

Conformational Preferences and Structural Characterization of Prolyl Cis/Trans Isomerization of Carbohydrate-Templated Proline Mimetics of Some Model Peptides Using Computational Methods

by

Robel Berhe Teklebrhan

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

Master of Science

Department of Chemistry

University of Manitoba

Winnipeg, Manitoba, Canada

Copyright © 2009 by Robel Berhe Teklebrhan

THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION

Conformational Preferences and Structural Characterization of Prolyl
Cis/Trans Isomerization of Carbohydrate-Templated Proline Mimetics of
Some Model Peptides Using Computational Methods

By

Robel Berhe Teklebrhan

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of
Manitoba in partial fulfillment of the requirement of the degree
Of

Master of Science

Robel Berhe Teklebrhan©2009

Permission has been granted to the University of Manitoba Libraries to lend a copy of this thesis/practicum, to Library and Archives Canada (LAC) to lend a copy of this thesis/practicum, and to LAC's agent (UMI/ProQuest) to microfilm, sell copies and to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

Abstract

Over the years a number of proline analogues have been developed to study the structural and biological properties of proline surrogates in peptides. This is due to the fact that the prolyl *N*-terminal *cis/trans* isomerization rate and equilibrium ratios of specific proline analogues are helpful in detecting and monitoring the local structure and environment of proline. Although these analogues have proven to be useful for inducing specific constraints on prolyl *N*-terminal amide isomerization, but lacks the ability to shift the prolyl amide equilibrium into both directions. In this study a computational analysis on some novel carbohydrate-templated proline mimetics model peptides were performed, in order to determine the effect of the carbohydrate moiety incorporation onto the proline based peptides. An extensive conformational analysis of these model peptides demonstrated that the carbohydrate moiety influences the *cis/trans* ratio and kinetics of the isomerization reaction.

A detailed Density Functional Theory analysis of these model peptides in water predicts that the stability of the *cis* population of Compound 1 and Compound 2 depends on the orientation or position of the C_δ -hydroxy-methylene substituent. Results on the structural characterization also show that the intramolecular hydrogen bond influences the *cis/trans* isomerization ratio. The puckering amplitude calculated for both Compound 1 and Compound 2 shows that the position of the C_δ primary hydroxyl group greatly distorts, in particular for Compound 2, the puckering behaviour of the five-membered ring, which is a key parameter in collagen stabilization.

In addition, a reliable computational protocol was developed for the computational calculations of these model peptides and all the results produced using this protocol were in excellent agreement with the experimental data. Finally, this quantum mechanical study of the sugar

proline analogues in addition to the extensive experimental data gave us a further insight and trend into prediction of the conformational equilibria and kinetics of the carbohydrate-templated proline-based peptidomimetics.

Acknowledgements

I would like first to thank my supervisor Dr. Georg Schreckenbach, for his unconditional help, intellectual guidance, and most of all, his never ending encouragement and emotional support over the years. To me his excellence in his scientific research so far, is second only to his marvellous human qualities, patients, sense of humour and full of wisdom. I really appreciate that and the memory of the time spent with you will leave with me for years to come. I would like to thank also Dr Frank Schweizer for his intellectual discussions and guidance in this collaborative research work. I valued and admired his overall friendly style of co-supervision. It will be a great ungratefulness if I fall short to acknowledge the computational resources that I got from Dr Stacey Wetmore, at the University of Lethbridge, I would not be able to complete this thesis without the giant computational cluster available from her, URACIL one of the biggest cluster in Canada. Also, thanks to Dr James Xidos for all the computational support from day one until the end of my thesis. Many thanks also to Prof Joe O'Neil, Kaidong Zhang, Neil Owens, Lsan Tzadu, Joel Berard, Titi Woldeab and Gregory A. Shamov for their kind and humble assistance.

Finally, I would like to thank my family for being always there for me, for their encouragement to pursue my own dreams and for their unconditional love. Thank you for that and may the almighty GOD bless you always, in particular my mother. I would like also to thank Efrem and his family for his kindness and help during my stay in Winnipeg. I also thank members of the staff of Chemistry department for their support. I am indeed very grateful to you all. I am grateful also to the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the University of Manitoba start-up funds and The Research Grants Program are also gratefully acknowledged.

Table of Content

Abstract	ii
Acknowledgements	iv
Table Of Content	v
List Of Tables	vii
List Of Figures	ix
List Of Copyrighted Material For Which Permission Was Obtained	xii
List Of Abbreviations	xiii

CHAPTER 1

INTRODUCTION	1
1.1. THE PROLINE (PRO) RESIDUE	1
1.2. PROLYL CIS-TRANS ISOMERIZATION AND PROTEIN FOLDING	8
1.3. NOMENCLATURE AND PSEUDOROTATIONAL PARAMETERS OF A PROLINE RESIDUE	14
1.4. ROLE OF PROLINE AND HYDROXYPROLINE IN THE COLLAGEN TRIPLE HELIX	26
1.5. PURPOSE OF THE CURRENT RESEARCH	30

CHAPTER 2

CARBOHYDRATE-TEMPLATED PROLINE MIMETICS	32
2.1. COMPUTATIONAL STUDIES OF PROLINE BASED NOVEL MODEL PEPTIDES	32
2.2. FUSED CARBOHYDRATE-TEMPLATE PROLINE ANALOGUE – N-ACETYL- GLCProH- NHMe	38
2.3. SPIROCYCLIC CARBOHYDRATE-TEMPLATE PROLINE ANALOGUE - N-ACETYL- GLCPro-OMe	41
2.4. CONCLUSIONS	44

CHAPTER 3

COMPUTATIONAL METHODS	45
3.1. COMPUTATIONAL THEORIES AND MODELS	45
3.1.1 The Schrödinger Equation	48
3.1.1.1 The Born-Oppenheimer Approximation	49

3.1.1.2 The Hartree-Fock Approximation.....	51
3.2. MOLECULAR MECHANICS	52
3.2.1. Molecular Mechanics Formulation.....	53
3.2.2. The MMFF94 force field method.....	59
3.3. CONFORMATIONAL SEARCHING	60
3.3.1. Systematic Searching.....	61
3.3.2. Monte Carlo Searching.....	63
3.4. DENSITY FUNCTIONAL THEORY	65
3.4.1. Mathematical Formulation	65
3.4.2. Hybrid Functionals	69
3.5 SOLVATION MODEL.....	70
3.6. COMPUTATIONAL METHODOLOGY	72
CHAPTER 4	
CONFORMATIONAL PREFERENCE OF SPIROCYCLIC CARBOHYDRATE-TEMPLATE PROLINE ANALOGUES	74
4.1. RESULTS AND DISCUSSION	74
4.1.1 Conformers and energies in gas and water.....	77
4.1.2 Conformational distribution of Compound 1 and 2 in gas and water	112
4.1.3 Intramolecular hydrogen bond and ring puckering in water	115
4.2 CONCLUSIONS	121
CHAPTER 5	
CONFORMATIONAL PREFERENCE OF FUSED CARBOHYDRATE-TEMPLATE PROLINE ANALOGUES	123
5.1. RESULTS AND DISCUSSION	123
5.1.1 Conformers and energies in gas and water.....	124
5.1.2 Conformational distribution of N-acetyl-GlcProH-NHMe in gas and water	136
5.1.3 Intramolecular hydrogen bond and ring puckering in water	138
5.2 CONCLUSIONS	140
CHAPTER 6	
GENERAL CONCLUSIONS.....	141
6.1 SUMMARY AND FUTURE WORKS.....	141
BIBLIOGRAPHY.....	144

List of Tables

Table 1.1	Properties of proline: covalent structures, average occurrences in proteins, residue mass and the pK_a value of its ionizable groups.	6
Table 1.2	Approximate torsion angles for some regular secondary structures	25
Table 4.1	Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of Compound 1, which are Optimized at the B3LYP/6-31+G(d, p) Level of Theory in the Gas Phase.	81
Table 4.2	Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of Compound 1, which are Optimized at the B3LYP/6-31+G(d, p) Level of Theory in the Water.	84
Table 4.3	Selected Angles (in deg) for Compound 1, at the B3LYP/6-31+G (d, p) level of theory in the water	91
Table 4.4	Selected bond length (in Å) for Compound 1, at the B3LYP/6-31+G (d, p) level of theory in the water	92
Table 4.5	Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of Compound 2, which are Optimized at the B3LYP/6-31+G(d, p) Level of Theory in the gas phase.	98
Table 4.6	Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of Compound 2, which are Optimized at the B3LYP/6-31+G(d, p) Level of Theory in Water	101
Table 4.7	Selected Angles (in deg) for Compound 2, at the B3LYP/6-31+G (d, p) level of theory in the water	110
Table 4.8	Selected bond ^a length (in Å) for Compound 1, at the B3LYP/6-31+G (d, p) level of theory in the water	111

Table 4.9	Backbone Dihedral angles (in degree), Pseudorotational Parameters (A and P, in degree), Relative Energies (ΔE , in Kcal/mole) and dipole moment for the minima energy representative conformations for Compound 1 and Compound 2 at the B3LYP/6-31+G (d, p) level of theory in the water	120
Table 5.1	Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of <i>N-acetyl-GlcProH-NHMe</i> , which are Optimized at the B3LYP/6-31+G (d, p) Level of Theory in the Gas Phase.....	126
Table 5.2	Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of <i>N-acetyl-GlcProH-NHMe</i> , which are Optimized at the B3LYP/6-31+G (d, p) Level of Theory in the water.....	128
Table 5.3	Selected Angles (in deg) for <i>N-acetyl-GlcProH-NHMe</i> , at the B3LYP/6-31+G (d, p) level of theory in the water.....	135
Table 5.4	Selected bond ^a length (in Å) for <i>N-acetyl-GlcProH-NHMe</i> , at the B3LYP/6-31+G (d, p) level of theory in the water.....	135
Table 5.5	Backbone Dihedral angles (in degree), Pseudorotational Parameters (A and P, in degree), Relative Energies (ΔE , in Kcal/mole) and dipole moment for <i>N-acetyl-GlcProH-NHMe</i> for the minima energy representative conformations at the B3LYP/6-31+G (d, p) level of theory in the water.....	139

List of Figures

Figure 1.1	The biosynthesis of the proline amino acid residue from the L-glutamate Amino acid.....	5
Figure 1.2	The conversion of prolyl residue to 4-hydroxyprolyl residue.	8
Figure 1.3a-b	Interactions in a non-proline (a) and proline (b) peptide models in a Conformation of cis and trans.....	10
Figure 1.4	Prolyl cis-trans isomerization as a molecular backbone switch	12
Figure 1.5	Conformational Characterization of a double rotor (ϕ , ψ).	17
Figure 1.6	Conformational equilibrium of the two puckering states in proline Containing peptide residues.....	18
Figure 1.7	The conformational space of a proline residue defined by ψ , χ_1 and ω	24
Figure 1.8	a) The triple helix collagen of Gly-Pro-Hyp sequence b) a single peptide chain of the collagen triple helix c) the Gly residue in the collagen triple helix	28
Figure 2.1	The cis-trans isomerization of the N-formyl-Prolinamide	36
Figure 2.2	The cis-trans isomerization of the Ac-azPro-NHMe.....	37
Figure 2.3	The synthesis of Ac-GlcProH-NHMe	39
Figure 2.4	The cis-trans isomerization of the Ac-GlcProH-NHMe.....	41
Figure 2.5	The synthesis of -N-acetyl-GlcPro-OMe.....	42
Figure 2.6	The cis-trans isomerization of Ac-Glc3(S)-HypH-OMe, in which the hydroxymethylene substituent is pointing <i>down</i>	43
Figure 2.7	The cis-trans isomerization of Ac-Glc3(S)-HypH-OMe, in which the hydroxymethylene substituent is pointing <i>up</i>	43

Figure 3.1	The PCM model.....	71
Figure 4.1	The definitions of the backbone and endocyclic torsion angles for Compound 1 and Compound 2	76
Figure 4.2	The Minima energy representative conformations of Compound 1 in a gas phase optimized at the B3LYP/6-31+G (d, p) Level of theory.	94
Figure 4.3	The Minima energy representative conformations of Compound 1 in water optimized at the B3LYP/6-31+G (d, p) Level of theory.....	95
Figure 4.4	The Minima energy representative conformations of Compound 2 in a gas phase optimized at the B3LYP/6-31+G (d, p) Level of theory.	107
Figure 4.5	The Minima energy representative conformations of Compound 2 in water optimized at the B3LYP/6-31+G (d, p) Level of theory.....	108
Figure 4.6	The cis and trans population distribution of Compound 1 & 2, each population was computed using the Boltzmann weight by the relative energy at the B3LYP/6-31+G (d, p) Level.	112
Figure 4.7	The backbone and endocyclic population distribution of Compound 1 and 2, each population was computed using the Boltzmann weight by the relative energy at the B3LYP/6-31+G (d, p) Level.	114
Figure 4.8	The three hydrogen bonds that exists in water for Compound 1	117
Figure 4.9	The two hydrogen bonds that exists in water for Compound 2.....	119
Figure 5.1	The Minima energy representative conformations of N-acetyl-GlcProH-NHMe in a gas phase optimized at the B3LYP/6-31+G (d, p) Level of theory.....	132
Figure 5.2	The Minima energy representative conformations of N-acetyl-GlcProH-NHMe In water optimized at the B3LYP/6-31+G (d, p) Level of theory	133
Figure 5.3	The cis and trans population distribution of N-acetyl-GlcProH-NHMe, each population was computed using the Boltzmann weight by the relative energy at the B3LYP/6-31+G (d, p) Level.	136

Figure 5.4	The backbone and endocyclic population distribution of N- <i>acetyl-GlcProH-NHMe</i> , each population was computed using the Boltzmann weight by the relative energy at the B3LYP/6-31+G (d, p) Level.	137
Figure 5.5	The hydrogen bond found in N- <i>acetyl-GlcProH-NHMe</i>	139

List of Copyrighted Material for which Permission was

Obtained

- Figure 1.8** a) The triple helix collagen of Gly-Pro-Hyp sequence b) a single peptide chain of the collagen triple helix c) the Gly residue in the collagen triple helix, Copyright permission obtained from Karl M. Oberholser, Messiah College, on June 02, 2009..... 28
- Figure 3.1** The PCM model, Copyright permission was obtained on May 27, 2009 from Georg Schreckenbach (GS private communications)..... 70

List of abbreviations

A	Puckering Amplitude
a.u.	Atomic units
ATP	Adenosine-5'-triphosphate
azPro	Azaprolines
B3LYP	non-local Becke exact exchange combined with the Lee, Yang and Parr correlation functional (hybrid method)
Boc	tert-butyloxycarbamate
CCSD(T)	coupled cluster singles, doubles and partial triples (method)
COSMO	conductor-like screening model (solvation)
CPCM	conductor polarized continuum model (solvation)
CsA	Cyclosporin A
CD	Cyclodextrin
Cyps	Cyclophilins
C _α	Alpha carbon
C ^γ -endo	the down puckering
C ^γ -exo	the up puckering
d	down puckering
DNA	Deoxyribonucleic acid
DFT	Density Functional Theory
ECM	Extracellular matrix
EVH1	Enabled/VASP homology 1
FKBPs	FK506-binding proteins
g03	Gaussian 2003 (software)
GGA	generalized gradient approximation (functional method)

Glc3(S)HypHs	spirocyclic glucose-3(S)-hydroxyproline hybrids
GlcProH	glucose proline hybrid
Gly	glycine
GYF	Glycine-tyrosine-phenylalanine
HF	the Hartree-Fock
Hyp	4-hydroxylproline
K-S	Kohn-Sham
LDA	Local density approximation
MC	Monte Carlo
MDCA	Multidimensional conformational analysis
MM	Molecular Mechanics
MMFF94	Molecular Mechanics Force Field
N	rotatable bonds
NADPH	Nicotinamide adenine dinucleotide phosphate
NRM	Nuclear Magnetic Resonance
P	the puckering phase angle
PCM	polarized continuum model
PEC	Potential energy curve
PEHS	Potential Energy Hypersurface
PES	Potential Energy Surface
PPI	Polyproline I
PPIases	Peptidyl prolyl cis-trans isomerases
PPII	Polyproline II
Pro	Proline
SCF	self-consistent field
SCRf	self-consistent reaction field
SH3	Src homology 3 domain

TBTU	O-benzotriazolyl-N, N, N', N'-tetramethyluronium tetrafluoroborate
u	up puckering
VDW	van der Waals
WW	Protein domain involved in binding to proline-rich peptide motifs
XC	exchange-correlation (potential)
X-Pro	The Proline residue
Xxx	any amino acid residue

Chapter 1

Introduction

1.1. The Proline (Pro) Residue

Although hundreds of different types of amino acid residues have been described in nature, only *twenty* (e.g. alanine, lysine and proline etc.) are commonly found in the proteins or peptides of most living organisms [1]. These common amino acids are called proteinogenic or standard amino acids, and are defined as those encoded by the standard genetic code [1, 2]. The transcription and translation of this genetic code of the Deoxyribonucleic acid (DNA) results in the formation of a specific linear L- α -amino acid residue sequence which is characteristic of proteins and peptides [1, 2, 3].

Proteins and peptides may also contain derived amino acid residues [2, 3]. These derived amino acid residues are usually formed by the posttranslational modification of the common amino acid residues in proteins, which is the modification of protein after translation, and are often important for the function of the proteins [3, 4]. One of the most prominent amino acid residue derivatives is 4-hydroxyproline (Hyp), which is found in the fibrous protein collagen. It results from the hydroxylation of the proline amino acid residue by the prolyl hydroxylase enzyme. Another interesting derivative amino acid residue is γ -Carboxyglutamic acid, which is one of the most important constituents of numerous proteins involved in blood clotting [3, 4]. Nevertheless, among all the

proteinogenic or derived amino acid residues that occur in proteins or peptides, proline (or its derivative hydroxyproline) exhibits very *distinctive* structural properties compared to most other amino acid residues [2, 3].

Proline (Pro) is unique in the realm of amino acids in that it incorporates the α -nitrogen atom of the peptide backbone in its side-chain via a covalent bond [2]. The incorporation of the α -nitrogen atom in its side-chain or the formation of a five-membered pyrrolidine ring around the peptide backbone has a significant structural consequence on the properties of the residue [5, 6]. In particular [5, 6], (i) the cyclic nature of the five-membered pyrrolidine ring of the proline restricts the conformational freedom of the proline residue about the rotation of the $-N_{\alpha}-C_{\alpha}-$ bond to a very confined value of dihedral angle $\sim 60^\circ$; (ii) it is found that there exists a firm dependence between the main-chain proline conformation and its side-chain conformation; (iii) contrary to all other proteinogenic amino acids which favor almost exclusively the *trans* structure in polypeptides, proline can exist in both *cis* ($\sim 5\%$) and *trans* ($\sim 95\%$) configurations in peptides or proteins; (iv) the lack of an amide hydrogen atom on the α -amino group prevents the proline backbone from forming a hydrogen bond with the preceding part of the peptide structure, as a result proline does not participate in forming a hydrogen bond in stabilizing an α -helix or a β -sheet structure. Consequently, proline always acts as a structural disrupter at the center of a regular secondary structure of α -helix and β -sheets [2-6].

Owing to these intrinsic structural properties, proline plays a determinant role in dictating and directing the secondary structures of proteins. It thereby strongly induces special motifs like reverse turns and bends that embody the proline residue [7, 8]. Indeed,

proline's specific positional preference at the beginning of the α -helices and generating kinked α -helical structures are also a direct result of the conformational rigidity of the five-membered pyrrolidine ring of proline [9]. As a consequence of these unique structural properties, proline is frequently found in turns, non-repetitive structures, and at the end of strands and α -helices of proteins [10].

Several structural studies indicated that the proline side-chain interaction also stabilizes the structures of β -turns [9, 11], and γ -turns [9, 12, 13]. The proline residue also plays an important role in the cell, like in receptor oligomerization [9, 14], receptor activation, and ligand binding [9, 15, 16], and molecular recognition specifically in intracellular signaling complexes and pathways [17]. Some modular recognition domains that bind to proline-rich motifs such as the Src homology 3 domain (SH3 domain), WW, Enabled/VASPI or EVH1 and Glycine-tyrosine-phenylalanine or GYF are used to mediate protein-protein interactions of signal transduction components by recognizing proline-containing peptide sequences [17].

Another fascinating role of proline is that it lowers the non-specific proteolytic degradations in peptide chains [9, 16]. A number of studies also indicated that the presence of proline (plus hydroxyproline) residues has an enormous influence not only on the conformation of the collagen strands, but also on the thermal stability of the collagen strands in fibrous protein collagen [7, 13].

With an average of 5% occurrences in proteins [1, 4], proline is one of the most important DNA-encoded amino acid residues in collagen [1, 2], a major structural component of cells and animals. Its codons are CCU, CCC, CCA, and CCG [4]. Unlike the other amino

acid residues in which the amino group is a primary, proline contains a secondary amino acid group [6]. It is a non-essential amino acid in that, unlike some other amino acids (e.g. Lysine, Valine etc.), humans can easily synthesize it with no any external supply. Generally, proline amino acid residues are biosynthetically derived, in both mammals and bacteria, from the conversion of the L-glutamate amino acid through several steps. In mammals, proline can also be synthesized via a different route, which is through the urea cycle [4].

Figure 1.1 shows a schematic representation of the proline biosynthesis route from L-glutamate amino acid. In this biosynthesis route, first the γ -carboxylate group of glutamate is activated by the Adenosine-5'-triphosphate (ATP)-driven phosphorylation to form an unstable product of γ - glutamylphosphate. Then, this unstable product is in a mixed anhydride reaction and presumed to be the substrate for the reduction that follows the next step. The reduction of the glutamate-5-phosphate with Nicotinamide adenine dinucleotide phosphate (NADPH) results in releasing the phosphate and producing glutamate-5-semialdehyde. The glutamate-5-semialdehyde is then spontaneously cyclized to produce internal Schiff base Δ^1 -pyrroline-5-carboxylate in a nonenzymatic reaction by eliminating H_2O . Finally, the reduction to proline is catalyzed by pyrroline-5-carboxylate reductase. It is also important to note that during the reduction of the Δ^1 -pyrroline-5-carboxylate to proline it is not clear whether the enzyme requires NADH or NADPH [4].

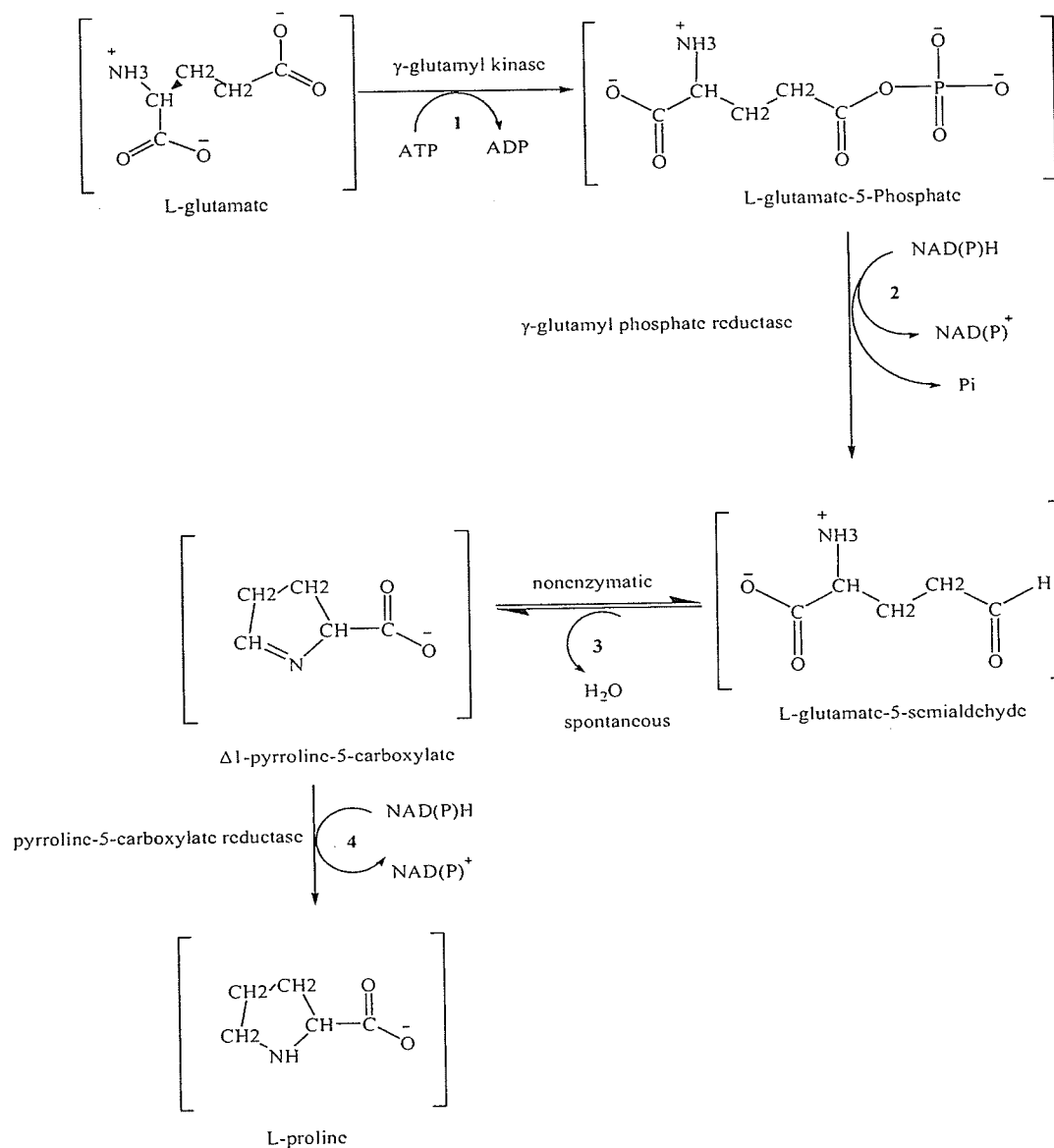
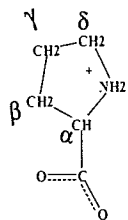


Figure 1.1 The biosynthesis of the proline amino acid residue from the L-glutamate amino acid. The different enzymes that are used in catalyzing proline biosynthesis at each stage are the following (1) γ -glutamyl kinase, (2) dehydrogenase, (3) nonenzymatic, and (4) pyrroline-5-carboxylate reductase. This figure has been drawn following reference, Voet, D.; Voet, G. J. 2004 [4].

Table 1.1 indicates that the pK_a values for the α -carboxylic acid group and the α -amino group of proline are 1.95 and 10.64 respectively, so above pH 3.5 and below pH 8 both these groups are in their ionic forms, which is in a carboxylate and ammonium ion respectively. This implies that in a normal physiological pH range both the α -carboxylic acid and α -amino group are completely ionized [4].

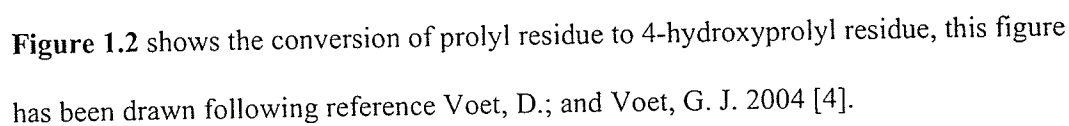
Table 1.1 Properties of proline: covalent structures, average occurrences in proteins, residue mass and the pK_a value of its ionizable groups.

Amino Acid residue	Structural Formula ^a	Residue Mass (D) ^b	Average Occurrence in proteins (%) ^c	pK_{a1} α -COOH ^d	pK_{a2} α -NH ^d	Side Chain	Three letter code
Proline		97.1	5.0	1.95	10.64	Non-polar	Pro
^a The structure of proline at pH 7.0 which is in an ionic form ^b The residue mass is for neutral proline residue ^c The average percent amino acid composition in the complete SWISS-PROT database ^d The pK_{a1} and pK_{a2} refers to the α -carboxylic acid and α -amino groups The following table was summarized and drawn following reference Voet, D.; and Voet, G. J., 2004 [4]							

One of the most essential contributions of L-proline is its connection to collagen. L-Proline (Pro) and (4R)-hydroxy-L-Proline are abundant amino acids in fibrous collagen proteins, and are used for strengthening the connective tissues, and refurbishing damage to tissue, skin and muscles [18]. It also improves skin texture and elasticity by reducing the loss of collagen through the aging process [2, 4].

Studies have also shown that improper production of collagen can create a great problem in the maintenance and healing of cartilage and the strengthening of joints, tendons, bones, lenses of the eye and cardiac muscles [4, 18]. Collagen in the skin contains a hydroxyproline, and this is the posttranslational amino acid formed from proline amino acid residue which is useful in the stability of collagen.

The 4-hydroxyprolyl (Hyp) (Figure 1.2) is formed or synthesized by the stereospecific posttranslational hydroxylation of the proline residue by the enzyme prolyl-4-hydroxylase [2, 4]. During this biosynthesis the ascorbic acid or Vitamin C plays an important role in activating the enzymatic capability of the prolyl hydroxylase; failing to do so results in losing its native conformation or denature. That is the main reason why deficiency of vitamin C in our body results in the formation of less fiber collagen or skin lesions, blood vessel fragility and poor wound healing, which are all symptoms of a disease called Scurvy. Abnormalities in collagen structure also leads to diseases such as osteogenesis imperfecta, Ehlers–Danlos syndrome, osteoporosis and arthritis [2, 4, 18].



In almost all proteins and peptides, the peptide bond of the amino acid residue is predominantly favoring the trans conformation [19]. However, analyses of a nonredundant set of 571 X-ray protein structures shows that proline residues have a higher percentage of cis conformation peptide bonds than any other amino acids [19]. Conformational analysis of these peptide bonds indicated that the cis population value ranges from ~ 0.03-0.04% for the non-prolyl containing structures, and ~5-6% for the prolyl containing peptide bonds [19, 20].

8

proline residue or X-Pro containing peptides due to proline's unique structural properties (Figure 1.3a-b).

In proline-containing peptides, the interaction between the two adjacent α -CH groups for both cis and trans conformers are almost similar, the $O_i \cdots C'_{i+1}$ electrostatic interactions also are greatly changed compared to that of non-prolyl containing peptides and finally there is only a small entropy loss in converting trans to cis conformation for the prolyl containing peptides relative to the non-prolyl peptides [21, 22, 23, 24].

Considerable theoretical [20, 24, 25, 26] and experimental [20, 23, 27] works indicated that the free energy differences between the two isomers (cis and trans) of proline containing peptide groups is very small or almost isoenergetic, with a slight energetical unfavorability in the cis conformer [21].

Small proline-containing peptides can be found in combinations of both cis and trans conformations, with an energy barrier of ~ 19 -20 Kcal/mole between them [22, 20]. This high-energy barrier is due to the partial double bond character of the peptide bond and this restricts the rate of interconversion between the cis and trans conformers [22, 28]. The rate of interconversion between cis and trans is also influenced by the polarity of the solvent, being higher in a non-polar solvent and lower in polar solvent [25].

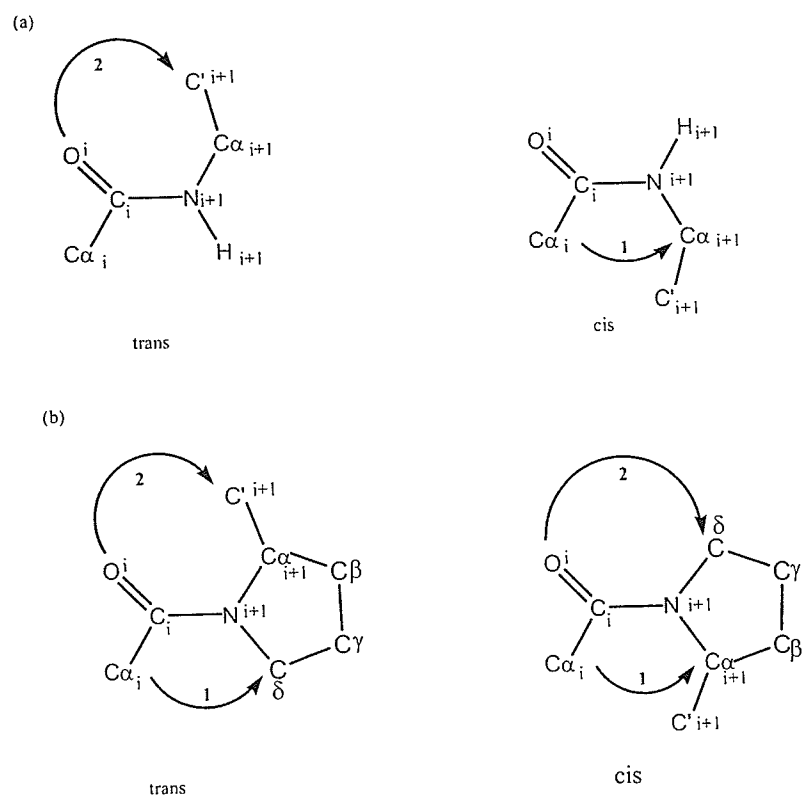


Figure 1.3a-b Interactions in a non-proline (a) and proline (b) peptide models in a conformation of cis and trans. (1) steric interaction between two adjacent Carbons (2) the electrostatic interaction between oxygen and carbon. This figure has been drawn following reference Wedemeyer, J. W. et al [20]

It is also reported that in native proteins a completely cis or a completely trans preference of proline containing peptide is observed and this is due to the favorable interactions with its surrounding groups [20]. The properties of proline emerge in various proteins, like oligopeptides [4], globular proteins [4] and even in polyproline I (PPI) and polyproline II (PPII), which are an all cis in the right-handed helical structure and an all trans in a left-handed helical structure respectively [23]. PPI can be easily converted to PPII or vice versa by simply changing the solvent [20, 23]. A large percentage of the observed cis

peptide bonds are found in surface accessible bend, coil or turn conformations, which are always solvent exposed [29].

Schmid and co-workers (1993) [25, 28], suggested that the prolyl cis-trans isomerizations of the proline containing peptide bonds are often participating in the rate-determining steps of the protein folding and refolding of a number of proteins. The prolyl cis and trans isomerization frequently affects also the heterogeneity of the unfolded states of the proteins, which directs the folding and refolding to multiple pathways [25]. Taking advantage of the importance of the structural and kinetic characterization, the cis-trans isomerization of proline can be exploited in two ways. (1) At 0-35 °C temperatures, the conformational folding of the proline containing peptides or proteins is much faster than the rate of proline cis-trans isomerization. When the circumstance of kinetic decoupling holds, the conformational folding of the different state species can be measured without difficulty, because the isomeric form of the two states are virtually fixed [20]. This condition is helpful in the prediction of the effect of the localized disruptions on the folding kinetics. That is, it gives some sort of indication of the role of the interactions within nearby chain segments during the conformational folding [20]. (2) Another important advantage of the cis-trans isomerization of a proline containing peptide is that it helps probe its local environment. The change in the cis to trans ratio may signal the development of local structures on the backbone of the protein residue. This is particularly important when the protein is partially folded or a portion of the protein has attained a stable structure; such a structure might not be identified by less localized probes such as circular dichroism absorbance of tyrosine [20].

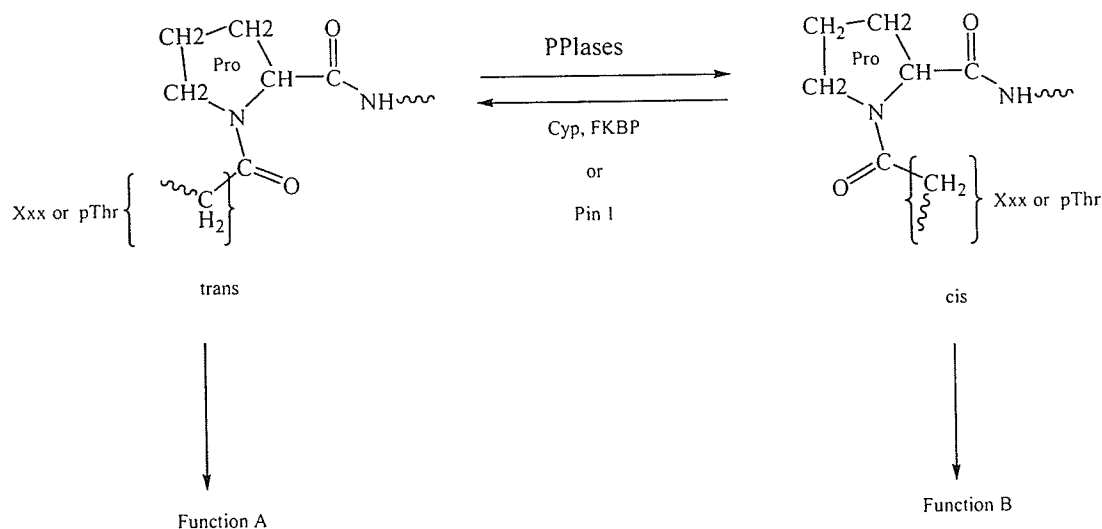


Figure 1.4 Prolyl cis-trans isomerization as a molecular backbone switch, Xxx refers to any amino acid residue. This figure has been drawn following reference Lu et al, 2007 [30]

Interestingly, the structural variation among the *cis* and *trans* conformers constitutes an essential molecular switch that can interconvert between two functional states of the protein or differentiate between two different positions of intermolecular binding partners (Figure 1.4) [30].

Several studies have suggested that this intrinsic but slow conformational switch or interconversion can be catalyzed by peptidyl prolyl cis-trans isomerases (PPIases) or rotamases. This *cis*-*trans* interconversion is catalyzed by a mechanism that apparently involves the disruption of the partial double-bond character of the peptide bond [20, 28, 31]. PPIases function as an accelerating agent, speeding up the isomerization rate significantly, and denature the proteins [30, 20, and 25]. Generally, PPIases can be

classified into four classes: Cyclophilins (Cyphs), FK506-binding proteins (FKBPs), parvulins and Ser/Thr phosphatase 2A (PP2A) activator PTPA [30, 32-34]. Cyphs and FKBPs are cellular targets for cyclosporin A (CsA) and FK506 (tacrolimus), respectively, which are potent immunosuppressive drugs commonly used in reducing the incidence and severity of allograft rejection after organ transplantation [35]. FKBPs are also targets for the immunosuppressant and anticancer agent rapamycin [30, 32-34].

The discovery of the PPIase Pin1 reveals the role of the prolyl cis-trans isomerization on cellular processes [30]. Contrary to all identified PPIases, Pin1 binds to and isomerizes specific phosphorylated Ser/Thr/-Pro motifs in certain proteins [30, 36]. In particular, Pin1-catalyzed prolyl-isomerization has an impact on a number of proteins in diverse cellular processes, like in cell growth regulation, genotoxic and cellular stress response, immune response, germ cell development and neuronal differentiation and survival [30, 37, 38]. It is very important to emphasize that, besides its wide range of physiological roles, the deregulation of Pin1 plays a critical role in cancer, Alzheimer's disease, aging, asthma and microbial infections, which signifies it as a therapeutic target [30, 37, 38].

The prolyl cis-trans isomerization has also been implicated to be involved in a number of biochemical roles, including the multiple elastic conformations of cardiac PEVK, the chemical-mechanical coupling within actomyosin, and as a switch for a neurotransmitter-gated ion channel. For these reasons the cis-trans isomerization is of paramount importance in a number of biological activities.

1.3. Nomenclature and Pseudorotational Parameters of a Proline residue

The torsion angle and structural parameters for the prolyl residue are defined in Figure 1.5c. Assuming that the backbone conformation of any amino acid residue in peptides or proteins is defined absolutely as a function of the three torsion angles, Φ , ψ and ω' (Figure 1.5b), which are characteristic of the rotation about the N—C $^\alpha$, C $^\alpha$ —C=O and O=C—N, respectively, the energy related to the backbone conformational change is given by the potential energy hypersurface (PEHS) as: [9, 39, 40, 41, 42, 43, 44].

$$E = f(\phi, \psi, \omega') \quad (1.1)$$

Fejer and co-workers [9, 45] pointed out that the $\chi_1(C_\alpha - C_\beta)$ torsional angle generally determines the orientation of the peptide side chain (Figure 1.5b, c), and suggested that ϕ , ψ , and χ_1 may be related significantly because these three torsional angles establish the internal rotation of the functional groups located on the alpha carbon (C_α) in the peptide backbone conformation. This leads to the conclusion that in *all cis or trans* peptide bonds ($\omega = \omega' = 0^\circ$ or 180°), the PEHS of the amino acid residue is described as: [9]

$$E = f(\phi, \psi, \chi_1) \quad (1.2)$$

This is the general description of any amino acid residue with an exception to glycine, in which the side chain group is a hydrogen ($R=H$), where the energy for glycine is given as a function of ϕ and ψ only [9].

$$E = f(\phi, \psi) \quad (1.3)$$

Nevertheless, the situation is quite different with the proline amino acid residue because the occurrence of the five-membered pyrrolidine ring restricts ϕ in the range of -60° (gauche⁻ or g⁻). As a result, ϕ is then considered to be a static parameter and the potential energy surface is given by, like a Ramachandran PES type, the following description [9]:

$$E = f(\psi, \chi_1) \quad (1.4)$$

This implies that not only ψ is important in the description of the PES of proline but also the puckering of the five-membered ring [9]. However, if we follow the traditional PEHS, the energy for proline is given by the potential energy curve (PEC) of equation 1.5 [9]. But this description does not fully describe the potential energy surface for proline, as the side chain has a greater impact on the stability of the conformation.

$$E = f(\psi) \quad (1.5)$$

Several theoretical calculations revealed that for each peptide residue, only a restricted number of distinct ψ , ϕ backbone torsional angle combinations exist [39, 40, 41]. With the exception of proline, nine conformational minima were predicted to exist, as can be seen in Figure 1.5b using a multidimensional conformational analysis (MDCA) [39, 41]. These nine different backbone conformations found on the potential energy surface of $E = f(\phi, \psi)$ of the most common amino acids are: γ_D , δ_D , α_L , ϵ_D , β_L , ϵ_L , α_D , δ_L , and γ_L . However, not all these nine valid conformers are energy minima on the PES of the Ramachandran map of most peptides and proteins. Conformers like α_L and ϵ_L often do not exist at all on the Ramachandran map [40, 41, 42, 43, 46, 47].

Figure 1.5 depicts the conformational characteristics of a double rotor following the recent systematic nomenclature of peptides [41, 43]. Due to the conformational restrictions imposed by its pyrrolidine ring, proline will only allow one ϕ value in the vicinity of the g- (i.e. -60° or 300°). This results in generating only three low energy conformers for each of cis and trans isomers with three different value of ψ of the proline residue (Figure 1.5c). The ψ value lies in the locality of gauche⁻ (g⁻); (i.e. $\psi = -60^\circ$), anti (a); (i.e. $\psi = 180^\circ$), and gauche⁺ (g⁺); (i.e. $\psi = 60^\circ$).

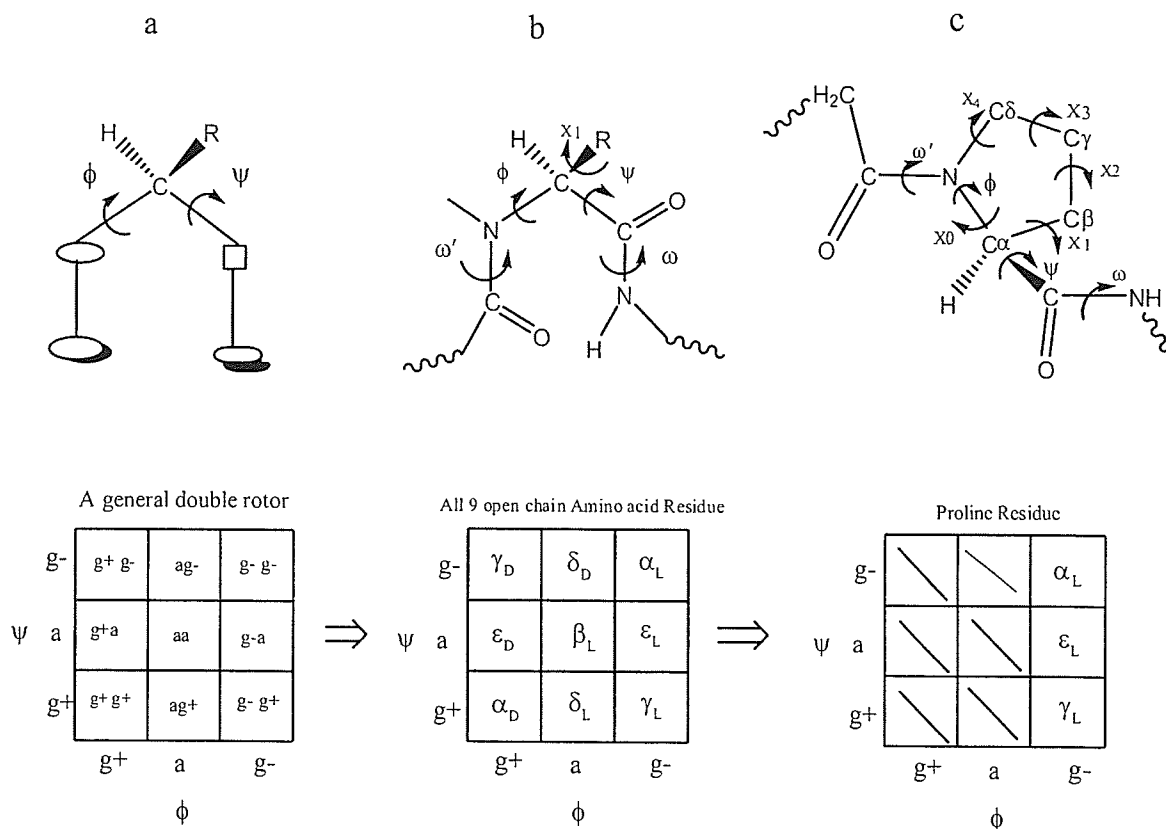


Figure 1.5 Conformational Characterization of a double rotor (ϕ , ψ). a) Conformational PES of a general double rotor, b) Conformational PES of an amino acid diamide c) Conformational PES for a proline diamide. This figure has been drawn following references Enriz et al. [44], Sahai et al et al [9] and Perczel et al [41].

These three backbone conformations are α_L , ϵ_L and γ_L [41]. These are (1) the (g^- , g^+) conformers, which corresponds to a reverse γ turn (γ_L), (2) the (g^- , g^-) conformer, which can lead to right-handed 3_{10} or α -helices in polypeptides (α_L), (3) the (g^- , a) conformer, which can lead to polyproline II (P_{II}) (ϵ_L) (Figure 1.5c). The trans γ_L conformer is the only one that allows the formation of an intermolecular H-bond in a dipeptide analogue. This leads to a pseudo 7-membered ring and is referred to as C_7 [10, 41].

The prolyl residue has a five-membered pyrrolidine ring (or side chain), which may adopt two distinct conformational states that are almost equally favorable. These two states are the “UP” sometimes called *syn*, and the “DOWN”, sometimes called *anti*. The “DOWN” and “UP” puckering of the five membered ring is defined as those of which the C_γ atom and the C=O group of the prolyl residue lie on the same and opposite sides, respectively, of the plane defined by the three atoms C_δ , N and C_α which are represented by “d” and “u”, respectively [10, 41]. In particular, the down ring puckering is characterized by positive values of χ_1 and χ_3 , and negative values of χ_2 and χ_4 , and vice versa for the up ring puckering, which is χ_1 and χ_3 are negative and positive χ_2 and χ_4 (Figure 1.6)

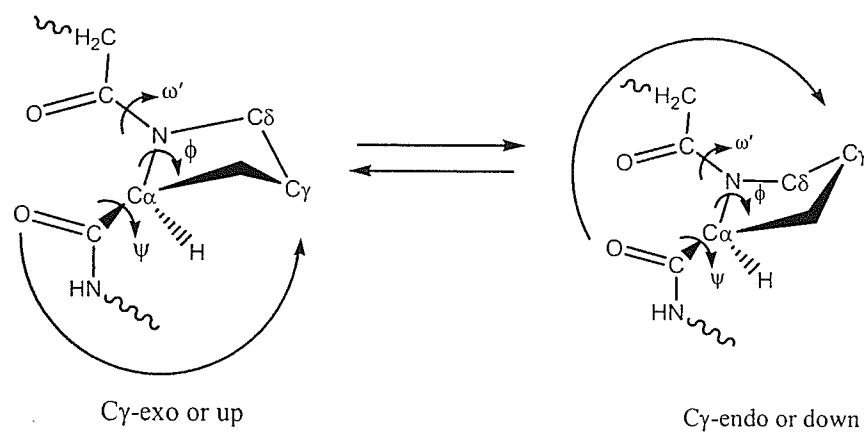


Figure 1.6 Conformational equilibrium of the two puckering states in proline containing peptide residues. This figure has been drawn following reference Enriz et al. [44], Sahai et al et al [9], Kang et al [10] and Perczel et al [41].

The puckering nature of the conformation of the five-membered ring in proline was measured through the concept of the classic pseudorotational algorithm [48, 49]. It is a

very useful method to explain the puckering amplitude and its phase angle for a non-planar proline ring [48, 49]. In this pseudorotation movement the five-membered ring can change its conformational structure to envelopes and twists along its path, which have some certain traits of a rotation [52]. Since the introduction of this concept in 1947, a number of experimental and theoretical [7, 50, 51, 52] studies have been done on the validity of this concept.

Kilpatrick et al. [49] were the first to introduce the concept of pseudorotation as an explanation for the high gas phase entropy of Cyclopentane. Because the five-membered ring is so flexible, unlike the aromatic rings which are planar, the maximum puckering rotates around the ring adopting the envelope, twist and half-twist structures, with a slight change to energy barriers. In their work they showed, the correlation of the displacement of the five carbon atoms perpendicular to the main plane of the ring using only two parameters.

The displacement of the j^{th} (Z_j) carbon atom perpendicular to the plane is given by:

$$Z_j = \sqrt{\frac{2}{5}} q \cos 2\left(\psi + \frac{2\pi}{5}(j-1)\right) \quad j = 1 \dots 5 \quad (1.6)$$

where q is a puckering amplitude and ψ is the phase angle that describes various types of puckering. This model is generally accepted with the exception that the application to a five-membered ring with unequal bond lengths and angles is not simple [53]. Cremer and Pople [53] further improved and refined the concept and made some corrections to the

original definition given by Kilpatrick. They described the puckering of Cyclopentane by pseudorotation on the basis of the displacement of each atom perpendicular to the average plane of the ring. Moreover, an attempt to define a generalized set of puckering coordinates was made by Geise et al. [54] which provides a description of puckering in five-membered rings that involves the torsion angles θ_j rather than a displacement perpendicular to some plane and avoids defining an average plane. They proposed a relationship of the form:

$$\theta_j = \theta_m \cos\left(P + \frac{4\pi(j-1)}{5}\right) \quad j = 1, \dots, 5 \quad (1.7)$$

where θ_m is the amplitude and P is a phase angle. The Phase angle is obtained in this expression using the following relationship:

$$\tan P = \frac{(\theta_3 + \theta_5 - \theta_2 - \theta_4)}{2\theta_1 \left(\sin \frac{1}{5\pi} + \sin \frac{2}{5\pi}\right)} \quad (1.8)$$

This method is very useful and can be applied directly to any five-membered ring given only the torsional angles, but the main drawback with this method is that it is very difficult to express the full set of torsion angles in terms of only two parameters [53]. This leads to a puckering amplitude that is changing its value depending on which atom

is chosen in equation 1.7. However, the general approximation of this method is good provided that the puckering is not very large [53].

Following the use of endocyclic torsion angles, which avoids the use of an average plane, Dunitz [9, 55] has shown that, for infinitesimal displacements of a pentagon from planarity, there is a direct linear relationship between the torsion angles and displacements. Hence the amplitudes and phases calculated can be related to the Kilpatrick parameters; however, for finite displacements this linear relationship deviates significantly. Han and Kang [9, 56] further modified the concept by introducing correction terms for describing nonequilateral rings [9, 56].

Han and Kang [56] proposed that to construct the geometry of a five-membered ring with two puckering parameters and five endocyclic bond lengths, they suggested a pseudorotation model expressed in terms of puckering amplitude and phase angle. Their proposed model is given by:

$$\alpha_j = q_\alpha \cos\left(\phi + \frac{4\pi(j-1)}{5}\right) \quad j = 1, \dots, 5 \quad (1.9)$$

where α_j is the angle between the x-y plane and the line joining the origin and the j^{th} atom, q_α is the puckering amplitude and ϕ is the phase angle.

In 2003, Hudaky and co-workers [57] outlined a method of calculating the pseudorotational parameters for a proline residue. In their work, the amplitude, A, gives

the maximum threshold to the five endocyclic torsional angle values, $\chi_0, \chi_1, \chi_2, \chi_3$, and χ_4 , while the phase angle, P , describe the state of pucker in the pseudorotation pathway, which characterize their ratios. These two parameters are calculated from the endocyclic torsional angle values as follows:

$$\chi_j = A \cos(P + 144^\circ j) \quad j = 0, \dots, 4 \quad (1.10)$$

In this case, the dimensions of both A and P are in degrees. The algorithm for calculating pseudorotational coordinates, A and P , from values $\chi_0, \chi_1, \chi_2, \chi_3$, and χ_4 , are as follows:

$$A = \sqrt{(A \sin P)^2 + (\chi_0)^2} \quad (1.11)$$

where
$$A \sin P = \frac{\chi_1 - \chi_2 + \chi_3 - \chi_4}{-2(\sin 144^\circ + \sin 72^\circ)} \quad (1.12)$$

and

$$P = \begin{cases} \cos^{-1}\left(\frac{\chi_0}{A}\right), & \text{if } A \sin P \geq 0 \\ -\cos^{-1}\left(\frac{\chi_0}{A}\right), & \text{if } A \sin P < 0 \end{cases} \quad (1.13)$$

It is to be noted that parameter A is always positive but the P value falls in the range between $-180^\circ, 180^\circ$.

Often the conformational analysis of the proline residue indicated that the peptide bond preceding the proline residue is far more important than the peptide bond following it. This is mainly due to the fact that a cis peptide bond only ever exists before the proline residue. [52, 58]. A cis peptide bond exists after the proline when the following residue is also a proline [52]. This results in selecting only three backbone torsional angles ϕ , ψ and ω being used for conformational studies [52]. It has been noted that the puckering of the five membered rings can adopt two states of pucker, and then a four-character code is used to label a proline residue [19, 52, 56].

The trans ($\omega = 180^\circ$) and cis ($\omega = 0^\circ$) conformations of proline containing peptides are defined by the orientation of the methyl carbon of the acetyl group, or peptide bond preceding the proline residue and the C_α of the proline residue, which are denoted by “t” and “c”, respectively. The backbone conformation is also denoted by the three conventional notations of γ_L , ϵ_L or α_L , which signals $\psi \approx 60^\circ$, 180° and -60° [19], respectively.

The down and up puckering conformations of the proline are denoted by “d” or “+” or “u” or “-”, respectively [19]. The sign “+” or “-” is generated from the sign of the χ_1 torsion angle, if the sign is positive it is down and if it is negative it is up. This results in a theoretically predicted total number of 12 conformers over the conformational hypersurface potential defined by ω , ψ and χ_1 parameters (Figure 1.7)

Figure 1.7 shows the 12 theoretically predicted conformers of a proline residue in the conformational space defined by ψ , ω and χ_1 variables.

ω' peptide bond isomer

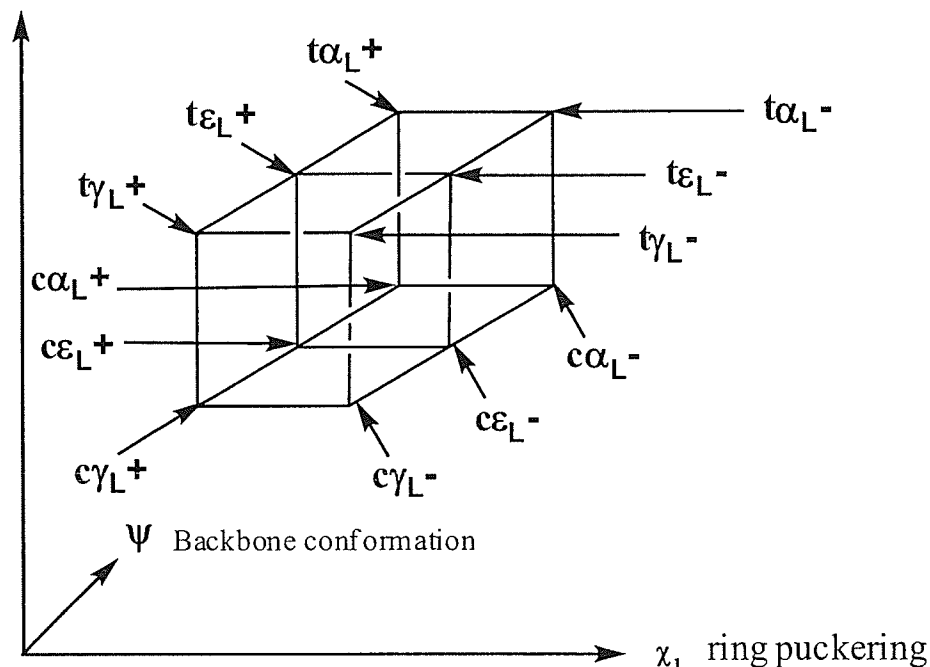


Figure 1.7 the conformational space of a proline residue defined by ψ , χ_1 and ω . The figure has been drawn following references Hudaky et al [57] and Perczel et al [41].

Then, $t\alpha_L^+$ or $t\alpha_L^-$ denoted the conformational structure of a trans proline peptide bond, with a backbone dihedral angle of ϕ , ψ falls in a region of the Ramachandran map α_L , a typical characteristic of 3_{10} and α helices; and with a down puckering ring. In this study figure 1.7 will be used as a guideline and also will follow the above procedure to characterize the conformational distribution of the novel proline containing peptides. Repeating the ϕ , ψ torsional angles always leads to some type of regular structure provided that they are sterically allowed. Table 1.2 shows the conformational dihedral angles of ϕ and ψ , of several secondary structures.

Table 1.2 Approximate torsion angles for some regular secondary structures

Secondary structures	Φ (deg)	Ψ (deg)
Right-handed α helix (α)	-57	-47
Left-handed α -helix (α_L)	57	47
Parallel β pleated sheet ($\uparrow\uparrow$)	-119	113
Antiparallel β pleated sheet ($\uparrow\downarrow$)	-139	135
Right-handed 3_{10} helix (3)	-49	-26
Right-handed π helix (π)	-57	-70
2.27 ribbon (2)	-78	59
Left -handed polyglycine II and poly-L- proline II helices (II)	-79	150
poly-L-proline I helices (II)	-83	153
Collagen (C)	-51, -76, -45	153, 127, 148

This Table has been drawn following reference Voet et al. [4]. Ref. IUPAC –IUB [47], and Flory, P, J. [46], as cited in [4]

1.4. Role of Proline and Hydroxyproline in the Collagen Triple Helix

Collagen (Greek name: *kolla*, which means glue) is the most abundant protein in animals or mammals, and is an essential proteinaceous component of the extracellular matrix (ECM) in vertebrates; it makes up about one quarter of all protein in the whole human body [4, 18, 64, 65, 66]. It is the most important structural protein, with an insoluble fiber of great tensile strength. It has a role as a major stress-bearing component of connective tissues, such as skin, basement membrane, bone, tendons, muscles, ligaments, cartilage, blood vessels, and teeth [4, 66]. Thus, it provides structure to the body, and maintains and connects organs to the skeleton [66]. To date, at least 28 different types of collagen have been identified, and several other proteins with triple-helical structures have been shown to contain collagenous or collagen-like domains [65, 67, 68].

Over 90% of the collagen content in our body is of Type I (i.e. in bone), Type II (i.e. in cartilage), Type III (i.e. in reticulate) and Type IV (i.e. basement membrane) in nature [4]. It is also important to note that collagen has a distinct amino acid composition, about 1/3 of the collagen residues are glycine or Gly, 15- 30% are either Pro or Hyp residues (Pro and Hyp have a distinct puckering nature in collagen), and the rest are other types of amino acids. The regular sequential arrangement of the amino acid residues is Gly-Xaa-Yaa, where Xaa and Yaa are any amino acid residues, but Pro and Hyp occur frequently [4].

Ramachandran and Kartha [67, 69, 70] were the first to propose a model for the collagen triple helix in the 1950s. This model was subsequently refined by Rich and Crick [67, 71] to its currently known structure after extensive X-ray fiber diffraction and modeling

studies. Rick and Crick proposed that the unique three-dimensional tertiary structure of collagen is to be a shallow right-handed triple helix formed by the parallel coiling of three left-handed polyproline II-type (PPII) strands about a common axis (Figure 1.8a-b) [4, 70]. However, in 1994, Berman, Brodsky and co-workers confirmed this model using the X-ray crystal structure determination of the triple-helical collagen mimic or Collagen-like polypeptide (Pro-Hyp-Gly)₁₀ [67, 71]. The strand in this collagen mimic had the sequence of (Pro-Hyp-Gly)₄ - (Pro-Hyp-Ala) - (Pro-Hyp-Gly)₅ in which one fifth of Gly is replaced by Ala [4, 67, 69].

Interestingly, in this structure every third residue of each strand passes through the center of the triple-helix, which is so tight that only the smallest of the twenty common amino acids, the Gly side chain, can fit in that small tight space (Figure 1.8c), unless the triple-helix structure will be distorted.

This suggested the unconditional requirement of a Gly at every third position of the collagen triple helix. In this triple helix each of the three strands are staggered relative to each other so that Gly, Pro, Hyp residues from the three chains occur at similar levels [4, 67, 71]. This staggered nature of the strands has the residues arranged in such a way that the N-H of the Gly is forming a hydrogen bond with the C=O of the X-Pro residue, which is helpful in stabilizing the triple helix [4]. In addition, the structural arrangement mentioned above also supports the hypothesis [67, 72, 73] that the hydroxyproline or Hyp stabilizes the collagen by forming a hydrogen bond with the water molecules around the triple helix [67, 74].

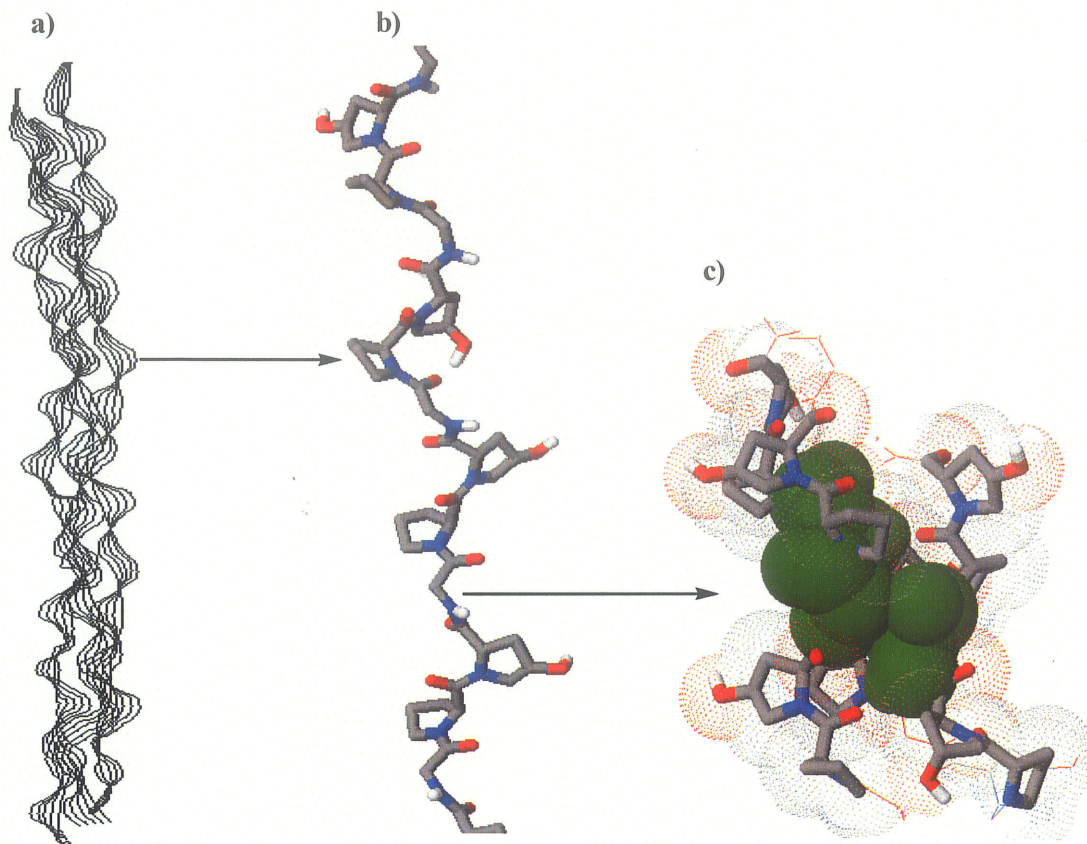


Figure 1.8 describes a) the triple helix collagen of Gly-Pro-Hyp sequence b) a single peptide chain of the collagen triple helix c) the Gly residue in the collagen triple helix and the close packing of the atoms along the helix axis. *Copyright permission obtained from Karl M. Oberholser, Messiah College, on June 02, 2009 [6].*

Several studies on the high-resolution structure of the collagen mimics have arrived at different conclusions regarding the factors that are most important to collagen stability [68]. Berman et al. [59] suggested that Hyp doesn't affect the triple-helix structure directly but the contribution of the Hyp to the collagen triple helix only arises from the Hyp-water interactions [59, 68]. In their work they demonstrated that the triple-helix that

interactions [59, 68]. In their work they demonstrated that the triple-helix that they refined from a number of crystals of (Pro-Pro-Gly)₁₀ and (Pro-Hyp-Gly)₄ - (Pro-Hyp-Ala) - (Pro-Hyp-Gly)₅ proves that the main chain torsion angles and the first hydration shell patterns around the peptide are very comparable, but the two structures show a different propensity of crystal packing [59]. In all structures they observed that the preference of the pyrrolidine ring in each Xaa position has a down pucker (C_γ -endo), but the preference of the pyrrolidine ring in Yaa position has an up pucker (C_γ -exo).

However, Okuyama et al. [60] arrived at a different conclusion. These authors found out that both (Pro-Pro-Gly)₁₀ and (Pro-Hyp-Gly)₁₀ have a certain structural similarity. This close structural similarity led them to conclude that not only Hyp doesn't influence the molecular structure directly but also the contribution of Hyp to the collagen stability may not be due to the water bridges, because nearly the same number of water molecules were found in both structures [68]. Similar conclusions were also drawn by Engle et al. [61], in which they said that even in anhydrous solution Hyp can still give extra stability.

Another interesting observation by Zagari et al. [62, 63, 68] is that the Pro residue has a unique ϕ torsion angle in the Xaa ($\sim 75^\circ$) and Yaa ($\sim 60^\circ$) positions, respectively, and these different angles are correlated with the puckering of the Pro residue. Pro in the Xaa position shows the (C_γ -endo) and Pro in the Yaa position exhibits a (C_γ -exo). They also indicate that having Hyp in the Yaa position is more stable than having Pro, because it decreases the number of conformations accessible to the unfolded state and provides a higher tendency to fold into a triple helix. However, Hyp in the Xaa position prevents triple-helix formation as it cannot adopt the C_γ -endo pucker and the unfavorable ϕ angle

[68]. This suggested that the impact of proline or hydroxyproline in the collagen triple-helix is very significant.

1.5. Purpose of the Current Research

In recent years, a number of proline analogues have been developed to study the structural and biological properties of proline surrogates in peptides. This is due to the fact that the prolyl N-terminal *cis/trans* isomerization rate and equilibrium ratios of specific proline analogues are helpful in detecting and monitoring the local structure and environment of proline. Although these analogues have proven to be useful for inducing specific constraints on prolyl N-terminal amide isomerization, they lack the ability to shift the prolyl amide equilibrium in both directions. To surmount this problem, the Schweizer group, University of Manitoba, recently incorporated carbohydrate-templated proline mimetics into some model peptides. This allows the equilibrium to be shifted in both directions. Therefore, the main goal of the current research is to investigate and understand the conformational preference and characterization of intramolecular hydrogen bonding of these model peptides so as to determine whether the carbohydrate moiety influences the *cis/trans* ratio, using computational studies.

Here we present a detailed computational analysis of the conformational preference (lowest energy conformer) and structural behaviour of these sugar proline analogues using density functional theory (DFT) methods in the gas phase and in aqueous solutions. This work also presents the thermodynamic properties and electronic structure of these sugar proline analogues. The thesis is organized into six chapters. In chapter 1 the *cis-trans* isomerization of the prolyl amide and its importance in a number of biological

applications are reviewed, with an emphasis on how collagen helix stability is affected by hydroxyproline. The fact that the improper production of collagen can cause a number of problems in the maintenance and healing of cartilage and the strengthening of joints, tendons, bones, lenses of the eye and cardiac muscles is discussed. Some important concepts such as the nomenclature and pseudorotational parameters of a proline residue are also introduced. Chapter 2 introduces the two novel carbohydrate-templated proline based mimetics that were synthesized by the Schweizer group, University of Manitoba. These are discussed and reviewed. The structural features and physical properties of these novel carbohydrate-templated proline mimetics are also explained in greater detail. A meticulous description of the DFT, MMFF94 and Monte Carlo methods and their mathematical concepts are available in Chapter 3. The computational results for these compounds are analyzed, discussed, and compared to the experimental observations. These results and discussions are presented in Chapter 4 and Chapter 5 separately for both synthesized compounds. Finally, the conclusion and summary of this research project with possible directions for future work are presented in Chapter 6. Hence, this quantum mechanical study of the sugar proline analogues, in addition to the experimental data will give further insight into prediction of the conformational equilibria and kinetics of carbohydrate-templated proline-based peptidomimetics.

Chapter 2

Carbohydrate-Templated Proline Mimetics

2.1. Computational studies of Proline based novel model peptides

The development of the understanding of the conformational preference and cis-trans isomerization of a proline residue in peptides or proteins has a prolonged record [44, 64]. In 1958, Katchalski et al. [44, 76, 77] anticipated that protonation of the amide nitrogen would alter the hindered rotation between the cis-trans conformations in peptides or proteins by changing the peptide's partial double bond nature to a virtually free rotation of the N-protonated peptide bond. They demonstrated this by using the conversion of poly-L-proline I to poly-L-proline II, and vice versa in different solutions e.g. benzyl alcohol and 2-chloroethanol. The idea was further investigated by Brandt and co-workers [44, 78] with respect to the O-protonation followed by N-deprotonation. This concept further supported the proposed occurrence of the cis-trans isomerization at the peptide bond, in particular for the proline residue, and since then, it has been reported that certain enzymes (or rotamases) can also catalyse the cis-trans isomerization to a greater extent [44, 79, 80].

Many researchers have developed a number of proline analogues to study the structural stability and biological properties of proline surrogates in peptides and proteins. Some of the experimental and computational work on proline analogs are performed on

C_β , C_γ and C_δ substituted prolines [81-84], azaproline [85], pseudoproline [86], silaproline [87], proline-amino acid chimera [88], and fused bicyclic proline [89-91]. These analogues induce specific constraints to the prolyl cis-trans isomerization. In particular, the pseudoproline bearing two substituents adjacent to the N-terminal prolyl nitrogen and endocyclic C_δ -substituted prolines that contain bulky substituents have been revealed to enhance the prolyl amide *cis* population in peptides [81, 92, 93]. Not only that, but also the incorporation of pseudoproline into peptides chains has been revealed to induce a “kink” kind of conformation in the backbone of the peptides. This “kink” formation on the backbone of the peptide results from the conformational preference of the prolyl residue for a *cis* amide conformation. Interestingly, this kind of structure prevents the formation of peptide aggregation, self- association and β -structure, thereby enhancing solvation and coupling kinetics of the increasing peptide chain [94].

In 1970, Maigret and co-workers [44, 95] pioneered the computational work of the cis-trans isomerization of the prolyl residue. They found that the energy difference between *cis* and *trans* structures is about 0.5 kcal/mol. Their work was mostly based on the gas phase, and the environmental factors were ignored (e.g. solvent effects). However, they were the first ones to report on the conformation of the *cis* and *trans* conformers using the Potential Energy Curve (PEC) and Potential Energy Surface (PES).

Farmer and Hopfinger [96] have also done some work on the conformational analysis to describe the molecular motion of the cis-trans isomerization in Poly (L-Proline). They suggested that the free energy barrier to both *cis-trans* and *trans-cis* transitions are

minimum when three residues are involved in the cooperative motions. Their calculations do not account for the solvent-solute interaction that can greatly affect the barrier heights of the isomerization but they provide a qualitative description of the isomerization molecular motions.

Karplus and co-workers [97], have also studied the cis-trans isomerization reaction of proline dipeptides (N-Acetylproline methylamide) extensively. They noticed that the reaction path for this isomerization is very complex and involves the imide torsion angle ω' ; the proline backbone dihedral angle ψ ; and the pyramidalization of the imide nitrogen (η). They emphasized the importance of the pyramidalization of the imide nitrogen in the process of isomerization, which is quite an interesting analysis for the stabilization of the cis population. The barriers predicted are 17.9 and 20.7 kcal/mol for the trans-cis isomerization for clockwise and anticlockwise rotations, respectively. A strong influence on the rotational barrier by ψ was also observed, and a change of ψ can alter the barrier value greatly (by as much as 12 kcal/mol). For $\psi \sim 0^\circ$, they showed that the interaction between the C-terminal -NH group and the N-terminal lone pair can also greatly reduce the energy barriers. This effect is called "autocatalysis". It is a factor in the peptidylprolyl isomerases like FKBP and Cyclophilin. Finally, they also indicated that rotamerase enzymes can reduce the enthalpy barrier by 5-6 kcal/mol [44, 97].

Kang and Choi [98] also reported extensive work on the conformational preference of N'-acetyl-L-proline-N'-methylamide (Ac-Pro-NHMe) and its prolyl cis-trans isomerization in gas phase and solutions using the HF/6-31+G(d) level of theory and the conductor-like polarizable continuum model (CPCM) which is a self-consistent reaction field solvation method (see below). They indicated that the free energy barriers in water

for the trans-cis and cis-trans isomerization are 19.0 and 18.8 kcal/mol, respectively, which is in accord with the experimental values of 20.4 and 19.8 kcal/mol, respectively [98]. They also mentioned that the overall barrier for rotation is greatly affected by the polarity of the solvent. It is higher in polar solvents compared to nonpolar ones. The rotational barrier to the ring flip or puckering from the down-puckered to the up-puckered conformer is approximately 2.5 and 3.2 kcal/mol for the trans and cis conformers, respectively. In addition, the torsion angles for the backbone and prolyl ring vary greatly around the transition states, when $\omega' \approx 120^\circ$ and -70° for the prolyl peptide bond.

One of the most effective models used for the computational studies of the proline residues was the N-formyl-prolinamide; see Figure 2.1 [44]. The main reason for choosing this model was that the cis-trans isomerization of N-formyl-Prolinamide model peptide is considered to be an equilibrium between structures of comparable stabilities. This is due to the fact that the $-\text{CONH}_2$ bulky substituent on the five-membered pyrrolidine ring or the endocyclic ring induces a relatively “*minor perturbation*” on the conformational equilibrium of the prolyl residue in the model peptide [44]. Enriz et al [44], presented a great work on this model peptide and reported that the barrier heights for the cis-trans isomerization and ring puckering are in the range of 19.58 to 24.6 kcal/mol, and 1.76-7.46 kcal/mol, respectively. They also reported that the backbone conformation interconversion spans the range from 0.61-5.56 kcal/mol. Their entire calculations were performed using the RHF/3-21G, RHF/6-31G (d) and RB3LYP/6-31G (d) levels of theory.

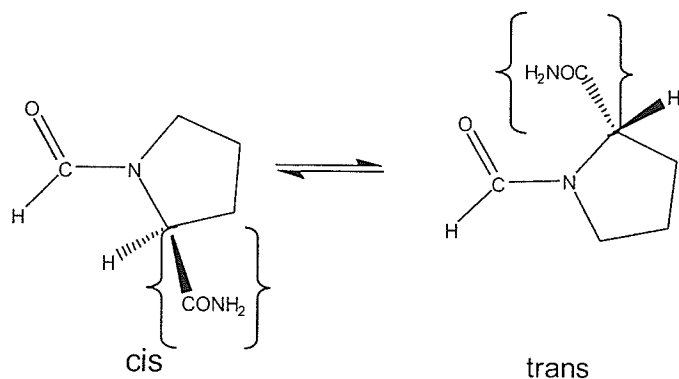


Figure 2.1 The cis-trans isomerization of the N-formyl-prolinamide, The figure has been drawn following reference Enriz et al [44]

In 2004, Che and Marshal [99] reported an extensive study of the amino acid analogue azaprolines (azPro) using NMR and computational studies (DFT, *ab initio*) for the exploration of the conformational preference. This amino acid analogue contains a nitrogen atom in place of the C_α of proline (Pro derivative). Their study indicated that peptides containing azPro stabilize the cis-amide conformer for the acyl-azPro bond, thereby preferring type VI beta-turns. In their report a clear conformational preference for the cis-amide conformer of azPro was observed, although the increased stability for the cis-amide conformers was relatively minor with respect to the trans-conformers.

In the azPro analog, the energy barrier for the cis-trans amide isomerization was found to be 6 kcal/mole less than in the Pro derivatives. They have also indicated that an intramolecular hydrogen bond was only observed in the trans-amide conformer that greatly reduced the conformational preference of the cis-amide conformer by 2.2

kcal/mole. From their work they concluded that azPro derivatives can stabilize the cis-amide bond and can mimic a type VI beta-turn without any additional input of steric bulk, and act as a conformational restriction for the peptide backbone when incorporated into the structure of selected bioactive peptides [99]. The model azPro peptides developed are shown in Figure 2.2.

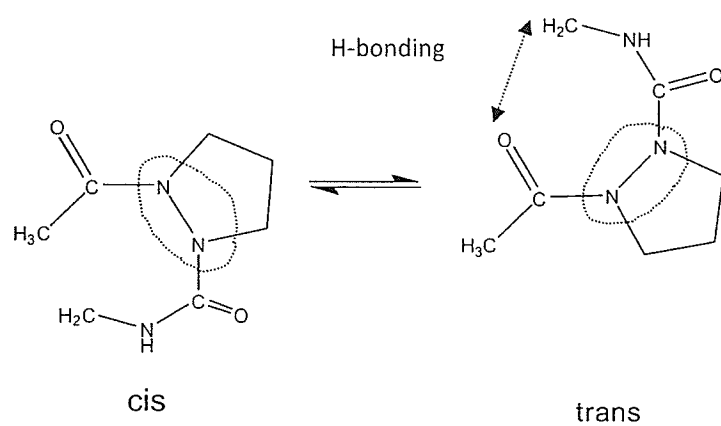


Figure 2.2 The cis-trans isomerization of the Ac-azPro-NHMe, This figure has been drawn following reference Che and Marshal [99]

Generally, although these analogues mentioned above have proven to be valuable for inducing specific constraints on prolyl N-terminal amide isomerization, they do not provide the capability of shifting the prolyl amide equilibrium between the cis and trans isomers. This means that different proline analogs are needed to induce a desired bias in the rate of equilibrium constant of the compounds. Therefore, in the following sections two new kinds of proline analogues which are synthesized by the incorporation of sugar/glucose into proline's five membered pyrrolidine rings, and which can shift the prolyl cis-trans isomerization in both directions, will be introduced. The structural

features and physical properties of these novel carbohydrate templated proline mimetics will be discussed in detail throughout in the next two sections.

2.2. Fused Carbohydrate-Template Proline Analogue – N-acetyl-

GlcProH- NHMe

As we have seen in section 2.1, a number of proline analogues have been synthesized and developed over the last few years. However, none of them have been developed with strategic functional groups positioned for further derivatization so that we can easily alter the amide equilibrium ratio of the cis-trans conformers. This drawback lead the Schweizer group at the University of Manitoba to develop novel carbohydrate templated proline analog building blocks in which prolyl cis-trans isomerization could be easily altered through simple synthetic routes, even after the incorporation of these analogues into the peptide chain. In other words postsynthetic modifications of these model peptides can be easily introduced through the derivatization of the carbohydrate scaffold [100].

Recently, Owens, Schweizer and co-workers [100], synthesized a fused bicyclic C-glucosyl proline hybrid compound, and then incorporated this compound into the model peptide Ac-GlcProH-NHMe. The synthetic route to this novel model peptide is very long, and since the main focus of this study is on the computational analysis of the final product, only a portion of their synthetic route will be shown in this work. They started from N-Boc-GlcPro-COOH (a), as shown in figure 2.3 which is synthesized through several steps from a commercially available 2,3,4,6-O-benzyl-D-glucopyranose [100]. In their synthesis the protection of the amino function was achieved through tert-butyloxycarbamate (Boc), as shown in Figure 2.3(a).

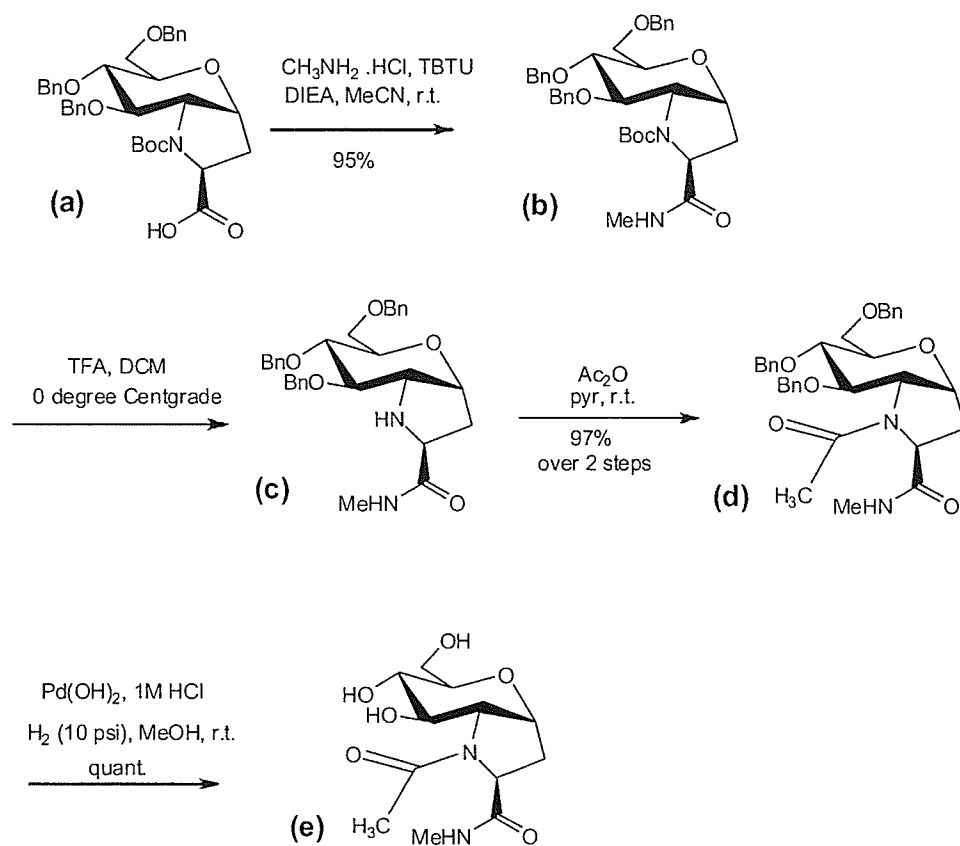


Figure 2.3 The synthesis of Ac-GlcProH-NHMe (e), This figure has been drawn following reference Owens and Schweizer [100]

The first step in their synthesis revealed that the compound (a) was directly coupled to methylamine by the use of O-benzotriazolyl-N,N,N',N'tetramethyluroniumtetrafluoroborate (TBTU) which is used as a coupling reagent in MeCN so that compound (b), an amide, was produced in a 95% yield. The deprotection of compound (b) and subsequent acylation of compound (c) to produce the diamide (d) was achieved in a 97%

yield over the two steps. Finally, the benylether protecting groups were removed to produce the polyhydroxylated diamide (e) by catalytic hydrogenolysis in methanol. In the last few years a number of quantum mechanical computational tools were deployed to investigate the structural and kinetic properties of proline analogues with results that were often in excellent agreement with the experimental results [7, 25, 41, 98]. As a result, computational methods are applied in this study to investigate the population distribution of the cis-trans isomerization, the peptide geometry and the intramolecular hydrogen bonding between the sugar substituents and the peptide backbone in the gas phase and aqueous solution in order to predict the effect of the sugar incorporation into the proline-based peptide, compound (e).

Figure 2.4 shows the general structure of the glucose proline hybrid (GlcProH) compound. The carbohydrate scaffold in this case is used as a template to constrain the five membered ring of the pyrrolidine ring of the L-proline, in particular for the C_β , C_γ , C_δ , and N, as described in Figure 1.5c. R denotes the hydroxyl-derivative substituents which are used to manipulate the steric and electronic properties of the glucose proline hybrid. In this research R= H.

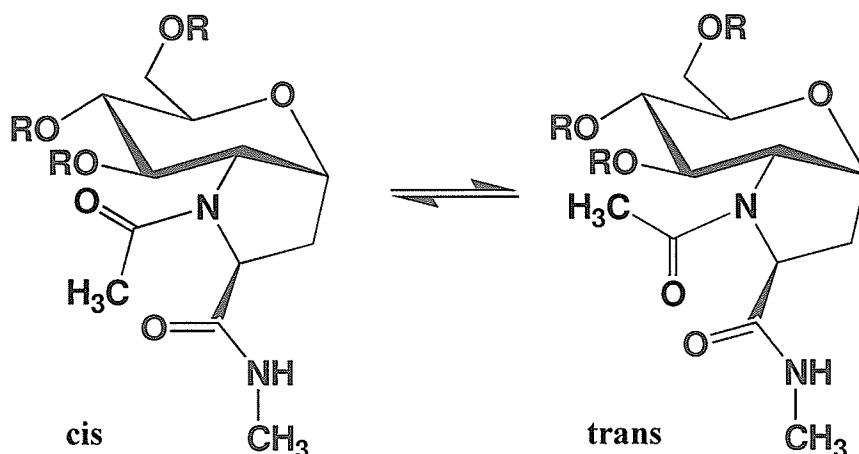


Figure 2.4 shows the cis-trans isomerization of the Ac-GlcProH-NHMe, compound (e).

2.3. Spirocyclic Carbohydrate-Template proline analogue - N-acetyl-GlcPro-OMe

In addition to the Fused Carbohydrate-template proline analogues, Zhang and Schweizer [101] have synthesized also the spirocyclic glucose-3-hydroxyproline hybrids recently. In particular, the spirocyclic nature of the gluco-derivative scaffold puts constraints onto the five-membered pyrrolidine ring of proline and also imposes non-natural post-translational modifications, which are the hydroxylation and glycosylations. Moreover, like the fused carbohydrate template proline analogues the chemical manipulations and derivatizations of the glucose-derived polyol scaffold provide an opportunity to modify the chemical, physical and pharmacodynamic properties of Glc3(S) Hyp-containing peptides [100]. The synthetic route to the compounds of our interest in this research from the spirocyclic glucose-3(S)-hydroxyproline hybrids (Glc3(S)HypHs) is described below,

and only a portion of the synthetic route is shown as the emphasis of this work is again on the computational studies of the final compound [101].

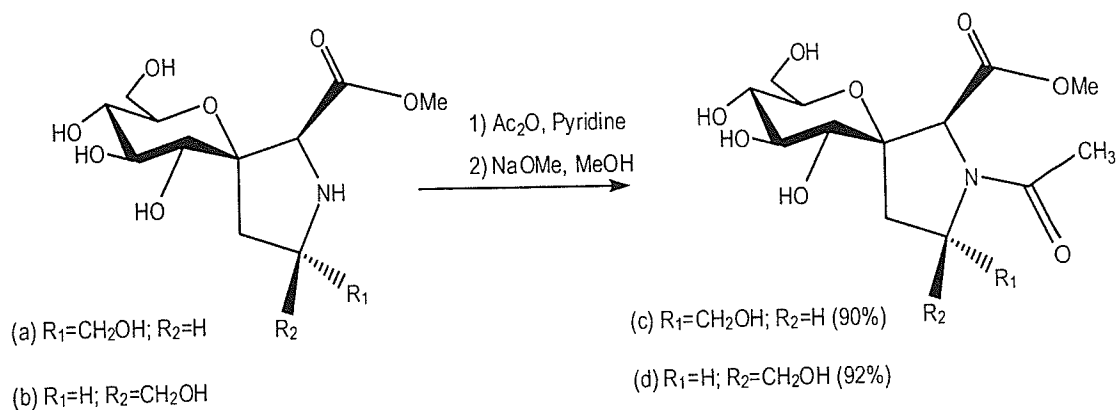


Figure 2.5 The synthesis of -N-acetyl-GlcPro-OMe, This figure has been drawn following reference Zhang and Schweizer [101]

The peptide mimics, compound (c) and (d) were obtained by acetylation of compounds (a) and (b) using acetic anhydride in pyridine followed by O-deacylation using sodium methoxide in methanol (Figure 2.5). The most fascinating property of compounds (c) and (d) is that they contain a hydroxymethylene substituent, which is adjacent to the imino function of proline. These substituents may contribute a control on the prolyl amide cis/trans isomerization through a strong hydrogen bond or some other forms of electrostatic or steric interactions.

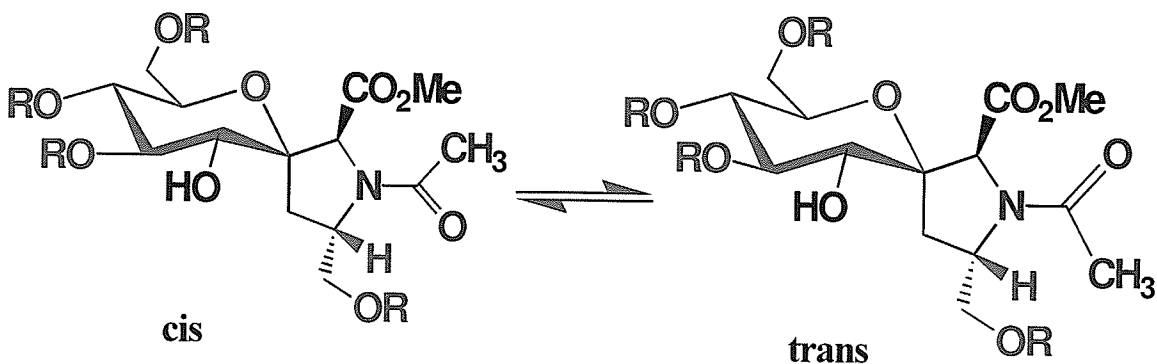


Figure 2.6 The *cis-trans* isomerization of Ac-Glc3(S)-HypH-OMe, in which the hydroxymethylene substituent is pointing *down*.

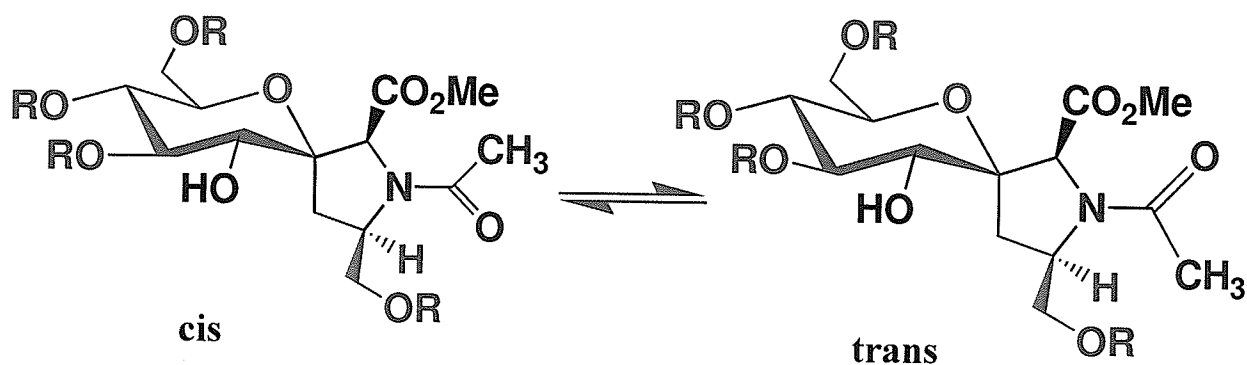


Figure 2.7 The *cis-trans* isomerization of Ac-Glc3(S)-HypH-OMe, in which the hydroxymethylene substituent is pointing *up*.

Therefore, in this thesis a conformational analysis of these novel compounds (c) and (d) are performed, and also investigate into the role of the hydroxymethylene substituent on the *cis-trans* population shift. The two compounds, (c) and (d) investigated in this research are shown in figures 2.6 and 2.7. The conformational distribution of the *cis-trans*, the backbone and the ring puckering of compounds (c) and (d) will be investigated

in both gas phase and in water. Moreover, the hydrogen bond will also be characterized to see if these compounds show some sort of intramolecular hydrogen bonding.

2.4. Conclusions

In this chapter the fundamentals of the carbohydrate-templated proline analogues were discussed along with their distinct advantages over the existing proline analogues. As indicated earlier, the presence of the unprotected glucose moiety in GlcProH provides opportunities to explore the effect of glycosylation in unusual glycopeptides while enrichment of the gluco-based polyol scaffold provides rich opportunities to alter the physical, chemical, hydrophobic, lipophilic nucleophilic, pharmacodynamic properties of proline mimetics and proline-containing peptidomimetics. A detailed computational analysis of the conformational preference of the structures depicted in figure 2.4, figure 2.6 and figure 2.7 is provided in Chapters 4 and 5.

Chapter 3

Computational Methods

3.1. Computational Theories and Models

Computational chemistry or molecular modeling [102], as the phrase states, is a branch of chemistry that deals with computations or theoretical aspects of chemistry. It is used in a number of applications, like in understanding chemical reactions or processes, structural determinations, chemical bonding, and other thermodynamic quantities. It is also used for various predictions prior to synthesizing a molecule or running the actual experiments [102]. Computational methods may not be accurate at all times but often reproduce or generate, depending on the model used, very accurate experimental results. They are often good enough to rule out 90% of possible compounds, as being inappropriate for their anticipated use, in conformational analysis [103-104].

Importantly, it is a very supportive tool in various organic and biochemical applications as synthesizing even a single molecule could cost a large amount of capital resources, labor time and generate toxic materials (e.g. radioactive elements) [104]. As a result, many experimental organic or biochemical synthetic chemists are now incorporating these computational models at a great pace into their research projects to gain an insight into their compounds, in particular in pharmaceutical applications [104].

The application of computational chemistry is very broad and applied to a variety of fields, however, in this study it will be applied to study the potential energy surface (PES)

of carbohydrate-template proline analogues, which were discussed in Chapter 2, in both gas phase and in solutions. A number of models with different approaches have been developed for the investigation or exploration of the potential energy surface of any given molecule, like molecular mechanics based models (MMFF94 [105-109], SYBYL [110]), *ab initio* methods (Hartree-Fock, post Hartree-Fock) [103], Density Functional Theory methods [103] and others.

In the last few years, the development of sophisticated and highly optimized computer codes together with the availability of supercomputing machines or clusters have enabled the full exploration of the potential energy surface of any given structure, in particular for small and medium size molecules [111-115]. This is due to the great innovation and introduction of high quality analytical energy gradients that makes possible the location of energy extrema of any given molecule on the potential energy surfaces [103]. As a result it is now a very common practice to carry out a full structural characterization of many stable molecules or equilibrium geometries by using *ab initio* quantum mechanical methods using the energy gradient methods [103]. For example, the use of *ab initio* methods which are entirely based on the Hartree-Fock (HF) scheme is always a common practice to apply for the potential energy surface calculations or full conformational explorations for compounds of organic and bioorganic molecules [105].

It is also possible to explore and characterize the potential energy surfaces of any given molecule by methods based on the density functional theory (DFT), which we will discuss in greater detail in the subsequent sections [115-120]. DFT is based on an electron density calculation framework rather than the wave-function approach like Hartree-Fock. In the last few years, this method has attracted a large number of

researchers and scientists due to its computational efficiency and its relatively sophisticated description of the electron correlation effects, which are missing in Hartree-Fock method. In fact the computational cost of DFT calculations are proportional to N^3 , where N is the number of basis functions. This is a dramatic reduction relative to N^4 and N^7 (or greater) for the HF and post-HF calculations, respectively [103, 113, 121]. A number of geometry optimizations for a given structure, frequency characterization, transition state structure predictions and thermodynamic calculations have indicated that the density functional theory based methods generate more accurate results and often agree with the experimental data very well, in particular for small and medium-size compounds containing C, H, and N or peptides and proteins[121].

In short, although DFT, HF and post-HF calculations are the most cited methods in computational chemistry, their use is still very restricted in the application of the conformational analysis of large molecules, peptides, and proteins [117]. Therefore, most often conformational analysis calculations for large molecules, like proteins and peptides are still based on Molecular Mechanics, due to the high computational cost of the quantum mechanical based methods (DFT and HF methods) [121].

Thus, in this research first Molecular Mechanics calculations are used to explore the potential energy surface of the carbohydrate template proline analogues using Systematic and Monto Carlo searching sampling techniques to extract all the possible local minima conformers in the conformational space. Once all the possible structures or conformers are obtained, the density functional theory method will be used for the remaining calculations in both gas phase and in solutions. A combination of both molecular

mechanics and density functional theory approaches are used to investigate the conformational preference of the novel carbohydrate-templated proline mimetics.

3.1.1 The Schrödinger Equation

This section is devoted to some of the mathematical aspects of the fundamental quantum chemical model framework of the Schrödinger Equation, and some of the approximations. The central “theme” of most of the conventional quantum chemical models is the wave function Ψ_i [116]. The main reason is that once we exactly know Ψ_i or a good approximation to it, then we have access to all the information or properties that can be known about this particular state of the system of interest [116]. The wave function is obtained as an (approximate) solution to the time-independent, non-relativistic Schrödinger equation, which describes in terms of many electron wavefunction [116, 121, 137].

$$\hat{H} \Psi_i(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N, \vec{R}_1, \vec{R}_2, \dots, \vec{R}_M) = E_i \Psi_i(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N, \vec{R}_1, \vec{R}_2, \dots, \vec{R}_M) \quad (3.1)$$

Where, \hat{H} is the Hamiltonian operator for a molecular system, M is the number of nuclei, N is the number of electrons, and Ψ_i is a many-electron wavefunction. The Hamiltonian operator is a representative of the total energy and is given by (atomic units) [116, 121]:

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \frac{1}{2} \sum_{A=1}^M \frac{1}{M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (3.2)$$

Z_A is the nuclear charge of nucleus A, M_A is the ratio of mass of nucleus A to the mass of an electron, R_{AB} is the distance between nuclei A and B, r_{ij} represents the distance between electrons i and j , and the term r_{iA} is the distance between electron i and nucleus A [116, 121]. The first and second terms in equation 3.2 refer to the kinetic energy of the electrons and nuclei respectively. The Laplacian operator ∇_i^2 is generally defined as follows:

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \quad (3.3)$$

Over the last decades the many-electron Schrödinger equation has not been solved exactly except for very few atoms like the hydrogen atom. Therefore, a number of approximations have been made to at least provide practical results. Some of the most notable approximations are the Born-Oppenheimer and the Hartree-Fock approximations [110].

3.1.1.1 The Born-Oppenheimer Approximation

The main assumption of the Born-Oppenheimer approximation is that if we assume that the nuclear and electronic separation is *approximately* correct by assuming that the nuclei are fixed and do not move at all relative to the electrons around it. Within this approximation, the kinetic energy of the nuclei is zero and the potential energy due to the nucleus-nucleus repulsions is constant [110, 115, 116]. This is based on the fact that the

nuclei are much more massive than the electrons around them, making it not unreasonable to claim that the nuclei are nearly fixed with respect to electron motion [116, 121]. As a result equation 3.2 reduces to only solving the electronic wavefunction at a fixed set of nuclear coordinates, and is given by:

$$\hat{H}_{elec} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (3.4)$$

Hence, the solution of the Schrödinger equation for equation 3.4 is generally given by the electronic wavefunction and electronic energy, which depends only on the coordinates of the electrons [115, 116]. Then the overall total energy is given as the sum of both the electronic energy and the nuclear energy and is given by the following equation [115, 121].

$$\hat{H}_{elec} \Psi_{elec} = E_{elec} \Psi_{elec} \quad (3.5)$$

And

$$E_{tot} = E_{elec} + E_{nuc} \quad (3.6)$$

The constant nuclear repulsion term is given by:

$$E_{nuc} = \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (3.7)$$

This approximation is generally valid and is helpful in the PES calculations. The PES is given by the E_{tot} as defined within the Born-Oppenheimer approximation. This helps us in understanding the equilibrium geometry and transition state structures of any give molecule, despite the fact that it does not tell a lot about the isotopic effect on molecular properties and etc. Therefore, this becomes the basis for a number of quantum mechanical models, like DFT [110, 121].

3.1.1.2 The Hartree-Fock Approximation

The fundamental assumption of Hartree-Fock theory is that each electron in a molecule sees all the other electrons as an average field or indirectly. Thus, it treats the motion of individual electrons as independent of each other. In this approximation the total wave function is given in the form of the Slater determinant, ensuring that it is antisymmetric upon interchange of electron coordinates [110, 116, 121], and is given by:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_{1(1)} & \chi_{2(1)} & \dots & \chi_{n(1)} \\ \chi_{1(2)} & \chi_{2(2)} & \dots & \chi_{n(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \chi_{1(N)} & \chi_{2(N)} & \dots & \chi_{n(N)} \end{vmatrix} \quad (3.8)$$

In this case, the χ_i is the spin orbital and is given as a product of the molecular orbital and spin function. In this approach the spin orbitals are varied under the constraint that they remain orthonormal, so that the energy generated via self-consistent-field or SCF of the Slater determinant is minimal [110, 121].

This assumption greatly simplified the electronic Schrödinger equation, but it neglects the electron correlation effects in molecules, which has a profound chemical consequence in any system [103, 116]. In particular, the underestimation or overestimation of the electron-electron repulsive and attractive energy often leads to inaccurate results, which have significant implications in particular in determining both structures and frequencies of any given molecule [114-115].

To overcome this drawback of the Hartree-Fock approach of modeling a system, the post-Hartree-Fock *ab initio* methods were developed that take into account the electron correlation effects [103]. These methods (e.g. CCSD (T), MP2 [103, 113]) provide good agreement to the experimental results [103]. However, the very high computational cost associated with these methods limits their application to very small molecules, hampering their usefulness in a wide range of biological areas [113, 116].

3.2. Molecular Mechanics

Molecular mechanics (MM), also known sometimes as force field methods, refers to the use of classical mechanics theory to model the conformation or geometry, and motions of molecules [122]. It is entirely based on the concept of a mathematical formulation of a molecule as a collection of spheres (i.e. atoms) held together by springs (i.e. bonds) [123]. Within the framework of this model the chemical bonds of any type of molecule are generally treated as springs whose hardness depends on the type of elements attached to the ends and the nature of the bond between these atoms (i.e. single, double, or triple) [122]. In addition to the springs used for bond length characterizations different types of

springs are also used to describe or represent the changes in bond angles, dihedral angles and so on [121].

The conceptual framework of MM treats or describes molecules as bonded atoms, which have been distorted by the non-bonded van der Waals and Coulombic interactions from their “ideal geometry” [121-123]. In general, MM models ignore all electrons in the core mathematical formulations and this makes it completely a different model or approach from most of the quantum chemical models (i.e. HF, DFT), in which there is no direct reference to chemical bonding [121].

It is worth noting that in quantum-mechanical calculations like *ab initio* and density functional, the molecules are purely defined by the relative positions of their atomic nuclei, the charge, and the multiplicity. However, these important parameters are totally ignored in molecular mechanics calculations. This leads to loss of information regarding electronic properties like charge distribution, nucleophilic, electrophilic behaviours of the molecule etc. [121-123].

3.2.1. Molecular Mechanics Formulation

The main idea or principle in molecular mechanics formulation is to express or convey the energy of any given molecule as a function of its resistance (strain energy) as a whole towards bond stretching (i.e. bond length), bond bending (i.e. bond angle), torsion angle (i.e. dihedral angle), and contributions from non-bonded terms. The strain energy equations were used to predict the actual bond length, bond angles and dihedral angles of the most stable conformation or a set of local minima conformers in the potential energy surface [121, 122]. Combining all these various types of spring constants associated with

the various defined springs (stretch, bend and torsion contributions), together with contributions due to non-bonded (van der Waals and Coulombic) interactions, the molecular mechanics energy of a molecule is given by [121, 122]:

$$\begin{aligned}
 E^{strain} = & \sum_A^{bonds} E_A^{stretch} + \sum_A^{bonds \ angles} E_A^{bend} \\
 & + \sum_A^{torsion \ angles} E_A^{torsion} + \sum_A^{non_bonded \ atoms} \sum_B E_{AB}^{non_bonded}
 \end{aligned} \tag{3.9}$$

The first three terms in equation 3.9 (bonds, bond angle and torsion angle summations) are defined or described as the sums over all the bonds, all the angles defined by three atoms, and all the dihedral angles defined by four atoms, respectively. The last term, which is the non bond summation term, refers to all pairs of significant nonbonded interactions (i.e. atoms which are not involved in bonding) [121-122]. Each summation term in equation 3.1 has a different mathematical form and meaning, in which the mathematical formulations have different parameters and equations.

The first two terms in equation 3.9, which are associated with contributions from bond stretching and bending angles are given in terms of a quadratic description.

$$E^{stretch}(r) = \frac{1}{2} K^{stretch} (r - r^{eq})^2 \tag{3.10}$$

$$E^{bend}(\alpha) = \frac{1}{2} K^{bend} (\alpha - \alpha^{eq})^2 \quad (3.11)$$

In the above equations, the terms r and α represent the bond distance and bond angle respectively, whereas the terms r^{eq} and α^{eq} are defined as the ideal equilibrium values for the bond length and bond angle respectively, which are often obtained from extensive experimental work or from very accurate quantum mechanical calculations. $K^{stretch}$ and K^{bend} refer to the stretch and bend force constants that need to be parameterized [113, 121].

The description above is only given for quadratic terms. Molecular mechanics may also include the cubic, fourth and other forms of higher-order contributions in addition to the cross terms to account for the correlations between bend and stretch terms. In practical applications, however, they are often ignored due to the increasingly complicated nature of these terms.

The energy of rotation is often expressed in terms of a truncated Fourier series type [121] as follow:

$$\begin{aligned} E^{torsion}(\omega) = & K^{torsion\ 1} [1 - \cos(\omega - \omega^{eq})] + K^{torsion\ 2} [1 - \cos 2(\omega - \omega^{eq})] \\ & + K^{torsion\ 3} [1 - \cos 3(\omega - \omega^{eq})] \end{aligned} \quad (3.12)$$

where $K^{torsion\ 1}$, $K^{torsion\ 2}$, and $K^{torsion\ 3}$ are parameters which are used to describe or account for the one-fold term, two-fold term and three-fold term difference in energy of each conformer they represent. ω and ω^{eq} refer to the torsion angle and an ideal torsion angle of a molecule, respectively. Like the stretch and bend terms in the summation in the molecular models, the torsion energy may also include higher-order terms and cross terms to account for asymmetric environments [103, 121, 122, 137].

The non-bonded interactions in molecular mechanics are usually given by the sum of the van der Waals interactions and Coulombic interactions. They generally represent the change in potential energy with distance apart of two atoms that are not bonded directly and also are not bonded to a common atom (i.e. they do not share a common atom.). These atoms are separated by at least two atoms. Hence, a simple way of approximation to these two terms (van der Waals interactions and Coulombic interactions) is given by [103, 121]:

$$E^{VDW}(r) = \epsilon \left[\left(\frac{r^o}{r} \right)^{12} - 2 \left(\frac{r^o}{r} \right)^6 \right] \quad (3.13)$$

$$E^{Coulombic}(r) = \frac{qq}{r} \quad (3.14)$$

$$E^{non-bonded}(r) = E^{VDW}(r) + E^{Coulombic}(r) \quad (3.15)$$

The VDW contribution is represented as a sum of repulsive and attractive terms, and is given in the form of the Lennard-Jones potential [103, 121]. r represents the non-bonded distance; r^o is the interatomic separation at which repulsive and attractive forces exactly balance; and ε is a parameter which defines the Lennard-Jones potential depth or energy well [110, 121]. The Coulombic term accounts for the interaction of charges, and q represents the atomic charges of the atoms.

As stated above, all of these equations and associated parameters which are used to calculate each of the energy terms (Equation 3.8) in the strain energy are collectively called the force fields methods. In the last few years a number of force fields have been developed for different molecular types. For example, CHARMM [124] is used for biomolecules and organics; Chem-X [125], is used for organics; MM3 [126], is used for organics and biomolecules; DREIDING [127] is used for main-group organic and inorganic compounds; GROMOS [128] is used for biomolecules; MMFF [105-109], is used for organics and biomolecules; SYBYL [110], is used for organics and proteins; AMBER [129], is used for biomolecules; and so on.

Most of the time, molecular mechanics calculations on a molecule are only used for qualitative description purposes, because a quantitative description of a molecule needs to include the electronic effects of a molecule [121]. However, the quantitative applications of molecular mechanics for the evaluation of minimum energy structures, heats of formation, and strain energies have been developed and tested to a high level of reliability. This method is generally well suited for providing and extracting excellent structural parameters in terms of bond distances, angles, etc., for the minimum energy conformation of a molecule [130]. Thus, this insures the molecule is in its lowest energy

state in a conformational space so that calculated results can be compared to those done experimentally within a reasonable time frame.

A big advantage of molecular mechanics is that it is very fast and computationally inexpensive; consequently this technique is well suited for studying large molecules, especially molecules such as proteins and peptides. Indeed, we can even say that it is still the only viable method for such systems. The computing effort scales proportional to N^2 , where N is the number of atoms in the structure under consideration. This makes molecular mechanics over an order of magnitude faster than most of the quantum mechanical calculations. [103, 121, 137].

In the last few years molecular mechanics calculations have been widely and intensively applied to a number of applications, for instance in the explanation of selector-selectand interactions in chiral chromatography systems [131]. Some of the early works and successful application of the molecular modeling are: the modeling of enantioselective binding of analytes in the system with cyclodextrin (CD) cavities by Armstrong and co-workers [131, 132]; the study of binding of tryptophan by α -CD, an investigation that shows the R-enantiomer of tryptophan is more tightly bound than its optical antipode, by Lipkowitz et al. [131, 133]; and the prediction of the chirally discriminating chromatographic behavior of β -CD, by Durham and Liang [131, 134]. Besides some limitations that were pointed out (i.e. it does not give electronic properties predictions) by some researchers, most of the papers cited above illustrate the great possibility of molecular mechanics calculations as a useful supportive tool to the actual laboratory experiments, for example, in the justification of the separation mechanism and in the simplifying and mounting systems optimization [131].

3.2.2. The MMFF94 force field method

Although molecular mechanics have a broad applications in organic and bioorganic systems, the most widely used application of molecular mechanics calculations is for its use in the extensive conformational searching of molecules containing a large number of atoms on the potential energy surface. Due to the above mentioned advantages, which mean reasonable structural predictions and being computationally inexpensive, conformational analysis became one of the “*strongest single*” applications of molecular mechanics [103, 121]. Molecular mechanics is currently the only practical method for calculations of the conformational preference of a very large molecule of a highly flexible structure. Of the existing force field methods of molecular mechanics, MMFF94, in particular, has proven to be one of the most reliable and fast tools for conformational analysis study [121].

MMFF94 is a very complex force field method, which was developed at Merck Pharmaceuticals, by Halgren in 1996, and is intended to be used in various pharmaceutical applications, like geometry optimization and conformational analysis [105-109]. It has been developed through quantum-mechanical *ab initio* techniques at its core and verified by extensive experimental data sets. As it is reported by Halgren, the guiding application in the development of MMFF94 was the study of receptor-ligand interactions involving proteins and nucleic acids as receptors and a large collection of chemical structures and ligands [105-109, 121]. In this case, the force field that is developed must be able to quantitatively describe the ligand and its target individually or all together [105-109]. Then, one of the requirements in doing so was that the force field

must be able to predict the conformational energies and molecular geometries, if it is to avoid modeling the wrong conformations of the ligand or receptor upon binding [121].

In short, MMFF94 is one of the most successful force field methods applied for the structural predictions of large molecules, and it still remains the most popular method for structural and conformational analysis of new molecules at the initial stage of synthesis due to its high computational speed and fairly accurate structural prediction [121]. The applications of MMFF94 ranges from complex drug design simulations to homogenous transition metal catalysis [105-109].

3.3. Conformational Searching

Almost all the existing molecules or compounds of interest to biochemists, as well as organic, pharmaceutical and medicinal chemists are very flexible in nature and can adopt more than one structure or conformer at one time depending on the conformational stability in the potential energy surface [121, 135]. Stable structures of a molecule are generally corresponding to the local minima positions of the structure in the potential energy surface. This potential energy surface is a function that relates the energy of any given molecule and its structure. Hence, as the conformation of the molecule changes in the PES, the energy associated with it also changes [125, 139].

Identifying the lowest energy conformer in the potential energy surface is a very critical process for any kind of chemical process or reaction. This is due to the fact that the conformation of any given molecule always dictates or directs the molecular size and shape of the molecule. As a result, it influences the overall molecular properties of the compound, like chemical reactivity and so on [121]. Therefore, it is vital to know the

conformation of the compound to predict its applications and other chemical and physical features. However, one of the most difficult tasks of the conformational searching is the lack of experimental data and the associated high computational cost as a result of the huge number of conformers to be evaluated to find the most stable conformer at the PES [110]. Moreover, identifying all the conformers in the PES and locating them is the most cumbersome task in the conformational searching research via the existing conformational searching tools, due to the fact that missing the most stable structure could lead to incorrect results [103, 110, 121].

There are plenty of conformational search energy methods nowadays, that generally fall into three categories, (i) systematic searching, (ii) Monte Carlo, and (iii) Genetic algorithms. In this research the first two methods are used, as each has its own advantage and disadvantage. The following sub-sections will introduce these methods in detail [110, 121].

3.3.1. Systematic Searching

Systematic conformational searching is one of the most powerful tools for the exploration of the potential energy surface [103, 121]. It always guarantees that no space was “*overlooked*” in the search for finding all the possible local minima conformers in the PES [136, 137]. The method is performed by varying systematically all the available dihedral angles of a molecule so that to generate all the probable conformations of the molecule. This means that when the torsion angles of the molecule are varied systematically with a fixed incremental value (e.g. 30°) over a range of 360°, a full and

complete image of the conformational energy landscape of any given molecule can be generated and examined [103, 121, 136, 137].

Mathematically, the number of generated conformations by the systematic search depends not only on the angle increment (M), but also on the number of rotatable bonds (N). Suppose we have a number of N rotatable bonds (or degrees of freedom) in the molecule and the angle increment is given by M. The theoretical number of possible generated conformations of any given molecule can be given by [110, 121]:

$$\text{Number of conformers} = \left(\frac{360}{M} \right)^N \quad (3.16)$$

As one can see from equation 3.16, the number of conformations generated by using the systematic search is directly correlated with M and N [136]. For instance, a systematic conformational search on a molecule with six rotatable bonds and an angle increment of 30° will generate 2,985,984 structures or conformers [121], which is an extremely large number of structures that cannot be easily handled and investigated at all via the existing supercomputing tools [121, 137]. This approach is not reasonable at all for proteins and peptides which contain a large number of rotatable bonds [121, 137]

Moreover, this method also suffers a lot in the application for cyclic molecules. This is due to the fact that its searching task is performed by sequentially bending the ring in the molecules which then leads to an improper searching for conformers since the search does not account for the entire rotatable bonds [121]. Therefore, the application of this method is very limited and applicable to small size and acyclic molecules only. To

examine the conformational behaviour of large molecules requires using random searching techniques like molecular dynamics, Monte Carlo, genetic algorithm or distance geometry. The latter method is based on interatomic distances allied to various mathematical procedures to generate structures for energy minimization [137].

3.3.2. Monte Carlo Searching

As we have already indicated in section 3.3.1, for large molecules a number of random conformational search methods have been formulated and implemented to overcome the drawbacks of the systematic searching methods [103, 110]. A basic concept of these methods is their random-based iterative procedure [136]. In this case, a hypothetical conformation of a molecule is chosen from a set of previously generated conformations arbitrarily and then subjected to modification and minimization [136]. In this regard, if the newly generated conformation of the molecule is not in the list of the conformers in the data set generated previously, then, the new conformer or structure obtained is added to the list of structures in a library, if not, the conformer will be rejected. This process of continued iteration is then repeated over and over again until either a pre-defined number of iterations have been performed (in our case 100 iterations) or until no new conformation is found or the desired number of conformations are already collected [121, 136].

The main difference between the systematic search methods and the random search is that, unlike the systematic search methods in which there is an upper limit to the conformational sampling of the molecule, in random search methods there is no usual termination point for its search of a conformer [121, 136, 137]. The upper limit for its

iteration is infinite [121, 136, 137]. Although all the existing random search methods share the same basic concept, they are completely different in the way the conformation is modified and selected for the next iteration process [136, 137].

In the Monte Carlo simulation technique, the selection criteria for the next iterative step are based on the conformational energy characterizations. The procedure is as follows, according to Leach and Gillet [137]: If the energy of the new conformer, V_{new} , is lower than the energy of the previous conformer, V_{old} , then the new conformer is used as the starting structure for the next iterative process in the procedure. However, if the V_{new} is greater than V_{old} then a Boltzmann factor which is given by equation 3.17 is applied:

$$\exp[- (V_{new} - V_{old}) / KT] \quad (3.17)$$

where k , is the Boltzmann constant, and T is the temperature of the system in Kelvin. If the calculated Boltzmann factor given in equation 3.16 is greater than a random number given between 0 and 1 then the new conformer generated via the Monte Carlo is retained, and if it is less than then the previous structure is used [136, 137].

There are numerous advantages to the Monte Carlo searching methods. Some of the most important advantages are, (i) all the possible higher energy conformers are sampled effectively (ii) it randomly samples all the conformational space, although it does not necessarily look everywhere in the conformational space, with less time relative to the systematic sampling; (iii) it is a very good method for cyclic and large molecules; and (iv) it is good because the global minimum configuration of a molecule may be very

different from the initial conformation. The biggest disadvantages of this method are that it does not guarantee that all the possible local minima or the global energy minimum will be found as the search technique method is based on random sampling framework and it might miss the structure(s) of interest [110, 121, 136, 137].

3.4. Density Functional Theory

A few years ago a remarkable theory, density functional theory (DFT), was developed. It is characterized by expressing the electronic properties as a function of the electron density. Since its inception this theory has gained a tremendous amount of support and fame for its practicality, reproducing high quality experimental results with low computational cost [116, 121]. This method is a variational method that is currently the most flourishing approach to calculate the electronic structure of matter [116, 138].

3.4.1. Mathematical Formulation

In DFT, the complicated many-electron wave function $\Psi_i(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)$ and the related Schrödinger equation are replaced by the much simpler electron density $\rho(\vec{r})$ and its related calculational framework [116, 138]. The electron density $\rho(\vec{r})$ is defined as:

$$\rho(\vec{r}) = N \int \dots \int \left| \Psi_i(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) \right|^2 ds_1 d\vec{x}_2 \dots d\vec{x}_N \quad (3.18)$$

This is the probability of finding any of the N electrons within the volume element $d\vec{r}_1$.

This leads to the following fundamental properties of the density:

$$\rho(\vec{r} \rightarrow \infty) = 0 \quad (3.19)$$

$$\int \rho(\vec{r}) d\vec{r} = N \quad (3.20)$$

The most interesting aspect of this approach is that the density $\rho(\vec{r})$ is a non-negative function with only three spatial variables that vanish at infinity and integrate to the total number of electrons according to equations 3.19 and 3.20 [116, 141].

Although, the concept of DFT has a long history which dates back to the 1920s, beginning with the works of Thomas and Fermi [139, 140], it was not until 1964 that the DFT found its wide spread theoretical foundation in the theorems of Hohenberg and Kohn [116, 141]. In their work they formulated and proved two fundamental theorems which helped to lay the foundation for the modern density functional theory. The first theorem states that: *“the external potential $V_{ext}(\vec{r})$ is (to within a constant) a unique functional of the $\rho(\vec{r})$; since, in turn $V_{ext}(\vec{r})$ fixes \hat{H} we see that the full many particle ground state is a unique functional of $\rho(\vec{r})$ ”*. This is given by equation 3.21: [116, 138, 141].

$$V_{ext}(\vec{r}) \Leftrightarrow \rho_0 \quad (3.21)$$

The second Hohenberg-Kohn theorem, which is formulated to answer questions about finding or calculating the real ground state density, states that for any given density $\tilde{\rho}$ which is associated with the N electron system in the external potential, the ground state energy is given by [116, 134]:

$$E_0 \leq E[\tilde{\rho}] = T[\tilde{\rho}] + E_{Ne}[\tilde{\rho}] + E_{ee}[\tilde{\rho}] \quad (3.22)$$

where E_0 is the ground state energy, $T[\tilde{\rho}]$ is the kinetic energy of the density $\tilde{\rho}$, $E_{ee}[\tilde{\rho}]$ is the electron-electron interaction and $E_{Ne}[\tilde{\rho}]$ is the external potential present in the system. After these groundbreaking theorem contributions by Hohenberg and Kohn in the treatment of the ground state density, Kohn and Sham, in 1965, paved an avenue for the practical application of the density functional theory [116, 142]. In the Kohn-Sham formalism, the ground-state electronic energy in any system can be written as:

$$E[\rho(\vec{r})] = T_s[\rho] + J[\rho] + E_{xc}[\rho] + E_{Ne}[\rho] \quad (3.23)$$

$$\begin{aligned} E[\rho(\vec{r})] = & -\frac{1}{2} \sum_i^N \langle \varphi_i | \nabla^2 | \varphi_i \rangle + \frac{1}{2} \sum_i^N \sum_j^N \iint \left| \varphi_i(\vec{r}) \right|^2 \frac{1}{r_{12}} \left| \varphi_j(\vec{r}_2) \right|^2 d\vec{r}_1 d\vec{r}_2 \\ & + E_{xc}[\rho(\vec{r})] - \sum_i^N \int \sum_A^M \frac{Z_A}{r_{1A}} \left| \varphi_i(\vec{r}_1) \right|^2 d\vec{r}_1 \end{aligned} \quad (3.24)$$

$$\rho_s(\vec{r}) = \sum_i^N \sum_s \left| \varphi_i(r, s) \right|^2 = \rho_0(\vec{r}) \quad (3.25)$$

Where the φ_i are the Kohn-Sham orbitals, Z_A is a nuclear charge, r_{1A} is the distance between the electron and nucleus, and r_{12} is the distance between two electrons. $J[\rho]$ is the classical Coulomb interaction of the system. The term $E_{xc}(\rho)$, which is called the exchange-correlation energy, contains the difference between the real kinetic energy and

that of the noninteracting electrons, as well as the two-body exchange and dynamic correlation contributions. This exchange-correlation energy is given by [116, 138]:

$$E_{xc}[\rho] = T_c[\rho] + E_{nc}[\rho] \quad (3.26)$$

Where T_c is the residual part of the real kinetic energy. Therefore, applying the variational principle (Eq. 3.21) to equation 3.22 results in a set of one-electron K-S equation [134]. They are given by [134]:

$$\left(-\frac{1}{2}\nabla^2 + \left[\int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2 + V_{xc}(\vec{r}) - \sum_A^M \frac{Z_A}{r_{1A}} \right] \right) \varphi_i$$

$$\left(-\frac{1}{2}\nabla^2 + V_{eff}(\vec{r}_1) \right) \varphi_i = \varepsilon_i \varphi_i \quad (3.27)$$

In this case, V_{xc} is the functional derivative of the E_{xc} , and V_{eff} contains the potential energy contributions from all the various terms in equation 3.23, except for the kinetic energy contribution. The Kohn-Sham equations are exact for any electronic systems at least *in principle*. The exact functional $E_{xc}(\rho)$ is not known at all. Therefore, an approximate exchange-correlation functional must always be used in solving the Kohn-Sham equations. This is why different models are used for the exchange-correlation functional, leading to different levels of theory in density functional theory modeling [116, 121, 138]. There are plenty of approximations for the $E_{xc}(\rho)$ used currently, like the general gradient approximations or GGA, the local density approximation (LDA), or the hybrid functionals (B3LYP, B97, B1, B3 etc).

3.4.2. Hybrid Functionals

These are the most popular and widely used models of approximation to the exchange-correlation energy functional, within the density functional theory scheme. The fundamental concept behind these models is the incorporation or mix of the exact Hartree-Fock exchange (typically 20-25% exact exchange) together with some of the other approximate functional like LDA or GGA.

The Kohn-Sham calculations which are expressed in terms of Kohn-Sham orbitals, are performed in a similar manner of that the Hartree-Fock computations [116].

The most often cited and most popular exchange-correlation functional, that is often used for organic and bio-organic molecular modeling, is a hybrid functional called B3LYP [116, 143]. The B3LYP exchange-correlation energy expression is given by [142]:

$$E_{XC}^{B3LYP} = (1-a)E_X^{LSD} + aE_{XC}^{\lambda=0} + bE_X^{B88} + E_C^{LYP} + (1-c)E_C^{LSD} \quad (3.28)$$

Where the a, b, and c are three empirical parameters due to Becke (B3), LSD is the local spin density approximation, E_X is energy exchange, and E_C is the energy correlation, LYP [143-145] is a Lee, Yang, and Parr correlation functional and B88 [146] is the Becke 1988 exchange functional. λ is the coupling strength parameter. Other hybrid functionals can be assembled in a similar manner by only changing the terms or functional components of the correlation energy. Some of the most widely used hybrid functionals are PBE1PBE, B3PW91 etc [116, 143].

3.5 Solvation Model

Most biological and bio-organic phenomena of any given system take place in solution, and therefore, modeling of the solvent together with the solute is a central aspect in particular in computing protein and peptide conformational structures and energetics [110, 144]. The determination of equilibrium bond lengths and angles, the transition state and other electrostatic interactions can greatly change in solutions, and thereby change the entire conformational free energies of the system. This could hinder our ability to design and synthesize a bioactive or effective molecule with valuable structures [110, 121, 143].

However, regardless of its significance, the computations of the solvation free energy pose a significant problem to most theoretical and computational experts in a number of fields of research. One of the main reasons is that although computing has been established to produce acceptable solvation free energy differences for a number of small molecules, in particular for peptides, most of these existing computational solvation methods are insurmountable for large molecules due to high computational cost.

There are a number of solvation models currently available like COSMO, CPCM and PCM approaches, but in this research only the polarizable continuum solvation model (PCM, see figure 3.1) is used [110, 143]. This solvation method utilizes or defines the solvent-excluded cavity as the union of a series of interlocking *van der Waals* atomic spheres and then calculates all the effects of the polarization of the solvent by numerically integration [110, 143].

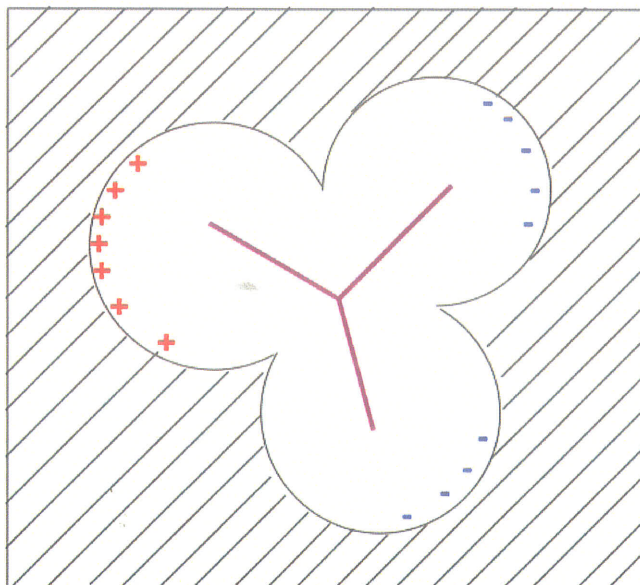


Figure 3.1 The PCM model. *Copyright permission was obtained on May 27, 2009 from Georg Schreckenbach (GS private communications)*

In the polarizable continuum model, the solvent is represented as a continuous dielectric without discrete internal structure and is described by its dielectric constant, ϵ . The dielectric is polarized when perturbed by the solute charge distribution and creating a reaction field on the cavity of the surface. The generated reaction field in turn interact with the solute in the defined cavity via electrostatics leading to net stabilization of the system [110]. The dielectric constant (ϵ) of water is 78. In solvation, it is to be noted that when a molecule is passed from the gas-phase into the solution, there is always a dipole moment effect as the solvent molecules start to re-orient themselves, thereby, can shift the gas-phase minimum energetics of the solution. This shift most of the time depends on its strength of the dipole moment.

3.6. Computational Methodology

Two programs were used for the entire calculations. The SPARTAN '02 [147] program was used for the initial empirical energy calculations to generate a number of trial structures or conformers using the MMFF94 force field [105-109], and Monte-Carlo or random searching methods for sampling the conformational space. This random conformational search was started with an initial energy minimized guess structure drawing in SPARTAN and then followed by the Monte-Carlo searching method. However, although Monte-Carlo does potentially sample out all the region of the conformational space fairly, one cannot be certain or guaranteed to locate all the lowest energy conformers in the conformational space. As a result of this limitation of the random sampling nature some very important conformers might be missed during the search. But to avoid this inherent problem we adopted a strategy of searching the conformational space by re-starting the search with different initial conformations and check for the completeness of the analysis. Once a redundant outcome is observed we can assume that the conformational space is fully covered and the desired conformations are gathered. Additionally, these initial conformations were also obtained using a build and search approach.

In this build and search approach, the searching started with a small molecule with few degrees of freedom. Its conformational space was searched using Systematic and Monte-Carlo methods. This searching was repeated five times to account for its completeness. A number of unique conformers or structures were identified for this molecule. Then we used these conformers as starting structures and built on them the next bigger molecule.

The same was then performed until we reached the desired molecule. During this course of conformational search a number of structures were identified as redundant structures but by superimposing them, only few unique structures were selected as the most stable conformers and the remaining redundant conformers were discarded.

In the build and search approach, for example Compound 1 and Compound 2 were started with the simple five-membered ring i.e. CH₂-CH₂-CH₂-CH₂-NH. A systematic and Monte-Carlo search was employed that generated 12 unique structures. Once these unique structures were identified, a six-membered ring (i.e. CH₂-CH₂-CH₂-CH₂-CH₂-O) was attached to each conformer in a Spirocyclic manner or some other way, depending on the structure of the molecule. These systems were again subjected to Monte-Carlo conformational search which generated around 27 unique local minima conformers. Finally the final compound (e.g. Compound 1) were built on each of them and subjected to further conformational search. At last a set of trial structures were gathered by superimposing them for each compound. Once a number of unique local minima conformers were obtained using the MMFF94 force field, all the remaining analysis of the conformers was performed using the GAUSSIAN 03 [148] program package. In this study, all the calculations were performed using DFT. On the basis of previous experience and literature reviews conducted, we selected the B3LYP level of theory, as it was good enough to provide an accurate prediction of the molecular structure, and the 6-31+G(d, p) basis set [121], which is a large basis set augmented by diffuse and polarization functions to account for the relatively large molecules.

Chapter 4

Conformational Preference of Spirocyclic Carbohydrate-Template Proline Analogues

4.1. Results and Discussion

In this chapter, a detailed structural assignment and analysis of the conformational distributions of Compound 1 (Figure 2.6) and Compound 2 (Figure 2.7) in gas and water are presented. As mentioned earlier in section 3.6, the build and search methodology was applied to generate a series of trial structures by allowing for all combinations of the internal single-bond rotamers, leading to a total of 5400 possible structures for both compounds (each 2700 structures). These trial structures were superimposed and characterized using their energies and a set of unique conformers were located. A total of 443 unique structures for Compound 1 and 457 structures for Compound 2 were collected.

Density Functional Theory (DFT) was used for the entire conformational searching calculations of these two compounds at the B3LYP/6-31+G (d, p) level of theory in both gas phase and water. First, these unique trial geometries were fully optimized at the B3LYP/6-31+G (d, p) level of theory with the GAUSSIAN 03 program package in the gas phase. All of these initial trial geometries were optimized in the gas phase without any difficulty. The zero-point energies and harmonic frequencies of these structures in gas phase were subsequently calculated in order to characterize all the stationary points

as real local minima. Except for four structures of Compound 2, no imaginary frequencies were observed in their frequency calculations for all structures of Compound 1 and Compound 2. This indicated that all these optimized structures were considered to be true local minima conformers in the gas phase. Once these gas-phase structures were identified, the conformers were further subjected to solvation to determine or quantify the extent of deviation from the gas phase predictions. After being re-optimized in H₂O using the same level of theory as for the gas phase only 355 and 258 conformers for Compound 1 and 2, respectively, were located as true local minima. The remaining structures had either big imaginary frequencies, which were very difficult to avoid, or having some sort of difficulty with convergence. Those structures with convergence problems were checked for their SCF convergence profile. In particular, we examined the energy change for each iterative step, and it was found that there was no dramatic change in their energy profile from one step to the next. In addition, these structures were at a very high energy level relative to the conformers already collected or identified. As a result, their impacts towards the conformational distribution calculation were considered to be negligible.

Finally, the percentage of conformers and cis/trans ratio of the local minima conformers at the equilibrium was calculated using the Boltzmann distribution (or Boltzmann statistics) for the entire set of conformers regardless of their energy level for both gaseous and water phases. In analyzing the structural characteristics, all the existing hydrogen bonds and the dipole moments of the most stable structures were also identified and characterized for some of the most stable conformers for both Compound 1 and Compound 2. In particular, the H-bonding features of the most stable conformers were analyzed in greater detail to try to understand the effect of the intramolecular hydrogen

analyzed in greater detail to try to understand the effect of the intramolecular hydrogen bond-controlled prolyl amide isomerization in glucosyl 3(S)-hydroxyproline hybrids, with special emphasis on the influence of the C_δ -hydroxy-methylene substituent on the kinetics of prolyl amide cis/trans isomerization. The backbone torsion and endocyclic angles were also characterized to determine the impact of C_δ -hydroxy-methylene on the peptide backbone chain and prolyl amide. The definitions of the backbone torsion and endocyclic angles are displayed in Figure 4.1. These definitions are similar to those used in the work of Song and Kang [25, 56].

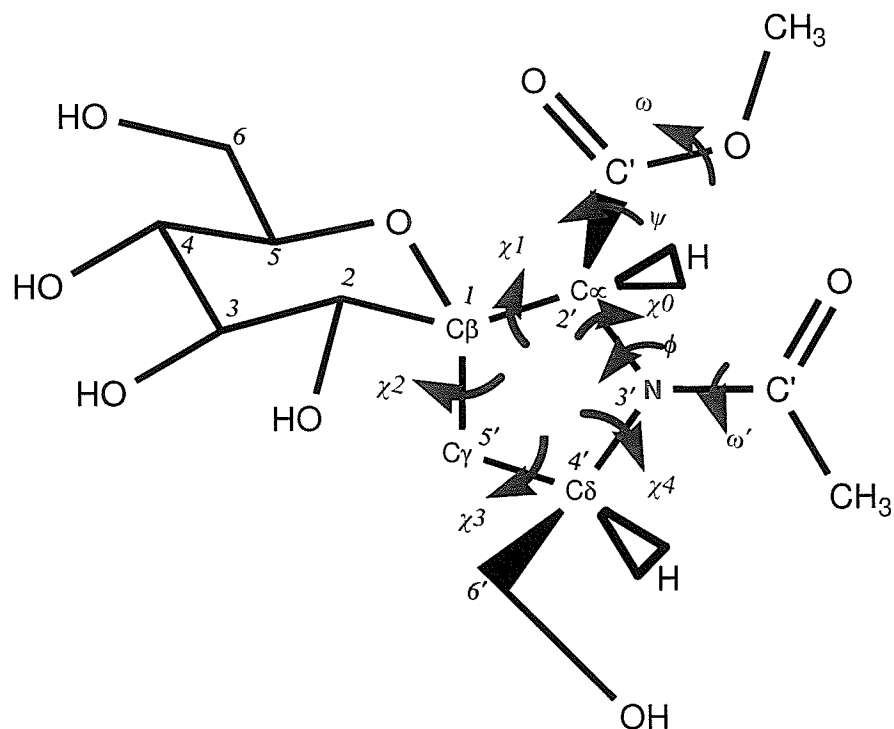


Figure 4.1 The definitions of the backbone and endocyclic torsion angles for Compound 1 and Compound 2 are as follows. Prolyl Carbon Alpha (C_α), Prolyl Carbon beta (C_β), Prolyl Carbon Gamma (C_γ), Prolyl Carbon Delta (C_δ) and Prolyl N-terminal Prolyl

carbon-prolyl N-terminal prolyl nitrogen-prolyl carbon alpha ($C-C'-N-C_\alpha$), ϕ = prolyl N-terminal carbonyl carbon-prolyl N-terminal prolyl nitrogen-prolyl carbon alpha-prolyl carbonyl carbon ($C'-N-C_\alpha-C'$), ψ = prolyl N-terminal prolyl nitrogen-prolyl carbon alpha-prolyl carbonyl carbon-prolyl ester oxygen ($N-C_\alpha-C'-O$), ω = prolyl carbon alpha-prolyl carbonyl carbon-prolyl ester oxygen-prolyl ester methyl group ($C_\alpha-C'-O-C$) for the backbone torsion angles, and $\chi_0 = C_\delta-3'N-C_\alpha-C_\beta$, $\chi_1 = 3'N-C_\alpha-C_\beta-C_\gamma$, $\chi_2 = C_\alpha-C_\beta-C_\gamma-C_\delta$, $\chi_3 = C_\beta-C_\gamma-C_\delta-3'N$, $\chi_4 = C_\gamma-C_\delta-3'N-C_\alpha$ for the endocyclic torsion angles.

4.1.1 Conformers and energies in gas and water

Table 4.1 and Table 4.2 (for Compound 1) and Table 4.5 and Table 4.6 (for Compound 2) show the relative energies, torsion and endocyclic angles, and the conformational distributions of the most stable structures of Compound 1 and 2 in gas and water, respectively. Although the population distributions were calculated for the entire sets of true local minima conformers, only those with a significant population contribution were selected. For the purpose of comparison, only that structures with a population distribution of > 0.5% were chosen in this discussion.

The conformers listed in Table 4.1, Table 4.2, Table 4.5 and Table 4.6 revealed that the conformation of the six-membered sugar residue is in a *chair* conformation. This has been confirmed by qualitatively examining the optimized structures of each conformer in the entire distribution. Because the hydroxyl groups in the sugar residue possess different orientations, two structures having the same backbone, endocyclic and sugar

conformations with different hydroxyl group orientations showed a different relative energy values.

To narrow the scope and analysis of this work, the comparisons of these conformers are made based on the classification of the backbone torsion angle, endocyclic torsion angle, cis-trans prolyl amide bond and the sugar residue conformation as an overall criteria to categorize structures into sub groups.

Compound 1. The local minima conformers and the geometric characterization of Compound 1 optimized at the B3LYP/6-31+G (d, p) level of theory shows that nine local minima $t-\epsilon_L[d]$, $t-\alpha_L[d]$, $t-\epsilon_L[u]$, $t-\epsilon_L[d]^*$, $c-\epsilon_L[u]$, $c-\alpha_L[d]$, $c-\epsilon_L[d]$, $c-\alpha_L[d]^*$, and $c-\epsilon_L[d]^*$ conformers in gas phase (Table 4.1), and four local minima $t-\epsilon_L[d]$, $t-\epsilon_L[d]$, $c-\epsilon_L[d]$ and $c-\alpha_L[d]$ conformers (Table 4.2) in *water* were identified, which are shown in Figures 4.1 and 4.2, respectively. The $[d]^*$ in the gas phase were introduced because these conformers resemble (mostly) the down puckering conformer but do not completely satisfy the criteria for down puckering definition, which is characterized by positive χ_1 and χ_3 , and negative χ_2 and χ_4 . For this reason, these conformers are considered in this research to be completely down puckering unless otherwise indicated.

Considering both the trans and cis peptide bonds in the gas phase, $c-\epsilon_L[u]$ is the most stable conformer and $t-\epsilon_L[u]$ is the least stable conformer for Compound 1. The relative stabilities of the conformations in the gas phase follows the order of $c-\epsilon_L[u] < c-\alpha_L[d] < c-\epsilon_L[d] < t-\epsilon_L[d] < t-\alpha_L[d] < t-\epsilon_L[u]$. The most stable conformer $c-\epsilon_L[u]$ is characterized by a cis prolyl amide bond, the ϵ_L backbone conformation of which leads to polyproline II (P_{II}), and “UP” puckering, which is C_γ -exo. It should be noted that the

relative energy difference between c- ϵ_L [u], c- α_L [d] and, c- ϵ_L [d] is very low (less than 0.1 kcal/mole) relative to the energy difference between t- ϵ_L [d], t- α_L [d] and t- ϵ_L [u] which is approximately 0.33 kcal/mol. This could be because the hydrogen bonding interactions are enhancing to the cis structures more than the trans structures. However, the corresponding relative energy difference between the most stable cis and trans conformers of Compound 1 is approximately 0.36 kcal/mol.

For Compound 1 in *water*, t- ϵ_L [d] is the most stable conformer according to the population distribution in Table 4.2, and the relative stabilities of the conformations are calculated in the order t- ϵ_L [d] < t- α_L [d] < c- ϵ_L [d] < c- α_L [d]. Conformer t- ϵ_L [d], which is the lowest-energy structure of Compound 1 in water, is characterized by a trans prolyl amide bond, the ϵ_L backbone conformation which leads to polyproline II (P_{II}), and the “DOWN” puckering (C_γ-endo). The relative energy difference between t- ϵ_L [d] and t- α_L [d] is about 0.12 kcal/mol, while there is a gap of 0.14 kcal/mol between c- ϵ_L [d] and c- α_L [d]. This shows that in water, unlike in the gas phase, the relative energy differences among cis and among trans conformers are almost identical. The relative energy difference between the most stable cis and trans conformers in water is about 0.54 kcal/mol. Interestingly, the most stable conformers in the gas phase, which are part the cis population, are not the most stable structures in water anymore; the whole scenario is reversed in water. The higher energy trans conformers in gas phase turn out to be the most stable conformers in water while the most stable cis conformers in gas phase became higher energy conformers in water. This trend is shown for almost all compounds. This could be due to the dipole moment effect, which helps in stabilizing the trans conformers or the effect of the intramolecular hydrogen bond. Furthermore, the

“UP” puckering conformer which is the most stable in the gas phase does even exist in water. All the conformers whether it is trans or cis shows a down puckering or C_γ -endo configuration.

For the gas-phase computed *cis*-conformations of Compound 1 listed in Table 4.1, the backbone torsion angles of ω' , ϕ , ψ , and ω generally fall into the following ranges, except for very few structures that do not abide by this generality. The ω' value ranges from $-7^\circ \leq \omega' \leq -1^\circ$, ϕ ranges from $-76^\circ \leq \phi \leq -53^\circ$, ψ falls into two different regions i.e. $139^\circ \leq \psi \leq 154^\circ$ and $-30^\circ \leq \psi \leq -12^\circ$, and the ω value falls in the regions of $\omega \approx 177^\circ$ and $\omega \approx -177^\circ$. The endocyclic torsion angles show spans as follows. The χ_0 ranges from $-28^\circ \leq \chi_0 \leq 0^\circ$, χ_1 ranges from $-19^\circ \leq \chi_1 \leq 36^\circ$, χ_2 ranges from $-35^\circ \leq \chi_2 \leq 30^\circ$, χ_3 ranges from $-30^\circ \leq \chi_3 \leq 23^\circ$, and χ_4 ranges from $-2^\circ \leq \chi_4 \leq 19^\circ$.

For the trans-conformations of Compound 1 in gas phase, the value of the backbone torsion angles of ω' , ϕ , ψ and ω , the following ranges have been found: The ω' value falls into two different regions $-179^\circ \leq \omega' \leq -153^\circ$ and $\omega' \approx 177^\circ$, ϕ falls in the range of $-68^\circ \leq \phi \leq -41^\circ$, ψ falls in regions $127^\circ \leq \psi \leq 143^\circ$ and $\psi \approx -16^\circ$, and finally the ω value falls at two distinct angles $\omega \approx -177^\circ$ and $\omega \approx 179^\circ$. The endocyclic torsion angles were also obtained in the ranges of, $-12^\circ \leq \chi_0 \leq 24^\circ$ for χ_0 , $-3^\circ \leq \chi_1 \leq 32^\circ$ for the χ_1 , $-17^\circ \leq \chi_2 \leq -38^\circ$ for χ_2 , $-31^\circ \leq \chi_3 \leq 37^\circ$ for χ_3 , and finally the χ_4 ranges from $-34^\circ \leq \chi_4 \leq 15^\circ$.

Table 4.1 Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of Compound 1, which are Optimized at the B3LYP/6-31+G(d, p) Level of Theory in the Gas Phase.

B3LYP/6-31+G(d, p)				Backbone Torsion Angles and Endocyclic Torsion Angles ^a								
Conformers		Gas (in kcal/mole)	Conformer Distribution	ω	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^0
		ΔE										
1	c- ϵ_L [u]	0.0 ^b	7.1266848039	-6.815	-54.616	139.747	177.322	-18.052	30.440	-30.672	19.328	-0.765
2	c- ϵ_L [u]	0.107391976	5.9451195845	-6.685	-53.85	140.101	177.149	-19.128	31.139	-30.728	18.636	0.365
3	c- α_L [d]	0.136803346	5.6571760562	-1.541	-72.084	-30.970	-176.183	32.976	-34.921	22.896	-1.370	-20.17
4	c- α_L [d]	0.136910023	5.6561574714	-1.514	-72.042	-30.949	-176.190	32.964	-34.941	22.939	-1.423	-20.128
5	c- ϵ_L [d]	0.165687608	5.3879704714	-1.889	-73.816	154.022	175.084	34.988	-35.965	22.702	0.239	-22.504
6	c- ϵ_L [d]*	0.165725259	5.3876280581	-1.904	-73.84	154.016	175.084	34.994	-35.952	22.676	0.272	-22.53
7	c- ϵ_L [d]*	0.165750359	5.3873997947	-1.916	-73.825	154.023	175.081	34.999	-35.953	22.672	0.28	-22.538
8	c- α_L [d]*	0.182975495	5.2330127803	-4.859	-76.399	-12.532	-179.587	36.82	-35.534	20.454	3.892	-26.045
9	c- α_L [d]*	0.265806749	4.5501930566	-4.294	-75.777	-14.334	-179.506	36.402	-35.605	20.959	3.098	-25.27
10	t- ϵ_L [d]	0.36194748	3.8685742309	179.242	-45.116	127.662	-176.788	8.587	-28.554	37.405	-33.449	15.748
11	t- ϵ_L [d]	0.361953755	3.8685332547	179.241	-45.114	127.665	-176.789	8.581	-28.547	37.401	-33.449	15.751
12	c- α_L [d]*	0.400112607	3.6272110513	-6.56	-75.596	-12.824	-179.737	36.462	-33.077	16.942	7.346	-28.049
13	t- α_L [d]	0.63313826	2.4476328788	-176.216	-48.366	129.429	-177.104	3.738	-22.055	31.661	-30.735	17.308

14	c- α_L [d]	0.636244432	2.4348331415	-2.302	-71.545	-29.636	-176.589	32.769	-34.174	21.95	-0.482	-20.628
15	c- α_L [d]	0.636269533	2.4347299823	-2.279	-71.596	-29.645	-176.587	32.767	-34.181	21.961	-0.494	-20.619
16	t- ε_L [d]	0.694728318	2.2059507499	-178.192	-46.217	129.003	-176.609	1.262	-20.245	31.217	-32.089	19.743
17	c- ε_L [d]*	0.741402474	2.0388247690	-2.363	-73.269	154.227	176.071	35.228	-35.884	22.325	0.813	-23.049
18	c- ε_L [d]*	0.741440125	2.0386951987	-2.352	-73.202	154.236	176.074	35.213	-35.89	22.349	0.777	-23.017
19	c- ε_L [d]*	0.741502876	2.0384792666	-2.312	-73.234	154.222	176.09	35.201	-35.898	22.372	0.745	-22.989
20	t- ε_L [u]	0.96209129	1.4047441983	-165.028	-54.953	143.067	-178.556	-24.541	34.299	-31.14	15.107	6.249
21	t- ε_L [d]*	0.962141491	1.4046251686	176.451	-41.724	128.623	-177.185	-1.053	-18.937	31.588	-34.097	22.37
22	t- ε_L [d]*	1.190354146	0.9555693539	176.366	-41.628	128.75	-176.97	-2.208	-17.656	30.659	-33.905	23.001
23	t- ε_L [d]	1.190360421	0.9555592322	175.091	-71.395	143.561	179.944	31.322	-38.638	31.37	-11.358	-12.933
24	c- α_L [d]*	1.304140444	0.7885835332	1.368	-81.028	-12.801	-179.667	36.604	-35.588	20.951	3.252	-25.518
25	t- ε_L [d]	1.306236325	0.7857986062	177.179	-44.511	129.259	-177.544	4.278	-21.286	29.964	-28.74	15.541
26	t- ε_L [d]*	1.316201176	0.7726916462	176.682	-41.839	129.075	-177.088	-3.542	-16.833	30.687	-34.846	24.426
27	t- ε_L [d]*	1.316320403	0.7725361557	-153.304	-41.863	129.074	-177.102	-3.443	-16.917	30.723	-34.815	24.345
28	t- ε_L [d]*	1.316370604	0.7724706954	176.647	-41.806	129.056	-177.081	-3.429	-16.923	30.719	-34.802	24.327
29	t- α_L [d]	1.4713529	0.5946583525	179.824	-68.281	-16.429	176.86	32.84	-37.89	28.228	-7.113	-16.442
Others*		< 0.5										

Total Population Distribution

	Gas (%)	Exp. H ₂ O (%) [101]
Total Cis Isomers	74.92	24
Total Trans Isomers	25.08	76

^a Defined in Figure 4.1; angles are in degrees, ^b $E = -1279.132219540$ a.u. The population distributions were calculated using the Boltzmann statistical weights at 25 °C.

Table 4.2 Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of Compound 1, which are Optimized at the B3LYP/6-31+G(d, p) Level of Theory in the Water.

B3LYP/6-31+G(d, p)				Backbone Torsion Angles and Endocyclic Torsion Angles ^a								
Conformers		H ₂ O (in kcal/mole)	Conformer Distribution	ω	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^0
		ΔE										
1	t- ϵ_L [d]	0.0 ^b	4.70017076	175.253	-64.198	153.736	177.502	28.915	-34.773	26.879	-8.378	-13.164
2	t- ϵ_L [d]	0.053338308	4.29548658	174.849	-63.973	153.458	177.498	29.437	-34.414	25.781	-6.852	-14.482
3	t- α_L [d]	0.121736843	3.82710817	175.55	-61.41	-28.681	-179.498	27.616	-33.469	26.027	-8.356	-12.356
4	t- α_L [d]	0.277359199	2.94297812	175.706	-61.659	-28.731	-179.405	28.154	-34.144	26.603	-8.633	-12.497
5	t- α_L [d]	0.352660339	2.59170305	175.751	-61.884	-28.605	-179.79	28.405	-34.608	27.1	-8.983	-12.439
6	t- α_L [d]	0.358935434	2.56439600	175.486	-61.371	-28.421	-178.709	27.881	-33.682	26.107	-8.272	-12.569
7	t- ϵ_L [d]	0.360190453	2.55896921	174.313	-64.422	153.418	177.149	30.688	-34.505	24.71	-4.861	-16.547
8	t- ϵ_L [d]	0.384663324	2.45541259	174.809	-63.748	153.48	177.554	29.199	-34.614	26.354	-7.622	-13.835
9	t- α_L [d]	0.454316878	2.18304566	176.059	-61.619	-29.02	-179.011	27.947	-34.48	27.349	-9.547	-11.775
10	t- ϵ_L [d]	0.545933265	1.87025420	177.627	-67.256	154.625	176.597	27.487	-33.487	26.314	-8.613	-12.088
11	c- ϵ_L [d]	0.546560775	1.86827423	-10.903	-63.708	154.37	177.909	31.884	-34.728	23.936	-3.193	-18.36
12	t- ϵ_L [d]	0.563503531	1.81560020	174.055	-64.29	153.729	177.029	30.79	-34.36	24.383	-4.444	-16.879
13	t- α_L [d]	0.595506515	1.72012283	175.399	-61.183	-29.02	-179.311	28.093	-33.767	26.084	-8.126	-12.779

14	c- ε_L [d]	0.633157086	1.61420453	-10.713	-64.058	153.755	177.811	31.792	-34.917	24.354	-3.726	-17.943
15	c- ε_L [d]	0.636922143	1.60397824	-10.972	-63.957	154.078	177.621	31.89	-35.01	24.411	-3.721	-18.01
16	t- ε_L [d]	0.659512485	1.54396672	177.548	-66.781	154.711	177.005	27.948	-33.772	26.306	-8.315	-12.573
17	c- α_L [d]	0.689632941	1.46742941	-10.68	-62.642	-27.379	-177.946	30.963	-34.117	23.815	-3.727	-17.428
18	c- ε_L [d]	0.758031476	1.30742140	-10.749	-63.601	153.891	177.725	31.447	-34.857	24.607	-4.229	-17.401
19	t- α_L [d]	0.771836685	1.27730691	175.365	-61.205	-29.374	-178.796	28.155	-33.865	26.186	-8.187	-12.779
20	t- α_L [d]	0.827057521	1.16362742	175.163	-61.389	-28.169	-179.668	28.332	-34.102	26.377	-8.261	-12.856
21	c- α_L [d]	0.834587635	1.14893052	-9.63	-62.972	-27.043	-177.84	30.548	-34.302	24.547	-4.794	-16.464
22	c- ε_L [d]	0.854667939	1.11063994	-1.733	-70.669	153.388	177.641	33.395	-35.71	23.807	-2.004	-20.045
23	t- α_L [d]	0.859060506	1.10243549	178.559	-64.409	-27.738	-179.68	26.06	-33.108	27.07	-10.435	-9.993
24	c- α_L [d]	0.863453072	1.09429164	-9.77	-62.77	-27.496	-177.25	30.428	-34.185	24.488	-4.807	-16.378
25	c- ε_L [d]	0.872865715	1.07704254	-1.876	-71.231	153.371	177.594	33.611	-35.857	23.82	-1.878	-20.26
26	c- α_L [d]	0.879140809	1.06569446	-9.889	-62.33	-27.827	-177.096	30.101	-33.876	24.28	-4.792	-16.197
27	t- α_L [d]	0.886043414	1.05334962	170.248	-56.04	-29.592	-178.56	26.461	-35.759	30.851	-14.224	-7.808
28	t- ε_L [d]	0.894828547	1.03784461	169.778	-59.198	153.487	176.826	28.366	-36.2	29.719	-11.737	-10.635
29	t- α_L [d]	0.905496209	1.01932360	175.556	-60.875	-29.808	-178.657	27.684	-33.658	26.29	-8.615	-12.196
30	t- ε_L [d]	0.912398813	1.00751591	177.233	-66.961	154.345	176.976	27.794	-33.45	25.946	-8.031	-12.662
31	t- ε_L [d]	0.927459041	0.98222629	174.626	-64.115	151.95	175.367	30.419	-34.763	25.404	-5.81	-15.773
32	c- ε_L [d]	0.938126703	0.96469783	-3.291	-70.535	153.12	177.54	33.483	-34.36	21.596	0.389	-21.651

33	t- α_L [d]	0.943146779	0.95655772	169.712	-55.97	-29.472	-178.615	26.193	-34.771	29.505	-12.975	-8.458
34	t- α_L [d]	0.950049383	0.94547711	165.447	-53.435	-29.854	-178.172	25.884	-34.453	29.298	-12.975	-8.285
35	t- α_L [d]	0.951304402	0.94347629	175.362	-61.039	-28.822	-178.998	27.718	-33.497	26	-8.297	-12.419
36	t- α_L [d]	0.982052368	0.89575709	166.129	-53.334	-30.287	-178.387	25.099	-34.786	30.63	-14.909	-6.521
37	t- α_L [d]	1.013427843	0.84955111	178.633	-63.984	-28.88	-178.987	25.955	-33.168	27.273	-10.711	-9.749
38	c- α_L [d]	1.01656539	0.84506368	-9.329	-62.914	-27.107	-178.164	30.355	-34.682	25.333	-5.75	-15.738
39	t- ε_L [d]	1.027233052	0.82998297	164.495	-56.344	152.993	176.97	28.09	-36.156	29.88	-12.072	-10.264
40	t- α_L [d]	1.027860561	0.82910430	171.153	-56.406	-29.716	-177.731	25.545	-35.246	30.974	-14.96	-6.759
41	t- α_L [d]	1.051078413	0.79723914	165.788	-53.402	-30.686	-177.616	25.705	-34.994	30.4	-14.258	-7.327
42	t- α_L [d]	1.062373584	0.78218295	175.849	-61.72	-28.977	-179.497	28.244	-34.288	26.771	-8.747	-12.481
43	c- α_L [d]	1.063628603	0.78052769	-9.462	-62.976	-27.188	-177.981	30.484	-34.706	25.239	-5.549	-15.954
44	t- α_L [d]	1.089356492	0.74735630	165.884	-53.481	-30.46	-177.374	25.903	-35.11	30.379	-14.095	-7.55
45	t- α_L [d]	1.112574344	0.71863299	170.369	-56.445	-29.143	-178.101	25.724	-34.231	29.135	-12.9	-8.21
46	c- ε_L [d]	1.114456872	0.71635303	-3.24	-70.809	153.498	177.347	33.778	-34.389	21.342	0.857	-22.13
47	t- α_L [d]	1.11696691	0.71332434	171.237	-56.629	-29.889	-177.841	25.746	-35.368	30.976	-14.84	-6.959
48	c- ε_L [d]	1.132654648	0.69468300	-3.404	-71.046	153.163	177.505	33.693	-34.427	21.477	0.671	-21.966
49	c- α_L [d]	1.135792195	0.69101360	-11.075	-63.19	-28.213	-177.649	31.159	-33.869	23.237	-2.969	-18.044
50	t- ε_L [d]	1.137674724	0.68882127	163.635	-56.031	153.751	176.631	28.087	-35.191	28.346	-10.447	-11.319
51	c- α_L [d]	1.158382537	0.66515990	-9.645	-62.458	-27.128	-177.416	30.071	-33.976	24.521	-5.094	-15.955

52	t- α_L [d]	1.160265065	0.66304959	176.054	-61.66	-28.593	-179.731	27.959	-34.37	27.172	-9.35	-11.914
53	c- ϵ_L [d]	1.169677708	0.65259808	-3.017	-71.674	153.034	177.422	34.082	-34.889	21.823	0.551	-22.121
54	t- α_L [d]	1.182855408	0.63824214	165.996	-53.495	-30.433	-177.748	25.711	-35.112	30.56	-14.43	-7.214
55	t- ϵ_L [d]	4.622234977	0.63756646	176.006	-41.743	132.539	178.252	-5.605	-14.717	29.215	-34.55	25.528
56	t- α_L [d]	1.203563221	0.61631819	178.168	-63.316	-28.592	-178.828	25.386	-32.416	26.629	-10.42	-9.571
57	c- ϵ_L [d]	1.204190731	0.61566571	-11.703	-64.354	153.877	177.66	32.655	-34.571	22.943	-1.625	-19.854
58	t- α_L [d]	1.223643525	0.59577806	165.159	-53.272	-30.275	-177.498	25.87	-34.411	29.246	-12.924	-8.307
59	t- α_L [d]	1.244978848	0.57470371	165.331	-53.511	-29.889	-177.981	25.949	-34.447	29.227	-12.867	-8.384
60	c- α_L [d]	1.248743905	0.57106285	-10.037	-62.966	-27.543	-177.851	30.931	-34.588	24.616	-4.611	-16.84
61	t- α_L [d]	1.25564651	0.56444775	165.703	-53.206	-29.521	-177.935	25.342	-34.881	30.585	-14.715	-6.792
62	t- α_L [d]	1.265059152	0.55555047	165.649	-53.595	-29.54	-178.976	25.893	-35.017	30.236	-13.973	-7.642
63	t- α_L [d]	1.295179608	0.52801080	170.236	-56.546	-29.551	-177.579	25.492	-34.176	29.308	-13.26	-7.839
64	c- α_L [d]	1.303964741	0.52023862	-1.77	-70.445	-28.226	-177.256	32.8	-35.235	23.58	-2.187	-19.525
Others*		< 0.5										

Total Population Distribution

	H ₂ O (%)	Exp. H ₂ O (%)
Total Cis Isomers	29.15	24
Total Trans Isomers	70.85	76

^a Defined in Figure 4.1; angles are in degrees, ^b $E = -1279.15805900$ a.u. The population distributions were calculated using the Boltzmann statistical weights at 25 °C.

Consequently, the values of the backbone torsion angles and endocyclic angles of Compound 1 in *water* show quite different ranges than the gas phase computations. This illustrates the importance of full optimizations in the presence of the solvent. For the *cis*-conformational structures of Compound 1, the ω' value ranges from $-11^\circ \leq \omega' \leq -1^\circ$, ϕ ranges from $-71^\circ \leq \phi \leq -62^\circ$, ψ falls into two different ranges $153^\circ \leq \psi \leq 154^\circ$ and $-28^\circ \leq \psi \leq -27^\circ$, and finally the ω value were obtained in the regions of $\omega \approx 177^\circ$ and $\omega \approx -177^\circ$. On the other hand, the endocyclic torsion angles span as follows, χ_0 in the range of $-22^\circ \leq \chi_0 \leq -15^\circ$, χ_1 in the range of $30^\circ \leq \chi_1 \leq 33^\circ$, χ_2 in the range of $-35^\circ \leq \chi_2 \leq -33^\circ$, χ_3 in the range of $21^\circ \leq \chi_3 \leq 25^\circ$, and finally the χ_4 ranges from $-5^\circ \leq \chi_4 \leq 0^\circ$. For the *trans*- minima conformations of Compound 1 in *water*, the values of the backbone torsion angles of ω' , ϕ , ψ and ω were obtained as follows, the ω' in the range of $163^\circ \leq \omega' \leq 178^\circ$, ϕ ranges from $-66^\circ \leq \phi \leq -53^\circ$, ψ falls into two different ranges $151^\circ \leq \psi \leq 154^\circ$ and $-30^\circ \leq \psi \leq -27^\circ$, and finally $\omega \approx -178^\circ$ and $\omega \approx 179^\circ$ for ω angle. The endocyclic torsion angles for the *trans*-Compound 1 structures in *water* are as follows: χ_0 ranges from $-16^\circ \leq \chi_0 \leq -6^\circ$, χ_1 ranges from $25^\circ \leq \chi_1 \leq 30^\circ$, χ_2 ranges from $-36^\circ \leq \chi_2 \leq -32^\circ$, χ_3 ranges from $24^\circ \leq \chi_3 \leq 30^\circ$, and χ_4 ranges from $-14^\circ \leq \chi_4 \leq -4^\circ$.

The representative conformations or the most stable structures for *t- ϵ_L [d]*, *t- α_L [d]*, *t- ϵ_L [u]*, *c- ϵ_L [u]*, *c- α_L [d]*, and *c- ϵ_L [d]* in gas phase, and the most stable structures *t- ϵ_L [d]*, *t- α_L [d]*, *c- ϵ_L [d]* and *c- α_L [d]* in water, optimized at the B3LYP/6-31+G (d, p) level of theory, are presented in Figure 4.2 and Figure 4.3, respectively. Since the conformational distribution data obtained from the gas phase calculation deviated largely from the experimental data, in the geometric inspections, only the structures of Compound 1 and 2

in water were considered. The general inspection of the geometric parameters of Compound 1 in water, listed in Table 4.3, indicated that the bond angles of the five-membered prolyl residue in cis and trans conformation deviated only by less than 0.5° from one another. This suggests that the cis-trans isomerization of the prolyl amide does not have a greater impact on the bond angle change of the five-membered pyrrolidine ring. Furthermore, the bond angle around the N-terminal, $\angle C^\delta - N - C^\alpha$, shows a larger deviation of 8° - 11° than those of the other bond angles.

Table 4.3 Selected Angles (in deg) for Compound 1, at the B3LYP/6-31+G (d, p) level of theory in the water

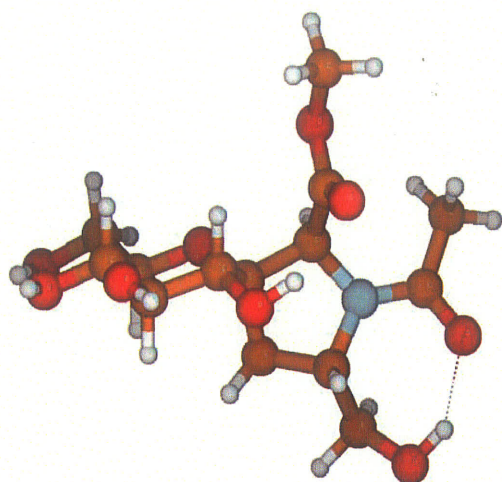
Conformers	$\angle N - C^{\alpha} - C^{\beta}$	$\angle C^{\alpha} - C^{\beta} - C^{\gamma}$	$\angle C^{\beta} - C^{\gamma} - C^{\delta}$	$\angle C^{\gamma} - C^{\delta} - N$	$\angle C^{\delta} - N - C^{\alpha}$
t- α_L [d]	103.95	102.49	105.54	102.92	113.24
t- ε_L [d]	104.1	102.77	105.89	102.99	113.35
c- ε_L [d]	103.38	102.28	105.63	103.42	112.9
c- α_L [d]	103.4	102.57	105.65	103.47	113.09

Table 4.4 Selected bond^a length (in Å) for Compound 1, at the B3LYP/6-31+G (d, p) level of theory in the water

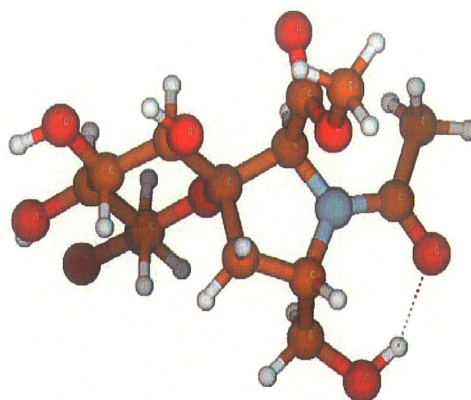
Conformers	$C_{Ac} = O$	$C_{Ac} - N$	$N - C^{\alpha}$	$C^{\alpha} - C_{XY}$	$C^{\alpha} - C^{\beta}$	$C^{\beta} - C^{\gamma}$	$C^{\gamma} - C^{\delta}$	$C^{\delta} - N$
t- α_L [d]	1.24	1.36	1.46	1.54	1.56	1.54	1.54	1.48
t- ε_L [d]	1.24	1.36	1.46	1.53	1.55	1.54	1.54	1.48
c- ε_L [d]	1.24	1.36	1.46	1.54	1.56	1.54	1.54	1.48
c- α_L [d]	1.24	1.36	1.47	1.54	1.56	1.54	1.54	1.48

^a C_{Ac} and C_{XY} denoted for the carbon atoms of the Prolyl N-terminal Carbonyl Carbon and Prolyl N-terminal Acetate carbon

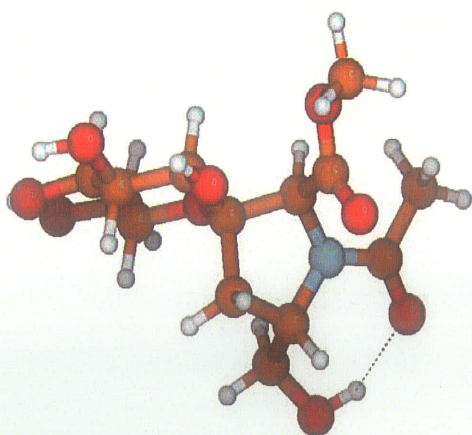
Thus, inspection of the bond lengths displayed in Table 4.4 also suggested that the cis-trans isomerization has little impact on the stretching and compressing of the bond length across the prolyl residue. Figure 4.2 and Figure 4.3 shows the geometries of the most stable conformations in a gas phase and water, respectively. Those low-lying conformers are displayed in accordance with their stability in the conformational distribution calculations in Table 4.1 and Table 4.2, respectively.



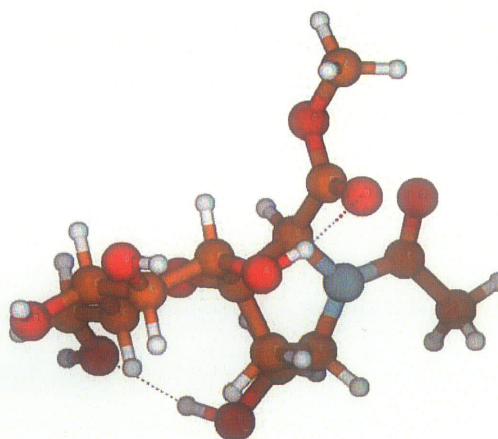
(a) c- ϵ_L [u]



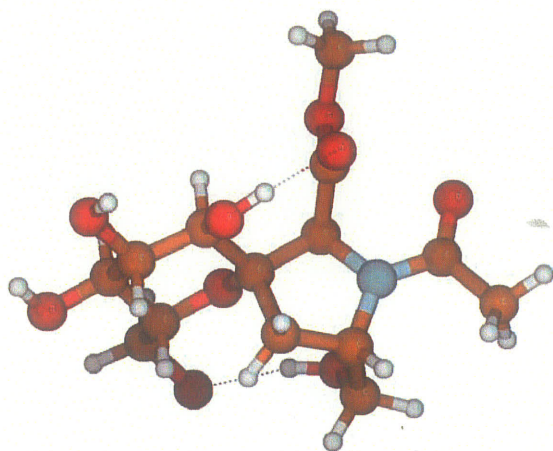
(b) c- α_L [d]



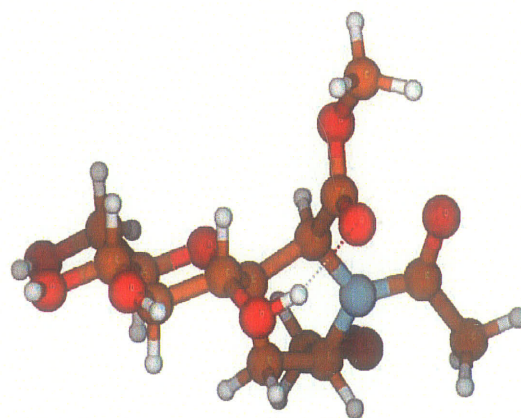
(c) c- ϵ_L [d]



(d) t- ϵ_L [d]

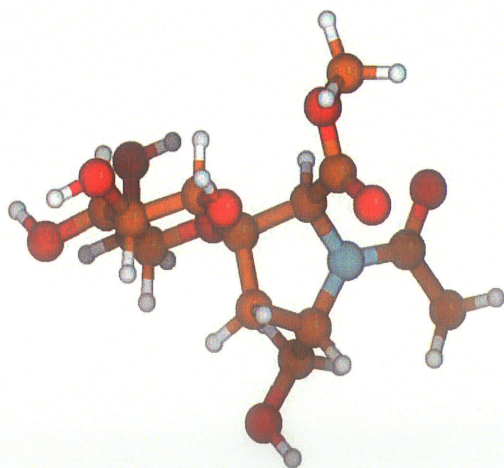


(e) t- α_L [d]

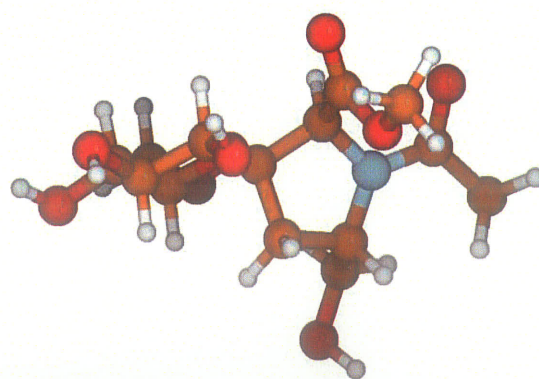


(f) t- ϵ_L [u]

Figure 4.2 The Minima energy representative conformations of Compound 1 in a gas phase optimized at the B3LYP/6-31+G (d) Level of theory.



(a) t- ϵ_L [d]



(b) t- α_L [d]

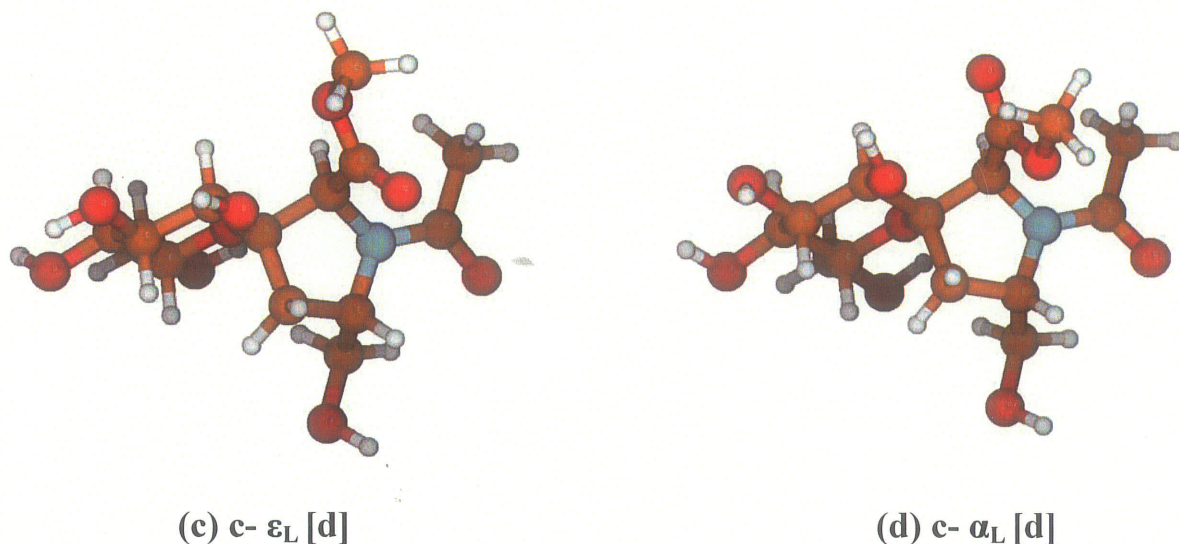


Figure 4.3 The Minima energy representative conformations of Compound 1 in water optimized at the B3LYP/6-31+G (d) Level of theory.

Compound 2. Similarly, the local minima conformers and the geometric classification of Compound 2 were also performed at the same level of theory, B3LYP/6-31+G (d), as shown in Table 4.5 and Table 4.6 in gas phase and in water, respectively. Eight local minima c- ϵ_L [u]*, t- ϵ_L [u]*, c- α_L [d], c- ϵ_L [u], c- ϵ_L [d], c- ϵ_L [d]*, t- ϵ_L [d]* and t- γ_L [d] structures in gas phase and seven local minima c- ϵ_L [d], t- ϵ_L [d]*, c- α_L [d], c- ϵ_L [d]*, t- α_L [d]*, t- α_L [d], and t- ϵ_L [d] in water were identified based on the backbone and endocyclic torsion angle classifications. A distorted down and up puckering was also observed in Compound 2 where the conformers do not follow the general rule of positive χ_1 and χ_3 and negative χ_2 and χ_4 for “Down” puckering and negative χ_1 and χ_3 and positive χ_2 and χ_4 for “UP” puckering, but resemble either the up or down puckering.

Therefore, in this analysis these structures, [u]* and [d]*, are considered as up puckering and down puckering, respectively, throughout the thesis unless otherwise specified.

The most stable conformers for Compound 2 obtained in gas phase and water are c- ϵ_L [u]* and c- ϵ_L [d] respectively. The relative stabilities of the conformations in the gas phase are in the order of c- ϵ_L [u]* < t- ϵ_L [u]* < c- α_L [d] < c- ϵ_L [u] < c- ϵ_L [d] < c- ϵ_L [d]* < t- ϵ_L [d]* < t- γ_L [d]. It is important to mention that the most stable conformer c- ϵ_L [u]* is characterized by a cis prolyl amide bond, the ϵ_L backbone conformation which leads to polyproline II (P_{II}), and distorted “UP” puckering (distorted C_γ-exo). The relative energy difference between c- ϵ_L [u]* and c- ϵ_L [u] of the cis-Compound 2 is approximately 1.1 kcal/mol. This indicates that the distortion of the up ring puckering stabilizes the structure by 1.1 kcal/mol if all the other torsional angles remain the same for both structures. The corresponding relative energy difference between the most stable cis conformer and the most stable trans conformers of Compound 2 is approximately 0.11 kcal/mol. Interestingly, Table 4.5 shows the appearance of backbone conformation γ_L in a trans conformer, which corresponds to a reverse γ turn. Results from the B3LYP/6-31+G (d, p) level of theory in water for Compound 2 shows that c- ϵ_L [d] is the most stable conformer, which is characterized by a cis prolyl amide bond, the ϵ_L backbone conformation which leads to polyproline II (P_{II}), and the “DOWN” puckering (C_γ-endo). The relative stabilities of the conformations in water were also obtained in the order of c- ϵ_L [d] < t- ϵ_L [d]* < c- α_L [d] < c- ϵ_L [d]* < t- α_L [d]* < t- α_L [d] < t- ϵ_L [d]. The relative energy difference between c- ϵ_L [d] and c- α_L [d] of the most two stable cis conformation of Compound 2 in water is approximately 0.42 kcal/mol. This indicates that the backbone ϵ_L stabilizes the structure much better than α_L , which means the formation of polyproline II

(P_{II}) is energetically favorable compared to the formation of right-handed 3₁₀ or α -helices in polypeptides. Surprisingly, it was found that the relative energy difference between the most stable cis (c- ϵ_L [d]) and the most stable trans (t- ϵ_L [d]*) conformers is almost identical, which is approximately 0.018 kcal/mol energy difference between the two. Both the gas phase and water based calculations show more or less the same prediction regarding the relative energy difference between the cis and trans conformers of Compound 2. Additionally, in water the “UP” puckering conformation is not found unlike in the gas phase. This indicates that the “Down” puckering conformers are much more stable in water than the “UP” puckering for Compound 2.

Table 4.5 Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of Compound 2, which are Optimized at the B3LYP/6-31+G(d, p) Level of Theory in the gas phase.

B3LYP/6-31+G(d, p)				Backbone Torsion Angles and Endocyclic Torsion Angles ^a								
Conformers		Gas phase (in kcal/mole)	Conformer Distribution	ω'	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^0
		ΔE										
1	c- ε_L [u]*	0.0 ^b	13.82680428	0.719	-77.888	162.085	177.281	7.924	5.259	-16.154	22.822	-19.776
2	t- ε_L [u]*	0.11346627	11.41673483	-174.027	-77.607	158.272	179.448	8.096	5.257	-16.335	23.115	-20.131
3	c- ε_L [u]*	0.13264923	11.05297678	0.264	-77.211	161.986	177.092	8.006	5.212	-16.154	22.88	-19.866
4	t- ε_L [u]*	0.34483530	7.725557087	-173.497	-77.392	158.35	179.546	8.175	5.189	-16.314	23.152	-20.202
5	c- ε_L [u]*	0.36361665	7.484478737	0.401	-77.445	162.242	176.88	8.197	5.111	-16.182	23.038	-20.097
6	t- ε_L [u]*	0.59632855	5.053182911	-173.822	-77.866	158.523	179.18	8.475	4.895	-16.134	23.166	-20.412
7	c- α_L [d]	0.87764106	3.142970622	1.827	-85.93	-6.503	179.181	35.298	-37.919	25.710	-2.656	-20.806
8	c- α_L [d]	0.87775401	3.14237144	1.85	-85.897	-6.633	179.196	35.286	-37.94	25.755	-2.712	-20.762
9	c- α_L [d]	0.91701728	2.94086023	1.81	-85.291	-7.902	179.386	35.592	-37.997	25.516	-2.242	-21.255
10	c- α_L [d]	0.91711768	2.940361867	1.885	-85.382	-7.816	179.372	35.587	-37.983	25.5	-2.229	-21.26
11	c- α_L [d]	1.08254174	2.223983924	4.154	-82.309	-30.642	-176.405	31.742	-36.630	26.867	-6.391	-16.123
12	c- α_L [d]	1.08258566	2.223819032	4.122	-82.247	-30.628	-176.403	31.736	-36.626	26.866	-6.394	-16.118
13	c- ε_L [u]	1.09696818	2.170480559	-3.526	-66.672	161.817	-179.433	-7.074	21.076	-27.305	23.261	-10.224

14	c- ε_L [d]	1.09705603	2.170158719	3.526	-84.145	154.614	175.347	34.033	-37.828	26.579	-4.467	-18.844
15	c- ε_L [d]*	1.11227314	2.115125474	14.203	-95.663	156.429	175.81	33.175	-33.047	20.096	1.714	-22.196
16	c- α_L [d]	1.14654771	1.996227894	2.19	-85.539	-8.925	179.449	35.341	-38.094	25.934	-2.841	-20.725
17	t- ε_L [d]*	1.45833208	1.17936191	-172.073	-87.689	155.868	178.441	32.985	-32.455	19.27	2.389	-22.547
18	t- ε_L [d]*	1.56017687	0.993085573	-171.601	-86.484	155.844	178.029	32.658	-32.202	19.164	2.29	-22.282
19	c- α_L [d]	1.57769066	0.964156772	3.81	-81.982	-29.944	-176.35	31.882	-36.672	26.811	-6.210	-16.349
20	c- ε_L [d]	1.64890671	0.854949692	3.117	-83.224	154.543	176.045	34.054	-38.057	26.927	-4.841	-18.622
21	c- ε_L [d]	1.64891926	0.854931581	3.105	-83.193	154.558	176.046	34.053	-38.058	26.931	-4.846	-18.618
22	c- ε_L [d]	1.64897574	0.854950084	3.26	-83.323	154.558	176.346	34.253	-38.258	26.951	-4.856	-18.647
23	t- γ_L [d]	1.72317246	0.754219562	-170.195	-89.793	1.503	175.591	35.127	-36.402	23.571	-0.669	-21.935
24	c- ε_L [d]*	1.74292019	0.729492989	14.014	-95.337	156.396	176.224	33.248	-33.074	20.078	1.799	-22.307
25	c- ε_L [d]	1.74297039	0.729431176	14.035	-95.342	156.409	176.215	33.251	-33.089	20.1	1.779	-22.297
26	t- γ_L [d]	1.95926036	0.506321503	-169.248	-90.054	1.68	175.264	35.236	-36.719	24.026	-1.079	-21.736
Others*		< 0.5										

Total Population Distribution

	Gas (%)	Exp. H ₂ O (%) [101]
Total Cis Isomers	68.63	52
Total Trans Isomers	31.37	48

^a Defined in Figure 4.1; angles are in degrees, ^b $E = -1279.13603782$ a.u. The population distributions were calculated using the Boltzmann statistical weights at 25 °C.

Table 4.6 Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of Compound 2, which are Optimized at the B3LYP/6-31+G(d, p) Level of Theory in Water

B3LYP/6-31+G(d, p)				Backbone Torsion Angles and Endocyclic Torsion Angles ^a								
Conformers		H ₂ O	Conformer	ω	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^0
		(in kcal/mole)	Distribution									
1	c- ϵ_L [d]	0.0 ^b	4.15042264	4.031	-81.347	154.516	176.927	32.683	-37.378	27.191	-6.035	-16.964
2	c- ϵ_L [d]	0.00753011	4.09800180	4.297	-80.845	154.838	177.406	32.539	-37.457	27.452	-6.399	-16.641
3	t- ϵ_L [d]*	0.01757027	4.02913580	-170.077	-85.486	154.199	177.889	32.851	-33.196	20.473	1.061	-21.596
4	c- ϵ_L [d]	0.10291156	3.48858953	3.902	-82.254	154.02	177.165	32.835	-37.137	26.661	-5.391	-17.464
5	t- ϵ_L [d]*	0.18699783	3.02696829	-172.554	-82.246	154.897	178.008	59.846	-30.292	18.116	2.039	-20.834
6	c- ϵ_L [d]	0.23406104	2.79580419	3.878	-79.31	154.78	177.459	31.566	-36.457	26.859	-6.415	-16.019
7	c- ϵ_L [d]	0.23845361	2.77515118	3.926	-78.782	153.309	177.545	31.75	-36.739	27.139	-6.59	-16.015
8	c- ϵ_L [d]	0.26543652	2.65158775	4.344	-81.073	154.429	176.76	32.373	-37.373	27.468	-6.529	-16.447
9	t- ϵ_L [d]*	0.26794656	2.64037702	-172.448	-82.232	154.292	179.353	30.499	-30.342	18.261	1.836	-20.648
10	t- ϵ_L [d]*	0.29492946	2.52281441	-172.917	-81.541	154.684	178.19	30.176	-29.97	18.004	1.879	-20.468
11	t- ϵ_L [d]*	0.35768041	2.26925765	-170.53	-86.917	153.504	177.346	33.723	-33.628	20.387	1.719	-22.572
12	c- ϵ_L [d]	0.41541129	2.05855481	4.2	-79.496	154.916	176.881	31.871	-37.151	27.642	-7.035	-15.814
13	c- α_L [d]	0.42733397	2.01753998	3.512	-81.193	-28.105	-177.812	32.361	-36.72	26.421	-5.464	-17.112

14	$t-\varepsilon_L[d]^*$	0.44427673	1.96065755	-170.208	-86.394	153.91	177.6	33.493	-33.659	20.655	1.279	-22.147
15	$c-\varepsilon_L[d]^*$	0.51769534	1.73212807	12.897	-91.14	154.666	177.338	32.641	-33.236	20.844	0.606	-21.173
16	$c-\varepsilon_L[d]^*$	0.53714813	1.67617568	12.757	-90.437	154.79	177.186	32.551	-33.191	20.854	0.538	-21.074
17	$c-\alpha_L[d]$	0.60491916	1.49498874	4.217	-79.003	-27.526	-177.594	31.29	-36.44	27.064	-6.843	-15.553
18	$c-\alpha_L[d]$	0.63064705	1.43145372	4.245	-80.608	-28.347	-177.15	31.629	-36.286	26.438	-5.955	-16.341
19	$c-\varepsilon_L[d]$	0.63127456	1.42993829	4.529	-82.05	152.858	175.989	32.954	-37.887	27.75	-6.464	-16.866
20	$t-\alpha_L[d]^*$	0.64005969	1.40889000	-170.052	-84.046	-27.779	-179.098	32.501	-32.811	20.21	1.031	-21.333
21	$c-\alpha_L[d]$	0.65198237	1.42993829	3.915	-80.778	-28.379	-177.398	31.934	-36.576	26.596	-5.929	-16.544
22	$c-\alpha_L[d]$	0.66641509	1.34758590	4.935	-78.149	-27.92	-177.228	31.189	-37.028	28.108	-8.014	-14.737
23	$c-\varepsilon_L[d]^*$	0.73230359	1.20574339	12.555	-90.425	154.631	177.253	32.497	-32.985	20.554	0.81	-21.207
24	$c-\alpha_L[d]$	0.73481362	1.20064559	4.62	-80.305	-28.143	-177.458	32.019	-37.135	27.458	-6.787	-16.038
25	$c-\alpha_L[d]$	0.74046121	1.18925420	3.744	-80.935	-28.491	-177.344	32.074	-36.59	26.508	-5.756	-16.736
26	$c-\alpha_L[d]$	0.78062182	1.18925420	4.276	-77.968	-27.479	-177.297	31.037	-36.672	27.67	-7.642	-14.882
27	$c-\varepsilon_L[d]^*$	0.78501438	1.10309683	12.314	-90.783	153.222	177.425	33.237	-34.026	21.54	0.256	-21.331
28	$t-\varepsilon_L[d]^*$	0.79191699	1.09031872	-171.984	-81.833	154.82	178.388	30.76	-31.456	19.83	0.343	-19.863
29	$t-\alpha_L[d]^*$	0.79944710	1.07654773	-170.08	-83.735	-28.796	-178.569	32.442	-33.185	20.865	0.318	-20.84
30	$c-\varepsilon_L[d]^*$	0.81827239	1.04287641	12.45	-90.565	154.9	177.356	32.588	-33.336	21.051	0.357	-20.977
31	$t-\alpha_L[d]^*$	0.82705752	1.02752556	-171.645	-79.531	-26.81	-179.607	30.626	-31.461	19.901	0.15	-19.636
32	$t-\varepsilon_L[d]^*$	0.86282556	0.96732360	-170.626	-84.75	154.602	177.906	32.227	-32.33	19.761	1.395	-21.428

33	c- α_L [d]	0.86470809	0.96425464	4.865	-79.122	-28.789	-176.825	31.377	-36.812	27.568	-7.32	-15.299
34	t- α_L [d]	0.89482855	0.91645474	-169.534	-84.422	-27.317	-179.049	32.379	-33.319	21.169	-0.052	-20.571
35	t- ε_L [d]	0.89545606	0.91548452	-171.356	-82.268	153.853	179.034	31.428	-32.728	21.177	-0.623	-19.657
36	t- ε_L [d]*	0.93436165	0.85729458	-170.132	-85.569	153.411	-179.818	32.805	-33.005	20.231	1.281	-21.716
37	c- ε_L [d]	0.94000923	0.84916081	3.931	-82.023	153.88	177.72	33.075	-37.619	27.186	-5.76	-17.401
38	t- ε_L [d]	0.95569697	0.82696965	-171.096	-82.727	154.418	177.761	31.508	-33.04	21.605	-1.014	-19.463
39	c- ε_L [d]*	0.96636463	0.81221184	11.487	-91.737	154.783	177.15	32.743	-32.909	20.221	1.309	-21.674
40	t- ε_L [d]*	1.00025014	0.76705854	-170.138	-85.506	154.067	178.014	33.061	-33.416	20.64	1.023	-21.699
41	t- ε_L [d]*	1.00652524	0.75897656	-171.742	-82.677	154.286	177.643	31.176	-31.666	19.763	0.694	-20.348
42	t- α_L [d]	1.05735351	0.69657383	-169.538	-83.9	-28.409	-178.713	32.12	-33.097	21.07	-0.119	-20.361
43	t- α_L [d]	1.06927619	0.69657383	-171.35	-80.052	-26.936	-179.325	30.15	-31.243	20.044	-0.33	-19.024
44	c- ε_L [d]	1.07617879	0.67478699	4.091	-81.838	154.139	177.361	32.967	-37.581	27.244	-5.904	-17.226
45	c- ε_L [d]*	1.08182638	0.66838481	11.538	-90.866	155.267	177.139	32.629	-32.982	20.448	0.998	-21.406
46	t- α_L [d]*	1.12073197	0.62590100	-169.576	-83.957	-27.802	-179.535	32.477	-33.131	20.735	0.47	-20.97
47	c- ε_L [d]	1.14959740	0.59613546	3.956	-81.412	153.965	177.561	32.564	-37.37	27.294	-6.205	-16.792
48	c- ε_L [d]	1.16026507	0.58549704	3.907	-82.207	154.239	177.143	32.742	-37.114	26.736	-5.529	-17.326
49	t- ε_L [d]*	1.16277510	0.58302160	-171.452	-84.225	154.551	178.295	31.992	-32.045	19.589	1.414	-21.303
50	t- ε_L [d]*	1.17971786	0.56658391	-170.544	-84.891	154.554	177.422	32.29	-32.307	19.654	1.548	-21.559
51	c- α_L [d]	1.18662046	0.56002069	3.759	-82.204	-27.891	-178.006	32.7	-36.942	26.448	-5.26	-17.46

52 c- ϵ_L [d]*	1.21046583	0.53792720	11.31	-91.845	154.642	176.438	32.758	-32.898	20.179	1.352	-21.704
53 t- ϵ_L [d]*	1.21297586	0.53565288	-171.622	-83.652	154.359	178.253	31.855	-31.854	19.37	1.568	-21.317
54 c- α_L [d]	1.22552605	0.52442471	12.852	-89.47	-28.305	-178.455	32.443	-33.58	21.54	-0.319	-20.44
55 t- ϵ_L [d]	1.23493870	0.51615831	4.236	-81.741	153.895	177.438	32.835	-37.567	27.355	-6.103	-17.018
Others*	< 0.5										

Total Population Distribution

	H ₂ O (%)	Exp. H ₂ O (%) [101]
Total Cis Isomers	57.75	52
Total Trans Isomers	42.25	48

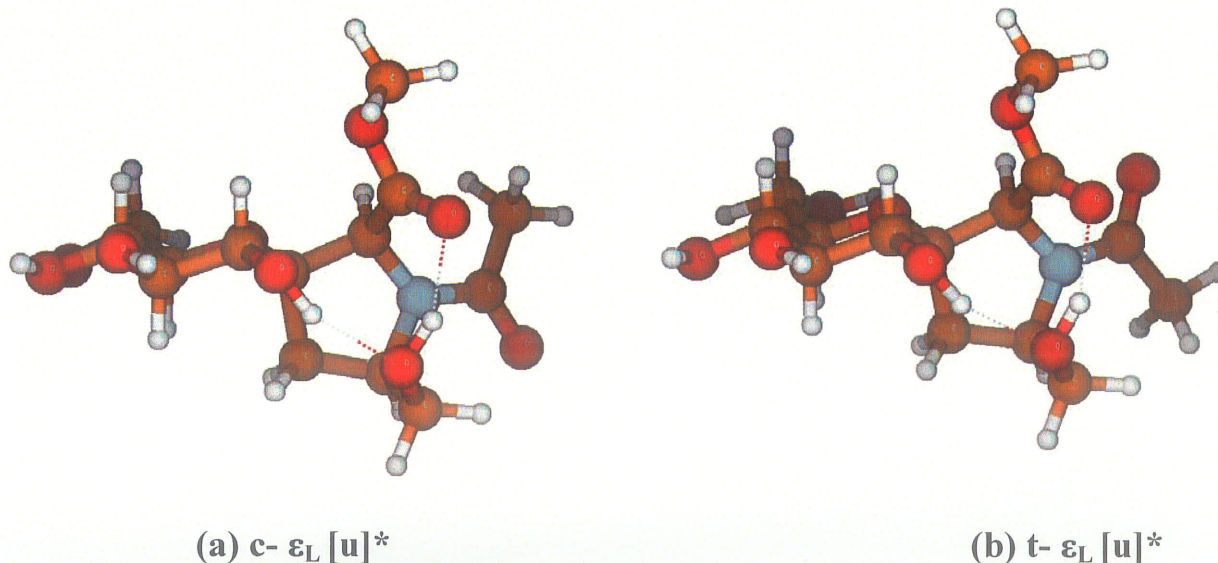
^a Defined in Figure 4.1; angles are in degrees, ^b $E = -1279.15838500$ a.u. The population distributions were calculated using the Boltzmann statistical weights at 25 °C.

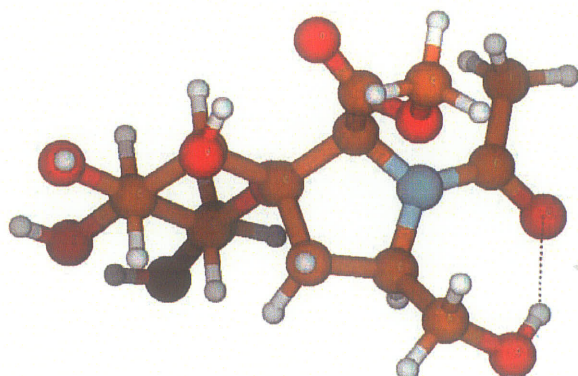
As reported earlier in Table 4.5, the gas phase calculations for cis-Compound 2 show that the backbone torsion angles of ω' , ϕ , ψ , and ω fall normally into the following ranges – except for few exceptional structures that do not bear by this generalization. The ω' value ranges from $-3.5^\circ \leq \omega' \leq 14.3^\circ$, ϕ ranges from $-95.7^\circ \leq \phi \leq -66.7^\circ$, ψ falls into two different regions i.e. $156.4^\circ \leq \psi \leq 161.8^\circ$ and $-30^\circ \leq \psi \leq -6.5^\circ$, and the ω value falls in the regions of $\omega \approx -176^\circ$ and $\omega \approx 179^\circ$. The results found for the endocyclic torsion angles exhibit ranges as follows: χ_0 ranges from $-10.2^\circ \leq \chi_0 \leq -22.2^\circ$, χ_1 ranges from $-7^\circ \leq \chi_1 \leq 35^\circ$, χ_2 ranges from $21^\circ \leq \chi_2 \leq -38^\circ$, χ_3 ranges from $-27^\circ \leq \chi_3 \leq 26^\circ$, and χ_4 ranges from $-2^\circ \leq \chi_4 \leq 23^\circ$. The trans-conformations of Compound 2 in the gas phase show the following ranges for ω' , ϕ , ψ and ω : ω' falls in the region $-174^\circ \leq \omega' \leq -169^\circ$, ϕ ranges $-90^\circ \leq \phi \leq -77^\circ$, ψ falls in the regions of $1^\circ \leq \psi \leq 2^\circ$ and $154^\circ \leq \psi \leq 158^\circ$, and ω in the range of $\omega \approx 177^\circ$ and 179° . The endocyclic torsion angles for the trans-Compound 2 are as follows: χ_0 ranges from $-22^\circ \leq \chi_0 \leq -20^\circ$, χ_1 ranges from $8^\circ \leq \chi_1 \leq 35^\circ$, χ_2 ranges from $-36^\circ \leq \chi_2 \leq 4^\circ$, χ_3 ranges for $-16^\circ \leq \chi_3 \leq 24^\circ$, and χ_4 ranges from $-1^\circ \leq \chi_4 \leq 23^\circ$.

In water, the value of the backbone torsion and endocyclic angles of cis-Compound 2 falls in the range of, ω' from $3^\circ \leq \omega' \leq 12^\circ$, ϕ ranges from $-77^\circ \leq \phi \leq -78^\circ$, ψ falls into two different ranges $153^\circ \leq \psi \leq 154^\circ$ and $-28^\circ \leq \psi \leq -27^\circ$, and ω falls either in the region of $\omega \approx 177^\circ$ or $\omega \approx -177^\circ$. The endocyclic torsion angles of χ_0 in the range of $-21 \leq \chi_0 \leq -14^\circ$, χ_1 in the range of $32^\circ \leq \chi_1 \leq 33^\circ$, χ_2 in the range of $-37^\circ \leq \chi_2 \leq -33^\circ$, χ_3 in the range of $20^\circ \leq \chi_3 \leq 27^\circ$, and χ_4 ranges from $-8^\circ \leq \chi_4 \leq 0^\circ$.

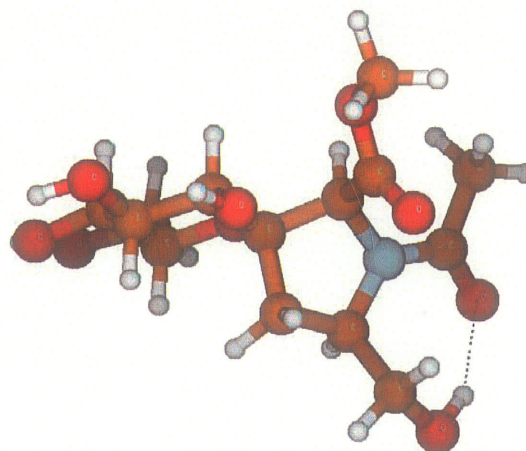
However, for the trans-Compound 2 minima conformations in *water*, ω' falls in the range of $-172^\circ \leq \omega' \leq 169^\circ$, ϕ ranges from $-86^\circ \leq \phi \leq -79^\circ$, ψ falls into two different ranges $153^\circ \leq \psi \leq 154^\circ$ and $-28^\circ \leq \psi \leq -27^\circ$, and $\omega \approx 178^\circ$ and $\omega \approx -179^\circ$. The endocyclic torsion angles of χ_0 ranges from $-21^\circ \leq \chi_0 \leq -19^\circ$, χ_1 ranges from $30^\circ \leq \chi_1 \leq 33^\circ$, χ_2 ranges from $-33^\circ \leq \chi_2 \leq -31^\circ$, χ_3 ranges from $19^\circ \leq \chi_3 \leq 20^\circ$, and χ_4 ranges from $0^\circ \leq \chi_4 \leq 2^\circ$.

The representative conformations for the most stable c- ϵ_L [u]*, t- ϵ_L [u]*, c- α_L [d], c- ϵ_L [d], t- ϵ_L [d]* and t- γ_L [d] in gas phase, and the most stable structures c- ϵ_L [d], t- ϵ_L [d]*, c- α_L [d], and t- α_L [d]* in water, optimized at the B3LYP/6-31+G (d) level of theory are displayed in Figure 4.4 and Figure 4.5, respectively.

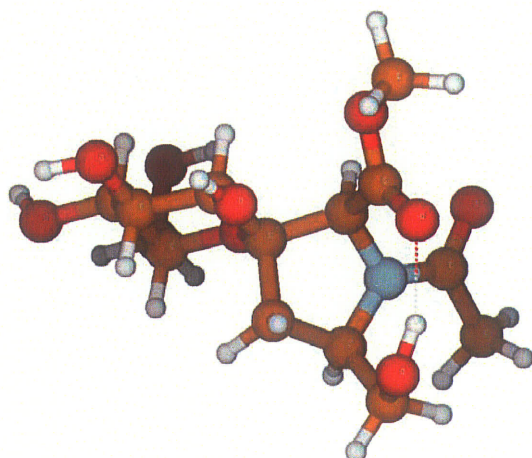




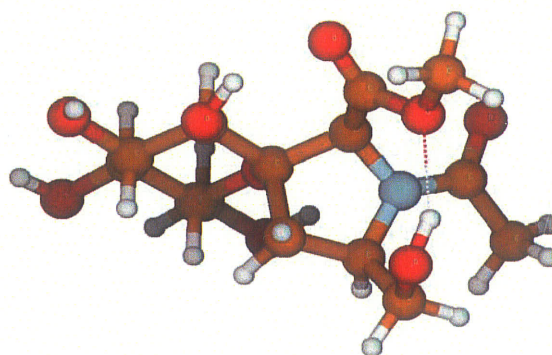
(c) c- α_L [d]



(d) c- ϵ_L [d]

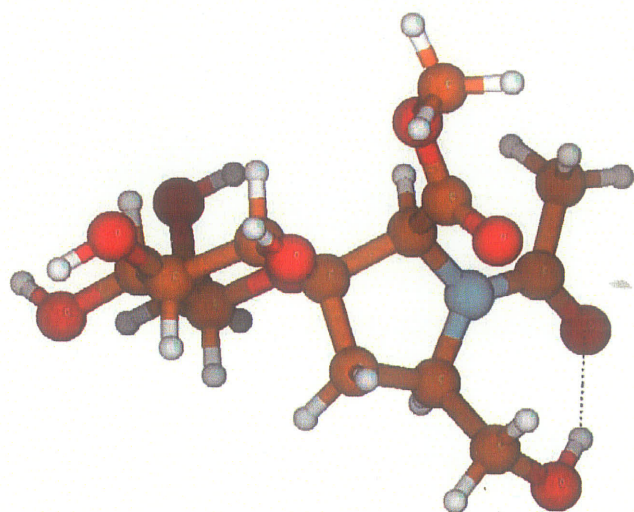


(e) t- ϵ_L [d]*

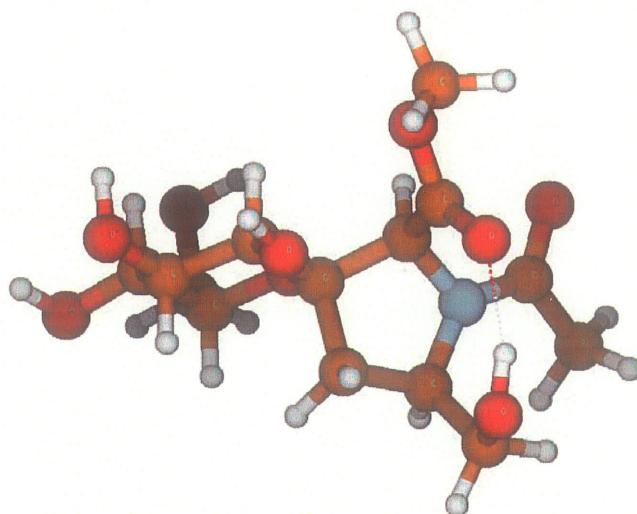


(f) c- γ_L [d]

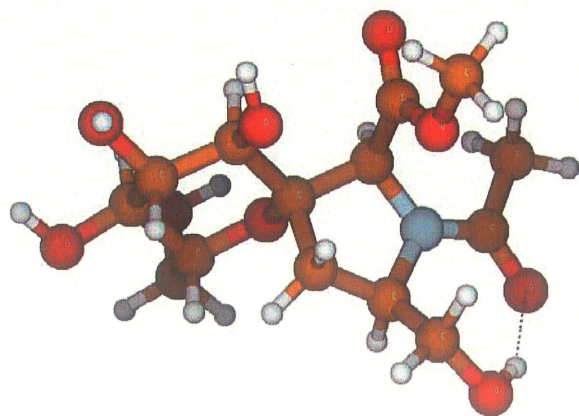
Figure 4.4 The Minima energy representative conformations of Compound 2 in a gas phase optimized at the B3LYP/6-31+G (d) Level of theory.



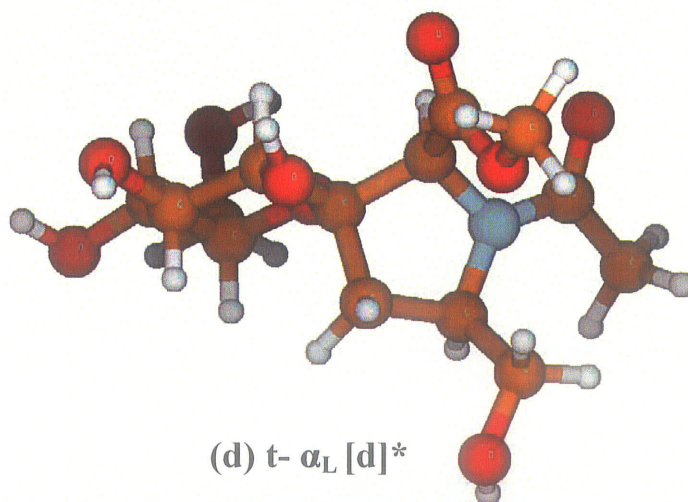
(a) c- ϵ_L [d]



(b) t- ϵ_L [d]*



(c) c- α_L [d]



(d) t- α_L [d]*

Figure 4.5 The Minima energy representative conformations of Compound 2 in water optimized at the B3LYP/6-31+G (d) Level of theory.

Inspection of the geometric parameters of the selected angles and bond lengths of the Compound 2 in water, which are listed in Table 4.7 and 4.8, respectively indicate that the bond angles of the pyrrolidine five-membered prolyl residue in the cis and trans conformations are similar. This suggests that the cis-trans isomerization of the prolyl amide do not have a greater influence on the bond angle change of the five-membered pyrrolidine ring. Similar conclusions were also drawn for Compound 2 the same as Compound 1.

Table 4.7: Selected Angles (in deg) for Compound 2, at the B3LYP/6-31+G (d, p) level of theory in the water

Conformers	$\angle N - C^\alpha - C^\beta$	$\angle C^\alpha - C^\beta - C^\gamma$	$\angle C^\beta - C^\gamma - C^\delta$	$\angle C^\gamma - C^\delta - N$	$\angle C^\delta - N - C^\alpha$
c- ϵ_L [d]	103.2	102.6	106.3	103.4	112.8
t- ϵ_L [d]*	103.1	102.7	106.01	103.3	112.9
c- α_L [d]	103.38	102.21	105.52	102.85	112.68
t- α_L [d]*	103.1	102.92	105.82	103.57	112.86

Table 4.8 Selected bond^a length (in Å) for Compound 2, at the B3LYP/6-31+G (d, p) level of theory in the water

Conformers	$C_{Ac} = O$	$C_{Ac} - N$	$N - C^{\alpha}$	$C^{\alpha} - C_{\alpha\chi}$	$C^{\alpha} - C^{\beta}$	$C^{\beta} - C^{\gamma}$	$C^{\gamma} - C^{\delta}$	$C^{\delta} - N$
c- ϵ_L [d]	1.24	1.35	1.46	1.54	1.56	1.54	1.54	1.48
t- ϵ_L [d]*	1.24	1.36	1.46	1.53	1.56	1.54	1.54	1.49
c- α_L [d]	1.25	1.35	1.47	1.54	1.56	1.54	1.54	1.49
t- α_L [d]*	1.24	1.36	1.46	1.53	1.56	1.54	1.54	1.48

^a C_{Ac} and $C_{\alpha\chi}$ denoted for the carbon atoms of the Prolyl N-terminal Carbonyl Carbon and Prolyl N-terminal Acetate carbon

4.1.2 Conformational distribution of Compound 1 and 2 in gas and water

The conformational distribution of the conformers of Compound 1 and Compound 2 in gas phase and water shows completely different results; this could be due to the dipole moment effect or the influence of the intramolecular hydrogen bond(s). Figure 4.6 shows the population distribution of the cis and trans conformers of Compound 1 (up) and Compound 2 (down), respectively.

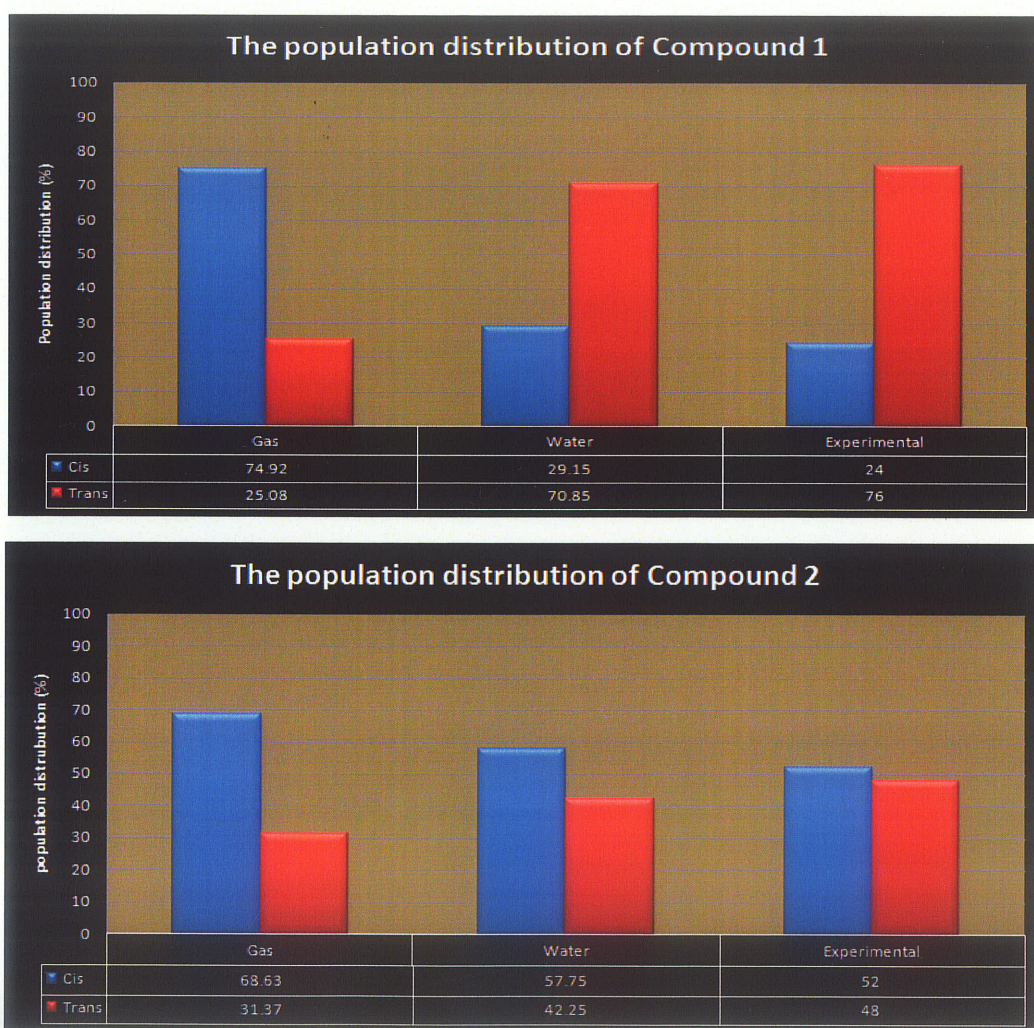


Figure 4.6 The cis and trans population distribution of Compound 1 & 2, each population was computed using the Boltzmann weight by the relative energy at the B3LYP/6-31+G (d) Level.

The population of the cis and trans conformers in gas phase are 74.92%, 25.08% for Compound 1, and 68.63%, 31.37% for Compound 2. Comparing this result with the actual experimental data [101] for the cis and trans structures of Compound 1 and Compound 2 leads to the conclusion that predicting the conformational preference of the carbohydrate templated proline mimetic compounds using the gas phase calculations would lead to erroneous results. In water, the population of the cis and trans conformers for Compound 1 and Compound 2 are 29.15%, 70.85%; and 57.75%, 42.25%, respectively. These results are in an excellent agreement with the experimental findings [101]. The population distributions of the backbone and endocyclic conformations, as shown in Table 4.2 and Table 4.6 for Compound 1 and 2 in water, are also shown in Figure 4.7.

In water, the population of the trans conformation $t-\alpha_L$ [d] and $t-\epsilon_L$ [d] for compound 1 are calculated to be approximately 35.4% and 22.6%, respectively. These conformations constitute the highest population distribution for Compound 1. The calculated populations of the cis backbone and endocyclic conformations for Compound 1 are very low relative to the trans counterpart, which comprises a population distribution of 12.2% for $c-\epsilon_L$ [d] and 8.9% for $c-\alpha_L$ [d].

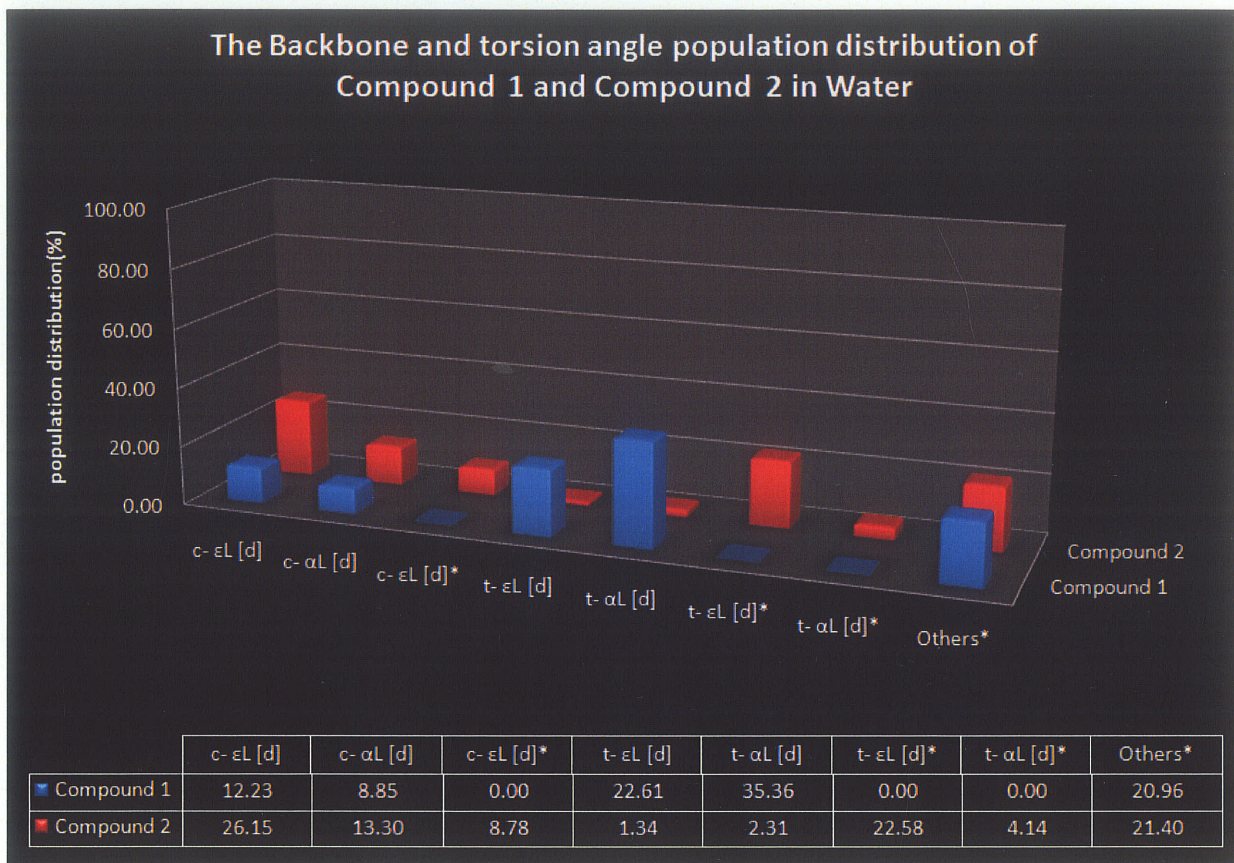


Figure 4.7 The backbone and endocyclic population distribution of Compound 1 and 2, each population was computed using the Boltzmann weight by the relative energy at the B3LYP/6-31+G (d) Level.

In figure 4.7, the entry “others*” indicated the population distribution of the backbone and endocyclic conformations which are not listed in Table 4.1 and Table 4.5, respectively. These conformations which are not listed in the above tables might be the same or different structures to those already listed. Figure 4.7, also revealed that there isn’t any distorted puckering for Compound 1; this indicated that the down or up puckering is favored to a greater extent.

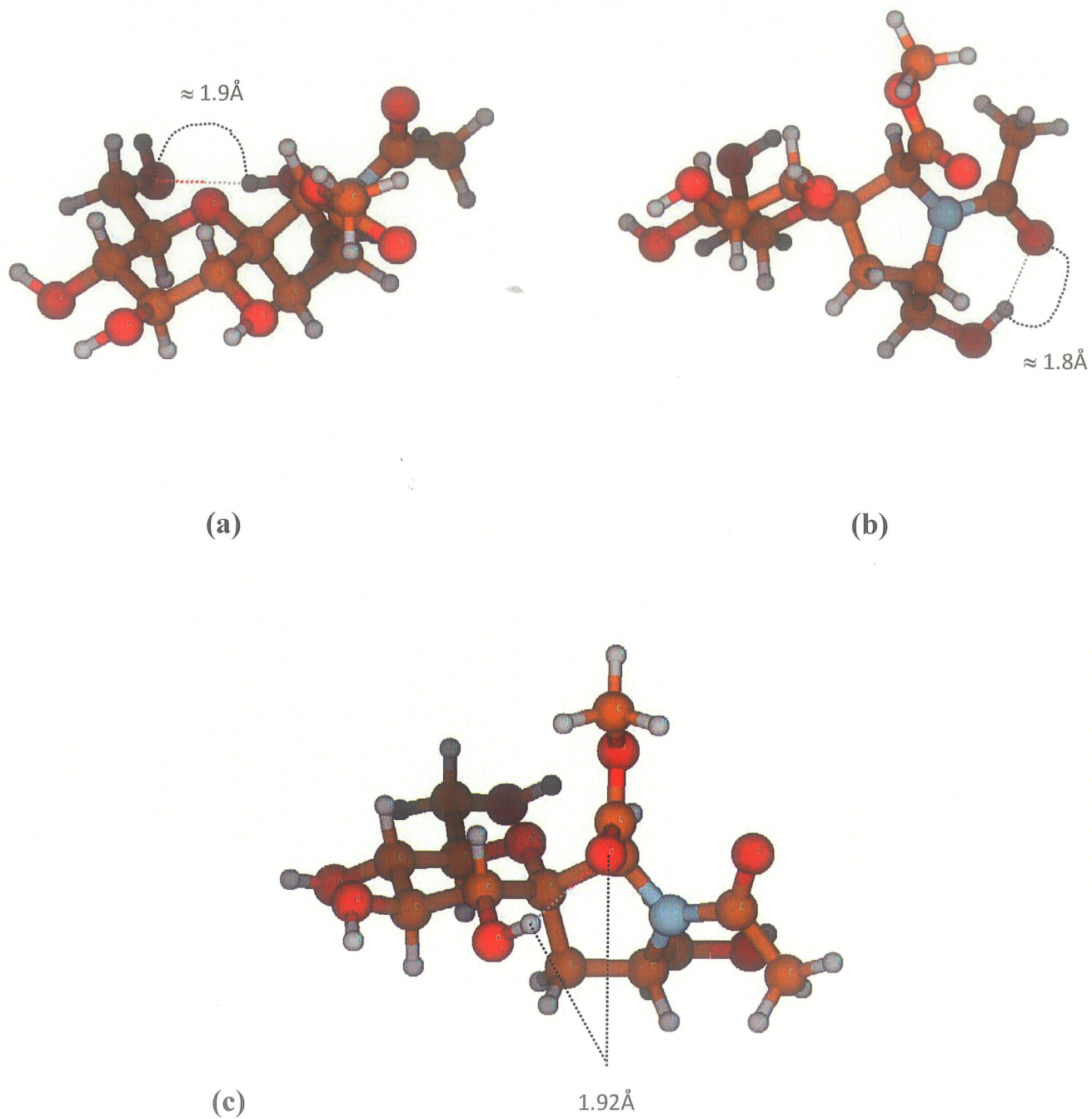


Figure 4.8 The three hydrogen bonds that exists in water for Compound 1

Using the same distance criteria and the same strategy as before, the conformers of Compound 2 were also investigated for internal hydrogen bonding. Two major types of hydrogen bonds were found (see Figure 4.9). The first major hydrogen bonding exists between the Prolyl C_δ primary hydroxyl group and Prolyl N-terminal Carbonyl Oxygen (C_δ -6'C-OH-----O=C'-3'N, see Figure 4.9 a) and the bond distance is $\approx 1.8\text{\AA}$. The conformers with this kind of hydrogen bond are conformers 1, 2, 4, 6, 7, 8, 12, 13, 17, 18, 19, 21, 22, 25, 33, 37, 44, 47, 48, 51, and 55, of a population approximately 37.3% of the total population distribution.

All these above mentioned twenty-one conformers are *cis* structures. However, the second hydrogen bonding exists between the Prolyl C_δ primary hydroxyl group and Prolyl Carbonyl Oxygen (i.e. C_δ -6'C-OH-----O=C'-C α , see Figure 4.9 b) and is found in *both* *cis* and *trans* conformations. The conformers with this kind of hydrogen bond that favors the *cis* structure are conformers 15, 16, 23, 30, 39, and 52, with a total population of approximately 7%; and the conformers that favor the *trans* structure are also conformers 3, 5, 10, 32, 36, 40, 49, 50, and 53, with a population of approximately 13.9%. The bond distance of this hydrogen bonding is $\approx 1.9\text{\AA}$.

It is important to mention that at least one of the above mentioned hydrogen bonds exists in the most stable conformers of Compound 2. From this we can speculate that *the existence of this intramolecular hydrogen bonding* may influence the population distribution of the *cis-trans* ratios by reducing the energy barrier for the isomerization, or indirectly it stabilizes the *cis-trans* isomerization of the prolyl residue.

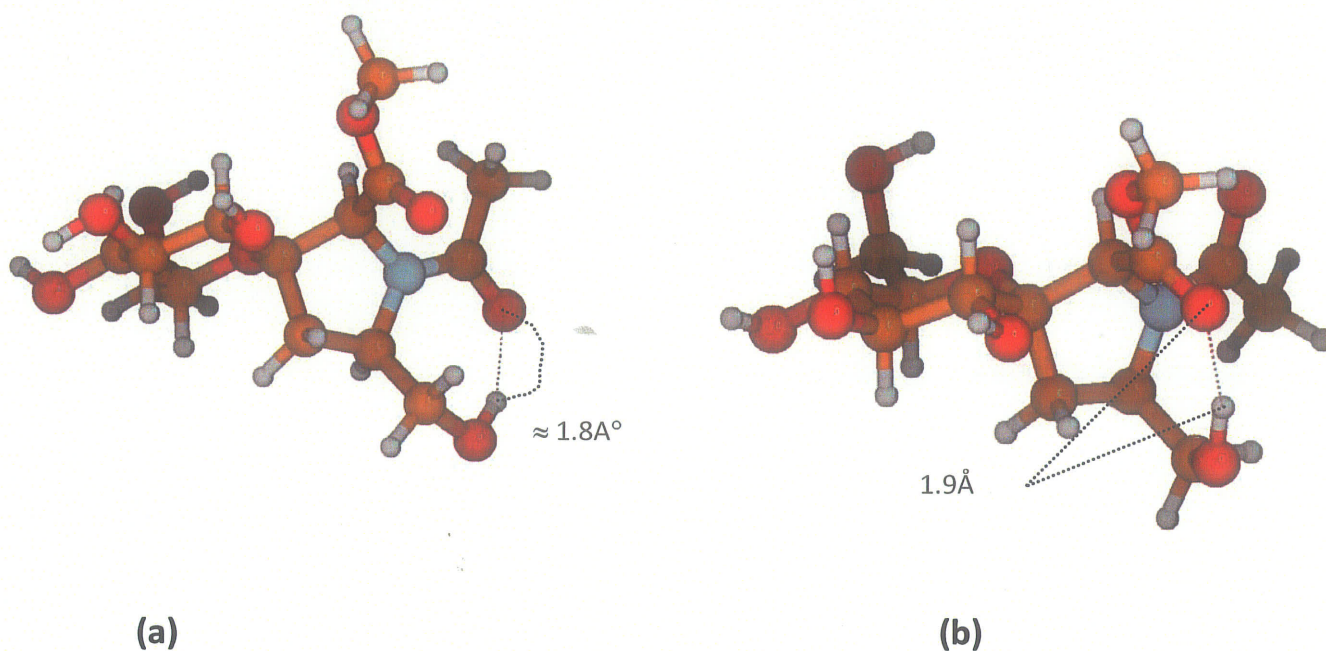


Figure 4.9 The two hydrogen bonds that exists in water for Compound 2

The puckering of the five-membered prolyl residue or ring was calculated using the classical pseudorotational parameters A and P of equation 1.11, which described the puckering amplitude and the state of the pucker in the pseudorotational pathway, respectively. Based on the formulas given earlier in Chapter 1 (equation 1.11), the A and P values of the backbone minima energy conformations of Compound 1 and 2 in water were only calculated, as listed in Table 4.9.

Table 4.9 Backbone Dihedral angles (in degree), Pseudorotational Parameters (A and P, in degree), Relative Energies (ΔE , in Kcal/mole) and dipole moment for the minima energy representative conformations for Compound 1 and Compound 2 at the B3LYP/6-31+G (d, p) level of theory in the water

Conformation	ω	ϕ	ψ	ω	(A, P)	ΔE	Dipole ^a Moment
--------------	----------	--------	--------	----------	--------	------------	-------------------------------

Compound 1

1	t- ϵ_L [d]	175.253	-64.198	153.736	177.502	(34.4,-67.7)	0.0	5.1329
3	t- α_L [d]	175.55	-61.41	-28.681	-179.498	(33.4,-68.3)	0.12	7.6872
11	c- ϵ_L [d]	-10.903	-63.708	154.37	177.909	(35.5,-58.9)	0.55	10.7905
17	c- α_L [d]	-10.68	-62.642	-27.379	-177.946	(34.7,-59.9)	0.69	8.3907

Compound 2

1	c- ϵ_L [d]	4.031	-81.347	154.516	176.927	(37.6,-63.2)	0.0	10.9153
5	t- ϵ_L [d]*	-172.554	-82.246	154.897	178.008	(40.3,-58.8)	0.19	6.5048
13	c- α_L [d]	3.512	-81.193	-28.105	-177.812	(37,-62.4)	0.43	8.3455
20	t- α_L [d]*	-170.052	-84.046	-27.779	-179.098	(34.7,-52.1)	0.64	7.2791

^a units in Debye (D).

Inspection of the pseudorotational parameters listed in Table 4.8 suggested that for Compound 1 the puckering amplitude for cis and trans show a similar trend while the values for the state of the pucker for cis and trans deviated by a large degree (approximately 8-10°). Similarly, for Compound 2 the pseudorotational parameters are completely different from one another among the cis and trans conformers. The puckering amplitude for trans and cis conformer deviated by at least 3-4°, while the state of the pucker for cis and trans shows a divergence of approximately 8-10°. Table 4.8 also indicated that the dipole moments for the cis conformers are higher relative to the trans conformers in both Compound 1 and Compound 2, although results show that there is no significant difference in dipole moment between these two compounds, be it in cis or trans conformers.

4.2 Conclusions

The present study has provided a number of important insights into the population distribution and the intramolecular hydrogen bonding of Compound 1 and Compound 2. Based on the results displayed in this chapter a number of conclusions can be drawn from this study. First, the population distribution obtained in gas phase does not match the experimental results for both compounds, but, the DFT calculations in water match the experimental data. Second, the position of the C_δ primary hydroxyl group does influence the cis-trans isomerization via the intramolecular hydrogen bonding. This results in a population shift by favoring more cis population in Compound 2. Third, the puckering amplitude calculated for both Compound 1 and Compound 2 shows that the position of

the C_δ primary hydroxyl group greatly distorts the puckering behaviour of the five-membered ring, which is a key parameter in collagen stabilization.

Chapter 5

Conformational Preference of Fused Carbohydrate-Template Proline Analogues

5.1. Results and Discussion

The conformational distribution for the Fused Carbohydrate-Template Proline Analogue - *N-acetyl-GlcProH-NHMe* in gas phase and water were also calculated, as shown in Table 5.1 and 5.2 respectively. The build and search methodology was also used for this compound to generate an initial trial structures, leading to a total of 1200 possible structures. These trial structures were superimposed and compared with respect to their energies, and a set of unique local minima conformers were identified. The initial results showed that a total of 101 unique local minima conformers were located using the SPARTAN software that uses the MMFF94 and Monte Carlo searching method. Following the same strategy as before, all the local minima conformer that was obtained using the MMFF94 were further subjected to density functional theory calculations using the GAUSSIAN 03 program package. The B3LYP level of theory with 6-31+G (d, p) basis set was used for the entire calculations in both gas phase and water.

All these trial geometries were fully optimized at the B3LYP/6-31+G (d, p) level of theory in the gas phase. The frequency calculations in the gas phase show that all of these

trial structures are real local minima as no imaginary frequencies were observed in their frequency calculations. Again, once these gas phase structures were identified, the conformers were further subjected to solvation to determine or quantify the extent of deviation from the gas phase predictions, using the exact same strategy as in Chapter 4.

After re-optimizing the gas phase structures in water using the same level of theory, the structures were found to be real local minima, as no imaginary frequencies were observed in their frequency calculations. The population distribution of each conformer and the cis-trans ratio of the local minima conformers at the equilibrium were calculated using the Boltzmann distribution for the entire set of conformers regardless of their energy level for both gaseous and water. In analyzing the structural characteristics, all the existing hydrogen bonds and dipole moment were also identified and characterized for some of the most stable conformers. In both of, the most stable cis and trans conformers in water, the sugar exists in the chair conformation, which is in agreement with the experimental NMR data [100]. The backbone torsion and endocyclic angles were also characterized to determine the impact of the fused carbohydrate on the peptide backbone chain and prolyl amide. As in the previous chapter, the definitions of the backbone torsion and endocyclic angles are displayed in Figure 4.1 except that the carbohydrate is attached in a fused rather than spirocyclic manner.

5.1.1 Conformers and energies in gas and water

The relative energies, backbone torsion and endocyclic angles, and the conformational distributions of the most stable structures of *N-acetyl-GlcProH-NHMe* in gas phase and water are listed in Tables 5.1 and Tables 5.2, respectively. It is important to note that,

although the population distributions were performed for the entire populations of local minima conformers, only those with significant population contributions (population distribution of $> 0.5\%$) were included in the tables. In this chapter, the comparisons of these conformers are made based on the classification of the backbone torsion angle, endocyclic torsion angle, cis-trans prolyl amide bond, and the sugar residue conformation as overall criteria to categorize structures into sub groups, in complete analogy to previous work (Chapter 4).

Table 5.1 Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of *N*-acetyl-GlcProH-NHMe, which are Optimized at the B3LYP/6-31+G (d, p) Level of Theory in the Gas Phase.

B3LYP/6-31+G(d, p)				Backbone Torsion Angles and Endocyclic Torsion Angles ^a								
Conformers		Gas Phase (in Kcal/mole)	Conformer Distribution	ω'	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^0
		ΔE										
1	c- α_L [d]	0.0 ^b	35.513	3.423	-81.070	-10.744	-179.845	33.536	-39.155	28.853	-7.807	-16.186
2	c- α_L [d]	0.274222	22.363	3.247	-80.864	-10.746	-179.894	33.491	-39.294	29.075	-8.018	-16.042
3	c- ϵ_L [d]	0.650727	11.840	0.990	-70.492	152.098	174.133	33.022	-38.474	28.461	-7.752	-15.948
4	c- α_L [d]	0.859061	8.330	3.721	-80.304	-13.048	179.693	33.085	-39.732	30.104	-9.376	-14.912
5	c- ϵ_L [d]	0.948167	7.172	1.807	-69.640	151.216	175.402	32.448	-38.724	29.386	-9.082	-14.748
6	t- γ_L [d]	1.534261	2.667	175.508	-74.665	81.617	-176.711	28.454	-40.822	36.798	-19.445	-5.538
7	t- γ_L [d]	1.553086	2.582	174.784	-74.185	82.285	-176.553	28.682	-40.677	36.355	-18.869	-6.052
8	c- ϵ_L [d]	1.735064	1.899	2.047	-69.603	150.574	176.045	32.167	-38.985	30.032	-9.945	-14.019
9	c- ϵ_L [u]	1.768322	1.794	-5.697	-48.365	144.231	176.238	-27.478	33.555	-26.180	9.209	11.229
10	t- γ_L [u]	1.920807	1.388	-3.452	-59.340	-31.173	-177.536	-23.466	30.764	-25.587	11.269	7.399
11	c- ϵ_L [u]*	2.076429	1.067	-6.190	-48.127	144.668	175.880	-27.299	33.567	29.965	9.695	10.791
12	c- ϵ_L [u]	2.080822	1.059	-6.195	-47.945	144.480	175.062	-27.513	33.711	-26.486	9.567	11.005

13	c- γ_L [d]	2.1191	0.992	3.422	-80.087	-12.934	-179.562	33.212	-39.103	29.065	-35.159	-15.662
14	c- γ_L [u]	2.291665	0.742	-3.066	-58.786	-31.891	-177.282	-24.518	31.332	-25.452	10.435	8.566
15	c- γ_L [u]	2.433482	0.584	-3.506	-59.046	-31.195	-177.065	-23.891	31.046	-25.674	-20.098	7.761
Others*			< 0.5									

Total Population Distribution

	Gas Phase (%)	Exp. H ₂ O (%)
Total Cis Isomers	94.75	13
Total Trans Isomers	5.25	87

^a Defined in Figure 4.1; angles are in degrees, ^b $E = -1030.210072$ a.u. The population distributions were calculated using the Boltzmann statistical weights at 25 °C.

Table 5.2 Backbone Torsion Angles, Endocyclic Torsion Angles, Relative Energies (ΔE in kcal/mole), and Conformational Distribution of Cis-Trans Isomers of *N*-acetyl-GlcProH-NHMe, which are Optimized at the B3LYP/6-31+G (d, p) Level of Theory in the water.

B3LYP/6-31+G(d, p)				Backbone Torsion Angles and Endocyclic Torsion Angles ^a								
Conformers		H ₂ O	Conformer	ω'	ϕ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^0
		(in Kcal/mole)	Distribution									
		ΔE										
1	t- ϵ_L [d]	0.0 ^b	14.366	171.291	-55.901	143.586	175.616	21.665	-36.790	37.275	-24.766	2.147
2	t- ϵ_L [d]	0.006275	14.140	171.356	-55.767	143.237	175.840	21.754	-36.853	37.284	-24.716	2.062
3	t- ϵ_L [d]	0.037651	13.482	171.109	-56.264	144.331	175.046	22.745	-37.621	37.468	-24.232	1.124
4	t- ϵ_L [d]	0.150602	11.059	171.508	-56.771	144.756	174.845	22.505	-37.492	37.519	-24.449	1.412
5	c- ϵ_L [d]	0.181978	10.489	1.892	-69.015	150.736	175.275	30.844	-38.748	30.874	-11.640	-12.063
6	t- ϵ_L [d]	0.188253	10.466	171.310	-56.231	143.252	175.251	21.716	-36.942	37.440	24.858	2.177
7	t- ϵ_L [d]	0.370231	7.657	171.163	-56.380	144.396	175.099	22.809	-37.702	37.530	-24.236	1.084
8	c- ϵ_L [u]	0.596134	5.219	-8.044	-51.252	145.107	175.024	-25.22	32.656	-26.84	11.393	8.414
9	t- ϵ_L [d]	0.947539	2.890	175.536	-59.143	144.493	175.173	21.889	-36.849	37.168	-24.443	1.804
10	c- ϵ_L [d]	1.073041	2.340	1.810	-69.419	150.631	175.658	31.082	-38.805	30.739	-11.357	-12.399
11	t- ϵ_L [d]	1.085591	2.286	171.452	-56.107	143.639	175.590	22.293	-37.325	37.521	-24.597	1.627
12	t- ϵ_L [d]	1.173443	1.984	170.000	-56.454	144.799	175.281	22.979	-37.701	37.390	-23.959	0.785
13	c- ϵ_L [d]	1.242469	1.749	1.792	-69.419	150.581	175.675	31.045	-38.753	30.694	-11.334	-12.392

14	t- ϵ_L [d]	1.255019	1.714	171.499	-56.770	145.010	175.084	22.610	-37.452	37.360	-24.209	1.190
15	c- ϵ_L [d]	2.692016	0.152	0.905	-70.164	151.872	174.868	31.836	-38.941	30.211	-10.294	-13.564
Others*		< 0.5										

Total Population Distribution

	H ₂ O (%)	Exp. H ₂ O (%)
Total Cis Isomers	19.95	13
Total Trans Isomers	80.05	87

^a Defined in Figure 4.3; angles are in degrees, ^b $E = -1030.23190$ a.u. The population distributions were calculated using the Boltzmann statistical weights at 25 °C

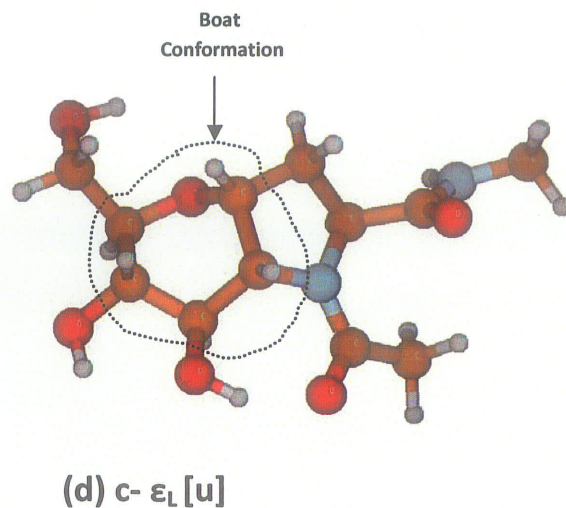
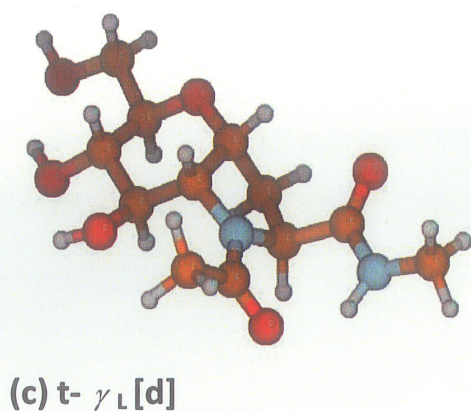
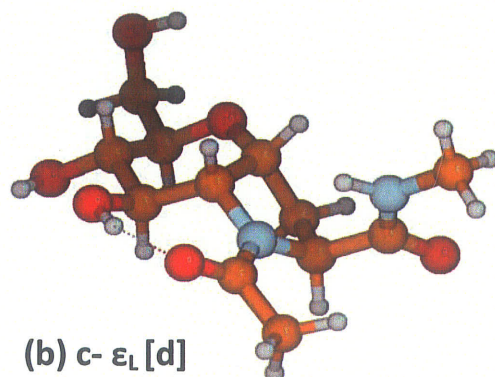
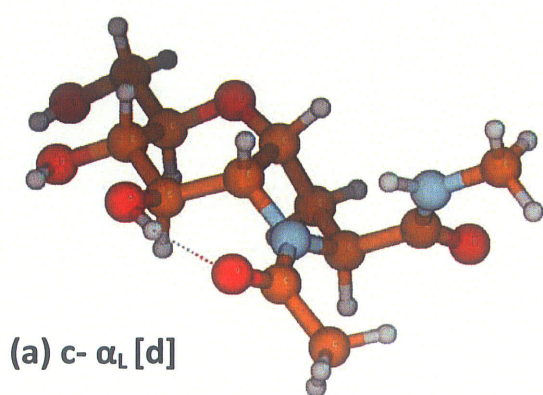
The population distributions in Table 5.1 and Table 5.2 of *N-acetyl-GlcProH-NHMe* optimized at the B3LYP/6-31+G (d, p) level of theory in gas phases and water, respectively, show that eight local minima $c-\alpha_L [d]$, $c-\epsilon_L [d]$, $t-\gamma_L [d]$, $c-\epsilon_L [u]$, $t-\gamma_L [u]$, $c-\epsilon_L [u]^*$, $c-\gamma_L [d]$, and $c-\gamma_L [u]$ in the gas phase calculation and only two local minima $t-\epsilon_L [d]$ and $c-\epsilon_L [d]$ in water were observed based on the classification of the backbone and endocyclic torsion angles, and the prolyl cis-trans isomerization. As in the previous analysis the $[u]^*$ again refers to the distorted puckering ring residue which resembles the up puckering. The conformations $c-\alpha_L [d]$ and $t-\gamma_L [d]$ are the two most stable cis and trans conformers in gas phase, respectively.

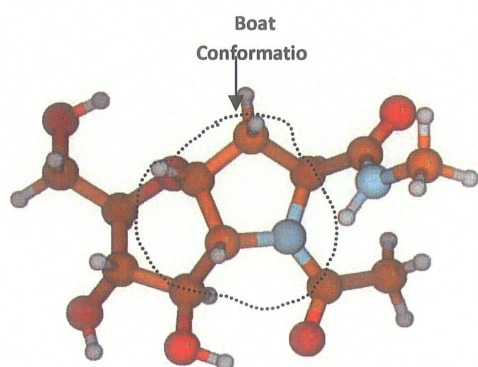
However, Table 5.2 shows different stable conformers of *N-acetyl-GlcProH-NHMe* in water. The two most stable cis and trans conformers obtained in water are $t-\epsilon_L [d]$ and $c-\epsilon_L [d]$, respectively. The relative stabilities of the conformations in the gas phase and water follow the order of $c-\alpha_L [d]$, $c-\epsilon_L [d] < t-\gamma_L [d] < c-\epsilon_L [u] < t-\gamma_L [u] < c-\epsilon_L [u]^* < c-\gamma_L [d] < c-\gamma_L [u]$, and $t-\epsilon_L [d] < c-\epsilon_L [d]$, respectively.

The relative energy difference between the most stable cis and trans conformers in gas phase was calculated to be 1.53 kcal/mole, which is by far larger than the difference observed in water, which is only approximately 0.18 kcal/mole. In water all the conformers listed in Table 5.2 confirm a down puckering or $C\gamma$ -endo configuration.

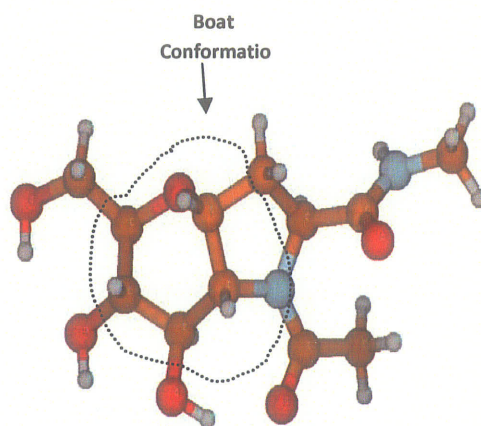
The backbone torsion angle of ω' , ϕ , ψ , and ω for the *cis* *N-acetyl-GlcProH-NHMe* conformers in gas phases spans the ranges of $-6^\circ \leq \omega' \leq 3^\circ$, $-47^\circ \leq \phi \leq -81^\circ$, $144^\circ \leq \psi \leq 152^\circ$ and $-31^\circ \leq \psi \leq -10^\circ$; and $\omega \approx -179^\circ$ and 175° , respectively. The endocyclic torsion angles χ_0 , χ_1 , χ_2 , χ_3 and χ_4 also range $-16^\circ \leq \chi_0 \leq 11^\circ$, $-27^\circ \leq \chi_1 \leq 33^\circ$, $-39^\circ \leq \chi_2 \leq$

33°, $-26^\circ \leq \chi_3 \leq 30^\circ$, $-35^\circ \leq \chi_4 \leq 11^\circ$, respectively. For the *trans*-N-acetyl-GlcProH-NHMe the backbone and endocyclic torsion angles ω' , ϕ , ψ , ω , χ_0 , χ_1 , χ_2 , χ_3 and χ_4 fall into the ranges of $174^\circ \leq \omega' \leq 175^\circ$, $\phi \approx -74^\circ$, $\psi \approx 82^\circ$, $\omega \approx -176^\circ$, $\chi_0 \approx -6^\circ$, $\chi_1 \approx 28^\circ$, $\chi_2 \approx -40^\circ$, $\chi_3 \approx 36^\circ$ and $\chi_4 \approx -18^\circ$, respectively. Correspondingly, in water the backbone and endocyclic torsion angles ω' , ϕ , ψ , ω , χ_0 , χ_1 , χ_2 , χ_3 and χ_4 for the *cis* N-acetyl-GlcProH-NHMe conformers span the ranges of $-8^\circ \leq \omega' \leq 0^\circ$, $-70^\circ \leq \phi \leq -51^\circ$, $145^\circ \leq \psi \leq 151^\circ$, $\omega \approx 175^\circ$, $-13^\circ \leq \chi_0 \leq 8^\circ$, $-25^\circ \leq \chi_1 \leq 31^\circ$, $-38^\circ \leq \chi_2 \leq 32^\circ$, $-26^\circ \leq \chi_3 \leq 30^\circ$, $-11^\circ \leq \chi_4 \leq 11^\circ$, respectively. For the *trans* N-acetyl-GlcProH-NHMe conformers, the angles span the ranges of $171^\circ \leq \omega' \leq 175^\circ$, $-59^\circ \leq \phi \leq -55^\circ$, $143^\circ \leq \psi \leq 145^\circ$, $\omega \approx 175^\circ$, $0^\circ \leq \chi_0 \leq 2^\circ$, $21^\circ \leq \chi_1 \leq 22^\circ$, $-38^\circ \leq \chi_2 \leq 36^\circ$, $36^\circ \leq \chi_3 \leq 37^\circ$, $-24^\circ \leq \chi_4 \leq 24^\circ$, respectively.

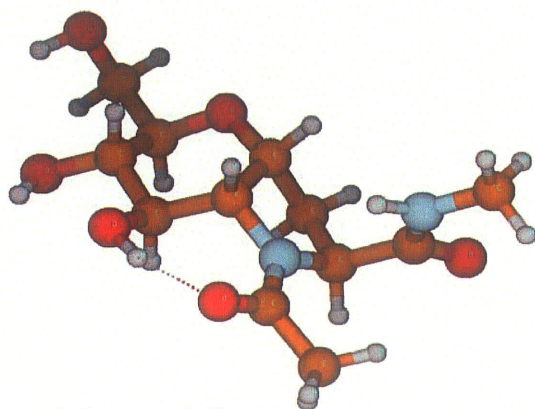




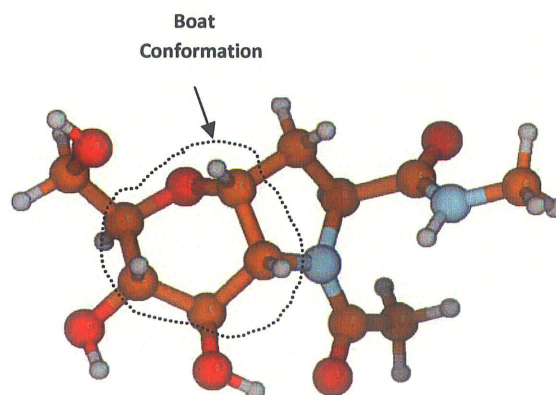
(e) $t\text{-}\gamma_L[u]$



(f) $c\text{-}\epsilon_L[u]^*$



(g) $c\text{-}\gamma_L[d]$



(h) $c\text{-}\gamma_L[u]$

Figure 5.1 The Minima energy representative conformations of *N*-acetyl-GlcProH-NHMe in a gas phase optimized at the B3LYP/6-31+G (d) Level of theory.

The representative conformations for the local minima conformers c- α_L [d], c- ϵ_L [d], t- γ_L [d], c- ϵ_L [u], t- γ_L [u], c- ϵ_L [u]*, c- γ_L [d], and c- γ_L [u] in the gas phase; and t- ϵ_L [d] and c- ϵ_L [d] in water optimized at the B3LYP/6-31+G (d) level of theory are displayed in Figure 5.1 and Figure 5.2, respectively.

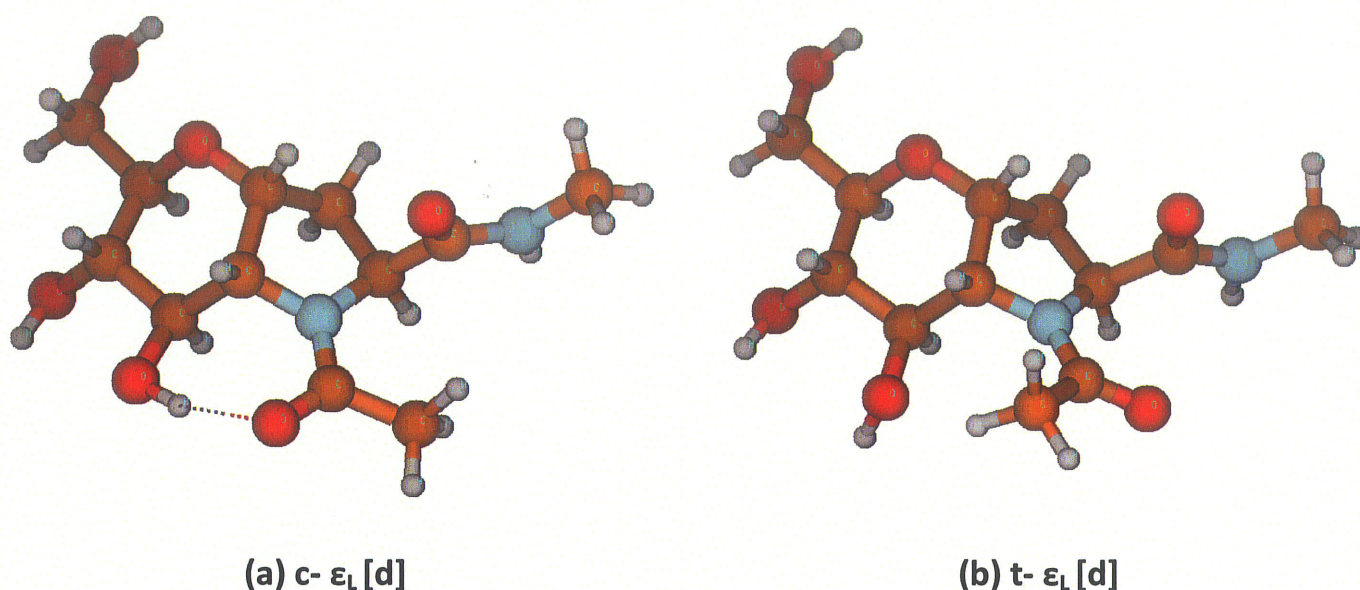


Figure 5.2 The Minima energy representative conformations of N-acetyl-GlcProH-NHMe in water optimized at the B3LYP/6-31+G (d) Level of theory

The geometric parameters for selected bond angles and lengths of N-acetyl-GlcProH-NHMe in water were also characterized to see if there is a big difference between the cis and trans conformers, as listed in Table 5.3 and Table 5.4. The bond angle and bond length show similar trends in the calculated values for both the bond angle and bond length despite a negligible deviation of 0.5° to 1° from one conformer to another. This

indicated that the incorporation of a fused carbohydrate template on the proline residue does not result in significant changes of the bond lengths and angles. Similar to the Compounds 1 and Compound 2 (Chapter 4) the bond angle $\angle C^{\delta} - N - C^{\alpha}$ generates a larger deviation of 8°-10° relative to other bond angles around the five membered ring, this is mainly due to the presence of the nitrogen atom. In short, the cis-trans isomerization has little influence on the stretching and compressing of the bond length around the five membered pyrrolidine ring.

Table 5.3 Selected Angles (in deg) for N-acetyl-GlcProH-NHMe, at the B3LYP/6-31+G (d, p) level of theory in the water

Conformers	$\angle N - C^{\alpha} - C^{\beta}$	$\angle C^{\alpha} - C^{\beta} - C^{\gamma}$	$\angle C^{\beta} - C^{\gamma} - C^{\delta}$	$\angle C^{\gamma} - C^{\delta} - N$	$\angle C^{\delta} - N - C^{\alpha}$
t- ϵ_L [d]	104.01	103.48	103.45	101.37	112.55
c- ϵ_L [d]	103.17	102.85	104.25	101.7	113.15

Table 5.4 Selected bond^a length (in Å) for N-acetyl-GlcProH-NHMe, at the B3LYP/6-31+G (d, p) level of theory in the water

Conformers	$C_{Ac} = O$	$C_{Ac} - N$	$N - C^{\alpha}$	$C^{\alpha} - C_{XX}$	$C^{\alpha} - C^{\beta}$	$C^{\beta} - C^{\gamma}$	$C^{\gamma} - C^{\delta}$	$C^{\delta} - N$
t- ϵ_L [d]	1.24	1.36	1.47	1.54	1.55	1.54	1.54	1.47
c- ϵ_L [d]	1.25	1.35	1.48	1.54	1.55	1.54	1.55	1.48

^a C_{Ac} and C_{XX} denoted for the carbon atoms of the Prolyl N-terminal Carbonyl Carbon and Prolyl N-terminal Acetate carbon

5.1.2 Conformational distribution of N-acetyl-GlcProH-NHMe in gas and water

The conformational distributions of the conformers of N-acetyl-GlcProH-NHMe in gas phase and water obtained show trends that are opposite to one another. Indeed, in the gas phase, the dominating conformation is trans, whereas in solution, cis is dominating (Figure 5.3). The DFT calculation in water came in close agreement with the experimental data compared to the DFT calculation obtained in the gas phase. Figure 5.3 shows the population distribution of the cis and trans conformers of N-acetyl-GlcProH-NHMe.

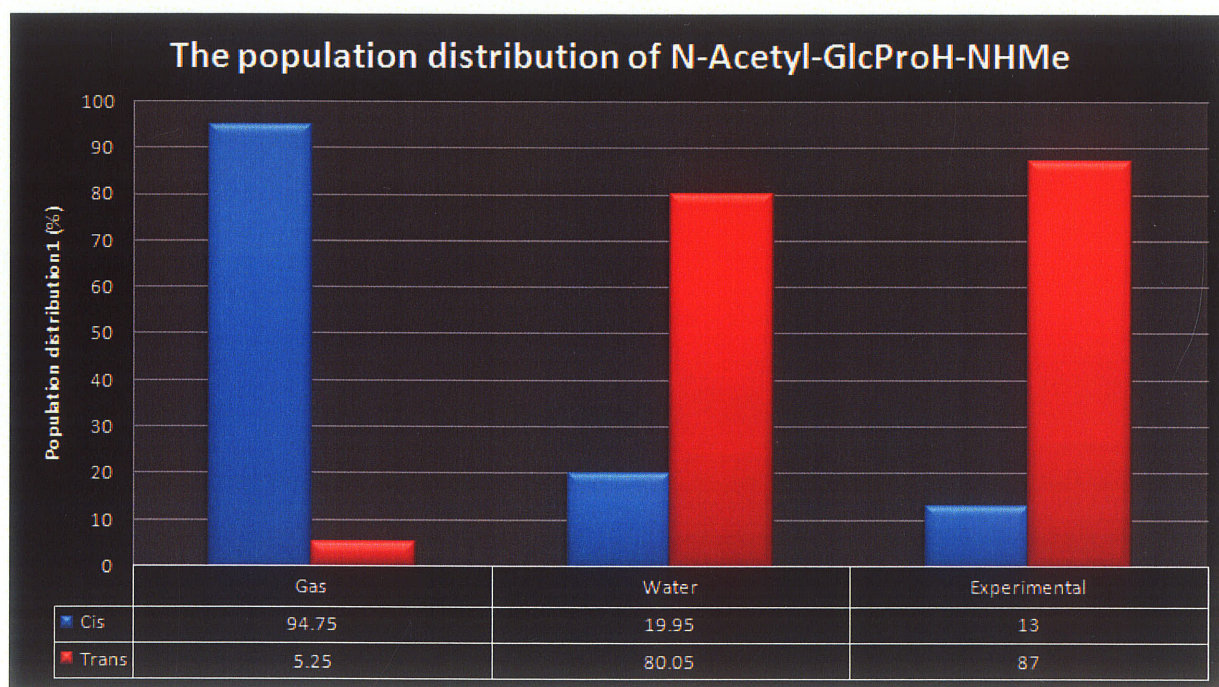


Figure 5.3 The cis and trans population distribution of N-acetyl-GlcProH-NHMe, each population was computed using the Boltzmann weight by the relative energy at the B3LYP/6-31+G (d) Level.

The population of the cis and trans conformers are 94.75%, 5.25% in the gas phase and 19.95%, 80.05% in water, respectively. In addition, the population distributions of the backbone and endocyclic conformations for *N-acetyl-GlcProH-NHMe* for the gas phase and for water calculation were also shown in Figure 5.4.

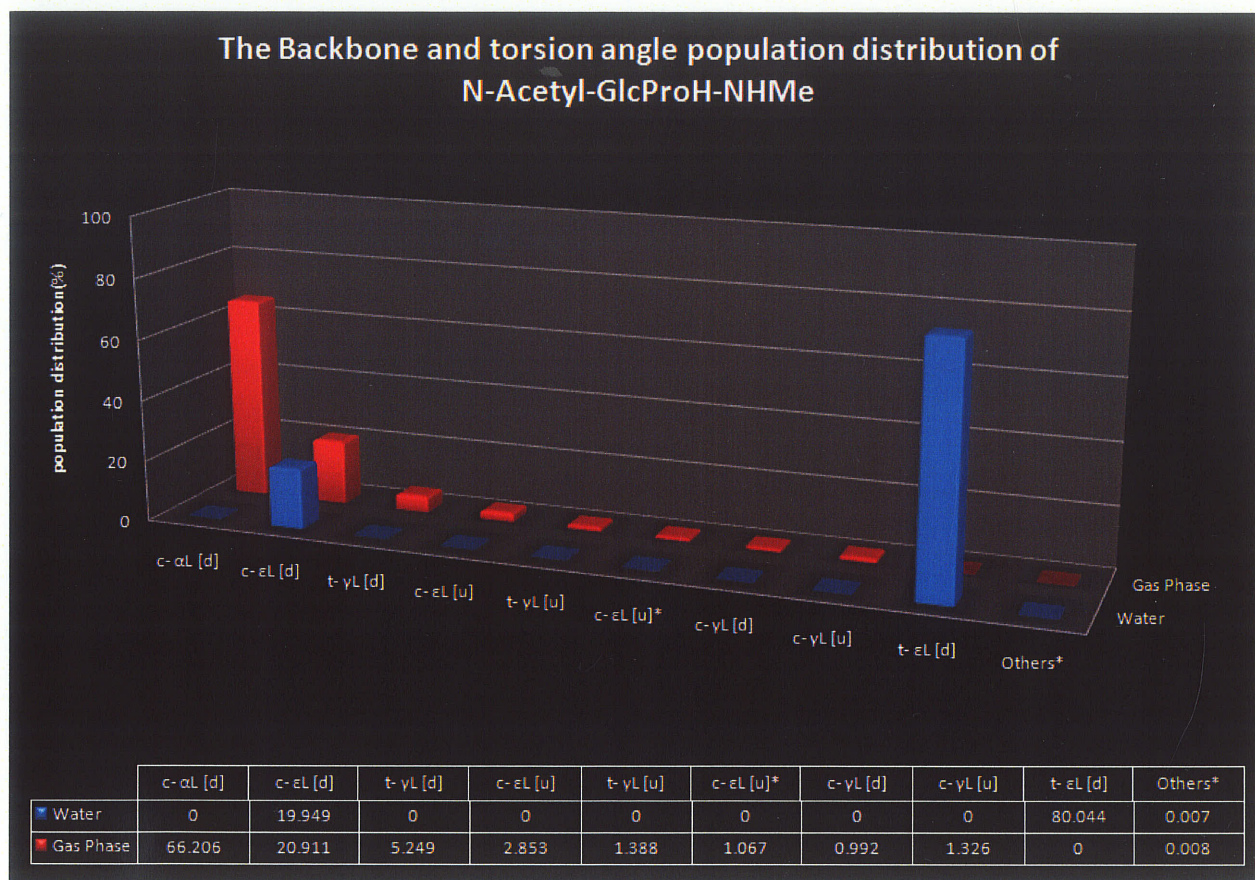


Figure 5.4 The backbone and endocyclic population distribution of *N-acetyl-GlcProH-NHMe*, each population was computed using the Boltzmann weight by the relative energy at the B3LYP/6-31+G (d) Level.

The population distribution of the trans conformation t- ϵ_L [d] and the cis conformation c- ϵ_L [d] constitute the majority of the populations in water, with a population distribution of approximately 80.00% and 19.95.6%, respectively. However, in the gas phase the most populated conformers are c- α_L [d] and c- ϵ_L [d] for the cis conformations and t- γ_L [d] for the

trans conformations. These conformer $c-\alpha_L$ [d], $c-\epsilon_L$ [d] and $t-\gamma_L$ [d] constitute a population distribution of approximately 66.2%, 20.9% and 5.25%, respectively. Similar to the calculations obtained for Compound 1 and Compound 2, the backbone and endocyclic torsion angles show an opposite trend in gas and water. Similarly, the entry “others*” (Figure 5.4) indicates those population distributions of the backbone and endocyclic conformations which are not listed in Table 5.1 and Table 5.2.

5.1.3 Intramolecular hydrogen bond and ring puckering in water

The internal hydrogen bonding for the most stable conformers of *N-acetyl-GlcProH-NHMe* in water were also investigated. Results showed that only one type of hydrogen bonding was observed that contribute a population distribution greater > 0.5% using a bond length cutoff of $\leq 2.5\text{\AA}$ for the conformations. This hydrogen bond exists between the Prolyl N-terminal Carbonyl Oxygen and the sugar 2C hydroxyl group ($C_\beta\text{-}2\text{C-OH} \text{-----O=C'-}3'\text{N}$, see figure 5.5). The bond distance is $\approx 1.79\text{\AA}$; the conformers with this type of hydrogen bond are conformers 5, 8, 23, 10, 13 and 15, of a population distribution of approximately 19.94% of the total distribution. All these six conformers are in favor of *cis* position and no hydrogen bonds were observed for the trans conformers.

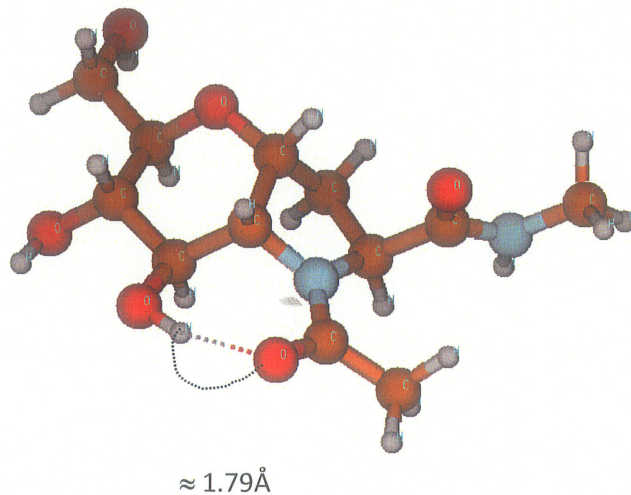


Figure 5.5 The hydrogen bond found in *N-acetyl-GlcProH-NHMe*

Table 5.5 Backbone Dihedral angles (in degree), Pseudorotational Parameters (A and P, in degree), Relative Energies (ΔE , in Kcal/mole) and dipole moment for *N-acetyl-GlcProH-NHMe* for the minima energy representative conformations at the B3LYP/6-31+G (d) level of theory in the water

Conformation	ω'	ϕ	ψ	ω	(A, P)	ΔE	Dipole ^a Moment
--------------	-----------	--------	--------	----------	--------	------------	-------------------------------

N-acetyl-GlcProH-NHMe

1	t- ϵ_L [d]	171.291	-55.901	143.586	175.616	(34.4,-67.7)	0.0	5.00
2	c- ϵ_L [d]	1.892	-69.015	150.736	175.275	(33.4,-68.3)	0.18	10.4

^a units in Debye (D)

The puckering of the five-membered prolyl residue was also calculated using the classical pseudorotational parameters, the puckering amplitude (A) and the state of the pucker in the pseudorotational pathway (P), as described in Table 5.5.

The puckering amplitude (A) and the state of the pucker (P), for both cis and trans conformers of *N-acetyl-GlcProH-NHMe* shows almost the same values, with a range of $\pm 1^\circ$. This indicated that the A, P values are not significantly changed by the cis to trans isomerization of the prolyl residue. From Table 5.5, it is clearly shown that there is a large difference in the dipole moment of the cis and trans conformers. The cis conformer shows a very large dipole moment compared to the trans conformer for the most stable conformers of *N-acetyl-GlcProH-NHMe*.

5.2 Conclusions

The conformational distribution for the *N-Acetyl-GlcProH-NHMe in water* for the cis isomers at 25° temperature were found to be approximately 19.95%, which is in good agreement with the observed experimental result of 13% for the proline residue. The population distribution obtained in gas phase does not match the experimental results and predicts the opposite of the experimental observed data. The main reason for this is not clear yet. The intramolecular hydrogen bond observed for *N-acetyl-GlcProH-NHMe* only exists for the cis conformation. There is no significant shift in the puckering amplitude between the cis and trans conformers for this compound.

Chapter 6

General Conclusions

6.1 Summary and Future Works

This research project was aimed at the conformational preference determination and characterization of some novel carbohydrate templated proline mimetics using computational tools. A number of mathematical models and tools were used for this study and compared based on their relevance and computational costs. One of the most important observations in this project is that it is very vital at the beginning of the project to develop some clear methodologies first by comparing different computational tools in order to determine which method is applicable or not for a particular system of interest. Sometimes, numbers generated from those computational tools could mean *nothing* and always need to be confident in the numerical results generated from these tools and the methodologies developed before any conclusion can be drawn. In this project, a reliable computational protocol was developed for the accurate prediction of the conformational distribution calculations in particular for the carbohydrates and peptides. The protocol developed in this study consist of a systematic scanning of the PES using Monte Carlo via the build and search methodology discussed in section 3.6.

A number of conclusions from this study could be drawn. The biggest surprise in this research is that the trend obtained using the gas phase calculations and water showed completely opposite results in the B3LYP/6-31+G (d, p) level of theory. So it is always important to use solvent for modeling peptides and proteins as the gas phase calculations completely contradict the

experimental results. Interestingly, the density functional theory calculations in water always produce results that are in a good agreement with the experimental data [100,101]. For example, the conformational distribution for the *N-acetyl-GlcProH-NHMe in water* for the cis isomers at 25°C temperature were found to be approximately 19.95%, which is in excellent agreement with the observed experimental result of 13%. While the gas phase calculation prediction for the cis conformer is 94.75%, which is almost equivalent with the trans population observed experimentally.

The DFT calculations in water for both Compound 1 and Compound 2 showed that the position of the C_6 primary hydroxyl group may play a great role in influencing the cis-trans isomerization via the intramolecular hydrogen bonding. In addition, it is also observed that the position of the C_6 primary hydroxyl group has a tremendous effect in distorting the five-membered ring.

Computational results for Compound 1 and Compound 2 showed that the most stable conformers obtained are t- ϵ_L [d] and c- ϵ_L [d], respectively. Both conformers exhibit a C_γ -endo puckering and a ϵ_L backbone conformation which leads to polyproline II (P_{II}). The conformers listed in Table 4.1, Table 4.2, Table 4.3 and Table 4.4 show that the conformation of the six-membered sugar residue in all these conformers is in a *chair* configuration.

The results obtained from this calculation produce encouraging results and drawn some conclusions based on these primary findings; however, more work should be done on the kinetic isomerization study of these conformers, like calculating the transition states, the enthalpy, Gibbs free energy etc. In particular, finding the rates of the cis-trans isomerization would also help in understanding how the position of the C_6 primary hydroxyl group affects the free energy barrier

to isomerization. Future work may also include extending this project to other novel proline based analogues or any peptide compounds. Finally, the future direction of this project will include developing a mathematical model that predicts the energy of a compound in different solutions directly from the gas phase calculations by estimating the energy of the gas phase and dipole moment.

Bibliography

1. Champe, C. P.; Harvey, A. R.; Ferrier, R. D. *Biochemistry: Lippincott's illustrated Reviews*, 3rd ed.; Lippincott Williams & Wilkins: New York, 2005.
2. Devlin, M. T. *Biochemistry: with clinical correlations*, 5th ed.; John Wiley & Sons: New York, 2002.
3. Bruice, Y. P. *Organic Chemistry*, 5th ed.; Pearson Education, Inc: London, 2004.
4. Voet, D.; Voet, G. J. *Biochemistry*, 3rd ed.; John Wiley & Sons, Inc: New Jersey, 2004.
5. Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. *Protein Science*. **2001**, 10, 2627.
6. Oberholser, K. M. Messiah College (2007). *The Collagen Structures, triple helix*. 3D Images Retrieved on Permission on June 2, 2009 from http://www.messiah.edu/departments/chemistry/molscilab/Jmol/collagen/collagen_index.htm
7. Benzi, C.; Improta, R.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2002**, 23, 341.
8. Sak, K.; Karelson, M.; Jarv, J. *Int J Quantum Chem.* **1998**, 66, 391.
9. Sahai, M. A.; Kehoe, T. A. K.; Koo, J. C. P.; Setiadi, D.H.; Chass, G. A.; Viskolcz, B.; Penke, B.; Pai, E. F.; Csizmadia, I. G. *J. Phys. Chem. A*. **2005**, 109, 2660.
10. Song, K.; Kang, K. J. *J. Phys. Chem. B*. **2005**, 109, 16982.
11. Gibbs, A.C.; Bjorndahl, T.C.; Hodges, R.S.; Wishart, D.S. *J. Am. Chem. Soc.* **2002**, 124, 1203.
12. Némethy, G.; Printz, P. M. *Macromolecules*. **1972**, 5, 755.
13. Madison, V. *Biopolymers*. **1977**, 31, 2671.
14. Chang, D. K.; Cheung, S. F.; Trivedi, V. D. *J. Struct. Biol.* **1999**, 128, 270.
15. Wess, J.; Nanavati, S.; Vogel, Z.; Maggio, R. *EMBO J.* **1993**, 12, 331.
16. Vanhoof, G.; Goosens, F.; De Meester, I.; Hendriks, D.; Scharpe, S. *FASEB J.* **1995**, 9, 736.
17. Zarrinpar, A.; Bhattacharyya, P. R.; Lim, W. A. *Sci.STKE*, **2003**, 179, re8.
18. Palifi, K. Villo.; Perczel, A. *J. Comput Chem.* **2008**, 29, 1374.
19. Kang, K. Y. *J. Phys. Chem. B*. **2006**, 110, 21338.
20. Wedemeyer, J. W.; Welker, E.; Scheraga, A. *Biochemistry*. **2002**, 41, 14639.

21. Zimmerman, S. S.; Scheraga, H. A. *Macromolecules*, **1976**, 9, 408.
22. Wüthrich, K.; Grathwohl, C.; Schwyzer, R. *Peptides, Polypeptides and Proteins*, Wiley: New York, 1974.
23. Steinberg, I. Z.; Harrington, W. F.; Berger, A.; Sela, M.; Katchalski, E. *J. Am. Chem. Soc.* **1960**, 82, 5263.
24. Pincus, M. R.; Gerewitz, F.; Wako, H.; Scheraga, H. A. *J. Protein Chem.* **1983**, 2, 131.
25. Kang, K. Y.; *J. Phys. Chem. B.* **2002**, 106, 2074.
26. Improta, R.; Benzi, C.; Barone, V.; *J. Am. Chem. Soc.* **2001**, 123, 12568.
27. Madison, V.; Kopple, K. D. *J. Am. Chem. Soc.* **1980**, 102, 4855.
28. Schmid, F. X.; Mayr, L.M.; Mücke, M.; Schönbrunner, E.R. *Advance Protein Chem.* **1993**, 44, 25.
29. Pahlke, D.; Leitner, D.; Wiedemann, U.; Labudde, D. *Bioinformatics.* **2005**, 21, 685.
30. Lu, K. P.; Finn, G.; Lee, H. T.; Nicholson, K. L. *Nature Chemical Biology.* **2007**, 3, 619.
31. Stein, R. L. *Advance Protein Che.* **1993**, 44, 1.
32. Harding, M. W.; Galat, A.; Uehling, D. E.; Schreiber, S. L. *Nature*, **1989**, 341, 758.
33. Fischer, G.; Wittmann-Liebold, B.; Lang, K.; Kiefhaber, T.; Schmid, F. X. *Nature*, **1989**, 337, 758.
34. Takahashi, N.; Hayano, T.; Suzuki, M. *Nature*. **1989**, 337, 473.
35. Almawi, W. Y.; Melemedjian, O. K. *Nephrol Dial Transplant.* **2000**, 15, 1916.
36. Zhuo, X. Z.; Lu, P. J.; Wulf, G.; Lu, K. P. *Cell. Mol. Life Sci.* **1999**, 56, 788.
37. Wulf, G.; Finn, G.; Suizu, F.; Lu, K. P. *Nat Cell Biol.* **2005**, 7, 435.
38. Lu, K. P.; Zhou, X. Z. *Nat Rev. Mol. Cell Biol.* **2007**, 11, 904.
39. Zimmerman, S.S.; Pottle, M.S.; Nemethy, G.; Scheraga, H. A. *Macromolecules.* **1977**, 10, 1.
40. Perczel, A.; Angyan, G. J.; Kajtar, M.; Viviani, W.; Rivail, J. -L.; Marcoccia, J. -F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, 113, 6256.
41. Perczel, A.; McAllister, A.; Császár, M.; Csizmadia, I. G. *Can J. Chem. Soc.* **1994**, 72, 2050.
42. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, J. V. *Mol. Biol.* **1963**, 7, 95
43. Perczel, A.; McAllister, A.; Császár M.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1993**, 115, 4849

44. Enriz, R. D.; Morales, M. E.; Baldoni, H. A.; Freile, M. L. *J. Argent. Chem. Soc.* **2006**, 94, 1.
45. Fejer, S. N. ; Jenei, Z. A. ; Paragi, G. *J. Mol. Struct. THEOCHEM.* **2003**, 303, 666.
46. Flory, P. J. *Statistical Mechanics of chain Molecules*, 1969, Interscience, p. 253. Ref4
47. IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry.* 1970, **9**, p. 3475.
ref4
48. Kilpatrick, J. E.; Spitzer, R. J. *Chem. Phys.* **1946**, 14, 463.
49. Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, K. S. *J. Am. Chem. Soc.* **1947**, 69, 2483.
50. Westhof, E.; Sundaralingam, M.; *J. Am. Chem. Soc.* **1977**, 99, 1232.
51. Ramek, M.; Kelterer, A. M.; Teppen, B. J.; Schafer, L. *J. Mol. Struct. (THEOCHEM).* **1995**, 352, 59
52. Hudaky, I.; Baldoni, H. A.; Perczel, A. *J. Mol. Struct (THEOCHEM).* **2002**, 582, 233.
53. Cremer, D., and Pople, J. A., *J. Am. Chem. Soc.* 1975, **97**, 1354.
54. Geise, H. J.; Altona, C.; Romers, C. *Tetrahedron Lett*, **1967**, 15, 1383.
55. Dunitz, J. D.; *Tetrahedron*, **1972**, 28, 5459.
56. Han, S. J.; Kang, Y. K. *J. Mol, Struct. (THEOCHEM).* **1996**, 362, 243.
57. Hudaky, I.; Perczel, A. *J. Mol, Struct. (THEOCHEM).* **2003**, 630, 135.
58. Jabs, A.; Weiss, M.S.; Hilgenfeld, R. *J. Mol. Biol.* **1999**, 286, 291.
59. Kramer, Z. R.; Vitagliano, L.; Bella, J.; Berisio, R.; Mazzarella, L.; Brodsky, B.; Zagari, A.; Berman, H. M.; *J. Mol, Biol.* **1998**, 280, 623.
60. Okuyama, V.; Nagarajan, V.; Kamitori. *Proc. Indian Acad. Sci (Chem. Sci.).* **1999**, 111, 19.
61. Engel, J.; Chen, T. H.; Prockop, J. D.; Klump, H.; *Biopolymers.* **1977**, 16, 601.
62. Vitagliano, L.; Berisio, R.; Mazarella, L.; Zagari, A.; *Biopolymers.* **2001**, 58, 459.
63. Berisio, R.; Vitagliano, L.; Sorrentino, G.; Carotenuto, L.; Piccolo, C.; Mazarella, L.; Zagari, A. *acta Crystallogr., Sect. D.* **2000**, 56, 55.
64. Deber, C. M.; Bovey, F. A.; Carver, J. P.; Blout, E. R. *J. Am. Chem. Soc.* **1970**, 92, 6191.
65. Bhattacharjee, A.; Bansal, M.; IUBMB Life. **2005**, 57, 161.
66. Shoulders, D. M.; Guzei, A. I.; Raines, T. R. *Biopolymers.* **2007**, 89, 5.
67. Myllyharju, J.; Kivirikko, I. K.; *Ann. med.* **2001**, 33, 7.
68. Cara, L. J.; Raines, L. J. *Nat. Prod. Rep.* **2002**, 19, 49.

69. Veit, G.; Kobbe, B.; Keene, D. R.; Paulsson, M.; Koch, M.; Wagener, R. *J Biol Chem.* **2006**, 281, 3494.
70. Ramachandran, G.N.; Kartha, G.; *Nature.* **1954**, 174, 269.
71. Rich, A.; Crick, F. H. C. *Nature*, **1955**, 176, 915.
72. Bella, J.; Eaton, M.; Brodsky, B.; Berman, M. H. *Science.* **1994**, 266, 75.
73. Gustavson, H. K. in *Connective Tissue*, ed. R. E. Tunbridge, Blackwell: Oxford, **1975**.
74. Suzuki, S.; Fraser, R. D. B.; MacRae, T. P. *Int. J. Biol. Macromol.* **1980**, 2, 54.
75. Bella, J.; Brodsky, B.; Berman, H. M.; *Structure.* **1995**, 3, 893.
76. Steimberg, I.Z.; Berger, A.; Katchalski, E.; *Biochem. Biophys. Acta.* **1958**, 28, 648.
77. Steimberg, I.Z.; Harrington, W.F.; Berger, A.; Sela M.; Katchalski, E. *J. Am. Chem. Soc.*, **1960**, 82, 5263.
78. Brandt, J.F.; Halvoson H.R.; Brennan, M. *Biochemistry.* **1978**, 14, 4953.
79. Hardin, M.W.; Galat, A.; Vehling D.E.; Schreiber, S.L. *Nature.* **1989**, 341, 758.
80. Siekierka, J.J.; Hung, S.H.; Poc, M.; Lin, C.S.; Sigal, N.H. *Nature.* **1989**, 341, 775.
81. Beausoleil, E.; Lubell, W.D. *J. Am. Chem. Soc.* **1996**, 118, 12902. And references therein.
82. Delaney, N. G.; Madison, V. J. *J. Am. Chem. Soc.* **1982**, 104, 6635.
83. Samanen, J.; Zuber, G.; Bean, J.; Eggleston, D.; Romoff, T.; Kopple, K.; Saunders, M.; Regoli, D. *Int. J. Pept. Protein. Res.* **1990**, 35, 501.
84. Quancard, J. ; Labonne, A. ; Jacquot, Y. ; Chassaing, G.; Lavielle, S. ; Karoyan, P. *J. Org.Chem.***2004**, 69, 7940
85. Che, Y.; Marshall, G. R. *J. Org. Chem.* **2004**, 69, 9030. And references therein.
86. Tam, J. P.; Miao, Z.; *J. Am. Chem. Soc.* **1999**, 121, 9013. And references therein
87. Cavelier, F. ; Vivet, B. ; Martinez. J.; Aubry, A.; Didierjean, C.; Vicherat, A.; Marraud, M. *J. Am. Chem. Soc.* **2002**, 124, 2917. And references therein.
88. Sharma, R.; Lubell, W.D. *J. Org. Chem.* **1996**, 61, 202. And references therein.
89. Jeannotte, G.; Lubell, W. D. *J. Org. Chem.* **2004**, 69, 4656.
90. (a) Wagaw, S.; Rennels, R. A.; Buchwald, S. L. *J. Am. Chem. Soc.* **1997**, 119, 8451; (b) Kuwano, R.; Sato, K.; Kurokawa, T.; Karube, D.; Ito, Y. *J. Am. Chem. Soc.* **2000**, 122, 7614; (c) Viswanathan, R.; Prabhakaran, E. N.; Plotkin, M. A.; Johnston, J. N.; *J. Am. Chem. Soc.* **2003**, 125, 163.

91. Koep, S.; Gais, H.-J.; Raabe, R. *J. Am. Chem. Soc.* **2003**, 125, 13243. And references therein.
92. Dumy, P.; Keller, M.; Ryan, D. E.; Rohwedder, B.; Wöhr, T.; Mutter, M. *J. Am. Chem. Soc.* **1997**, 119, 918.
93. Keller, M.; Sager, C.; Dumy, P.; Schutkowski, M.; Fisher, G. S.; Mutter, M. *J. Am. Chem. Soc.* **1998**, 120, 2714. And references therein.
94. Wöhr, T.; Wahl, F.; Nefzi, A.; Rohwedder, B.; Sato, T.; Sun, X.; Mutter, M. *J. Am. Chem. Soc.* **1996**, 118, 9218.
95. Maigret, B.; Perahia D.; Pullman, B. *J. Theor. Biol.* **1970**, 29, 275.
96. Farmer B.L.; Hopfinger, A. *J. Macromolecules*, **1974**, 7, 793.
97. Fischer, S.; Dunbrack Jr, R.L.; Karplus, M. *J. Am. Chem. Soc.* **1994**, 116, 11931.
98. Kang, Y.K.; Choi, H.Y. *Biophys Chem.* **2004**, 111, 135.
99. Che, Y.; Marshall, G. R. *J Org Chem.* **2004**, 69, 9030.
100. Owens, N.; Braun, C.; Schweizer, F. J. *Org. Chem.* **2007**, 72, 4635.
101. Zhang, K.; Schweizer, F. *Syntett.* **2005**, 3111.
102. Jensen. F. *Introduction to Computational Chemistry*. 1st ed.; Wiley & Sons: New York, 1998.
103. Cramer. C. *Essentials of Computational Chemistry: Theories and Models*. 2nd ed.; Wiley & Sons: New York, 2004.
104. Young, C. D. *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*, Wiley & Sons: New York, 2001.
105. Hagler, A. T.; Huler, E.; Lifson, S. *J. of Amer. Chem. Soc.* **1977**, 96, 5319.
106. Halgren, T. A. *C. Opin. Struct. Bio.* **1995**, 5, 205.
107. Halgren, T. A. *J. Comput. Chem.* **1996**, 17, 490.
108. Halgren, T. A. *J. Comput. Chem.* **1996**, 17, 520.
109. Halgren, T. A. *J. Comput. Chem.* **1996**, 17, 553.
110. Clark, M.; Cramer III, R.; Opdensch, N. *J. Computational Chem.* **1989**, 10, 982
111. Clark, T. *A Handbook of computational Chemistry: A Practical Guide to Chemical Structure and Energy Calculations*. Wiley-Interscience: New York, 1985.
112. Fletcher, R. *Practical Methods of Optimization*. 2nd ed.; Wiley & Sons: New York: 1980.

113. Hehre, W.; Radom, L.; Schleyer, P.; Pople, J. *Ab-initio Molecular Orbital Theory*. John Wiley & Sons: New York, 1986.
114. Faegri, K.; Almlöf, J. *J. Chem. Phys. Lett.* **1984**, 107, 121.
115. Lüthi, H. P.; Siegbahn, P.; Almlöf, J.; Faegri, K. *J. Chem. Phys.* **1982**, 77, 2002.
116. Koch, W.; Holthausen, C. *A Chemist's Guide to Density Functional Theory*. 2nd ed.; Weinheim: Wiley-VCH, 2001.
117. Truhlar, G. D. *Potential Energy Surface and Dynamics calculations for Chemical Reactions and Molecular Energy Transfer*. Plenum Press, 1981.
118. Parr, G.; Yang, W. *Density Functional Theory of Atoms and Molecules*. Oxford University Press: Oxford, 1989.
119. Kryachko, E. S.; Ludeña, E. V.; *Density Functional Theory of Many Electron Systems*, Dordrecht: Kluwer Academics Press, 1990.
120. Ziegler, T. *Chem. Rev.* **1991**, 91, 651
121. Hehre, W. *A Guide to Molecular Mechanics and Quantum Chemical Calculations*. Wavefunction, Inc: CA, 2003.
122. Burkert, U.; Allinger, N. *Molecular Mechanics: American Chemical Society*. Vol. 177, Washington, DC, 1982.
123. Lewars, E. G. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, Springer: London, 2003.
124. Momany, F.; Rone, R. *J. Comput. Chem.* **1992**, 13, 188.
125. Davies, E. K.; Marrall, N. W. *J. Comput. Chem.* **1989**, 13, 149.
126. Allinger, N.; Yuh, Y.; and Lii, J. *J. Am. Chem. Soc.* **1989**, 111, 8551.
127. Mayo, S. L.; Olafson, B.D.; Goddard, W. *J. Phys. Chem.* **1990**, 94, 8897.
128. Daura, X.; Mark, A. E.; van Gunsteren, W. *J. Comput. Chem.*, **1998**, 19, 535.
129. Weiner, S.; Kollman, P.; Nguyen, D.; Case, D. *J. Comput. Chem.* **1986**, 7, 230.
130. Sunggyu, L.; Lee, L. *Encyclopedia of Chemical Processing*. Vol 1, CRC Press, 2005.
131. Chankvetadze, B.; *Capillary Electrophoresis in Chiral Analysis*. Wiley & Sons: London, 1997.
132. Armsrong, D.; Ward, R.; Armstrong, R.; Beesley, T. *Science*, 1986, 232, 1132.
133. Lipkowitz, B.; Ragothama, S.; Yang, J. *J. Am. Chem. Soc.*, **1992**, 114, 1554.
134. Durham, G.; Liang, G. *Chirality*. **1994**, 6, 239.

135. Gardner, M.; Vinter, G. *Molecular Modelling and Drug Design*. CRC Press LLC, 1994.
136. Lee, G. D. *Global Drug Enforcement*. CRC Press, Taylor and Francis group, 2003.
137. Leach, A.; Gillet, V. *An introduction to Chemoinformatics*, Springer: London, 2003.
138. Foresman, B. J.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*. 2nd ed.; Gaussian, Inc: Pittsburgh, PA, 1993.
139. Thomas, L. *Proc. Cambridge Philos. Soc.* **1926**, 23, 542.
140. Fermi, E. *Phys.* **1928**, 48, 73.
141. Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, B 136, 864.
142. Kohn, W.; Sham, L. *Phys. Rev.* **1965**, A 140, 1133.
143. Beck, A. D. *Phys. Rev. A.* **1998a**, 88, 1058.
144. Beck, A. D. *Phys. Rev. A.* **1998b**, 38, 3098.
145. Beck, A. D. *Phys. Rev. A.* **1998c**, 38, 2547.
146. Lee, C.; Yang, W.; Parr, R.G. *Phys. Rev. B.* **1988**, 37, 785.
147. SPARTAN '02 Build 119. Wavefunction, Inc: Irvine, CA, 2002.
148. Gaussian 03, Revision C.02, Frisch, M.J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr. T.; Kudin, K. N.; Burant, J. C.; Millam J. M.; Iyengar, S. S.; Tomasi, J.; Barone V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda R.; Hasegawa J.; Ishida M.; Nakajima T.; Honda Y.; Kitao O.; Nakai H.; Klene, M.; Li X.; Knox J. E.; Hratchian H. P.; Cross J. B.; Bakken V.; Adamo C.; Jaramillo, J.; Gomperts, R.; Stratmann R. E.; Yazyev O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004.
149. Morozov, A.; Kortemme, T.; Tsemekhman, K.; Baker, D. *Proc.Natl.Acad. Sci. U S A* **2004**.101, 6946.