UNIVERSITAT POLITÈCNICA DE CATALUNYA
DEPARTAMENT D'ENGINYERIA QUÍMICA

"CONFORMATIONAL PROPERTIES OF CONSTRAINED PROLINE ANALOGUES AND THEIR APPLICATION IN NANOBIOLOGY"

Alejandra Flores Ortega

Supervisors: Dr. Carlos Alemán Llansó and Dr. Jordi Casanovas Salas.

Barcelona, 27th January 2009

“Chance is a word void of sense; nothing can exist without a cause”.

François-Marie Arouet, Voltaire

“*Imagination will often carry us to worlds that never were.*

But without it, we go nowhere”.

Carl Sagan

ACKNOWLEDGEMENTS

I would like to acknowledge to Dr. Carlos Aleman and Dr. Jordi Cassanovas Salas for an interesting research theme, and scientific support.

I gratefully acknowledge to Dr. David Zanuy for interesting suggestions and strong discussions, without their support this would be an unfulfilled task.

Also I, would like to address my thanks to all my colleagues in my group and department, specially Elaine Armelin for assiting me in many different ways. I thank not only my friends, but also colleagues for helping me to overcome the stressful time, without whom it would have been difficult to cope up.

I wish to express my gratefulness to my parents, specially to my mother, María Esther, for all his care, and support. Also I will like to thanks to my friends and specially Jesus, Merches, Laura y Arturo. My PhD thesis have been finished for all this support.

I am greatly indepted to Dr. Ruth Nussinov at NCI, Dr. Carlos Cativiela at the University of Zaragoza and Ana I. Jiménez at the "Instituto de Ciencias de Materiales de Aragon" for a collaborative effort.

I wish to thank all my colleague in the *"Chimie et Biochimie Théoriques, Faculté des Sciences et Techniques"* in Nancy France, I will be grateful to have worked with : Pr. Xavier.Assfeld and PhD Adele Laurent.

I gratefully acknowledge the financial support provided by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

OBJECTIVES

(1) Examine the conformational preferences of proline analogs having one or more double bonds in the pyrrolidine ring. Analyze the influence of the insaturations on: (i) the stability of the *cis* arrangement of the peptide bond involving the pyrrolidine nitrogen; and (ii) the conformational flexibility of the backbone.

(2) Analyze the intrinsic conformational preferences of two representative α-tetrasubstituted proline analogs (α-methylproline and α-phenylproline) and compare them with those of conventional proline. Understand the effects of the substituent incorporated at the α position on the preferred backbone conformation, the puckering of the pyrrolidine ring and the *cis*/*trans* disposition of the amide bonds.

(3) Compare the conformational properties of different aminated and dimethylaminated derivatives of proline. Examine how the formation of side chain···backbone hydrogen bonds affects not only the conformational flexibility but also the *trans*/*cis* disposition of the peptide bond involving the pyrrolidine nitrogen.

(4) Determine the conformational preferences of the aminoproline analogs protonated at the amino side group. Analyze the influence of the pH on the relative stability of the different possible isomers, the backbone flexibility and the disposition of the peptide bond.

(5) Characterize the conformational profile of the CREKA sequence, which defines a very efficient tumor-homing pentapentide, and identify the corresponding bioactive conformation. A satisfactory achievement of this objective is essential for designing of synthetic analogs able to provide protection from proteases, which is an important step before the development of potential applications of tumor-homing peptides.

(6) Improve the biological performance and pharmacological profile of CREKA by engineering an analogue that incorporates a non-proteinogenic amino acid. This residue should be conceived to retain the most relevant characteristics of the conformational profile of the natural peptide and simultaneously impart stability against proteolytic cleavage.

GLOSSARY.

A	Puckering Amplitude
αMePro	Methylproline
Amp	Aminoproline Dipeptides
αPhPro	Phenylproline
Aze	L-azetidine-2-carboxylic acid
azPro	azaproline
B3	Becke´s three-parameter hybrid functional
C_7	Seven Memberedintramolecular Hydrogen Bond
ΔEgp	Relative Energy
DFT	Density Functional Theory
ΔGgp	Gibbs free energies in the gas phase
Dmp	Dimethylaminoproline
E	Energy of the System
HF	Hartree-Fock
Hyp	4R-Hydroxyproline
φ	Fi
LYP	Lee, Yang and Parr
MP	Moller-Plesset
NHMe	N-Methyl Amide Group
Oxa	(S)-oxazolidine-4-carboxylicacid
P	State of Puckering
Pip	(S)-piperidine-2-carboxylic acid
Pro	Proline
SA	Simulated Annealing - Molecular Dynamics
SCF	Self Consist Field
SPIO	Dextran-Coated Iron Oxide
Thz	((R)-thiazolidine-4-carboxylicacid
UHF	Unrestricted Hartree-Fock
ZPVE	Zero-Point Vibrational Energies
ρ	electronic density
Ψ	Wave Function
ψ	Psi

TABLE OF CONTENTS

1 Introduction 1

1.1 Prolin estructure and properties 1
1.2Survey of modified Proline residues. Conformational features 3
1.3 Peptide design: improving nature for bionanotechnological applications 5
1.4 References 8

2 Methods 11

2.1 Introduction 11
2.2 Quantum Mechanical Methods 11
2.2.1 Ab Initio Methods 12
2.2.2 Hartree-Fock Method 13
2.2.3 Other Ab Initio Methods 14
2.2.4 DFT Methods 15
2.2.4.1 Correlation Term 16
2.2.4.2 Hybrid Functionals 16
2.2.5 Solvent Effects 16
2.3 Molecular Dynamics Simulation 18
2.3.1 Force Field 18
2.3.2 Classical Dynamics 19
2.3.3 Periodic boundary conditions 20
2.3.4 Temperature and pressure 20
2.4 Conformational Search Methods 21
2.5 References 24

3 Intrinsic Conformational Properties of synthetic Proline Analogues 27

3.1 Conformation of Proline Analogs Having Double Bonds in the Ring 29
3.1.1 Introduction 29
3.1.2 Methods 32
3.1.2.1 Computational Details 32
3.1.2.2 Nomenclature and Pseudorotational Parameters 33
3.1.3 Results and Discussion 34
3.1.3.1 Ac-L-Pro-NHMe 34
3.1.3.2 Ac-$\Delta^{\alpha,\beta}$Pro-NHMe 36
3.1.3.3 Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe 41
3.1.3.4 Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe 42
3.1.3.5 Ac-Py-NHMe 43
3.1.3.6 Relative Stability of the Cis Conformers 44
3.1.4 Conclusions 47
3.1.5 References 48

3.2 Conformational Preferences of α-Substituted Proline Analogues 53
3.2.1 Introduction 53
3.2.2 Methods 56
3.2.2.1 Computational Details 56

3.2.2.2 Nomenclature and Pseudorotational Parameters. 59
3.2.3 Results and Discussion 60
3.2.3.1 Ac-L-Pro-NHMe. 60
3.2.3.2 Ac-L-αMePro-NHMe 66
3.2.3.3 Ac-L-αPhPro-NHMe. 72
3.2.4 Conclusions 75
3.2.5 References 76

3.3 Conformational Preferences of β- and γ- Aminated Proline Analogues ... 83
3.3.1 Introduction 83
3.3.2 Methods 86
3.3.2.1 Nomenclature and Pseudorotational Parameters 89
3.3.3 Results and Discussion 90
3.3.3.1 Aminoproline (Amp) dipeptides 90
3.3.3.2 Dimetylaminoproline (Dmp) dipeptides 101
3.3.4 Conclusions 112
3.3.5 References 114

3.4 Protonation of the side group in β-and γ-Aminated Proline Analogues: Effects on the Conformational Preferences 119
3.4.1 Introduction 119
3.4.2 Methods 122
3.4.3 Results and Discussion 124
3.4.4 Conclusions 138
3.4.5 References 140

4 Targeted replacement with constrained amino acid in tumor homing peptides 143

4.1 The energy landscape of a selective tumor-homing pentapeptide 145
4.1.1 Introduction 145
4.1.1.1 Simulated Systems and Bioactive Conformation 147
4.1.1.1.1 System I 147
4.1.1.1.2 System II 147
4.1.1.1.3 System III 148
4.1.1.1.4 System IV 148
4.1.2 Methods 149
4.1.2.1 Conformational Search 149
4.1.2.2 Molecular Dynamics and Computational Details. 150
4.1.2.3 Conformation Classification and Clustering Analysis. 151
4.1.3 Results 152
4.1.3.1 Free CREKA in Unionized and Ionized Aqueous Solution (Systems I and II) 152
4.1.3.2 CREKA Attached to the Surface of a SPIO-Nanoparticle (System III) 159
4.1.3.3 CREKA Inserted in a Phage Display Protein (System IV) 162
4.1.3.4 Stability of the β-Turn Motif in CREKA. 165
4.1.4 Discussion and Conclusions 168
4.1.5 References 171

4.2 In silico molecular engineering for a targeted replacement in a tumor-homing peptide 175
4.2.1 Introduction 175
4.2.2 Computational Methods 177
4.2.2.1 Quantum mechanical calculations 177
4.2.2.2 Conformational search and force-field calculations. 178
4.2.2.3 Conformation classification and clustering analysis. 180
4.2.3 Results and Discussion 182
4.2.3.1. Design of the Chemical Modifications in CREKA 182
4.2.3.2 Conformational Profile and Force-Field Parametrization 184
4.2.3.3 Conformational Search of the CREKA Analogue Attached to a Nanoparticle 193
4.2.4 Conclusions 201
4.2.5 References 203

5 Discussion of the Results 209

6 Conclusions 215

1 Introduction

1.1 Proline: structure and properties

Among the 20 amino acids that universally constitute all proteins, Proline (Pro) is the only one that presents specific conformational features: its chemical constitution is characterized by presenting its aliphatic side chain fused to the molecule back bone[1]. Thus, the characteristic primary amino group, that the other 19 amino acids present in their free form, becomes a secondary amino group by the bond formed between its nitrogen atom and the delta carbon of the side chain (Chart I). This structural feature first implies that Proline loses one of the two main chain conformational degrees of freedom, since the cyclized side chain locks the backbone torsion defined by the bond N-C^{α} (ψ at Chart I) around -60°. This conformational restriction limits its presence in protein structural motifs, as it will be pointed bellow. The five member ring (*pyrrolidine* ring) also features its own conformational profile. In the available crystal structures of peptides and proteins, two particular puckered conformations are the most populated and equally probable: The down and up ring puckerings are defined as those in which the C^{γ} and C=O group of the Pro residue lie on the same and opposite sides, respectively, of the plane defined by C^{δ}, N, and C^{α} atoms.[2]

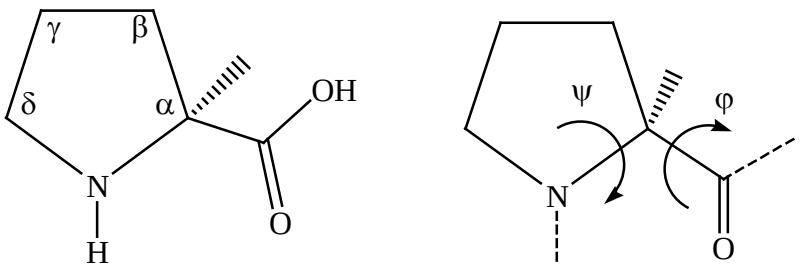


Figure 1.1: *Chart I*

Another unique property associated with this residue is the relatively high impact that the *cis* configuration has on the peptide bonds preceding the Pro in many naturally occurring protein systems (5-6%) in front of the almost negligible impact for the other 19 residues (0.03 %).[3-5] The cis-trans isomerization of Pro (Chart II) is often involved in the rate-determining steps for folding and refolding of some proteins.[6-7]

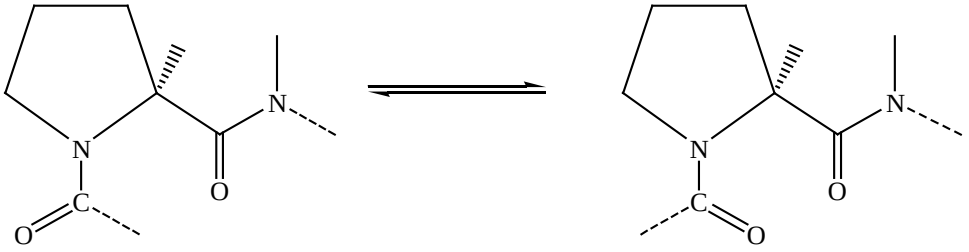


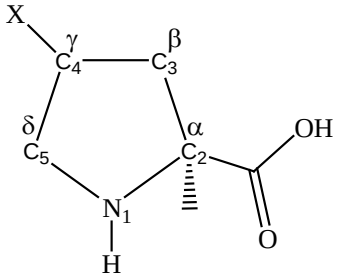
Figure 1.2: *Chart II*

The conformational restrictions imposed by its constitution determine the structural impact of Pro in natural peptides and proteins. Hence, Pro is usually placed as secondary structure disruptor, whereas its prominent presence in the first positions of turns is an especial case of such feature. Its presence in helical segments generally is limited to the forward positions, since the lack of hydrogen bonded to the main chain nitrogen when forming peptide bonds only habilitates Proline as hydrogen bond acceptor. As mentioned before, Pro has a remarkable impact in turns, since its conformational restrictions make Pro an excellent hinge for changing the direction in β strands.[8] Finally, the fixed value for ψ dihedral angle around *gauche*$^-$ conformation aims the formation of a special helical arrangement, the *polyproline* helix, when Pro (and 4-hydroxiproline) polymerize. This type of regular arrangement is the predominant secondary structure in collagen fibers, which are constituted by the lateral assembly of 3 polypeptide chains folded in the aforementioned helical conformation.

1.2 Survey of modified Proline residues. Conformational features

Among other strategies, the chemical modification of natural proteogenic amino acids has proven to be an efficient approach for restricting the conformational space of such molecular species.[9] This feature is of great interest for further nanotechnological applications, since a major part of the work devoted to redesign natural biomacromolecules implies exercising conformational control over such systems (either as a whole or small parts of them).[10] The particular conformational features of Proline has made of it a major target for potential molecular engineering modifications, since its inner constitutional restrictions aids limiting its low energy accessible conformations, helping biasing the conformational freedom of the peptide in which this residue is included.

The most characterized chemical substitution in Proline is the presence of a functional group at the position 4 of the ring (C^{γ} at chart II). This modification has special relevance in biological systems since the substitution of the Prochiral hydrogen *R* in such carbon by a hydroxyl group (4*R*-hydroxy-L-proline, *Hyp*) is detected in about the 50% of the Prolines present in natural collagen (Chart III).[1] The presence electron-withdrawing groups significantly affects the conformational dynamics of the pyrrolidine ring puckering transition. Thus, the presence of either a hydroxyl group or a Fluoride atom (4*R*-fluoro-L-proline, *Flp*) affect the transition barriers between the up and down fall puckering.[11] Moreover, these changes also have a reflection over the main chain conformational preferences, since an increment in the population on the up-puckered conformation aids the stabilization of the ε_L conformation (also denoted polyproline II-like or PP_{II}), which is very unfavored in the unsubstituted Pro.[12]

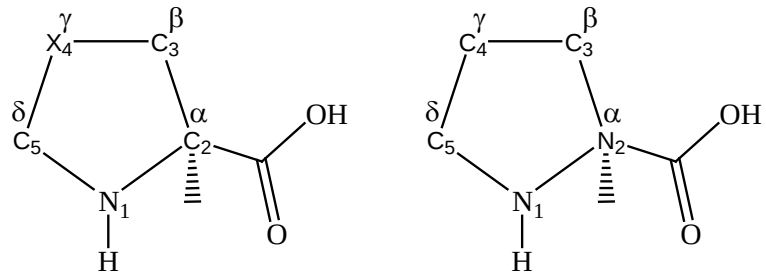


X= -OH, *Hyp* X= -F, *Flp*

Figure 1.3: *Chart III*

Other ways of altering the conformational features of Proline is playing with its chemical constitution. The ring size can be changed, either reducing its flexibility by diminishing the ring size (eliminating a methylene group, the L-azetidine-2-carboxylic acid, *Aze*) or by inducing the opposite effect by enlarging the cyclized segment (inserting a new methylene group, the (*S*)-piperidine-2-carboxylic acid, *Pip*).[13] In both systems, when comparing with the native proline, there is a reduction of the energy difference between C_7 arrangement (also denoted γ_L) and ε_L, becoming the former the most favored conformation as the solvent polarity increases. It is worth noting that despite the notable differences between the new constitutions and the Pro there isn't great differences respect to both the main chain conformational preferences and the *cis-trans* equilibrium in the peptide bond.[13]

Another possibility is changing the chemical nature of the pyrrolidine ring: it is possible to think about replacing the methylene unit of the gamma position by another functional group. For instance Kang and Park[14] explored the possibility of converting the Pro ring into and heterocyclic species (see Chart IV) by introducing in such position either an oxygen atom ((S)-oxazolidine-4-carboxylicacid, named Oxa) or a thiol group ((R)-thiazolidine-4-carboxylicacid, named Thz). In these pseudo proline residues the main chain conformation preferences are very influenced by the polarity of the solvent, as also observed for the unmodified proline.[15] Thus, in both Oxa and Thz the population of γ_L conformation decreases as the medium polarity increases, favoring ε_L arrangement. The puckering preferences though invert their tendencies in the case of the new rings respect to the natural residue, since the polarity promotes the presence of up arrangements for the latter but unfavors them for the former residues.



X= O, *Oxa* *azPro*

X= S-H, *Thz*

Figure 1.4: *Chart IV*

Finally, the proper backbone of the amino acid can also be modified: if the C^{α} is replaced by a nitrogen atom, the azaproline (*azPro*) is obtained (Chart IV).[16] In this case the new electronic structure resulting from the insertion of the azo moiety drastically changes the new residue conformational preferences, highly stabilizing β-sheet like arrangements (δ_L conformation). Furthermore, the effect of the non-bonded electron pairs of the new nitrogen atom aids over stabilizing the *cis* arrangement for the second peptide bond, when azPro is inserted between a dipeptide moiety. As the polarity of the medium increases the preferences of this residue tend to meet those of the natural proline by favoring the ε_L arrangement over the δ_L one.

1.3 Peptide design: improving nature for bionanotechnological applications

The relationship between folding and function among proteins has long been a source of fascination to the molecularly inclined scientist. The interplay between R-amino acid residue sequence and the three-dimensional arrangement of these subunits that results from adoption of a specific conformation enables proteins to manifest an extraordinary range of functions. Among such sequences, short peptides have found tremendous attention in diverse aspects of science ranging from rational drug design[17] to nanomaterials.[18] These diverse applications are due

to their distinctive properties, such as ease of synthesis and characterization, introduction of chemical diversity by simple amino acid substitution, and modulation of 3D structure by chemical modification. The application of peptides as drugs stems from their key role in many signal transduction pathways, which makes them an attractive avenue to target diseases. Despite the high activity and receptor selectivity of naturally occurring bioactive peptides (or active protein fragments), they have distinct disadvantages for practical application in medicine, such as short half-life *in vivo* and lack of oral availability. The initial step in drug research of peptides is usually simplification (e.g., reduction in size), followed by peptidomimetic approaches to ensure metabolic stability, with the final goal of an orally available, highly active, and selective drug. Whereas the preliminary steps can be done in a rational way with relatively high probability of success, the final, crucial step of conversion from peptide into a drug is often more problematic.

On the other hand, not all of the amino acids in a peptide sequence are essential to achieve the biological effect. The initial identification of the "bioactive sequence", the minimal sequence[19] required to achieve the biological activity, is often done by alanine scanning. This is the systematic substitution of each amino acid by alanine to identify the key residues, that is, those whose substitution results in reduced activity. The next important factor is the conformation of the peptide. In the majority of such peptides, a major obstacle in the study of the "bioactive sequence" is intrinsic flexibility. Thus, the active sequence must be rigidified in a defined conformation in order to achieve the desired activity and selectivity. Reduction of conformational space can be achieved by cyclization, resulting in highly active and selective derivatives when the bioactive conformation is matched.[20] This search of matching is done by "spatial screening".[21] An alternative strategy can be the systematic exploration of the hypersurface of potential energy of the studied segment. Hence, the first step would consist of *in silico* exploring all the accessible conformations that the peptide can adopt under physiological (or under those conditions that experimental information has been recorded).[22] Once the accessible conformations have been identified, specific chemical modifications can be targeted to enhance those conformations that contribute to the collective of bioactive conformations.

Another major problem in developing peptidic drugs is their enzymatic degradation *in vivo*, which eventually results in the lowering of the pharmacokinetic profile (half-life, bioavailability, etc.). Medicinal chemists have developed an array of strategies over the years to confront this problem, such as incorporating peptide bond isosters,[23] peptoids,[24] retro-inverso peptides,[25] and peptidomimetics.[26] Although these strategies have elegant properties of their own, they demand careful design with challenging syntheses. Despite this complexity, targeted modifications of the natural peptides present high potential, since they can allow to both enhance the peptide inner properties (specificity and activity) and to provide new physicochemical features to the modified segment, such as resistance to endogen proteases.

1.4 References

1. Richardson, J. S.; Richardson, D. C. Principles and patterns of protein conformation. In *Prediction of Protein Structure and the Principles of Protein Conformation*; Fasman, G. D., Ed.; Plenum Press: New York, **1989**, 98.
2. Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. *Protein Sci.* **2001**, *10*, 2632.
3. Stewart, D. E.; Sarkar, A.; Wampler, J. E. *J. Mol. Biol.* **1990**, *214*, 253.
4. Jabs, A.; Weiss, M. S.; Hilgenfeld, R. *J. Mol. Biol.* **1999**, *286*, 291.
5. Pal, D.; Chakrabarti, P. *J. Mol. Biol.* **1999**, *294*, 271.
6. Wedemeyer, W. J.; Welker, E.; Scheraga, H. A. *Biochemistry* **2002**, *41*, 14637.
7. Dugave, C.; Demange, L. *Chem. Re*V. **2003**, *103*, 2475.
8. Gibbs, A. C.; Bjorndahl, T. C.; Hodges, R. S.; Wishart, D. S. *J. Am. Chem. Soc.* **2002**, *124*, 1203.
9. Crisma, M.; Formaggio, F.; Moretto, A.; Toniolo, C. *Biopolymers*, **2006**, *84*, 12.
10. Alemán C.; Zanuy, D.; Jiménez A.I.; Cativiela, C.; Haspel, N.; Zheng, J.; Casanovas, J.; Wolfsom, H.; Nussinov, N. *Phys. Biol.* **2006**, *3*, S62.
11. Song, I.K.; Kang, Y.K. *J.Phys.Chem.B* **2005**, *109*, 16987.
12. Song, I.K.; Kang, Y.K. *J.Phys.Chem.B* **2006**, *110*, 1927.
13. Jhon, J.S.; Kang Y.K. *J. Phys. Chem. B* **2007**, *111*, 3507.
14. Kang, Y.K.; Park H.S. *J.Phys. Chem. B* **2007**, *111*, 12562.
15. Improta, R.; Benzi, C.; Barone, V. *J. Am. Chem. Soc.* **2001**, *123*, 12577.
16. Kang K.Y.; Byun B.J. *J. Phys. Chem. B* **2007**, *111*, 5385.
17. Marx, V. *Chem. Eng. News* **2005**, *83*, 21.
18. Teixido, M.; Giralt, E. *J.Pept. Sci.* **2008**, *14*, 173.
19. Gurrath, M. *Curr. Med. Chem.* **2001**, *8*, 1648.
20. Kessler, H. *Angew. Chem., Int. Ed.* **1982**, *21*, 523.
21. Kessler, H.; Gratias, R.; Hessler, G.; Gurrath, M.; Muller *Pure Appl. Chem.* **1996**, *68*, 1205.

22. Agrafiotis, D. M.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E. *J. Chem. Inf. Model.* **2007**, *47*, 1086.
23. *Houben-Weyl Methods of Organic Chemistry*; Goodman, M.; Felix, A.; Moroder, L.; Tonolio, C. Eds.; Georg Thieme Verlag: **2002**, *E22c*, 633.
24. Kessler, H. *Angew. Chem., Int. Ed.* **1993**, *32*, 544.
25. Fletcher, M. D.; Campbell, M. M. *Chem. Rev.* **1998**, *98*, 795.
26. Giannis, A. *Angew. Chem., Int. Ed.* **1993**, *32*, 1267.

2 Methods

2.1 Introduction

In this chapter the methods used throughout this Thesis, which can be organized in quantum mechanical methods and molecular dynamics simulations, will be discussed. In the section 2.2, the basic elements of quantum mechanical methods are presented. Specifically, the more essential trends of ab initio and DFT methods are described. In section 2.3 classical methods based on Molecular Dynamics simulations are briefly discussed. Finally, section 2.4. presents the basic concepts of the conformational search procedures used in this Thesis.

2.2 Quantum Mechanical Methods

These methods provide a reliable description of the energies, geometries and electronic properties of the systems under study. In this approach, nuclei are arranged in the space while the corresponding electrons are spread all over the system in continuous electronic density and computed using the Schrödinger equation.

When the Schrödinger equation 2.1 is solved, quantum mechanical methods postulate the existence of a wave function, Ψ, that contains all the information of the system:

$$\hat{H}\,\Psi = E\Psi \qquad (2.1)$$

where $\hat{H}$ is the Hamiltonian operator that includes the kinetic and potential energy of the nuclei and electrons, and E is the energy of the system. Two basic

quantum mechanical methodologies are currently used to study chemical problems: ab initio and Density Functional Theory (DFT), which differ in the procedure to obtain Ψ.

2.2.1 Ab Initio Methods

For a system of N nuclei and M electrons, the $\hat{H}$ is expressed as (in atomic units):

$$\hat{H} = \sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{1}{2m_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N} \frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{2.2}$$

where m_A is the relation of the nucleus mass with respect to the electron mass, Z_A is the atomic number of the nucleus A, $\nabla_i^2 and \nabla_A^2$ are operators that refer to the differentiation between the coordinates of electron *i* and the atom A, respectively. In equation 2.2 each term is an operator defining the energy components of the system: the first term represents the kinetic energy of the electrons, the second term is the kinetic energy of the nucleus, the third one is the electrostatic attraction between nucleus and electrons, the fourth one is the electrostatic repulsion between the electrons and, finally, the fifth one corresponds to the electrostatic repulsion between nuclei. Taking into account the high ratio between nuclear and electronic masses, the Born-Oppenheimer approach allows discard the second and the fifth terms, the general Hamiltonian being transformed into an electronic Hamiltonian ($\hat{H}_{el}$):

$$\hat{H}_{el} = \sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N} \frac{1}{r_{ij}} \tag{2.3}$$

In principle, it is possible to describe all chemical systems by solving the Schrödinger equation. However, in practice, only the simplest ones may be studied exactly using this level of theory, and the introduction of some

approximations necessary (Hartree-Fock, post-Hartree-Fock or Density Functional).

The reliability of the results obtained by applying the Schrödinger equation depends on the quality of quantum chemical procedure but also on the number of atomic orbitals employed to describe the chemical system (see next section). It should be noted that the more sophisticated the basis set is, the longer calculation time is needed. The 3-21G[1], 6-31G (d)[2], 6-31+G (d,p)[3]and 6-311G(d,p)[4] are, in increasing order of complexity, the basis sets more frequently used.

2.2.2 Hartree-Fock Method

The Hartree-Fock method (HF) is an approximated process for the determination of the ground state wavefunction and the energy of a chemical system. The HF method assumes that the wavefunction of the system can be approximated by a single Slater determinant of spin-orbitals and solves the Schrödinger equation through an iterative process applying the variational principle. In addition, the molecular orbitals are required to be orthogonal. The first step of this iterative process consists on the generation of a test wave function, and then the Fock operator used to describe how each particle is subjected to the mean field created by all the other particles is build. The equation's values and their own functions are solved and the procedure is repeated till reach the convergence. For this reason, the HF method is also known with the name of Self Consistent Field (SCF). The functions that the HF method uses are expressed as an antisymmetrical linear combination of spin-orbital products (*i.e.* a Slater determinant; equation 2.4):

$$\left|\Psi_o\right| = \chi_1\chi_2\cdots\chi_i\cdots\chi_N \qquad (2.4)$$

The spin-orbitals (χ_i) are represented as a linear combination of monoelectronic orbitals (also called base functions), which often correspond to atomic orbitals. This approximation is known as Linear Combination of Atomic Orbitals:

$$\chi_i = \sum_{v=1}^{k} C_{vi}\phi_v \qquad (2.5)$$

where *k* is the number of base functions ϕ used to describe the spin-orbital χ_i, and C_{vi} are the coefficients multiplying each atomic orbital.

The base functions are usually Gaussian-type because of their computational efficiency. The results provided by the HF method strongly depend on the number of base functions used. The smallest number of functions that can accommodate all the electrons of the system is denoted minimum basis set. In the split-valence basis sets, which are have been used to study the bio-organic chemical systems reported in this work, the internal electrons are described with a single base function, whereas two or more basis are used for valence electrons. Furthermore, additional functions, as polarization and diffuse functions, can be added to improve the description of the electronic density. As was mentioned above, the quality of the results increases with the size of the basis set. Accordingly, the energy of the system decreases with the number of base functions until it reaches the HF limit. Once this limit is reached, the reduction of the energy cannot continue by enlarging the basis set.

Many chemical systems bring all their electrons in a coupled form. Such kinds of systems, which are denoted closed shell systems, are determined by means of the Restricted Hartree-Fock (RHF) approach. In this approach spin-orbitals χ_i χ_{i+1} are described as having the same space function but different spin function. However, charged and excited systems require the Unrestricted Hartree-Fock (UHF) approach in which spin-orbitals also differ in the space function.

2.2.3 Other Ab Initio Methods

An intrinsic approach of the HF method lies in the assumption that each electron moves in a field generated by the average electronic distribution produced by the remaining electrons. Since the repulsion between electrons is not explicitly considered in this approach, the HF method cannot provide good results for some properties, as for example excited states and chemical reactions. However, more sophisticated ab initio methods can be used to study such properties, in which the electronic correlation is crucial.

An easy way to include the electronic correlation is through the configuration interaction method. In this procedure the ground electronic state is described as a

mixture of all the possible electronic states (full-CI). In practice, this expansion is usually truncated to a few accessible electronic states (truncated-CI). The CI method allows obtain the most complete non-relativistic description of chemical systems, but is computationally prohibitive for medium and large size systems. Another way to include the electronic correlation is through perturbational methods. Without any doubt, the Møller-Plesset (MP) procedure[5], which is based on the theory of Rayleigh and Schrödinger, is the most frequently used.

2.2.4 DFT Methods

The DFT approach is based on the theorems of Hohenberg-Kohn[10] which demonstrate the existence of an electronic density functional able to provide the energy of the fundamental state, even although the form of such functional cannot be determined.

The DFT methods are based in the theory developed by Kohn and Sham: the electronic density of a system can be represented as the sum of N monoelectric orbital densities, which ensures the DFT calculations to be easily applicable. Using these ideas, the electronic energy, which includes electron correlation, is calculated as the sum of several terms that depend on the electronic density. In this context the electronic energy can be calculated using the equation 2.6:

$$E = E^T + E^V + E^J + E^{XC} \tag{2.6}$$

where E^T is the kinetic energy, E^v contains the terms of the potential energy nucleus-electron and the repulsive term between nuclei, E^J represents the Coulombic repulsion energy between electrons and, finally, E^{XC} is interchange-correlation energy. All of these energetic terms, with exception of the nucleus-nucleus repulsion, depend on the electronic density ρ.

The term E^{XC} is not determined directly in this approximation, due to its unknown mathematical formulation. Usually this term is described as a sum of an exchange term E^X and another of electronic correlation E^C (equation 2.7).

$$E^{XC} = E^X(\rho) + E^C(\rho) \tag{2.7}$$

2.2.4.1 Correlation Term

This term can be calculated by local or non local functionals. In the present thesis we have used the non local functional proposed by Lee, Yang and Parr (LYP)[11].

2.2.4.2 Hybrid Functionals

In order to calculate more efficiently the term E^X, it is possible to use a mixture of the HF exchange with the DFT exchange-correlation, what is known as *hybrid functional.* The hybrid functional more frequently used corresponds to that derived by Becke (proposed in 1988)[12].

Nowadays, one of the most known combinations, and used in this thesis, is the combination of the Becke three parameters hybrid functional (B3) with the non local correlation provided by the LYP expression (B3LYP) It is well established that B3LYP calculations provide a reliable description of a large number of organic chemical systems from both qualitative and quantitative points of view.

2.2.5 Solvent Effects

Evaluate the solvent effects is one of the majors goals in computational chemistry, since many important chemical processes take place in solution rather than in the gas-phase.

The continuum models, describe the solvent as a dielectric continuum, which can interact with the solute molecules. Within these models, the Self-Consistent Reaction Field (SCRF), have been the most successful. This method treats the solute at the quantum mechanical level, while the solvent is represented as a dielectric continuum.

In particular, the Polarizable Continuum Model developed by Tomasi and co-workers,[13] and used in this Thesis, calculates the molecular free energy of a system in solution as a sum over three terms:

$$G_{sol} = G_{es} + G_{dr} + G_{cav} \qquad (2.8)$$

where G_{es} represents the electrostatic, G_{dr} the dispersion-repulsion and the G_{cav} the cavitation energy contributions to the free energy.

All three terms are calculated using a solvent cavity defined through interlocking *van der Waals*-spheres centered at the atomic positions of the solute and involves the calculation of virtual point charges on the cavity surface. The van der Waals radius of each atom is a function of the atom type, connectivity, overall charge of the molecule, and the number of attached hydrogen atoms. The magnitude of the charges is proportional to the derivative of the solute electrostatic potential at each point calculated from the molecular wavefunction. Then, the point charges may be introduced in the electron Hamiltonian what represents the polarization of the solute. An iterative calculation is carried out until the wavefunction and the surface charges are self-consistent.

2.3 Molecular Dynamics Simulation

2.3.1. Force Field

The basis of all molecular simulation methods is a potential energy function Φ (R). This function gives the potential energy of a system containing N atoms and depends on the position $\boldsymbol{r}_i$ of each atom i, which we denote by the 3N-dimensional vector $R = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)$. Ideally one would like to obtain this energy quantum chemically in which case Φ (R) would describe the Born-Oppenheimer surface for the motion of the nuclei (see section 2.2). In practice this would only be feasible for systems containing a few tenths of atoms, whose dynamics could then only be simulated for very short times due to the large computational time spent to evaluate Φ (R) throughout the simulation. Unfortunately, such approach is not useful for peptides or macromolecular systems containing hundreds up to thousands of atoms. Instead, Φ (R) is expressed as series of analytical terms: a so-called force field. The analytical form can, in principle, be chosen freely; in practice, however, it expresses the chemical common sense. A typical force field of a system of N atoms looks like:

$$\begin{aligned}\Phi(R) = & \sum_{\text{bonds}} \tfrac{1}{2} k_b [b - b_o]^2 + \sum_{\text{angles}} \tfrac{1}{2} k_\theta [\theta - \theta_o]^2 \\ & + \sum_{\text{impropers}} \tfrac{1}{2} k_\varepsilon [\varepsilon - \varepsilon_o]^2 + \sum_{\text{dihedrals}} k_\varphi [1 + \cos(n\varphi - \delta)] \\ & + \sum_{\text{pairs}(i<j)} \left[c12(i,j) / r_{ij}^{12} - c6(i,j) / c_{ij}^{6} + q_i q_j / 4\pi\pi_0 r_{ij} \right]\end{aligned} \tag{2.9}$$

The first two terms in Eq.(2.9) represent stretching and bond angle bending energies of the covalent bond, respectively. The next two terms describe the so-called improper dihedrals – torsion angles that do not undergo transitions (*e.g.* dihedral within aromatic rings) – and a sinusoidal term for the other dihedral angles, which may make 360^o turns. The last term, the nonbonded part, models the interactions between atoms that are three or more bonds away plus all intermolecular interactions. It is composed of the van der Waals interaction, consisting of hard core repulsion ($\boldsymbol{r}^{-12}$ *–term)* and dispersive attraction ($\boldsymbol{r}^{-6}$ *–term*), and the Coulomb interaction ($\boldsymbol{r}^{-1}$ *–term)*.

The parameters K_b, b_o K_θ, θ_o, K_ε, ε_o, K_ε, ε_o, K_φ , η , δ, C_6, C_{12} and q are the force field parameters. Quantum chemical calculations can be used to obtain some of these, *i.e.* the molecular geometries (angles, bond lengths), torsional potentials and atomic charges, whereas others are usually obtained empirically. In particular, the non-bonded terms are often parametrized to properly describe the liquid state properties, *e.g.* density, second virial coefficient, heat of vaporization, radial distribution function and/or time dependent properties such as the diffusion coefficient or rotational correlation time.

Many force fields are available. A proper choice should be made bearing in mind the properties that were used for parametrization. If one intends to calculate the excess free energy of liquid water, *i.e.* the free energy for turning one mole of water into an ideal gas, a force field designed to properly describe the heat of vaporization will be a good candidate.

2.3.2. Classical Dynamics

The motion of atoms in a molecular system can be simulated using the classical equation of motion provided that the force field $\Phi(R)$ is available. The classical equation read

$$dv_i(t)/dt = m_i^{-1}F_i(R(t)) \tag{2.10}$$

$$dr_i(t)/dt = v_i(t) \tag{2.11}$$

where $v_i(t)$ is the velocity of atom *i* at time t, m_i is its mass and F_i is the force on *i*,

$$F_i = \frac{\partial\Phi(R)}{\partial r_i} \tag{2.12}$$

Equation (2.10) and (2.11) cannot be solved analytically and one therefore is forced to use finite difference methods. A simple finite Taylor expansion of $\mathbf{v}_i(t)$ at time point $t = t_n$ yields[14,15].

$$v_i(t_n \pm \Delta t/2) = v_i(t_n) \pm \left.\frac{dv_i}{dt}\right|_{t_n} (\Delta t/2) + \left.\frac{d^2v_i}{dt^2}\right|_{t_n} (\Delta t/2)^2/2! + O(\Delta t^3) \tag{2.13}$$

Subtraction of these two expressions and using Eq. (2.10) gives

$$v_i(t_n + \Delta t/2) = v_i(t_n - \Delta t/2) + m_i^{-1} F(R(t_n))\Delta t + O(\Delta t^3) \quad (2.14)$$

Using the same procedure for Taylor expansions of $\boldsymbol{r}_i$.(t) at time point t = t_n + Δt/2 gives in combination with Eq. (2.11)

$$r_i(t_n + \Delta t) = r_i(t) + v_i(t_n + \Delta t/2)\,\Delta t + O(\Delta t^3) \quad (2.15)$$

In this work bond lengths have been constrained using the method proposed by Ryckaert et al., which is known as the SHAKE algorithm[16].

2.3.3. Periodic boundary conditions

The MD simulation of a liquid is performed using a simulation box that is typically filled with a few thousands of atoms. Dealing with macromolecular systems this means that we usually simulate chains each containing several hundreds of atoms. To remove the otherwise significant wall effects, a practical trick, known as the periodic boundary condition, is applied. This consists on surround the simulation box by identical copies to form a bulk system. This condition ensures that atoms moving out of the box at one side are able to re-enter the box at the opposite side since the replicas of this particle in the neighbouring boxes move in exactly the same way. This strategy eliminates the walls at the boundary of the central box and the surface molecules.

2.3.4. Temperature and pressure

The algorithm for molecular dynamics described above generates the time evolution of the system with fixed number of particles N, volume V and energy U, where the latter is the sum of the kinetic energy and the potential energy. Experimentally one often obtains information at constant volume and temperature (NVT) or at constant pressure and temperature (NPT). One therefore would like to have an algorithm that simulates the dynamics of the system at fixed values of P and or T.

Temperature is controlled by adjusting the velocities of the atoms during the simulation. This is done because the average kinetic energy of the molecules in the system defines temperature. The extent to which the velocities need to be adjusted is made to depend on the instantaneous value of the kinetic energy at a

given time together with the desired kinetic energy (temperature) of the system (a large difference causes a larger adjustment).

A fixed pressure means that volume must be able to fluctuate; *e.g.* a torsional transition in a molecule in solution at fixed pressure causes a small sudden volume fluctuation, or, dissolving a molecule in a membrane at fixed pressure causes the membrane to swell. In constant pressure simulations one therefore needs to adjust the volume of the simulation box by multiplying the cartesian coordinates of the atoms with an appropriate value (which is very close to 1) after each time step in the numerical integration scheme. The extent to which this is done again depends on the instantaneous and the desired values; a large difference between the desired pressure and the actual one requires a larger volume adjustment. The pressure is calculated using the virial expression.

2.4 Conformational Search Methods

Conformational analysis consists on the characterization of the structures that a molecule is able to adopt and how these influence its properties. A key component of the conformational analysis is the conformational search, the object of which is to identify the preferred conformations of a molecule, *i.e.* those conformations that determine its behavior. This usually requires the characterization of conformations that are minima on the potential energy surface. For a peptide, due to its high conformational flexibility in solution, there is a so large number of minima on the potential energy surface that is impractical to characterize all them. Specifically, most of the peptides exist in physiological conditions as a mixture of interchangeable conformations with similar energies populated according to the Boltzmann distribution. It is important to remember that the statistical weights of the different conformations involve also entropic contributions. Solvation effects may also be important, and various schemes are now available for calculating the solvation free energy of a conformation, that may be added as an additional term to the intramolecular energy. Under such circumstances, it is often assumed that the native (*i.e.* naturally occurring) conformation is the one with the very lowest value of energy. This conformation is usually referred to as the global minimum. Although the global minimum

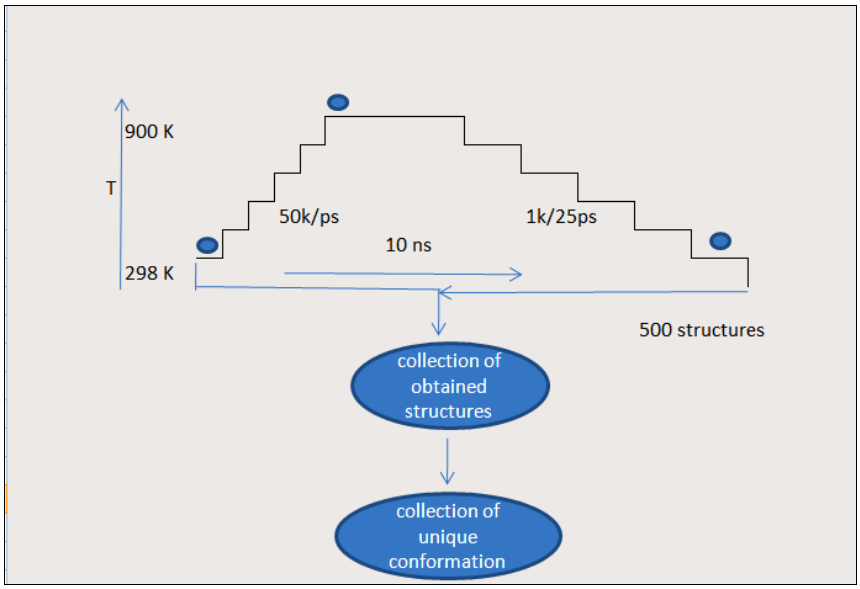
exhibits the lowest energy value, it may not be highly populated because of the contribution of the vibrational entropy to the statistical weight of each structure. Moreover, the global minimum may not be the active (*i.e.* the functional) structure. In this case, it may be even necessary for a molecule to adopt more than one conformation. For example, a substrate might bind in one conformation to an enzyme and then adopt a different conformation prior to reaction is produced. Indeed, in some cases it is possible that the active conformation does not correspond to any minimum on the energy surface of the isolated molecule.

Computational methods for the exploration of the conformational space of a peptide started about thirty years ago[17]. From then different strategies have been described and reviewed [18-20], and, although many efforts have been devoted, this field of research still remains open. Conformational search methods can be divided into the following categories: systematic search algorithms, model-building methods, random approaches, distance geometry and MD. Independently from the strategy selected, four key elements are needed to carry out the exploration of a peptide conformational space. The first consists of employing a peptide model description based on classic mechanics, *i.e.* a force field that permits to calculate the energy of a determined conformation. The second is to find a method capable of generating different conformations, in order to explore all the low energy regions of the conformational space. The third key element consists of minimizing the different conformations, whereas the fourth and last element is to find a convergence criterion to assess if the conformational space has been sufficiently explored.

A conformational search method that has shown to be particularly effective for the exploration of the conformational space of peptides is the iterative simulated annealing[21]. The method has been used in the present thesis work in chapter 4. The simulated annealing method was first described in 1983[22]. This method is based on the similarity that exists between locating the global minimum of the potential energy function of a molecule and the slow cooling required to obtain a perfect crystal (Figure 2.1). In fact, crystal growing will probably be perfect if the system is cooled very slowly by reaching the thermodynamic equilibrium when passing through restrained regions of the phase space. Application of this concept to the exploration of the conformational space can be translated in terms of starting the simulation at a sufficiently high temperature and subsequently

decreasing it gradually until the system is frozen in the global minimum. All the studies carried out using the simulated annealing method have demonstrated that although the cooling scheme is not sufficiently slow to find the global minimum, it is capable to find local energy minima of the regions explored. This means that simulated annealing combined with a searching strategy, which permits to cross different potential energy barriers and to reach the low energy regions, is a very efficient method to explore the conformational space.

Under such circumstances, a protocol based on the simulated annealing method combined with MD (SA-MD) have been used in this work for the exploration of the conformational space of peptides.. This strategy is schematically shown in Figure 2.1. The method, which is particularly efficient in the case of large peptide sequences, consists of performing independent SA-MD cycles and selection of a large number of structures (500) from each SA-MD cycle for energy minimization. This procedure is robust enough to locate the lower-energy minimum structures of the system under study, *i.e.* structures that are quasi-degenerate with the global minimum but situated in different valleys of the peptide landscape. The development of this sampling technique was inspired not only in the work of Filizola *et al.*[21] but also in other recent studies, which demonstrated that very low energies are obtained by minimizing the energy of structures generated at the initial and intermediate states of conventional SA-MD[23,24].



2.1 Schematic diagram of the simulated annealing protocol

2.5 References

1. Binkley, J.S.; Pople, J.A.; Hehre, W.J. *J.Am. Chem. Soc.* **1980**, *102*,939.
2. Hariharan, P.C.; Pople, J.A.; *Theor. Chim. Acta.* **1973**, *28*,213.
3. McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639.
4. Frisch, M.J.; Pople, J.A.; Binkle, J.S. *J. Chem. Phys.* **1984**, *18*, 3265.
5. Møller, C.; Plesst, M.S.; *Phys, Rev.* **1934**, *46*, 618.
6. Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.;Stewart, J.J.P. *J.Am. Chem. Soc.* **1985**, *107*, 3902.
7. Stewart, J.J.P. *J. Comput. Chem.* **1989a,** *10*, 209.
8. Stewart, J.J.P. *J. Comput. Chem.* **1989b**, *10*, 209.
9. Dewar, M.J.S.; Thiel, W. *J.Am. Chem. Soc.* **1977**, *99*, 4899.
10. Hohenberg, P.; Kohn, W. *Phys. Rev. B.* **1964**,*136*, B864.
11. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1993**, *37*, 785.
12. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
13. (a) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999. (b) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027. (c) Miertus, S.; Tomasi, *J. Chem. Phys.* **1982**, *65*, 239. (d) Miertus, M.; Scrocco, E.; Tomasi, *J. Chem. Phys.* **1981**, *55*, 117.
14. Verlet, L. *Phys, Rev.* **1967**, *159*, 98
15. Verlet, L. *Phys, Rev.* **1968**, *165*, 201
16. Ryckaert, J.P.; Ciccotti, G. ; Berendsen. H.J. C. *J. Comput, Phys.,* **1977,** *23*,237
17. (a) Endres, G.F.; Scheraga, H.A. *Biochemistry* **1968,** 7, 4219. (b) Epand, RF.; Scheraga, H.A. *Biochemistry* **1968,** 6, 1383. (c) Epand, R.F.; Scheraga, H.A. *Biochemistry*, **1968,** *6*, 1551 (d) Ingwall, Rt. ; Scheraga, H.A.; Lotan, N. *Biochemistry*, **1968,** *6*, 1968.
18. (a) Howard, A.E; Kollman, P.A. **1988,** *110*, 7195. (b) Howard, A.E; Kollman, P.A. *J.Med. Chem.* **1988,** *31*, 1669.
19. (a) Van Gunsteren, W.F.; Berendsen, H.J.C. *Angew. Chem. Int. Ed.,* **1990,** *29*, 992.

20. (a) Scheraga, H.A. *International Journal of Quantum Chemistry,* **1992**, *42*, 1529. (b) Scheraga, H.A**.** *Prot. Sci.*, **1992**, *1*, 691. (c) Scheraga, H.A. *Abstracts of papers of the American Chemical Society***, 1992,** *203*, 231.
21. (a) Filizola, M.; Perez, J.J; Palomer, A, et al. *J. of Mol. Graph. Model.*, **1997**, *15*, 290. (b) Filizola, M; CarteniFarina, M; Perez, J.J. *Journal of Pept. Res.*, **1997**, *50*, 55. **(**c) Filizola, M; Centeno, N.B; Perez, J.J. *J. of Pept. Sci.*, **1997**, *3*, 85.
22. Kirkpatrick, S.; Gelatt C.D.; Vecchi, M.P. *Science* **1983**, *220*, 671.
23. Baysal, C.; Meirovitch, H. J. Comput. Chem. 1999, 20, 1659.
24. Simmerling, C.; Elber, R. J. Am. Chem. Soc. 1994, 116, 2534.

3. Intrinsic Conformational Properties of synthetic Proline Analogues

3.1 Conformation of Proline Analogs Having Double Bonds in the Ring

The intrinsic conformational preferences of proline analogs having double bonds between carbon atoms in the ring have been investigated using quantum mechanical calculations at the B3LYP/6-31+G(d,p) level. For this purpose, the potential energy surface of the N-acety-N'-methylamide derivatives of three dehydroprolines (proline analogs unsaturated at α,β; β,γ; and γ,δ) and pyrrole (proline analog with unsaturations at both α,β and γ,δ) have been explored, and the results are compared with those obtained for the derivative of the non-modified proline. We found that the double bonds affect the ring puckering and the geometric internal parameters, even though the backbone conformation was influenced the most. Results indicate that the formation of double bonds between carbon atoms in the pyrrolidine ring should be considered as an effective procedure to restrict the conformational flexibility of prolines. Interestingly, we also found that the N-acetyl-N'-methylamide derivative of pyrrole shows a high probability of having a cis peptide bond preceding the proline analog.[*]

3.1.1 Introduction

Among the amino acids that occur in proteins, proline (Pro) is unique due to the presence of a five-membered cycle that produces distinct conformational restrictions, i.e. the dihedral angle φ (N-C^{α} rotation) is restrained to about –60º, and to the lack of amide hydrogen atom, which precludes the participation of the backbone in intramolecular hydrogen bonds. These structural features play a crucial role in the secondary structure of peptides and proteins, with Pro usually involved in special motifs like turns and bends.[1-5]

Pro has two additional important structural characteristics. First, the five-membered cycle, i.e. pyrrolidine ring, may adopt two distinct puckered conformations, known to be almost equally probable from X-ray crystallography of peptides and proteins.[6-9] The down and up ring puckerings are defined as those in which the C^{γ} and C=O group of the Pro residue lie on the same and opposite sides, respectively, of the plane defined by C^{δ}, N and C^{α} atoms. Second, the Pro

[*] *The work described in this chapter previously appeared in J. Phys. Chem. B* **2007**, *111*, 5475-5482

residue relatively often exhibits a cis peptide bond. Analyses of X-ray protein structures show that the peptide bond preceding the Pro has a relatively high intrinsic probability (~ 5-6%) of having a cis arrangement as compared with other amino acids (~0.03%).10-12 The cis-trans isomerization of Pro is often involved in the rate determining steps for folding and refolding of some proteins.[13,14] The unique structural properties of Pro have been examined in detail through both experiments[15-21] and calculations[22-35] on small blocked model peptides, i.e. mono-, di- and tripeptides. Interestingly, gas-phase experiments on Pro indicated that some of the conformational features previously discussed, as for instance the existence of distinct low-energy puckered arrangements, are intrinsic to this amino acid and should not be attributed to the environment when it is contained into a peptide or protein.[20,30]

During the last few years we have devoted a great deal of attention to non-proteogenic amino acids with restricted conformational properties, in particular 1-aminocycloalkanecarboxylic acid derivatives, using theoretical methodologies.[36-40] We used quantum chemical methodologies to study the conformational properties of amino acids with a significant tendency to promote γ- and/or β-turn motifs. Our overall goal has been to enhance the thermodynamic stability of nanotubular structures constructed by self-assembling protein fragments with a β-helical conformation, using targeted replacements in the loop region of the β -helix by amino acids with a high tendency to promote turn conformations.[41-43] Specifically, in some cases we found that the stability of β-helix nanotubes increases significantly by introducing 1-aminocyclopropanecarboxylic acid derivatives.[42,43]

In this work we propose to enhance the intrinsic ability of Pro to adopt turn motifs[1-5] by introducing additional constraints in the cyclic side chain of this amino acid. Specifically, we introduced a double bond between two carbon atoms to generate dehydro amino acids. Unsaturation of the pyrrolidine ring can occur in three different locations: between the C^{α} and C^{β} atoms ($\Delta^{\alpha,\beta}$Pro), C^{β} and C^{γ} atoms ($\Delta^{\beta,\gamma}$Pro), and C^{γ} and C^{δ} atoms ($\Delta^{\gamma,\delta}$Pro). In the present study we consider N-acetyl-N'-methylamide derivatives of these three Pro analogs having a double bond in the ring, hereafter denoted Ac-$\Delta^{\alpha,\beta}$Pro-NHMe, Ac-$\Delta^{\beta,\gamma}$Pro-NHMe and Ac-$\Delta^{\gamma,\delta}$Pro-NHMe, respectively (Figure 3.1.1). The N-acetyl-N'-methylamide analog

with two double bonds located between the C^{α} and C^{β} atoms and C^{γ} and C^{δ} atoms have been also considered (Figure 3.1.1). In the latter compound, abbreviated Ac-Py-NHMe, the pyrrolidine transforms into the pyrrole ring. Modification of the Pro structure by introducing double bonds in the side chain may have some important structural consequences for the following three reasons: (i) the double bond acts as an additional stereochemical constraint; (ii) in some cases (Ac-$\Delta^{\alpha,\beta}$Pro-NHMe and Ac-Py-NHMe) the chirality gets lost; and (iii) in some cases a cross-conjugated system with the peptide bond is formed.

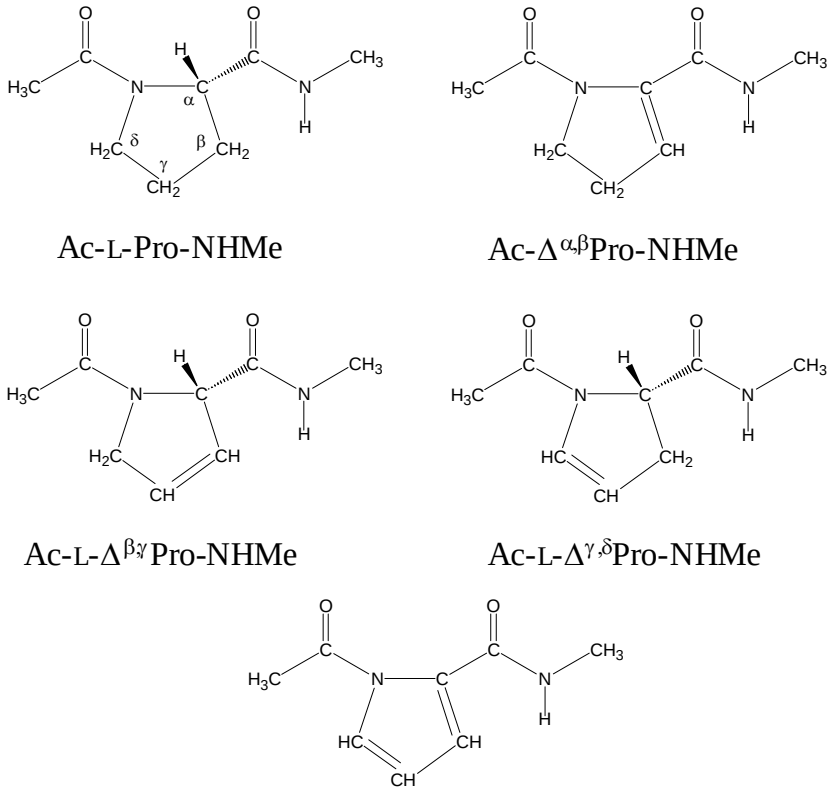


Figure 3.1.1: *Compounds studied in this work*

Although analogs of Pro having double bonds in the ring were synthetized and characterized a long time ago[44] and they have some biomedical applications,[45-47] their conformational preferences have not been reported. In this work we investigate the conformational preferences of Ac-Δα,βPro-NHMe, Ac-L-Δβ,γPro-

NHMe, Ac-L-Δγ,δPro-NHMe and Ac-Py-NHMe using Density Functional Theory (DFT) calculations. The influence of double bonds between carbon atoms on the conformational properties have been examined by comparing the minimum energy conformations of these dipeptides with those of the *N*-acetyl-*N*'-methylamide derivative of Pro (Ac-L-Pro-NHMe), which has been investigated using the same quantum mechanical method. Specifically, we have examined how the double bonds affect the backbone conformation, the puckering of the ring, the internal geometric parameters and the relative stabilities of the cis conformers.

3.1.2 Methods

3.1.2.1 Computational Details

All calculations were carried out using the Gaussian 03 computer program.[48] DFT calculations were performed using the following combination: the Becke´s three-parameter hybrid functional (B3)[49] with the Lee, Yang and Parr (LYP)[50] expression for the nonlocal correlation (B3LYP). Thus, all the calculations presented in this work were performed using the B3LYP method combined with the 6-31+G(d,p) basis set.[51]

The backbone ($\omega_0,\varphi,\psi,\omega$) and side chain ($\chi^j$; endocyclic) dihedral angles of the *N*-acetyl-*N'*-methylamide derivatives of Pro and its analogs studied in this work are defined in Figure 3.1.2. Because φ is severely restrained by the geometry of the five-membered ring, only three minima may be anticipated for the potential energy surfaces $E=E(\psi)$ of the dipeptides if all peptide bonds (ω_0 and ω) are considered in trans conformation.[52] Thus, the flexible backbone dihedral angle ψ is expected to have three minima, *i.e. gauche*$^+$ (60º), *trans* (180º) and *gauche*$^-$ (–60º). Cyclic side chains can adopt two different conformational states, which correspond to the up and down ring puckerings. Accordingly, for each dipeptide with the two peptide bonds arranged in trans, 3(backbone)×2(cyclic side chain)= 6 structures were considered as starting points for complete geometry optimizations at the B3LYP/6-31+G(d,p) level. Frequency analyses were carried out to verify the nature of the minimum state of all the stationary points obtained and to calculate the zero-point vibrational energies (ZPVE) with both thermal and entropic corrections. These statistical terms were used to compute the

conformational Gibbs free energies (ΔG) at the B3LYP/6-31+G(d,p) level. No frequency scale factor has been used for thermochemistry analysis.

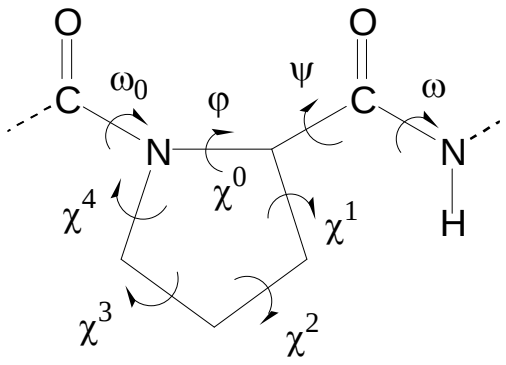


Figure 3.1.2: *Dihedral angles used to identify the conformations of the N-acetyl-N'-methyamide derivatives of Pro and Pro analogs having double bonds between carbon atoms. The dihedral angles ω_0, φ, ψ and ω have been defined using backbone atoms while the atoms of the ring have been employed for the endocyclic dihedral angles χ^i. In particular, the sequence of atoms used to define the dihedral angles φ and χ^0 are C(=O)-N-C^α-C(=O) and C^δ-N-C^α-C^β, respectively.*

Furthermore, all the minimum energy conformations obtained with the two peptide bonds in trans were re-optimized, and subsequently characterized as mimima, following the change in the dihedral angle associated with the first peptide bond (ω_0) from trans to cis. These calculations provided information relating to how the cis-trans isomerization of Pro is affected by the incorporation of double bonds between carbon atoms in the pyrrolidine ring.

3.1.2.2 Nomenclature and Pseudorotational Parameters

The minimum energy conformations of the dipeptides studied in this work have been denoted using a three-labels code that specifies the arrangement of the first peptide bond, the backbone conformation and the puckering of the five membered ring. The first letter refers to the trans (t) or cis (c) arrangement of ω_o. The second label identifies the backbone conformation using the nomenclature introduced by Perczel *et al.*[53] more than fifteen years ago. Accordingly, nine different backbone conformations can be found in the potential energy surface $E=E(\varphi,\psi)$ of amino acids: γ_D, δ_D, α_L, ε_D, β_L, ε_L, α_D, δ_L and γ_L. Finally, the up or down puckering of the five-membered ring is indicated using the labels [u] and [d], respectively. In particular, the [d] ring puckering was identified when χ^1 and χ^3 are positive while χ^2 and χ^4 are negative. Therefore, the [u] ring puckering is characterized by negative values of χ^1 and χ^3 and positive values of χ^2 and χ^4. We

note that the cyclic side chain adopts a planar arrangement in two of the studied molecules, and no indication of the puckering was included in the code used for these cases.

The puckering of the five-membered ring was described using the classical pseudorotational algorithm, which uses a very simple model based on only two parameters, as was previously applied to Pro by Hudaky and Perczel.[28,29] The pseudorotational parameters are A and P, which describe the puckering amplitude and the state of the pucker in the pseudorotation pathway, respectively. The parameters are derived from the endocyclic dihedral angles as follows:

$$A = \sqrt{(A\sin P)^2 + (\chi^0)^2}\ , \text{ where } A\sin P = \frac{\chi^1 - \chi^2 + \chi^3 - \chi^4}{-2(\sin 144^o + \sin 72^o)} \qquad (3.1.1)$$

and

$$P = \begin{cases} \arccos\dfrac{\chi^0}{A}, & \text{if} \quad A\sin P \geq 0 \\ -\arccos\dfrac{\chi^0}{A}, & \text{if} \quad A\sin P < 0 \end{cases} \qquad (3.1.2)$$

Accordingly, parameter A is defined to be positive while P falls between –180º and 180º.

3.1.3 Results and Discussion

3.1.3.1 Ac-L-Pro-NHMe

Considering only trans peptide bonds, three minimum energy conformations have been characterized for Ac-L-Pro-NHMe at the B3LYP/6-31+G(d,p) level (Table 3.1.1). Two of them, the global minimum and the most stable local minimum, correspond to the t-γ_L conformation differing only in the puckering of the pyrrolidine ring (Figures 3.1.3a and 3b). Specifically, the local minimum t-γ_L[u] is 1.3 kcal/mol less stable than the global minimum t-γ_L[d]. Both the t-γ_L[d] and t-γ_L[u] conformations are stabilized by an intramolecular hydrogen bond that

takes place in the seven-membered hydrogen bonded ring. The parameters of this interaction are [d(H···O)= 1.971 Å, ∠N-H···O= 146.4º] and [d(H···O)= 1.984 Å, ∠N-H···O= 143.6º], respectively. Thus, the γ_L conformation is equivalent to the typical γ-turn arrangement. Finally, the third minimum was found to be the t-α_L[u] (Figure 3.1.3c). This conformation, which does not involve any intramolecular hydrogen bond, is unfavored with respect to the t-γ_L[d] by 4.0 kcal/mol. The results presented in Table 3.1.1 are in excellent agreement with those reported by Csizmadia and co-workers[32] and Kang[34b] for the same compound. Thus, these authors found the same three minima using the HF/6-31G(d), HF/6-31+G(d), B3LYP/6-31G(d) and B3LYP/6-311++G(d,p) methods, the relative free energies (ΔG) of the t-γ_L[u] and t-α_L[u] at the latter level of theory being 1.2 and 4.0 kcal/mol, respectively.

The five endocyclic bond angles associated with the pyrrolidine ring and selected bond distances for the three minimum energy conformations of Ac-L-Pro-NHMe are listed in Tables 2 and 3, respectively. These parameters, which will be compared with those obtained for the analogs studied in this work (see below), do not show any significant variation with the conformation.

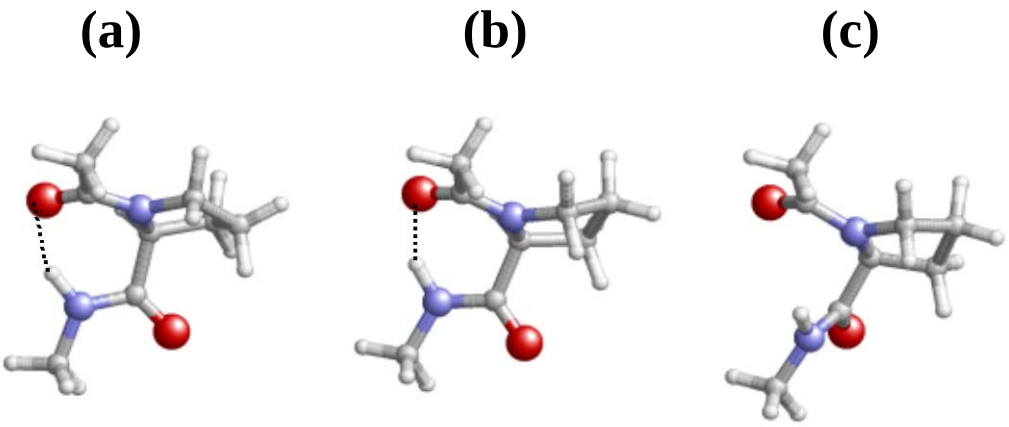


Figure 3.1.3: *Minimum energy conformations of Ac-L-Pro-NHMe at the B3LYP/6-31+G(d,p) level: (a) t-γ_L[d]; (b) t-γ_L[u]; and (c) t-α_L[u].*

Table 3.1.1 *Backbone dihedral angles (in degrees), pseudorotational parameters (A and P, in degrees), relative energy (ΔE; in kcal/mol) and relative free energy (ΔG; in kcal/mol) for the minimum energy conformations of Ac-L-Pro-NHMe with the two peptide bonds in trans calculated at the B3LYP/6-31+G(d,p) level.*

# Conf.	ω_0	φ	ψ	ω	(A, P)	ΔE	ΔG
t-γ_L[d]	-172.6	-83.4	70.3	-177.7	(37.4, -111.9)[a]	0.0[b]	0.0[c]
t-γ_L[u]	-173.9	-81.6	77.3	-175.9	(37.5, 75.8)[d]	1.0	1.3
t-α_L[u]	-171.0	-77.5	-11.5	175.9	(37.8, 89.2)[e]	4.9	4.0

[a] χ^0= -13.9º, χ^1= 31.4º, χ^2= -37.6º, χ^3= 28.7º and χ^4= -9.3º. [b] E= -573.315217 a.u. [c] G= -573.132049 a.u. [d] χ^0= -10.3º, χ^1= -13.4º, χ^2= 31.0º, χ^3= -36.6º and χ^4= 29.8º. [e] χ^0= 0.5º, χ^1= -22.9º, χ^2= 36.1º, χ^3= -35.4º and χ^4= 22.0º.

3.1.3.2 Ac-$\Delta^{\alpha,\beta}$Pro-NHMe

Conformational parameters of the two minimum energy conformations found for Ac-$\Delta^{\alpha,\beta}$Pro-NHMe with the peptide bond in trans are listed in Table 3.1.4. As can be seen, these two minima, which are separated by 2.7 kcal/mol, are significantly different from those previously described for Ac-L-Pro-NHMe. The global minimum corresponds to the t-γ_L[u] (Figure 3.1.4a), which is stabilized by a seven membered intramolecular hydrogen bonded ring with parameters [d(H···O)= 1.722 Å, ∠N-H···O= 156.2º]. The dihedral angles φ,ψ of this minimum are significantly closer to zero than those found for the t-γ_L[d] and t-γ_L[u] conformations of Ac-L-Pro-NHMe. On the other hand, interestingly the seven atoms involved in the intramolecular hydrogen bonded ring of the t-γ_L[u] conformation of Ac-$\Delta^{\alpha,\beta}$Pro-NHMe are almost in the same plane, *i.e.* the hydrogen bonded ring is planar. These striking conformational features have not been detected in other dehydroamino acids.[54,55] For instance, the dihedral angles φ,ψ of the γ_L minimum detected for the N-acetyl-N'-methyl-dehydroalanineamide are –66º,27º, the geometry of the seven-membered hydrogen bonded ring being similar to that detected for Ac-L-Pro-NHMe.[54,55]

Table 3.1.4 *Backbone dihedral angles (in degrees), pseudorotational parameters (A and P, in degrees), relative energy (ΔE; in kcal/mol) and relative free energy (ΔG; in kcal/mol) for the minimum energy conformations of the N-acetyl-N'-methylamide derivatives of the proline analogs having double bonds in the ring calculated at the B3LYP/6-31+G(d,p). In all cases the two peptide bonds are in trans.*

# Conf.	ω_0	φ	ψ	ω	(A, P)	ΔE	ΔG
				Ac-$\Delta^{\alpha,\beta}$Pro-NHMe			
t-γ_L[u]	180.0	-22.1	7.4	179.3	(21.3, 123.7)[a]	0.0[b]	0.0[c]
t-ε_L[u]	179.1	-34.5	126.7	-176.4	(19.3, 125.6)[d]	3.7	2.7
				Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe			
t-γ_L	-172.4	-80.2	67.0	-179.4	(6.7, 164.3)[e]	0.0[f]	0.0[g]
t-α_L	-169.6	-81.9	-6.6	175.6	(2.9, -180.0)[h]	1.9	1.6
				Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe			
t-γ_L[d]	-172.0	-81.3	67.2	-178.0	(10.0,-162.2)[i]	0.0[j]	0.0[k]
				Ac-Py-NHMe			
t-γ	180.0	0.0	0.1	179.9	-[l]	0.0[m]	0.0[n]
t-ε_L	166.5	-24.3	137.8	179.3	(0.4, 80.6)[o]	3.1	2.1

[a] χ^0= -11.8º, χ^1= -1.6º, χ^2= 13.6º, χ^3= -19.6º and χ^4= 19.4º. [b] E= -572.078301 a.u. [c] G= -571.918781 a.u. [d] χ^0= -11.1º, χ^1= -0.9º, χ^2= 11.8º, χ^3= -17.3º and χ^4= 17.8º. [e] χ^0= -6.4º, χ^1= 4.1º, χ^2= -0.3º, χ^3= -3.6º and χ^4= 6.3º. [f] E= -572.078932 a.u.. [g] G= -571.920148 a.u. [h] χ^0= -2.9º, χ^1= 2.4º, χ^2= -1.1º, χ^3= -0.8º and χ^4= 2.4º. [i] χ^0= -9.5º, χ^1= 9.1º, χ^2= -6.1º, χ^3= 0.2º and χ^4= 6.1º. [j] E= -572.084141 a.u.. [k] G= -571.924761 a.u. [l] χ^0= 0.0º, χ^1= 0.0º, χ^2= 0.0º, χ^3= 0.0º and χ^4= 0.0º. In this case no pseudorotational parameter has been provided because the ring is ideally planar. [m] E= -570.873298 a.u.. [n] G= -570.736367 a.u. [o] χ^0= 0.1º, χ^1= -0.3º, χ^2= 0.4º, χ^3= -0.4º and χ^4= 0.2º.

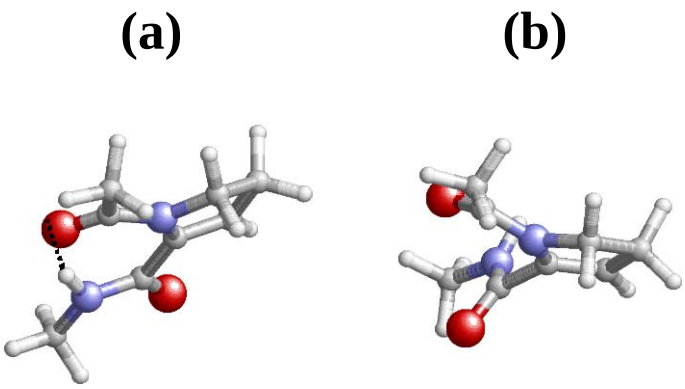


Figure 3.1.4: *Minimum energy conformations of Ac-$\Delta^{\alpha,\beta}$Pro-NHMe at the B3LYP/6-31+G(d,p) level: (a) t-γ_L[u]; and (b) t-ε_L[u].*

The most relevant characteristic of the second minimum, t-ε_L[u] (Figure 3.1.4b), is that the two amide groups are arranged perpendicularly with respect to each other. On the other hand, the extraction of two hydrogen atoms alters not only the conformational preferences of the backbone but also the puckering of the cyclic side chain. Thus, the pseudorotational pathway obtained for the minima of Ac-$\Delta^{\alpha,\beta}$Pro-NHMe is different from those calculated for the minima of Ac-L-Pro-NHMe. Furthermore, the puckering amplitude is significantly smaller for the minima of the former dipeptide than for those of the latter one. On the other hand, it should be noted that due to the achiral nature of Ac-$\Delta^{\alpha,\beta}$Pro-NHMe the t-γ_D[d] and t-ε_D[d] are degenerated minima of those mentioned above.

Inspection of the geometric parameters listed in Table 3.1.2 indicates that, as expected, the bond angles $\angle$N-C^{α}-C^{β} and $\angle C^{\alpha}$-C^{β}-C^{γ} are about 5º-8º larger for Ac-$\Delta^{\alpha,\beta}$Pro-NHMe than for Ac-L-Pro-NHMe due to the double bond between C^{α} and C^{β}. Furthermore, $\angle C^{\delta}$-N-C^{α} is a few degrees smaller in the former dipeptide than in the latter one. This suggests a change in the conjugation pattern of the modified residue. Thus, inspection of the bond lengths displayed in Table 3.1.3 for Ac-$\Delta^{\alpha,\beta}$Pro-NHMe indicates that the peptide bond extends the conjugation to the double bond of the cycle. d(N-C^{α}) and d(C^{α}-C_{XX}) are significantly smaller than in Ac-L-Pro-NMe, while the value of d(C^{α}-C^{β}) is slightly larger than the value typically expected for a C_{sp2}=C_{sp2} bond, *i.e.* 1.332 Å for C_2H_4 at the B3LYP/6-31+G(d,p) level.

Table 3.1.2 *Selected angles (in degrees) for the minimum energy conformations of the N-acetyl-N'-methyl derivatives of proline and its analogs having double bonds in the ring characterized at the B3LYP/6-31+G(d,p) level.*

	$\angle$N-C$^{\alpha}$-C$^{\beta}$	$\angle$C$^{\alpha}$-C$^{\beta}$-C$^{\gamma}$	$\angle$C$^{\beta}$-C$^{\gamma}$-C$^{\delta}$	$\angle$C$^{\gamma}$-C$^{\delta}$-N	$\angle$C$^{\delta}$-N-C$^{\alpha}$
			Ac-L-Pro-NHMe		
t-γ_L[d]	103.1	103.6	103.6	103.7	112.1
t-γ_L[u]	104.4	105.4	103.1	102.8	110.7
t-α_L[u]	104.0	104.2	102.9	103.3	111.5
			Ac-$\Delta^{\alpha,\beta}$Pro-NHMe		
t-γ_L[u]	109.7	112.1	102.1	104.5	107.2
t-ε_L[u]	111.3	110.4	102.5	103.8	108.5
			Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe		
t-γ_L	102.1	111.8	111.7	102.4	111.8
t-α_L	102.3	111.7	111.6	102.5	111.6
			Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe		
t-γ_L[d]	104.0	103.7	110.7	111.8	108.8
			Ac-Py-NHMe		
t-γ	106.5	109.5	107.2	109.3	109.3
t-ε_L	107.6	108.5	107.3	108.5	108.5

Table 3.1.3 *Selected distances[a] (in Å) for the minimum energy conformations of the N-acetyl-N'-methyl derivatives of proline and its analogs having double bonds in the ring characterized at the B3LYP/6-31+G(d,p) level.*

	C_{Ac}=O	C_{Ac}-N	N-C^{α}	C^{α}-C_{XX}	C^{α}-C^{β}	C^{β}-C^{γ}	C^{γ}-C^{δ}	C^{δ}-N
Ac-L-Pro-NHMe								
t-γ_L[d]	1.241	1.360	1.483	1.553	1.533	1.538	1.536	1.4770
t-γ_L[u]	1.241	1.361	1.485	1.554	1.543	1.540	1.533	1.471
t-α_L[u]	1.241	1.375	1.480	1.534	1.550	1.535	1.534	1.476
Ac-$\Delta^{\alpha,\beta}$Pro-NHMe								
t-γ_L[u]	1.239	1.367	1.449	1.520	1.343	1.500	1.539	1.490
t-ε_L[u]	1.228	1.378	1.415	1.507	1.342	1.514	1.551	1.481
Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe								
t-γ_L	1.240	1.360	1.484	1.557	1.506	1.333	1.504	1.476
t-α_L	1.230	1.375	1.476	1.545	1.510	1.334	1.505	1.478
Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe								
t-γ_L[d]	1.239	1.366	1.493	1.558	1.549	1.509	1.336	1.417
Ac-Py-NHMe								
t-γ	1.223	1.403	1.433	1.508	1.374	1.425	1.363	1.404
t-ε_L	1.212	1.422	1.397	1.497	1.376	1.431	1.370	1.394

[a] C_{Ac} and C_{XX} denote the carbon atoms of the carbonyl groups in the acetyl and the Pro residue (or its analogue), respectively.

3.1.3.3 Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe

Two minimum energy conformations have been characterized for Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe when the two peptide bonds are in trans (Table 3.1.4). The lowest energy minimum corresponds to a γ_L backbone conformation (Figure 3.1.5a) with backbone dihedral angles similar to those of the global minimum of Ac-L-Pro-NHMe. The hydrogen bonding parameters associated with the stabilizing intramolecular interaction found in the t-γ_L conformation are [d(H···O)= 1.939 Å, ∠N-H···O= 147.0º]. The second minimum was found to be t-α_L (Figure 3.1.5b), which was previously detected in Ac-L-Pro-NHMe but not in Ac-$\Delta^{\alpha,\beta}$Pro-NHMe. This structure is destabilized by 1.6 kcal/mol with respect to the global minimum. On the other hand, a detailed inspection of Figure 3.1.5 indicates that the cyclic side chain adopts an almost planar arrangement in the two minima. This is confirmed by the low puckering amplitude parameters calculated from the endocyclic dihedral angles (see Table 3.1.4).

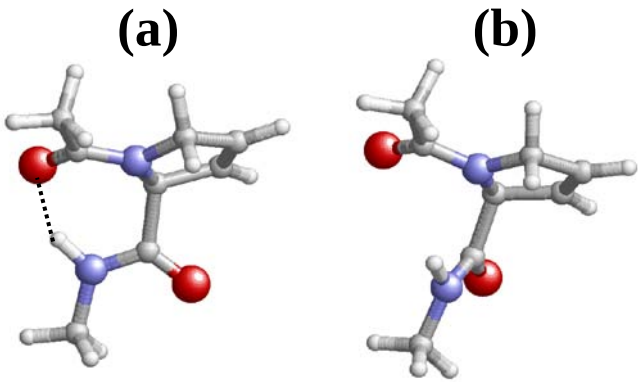


Figure 3.1.5: *Minimum energy conformations of Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe at the B3LYP/6-31+G(d,p) level: (a) t-γ_L; and (b) t-α_L.*

On the other hand, Table 3.1.2 indicates that only the angles centered at the C^{β} and C^{γ} atoms differ from those obtained for Ac-L-Pro-NHMe, with no resonance between the backbone amide group and the side chain double bond being detected. The latter feature is fully consistent with the bond lengths listed in Table 3.1.3. Specifically, d(N-C^{α}) and d(C^{α}-C_{XX}) are very similar to those found for Ac-L-Pro-NMe, while the d(C^{β}-C^{γ}) is larger than the d(C^{α}-C^{β}) calculated for Ac-$\Delta^{\alpha,\beta}$Pro-NHMe. Tables 5 analyzes the relative stabilities between Ac-$\Delta^{\alpha,\beta}$Pro-NHMe and Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe isomers in terms of energies and free energies. The t-γ_L conformation of the latter isomer is 0.9 kcal/mol more stable than the t-γ_L[u] of the former one.

3.1.3.4 Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe

The first noticeable result for this isomer is that only one minimum energy conformation was found when the two peptide bonds are arranged in trans. This corresponds to a γ_L (Figure 3.1.6) conformation, which is characterized by a seven-membered hydrogen bonded ring with the parameters [d(H···O)= 1.938 Å, ∠N-H···O= 147.3º]. Thus, the position of the double bond restricts the conformational flexibility of the Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe with respect to the other isomers. The structural parameters listed in Table 3.1.4 indicate that the ring presents an incipient [d] puckering, which is manifested by the low value of A.

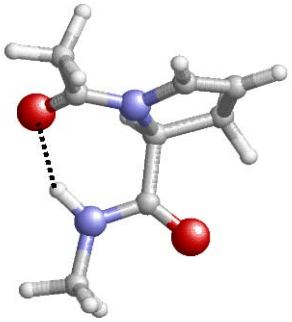


Figure 3.1.6: *Minimum energy conformation (t-γ_L[d]) of Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe at the B3LYP/6-31+G(d,p) level.*

Inspection of the bond angles and distances listed in Tables 2 and 3 suggests that the partial sp^2 character of the amide nitrogen is slightly smaller in Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe than in the dehydroproline isomers presented above. This is particularly evidenced by the d(N-C^{α}) bond length, which is even larger than those found for the minimum energy conformations of Ac-L-Pro-NHMe. Inspection of Table 3.1.5 reveals that Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe is the most stable N-acetyl-N'-methylamide derivative of Pro analogs having one double bond in the ring. Thus, the global minimum of Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe and Ac-$\Delta^{\alpha,\beta}$Pro-NHMe is destabilized by 2.9 and 3.8 kcal/mol, respectively, with respect to the t-γ_L[d] conformation of Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe.

Table 3.1.5 *Relative stability at the B3LYP/6-31+G(d,p) among the three isomers calculated in this work. Relative energy (ΔE; in kcal/mol) and relative free energy (ΔG; in kcal/mol) of the minimum energy conformations obtained for the N-acetyl-N'-methylamide derivatives of proline analogs having one double bond in the ring.*

Compound		ΔE	ΔG
Ac-$\Delta^{\alpha,\beta}$Pro-NHMe	t-γ_L[u]	3.7	3.8
	t-ε_L[u]	7.4	6.4
Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe	t-γ_L	3.3	2.9
	t-α_L	5.2	4.5
Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe	t-γ_L[d]	0.0	0.0

3.1.3.5 Ac-Py-NHMe

As can be seen in Table 3.1.4, the conformational preferences of Ac-Py-NHMe are relatively similar to those found for Ac-$\Delta^{\alpha,\beta}$Pro-NHMe. Thus, two minimum energy conformations were characterized when the two peptide bonds are arranged in trans. The φ,ψ dihedral angles of the global minimum are zero, which represents a small but non-negligible reduction of the dihedrals found for the global minimum of Ac-$\Delta^{\alpha,\beta}$Pro-NHMe. As evidenced in Figure 3.1.7a, this minimum presents a perfectly planar seven-membered hydrogen bonded ring. The corresponding geometric parameters, [d(H···O)= 1.767 Å, ∠N-H···O= 153.5º], suggests that the intramolecular hydrogen bond is very strong in this case. Furthermore, the global minimum of Ac-Py-NHMe, which is not found in proteogenic amino acids, is between the γ_L and γ_D arrangements. Accordingly, this conformation has been denoted hereafter t-γ, *i.e.* the L or D character typically attributed to the γ conformation has been omitted. The second minimum is the t-ε_L (Figure 3.1.7b), which is disfavored by 2.1 kcal/mol with respect to the t-γ. The most important difference between Ac-Py-NHMe and Ac-$\Delta^{\alpha,\beta}$Pro-NHMe was found to be the arrangement of the cyclic side chain. Thus, as expected, the pyrrole ring adopts a planar conformation in the two minima. On the other hand, it

should be noted that t-ε_D is the only degenerated minimum expected for Ac-Py-NHMe because of the particular dihedral angles found for the t-γ global minimum.

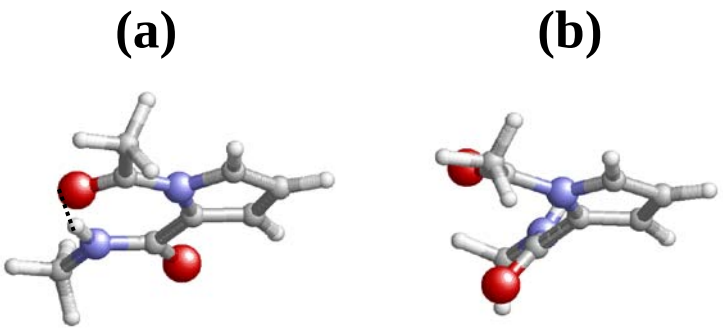


Figure 3.1.7: *Minimum energy conformation of Ac-Py-NHMe at the B3LYP/6-31+G(d,p) level: (a) t-γ; and (b) t-ε_L.*

On the other hand, inspection of Table 3.1.2 indicates that the bond angles are essentially those expected for a planar pyrrole ring, no significant conformational dependence being detected. Indeed, they do not differ too much from those calculated for Ac-L-Pro-NHMe with the exception of $\angle C^{\gamma}$-C^{δ}-N, which is about 6º larger in Ac-Py-NHMe. However, bond lengths illustrate an opposite behavior in Table 3.1.3. Thus, the bond length between the carbonyl carbon atom of the Ac group and the nitrogen of the pyrrole ring, d(C_{Ac}-N) is very large, whereas the d(N-C^{α}) and d(C^{α}-C_{XX}) are very short. These features combined with the values of the endocyclic bonds lengths suggest that conjugation effects extend from the pyrrole ring to the dipeptide backbone. Furthermore, the differences found between the bond lengths of the t-γ and t-ε_L minima indicate that this electronic process depends on the conformation.

3.1.3.6 Relative Stability of the Cis Conformers

The dihedral angle ω_0 of the three minimum energy conformations characterized for Ac-L-Pro-NHMe was changed from the values displayed in Table 3.1.1 to 0º. The resulting conformations were used as starting points for full geometry optimizations at the B3LYP/6-31+G(d,p) level. The conformational parameters of the new minima are listed in Table 3.1.6. Results indicate that the starting conformations c-γ_L[d] and c-γ_L[u] evolve towards two completely different conformations. These are the c-α_L[d] and c-ε_L[u], which are disfavored with respect to the global minimum of Ac-L-Pro-NHMe (t-γ_L[d] in Table 3.1.1) by 2.3 and 5.0 kcal/mol, respectively. On the other hand, the starting c-α_L[u] conformation was retained as energy minimum after complete geometry

optimization. This conformation is 3.6 kcal/mol less stable than the t-γ_L[d]. However, the most noticeable feature is that c-α_L[u] is slightly more stable than t-α_L[u], *i.e.* 0.4 kcal/mol. These results are in excellent agreement with those reported by Csizmadia and co-workers[32] and Kang[34b] at similar levels of theory.

Table 3.1.6 *Backbone dihedral angles (in degrees), pseudorotational parameters (A and P, in degrees), relative energy (ΔE; in kcal/mol) and relative free energy (ΔG; in kcal/mol) for the minimum energy conformations of Ac-L-Pro-NHMe with the first peptide bond in cis calculated at the B3LYP/6-31+G(d,p) level. The initial conformation[a] (see text) is also indicated.*

Conformation # Starting.	# Final	ω_0	φ	ψ	ω	(A, P)	ΔE^b	ΔG^b
c-γ_L[d]	c-α_L[d]	10.2	-90.9	-5.1	-179.6	(37.5, -111.6)[c]	3.3	2.3
c-γ_L[u]	c-ε_L[u]	-0.1	-61.2	145.3	177.5	(37.4, 88.0)[d]	6.6	5.0
c-α_L[u]	c-α_L[u]	8.00	-79.2	-18.4	-177.1	(37.5, 89.8)[e]	4.2	3.6

[a] The initial conformations were generated by changing from trans to cis the dihedral angle ω_0 of the minimum energy conformations listed in Table 3.1.1. [b] ΔE and ΔG are relative to the t-γ_L[d] conformation of Ac-L-Pro-NHMe (Table 3.1.1). [c] χ^0= -13.8º, χ^1= 31.4º, χ^2= -37.6º, χ^3= 29.0º and χ^4= -9.5º. [d] χ^0= 1.3º, χ^1= -23.3º, χ^2= 36.0º, χ^3= -34.6º and χ^4= 21.2º. [e] χ^0= -0.1º, χ^1= -22.1º, χ^2= 35.7º, χ^3= -35.2º and χ^4= 22.3º.

The same strategy was applied to the *N*-acetyl-*N'*-methyl derivatives of proline analogs having double bonds in the ring, with results displayed in Table 3.1.7. For all dipeptides, geometry optimization of the starting point derived from the global minimum, *i.e.* modifying the dihedral angle ω_0, leads to a new minimum with different backbone conformation that is at least 2.4 kcal/mol less stable than the lowest energy minimum (Table 3.1.4). On the other hand, optimization of the conformations generated by changing the local minimum of Ac-$\Delta^{\alpha,\beta}$Pro-NHMe and Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe produce structures with the same backbone conformation but more stable than those with the two peptide bonds arranged in trans. Thus, the relative free energies of the t-ε_L[u] and c-ε_L[u] minima of Ac-$\Delta^{\alpha,\beta}$Pro-NHMe are 2.7 and 1.0 kcal/mol, respectively, while the relative free energies of the t-α_L and c-α_L minima of Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe are 1.6 and 1.1 kcal/mol.

Table 3.1.7 *Backbone dihedral angles (in degrees), pseudorotational parameters (A and P, in degrees), relative energy (ΔE; in kcal/mol) and relative free energy (ΔG; in kcal/mol) for the minimum energy conformations of the N-acetyl-N'-methyl derivatives of proline analogs having double bonds in the ring with the first peptide bond in cis. The initial conformation (see text) is also indicated.*[a]

Conformation # Starting.	# Final	ω_0	φ	ψ	ω	(A, P)	ΔE	ΔG
				Ac-$\Delta^{\alpha,\beta}$Pro-NHMe[b]				
c-γ_L[u]	c-α_L[u]	11.7	-55.3	-29.0	-173.4	(23.1, 127.0)[c]	3.2	2.4
c-ε_L[u]	c-ε_L[u]	-0.4	-27.8	140.9	175.3	(16.8, 125.6)[d]	2.3	1.0
				Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe[e]				
c-γ_L	c-ε_L	-0.5	-66.0	155.7	177.9	(6.3, 172.8)[f]	4.7	3.2
c-α_L	c-α_L	9.5	-83.0	-14.6	-176.0	(2.8,-175.2)[g]	1.4	1.1
				Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe[h]				
c-γ_L[d]	c-α_L[d]	1.3	-77.9	-17.1	-176.5	(11.5,-153.6)[i]	3.4	2.4
				Ac-Py-NHMe[j]				
c-γ	c-α_L	-11.4	-29.7	-39.7	-173.6	(0.2,180.0)[k]	4.1	3.2
c-ε_L	c-ε_L	-12.6	-21.1	151.8	175.0	(0.4,90.0)[l]	0.7	0.0

[a] *The initial conformations were generated by changing from trans to cis the dihedral angle ω_0 of the minimum energy conformations listed in Table 3.1.4.* [b] *ΔE and ΔG are relative to the t-γ_L[u]conformation of Ac-$\Delta^{\alpha,\beta}$Pro-NHMe (Table 3.1.4).* [c] *χ^0= -13.9º, χ^1= -0.4º, χ^2= 13.8º, χ^3= -20.9º and χ^4= 21.6º.* [d] *χ^0= -11.8º, χ^1= -1.0º, χ^2= 11.9º, χ^3= -17.4º and χ^4= 17.9º.* [e] *ΔE and ΔG are relative to the t-γ_L conformation of Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe (Table 3.1.4).* [f] *χ^0= -6.3º, χ^1= 4.5º, χ^2= -1.2º, χ^3= -2.6º and χ^4= 5.6º.* [g] *χ^0= -2.8º, χ^1= 2.3º, χ^2= -1.1º, χ^3= -0.6º and χ^4= 2.2º.* [h] *ΔE and ΔG are relative to the t-γ_L[d]conformation of Ac-$\Delta^{\gamma,\delta}$Pro-NHMe (Table 3.1.4).* [i] *χ^0= -10.3º, χ^1= 10.8º, χ^2= -8.3º, χ^3= 2.1º and χ^4= 5.6º.* [j] *ΔE and ΔG are relative to the t-γ conformation of Ac-Py-NHMe (Table 3.1.4).* [k] *χ^0= -0.2º, χ^1= 0.1º, χ^2= 0.0º, χ^3= -0.1º and χ^4= 0.2º.* [l] *χ^0= 0.0º, χ^1= -0.2º, χ^2= 0.4º, χ^3= -0.3º and χ^4= 0.2º.*

The most noticeable feature has been found for the Ac-Py-NHMe dipeptide. Thus, geometry optimization of the generated c-ε_L conformation led to a minimum with similar backbone conformation, which is not only more stable that

the t-ε_L local minimum but also isoenergetic with the t-γ global minimum. This indicates that the intrinsic probability of having a cis peptide bond is higher in peptides containing the analog with two double bonds than in peptides containing unmodified Pro or analogs with one double bond.

Finally, it should be mentioned that analyses of the bond lengths and angles (data not shown) for the minima with one peptide bond in cis do not provide any finding different from those discussed above.

3.1.4 Conclusions

DFT calculations on Ac-L-Pro-NHMe, Ac-$\Delta^{\alpha,\beta}$Pro-NHMe, Ac-L-$\Delta^{\beta,\gamma}$Pro-NHMe, Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe and Ac-Py-NHMe at the B3LYP/6-31+G(d,p) level allow us to draw the following conclusions about the conformational impact of the incorporated double bonds:

(i) The incorporation of double bonds at the ring is an effective way to reduce the conformational flexibility. Interestingly, the results obtained for Ac-L-$\Delta^{\gamma,\delta}$Pro-NHMe, which only presents one minimum energy conformation when the two peptide bonds are in trans, indicate that the analog γ,δ-unsaturated is the most constrained one.

(ii) The backbone conformation of Pro is clearly affected by the incorporation of double bonds in the ring, especially when the two peptide bonds are arranged in trans. This is particularly evidenced in Ac-$\Delta^{\alpha,\beta}$Pro-NHMe and Ac-Py-NHMe, where the φ,ψ values found for the lowest minimum energy conformation found of these dipeptides are significantly lower than those obtained for Ac-L-Pro-NHMe.

(iii) Some of the *N*-acetyl-*N'*-methyl derivatives of Pro analogs having double bonds in the ring, in particular Ac-$\Delta^{\alpha,\beta}$Pro-NHMe and Ac-Py-NHMe, show interesting electronic effects. The geometric parameters of these dipeptides evidence conjugation between the backbone amide group and the side chain double bonds.

(iv) For the *N*-acetyl-*N'*-methyl derivatives of Pro and its analogs having one double bond in the ring, the conformations with one peptide bond in cis are less stable than those with the two peptide bonds in trans.

However, the global minimum of Ac-Py-NHMe, t-γ, is isoenergetic with the c-ε_L conformation. Accordingly, a high probability of having a cis peptide bond is expected for peptides containing Py. This is a very important feature for both protein engineering and nanoconstruct design.

3.1.5 References

1. Mac Arthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397.
2. Gibbs, A. C.; Bjorndahl, T. C.; Hodges, R. S.; Wishart, D. S. *J. Am. Chem. Soc.* **2002**, *124*, 1203.
3. Quancard, J.; Karoyan, P.; Lequin, O.; Wenger, E.; Aubry, A.; Lavielle. S.; Chassaing, G. *Tetrahedron Lett.* **2004**, *45*, 623.
4. Némethy, G.; Printz, M. P. *Macromolecules* **1972**, *6*, 755.
5. Richardson, J. S.; Richardson, D. C. "Principles and patterns of protein conformation"; in: Prediction of Protein Structure and the Principles of Protein Conformation, Fasman, G.D. (Ed.), Plenum, New York, **1989**, 98.
6. Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. *Protein Sci.* **2001**, *10*, 2632.
7. Milner-White, E. J.; Bell, L. H.; Maccallum, P. H. *J. Mol. Biol.* **1992**, *228*, 725.
8. DeTar, D. F.; Luthra, N. P. *J. Am. Chem. Soc.* **1977**, *99*, 1232.
9. Balasubramanian, R.; Lakshminarayanan, A. V.; Sabesan, M. N.; Tegoni, G.; Venkatesan, K. Ramachandran, G. N. *Int. J. Pept. Protein Res.* **1971**, *3*, 25.
10. Stewart, D. E.; Sarkar, A.; Wampler, J. E. *J. Mol. Biol.* **1990**, *214*, 253.
11. Jabs, A.; Weiss, M. S.; Hilgenfeld, R. *J. Mol. Biol.* **1999**, *286*, 291.
12. Pal, D.; Chakrabarti, P. *J. Mol. Biol.* **1999**, *294*, 271.
13. Wedemeyer, W. J.; Welker, E.; Scheraga, H. A. *Biochemistry* **2002**, *41*, 14637.
14. Dugave, C.; Demange, L. *Chem. Rev.* **2003**, *103*, 2475.
15. Madison, V.; Kopple, K. D. *J. Am. Chem. Soc.* **1980**, *102*, 4855.
16. Madison, V.; Schellman, J. *Biopolymers* **1970**, *9*, 511.

17. Matsuzaki, T.; Iitaka, Y. *Acta Crystallogr., Sect. B* **1971**, *27*, 507.
18. Delaney, N. G.; Madison, V. *J. Am. Chem. Soc.* **1982**, *104*, 6635.
19. Beausoleil, E.; Lubell, W. D. *J. Am. Chem. Soc.* **1996**, *118*, 12902.
20. Lesarri, A.; Mata, S.; Cocinero, E. J.; Blanco, S.; Lopez, J. C.; Alonso, J. L. *Angew. Chem*. **2002**, *114*, 4867; *Angew. Chem. Int. Ed.* **2002**, *41*, 4673.
21. Taylor, C. M.; Hardre, R.; Edwards, P. J. B.; Park, J. H. *Org. Lett.* **2003**, *5*, 4413.
22. Zimmerman, S. S.; Pottle, M. S.; Nemethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1.
23. Fischer, S.; Dunbrack, R. L., Jr.; Karplus, M. *J. Am. Chem. Soc.* **1994**, *116*, 11931.
24. Kang, Y. K. *J. Phys. Chem.* **1996**, *100*, 11589.
25. Improta, R.; Benzi, C.; Barone, V. *J. Am. Chem. Soc.* **2001**, *123*, 12568.
26. Benzi, C.; Improta, R.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2002**, *23*, 341.
27. Kang, Y. K. *J. Phys. Chem. B* **2002**, *106*, 2074.
28. Hudáky, I.; Baldoni, H. A.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2002**, *582*, 233.
29. Hudáky, I.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2003**, *630*, 135.
30. Allen, W.D.; Czinki, E.; Császár, A. G. *Chem. Eur. J.* **2004**, *10*, 4512.
31. Kang, Y. K.; Park, H. S. *J. Mol. Struct. (Theochem)* **2005**, *718*, 17.
32. Sahai, M.A.; Kehoe, A. K.; Koo, J. C. P.; Setiadi, D. H.; Chass, G. A.; Viskolcz, B.; Penke, B.; Pai, E. F.; Csizmadia, I. G. *J. Phys. Chem. A* **2005**, *109*, 2660.
33. Kang, Y. K.; Park, H. S. *Biophys. Chem.* **2005**, *113*, 93.
34. (a) Kang. Y. K., Jhon, J. S.; Park, H. S. *J. Phys. Chem. B* **2006**, *110*, 17645. (b) Kang, Y. K. *J. Phys. Chem. B* **2006**, *110*, 21338. (c) Kang, Y. K. *J. Mol. Struct.* **2004**, *675*, 37. (d) Kang, Y. K.; Choi, H. Y. *Biophys. Chem.* **2004**, *111*, 135.
35. Czinki, E.; Csaszar, A. G. *Chem. Eur. J.* **2003**, *9*, 1008.
36. Alemán, C.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Casanovas, J. *J. Phys. Chem. B* **2002**, *106*, 11849.
37. Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *Chem. Phys. Lett.* **2006**, *429*, 558.

38. Alemán, C.; Zanuy, D.; Casanovas, J.; Cativiela, C.; Nussinov, J. *J. Phys. Chem. B* **2006**, *110*, 21264.
39. Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Alemán, C. *J. Org. Chem.* **2003**, *68*, 7088.
40. Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Alemán, C. *J. Phys. Chem. B* **2006**, *110*, 5762.
41. Haspel, N.; Zanuy, D.; Alemán, C.; Wolfson, H.; Nussinov, R. *Structure* **2006**, *14*, 1137.
42. Zheng, J.; Zanuy, D.; Haspel, N.; Tasi, C.-J.; Alemán, C.; Nussinov, R. *Biochemistry* **2007**, *46*, 1205.
43. Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2007**, in press.
44. Mauger, A. B.; Witkop, B. *Chem. Rev.* **1966**, *66*, 47.
45. Moon, S. J.; Han, C. T. *J. Biochem. Mol. Biol.* **1998**, *31*, 201.
46. Li, D. X.; Hirsila, M.; Koiyunen, P.; Brenner, M. C; Xu, L.; Yang, C.; Kivirikko, K. I.; Myllyhariu, J. *J. Biol. Chem.* **2004**, *279*, 55051.
47. Hacker, M. P.; Newman, R. A.; Hong, C. B. *Tox. Appl. Pharmacol.* **1983**, *69*, 102.
48. Gaussian 03, Revision B.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr., T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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.; C. Strain, M.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.;

Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Pittsburgh PA, 2003.

49. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
50. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1993**, *37*, 785.
51. McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639.
52. Baldoni, H. A.; Rodriguez, A. M.; Zamarbide, G.; Enriz, R. D.; Farkas, O¨ .; Csaszar, P.; Torday, L. L.; Sosa, C. P.; Jakli, I.; Perczel, A.; Hollosi, M.; Csizmadia, I. G. *J. Mol. Struct. (THEOCHEM)* **1999**, *465*, 79.
53. Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L-; Marcoccia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, *113*, 6256.
54. Alemán, C.; Casanovas, J. *Biopolymers* **1995**, *36*, 71.
55. Thormann, M.; Hofmann, H.-J. *J. Mol. Struct. (THEOCHEM)* **1998**, *431*, 79.

3.2 Conformational Preferences of α-Substituted Proline Analogues

DFT calculations at the B3LYP/6-31+G(d,p) level have been used to investigate how the replacement of the α hydrogen by a more sterically demanding group affects the conformational preferences of proline. Specifically, the N-acetyl-N'-methylamide derivatives of L-proline, L-α-methylproline and L-α-phenylproline have been calculated, with both the cis/trans isomerism of the peptide bonds and the puckering of the pyrrolidine ring being considered. The effects of solvation have been evaluated using a Self Consistent Reaction Field model. As expected, tetrasubstitution at the α carbon destabilizes the conformers with one or more peptide bonds arranged in cis. The lowest energy minimum has been found to be identical for the three compounds investigated, but important differences are observed regarding other energetically accessible backbone conformations. The results obtained provide evidence that the distinct steric requirements of the substituent at C^{α} may play a significant role in modulating the conformational preferences of proline.[*]

3.2.1 Introduction

The incorporation of conformationally constrained amino acids into a peptide chain is a powerful tool to reduce its intrinsic flexibility. Among the residues whose structural rigidity can be exploited in the design of peptides with well-defined backbone conformations are α-tetrasubstituted α-amino acids.[1]

The simplest α-tetrasubstituted analogue of a proteinogenic amino acid that can be considered is that resulting from the replacement of the α hydrogen by a methyl group. In the last two decades, extensive efforts have been directed at the development of efficient methodologies for the synthesis of the α-methyl derivatives of all genetically coded amino acids[2] (glycine excluded, since it leads to alanine). The simplest one is α-methylalanine (α-aminoisobutyric acid, Aib), whose conformational properties have been deeply investigated and are well established.[1,3,4] In comparison, the α-methylated analogues of all other

[*] *The work described in this chapter previously appeared in J. Org. Chem,* 2008, *73,* 3418-3427

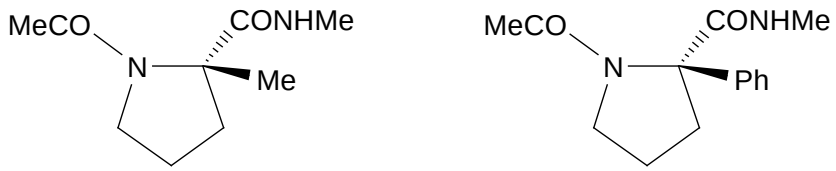
proteinogenic amino acids have been much less studied, mainly due to synthetic difficulties: α-methylation of Ala gives rise to a symmetric achiral residue, whereas two enantiomeric forms are possible for all other residues. Although not as extensively as for Aib, the study of the conformational properties of the α-methyl derivatives of other proteinogenic amino acids (mainly valine, leucine and phenylalanine) has been addressed.[1c,e,5] In general, these α-methylated residues behave as the prototype Aib, although they present particular conformational features derived from their chiral nature.

The unique properties of proline make the study of its α-methylated derivative (in general, α-substituted analogues) particularly intriguing. The singularity of proline lies in its cyclic structure, which includes the amino function. As a consequence, rotation about the N—C^{α} bond is prohibited and the φ torsion angle is confined to values around –60º. Accordingly, proline is overwhelmingly found in the α-helical [$(\varphi,\psi) \approx (-60º,-30º)$] and semi-extended [$(\varphi,\psi) \approx (-60º,140º)$] regions of the conformational map.[6] In addition, proline shows a higher propensity to promote γ-turn conformations [$(\varphi,\psi) \approx (-70º,60º)$] than other proteinogenic amino acids.[6d,7] Another effect derived from its cyclic structure is that the peptide bond preceding proline (that involving the pyrrolidine nitrogen) has a relatively high probability of accommodating a *cis* arrangement[8] as compared to other peptide bonds, for which the *cis* form is almost inexistent. Recent studies in proline dipeptides evidenced that the *cis*/*trans* isomerization is a enthalpy driven process that depends on the polarity of the environment.[9] Thus, although the electronic effects that stabilize the *cis* form become enhanced in polar environments, the *cis*/*trans* rotational barriers increases with the polarity of the enviroment.

Due to its particular structural properties, proline plays a key role in the structure and biology of peptides and proteins, and, hence, α-substituted derivatives are of great interest. The conformational preferences of the α-methylated analogue (αMePro) remain little explored.[10,11] Studies on the *N*-acetyl-*N*'-methylamide derivative indicated a preference for the γ-turn conformation in solution,[10c,d] whereas an α-helical structure was found in the solid state.[10b] Spectroscopic and computational studies on other peptides containing αMePro suggested a stabilization of the βI-turn in comparison with proline.[11] In contrast to

the scarce structural studies, the large number of papers[11,12] and patents[13] dealing with the incorporation of αMePro into bioactive peptides and other biologically relevant systems provide evidence for the enormous potential of this amino acid. However, the exploitation of αMePro and other α-tetrasubstituted proline analogues in the design of peptides with controlled fold in the backbone relies on the previous knowledge of their conformational propensities.

In this work, we have investigated the intrinsic conformational preferences of α-methylproline (αMePro) and α-phenylproline (αPhPro) using Density Functional Theory (DFT) methods. Calculations were performed on the *N*-acetyl-*N*'-methylamide derivatives of the L-amino acids, hereafter denoted as Ac-L-αMePro-NHMe and Ac-L-αPhPro-NHMe(Figure 3.2.1), respectively. The influence of the methyl and phenyl groups has been determined by comparison with the proline derivative Ac-L-Pro-NHMe, which has been investigated for comparative purposes using the same quantum mechanical method. Specifically, we have examined how the substituent incorporated at the α position affects the preferred backbone conformation, the puckering of the pyrrolidine ring and the *cis*/*trans* disposition of the amide bonds. On the other hand, as was mentioned above, the role of the environment, in particular of the solvent, in the *cis*/*trans* rotational isomerism of proline was reported to be crucial.[9] In spite of this, no information about the solvent effects on the isomerization of the α-substituted proline analogs has been provided yet. Accordingly, we decided to evaluate the influence of the solvent polarity on the conformational preferences of the compounds under study using a Self Consistent Reaction Field method.



Ac-L-α-MePro-NHMe Ac-L-α-PhPro-NHMe

Figure 3.2.1: *α-substituted Proline Analogues studied in this work*

3.2.2 Methods

3.2.2.1 Computational Details.

DFT calculations were carried out using the Gaussian 03 computer program,[14] combining the Becke´s three-parameter hybrid functional (B3)[15] with the Lee, Yang and Parr (LYP)[16] expression for the nonlocal correlation (B3LYP). This method provides a very satisfactory description of the conformational properties of cyclic constrained amino acids, including Pro and pseudoprolines.[17,18] Accordingly, all the calculations presented in this work were performed using the B3LYP method combined with the 6-31+G(d,p) basis set,[19] even although some additional single point calculations on selected conformations were performed using the aug-cc-pVTZ[20] basis set.

The backbone ($\omega_0,\varphi,\psi,\omega$) and side chain ($\chi^i$; endocyclic) dihedral angles of the *N*-acetyl-*N*'-methylamide derivatives of Pro, αMePro and αPhPro are defined in Figure 3.2.2. Since φ is fixed by the geometry of the five-membered ring, only three minima may be anticipated for the potential energy surfaces $E=E(\psi)$ of the dipeptides for a given arrangement of the peptide bonds. The flexible angle ψ is expected to have three minima, *i.e. gauche*$^+$ (60º), *trans* (180º) and *gauche*$^-$ (–60º), while each amide bond (ω_0, ω) can be arranged in *cis* or *trans*. It should be noted that only the peptide bond preceding proline (that involving the pyrrolidine nitrogen, corresponding to the ω_0 torsion angle) is likely to adopt a *cis* configuration. However, we considered also the *cis* and *trans* states of the amide bond formed by the proline carbonyl (the methylcarboxamide group, –CONHMe, given by ω) with the aim of exploring how α-methylation affects the isomerism of this amide linkage. For the αPhPro derivative, only the *cis*/*trans* arrangement of ω_0 was considered.

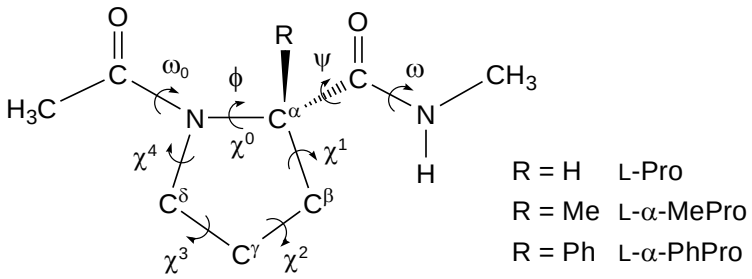


Figure 3.2.2: *Dihedral angles used to identify the conformations of the N-acetyl-N'-methylamide derivatives of proline and its α-substituted analogues studied in this work. The dihedral angles ω_0, φ, ψ and ω are defined using backbone atoms while the endocyclic dihedral angles χ^j are given by the atoms of the five-membered ring. In particular, the sequence of atoms used to define φ and χ^0 are C(=O)–N–C^α–C(=O) and C^δ–N–C^α–C^β, respectively. compounds studied in this work.*

The cyclic side chains of the compounds under study may adopt two main different conformational states, corresponding to the *down* and *up* puckering of the five-membered ring. They are defined as those in which the C^γ atom and the carbonyl group of the Pro residue (or analogue) lie on the same and opposite sides, respectively, of the plane defined by the C^δ, N and C^α atoms.

Accordingly, for Ac-L-Pro-NHMe and Ac-L-αMePro-NHMe, 3(ψ backbone) × 2(ω_0 *cis*-or-*trans*) × 2(ω *cis*-or-*trans*) × 2(cyclic side chain) = 24 structures were considered as starting points for complete geometry optimizations at the B3LYP/6-31+G(d,p) level. Regarding Ac-L-αPhPro-NHMe, ω was kept in the *trans* configuration, while for the arrangement of the phenyl substituent three different orientations were considered. Therefore, the number of starting structures for geometry optimizations were 3(ψ backbone) × 2(ω_0 *cis*-or-*trans*) × 2(cyclic side chain) × 3(Ph substituent) = 36. Frequency analyses were carried out to verify the nature of the minimum state of all the stationary points obtained and to calculate the zero-point vibrational energies (ZPVE) with both thermal and entropic corrections, the latter statistical terms being used to compute the conformational Gibbs free energies in the gas phase (ΔG^{gp}) at the B3LYP/6-31+G(d,p) level.

To obtain an estimation of the solvation effects on the relative stability of the different minima, single point calculations were also conducted on the B3LYP/6-31+G(d,p) optimized structures using a Self-Consistent Reaction Field (SCRF) model. SCRF methods treat the solute at the quantum mechanical level, while the solvent is represented as a dielectric continuum. Specifically, the Polarizable Continuum Model (PCM) developed by Tomasi and co-workers was used to describe the bulk solvent.[21] This method involves the generation of a solvent cavity from spheres centered at each atom in the molecule and the calculation of virtual point charges on the cavity surface representing the polarization of the solvent. The magnitude of these charges is proportional to the derivative of the solute electrostatic potential at each point calculated from the molecular wave function. The point charges may, then, be included in the one-electron Hamiltonian, thus inducing polarization of the solute. An iterative calculation is carried out until the wave function and the surface charges are self-consistent. PCM calculations were performed using the standard protocol and considering the dielectric constants of carbon tetrachloride (ε = 2.228), chloroform (ε = 4.9), methanol (ε= 32.6) and water (ε= 78.4). The conformational free energies in solution ($\Delta G^{\#sol\#}$, where #sol# refers to the solvent) were computed using the classical thermodynamics scheme, that is, the free energies of solvation provided by the PCM model were added to the ΔG^{gp} values.

3.2.2.2 Nomenclature and Pseudorotational Parameters.

The minimum energy conformations of the three dipeptides studied in this work have been denoted using a four-label code that specifies the arrangement of the two peptide bonds, the (φ,ψ) backbone conformation and the puckering of the five-membered ring. The first letter refers to the *trans* (t) or *cis* (c) arrangement of the peptide bond preceding proline (ω_0). The second label identifies the backbone conformation using the nomenclature introduced by Perczel *et al.*[22] more than fifteen years ago. Accordingly, nine different backbone conformations can be distinguished in the potential energy surface $E=E(\varphi,\psi)$ of amino acids: γ_D, δ_D, α_D, ε_D, β_L, ε_L, α_L, δ_L and γ_L. In the case of proline, only the γ_L (γ-turn or C_7), α_L (α-helical), and ε_L (polyproline II-like) conformations are accessible due to φ being fixed in the neighborhood of –60º. Next, the *up* or *down* puckering of the five-membered ring is indicated using the [u] and [d] labels, respectively. In particular, the *down* ring puckering was identified when χ^1 and χ^3 were positive while χ^2 and χ^4 were negative. Conversely, the *up* ring puckering is characterized by negative values of χ^1 and χ^3 and positive values of χ^2 and χ^4. Finally, the last letter indicates the *trans* (t) or *cis* (c) arrangement of the amide bond involving the proline carbonyl group (ω).

The puckering of the five-membered ring was described using the classical pseudorotational algorithm, which uses a very simple model based on only two parameters, as previously applied to proline by Perczel *et al.*[23] The pseudorotational parameters A and P, which describe the puckering amplitude and the state of the pucker in the pseudorotation pathway, respectively, are derived from the endocyclic dihedral angles as follows:

$$\mathrm{A} = \sqrt{(\mathrm{A}\sin\mathrm{P})^2 + (\chi^0)^2}\ , \text{ where } \mathrm{A}\sin\mathrm{P} = \frac{\chi^1 - \chi^2 + \chi^3 - \chi^4}{-2(\sin 144^\circ + \sin 72^\circ)} \qquad (3.2.1)$$

and

$$\mathrm{P} = \begin{cases} \arccos\dfrac{\chi^0}{\mathrm{A}}, & \text{if} \quad \mathrm{A}\sin\mathrm{P} \geq 0 \\ -\arccos\dfrac{\chi^0}{\mathrm{A}}, & \text{if} \quad \mathrm{A}\sin\mathrm{P} < 0 \end{cases} \qquad (3.2.2)$$

Accordingly, parameter A is defined to be positive while P falls between –180º and 180º.

3.2.3 Results and Discussion

3.2.3.1 Ac-L-Pro-NHMe.

Table 3.2.1 lists the most relevant structural parameters together with the relative energy (ΔE^{gp}) and free energy (ΔG^{gp}) in the gas phase for the 14 minimum energy conformations characterized for Ac-L-Pro-NHMe (Figure 3.2.3). These minima are distributed according to the disposition of the peptide bonds (defined by the ω_0 and ω angles, Figure 3.2.2) as follows: both amide moieties adopt a *trans* arrangement in 3 minima (*trans-trans* conformers), one peptide bond is *cis* in 7 minima (4 *cis-trans* and 3 *trans-cis* conformers) and, finally, both peptide bonds exhibit a *cis* configuration in 4 minima (*cis-cis* conformers). It is worth noting that the structural data and ΔE^{gp} values displayed in Table 3.2.1 for the 14 minima characterized for Ac-L-Pro-NHMe are in excellent agreement with the results recently reported by Csizmadia[24] and Kang[8b] at the B3LYP/6-31G(d) and B3LYP/6-311++G(d,p) levels, respectively.

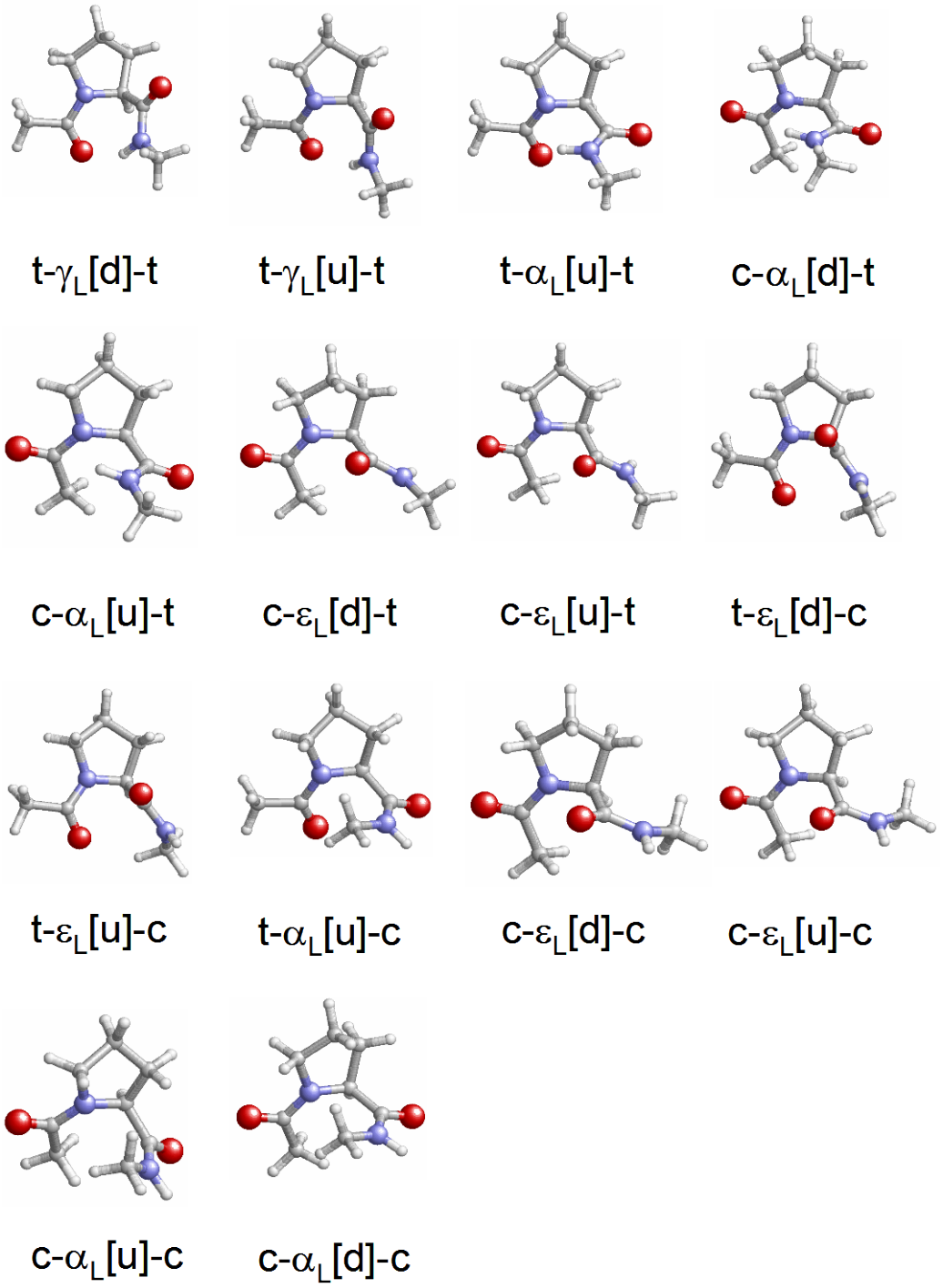


Figure 3.2.3: *Representation of the minimum energy conformations characterized for Ac-L-Pro-NHMe at the B3LYP/6-31+G(d,p) level.Table of contents graphics.*

Table 3.2.1:*Backbone dihedral angles (in degrees), pseudorotational parameters (A and P; in degrees) and relative energy (ΔE^{gp}; in kcal/mol) and free energy (ΔG^{gp}; in kcal/mol) of the minimum energy conformations characterized for Ac-L-Pro-NHMe at the B3LYP/6-31+G(d,p) level in the gas phase.*

# Conf.	ω_0	φ	ψ	ω	(A, P)	ΔE^{gp}	ΔG^{gp}
t-γ_L[d]-t	-172.6	-83.4	70.3	-177.7	(37.4, -111.9)[a]	0.0[b]	0.0[c]
t-γ_L[u]-t	-173.9	-81.6	77.3	-175.9	(37.5, 75.8)[d]	1.0	1.3
t-α_L[u]-t	-171.0	-77.5	-11.5	175.9	(37.8, 89.2)[e]	4.9	4.0
c-α_L[d]-t	10.2	-90.9	-5.1	-179.6	(37.5, -111.5)[f]	3.3	2.3
c-α_L[u]-t	8.0	-79.2	-18.4	-177.1	(37.5, 90.1)[g]	4.2	3.6
c-ε_L[d]-t	1.2	-75.2	147.9	174.6	(36.8, -114.4)[h]	6.3	4.8
c-ε_L[u]-t	-0.1	-61.2	145.3	177.5	(37.4, 88.0)[i]	6.6	5.0
t-ε_L[d]-c	-179.9	-76.7	125.9	-14.6	(35.2, -114.4)[j]	6.1	6.2
t-ε_L[u]-c	177.9	-63.3	128.1	-18.6	(37.6, -48.1)[k]	6.5	6.7
t-α_L[u]-c	-173.4	-61.6	-35.4	12.0	(37.8, 77.9)[l]	11.3	11.3
c-ε_L[d]-c	0.6	-75.6	159.3	-6.0	(37.2, -116.4)[m]	9.3	8.5
c-ε_L[u]-c	-1.6	-61.9	155.1	-5.8	(37.4, 91.5)[n]	9.8	9.2
c-α_L[u]-c	4.3	-63.5	-38.7	0.3	(37.2, 81.4)[o]	10.1	9.7
c-α_L[d]-c	6.9	-82.5	-16.2	-0.8	(36.1, -111.2)[p]	10.8	10.7

[a] χ^0= -13.9º, χ^1= 31.4º, χ^2= -37.6º, χ^3= 28.7º and χ^4= -9.3º. [b] E= -573.315217 a.u. [c] G= -573.132049 a.u. [d] χ^0= -10.3º, χ^1= -13.4º, χ^2= 31.0º, χ^3= -36.6º and χ^4= 29.8º. [e] χ^0= 0.5º, χ^1= -22.9º, χ^2= 36.1º, χ^3= -35.4º and χ^4= 22.0º. [f] χ^0= -13.8º, χ^1= 31.4º, χ^2= -37.6º, χ^3= 29.0º and χ^4= -9.5º. [g] χ^0= -0.1º, χ^1= -22.1º, χ^2= 35.7º, χ^3= -35.2º and χ^4= 22.3º. [h] χ^0= -15.2º, χ^1= 31.6º, χ^2= -36.7º, χ^3= 27.4º and χ^4= -7.5º. [i] χ^0= 1.3º, χ^1= -23.3º, χ^2= 36.0º, χ^3= -34.6º and χ^4= 21.2º. [j] χ^0= -14.5º, χ^1= 30.3º, χ^2= -35.1º, χ^3= 26.0º and χ^4= -7.2º. [k] χ^0= 25.1º, χ^1= -3.6º, χ^2= -19.5º, χ^3= 34.4º and χ^4= -35.8º. [l] χ^0= 7.9º, χ^1= -28.1º, χ^2= 37.7º, χ^3= -32.5º and χ^4= 15.5º. [m] χ^0= -16.5º, χ^1= 32.5º, χ^2= -36.9º, χ^3= 26.8º and χ^4= -6.3º. [n] χ^0= -1.0º, χ^1= -21.5º, χ^2= 35.3º, χ^3= -35.3º and χ^4= 23.1º. [o] χ^0= 5.5º, χ^1= -26.1º, χ^2= 36.8º, χ^3= -33.1º and χ^4= 17.3º. [p] χ^0= -13.1º, χ^1= 30.1º, χ^2= -36.2º, χ^3= 28.0º and χ^4= -9.3º.

As expected, the lowest energy minimum, denoted as t-γ_L[d]-t, was found to be *trans-trans* with the backbone defining a γ_L conformation, *i.e.* a seven-membered intramolecularly hydrogen-bonded ring [d(H···O) = 1.971 Å, ∠N–H···O = 146.4º] and the pyrrolidine moiety exhibiting a *down* puckering. The next minimum, t-γ_L[u]-t, only differs in the ring puckering and this change produces a

destabilization of 1.0 and 1.3 kcal/mol in terms of ΔE^{gp} and ΔG^{gp}, respectively. The last *trans-trans* conformer combines the α_L backbone conformation with an *up*-puckered pyrrolidine and is disfavored by more than 4 kcal/mol.

The *cis-trans* conformer of lowest energy corresponds to c-α_L[d]-t, which is destabilized with respect to the global minimum by 2.3 kcal/mol in terms of ΔG^{gp}. A *down*-to-*up* transition of the pyrrolidine puckering leads to a further destabilization of 1.3 kcal/mol so that the ΔG^{gp} of the c-α_L[u]-t minimum is 3.6 kcal/mol. The remaining *cis-trans* minima as well as all the *trans-cis* and *cis-cis* conformers are highly destabilized with respect to the global minimum, with their ΔG^{gp} values ranging from 4.8 to 11.3 kcal/mol. Surprisingly enough, the least stable minimum, t-α_L[u]-c, corresponds to a *trans-cis* rather than to a *cis-cis* conformer. In order to check that the 6-31+G(d,p) basis set describes satisfactorily the stability of the different conformers, single point calculations were performed at the B3LYP/aug-cc-pVTZ level on all the *trans-trans* and *cis-trans* conformers of Ac-L-Pro-NHMe. As expected, differences between the relative energies provided by these two basis sets are very small (see Supporting Information) evidencing the suitability of the 6-31+G(d,p) one.

A detailed inspection of Table 3.2.1 allows the establishment of a clear relationship between the characterization of certain backbone conformations as energy minima and the *cis*/*trans* state of the amide bonds. Specifically, the γ_L structure appears as an energy minimum only when the two peptide bonds adopt a *trans* disposition (necessary for the formation of the intramolecular hydrogen bond), while minima in the ε_L region are located provided that at least one of the amide linkages is *cis*. Moreover, the latter becomes the preferred backbone conformation when the *cis* peptide bond is that involving the proline carbonyl group (–CONHMe, $\omega \approx 0^o$). In contrast, if the acetamido group adopts a *cis* arrangement ($\omega_0 \approx 0^o$) while ω remains close to 180º, the ε_L backbone conformation is an energy minimum but becomes less stable than the α_L arrangement. It is noteworthy that, even if the ε_L conformation was not detected as an energy minimum for all *trans* peptide bonds, proline is experimentally found to accommodate this disposition with high frequency.[6] Indeed, (φ,ψ) values in the ε_L region correspond to the *i*+1 position of a βII-turn,[6d] which is known to be among those preferred by proline.[6] Interestingly, calculations on Ac-L-Pro-NHMe

at levels of theory lower than that used in this study, locate the *trans-trans* ε_L conformation as an energy minimum,[24] but it disappears when going to larger basis sets, as seen in the present and previous[8b,24,,25] works.

Table 3.2.2 lists the relative free energies in carbon tetrachloride, chloroform, methanol and water solutions for the 14 minima mentioned above. PCM calculations were performed using the geometries optimized in the gas phase. It should be noted that previous studies indicated that solute geometry relaxations in solution and single point calculations on the optimized geometries in the gas phase give almost identical free energies of solvation.[26]

The solvent introduces significant changes in the relative stability of the different minima (Table 3.2.2). Carbon tetrachloride was found to considerably stabilize conformers with at least one *cis* amide bond. Thus, in this solvent, the c-α_L[d]-t conformer becomes almost isoenergetic with the global minimum, t-γ_L[d]-t. Furthermore, the c-ε_L[d]-t and c-ε_L[u]-t minima are 3.2 and 2.5 kcal/mol, respectively, more stable than in the gas phase. In spite of the stabilization produced by this solvent in conformers with *cis* amide bonds, the *trans-cis* and *cis-cis* minima are not within the set of conformations energetically accessible at room temperature.

Table 3.2.2: *Relative free energy in the gas-phase (ΔG^{gp}; in kcal/mol) and in carbon tetrachloride, chloroform, methanol and aqueous solutions (ΔG^{CCl4}, ΔG^{CHCl3}, ΔG^{CH3OH} and ΔG^{H2O}, respectively; in kcal/mol) for the minimum energy conformations of Ac-L-Pro-NHMe at the B3LYP/6-31+G(d,p) level.*

# Conf.	ΔG^{gp}	ΔG^{CCl4}	ΔG^{CHCl3}	ΔG^{CH3OH}	ΔG^{H2O}
t-γ_L[d]-t	0.0	0.0	0.3	4.5	1.3
t-γ_L[u]-t	1.3	1.4	1.7	5.9	2.8
t-α_L[u]-t	4.0	3.0	1.9	4.0	0.3
c-α_L[d]-t	2.3	0.1	1.0	3.8	0.2
c-α_L[u]-t	3.6	2.7	2.1	4.9	0.8
c-ε_L[d]-t	4.8	1.6	0.2	0.3	0.7
c-ε_L[u]-t	5.0	2.5	0.0	0.0	0.0
t-ε_L[d]-c	6.2	4.5	3.8	6.8	3.2
t-ε_L[u]-c	6.7	4.9	3.9	6.6	3.1
t-α_L[u]-c	11.3	7.3	4.8	5.7	1.7
c-ε_L[d]-c	8.5	5.3	3.2	3.3	0.4
c-ε_L[u]-c	9.2	6.5	3.7	4.8	0.9
c-α_L[u]-c	9.7	6.4	4.9	7.0	3.3
c-α_L[d]-c	10.7	7.4	5.9	8.2	4.6

The higher polarity of chloroform results in a further stabilization of conformers with *cis* peptide bonds. In fact, the lowest energy minimum in this solvent is c-ε_L[u]-t, with the c-ε_L[d]-t and t-γ_L[d]-t conformations being disfavored by only 0.2 and 0.3 kcal/mol, respectively. Moreover, the ΔG^{CHCl3} value of the least stable *cis-trans* conformer is 2.1 kcal/mol, which provides evidence for the strong stabilizing effect of this solvent on *cis* peptide bonds. Although *trans-cis* and *cis-cis* conformers are not energetically accessible in chloroform solution, their ΔG^{CHCl3} are about half the values in the gas phase.

The c-ε_L[u]-t is again the most stable conformation in both methanol and aqueous solutions, even although in the latter environment the c-α_L[d]-t and t-α_L[u]-t are unfavored by only 0.2 and 0.3 kcal/mol, respectively. However, the most remarkable result in polar environments is the stabilization of the *cis-trans*

conformers, which is consistent with the theoretical estimation previously reported for the Pro dipeptide.[8] Thus, in such studies it was found that polar environments favor the *cis* conformation in Pro peptidic bonds.

The overall of the results obtained in solution suggests that the conformational flexibility induced by the environment is considerable in terms of both the ψ dihedral angle and *cis*/*trans* isomerism. However, the latter effect seems to be overestimated by the PCM solvation model, this effect being very large when polar solvents like methanol and water are considered. According to the results in Table 3.2.2, conformers exhibiting a *cis* configuration for the peptide bond involving the Pro nitrogen ($\omega_0 \approx 0º$) should predominate over *trans* conformers, which contradicts experimental data.[6-9] Thus, although in condensed phases, *i.e.* solid state or solution, the peptide bond preceding Pro has a relatively high probability of adopting a *cis* arrangement, the *trans* is by far preferred. Accordingly, solvent-induced stabilizations seem to be considerably overestimated by the PCM method. Caution is therefore required when interpreting the results provided by PCM calculations, which should be considered only qualitatively.

3.2.3.2 Ac-L-αMePro-NHMe.

The 17 minimum energy conformations characterized for Ac-L-αMePro-NHMe in the gas phase are displayed in Figure 3.2.4. Their structural and energy data are given in Table 3.2.3. According to the *cis*/*trans* state of the peptide bonds, these minima are distributed as 4 *trans-trans*, 4 *cis-trans*, 5 *trans-cis* and 4 *cis-cis*. Interestingly, α-methylation of Ac-L-Pro-NHMe produces an enlargement of the number of minimum energy conformations, which suggests a higher flexibility for the non-proteinogenic residue. However, a detailed inspection of the energy data in Table 3.2.3 reveals that the replacement of the α hydrogen in proline by a methyl group results in a general destabilization of the minimum energy conformations, particularly, of the *trans-cis* and *cis-cis* subgroups.

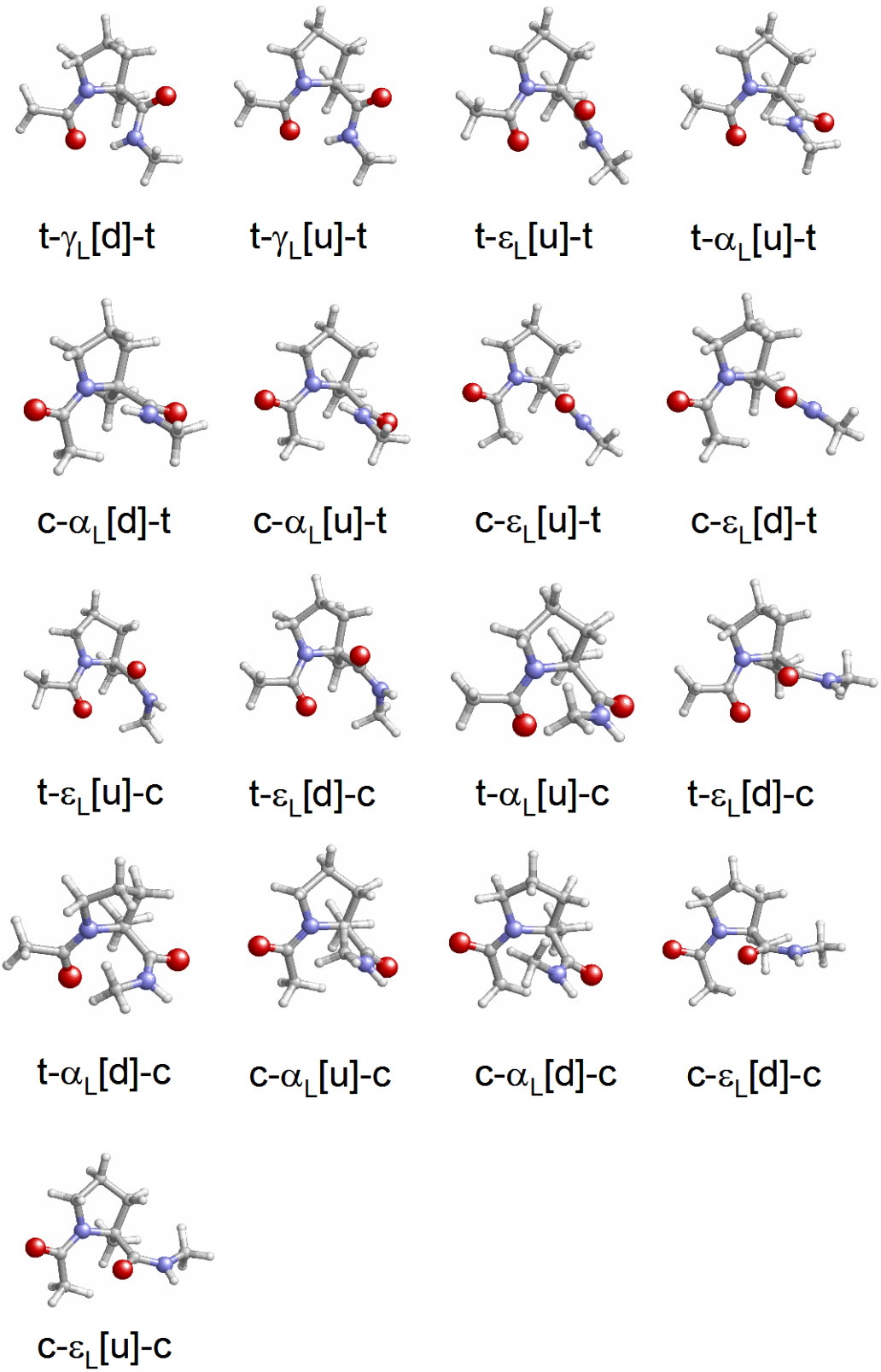


Figure 3.2.4: *Representation of the minimum energy conformations characterized for Ac-L-αMePro-NHMe at the B3LYP/6-31+G(d,p) level.*

Table 3.2.3: *Backbone dihedral angles (in degrees), pseudorotational parameters (A and P; in degrees) and relative energy (ΔE^{gp}; in kcal/mol) and free energy (ΔG^{gp}; in kcal/mol) of the minimum energy conformations characterized for Ac-L-αMePro-NHMe at the B3LYP/6-31+G(d,p) level in the gas phase.*

# Conf.	ω_0	φ	ψ	ω	(A, P)	ΔE^{gp}	ΔG^{gp}
t-γ_L[d]-t	-172.2	-77.2	57.1	178.7	(35.0, -108.7)[a]	0.0[b]	0.0[c]
t-γ_L[u]-t	-176.2	-69.8	60.7	178.8	(31.2, 86.9)[d]	1.6	1.7
t-ε_L[u]-t	175.4	-55.1	123.4	-174.0	(31.3, 91.7)[e]	3.3	2.8
t-α_L[u]-t	-172.6	-64.2	-20.1	177.6	(33.6, 72.8)[f]	4.2	3.2
c-α_L[d]-t	10.8	-77.7	-17.1	-179.1	(37.3, -94.3)[g]	4.5	3.3
c-α_L[u]-t	3.0	-63.5	-28.1	-177.9	(37.6, 74.9)[h]	4.6	3.1
c-ε_L[u]-t	-3.0	-48.5	141.5	176.4	(37.8, 74.9)[i]	7.6	6.9
c-ε_L[d]-t	0.0	-66.3	149.5	178.3	(36.8, -100.0)[j]	7.7	6.6
t-ε_L[u]-c	175.3	-55.5	125.9	-32.4	(38.3, 88.5)[k]	10.3	11.2
t-ε_L[d]-c	177.8	-61.1	127.7	-32.4	(36.1, -94.3)[l]	10.6	11.5
t-α_L[u]-c	-174.4	-54.6	-41.1	15.8	(36.8, 68.3)[m]	10.6	11.1
t-ε_L[d]-c	173.9	-67.2	164.4	-2.0	(36.1, -116.5)[n]	11.6	11.7
t-α_L[d]-c	-170.2	-73.7	-18.9	18.6	(37.7, -95.5)[o]	11.8	11.7
c-α_L[u]-c	-0.2	-56.5	-40.0	3.5	(36.4, 71.2)[p]	10.6	10.8
c-α_L[d]-c	12.9	-64.9	-38.6	-8.4	(32.2, -68.8)[q]	12.8	12.7
c-ε_L[d]-c	-2.5	-70.0	177.0	-4.1	(37.7, -111.0)[r]	13.9	14.1
c-ε_L[u]-c	-11.7	-57.5	177.3	3.7	(37.9, 94.5)[s]	14.6	14.4

[a] χ^0= -11.2º, χ^1= 30.9º, χ^2= -39.4º, χ^3= 31.6º and χ^4= -12.8º. [b] E= -612.629968 a.u. [c] G= -612.419998 a.u. [d] χ^0= -0.9º, χ^1= -22.5º, χ^2= 37.1º, χ^3= -36.8º and χ^4= 23.8º. [e] χ^0= 1.7º, χ^1= -24.0º, χ^2= 37.0º, χ^3= -35.1º and χ^4= 21.2º. [f] χ^0= 9.9º, χ^1= -29.4º, χ^2= 38.1º, χ^3= -31.4º and χ^4= 13.4º. [g] χ^0= -2.8º, χ^1= 24.4º, χ^2= -36.5º, χ^3= 33.9º and χ^4= -19.7º. [h] χ^0= 9.8º, χ^1= -29.1º, χ^2= 37.9º, χ^3= -31.2º and χ^4= 13.4º. [i] χ^0= 9.8º, χ^1= -29.3º, χ^2= 38.0º, χ^3= -31.4º and χ^4= 13.6º. [j] χ^0= -6.4º, χ^1= 26.6º, χ^2= -36.7º, χ^3= 32.1º and χ^4= -16.2º. [k] χ^0= 1.0º, χ^1= -23.5º, χ^2= 36.9º, χ^3= -35.4º and χ^4= 21.8º. [l] χ^0= -2.3º, χ^1= 22.5º, χ^2= -33.7º, χ^3= 31.3º and χ^4= -18.2º. [m] χ^0= 13.6º, χ^1= -30.9º, χ^2= 317.1º, χ^3= -28.3º and χ^4= 9.1º. [n] χ^0= -16.1º, χ^1= 31.7º, χ^2= -36.0º, χ^3= 25.7º and χ^4= -5.9º. [o] χ^0= -3.6º, χ^1= 25.2º, χ^2= -37.1º, χ^3= 33.9º and χ^4= -19.2º. [p] χ^0= 11.3º, χ^1= -29.3º, χ^2= 36.7º, χ^3= -29.3º and χ^4= -11.2º. [q] χ^0= 11.6º, χ^1= 8.8º, χ^2= -24.9º, χ^3= 31.3º and χ^4= -27.5º. [r] χ^0= -13.5º, χ^1= 31.5º, χ^2= -38.0º, χ^3= 29.2º and χ^4= -9.7º. [s] χ^0= -3.0º, χ^1= -20.1º, χ^2= 35.1º, χ^3= -36.1º and χ^4= 25.0º.

The lowest energy conformation characterized for Ac-L-αMePro-NHMe in the gas phase corresponds to a t-γ_L[d]-t conformer, which was also identified as the global minimum for Ac-L-Pro-NHMe. The geometric parameters of the hydrogen bond associated with this conformation [d(H···O) = 1.874 Å, ∠N–H···O = 151.4º] indicate that this intramolecular interaction is stronger in the α-methyl derivative. The other two *trans-trans* conformers found for Ac-L-Pro-NHMe, t-γ_L[u]-t and t-α_L[u]-t (Table 3.2.1), were also located as energy minima for Ac-L-αMePro-NHMe (Table 3.2.3), with similar geometries and energies. Thus, the main difference between Pro and αMePro when both peptide bonds exhibit a *trans* arrangement is the characterization of a minimum in the ε_L region for the α-methylated compound. No such semi-extended backbone conformation was detected as an energy minimum for Ac-L-Pro-NHMe. This could be indicative of this backbone conformation being more favorable for αMePro than for the parent amino acid, which is contrary to the general observation that semi-extended and fully extended conformations are more stable for proteinogenic amino acids than for their α-methylated counterparts.[1c,e,3-5]

This singularity is specifically evidenced in Figure 3.2.5, where the potential energy curves $E = E(\psi)$ of Ac-L-Pro-NHMe and Ac-L-αMePro-NHMe for *trans* peptide bonds and an *up*-puckered ring are compared. As can be seen, the two profiles differ almost uniquely in the flat region that appears for the latter compound at ψ values ranging from 120º to 150º, that is, where the ε_L semi-extended conformation is located. However, as already mentioned, conformations in the ε_L region are very often observed experimentally[6] for Pro-containing peptides longer than that considered in the present work.

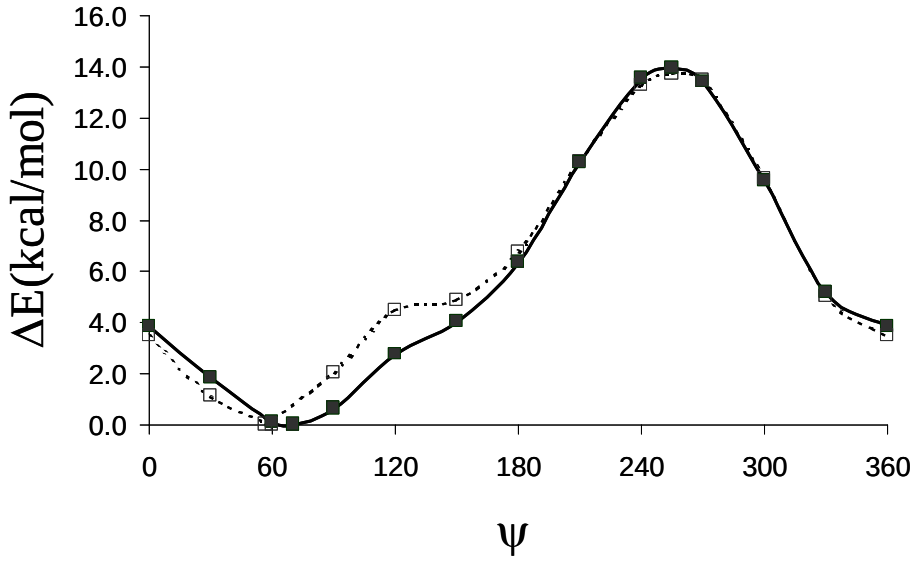


Figure 3.2.5: *Potential energy curves E=E(ψ) cross sections of the conformational potential energy surfaces of Ac-L-Pro-NHMe (filled squares and solid lines) and Ac-L-αMePro-NHMe (empty squares and dashed lines). In both compounds, the pyrrolidine ring is up-puckered and the peptide bonds are arranged in trans.*

Also *cis-trans* α_L and ε_L conformers similar to those observed for proline were characterized for αMePro. They are disfavored with respect to the global minimum by about 3 and 7 kcal/mol, respectively, the influence of the pyrrolidine ring puckering being negligible (Table 3.2.3). Comparison between the ΔG^{gp} values obtained for the *cis-trans* conformers of Ac-L-Pro-NHMe and Ac-L-αMePro-NHMe indicates that, in general, α-methylation produces a destabilization of 1–2 kcal/mol. This result is not unexpected since the α-methyl group increases the steric hindrance around C^{α}, thus disfavoring the *cis* disposition between the acetyl methyl group and the α carbon ($\omega_0 \approx 0^o$). The effect of α-methylation in the destabilization of *cis* amide bonds becomes more evident for the –CONHMe moiety (corresponding to ω). In fact, all *trans-cis* and *cis-cis* conformers exhibit ΔG^{gp} values above 10.8 kcal/mol (Table 3.2.3) and significantly higher than those obtained for the equivalent conformers of Ac-L-Pro-NHMe (Table 3.2.1).

Table 3.2.4 shows the effects of solvation on the 17 minima of Ac-L-αMePro-NHMe. As can be seen, the t-γ_L[d]-t is the most stable conformation not only in

the gas phase but also in carbon tetrachloride and chloroform solutions. Significant differences are observed between the results obtained for Ac-L-Pro-NHMe (Table 3.2.2) and Ac-L-αMePro-NHMe (Table 3.2.4) in chloroform. Specifically, for the latter peptide, the ΔG^{CHCl3} values of all four *trans-trans* conformers lie below 1.6 kcal/mol, whereas all the *cis-trans* conformers show ΔG^{CHCl3} values above this limit, indicating that only *trans-trans* arrangements are energetically accessible in chloroform. This is in sharp contrast with the results obtained for Ac-L-Pro-NHMe, for which certain *cis-trans* conformers were found to exhibit a high stability.

Table 3.2.4: *Relative free energy in the gas-phase (ΔG^{gp}; in kcal/mol) and in carbon tetrachloride, chloroform. methanol and aqueous solutions (ΔG^{CCl4}, ΔG^{CHCl3}, ΔG^{CH3OH} and ΔG^{H2O}, respectively; in kcal/mol) for the minimum energy conformations of Ac-L-αMePro-NHMe at the B3LYP/6-31+G(d,p) level.*

# Conf.	ΔG^{gp}	ΔG^{CCl4}	ΔG^{CHCl3}	ΔG^{CH3OH}	ΔG^{H2O}
t-γ_L[d]-t	0.0	0.0	0.0	0.9	1.4
t-γ_L[u]-t	1.7	1.5	1.4	2.4	2.8
t-ε_L[u]-t	2.8	2.1	1.6	1.7	2.0
t-α_L[u]-t	3.3	2.4	1.2	0.2	0.3
c-α_L[d]-t	3.3	1.1	1.9	1.1	0.5
c-α_L[u]-t	3.1	0.8	1.7	0.4	0.0
c-ε_L[u]-t	6.9	2.7	4.5	0.0	0.6
c-ε_L[d]-t	6.6	2.3	4.3	0.5	0.6
t-ε_L[u]-c	11.2	8.9	8.0	7.7	7.5
t-ε_L[d]-c	11.5	9.4	8.6	8.5	8.4
t-α_L[u]-c	11.1	7.2	4.7	2.6	2.2
t-ε_L[d]-c	11.7	8.5	6.7	5.1	5.1
t-α_L[d]-c	11.7	8.2	6.2	4.8	4.5
c-α_L[u]-c	10.8	7.1	5.4	4.8	4.5
c-α_L[d]-c	12.7	9.0	7.2	6.7	6.2
c-ε_L[d]-c	14.1	9.8	7.2	5.2	4.8
c-ε_L[u]-c	14.4	10.0	7.4	5.5	5.0

Although the stability of the *cis-trans* conformers in Table 3.2.4 is probably overestimated in carbon tetrachloride, the general tendencies derived from PCM calculations in non-polar environments are fully consistent with data from NMR experiments, which showed no *cis* conformers for Ac-L-αMePro-NHMe in chloroform solution.[10d] In good agreement, our calculations predict the peptide involving the pyrrolidine nitrogen (ω_0) to exhibit a considerably smaller probability of adopting a *cis* disposition in αMePro than in Pro.

Finally, analysis of the results obtained for Ac-L-αMePro-NHMe in methanol and aqueous solution indicates that the *cis-trans* conformers are the most favored in these polar environments. Thus, the c-ε_L[u]-t and c-α_L[u]-t are the lowest energy minimum in methanol and water, respectively, and, in addition, the ΔG^{CH3OH} and ΔG^{H2O} of the remaining three *cis-trans* conformers are lower than 1.5 kcal/mol. These results clearly evidence that the stability of the *cis* configuration for the peptide bond involving the αMePro nitrogen ($\omega_0 \approx 0^o$) is significantly overestimated by the PCM method when polar solvents are considered.

3.2.3.3 Ac-L-αPhPro-NHMe.

Table 3.2.5 shows the structural parameters together with the ΔE^{gp} and ΔG^{gp} values for the 8 minimum energy conformations found for the αPhPro-containing peptide (Figure 3.2.6). Specifically, 4 minima with two *trans* amide bonds were characterized, while the other 4 correspond to *cis-trans* conformers. It should be noted that the *cis* arrangement of the –CONHMe peptide bond (corresponding to the ω angle) was not considered for this compound.

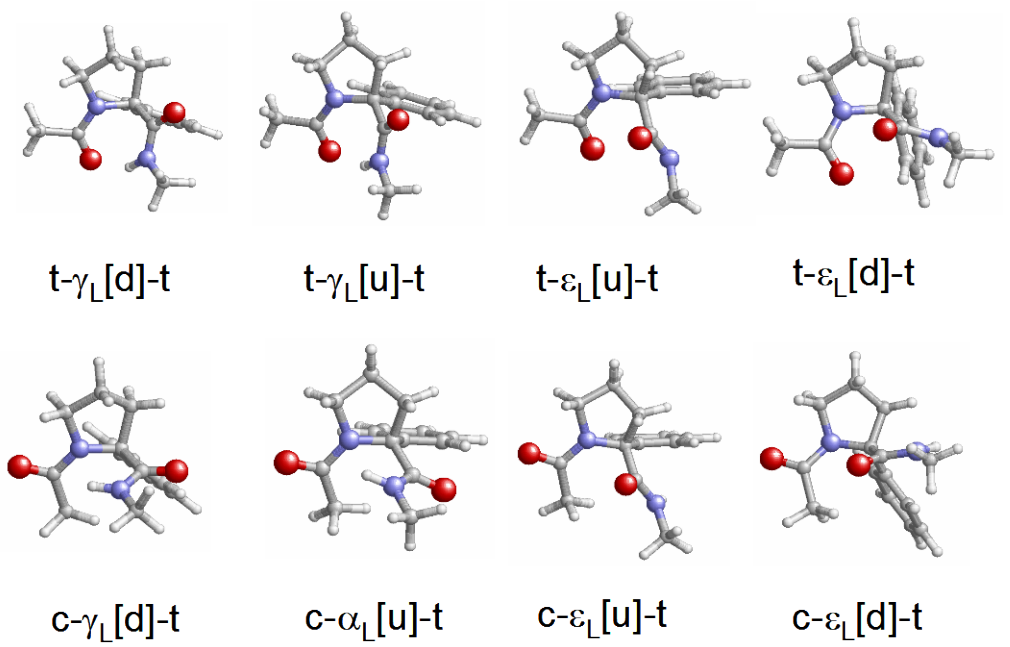


Figure 3.2.6: *Representation of the minimum energy conformations characterized for Ac-L-αPhPro-NHMe at the B3LYP/6-31+G(d,p) level.*

Table 3.2.5: *Backbone dihedral angles (in degrees), pseudorotational parameters (A and P; in degrees) and relative energy (ΔE^{gp}; in kcal/mol) and free energy (ΔG^{gp}; in kcal/mol) of the minimum energy conformations characterized for Ac-L-αPhPro-NHMe at the B3LYP/6-31+G(d,p) level in the gas phase.*

# Conf.	ω_0	φ	ψ	ω	(A, P)	ΔE^{gp}	ΔG^{gp}
t-γ_L[d]-t	-174.9	-75.4	59.1	178.9	(39.1, -107.0)[a]	0.0[b]	0.0[c]
t-γ_L[u]-t	-179.6	-67.9	66.2	-179.9	(39.2, 89.8)[d]	1.7	1.3
t-ε_L[u]-t	172.0	-46.5	120.3	-172.0	(39.9, 76.6)[e]	2.6	2.6
t-ε_L[d]-t	176.7	-61.2	161.7	176.6	(37.8, -98.4)[f]	7.8	7.5
c-γ_L[d]-t	10.8	-85.5	5.5	-176.4	(38.7, -103.8)[g]	4.5	4.0
c-α_L[u]-t	3.2	-59.9	-28.1	-175.9	(38.6, 68.2)[h]	5.1	3.8
c-ε_L[u]-t	-5.0	-43.3	129.6	-179.5	(39.9, 66.0)[i]	5.6	5.5
c-ε_L[d]-t	-4.7	-73.3	-162.6	-177.4	(39.7, -118.2)[j]	8.2	7.6

[a] χ^0= -11.4º, χ^1= 31.2º, χ^2= -39.5º, χ^3= 31.7º and χ^4= -12.7º. [b] E= -804.371070 a.u. [c] G= -804.113301 a.u. [d] χ^0= 0.1º, χ^1= -23.4º, χ^2= 37.5º, χ^3= -36.6º and χ^4= 23.1º. [e] χ^0= 9.3º, χ^1= -30.3º, χ^2= 40.2º, χ^3= -33.7º and χ^4= 15.3º. [f] χ^0= -5.5º, χ^1= -26.6º, χ^2= -37.4º, χ^3= 34.0º and χ^4= -17.6º. [g] χ^0= -9.2º, χ^1= 29.5º, χ^2= -38.9º, χ^3= 32.6º and χ^4= -14.6º. [h] χ^0= 14.3º, χ^1= -32.4º, χ^2= 38.9º, χ^3= -29.7º and χ^4= 9.4º. [i] χ^0= 16.2º, χ^1= -34.3º, χ^2= 40.1º, χ^3= -29.7º and χ^4= 8.2º. [j] χ^0= -18.8º, χ^1= 35.4º, χ^2= -39.5º, χ^3= 27.6º and χ^4= -5.2º.

As observed before for Pro and αMePro, the γ_L backbone conformation with all *trans* peptide bonds is the most stable arrangement for αPhPro, with the *down* puckering of the pyrrolidine ring being preferred. Thus, t-γ_L[d]-t appears as the global minimum while t-γ_L[u]-t is destabilized by 1.3 kcal/mol. In spite of this parallelism, the conformational profile of the α-phenyl derivative shows important differences with respect to those described above for Pro and αMePro. The semi-extended structure t-ε_L[u]-t characterized as an energy minimum for the α-methylated compound, but not for the parent amino acid, was also located for αPhPro, 2.6 kcal/mol above the global minimum (Table 3.2.5). Moreover, an additional ε_L minimum with a *down* puckering was found for the latter compound, although this arrangement of the five-membered ring proved very unfavorable energetically. The overall of these results suggests that conformations in the ε_L region could be more favored for α-substituted proline derivatives than for proline itself, contrary to the general behavior expected for α-tetrasubstituted amino acids in comparison with their proteinogenic counterparts.[1c,e,3–5]

Another distinct feature in the conformational map of Ac-L-αPhPro-NHMe is the disappearance of *trans-trans* minima of the α-helical type. Thus, the t-α_L[u]-t, which was characterized for both Ac-L-Pro-NHMe and Ac-L-αMePro-NHMe, was not a minimum in the potential energy hypersurface of Ac-L-αPhPro-NHMe. Although calculations on small peptide systems like these in the present study are known to underestimate the stability of α-helical conformations (in general, of those lacking an intramolecular hydrogen bond) in favor of γ-turns, this finding is highly remarkable.

The effect of α-substitution on the *cis*/*trans* isomerism described above for Ac-L-αMePro-NHMe is also observed for αPhPro. The *cis-trans* conformers in Table 3.2.5 exhibit ΔG^{gp} values ranging from 3.8 to 7.6 kcal/mol, evidencing a destabilization of the *cis* disposition of the ω_0 amide bond with reference to that observed for Pro.

Table 3.2.6 compares the solvation effects estimated for the 8 minima characterized for Ac-L-αPhPro-NHMe. As can be seen, the conformational properties predicted in carbon tetrachloride and chloroform solutions are very similar to those obtained in the gas phase. The *trans-trans* conformers are scarcely

affected by solvation, while the relative free energy of the *cis-trans* conformers decreases on going from the gas phase to solution, and with the solvent polarity. In spite of such stabilization, minima with a *cis* peptide bond remain inaccessible at room temperature. In fact, only the t-γ_L[d]-t and t-γ_L[u]-t conformers present energies below 2.0 kcal/mol in both solvents and are therefore predicted to be populated. However, results in methanol and aqueous solutions reflect again the limitations of the PCM model to describe the *cis*/*trans* isomerism of ω_0 in polar environments.

Table 3.2.6: *Relative free energy in the gas-phase (ΔG^{gp}; in kcal/mol) and in carbon tetrachloride, chloroform, methanol and aqueous solutions (ΔG^{CCl4}, ΔG^{CHCl3}, ΔG^{CH3OH} and ΔG^{H2O}, respectively; in kcal/mol) for the minimum energy conformations of Ac-L-αPhPro-NHMe at the B3LYP/6-31+G(d,p) level.*

# Conf.	ΔG^{gp}	ΔG^{CCl4}	ΔG^{CHCl3}	ΔG^{CH3OH}	ΔG^{H2O}
t-γ_L[d]-t	0.0	0.0	0.0	0.0	0.3
t-γ_L[u]-t	1.3	1.2	1.3	1.4	1.6
t-ε_L[u]-t	2.6	2.4	2.2	1.6	1.3
t-ε_L[d]-t	7.5	6.9	6.1	4.5	4.4
c-γ_L[d]-t	4.0	2.8	2.5	2.3	2.1
c-α_L[u]-t	3.8	2.4	2.0	1.4	0.9
c-ε_L[u]-t	5.5	3.5	2.5	1.3	2.8
c-ε_L[d]-t	7.6	4.8	2.8	0.5	0.0

3.2.4 Conclusions

Quantum mechanical calculations at the B3LYP/6-31+G(d,p) level have been used to explore the conformational preferences of Ac-L-αMePhe-NHMe and Ac-L-αPhPro-NHMe. Comparison of the results with those obtained for Ac-L-Pro-NHMe at the same theoretical level allows us to draw the following conclusions:

(i) Replacement of the α hydrogen in proline by a more bulky group destabilizes the *cis* configuration of the amide bond involving the pyrrolidine nitrogen. The percentage of *cis* conformers usually

observed for the peptide bond preceding proline, if any, is thus predicted to be much inferior for α-tetrasubstituted proline derivatives.

(ii) Another general structural trend associated with C^{α}-tetrasubstitution seems to be the stabilization of the semi-extended polyproline II conformation (ε_L), which was identified as an energy minimum for both αMePro and αPhPro but not for the proteinogenic amino acid.

(iii) Although α-tetrasubstitution results in general conformational changes like those outlined above, more subtle but equally important differences seem to be associated with the particular nature of the substituent incorporated at C^{α}. Thus, even if the γ-turn (γ_L) is the lowest energy minimum for both Ac-L-αMePhe-NHMe and Ac-L-αPhPro-NHMe in all the environmental conditions examined, the α-helical conformation (α_L) with *trans* amide bonds was also found to be accessible for the α-methyl derivative but was not located as an energy minimum for the αPhPro-containing peptide.

(iv) PCM calculations in solution indicate that the stability of the conformers with a *cis* configuration for the peptide bond involving Pro nitrogen increases with the polarity of the environment. However, in this case results in solution must be analyzed with caution since SCRF calculations overestimate this effect significantly, especially when polar solvents (as water or methanol) are considered.

3.2.5 References

1. (a) Toniolo, C.; Formaggio, F.; Kaptein, B.; Broxterman, Q. B. *Synlett* **2006**, 1295. (b) Venkatraman, J.; Shankaramma, S. C.; Balaram, P. *Chem. Rev.* **2001**, *101*, 3131. (c) Toniolo, C.; Crisma, M.; Formaggio, F.; Peggion, C. *Biopolymers (Pept. Sci.)* **2001**, *60*, 396. (d) Kaul, R.; Balaram, P. *Bioorg. Med. Chem.* **1999**, *7*, 105. (e) Benedetti, E. *Biopolymers (Pept. Sci.)* **1996**, *40*, 3. (f) Toniolo, C.; Benedetti, E. *Macromolecules* **1991**, *24*, 4004.
2. (a) Cativiela, C.; Díaz-de-Villegas, M. D. *Tetrahedron: Asymmetry* **2007**, *18*, 569. (b) Vogt, H.; Bräse, S. *Org. Biomol. Chem.* **2007**, *5*, 406. (c) Park,

K.-H.; Kurth, M. J. *Tetrahedron* **2002**, *58*, 8629. (d) Cativiela, C.; Díaz-de-Villegas, M. D. *Tetrahedron: Asymmetry* **2000**, *11*, 645. (e) Cativiela, C.; Díaz-de-Villegas, M. D. *Tetrahedron: Asymmetry* **1998**, *9*, 3517.

3. (a) Karle, I. L.; Baralam, P. *Biochemistry* **1990**, *29*, 6747. (b) Karle, I. L.; Flippen-Anderson, J. L.; Uma, K.; Balaram, P. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 7921. (c) Karle, I. L.; Flippen-Anderson, J. L.; Uma, K.; Balaram, H.; Balaram, P. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 765. (d) Benedetti, E.; Bavoso, A.; Diblasio, B.; Pavone, V.; Pedone, C.; Crisma, M.; Bonora, G. M.; Toniolo, C. *J. Am. Chem. Soc.* **1982**, *104*, 2437.
4. (a) Improta, R.; Barone, V.; Kudin, K. N.; Scuseria, G. E. *J. Am. Chem. Soc.* **2001**, *123*, 3311. (b) Improta, R.; Rega, N.; Alemán, C.; Barone, V. *Macromolecules* **2001**, *34*, 7550. (c) Alemán, C. *Biopolymers* **1994**, *34*, 841. (d) Huston, S. E.; Marshall, G. R. *Biopolymers* **1994**, *34*, 75. (e) Paterson, Y.; Rumsey, S. M.; Benedetti, E.; Nemethy, G.; Scheraga, H. A. *J. Am. Chem. Soc.* **1981**, *103*, 2947.
5. (a) Toniolo, C.; Crisma, M.; Formaggio, F.; Peggion, C.; Broxterman, Q. B.; Kaptein, B. *J. Inclusion Phenom. Macrocyclic Chem.* **2005**, *51*, 121. (b) Pengo, P.; Pasquato, L.; Moro, S.; Brigo, A.; Fogolari, F.; Broxterman, Q. B.; Kaptein, B.; Scrimin, P. *Angew. Chem. Int. Ed.* **2003**, *42*, 3388. (c) Gratias, R.; Konat, R.; Kessler, H.; Crisma, M.; Valle, G.; Polese, A.; Formaggio, F.; Toniolo, C.; Broxterman, Q. B.; Kamphuis, J. *J. Am. Chem. Soc.* **1998**, *120*, 4763. (d) Benedetti, E.; Saviano, M.; Iacovino, R.; Pedone, C.; Santini, A.; Crisma, M.; Formaggio, F.; Toniolo, C.; Broxterman, Q. B.; Kamphuis, J. *Biopolymers* **1998**, *46*, 433. (e) Yoder, G.; Polese, A.; Silva, R. A. G. D.; Formaggio, F.; Crisma, M.; Broxterman, Q. B.; Kamphuis, J.; Toniolo, C.; Keiderling, T. A. *J. Am. Chem. Soc.* **1997**, *119*, 10278. (f) Formaggio, F.; Crisma, M.; Toniolo, C.; Kamphuis, J. *Biopolymers* **1996**, *38*, 301. (g) Valle, G.; Pantano, M.; Formaggio, F.; Crisma, M.; Toniolo, C.; Précigoux, G.; Sulzenbacher, G.; Boesten, W. H. J.; Broxterman, Q. B.; Schoemaker, H. E.; Kamphuis, J. *Biopolymers* **1993**, *33*, 1617. (h) Toniolo, C.; Crisma, M.; Formaggio, F.; Valle, G.; Cavicchioni, G.; Precigoux, G.; Aubry, A.; Kamphuis, J. *Biopolymers* **1993**, *33*, 1061. (i) Pantano, M.; Formaggio, F.; Crisma, M.; Bonora, G. M.; Mammi, S.; Peggion, E.; Toniolo, C.; Boesten, W. H. J.;

Broxterman, Q. B.; Schoemaker, H. E.; Kamphuis, J. *Macromolecules* **1993**, *26*, 1980.

6. (a) Chakrabarti, P.; Pal, D. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1. (b) Marraud, M.; Aubry, A. *Biopolymers* **1996**, *40*, 45. (c) MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397. (d) Rose, G. D.; Gierasch, L. M.; Smith, J. A. *Adv. Protein Chem.* **1985**, *37*, 1.
7. (a) Boussard, G.; Marraud, M.; Aubry, A. *Biopolymers* **1979**, *18*, 1297. (b) V. Madison, K. D. Kopple, *J. Am. Chem. Soc.* **1980**, *102*, 4855. (c) Liang, G. B.; Rito, C. J.; Gellman, S. H. *Biopolymers* **1992**, *32*, 293. (d) Beausoleil, E.; Lubell, W. D. *J. Am. Chem. Soc.* **1996**, *118*, 12902.
8. (a) Kang, Y. K. *J. Phys. Chem. B* **2006**, *110*, 21338. (b) Kang, Y. K.; Byun, B. J. *J. Phys. Chem. B* **2007**, *111*, 5377. (c) Rivail, J. L.; Bouchy, A.; Loos, P. F.*J. Arg. Chem. Soc.* **2006**, *94*, 19.
9. (a) Dugave, C.; Demange, L. *Chem. Rev.* **2003**, *103*, 2475. (b) Pal, D.; Chakrabarti, P. *J. Mol. Biol.* **1999**, *294*, 271. (c) Stewart, D. E.; Sarkar, A.; Wampler, J. E. *J. Mol. Biol.* **1990**, *214*, 253. (d) Grathwohl, C.; Wüthrich, K. *Biopolymers* **1981**, *20*, 2623.
10. (a) Baures, P. W.; Ojala, W. H.; Gleason, W. B.; Johnson, R. L. *J. Pept. Res.* **1997**, *50*, 1. (b) Flippen-Anderson, J. L.; Gilardi, R.; Karle, I. L.; Frey, M. H.; Opella, S. J.; Gierasch, L. M.; Goodman, M.; Madison, V.; Delaney, N. G. *J. Am. Chem. Soc.* **1983**, *105*, 6609. (c) Madison, V.; Delaney, N. G. *Biopolymers* **1983**, *22*, 869. (d) Delaney, N. G.; Madison, V. *J. Am. Chem. Soc.* **1982**, *104*, 6635.
11. (a) Nanzer, A. P.; Torda, A. E.; Bisang, C.; Weber, C.; Robinson, J. A.; van Gunsteren, W. F. *J. Mol. Biol.* **1997**, *267*, 1012. (b) Bisang, C.; Weber, C.; Inglis, J.; Schiffer, C. A.; van Gunsteren, W. F.; Jelesarov, I.; Bosshard, H. R.; Robinson, J. A. *J. Am. Chem. Soc.* **1995**, *117*, 7904. (c) Welsh, J. H.; Zerbe, O.; von Philipsborn, W.; Robinson, J. A. *FEBS Lett.* **1992**, *297*, 216. (d) Hinds, M. G.; Welsh, J. H.; Brennand, D. M.; Fisher, J.; Glennie, M. J.; Richards, N. G. J.; Turner, D. L.; Robinson, J. A. *J. Med. Chem.* **1991**, *34*, 1777. (e) Richards, N. G. J.; Hinds, M. G.; Brennand, D. M.; Glennie, M. J.; Welsh, J. H.; Robinson, J. A. *Biochem. Pharmacol.* **1990**, *40*, 119.

12. (a) Alonso De Diego, S. A.; Muñoz, P.; González-Muñiz, R.; Herranz, R.; Martín-Martínez, M.; Cenarruzabeitia, E.; Frechilla, D.; Del Río, J.; Jimeno, M. L.; García-López, M. T. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2279. (b) Kobayashi, S.; Chikushi, A.; Tougu, S.; Imura, Y.; Nishida, M.; Yano, Y.; Matsuzaki, K. *Biochemistry* **2004**, *43*, 15610. (c) Sem, D. S.; Baker, B. L.; Victoria, E. J.; Jones, D. S.; Marquis, D.; Yu, L.; Parks, J.; Coutts, S. M. *Biochemistry* **1998**, *37*, 16069. (d) Jones, D. S.; Gamino, C. A.; Randow, M. E.; Victoria, E. J.; Yu, L.; Coutts, S. M. *Tetrahedron Lett.* **1998**, *39*, 6107. (e) Alig, L.; Edenhofer, A.; Hadváry, P.; Hürzeler, M.; Knopp, D.; Müller, M.; Steiner, B.; Trzeciak, A.; Weller, T. *J. Med. Chem.* **1992**, *35*, 4393. (f) Thaisrivongs, S.; Pals, D. T.; Lawson, J. A.; Turner, S. R.; Harris, D. W. *J. Med. Chem.* **1987**, *30*, 536. (g) Pinnock, R. D.; Woodruff, G. N.; Turnbull, M. J. *Neuropharmacology* **1983**, *22*, 687.

13. (a) Haque, T. S.; Ewing, W. R.; Mapelli, C.; Lee, V. G.; Sulsky, R. B.; Riexinger, D. J.; Martinez, R. L.; Zhu, Y.; Ruan, Z. PCT Int. Appl. 2007, WO 2007082264. (b) Gluckman, P. D.; Guan, J.; Woodnorth, M.-A.; Brimble, M. A. US Pat. Appl. 2007, US 2007004641. (c) Ewing, W. R.; Mapelli, C.; Riexinger, D. J.; Lee, V. G.; Sulsky, R. B.; Zhu, Y. PCT Int. Appl. 2006, WO 2006127948. (d) Bo, Y. Y.; Chakrabarti, P. P.; Chen, N.; Liao, H.; Norman, M. H.; Stec, M.; Tamayo, N.; Wang, X. US Pat. Appl. 2006, US 2006183745. (e) Chao, B.; Deckwerth, T. L.; Furth, P. S.; Linton, S. D.; Spada, A. P.; Ullman, B. R.; Weinhouse, M. I. PCT Int. Appl. 2006, WO 2006017295. (f) Brimble, M. A.; Harris, P. W. R.; Sieg, F. PCT Int. Appl. 2006, WO 2006127702. (g) Grotta, J.; Gluckman, P. D. PCT Int. Appl. 2005, WO 2005097161. (h) Hennequin, L. F. A.; Halsall, C. T. PCT Int. Appl. 2005, WO 2005026156. (i) Natarajan, S. I.; Bastos, M. M.; Bernatowicz, M. S.; Mapelli, C.; Lee, V.; Ewing, W. R. PCT Int. Appl. 2003, WO 2003033671. (j) Chu, S. S.; Alegria, L. A.; Bleckman, T. M.; Chong, W. K. M.; Duvadie, R. K.; Li, L.; Reich, S. H.; Romines, W. H.; Wallace, M. B.; Yang, Y. PCT Int. Appl. 2003, WO 2003004467. (k) Durette, P. L.; Hagmann, W. K.; Maccoss, M.; Mills, S. G.; Mumford, R. A.; Van Riper, G. M.; Schmidt, J. A.; Kevin, N. J. PCT Int. Appl. 1998, WO 9853814. (l) Victoria, E. J.; Marquis, D. M.; Jones, D. S.; Yu, L. PCT Int. Appl. 1996, WO 9640197. (m) Horwell, D. C.; Howson, W.; Hugues,

J.; Richardson, R. S. PCT Int. Appl. 1994, WO 9409031. (n) Patel, D. V.; Kline, T. B.; Meyers, C. A.; Leftheris, K.; Bhide, R. S. Eur. Pat. Appl. 1994, EP 618221. (o) Politi, V.; Di Stazio, G.; Margonelli, A.; De Luca, G. PCT Int. Appl. 1992, WO 9206108. (p) Rosenberg, S. H.; Sham, H. L.; Baker, W. R.; Dellaria, J. F. Jr.; Kempf, D. J. Eur. Pat. Appl. 1989, EP 307837. (q) Marshall, G. R.; Moore, M. L. Eur. Pat. Appl. 1987, EP 234935. (r) Hester, J. B.; Pals, D. T.; Saneii, H. H.; Sawyer, T. K.; Schostarez, H. J.; Tenbrink, R. E.; Thaisrivongs, S. Eur. Pat. Appl. 1986, EP 173481. (s) Albrecht, H. P.; Kreiskott, H. German Pat. 1984, DE 3226242. (t) Morita, Y.; Oya, J.; Wagatsuma, K.; Shirasaka, T. Jpn. Pat. 1978, JP 53073591.

14. Gaussian 03, Revision B.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr., T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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.; C. Strain, M.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Pittsburgh PA, 2003.
15. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
16. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1993**, *37*, 785.
17. (a) Alemán, C.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Casanovas, J. *J. Phys. Chem. B* **2002**, *106*, 11849. (b) Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *Chem. Phys. Lett.* **2006**, *429*, 558. (c) Casanovas, J.;

Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Alemán, C. *J. Phys. Chem. B* **2006**, *110*, 5762.

18. (a) Allen, W. D.; Czinki, E.; Csaszar, A. G. *Eur. Chem. J.* **2004**, *10*, 4512. (b) Kang, Y. K. *J. Phys. Chem. B* **2002**, *106*, 2074. (c) Kang, Y. K. *J. Mol. Struct. (Theochem)* **2002**, *585*, 209.
19. McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639.
20. Kendall, R. A.; Dunning Jr., T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *92*, 6796.
21. (a) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999. (b) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027. (c) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239. (d) Miertus, M.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
22. Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, *113*, 6256.
23. (a) Hudáky, I.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2003**, *630*, 135. (b) Hudáky, I.; Baldoni, H. A.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2002**, *582*, 233.
24. Sahai, M. A.; Kehoe, T. A. K.; Koo, J. C. P.; Setiadi, D. H.; Chass, G. A.; Viskolcz, B.; Penke, B.; Pai, E. F.; Csizmadia, I. G. *J. Phys. Chem. A* **2005**, *109*, 2660.
25. Flores-Ortega, A.; Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2007**, *111*, 5475.
26. (a) Iribarren, J. I.; Casanovas, J.; Zanuy, D.; Alemán, C. *Chem. Phys.* **2004**, *302*, 77. (b) Jang, Y. H.; Goddard III, W. A.; Noyes, K. T.; Sowers, L. C.; Hwang, S.; Chung, D. S. *J. Phys. Chem. B* **2003**, *107*, 344. (c) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys. B* **1998**, *102*, 3257. (d) Orozco, M.; Luque, F. J. *J. Am. Chem. Soc.* **1995**, *117*, 1378. (e) Alemán, C.; Navarro, E.; Puiggalí, J. *J. Org. Chem.* **1995**, *60*, 6135.

3.3 Conformational Preferences of β– and γ–Aminated Proline Analogues

Quantum mechanical calculations have been used to investigate how the incorporation of an amino group to the C^{β}- or C^{γ}-positions of the pyrrolidine ring affects the intrinsic conformational properties of the proline. Specifically, a conformational study of the N-*acetyl*-N'-*methylamide derivatives of four isomers of aminoproline, which differ not only in the β- or γ-position of the substituent but also in its* cis *or* trans *relative disposition, has been performed. In order to further understand the role of the intramolecular hydrogen bonds between the backbone carbonyl groups and the amino side group, a conformational study was also performed on the corresponding four analogues of dimethylaminoproline. In addition, the effects of solvation on aminoproline and dimethylaminoproline dipeptides have been evaluated using a Self Consistent Reaction Field model, and considering four different solvents (carbon tetrachloride, chloroform, methanol and water). Results indicate that the incorporation of the amino substituent into the pyrrolidine ring affects the conformational properties, with backbone···side chain intramolecular hydrogen bonds detected when it is incorporated in a* cis *relative disposition. In general, the incorporation of the amino side group tends to stabilize those structures where the peptide bond involving the pyrrolidine nitrogen is arranged in* cis. *The aminoproline isomer with the substituent attached to the C^{γ}-position with a* cis *relative disposition is the most stable in the gas-phase and in chloroform, methanol and water solutions. Replacement of the amino side group by the dimethylamino substituent produces significant changes in the potential energy surfaces of the four investigated dimethylaminoproline-containing dipeptides. Thus, these changes affect not only the number of minima, which increases considerably, but also the backbone and pseudorotational preferences. In spite of these effects, comparison of the conformational preferences,* i.e. *the more favored conformers, calculated for different isomers of aminoproline and dimethylaminoproline dipeptides showed a high degree of consistency for the two families of compounds.*[*]

3.3.1 Introduction

Proline (Pro) is unique among naturally occurring amino acids in that its side chain is bonded to both the α-carbon and its preceding amide nitrogen. As a consequence, rotation about the N—C^{α} bond is prohibited and the φ torsion angle

[*] *The work described in this chapter previously appeared in J. Phys. Chem. B* **2008**, *112,* 14045–14055

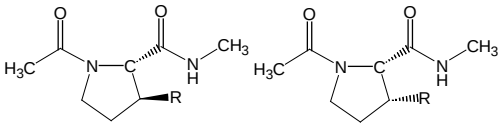
is confined to values around –60º. Accordingly, Pro is overwhelmingly found in the α-helical ($\varphi,\psi\approx$ –60º,–30º) and semi-extended ($\varphi,\psi\approx$ –60º,140º) regions of the conformational map.[1] In addition, Pro shows a higher propensity to promote γ-turn conformations ($\varphi,\psi\approx$ –70º,60º) than other proteogenic amino acids.[1d,2] Another effect derived from its cyclic structure is that the peptide bond preceding Pro (that involving the pyrrolidine nitrogen) has a relatively high probability of accommodating a *cis* arrangement[3] as compared to other peptide bonds, for which the *cis* form is almost inexistent. Recent studies in Pro dipeptides observed that the *cis*/*trans* isomerization is an enthalpy driven process that depends on the polarity of the environment.[4] Thus, although the electronic effects that stabilize the *cis* form are enhanced in polar environments, the *cis*/*trans* rotational barriers increase with the polarity of the environment. These structural features play a fundamental role in directing the secondary structure of proteins,[5] inducing special motifs like reverse turns and bends.[6] Furthermore, the *cis*-*trans* isomerization of Pro has been speculated to play a role not only in important biological processes[7] but also in the rate determining steps for folding and refolding of some proteins.[8]

4R-Hydroxyproline (Hyp) is a hydroxylated derivative of Pro that shares the same features as its parent amino acid. It is formed by a post-translational modification where a Pro residue is converted to Hyp by an enzyme with a ferrous ion at its active site, called prolyl hydroxylase. Both Hyp and Pro, along with glycine, are found in collagen, the most abundant protein in vertebrates. As a consequence of their importance, the intrinsic conformational preferences of Pro[3a,9-12] and Hyp[13] have been examined in detail on the corresponding dipeptide analogues using advanced theoretical methods. Interestingly, in spite of the capabilities of the hydroxyl side group to form intramolecular hydrogen bonds able to induce significant structural distortions, the minimum energy conformations found for the *N*-acetyl-*N*'-methylamide derivatives of Pro and Hyp (Ac-Pro-NHMe and Ac-Hyp-NHMe dipeptides, respectively) were very similar. Specifically, a strong correlation was observed between the optimized dihedral angles of these dipeptides. Indeed, the largest effect produced by hydroxylation of Pro was detected in the puckering of the pyrrolidine ring. Thus, the down puckering is preferred for Ac-Pro-NHMe, while the up puckering with the

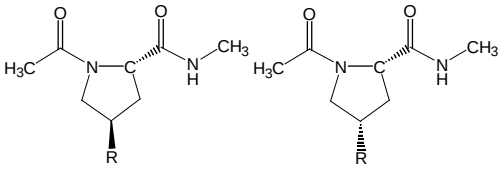
hydroxyl group occupying an equatorial position is the most stable for Ac-Hyp-NHMe.

In recent years we have been involved in a broad project devoted to the design and application of synthetic amino acids with restricted conformational mobility in different fields of nanobiology. Non-proteogenic amino acids have been found to be very useful for the re-engineering of physical protein modules and the generation of nanodevices.[14] More specifically, we observed that insertion of chemically constrained residues with suitable backbone conformational tendencies enhance the thermodynamic stability of the nanotubular structures constructed by self-assembling protein fragments with a β-helical conformation.[15] We have further selectively incorporated synthetic amino acids to impart resistance against proteases not only at the mutated position but also at neighboring amino acids.[16] In this work, we investigate the intrinsic conformational preferences of different aminated derivates of Pro. These non-proteogenic amino acids, which have been already used to construct β-peptides with helical secondary structures,[17] are expected to be of potential interest in many nanobiological applications. This is because the topological characteristics of the amino and hydroxyl groups are different and, therefore, intramolecular hydrogen bonds in aminoproline (Amp) derivatives are expected to alter significantly the structural properties of Pro.

Theoretical calculations based on Density Functional Theory (DFT) methods have been used to investigate the conformational properties of the *N*-acetyl-*N*'-methylamide derivatives of Amp that incorporate an amino group to the C^{β}- or C^{γ}-positions of the pyrrolidine ring, both the *cis* and *trans* isomers being considered in each case. Accordingly, calculations were performed on the four compounds displayed in Figure 3.3.1: Ac-β*t*Amp-NHMe, Ac-β*c*Amp-NHMe, Ac-γ*t*Amp-NHMe and Ac-γ*c*Amp-NHMe.



R= NH_2	Ac-β*t*Amp-NHMe	Ac-β*c*Amp-NHMe
R= $N(CH_3)_2$	Ac-β*t*Dmp-NHMe	Ac-β*c*Dmp-NHMe



R= NH_2	Ac-γ*t*Amp-NHMe	Ac-γ*c*Amp-NHMe
R= $N(CH_3)_2$	Ac-γ*t*Dmp-NHMe	Ac-γ*c*Dmp-NHMe

Figure 3.3.1: *Compounds studied in this work*

In order to provide a better understanding of the crucial role of intramolecular hydrogen bonds, the study has been further extended to the four dipeptides constructed by replacing the Amp residue by the corresponding dimethylaminoproline (Dmp) analogue: Ac-β*t*Dmp-NHMe, Ac-β*c*Dmp-NHMe, Ac-γ*t*Dmp-NHMe and Ac-γ*c*Dmp-NHMe in Figure 3.3.1. In addition we have examined how the incorporation of amino and dimethylamino substituents at the β and γ positions of Pro affects the *trans*/*cis* disposition of the peptide bond involving the pyrrolidine nitrogen. Finally, the influence of the environment, in particular of the solvent, on the conformational preferences of the different Amp- and Dmp-containing dipeptides has been evaluated using a Self Consistent Reaction Field (SCRF) method. Results have been compared with those recently reported for Ac-Pro-NHMe,[12b] which were calculated using the same theoretical procedures.

3.3.2 Methods

All calculations were carried out using the Gaussian 03 computer program.[18] DFT calculations were performed using the following combination: the

Becke´s three-parameter hybrid functional (B3)[19] with the local functional developed by Lee, Yang and Parr (LYP),[20] which is gradient corrected. Thus, all the calculations presented in this work were performed using the B3LYP method combined with the 6-31+G(d,p) basis set.[21] This computational procedure provided a very satisfactory description of the conformational properties of cyclic constrained amino acids, including Pro and its dehydro- and α-substituted derivatives.[22,12]

The backbone (ω_0,φ,ψ,ω) and side chain (χ^i; endocyclic) dihedral angles of the *N*-acetyl-*N'*-methylamide derivatives of conventional Pro, Amp and Dmp are defined in Figure 3.3.2. Since φ is fixed by the geometry of the five-membered ring, only three minima may be anticipated for the potential energy surfaces $E=E(\psi)$ of the dipeptides for a given arrangement of the peptide bonds. Thus, the flexible angle ψ is expected to have three minima, *i.e. gauche*$^+$ (60º), *trans* (180º) and *gauche*$^-$ (–60º), while each amide bond (ω_0, ω) can be arranged in *trans* or *cis*. It should be noted that only the peptide bond involving the pyrrolidine nitrogen, which corresponds to the ω_0 torsion angle, is likely to adopt a *cis* configuration. Therefore, both the *trans* and *cis* states were considered for ω_0, while the amide bond involving the *N*-methylamide blocking group (given by ω) was arranged in *trans* only. The cyclic side chains of the compounds under study may adopt two main different conformational states that correspond to the *down* and *up* puckering of the five-membered ring. They are defined as those in which the C^γ atom and the carbonyl group of the Pro residue (or analogue) lie on the same and opposite sides, respectively, of the plane defined by the C^δ, N and C^α atoms.

Accordingly, for each of the eight dipeptides under study (Figure 3.3.1), 3(ψ backbone) × 2(ω_0 *trans*-or-*cis*) × 2(cyclic side chain) = 12 structures were considered as starting points for complete geometry optimizations at the B3LYP/6-31+G(d,p) level. Frequency analyses were carried out to verify the nature of the minimum state of all the stationary points obtained and to calculate the zero-point vibrational energies (ZPVE) as well as both thermal and entropic corrections, these statistical terms being used to compute the conformational Gibbs free energies in the gas phase (ΔG^{gp}) at 298 K.

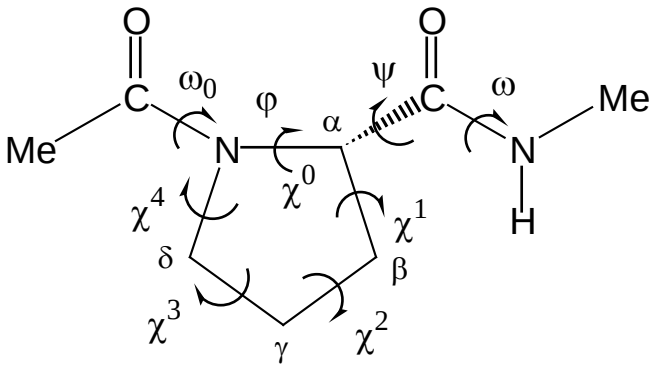


Figure 3.3.2: *Dihedral angles used to identify the conformations of the N-acetyl-N'-methylamide derivatives of the Amp and Dmp analogues studied in this work. The dihedral angles ω_0, φ, ψ and ω are defined using backbone atoms while the endocyclic dihedral angles χ^j are given by the atoms of the five-membered ring. In particular, the sequence of atoms used to define φ and χ^0 are C(=O)–N–C^α–C(=O) and C^δ–N–C^α–C^β, respectively*

To obtain an estimation of the solvation effects on the relative stability of the different minima, single point calculations were conducted on the B3LYP/6-31+G(d,p) optimized structures using a Self-Consistent Reaction Field (SCRF) model. SCRF methods treat the solute at the quantum mechanical level, while the solvent is represented as a dielectric continuum. Specifically, the Polarizable Continuum Model (PCM) developed by Tomasi and co-workers was used to describe the bulk solvent.[23] This method involves the generation of a solvent cavity from spheres centered at each atom in the molecule and the calculation of virtual point charges on the cavity surface representing the polarization of the solvent. The magnitude of these charges is proportional to the derivative of the solute electrostatic potential at each point calculated from the molecular wave function. The point charges may, then, be included in the one-electron Hamiltonian, thus inducing polarization of the solute. An iterative calculation is carried out until the wave function and the surface charges are self-consistent. PCM calculations were performed using the standard protocol implemented in Gaussian 03[18] and considering the dielectric constants of carbon tetrachloride (ε = 2.228), chloroform (ε = 4.9), methanol (ε= 32.6) and water (ε= 78.4). The conformational free energies in solution ($\Delta G^{\#sol\#}$, where #sol# refers to the

solvent) were computed using the classical thermodynamics scheme, *i.e.* the free energies of solvation provided by the PCM model were added to the ΔG^{gp} values..

3.3.2.1 Nomenclature and Pseudorotational Parameters

The minimum energy conformations of the dipeptides studied in this work have been denoted using a three-labels code that specifies the arrangement of the first peptide bond, the backbone conformation and the puckering of the five membered ring. The first letter refers to the trans (t) or cis (c) arrangement of ω_o. The second label identifies the backbone conformation using the nomenclature introduced by Perczel *et al.*[53] more than fifteen years ago. Accordingly, nine different backbone conformations can be found in the potential energy surface $E=E(\varphi,\psi)$ of amino acids: γ_D, δ_D, α_L, ε_D, β_L, ε_L, α_D, δ_L and γ_L. Finally, the up or down puckering of the five-membered ring is indicated using the labels [u] and [d], respectively. In particular, the [d] ring puckering was identified when χ^1 and χ^3 are positive while χ^2 and χ^4 are negative. Therefore, the [u] ring puckering is characterized by negative values of χ^1 and χ^3 and positive values of χ^2 and χ^4. We note that the cyclic side chain adopts a planar arrangement in two of the studied molecules, and no indication of the puckering was included in the code used for these cases.

The puckering of the five-membered ring was described using the classical pseudorotational algorithm, which uses a very simple model based on only two parameters, as was previously applied to Pro by Hudaky and Perczel.[28,29] The pseudorotational parameters are A and P, which describe the puckering amplitude and the state of the pucker in the pseudorotation pathway, respectively. The parameters are derived from the endocyclic dihedral angles as follows:

$$A = \sqrt{(A\sin P)^2 + (\chi^0)^2}\ , \text{ where } A\sin P = \frac{\chi^1 - \chi^2 + \chi^3 - \chi^4}{-2(\sin 144^o + \sin 72^o)} \qquad (3.1.1)$$

and

$$P = \begin{cases} \arccos\dfrac{\chi^0}{A}, & \text{if} \quad A\sin P \geq 0 \\ -\arccos\dfrac{\chi^0}{A}, & \text{if} \quad A\sin P < 0 \end{cases} \qquad (3.1.2)$$

Accordingly, parameter A is defined to be positive while P falls between –180º and 180º.

3.3.3 Results and Discussion

3.3.3.1 Aminoproline (Amp) dipeptides

This section reports the results obtained at the B3LYP/6-31+G(d,p) level for the four Amp-containing dipeptides displayed in Figure 3.3.1, which have been compared to the dipeptide of conventional Pro that was recently reported at the same level of theory.[12b] Table 3.3.1 lists the more relevant structural parameters together with the relative energy (ΔE^{gp}) for the 4, 6, 3 and 7 minimum energy conformations characterized for Ac-β*t*Amp-NHMe, Ac-β*c*Amp-NHMe, Ac-γ*t*Amp-NHMe and Ac-γ*c*Amp-NHMe, respectively, selected minima being displayed in Figures 3.3.3 and 3.3.4. The relative stability of the four dipeptides is indicated in Table 3.3.1 through $\Delta E^{\#gp\#}$, which corresponds to the energy relative to the lowest energy conformation of the most stable isomer. Table 3.3.2 compares the relative free energies in the gas-phase (ΔG^{gp}), carbon tetrachloride (ΔG^{CCl4}), chloroform (ΔG^{CHCl3}), methanol (ΔG^{MeOH}) and water (ΔG^{H2O}) solutions for the minima of the four dipeptides mentioned above. Calculations in solution were performed by applying the PCM method to the geometries optimized in the gas phase. Thus, previous studies on simple organic and bio-organic compounds indicated that solute geometry relaxations in solution and single point calculations on the optimized geometries in the gas phase give almost identical free energies of salvation,[26] even although nuclear relaxation in solute has been found to be essential in some specific cases.[27] Finally, Table 3.3.3 compares the relative stability of the four Amp-containing dipeptides by showing the free energies in the different environments calculated in each case with respect to the

conformation of lowest free energy of the most stable isomer: $\Delta G^{\#gp\#}$, $\Delta G^{\#CCl4\#}$, $\Delta G^{\#CHCl3\#}$, $\Delta G^{\#MeOH\#}$ and $\Delta G^{\#H2O\#}$.

Table 3.3.1 *Backbone dihedral angles (in degrees), pseudorotational parameters (A and P; in degrees), relative energy (ΔE^{gp}; in kcal/mol) and relative energy with respect to the lowest energy conformation of the most stable dipeptide ($\Delta E^{\#gp\#}$; in kcal/mol) of the minimum energy conformations characterized for Ac-βtAmp-NHMe, Ac-βcAmp-NHMe, Ac-γtAmp-NHMe and Ac-γcAmp-NHMe at the B3LYP/6-31+G(d,p) level in the gas phase.*

# Conf.	ω_0	φ	ψ	ω	*(A, P)*	ΔE^{gp}	$\Delta E^{\#gp\#}$
				Ac-β*t*Amp-NHMe			
t-γ_L[d]	-174.0	-83.1	71.9	-177.6	(38.4, -117.2)[a]	0.0[b]	1.4
c-ε_L[u]	1.2	-66.7	178.8	175.9	(38.2, 92.2)[c]	2.8	4.2
c-α_L[u]	6.8	-69.6	-35.1	179.4	(37.8, 82.2)[d]	5.9	7.3
c-ε_L[d]	-0.6	-77.5	145.8	176.0	(37.5, 75.8)[e]	5.9	7.3
				Ac-β*c*Amp-NHMe			
t-γ_L[d]	-174.1	-83.6	78.6	-175.5	(39.8,118.8)[f]	0.0[g]	0.5
t-α_L[d]	-170.5	-88.7	-7.5	173.9	(39.2, -112.5)[h]	4.0	4.5
c-α_L[d]	10.4	-84.3	-17.4	-177.2	(36.2, 104.0)[i]	4.3	4.8
c-ε_L[d]	-3.0	-74.5	172.9	177.8	(39.3, -126)[j]	5.0	5.5
c-α_L[u]	8.2	-96.2	-0.7	178.6	(42.8, -122.2)[k]	5.6	6.1
t-α_L[u]	-170.8	-70.5	-19.3	175.2	(40.2, 78.1)[l]	8.6	9.1
				Ac-γ*t*Amp-NHMe			
t-γ_L[u]	-170.8	-83.7	75.2	-176.6	(37.3, 102.6)[m]	0.0[n]	1.3
c-α_L[d]	10.5	-91.6	-3.9	-179.9	(37.8, -112.5)[o]	2.8	4.1
c-ε_L[u]	-0.8	-62.5	147.7	175.7	(39.3, 93.1)[p]	5.9	7.2
				Ac-γ*c*Amp-NHMe			

t-γ_L[d]	-172.8	-82.9	76.6	-176.0	(32.73,-116.5)[q]	0.0[r]	0.0
t-γ_L[u]	-174.5	-81.8	77.4	-176.2	(37.8, 109.3)[s]	1.5	1.5
c-α_L[d]	8.5	-68.5	-47.6	-176.4	(34.1, -93.2)[t]	2.7	2.7
c-α_L[u]	8.0	-79.0	-18.9	-177.6	(37.2, 89.4)[u]	5.2	5.2
t-α_L[u]	-170.9	-79.3	-9.2	175.5	(37.9, 93)[v]	5.7	5.7
c-ε_L[d]	3.8	-70.8	148.6	174.6	(35.4, -104.1)[w]	6.0	6.0
c-ε_L[u]	0.9	-62.5	148.2	176.9	(39, 89.3)[x]	7.9	7.9

[a] χ^0= -17.5º, χ^1= 33.5º, χ^2= -38.0º, χ^3= 27.4º and χ^4= -6.0º. [b] E= -628.667446 a.u. [c] χ^0= -1.4º, χ^1= -21.6º, χ^2= 35.7º, χ^3= -36.2º and χ^4= 23.9º. [d] χ^0= 5.1º, χ^1= -26.01º, χ^2= 37.3º, χ^3= -33.9º and χ^4= 18.1º. [e] χ^0= -10.3º, χ^1= -13.4º, χ^2= 31.0º, χ^3= -36.6º and χ^4= 29.8º. [f] χ^0= -18.7º, χ^1= 34.5º, χ^2= -38.3º, χ^3= 27.0º and χ^4= -5.0º. [g] E= -628.668896 a.u. [h] χ^0= -15.0º, χ^1= 32.9º, χ^2= -39.2º, χ^3= 30.2º and χ^4= -9.2º. [i] χ^0= -8.8º, χ^1= 27.5º, χ^2= -36.3º, χ^3= 30.7º and χ^4= -13.7º. [j] χ^0= -23.1º, χ^1= 36.5º, χ^2= -37.4º, χ^3= 23.8º and χ^4= -0.1º. [k] χ^0= -22.8º, χ^1= 39.0º, χ^2= -41.6º, χ^3= 28.0º and χ^4= -3.0º. [l] χ^0= 8.3º, χ^1= -29.5º, χ^2= 39.9º, χ^3= -35.0º and χ^4= 16.7º. [m] χ^0= -8.2º, χ^1= -15.5º, χ^2= 31.9º, χ^3= -36.3º and χ^4= 28.4º. [n] E= -628.667515 a.u. [o] χ^0= -14.4º, χ^1= 31.0º, χ^2= -37.7º, χ^3= 28.7º and χ^4= -9.1º. [p] χ^0= -2.1º, χ^1= -21.9º, χ^2= 36.6º, χ^3= -37.1º and χ^4= 25.1º. [q] χ^0=-14.6º, χ^1= 28.9º, χ^2= -32.4º, χ^3= 23.1º and χ^4= -5.6º. [r] E= -628.669674 a.u. [s] χ^0= -12.5º, χ^1= -11.8º, χ^2= 29.9º, χ^3= -36.8º and χ^4= 31.3º. [t] χ^0= -1.9º, χ^1= 21.8º, χ^2= -33.0º, χ^3= 31.1º and χ^4= -18.7º. [u] χ^0= 0.4º, χ^1= -22.5º, χ^2= 35.4º, χ^3= -34.6º and χ^4= 21.9º. [v] χ^0= -2.0º, χ^1= -21.2º, χ^2= 35.2º, χ^3= -35.9º and χ^4= 24.1º. [w] χ^0= -8.6º, χ^1= 27.4º, χ^2= -35.3º, χ^3= 29.6º and χ^4= -13.4º. [x] χ^0= 0.5º, χ^1= -23.7º, χ^2= 37.2º, χ^3= -36.3º and χ^4= 22.8º.

Table 3.3.2 *Relative free energy in the gas-phase (ΔG^{gp}; in kcal/mol) and in carbon tetrachloride, chloroform, methanol and aqueous solutions (ΔG^{CCl4}, ΔG^{CHCl3}, ΔG^{CH3OH} and ΔG^{H2O}, respectively; in kcal/mol) for the minimum energy conformations of Ac-βtAmp-NHMe, Ac-βcAmp-NHMe, Ac-γtAmp-NHMe and Ac-γcAmp-NHMe at the B3LYP/6-31+G(d,p) level.*

# Conf.	ΔG^{gp}	ΔG^{CCl4}	ΔG^{CHCl3}	ΔG^{CH3OH}	ΔG^{H2O}
		Ac-β*t*Amp-NHMe			
t-γ_L[d]	0.0[a]	0.0	1.3	5.8	6.4
c-ε_L[u]	2.7	1.0	0.8	3.2	3.5
c-α_L[u]	4.7	3.0	2.5	4.1	4.2
c-ε_L[d]	3.9	1.6	0.0	0.0	0.0
		Ac-β*c*Amp-NHMe			
t-γ_L[d]	0.0[b]	1.0	0.0	2.7	3.3
t-α_L[d]	2.7	0.0	0.2	1.1	1.4
c-α_L[d]	3.3	0.6	1.3	2.5	2.3
c-ε_L[d]	4.4	3.3	0.3	0.0	0.0
c-α_L[u]	4.4	1.6	2.6	3.9	3.7
t-α_L[u]	7.6	6.4	2.9	1.5	1.2
		Ac-γ*t*Amp-NHMe			
t-γ_L[u]	0.0[c]	0.0	0.5	4.4	5.1
c-α_L[d]	1.7	4.2	0.4	2.9	2.9
c-ε_L[u]	4.4	5.4	0.0	0.0	0.0
		Ac-γ*c*Amp-NHMe			
t-γ_L[d]	0.0[d]	0.4	2.3	7.0	7.9
t-γ_L[u]	1.3	0.8	1.7	5.4	6.3
c-α_L[d]	2.2	0.0	0.0	2.7	3.4
c-α_L[u]	3.9	2.3	2.2	4.4	4.3
t-α_L[u]	3.7	1.2	1.5	2.9	3.2
c-ε_L[d]	4.6	2.3	1.3	1.8	2.1
c-ε_L[u]	6.0	2.8	0.8	0.0	0.0

[a] G=-628.468989 a.u. [b] G= -628.469683 a.u. [c] G=-628.468329 a.u. [d] G= -628.470592 a.u.

Table 3.3.3 *Relative free energy in the gas-phase ($\Delta G^{\#gp\#}$; in kcal/mol) and in carbon tetrachloride, chloroform, methanol and aqueous solutions ($\Delta G^{\#CCl4\#}$, $\Delta G^{\#CHCl3\#}$, $\Delta G^{\#CH3OH\#}$ and $\Delta G^{\#H2O\#}$, respectively; in kcal/mol) calculated at the B3LYP/6-31+G(d,p) level for the minimum energy conformations of Ac-βtAmp-NHMe, Ac-βcAmp-NHMe, Ac-γtAmp-NHMe and Ac-γcAmp-NHMe with respect to the lowest energy minimum of the most stable isomer.*

# Conf.	$\Delta G^{\#gp\#}$	$\Delta G^{\#CCl4\#}$	$\Delta G^{\#CHCl3\#}$	$\Delta G^{\#CH3OH\#}$	$\Delta G^{\#H2O\#}$
			Ac-β*t*Amp-NHMe		
t-γ_L[d]	1.0	2.6	2.0	5.8	6.6
c-ε_L[u]	3.7	3.7	1.5	3.2	3.7
c-α_L[u]	5. 7	5.6	3.2	4.1	4.4
c-ε_L[d]	4. 9	4.2	0.7	0.0	0.2
			Ac-β*c*Amp-NHMe		
t-γ_L[d]	0.6	2.7	2.5	7.0	7.8
t-α_L[d]	3.2	1.7	2.8	5.4	6.0
c-α_L[d]	3.8	2.3	3.8	6.8	6.9
c-ε_L[d]	5.0	5.0	2.8	4.3	4.6
c-α_L[u]	5.0	3.3	5.1	8.2	8.3
t-α_L[u]	8.2	8.1	5.4	5.6	5.9
			Ac-γ*t*Amp-NHMe		
t-γ_L[u]	1.4	0.0	2.1	5.9	6.8
c-α_L[d]	3.2	3.7	2.0	4.4	4.6
c-ε_L[u]	5.8	4.9	1.6	1.5	1.7
			Ac-γ*c*Amp-NHMe		
t-γ_L[d]	0.0	2.2	2.3	7.0	7.9
t-γ_L[u]	1.3	2.6	1.7	5.4	6.3
c-α_L[d]	2.2	1.8	0.0	2.3	3.4
c-α_L[u]	3.9	4.1	2.2	4.4	4.3
t-α_L[u]	3.7	3.0	1.5	2.9	3.2
c-ε_L[d]	4.6	4.1	1.3	1.8	2.1
c-ε_L[u]	6.0	4.6	0.8	0.0	0.0

Our previous calculations revealed three minimum energy conformations for Ac-Pro-NHMe when the two amide bonds are arranged in *trans*: t-γ_L[d], t-γ_L[u] and t-α_L[u], the two latter being 1.0 and 4.9 kcal/mol, respectively, less stable than the former.[12b] These results, which are in excellent agreement with those reported by other authors,[10,3b] indicate that the backbone of conventional Pro tends to adopt a turn-like conformation that is compatible with both the *down* and *up* puckerings of the pyrrolidine ring, even though the former is the most favored. Table 3.3.1 indicates that β- and γ-amination reduce the intrinsically restricted conformational flexibility of conventional Pro, especially when the amino group is introduced in *trans* position. Thus, only one minimum energy conformation was found for Ac-β*t*Amp-NHMe and Ac-γ*t*Amp-NHMe when ω_0 is arranged in *trans*, which corresponds to the t-γ_L[d] (Figure 3.3.3a) and t-γ_L[u] (Figure 3.3.3b), respectively. Moreover, although three minima were characterized for Ac-β*c*Amp-NHMe, the t-γ_L[d] conformation (Figure 3.3.3c) is favored with respect to the t-α_L[d] and t-α_L[u] ones by 4.0 and 8.6 kcal/mol, respectively. Finally, the t-γ_L[d] (Figure 3.3.3d), t-γ_L[u] and t-α_L[u] structures were identified as minimum energy conformations of Ac-γ*c*Amp-NHMe, the two latter being unfavored with respect to the former by 1.5 and 5.7 kcal/mol, respectively. It is worth noting that these energy differences are larger by 0.5 and 0.8 kcal/mol, respectively, than those calculated for the same structures in Ac-Pro-NHMe.

As a consequence of their γ_L conformation, the lowest energy minimum of all the Amp-containing dipeptides studied in this work is stabilized by a seven-membered intramolecular hydrogen bond (typically denoted as C_7) between the NH and O=C moieties of the NHMe and Ac blocking groups (Figure 3.3.3). However, it is worth noting that only the t-γ_L conformers of the dipeptides with the amino group arranged in *cis* are able to form simultaneously an intramolecular hydrogen bond between the side amino group and the O=C of the own Amp residue. Thus, the stability of the Ac-γ*c*Amp-NHMe and Ac-β*c*Amp-NHMe dipeptides, which is higher than that of the corresponding analogues with the amino group in *trans*, should be attributed to the formation of this intraresidue hydrogen bond (Figures 3.3.3c and 3.3.3d). This feature is reflected by the $\Delta E^{\#gp\#}$ values displayed in Table 3.3.1. On the other hand, it should be mentioned that, independently of both the arrangement of the peptide bond ω_0 and the puckering

of the five membered ring, the minima with an α_L conformation of the four Amp-containing derivatives are typically stabilized by an intramolecular hydrogen bond between the N-H moiety of the NHMe blocking group and the nitrogen atom of the pyrrolidine ring. This interaction is illustrated in Figure 3.3.3 for the c-α_L[d] conformation of both Ac-γ*t*Amp-NHMe and Ac-β*c*Amp-NHMe.

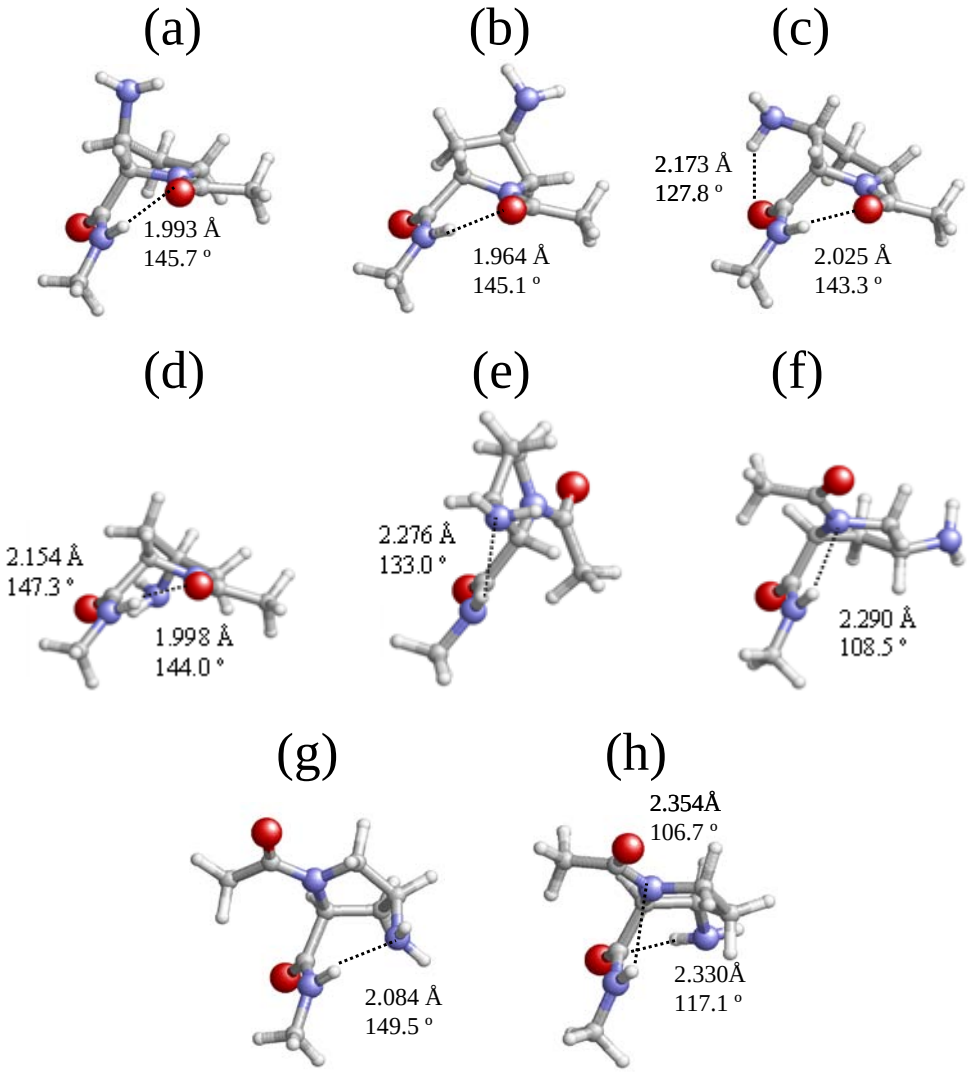


Figure 3.3.3: *Representation of selected minimum energy conformations characterized in the gas-phase for the Amp-containing dipeptides studied in this work: (a) t-γ_L[d] for Ac-β*t*Amp-NHMe; (b) t-γ_L[u] for Ac-γ*t*Amp-NHMe; (c) t-γ_L[d] for Ac-β*c*Amp-NHMe; (d) t-γ_L[d] for Ac-γ*c*Amp-NHMe; (e) c-ε_L[u] for Ac-β*t*Amp-NHMe; (f) c-α_L[d] for Ac-γ*t*Amp-NHMe; (g) c-α_L[d] for Ac-γ*c*Amp-NHMe; and (h) c-α_L[d] for Ac-β*c*Amp-NHMe.*

Regarding the structures with ω_0 arranged in *cis*, four minimum energy conformations were characterized for the Ac-Pro-NHMe dipeptide: c-α_L[d], c-α_L[u], c-ε_L[d] and c-ε_L[u], which were destabilized by 3.3, 4.2, 6.3 and 6.6

kcal/mol with respect to the global minimum.[12b] Interestingly, the energy difference between the global minimum and most stable conformations with ω_0 arranged in *cis* is about 0.5 kcal/mol lower for Ac-β*t*Amp-NHMe, Ac-γ*t*Amp-NHMe and Ac-γ*c*Amp-NHMe than for Ac-Pro-NHMe. Three minima were characterized for Ac-β*t*Amp-NHMe, the c-ε_L[u] structure (Figure 3.3.3e) being more stable than the c-α_L[u] and the c-ε_L[d] by 3.1 kcal/mol. The c-ε_L[u] presents an intramolecular hydrogen bond between the amino group, which acts as hydrogen bonding acceptor, and the NHMe blocking group. Only two minimum energy conformations were detected for Ac-γ*t*Amp-NHMe, the c-α_L (Figure 3.3.3f) being more stable than the c-ε_L by 3.1 kcal/mol. Regarding Ac-γ*c*Amp-NHMe, four structures with the dihedral angle ω_0 arranged in *cis* were characterized as energy minima. The most favored one is the c-α_L[d] (Figure 3.3.3g), while the c-α_L[u], c-ε_L[d] and c-ε_L[u] are higher in energy by 2.5, 3.3 and 5.6 kcal/mol. In this case, the c-α_L[u] and c-ε_L[u] do not show intramolecular hydrogen bonding interactions, while such type of interactions were found in the c-α_L[d] and c-ε_L[d] structures. Specifically, the amino nitrogen acts as hydrogen bonding acceptor with respect to the NHMe blocking group in the former conformation, whereas in the latter one the amino side group interacts with the carboxyl oxygen of the γ*c*Amp residue.

The situation is completely different for Ac-β*c*Amp-NHMe since, in this case, the minimum of lowest energy with ω_0 in *cis*, c-α_L[d] (Figure 3.3.3h), is less stable than the global minimum by 4.3 kcal/mol. This result indicates that the *cis*-amination of the β-carbon atom produces a significant destabilization of the *cis* disposition of the ω_0 amide bond with reference to that observed for Pro. As can be seen in Figure 3.3.3, the c-α_L[d] structure shows two intramolecular hydrogen bonds. The first is between the amino side group, which acts as donor, and the carboxyl oxygen of the β*c*Amp residue, and the second corresponds to the interaction between the N-H of NHMe and the nitrogen of the pyrrolidine. The other two minimum energy conformations with ω_0 arranged in *cis*, c-ε_L[d] and c-α_L[u], are unfavored with respect to the c-α_L[d] one by 1.0 and 1.6 kcal/mol, respectively. In spite of their high ΔE^{gp} values (Table 3.3.1), these two structures

show stabilizing intramolecular hydrogen bonds between the amino side group and the NHMe (c-ε_L[d]) or Ac (c-α_L[u]) blocking groups.

Inspection of the free energies listed in Table 3.3.2 shows the importance of the ZPVE, thermal and entropic corrections to the relative stability of the minimum energy conformations found for the Amp-containing dipeptides. Consideration of these statistical terms for the transformation of ΔE^{gp} into ΔG^{gp} represents relative variations usually higher than 1 kcal/mol. Specifically, the largest change which was detected in both the c-α_L[u] of Ac-β*t*Amp-NHMe and the t-α_L[u] of Ac-γ*c*Amp-NHMe, is 2.0 kcal/mol. Nonetheless, the t-γ_L is still the most favored conformer for the four dipeptides, the ΔG^{gp} of the other structures being higher than 1.5 kcal/mol. Thus, according to a Bolzmann distribution of minima, the ΔG^{gp} values indicate that the population at room temperature of all the local minima is negligible for the four dipeptides under study, *i.e.* the t-γ_L is the only populated conformation.

Table 3.3.2 includes the relative free energies in carbon tetrachloride, chloroform, methanol and water solutions. The solvent introduces significant changes in the relative stability of the different minima characterized for Amp-containing dipeptides. In general, carbon tetrachloride was found to considerably stabilize conformers with the amide bond ω_0 arranged in *cis*. Thus, although the most stable conformations of Ac-β*t*Amp-NHMe and Ac-β*c*Amp-NHMe, t-γ_L[d] and t-α_L, respectively show the two amide bonds in *trans*, yet ω_0 presents a *cis* arrangement in the most favored conformation of Ac-γ*c*Amp-NHMe, c-α_L[d] (Figure 3.3.3g). As a consequence of the stabilization produced by this solvent in conformers with a *cis* peptide bond, these structures are within the set of conformations energetically accessible at room temperature. Interestingly, the conformational behavior detected for Ac-γ*t*Amp-NHMe in carbon tetrachloride is completely different from that discussed for the other three dipeptides, in this case the relative stability of the c-α_L[d] and c-ε_L[u] structures being smaller than in the gas-phase.

The higher polarity of chloroform results in a further stabilization of conformers with *cis* peptide bonds. Thus, the most favored conformation in this environment shows the amide bond ω_0 arranged in *cis* for three of the dipeptides under study, *i.e.* c-ε_L[d] (Figure 3.3.4a), c-ε_L[u] (Figure 3.3.4b) and c-α_L[d]

(Figure 3.3.4c) for Ac-β*t*Amp-NHMe, Ac-γ*t*Amp-NHMe and Ac-γ*c*Amp-NHMe, respectively. The only exception to this behaviour was for Ac-β*c*Amp-NHMe, in which the lowest energy minimum corresponds to the t-γ_L[d] conformation (Figure 3.3.3c). However, it should be noted that in this case the t-α_L[d] and c-ε_L[d] structures are destabilized by only 0.2 and 0.3 kcal/mol, respectively. On the other hand, the ΔG^{CHCl3} value of the least stable conformer is 2.5, 2.9, 0.5 and 2.3 kcal/mol for Ac-β*t*Amp-NHMe, Ac-β*c*Amp-NHMe, Ac-γ*t*Amp-NHMe and Ac-γ*c*Amp-NHMe, respectively, suggesting that chloroform induces a strong stabilizing effect in all the structures.

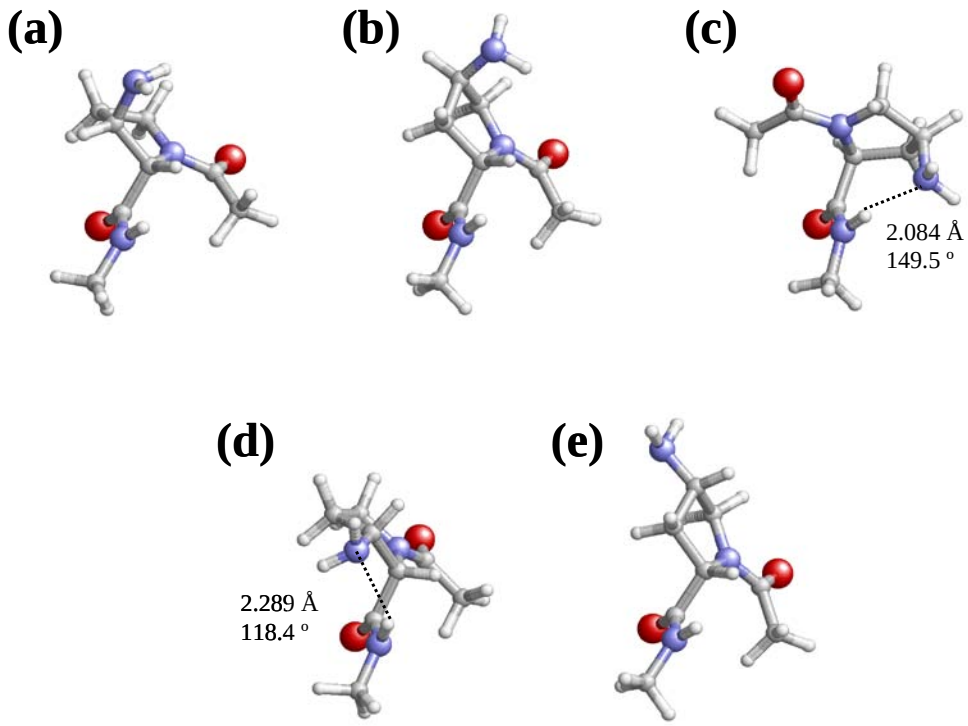


Figure 3.3.4: *Representation of selected minimum energy conformations characterized for the Amp-containing dipeptides studied in this work: (a) c-ε_L[d] for Ac-β*tAmp-NHMe; (b) c-ε_L[u] for Ac-γtAmp-NHMe; (c) c-α_L[d] for Ac-γcAmp-NHMe; (d) c-ε_L[d] for Ac-βcAmp-NHMe; (e) c-ε_L[u] for Ac-γtAmp-NHMe. These minima are especially relevant because they are relatively stable in chloroform, methanol and/or aqueous solution.*

The c-ε_L is the most stable conformation in both methanol and aqueous solutions for all the Amp-containing dipeptides, the only difference between them being the puckering of the ring. Thus, the two β-aminated dipeptides prefer a

down puckering, while the ring is arranged *up* when the substituent is introduced at the γ-carbon atom. The conformational characteristics of these structures are displayed in Figure 3.3.4. However, the most remarkable result in polar environments is the destabilization of the remaining structures, especially those with ω_0 arranged in *trans*. This feature is fully consistent with theoretical estimations previously reported for the Pro dipeptide.[3] Thus, it was found that the electronic effect that stabilize the *cis* form of the peptide bond become enhanced in polar environments, even though the *cis*/*trans* rotational barriers increase with the polarity of the solvent.

Although the stability of the *cis* conformers in solution was found to be overestimated by PCM for Pro derivatives, especially in protic solvents able to form specific hydrogen bonds with the solute, the general tendencies provided by this theoretical method describe very satisfactorily the experimental observations from a qualitative point of view.[12] Thus, in a recent study PCM calculations predicted that ω_0 exhibits a considerably smaller probability of adopting a *cis* disposition in α-methylproline and α-phenylproline than in Pro,[12b] which was in good agreement with experimental information.[29] Comparison of the results provided in Table 3.3.2 for Amp-containing dipeptides with those reported for Ac-Pro-NHMe at the same theoretical level suggests that the incorporation of the substituent to the pyrrolidine ring enhances, in general, the stability of the conformers with ω_0 arranged in *cis*. Thus, although the c-ε_L[u] conformation was predicted as the most favored for Ac-Pro-NHMe in both chloroform and water, the lowest energy structure with ω_0 in trans was unfavored by only 0.3 kcal/mol (t-γ_L[d]) and t-α_L[u], respectively).[12b] Table 3.3.2 illustrates that this energy difference is higher for the investigated Amp-containing dipeptides. However, caution is required in the analysis of PCM results, especially when protic solvents able to form specific solute-solvent interactions are considered.

Table 3.3.3 shows the free energies relative to the lowest energy minimum of the most stable Amp-containing isomer for each environment. As can be seen the most favored isomer in the gas-phase, chloroform, methanol and water solutions is the Ac-γ*c*Amp-NHMe dipeptide, even though as reflected in Table 3.3.2, the preferred conformation depends on the polarity of the environment. Moreover, in the gas-phase the most stable conformation of each isomer shows $\Delta G^{\#gp\#} < 1.5$

kcal/mol indicating that the relative stability of the other three dipeptides is still significant. However, in chloroform, methanol and aqueous solutions only one isomer, the Ac-β*t*Amp-NHMe dipeptide, satisfies such condition. These results clearly indicate that the stability of Ac-β*c*Amp-NHMe and Ac-γ*t*Amp-NHMe decreases with the polarity of the environment. Finally, it should be noted that the Ac-γ*t*Amp-NHMe is the lowest energy isomer in carbon tetrachloride solution. Accordingly, it can be concluded that Ac-β*t*Amp-NHMe is stabilized by the favorable electrostatic interactions between the solute and the solvent, while Ac-γ*t*Amp-NHMe is preferred in non-polar organic solvent where solute-solvent interactions are dominated by non-electrostatic terms, *i.e.* van der Waals and cavitation.

3.3.3.2 Dimetylaminoproline (Dmp) dipeptides

Results provided by B3LYP/6-31+G(d,p) calculations for the four Dmp-containing dipeptides (Figure 3.3.1) are reported in Tables 3.3.4, 3.3.5 and 3.3.6, atomistic pictures of the more relevant minima being displayed in Figures 3.3.5 and 3.3.6.

The ΔE^{gp} values displayed in Table 3.3.4 indicate that the conformational preferences of the Dmp-containing dipeptides are completely different from those described in the previous section for the Amp-containing ones. Seven minimum energy conformations, including those with the peptide bond ω_0 arranged in *cis*, were obtained for Ac-β*t*Dmp-NHMe, while four were found for Ac-β*t*Amp-NHMe. The only structures detected for the former dipeptide below a relative energy threshold value of 1.5 kcal/mol were the t-γ_L[d] and t-γ_L[u] (Figures 3.3.5a and 3.3.5b), which are almost isoenergetic and present a seven-membered hydrogen bonded ring. This represents another important difference with respect to Ac-β*t*Amp-NHMe, since for this compound the only structure with $\Delta E^{gp} < 1.5$ kcal/mol was the t-γ_L[d]. The relative stability of the c-ε_L[u] conformation (Figure 3.3.5c), which is the most stable structure with ω_0 in *cis*, with respect to the global minimum is similar for both Ac-β*t*Dmp-NHMe and Ac-β*t*Amp-NHMe. This is a surprising feature since in the former dipeptide, this conformation presents a stabilizing hydrogen bond between the N-H of the NHMe blocking group and the nitrogen of the dimethylamino substituent that was not detected in the latter.

Furthermore, remarkable differences appear in the ΔE^{gp} of the minima with c-α_L backbone conformation. Thus, these are more stable in Ac-β*t*Dmp-NHMe than in the corresponding β*t*Amp-containing analogue by about 3 kcal/mol.

Table 3.3.4 *Backbone dihedral angles (in degrees), pseudorotational parameters (A and P; in degrees), relative energy (ΔE^{gp}; in kcal/mol) and relative energy with respect to the lowest energy conformation of the most stable dipeptide ($\Delta E^{\#gp\#}$; in kcal/mol) of the minimum energy conformations characterized for Ac-βtDmp-NHMe, Ac-βcDmp-NHMe, Ac-γtDmp-NHMe and Ac-γcDmp-NHMe at the B3LYP/6-31+G(d,p) level in the gas phase.*

# Conf.	ω_0	φ	ψ	ω	*(A, P)*	ΔE^{gp}	$\Delta E^{\#gp\#}$
				Ac-β*t*Dmp-NHMe			
t-γ_L[d]	-170.6	-85.0	72.0	-176.9	(40.5, -110.2)[a]	0.0[b]	2.3
t-γ_L[u]	-175.1	-81.5	78.8	-175.8	(37.5, 110.2)[c]	0.2	2.5
c-ε_L[u]	0.1	-67.4	179.9	176.0	(38.7, 95.0)[d]	2.9	5.2
c-α_L[d]	12.0	-94.1	-4.9	179.3	(39.4, -111.2)[e]	3.2	5.5
c-α_L[u]	3.9	-84.1	-17.2	179.8	(35.6, 113.7)[f]	3.4	5.7
c-ε_L[d]	3.6	-79.6	141.4	177.8	(39.1, -111.9)[g]	5.0	7.3
c-ε_L[u]	-0.2	-74.8	120.7	-178.5	(38.4, 106.5)[h]	6.3	8.6
				Ac-β*c*Dmp-NHMe			
t-γ_L[d]	179.1	-77.8	122.0	-174.1	(39.2, -123.2)[i]	0.0[j]	4.0
c-ε_L[d]	0.9	-87.3	-137.3	-177.6	(43.1, -134.6)[k]	1.6	5.6
t-α_L[d]	-171.8	-81.7	-23.9	174.4	(39.6, -106.7)[l]	1.7	5.7
t-ε_L[d]	175.7	-92.9	-159.7	180.0	(46.2, -142.2)[m]	1.8	5.8
c-α_L[d]	7.6	-83.7	-27.2	178.0	(38.2, -110.9)[n]	2.0	6.0
t-ε_L[u]	179.5	-65.1	143.8	-176.5	(32.5, 99.3)[o]	4.2	8.2
c-ε_L[d]	-1.5	-81.8	138.3	174.1	(40.6, -125.2)[p]	4.4	8.4
t-α_L[u]	-172.1	-65.9	-26.9	175.9	(42.6, 75.3)[q]	5.6	9.6

c-α_L[u]	9.8	-66.0	-32.0	-175.4	(41.8, 76.2)[r]	5.6	9.6
c-ε_L[u]	2.3	-62.0	156.9	171.6	(31.8, 88.3)[s]	6.6	10.6
Ac-γ*t*Dmp-NHMe							
t-γ_L[d]	-171.6	-83.9	69.5	-178.1	(37.1, -109.5)[t]	0.0[u]	0.0
t-γ_L[u]	-171.0	-83.1	75.7	-176.5	(40.3, 98.3)[v]	3.1	3.1
c-α_L[u]	8.9	-89.4	-4.5	-179.3	(37.2, -111.4)[w]	3.4	3.4
c-ε_L[d]	1.0	-73.1	148.6	-175.9	(36.1, -112.2)[x]	6.5	6.5
t-α_L[u]	-168.1	-81.0	-6.9	175.3	(40.8, 87.9)[y]	6.7	6.7
c-ε_L[u]	4.1	-65.0	143.8	177.7	(41.8, 76.2)[z]	8.0	8.0
Ac-γ*c*Dmp-NHMe							
t-γ_L[u]	-173.6	-81.5	74.5	-176.9	(38.1, 105.2)[aa]	0.0[bb]	1.3
t-γ_L[d]	-171.1	-79.2	47.8	175.4	(36.7, -105.2)[cc]	2.0	3.3
c-α_L[u]	8.6	-79.0	-18.8	-176.1	(37.1, 89.1)[dd]	2.7	4.0
c-α_L[d]	7.8	-67.5	-44.5	-176.3	(38.1, -92.7)[ee]	3.2	4.5
t-α_L[u]	-170.8	-77.5	-11.3	-176.0	(37.5, 88.6)[ff]	3.5	4.8
c-α_L[d]	9.7	-90.8	3.7	-175.5	(38.1, -111.7)[gg]	3.6	4.9
c-ε_L[u]	1.1	-62.1	147.5	177.0	(37.4, 87.6)[hh]	5.5	6.8
c-ε_L[d]	-0.1	-76.4	-178.8	178.6	(40.8, -117.9)[ii]	8.9	10.2

[a] χ^0= -14.0º, χ^1= 32.9º, χ^2= -40.5º, χ^3= 32.3º and χ^4= -11.3º. [b]E= -707.281408 a.u. [c] χ^0= -13.0º, χ^1= -10.8º, χ^2= 29.3º, χ^3= -36.7º and χ^4= 31.6º. [d] χ^0= -3.4º, χ^1= -20.3º, χ^2= 35.4º, χ^3= -37.0º and χ^4= 25.8º. [e] χ^0= -14.3º, χ^1= 32.4º, χ^2= -39.3º, χ^3= 31.1º and χ^4= -10.3º. [f] χ^0= -14.3º, χ^1= -8.1º, χ^2= 26.5º, χ^3= -34.7º and χ^4= 30.9º. [g] χ^0= -14.6º, χ^1= 32.3º, χ^2= -39.0º, χ^3= 30.6º and χ^4= -9.7º. [h] χ^0= -10.9º, χ^1= -13.3º, χ^2= 31.6º, χ^3= -37.7º and χ^4= 30.7º. [i] χ^0= -21.4º, χ^1= 35.9º, χ^2= -37.9º, χ^3= 25.1º and χ^4= -2.1º. [j] E= -707.278763 a.u. [k] χ^0= -30.3º, χ^1= 41.5º, χ^2= -38.6º, χ^3= 20.8º and χ^4= 6.4º. [l] χ^0= -11.4º, χ^1= 31.2º, χ^2= -39.8º, χ^3= 32.7º and χ^4= -13.2º. [m] χ^0= -36.5º, χ^1= 44.8º, χ^2= -38.0º, χ^3= 17.0º and χ^4= 12.7º. [n] χ^0= -13.6º, χ^1= 31.6º, χ^2= -38.3º, χ^3= 30.0º and χ^4= -10.0º. [o] χ^0= -5.3º, χ^1= -14.8º, χ^2= 28.8º, χ^3= -31.8º and χ^4= 23.4º. [p] χ^0= -23.4º, χ^1= 37.6º, χ^2= -38.8º, χ^3= 25.0º and χ^4= -0.7º. [q] χ^0= 10.8º, χ^1= -32.3º, χ^2= 42.4º, χ^3= -36.4º and χ^4= 15.6º. [r] χ^0= 10.0º, χ^1= -31.3º, χ^2= 41.7º,

χ^3= -36.2º and χ^4= 15.9º. [s] χ^0= 0.9º, χ^1= -19.3º, χ^2= 30.4º, χ^3= -29.9º and χ^4= 18.2º. [t] χ^0= -12.4º, χ^1= 30.6º, χ^2= -37.2º, χ^3= 29.2º and χ^4= -10.7º. [u] E=-707.285136 a.u. [v] χ^0= -5.8º, χ^1= -19.6º, χ^2= 35.8º, χ^3= -38.6º and χ^4= 28.6º. [w] χ^0= -13.6º, χ^1= 31.2º, χ^2= -37.1º, χ^3= 28.6º and χ^4= -9.6º. [x] χ^0= -13.7º, χ^1= 30.6º, χ^2= -36.0º, χ^3= 27.5º and χ^4= -8.8º. [y] χ^0= 1.5º, χ^1= -25.8º, χ^2= 38.9º, χ^3= -37.5º and χ^4= 23.2º. [z] χ^0= 10.0º, χ^1= -31.3º, χ^2= 41.7º, χ^3= -36.2º and χ^4= 15.9º. [aa] χ^0= -10.0º, χ^1= -14.3º, χ^2= 31.7º, χ^3= -37.1º and χ^4= 30.0º. [bb] E= -707.283071 a.u. [cc] χ^0= -9.6º, χ^1= 29.01º, χ^2= -36.7º, χ^3= 30.1º and χ^4= -13.3º. [dd] χ^0= 0.6º, χ^1= -22.7º, χ^2= 35.3º, χ^3= -34.5º and χ^4= 21.6º. [ee] χ^0= -1.8º, χ^1= 24.4º, χ^2= -36.6º, χ^3= 34.7º and χ^4= -21.4º. [ff] χ^0= 0.9º, χ^1= -23.2º, χ^2= 35.8º, χ^3= -34.8º and χ^4= 21.6º. [gg] χ^0= -14.1º, χ^1= 32.4º, χ^2= -38.0º, χ^3= 29.0º and χ^4= -9.6º. [hh] χ^0= 1.6º, χ^1= -23.7º, χ^2= 36.0º, χ^3= -34.4º and χ^4= 21.0º. [ii] χ^0= -19.0º, χ^1= 36.4º, χ^2= -40.0º, χ^3= 28.5º and χ^4 = -6.1 º

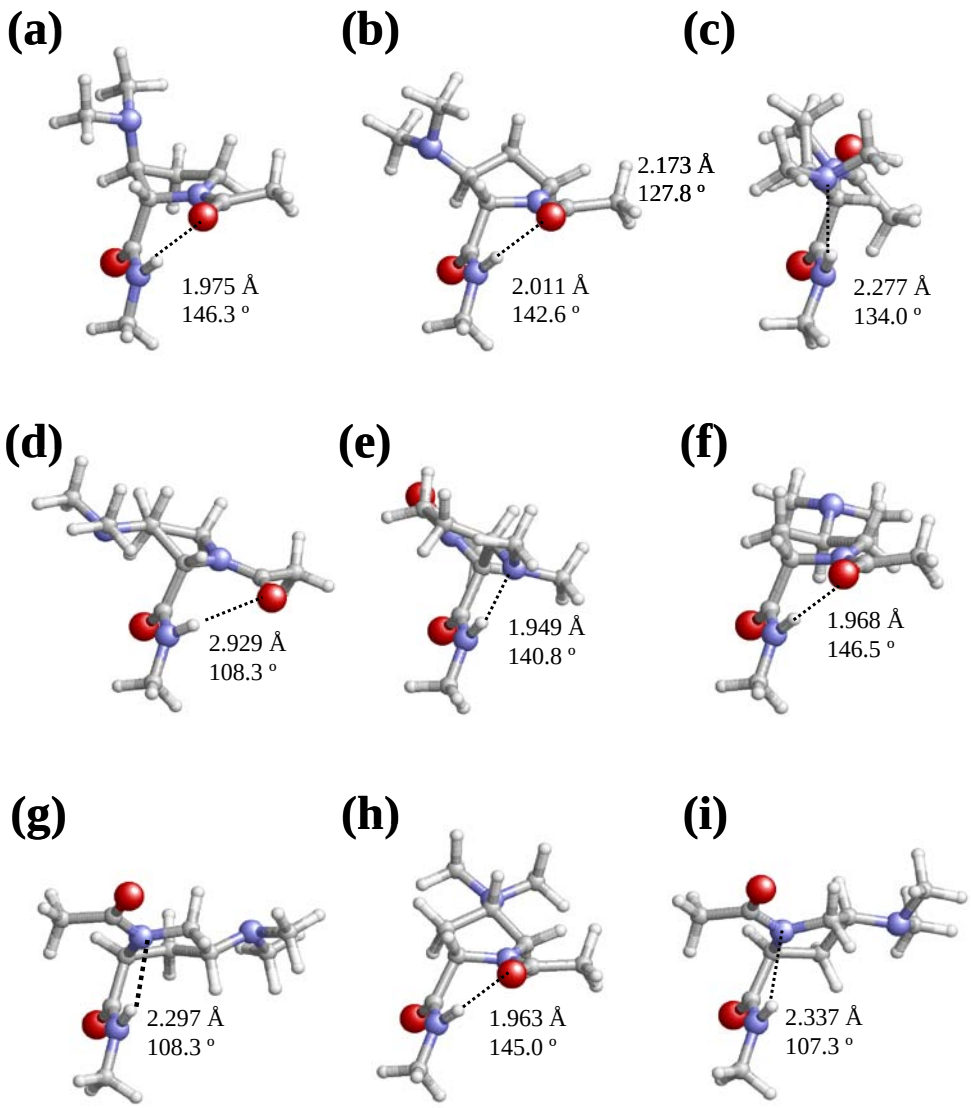


Figure 3.3.5: *Representation of selected minimum energy conformations characterized in the gas-phase for the Dmp-containing dipeptides studied in this work: (a) t-γ_L[d], (b) t-γ_L[u] and (c) c-ε_L[u] for Ac-β*t*Dmp-NHMe; (d) t-γ_L[d] and (e) c-ε_L[d] for Ac-β*c*Dmp-NHMe; (f) t-γ_L[d] and (g) c-α_L[u] for Ac-γ*t*Dmp-NHMe; (h) t-γ_L[u] and (i) c-α_L[u] for Ac-γ*c*Dmp-NH.*

A total of 10 minimum energy conformations were characterized for Ac-β*c*Dmp-NHMe, 5 for each arrangement of ω_0. Surprisingly, the lowest energy conformation corresponds to a t-γ_L[d] with the dihedral angles φ and ψ significantly distorted towards those of a conventional t-ε_L (Figure 3.3.5d). As indicated by the corresponding geometric parameters, *i.e.* $d_{H\cdots O}$= 2.929 Å and

∠N-H···O= 108.3º, this structure is stabilized by a very weak intramolecular interaction that defines a seven-membered hydrogen bonded ring between the N-H of NHMe and the C=O of the Ac. Indeed, a standard t-γ_L conformation with a strong intramolecular hydrogen bond forming a seven-membered ring is not possible for the Ac-β*c*Dmp-NHMe dipeptide. This is because such combination of φ and ψ dihedral angles leads to a strong repulsive interaction between the lone pair of the dimethylamine group and the carbonyl oxygen of the Dmp residue. The ΔE^{gp} of the other four conformations with ω_0 arranged in *trans* ranges from 1.7 (t-α_L[d]) to 5.6 (t-α_L[u]) kcal/mol, these energy values being significantly lower than those found for Ac-β*c*Amp-NHMe, *i.e.* in the latter dipeptide the ΔE^{gp} of the first (t-α_L[d]) and the last (t-α_L[u]) local minimum were 4.5 and 9.1 kcal/mol, respectively. On the other hand, the c-ε_L[d] is the lowest energy conformation with ω_0 arranged in *cis*, this structure being destabilized with respect to the global minimum by 1.6 kcal/mol. Figure 3.3.5e reveals that this conformation is stabilized by an intramolecular hydrogen bond between the N-H of the NHMe blocking group and the nitrogen of the dimethylamino substituent. Comparison with the results listed in Table 3.3.1 for Ac-β*c*Amp-NHMe indicates that the replacement of the amino substituent by the dimethylamino group also alters the potential energy hypersurface of the dipeptide with the peptide bond arranged in *cis*. Thus, the least favored *cis* minimum of Ac-β*c*Dmp-NHMe (c-ε_L[u]) is destabilized by 5.0 kcal/mol with respect to that of c-ε_L[d], whereas in Ac-β*c*Amp-NHMe the most (c-α_L[d]) and the least (c-α_L[u]) stable conformations with ω_0 arranged in *cis* are separated by only 1.3 kcal/mol.

Six minimum energy conformations, three with ω_0 arranged in *trans*, were found for Ac-γ*t*Dmp-NHMe. The lowest energy one corresponds to the t-γ_L[d] (Figure 3.3.5f), the t-γ_L[u], which was the global minimum of Ac-γ*t*Amp-NHMe, being unfavored by 3.1 kcal/mol. Interestingly, the lowest energy conformation and the t-α_L[u], which is destabilized by 6.7 kcal/mol, were not found as energy minima in Ac-γ*t*Amp-NHMe. Regarding the three conformations with ω_0 in *cis*, the most stable, c-α_L[u] (Figure 3.3.5g), is stabilized by a five-membered intramolecular hydrogen bonded ring involving the backbone nitrogen atom of the γ*t*Dmp residue and the N-H of the NHMe blocking group. This structure is

unfavored by 3.4 kcal/mol with respect to the global minimum, while the ΔE^{gp} values of the other two *cis* conformers are 6.5 (c-ε_L[d]) to 8.0 kcal/mol (c-ε_L[u]).

Eight minimum energy conformations were characterized for Ac-γ*c*Dmp-NHMe. The lowest energy one corresponds to the t-γ_L[u] (Figure 3.3.5h), while the other structures with ω_0 in *trans*, t-γ_L[d] and t-α_L[u], are unfavored by 2.0 and 3.0 kcal/mol, respectively. These three conformations are stabilized by an intramolecular hydrogen bond. Thus, a seven-membered hydrogen bonded ring involving the NHMe and Ac blocking groups is shown by the two t-γ_L structures, whereas in the t-α_L[u] conformation the nitrogen of the γ*c*Dmp residue and the N-H of the NHMe group forms a five-membered intramolecular hydrogen bonded ring. Comparison with the minima listed in Table 3.3.1 for Ac-γ*c*Amp-NHMe indicates that the incorporation of the methyl groups into the amino substituent mainly affects to the puckering of the pyrrolidine ring, *i.e.* the relative stability between t-γ_L[u] and t-γ_L[d] is exchanged and the t-α_L[u] minimum transform into the t-α_L[d].

On the other hand, ω_0 is arranged in *cis* in the remaining five minima of Ac-γ*c*Dmp-NHMe, the most stable *cis* structure being the c-α_L[u] (Figure 3.3.5i). This conformation is 2.7 kcal/mol less stable than the global minimum, and is stabilized by an intramolecular hydrogen bond similar to that described for the t-α_L[u] minimum. Interestingly, the Ac-γ*c*Dmp-NHMe dipeptide shows two minima with c-α_L[d] conformation, which differ in the nitrogen atom that acts as acceptor in the stabilizing intramolecular hydrogen bond. In the most favored conformation that is 3.2 kcal/mol less stable than the global minimum, the nitrogen of the γ*c*Dmp residue participates in such interaction, whereas the interaction in the second conformation which is 0.4 kcal/mol less favored than the first one, involves the nitrogen of the dimethylamino side group. Finally, the ΔE^{gp} values of the c-ε_L[u] and c-ε_L[d] structures are 5.5 and 8.9 kcal/mol, respectively.

The ΔG^{gp} values listed in Table 3.3.5 show that the effects produced by the incorporation of the ZPVE and the thermal and entropic corrections are less dramatic for Dmp-containing dipeptides than for the Amp ones. Thus, the addition of these statistical terms represents relative variations typically smaller than 1 kcal/mol. Specifically, the largest change found in Ac-β*t*Dmp-NHMe, Ac-

β*c*Dmp-NHMe, Ac-γ*t*Dmp-NHMe and Ac-γ*c*Dmp-NHMe, is -0.9 (c-ε_L[u]), -1.1 (c-α_L[d]), -1.3 (c-ε_L[u]) and -1.4 kcal/mol (c-ε_L[u]), respectively. Inspection of the relative free energies in carbon tetrachloride, also displayed in Table 3.3.5, indicates that in general, solute-solvent interactions tend to stabilize the minimum energy conformations with ω_0 in *cis*. In spite of this, the lowest energy conformation in carbon tetrachloride solution and in the gas-phase is the same for the four Dmp-containing dipeptides. This is an important difference with respect to the four Amp-containing dipeptides since, as we previously showed, this organic solvent is able to alter the conformational preferences of Ac-β*c*Amp-NHMe and Ac-γ*c*Amp-NHMe (Table 3.3.2).

Table 3.3.5 *Relative free energy in the gas-phase (ΔG^{gp}; in kcal/mol) and in carbon tetrachloride, chloroform. methanol and aqueous solutions (ΔG^{CCl4}, ΔG^{CHCl3}, ΔG^{CH3OH} and ΔG^{H2O}, respectively; in kcal/mol) for the minimum energy conformations of Ac-β*t*Dmp-NHMe, Ac-β*c*Dmp-NHMe, Ac-γ*t*Dmp-NHMe and Ac-γ*c*Dmp-NHMe at the B3LYP/6-31+G(d,p) level.*

# Conf.	ΔG^{gp}	ΔG^{CCl4}	ΔG^{CHCl3}	ΔG^{CH3OH}	ΔG^{H2O}
		Ac-β*t*Dmp-NHMe			
t-γ_L[d]	0.0[a]	0.0	0.4	3.0	4.0
t-γ_L[u]	0.4	0.6	0.8	3.0	4.0
c-ε_L[u]	2.6	1.0	0.0	0.7	1.4
c-α_L[d]	2.6	1.8	1.4	0.7	3.0
c-α_L[u]	2.7	1.7	1.6	3.1	3.5
c-ε_L[d]	5.3	3.4	4.3	0.5	0.6
c-ε_L[u]	5.4	3.6	1.0	0.0	0.0
		Ac-β*c*Dmp-NHMe			
t-γ_L[d]	0.0[b]	0.0	0.0	0.4	1.6
c-ε_L[d]	0.8	0.2	0.2	0.7	1.9
t-α_L[d]	2.6	2.4	2.0	1.5	2.2
t-ε_L[d]	2.4	1.7	1.2	0.8	2.4
c-α_L[d]	0.9	1.1	1.5	2.1	2.9
t-ε_L[u]	4.8	4.0	3.4	2.7	3.8
c-ε_L[d]	2.9	2.0	1. 2	0.0	0.4

t-α_L[u]	5.9	5.0	4.6	4.2	0.0
c-α_L[u]	5.1	4.0	4.3	5.0	5.9
c-ε_L[u]	5.5	4.0	2.8	1.4	2.1
			Ac-γ*t*Dmp-NHMe		
t-γ_L[d]	0.0[c]	0.0	0.6	4.8	5.5
t-γ_L[u]	2.9	3.1	3.8	8.2	8.9
c-α_L[u]	2.3	1.7	1.4	4.1	3.9
c-ε_L[d]	4.5	2.1	0.0	0.0	0.0
t-α_L[u]	5.6	5.2	4.5	6.9	6.7
c-ε_L[u]	6.7	4.6	2.8	3.1	3.1
			Ac-γ*c*Dmp-NHMe		
t-γ_L[u]	0.0[d]	0.0	1.3	5.6	6.3
t-γ_L[d]	2. 6	2.8	4.5	8.9	3.7
c-α_L[u]	1.8	0.9	1.3	4.6	4.8
c-α_L[d]	2.5	0.8	1.2	5.1	5.5
t-α_L[u]	2.4	1.5	1.6	4.2	4.6
c-α_L[d]	3.5	2.7	3.5	7.1	7.4
c-ε_L[u]	4.1	1.5	0.0	0.0	0.0
c-ε_L[d]	7.7	6.8	9.1	7.6	0.5

[a] G= -707.030159 a.u. [b] G= -707.026345 a.u. [c] G= -707.033869 a.u. [d] G= -707.031747 a.u.

The stabilization of conformers with the peptide bond in *cis* is enhanced in chloroform solution (Table 3.3.5). Thus, the lowest energy minimum in this solvent for Ac-β*t*Dmp-NHMe, Ac-γ*t*Dmp-NHMe and Ac-γ*c*Dmp-NHMe corresponds to the c-ε_L[u] (Figure 3.3.5c), c-ε_L[d] (Figure 3.3.6a) and c-ε_L[u] (Figure 3.3.6b), respectively. It is worth noting that the tendency to adopt the c-ε_L conformer in chloroform solution was also detected in Ac-β*t*Amp-NHMe and Ac-γ*t*Amp-NHMe (Table 3.3.2), while Ac-γ*c*Amp-NHMe preferred a c-α_L helical structure. In contrast, the t-γ_L[d] (Figure 3.3.5d) is the lowest energy conformation of Ac-β*c*Dmp-NHMe, which was also the global minimum in both the gas-phase and carbon tetrachloride solution. Amazingly this exceptional behavior is identical to that observed in the previous section for Ac-β*c*Amp-NHMe.

The most stable conformation in both methanol and aqueous solutions of Ac-β*t*Dmp-NHMe, Ac-γ*t*Dmp-NHMe and Ac-γ*c*Dmp-NHMe also corresponds to the c-ε_L conformer (Figures 3.3.6c, 3.3.6a and 3.3.6b, respectively), even though the puckering of the pyrrolidine ring depends on the position of the substituent. Again, this result is fully consistent with that obtained for the Amp-containing dipeptides. Regarding the Ac-β*c*Dmp-NHMe dipeptide, the lowest energy conformations in methanol and aqueous solutions are the c-ε_L[d] (Figure 3.3.6d) and t-α_L[u] (Figure 3.3.6e), respectively. However, it should be noted that in the latter environment the former conformation is destabilized by only 0.4 kcal/mol. This indicates that in the presence of polar solvents the four Dmp-containing dipeptides follow a similar behavior.

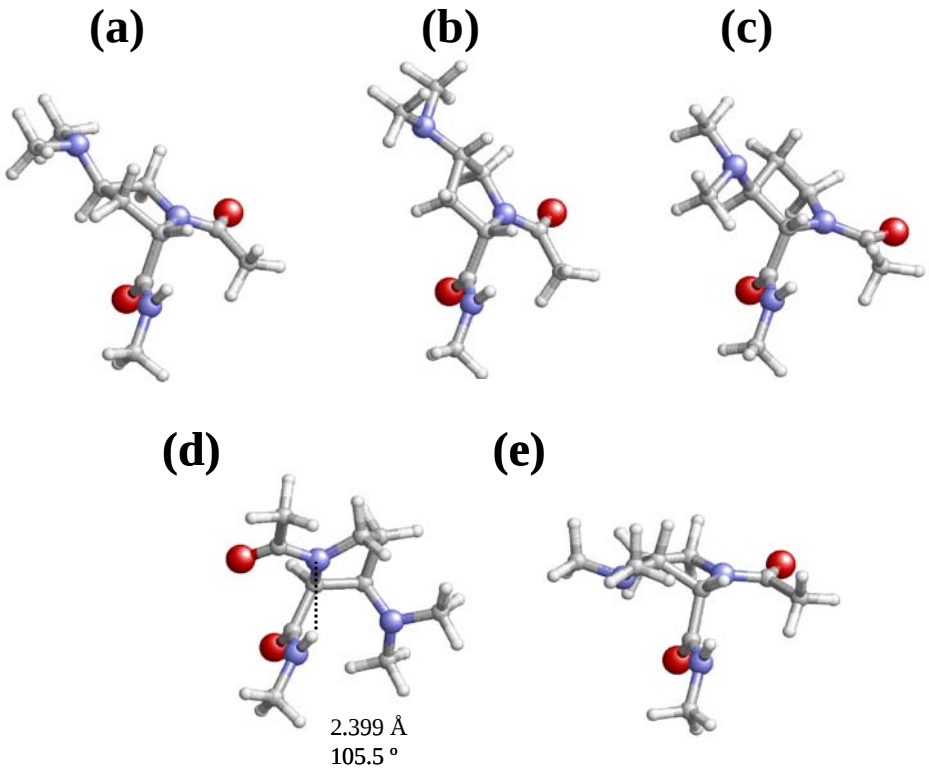


Figure 3.3.6: *Representation of selected minimum energy conformations characterized for the Dmp-containing dipeptides studied in this work: (a) c-ε_L[d] for Ac-γtDmp-NHMe; (b) c-ε_L[u] for Ac-γcDmp-NHMe; (c) c-ε_L[u] for Ac-βtDmp-NHMe; (d) c-ε_L[d] and (e) t-α_L[u] for Ac-βcDmp-NHMe. These minima are especially relevant because they are relatively stable in chloroform, methanol and/or aqueous solution.*

The free energies relative to the lowest energy minimum of the most stable Dmp-containing isomer are listed in Table 3.3.6. The most stable isomer in gas-

phase is Ac-γ*t*Dmp-NHMe, even though the Ac-γ*c*Dmp-NHMe is destabilized by only 1.0 kcal/mol. However, the latter is the most favored isomer in solution, independent of the polarity of the environment. Moreover, the relative order in terms of stability is: Ac-γ*c*Dmp-NHMe > Ac-γ*t*Dmp-NHMe > Ac-β*c*Dmp-NHMe > Ac-β*t*Dmp-NHMe. A detailed comparison of these results with those reported in the previous section (see Table 3.3.3) for the Amp-containing dipeptides reveals some important differences. Thus, the relative order found for the aminated dipeptides in aqueous and chloroform solution was Ac-γ*c*Amp-NHMe > Ac-β*t*Amp-NHMe > Ac-γ*t*Amp-NHMe > Ac-β*c*Amp-NHMe, while in methanol and carbon tetrachloride solutions the relative orders were Ac-γ*c*Amp-NHMe ≅ Ac-β*t*Amp-NHMe > Ac-γ*t*Amp-NHMe > Ac-β*c*Amp-NHMe and Ac-γ*t*Amp-NHMe > Ac-β*c*Amp-NHMe > Ac-γ*c*Amp-NHMe > Ac-β*t*Amp-NHMe, respectively.

Table 3.3.6 *Relative free energy in the gas-phase ($\Delta G^{\#gp\#}$; in kcal/mol) and in carbon tetrachloride, chloroform, methanol and aqueous solutions ($\Delta G^{\#CCl4\#}$, $\Delta G^{\#CHCl3\#}$, $\Delta G^{\#CH3OH\#}$ and $\Delta G^{\#H2O\#}$, respectively; in kcal/mol) calculated at the B3LYP/6-31+G(d,p) level for the minimum energy conformations of Ac-βtDmp-NHMe, Ac-βcDmp-NHMe, Ac-γtDmp-NHMe and Ac-γcAmp-NHMe with respect to the lowest energy minimum of the most stable isomer.*

# Conf.	$\Delta G^{\#gp\#}$	$\Delta G^{\#CCl4\#}$	$\Delta G^{\#CHCl3\#}$	$\Delta G^{\#CH3OH\#}$	$\Delta G^{\#H2O\#}$
			Ac-β*t*Dmp-NHMe		
t-γ_L[d]	2.3	3.8	6.9	11.8	12.7
t-γ_L[u]	2.7	3.9	6.9	11.5	12.4
c-ε_L[u]	4.9	2.1	3.9	6.9	7.5
c-α_L[d]	4.9	2.9	5.3	6.5	9.2
c-α_L[u]	5.0	2.7	5.4	9.2	9.5
c-ε_L[d]	7.6	1.8	5.5	4.0	4.1
c-ε_L[u]	7.7	1.9	2.1	3.5	3.4
			Ac-β*c*Dmp-NHMe		
t-γ_L[d]	4.7	2.2	4.0	7.0	7.5
c-ε_L[d]	5.5	1.6	3.4	6.6	7.0
t-α_L[d]	7.3	1.9	3.4	5.5	5.9

t-ε_L[d]	7.1	1.5	2.8	5.0	5.8
c-α_L[d]	5.6	2.4	4.7	7.9	7.9
t-ε_L[u]	9.5	1.3	2.6	4.5	4.8
c-ε_L[d]	7.6	1.3	2.4	3.7	3.4
t-α_L[u]	10.6	1.2	2.7	4.9	4.7
c-α_L[u]	9.8	1.1	3.3	6.5	6.6
c-ε_L[u]	10.2	0.6	1.3	2.4	2.4
		Ac-γ*t*Dmp-NHMe			
t-γ_L[d]	0.0[a]	2.8	2.6	10.5	5.1
t-γ_L[u]	2.9	3.0	2.4	10.9	4.6
c-α_L[u]	2.3	2.1	4.2	7.3	7.4
c-ε_L[d]	4.5	0.4	0.7	1.1	1.4
t-α_L[u]	5.6	2.4	4.1	6.9	7.0
c-ε_L[u]	6.7	0.7	1.2	2.1	2.3
		Ac-γ*c*Dmp-NHMe			
t-γ_L[u]	1.0	2.7	5.4	9.7	10.5
t-γ_L[d]	3. 6	2.9	6.1	10.5	5.3
c-α_L[u]	2.8	1.8	3.7	7.0	7.2
c-α_L[d]	3.5	0.9	2.9	6.6	7.1
t-α_L[u]	3.4	1.7	3.3	5.9	6.3
c-α_L[d]	4.5	1.9	4.2	7.8	8.1
c-ε_L[u]	5.1	0.0	0.0	0.0	0.0
c-ε_L[d]	8.7	1.1	2.1	3.3	3.1

[a] G= -707.033869 a.u.

3.3.4 Conclusions

Quantum mechanical calculations at the B3LYP/6-31+G(d,p) level have been used to explore the conformational preferences of Ac-β*t*Amp-NHMe, Ac-β*c*Amp-NHMe, Ac-γ*t*Amp-NHMe, Ac-γ*c*Amp-NHMe, Ac-β*t*Dmp-NHMe, Ac-β*c*DDmp-NHMe, Ac-γ*t*Dmp-NHMe and Ac-γ*c*Dmp-NHMe. Comparison of the results with

those obtained for Ac-L-Pro-NHMe at the same theoretical level allows us to draw the following conclusions:

(i) Incorporation of the amino group into the β- or γ-position of the pyrrolidine ring reduces the intrinsically low conformational flexibility of conventional Pro. Specifically, the four Amp-containing dipeptides investigated in this work only show one energetically accessible minimum in the gas-phase, *i.e.* $\Delta E^{gp} < 1.5$ kcal/mol, when ω_0 is arranged in *trans*. On the other hand, the stability of conformations with ω_0 in *cis* is, in general, higher for Amp than for Pro. This is because some such conformations are stabilized by intramolecular hydrogen bonds in which the nitrogen of the amino substituent acts as acceptor.

(ii) In general the solvent enhances the stability of the conformers with ω_0 in *cis*. Thus, the c-ε_L was the most stable conformation for Ac-β*t*Amp-NHMe, Ac-β*c*Amp-NHMe and Ac-γ*t*Amp-NHMe in chloroform, methanol and aqueous solution, whereas for Ac-γ*t*Amp-NHMe the most stable conformer in solution was the c-α_L (chloroform) or the c-ε_L (methanol and water). However, PCM results must be analyzed with caution since this SCRF method overestimates the stability of the *cis* structures significantly, especially when protic solvents (as water or methanol) are considered.

(iii) Characterization of the minimum energy conformations of the four Dmp-containing peptides indicates that substitution of the amino by the dimethylamino reduces considerably the conformational restriction found in the *N*-acetyl-*N'*-methyl derivatives of Amp. This is indicated by both the increase in the number of minimum energy conformations and the decrease of relative conformational energies.

(iv) The stabilization of the structures with ω_0 in *cis* in the Dmp-containing dipeptides is similar to that found for the Amp derivatives. Thus, this type of structures is the most favored in chloroform, methanol and aqueous solutions. However, the conformational preferences of the Dmp-containing dipeptides in the gas-phase and in carbon tetrachloride solution are identical.

(v) A detailed energetic comparison of the dipeptides studied in this work indicates that the most favored Amp isomer in the gas-phase, chloroform, methanol and water solutions is the Ac-γ*c*Amp-NHMe dipeptide. In contrast, Ac-γ*t*Dmp-NHMe is the most stable Dmp-containing dipeptide in the gas-phase, even although the Ac-γ*c*Dmp-NHMe is most favored isomer in solution, independently of the polarity of the environment.

3.3.5 References

1. (a) Chakrabarti, P.; Pal, D. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1. (b) Marraud, M.; Aubry, A. *Biopolymers* **1996**, *40*, 45. (c) MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397. (d) Rose, G. D.; Gierasch, L. M.; Smith, J. A. *Adv. Protein Chem.* **1985**, *37*, 1.
2. (a) Boussard, G.; Marraud, M.; Aubry, A. *Biopolymers* **1979**, *18*, 1297. (b) V. Madison, K. D. Kopple, *J. Am. Chem. Soc.* **1980**, *102*, 4855. (c) Liang, G. B.; Rito, C. J.; Gellman, S. H. *Biopolymers* **1992**, *32*, 293. (d) Beausoleil, E.; Lubell, W. D. *J. Am. Chem. Soc.* **1996**, *118*, 12902.
3. (a) Kang, Y. K. *J. Phys. Chem. B* **2006**, *110*, 21338. (b) Kang, Y. K.; Byun, B. J. *J. Phys. Chem. B* **2007**, *111*, 5377. (c) Rivail, J. L.; Bouchy, A.; Loos, P. F.*J. Arg. Chem. Soc.* **2006**, *94*, 19.
4. (a) Dugave, C.; Demange, L. *Chem. Rev.* **2003**, *103*, 2475. (b) Pal, D.; Chakrabarti, P. *J. Mol. Biol.* **1999**, *294*, 271. (c) Stewart, D. E.; Sarkar, A.; Wampler, J. E. *J. Mol. Biol.* **1990**, *214*, 253. (d) Grathwohl, C.; Wüthrich, K. *Biopolymers* **1981**, *20*, 2623.
5. Mac Arthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397.
6. (a) Sapse, A.-M.; Mallah-Levy, L.; Daniels, S. D.; Erickson, B. W. *J. Am. Chem. Soc.* **1987**, *109*, 3526. (b) Ramachandran, G. N. *Int. J. Peptide Prot. Res.* **1988**, *31*, 1.
7. Andreotti, A. H. *Biochemistry* **2003**, *42*, 9515.
8. (a) Wedemeyer, W. J.; Welker, E.; Scheraga, H. A. *Biochemistry* **2002**, *41*, 14637. (b) Dugave, C.; Demange, L. *Chem. Rev.* **2003**, *103*, 2475.

9. (a) Zimmerman, S. S.; Pottle, M. S.; Nemethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1. (b) Fischer, S.; Dunbrack, R. L., Jr.; Karplus, M. *J. Am. Chem. Soc.* **1994**, *116*, 11931. (c) Kang, Y. K. *J. Phys. Chem.* **1996**, *100*, 11589. (d) Improta, R.; Benzi, C.; Barone, V. *J. Am. Chem. Soc.* **2001**, *123*, 12568. (e) Benzi, C.; Improta, R.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2002**, *23*, 341. (f) Kang, Y. K. *J. Phys. Chem. B* **2002**, *106*, 2074. (g) Hudáky, I.; Baldoni, H. A.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2002**, *582*, 233. (h) Hudáky, I.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2003**, *630*, 135. (i) Allen, W.D.; Czinki, E.; Császár, A. G. *Chem. Eur. J.* **2004**, *10*, 4512. (j) Kang, Y. K.; Park, H. S. *J. Mol. Struct. (Theochem)* **2005**, *718*, 17. (k) Czinki, E.; Csaszar, A. G. *Chem. Eur. J.* **2003**, *9*, 1008.
10. Sahai, M.A.; Kehoe, A. K.; Koo, J. C. P.; Setiadi, D. H.; Chass, G. A.; Viskolcz, B.; Penke, B.; Pai, E. F.; Csizmadia, I. G. *J. Phys. Chem. A* **2005**, *109*, 2660.
11. (a) Kang, Y. K.; Park, H. S. *Biophys. Chem.* **2005**, *113*, 93. (b) Kang. Y. K., Jhon, J. S.; Park, H. S. *J. Phys. Chem. B* **2006**, *110*, 17645. (c) Kang, Y. K. *J. Mol. Struct.* **2004**, *675*, 37. (e) Kang, Y. K.; Choi, H. Y. *Biophys. Chem.* **2004**, *111*, 135.
12. (a) Flores-Ortega, A.; Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2007**, *111*, 5475. (b) Flores-Ortega, A.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C.; Casanovas, J. *J. Org. Chem.* **2008**, *73*, 3418.
13. (a) Benzi, C.; Improta, R.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2002**, *23*, 341. (b) Lam, J. S. W.; Koo, J. C. P.; Hudáky, I.; Varro, A.; Papp, J. G.; Penke, B.; Csizmadia, I. G. *J. Mol. Struct. (Theochem)* **2003**, *666-667*, 285-289.
14. (a) Alemán, C.; Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Haspel, N.; Zheng, J.; Casanovas, J.; Wolfson, H.; Nussinov, R. *Phys. Biol.* **2006**, *3*, S54. (b) Tsai, C.-J.; Zheng, J.; Zanuy, D.; Haspel, N.; Wolfson, H.; Alemán, C.; Nussinov, R. *Proteins* **2007**, *68*, 1.
15. (a) Haspel, N.; Zanuy, D.; Alemán, C.; Wolfson, H.; Nussinov, R. *Structure* **2006**, *14*, 1137. (b) Zheng, J.; Zanuy, D; Haspel, N.; Tsai, C.-J.; Alemán, C.; Nussinov, R. *Biochemistry* **2007**, *46*, 1205. (c) Zanuy, D.;

Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2007**, *111*, 3236.

16. Zanuy, D.; Flores-Ortega, A.; Casanovas, J.; Curcó, D.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2008**, *112*, 8692.
17. (a) Porter, E. A.; Wang, X.; Schmitt, M. A.; Gellman, S. H. *Org. Lett.* **2002**, *4*, 3317. (b) Farrera-Sinfreu,J.; Zaccaro, L.; Vidal, D.; Salvatella, X.; Giralt, E.; Pons, M.; Albericio, F; Royo, M. *J. Am. Chem. Soc.* **2004**, *126*, 6048.
18. Gaussian 03, Revision B.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr., T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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.; C. Strain, M.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Pittsburgh PA, 2003.
19. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
20. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1993**, *37*, 785.
21. McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639.
22. (a) Alemán, C.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Casanovas, J. *J. Phys. Chem. B* **2002**, *106*, 11849. (b) Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *Chem. Phys. Lett.* **2006**, *429*, 528. (c) Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Alemán, C. *J. Phys. Chem. B* **2006**, *110*, 5762.

23. (a) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999. (b) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027. (c) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239. (d) Miertus, M.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
24. Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, *113*, 6256.
25. (a) Hudáky, I.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2003**, *630*, 135. (b) Hudáky, I.; Baldoni, H. A.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2002**, *582*, 233.
26. (a) Iribarren, J. I.; Casanovas, J.; Zanuy, D.; Alemán, C. *Chem. Phys.* **2004**, *302*, 77. (b) Jang, Y. H.; Goddard III, W. A.; Noyes, K. T.; Sowers, L. C.; Hwang, S.; Chung, D. S. *J. Phys. Chem. B* **2003**, *107*, 344. (c) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys. B* **1998**, *102*, 3257. (d) Orozco, M.; Luque, F. J. *J. Am. Chem. Soc.* **1995**, *117*, 1378. (e) Alemán, C.; Navarro, E.; Puiggalí, J. *J. Org. Chem.* **1995**, *60*, 6135.
27. (a) Assfeld, X.; Ruiz-Lopez, M. F.; González, J.; Lopez, R.; Sordo, J. A.; Tomas, T. L. *J. Comp. Chem.* **1994**, *15*, 479. (b) Assfeld, X.; Ruiz-Lopez, M. F.; García, J. I.; Mayoral, J. A.; Salvatella, L. *J. Chem. Soc. Chem. Comm.* **1995**, *13*, 1371.
28. Delaney, N. G.; Madison, V. *J. Am. Chem. Soc.* **1982**, *104*, 6635.

3.4 Protonation of the side group in β–and γ–Aminated Proline Analogues: Effects on the Conformational Preferences

Density Functional Theory calculations have been performed on the N-*acetyl*-N*'-methylamide derivatives of the four possible isomers of aminoproline protonated at the amino side group. Comparison of the results obtained for these isomers, which differ in the β- or γ-position of the substituent and its* cis *or* trans *relative disposition, with those reported for the corresponding neutral analogues [*J. Phys. Chem. B ***2008***, 112, *14045] has allowed to reach the following conclusions: (i) protonation of the amino group produces a reduction of the backbone conformational flexibility and a destabilization of the* cis *configuration of the amide bond involving the pyrrolidine nitrogen; (ii) the planarity of the peptide bond is broken in some cases to form strong side chain···backbone interactions, which induce a very significant pyramidilization at the amide nitrogen atom; (iii) as was also detected for the neutral analogues, the formation of side chain···backbone intraresidue interactions favour the* cis *disposition of the substituent; and (iv) protonation of the amino side group increases the energy gaps that separate the different investigated isomers, which results in an enhancement of the destabilization of the dipeptides with the substituent attached in* trans[*].

3.4.1 Introduction

The design and application of synthetic amino acids with restricted conformational mobility in different fields of nanobiology, *e.g.* the re-engineering of physical protein modules and the generation of nanodevices,[1,2] is a topic of growing interest. Within this context, we recently observed that the insertion of chemically constrained residues with suitable backbone conformational tendencies enhance the thermodynamic stability of the nanotubular structures constructed by self-assembling protein fragments.[3]

Among the large variety of amino acids that can be designed, those achieved by introducing chemical modifications to Proline (Pro) are particularly attractive. This is because the side chain of Pro is bonded to both the α-carbon and its

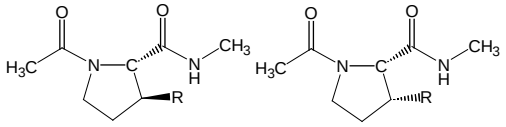
[*] *Submitted for Publication.*

preceding amide nitrogen providing conformational properties that are unique among naturally occurring amino acids. As a consequence, rotation about the N—C^{α} bond is prohibited and the φ torsion angle is confined to values of around –60º. Accordingly, Pro is overwhelmingly found in the α-helical ($\varphi,\psi\approx$ –60º,–30º) and semi-extended ($\varphi,\psi\approx$ –60º,140º) regions of the conformational map.[4] In addition, Pro shows a higher propensity to promote γ-turn conformations ($\varphi,\psi\approx$ –70º,60º) than other proteogenic amino acids.[4d,5] Another effect derived from its cyclic structure is that the peptide bond preceding Pro (that involving the pyrrolidine nitrogen) has a relatively high probability of accommodating a *cis* arrangement[6] as compared to other peptide bonds, for which the *cis* form is almost inexistent.

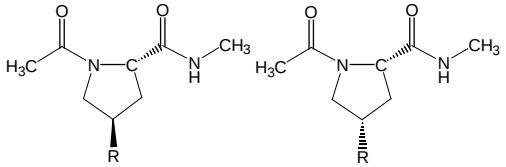
The conformational properties of a relative wide number of synthetic Pro derivatives have been reported. These compounds were obtained by incorporating a substituent at the C^{α} atom (α-substituted Pro analogues)[7,8] or in the pyrrolidine ring (*e.g.* hydroxylated, fluorinated and aminated Pro analogues),[9-11] or altering the chemical nature of the own pyrrolidine ring (*e.g.* diminishing or enlarging the ring size,[12] replacing a carbon atom by an heteroatom,[13] and incorporating a double bond through a deshydrogenation[14]). Within this context, we recently reported the intrinsic conformational properties of different aminoproline (Amp) derivatives,[11] which have been already used to construct β- and γ-peptides with helical secondary structures.[15] Specifically, we investigated the *N*-acetyl-*N*'-methylamide derivatives of both the *cis* and *trans* Amp isomers that incorporate an amino group to the C^{β}- or C^{γ}-positions of the pyrrolidine ring. Theoretical calculations based on Density Functional Theory (DFT) methods on these four compounds, which were denoted Ac-β*t*Amp-NHMe, Ac-β*c*Amp-NHMe, Ac-γ*t*Amp-NHMe and Ac-γ*c*Amp-NHMe (Figure 3.4.1), evidenced that the incorporation of the amino group reduces the intrinsically low conformational flexibility of conventional Pro. Furthermore, the stability of conformations with the peptide bond involving the pyrrolidine nitrogen arranged in *cis* was, in general, higher for Amp derivatives than for Pro. This was attributed to the formation in such conformations of stable intramolecular hydrogen bonds with the nitrogen of the amino substituent acting as acceptor.

In Although these results suggested that Amp derivatives have potential interest for many nanobiological applications, the conformational properties of

these amino acids may be easily altered by transforming the amino group, a reasonably weak base, into the positively charged ammonium group, *i.e.* amines react readily with acids. This proteolitic equilibrium is especially important in aqueous solution, in which the pH can be used to control the conformation of the Amp derivatives. However, the conformational preferences of the protonated Amp derivatives, hereafter denoted Ac-β*t*AmH$^+$p-NHMe, Ac-β*c*AmH$^+$p-NHMe, Ac-γ*t*AmH$^+$p-NHMe and Ac-γ*c*AmH$^+$p-NHMe (see Figure 3.4.1) remain totally unknown yet. In this work we have used DFT calculations to explore systematically the potential energy hypersurfaces of these four dipeptides, the influence of the solvent being been evaluated through a Self Consistent Reaction Field (SCRF) method. Furthermore, we have also examined the influence of the proteolitic equilibrium on both the *trans*/*cis* disposition of the peptide bond involving the pyrrolidine nitrogen and the relative stability of the four isomers generated by the substitution at different positions.



R= NH_2	Ac-β*t*Amp-NHMe	Ac-β*c*Amp-NHMe
R= NH_3^+	Ac-β*t*AmH$^+$p-NHMe	Ac-β*c*AmH$^+$p-NHMe



R= NH_2	Ac-γ*t*Amp-NHMe	Ac-γ*c*Amp-NHMe
R= NH_3^+	Ac-γ*t*AmH$^+$p-NHMe	Ac-γ*c*AmH$^+$p-NHMe

Figure 3.4.1: *Compounds studied in this work*

3.4.2 Methods

All calculations were carried out using the Gaussian 03 computer program.[16] DFT calculations were performed using the B3LYP method[17] combined with the 6-31+G(d,p) basis set,[18] which was previously employed to study the *N*-acetyl-*N'*-methylamide derivatives of conventional Pro[7] and Amp.[11]

Torsion angles for the backbone and side chain of the AmH^+p derivatives studied in this work are defined in Figure 3.4.2. Since each flexible backbone dihedral angle is expected to have three minima, *i.e. gauche*$^+$ (60º), *trans* (180º) and *gauche*$^-$ (–60º), and φ is fixed by the geometry of the five-membered ring, the number of minima that may be anticipated for the potential energy surface E= E(ψ) of each AmH^+p dipeptide is 3. Additionally, each amide bond (given by the torsional angles ω_0 and ω) can be arranged in *trans* or *cis*, even though only the peptide bond involving the pyrrolidine nitrogen (ω_0) is likely to adopt a *cis* configuration. Therefore, both the *trans* and *cis* states were considered for ω_0, while the amide bond involving the *N*-methylamide blocking group (ω) was arranged in *trans* only. Furthermore, due to the cyclic nature of the side chain, two puckering states (denoted *down* and *up*) are expected for each backbone minimum energy conformation. The *down* and *up* arrangements are defined as those in which one atom of the pyrrolidine ring and the carbonyl group of the AmH^+p residue lie on the same and opposite sides, respectively, of the plane defined by the remaining four atoms of the pyrrolidine ring.

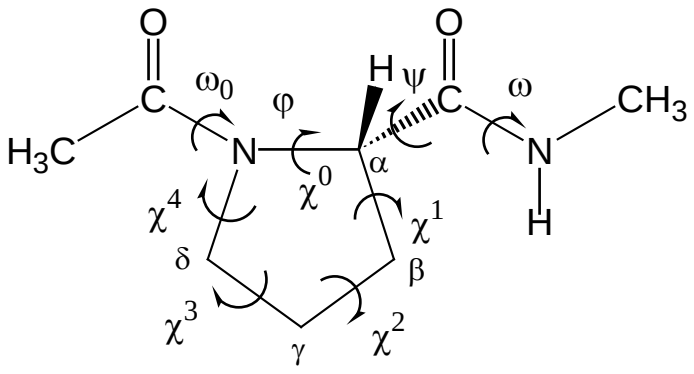


Figure 3.4.2: *Dihedral angles used to identify the conformations of the* N-*acetyl*-N'*-methylamide derivatives of the AmH^+p analogues studied in this work. The dihedral angles* ω_0, φ, ψ *and* ω *are defined using backbone atoms while the endocyclic dihedral angles* χ^i *are given by the atoms of the five-membered ring. In particular, the sequence of atoms used to define* φ *and* χ^0 *are C(=O)–N–*C^α*–C(=O) and* C^δ*–N–*C^α*–*C^β*, respectively.*

Accordingly, for each of the four dipeptides under study (Figure 3.4.1), 3(ψ backbone) × 2(ω_0 *trans*-or-*cis*) × 2(cyclic side chain) = 12 structures were considered as starting points for complete geometry optimizations at the B3LYP/6-31+G(d,p) level.

Frequency analyses were carried out to verify the nature of the minimum state of all the stationary points obtained and to calculate the zero-point vibrational energies (ZPVE) as well as both thermal and entropic corrections. These statistical terms were used to evaluate the conformational Gibbs free energies in the gas phase (ΔG^{gp}) at 298 K.

To examine the solvation effects on the molecular geometry and conformational stability, single point calculations were conducted on the B3LYP/6-31+G(d,p) optimized structures using a self-consistent reaction-field (SCRF) model. SCRF methods treat the solute at the quantum mechanical level, and the solvent is represented as a dielectric continuum. Specifically, we chose the polarizable continuum model (PCM) developed by Tomasi and co-workers to describe the solvent.[19] The PCM represents the polarization of the liquid by a charge density appearing on the surface of the cavity created in the solvent, *i.e.* the solute/solvent interface. This cavity is built using a molecular shape algorithm.

PCM calculations were performed in the framework of the B3LYP/6-31+G(d,p) level using the standard protocol, and considering the dielectric constant of water (ε= 78.1). The conformational free energies in solution (ΔG^{WAT}) were estimated by adding the free energies of solvation to the ΔG^{gp} values.

The minimum energy conformations of the four dipeptides studied in this work have been denoted using the same three-label code that was used for the Amp derivatives,[11] which specifies the arrangement of the ω_0 peptide bond, the (φ,ψ) backbone conformation and the puckering of the five-membered ring. The first letter refers to the *trans* (t) or *cis* (c) arrangement of the peptide bond involving the pyrrolidine nitrogen. The second label identifies the backbone conformation using the nomenclature introduced by Perczel *et al.*[20] more than fifteen years ago. In the case of Pro and its derivatives, only the γ_L (γ-turn or C_7), α_L (α-helical), and ε_L (polyproline II-like) conformations are accessible to the backbone since φ is fixed in the neighborhood of –60º. Finally, the *up* or *down* puckering of the five-membered ring is indicated using the [u] and [d] labels, respectively. The puckering of the ring was described using the classical pseudorotational algorithm, which uses a very simple model based on the puckering amplitude and the state of the pucker in the pseudorotation pathway. This model was previously applied by Perczel *et al.*[21] to describe conventional Pro.

3.4.3 Results and Discussion

Calculations at the B3LYP/6-31+G(d,p) level led to 3, 5, 5 and 3 minimum energy conformations characterized for Ac-β*t*AmH$^+$p-NHMe, Ac-β*c*AmH$^+$p-NHMe, Ac-γ*t*AmH$^+$p-NHMe and Ac-γ*c*AmH$^+$p-NHMe, respectively. Table 3.4.1 lists the more relevant structural parameters together with the relative energy (ΔE^{gp}) of all these structures, which are displayed in Figures 3.4.3-3.4.6. Table 3.4.1 also shows the relative stability of the four dipeptides ($\Delta E^{\#gp\#}$), which corresponds to the energy relative to the lowest energy conformation of the most stable isomer. Table 3.4.2 compares the relative free energies in the gas-phase (ΔG^{gp}) and aqueous solution (ΔG^{H2O}) for the minima of the four dipeptides.

Furthermore, the relative free energies in these two environments calculated in each case with respect to the conformation of lowest free energy of the most stable isomer, $\Delta G^{\#gp\#}$ and $\Delta G^{\#H2O\#}$, have been also included in Table 3.4.2.

Table 3.4.1 *Backbone dihedral angles (in degrees), pseudorotational parameters (A and P; in degrees), relative energy (ΔE^{gp}; in kcal/mol) and relative energy with respect to the lowest energy conformation of the most stable dipeptide ($\Delta E^{\#gp\#}$; in kcal/mol) of the minimum energy conformations characterized for Ac-βtAmH$^+$p-NHMe, Ac-βcAmH$^+$p-NHMe, Ac-γtAmH$^+$p-NHMe and Ac-γcAmH$^+$p-NHMe at the B3LYP/6-31+G(d,p) level in the gas phase.*

# Conf.	ω_0	φ	ψ	ω	(A, P)	ΔE^{gp}	$\Delta E^{\#gp\#}$
				Ac-β*t*AmH$^+$p-NHMe			
t-γ_L[u]	-170.5	-66.2	33.5	176.0	(42.3, 73.0)[a]	0.0[b]	6.2
c-α_L[u]	15.5	-88.7	4.1	-178.5	(42.3, 83.5)[c]	6.8	13.0
c-α_L[d]	7.2	-97.1	12.1	178.0	(35.6, -131.9)[d]	12.7	18.9
				Ac-β*c*AmH$^+$p-NHMe			
t-γ_L[d]	-171.5	-83.7	65.3	-179.9	(34.0,-114.0)e	0.0[f]	2.8
t-γ_L[u]	-170.9	-79.2	68.6	179.1	(38.2, 90.0)g	1.6	4.4
c-α_L[d]	14.3	-109.9	22.0	176.8	(42.6,-128.7)h	7.2	10.0
c-α_L[u]	21.2	-113.1	20.4	175.2	(39.1, 133.1)i	10.3	13.1
c-ε_L[u]	13.0	-45.9	117.2	178.6	(43.2, 41.7)j	11.0	13.8
				Ac-γ*t*AmH$^+$p-NHMe			
g$^+$-δ_L[u]	63.0	-147.6	33.8	175.9	(43.8, 154.7)k	0.0[l]	13.1
g$^+$-δ_L[u]	60.0	-149.0	104.6	-177.6	(44.1,154.6)m	0.2	13.3
t-γ_L[d]	-177.5	-83.6	85.8	-176.8	(40.8,-118.3)n	2.0	15.1
t-γ_L[u]	172.4	-75.2	87.0	-176.9	(42.8, 128.8)o	2.8	15.9
c-ε_L[d]	-2.3	-91.2	115.8	-179.2	(38.5, 145.0)p	4.7	17.8
				Ac-γ*c*AmH$^+$p-NHMe			

t-γ_L[d]	-173.5	-83.5	95.9	176.8	(33.1,-113.9)q	0.0[r]	0.0
c-ε_L[d]	7.2	-81.5	121.0	178.9	(33.6,-104.6)s	4.3	4.3
c-α_L[d]	13.6	-98.4	-12.4	172.9	(36.1,-112.7)t	15.9	15.9

[a] χ^0= 12.4º, χ^1= -34.4º, χ^2= 42.6º, χ^3= -33.9º and χ^4= 13.7º. [b] E= -629.0390489 a.u. [c] χ^0= 4.8º, χ^1= -29.4º, χ^2= 41.2º, χ^3= -37.6º and χ^4= 21.0º. [d] χ^0= -23.8º, χ^1= 34.1º, χ^2= -32.6º, χ^3= 18.3º and χ^4= 3.6º. [e] χ^0= -16.0º, χ^1= 34.0º, χ^2= -39.7º, χ^3= 29.0º and χ^4= -8.1º. [f] E= -629.044445 a.u. [g] χ^0= 0.0º, χ^1= -22.8º, χ^2= 36.4º, χ^3= -36.0º and χ^4= 22.6º. [h] χ^0= -22.6º, χ^1= 40.8º, χ^2= -39.9º, χ^3= 23.8º and χ^4= 2.2º. [i] χ^0= -26.8º, χ^1= 3.9º, χ^2= 18.6º, χ^3= -34.3º and χ^4= 38.9º. [j] χ^0= 32.2º, χ^1= -42.3º, χ^2= 37.5º, χ^3= -18.1º and χ^4= -9.4º. [k] χ^0= -39.6º, χ^1= 20.5º, χ^2= 5.1º, χ^3= -29.4º and χ^4= 43.5º. [l] E= -629.028031 a.u. [m] χ^0= -39.8º, χ^1= 20.4º, χ^2= 5.2º, χ^3= -29.3º and χ^4= 43.9º. [n] χ^0= -19.3º, χ^1= 35.4º, χ^2= 40.4º, χ^3= 28.8º and χ^4= -5.6º. [o] χ^0= -26.8º, χ^1= 0.8º, χ^2= 23.4º, χ^3= -38.6º and χ^4= 41.4º. [p] χ^0=-31.5º, χ^1= 36.9º, χ^2= -30.9º, χ^3= 12.6º and χ^4= 12.3º. [q] χ^0=-13.4º, χ^1= 28.3º, χ^2= -33.1º, χ^3= 24.6º and χ^4= -7.0º. [r] E= -629.049002 a.u. [s] χ^0= -8.5º, χ^1= 25.9º, χ^2= -33.8º, χ^3= 28.1º and χ^4= -12.4º. [t] χ^0= -13.9º, χ^1= 30.5º, χ^2= -36.1º, χ^3= 27.5º and χ^4= -8.4º.

Table 3.4.2 *Relative free energy in the gas-phase and in aqueous solutions (ΔG^{gp} and ΔG^{H2O}, respectively; in kcal/mol) for the minimum energy conformations characterized for Ac-βtAmH$^+$p-NHMe, Ac-βcAmH$^+$p-NHMe, Ac-γtAmH$^+$p-NHMe and Ac-γcAmH$^+$p-NHMe at the B3LYP/6-31+G(d,p) level. The relative free energy in the gas-phase and in aqueous solutions calculated with respect to the lowest energy minimum of the most stable isomer ($\Delta G^{\#gp\#}$ and $\Delta G^{\#H2O\#}$, respectively; in kcal/mol) are also displayed.*

# Conf.	ΔG^{gp}	$\Delta G^{\#gp\#}$	ΔG^{H2O}	$\Delta G^{\#H2O\#}$
		Ac-β*t*AmH$^+$p-NHMe		
t-γ_L[u]	0.0[a]	7.6	0.5	8.1
c-α_L[u]	8.2	15.8	1.7	9.3
c-α_L[d]	14.8	22.4	0.0	7.6
		Ac-β*c*AmH$^+$p-NHMe		
t-γ_L[d]	0.0[b]	1.7	0.0	3.6
t-γ_L[u]	2.2	3.9	3.2	6.8
c-α_L[d]	9.3	11.0	3.9	7.5
c-α_L[u]	9.1	10.8	3.7	7.3
c-ε_L[u]	13.8	15.5	3.7	7.3
		Ac-γ*t*AmH$^+$p-NHMe		
g^+-δ_L[u]	0.0[c]	17.4	11.9	16.3

g^+-δ_L[u]	0.2	17.6	1.0	5.4
t-γ_L[d]	2.0	19.4	1.9	6.3
t-γ_L[u]	2.4	19.8	3.3	7.6
c-ε_L[d]	2.9	20.3	0.0	4.4
		Ac-γ*c*AmH$^+$p-NHMe		
t-γ_L[d]	0.0[d]	0.0	1.3	1.3
c-ε_L[d]	5.3	5.3	0.0	0.0
c-α_L[d]	20.9	20.9	10.5	10.5

[a] G=-628.831142 a.u. [b] G= -628.840562 a.u. [c] G=-628.815530 a.u. [d] G= -628.843287 a.u.

Ac-βtAmH$^+$p-NHMe. The lowest energy conformation found for this dipeptide corresponds to the t-γ_L[u] (Figure 3.4.3a), which is the only minimum identified with ω_0 arranged in *trans*. This structure is stabilized by a backbone···backbone hydrogen bonding interaction involving the NHMe and Ac blocking groups and by an intraresidue interaction between the ammonium side group and the carbonyl oxygen atom of the β*t*AmH$^+$p. The latter side chain···backbone interaction is consequence of the change in the pyrrolidine puckering produced by the protonation of the lowest energy minimum of Ac-β*t*Amp-NHMe, which was the t-γ_L[d].[11] Thus, the C^β-exo ($_\beta$E) conformation found for the pyrrolidine ring in this minimum transforms into C^β-endo ($^\beta$E) when the neutral peptide transforms into Ac-β*t*AmH$^+$-NHMe.

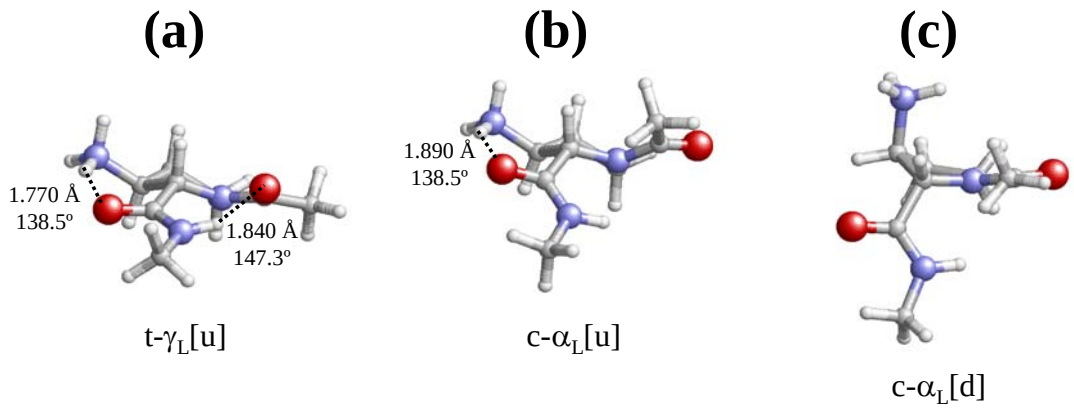


Figure 3.4.3: *Representation of the minimum energy conformations characterized in the gas-phase for the Ac-βtAmH$^+$p-NHMe dipeptide.*

The peptide bond ω_0 shows a *cis* arrangement in the other two minima characterized for Ac-β*t*AmH$^+$p-NHMe, c-α_L[u] (Figure 3.4.3b) and c-α_L[d] (Figure 3.4.3c), which only differ in the puckering of the five membered ring. Thus, the *up* puckering ($^{\beta}$E) favors the formation of a side chain···backbone interaction in the former structure, while this interaction is precluded by the down arrangement ($_{\beta}$E) showed in the latter conformation. The energy difference between such two minima, 5.9 kcal/mol, provides an estimation of the strength of the interaction between the charged ammonium group and the carbonyl oxygen atom of the same residue. Comparison with the results obtained for the Ac-β*t*Amp-NHMe evidences the annihilation of the ε_L backbone conformation, which was identified in two minima of the neutral peptide, upon protonation of the amino side group. On the other hand, the c-α_L[u] minimum was detected for both Ac-β*t*Amp-NHMe and Ac-β*t*AmH$^+$p-NHMe, even although the two dipeptides differ significantly in the φ,ψ backbone dihedral angles *i.e.* φ,ψ= -69.6º,-35.1º for the neutral compound. Indeed, the dihedral angles of the two c-α_L minima found for Ac-β*t*AmH$^+$p-NHMe are relative close to those usually observed in the so-called *bridge* region (φ,ψ= -80º,0º),[22] which are typically adopted by amino acids accommodated in the *i*+2 position of types I and II β-turns.

Inspection of the free energies listed in Table 3.4.2 shows the importance of the ZPVE, thermal and entropic corrections to the relative stability of the minima characterized for the Ac-β*t*AmH$^+$p-NHMe dipeptide. Consideration of these statistical terms for the transformation of ΔE^{gp} into ΔG^{gp} produces a relative

destabilization of 1.4 and 2.1 kcal/mol for the c-α_L[u] and c-α_L[d] conformations, respectively. According to a Bolzmann distribution of minima, the ΔG^{gp} values indicate that the only populated conformation at room temperature in the gas-phase is the t-γ_L[u].

Table 3.4.2 includes the relative free energies in aqueous solution. As can be seen, this solvent introduces significant changes in the relative stability of the three minima. Thus, the c-α_L[d] conformation, which was the least stable in the gas-phase, becomes the most favored in water. However, it is worth noting that the ΔG^{H2O} values of the other two conformations are relatively close. Indeed, the population predicted at room temperature for the c-α_L[d], t-γ_L[u] and c-α_L[u] conformations in water is 67.3%, 28.9% and 3.7%, respectively. Our previous calculations on Amp derivatives also evidenced a significant stabilization of the conformers with the peptide bond ω_0 arranged in *cis*,[11] *i.e.* the most stable conformation found for Ac-β*t*Amp-NHMe in aqueous solution was the c-ε_L[d], the other minima being destabilized by more than 3.5 kcal/mol in this environment.

Previous studies[7,11] evidenced that, although the stability of the *cis* conformers in protic solvents able to form specific hydrogen bonds with the solute is overestimated by PCM calculations, the general tendencies provided by this theoretical method for Pro derivatives describe very satisfactorily the experimental observations from a qualitative point of view. Thus, PCM calculations recently predicted that ω_0 exhibits a considerably smaller probability of adopting a *cis* disposition in α-methylproline and α-phenylproline than in Pro,[7] which was in good agreement with experimental information. Comparison of the results provided in Table 3.4.2 for Ac-β*t*AmH^+p-NHMe with those reported for Ac-β*t*Amp-NHMe[11] and Ac-Pro-NHMe[7] at the same theoretical level suggests that, although the incorporation of the amino group to the pyrrolidine ring enhances the stability of the conformers with ω_0 arranged in *cis*, its protonation tends to compensate this effect. Thus, although the c-ε_L[u] conformation was predicted as the most favored for Ac-Pro-NHMe in water, the t-α_L and t-γ_L were unfavored by only 1.3 and 0.3 kcal/mol, respectively.[7] These differences are similar to that found for the t-γ_L[u] conformation of Ac-β*t*AmH^+p-NHMe (0.5

kcal/mol in Table 3.4.2) and significantly smaller than that reported for the t-γ_L[d] minimum of Ac-β*t*Amp-NHMe (6.4 kcal/mol in reference 11). However, caution is required in the analysis of PCM results, when protic solvents able to form specific solute-solvent interactions are considered.

Ac-βcAmH$^+$p-NHMe. Five minimum energy conformations, including the three with the peptide bond ω_0 arranged in *cis*, were obtained for Ac-β*c*AmH$^+$p-NHMe, while six were found for the non-protonated Ac-β*c*Amp-NHMe. In both cases the lowest energy minimum corresponds to the t-γ_L[d] (Figure 3.4.4a) evidencing that the stability provided by the side chain···backbone interaction, which forms simultaneously to the backbone···backbone hydrogen bond typically associated to the γ_L conformation, is independent of the protonation state. In contrast, the t-γ_L[u] (Figure 3.4.4b) was only characterized as minimum in the potential energy surface of Ac-β*c*AmH$^+$p-NHMe dipeptide. This structure, which also presents an intramolecular hydrogen bond between the side ammonium group and the O=C of the own AmH$^+$p residue, only differ from the global minimum in the arrangement of the five membered ring, *i.e.* $_\beta$E and $^\beta$E for the t-γ_L[d] and t-γ_L[u], respectively. The t-α_L[d] and t-α_L[u] conformations of Ac-β*c*Amp-NHMe, which were found to be destabilized by 4.0 and 8.8 kcal/mol, respectively,[11] annihilate upon protonation of the side amino group.

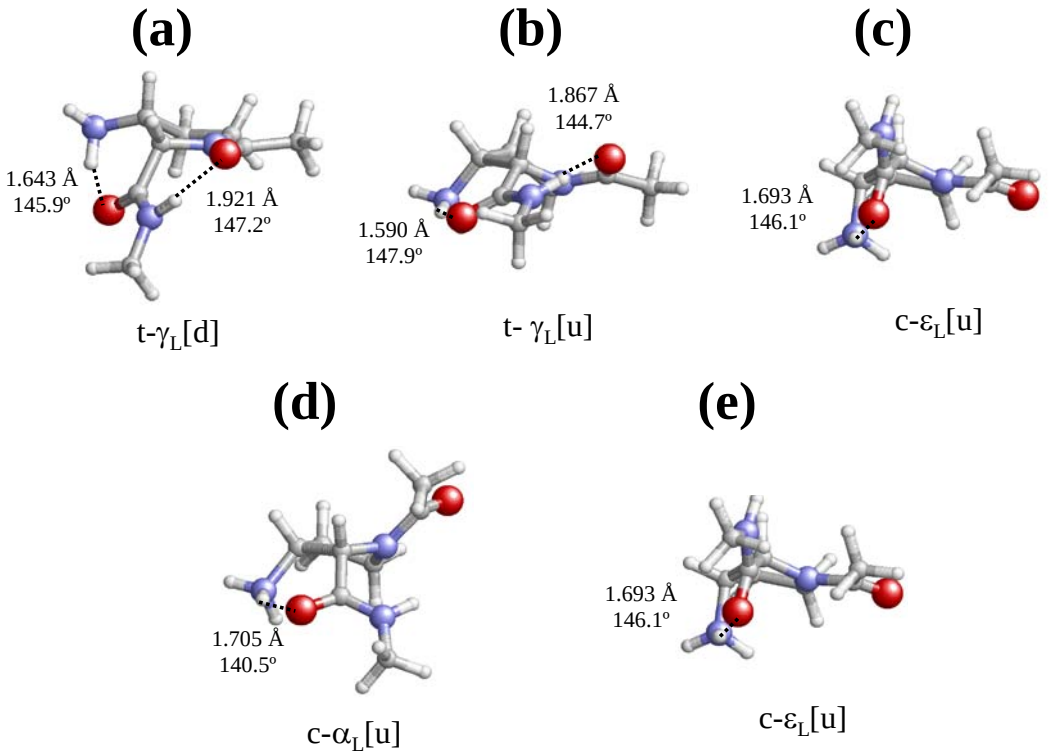


Figure 3.4.4: *Representation of the minimum energy conformations characterized in the gas-phase for the Ac-βcAmH$^+$p-NHMe dipeptide.*

The c-α_L[d] (Figure 3.4.4c) is the most stable minimum of Ac-β*c*AmH$^+$p-NHMe with the peptide bond ω_0 arranged in *cis*. Although this conformation presents a strong side chain···backbone intraresidue interaction, it is unfavored by 7.2 kcal/mol with respect to the lowest energy minimum. These results are fully consistent with those showed above for Ac-β*t*AmH$^+$p-NHMe, which indicated that the structures with ω_0 in *cis* become strongly destabilized upon protonation of the amino group attached to the C^β atom. Thus, for Ac-β*c*Amp-NHMe the energy of the c-α_L[d] was higher than that of the global minimum by 4.3 kcal/mol only. On the other hand, the other two minima found for Ac-β*c*AmH$^+$p-NHMe correspond to the c-α_L[u] (Figure 3.4.4d) and c-ε_L[u] (Figure 3.4.4e) with ΔE^{gp} values of 10.3 and 11.0 kcal/mol, respectively. Interestingly, the φ,ψ values of the two minima that show a α_L conformation are significantly distorted with respect to those expected for an ideal conformation. These deformations, which were also

detected in the two c-α_L minima of Ac-β*t*AmH$^+$p-NHMe, are consequence of the interaction between the ammonium group and the peptide bond ω.

Inspection to the ΔG^{gp} values indicates that the statistical corrections added to the electronic energies produce a relative destabilization of the t-γ_L[u], c-α_L[d] and c-ε_L[u] structures, which range from 0.6 to 2.8 kcal/mol. In opposition, the c-α_L[u] conformation becomes more stable by 1.2 kcal/mol, even although its population in the gas-phase is negligible. Thus, the only structure of Ac-β*c*AmH$^+$p-NHMe that presents a significant population (97.6%) in the gas-phase is the t-γ_L[d]. Similarly, this minimum is the only populated conformation in aqueous solution, the ΔG^{H2O} of the other four conformers ranging between 3.2 and 3.9 kcal/mol. The fact that the global minimum in the gas-phase is also the most favored conformation in water represents a remarkable difference with respect to the Ac-β*c*Amp-NHMe dipeptide. Thus, for the latter system the t-γ_L[d] was destabilized by 3.3 kcal/mol in aqueous solution, the c-ε_L[d] becoming the most favored in this polar environment. These results are fully consistent with the destabilization of the structures with ω_0 arranged in *cis* discussed above for Ac-β*t*AmH$^+$p-NHMe.

Ac-γtAmH$^+$p-NHMe. The conformational preferences displayed in Table 3.4.1 for Ac-γ*t*AmH$^+$p-NHMe are unique. As can be seen, two almost isoenergetic minima with the peptide bond ω_0 arranged in a *gauche*$^+$ conformation (labeled as g$^+$) are the most favored for this dipeptide. The distortion from the planarity of the peptide bond in these structures, g$^+$-δ_L[u] (Figures 3.4.5a and 3.4.5b), must be attributed to the strength of the side chain···backbone interaction, which shows a favorable geometry because of the N-exo ($_N$E) conformation of the pyrrolidine ring, *i.e.* this envelope conformation breaks the planar disposition of the peptide bond. This distortion is evidenced by a notable pyramidalization of the amide nitrogen, the sum of the valence angles around this atom (θ) being 338.5 and 340º for the two g$^+$-δ_L[u] minima. This large deformation evidences that the pyramidalization of ω_0 in the g$^+$-δ_L[u] structures is similar, or even higher, than that observed for the bicyclic amide nitrogen of highly constrained Pro analogues.[23] Interestingly, the two g$^+$-δ_L[u] minima only differ in the dihedral angle ψ, which defines the orientation of peptide bond ω. The orientation of the

polar -CONH- moiety in such minima does not affect to their intrinsic stability in the gas-phase, *i.e.* the ΔG^{gp} values differ by 0.2 kcal/mol only, even though their relative stabilities in aqueous solution are completely different. Thus, the strength of the solute···solvent attractive interactions increases with the accessibility of this peptide group to the solvent, *i.e.* the g^{+}-δ_L[u] conformation with ψ= 104.6º is favored by 11.1 kcal/mol, which explains the difference found in their ΔG^{H2O} values (Table 3.4.2).

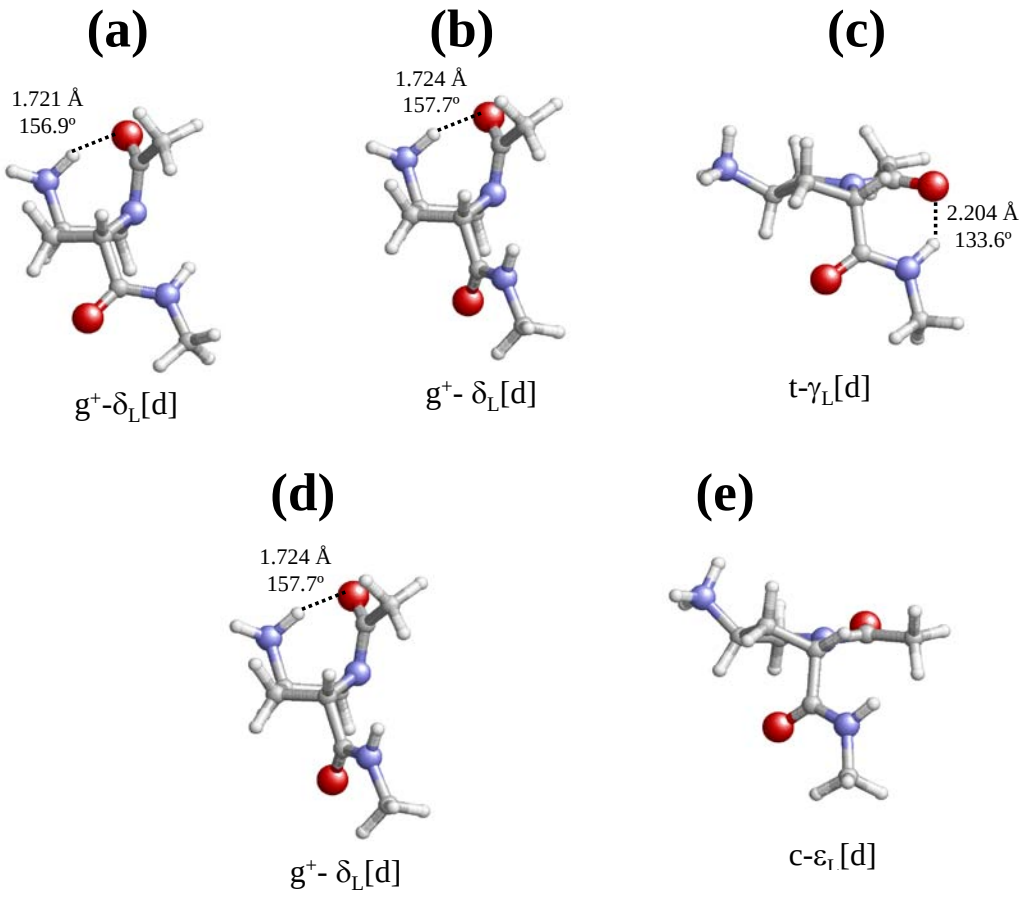


Figure 3.4.5: *Representation of the minimum energy conformations characterized in the gas-phase for the Ac-γtAmH^{+}p-NHMe dipeptide.*

The next two minima, which correspond to the conventional t-γ_L[d] (Figure 3.4.5c) and t-γ_L[u] (Figure 3.4.5d) conformations, are destabilized by 2.0 and 2.8 kcal/mol, respectively. These structures, in which the two peptide bonds adopt a planar *trans* arrangement, are stabilized by the backbone···backbone intramolecular hydrogen bond only, no side chain···backbone interaction being detected. This feature explains the lower stability of the two t-γ_L conformations

with respect to the two g^+-δ_L minima. Thus, the strong side chain···backbone interaction of the latter, which is more attractive than the seven-membered intramolecular hydrogen bond, almost compensates the energy penalty associated to the geometric deformation of the peptide bond ω_0. Finally, the least stable conformation, c-ε_L[d] (Figure 3.4.5e), which is unfavored by 4.7 kcal/mol with respect to the global minimum, corresponds to the only conformation with ω_0 arranged in *cis*. As can be seen, this structure does not show any N-H···O intramolecular interaction.

Comparison of the ΔE^{gp} and ΔG^{gp} values reveals that, in this case, the influence of the statistical corrections is very small. Thus, with exception of two minima of higher energy, which become stabilized by 0.4 (t-γ_L[u]) and 1.8 kcal/mol (c-ε_L[d]) by the addition of the thermal and entropic terms, the relative stability of the other three structures remained unaltered. On the other hand, the relative free energy order undergoes a drastic change in aqueous solution. Specifically, the most stable structure in water is the c-ε_L[d], which was the least favored in the gas-phase. This feature is consistent with the overestimation of the conformers with ω_0 arranged in *cis* previously attributed to the PCM solvation model. Furthermore, the lowest energy minimum in the gas-phase is the least favored in aqueous solution, this feature being consequence of the poor interaction between the C=O group of the peptide bond ω and the solvent (Figure 3.4.5a). In contrast, the second minimum with a distorted peptide bond results less favored in aqueous solution than the c-ε_L[d] by 1.0 kcal/mol only, which is due to the very favorable interactions of the peptide bond ω with the environment. Finally, the stability in water of the two structures with the t-γ_L backbone conformation is similar to that obtained in the gas-phase. The overall of these results indicate that the conformational populations predicted for Ac-γ*t*AmH^+p-NHMe in aqueous solution, considering Boltzmann distribution of the identified minima, are: 81.7 % c-ε_L[d], 15.0 % g^+-δ_L[d], and 3.3% t-γ_L[d].

Ac-γcAmH$^+$p-NHMe. Only three minima were detected for Ac-γ*c*AmH^+p-NHMe evidencing that the protonation of the amino side group reduces drastically the conformational flexibility of neutral Ac-γ*c*Amp-NHMe.[11] Thus, seven minima with relative energies of up 7.9 kcal/mol were found for the latter dipeptide: three

with ω_0 arranged in *trans* and four in *cis*. The lowest energy minimum found for Ac-γ*c*AmH$^+$p-NHMe corresponds to the t-γ_L[d] (Figure 3.4.6a), which presents both backbone···backbone and side chain···backbone favorable interactions, the latter being facilitated by the $_\beta$E arrangement of the pyrrolidine ring. This conformation was also identified as the most stable conformation of Ac-γ*c*Amp-NHMe, even although the t-γ_L[u] and t-α_L[d] minima annihilate when the amino group of this dipeptide transforms into ammonium.

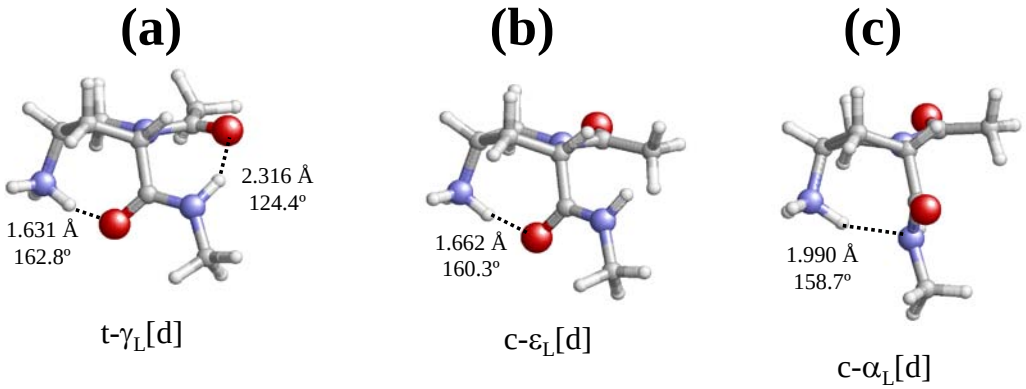


Figure 3.4.6: *Representation of the minimum energy conformations characterized in the gas-phase for the Ac-γcAmH$^+$p-NHMe dipeptide.*

The second minimum of Ac-γ*c*AmH$^+$p-NHMe shows a c-ε_L[d] structure (Figure 3.4.6b) with the pyrrolidine ring arranged like in the global minimum, *i.e.* $_\beta$E conformation. This structure, which is unfavored by 4.3 kcal/mol, is similar to the least stable minimum of the Ac-γ*t*AmH$^+$p-NHMe dipeptide, even although the *cis* disposition of the substituent allows the formation of an attractive side chain···backbone interaction that was not possible in the latter compound. Finally, the last minimum, c-α_L[d] (Figure 3.4.6c), is strongly destabilized, *i.e.* ΔE^{gp}= 15.9 kcal/mol. This should be attributed to the simultaneous combination of a number of factors: (i) the *cis* arrangement of ω_0; (ii) the lack of backbone···backbone intramolecular hydrogen bond; and, especially, (iii) the nature of side chain···backbone interaction, which is of the N-H···N type. Thus, the stabilizing effect provided by the latter interaction is lower than that achieved through the N-H···O(=C) one.[24] It is worth noting that the c-ε_L[d] and c-α_L[d] minima of Ac-γ*c*Amp-NHMe were unfavored by 6.0 and 2.7 kcal/mol,[11] respectively, which

reflects the large change that the ionization of the side group produces in the potential energy surface of this dipeptide.

The most significant change produced by the transformation of ΔE^{gp} into ΔG^{gp} is the destabilization of the c-α_L[d] conformation, which increases 5.0 kcal/mol. Accordingly, the t-γ_L[d] is predicted to be the only populated conformation in the gas-phase. On the other hand, inspection to the ΔG^{H2O} values indicates again that the PCM model produces a considerable stabilization of the c-ε_L[d] structure. Thus, the latter conformation becomes the most favored in aqueous solution, whereas the t-γ_L[d] is higher in energy by 1.3 kcal/mol, *i.e.* the populations of the c-ε_L[d] and t-γ_L[d] conformations in aqueous solution are 90.1% and 9.9%, respectively. These preferences are significantly different from those reported for Ac-γ*c*Amp-NHMe, in which the c-ε_L[d] was predicted to be only conformation with a significant population in aqueous solution,[11] *i.e.* all the other minima were destabilized by more than 2 kcal/mol.

Relative Stability of the four isomers. The $\Delta E^{\#gp\#}$ and $\Delta G^{\#gp\#}$ values displayed in Tables 3.4.1 and 3.4.2, respectively, indicate that the Ac-γ*c*AmH$^+$p-NHMe is the most stable isomer, the Ac-β*c*AmH$^+$p-NHMe being unfavored by only 1.7 kcal/mol (2.8 kcal/mol in terms of $\Delta E^{\#gp\#}$). The stability of these isomers, which is significantly higher than that of the analogues with a *trans* disposition of the charged side group, should be attributed to the formation of side chain···backbone interactions. Thus, intraresidue interactions are clearly stronger when the substitution is attached in *cis*. This feature is reflected by the H···O distances displayed in Figures 3.4.4 and 3.4.6, which are ~1.65 Å for the minima of the γ*c*AmH$^+$p- and β*c*AmH$^+$p-containing dipeptides. In contrast, these distances are 1.770 and 1.890 Å for the two minima of Ac-β*t*AmH$^+$p-NHMe that present a side chain···backbone interaction (Figures 3.4.3a and 3.4.3b), this isomer being unfavored by 7.6 kcal/mol (6.2 kcal/mol in terms of $\Delta E^{\#gp\#}$) with respect to the most stable. Similarly, the values of the ∠N-H···O angles are consistent with a more favorable interaction when the substituent is attached in *cis*. The Ac-γ*t*AmH$^+$p-NHMe is not able to form side chain···backbone interactions without induce a remarkable distortion of the peptide bond, which produces a significant

energy penalty. Consequently, this isomer is unfavored by 17.4 kcal/mol (13.1 kcal/mol in terms of $\Delta E^{\#gp\#}$) with respect to the Ac-γ*c*AmH^+p-NHMe dipeptide.

The relative stability order obtained for the neutral Amp-containing dipeptides was identical, *i.e.* Ac-γ*c*Amp-NHMe > Ac-β*c*Amp-NHMe > Ac-β*t*Amp-NHMe > Ac-γ*t*Amp-NHMe.[11] However, in this case the energy differences among the different isomers were significantly lower than those obtained for the AmH^+p derivatives. Thus, the lowest $\Delta G^{\#gp\#}$ value of the β*c*Amp-, β*t*Amp- and γ*t*Amp-containing dipeptides was 0.6, 1.0 and 1.4 kcal/mol, respectively. The remarkable energetic difference between Amp and AmH^+p dipeptides should be attributed to the strength of the intraresidue interaction, which is significantly higher when the side group is ionized.

On the other hand, $\Delta G^{\#H2O\#}$ values indicate that the solvent does not alter the preferences by the isomers with the ammonium group attached in *cis* with respect to those in *trans*. Thus, the Ac-γ*c*AmH^+p-NHMe is the most stable in water followed by the Ac-β*c*AmH^+p-NHMe, which is unfavored by 3.6 kcal/mol. Accordingly, the solvent destabilizes the latter isomer 1.9 kcal/mol with respect to the gas-phase. Regarding to the Ac-β*t*Amp-NHMe and Ac-γ*t*Amp-NHMe, their relative stabilities are exchanged with respect to the gas-phase, the latter being favored with respect to the former by 3.2 kcal/mol. Thus, these isomers are 5.4 and 7.6 kcal/mol, respectively, less stable than the Ac-γ*c*AmH^+p-NHMe one. Comparison with the results reported in aqueous solution for the neutral Amp-containing dipeptides reveals considerable differences. Specifically, the relative energy order reported for these dipeptides were:[11] Ac-γ*c*Amp-NHMe ≈ Ac-β*t*Amp-NHMe ($\Delta G^{\#H2O\#}$= 0.2 kcal/mol) > Ac-γ*t*Amp-NHMe ($\Delta G^{\#H2O\#}$= 1.7 kcal/mol) > Ac-β*c*Amp-NHMe ($\Delta G^{\#H2O\#}$= 5.9 kcal/mol). Accordingly, the protonation of the amino substituent in water produces a pronounced stabilization of the isomers substituted in *cis* with respect to those with the substituent in *trans*.

3.4.4 Conclusions

DFT calculations at the B3LYP/6-31+G(d,p) level have been used to examine the conformational preferences of Ac-β*t*AmH$^+$p-NHMe, Ac-β*c*AmH$^+$p-NHMe, Ac-γ*t*AmH$^+$p-NHMe and Ac-γ*c*AmH$^+$p-NHMe, both in the gas-phase phase and aqueous solution, which have been compared with those reported for the neutral analogues Ac-β*t*Amp-NHMe, Ac-β*c*Amp-NHMe, Ac-γ*t*Amp-NHMe and Ac-γ*c*Amp-NHMe. Results allow us to draw the following conclusions:

(i) Protonation of the amino group attached to the β- or γ-position of the pyrrolidine ring reduces the backbone conformational flexibility of the Amp derivatives, which was low compared to that conventional Pro. Specifically, the number of minima identified for the four AmH$^+$p-containing dipeptides was smaller than that found for their Amp analogues. Furthermore, the relative energies and free energies increase upon protonation of the side group.

(ii) The stability of conformations with ω_0 in *cis* is significantly lower for AmH$^+$p than for Amp. This is a very remarkable result because the incorporation of the non-protonated amino group to the pyrrolidine ring of conventional Pro produced a stabilization of such conformations. Accordingly, the population of *cis* conformers in Amp/AmH$^+$p derivatives could be easily controlled with the pH.

(iii) The intrinsic conformational preferences of the Ac-γ*t*AmH$^+$p-NHMe dipeptides show that the strength of the side chain···backbone attractive interaction allows compensate the energy penalty associated to the deformation of the peptide bond. Thus, in order to reach such interactions, this compound tends to break the planarity of the peptide bond inducing a large pyramidalization of the amide nitrogen atom.

(iv) The Ac-γ*c*AmH$^+$p-NHMe dipeptide, followed by the Ac-β*c*AmH$^+$p-NHMe, are the more stable isomers in both the gas-phase and aqueous solution. The *cis* disposition of the substituent is favored because of the attractive side chain···backbone interactions. Comparison, between

Amp and AmH^+p derivatives reveals that the energy gap between the different isomers increases upon ionization of the side group. This should be attributed to the strength of the intraresidue interactions, which is higher when a positively charged ammonium group is involved.

3.4.5 References

1. Alemán, C.; Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Haspel, N.; Zheng, J.; Casanovas, J.; Wolfson, H.; Nussinov, R. Phys. Biol. 2006, 3, S54.
2. Tsai, C.-J.; Zheng, J.; Zanuy, D.; Haspel, N.; Wolfson, H.; Alemán, C.; Nussinov, R. Proteins 2007, 68, 1.
3. (a) Zheng, J.; Zanuy, D; Haspel, N.; Tsai, C.-J.; Alemán, C.; Nussinov, R. Biochemistry 2007, 46, 1205. (b) Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. J. Phys. Chem. B 2007, 111, 3236. (c) Ballano, G.; Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. J. Phys. Chem. B 2008, 112, 13101.
4. (a) Chakrabarti, P.; Pal, D. Prog. Biophys. Mol. Biol. 2001, 76, 1. (b) Marraud, M.; Aubry, A. Biopolymers 1996, 40, 45. (c) MacArthur, M. W.; Thornton, J. M. J. Mol. Biol. 1991, 218, 397. (d) Rose, G. D.; Gierasch, L. M.; Smith, J. A. Adv. Protein Chem. 1985, 37, 1.
5. (a) Boussard, G.; Marraud, M.; Aubry, A. Biopolymers 1979, 18, 1297. (b) V. Madison, K. D. Kopple, J. Am. Chem. Soc. 1980, 102, 4855. (c) Liang, G. B.; Rito, C. J.; Gellman, S. H. Biopolymers 1992, 32, 293. (d) Beausoleil, E.; Lubell, W. D. J. Am. Chem. Soc. 1996, 118, 12902.
6. (a) Kang, Y. K. J. Phys. Chem. B 2006, 110, 21338. (b) Kang, Y. K.; Byun, B. J. J. Phys. Chem. B 2007, 111, 5377. (c) Rivail, J. L.; Bouchy, A.; Loos, P. F.J. Arg. Chem. Soc. 2006, 94, 19.
7. Flores-Ortega, A.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C.; Casanovas, J. J. Org. Chem. 2008, 73, 3418.
8. (a) Baures, P. W.; Ojala, W. H.; Gleason, W. B.; Johnson, R. L. J. Pept. Res. 1997, 50, 1. (b) Flippen-Anderson, J. L.; Gilardi, R.; Karle, I. L.; Frey, M. H.; Opella, S. J.; Gierasch, L. M.; Goodman, M.; Madison, V.; Delaney, N. G. J. Am. Chem. Soc. 1983, 105, 6609. (c) Madison, V.; Delaney, N. G. Biopolymers 1983, 22, 869. (d) Delaney, N. G.; Madison, V. J. Am. Chem. Soc. 1982, 104, 6635.
9. (a) Benzi, C.; Improta, R.; Scalmani, G.; Barone, V. J. Comput. Chem. 2002, 23, 341. (b) Lam, J. S. W.; Koo, J. C. P.; Hudáky, I.; Varro, A.;

Papp, J. G.; Penke, B.; Csizmadia, I. G. J. Mol. Struct. (Theochem) 2003, 666-667, 285-289.

10. Song, I. K.; Kang, Y. K. J.Phys.Chem.B 2005, 109, 16982.
11. Flores-Ortega, A.; Casanovas, J.; Nussinov, R.; Alemán, C. J. Phys. Chem. B 2008, 112, 14045
12. Jhon, J. S.; Kang Y. K. J. Phys. Chem. B 2007, 111, 3496.
13. (a) Kang, Y. K.; Park H. S. J.Phys. Chem. B 2007, 111, 12551. (b) Kang K. Y.; Byun B. J. J. Phys. Chem. B 2007, 111, 5377.
14. Flores-Ortega, A.; Casanovas, J.; Nussinov, R.; Alemán, C. J. Phys. Chem. B 2007, 111, 5475.
15. (a) Porter, E. A.; Wang, X.; Schmitt, M. A.; Gellman, S. H. Org. Lett. 2002, 4, 3317. (b) Farrera-Sinfreu,J.; Zaccaro, L.; Vidal, D.; Salvatella, X.; Giralt, E.; Pons, M.; Albericio, F; Royo, M. J. Am. Chem. Soc. 2004, 126, 6048.
16. Gaussian 03, Revision B.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr., T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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.; C. Strain, M.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Pittsburgh PA, 2003.
17. (a) Becke, A. D. J. Chem. Phys. 1993, 98, 1372. (b) Lee, C.; Yang, W.; Parr, R. G. Phys. Rev. B 1993, 37, 785.

18. McLean, A. D.; Chandler, G. S. J. Chem. Phys. 1980, 72, 5639.
19. (a) Miertus, M.; Scrocco, E.; Tomasi, J. Chem. Phys. 1981, 55, 117. (b) Tomasi, J.; Mennucci, B.; Cammi, R. Chem. Rev. 2005, 105, 2999.
20. Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadia, I. G. J. Am. Chem. Soc. 1991, 113, 6256.
21. (a) Hudáky, I.; Perczel, A. J. Mol. Struct. (THEOCHEM) 2003, 630, 135. (b) Hudáky, I.; Baldoni, H. A.; Perczel, A. J. Mol. Struct. (THEOCHEM) 2002, 582, 233.
22. (a) Valle, G.; Crisma, M.; Toniolo, C.; Holt, E. M.; Tamura, M.; Bland, J.; Stammer, C. H. Int. J. Pept. Protein Res. 1989, 34, 56. (b) Benedetti, E.; Di Blasio, B.; Pavone, V.; Pedone, C.; Santini, A.; Crisma, M.; Valle, G.; Toniolo, C. Biopolymers 1989, 28, 175. (c) Benedetti, E.; Di Blasio, B.; Pavone, V.; Pedone, C.; Santini, A.; Barone, V.; Fraternalli, F.; Lelj, F.; Bavoso, A.; Crisma, M.; Toniolo, C. Int. J. Biol. Macromol. 1989, 11, 353. (d) Fabiano, N.; Valle, G.; Crisma, M.; Toniolo, C.; Saviano, M.; Lombardi, A.; Isernia, C.; Pavone, V.; Di Blasio, B.; Pedone, C.; Benedetti, E. Int. J. Pept. Protein Res. 1993, 42, 459..
23. Alemán, C.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Casanovas, J. J. Phys. Chem. B 2005, 109, 11836.
24. (a) Alemán, C.; Zanuy, D.; Casanovas, J. J. Phys. Chem. A 2003, 107, 4151. (b) Barril, X.; Alemán, C.; Orozco, M.; Luque, F. J. Proteins 1998, 32, 67. (c) Alemán, C.; Navas, J. J.; Muñoz-Guerra, S. J. Phys. Chem. 1995, 99, 17653.

4. Targeted replacement with constrained amino acid in tumor homing peptides

4.1 The energy landscape of a selective tumor-homing pentapeptide

Recently, a potentially powerful strategy based on the of phage-display libraries has been presented to target tumors via homing peptides attached to nanoparticles. The Cys-Arg-Glu-Lys-Ala (CREKA) peptide sequence has been identified as a tumor-homing peptide that binds to clotted plasmas proteins present in tumor vessels and interstitium. The aim of this work consists of mapping the conformational profile of CREKA to identify the bioactive conformation. For this purpose, a conformational search procedure based on modified Simulated Annealing combined with Molecular Dynamics was applied to three systems that mimic the experimentally used conditions: (i) the free peptide; (ii) the peptide attached to a nanoparticle; and (iii) the peptide inserted in a phage display protein. In addition, the free peptide was simulated in an ionized aqueous solution environment, which mimics the ionic strength of the physiological medium. Accessible minima of all simulated systems reveal a multiple interaction pattern involving the ionized side chains of Arg, Glu and Lys, which induces a β-turn motif in the backbone observed in all simulated CREKA systems.[*]

4.1.1 Introduction

Over the last few years Ruoslahti and coworkers[1-3] have identified a series of tumor-homing peptides by using *in vivo* screening of peptide libraries. These tumor-homing peptides were obtained from phage display libraries targeting highly-expressed molecular receptors in tumor blood vessels. Nanoparticles coated by the homing peptides were shown to efficiently focus at these sites and the potential of these nanoparticles for imaging and as drug-carriers has already been indicated. Among the homing peptides discovered by *in vivo* phage screening is a linear peptide that contains only 5 amino acids with the sequence CREKA (Cys-Arg-Glu-Lys-Ala). Synthesized CREKA peptide labeled with an attached fluorescent dye was detectable in human tumors from minutes to hours after intravenous injection, while it was essentially undetectable in normal tissues.[3] *In vivo* experiments revealed that CREKA bound to clotted plasma

[*] *The work described in this chapter previously appeared in J. Phys. Chem. B,* 2008,*112,* 8692-8700

proteins, establishing its being a clot-binding peptide. The CREKA peptide was further linked to amino dextran-coated iron oxide (SPIO) nanoparticles.[3] Strikingly, experiments with CREKA-SPIO nanoparticles indicated that they not only bind to blood and plasma clots but can also effectively induce further localized tumor clotting, thus amplifying the nanoparticle homing. It was further found that the chemical nature of the nanoparticle is not important for this activity since both CREKA-SPIO nanoparticles and CREKA-coated liposomes were found to cause clotting in tumor vessels. A similar activity, *i.e.* binding to plasma clot, was observed when CREKA was inserted into phage display proteins.[4] The clotting caused by CREKA in tumor vessels is expected to be useful in improving tumor detection by microscopic- and whole body- imaging techniques as well to tumor treatment by physical blockage of vessels through local embolism and drug-carrying nanoparticles.

At the same time, in spite of the potential interest in CREKA in cancer diagnostics and therapeutics, the application of this and other homing peptide sequences might be endangered by short half life times; after injection endogenous proteolytic enzymes could rapidly digest the peptides. Thus, protection from proteases is an important step in the development of potential applications of tumor-homing peptides. Several strategies have been proposed to protect proteogenic peptide sequences from proteases; among these, targeted replacements with synthetic amino acids have been the most successful.[5-12] Selective incorporation of synthetic amino acids induces a significant resistance against proteases not only at the mutated position but also at neighboring amino acids. Non-proteogenic amino acids are compounds with many potential applications in Nanobiology, and have been found to be very useful for the re-engineering of physical protein modules and the generation of nanodevices.[13-15].

Within the framework of a broad project devoted to improve the affinity of CREKA to its targets as well as to impart resistance against proteases, we have started by performing a computational analysis on such peptide sequence. Thus, the first step prior to the design of synthetic analogs which incorporate non-proteogenic amino acids, consists of the identification of the conformational profile and, if possible, the bioactive conformation of the peptide. Although no information about the conformational profile of this peptide is available, that is, the CREKA sequence does not occur in the Protein Data Base, experiments using

the free peptide, the CREKA-SPIO nanoparticles and phage display proteins containing expressed CREKA suggest that the bioactive conformation for this sequence does not depend on the molecular environment.[1-4] Identification of the low-energy regions of the conformational profiles of the natural CREKA sequence is important since any candidate synthetic analogue should not present significant differences with respect to the native. Theoretical methods based on computer simulations of atomic models have proven to be powerful techniques to sample the conformational space of biomolecules.[16-18]

In this work we use computer simulation methods based on Molecular Dynamics (MD) to characterize the conformational profile of the CREKA sequence and identify the corresponding bioactive conformation. For this purpose we used a multiple conformational search strategy considering the three experimentally tested environments: free peptide, peptide attached to a nanoparticle, and a peptide inserted in a phage display protein. The influence of the ionic strength on the conformational preferences of CREKA has also been examined. We note that computer simulation methods are usually applied considering a dilute solution of pure water while physiological media consists of ionized aqueous solutions. Accordingly, in this work we decided to examine how the addition of NaCl molecules to the aqueous environment affects the three-dimensional organization of the peptide.

4.1.1.1 Simulated Systems and Bioactive Conformation

The conformational profiles of four different molecular systems based on the CREKA peptide sequence have been determined using atomistic computational methods. The four systems under consideration are the following:

4.1.1.1.1 System I

The CREKA linear peptide was capped at the N- and C-terminus with acetyl (Ac) and *N*-methyl amide (NHMe) groups respectively, and was immersed in a dilute aqueous solution.

4.1.1.1.2 System II

CREKA was blocked at the ends with Ac and NHMe and was immersed in a solution that mimics the ionic strength of the physiological medium. The ionized solution was obtained by adding sodium and chloride ions to the environment.

Comparison between the conformational profiles obtained for systems I and II should permit testing the influence of the simulation conditions on the conformational preferences of the peptide.

4.1.1.1.3 System III

CREKA was attached to a surface through the sulfhydryl group of the Cys residue. This surface was formed by 100 spherical particles, which were distributed in a 10×10 square (47.5 $Å^2$), with van der Waals parameters [R= 2.35 Å and ε= 0.90 kcal/mol] and without electric charge. The whole system was immersed in a dilute aqueous solution. System III mimics the situation in which CREKA is bound to the surface of a SPIO nanoparticle. It should be noted that, in this simple model, electrostatic charges were omitted for the particles used to represent the surface. This is because many other chemical species that should induce a significant screening in the electrostatic interaction between the surface and the peptide have been also neglected in this model, *e.g.* neighboring peptides attached to the particle (about 8000 peptides were attached to each particle), ionic salts. The conformational freedom of the peptide is expected to be lower in III than in I and II because of the effect of having the end tethered to the surface.

4.1.1.1.4 System IV

This molecular system mimics the CREKA peptide when it is inserted in a phage display protein. Copies of the expressed peptides are usually presented on the flexible regions of the phage display proteins, for example in loops. In order to mimic such flexible regions, two Gly were added to each termini of CREKA, making the peptide sequence of system IV GGCREKAGG. The two Gly located at both ends of the chain were attached to spherical particles similar to those used in III to construct the surface. Such particles represent the secondary structure motifs that tether the flexible regions to the protein. The separation between the two added particles was restrained to a distance D through a force constant, k= 5 $kcal \cdot mol^{-1} \cdot Å^{-2}$, and the whole system was immersed in a dilute aqueous solution. The conformational profiles of IV were characterized considering D= 5.0, 7.0, 10.0, 15.0 and 20.0 Å. It is expected that such wide range of values will allow sampling a large number of potential positions and configurations in which the CREKA peptide is inserted in the flexible regions of phage display proteins.

Although the conformational profiles of I, II, III and IV may be significantly different, the bioactive conformation of the CREKA peptide sequence should appear within the set of lower-energy structures of each system. As stated in the Introduction, the experimental results suggest that the bioactive conformation is unique for this sequence independently of the molecular environment, *i.e.* the bioactive conformations when the peptide is free, attached to a SPIO nanoparticle or inserted in a phage display protein are practically identical. At the same time, the bioactive conformation does not necessarily correspond to the lowest minimum energy conformation, even though it must be contained within the set of accessible energy minima, *i.e.* the energy of the bioactive conformation is relatively close to that of the global minimum. Accordingly, a thorough exploration of the conformational space of the four CREKA-based systems is necessary to identify such a bioactive conformation.

4.1.2 Methods

4.1.2.1 Conformational Search

The structural features of the four CREKA-based systems studied in this work were explored using a modified simulated annealing combined with MD (SA-MD) as a sampling technique.[19,20] In the SA-MD strategy, which relies on temperature evolution in time, the starting temperature is gradually reduced during the simulation. Thus, the system is given the opportunity to surmount energy barriers in a search for conformations with energies lower than the local-minimum typically found by energy minimization. However, in practice, when used with a regular force-field, SA-MD does not always lead the system to the most stable region at the end of the simulation. Recent studies demonstrated that very low energies are obtained by minimizing the energy of structures generated at the initial and intermediate states of a SA-MD process.[21,22] In the present study we applied the latter strategy to locate the lower-energy minimum structures of the CREKA-based systems. A large number of structures (500) were selected for energy minimization from each SA-MD cycle, which is justified since the temperature was decreased very slowly along a relatively large trajectory (see below). As mentioned above, the bioactive conformation does not necessarily

correspond to the lowest energy minimum, even though it should be a low-energy structure. Accordingly, the sampling technique applied to the four systems under study must be robust enough to obtain low-energy structures that may be quasi-degenerate with the global minimum but situated in different valleys of the peptide landscape.

4.1.2.2 Molecular Dynamics and Computational Details.

Each simulated system was placed in the center of a cubic simulation box (*a*= 47.5Å) filled with at least 3405 explicit water molecules, which were represented using the TIP3 model.[23] One negatively charged chloride and two positively charged sodium atoms were added to the simulation box to reach electric neutrality (net charges were considered for Arg, Lys and Glu residues at neutral pH). For the conformational search of free CREKA under physiological conditions (system II), four additional NaCl molecules were added to the simulation box.

Before the production cycles with the modified SA-MD, the content and the density of the simulation box were equilibrated. For this purpose, each system was initially minimized to relax the conformational and structural tension using the conjugate gradient method. Next, different consecutive rounds of short MD runs were performed. Thus, 0.5 ns of NVT-MD at 500 K were used to homogeneously distribute the solvent and ions in the box. Next, 0.5 ns of NVT-MD at 298 K (thermal equilibration) and 0.5 ns of NPT-MD at 298 K (density relaxation) were run. The last snapshot of the NPT-MD was used as the starting point for the conformational search process. This initial structure was quickly heated to 900 K at a rate of 50 K/ps to force the molecule to jump to a different region of the conformational space. Along 10 ns, the 900 K-structure was slowly cooled to 500 K at a rate of 1 K per 25 ps. A total of 500 structures were selected and subsequently minimized during the first cycle of modified SA-MD. The resulting minimum energy conformations were stored in a rank-ordered library of low energy structures. The lowest energy minimum generated in a modified SA-MD cycle was used as starting conformation of the next cycle.

The 100 spherical particles used to construct the surface in system III were kept fixed at the initial positions in all MD simulations and energy minimizations.

The energy was calculated using the AMBER force-field,[24,25] with the required parameters taken from the AMBER libraries. Atom pair distance cutoffs were applied at 14.0 Å to compute the van der Waals and electrostatic interactions. Both temperature and pressure were controlled by the weak coupling method, the Berendsen thermo-barostat,[26] using a time constant for heat bath coupling and a pressure relaxation time of 1 ps. Bond lengths were constrained using the *SHAKE* algorithm[27] with a the numerical integration step of 2 fs.

4.1.2.3 Conformation Classification and Clustering Analysis.

In order to construct a list of unique minimum energy conformations, the 500 minimized structures provided by each cycle of modified SA-MD were compared not only among themselves but also with unique minima generated in previous cycles. The list was organized by rank ordering all the unique minimum energy conformations found following an increasing order or energy. Previously listed conformations were discarded. After testing different criteria, the one selected to identify unique minimum energy conformations was based on the values of virtual dihedral angles, which were defined using main-chain atoms, and the presence of interaction patterns, *i.e.* salt bridges, hydrogen bonds and dipole-dipole interactions.

Five virtual dihedral angles were defined considering the α-carbon atoms of the CREKA peptide and one methyl hydrogen atom of the Ac and NHMe capping groups. The hydrogen atoms of Ac and NHMe were chosen instead of the methyl carbon atoms, to retain the two-particle separation imposed by the consecutive α-carbon atoms of the peptide sequence. Each of these virtual dihedral angles was associated with the conformation of one amino acid of the CREKA peptide. The existence of different interactions was accepted on the basis of the following geometric criteria: a) salt bridges: distance between the centers of the interacting groups shorter than 4.50 Å; b) hydrogen bonds: distance H···O ($d_{H\cdots O}$) shorter than 2.50 Å and angle ∠N-H···O higher than 120.0º; and (c) dipole-dipole: distance between dipoles shorter than 3.00 Å and the interaction has not been counted as hydrogen bond. Two structures were considered different when they differ in at least one of their virtual dihedral angles by more than 60º or in at least one of the interactions counted.

For each system, all the structures categorized as different were subsequently clustered according to a criterion based on the presence of the intramolecular interactions mentioned above: salt bridges and hydrogen bonds. For each studied system, a representation of the number of clusters *versus* the number of structures within a given cluster is provided in the Supporting Information.

4.1.3 Results

4.1.3.1 Free CREKA in Unionized and Ionized Aqueous Solution (Systems I and II)

Thirteen production cycles of modified SA-MD were run for each of the two systems, with 500 selected structures minimized from each cycle. This procedure led to 3130 and 3114 unique minimum energy conformations for systems I and II, respectively. Figure 4.1.1, which represents the evolution of the number of unique conformations against the number of SA-MD cycles, clearly indicates that the conformational search converges for the two systems. However, interestingly, the molecular dynamics of the conformational search procedure was influenced by the chemical nature of the aqueous solution. Thus, inspection of the energies of the unique conformations obtained in each SA-MD cycle indicates that in unionized aqueous solution (system I) the energy of at least one new unique conformation is lower than those of the conformations previously generated for cycles 2, 3, 4 and 8, *i.e.* the global minimum appears in the 8^{th} cycle of modified SA-MD. In contrast, in the ionized solution (system II) the lowest energy conformation of free CREKA appears after only 3 cycles, and no improvement in the energy has been detected subsequently. Analysis of the number of new unique conformations generated by each cycle reveals that it is lower than 1% after the 9^{th} and 11^{th} cycle for I and II, respectively.

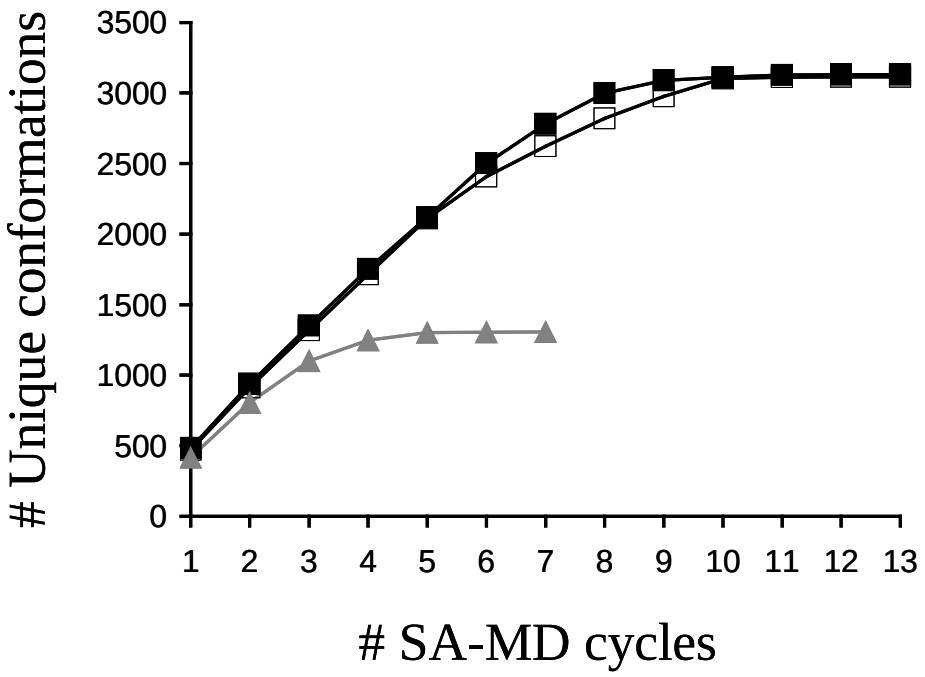


Figure 4.1.1: *Number of unique minimum energy conformations found for systems I (black line, solid squares), II (black line, empty squares) and III (grey line, solid triangles) against the number of modified SA*

Figure 4.1.2 shows the distribution of energies, which are relative to that of the global minimum, for the unique minimum energy conformations of I and II. As can be seen, the two profiles are very similar, with the highest energy bordering ~23 kcal/mol in both cases. The histograms displayed in Figure 4.1.3, which represent the distribution of the virtual dihedral angles used to define the conformation of the five CREKA residues (see Methods), permit a comparison of the similarity between the sets of unique minima detected for the free peptide in aqueous and ionized solutions. Thus, the incorporation of additional NaCl molecules into the aqueous medium does not alter, in general terms, the conformational profile of the free CREKA peptide. A more detailed comparison between the energy minima of systems I and II is provided in the Supporting Information section (Figure S1), which shows the distribution of the main chain dihedral angles φ,ψ of the five CREKA residues. It is worth noting that special attention has been paid to the comparison of the structures with a relative energy lower than 2.0 kcal/mol, *i.e.* 27 and 23 structures for I and II respectively, which

are expected to be the more significant ones. Accordingly, the values of φ,ψ for these accessible energy minima are highlighted in the Ramachandran maps of Figure S1 (see Supporting Information). As can be seen, a very notable resemblance between the results reached for the two systems is indicated again. Overall, these results allow us to conclude that the conformational profile of the peptide is not affected by the increase of the ionic concentration.

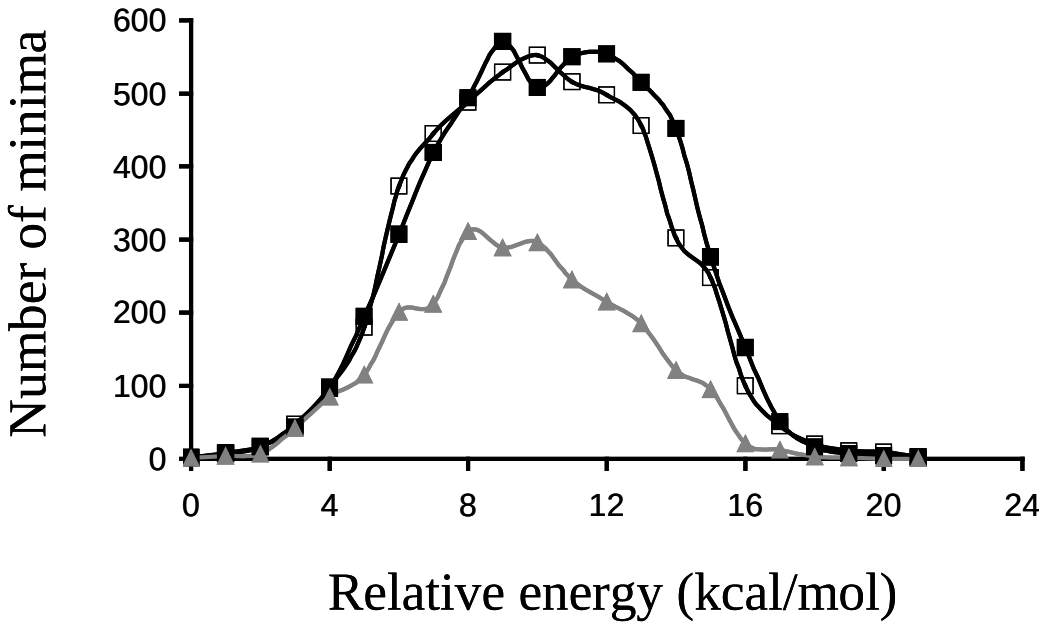


Figure 4.1.2: *Distribution of energies for the minimum energy conformations found for systems I (black line, solid squares), II (black line, empty squares) and III (grey line, solid triangles). In each case represented energies are relative to the corresponding global minimum.*

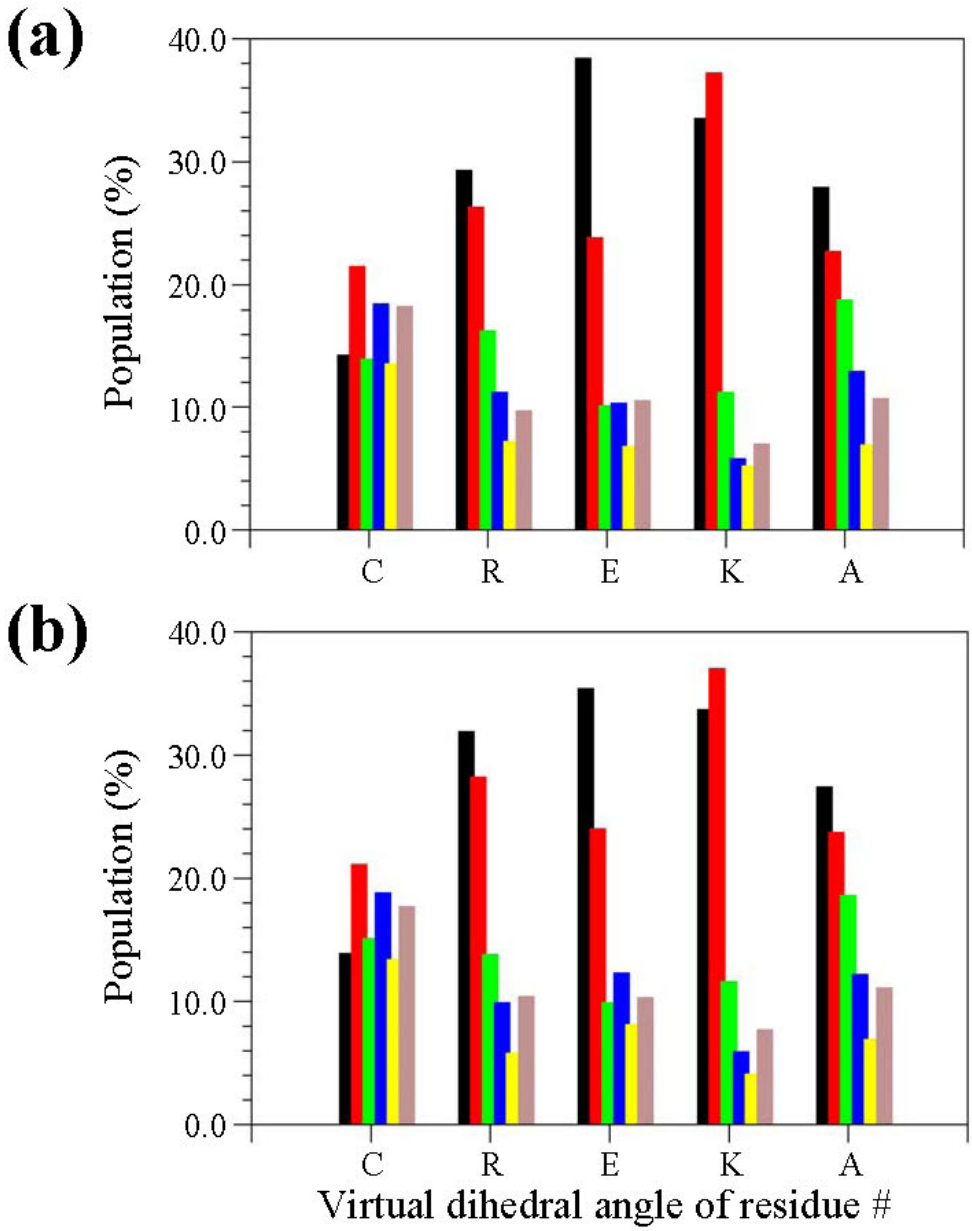


Figure 4.1.3: *Distribution of the virtual dihedral angles (see methods) used to define the conformation of five CREKA residues in all the unique minima obtained for systems I (a) and II (b). Colour code for the bars is: black for dihedral angle values ranging from 0º-to-60º, red from 60º-to-120º, green from 120º-to-180º, blue from 180º-to-240º, yellow from 240º-to-300º and grey from 300º-to-360º.*

Figure 4.1.4a shows the lowest energy minimum obtained for system I. In this structure the backbone amide groups form intramolecular hydrogen bonds leading to the formation of two γ-turns and one β-turn. Specifically, γ-turns (7-membered intramolecular hydrogen bonds) are formed between the (Ac)C=O···H-N(Arg) [$d_{H\cdots O}$= 1.999 Å, ∠N-H···O= 149.2º] and (Lys)C=O···H-N(NHMe) [$d_{H\cdots O}$= 1.828 Å, ∠N-H···O= 153.8º], where Ac and NHMe are the blocking groups at the N_t- and C_t-termini, respectively. The β-turn (10-membered intramolecular hydrogen bond) involves the Cys and Lys residues: (Cys)C=O···H-N(Lys) [$d_{H\cdots O}$= 1.917 Å, ∠N-H···O= 168.0º]. The view provided in Figure 4.1.4a of the global minimum suggests that the (Cys)C=O and H-N(Ala) moieties form a 13-membered intramolecular hydrogen bond, which is the interaction typically found α-helices. However, the poor geometric parameters found in this case, [$d_{H\cdots O}$= 2.375 Å, ∠N-H···O= 112.3º], indicates that it should considered as a simple dipole-dipole interaction. Inspection of the side chains reveals a multiple interaction pattern formed by salt bridges between the negatively charged carboxylate group of Glu and the positively charged side chains of Arg and Lys. Furthermore, the carboxylate group of Glu also forms a hydrogen bond with its own backbone N-H amide group, *i.e.* one oxygen atom of the carboxylate forms two interactions with the guanidium group of Arg while the second oxygen participates in one salt bridge with the Lys side chain and one intra-residue hydrogen bond (Figure 4.1.4a).

Figure 4.1.4b shows the superposition of the global minimum found for systems I and II. As can be seen, the two structures show a remarkable similarity in the backbone conformation, the only difference between them corresponding to the orientation of the Lys side chain. Specifically, in the global minimum of system II no salt bridge interaction between the Glu(C=O) and H-N(Lys) was detected. However, this small difference does not allow us to conclude that ionic concentration leads to structural changes in the accessible conformations.

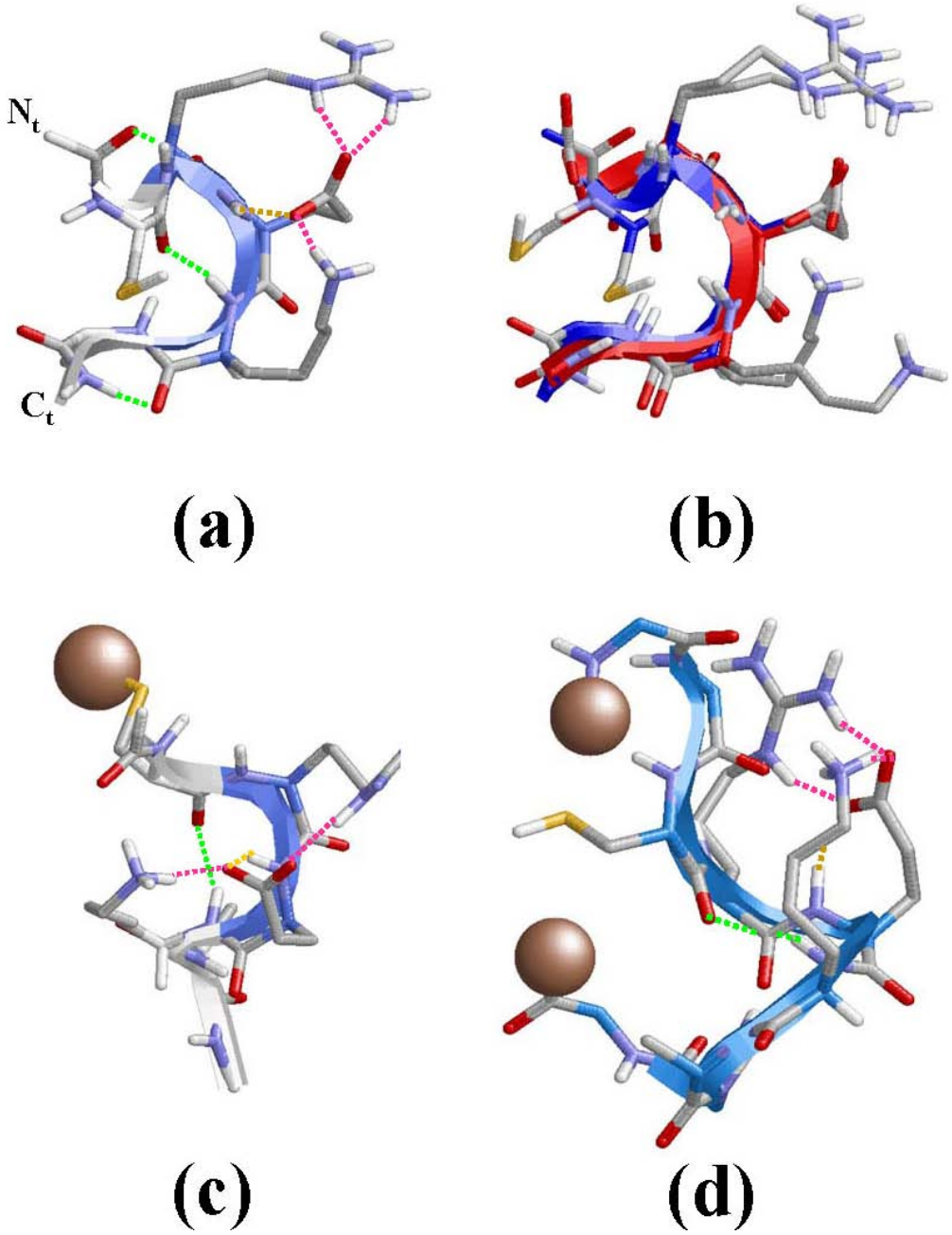


Figure 4.1.4: *(a) Lowest energy minimum obtained for free CREKA in unionized aqueous solution (system I). (b) Superposition of the global minimum found for systems I (blue) and II (green). (c) Lowest energy minimum obtained for CREKA attached to a SPIO nanoparticle (system III). (d) Lowest energy minimum obtained for GGCREKAGG inserted in a phage display protein (system IV; D= 20.0 Å). The surface used to mimic the nanoparticle in system III and the connections to the phage display protein in system IV have been represented using a single grey ball. Intramolecular hydrogen bonds between the backbone amide groups are displayed using green dashed lines, while side chain···side chain and side chain···backbone interactions are marked with pink (salt bridges) and orange (hydrogen bonds) dashed lines.*

A detailed analysis of the minima obtained for systems I and II below a relative energy threshold of 2 kcal/mol shows two common structural patterns for at least 90% of the relevant conformations: (i) a β-turn involving the Cys and Lys residues is systematically formed; and (ii) the ionized side chains of the three central residue are arranged to form a multiple interaction pattern, which is dominated predominantly by salt bridges.

Clustering analyses of systems I and II indicate that 69.5% (2165 structures) and 76.4% (2380 structures) minima present at least one intramolecular interaction (see Supporting Information). As can be seen, only 13 (I) and 19 (II) clusters contain more than 25 structures grouping 37.9% and 56.9% of the minima stabilized by intramolecular interactions, *i.e.* many clusters are constituted by a very small number of minimum. A more detailed analysis of the structures contained in the different clusters reveals that a total of 4243 and 4522 interactions were detected in the 2165 and 2380 minima of I and II, respectively, stabilized by intramolecular interactions, which is consistent with ~ 2 interactions per minimum. These interactions are distributed as follows:

(a) System I. Main chain⋯main chain hydrogen bonds: 84.7% (3593 interactions); main chain⋯side chain hydrogen bonds: 6.7% (285 interactions); and side chain⋯side chain salt bridges: 8.6% (365 interactions).

(b) System II. Main chain⋯main chain hydrogen bonds: 80.9% (3658 interactions); main chain⋯side chain hydrogen bonds: 5.7% (258 interactions); and side chain⋯side chain salt bridges: 13.4% (605 interactions).

Figure 4.1.5 compares the number of minimum energy conformations found for the more populated clusters of I and II. As can be seen, there is a quantitative agreement between the interaction patterns detected in the two systems. Thus, the more populated patterns are clearly conserved in the two systems. The overall of the results reflects a significant resemblance between the conformational profiles of I and II, which is also consistent with the small influence of the ionic concentration in the conformational profile of CREKA (see above).

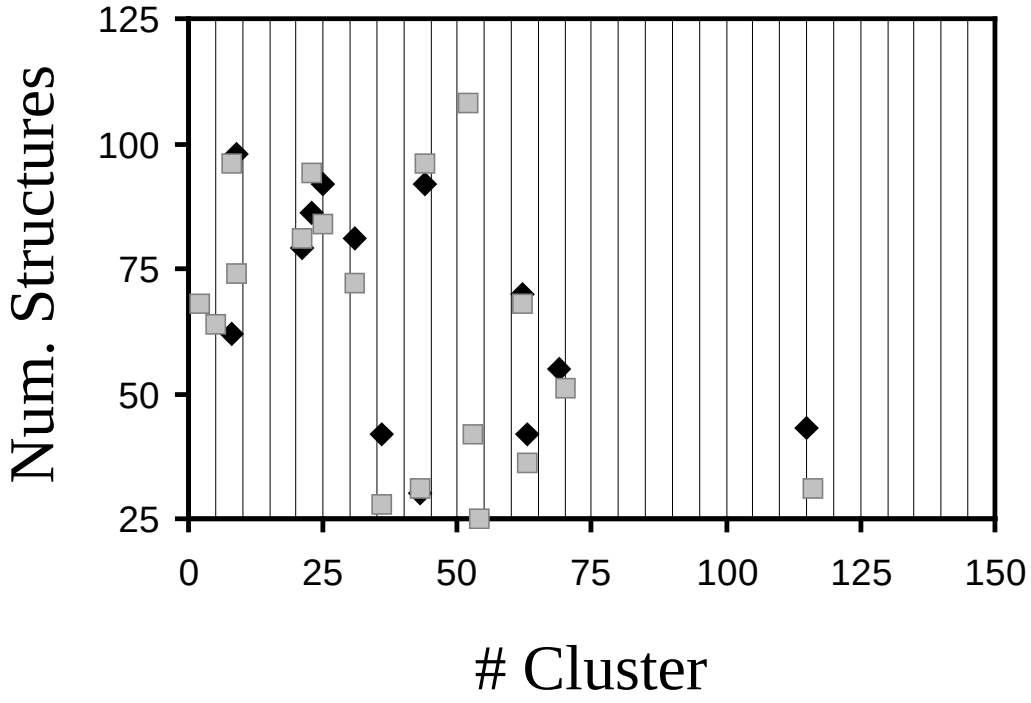


Figure 4.1.5: *Number of minimum energy conformations found for the more populated clusters, i.e. those containing more than 25 structures, of I and II.*

4.1.3.2 CREKA Attached to the Surface of a SPIO-Nanoparticle (System III).

After seven production cycles of modified SA-MD, which provided 1306 unique minima, the conformational search of system III fully converged. This is seen in Figure 4.1.1, which indicates that the last 3 cycles (minimization of 1500 selected structures) only provided 4 new unique conformations. The global minimum was obtained in the 2nd cycle of SA-MD reflecting the efficiency of the conformational search procedure. The number of unique minimum energy conformations is about 40% of those obtained for systems I and II indicating that, as expected, the conformational flexibility of the peptide becomes severely restricted when the sulfhydryl group of the Cys residue is attached to a surface. On the other hand, inspection of the distribution of the relative energies displayed in Figure 4.1.2 shows that the tethering of the peptide to the SPIO nanoparticle induces a significant reduction in both the height and width of the peak. The number of

minima with a relative energy smaller than 2 kcal/mol is only 12, which means that the number of significant minima decreases by about 50-55% with respect to the free CREKA peptide.

The distribution of the virtual dihedral angles used to define the conformation of CREKA attached to a SPIO nanoparticle is displayed in Figure 4.1.6. Comparison with the histograms obtained for free CREKA (Figure 4.1.3) reveals very interesting features. First, the conformation of the Cys residue is completely different in the free peptide as compared to that of system III. Thus, the linking of the sulfhydryl group to the surface used to represent the SPIO nanoparticle induces significant perturbations in the backbone conformation of the Cys. Second, and most importantly, the conformational distribution of the backbone for the remaining four residues is very similar in systems I, II and III. Thus, the surface of the nanoparticle does not alter the conformational preferences of the polar amino acids, which are dominated by the strong interactions between the ionized side chains. This interesting result is fully consistent with the binding to plasma clot of both free CREKA and CREKA-SPIO nanoparticle found by Ruoslahti and coworkers.[1-3] The resemblance between the conformational preferences of the free peptide and system III is also reflected in Figure S2 (see Supporting Information), which shows the distribution of the dihedral angles φ,ψ for the five residues for the latter system. As can be seen, the regions of the Ramachandran map visited by the unique minimum energy conformations are very similar to those displayed in Figure S1. In addition, the tethering to the surface does not influence the positions of the more significant minima of CREKA, which are highlighted in the same Figure. Thus, with the exception of the Cys residue, the positions of the accessible minima for the other four residues is similar in Figures 4.1.4 and 4.1.8, even though the dispersion is significantly lower in the latter. The latter feature is a consequence of the reduction of the conformational flexibility induced by the linking to the nanoparticle.

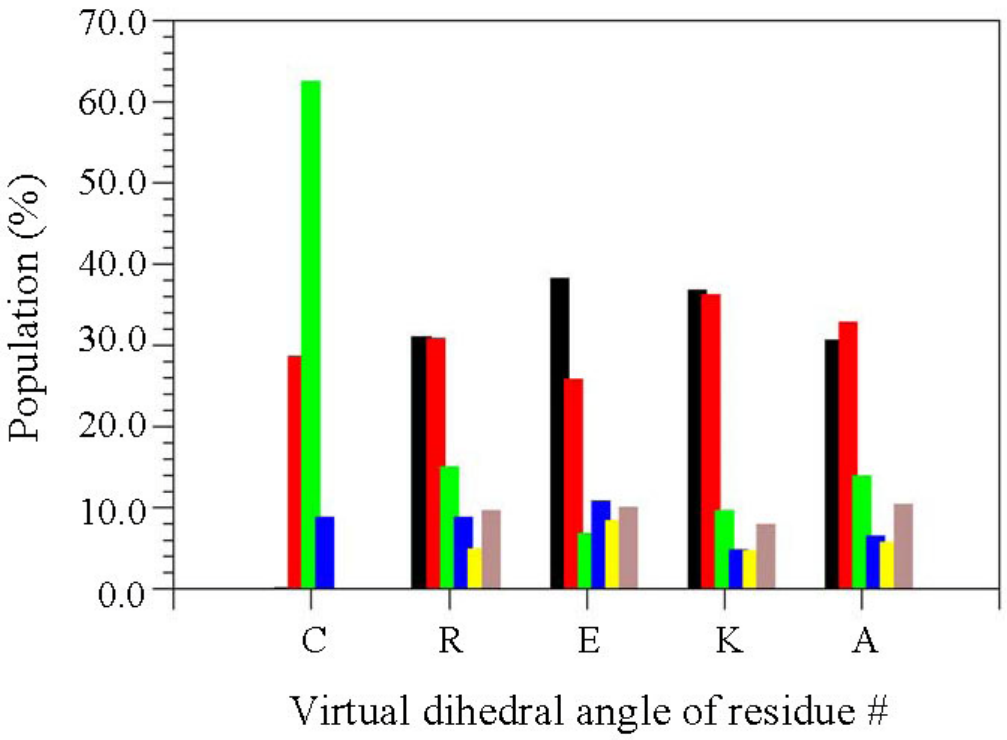


Figure 4.1.6: *Distribution of the virtual dihedral angles (methods) used to define the conformation of five CREKA residues in all the unique minima obtained for system III. Colour code for the bars is identical to that of Figure 4.1.3.*

The global minimum found for system III is depicted in Figure 4.1.4c. As can be seen, there is a notable resemblance between the global minimum of the three systems. In spite of this, there is a critical difference between the most stable structure of CREKA attached to the nanoparticle and that obtained for the free peptide: the two γ-turns found at the ends of the latter system are not detected when the peptide is tethered to the surface. Thus, the only backbone···backbone intramolecular interaction detected in the global minimum of system III corresponds to a β-turn, which involves (Cys)C=O···H-N(Lys) [$d_{H\cdots O}$= 2.048 Å, ∠N-H···O= 147.0º]. This structural motif, which is topologically identical to that obtained for I and II, is also detected in many of the minima with relative energies below 2 kcal/mol. On the other hand, the carboxylate group of the Glu is involved in a multiple interaction pattern similar to that described above for the free

peptide. The only difference with respect to the latter is that the salt bridge between the oxygen atom of the carboxylate group and the guanidinium group of Arg involves two centers rather than three. Accordingly, one oxygen atom of the carboxylate forms a salt bridge with the Arg while the second oxygen participates in one salt bridge and one hydrogen bond (Figure 4.1.4c). Furthermore, multiple interaction patterns involving the ionized side chains of Arg, Glu and Lys were also found in the 12 significant minima, the main difference among them being the number of centers involved in the salt bridges.

Clustering analysis of system III indicates that 82.1% (1072 structures) of the minimum energy conformations present at least one intramolecular interaction (see Supporting Information). Interestingly, only 8 clusters contain more than 25 structures grouping 25.5% of the minima stabilized by intramolecular interactions. A more detailed analysis of the structures contained in the different clusters reveals that a total of 2198 interactions were detected in the 1072 minima of III stabilized by intramolecular interactions, *i.e.* ~ 2 interactions per minimum. These interactions are distributed as follows: (i) main chain⋯main chain hydrogen bonds, 83.6% (1839 interactions); (ii) main chain⋯side chain hydrogen bonds, 4.7% (102 interactions); and side chain⋯side chain salt bridges, 11.7% (257 interactions). It is worth noting that these results, *i.e.* both the average number of interactions per minima and the distribution of the interactions, are fully consistent with those discussed above for I and II.

4.1.3.3 CREKA Inserted in a Phage Display Protein (System IV).

After four cycles of modified SA-MD for each value of D (minimization of 2000 selected structures), we observed that the energies of the minima obtained for system IV are considerably higher when D= 5.0, 7.0, 10.0 and 15.0 Å. Thus, a significant strain is induced in the peptide if the distance between the two ending particles is too short. The reduction of the strain of the peptide when the distance between the particles that tether CREKA to the phage display protein increases is clearly reflected in Figure 4.1.7, which depicts the lowest energy minimum for each such D value. For the unique minima obtained for D= 20.0 Å, the corresponding energies show reasonable values although still higher than those of systems I, II and III. This led us to focus on the analysis of the 527 resulting

unique minimum energy conformations obtained using the latter distance. However, due to the extra energy penalty observed for these structures with respect to those of the previously discussed systems, no additional modified SA-MD cycle was performed. Despite this shortcoming, the high efficiency shown by this conformational search procedure combined with the restricted conformational freedom of system IV, which is imposed by the restraint at the ends, suggests that the number of minima provided in four cycles is enough to provide qualitative information about the influence of the phage display protein on the conformational preferences of CREKA. On the other hand, the lack of experimental information about the structure of the loops of phage display proteins precludes a more realistic modeling of system IV.

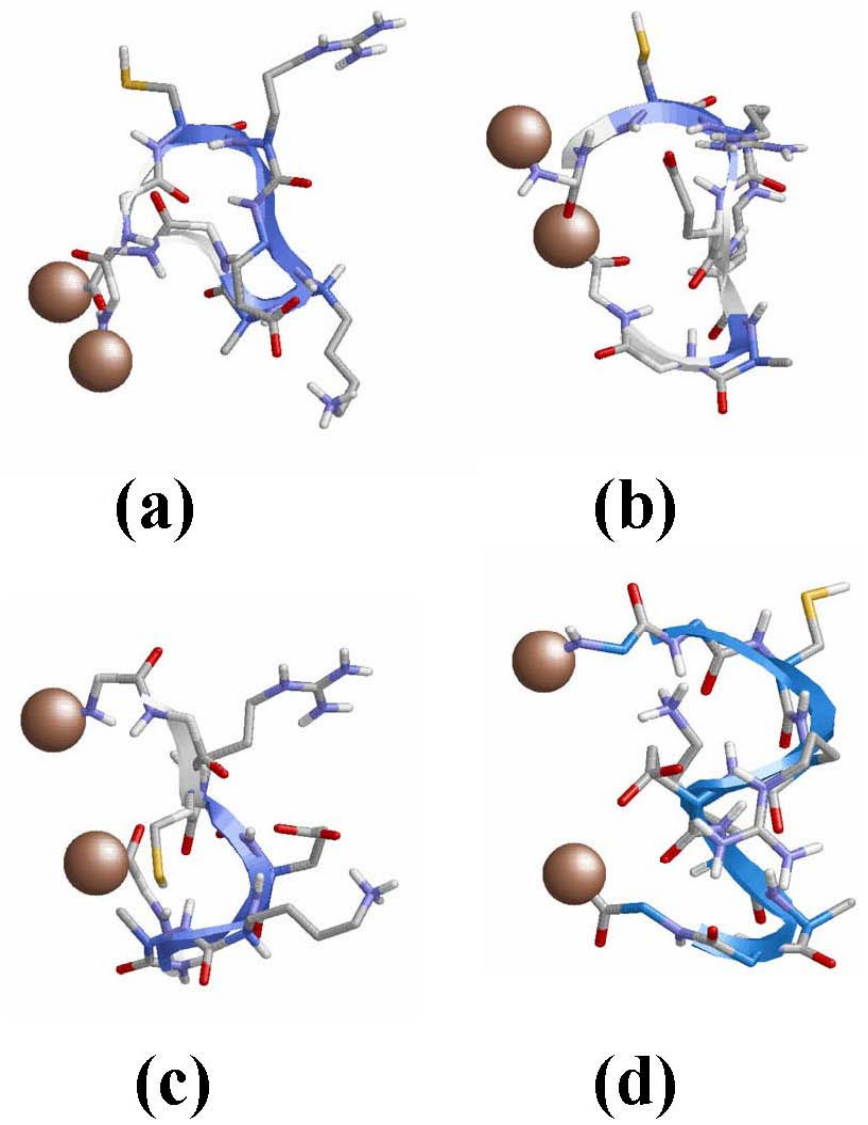


Figure 4.1.7: *Lowest energy minimum obtained for GGCREKAGG attached to two particles (system IV), which represents the tethering to a phage display protein, separated by D= 5 (a), 7 (b), 10 (c) and 15 Å (d).*

The distribution of the backbone dihedral angles φ,ψ for the CREKA peptide obtained using D= 20.0 Å is provided in the Supporting Information section (Figure S3), omitting the analysis of the four Gly residues included in system IV. As can be seen, the Ramachandran maps are very similar to those reported for the free peptide (Figure S1) and the peptide attached to a SPIO-nanoparticle (Figure S2). This coincidence, which also extends to the accessible energy minima, allows us to conclude that the chemical environment does not influence the conformational preferences of CREKA. Thus, analysis of the unique energy minima obtained for the four systems under study provides evidence that the conformational profile of this peptide is mainly characterized by both the β-turn motif and the interactions that form the side chains of the three charged residues.

Figure 4.1.4d depicts the lowest energy minimum obtained for system IV with D= 20.0 Å. Again, this structure is dominated by the (Cys)C=O···H-N(Lys) [$d_{H\cdots O}$= 2.226 Å, ∠N-H···O= 152.1º] β-turn and the multiple interaction that involve the ionized side chains of Arg, Glu and Lys. Furthermore, the intraresidue hydrogen bond between the amide N-H and the oxygen atom of the carboxylate group of Glu is also detected in this conformation. The noticeable agreement between the lowest energy minimum of systems I, II, III and IV is graphically displayed in Figure 4.1.8, which provides the superposition of all structures.

Clustering analysis of system IV reveals that 75.0% (395 structures) minima show one or more intramolecular interactions (see Supporting Information). Analysis of these minima evidence a total of 715 interactions distributed in such 395 structures, *i.e.* ~ 2 interactions per minimum, which can be categorized as follows: (i) main chain···main chain hydrogen bonds, 89.0% (636 interactions); (ii) main chain···side chain hydrogen bonds, 2.2% (16 interactions); and side chain···side chain salt bridges, 8.8% (63 interactions). The overall of these results are in excellent agreement with those reported for I, II and III.

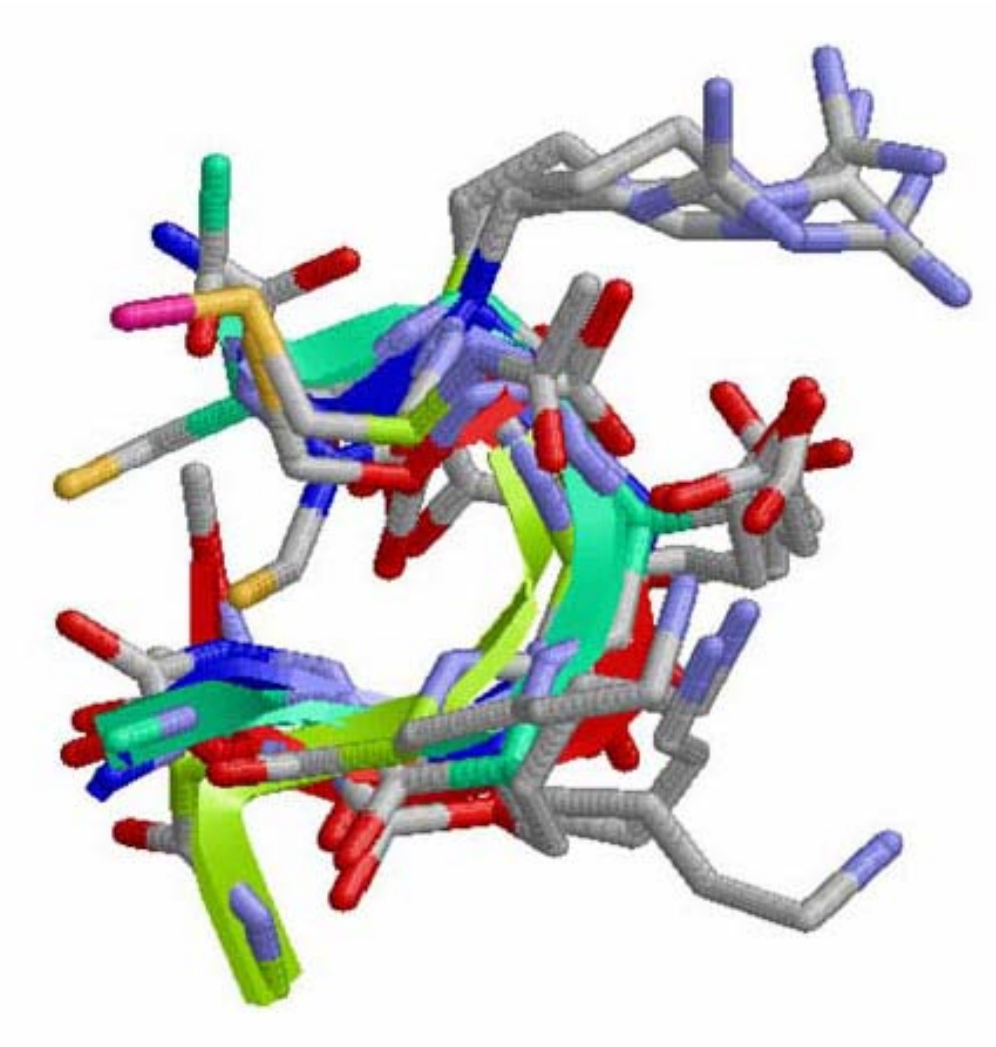


Figure 4.1.8: *Superposition of the lowest energy minimum obtained for systems I, II, III and IV (D= 20.0 Å). The CREKA peptide is the only represented, the other elements included in simulations of systems III and IV being omitted for clarity.*

4.1.3.4 Stability of the β-Turn Motif in CREKA.

Results obtained in previous sections indicate that the β-turn motif is significantly favored for the CREKA peptide sequence. Thus, this conformation was found to be present in many of the minima found for systems I-IV evidencing its accessibility. In this section we demonstrate that the β-turn motif is not only an accessible conformation but also a stable one. For this purpose, *NPT* MD simulations were performed in aqueous solution at 310 K (physiological temperature) and P= 1 atm using as starting points three of the minima found for I.

Specifically, we considered two different minima arranged in a β-turn conformation, hereafter denoted Ia and Ib, and one minimum with a helical-like conformation, Ic. These three minima are schematically depicted in Figure 4.1.9a. As can be seen, minimum Ia corresponds to the lowest energy minimum of system I (Figure 4.1.4a).

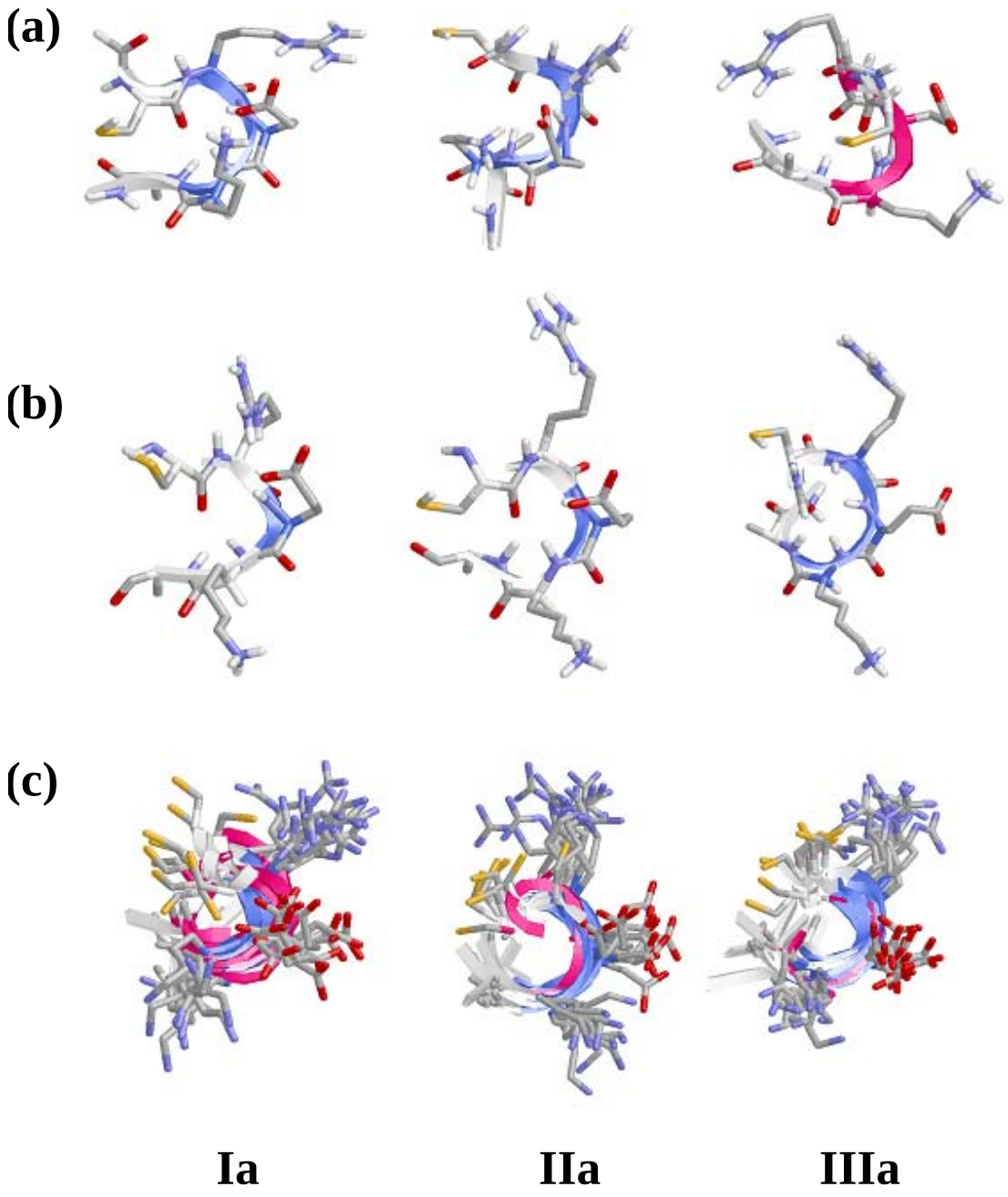


Figure 4.1.9: *(a) Minimum energy conformations (labelled as Ia, Ib and Ic) of free CREKA in unionized aqueous solution (system I) selected for MD studies at 310 K (physiological temperature). (b) Structure of Ia, Ib and Ic after the equilibration period. (c) Superimposition of the snapshots recorded at intervals of 1 ns during the MD simulations of Ia, Ib and Ic.*

Before the production series, the thermodynamic variables of the system were equilibrated using the protocol indicated in the Methods section. Specifically, consecutive rounds of short MD runs were performed as follows: (i) 0.5 ns of NVT-MD at 500 K were used to homogeneously distribute the solvent and ions in the box; (ii) 0.5 ns of NVT-MD at 298 K (thermal equilibration); and (iii) 0.5 ns of NPT-MD at 298 K (density relaxation) were run. The conformations obtained for Ia, Ib and Ic after equilibration are displayed in Figure 4.1.9b. It is worth noting that both dynamical and thermal effects induce some re-arrangements in the three minima, even although the more important motifs of the three initial conformations are retained, especially of Ia and Ib. After equilibration, a NPT-MD in aqueous solution was run during 10 ns for each structure.

Figure 4.1.10a compares the root mean square deviation (RMSD) of the three minima, while Figure 4.1.10b displays the root mean square fluctuation (RMSF) of individual residues averaged over the whole simulation. Both RMSD and RMSF were computed with respect to the backbone atoms (-N-C^{α}-C-). Inspection to the RMSD, which were calculated without considering the Ac and NHMe capping groups, indicates that the three minima are very stable, the average RMSD being around 1.5 Å in all cases. Similarly, the RMSF profiles clearly show that the only distortions are located at the capping groups. This feature is also reflected in Figure 4.1.9c, which shows the superimposed snapshots that were recorded every 1 ns. These results indicate that the more main structural characteristics of these accessible minima are retained through the whole trajectories evidencing the high stability of the β-turn motif.

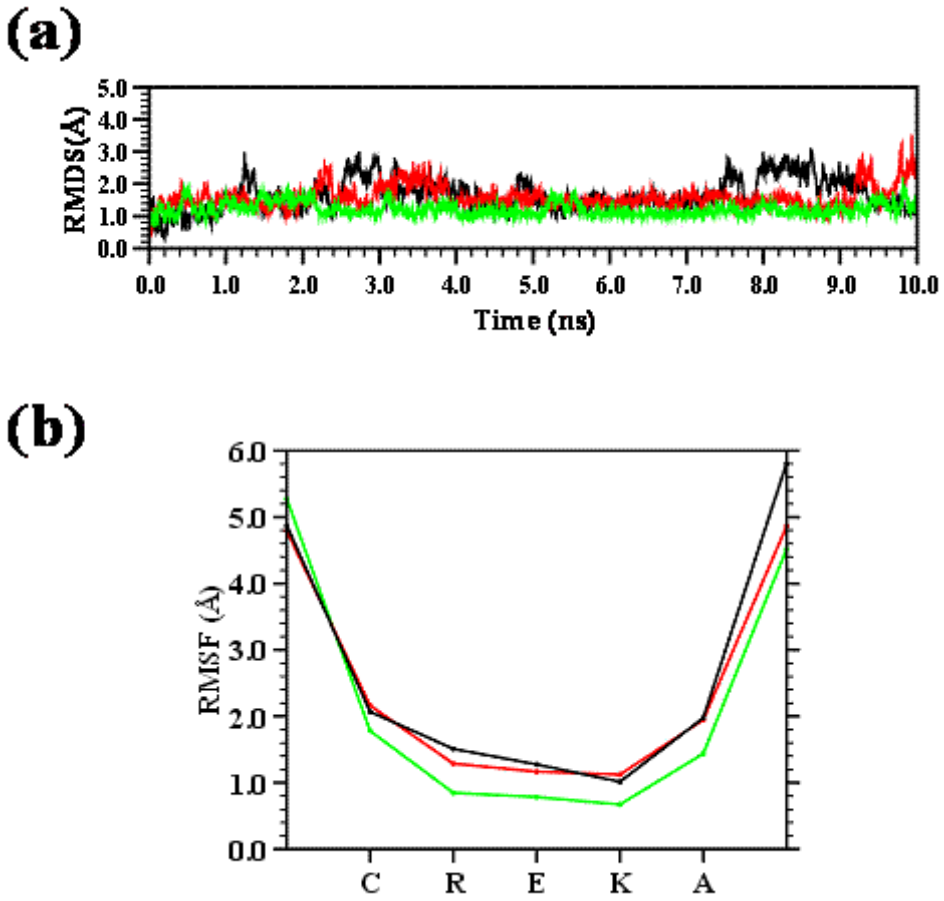


Figure 4.1.10: *(a) Temporal evolution of the backbone RMSD and (b) RMSF of the minimum energy conformations Ia (black), Ib (green) and Ic (red) of the CREKA peptide.*

4.1.4 Discussion and Conclusions

Considering all the results reported in this work, it is clear that the modified SA-MD is a powerful method to explore the conformational space of peptides. Complete convergence was reached for systems I, II and III after a few simulation cycles. The limited influence exerted, in this case, by the NaCl molecules used to mimic the ionic strength of the physiological medium was confirmed by comparing the conformational profiles of systems I and II. Unique structures produced by independent searches for these two systems reveal a strong degree of

similarity, as was also evidence by the clustering analyses. Accordingly, these results demonstrate that simulations in unionized aqueous solution are able to represent the behavior of CREKA in the ionized physiological medium.

The conformational space of both the free peptide and the peptide attached to a nanoparticle has been explored exhaustively. Results indicate that the conformational profile of the REKA sequence is very similar in both cases, the only difference being the Cys residue. This coincidence is particularly evident for the accessible minima, as can be deduced from the Figures 4.1.4 and 4.1.8. Thus, comparison between all the minimum energy conformations with relative energies lower than 2.0 kcal/mol reveals a high degree of concordance. Furthermore, complete clustering analyses also reflect to a significant agreement, as is demonstrated by the similarities in the average number of interactions per minimum and the distribution of hydrogen bonding interactions and salt bridges. The overall of these results allows us to conclude that the conformational profile of the peptide is independent of the chemical environment, *i.e.* the conformational preferences of the peptide are very similar when it is free or attached to a SPIO nanoparticle. This agreement is expected to facilitate the future design of synthetic analogs to protect the peptide from the attack of proteases since the choice of the synthetic amino acid will be based on a general conformational profile rather than on a specific conformation. It should be noted that the conformational preferences of synthetic amino acids are typically controlled by the introduction of chemical restraints, *e.g.* $C^{\alpha}\leftrightarrow C^{\alpha}$ cyclization and/or the introduction of side groups. Therefore, in recent years it has been demonstrated that synthetic amino acids with restrained and well-defined conformational preferences can be designed for a given application.[13,15,28]

On the other hand, the pattern of multiple interactions which involves salt bridges and hydrogen bonds, defined by the ionized side chains of Arg, Glu and Lys determines the backbone conformation of the peptide. This is clearly reflected in the lowest energy minimum, which is very similar for both the free CREKA and the CREKA-SPIO nanoparticle. Thus, the β-turn motif that characterizes the global minimum and many of the accessible minima is a consequence of such interactions. Moreover, the ionized side chains of Arg, Glu and Lys are probably essential for the binding activity of CREKA. Thus, peptide sequences used to recognize tumors are frequently formed by charged amino acids.[29] Therefore, the

multiple interaction pattern mentioned above together with the backbone β-turn should be preserved in the engineered CREKA synthetic analogs.

Calculations on system IV have shown that it is not easy to mimic peptide sequences expressed in phage display proteins. Thus, the lack of available structural and chemical information about such proteins makes their simulation extremely difficult. Our results indicate that, when the dimensions used mimic the loop that contains the expressed peptide are too short, the produced minimum energy conformations are significantly strained. Accordingly, the results obtained for system IV should be considered with caution. In spite of these limitations, the results are fully consistent with those obtained for the free peptide and the peptide attached to a nanoparticle when the separation between the two extremes of the loop that contains CREKA is 20.0 Å. This separation provides the lowest strain, even though an unfavorable energy penalty is still involved in the resulting minima. However, the good agreement with the conformational profiles of systems I, II and III supports the conclusion reached above: the conformational preferences of CREKA are not determined by the chemical environment but by the strong interactions between the charged residues.

The coincidence between the global minimum of the four studied systems, which is reflected in Figure 4.1.8, together with the analysis of the accessible conformations suggest that the bioactive conformation of CREKA involves both a β-turn motif and strong interactions involving the side chains of Arg, Glu and Lys. Furthermore, MD simulations show that this is a very stable conformation. Thus, overall, the results indicate that the shape of the global minimum is appropriate to form intermolecular interactions with a receptor, *i.e.* it is like a pocket with the charged groups pointing outwards. Accordingly, the future design of CREKA synthetic analogs should take care to retain these elements to preserve the peptide binding activity.

4.1.5 References

1. Pasqualini, R.; Ruoslahti, E. Organ Targeting in Vivo Using Phage Display Peptide Libraries. *Nature* **1996**, *380*, 364-366.
2. Hoffman, J. A.; Giraudo, E.; Singh, M.; Zhang, L.; Inoue, M.; Porkka, K.; Hanahan, D.; Ruoslathi, E. Progressive Vascular Changes in a Transgenic Mouse Model of Squamous Cell Carcinoma. *Cancer Cell* **2003**, *4*, 383-391.
3. Ruoslahti, E. Personal communication (unpublished results) 2007.
4. Simberg, D.; Duza, T.; Park, J. H.; Essier, M.; Pilch, J.; Zhang, L.; Derfus, A. M.; Yang, M.; Hoffman, R. M.; Bathia, S.; Sailor, M. J.; Ruoslahti, E. Biomimetic Amplification of Nanoparticle Homing to Tumors. *Proc. Natl. Acad. Sci.* **2007**, *104*, 932-936.
5. Adessi, C.; Soto, C. Converting a Peptide into a Drug: Strategies to improve Stability and Bioavailability. *Curr. Med. Chem.* **2002**, *9*, 963-978.
6. Lelais, G.; Seebach, D. Beta(2)-Amino Acids - Syntheses, occurrence in Natural Products, and Components of Bbeta-Peptides. *Biopolymers* **2004**, *76*, 206-243.
7. Yamaguchi, H.; Kodama, H.; Osada, S.; Kato, F.; Jelokhani-Niaraki, M.; Kondo, M. Effect of alpha, alpha-dialkyl Amino Acids on the Protease Resistance of Peptides**.** *Biosci. Biotech. Biochem.* **2003**, *67*, 2269-2272.
8. Sadowsky, J. D.; Murray, J. K.; Tomita, Y.; Gellman, S. H. Exploration of backbone space in foldamers containing alpha- and beta-amino acid residues: Developing protease-resistant oligomers that bind tightly to the BH3-recognition cleft of Bcl-x(L) *Chem. Bio. Chem.* **2007**, *8*, 903-916.
9. Talele, T. T.; McLaughlin, M. L. Asymmetric syntheses of enantiomerically pure alpha,alpha-dialkylated glycines as core structural units for novel HIV-1 protease inhibitors *Biopolymers* **2003**, *71*, 344-344.
10. Banerjee, R.; Basu, G.; Chene, P.; Roy, S. Aib-based Peptide Backbone as Scaffolds for Helical Peptide Mimics. *J. Pept. Res.* **2002**, *60*, 88-94.

11. Webb, A. I.; Dunstone, M. A.; Williamson, N. A.; Price, J. D.; de Kauwe, A.; Chen, W. S.; Oakley, A.; Perlmutter, P.; McCluskey, J.; Aguilar, M. I.; Rossjohn, J.; Purcell, A. W. T Cell determinants incorporating Beta-Amino Acid residues are Protease resistant and remain Immunogenic in Vivo. *J. Immun.* **2005**, *175*, 3810-3818.
12. Bannwarth, L.; Kessler, A.; Pethe, S.; Collinet, B.; Merabet, N.; Boggetto, N.; Sicsic, S.; Reboud-Ravaux, M.; Ongeri, S. Molecular Tongs containing Amino Acid Mimetic Fragments: New Inhibitors of Wild-Type and Mutated HIV-1 Protease Dimerization. *J. Med. Chem.* **2006**, *49*, 4657-4664.
13. Alemán, C.; Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Haspel, N.; Zheng, J.; Casanovas, J.; Wolfson, H.; Nussinov, R. Concepts and Schemes for the Re-Engineering of Physical Protein Modules: generating Nanodevices via targeted replacements with Constrained Amino Acids. *Phys. Biol.* **2006**, *3*, S54-S62.
14. Tsai, C. J.; Zheng, J.; Zanuy, D.; Haspel, N.; Wolfson, H.; Alemán, C.; Nussinov, R. Principles of Nanostructure Design with Protein Building Blocks**.** *Proteins* **2007**, *68*, 1-12.
15. Zheng, J.; Zanuy, D.; Haspel, N.; Tsai, C.-J.; Alemán, C.; Nussinov, R. Nanostructure Design using Protein Building Blocks enhanced by Conformationally Constrained Synthetic Residues. *Biochemistry* **2007**, *46*, 1205-1218.
16. Agrafiotis, D. M.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E. Conformational Sampling of Bioactive Molecules: A Comparative Study. *J. Chem. Inf. Model.* **2007**, *47*, 1067-1086.
17. Steinbach, P. J. Exploring Peptide Energy Landscapes: A test of Force Fields and Implicit Solvent Models. *Proteins* **2004**, *57*, 665-677.
18. Swaminathan, P.; Hariharan, M.; Murali, R.; Singh, C. U. Molecular Structure, Conformational Analysis, and Structure-Activity studies of Dendrotoxin and its Homologues using Molecular Mechanics and Molecular Dynamics Techniques. *J. Med. Chem.* **1996**, *39*, 2141-2155.
19. Kirkpatrick, S.; Gelatt Jr.; C.D. Vecchi, M.P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671-680.

20. Steinbach, P. J.; Brooks, B. R. Protein Simulation below the Glass-Transition Temperature – Dependence on Cooling Protocol. *Chem. Phys. Letters* **1994**, *226*, 447-452.
21. Baysal, C.; Meirovitch, H. Efficiency of Simulated Annealing for Peptides with Increasing Geometrical Restrictions. *J. Comput. Chem.* **1999**, *20*, 1659-1670.
22. Simmerling, C.; Elber, R. Hydrophobic Collapse in a Cyclic Hexapeptide - Computer-Simulations of Chdlfc and Caaaac in Water. *J. Am. Chem. Soc.* **1994**, *116*, 2534-2547.
23. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.
24. Wang, J.; Cieplak, P.; Kollman, P.A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules?. *J. Comput. Chem.* **2000**, *21*, 1049-1074.
25. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
26. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
27. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints – Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.
28. Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. Use of Constrained Synthetic Amino acids in β-Helix Proteins for Conformational Control. *J. Phys. Chem. B* **2007**, *111*, 3236-3242.
29. Zhang, L.; Giraudo, E.; Hoffman, J. A.; Hanahan, D.; Ruoslahti, E. Lymphatic Zip Codes in Premalignant Lesions and Tumors. *Cancer Res.* **2006**, *66*, 5696- 5706.

4.2 In silico molecular engineering for a targeted replacement in a tumor-homing peptide

A new amino acid has been designed as a replacement for arginine (Arg, R) to protect the tumor-homing pentapeptide CREKA from proteases. This amino acid, denoted (Pro)hArg, is characterized by a proline skeleton bearing a specifically oriented guanidinium side chain. This residue combines the ability of Pro to induce turn-like conformations with the Arg side-chain functionality. The conformational profile of the CREKA analogue incorporating this Arg substitute has been investigated by a combination of simulated annealing and Molecular Dynamics. Comparison of the results with those previously obtained for the natural CREKA shows that (Pro)hArg significantly reduces the conformational flexibility of the peptide. Although some changes are observed in the backbone···backbone and side chain···side chain interactions, the modified peptide exhibits a strong tendency to accommodate turn conformations centered at the (Pro)hArg residue and the overall shape of the molecule in the lowest energy conformations characterized for the natural and the modified peptide exhibit a high degree of similarity. In particular, the turn orients the backbone such that the Arg, Glu and Lys side chains face the same side of the molecule, which is considered essential for bioactivity. These results suggest that replacement of Arg by (Pro)hArg in CREKA may be useful in providing resistance against proteolytic enzymes while retaining conformational features which are essential for tumor-homing activity.[*]

4.2.1 Introduction

The linear pentapeptide CREKA (Cys-Arg-Glu-Lys-Ala) was recently discovered[1] by *in vivo* screening of phage-display peptide libraries[2] for tumor-homing in tumor-bearing MMTV-PyMT transgenic breast cancer mice.[3] After intravenous injection of synthetic CREKA, the peptide was detected in human tumors; but not in normal tissues.[4] *In vivo* experiments revealed that this tumor-homing pentapeptide binds to clotted plasma proteins, thus establishing its behavior as a clot-binding peptide. In addition, experiments carried out with

[*] *Submitted for Publication.*

CREKA linked to amino dextran-coated iron oxide (SPIO) nanoparticles indicated that these systems not only bind to blood and plasma clots but also effectively induce further localized tumor clotting, thus amplifying the nanoparticle homing.[1] It has been established that the chemical nature of the nanoparticle is not essential for this activity since both CREKA-SPIO nanoparticles and CREKA-coated liposomes were found to cause clotting in tumor vessels.[1]

In a recent study,[5] we determined the energy landscape and proposed the bioactive conformation of CREKA by using a multiple conformational search strategy based on Molecular Dynamics (MD) simulations. Two experimentally tested environments were considered: (*i*) the free peptide immersed in either a dilute aqueous solution or an ionized aqueous solution mimicking the physiological medium; (*ii*) the peptide attached to a nanoparticle through the sulfhydryl group of cysteine. In addition, we considered a molecular system that mimics the CREKA peptide when inserted in a phage display protein, although in this case the search strategy did not converge completely and the conformational space was not explored exhaustively. Results indicated[5] that the conformational profile of the REKA sequence is very similar in all cases; significant differences were observed only for the Cys residue. Complete clustering analyses based on the number and distribution of intramolecular interactions showed that such coincidence was particularly remarkable for the accessible minima, *i.e.* those with relative energies lower than 2.0 kcal/mol.

Despite the high potential of CREKA in cancer diagnostics and therapeutics, the application of this and other tumor-homing peptide sequences may be hampered by short half-life times since endogenous proteases rapidly digest such peptides. Protection from proteolytic cleavage is therefore crucial for tumor-homing peptide applications, similar to most peptides exhibiting biological activity. Several strategies have been proposed for this purpose, targeted replacements with non-coded amino acids being among the most successful ones.[6–13]

Within the framework of a project aimed at improving the biological performance and pharmacological profile of CREKA, we have undertaken the *in silico* study of a CREKA analogue incorporating a non-proteinogenic amino acid. This residue has been conceived to retain the most relevant characteristics of the

conformational profile of the natural peptide and simultaneously impart stability against proteolytic cleavage. One should bear in mind that any possible analogue able to overcome the limitations of CREKA while retaining its tumor-homing activity should exhibit a conformational profile with low-energy regions close to those characterizing the natural peptide.[5] Computer simulations based on the application of theoretical methodologies have proven to be powerful techniques not only for sampling the conformational space of biomolecules[5,14–16] as small peptides, but also for designing, engineering and testing non-proteinogenic amino acids with potential applications in nanobiology.[17–19]

In this work, we report the explicit design and subsequent conformational study of a CREKA analogue generated by replacing an amino acid in the sequence by a non-proteinogenic counterpart. The work is divided into three parts. First, the new amino acid is engineered. Then, force-field parameters for this non-coded residue are determined using quantum mechanical calculations. Finally, the conformational profile of the modified CREKA sequence attached to a nanoparticle is examined using as the sampling technique a computational procedure that combines a modified simulated annealing[20,21] with molecular dynamics (SA-MD). A detailed comparison between the conformational properties exhibited by the natural and modified CREKA sequences is provided.

4.2.2 Computational Methods

4.2.2.1 Quantum mechanical calculations

Density Functional Theory (DFT) methods were applied for quantum mechanical calculations, which were carried out using the Gaussian 03 computer program.[22] Specifically, calculations were performed by combining the Becke´s three-parameter hybrid functional (B3)[23] with the Lee, Yang and Parr (LYP)[24] expression for the nonlocal correlation (B3LYP). Accordingly, all the quantum mechanical calculations presented in this work were performed using the unrestricted formalism of the B3LYP method combined with the 6-31+G(d,p) basis set.[25] Frequency analyses were carried out to verify the nature of the minimum state of all the stationary points provided by geometry optimizations

and to calculate the zero-point vibrational energies (ZPVE) with both thermal and entropic corrections, the latter statistical terms being used to compute the conformational Gibbs free energies in the gas phase (ΔG^{gp}).

To obtain an estimation of the solvation effects on the relative stability of the different minima, single point calculations were also conducted using a Self-Consistent Reaction Field (SCRF) model. SCRF methods treat the solute at the quantum mechanical level, while the solvent is represented as a dielectric continuum. Specifically, the Polarizable Continuum Model (PCM) developed by Tomasi and co-workers was used to describe the bulk solvent.[26] This method involves the generation of a solvent cavity from spheres centered at each atom in the molecule and the calculation of virtual point charges on the cavity surface representing the polarization of the solvent. The magnitude of these charges is proportional to the derivative of the solute electrostatic potential at each point calculated from the molecular wave function. The point charges may then be included in the one-electron Hamiltonian, thus inducing polarization of the solute. An iterative calculation is carried out until the wave function and the surface charges are self-consistent. PCM calculations were performed using the standard protocol and considering the dielectric constants of carbon tetrachloride (ε = 2.228), chloroform (ε = 4.9) and water (ε = 78.4). The conformational free energies in solution (ΔG^{sol}, where *sol* refers to the solvent) were computed using the classical thermodynamics scheme, that is, the free energies of solvation provided by the PCM model were added to the ΔG^{gp} values.

4.2.2.2 Conformational search and force-field calculations.

The conformational preferences of the CREKA analogue studied in this work were explored using the sampling strategy previously employed for the natural peptide.[5] This procedure is based on the minimization of structures generated at the initial and intermediate states of several SA-MD cycles. The starting temperature is gradually reduced during the simulation, thus allowing the system to surmount energy barriers. In spite of this, in practice, this sampling technique does not always lead the system to the most stable region at the end of the simulation. However, recent studies showed that very low energies are obtained by minimizing the structures extracted from SA-MD processes.[27,28] According to

our previous study,[5] this sampling technique, which was found to be robust enough to obtain low-energy structures that may be quasi-degenerate with the global minimum but situated in different valleys of the peptide landscape, was applied by submitting 500 structures to energy minimization from each SA-MD cycle.

The simulated system consisted of the CREKA analogue attached to a surface through the sulfhydryl group of the Cys residue. The *N*- and *C*-termini of the peptide backbone were capped with acetyl (Ac, MeCO–) and methylamide (–NHMe) groups, respectively. The surface was formed by 100 spherical particles distributed in a 10×10 square (47.5 Å^2), with van der Waals parameters R = 2.35 Å and ε = 0.90 kcal/mol and no electric charge. This system is identical to that considered for natural CREKA in our previous work[5] and mimics the experimental conditions[1,4] in which the peptide is bound to the surface of a nanoparticle. Therefore, the results obtained in this work for the CREKA analogue are compared with those reported for natural CREKA attached to an identical surface (system III in ref. 5), unless otherwise indicated.

The CREKA analogue attached to the surface was placed in the center of a cubic simulation box (a = 47.5Å) filled with 3405 explicit water molecules, which were represented using the TIP3 model.[29] Two negatively charged chloride atoms and one positively charged sodium atom were added to the simulation box to reach electric neutrality (net charges were considered for Arg, Lys and Glu at neutral pH). Force-field parameters for natural residues were extrapolated from AMBER libraries,[30] while those for the non-coded amino acid proposed in this work were explicitly developed (see below).

Prior to the production cycles with the modified SA-MD, the simulation box was equilibrated. Thus, 0.5 ns of NVT-MD at 500 K were used to homogeneously distribute the solvent and ions in the box. Next, 0.5 ns of NVT-MD at 298 K (thermal equilibration) and 0.5 ns of NPT-MD at 298 K (density relaxation) were run. The last snapshot of the NPT-MD was used as the starting point for the conformational search process. This initial structure was quickly heated to 900 K at a rate of 50 K/ps to force the molecule to jump to a different region of the conformational space. Along 10 ns, the 900 K-structure was slowly cooled to 500 K at a rate of 1 K per 25 ps. A total of 500 structures were selected and

subsequently minimized during the first cycle of modified SA-MD. The resulting minimum energy conformations were stored in a rank-ordered library of low energy structures. The lowest energy minimum generated in a modified SA-MD cycle was used as starting conformation of the next cycle.

The energy was calculated using the AMBER potential.[30] Atom pair distance cut-offs were applied at 14.0 Å to compute the van der Waals and electrostatic interactions. In order to avoid discontinuities in the potential energy function, non-bonding energy terms were forced to slowly converge to zero, by applying a smoothing factor from a distance of 12.0 Å. Both temperature and pressure were controlled by the weak coupling method, the Berendsen thermobarostat,[31] using a time constant for heat bath coupling and a pressure relaxation time of 1 ps. Bond lengths were constrained using the *SHAKE* algorithm[32] with a numerical integration step of 2 fs.

4.2.2.3 Conformation classification and clustering analysis.

The list of unique minimum energy conformations was generated by comparing the 500 minimized structures provided by each cycle of modified SA-MD not only among themselves but also with unique minima generated in previous cycles. The list was organized by rank ordering all the unique minimum energy conformations found in an increasing energy, with the previously listed conformations discarded. Unique minimum energy conformations were identified based on the virtual dihedral angles which characterize the peptide backbone conformation and on hydrogen-bond and salt-bridge interactions. Five virtual dihedral angles were defined considering the α-carbon atoms of the five residues, the methyl carbon atom of the acetyl and methylamide capping groups and one acetyl hydrogen atom (Figure 4.2.1a). The existence of interactions was accepted on the basis of the following geometric criteria: a) for salt bridges: distance between the centers of the interacting groups shorter than 4.50 Å; b) for hydrogen bonds: H···O distance shorter than 2.50 Å and N–H···O angle higher than 120.0º. Two structures were considered different when differing in at least one of their virtual dihedral angles by more than 60º or in at least one of the above interactions. All the structures classified as different were subsequently clustered based on salt bridges and hydrogen bonds.

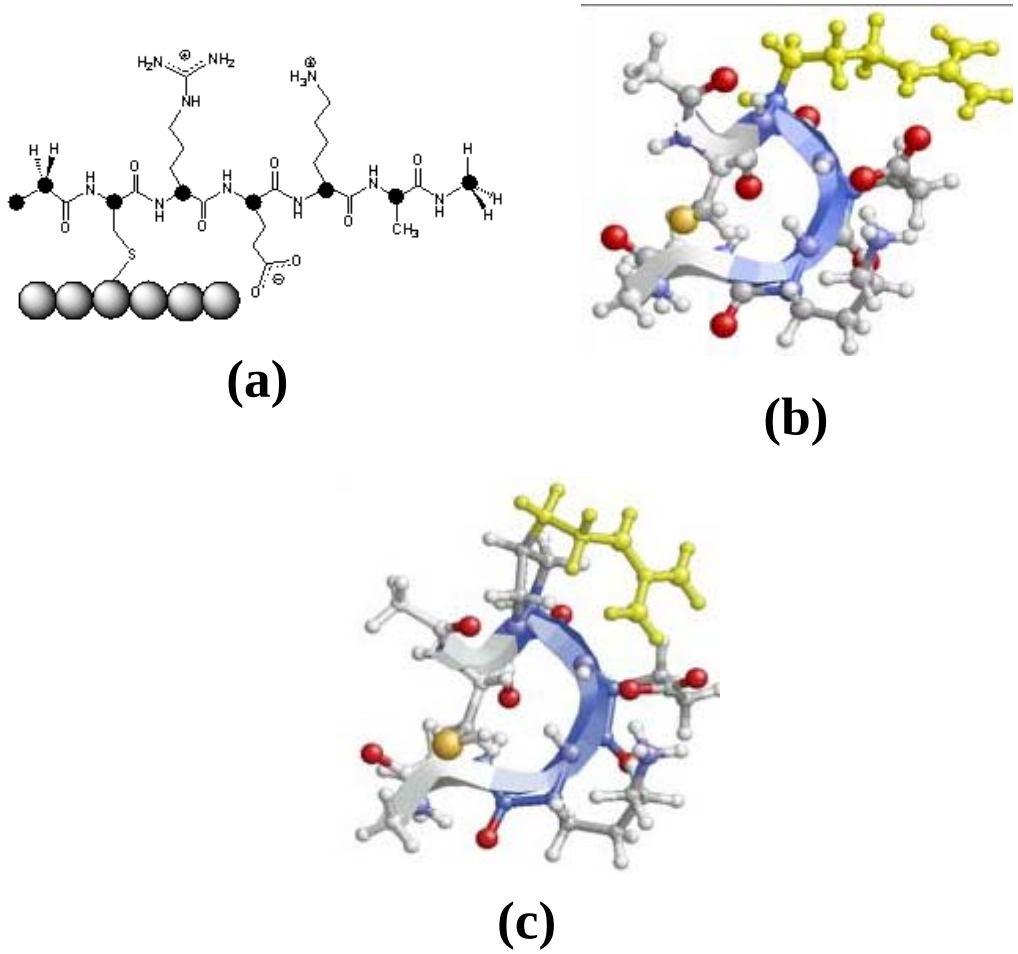


Figure 4.2.1: *(a) Schematic representation of the system under study showing the position of the particles (black dots) used to define the virtual dihedral angles (see Methods Section). (b) Bioactive conformation proposed for CREKA (according to ref. 5). (c) CREKA analogue constructed by replacing the Arg residue by (Pro)hArg within the bioactive conformation. The Arg and (Pro)hArg side chains are depicted in yellow.*

4.2.3 .Results and Discussion

4.2.3.1 Design of the Chemical Modifications in CREKA

The bioactive conformation proposed for natural CREKA on the basis of theoretical calculations[5] is depicted in Figure 4.2.1b. In this structure, the backbone defines a β type turn motif. The functionalized side chains of the central residues (Arg, Glu and Lys) face the same side of the molecule, and the backbone Cys CO and Lys NH groups are hydrogen bonded. This structural motif was identified in the global minimum of the free peptide, when inserted in a phage display protein, and in many of the accessible minima. Inspection of the side chains reveals salt bridges involving the negatively charged carboxylate group of Glu and the contiguous positively charged side chains of Arg and Lys. These side-chain interactions are made possible by the peptide backbone turn conformation. An extended backbone would place the positively and negatively charged side chains pointing towards opposite sides (Figure 4.2.1a). The salt-bridges presented multiple interaction patterns, with the specific atoms involved being dependent on the chemical environment considered, *i.e.* the peptide in the free state, attached to a nanoparticle or inserted in a phage display protein. Accordingly, the salt bridges identified in many of the significant minima were found to differ in the atoms involved in these interactions.

The structural features observed[5] for CREKA therefore lie in the β-turn conformation adopted by the peptide backbone. In this turn motif, Arg occupies the *i*+1 position. Proline (Pro) is known to exhibit a high propensity to induce turns (of either the β- or the γ-type) and, in particular, to occupy position *i*+1 of β-turns.[33–35] Moreover, the cyclic structure of Pro, which is unique among naturally-occurring amino acids, is known to impart stability against enzymatic hydrolysis.[36–39] Thus, replacing the Arg in CREKA by a proline-like derivative of this amino acid could lead to an increase in stability with a simultaneous retention, or even enhancement, of the propensity to adopt a folded turn-like conformation.

We therefore considered the replacement of Arg in CREKA by a proline derivative bearing the guanidinium group present in the Arg side chain at the γ-

position of the pyrrolidine ring. Proline derivatives of this type may be synthetically accessible by transformation of γ-hydroxyproline, a starting material available in enantiomerically pure form from commercial sources. The exact chemical structure of the newly designed amino acid needs however to be selected: an Arg side chain incorporated at the C^{γ} in proline can exhibit either a *cis* or a *trans* conformation with respect to the carbonyl function. In addition, the guanidinium group may be attached to the C^{γ} atom of the pyrrolidine ring through a variable number (*n*) of methylene units, whose optimal value for formation of salt-bridge interactions with the proximal Glu side chain should be determined.

For this purpose, graphical molecular modeling was performed using the proposed bioactive conformation[5] of natural CREKA (Figure 4.2.1b) as a template. This qualitative analysis provided the best fitting when the Arg side chain was arranged in *cis* with respect to the proline carbonyl group and for a chain length corresponding to $n = 2$ (Figure 4.2.1c). The amino acid thus designed was denoted as (Pro)hArg, where hArg stands for *homo*arginine, that is, the Arg homologue containing one more methylene unit. Figure 4.2.2 shows the structures of Arg, hArg and their corresponding *cis* proline-like derivatives (Pro)Arg and (Pro)hArg, respectively, for comparison.

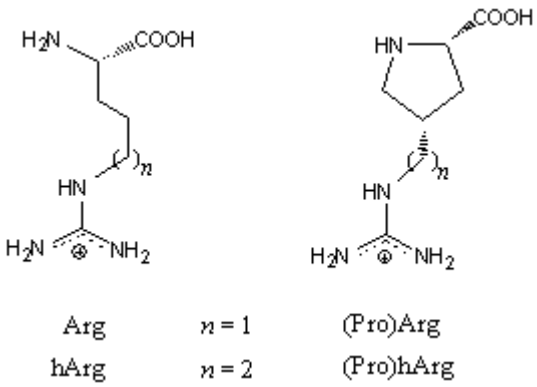
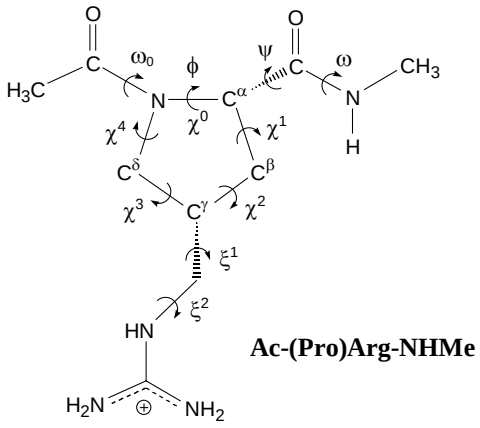


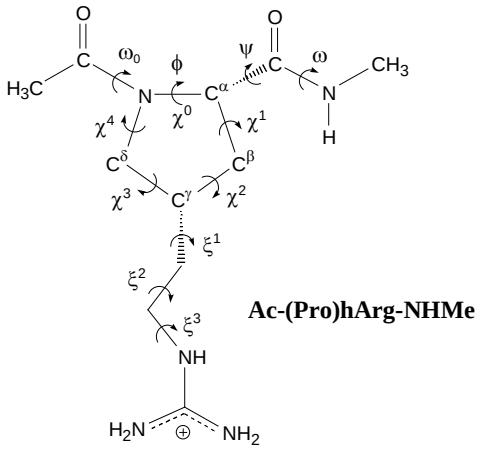
Figure 4.2.2: *Structure of arginine (Arg) and its homologue containing an additional methylene group (hArg), as well as that of their respective proline-like derivatives considered in this work.*

4.2.3.2 Conformational Profile and Force-Field Parametrization.

Before analyzing the conformational impact derived from the incorporation of (Pro)hArg into CREKA, parametrization of this non-proteinogenic residue is necessary. For this purpose, the conformational properties of its *N*-acetyl-*N'*-methylamide derivative, Ac-(Pro)hArg-NHMe (Figure 4.2.3b), have been analyzed. Given the large number of dihedral angles in this molecule and the subsequent huge number of starting geometries to be considered, a simplified methodology has been applied. Thus, the minimum energy conformations of this compound have been derived from those obtained for the residue containing one less methylene unit, Ac-(Pro)Arg-NHMe (Figure 4.2.3a).



(a)



(b)

Figure 4.2.3: *Dihedral angles used to identify the conformations of the* N-*acetyl*-N*'-methylamide derivatives of (Pro)Arg (a) and (Pro)hArg (b) studied in this work. The dihedral angles ω_0, φ, ψ and ω are defined using backbone atoms, whereas the endocyclic dihedral angles (χ^i) and ξ^1, respectively, are given by the atoms of the five-membered ring. In particular, the sequence of atoms used to define φ, χ^0 and ξ^1 are C(O)–N–C^α–C(O), C^δ–N–C^α–C^β and C^β–C^γ–CH_2–N, respectively.*

The backbone (ω_0,φ,ψ,ω) and side chain (χ^i, endocyclic; ξ^i exocyclic) dihedral angles considered for Ac-(Pro)Arg-NHMe and Ac-(Pro)hArg-NHMe are defined in Figure 4.2.3. The minimum energy conformations of these compounds have been denoted using a three-label code describing the backbone conformation, the puckering of the five-membered ring and the conformation of the exocyclic substituent. Specifically, the first label identifies the backbone conformation, defined by the φ,ψ dihedral angles, using Perczel's nomenclature.[40] Accordingly, nine different backbone conformations can be distinguished in the potential energy surface $E=E(\varphi,\psi)$ of an α-amino acid: γ_D, δ_D, α_D, ε_D, β_L, ε_L, α_L, δ_L and γ_L. In the case of proline, only the γ_L (γ-turn or C_7), α_L (α-helical), and ε_L (polyproline II) conformations are possible because φ is fixed around –60º.[33–35] The *up* or *down* puckering of the five-membered pyrrolidine ring is next indicated using the [u] and [d] labels, respectively. These conformational states are defined as those in which the C^γ atom and the carbonyl group of proline (or the proline-like residue) lie on the same or opposite sides, respectively, of the plane defined by the C^δ, N and C^α atoms. In particular, the *down* ring puckering is identified when χ^1 and χ^3 are positive while χ^2 and χ^4 are negative. Conversely, the *up* ring puckering is characterized by negative values of χ^1 and χ^3 and positive values of χ^2 and χ^4. Finally, the last set of letters indicates the *gauche*$^+$ (g^+), *skew*$^+$ (s^+), *trans* (t), *skew*$^-$ (s^-) or *gauche*$^-$ (g^-) arrangement of each exocyclic dihedral angle ξ^i.

In a first step, the intrinsic conformational properties of Ac-(Pro)Arg-NHMe (Figure 4.2.3a) were investigated, using DFT calculations at the B3LYP/6-31+G(d,p) level. The conformational search was performed considering that this compound retains the restrictions imposed by the five-membered pyrrolidine ring in proline. Thus, the three minimum energy conformations previously characterized for Ac-Pro-NHMe[41] with all *trans* amide bonds (ω_0 and $\omega \approx 180º$), namely γ_L[d], γ_L[u] and α_L[u], were used to generate the starting structures of Ac-(Pro)Arg-NHMe. The arrangement of the side group in (Pro)Arg is defined by the flexible dihedral angles ξ^1 and ξ^2 (Figure 4.2.3a), which are expected to exhibit three different minima: *trans* (180º), *gauche*$^+$ (60º) and *gauche*$^-$ (–60º). Accordingly, 3 (minima of Ac-Pro-NHMe) × 3 (minima of ξ^1) × 3 (minima of ξ^2)

= 27 minima were anticipated for the potential energy hypersurface (PEH) $E = E(\varphi,\psi,\chi^i,\xi^j)$ of Ac-(Pro)Arg-NHMe. All these structures were used as starting points for subsequent full geometry optimizations.

Table 4.2.1 lists the geometric parameters and the relative energies (ΔE^{gp}) of the five minima characterized for this compound in the gas phase, which are displayed in Figure 4.2.4. In the lowest energy minimum (γ_L[u]g^+t, Figure 4.2.4a), the backbone acetyl CO and methylamide NH sites form an intramolecular hydrogen bond defining a seven-membered cycle (γ-turn or C_7 conformation), while the pyrrolidine ring exhibits an *up* puckering and the exocyclic dihedral angles ξ^1 and ξ^2 are arranged in *gauche*$^+$ and *trans*, respectively. This side chain disposition allows formation of a strong hydrogen bond involving the (Pro)Arg carbonyl oxygen and the NH moiety in the guanidinium group. Similar backbone···backbone and backbone···side chain hydrogen-bond interactions are present in the third minimum (γ_L[d]s^+s^+, Figure 4.2.4c) although, in this case, a *down* puckering of the pyrrolidine ring and a *skew*$^+$ conformation of both ξ^1 and ξ^2 are required, and this side chain rearrangement is associated with an energy penalty of 3.6 kcal/mol.

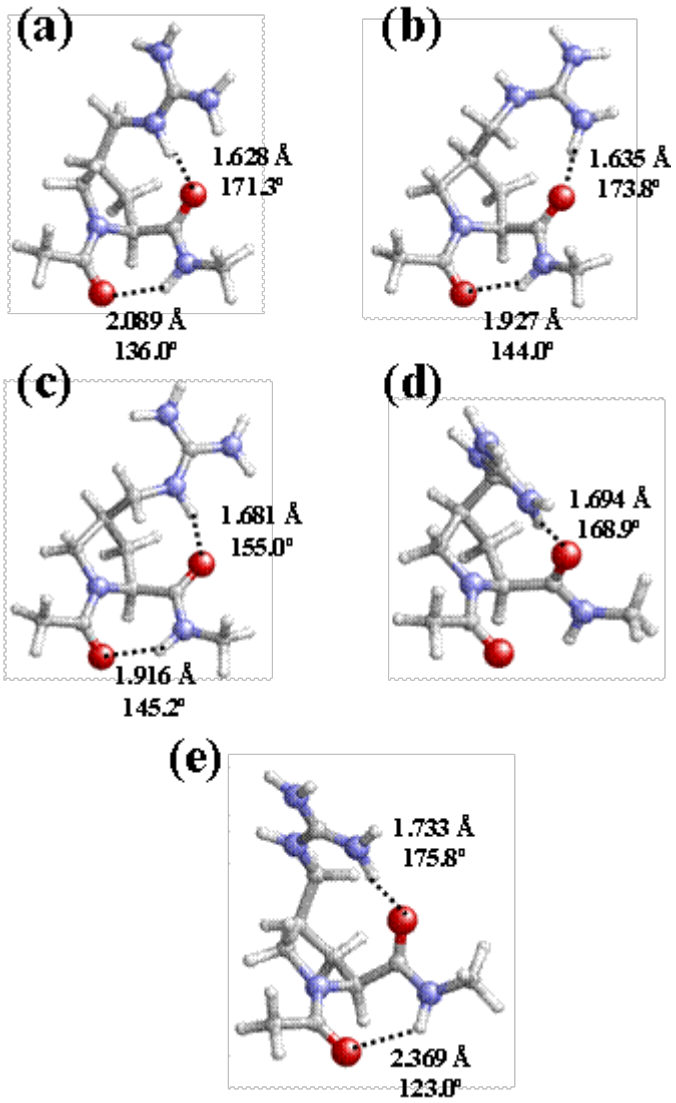


Figure 4.2.4: *Minimum energy conformations of Ac-(Pro)Arg-NHMe obtained from B3LYP/6-31+G(d,p) calculations: (a) γL[u]g+t; (b) γL[d]ts-; (c) γL[d]s+s+; (d) εL[d]g-s+; (e) γL[d]g-s+ (see Table 1 for geometries). Distances (H···O) and angles (N–H···O) associated with the backbone···backbone and backbone···side chain interactions (dashed lines) are given.*

Table 4.2.1 Dihedral angles[a,b] of the backbone and the exocyclic side group, pseudorotational parameters[a] of the pyrrolidine ring (A and P), and relative energy[c] (ΔE^{gp}) of the minimum energy conformations characterized for Ac-(Pro)Arg-NHMe at the B3LYP/6-31+G(d,p) level in the gas phase.

	ω_0	φ	ψ	ω	(P, A)	ξ_1	ξ_2	ΔE^{gp}
$\gamma_L[u]g^+t$	-175.9	-82.3	84.4	-176.7	(177.2, 27.7)[d]	67.7	-164.5	0.0[e]
$\gamma_L[d]ts^-$	-170.2	-82.9	74.0	-178.8	(-108.9, 30.7)[f]	-171.4	-115.3	0.4
$\gamma_L[d]s^+s^+$	-169.0	-82.0	71.4	179.4	(-97.4, 38.3)[g]	149.8	110.3	3.6
$\varepsilon_L[d]g^-s^+$	-172.9	-69.2	137.4	-178.5	(-91.1, 33.3)[h]	-38.5	104.7	5.6
$\gamma_L[d]g^-s^+$	-176.4	-82.4	98.4	-175.8	(-152.0, 29.0)[i]	-36.4	108.8	6.4

[a] In degrees. [b] See Figure 4.2.3a for definition. [c] In kcal/mol. [d] χ^0= -27.7º, χ^1= 20.4º, χ^2= -7.3º, χ^3= -8.7º and χ^4= 23.2º. [e] E= -817.237062 a.u.. [f] χ^0= -9.9º, χ^1= 25.2º, χ^2= -30.8º, χ^3= 24.3º and χ^4= -9.2º. [g] χ^0= -4.9º, χ^1= 26.7º, χ^2= -37.8º, χ^3= 33.8º and χ^4= -18.6º. [h] χ^0= -0.6º, χ^1= 20.4º, χ^2= -31.7º, χ^3= 30.8º and χ^4= -19.5º. [i] χ^0= -25.6º, χ^1= 27.8º, χ^2= -20.9º, χ^3= 5.8º and χ^4= 12.6º.

Indeed, for a γ_L backbone conformation, the most favorable backbone···side chain interaction when the pyrrolidine ring is *down*-puckered seems to involve the NH_2 rather than the NH site in the guanidinium side chain. This is inferred from the geometry of the second conformer in Table 4.2.1 ($\gamma_L[d]ts^-$, Figure 4.2.4b), which is destabilized with respect to the global minimum by only 0.4 kcal/mol. Interestingly, deviation of the ψ angle in this second conformer to values around 100º results in a new minimum energy structure ($\gamma_L[d]g^-s^+$, Figure 4.2.4e) where the backbone···backbone and backbone···side chain hydrogen-bond interactions are retained. However, the large ψ angle leads to a much less favorable geometry for the hydrogen bond stabilizing the γ-turn conformation and this results in a ΔE^{gp} value of 6.4 kcal/mol.

The only conformer in Table 4.2.1 with a backbone structure other than γ_L is the fourth minimum ($\varepsilon_L[d]g^-s^+$, Figure 4.2.4d), which corresponds to a polyproline II conformation. Despite the presence of a strong interaction involving the (Pro)Arg carbonyl and one guanidinium NH_2 moiety, this conformer is unfavored

by 5.6 kcal/mol, due to the absence of hydrogen bonds linking the backbone amide groups.

The free energies in the gas phase (ΔG^{gp}) calculated for the five minimum energy conformations of Ac-(Pro)Arg-NHMe are displayed in Table 4.2.2. As can be seen, consideration of the ZPVE, thermal and entropic corrections for the transformation of ΔE^{gp} into ΔG^{gp} represents relative variations lower than 0.2 kcal/mol in all cases with the exception of the γ_L[d]ts$^-$ disposition, for which these statistical corrections produce a destabilization of 1.2 kcal/mol. Accordingly, γ_L[u]g$^+$t is the only conformation significantly populated in the gas phase at room temperature according to a Boltzmann distribution since all the local minima exhibit ΔG^{gp} values higher than 1.5 kcal/mol.

Table 4.2.2 *Relative free energy[a] in the gas phase (ΔG^{gp}) and in carbon tetrachloride, chloroform and aqueous solutions (ΔG^{CCl4}, ΔG^{CHCl3} and ΔG^{H2O}, respectively) for the minimum energy conformations of Ac-(Pro)Arg-NHMe at the B3LYP/6-31+G(d,p) level.*

	ΔG^{gp}	ΔG^{CCl4}	ΔG^{CHCl3}	ΔG^{H2O}
γ_L[u]g$^+$t	0.0[b]	1.1	2.6	6.2
γ_L[d]ts$^-$	1.6	0.0	0.0	1.7
γ_L[d]s$^+$s$^+$	3.8	3.7	4.3	6.6
ε_L[d]g$^-$s$^+$	5.7	2.4	0.6	0.0
γ_L[d]g$^-$s$^+$	6.4	4.1	3.3	4.2

[a] In kcal/mol. [b] G= -816.963546 a.u.

In order to obtain an estimation of the solvation effects on the relative stability of the different minima, single point calculations were conducted on the optimized structures using the PCM method. Table 4.2.2 includes the relative free energies in carbon tetrachloride, chloroform and water solutions (ΔG^{CCl4}, ΔG^{CHCl3} and ΔG^{H2O}, respectively). The solvent introduces significant changes in the relative stability of the different minima characterized for Ac-(Pro)Arg-NHMe. Carbon tetrachloride was found to considerably stabilize the γ_L[d]ts$^-$, ε_L[d]g$^-$s$^+$ and γ_L[d]g$^-$

s^+ conformations, to the point that the former becomes the most stable structure and the two latter are stabilized by 3.3 and 2.3 kcal/mol, respectively, with respect to the gas phase. The higher polarity of chloroform results in a further stabilization of the $\varepsilon_L[d]g^-s^+$ conformation, even though the lowest energy structure remains $\gamma_L[d]ts^-$, as in carbon tetrachloride. Finally, $\varepsilon_L[d]g^-s^+$ becomes the most stable structure in aqueous solution, the ΔG^{H2O} values of the other four conformers being higher than 1.5 kcal/mol.

The $\gamma_L[u]g^+t$ and $\varepsilon_L[d]g^-s^+$ arrangements, corresponding to the lowest free energy conformations of Ac-(Pro)Arg-NHMe in the gas phase and in aqueous solution, respectively (Table 4.2.2), were used as starting structures for the conformational study of Ac-(Pro)hArg-NHMe (Figure 4.2.3b). Thus, the starting geometries for the latter compound were prepared by including an additional methylene group in the side chain of (Pro)Arg for such two representative structures. The conformation of the new methylene group is defined by the dihedral angle ξ^3 (Figure 4.2.3b), for which three different arrangements were considered: *gauche*$^+$, *trans* and *gauche*$^-$. Accordingly, the conformational search of Ac-(Pro)hArg-NHMe was carried out considering 2 [representative minima of Ac-(Pro)Arg-NHMe] × 3 (minima of ξ^3) = 6 starting geometries. Energy minimizations at the B3LYP/6-31+G(d,p) level led to five different minimum energy structures, the three most stable being given in Table 4.2.3 and Figure 4.2.5. The remaining two are not included since their relative energies were found to be above 10 kcal/mol and therefore were not considered representative.

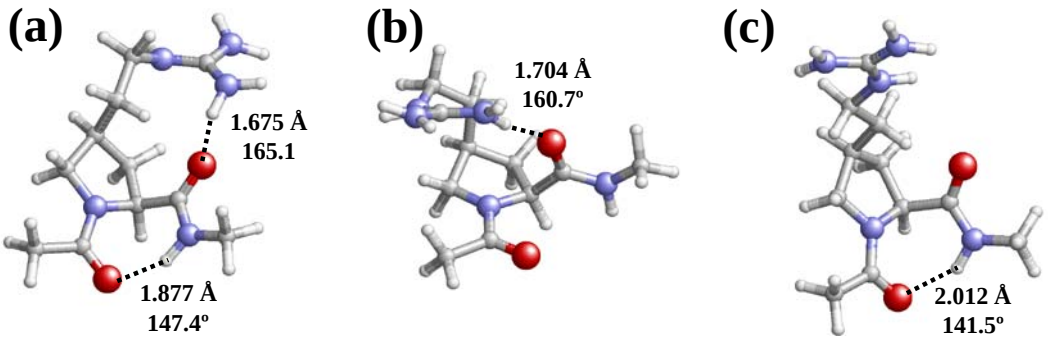


Figure 4.2.5: *Minimum energy conformations of Ac-(Pro)hArg-NHMe obtained from B3LYP/6-31+G(d,p) calculations: (a) $\gamma_L[d]tg^+g^-$; (b) $\varepsilon_L[d]g^-g^+g^+$; (c) $\gamma_L[d]g^-tt$ (see Table 3 for geometries). Distances (H···O) and angles (N–H···O) associated with the backbone···backbone and backbone···side chain interactions (dashed lines) are given.*

Table 4.2.3. *Dihedral angles[a,b] of the backbone and the exocyclic side group, pseudorotational parameters[a] of the pyrrolidine ring (A and P), and relative energy[c] (ΔE^{gp}) of the minimum energy conformations characterized for Ac-(Pro)hArg-NHMe at the B3LYP/6-31+G(d,p) level in the gas phase.*

	ω_0	φ	ψ	ω	(P, A)	ξ_1	ξ_2	ξ_3	ΔE^{gp}
γ_L[d]tg$^+$g$^-$	-169.4	-82.5	68.8	-179.5	(-105.0, 35.2)[d]	-159.3	52.3	-85.3	0.0[e]
ε_L[d]g$^-$g$^+$g$^+$	-171.0	-68.9	139.9	-179.4	(-86.6, 37.1)[f]	-51.8	70.0	69.8	5.5
γ_L[d]g$^-$tt	-170.9	-83.7	78.1	-178.4	(-105.5, 32.9)[g]	-72.2	179.3	-176.2	9.6

[a] In degrees. [b] See Figure 4.2.3b for definition. [c] In kcal/mol. [d] χ^0= -9.1º, χ^1= 27.6º, χ^2= -35.1º, χ^3= 29.0º and χ^4= -12.9º. [e] E= -856.547433 a.u. [f] χ^0= -10.4º, χ^1= 26.8º, χ^2= -32.9º, χ^3= 26.0º and χ^4= -10.2º. [g] χ^0= 2.2º, χ^1= 20.5º, χ^2= -34.3º, χ^3= 35.0º and χ^4= -24.1º.

The lowest energy conformation (γ_L[d]tg$^+$g$^-$, Figure 4.2.5a) corresponds to a γ-turn conformation stabilized by a hydrogen bond linking the backbone terminal CO and NH sites. The (Pro)hArg CO group and one of the NH_2 moieties in the guanidinium side chain are also involved in a strong hydrogen-bond interaction. The latter is lost in the γ_L[d]g$^-$tt minimum (Figure 4.2.5c), which explains its high relative energy. Conversely, the second conformer in Table 4.2.3 (ε_L[d]g$^-$g$^+$g$^+$, b Figure 4.2.5) exhibits no backbone···backbone hydrogen bond, as expected for an ε_L conformation, while being stabilized by a strong side chain···backbone interaction. Accordingly, it is destabilized by 5.5 kcal/mol with respect to the lowest energy minimum.

Table 4.2.4 lists the ΔG^{gp}, ΔG^{CCl4}, ΔG^{CHCl3} and ΔG^{H2O} values calculated for the three Ac-(Pro)hArg-NHMe minima described above. Notably, the ZPVE, thermal and entropic corrections stabilize the γ_L[d]g$^-$tt structure in the gas phase by 3.5 kcal/mol. In spite of this, the γ_L[d]tg$^+$g$^-$ remains the only accessible conformation both in the gas phase and in carbon tetrachloride solution. In contrast, γ_L[d]g$^-$tt becomes the most stable structure in the presence of chloroform or water.

These results indicate that (Pro)hArg retains the most important structural features of proline,[41] reflecting a high tendency to induce peptide turns. It should also be noted that, even if the γ_L and ε_L conformations are characterized by different values of the ψ dihedral angle, both of them correspond to turn-like

backbone conformations, namely the *i*+1 position of a γ- and a βII-turn, respectively.[33–35]

Table 4.2.4 *Relative free energy*[a] *in the gas phase (ΔG^{gp}) and in carbon tetrachloride, chloroform and aqueous solutions (ΔG^{CCl4}, ΔG^{CHCl3} and ΔG^{H2O}, respectively) for the minimum energy conformations of Ac-(Pro)hArg-NHMe at the B3LYP/6-31+G(d,p) level.*

	ΔG^{gp}	ΔG^{CCl4}	ΔG^{CHCl3}	ΔG^{H2O}
γ_L[d]tg$^+$g$^-$	0.0[b]	0.0	2.1	5.3
ε_L[d]g$^-$g$^+$g$^+$	5.1	3.8	4.3	5.1
γ_L[d]g$^-$tt	6.1	1.7	0.0	0.0

[a] In kcal/mol. [b] G= -856.244850 a.u.

The stretching, bending, torsion and van der Waals parameters used in the AMBER force-field[30] to describe Pro and Arg were directly transferred to (Pro)hArg. Atomic centered charges for the minimum energy conformations listed in Table 4.2.3 were calculated by fitting the UHF/6-31G(d) quantum mechanical and the Coulombic molecular electrostatic potentials (MEPs) to a large set of points placed outside the nuclear region. It should be noted that the electrostatic parameters derived at this level of theory are fully compatible with the current AMBER force-field.[30] On the other hand, electrostatic force-field parametrization using a strategy based on weighted multiple conformations through a Boltzmann distribution in the gas phase, which was originally proposed by different authors,[42] has been demonstrated to be especially successful for non-proteinogenic residues.[42b,43] Accordingly, parameters for (Pro)hArg (Figure 4.2.6) have been obtained considering the atomic charges of the lowest energy minimum only (γ_L[d]tg$^+$g$^-$, Table 4.2.4) since the other two local minima are unfavored by more than 5.1 kcal/mol and, therefore, their contribution to a normalized Boltzmann distribution can be considered negligible.

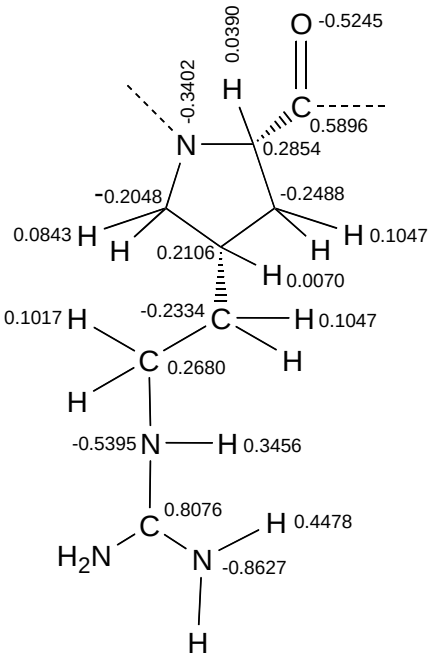


Figure 4.2.6: *Electrostatic parameters determined for the (Pro)hArg residue*

4.2.3.3 Conformational Search of the CREKA Analogue Attached to a Nanoparticle.

Comparison with the Natural Peptide. The conformational preferences of the CREKA analogue incorporating (Pro)hArg as an Arg substitute, hereafter denoted as CR*EKA, have been explored using the sampling technique previously employed for the study of the natural pentapeptide.[5] Figure 4.2.7a represents the evolution of the number of unique minimum energy conformations against the number of modified SA-MD production cycles necessary for the conformational search of CR*EKA to converge. As can be seen, the exploration is completed after seven cycles, the last one providing only 3 new structures to the list of unique conformations. Interestingly, the replacement of Arg by (Pro)hArg did not lead to a reduction in the amount of modified SA-MD cycles required to complete the conformational search, but the number of accessible low energy conformations diminished dramatically, *i.e.* 612 and 1305 minima were obtained for CR*EKA and CREKA, respectively. Moreover, the energetic distribution of the minima generated was also affected by this targeted replacement, as indicated in Figure 4.2.7b. Thus, the conformational restrictions imposed by the presence of (Pro)hArg eliminated a large number of minima with relative energies in the interval between 4 and 11 kcal/mol, which was found to be the most populated for

CREKA. This effect is shown by the plateau distribution obtained for CR*EKA, which contrasts with the gaussian-like distribution achieved for the natural peptide.

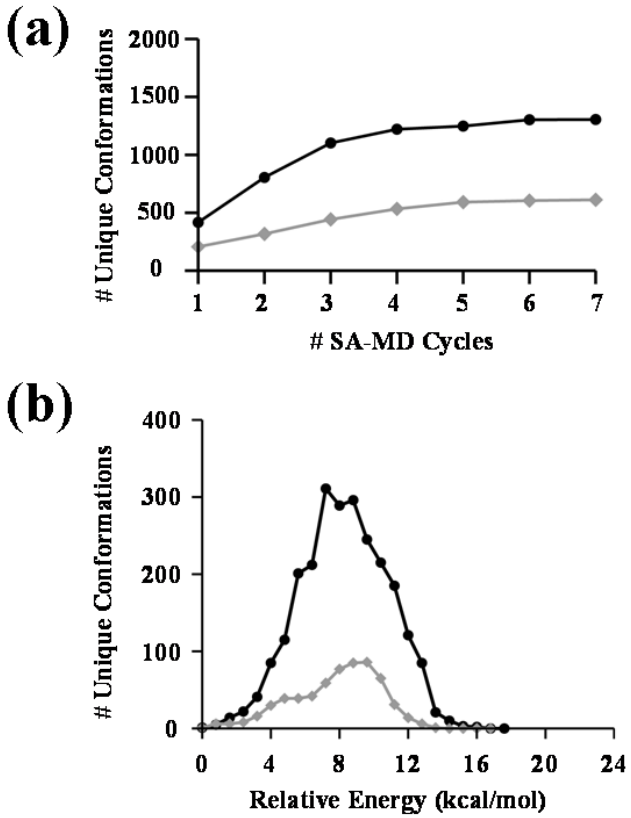


Figure 4.2.7: *(a) Number of unique minimum energy conformations found for CR*EKA (grey line and diamonds) and natural CREKA (black line and circles) against the number of modified SA-MD cycles used for the conformation search. (b) Distribution of energies for the unique minimum energy conformations of CR*EKA and natural CREKA. Energies are computed relative to the corresponding lowest energy minimum.*

The conformational preferences of the backbone in the two pentapeptides have been compared by analyzing the virtual dihedral angles used to define the specific arrangement of each residue. Results are given in Figure 4.2.8, which represents the distribution of such dihedrals through histograms. As can be seen, the incorporation of (Pro)hArg produces some changes in the general conformational profile of the peptide. The distributions obtained for the (Pro)hArg and Glu residues in CR*EKA are very similar to those found for their Arg and Glu counterparts in CREKA, while important variations are detected for the other

three residues, especially Lys and Ala. Indeed, the conformational space of the two latter amino acids is significantly narrower in CR*EKA, indicating that the incorporation of (Pro)hArg reduces the conformational flexibility of the whole peptide. This is also reflected in Figure 4.2.9, which compares the distribution of the φ,ψ backbone dihedral angles of the five residues for the minima with relative energies lower than 2 kcal/mol.

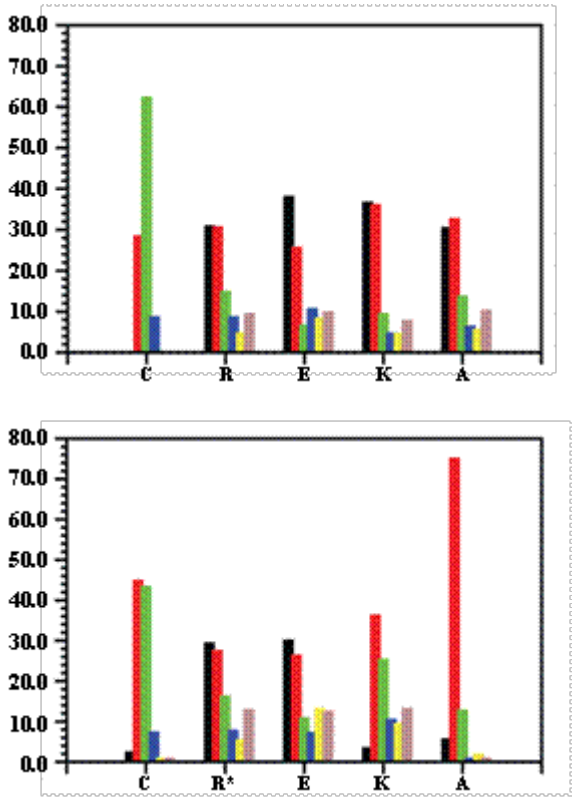


Figure 4.2.8: *Comparison of the distribution of virtual dihedral angles used to define the backbone conformation of CREKA (a) and CR*EKA (b) in each unique minimum energy structure obtained. Color code for the bars: black for dihedral angle values ranging from 0° to 60°, red from 60° to 120°, green from 120° to 180°, blue from 180° to 240°, yellow from 240° to 300° and grey from 300° to 360°.*

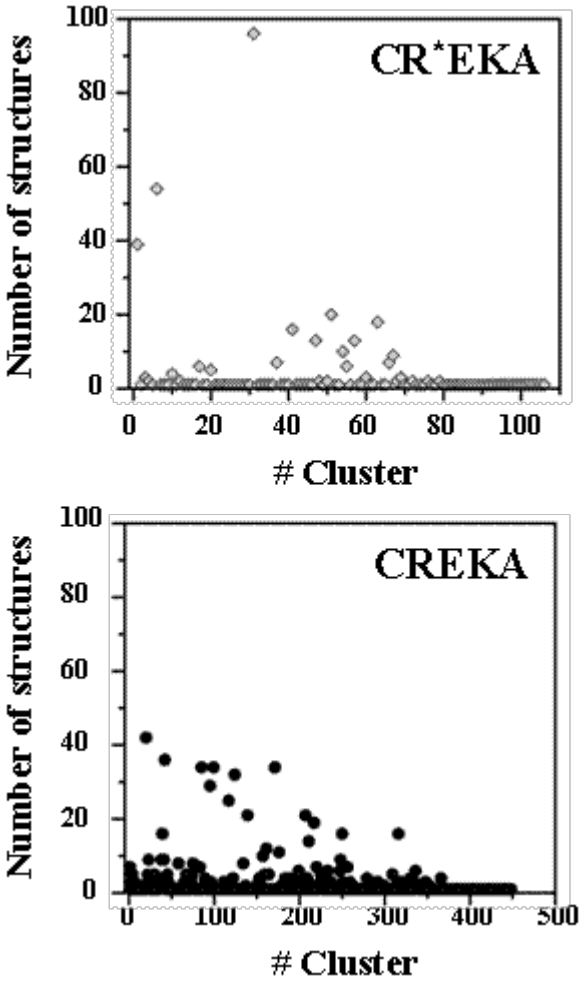


Figure 4.2.9: *Distribution of the unique minimum energy structures generated for CR*EKA (top) and CREKA (bottom) in clusters, which have been grouped on the basis of the formation of hydrogen bonds and salt bridges.*

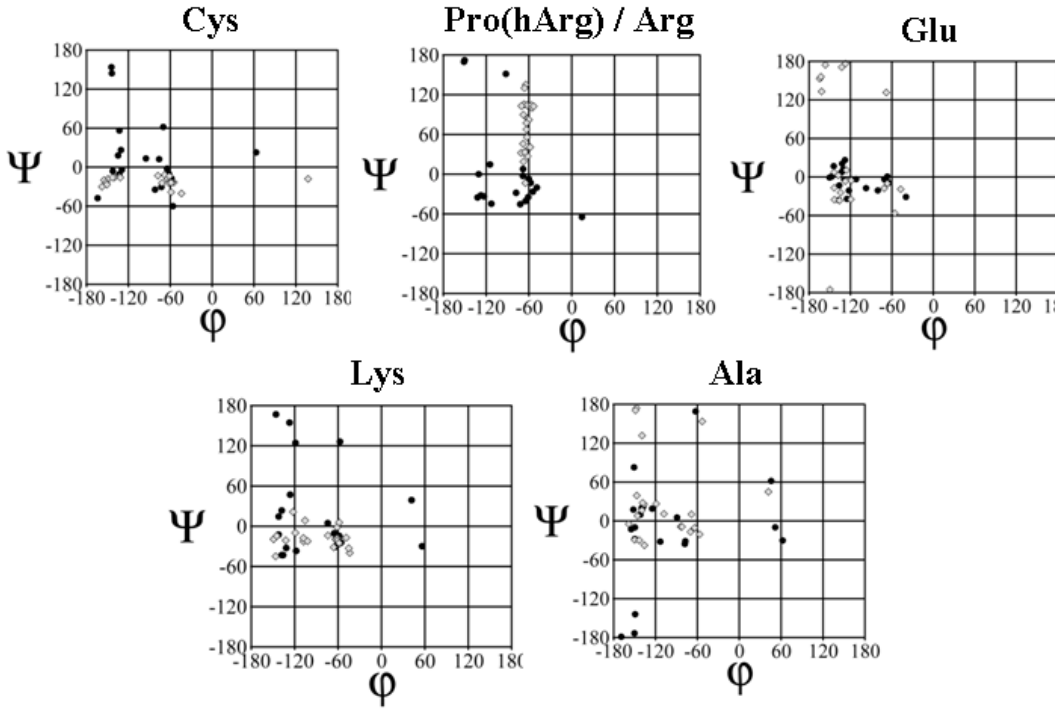


Figure 4.2.10: *Ramachandran plot distribution for the five residues of CR*EKA (open circles) and CREKA (filled black circles) considering the more representative minimum energy structures, i.e. those within a relative energy interval of 2 kcal/mol.*

A clustering analysis based on hydrogen bonds and salt bridges indicated that 68% of the CR*EKA minima (417 structures) present at least one such interaction, while this value reaches 82% (1092 structures) for natural CREKA. The total number of interactions (either of the hydrogen-bond or salt-bridge type) detected in these minima is 667 (1.6 interactions per structure) and 2198 (2.1 interactions per structure) for CR*EKA and CREKA, respectively, which are distributed in 106 and 448 clusters (Figure 4.2.10). It is interesting to note that only 3 clusters in CR*EKA and 8 in CREKA contain more than 25 structures, respectively grouping 31% and 25% of the minima. Clusters are organized as follows: (*i*) backbone···backbone hydrogen bonds: 79.3% (529 interactions) for CR*EKA and 83.6% (1839 interactions) for CREKA; (*ii*) backbone···side chain hydrogen bonds: 0.0% (no interaction) for CR*EKA and 4.7% (102 interactions) for CREKA; (*iii*) side chain···side chain salt bridges: 20.7% (138 interactions) for CR*EKA and 11.7% (257 interactions) for CREKA.

The most frequent interactions in the modified peptide are the Glu(N–H)···(O=C)Cys hydrogen bond (219 interactions; 32.8%) and the Glu···Lys salt bridge (138 interactions; 20.7%), while in natural CREKA they are the Glu(N–H)···(O=C)Ac, Lys(N–H)···(O=C)Cys and Ala(N–H)···(O=C)Arg backbone···backbone hydrogen bonds (251, 225 and 207 interactions, respectively; 11.4%, 10.2% and 9.4%, respectively) and the Arg···Glu salt bridge (187 interactions; 8.5%). These distinct interaction patterns suggest important differences between the two peptides, which are confirmed upon a cross-comparison of the frequency in which a particular interaction is observed. Thus, the backbone···backbone hydrogen bond most frequently detected in CR*EKA, Glu(N–H)···(O=C)Cys, accounts for only 3.4% of the CREKA interactions. Conversely, the Glu(N–H)···(O=C)Ac, Lys(N–H)···(O=C)Cys and Ala(N–H)···(O=C)Arg hydrogen bonds are, respectively, 0.8%, 3.6% and 5.1% of the CR*EKA interactions. Regarding side chain···side chain contacts, the frequency of the Glu···Lys salt bridge is 7.5% in CREKA, while the Arg···Glu interaction was not identified in any of the CR*EKA minima.

These results indicate that the conformational restrictions imposed by the presence of (Pro)hArg lead to some alterations in the conformational profile of the whole peptide that affect the turn type generated and the residues involved in this turn, as well as the interactions between adjacent ionized side chains. As expected, the conformational propensities of the (Pro)hArg-containing peptide seem to be more clearly defined than those of the natural sequence, as shown by the smaller variation observed for the preferred backbone···backbone and side chain···side chain interactions in CR*EKA with respect to CREKA. This is deduced from the cluster analysis described above and it becomes even clearer when the interaction schemes of the three conformations of lowest energy generated for the two peptides are compared (Table 4.2.5). Thus, up to three different β-turns are detected in these CREKA conformations, namely those centered at Arg-Glu [stabilized by a Lys(N–H)···(O=C)Cys hydrogen bond; minima # 1 and 3], Glu-Lys [stabilized by a Ala(N–H)···(O=C)Arg hydrogen bond; minimum # 2] and Lys-Ala [stabilized by a NHMe(N–H)···(O=C)Glu hydrogen bond; minimum # 2]. In comparison, all three CR*EKA conformers in Table 4.2.5 share a γ-turn centered at the (Pro)hArg residue and stabilized by a

Glu(N–H)···(O=C)Cys hydrogen bond. This is due to the much higher propensity of proline, when compared to other proteinogenic amino acids, in nucleating turns and therefore occupying the central turn positions.[33–35] The fact that the turn observed in the lowest energy conformation in CREKA is of the β type whereas it is of the γ type in CR*EKA, suggests that the difference in the overall shapes of the molecules is rather small. In both turns the *i*+1 position is occupied by either Arg (in CREKA) or its substitute (in CR*EKA) and both turn conformations orient the charged side chains of the three central residues toward the same region of the molecule. Thus, the main structural requirements considered to be essential for the bioactivity of CREKA[5] are also present in CR*EKA.

Table 4.2.*5 Comparison between the interaction pattern[a] of the three minima of lowest energy generated for natural CREKA[b] and its analogue CR*EKA. Relative energies[c] (ΔE) are also given.*

#	*ΔE*	*CR*EKA*	*ΔE*	*CREKA*
1	0.0	HB Glu(N–H)···(O=C)Cys SB Lys···Glu	0.0	HB Lys(N–H)···(O=C)Cys SB Arg···Glu SB Lys···Glu
2	1.0	HB Glu(N–H)···(O=C)Cys HB Ala(N–H)···(O=C)Cys HB NHMe (N–H)···(O=C)Glu SB Lys···Glu	1.0	HB Ala(N–H)···(O=C)Arg HB NHMe(N–H)···(O=C)Glu SB Arg···Glu
3	1.0	HB Glu(N–H)···(O=C)Cys HB NHMe(N–H)···(O=C)Lys SB Lys···Glu	2.0	HB Lys(N–H)···(O=C)Cys HB Ala(N–H)···(O=C)Cys SB Arg···Glu

[a] Hydrogen bonds and salt bridges are labeled as HB and SB, respectively. [b] According to reference [5]. [c] In kcal/mol.

In spite of this similarity, the natural peptide and its analogue differ in the interactions involving the three charged side chains. Specifically, the only salt-bridge type detected in the three CR*EKA minima in Table 4.2.5 involves Glu and Lys, while the natural CREKA presents multiple salt bridges Arg···Glu···Lys (minimum # 1) or an Arg···Glu only (minima # 2 and 3). The Arg···Glu salt bridge seems to be disfavored when (Pro)hArg replaces Arg, but the biological consequences of this change are difficult to predict since this specific interaction may not be essential to bioactivity. In the peptide-receptor recognition process, the Arg, Gly and Lys side chains are likely to interact with the complementary groups in the receptor, rather than among themselves.

Overall, these results suggest that, although the targeted replacement of Arg by (Pro)hArg in CREKA induces some changes in the conformational profile of the peptide, the most important structural trends of the natural compound are retained. This general resemblance is reflected in the lowest energy conformations found for CR*EKA and for CREKA, with the backbones within 1.5 Å of each other. Figure 4.2.11 shows the superposition of the two structures. This shape similarity between the most stable minima suggests that the incorporation of

(Pro)hArg to provide resistance against the proteolytic enzymes would not induce major alterations in the conformational features considered to be essential for the tumor-homing activity of the pentapeptide.

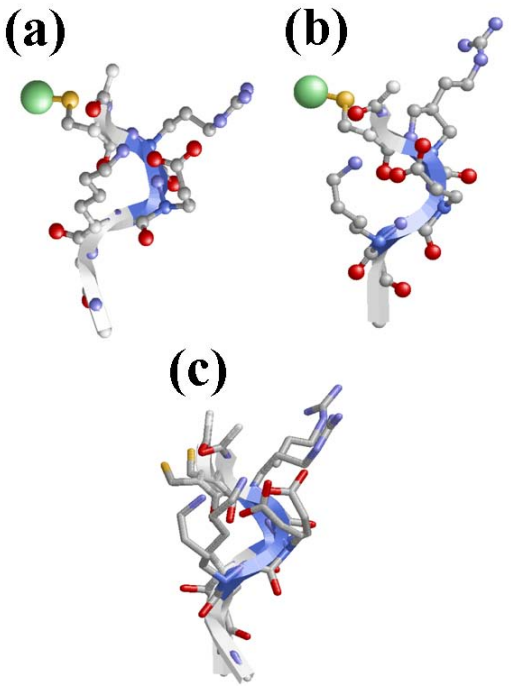


Figure 4.2.11: *Lowest energy minimum obtained for CR*EKA (a) and CREKA (b) attached to a nanoparticle, and superposition of both structures (c). The surface used to mimic the nanoparticle is represented by a single green ball*

4.2.4 Conclusions

A targeted replacement strategy has been followed to protect CREKA, a promising tumor-homing pentapeptide for cancer diagnostics and treatment, from proteolytic cleavage. For this purpose, a proline-like derivative of arginine, denoted as (Pro)hArg, has been designed, parameterized and incorporated into CREKA, replacing Arg. The conformational flexibility of the CREKA analogue incorporating (Pro)hArg is significantly reduced with respect to the natural CREKA. We attribute this reduction to the conformational restrictions imposed by the cyclic residue. The presence of this arginine surrogate produces some alterations in the interaction patterns as observed in the clustering analysis based on hydrogen bonds and salt bridges and in the detailed comparison of the interactions in the lowest energy conformations in the natural and modified peptide. In spite of this, the most important features proposed for the bioactive conformation of CREKA, *i.e.* a turn motif orienting the charged side chains towards the same side of the molecule, are retained in the (Pro)hArg-containing

analogue. The shape of the lowest energy conformation of the modified peptide shows a significant degree of similarity with that proposed for the bioactive conformation of the natural peptide. Thus, our results suggest that the incorporation of (Pro)hArg into CREKA as an arginine substitute could be useful to protect the peptide from proteolytic attack while retaining, or even enhancing, those structural features of the parent peptide considered to be essential for its tumor-homing activity.

4.2.5 References

1. Simberg, D.; Duza, T.; Park, J. H.; Essier, M.; Pilch, J.; Zhang, L.; Derfus, A. M.; Yang, M.; Hoffman, R. M.; Bathia, S.; Sailor, M. J.; Ruoslahti, E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 932.
2. (a) Hoffman, J. A.; Giraudo, E.; Singh, M.; Zhang, L.; Inoue, M.; Porkka, K.; Hanahan, D.; Ruoslathi, E. *Cancer Cell* **2003**, *4*, 383. (b) Pasqualini, R.; Ruoslahti, E. *Nature* **1996**, *380*, 364.
3. Hutchinson, J. N.; Muller, W. J. *Oncogene* **2000**, *19*, 6130.
4. Ruoslahti, E. Personal communication (unpublished results) 2007.
5. Zanuy, D.; Flores-Ortega, A.; Casanovas, J.; Curcó, D.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2008**, *112*, 8692.
6. Adessi, C.; Soto, C. *Curr. Med. Chem.* **2002**, *9*, 963.
7. Lelais, G.; Seebach, D. *Biopolymers* **2004**, *76*, 206.
8. Yamaguchi, H.; Kodama, H.; Osada, S.; Kato, F.; Jelokhani-Niaraki, M.; Kondo, M. *Biosci. Biotech. Biochem.* **2003**, *67*, 2269.
9. Sadowsky, J. D.; Murray, J. K.; Tomita, Y.; Gellman, S. H. *ChemBioChem* **2007**, *8*, 903.
10. Biron, E.; Chatterjee, J.; Ovadia, O.; Langenegger, D.; Brueggen, J.; Hoyer, D.; Schmid, H. A.; Jelinek, R.; Gilon, C.; Hoffman, A.; Kessler, H. *Angew. Chem. Int. Ed.* **2008**, *47*, 2595.
11. Banerjee, R.; Basu, G.; Chene, P.; Roy, S. *J. Pept. Res.* **2002**, *60*, 88.

12. Webb, A. I.; Dunstone, M. A.; Williamson, N. A.; Price, J. D.; de Kauwe, A.; Chen, W.; Oakley, A.; Perlmutter, P.; McCluskey, J.; Aguilar, M. I.; Rossjohn, J.; Purcell, A. W. *J. Immunol.* **2005**, *175*, 3810.
13. Horne W. S.; Gellman. S. H. *Acc. Chem. Res.* **2008**, *41*, 1399.
14. Agrafiotis, D. M.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E. *J. Chem. Inf. Model.* **2007**, *47*, 1067.
15. Steinbach, P. J. *Proteins* **2004**, *57*, 665.
16. Swaminathan, P.; Hariharan, M.; Murali, R.; Singh, C. U. *J. Med. Chem.* **1996**, *39*, 2141.
17. Zheng, J.; Zanuy, D.; Haspel, N.; Tsai, C.-J.; Alemán, C.; Nussinov, R. *Biochemistry* **2007**, *46*, 1205.
18. Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2007**, *111*, 3236.
19. Ballano, G.; Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2008**, *112*, 13101.
20. Kirkpatrick, S.; Gelatt Jr., C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
21. Steinbach, P. J.; Brooks, B. R. *Chem. Phys. Lett.* **1994**, *226*, 447.
22. Gaussian 03, Revision B.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, Jr., T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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.; C. Strain, M.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Pittsburgh PA, 2003.

23. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
24. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1993**, *37*, 785.
25. McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639.
26. (a) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999. (b) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027. (c) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239. (d) Miertus, M.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
27. Baysal, C.; Meirovitch, H. *J. Comput. Chem.* **1999**, *20*, 1659.
28. Simmerling, C.; Elber, R. *J. Am. Chem. Soc.* **1994**, *116*, 2534.
29. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983,** *79*, 926.
30. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
31. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

32. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

33. MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397.

34. Rose, G. D.; Gierasch, L. M.; Smith, J. A. *Adv. Protein Chem.* **1985**, *37*, 1.

35. Chakrabarti, P.; Pal, D. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1.

36. Vanhoof, G.; Goossens, F.; De Meester, I.; Hendriks, D.; Scharpé S. *Faseb J.* **1995**, *9*, 736.

37. Walker, J. R.; Altman, R. K.; Warren, J. W.; Altman, E. *J. Pept. Res.* **2003**, *62*, 214.

38. Green, B. D.; Gault, V. A; Irwin, N.; Mooney, M. H; Bailey, C. J.; Harriott, P.; Greer, B.; Flatt, P. R.; O'Harte, F. P. M. *Biol. Chem.* **2003**, *384*, 1543.

39. (a) Markert, Y.; Köditz. J.; Ulbrich-Hofmann, R.; Arnold, U. *Protein Eng.* **2003**, *16*, 1041. (b) Frenken, L. G. J.; Egmond, M. R.; Batenburg, A. M.; Verrips, C. T. *Protein Eng.* **1993**, *6*, 637. (c) Reidhaar-Olson, J. F.; Parsell, D. A.; Sauer, R. T. *Biochemistry* **1990**, *29*, 7563.

40. Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, *113*, 6256.

41. Flores-Ortega, A.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C.; Casanovas, J. *J. Org. Chem.* **2008**, *73*, 3418.

42. (a) Reynolds, C. A.; Essex, J. W.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 9075. (b) Alemán, C.; Casanovas, J. *J. Chem. Soc. Perkin Trans 2*, **1994**, 563. (c) Cieplak, P.; Cornell, W. D.; Bayly, C. I.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357.

43. (a) Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *J. Org. Chem.* **2007**, *72*, 2174. (b) Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *Chem. Phys. Lett.* **2006**, *429*, 558. (c) Alemán, C.; Zanuy, D.; Casanovas, J.; Cativiela, C.; Nussinov, R. *J. Phys. Chem.* **2006**, *110*, 21264.

5 Discussion of the Results

The major aims of this work are: (i) the understanding of the conformational preferences of synthetic proline derivatives; (ii) the design of new proline derives useful for well-defined applications within the field of nanobiology, and (iii) the "in vitro" application of the engineered proline derivative to demonstrate the reliability of the design process. These aims have been led to divide the presentation of the results achieved in this work in two different chapters. The first one involves the conformational study using DFT methods of different proline derivatives that were obtained by introducing chemical modifications on the pyrrolidine ring (dehydroproline analogs) and the amino acid backbone (α-alkyl-proline analogs) or by incorporating new substituents (aminated proline analogs). The second chapter of results presents the molecular engineering of a new proline derivative that has been subsequently used to protect a very efficient tumor-homing peptide from the attack of proteases..

Initially, a highly constrained family of proline analogs, which was obtained by introducing one or more double bonds in the cyclic side chain, was studied. Specifically, the dehydro-derivatives of proline, which incorporate a single insaturation in the pyrrolidine ring, investigated in this work were: $\Delta^{\alpha,\beta}$Pro (double bond betwee C^{α} and C^{β} atoms), $\Delta^{\beta,\gamma}$Pro (double bond between C^{β} and C^{γ} atoms) and $\Delta^{\gamma,\delta}$Pro (double bond between C^{γ} and C^{δ} atoms). Furthermore, the analog with two double bonds located between the C^{α} and C^{β} atoms and C^{γ} and C^{δ} atoms (Py) atoms was also examined. The objective of this study was to predict if this type of constraint enhance the ability of proline to adopt turn motifs. Results evidenced that the insaturations not only reduced the intrinsically low conformational flexibility of proline but also affect to the values adopted by the backbone dihedral angles φ,ψ in the minimum energy conformations. Interestingly, the $\Delta^{\alpha,\beta}$Pro and Py derivatives show striking conformational

properties that are consequence of the electronic effects produced by the side chain – backbone conjugation. Thus, these effects induce the formation of planar intramolecular hydrogen rings, which have not reported in any other synthetic nor natural amino acid. As expected, the restriction imposed by a double bond in the pyrrolidine ring produce a remarkable destabilization of the *cis* arrangement of the amide bond involving the pyrrolidine nitrogen. However, this effect disappears when the second double bond is incorporated to the pyrrolidine ring. Thus, conformations with the peptide bond arranged in *trans* and *cis* resulted isoenergetic for the Py derivative.

In the next step, a detailed conformational study of two α-substituted proline analogs was developed. Specifically, we used DFT calculations to examine the intrinsic preferences of α-methylproline (αMePro) and α-phenylproline (αPhPro). It should be noted that α-substituted derivatives are of great interest because a large number of papers and patents dealing with the incorporation of α-tetrasubstituted proline analogs into bioactive peptides and other biologically relevant systems evidenced the potential of these type of synthetic amino acids. Results showed that the restrictions introduced by the replacement of the α hydrogen by a more bulky group also destabilize the *cis* configuration of the peptide bond involving the pyrrolidine nitrogen. A distinctive structural feature identified for both αMePro and αPhPro is the stabilization of the semi-extended polyproline II conformation, which is clearly due to the C^{α}-tetrasubstitution. Thus, the ε_L was identified as an energy minimum for both αMePro and αPhPro but not for the natural amino acid. Furthermore, the results obtained provide evidence that the distinct steric requirements of the substituent at C^{α} may play a significant role in modulating the conformational preferences. Thus, although the γ-turn was found to be the lowest energy minimum of both αMePro- and αPhPro-containing dipeptides, independently of the environment, the α-helical conformation was accessible for the former but not for the latter.

On the other hand, the intrinsic conformational properties of aminated derivates of Pro (Amp), which were constructed by incorporating the amino functional group at the C^{β}- or C^{γ}-positions of the pyrrolidine ring (β*t*Amp / β*c*Amp and γ*t*Amp / γ*c*Amp, respectively, where *c* / *t* refer to the *cis* and *trans* isomers), were investigated because they can be considered as an alternative to the

conventional hydroxylated proline derivatives, *i.e.* the amino and hydroxyl groups can act as hydrogen bonding donors and acceptors. Interestingly, as was previously found for the other families of proline derivatives, the incorporation of the amino group in the pyrrolidine ring reduced the conformational flexibility of the amino acid. Specifically, the γ-turn was found to be the only accessible minimum of the four aminated prolines with the peptide bond involving the pyrrolidine nitrogen atom in *trans*, independently of the β- or γ-position of the amino group and its *cis* or *trans* isomerization state in the cycle. However, comparison between Amp and α-alkylated and dehydro proline derivatives reveals a very important difference. Thus, the stability of the conformations with the peptide bond in *cis* was, in general, higher for Amp than for conventional proline. This is because the amino substituent forms intramolecular hydrogen bonds, in which the nitrogen atom acts as an acceptor, when the peptide bond adopts such conformation. A better understanding of the role of the hydrogen bond has been provided by extending the conformational study to the intrinsic properties of the corresponding dimethylaminoproline (Dmp) analogs: β*t*Dmp, β*c*Dmp, γ*t*Dmp and γ*c*Dmp. As expected, the elimination of the side chain···backbone hydrogen bonds reduces the drastic conformational restrictions found in the Amp analogs. However, the stability of the conformers with the peptide bond in *cis* was similar for Dmp and Amp derivatives, this type of structures being the most favored when the polarity of the environment is slightly higher than that of a carbon tetrachloride solution.

Obviously, the conformational properties of the Amp analogs can be drastically altered by the proteolitic equilibrium of the side amino group, which will be particularly relevant in aqueous solution. This feature led to examined the conformational preferences of the four protonated Amp (AmH^+p) derivatives: β*t*AmH^+p, β*c*AmH^+p, γ*t*AmH^+p and γ*c*AmH^+p. Results evidences that the protonation of the amino side group produces a reduction of the backbone conformational flexibility and a des destabilization of the *cis* configuration of the amide bond involving the pyrrolidine nitrogen. Interestingly, the isomers with the amino group disposed in *cis* were the most stable independently of the ionization state because of the stabilizing effect produced by attractive side chain···backbone

intraresidue interactions. In contrast, the analogs with the side group disposed in *trans* become destabilized by the protonation of the amino side group.

After investigate the conformational preferences of some proline derivatives, a relevant application in the field of the nanotechnology was developed. Specifically, the resistance against proteases of CREKA, a tumor-homing peptide recently discovered, has been developed using a targeted replacement strategy by a synthetic proline derivative. Synthesized CREKA peptide labeled with an attached fluorescent dye has been found to be detectable in human tumors from minutes to hours after intravenous injection, while it is essentially undetectable in normal tissues. However, in spite of the potential interest of CREKA in cancer diagnostics and therapeutics, the application of this and other homing peptide sequences was endangered by short half life times; after injection endogenous proteolytic enzymes rapidly digested the peptides. Thus, protection from proteases is an important step in the development of potential applications of tumor-homing peptides. It is worth noting that, among the several strategies proposed to protect proteogenic peptide sequences from proteases; among these, targeted replacements with synthetic amino acids have been the most successful. Thus, selective incorporation of synthetic amino acids induces a significant resistance against proteases not only at the mutated position but also at neighboring amino acids.

The first step prior to the design of synthetic analogs which incorporate non-proteogenic amino acids, consists of the identification of the conformational profile and, if possible, the bioactive conformation of the peptide. This was performed for CREKA using computer simulation methods based on Molecular Dynamics (MD), which allowed characterize the conformational profile of the CREKA sequence. Specifically, the procedure followed in this investigation was based on a multiple conformational search strategy considering the three experimentally tested environments: free peptide, peptide attached to a nanoparticle, and a peptide inserted in a phage display protein. The influence of the ionic strength on the conformational preferences of CREKA was also examined. Results indicated that the conformational profile of the REKA sequence is very similar in allcases, the only difference being the Cys residue. This coincidence allowed to conclude that the conformational profile of the peptide is independent of the chemical environment, *i.e.* the conformational

preferences of the peptide are very similar when it is free or attached to a nanoparticle. In the bioactive conformation proposed for CREKA, the backbone defines a β type turn motif. The functionalized side chains of the central residues (Arg, Glu and Lys) face the same side of the molecule, and the backbone Cys CO and Lys NH groups are hydrogen bonded. Inspection of the side chains reveals salt bridges involving the negatively charged carboxylate group of Glu and the contiguous positively charged side chains of Arg and Lys. These side-chain interactions are made possible by the peptide backbone turn conformation.

Finally, the last part of this Thesis presents the targeted replacement strategy followed to protect CREKA from proteolytic cleavage. Specificially, a proline-like derivative of arginine, denoted as (Pro)hArg, was designed, parameterized and incorporated into CREKA, replacing Arg. The conformational flexibility of the CREKA analogue incorporating (Pro)hArg was found to be smaller than that of natural CREKA, which was attributed to the conformational restrictions imposed by the cyclic residue. On the other hand, although the presence of this arginine surrogate produces some alterations in the interaction patterns, the most important features proposed for the bioactive conformation of CREKA, *i.e.* a turn motif orienting the charged side chains towards the same side of the molecule, were retained in the (Pro)hArg-containing analogue. The shape of the lowest energy conformation of the modified peptide showed a significant degree of similarity with that proposed for the bioactive conformation of the natural peptide. Thus, our results suggested that the incorporation of (Pro)hArg into CREKA as an arginine substitute could be useful to protect the peptide from proteolytic attack while retaining, or even enhancing, those structural features of the parent peptide considered to be essential for its tumor-homing activity.

6 Conclusions

(1) The modification of conventional proline by introducing double bonds or an amino group in the pyrrolidine ring or by replacing the α-hydrogen atom by an alkyl group has an impact on the conformational flexibility, which is severely reduced. In the case of aminoproline derivatives, the conformational flexibility decreases by protonating the amino side group.

(2) Conjugation effects between the backbone and the side chain have been detected in the α,β-unsaturated proline analog and the proline incorporating two double bonds.

(3) The restrictions introduced by the incorporation of a single double bond in the pyrrolidine ring and the substitution of the α hydrogen atom by an alkyl group produce a destabilization of the *cis* configuration of the peptide bond involving the pyrrolidine nitrogen.

(4) The incorporation of two double bonds located between the C^{α} and C^{β} atoms and C^{γ} and C^{δ} atoms increases significantly the probability of having a *cis* peptide bond.

(5) The incorporation of an amino or dimethylamino substituent in the pyrrolidine ring stabilizes the conformations with the peptide bond involving the pyrrolidine nitrogen arranged in *cis*

(6) The protonation of the side group in aminoproline analogs destabilizes the conformations with the peptide bond arranged in *cis*. Accordingly, the isomerism of the peptide group in this family of proline derivatives could be controlled with the pH.

(7) Replacement of the α hydrogen in proline by a more bulky alkyl group results in the stabilization of the semi-extended polyproline II conformation. This conformation was identified as an energy minimum for both α-methyl and α-phenyl proline derivatives but not for the proteinogenic amino acid.

(8) The conformational preferences of α-tetrasubstituted proline derivatives depend on the particular nature and size of the substituent incorporated at C^{α}. This is particularly evident for the α-helical conformation which was identified as a minimum for the α-methylproline dipeptide but not for the α-phenylproline one.

(9) The aminoproline analogs with the side group attached in a *cis* disposition are energetically favored with respect to the isomers with the amino group in *trans*. This feature, which does not depend on the protonation of the amino group, is consequence of the attractive side chain···backbone interactions found in the *cis* isomers.

(10) Rigorous multiple conformational analyses of CREKA indicated that the conformational profile of the REKA sequence is very similar in the different cases investigated: the free peptide, the peptide attached to a nanoparticle and the peptide inserted in a phage display protein. This coincidence, which was particularly remarkable for the accessible minima, evidenced that the conformational profile of CREKA does not depend on the chemical environment.

(11) The bioactive conformation presents a β-turn motif, strong interactions involving the functionalized side chains of the central residues (Arg, Glu and Lys), which face the same side of the molecule, and the backbone Cys CO and Lys NH groups are hydrogen bonded.

(12) A new amino acid has been designed as a replacement for arginine to protect the tumor-homing pentapeptide CREKA from proteases. This amino acid, denoted (Pro)hArg, is characterized by a proline skeleton bearing a specifically oriented guanidinium side chain. This residue combines the ability of Pro to induce turn-like conformations with the Arg side-chain functionality.

(13) The substitution of Arg by (Pro)hArg in CREKA produces a significant reduction of the conformational flexibility. Although some changes are observed in the interaction pattern, the peptide analog exhibits a strong tendency to accommodate turn conformations centered at the (Pro)hArg residue and the overall shape of the molecule in the lowest energy conformations characterized for the natural and the modified peptide exhibit a high degree of similarity. These results suggest that replacement of Arg by (Pro)hArg in CREKA may be useful in providing resistance against proteolytic enzymes while retaining conformational features which are essential for tumor-homing activity.