# Molecular Orbital Studies of Collagen and Tri-Alanine Peptides

by

**I-Hsien (Midas) Tsai**

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

**2012**

This manuscript has been read and accepted for the Graduate Faculty in Chemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

**Prof. Joseph J. Dannenberg**

Date — Chair of Examining Committee

**Prof. Maria Tamargo**

Date — Executive Officer

**Prof. Alexander Greer**

**Prof. Yujia Xu**

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

# Abstract

Molecular Orbital Studies of Collagen and
Tri-Alanine Peptides

By

I-Hsien (Midas) Tsai

Advisor: Prof. Joseph J. Dannenberg

The purpose of this thesis is to study the stability of the triple helix (collagen) like peptide structure using computational methods. I am focusing on the collagen stability by using molecular orbital (MO) methods to compare the energies of interaction of Collagen and Tri-Alanine peptides using density functional theory at the B3LYP/D95(d,p) level on the gas phase, aqueous solvation and solvated energies. I present the overall interaction energies as broken down into pure H-bonding between the strands at the geometries they assume in the triple helix.

## Acknowledgments

I would like to express my heartfelt gratitude to my mentor Professor Joseph J. Dannenberg for his support, guidance and patience during this thesis studies. I am also thankful to my Committee members, Prof. Yujia Xu and Prof. Alexander Greer, for their expertise and suggestions. To addition, I would like to appreciate to Prof. Antoni Oliva and Dr. Dipankar Roy for their patience to check my grammar of this thesis. To Prof. Klaus Grohmann and Prof. Pamela Mills, the Graduate Student Advisor at the Chemistry Department of Hunter College – for all their support during all these years

# Table of Contents

**Chapter** **Page**

# List of Tables

## *Chapter 2*



## *Chapter 3*



## *Chapter 4*

## *Chapter 5*

# List of Figures

## *Chapter 1*



## *Chapter 2*

## *Chapter 3*



## *Chapter 4*

## *Chapter 5*

# Chart

# Chapter 1

## 1. Introduction

Collagen is an abundant fibrous protein [1] which provides mechanical strength to skin, tendon, bone, dentine, cornea and sclera [2]. It is about 25 percent of the total amount of proteins in the human body [3]. Collagen molecules consist of three-polyproline II-like left-handed helices wind around one another to form a right-handed super helical structure (triple helix) [4-8]. The three dimensional structure of collagen was first proposed based on studies using fiber diffraction and electron microscope in the 1960s [5, 9]. The first X-ray structure of the triple helix was reported using short synthetic peptides [7]. Each polypeptide chain of collagen is composed of approximately 300 repeats of the amino acid sequence and it requires a glycine(Gly) amino acid at every third residue to form a unique $(X\text{-}Y\text{-}Gly)_N$ repeating sequence pattern, where X and Y can be any amino acids. Gly, the smallest amino acid, presents at every third position of the triple helix and essential for the stability of the triple helix as it involves in strong hydrogen bonding interaction with the carbonyl oxygen (C=O) of the X position residue on the adjacent strand. It is the only amino acid packed into the

restricted space in the center of triple helix. The side chains of other residues on the X and the Y positions are largely exposed to solvent [10]. There are several amino acids frequently occurring in the collagen, such as Ala, Lys, Arg, Leu, Val, Ser, Pro and Thr [11]. Particularly, proline(Pro) is usually found in the position X and 4(R)-hydroxyproline (Hyp = O) is normally located in the position Y, respectively.

Hyp also appears to play an important role in the thermal stability of collagen triple helix [12-14]. The stability arises by a network of bridging water molecules of hydrogen bonds involving hydroxyl group of Hyp at the Y position and backbone carbonyl groups. Water bridges have been detected by high-resolution X-ray diffraction analysis of crystalline collagen [15]. These studies propose that Hyp mediated collagen stability can occur due to inter-strand hydrogen bonds with other carbonyl groups in the helix and properly oriented water molecules [13, 16-19]. In addition, the electronegative oxygen in a hydroxyl group is effective at withdrawing electron density by through-bond and through-space interactions [20]. The electron withdrawing effect of the hydroxyl group could stabilize the trans configuration of the proline amide bond. Raines distinguished the contributions between hydrogen bonding and inductive effects by replacing the hydroxyl groups with fluorine atoms in the Hyp[21]. He thought fluorine could elicit a large inductive effect since fluorine is the most electronegative atom. Recent studies also show that 4(S)-hydroxyproline (hyp) and 4(S)-fluoroproline (flp) decrease the preference for the S-trans configuration of the amide bond [22].

There are several articles in the literature dealing with the stability of collagen [3, 17, 21, 23-27]. However, the conformational preference of proline, hydroxyproline and flouroproline peptides was compared with each other in these studies. They were not

compared directly the stability of secondary structures of these amino acids or solvation effect.

In this thesis, I am focusing on the collagen stability by using computational methods. In the past half century, novel bioinformatic procedures and computational methods have been used to analyze, characterize and provide more detailed descriptions of proteins. In this regard, molecular orbital (MO) calculations have become an essential tool for structure-property investigations of an array of systems of simple reactions to larger processes of biological significance. The purpose of this thesis is to study the stability of the triple helix (collagen) like peptide structure using molecular orbital (MO) methods. It will also help us to understand the important structural aspects of peptide structure and functions.

In this thesis, molecular orbital (MO) calculations were performed using a hybrid density functional theory (DFT) at the B3LYP/D95(d,p) level using GAUSSIAN 03and GAUSSIAN 09 [28] suite of computer programs. Our lab has used this combination in several previous studies involving H-bonds in peptides and similar systems [29-35]. This combination performs well for the water dimer [36], while several other combinations of functionals and medium-sized basis sets have significant problems. This method combines Becke's three-parameter functional [37], with the nonlocal correlation provided by the correlation functional of Lee, Yang, and Parr [38].

In Chapter 2 presents the energies level of mutation Gly for collagen to DAla, LAla and DSer. The reason is that the most common genetic disorder mutation in collagen is a single base substitution that leads to the replacement of an obligate Gly by another amino acid [39, 40]. This Gly replacement mutations in collagen will destabilize the triple helix

structure and cause a series of genetic disorder, including Osteogenesis Imperfecta (OI). Osteogenesis Imperfecta (OI) is a genetic disorder characterized by unusually fragile bones that break easily, often under loads that normal bones bear daily. This inherent weakness of the bones is due to a malfunction in the patient's production of the collagen. I am using DFT and ONIOM [41, 42] calculations using B3LYP/D95(d,p) and AM1 method. ONIOM calculations can define two or three layers within the structure that are treated at different levels of accuracy. The backbone of the system is calculated at the DFT B3LYP/D95** which is high level. While the rest of the system is calculated using the AM1 semiempirical molecular orbital method which is calculated for proline, alanine side chains.

4(R)-hydroxyproline (Hyp = O) also appears to play an important role in the thermal stability of collagen triple helix [12-14]. To understand the stability of the triple helix on the X and Y positions, density functional theory (DFT) and ONIOM [41, 42] calculations are using at B3LYP/D95(d,p) and AM1 method to calculate the internal hydrogen bonding at triple helix based on mutations at the X or the Y positions. The details will be described on Chapter 2.

In chapter 3, the detailed analysis of energies were calculated for single proline (Pro), hydroxyproline (Hyp), fluoroproline (Flp) and methoxyproline (Mop) isomers (R or S structures) in the gas phase as well as in solvated state by using density functional theory and the solvation model SM5.2 of Cramer/Truhlar [43] to compare with previous study. Raines and co-workers have attributed the stability of collagen triple helices containing 4(R)-substituents in the Y position solely to stereoelectronic effects. My results show the stabilization is mainly due to differential aqueous solvation effect of the modified

prolines. The results will be discussed in Chapter 3.

Polyglycine crystalizes in three forms, polyglycine I (actually a dimorph) and II. Polyglycine I includes two different crystal structures each containing a different type of β-sheet [44-48]. The structure of polyglycine II (PII) consists of parallel β-sheets oriented 120 degrees to each other [49, 50] (Figure 1.1). Three of these sheets intersect at each polyglycine strand. Each triad of nearest neighbor strands bears a striking resemblance to collagen. The principal differences between collagen-like structures and PII lie in the inability of collagen to form the infinite pattern of H-bonds normal to the peptide backbone (found in PII) beyond the three strands of the collagenic triple helix, due to a) the lack of N-H donors on the proline and hydroxyproline residues and b) the steric impediments caused by the side chains of all the amino acids except for glycine. In Chapter 4, I will compare the same size of PII and collagen structures.

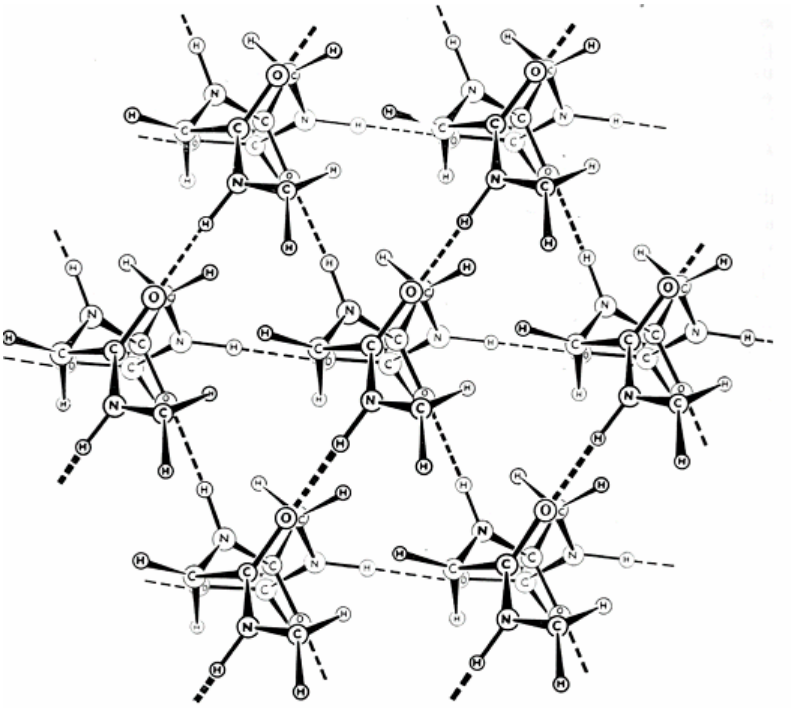


**Figure1.1**: Polyglycine II structure contains three parallel-sheets. Hydrogen bonds drawn as dashed lines.

In Chapter 4, I kept each strand of formyl-$Gly_6NH_2$ equivalent. The β-sheets were also optimized with the same constraints for comparison. The single-point a postieri counterpoise corrections (CP) were calculated using the procedure incorporated in GAUSSIAN 09. Optimization on the CP-corrected potential energy surfaces (CP-OPT) [51] were not completed due to the excessive CPU time required.

While there have been several reports of problems with the B3LYP functional in other applications [52], this functional has been tested with the D95(d,p) and D95++(d,p) basis sets and shown to give reasonable results for this thesis [53] dihedrals and for H-bonding interactions [29, 36].

In the 1960's, Ramachandran published his classic studies of the energy landscapes

of peptides as a function of the two generally mobile dihedral angles in the peptide backbone [6, 9]. Since no useful theoretical methods were available at the time of his study, he used molecular models to define those dihedral angles for which unfavorable steric interactions make the conformations unstable. He was not able to differentiate between those conformations that had no steric impediments, nor did he take the effects of weak ($C_5$) or other stronger *intra*molecular hydrogen bonds into account. The last chapter of this thesis deals with energy surfaces of diastereomeric trialanine peptides in gas phase and aqueous solution. The density functional theory has been show the energy landscapes of two diasteriomeric capped trialanine peptides in the gas phase and aqueous solution.

In Chapter 5, I will present energy surfaces of Ramachandran polt by using DFT calculations base on the two dihedral angles, $\phi$ and $\varphi$, of the acetyl-$^{L}$Ala-$^{L}$Ala-$^{L}$Ala-$NH_2$ (3AL) and its diastereomer, acetyl-$^{L}$Ala-$^{D}$Ala-$^{L}$Ala-$NH_2$ (3AD) as defined in figure **1.2.** Those two dihedral angles were fixed for five degree intervals of each, while all the other internal coordinates were relaxed. In a preliminary study, I found that using 15 degree intervals produced too course and energy surface to be useful. Thus, 5184 ($72^2$) individual geometrical optimizations (with two fixed dihedral angles) were performed on each of the two diasteromers (10368 optimizations, in all). The aqueous solvation free energies for each of the 5184 geometries per diasteromer were calculated using Cramer/Truhlar solvation scheme [43] using the AMPAC 8.15 program [54], which contains the SM 5.2 version of the AMSOL [43] solvent model. All AMPAC calculations used the AM1 [55] semiempirical Hamiltonian together with the SM 5.2 solvation method at each of the fixed 5184 molecular geometries previously determined using DFT.

In this manner, we obtained separate energy landscapes for the total DFT *energy* (gas phase) calculation and the *free energy* of solvation as a function of the two dihedral angles.

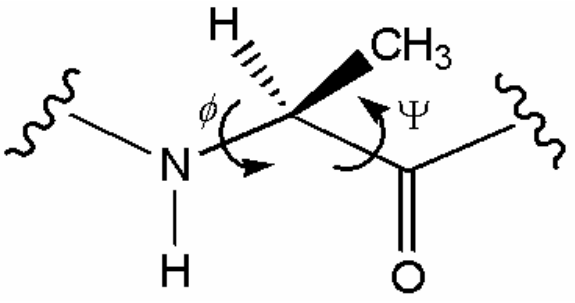



**Figure 1.2:** Definitions of the dihedral angles considered in this study.

I obtained the energy landscapes for the solvated structures by adding the corresponding points on each of these two surfaces. Thus, the solvated landscapes represent a combination of the gas phase potential energies and the free energies of solution.

I completely optimized all the local minima on the gas phase surface including all the internally H-bonded structures. I also calculated the vibrational frequencies at these minima to obtain the vibrational corrections to the enthalpies and free energies. The vibrational calculation (when performed) used the standard harmonic approximations used in the Gaussian 03 program.

Counterpoise corrections [56, 57] (CP) to the basis set superposition error (BSSE) were approximated by from formamide dimers taken from the geometries of the cyclic H-bonding systems. The results will be shown in Chapter 5.

**Referances:**


1. Nimni, M.E. and Editor, *Collagen: Biochemistry, Biomechanics, Biotechnology. A 3-Volume Set*. 1988. 1088 pp.
2. Kielty, C.M., I. Hopkinson, and M.E. Grant, *Collagen: the collagen family: structure, assembly, and organization in the extracellular matrix.* 1993: p. 103-47.
3. Bhattacharjee, A. and M. Bansal, *Collagen structure: The Madras triple helix and the current scenario.* IUBMB Life, 2005. **57**(3): p. 161-172.
4. Cowan, P.M. and S. McGavin, *The structure of poly-L-proline.* Nature (London, U. K.), 1955. **176**: p. 501-3.
5. Rich, A. and F.H. Crick, *The molecular structure of collagen.* Journal of molecular biology, 1961. **3**: p. 483-506.
6. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* Journal of Molecular Biology, 1963. **7**(1): p. 95-9.
7. Bella, J., et al., *Crystal and molecular structure of a collagen-like peptide at 1.9 .ANG. resolution.* Science (Washington, DC, United States), 1994. **266**(5182): p. 75-81.
8. Sasisekharan, V., *Structure of poly-L-proline II.* Acta Crystallographica, 1959. **12**: p. 897-903.
9. Ramachandran, G.N. and V. Sasisekharan, *Conformation of polypeptides and proteins.* Advances in Protein Chemistry, 1968. **23**: p. 283-438.
10. Jones, E.Y. and A. Miller, *Analysis of structural design features in collagen.* Journal of Molecular Biology, 1991. **218**(1): p. 209-19.
11. Woodhead-Galloway, J., *Collagen: the anatomy of a protein.* Institute of Biology's Studies in Biology, 1980. **117**: p. 64 pp.
12. Chopra, R.K. and V.S. Ananthanarayanan, *Conformational implications of enzymic proline hydroxylation in collagen.* Proceedings of the National Academy of Sciences of the United States of America, 1982. **79**(23): p. 7180-4.
13. Privalov, P.L., *Stability of proteins. Proteins which do not present a single cooperative system.* Advances in Protein Chemistry, 1982. **35**: p. 1-104.
14. Berg, R.A. and D.J. Prockop, *Thermal transition of a nonhydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple helix of collagen.* Biochemical and Biophysical Research Communications, 1973. **52**(1): p. 115-20.
15. Bella, J., et al., *Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution.* Science (New York, N.Y.), 1994. **266**(5182): p. 75-81.
16. Bella, J., B. Brodsky, and H.M. Berman, *Hydration structure of a collagen peptide.* Structure (London), 1995. **3**(9): p. 893-906.
17. Vitagliano, L., et al., *Preferred proline puckerings in cis and trans peptide groups: implications for collagen stability.* Protein Science, 2001. **10**(12): p. 2627-2632.
18. Vitagliano, L., et al., *Structural bases of collagen stabilization induced by proline hydroxylation. [Erratum to document cited in CA135:15652].* Biopolymers, 2001. **59**(6): p. 465.
19. Berisio, R., et al., *Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)$_{10}$]$_3$.* Protein Science, 2002. **11**(2): p. 262-270.
20. Stock, L.M., *Origin of the inductive effect.* Journal of Chemical Education, 1972.

**49**(6): p. 400-4.
21. Raines, R.T., *2005 Emil Thomas Kaiser Award.* Protein Science, 2006. **15**(5): p. 1219-1225.
22. Bretscher, L.E., et al., *Conformational Stability of Collagen Relies on a Stereoelectronic Effect.* Journal of the American Chemical Society, 2001. **123**(4): p. 777-778.
23. Schumacher, M., K. Mizuno, and H.P. Baechinger, *The Crystal Structure of the Collagen-like Polypeptide (Glycyl-4(R)-hydroxyprolyl-4(R)-hydroxyprolyl)$_9$ at 1.55 A Resolution Shows Up-puckering of the Proline Ring in the Xaa Position.* Journal of Biological Chemistry, 2005. **280**(21): p. 20397-20403.
24. Doi, M., et al., *Collagen-like triple helix formation of synthetic (Pro-Pro-Gly)$_{10}$ analogs: (4(S)-hydroxyprolyl-4(R)-hydroxyprolyl-Gly)$_{10}$, (4(R)-hydroxyprolyl-4(R)-hydroxyprolyl-Gly)$_{10}$ and (4(S)-fluoroprolyl-4(R)-fluoroprolyl-Gly)$_{10}$.* Journal of Peptide Science, 2005. **11**(10): p. 609-616.
25. Persikov, A.V., et al., *Amino acid propensities for the collagen triple-helix.* Biochemistry, 2000. **39**(48): p. 14960-14967.
26. Improta, R., C. Benzi, and V. Barone, *Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution.* Journal of the American Chemical Society, 2001. **123**(50): p. 12568-12577.
27. Improta, R., et al., *Understanding the role of stereoelectronic effects in determining collagen stability. 2. A quantum mechanical/molecular mechanical study of (proline-proline-glycine)n polypeptides.* Journal of the American Chemical Society, 2002. **124**(26): p. 7857-7865.
28. Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, 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, K.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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; 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 03*, 2003.
29. Plumley, J.A. and J.J. Dannenberg, *The Importance of Hydrogen Bonding between the Glutamine Side Chains to the Formation of Amyloid VQIVYK Parallel β-Sheets: An ONIOM DFT/AM1 Study.* Journal of the American Chemical Society. **132**(6): p. 1758-1759.
30. Viswanathan, R. and J.J. Dannenberg, *A density functional theory study of*

*vibrational coupling in the amide I band of β-sheet models.* J. Phys. Chem. B, 2008. **112**(16): p. 5199-5208.
31. Chen, Y.-F. and J.J. Dannenberg, *Cooperative 4-pyridone H-bonds with extraordinary stability. A DFT molecular orbital study.* Journal of the American Chemical Society, 2006. **128**(25): p. 8100-8101.
32. Wieczorek, R. and J.J. Dannenberg, *Comparison of Fully Optimized alpha - and $3_{10}$-Helices with Extended beta -Strands. An ONIOM Density Functional Theory Study.* Journal of the American Chemical Society, 2004. **126**(43): p. 14198-14205.
33. Tsai, M., Y. Xu, and J.J. Dannenberg, *Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProProDAla, and ProProDSer Triads.* Journal of the American Chemical Society, 2005. **127**(41): p. 14130-14131.
34. Kobko, N. and J.J. Dannenberg, *Cooperativity in Amide Hydrogen Bonding Chains. Relation between Energy, Position, and H-Bond Chain Length in Peptide and Protein Folding Models.* Journal of Physical Chemistry A, 2003. **107**(48): p. 10389-10395.
35. Kobko, N., et al., *Cooperativity in Amide Hydrogen Bonding Chains: Implications for Protein-Folding Models.* Journal of the American Chemical Society, 2001. **123**(18): p. 4348-4349.
36. Simon, S., M. Duran, and J.J. Dannenberg, *Effect of Basis Set Superposition Error on the Water Dimer Surface Calculated at Hartree-Fock, Moller-Plesset, and Density Functional Theory Levels.* Journal of Physical Chemistry A, 1999. **103**(11): p. 1640-1643.
37. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange.* Journal of Chemical Physics, 1993. **98**(7): p. 5648-52.
38. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B Condensed Matter and Materials Physics, 1988. **37**(2): p. 785-9.
39. Byers, P.H., *Osteogenesis imperfecta.* 1993: p. 317-50.
40. Marini, J.C., et al., *Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans.* Human Mutation, 2007. **28**(3): p. 209-221.
41. Morokuma, K., *New challenges in quantum chemistry: quests for accurate calculations for large molecular systems.* Philosophical Transactions of the Royal Society of London, Series A Mathematical, Physical and Engineering Sciences, 2002. **360**(1795): p. 1149-1164.
42. Dapprich, S., et al., *A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives.* Theochem, 1999. **461-462**: p. 1-21.
43. Cramer, C.J. and D.G. Truhlar, *Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics.* Chemical Reviews (Washington, D. C.), 1999. **99**(8): p. 2161-2200.
44. Lotz, B., *Crystal structure of polyglycine I.* Comptes Rendus des Seances de l'Academie des Sciences, Serie C Sciences Chimiques, 1972. **274**(23): p. 1907-10.
45. Colonna-Cesari, F., S. Premilat, and B. Lotz, *Structure of polyglycine I: a*

*comparison of the antiparallel pleated and antiparallel rippled sheets.* Journal of Molecular Biology, 1974. **87**(2): p. 181-91.
46. Lotz, B., *Crystal structure of polyglycine I.* Journal of Molecular Biology, 1974. **87**(2): p. 169-80.
47. Munoz-Guerra, S., et al., *Crystals of polyglycine in the beta form.* Journal of Molecular Biology, 1983. **167**(1): p. 223-5.
48. Kajava, A.V., *Dimorphism of polyglycine I: structural models for crystal modifications.* Acta crystallographica. Section D, Biological crystallography, 1999. **55**(Pt 2): p. 436-42.
49. Crick, F.H.C. and A. Rich, *Structure of polyglycine II.* Nature (London, U. K.), 1955. **176**: p. 780-1.
50. Ramachandran, G.N., C. Ramakrishnan, and C.M. Venkatachalam, *Structure of polyglycine II with direct and inverted chains.* Conformation of Biopolymers, Papers read at an International Symposium, 1967. **2**: p. 429-38.
51. Simon, S., M. Duran, and J.J. Dannenberg, *How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers?* Journal of Chemical Physics, 1996. **105**(24): p. 11024-11031.
52. Wodrich, M.D., et al., *How accurate are DFT treatments of organic energies?* Organic Letters, 2007. **9**(10): p. 1851-1854.
53. Improta, R., et al., *The conformational behavior of polyglycine as predicted by a density functional model with periodic boundary conditions.* Journal of Chemical Physics, 2001. **114**(6): p. 2541-2549.
54. Dewar, M.J., MS. Mod. Technol. Comput. Chem. Motecc, 1991.
55. Dewar, M.J.S., et al., *Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model.* Journal of the American Chemical Society, 1985. **107**(13): p. 3902-9.
56. Jansen, H.B. and P. Ros, *Nonempirical molecular orbital calculations on the protonation of carbon monoxide.* Chemical Physics Letters, 1969. **3**(3): p. 140-3.
57. Boys, S.F. and F. Bernardi, *The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors.* Molecular Physics, 1970. **19**(4): p. 553-566.

# Chapter 2

# Triple-Helical Collagen-like Structures

## 2.1 Containing ProProGly, ProProAla, ProPro$^{D}$Ala, and ProPro$^{D}$Ser Structures

### 2.1.1 Introduction

The structural protein, collagen, recently reviewed [1], consists of triple helical peptides whose structure was proposed by Rich and Crick [2], in which each of three peptide strands contains the repeating amino acid triad X-Y-Gly, where X and Y can be any amino acid residues. In this structure, each peptide strand H-bonds to one of the other two using its Gly N-H's as donors, and to the third, using the C=O's on other amino acid residues as acceptors. These triple helices are rich in proline and 4-(*R*)-hydroxyproline (Hyp), which constitute about 20% of the amino acid content, while the most common repeating unit is Pro-Hyp-Gly [3]. Various diseases, such as osteogenesis imperfecta, Ehlers-Danlos syndrome, Alport syndrome, Schmid

metaphyseal chondrodysplasia, and dystrophic epidermolysis bullosa, result from amino acid residue mutations in the collagens, mostly mutations of the Gly [4, 5]. These lead to misfolding of the collagen, which can be modeled with short synthetic peptide chains [6].

### 2.1.2 Method

I construct the smallest structure of the triple helix repeating $(Pro\text{-}Pro\text{-}Gly)_2$ in each strand and that are capped with acetyl for N-termanial and dimethyl amide for C-terminal to prevent the formation of H-bonds involving the COOH's and $NH_2$'s of the peptide strands. There are three strands in the helix; the total is 18 amino acids in our structure. The triple helical structure contains two repeats of the ProProGly triad in each strand. One strand had its sequence dislocated by one amino acid from the other two so that six H-bonds could form (ProGlyPro instead of ProProGly).

To probe the effect of Gly mutations, I present DFT and ONIOM [7, 8] calculations using B3LYP/D95(d,p) and AM1 method. That ONIOM can define two or three layers within the structure that are treated at different levels of accuracy. I have treated the most important part of the system, the peptide backbone, at the high level (B3LYP/D95**), while the rest of the system (side chains) is calculated using the AM1 semiempirical molecular orbital method.

All calculations used the GAUSSIAN 03 suite of programs [9] using a procedure fully described elsewhere [10]. In the beginning, I took the protein data bank structures 1A3J as the starting point. The triple helical structures were corrected for basis set superposition error using the posteri counterpoise correction (CP).

One can use several approaches towards evaluating stabilities of a triple helical structure containing PPG repeats. For example, the stability of a triple helix can be computed as a difference of energies from the constituting strands (equation 1 and 2). On the other hand one can use the difference of polymerization energy between three $(PPG)_2$ strands and that of triple helix (equation 3).

3 x Separated optimized individual peptide strands → Triple Helical (eq 1)

3 x $(PPG)_2$ strands (frozen in their triple helical geometries) → Triple Helix (eq 2)

3 x [4 Proline + 2 Glycine + $(CH_3)_2NCOCH_3$ ]→ 3 [$(CH_3)_2$N-PPGPPG-$COCH_3$ ] + 6 Water (eq3)

### 2.1.3 Results and Discussions

The structure of the extended single strand of ProProGly is distorted from a polyglycine structure as evident from Figure 2.1.1. Consequently, the formation of the three ProProGly strands from the component amino acids is calculated to be

energetically unfavorable by 53.0 kcal/mol. The value is consistent with other calculations from our laboratory [11], which indicate that the substitution of the Pro for an Ala in $Ac(Ala)_{17}NH_2$ destabilizes the peptide by about 7.5 kcal/mol (there are 12 P's in the combined three strands).

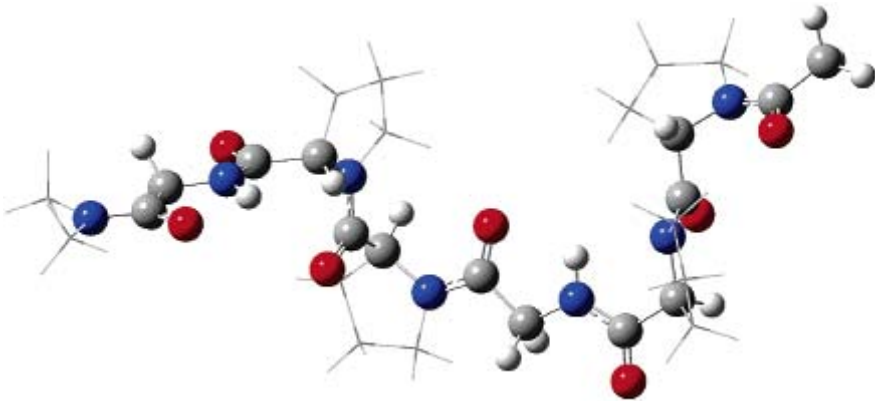


**Figure 2.1.1:** Optimized ProProGly single strand

The H-bonding interaction energy in the triple helix(PPG) is -20.8 kcal/mol. This relative energy (stability) of a triple helix is calculated with respect to complete single strands. After counterpoise (CP) but no vibrational correction(Figure 2.1.2), while the average O···N distance across the six H-bonds is 2.961 Å compared to the protein data bank structures 3.012 (1A3I) , 2.960(1A3J) , 2.971(1G9W) and 2.968(1ITT) Å respectively. Thus the optimized calculations before CP correction can provide a lower limit for both the H-bonding stabilizations and the H-bonding distances for this

level of calculation. On the other hand, the experimental H-bond distances come from much longer chains, so they are less influenced by end effects (where the H-bonds are longer).

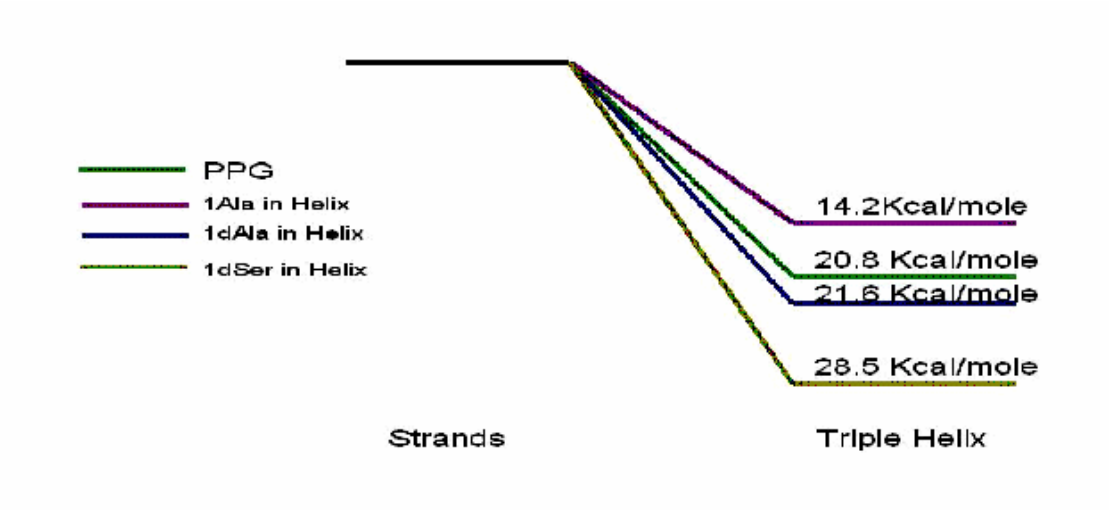



**Figure 2.1.2:** The H-bonding interaction in the triple helix after counterpoise (CP) but no vibrational correction

From inspection of Figures 2.1.1 and 2.1.3, one can easily see that the requirement for the Gly as every third amino acid in the collagen and the collage-like triple helices derives from its enantiomorphic property, unique among naturally occurring amino acids. Had an $^{L}$amino acid been substituted for the Gly, the side chain would sit in the center of the triple helix, causing considerable strain and weakening the structure. However, if a $^{D}$amino acid were substituted for the Gly, its side chain would extend peripherally from the helix, seemingly avoiding this steric problem. To test this hypothesis, I considered ProProGly triple helices, where a single Gly near the

center of the structure was mutated either for $^{L}$Ala or $^{D}$Ala. The mutated triple helices had H-bonding interactions of -14.2 and -21.6 kcal/mol for $^{L}$ and $^{D}$Ala, respectively. Thus, $^{L}$Ala *reduces* the H-bonding stabilization by 6.6 kcal/mol while increasing the average N···O distance to 3.039 Å, and $^{D}$Ala *increases* the stabilization by 0.8 kcal/mol but slightly increases the average N···O distance (to 2.999 Å).

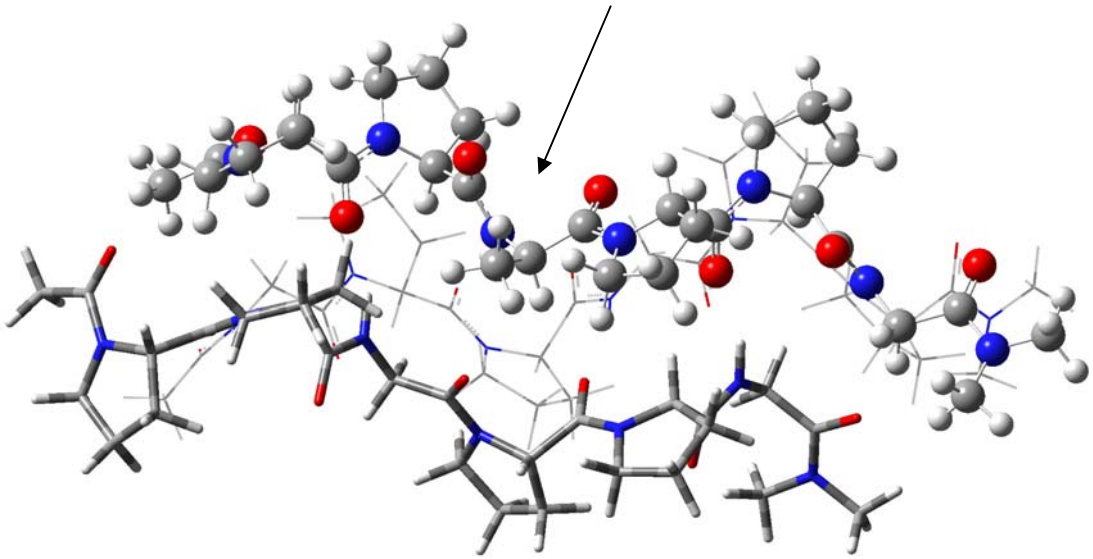


**Figure 2.1.3:** The ProProGly triple helix with one $^{D}$Ala (indicated by arrow). Each of the three strands is rendered differently (ball and stick, tube, and wire frame).

The increased H-bonding in the structure with the Gly → $^{D}$Ala mutation encouraged us to evaluate the effect of a Gly → $^{D}$Ser mutation. Ser is simply Ala with one methyl H transformed to an OH, which would be near enough to the C=O of a neighboring strand to form an additional H-bond, thus increasing the triple helical stability. In fact, this triple helix had the most stabilizing H-bonding energy of all: -28.5 kcal/mol (-7.7 more than Pro-Pro-Gly) with respect to the optimized single

peptide strands with an average N···O of 3.029 Å (see Figure 2.1.4).

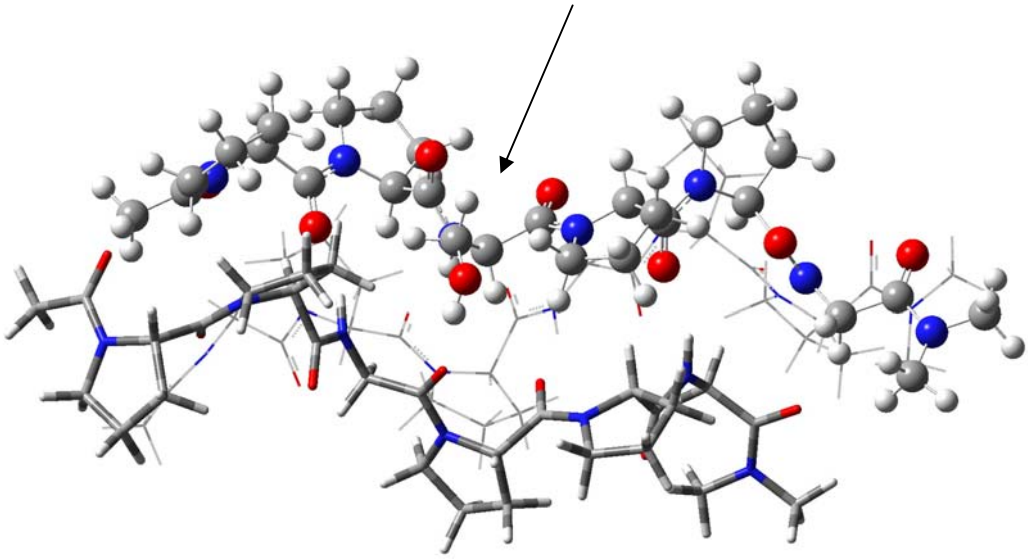


**Figure 2.1.4.** Structure of the triple helix containing one $^{D}$Ser in place of the Gly. The additional H-bond is indicated.

When considering the interaction energies upon going from the single strands to the triple helix, one must realize that there is a bias inherent in the choice of the optimized single strands as a reference point, and that the interaction energy calculated in this manner can be conceived of as a combination of both a distortion energy from the optimized strand to the conformation it assumes in the triple helix and the interactions of the three distorted strands into the helix. When approached this way, it becomes immediately clear that the closer the geometry of an optimized strand is to that of its distorted geometry in the triple helix, the greater the stabilization upon

forming the triple helix from those individual strands. Further the inherent stabilizations of the strands will vary depending upon the component amino acids. The capped (single) strand ProPro$^{L}$AlaProProGly is 4.1 kcal/mol more stable than ProPro$^{D}$AlaProProGly, its diastereomer, as is reflected in the total energies of the three optimized strands that differ by the configuration of the Ala. However, when comparing nonisomeric peptides, I can use the polycondensation reaction to evaluate their energies relative to the component amino acids. As seen from Figure 2.1.5, the energies of the three states considered (optimized and distorted strands and triple helices) assume a different order. This provides an extraordinary example of the importance of the choices of reference state. Thus, the ProProGly triple helix is 4.6, 6.8 and 10.0 kcal/mole most stable than the triple helix ProPro$^{D}$SerProProGly, ProPro$^{D}$AlaProProGly and ProPro$^{L}$AlaProProGly respectively with reference to the amino acids, but the triple helix with the $^{D}$Ser is the most stable (28.5 kcal/mole) with reference to the optimized strands.

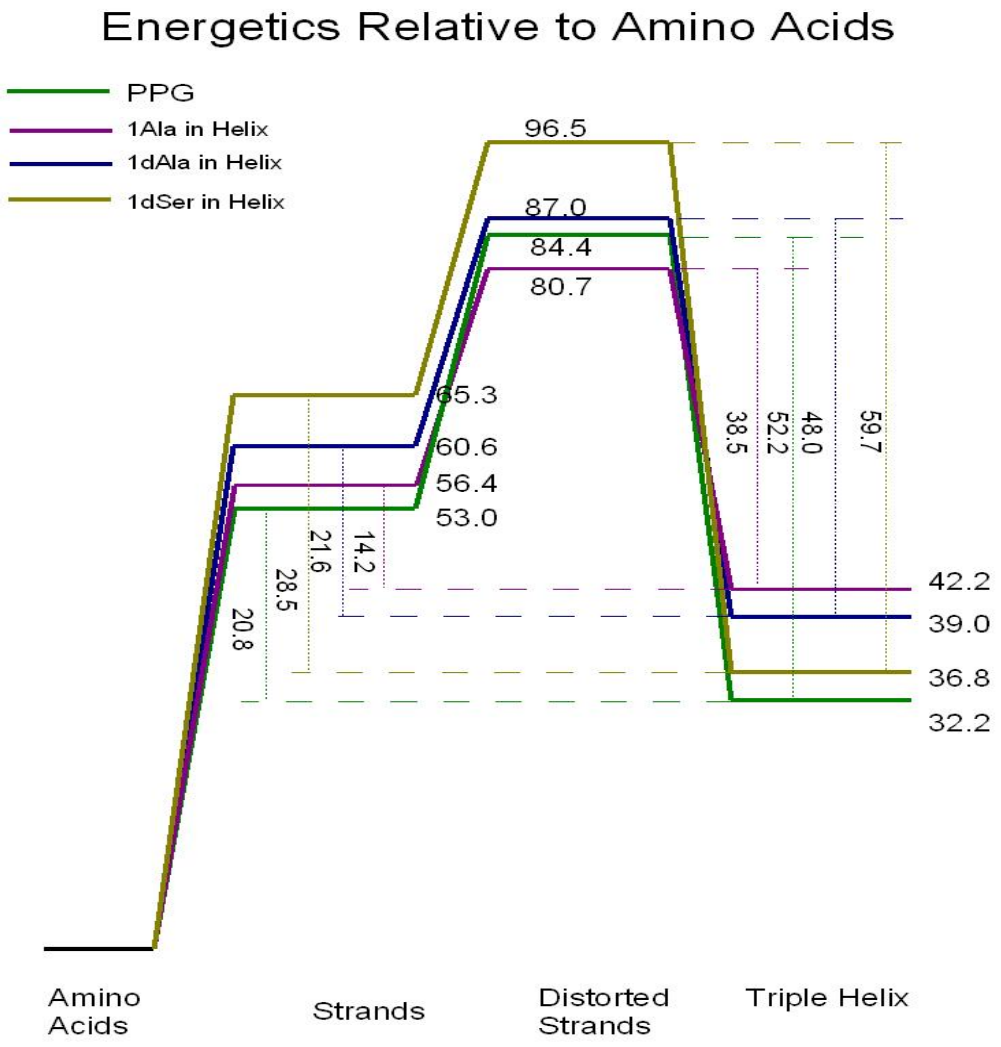



**Figure 2.1.5.** Energies of the (combined) three optimized strands, strands distorted to their triple helical geometries, and optimized triple helices (including CP correction). PPG represent ProProGly.

Notice that using the DFT calculation for the energy of triple helix containing $^{L}$Ala has a significantly larger (around 4.7 kcal/mol) counterpoise correction than that containing $^{D}$Ala (due to the greater effect of the orbitals of the methyl buried in the middle of the triple helix containing an $^{L}$Ala strand compared to those of the exterior

methyl in the $^{D}$Ala strand), and that this contributes significantly to the difference in interaction energies.(Figure 2.1.6)

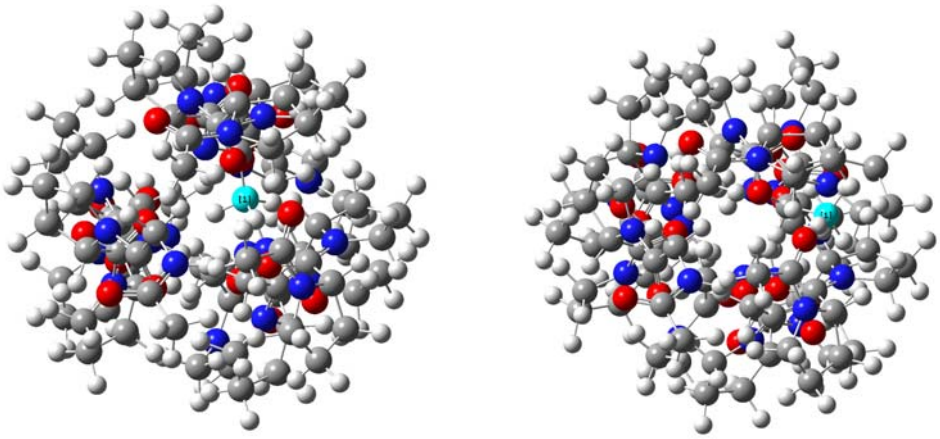


**Figure 2.1.6.** Triple helix containing $^{D}$Ala(left) and $^{L}$Ala(right)

In summary, the enantiomorphic Gly plays the role of a $^{D}$amino acid in the repeating triad XYG. Thus, $^{D}$amino acids can potentially replace it to energetic advantage.

The results indicate distinguishing the energetic of a variety of triple helix collagen models of six amino acids in length at the DFT level of theory for the backbone, and at the AM1 level of theory for the side-chain atoms. An important point that emerges from my comprehensive study is that the relative energy (and hence stability) of a triple helix can be calculated with respect to more than one reference state, using either the condensation energy referenced to the gas phase amino acids, or the single strands that form the triple helix. Comparing these relative energies

(stabilities) for two or more triple helices with different amino acid composition the relative stability of two triple helices can differ with the choice of the reference state. Another fascinating suggestion of our work is to use a D-amino acid instead of glycine. I also found that $^{D}$Ala or $^{D}$Ser at the right sequential position does stabilize a collagen triple helix structure.

## 2.2 Mutation of X and Y positions of triple-helical collagen-like structures

The previous section 2.1 shows that the mutation of Gly (→) to $^{D}$Ala or $^{D}$Ser increases the overall stability of the triple helix by H-bonding whereas the stability decreases through the incorporation of $^{L}$Ala. In this section, X and Y positions have been mutated, one position at a time, using derivatives of proline (*viz*. Fluoroproline, Hydrodyproline, and Methoxyproline) to compare and contrast the resultant structures (e.g. H-bonding distance and stability).

### 2.2.1 Introduction

The 4R-hydroxyproline(Hyp; O) amino acid residue is usually observed in the Y position and it can stabilize the triple helix while present in this position only [12]. It also appears to play an important role in the thermal stability of collagen triple helix [13-15]. Several studies have showed that Hyp (mediates) increases collagen stability

by orienting water molecules to form interstrand hydrogen bonding [16, 17]. My objective is to understand the reason behind enhanced stability of the collagen's triple helix formed from repeating POG triad. A second objective is to compare and contrast the preference of Proline and three of its derivatives mentioned before for X and Y-positions in the triple helices. In this way, the stability of this different proline-derivative-rich collagen triple helix can be compared with that of wild type.

There are several articles present in literature dealing with the stability of collagen helices comprising of proline derivatives [1, 18-24]. Though the individual conformational preferences of proline, hydroxyproline, and fluoroproline residues were compared with each other in these studies, a direct comparison of the stability of the corresponding secondary structures and the role of solvation were lacking.

Raines *et al.* [25, 26], have suggested that the dominant effect might be stereoelectronic rather than H-bonding, since peptides in which 4-fluoroproline (Flp) replaces Hyp form more stable triple helices of the unfolded peptides in aqueous solution. It is interesting to note that 4-methoxyproline (Mop), which can participate in H-bonding as an acceptor, has a stereoelectronic effect similar to Hyp [27].

In this section, my focus is on the collagen stability when proline, in either the X or the Y positions, is mutated to Flp, Hyp, or Mop. Barone *et al*. [21] have reported that electronegative substituents in the 4R position of proline stabilize the up

pucker of the pyrrolidine ring (i.e. the exo conformations of the proline ring), while 4S substitutions favor down pucker of the pyrrolidine ring (endo conformations of the proline ring)(Figure 2.2.1). Recent studies have also shown that 4(S)-hydroxyproline (Hyp) and 4(S)-fluoroproline (Flp) decrease the preference for the trans-configuration of the amide bond [28]. My goal will be to understand the effect of amino acid mutations in the stability of the various conformers of fluorine derivatives in triple helix, with special emphasis in the aqueous solvation effect.

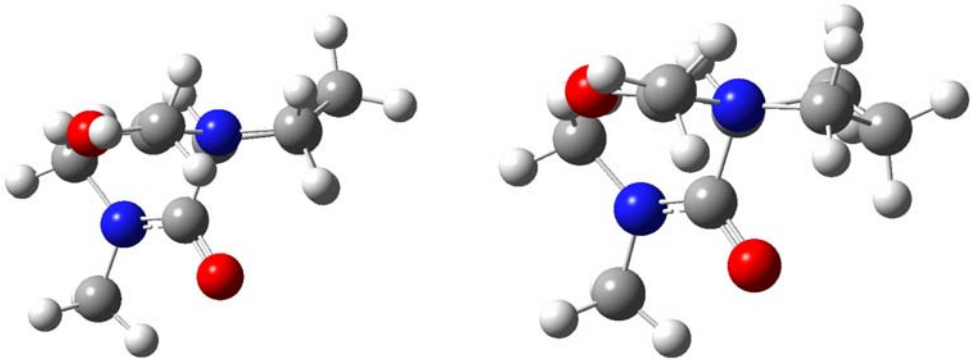


**Figure 2.2.1**. Exo(Left) and Endo(right) conformations of the proline ring

### 2.2.2 Method

I have carried out DFT calculations at the B3LYP/D95(d,p) level with the GAUSSIAN 03[29] and GAUSSIAN 09 [30] suite of computer programs [31]. This method combines Becke's 3-parameter functional [32] with the non-local correlation provided by the correlation functional of Lee, Yang and Parr [33]. I treated the backbone of the middle six amino acids of the system and the mutated amino acids in

the higher B3LYP/D95** level. The rest of the system (N and C-terminals) is calculated using the AM1 semiempirical molecular orbital method. All the mutated structures of triple helix on the gas phase have been completely optimized. Due to prohibitively large CPU time, I have not calculated the vibrational frequencies of these optimized structures. The aqueous solvation free energies for each of the optimized geometries were calculated using Cramer and Truhlar solvation scheme[34] within the AMPAC 8.16 program [35] which contains the SM 5.2 version of the AMSOL solvent model. All AMPAC calculations used the AM1[36] semiempirical Hamiltonian together with the SM 5.2 solvation method at the fixed optimized triple helix geometries previously optimized using DFT. The relative solvated energies represent a combination of the gas phase potential energies and the free energies of solution.

**2.3 Results**

Table 2.3.1 summarizes the results for mutations in the Y position, whereas mutations in the X position are shown in Table 2.3.2. I will first compare the hydrogen bonding energies for the fluorine derivatives in triple helix conformations. One can observe that all the exo-structures have stronger hydrogen bonding energies than the endo ones, and that the 4R conformations for Y positions are more stable than the 4S

ones. It is worth mentioning that the 4S-Hyp triple helix conformations have two isomers for both X and Y positions which differ by one H-bonding interaction (see Figure 2.3.2). Only the lowest energy of all the proline derivatives conformation are discussed here.

It is important to note that my findings are in good agreement with previously reported preferred conformations of proline derivatives in the context of collagen triple helices [21]. For example, introduction of electronegative substituent in the 4R and 4S positions of pyrollidine ring in the Y position, stabilize the up-puckering (exo), or the down-puckering (endo) of the pyrrolidine ring, respectively.

## Table 2.3.1 : Mutation of Y position

| Mutation of Y position (kcal/mol) | HB energies (CP) | $\Delta E_3$ from optimized strand | Helix Solvation Energies | fix geom. single strands solvation free Energies | Opt. single strands solvation free Energies | Relative Solvated energies ($\Delta E_3$ in gas phase + free energy of solvation) | Melting temperature (°C) |
|---|---|---|---|---|---|---|---|
| Proline(endo) | -48.20 | -27.54 | -127.40 | -178.37 | -165.06 | 1.68 | 41[b] |
| 4R-Flp (endo) | -48.42 | -29.45 | -126.67 | -177.55 | -164.75 | 0.19 | 91[a] |
| 4R-Hyp (endo) | -48.29 | -28.83 | -132.37 | -183.32 | -170.36 | 0.72 | 62[a], 69[b], 76[c] |
| 4R-Mop (endo) | -48.33 | -27.34 | -128.57 | -179.52 | -165.95 | 1.59 | 70[a] |
| 4S-Flp (endo) | -47.93 | -27.40 | -129.07 | -179.83 | -166.73 | 1.81 | no helix |
| 4S-Hyp (endo)* | -49.44 | -25.82 | -130.37 | -181.73 | -164.37 | -0.27 | no helix |
| 4S-Mop (endo) | -48.06 | -26.17 | -129.83 | -180.93 | -166.51 | 2.06 | - |
| Proline(exo) | -48.43 | -26.35 | -127.57 | -178.68 | -163.68 | 1.31 | 41[b] |
| 4R-Flp (exo) | -49.10 | -28.09 | -127.86 | -179.36 | -164.40 | 0.00 | 91[a] |
| 4R-Hyp (exo) | -49.13 | -28.63 | -132.77 | -184.63 | -170.84 | 0.99 | 62[a], 69[b], 76[c] |
| 4R-Mop (exo) | -49.30 | -27.56 | -128.78 | -180.11 | -167.22 | 2.43 | 70[a] |
| 4S-Flp (exo) | -48.79 | -26.41 | -127.06 | -178.11 | -162.89 | 0.97 | no helix |
| 4S-Hyp (exo) | -48.93 | -26.45 | -132.76 | -184.73 | -168.52 | 0.86 | no helix |
| 4S-Mop (exo) | -49.04 | -26.40 | -128.87 | -179.83 | -164.84 | 1.12 | - |

*One additional HB


a. Kotch, F. W.; Guzei, I. A.; Raines, R. T. *J. Am. Chem. Soc.* **2008**, *130*, 2952
b. Vitagliano, L.; Nemethy, G.; Zagari, A.; Scheraga, H. A. *Biochemistry* **1993**, *32*, 7354
c. Mizuno, K.; Hayashi, T.; Peyton, D. H.; Bachinger, H. P. *J. Biol. Chem.* **2004**, *279*, 282

## Table 2.3.2 : Mutation of X position

| Mutation of X position (kcal/mole) | HB energies (CP) | $\Delta E_2$ from optimized strand | Helix Solvation Energies | fix geom. single strands solvation free Energies | Opt. single strands solvation free Energies | Relative Solvated energies ($\Delta E_2$ in gas phase + free energy of solvation) | Melting temperature (°C) |
|---|---|---|---|---|---|---|---|
| Proline(endo) | -47.55 | -33.49 | -126.89 | -169.11 | -163.67 | 2.93 | 41[b] |
| 4R-Flp (endo) | -47.53 | -33.55 | -126.43 | -164.83 | -162.65 | 2.31 | no helix |
| 4R-Hyp (endo) | -47.93 | -33.52 | -131.97 | -170.89 | -168.86 | 3.01 | no helix |
| 4R-Mop (endo) | -47.56 | -36.34 | -128.27 | -167.02 | -164.97 | 0.00 | - |
| 4S-Flp (endo) | -49.61 | -33.43 | -127.65 | -170.77 | -168.18 | 6.74 | 33[a] |
| 4S-Hyp (endo)* | -46.07 | -34.89 | -131.50 | -170.51 | -166.56 | -0.21 | no helix |
| 4S-Mop (endo) | -48.72 | -33.13 | -127.06 | -169.86 | -166.63 | 6.08 | - |
| Proline(exo) | -48.26 | -32.09 | -126.93 | -168.14 | -164.14 | 4.75 | 41[b] |
| 4R-Flp (exo) | -48.19 | -32.77 | -127.51 | -168.41 | -164.39 | 3.74 | no helix |
| 4R-Hyp (exo) | -48.67 | -34.10 | -132.44 | -174.21 | -169.75 | 2.83 | no helix |
| 4R-Mop (exo) | -48.32 | -35.98 | -128.62 | -170.25 | -165.61 | 0.64 | - |
| 4S-Flp (exo) | -48.15 | -32.28 | -126.53 | -167.74 | -163.93 | 4.75 | 33[a] |
| 4S-Hyp (exo) | -48.66 | -33.65 | -132.09 | -173.50 | -169.37 | 3.26 | no helix |
| 4S-Mop (exo) | -48.78 | -36.01 | -128.03 | -169.86 | -165.40 | 1.00 | - |

*One additional HB

a. Hodges, J. A.; Raines, R. T. *JACS*. **2003**, *125*, 9262.

b. Vitagliano, L.; Nemethy, G.; Zagari, A.; Scheraga, H. A. *Biochemistry* **1993**, *32*, 7354

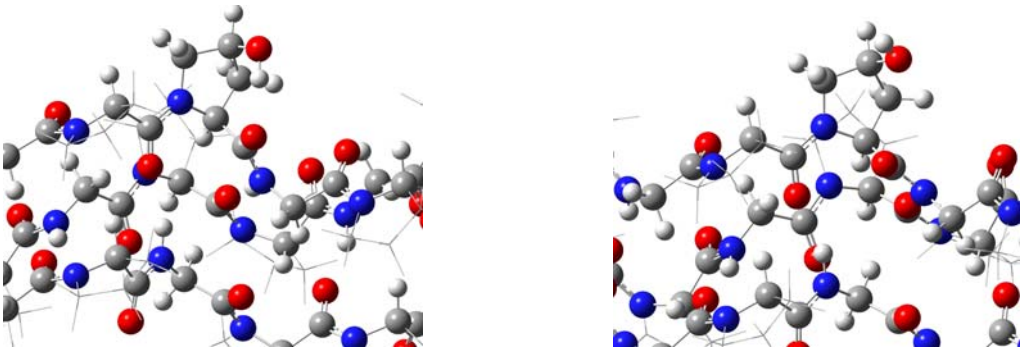


**Figure 2.3.2**: Two 4S-Hyp triple helix conformations. The left one has an internal HB

Figures 2.3.3 and 2.3.4 present the hydrogen bond distance in the triple helix of each mutation at the X and Y positions. For the sake of simplicity, the conformations can be divided into four groups R-endo, S-endo, R-exo, and S-exo. In the Y position, the hydrogen bonding energies for Mop are more stable for the R-exo and S-exo conformations because they have the shortest hydrogen bonding distance. However, in the case of Hyp, the most favorable conformation is the S-endo which is due, in this case, to an extra hydrogen bond. In contrast, Flp makes R-endo conformation the most stable because the average CH···O hydrogen bonding distance is the shortest for R-endo while the average NH···O distance is almost the same in the different conformations.

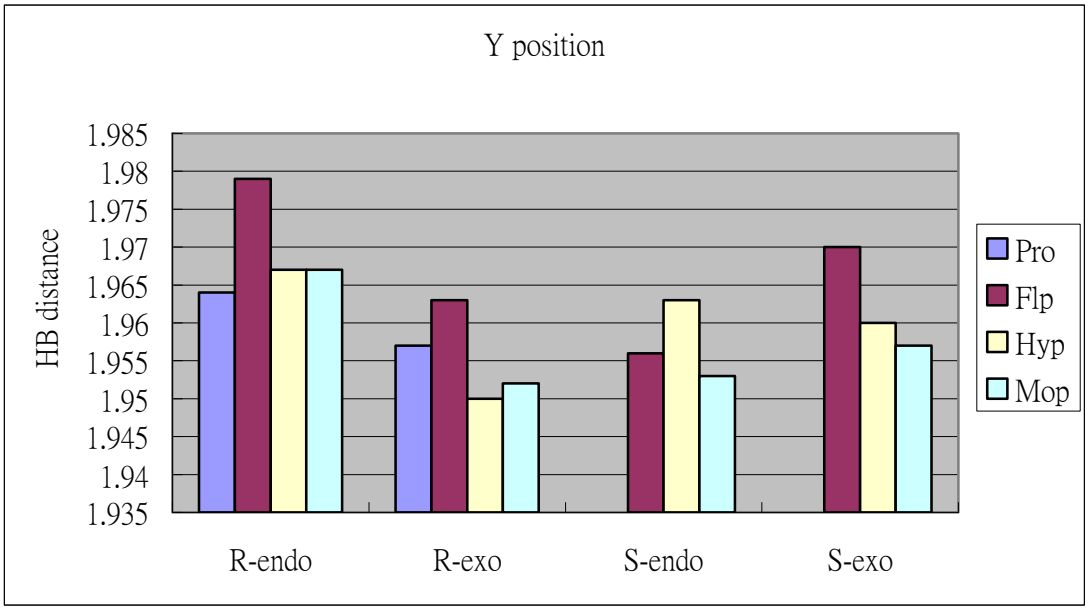



**Figure 2.3.3.** Hydrogen bond distance of mutated amino acids at the Y position

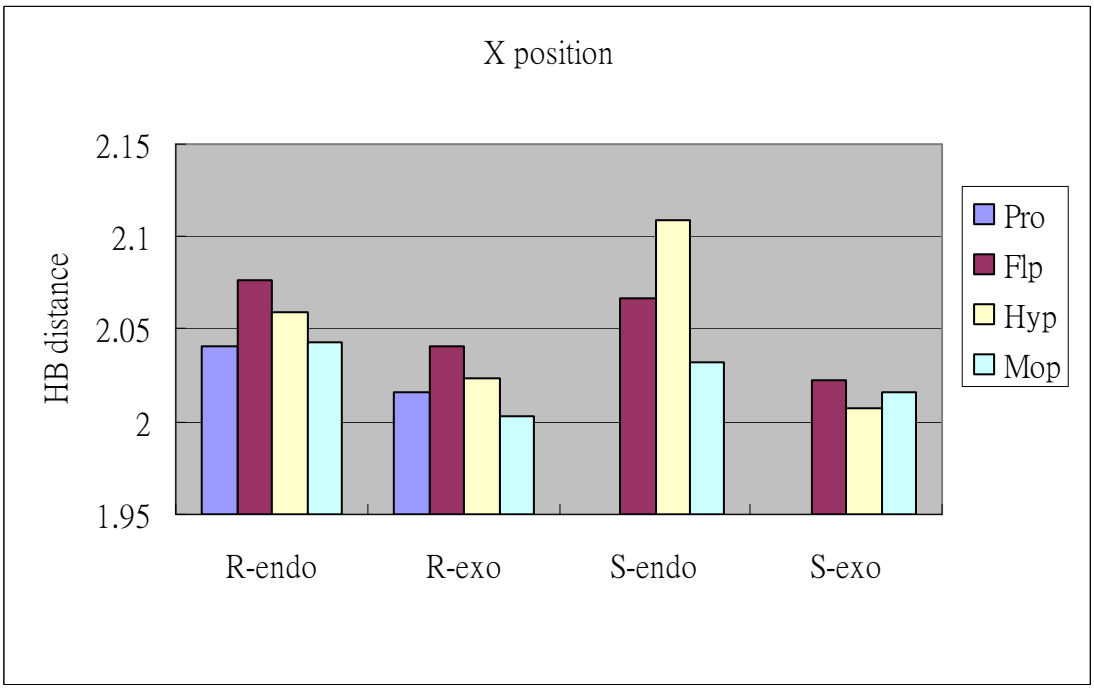



**Figure 2.3.4.** Hydrogen bond distance of mutated amino acids at the X position

The hydrogen bonds (HB) distance of mutated amino acids at 4R-exo Hyp are 0.002Å shorter than those in 4R-exo Mop in the Y position, while the average HB

they are 0.008Å longer in the Y position. In contrast, the average HB distances for Hyp S-endo conformation are 0.01 Å shorter than Flp and Mop S-endo conformation. Therefore, Hyp is more HB favorable in the S-endo conformation.

In the X position, the R-endo conformation is the most favorable for Hyp, whereas the S-exo conformer is found to be the most favorable for Mop. The additional stabilization of triple helices incorporating these two derivatives can be attributed to several weak CO···HN and CO···HC hydrogen bonding involving the peptide backbone. Notice that, S-endo-Hyp becomes less stable at X position because 4S-Hyp needs more space to accommodate the –OH group. This particular orientation of –OH group probably affects the inter-strand hydrogen bonding interaction (longer inter-strand H-bond distance; Figures 2.3.5).

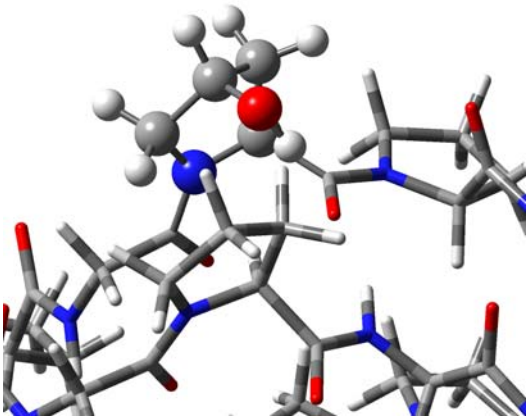


**Figure 2.3.5.** The triple helix conformation of 4S-Hyp at X position

The average distances for those weak HBs (CO···HN and CO···HC) at Hyp in the X position are found to be shorter for Flp, Mop and Pro. The mutated amino acid

HB distance for 4S-endo Flp is 0.04Å longer than for 4S-endo Mop, but the HB energy for 4S-endo Flp is 0.89 kcal/mol lower than for 4S-endo Mop because the average HB (NH···CO) distance is shorter for 4S-endo Flp.

Let us now compare the gas phase relative energies of the optimized strands with that of the triple helix (The second column of the Table 2.3.1 and 2.3.2). Interestingly, I have found an additional internal HB on the 4S-Hyp optimized strand (Figure 2.3.6) for both X and Y positions. Thus I can conclude that 4S-Hyp at these positions will not favor formation of triple helix. The gas phase relative energy for R conformation is lower than that of the S one and the endo conformation is more stable than the exo one for Flp, Hyp, Mop and Proline expect 4S-Mop (exo) at X position. However, as the published reports for the crystal structure do not confirm the preferred conformation of the ring, the lowest relative energies computed here have been used to compare with the experimental data [25].

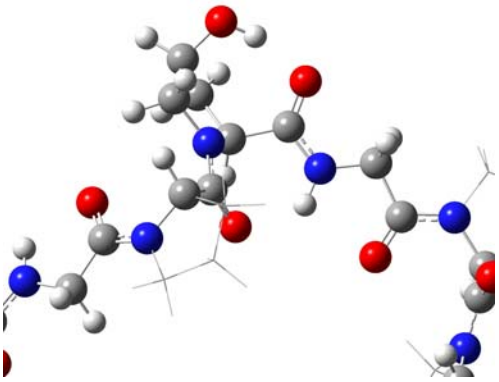


**Figure 2.3.6**. Additional internal HB on the 4S-Hyp optimized strand

The gas phase relative energy of optimized strand and triple helix show that R-endo-Flp is the most stable in the Y position amongst all the proline derivatives investigated here. The experimental melting temperature for 4R-Flp is 91 degrees [27]. According to experimental results, the second most stable structure is 4R-Mop [27] or 4R-Hyp [27, 37], while the less stable one is Pro [38]. So, the experimental stability ordering is Flp > Mop ≈ Hyp > Pro, in good agreement with the results found in my calculations, according to which the gas phase relative energies for 4R-Flp-endo, 4R-Hyp-endo, 4R-Mop-exo and Proline are -29.45, -28.83, -27.56 and -27.54 kcal/mol, respectively. The experimental results indicate that 4S-Flp and 4S-Hyp do not form the triple helix at the Y position, a fact consistent with the calculated small stability for those two conformations. There is however no experimental data for 4-Mop in order to compare the trends observed from my calculations. The optimization of the single strand with mutation at the X position leads to a more distorted structure than when mutation is in the Y position in such a way that the gas phase relative energy is much lower at the X position.

As described in the preceding chapter, I have calculated the aqueous solvation free energies for each of the optimized geometries using the Cramer and Truhlar solvation scheme SM5.2 [34]. Hyp is the most solvated triple helical due to the –OH group on the γ carbon which is able to form a hydrogen bond with water. The data in

tables 2.3.1 and 2.3.2 clearly show that 4-Hyp is the best solvated of the proline derivatives I have considered, since water stabilization is about 5 kcal/mol larger in 4-Hyp than in 4-Flp 4-Mop or Pro. In most of the cases, the order of the solvated free energies of triple helix conformation is Hyp > Mop > Flp > Pro. I have also calculated the single strand solvation free energies by fixing the geometry of the triple helix conformation. The results I obtained show that the order is the exactly same.

The solvated free energies for fixed geometry single strands are ~50 kcal/mol higher than the ones for triple helix in the Y position, this value being ~40 kcal/mol in the X position. These results are reasonable, since single strands have much more amide hydrogen atoms (N-H) in each Gly and carbonyl oxygen atoms (C=O) available to form hydrogen bonds with each other in triple helix structures. The fact that the X position is around 10 kcal/mol less solvated than the Y position means that the solvation energies of 4R substitution of proline at the Y position are higher than in the X position. It might be the reason why the stability arises by a network of bridging water molecules at the Y position not the X position. The solvation free energy is decreased by ~12.8 to ~17.4 kcal/mol for the Y position when it put in aqueous solution from fixed geometry of single strand to optimized single strand. The same change for X-position if found to be in the range of ~2.04 to ~5.44 kcal/mol depending on proline derivative.

It has been recently reported [27] that replacing Hyp by Mop in the Y-position increases the melting temperature of triple-helices less in comparison to replacement by Flp. This is in a good agreement with my suggestion that the value of ΔG for the folding/melting process is driven primarily by the stereoelectronic effect, which should be about the same for the methoxy and hydroxy groups. However, my data in table 2.3.1 show that Mop and Hyp are better solvated than Flp (More detail will be discussed in Chapter 3).

The energies, which are being discussed here, are defined as the energies in gas phase + salvation free energy of the triple helix. The 4S-Hyp (endo) triple helix and the single strand conformations have extra hydrogen bonding at both X and Y positions. However, the –OH group of Hyp can form a bridge with the water molecules. This internal HB changes the stability of the triple helix. Therefore, it is possible that my calculations could have neglected the relative energy for this conformation. The most stable structure is 4R-Flp (exo) in the Y position, in good agreement with the experimental results [27] (Melting temperature of 91°). The overall stability results of my study show that 4R-Flp > 4R-Hyp > 4S-Flp > 4S-Mop > Pro. The experimental ordering based on the melting points is 4R-Flp > 4R-Hyp ≈ 4R-Mop > Pro. (In the case of the less stable Pro structure, the melting point is 41 degree). However, it is worth mentioning that there are some differences between the

experimental data and my theoretical calculations. For instance, the experimental melting points were determined at a very low pH of 3, using HOAc as the solvent [27] or at a low temperature (4°) [25]. Notice also that, there is no experimental data for Mop. In my thesis, I was using water (neutral) as the solvent. In addition, I only mutate one amino acid in the triple helix, but in the experiments as high as 18 mutations were done at one time. Therefore, 4S-Flp (exo) and 4S-Hyp (exo) could not form the triple helix in the experiments, but they are more stable than Proline at the Y position in my study.

In the X position, 4R-Mop becomes the most stable than the various conformers of fluorine derivatives in triple helix. The ordering in my study is 4R-Mop > 4S-Mop > 4R-Hyp > 4R-Flp > Pro > 4S-Flp. In this chapter I have been able to explore the reasons behind preference of different proline derivatives in forming collagen triple helix. The results and trends reported here are well in agreement with available experimental reports. For instance, the experimental melting points of Pro and 4S-Flp are 41° and 33°, respectively, confirming that Pro is more stable than 4S-Flp, as I have found in this work.

**References:**


1. Bhattacharjee, A. and M. Bansal, *Collagen structure: The Madras triple helix and the current scenario.* IUBMB Life, 2005. **57**(3): p. 161-172.
2. Rich, A. and F.H. Crick, *The molecular structure of collagen.* Journal of molecular biology, 1961. **3**: p. 483-506.
3. Bella, J., et al., *Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution.* Science (New York, N.Y.), 1994. **266**(5182): p. 75-81.
4. Chan, D., M.S.P. Ho, and K.S.E. Cheah, *Aberrant signal peptide cleavage of collagen X in Schmid metaphyseal chondrodysplasia. Implications for the molecular basis of the disease.* Journal of Biological Chemistry, 2001. **276**(11): p. 7992-7997.
5. Persikov, A.V., et al., *Stability related bias in residues replacing glycines within the collagen triple helix (Gly-Xaa-Yaa) in inherited connective tissue disorders.* Human Mutation, 2004. **24**(4): p. 330-337.
6. Baum, J. and B. Brodsky, *Folding of peptide models of collagen and misfolding in disease.* Current Opinion in Structural Biology, 1999. **9**(1): p. 122-128.
7. Morokuma, K., *New challenges in quantum chemistry: quests for accurate calculations for large molecular systems.* Philosophical Transactions of the Royal Society of London, Series A Mathematical, Physical and Engineering Sciences, 2002. **360**(1795): p. 1149-1164.
8. Dapprich, S., et al., *A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives.* Theochem, 1999. **461-462**: p. 1-21.
9. Frisch, M.J., et al. , , *GAUSSIAN 03.* Gaussian, Inc.: Pittsburgh PA,, 2003.
10. Wieczorek, R. and J.J. Dannenberg, *Comparison of Fully Optimized alpha - and 310-Helices with Extended beta -Strands. An ONIOM Density Functional Theory Study.* Journal of the American Chemical Society, 2004. **126**(43): p. 14198-14205.
11. Wieczorek, R. and J.J. Dannenberg, *The Energetic and Structural Effects of Single Amino Acid Substitutions upon Capped alpha -Helical Peptides Containing 17 Amino Acid Residues. An ONIOM DFT/AM1 Study.* Journal of the American Chemical Society, 2005. **127**(49): p. 17216-17223.
12. Vitagliano, L., et al., *Structural bases of collagen stabilization induced by proline hydroxylation.* Biopolymers, 2001. **58**(5): p. 459-464.
13. Chopra, R.K. and V.S. Ananthanarayanan, *Conformational implications of enzymic proline hydroxylation in collagen.* Proceedings of the National Academy of Sciences of the United States of America, 1982. **79**(23): p.

7180-4.

14. Privalov, P.L., *Stability of proteins. Proteins which do not present a single cooperative system.* Advances in Protein Chemistry, 1982. **35**: p. 1-104.
15. Berg, R.A. and D.J. Prockop, *Thermal transition of a nonhydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple helix of collagen.* Biochemical and Biophysical Research Communications, 1973. **52**(1): p. 115-20.
16. Ramachandran, G.N., M. Bansal, and R.S. Bhatnager, *Hypothesis on the role of hydroxyproline in stabilizing collagen structure.* Biochim. Biophys. Acta, Protein Struct., 1973. **322**(1): p. 166-71.
17. Suzuki, E.F., R. D. B.; MacRae, T. P., Int. J. Biol. Macromol., 1980. **2**: p. 54-56.
18. Schumacher, M., K. Mizuno, and H.P. Baechinger, *The Crystal Structure of the Collagen-like Polypeptide (Glycyl-4(R)-hydroxyprolyl-4(R)-hydroxyprolyl)$_9$ at 1.55 Å Resolution Shows Up-puckering of the Proline Ring in the Xaa Position.* Journal of Biological Chemistry, 2005. **280**(21): p. 20397-20403.
19. Doi, M., et al., *Collagen-like triple helix formation of synthetic (Pro-Pro-Gly)$_{10}$ analogs: (4(S)-hydroxyprolyl-4(R)-hydroxyprolyl-Gly)$_{10}$, (4(R)-hydroxyprolyl-4(R)-hydroxyprolyl-Gly)$_{10}$ and (4(S)-fluoroprolyl-4(R)-fluoroprolyl-Gly)$_{10}$.* Journal of Peptide Science, 2005. **11**(10): p. 609-616.
20. Persikov, A.V., et al., *Amino acid propensities for the collagen triple-helix.* Biochemistry, 2000. **39**(48): p. 14960-14967.
21. Improta, R., C. Benzi, and V. Barone, *Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution.* Journal of the American Chemical Society, 2001. **123**(50): p. 12568-12577.
22. Vitagliano, L., et al., *Preferred proline puckerings in cis and trans peptide groups: implications for collagen stability.* Protein Science, 2001. **10**(12): p. 2627-2632.
23. Raines, R.T., *2005 Emil Thomas Kaiser Award.* Protein Science, 2006. **15**(5): p. 1219-1225.
24. Improta, R., et al., *Understanding the role of stereoelectronic effects in determining collagen stability. 2. A quantum mechanical/molecular mechanical study of (proline-proline-glycine)n polypeptides.* Journal of the American Chemical Society, 2002. **124**(26): p. 7857-7865.
25. Hodges, J.A. and R.T. Raines, *Stereoelectronic Effects on Collagen Stability:*

*The Dichotomy of 4-Fluoroproline Diastereomers.* Journal of the American Chemical Society, 2003. **125**(31): p. 9262-9263.

26. Horng, J.-C. and R.T. Raines, *Stereoelectronic effects on polyproline conformation.* Protein Science, 2006. **15**(1): p. 74-83.
27. Kotch, F.W., I.A. Guzei, and R.T. Raines, *Stabilization of the Collagen Triple Helix by O-Methylation of Hydroxyproline Residues.* Journal of the American Chemical Society, 2008. **130**(10): p. 2952-2953.
28. Bretscher, L.E., et al., *Conformational Stability of Collagen Relies on a Stereoelectronic Effect.* Journal of the American Chemical Society, 2001. **123**(4): p. 777-778.
29. Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, 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, K.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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; 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 03*, 2003.
30. Frisch, M.J., et al. , , *GAUSSIAN 09, Revision A.1.* Gaussian, Inc.: Pittsburgh PA,, 2009.
31. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange.* Journal of Chemical Physics, 1993. **98**(7): p. 5648-52.
32. Anil, B., et al., *Exploiting the right side of the Ramachandran plot: Substitution of glycines by D-alanine can significantly increase protein stability.* Journal of the American Chemical Society, 2004. **126**(41): p. 13194-13195.
33. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B Condensed Matter and Materials Physics, 1988. **37**(2): p. 785-9.
34. Cramer, C.J. and D.G. Truhlar, *Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics.* Chemical Reviews (Washington, D. C.), 1999. **99**(8):

p. 2161-2200.

35. AMAPC, Semichem, Inc.
36. Dewar, M.J.S., et al., *Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model.* Journal of the American Chemical Society, 1985. **107**(13): p. 3902-9.
37. Mizuno, K., et al., *The Peptides Acetyl-(Gly-3(S)Hyp-4(R)Hyp)$_{10}$-NH$_2$ and Acetyl-(Gly-Pro-3(S)Hyp)$_{10}$-NH$_2$ Do Not Form a Collagen Triple Helix.* Journal of Biological Chemistry, 2004. **279**(1): p. 282-287.
38. Holmgren, S.K., et al., *Code for collagen's stability deciphered.* Nature (London), 1998. **392**(6677): p. 666-667.

# Chapter 3

## Aqueous Solvation affects Preferred Conformations of Proline, and its 4-Hydroxy, Methoxy, and Fluoro Derivatives

### 3.1 Introduction

Collagen is the most abundant protein in the human body. Its structure consists of triple helices in which every third amino acid residue is glycine. Each of the three individual peptide strands can be thought of as $(X\text{-}Y\text{-}G)_N$, where X and Y can be any amino acid, G is glycine and N can be > 1000 depending upon the source. While X and Y can be any amino acid, proline (Pro) and its derivative 4R-hydroxyproline (Hyp) are the most abundant by far. Hyp, not one of the 20 'normal' amino acids listed in most texts, forms as a metabolite of Pro and has two possible diastereomers (assuming I only consider $^{L}$proline as a precursor) which I shall refer to as 4R and 4S (the configuration at the α-carbon always assumed to be that of $^{L}$proline). Hyp occurs

preferentially at the Y-position and Pro at the X-position.

There appears to be a consensus that the presence of Hyp in the Y position enhances the stability of the triple helical structures in which it is found. However, the cause of the enhanced stabilization remains controversial. The presence of the hydroxyl group in Hyp but not Pro obviously suggests that hydrogen-bonding might be instrumental in the stabilization. Not surprisingly, several suggestions that use hydrogen-bonds either simply or via water bridges have been promulgated to explain the stability [1, 2]. More recently, Vitagliano, et al. have suggested that the Hyp better accommodates the preferred ‘up’ (exo) pucker in the Y position than the preferred ‘down’ (endo) pucker in the X-position of the XYG traid [3]. On the basis of Hartree-Fock (HF) calculations optimized using the polarizable continuum model (CPCM) Barone, et al, have reported that electronegative substituents in the 4R position of proline stabilize up-puckering (exo), while 4S substitutions favor down-puckering (endo) of the proline ring [4]. Raines, et al., have subsequently suggested that the dominant effect might be stereoelectronic rather than H-bonding since peptides in which 4-fluoroproline (Flp) replaces Hyp form more stable triple helices from aqueous solutions of the unfolded peptides [5, 6]. They suggest that replacing Hyp with Flp imparts a stereoelectronic effect that changes the preferred conformation of the proline derivative to one that is preorganized to form a triple

helix. While experimental determinations of melting temperatures on the small homogenous triple helices $(Pro\text{-}Flp\text{-}Gly)_N$ (N=7 or 10) melt at higher temperatures than those of $(Pro\text{-}Hyp\text{-}Gly)_N$ , this effect appears less clear for single Flp→Hyp mutations in triple helices [7, 8].

Since the energy of folding represents the difference in energies between the triple helix and the three unfolded peptides from which it forms, one should consider the effects of the mutation Flp → Hyp upon the energies of the unfolded structures and the effect of differential solvation upon the equilibria. Thus, the change in equilibrium towards the triple helix upon Flp > Hyp mutation could be due either to stabilization of the triple helix, destabilization of the unfolded form, or a combination of both. In this chapter, I consider the relative energies of various conformations of Pro, Hyp and Flp and 4-methoxyproline (Mop) in the gas phase and the relative aqueous solvation energies of each. Mop has been included as it contains an oxygen but can not form an H-bond as a donor. A recent report shows that Mop (which can participate in H-bonding as an acceptor) has a stereoelectronic effect similar to Hyp [9].

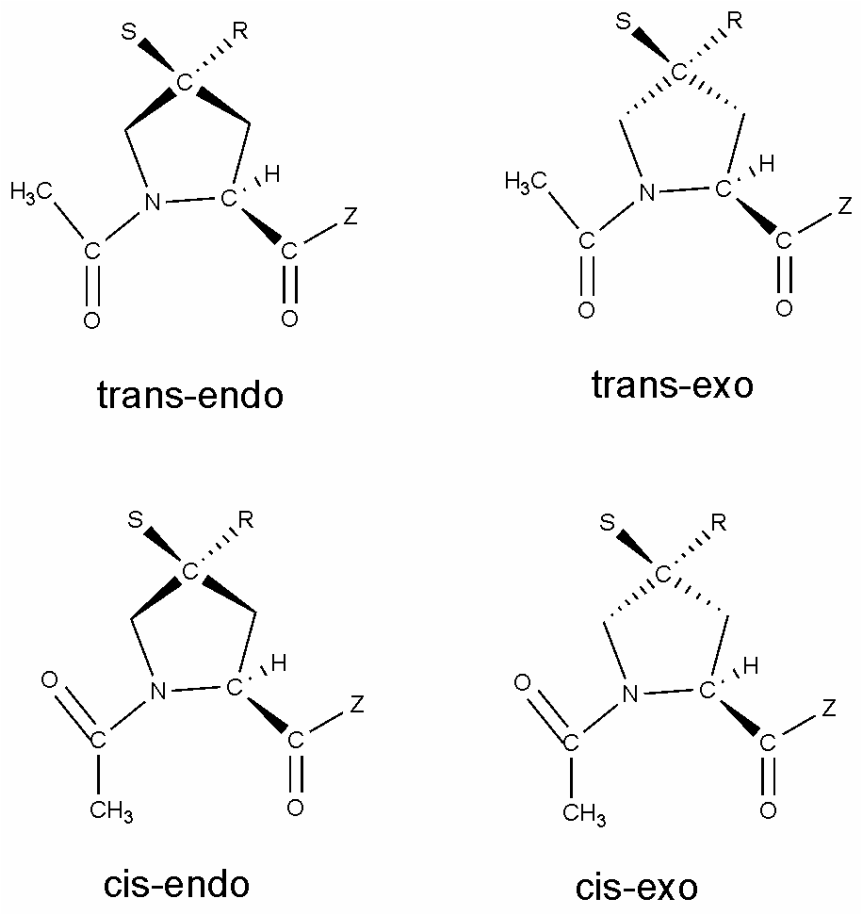



**Figure 3.1.** Structures for the various substituted capped prolines. The substituent (OH, F, OMe) is in the R- or S- position, while the capping group (OMe, $NH_2$, cis or trans NHMe, or $NMe_2$) is in the Z-position.

### 3.2 Results

Figure 3.1 presents the structures of the compounds considered, while table 3.1 presents the relative enthalpies for acetylPro, acetlyHyp, acetylFlp and acetylMop capped with five different groups (OMe, $NH_2$ , cis-NHMe, trans-NHMe, and $NMe_2$) at the O-terminus. Notice that, all structures capped with $NH_2$ and NHMe(trans) contain NH⋯O H-bonds. All structures capped with NHMe(cis) and $NMe_2$ contain a CH⋯OH bond. I shall subsequently refer to them simply as Pro, Hyp, Flp and Mop. I

initially focused on the derivatives with amino capping groups, as these seem to best mimic the environment of a proline within a peptide. Much to our surprise, I found that our results differed substantially from those from earlier calculations by DeRider, et al., who used methyl esters (rather than amides) of the prolines as the models for their calculations[10]. Since the DFT methods were slightly different (they used B3LYP/6-311+G(2d,p)), I repeated their calculations using our methods. As can be seen from Table 3.1, my results do not differ substantially from theirs. Rather, it appears that the nature of the capping group at the O-terminus has a significant effect upon the relative stabilities of the conformations of these molecules. Most notably, the two cis conformations (endo and exo) become significantly less stable than the trans isomers when capped by amides than by the methyl ester. This trend holds for proline and its three derivatives.

**Table 3.1.** Calculated (B3LYP-d95(d,p)) relative enthalpies at 298 K (kcal/mol) of isomers of proline and its derivatives without solvation.

| | Capping Group | | | | | |
|---|---|---|---|---|---|---|
| Isomer | OMe[a] | OMe | $NH_2$ | NHMe(Cis) | NHMe(Trans) | $NMe_2$ |
| | | | **Proline** | | | |
| trans-exo | 0.41 | 0.49 | 1.11 | 1.15 | 0.46 | 0.50 |
| cis-exo | 1.80 | 1.87 | 7.06 | 6.84 | 3.76 | 3.56 |
| trans-endo | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| cis-endo | 1.20 | 1.26 | 6.57 | 6.43 | 3.07 | 2.78 |
| | | | **4R-Fluoropline** | | | |
| trans-exo | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| cis-exo | 1.30 | 1.29 | 6.23 | 5.98 | 3.37 | 3.30 |
| trans-endo | 0.85 | 0.75 | 0.85 | 0.86 | 0.51 | 0.34 |
| cis-endo | 2.48 | 2.36 | 7.44 | 7.18 | 4.41 | 4.17 |

**Table 3.1. Continue**

| | | | | | | |
|---|---|---|---|---|---|---|
| **4S-Fluoropline** | | | | | | |
| trans-exo | 0.61 | 0.54 | 0.20 | 0.04 | 0.00 | 0.00 |
| cis-exo | 2.12 | 2.06 | 6.37 | 6.01 | 3.41 | 3.15 |
| trans-endo | 0.00 | 0.03 | 0.00 | 0.00 | 1.62 | 1.32 |
| cis-endo | 0.13 | 0.00 | 5.33 | 5.26 | 2.67 | 2.10 |
| **4R-Hydroxyproline** | | | | | | |
| trans-exo | | 0.04 | 0.42 | 0.41 | 0.37 | 0.45 |
| cis-exo | | 0.85 | 6.21 | 6.00 | 3.01 | 2.91 |
| trans-endo | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| cis-endo | | 1.45 | 6.98 | 6.81 | 3.63 | 3.43 |
| **4S-Hydroxyproline** | | | | | | |
| trans-exo | | 4.93 | 6.77 | 7.09 | 7.27 | 7.46 |
| cis-exo | | 6.01 | 12.23 | 12.35 | 10.04 | 9.95 |
| trans-endo | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| cis-endo | | 2.38 | 7.44 | 7.19 | 5.28 | 5.20 |
| **4R-Methoxyproline** | | | | | | |
| trans-exo | | 0.27 | 0.65 | 0.65 | 0.49 | 0.55 |
| cis-exo | | 1.12 | 6.39 | 6.18 | 3.27 | 3.14 |
| trans-endo | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| cis-endo | | 1.38 | 6.83 | 6.68 | 3.49 | 3.24 |
| **4S-Methoxyproline** | | | | | | |
| trans-exo | | 0.00 | 0.00 | 1.15 | 0.00 | 0.00 |
| cis-exo | | 1.15 | 5.53 | 6.49 | 2.88 | 2.60 |
| trans-endo | | 0.72 | 1.15 | 0.00 | 2.96 | 3.24 |
| cis-endo | | 0.70 | 5.46 | 7.60 | 3.39 | 3.63 |

For proline, the trans-endo conformation remains the most stable of the isomers considered. The results for 4R-Hyp and Mop are consistent with those for Pro, as the trans-endo is most stable for derivatives with all five capping groups. This situation changes for 4R-Flp, where the R-trans-exo is consistently the most stable. The situation for the 4S isomers becomes a bit more complex. Only 4S-Hyp has the same

most stable conformation (trans-endo) for all five capping groups, probably due to the additional H-bond found only in this conformation (Figure3.2). For 4S-Flp, the trans and cis endo have equivalent energies (within 0.03 kcal/mol) when capped with OMe, the trans endo and exo have equivalent energies (within 0.04 kcal/mol) when capped with trans-NHMe, the trans-endo is most stable when capped with $NH_2$, and the trans-exo when capped by cis-NHMe or $NMe_2$ (these last two structures lack an H-bond present when capped by the other groups). Finally, 4S-Mop prefers the trans-exo conformation for all capping groups but trans-NHMe despite the fact that the structures lack a H-bond when capped by cis-NHMe or $NMe_2$ (as is the case for 4S-Flp).

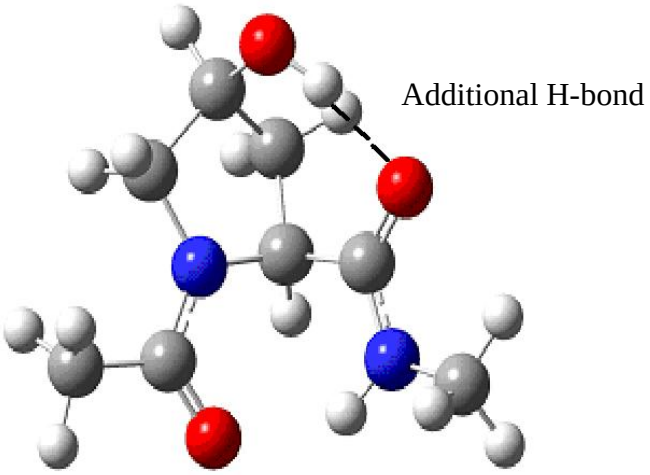



**Figure 3.2**: Additional H-bond only shows on 4S-Hyp trans-endo conformation

### 3.3.1 Aqueous solvation

Table 3.2 presents the results for the aqueous solvation energies of each the

species considered. Notice that, all structures capped with $NH_2$ and NHMe(trans) contain NH…O H-bonds. All structures capped with NHMe(cis) and $NMe_2$ contain a C-H…O H-bond. One should immediately note that Hyp is generally the best solvated for each conformation/capping group considered with the exception of the S-trans-endo isomers. The Mop molecules are not so well stabilized by aqueous solvation as the Hyp's, but they are better solvated than the Pro's or Flp's. These results accord well with the general experience that alcohols (Hyp) are reasonably soluble in water, while hydrocarbons (Pro) and fluorocarbons (Flp) are not. Ethers, while not generally water soluble, can behave as a hydrogen-bond acceptors, so the intermediate behavior of Mop would be expected.

**Table 3.2** Solvation free energies (kcal/mol) calculated using SM5.2 for each conformation of each proline derivative considered.

| | Capping Group | | | | |
|---|---|---|---|---|---|
| Isomer | OMe | $NH_2$ | NHMe(Cis) | NHMe(Trans) | $NMe_2$ |
| **Proline** | | | | | |
| trans-exo | -13.30 | -16.45 | -17.04 | -14.32 | -14.16 |
| cis-exo | -11.42 | -20.56 | -18.32 | -17.48 | -14.94 |
| trans-endo | -12.82 | -16.49 | -16.12 | -14.26 | -13.18 |
| cis-endo | -11.54 | -20.47 | -18.26 | -17.34 | -14.84 |
| **4R-Fluoropline** | | | | | |
| trans-exo | -13.58 | -16.85 | -16.86 | -14.70 | -13.78 |
| cis-exo | -11.71 | -20.70 | -18.35 | -17.60 | -14.98 |
| trans-endo | -11.55 | -15.62 | -14.68 | -13.31 | -11.54 |
| cis-endo | -10.96 | -20.14 | -17.78 | -16.91 | -14.33 |
| **4S-Fluoropline** | | | | | |
| trans-exo | -12.62 | -16.32 | -16.69 | -14.08 | -13.81 |

**Table 3.2 Continue**

| | | | | | |
|---|---|---|---|---|---|
| cis-exo | -11.22 | -20.65 | -18.24 | -17.56 | -14.75 |
| trans-endo | -14.17 | -18.81 | -19.44 | -16.55 | -16.22 |
| cis-endo | -12.98 | -21.48 | -19.73 | -18.54 | -16.45 |
| **4R-Hydroxyproline** | | | | | |
| trans-exo | -19.02 | -22.13 | -22.41 | -19.96 | -19.36 |
| cis-exo | -16.80 | -25.82 | -23.46 | -22.72 | -20.04 |
| trans-endo | -17.34 | -21.24 | -20.53 | -18.97 | -17.56 |
| cis-endo | -16.64 | -25.81 | -23.43 | -22.64 | -20.03 |
| **4S-Hydroxyproline** | | | | | |
| trans-exo | -18.32 | -21.87 | -22.37 | -19.68 | -19.49 |
| cis-exo | -16.77 | -26.02 | -23.75 | -23.01 | -20.26 |
| trans-endo | -16.90 | -19.42 | -18.92 | -17.08 | -16.05 |
| cis-endo | -15.73 | -24.08 | -21.93 | -20.71 | -18.86 |
| **4R-Methoxyproline** | | | | | |
| trans-exo | -15.06 | -18.26 | -18.53 | -16.05 | -15.48 |
| cis-exo | -12.98 | -22.01 | -19.67 | -18.91 | -16.26 |
| trans-endo | -13.44 | -17.35 | -16.72 | -15.09 | -13.69 |
| cis-endo | -12.89 | -22.07 | -19.73 | -18.93 | -16.27 |
| **4S-Methoxyproline** | | | | | |
| trans-exo | -14.36 | -17.99 | -18.49 | -15.77 | -15.66 |
| cis-exo | -12.94 | -22.20 | -19.96 | -19.24 | -16.37 |
| trans-endo | -14.87 | -19.80 | -19.14 | -16.54 | -15.97 |
| cis-endo | -13.80 | -22.17 | -20.76 | -19.35 | -17.23 |

The data in table 3.3 (notice that, all structures capped with $NH_2$ and NHMe(trans) contain NH···O H-bonds. All structures capped with NHMe(cis) and $NMe_2$ contain a C-H···O H-bond.) indicate that upon aqueous solvation, the trans-exo conformations of all the 4R-substituted prolines become most stable for all the capping groups considered, while the trans-endo conformation becomes most stable for 4S-Flp and 4S-Hyp for all capping groups. Solvated 4S-Mop is most stable in the

trans-endo conformation for the capping groups OMe, $NH_2$, and trans-NHMe, but the additional H-bonds made by the hydroxyl group reduce the solvation energies of the trans-endo conformation for cis-NHMe and $NMe_2$, rendering the trans-exo conformations slightly (0.3 kcal/mol) more stable for these capping groups. Solvated proline, itself, has similar stabilities for both trans conformations, with the trans endo more stable for two of the amido capping groups and the trans exo for the other two. Solvated proline capped with OMe has essentially the same energy in either of the trans conformations.

**Table 3.3** Relative energies (enthalpy in gas phase + free energy of solvation) in kcal/mol.

| | Capping Group | | | | |
|---|---|---|---|---|---|
| Isomer | OMe | $NH_2$ | NHMe(Cis) | NHMe(Trans) | $NMe_2$ |
| | | | **Proline** | | |
| trans-exo | 0.02 | 1.15 | 0.22 | 0.42 | 0.00 |
| cis-exo | 3.28 | 2.99 | 4.63 | 0.55 | 2.26 |
| trans-endo | 0.00 | 0.00 | 0.00 | 0.01 | 0.47 |
| cis-endo | 2.54 | 2.58 | 4.29 | 0.00 | 1.59 |
| | | | **4R-Fluoropline** | | |
| trans-exo | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| cis-exo | 3.17 | 2.38 | 4.50 | 0.47 | 2.10 |
| trans-endo | 2.78 | 2.08 | 3.04 | 1.90 | 2.58 |
| cis-endo | 4.98 | 4.16 | 6.27 | 2.19 | 3.63 |
| | | | **4S-Fluoropline** | | |
| trans-exo | 2.07 | 2.69 | 2.79 | 1.79 | 1.09 |
| cis-exo | 4.98 | 4.53 | 7.20 | 1.72 | 3.30 |
| trans-endo | 0.00 | 0.00 | 0.00 | 0.94 | 0.00 |
| cis-endo | 1.16 | 2.66 | 4.96 | 0.00 | 0.55 |
| | | | **4R-Hydroxyproline** | | |
| trans-exo | 0.00 | 0.00 | 0.00 | 0.12 | 0.00 |

**Table 3.3 Continue**

| | | | | | |
|---|---|---|---|---|---|
| cis-exo | 3.04 | 2.10 | 4.54 | 0.00 | 1.77 |
| trans-endo | 1.64 | 0.47 | 1.47 | 0.74 | 1.34 |
| cis-endo | 3.79 | 2.89 | 5.37 | 0.70 | 2.30 |
| | | **4S-Hydroxyproline** | | | |
| trans-exo | 3.52 | 4.32 | 3.64 | 4.67 | 4.02 |
| cis-exo | 6.14 | 5.63 | 7.52 | 4.11 | 5.73 |
| trans-endo | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| cis-endo | 3.55 | 2.78 | 4.18 | 1.65 | 2.39 |
| | | **4R-Methoxyproline** | | | |
| trans-exo | 0.00 | 0.00 | 0.00 | 0.08 | 0.00 |
| cis-exo | 2.93 | 2.00 | 4.39 | 0.00 | 1.80 |
| trans-endo | 1.34 | 0.26 | 1.16 | 0.55 | 1.24 |
| cis-endo | 3.27 | 2.38 | 4.83 | 0.20 | 1.89 |
| | | **4S-Methoxyproline** | | | |
| trans-exo | 0.00 | 0.66 | 1.80 | 0.59 | 0.00 |
| cis-exo | 2.58 | 1.99 | 5.67 | 0.00 | 1.90 |
| trans-endo | 0.21 | 0.00 | 0.00 | 2.78 | 2.93 |
| cis-endo | 1.25 | 1.95 | 5.98 | 0.40 | 2.07 |

The results of Barone's earlier study [4] differed from ours in some important aspects. As his paper only reports the energies calculate using the CPCM solvent model (no gas phase calculations were reported), I could not determine if the differences resulted from discrepancies in the gas phase energies, the solvation energies, or both. I repeated the CPCM calculations reported [4] using both HF/6-31G(d) (Barone's method) and B3LYP/d95(d,p). The implementation of the CPCM model in Gaussian 03 was not robust. However, this was greatly improved in Gaussian 09[11]. All the CPCM calculation reported here used GAUSSIAN 09 unless otherwise indicated. As seen from Table 3.4, the apparent anomalies disappear when

the B3LYP/D95(d,p) CPCM calculations are reproduced with this later version of GAUSSIAN.

**Table 3.4** Comparison of relative energies (enthalpy in gas phase + free energy of solvation) in kcal/mol for NHMe(trans) capping group using DFT/AM1(SM5.2) and CPCM with those previously reported

| Isomer | DFT/AM1(SM5.2) | DFT/CPM(G09) | HF/CPM (Barone) | HF/CPM (G03) |
|---|---|---|---|---|
| **Proline** | | | | |
| trans-exo | 0.4 | 0.1 | 0.4 | 0.3 |
| cis-exo | 0.5 | 1.2 | 1.2 | 1.8 |
| trans-endo | 0.0 | 0.0 | 0.0 | 0.0 |
| cis-endo | 0.0 | 0.7 | 0.9 | 1.1 |
| **4R-Fluoroproline** | | | | |
| trans-exo | 0.0 | 0.0 | 0.0 | 0.0 |
| cis-exo | 0.5 | 1.1 | 2.6 | 1.5 |
| trans-endo | 1.9 | 1.7 | 1.1 | 1.5 |
| cis-endo | 2.2 | 2.4 | 2.3 | 2.6 |
| **4S-Fluoroproline** | | | | |
| trans-exo | 1.8 | 1.3 | 1.1 | 1.1 |
| cis-exo | 1.7 | 2.4 | 2.4 | 2.6 |
| trans-endo | 0.9 | 0.0 | 0.2 | 0.0 |
| cis-endo | 0.0 | 0.4 | 0.0 | 0.4 |
| **4R-Hydroxyproline** | | | | |
| trans-exo | 0.1 | 0.0 | 0.0 | 0.0 |
| cis-exo | 0.0 | 1.1 | 2.0 | 1.4 |
| trans-endo | 0.7 | 0.6 | 0.5 | 0.6 |
| cis-endo | 0.7 | 1.3 | 1.3 | 1.8 |
| **4S-Hydroxyproline** | | | | |
| trans-exo | 0.6 | 0.0 | 1.5 | 0.0 |
| cis-exo | 0.0 | 1.1 | 2.7 | 1.4 |
| trans-endo | 2.8 | 0.3 | 0.0 | 0.1 |
| cis-endo | 0.4 | 0.8 | 1.8 | 1.1 |

### 3.4 Discussion

The equilibrium constant for folded collagen-like triple helices and unfolded peptide strands depends upon the ΔG between the two states in the solvent used for the experiment. As my mentor and I have previously emphasized, [12, 13] one must consider the effect of perturbations (i.e., a Flp → Hyp mutation) upon both states (folded and unfolded) involved in the equilibria. Most of the previous analyses of the added stability afforded by Hyp, Flp and Mop have focused upon the presumed stabilizing effects provided to the triple helical structure by the energies required to distort the substituted prolines from their preferred conformation to those they assume in the collagenic triple helical structure[10, 14], while others concerned with the effects of Hyp → Pro mutations have concentrated on possible H-bonding [1, 2]. To our knowledge the effects of the solvation upon the unfolded peptide structures have not previously been reported.

According to experimental reports based upon melting temperatures of collagen models containing triple helices formed from strands of repeating seven or ten triads of XYG's, 4R-Hyp and 4R-Flp [15, 16] and 4R-Mop [9] in the Y position stabilize the respective triple helices compared to the all PPG [17] standard. Experimental reports have also shown the conformation of Pro residues in triple helices made for PPG triads to be trans-endo in the X-position and trans-exo in the Y position [3].

The calculations of the gas phase (unsolvated) energies for the 4R derivatives show that only 4R-Flp prefers the trans-exo conformation that it must assume in the Y-positions of the triple helices (in accord with NMR data in a nonpolar solvent [10]). Based upon these gas phase calculations, one would assume that an additional distortion energy would be necessary for 4R-Hyp (and 4R-Mop) to assume this conformation, which would suggest that the two latter residues would form less stable triple helices than Flp. However, these relative distortion energies disappear when aqueous solvation is considered, as all three 4R-substituted prolines rather robustly prefer the trans-exo conformations. Thus, the suggestion that an energy associated with conformational distortion must be added to the $\Delta G$ of folding/melting, while a reasonable hypothesis based upon gas phase calculations, appears to dissolve in aqueous solution. One should note that the NMR data that agreed with the calculated gas phase conformations reported by DeRider, et al. were obtained in dioxane (rather than aqueous) solvent to better mimic the gas phase conformations [10]. Thus, these experiments do not contradict our findings that the conformations in aqueous solution may differ from those of the gas phase.

Of the 4S-substituted prolines, 4S-Flp prefers the trans-endo conformation in the gas phase with the capping groups $NH_2$ and trans-NHMe, but the trans-exo with cis-NHMe and $NMe_2$ (presumably, because these structures lack an H-bond common

to the others), while the trans and cis-endo conformations have roughly the same energy for OMe. For 4S-Hyp, the trans-endo is the most stable for all capping groups, while 4S-Mop prefers trans -exo for all capping groups except trans-NHMe. When put in aqueous solution, both 4S-Flp and 4S-Hyp robustly prefer the trans-endo conformations, while 4S-Mop prefers this conformation except when capped by cis-NHMe and $NMe_2$, both of which lack an H-bond that the others have.

Proline, itself, prefers the trans-endo conformation in both the gas phase and solution when capped with OMe, $NH_2$, or trans-NHMe. When capped by cis-NHMe or $NMe_2$, proline prefers the trans-endo in the gas phase, but the trans-exo in solution.

The preferred conformation in the gas and aqueous phases together with the relative solvation free energies of the most stable solution conformation when capped with $NH_2$ (the values for the other capping groups can be computed from tables 3.1-3.3) are collected in table 3.5. As can be seen from tables 3.3 and 3.5, 4R-Flp, Hyp and Mop all prefer the trans-exo conformation that is found in the Y position of collagen triple helices. These data correlate well with the experimental findings that 4R-Flp and Hyp and Mop form higher melting triple helices when they are in the Y position with Pro in the X-position. Experimentally, 4S-Flp also forms a relatively high melting triple helix when in the X-position (with Pro in the Y-position) in agreement with our finding that solvated 4S-Flp prefers the endo conformation that is

generally found in the X-position of triple helices. Although Raines, et al. could not find a triple helix for peptides with the 4(R or S)-Hyp-Pro-Gly sequence[15], Bachinger, et al. reported that acetyl(4R-Hyp-4R-Hyp-Gly)$_{10}$NH$_2$ has a slightly higher melting temperature than acetyl(Pro-4R-Hyp-Gly)$_{10}$NH$_2$ [18].

Table 3.5 recapitulates the relevant theoretical and experimental data. The relative energy vs. the gas phase is calculated as the energy required to distort the molecule in question from its preferred gas phase conformation to that preferred in solution plus the solvation energy for the latter conformation. In the table zero is set as the lowest relative energy vs. the gas phase. The data of table 3.2 clearly show 4R-Hyp to be the best solvated of the proline derivatives considered. Water stabilizes it about 5 kcal/mol more than 4R-Flp or Pro. This behavior can be expected, as alcohols are generally better solvated than hydrocarbons or fluorocarbons by water. If I assume that the data for single proline derivatives can be applied to unfolded peptides, the unfolded peptides containing 4R-Flp will be destabilized in water relative to 4R-Hyp by about 5 kcal/mol per Hyp $\rightarrow$ Flp mutation. One might expect a smaller, but qualitatively similar effect upon the solvation of the triple helices, as much of the surface of the residues in the triple helices will not be directly exposed to solvent. Thus, the ~ 5 kcal/mol destabilization of 4R-Flp relative to 4R-Hyp should be taken as an upper limit. Consequently, the cause of the added stability afforded by

4R-Flp over 4R-Hyp might be substantially due to the destabilization of 4R-Flp in water relative to 4R-Hyp.

The recent report that replacing Hyp with Mop in the Y-position increases the melting temperature of triple-helices less than Flp agrees with the suggestion that the $\Delta G$ for the folding/melting process is driven primarily by the relative solvation energies of the unfolded state as the stereoelectronic effect of the methoxy group should be about the same as that of the hydroxy group. However, the data in table 3.2 show Mop to be better solvated than Flp, but not so much as Hyp. If one makes an overly simplified assumption (for the purpose of illustration) that the differential solvation of the unfolded structures be the only operative effect, one would imagine an energy diagram similar to figure 3.3. This figure assumes that triple helical states be more stable that the unfolded states, which would be the case below the melting temperatures.

### 3.5 Conclusions

The stabilization due to aqueous solvation of 4R-substituted acetylprolines follows the order Hyp > Mop > Flp, which is the reverse order of the stabilization they impart when incorporated in the Y-position for stabilizing collegenic triple helices. This observation suggests that aqueous solvent destabilization of Flp in the

unfolded peptide (where it is more exposed to solvent than in the triple helix) might be the dominant factor in the enhanced stability of triple helices that contain 4R-Flp in the Y-position.

While a stereoelectronic effect may cause 4R-Flp (but not 4R-Hyp or 4R-Mop) to prefer the exo conformation necessary for incorporation into the triple helix in the gas phase, all three prefer the exo conformation in aqueous solution. As the energetic hurdle for Hyp to achieve the requisite conformation in the gas phase disappears once placed in aqueous solution, it should not be a determining factor in the relative preferences of the 4R-substituted prolines for incorporation into the Y-position of the triple helices.

**References**:


1. Bella, J., et al., *Crystal and molecular structure of a collagen-like peptide at 1.9 .ANG. resolution.* Science (Washington, DC, United States), 1994. **266**(5182): p. 75-81.
2. Kramer, R.Z., et al., *Sequence dependent conformational variations of collagen triple-helical structure.* Nat. Struct. Biol., 1999. **6**(5): p. 454-457.
3. Vitagliano, L., et al., *Structural bases of collagen stabilization induced by proline hydroxylation.* Biopolymers, 2001. **58**(5): p. 459-464.
4. Improta, R., C. Benzi, and V. Barone, *Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution.* Journal of the American Chemical Society, 2001. **123**(50): p. 12568-12577.
5. Hodges, J.A. and R.T. Raines, *Stereoelectronic Effects on Collagen Stability: The Dichotomy of 4-Fluoroproline Diastereomers.* Journal of the American Chemical Society, 2003. **125**(31): p. 9262-9263.
6. Horng, J.-C. and R.T. Raines, *Stereoelectronic effects on polyproline conformation.* Protein Science, 2006. **15**(1): p. 74-83.
7. Persikov, A.V., et al., *Triple-Helix Propensity of Hydroxyproline and Fluoroproline: Comparison of Host-Guest and Repeating Tripeptide Collagen Models.* Journal of the American Chemical Society, 2003. **125**(38): p. 11500-11501.
8. Malkar, N.B., et al., *Modulation of Triple-Helical Stability and Subsequent Melanoma Cellular Responses by Single-Site Substitution of Fluoroproline Derivatives.* Biochemistry, 2002. **41**(19): p. 6054-6064.
9. Kotch, F.W., I.A. Guzei, and R.T. Raines, *Stabilization of the Collagen Triple Helix by O-Methylation of Hydroxyproline Residues.* Journal of the American Chemical Society, 2008. **130**(10): p. 2952-2953.
10. DeRider, M.L., et al., *Collagen Stability: Insights from NMR Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations.* Journal of the American Chemical Society, 2002. **124**(11): p. 2497-2505.
11. Frisch, M.J., et al. , , *GAUSSIAN 09, Revision A.1.* Gaussian, Inc.: Pittsburgh PA,, 2009.
12. Tsai, M., Y. Xu, and J.J. Dannenberg, *Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProPro$^{D}$Ala, and ProPro$^{D}$Ser Triads.* Journal of the

American Chemical Society, 2005. **127**(41): p. 14130-14131.

13. Dannenberg, J.J., *The importance of cooperative interactions and a solid-state paradigm to proteins: what peptide chemists can learn from molecular crystals.* Advances in Protein Chemistry, 2006. **72**(Peptide Solvation and H-Bonds): p. 227-273.
14. Vitagliano, L., et al., *Stabilization of the triple-helical structure of natural collagen by side-chain interactions.* Biochemistry, 1993. **32**(29): p. 7354-9.
15. Holmgren, S.K., et al., *Code for collagen's stability deciphered.* Nature (London), 1998. **392**(6677): p. 666-667.
16. Holmgren, S.K., et al., *A hyperstable collagen mimic.* Chemistry & Biology, 1999. **6**(2): p. 63-70.
17. Shaw, B.R. and J.M. Schurr, *Association reaction of collagen model polypeptides (Pro-Pro-Gly)n.* Biopolymers, 1975. **14**(9): p. 1951-85.
18. Mizuno, K., et al., *The Peptides Acetyl-(Gly-3(S)Hyp-4(R)Hyp)$_{10}$-NH$_2$ and Acetyl-(Gly-Pro-3(S)Hyp)$_{10}$-NH$_2$ Do Not Form a Collagen Triple Helix.* Journal of Biological Chemistry, 2004. **279**(1): p. 282-287.

# Chapter 4

## Hydrogen Bonding and Relative Stability of Polyglycine Conformers

### 4.1 Introduction:

Polyglycine crystalizes in three forms, polyglycine I (actually a dimorph) and II. Polyglycine I includes two different crystal structures each containing a different type of β-sheet [1-5]. The structure of polyglycine II (PII) consists of parallel β-sheets oriented 120 degrees to each other [6, 7]. Three of these sheets intersect at each polyglycine strand. Each triad of nearest neighbor strands bears a striking resemblance to collagen, the most abundant protein in the human body. In collagen, the three strands of polyglycine taken from the crystal structure are replaced with strands comprised of triads of XYG (where X and Y can be any amino acid, most often proline and 4-hydroxyproline) and G is glycine. These are connected by H-bonds from the glycines (donors) on one strand to C=O's (acceptors) of the amide

couplings on another peptide strand. The "sheets" each contain only two strands in collagen-like structures. The principal differences between collagen-like structures and PII lie in the inability of collagen to form the infinite pattern of H-bonds normal to the peptide backbone (found in PII) beyond the three strands of the collagenic triple helix, due to a) the lack of N-H donors on the proline and hydroxyproline residues and b) the steric impediments caused by the side chains of all the amino acids except for glycine. Comparison of the PII and collagen structures begs the questions: a) Does the triple helical $(XYG)_N$ structure of collagen result from an evolutionary selection against structures that could form additional H-bonds such as those present in polyglycine? b) How competitive are the PII structures with β-sheets of the same size? Recent Hartree-Fock and DFT calculations have found small models of infinite strands of polyglycine [8] and antiparallel β-sheets to be essentially planar [9, 10]. Earlier MO calculations were based upon pleated sheet structures without geometry optimization [11-14].

Polyglycine is the only short peptide that can form H-bonding motifs that resemble all the H-bonding motifs studied here. As such, it is the only short peptide that can be used to compare the "natural" relative energies of all these motifs without the influence of the side chains that differentiate between the amino acids. In this chapter, I compare the several possible ways in which polyglycine strands can

aggregate. I base our calculations upon the capped peptide, formyl($gly_6$)$NH_6$. Use of the formyl (rather than acetyl) capping group prevents the methyl of the acetyl from becoming an obstacle to achieving planarity. I consider trimers and heptamers of the peptides in PII-like and antiparallel and parallel β-sheet-like conformations which I shall designate as $PII_N$, β-$AP_N$, or β-$P_N$, respectively, where N=3 or 7.

## 4.2 Results and Discussion

Table 4.1 presents the calculated interaction energies (ΔE, simple DFT energies) for the $PII_N$, β-$P_N$ and β-$AP_N$ species relative to the equivalent number of optimized extended strands of formyl($gly)_6NH_2$. The figures 4.2~ 4.4 presented below show the structures of the seven-stranded optimized species. The comparisons made all refer to the CP corrected results.

**Table 4.1:** Relative Energies of the Species Considered in Kilocalories per mole

| Species | H-bonds | ΔE/H-bond | ΔE/H-bond (CP) | Distortion /Residue | ΔE/H-bond (CP) from optimized strand | Relative total energies (CP) |
|---|---|---|---|---|---|---|
| | | | **Three Strands** | | | |
| $PII_3$ | 7 | -8.6 | -6.0 | 3.0 | 1.8 | 60.5 |
| β-$P_3$ | 14 | -7.6 | -5.9 | 2.5 | -2.7 | 9.8 |
| β-$AP_3$ | 14 | -5.8 | -4.0 | 0.4 | -3.4 | 0.0 |
| | | | **Seven Strands** | | | |
| $PII_7$ | 28 | -9.5 | -6.8 | 3.4 | -1.7 | 130.7 |
| β-$P_7$ | 42 | -8.2 | -6.4 | 2.8 | -3.5 | 29.8 |
| β-$AP_7$ | 42 | -6.5 | -4.9 | 0.6 | -4.2 | 0.0 |

### 4.2.1 Extended formyl$(gly)_6NH_2$

Extended formyl$(gly)_6NH_2$ exhibits many of the characteristics of the extended acetyl(ala)$NNH_2$ strands that have been reported in previous studies [15, 16]. In particular, slightly cooperative $C_5$ H-bonding interactions stabilize these strands. The cooperativity can be deduced from the shorter $C_5$ O···H distances found near the center of the strand (see figure 4.1). The necessary sacrifice of this stabilizing factor upon formation of β-sheets contributes to the overall low cooperativity of simple β-sheets [9, 10].

I have calculated the energies of the extended strands in fixed the geometries of the appropriate cluster of strands (sheets collagenic helix or PII model). The energy differences between the latter and the optimized strands measures the distortion

energies required for the optimized strands to achieve the requisite geometries of the aggregates. The sum of the distortion energies of the individual strands constitutes the total distortion energy for the aggregate. The fixed geometries for those calculations where the strands are kept equivalent are necessarily identical, as are their distortion energies.


**Figure 4.1**: Optimized single strand with C5O-Hdistances (angstroms) noted.

The interaction energies of the strands to form the trimeric and heptameric structures can be broken down into two parts: a) the energy required to distort the relaxed strand into the geometry it assumes as part of the larger aggregate, and b) the interaction energy between the (already distorted) fragments. The interaction energy defined by part (b) above will not depend upon the reference geometry of the strand, but the distortion energy will. Thus, the total energies of interaction for the aggregates considered will depend upon the (arbitrary) choice of the relaxed extended single strand. In reality, the unassociated strands can almost certainly exist in many different conformations which will differ in different environments. Many such conformations

may have one or more intramolecular H-bonds. These may be higher or lower in energy than the fully extended strand. The distortion energy from any optimized structure of the single strand to that taken in the optimized aggregate must always be > 0 unless the relaxed geometry is exactly the same as that of the strand in the optimized structure (in which case it will be zero). Thus, the distortion energy for each strand makes a positive contribution to the interaction energy. The difference in the distortion energy for a strand with a known structure different from the fully extended structure used here as a standard can easily be calculated. This calculated difference will give the energy of forming the aggregate from any other conformation of the strand that may be relevant including multiple conformations that might exist in solution.

### 4.2.2 β-Sheet

Both the parallel and antiparallel β-sheets form seven H-bonds between each pair of adjacent strands (six between the glys and another involving the capping formyl group), totaling14 for the(β-$AP_3$) and parallel (β-$P_3$) sheet; 42 for the seven stranded β-antiparallel (β-$AP_7$) and parallel (β-$P_7$) sheets. Antiparallel β-sheets can form quasi planar structures as expected from previous theoretical studies [9, 10].

### 4.2.3 $PII_N$

In order to keep the geometry as close to what one would expect from a seven strand fragment of an infinite crystal, I constrained the structure so that each strand has the same structure and the six outer strands of $PII_7$ are symmetrically disposed around the central strand (see figure 4.2). All attempts to achieve a completely optimized $PII_3$ structure failed as the strands moved apart. I recalculated the $AP_N$ and $P_N$ using the same constraint for comparison. The sheets structures discussed in this section refer to those optimized with this constraint. The positive $\Delta$E/H-bond of $PII_3$ (see table 4.1) indicates that this structure is repulsive compared to three separate strands.

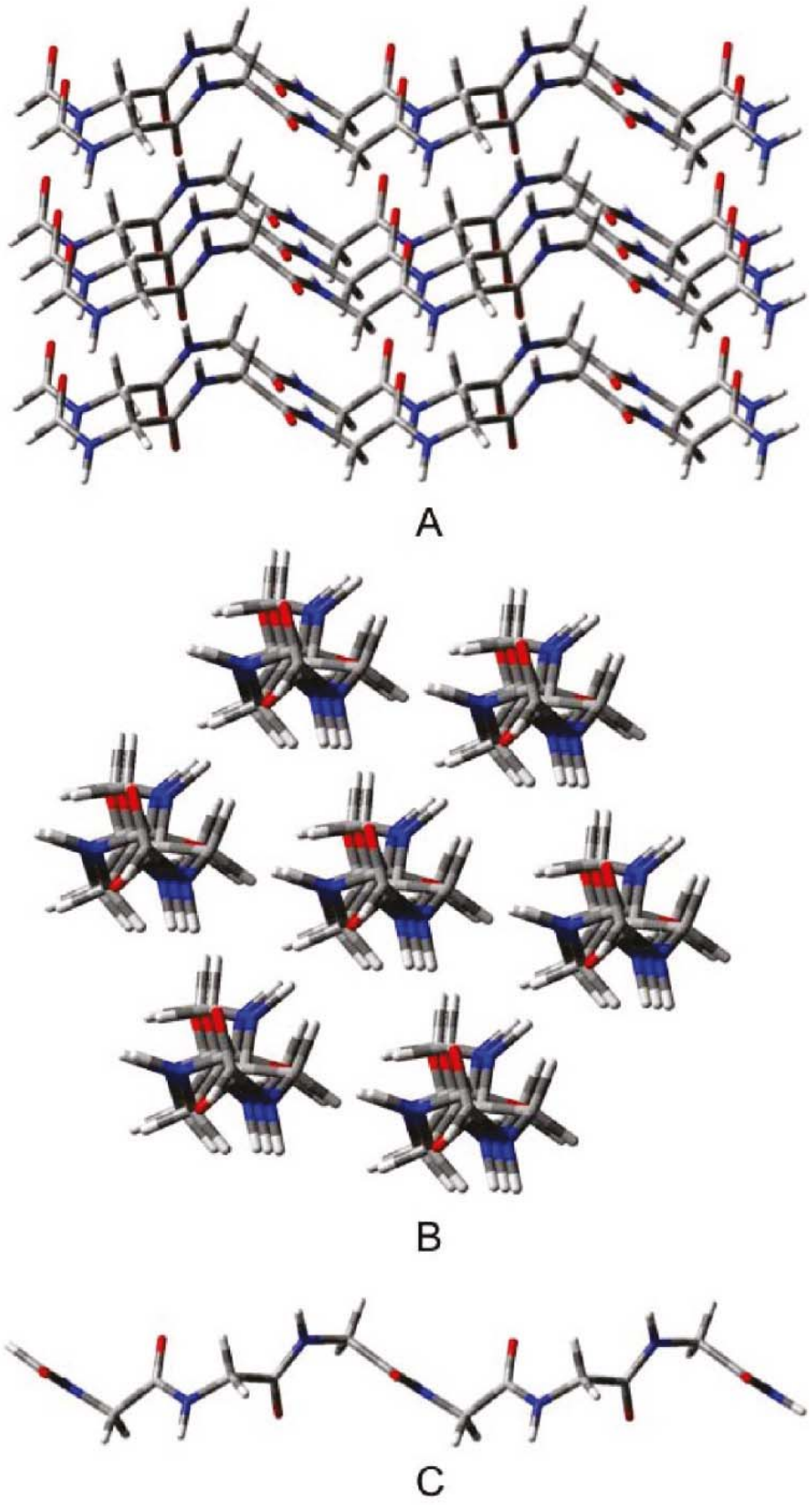



**Figure 4.2:** Seven-stranded PII-like structure, $PII_7$: (A) top or edge, (B) end, and (C) individual strand (all strands equivalent).

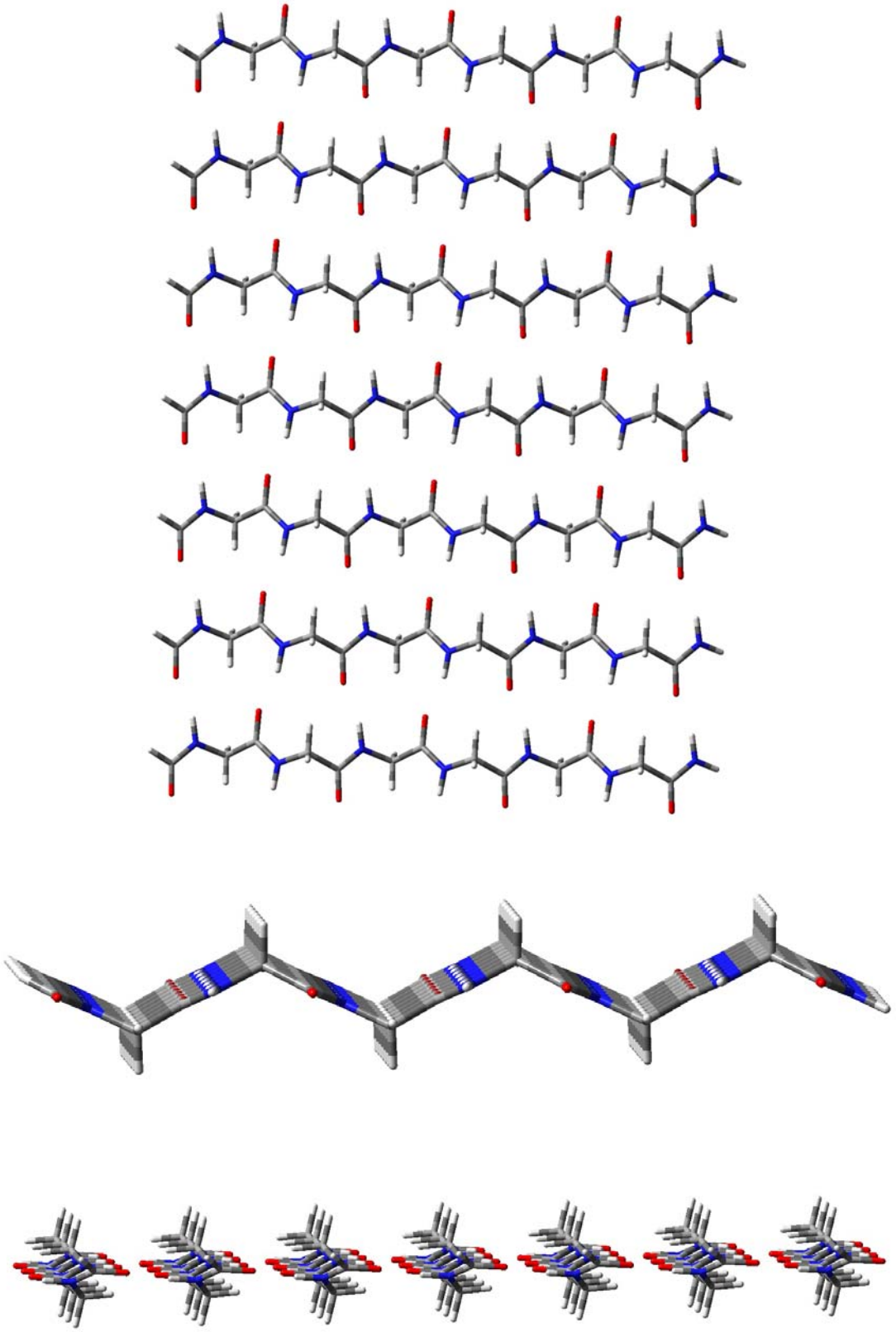


**Figure 4.3:** Seven-stranded parallel pleated sheet β-$AP_7$: (A) top view, (B) end view, (C) edge view

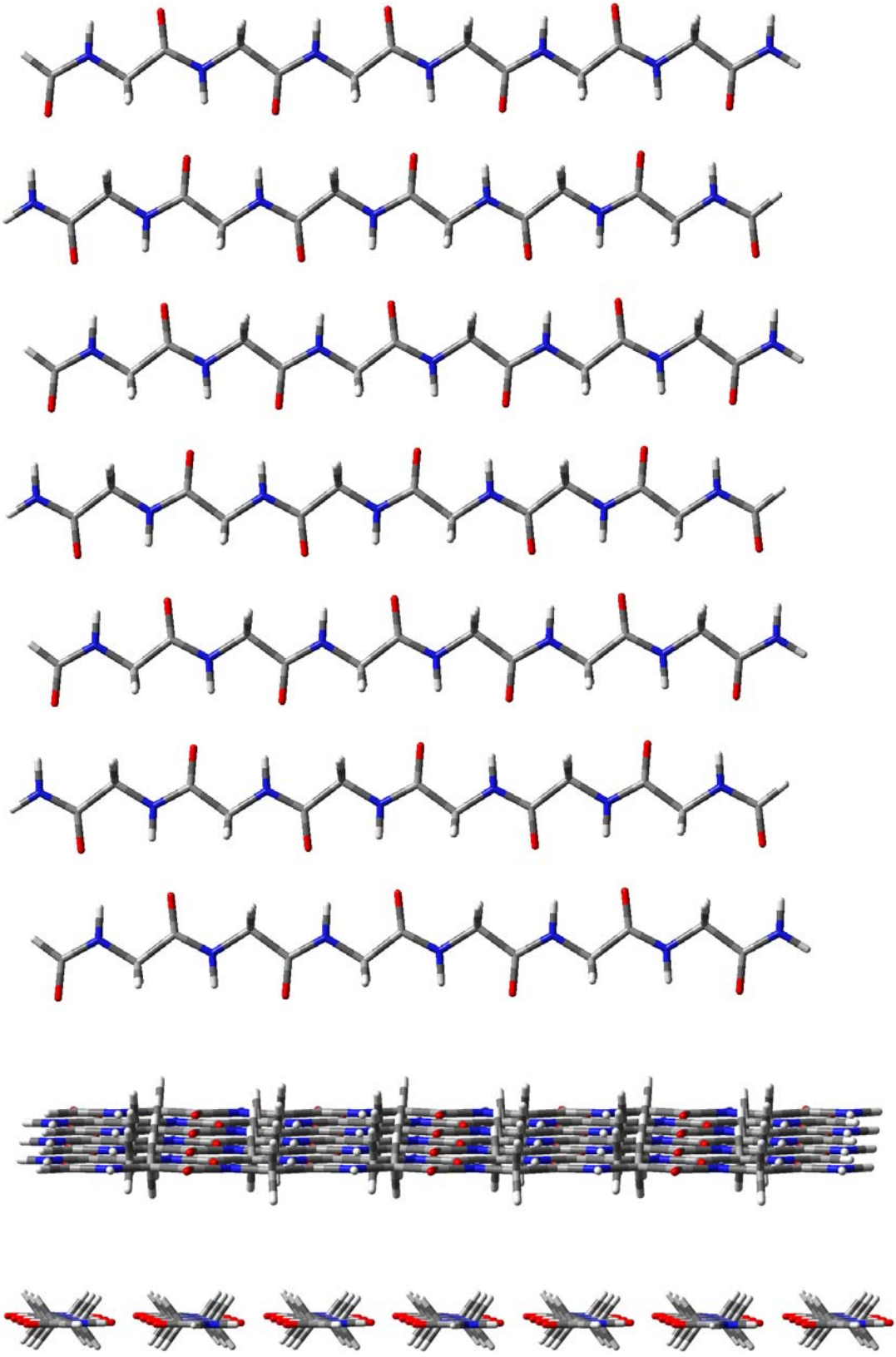


**Figure 4.4:** Seven-stranded anti parallel pleated sheet β-$AP_7$: (A) top view, (B) end view, (C) edge view

The $PII_3$ structure (roughly equivalent to a collagen triple helix), cannot H-bond cooperatively as no H-bonding chains (longer than one H-bond) exist. However the $PII_7$ structure can be viewed as three $\beta$-$P_3$ sheets passing through the same central strand at 120 degree angles to each other. If one were to expand the $PII_7$ structure to a polyglycine crystal, the three β-sheets that pass through the central strand would become quasi-infinite in size and exhibit cooperativity equal or greater to that found in $PII_7$. The average H-bond energy (after CP) for $PII_7$ is lower than that for $\beta$-$P_7$ by 0.4 kcal/mol despite the fact that each sheet within the $PII_7$ structure contains only three strands. The distortion energies favor the $\beta$-$P_7$ structure by 0.6 kcal/mol/residue, however, the $PII_7$ structure has 14 fewer H-bonds but the same number of residues as $\beta$-$P_7$. The higher relative total energy of $PII_7$ compared to $\beta$-$P_7$ (130.7 vs. 29.8, for a difference of 100.9 kcal/mol) should be corrected by the added stabilization of the additional H-bonds that would be formed in a crystal of PII. If one adds a stabilization of 14 H-bonds at -6.8kcal/mol each to compare with the $\beta$-$P_7$ structure, the difference in the relative energies would decrease by 95.2 to 5.7 kcal/mol. However, the 'sheets' in the $PII_7$ structure contain only three strands, rather than the seven in $\beta$-$P_7$. I see from table 4.1 that the $\Delta$E/H-bond becomes more negative by 0.5 kcal/mol upon going $\beta$-$P_7$ to $\beta$-$P_7$, presumably due to increased cooperativity. If one adds this cooperative stabilization for each of the 42 H-bonds (28+14), the $PII_7$ structure would

become relatively more stable by about another 21 kcal/mol, making the $PII_7$ structure more stable than either the β-P7 or even the β-AP7 structure by about 15 kcal/mol or roughly 0.4 kcal/mol/residue. While the above demonstrates the plausibility that the PII structure could be the most stable if an infinite crystal be considered, one must remember that all interactions between strands are considered in the PII structures, while the structures considered for the sheets do not take the stabilizing interaction between the sheets into account. Also, the calculations using the PII structures use equivalent strands and do not include vibrational corrections.

### 4.2.4 Trimeric structures

For the structures containing three peptide units, the antiparallel sheets ($\beta$-$AP_3$) are the most stable based both upon relative energies and enthalpies. However, this structure contains the weakest H-bonds based upon the energies of the optimized trimeric structures and the single strands frozen in the geometries they assume in the trimer. $PII_3$ contains the strongest H-bonds despite being the least stable of the three structures. In fact, $PII_3$ is the only structure that is unstable relative to three unassociated extended strands, when completely optimized. Palfi and Perczel have reported similar behavior for strands containing the AAG triad [17]. The low stability of the $PII_3$ structure must largely be due to the fewer H-bonds (seven) than in the

sheets (14). As seen from table 4.1, the distortion energy/residue of $PII_3$ is 3.0 kcal per residue (or about 63 kcal/mol for 18 residues), while the H-bonding energy is -6.0 kcal/mol per H-bond (or about -42 kcal/mol for seven H-bonds).

### 4.2.5 Heptameric structures

The relative energies of the heptameric peptides, as well as the individual contributions to these, follow the same qualitative order as those of the trimers. Nevertheless, several important differences become apparent. The CP-corrected energies per H-bond all become more negative compared to those of the trimers. At the same time, the distortion energies per residue all increase. The $\Delta$E/H-bond compared to optimized strands for the $PII_N$ structures goes from being repulsive (1.8 kcal/mol in $PII_3$) to attractive (-1.7 kcal/mol in $PII_7$) which corresponds to a relative increase in H-bond stability of 3.5 kcal/mol. The relative instability of $PII_7$ compared to its isomers appears to be due to the reduced number of H-bonds in this structure (28) as compared to the two sheets (42). The data suggest that larger isomeric structures will eventually favor the $PII_N$ over the $\beta$-$P_N$ form, which would be consistent with the fact that all three crystalline forms of polyglycine have the same number of H-bonds

### 4.2.6 Distortion Energies

As seen from table 4.1, the distortion energies per residue contribute most to the

difference in ΔE's/H-bond for the different structures (compare figure 4.1 with figure 4.2~4.4). One must bear in mind that the number of residues is *coincidentally* the same as the number of H-bonds (including those involving the formyl and $NH_2$ caps) for the sheet structures, but exceeds the number of H-bonds for the PII structures. The distortion energies per residue increase upon increasing the aggregate from three to seven strands. As noted above, the seven stranded structure has stronger H-bonds than their three stranded analogs. The stronger H-bonds allow the individual strands to accept more strain. The geometries of the frozen strands taken from the β-$P_7$ and $PII_7$ structures exhibit considerable changes in the backbone dihedral angles compared to the optimized strand in addition to the breaking of the $C_5$ H-bonding interactions. The distorted strand taken from $PII_7$ assumes the polyproline II-like backbone structure, as expected from the crystal structures [6, 18]. For the $PII_7$ structure, the relative values of distortion energy/residue and ΔE/H-bond can be somewhat misleading, as in the polyglycine crystal each strand would have seven H-bonds (rather than the four in the model) while the distortion energy per strand should not change appreciably. Assuming that the $PII_7$ structure had 49 H-bonds (instead of 28) would lower the distortion energy/H-bond (not per residue) from 5.1 to about 2.9 kcal/mol. Of course, the β-$AP_7$ and β-$P_7$ structures should also form 49 H-bonds if inserted in infinite sheets. However, the resulting reduction in the distortion energies/H-bond would be

much smaller.

### 4.2.7 Implications for collagen

Our results show the collagen-like triple helix to be unstable relative to separate strands of polyglycine largely due to the distortion energy required for each strand to assume the geometry of the collagen-like structure. As collagen contains substantial amounts of proline and hydroxyproline, the individual isolated strands should have a structure more similar to that of polyproline II, which is close to the geometries of the strands in the collagen triple helix. Comparing the distortion energies per residue for the glycine strands (3.0 kcal/mol) of the current study with those that can be derived from capped strands of the PPG (about 1.7 kcal/mol based upon 31.4 kcal/mol for three six stranded strands, or 18 residues) calculated using similar methods, confirms this hypothesis [19]. One sees that a major reason for the energetic preference of proline and hydroxyproline in the X and Y positions of the proline XYG triad must be that the strands containing the amino acids have small distortion energies for forming the triple helix. Furthermore, in a collagen triple helix, the X and Y residues (when proline or hydroxyproline) cannot form the additional H-bonds required for a structure like $PII_7$. Palfi and Perczel's comparison of strands containing AAG with those containing PPG gives results in agreement with those presented here [17]. I

suggest that these two effects combine to cause a preference for the sheets in $G_6$ and a preference for collagen in $(PPG)_2$.

### 4.3 Conclusions

Strands of polyglycines can interact via hydrogen-bonding in several different motifs. Of those studied here, the antiparallel β-sheets are most stable for aggregates containing both three and seven individual strands. The relative stability of these antiparallel β-sheets derives principally from the lower strain in the individual strands rather than from strong H-bonding interactions, as indicated in table 4.1. In fact, these sheets have the smallest H-bonding interactions. The polyglycine II-like heptamer has the strongest H-bonds, but is the least stable of the heptameric structures as it has fewer H-bonds than its isomeric structures and a relatively large amount of distortion energy/residue. When I make approximate corrections to make the number of H-bonds equivalent in number and cooperative enhancement to the isomeric structures, the relative energy of the $PII_7$ implies a stable PII crystal. The parallel b-sheets are intermediate between the other two forms, both in H-bonding energies and strain within the individual strands.

**References:**


1. Lotz, B., *Crystal structure of polyglycine I.* Comptes Rendus des Seances de l'Academie des Sciences, Serie C Sciences Chimiques, 1972. **274**(23): p. 1907-10.
2. Colonna-Cesari, F., S. Premilat, and B. Lotz, *Structure of polyglycine I: a comparison of the antiparallel pleated and antiparallel rippled sheets.* Journal of Molecular Biology, 1974. **87**(2): p. 181-91.
3. Lotz, B., *Crystal structure of polyglycine I.* Journal of Molecular Biology, 1974. **87**(2): p. 169-80.
4. Munoz-Guerra, S., et al., *Crystals of polyglycine in the beta form.* Journal of Molecular Biology, 1983. **167**(1): p. 223-5.
5. Kajava, A.V., *Dimorphism of polyglycine I: structural models for crystal modifications.* Acta crystallographica. Section D, Biological crystallography, 1999. **55**(Pt 2): p. 436-42.
6. Crick, F.H.C. and A. Rich, *Structure of polyglycine II.* Nature (London, U. K.), 1955. **176**: p. 780-1.
7. Ramachandran, G.N., C. Ramakrishnan, and C.M. Venkatachalam, *Structure of polyglycine II with direct and inverted chains.* Conformation of Biopolymers, Papers read at an International Symposium, 1967. **2**: p. 429-38.
8. Improta, R., et al., *The conformational behavior of polyglycine as predicted by a density functional model with periodic boundary conditions.* Journal of Chemical Physics, 2001. **114**(6): p. 2541-2549.
9. Zhao, Y.-L. and Y.-D. Wu, *A Theoretical Study of β-Sheet Models: Is the Formation of Hydrogen-Bond Networks Cooperative?* Journal of the American Chemical Society, 2002. **124**(8): p. 1570-1571.
10. Viswanathan, R., A. Asensio, and J.J. Dannenberg, *Cooperative Hydrogen-Bonding in Models of Antiparallel β-Sheets.* Journal of Physical Chemistry A, 2004. **108**(42): p. 9205-9212.
11. Kleier, D.A. and W.N. Lipscomb, *Molecular orbital study of polypeptides. Conformational and electronic structure of polyglycine.* International Journal of Quantum Chemistry, Quantum Biology Symposium, 1977. **4**(Proc. Int. Symp. Quantum Biol. Quantum Pharmacol., 4th): p. 73-86.
12. Fleck, O., M. Seel, and J. Ladik, *Calculation of the band structure of polyglycine as a two-dimensional periodic system.* Solid State Communications, 1988. **65**(7): p. 701-4.
13. Ladik, J., A. Sutjianto, and P. Otto, *Improved band structures of some homopolypeptides with aliphatic side chains and of the four nucleotide base stacks: estimation of their fundamental gap.* Journal of Molecular Structure

THEOCHEM, 1991. **74**: p. 271-6.

14. Suhai, S., *The electronic structure of parallel β-pleated sheets in proteins: an ab initio computation including electron correlation.* International Journal of Quantum Chemistry, 1991. **40**(4): p. 559-76.
15. Wieczorek, R. and J.J. Dannenberg, *Comparison of Fully Optimized α- and $3_{10}$-Helices with Extended β-Strands. An ONIOM Density Functional Theory Study.* Journal of the American Chemical Society, 2004. **126**(43): p. 14198-14205.
16. Horvath, V., Z. Varga, and A. Kovacs, *Long-Range Effects in Oligopeptides. A Theoretical Study of the β-Sheet Structure of Glyn (n = 2-10).* Journal of Physical Chemistry A, 2004. **108**(33): p. 6869-6873.
17. Palfi, V.K. and A. Perczel, *How stable is a collagen triple helix? An ab initio study on various collagen and β-sheet forming sequences.* Journal of Computational Chemistry, 2008. **29**(9): p. 1374-1386.
18. Rich, A. and F.H. Crick, *The molecular structure of collagen.* Journal of Molecular Biology, 1961. **3**: p. 483-506.
19. Tsai, M., Y. Xu, and J.J. Dannenberg, *Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProPro$^{D}$Ala, and ProPro$^{D}$Ser Triads.* Journal of the American Chemical Society, 2005. **127**(41): p. 14130-14131.

# Chapter 5

## DFT Energy Surfaces of the Ramachandran Polts

### 5.1 Indruction:

Interest in the potential energy landscapes of peptide oligomers has intensified in the last few years as energetic comparisons of folded and unfolded peptides have been extensively studied. For example, comparisons of the stabilities of α-helical peptides with their unfolded counterparts have been studied both experimentally [1-3] and theoretically [4-6]. While the structure of the α-helix is well defined both in the gas phase and in solution, that of the unfolded state is not in either phase.

In other studies, mutations of one amino acid residue in a peptide to its enantiomer (to form a *diastereomer* of the original peptide) have been suggested to stabilize a folded state. Similarly, mutation of the enantiomorphic residue, Gly, to $^{D}$Ala sometimes leads to more stable folding [7, 8]. Since the folded structures are much better understood than the unfolded ones, the interpretation of the effects of the structural differences between the folded and unfolded structures have concentrated on those of the former.

Historically, the unfolded peptides were thought to be ‘random coils’ which have unrestricted rotation about two of the three the kinds of dihedral angles in the peptide backbone. More recently, experimental [9-11], dynamics using empirical force fields[12, 13], and theoretical [6] evidence has suggested that the polyproline II (PII) structure may predominate for small polyalanines in aqueous solution. Nevertheless, Scheraga, et al. [14] have maintained that the conformational composition of peptides in aqueous solution might be more complex.

In the 1960's, Ramachandran published his classic studies of the energy landscapes of peptides as a function of the two generally mobile dihedral angles in the peptide backbone [15, 16]. Since no useful theoretical methods were available at the time of his study, he used molecular models to define those dihedral angles for which unfavorable steric interactions make the conformations unstable. He was not able to differentiate between those conformations that had no steric impediments, nor did he take the effects of weak ($C_5$) or other stronger *intra*molecular hydrogen bonds into account. Since that time, several theoretical studies of the effects of these dihedral angles upon specific sequences have been published [17-19]. However, to the best of our knowledge, there have been no published molecular orbital or density functional theory studies including the effects of solvation nor the effects of mutation of one residue to its enantiomer upon the energy landscapes defined by these two dihedral angles in di- or tripeptides. However, one reported study on capped glycine and capped alanine [20] and one Hartree-Fock study on capped alanine have appeared [21]. Some such studies using empirical force fields and molecular dynamics have appeared[22].

In this chapter show that the energy landscapes of two diasteriomeric capped trialanine peptides, acetyl-$^{L}$Ala-$^{L}$Ala-$^{L}$Ala-$NH_2$ (**3AL**) and acetyl-$^{L}$Ala-$^{D}$Ala-$^{L}$Ala-

$NH_2$ (**3AD**) as a function of these dihedral angles both in the gas phase and in aqueous solution. I also explicitly consider each of the local minima that contain one or more $C_7$, $C_8$, $C_{10}$, $C_{11}$, or $C_{13}$ *intra*molecular H-bonds in both the gas phase and aqueous solution. Experimental observation of related (but not identical) tripeptide structures containing such interactions have been reported both in the gas phase [23-25] and in solution [26].

## 5.2 Results and Disscussion

I first present the energy landscapes for the conformations of the two diastereomeric tripeptides that do not contain intramolecular H-bonds (other than $C_5$) in the gas phase, in aqueous solution and that of the solvation free energy alone. The schematic structures of $C_N$ (where *N* represents the number of atoms in the smallest cyclic H-bonding structure) H-bonds are illustrated in Figure 5.1. I then present data for the completely optimized structures, including those with intramolecular H-bonds, in the gas phase and in aqueous solution. Finally, I compare and discuss the data for the two diastereomers.

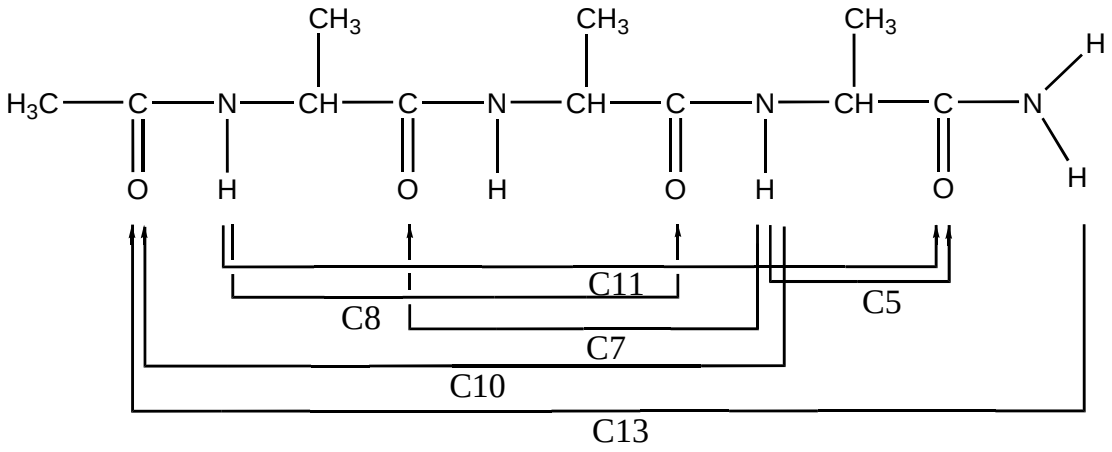



**Figure 5.1**: Schematic of $C_N$

### 5.2.1 Acetyl-$^{L}$Ala-$^{L}$Ala-$^{L}$Ala-$NH_2$ (3AL)

The energy landscape for **3AL** in the gas phase is depicted in Figure5.2. The energies are relative to that of the global minimum for structures without intramolecular H-bond(s) other than $C_5$ (with the exception noted below). This global minimum, found near φ = 170°, ϕ = -160° represents an extended β-strand. This structure is stabilized by $C_5$ intramolecular H-bonds, as has been previously reported [4, 27, 28]. As this region of the dihedral map is rather flat, the energy increases only gradually as the dihedrals change for values close to the minimum. A shallow trough about a local minimum centered around φ = 0°, ϕ = -100° includes the α-helical conformation near φ = -50°, ϕ = -60°.

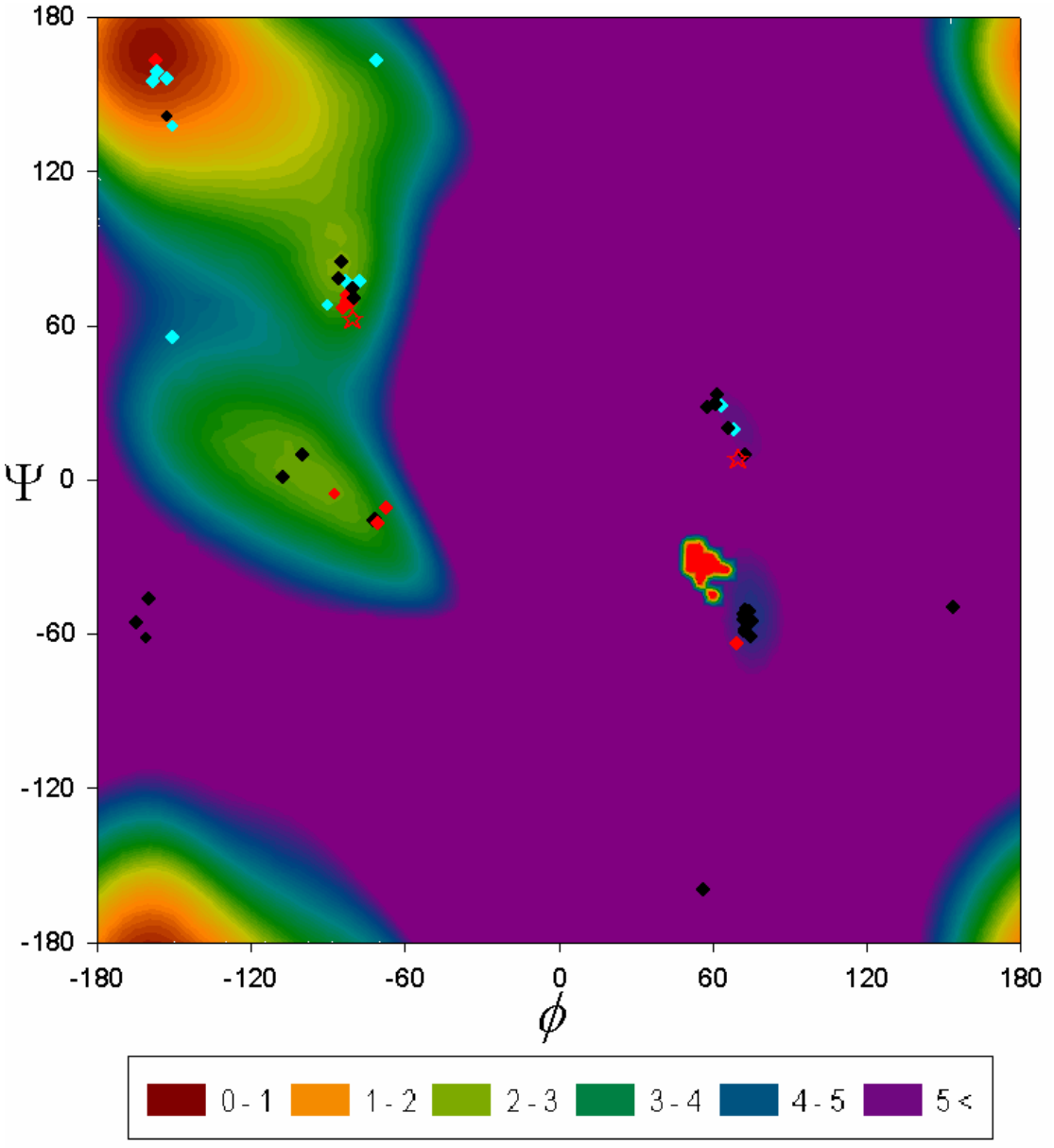



**Figure 5.2:** Ramachandran plot for **3AL** in the gas phase. H-bonding structures are indicated by colored points. Red points represent H-bonding structures that have lower energies than the global minimum for non-H-bonding structures; black points have lower energies than the optimized non-H-bonding structures with the same dihedral angles but higher than the global minimum for non-H-bonding structures; light blue points have higher energies than the non-H-bonding structures with the same dihedrals; open stars correspond to gas phase H-bonding structures reported by Mons for a related (but different) structure. Energies are relative to the "global" minimum (see text).

Another, much smaller, region of low energy appears near φ = -30°, ϕ = 50°. All the structures in this region contain $C_7$ H-bonds, as I was not able to find partially optimized structures in the region of dihedral space without these interactions. These structures are typical of β-turns. Unlike the surface near the first minimum, the energy changes quite abruptly with dihedral angle as the H-bond is either formed or not. This behavior occurs as the purely conformational energy (without the stabilization due to the H-bonds) is significantly higher in this region (compared to the β-sheet), while the $C_7$ H-bond is stronger than the $C_5$. The energy difference between the β-strand global minimum and the α-helical region are consistent with the estimated distortion energies of the β-strand to form and α-helix that in our lab have previously reported [4].

Several conformations of **3AL** can be formed that are stabilized by internal H-bonds (other than $C_5$). Among the cyclic H-bonds that are possible, I have identified structures containing $C_7$, $C_8$, $C_{10}$, and $C_{11}$ cyclic H-bonds. I have optimized the structures for each of these (see table 5.1) and indicated each in the landscape of Figure 7.3 as color-coded points at the appropriate φ and ϕ coordinates. Mons [23-25] has observed two H-bonding structures (among several that were considered) for similar capped three residue peptides. These are included in the figure. They both have lower energies than the global non-H-bonding minimum, in agreement with his report. These internally H-bonded structures, as well as those for the diastereomeric **3AD**, are discussed in more detail below.

**Table 5.1:** Data for All Optimized Structures of 3AL [Energies in kcal/mol, Dipole Moments in Debyes; $\Delta E$ and $\Delta E + \Delta G$ Are Relative to 1L (the Global Minimum)]

| 3AL | Internal H-Bonding | | | | | ψ | ϕ | dipole | E | $\Delta G_{solv}$ | $\Delta E + \Delta G_{solv}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $C_7$ | $C_8$ | $C_{10}$ | $C_{11}$ | $C_{13}$ | | | | | | |
| **1L** | | | | | | 167 | -159 | 3.0 | 0.00 | -36.20 | 0.00 |
| **2L** | 1 | | | | | 164 | -158 | 2.8 | -0.16 | -34.89 | 1.15 |
| **3L** | 1 | | | | | 161 | -158 | 3.3 | 2.38 | -35.97 | 2.60 |
| **4L** | 1 | | | | | 85 | -85 | 4.4 | 1.94 | -36.99 | 1.14 |
| **5L** | 1 | | | | | -52 | 75 | 2.7 | 4.42 | -37.36 | 3.25 |
| **6L** | 1 | | | | | 157 | -159 | 2.6 | 1.64 | -37.42 | 0.42 |
| **7L** | 1 | | | | | 158 | -153 | 4.4 | 4.73 | -38.33 | 2.60 |
| **8L** | 2 | | | | | 74 | -83 | 2.4 | -0.78 | -34.08 | 1.34 |
| **9L** | 2 | | | | | -62 | 73 | 4.2 | 1.15 | -34.84 | 2.51 |
| **10L** | 2 | | | | | 76 | -81 | 3.8 | 1.19 | -34.24 | 3.16 |
| **11L** | 2 | | | | | -56 | 74 | 3.9 | 3.38 | -35.26 | 4.32 |
| **12L** | 2 | | | | | 142 | -152 | 4.1 | 0.48 | -35.04 | 1.63 |
| **13L** | 2 | | | | | 139 | -150 | 5.3 | 2.77 | -35.96 | 3.01 |
| **14L** | 2 | | | | | 78 | -87 | 2.0 | 0.93 | -36.80 | 0.32 |
| **15L** | 2 | | | | | 78 | -83 | 3.8 | 3.21 | -37.32 | 2.08 |
| **16L** | 3 | | | | | 69 | -84 | 7.2 | -2.84 | -32.94 | 0.42 |
| **17L** | 3 | | | | | 71 | -83 | 6.9 | -1.01 | -33.09 | 2.10 |
| **18L** | 3 | | | | | -62 | 70 | 1.0 | -2.42 | -28.86 | 4.92 |
| **19L** | 3 | | | | | 70 | -81 | 6.7 | -0.74 | -33.52 | 1.94 |
| **20L** | 3 | | | | | 70 | -81 | 4.2 | 0.70 | -32.58 | 4.32 |
| **21L** | 3 | | | | | -59 | 72 | 7.4 | 0.69 | -33.69 | 3.20 |
| **22L** | 3 | | | | | -55 | 70 | 6.6 | 0.89 | -33.35 | 3.74 |
| **23L** | 3 | | | | | -54 | 73 | 9.0 | 2.99 | -34.07 | 5.12 |
| **24L** | 1 | | | | | 3 | -106 | 2.8 | 0.45 | -36.11 | 0.54 |
| **25L** | 1 | | | | | 19 | 68 | 3.6 | 4.88 | -36.79 | 4.28 |
| **26L** | 1 | | | | | 20 | 67 | 5.5 | 1.07 | -33.87 | 3.39 |
| **27L** | 1 | | | 1 | | 77 | -78 | 6.3 | 2.24 | -34.49 | 3.95 |
| **28L** | 1 | | | | 1 | 68 | -90 | 9.9 | 4.64 | -35.53 | 5.31 |
| **29L** | 1 | | | | 1 | -51 | 73 | 9.2 | 1.73 | -34.93 | 2.99 |
| **30L** | 1 | | | | 1 | -55 | -165 | 8.5 | 2.93 | -33.22 | 5.90 |
| **31L** | 1 | | | | 1 | -47 | -161 | 7.4 | 4.26 | -33.18 | 7.28 |
| **32L** | 1 | | 1 | | | -5 | -88 | 9.4 | -0.42 | -36.48 | -0.70 |
| **33L** | 1 | | 1 | | | 9 | -99 | 8.0 | 0.87 | -34.57 | 2.49 |
| **34L** | 1 | | 1 | | | 9 | 73 | 9.4 | 1.21 | -35.45 | 1.95 |
| **35L** | 1 | | 1 | | | 33 | 61 | 8.0 | 0.73 | -32.54 | 4.38 |
| **36L** | 1 | | 1 | | | 29 | 60 | 9.1 | 3.18 | -35.06 | 4.32 |
| **37L** | 1 | | 1 | | | 31 | 62 | 10.7 | 4.90 | -35.50 | 5.60 |

**Table 5.1 Continue**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **38L** | 1 | 1 | -18 | -72 | 9.4 | -1.39 | -33.54 | 1.27 |
| **39L** | 1 | 1 | -15 | -71 | 9.1 | 1.08 | -33.93 | 3.35 |
| **40L** | | 1 | 29 | 63 | 5.6 | 5.67 | -38.06 | 3.80 |
| **41L** | | 1 | -16 | -72 | 4.6 | 1.75 | -36.50 | 1.45 |
| **42L** | | 2 | -11 | -67 | 12.5 | -0.79 | -36.76 | -1.35 |
| **43L** | | 2 | 28 | 58 | 9.8 | 2.20 | -36.25 | 2.15 |
| **44L** | 1 | | -62 | -161 | 6.4 | 4.35 | -36.05 | 4.49 |
| **45L** | 1 | | 56 | -151 | 6.0 | 5.28 | -35.27 | 6.20 |
| **46L** | 1 | | -159 | 58 | 7.1 | 5.43 | -35.58 | 6.05 |
| **47L** | 1 | | 163 | -71 | 8.3 | 4.48 | -36.68 | 4.00 |
| **48L** | 1 | 1 | -49 | 152 | 5.5 | 4.04 | -34.94 | 5.31 |

The landscape for the solvation free energy is depicted in Figures 5.3. Solvation best stabilizes structures that differ from the gas phase minima of Figure 5.2. Water solvates the β-strand (gas phase) global minimum roughly 4 kcal/mol less than the best solvated structures. One expects this behavior as the internal $C_5$ H-bonds which stabilize the gas phase structures should weaken or break in aqueous solution to better accommodate the H-bonds with water that form. The regions of the energy landscape most stabilized by aqueous solvation include structures close to the regions of both right- and left-handed α-helices, However, they are not minima on the gas phase surface, nor do they possess the internal H-bonds. While the two dihedral angles used to construct the figures are typical of α-helices, the other dihedral angles are not. These structures have all four C=O bonds pointing in parallel directions, thus increasing the magnitudes of the dipole moments of the structures. Increased stabilization would come form the stabilizing effects of a high dielectric medium upon the dipole moments. This point will be further discussed below. I note from Figure 5.3 that water solvates the reverse α-helix (which should also have a large dipole moment), but not sufficiently to overcome the high energy for this structure in the gas phase.

Figure 5.4 depicts the landscape of the solvated peptide obtained by adding the corresponding points of the previous two figures. From this figure (a combination of gas phase *potential energy* and solvation *free energy*, as noted above), one can identify three general regions of dihedral space that have large, relatively flat surfaces and a fourth with a very localized minimum corresponding to a specific geometry containing a $C_7$ H-bond as in a β-turn. The three larger minima correspond to a β-sheet (similar to that found in the gas phase), a **PII** structure near φ = 140° , ϕ = -50° (similar to that found in earlier studies)[6, 9, 10, 12, 13], and a structure similar to an α-helix at φ = -40° , ϕ = -50°, in agreement with an earlier report by Head-Gordon[29], The β-strand structure is relatively destabilized by solvation in comparison to the **PII** and α-helix-like structure, as noted above.

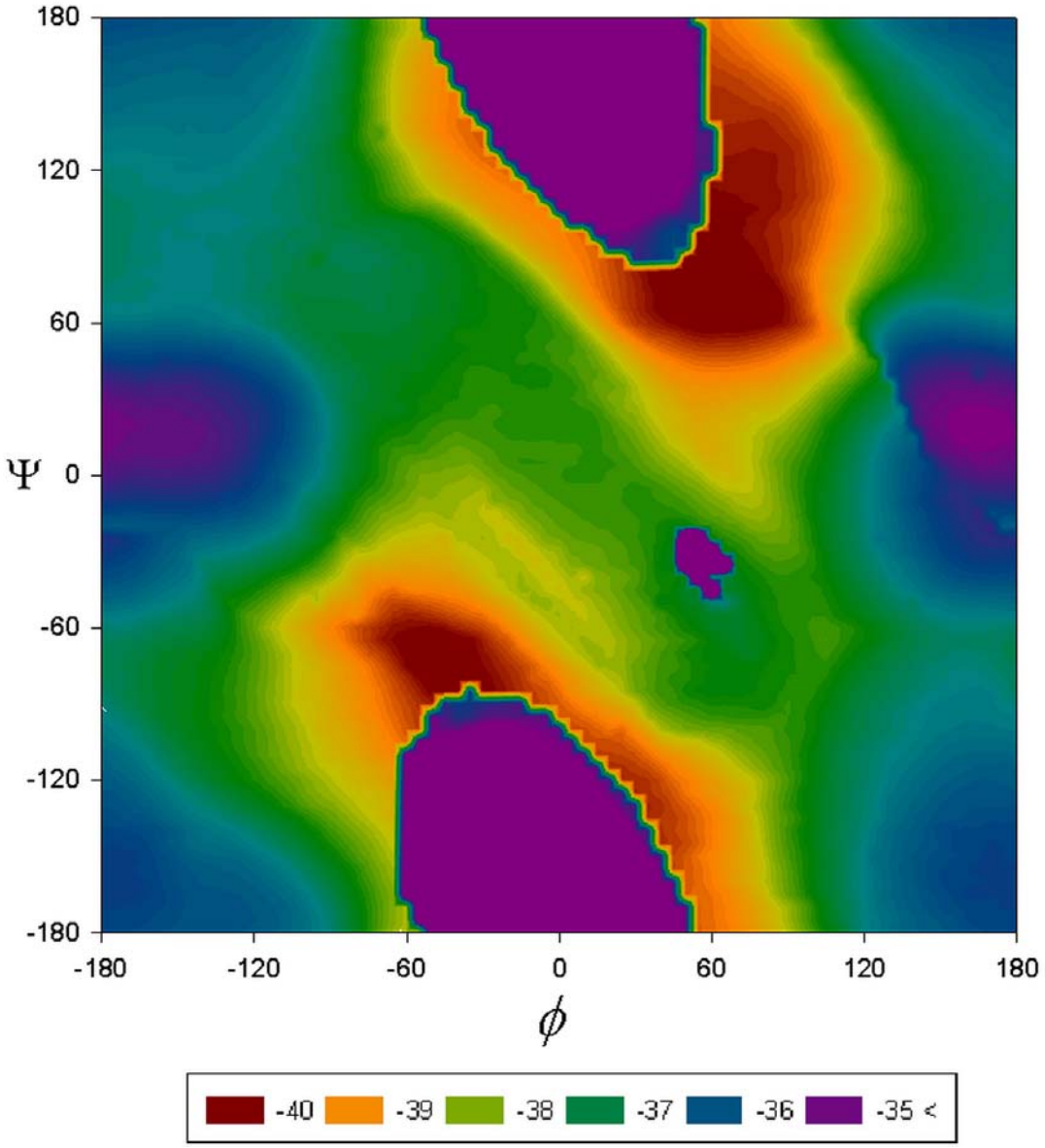



**Figure 5.3:** Plot of the aqueous solvation energies of **3AL** as a function of the Ramachandran dihedral angles. Only those H-bonding structures for which no non-H-bonding structure (with the same dihedrals) could be found (β-turn region) are included.

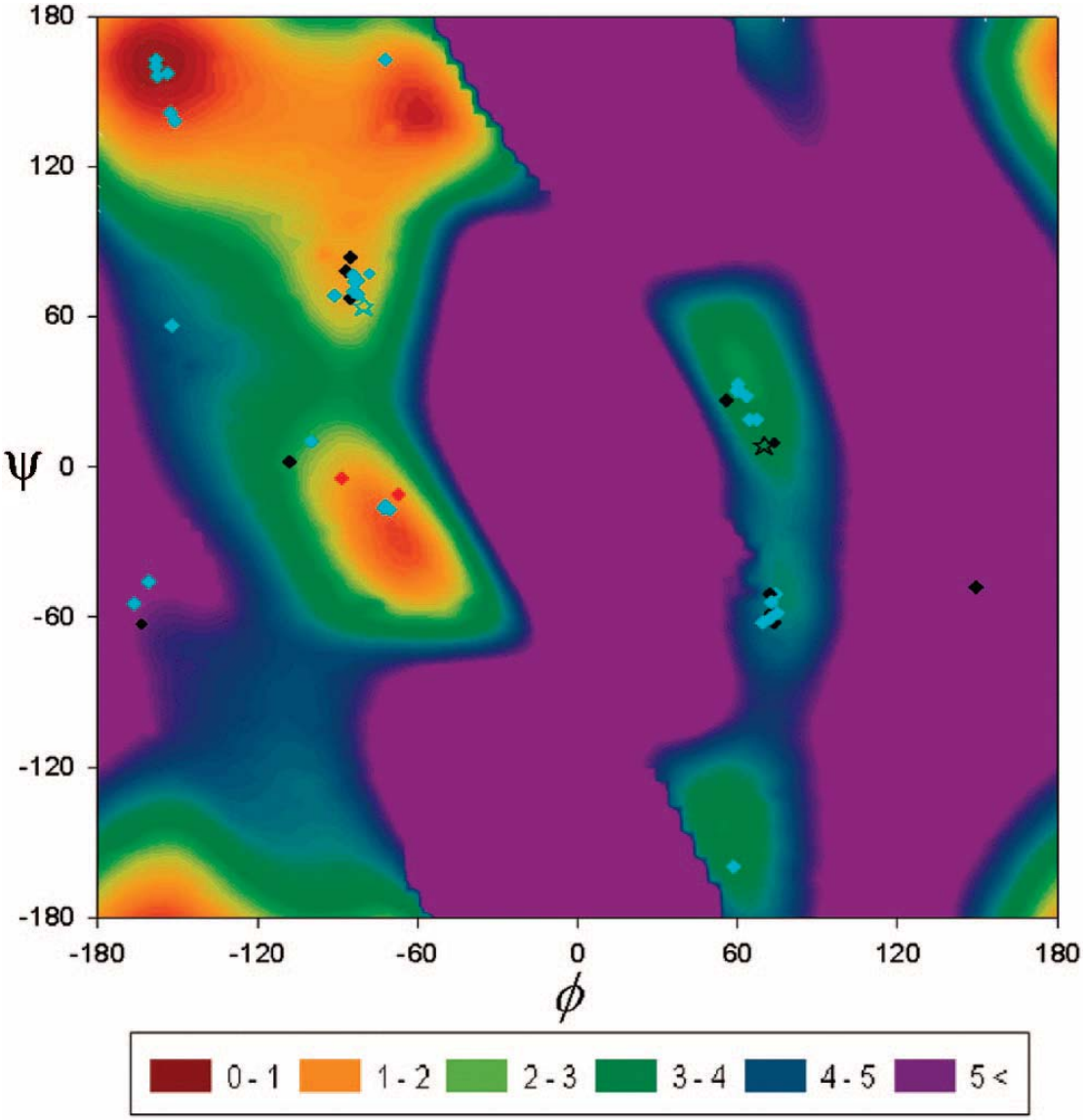



**Figure 5.4:** Ramachandran plot for **3AL** including aqueous solvation using the SM5.2 method as described in the text. H-bonding structures are indicated by colored points. Red points represent H-bonding structures that have lower energies than the global minimum for non-H-bonding structures; structures indicated by black points have lower energies than the optimized non-H-bonding structures with the same dihedral angles but higher than the global minimum for non-H-bonding structures; light blue points represent structures with higher energies than the non-H-bonding structures with the same dihedrals; open stars correspond to gas phase H-bonding structures reported by Mons for a related (but different) structure. Energies are relative to the "global" minimum.

The two red diamonds in Figure 5.4 indicate the two lowest energies on the solvated energy surface. They correspond to structures with internal H-bonds. The more stable structure ($\varphi$ = -10.8° , $\phi$ = -67.3°) resembles a small $3_{10}$ -helix containing two $C_{10}$ H-bonds and is 1.3 kcal/mol below the 'global' minimum. Wieczorek previous report had indicated this structure to be a local gas phase minimum [4]. The solvation landscape of Figure 5.3, indicates this region to be a bit better solvated than the global minimum (β-strand). As noted above, I attribute this increased stabilization to the higher dipole moment of the $3_{10}$-helical structure (12.5 D) over that of the β-strand (3.0 D). The second of these two structures ($\varphi$ = -5.1°, $\phi$ = -88.0°) lies 0.7 kcal/mol below the 'global' minimum and contains one $C_{10}$ and one $C_7$ H-bond and has a dipole moment of 9.4 D.

Aside from these two, all the other structures containing internal H-bonds become less stable than the β-strand global minimum after solvation.

The relative energies, enthalpies and free energies of theses two structures, the b-turn-like structures and others containing internal cyclic H-bonding motifs require further discussion. Clearly, the enthalpies and free energies of these structures relative to the non-H-bonding structures will be affected by the stronger vibrational force constants that will restrict conformational mobility of these H-bonding structures compared to the extended ones. Also, the stabilization energies of H-bonds are generally overestimated by *ab initio* molecular orbital calculations with less than complete basis sets because of the basis set superposition error (BSSE) [30]. While the effect of this error upon both the energy [31, 32] and the geometry [33] can be corrected using the counterpoise correction (CP), this is not easily accomplished for internal H-bonds. The CP correction for BSSE affects the relative energy, enthalpy

and free energies, while the vibrational corrections will affect only the relative enthalpies and free energies.

In order to obtain estimates of these errors, I have A) performed vibrational frequency calculations on completely optimized structures representative of the extended β-strands and the β-turns and other H-bonding structures which represent local minima on the potential energy surface (vibrations can only be properly calculated at stationary points on the surface); and B) calculated the CP error for single point calculations of H-bonds between two formamide molecules derived from the structures with the corresponding H-bond. In these structures all bond lengths and angles in the formamide dimers are constrained to the optimized structure, but the relevant C-C and N-C bonds are replaced with C-H and N-H bonds. The relative energies of the two low energy H-bonding structures discussed above (compared to the extended β-strand global minimum) need to be increased by 1.5 and 4.4 kcal/mol, respectively, due to the CP correction, while the relative enthalpies and free energies of all the H-bonded structures need to be increased by a constant 0.35 and 0.28 kcal/mol, respectively, due to the vibrational corrections. After these corrections, the two structures discussed above no longer fall below the global minimum. However, the first of the two ($3_{10}$-helix) remains only about 0.1 kcal/mol higher. Consequently, it should be counted as one of the structures in the solvated equilibrium mixture.

Thus, there are at least four kinds of structures that should contribute to the solvated equilibrium: 1) the gas-phase global minim (extended β-strand), 2) the **PII** structure identified by Kallenbach [9, 10] as one of the major components of the equilibrium; 3) a structure that has dihedrals similar to those of an α-helix (although no H-bonds); and 4) a $3_{10}$-helix containing two $C_{10}$ H-bonds. The other H-bonding structures in the β-turn region and elsewhere are all of higher energy than the β-strand

when solvated even before the CP and vibrational corrections. In general, the reports of multiple conformations for small solublized poly-alanines are consistent with our results [14, 34, 35].

These results strongly suggest that simply considering the interaction energies between simple amides or extended dipeptides using explicit waters [36, 37] or electrostatic models that mimic the explicit waters [38, 39] will be insufficient to understand the conformation preferences of peptides containing as few as three residues, because these models neglect the stabilization energy derived from the interaction of the peptide dipole moment and the dielectric properties of bulk water.

### 5.2.2 Acetyl-$^{L}$Ala-$^{D}$Ala-$^{L}$Ala-$NH_2$ (3AD)

Figure 5.5 depicts the gas phase landscape of relative energy for this structure. The colored points indicate structures with internal H-bonds, whose details are collected in table 5.2. Figure 5.5 bears certain qualitative resemblance to that for the mirror image of its all-L diastereomer, **3AL**, as might be expected since the dihedral angles considered are those nearest the inverted chiral center. Nevertheless, there are important qualitative and quantitative differences (see discussion of differences below). The most striking is that the region near $\varphi$ = -30°, $\phi$ = 50.0°, which corresponds to the β-turn-like region in the **3AL** (*not its mirror image*), is both larger (in dihedral space) and much more stable relative to the other minima than in the diastereomer. Since the $C_7$ H-bond in the β-turn involves the $^{L}$Ala's on either side of the central $^{D}$Ala, one should expect this feature of the landscape to be similar to the diastereomer, not its mirror image, as is the case. The inversion of the chiral center on the central Ala from L to D relieves some steric strain, as might be expected from the fact that the β-turn comes in a region of otherwise low energy in **3AD**, but high

energy in **3AL**. Consequently, and more importantly, most of the β-turn structures for **3AD** contain two $C_7$ H-bonds in contrast to only one in the **3AL** structures. The existence of such structures has previously been suggested on the basis of earlier Hartree-Fock calculations on dipeptides[46] although multiple $C_7$ H-bonds would not be possible in dipeptides.

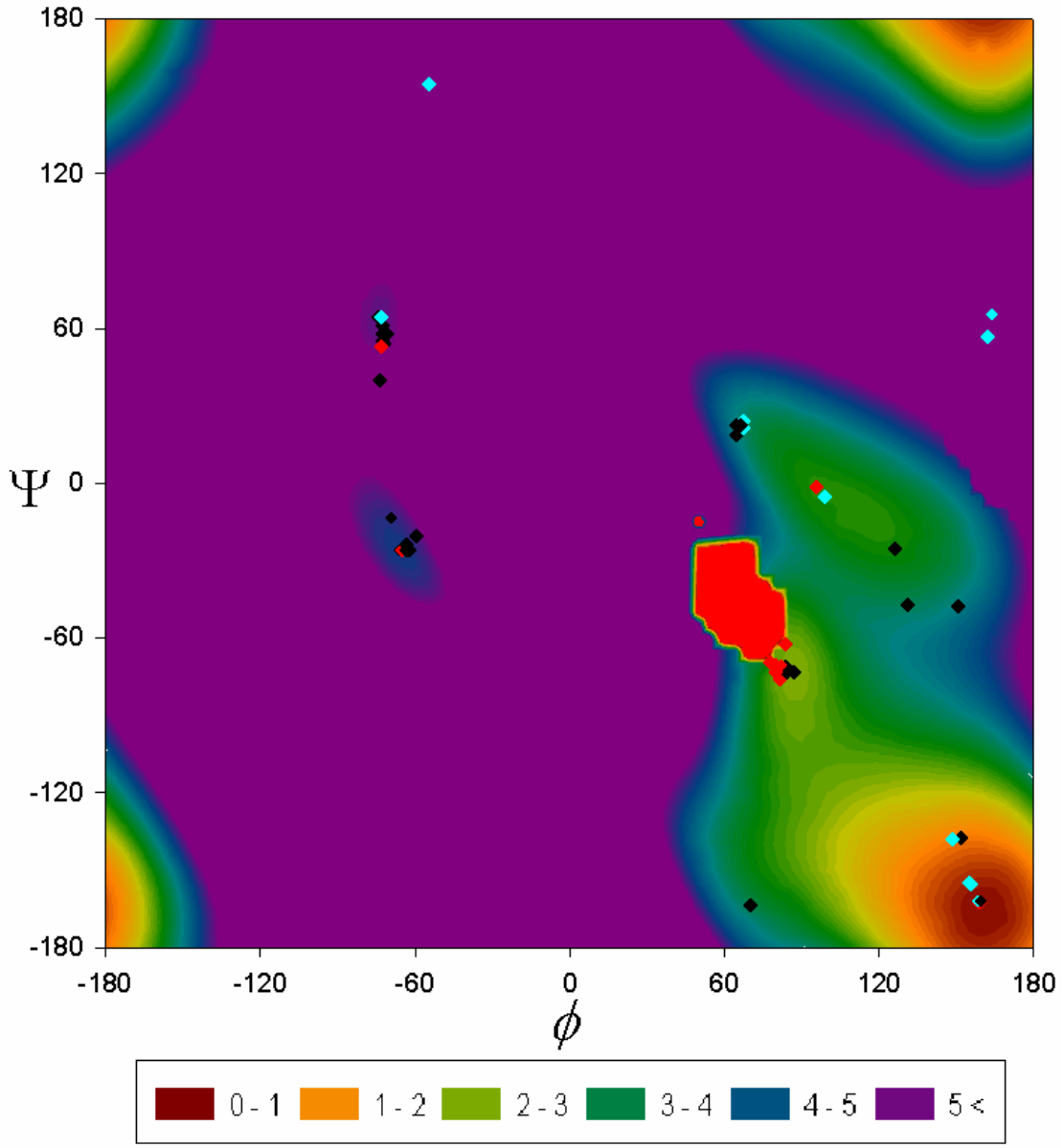



**Figure 5.5:** Ramachandran plot for **3AD** in the gas phase. H-bonding structures are indicated by colored points. Red points represent H-bonding structures that have lower energies than the global minimum for non-H-bonding structures; black points indicate structures that have lower energies than the optimized non-H-bonding structures with the same dihedral angles but higher than the global minimum for non-Hbonding structures; light blue points represent structures that have higher energies than non-H-bonding structures with the same dihedrals. Energies are relative to the "global" minimum.

**TABLE 5.2:** Data for All Optimized Structures of 3AD [Energies in kcal/mol, Dipole Moments in Debyes; $\Delta E$ and $\Delta E + \Delta G_{solv}$ Are Relative to 1D (the Global Minimum)]

| 3AD | Internal H-Bonding | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $C_7$ | $C_8$ | $C_{10}$ | $C_{11}$ | $C_{13}$ | ψ | ϕ | dipole | $\Delta E$ | $\Delta G_{solv}$ | $\Delta E + \Delta G_{solv}$ |
| 1D | | | | | | -167 | 161 | 7.2 | 0.00 | -36.41 | 0.00 |
| 2D | 1 | | | | | -163 | 159 | 3.9 | -0.20 | -35.33 | 0.89 |
| 3D | 1 | | | | | -161 | 160 | 2.4 | 2.53 | -35.65 | 3.29 |
| 4D | 1 | | | | | -74 | 88 | 3.0 | 1.92 | -36.78 | 1.55 |
| 5D | 1 | | | | | 64 | -74 | 3.7 | 4.81 | -37.65 | 3.57 |
| 6D | 1 | | | | | -161 | 160 | 3.8 | 1.97 | -37.48 | 0.90 |
| 7D | 1 | | | | | -154 | 156 | 2.5 | 4.60 | -38.36 | 2.66 |
| 8D | 2 | | | | | -76 | 82 | 4.1 | -1.10 | -34.21 | 1.10 |
| 9D | 2 | | | | | 55 | -72 | 2.6 | 1.41 | -34.86 | 2.96 |
| 10D | 2 | | | | | -73 | 84 | 3.9 | 1.31 | -34.54 | 3.17 |
| 11D | 2 | | | | | 61 | -72 | 3.7 | 3.51 | -34.86 | 5.06 |
| 12D | 2 | | | | | -137 | 152 | 2.4 | 0.42 | -34.62 | 2.21 |
| 13D | 2 | | | | | -138 | 149 | 5.0 | 3.04 | -35.97 | 3.48 |
| 14D | 2 | | | | | -156 | 155 | 4.5 | 3.46 | -35.75 | 4.13 |
| 15D | 2 | | | | | 63 | -73 | 3.5 | 5.47 | -37.46 | 4.42 |
| 16D | 2 | | | | | 62 | -76 | 2.6 | 3.23 | -36.89 | 2.76 |
| 17D | 3 | | | | | -70 | 78 | 0.8 | -5.06 | -28.52 | 2.83 |
| 18D | 3 | | | | | -72 | 80 | 6.1 | -0.85 | -33.14 | 2.42 |
| 19D | 3 | | | | | 52 | -74 | 7.5 | -1.17 | -33.22 | 2.02 |
| 20D | 3 | | | | | -71 | 83 | 6.6 | -1.24 | -33.45 | 1.72 |
| 21D | 3 | | | | | 58 | -73 | 7.6 | 0.85 | -33.31 | 3.96 |
| 22D | 3 | | | | | 55 | -71 | 7.3 | 0.92 | -33.78 | 3.55 |
| 23D | 3 | | | | | -71 | 83 | 8.7 | 1.35 | -33.85 | 3.91 |
| 24D | 3 | | | | | 58 | -71 | 5.2 | 2.61 | -33.08 | 5.94 |
| 25D | 1 | | | | | -26 | -67 | 3.1 | 2.48 | -36.59 | 2.31 |
| 26D | 1 | | | | | -7 | 99 | 3.9 | 2.80 | -36.35 | 2.86 |
| 27D | 1 | | | | | -2 | 95 | 5.6 | -1.20 | -33.71 | 1.50 |
| 28D | 1 | | | 1 | | -75 | 79 | 5.8 | -2.53 | -31.88 | 2.00 |
| 29D | 1 | | | | 1 | -62 | 84 | 8.3 | -0.58 | -34.27 | 1.56 |
| 30D | 1 | | | | 1 | 40 | -74 | 11.5 | 5.17 | -36.12 | 5.46 |
| 31D | 1 | | | | 1 | -26 | 126 | 7.6 | 0.65 | -31.28 | 5.79 |
| 32D | 1 | | | | 1 | 57 | 162 | 10.0 | 8.63 | -38.16 | 6.88 |
| 33D | 1 | | 1 | | | -24 | -63 | 9.7 | 2.21 | -37.62 | 1.00 |
| 34D | 1 | | 1 | | | -13 | -69 | 9.4 | 3.78 | -36.30 | 3.89 |
| 35D | 1 | | 1 | | | 23 | 66 | 9.0 | 1.89 | -35.34 | 2.96 |
| 36D | 1 | | 1 | | | 24 | 68 | 10.8 | 3.68 | -35.80 | 4.29 |
| 37D | 1 | | 1 | | | -26 | -65 | 9.3 | -0.05 | -33.49 | 2.87 |

**Table 5.2 Continue**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 38D | 1 | 1 | -26 | -63 | 9.3 | 2.46 | -34.01 | 4.86 |
| 39D | 1 | 1 | 23 | 66 | 9.4 | 1.89 | -35.34 | 2.96 |
| 40D | 1 | 1 | 24 | 68 | 10.8 | 3.68 | -35.80 | 4.29 |
| 41D | | 1 | -26 | -64 | 4.6 | 3.35 | -36.76 | 3.00 |
| 42D | | 1 | 21 | 69 | 5.7 | 4.25 | -37.89 | 2.78 |
| 43D | | 2 | -20 | -59 | 12.6 | 1.05 | -36.56 | 0.89 |
| 44D | | 2 | 19 | 64 | 9.9 | 0.42 | -36.63 | 0.20 |
| 45D | 1 | | 156 | -57 | 7.5 | 7.07 | -37.38 | 6.10 |
| 46D | 1 | | -164 | 71 | 7.7 | 2.91 | -35.20 | 4.12 |
| 47D | 1 | | -48 | 150 | 4.9 | 2.93 | -34.77 | 4.57 |
| 48D | 1 | | 67 | 164 | 6.6 | 7.10 | -37.45 | 6.06 |
| 49D | 1 | 1 | -49 | 130 | 5.5 | 0.28 | -35.5 | 1.20 |

As in the case of the **3AL**, solvation most stabilizes those regions of dihedral space that does not contain H-bonds (see Figure 5.6). The stabilization in the region of the reverse α-helix is particularly large. The plot for the solvated **3AD** shows many features similar to the (inverted) plot of **3AL**, especially since the H-bonded structures of the β-turn regions (which provide the largest difference in the gas phase) no longer are relevant as they are poorly solvated. This effect tends to flatten the combined (solvated) landscape (Figure 5.7). Corrections for CP and vibrations discussed above for the other diastereomer must, also, be made here. The corrections have a qualitatively different effect as the depth of the relative energy well for the $C_7$β-turn H-bonding region exceed that of the all-L diastereomer due to the formation of an additional $C_{11}$ H-bond that is not possible in **3AL**, thus the CP corrections will be greater.

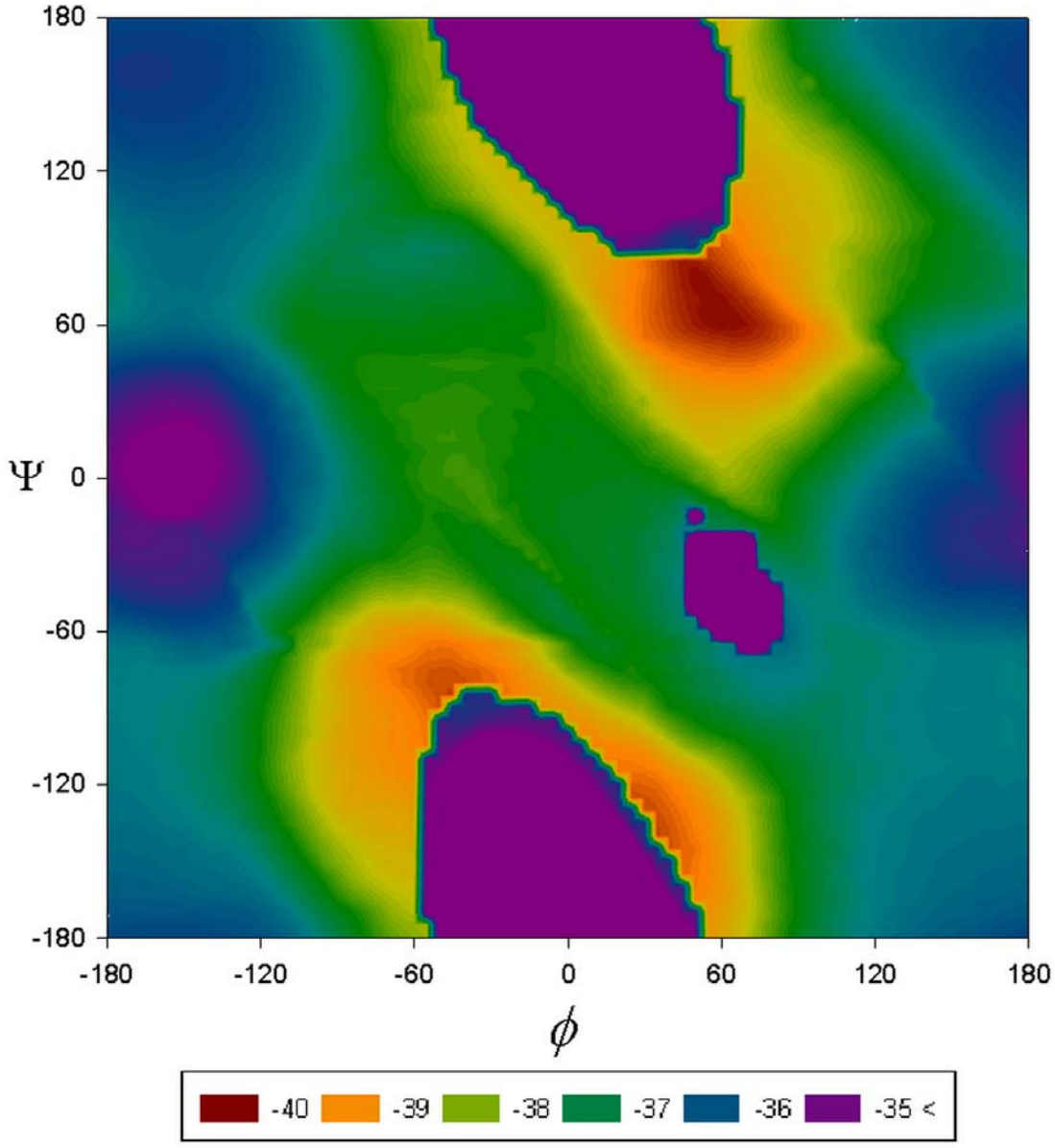



**Figure 5.6.** Plot of the aqueous solvation energies of **3AD** as a function of the Ramachandran dihedral angles. Only those H-bonding structures for which no non-H-bonding structure could be found ($\beta$-turn region) are included.

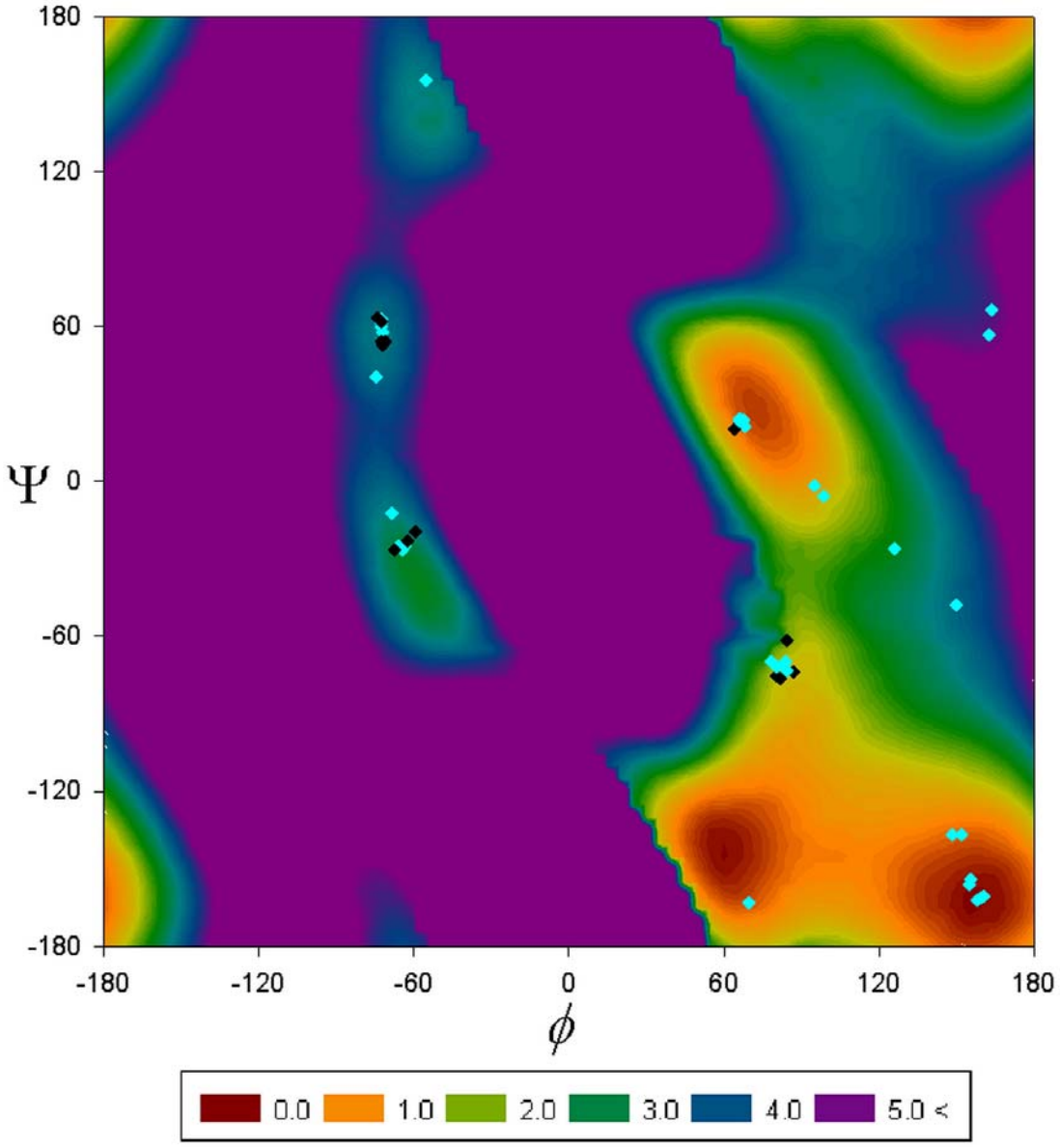



**Figure 5.7.** Ramachandran plot for **3AD** including aqueous solvation using the SM 5.2 method as described in the text. H-bonding structures are indicated by colored points. Black points represent structures that have lower energies than the optimized non-H-bonding structures with the same dihedral angles but higher than the global minimum for non-H-bonding structures; light blue points represent structures with energies higher than non-H-bonding points with the same dihedrals. Energies are relative to the "global" minimum.

The results of recent vibrational studies on tripeptides in aqueous solution indicate that the (uncapped) $^{L}$Ala-$^{D}$Ala-$^{L}$Ala adopts a structure with a negative φ and positive ϕ[26]. Our calculations suggest these structures to be higher than the global

minimum in aqueous solution, but they might reasonably contribute to the equilibrium mixture.

### *5.3 Structures with intramolecular H-bonds*

Even with the φ and ϕ dihedrals fixed at one of the 5184 individual combinations on the grids I used, the rest of the tripeptides retain considerable conformational freedom. I located 47 such structures for **3AL** and 48 for **3AD**. These are depicted in chart S1 together with the global minima. I completely geometrically optimized each of these gas phase structures, calculated their vibrational frequencies (leading to vibrational corrections to enthalpies and free energies), and calculated the solvation energies of these optimized geometries. Tables 5.1 and 5.2 present the relevant data, while structural drawings are presented in the chart S1. I note their positions on the appropriate figures with colored marks at the points for the corresponding dihedral angles, φ and ϕ, (red if the structure is of lower energy than that of the global minimum of the non-Hbonding structures, black if the structures is lower than the non-H-bonding structure with the same φ and ϕ, and light blue if the energy is higher). I include all the structures related (but not identical) to those discussed by Mons et al. [23-25] (Those that were observed are denoted in the figure as open stars) as well as others that are potential minima but have not been observed experimentally.

Among the structures containing internal H-bonds, I find (in addition to $C_5$, which was not considered), $C_7$, $C_8$, $C_{10}$, $C_{11}$, and $C_{13}$ H-bonds. These H-bonds can form coils of the right handed ($C_7$, $C_{10}$ and $C_{13}$ ), left handed ($C_8$ and $C_{10}$) or mixed (one from each group) varieties. Structures containing the most H-bonds contain three (all $C_7$). Several structures contain $C_{13}$ H-bonds, which are found in α-helices.

However, none of these correspond to α-helices, as they all contain a second internal H-bond ($C_7$) and have values of φ and ϕ that are not in the α-helical region.

The comparison of the energies of the H-bonded and non-H-bonded structures requires additional discussion as the H-bonded structures will have more stiffer vibrations and more vibrations frozen (leading to larger vibrational corrections to the enthalpies and free energies) than the others. Also, the stabilizations of H-bonds are generally overestimated by molecular orbital methods due to the basis set superposition error (BSSE).

The enthalpies and free energies of structures are both thermodynamic functions of state, which are only defined at minima on the potential energy surfaces. The enthalpies and (part of) the free energies can be calculated from the normal vibrational modes at these minima. Since, these vibrational modes will be qualitatively different at the H-bonding minima from those of the other structures, one can expect a systematic error that will make the H-bonding structures seem relatively more stable as they should have higher vibrational corrections. In order to estimate the magnitude of this systematic error, I calculated the vibrational frequencies at all of the H-bonded (gas-phase) minima and that of the global non H-bonding minimum (extended b-strand). The differential vibrational corrections to the enthalpies and free energies (compared to the β-strand global minimum) are remarkably constant at 0.35 and 0.27 kcal/mol for all H-bonding structures considerd.

BSSE correction for intramolecular H-bonds poses a problem as there is no obvious and accepted way to break the H-bond donor and acceptor into two individual fragments. In order to avoid this problem, I have simply cut pairs of formamides from the optimized geometries of the internally H-boning peptides. The valences at the cuts were satisfied by playing hydrogens at standard distances from the heavy atoms.

Since these hydrogens are remote from the H-bonds, I do not expect them to appreciably affect the magnitude of the BSSE. Single point counterpoise (CP) calculations performed on these formamide dimers are used to approximate the BSSE's. I did not reoptimize the geometries on the CP-corrected surfaces[33] as the dimers would not be constrained by the peptide backbone, so the optimized geometries would lack any physical meaning. Wieczorek has previously used this procedure to calculate the enthalpies of α-helical poly-alanines. Using the single point CP values overestimates the BSSE. Thus, this calculation provides an upper limit to the BSSE. The CP corrections varied from1 to 2 kcal/mol/H-bond.

Mons et al. studied several capped tripeptides [23-25, 40] but not specifically trialanines, and only in the all-L diastereomer using gas phase infrared spectroscopy. Not all the structures he studied were actually observed. Some were only calculated. I find all of the **3AL** structures that I calculated that correspond to those *observed* by him to be more stable than those without *intra*molecular H-bonds in the gas phase, in good agreement with his reports. When solvation energies are included, only two structures remain more stable than the global minimum for non H-bonded structures, and both become higher than the global minimum after vibrational and CP correction (although one remains very close, see above).

### *5.4 Comparison of the diastereomers*

Comparison of the energies of these diastereomers can be made either with respect to the global minimum for each or with respect to an external standard. Any external standard must necessarily be arbitrary. My mentor and I have discussed this problem previously [8, 41]. I have chosen the total energy (relative to dissociated nuclei and electrons) as this standard as it is straightforward and equivalent (in

relative energy) to the polycondensation reactions starting from component amino acids (which in our lab have used previously) [4, 8, 28] as the energies of the two enantiomers of free alanine are necessarily the same. The particular energetic comparison made can be very important for the interpretation of experimental results. For example, if one compares the equilibria of folded vs. unfolded diastereomeric peptides in aqueous solution, one obtains a result based on the differences in free energies of the solvated folded and unfolded species of each diastereomer. Typically, only the effect of the inversion of the chiral center upon the energy of the folded structure is discussed while that of the unfolded structure is ignored, since its structure and solvated energy was not known. On the other hand, the data in this paper provide information with regard to the relative energies of the solvated capped trialanine diastereomers considered, but not upon those of folded structures (these peptides are too short to fold into most secondary structures), nor of other pairs of diastereomeric peptides, although they may be illustrative of the latter.

I shall begin by considering the differences between the diastereomers with reference to the global minima for that diastereomer and then discuss the differences with respect to the overall global minima for the two. The global minimum for **3AD** is very slightly lower in energy (0.014 kcal/mol) than that for **3AL** in the gas phase. The solvation free energy stabilizes **3AD** more than **3AL** by 0.214 kcal/mol, rendering the solvated energy of **3AD** lower by 0.227 kcal/mol.

The differences in the energy landscapes for the two diastereomers are compared in several figures. In each of these figures, the blue region depicts the part of the surface where **3AL** is favored, while the red region depicts that part where **3AD** is favored. One should bare in mind that the structures of **3AL** and **3AD** that correspond to the same values of φ and ϕ can be significantly different geometrically. The first of

these (Figure 5.8) depicts the difference in the gas phase energies of the diastereomers. As expected, the left side (particularly the upper left) is more stable for **3AL** while the reverse is true for **3AD**. In the region for the β-turn, **3AD** is much more stable than **3AL**. Perhaps a more informative comparison is depicted in Figure 5.9 witch depicts the landscape for the difference in gas phase energy for **3AL** with the *mirror image* of that for **3AD** (which would be zero throughout if the two structures were enantiomers). The **3AL** structure is significantly more stable than the **3AD** in the regions for the right- (of **3AL**) and the left-handed (of **3AD)** α- and $3_{10}$ helices, as well as the PII region, all of which are superimposed when the mirror image of **3AD** is plotted against **3AL**. The **3AD** diastereomer is more stable for the β-turn structures (which are not superimposed), in the reverse helical regions (left for **3AL** and right for **3AD**) and in the relatively unimportant regions where neither is very stable.

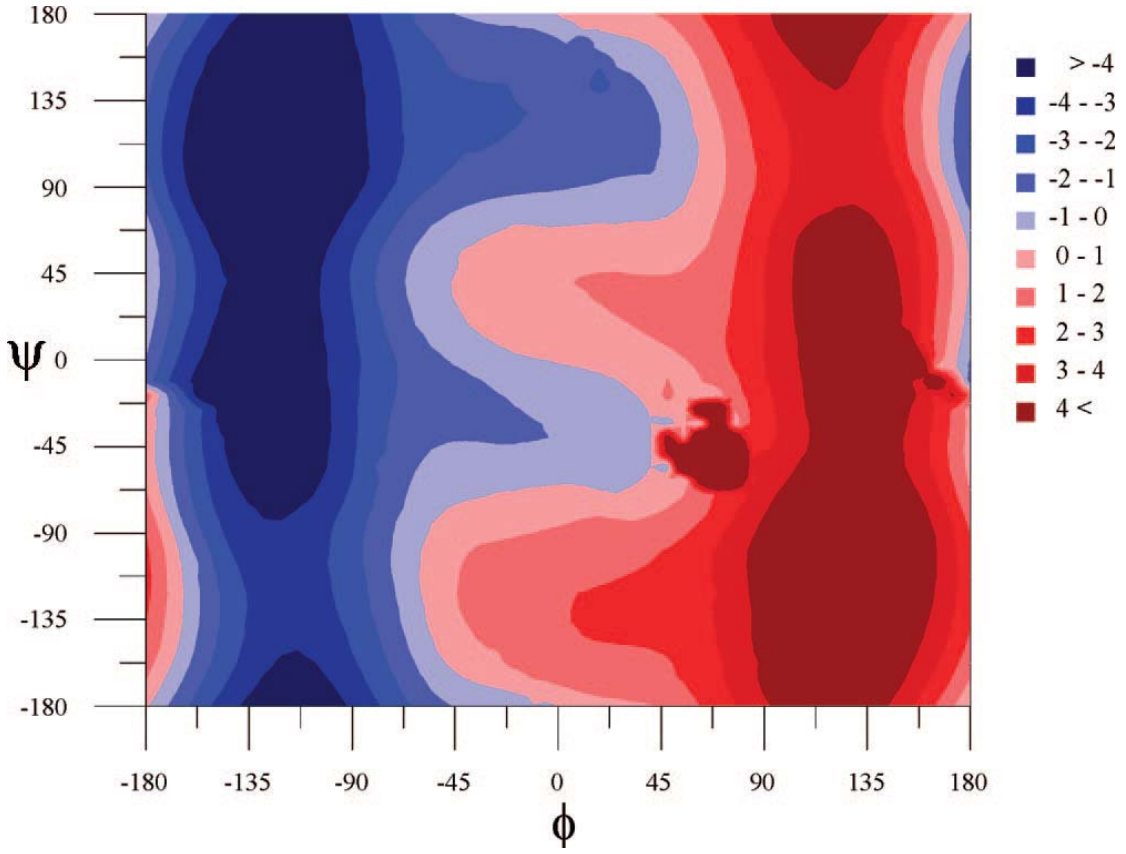



**Figure 5.8.** Potential energy difference for **3AL-3AD** in the gas phase

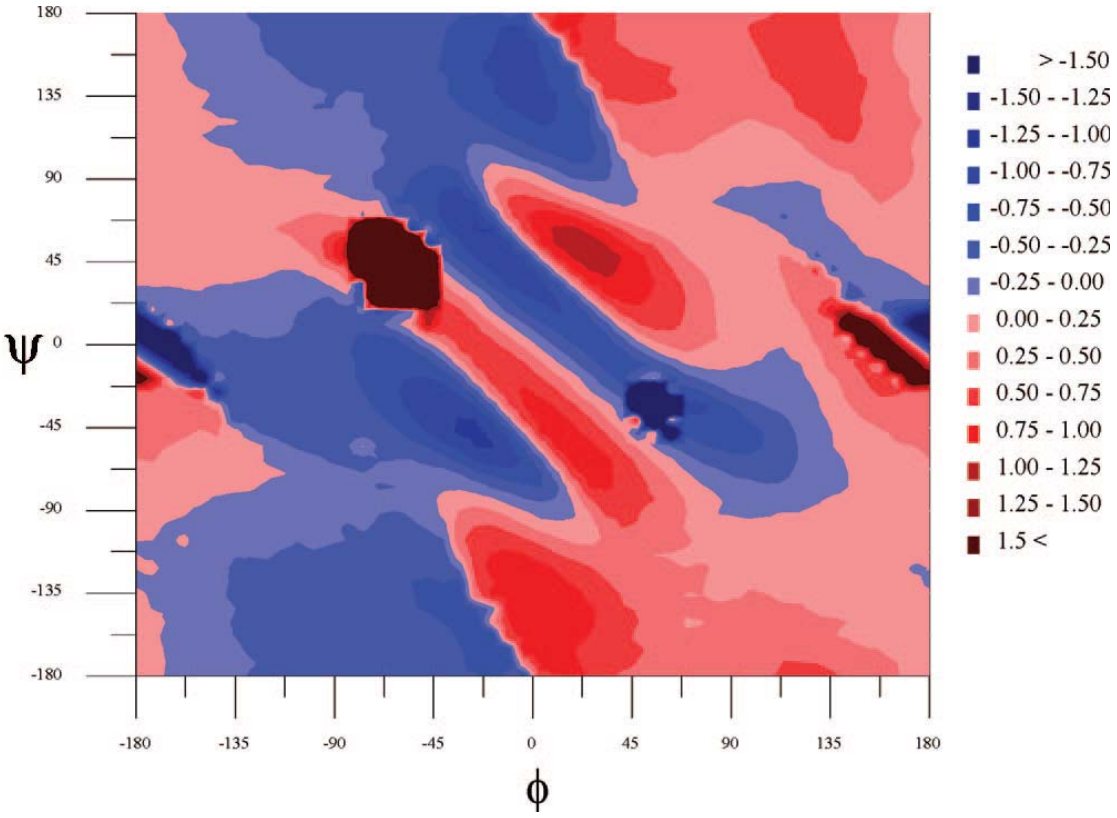



**Figure 5.9:** Potential energy difference plot of **3AL**-*mirror image* of **3AD** in the gas phase

Figures 5.10 and 5.11 depict the differences in the solvation free energies between **3AL** and **3AD** and between **3AL** and the mirror image of **3AD**. Here we see that the β-strand, most of the PII, α- and $3_{10}$-helical regions are all better solvated for **3AD**, while the **3AL** b-turn is better solvated.

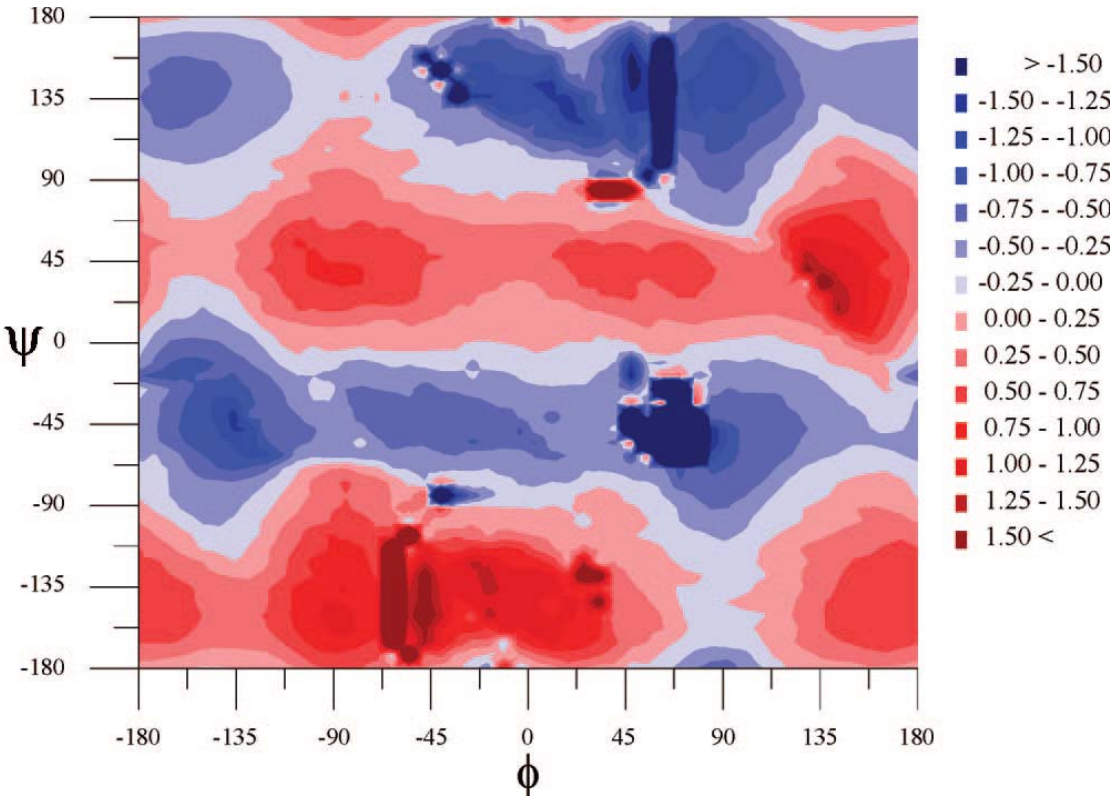



**Figure 5.10:** Difference map for the aqueous solvation free energies for **3AL**-**3AD**.

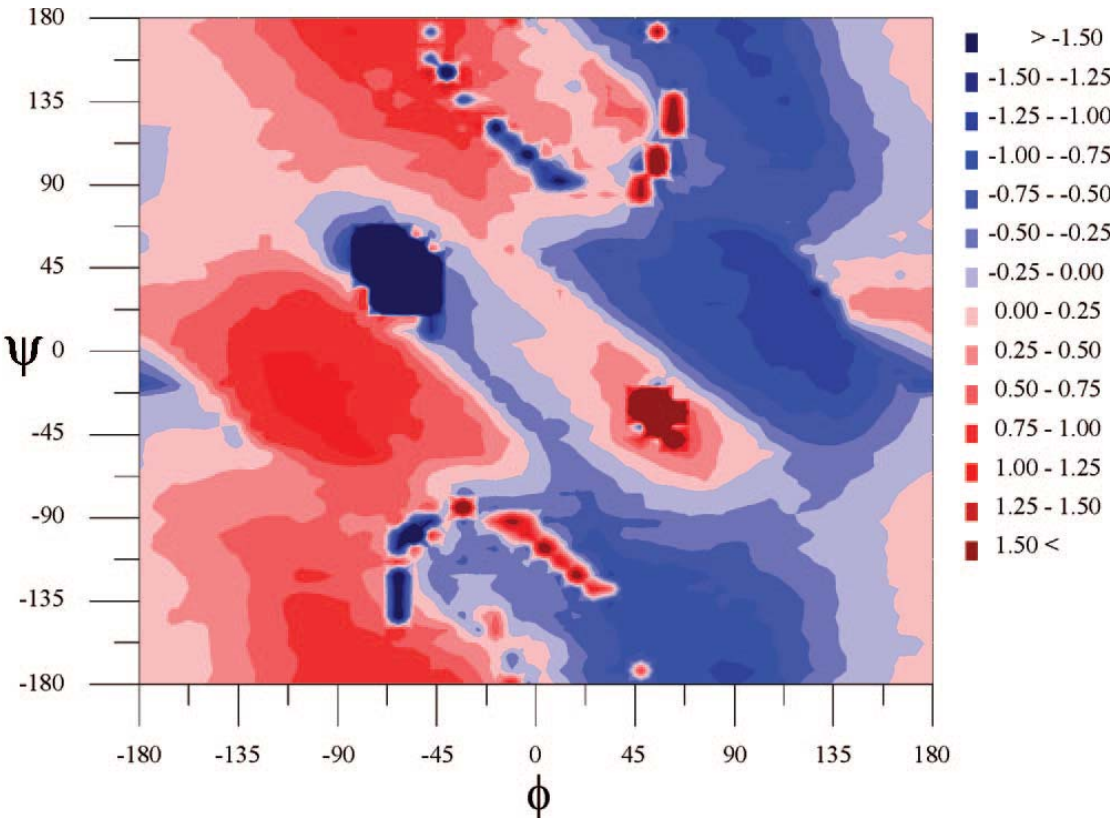



**Figure 5.11:** Difference plot for the solvation free energies for **3AL**-*mirror image* of **3AD**.

Finally Figures 5.12 and 5.13, depict the solution energy (gas phase energy + solvation free energy) plots of **3AL -3AD** and **3AL** - the mirror image of 3AD. From the latter, one can see **3AD** to be more stable in aqueous solution, in the β-strand, PII and helical regions, basically the reverse of the gas phase relative stabilities. Thus, the differential solvation free energies dominate the combined surface for energy plus solvation free energy.

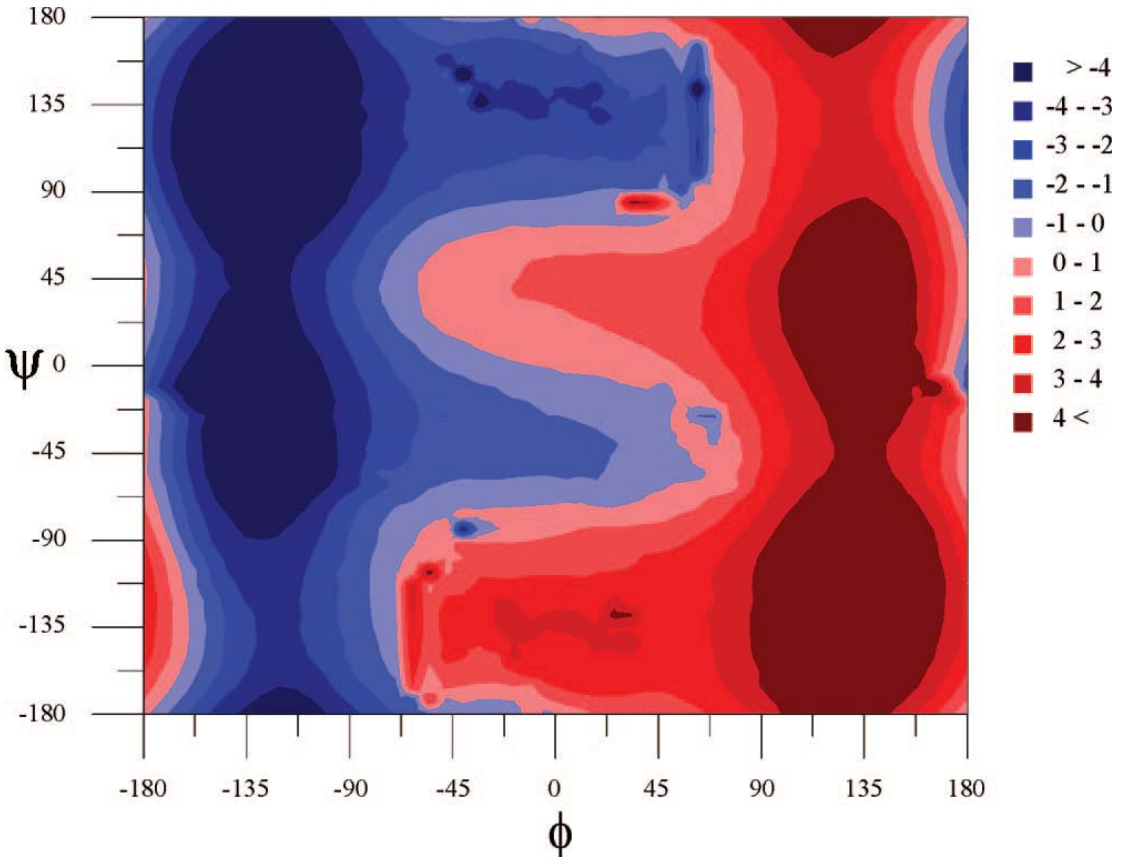



**Figure 5.12:** Difference plot for the gas phase potential energies **+** the solvation free energies for **3AL-3AD**.

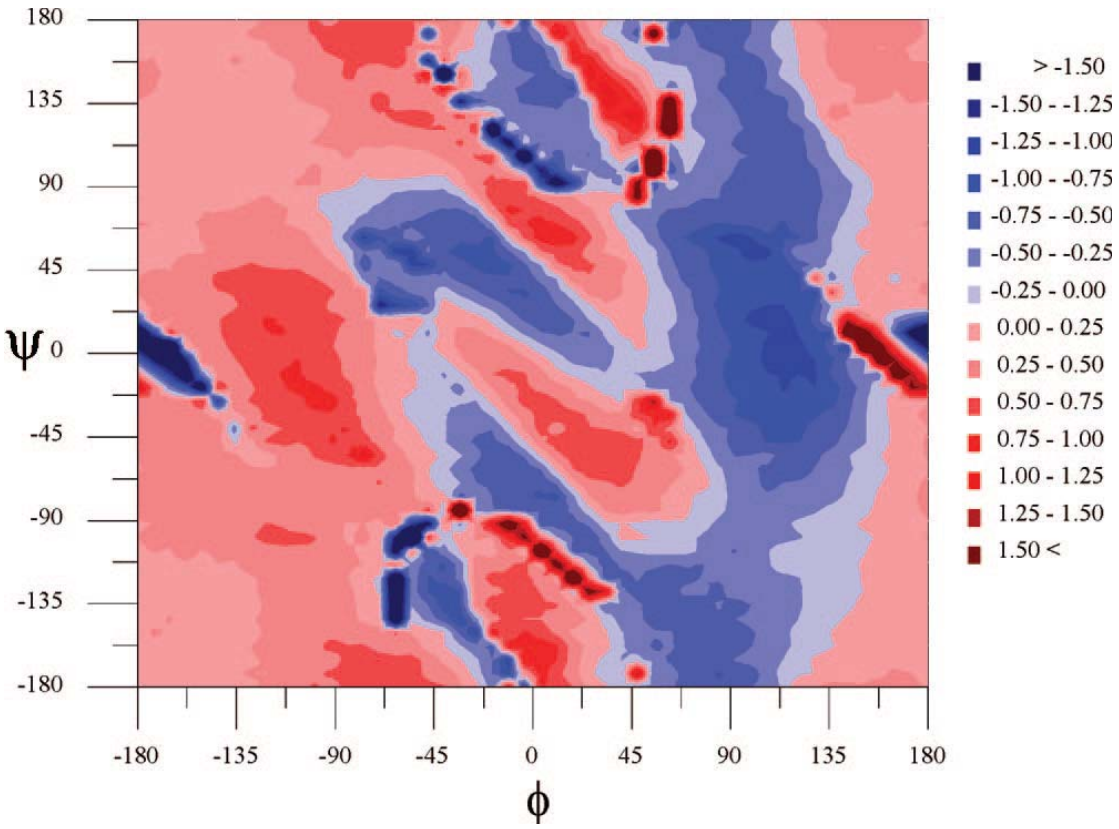



**Figure 5.13:** Difference plot for the gas phase potential energies **+** the solvation free energies for **3AL**-*mirror image* of **3AD**.

### *5.5 Helices*

The observation that aqueous solvation stabilizes the $3_{10}$-helix and the conformation that resembles an α-helix (but contains no internal H-bond) bears further discussion. Several groups have reported that poly-alanine α-helices containing more than ~10 residues form spontaneously in water and that they are enthalpy driven [1, 3, 42]. Gas phase DFT calculations predict enthalpy driven helix formation [4] and suggest that aqueous solvation does not appreciably change this prediction [6]. Discussions of the physical bases behind these observation typically focus on H-bonding and hydrophobic effects. The importance of stabilization of the large dipole moments of α-helices has not received much attention, although is has been invoked in the solvation of protonated α-helices [43]. The apparent contributions

of H-bonding and hydrophobic interactions would be diminished to the extent that the dipole stabilization contributes to the overall energy.

### 5.6 β-*Turns*

I noted that β-turns are more stable for **3AD** than for **3AL**, due to the additional $C_{11}$ H-bond that can form in the latter. In both cases, solvation destabilized these structures with respect to the non-H-bonding conformations. However, this might not be the case for other tripeptides, in particular, those containing proline, which has a natural propensity to form turns. I have previously suggested that when Gly is replaced by $^{D}$Ala in collagen-like structures, the triple helix might be stabilized. A recent experimental study failed to confirm this by measuring the equilibria between folded and unfolded states in water [44]. Equilibrium constants correspond to the ΔG's between the two thermodynamic states. One of the possible reasons for this failure might be the increased stability of the unfolded mutated peptide rather than the lack of stability of the folded state. The increased relative stability of the **3AD** β-turn structures suggests this might be a possibility. The (albeit) small overall preference for **3AD** in its solvated global minimum might also be indicative for a slight preference for other unfolded peptide containing individual $^{D}$Ala's.

### 5.7 Conclusions

Ramaschandran plots of capped trialanine peptides based upon B3LYP/d95(d,p) DFT calculations differ significantly from those based upon smaller peptides. These tripeptides can form several kinds of internal H-bonds ($C_5$, $C_7$, $C_8$, $C_{10}$, $C_{11}$, and $C_{13}$), which are not possible for the shorter peptides that have been traditionally studied in dihedral space. Consideration of these H-bonding structures is important in evaluating

relative energies, as they particularly stabilize the $3_{10}$helix and the region of dihedral space that corresponds to β-turns. The $3_{10}$-helix is particularly stable for **3AL** and the β-turns for **3AD** (as they contain a $C_{11}$H-bond absent in the corresponding structures of **3AL**).

Aqueous solvation considerably changes the Ramachandran plots. The β-strand that corresponds to the global minimum (of non-H-bonded structures) remains the global minimum in solution, but three other structures - polyproline II, $3_{10}$-helix and a conformation whose dihedral angles resemble those of an α-helix become almost equivalent in energy (withn ~ 1 kcal/mole); thus, all four should significantly contribute to the equilibrium mixture. Solvation, unsurprisingly, destabilizes most structures containing internal H-bonds relative to those that do not. The $3_{10}$-helix provides a major exception to this generality, as its free energy of solvation exceeds that of the β-strand due to the interaction of its large dipole moment with the high dielectric medium (water). After consideration of the effects of BSSE and vibrational corrections, among the structure with internal H-bonds, only the $3_{10}$-helix remains energetically competitive in aqueous solution.

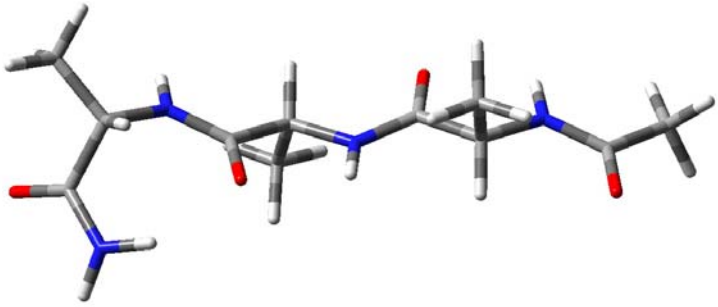
The importance of the dipole moment to the solvation energy becomes readily apparent with these trialanine models. Smaller models can not develop as large dipole moments as there are fewer C=O bonds that can align in parallel directions. Solvation models that do not account for the interactions of the solute dipole moments with bulk solvent will likely overlook the importance of the higher dipole conformations.

Our calculations show that the diastereomeric capped trialanine peptides, **3AL** and **3AD** have significantly different Ramachandran plots in both the gas phase and in aqueous solution. The energy differences of the several non-H-bonding conformations

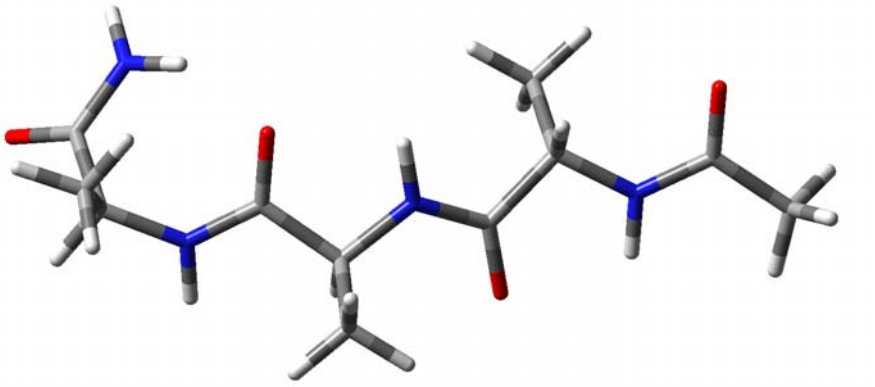
of each that correspond to local minima are significantly smaller in aqueous solution than in the gas phase.

**Chart S1: Structures for 3AL (L) and 3AD (D)**

1 L

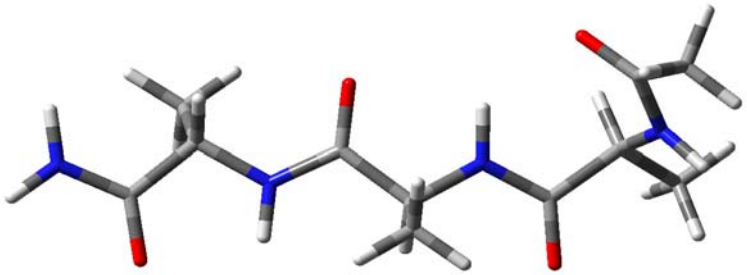


2 L

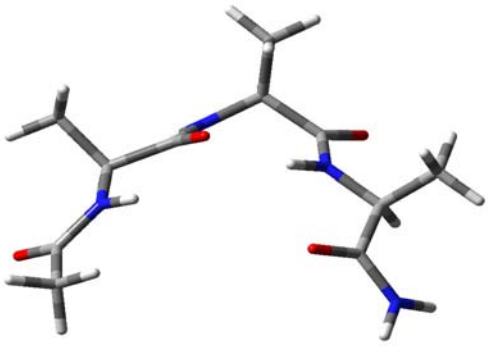


3 L

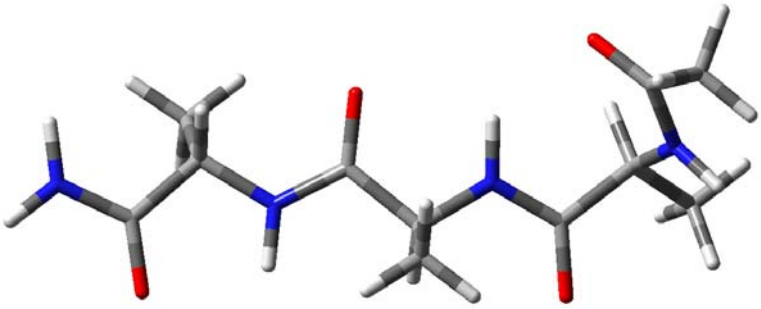
**Chart S1 Continue**



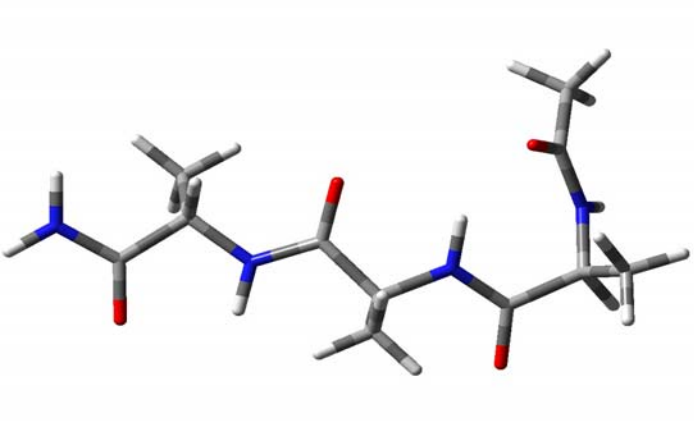
4 L



5 L

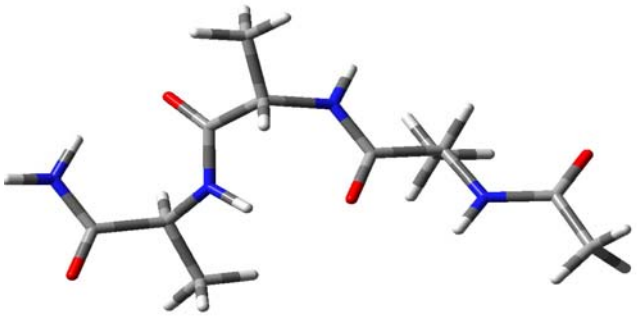


6 L

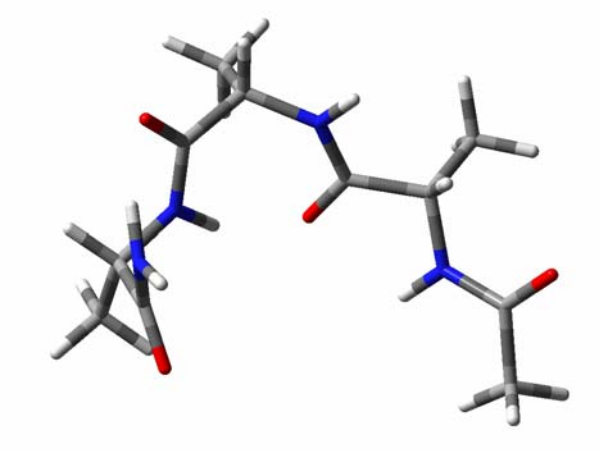


7 L

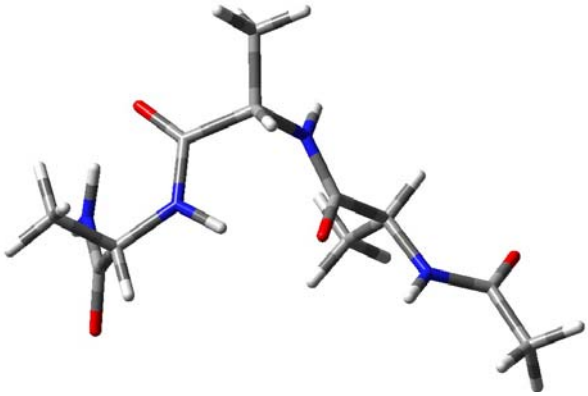
**Chart S1 Continue**



8 L

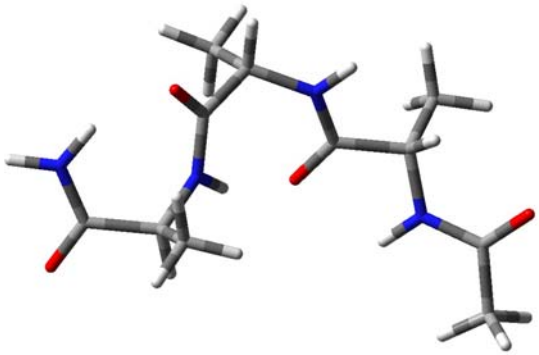


9 L

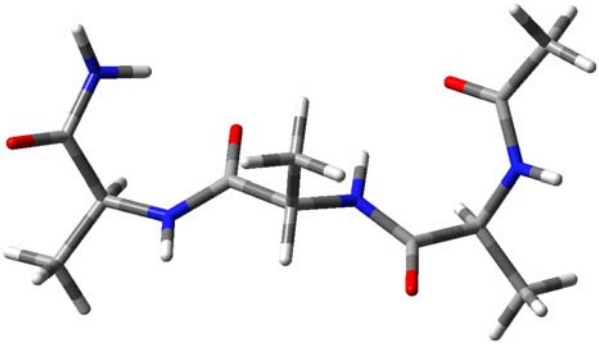


10 L

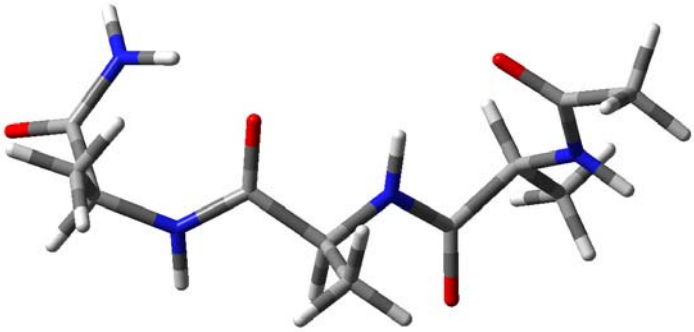
**Chart S1 Continue**



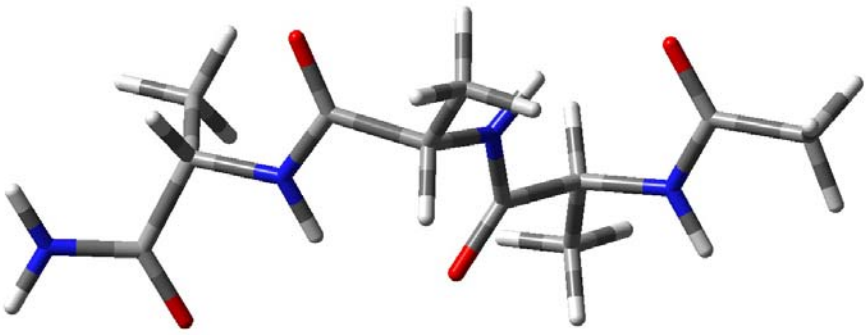
11 L



12 L

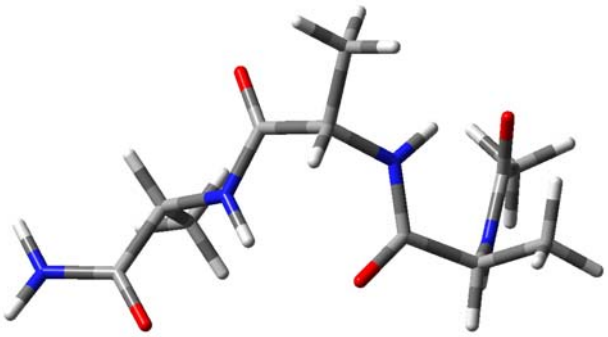


13 L

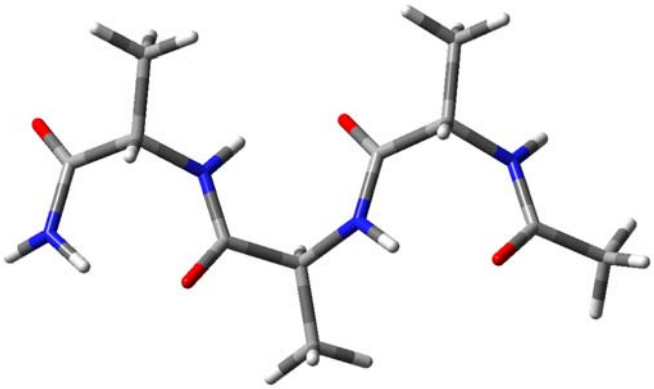


14 L

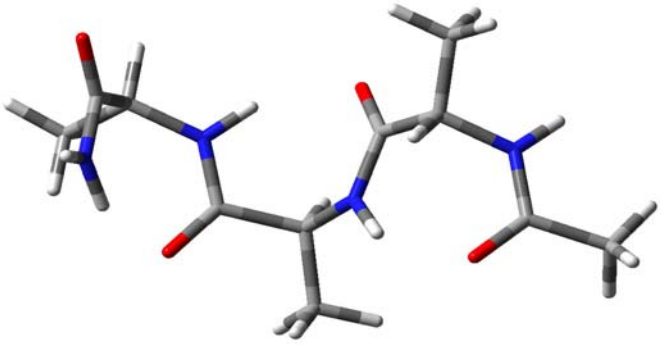
**Chart S1 Continue**



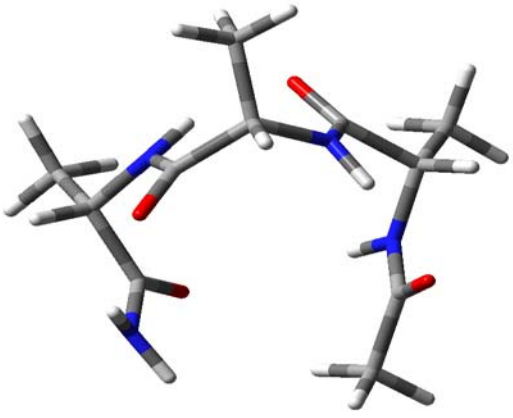
15 L



16 L

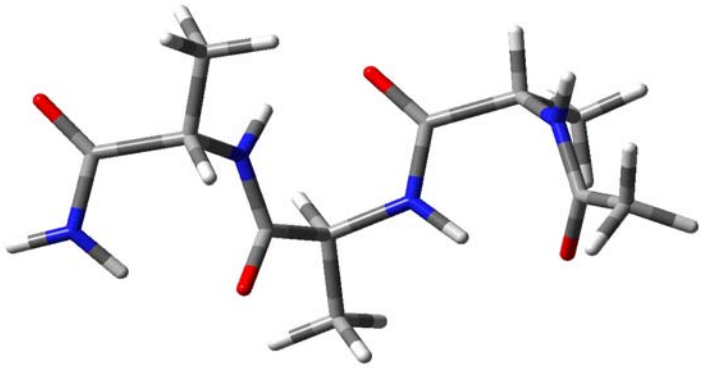


17 L

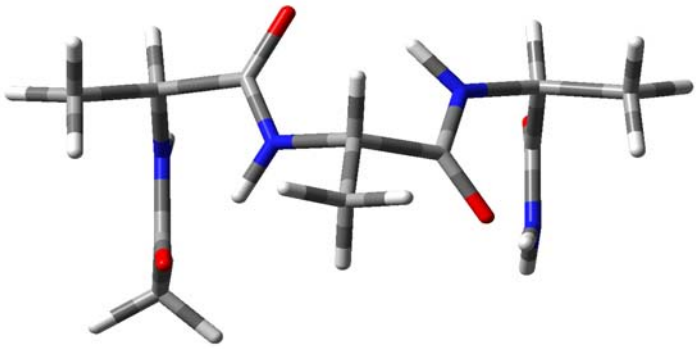


18 L

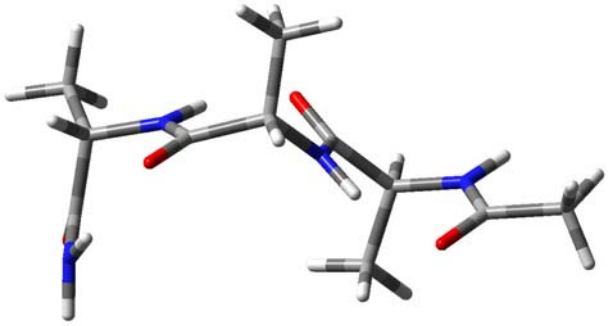
**Chart S1 Continue**



19 L

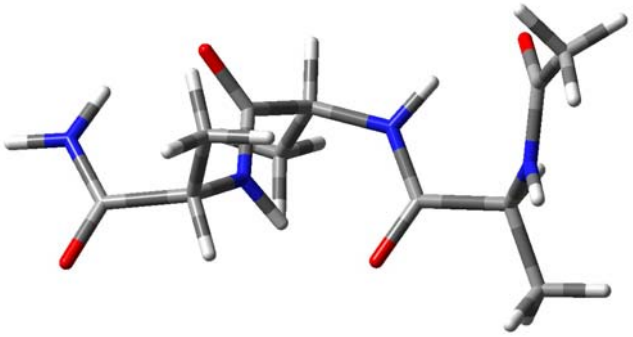


20 L

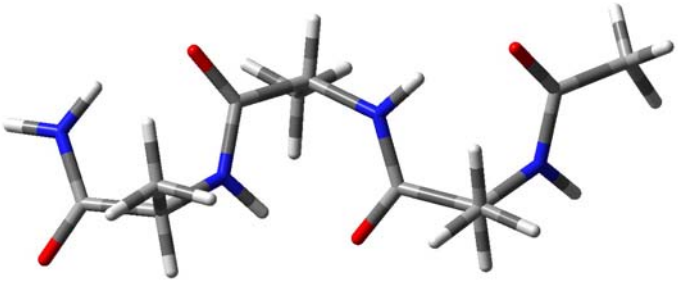


21 L

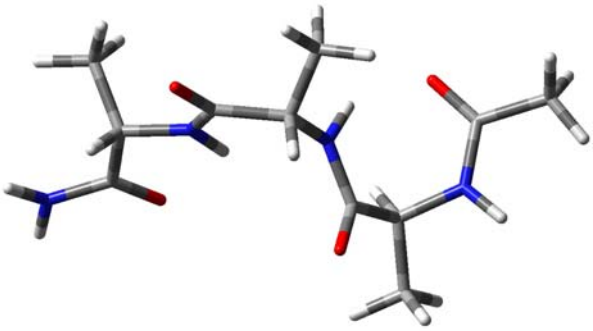
**Chart S1 Continue**



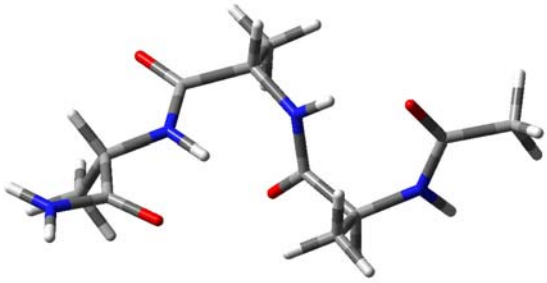
22 L



23 L

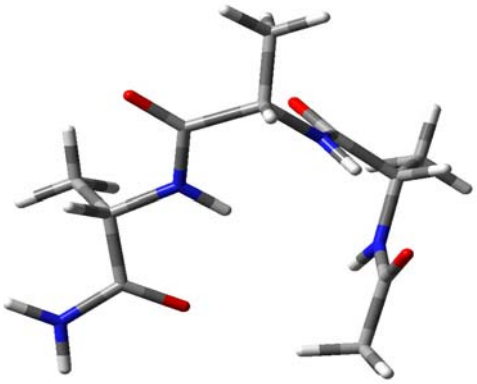


24 L

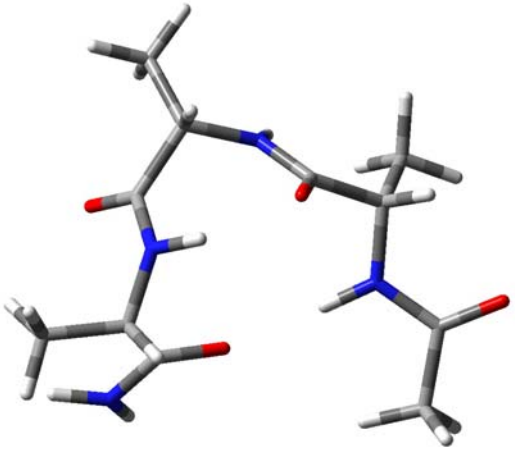


25 L

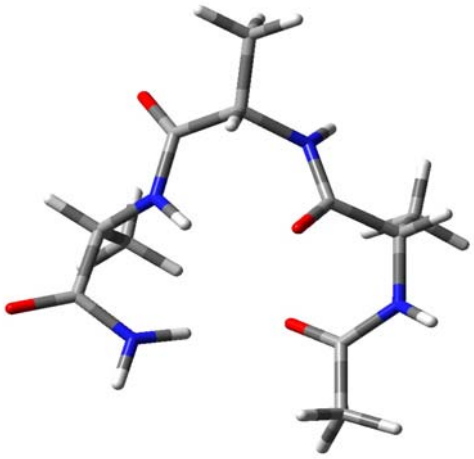
**Chart S1 Continue**



26 L

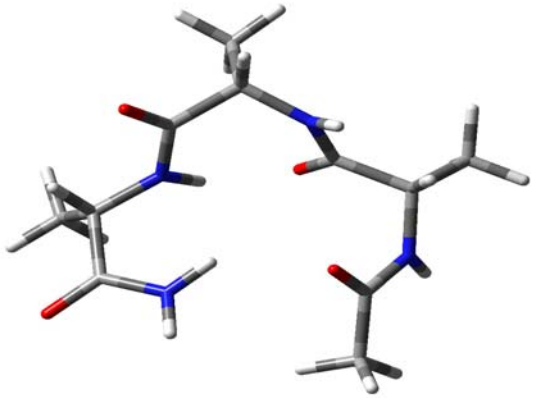


27 L

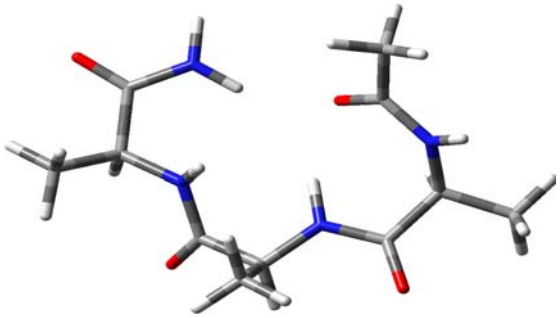


28 L

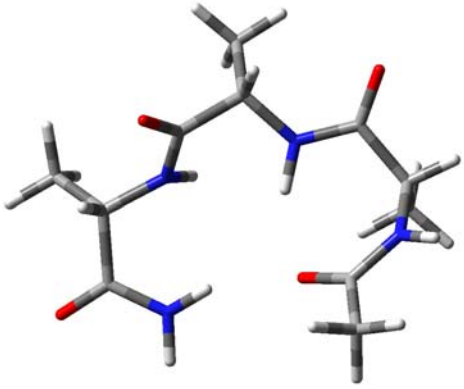
**Chart S1 Continue**



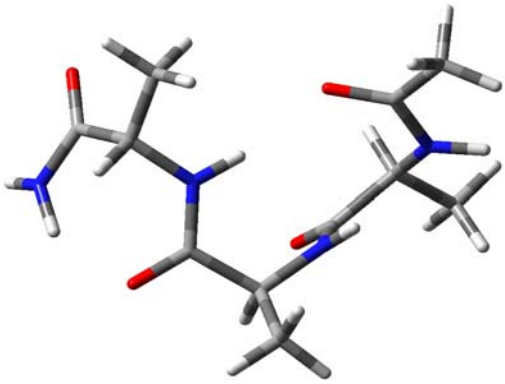
29 L



30 L

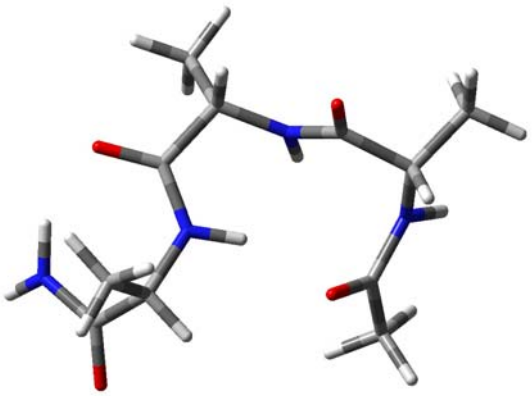


31 L

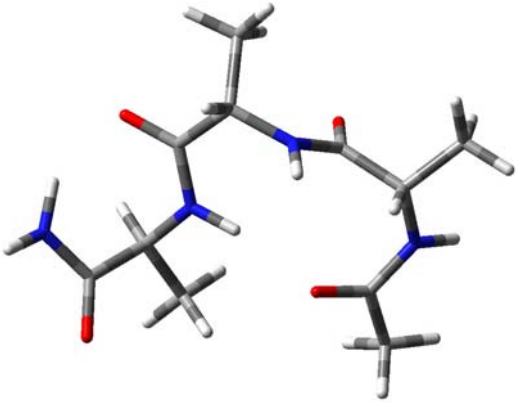


32 L

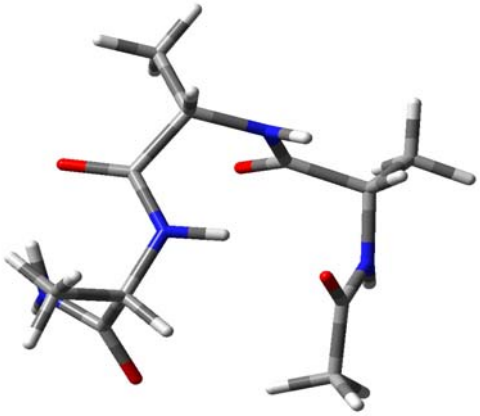
**Chart S1 Continue**



33 L

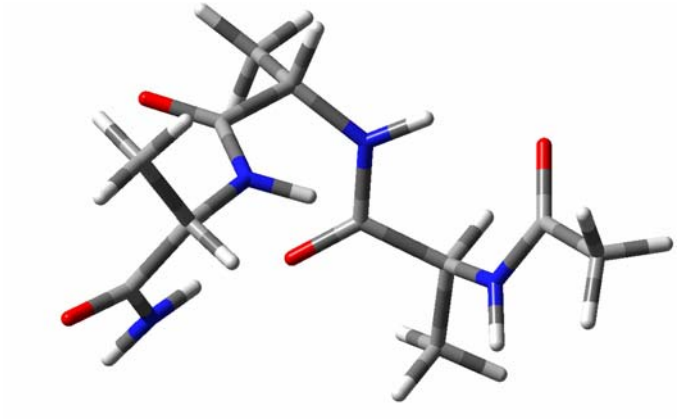


34 L

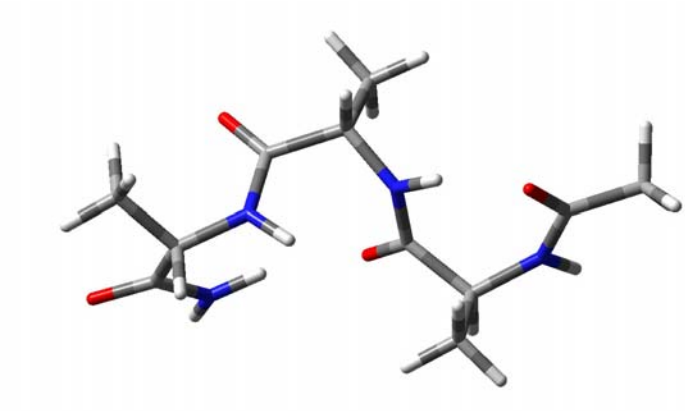


35 L

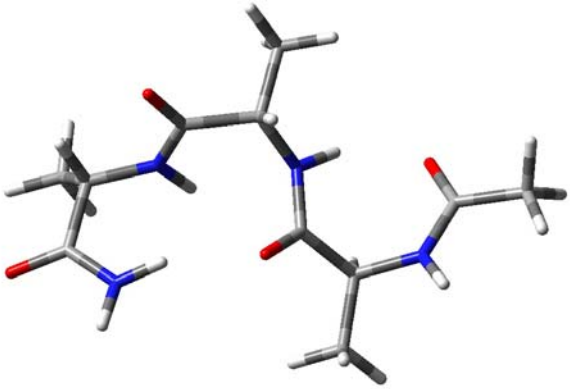
**Chart S1 Continue**



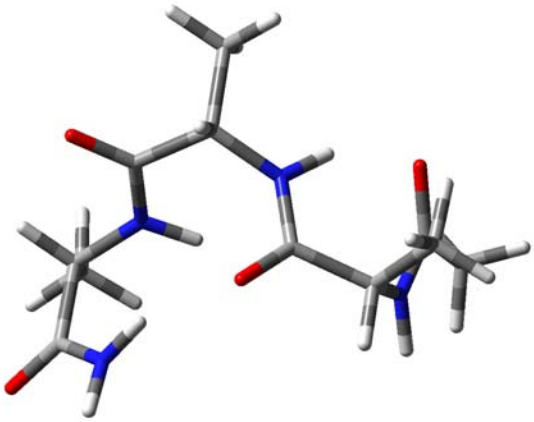
36 L



37 L

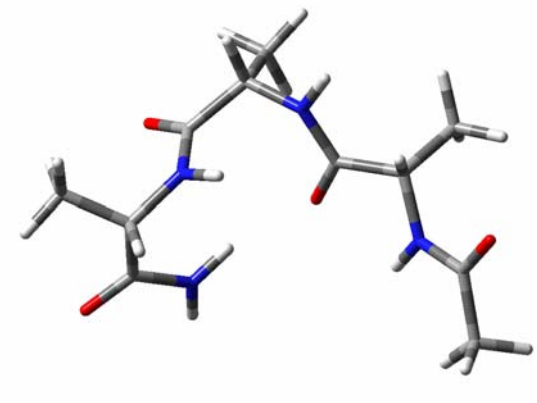


38 L

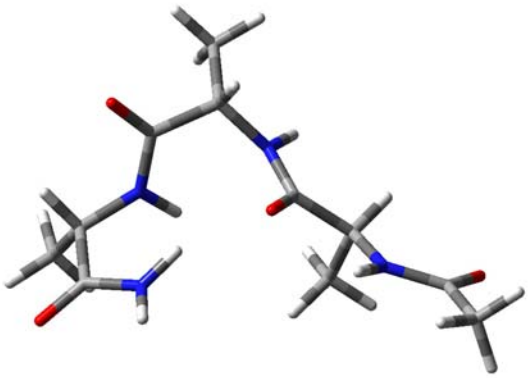


39 L

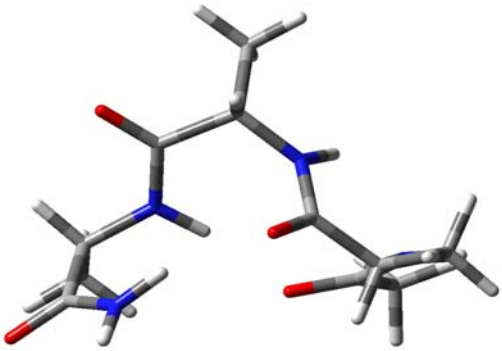
**Chart S1 Continue**



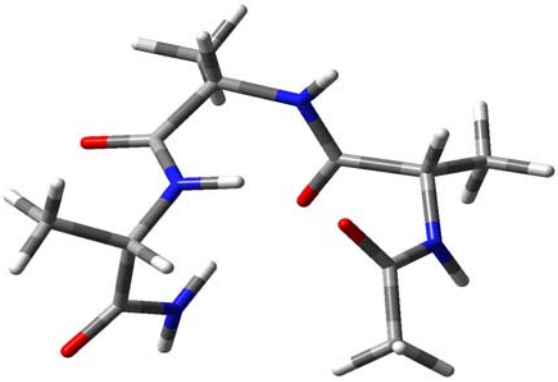
40 L



41 L

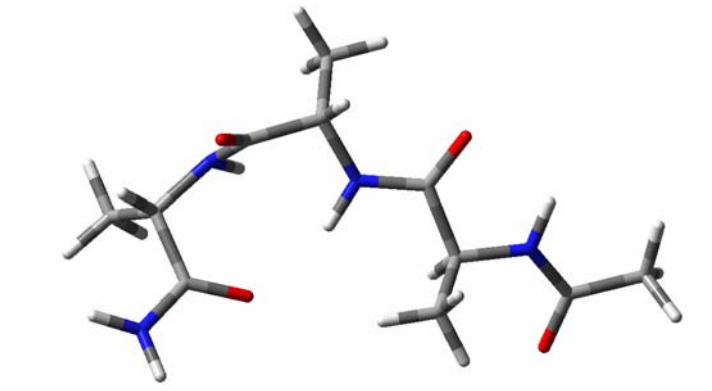


42 L

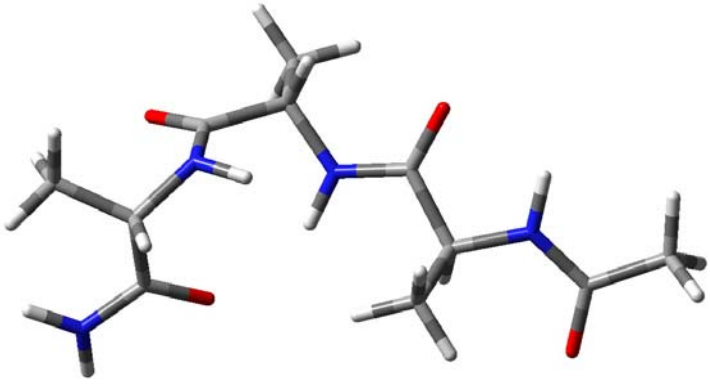


43 L

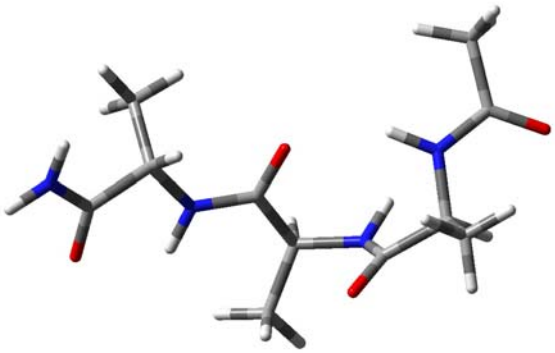
**Chart S1 Continue**



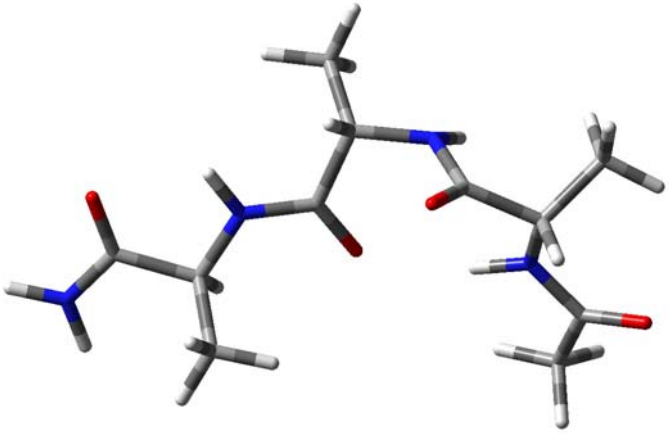
44 L



45 L

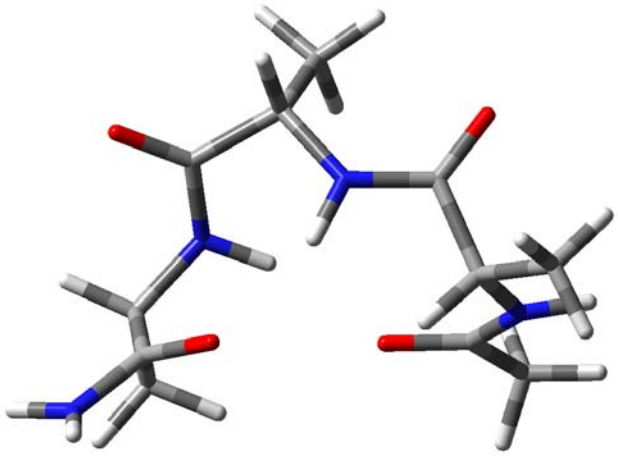


46 L

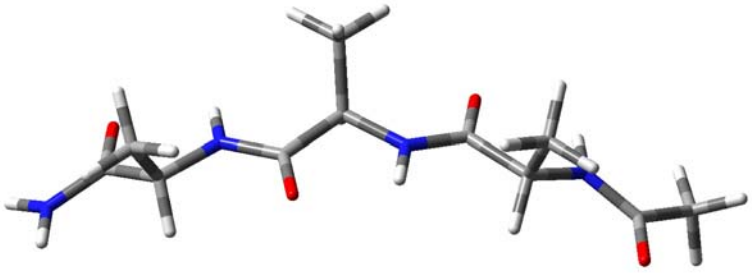


47 L

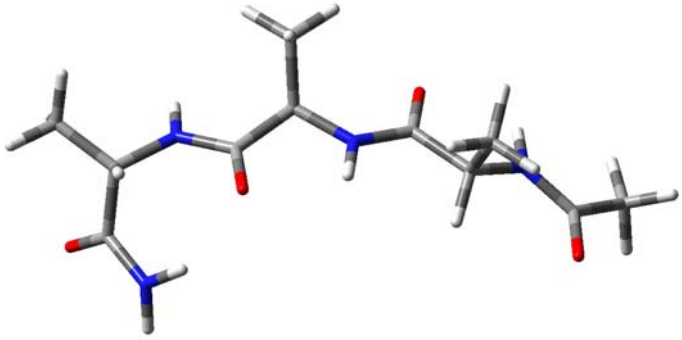
**Chart S1 Continue**



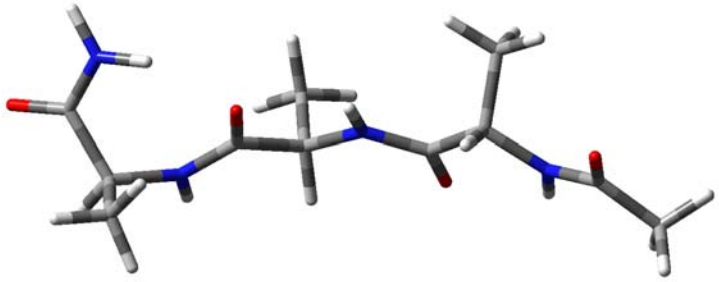
48 L



1 D

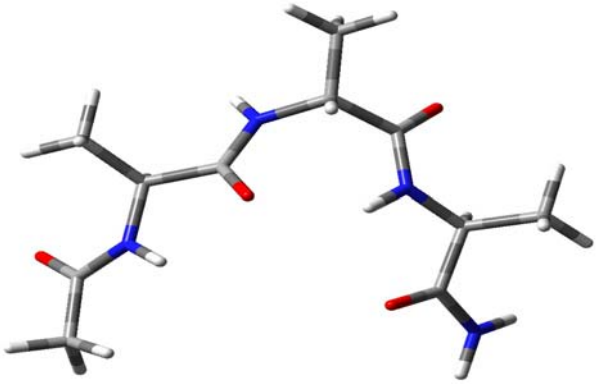


2 D

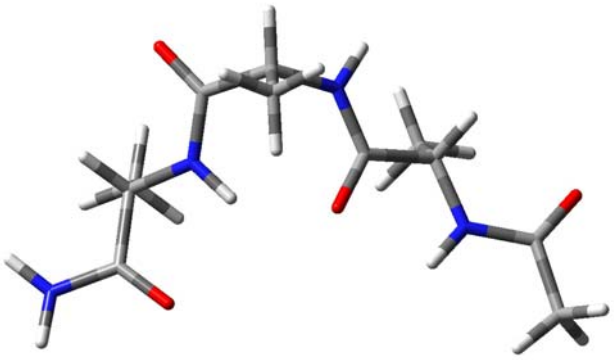


3 D

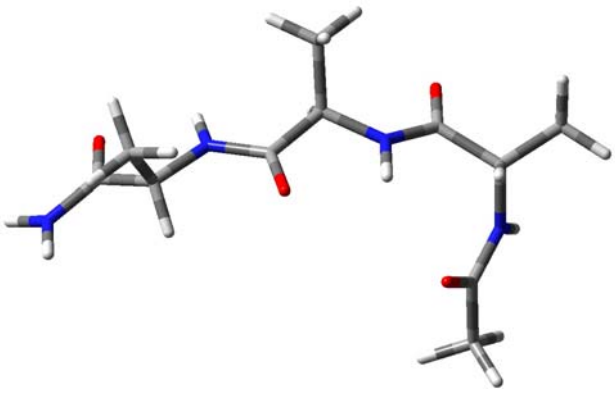
**Chart S1 Continue**



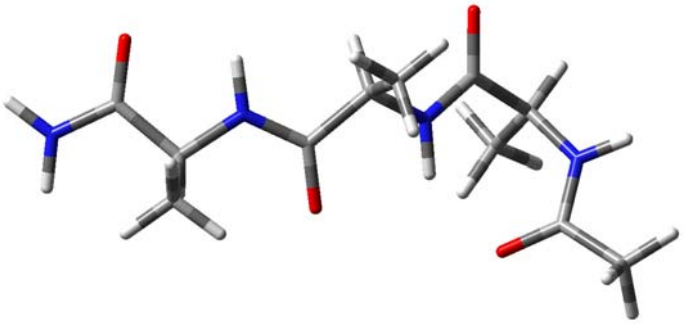
4 D



5 D

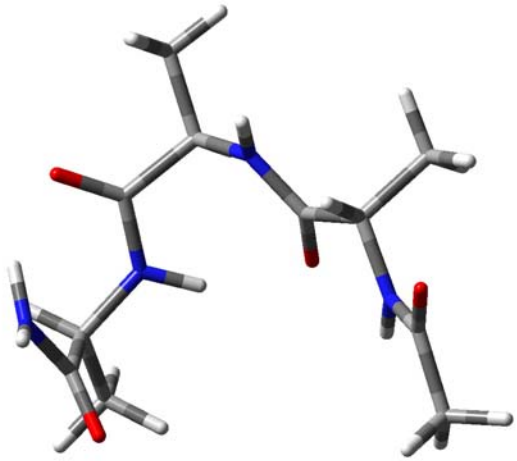


6 D

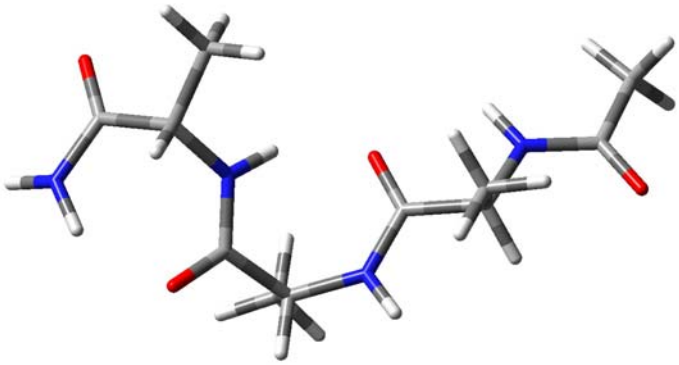


7 D

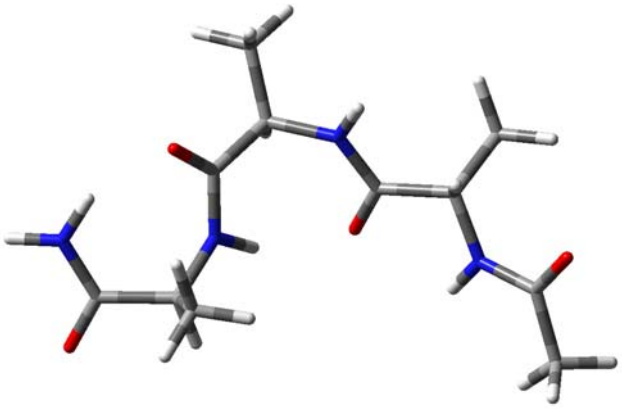
**Chart S1 Continue**



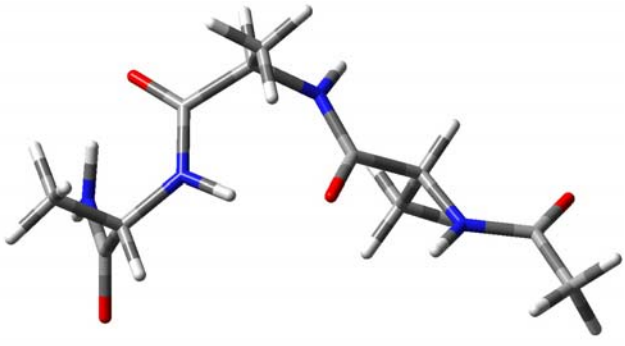
8 D



9 D

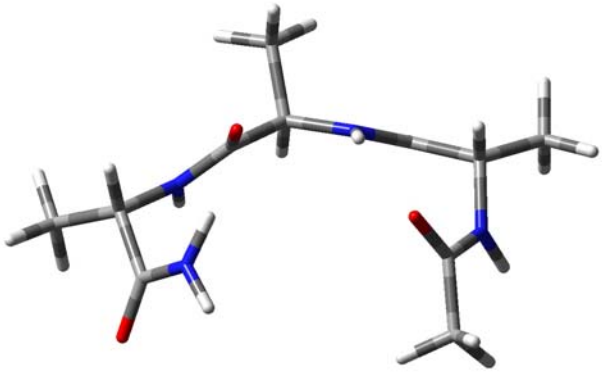


10 D

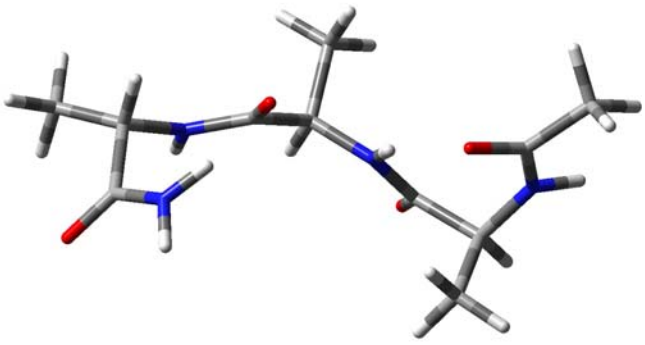


11 D

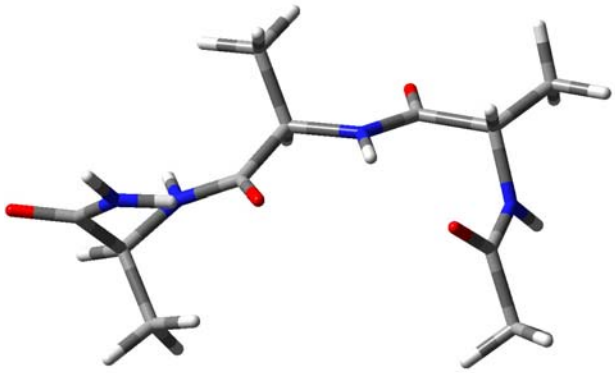
**Chart S1 Continue**



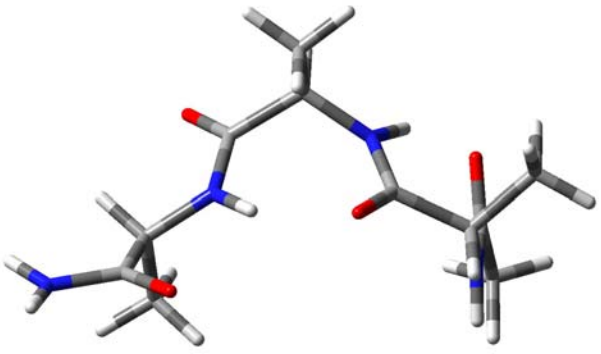
12 D



13 D

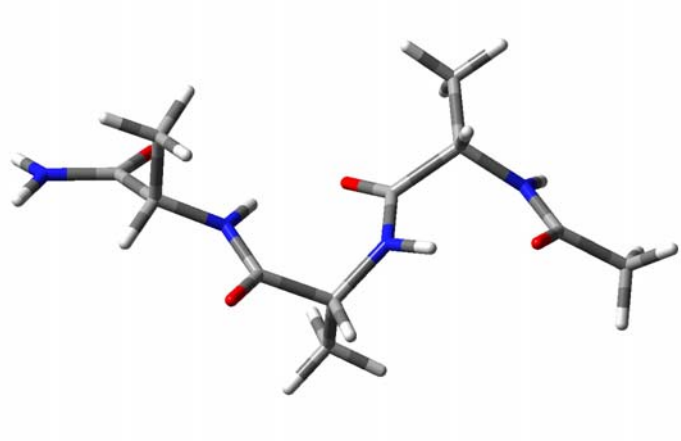


14 D

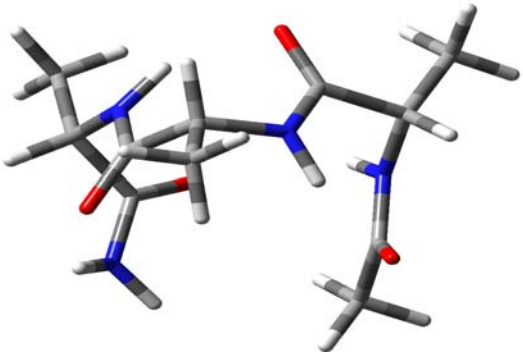


15 D

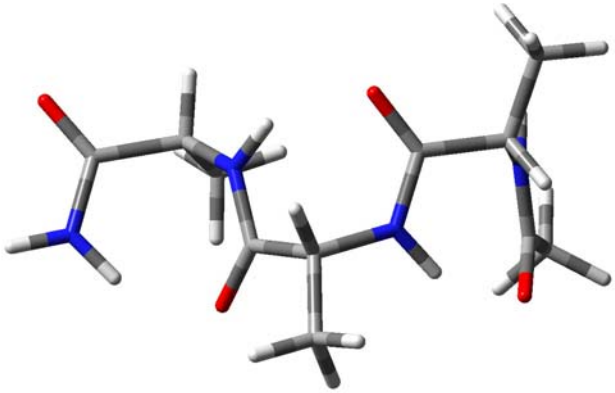
**Chart S1 Continue**



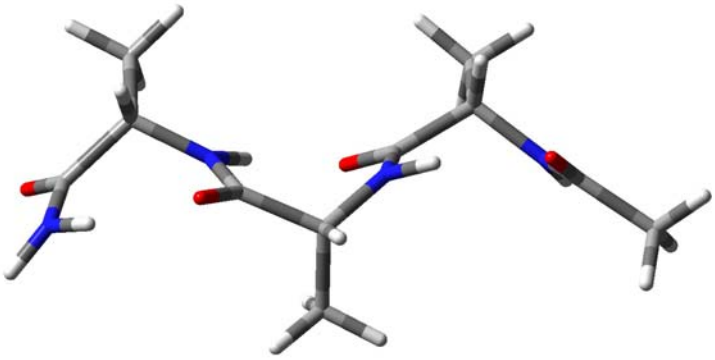
16 D



17 D

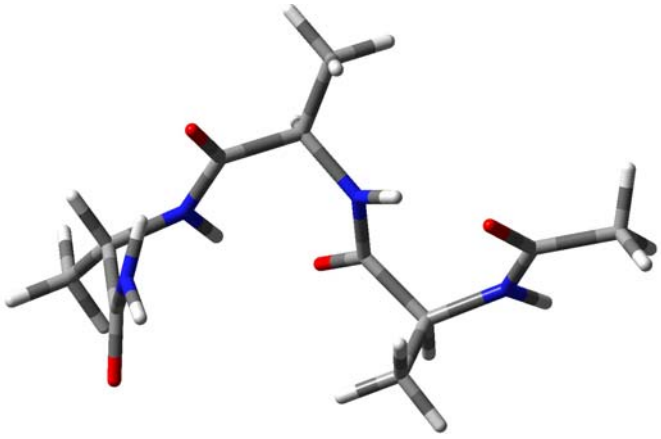


18 D

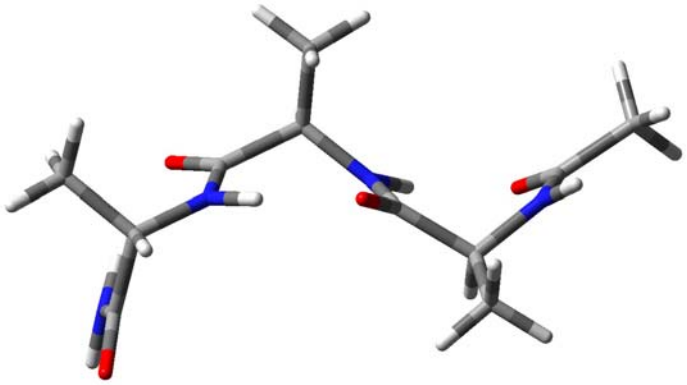


19 D

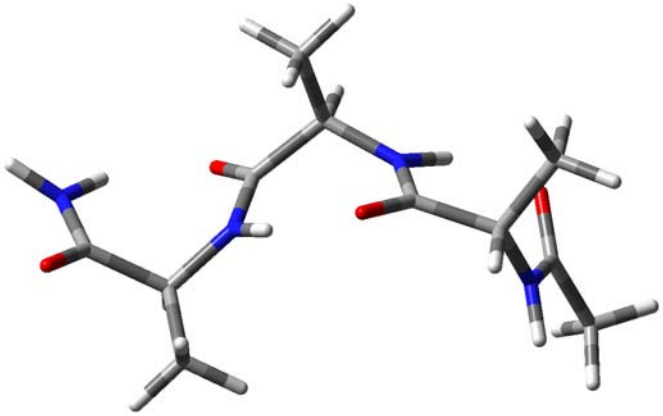
**Chart S1 Continue**



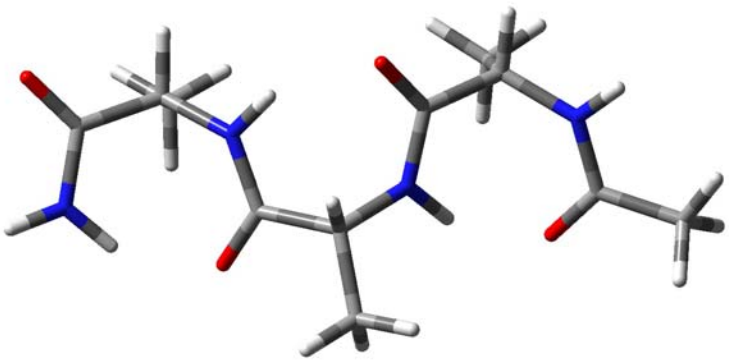
20 D



21 D

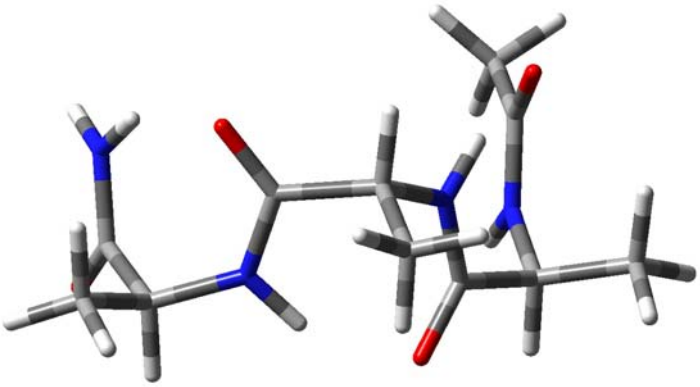


22 D

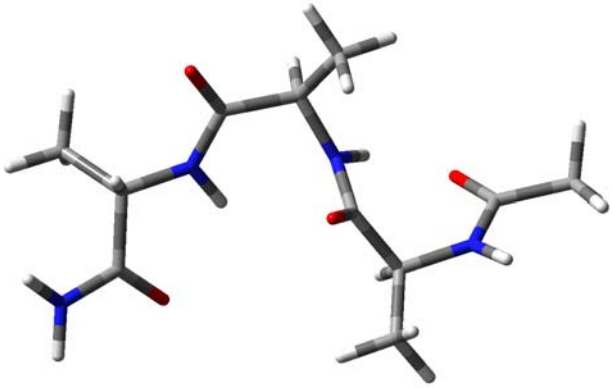


23 D

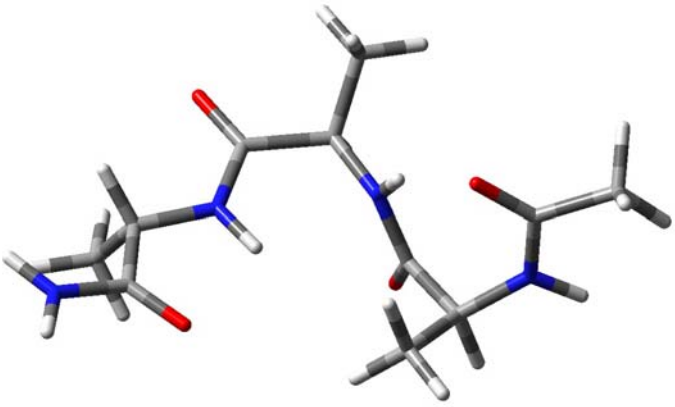
**Chart S1 Continue**



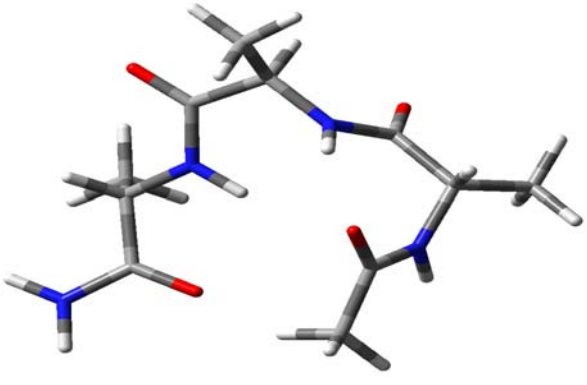
24 D



25 D

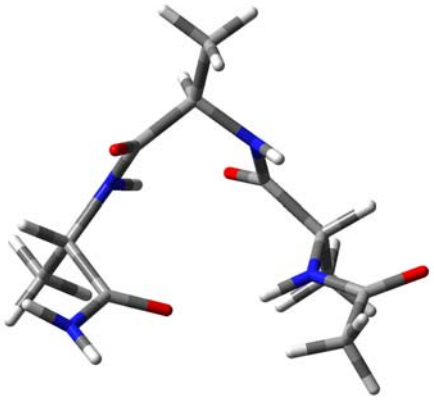


26 D

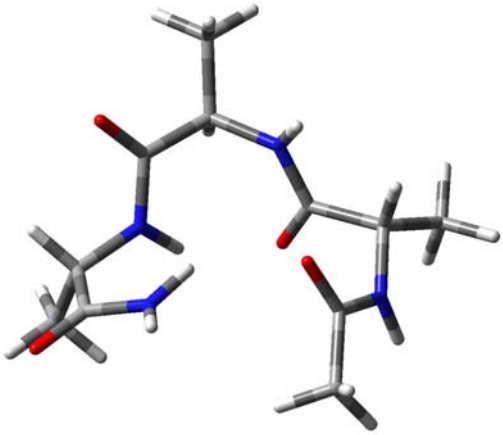


27 D

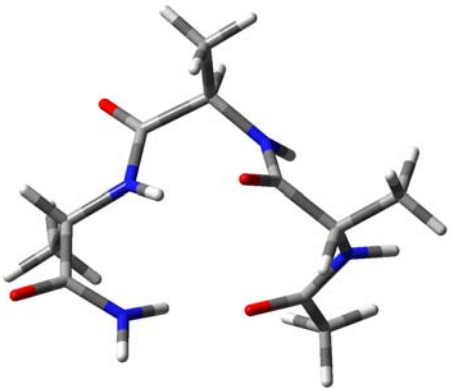
**Chart S1 Continue**



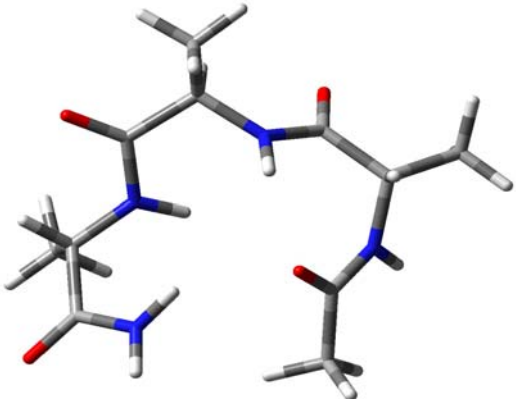
28 D



29 D

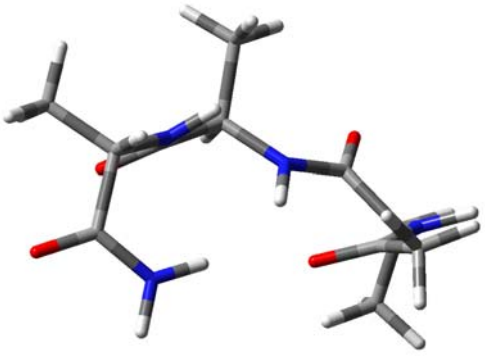


30 D

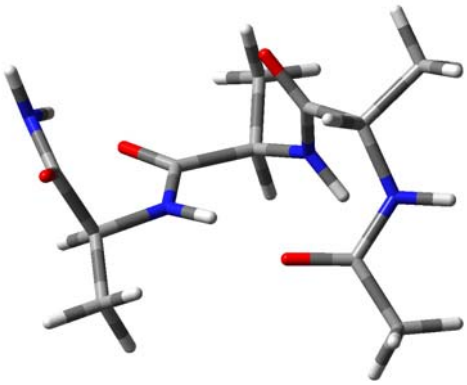


31 D

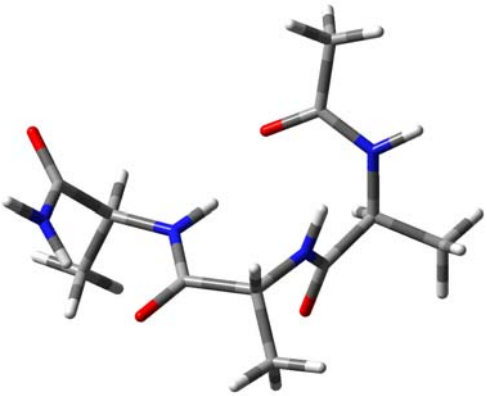
**Chart S1 Continue**



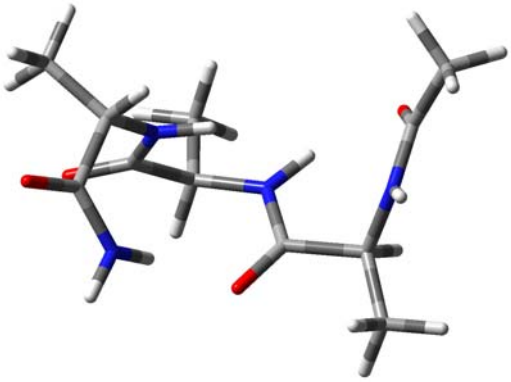
32 D



33 D

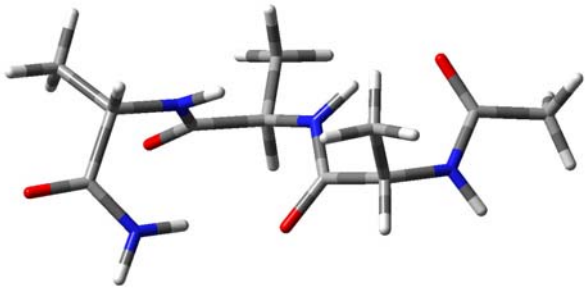


34 D

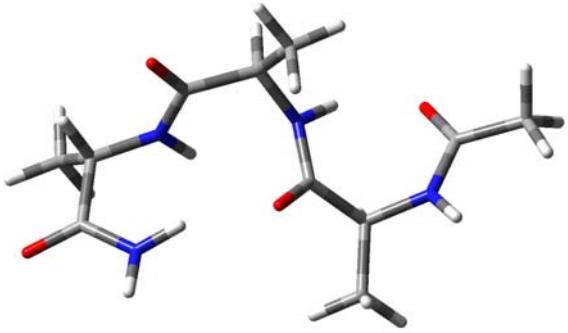


35 D

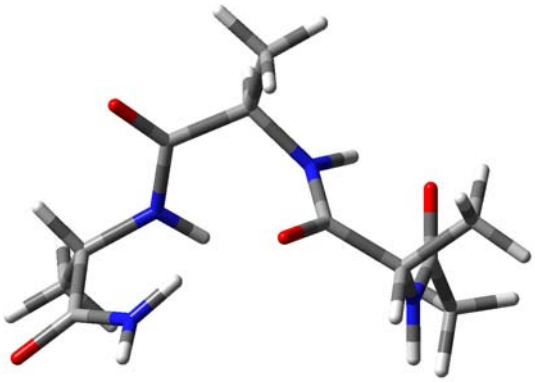
**Chart S1 Continue**



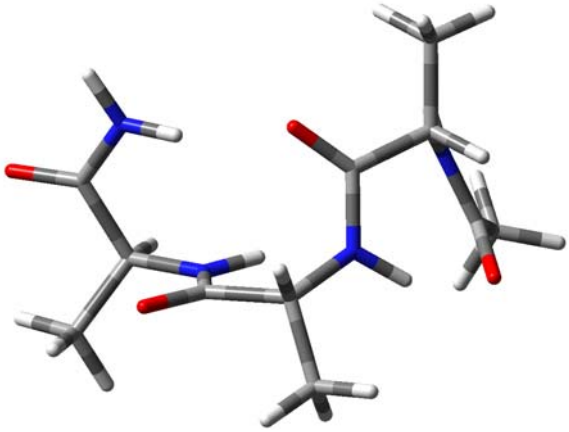
36 D



37 D

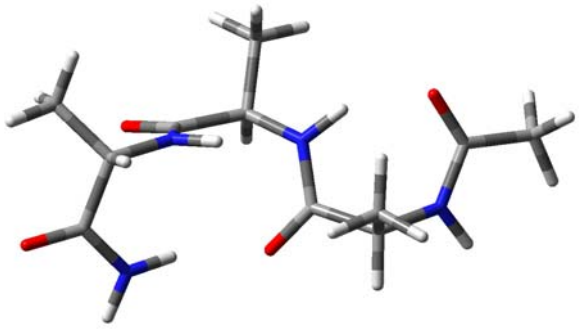


38 D

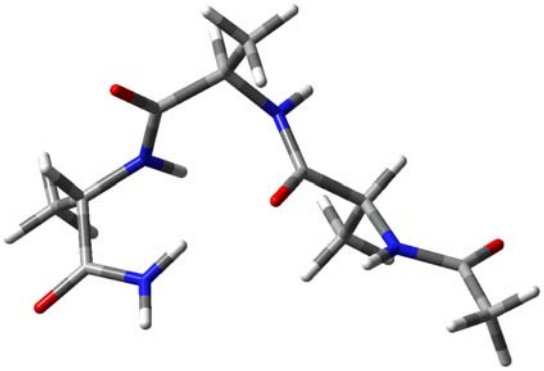


39 D

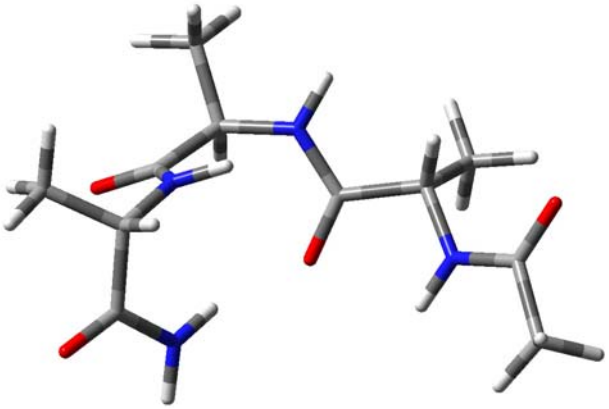
**Chart S1 Continue**



40 D

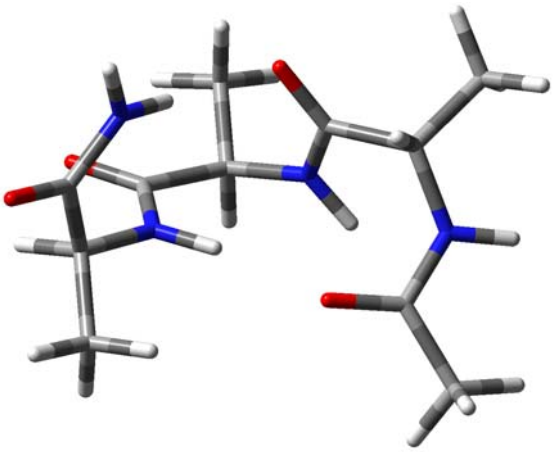


41 D

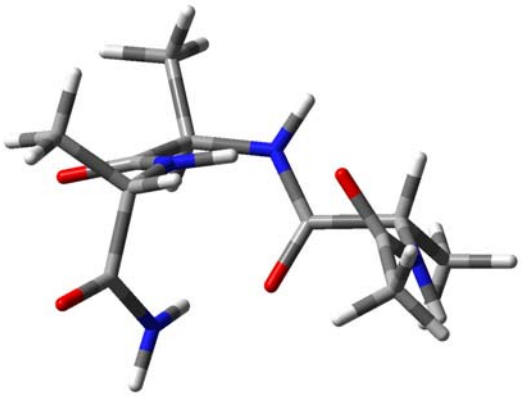


42 D

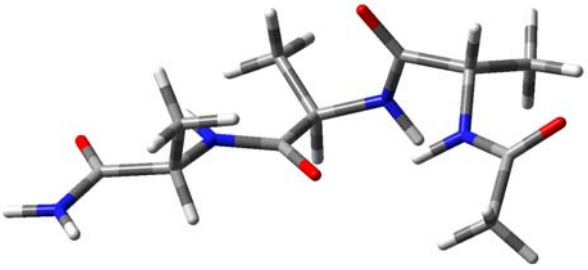
**Chart S1 Continue**



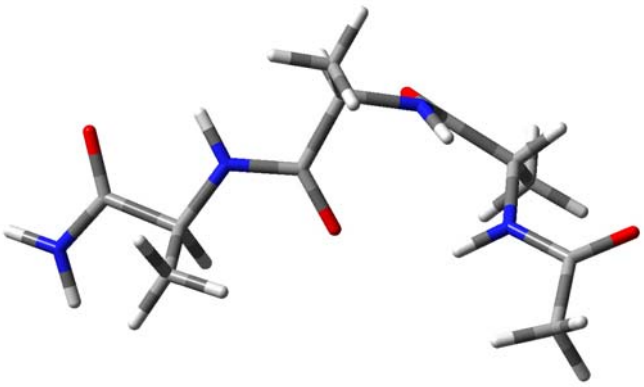
43 D



44 D

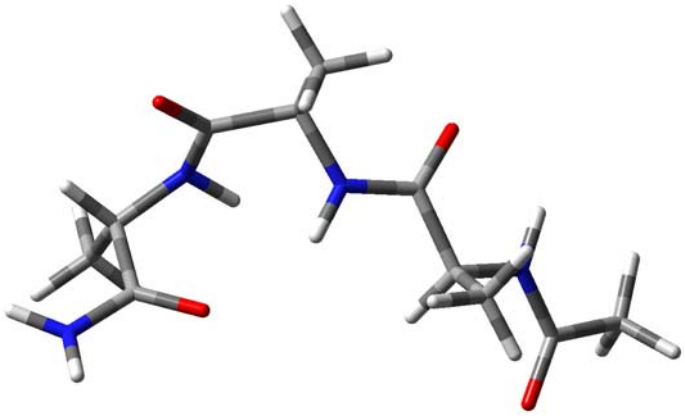


45 D

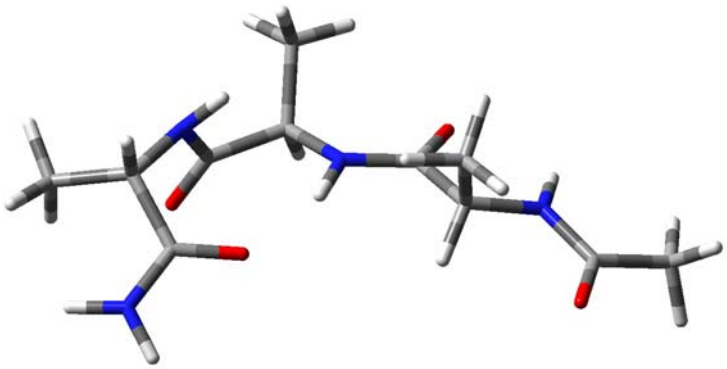


46 D

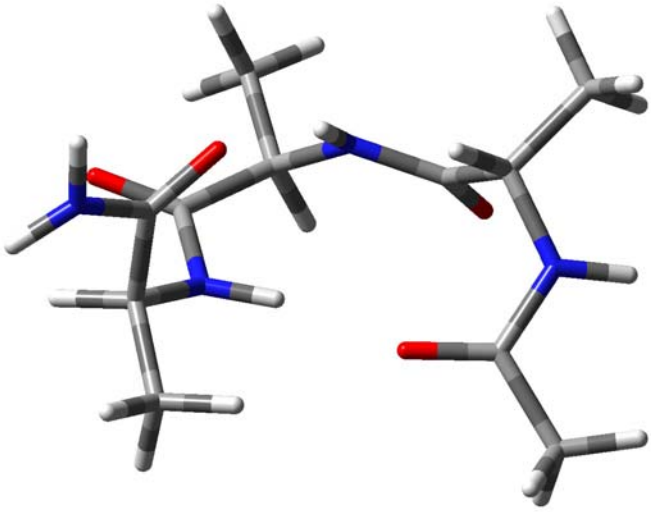
**Chart S1 Continue**



47 D



48 D



49 D

## References


1. Marqusee, S., V.H. Robbins, and R.L. Baldwin, *Unusually stable helix formation in short alanine-based peptides.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(14): p. 5286-90.
2. Miller, J.S., R.J. Kennedy, and D.S. Kemp, *Solubilized, Spaced Polyalanines: A Context-Free System for Determining Amino Acid alpha -Helix Propensities.* Journal of the American Chemical Society, 2002. **124**(6): p. 945-962.
3. Job, G.E., et al., *Temperature- and Length-Dependent Energetics of Formation for Polyalanine Helices in Water: Assignment of wAla(n,T) and Temperature-Dependent CD Ellipticity Standards.* Journal of the American Chemical Society, 2006. **128**(25): p. 8227-8233.
4. Wieczorek, R. and J.J. Dannenberg, *Comparison of Fully Optimized alpha - and $3_{10}$-Helices with Extended beta -Strands. An ONIOM Density Functional Theory Study.* Journal of the American Chemical Society, 2004. **126**(43): p. 14198-14205.
5. Wieczorek, R. and J.J. Dannenberg, *Enthalpies of hydrogen-bonds in alpha -helical peptides. An ONIOM DFT/AM1 study.* Journal of the American Chemical Society, 2005. **127**(42): p. 14534-14535.
6. Salvador, P., A. Asensio, and J.J. Dannenberg, *The Effect of Aqueous Solvation upon alpha -Helix Formation for Polyalanines.* J. Phys. Chem. B, 2007. **111**(25): p. 7462-7466.
7. Anil, B., et al., *Exploiting the right side of the Ramachandran plot: Substitution of glycines by D-alanine can significantly increase protein stability.* Journal of the American Chemical Society, 2004. **126**(41): p. 13194-13195.
8. Tsai, M., Y. Xu, and J.J. Dannenberg, *Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProPro$^{D}$Ala, and ProPro$^{D}$Ser Triads.* Journal of the American Chemical Society, 2005. **127**(41): p. 14130-14131.
9. Shi, Z., et al., *Polyproline II structure in a sequence of seven alanine residues.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(14): p. 9190-9195.
10. Shi, Z., R.W. Woody, and N.R. Kallenbach, *Is polyproline II a major backbone conformation in unfolded proteins?* Advances in Protein Chemistry, 2002. **62**(Unfolded Proteins): p. 163-240, 1 plate.
11. Kim, Y.S., J. Wang, and R.M. Hochstrasser, *Two-Dimensional Infrared Spectroscopy of the Alanine Dipeptide in Aqueous Solution.* J. Phys. Chem. B, 2005. **109**(15): p. 7511-7521.
12. Kentsis, A., et al., *Unfolded state of polyalanine is a segmented polyproline II helix.* Proteins Structure, Function, and Bioinformatics, 2004. **55**(3): p. 493-501.
13. Mezei, M., et al., *Polyproline II helix is the preferred conformation for unfolded polyalanine in water.* Proteins Structure, Function, and Bioinformatics, 2004. **55**(3): p. 502-507.
14. Makowska, J., et al., *Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(6): p. 1744-1749.

15. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* Journal of Molecular Biology, 1963. **7**(1): p. 95-9.
16. Ramachandran, G.N. and V. Sasisekharan, *Conformation of polypeptides and proteins.* Advances in Protein Chemistry, 1968. **23**: p. 283-438.
17. Perczel, A., O. Farkas, and I.G. Csizmadia, *Peptide Models. 18. Hydroxymethyl Side-Chain Induced Backbone Conformational Shifts of L-Serine Amide. All ab Initio Conformers of For-L-Ser-$NH_2$.* Journal of the American Chemical Society, 1996. **118**(33): p. 7809-7817.
18. Perczel, A., O. Farkas, and I.G. Csizmadia, *Peptide models. XVI. The identification of selected HCO-L-Ser-$NH_2$ conformers via a systematic grid search using ab initio potential energy surfaces.* Journal of Computational Chemistry, 1996. **17**(7): p. 821-34.
19. Perczel, A., et al., *Peptide models. XIV. Ab initio study on the role of side-chain backbone interaction stabilizing the building unit of right- and left-handed helixes in peptides and proteins.* International Journal of Quantum Chemistry, 1997. **61**(5): p. 797-814.
20. Iwaoka, M., M. Okada, and S. Tomoda, *Solvent effects on the j-y potential surfaces of glycine and alanine dipeptides studied by PCM and I-PCM methods.* Theochem, 2002. **586**: p. 111-124.
21. Grant, J.A., R.L. Williams, and H.A. Scheraga, *Ab initio self-consistent field and potential-dependent partial equalization of orbital electronegativity calculations of hydration properties of N-acetyl-N'-methyl-alanineamide.* Biopolymers, 1990. **30**(9-10): p. 929-49.
22. Asciutto, E.K., et al., *Computational and Experimental Determination of the alpha -Helix Unfolding Reaction Coordinate.* Biochemistry, 2008. **47**(7): p. 2046-2050.
23. Chin, W., et al., *Spectroscopic Evidence for Gas-Phase Formation of Successive beta -Turns in a Three-Residue Peptide Chain.* Journal of the American Chemical Society, 2005. **127**(5): p. 1388-1389.
24. Chin, W., et al., *Gas Phase Formation of a $3_{10}$-Helix in a Three-Residue Peptide Chain: Role of Side Chain-Backbone Interactions as Evidenced by IR-UV Double Resonance Experiments.* Journal of the American Chemical Society, 2005. **127**(34): p. 11900-11901.
25. Chin, W., et al., *Probing the competition between secondary structures and local preferences in gas phase isolated peptide backbones.* Physical Chemistry Chemical Physics, 2006. **8**(9): p. 1033-1048.
26. Eker, F., et al., *Tripeptides Adopt Stable Structures in Water. A Combined Polarized Visible Raman, FTIR, and VCD Spectroscopy Study.* Journal of the American Chemical Society, 2002. **124**(48): p. 14330-14341.
27. Horvath, V., Z. Varga, and A. Kovacs, *Long-Range Effects in Oligopeptides. A Theoretical Study of the beta -Sheet Structure of Glyn (n = 2-10).* Journal of Physical Chemistry A, 2004. **108**(33): p. 6869-6873.
28. Wieczorek, R. and J.J. Dannenberg, *The Energetic and Structural Effects of Single Amino Acid Substitutions upon Capped alpha -Helical Peptides Containing 17 Amino Acid Residues. An ONIOM DFT/AM1 Study.* Journal of the American Chemical Society, 2005. **127**(49): p. 17216-17223.
29. Shang, H.S. and T. Head-Gordon, *Stabilization of Helixes in Glycine and Alanine Dipeptides in a Reaction Field Model of Solvent.* Journal of the American Chemical Society, 1994. **116**(4): p. 1528-32.

30. Van Duijneveldt, F.B., *Basis set superposition error.* Molecular Interactions, 1997: p. 81-104.
31. Jansen, H.B. and P. Ros, *Nonempirical molecular orbital calculations on the protonation of carbon monoxide.* Chemical Physics Letters, 1969. **3**(3): p. 140-3.
32. Boys, S.F. and F. Bernardi, *The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors.* Molecular Physics, 1970. **19**(4): p. 553-566.
33. Simon, S., M. Duran, and J.J. Dannenberg, *How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers?* Journal of Chemical Physics, 1996. **105**(24): p. 11024-11031.
34. Tobias, D.J. and C.L. Brooks, III, *Thermodynamics and mechanism of alpha helix initiation in alanine and valine peptides.* Biochemistry, 1991. **30**(24): p. 6059-70.
35. Vila, J.A., et al., *Fast and accurate computation of the 13C chemical shifts for an alanine-rich peptide.* Proteins Structure, Function, and Bioinformatics, 2004. **57**(1): p. 87-98.
36. Rablen, P.R., J.W. Lockman, and W.L. Jorgensen, *Ab Initio Study of Hydrogen-Bonded Complexes of Small Organic Molecules with Water.* Journal of Physical Chemistry A, 1998. **102**(21): p. 3782-3797.
37. Dannenberg, J.J., *Enthalpies of Hydration of N-Methylacetamide by One, Two, and Three Waters and the Effect upon the C:O Stretching Frequency. An Ab Initio DFT Study.* Journal of Physical Chemistry A, 2006. **110**(17): p. 5798-5802.
38. Avbelj, F. and R.L. Baldwin, *Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: Distributions of phi.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(10): p. 5742-5747.
39. Avbelj, F. and R.L. Baldwin, *Limited validity of group additivity for the folding energetics of the peptide group.* Proteins Structure, Function, and Bioinformatics, 2006. **63**(2): p. 283-289.
40. Mons, M., et al., *Energetics of the Gas Phase Hydrates of trans-Formanilide: A Microscopic Approach to the Hydration Sites of the Peptide Bond.* Journal of Physical Chemistry A, 2001. **105**(6): p. 969-973.
41. Dannenberg, J.J., *The importance of cooperative interactions and a solid-state paradigm to proteins: what peptide chemists can learn from molecular crystals.* Advances in Protein Chemistry, 2006. **72**(Peptide Solvation and H-Bonds): p. 227-273.
42. Ingwall, R.T., et al., *Conformational studies of poly-L-alanine in water.* Biopolymers, 1968. **6**(3): p. 331-68.
43. Wieczorek, R. and J.J. Dannenberg, *alpha -Helical Peptides Are Not Protonated at the N-Terminus in the Gas Phase.* Journal of the American Chemical Society, 2004. **126**(39): p. 12278-12279.
44. Horng, J.-C., F.W. Kotch, and R.T. Raines, *Is glycine a surrogate for a D-amino acid in the collagen triple helix?* Protein Science, 2007. **16**(2): p. 208-215.

# Bibliography

## References for Chapter 1.


1. Nimni, M.E. and Editor, *Collagen: Biochemistry, Biomechanics, Biotechnology. A 3-Volume Set*. 1988. 1088 pp.
2. Kielty, C.M., I. Hopkinson, and M.E. Grant, *Collagen: the collagen family: structure, assembly, and organization in the extracellular matrix.* 1993: p. 103-47.
3. Bhattacharjee, A. and M. Bansal, *Collagen structure: The Madras triple helix and the current scenario.* IUBMB Life, 2005. **57**(3): p. 161-172.
4. Cowan, P.M. and S. McGavin, *The structure of poly-L-proline.* Nature (London, U. K.), 1955. **176**: p. 501-3.
5. Rich, A. and F.H. Crick, *The molecular structure of collagen.* Journal of molecular biology, 1961. **3**: p. 483-506.
6. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* Journal of Molecular Biology, 1963. **7**(1): p. 95-9.
7. Bella, J., et al., *Crystal and molecular structure of a collagen-like peptide at 1.9 .ANG. resolution.* Science (Washington, DC, United States), 1994. **266**(5182): p. 75-81.
8. Sasisekharan, V., *Structure of poly-L-proline II.* Acta Crystallographica, 1959. **12**: p. 897-903.
9. Ramachandran, G.N. and V. Sasisekharan, *Conformation of polypeptides and proteins.* Advances in Protein Chemistry, 1968. **23**: p. 283-438.
10. Jones, E.Y. and A. Miller, *Analysis of structural design features in collagen.* Journal of Molecular Biology, 1991. **218**(1): p. 209-19.
11. Woodhead-Galloway, J., *Collagen: the anatomy of a protein.* Institute of Biology's Studies in Biology, 1980. **117**: p. 64 pp.
12. Chopra, R.K. and V.S. Ananthanarayanan, *Conformational implications of enzymic proline hydroxylation in collagen.* Proceedings of the National Academy of Sciences of the United States of America, 1982. **79**(23): p. 7180-4.
13. Privalov, P.L., *Stability of proteins. Proteins which do not present a single cooperative system.* Advances in Protein Chemistry, 1982. **35**: p. 1-104.
14. Berg, R.A. and D.J. Prockop, *Thermal transition of a nonhydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple helix of collagen.* Biochemical and Biophysical Research Communications, 1973.

**52**(1): p. 115-20.

15. Bella, J., et al., *Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution.* Science (New York, N.Y.), 1994. **266**(5182): p. 75-81.
16. Bella, J., B. Brodsky, and H.M. Berman, *Hydration structure of a collagen peptide.* Structure (London), 1995. **3**(9): p. 893-906.
17. Vitagliano, L., et al., *Preferred proline puckerings in cis and trans peptide groups: implications for collagen stability.* Protein Science, 2001. **10**(12): p. 2627-2632.
18. Vitagliano, L., et al., *Structural bases of collagen stabilization induced by proline hydroxylation. [Erratum to document cited in CA135:15652].* Biopolymers, 2001. **59**(6): p. 465.
19. Berisio, R., et al., *Crystal structure of the collagen triple helix model $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$.* Protein Science, 2002. **11**(2): p. 262-270.
20. Stock, L.M., *Origin of the inductive effect.* Journal of Chemical Education, 1972. **49**(6): p. 400-4.
21. Raines, R.T., *2005 Emil Thomas Kaiser Award.* Protein Science, 2006. **15**(5): p. 1219-1225.
22. Bretscher, L.E., et al., *Conformational Stability of Collagen Relies on a Stereoelectronic Effect.* Journal of the American Chemical Society, 2001. **123**(4): p. 777-778.
23. Schumacher, M., K. Mizuno, and H.P. Baechinger, *The Crystal Structure of the Collagen-like Polypeptide $(Glycyl\text{-}4(R)\text{-}hydroxyprolyl\text{-}4(R)\text{-}hydroxyprolyl)_9$ at 1.55 A Resolution Shows Up-puckering of the Proline Ring in the Xaa Position.* Journal of Biological Chemistry, 2005. **280**(21): p. 20397-20403.
24. Doi, M., et al., *Collagen-like triple helix formation of synthetic $(Pro\text{-}Pro\text{-}Gly)_{10}$ analogs: $(4(S)\text{-}hydroxyprolyl\text{-}4(R)\text{-}hydroxyprolyl\text{-}Gly)_{10}$, $(4(R)\text{-}hydroxyprolyl\text{-}4(R)\text{-}hydroxyprolyl\text{-}Gly)_{10}$ and $(4(S)\text{-}fluoroprolyl\text{-}4(R)\text{-}fluoroprolyl\text{-}Gly)_{10}$.* Journal of Peptide Science, 2005. **11**(10): p. 609-616.
25. Persikov, A.V., et al., *Amino acid propensities for the collagen triple-helix.* Biochemistry, 2000. **39**(48): p. 14960-14967.
26. Improta, R., C. Benzi, and V. Barone, *Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution.* Journal of the American Chemical Society, 2001. **123**(50): p. 12568-12577.
27. Improta, R., et al., *Understanding the role of stereoelectronic effects in determining collagen stability. 2. A quantum mechanical/molecular*

*mechanical study of (proline-proline-glycine)n polypeptides.* Journal of the American Chemical Society, 2002. **124**(26): p. 7857-7865.

28. Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, 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, K.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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; 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 03*, 2003.
29. Plumley, J.A. and J.J. Dannenberg, *The Importance of Hydrogen Bonding between the Glutamine Side Chains to the Formation of Amyloid VQIVYK Parallel β-Sheets: An ONIOM DFT/AM1 Study.* Journal of the American Chemical Society. **132**(6): p. 1758-1759.
30. Viswanathan, R. and J.J. Dannenberg, *A density functional theory study of vibrational coupling in the amide I band of β-sheet models.* J. Phys. Chem. B, 2008. **112**(16): p. 5199-5208.
31. Chen, Y.-F. and J.J. Dannenberg, *Cooperative 4-pyridone H-bonds with extraordinary stability. A DFT molecular orbital study.* Journal of the American Chemical Society, 2006. **128**(25): p. 8100-8101.
32. Wieczorek, R. and J.J. Dannenberg, *Comparison of Fully Optimized alpha - and $3_{10}$-Helices with Extended beta -Strands. An ONIOM Density Functional Theory Study.* Journal of the American Chemical Society, 2004. **126**(43): p. 14198-14205.
33. Tsai, M., Y. Xu, and J.J. Dannenberg, *Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProProDAla, and ProProDSer Triads.* Journal of the American Chemical Society, 2005. **127**(41): p. 14130-14131.
34. Kobko, N. and J.J. Dannenberg, *Cooperativity in Amide Hydrogen Bonding Chains. Relation between Energy, Position, and H-Bond Chain Length in Peptide and Protein Folding Models.* Journal of Physical Chemistry A, 2003.

**107**(48): p. 10389-10395.

35. Kobko, N., et al., *Cooperativity in Amide Hydrogen Bonding Chains: Implications for Protein-Folding Models.* Journal of the American Chemical Society, 2001. **123**(18): p. 4348-4349.
36. Simon, S., M. Duran, and J.J. Dannenberg, *Effect of Basis Set Superposition Error on the Water Dimer Surface Calculated at Hartree-Fock, Moller-Plesset, and Density Functional Theory Levels.* Journal of Physical Chemistry A, 1999. **103**(11): p. 1640-1643.
37. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange.* Journal of Chemical Physics, 1993. **98**(7): p. 5648-52.
38. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B Condensed Matter and Materials Physics, 1988. **37**(2): p. 785-9.
39. Byers, P.H., *Osteogenesis imperfecta.* 1993: p. 317-50.
40. Marini, J.C., et al., *Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans.* Human Mutation, 2007. **28**(3): p. 209-221.
41. Morokuma, K., *New challenges in quantum chemistry: quests for accurate calculations for large molecular systems.* Philosophical Transactions of the Royal Society of London, Series A Mathematical, Physical and Engineering Sciences, 2002. **360**(1795): p. 1149-1164.
42. Dapprich, S., et al., *A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives.* Theochem, 1999. **461-462**: p. 1-21.
43. Cramer, C.J. and D.G. Truhlar, *Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics.* Chemical Reviews (Washington, D. C.), 1999. **99**(8): p. 2161-2200.
44. Lotz, B., *Crystal structure of polyglycine I.* Comptes Rendus des Seances de l'Academie des Sciences, Serie C Sciences Chimiques, 1972. **274**(23): p. 1907-10.
45. Colonna-Cesari, F., S. Premilat, and B. Lotz, *Structure of polyglycine I: a comparison of the antiparallel pleated and antiparallel rippled sheets.* Journal of Molecular Biology, 1974. **87**(2): p. 181-91.
46. Lotz, B., *Crystal structure of polyglycine I.* Journal of Molecular Biology, 1974. **87**(2): p. 169-80.
47. Munoz-Guerra, S., et al., *Crystals of polyglycine in the beta form.* Journal of Molecular Biology, 1983. **167**(1): p. 223-5.

48. Kajava, A.V., *Dimorphism of polyglycine I: structural models for crystal modifications.* Acta crystallographica. Section D, Biological crystallography, 1999. **55**(Pt 2): p. 436-42.
49. Crick, F.H.C. and A. Rich, *Structure of polyglycine II.* Nature (London, U. K.), 1955. **176**: p. 780-1.
50. Ramachandran, G.N., C. Ramakrishnan, and C.M. Venkatachalam, *Structure of polyglycine II with direct and inverted chains.* Conformation of Biopolymers, Papers read at an International Symposium, 1967. **2**: p. 429-38.
51. Simon, S., M. Duran, and J.J. Dannenberg, *How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers?* Journal of Chemical Physics, 1996. **105**(24): p. 11024-11031.
52. Wodrich, M.D., et al., *How accurate are DFT treatments of organic energies?* Organic Letters, 2007. **9**(10): p. 1851-1854.
53. Improta, R., et al., *The conformational behavior of polyglycine as predicted by a density functional model with periodic boundary conditions.* Journal of Chemical Physics, 2001. **114**(6): p. 2541-2549.
54. Dewar, M.J., MS. Mod. Technol. Comput. Chem. Motecc, 1991.
55. Dewar, M.J.S., et al., *Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model.* Journal of the American Chemical Society, 1985. **107**(13): p. 3902-9.
56. Jansen, H.B. and P. Ros, *Nonempirical molecular orbital calculations on the protonation of carbon monoxide.* Chemical Physics Letters, 1969. **3**(3): p. 140-3.
57. Boys, S.F. and F. Bernardi, *The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors.* Molecular Physics, 1970. **19**(4): p. 553-566.

## References for Chapter 2.

1. Bhattacharjee, A. and M. Bansal, *Collagen structure: The Madras triple helix and the current scenario.* IUBMB Life, 2005. **57**(3): p. 161-172.
2. Rich, A. and F.H. Crick, *The molecular structure of collagen.* Journal of molecular biology, 1961. **3**: p. 483-506.
3. Bella, J., et al., *Crystal and molecular structure of a collagen-like peptide at 1.9* Å *resolution.* Science (New York, N.Y.), 1994. **266**(5182): p. 75-81.
4. Chan, D., M.S.P. Ho, and K.S.E. Cheah, *Aberrant signal peptide cleavage of collagen X in Schmid metaphyseal chondrodysplasia. Implications for the molecular basis of the disease.* Journal of Biological Chemistry, 2001. **276**(11): p. 7992-7997.

5. Persikov, A.V., et al., *Stability related bias in residues replacing glycines within the collagen triple helix (Gly-Xaa-Yaa) in inherited connective tissue disorders.* Human Mutation, 2004. **24**(4): p. 330-337.
6. Baum, J. and B. Brodsky, *Folding of peptide models of collagen and misfolding in disease.* Current Opinion in Structural Biology, 1999. **9**(1): p. 122-128.
7. Morokuma, K., *New challenges in quantum chemistry: quests for accurate calculations for large molecular systems.* Philosophical Transactions of the Royal Society of London, Series A Mathematical, Physical and Engineering Sciences, 2002. **360**(1795): p. 1149-1164.
8. Dapprich, S., et al., *A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives.* Theochem, 1999. **461-462**: p. 1-21.
9. Frisch, M.J., et al. , , *GAUSSIAN 03.* Gaussian, Inc.: Pittsburgh PA,, 2003.
10. Wieczorek, R. and J.J. Dannenberg, *Comparison of Fully Optimized alpha - and 310-Helices with Extended beta -Strands. An ONIOM Density Functional Theory Study.* Journal of the American Chemical Society, 2004. **126**(43): p. 14198-14205.
11. Wieczorek, R. and J.J. Dannenberg, *The Energetic and Structural Effects of Single Amino Acid Substitutions upon Capped alpha -Helical Peptides Containing 17 Amino Acid Residues. An ONIOM DFT/AM1 Study.* Journal of the American Chemical Society, 2005. **127**(49): p. 17216-17223.
12. Vitagliano, L., et al., *Structural bases of collagen stabilization induced by proline hydroxylation.* Biopolymers, 2001. **58**(5): p. 459-464.
13. Chopra, R.K. and V.S. Ananthanarayanan, *Conformational implications of enzymic proline hydroxylation in collagen.* Proceedings of the National Academy of Sciences of the United States of America, 1982. **79**(23): p. 7180-4.
14. Privalov, P.L., *Stability of proteins. Proteins which do not present a single cooperative system.* Advances in Protein Chemistry, 1982. **35**: p. 1-104.
15. Berg, R.A. and D.J. Prockop, *Thermal transition of a nonhydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple helix of collagen.* Biochemical and Biophysical Research Communications, 1973. **52**(1): p. 115-20.
16. Ramachandran, G.N., M. Bansal, and R.S. Bhatnager, *Hypothesis on the role of hydroxyproline in stabilizing collagen structure.* Biochim. Biophys. Acta, Protein Struct., 1973. **322**(1): p. 166-71.
17. Suzuki, E.F., R. D. B.; MacRae, T. P., Int. J. Biol. Macromol., 1980. **2**: p.

54-56.

18. Schumacher, M., K. Mizuno, and H.P. Baechinger, *The Crystal Structure of the Collagen-like Polypeptide (Glycyl-4(R)-hydroxyprolyl-4(R)-hydroxyprolyl)$_9$ at 1.55* Å *Resolution Shows Up-puckering of the Proline Ring in the Xaa Position.* Journal of Biological Chemistry, 2005. **280**(21): p. 20397-20403.
19. Doi, M., et al., *Collagen-like triple helix formation of synthetic (Pro-Pro-Gly)$_{10}$ analogs: (4(S)-hydroxyprolyl-4(R)-hydroxyprolyl-Gly)$_{10}$, (4(R)-hydroxyprolyl-4(R)-hydroxyprolyl-Gly)$_{10}$ and (4(S)-fluoroprolyl-4(R)-fluoroprolyl-Gly)$_{10}$.* Journal of Peptide Science, 2005. **11**(10): p. 609-616.
20. Persikov, A.V., et al., *Amino acid propensities for the collagen triple-helix.* Biochemistry, 2000. **39**(48): p. 14960-14967.
21. Improta, R., C. Benzi, and V. Barone, *Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution.* Journal of the American Chemical Society, 2001. **123**(50): p. 12568-12577.
22. Vitagliano, L., et al., *Preferred proline puckerings in cis and trans peptide groups: implications for collagen stability.* Protein Science, 2001. **10**(12): p. 2627-2632.
23. Raines, R.T., *2005 Emil Thomas Kaiser Award.* Protein Science, 2006. **15**(5): p. 1219-1225.
24. Improta, R., et al., *Understanding the role of stereoelectronic effects in determining collagen stability. 2. A quantum mechanical/molecular mechanical study of (proline-proline-glycine)n polypeptides.* Journal of the American Chemical Society, 2002. **124**(26): p. 7857-7865.
25. Hodges, J.A. and R.T. Raines, *Stereoelectronic Effects on Collagen Stability: The Dichotomy of 4-Fluoroproline Diastereomers.* Journal of the American Chemical Society, 2003. **125**(31): p. 9262-9263.
26. Horng, J.-C. and R.T. Raines, *Stereoelectronic effects on polyproline conformation.* Protein Science, 2006. **15**(1): p. 74-83.
27. Kotch, F.W., I.A. Guzei, and R.T. Raines, *Stabilization of the Collagen Triple Helix by O-Methylation of Hydroxyproline Residues.* Journal of the American Chemical Society, 2008. **130**(10): p. 2952-2953.
28. Bretscher, L.E., et al., *Conformational Stability of Collagen Relies on a Stereoelectronic Effect.* Journal of the American Chemical Society, 2001. **123**(4): p. 777-778.
29. Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.;

Cheeseman, J. R.; Montgomery, J. A.; Vreven, 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, K.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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; 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 03*, 2003.

30. Frisch, M.J., et al. , , *GAUSSIAN 09, Revision A.1*. Gaussian, Inc.: Pittsburgh PA,, 2009.
31. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange.* Journal of Chemical Physics, 1993. **98**(7): p. 5648-52.
32. Anil, B., et al., *Exploiting the right side of the Ramachandran plot: Substitution of glycines by D-alanine can significantly increase protein stability.* Journal of the American Chemical Society, 2004. **126**(41): p. 13194-13195.
33. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B Condensed Matter and Materials Physics, 1988. **37**(2): p. 785-9.
34. Cramer, C.J. and D.G. Truhlar, *Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics.* Chemical Reviews (Washington, D. C.), 1999. **99**(8): p. 2161-2200.
35. AMAPC, Semichem, Inc.
36. Dewar, M.J.S., et al., *Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model.* Journal of the American Chemical Society, 1985. **107**(13): p. 3902-9.
37. Mizuno, K., et al., *The Peptides Acetyl-(Gly-3(S)Hyp-4(R)Hyp)$_{10}$-NH$_2$ and Acetyl-(Gly-Pro-3(S)Hyp)$_{10}$-NH$_2$ Do Not Form a Collagen Triple Helix.* Journal of Biological Chemistry, 2004. **279**(1): p. 282-287.
38. Holmgren, S.K., et al., *Code for collagen's stability deciphered.* Nature (London), 1998. **392**(6677): p. 666-667.

## References for Chapter 3.


1. Bella, J., et al., *Crystal and molecular structure of a collagen-like peptide at 1.9 .ANG. resolution.* Science (Washington, DC, United States), 1994. **266**(5182): p. 75-81.
2. Kramer, R.Z., et al., *Sequence dependent conformational variations of collagen triple-helical structure.* Nat. Struct. Biol., 1999. **6**(5): p. 454-457.
3. Vitagliano, L., et al., *Structural bases of collagen stabilization induced by proline hydroxylation.* Biopolymers, 2001. **58**(5): p. 459-464.
4. Improta, R., C. Benzi, and V. Barone, *Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution.* Journal of the American Chemical Society, 2001. **123**(50): p. 12568-12577.
5. Hodges, J.A. and R.T. Raines, *Stereoelectronic Effects on Collagen Stability: The Dichotomy of 4-Fluoroproline Diastereomers.* Journal of the American Chemical Society, 2003. **125**(31): p. 9262-9263.
6. Horng, J.-C. and R.T. Raines, *Stereoelectronic effects on polyproline conformation.* Protein Science, 2006. **15**(1): p. 74-83.
7. Persikov, A.V., et al., *Triple-Helix Propensity of Hydroxyproline and Fluoroproline: Comparison of Host-Guest and Repeating Tripeptide Collagen Models.* Journal of the American Chemical Society, 2003. **125**(38): p. 11500-11501.
8. Malkar, N.B., et al., *Modulation of Triple-Helical Stability and Subsequent Melanoma Cellular Responses by Single-Site Substitution of Fluoroproline Derivatives.* Biochemistry, 2002. **41**(19): p. 6054-6064.
9. Kotch, F.W., I.A. Guzei, and R.T. Raines, *Stabilization of the Collagen Triple Helix by O-Methylation of Hydroxyproline Residues.* Journal of the American Chemical Society, 2008. **130**(10): p. 2952-2953.
10. DeRider, M.L., et al., *Collagen Stability: Insights from NMR Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations.* Journal of the American Chemical Society, 2002. **124**(11): p. 2497-2505.
11. Frisch, M.J., et al. , , *GAUSSIAN 09, Revision A.1.* Gaussian, Inc.: Pittsburgh PA,, 2009.
12. Tsai, M., Y. Xu, and J.J. Dannenberg, *Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProProDAla, and ProProDSer Triads.* Journal of the American Chemical Society, 2005. **127**(41): p. 14130-14131.

13. Dannenberg, J.J., *The importance of cooperative interactions and a solid-state paradigm to proteins: what peptide chemists can learn from molecular crystals.* Advances in Protein Chemistry, 2006. **72**(Peptide Solvation and H-Bonds): p. 227-273.
14. Vitagliano, L., et al., *Stabilization of the triple-helical structure of natural collagen by side-chain interactions.* Biochemistry, 1993. **32**(29): p. 7354-9.
15. Holmgren, S.K., et al., *Code for collagen's stability deciphered.* Nature (London), 1998. **392**(6677): p. 666-667.
16. Holmgren, S.K., et al., *A hyperstable collagen mimic.* Chemistry & Biology, 1999. **6**(2): p. 63-70.
17. Shaw, B.R. and J.M. Schurr, *Association reaction of collagen model polypeptides (Pro-Pro-Gly)n.* Biopolymers, 1975. **14**(9): p. 1951-85.
18. Mizuno, K., et al., *The Peptides Acetyl-(Gly-3(S)Hyp-4(R)Hyp)10-NH2 and Acetyl-(Gly-Pro-3(S)Hyp)10-NH2 Do Not Form a Collagen Triple Helix.* Journal of Biological Chemistry, 2004. **279**(1): p. 282-287.

## References for Chapter 4.

1. Lotz, B., *Crystal structure of polyglycine I.* Comptes Rendus des Seances de l'Academie des Sciences, Serie C Sciences Chimiques, 1972. **274**(23): p. 1907-10.
2. Colonna-Cesari, F., S. Premilat, and B. Lotz, *Structure of polyglycine I: a comparison of the antiparallel pleated and antiparallel rippled sheets.* Journal of Molecular Biology, 1974. **87**(2): p. 181-91.
3. Lotz, B., *Crystal structure of polyglycine I.* Journal of Molecular Biology, 1974. **87**(2): p. 169-80.
4. Munoz-Guerra, S., et al., *Crystals of polyglycine in the beta form.* Journal of Molecular Biology, 1983. **167**(1): p. 223-5.
5. Kajava, A.V., *Dimorphism of polyglycine I: structural models for crystal modifications.* Acta crystallographica. Section D, Biological crystallography, 1999. **55**(Pt 2): p. 436-42.
6. Crick, F.H.C. and A. Rich, *Structure of polyglycine II.* Nature (London, U. K.), 1955. **176**: p. 780-1.
7. Ramachandran, G.N., C. Ramakrishnan, and C.M. Venkatachalam, *Structure of polyglycine II with direct and inverted chains.* Conformation of Biopolymers, Papers read at an International Symposium, 1967. **2**: p. 429-38.
8. Improta, R., et al., *The conformational behavior of polyglycine as predicted by a density functional model with periodic boundary conditions.* Journal of Chemical Physics, 2001. **114**(6): p. 2541-2549.
9. Zhao, Y.-L. and Y.-D. Wu, *A Theoretical Study of β-Sheet Models: Is the*

*Formation of Hydrogen-Bond Networks Cooperative?* Journal of the American Chemical Society, 2002. **124**(8): p. 1570-1571.

10. Viswanathan, R., A. Asensio, and J.J. Dannenberg, *Cooperative Hydrogen-Bonding in Models of Antiparallel β-Sheets.* Journal of Physical Chemistry A, 2004. **108**(42): p. 9205-9212.
11. Kleier, D.A. and W.N. Lipscomb, *Molecular orbital study of polypeptides. Conformational and electronic structure of polyglycine.* International Journal of Quantum Chemistry, Quantum Biology Symposium, 1977. **4**(Proc. Int. Symp. Quantum Biol. Quantum Pharmacol., 4th): p. 73-86.
12. Fleck, O., M. Seel, and J. Ladik, *Calculation of the band structure of polyglycine as a two-dimensional periodic system.* Solid State Communications, 1988. **65**(7): p. 701-4.
13. Ladik, J., A. Sutjianto, and P. Otto, *Improved band structures of some homopolypeptides with aliphatic side chains and of the four nucleotide base stacks: estimation of their fundamental gap.* Journal of Molecular Structure THEOCHEM, 1991. **74**: p. 271-6.
14. Suhai, S., *The electronic structure of parallel β-pleated sheets in proteins: an ab initio computation including electron correlation.* International Journal of Quantum Chemistry, 1991. **40**(4): p. 559-76.
15. Wieczorek, R. and J.J. Dannenberg, *Comparison of Fully Optimized α- and $3_{10}$-Helices with Extended β-Strands. An ONIOM Density Functional Theory Study.* Journal of the American Chemical Society, 2004. **126**(43): p. 14198-14205.
16. Horvath, V., Z. Varga, and A. Kovacs, *Long-Range Effects in Oligopeptides. A Theoretical Study of the β-Sheet Structure of Glyn (n = 2-10).* Journal of Physical Chemistry A, 2004. **108**(33): p. 6869-6873.
17. Palfi, V.K. and A. Perczel, *How stable is a collagen triple helix? An ab initio study on various collagen and β-sheet forming sequences.* Journal of Computational Chemistry, 2008. **29**(9): p. 1374-1386.
18. Rich, A. and F.H. Crick, *The molecular structure of collagen.* Journal of Molecular Biology, 1961. **3**: p. 483-506.
19. Tsai, M., Y. Xu, and J.J. Dannenberg, *Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProPro$^{D}$Ala, and ProPro$^{D}$Ser Triads.* Journal of the American Chemical Society, 2005. **127**(41): p. 14130-14131.

## References for Chapter 5.

1. Marqusee, S., V.H. Robbins, and R.L. Baldwin, *Unusually stable helix formation in short alanine-based peptides.* Proceedings of the National

Academy of Sciences of the United States of America, 1989. **86**(14): p. 5286-90.

2. Miller, J.S., R.J. Kennedy, and D.S. Kemp, *Solubilized, Spaced Polyalanines: A Context-Free System for Determining Amino Acid alpha -Helix Propensities.* Journal of the American Chemical Society, 2002. **124**(6): p. 945-962.
3. Job, G.E., et al., *Temperature- and Length-Dependent Energetics of Formation for Polyalanine Helices in Water: Assignment of wAla(n,T) and Temperature-Dependent CD Ellipticity Standards.* Journal of the American Chemical Society, 2006. **128**(25): p. 8227-8233.
4. Wieczorek, R. and J.J. Dannenberg, *Comparison of Fully Optimized alpha - and $3_{10}$-Helices with Extended beta -Strands. An ONIOM Density Functional Theory Study.* Journal of the American Chemical Society, 2004. **126**(43): p. 14198-14205.
5. Wieczorek, R. and J.J. Dannenberg, *Enthalpies of hydrogen-bonds in alpha -helical peptides. An ONIOM DFT/AM1 study.* Journal of the American Chemical Society, 2005. **127**(42): p. 14534-14535.
6. Salvador, P., A. Asensio, and J.J. Dannenberg, *The Effect of Aqueous Solvation upon alpha -Helix Formation for Polyalanines.* J. Phys. Chem. B, 2007. **111**(25): p. 7462-7466.
7. Anil, B., et al., *Exploiting the right side of the Ramachandran plot: Substitution of glycines by D-alanine can significantly increase protein stability.* Journal of the American Chemical Society, 2004. **126**(41): p. 13194-13195.
8. Tsai, M., Y. Xu, and J.J. Dannenberg, *Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProPro$^{D}$Ala, and ProPro$^{D}$Ser Triads.* Journal of the American Chemical Society, 2005. **127**(41): p. 14130-14131.
9. Shi, Z., et al., *Polyproline II structure in a sequence of seven alanine residues.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(14): p. 9190-9195.
10. Shi, Z., R.W. Woody, and N.R. Kallenbach, *Is polyproline II a major backbone conformation in unfolded proteins?* Advances in Protein Chemistry, 2002. **62**(Unfolded Proteins): p. 163-240, 1 plate.
11. Kim, Y.S., J. Wang, and R.M. Hochstrasser, *Two-Dimensional Infrared Spectroscopy of the Alanine Dipeptide in Aqueous Solution.* J. Phys. Chem. B, 2005. **109**(15): p. 7511-7521.
12. Kentsis, A., et al., *Unfolded state of polyalanine is a segmented polyproline II helix.* Proteins Structure, Function, and Bioinformatics, 2004. **55**(3): p.

493-501.

13. Mezei, M., et al., *Polyproline II helix is the preferred conformation for unfolded polyalanine in water.* Proteins Structure, Function, and Bioinformatics, 2004. **55**(3): p. 502-507.
14. Makowska, J., et al., *Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(6): p. 1744-1749.
15. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* Journal of Molecular Biology, 1963. **7**(1): p. 95-9.
16. Ramachandran, G.N. and V. Sasisekharan, *Conformation of polypeptides and proteins.* Advances in Protein Chemistry, 1968. **23**: p. 283-438.
17. Perczel, A., O. Farkas, and I.G. Csizmadia, *Peptide Models. 18. Hydroxymethyl Side-Chain Induced Backbone Conformational Shifts of L-Serine Amide. All ab Initio Conformers of For-L-Ser-$NH_2$.* Journal of the American Chemical Society, 1996. **118**(33): p. 7809-7817.
18. Perczel, A., O. Farkas, and I.G. Csizmadia, *Peptide models. XVI. The identification of selected HCO-L-Ser-$NH_2$ conformers via a systematic grid search using ab initio potential energy surfaces.* Journal of Computational Chemistry, 1996. **17**(7): p. 821-34.
19. Perczel, A., et al., *Peptide models. XIV. Ab initio study on the role of side-chain backbone interaction stabilizing the building unit of right- and left-handed helixes in peptides and proteins.* International Journal of Quantum Chemistry, 1997. **61**(5): p. 797-814.
20. Iwaoka, M., M. Okada, and S. Tomoda, *Solvent effects on the j-y potential surfaces of glycine and alanine dipeptides studied by PCM and I-PCM methods.* Theochem, 2002. **586**: p. 111-124.
21. Grant, J.A., R.L. Williams, and H.A. Scheraga, *Ab initio self-consistent field and potential-dependent partial equalization of orbital electronegativity calculations of hydration properties of N-acetyl-N'-methyl-alanineamide.* Biopolymers, 1990. **30**(9-10): p. 929-49.
22. Asciutto, E.K., et al., *Computational and Experimental Determination of the alpha -Helix Unfolding Reaction Coordinate.* Biochemistry, 2008. **47**(7): p. 2046-2050.
23. Chin, W., et al., *Spectroscopic Evidence for Gas-Phase Formation of Successive beta -Turns in a Three-Residue Peptide Chain.* Journal of the American Chemical Society, 2005. **127**(5): p. 1388-1389.

24. Chin, W., et al., *Gas Phase Formation of a $3_{10}$-Helix in a Three-Residue Peptide Chain: Role of Side Chain-Backbone Interactions as Evidenced by IR-UV Double Resonance Experiments.* Journal of the American Chemical Society, 2005. **127**(34): p. 11900-11901.
25. Chin, W., et al., *Probing the competition between secondary structures and local preferences in gas phase isolated peptide backbones.* Physical Chemistry Chemical Physics, 2006. **8**(9): p. 1033-1048.
26. Eker, F., et al., *Tripeptides Adopt Stable Structures in Water. A Combined Polarized Visible Raman, FTIR, and VCD Spectroscopy Study.* Journal of the American Chemical Society, 2002. **124**(48): p. 14330-14341.
27. Horvath, V., Z. Varga, and A. Kovacs, *Long-Range Effects in Oligopeptides. A Theoretical Study of the beta -Sheet Structure of Glyn (n = 2-10).* Journal of Physical Chemistry A, 2004. **108**(33): p. 6869-6873.
28. Wieczorek, R. and J.J. Dannenberg, *The Energetic and Structural Effects of Single Amino Acid Substitutions upon Capped alpha -Helical Peptides Containing 17 Amino Acid Residues. An ONIOM DFT/AM1 Study.* Journal of the American Chemical Society, 2005. **127**(49): p. 17216-17223.
29. Shang, H.S. and T. Head-Gordon, *Stabilization of Helixes in Glycine and Alanine Dipeptides in a Reaction Field Model of Solvent.* Journal of the American Chemical Society, 1994. **116**(4): p. 1528-32.
30. Van Duijneveldt, F.B., *Basis set superposition error.* Molecular Interactions, 1997: p. 81-104.
31. Jansen, H.B. and P. Ros, *Nonempirical molecular orbital calculations on the protonation of carbon monoxide.* Chemical Physics Letters, 1969. **3**(3): p. 140-3.
32. Boys, S.F. and F. Bernardi, *The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors.* Molecular Physics, 1970. **19**(4): p. 553-566.
33. Simon, S., M. Duran, and J.J. Dannenberg, *How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers?* Journal of Chemical Physics, 1996. **105**(24): p. 11024-11031.
34. Tobias, D.J. and C.L. Brooks, III, *Thermodynamics and mechanism of alpha helix initiation in alanine and valine peptides.* Biochemistry, 1991. **30**(24): p. 6059-70.
35. Vila, J.A., et al., *Fast and accurate computation of the 13C chemical shifts for an alanine-rich peptide.* Proteins Structure, Function, and Bioinformatics, 2004. **57**(1): p. 87-98.
36. Rablen, P.R., J.W. Lockman, and W.L. Jorgensen, *Ab Initio Study of*

*Hydrogen-Bonded Complexes of Small Organic Molecules with Water.* Journal of Physical Chemistry A, 1998. **102**(21): p. 3782-3797.

37. Dannenberg, J.J., *Enthalpies of Hydration of N-Methylacetamide by One, Two, and Three Waters and the Effect upon the C:O Stretching Frequency. An Ab Initio DFT Study.* Journal of Physical Chemistry A, 2006. **110**(17): p. 5798-5802.
38. Avbelj, F. and R.L. Baldwin, *Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: Distributions of phi.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(10): p. 5742-5747.
39. Avbelj, F. and R.L. Baldwin, *Limited validity of group additivity for the folding energetics of the peptide group.* Proteins Structure, Function, and Bioinformatics, 2006. **63**(2): p. 283-289.
40. Mons, M., et al., *Energetics of the Gas Phase Hydrates of trans-Formanilide: A Microscopic Approach to the Hydration Sites of the Peptide Bond.* Journal of Physical Chemistry A, 2001. **105**(6): p. 969-973.
41. Dannenberg, J.J., *The importance of cooperative interactions and a solid-state paradigm to proteins: what peptide chemists can learn from molecular crystals.* Advances in Protein Chemistry, 2006. **72**(Peptide Solvation and H-Bonds): p. 227-273.
42. Ingwall, R.T., et al., *Conformational studies of poly-L-alanine in water.* Biopolymers, 1968. **6**(3): p. 331-68.
43. Wieczorek, R. and J.J. Dannenberg, *alpha -Helical Peptides Are Not Protonated at the N-Terminus in the Gas Phase.* Journal of the American Chemical Society, 2004. **126**(39): p. 12278-12279.
44. Horng, J.-C., F.W. Kotch, and R.T. Raines, *Is glycine a surrogate for a D-amino acid in the collagen triple helix?* Protein Science, 2007. **16**(2): p. 208-215.