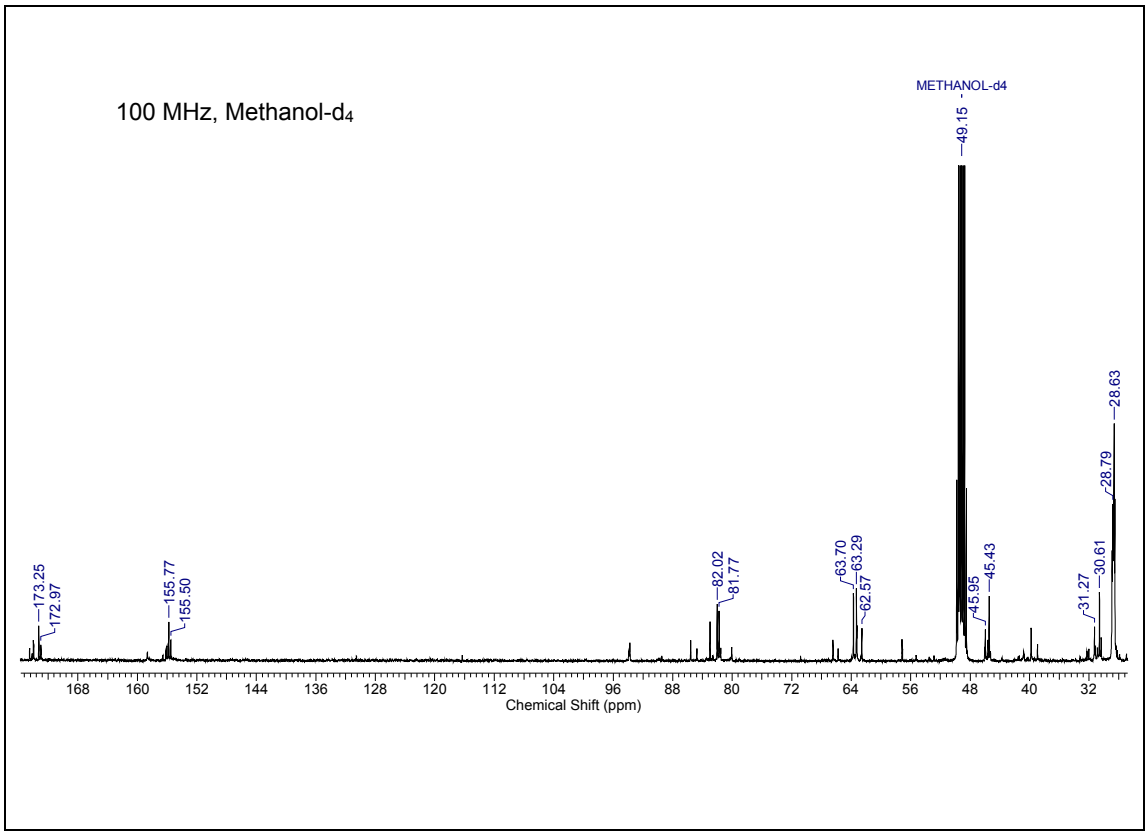
IR Spectra of compound 53



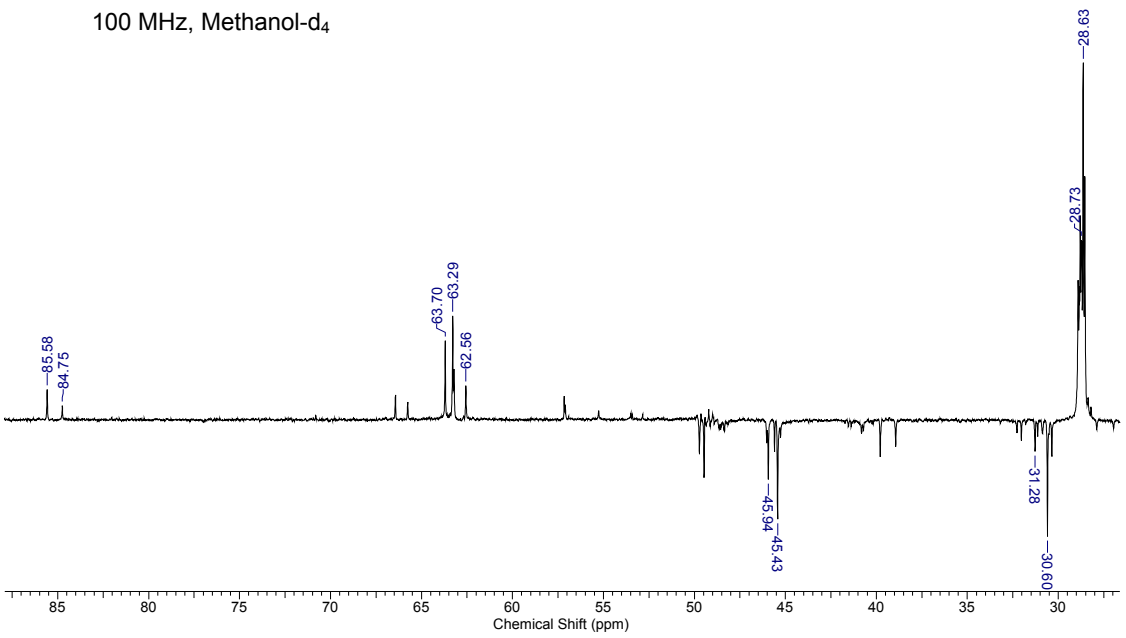
^{1}H NMR of compound 54



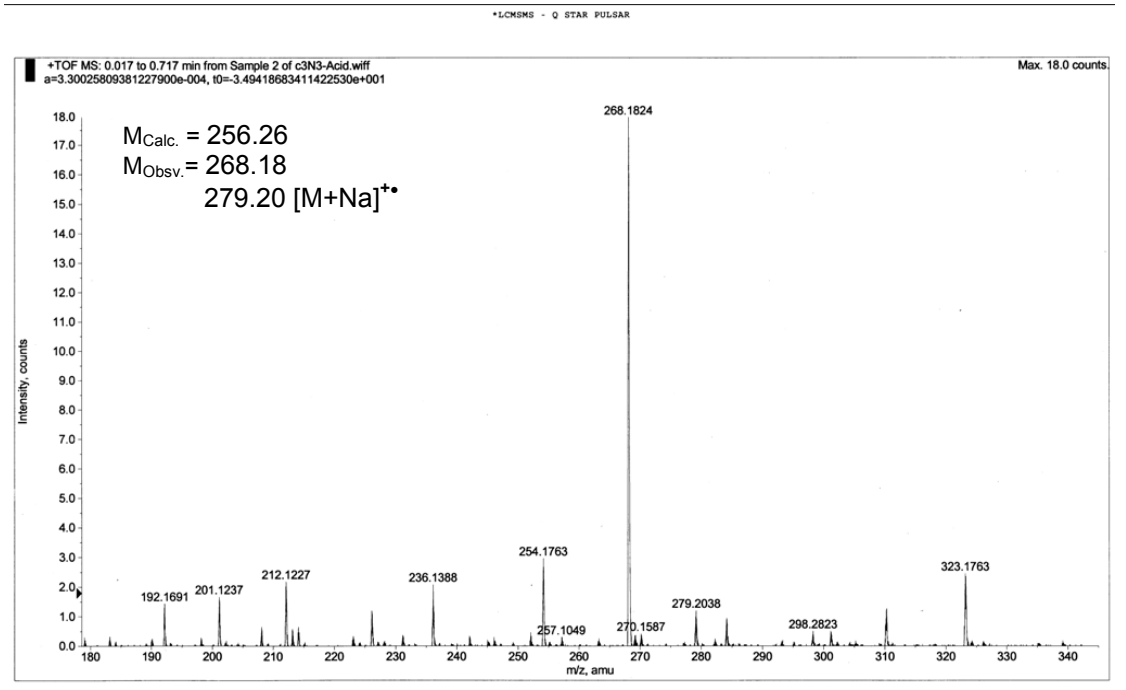
^{13}C NMR of compound 54



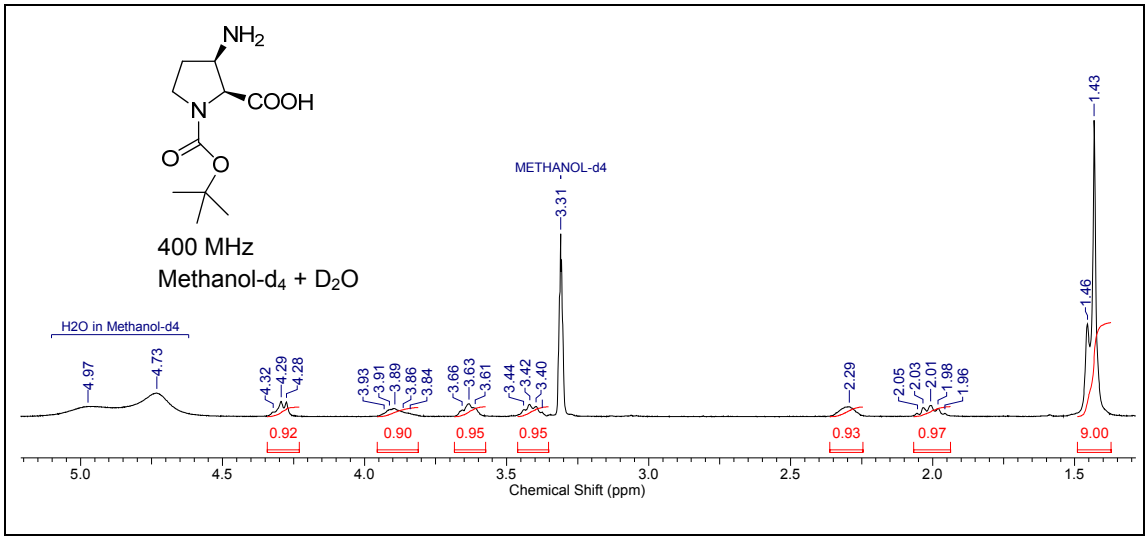
DEPT-^{13}C NMR of compound 54



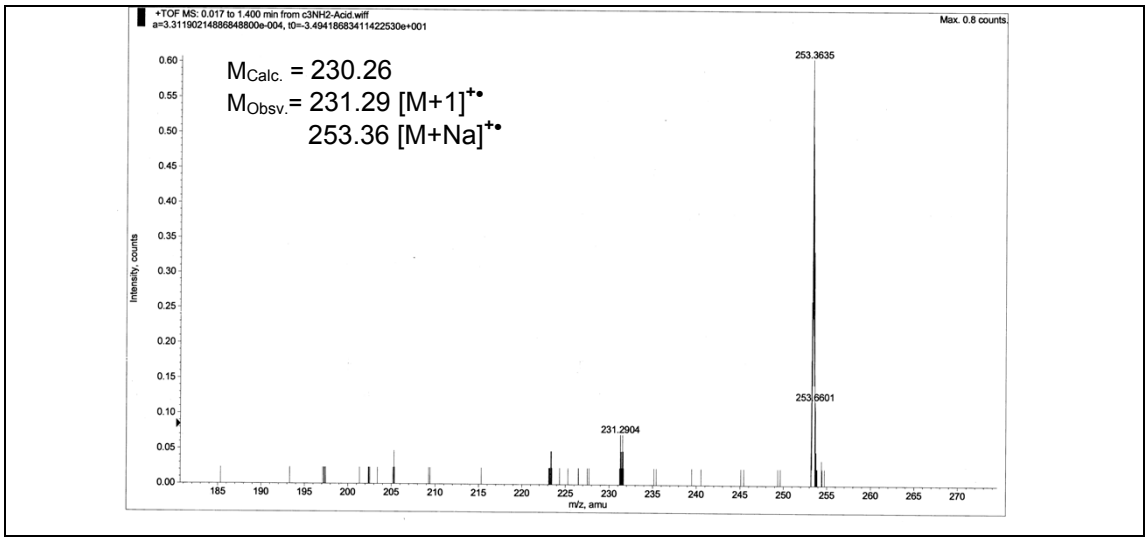
ESI-MS of compound 54



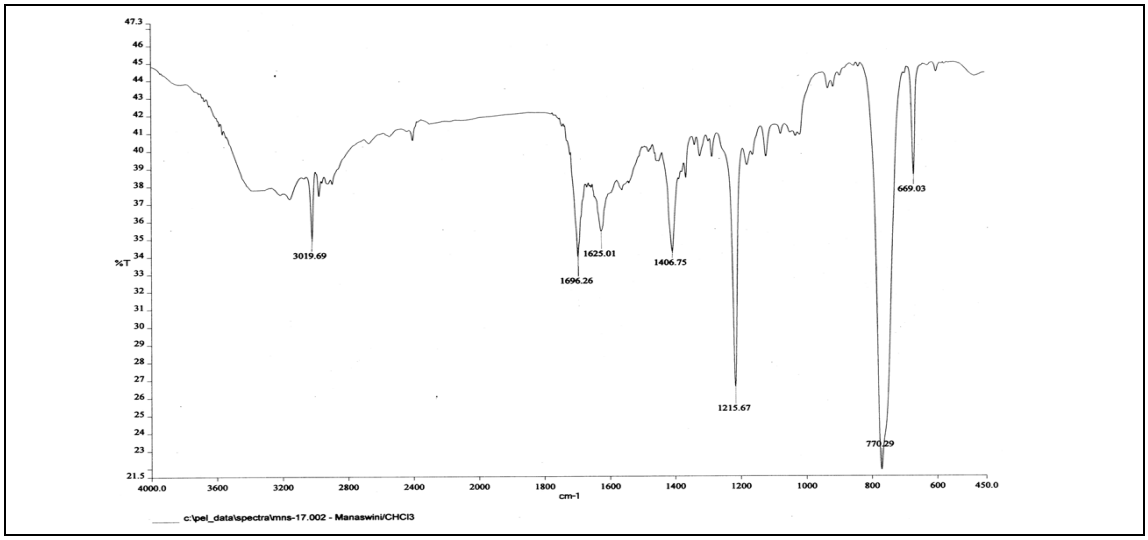
^{1}H NMR of compound 55



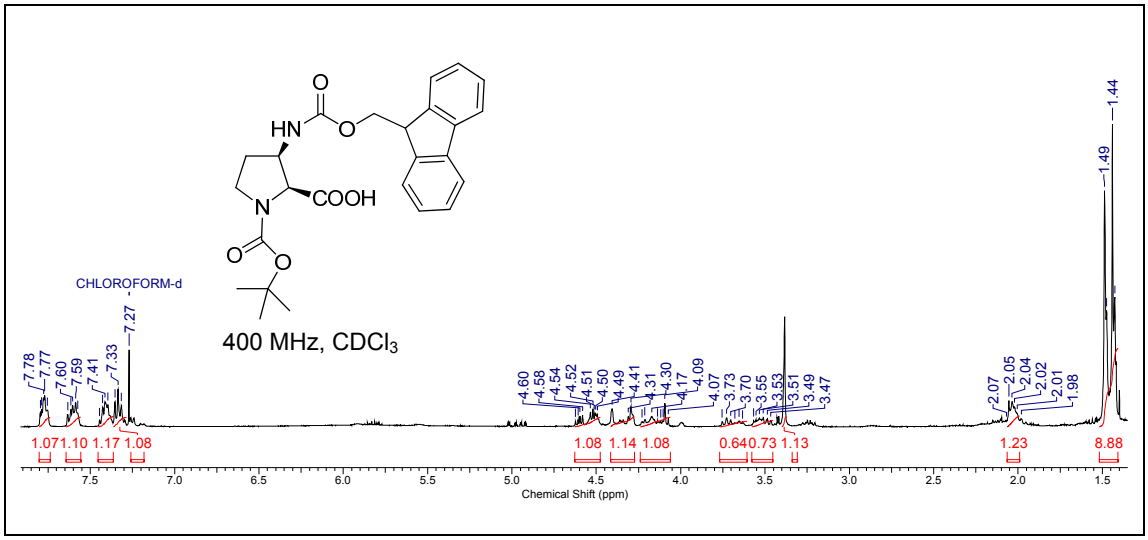
ESI-MS of compound 55



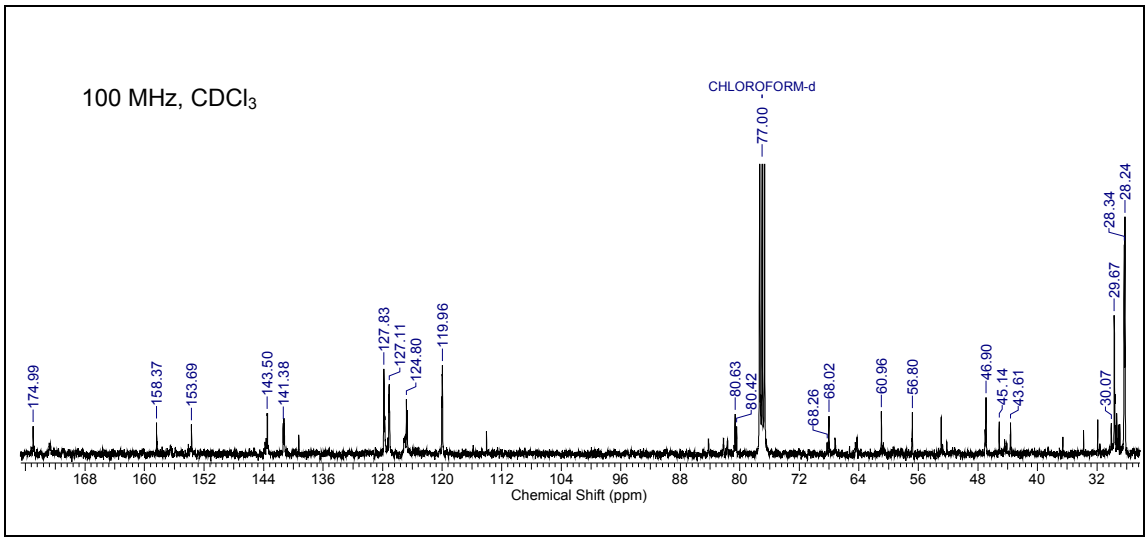
IR Spectra of compound 55



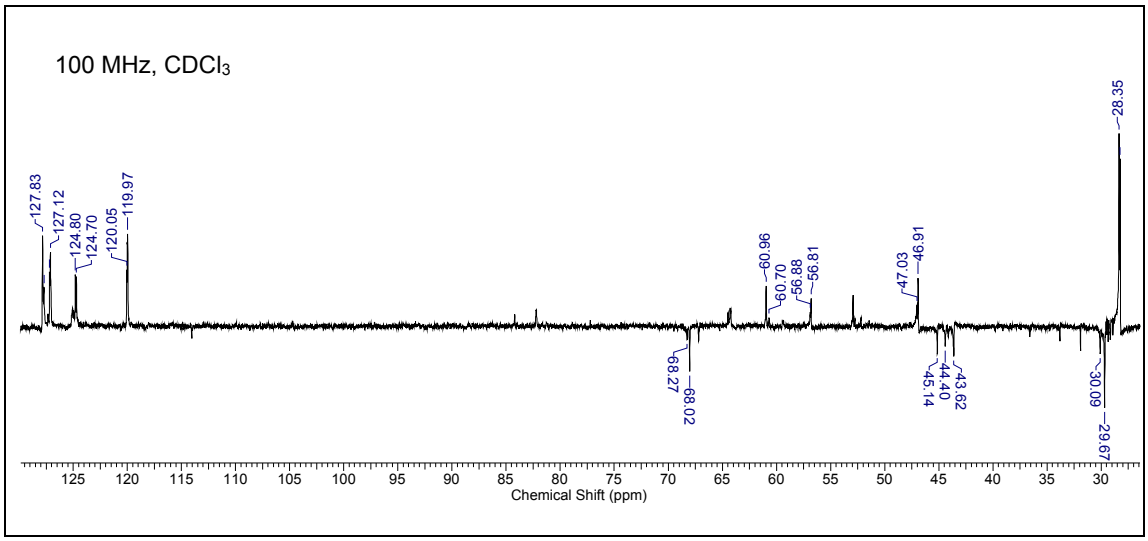
^{1}H NMR of compound 56



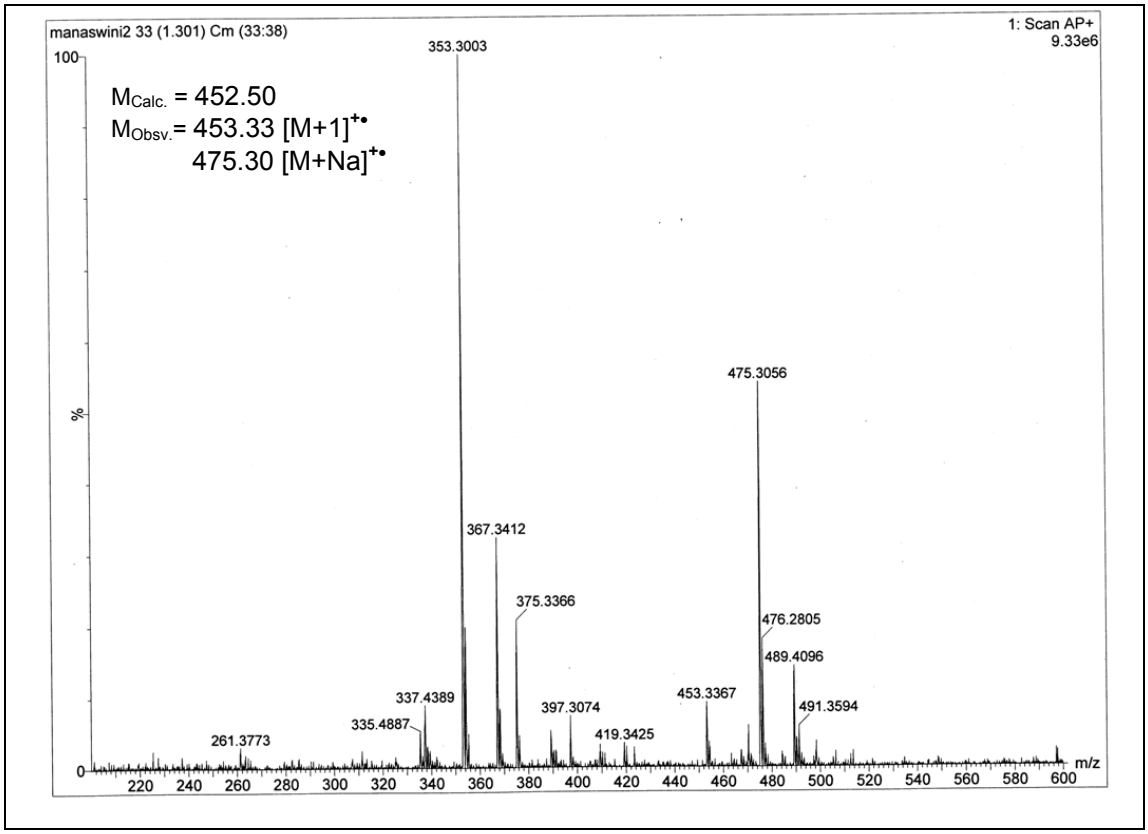
^{13}C NMR of compound 56



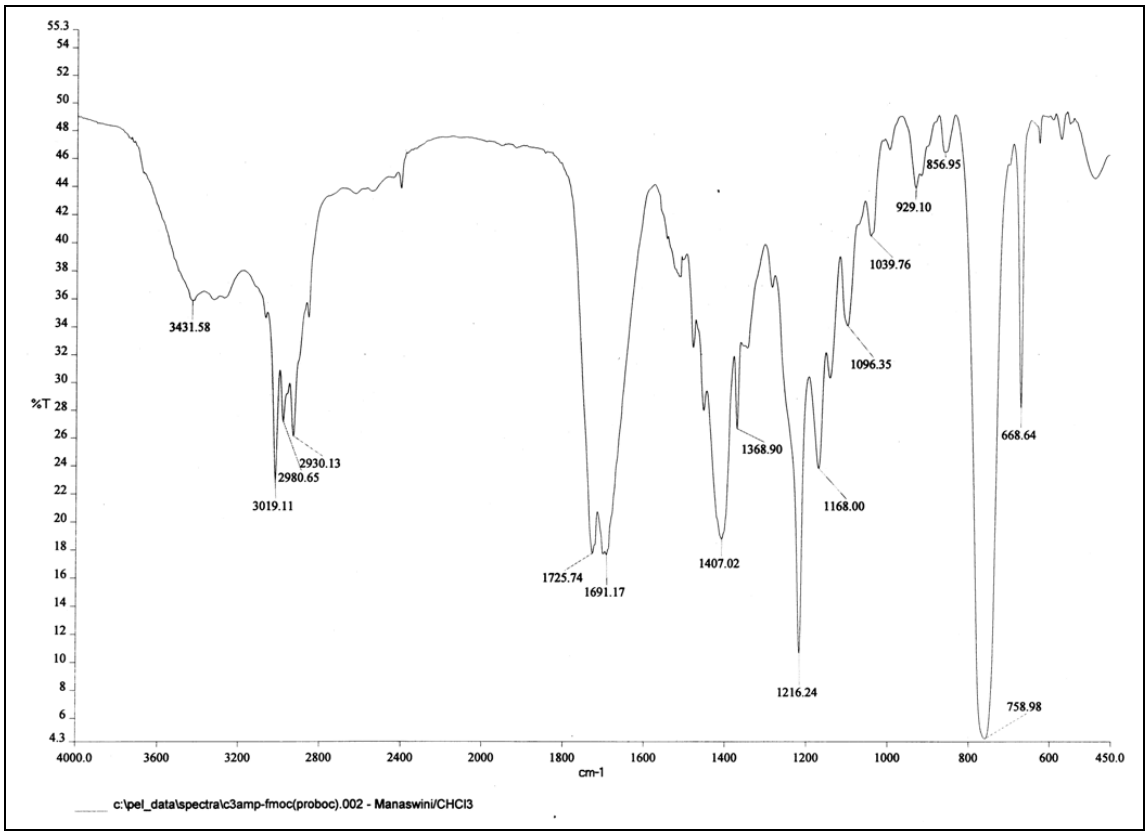
DEPT-^{13}C NMR of compound 56



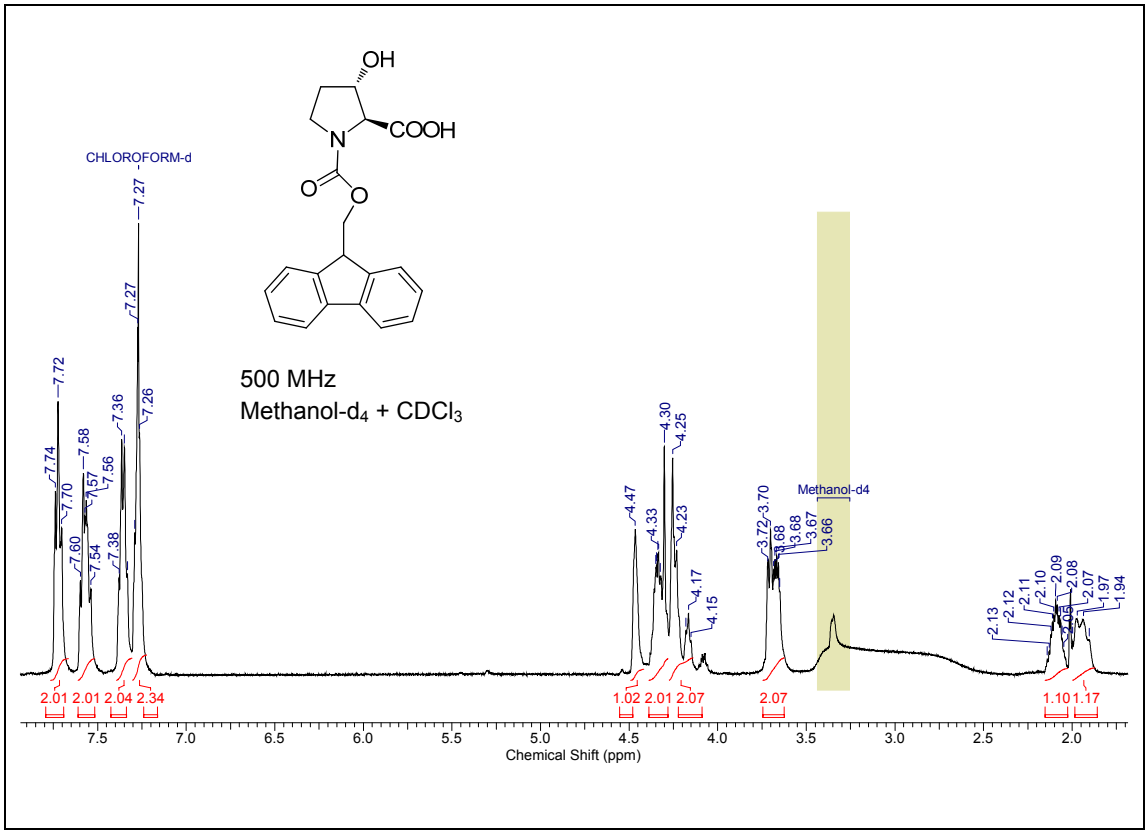
ESI-MS of compound 56



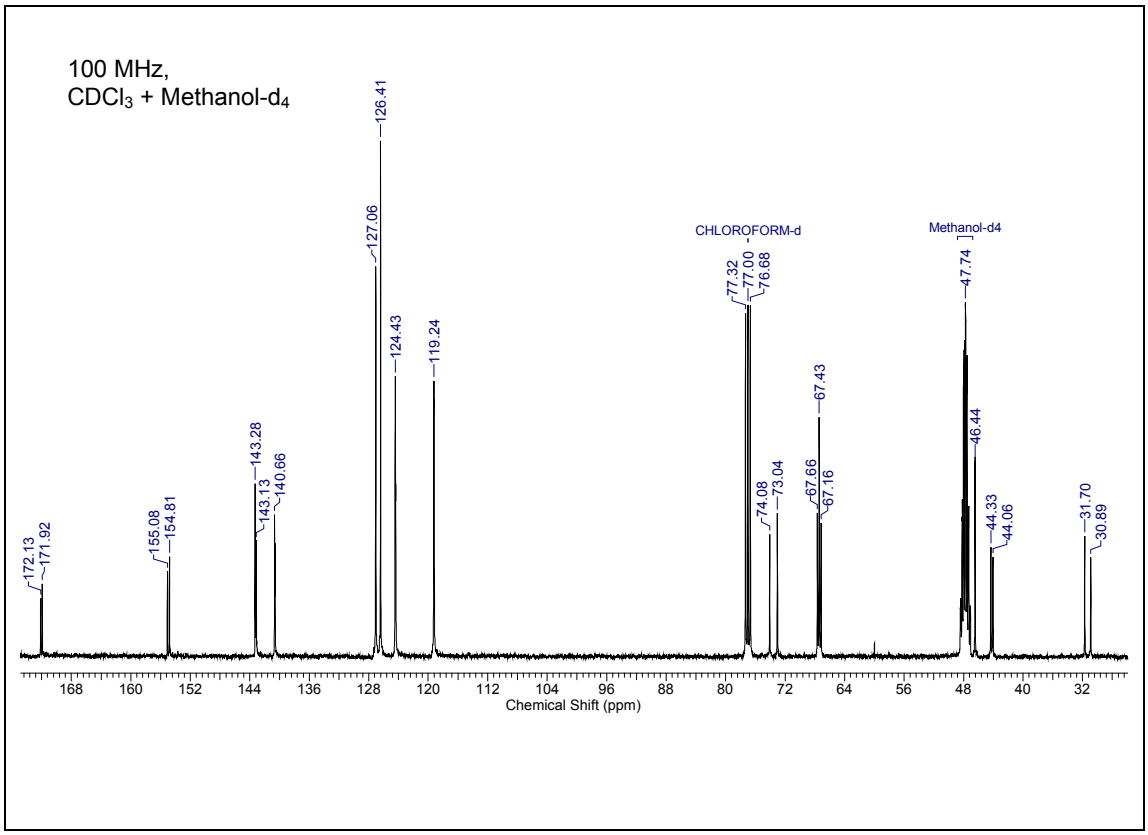
IR Spectra of compound 56



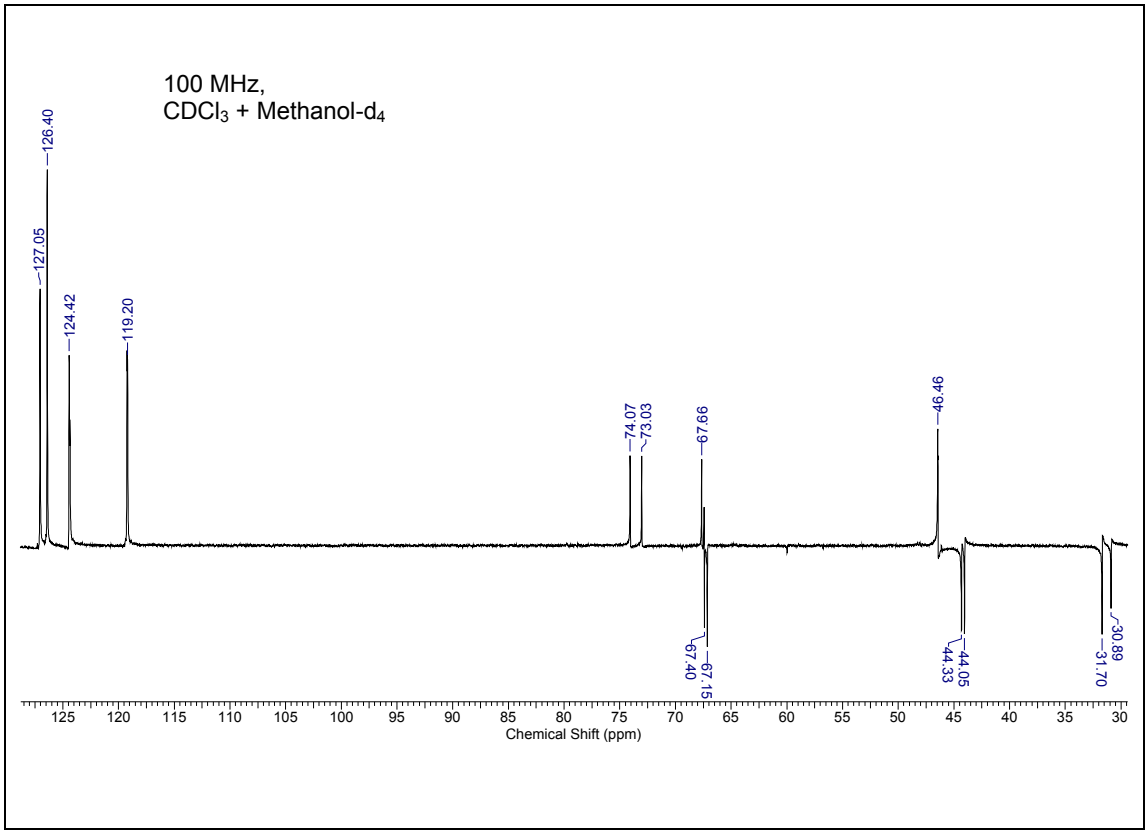
^{1}H NMR of compound 57



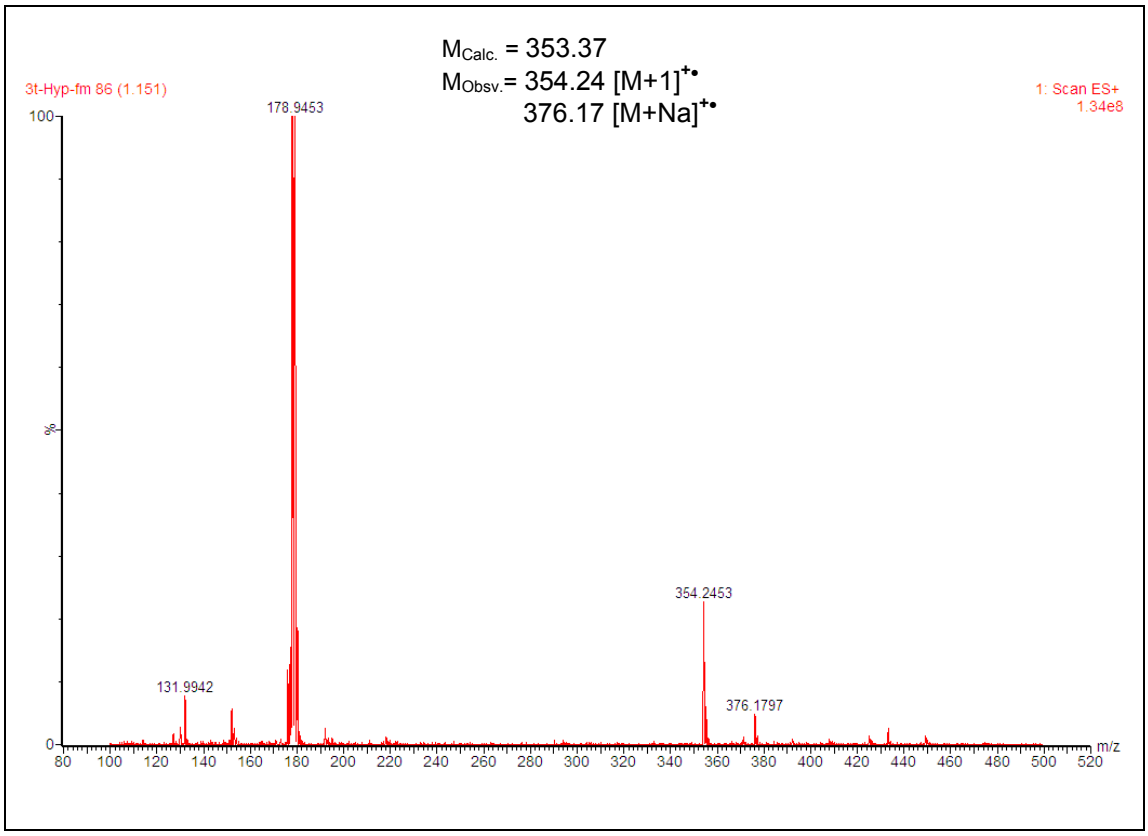
^{13}C NMR of compound 57



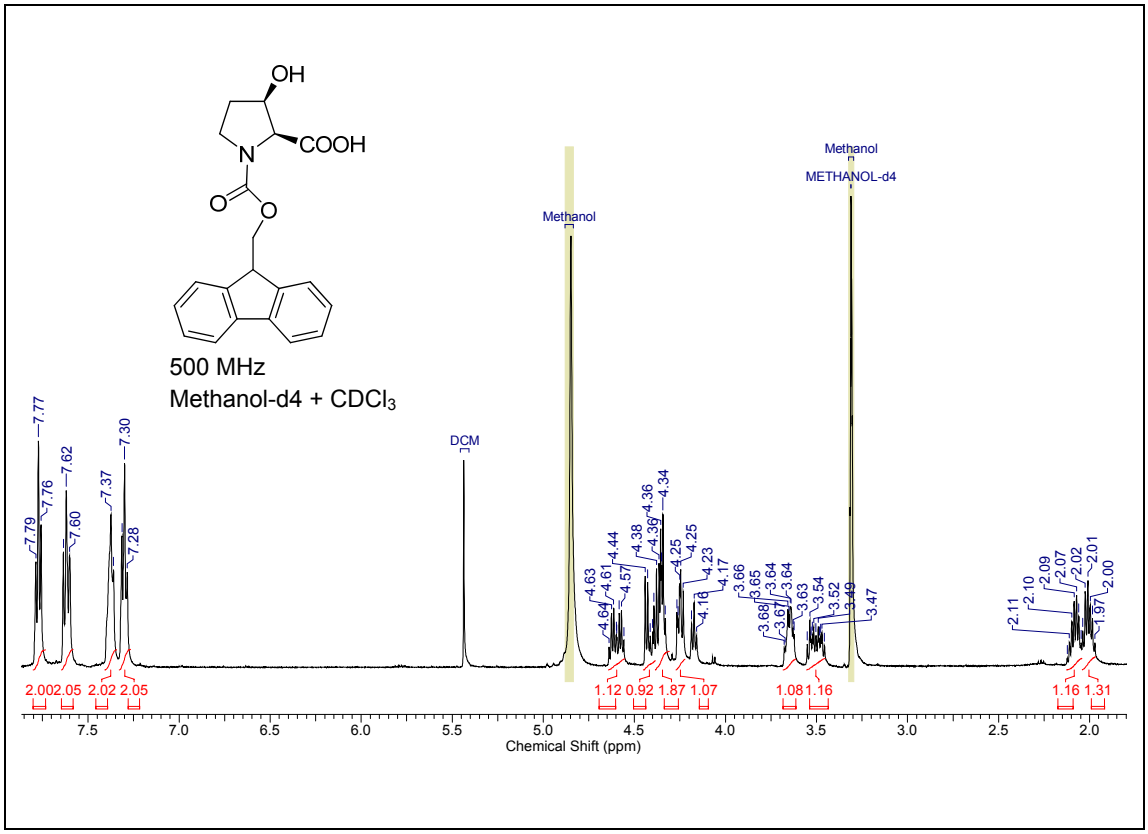
DEPT-^{13}C NMR of compound 57



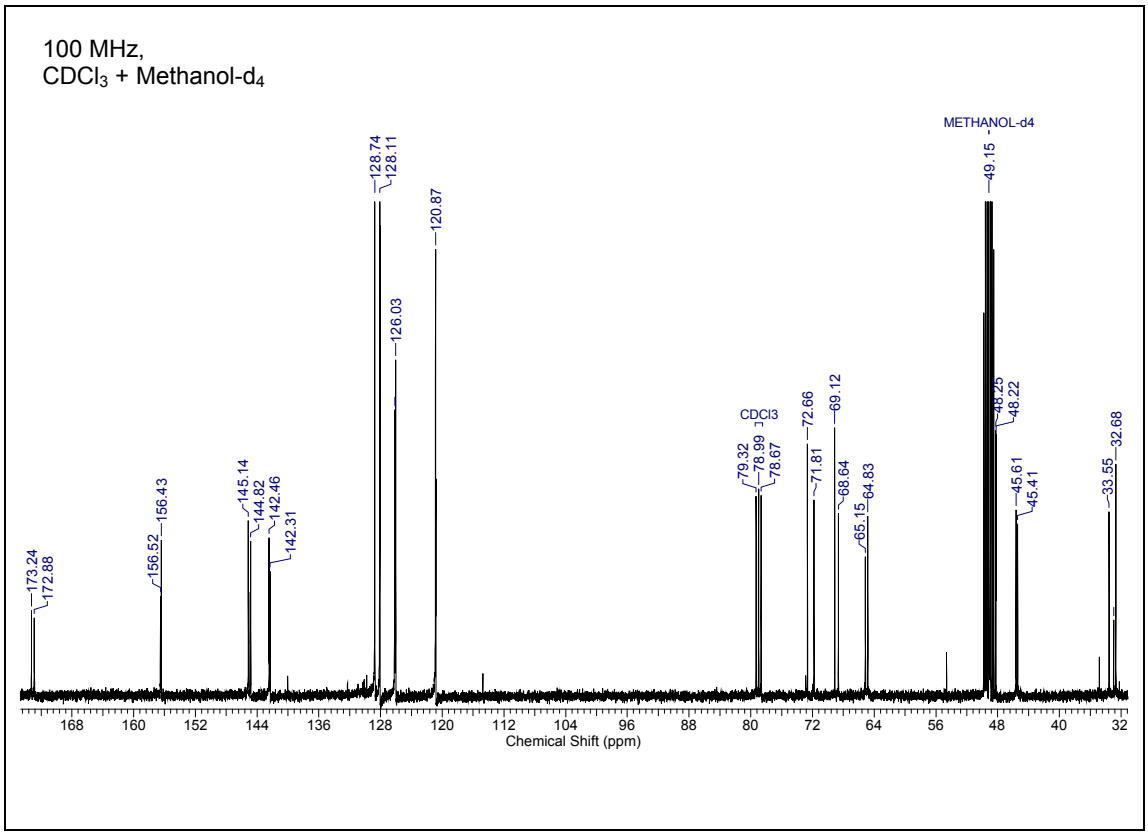
ESI-MS of compound 57



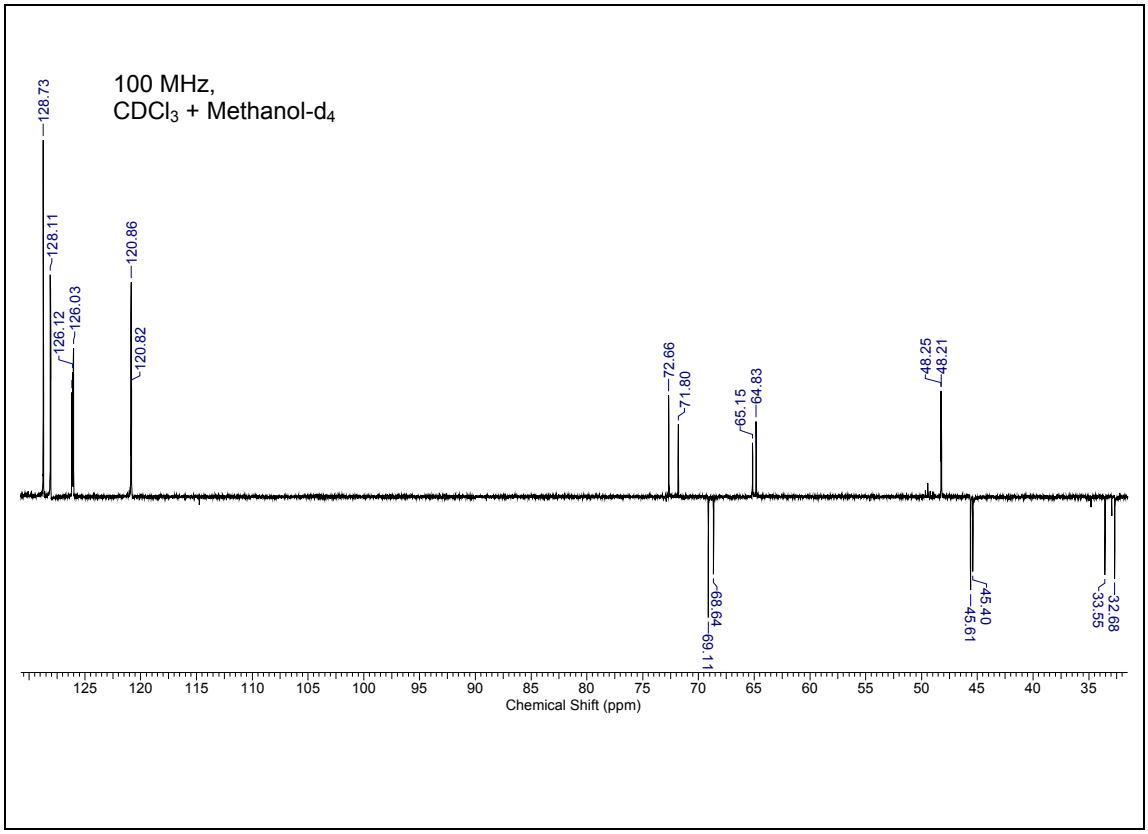
^{1}H NMR of compound 58



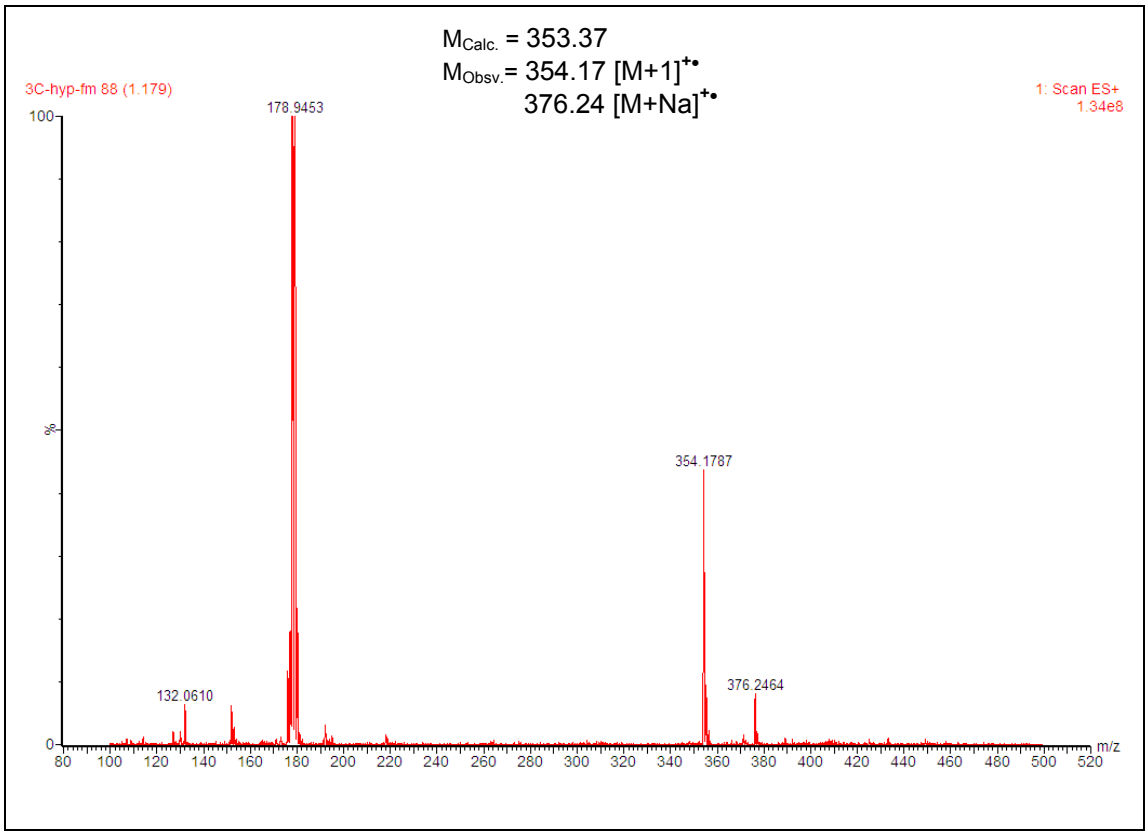
^{13}C NMR of compound 58



DEPT-^{13}C NMR of compound 58



ESI-MS of compound 58



Erratum