

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tesisenxarxa.net) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tesisenred.net) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tesisenxarxa.net) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author

GUILLEM REVILLA LOPEZ
UNIVERSITAT POLITECNICA DE CATALUNYA
MMXI

***A THEORETICAL APPROACH TO THE
ENGINEERING OF BIOINSPIRED SYSTEMS:
DESIGN, APPLICATIONS & INFORMATION
MANAGEMENT***

Presentat pel:

M.Sc. Guillem Revilla López

Sota la direcció de:

Dr Joan Torras Costa

Dr David Zanuy Gómara

Realitzada al Departament d'Enginyeria Química (DEQ) de l'Escola Tècnica Superior d'Enginyeria Industrial de Barcelona (ETSEIB) de la Universitat Politècnica de Catalunya (UPC)

A Barcelona el Novembre del 2010



Agraïments

Vet ací una de les parts més difícils d'escriure una tesis doctoral: el agraïments. Pensar i jutjar què i qui ha estat important al llarg d'aquest anys a la teva vida i que consti que no només parlo dels anys de tesis sinó també de molt abans, dels orígens de tot plegat, de qui t'ha encoratjat i ensenyat a seguir endavant, etc... Queda clar que són moltes les persones que han influït i deixat, d'una manera o d'una altra, la seva empremta. Per això mateix, segur que mai podré agrair a tothom qui s'ho mereix l'haver-nos creuat i fet una part del camí junts i, de ben segur, els que apareguin en els agraïments no rebran ni la desena part de la meitat del reconeixement que mereixen.

La primera injustícia és posar només el meu nom en aquest llibre, a molts se'ls ha d'agrair molta feina i els primers els meus dos directors: el Dr. Joan Torras Costa i el Dr. David Zanuy Gómara, sense el quals aquest treball hagués estat impossible.

Al Prof. Carles Alemán agraeixo sincerament haver fet possible aquesta tesis donant-me allò que a vegades sembla tan òbviament necessari i tot sovint és tan difícil de trobar: l'oportunitat. Gràcies pels inicis i pel suport constant.

Atès que fer un doctorat és també treballar en equip, no puc deixar passar l'ocasió d'expressar el meu agraïment a tots els membres del grup d'investigació IMEM. Puc dir que aquí no només he tingut companys de feina sinó amics, models i mestres: amb l'Elaine, el Jordi, l'Òscar, el Bruno (gracias por los discos de buena música portuguesa, ¡comegatos!), el David Aradilla (l'única cosa sobre la que no vull aprendre de tu és sobre el teu equip de futbol), en David Curcó, la Georgina (o l'art d'encaixar conyes) , l'Ester (mai hagués pensat que hi hagués algú amb gustos humorístics tan semblants al meus) , en Daniel i la Mar (tots dos comenceu ara però estic segur que demostrareu el que valeu) i en Francesc (els sopars amb ell eren una de les coses bones que portaven les primaveres) puc dir que els he gaudit com a companys. No només els presents sinó també els passats han estat importants: Al Dr. Rodríguez-Ropero (que estrany se'm fa dir-li així a en Frankie) li haig d'agrair haver pogut treballar, aprendre, prendre cafès i birres i visitar bars "veterans" amb ell.

No només de feina viu el doctorant, encara que no ho sembli, i per això no em puc oblidar de "la gente de arriba" amb els que un cafè, un dinar o sopar es converteixen sempre en quelcom memorable, surrealista i agradable. Abdel (Dr Allah disponible sempre per a una discussió excepte en opinions en fórmula 1 a on coincidim totalment ;)), Antxón (el Dr amb el cognom més llarg que conec i especialista en rescat de vehicles ficats en clots), Ainhoa (el teu avinagrament i la teva companyia han estat de

les coses bones d'aquest període), Cristina Lavilla (impressionantment brillant i a més a més amb sentit de l'humor, et dec bons consells per a la portada de la tesis i molt bones sobretalles) i els que ja no hi són per haver passat a millor vida (es a dir, per haver acabat la tesi): Robert (Dr Quintana, el belga), Nathalie (Dra Antibolivariana), Romina (Dra Miss ETSEIB). A tota aquesta gent "de arriba" gràcies per uns dinars que feien el dia millor independentment de com hagués anat el matí o de si el tema de la conversa tractava de "propietats enzimàtiques digestives" o de la primera bajanada o acudit de gran qualitat que em passava pel cap i que vosaltres aguantàveu. Aquí David (Zanuy i sense el Dr al davant) tornes a sortir tu. Gràcies a tots, gent.

It is absolutely necessary to express my gratitude to some people I have met during my stay abroad while doing the thesis or here in Barcelona during their long/short stays like Julien Preat, the Belgian "colchonero": taking some Belgian beers in Barcelona with you was quite an experience, Adèle: we have worked together just for a few weeks but I had enough time to value you as a hard-working researcher and a nice person. Gratitude must be also expressed to Prof Èric Perpete and Prof Dennis Jacquemin for hosting me during my stay in Namur. Y de Namur no puedo marcharme sin recordar a José (Dr. Cerón), y las andadas por tierras valonas y alemanas además de un compañero durante mi estancia en Bélgica. Por último, pero sin ser menos importante que ninguno de los anteriores, destaco al Sr Javi Giménez Warren con quien he tenido el gusto de compartir trabajo, calendario del doctorado y un tiempo maravilloso en Barcelona (razones futbolísticas hicieron que para él el tiempo fuera un poco menos agradable que para mí). Ahora que ya estamos acabando ¡ánimos chaval! .

Aquesta tesis s'ha fet amb (molt) treball, una part al laboratori (de tetris com diria l'Ainhoa) i una altra a un altre lloc els caps de setmana. D'aquesta experiència em quedo amb les persones que hi vaig trobar en especial la Noemí i la Daniela perquè m'han fet costat i he pogut comptar sempre amb elles.

Com que tot passa en un espai físic concret, no puc oblidar-me dels companys del departament d'Enginyeria Química, especialment de l'Elsa a qui vaig conèixer en començar el màster i des de llavors, i de trobada en trobada pel passadís, ha estat sempre un suport.

El último párrafo está reservado a las personas más importantes para mí. Cuando uno empieza un doctorado se supone que elige un grupo de investigación, un tema... En

definitiva, es “libre” de elegir. Lo que no se suele ver es que esta decisión implica que otras personas se vean matriculadas *de facto* en el mismo doctorado: aquellas que llevan años dándote ánimos, ayudándote, manteniéndote y siendo el refugio para cuando todo va mal se ven envueltas, aún sin quererlo, en el mismo “tinglado”. En mi caso éstos han sido mi familia, los que ya no están: abuelos Tomás y Engracia y los que han estado y están: iaios Dositeo y Antonia, mis padres Jesús y Montserrat y mi hermano Sergi. Si hay alguien que sin ser de la parte “técnica” debiera aparecer como autor son ellos. A mis abuelos por dejarse la piel trabajando para que otros pudiéramos tenerlo un poco más fácil. A mis padres por enseñarme que la mejor manera de subir una escalera es peldaño a peldaño teniendo claro dónde se pisa, valorando lo andado y con la vista puesta en lo que queda, y a mi hermano por ser mi hermano, por ser aquel en quien podía confiar y apoyarme.

Només em queda, per a acabar, dir que si alguna cosa bona ha sortit d'aquesta tesi és també mèrit vostre.

Gràcies a tots.

Piera, 22 de Novembre del 2010.

“Science is a very human form of knowledge. We are always at the brink of the known; we always feel forward for what is hoped. Every judgement in science stands on the edge of error and is personal: Science is a tribute to what we can know although we are fallible”

J. Boronowsky, extracted from a Gaussian09 quote.

“Bueno es saber que los vasos sirven para beber, lo malo es que no sabemos para qué sirve la sed”

Antonio Machado, “Proverbios y cantares”.

TABLE OF CONTENTS

1.Introduction	1
1.1. On the concept of nanotechnology and its history	3
1.2. On the importance of nanotechnology and its impact	3
1.3. On the applications of nanotechnology	5
1.3.1. <i>Nanoelectronics</i>	5
1.3.2. <i>Nanobiomedicine and bionanotechnology</i>	5
1.3.3. <i>Diagnostics</i>	6
1.3.4. <i>Protein engineering</i>	7
1.4. On the theoretical approach to nanotechnology	7
1.5. References	9
2.Objectives	15
3.Methodology	21
3.1. Computational chemistry	23
3.2. Quantum Mechanical Methods	23
3.2.1. <i>The Hamiltonian Operator</i>	24
3.2.2. <i>Basis sets</i>	24
3.2.3. <i>Ab initio calculations</i>	27
3.2.4. <i>Semiempirical calculations</i>	29
3.2.5. <i>Density Functional Theory calculations</i>	30
3.2.6. <i>Time dependent DFT calculations</i>	31
3.2.7. Solvent effects	32
3.3. Molecular mechanics (MM) simulations	34
3.3.1. <i>The force field</i>	34
3.3.2. <i>Molecular Dynamics (MD)</i>	36
3.3.3. <i>Thermodynamic Ensembles</i>	38
3.3.4. <i>Periodic boundary conditions and PME Summation</i>	39
3.3.5. <i>QM/MM calculations</i>	40
3.3.6. <i>Conformational exploration</i>	41
3.3.6.1. <i>Simulated Annealing (SA)</i>	42
3.3.6.2. <i>Replica Exchange Molecular Dynamics (REMD)</i>	42
3.4. References	44
4.Characterization of non-coded amino acids	49
4.1. Side-chain to backbone interactions dictate the conformational preferences of a cyclopentane arginine analogue	51
4.1.1. <i>Introduction</i>	51
4.1.2. <i>Methods</i>	53
4.1.3. <i>Results and discussion</i>	55
4.1.4. <i>Conclusions</i>	75
4.1.5. <i>References</i>	76

4.2. Conformational profile of a proline-arginine hybrid	81
4.2.1. <i>Introduction</i>	81
4.2.2. <i>Methods</i>	85
4.2.3. <i>Results and discussion</i>	88
4.2.4. <i>Conclusions</i>	101
4.2.5. <i>References</i>	102
4.3. Conformational Preferences of Proline Analogues with a Fused Benzene Ring	107
4.3.1. <i>Introduction</i>	107
4.3.2. <i>Methods</i>	110
4.3.3. <i>Results and discussion</i>	113
4.3.4. <i>Conclusions</i>	133
4.3.5. <i>References</i>	135
4.4. Phenylazophenylalanine: a key building block of photoresponsive biomimetic systems	141
4.4.1. <i>Introduction</i>	141
4.4.2. <i>Methods</i>	143
4.4.3. <i>Results and discussion</i>	146
4.4.4. <i>Conclusions</i>	163
4.4.5. <i>References</i>	164
5. Engineering of bioinspired systems	171
5.1. Exploring the Energy Landscape of a Molecular Engineered Analog of a Tumor – Homing Peptide	173
5.1.1. <i>Introduction</i>	173
5.1.2. <i>Methods</i>	176
5.1.3. <i>Results and discussion</i>	179
5.1.4. <i>Conclusions</i>	192
5.1.5. <i>References</i>	194
5.2. Atomistic modelling of peptides bound to a chemically active surface: conformational implications	199
5.2.1. <i>Introduction</i>	199
5.2.2. <i>Methods</i>	201
5.2.3. <i>Results and discussion</i>	206
5.2.4. <i>Conclusions</i>	215
5.2.5. <i>References</i>	216
5.3. Self-Assembly of a Designed Amyloid Peptide Containing the Functional Thienylalanine Unit	219
5.3.1. <i>Introduction</i>	219
5.3.2. <i>Methods</i>	221
5.3.3. <i>Results and discussion</i>	226
5.3.4. <i>Conclusions</i>	239

5.3.5. <i>References</i>	240
6.Design of a non-coded amino acid data base	247
6.1. NCAD, a data base integrating the intrinsic conformational preferences of noncoded amino acids	249
6.1.1. <i>Introduction</i>	249
6.1.2. <i>Methods</i>	252
6.1.3. <i>Results and discussion</i>	264
6.1.4. <i>Conclusions</i>	271
6.1.5. <i>References</i>	272
6.2.Integrating the intrinsic conformational preferences of non-coded α -amino acids modified at the peptide bond into the NCADB data base	277
6.2.1. <i>Introduction</i>	277
6.2.2. <i>Methods</i>	280
6.2.3. <i>Results and discussion</i>	289
6.2.4. <i>Conclusions</i>	295
6.2.5. <i>References</i>	296
7.Summary and discussion of the results	303
8.Conclusions	311

LIST OF ABBREVIATIONS

2-Thie	β -2-Thienylalanine
α MeInc	(α -methyl)indoline-2-carboxylic acid
Aib	2-amino-2-methylpropanoic acid (α -aminoisobutyric acid)
AB	Azobenzene
Ac	Acetamide blocking group
Ac ₃ c	1-aminocyclopropanecarboxylic acid
Ac ₄ c	1-aminocyclobutanecarboxylic acid
Ac ₅ c	1-aminocyclopentanecarboxylic acid
Ac ₆ c	1-aminocyclohexanecarboxylic acid
Ac _n c	Cyclic amino acid generalized formula
Adt	4-amino-1,2-dithiolane-4-carboxylic acid
AFM	Atomic Force Microscopy
AMBER	Assisted Model Building and Energy Refinement
AO	Atomic Orbitals
B3LYP	Functional of three parameters of Becke, Lee, Yang and Parr
B3PW91	Functional of three parameters of Becke, Perdew and Wang 91
B95	Functional Becke95
BMK	Boese and Martin τ -dependent functional
CD	Circular Dichroism
CSD	Cambridge Structural Data bank
Cryo-TEM	Cryogenic Transmission Electron Microscopy
<i>cis</i> -(β Pro)Arg	Cis isomer of 3-(guanidinomethyl)-pyrrolidine-2-carboxylic acid
<i>cis</i> -(γ Pro)Arg	Cis isomer of 4-(guanidinomethyl)-pyrrolidine-2-carboxylic acid
(1 <i>S</i> ,2 <i>S</i>)c ₃ Phe	(1 <i>S</i> ,2 <i>S</i>)-1-amino-2-phenylcyclopropanecarboxylic acid
(1 <i>S</i> ,2 <i>R</i>)c ₃ Phe	(1 <i>S</i> ,2 <i>R</i>)-1-amino-2-phenylcyclopropanecarboxylic acid
(2 <i>S</i> ,3 <i>S</i>)c ₃ diPhe	(2 <i>S</i> ,3 <i>S</i>)-1-amino-2,3-diphenylcyclopropanecarboxylic acid

(1 <i>S</i> ,2 <i>S</i>)c ₅ Arg	(1 <i>S</i> ,2 <i>S</i>)-1-amino-2-(guanidinomethyl)-cyclopentanecarboxylic acid
(1 <i>S</i> ,2 <i>R</i>)c ₅ Arg	(1 <i>S</i> ,2 <i>R</i>)-1-amino-2-(guanidinomethyl)-cyclopentanecarboxylic acid
(1 <i>S</i> ,2 <i>S</i>)c ₅ Phe	(1 <i>S</i> ,2 <i>S</i>)-1-amino-2-phenylcyclopentanecarboxylic acid
(1 <i>S</i> ,2 <i>R</i>)c ₅ Phe	(1 <i>S</i> ,2 <i>R</i>)-1-amino-2-phenylcyclopentanecarboxylic acid
(1 <i>S</i> ,2 <i>S</i>)c ₆ Phe	(1 <i>S</i> ,2 <i>S</i>)-1-amino-2-phenylcyclohexanecarboxylic acid
(1 <i>S</i> ,2 <i>R</i>)c ₆ Phe	(1 <i>S</i> ,2 <i>R</i>)-1-amino-2-phenylcyclohexanecarboxylic acid
conf.	Conformational
DB	Data Base
ERD	Entity Relationship Diagram
ΔE	Energy difference
ΔG	Gibbs energy difference
ΔG _{solv}	Solvation Gibbs energy difference
DFTB	Density Functional Tight Binding
ΔS	Entropy difference
ΔT	Temperature difference
Dbg	2-amino-2-benzyl-3-phenylpropanoic acid (dibenzylglycine)
Dpg	2-amino-2,2-diphenylacetic acid (diphenylglycine)
dc-aa	retro/inverso amino acid (dicarboxylate)
DFT	Density Functional Theory
dn-aa	retro/inverso amino acid (diamine)
FTIR	Fourier Transform Infra-Red spectroscopy
FMA	5(6)-carboxyfluorescein
GAFF	Generalized Amber Force Field
gp	Gas Phase
GTO	Gaussian Type Orbital
HF	Hartree-Fock
IMOMM	Integrated Molecular Orbital + Molecular Mechanics

IMOMO	Integrated Molecular Orbital + Molecular Orbital
GGA	Generalised Gradient Approximation
Inc	indoline-2-carboxylic
L-c ₃ Dip	(<i>S</i>)-1-amino-2,2-diphenylcyclopropanecarboxylic acid
LDA	Local Density Approximation
LSDA	Local Spin Density Approximation
LCAO	Linear Combination of Atomic Orbitals
LYP	Functional of Lee, Yang and Parr
MEP	Molecular Electrostatic Potential
MD	Molecular Dynamics
MM	Molecular Mechanics
MST	Mirteus-Scrocco-Tomasi solvation model
nc-aa	Non-Coded Amino Acid
NCADB	Non-Coded Amino Acid Data Base
NHMe	N-methylamide blocking group
NMR	Nuclear Magnetic Resonance
NVE	Microcanonical ensemble
NVT	Canonical ensemble
NPT	Isobaric, isothermal, constant number of molecules molecular dynamics
NPH	Isobaric, isoenthalpic, constant number of molecules molecular dynamics
PAP	Phenylazophenylalanine
PDB	Protein Data Bank
PME	Particle Mesh-Ewald summation
L-(α Me)Pro	(<i>2S</i>)-2-methylpyrrolidine-2-carboxylic acid [(α -methyl)proline]
L-(α Ph)Pro	(<i>2R</i>)-2-phenylpyrrolidine-2-carboxylic acid [(α -phenyl)proline]
QM	Quantum Mechanics

QM/MM	Quantum Mechanics/Molecular Mechanics
P	Functional Pedrew86
PBE	Functional of Pedrew, Burke and Emzerhof
PCM	solvation Polarizable Continuum Model
PL	Functional of Pedrew
REMD	Replica Exchange Molecular Dynamics
RESP	Restrained Electrostatic Potential
RMSD	Root Mean Square Deviation
SA	Simulated Annealing
SAXS	Small Angle X-ray Scattering
SCRF	Self-Consistent Reaction Field
SCF	Self-Consistent Field
SLBO	Strictly Localized Bond Orbitals
SPIO	Super Para-magnetic Iron Oxide
STO	Slater Type Orbital
<i>t</i> -L-c ₅ Arg	(1 <i>S</i> ,2 <i>R</i>)-1-amino-2-(guanidinomethyl)-cyclopentanecarboxylic acid
<i>trans</i> -(^β Pro)Arg	Cis isomer of 3-(guanidinomethyl)-pyrrolidine-2-carboxylic acid
<i>c</i> -L-c ₅ Arg	(1 <i>S</i> ,2 <i>S</i>)-1-amino-2-(guanidinomethyl)-cyclopentanecarboxylic acid
td-DFT	Time-Dependant Density Functional Theory
Toac	4-amino-2,2,6,6-tetramethylpiperidine-1-oxyl-4-carboxylic acid
UAff	United Atom force field
VSXC	van Horis and Scuseria τ -dependent gradient corrected correlation functional
VWN	Functional of Vosco, Wilk and Nusair
wat	Water
WT	Wild Type
X3LYP	Functional of three parameters of Xu, Lee, Yang and Parr
XRD	X-Ray Diffraction

ZPVE

Zero Point Vibrational Energy

1.Introduction

1.1. On the concept of nanotechnology and its history

The study of systems in a scale of nanometers is a keystone for understanding the properties of matter, not only the properties in such scale (hereafter nanoscale) but also in the macroscopic one. The disciplines involved in this field of science are diverse and they include chemistry, physics, material science, biology, and statistics among others. All these different fields converge in a central branch of science and technology called nanotechnology. The term nanotechnology was first coined by professor Norio Taniguchi in 1974 in a publication called "*On the basic concept of nano-tachnology*"¹. Although, the concept of nanotechnology is older. To be precise, it dates back from 1958 when Nobel prize-awarded Professor Richard Feynman gave a lecture² to the west coast section of the American Physical Society; this lecture was titled "*Plenty of room at the Bottom*". Prof. Feynman enounced the possibility of manipulating individual atoms in order to achieve self-assembling and self-reproducing devices. That constituted the very first idea of nanotechnology.³ Despite introducing the idea that atoms by themselves can store information² and form new devices that work like "little hands",² Prof. Feynman establish no consideration about the size of the scale that nanotechnology deals with. Professor Taniguchi set this limit around 1 nanometer.¹ Nowadays the most widely accepted interval ranges from 1 to 100 nanometers³ and not only it considers the formation of self-assembling and self-reproducing devices the main goal of nanotechnology, but also embraces a growing number of disciplines and technologies, like computational chemistry,⁴ that work at the aforementioned scale whichever their goal is.³

1.2. On the importance of nanotechnology and its impact

Despite its early beginnings, nanotechnology did not soar as a hotspot in modern science until years later.⁵ Its development was specially thrived by a key development: the invention of the Scanning Tunnelling Microscope (STM) in 1982,⁶ this innovation allowed to see, for the first time, an image of surfaces at an atomic level, this resolution neared the work scale to that defined by Prof. Feynman² and Prof. Taniguchi¹ as the optimal one for nanotechnology. Many other novelties arrived after STM, among them, the most outstanding were: synthesis of buckminsterfullerene C-60⁷, the finding of carbon nanotubes⁸, and the Atomic Force Microscope (AFM) invention⁹.

All these advances increased considerably the importance of nanotechnology, thus causing an upturn in its impact in our daily life. The influence of nanotechnology is clearly expressed in the growing number of marketplace products that nowadays contain “nanotechnology innovations”¹⁰ which totals 1,015 referenced products¹⁰. This great surge is also found when one looks at the public budgets for nanotechnology research and development: The U.S. government supports since 2000 the National Nanotechnology Initiative (NNI),¹¹ this program has provided funding for nanotechnology-related research projects for more than 14,000 million USD¹² from 2000 to 2011, being the budget for 2011 of 1,761.6 USD;¹² the European Union¹³ (EU) and Japan¹⁴ altogether with emerging countries¹⁶ also devote huge efforts to nanotechnology. This trend in public budget is followed and even overpassed by the business sector investments. This conjunction makes a global spending of 3,000 million USD in 2003¹⁵ and 9,000 million USD in 2005.¹⁹ All expenditures resulted in a dramatic growth of publications and made patents: rocketing from 224 non-overlapping patent applications in 1991 to 10,067 in 2008 (representing at that year a 1% of the total patent applications)¹⁶. Thus, in the same period of time the total number of referenced scientific publications found in ISI Web of Science¹⁷ that do have the word *nanotechnology* in their title or as a topic passed from 3,630 (year 1991) to 7,192 (year 2008).

The phenomenal increase in the importance of nanotechnology, clearly shown in its faster-than-ever-growing applications (see number of patents¹⁶ and commercial products¹⁰), has raised fears among the scientific community about the safety of some nanotechnological innovations.¹⁸ One of the first concerns about nanotechnology was the possibility of introducing self-replicating devices inside living beings³, this resembles a “new parasitic life-form”³ that needs to be controlled. These first fears about nanotechnology were soon surpassed by other emerging concerns about nanotechnology products such as carbon nanotubes (CNTs) and nano-onions.¹⁹ On the other side, nanotechnology-product-related toxicity remains unlighted since the products involved in it are of recent creation. Nonetheless, the works on toxicokinetics and toxicodynamics in the long and the short term are growing in number and significance.²⁰ The environmental impact of nanotechnology²¹ products is another key fact when we consider the social impact of nanotechnology and it is closely correlated with the toxicity of nanotechnology although nanotechnology also plays a key role in bioremediation.^{22,23,24,25} All the aforementioned problems have led

the scientific community as well as official regulatory agents to stress the special importance of carrying on further research^{20,21} to assess potential hazards and implementing new ethical codes for nanotechnology.²¹

1.3. On the applications of nanotechnology

The impact of nanotechnology in our daily life has grown quickly thanks to the number of applications available. These applications belong to daily life products although it is far more remarkable the role that nanotechnology plays in high added value products in electronics, biomedicine, and biotechnology among others. Some of the most relevant applications are made in fields such as nanoelectronics, nanobiomedicine, and bionanotechnology.

1.3.1. Nanoelectronics

Electronics has been focused mainly on two goals since its creation: increasing the power of the devices and reducing their size. These two streamlines produced every time new smaller components smaller and more powerful than the previous ones. Nonetheless, this race found a serious drawback as soon as the size of some components reached the 100 nanometre border. Below this critical size, quantum mechanics plays an important role and the main forces in macroscopic world have lower significance.² This was the time for a nanotechnology-based approach.²⁶

Information storage devices are probably the first one to use nanoelectronics and being benefited from it. The reason lies on the ever-growing demand for information mass storage memories constrained, at the same time, by room availability. The use of carbon nanotubes (CNTs) as electromechanical devices²⁷ in non-volatile memory allow CNTs to work like switch elements²⁶ or field-effect transistor (FET).²⁶ This application produces new devices with memory densities around several terabits per square centimetre although this density depends on the spatial arrangement of CNTs.²⁸ Other molecules have also been tested as single-molecule based memories since it can be assumed that the size of the molecule is the limit for the physical condensation of a bit of information²⁹. Although the results have not been as promising as with CNTs except for nanowires.³⁰

1.3.2 Nanobiomedicine and bionanotechnology

Bionanotechnology (or nanobiotechnology) and nanobiomedicine have gone hand-in-hand since their beginnings, the research projects labelled “nanobio” attracted more than a half¹⁵ of the venture capital funding raised between 1998 until 2003. This

provides us with an idea of the importance and strength of these particular branches of nanotechnology.

Bionanotechnology began at the same time that modern biotechnology: both disciplines work with elements of the same scale (from one to several nanometers) although, biotechnology initially paid little attention to neither the details neither in the nanometer scale nor to engineering-based focus.³¹ The dawn of modern biotechnology can be set in 1973 when the first *in vitro* bacterial plasmid was built,³² opening thus the possibility for mass production of recombinant proteins, one of the first successfully cloned and commercially exploited proteins was recombinant human insulin in 1978.³³

This production of proteins was tightly related to commercial biomedical applications in different fields and nanotechnology came to supply an engineering vision of biotechnology and biomedicine that boosted the number of applications and the number of fields in which they could be used.³⁴ Some of the most outstanding fields revolutionized by nanotechnology are diagnostics and protein engineering.

1.3.3 Diagnostics

The usefulness of a diagnostic method falls on the ability of a detector to interact with a certain substance that reveals the presence and/or state of a physiological condition. Given this definition, it seems clear that details in the nanotechnology scale for both detector and detected substance are of vital importance since the atomic/molecular factors will be the leading ones in the interaction. Quantum Dots (QDs) are key detection tools³¹ for DNA and proteins³⁵ thanks to the changes in its absorption spectra when they bind to a ligand,³¹ these properties have been used to build planar arrays that allow multi detection.³¹

Probe-attached CNTs-based electrodes are also of relevant importance in the diagnostics field since they show an electric variation as soon as the probe interacts with its ligand³⁶. Due to the huge variety of feasible probes³⁷ the applications for these devices are numerous and keep growing. Furthermore, the success of the aforementioned electrodes lead to the development of other non-carbon made nanotubes-based electrodes which also have excellent performances in the same field.³⁸

Moreover, imaging becomes crucial for diagnosing since a high resolution image helps assessment and detection of physiological dysfunctions. For instance, cysteine-arginine-glutamate-lysine-alanine homing peptide³⁹ (CREKA) has a remarkable

interest in this field thanks to its high affinity for tumour blood vessels where it can cause clotting.³⁹ This high affinity towards tumours can be used as an imaging probe when the peptide is attached to a SPIO nanoparticle³⁹ and as a therapeutic tool when the peptide is bound to nanoliposomes as a coating.³⁹

1.3.4 Protein engineering

Proteins and peptides are of the utmost importance in bionanotechnology and nanobiomedicine due to the significant functions they do in living organism. Besides, protein and peptides are a part of a number of nanosystems where their interactions with ligands strongly affect the mechanism of action.^{36, 37, 38, 39} These interactions are allowed or hindered by the 3-dimensional structure of the protein, the modification of this structure strongly affects the functionality of the protein and, thus, the manipulation of protein structure reveals itself as one of the most promising fields in nanotechnology since these changes may modulate activity or even create new functions.

Amino acids form the primary structure of proteins, which folds to form secondary structure that adopts a spatial distribution called tertiary structure. Some proteins may have a supra tertiary organization called quaternary structure. This hierarchy of protein structures highlights the influence of the amino acid sequence on the downstream protein conformations.⁴⁰ Therefore, this leading role of amino acids can be used to modulate protein structure by replacing natural-occurring amino acids by other natural-occurring amino acids or by non-coded (non-naturally found) amino acids.^{41, 42, 43}

1.4. On the theoretical approach to nanotechnology

The quick take-off of nanotechnology in recent years^{10,16,44} has been accompanied by an increase in expenditure,^{15,19} although there is a limit to the amount of money devoted to this end from both public and business sectors. This budgetary constraint let us know that refinement and innovation is also required in the proceedings we use to develop knowledge and products in nanotechnology. We need to be able to foresee the potential applications of a certain discovery or, conversely, discard it as non-feasible based on solid, reproducible, reality-reproducing, and trustable methods. In the face of these endeavours, the theoretical approach to nanotechnology must be taken into account.

Among different theoretical approaches that can be found, theoretical chemistry based on quantum and classical formalisms is probably the most outstanding one. This family of *in silico* methods embraces several simulations to understand the structure and function of matter at atomic and nanometre scales: quantum mechanics (QM), molecular mechanics (MM), and molecular dynamics (MD) if we consider the dimension of time.

QM methodology holds the highest level of accuracy describing atoms and molecules behaviour, although it has a serious limitation at increasing system size. Thus, systems made of several dozens of atoms are the only realistically calculable ones, being bigger systems computationally unattainable by this method. In spite of this drawback, this formalism is specially useful for characterizing small molecules such as amino acids,⁴⁵ conducting polymer's monomers,⁴⁶ dyes,⁴⁷ or even short oligomers.^{48,49,50} This tool has two key focus for this thesis: DFT^{45,46} and TD-DFT.⁴⁷

Classical formalisms, such as MM and MD do not depict the electronic nature of the atomic level as QM does, but it has a great advantage over the latter. MM and MD, the time-considering variant of MM, methods can cope with systems far bigger than QM, reaching up to several hundred thousands atoms. This enables these techniques to realistically reproduce the properties of big amounts of matter such as solvents,⁵¹ biological⁵² and synthetic⁵¹ membranes and also entire proteins.⁵³ This increment in the amount of atoms is due to a correct treatment of the electronic nature of atoms. In order to better fit reality, some of the parameters used in the classical formalism can be derived from experiments and/or QM.⁵⁴ Sometimes, it results greatly useful to combine both QM and MM in the so called QM/MM methods. This technique treats one part of the system of interest with QM formalism whereas the rest is treated with MM formalism; this allows better reproductions of phenomena that can just be studied by QM when there are embedded in bigger systems that are just attainable at by MM.

The main conclusion one can obtain from the extensive bibliography on approaches to nanotechnology, just scarcely referenced here, is that theoretical methods, specially those based on MM and QM are major *in silico* tools that fit reality properly saving time and economic resources for research and also contribute significantly to a better understanding and knowledge of nanotechnology.

1.5. References

- [1] Taniguchi N. “On the Basic Concept of Nano-Technology”, **1974**, *Proc. Intl. Conf. Prod.* Part II, Japan Society of Precision Engineering, Tokyo.
- [2] Feynman R. P. “Plenty of room at the Bottom”, **1959**, Presentation to the American Physical Society.
- [3] Drexler E. K. “Nanotechnology: From Feynman to Funding”, **2004**, *Nanotechnology: Risk, Ethics and Law*, Pages: 25-34.
- [4] Rodríguez-Ropero F. “Application of molecular simulation techniques to the design of nanosystems”, **2009**. Doctoral thesis. Department of chemical engineering, Universitat Politècnica de Catalunya (UPC).
- [5] Drexler K. E. “Nanosystems: Molecular Machinery, Manufacturing and Computation”, **1992**, John Wiley & Sons Inc., New York.
- [6] Binning G., Rohrer H., Gerber C. *Phys. Rev. Lett.*, **1982**, 49, 57.
- [7] Kroto H. W., Heath J. R., O’Brien S. C., Curl R. F., Stanley R. E. *Nature*, **1985**, 318, 162.
- [8] Iijima S. *Nature*, **1991**, 354, 56.
- [9] Binning G., Quate C. F., Gerber C., Weibel E. *Phys. Rev. Lett.*, **1987**, 56, 930-
- [10] “The project on Emerging Nanotechnologies: inventories” [on line] [consulted on June 14th **2010**] <http://www.nanotechproject.org/inventories/>
- [11] “National Nanotechnology Initiative: History” [on line] consulted on June 15th **2010**] <http://www.nano.gov/html/about/history.html>
- [12] “National Nanotechnology Initiative: Funding” [on line] consulted on June 15th **2010**] <http://www.nano.gov/html/about/funding.html>
- [13] “CORDIS: Nanotechnology: EU Funding” [on line] consulted on June 15th **2010**] http://cordis.europa.eu/fp7/home_en.html
- [14] Kishi T., Bando Y. *Nat. Mat.* **2004**, 3, 129-131
- [15] Paull R., Wolfe J., Hébert P., Sinkula M. *Nat. Biotechnol.* **2003**, 21, 1144
- [16] Dang Y., Zhang Y., Fan L., Chen H., Roco M. C. *J. Nanopart. Res.* **2010**, 12, 687-706
- [17] “ISI Web of Knowledge” [on line] [consulted on June 15th **2010**] <http://science.thomsonreuters.com/>
- [18] Eckelman M. J., Zimmerman J. B., Anastas P. T. *J. Ind. Ecol.* **2008**, 12, 316

- [19] Ding L., Stilwell J., Zhang T., Elboudwarej O., Jiang H., Selegue J. P., Cooke P. A., Gray J. W., Chen F. F. *Nano Lett.* **2005**, 12, 2448-2464
- [20] Hoet P., Legiest B., Geys J., Nemery B. *Drug Safety.* **2009**, 12, 625-636
- [21] Matsuda M., Hunt G. *Bio-Med. mater.Eng.* **2009**, 19, 259-267
- [22] Ahuja S. K., Ferreira G. M., Moreira F. *Crit. Rev. Biotechnol.* **2004**, 2-3, 125-134
- [23] Liu W. T. *J. Biosci. Bioeng.* **2006**, 1, 1-7
- [24] Duran R. *R.C.SueloNutr. Veg.* **2008**, 3, -
- [25] Gelman D. L., Wolf D. G. *Journal of the Homeland Security and Emergency Management.* **2009**, 1, -
- [26] Lu W., Lieber C. M. *Nat. Mater.* **2007**, 6, 841
- [27] Rueckes T., Kim K. *Science.* **2000**, 289, 94-97
- [28] Strukov D. B., Likharev K. K. *Nanotechnology.* **2005**, 16, 137-148
- [29] Heath J. R., Kuekes P. J., Snider G. S., Williams R. S. *Science.* **1998**, 280, 1716
- [30] Waser R, Aono M. *Nat. Mater.* **2007**, 6, 833-840
- [31] Fortina P., Kricka L. J., Surrey S., Grodzinsky P. 2005, *TrendsBiotechnol.*, 24, 168-173
- [32] Cohen S. N., Chang A. C. Y., Boyer H. W., Helling R. B. **1973**, *P. Nat. Acad. Sci. USA*, 70, 3240-3244
- [33] Villakomaroff L., Efstratiadis A., Broome S., Lomedico P., Tizard R., Naber S. P., Chick W. L., Gilbert W. **1978**, *P. Nat. Acad. Sci. USA*, 75, 3727-3731
- [34] Emerich D. F., Thanos C. G. **2006**, *Biomol. Eng.*, 23, 171-184
- [35] Gao X., Chan W. C. W., Nie S. **2002**, *J. Biomed.l Opt.*, 7, 532-537
- [36] Li J., Ng H. T., Cassell A., Fan W., Chen H., Koehene J., Han J., Meyyappan M. **2003**, *Nano Lett.*, 3, 597-602
- [37] Wang J., Liu G. D., Jan M. R. **2004**, *J. Am. Chem. Soc.*, 126, 3010-3011
- [38] Qi P. F., Javey A., Rolandi M., Wang Q., Yenilmez E., Dai H. J. **2004**, *J.Am.Chem.Soc.*,126, 11774-11775
- [39] 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. **2007**, *P. Nat. Acad. Sci. USA*, 104, 932
- [40] Nelson D. L., Cox, "Lehninger Principles of Biochemistry", **2007**, 5ed
- [41] Solanas C., de la Torre B. G., Fernandez-Reyes M., Santiveri C. M., Jimenez M. A., Rivas L., Jimenez A. I., Cativiela C. **2010**, *J.Med. Chem.*, 53, 4119-4129

- [42] Zanuy D., Ballamo G., Jimenez A. I., Casanovas J., Haspel N., Cativiela C., Curco D., Nussinov R., Aleman C. **2009**, *J.Chem.Inf.Model.*, 49, 1623-1629
- [43] Ballamo G., Zanuy D., Jimenez A. I., Cativiela C., Nussinov R., Aleman C. **2009**, *J.Phys.Chem. B*, 41, 13101-13115
- [44] Takeda Y., Mae S., Kajikawa Y., Matsushima K. **2009**, *Scientometrics*, 80, 23-38
- [45] Rodriguez-Ropero F., Zanuy D., Casanovas J., Nussinov R., Aleman C. **2008**, *J.Chem.Inf.Model.*, 48, 333-343
- [46] Carrion S., Rodriguez-Ropero F., Aradilla D., Zanuy D., Casanovas J., Aleman C. **2010**, *J.Phys.Chem. B*, 114, 3494-3499
- [47] Perpete E. A., Jacquemin D. **2009**, *J.Mol. Struct.-Theochem*, 914, 100-105
- [48] Bertran O., Pfeiffer P., Torras J., Armelin E., Estrany F., Aleman C. **2007**, *Polymers*, 48, 6955-6964
- [49] Bertran O., Armelin E., Estrany F., Gomes A., Torras J., Aleman C. **2010**, *J. Phys. Chem. B*, 114, 6281-6290
- [50] Torras J., Bertran O., Aleman C. **2009**, *J. Phys. Chem. B*, 113, 15196-15203
- [51] Chang K. S., Huang Y. H., Lee K. R., Tung K. L., **2010**, *J. Membrane Sci.*, 354, 93-100
- [52] Cordomi A., Edholm O., Perez J. J., **2008**, *J.Phys.Chem. B*, 112, 1397-1408
- [53] Zanuy D., Flores-Ortega A., Jimenez A. I., Calaza M. I., Cativiela C., Nussinov R., Rouslathi E., Aleman C. **2009**, *J.Phys.Chem. B*, 113, 7879-7889
- [54] Preat J., Rodriguez-Ropero F., Torras J., Bertran O., Zanuy D., Aleman C. **2010**, *Journal of Computational Chemistry*, 31, 1741-1751

2.Objectives

(1) To design and characterize the conformational properties of arginine surrogates. Among them, the conformational properties of those derived from proline and from Ac₅c will be assessed. The role of the solvent polarity and the guanidium group over their structural preferences will be studied.

(2) To characterize structurally indoline carboxylic acid and α -methyl indoline carboxylic acid as proline surrogates and to study the influence of the used theoretical method in the energy assessment of conformational preferences.

(3) To elucidate the preferred conformations of phenylazophenylalanine (PAP) and its optical behavior. In order to correlate the optical properties with the conformational preferences, such features for the new amino acid will be compared with those for azobenzene.

(4) To improve the homing peptide activity of natural home peptide CREKA by selective replacement using non-coded engineered amino acid previously studied (C₅REKA). The homing activity of the synthetic peptide will be evaluated by exploring its conformational behavior under different conditions and compared to those of the wild type peptide.

(5) To simulate the effect of the hypercrowded environment over the conformational landscape of CREKA linked to different kinds of surfaces. The detected structural differences between isolation, crowded environment and surface nature are evaluated in terms of the conformational variability shown by CREKA in every case.

(6) To study the conformational dynamics of an engineered peptide. This new molecular hybrid is constituted by an engineered β -amyloid fragment that features thienylalanines instead of phenylalanines. The inner structural properties of the new material will be compared with experimental data in terms of conformational restrictions coming from NMR experiments.

(7) To develop a coherent systematic data base that compiles all available structural information of non-coded amino acids, which will be acquired from high level theoretical calculations and, if available, high resolution experimental data. The information will be graphically presented in terms of structural features of each amino acid, easing its use in material science and biomedicine.

3.Methodology

3.1. Computational chemistry

All science insight explored and studied in this thesis comes from the application of a well-known set of theoretical methodologies defined as computational chemistry. As previously seen, theoretical chemistry includes two main formalisms: quantum mechanics (QM) and molecular mechanics (MM). The main difference between these formalisms is the level of detail in the treatment of matter. This difference has its own expression in the amount of computational effort needed for systems of similar size.

The presented work of this thesis includes a methods section so this section must be understood as a general introduction to the aforementioned techniques. A good deeper focus on the question can be found in “Quantum Chemistry” by Ira N. Levine¹ and “Molecular modeling for beginners” by Alan Hinchliffe.²

3.2. Quantum Mechanical Methods

Quantum mechanics focus on the atomic level of matter by dealing with the motion of electrons under the influence of the electromagnetic force exerted by nuclear charges and other electrons. Most of the QM methods assume that nuclei and electron are distinguished from each other and both electron-electron and electron-nucleus interactions are explicitly described. These interactions lead the contribution to both kinetic and potential energies. These energies are computed in the Schrödinger equation (equation 3.1.). This equation treats electrons as if these exhibit dual wave-particle behaviour.

$$\hat{H} \psi = E \psi \quad (3.1.)$$

Where \hat{H} is the Hamiltonian operator containing the kinetic and potential energy of the nuclei and electrons, E is the energy of the system and ψ is the wave function that describes the molecular orbitals.

A wave function (ψ) and a probability function (ψ^2) can be found for monoelectronic atoms. Unfortunately, Schrödinger’s equation is just analytically solvable for those atoms having one electron. In case of polyelectronic atoms, one must use some simplifying approaches, for instance, considering the molecular wave function as a product of unielectronic Atomic Orbitals (AO’s)

3.2.1. The Hamiltonian Operator

The Hamiltonian operator (\hat{H}) of a polyatomic system, made of M nuclei and N electrons, can be expressed as:

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

where r_{ij} , r_{iA} , and R_{AB} are the distances between electron-electron, electron-nucleous, and nucleus-nucleus, respectively. M_A is the relation of the mass of the nucleus A with respect to the mass of the electron, Z_A is the atomic number of the nucleus A and ∇_i^2 and ∇_A^2 are the Laplace operators referred to the differentiation between the coordinates of the electron i and nucleus A , respectively. The first two terms of equation 3.2. show the kinetic energy of the electrons and nuclei respectively, the third term deals with the electrostatic attraction between nuclei and electrons, and the last two terms take into account, respectively, the electrostatic repulsion between electrons and nuclei. The second and fifth terms of equation 3.2. can be added as constant by taking into account the Born-Oppenheimer approach. This approximation states the motion of nuclei, when compared to the motion of electrons, as neglectable. This assumption renders a Hamiltonian operator such as:

$$\hat{H}_{el} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>1}^N \frac{1}{r_{ij}} \quad (3.3.)$$

3.2.2. Basis sets

Molecular orbitals can be represented as linear combinations of m basis functions (equation 3.4.).³

$$\psi(\varphi_s) = E(\varphi_s) \quad ; \quad \Psi_i = \sum_{s=1}^m c_{si} \varphi_s \quad (3.4)$$

where ψ_i is the i^{th} molecular orbital, φ_s is the s^{th} basis function and c_{si} are weighting coefficients that must be adjusted to get the best molecular orbital.

Linear Combination of Atomic Orbitals (LCAO) is the approach to generate molecular orbitals. In spite that these basis functions are not necessarily atomic

orbitals but any set of mathematical functions whose linear combination yields useful representations of molecular orbitals. The whole set of those mathematical functions that will be used to generate the proper molecular orbitals, is called basis set.

Slater functions (equation 3.5.) are one of those mathematical functions, which allow a good approximation to atomic wave functions. However, large computational time is required to obtain their LCAO representation. Using Cartesian Gaussian functions (equation 3.6.) reduce considerably the CPU time due to the mathematical properties of Cartesian Gaussian functions, these properties allow us to reduce the complexity of the integrals (interactions) to be calculated.

$$\varphi = \mathbf{a} \cdot \mathbf{e}^{(-br)} \quad (3.5.)$$

$$\varphi = \mathbf{a} \cdot \mathbf{e}^{(-br^2)} \quad (3.6.)$$

The total wave function ψ is expressed as several combinations of spin-orbitals. A spin-orbital is the product of a one-electron spatial orbital and a one-electron spin function. This simple approximation allows building total wave function which accomplishes with the Pauli's exclusion principle, which states as following: *No two electrons can occupy the same spin-orbital*. Thus, since the exclusion principle is based on approximate (zeroth-order) wave function, a determinant of the form (3.7), which satisfies the antisymmetry requirement for a many-electron atom, is called a Slater determinant. Consequently, the wave function of the ground-state and excite-state configurations for any polyelectronic system can be written as Slater determinants.

$$\Psi = \frac{1}{\sqrt{(2n)!}} \begin{vmatrix} 1s(1)\alpha(1) & 1s(1)\beta(1) & 2s(1)\alpha(1) & 2s(1)\beta(1) & \dots & nm_1(1)\beta(1) \\ 1s(2)\alpha(2) & 1s(2)\beta(2) & 2s(2)\alpha(2) & 2s(2)\beta(2) & \dots & nm_1(2)\beta(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1s(2n)\alpha(2n) & 1s(2n)\beta(2n) & 2s(2n)\alpha(2n) & 2s(2n)\beta(2n) & \dots & nm_1(2n)\beta(2n) \end{vmatrix} \quad (3.7.)$$

STO-nG basis sets

In spite of being easy to use due to their mathematical properties, Gaussian Type Orbitals (GTO) have 2 basic problems:

- The assessment of energy is often biased

- The shape of the orbitals does not fit STO. Unlike the Slater functions the Gaussian functions do not have a cusp at the nuclei and they are too short since they decay towards zero more quickly.

In order to combine both properties, easiness for calculations and realistic representation of orbitals, one can use a linear combination of Gaussian type orbitals (GTO) whose exponent constant (b) and contraction coefficient (a) (eq. 3.6.) are adjusted to make the whole orbital fit the shape of a desired STO.⁴ In this sense, one can distinguish between STO-3G, STO-4G or STO-6G basis sets, in which three, four, and six Gaussian functions have been used respectively.

Split-valence basis sets

As previously seen, STO-nG basis set show serious limitations. One of the most important is an incorrect assessment of energy. The rigidity of the minimal basis set does not allow incorporating any anisotropy from the molecular environment. The core orbitals, unlike the valence orbitals, do not affect chemical properties very much and vary only slightly from one molecule to another. Consequently, the split-valence approach increases the number of functions used to describe the valence electrons but keeping a single function for the inner shells.

In the present thesis one kind of split-valence basis sets has been used: Pople basis sets. Split-valence double-zeta Pople basis sets nomenclature follows the *K-MNG* scheme. *K* is the number of Gaussian functions used to represent the basis set functions of the core electrons, *M* and *N* indicate that each valence orbital is split into two parts being each part represented by *M* and *N* Gaussians respectively. Examples of these basis sets are 3-21G⁵ or 6-31G.⁶ Also, split-valence triple-zeta Pople basis set such as 6-311G, in which each valence orbital is split into three parts, have been used in the present thesis.

When atoms bind each other in order to form molecules, atomic charge distribution is often distorted from its assumed spherical form, specially for molecules with a strongly anisotropic charge distribution. This effect would be poorly depicted by the mentioned basis set since they do not allow electrons to move out further of that preconceived form. The addition of polarization functions let electrons move in a bigger room, thus being able to reproduce the distortion effect. These polarization functions have a higher angular quantum number from those atomic orbitals that is being adjusted. So, an extra p orbital for the hydrogen, d orbitals for the first- and

second-row elements, and f orbitals for the transition metals, indicated by * or ** symbols after the basis set, will be added.

Another drawback to be solved when considering a basis set is the inability to deal with those molecules that present a significant amount of electron density located away from the nuclei, such as anions and molecules with lone pairs. This is solved by adding diffuse functions, denoted using a “+” for heavy atoms and “++” when involve heavy and hydrogen atoms. These diffuse functions are Gaussian Cartesian functions and the difference between these and the rest of functions lie in their lower exponent constant (b in eq 3.6.). This causes a longer shape of the orbitals.

3.2.3. *Ab initio* calculations

Ab initio calculations are grounded on solving the Schrödinger equation (eq. 3.1.) without using any empirical adjustment, i.e. from first principles. Thus *ab initio* calculations are based only on basic physical theory. As we mentioned before, the poly-electronic Schrödinger equation is unsolvable due to the repulsion between electrons. The atomic orbital approximation (AO) tries to solve this problem by assuming that orbitals have the same shape and size in polyelectronic atoms and in the mono-electronic ones, this lead to a missassessment of energy. The Hartrees (father and son) were the first to suggest a model that takes into account the AO approximation and, at the same time, the repulsion between electrons. By assuming a smeared electric potential caused by the rest of electrons, the affection on the electron whose spin-orbital was calculated. This new approach did not consider Pauli’s principle (exchange energy) and was completed by Fock who introduced the Slater’s determinant; the new *ab initio* method was the Hartree-Fock (HF) method or Self Consistent Field (SCF) procedure.⁷ The HF wave function is approximated to a single Slater determinant of spin-orbitals (eq. 3.7.) which makes this function antisymmetrized (exchange is effectively considered). In the HF method spin-orbitals are split in a harmonic spherical part, an orbital part and a radial function, the radial function is optimized by minimizing the variational energy:

$$E(\psi) = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \quad (3.8.)$$

Since HF method assumes that every electron moves in an electrostatic field generated by the rest of the electrons, the assessment of the energy of the system is

biased. HF-SCF wave function takes into account the interactions between electrons only in an average way. Consequently, the instantaneous interactions between electrons (electron correlation) must be considered since electrons repel each other, they tend to keep out of each others's way. Electron correlation plays a key role in some important phenomena such as hydrogen bonding, π -stacking, chemical reaction (bond breaking and making) and aromatic systems among others. This limitation is overpassed by post-HF methods; these methods include different kind of corrections that reproduce electron correlation. Two of the most used post-HF methods are Møller Plesset (MP) procedure⁸ and Coupled Cluster (CC) method.⁹ Møller Plesset is based on perturbation theory. The philosophy behind MP methods is that a correction term handles electron correlation by promoting electrons from occupied to virtual molecular orbitals providing electrons with more room to move and thus making it easier for them to show correlation. The Hamiltonian operator is defined as the addition of a perturbation operator (\hat{V}) to the unperturbed HF Hamiltonian (\hat{H}^0):

$$\hat{H} = \hat{H}^0 + \lambda \cdot \hat{V} \quad (3.9.)$$

where λ is a dimensionless parameter taking values compressed between 0 and 1. Thus both wave function and energy are described by this perturbed Hamiltonian operator.

$$\psi = \psi^{(0)} + \lambda \cdot \psi^{(1)} + \lambda^2 \cdot \psi^{(2)} + \dots + \lambda^n \cdot \psi^{(n)} \quad (3.10.)$$

$$E = E^{(0)} + \lambda \cdot E^{(1)} + \lambda^2 \cdot E^{(2)} + \dots + \lambda^n \cdot E^{(n)} \quad (3.11.)$$

where $\psi^{(n)}$ is the n^{th} correction of the wave function related to the number of virtual spin orbitals and $E^{(n)}$ is the n^{th} correction energy term. Since HF energy is the sum of $E^{(0)}$ and $E^{(1)}$ terms, electronic correlation corrections are taken into account from the second term (MP2). The MP approach has a serious limit in the perturbation order since it consumes a lot of computer resources; the most frequent limit in the available commercial software is MP5. MP method shows problems for geometry optimization when the starting point is far away from the equilibrium, higher perturbation orders are necessary in those cases, for instance MP48 in extreme cases².

The Coupled Cluster method expresses the correlated wave function as a sum of the HF ground state determinant plus determinants representing the promotion of electrons to virtual molecular orbitals:

$$\Psi = \Psi_{\text{HF}} \cdot \sum_{k=0}^{\infty} \frac{\hat{T}^k}{k!} = e^{\hat{T}} \cdot \Psi_{\text{HF}} \quad (3.12.)$$

$$\hat{T}_1 \Psi_{\text{HF}}^0 = \sum_{a=n+1}^{\infty} \sum_{i=1}^{\infty} t_i^a \psi_i^a \quad (3.13.)$$

$$\hat{T}_2 \Psi_{\text{HF}}^0 = \sum_{b=a+1}^{\infty} \sum_{a=n+1}^{\infty} \sum_{j=i+1}^n \sum_{i=1}^{n-1} \psi_{ij}^{ab} t_{ij}^{ab} \quad (3.14.)$$

where $\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_n$ is the cluster operator and the operators \hat{T}_n are excitation operators and have the effect of promoting n electrons into virtual spin orbitals. It is widely accepted that double excitation operator ($n=2$) (eq.3.14.) represents most of the excitatory contribution, that is why coupled cluster double method, CCD, is among the most used ones altogether with CCSD (including single and double excitation) and CCSDT (single, double and triple excitations).CCS has been used as a rudimentary way to study single excitations. As the number of the considered excitations grows, so does the similarity between CC method and full CI.

3.2.4. Semiempirical calculations

Ab initio calculations offer reliability, as they pursue a solution based on the Schrödinger's equation. Unfortunately, a solution with results very close to the experimental values is very costly in terms of computational resources, specially for large systems. Since the early beginnings of theoretical chemistry, scientists on this field have tried to find ways to perform calculations lighter in terms of resource consumption. Semiempirical methods (SE) became the key to reduce the computational load by introducing the following approximations:

- Only valence electrons are considered for a *Ab initio* treatment.
- Minimal basis set is used for describing valence electrons. Integrals are calculated in a simpler way by using experimental data, analytical formulae or parameterization.

Parameters used in SE methods are adjusted to reproduce experimental data.

Nowadays, the most used family of semiempirical methods is MINDO family (Modified Intermediated Neglect of Differential Overlap), which is very reliable for geometry and reaction thermodynamics. Among this family, two of the most used methods are PM3,¹⁰ which follows the same formalism as the previous SE method AM1,¹¹ differences between them are found in the parameterization employed as well as in the way to describe functions for the core-core repulsion.

3.2.5. Density Functional Theory calculations

The Density Functional Theory (DFT) is based on the theorems postulated by Hohenberg and Kohn¹² which state that the ground state energy of a non degenerate electronic system is solely defined by an external potential, the electron density and a functional of the electron density. Besides, DFT is also based on the variation principle which enounces that the energy of any calculated density will always be higher than the energy of the actual density. In consequence DFT does not use the wave function but an electron probability density function $\rho(x,y,z)$, i.e. the probability of finding an electron in a volume element $dx dy dz$ centered on a point with coordinates (x, y, z) . Assuming r as the position vector of the point with coordinates (x, y, z) and taking into account Born and Pauli interpretation, which states that the square of a one electron wave function (Kohn-Sham orbitals, ψ_i in eq. 3.15.) at any point is the probability density at that point, the electron density can be written as follows:

$$\rho(\mathbf{r}) = \sum_{i=1}^n |\psi_i(\mathbf{r})|^2 \quad (3.15.)$$

where n is the number of occupied molecular orbitals. Electronic energy $E(\rho)$ is calculated as a simple sum of different contributions that depend on the electronic density:

$$E(\rho) = E^T(\rho) + E^V(\rho) + E^J(\rho) + E^{XC}(\rho) \quad (3.16)$$

being $E^T(\rho)$ the kinetic energy, $E^V(\rho)$ the term containing the potential attractive energy electron-nucleus and the repulsive term between nuclei, $E^J(\rho)$ the Coulombic repulsion between electrons, and $E^{XC}(\rho)$ the interchange-correlation energy. Usually $E^{XC}(\rho)$ is described as a sum of two terms; the exchange term, $E^X(\rho)$, and the

electronic correlation term, $E^C(\rho)$, depending on both terms of the electron density and its α and β spin electron density gradient.

$$E^{XC}(\rho) = E^X(\rho) + E^C(\rho) \quad (3.17.)$$

The exchange term can be calculated by using approximations that applies a homogeneous electron density, such as the Local Density Approximation (LDA)¹³ and the Local Spin Density Approximation (LSDA)¹⁴. The local density approximation performs quite well results. However, it has been shown inadequate for some problems. Thus, some new extensions has been developed, i.a., gradient corrected functionals such as the so called Generalised Gradient Approximation (GGA) methods, which depend upon the gradient of the density at each point in space and not just on its value. Examples of functionals using a homogeneous electron density are the VWN¹⁵ or the local correlation functional of Perdew (PL).¹⁶ Several GGA based functionals are reported such as Becke95 (B95),¹⁷ Perdew 86 (P),¹⁸ Perdew-Burke-Ernzerhof (PBE),¹⁹ Perdew-Wang 91 (PW91)²⁰ and the extremely popular Lee-Yang-Parr (LYP).²¹ The exchange-correlation term can also be calculated using the so called hybrid density functionals, which combine a conventional GGA method with a percentage of Hartree-Fock exchange. Examples of hybrid density functional that combine a part of the HF exchange with DFT exchange-correlation include B3LYP,^{21,22} B3PW91,^{20,22} MPW1K,²³ O3LYP,^{21,24} X3LYP^{20,21,25} and BMK.²⁶ B3LYP is one of the most used hybrid methods in this thesis due to its high correlation with experimental values, i.a., geometric and thermodynamic parameters. B3LYP obtains its energy from eq 3.18.

$$E_{XC}^{B3LYP} = E_{XC}^{LDA} + a_0(E_X^{HF} - E_X^{LDA}) + a_X(E_X^{GGA} - E_X^{LDA}) + a_C(E_C^{GGA} - E_C^{LDA}) \quad (3.18.)$$

Where E_{XC} is the exchange-correlation energy, E_C is the correlation energy and E_X is the exchange energy. The three different parameter (a_0 , a_X , a_C) are coefficients determined to fit experimental values of energy whose usual values are 0.2, 0.72 and 0.8 respectively.

3.2.6. Time dependent DFT calculations

After the introduction of DFT methods which, originally, were time independent, the elucidation of DFT for systems with densities evolving with time and position (Eq

3.18) was developed. It was E. Runge and E. K. U. Gross who, in 1984, enounced the four theorems that were the physical foundation of *Time Dependent Density Functional Theory* (hereafter TD-DFT).²⁷

- For any external potential, $v(\mathbf{r},t)$, that can be expressed as a Taylor series around t_0 (initial moment), the time-dependant Schrödinger equation can be solved if the initial state, $\varphi_0 = \varphi(t_0)$, is invariable, thus generating a unique density map.
 - There is a time-dependent density functional, which is parametrically dependant on \mathbf{r} and t .
 - Action integral (A) can be represented as a functional dependent on density, the density functional and time, $A(n,B,t)$.
-
- The time-dependence of electron density can be expressed as:

$$\rho(\mathbf{r}) = \sum_{i=1}^n |\psi_i(\mathbf{r},t)|^2 \quad (3.19.)$$

This method opens huge possibilities for studying any systems that is under influence of time-depending varying potential fields. These potential fields can be due to moving electrons, moving atoms or an electromagnetic wave like light. These possibilities make TD-DFT a major tool for the study of electron excitation and relaxation altogether with the accompanying events in chemical systems, that is light adsorption and emission. TD-DFT has been used in the present work for studying the spectroscopic properties of molecules.

3.2.7. Solvent effects

The solvent effects on the solute must be taken into account since many chemical and physical reactions and phenomena happen in condensed phases. There are two basic approaches to treat those solvent effects:

- Explicit solvent models. This approach takes into account all the local anisotropies due to solvent molecules by considering all or just a part of solvent molecules explicitly. It is probably the most accurate way but it is computationally inaccessible for large systems by using *ab initio* calculations. Explicit solvent are more frequent in MM and MD approaches.

- Implicit solvent models. This is the most frequent approach in *ab initio* calculations. To describe implicitly the solvent as it is done in the continuum models, i.e. while the solvent is described as an infinite dielectric medium, the solute is treated at the QM level. In the present thesis the so called Polarizable Continuum Model (PCM or MST) has been the most used methodology.²⁸ The PCM method generates spheres centered on each atom of the molecule on the surface of which point charges are added to simulate the polarization of the solvent due to the nuclear charges and electron density of the solute, i.e.. virtual point charges are created on the surface of the cavity to fit the derivative of the solute electrostatic potential calculated from the molecular wave function. PCM is a Self Consistent Reaction Field (SCRF) method that adjust virtual cavity charges altogether with solute charges .

All this process enables us to decompose the solvation energy into three components:

- Creation of the solute cavity inside bulk solvent.
- Generation of the van der Waals particle inside the cavity.
- Generation of the solute charge distribution in solution.

$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{vdW}} + \Delta G_{\text{elec}} \quad (3.20.)$$

where ΔG_{cav} is the work involved to increase the cavity, ΔG_{vdW} is the contribution due to the van der Waals interactions and ΔG_{elec} is the electrostatic component of the ΔG_{solv} and it corresponds to the work required in the polarization process. Interaction potential between solvent and solute is introduced in the solute electronic Hamiltonian as a perturbation operator (\hat{V}_R). Thus, the Schrödinger equation is expressed as follows:

$$\left(\hat{H}^0 + \hat{V}_R \right) \psi = E \psi \quad (3.21.)$$

where \hat{H}^0 is the solute's Hamiltonian operator.

3.3. Molecular mechanics (MM) simulations

3.3.1. The force field

The previously mentioned quantum mechanical methodology (QM) treats matter with a high accuracy level, considering atom nuclei and electrons. Unfortunately, those methods are highly resource-consuming in terms of computation. Thus, on large systems (over 100 atoms), both a big number of processors and a considerable amount of computing time are required. This situation makes large systems unrealistically attainable by these methods. Clearly, we need of new approaches to downsize the necessary amount of time and processors. Molecular mechanics (hereafter MM) comes to provide a powerful tool for systems made up to several hundred thousand atoms that do not require of the electronic properties study.

In MM potential energy of a given system is evaluated by a set of equations and parameters named “Force Field” (FF), and based on classical mechanics laws. MM considers atoms as non-deformable balls, and atomic bonds as springs with harmonic potential behaviour. Also, angle bond alterations from ideal values are also represented with harmonic potentials, whereas the torsion angle potential function is a rotation described by a dihedral angle and a coefficient of symmetry n , around the middle bond, as can be seen in eq. 3.22.

$$E(r) = \sum_{\text{bonds}} k_{\text{stretch}} (r - r_{\text{eq}})^2 + \sum_{\text{angles}} k_{\text{bend}} (\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} k_{\text{torsion}} [1 + \cos(n \cdot \phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right] + \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

(3.22.)

Equation 3.22. is the equation for the *Assisted Model Building and Energy Refinement* (AMBER) force field (FF).²⁹ $E(r)$ is the expression for the potential energy of an atom at position r . AMBER is the force field used for the MM simulations of this thesis. The terms included in the equation are:

- Bond stretching is the first term, that is bond elongation and contraction from its equilibrium value (r_{eq}). This potential is computed as a harmonic potential, though some other FFs used anharmonic expressions for this term. Despite this

variety of available equations, all models (harmonic and anharmonic) have the same behaviour around the equilibrium value.

- The second term is the harmonic potential for bond angle bending from its equilibrium value (θ_{eq}).
- The third term is the torsion angle potential function which models steric barriers between atoms separated by 3 covalent bonds (1,4 pairs). This term is described by a dihedral torsion from equilibrium value γ to its non-equilibrium value (ϕ) This potential is assumed to be periodic and is often expressed as a cosine function.

The fourth term expresses Van der Waals interactions which is one of the two considered non-bonding interactions. Nonetheless, we can reproduce other interactions by modifying the exponents of the dividends, for instance, using the 12-6 potential for Van der Waals interactions and 12-10 for hydrogen bonding (rarely used). The r_{ij} variable is the distance between the two considered atoms and A_{ij} and B_{ij} are parameters that depend of the pair of atoms, these two parameters refer to hard core repulsion and dispersive attraction respectively.

- The fifth term accounts for the other non-bonding interactions called electrostatic potential where q_i and q_j are the point charges of atoms i and j respectively, r_{ij} is the distance between them, and ϵ_0 is vacuum permittivity.

It is highly remarkable that, in MM simulations, each atom is defined by three main characteristics, which are constituted by:

- Its hybridization
- Its formal atomic charge
- Its nearest neighbours

These characteristics are represented by the constants written in equation 3.22., which altogether with the equilibrium values are taken from AMBER libraries. All the terms not found in AMBER libraries must be parameterized either using experimental data or by means of *ab initio* or DFT calculations. In the present thesis charges have been obtained from QM *ab initio* calculations according to either the so-called Restrained Electrostatic Potential (RESP)³⁰ fitting or by fitting the quantum mechanical and the Coulombic Molecular Electrostatic Potential (MEP).³¹

The aforementioned potentials lead to a high number of constants and interactions. In spite of the high power of MM for huge systems, it is necessary,

sometimes, to reduce the complexity of the system in terms of parameters and interactions number. This reduction entails less resource consumption. There are several methods to reduce complexity of the systems:

- United Atom (UA) FFs embed light atoms into the heavy ones they are bound to. This is made by adding the Van der Waals radii and charges of the light atoms to the heavy ones. By doing this, we can slash down dramatically the number of constants and potential energy functions to be calculated.
- Cut-off is a threshold distance beyond which we do not take into account neither non-bonded interactions nor electrostatic potential. This is useful to cut down the number of the aforementioned interactions to be calculated. It is assumed that, at distances over the cut-off, the contribution of electrostatic and non-bonded interaction to the total potential energy is negligible for Van der Waals interaction. The remaining electrostatic contribution in distances beyond cut-off, which may be significant, can be modelled by Particle Mesh of Ewald (PME) method.
- SHAKE algorithm takes as rigid (non-vibrating) the bonds between hydrogen and carbon atoms.

3.3.2. *Molecular Dynamics (MD)*

MD uses Newton's equations of movement with classic potentials (coming from MM) to include time.

Given an atom i with mass m_i and considering that its position is described by a three dimensional vector r_i , its motion is ruled by Newton's law

$$\frac{d\mathbf{v}_i(t)}{dt} = \frac{\mathbf{F}_i}{m_i} \quad (3.23.)$$

$$\mathbf{v}_i(t) = \frac{d\mathbf{r}_i(t)}{dt} \quad (3.24.)$$

where v_i and F_i are respectively the velocity and the force acting on the atom i in a given moment. F_i can be obtained through equation 3.23.

$$\mathbf{F}_i = -\frac{\partial E(\mathbf{r}_N)}{\partial \mathbf{r}_i} \quad (3.25.)$$

where $E(r_N)$ is given by eq. 3.22. and N is the number of particles in the system.

Equation 3.24. can be expressed as a Taylor expansion around t , this leads to a mathematical expression that allows us to calculate the velocity at each time step (Δt) (Equation 3.26.)

$$\mathbf{v}_i\left(t_n + \frac{\Delta t}{2}\right) = \mathbf{v}_i\left(t_n - \frac{\Delta t}{2}\right) - \frac{\Delta t}{m_i} \frac{\partial E(\mathbf{r}_N)}{\partial \mathbf{r}_i} + o(\Delta t^3) \quad (3.26.)$$

By calculating the same Taylor's expansion for the position vectors (\mathbf{r}_i) at $t + \Delta t$ and $t - \Delta t$, assuming that expansion terms over the third one are negligible we can develop the so called leap-frog Verlet algorithm³² (Equations 3.27. and 3.28.)

$$\mathbf{r}_i(t_n + \Delta t) = \mathbf{r}_i(t_n) + \mathbf{v}_i \Delta t + \frac{(\Delta t)^2 \mathbf{a}(t_n)}{2} \quad (3.27.)$$

$$\mathbf{r}_i(t_n + \Delta t) = \mathbf{r}_i(t_n) - \mathbf{v}_i \Delta t + \frac{(\Delta t)^2 \mathbf{a}(t_n)}{2} \quad (3.28.)$$

That which can be rearranged as:

$$\mathbf{r}_i(t_n + \Delta t) = 2\mathbf{r}_i(t_n) - \mathbf{r}_i(t_n - \Delta t) + (\Delta t)^2 \mathbf{a}(t_n) \quad (3.29.)$$

As we can see there is no expression for velocity in this formula, though it can be calculated as:

$$\mathbf{v}_i(t_n) = \frac{\mathbf{r}_i(t_n + \Delta t) - \mathbf{r}_i(t_n - \Delta t)}{2 \Delta t} \quad (3.30.)$$

Time step Δt might be small enough to simulate those movements with the highest frequencies, which normally are the bond vibrations. Consequently time step takes values of the order of the femtosecond. It is especially remarkable the C-H vibrating frequency since it is one of the smallest frequencies that can be found. Small time steps lead to more expensive simulations.

One of the biggest drawbacks of this method is that it needs to store information for position and acceleration for time t_n and $t_n + \Delta t$. This increases significantly

the computational resources in terms of memory space. However, there is another version of the Verlet algorithm named *velocity Verlet algorithm* that works with velocities, positions and accelerations at time t_n .

The CPU time needed for a MD simulation depends on several factors as the number of the explicit particles in the system, the time step or the cut-off radii. In order to speed up the MD simulation some simplifications might be done such as freezing the fastest modes of vibration by constraining the bonds to hydrogen atoms to fixed lengths. In this sense, SHAKE or constrained Verlet method³³ has been used in this thesis.

3.3.3. Thermodynamic Ensembles

An ensemble is a collection of all microscopic states that belong to a single macroscopic or thermodynamic state. There are five ensembles that are used in MD simulations.³⁴

- The canonical or NVT ensemble, whose thermodynamic state is characterized by a fixed number of atoms N , fixed volume V and fixed temperature T . This ensemble is widely used to simulate biological systems.
- The isobaric-isoenthalpic or NPH ensemble, where the number of atoms N , the pressure P and enthalpy H are fixed.
- The isobaric-isothermal or NPT ensemble, with fixed values of number of atoms N , pressure P and temperature T .
- The microcanonical or NVE ensemble, which corresponds to a closed or isolated system since energy E , besides the number of atoms N and volume V , is fixed.

In the present thesis NVT and NPT ensembles have been employed. Temperature is mainly set by the average kinetic energy of the molecules of the system. Consequently temperature can be controlled by adjusting the velocities of each atom. Berendsen thermostat rescales velocities of atoms assuming a scale factor λ defined as:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T}{T_0} - 1 \right) \right]^{\frac{1}{2}} \quad (3.31.)$$

where τ_T is the heat bath coupling time and T_0 is the set point temperature.

Concerning the pressure, a fixed value of it implies that volume must be able to fluctuate by dynamically adjusting the size of the unit cell and rescaling all atomic coordinates. In the case of Berendsen barostat, a scale factor μ is used to scale lengths of the box of the system.

$$\mu = \left[1 - \frac{\Delta t}{\tau_p} (P - P_0) \right]^{\frac{1}{3}} \quad (3.32.)$$

where τ_p is the pressure relaxation time of the barostat and P_0 is the set point pressure. Equations 3.31. and 3.32. define the so called Berendsen thermobarostat.³⁵

3.3.4. Periodic boundary conditions and Particle Mesh-Ewald Summation

MD simulations are performed in finite systems, this is in a box. In spite using a proper size for the box, it is absolutely unavoidable to suffer “wall effects”. These effects are caused by the crash of particles against the edges of the box. In order to prevent this from happening, MD simulations are usually performed with periodic boundary conditions; this is by representing an infinite system. The whole system is replicated periodically in all directions so atoms outside the simulation box are simply images of the atoms simulated in that box. Periodic boundary conditions ensure that all simulated atoms are surrounded by neighboring atoms, either images or not. This condition guarantees that atoms moving out of the box at one side are able to enter again inside the box at the opposite side because replicas of each particle in all duplicated boxes move exactly the same way. Minimum image convention avoid duplicate interactions between atoms i and j by taking into account only the interaction of atom i with the closest j atom, either original or copy. Thus, periodic boundary conditions are useful to correct potential energy errors in Coulombic and van der Waals non bonded terms for those systems that none of the molecules inside simulation box is covalently bound to another molecular moiety outside the box. Concerning the Coulombic non bonded term, the main problem to evaluate it correctly is that a sudden cut-off lead to large errors. This entanglement can be solved with the use of the so called Particle Mesh-Ewald summation (PME),³⁶ which calculate the infinite electrostatic interactions by splitting the summation into short and long range parts. For PME, the cut-off only determines the balance between the

two parts, and the long-range part is treated by assigning charges to a grid that is solved in reciprocal space through Fourier transforms.

3.3.5. QM/MM calculations

MM methodology has reached an enormous success in theoretical structural biology, polymer and material science. This is due to the ability of MM to cope with big systems with a number of atoms up to several hundred thousand. This formidable achievement was made at expense of the accuracy for electronic description of the system. This lack of accuracy has its highest expression in the lack of electron correlation and, as it was mentioned previously in the text, this is major problem for reproducing chemical reaction, π -stacking, electron conduction, substrate polarization and a long etcetera.

This limitation forced researchers that studied enzymatic biochemical reactions to use simpler models that tried to reproduce the entire system with a lower number of atoms to make it computationally accessible. A powerful alternative solution is QM/MM methods. They are based in splitting the system in two parts, a small one of special interest described at the QM level and the rest of the system described at the MM level. QM/MM philosophy was firstly introduced by Warshel and Levitt³⁷ and developed by Field, Bash and Karplus.³⁸ QM/MM methods differ, besides the classical force field and the QM level of theory employed, in how the two parts are connected. There are two way to connect QM and MM parts:

- *Integrated Molecular Orbital + Molecular Mechanics* (IMOMM) methods are the so called link atom scheme, which uses a monovalent atom to cap the unsaturated QM atoms. Normally hydrogen atom is used as link atom.³⁹ The use of these fictitious atoms caps the QM part to avoid calculating radical forms and has a less complex mathematical treatment. The fictitious atom binds through an harmonic potential to the MM part.
- *Integrated Molecular Orbital + Molecular Orbital* (IMOMO) are methods based in the use of localized orbitals in the boundaries separating the QM and MM parts. They are more realistic than the previous method but more “expensive” in computational terms. One example is the Local Self-Consistent Field (LSCF) method⁴⁰ in which the bonds connecting both parts are represented by a set of strictly localized bond orbitals (SLBOs) that are determined by calculations on small model compounds and assumed to be transferable.

Within the framework of QM/MM calculations the total Hamiltonian operator is defined as the sum of the QM, QM/MM and MM Hamiltonians:

$$\hat{H} = \hat{H}^{\text{QM}} + \hat{H}^{\text{QM/MM}} + \hat{H}^{\text{MM}} \quad (3.33.)$$

where $\hat{H}^{\text{QM/MM}}$ is expressed as the sum of electron-charge, nuclei-charge and van der Waals interaction potentials between QM and MM atoms:

$$\hat{H}^{\text{QM/MM}} = V_{\text{elec}}^{\text{QM/MM}} + V_{\text{nucl}}^{\text{QM/MM}} + V_{\text{vdW}}^{\text{QM/MM}} \quad (3.34.)$$

Equation 3.34. can be expanded for a system containing N QM atoms and M MM atoms as follows:

$$\hat{H}^{\text{QM/MM}} = - \sum_{i=1}^N \sum_{j=1}^M \frac{q_j}{r_{ij}} + \sum_{i=1}^N \sum_{j=1}^M \frac{Z_i q_j}{r_{ij}} + \sum_{i=1}^N \sum_{j=1}^M \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right] \quad (3.35.)$$

Thus energy in a QM/MM system is defined as shown in Equation 3.36.

$$E = E^{\text{QM}} + E^{\text{MM}} + E^{\text{QM/MM}} \quad (3.36.)$$

where $E^{\text{QM/MM}}$ is defined as:

$$E^{\text{QM/MM}} = \langle \Psi | - \sum_{i=1}^N \sum_{j=1}^M \frac{q_j}{r_{ij}} | \Psi \rangle + V_{\text{nucl}}^{\text{QM/MM}} + V_{\text{vdW}}^{\text{QM/MM}} \quad (3.37.)$$

3.3.6. Conformational exploration

One of the fields where MM has developed intensively is the conformational exploration for macromolecules. These molecules are usually so big that the QM approach is unfeasible, leaving the MM approach as the only possible one. Conformational exploration consists on characterizing all the possible conformations a molecule is able to adopt.

For small molecules one considers of the possible values a dihedral may adopt and all the possible combinations of values to generate starting points to be optimized with QM. This optimization should render the lowest energy minima for that molecule. For big molecules such as peptides, proteins, polymers and DNA conformational exploration is not so simple. In fact, the high number of dihedrals to

be considered, the number of possible values for those dihedrals, and the number of combinations makes it unrealistic that combinatorial approach. Besides, for large molecules we find problems to locate the global minimum since other quasi degenerate minima and high energy conformational barriers might exist, entropic and solvation effects play a key role too. Putting it all together: we need a methodology to overcome that conformational barriers and, at the same time, to find the lowest energy minima. Two of these methodologies have been used in this thesis:

3.3.6.1. Simulated Annealing (SA)

The conformational space for large systems is difficult to explore due to an exacerbation of the aforementioned difficulties, high energy conformational barriers are highly remarkable stressed. Simulated Annealing (SA) has been widely used to search the conformational space of peptides.^{41,42,43} This optimization algorithm was explored very early in the seminal work by Kirkpatrick and Gelatt.⁴⁴

SA consists on running MD simulations in systems that are gradually heated. As soon as the system reaches a temperature in which kinetic energy is high enough to cross through the conformational barriers, temperature is hold. After a period of time running at that high temperature, the system is gradually cooled down. This system is believed to emulate the natural folding of protein from a random-coil conformation to a minimal energy conformation.

After several cycles, one can obtain series of low energy structures in which we can find quasi degenerated structures with the global minimum.

3.3.6.2. Replica Exchange Molecular Dynamics (REMD)

REMD^{45,46} is a methodology to search the conformational space of macromolecules by running several non-interacting replica MD simulations simultaneously at different temperatures each one. The potential energy of the replicas is assessed each certain period of time and then compared pair wise between consecutive replicas (consecutive temperatures).

Out of this comparison, one can obtain one of the two options: an exchange between two replicas of the heat bath, followed by a concomitant re-scaling of kinetic energy, or no exchange. To assess the probability for this exchange, REMD uses a metropolis criterion:

$$\rho = \min\left(1, \exp\left[\left(\frac{1}{K_B T_M} - \frac{1}{K_B T_N}\right)(E(q_i) - E(q_j))\right]\right) \quad (3.38.)$$

Where ρ is the probability of exchange, $E(q_i)$ and $E(q_j)$ are the potential energy for the two replicas, K_B is the Boltzmann constant, and, T_M and T_N are the temperatures of replicas i and j respectively.

From equation 3.38. we can deduce that the probability of acceptance of the exchange is as follows:

$$P_{acc} = K \exp\left[-\left(\frac{\Delta T^2}{T^2}\right)\right] \quad (3.39.)$$

Equation 3.39 let be that the smallest the ΔT^2 is, the highest the probability of acceptance probability is. This relationship collides severely with the fact that it is necessary to use big ΔT^2 in order to reach high temperatures, in a few number of replicas, to overcome energy barriers. Unfortunately, REMD needs an exchange ratio over 10% to be a useful technique and this makes mandatory to use a number of replicas for big systems. Another way to understand the importance of ΔT is to consider that to have a successful exchange the value of the quotient $\Delta E / \delta E$ must be around 1; ΔE is the energy gap between replicas and δE is energy fluctuation of each replica during MD. This quotient should have values around 1, then normal fluctuations in the energy of the system would be enough to match the energy difference between the two replicas at some moment, allowing its exchange.

The determination of exchange rate (probability) follows a multi-factorial scheme: potential energy, temperature, and differences of the two previous factors between replicas are keystones. This complexity makes temperature distribution to be determined through several algorithms,⁴⁷ though a trial and error approach is necessary for a more precise adjust .

3.4. References

- [1] Levine N. I. "Química cuántica", **2005**, Pearson Prentice-Hall
- [2] Hinchliffe A. "Molecular modeling for beginners", **2008**, Wiley.
- [3] a) Roothaan C.C.J. *Rev. Mod. Phys.* **1951**, *23*, 69. b) Hall G.G., *Proc. Roy. Soc.* **1951**, *A205*, 541.
- [4] Hehre W.J., Stewart R.F., Pople J.A. *J. Chem. Phys.* **1969**, *51*, 2657.
- [5] Binkley J.S., Pople J.A., Hehre W.J. *J. Am. Chem. Soc.* **1980**, *102*, 939.
- [6] Ditchfield R., Hehre W.J., Pople J.A. *J. Chem. Phys.* **1971**, *54*, 724.
- [7] a) Hartree D.R. *Proc. Cambridge Phil. Soc.* **1928**, *24*, 89 b) Slater J.C. *Phys. Rev.* **1930**, *35*, 210 c) Fock V. *Z. Physik* **1930**, *61*, 126.
- [8] a) Møller C., Plesset M.S. *Phys. Rev.* **1934**, *46*, 618 b) Binkley J.S., Pople J.A. *Int. J. Quantum Chem.*, **1975**, *9*, 229.
- [9] a) Coester F., Kümmel H. *Nucl. Phys.* **1960**, *17*, 477 b) Paldus J., Čížek J. Shavitt I. *Phys. Rev. A* **1972**, *5*, 50.
- [10] a) Stewart J.J.P. *J. Comp. Chem.* **1989**, *10*, 209 b) Stewart J.J.P. *J. Comp. Chem.* **1989**, *10*, 221 c) Stewart J.J.P. *J. Comp. Chem.* **1991**, *12*, 320.
- [11] Dewar M.J.S., Zoebisch E.G., Healy E.F., Stewart J.J.P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
- [12] Hohenberg P., Kohn W. *Phys. Rev. B* **1964**, *136*, B864.
- [13] Slater J.C. *Int. J. Quantum Chem. Symp.* **1975**, *9*, 7.
- [14] Painter G.S. *Phys. Rev. B* **1981**, *24*, 4264.
- [15] Vosko S.H., Wilk L., Nusair M. *Can. J. Phys.* **1980**, *58*, 1200.
- [16] Perdew J.P., Zunger A. *Phys. Rev. B* **1981**, *23*, 5048.
- [17] Becke A.D. *J. Chem. Phys.* **1996**, *104*, 1040.
- [18] Perdew J.P. *Phys. Rev. B* **1986**, *33*, 8822.
- [19] Perdew, J.P., Burke, K., Ernzerhof M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- [20] Perdew J. P., Wang Y. *Phys. Rev. B* **1992**, *45*, 13244.
- [21] Lee C., Yang W., Parr R.G. *Phys. Rev. B* **1988**, *37*, 785.
- [22] Becke A.D. *J. Chem. Phys.* **1993**, *98*, 1372.
- [23] a) Adamo C., Barone V. *J. Chem. Phys.* **1998**, *108*, 664 b) Lynch B.J., Fast P.L., Harris M., Truhlar D.G. *J. Phys. Chem. A* **2000**, *104*, 4811 c) Lynch B.J., Zhao Y., Truhlar D.G. *J. Phys. Chem. A* **2003**, *107*, 1384.
- [24] Schultz N.E., Zhao Y., Truhlar D.G. *J. Phys. Chem. A* **2005**, *109*, 11127.

- [25] a) Becke A.D. *Phys. Rev. A* **1988**, 38, 3098 b) Xu X., Goddard W.A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101, 2673.
- [26] Boese A.D., Martin J.M.L. *J. Chem. Phys.* **2004**, 121, 3405-3416.
- [27] Runge K., Gross E.K.U. *Phys. Rev. Lett.* **1984**, 52, 997-1000.
- [28] a) Miertuš S., Scrocco E., Tomasi J. *Chem. Phys.* **1981**, 55, 117 b) Miertuš S., Scrocco E., Tomasi J. *Chem. Phys.* **1982**, 65, 239.
- [29] a) 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 b) Wang J., Cieplak P., Kollman P.A. *J. Comput. Chem.* **2000**, 21, 1049.
- [30] a) Bayly C.I., Cieplak P., Cornell W., Kollman P.A. *J. Phys. Chem.* **1993**, 97, 10269 b) Cornell W.D., Cieplak P., Bayly C.I., Kollmann P.A. *J. Am. Chem. Soc.* **1993**, 115, 9620.
- [31] Alemán C., Luque F.J., Orozco M. *J. Comput. Aided Mol. Design* **1993**, 7, 721.
- [32] Verlet J. *Phys. Rev.* **1967**, 159, 98.
- [33] a) van Gunsteren W.F., Berendsen H.J.C. *Mol. Phys.* **1977**, 34, 1311 b) Ryckaert J. P., Ciccotti G., Berendsen H.J.C. *J. Comput. Phys.* **1977**, 23, 327.
- [34] Adcock S.A., McCammon J.A. *Chem. Rev.* **2006**, 106, 1589.
- [35] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. *J. Chem. Phys.* **1984**, 81, 3684.
- [36] Darden T., York D., Pedersen L. *J. Chem. Phys.* **1993**, 98, 10089.
- [37] Warshel A., Levitt M. *J. Mol. Biol.* **1976**, 103, 227.
- [38] Field M., Bash P., Karplus M. *J. Comp. Chem.* **1990**, 11, 700.
- [39] Maseras F., Morokuma K. *J. Comput. Chem.* **1995**, 16, 1170.
- [40] Ferré N., Assfeld X., Rivail J.L. *J. Comput. Chem.* **2002**, 23, 610.
- [41] Filizola M., Perez J.J., Palomer A., et al. *J. of Mol. Graph. Model.* **1997**, 15, 290.
- [42] Filizola M., CarteniFarina M., Perez J.J. *Journal of Pept. Res.* **1997**, 50, 55.
- [43] Filizola M., Centeno N.B., Perez J.J. *J. of Pept. Sci.* **1997**, 3, 85.
- [44] Kirkpatrick S., Gelatt C.D., Vecchi M.P. *Science* **1983**, 220, 671.
- [45] Cheng X., Cui G., Hornak V., Simmerling C. *J. Phys. Chem. B* **2005**, 109, 8220.
- [46] Sugita Y., Okamoto Y. *Chem. Phys. Lett.* **1999**, 314, 141.
- [47] Patriksson A., Van der Spoel D. *Phys. Chem. Chem. Phys.* **2008**, 10, 2073

4. Characterization of non-coded amino acids

4.1. Side-chain to backbone interactions dictate the conformational preferences of a cyclopentane arginine analogue

4.1.1. Introduction

Among non-proteinogenic amino acids the conformational propensities of which can be exploited in the design of peptides analogues with well-defined backbone conformations are 1-aminocycloalkane-1-carboxylic acids¹ (known in the abbreviated form as $Ac_n c$, with n referring to the ring size). Within this series, the cyclopropane ($Ac_3 c$),^{1,2} cyclobutane ($Ac_4 c$),^{1,2c,3} cyclopentane ($Ac_5 c$)^{1,2a,3b,4} and cyclohexane ($Ac_6 c$)^{1,3b,5} members have been deeply investigated and shown to exhibit a restricted conformational space characterized by a high propensity to adopt ϕ, ψ backbone angles typical of the 3_{10} -/ α -helix (with some distortion in the case of $Ac_3 c$).

When considering a bioactive peptide, amino acids of the $Ac_n c$ family have proven appropriate replacements for proteinogenic residues bearing aliphatic or aromatic side chains.⁶ However, the $Ac_n c$ series may not be convenient to replace a proteinogenic amino acid containing a functionalized side chain which is directly involved in the peptide-receptor recognition process and is therefore essential for bioactivity. One may yet consider a new family of non-coded amino acids generated by attaching the functionalized side chain of a natural residue to the cycloalkane moiety in $Ac_n c$. This allows the combination of the necessary functionality with the particular conformational properties of the $Ac_n c$ residues (Figure 4.1.1.). Moreover, this may enable the specific orientation of the side chain functionality by selecting the appropriate cycloalkane size and stereochemistry.

Specifically, we have been working on the synthesis⁷ and structural study⁸ –both theoretical and experimentally– of the amino acids obtained by incorporating a phenyl substituent at one of the α carbons of $Ac_n c$ ($n = 3-6$). The compounds thus obtained can be considered as phenylalanine (Phe) analogues and we denote them as $c_n Phe$, where n indicates the size of the cycle, as in $Ac_n c$. Since the phenylalanine side chain in $c_n Phe$ is included in a cyclic structure, the $C^\alpha-C^\alpha$ bond can not rotate freely and, as a consequence, the orientation of the aromatic group is dictated by the size (n value) and stereochemistry of the cycloalkane ring. It should be noted that the additional phenyl substituent may exhibit a *cis* or a *trans* relative disposition with respect to the amino function. Accordingly, the different $c_n Phe$ stereoisomers can be regarded as a series of phenylalanine analogues with distinct well-defined side-chain orientations.

Indeed, the different spatial arrangement attained by the aromatic substituent has proven useful in several applications related to the stabilization of particular peptide backbone conformations.^{8,9}

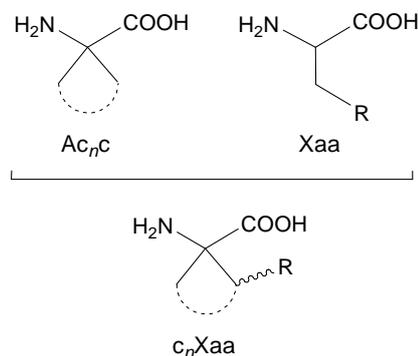


Figure 4.1.1. Structure of 1-aminocycloalkane-1-carboxylic acids (Ac_nC , n : cycle size) and a proteinogenic amino acid (represented, in general, as Xaa). Combination of the cyclic structure of Ac_nC with the side-chain functionality in Xaa gives rise to c_nXaa residues.

Within a project aimed at imparting protection against proteolytic cleavage to a bioactive peptide with simultaneous stabilization of a folded conformation, we became interested in the replacement of an arginine residue (Arg) by a non-natural analogue. In particular, we have focused our attention on the arginine analogue bearing an Ac_5c skeleton, that is, c_5Arg according to the nomenclature described above (Figure 4.1.1.). As it can be seen, in c_5Arg the α carbon is separated from the guanidinium group by two carbon atoms, while this segment involves three carbon atoms in Arg . Therefore, from a rigorous point of view c_5Arg is a substituted Ac_5c -like derivative of nor-Arginine, where nor refers to a reduction of one carbon atom with respect to the side chain of conventional Arg . For an L configuration at the α carbon, the guanidilated side chain of arginine may exhibit a *trans* or a *cis* disposition relative to the amino moiety, respectively giving rise to *trans*- and *cis*- c_5Arg (Figure 4.1.2.). It should be considered that the charged side chain of c_5Arg may interact with the backbone not only sterically but also electronically, and this may have a strong impact on the structural preferences of the peptide chain. In order to evaluate the behavior of the L enantiomer of both *trans*- and *cis*- c_5Arg , we report a conformational study of the corresponding *N*-acetyl-*N'*-methylamide derivatives, hereafter denoted as *Ac-t-L-c₅Arg-NHMe* and *Ac-c-L-c₅Arg-NHMe*, respectively (Figure 4.1.2). Density Functional Theory (DFT) calculations at the B3LYP/6-311+G(d,p) level have been

used to locate and characterize the minimum energy conformations. The influence of the solvent polarity on the conformational preferences has been examined using a Self Consistent Reaction Field (SCRF) method and molecular dynamics (MD) simulations with explicit solvent molecules.

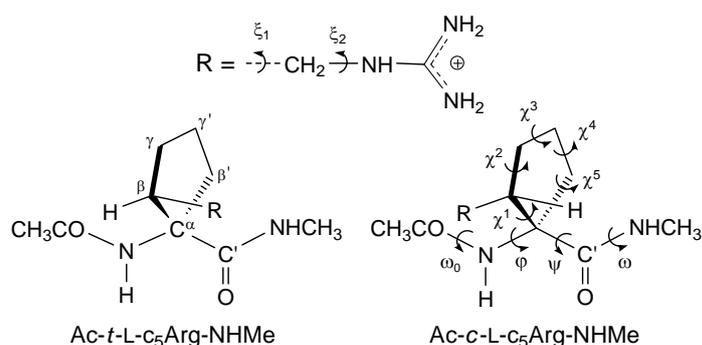


Figure 4.1.2. Structure of the compounds investigated, containing the *trans* (*t*) and *cis* (*c*) cyclopentane analogues of L-arginine. The backbone and side-chain dihedral angles are indicated.

4.1.2. Methods

The conformational properties of Ac-*t*-L-c₅Arg-NHMe and Ac-*c*-L-c₅Arg-NHMe have been investigated using the Gaussian-03 computer program.¹⁰ The structural search was performed considering that the compounds under study retain the restrictions imposed by the cyclopentane ring on the backbone in Ac₅c.^{4a} Accordingly, the five minimum energy conformations characterized for Ac-Ac₅c-NHMe in ref. 4a were used to generate the starting structures for Ac-*t*-L-c₅Arg-NHMe and Ac-*c*-L-c₅Arg-NHMe. Although for Ac-Ac₅c-NHMe such five minima were two-fold degenerate due to the symmetry of the molecule, *i.e.* $\{\varphi, \psi, \chi^i\} = \{-\varphi, -\psi, -\chi^i\}$, the chiral nature of the two c₅Arg derivatives under study requires explicit consideration of both $\{\varphi, \psi, \chi^i\}$ and $\{-\varphi, -\psi, -\chi^i\}$ possibilities. The arrangement of the side group is defined by the flexible dihedral angles ζ_1 and ζ_2 , which are expected to exhibit three different minima: *trans* (180°), *gauche*⁺ (60°) and *gauche*⁻ (-60°). Consequently, 5 (minima of Ac-Ac₅c-NHMe) × 2 (chiral nature of c₅Arg) × 3 (minima of ζ_1) × 3 (minima of ζ_2) =

90 minima can be anticipated for the potential energy hypersurface (PEH) $E = E(\phi, \psi, \chi^i, \zeta_1, \zeta_2)$ of each c_5 Arg-containing derivative. All these structures were used as starting points for subsequent full geometry optimizations.

All geometry optimizations were performed using the B3LYP functional^{11,12} combined with the 6-311+G(d,p) basis set.¹³ 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) and both thermal and entropic corrections. These statistical terms were then used to compute the conformational Gibbs free energies in the gas phase at 298K (ΔG_{gp}).

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 a SCRF model. Specifically, the Polarizable Continuum Model (PCM) developed by Tomasi and co-workers¹⁴ was used to describe water and chloroform as solvents. The PCM model represents the polarization of the liquid by a charge density appearing on the surface of the cavity created in the solvent. This cavity is built using a molecular shape algorithm. PCM calculations were performed in the framework of the B3LYP/6-311+G(d,p) level using the standard protocol and considering the dielectric constants of water ($\epsilon = 78.4$) and chloroform ($\epsilon = 4.9$) to obtain the free energies of solvation (ΔG_{solv}) of the minimum energy conformations. Within this context, it should be emphasized 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 ΔG_{solv} values.¹⁵ The conformational free energies in solution (ΔG^{conf}) at the B3LYP/6-311+G(d,p) level were estimated using the classical thermodynamics scheme: $\Delta G^{conf} = \Delta G_{gp} + \Delta G_{solv}$.

MD simulations in water solution were performed using the NAMD program.¹⁶ The simulated peptides Ac-*t*-L- c_5 Arg-NHMe and Ac-*c*-L- c_5 Arg-NHMe were placed in the center of a cubic simulation box ($a = 31.1 \text{ \AA}$) filled with 338 explicit water molecules, which were represented using the TIP3 model,¹⁷ and a negatively charged chloride atom as counterion. Atom pair distance cutoffs were applied at 14 \AA to compute van der Waals interactions. The electrostatic interactions were computed using the nontruncated electrostatic potential by means of Ewald Summations. The real space term was determined by the van der Waals cutoff (14 \AA), while the reciprocal term was estimated by interpolation of the effective charge into a charges mesh with a grid thickness of 5 points per volume unit, *i.e.* the Particle-Mesh Ewald

(PME) method.¹⁸ Bond lengths were constrained using the *SHAKE* algorithm¹⁹ and the numerical integration step was 2 fs.

Before the MD run series was started, 5×10^3 steps of energy minimization were performed to relax conformational and structural tensions. Different consecutive rounds of short MD runs were performed to equilibrate the density, temperature, and pressure: 0.50 ns of *NVT*-MD at 298 K (thermal relaxation) followed by 0.25 ns of isobaric relaxation (*NPT*-MD). Both temperature and pressure were controlled by the weak coupling method, the Berendsen thermo-barostat²⁰ using a time constant for heat bath coupling, and a pressure relaxation time of 1 ps. The coordinates of the *NPT*-MD production runs, which were 10 ns long, were saved every 500 steps (1 ps intervals) for subsequent analysis.

4.1.3. Results and discussion

Geometry optimization at the B3LYP/6-311+G(d,p) level led to the characterization of 28 and 23 different minimum energy structures for *Ac-t-L-c₅Arg-NHMe* and *Ac-c-L-c₅Arg-NHMe*, respectively. These minima are within relative energy (ΔE) intervals of 36.7 and 20.8 kcal/mol, respectively.

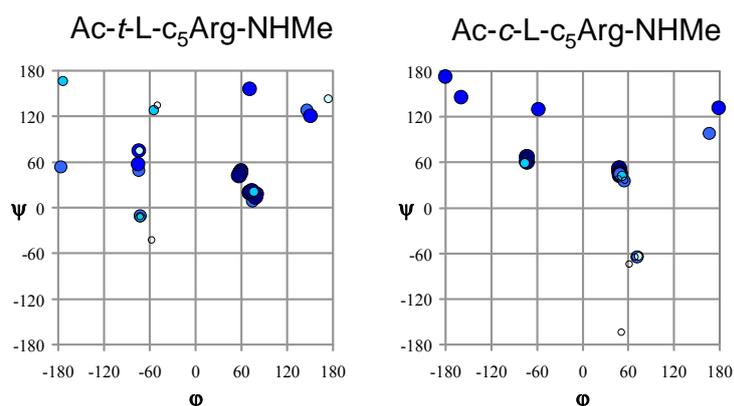
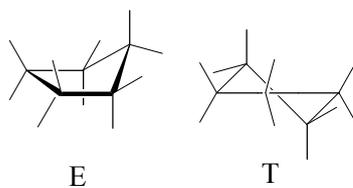


Figure 4.1.3. Distribution on the Ramachandran map of the minimum energy conformations characterized at the B3LYP/6-311+G(d,p) level for the two *c₅Arg* derivatives under study. The color and size of the symbols used to represent the backbone conformations depend on the relative energy (ΔE) values. Specifically, large and dark blue circles correspond to the more stable minima, while small and empty circles are the least stable ones, *i.e.* both the intensity of the color and the size of the circles decrease when the relative energy increases.

Figure 4.1.3. represents the φ, ψ backbone dihedral angles of these minima using a color scale (dark blue-to-white) to show the ΔE increase through intervals of 5 kcal/mol. As it can be seen, almost all regions of the Ramachandran map are visited because of the large number of minima characterized. However, the conformational space available to the compounds investigated is relatively restricted, especially that corresponding to the *trans*- c_5 Arg derivative. Thus, only 8 out of the 28 minima found for Ac-*t*-L- c_5 Arg-NHMe have ΔE values lower than 5 kcal/mol, and all 8 exhibit conformations in the α_L region (left-handed α -helix; $\varphi, \psi \approx 60^\circ, 50^\circ$) of the Ramachandran map. Regarding *cis*- c_5 Arg, 7 (out of 23) minimum energy conformations have $\Delta E < 5$ kcal/mol and correspond to three different backbone conformations, namely α_L , C_7^{eq} (equatorial C_7 or inverse γ -turn; $\varphi, \psi \approx -60^\circ, 60^\circ$) and C_5 (fully extended $\varphi, \psi \approx \pm 180^\circ, \pm 180^\circ$). Accordingly, the relative stability of the minimum energy conformations characterized for these c_5 Arg derivatives is strongly influenced by the *cis* / *trans* disposition of the charged substituent.

The next two sections present a detailed description of those minimum energy conformations characterized for the compounds under study that are more favored, not only in the gas phase but also in chloroform and water solutions. These minima are denoted using three labels. The first one refers to the backbone conformation type, defined by the φ, ψ dihedral angles. The second label corresponds to the puckering of the cyclopentane ring, *i.e.* endo/exo-envelope (E) or twist (T) conformations (Scheme 1). Finally, the third label indicates the conformation of the guanidinium side chain, that is, the *trans* (t), *gauche*⁺ (g^+) or *gauche*⁻ (g^-) arrangement of the dihedral angles ζ_1 and ζ_2 .



Scheme 1

Ac-*t*-L-c₅Arg-NHMe. Table 4.1.1. lists the backbone and side-chain dihedral angles of the 13 minimum energy conformations calculated for the *trans*-c₅Arg derivative with $\Delta E < 7$ kcal/mol. The global minimum corresponds to an $\alpha_L/\gamma'E/g^+t$ conformation (Figure 4.1.4.a), in which the backbone adopts an α -helical structure and the cyclopentane ring accommodates a C γ' -exo envelope ($\gamma'E$) arrangement. This geometry, combined with the *gauche*⁻/*trans* disposition of ζ_1/ζ_2 , allows the formation of a strong hydrogen bond between the guanidinium NH and the carbonyl oxygen of the acetyl blocking group [$d(H\cdots O) = 1.706$ Å, $\angle N-H\cdots O = 170.0^\circ$]. Modification of the envelope arrangement of the cyclopentane moiety from C γ' -exo ($\gamma'E$) to C γ' -endo ($\gamma'E$) gives rise to a new minimum ($\alpha_L/\gamma'E/g^+t$, Figure 4.1.4.b), that maintains all other conformational features present in the global minimum, including the side-chain \cdots backbone interaction. This $\gamma'E$ -to- $\gamma'E$ transition is associated with an energy penalty of 1.6 kcal/mol.

The conformation adopted by both the peptide backbone and the cyclopentane ring in the lowest energy conformer is maintained in the third ($\alpha_L/\gamma'E/tg^+$, Figure 4.1.4.c), fourth ($\alpha_L/\gamma'E/g^-g^-$, Figure 4.1.4.d), sixth ($\alpha_L/\gamma'E/g^+t$, Figure 4.1.4.f) and eighth ($\alpha_L/\gamma'E/g^-g^-$, Figure 4.1.4.h) minima, which present ΔE values ranging from 2.5 to 4.4 kcal/mol. Thus, minima # 1, 3, 4, 6 and 8 differ mainly in the arrangement of the guanidinium substituent. The diverse orientations exhibited by this group translate into different hydrogen-bonding schemes involving the donor sites in the side chain (NH/NH₂) and the backbone carbonyl groups. Notably enough, two of such side-chain \cdots backbone interactions exist simultaneously in minimum # 4 (Figure 4.1.4.d), although none of them present an optimal geometry. Moreover, to allow their formation, the ψ angle deviates by around 25° from the value characterizing the rest of $\alpha_L/\gamma'E$ conformers (Table 4.1.1.).

#	Conformer	Backbone dihedral angles				Cyclopentane dihedral angles					Side group	ΔE^b
		ω_0	φ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	$\zeta_1 \zeta_2$	
1	$\alpha_L/\gamma/E/g^+t$	174.8	71.9	19.8	174.5	-0.1	25.3	-40.9	41.2	-25.2	-63.3, 166.9	0.0 ^c
2	$\alpha_L/\gamma^+/E/g^+t$	174.6	71.2	19.6	173.6	-3.9	-21.2	38.6	-41.3	27.8	-59.3, 164.3	1.6
3	$\alpha_L/\gamma/E/tg^+$	177.5	79.0	14.3	177.2	-7.4	31.2	-42.8	38.3	-18.9	-162.8, 96.8	2.5
4	$\alpha_L/\gamma/E/g^+g^-$	176.2	59.1	45.5	177.9	4.6	21.0	-38.8	42.0	-28.6	-62.7, -94.1	2.8
5	$\alpha_L/\alpha/E/g^+t$	172.2	56.9	42.2	178.8	-39.0	22.3	4.1	-29.2	41.7	68.4, 174.4	3.0
6	$\alpha_L/\gamma/E/g^+t$	173.0	74.0	21.1	176.6	-8.7	32.0	-42.9	37.8	-17.7	75.1, 148.8	3.2
7	$\alpha_L/\gamma^+/E/tg^+$	177.8	79.0	16.3	176.5	-11.1	-13.3	33.0	-40.4	31.7	-161.8, 98.1	3.8
8	$\alpha_L/\gamma/E/g^+g^-$	174.7	80.5	17.5	174.3	10.4	15.4	-35.5	42.4	-32.6	-50.3, -87.0	4.4
9	$C_7^{eq}/\alpha/E/tg^+$	-177.8	-73.8	74.8	-175.5	45.0	-34.8	11.1	17.0	-38.2	-155.2, 85.4	5.2
10	$\beta_L/\alpha/E/g^+t$	164.7	150.9	119.9	179.0	-44.9	36.8	-13.5	15.2	37.0	-68.5, 125.5	5.6
11	$\alpha_L/\alpha/E/tg^+$	174.9	60.0	48.4	-179.4	45.3	-38.2	16.6	11.9	-35.4	-143.9, 82.9	6.2
12	$P_{II}/\gamma^+/E/g^+t$	-170.5	71.0	155.3	-176.0	-5.1	17.8	34.0	-37.7	26.4	-65.0, 159.9	6.5
13	$C_7^{eq}/\alpha/E/tg^-$	-170.5	-75.2	57.3	179.2	-41.0	24.8	15.6	-27.4	41.9	147.0, -98.3	6.7

Table 4.1.1. Dihedral angles^a and relative energies in the gas phase (ΔE) for the minimum energy conformations with $\Delta E < 7.0$ kcal/mol characterized for Ac-*t*-L-c₅Arg-NHMe at the B3LYP/6-311+G(d,p) level.^a In degrees; see Figure 4.1.2. for definition. ^b In kcal/mol. ^c E=-856.744293 a.u.

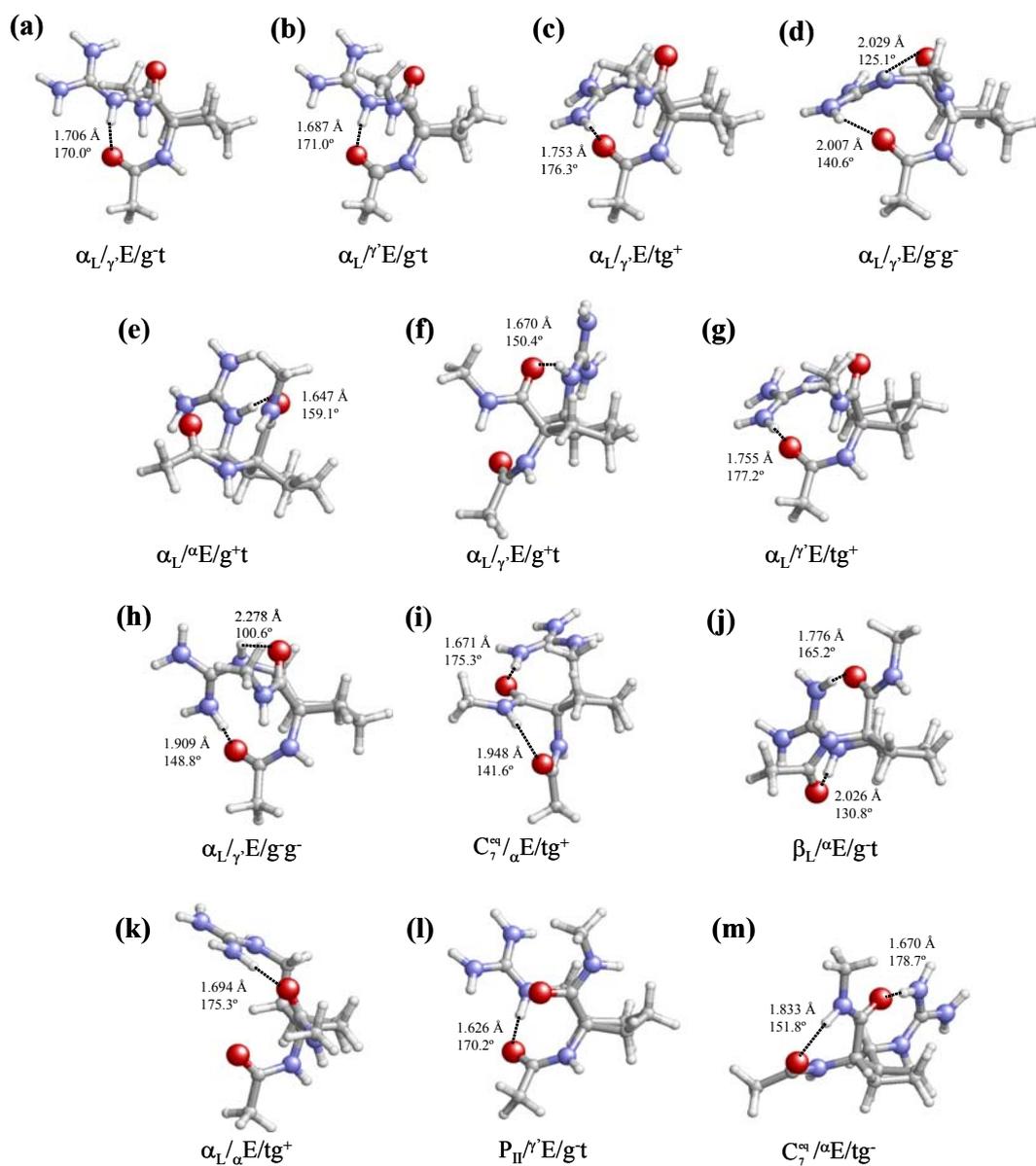


Figure 4.1.4. Lower minimum energy conformations of Ac-*t*-L-c₃Arg-NHMe obtained from B3LYP/6-311+G(d,p) calculations. The 13 structures depicted correspond to the minima listed in Table 4.1.1., *i.e.* minimum energy conformations with $\Delta E < 7$ kcal/mol.

On the other hand, changing the cyclopentane envelope arrangement in minimum # 3 from C^{γ} -exo to C^{γ} -endo results in a new minimum (# 7, $\alpha_L/\gamma'E/tg^+$, Figure 4.1.4.g), which is 1.3 kcal/mol higher in energy. This destabilization is similar to that produced on going from the first to the second minima, which implies the same conformational change.

Finally, two other minimum energy conformations with an α_L backbone structure and $\Delta E < 7$ kcal/mol were located for Ac-*t*-L-c₅Arg-NHMe, namely conformers # 5 ($\alpha_L/\alpha'E/g^+t$, Figure 4.1.4.e) and # 11 ($\alpha_L/\alpha'E/tg^+$, Figure 4.1.4.k). They present ΔE values of 3.0 and 6.2 kcal/mol, respectively, and are the only minima in Table 4.1.1. exhibiting an α_L backbone conformation in which the flap of the cyclopentane envelope is occupied by the α instead of the γ' carbon. The fact that no minima with an atom other than C^{γ} at the envelope flap is located below $\Delta E = 3$ kcal/mol is a significant difference with respect to the unsubstituted derivative Ac-Ac₃c-NHMe, for which the global minimum was found to exhibit an $\alpha'E$ cyclopentane arrangement.^{4a}

The most stable conformer with a backbone disposition other than α_L is # 9, which is unfavored by 5.2 kcal/mol with respect to the global minimum. In this structure ($C_7^{eq}/\alpha'E/tg^+$, Figure 4.1.4.i), the terminal backbone CO and NH sites form an intramolecular hydrogen bond [$d(H\cdots O) = 1.948 \text{ \AA}$, $\angle N-H\cdots O = 141.6^\circ$] defining a seven-membered cycle (C_7 or γ -turn conformation) and the cyclopentane ring adopts a C^α -exo envelope arrangement. Furthermore, the side chain orientation allows the formation of an additional hydrogen bond involving the c₅Arg CO and the guanidinium NH₂ [$d(H\cdots O) = 1.671 \text{ \AA}$, $\angle N-H\cdots O = 175.3^\circ$]. The other minimum with a C_7 backbone structure ($C_7^{eq}/\alpha'E/tg^-$, Figure 4.1.4.m.) exhibits similar backbone \cdots backbone and side-chain \cdots backbone interactions and differs from the former in the endo position occupied by C^α within the cyclopentane envelope.

Two additional types of peptide backbone conformation, corresponding to extended (β -pleated sheet) or semi-extended (polyproline II) structures were detected among the minima characterized for Ac-*t*-L-c₅Arg-NHMe with $\Delta E < 7$ kcal/mol, namely minima # 10 ($\beta_L/\alpha'E/g^+t$, Figure 4.1.4.j) and # 12 ($P_{II}/\gamma'E/g^+t$, Figure 4.1.4.l). They share a common disposition for the guanidinium side group (*gauche*/*trans* arrangement of ζ_1/ζ_2), that, combined with the different backbone and cyclopentane

conformations, leads to the formation of two and one side-chain...backbone interactions, respectively, for minima # 10 and 12.

Table 4.1.2. shows the conformational Gibbs free energies in the gas phase at 298K (ΔG_{gp}) for the minima listed in Table 4.1.1.. It is worth noting that the addition of the ZPVE, thermal and entropic contributions to the ΔE values does not produce significant changes in the relative stability order outlined above. Thus, assuming a Boltzmann distribution, $\alpha_L/\gamma/E/g/t$ is the only conformation with a significant population at room temperature in the gas phase, since the ΔG_{gp} values of all other minima are above 1.5 kcal/mol (Table 4.1.2.). Furthermore, the backbone adopts an α_L conformation in the seven structures with lower ΔG_{gp} values, indicating that the preference for this helical fold is not altered by the addition of the statistical corrections. Not surprisingly, ΔG_{gp} values above 7 kcal/mol were obtained for all the minima with $\Delta E > 7$ kcal/mol and, therefore, the contribution of these structures to the conformational preferences of Ac-*t*-L-c₅Arg-NHMe are completely negligible.

Table 4.1.2. Relative conformational Gibbs free energies^a at 298K for selected^b minimum energy conformations of Ac-*t*-L-c₅Arg-NHMe in the gas phase (ΔG_{gp}), chloroform ($\Delta G_{\text{CHL}}^{\text{conf}}$) and aqueous solution ($\Delta G_{\text{WAT}}^{\text{conf}}$) characterized at the B3LYP/6-311+G(d,p) level.

#	Conformer	ΔG_{gp}	$\Delta G_{\text{CHL}}^{\text{conf}}$	$\Delta G_{\text{WAT}}^{\text{conf}}$
1	$\alpha_L/\gamma'E/g^+t$	0.0 ^c	2.0	3.2
2	$\alpha_L/\gamma'E/g^+t$	1.6	3.1	4.8
3	$\alpha_L/\gamma'E/tg^+$	2.8	0.0	0.0
4	$\alpha_L/\gamma'E/g^-g^-$	4.1	4.0	5.2
5	$\alpha_L/\alpha'E/g^+t$	2.4	3.6	4.1
6	$\alpha_L/\gamma'E/g^+t$	3.8	3.9	4.7
7	$\alpha_L/\gamma'E/tg^+$	4.2	1.2	1.3
8	$\alpha_L/\gamma'E/g^-g^-$	6.2	5.9	7.4
9	$C_7^{\text{eq}}/\alpha'E/tg^+$	6.4	1.5	1.3
10	$\beta_L/\alpha'E/g^+t$	6.9	6.7	8.4
11	$\alpha_L/\alpha'E/tg^+$	6.4	3.6	3.1
12	$P_{II}/\gamma'E/g^+t$	5.7	7.7	8.6
13	$C_7^{\text{eq}}/\alpha'E/tg^-$	8.2	3.8	4.2

^a In kcal/mol. ^b Minimum energy conformations with $\Delta E < 7.0$ kcal/mol (see Table 4.1.1.). ^c G=-856.436015 a.u

The conformational free energies estimated in chloroform ($\Delta G_{\text{CHL}}^{\text{conf}}$) and water ($\Delta G_{\text{WAT}}^{\text{conf}}$) solutions at the same temperature are also included in Table 4.1.2. As it can be seen, $\alpha_L/\gamma\text{E}/\text{tg}^+$ (Figure 4.1.4.c) becomes the most favored conformation in both solvents, with $\alpha_L/\gamma'\text{E}/\text{tg}^+$ (Figure 4.1.4.g) being destabilized by only 1.2 and 1.3 kcal/mol in chloroform and water, respectively. These results indicate that the solvent affects the arrangement of the side chain and, to some extent, the puckering of the cyclopentane ring but not the backbone, which in solution retains the preference for the α -helical conformation previously detected in the gas phase.

At this point, it is interesting to establish a comparison between the results obtained in this work for Ac-*t*-L-c₅Arg-NHMe and those recently reported for the analogous phenylalanine derivative, Ac-*t*-L-c₅Phe-NHMe.^{8a} Thus, for the *trans*-c₅Phe-containing compound, four different peptide backbone conformations were found to be energetically accessible in the gas phase. They correspond to C₇^{eq}, C₅, C₇^{ax} and α_R structures, the three latter being destabilized with respect to the former by 0.6, 1.0, and 1.5 kcal/mol, respectively. In comparison, only the α_L backbone conformation is accessible to Ac-*t*-L-c₅Arg-NHMe, with no other being located within $\Delta E < 5$ kcal/mol (Table 4.1.1.). The cyclopentane ring puckering propensities are also significantly different for the two compounds. Thus, αE , γE and $\alpha'\text{E}$ arrangements were characterized in the accessible minima of the *trans*-c₅Phe derivative,^{8a} whereas for Ac-*t*-L-c₅Arg-NHMe only the γE disposition is detected. Regarding the behavior in solution, the environment was found to alter the conformational preferences of the *trans*-c₅Phe derivative from a quantitative point of view but not qualitatively, that is, both compounds retain the main conformational trends observed in the gas phase.

This comparison provides evidence for the different roles played by the guanidinium and phenyl side groups in directing the conformational preferences of Ac₅c and indicates that the presence of a charged guanidinium group in the neighborhood of the carbonyl terminus (*trans*-c₅Arg) imposes more severe conformational constraints than those induced when an aromatic group is incorporated in the same position (*trans*-c₅Phe). This distinct behavior should be attributed to the different types of interactions that each side group may establish with the backbone. Thus, the phenyl substituent may affect the conformational propensities of the rest of the molecule by steric reasons or through the establishment of weak

attractive interactions of the N–H··· π type²¹ with the NH groups in the peptide backbone. In comparison, the guanidinium side chain mainly interacts with the backbone through the formation of hydrogen bonds. The latter interactions have a more marked directional character, involve the CO instead of the NH backbone groups and are associated with a much higher energy. As a consequence, the guanidinium side chain specifically oriented by the cyclopentane ring towards the carbonyl terminus has a greater impact on the conformational properties of the peptide backbone than a phenyl substituent, and induces conformations different to those encountered for peptides incorporating the unsubstituted Ac₅c^{4a} or the phenylalanine counterpart, *trans*-c₅Phe.^{8a}

Ac-c-L-c₅Arg-NHMe. The relevant structural parameters of the minimum energy conformations with $\Delta E < 7$ kcal/mol characterized for Ac-c-L-c₅Arg-NHMe are listed in Table 4.1.3. As it can be seen, only 9 minima satisfy this energetic criterion. It should be noted that, at variance with the compound described in the previous section, in this c₅Arg derivative the charged guanidinium group exhibits a *cis* relative orientation with respect to the amino substituent (Figure 4.1.2.). This should reflect in different interactions between the side chain and the rest of the molecule (both the cyclopentane ring and the peptide backbone) and therefore lead to different conformational propensities.

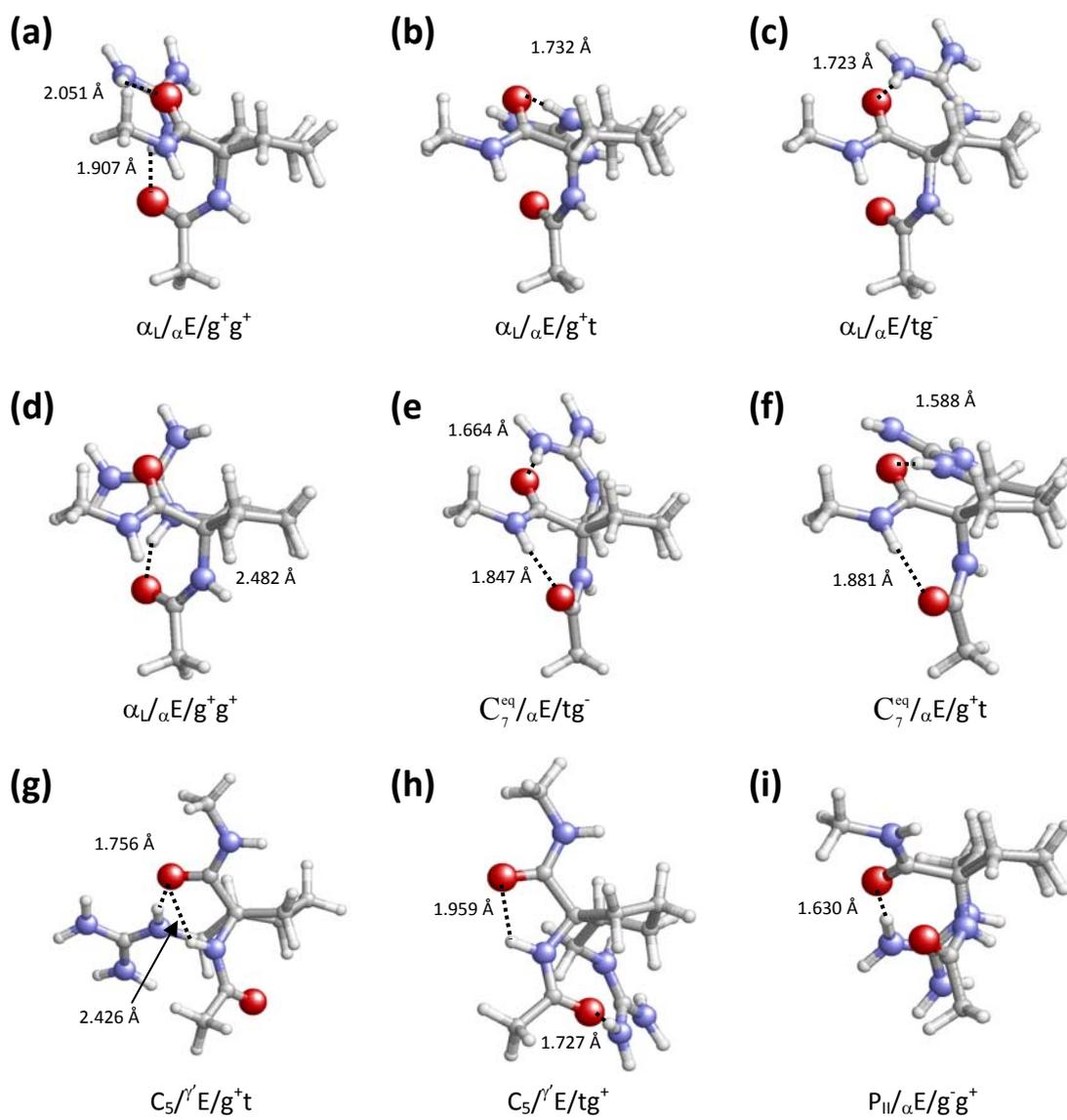


Figure 4.1.5. Lower minimum energy conformations of Ac-c-L-c₅Arg-NHMe obtained from B3LYP/6-311+G(d,p) calculations. The 9 structures depicted correspond to the minima listed in Table 4.1.3., *i.e.* minimum energy conformations with $\Delta E < 7$ kcal/mol.

Table 4.1.3. Dihedral angles^a and relative energies in the gas phase (ΔE) for the minimum energy conformations with $\Delta E < 7.0$ kcal/mol characterized for Ac-c-L-c₅Arg-NHMe at the B3LYP/6-311+G(d,p) level.

#	Conformer	Backbone dihedral angles				Cyclopentane dihedral angles					Side group	ΔE^b
		ω_0	φ	ψ	ω	χ^1	χ^2	χ^3	χ^4	χ^5	$\zeta_1\zeta_2$	
1	$\alpha_1/\alpha E/g^+g^+$	-178.5	48.0	48.6	180.0	40.4	-26.7	2.0	23.8	-39.6	63.0, 84.6	0.0 ^c
2	$\alpha_1/\alpha E/g^+t$	-178.5	48.6	51.6	179.4	43.4	-32.1	7.7	19.9	-38.9	69.5, -163.2	0.5
3	$\alpha_1/\alpha E/tg^-$	180.0	49.8	45.0	-178.7	44.6	-35.2	11.5	16.5	-37.3	151.6, -86.3	1.4
4	$\alpha_1/\alpha E/g^+g^+$	-178.3	49.6	42.5	178.1	41.3	-28.6	4.0	22.2	-39.0	54.2, 84.6	1.5
5	$C_7^{eq}/\alpha E/tg^-$	-172.0	-72.9	60.3	179.2	42.3	-29.9	5.4	21.0	-38.8	152.8, -84.2	2.3
6	$C_7^{eq}/\alpha E/g^+t$	-173.0	-72.8	67.1	-179.6	42.6	-28.7	2.9	24.3	-41.1	70.3, 161.9	3.1
7	$C_5/\alpha E/g^+t$	169.4	179.4	131.4	176.9	11.7	-33.6	42.7	-35.7	14.6	61.9, 166.4	4.9
8	$C_5/\alpha E/tg^+$	169.5	-179.9	172.2	-177.3	13.9	-34.6	42.1	-33.5	11.9	-156.0, 99.8	6.1
9	$P_{II}/\alpha E/g^+g^+$	-172.0	-57.8	129.1	-177.5	43.8	-29.4	3.3	24.5	-41.8	-66.6, 111.4	6.9

^a In degrees; see Figure 4.4.2. for definition. ^b In kcal/mol. ^c E=-856.7433141 a.u.

The lowest energy minimum characterized for Ac-*c*-L-*c*₅Arg-NHMe ($\alpha_L/\alpha E/g^+g^+$, Figure 4.1.5.a) corresponds to an α_L backbone conformation with the cyclopentane ring arranged as a C ^{α} -exo envelope and the two side-chain dihedral angles in *gauche*⁺. This spatial organization orientates both backbone carbonyl oxygens (those in the acetyl group and the *c*₅Arg residue) towards the guanidinium side chain, thus allowing the existence of two side-chain···backbone hydrogen bonds. The second ($\alpha_L/\alpha E/g^+t$, Figure 4.1.5.b), third ($\alpha_L/\alpha E/tg^-$, Figure 4.1.5.c) and fourth ($\alpha_L/\alpha E/g^+g^+$, Figure 4.1.5.d) minima only differ from the global one in the orientation of the guanidinium side group. These conformational transitions bring about significant changes in the hydrogen bonding scheme and an energy destabilization ranging from 0.5 to 1.5 kcal/mol.

The next minimum ($C_7^{eq}/\alpha E/tg^-$, Figure 4.1.5.e) adopts a different backbone conformation. Specifically, the backbone acetyl CO and methylamide NH groups form a seven-membered hydrogen-bonded ring typical of the C₇ or γ -turn conformation. Additionally, this structure is stabilized by a strong side-chain···backbone interaction involving one guanidinium NH₂ and the *c*₅Arg CO group. The arrangement of the cyclopentane ring is identical to that observed for the four preceding minima. Conformer # 6 ($C_7^{eq}/\alpha E/g^+t$, Figure 4.1.5.f) differs from # 5 in the orientation of the side group only, and this is associated with a change in the side-chain···backbone hydrogen-bonding pattern and an energy cost of 0.8 kcal/mol.

The next two structures, minima # 7 ($C_5'/E/g^+t$, Figure 4.1.5.g) and # 8 ($C_5'/E/tg^+$, Figure 4.1.5.h), correspond to a C₅ peptide backbone conformation, characterized by the presence of a hydrogen bond linking the *c*₅Arg NH and CO sites and closing a five-membered cycle. In the case of minimum # 7, the geometry of this pseudocycle is severely distorted –as evidenced by the small ψ angle– to allow the involvement of the same CO group in a strong hydrogen bond with the guanidinium NH site. The different orientation of the guanidinium side chain in minimum # 8 leads to an interaction with the acetyl CO group (instead of the *c*₅Arg CO), and the C₅ conformation accommodated by the peptide backbone is completely regular. It is also noteworthy that # 7 is the most stable minimum of the *cis*-*c*₅Arg derivative in which the flap of the cyclopentane envelope is not occupied by the α carbon. This means a

significant difference with reference to the behavior described above for the *trans*-*c*₅Arg derivative.

Finally, the last conformer in Table 4.1.3. ($P_{II/a}E/g^{\sim}g^+$, Figure 4.1.5.i.) is unfavored by 6.9 kcal/mol with respect to the global minimum and corresponds to a polyproline II conformation stabilized by a single side-chain···backbone interaction.

Inspection of Table 4.1.4. indicates that the lowest ΔG_{gp} value corresponds to the $\alpha_L/\alpha E/g^+t$ conformation (Figure 4.1.5.b), while the $\alpha_L/\alpha E/g^+g^+$ minimum (Figure 4.1.5.d) is destabilized by 1.8 kcal/mol. Accordingly, if a Boltzmann distribution is assumed, only the α_L backbone conformation and the C^{α} -exo envelope arrangement of the cyclopentane ring are populated in the gas phase at 298 K. For all the minima with $\Delta E > 7.0$ kcal/mol, ΔG_{gp} values above 5.1 kcal/mol were obtained and, therefore, the contribution of these structures to describe the conformational preferences of *cis*-*c*₅Arg can be considered as negligible.

Table 4.1.4. Relative conformational Gibbs free energies^a at 298K for selected^b minimum energy conformations of Ac-c-L-c₅Arg-NHMe in the gas phase (ΔG_{gp}), chloroform (ΔG_{CHL}^{conf}) and aqueous solution (ΔG_{WAT}^{conf}) characterized at the B3LYP/6-311+G(d,p) level.

#	Conformer	ΔG_{gp}	ΔG_{CHL}^{conf}	ΔG_{WAT}^{conf}
1	$\alpha_L/\alpha E/g^+g^+$	2.0	2.4	4.6
2	$\alpha_L/\alpha E/g^+t$	0.0 ^c	3.1	4.9
3	$\alpha_L/\alpha E/tg^-$	2.1	0.0	0.0
4	$\alpha_L/\alpha E/g^+g^+$	1.8	3.4	5.4
5	$C_7^{eq}/\alpha E/tg^-$	3.4	1.0	3.3
6	$C_7^{eq}/\alpha E/g^+t$	2.9	4.5	7.5
7	$C_5/\alpha' E/g^+t$	8.2	5.6	6.9
8	$C_5/\alpha' E/tg^+$	5.5	3.9	6.1
9	$P_{II}/\alpha E/g^-g^+$	7.7	3.8	4.3

^a In kcal/mol. ^b Minimum energy conformations with $\Delta E < 7.0$ kcal/mol (see Table 3). ^c G=-856.429193 a.u.

The values of ΔG_{CHL}^{conf} and ΔG_{WAT}^{conf} are also listed in Table 4.1.4. As it can be seen, the only conformations with $\Delta G_{CHL}^{conf} < 1.5$ kcal/mol are $\alpha_L/\alpha E/tg^-$ (Figure 4.1.5.c), which is the global minimum in chloroform solution, and $C_7^{eq}/\alpha E/tg^-$ (Figure 4.1.5.e), that is destabilized by 1.0 kcal/mol. Accordingly, both the α_L and C_7^{eq} backbone conformations are expected to exhibit significant populations in this organic solvent, while the disposition of the cyclopentane ring and the guanidinium side chain seems to be restricted to the C ^{α} -exo envelope and tg^- arrangements, respectively. On the other hand, the lowest energy minimum in aqueous solution is $\alpha_L/\alpha E/tg^-$ (Figure

4.1.5.c), all the other structures being destabilized by more than 3.3 kcal/mol. It is worth noting that the $\alpha_L/\alpha E/g^+t$ conformation (Figure 4.1.5.b), which presented the lowest ΔG_{gp} value, is disfavored by 3.1 and 4.9 kcal/mol in chloroform and water, respectively.

Again, significant differences are detected when the results obtained in this work for Ac-*c*-L-c₅Arg-NHMe are compared to those previously described for Ac-*c*-L-c₅Phe-NHMe.^{8a} Notably, the conformational space available to the latter compound was found to be more restricted than that of the *trans* derivative as a consequence of the high proximity between the amino terminus and the bulky, rigidly held aromatic substituent. Thus, only equatorial C₇ conformers were found to be accessible at room temperature for Ac-*c*-L-c₅Phe-NHMe. Indeed, for this compound, minima with an α_L backbone arrangement were destabilized by more than 5 kcal/mol, whereas this is the only backbone structure energetically accessible to *cis*-c₅Arg (Table 4.1.3.). Also the cyclopentane arrangement was found to be substantially different. Thus, the βE disposition, which places the β carbon bearing the bulky phenyl ring out of the plane defined by the other cyclopentane atoms, proved the most favorable one for the *cis*-c₅Phe derivative,^{8a} whereas *cis*-c₅Arg largely prefers the αE arrangement. In chloroform and aqueous solution, C₇^{eq}/ βE remained the only structure energetically accessible to Ac-*c*-L-c₅Phe-NHMe. Accordingly, the conformational preferences of *cis*-c₅Phe and *cis*-c₅Arg are substantially different since they depend, to a large extent, on the need to relieve steric hindrance in the former case and on the ability of the side chain to form hydrogen bonds with the backbone in the latter.

Classical Molecular Dynamics simulations in aqueous solution. In the absence of experimental data, classical MD simulations with explicit solvent molecules are valuable for describing the favored low-energy conformations of peptides. In order to explore the conformational energy surfaces of Ac-*t*-L-c₅Arg-NHMe and Ac-*c*-L-c₅Arg-NHMe in aqueous solution using this methodology, a specific force-field parametrization to represent the stretching, bending, torsional, van der Waals and electrostatic interactions of these constrained peptides was required. In a previous study we showed that no special electronic effect is present in Ac₅c^{4a} and, therefore, the stretching, bending, torsional and van der Waals parameters for Ac₅c and its derivatives can be directly transferred from the Amber force-field.²² Accordingly,

electrostatic charges have been the only force-field parameters specifically developed for *trans*- and *cis*- c_5 Arg.

Atomic charges for the five minimum energy conformations listed in Table 4.1.1. and 3 were calculated by fitting the HF/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.²² Electrostatic potential (ESP) fitting atomic centered charges for *trans*- and *cis*- c_5 Arg were derived by weighting the charges calculated for the corresponding minimum energy conformations according to Boltzmann populations.²³ The weights were given by the standard Boltzmann formula using the ΔG_{gp} values given in Tables 4.1.2. and 4.1.4. As the charges obtained for *trans*- and *cis*- c_5 Arg were similar, *i.e.* the absolute value of the largest was lower than 0.08 e.u., we decided to simplify the force-field by providing a unique set of electrostatic parameters for the two amino acids (Figure 4.1.6.).

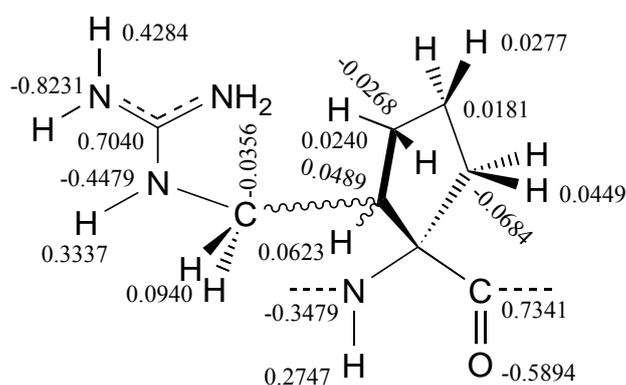


Figure 4.1.6. Electrostatic parameters determined for c_5 Arg residues.

MD simulations of Ac-*t*-L- c_5 Arg-NHMe and Ac-*c*-L- c_5 Arg-NHMe were performed at 350 K. The lowest energy conformation was used as starting point of a trajectory that was 10 ns long for each compound. Figure 4.1.7. represents the accumulated Ramachandran plot for the *trans*- and *cis*- c_5 Arg dipeptides. In both cases the most populated conformation in aqueous solution corresponds to the α_L , which is visited much more frequently than the C_7^{eq} conformation during the trajectory. This fact is in excellent agreement with the results displayed in Table 4.1.2. and 4.1.4.,

which indicate that the α_L conformation is the lowest energy minimum. These evidences clearly confirm that the conformational space of both *trans*- and *cis*- c_5 Arg is severely restricted by the constrains imposed not only by the cyclopentane ring but also by the charged guanidinium group, which establish hydrogen-bond interactions with the peptide backbone.

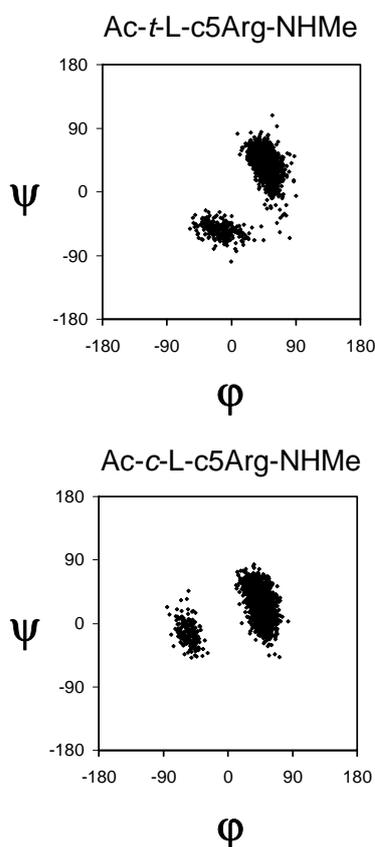
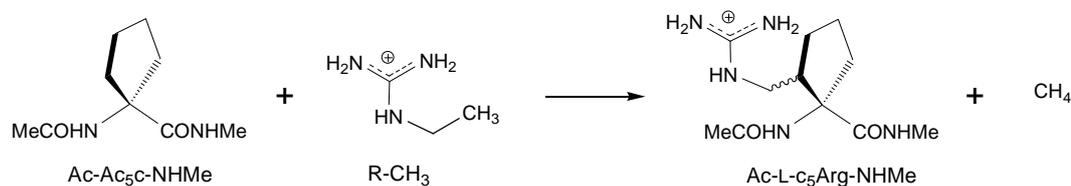


Figure 4.1.7. Accumulated Ramachandran plot for Ac-*t*-L- c_5 Arg-NHMe and Ac-*c*-L- c_5 Arg-NHMe derived from a MD trajectory 10 ns long in aqueous solution.

Influence of the guanidinium side group in the conformational properties. In order to evaluate quantitatively the consequences arising from the incorporation of the guanidinium side group in Ac $_5$ c to generate *cis*- and *trans*- c_5 Arg, the isodesmic reaction displayed in Scheme 1 has been considered. For different backbone conformations and cyclopentane ring arrangements, the energy (E^{SB}) and free energy

(G^{sg}) contribution associated with the side group in Scheme 1 were estimated according to equations (1) and (2), respectively.



Scheme 1

$$E^{\text{sg}} = E^{\text{Ac-L-c}_5\text{Arg-NHMe}} + E^{\text{CH}_4} - (E^{\text{Ac-Ac}_5\text{c-NHMe}} + E^{\text{CH}_3\text{-R}}) \quad (1)$$

$$G^{\text{sg}} = G^{\text{Ac-L-c}_5\text{Arg-NHMe}} + G^{\text{CH}_4} - (G^{\text{Ac-Ac}_5\text{c-NHMe}} + G^{\text{CH}_3\text{-R}}) \quad (2)$$

In these equations, E^{sg} and G^{sg} provide an estimation of the energy and free energy contribution, respectively, associated to the incorporation of the guanidinium side group for a given backbone conformation and cyclopentane ring puckering of Ac_5c . Table 4.1.5. shows the values calculated considering selected minimum energy conformations of both $\text{Ac-}t\text{-L-c}_5\text{Arg-NHMe}$ (Table 4.1.1.) and $\text{Ac-}c\text{-L-c}_5\text{Arg-NHMe}$ (Table 4.1.3.), namely the most stable ones among those for which $\text{Ac-Ac}_5\text{c-NHMe}$ was found^{4a} to exhibit a minimum with similar backbone conformation and cyclopentane puckering. As no α_{L} minimum with an ${}^{\alpha}\text{E}$ cyclopentane arrangement was characterized for the *cis-c*₅Arg derivative, the $\alpha_{\text{L}}/\alpha\text{E}/g^+g^+$ conformation (the global minimum in Table 4.1.3.) was considered in this case. On the other hand, it should be noted that the minimum energy conformations previously obtained for $\text{Ac-Ac}_5\text{c-NHMe}$ through B3LYP/6-311G(d,p) calculations^{4a} have been re-optimized at the B3LYP/6-311+G(d,p) level.

The negative values obtained for both E^{sg} and G^{sg} (Table 4.1.5.) reveal significant favorable interactions for all the conformations of Ac-*t*-L-c₅Arg-NHMe and Ac-*c*-L-c₅Arg-NHMe considered. Specifically, the attractive interactions between the charged side group and the polar backbone amide groups produce a significant stabilization for the α_{L} , C₇^{eq} and C₅ backbone conformations. The strength of this effect is fully consistent with the relative energies and free energies obtained for such conformations, the most and least attractive interaction being obtained for the α_{L} and C₅ structures, respectively. Overall, these results indicate that the remarkable preference of c₅Arg towards the α_{L} helical conformation is due to the formation of strong side-chain···backbone interactions, which are more attractive than those established for other backbone conformations. As expected, G^{sg} is higher than E^{sg} in all cases, which should be attributed to the unfavorable entropic contribution associated with the disappearance of the strong side chain···backbone interactions.

Table 4.1.5. Energy (E^{sg}) and free energy (G^{sg}) contributions associated with the guanidinium side group for selected backbone conformations of Ac-*t*-L-c₅Arg-NHMe and Ac-*c*-L-c₅Arg-NHMe.

<i>Compound</i>	<i>Conf. L-c₅Arg</i>	<i>Conf. Ac₅c^a</i>	E^{sg}	G^{sg}
Ac- <i>t</i> -L-c ₅ Arg-NHMe	$\alpha_{\text{L}}/\alpha\text{E}/\text{g}^+\text{t}$	$\alpha_{\text{L}}/\alpha\text{E}$	-17.1	-13.8
	$\text{C}_7^{\text{eq}}/\alpha\text{E}/\text{tg}^+$	$\text{C}_7^{\text{eq}}/\alpha\text{E}$	-12.5	-8.4
	$\text{C}_7^{\text{eq}}/\alpha\text{E}/\text{tg}^-$	$\text{C}_7^{\text{eq}}/\alpha\text{E}$	-11.1	-7.3
Ac- <i>c</i> -L-c ₅ Arg-NHMe	$\alpha_{\text{L}}/\alpha\text{E}/\text{g}^+\text{g}^+$	$\alpha_{\text{L}}/\alpha\text{E}$	-19.5	-13.7
	$\text{C}_7^{\text{eq}}/\alpha\text{E}/\text{tg}^-$	$\text{C}_7^{\text{eq}}/\alpha\text{E}$	-14.7	-10.9
	$\text{C}_5/\beta'\text{E}/\text{g}^+\text{t}$	$\text{C}_5/\beta'\text{E}$	-13.2	-9.8

^a From ref. 4a [minima re-optimized at the B3LYP/6-311+G(d,p) level]

4.1.4. Conclusions

The conformational preferences of Ac-*t*-L-c₅Arg-NHMe and Ac-*c*-L-c₅Arg-NHMe have been explored using quantum mechanical calculations at the B3LYP/6-311+G(d,p) level. Results indicate that the *cis* and, particularly, the *trans* stereoisomers of c₅Arg prefer an α -helical conformation. Thus, all the minima found for Ac-*t*-L-c₅Arg-NHMe and Ac-*c*-L-c₅Arg-NHMe with $\Delta E \leq 4.4$ and 1.5 kcal/mol, respectively, exhibit this peptide backbone structure. Furthermore, the preference for the α -helical conformation is retained in solution. Also the cyclopentane ring puckering is significantly affected by the presence and orientation of the guanidinium side chain, and thus, *cis*- and *trans*-c₅Arg show a marked preference for the C ^{γ} -exo and C ^{α} -exo envelope arrangements, respectively, in all environmental conditions considered.

The structural preferences exhibited by the c₅Arg derivatives are in high contrast with those previously observed for the analogous phenylalanine derivatives, Ac-*t*-L-c₅Phe-NHMe and Ac-*c*-L-c₅Phe-NHMe, which have been shown to prefer the C₇^{eq} arrangement. The unique conformational properties observed for c₅Arg should be attributed to the ability of the side-chain guanidinium group to establish hydrogen-bond interactions with the peptide backbone, which are particularly attractive when the backbone adopts a helical conformation. The present work provides evidence for the ability of the side chain to influence the peptide backbone conformation and, specifically, illustrates how the latter may be affected by the side-chain nature and orientation.

4.1.5. 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) Benedetti, E. *Biopolymers (Pept. Sci.)* 1996, 40, 3. (e) Toniolo, C.; Benedetti, E. *Macromolecules* 1991, 24, 4004.
- [2] (a) Ballano, G.; Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* 2008, 112, 13101. (b) Gómez-Catalán, J.; Alemán, C.; Pérez, J. J. *Theor. Chem. Acc.* 2000, 103, 380. (c) Alemán, C. *J. Phys. Chem. B* 1997, 101, 5046. (d) Valle, G.; Crisma, M.; Toniolo, C.; Holt, E. M.; Tamura, M.; Bland, J.; Stammer, C. H. *Int. J. Pept. Protein Res.* 1989, 34, 56. (e) Benedetti, E.; Di Blasio, B.; Pavone, V.; Pedone, C.; Santini, A.; Crisma, M.; Valle, G.; Toniolo, C. *Biopolymers* 1989, 28, 175. (f) Crisma, M.; Bonora, G. M.; Toniolo, C.; Barone, V.; Benedetti, E.; Di Blasio, B.; Pavone, V.; Pedone, C.; Santini, A.; Fraternali, F.; Bavoso, A.; Lelj, F. *Int. J. Biol. Macromol.* 1989, 11, 345. (g) Benedetti, E.; Di Blasio, B.; Pavone, V.; Pedone, C.; Santini, A.; Barone, V.; Fraternali, F.; Lelj, F.; Bavoso, A.; Crisma, M.; Toniolo, C. *Int. J. Biol. Macromol.* 1989, 11, 353. (h) Barone, V.; Fraternali, F.; Cristinziano, P. L.; Lelj, F.; Rosa, A. *Biopolymers* 1988, 27, 1673.
- [3] (a) Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *Chem. Phys. Lett.* 2006, 429, 558. (b) Rao, S. N.; Chan, M. F.; Balaji, V. N. *Bull. Chem. Soc. Jpn.* 1997, 70, 293. (c) Gatos, M.; Formaggio, F.; Crisma, M.; Toniolo, C.; Bonora, G. M.; Benedetti, E.; Di Blasio, B.; Iacovino, R.; Santini, A.; Saviano, M.; Kamphuis, J. J. *Pept. Sci.* 1997, 3, 110. (d) Balaji, V. N.; Ramnarayan, K.; Chan, M. F.; Rao, S. N. *Peptide Res.* 1995, 8, 178.
- [4] (a) Alemán, C.; Zanuy, D.; Casanovas, J.; Cativiela, C.; Nussinov, R. *J. Phys. Chem. B* 2006, 110, 21264. (b) Aschi, M.; Lucente, G.; Mazza, F.; Mollica, A.; Morera, E.; Nalli, M.; Paglialunga Paradisi, M. *Org. Biomol. Chem.* 2003, 1, 1980. (c) Valle, G.; Crisma, M.; Toniolo, C. *Can. J. Chem.* 1988, 66, 2575. (d) Crisma, M.; Bonora, G. M.; Toniolo, C.; Benedetti, E.; Bavoso, A.; Di Blasio, B.; Pavone, V.; Pedone, C. *Int. J. Biol. Macromol.* 1988, 10, 300. (e) Santini, A.; Barone, V.; Bavoso, A.; Benedetti, E.; Di Blasio, B.; Fraternali, F.; Lelj, F.; Pavone, V.; Pedone, C.; Crisma, M.; Bonora, G. M.; Toniolo, C. *Int. J. Biol. Macromol.* 1988, 10, 292.
- [5] (a) Harini, V. V.; Aravinda, S.; Rai, R.; Shamala, N.; Balaram, P. *Chem. Eur. J.* 2005, 11, 3609. (b) Crisma, M.; Bonora, G. M.; Toniolo, C.; Bavoso, A.; Benedetti,

E.; Di Blasio, B.; Pavone, V.; Pedone, C. *Macromolecules* 1988, 21, 2071. (c) Pavone, V.; Benedetti, E.; Barone, V.; Di Blasio, B.; Lelj, F.; Pedone, C.; Santini, A.; Crisma, M.; Bonora, G. M.; Toniolo, C. *Macromolecules* 1988, 21, 2064. (d) Paul, P. K. C.; Sukumar, M.; Bardi, R.; Piazzesi, A. M.; Valle, G.; Toniolo, C.; Balaram, P. *J. Am. Chem. Soc.* 1986, 108, 6363.

[6] (a) Prasad, S.; Mathur, A.; Jaggi, M.; Singh, A. T.; Mukherjee, R. *J. Pept. Sci.* 2007, 13, 544. (b) Labudda-Dawidowska, O.; Wierzba, T. H.; Prahl, A.; Kowalczyk, W.; Gawinski, L.; Plackova, M.; Slaninová, J.; Lammek, B. *J. Med. Chem.* 2005, 48, 8055. (c) Kowalczyk, W.; Prahl, A.; Dawidowska, O.; Derdowska, I.; Sobolewski, D.; Hartrodt, B.; Neubert, K.; Slaninová, J.; Lammek, B. *J. Pept. Sci.* 2005, 11, 584. (d) Kowalczyk, W.; Prahl, A.; Derdowska, I.; Dawidowska, O.; Slaninová, J.; Lammek, B. *J. Med. Chem.* 2004, 47, 6020. (e) García-Echeverría, C.; Gay, B.; Rahuel, J.; Furet, P. *Bioorg. Med. Chem. Lett.* 1999, 9, 2915. (f) Breveglieri, A.; Guerrini, R.; Salvadori, S.; Bianchi, C.; Bryant, S. D.; Attila, M.; Lazarus, L. H. *J. Med. Chem.* 1996, 39, 773. (g) Sukumar, M.; Raj, P. A.; Balaram, P.; Becker, E. L. *Biochem. Biophys. Res. Commun.* 1985, 128, 339.

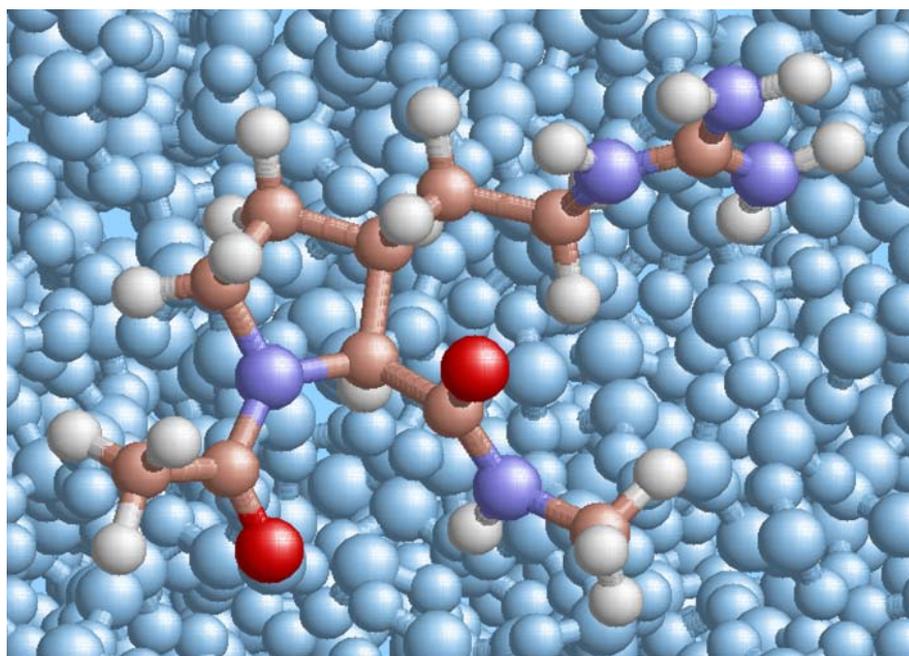
[7] (a) Lasa, M.; Cativiela, C. *Synlett* 2006, 2517. (b) Lasa, M.; López, P.; Cativiela, C. *Tetrahedron: Asymmetry* 2005, 16, 4022. (c) Cativiela, C.; Lasa, P.; López, P. *Tetrahedron: Asymmetry* 2005, 16, 2613. (d) Cativiela, C.; López, M.; Lasa, M. *Eur. J. Org. Chem.* 2004, 3898. (e) Alías, M.; Cativiela, C.; Jiménez, A. I.; López, P.; Oliveros, L.; Marraud, M. *Chirality* 2001, 13, 48. (f) Cativiela, C.; Díaz-de-Villegas, M. D.; Jiménez, A. I.; López, P.; Marraud, M.; Oliveros, L. *Chirality* 1999, 11, 583.

[8] (a) Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Org. Chem.* 2008, 73, 644. (b) Lasa, M.; Jiménez, A. I.; Zurbano, M. M.; Cativiela, C. *Tetrahedron Lett.* 2005, 46, 8377. (c) Alemán, C.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Casanovas, J. *J. Phys. Chem. B* 2002, 106, 11849. (d) Gomez-Catalan, J.; Jiménez, A. I.; Cativiela, C.; Perez, J. J. *J. Pept. Res.* 2001, 57, 435. (e) Jiménez, A. I.; Cativiela, C.; Gómez-Catalán, J.; Pérez, J. J.; Aubry, A.; París, M.; Marraud, M. *J. Am. Chem. Soc.* 2000, 122, 5811. (f) Jiménez, A. I.; Cativiela, C.; París, M.; Peregrina, J. M.; Avenoza, A.; Aubry, A.; Marraud, M. *Tetrahedron Lett.* 1998, 39, 7841. (g) Jiménez, A. I.; Cativiela, C.; Aubry, A.; Marraud, M. *J. Am. Chem. Soc.* 1998, 120, 9452. (h) Jiménez, A. I.; Vanderesse, R.; Marraud, M.; Aubry, A.; Cativiela, C. *Tetrahedron Lett.* 1997, 38, 7559.

- [9] (a) Crisma, M.; Toniolo, C.; Royo, S.; Jiménez, A. I.; Cativiela, C. *Org. Lett.* 2006, 8, 6091. (b) Crisma, M.; De Borggraeve, W. M.; Peggion, C.; Formaggio, F.; Royo, S.; Jiménez, A. I.; Cativiela, C.; Toniolo, C. *Chem. Eur. J.* 2006, 12, 251. (c) Jiménez, A. I.; Ballano, G.; Cativiela, C. *Angew. Chem. Int. Ed.* 2005, 44, 396. (d) Royo, S.; De Borggraeve, W. M.; Peggion, C.; Formaggio, F.; Crisma, M.; Jiménez, A. I.; Cativiela, C.; Toniolo, C. *J. Am. Chem. Soc.* 2005, 127, 2036.
- [10] 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.
- [11] Becke, A. D. *J. Chem. Phys.* 1993, 98, 1372.
- [12] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* 1993, 37, 785.
- [13] McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* 1980, 72, 5639.
- [14] (a) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* 1981, 55, 117. (b) Miertus, S.; Tomasi, J. *Chem. Phys.* 1982, 65, 239. (c) Tomasi, J.; Persico, M. *Chem. Phys.* 1994, 94, 2027. (b) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* 2005, 105, 2999.
- [15] Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys. B* 1998, 102, 3257. (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) Iribarren, J. I.; Casanovas, J.; Zanuy, D.; Alemán, C. *Chem. Phys.* 2004, 302, 77.

- [16] Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* 1999, *151*, 283.
- [17] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* 1982, *79*, 926.
- [18] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* 1993, *98*, 10089.
- [19] Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Comput. Phys.* 1990, *94*, 1683.
- [20] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* 1984, *81*, 3684.
- [21] (a) Steiner, T.; Koellner, G. *J. Mol. Biol.* 2001, *305*, 535. (b) Worth, G. A.; Wade, R. C. *J. Phys. Chem.* 1995, *99*, 17473. (c) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* 1994, *239*, 315. (d) Mitchell, J. B. O.; Nandi, C. L.; Ali, S.; McDonald, I. K.; Thornton, J. M. *Nature* 1993, *366*, 413.
- [22] 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.
- [23] Alemán, C.; Casanovas *J. Chem. Soc. Perkin Trans. 2* 1994, 563.

4.2. Conformational profile of a proline-arginine hybrid



4.2.1. Introduction

Design of specific chemical modifications in natural amino acids is a powerful strategy to control the conformational properties of short peptides.¹ Moreover, non-coded residues in a peptide chain may result in increased resistance to proteolytic degradation.² Non-proteinogenic amino acids are useful in engineering peptide analogues with improved pharmacokinetics and medicinal applications.^{2,3}

A non-coded amino acid can replace residues in natural peptide sequences if it does not disrupt the *bioactive* conformation of the preplaced segment. The resulting peptide must preserve the native shape that interacts with the receptor. When dealing with small flexible peptides, a non-coded residue can improve the targeted peptide by biasing its conformational equilibrium to a conformational set that guarantees function. Thus, the non-proteinogenic amino acid should exhibit a high preference for the conformation adopted by the natural residue that is to be replaced.

Theoretical methods can assist so that only those candidates yielding satisfactory results *in silico* are selected for experimental studies. This evaluation requires (i) theoretical study of the wild-type bioactive conformation (if not available experimentally); (ii) assessing which amino acid should be replaced; and (iii) design

of a new non-coded amino acid adequate to replace the targeted position of the peptide. The conformational preferences of the new non-coded amino acid need to be determined before the replacement is performed. If such conformational preferences do not match those of the targeted position, the replacement will not be successful.

We are involved in a project aimed at improving the bioactivity of the pentapeptide Cys-Arg-Glu-Lys-Ala (CREKA). This peptide has biotechnological interests because it recognizes molecular markers that are present in tumor blood vessels but not in the vasculature of normal tissues,⁴ thus showing promising applications in cancer diagnosis and therapy. Medicinal use is however hampered by the poor stability against proteases and short half-life time typically exhibited by small and medium-size natural peptides. The conformational landscape of CREKA was explored by computational methods under different environmental conditions.⁵ This analysis led to a bioactive conformation exhibiting a turn motif, with the charged side chains of Arg, Glu, and Lys oriented toward the same side of the molecule.⁵ The peptide backbone is folded in β -turn centered at the Arg and Glu residues. Arginine occupies the first corner position of the β -turn ($i+1$) and adopts dihedral angles corresponding to the α_L -helical (α_L) region of the Ramachandran map.

The first attempt to decrease the peptide sensitivity to proteases was made though by introducing unspecific chemical modifications that did not affect its overall conformational properties.⁶ Hence, a methyl group replaced the hydrogen atom at either N position or C $^\alpha$ position. This simple approach increased *in vivo* the half-life time of the CREKA coated nanoparticles in the tumor vessels.⁶ However, no effort had been made to enhance the stability of the bioactive conformation.⁵ Then in a second stage, we undertook the design of new non-coded amino acids that could bias the folding of CREKA towards its bioactive organization, focusing our efforts on arginine surrogates. The main goal was to incorporate the side-chain functionality of arginine, which is essential for CREKA's activity, in residue that presented clear conformational preferences for the α_L region of the Ramachandran map.

A first amino acid was designed by combining a non-coded amino acid of the family 1-aminocycloalkane-1-carboxylic acids (Ac $_n$ c, where n refers to the size of the cycle) and the side chain functionality of arginine (Figure 4.2.1.).⁷ These amino acid series had been previously investigated and shown to exhibit a restricted conformational space characterized by a high propensity to adopt φ, ψ backbone angles typical of the 3_{10} - α -helix (with some distortion in the case of Ac $_3$ c).⁸ The new amino

acid (denoted as $c_5\text{Arg}$) was built by incorporating the side chain of arginine at the β -carbon atom of Ac_5c .⁷ The intrinsic conformational preferences of the new amino acid were studied using theoretical methods, showing that α -helical conformation was favored both in the gas phase and in solution. It was remarkable that the ability of the guanidinium moiety to form hydrogen bonds with the peptide backbone conditioned the conformational features of the parent Ac_5c ,^{8d} which tends to favor the formation γ -turn based conformations before α -helix like arrangements.

The latter challenge was though to achieve similar results with a backbone constitution closer to that of coded amino acids. Among proteinogenic amino acids, proline is known to impart protection against proteolytic cleavage⁹ as well as to nucleate peptide turns,¹⁰ with a marked propensity to occupy the $i+1$ position of β -turns. Accordingly, following the previous strategy we generated a new residue by attaching the arginine side chain to the proline skeleton (see the Supporting Information for details). It should be noted that the cyclic nature of proline, that includes the amine nitrogen atom in the ring constitution, facilitates a *cis* arrangement of the peptide bond involving the pyrrolidine nitrogen, as compared to other peptide bonds, for which the *cis* form is almost nonexistent.¹¹ Here, however, this issue has not been addressed since the targeted arginine in wild-type CREKA presents both peptide bonds in *trans*⁵ and the new residue is therefore useful only for the latter geometry.

In a previous work,¹² the guanidilated side chain was attached to the γ -carbon of the pyrrolidine ring in a *cis* configuration with the carboxylic acid moiety, thus giving rise to the residue denoted *cis*-(γ Pro)Arg in Figure 4.2.1.. The chain length of this arginine analogue proved insufficient to reproduce the interactions observed for the guanidinium group in the bioactive conformation of CREKA.^{5,12} Although the addition of another methylene unit to the exocyclic guanidilated substituent was favorable, incorporation of the resulting residue into CREKA led to the disruption of the β -turn conformation of the natural peptide.¹²

These results led us to design a new arginine surrogate built on a proline skeleton. Here, it is the β pyrrolidine carbon that bears the guanidilated arginine side chain and the resulting residue is termed *cis*-(β Pro)Arg, where *cis* refers to the position of the guanidilated substituent relative to the carboxylic acid and β denotes the carbon atom of the five-membered ring where this substituent is placed (Figure 4.2.1.). Prior to testing the modified pentapeptide Cys-*cis*(β Pro)Arg-Glu-Lys-Ala, the conformational

propensities of the single amino acid have been investigated in depth by theoretical methods. As noted above, *cis*-(β Pro)Arg presents a *cis* orientation between the guanidinium and carbonyl moieties –as the previously studied¹² *cis*-(γ Pro)Arg– in agreement with the spatial relationship characterized for the guanidilated segment of natural arginine in CREKA.⁵ However, beyond the CREKA project, other peptides incorporating key arginine residues may present the guanidilated side chain oriented away from the carbonyl group in the bioactive form. This consideration prompted us to also evaluate in this work the conformational propensities of *trans*-(β Pro)Arg (Figure 4.2.1.).

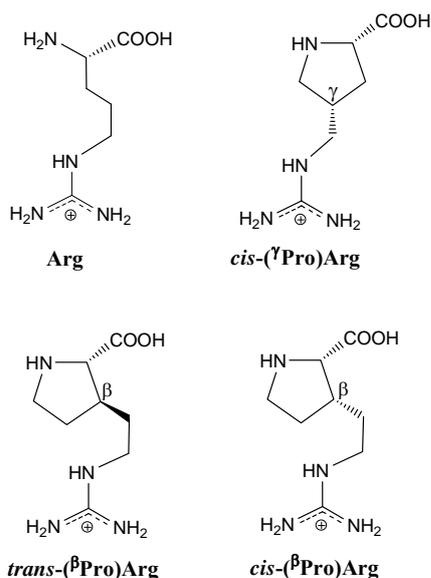


Figure 4.2.1.. Structure of arginine (Arg) and its proline-like analogues studied previously, *cis*-(γ Pro)Arg (ref. 8), and in the present work, *trans*-(β Pro)Arg and *cis*-(β Pro)Arg. The arginine surrogates are named according to the γ/β position of the proline skeleton bearing the guanidilated side chain and to the *trans/cis* relative orientation between this side chain and the carboxylic acid.

We have therefore performed quantum mechanics calculations on the *N*-acetyl-*N'*-methylamide derivatives of both *cis*-(β Pro)Arg and *trans*-(β Pro)Arg, hereafter denoted Ac-*c*-(β Pro)Arg-NHMe and Ac-*t*-(β Pro)Arg-NHMe, respectively. The results are compared with those reported previously¹² for the analogous *cis* γ -substituted derivative, Ac-*c*-(γ Pro)Arg-NHMe. Additionally, parameterization of the two non-proteinogenic amino acids under study has been carried out before analyzing the

conformational impact derived from their incorporation into biologically active peptides. The dynamical conformational features of the two residues have been explored in aqueous solution at room temperature using classical Molecular Dynamics (MD) simulations with explicit water molecules.

4.2.2. Methods

Quantum mechanical calculations. Density Functional Theory (DFT) methods were applied for quantum mechanical calculations, which were performed using the Gaussian 03 computer program.¹³ Specifically, calculations were carried out by combining the unrestricted formalism of the B3LYP functional^{14,15} with the 6-31+G(d,p) basis set.¹⁶ 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) and both thermal and entropic corrections. These statistical terms were then used to compute the conformational Gibbs free energies in the gas phase (ΔG^{gp}) at 298K.

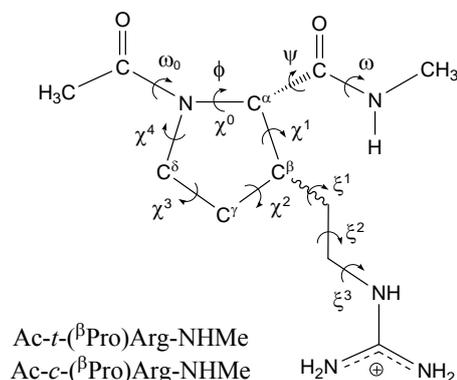


Figure 4.2.2.. Dihedral angles used to identify the conformations of the *N*-acetyl-*N'*-methylamide derivatives of *trans*- $(\beta\text{Pro})\text{Arg}$ and *cis*- $(\beta\text{Pro})\text{Arg}$ studied in this work. The (φ, ψ) dihedral angles are defined by the atoms in the backbone, whereas the side-chain dihedral angles χ^i and ξ^i are given by the pyrrolidine atoms and the exocyclic side-chain atoms, respectively. In particular, φ and χ^0 are defined as $\text{C}(\text{O})\text{-N-C}^\alpha\text{-C}(\text{O})$ and $\text{C}^\delta\text{-N-C}^\alpha\text{-C}^\beta$, respectively. The dihedral angle ξ^1 is given by $\text{C}^\alpha\text{-C}^\beta\text{-C}(\text{H}_2)\text{-C}(\text{H}_2)$.

Figure 4.2.2. shows the backbone ($\omega_0, \varphi, \psi, \omega$) and side chain (χ^i, ξ^i) dihedral angles that define the conformations adopted by *Ac-t*- $(\beta\text{Pro})\text{Arg-NHMe}$ and *Ac-c*- $(\beta\text{Pro})\text{Arg-NHMe}$. The minimum energy structures characterized for these compounds have been denoted using a three-label code that specifies the arrangement of the peptide backbone, the puckering of the five-membered cycle and the conformation adopted by

the exocyclic substituent. The first label identifies the backbone conformation type according to Perczel's nomenclature,¹⁷ which categorizes the potential energy surface $E = E(\varphi, \psi)$ of α -amino acids in nine different regions: γ_D , δ_D , α_D , ε_D , β_{DL} , ε_L , α_L , δ_L , and γ_L . The presence of the pyrrolidine ring in proline fixes the φ angle near -60° and, accordingly, only three of such regions can be accessed,¹⁰ namely, γ_L (γ -turn), α_L (α -helix), and ε_L (polyproline II). Identical geometric restrictions should apply to the arginine surrogates under study since they have a proline skeleton. A *trans* configuration was considered for the amide bonds (ω_0 , $\omega \approx 180^\circ$). The second label describes the *down* [d] or *up* [u] puckering of the five-membered pyrrolidine ring.^{10a,18} Such conformational states are also called C^{γ}_{endo} and C^{γ}_{exo} , respectively, and correspond to those in which the C^{γ} atom and the carbonyl group of proline (or the proline-like residue) lie on the same and opposite sides of the plane defined by C^{δ} , N, and C^{α} . Specifically, a *down* puckering was assigned when χ^1 and χ^3 were positive while χ^2 and χ^4 were negative. Conversely, negative values of χ^1 and χ^3 and positive values of χ^2 and χ^4 correspond to an *up*-puckered pyrrolidine ring. Finally, the orientation of the polar exocyclic substituent is described by the third label, which indicates the *gauche*⁺ (g^+), *skew*⁺ (s^+), *trans* (t), *skew*⁻ (s^-), *gauche*⁻ (g^-), or *cis* (c) state of each ξ^i dihedral angle.

The conformational search was performed following the strategy used in our previous work on *cis*-($^{\gamma}$ Pro)Arg.¹² It was assumed that the two ($^{\beta}$ Pro)Arg derivatives under study maintain the geometric restrictions derived from the cyclic nature of proline. Thus, the three minimum energy conformations characterized¹⁹ for Ac-Pro-NHMe with *trans* amide bonds, *i.e.* $\gamma_L[d]$, $\gamma_L[u]$, and $\alpha_L[u]$, were considered as starting geometries for Ac-*t*-($^{\beta}$ Pro)Arg-NHMe and Ac-*c*-($^{\beta}$ Pro)Arg-NHMe in the present work. Regarding the substituent attached to the β -carbon of the pyrrolidine moiety, each ξ^i dihedral was expected to exhibit minima of the *gauche*⁺, *trans* and *gauche*⁻ type. Accordingly, 3 (minima of Ac-Pro-NHMe¹⁶) \times 3 (minima of ξ^1) \times 3 (minima of ξ^2) \times 3 (minima of ξ^3) = 81 minima were anticipated for the potential energy hypersurface $E = E(\varphi, \psi, \chi^i, \xi^i)$ of each ($^{\beta}$ Pro)Arg derivative. All these structures were used as starting points for subsequent full geometry optimizations.

The influence of the solvent on the conformational preferences of the compounds under study was quantified by performing Self-Consistent Reaction Field (SCRF) calculations on the optimized geometries. Under this formalism, the solute is treated at

the quantum mechanical level, while the solvent is represented as a dielectric continuum. In particular, we used the Polarizable Continuum Model (PCM) developed by Tomasi and co-workers to describe the bulk solvent.²⁰ PCM calculations were performed following the standard protocol and considering the dielectric constants of carbon tetrachloride ($\epsilon = 2.228$), chloroform ($\epsilon = 4.9$), and water ($\epsilon = 78.4$). The conformational free energies in solution (ΔG^{sol} , where *sol* refers to the solvent) were computed using the classical thermodynamics scheme, *i.e.* for each minimum, the free energy of solvation provided by the PCM model was added to the ΔG^{gp} value.

Force-field parameterization. The stretching, bending, torsion and van der Waals interactions of *trans*-(β Pro)Arg and *cis*-(β Pro)Arg were described classically by extrapolating the force-field parameters contained in the AMBER libraries²¹ for proline and arginine. For selected minimum energy conformations, electrostatic atomic centered charges were calculated by fitting the HF/6-31G(d) quantum mechanical and the Coulombic Molecular Electrostatic Potentials (MEPs) to a large set of points placed outside the nuclear region. The electrostatic parameters derived at this level of theory are fully compatible with the current parameters of the AMBER force-field.²¹ Electrostatic force-field parameters for the two (β Pro)Arg isomers were obtained by applying to such atomic charges a strategy based on a Boltzmann distribution of multiple conformations, which was originally proposed by different authors²² and has been shown to be specially suitable for non-proteinogenic residues.^{7,8c,8d,22b,23} Moreover, this strategy provides conformationally independent electrostatic parameters.

Force-field calculations. MD simulations in water solution were performed using the NAMD program.²¹ The Ac-*t*-(β Pro)Arg-NHMe or Ac-*c*-(β Pro)Arg-NHMe molecules were placed in the center of a cubic simulation box ($a = 30.6 \text{ \AA}$) filled with 955 explicit water molecules, which were represented using the TIP3 model.²⁵ Negatively charged chloride atoms were added to reach electron neutrality. Before the production runs, the simulation box was equilibrated for each compound. Thus, 0.5 ns of NVT-MD at 500K were used to homogeneously distribute the solvent and ions in the box. Next, 0.5 ns of NVT-MD at 298K (thermal equilibration) and 0.5 ns of NPT-MD at 298K (density relaxation) were carried out. The last snapshot of the NPT-MD was used as the starting point for production NVT-MD runs at standard conditions.

The energy was calculated using the AMBER potential.²¹ Atom pair distance cut-offs were applied at 12.0 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 slowly converged to 0 by applying a smoothing factor from a distance of 10.0 Å. Both temperature and pressure were controlled using the weak coupling method²⁶ applying a time constant for heat bath coupling and a pressure relaxation time of 1 ps. Bond lengths were constrained using the *SHAKE* algorithm²⁷ with a numerical integration step of 2 fs.

4.2.3. Results and discussion

A total of 21 minimum energy conformations were found and characterized for Ac-*t*-(^βPro)Arg-NHMe in the gas phase. The conformational parameters of those with relative energies (ΔE^{gp}) below 5.0 kcal/mol are listed in Table 4.2.1. (the complete list is provided as Supporting Information). In the global minimum ($\gamma_{\text{L}}[\text{u}]s^{-}g^{+}t$, Figure 4.2.3.a), the terminal acetyl CO and methylamide NH groups are linked by a hydrogen bond [$d_{\text{H}\cdots\text{O}} = 1.793$ Å, $\angle \text{N-H}\cdots\text{O} = 151.3^{\circ}$] closing a seven-membered cycle (γ -turn or C₇ conformation), and the pyrrolidine ring adopts an *up* puckering. The orientation of the exocyclic guanidilated side chain, which is defined by the *skew*⁻, *gauche*⁺ and *trans* arrangement of ξ^1 , ξ^2 and ξ^3 , respectively, enables the formation of a strong hydrogen bond between the carbonyl oxygen of the *trans*-(^βPro)Arg residue and the guanidinium NH site [$d_{\text{H}\cdots\text{O}} = 1.628$ Å, $\angle \text{N-H}\cdots\text{O} = 176.2^{\circ}$]. The second ($\gamma_{\text{L}}[\text{u}]g^{-}g^{-}s^{+}$, Figure 4.2.3.b) and third ($\gamma_{\text{L}}[\text{u}]g^{-}g^{-}s^{-}$, Figure 4.2.3.c) minima also exhibit the seven-membered hydrogen-bonded ring typical of a C₇ conformation and an *up*-puckered pyrrolidine moiety, while they differ from the global minimum in the orientation of the guanidilated side chain. In the $\gamma_{\text{L}}[\text{u}]g^{-}g^{-}s^{+}$ conformer, the side chain \cdots backbone interaction involves an NH₂ group in the guanidinium substituent instead of the NH site. The less favorable arrangement of the exocyclic side chain in these two conformers produces a destabilization of 1.1–1.5 kcal/mol with respect to the global minimum. A similar but more pronounced effect is observed for the last minimum listed in Table 4.2.1. ($\gamma_{\text{L}}[\text{u}]s^{-}tg^{-}$, Figure 4.2.3.g). This conformer presents identical shapes for both the peptide backbone and the pyrrolidine moiety to those described above for the first, second and third minima, but a much higher energy ($\Delta E^{\text{gp}} = 3.7$ kcal/mol). This destabilization should be attributed to the unfavorable steric interactions produced within the methylene groups in the side chain to allow the

formation of a hydrogen bond between the *trans*-(β Pro)Arg CO and the guanidinium NH₂.

The most stable structure in Table 4.2.1. exhibiting a backbone conformation other than a γ -turn is $\alpha_L[u]s^-g^+t$, which is unfavored by 2.1 kcal/mol with respect to the global minimum. This is noteworthy, since the most stable α_L conformation with *trans* amide bonds characterized for Ac-Pro-NHMe exhibits a ΔE^{sp} value of 4.9 kcal/mol.¹⁹ The $\alpha_L[u]s^-g^+t$ minimum of Ac-*t*-(β Pro)Arg-NHMe (Figure 4.2.3.d) presents no hydrogen-bonding interaction within the backbone amide groups, but is stabilized by a strong backbone···side chain hydrogen bond. The two additional α_L conformers in Table 4.2.1., $\alpha_L[u]g^+g^-t$ (Figure 4.2.3.e) and $\alpha_L[u]g^-tg^-$ (Figure 4.2.3.f), exhibit similar arrangements for the peptide backbone and the pyrrolidine ring, while differing in the orientation of the guanidylated substituent and the topology of the associated backbone···side chain interaction. Local repulsions within the aliphatic segment in this exocyclic side chain produce a destabilization of 0.9 and 1.4 kcal/mol, respectively, relative to the most stable α_L conformer.

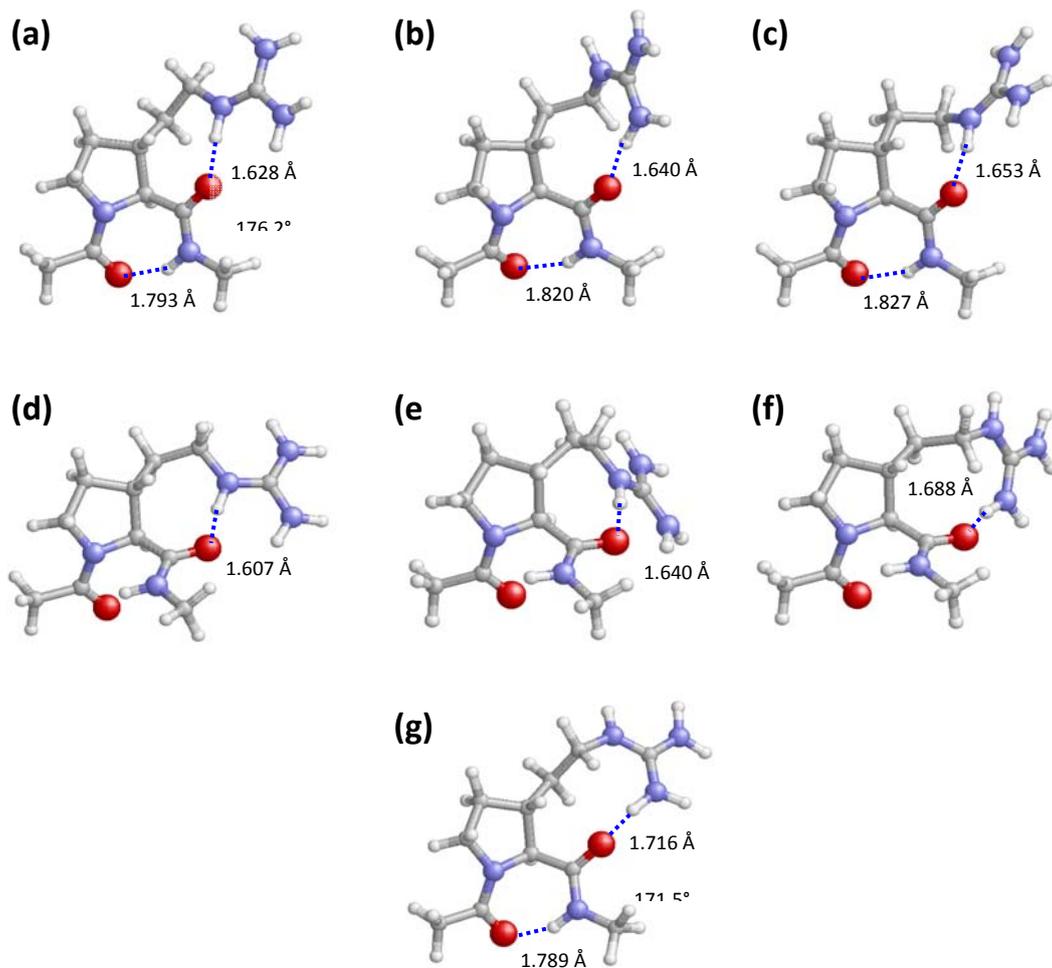


Figure 4.2.3.. Selected minimum energy conformations of Ac-*t*-(β Pro)Arg-NHMe obtained from B3LYP/6-31+G(d,p) calculations ($\Delta E^{\text{sp}} < 5.0$ kcal/mol, see Table 4.2.1.): (a) $\gamma_L[u]s^-g^+t$; (b) $\gamma_L[u]g^-g^-s^+$; (c) $\gamma_L[u]g^-g^-s^-$; (d) $\alpha_L[u]s^-g^+t$; (e) $\alpha_L[u]g^+g^-t$; (f) $\alpha_L[u]g^-tg^-$; and (g) $\gamma_L[u]s^-tg^-$. Intramolecular hydrogen bonds are indicated by dashed lines (H...O distances and N-H...O angles are given).

Table 4.2.1. Dihedral angles (see Figure 4.2.2.; in degrees), pseudorotational parameters of the pyrrolidine ring (A, P; in degrees), and relative energy (ΔE^{gp} ; in kcal/mol) of the minimum energy conformations with $\Delta E^{\text{gp}} < 5.0$ kcal/mol characterized for Ac-*t*-(β Pro)Arg-NHMe at the B3LYP/6-31+G(d,p) level.

<i>Conformer</i>	ω_0	φ	ψ	ω	A, P ^a	ξ^1	ξ^2	ξ^3	ΔE^{gp}
$\gamma_{\text{L}}[\text{u}]\text{s}^- \text{g}^+ \text{t}$	-169.6	-77.8	54.8	178.2	40.1, 92.5 ^b	-127.0	61.1	162.8	0.0 ^c
$\gamma_{\text{L}}[\text{u}]\text{g}^- \text{g}^- \text{s}^+$	-170.3	-79.9	59.5	179.1	39.0, 98.3 ^d	-84.2	-77.5	117.9	1.1
$\gamma_{\text{L}}[\text{u}]\text{g}^- \text{g}^- \text{s}^-$	-170.3	-80.5	56.1	179.3	38.0, 103.3 ^e	-59.8	-49.6	-115.6	1.5
$\alpha_{\text{L}}[\text{u}]\text{s}^- \text{g}^+ \text{t}$	-169.7	-89.7	-3.5	176.5	36.0, 112.0 ^f	-92.4	61.2	156.5	2.1
$\alpha_{\text{L}}[\text{u}]\text{g}^+ \text{g}^- \text{t}$	-170.2	-73.1	-20.2	179.0	40.5, 76.2 ^g	39.5	-77.7	-173.7	3.0
$\alpha_{\text{L}}[\text{u}]\text{g}^- \text{tg}^-$	-169.9	-76.5	-16.0	177.2	38.0, 85.5 ^h	-88.4	155.2	-86.1	3.5
$\gamma_{\text{L}}[\text{u}]\text{s}^- \text{tg}^-$	-169.4	-73.5	44.0	177.3	40.8, 85.5 ⁱ	-126.2	157.4	-87.1	3.7

^a See ref. 16 for definition. ^b $\chi^0 = -1.7^\circ$, $\chi^1 = -22.3^\circ$, $\chi^2 = 37.5^\circ$, $\chi^3 = -38.2^\circ$, $\chi^4 = 25.2^\circ$. ^c $E = -856.5567162$ a.u. ^d $\chi^0 = -5.7^\circ$, $\chi^1 = -18.4^\circ$, $\chi^2 = 34.9^\circ$, $\chi^3 = -38.0^\circ$, $\chi^4 = 27.5^\circ$. ^e $\chi^0 = -8.7^\circ$, $\chi^1 = -15.1^\circ$, $\chi^2 = 32.5^\circ$, $\chi^3 = -37.3^\circ$, $\chi^4 = 29.1^\circ$. ^f $\chi^0 = -13.5^\circ$, $\chi^1 = -9.2^\circ$, $\chi^2 = 27.4^\circ$, $\chi^3 = -35.4^\circ$, $\chi^4 = 30.7^\circ$. ^g $\chi^0 = 9.7^\circ$, $\chi^1 = -30.7^\circ$, $\chi^2 = 40.2^\circ$, $\chi^3 = -34.5^\circ$, $\chi^4 = 15.5^\circ$. ^h $\chi^0 = 3.0^\circ$, $\chi^1 = -24.6^\circ$, $\chi^2 = 36.8^\circ$, $\chi^3 = -35.1^\circ$, $\chi^4 = 20.0^\circ$. ⁱ $\chi^0 = 3.2^\circ$, $\chi^1 = -26.6^\circ$, $\chi^2 = 39.6^\circ$, $\chi^3 = -37.4^\circ$, $\chi^4 = 21.5^\circ$.

Comparison with the results previously reported¹² for Ac-*c*-(γ Pro)Arg-NHMe provides evidence for the higher flexibility of the β -substituted derivative Ac-*t*-(β Pro)Arg-NHMe studied in the present work. This is due to the presence of an additional exocyclic methylene unit in the latter case (Figure 4.2.1.), which broadens the conformational space that may be explored by the guanidilated side chain. Yet, the number of energetically accessible conformers is small in both cases, as expected from the restrictions imposed by the proline skeleton. The two compounds share the main structural features of the global minimum, which belongs to the $\gamma_L[u]$ category and exhibits identical patterns for both the backbone···backbone and side chain···backbone hydrogen-bonding interactions. However, a highly stable $\gamma_L[d]$ conformer was located¹² for Ac-*c*-(γ Pro)Arg-NHMe at only 0.4 kcal/mol, whereas no $\gamma_L[d]$ structure appears in Table 4.2.1. for Ac-*t*-(β Pro)Arg-NHMe. Indeed, the only $\gamma_L[d]$ minimum located for the latter compound presents $\Delta E^{\text{sp}} = 19.0$ kcal/mol (see Supporting Information). Another important difference is the presence of α_L conformers in Table 4.2.1., whereas no Ac-*c*-(γ Pro)Arg-NHMe minima¹² presented this backbone conformation.

Table 4.2.2. gives the conformational parameters of the five minima characterized for Ac-*c*-(β Pro)Arg-NHMe with ΔE^{sp} values below 5.0 kcal/mol (see the Supporting Information for a complete list of minima). Interestingly, the backbone shape preferred by this compound corresponds to the α -helix, whereas the γ -turn is disfavored by at least 2.0 kcal/mol. This is in sharp contrast with that described above for the *trans*-(β Pro)Arg derivative, as well as with the behavior observed before for the analogous γ -substituted compound¹² [Ac-*c*-(γ Pro)Arg-NHMe] and for proline itself¹⁹ (Ac-Pro-NHMe). Indeed, the most stable α -helical minimum characterized for Ac-Pro-NHMe with *trans* amide bonds lies 4.9 kcal/mol above the preferred γ -turn conformer,¹⁹ and no minimum energy structure was characterized in the α -helix region for Ac-*c*-(γ Pro)Arg-NHMe.¹² This comparative analysis provides evidence for the enormous impact that the incorporation of a functionalized side chain able to establish hydrogen-bonding interactions with the main-chain amide groups may have on the conformational preferences of the peptide backbone. It also illustrates the fact that the conformational profile of arginine analogues bearing a proline skeleton depends dramatically on the specific position of the guanidilated side chain, that is,

the pyrrolidine carbon bearing it and its relative orientation with respect to the carboxyl terminus.

Table 4.2.2.. Dihedral angles (see Figure 4.2.2.; in degrees), pseudorotational parameters of the pyrrolidine ring (A, P; in degrees), and relative energy (ΔE^{sp} ; in kcal/mol) of the minimum energy conformations with $\Delta E^{\text{sp}} < 5.0$ kcal/mol characterized for Ac-*c*-($^{\beta}$ Pro)Arg-NHMe at the B3LYP/6-31+G(d,p) level.

<i>Conformer</i>	ω_0	φ	ψ	ω	A, P ^a	ξ^1	ξ^2	ξ^3	ΔE^{sp}
$\alpha_{\text{L}}[\text{d}]\text{g}^+\text{g}^-\text{t}$	-169.7	-86.3	-11.6	175.3	39.6, -108.3 ^b	75.3	-71.4	178.9	0.0 ^c
$\alpha_{\text{L}}[\text{d}]\text{g}^+\text{ts}^+$	-169.2	-90.1	-4.3	174.5	40.1, -111.7 ^d	69.2	-172.0	95.2	1.3
$\gamma_{\text{L}}[\text{d}]\text{s}^-\text{g}^+\text{t}$	-171.1	-84.1	75.5	-176.7	39.8, -117.0 ^e	-116.3	70.4	164.2	2.0
$\gamma_{\text{L}}[\text{d}]\text{s}^-\text{ts}^-$	-170.3	-83.1	66.4	-178.5	40.0, -113.6 ^f	-104.8	168.1	-93.4	3.1
$\gamma_{\text{L}}[\text{u}]\text{s}^+\text{g}^-\text{t}$	-168.7	-66.8	31.8	174.5	43.7, 78.2 ^g	127.7	-65.4	178.3	3.9

^a See ref. 16 for definition. ^b $\chi^0 = -12.4^\circ$, $\chi^1 = 31.7^\circ$, $\chi^2 = -39.6^\circ$, $\chi^3 = 32.0^\circ$, $\chi^4 = -12.2^\circ$. ^c $E = -856,554209$ a.u. ^d $\chi^0 = -14.8^\circ$, $\chi^1 = 33.5^\circ$, $\chi^2 = -40.1^\circ$, $\chi^3 = 31.0^\circ$, $\chi^4 = -10.0^\circ$. ^e $\chi^0 = -18.1^\circ$, $\chi^1 = 35.0^\circ$, $\chi^2 = -39.5^\circ$, $\chi^3 = 28.3^\circ$, $\chi^4 = -6.3^\circ$. ^f $\chi^0 = -16.0^\circ$, $\chi^1 = 34.1^\circ$, $\chi^2 = -40.0^\circ$, $\chi^3 = 30.0^\circ$, $\chi^4 = -8.8^\circ$. ^g $\chi^0 = 8.9^\circ$, $\chi^1 = -32.1^\circ$, $\chi^2 = 43.4^\circ$, $\chi^3 = -38.2^\circ$, $\chi^4 = 18.1^\circ$.

As expected for an α -helical conformer, the lowest energy minimum of Ac-*c*-($^{\beta}$ Pro)Arg-NHMe ($\alpha_L[d]g^+g^-t$, Figure 4.2.4.a) exhibits no hydrogen-bonding interaction involving the backbone amide groups. However, a very strong hydrogen bond is established between the *cis*-($^{\beta}$ Pro)Arg CO and the guanidinium NH sites [$d_{H\cdots O} = 1.595 \text{ \AA}$, $\angle N-H\cdots O = 173.4^\circ$]. The α_L backbone conformation and the *down* pyrrolidine puckering are also present in the second minimum ($\alpha_L[d]g^+ts^+$, Figure 4.2.4.b). However, the less favorable backbone \cdots side chain interaction in this case, which involves a guanidinium NH₂ moiety, produces a destabilization of 1.3 kcal/mol.

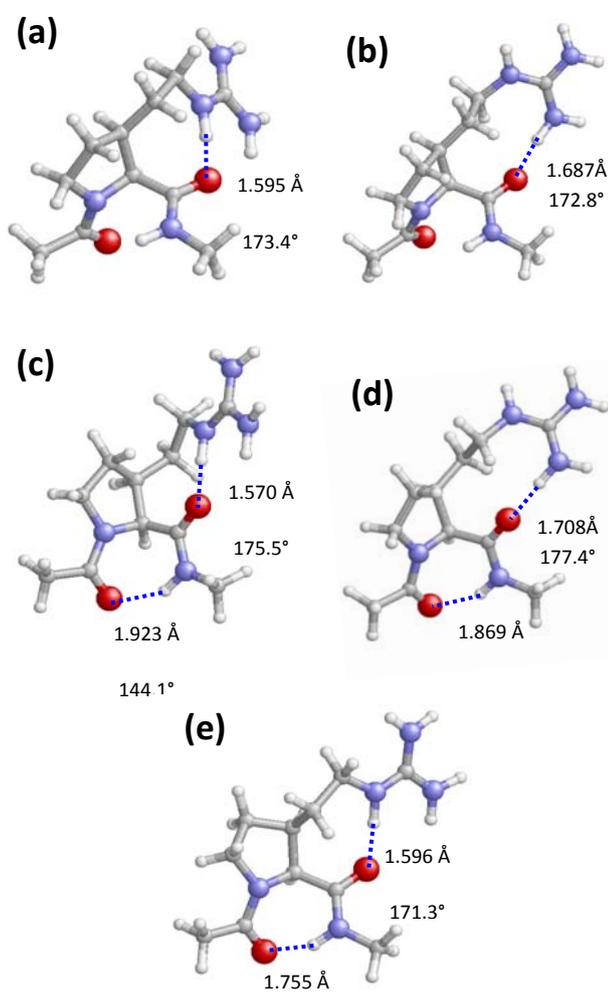


Figure 4.2.4. Selected minimum energy conformations of Ac-*c*-($^{\beta}$ Pro)Arg-NHMe obtained from B3LYP/6-31+G(d,p) calculations ($\Delta E^{zp} < 5.0$ kcal/mol, see Table 4.2.2.): (a) $\alpha_L[d]g^+g^-t$; (b) $\alpha_L[d]g^+ts^+$; (c) $\gamma_L[d]s^-g^+t$; (d) $\gamma_L[d]s^-ts^-$; and (e) $\gamma_L[u]s^+g^-t$. Intramolecular hydrogen bonds are indicated by dashed lines ($H\cdots O$ distances and $N-H\cdots O$ angles are given).

The remaining Ac-*c*-(β Pro)Arg-NHMe minima in Table 4.2.2. exhibit a γ -turn conformation stabilized by the corresponding hydrogen bond connecting the acetyl CO and methylamide NH groups. The two most stable conformers of this type, $\gamma_L[d]s^-g^+t$ (Figure 4.2.4.c) and $\gamma_L[d]s^-ts^-$ (Figure 4.2.4.d), retain the *down* pyrrolidine puckering observed in the helical minima and differ from each other in the guanidinium site (NH/NH₂) that is hydrogen-bonded to the carbonyl group of *cis*-(β Pro)Arg. Comparison of their relative energies (2.0 and 3.1 kcal/mol, respectively) suggests that the guanidinium NH provides a better geometry for hydrogen bonding to the main chain and therefore produces a higher stabilizing effect, as observed before for the helical conformers (Figures 4.2.4.a and 4.2.4.b). It should be noted that the most stable α_L and γ_L conformers characterized for the *trans*-(β Pro)Arg derivative (Table 4.2.1.) also present backbone...side chain interactions involving the NH guanidinium site. Therefore, this seems to be a general trend of (β Pro)Arg, independently of the *cis/trans* relative orientation of the guanidilated side chain.

The *trans*-(β Pro)Arg and *cis*-(β Pro)Arg derivatives investigated in this work not only differ in their respective preferences to accommodate γ -turn or α -helical backbone arrangements. These compounds also exhibit different conformational propensities in their five-membered ring. Thus, all minima in Table 4.2.1. exhibit an *up*-puckered pyrrolidine unit, whereas the most stable Ac-*c*-(β Pro)Arg-NHMe minima (Table 4.2.2.) present a *down* puckering. Moreover, the preference for a particular arrangement of the five-membered ring is much more pronounced in the former case as evidenced by the fact that the first *up*-puckered minimum of *cis*-(β Pro)Arg lies 3.9 kcal/mol above the global minimum (Table 4.2.2.) whereas no minima exhibiting a conformation other than *up* was identified below 7.0 kcal/mol for the *trans* isomer. Indeed, arrangements of the pyrrolidine ring rarely observed in proline and proline-like residues were found to be preferred over the *down* conformation for this compound. The contrast between the puckering tendencies of the five-membered ring in *trans*-(β Pro)Arg and *cis*-(β Pro)Arg becomes most evident when minima in the γ_L region are compared. Thus, for the *cis* isomer (Table 4.2.2.), the most stable $\gamma_L[d]$ and $\gamma_L[u]$ minima are separated by an energy gap of 1.9 kcal/mol only, whereas the corresponding energy difference for the *trans* derivative amounts to 19.0 kcal/mol. This distinct behavior should be ascribed to the different spatial relative disposition between the (β Pro)Arg carbonyl and guanidinium groups in the compounds under

study. In both cases, the pyrrolidine puckering providing optimal geometry for the establishment of hydrogen-bonding interactions between the two groups mentioned is preferred. For *cis*-(β Pro)Arg, both substituents lie on the same face of the five-membered cycle and are therefore close enough to interact for any puckering state. In contrast, the two interacting groups exhibit a *trans* relative orientation in *trans*-(β Pro)Arg, and it is the *up* (but not the *down*) arrangement of the pyrrolidine ring that brings them in close proximity, thus allowing for strong hydrogen bonding. For the *trans* isomer, the *up* puckering is particularly favorable in the case of γ_L conformers, because positive ψ values make the carbonyl oxygen of *trans*-(β Pro)Arg point in the opposite direction to where the guanidylated side chain is located. It should be noted that no hydrogen bond between the guanidinium group and the backbone exists in the only $\gamma_L[d]$ minimum characterized for this compound.

Table 4.2.3. Relative free energy^a in the gas phase (ΔG^{gp}) and in carbon tetrachloride, chloroform and aqueous solutions (ΔG^{CCl_4} , ΔG^{CHCl_3} , and $\Delta G^{\text{H}_2\text{O}}$, respectively) at 298K for selected^b minimum energy conformations of Ac-*t*-(β Pro)Arg-NHMe and Ac-*c*-(β Pro)Arg-NHMe at the B3LYP/6-31+G(d,p) level.

Conformer	ΔG^{gp}	ΔG^{CCl_4}	ΔG^{CHCl_3}	$\Delta G^{\text{H}_2\text{O}}$
Ac- <i>t</i> -(β Pro)Arg-NHMe				
$\gamma_{\text{L}}[\text{u}]\text{s}^{-}\text{g}^{+}\text{t}$	0.9	1.2	2.5	6.5
$\gamma_{\text{L}}[\text{u}]\text{g}^{-}\text{g}^{-}\text{s}^{+}$	2.6	0.6	0.3	2.6
$\gamma_{\text{L}}[\text{u}]\text{g}^{-}\text{g}^{-}\text{s}^{-}$	2.0	1.5	1.9	4.9
$\alpha_{\text{L}}[\text{u}]\text{s}^{-}\text{g}^{+}\text{t}$	0.0 ^c	0.0	0.5	2.7
$\alpha_{\text{L}}[\text{u}]\text{g}^{+}\text{g}^{-}\text{t}$	3.1	3.2	3.6	4.9
$\alpha_{\text{L}}[\text{u}]\text{g}^{-}\text{tg}^{-}$	4.1	1.2	0.0	0.0
$\gamma_{\text{L}}[\text{u}]\text{s}^{-}\text{tg}^{-}$	5.1	2.4	1.9	4.2
Ac- <i>c</i> -(β Pro)Arg-NHMe				
$\alpha_{\text{L}}[\text{d}]\text{g}^{+}\text{g}^{-}\text{t}$	0.0 ^d	0.0	1.1	2.1
$\alpha_{\text{L}}[\text{d}]\text{g}^{+}\text{ts}^{+}$	3.0	0.2	0.0	0.0
$\gamma_{\text{L}}[\text{d}]\text{s}^{-}\text{g}^{+}\text{t}$	3.9	4.4	6.5	9.4
$\gamma_{\text{L}}[\text{d}]\text{s}^{-}\text{ts}^{-}$	5.9	3.6	4.1	5.7
$\gamma_{\text{L}}[\text{u}]\text{s}^{+}\text{g}^{-}\text{t}$	5.8	6.0	7.9	10.2

^a In kcal/mol. ^b Those given in Tables 4.2.1 and 4.2.2 ($\Delta E^{\text{gp}} < 5.0$ kcal/mol). ^c $G = -856.257478$ a.u. ^d $G = -856.256685$ a.u.

Table 4.2.3. displays the free energies in the gas phase (ΔG^{gp}) at 298K for the Ac-*t*-(β Pro)Arg-NHMe and Ac-*c*-(β Pro)Arg-NHMe minima described above. Consideration of the ZPVE, thermal, and entropic corrections to transform ΔE^{gp} into ΔG^{gp} affects substantially the relative energy order of the minimum energy conformations characterized for the *trans*-(β Pro)Arg derivative. Specifically, $\alpha_{\text{L}}[\text{u}]\text{s}^{-}\text{g}^{+}\text{t}$ becomes the most stable conformer, with $\gamma_{\text{L}}[\text{u}]\text{s}^{-}\text{g}^{+}\text{t}$ being destabilized by 0.9 kcal/mol. Regarding the *cis* isomer, the relative stability of the $\alpha_{\text{L}}[\text{d}]\text{g}^{+}\text{g}^{-}\text{t}$ minimum is enhanced upon addition of these statistical contributions. The ΔG^{gp} values in Table 4.2.3. therefore indicate that α_{L} is the preferred backbone arrangement for both

compounds in the gas phase. Assuming a Boltzmann distribution, the population of α_L conformers at room temperature is about 80% and 100% for the *trans* and *cis* compound, respectively. In contrast, minima of the γ_L type showed the lowest ΔG^{sp} value for both Ac-*c*-(γ Pro)Arg-NHMe,¹² and Ac-Pro-NHMe,¹⁹ thus indicating a substantially higher tendency to adopt conformations in the α_L region for the arginine surrogates investigated in the present work, especially, the *cis* isomer.

The effect of solvation was next evaluated by performing single point calculations on the optimized structures through the PCM method. The presence of chlorinated solvents results in the stabilization of several Ac-*t*-(β Pro)Arg-NHMe minima (Table 4.2.3.), in particular, $\gamma_L[\text{u}]\text{g}^- \text{g}^- \text{s}^+$ and $\alpha_L[\text{u}]\text{g}^- \text{tg}^-$. The relative stability of the latter notably increases with the polarity of the solvent. Indeed, it becomes the preferred structure in chloroform, even if two other minima exhibit relative free energies within a 0.5 kcal/mol interval, and is the only accessible conformation at room temperature in aqueous solution. Regarding Ac-*c*-(β Pro)Arg-NHMe, only α_L structures are predicted to be populated either in the gas phase or in the different solvents considered (Table 4.2.3.). For this compound, solvation seems to affect mainly the arrangement of the exocyclic substituent, with the $\text{g}^+ \text{ts}^+$ disposition being favored with increasing polarity. Accordingly, $\alpha_L[\text{d}]\text{g}^+ \text{ts}^+$ becomes the preferred conformation in chloroform and, to a larger extent, in aqueous solution.

Comparison of the solvation effects described above with those observed before¹² for the γ -substituted compound Ac-*c*-(γ Pro)Arg-NHMe provides further evidence for the higher stability of the α_L conformation in the arginine surrogates investigated in the present work. Indeed, minima of the γ_L type showed the lowest ΔG value for the *cis*-(γ Pro)Arg derivative not only in the gas phase but also in carbon tetrachloride and chloroform solutions.¹² In water, conformations devoid of intramolecular hydrogen bonds between the backbone amide groups are usually favored for small peptides like the ones considered and, accordingly, an ε_L conformer became the most populated structure for Ac-*c*-(γ Pro)Arg-NHMe in this solvent.¹² Interestingly, Ac-Pro-NHMe was predicted¹⁹ to prefer the α -helical structure in water. The latter point shows that the conformational preferences of proline in aqueous solution are retained to a larger extent when the arginine side-chain is attached to the β position of the pyrrolidine moiety.

As stated in the Introduction, *trans*-(β Pro)Arg and *cis*-(β Pro)Arg are conceived as arginine substitutes in biologically active peptides. The conformational consequences arising from the incorporation of these arginine surrogates into such peptides may be performed by methods like molecular dynamics (MD) simulations. For this purpose, previous parameterization of the non-proteinogenic residues is necessary. A specific set of force-field parameters was developed for *trans*-(β Pro)Arg and *cis*-(β Pro)Arg to describe the inter- and intramolecular interactions within the classical formalism. Our previous work showed that there is no special electronic effect that might condition the conformational preferences of proline upon addition of the arginine side chain¹² and, therefore, the stretching, bending, torsional, and van der Waals parameters for *trans*-(β Pro)Arg and *cis*-(β Pro)Arg were transferred directly from the AMBER force-field.²¹ Accordingly, electrostatic charges were the only force-field parameters specifically developed for these non-proteinogenic residues.

Atomic charges were calculated by fitting the HF/6-31G(d) quantum mechanical and the Coulombic MEPs to a large set of points placed outside the nuclear region. The electrostatic parameters were obtained by weighting the charges calculated for the low-energy conformers of each compound according to a Boltzmann distribution.^{22,23} The latter was estimated with the ΔG^{BP} values listed in Table 4.2.3.. The $\alpha_{\text{L}}[\text{u}]s^{-}g^{+}t$, $\gamma_{\text{L}}[\text{u}]s^{-}g^{+}t$ and $\gamma_{\text{L}}[\text{u}]g^{-}g^{-}s^{-}$ structures were considered for the *trans* isomer, whereas $\alpha_{\text{L}}[\text{d}]g^{+}g^{-}t$ was the only conformer used for the *cis* derivative since all the local minima are destabilized by at least 3.0 kcal/mol. The electrostatic parameters obtained for *trans*-(β Pro)Arg and *cis*-(β Pro)Arg are given in Figure 4.2.5..

To check the validity of classical MD simulations in describing the conformational properties of the arginine analogues under study, MD with explicit solvent molecules were performed on Ac-*t*-(β Pro)Arg-NHMe and Ac-*c*-(β Pro)Arg-NHMe in aqueous solution at 298K. For each compound, the lowest-energy conformation was used as the starting point of a 10 ns trajectory. Figure 4.2.6. represents the accumulated Ramachandran plot obtained for each derivative. In both cases, the most populated backbone structure corresponds to α_{L} , which is visited much more frequently than the γ_{L} region during the trajectory. This finding is in excellent agreement with the results displayed in Table 4.2.3., which indicate that α_{L} is the most favored conformation in aqueous solution for both compounds.

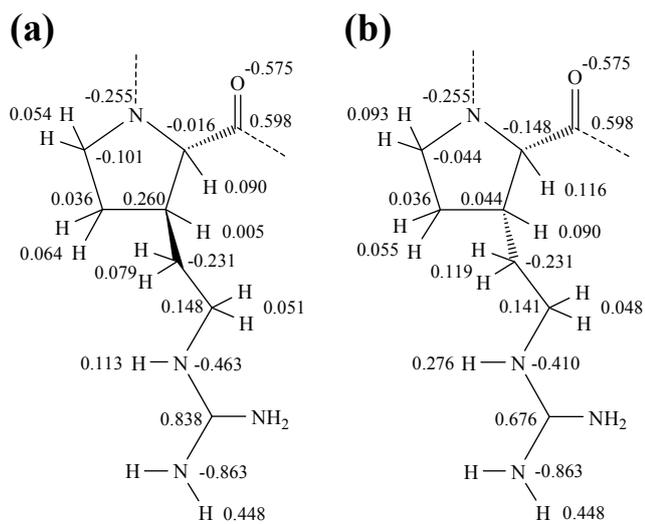


Figure 4.2.5. Electrostatic parameters determined for the (a) *trans*-(β Pro)Arg and (b) *cis*-(β Pro)Arg residues.

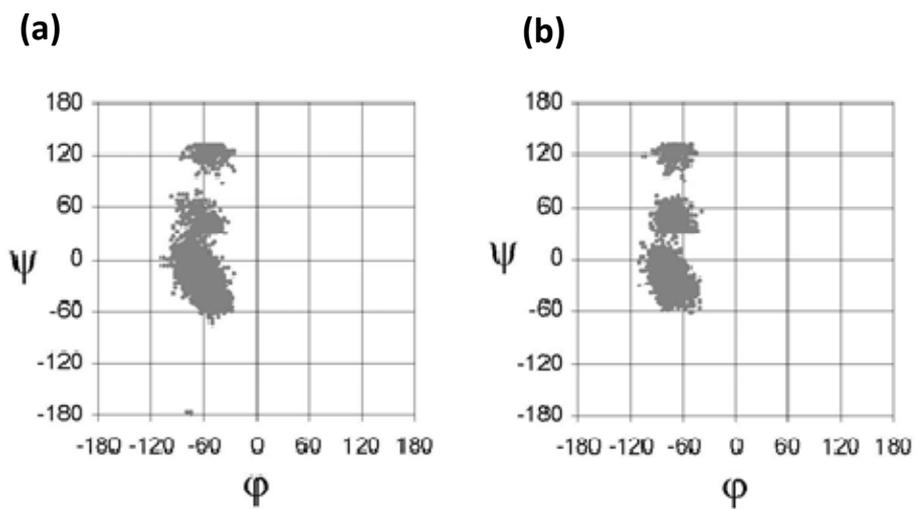


Figure 4.2.6. Accumulated Ramachandran plots for (a) *Ac-t*-(β Pro)Arg-NHMe and (b) *Ac-c*-(β Pro)Arg-NHMe derived from MD simulations in aqueous solution.

4.2.4. Conclusions

Two isomers of an arginine surrogate have been built by attaching the arginine side chain to the proline β -carbon in either a *trans* or a *cis* disposition relative to the carboxylic acid. The resulting amino acids, respectively denoted *trans*-(β Pro)Arg and *cis*-(β Pro)Arg, combine the conformational restrictions associated with the cyclic nature of proline with the side-chain functionality of arginine. Quantum mechanics calculations on the *N*-acetyl-*N'*-methylethylamide derivatives of these arginine surrogates show that the conformational space available is highly restricted, as expected from their proline-like character. Their conformational preferences are essentially determined by their cyclic structure and the capacity of the guanidylated side chain to establish hydrogen-bonding interactions with the peptide backbone. The latter factor is especially significant for the α_L conformation, which is stabilized with respect to natural proline¹⁶ and is predicted to be the most populated structure for both (β Pro)Arg isomers not only in the gas phase but also in aqueous solution. MD simulations show that the restricted flexibility and the preference for α -helical conformations are kept when thermal agitation is included.

The two non-coded amino acids studied in the present work are suitable candidates to replace arginine in bioactive peptides when the natural residue is found in the α_L region, and more specifically, occupying the *i+1* position of a β -turn of type I. In particular, *cis*-(β Pro)Arg may be an excellent replacement for arginine in CREKA, and more appropriate than the arginine surrogate considered in a previous work.¹² Thus, *cis*-(β Pro)Arg is expected not only to increase resistance to proteases but also to greatly stabilize the type I β -turn found for CREKA. For other biologically relevant peptides, either the *cis* or the *trans* isomer of (β Pro)Arg may be adequate to replace arginine depending on the orientation attained by the guanidylated side chain in the bioactive conformation.

4.2.5. References

- [1] (a) Venkatraman, J.; Shankaramma, S. C.; Balaram, P. Design of Folded Peptides. *Chem. Rev.* 2001, *101*, 3131–3152. (b) Toniolo, C.; Crisma, M.; Formaggio, F.; Peggion, C. Control of Peptide Conformation by the Thorpe-Ingold Effect (C^α-Tetrasubstitution). *Biopolymers (Pept. Sci.)* 2001, *60*, 396–419.
- [2] (b) Nestor, J. J. The Medicinal Chemistry of Peptides. *Curr. Med. Chem.* 2009, *16*, 4399–4418. (b) Horne, W. S.; Gellman, S. H. Foldamers with Heterogeneous Backbones. *Acc. Chem. Res.* 2008, *41*, 1399–1408. (c) Chatterjee, J.; Gilon, C.; Hoffman, A.; Kessler, H. N-Methylation of Peptides: A New Perspective in Medicinal Chemistry. *Acc. Chem. Res.* 2008, *41*, 1331–1342. (d) Lelais, G.; Seebach, D. α^2 -Amino Acids – Syntheses, Occurrence in Natural Products, and Components of α -Peptides. *Biopolymers (Pept. Sci.)* 2004, *76*, 206–243. (e) Adessi, C.; Soto, C. Converting a Peptide into a Drug: Strategies to Improve Stability and Bioavailability. *Curr. Med. Chem.* 2002, *9*, 963–978.
- [3] (a) Peptide and Protein Design for Biopharmaceutical Applications; Jensen, K. J., Ed.; John Wiley & Sons: Chichester, 2009. (b) Pseudo-Peptides in Drug Discovery; Nielsen, P. E., Ed.; Wiley-VCH: Weinheim, 2004.
- [4] Simberg, D.; Duza, T.; Park, J. H.; Essler, 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. USA* 2007, *104*, 932–936.
- [5] Zanuy, D.; Flores-Ortega, A.; Casanovas, J.; Curcó, D.; Nussinov, R.; Alemán, C. The Energy Landscape of a Selective Tumor-Homing Pentapeptide. *J. Phys. Chem. B* 2008, *112*, 8692–8700.
- [6] Agemy, L.; Sugahara, K.N.; Kotamraju, V.R.; Gujrati, K.; Girard, O.M.; Kono, Y.; Mattrey, R. F.; Park, J.H.; Sailor, M.J.; Jimenez, A.I.; Cativiela, C.; Zanuy, D.; Sayago, F.J.; Aleman, C.; Nussinov, R.; Ruoslahti E. Nanoparticle-induced vascular blockade in human prostate cancer. *Blood*, 2010, DOI 10.1182/blood-2010-03-274258.
- [7] Revilla-Lopez, G.; Torras, J.; Jimenez, A.I.; Cativiela, C.; Nussinov, R.; Aleman, C. Side-Chain to Backbone Interactions Dictate the Conformational Preferences of a Cyclopentane Arginine Analogue. *J. Org. Chem.*, 2009, *74* (6), 2403-2412.
- [8] (a) Alemán, C. Conformational properties of alpha-amino acids disubstituted at the alpha-carbon. *J. Phys. Chem. B* 1997, *101*, 5046-5050. (b) Alemán, C.; Jiménez,

A.I.; Cativiela, C.; Perez, J.J.; Casanovas, J. Influence of the phenyl side chain on the conformation of cyclopropane analogues of phenylalanine *J. Phys. Chem. B* 2002, *106*, 11849-11858. (c) Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. Identification of the intrinsic conformational properties of 1-aminocyclobutane-1-carboxylic acid *Chem. Phys. Lett.* 2006, *429*, 558-562 (d) Alemán, C.; Zanuy, D.; Casanovas, J. Cativiela, C.; Nussinov, R. Backbone Conformational Preferences and Pseudorotational Ring Puckering of 1-Aminocyclopentane-1-carboxylic Acid *J. Phys. Chem. B* 2006, *110*, 21264-21271. (e) Rodriguez-Roperero, F.; Zanuy, D.; Casanovas, J.; Nussinov, R.; Alemán, C. Application of 1-Aminocyclohexane Carboxylic Acid to Protein Nanostructure Computer Design *J. Chem. Inf. Model.* 2008, *48*, 333-343.

[9] (a) Markert, Y.; Köditz, J.; Ulbrich-Hofmann, R.; Arnold, U. Proline versus Charge Concept for Protein Stabilization against Proteolytic Attack. *Protein Eng.* 2003, *16*, 1041-1046. (b) Walker, J. R.; Altman, R. K.; Warren, J. W.; Altman, E. Using Protein-Based Motifs to Stabilize Peptides. *J. Pept. Res.* 2003, *62*, 214-226. (c) Vanhoof, G.; Goossens, F.; De Meester, I.; Hendriks, D.; Scharpé, S. Proline Motifs in Peptides and Their Biological Processing. *Faseb J.* 1995, *9*, 736-744. (d) Frenken, L. G. J.; Egmond, M. R.; Batenburg, A. M.; Verrips, C. T. *Pseudomonas-Glumae* Lipase: Increased Proteolytic Stability by Protein Engineering. *Protein Eng.* 1993, *6*, 637-642.

[10] (a) Chakrabarti, P.; Pal, D. The Interrelationships of Side-Chain and Main-Chain Conformations in Proteins. *Prog. Biophys. Mol. Biol.* 2001, *76*, 1-102. (b) MacArthur, M. W.; Thornton, J. M. Influence of Proline Residues on Protein Conformation. *J. Mol. Biol.* 1991, *218*, 397-412. (c) Rose, G. D.; Gierasch, L. M.; Smith, J. A. Turns in Peptides and Proteins. *Adv. Protein Chem.* 1985, *37*, 1-109.

[11] (a) Che, Y.; Marshall, G. R. Impact of *Cis*-Proline Analogs on Peptide Conformation. *Biopolymers* 2006, *81*, 392-406. (b) Sahai, M. A.; Fejer, S. N.; Viskolcz, B.; Pai, E. F.; Csizmadia I. G. First-Principle Computational Study on the Full Conformational Space of L-Threonine Diamide, the Energetic Stability of *Cis* and *Trans* Isomers. *J. Phys. Chem. A* 2006, *110*, 11527-11536.

[12] Zanuy, D.; Flores-Ortega, A.; Jiménez, A. I.; Calaza, M. I.; Cativiela, C.; Nussinov, R.; Ruoslahti, E.; Alemán, C. *In Silico* Molecular Engineering for a Targeted Replacement in a Tumor-Homing Peptide. *J. Phys. Chem. B* 2009, *113*, 7879-7889.

[13] 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.

[14] Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.* 1993, *98*, 1372–1377.

[15] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B* 1988, *37*, 785–789.

[16] McLean, A. D.; Chandler, G. S. Contracted Gaussian Basis Sets for Molecular Calculations. I. Second Row Atoms, $Z=11-18$. *J. Chem. Phys.* 1980, *72*, 5639–5648.

[17] Perczel, A.; Angyán, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadia, I. G. Peptide Models. 1. Topology of Selected Peptide Conformational Potential Energy Surfaces (Glycine and Alanine Derivatives). *J. Am. Chem. Soc.* 1991, *113*, 6256–6265.

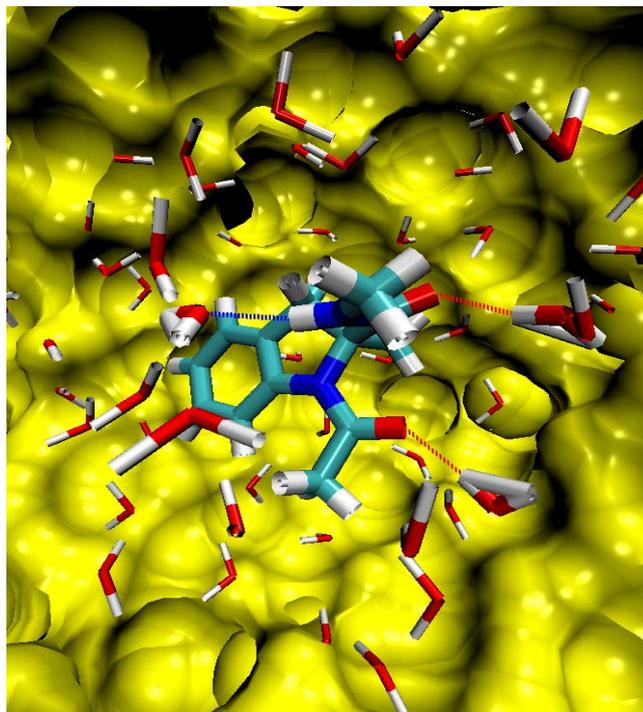
[18] (a) Milner-White, E. J.; Bell, L. H.; Maccallum, P. H. Pyrrolidine Ring Puckering in *cis* and *trans*-Proline Residues in Proteins and Polypeptides – Different Puckers are Favoured in Certain Situations. *J. Mol. Biol.* 1992, *228*, 725–734. (b) Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. Energy Parameters in Polypeptides. 10. Improved Geometrical Parameters and Nonbonded Interactions for Use in the ECEPP/3 Algorithm, with Application to Proline-Containing Peptides. *J. Phys. Chem.* 1992, *96*, 6472–6484.

- [19] Flores-Ortega, A.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C.; Casanovas, J. Conformational Preferences of α -Substituted Proline Analogues. *J. Org. Chem.* 2008, *73*, 3418–3427.
- [20] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* 2005, *105*, 2999–3093. (b) Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* 1994, *94*, 2027–2094. (c) Miertus, S.; Tomasi, J. Approximate Evaluations of the Electrostatic Free Energy and Internal Energy Changes in Solution Processes. *Chem. Phys.* 1982, *65*, 239–245. (d) Miertus, M.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of Ab Initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* 1981, *55*, 117–129.
- [21] 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. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* 1995, *117*, 5179–5197.
- [22] (a) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* 1995, *16*, 1357–1377. (b) Alemán, C.; Casanovas, J. Ab Initio SCF and Force-Field Calculations on Low-Energy Conformers of 2-Acetylamino-2,N-Dimethylpropanamide. *J. Chem. Soc. Perkin Trans 2*, 1994, 563–568. (c) Reynolds, C. A.; Essex, J. W.; Richards, W. G. Atomic Charges for Variable Molecular Conformations. *J. Am. Chem. Soc.* 1992, *114*, 9075–9079.
- [23] Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. Intrinsic Conformational Characteristics of α,α -Diphenylglycine. *J. Org. Chem.* 2007, *72*, 2174–2181.
- [24] Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* 1999, *151*, 283–312.
- [25] 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.

[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 the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* 1977, *23*, 327–341.

4.3. Conformational Preferences of Proline Analogues with a Fused Benzene Ring



4.3.1. Introduction

Conformationally constrained amino acids are of growing interest due to their powerful capability to stabilize structural motifs in peptides and proteins by selective replacements in native sequences. Introduction of rigidity into peptide chains through the incorporation of residues with well-defined conformational properties has proven useful in the design of bioactive peptides with improved pharmacological profile¹ and other biologically relevant systems.²

Proline (Pro) is the only proteinogenic amino acid that can be viewed as conformationally constrained. The uniqueness of Pro derives from the presence of the five-membered pyrrolidine ring that includes the α -carbon and the amino function. As a consequence, rotation about the N-C $^{\alpha}$ bond is prohibited and the ϕ dihedral angle is forced to adopt values near -60° . This is at the basis of the well-known tendency of Pro to act as a turn inductor in peptide chains.³ Another distinct feature of Pro associated to its cyclic nature is that the *cis* and *trans* states of the peptide bond

involving the pyrrolidine nitrogen are closer in energy than for any other proteinogenic amino acid. Accordingly, the peptide bond preceding Pro has a non-negligible probability of accommodating a *cis* arrangement.^{3,4}

The high significance of Pro in peptide conformation and biology has stimulated the search for new Pro analogues endowed with tailored properties. The incorporation of the functional groups present in the side chains of other proteinogenic amino acids is particularly appealing in this context. This allows the combination of the particular structural properties of Pro with the functionality of other residues. This is the case of indoline-2-carboxylic acid (Inc, Figure 4.3.1.), which results from the fusion of a benzene ring to the pyrrolidine bond linking the γ and δ carbons in Pro. In fact, Inc may be considered as being simultaneously a Pro and a phenylalanine (Phe) analogue. Such combination of structural and functional properties may be synergistic and allow optimal interaction with the complementary groups in the receptor binding pocket. Moreover, at variance with Phe, the aromatic side chain in Inc is anchored in a particular orientation with respect to the peptide backbone and this may be exploited to investigate the conformational requirements for optimal binding when aromatic groups are directly involved in the peptide-receptor recognition process.

The peculiar conformational and functional properties of Inc have attracted the interest of many researchers and, thus, this amino acid has been incorporated –most frequently, as a replacement for either Pro or Phe– in a variety of peptides⁵ exhibiting medically relevant properties. Moreover, the great deal of patents related to biologically active Inc-containing peptides⁶ provides unequivocal proof of the high potential that this amino acid offers in the design of pharmacologically useful compounds.

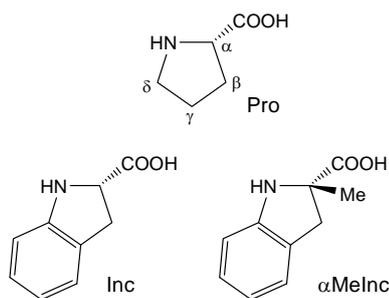


Figure 4.3.1. Structure of the proline analogues investigated in the present work, indoline-2-carboxylic acid (Inc) and (α -methyl)indoline-2-carboxylic acid (α MeInc).

However, no work has been devoted to investigate to which extent the additional benzene ring in Inc affects the conformational preferences exhibited by Pro. Knowledge of the intrinsic structural properties of Inc is yet essential to satisfactorily exploit this amino acid in the peptides field. The present paper describes the conformational profile of Inc established by quantum mechanical methods. The study has been extended to its α -methyl derivative (α MeInc, Figure 4.3.1.) because of the structural significance and usefulness of α -methyl amino acids in the design of peptides with well-defined conformational properties.⁷ Specifically, we have performed Density Functional Theory (DFT) calculations on the *N*-acetyl-*N'*-methylamide derivatives of the L enantiomers of Inc and α MeInc, hereafter denoted as Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe, respectively (Figure 4.3.2.). Their behavior has been compared to that exhibited by the analogous Pro and α MePro derivatives investigated in a previous work.⁸ Moreover, the influence of the environment has been examined using both implicit and explicit solvation models through the Self Consistent Reaction Field (SCRF) method and a hybrid quantum mechanics/molecular mechanics (QM/MM) approach, respectively.

A new interface NWChem⁹/PUPIL¹⁰/Amber¹¹ was built to carry out the multi-scale simulations in the latter case.

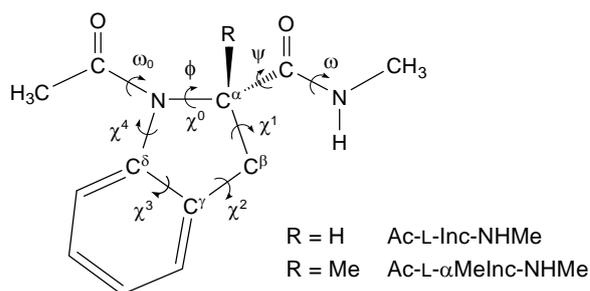


Figure 4.3.2. Dihedral angles used to identify the conformations of Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe. The ω_0 , ϕ , ψ , and ω angles are defined by using backbone atoms, while the endocyclic dihedral angles χ^i are given by the five-membered ring atoms. In particular, ϕ and χ^0 are defined by C(O)-N-C $^\alpha$ -C(O) and C $^\delta$ -N-C $^\alpha$ -C $^\beta$, respectively.

4.3.2. Methods

Quantum Mechanical Calculations. Density Functional Theory (DFT) calculations have been carried out using the Gaussian 03 computer program,¹² by combining the Becke's three parameter hybrid functional (B3)¹³ with the Lee, Yang and Parr (LYP)¹⁴ expression for the nonlocal correlation (B3LYP). Although this methodology is known to underestimate dispersive interactions and to overestimate bonding energies,¹⁵ it has been demonstrated to provide a very satisfactory description of the conformational properties of constrained amino acids, including Pro^{8,16} and some non-proteinogenic analogues of this residue.¹⁷ All DFT calculations in this work have been performed by combining the B3LYP method with the 6-31+G(d,p) basis set.¹⁸

The backbone ($\omega_0, \phi, \psi, \omega$) and side chain (χ^i) dihedral angles of Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe are defined in Figure 4.3.2. The ϕ value (rotation about N-C $^\alpha$) is fixed by the geometry of the five-membered ring, whereas the flexible dihedral angle ψ is expected to exhibit three minima: *gauche*⁺ (60°), *trans* (180°) and *gauche*⁻ (-60°). Therefore, only three minima can be anticipated for the potential energy surface $E = E(\phi, \psi)$ of the compounds investigated for a given arrangement of ω_0 and ω . Regarding the peptide bonds, the methylcarboxamide group (CONHMe) was kept in the *trans* state ($\omega \approx 180^\circ$), whereas both the *cis* and *trans* arrangements ($\omega_0 \approx 0^\circ$ or 180° , respectively) were considered for the amide moiety involving the pyrrolidine nitrogen (acetamido, MeCON) due to the well-known tendency of this peptide bond to accommodate the *cis* and *trans* dispositions in Pro and proline-like residues.^{3,4} In contrast, the *up/down* conformations typically adopted by the five-membered ring in Pro,^{3a,19} and corresponding to a deviation of the C $^\gamma$ atom from the plane formed by C $^\delta$, N, and C $^\alpha$, are not applicable to Inc or α MeInc because the fused benzene ring imposes coplanarity to the C $^\beta$, C $^\gamma$, C $^\delta$ and N atoms.

Accordingly, 3 (ψ minima) \times 2 (ω_0 *cis/trans*) = 6 structures were considered as starting points for complete geometry optimizations at the B3LYP/6-31+G(d,p) level for both Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe. 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 the thermal and entropic corrections. These statistical terms were used to compute the conformational Gibbs free energies in the gas phase (ΔG^{gp}) at 298K.

The minimum energy conformations thus characterized have been denoted using a two-label code that specifies the ω_0 arrangement as *cis* (c) or *trans* (t) followed by the type of backbone conformation. The latter is defined using the nomenclature introduced by Perczel *et al.*²⁰ that categorizes the potential energy surface $E = E(\varphi, \psi)$ of α -amino acids in nine different regions: γ_D , δ_D , α_D , ε_D , β_{DL} , ε_L , α_L , δ_L and γ_L . In the case of Pro or proline-like residues, only the γ_L (γ -turn or C₇), α_L (α -helix), and ε_L (polyproline II) conformations are accessible^{3,8} because φ is confined to values around -60° .

Even if the benzene ring in Inc and α MeInc significantly restricts the flexibility of the five-membered cycle, this cycle is not completely flat. Its puckering state has been described using the classical pseudorotational algorithm, which uses a very simple model based on only two parameters, as previously applied to Pro by Perczel *et al.*²¹ 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 χ^i (Figure 4.3.2.) as follows:

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

$$P = \begin{cases} \arccos \frac{\chi^0}{A}, & \text{if } A \sin P \geq 0 \\ -\arccos \frac{\chi^0}{A}, & \text{if } A \sin P < 0 \end{cases}$$

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

To obtain an estimation of the solvation effects on the relative stability of the minimum energy conformations, two different approaches were initially followed. First, the effect of different solvents covering a wide range of polarities was analyzed by performing single-point calculations on the structures optimized in the gas phase using a Self-Consistent Reaction Field (SCRF) model at the B3LYP/6-31+G(d,p) level. Next, to obtain more precise results in a high-polarity solvent such as water, the estimation of the solvation effects was carried out by allowing the previously gas-phase optimized structures to relax using a SCRF model. SCRF methods treat the solute at the quantum mechanical level, while the solvent is represented as a dielectric

continuum. In particular, the Polarizable Continuum Model (PCM) developed by Tomasi and co-workers was used to describe the bulk solvent.²² This method involves the generation of a solvent cavity from spheres centred at each atom in the molecule and, subsequently, the calculation of virtual point charges on the cavity surface representing the effect of 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. Therefore, the point charges may be included in the one-electron Hamiltonian, thus inducing polarization on the solute. At that moment, 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 ($\epsilon = 2.228$), chloroform ($\epsilon = 4.9$), and water ($\epsilon = 78.4$). The conformational free energies in solution ($\Delta G^{\text{sol}\#}$, where #sol# refers to the solvent) were estimated by using the classical thermodynamics scheme, that is, the free energies of solvation provided by the PCM model were added to the ΔG^{sp} values.

Multi-Scale Approach (Hybrid QM/MM Calculations). The advantage of using the SCRF model described in the previous section lies in the omission of explicit solvent molecules, which produces a computational speedup. Nevertheless, the results obtained with this method may be significantly influenced by an important assumption: both the solvent configurational sampling and the local anisotropies around the solute are implicit in the continuum model. The influence of this assumption was evaluated by applying a multi-scale approach based on an explicit solvation model to investigate the conformational preferences of Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe in aqueous solution. In this method, the atomic motions are handled by molecular dynamics (MD), with energies and forces being calculated by dividing the system into two different parts. The compound investigated (solute) is treated at the quantum mechanical (QM) level, while molecular mechanics (MM) are applied to the rest of the system (solvent) by using a classical potential energy function.

For each compound, two structures were selected for this study, namely the lowest energy minima exhibiting ω_0 arranged in either *cis* or *trans*. Such structures were solvated with 1350 and 1408 water molecules for Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe, respectively, thus defining a 12 Å buffer region around each compound. The solvent and solute molecules were described using the TIP3P²³ model

and the General Amber Force Field (GAFF),²⁴ respectively. The Amber 10¹¹ code was used for all classical calculations. The four resulting systems (two structures for each compound) were fully minimized at the MM level, first heated up to 298K using the Langevin thermostat during 25 ps at a constant volume (1 fs time steps), and then equilibrated using a NPT ensemble for 250 ps at 1 atm and 298K. Subsequently, the amino acid derivatives were changed to a QM description, and treated at the SCC-DFTB²⁵ level, while the water molecules remained within the MM framework. After that, the systems were allowed to relax for 100 ps using a constant pressure simulation with the parameters previously used for fully classical MD simulations. In all cases, atom pair distance cutoffs were applied at 10 Å to compute van der Waals interactions. The electrostatic interactions were computed by using the nontruncated electrostatic potential by means of Ewald Summations.

The hybrid QM/MM calculations were run using a new interface, which was specially programmed for this work and that allows the incorporation of the NWChem⁹ program as a new QMWorker into the PUPIL²⁶ package (Program for User Package Interfacing and Linking). The starting structures for NWChem-PUPIL-Amber calculations for the *cis* and *trans* conformations of each compound were extracted from the SCC-DFTB/MD simulation (last snapshot). In all the production trajectories, the QM region involving the amino acid derivative was described by combining the B3LYP functional with the 6-31+G(d) basis set. Simulations were carried out in the NVT ensemble at 298K using a Langevin thermostat with a friction coefficient of 10 ps for 6 ps (1 fs time step). Periodic boundary conditions were applied in the preparation of the NWChem input so as to wrap neighboring point charges around the quantum region. The QM region comprises a total of 268 and 287 basis functions for Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe, respectively.

4.3.2. Results and discussion

Ac-L-Inc-NHMe. Table 4.3.1. lists the most relevant structural parameters together with the relative energy (ΔE^{gp}) and free energy (ΔG^{gp}) of the three minimum energy conformations characterized for Ac-L-Inc-NHMe in the gas phase (Figure 4.3.3.). As can be seen, the three types of (ϕ, ψ) backbone structures accessible to Pro,⁸ namely γ_{L} , α_{L} , and ϵ_{L} , were also characterized as energy minima for Inc. However, as explained in the following, important differences are observed between Pro and Inc regarding the arrangement of both the five-membered cycle and the acetamido

moiety, that is, the peptide bond formed by the acetyl group and the pyrrolidine nitrogen (MeCON, given by ω_0 in Figure 4.3.2.). Such differences derive from the presence of a fused benzene moiety in the latter residue, which produces an important restriction of the conformational space. This is clearly evidenced when the three Ac-L-Inc-NHMe minima in Table 4.3.1. are compared to the seven minimum energy conformers located for the analogous Pro derivative (Ac-L-Pro-NHMe) at the same level of theory.⁸

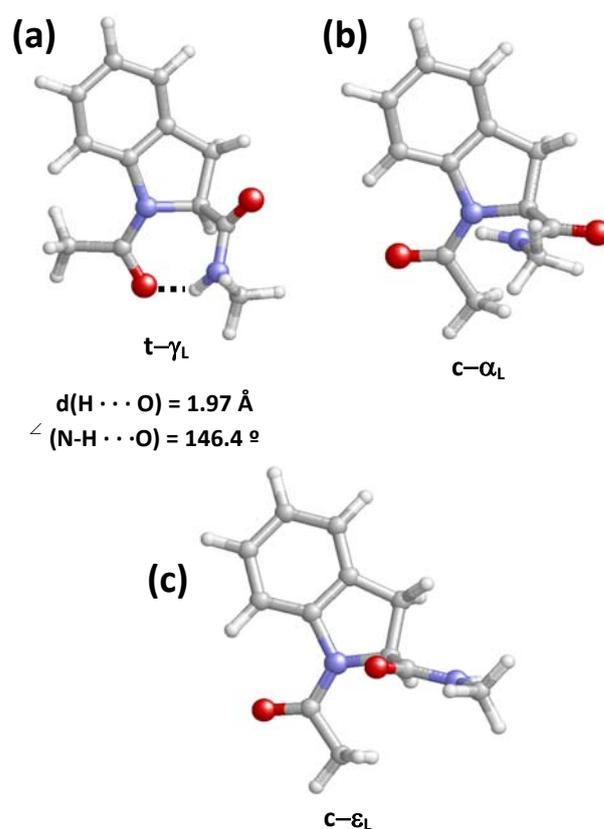


Figure 4.3.3. Representation of the minimum energy conformations characterized for Ac-L-Inc-NHMe at the B3LYP/6-31+G(d,p) level. Structural parameters and relative energies are provided in Table 4.3.1..

The lowest energy conformation characterized for Ac-L-Inc-NHMe, $t\text{-}\gamma_L$ (Figure 4.3.3.a), presents the acetyl and methylamide terminal groups connected by an intramolecular hydrogen bond [$d(\text{H}\cdots\text{O}) = 1.97 \text{ \AA}$, $\angle\text{N-H}\cdots\text{O} = 146.4^\circ$] that closes a seven-membered cycle. The existence of this hydrogen-bonding interaction typical of the γ -turn (also called C_7 conformation) requires a *trans* ω_0 angle, which is yet severely distorted. Thus, ω_0 deviates significantly (-166.6° , Table 4.3.1.) from the value corresponding to a *trans* planar amide linkage (180°). Non surprisingly, the out-of-plane deviation of this peptide bond is much more pronounced than that observed for the γ -turn structure characterized as the global minimum for Ac-L-Pro-NHMe ($\omega_0 = -172.6^\circ$)⁸ and this difference should be attributed to the additional steric hindrance introduced by the aromatic system in Inc. It is worth noting that, for a *trans* arrangement of ω_0 , the methyl group of the acetyl moiety lies in close proximity to the δ position of the pyrrolidine cycle. In Pro, C^δ bears two hydrogen atoms whereas they are replaced by a fused benzene ring in Inc (Figure 4.3.1.). Therefore, values of ω_0 near 180° produce a more severe steric conflict in the latter amino acid, which is partially alleviated by a distortion of ω_0 of almost 14° (Table 4.3.1.).

Table 4.3.1. Dihedral Angles,^{a,b} Pseudorotational Parameters^a (A, P), and Relative Energy^c (ΔE^{gp}) and Free Energy^c (ΔG^{gp}) for the Minimum Energy Conformations Characterized for Ac-L-Inc-NHMe in the Gas Phase at the B3LYP/6-31+G(d,p) Level.

conf.	ω_0	ϕ	ψ	ω	χ^0	χ^1	χ^2	χ^3	χ^4	(A, P)	ΔE^{gp}	ΔG^{gp}
t- γ_L	-166.6	-88.2	64.8	-179.9	-19.2	18.0	-11.4	-0.3	12.6	(19.9, 164.3)	0.0 ^d	0.0 ^e
c- α_L	-1.1	-82.6	-10.7	-177.3	-18.6	19.6	-14.4	3.2	10.1	(20.6, 154.7)	3.1	1.9
c- ε_L	-3.1	-69.5	158.0	175.3	-14.6	14.8	-10.3	1.6	8.6	(15.7, 158.1)	7.3	5.5

^a In degrees. ^b See Figure 4.3.2. for definition. ^c In kcal/mol. ^d $E = -725.747154$ au. ^e $G = -725.543471$ au.

Besides the high deviation from planarity observed for the *trans* amide function involving the acetyl group and the pyrrolidine nitrogen (*i.e.* the acetamido moiety) in Ac-L-Inc-NHMe, another effect that should be associated with the presence of the benzene ring is the stabilization of the *cis* state for this peptide bond with respect to that found for the analogous Pro derivative. In fact, the different steric interactions established between the acetyl substituent (the methyl group or the carbonyl oxygen) and the contiguous α and δ positions of the five-membered pyrrolidine ring govern the *cis-trans* equilibrium of this peptide bond. An increase in the steric hindrance around C^δ , as occurs on going from Pro to Inc, should translate into a higher preference of the more sterically demanding fragment of the acetyl group (that is, the methyl unit) to situate far away from C^δ , that is, of this peptide bond to accommodate a *cis* arrangement. Accordingly, a higher percentage of *cis* conformers is expected for Inc with reference to Pro.

It is therefore not surprising that all other minima characterized for Ac-L-Inc-NHMe (Table 4.3.1.) exhibit the acetamido moiety in the *cis* form irrespective of the α_L or ε_L arrangement adopted by the peptide backbone (minima c- α_L and c- ε_L , Figures 4.3.3.b and 4.3.3.c, respectively). It is worth noting that, at variance with the γ -turn structure exhibited by the global minimum, the α_L and ε_L conformations are not stabilized by any intramolecular hydrogen-bonding interaction requiring a *trans* configuration of the acetamido group. It seems therefore that, in the absence of intramolecular interactions forcing ω_0 to be *trans*, only the *cis* arrangement is favorable. In comparison, both the *cis* and *trans* states of ω_0 were found to be compatible with the ε_L and α_L structures for Ac-L-Pro-NHMe,⁸ even if *trans* ε_L was located as a minimum only when levels of theory lower than B3LYP/6-31+G(d,p) were used.^{16b}

As far as *cis* ω_0 conformers are concerned, Ac-L-Pro-NHMe and Ac-L-Inc-NHMe present the same stability order, *i.e.* the c- α_L structure was found to be much more stable than c- ε_L for both compounds.

The influence of the benzene moiety on the conformation adopted by the five-membered cycle also deserves some comment. The pyrrolidine ring in Pro is known to mainly accommodate two puckering states corresponding to envelope conformations with C^γ at the flap pointing to the same (C^γ_{endo} or *down*) or the opposite (C^γ_{exo} or *up*) side of the molecule where the carboxylic terminus is located.^{3a,19} Both side-chain conformational states were found to be compatible with most of the backbone structures characterized as energy minima for Ac-L-Pro-

NHMe, which resulted in a total of 7 minima.⁸ In Inc, C^γ and C^δ are part of the aromatic system and, as a consequence, the C^β, C^γ, C^δ and N atoms are forced to be coplanar. This is shown by the χ^3 dihedral, which is defined by the C^β-C^γ-C^δ-N torsion (Figure 4.3.2.) and exhibits values close to 0° for all the minima in Table 4.3.1.. The α carbon is therefore the only atom in the five-membered ring of Ac-L-Inc-NHMe that may deviate from planarity, even if its flexibility is partially limited by its involvement in two peptide bonds. As a consequence, the five-membered cycle in the three Inc minima accommodates an envelope-like shape with C^α at the flap. This conformation may be regarded as intermediate between a planar structure and the typical envelope adopted by an unrestricted pyrrolidine, as evidenced by the relatively small puckering amplitude $A = 16\text{--}21^\circ$ (Table 4.3.1.), which is about half the value observed for the Ac-L-Pro-NHMe minima⁸ ($A \approx 37^\circ$). For all three conformers in Table 4.3.1., deviation of the α carbon from planarity occurs so that to increase the distance between the bulkier substituent at C^α, namely the methylcarboxamide group, and the vicinal acetyl group.

Table 4.3.2.. Relative Free Energies^a in the Gas Phase (ΔG^{gp}), and in Carbon Tetrachloride, Chloroform, and Aqueous Solutions^b (ΔG^{CCl_4} , ΔG^{CHCl_3} , and $\Delta G^{\text{H}_2\text{O}}$, respectively) for the Minimum Energy Conformations Characterized for Ac-L-Inc-NHMe at the B3LYP/6-31+G(d,p) Level.

conf.	ΔG^{gp}	ΔG^{CCl_4}	ΔG^{CHCl_3}	$\Delta G^{\text{H}_2\text{O}}$
t- γ_{L}	0.0	0.0	0.0	5.5
c- α_{L}	1.9	1.8	1.3	5.7
c- ε_{L}	5.5	3.2	0.3	0.0

^a In kcal/mol. ^b Derived from PCM single-point calculations on the structures optimized in the gas phase.

Table 4.3.2. compares the relative free energies in the gas phase (ΔG^{gp}) to those computed in carbon tetrachloride (ΔG^{CCl_4}), chloroform (ΔG^{CHCl_3}), and water ($\Delta G^{\text{H}_2\text{O}}$) solutions for the three Ac-L-Inc-NHMe minima displayed in Figure 4.3.3.. Such PCM calculations in solution were performed using the geometries optimized in the gas phase. The results obtained indicate that solvation introduces significant changes

in the relative energy order of these structures. In particular, the stability of the $c\text{-}\epsilon_L$ conformer increases dramatically with the polarity of the environment. Thus, although $t\text{-}\gamma_L$ remains the preferred conformation in carbon tetrachloride and chloroform solutions, these solvents increase the stability of $c\text{-}\epsilon_L$ by 2.3 and 5.2 kcal/mol, respectively. Moreover, $c\text{-}\epsilon_L$ becomes the most stable conformation in water, the other two structures being disfavored by more than 5 kcal/mol. The fact that the preferred conformation by Ac-L-Inc-NHMe moves from the γ -turn ($t\text{-}\gamma_L$) in the gas phase to polyproline II ($c\text{-}\epsilon_L$) in aqueous solution is not surprising. Indeed, conformations devoid of intramolecular hydrogen bonds provide better interactions with the solvent, and the relative stabilization produced upon solvation increases with the polarity of the environment.

The greater stabilization observed for the $c\text{-}\epsilon_L$ conformation in comparison to $c\text{-}\alpha_L$ could be associated to a larger solvent-induced polarization effect in the former case, particularly significant in polar media. According to the results in Table 4.3.2., $c\text{-}\epsilon_L$ is the only Ac-L-Inc-NHMe conformer expected to be populated at room temperature in aqueous solution. Even if $c\text{-}\epsilon_L$ was also predicted to be the global minimum for Ac-L-Pro-NHMe in water,⁸ $c\text{-}\alpha_L$, $t\text{-}\alpha_L$, and $t\text{-}\gamma_L$ conformers were found to be energetically accessible at room temperature for the latter compound. These results suggest a much superior restriction of the conformational space available to the non-proteinogenic amino acid both in terms of ω_0 and ψ .

Ac-L- α MeInc-NHMe. The five minimum energy conformations characterized for Ac-L- α MeInc-NHMe in the gas phase are displayed in Figure 4.3.4.. The most relevant structural and energy data are given in Table 4.3.3.. Minima exhibiting backbone conformations similar to those found for the unmethylated compound (Table 4.3.1.), *i.e.* $t\text{-}\gamma_L$, $c\text{-}\alpha_L$, and $c\text{-}\epsilon_L$, were also located for the α MeInc derivative and found to exhibit the same stability order. Thus, the global minimum of Ac-L- α MeInc-NHMe corresponds to $t\text{-}\gamma_L$ (Figure 4.3.4.a), as described above for the Inc derivative. The parameters associated to the intramolecular hydrogen bond [$d(\text{H}\cdots\text{O}) = 1.91 \text{ \AA}$, $\angle\text{N-H}\cdots\text{O} = 151.8^\circ$] indicate that this stabilizing interaction is stronger in the methylated compound. The $c\text{-}\alpha_L$ (Figure 4.3.4.b) and $c\text{-}\epsilon_L$ (Figure 4.3.4.e) minima are destabilized in terms of ΔG^{sp} by 2.2 and 6.0 kcal/mol, respectively, and therefore present a stability order similar to that found for the corresponding Inc minima (Table 4.3.1.).

In spite of this parallelism, significant differences are observed between the conformational profiles of the two compounds under study. Notably enough, the

presence of the α -methyl group led to the characterization of *trans* minima for the α_L and ε_L backbone conformations (Table 4.3.3.), which were not located for the unmethylated compound (Table 4.3.1.). Thus, non unexpectedly, α -methylation results in the stabilization of the *trans* arrangement of the amide bond involving the pyrrolidine nitrogen, so that it becomes compatible with all the (φ, ψ) backbone conformations accessible to a proline-like residue (γ_L , α_L , and ε_L). This effect is produced by the increase in the bulkiness at C^α associated to the replacement of the α hydrogen by a methyl group, which operates in the opposite direction to that described above for the Pro-to-Inc transition. As a matter of fact, going from Pro to Inc results in an enhancement of the steric hindrance at C^δ and a concomitant stabilization of the *cis* disposition of the acetamido moiety, whereas, α -methylation of Inc increases the bulkiness at C^α and is therefore expected to produce the opposite effect.

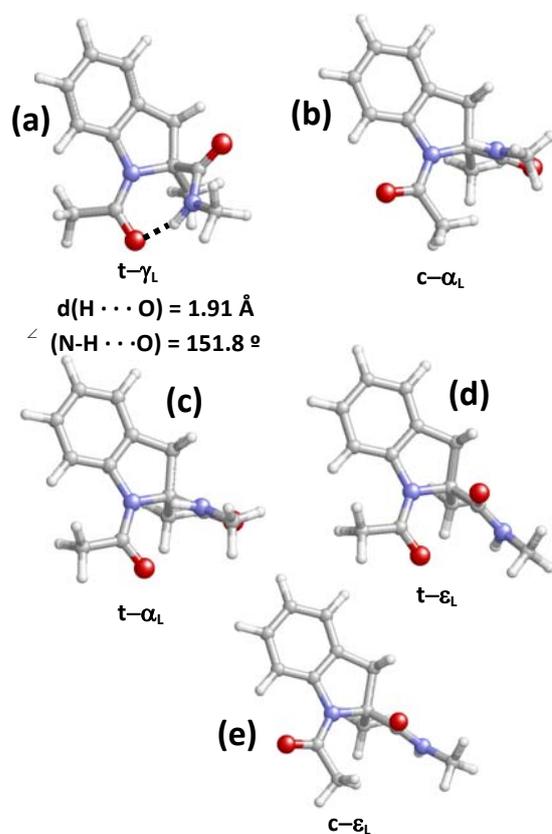


Figure 4.3.4. Representation of the minimum energy conformations characterized for Ac-L- α MeInc-NHMe at the B3LYP/6-31+G(d,p) level. Structural parameters and relative energies are provided in Table 4.3.3..

The new minimum energy structures characterized upon α -methylation, t- α_L (Figure 4.3.4.c) and t- ε_L (Figure 4.3.4.d), present ΔG^{gp} values 3.7 and 3.8 kcal/mol above the global minimum, respectively. Hence, for a *trans* arrangement of ω_0 , the α_L and ε_L backbone conformations are almost isoenergetic whereas they differ in about 4 kcal/mol when ω_0 exhibits a *cis* state (Table 4.3.3.). This result is particularly significant when the ε_L backbone structure is concerned. In this case, α -methylation not only leads to the characterization of a *trans* minimum without counterpart in the unmethylated compound; besides, this new low-energy conformer is much more stable than the one exhibiting a *cis* ω_0 arrangement ($\Delta G^{\text{gp}} = 3.8$ and 6.0 kcal/mol for t- ε_L and c- ε_L , respectively; Table 4.3.3.).

Table 4.3.3. Dihedral Angles,^{a,b} Pseudorotational Parameters^a (A, P), and Relative Energy^c (ΔE^{gp}) and Free Energy^c (ΔG^{gp}) for the Minimum

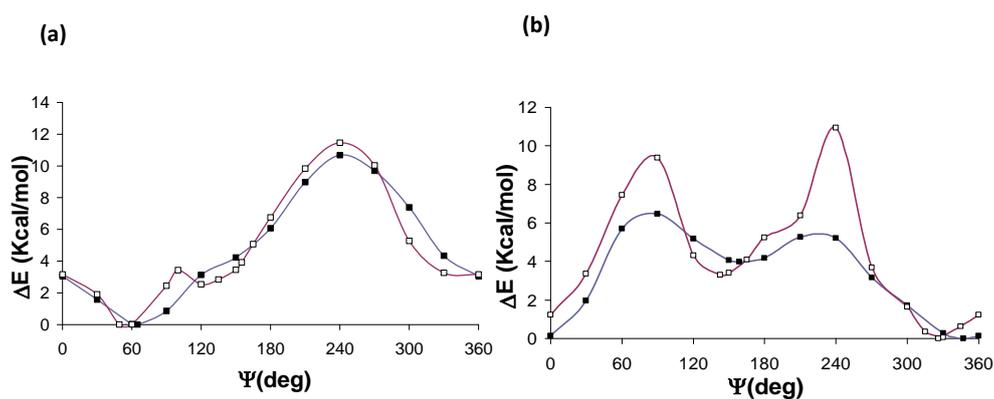
conf.	ω_0	ϕ	ψ	ω	χ^0	χ^1	χ^2	χ^3	χ^4	(A, P)	ΔE^{gp}	ΔG^{gp}
t- γ_L	-159.2	-84.8	49.2	176.1	-21.4	20.9	-14.2	0.9	13.5	(22.6, 161.2)	0.0 ^d	0.0 ^e
c- α_L	5.2	-62.0	-32.8	-175.9	12.8	-13.2	9.5	-1.7	-7.4	(14.0, -23.5)	3.5	2.2
t- α_L	-174.5	-58.3	-33.5	177.5	18.0	-18.0	12.6	-1.6	-10.8	(19.3, -21.2)	4.6	3.7
t- ε_L	174.2	-50.2	126.7	-174.9	18.3	-19.0	13.8	-2.7	-10.3	(20.0, -24.1)	4.2	3.8
c- ε_L	8.0	-52.0	140.8	177.9	20.1	-21.3	15.8	-3.6	-11.0	(22.3, -25.6)	7.0	6.0

Energy Conformations Characterized for Ac-L- α MeInc-NHMe in the Gas Phase at the B3LYP/6-31+G(d,p) Level.^a In degrees. ^b See Figure 4.3.2. for definition. ^c

In kcal/mol. ^d E = -765.060772 au. ^e G = -764.830288 au.

The conformational features outlined above are clearly seen in Figure 4.3.5., which compares the potential energy curves $E=E(\psi)$ calculated for Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe with either a *trans* (Figure 4.3.5.a) or a *cis* (Figure 4.3.5.b) arrangement of the acetamido peptide bond. Such curves were obtained following the flexible rotor approximation, by fixing the value of the dihedral angle ψ while allowing the rest of the molecule to relax. Thus, each point of the path was derived from a constrained geometry optimization, with ψ being the only variable fixed. The ψ space (0° to 360°) was explored in steps of 30° . The energy curves obtained for the Inc and α MeInc derivatives with all *trans* peptide bonds (Figure 4.3.5.a) differ in the ψ regions corresponding to the α_L and ϵ_L backbone conformations, where energy minima appear only for the latter compound. In comparison, when the acetamido moiety is arranged in *cis* (Figure 4.3.5.b), the profiles of both compounds exhibit minima in the same regions (α_L and ϵ_L), with extremely high energy barriers being observed for the methylated derivative.

Figure 4.3.5.



Another general effect arising from α -methylation is the change in the conformation adopted by the five-membered ring. As observed before for Inc, the five-membered cycle accommodates an incipient envelope conformation with the α carbon deviating slightly from the plane formed by the four other atoms. However, in all the minima located for the α MeInc derivative with the exception of the $t\text{-}\gamma_L$ one, the α carbon atom protrudes from the plane in the opposite direction to that observed for the Inc minima. This is clearly illustrated by the value adopted by the χ^2 dihedral, which is close to -12° for the three Inc minima (Table 4.3.1.) and the $t\text{-}\gamma_L$ α MeInc structure (Table 4.3.3.), whereas it approaches 12° for the remaining α MeInc conformers in Table 4.3.3.. The reason for this change is that the substituent at C^α introducing stronger steric repulsions with the rest of the system in α MeInc is not the methylcarboxamide moiety (as was in Inc) but the methyl group. Accordingly, in all the α MeInc minima but $t\text{-}\gamma_L$, it is the more sterically demanding substituent at C^α (methyl) that moves away from the neighboring acetyl group. In the $t\text{-}\gamma_L$ conformer, the five-membered ring probably adopts the shape that provides an optimal geometry for intramolecular hydrogen-bonding.

Finally, comparison of the geometric parameters in Tables 4.3.1. and 4.3.3. reveals some non-negligible differences between minima of the same type characterized for the two compounds investigated, that is, $t\text{-}\gamma_L$, $c\text{-}\alpha_L$, and $c\text{-}\varepsilon_L$. In particular, the deviation from planarity observed for ω_0 in the three Inc minima is enhanced in the α MeInc counterparts for both a *cis* and a *trans* arrangement. Thus, the $t\text{-}\gamma_L$ ω_0 angle deviates by more than 20° from the ideal 180° value in Ac-L- α MeInc-NHMe, and the *cis* low-energy conformers also present a larger out-of-plane deviation ($5\text{--}8^\circ$) than that observed for the unmethylated compound ($1\text{--}3^\circ$). The (φ, ψ) backbone angles also exhibit some remarkable differences. Specifically, the absolute value of the φ dihedral is about 20° lower for the $c\text{-}\alpha_L$ and $c\text{-}\varepsilon_L$ conformers of the α -methyl derivative. The same holds true when comparing the ψ absolute value of the $c\text{-}\varepsilon_L$ and $t\text{-}\gamma_L$ conformers, while the $c\text{-}\alpha_L$ structure varies in the opposite direction. All these changes should be attributed to the additional steric repulsions introduced by the α -methyl group in Ac-L- α MeInc-NHMe. Interestingly, similar differences in molecular geometry are observed when Pro and α MePro are compared.⁸

Table 4.3.4. Relative Free Energies^a in the Gas Phase (ΔG^{gp}), and in Carbon Tetrachloride, Chloroform, and Aqueous Solutions^b (ΔG^{CCl_4} , ΔG^{CHCl_3} , and $\Delta G^{\text{H}_2\text{O}}$, respectively) for the Minimum Energy Conformations Characterized for Ac-L- α MeInc-NHMe at the B3LYP/6-31+G(d,p) Level.

conf.	ΔG^{gp}	ΔG^{CCl_4}	ΔG^{CHCl_3}	$\Delta G^{\text{H}_2\text{O}}$
t- γ_L	0.0	0.0	0.0	2.8
c- α_L	2.2	0.9	0.2	0.0
t- α_L	3.7	2.8	1.5	1.6
t- ε_L	3.8	2.9	2.3	3.4
c- ε_L	6.0	3.8	1.6	1.2

^a In kcal/mol. ^b Derived from PCM single-point calculations on the structures optimized in the gas phase.

Table 4.3.4. illustrates the effects of the environment on the relative stability of the five Ac-L- α MeInc-NHMe minima, as calculated by PCM methods considering the molecular geometries obtained in the gas phase. The most remarkable result is the stabilization of conformers exhibiting a *cis* disposition of ω_0 . Thus, even if t- γ_L remains the preferred conformation in chlorinated solvents, the energy gap between this structure and either the c- α_L or the c- ε_L minimum is substantially reduced. This effect is even more intense in water, where c- α_L becomes the global minimum and all other conformers are destabilized by at least 1.2 kcal/mol. The c- α_L structure was also found to be the most stable conformation for Ac-L- α MePro-NHMe in aqueous solution⁸ but, in this case, both the c- ε_L and t- α_L structures were found within a free energy interval of 0.6 kcal/mol. These results evidence a higher restriction of the conformational space available to α MeInc in comparison to α MePro in this environment. Similar differences –although less marked– to those commented above for Inc and Pro are therefore observed when their respective α -methyl derivatives are compared.

PCM Geometry Optimizations. In the previous sections, the effect of the environment on the conformational preferences of Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe was evaluated by considering the molecular geometries obtained in the gas phase, that were submitted to single-point calculations using the PCM model. This SCRF method has been shown to provide reliable results for a number of solutes, for which geometry relaxation in solution gave very similar free energies of solvation.²⁷ However, an overestimation of the relative stability of conformers with ω_0 arranged in *cis* was detected for Pro and α MePro when using PCM single-point calculations.⁸ This effect, which seemed to be particularly marked in high-polarity solvents, might be related to the existence of some internal geometric stress in such constrained cyclic solutes. In order to test this possibility, PCM geometry optimizations were performed in aqueous solution using as starting geometries all the minima characterized for Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe in the gas phase (Tables 4.3.1. and 4.3.3.). In addition, the t- ϵ_L conformation, which was derived from QM/MM calculations (see below), was also used as starting geometry for Ac-L-Inc-NHMe. The results obtained are summarized in Table 4.3.5., which shows the relative free energies and the most representative dihedral angles.

Comparison of the $\Delta G^{\text{H}_2\text{O}}$ values displayed in Tables 4.3.2. and 4.3.5. for Ac-L-Inc-NHMe indicates that, even if molecular geometry optimization does not alter the structure characterized as global minimum in aqueous solution, the energy gap between the most and least stable conformer is substantially reduced. Specifically, geometry relaxation results in the stabilization of the c- α_L and t- γ_L minima by 4.1 and 2.0 kcal/mol, respectively. Thus, although c- ϵ_L remains the preferred backbone conformation for Ac-L-Inc-NHMe, the c- α_L structure is destabilized by 1.6 kcal/mol (5.7 kcal/mol in Table 4.3.2.) and becomes more stable than t- γ_L . Moreover, according to the results in Table 4.3.5., both the *cis* and *trans* states of ω_0 should be accessible to Ac-L-Inc-NHMe at room temperature. In particular, the c- ϵ_L and t- ϵ_L conformers differ only in 0.8 kcal/mol. On the other hand, the backbone dihedrals in Table 4.3.5. are very similar to those listed in Table 4.3.1., the largest difference being detected for the ψ angle of the c- ϵ_L structure (6.3°).

Analysis of the results obtained for Ac-L- α MeInc-NHMe reveals some remarkable changes in molecular geometry. Specifically, the ψ dihedrals in Tables 4.3.3. and 4.3.5. differ in more than 10° for three of the five minima (above 15° for the t- ϵ_L structure) and a significant variation (near 7°) is also observed for ω_0 in the c- ϵ_L

conformer. Conversely, the differences in the $\Delta G^{\text{H}_2\text{O}}$ values displayed in Tables 4.3.4. and 4.3.5. are quantitatively inferior to those commented for the unmethylated compound, although qualitatively more significant. Thus, the greatest $\Delta G^{\text{H}_2\text{O}}$ difference observed for the αMeInc minima corresponds to the $t\text{-}\varepsilon_{\text{L}}$ structure, which is stabilized by 2.0 kcal/mol in Table 4.3.5.. However, geometry optimization brings about a change in the relative energy order of the minima. PCM single-point calculations in aqueous solution predicted any of the *cis* αMeInc minima to be more stable than any of the *trans* conformers (relative stability: $c\text{-}\alpha_{\text{L}} > c\text{-}\varepsilon_{\text{L}} > t\text{-}\alpha_{\text{L}} > t\text{-}\gamma_{\text{L}} > t\text{-}\varepsilon_{\text{L}}$, Table 4.3.4.), whereas geometry relaxation resulted in the $t\text{-}\alpha_{\text{L}}$ and $t\text{-}\varepsilon_{\text{L}}$ structures becoming lower in energy than the $c\text{-}\varepsilon_{\text{L}}$ one (relative stability: $c\text{-}\alpha_{\text{L}} > t\text{-}\alpha_{\text{L}} > t\text{-}\varepsilon_{\text{L}} > c\text{-}\varepsilon_{\text{L}} > t\text{-}\gamma_{\text{L}}$, Table 4.3.5.). In fact, the latter model suggests an almost equal preference for the *cis* and *trans* disposition of ω_0 for both the α_{L} and ε_{L} backbone arrangements as denoted by the differences in the $\Delta G^{\text{H}_2\text{O}}$ values between the corresponding pair of minima (0.5 and 0.2 kcal/mol, respectively; Table 4.3.5.). In comparison, such differences reach 1.6 and 2.2 kcal/mol, respectively, in Table 4.3.4., with the *cis* form being preferred in both cases.

Table 4.3.5. Backbone Dihedral Angles (in deg), Pseudorotational Parameters (A , P ; in deg) and Relative Free Energy in Aqueous Solution ($\Delta G^{\text{H}_2\text{O}}$; in kcal/mol) of Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe

conf.	ω_0	ϕ	ψ	ω	(A , P)	$\Delta G^{\text{H}_2\text{O}}$
Ac-L-Inc-NHMe						
t- γ_{L}	-166.7	-90.1	68.3	-178.3	(22.1, 163.7) ^a	3.5
c- α_{L}	4.4	-81.5	-13.1	-178.6	(15.0, 155.1) ^b	1.6
c- ε_{L}	-1.6	-73.9	151.7	175.3	(15.1, 157.6) ^c	0.0 ^d
t- ε_{L} ^e	-173.3	-76.5	148.8	176.6	(15.3, 162.1) ^f	0.8
Ac-L- α MeInc-NHMe						
t- γ_{L}	-158.3	-86.6	50.0	177.0	(24.1, 160.8) ^g	2.3
c- α_{L}	3.3	-60.9	-34.0	-175.0	(11.2, -23.4) ^h	0.0 ⁱ
t- α_{L}	-177.7	-51.7	-43.1	-176.8	(19.2, -19.2) ^j	0.5
t- ε_{L}	-174.4	-50.9	141.8	176.3	(17.9, -20.2) ^k	1.4
c- ε_{L}	1.1	-52.3	151.2	175.2	(17.3, -25.3) ^l	1.6

^a $\chi^0 = -21.2^\circ$, $\chi^1 = 20.0^\circ$, $\chi^2 = -12.9^\circ$, $\chi^3 = -0.1^\circ$ and $\chi^4 = 13.7^\circ$. ^b $\chi^0 = -13.0^\circ$, $\chi^1 = 13.7^\circ$, $\chi^2 = -10.2^\circ$, $\chi^3 = 2.5^\circ$ and $\chi^4 = 7.0^\circ$. ^c $\chi^0 = -14.0^\circ$, $\chi^1 = 14.1^\circ$, $\chi^2 = -10.1^\circ$, $\chi^3 = 1.7^\circ$ and $\chi^4 = 8.1^\circ$. ^d $G = -725.568457$ au. ^e The coordinates used as starting point were derived from hybrid QM/MM simulations. ^f $\chi^0 = -14.6^\circ$, $\chi^1 = 14.0^\circ$, $\chi^2 = -9.3^\circ$, $\chi^3 = 0.5^\circ$ and $\chi^4 = 9.3^\circ$. ^g $\chi^0 = -22.8^\circ$, $\chi^1 = 22.4^\circ$, $\chi^2 = -15.4^\circ$, $\chi^3 = 1.1^\circ$ and $\chi^4 = 14.4^\circ$. ^h $\chi^0 = 10.3^\circ$, $\chi^1 = -10.6^\circ$, $\chi^2 = 7.7^\circ$, $\chi^3 = -1.4^\circ$ and $\chi^4 = -6.0^\circ$. ⁱ $G = -764.850972$ au. ^j $\chi^0 = 18.1^\circ$, $\chi^1 = -17.6^\circ$, $\chi^2 = 12.1^\circ$, $\chi^3 = -1.0^\circ$ and $\chi^4 = -11.3^\circ$. ^k $\chi^0 = 16.8^\circ$, $\chi^1 = -16.5^\circ$, $\chi^2 = 11.5^\circ$, $\chi^3 = -1.3^\circ$ and $\chi^4 = -10.3^\circ$. ^l $\chi^0 = 15.6^\circ$, $\chi^1 = -16.3^\circ$, $\chi^2 = 12.2^\circ$, $\chi^3 = -2.8^\circ$ and $\chi^4 = -8.6^\circ$.

Multi-scale Approach (Hybrid QM/MM). Although PCM geometry optimizations decreased the ΔG^{H2O} values of conformations with ω_0 arranged in *trans*, the relative stability of *cis* structures might be still overestimated. It should be noted that only the *trans* arrangement of ω_0 has been detected experimentally for Ac-L- α MePro-NHMe.²⁸ Even if the fused aromatic ring in α MeInc is expected to increase the population of *cis* conformers with respect to α MePro, the fact that an α -tetrasubstituted proline derivative presents similar propensities to accommodate the *cis* and *trans* states of ω_0 is *a priori* surprising.

In order to analyze whether the limitations of the PCM approach are due to the solvation protocol used in this method or to the omission of explicit solute-solvent interactions, additional calculations were performed. In particular, the effects of the first solvation shell, which are particularly important in protic solvents, on the conformational preferences of Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe were examined by performing hybrid QM/MM calculations in explicit water. Two trajectories were run for each compound differing in the *cis/trans* arrangement of the acetamido moiety (ω_0). Table 4.3.6. lists the average values of the backbone dihedral angles and the populations for the conformers calculated using hybrid QM/MM simulations, in which the solute, *i.e.* Ac-L-Inc-NHMe or Ac-L- α MeInc-NHMe, was described at the B3LYP/6-31+G(d) level. The ΔG^\ddagger values listed in Table 4.3.6., which were derived from the populations using a Boltzmann distribution, refer to the relative stability of the different backbone conformations detected during each trajectory. All snapshots obtained from the QM/MM calculations were processed in order to characterize the most visited conformers during the simulations, the backbone conformations being categorized according to Perczel's nomenclature.²⁰

The most populated conformation of Ac-L-Inc-NHMe corresponds to ϵ_L , independently of the *cis/trans* initial state of ω_0 . This backbone structure combined with a *trans* disposition of ω_0 ($t\text{-}\epsilon_L$) was not identified as an energy minimum in the gas phase (Table 4.3.1.). However, subsequent geometry optimizations using the PCM method suggested this conformation to be very stable in solution (Table 4.3.5.). The populations observed for the $c\text{-}\epsilon_L$ and $c\text{-}\alpha_L$ conformations are very similar and correspond to a relative free energy difference ($\Delta G^\ddagger = 0.2$ kcal/mol) significantly smaller than that obtained with the PCM approach, particularly when single-point

calculations were performed ($\Delta G^{\text{H}_2\text{O}} = 5.7$ and 1.6 kcal/mol in Tables 4.3.2. and 4.3.5., respectively).

Table 4.3.6.. Backbone Dihedral Angles (in deg), Population (in %) and Relative Free Energy^a (ΔG^\ddagger ; in kcal/mol) Derived from the Trajectories^b Run for Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe in Explicit Water Using Hybrid QM/MM Simulations. Standard Deviations are Shown.

conf.	ω_0	ϕ	ψ	ω	population	ΔG^\ddagger
Ac-L-Inc-NHMe (<i>cis</i>) ^b						
c- ϵ_L	-8.0±10.6	-73.6±9.3	161.5±11.1	176.6±12.5	51	0.0
c- α_L	3.4±12.9	-72.6±10.8	-27.1±15.4	-175.7±8.1	42	0.2
c- γ_L	8.4±8.8	-91.6±7.4	22.1±14.8	-175.7±8.1	7	2.3
Ac-L-Inc-NHMe (<i>trans</i>) ^b						
t- ϵ_L	-178.3±10.6	-77.5±14.3	140.2±12.8	-178.3±10.9	90	0.0
t- γ_L	-179.4±8.0	-79.7±11.8	113.7±4.5	-176.5±8.0	10	1.3
Ac-L- α MeInc-NHMe (<i>cis</i>) ^b						
c- α_L	3.9±13.3	-59.7±13.0	-47.2±8.7	-173.9±8.9	100	0.0
Ac-L- α MeInc-NHMe (<i>trans</i>) ^b						
t- α_L	-176.5±10.1	-59.6±8.6	-33.6±14.3	-179.4±10.0	79	0.0
t- γ_L	-152.7±11.0	-85.5±8.4	43.1±19.1	-174.9±10.2	21	0.8

^a ΔG^\ddagger values were estimated for each trajectory from the population values assuming a Boltzmann distribution of conformers. ^b Two trajectories differing in the *cis* or *trans* arrangement of ω_0 were run for each compound.

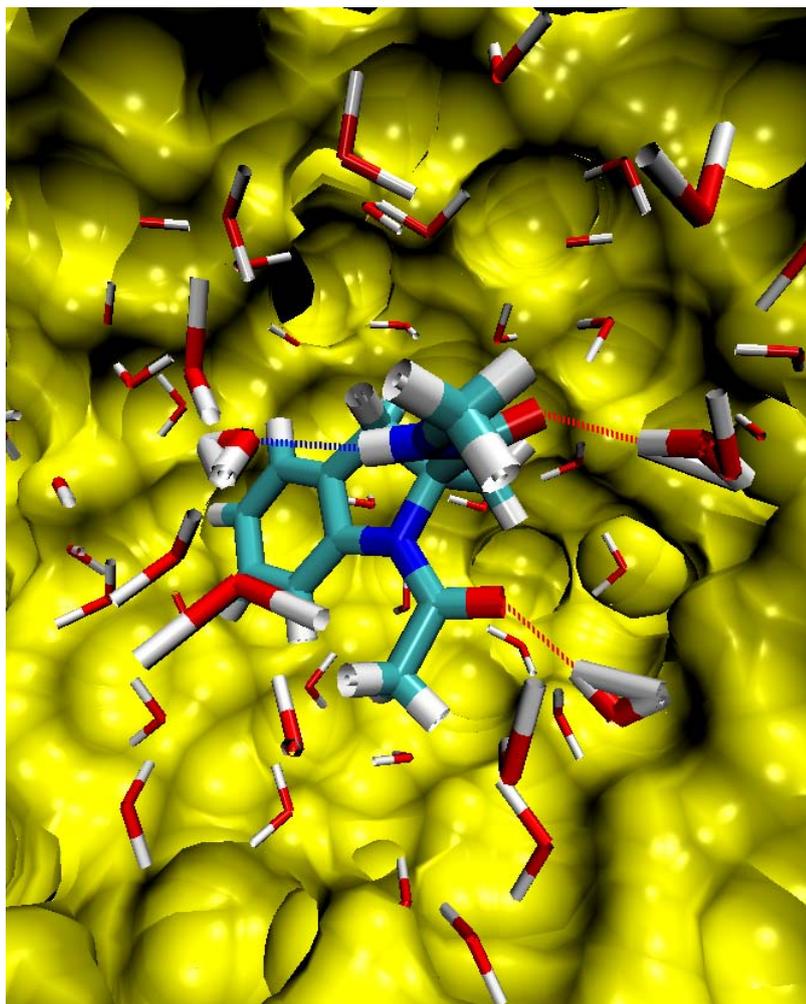


Figure 4.3.6.. Model of a selected snapshot from a trajectory of Ac-L- α MeInc-NHMe in explicit water obtained from multi-scale simulations (hybrid QM/MM). The hydrogen bonds established between the water molecules and the CO or NH moieties of the peptide backbone are represented by red and blue dotted lines, respectively.

The energy difference between *cis* and *trans* conformers for Ac-L-Inc-NHMe was estimated by averaging the energies recorded during the hybrid QM/MM trajectories, which included the electronic energy of the solute, the classical energy of the solvent and the interaction energy of the quantum and classical regions. In order to obtain such average values, the first 2 ps of each QM/MM trajectory were rejected, the

resulting energy difference being derived from the data recorded during a total simulation time of 8 ps (4 ps for each conformation). In this way, the *trans* disposition of ω_0 was predicted to be 4.5 kcal/mol lower in energy than the *cis* state for this compound. Taking as a basis this high energy difference and assuming comparable entropy contributions to the free energy for both peptide bond arrangements, a very high population of *trans* conformers may be expected for Ac-L-Inc-NHMe according to the QM/MM approach. This result differs significantly from those derived from PCM calculations, which suggested the *cis* state of ω_0 to be preferred (Tables 4.3.2. and 4.3.5.).

For the α -methylated derivative, the most populated backbone conformation falls in the α_L region for both a *cis* and a *trans* arrangement of ω_0 (Table 4.3.6.). This result is in excellent agreement with the data displayed in Table 4.3.5.. Surprisingly enough, conformers exhibiting an ϵ_L backbone structure were not detected in the QM/MM simulation for this compound. The results from hybrid QM/MM and PCM calculations with geometry relaxation also show discrepancies with regard to the energy gap between the $t\text{-}\gamma_L$ and $t\text{-}\alpha_L$ conformations, which was estimated to be 0.8 (Table 4.3.6.) and 1.8 (Table 4.3.5.) kcal/mol, respectively. The backbone dihedral angles in Table 4.3.6. do not differ substantially from those obtained when the molecular geometry is relaxed through a SCRF method, with the highest deviation being detected for the ψ dihedral in the $c\text{-}\alpha_L$ minimum ($\approx 13^\circ$). Wider variations (of up to 45°) were observed for the unmethylated compound. The energy difference between the *cis* and *trans* conformers of Ac-L- α MeInc-NHMe was obtained from the data recorded during the last 4 ps of each QM/MM simulation and indicated that the *trans* arrangement is 6.0 kcal/mol more stable than the *cis* one. This finding indicates that α -methylation of Inc increases the stability of the *trans* disposition of ω_0 , as expected. However, the hybrid QM/MM calculations performed seem to overestimate the stability of the *trans* arrangement of ω_0 , since a non-negligible percentage of *cis* conformers (superior to that exhibited by Pro) is expected for the Inc-containing compound. Clearly, further work in this context, including experimental studies –to the best of our knowledge, non available to date–, is necessary before the conformational propensities of Inc and α MeInc regarding the *cis-trans* arrangement of ω_0 is unambiguously established.

The discrepancies between the results provided by the implicit and the explicit solvation models applied could be caused by imperfections of the PCM protocol. In

particular, the separation of the solute from the nearest solvent molecules is not large enough for the solute to see a continuous solvent. Accordingly, the solute fails to see a uniform dielectric constant,²⁹ and this deficiency becomes particularly significant in the presence of a polar solvent like water. Additionally, QM/MD allows the whole system to relax considering the anisotropies around the first solvation shell.

The differences in predicting the relative stability order of *cis* and *trans* conformers provided by the PCM and QM/MM methods should be attributed mainly to the local anisotropies and the hydrogen-bonding network characterizing the latter type of calculation, which results in a higher restriction of the degrees of freedom and therefore leads to a limited number of accessible conformations. Figure 4.3.6. illustrates the main hydrogen bonding interactions established between the water molecules of the first solvation shell and the CO and NH moieties in the peptide backbone for Ac-L- α MeInc-NHMe exhibiting a *trans* acetamido moiety. It should be noted that no significant differences were observed for the four systems studied regarding the number of water molecules arranged in the first solvation shell, which ranged from 33 to 35 in all cases. However, averaged life-time increments in the most populated hydrogen-bonds were observed for both the unmethylated (23.2 and 24.9 ps for *cis* and *trans* Ac-L-Inc-NHMe, respectively) and methylated (21.5 and 32.8 ps for *cis* and *trans* Ac-L- α MeInc-NHMe, respectively) compounds. Accordingly, the Inc and, particularly, the α MeInc derivative seem to be stabilized by a stronger hydrogen-bonding network when the peptide bonds are arranged in *trans*. This could account, at least in part, for the higher stability predicted by the QM/MM approach for the *trans* conformers of Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe.

4.3.3. Conclusions

The conformational preferences of Ac-L-Inc-NHMe and its α -methyl derivative, Ac-L- α MeInc-NHMe, have been explored in the gas phase by quantum mechanical calculations at the B3LYP/6-31+G(d,p) level. Such amino acids are proline analogues bearing a benzene ring fused to the C ^{γ} -C ^{δ} bond of the pyrrolidine moiety. Moreover, the influence of the environment on the conformational preferences of these compounds has been examined using both a PCM SCRF method and a hybrid QM/MM approach with explicit solvent molecules. The following conclusions have been drawn:

The presence of the fused aromatic moiety further reduces the intrinsically low conformational flexibility of Pro. This feature is reflected by the annihilation in Ac-L-Inc-NHMe of some of the minima previously detected for Ac-L-Pro-NHMe in the gas phase at the same level of theory. In particular, a restriction of the flexibility of the five-membered ring and a higher preference for the *cis* state of the peptide bond involving the pyrrolidine nitrogen is observed for the non-proteinogenic amino acid. α -Methylation of Inc results in the appearance of new minima exhibiting a *trans* arrangement of this peptide bond. Regarding the (φ, ψ) backbone preferences, the γ -turn motif ($t\text{-}\gamma_L$) was found to be the global minima for the two compounds investigated in the gas phase and in chlorinated solvents.

The presence of a polar solvent like water leads to the stabilization of conformations devoid of intramolecular hydrogen-bonding interactions. In particular, the semiextended polyproline II (ϵ_L) and the folded α -helical (α_L) conformations were found to be favored for Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe, respectively. This result is in accordance with the higher propensity to accommodate folded conformations typically exhibited by α -methyl amino acids when compared to their non-methylated counterparts. The results derived from PCM and QM/MM methods in this respect are in good agreement.

PCM single-point calculations on the geometries optimized in the gas phase seem to overestimate the stability of conformations exhibiting the peptide bond involving the pyrrolidine nitrogen arranged in *cis*, especially in aqueous solution. Relaxation of the molecular geometries in the latter environment produces a significant reduction of the energy gap between *cis* and *trans* conformers for a given (φ, ψ) backbone type. Elimination of the internal geometric stress seems essential to improve the description of the solvent effects on solutes with high conformational constraints like the ones investigated. Hybrid QM/MM simulations using explicit water molecules predict the *trans* arrangement of this peptide bond to be favored over the *cis* one by 4.5 and 6.0 kcal/mol for Ac-L-Inc-NHMe and Ac-L- α MeInc-NHMe, respectively. In this case, the stability of *trans* conformers is probably overestimated.

4.3.4. References

- [1] (a) *Peptide and Protein Design for Biopharmaceutical Applications*. Jensen, K. J., Ed.; John Wiley & Sons: Chichester, 2009. (b) Nestor, J. J. *Curr. Med. Chem.* **2009**, *16*, 4399–4418. (c) Chatterjee, J.; Gilon, C.; Hoffman, A.; Kessler, H. *Acc. Chem. Res.* **2008**, *41*, 1331–1342. (d) *Pseudo-peptides in Drug Discovery*. Nielsen, P. E., Ed.; Wiley-VCH: Weinheim, 2004. (e) Adessi, C.; Soto, C. *Curr. Med. Chem.* **2002**, *9*, 963–978.
- [2] (a) Zanuy, D.; Ballano, G.; Jiménez, A. I.; Casanovas, J.; Haspel, N.; Cativiela, C.; Curcó, D.; Nussinov, R.; Alemán, C. *J. Chem. Inf. Model.* **2009**, *49*, 1623–1629. (b) Wu, X.; Schultz, P. G. *J. Am. Chem. Soc.* **2009**, *131*, 12497–12515. (c) Alemán, C.; Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Haspel, N.; Zheng, J.; Casanovas, J.; Wolfson, H.; Nussinov, R. *Phys. Biol.* **2006**, *3*, S54–S62. (d) Tu, R. S.; Tirrell, M. *Adv. Drug Delivery Rev.* **2004**, *56*, 1537–1563.
- [3] (a) Chakrabarti, P.; Pal, D. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1–102. (b) MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397–412. (c) Rose, G. D.; Gierasch, L. M.; Smith, J. A. *Adv. Protein Chem.* **1985**, *37*, 1–109.
- [4] (a) Dugave, C.; Demange, L. *Chem. Rev.* **2003**, *103*, 2475–2532. (b) Stewart, D. E.; Sarkar, A.; Wampler, J. E. *J. Mol. Biol.* **1990**, *214*, 253–260. (c) Grathwohl, C.; Wüthrich, K. *Biopolymers* **1981**, *20*, 2623–2633.
- [5] (a) Bourgault, S.; Vaudry, D.; Ségalas-Milazzo, I.; Guilhaudis, L.; Couvineau, A.; Laburthe, M.; Vaudry, H.; Fournier, A. *J. Med. Chem.* **2009**, *52*, 3308–3316. (b) Mandal, P. K.; Limbrick, D.; Coleman, D. R.; Dyer, G. A.; Ren, Z.; Birtwistle, J. S.; Xiong, C.; Chen, X.; Briggs, J. M.; McMurray, J. S. *J. Med. Chem.* **2009**, *52*, 2429–2442. (c) Willoughby, C. A.; Bull, H. G.; Garcia-Calvo, M.; Jiang, J.; Chapman, K. T.; Thornberry, N. A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2197–2200. (d) Pahl, A.; Zhang, M.; Török, K.; Kuss, H.; Friedrich, U.; Magyar, Z.; Szekely, J.; Horvath, K.; Brune, K.; Szelenyi, I. *J. Pharmacol. Exp. Ther.* **2002**, *301*, 738–746. (e) Ingallinella, P.; Fattori, D.; Altamura, S.; Steinkühler, C.; Koch, U.; Cicero, D.; Bazzo, R.; Cortese, R.; Bianchi, E.; Pessi, A. *Biochemistry* **2002**, *41*, 5483–5492. (f) Ohmoto, K.; Yamamoto, T.; Okuma, M.; Horiuchi, T.; Imanishi, H.; Odagaki, Y.; Kawabata, K.; Sekioka, T.; Hirota, Y.; Matsuoka, S.; Nakai, H.; Toda, M. *J. Med. Chem.* **2001**, *44*, 1268–1285. (g) Wiczorek, M.; Gyorkos, A.; Spruce, L. W.; Ettinger, A.; Ross, S. E.; Kroona, H. S.; Burgos-Lepley, C. E.; Bratton, L. D.; Drennan, T. S.; Garnert, D. L.; Von Burg, G.; Pilkington, C. G.; Cheronis, J. C. *Arch. Biochem. Biophys.* **1999**, *367*,

193–201. (h) Leftheris, K.; Kline, T.; Vite, G. D.; Cho, Y. H.; Bhide, R. S.; Patel, D. V.; Patel, M. M.; Schmidt, R. J.; Weller, H. N.; Andahazy, M. L.; Carboni, J. M.; Gullo-Brown, J. L.; Lee, F. Y. F.; Ricca, C.; Rose, W. C.; Yan, N.; Barbacid, M.; Hunt, J. T.; Meyers, C. A.; Seizinger, B. R.; Zahler, R.; Manne, V. *J. Med. Chem.* **1996**, *39*, 224–236. (i) Flouret, G.; Briehar, W.; Majewski, T.; Mahan, K.; Wilson, L., Jr. *J. Med. Chem.* **1991**, *34*, 2089–2094.

[6] (a) Tong, X.; Malcolm, B. A.; Huang, H.-C. PCT Int. Appl. 2007, WO 2007092616. (b) Jeong, J. U.; Marquis, R. W. PCT Int. Appl. 2007, WO 2007030761. (c) Zheleva, D. I.; Fischer, P. M.; McInnes, C.; Andrews, M. J. I.; Chan, W. C.; Atkinson, G. E. PCT Int. Appl. 2005, WO 2005040802. (d) Mortimore, M.; Golec, J. M. C. PCT Int. Appl. 2003, WO 2003068242. (e) Skinner, J. E.; Anchin, J. M. PCT Int. Appl. 2003, WO 2003065997. (f) Miller, K.; Diu-Hercend, A.; Hercend, T.; Lang, P.; Weber, P.; Golec, J.; Mortimore, M. PCT Int. Appl. 2003, WO 2003088917. (g) Karanewsky, D. S.; Kalish, V. J.; Robinson, E. D.; Ullman, B. R. PCT Int. Appl. 2000, WO 2000023421. (h) Karanewsky, D. S.; Ternansky, R. J. PCT Int. Appl. 2000, WO 2000001666. (i) Spruce, L.; Gyorkos, A. PCT Int. Appl. 1999, WO 9962514. (j) Reichert, D.; Kutscher, B.; Bang, H.; Brune, K.; Quinkert, G.; Schaible, H.-G. PCT Int. Appl. 1997, WO 9741148. (k) Stewart, J. M.; Gera, L.; Whalley, E. T. PCT Int. Appl. 1996, WO 9616081. (l) Patel, D. V.; Kline, T. B.; Meyers, C. A.; Leftheris, K.; Bhide, R. S. Eur. Pat. Appl. 1994, EP 618221. (m) Ryder, H.; Kendrick, D. A.; Semple, G.; Miyata, K.; Batt, A. R.; Mathews, E. A.; Rooker, David P.; Nishida, A. PCT Int. Appl. 1993, WO 9320099. (n) Vincent, M.; Remond, G.; Cudennec, C. US Pat. Appl. 1988, US 4720484. (o) Gruenfeld, N. US Pat. Appl. 1984, US 4479963.

[7] Toniolo, C.; Crisma, M.; Formaggio, F.; Peggion, C. *Biopolymers (Pept. Sci.)* **2001**, *60*, 396–419.

[8] Flores-Ortega, A.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C.; Casanovas, J. *J. Org. Chem.* **2008**, *73*, 3418–3427.

[9] Bylaska, E. J.; Jong, W. A. d.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Apra, E.; Windus, T. L.; Hammond, J.; Nichols, P.; Hirata, S.; Hackler, M. T.; Zhao, Y.; Fan, P.-D.; Harrison, R. J.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Wu, Q.; Voorhis, T. V.; Auer, A. A.; Nooijen, M.; Brown, E.; Cisneros, G.; Fann, G. I.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J. A.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt,

D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Pollack, L.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; Lenthe, J. v.; Wong, A.; Zhang, Z.; 5.1 ed.; Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA.: 2007.

[10] Torras, J.; Deumens, E.; Trickey, S. B.; Cheng, H.-P.; Cao, C.; He, Y.; Muralidharana, K.; Roitberg, A. E.; Seabra, G. M.; University of Florida Quantum Theory Project: 2008.

[11] Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; B.Wang; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; K.F.Wong; Paesani, F.; Vanicek, J.; X.Wu; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. In *AMBER 10*; University of California: San Francisco, 2008.

[12] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; J. A. Montgomery, J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004: 2004.

[13] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

[14] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

- [15] (a) Holroyd, L. F.; van Mourik, T. *Chem. Phys. Lett.* **2007**, *442*, 42–46. (b) Zhao, Y.; González-García, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 2012–2018.
- [16] (a) Kang, Y. K. *J. Phys. Chem. B* **2006**, *110*, 21338–21348. (b) 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*, 2660–2679.
- [17] (a) Flores-Ortega, A.; Casanovas, J.; Assfeld, X.; Alemán, C. *J. Org. Chem.* **2009**, *74*, 3101–3108. (b) Zanuy, D.; Flores-Ortega, A.; Jiménez, A. I.; Calaza, M. I.; Cativiela, C.; Nussinov, R.; Ruoslahti, E.; Alemán, C. *J. Phys. Chem. B* **2009**, *113*, 7879–7889. (c) Flores-Ortega, A.; Casanovas, J.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2008**, *112*, 14045–14055. (d) Flores-Ortega, A.; Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2007**, *111*, 5475–5482. (e) Song, I. K.; Kang, Y. K. *J. Phys. Chem. B* **2006**, *110*, 1915–1927.
- [18] McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639–5648.
- [19] (a) Milner-White, E. J.; Bell, L. H.; Maccallum, P. H. *J. Mol. Biol.* **1992**, *228*, 725–734. (b) Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472–6484.
- [20] Perczel, A.; Angyán, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, *113*, 6256–6265.
- [21] (a) Hudáky, I.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2003**, *630*, 135–140. (b) Hudáky, I.; Baldoni, H. A.; Perczel, A. *J. Mol. Struct. (THEOCHEM)* **2002**, *582*, 233–249.
- [22] (a) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3094. (b) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094. (c) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239–245. (d) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117–129.
- [23] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- [24] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- [25] Seabra, G. M.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E. *J. Phys. Chem. A* **2007**, *111*, 5655–5664.
- [26] (a) Torras, J.; Seabra, G. d. M.; Deumens, E.; Trickey, S. B.; Roitberg, A. E. *J. Comput. Chem.* **2008**, *29*, 1564–1573. (b) Torras, J.; He, Y.; Cao, C.; Muralidharan,

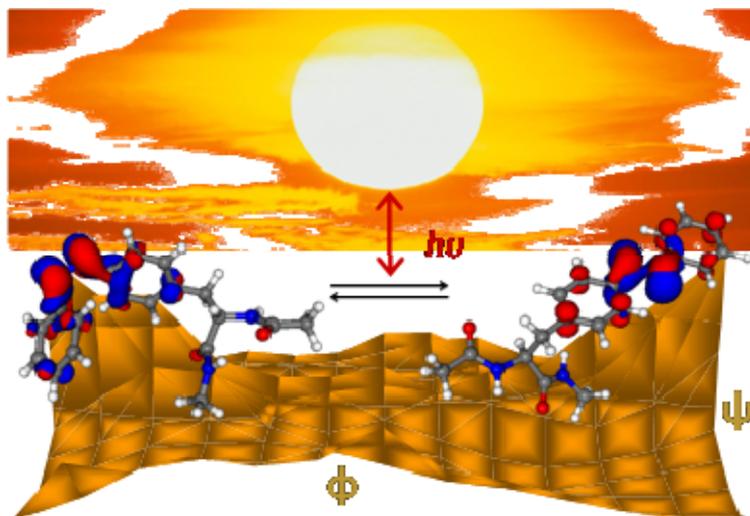
K.; Deumens, E.; Cheng, H.-P.; Trickey, S. B. *Comput. Phys. Commun.* **2007**, *177*, 265–279.

[27] (a) Iribarren, J. I.; Casanovas, J.; Zanuy, D.; Alemán, C. *Chem. Phys.* **2004**, *302*, 77–83. (b) Jang, Y. H.; Goddard III, W. A.; Noyes, K. T.; Sowers, L. C.; Hwang, S.; Chung, D. S. *J. Phys. Chem. B* **2002**, *107*, 344–357. (c) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys. B* **1998**, *102*, 3257–3271.

[28] (a) 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*, 6609–6614. (b) Delaney, N. G.; Madison, V. *J. Am. Chem. Soc.* **1982**, *104*, 6635–6641.

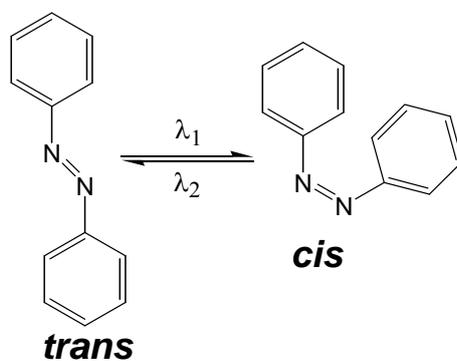
[29] Das, G. P.; Dudis, D. S. *J. Phys. Chem. A* **2000**, *104*, 4767–4771.

4.4. Phenylazophenylalanine: a key building block of photoresponsive biomimetic systems



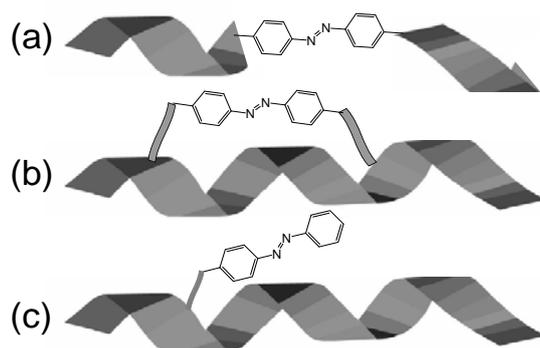
4.4.1. Introduction

A number of photoresponsive peptides and polypeptides, reactive to light to give reversible variations of their structure and conformation, have been designed and studied in the last decades.¹⁻⁹ Such reactions are accompanied by changes of physical properties, which can be exploited to develop potential applications, *e.g.* photomechanical biotransducers and actuators, or regulators in biological processes.



Scheme 1

Photoresponsive peptides are obtained by introducing photoactive ligands (chromophores) on the peptide chain, through chemical modification. Azobenzene (AB), which undergoes a reversible *cis-trans* photochemical isomerization (Scheme 1), is often considered as an optimal photochrome.¹⁰ Irradiation at $\lambda_1 \approx 320\text{-}340\text{ nm}$ converts the thermodynamically most stable *trans* isomer to the *cis* isomer. The latter reverts thermally, or upon irradiation at $\lambda_2 \geq 420\text{ nm}$.¹¹ Because the structure and dipole moments of the two isomers strongly differ, the chemical incorporation of AB at strategic positions can be used to reversibly modulate not only the peptide conformation but also the binding affinity, thereby creating semisynthetic biomaterials whose activity can be controlled photochemically. Three effective strategies in terms of conformational control have been devised to incorporate the AB chromophore: (i) the peptide backbone approach, in which the chromophore is inserted into the backbone (Scheme 2a); (ii) the side-chain cyclization approach, in which non-consecutive residues are cross-linked through their side chains with suitable AB derivatives (Scheme 2b); and (iii) the selective incorporation of unnatural amino acids that contains the chromophore in the side chain (Scheme 2c).



Scheme 2

Amongst AB-containing unnatural amino acids, phenylazophenylalanine, hereafter abbreviated PAP (Figure 4.4.1.) is particularly attractive due to its simplicity, the side chain chromophore being separated from the backbone by a single methylene unit. PAP has been successfully used to photomodulate the binding affinity and activity of peptides and proteins.^{3,12-17} Whilst AB and several of its derivatives have been the focus of extensive experimental¹⁸⁻²² and theoretical²³⁻²⁹ investigations,

studies aiming at understanding the structural and optical properties of AB-containing unnatural amino acids remain extremely limited. Thus, although the relationships between the secondary structures and the UV-Vis spectra of poly(α -L-glutamic acid) featuring AB side chain were reported,³⁰⁻³³ the intrinsic properties of its photoresponsive building blocks have yet to be tackled.

This work intends to fill this gap by presenting a comprehensive study of the *N*-acetyl-*N'*-methylamide derivative of L-PAP, hereafter denoted Ac-L-PAP-NHMe, in which the AB side group may adopt both *trans* and *cis* conformations. More specifically, Density Functional Theory (DFT) calculations have been used to provide a complete characterization of the *trans* and *cis* isomers, abbreviated Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe, respectively. Calculations have allowed to ascertain not only the influence of the isomerization of the AB core on the φ, ψ backbone dihedral angles, but also the impact of the latter on the conformation of the chromophore, which is defined by the dihedral angle θ (Figure 4.4.1.). For this purpose, the potential energy surface $E=E(\varphi, \psi)$, *i.e.* the φ, ψ -Ramachandran map, and the minimum energy conformations have been determined for both Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe. Furthermore, the changes of dipole moment and electronic transition wavelength induced by the AB isomerization (*i.e.* the $\mu=\mu(\varphi, \psi)$ and $\lambda=\lambda(\varphi, \psi)$ maps, respectively) have been examined, to quantify the influence of the molecular conformation on the electrostatic features of the system. We wish to emphasize that, to the best of our knowledge, the interplay between the spectral properties of the photoresponsive amino acid and the backbone conformation has never been systematically investigated before. This is performed for the first time, by using Time-Dependent Density Functional Theory (TD-DFT).³⁴

4.4.2. Methods

Nomenclature. The minimum energy conformations of the two isomers of Ac-PAP-NHMe have been denoted using three labels. The first one corresponds to the *trans* (t) or *cis* (c) disposition of the AB chromophore and refer to the Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe isomers, respectively. The second label identifies the backbone conformation, which is defined by the φ, ψ dihedral angles, using the nomenclature introduced by Perczel *et al*³⁵ more than 15 years ago. Finally, the third label indicates the conformation of the dihedral angles χ^1 and χ^2 , that is, the *trans* (t), *gauche*⁺ (g^+), *gauche*⁻ (g^-), *skew*⁺ (s^+), *skew*⁻ (s^-) and *cis* (c).

Geometry optimizations. All calculations were performed using the Gaussian 03³⁶ computer program. All geometry optimizations were performed in gas phase at the B3LYP/6-311++G(d,p) level.^{37,38} A complete exploration of the potential energy surface $E(\varphi, \psi)$ was performed by mapping the Ramachandran plot for both *trans* ($\theta \approx 180^\circ$) and *cis* ($\theta \approx 0^\circ$) isomers of the L-Pap dipeptide. Calculations were performed on a grid of points on the (φ, ψ) space with 30° intervals, the dihedral angles ω_0 , ω , χ^1 and χ^2 being initially positioned at 180° in all cases. At each point of the grid, the geometry was optimized by keeping the dihedral angles φ and ψ constrained during the minimization process. Possible energy minima on the surface were investigated for all low-energy regions of the *trans* and *cis* maps by performing fully-relaxed geometry optimization. Thus, starting points for complete geometry optimization were generated by combining selected low energy backbone arrangements for the *trans* and *cis* isomers with different dispositions of the dihedral angles χ^1 and χ^2 . Frequency analyses were carried out to verify the nature of the minimum state of all the obtained stationary points, and to determine the zero-point vibrational energies (ZPVE) as well as the thermal and the entropic corrections. These statistical terms were used to compute the ΔG^{gp} values at 298 K.

Calculations in Aqueous Solution. To obtain an estimation of the hydration effects on the relative stability of the different minima, single point calculations were conducted on the optimized structures with a SCRF model. Specifically, the Polarizable Continuum Model (PCM) developed by Tomasi and co-workers³⁹ was used in this work. The PCM model represents the polarization of the liquid by a charge density appearing on the surface of the molecular-shaped cavity created in the solvent. PCM calculations were performed with B3LYP/6-311++G(d,p) method using the standard protocol and considering the dielectric constant of water ($\epsilon = 78.4$) to obtain the free energy of solvation (ΔG_{solv}) of the minimum energy conformations. The conformational free energies in solution ($\Delta G^{\text{H}_2\text{O}}$) were estimated at the same level by using the classical thermodynamics scheme: $\Delta G^{\text{H}_2\text{O}} = \Delta G^{\text{gp}} + \Delta G_{\text{solv}}$.

TD-DFT Calculations. Electronic transition were evaluated by determining the vertical electronic energies by means of TD-DFT calculations.^[34] Using the 6-311+G(d,p) basis set,³⁸ the following functionals were tested for these calculations:

VSXC⁴⁰, B3LYP³⁷, PBE0⁴¹, BMK⁴² and BHandHLYP³⁷. Molecular geometries used for TD-DFT calculations were derived from both partial and complete geometry optimizations at the B3LYP/6-311+G(d,p) level (*i.e.* geometries calculated to construct the $E=E(\varphi,\psi)$ maps and minimum energy conformations, respectively).

4.4.3. Results and discussion

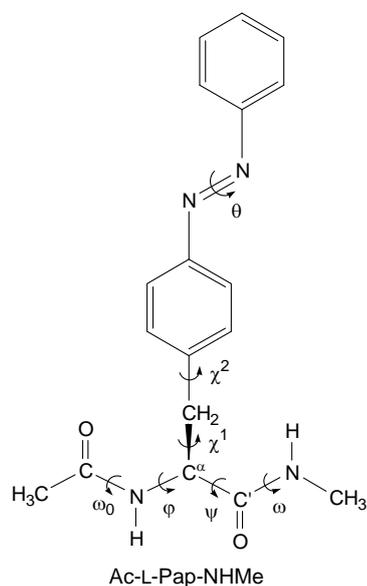


Figure 4.4.1. Structure of the compound investigated, Ac-PAP-NHMe. The backbone and side chain dihedral angles are indicated.

Figure 4.4.1. presents the nomenclature used for both the side chain and backbone dihedral angles of Ac-L-PAP-NHMe, the dihedral angle θ being used to define the *trans* and *cis* isomers of this dipeptide (*i.e.* $\theta \approx 180^\circ$ and 0° for Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe, respectively). The Ramachandran maps $E=E(\varphi, \psi)$ of the two isomers of Ac-L-Pap-NHMe were calculated at the B3LYP/6-311++G(d,p) level by keeping the backbone dihedral angles φ, ψ constrained during the minimization process. In all cases the side chain dihedral angles χ^1 and χ^2 and the two peptide bonds (ω_0 and ω) were initially positioned at 180° , even although they were allowed to relax. The $E=E(\varphi, \psi)$ maps obtained for the dipeptide with the AB arranged in *trans* and *cis* are displayed in Figures 4.4.2.a and 4.4.2.b, respectively, whereas Figure 4.4.2.c depicts the difference between the two surfaces (*i.e.* $\Delta E=E[\varphi, \psi; \text{Ac-L-(t)PAP-NHMe}] - E[\varphi, \psi; \text{Ac-L-(c)PAP-NHMe}]$).

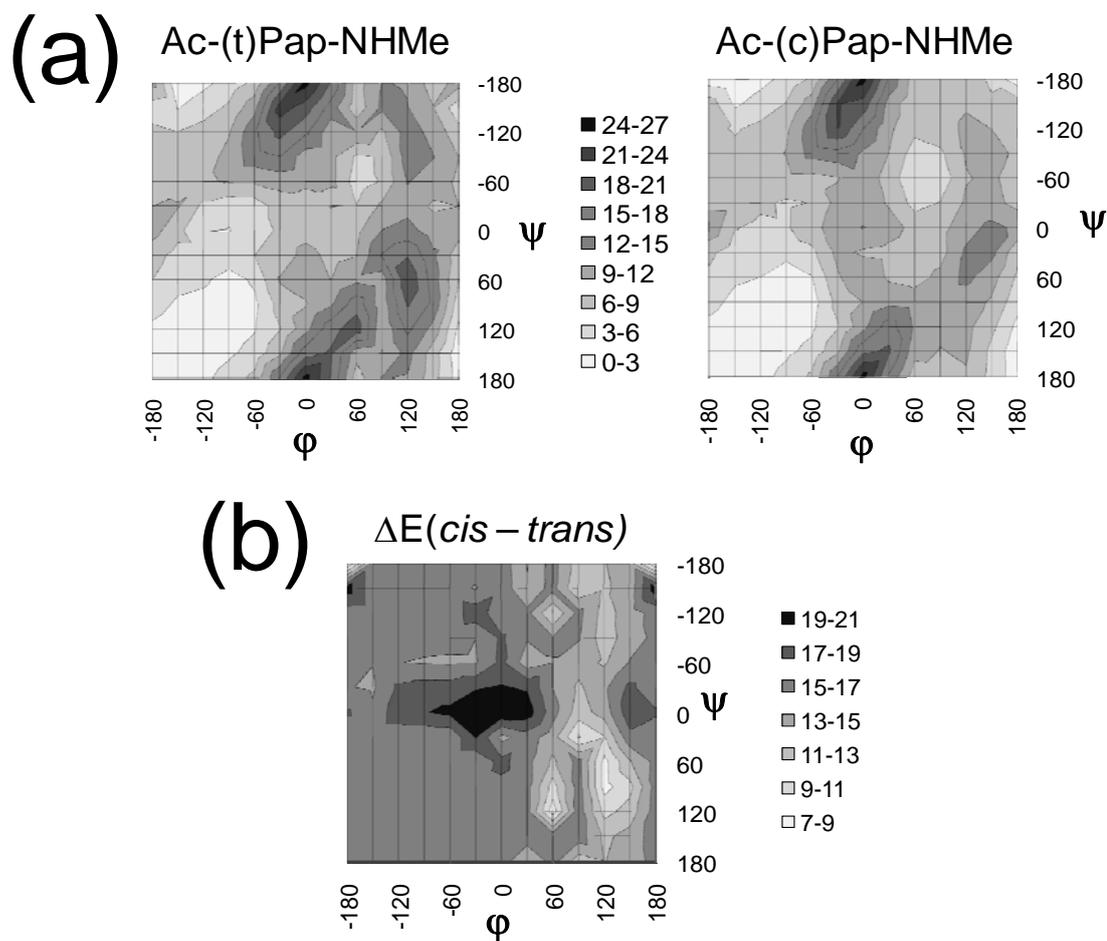


Figure 4.4.2. $E=E(\varphi, \psi)$ maps calculated at the B3LYP/6-311++G(d,p) level for the Ac-PAP-NHMe dipeptide with the AB arranged in (a) *trans* and (b) *cis*, i.e. Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe isomers, respectively. The difference between the two surfaces, $E[\varphi, \psi; \text{Ac-L-(t)PAP-NHMe}] - E[\varphi, \psi; \text{Ac-L-(c)PAP-NHMe}]$, is represented in (c).

As can be seen, the regions corresponding to relative energies within $6 \text{ kcal} \cdot \text{mol}^{-1}$ of the global minimum are closely confined for both *trans* and *cis* conformations. Furthermore, it is noticeable that the position of the energetically accessible regions in the $E=E(\varphi, \psi)$ map are not significantly influenced by the *trans* or *cis* arrangement of

the AB chromophore, even though the θ value affects the regions of higher energy. As a consequence, in both cases the regions of lower energy are located around the semi-extended conformations with $\varphi, \psi \approx -120^\circ, 120^\circ$, and $\varphi, \psi \approx 65^\circ, -60^\circ$. These values are typically found in the β -sheet and the γ -turn motifs, respectively. In opposition, the helical conformations ($\varphi, \psi \approx \pm 60^\circ, \pm 60^\circ$) are destabilized by at least 4 kcal/mol.

The $\Delta E = \Delta E(\varphi, \psi)$ map indicates that the Ac-L-(t)PAP-NHMe isomer is favored with respect to Ac-L-(c)PAP-NHMe in the whole potential energy surface, the lowest energy difference, at $\varphi, \psi \approx 120^\circ, 75^\circ$, being of approximately 7 kcal/mol. However, this energy difference ranges from 15 to 17 kcal/mol over more than a half of the map, including the regions associated to the semi-extended conformations. The analysis of the dihedral angle θ for all the structures indicates variations smaller than 1.5° , *i.e.* θ ranges from 179.4° to -179.5° and from 8.7° to 10.1° for the *trans* and *cis* isomer, respectively. These values are very similar to those predicted for the free AB (*i.e.* when the chromophore does not act as a substituent of the amino acid) at the same level of theory: 179.9° and -9.9° for the *trans* and *cis* isomer, respectively.

Starting geometries for complete optimizations were selected from the $E = E(\varphi, \psi)$ maps (the conformational search process is described in the Methods section). Full gas-phase geometry optimizations, followed by frequency calculations allowed to locate 14 and 18 minimum energy conformations for Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe, respectively, which were distributed within a relative energy interval of 7.4 and 7.7 kcal/mol. Tables 4.4.1. and 4.4.2. list the backbone and side chain dihedral angles, as well as the relative total (ΔE^{sp}) and free (ΔG^{sp}) energies of all minima calculated for *trans* and *cis* isomers, respectively. Five and seven of them present $\Delta G^{\text{sp}} \leq 1.5$ kcal/mol for the *trans* and *cis* isomer, respectively, being sketched in Figures 4.4.3. and 4.4.4.. The minimum energy conformations have been denoted using a three-label code, which specifies the AB isomerism, the arrangement of the backbone conformation, and the dihedral angles χ^1 and χ^2 (see the Methods section for details).

Table 4.4.1. Backbone and side chain dihedral angles (Figure 4.4.1.; in degrees), relative energy and free energy in the gas-phase (ΔE^{sp} and ΔG^{sp} ; in kcal/mol) and relative free energy in aqueous solution ($\Delta G^{\text{H}_2\text{O}}$; in kcal/mol) for the minimum energy conformations of Ac-L-(t)PAP-NHMe calculated at the B3LYP/6-311++G(d,p) level.

#	ω_0	φ	ψ	ω	χ_1	χ_2	θ	ΔE^{sp}	ΔG^{sp}	$\Delta G^{\text{H}_2\text{O}}$
t- β_{DL} -tg ⁺	177.6	-155.8	154.6	174.3	-169.4	71.8 / -107.8	-179.8	0.0 ^a	0.0 ^b	3.0
t- γ_{L} -g ⁻ s ⁺	-174.3	-84.5	75.0	-175.5	-57.8	113.0 / -67.1	180.0	0.4	0.6	2.8
t- γ_{L} -tg ⁺	-179.5	-83.7	82.4	-174.5	-165.0	86.0 / -93.8	-179.9	0.4	1.1	2.7
t- δ_{L} -g ⁺ g ⁺	-170.1	-120.6	16.1	173.8	55.1	82.3 / -97.3	-179.6	1.6	1.1	3.3
t- γ_{L} -g ⁺ g ⁺	-175.0	-82.6	58.7	178.5	43.0	78.8 / -101.6	-179.9	0.4	1.5	8.3
t- γ_{L} -g ⁺ g ⁺	-170.	-117.3	15.9	174.1	54.3	80.0 / -99.8	179.7	1.5	1.9	4.2
t- γ_{L} -g ⁻ s ⁺	-168.0	-106.8	6.1	174.4	-61.0	115.2 / -65.2	-179.8	2.7	2.0	0.0
t- β_{DL} -g ⁺ s ⁺	175.7	-154.2	165.5	-178.5	61.3	92.0 / -87.9	-179.9	2.2	2.1	0.6
t- γ_{D} -g ⁻ s ⁺	174.1	73.4	-54.2	-178.6	-59.6	103.6 / -75.6	180.0	1.8	2.2	3.1
t- γ_{D} -tg ⁺	173.1	73.3	-65.9	177.4	-170.2	77.4 / -102.6	180.0	3.8	4.5	3.4
t- δ_{D} -g ⁺ s ⁺	170.6	-162.3	-38.0	-174.3	62.7	97.3 / -83.0	-179.9	6.2	6.0	6.4
t- ϵ_{D} -tg ⁺	-164.5	63.8	-157.8	-176.8	-160.4	56.5 / -125.4	179.8	6.9	6.8	6.5
t- γ_{D} -g ⁺ g ⁺	173.3	56.3	-29.9	-177.8	69.6	81.0 / -101.0	179.8	7.0	7.9	8.1
t- δ_{D} -tg ⁺	174.7	-153.4	-65.0	-177.1	-174.6	71.2 / -108.7	180.0	7.4	6.6	0.8

^a E= -1067.714321 a.u. ^b G= -1067.416009 a.u.

Table 4.4.2. Backbone and side chain dihedral angles (Figure 4.4.1.; in degrees), relative energy and free energy in the gas-phase (ΔE^{gp} and ΔG^{gp} ; in kcal/mol) and relative free energy in aqueous solution ($\Delta G^{\text{H}_2\text{O}}$; in kcal/mol) for the minimum energy conformations of Ac-L-(c)PAP-NHMe calculated at the B3LYP/6-311++G(d,p) level.

#	ω_0	φ	ψ	ω	χ_1	χ_2	θ	ΔE^{gp}	ΔG^{gp}	$\Delta G^{\text{H}_2\text{O}}$
c- $\beta_{\text{DL}}-tg^+$	176.3	-157.3	165.7	176.1	-158.3	71.8 / -109.1	9.8	0.2	0.0 ^b	3.5
c- $\gamma_{\text{L}}-tg^+$	178.8	-84.0	81.3	-174.2	-164.7	89.9 / -91.0	9.2	0.0 ^a	0.5	1.6
c- $\gamma_{\text{L}}-g^+s^+$	-175.9	-84.2	75.0	-175.4	-60.0	110.6 / -70.6	9.1	0.4	0.5	1.4
c- $\beta_{\text{DL}}-g^+s^+$	174.4	-126.5	144.1	178.6	-64.8	91.1 / -87.3	8.2	1.6	0.5	0.0
c- $\gamma_{\text{L}}-g^+g^+$	-170.5	-117.1	14.8	173.6	56.0	83.4 / -97.2	9.4	1.4	1.3	4.5
c- $\beta_{\text{DL}}-g^+s^+$	175.6	-153.8	165.0	-171.3	61.9	93.6 / -86.7	9.4	1.5	1.3	0.7
c- $\gamma_{\text{L}}-g^+g^+$	-174.6	-82.8	59.5	178.9	43.5	81.9 / -99.4	9.3	0.1	1.4	7.7
c- $\gamma_{\text{D}}-g^+s^+$	174.6	73.1	-55.2	-178.7	-60.8	104.7 / -75.8	9.6	1.7	2.3	2.0
c- $\gamma_{\text{D}}-tg^+$	176.6	73.1	-64.2	178.2	-172.2	76.2 / -104.1	9.0	2.9	3.5	3.6
c- $\alpha_{\text{D}}-g^+s^+$	168.8	73.4	16.5	-177.9	-60.9	105.7 / -74.9	9.4	4.3	4.3	0.6
c- $\alpha_{\text{L}}-tg^+$	-170.9	-88.6	-19.5	176.5	-172.2	72.4 / -109.3	8.9	5.6	5.0	1.1
c- $\delta_{\text{D}}-tg^+$	174.2	-154.0	-62.3	-176.7	-174.6	74.2 / -105.7	8.9	6.5	6.2	0.9
c- $\delta_{\text{D}}-g^+s^+$	1706	-162.5	-39.2	-174.4	62.5	100.1 / -81.3	10.0	6.4	6.4	7.3
c- $\alpha_{\text{D}}-s^+s^+$	170.1	68.0	31.5	-177.9	-133.9	102.5 / -78.2	9.3	6.6	6.7	1.1
c- $\delta_{\text{D}}-g^+s^+$	173.9	-122.1	-92.8	-179.2	-67.1	94.2 / -83.9	7.5	7.6	6.8	2.3
c- $\varepsilon_{\text{D}}-tg^+$	-164.8	63.8	-156.6	-177.5	-159.2	57.4 / -125.5	9.1	6.8	7.1	5.8
c- $\gamma_{\text{D}}-g^+g^+$	171.7	54.3	-26.3	-178.4	64.3	80.9 / -100.8	8.8	6.1	8.3	9.3
c- $\alpha_{\text{D}}-g^+g^+$	168.5	51.8	37.4	-176.2	54.8	83.5 / -98.8	9.6	7.7	9.2	5.3

^a E = -1067.689001 a.u. ^b G = -1067.390678 a.u.

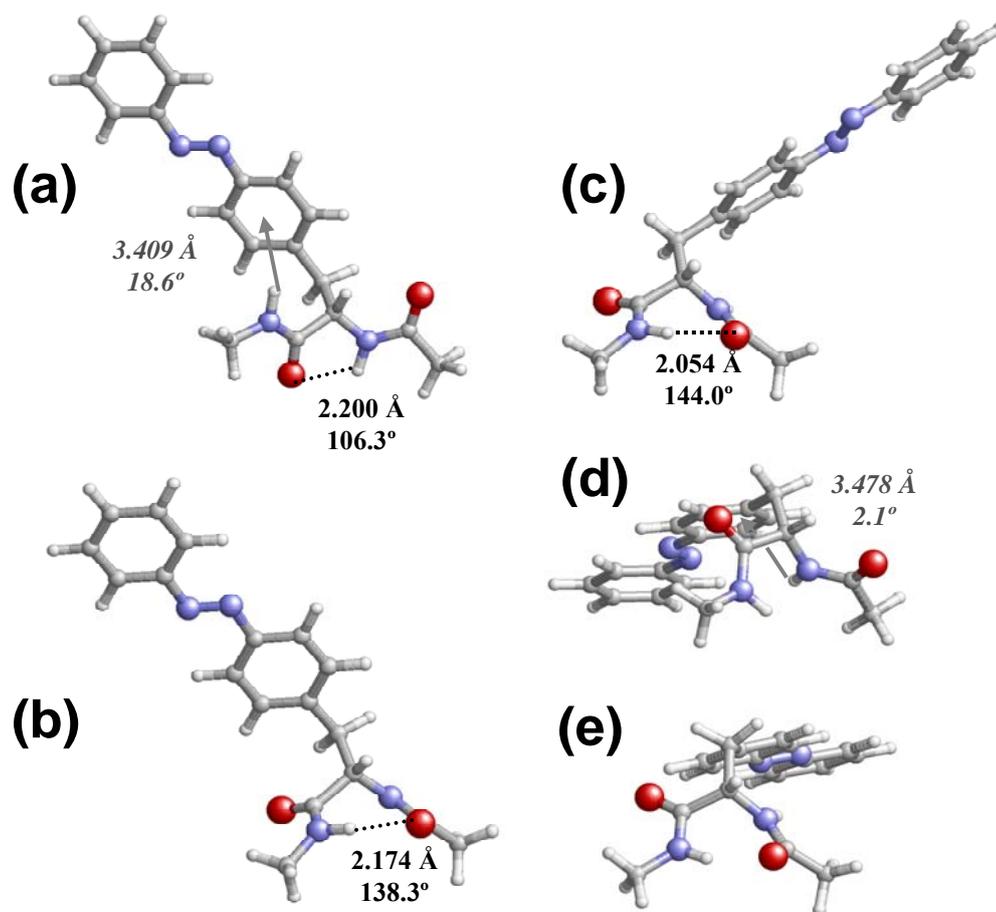


Figure 4.4.3.. Ac-L-(t)PAP-NHMe minimum energy conformations with with $\Delta G^{\text{BP}} \leq 1.5$ kcal/mol at the B3LYP/6-311++G(d,p) level: (a) t- β_{DL} - tg^+ ; (b) t- γ_{L} - g^-s^+ ; (c) t- γ_{L} - tg^+ ; (d) t- δ_{L} - g^+g^+ ; and (e) t- γ_{L} - g^+g^+ .

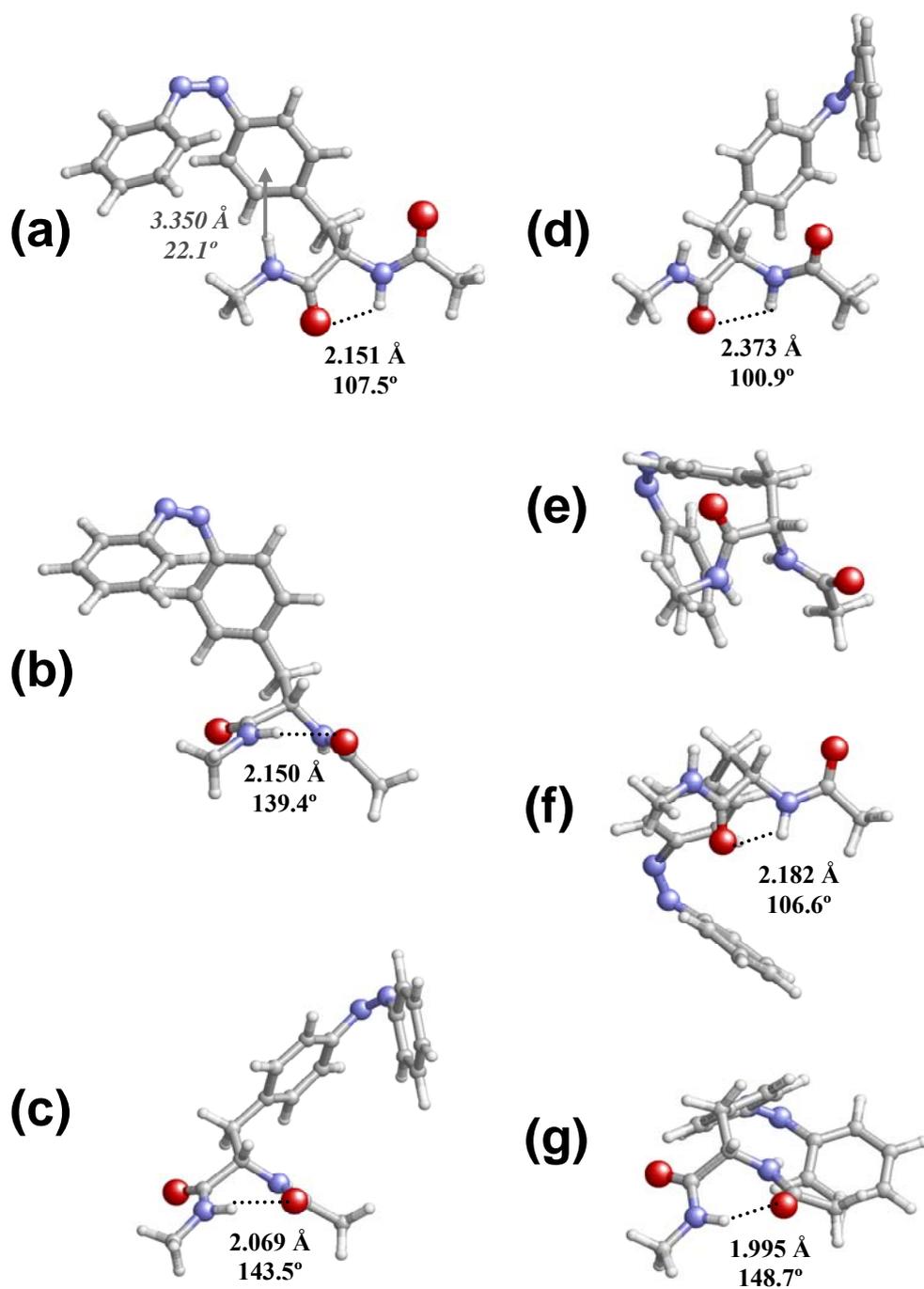


Figure 4.4.4. Ac-L-(c)PAP-NHMe minimum energy conformations with $\Delta G^{\text{BP}} \leq 1.5$ kcal/mol at the B3LYP/6-311++G(d,p) level: (a) c- $\beta_{\text{DL}}\text{-}t\text{g}^+$; (b) c- $\gamma_{\text{L}}\text{-}t\text{g}^+$; (c) c- $\gamma_{\text{L}}\text{-}g^+s^+$; (d) c- $\beta_{\text{DL}}\text{-}g^+s^+$; (e) c- $\gamma_{\text{L}}\text{-}g^+g^+$; (f) c- $\beta_{\text{DL}}\text{-}g^+s^+$; and (g) c- $\gamma_{\text{L}}\text{-}g^+g^+$. The c- $\beta_{\text{DL}}\text{-}g^+s^+$ (d) is the conformation with lowest free energy in aqueous solution.

The lowest energy minimum of Ac-L-(t)PAP-NHMe, t- β_{DL} - tg^+ (Figure 4.4.3.a), adopts a semi-extended backbone arrangement (C_5 conformation), in which the CO and NH moieties of the PAP residue defines a five-member cycle with parameters $d(H\cdots O)= 2.200 \text{ \AA}$ and $\angle N-H\cdots O= 106.3^\circ$. In this case, the orientation of the substituent leads to the formation of a weak N-H $\cdots\pi$ attractive interaction between the amino of the N-methylamide blocking group and the photochrome. This interaction may be defined by, on the one hand, the distance between the amide hydrogen and the center of the aromatic ring [$d(H\cdots Ph)= 3.409 \text{ \AA}$] and, on the other hand, the angle formed by the N-H bond and the phenyl ring plane ($\theta_{ph}= 18.6^\circ$). The ability of the π electron density to interact with proton donors has long been recognized⁴³⁻⁴³ and has recently been identified as a stabilizing factor of peptide and protein structures.⁴⁶⁻⁴⁹ The next two minima, t- γ_L - g^-s^+ and t- γ_L - tg^+ (Figures 4.4.3.b and 4.4.3.c, respectively), present a γ_L backbone conformation with the terminal acetyl CO and methylamide NH sites forming a seven-member hydrogen bonded ring. These two minima lack of stabilizing interactions of the N-H $\cdots\pi$ type and, the main differences between them is the orientation of the AB group, as well as the strength of the intramolecular hydrogen bond.

The backbone conformation of the t- δ_L - g^+g^+ (Figure 4.4.3.d) geometry is actually in between the δ_L and γ_L regions, and though no intramolecular hydrogen bond was detected in this structure, it is isoenergetic to t- γ_L - tg^+ in terms of ΔG^{gp} . Such an outcome can be attributed to the orientation of the side group that allows the formation of a weak N-H $\cdots\pi$ interaction. The t- γ_L - g^+g^+ (Figure 4.4.3.e) minimum, which shows a ΔG^{gp} value of 1.5 kcal/mol, presents both backbone \cdots backbone (hydrogen bond) and backbone \cdots side chain (N-H $\cdots\pi$) interactions. This structure, isoenergetic to both t- γ_L - g^-s^+ and t- γ_L - tg^+ ones in terms of ΔE^{gp} , is nevertheless destabilized because of the unfavorable vibrational corrections obtained for the g^+g^+ arrangement of the side group. The backbone conformations of the remaining 9 minima, in which ΔG^{gp} ranges from 1.9 to 7.9 kcal/mol, is distributed as follows: 2 γ_L , 1 β_{DL} , 2 δ_D , 3 γ_D and 1 ϵ_D . It is worth noting that the conformations in the helical (α_L and α_D) and ϵ_L regions are sterically inaccessible for Ac-L-(t)PAP-NHMe. In short, our results indicate that the *trans* disposition of AB group reduces the conformational flexibility with respect to alanine.^{50,51}

This effect is mainly due to the broadening of the gap that separates the energy minima, as well as the emergence of backbone···chromophore steric clashes in the helical structures.

Regarding to Ac-L-(c)PAP-NHMe, the most stable conformation in terms of ΔG^{gp} corresponds to the c- $\beta_{\text{DL}}\text{-}tg^+$ (Figure 4.4.4.a), that is completely similar to the most favored *trans* conformer NHMe (t- $\beta_{\text{DL}}\text{-}g^+$ in Figure 4.4.3.a), but for the θ angle. As expected, the *trans*-to-*cis* isomerization of the AB group increases the free energy by 15.9 kcal/mol. This is in clear agreement for the *trans*-to-*cis* of free AB, which shows an energy difference of 15.8 kcal/mol at the same theoretical level. The geometric parameters indicate that the strength of both the hydrogen bond and N-H··· π interactions is only slightly larger for Ac-L-(c)PAP-NHMe than for Ac-L-(t)PAP-NHMe. The next two minima, c- $\gamma_{\text{L}}\text{-}tg^+$ and c- $\gamma_{\text{L}}\text{-}g^+s^+$ (Figures 4.4.4.b and 4.4.4.c, respectively), differ in the orientation of the side group, and are both disfavored by 0.5 kcal/mol in terms of ΔG^{gp} , suggesting that the strength of the backbone···side chain interactions is similar in each case. The dihedral angles (χ_1 and χ_2) of the fourth minimum at 0.5 kcal/mol, c- $\beta_{\text{DL}}\text{-}g^+s^+$ (Figure 4.4.4.d), preclude the formation of the N-H··· π interaction found in the global minimum. Finally, the c- $\gamma_{\text{L}}\text{-}g^+g^+(a)$, c- $\beta_{\text{DL}}\text{-}g^+s^+$ and c- $\gamma_{\text{L}}\text{-}g^+g^+(b)$ conformations (Figures 4.4.4.e, 4.4.4.f and 4.4.4.g, respectively), which present ΔG^{gp} values close to 1.3-1.4 kcal/mol, do not benefit from interactions between the backbone and the side chain, as the d(H···Ph) systematically exceeds 5 Å. The backbone conformation of the remaining 11 minima, presenting ΔG^{gp} ranging from 2.3 to 9.2 kcal/mol, is distributed as follows: 3 γ_{D} , 3 α_{D} , 1 α_{L} , 3 δ_{D} and 1 ε_{D} . As it can be seen, both the α_{L} and α_{D} were identified as energy minima for the Ac-L-(c)PAP-NHMe isomer, indicating that the helical arrangement is only forbidden when the chromophore adopts a *trans* disposition.

The values of the relative free energy in aqueous solution ($\Delta G^{\text{H}_2\text{O}}$), obtained with the PCM method, of all the minima calculated for Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe are also listed in Tables 4.4.1. and 4.4.2., respectively. The relative free energy intervals are slightly larger than those found in the gas-phase, reproducing the tendency recently pointed out for dipeptides made of amino acids bearing one or two phenyl side groups.⁵²⁻⁵⁴ Solvation induces important changes in the energy ranking of the conformers. Specifically, the t- $\gamma_{\text{L}}\text{-}g^+s^+$ (Figure 4.4.5.a), t- $\beta_{\text{DL}}\text{-}g^+s^+$ (Figure 4.4.5.b) and t- $\delta_{\text{D}}\text{-}tg^+$ (Figure 4.4.5.c), which are disfavored in the gas-phase by

2.0, 2.1 and 6.6 kcal/mol, respectively, become the most stabilized conformers in water, being the only three structures of Ac-L-(t)PAP-NHMe with $\Delta G^{\text{H}_2\text{O}} \leq 1.5$ kcal/mol. The most remarkable characteristics of these three conformers are: 1) the NH and CO moieties of the polar amide groups are highly exposed, allowing attractive electrostatic interactions with water; and 2) the accessibility of the hydrophobic side group is partially hindered by the backbone. This is especially true for the most stable geometry since its backbone dihedral angles are in the border between the γ_{L} and δ_{L} regions, therefore precluding the formation of the seven-member hydrogen bonded ring.

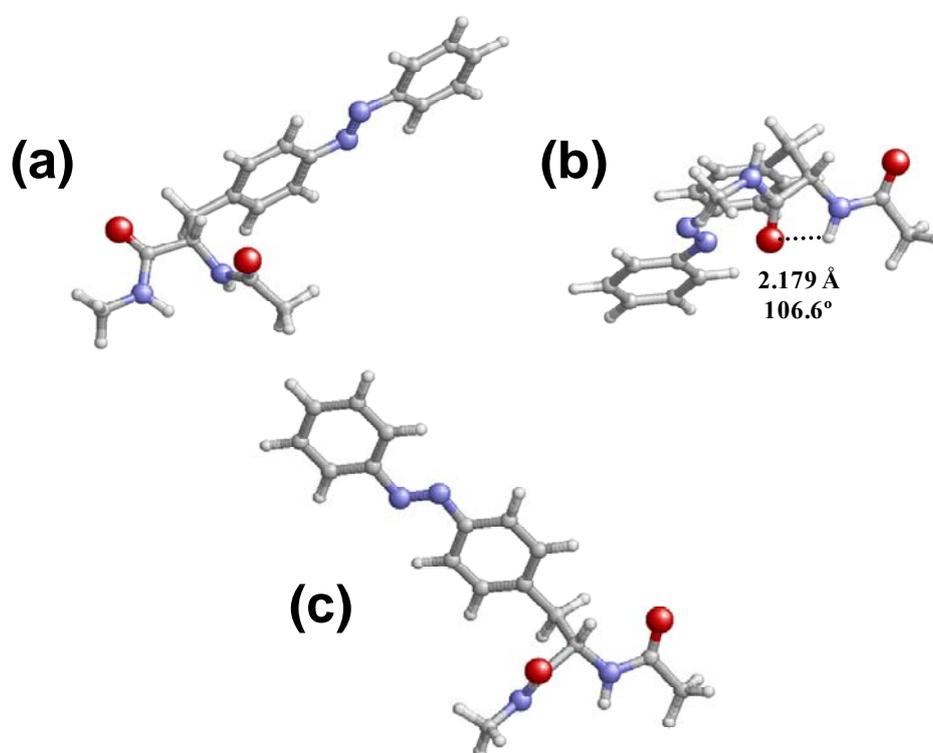


Figure 4.4.5.. Conformations of Ac-L-(t)PAP-NHMe with $\Delta G^{\text{H}_2\text{O}} \leq 1.5$ kcal/mol calculated using the PCM model: (a) t- γ_{L} -g⁻s⁺; (b) t- β_{DL} -g⁺s⁺; and (c) t- δ_{D} -tg⁺.

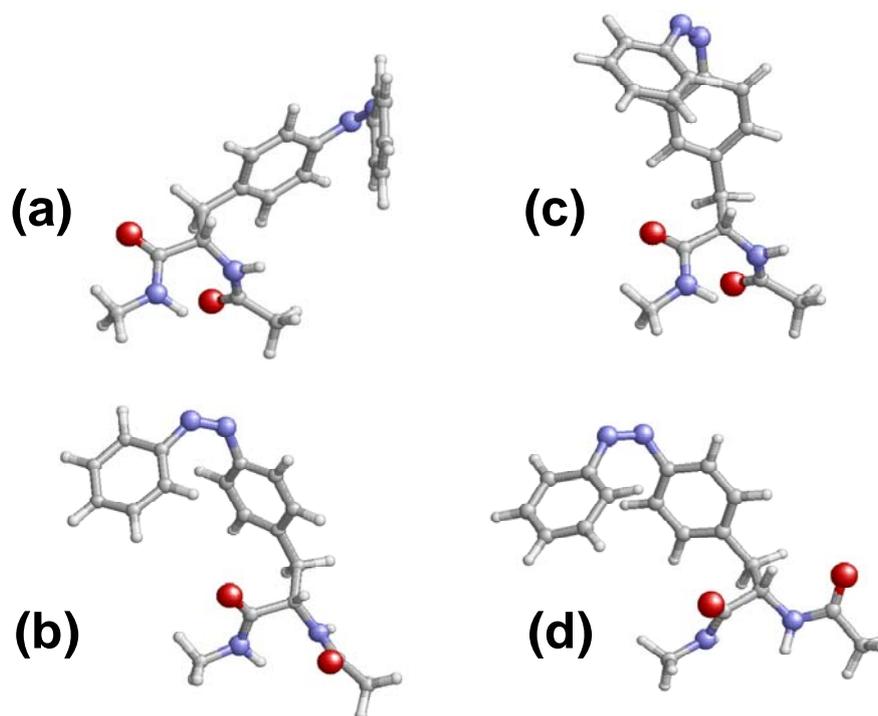


Figure 4.4.6.. Selected conformations of Ac-L-(c)PAP-NHMe with $\Delta G^{\text{H}_2\text{O}} \leq 1.5$ kcal/mol calculated using the PCM model: (a) $c\text{-}\alpha_{\text{D}}\text{-g}^-s^+$; (b) $c\text{-}\alpha_{\text{L}}\text{-tg}^+$; (c) $c\text{-}\alpha_{\text{D}}\text{-s}^+s^+$; and (d) $c\text{-}\delta_{\text{D}}\text{-tg}^+$.

Regarding to Ac-L-(c)PAP-NHMe, seven conformers with ΔG^{BP} ranging from 0.5 to 6.7 kcal/mol show $\Delta G^{\text{H}_2\text{O}} \leq 1.5$ kcal/mol. The larger conformational flexibility of the *cis* isomer with respect to its *trans* counterpart can be mainly explained by the lower accessibility of the aromatic rings in the *cis* conformation, therefore reducing the repulsive interactions between side group and the solvent. This point is clearly evidenced in Figure 4.4.4.d, which shows that the two internal sides of the aromatic rings are protected from the solvent in the most stable conformer ($c\text{-}\beta_{\text{DL}}\text{-g}^-s^+$) in aqueous solution. For the other low energy conformations in water environment, the helical arrangements, namely $c\text{-}\alpha_{\text{D}}\text{-g}^-s^+$ (Figure 4.4.6.a), $c\text{-}\alpha_{\text{L}}\text{-tg}^+$ (Figure 4.4.6.b) and $c\text{-}\alpha_{\text{D}}\text{-s}^+s^+$ (Figure 4.4.6.c) with respective $\Delta G^{\text{H}_2\text{O}}$ of 0.6, 1.1 and 1.1 kcal/mol, deserve special attention as they are significantly unfavoured in the gas-phase. Previous theoretical studies on model dipeptides of both proteinogenic⁵⁵ and synthetic^{56,57}

amino acids indicated that, in general, helical conformations are significantly stabilized in aqueous solution due to a relatively high dipole moment that yields attractive electrostatic interactions. Our present results confirm that trend. The three remaining low energy conformations correspond to the $c\text{-}\beta_{\text{DL}}\text{-}g^+s^+$ (Figure 4.4.4.f), $c\text{-}\delta_{\text{D}}\text{-}tg^+$ (Figure 4.4.6.d) and $c\text{-}\gamma_{\text{L}}\text{-}g^-s^+$ (Figure 4.4.4.c) with $\Delta G^{\text{H}_2\text{O}} = 0.7, 0.9$ and 1.4 kcal/mol, respectively. The stability of the former and the latter are relatively similar in gas-phase ($\Delta G^{\text{SP}} = 1.3$ and 0.5 kcal/mol, respectively) and water, while the $c\text{-}\delta_{\text{D}}\text{-}tg^+$ conformation gains 5.3 kcal/mol.

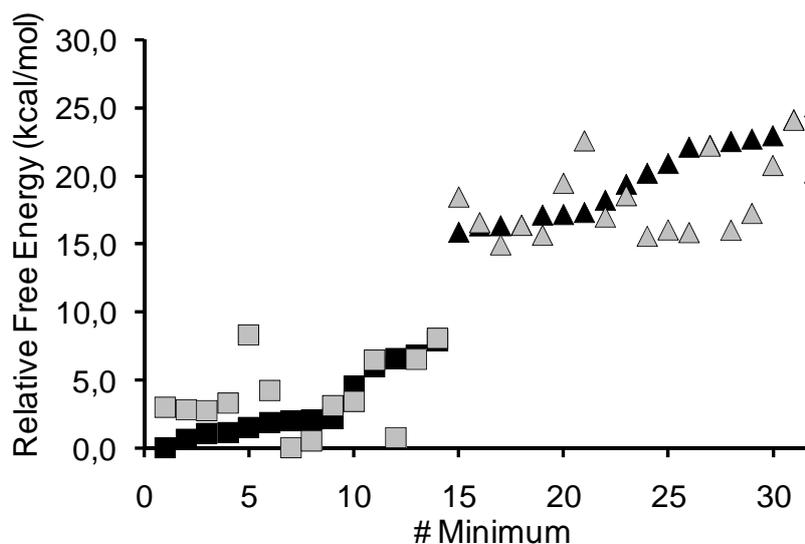


Figure 4.4.7. Free energies in the gas-phase (black symbols) and in aqueous solution (gray symbols) of the minimum energy conformations obtained for the *trans* (squares) and *cis* (triangles) stereoisomers of Ac-PAP-NHMe relative to the structure of lowest energy independently of the disposition of the AB group, *i.e.* $c\text{-}\beta_{\text{DL}}\text{-}tg^+$ and $t\text{-}\gamma_{\text{L}}\text{-}g^-s^+$ in the gas-phase and aqueous solution, respectively. The 32 conformers have been ordered following the increasing relative free energy in the gas-phase.

Figure 4.4.7. compares the relative free energy in the gas phase and water of the driven out minimal energy conformers, the global minimum of the *trans* Ac-L-PAP-NHMe being used as benchmark. As previously mentioned, the solvent induces significant modifications of the relative energy order of the minima calculated for both Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe. However, the energy gap separating the *trans* and *cis* geometries remains almost constant. Indeed, the free

energy difference between the most stable conformation of each isomer is 15.9 kcal/mol (gas-phase) and 14.9 kcal/mol (water). The relative stability of the two isomers is essentially independent of the external forces, even though the conformational preferences of the backbone are tuned by the medium.

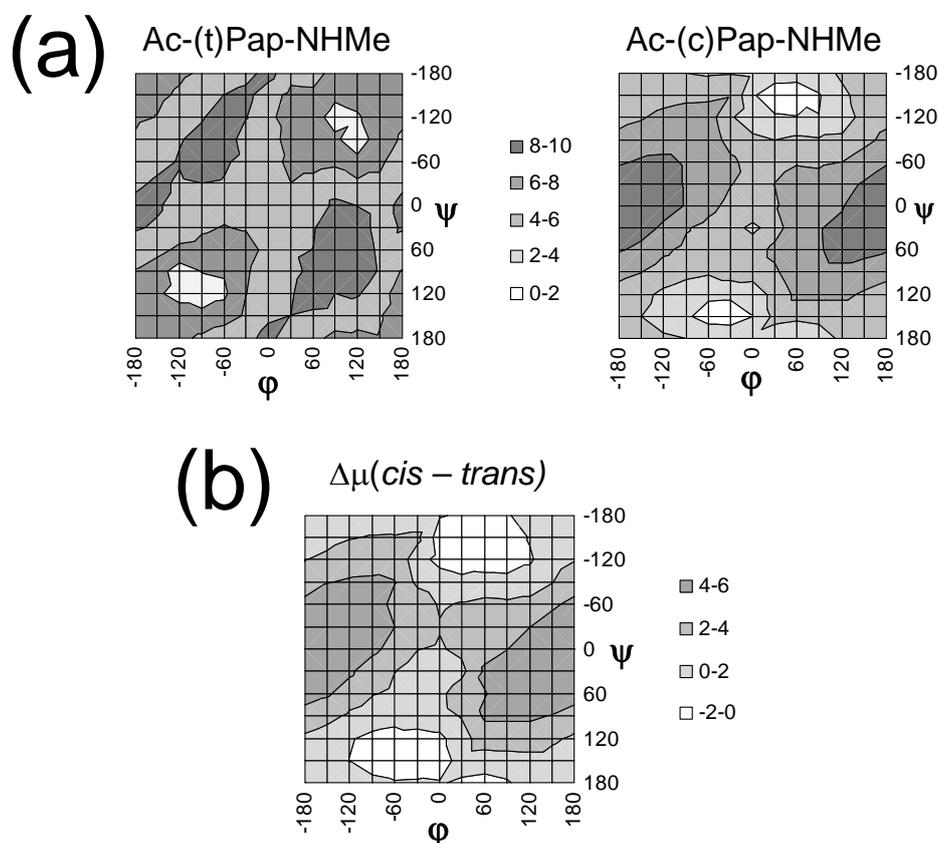


Figure 4.4.8. $\mu=\mu(\phi,\psi)$ (in Debyes) maps calculated at the B3LYP/6-311++G(d,p) level for the Ac-PAP-NHMe dipeptide with the AB arranged in (a) *trans* and (b) *cis*, i.e. Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe isomers, respectively. The difference between the maps calculated for the two isomers, $\mu[\phi,\psi;\text{Ac-L-(t)PAP-NHMe}] - \mu[\phi,\psi;\text{Ac-L-(c)PAP-NHMe}]$, is represented in (c)

Figure 4.4.8. compares the $\mu=\mu(\phi,\psi)$ maps calculated for Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe. The latter isomer presents the lowest [highest] dipole value in the semi-extended ($\phi,\psi\approx 120^\circ,-120^\circ$) [helical ($\phi,\psi\approx \pm 60^\circ, \pm 60^\circ$)] region, while the

polyproline II ($\varphi, \psi \approx \pm 60^\circ, 150^\circ$) [semi-folded ($\varphi, \psi \approx \pm 180^\circ, 0^\circ$)] region displays the lowest [highest] dipole when the chromophore is arranged in *trans*. The $\Delta\mu = \Delta\mu(\varphi, \psi)$ (Figure 4.4.8.c), clearly describes the impact of AB photoisomerization on the μ of Ac-L-PAP-NHMe, that can be as large as ~ 5 Debyes. These results allow to conclude that when the chromophore-containing amino acid is placed in close proximity to a substrate or ligand binding site in an enzyme, receptor, or ion channel, the variation in the μ accompanying the geometric changes can be used to regulate and control the binding affinity and, consequently, the activity of peptides and proteins.

The complete $\lambda = \lambda(\varphi, \psi)$ absorption maps for Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe have been determined using TD-DFT calculations,³⁴ which are known to provide accurate results for a reasonable computational effort, if a suitable functional is selected.⁵⁸⁻⁴⁰ In order to select the most appropriate TD-DFT level, calculations were first conducted on five minima of each isomer, which included the structures of lowest and highest energy as well as three additional minima separated by approximately regular ΔE^{sp} intervals. Specifically, we collated the UV-vis performance of five typical functionals (see Method Section) combined with the 6-311+G(d,p) basis set, which was reported to be adequate for the AB and its derivatives.^{29,64} The wavelengths calculated for the first $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions (hereafter denoted λ^{np} and λ^{pp}), that are relevant for isomerization of the AB, were compared with the experimental data determined for the *trans* and *cis* isomer of free AB.⁶⁵ The closest agreement between measurements and simulations was provided by the BMK functional, which was consequently selected to compute the λ^{np} and λ^{pp} of all the optimized geometries used to construct the $E = E(\varphi, \psi)$ maps.

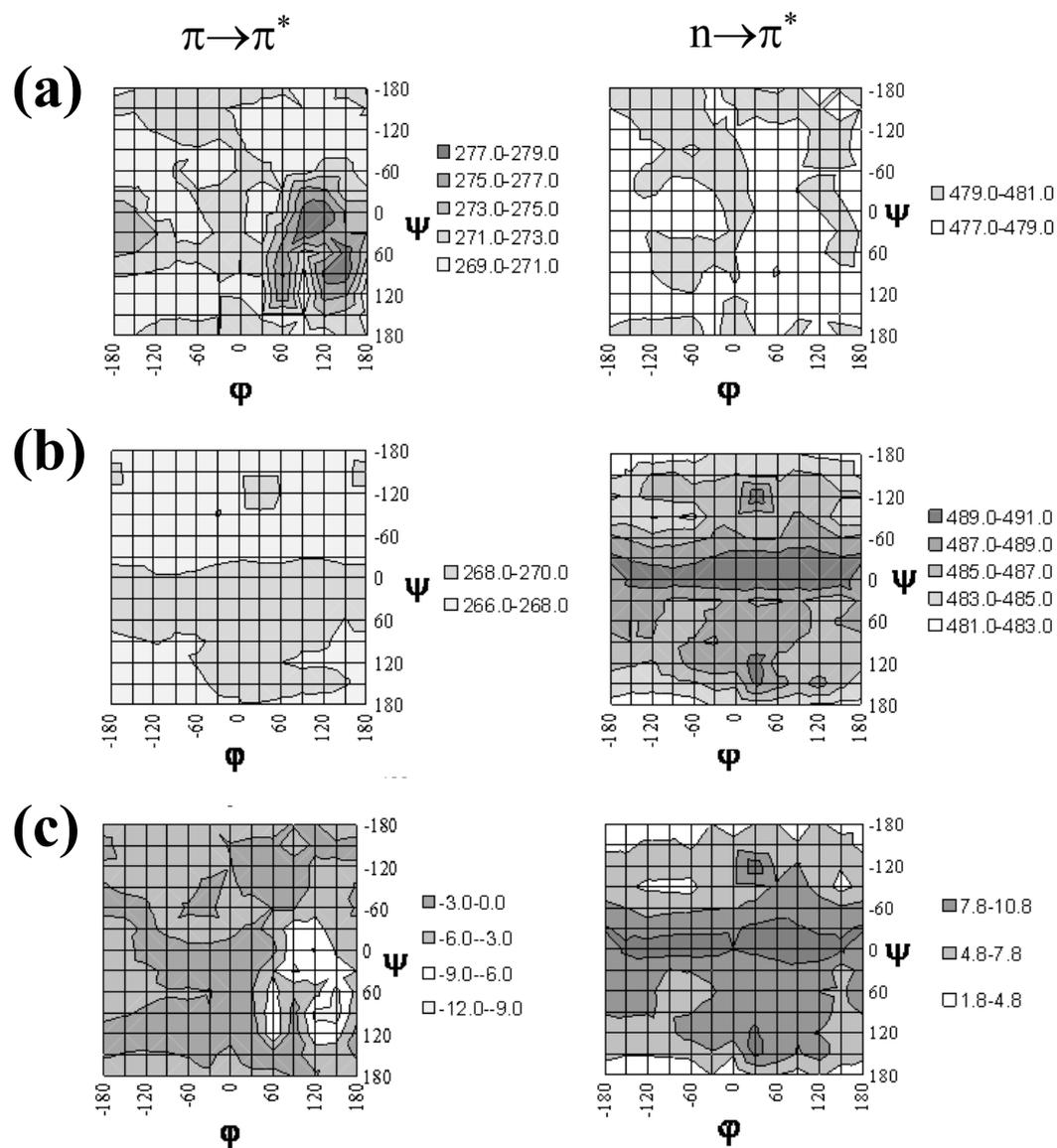


Figure 4.4.9. $\lambda^{np} = \lambda^{np}(\phi, \psi)$ and $\lambda^{pp} = \lambda^{pp}(\phi, \psi)$ maps ($n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions, respectively) calculated at the TD-BMK/6-311+G(d,p) level for the Ac-PAP-NHMe dipeptide with the AB arranged in (a) *trans* and (b) *cis*. The difference between the maps calculated for the two isomers, is represented in (c). Units in nm.

Figure 4.4.9. shows the $\lambda=\lambda(\varphi,\psi)$ absorption maps calculated for Ac-L-(c)PAP-NHMe and Ac-L-(t)PAP-NHMe, respectively. It turns out that, in all cases, the variations of the transition wavelengths with the backbone dihedral angles φ,ψ are relatively limited, independently of the *trans* or *cis* arrangement of the AB side group. More specifically, the largest variation, which is as small as 10 nm, occurs in the $\lambda^{np}(\varphi,\psi)$ and $\lambda^{pp}(\varphi,\psi)$ maps of Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe, respectively, while the variations in the other two computed maps are smaller than 4 nm. The λ^{np} and λ^{pp} values for the two Ac-L-PAP-NHMe isomers were determined considering a Boltzmann distribution of the calculated peptide backbone conformations (*i.e.* using the populations derived from the $E=E(\varphi,\psi)$ potential energy hypersurfaces to weight the $\lambda=\lambda(\varphi,\psi)$ values). The resulting values are $\lambda^{np}= 479$ nm and $\lambda^{pp}= 270$ nm for Ac-L-(t)PAP-NHMe, and $\lambda^{np}= 484$ nm and $\lambda^{pp}= 267$ nm for Ac-L-(c)PAP-NHMe. Thus, small shifts of $\Delta \lambda^{np}= +5$ nm (red shift) and $\Delta \lambda^{pp}= -3$ nm (blue shift) are predicted for the *trans* to *cis* isomerization of Ac-L-PAP-NHMe. These shifts are practically identical to those found for free AB calculated in the gas-phase at the same theoretical level: $\Delta \lambda^{np}= +4$ nm and $\Delta \lambda^{pp}= -3$ nm. These results are consistent with those calculated for poly(glutamic acid) featuring a photochromic side chain.³³ Specifically, TD-DFT calculations predicted bathochromic and hypsochromic shifts for the $n\rightarrow\pi^*$ and $\pi\rightarrow\pi^*$ transitions, respectively, when the dye changes from *trans* to *cis*, even though $\Delta \lambda^{np}$ and $\Delta \lambda^{pp}$ changed very few nm with the conformation of the polypeptide.

In order to ascertain the influence of the peptide backbone on the photochromic behavior of the AB moiety, excitation energies have been compared with those calculated for free AB: $\lambda^{np}= 480$ nm and $\lambda^{pp}= 269$ nm for the *trans* isomer, and $\lambda^{np}= 484$ nm and $\lambda^{pp}= 266$ nm for the *cis* one). Bathochromic shifts of around 0.5 nm are predicted for the $n\rightarrow\pi^*$ transition of both the *trans* and *cis* isomers, while the $\pi\rightarrow\pi^*$ transition shows hypsochromic shifts of around 1 nm. These small changes indicate that the incorporation of the AB chromophore to a peptide does not induce relevant changes neither in λ^{np} nor in λ^{pp} values, the spectra of free AB being preserved.

On the other hand, the value of the oscillator strength associated to the $n\rightarrow\pi^*$ transition is lower for the *trans* isomer than for the *cis* one in the whole φ,ψ map (*i.e.* it varies from $\sim 1\times 10^{-4}$ to 8×10^{-4} and from 0.025 to 0.030 for the *trans* and *cis* isomer, respectively). The average value weighted according to a Boltzmann distribution of

conformations is 2×10^{-4} and 0.035 for the *trans* and *cis* isomers, respectively. For the $\pi \rightarrow \pi^*$ transition the values of oscillator strength range from 0.013 to 0.037 for the *trans* isomer and from 0.025 to 0.070 for the *cis* one, while the weighted averages are 0.024 and 0.052, respectively.

4.4.5. Conclusions

The conformational and optical properties of Ac-L-(c)PAP-NHMe and Ac-L-(t)PAP-NHMe have been studied using DFT and TD-DFT calculations, respectively. The calculated $E=E(\varphi, \psi)$ maps indicate that the regions of lower energy correspond to those typically associated to the β -sheet and γ -turn secondary structures for the two isomers. In spite of this resemblance, the Ac-L-(t)PAP-NHMe isomer is favored with respect to Ac-L-(c)PAP-NHMe in the whole map, the energy difference between the two isomers ranging from 7 to 17 kcal/mol. A conformational search process led to identify 14 and 18 minimum energy conformations for Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe, respectively, even though only 5 and 7 of such minima showed ΔG^{gp} values lower than 1.5 kcal/mol. The lowest energy minimum found for the two isomers in the gas-phase corresponds to a semi-extended conformation stabilized by: an intramolecular N-H \cdots O=C hydrogen bond and a N-H \cdots π interaction. Thus, the only difference involves the dihedral angle θ of the AB side group, which increases the free energy 15.9 kcal/mol when changes from *trans* to *cis*. On the other hand, we found that helical conformations are sterically forbidden for Ac-L-(t)PAP-NHMe but not for Ac-L-(c)PAP-NHMe, even though they are destabilized by at least 4.3 kcal/mol in the gas-phase.

Aqueous solvent produced important changes in the conformational preferences of Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe, the number of conformations with $\Delta G^{\text{H}_2\text{O}}$ values lower than 1.5 kcal/mol being 3 and 7, respectively. Solvation stabilizes conformations in which the accessibility of the polar amide groups and the aromatic rings of the AB side group is maximum and minimum, respectively. The lowest energy conformation in aqueous solution is the t- γ_L - g^-s^+ and c- β_{DL} - g^-s^+ for the *trans* and *cis* isomers, respectively, which are disfavored by 2.0 and 0.5 in the gas-phase. Interestingly, the relative stability of all the helical conformations identified for the *cis* isomer is significantly enhanced in solution.

The calculated $\lambda^{\text{pp}}=\lambda^{\text{pp}}(\varphi, \psi)$ and $\lambda^{\text{pp}}=\lambda^{\text{pp}}(\varphi, \psi)$ absorption maps indicate that the variations of the transition wavelengths with the backbone dihedral angles φ, ψ is

reduced (*i.e.* smaller than 10 nm) for both Ac-L-(t)PAP-NHMe and Ac-L-(c)PAP-NHMe. The transition wavelength predicted by averaging the values of the map according to a Boltzmann distribution are: $\lambda^{np} = 479$ nm and $\lambda^{pp} = 270$ nm for Ac-L-(t)PAP-NHMe, and $\lambda^{np} = 484$ nm and $\lambda^{pp} = 267$ nm for Ac-L-(c)PAP-NHMe. The shifts predicted for the *trans-to-cis* isomerization of Ac-L-PAP-NHMe, $\Delta \lambda^{np} = +5$ nm (red shift) and $\Delta \lambda^{pp} = -3$ nm (blue shift), are very similar to those obtained for the free AB at the same theoretical level ($\Delta \lambda^{np} = +4$ nm and $\Delta \lambda^{pp} = -3$ nm). In opposition, the calculated $\mu = \mu(\phi, \psi)$ maps evidenced that the isomerization of AB has a significant impact on the μ of Ac-L-PAP-NHMe (*i.e.* around 5 Debyes). Accordingly, μ has been proposed as key property to regulate and control the activity of L-PAP containing peptides and proteins.

4.4.6. References

- [1] Woolley, A. *Acc. Chem. Res.* 2005, 38, 486.
- [2] Yang, S.; Li, L.; Cholli, A. L.; Kumar, J.; Tripathy, S. K. *Biomacromolecules* 2003, 4, 366.
- [3] Bose, M.; Groff, D.; Xie, J.; Brustad, E.; Schultz, P. G. *J. Am. Chem. Soc.* 2006, 128, 388.
- [4] Behrendt, R. Renner, C.; Schenk, M.; Wang, F. Q.; Wachtweil, J.; Oesterhelt, D.; Moroder, L. *Angew. Chem. Int. Ed. Engl.* 1999, 38, 2771.
- [5] Löweneck, M.; Milbradt, A. G.; Root, C.; Satzger, H.; Zinth, W.; Moroder, L.; Renner, C. *Biophys. J.* 2006, 90, 2099.
- [6] Renner, C.; Moroder, L. *Chem.Bio.Chem.* 2006, 7, 868.
- [7] Willner, I.; Rubin, S. *Angew. Chem. Int. Ed. Engl.* 1996, 35, 367.
- [8] Willner, I. *Acc. Chem. Res.* 1997, 30, 347.
- [9] Pieroni, O.; Fissi, A.; Angelini, N.; Lenci, F. *Acc. Chem. Res.* 2001, 34, 9.
- [10] Sisido, M.; Ishikawa, Y.; Itoh, K.; Tazuket, S. *Macromolecules* 1991, 24, 3993.
- [11] Rau, H. In *Photochromism: Molecules and Systems*; Duerr, H., Bouas-Laurent, H., Eds.; Elsevier Science B.V.: Amsterdam, The Netherlands, 1990; pp 165-192.
- [12] Nakayama, K.; Endo, M.; Majima, T. *Chem. Commun.* 2004, 2386.
- [13] Pearson, D.; Downard, A. J.; Muscroft-Taylor, A.; Abell, D. *J. Am. Chem. Soc.* 2007, 129, 14862.
- [14] Hamachi, I.; Hiraoka, T.; Yamada, Y.; Shinkai, S. *Chem. Lett.* 1998, 537.
- [15] Liu, D.; Karanicolas J.; Yu, C.; Zhang, Z. Woolley, A. *Biorg. Med. Chem. Lett.* 1997, 7, 2762.
- [16] James, D. A.; Burns, D. C.; Woolley, G. A. *Prot. Engin.* 2001, 14, 983.
- [17] Muranaka, N.; Hoshaka, T.; Sisido, M. *FEBS Lett.* 2002, 510, 10.
- [18] Kumar, G. S.; Neckers, D. C. *Chem. Rev.* 1989, 89, 1915.
- [19] Lednev, I. K.; Ye, T. Q.; Matousek, P.; Towrie, M.; Foggi, P.; Neuwahl, F. V. R.; Umapathy, S.; Hester, R. E.; Moore, J. N. *Chem. Phys. Lett.* 1998, 290, 68.
- [20] Nägele, T.; Hoche, R.; Zinth, W.; Wachtveitl, J. *Chem. Phys. Lett.* 1997, 272, 489.
- [21] Fujino, T.; Arzhantsev, Y. S.; Tahara, T. *J. Phys. Chem. A* 2001, 105, 8123.
- [22] Chan, C. W.; Lu, Y. C.; Wang, T. T.; Diao, E. G. W. *J. Am. Chem. Soc.* 2004, 126, 10109.
- [23] Ishikawa, T.; Noro, T.; Shoda, T. *J. Chem. Phys.* 2001, 115, 7503.

- [24] Fliegl, H.; Köhn, A.; Hättig, C.; Ahlrichs, R. *J. Am. Chem. Soc.* 2003, *125*, 9821.
- [25] Cembran, A.; Bernadi, F.; Garavelli, M.; Gagliardi, L.; Orlandi, G. *J. Am. Chem. Soc.* 2004, *126*, 3234.
- [26] Ciminelli, C.; Granucci, G.; Persico, M. *Chem. Eur. J.* 2004, *10*, 2327.
- [27] Tiago, M. L.; Ismail-Beigi, S.; Louie, S. G. *J. Chem. Phys.* 2005, *122*, 094311.
- [28] Jacquemin, D.; Perpète, E.; Scuseria, G. E.; Ciofini, I.; Adamo, C. *Chem. Phys. Lett.* 2008, *465*, 226.
- [29] Briquet, L.; Vercauteren, D. P.; André, J. M.; Perpète, E. A.; Jacquemin, D. *Chem. Phys. Lett.* 2007, *435*, 257.
- [30] Pieroni, O.; Houben, J. L.; Fissi, A.; Constantino, P.; Ciardelli, F. *J. Am. Chem. Soc.* 1980, *102*, 5913.
- [31] Houben, J. L.; Fissi, A.; Bacciola, D.; Rosato, N.; Pieroni, O.; Ciardelli, F. *Int. J. Biol. Macromol.* 1983, *5*, 94.
- [32] Ciardelli, F.; Pieroni, O.; Fissi, A.; Houben, J. L. *Biopolymers* 1984, *23*, 1423.
- [33] Loos, P.-F.; Preat J.; Laurent, A. D.; Michaux, C.; Jacquemin, D.; Perpète, E. A.; Assfeld, X. *J. Chem. Theory Comput.* 2008, *4*, 637.
- [34] Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* 1984, *52*, 997.
- [35] 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.
- [36] Gaussian 03, Revision B.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven Jr., K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Pittsburgh, PA, 2003.

- [37] (a) A. D. Becke, *J. Chem. Phys.* 1993, 98, 1372. (b) A. D. Becke, *J. Chem. Phys.* 1993, 98, 5648. (c) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* 1993, 37, 785.
- [38] A. D. McLean, G. S. Chandler, *J. Chem. Phys.* 1980, 72, 5639.
- [39] (a) Miertus, S.; Scrocco, E.; Tomasi, J. *J. Chem. Phys.* 1981, 55, 117. (b) Miertus, S.; Tomasi, J. *J. Chem. Phys.* 1982, 65, 239. (c) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* 2005, 105, 2999.
- [40] Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.*, 1998, 109, 400.
- [41] (a) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* 1999, 110, 5029. (b) Adamo, C.; Barone, V. *J. Chem. Phys.* 1999, 110, 6158.
- [42] Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* 2004, 121, 3405.
- [43] Kopple, K. D.; Marr, D. H. *J. Am. Chem. Soc.* 1967, 89, 6193.
- [44] Robinson, D. R.; Jencks, W. P. *J. Am. Chem. Soc.* 1965, 87, 2470.
- [45] Cubero, E.; Luque, F. J.; Orozco, M. *Proc. Natl. Acad. Sci.* 1998, 95, 5976.
- [46] Halab, L.; Lubell, W. D. *J. Am. Chem. Soc.* 2002, 124, 2474.
- [47] Steiner, T.; Koellner, G. *J. Mol. Biol.* 2001, 305, 535.
- [48] Steiner, T.; Schreurs, A. M. M.; Kanters, J. A.; Kroon, J. *Acta Crystallogr., Sect. D* 1998, 54, 25.
- [49] Mitchell, J. B. O.; Nandi, C. L.; Ali, S.; McDonald, I. K.; Thornton, J. M. *Nature* 1993, 366, 413.
- [50] Vargas, R.; Garza, J.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* 2002, 106, 3213.
- [51] Gould, I. R.; Cornell, W. D.; Hillier, I. H. *J. Am. Chem. Soc.* 1994, 116, 9250.
- [52] Casanovas, J.; Nussinov, R.; Alemán, C. *J. Org. Chem.* 2008, 73, 4205.
- [53] Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *J. Org. Chem.* 2007, 72, 2174.
- [54] Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Org. Chem.* 2008, 73, 644.
- [55] Hudaky, I.; Hudaky, P.; Perzel, A. *J. Comput. Chem.* 2004, 25, 1522.
- [56] Revilla, G.; Torras, J.; Giménez, A. I.; Cativiela, C.; Nussinov, C.; Alemán, C. *J. Org. Chem.* 2009, 74, 2403.
- [57] Alemán, C.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Casanovas, J. *J. Org. Chem.* 2009, 74, 7834.
- [58] Jamorski-Jödicke, C.; Lüthi, H. P. *J. Am. Chem. Soc.* 2002, 125, 252.

- [59] Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. *J. Chem. Theory. Comput.* 2009, 5, 2420.
- [60] Le Bahers, T.; Pauporté, T.; Scalmani, G.; Adamo, C.; Ciofini, I. *Phys. Chem. Chem. Phys.* 2009, 11, 11276.
- [61] Cossi, M.; Barone, V. *J. Chem. Phys.* 2001, 115, 4708.
- [62] Adam, W.; Krebs, O. *Chem. Rev.* 2003, 103, 4131.
- [63] Jacquemin, D.; Preat, J.; Wathelet, V.; Fontaine, M.; Perpète, E. A. *J. Am. Chem. Soc.* 2006, 128, 2072.
- [64] Briquet, L.; Vercauteren, D. P.; Andre, J.-M.; Perpetè, E.; Jacquemin, D. *Chem. Phys. Lett.* 2007, 435, 257.
- [65] Andersson, J.-A.; Petterson, R.; Tegner, L. *J. Photochem.* 1982, 20, 17.

5. Engineering of bioinspired system

5.1. Exploring the Energy Landscape of a Molecular Engineered Analog of a Tumor – Homing Peptide

5.1.1. Introduction

The systematic use of nanostructures for clinical applications is rapidly becoming a central milestone in modern nanomedicine. Among the possible nano-approaches, treatment systems based upon nanoparticles able to target tumor cells, rapidly took the lead.¹ Even though the first generation of nanoparticles relied on vessel “leakiness” for preferential accumulation in tumors - which limited the efficiency of such particles because of their low penetration capability into the targeted tumors - they were trapped by the sustained high pressure of the interstitial tumor fluid.²

An alternative approach is targeting the nanoparticles to specific molecular receptors in blood vessels: tumors express many molecules that are not significantly populated in normal tissues and the receptors are available for direct binding of species in the blood stream.³⁻⁵ Different nanosystems have been proposed, such as the super-paramagnetic iron oxide (SPIO) nanoparticles coated with dextran, by Ruoslahti and co-workers. These nanoparticles are able to home to tumors while amplifying their homing activity through a mechanism that resembles formation of blood clots by platelets.⁶ The bioactive part of these nanoparticles is the coating peptide, which can recognize selectively clotted plasma proteins.⁶ This short linear peptide (CREKA, Cys-Arg-Glu-Lys-Ala) was recently discovered by *in vivo* screening of phage-display peptide libraries^{3,7} for tumor-homing peptides in tumor-bearing MMTV-PyMT transgenic breast cancer mice.⁸ Synthesized CREKA labeled with the fluorescent dye 5(6)-carboxyfluorescein (FMA) was detectable in human tumors from minutes to hours after intravenous injection, whereas it was essentially undetectable in normal tissues.⁹

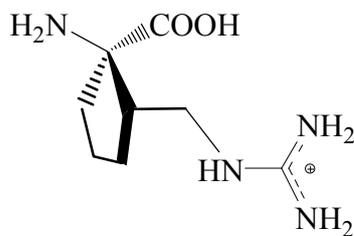
For the last three years we have devoted our efforts to devise protection strategies of CREKA against proteolytic cleavage with concomitant stabilization of the bioactive folded conformation.¹⁰⁻¹² Following determination of the bioactive conformation, specific replacements were designed using non-coded amino acids. This strategy is expected to grant proteolytic protection and, simultaneously, allow control of the conformational preferences. Exploration of CREKA conformational preferences was performed under conditions in which the homing peptide was

experimentally shown to bind to the clotted plasma proteins: as a free peptide, attached to a nanoparticle, and inserted into a viral capsid protein (in a phage display library¹⁰). Independently of the simulation conditions, the most favored organizations often featured turn motifs, which generally involved Cys and either Glu or Lys. Among this variety of turn organizations, a common pattern was observed in all simulated cases: a β -turn arrangement that involves the formation of a hydrogen bond between Cys and Lys residues. This organization was identified as the bioactive motif, which enabled the formation of salt bridges between the side-chains of all ionized amino acids, since the curvature site made them protrude outward. Once the bioactive structural profile was delineated, non-coded amino acids were used in order to enhance the protection against proteolytic cleavage and the homing activity.¹³

A key question relates to the final choice of the non-coded amino acid in the substituted position which should satisfy the bioactive conformational requirements. Examination of the conformation adopted by Arg in the bioactive organization shows that it is located at position $i+1$ of a β -turn.¹⁰ If one searches the conformational preferences of small peptides and proteins, this position is usually occupied by Pro.¹⁴⁻¹⁶ Hence, initially, a new amino acid with a Pro scaffold and Arg side chain, named (⁷Pro)Arg, was modeled.¹² Subsequent work¹² showed that inclusion of this residue significantly reduced the conformational flexibility of the peptide, enhancing its tendency to accommodate partially folded conformations centered at the Arg site. In addition, inspection of the lowest energy conformation indicated that the ionized side chains faced the same side of the molecule, a feature that seems to have a role in the peptide bioactivity. However, insertion of this Arg surrogate presented drawbacks: despite the similarities described above, the turn conformations adopted when (⁷Pro)Arg replaced Arg did not match the most favored conformations previously described,¹⁰ since it shifted the conformational preferences of the CREKA analog towards γ -turn formation. Therefore, the search for appropriate amino acid backbones continued.

Within that context, a specific family of non-coded amino acids, the 1-aminocycloalkane-1-carboxylic acids¹⁸⁻¹⁹ (known in the abbreviated form as Ac_{*n*}c, with *n* referring to the ring size), have been recently proposed as a structural tool for conformational control and enhancement of naturally occurring structural motifs.¹⁹⁻²³ This series of residues exhibits a restricted conformational space

characterized by a high propensity to adopt φ, ψ backbone angles typical of the 3_{10} - α -helix (with some distortion in the case of Ac_3c).²³⁻²⁵ As we previously mentioned, in the majority of the most favored CREKA conformations Arg is at position $i+1$ of a β -turn motif.¹⁰ In such arrangement, the backbone dihedral angles adopted by the residue are in the regions corresponding to the aforementioned helical structures. Hence, in order to take advantage of Ac_nc conformational preferences, a new amino acid was designed by attaching the Arg functionalized side-chain to the cycloalkane moiety in Ac_5c , the resulting compound being denoted c_5Arg (see scheme I).²⁶ c_5Arg is a substituted Ac_5c -like derivative of *nor*-arginine, where *nor* refers to a chain-length reduction of one carbon atom with respect to the Arg side-chain.²⁶ A complete conformational characterization showed that the new residue preserved the conformational preferences corresponding to those expected for its Ac_5c skeleton,²⁵ which makes it suitable for tests as a CREKA homing peptide-enhancing modification.



Scheme I

This work reports the conformational preferences of a new CREKA analog, hereafter denoted Cc_5REKA , obtained by substituting Arg by the *cis* enantiomer of c_5Arg , where *cis* refers to the relative position of the guanidinium group in the cycloalkyl segment. Specifically, results obtained in previous quantum mechanical calculations on both the *cis* and *trans* isomers of c_5Arg suggest that, *a priori*, the former is the best to fulfill the structural requirements of the peptide.²⁶ The energy landscape of the new peptide analog has been explored by techniques based upon molecular dynamics (MD) simulations, *i.e.*, modified simulated annealing MD (SA-MD)²⁷ and replica exchange MD (REMD)²⁸.

5.1.2. Methods

Conformational Search and Force-Field Calculations. The conformational preferences of *Cc₃REKA* were explored using both modified SA-MD and REMD sampling strategies. The first methodology, which was previously used to study the conformational preferences of the parent peptide,^{10,11} is based on the minimization of conformations generated at the initial and intermediate states of several SA-MD cycles. In the SA-MD technique the high starting temperature is gradually reduced during the simulation, allowing the system to surmount energy barriers. In spite of this, in practice, it is known that such sampling technique does not always lead the system to the most stable region at the end of the simulation. However, recent studies showed that this limitation can be overcome, since very low energy structures can be reached by minimizing structures generated during the SA-MD cycles.²⁷ Accordingly, such modification of this sampling technique was found to be robust enough to obtain conformations close in energy to the global minimum but located in different regions of the potential energy hypersurface, when the extracted conformations from each SA-MD cycle are minimized.¹⁰⁻¹²

The same conformational profile was studied using REMD.²⁸ This MD based technique speeds convergence relative to brute force conventional MD.²⁷ The method is based on the generation of a number of copies (“replicas”) of the system that span from the temperature of interest (e.g., physiological temperature) to heated states, which facilitates overcoming the free energy barriers. Periodic swaps of neighboring replicas, which are performed while preserving an overall Boltzmann-weighted ensemble at each temperature (Monte Carlo based criterion of swap acceptance), enable conformations to heat up and cool down. Nonetheless, recent work has indicated that for short peptides the optimum convergence of the sampling can be obstructed by a combination of two factors:²⁹ the intrinsically high flexibility of these systems and the limited thermal agitation associated with the reference temperature. In this work we show that under normal conditions flexible systems may not reach some conformations of medium energy rank, affecting significantly the final bioactive ensemble of conformations.

Computational details. Energy calculation. The conformational energy for all the simulated systems was calculated using the Amber force field.³⁰ All bonding and non bonding parameters were extracted from Amber libraries³⁰ except for those describing the non coded residue which were previously optimized and tested.²⁶

Molecular models. The simulated system consisted of the Cc_5REKA peptide 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 ($45.90 \times 45.90 \text{ \AA}^2$), with van der Waals parameters $R = 2.35 \text{ \AA}$ and $\epsilon = 0.90 \text{ kcal} \cdot \text{mol}^{-1}$ and no electric charge. This system, which is identical to that considered for CREKA in our previous work,¹⁰ mimics the experimental conditions,^{3,7} *i.e.* the peptide linked to the surface of a nanoparticle. Therefore, the results obtained in this work for the CREKA analog have been compared with those reported for natural CREKA attached to an identical surface (system III in ref. 5), unless otherwise indicated. The Cc_5REKA analog attached to the surface was placed in the center of a cubic simulation box ($a = 45.90 \text{ \AA}$) filled with 2663 explicit water molecules, which were represented using the TIP3 model.³¹ Two chloride ions and one sodium ion were added to the simulation box to reach electric neutrality (net charges were considered for Arg, Lys, and Glu at neutral pH).

Simulated annealing. Prior to the production cycles with the modified SA-MD, the system was equilibrated. 0.5 ns of NVT-MD at 500 K were used to homogeneously distribute the solvent and ions in the box. Next, thermal equilibration for 0.5 ns in the constant-NVT ensemble at 298 K, followed by density relaxation for 0.5 ns in the constant-NPT ensemble at 298 K, were performed. 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/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 structure generated in a modified SA-MD cycle was used as starting conformation of the next cycle.

Atom pair distance cut-offs were applied at 14.0 \AA 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 \AA . Both

temperature and pressure were controlled by the weak coupling method, the Berendsen thermobarostat,³² using a time constant for heat bath coupling and a pressure relaxation time of 1 ps. Bond lengths were constrained using the SHAKE algorithm³³ with a numerical integration step of 2 fs. All MD simulations were performed using the NAMD program.³⁴

Replica Exchange Molecular Dynamics. Simulations were performed using the AMBER 10 program.³⁵ In order to use the low complexity of the system to ensure a reasonable replica swaps ratio of acceptance (approaching a Partial Replica Exchange Molecular Dynamics model), 8 replicas were exponentially distributed in the temperature range from 283.8 K to 418.7 K.³⁶ Exchanges were attempted every 40 ps between all neighboring replicas with an average acceptance rate of 16%, above the minimal acceptance rate for the complexity of the studied peptide.³⁶ The REMD trajectories resulted in a cumulative simulation time of 41 ns. Between replica exchanges, the system was evolved using NVT Langevin MD³⁷ with a damping coefficient of $\gamma = 2.5 \text{ ps}^{-1}$ and an integration step of 2 fs. The replicas were previously equilibrated by a set of short runs (isothermal and isobaric equilibration), and completed with a final NVT run of 0.5 ns to ensure that each replica reached the target temperature. In all cases the surface particles were fixed at the initial positions and only peptide, water molecules, and ions were allowed to move. Atom pair distance cut-offs were applied at 14.0 Å to compute the van der Waals and electrostatic interactions. Bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm³³ and explicit water molecules of TIP3 model³¹ were used in every replica except for evaluating the replicas energies before the swapping process. In order to increase the global efficiency of the technique, a hybrid solvent model was used to evaluate each replica total energy.³⁸ Specifically, 118 explicit water molecules were taken into account as a first hydration shell while the rest of solvent contribution was computed using Generalized Born Model implemented by Hawkins and coworkers,³⁹ using the late parametrization of Tsue and Case.⁴⁰ The REMD data were collected from the last 10 ns of simulation at the targeted temperature of 300 K to perform clustering analysis.

Conformation Classification and Clustering Analysis. In order to construct a list of unique minimum energy conformations, each set of structures provided by either modified SA-MD or REMD were compared among them. The list was

organized by rank ordering all the unique minimum energy conformations found following an increasing order of energy. In the case of SA-MD, previously listed conformations that appear at a new cycle were discarded. The criterion to identify unique minimum energy conformations was already developed in previous work, and it is based on defining virtual dihedral angles combined with the computation of the interaction pattern, *i.e.* identification of salt bridges, hydrogen bonds and dipole-dipole interactions.¹⁰

Five virtual dihedral angles were defined considering the α -carbon atoms of the five residues, the methyl carbon atom of the acetyl and *N*-methylamine capping groups, and one acetyl hydrogen atom. The existence of different interactions is 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: H \cdots O distance ($d_{\text{H}\cdots\text{O}}$) shorter than 2.50 Å and $\angle\text{N-H}\cdots\text{O}$ angle higher than 120.0°; and (c) dipole-dipole: distance between dipoles shorter than 3.00 Å and the interaction has not been counted as a hydrogen bond. Two structures were considered different if they differ in at least one of their virtual dihedral angles by more than 60° or in at least one of the interactions counted. All the structures categorized as different were subsequently clustered according to a criterion based on the presence of the intramolecular interactions mentioned above.

5.1.3. Results and discussion

Clustering Analysis Following the previously developed analysis strategy, we classified the microstructures generated by both SA-MD and REMD following a hierarchical clustering.¹⁰ In this process, each structure is compared with the rest of the structures obtained using the same strategy and included in a cluster that presents the same main-chain conformation and the same polar interaction types. This analysis provides important information about the population of conformations generated for the studied system and, additionally, it serves as a simple and efficient criterion of convergence for SA-MD (*i.e.* when no new conformations are obtained in the annealing cycles the search is concluded). In Figure 5.1.1. the convergence of the exploration clearly shows that after 9 cycles the number of new conformations does not increase. Moreover, the number of different structures produced for the peptide analog, 1125, is close to the number of different structures described for the wild type peptide, 1306, after an equivalent

number of SA-MD cycles.¹⁰ However, the significant reduction of the number of different minima, 14%, seems to indicate that the replacement has succeeded, at least in terms of reduction of the conformational flexibility. On the other hand, much faster convergence was reached by REMD considering the same criterion (*i.e.* the number of newly explored conformations), a plateau being obtained after only 8 ns of cumulative simulation time, and the maximum number of explored conformations was found to be only 550. Below, we further show that this relatively small number of different structures may be enough to represent the most relevant conformational features of the bioactive organization for *Cc*₅REKA peptide.

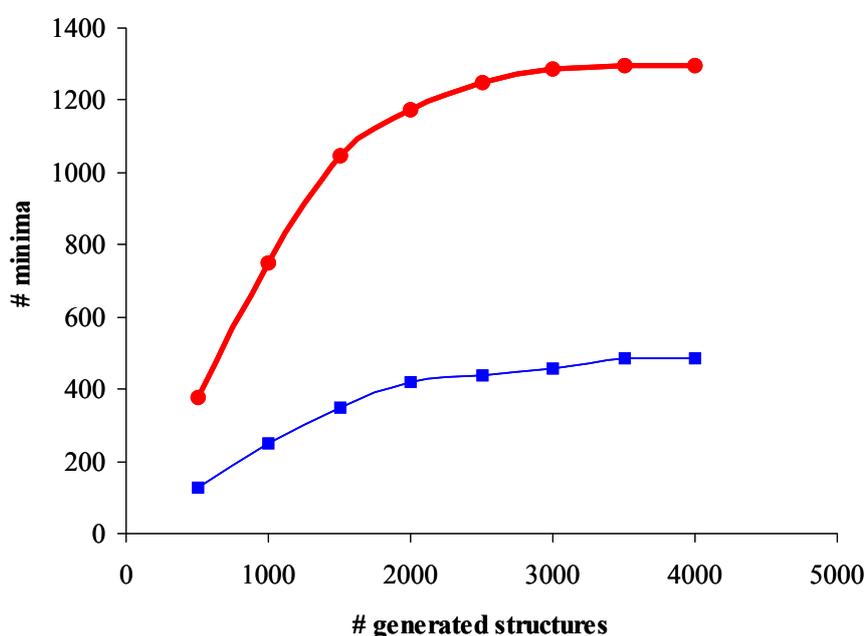


Figure 5.1.1.- Number of unique minimum energy conformations found using SA-MD (red line, solid squares) and REMD (blue line, empty squares) the number of minimum energy structures in the conformational search process.

Although the main goal of the current study was assessing the suitability of *c*₅Arg as a homing enhancer, comparison between two different conformational search strategies for short peptides is necessary. Despite the general perception that REMD is the most suitable strategy to explore the conformational space of any

given molecule, recent works have pointed out the flaws of the method when dealing with highly flexible peptides.⁴¹ These limitations were attributed to the inherent efficiency of REMD for locating deep basins. Thus, when the energy profile that separates the lowest energy structure and other low energy arrangement is not sufficiently rough, the exploration can be confined to those structures of lower conformation energy. In this study, the rapid convergence is apparently reached by biasing the conformational ensemble towards those structures that presented lower potential energy.

Figure 5.1.2., which depicts the energy distribution of all the generated structures, shows a Gaussian and a bimodal function for the SA-MD and REMD methods, respectively. These profiles indicate that for REMD the lower energy basins are overpopulated while zones that might represent intermediate potential energy ranges are not explored. Despite the apparent independent evolution of each trajectory, these results show that replicas close to the target temperature may present some degree of structural correlation. Thus, in cases of very flexible peptides longer cumulative time should be used to complete a proper exploration of the conformation space (*i.e.*, higher computational effort). Taking into account that only the last 10 ns of simulation were collected to obtain statistical information about the conformational distribution at the given temperature, the computational resources used by REMD for obtaining 40 ns of cumulative simulated time were significantly larger than those committed for reaching the search convergence using SA-MD. It is worth noting that in our case this methodology has allowed us to obtain an energy distribution closer to a canonical ensemble than the one reached by REMD. The main difference between the two methods is the introduction of the temperature as a variable in the conformational search process, which imposes a heavier thermal hindrance for reaching those conformations that are kinetically less favored. As we show later, the first peak of the bimodal distribution derived from the REMD results corresponds to those conformations that are similar to the wild type bioactive motif, while the second peak corresponds to organizations that do not resemble those of lower energy. The central void in the population of medium energy structures may denote kinetic traps that preclude the formation of transient structures at 300 K. These structures, which are needed to reach the conformations that belong to the lowest energy rank, are only described by the SA-MD method.

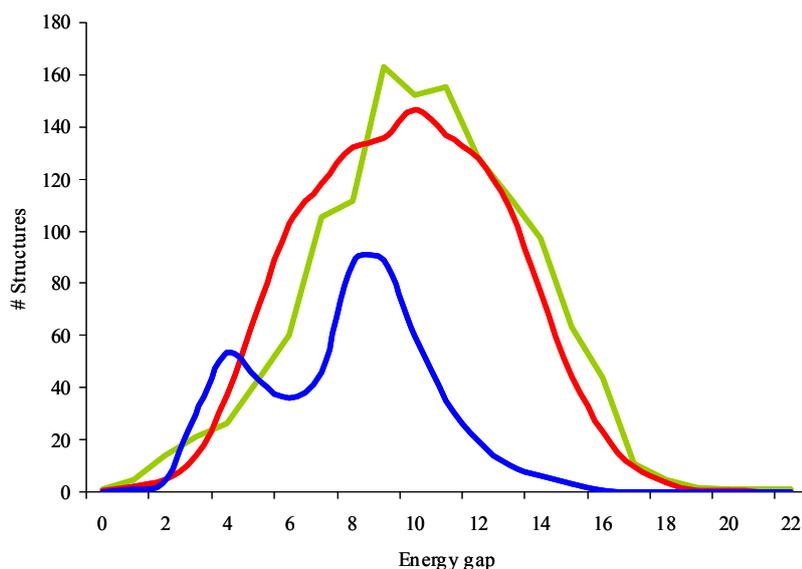


Figure 5.1.2.- Distribution of energies for the minimum energy conformations of Cc_5REKA found using SA-MD (red line) and REMD (blue line). The profile obtained with CREKA using SA-MD (green line, ref. 10) has also been enclosed for comparison. All energies are relative to the corresponding global minimum.

It was previously demonstrated that the bioactive profile of CREKA mainly corresponds to the lowest energy conformations.¹² Because both the brute force-based simulations (SA-MD) and the statistically rigorous exploration (REMD) techniques drove toward the most energetically relevant structures, they should be equally efficient in assessing the conformational preferences of the Cc_5REKA analog. Our REMD results still might suggest the presence some degree of deficient exploration. These apparent restrictions though can be surpassed by considering a higher number of replicas, which would allow overcoming the REMD limitations when it is used to search the energy landscape of small flexible peptides. Nevertheless, it should be emphasized that this option increases remarkably the computer resources devoted to the conformational search.⁴¹

Conformational features. A detailed analysis of the conformational exploration performed by both SA-MD and REMD shows interesting features and differences that are inherent to these techniques. Figure 5.1.3., which shows the Ramachandran plots for the five residues of Cc_5REKA , suggests that SA-MD

covers wider extent of the main chain conformational space than REMD. However, the former method explores conformations that are accessible, without representing the effect of the temperature agitation (*i.e.* the kinetics accessibility of these conformations at a determined temperature). Thus, most of the conformations obtained by SA-MD might not be reachable at the studied reference temperature. As we have previously indicated, this is not a surprising feature since this technique was designed to speed up the search of the most relevant conformational minima at a specific temperature. Despite this difference, the lower energy structures derived from the two techniques are almost coincident (Figure 5.1.3.). In other words, at 300 K the most relevant conformations for the studied pentapeptide analog are well represented with relatively short trajectories and with a limited amount of replicas. It is worth noting that at such temperature some conformations that are separated by 4 or 5 kcal·mol⁻¹ from the lowest energy structure, which were detected via SA-MD, are not found by REMD. These conformations that had a meaningful contribution in the conformational profile of the wild type segment¹⁰⁻¹² require higher thermal agitation to be reached. This question though does not invalidate our bioactive motif of both wild type and the peptide analog, since the physiological conditions are represented by temperatures that feature almost 10 degrees above the studied conditions.

Regarding the conformational space accessible by c₅Arg (position 2 in Cc₅REKA), Figure 5.1.3. shows that the non-coded amino acid has achieved its primary goal: it restricts the conformational freedom of the peptide and, simultaneously, favors specific conformation profiles. The flexibility reduction is selective because the cyclic nature of the Arg surrogate favors the adoption of the conformation that the coded Arg presents in the bioactive conformation.¹⁰ Independently of the exploration method, the main-chain conformations found for the second residue of the Cc₅REKA analog mainly clustered at the two α regions of the Ramachandran plot, which correspond to the more favored conformations found for a single c₅Arg residue in aqueous solution.²⁶ In both cases the structures of lower energy (blue and black dots in Figure 5.1.3.) present conformations for Cc₅REKA's 2nd position that are located at the α_L region, which corresponds to that adopted by Arg in the parent CREKA peptide. On the other hand, the lower efficiency exhibited by the SA-MD method for locating the more stable conformations widens the range of explored conformations for c₅Arg, with a

significant number of Cc_5REKA arrangements with relative energies higher than 5 kcal·mol⁻¹ versus the tight conformation distribution observed using REMD. However, in the case of the SA-MD exploration, those Cc_5REKA structures that were within 2-3 kcal·mol⁻¹ from the absolute minimum again showed conformations at the α_L region, which was the main objective of introducing c_5Arg as Arg surrogate.

Despite the initial success, in order to enhance the homing activity, the biased conformation of c_5Arg in the peptide analog should also favor the formation of turn conformations beyond the residue position. As we mentioned before, we had already tried enhancing the stability of the β -turn featured by the bioactive conformation of CREKA replacing the Arg by another surrogate of low conformational flexibility, (^lPro)Arg.¹² In fact, the incorporation of this Pro derivative into CREKA also reduced the conformational flexibility, especially at the position 2. However, it caused the disruption of the desired β -turn conformation, apparently because of the drastic structural changes that the Pro skeleton introduced in the dynamics of the adjacent residues. In the case of c_5Arg , despite fixing the conformational preferences of position 2, the higher flexibility of the Ac_5c skeleton²⁵ facilitates the adjacent Glu adopting the proper conformation for the continuation of the turn (Figure 5.1.3.). Thus, in the minimum energy conformations found for the Cc_5REKA analog, Glu usually falls in the same region that was featured for the parent peptide, making possible the formation of the β -turn.¹⁰⁻¹²

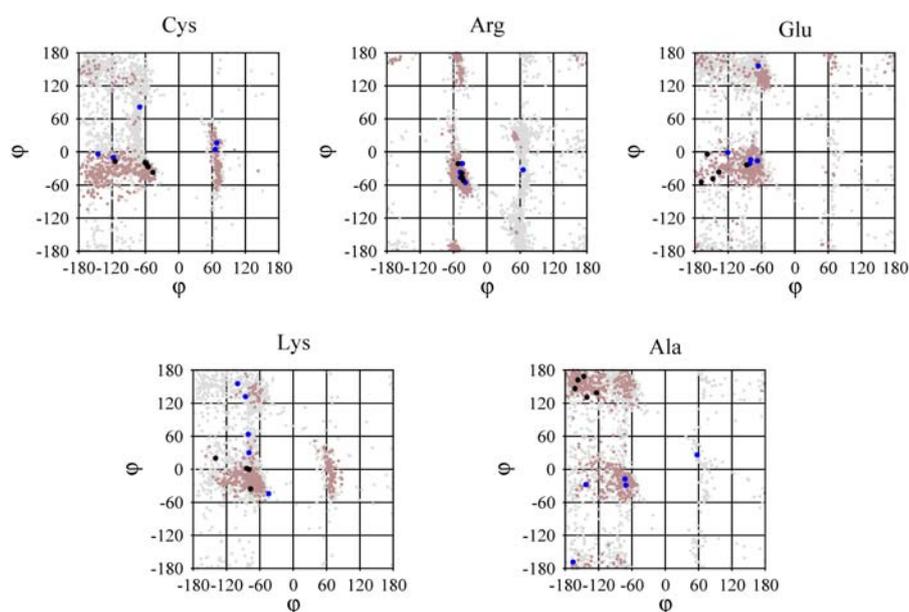


Figure 5.1.3. Ramachandran Plot of the main chain dihedral angles (ϕ, ψ) for the five residues of Cc_5REKA for the structures of low energy generated by both SA-MD and REMD. Grey dots represent all the structures generated by SA-MD, whereas blue dots correspond to the 5 minima of lowest energy. Dark brown dots depict all the structures generated by REMD, whereas black dots indicate the 5 minima of lowest energy.

Structural comparison between the lowest energy conformations.

Examination of the backbone conformations for the most favored arrangements of Cc_5REKA reveals several features that are closely related with the exploration efficiency and the effect of c_5Arg on the conformational properties of CREKA. Figure 5.1.4. plots the Root Mean Square Difference (RMSD) correlation between the 10 lowest energy structures obtained in each landscape exploration (Cc_5REKA generated by SA-MD and REMD, named SA_n and RE_n , respectively, where n refers to rank position and number 01 the lowest energy structure) and their structural correlation with the previously determined bioactive conformational pool of the wild type homing peptide¹⁰ (named WT_n following an analogous nomenclature). As can be seen, the most favored conformations generated by REMD are significantly similar between themselves. The largest main-chain difference between the structures is 1.83 Å, which corresponds to a pair of almost iso-energetic conformations, RE04 and RE05. Again, this feature indicates that the

conformational space explored by REMD is essentially restricted to peptide organizations with low energy, which is consistent with the results showed in the previous sections.

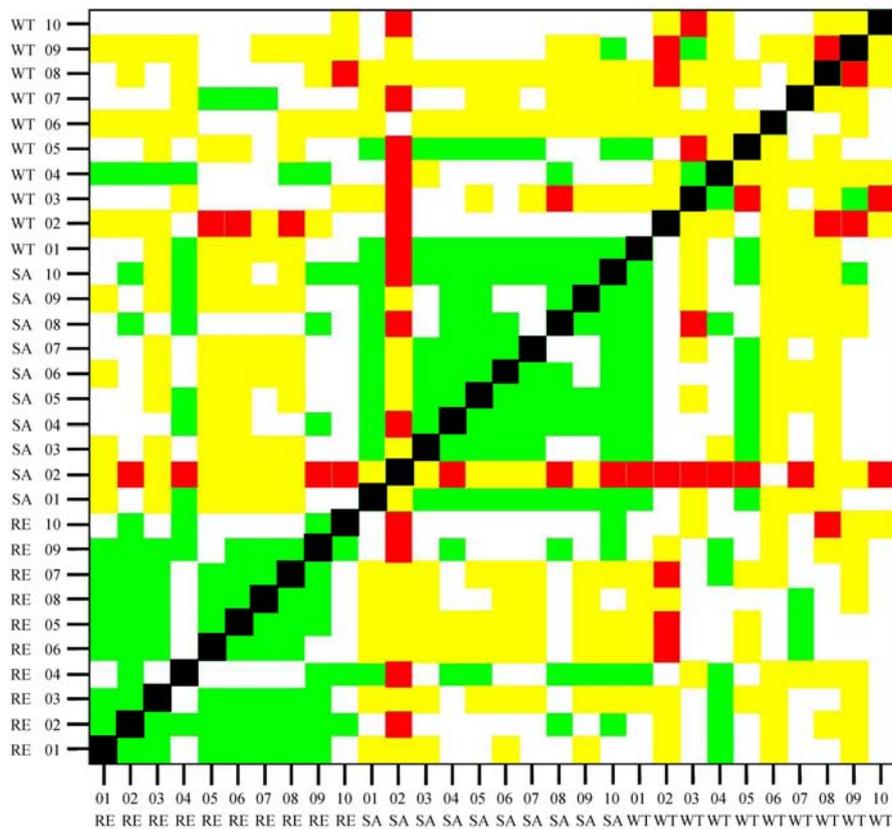


Figure 5.1.4.- Main chain structural correlation between each 10 structures of lowest energy generated by SA-MD for CREKA (*WT_n*, ref .10), using SA-MD for *Cc₃REKA* (*SAn*) and REMD for *Cc₃REKA* (*REn*). The correlation is computed as the structures pair Root Mean Square Deviation (RMSD). Each color represents a specific RMSD interval of values in which the deviation value of two intersecting structures is situated: green color means RMSD between two structures is lower than 1.3 Å, white is comprised between 1.3 Å and 2.0 Å, yellow between 2.0 Å and 3.0 Å and, finally, orange higher than 3.0Å.

On the other hand, SA-MD appears a more efficient tool when exploring the conformation of small peptides since high structural diversity is observed among those structures of lower energy. This feature becomes especially remarkable for the *SA02* conformation, which features a γ -turn that does not correspond to the β -turn motif frequently found in *Cc₅REKA*. However, differences among the rest of the conformations of *Cc₅REKA* obtained using both methodologies are relatively small, essentially reflecting small variations in the β -turn motif with respect to the most favored organization. Hence, the use of *c₅Arg* has achieved another primary goal: its incorporation in CREKA instead of Arg drastically restricts the potential organizations to those that characterized the bioactive structure for the parent pentapeptide (*i.e.* the β -turn). Thus, Figure 5.1.4. reveals that the majority of backbone conformations are very similar to that described for the bioactive organization of CREKA, with exception of the aforementioned structure *SA02*.

Table 5.1.1. lists the main electrostatic interactions detected in the four arrangements of lower energy for both the parent peptide and *Cc₅REKA*. It is worth noting that *WT01*, *WT03* and *WT04*, despite their evident differences, share important features that make them part of the bioactive profile of CREKA. The three conformations are folded into a β -turn and all expose the charged side chain to the solvent on the same side of the peptide chain.¹⁰⁻¹² When comparing their interaction patterns, both present remarkable similarities in the central segment, -REK-. Specifically, in *WT01* (Figure 5.1.5.a) and *WT04* the backbone of Lys (N-H) and Cys (C=O) interact through a hydrogen bond, closing a type II β -turn. On the other hand, in *WT03* the Lys backbone (N-H) forms a hydrogen bond with the terminal acetyl blocker (Ac), defining an α -turn that may easily transform into a β -turn if the competing Ala C=O forces tightening the turn (Ac and Ala are mutually blocked by both amide groups forming two interlocked hydrogen bond). The topological outcome of this complex arrangement resembles *WT01* and places the ionized side-chains in the proper position for the formation of salt bridges. In summary, with the exception of *WT02*, the structures mainly differ in the conformation adopted by both terminal group.

Table 5.1.1.- Comparison between the Interaction Pattern of the four Minima of Lowest Energy Generated for Natural CREKA and its Analogue *Cc₅REKA* using both SA-MD and REMD ^a

	<i>Cc₅REKA</i> (SA-MD)	<i>Cc₅REKA</i> (REMD)	CREKA (SA-MD) ^d
<i>SA01</i>	HB Lys(N-H)···(O=C)Cys ¹ HB Ala(N-H)···(O=C)Glu ² HB <i>c₅R</i> (C=O)···(⁺ H ₃ N,side chain)Lys SB Lys···Glu	<i>RE01</i> HB Glu(N-H)···(O=C)Ace ¹ HB Lys(N-H)···(OOC ⁻ ,side chain)Glu SB Lys···Glu	<i>WT01</i> HB Lys(N-H)···(O=C)Cys ¹ SB Arg···Glu SB Lys···Glu
<i>SA02</i>	HB <i>c₅R</i> ((N-H)···(O=C)Ace ² HB Glu(N-H)···(O=C)Cys ² HB <i>c₅R</i> (N-H ₂ guanidinium)···(O=C)Glu HB Ala(N-H)···(OOC ⁻ ,side chain)Glu HB Nem(N-H)···(OOC ⁻ ,side chain)Glu SB Lys···Glu	<i>RE02</i> HB Glu(N-H)···(O=C)Ace ¹ HB Lys(N-H)···(O=C)Ace ³ HB <i>c₅R</i> (N-H ₂ guanidinium)···(O=C)Ala SB Lys···Glu	<i>WT02</i> HB Ala(N-H)···(O=C)Arg ¹ HB NEM(N-H)···(O=C)Glu ¹ SB Arg···Glu
<i>SA03</i>	HB Lys(N-H)···(O=C)Cys ¹ HB <i>c₅R</i> (N-H)···(O=C)Lys ² SB Lys···Glu	<i>RE03</i> HB Glu(N-H)···(O=C)Ace ¹ HB Lys(N-H)···(O=C)Ace ³ SB Lys···Glu	<i>WT03</i> HB Lys(N-H)···(O=C)Ace ³ HB Ala(N-H)···(O=C)Ace ⁴ HB Nme(N-H)···(O=C)Cys ⁴ SB Arg···Glu
<i>SA04</i>	HB Lys(N-H)···(O=C)Cys ¹ HB <i>c₅R</i> (N-H ₂ guanidinium)···(O=C)Cys HB <i>c₅R</i> (N-H ₂ guanidinium)···(O=C)Lys SB Lys···Glu	<i>RE04</i> HB Ala(N-H)···(O=C)Cys ³ SB Lys···Glu	<i>WT04</i> HB Lys(N-H)···(O=C)Cys ¹ HB Ala(N-H)···(O=C)Cys ³ HB Nme(N-H)···(O=C)Cys ⁴ SB Arg···Glu SB Lys···Glu

^a Interaction pattern: hydrogen bonds and salt bridges are labeled as HB and SB, respectively. ^b Four minima generated according to reference 10 Classification of tight turns as function of the distance (in number of residues) between hydrogen bond donor and acceptor(refe?):¹ β-turn (*i,i+3*),² γ-turn(*i,i+2*),³ α-turn (*i,i+4*),⁴ π-turn, (*i,i+5*)

Analysis of the four lower energy conformations obtained for Cc_5REKA using REMD (Table 5.1.1.) demonstrates that they all have identifiable counterparts in the structures characterized for the parent peptide, with the exception of structure $RE01$ (Figure 5.1.5.b). The latter conformation features a single β -turn at the C terminus, which involves the N-H of Glu and the C=O of the acetyl blocking group, that was not observed in any CREKA conformation. This organization is partially due to the interaction between the guanidinium group of c_5Arg and the C=O of Ala, which precludes the formation of other hydrogen bonding patterns more similar to those observed in CREKA. $R02$ and $R03$ define arrangements intermediate between the α - and β -turns, leading to turns at the C termini with geometries slightly wider than that typically expected for a conventional β -turn. However, a simple rearrangement of the N terminal group would easily facilitate the formation of the proper turn since in those two conformations the hydrogen bonding acceptor is the C=O of the terminal Ac, whereas in the bioactive conformation of CREKA it is the C=O of Cys. Finally, $R04$ features an α -turn that presents a similar topology to that observed in the parent peptide (Figure 5.1.5.c). Thus, c_5Arg favors the formation of a turn that makes possible the correct spatial location of the Glu and Lys ionized side chains allowing their interaction (Figure 5.1.6.a).

Results obtained with SA-MD show a better match with the expected arrangements for Cc_5REKA . This search strategy allows locating low energy structures that adopt the β -turn featured by $WT01$ and $WT04$. Thus, the N-termini organization adopted by $SA01$, $SA03$ and $SA04$ (Figures 5.1.6.b and 5.1.6.c, respectively) correspond to that found for $WT01$, even though some differences are observed at the C terminus. Finally, $SA02$ presents a double γ -turn conformation (Figure 5.1.5.d), which was also predicted for other Ac_3c -containing peptides.^{21,23}

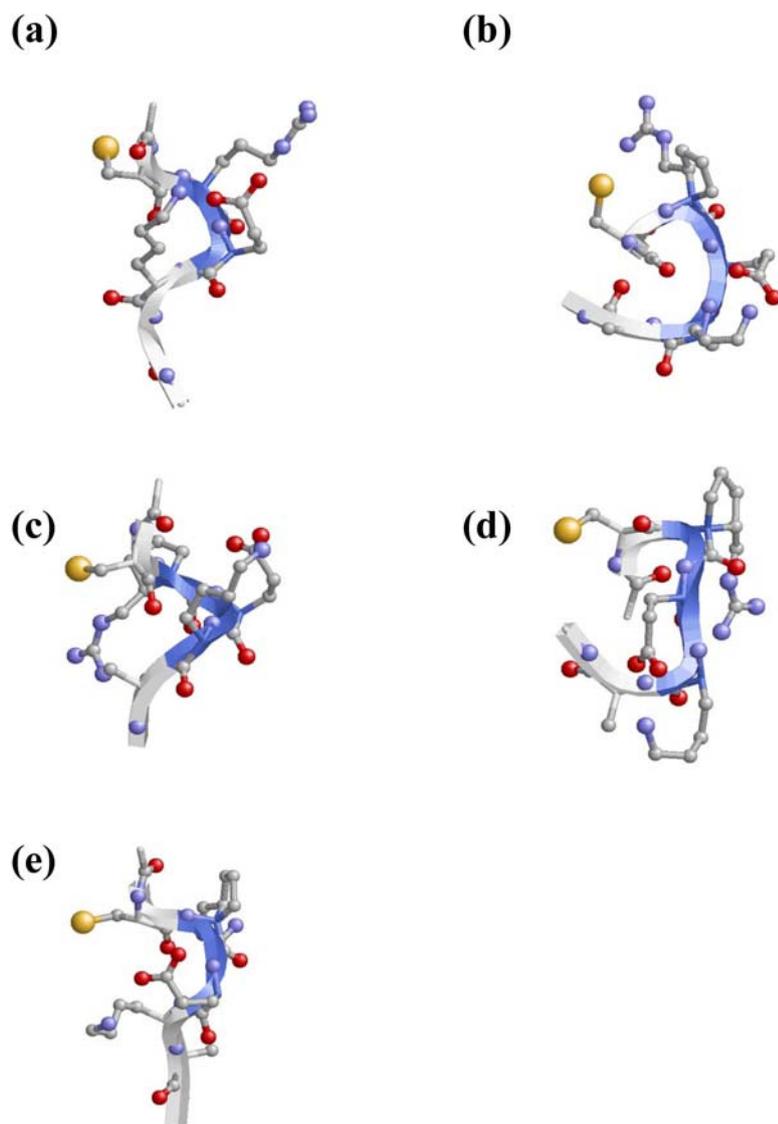
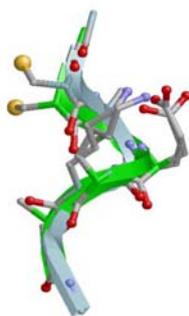
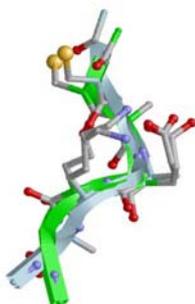


Figure 5.1.5.- Detailed representation of selected low energy conformations obtained for CREKA and *Cc*₃REKA. Main chain is represented by solid cartoons, with standard protein conformational colors (blue and white for turn motifs). Heavy are represented with solid scaled balls and its bonds with solid sticks following CPK colors convention. Hydrogen atoms were removed for clarity. The depicted structures are: (a) WT01 (extracted from ref. 10) (b) RE01 (c) RE04 (d) SA02 (e) SA03 (name codes are explained in text).

(a)



(b)



(c)

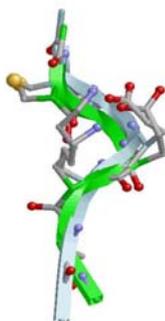


Figure 5.1.6.- Superimposition of selected structures of *Cc*₃REKA with the lowest energy conformation of CREKA (named in the text WT01, ref. 10). In all cases main chains are represented by solid cartoons, green color depicting WT01 structure and cyan color corresponding to each particular *Cc*₃REKA structure. Heavy atoms are represented with solid scaled balls and their bonds with solid sticks following CPK colors convention. Hydrogen atoms were removed for clarity. The superimposed structures are: **(a)** RE04 **(b)** SA01 **(c)** SA03 (name codes are explained in text).

5.1.4. Conclusions

The conformational preferences of a homing peptide synthetic analog have been widely explored by different techniques based on MD simulations. Independent of the strategy, our results have indicated that the recently engineered residue, $c_5\text{Arg}$,²⁶ does accomplish its design aim, which is to reduce the conformational freedom of the CREKA peptide,¹⁰ biasing the accessible conformations towards those that feature the bioactive one. The lower energy conformations of $Cc_5\text{REKA}$ involve the formation of β -turn motifs centered around the second and third residues of the synthetic analog, similar to those observed in the parent peptide.¹⁰ At the same time, this conformational bias aids the formation of an interaction network between the side-chains of the central residues, which was described as a necessary feature for the bioactivity of CREKA (*i.e.*, the orientation of the ionized central side chains towards the same side of the turn allows their interaction with the tumor vessel receptors). In this context, it is worth noting that even in those conformations that were not described by the parent peptide, such as *RE01* and *SA02*, the final topological distribution of the side chains guarantees that the formation of the required interaction pattern.

On the other hand, differences between the two methodologies demonstrate that the temperature is an important variable in the $Cc_5\text{REKA}$ conformational preferences. Even though the energy differences among the ten absolute lowest energy conformations are less than $5 \text{ kcal}\cdot\text{mol}^{-1}$ some of them are not detected at the REMD reference temperature. Thus, the lower thermal agitation featured under normal conditions precludes the complete coverage of the bioactive conformational profile. In contrast, the SA-MD, which is a technique that explores the most relevant regions of the energy landscape without help from higher temperatures, allows obtaining conformations of $Cc_5\text{REKA}$ that were previously reported as relevant for CREKA's bioactivity. Both the lack of thermal distribution in the generated ensembles and the redundant production of conformations starting from those that are already located in low energy regions facilitated the identification of quasi-degenerated energy arrangements that were not detected at the REMD targeted temperature.

In summary, results derived from the exploration of the conformational bioactive profile of $Cc_5\text{REKA}$ demonstrate that inclusion of the engineered residue $c_5\text{Arg}$ as Arg replacement does achieve the objectives of its design: not only could

the new surrogate protect the homing peptide from the protease activity but also increase the stability of the bioactive conformation previously determined for the wild type peptide CREKA. These results confirm our strategy^{12,13,19-23,26} that is aimed to re-design natural amino acids to improve properties of biotechnological interest in natural peptides.

5.1.5. Bibliography

- [1] Weissleder, R.; Bogdanov, A. Jr; Neuwelt, E.A.; Papisov M. Long-Circulating Iron-Oxides for MR-Imaging *Adv Drug Delivery Rev* 1995, 16, 321–334.
- [2] Desai, N; Trieu, V; Yao, ZW; Louie, L; Ci, S; Yang, A; Tao, CL; De, T; Beals, B; Dykes, D; Noker, P; Yao, R; Labao, E; Hawkins, M; Soon-Shiong, P. Increased antitumor activity, intratumor paclitaxel concentrations, and endothelial cell transport of Cremophor-free, albumin-bound paclitaxel, ABI-007, compared with Cremophor-based paclitaxel (2006) *Clin Cancer Res* 2006, 12,1317–1324.
- [3] Hoffman, J.A.; Giraudo, E.; Singh, M.; Zhang, L.; Inoue, M.; Porkka, K.; Hanahan, D.; Ruoslahti, E. Progressive vascular changes in a transgenic mouse model of squamous cell carcinoma *Cancer Cell* 2003, 4, 383–391.
- [4] Oh, P.; Li, Y.; Yu, J.; Durr, E.; Krasinska, K.M.; Carver, L.A.; Testa, J.E.; Schnitzer J.E. Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy *Nature* 2004, 429, 629–635.
- [5] Ruoslahti E Specialization of tumour vasculature *Nat Rev Cancer* 2002, 2, 83–90.
- [6] Simberg, D.; Duza, T.; Park, J. H.; Essler, M.; Pilch, J.; Zhang, L.; Derfus A. M.; Yang, M.; Hoffman, R. M.; Bhatia, S.; Sailor, M .J.; Ruoslahti E. Biomimetic amplification of nanoparticle homing to tumors *Proc. Natl. Acad. Sci. USA* 2007, 104, 932-936.
- [7] Pasqualini, R.; Ruoslahti, E. Organ targeting in vivo using phage display peptide libraries *Nature* 1996, 380, 364.
- [8] Hutchinson, J. N.; Muller, W.J. Transgenic mouse models of human breast cancer *Oncogene* 2000, 19, 6130-6137.
- [9] Karmali, P.P.; Kotamraju, V.R.; Kastantin, M.; Black, M.; Missirlis, D.; Tirrell, M.; Ruoslahti, E. Targeting of albumin-embedded paclitaxel nanoparticles to tumors. *Nanomedicine* 2009, 5, 73-82.
- [10] Zanuy, D.; Flores-Ortega, A.; Casanovas, J.; Curco, D.; Nussinov, R.; Aleman, C. The Energy Landscape of a Selective Tumor-Homing Pentapeptide *J. Phys. Chem. B* 2008, 112, 8692–8700.
- [11] Zanuy, D.; Curc6, D.; Nussinov, R.; Alem6n, C. Influence of the dye presence on the conformational preferences of CREKA, a tumor homing linear pentapeptide. *Biopolymers, Pept. Sci.* 2009, 92, 83-93.
- [12] Zanuy, D.; Flores-Ortega, A.; Jimenez, A.I. ; Calaza, M.I.; Cativiela, C.; Nussinov, R.; Ruoslahti, E.; Aleman, C. In Silico Molecular Engineering for a

Targeted Replacement in a Tumor-Homing Peptide *J. Phys. Chem. B* 2009, *113*, 7879–7889.

[13] Agemy L., Sugahara, KN, Kotamraju V.R., Gujraty, K., Girard, O.M., Kono, Y., Mattrey, R.F., Park, J.H., Sailor M. Jimenez, A.I., Cativieli, C, Zanuy, D Sayago, F.J., Aleman, C., Nussinov, R., Ruoslahti, E. “Nanoparticle-induced vascular blockade in human prostate cancer” *Blood*, *in press* (doi:10.1182/blood-2010-03-274258)

[14] Chakrabarti, P.; Pal, D. Interrelationships of side-chain and main-chain conformations in proteins *Prog. Biophys. Mol. Biol.* 2001, *76*, 1-102.

[15] Marraud, M.; Aubry, A. Crystal structures of peptides and modified peptides *Biopolymers* 1996, *40*, 45-83.

[16] MacArthur, M. W.; Thornton, J. M. Influence of Proline residues on protein conformation *J. Mol. Biol.* 1991, *218*, 397-412.

[17] Venkatraman, J.; Shankaramma, S.C.; Balaram, P. Design of Folded Peptides. *Chem. Rev.* 2001, *101*, 3131-3152.

[18] Toniolo, C.; Formaggio, F.; Kaptein, B.; Broxterman, Q.B. You are sitting on a gold mine! *Synlett* 2006, 1295-1310.

[19] Zanuy, D.; Jimenez, A.I.; Cativiela, C.; Nussinov, R.; Aleman, C. Use of Constrained Synthetic Amino Acids in β -Helix Proteins for Conformational Control *J. Phys. Chem. B* 2007, *111*, 3236-3242

[20] Zheng, J.; Zanuy, D.; Haspel, N.; Tsai, C.J.; Aleman, C.; Nussinov, R. Nanostructure Design Using Protein Building Blocks Enhanced by Conformationally Constrained Synthetic Residues. *Biochemistry* 2007, *46*, 1205-1218.

[21] Zanuy, D.; Rodriguez-Ropero, F.; Haspel, N.; Zheng, J.; Nussinov, R.; Aleman, C. Stability of Tubular Structures Based on β -Helical Proteins: Self-Assembled versus Polymerized Nanoconstructs and Wild-Type versus Mutated Sequences. *Biomacromolecules* 2007, *8*, 3135-3146.

[22] Rodriguez-Ropero, F.; Zanuy, D.; Casanovas, J.; Nussinov, R.; Aleman, C. Application of 1-Aminocyclohexane Carboxylic Acid to Protein Nanostructure Computer Design *J. Chem. Inf. Model.* 2008, *48*, 333-343.

[23] Zanuy, D.; Ballano, G.; Jimenez, A.I.; Casanovas, J.; Haspel, N.; Cativiela, C.; Curco, D., Nussinov, R.; Aleman, C. Protein Segments with Conformationally Restricted Amino Acids Can Control Supramolecular Organization at the Nanoscale. *J. Chem. Inf. Model.* 2009, *49*, 1623–1629.

- [24] Alemán, C. Conformational properties of alpha-amino acids disubstituted at the alpha-carbon. *J. Phys. Chem. B* 1997, *101*, 5046-5050.
- [25] Alemán, C.; Zanuy, D.; Casanovas, J. Cativiela, C.; Nussinov, R. Backbone Conformational Preferences and Pseudorotational Ring Puckering of 1-Aminocyclopentane-1-carboxylic Acid. *J. Phys. Chem. B* 2006, *110*, 21264-21271.
- [26] Revilla-Lopez, G.; Torras, J.; Jimenez, AI; Cativiela, C.; Nussinov, R.; Aleman, C. Side-Chain to Backbone Interactions Dictate the Conformational Preferences of a Cyclopentane Arginine Analogue *J. Org. Chem.* 2009, *74*, 2403–2412.
- [27] Baysal, C.; Meirovitch, H. Efficiency of simulated annealing for peptides with increasing geometrical restrictions. *J. Comput. Chem.* 1999, *20*, 1659–1670.
- [28] Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 1999, *314*, 141–151.
- [29] Lin, E.; Shell, M. S. Convergence and Heterogeneity in Peptide Folding with Replica Exchange Molecular Dynamics *J. Chem. Theory Comput.* 2009, *5*, 2062–2073.
- [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. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules *J. Am. Chem. Soc.* 1995, *117*, 5179.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* 2005, *26*, 1781-1802.
- [35] Case, D.A.; Darden, T.A.; Cheatham, III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R C.; Zhang, W.; Merz, K.M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K.F.; Paesani, F.;

Vanicek, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D.H.; Seetin, M.G.; Sagui, C.; Babin, V.; Kollman P.A.; *AMBER 10*; University of California, San Francisco, 2008.

[36] Cheng, X.; Cui, G.; Hornak, V.; Simmerling, C. Modified Replica Exchange Simulation Methods for Local Structure Refinement *J. Phys. Chem. B* 2005, *109*, 8220-8230

[37] Pastor, R.W.; Brooks, B.R.; Szabo, A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* 1988, *65*, 1409-1419.

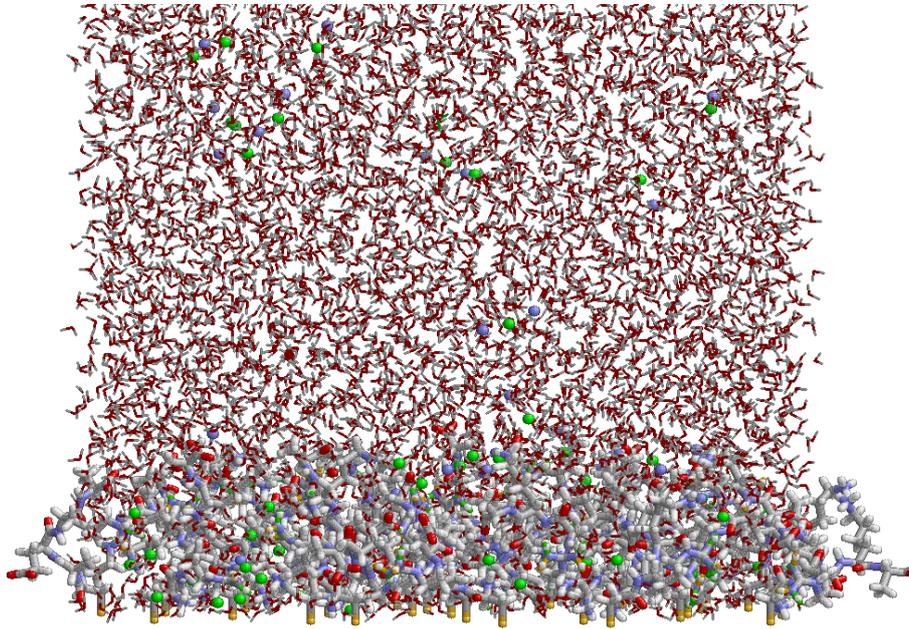
[38] Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model *J. Chem. Theory Comput.* 2006, *2*, 420-433.

[39] Hawkins, G.D.; Cramer C. J.; Truhlar, D.G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium *J. Phys. Chem.* 1996, *100*, 19824-19839.

[40] Tsui, V.; Case D A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations *Inc. Biopoly (Nucleic Acid Sci)* 2001, *56*, 275-291.

[41] Lin, E.; Shell, M. S. Convergence and Heterogeneity in Peptide Folding with Replica Exchange Molecular Dynamics *J. Chem. Theory Comput.* 2009, *5*, 2062-2073.

5.2. Atomistic modelling of peptides bound to a chemically active surface: conformational implications



5.2.1. Introduction

Nanotechnology approaches represent the ultimate frontier in diagnosis and treatment of diseases. Among different possible strategies, nanoparticle-based systems have taken the lead in tumor treatments.¹⁻⁹ Tumor chemical markers direct the nanoparticle-based device towards the malignant cells, being specially efficient if the molecular targets are receptors located in blood vessels grown around tumors.³⁻⁹ Within this context, nanoparticles recently proposed by Ruoslahti and co-workers deserves special attention since they home to tumors amplifying their own homing activity by stimulating platelet clots formation.⁸ These biomimetic nanoparticles were coated with a short linear peptide that recognizes clotted plasma proteins and selectively homes to tumors.⁸ This peptide, with sequence Cys-Arg-Glu-Lys-Ala (CREKA, Cys-Arg-Glu-Lys-Ala), was recently discovered by *in vivo* screening of phage-display peptide libraries^{4,10} for tumor-homing in tumor-bearing MMTV-PyMT transgenic breast cancer mice.¹¹ CREKA peptide labeled with the

fluorescent dye 5(6)-carboxyfluorescein (FMA) was detectable in human tumors from minutes to hours after intravenous injection, while it was essentially undetectable in normal tissues.¹²

On the other hand, enhancement of the homing activity of peptides or improvement of their resistance towards proteolytic activity require explicit knowledge of their conformational preferences, in particular of the bioactive conformation/s. Experimental determination of the bioactive conformation is frequently a difficult task due to the peptide inner flexibility. A feasible alternative is based on the use of computer simulation techniques, which allow explore the conformational preference of peptides and proteins. Recently, we determined the bioactive conformation of CREKA using a multiple conformational search strategy based on Molecular Dynamics (MD) simulations.^{13,14} Calculations were performed using a simplified description of different experimentally-tested environments. In one of these, a single CREKA molecule was bound to the surface of a nanoparticle, which was mimicked by attaching the sulfhydryl group of the Cys residue to the center of a square surface formed by 100 rigid van der Waals spherical particles (10×10). Clustering analyses of the results for this environment showed that the bioactive conformation of CREKA presents both a β -turn motif and strong electrostatic interactions involving the side chains of Arg, Glu and Lys.^{13,14} These results were used to suggest the synthesis of different CREKA analogs based on the chemical modification of single residues, which led to improve the homing peptide activity and to prove the turn shape of the bioactive conformation.¹⁵

Despite of the successful outcome of the bioactive conformation assessment, we were aware that a single peptide molecule covalently tethered to a surface is a very rough description of a nanoparticle delivery system. For example, the nanoparticles used Ruoslahti and co-workers were spherical (~50 nm of diameter) and each one was coated by a large number of peptide molecules, *i.e.* around 8000 peptides per particle.⁸ In order to improve the reliability of our molecular models and to overcome the serious limitations of atomistic MD simulations when dealing with surfaces with relatively high concentrations of flexible molecules covalently tethered (*i.e.* MD is very inefficient for dense systems since the motion of individual entities is easily hampered by steric hindrances), we recently developed a simple methodology to generate uncorrelated, energetically relaxed and representative structures.¹⁶

The new method¹⁶ consisted of a combination of an algorithm that generates representative atomistic microstructures of the coated surface, a procedure to construct energetically reliable configurations of the whole simulation box (including solvent molecules and ions, if present), and a relaxation method to minimize non-bonding interactions. We successfully investigated the conformation of CREKA peptides covalently tethered to a surface under diverse conditions, such as different density of peptides per Å² (ranged from $3 \cdot 10^{-4}$ to $1.67 \cdot 10^{-2}$) or ionic strengths. Clustering analysis of the peptide conformations revealed that the structure identified as bioactive is the most stable and populated cluster in all cases. However, our initial approach was very simple and did not account for the interactions between the surface and the tethered molecules (*i.e.* the surface was considered as an inert element), even although they may play a major role in biasing the conformation of the peptides.

In this work, we present a significant improvement of this new methodology by introducing a chemical descriptor for the surface and considering the peptide-surface interactions in the construction and relaxation algorithms. The strategy allows describe active surfaces made of crystalline materials, like metals, which are the frequently used in nanomedicine. Moreover, the method can be also applied to other surfaces (*e.g.* clays and carbon – both graphite and diamond allotropic forms –). The influence of the active surface on the peptide conformation has been examined on CREKA, results obtained in the present work being compared with those reported for a completely inert surface.¹⁶ However, it should be remarked that this is not intended to provide a deep and rigorous investigation of CREKA peptide, but to illustrate the potential utility of the new methodology in applications involving peptides attached to an active surface through a test case.

5.2.2. Methods

The principles followed to design the generation, construction and relaxation algorithms were described in detail in our previous work.^[16] Accordingly, in this section we only provide a brief background and describe the modifications that have been introduced to transform the inert surface into active.

The main steps followed to model the systems under study are summarized in Figure 5.2.1. The method consists of a three steps procedure. First, molecules tethered to the surface are generated without atomic overlaps or, in the case of

systems with a very high density of molecules per \AA^2 , with atomic overlaps lower than a given threshold. In the second step the simulation box is constructed by filling it with the counter-ions (in the case of charged molecules), solvent (if desired) and salt molecules (if the ionic strength is increased). In order to effectively eliminate the overlaps created in the first step for systems with a very high density of peptides per \AA^2 , the positions of the tethered molecules and the counter-ions are relaxed prior to filling the simulation box with explicit solvent molecules. Finally, the conformation and position of the molecules attached to the surface as well as the position of the ions and solvent molecules are relaxed.

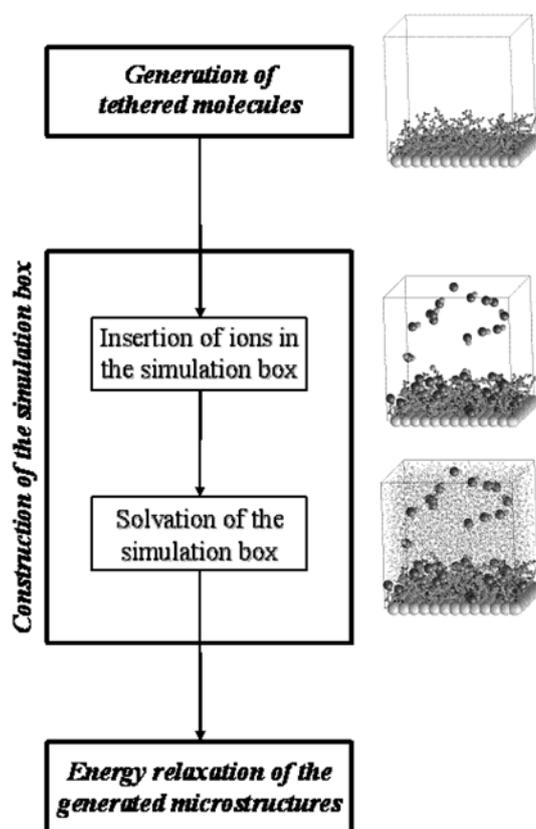


Figure 5.2.1. General flowchart of the theoretical strategy used to build peptides tethered to active surfaces.

Generation Algorithm. The generation algorithm is inspired in previously reported algorithms which provide atomistic models of amorphous polymers.^[17-19] The positions of the atoms linked to the surface are randomly selected with the following restriction: the minimum distance between two of such atoms must be larger than $2.1 \cdot R$, R being the van der Waals radius. Each tethered chain is built following the strategy previously developed for comb-like polymers,^[18] which is based on two principles:

(a) The radii of atoms are reduced multiplying them by a factor $\lambda < 1$. The use of scaled radii allows increase the computational efficiency, especially for dense systems, without produce any detriment in the reliability of the microstructures obtained at the end of the process (*i.e.* unfavorable van der Waals interactions are easily relaxed).

(b) Atomic positions are generated residue-by-residue using a stochastic procedure. For each residue the backbone is generated before the side chain.

It should be noted that no definition of the peptide-surface interactions is required in this step. This is because the only energy evaluation performed during the generation of the peptide molecules, which is used to accept or reject atomic positions in (b) through a Monte Carlo (MC) criterion, is exclusively based on short-range interactions with the immediate neighbors.

Construction of the Simulation Box. The simulation box is filled using a two-step process. The first involves the insertion of both counter-ions and ions forming salt molecules while the box is filled with solvent molecules in the second. The initial positions of the ions inside the simulation box are randomly selected but avoiding atomic overlaps with the previously generated peptides. These positions are further relaxed by applying random displacements moves, which are accepted if the energy of the configuration obtained after the displacement of the ion is lower than that of the configuration before such displacement. Before adding the solvent molecules, tethered peptides and ions are relaxed together, 80% of the moves corresponding to peptides. Finally, each solvent molecule is positioned at a position that produces attractive interaction energy with both the previously generated solvent molecules and the peptides. No scaling of the van der Waals radii is used in this algorithm.

In this case the relaxation of both ions and peptides is performed considering the influence of the surface. For this purpose, the surface is parceled in a grid of

points that represent the atom nuclei forming the ultra structure of the surface. The position of these points is equivalent to those atoms in the corresponding crystal structure. The grid of points defines the reference positions to calculate the interaction energy between the molecules located in the simulation box (*i.e.* peptides, solvent molecules and ions) and the surface. In this work we have considered a metallic surface, the evaluation of the interaction energy being described through a conventional Lennard-Jones potential. Thus, the contribution of the surface is evaluated by computing atom pair distances between any particle in the simulation box and all the surface points that are within a previously defined cutoff distance.

Relaxation. The energy of the peptides tethered to the surface, the ions and the solvent molecules is minimized through a relaxation algorithm, which is applied using periodic boundary conditions at the x - and y -directions (*i.e.* those used to define the surface). In order to avoid an erroneous description of the system in the z -direction, the c axis of the simulation box is divided into two different regions. All the solvent molecules contained in the simulation box that are placed below a specific distance to the c axis edge are allowed to move during the relaxation process, while the solvent molecules above such distance are kept at fixed positions. Each relaxation cycle consists of the following three steps:

(a) The interaction energy is evaluated for each movable solvent molecule. The positions of the molecules with higher interaction energies are varied by introducing a random displacement and a random rotation. The moves are accepted or rejected using a typical MC criterion.

(b) The positions of all the ions are improved by applying random displacements that are accepted when the energy decreases.

(c) The values of the dihedral angles (90% of the moves) and the position (10% of the moves) of the tethered peptides are relaxed applying the same procedure that was used for the construction of the simulation box. The only difference is that in this case several randomly chosen dihedral angles move simultaneously.

As in the construction algorithm, the influence of the surface has been considered in the relaxation algorithm. The procedure used to define the position of the metallic atoms in the surface and the potential employed to define the

interaction with all the particles contained in the simulation box were identical to those described above.

Molecular models and simulation details

The theoretical strategy presented in this work was applied to explore the conformational space of CREKA peptides tethered to a metallic surface (see below) and surrounded by water molecules. In all cases the sulfur of the Cys was used to form the covalent linkage to the surface. The dimensions of the simulation box were $a=b=60$ Å and $c=90$ Å, where the regions along the c -axis defining the movable and fixed solvent molecules during the relaxation were defined by $c_1=60$ Å and $c_2=30$ Å, respectively. The MC criterion involved in the generation and relaxation algorithms was applied considering a temperature of 300 K.

The molecular model was built using the density of tethered CREKA peptides (σ , in peptides per Å²) experimentally used by Roushlati and co-workers: $\sigma=0.083$ peptides·Å², which is equivalent to thirty peptides attached to the surface. The system was completed with ninety counter-ions (320 Na⁺ + 60 Cl⁻) and 9621 water molecules. The chemical nature of the surface was represented using gold (Au), which is frequently used as sustaining metal for peptides.

The energy of the peptides was calculated using the following analytical potential function:

$$E = \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{nonbonded} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{R_{ij}} \right] \quad \text{Eq.1}$$

where the first sum represents a series expansion for the torsional term followed by the Lennard-Jones and electrostatic terms. Non-bonding interactions of atoms connected by three atoms (1-4 interactions) were treated as in the AMBER force-field,^[20] applying a scaling factor of 0.5. The energy involving the solvent molecules and the ions was calculated using the non-bonding terms of Eqn (1), while the Lennard-Jones was the only term used to evaluate the interaction between the surface and the rest of species. Non-bonding interactions were evaluated using a cutoff of 14.0 Å. Force-field parameters for the CREKA peptide and the ions were taken from the AMBER libraries,^[20,21] while OPLS van der Waals parameters for Au particles^[22] were used to compute the interaction between the surface and the rest of species ($\sigma_{Au}=1.789$ Å and $\epsilon_{Au}=0.193$ kcal·mol⁻¹). The Au nuclei were placed at the (100) lattice positions of a cubic F unit cell with crystallographic parameter

$a=2.93 \text{ \AA}$. Although this parameter is 1.4% larger than the experimental one ($a=2.89 \text{ \AA}$), it was chosen for consistency with the van der Waals parameters of Au (*i.e.* these parameters predict that the closest distance is 2.93 \AA).^[22] Finally, water molecules were represented using the TIP3P model.^[23] Bond lengths and bond angles, which were also taken from the AMBER libraries,^[20] were kept fixed. A total of 1000 atomistic microstructures were generated and relaxed.

The results obtained for this system, hereafter denoted Au-S/30, have been compared with those previously obtained for an identical molecular model with exception of the surface,^[16] which was considered inert (*i.e.* the interaction between the surface and the rest of chemical species was neglected). In next sections the system with an inert surface has been denoted Inert-S/30. It should be noted that the computational resources required to simulate Au-S/30 were one order or magnitude larger than those used for Inert-S/30, which were also significant due to large density of peptides per \AA^2 . This is because the number of interactions between atom pairs increased considerably when the surface participates in the construction and relaxation algorithms.

5.2.3. Results and discussion

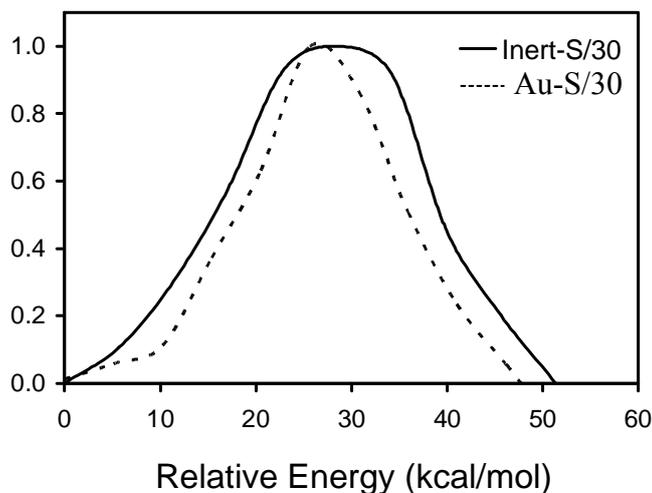


Figure 5.2.2. Normalized distribution of relative energies (in kcal/mol) for the microstructures of Au-S/30 and Inert-S/30 models obtained using the computational strategy presented in Figure 5.2.1..

Figure 5.2.2. depicts the distribution of energies associated to the relaxed microstructures produced for Au-S/30 and Inert-S/30 models. The shape of the two distributions fits to a Gaussian function, which reflects the reliability of our theoretical approach. It is well-known that the shape of the distribution of minima for a given model resembles a Gaussian functions when the conformations of the can be described in terms of rotational isomers.^[24,25] Although the width of the Gaussian function depends on the number generated microstructures, such number was identical for the two surfaces under study. Furthermore, as was demonstrated in our previous studies on amorphous polymers,^[17-19] the generation method used in this work tends to provide representative microstructures rather than random ones. This feature guarantees that the left side of the Gaussian distribution is well described after generating a relatively low number of microstructures. The Boltzmann distributions displayed in Figure 5.2.2. clearly depend on the surface, the function becomes more narrow when the surface is allowed to interact with the chemical species contained in the simulation box. Thus, in general terms, it can be state that the conformational flexibility is smaller for Au-S/30 than for Inert-S/30.

The conformational features of the CREKA peptide in Au-S/30 and Inert-S/30 models have been analyzed by considering each molecule as an independent case. A list of unique peptide conformations was constructed for each model by comparing all the molecules for all the produced microstructures. The list was organized by ranking all the unique peptide conformations according to their energies. Previously listed conformations were discarded from the list. A unique peptide conformation was identified if it did not share a set of selected structural parameters, which involve virtual backbone dihedral angles and intramolecular interactions. Virtual dihedral angles were defined considering the α -carbon atoms in the CREKA peptide, whereas the presence of polar interactions was accepted on the basis of geometric criteria: (a) hydrogen bonds: distance $H \cdots O$ shorter than 2.50 Å and angle $\angle N-H \cdots O$ larger than 120.0°; (b) salt bridges: distance between the geometric centers of the interacting groups shorter than 4.50 Å. The interaction pattern was also used to discriminate between side chain and main chain interacting groups. Two conformations were considered different if they differ at least in one virtual dihedral angle by more than 60° or in one of the interactions counted. For each model all the conformations categorized as unique were clustered according to

a criterion based on the presence of the interactions mentioned above: salt bridges and hydrogen bonds.

The number of unique conformations (*clusters*) and the percentage of structures present in each cluster for the two models are depicted in Figure 5.2.3.. These percentages refer to the total amount of peptide conformations present in the 1000 microstructures (*i.e.* 30 peptides \times 1000 microstructures= 30000 conformations) produced for the Au-S/30 and Inert-S/30 models. Figure 5.2.3. also includes the conformational preferences of a single CREKA molecule attached to a surface (hereafter referred as MD-S/1), which were previously determined by MD simulations.^[13] For this purpose, the results derived from such simulations have been clustered in this work using the procedure discussed above. The different environments in which CREKA was simulated led to a distribution of conformations with 45, 40 and 89 clusters for Au-S/30, Inert-S/30 and MD-S/1, respectively, although only a small number of such clusters contained a significant number of structures. Specifically, the number of clusters that involve more than 1% of the total amount of conformations was 11, 16 and 20, respectively. Moreover, this number ranges from 2 to 5 if the threshold is increased to 10%. The results displayed in Figure 5.2.3. indicate that the conformation of CREKA is significantly restricted when the effects of the neighboring molecules, which are associated to the density of peptide per \AA^2 , are included in the model. On the other hand, independently of nature of the surface the results of the clustering are very similar, suggesting that the conformation of CREKA is dominated by the intramolecular interactions rather than by the intermolecular ones. Moreover, further analysis of each cluster shows that the conformation proposed as bioactive is predicted among the most favored ones independently of the characteristics of the surface (see below). However, significant differences have been observed depending on the activity of the surface since the C termini of the peptides may be attracted by it.

The results presented in Figure 5.2.3. clearly narrow the conformational characterization of CREKA to the most representative clusters obtained for the different models. In the case of Au-S/30 and Inert-S/30 more than 80% of the total amount of produced peptide conformations was clustered within the first seven clusters. The energy difference between the lead conformations of the most and least favored of such clusters was of only 3.0 kcal mol⁻¹. Accordingly, the most

relevant conformational descriptors have been focused on those seven lead conformations. On the other hand, these results have been compared with the features observed using an isolated peptide tethered to a surface (MD-S/1). In the latter case the first ten clusters concentrate 80% of the produced conformations.

Figure 5.2.4. depicts the Ramachandran plot of the Arg, Glu and Lys residues in the most populated clusters of Au-S/30 and Inert-S/30 models. As it can be seen, most of the conformations clustered around a narrow region of the Ramachandran map, this feature being particularly evident for the Glu. Indeed, this trend was already described as one of most remarkable conformational characteristics of CREKA peptide, since the bioactive ensemble was based on tight-turns nucleated around Glu residue.^[13] On the other hand, the conformational preferences of MD-S/1 are similar to those of Au-S/30 and Inert-S/30 suggesting again that the conformational preferences exhibited by the central segment of CREKA are driven by intramolecular interactions. For the three studied systems the most relevant differences are found at the C-termini of the peptide (Lys), while the central core (Arg and Glu) tends to present very low conformational variability (*i.e.* the ϕ, ψ values observed for Au-S/30 and Inert-S/30 are close to those obtained for MD-S/1).

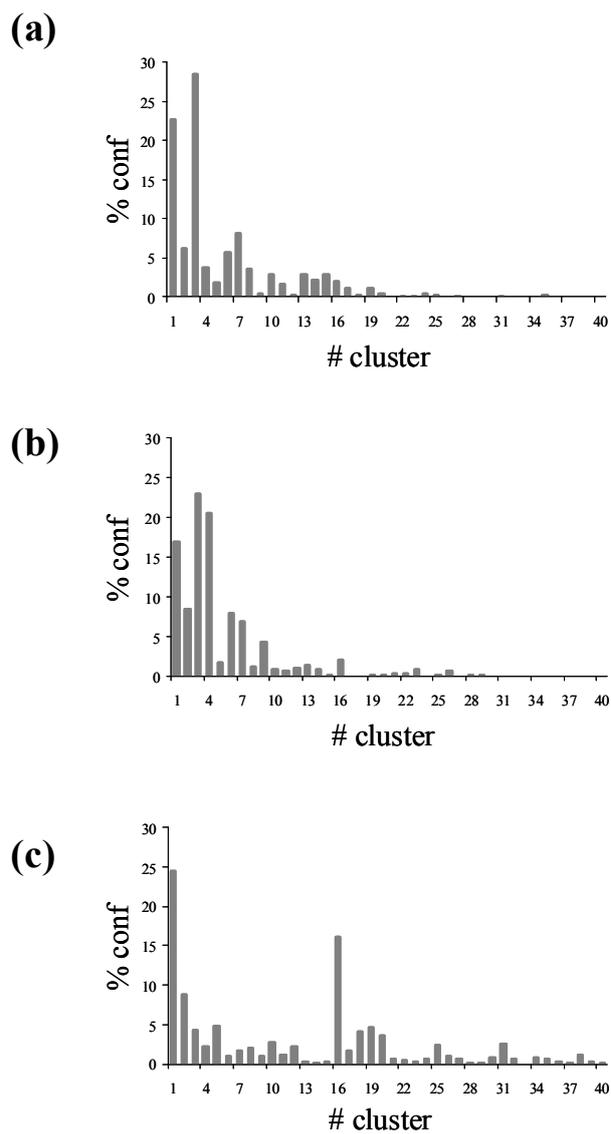


Figure 5.2.3. Percentage of structures in each of the 40 and 45 clusters found in the simulation conditions corresponding to (a) Inert-S/30 and (b) Au-S/30, respectively. Clusters are ordered considering the energy of the lead conformation (from the lowest to the highest), which corresponds to the most stable microstructure within the cluster. Results derived from MD simulations on a single CREKA molecule attached to a surface (reference 12) are displayed in (c) for comparison (MD-S/1 set in text).

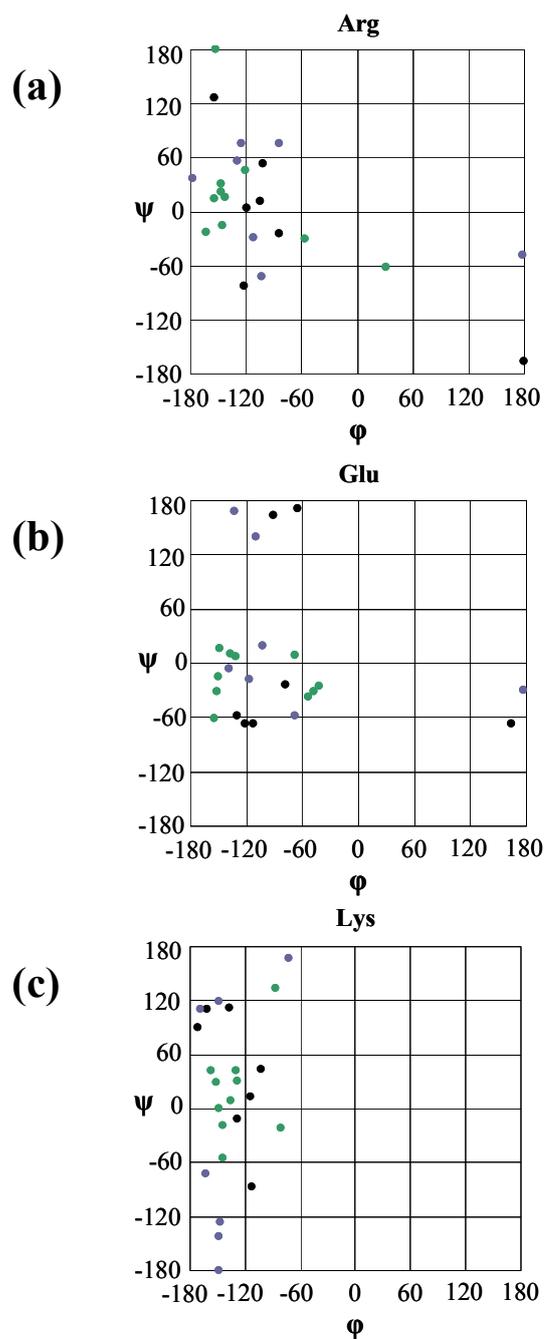


Figure 5.2.4. Ramachandran plot distribution for the three central residues of CREKA, considering the most representative minimum energy structures for the **Inert-S/30** (*black dots*), **Au-S/30** (*green dots*) and **MD-S/1** (*blue dots*) models.

In spite of Figure 5.2.4. suggests that the conformational preferences of the CREKA are independent of the surface, there are some differences among the conformations produced for Au-S/30 and Inert-S/30 that deserve consideration. The organizations obtained for the C-termini of the peptide when the surface participates or not in the interaction energy are significantly different. Thus, in Au-S/30 the edge of the peptide that is far from the group directly tethered to the surface (*i.e.* Cys side chain) is able to bend, interacting not only with other peptide chains but also with the surface itself. Accordingly, the metallic surface partially affects the conformational properties of the tethered peptides. Specifically, unspecific peptide···surface van der Waals interactions induce chains to move away when the intrinsic conformational tendencies of the peptides allow them to bend and intermolecular interactions do not reduce their flexibility (Figure 5.2.5.a). In contrast, the lack of peptide···surface interactions facilitates that a large amount of molecules lay on the surface of the simulated barrier, as is reflected in Figure 5.2.5.b for Inert-S/30. However, it should be emphasized that for Au-S/30 the interactions with the surface are not strong enough to revert decisively the intrinsic conformational preferences of the peptide core, as is evidenced in Figure 5.2.4. for the Arg and Glu residues.

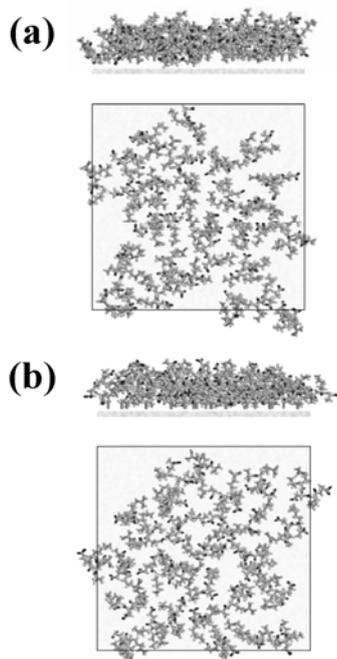


Figure 5.2.5.. Axial (top panels) and equatorial (bottom panels) projections of a representative microstructure for the (a) **Au-S/30** and (b) **Inert-S/30** models. Ions, solvent molecules and the *c*-axis have been omitted to clarify the picture.

In order to extend the analysis over the conformational particularities of each studied model, the structural correlation among all the low energy conformations has been established. Thus, the produced conformations have been compared transversally, beyond the theoretical methodology used to obtain them or the characteristics of the surface. Specifically, the backbone Root Mean Square Deviation (RMSD) of the last four residues has been computed for all possible pairs of conformations. The results of this analysis, which are plotted in Figure 5.2.6., confirm the influence of the surface activity in the more favored arrangements of the peptide molecules. Thus, inspection of the similarities among the different models indicates that, in general, the conformations obtained for MD-S/1 are closer to those produced for Au-S/30 than for Inert-S/30. The Au-S/30 conformations present 48% of the computed chain pair RMSDs below 2.0 Å, while this percentage reduces to 35% for the Inert-S/30 conformations. As it was mentioned above, the MD-S/1 model allowed consider peptide···surface interactions through a surface descriptor (*i.e.* the spheres used to define the nanoparticle were allowed to interact with the peptide molecule through a Lennard-Jones potential, even though they did not represent any realistic material).^[13] Accordingly, the two models with active surfaces reached similar molecular conformations, even though the molecular environments were different (*i.e.* isolation and high density of molecules per Å² for MD-S/1 and Au-S/30, respectively).

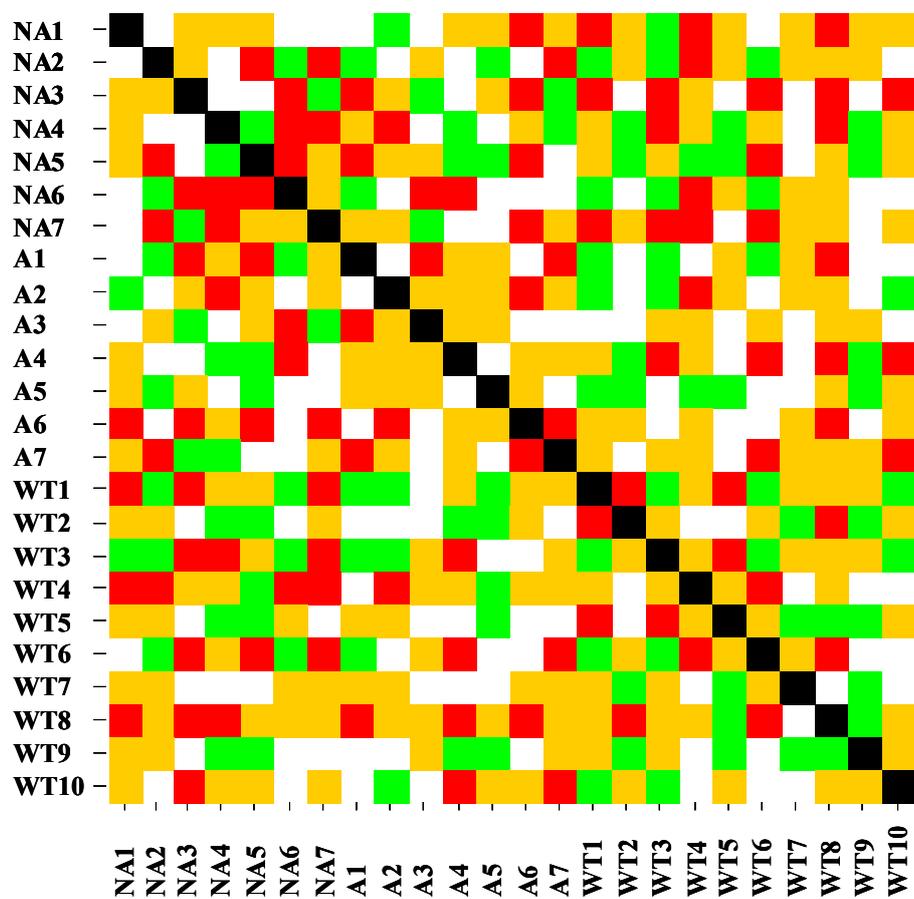


Figure 5.2.6.. Conformational correlation expressed through the RMDS of the backbone atoms for the most representative conformations produced for the three models. Filling colors represent different degrees of similarity referred to the value of RMSD: less than 1.3 Å *green*; higher than 1.3 Å and lower than 2.0 Å *white*; higher than 2.0 Å and lower than 3.0 Å *orange*; and higher than 3.0 Å *red*. The labels NA#, A# and WT# refer to the conformations of the more populated clusters produced for Inert-S/30, Au-S/30 and MD-S/1, respectively.

5.2.4. Conclusions

The previously developed strategy has been extended by including a potential that describes the attractive and repulsive interactions between the surface and the rest of chemical species in the construction and relaxation algorithms. The surface descriptor is based on a grid of points that represents the geometrical organization of atoms or molecules in a crystalline structure. Accordingly, the new strategy is able to produce energetically representative structures of flexible peptides tethered to a surface with chemical identity. Although the description of the systems modeled using this new procedure is more realistic, the computational resources required for simulations also increase significantly.

In order to examine the influence of the surface activity on the conformational preferences of peptide, a system formed by 30 CREKA molecules covalently linked to an Au surface of $60 \times 60 \text{ \AA}^2$ has been simulated using this new strategy. Results have been compared with those obtained for: (1) a system with identical chemical composition with exception of the surface, which is not able to interact with peptides, solvent molecules nor ions (*i.e.* inert surface); and (2) a system formed by a single peptide molecule attached to a van der Waals surface, which was previously studied using conventional MD simulations. .

Analysis of the results indicate that the omission of the chemical definition of the surface produce spatial arrangements for the peptides that might preclude the proper organization of the side chains. Specifically, the Au surface affects the orientation of the CREKA molecules by influencing the conformation of the edge residues and the subsequent organization of the ionized side chains. Even though the intrinsic preferences of the peptide main chain remain practically unaltered by the active surface, a proper description of the ionized side chains requires a reliable definition of the interactions with the surface. Thus, an erroneous representation of the ionized side chains organization would lead to a misunderstanding of the nanoparticle function hampering potential further chemical improvements.

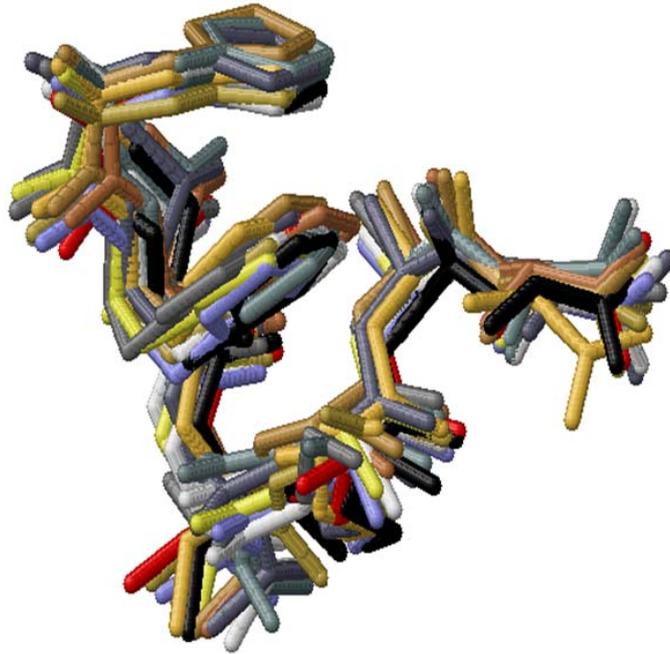
5.2.5. Bibliography

- [1] Weissleder R., Bogdanov A.Jr., Neuwelt E.A., Papisov M. Long-circulating iron oxides for MR Imaging. *Adv. Drug Delivery Rev.* 1995;16:321-334.
- [2] Sinek J., Frieboes H., Zheng X, Cristini V. Two-dimensional chemotherapy simulations demonstrate fundamental transport and tumor response limitations involving nanoparticles. *Biomed. Microdev.* 2004;6:297-309.
- [3] Hoffman J.A., Giraud E., Singh M., Zhang L., Inoue M., Porkka K., Hanahan D., Ruoslahti E. Progressive vascular changes in a transgenic mouse model of squamous cell carcinoma. *Cancer Cell.* 2003;4:383-391.
- [4] Oh P., Li Y., Yu J., Durr E., Krasinska K.M., Carver L.A., Testa J.E., Schnitzer J.E. Subtractive proteomic mapping of the endothelial surface in lung and solid tumors for tissue-specific therapy. *Nature* 2004;429:629-635.
- [5] Ruoslahti E. Specialization of tumour vasculature. *Nat. Rev. Cancer* 2002;2:83-90.
- [6] Akerman M.E., Chan W.C., Laakkonen P., Bhatia S.N., Ruoslahti E. Nanocrystal targeting in vivo. *Proc. Natl. Acad. Sci. USA* 2002;99:12617-12621.
- [7] Cai W., Shin D.W., Chen K, Gheysens O., Cao Q., Wang S.X., Gambhir S.S., Chen X. Peptide-labeled near-infrared quantum dots for imaging tumor vasculature in living subjects. *Nano Lett.* 2006;6:669-676.
- [8] Simberg D., Duza T., Park J.H., Essler M., Pilch J., Zhang L., Derfus A.M., Yang M., Hoffman R.M., Bhatia S., Sailor M J, Ruoslahti E. Biomimetic amplification of nanoparticle homing to tumors. *Proc Natl Acad Sci USA* 2007;104,932-936.
- [9] Riehemann K., Schneide SW., Luger T.A., Godin B, Ferrari M., Fuchs H. Nanomedicine – Challenge and perspectives. *Angew. Chem. Int. Ed.* 2009;48,872-897..
- [10] Pasqualini R., Ruoslahti E. Organ targeting in vivo using phage display peptide libraries. *Nature* 1996;380:364-366.
- [11] Hutchinson J.N., Muller W.J. Transgenic mouse models of human breast cancer. *Oncogene* 2000;19:6130-6137.
- [12] Ruoslahti E. Personal communication (unpublished results) 2007.
- [13] Zanuy D., Flores-Ortega A., Casanovas J., Curcó D, Nussinov R, Alemán C. The energy landscape of a selective tumor-homing pentapeptide. *J. Phys. Chem. B* 2008;112:8692-8700.

- [14] Zanuy D., Curcó D., Nussinov R., Alemán C. Influence of the dye presence on the conformational preferences of CREKA, a tumor homing linear pentapeptide. *Biopolymers (Pept. Sci.)* 2009;90:83-93.
- [15] Agemy L., Sugahara K.N., Kotamraju V.R., Gujraty K., Girard O.M., Kono Y., Mattrey R.F., Park J.H., Sailor M., Jimenez A.I., Cativiela C, Zanuy D., Sayago F.J., Aleman C., Nussinov R., Ruoslahti E. Nanoparticle-induced vascular blockade in human prostate cancer” *Blood* 2010; in press (doi:10.1182/blood-2010-03-274258).
- [16] Curcó D., Zanuy D., Nussinov R., Alemán C. A simulation strategy for the atomistic modeling of flexible molecules covalently tethered to rigid surfaces: Application to peptides. *J. Comput. Chem.* 2010; in press.
- [17] Curcó D., Alemán C. Simulation of dense amorphous polymers by generating representative atomistic models. *J. Chem. Phys.* 2003;119:2915-2922.
- [18] Curcó D., Alemán C. A theoretical strategy to provide atomistic models of comblike polymers: A generation algorithm combined with configurational bias Monte Carlo. *J. Chem. Phys.* 2004;121:9744-9752.
- [19] Curcó D., Alemán C. A computational tool to model the packing of polycyclic chains: Structural analysis of amorphous polythiophene. *J. Comput. Chem.* 2007;28: 1743-1749.
- [20] 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.
- [21] 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 second generation force field for the simulation of proteins, nucleic Acids, and organic molecules. *J. Am. Chem. Soc.* 1995;117:5179-5197.
- [22] Iori F., Di Felice R. Molinari E., Corni S. GolP: An atomistic force-field to describe the interaction of proteins with Au(111) surfaces in water. *J. Comput. Chem.* 2009;30:1465-1476.
- [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] Pérez J.J.; Villar H.O., Loew G.H. Characterization of low-energy conformational domains for Met-enkephalin *J. Comput. Aided Mol. Design* 1992;6:175-190.

[25] Pérez J.J., Villar H.O., Arteca G.A. Distribution of conformational energy minima in molecules with multiple torsional degrees of freedom. *J. Phys. Chem.* 1994;98:2318-2324.

5.3. Self-Assembly of a Designed Amyloid Peptide Containing the Functional Thienylalanine Unit



5.3.1. Introduction

Peptide-based systems are the focus of immense research activity, for applications in bionanotechnology. The palette of natural amino acids already offers an impressive scope for the design of novel structures and functionalities. Beyond this, the incorporation of non-natural amino acids into peptides and proteins has attracted huge interest, with the potential to develop novel biomaterials with properties distinct from those available in the natural world. Tirrell and co-workers have pioneered the use of protein engineering to incorporate non-natural amino acids into proteins via the use of stop (nonsense) codons or 4-base codons.^{1,2} The non-natural amino acids are coupled to tRNA using a suppressor tRNA that is not recognized by the host synthetase enzymes that couple the amino acids to tRNA. This method has been used to incorporate analogues of many amino acids^{1,3} including β -

2-thienylalanine^{3,4} and β -3-thienylalanine,⁵ abbreviated 2-Thi and 3-Thi, respectively. Incorporation of thienylalanine residues is of considerable interest in the development of polypeptides with conductivity, arising from the creation of extended conjugated electronic systems.⁵ A recent computational investigation, using quantum mechanical/molecular mechanics (QM/MM) calculations and molecular dynamics (MD) simulations, has explored delocalization and π -stacking in peptides containing 3-Thi.⁶ The calculated peptides were based on mutations (isoleucine and/or valine substitutions with 3-Thi) of peptides (based on a motif from *E. coli* galactoside acetyltransferase) that MD simulations indicated form β -helices.^{7,8} Charge transfer properties were also investigated.⁶

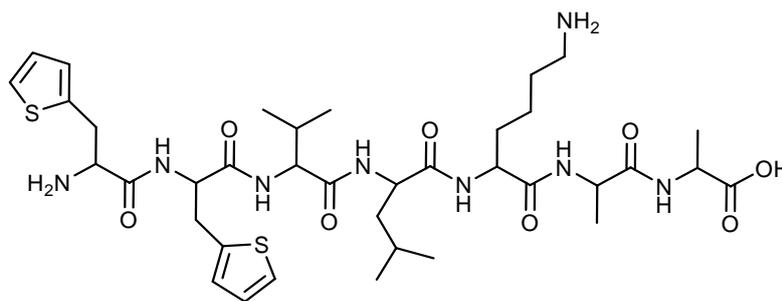
Several groups have reported the linking of thiophene moieties into peptide-based hybrids. Klok and co-workers first reported such conjugates comprising a head-to-tail coupled tetra(3-hexylthiophene) linked to the silk-mimic peptide GAGAG.⁹ Antiparallel β -sheet structures were observed, pointing to the ability of the peptide to guide the self-assembly of the thiophene units. The self-assembly into fibrils of oligopeptides containing a dithiophene central unit has been examined.¹⁰ This was shown to be due to π - π stacking interactions, and strong photoluminescence was observed, ascribed to possible twisted H-aggregates. A tetra-oligothiophene has been conjugated at both ends to a G(TV)3G(aF)G peptide sequence (aF) azidophenylalanine) attached at the N termini to a PEG chain to enhance solubility.¹¹ The (TV)3 repeat favors the formation of β -sheets (to disrupt this during synthesis, an O to N-acyl transfer switch strategy was used). Fibril formation in the hybrid system was observed; however, electronic properties were not reported.

In the present paper, we exploit a peptide motif previously developed in our group as a basis to present 2-Thi in β -sheet based fibrils and nanotubes. In contrast to previous work, the thienyl group is incorporated within the peptide sequence within a non-natural amino acid. The peptide is based on AAKLVFF, which contains the KLVFF core motif A β (16-20) from the amyloid β peptide, extended at the N terminus by two alanine residues. It has previously been shown that this peptide forms twisted fibrils in water¹² and nanotubes in methanol,¹³ and the structure of these has been investigated in detail using transmission electron microscopy, X-ray scattering, and multiple spectroscopic methods^{14,15} as well as computer simulation.¹⁵ Here, we

investigate the self-assembly of (2-Thi)(2-Thi)VLKAA in which the terminal phenylalanine residues are replaced by 2-Thi and the sequence is reversed (i.e., N and C termini are switched). The sequence is reversed because in FFVLKAA the large distance between the negatively charged C terminus and the π -electron density of side phenyl groups at the F residues would be expected to lead to intramolecular folding (β -turn structure), promoting the formation of π - π interactions between the two aromatic groups. This arrangement is expected to be preserved upon substitution of F by 2-Thi, as indicated by MD simulations, which is essential to induce intramolecular electron transfer processes. It should be noted that this is not the case for AAKLVFF in which β -sheet self-assembly is observed even in dilute solution.¹⁵ On the other hand, the closeness of the 2-Thi residues to the C terminus in (2-Thi)(2-Thi)VLKAA is expected to enhance the stacking of the two thienyl groups. The substitution of aromatic residues was expected not to disrupt β -sheet formation; however, the 2-Thi residues are expected to confer interesting π -stacking and, potentially, conductivity properties.

5.3.2. Methods

Fmoc-alanine-OH, Fmoc-lysine(Boc)-OH, Fmoc-leucine-OH and Fmoc-valine-OH, Fmoc-alanine-Wang resin (0.72 mmol g⁻¹ substitution), HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) were purchased from Novabiochem (UK). Fmoc-L-2-Thienylalanine-OH, trifluoroacetic acid (TFA), piperidine and triisopropylsilane were purchased from Sigma-Aldrich. HOBt/DMF (a mixture of 1-hydroxybenzotriazole and dimethylformamide), DIEA/NMP (a mixture of diisopropylethylamine and N-methylpyrrolidone), and NMP were obtained from Applied Biosystems(UK). Water (HPLC grade), acetonitrile (HPLC grade) and diethyl ether were purchased from Fisher Scientific (UK).



Scheme 1. Molecular structure of (2-Thi)(2-Thi)VLKAA

The peptide H-(2-Thi)(2-Thi)VLKAA-COOH (Scheme 1) was synthesized by solid phase methods using standard FastMoc chemistry [Fmoc (9-fluorenylmethyloxycarbonyl) protecting group and activation by HBTU /HOBt] on a 0.25 mmol scale with a fully automated peptide synthesizer (433A Applied Biosystems), which allowed for direct conductivity monitoring of Fmoc deprotection.

Peptide H-(2-Thi)(2-Thi)VLKAA-COOH was assembled from the C-terminus toward the N-terminus and was attached to the resin at the C-terminus by the α -carbonyl group of the amino acid. According to the standard FastMoc protocol, the first step of the reaction was to remove the Fmoc protecting group from the preloaded amino acid using a solution of piperidine in NMP. The next step was activation of the carbonyl group of the new amino acid (dissolved in NMP) using HBTU (dissolved in HOBt, DIEA, and DMF). The activated amino acid was transferred from the activation vessel to the reaction vessel containing the previously deprotected amino terminal group of the peptide chain, and coupling was performed. To obtain the highest coupling efficiency, four times excess of each amino acid was used in 0.25 mmol cycles. Peptide attached to the solid support was obtained from the synthesizer. In the cleavage step, the peptide attached to the resin was treated with a mixture of 95% TFA, 2.5% triisopropylsilane, and 2.5% water, and was stirred at room temperature for approximately 4 h, followed by filtration. During the cleavage the side chain protecting groups (Boc) were removed by TFA. The obtained peptide solution was precipitated in cold diethyl ether and the crude product was separated by centrifugation and decanting the supernatant. The crude peptide was redissolved in HPLC grade water/acetonitrile co-solvent (1:1) and subjected to reverse phase HPLC (Perkin Elmer 200) for purification. The following conditions were employed: a mobile phase of HPLC grade water and acetonitrile with 0.1% trifluoroacetic acid

(TFA) was used. During HPLC process the acetonitrile portion increased linearly from 0% to 90% over 20 min and then decreased linearly to 0% over 10 min. A sample injection (injected with Perkin Elmer Series 200 Autosampler) of 100-150 μL was run on a C18 semi-preparative column (Macherey-Nagel) for 30 min with flow rates 4 ml/min at 35°C (Perkin Elmer Series 200 Peltier Column Oven), where a UV/vis detector (Perkin Elmer Series 200 UV/VIS Detector) monitored the sample elution at 254 nm. The fractions of (2-Thi)(2-Thi)VLKAA were collected, followed by lyophilization to give a white solid (RP-HPLC retention time: 11.97 min). ESIMS m/z for $\text{C}_{37}\text{H}_{58}\text{N}_8\text{O}_8\text{S}_2$ $[\text{M}+\text{H}]^+$ calcd 807.38, found $[\text{M}+\text{H}]^+$, 807.38; $[\text{M}+2\text{H}]^{2+}$ /2 calcd 404.20, found 404.20. ^1H NMR (400 MHz methanol- d_4 , ppm): 0.87 (m, 12H), 1.30 (dd, $J = 11.6$ Hz, 7.2 Hz, 6H), 1.37 (q, $J = 7.2$ Hz, 2H), 1.47-1.64 (m, 6H), 1.72 (m, 1H), 1.94 (sep, $J = 7.2$ Hz, 1H), 2.82 (t, $J = 7.2$ Hz, 2H), 3.10 (dd, $J = 21.6$ Hz, 8.4 Hz, 1H), 3.23 (m, 2H), 3.40 (dd, $J = 15.4$ Hz, 4.8 Hz, 1H), 3.98 (dd, $J = 8.4$ Hz, 4.4 Hz, 1H), 4.10 (m, 1H), 4.21- 4.35 (m, 4H), 4.66 (m, 1H), 6.82 (m, 2H), 6.90 (m, 2H), 7.13 (t, $J = 3.2$ Hz, 1H), 7.25 (dd, $J = 4.4$ Hz, 2 Hz, 1H), 8.02-8.06, 8.11-8.17 and 8.26-8.28 [NH, NH_2 10 H].

Fourier Transform Infra-red (FTIR) spectroscopy. Spectra were measured on a Nicolet Nexus spectrometer with DTGS detector. Solutions of (2-Thi)(2-Thi)VLKAA in D_2O (1 wt %) or MeOD (1 wt%) were sandwiched between two CaF_2 plate windows (spacer 0.006 mm) Spectra were scanned 128 times over the range of 4000-900 cm^{-1} . Spectra of dried films were obtained with an Attenuated Total Reflectance (ATR) cell after deposition of a 0.03% water solution on a ZnSe crystal (128 scans, 4700-700 cm^{-1}).

Circular Dichroism (CD). Spectra were recorded on a Chirascan spectropolarimeter (Applied Photophysics, UK). CD for 1 wt% solutions was performed using (2-Thi)(2-Thi)VLKAA dissolved in water or methanol and loaded into 0.01 mm thick quartz cover slip cuvettes. Spectra were measured with a 0.5 nm step, 1 nm bandwidth and 1 second collection time per step at 20 °C. Measurements were repeated four times. CD spectra on dilute conditions (0.03%) were performed using a 0.1 cm quartz cell and reported as an average of 8 consecutive scans (2 nm bandwidth, 2 s time constant, 0.2 nm step).

Fluorescence Spectroscopy. Spectra were recorded on a Cary Eclipse Varian Fluorescence Spectrometer with samples in a 1.0 cm quartz cuvette. Spectra were

measured for (2-Thi)(2-Thi)VLKAA in water or methanol (0.03 wt.%). The solvent spectra were recorded for reference. The spectra were recorded from 279 to 490 nm using an excitation wavelength $\lambda_{\text{ex}} = 265$ nm. Excitation spectra were recorded at an emission wavelength. Excitation spectra were measured at $\lambda_{\text{em}} = 300$ nm for the same samples. Background spectra from the corresponding solvents were subtracted.

X-Ray Diffraction (XRD). Diffraction patterns were obtained for stalks prepared by drying filaments of the peptide. Solutions of peptide (0.5 wt% in water of 2 wt% in methanol) were suspended between the ends of a wax-coated capillary and dried. The stalk was mounted (vertically) onto the four axis goniometer of a RAXIS IV++ x-ray diffractometer (Rigaku) equipped with a rotating anode generator. The XRD data was collected using a Saturn 992 CCD camera.

Cryogenic-Transmission Electron Microscopy (Cryo-TEM). Experiments were performed at Unilever Research, Colworth, Bedford, UK. A solution of (2-Thi)(2-Thi)VLKAA in water (1.0 wt%) was blotted and vitrified using a Gatan Cp3 cryoplunge system. Samples were prepared at a controlled temperature of 22 °C and at a relative humidity around 90%. A 3 μl drop of the solution was placed on a 400-mesh copper TEM grid (Agar) covered with a perforated carbon film (plasma treated). The drop was automatically blotted and the sample was plunged into liquid ethane (-183 °C) to form a vitrified specimen,^{16, 17} then transferred to liquid nitrogen (-196 °C) for storage. Specimens were examined in a JEOL JEM-2100 electron microscope at 200 kV, at temperatures below -175 °C. Images were recorded digitally on a Gatan UltraScan 1000 cooled CCD camera using DigitalMicrograph (Gatan) in the low-dose imaging mode to minimize beam exposure and electron-beam radiation damage.

Negative Stain TEM. High resolution TEM (HR-TEM) was done using a JEOL JEM-2010 microscope operated at 200 kV. Droplets of the peptide solution (1 wt % (2-Thi)(2-Thi)VLKAA in methanol) were placed on Cu grids coated with a carbon film (Agar Scientific, UK), stained with uranyl acetate (1 wt %) (Agar Scientific, UK) and dried.

Atomic Force Microscopy (AFM). Measurements were performed in air using a Veeco Multiprobe IIIa instrument, at room temperature (20°C) in tapping mode by using non-conductive Si tips with a force constant of about 40 N/m and a typical curvature radius on the tip of 7 nm.

Quantum Mechanical Calculations. QM calculations on the *N*-acetyl-*N*⁷-methylamide derivative of 2-Thi, abbreviated Ac-(2-Thi)-NMHe, were performed using the B3LYP^{18, 19} functional combined with the 6-31+G(d,p) basis set,²⁰ *i.e.* B3LYP/6-31+G(d,p). Classical electrostatic charges were derived for the 3-Thi residue by fitting the QM and the Coulombic molecular electrostatic potentials (MEPs), which were calculated on a large sets of point outside the nuclear region. The QM-MEP was calculated at the HF level combined with the 6-31G(d) basis set,²¹ *i.e.* HF/6-31G(d). All the calculations were carried out using the Gaussian 03 computer program.²²

Molecular Dynamics Simulations. The simulated peptide was placed in the center of a cubic simulation box ($a= 37.97 \text{ \AA}$) filled with 1781 explicit water molecules, which were represented using the TIP3 model.²³ One negatively charged chloride atom was added to the simulation box to reach electric neutrality (one positive net charge was considered for the Lys residue at neutral pH). The energy was calculated using the AMBER force-field.^{24, 25} All parameters were taken from the AMBER libraries with exception of the partial atomic charges for the 2-Thi residue, which have been explicitly developed in this work (see below and Figure 5.3.S2). Atom pair distance cut-offs were applied at 14.0 \AA to compute the van der Waals interactions. Electrostatic interactions were computed using the non-truncated electrostatic potential with Ewald Summations.²⁶ The real space term was determined by the van der Waals cut off (14.0 \AA), while the reciprocal term was estimated by interpolation of the effective charge into a charge mesh with a grid thickness 5 points per volume unit, *i.e.* using the particle-mesh Ewald (PME) method.²⁶ Both temperature and pressure were controlled by the weak coupling method, the Berendsen thermostat,²⁷ using a time constant for heat bath coupling and pressure relaxation time of 1.0 ps. Bond lengths were constrained using the SHAKE algorithm²⁸ with a numerical integration step of 2 fs. All classical calculations were performed using the Amber 10 computer program.²⁹

Before the production series, the thermodynamic variables of the system were equilibrated. The energy of the system was initially minimized to relax conformational and structural tensions using the conjugate gradient method for 5×10^3 steps. Next, different consecutive rounds of short MD runs were performed in order to equilibrate the density, temperature and pressure. First, solvent and the charged chloride atom were thermally relaxed by three consecutives runs, while the peptide

was kept frozen: 0.5 ns of NVT-MD at 500 K were used to homogeneously distribute the solvent and ions in the box. Second, 0.5 ns of isothermal and 0.5 ns isobaric relaxation were run. Finally, all the atoms of the system were submitted to 0.15 ns of steady heating until the target temperature was reached (298 K), 0.25 ns of NVT-MD at 298 K (thermal equilibration) followed by 0.5 ns of density relaxation (NPT-MD). The end of the density relaxation simulation was the starting point of the molecular simulations presented in this work. All the simulations were performed at 298 K and constant pressure of 1 atm. The coordinates of all the production runs, which were 15 ns long, were saved every 500 steps (1 ps intervals) for subsequent analysis.

5.3.2. Results and discussion

FTIR spectra (Figure 5.3.1) reveal features associated with β -sheets, in particular for the case of water as solvent, a prominent peak is observed the amide I region at 1619 cm^{-1} in solution (shifting slightly to 1625 cm^{-1} for a dried film).³⁰ The peak at 1672 cm^{-1} is due to residual TFA from the peptide synthesis,^{31, 32} however there is a small shoulder peak at 1684 cm^{-1} for the solution and more pronounced at 1693 cm^{-1} for the dried film. This latter feature is often associated with antiparallel ordering of β -sheets.^{30, 33, 34} Peaks associated with β -sheets at 1624 cm^{-1} and 1695 cm^{-1} are observed for the methanol solution, with a shoulder at 1677 cm^{-1} . In contrast to water, features in the amide II region at 1419 cm^{-1} and 1450 cm^{-1} are stronger than those in the amide I region.³⁰ These peaks are assigned to a CH_2 deformation mode and N-H in plane bend/C-N stretch modes respectively.^{30, 35, 36}

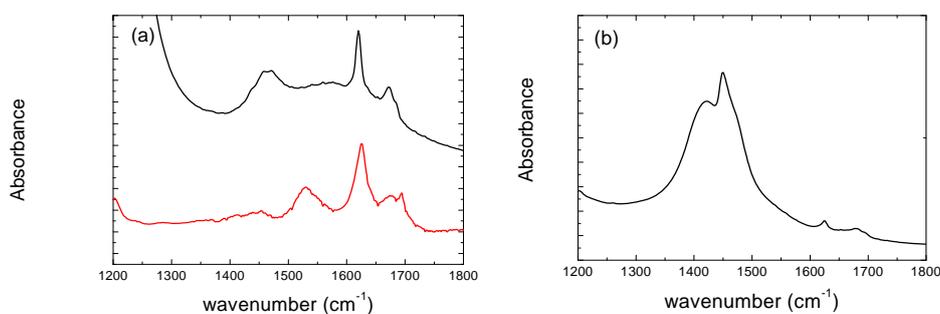


Figure 5.3.1. FTIR data for peptide in (a) 1 wt% D₂O solution (top: black line) and a dried film (bottom: red line) prepared from an 0.03 wt% solution, (b) 1 wt% methanol solution.

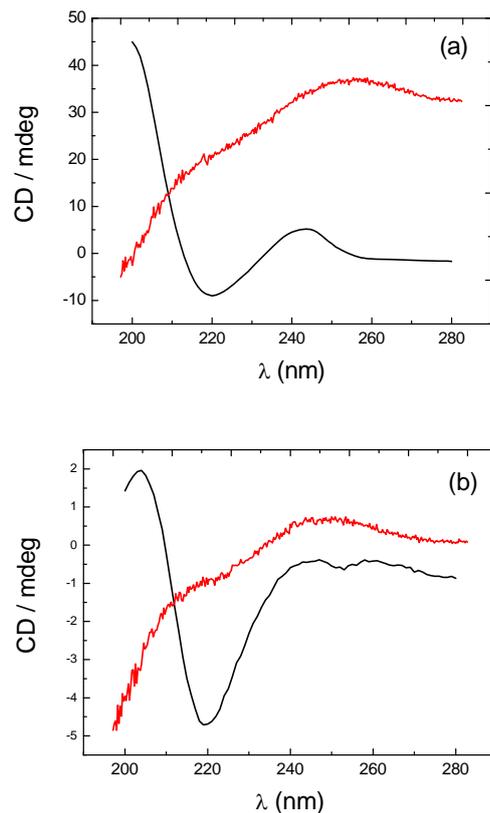


Figure 5.3.2. Circular dichroism spectrum, obtained for (2-Thi)(2-Thi)VLKAA in (a) water, (b) in methanol. The black lines are for 1 wt% solutions, the red lines are for 0.03 wt% solutions.

The circular dichroism spectra (Figure 5.3.2.) show pronounced differences depending on concentration of (2-Thi)(2-Thi)VLKAA. In dilute solution (0.03 wt% peptide in either water or methanol), there is no evidence for the usual CD spectrum from β -sheets, however at higher concentration the spectra show a minimum just below 220 nm, close to the position expected for a β -sheet structure.³⁰ The maximum at 244 nm is almost certainly from the β -thienylalanine chromophore. It is known that natural aromatic amino acids give maxima in the range 250-270 nm, close to the location of the observed maximum.^{37, 38} We are unaware of previous reports of CD spectra for peptides containing thienylalanine. Since FTIR shows features from β -sheets for 0.03 wt% solutions, we conclude that CD at the same concentration is simply not sensitive to these structures. The extent of β -sheet secondary structure

may increase with concentration, leading to the clear evidence of β -sheets in the CD for 1 wt% solutions, which is consistent with the results from FTIR.

Both CD and FTIR confirm that (2-Thi)(2-Thi)VLKAA forms β -sheets. Further support is provided by Congo red staining, a characteristic diagnostic of amyloid formation. Figure 5.3.S1 shows polarized optical microscopy images of a sample before and after staining with Congo red. The uptake of the dye leading to characteristic green birefringence is clear. The presence of birefringence in the sample without dye indicates that this peptide forms a liquid crystal phase in 1 wt% aqueous solution

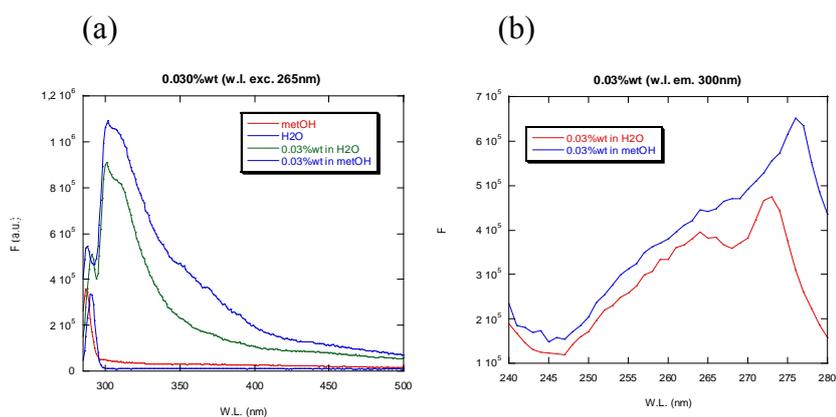


Figure 5.3.3. Fluorescence spectra measured for 0.03 wt% solutions in water and methanol, (a) emission, (b) excitation.

Background subtracted fluorescence spectra are presented in figure 5.3.3.. Raman scattering from water³⁹ gave rise to a peak in the emission spectrum that was removed when the background was subtracted. We use 265 nm as an excitation wavelength because this is a standard wavelength to reveal aromatic residues in amyloid peptides.^{40, 41} The emission spectrum shows a peak with a maximum at 301 nm. In the excitation spectrum, the two peaks at 273 nm (water) and 275 nm (methanol) can be safely assigned to solvent Raman scattering. The broad shape of the absorption at 260 nm and the shape of the emission spectrum (peaked at 301 nm) leads to an assignment of these features to the presence of residual Fmoc groups.

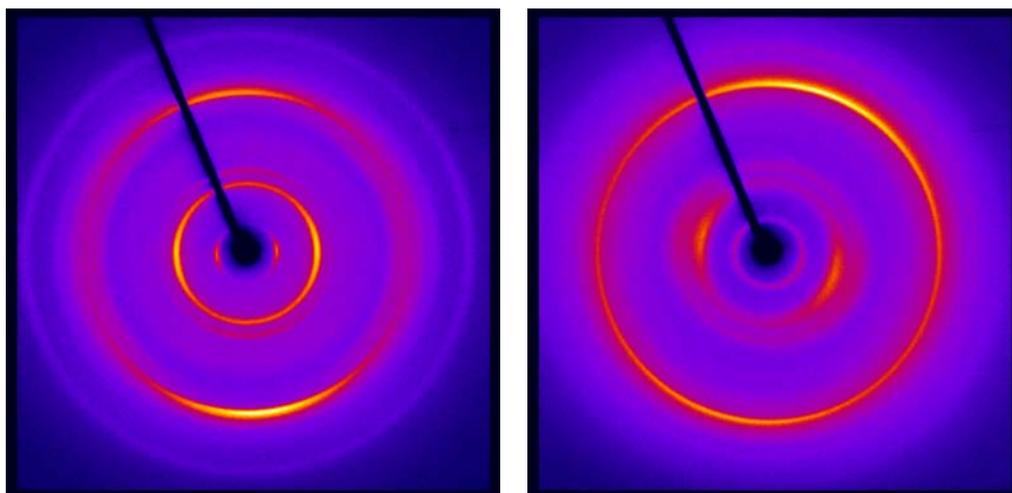


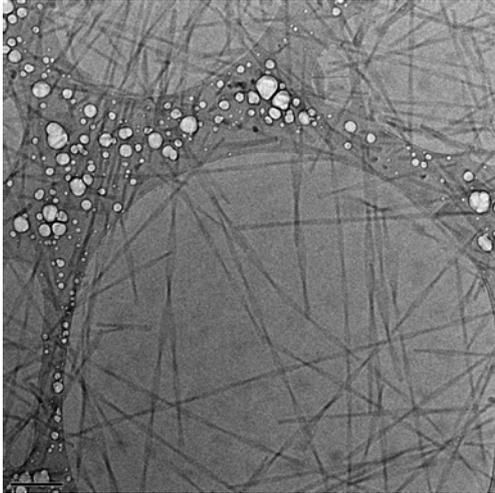
Figure 5.3.4. XRD patterns from stalks dried from (a) 0.5 wt% aqueous solution, (b) 2 wt% methanol solution.

The x-ray diffraction data shows a cross-beta pattern (figure 5.3.4.) with meridional 4.8 \AA reflections and main equatorial reflections for a sample dried from an 0.5wt% solution at 24.8 \AA and 10.7 \AA . The positions of these reflections (arising from the stacking of β -sheets) were 23.2 \AA and 9.2 \AA for the sample dried from a 2 wt% solution. The reduction in these spacings may reflect more dense packing of β -sheets as the concentration of peptide is increased. However there is better alignment of the fibres within the stalk for the latter sample. The XRD data thus supports the presence of β -sheet structures in stalks dried from solutions at different concentration.

Cryo-TEM is a powerful method to determine the structure of self-assembled structures such as peptide fibrils. The aqueous solvent is vitrified, thus trapping the in situ structure and avoiding possible drying effects that result from sample preparation for conventional TEM experiments. Figure 5.3.5.a presents a representative cryo-TEM image from a 1 wt% solution of (2-Thi)(2-Thi)VLKAA in water. There is clear evidence for twisted tapes, as observed for the related peptide AAKLVFF in water.^{12, 14} The maximum width is typically around 30 – 40 nm, it is difficult to estimate the thickness of the tapes since the “edge-on” images of tapes approach the resolution of the images. Since cryo-TEM proved problematic using methanol as solvent (due to

its high volatility), conventional TEM was performed. The image in figure 5.3.5.b shows a twisted tape structure for (2-Thi)(2-Thi)VLKAA in methanol, in contrast to AAKLVFF which forms nanotubes in methanol.¹³⁻¹⁵

(a)



(b)

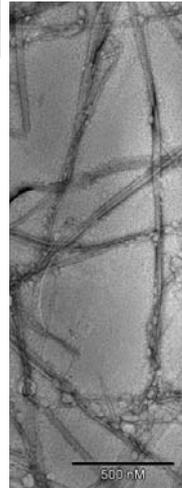
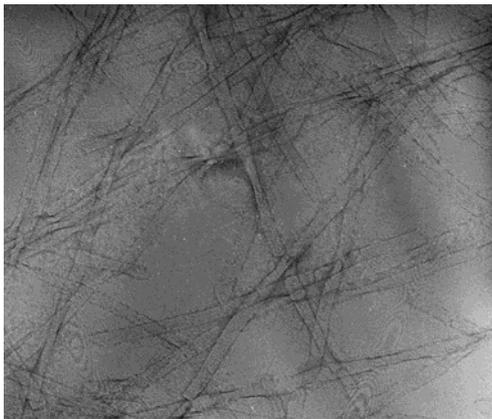


Figure 5.3.5. (a) Cryo-TEM image from a 1 wt% solution of (2-Thi)(2-Thi)VLKAA in water, (b) TEM image from the peptide in methanol.

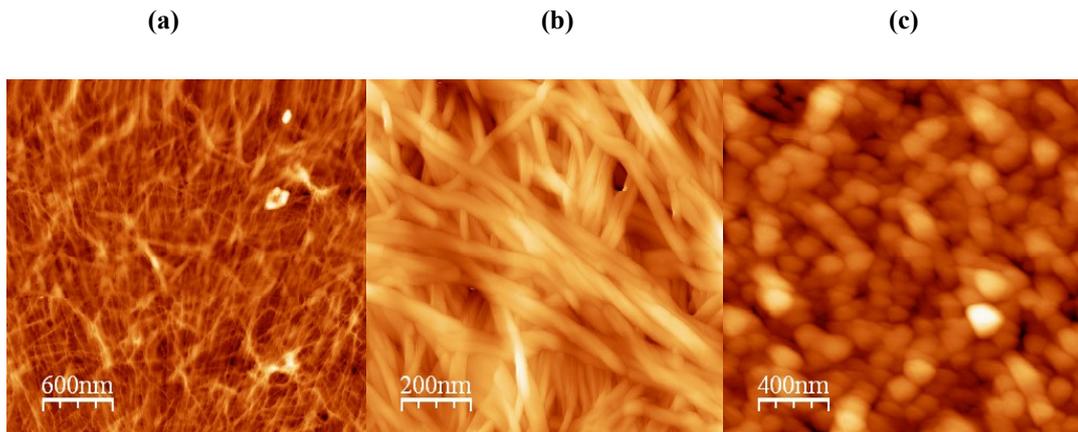


Figure 5.3.6. Representative AFM images for (2-Thi)(2-Thi)VLKAA dried onto silicon (a) from 0.03 wt% solution, (b) from 1 wt% solution, (c) blobs structure from an 0.03 wt% solution.

AFM experiments were carried out on dried films supported on mica, after deposition and overnight drying of diluted (0.03%) and concentrated (1%) peptide water solutions. For diluted solutions, regions with peptide filaments several microns long and 50 ± 10 nm thick were observed (figure 5.3.6.a). However, regions showing almost globular structures with $170(\pm 50) \times 140(\pm 40)$ nm² surface dimensions were also detected (figure 5.3.6.b). In the case of films obtained from 1% aqueous solutions only regions densely covered by peptide interlaced fibrils of the same dimensions (several microns long and 56 ± 8 nm thick) can be observed (figure 5.3.6.c).

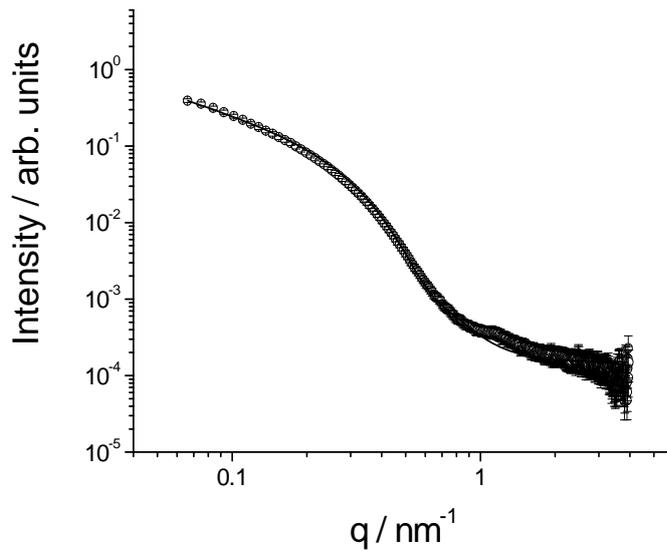


Figure 5.3.7. SAXS data (open symbols) for an 0.5 wt% solution of (2-Thi)(2-Thi)VLKAA in water, with model fit (solid line) described in text.

A SAXS intensity profile obtained for an 0.5 wt% solution in water is shown in figure 5.3.7., along with a fit that enabled fibril dimensions to be obtained. The data were fitted to a form factor of a polydisperse uniform long cylinder, using the software SASfit.⁴² The expression for the form factor is provided elsewhere.^{42, 43} The fit parameters were as follows: fibril radius $R = 4.95$ nm, Gaussian polydispersity $\sigma = 31\%$, cylinder length $L = 1652$ nm, electron density contrast with respect to solvent $\eta = 2.3 \times 10^{-5}$ (intensity units), flat background = 0.00015 (intensity units). The fit is much more sensitive to cylinder radius than cylinder length (since $L \gg R$). This model is seen to describe the SAXS data very well, and more complex models that allow, for example, for the observed twisting of the fibrils⁴⁴ are not justified because multiple form factor minima are not observed. These are probably washed out by polydispersity. The fibril radius corresponds to a diameter around 10 nm, which can be considered an effective value considering the twisted nature of the fibrils. The apparent large polydispersity in radius is consistent with the variation in fibril width and thickness as observed by cryo-TEM.



Fig 5.3.8. Self-supporting hydrogel of peptide (2-Thi)(2-Thi)VLKAA (2% wt in D₂O).

At higher concentration, (2-Thi)(2-Thi)VLKAA was observed to form a hydrogel (figure 5.3.8.). Peptide hydrogels are of interest as responsive delivery and sensor systems,⁴⁵⁻⁴⁸ and incorporation of the 2-Thi moiety may ultimately lead to responsive conductivity properties which will be investigated in future work.

Table 5.3.1. Backbone Conformation, Backbone and Side Chain Torsion Angles (figure 5.3.S2),^a and Relative Energy (ΔE)^b for the Minimum Energy Conformations of Ac-(2-Thi)-NHMe and Ac-(3-Thi)-NHMe^c at the B3LYP/6-31+G(d,p) Level

# ^c	ω_0	ω	ϕ	ψ	χ_1	$\Delta E(2\text{-Thi})$	$\Delta E(3\text{-Thi})$
$C_7^{\text{eq}} - \text{I}$	-172.8	-84.4	74.4	-175.9	-55.2	0.0	0.2
$C_7^{\text{eq}} - \text{II}$	-174.7	-82.9	59.8	179.0	44.8	0.1	0.0
C_5	177.8	-156.6	157.2	174.4	-170.2	0.2	0.4
$C_7^{\text{eq}} - \text{III}$	-179.2	-82.9	80.9	-174.7	-165.4	1.1	0.4
C_7^{ax}	172.9	74.3	-53.5	-178.4	-59.1	1.8	1.7

^a In degrees. Dihedral angles are identical for the two dipeptides (see text). ^b In kcal/mol. ^c The three minima with a C_7^{eq} backbone conformation has been labeled as $C_7^{\text{eq}} - \text{I}$, $C_7^{\text{eq}} - \text{II}$, and $C_7^{\text{eq}} - \text{III}$ following the order of stability.

QM calculations on Ac-(2-Thi)-NMHe indicated that intrinsic conformational preferences of 2-Thi are very similar to those previously reported for 3-Thi.⁶ More specifically, the linkage between methylene and the thienyl ring does not produce any change in the backbone and side chain dihedral angles (Table 5.3.1.), which are practically identical in the 2-Thi and 3-Thi containing dipeptides. The only difference between the two compounds corresponds to the relative energies (ΔE in Table 5.3.1.) that undergo some small variations. However, the C_7^{eq} and C_7^{ax} are clearly the most and least favorable backbone conformations, respectively, in both cases. Atomic centered electrostatic charges were calculated for all the minimum energy conformations shown in Table 5.3.1. and, subsequently, weighted according to the Boltzmann population of each minimum to derive the electrostatic force-field parameters for the 2-Thi residue.

MD simulations on (2-Thi)(2-Thi)VLKAA were performed at 298 K considering two different protocols. In the first one (MD1) the temporal evolution of the peptide was examined without imposing any restraint, while in the second one (MD2) the restraints derived from NMR experiments were included in the potential through a harmonic expression with a soft constant ($k_r = 10 \text{ kcal/mol}\cdot\text{\AA}^2$). The restraints imposed in MD2 correspond to 5 main chain – main chain, 14 main chain – side chain and 19 side chain – side chain distances. In both cases the fully extended conformation was used as starting point. Despite the fact that QM calculations predict that C_7^{eq} (also denoted inverse γ -turn) is the most stable conformation of the 2-Thi residue, comparison between theoretical results and available experimental data reported for other amino acids show that the stability of this conformation is overestimated when computational methods are applied to small model systems (dipeptides) in the gas-phase.⁴⁹⁻⁵¹ This led us to consider the β (or extended conformation), which is destabilized by only 0.2 kcal/mol with respect to the γ_L (Table 5.3.1.) as a suitable alternative for the starting conformation.

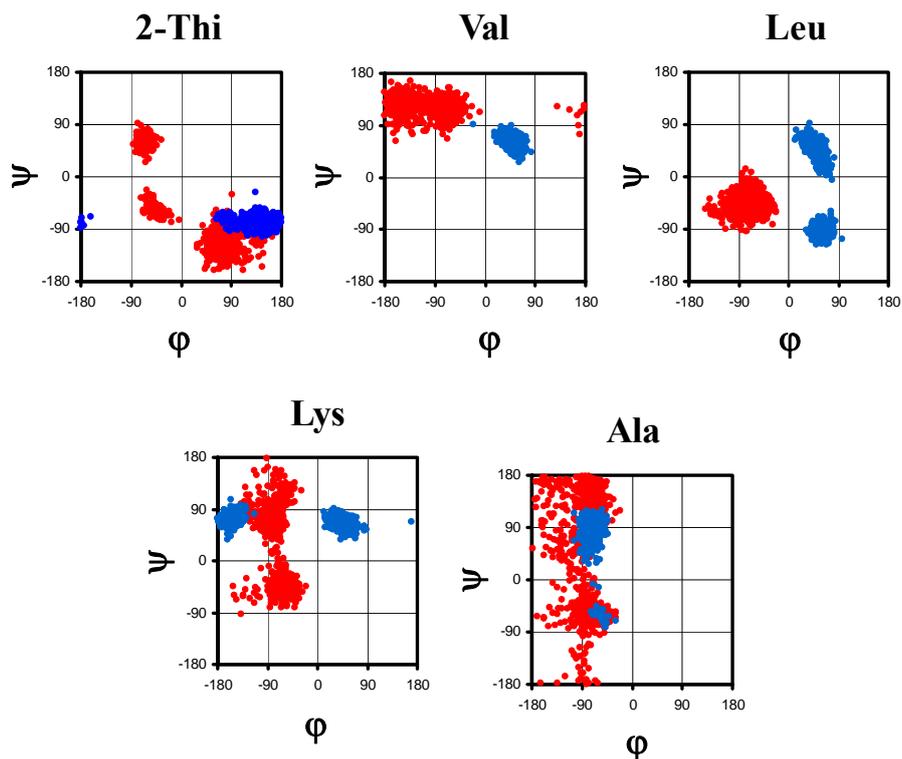


Fig 5.3.9. Backbone dihedral angle distributions for the peptide residues in MD1 (red) and MD2 (blue) simulations. Trajectories were 15 ns long and snapshots were saved at 1 ps intervals.

Analysis of the distribution for the five central residues (figure 5.3.9.) reveals some differences in the secondary structures reached by MD1 and MD2. For MD1 the dihedral angle ψ of Ala and Lys exhibits a large degree of flexibility, while the $\{\phi, \psi\}$ distribution of the 2-Thi residue falls into four different clusters. Specifically, the four regions visited by the 2-Thi residue in the simulation without restraints correspond to the C_7^{eq} ($\phi, \psi \approx -70^\circ, 65^\circ$), C_7^{ax} ($\phi, \psi \approx 70^\circ, -65^\circ$), α_L ($\phi, \psi \approx -50^\circ, -50^\circ$) and P_{II} ($\phi, \psi \approx 70^\circ, -120^\circ$) with populations of 30%, 27%, 21% and 14%. In contrast, the two backbone dihedral angles occupy well-defined clusters in MD2 indicating that once the secondary structure compatible with the imposed NMR restraints has been reached it is preserved during the rest of the trajectory. The flexibility of the peptide in MD1 as well as its well-defined secondary structure in MD2 are also evidenced in figure 5.3.10., which displays the temporal evolution of both the end-to-end distance

(d_{e-e}) and the radius of gyration (R_g). Thus, d_{e-e} ranges from 4.939 to 18.838 Å and from 8.435 to 13.570 Å in MD1 and MD2, respectively, the average value being 14.006 ± 2.840 and 12.187 ± 0.649 Å. The average value of R_g , which is 6.110 ± 0.319 and 5.144 ± 0.042 Å for MD1 and MD2, respectively, indicates that the conformation found for (2-Thi)(2-Thi)VLKAA in dilute aqueous solution does not correspond to that of a fully extended nor semi-extended strand.

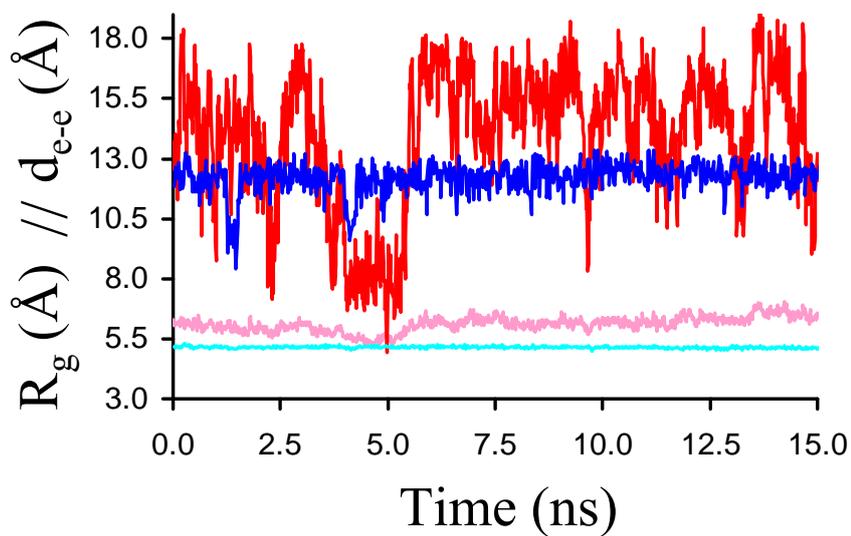


Fig 5.3.10. Temporal evolution of the end-to-end distance (d_{e-e}) and the radius of gyration (R_g) for MD1 (red and pink, respectively) and MD2 (dark blue and light blue, respectively) simulation

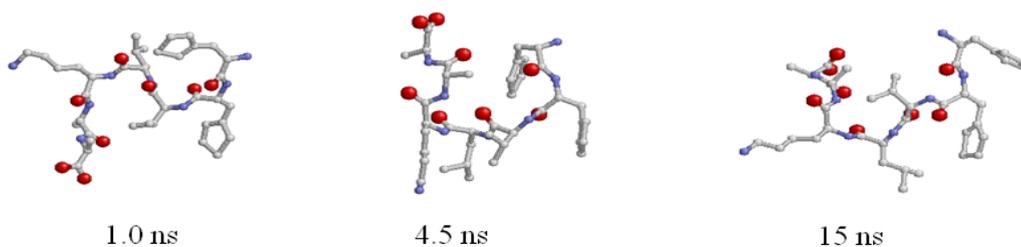


Fig 5.3.11. Snapshots of (2-Thi)(2-Thi) VLKAA taken from the MD1 simulation at 1.0, 4.5 and 15.0 ns.

Figure 5.3.11. shows the structure of the peptide at selected snapshots (1.0, 4.5 and 15.0 ns) recorded during the MD1 trajectory. It can be seen that the most essential conformational trends of the peptide are preserved during the whole simulation: the backbone adopts a folded conformation in which the polar amide groups and the positively charged side group of Lys tend to be exposed to the solvent molecules. Although the side groups of the 2-Thi residues exhibit an enormous flexibility, the distance between the centers of mass of the two aromatic rings was larger than 6.2 Å during the whole trajectory. As attractive interactions between two thiophene rings have been reported to exist at distances of around 3.8 and 4.8 Å for the planar π -stacked and perpendicular T-shaped arrangements,^{52, 53} respectively, we conclude that the side chains of the 2-Thi residues are not able to interact promoting intramolecular π -electron transfer processes. On the other hand, inspection of the backbone conformation in figure 5.3.11. allows rationalize the large fluctuations found for MD1 in both the d_{e-e} and R_g (figure 5.3.10.). Although in general terms the molecular shape is retained during the whole simulation, breathing movements result in the continuous formation and disruption of intramolecular hydrogen bonds altering the molecular dimensions. For example, while the structure at 1.0 ns does not present any intramolecular hydrogen bond, those at 4.5 and 15.0 ns exhibit two and one 10-membered hydrogen bonded rings, respectively, which correspond to the amide...amide intramolecular interactions typically associated with the β -turn motif. Detailed analysis of the stored coordinates indicates that the changes in the hydrogen bonding pattern occur during the whole MD1 trajectory.

In contrast, the structure reached for (2-Thi)(2-Thi)VLKAA by imposing NMR restrictions in MD2 was preserved during the whole trajectory. This is evidenced in figure 5.3.12.a, which shows the superimposition of 15 peptide conformers taken from snapshots separated by intervals of 1ns. figure 5.3.12.b provides details of the averaged conformation, which allows describe at the atomistic level the structure compatible with the NMR data. As it can be seen, two N-H...O intramolecular interactions defining a 7- and 13-membered hydrogen bonded rings, which correspond to the γ -turn and α -helix motifs, respectively, have been identified in this conformation. Furthermore, as was suggested by the low value of R_g (figure 5.3.10.), it correspond to a highly folded conformation, even though the polar peptide groups are exposed to the solvent with the only exception of those involved in intramolecular

hydrogen bonds. Another important finding is that the aromatic groups of the two 2-Thi residues are separated by only ~ 4.6 Å, their relative orientation being like a distorted T-shape. Accordingly, this structure seems to be stabilized by an attractive non-covalent interaction between the two aromatic rings, suggesting that it may be used as a building block to create a π -stacking ladder through a self-assembly processes. It is worth noting that these results confirm our initial assumption (see Introduction), according to which the reversal of the sequence promotes the formation of π - π interactions. Thus, the carboxylate group at the C terminus is completely exposed to the solvent without inducing any repulsive interaction with the π -electron density of the thienyl groups. In addition, the positively charged N terminus enhances the stability of the π - π interaction since the distance between the NH_3^+ moiety is located at about ~ 4.8 Å from the center of mass of the second 2-Thi residue.

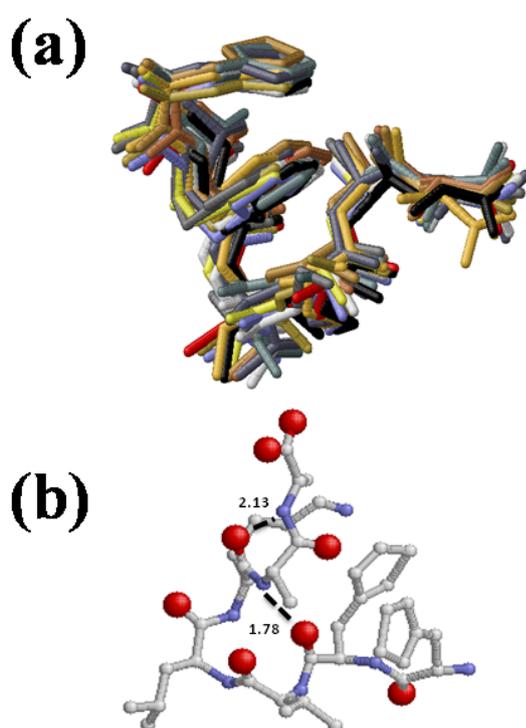


Fig 5.3.12. (a) Superimposition of snapshots saved from the MD2 simulation at 1 ns intervals. (b) Averaged coordinates of **(2-Thi)(2-Thi)VLKAA** derived from the MD2 simulation.

5.3.3. Conclusions

The self-assembly of a peptide incorporating the non-natural amino acid β -2 thienylalanine has been investigated, in water and methanol. The peptide has been designed on the basis of a previously studied motif AAKLVFF which forms twisted fibrils in water and nanotubes in methanol. For (2-Thi)(2-Thi)VLKAA, a twisted fibril morphology is observed in both water and methanol and CD and FTIR confirm a β -sheet structure. Hydrogelation is observed at higher concentration. Quantum mechanics calculations are used to examine the conformation and charge distribution within the β -2-thienylalanine amino acid. This is complemented with molecular dynamics simulations on isolated (2-Thi)(2-Thi)VLKAA peptides which reveals turn-like structures. The aggregation of these peptides into β -sheet fibrils is currently being investigated, as are the electronic properties which are expected to be of interest due to charge delocalization and π -stacking. The ability to form hydrogels in a concentration dependent manner is noteworthy for potential applications as responsive nanomaterials incorporating electronic functionality.

5.3.4. Bibliography

- [1] van Hest, J. C. M.; Tirrell, D. A. *Chem. Comm.* 2001, (19), 1897-1904.
- [2] Wang, L.; Schultz, P. *Angew. Chem.* 2005, 44, 34-66.
- [3] Connor, R. E.; Tirrell, D. A. *J. Macromol. Sci., Part C, Polym. Rev.* 2007, 47, 928.
- [4] Munier, R.; Cohen, G. N. *Biochim. Biophys. Acta* 1959, 31, (2), 378-391.
- [5] Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *J. Am. Chem. Soc.* 1995, 117, 536-537.
- [6] Rodríguez-Ropero, F.; Zanuy, D.; Assfeld, X.; Alemán, C. *Biomacromolecules* 2009, 10, 2338-2343.
- [8] Haspel, N.; Zanuy, D.; Aleman, C.; Wolfson, H.; Nussinov, R. *Structure* 2006, 14, (7), 1137-1148.
- [9] Haspel, N.; Zanuy, D.; Zheng, J.; Aleman, C.; Wolfson, H.; Nussinov, R. *Biophys. J.* 2007, 93, (1), 245-253.
- [10] Klok, H.-A.; Rösler, A.; Götz, G.; Mena-Osteritz, E.; Bäuerle, P. *Org. Biomol. Chem.* 2004, 2, (2541-2544).
- [11] Digelmann, S. R.; Gorham, J. M.; Tovar, J. D. *J. Am. Chem. Soc.* 2008, 130, 13840-13841.
- [12] Schillinger, E.-K.; Mena-Osteritz, E.; Hentschel, J.; Börner, H. G.; Bäuerle, P. *Adv. Mater.* 2009, 21, 1562-1567.
- [13] Castelletto, V.; Hamley, I. W.; Harris, P. J. F. *Biophys. Chem.* 2008, 139, 29-35.
- Krysmann, M. J.; Castelletto, V.; McKendrick, J. M. E.; Hamley, I. W.; Stain, C.; Harris, P. J. F.; King, S. M. *Langmuir* 2008, 24, 8158-8162.
- [14] Castelletto, V.; Hamley, I. W.; Harris, P. J. F.; Olsson, U.; Spencer, N. *J. Phys. Chem. B* 2009, 113, 9978-9987.
- [15] Hamley, I. W.; Nutt, D. R.; Brown, G. D.; Miravet, J. F.; Escuder, B.; Rodríguez-Llansola, F. *J. Phys. Chem. B* 2010, 114, 940-951.
- [16] Talmon, Y., Cryogenic transmission electron microscopy in the study of surfactant systems. In *Modern Characterization Methods of Surfactant Systems*, Binks, B. P., Ed. Marcel Dekker: New York, 1999; pp 147-178.
- [17] Cui, H.; Hodgdon, T. K.; Kaler, E. W.; Abezgaous, L.; Danino, D.; Lubovsky, M.; Talmon, Y.; Pochan, D. J. *Soft Matter* 2007, 3, 945-955.
- [18] Becke, A. D. *J. Chem. Phys.* 1993, 98, 1372.
- [19] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* 1993, 37, 785.

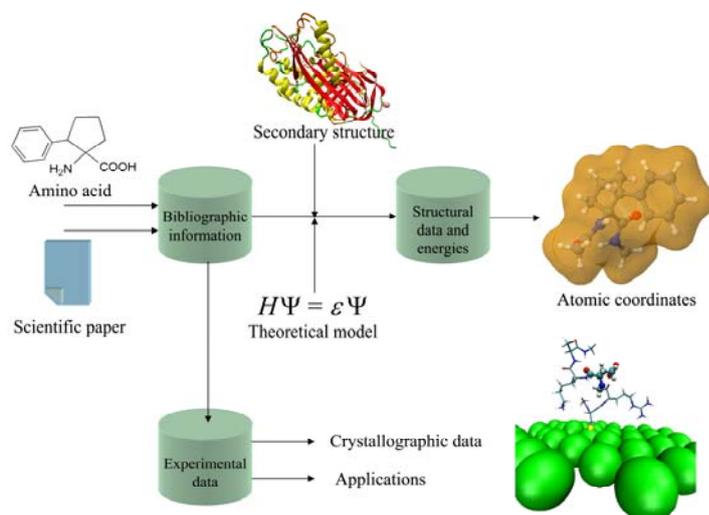
- [20] McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* 1980, 72, (10), 5639-5648.
- [21] Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* 1973, 28, (3), 213-222.
- [22] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T., Jr.; 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 03*, Revision B.02; Gaussian, Inc.: Pittsburgh PA, 2003.
- [23] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* 1983, 79, (2), 926-935.
- [24] Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* 2000, 21, (12), 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. *J. Am. Chem. Soc.* 1995, 117, (19), 5179-5197.
- [26] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* 1993, 98, (12), 10089-10092.
- [27] Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* 1984, 81, (8), 3684-3690.
- [28] Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* 1977, 23, (3), 327-341.
- [29] Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.;

- Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. AMBER 10, 2008, University of California, San Francisco.
- [30] Stuart, B., *Biological Applications of Infrared Spectroscopy*. Wiley: Chichester, 1997.
- [31] Gaussier, H.; Morency, H.; Lavoie, M. C.; Subirade, M. *Appl. Environ. Microbiol.* 2002, 68, (10), 4803-4808.
- [32] Pelton, J. T.; McLean, L. R. *Anal. Biochem.* 2000, 277, 167-176.
- [33] Haris, P.; Chapman, D. *Biopolymers* 1995, 37, 251-263.
- [34] Miyazawa, T.; Blout, E. R. *J. Am. Chem. Soc.* 1961, 83, 713-719.
- [35] Krimm, S.; Bandekar, J. *Adv. Protein Chem.* 1986, 38, 181-364.
- [36] Lin, S.-Y.; Chu, H.-L. *Int. J. Biol. Macromolecules* 2003, 32, 173-177.
- [37] Woody, R. W., Circular dichroism of peptides and proteins. In *Circular dichroism. Principles and applications*, Nakanishi, K.; Berova, N.; Woody, R. W., Eds. VCH: New York, 1994.
- [38] Woody, R. W. *Methods Enzymol.* 1995, 246, 34-71.
- [39] Lacowicz, J. R., *Principles of Fluorescence Spectroscopy*. 2nd ed.; Kluwer: New York, 1999.
- [40] Teale, F. W. J.; Weber, G. *Biochem. J.* 1957, 65, (3), 476-482.
- [41] Giri, K.; Bhattacharyya, N. P.; Basak, S. *Biophys. J.* 2007, 92, 293-302.
- [42] Kohlbrecher, J., <http://kur.web.psi.ch/sans1/SANSSoft/sasfit.html>. In 2009.
- [43] Pedersen, J. S. *Adv. Colloid Interface Sci.* 1997, 70, 171-210.
- [44] Hamley, I. W. *Macromolecules* 2008, 41, 8948-8950.
- [45] Zhang, Y.; Gu, H.; Yang, Z.; Xu, B. *J. Am. Chem. Soc.* 2003, 125, 13680-13681.
- [46] Liang, G.; Yang, Z.; Zhang, R.; Li, L.; Fan, Y.; Kuang, Y.; Gao, Y.; Wang, T.; Lu, W. W.; Xu, B. *Langmuir* 2009, 25, (15), 8419-8422.
- [47] Yang, Z.; Xu, B. *Chem. Comm.* 2004, 2424-2425.
- [48] Yang, Z.; Xu, B. *Adv. Mater.* 2006, 18, 3043-3046.
- [49] Aleman, C.; Jimenez, A. I.; Cativiela, C.; Perez, J. J.; Casanovas, J. *J. Phys. Chem. B* 2002, 106, (45), 11849-11858.
- [50] Aleman, C.; Zanuy, D.; Casanovas, J.; Cativiela, C.; Nussinov, R. *J. Phys. Chem. B* 2006, 110, (42), 21264-21271.
- [51] Casanovas, J.; Zanuy, D.; Nussinov, R.; Aleman, C. *J. Org. Chem.* 2007, 72, (6), 2174-2181.

- [52] Tsuzuki, S.; Honda, K.; Azumi, R. *J. Am. Chem. Soc.* 2002, 124, (41), 12200-12209.
- [53] Rodríguez-Ropero, F.; Casanovas, J.; Alemán, C. *J. Comput. Chem.* 2007, 29, 69.

6.Design of a non-coded amino acid data base

6.1. NCAD, a data base integrating the intrinsic conformational preferences of non-coded amino acids



6.1.1 Introduction

The repertoire of amino acids currently available for application in life and materials sciences is rapidly expanding. Besides the 20 genetically coded α -amino acids contained in proteins, more than 700 different amino acids have been found in Nature¹ (most of them, also α -amino acids), and many others have been synthesized by organic chemists.^{1a,2} These residues are joined under the common name of *non-proteinogenic* or *non-coded* amino acids (nc-aa). Although they are rigorously excluded from ribosomally-synthesized native peptides and proteins, several naturally occurring peptides (produced non-ribosomally) have been found to contain nc-aa.^{2i,3} Additionally, they are increasingly used to improve the pharmacological profile of natural peptides endowed with biological activity⁴ (to confer resistance against enzymatic degradation, enhance membrane permeability, or increase selectivity and affinity for a particular receptor). Nc-aa have also led to the development of efficient drugs of non-peptidic nature.⁵ Moreover, recent advances in biotechnology have paved the way for protein engineering with nc-aa.⁶ Thus, proteins

containing residues with fluorescent,⁷ redox-active,⁸ photosensitive groups,^{7c,9} or other side chains endowed with specific chemical reactivity¹⁰ or spectroscopic properties,¹¹ may serve as biosensors and spectroscopic or biophysical probes, as well as to construct systems for drug delivery and diagnosis through imaging useful in medicine.⁶⁻¹¹ Nc-aa have also found application in nanobiology to promote the self-assembly of nanostructures.¹² Other remarkable applications of nc-aa include the development of bio-inspired synthetic organic polymers that emulate the shape and properties of natural peptides and proteins.¹³

In spite of the great significance and potential utility of nc-aa in modern biology, biomaterials engineering and medicine, structural information is only available for a limited number of them. Specifically, the intrinsic conformational preferences of nc-aa, which are typically obtained by sampling the potential energy hypersurface of small peptide systems incorporating them through sophisticated quantum mechanical calculations at the *ab initio* or Density Functional Theory (DFT) levels in the absence of external forces, have been reported for only several tenths of these compounds. Knowledge of such intrinsic conformational preferences is however essential to understand the features that distinguish these nc-aa from the coded ones and, therefore, for a complete and satisfactory exploitation of their capabilities. Structural data yielded by crystallographic or spectroscopic studies on peptide sequences incorporating nc-aa are also available for an increasing number of cases. Comparison of such experimental information with that derived from high-level theoretical calculations helps establish the influence of the external packing forces, the solvent effects, and the chemical environment on the conformational propensities of nc-aa.

In practice, the use of nc-aa with well-characterized conformational properties is frequently limited because the relevant information is highly dispersed. Indeed, the quantum mechanical calculations describing the intrinsic conformational preferences of these compounds are typically reported in physical chemistry journals, while their synthesis appears in organic chemistry publications. Crystallographic and spectroscopic studies of small peptide sequences containing nc-aa are usually developed by organic and peptide chemists and published in journals specialized in these fields. In contrast, many applications of nc-aa are tested by researchers working on protein engineering, medicine, or materials science. This highly dispersed information together with the wide and diverse potential applications of nc-aa made us realize that the design of a simple informatics tool integrating the contributions

supplied by the different research fields would facilitate the use of these compounds in many practical applications. To this aim, we have built a data base of nc-aa containing the fundamental conformational descriptors and relevant bibliographic information about experimental studies and already developed practical applications. This user-friendly on-line facility is intended as an aid to the development of new applications by combining the structural descriptors with thermodynamics criteria for a proper selection of nc-aa compatible with the user-defined requirements.

In this work, we present NCAD (Non-Coded Amino acids Data base), a data base conceived and created to identify the nc-aa that are compatible with a given structural motif, which is the key requirement for the application of these compounds in life and materials sciences. The data base integrates structural and energetic descriptors of those nc-aa whose intrinsic conformational properties have been previously studied using *ab initio* or DFT quantum mechanical calculations and reported in the literature. The information about NCAD presented in the following sections is organized as explained next. First, the most relevant aspects of the data base are described, such as structure, contents, technical features, and relationships allowed among the different descriptors. Next, α -tetrasubstituted α -amino acids, being the first family of nc-aa integrated into the data base, are presented. The conformational preferences of these compounds, which have been the subject of extensive study by our research groups, are examined and discussed using the overall information captured in the data base. Finally, the utility and applicability of the data base is illustrated with a simple example. Specifically, conformationally restricted analogues of the neuropeptide Met-enkephalin are designed through targeted replacements with nc-aa, using the information incorporated in the data base.

6.1.2. Methods

NCAD: Contents, Tools and Technical Features

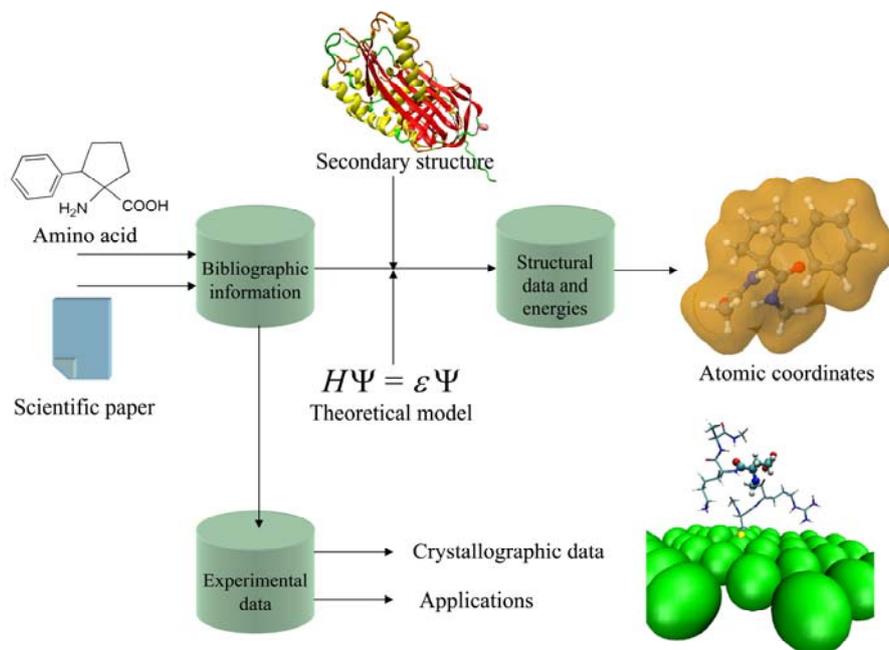


Figure 6.1.1. Schematic graphic algorithm showing the structure of the NCAD data base.

Figure 6.1.1. portrays the NCAD contents. The information contained in the data base for each nc-aa is the following:

- A complete description (dihedral angles, three-dimensional structure, relative energy, etc) of all the minimum energy conformations found for each amino acid using *ab initio* or DFT quantum mechanical calculations. It should be noted that such theoretical studies are carried out on small peptide systems of general formula RCO-Xaa-NHR' (Xaa = amino acid; R and R' = Me or H). When several theoretical studies were available in the literature for a particular amino acid, the information was extracted from that using the highest level of calculation. Conformational studies based on calculations less accurate than *ab initio* or DFT methods, *i.e.* molecular mechanics based on classical force-fields,

semiempirical procedures, etc, have not been considered for NCAD. It is worth noting that only in a few cases are the atomic coordinates of the minimum energy conformations supplied by the authors as Supporting Information of the original article. Accordingly, in order to incorporate such coordinates into the data base for all nc-aa, all minimum energy conformations for which coordinates were not available have been re-calculated using the same theoretical level as in the original study. Although this is an arduous and extremely time-consuming task, we considered it essential for the usefulness of the data base. Throughout the theoretical works describing the conformational propensities of nc-aa, different nomenclature systems are used to term the energy minima located. We have unified them so that only the nomenclature introduced by Perczel *et al.*¹⁴ is used in NCAD to identify the backbone conformation of the different minima characterized for each amino acid. According to it, nine different conformations can be distinguished in the potential energy surface $E=E(\phi,\psi)$ of α -amino acids,¹⁴ namely γ_D , δ_D , α_D , ε_D , β_{DL} , ε_L , α_L , δ_L , and γ_L . This information is particularly useful to check the compatibility of the minima energetically accessible to a particular nc-aa with the secondary structure motifs typically found in peptides and proteins (helices, sheets, turns, etc).

b) Bibliographic information related to experimental data, when available. Selected references describing the synthesis and characterization of each particular amino acid in its free form, as a salt, or with the amino group adequately protected for use in peptide synthesis are given. Also references providing experimental information about the conformational preferences of the amino acid are included in NCAD. These concern mainly structural studies of peptides incorporating nc-aa using either spectroscopic or crystallographic techniques. Furthermore, in selected cases, atomic coordinates extracted from X-ray crystal structures have also been included in the data base. Such three-dimensional structures allow a direct comparison with the minimum energy conformations predicted by theoretical calculations. Upon request and permission of the authors, we plan to include in NCAD examples of crystal structures of peptide sequences incorporating different nc-aa. This does not intend to be comprehensive, since we are aware that the crystal structures of peptides and proteins containing nc-aa are already stored in specialized data

bases, like the Cambridge Structural Data base¹⁵ (CSD) and the Protein Data Bank¹⁶ (PDB).

c) Applications reported for each particular nc-aa divided into two main categories, namely, those related to the biological properties of the amino acid (or compounds incorporating it), and applications in materials science. The most relevant publications in each field, either papers in scientific journals or patents, are given.

As commented above, NCAD has been conceived to contain all those nc-aa whose intrinsic conformational propensities have been determined through *ab initio* or DFT quantum mechanical calculations. Such amino acids may be chiral or not. In the former case, two enantiomeric forms do exist for a given amino acid. Obviously, the conformational preferences of only one enantiomer have been calculated and are included in the data base, since those of the enantiomeric species may be obtained by simply changing the signs of the dihedral angles. The bibliographic information concerning synthesis, spectroscopic characterization, experimental structural properties, and applications included in the data base covers the two enantiomers as well as the racemic form, when applicable.

Figure 6.1.2. shows the entity-relationship diagram¹⁷ (ERD) of NCAD. This diagram summarizes the conceptual representation of the information stored in the data base, as well as their inter-relational scheme. The entities are shown in squares, the relationships are lines between boxes, and their multiplicities are displayed using numbers or stars following ERD standards. All the experimental and theoretical information in the data base has been extracted from relevant literature sources and, accordingly, all data available in NCAD for a given amino acid are connected with the bibliographic references. Labels (*label_{exp}*, *label_{appli}* entities) allow the identification of the applications described for each amino acid as well as the experimental data. The data base also stores the atomic coordinates of the minimum energy conformations (*minimum*) predicted by a particular quantum mechanical method (*theory*), which have been extracted from the literature (*bibliography*), and its correspondence with secondary structure motifs (*conformation-secondarystructure*). Furthermore, internal coordinates, such as dihedral angles (*dihedrals*) for both the backbone and the side chain, can be easily obtained from the data base for each minimum energy conformation of a given amino acid.

investigation (author); (vii) the experimental studies reported in the literature relative to synthesis and characterization or to conformational properties; and (ix) the applications (in medicine or materials science) described.

The amino acids fulfilling all the selected criteria are displayed in a list (Figure 6.1.3.a) (by default, when no criterion is specified, all residues included in NCAD are listed) and a report with the search criteria and results may be generated. The user can select three different nomenclatures to indicate the amino acids names: systematic, common or abbreviated (usually resembling the three-letter code used for proteinogenic amino acids), e.g. 2-amino-2-methylpropanoic acid, α -aminoisobutyric acid or Aib, respectively. It should be noted that the common and abbreviated names are not available for all residues.

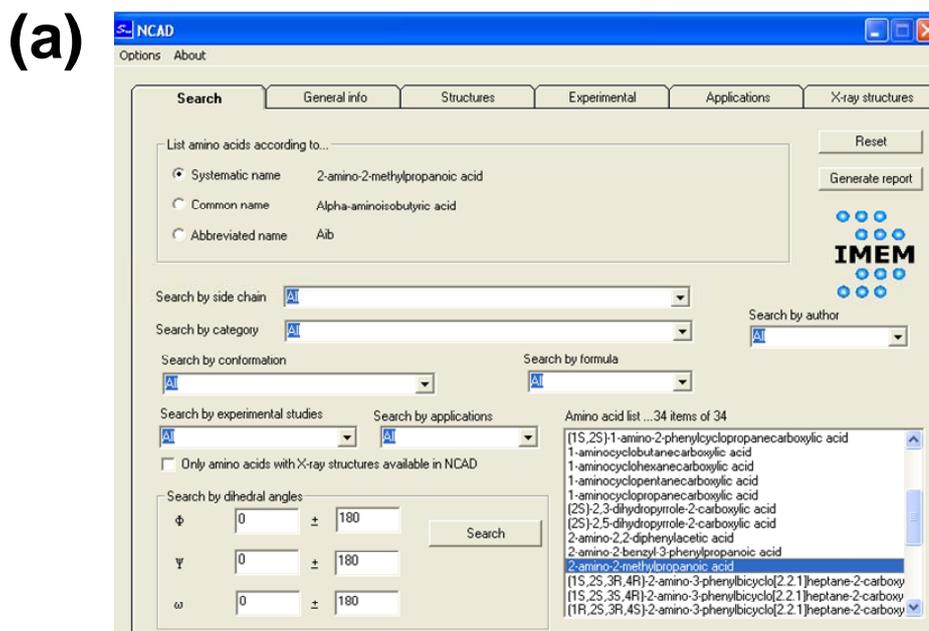
Selection of an amino acid from this list allows access to all the information stored in the data base for this particular residue, which is presented in the remaining five folders. The “general info” folder provides the chemical description of the amino acid, including a graphical scheme of its chemical structure and the types of experimental studies and applications available. These data together with the corresponding bibliographic references stored in the data base for the selected amino acid may be exported to a report. The “conformations” folder (Figure 6.1.3.b) displays information extracted from the theoretical study performed at the highest quantum mechanical level reported in the literature. Specifically, it includes the bibliographic source, the number of minima characterized, the list of dihedral angles and energy of each minimum, the position of each minima within the Ramachandran map, and the quantum mechanical method used in the theoretical study. A report containing this information may be generated. Furthermore, this folder allows a graphical inspection of each minimum energy conformation with the molecular visualization software Rasmol.¹⁸ This is made possible by accessing the atomic coordinates stored in the data base through an internal call of the interface. The atomic coordinates of each minimum energy conformation may be also extracted. The next two folders provide bibliographic data about experimental studies (synthesis and characterization, conformational propensities) and reported applications in medicine and materials science. Such bibliographic information is not intended to be comprehensive but rather present some significant publications in the field. Thus, these folders supply essential information reported for the amino acid considered, but a systematic literature search is recommended if this amino acid is finally selected by

the user for practical applications. Finally, the last folder allows access to the X-ray crystal structures of peptide sequences containing the selected amino acid that are available in the data base. Three-dimensional visualization of these structures is made through the Rasmol software.¹⁸

α -Tetrasubstituted α -Amino Acids

Because of our interest in α -tetrasubstituted α -amino acids and their relevance in the design of peptides with well-defined conformational properties,¹⁹ this is the first family of nc-aa integrated into NCAD. Table 6.1.1. lists the 29 α -tetrasubstituted α -amino acids whose intrinsic conformational propensities have been determined to date using *ab initio* or DFT quantum mechanical methods²⁰⁻³⁷ according to a literature search. This list is expected to be dynamic and thus be updated when studies on new compounds in this category are reported. Other families of nc-aa will be integrated into NCAD in the near future.

α -Tetrasubstituted amino acids are characterized by an α -carbon atom bearing four substituents (different from hydrogen) and hence are also called quaternary amino acids. They are usually divided into two subtypes, depending on whether the α carbon is involved in a cyclic structure or not. Notably, the former has been studied much more frequently, the number of entries in the data base being 24 and 5, respectively. Regarding chirality, 9 residues in Table 6.1.1. present two identical substituents at the α carbon and are therefore achiral, namely Aib, Ac₃c, Dpg, Dbg, Ac₄c, Ac₅c, Adt, Ac₆c, and Toac. The remaining 20 amino acids present at least one chiral center. In the latter case, only the low energy minima of one enantiomer (usually L) have been considered for incorporation into NCAD, those of the other enantiomer being easily deducible by simply changing the sign of both (ϕ, ψ) angles. When two or more chiral centers are present in the molecule, the stereochemistry is indicated through the *R/S* instead of the *L/D* nomenclature (the *S* configuration at the α carbon usually corresponding to L).



(b)

(1S,2F)-1-amino-2-(guanidinomethyl)-cyclopentanecarboxylic acid

This amino acid is found in the following conformations (BACKBONE/SIDE CHAIN):
 [Alpha-D,Envelope-gamma-pi-down,phi_1]

Paper reporting this information:
 SIDE CHAIN TO BACKBONE INTERACTIONS DICTATE THE CONFORMATIONAL PREFERENCES OF A CYCLOPENTANE ARGININE ANALOGUE
 URL: <http://pubs.acs.org/doi/10.1021/ol002206>
 Journal: JOURNAL OF ORGANIC CHEMISTRY
 volume: 74 first page: 2403 year: 2009
 Authors: REVILLA-LÓPEZ, G.

Relative energy: 0 kcal/mol
 Level of theory: B3LYP/6-31+G(d,p)

Dihedral angles (degrees):

Angle	Value
omega1	174.8
phi	71.9
psi	18.8
omega2	174.5
chi1	-41.1
chi2	-20.3
chi3	-40.9
chi4	-41.2

RasMol

Structural information for the amino acid: Aib

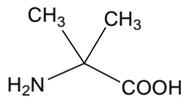
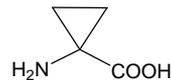
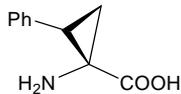
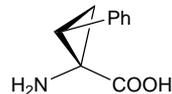
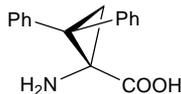
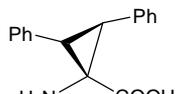
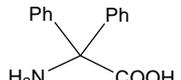
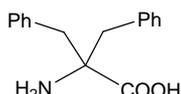
The following information comes from the paper:
 CONFORMATIONAL PREFERENCES OF ALPHA-AMINOCYCLOPENTANECARBOXYLIC ACID SUBSTITUTED AT THE ALPHA CARBON.
 J PHYS CHEM B (2012) 16(4) 1048-1054
 URL: <http://pubs.acs.org/doi/10.1021/jp121394n>

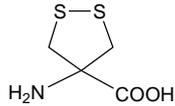
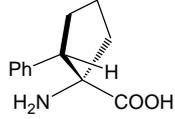
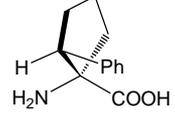
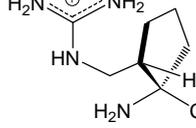
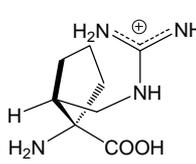
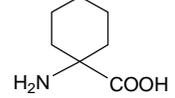
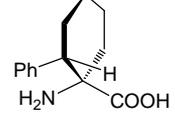
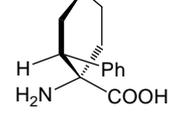
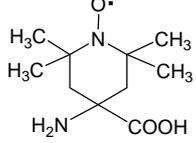
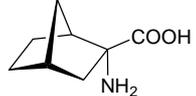
The minima of this amino acid are found in the following conformations (BACKBONE/SIDE CHAIN):
 conformation 1: Beta-D/L/N, A... with relative energy 0 kcal/mol
 conformation 2: Gamma-L/N, A... with relative energy 0.74 kcal/mol
 conformation 3: Alpha-L/N, A... with relative energy 1.6 kcal/mol
 conformation 4: Epsilon-L/N, A... with relative energy 1.7 kcal/mol
 conformation 5: Delta-D/N, A... with relative energy 3.4 kcal/mol

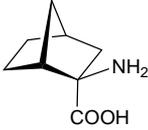
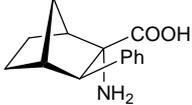
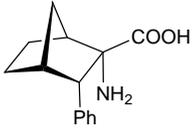
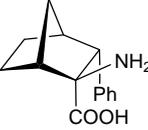
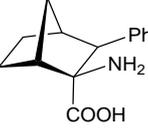
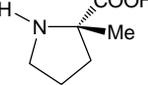
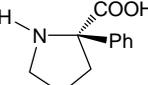
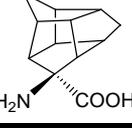
The dihedral values (degrees) are the following ones:
 For conformation (1) omega1=174.8, phi=71.9, psi=18.8, omega2=174.5
 For conformation (2) omega1=174.8, phi=75.78, psi=17.81, omega2=174.81

Figure 6.1.3. Interface system used to interact with the NCAD data base. (a) Window used to perform the search. (b) Window showing details about the theoretical study at the highest quantum mechanical level reported in the literature for the nc-aa selected; the RasMol 3D-view of a selected minimum energy conformation and the report are also displayed (both are generated through internal calls of the interface upon the user's request).

Table 6.1.1.. α -Tetrasubstituted α -amino acids stored in NCAD. The systematic name (and the common and abbreviated names,^a when available), the chemical structure, and the reference reporting the quantum mechanical study are given for each amino acid.

Abbreviated name	Systematic name (common name)	Structure	Ref.
Aib	2-amino-2-methylpropanoic acid (α -aminoisobutyric acid)		20
Ac ₃ c	1-aminocyclopropanecarboxylic acid		21
(1 <i>S</i> ,2 <i>S</i>) _c Phe	(1 <i>S</i> ,2 <i>S</i>)-1-amino-2-phenylcyclopropanecarboxylic acid		21
(1 <i>S</i> ,2 <i>R</i>) _c Phe	(1 <i>S</i> ,2 <i>R</i>)-1-amino-2-phenylcyclopropanecarboxylic acid		21
L-c ₃ Dip	(<i>S</i>)-1-amino-2,2-diphenylcyclopropanecarboxylic acid		22
(2 <i>S</i> ,3 <i>S</i>) _c diPhe	(2 <i>S</i> ,3 <i>S</i>)-1-amino-2,3-diphenylcyclopropanecarboxylic acid		23
Dpg	2-amino-2,2-diphenylacetic acid (diphenylglycine)		24
Dbg	2-amino-2-benzyl-3-phenylpropanoic acid (dibenzylglycine)		25
Ac ₄ c	1-aminocyclobutanecarboxylic acid		26
Ac ₅ c	1-aminocyclopentanecarboxylic acid		27

Adt	4-amino-1,2-dithiolane-4-carboxylic acid		28
(1 <i>S</i> ,2 <i>S</i>) _c Phe	(1 <i>S</i> ,2 <i>S</i>)-1-amino-2-phenylcyclopentanecarboxylic acid		29
(1 <i>S</i> ,2 <i>R</i>) _c Phe	(1 <i>S</i> ,2 <i>R</i>)-1-amino-2-phenylcyclopentanecarboxylic acid		29
(1 <i>S</i> ,2 <i>S</i>) _c Arg	(1 <i>S</i> ,2 <i>S</i>)-1-amino-2-(guanidinomethyl)-cyclopentanecarboxylic acid		30
(1 <i>S</i> ,2 <i>R</i>) _c Arg	(1 <i>S</i> ,2 <i>R</i>)-1-amino-2-(guanidinomethyl)-cyclopentanecarboxylic acid		30
Ac ₆ c	1-aminocyclohexanecarboxylic acid		31
(1 <i>S</i> ,2 <i>S</i>) _c Phe	(1 <i>S</i> ,2 <i>S</i>)-1-amino-2-phenylcyclohexanecarboxylic acid		32
(1 <i>S</i> ,2 <i>R</i>) _c Phe	(1 <i>S</i> ,2 <i>R</i>)-1-amino-2-phenylcyclohexanecarboxylic acid		32
Toac	4-amino-2,2,6,6-tetramethylpiperidine-1-oxyl-4-carboxylic acid		33
—	(1 <i>S</i> ,2 <i>R</i> ,4 <i>R</i>)-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid		34

—	(1 <i>R</i> ,2 <i>R</i> ,4 <i>S</i>)-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid		34
—	(1 <i>S</i> ,2 <i>S</i> ,3 <i>R</i> ,4 <i>R</i>)-2-amino-3-phenylbicyclo[2.2.1]heptane-2-carboxylic acid		34
—	(1 <i>S</i> ,2 <i>S</i> ,3 <i>S</i> ,4 <i>R</i>)-2-amino-3-phenylbicyclo[2.2.1]heptane-2-carboxylic acid		34
—	(1 <i>R</i> ,2 <i>S</i> ,3 <i>R</i> ,4 <i>S</i>)-2-amino-3-phenylbicyclo[2.2.1]heptane-2-carboxylic acid		34
—	(1 <i>R</i> ,2 <i>S</i> ,3 <i>S</i> ,4 <i>S</i>)-2-amino-3-phenylbicyclo[2.2.1]heptane-2-carboxylic acid		34
L-(α Me)Pro	(2 <i>S</i>)-2-methylpyrrolidine-2-carboxylic acid [(α -methyl)proline]		35
L-(α Ph)Pro	(2 <i>R</i>)-2-phenylpyrrolidine-2-carboxylic acid [(α -phenyl)proline]		35
—	(1 <i>S</i> ,2 <i>R</i> ,3 <i>R</i> ,5 <i>R</i> ,6 <i>R</i> ,7 <i>S</i> ,8 <i>R</i> ,9 <i>R</i> ,10 <i>R</i>)-8-aminopentacyclo[5.4.0.0 ^(2,6) .0 ^(3,10) .0 ^(5,9)]-undecane-8-carboxylic acid		36
—	(<i>S</i>)-4-aminopentacyclo[6.3.0.0 ^(2,6) .0 ^(3,10) .0 ^(5,9)]-undecane-4-carboxylic acid		37

^a In the abbreviated name, the configuration is defined by the *L/D* nomenclature when only the α carbon is chiral (*L* usually corresponding to *S*). When two or more chiral centers are present, the *R/S* stereochemical descriptors are used instead.

Figure 6.1.4. presents Ramachandran maps corresponding to the (ϕ, ψ) backbone dihedral angles of all the energy minima reported for the 29 α -tetrasubstituted α -amino acids included in the data base.²⁰⁻³⁷ Those chiral in nature (Figure 6.1.4.a) are

represented together with the minima calculated by *ab initio* quantum mechanical methods for the *N*-acetyl-*N*'-methylamide derivative of L-alanine³⁸ (MeCO-Ala-NHMe), as a reference for proteinogenic amino acids. The non-chiral α -tetrasubstituted residues (Figure 6.1.4.b) are compared with the glycine minima (calculated *ab initio* on MeCO-Gly-NHMe³⁸). The latter map is centrosymmetric, since two points (ϕ, ψ) and $(-\phi, -\psi)$ are equivalent for achiral amino acids.

As can be seen, tetrasubstitution at the α carbon affects the location of the minimum energy conformations typically found for proteinogenic residues. Specifically, the γ_L conformation (intramolecularly hydrogen-bonded seven-membered ring, also denoted C₇ conformation), which was found³⁸ at $(\phi, \psi) = (-86, 79)$ and $(-86, 72)$ for Ala and Gly, respectively, tends to evolve towards lower absolute ψ values upon replacement of the α -hydrogen atom by a substituent. The distortion of the ψ dihedral angle shown by quaternary α -amino acids is maximal for 1-aminocyclopropanecarboxylic acid (Ac₃c, Table 6.1.1.), which presents a γ_L minimum at $(\phi, \psi) = (-77, 34)$,²¹ as a consequence of the peculiar stereoelectronic properties of the three-membered ring.³

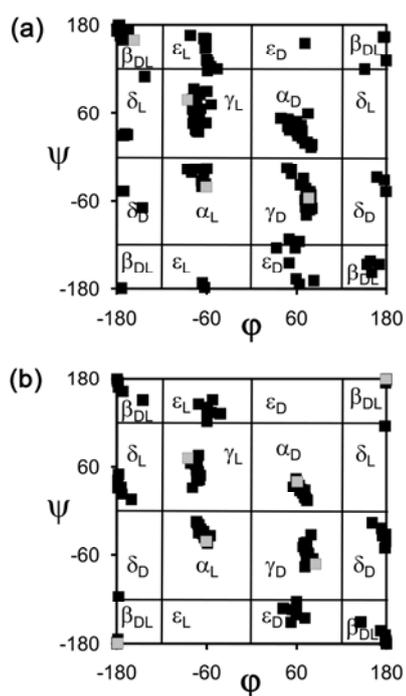


Figure 6.1.4. Ramachandran maps showing all the minimum energy conformations predicted by quantum mechanical calculations for the α -tetrasubstituted α -amino acids included in NCAD (see Table 6.1.1.): (a) chiral residues (black squares) compared with Ala (grey squares); (b) achiral residues (black squares) compared with Gly (grey squares).

The ψ backbone dihedral angle characterized for the α -helix conformation also deviates to lower absolute values when the α carbon is tetrasubstituted. Thus, the α_L conformation of Ala was located at $(\phi, \psi) = (-61, -41)$,³⁸ (as was for Gly³⁸) whereas that of its α -methylated counterpart, α -aminoisobutyric acid (Aib, Table 6.1.1.) is found at $(-65, -31)$.²⁰ A similar trend, or even more pronounced, is observed for most amino acids in Table 6.1.1., either chiral or not. Thus, for example, Ac₅c, Ac₆c, and L-c₃Dip exhibit α_L minima with (ϕ, ψ) values $(-73, -15)$,²⁷ $(-70, -20)$,³¹ and $(-80, -20)$,²² respectively. This apparently minor change brings about important consequences and, indeed, Aib and other quaternary α -amino acids are known to stabilize the 3_{10} -helix¹⁹ rather than the α -helix typically found for proteinogenic amino acids. The slight difference in the backbone dihedral angles translates into significant variations in the parameters associated to each helical structure, including the hydrogen bonding pattern, which involves residues i and $i+3$ in the 3_{10} -helix and residues i and $i+4$ in the α -helix.⁴⁰ Quaternary α -amino acids can therefore be used to design helical structures exhibiting hydrogen-bonding schemes and geometries different from those formed by standard amino acids.

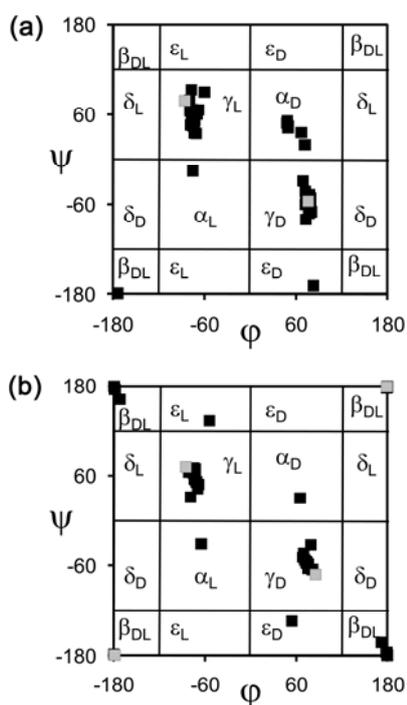


Figure 6.1.5. As in Figure 6.1.4., but only minimum energy conformations with a relative energy below 2.0 kcal/mol are considered.

Another effect of α -tetrasubstitution is the stabilization of minimum energy conformations that are not usually detected experimentally for coded amino acids because they are energetically unfavored. This feature is clearly observed when only the minima found for quaternary amino acids within a relative energy interval of 2.0 kcal/mol, *i.e.* those energetically accessible, are considered and compared with Ala or Gly (see Figures 6.1.5.a and 6.1.5.b for chiral and achiral residues, respectively). Specifically, the ϵ_D and ϵ_L regions of the (ϕ, ψ) map are visited by some α -tetrasubstituted amino acids, like Aib²⁰ and (1*S*,2*R*)_{c3}Phe²¹ but not by Ala³⁸ or Gly.³⁸ Furthermore, the α_L conformation is energetically accessible to several of the nc-aa considered, whereas the Ala³⁸ α_L minimum is unfavored by near 4.0 kcal/mol with respect to the global minimum and the same holds true for Gly.³⁸ Moreover, minima in the α_D region are found under the threshold of 2.0 kcal/mol for several amino acids in Table 6.1.1., among which are not only those of achiral nature, and therefore presenting an identical propensity to accommodate α_L and α_D conformations, but also some exhibiting an L configuration, like (1*S*,2*S*)_{c5}Arg.³⁰ Therefore, tetrasubstitution at the α carbon may result in the stabilization of conformations typically accessed by D amino acids even if the L chirality is maintained.

The features commented above illustrate the great conformational impact that may derive from the presence of α -tetrasubstituted amino acids in a peptide chain, which has been demonstrated in numerous investigations.^{19,40} Nowadays, the usefulness of this family of nc-aa in the control of peptide conformation is beyond doubt, but the practical application to the design of biologically relevant molecules or peptide-based materials is only in its earliest stages. NCAD is intended to mean a significant contribution to its development.

6.1.3. Results and discussion

***In Silico* Molecular Engineering for Targeted Replacements in Met-Enkephalin Using NCAD: A Test Case**

We provide in this section a computational design study aimed at constructing mutants of Met-Enkephalin using information extracted from NCAD. This is not intended to provide a deep and rigorous investigation of the resulting peptides, but to illustrate the potential utility of the data base in applications involving peptide and protein engineering through a test case.

The Met-enkephalin neurotransmitter is a pentapeptide (Tyr-Gly-Gly-Phe-Met) that interacts with opioid receptors.⁴¹ Opioid systems play a significant role in pain medication and opiate dependence, regulate dopamine release and are thought to be important for drug-induced reward. Interestingly, no unique native structure has been found for Met-enkephalin, whose conformational flexibility has been examined using a wide variety of experimental techniques (*e.g.* X-ray crystallography, NMR, circular dichroism, and infrared, ultraviolet and fluorescence spectroscopies).⁴² Accordingly, this short peptide is able to adopt different conformations depending on the biological context, which explains its ability to interact with opioid receptors of the μ , κ , and δ types.⁴¹ The multiple conformational states accessible to Met-enkephalin are therefore at the basis of its poor selectivity of action. A number of theoretical studies using both conventional and advanced simulation methods based on molecular dynamics (MD) and Monte Carlo (MC) algorithms have been devoted to explore the energy landscape of Met-enkephalin and have shown its ill-defined conformational state.^{43–45}

We have examined the information stored in NCAD to propose targeted replacements able to stabilize one of the conformations experimentally detected for Met-enkephalin. Specifically, we considered the conformation identified by NMR for Met-enkephalin in fast-tumbling bicelles.⁴⁶ The atomic coordinates of this conformation were extracted from the Protein Data Bank¹⁶ (PDB code: 1PLW).

The peptide was placed in the center of a cubic simulation box ($a = 44.6 \text{ \AA}$) filled with 2916 explicit water molecules, which were represented using the TIP3 model.⁴⁷ The *N*- and *C*-termini were described using the positively charged ammonium and the negatively charged carboxylate groups, respectively. All the MD simulations were performed using the NAMD computer program.⁴⁸ In order to equilibrate the density of the simulation box, different consecutive rounds of short MD runs were performed. Thus, 0.5 ns of NVT-MD at 350K were used to homogeneously distribute the solvent in the box. Next, 0.5 ns of NVT-MD at 298K (thermal equilibration) and 0.5 ns of NPT-MD at 298K (density relaxation) were run. After this, the peptide conformation was equilibrated by running a 3 ns trajectory of NVT-MD at 298K. The last snapshot of this trajectory was used to determine the (ϕ, ψ) dihedral angles of the three central residues of Met-enkephalin (Gly-Gly-Phe), which define the most flexible region of the peptide. The (ϕ, ψ) values thus obtained were (83,77) for Gly2, (61,-81) for Gly3, and (-66,-38) for Phe4, which correspond

to the α_D , γ_D , and α_L conformations, respectively, according to Perczel's nomenclature.¹⁴

These structural parameters have been used to design targeted substitutions in Met-enkephalin aimed at reducing the flexibility of the peptide backbone so that the particular conformation selected is stabilized. This molecular design pursues the validation of the performance and usefulness of the NCAD data base. The following criteria were considered in the design: (i) the side chain functionality of the nc-aa selected should be similar to that of the natural residue to be replaced, which is essential to preserve the biological functions of the peptide; and (ii) the intrinsic conformational properties of the selected nc-aa must be consistent with the conformation adopted by the targeted residue in the parent peptide. Based on these selection criteria and the structural parameters provided by MD for the Gly-Gly-Phe peptide fragment, three different mutations were proposed using the information in NCAD:

(a) Replacement of Gly2 or Gly3 by α -aminoisobutyric acid (Aib, see Table 6.1.1.). This amino acid bears two methyl groups at the α position (instead of the two α -hydrogen atoms in Gly) and presents low-energy minima²⁰ located at the α_L and γ_L regions of the Ramachandran map, with (ϕ, ψ) angles $(-65, -31)$ and $(-76, 58)$, and relative energies 1.7 and 0.0 kcal/mol, respectively, at the MP2/6-31G(d)//HF/6-31G(d) level. Given the non-chiral nature of Aib, these minimum energy conformations are equivalent to the enantiomeric α_D and γ_D , characterized by (ϕ, ψ) dihedrals $(65, 31)$ and $(76, -58)$, respectively. The latter values fit very well with those provided by MD for the Gly2 $[(83, 77)]$ and Gly3 $[(61, -81)]$ residues of Met-enkephalin. The mutants resulting from these substitutions are denoted G2 and G3.

(b) Replacement of Phe4 by $(2R, 3R)$ -1-amino-2,3-diphenylcyclopropanecarboxylic acid, known in the abbreviated form as $(2R, 3R)c_3diPhe$. This cyclic α -tetrasubstituted residue contains a cyclopropane ring bearing two vicinal phenyl substituents in a *trans* relative disposition [see the $(2S, 3S)$ enantiomer in Table 6.1.1.]. It retains the side chain of Phe and shows an energy minimum of the α_L type characterized by (ϕ, ψ) values $(-67, -27)$,²³ which closely resemble those obtained for Phe4 in Met-enkephalin in the MD trajectory described above, $(-66, -38)$. Moreover, $(2R, 3R)c_3diPhe$ has been

shown experimentally to exhibit a high propensity to adopt α_L conformations due to the interaction between the two rigidly held aromatic groups and the peptide backbone.⁴⁹ The mutant resulting from the replacement of Phe4 by (2*R*,3*R*)₃diPhe in Met-enkephalin is denoted F4.

(c) MD trajectories of 6 ns were performed for the wild-type peptide and the three mutants designed. The starting geometries for the mutants were constructed using the coordinates stored in the data base for the minimum energy conformation selected for each nc-aa. The protocols applied for thermal equilibration and density relaxation were identical to those described above. All each nc-aa. The protocols applied for thermal equilibration and density relaxation were identical to

those described above. All the force-field parameters for the MD simulations were extracted from the AMBER libraries,⁵⁰ with the exception of the electrostatic charges for the nc-aa, which were taken from the papers quoted in the data base.

Figure 6.1.6. shows the accumulated Ramachandran plots obtained through 6 ns of simulation corresponding to the Gly2, Gly3 and Phe4 residues for the wild-type peptide. As can be seen, the two Gly residues sample all the accessible regions of the (ϕ , ψ) space, which is consistent with the large conformational flexibility observed for Met-enkephalin in aqueous solution both by experimental^{42a,b,d-f} and theoretical techniques.⁴³⁻⁴⁵ The Phe4 residue also exhibits a great flexibility, although it only explores conformations in the left-half part of the map, as expected from its L configuration (in comparison, for the non-chiral Gly residue, the left and right parts of the map are energetically equivalent). Therefore, all three residues are actually found to adopt (ϕ , ψ) values in all the regions that are energetically accessible to them. In other words, no selectivity is detected towards a particular structure among those allowed to L (in the case of Phe) or achiral (in the case of Gly) coded residues. It is worth noting that the sampling of the conformational space shown in Figure 6.1.6. is in excellent agreement with that obtained for the same peptide using replica exchange MD.⁴⁵

Figure 6.1.6.

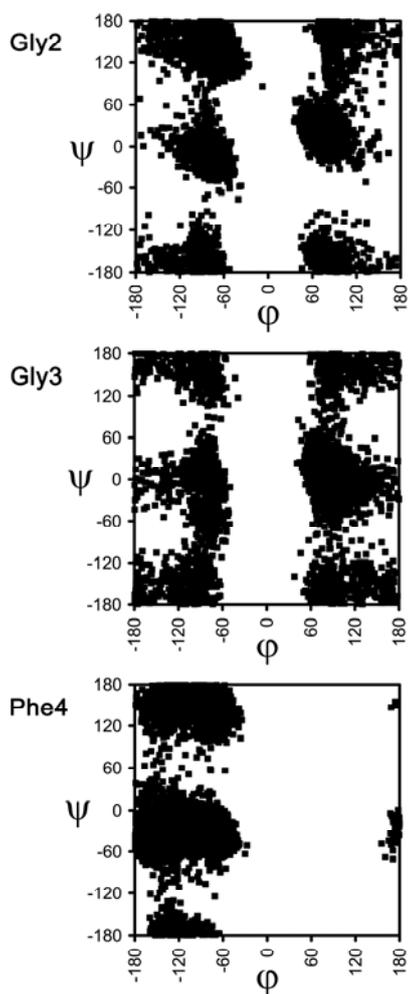


Figure 6.1.7.

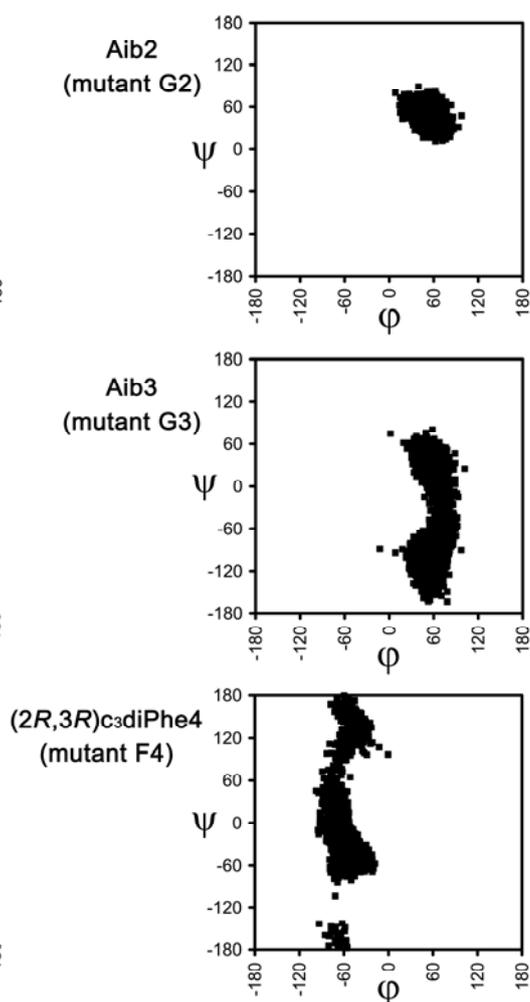


Figure 6.1.6. Backbone (ϕ, ψ) dihedral angles distributions obtained for the Gly2, Gly3, and Phe4 residues in wild-type Met-enkephalin.

Figure 6.1.7. Backbone (ϕ, ψ) dihedral angles distributions obtained for the Aib2, Aib3, and (2R,3R) c_3 diPhe4 residues in the G2, G3, and F4 mutants of Met-enkephalin.

Figure 6.1.7. shows the (ϕ, ψ) distributions obtained for the nc-aa included in mutants G2, G3 and F4 of Met-enkephalin. Table 6.1.2. displays the population of the different backbone conformations, which is expressed in terms of number of visits to a certain region of the Ramachandran map during the MD trajectories. These

data provide evidence for the great conformational selectivity induced by the α -tetrasubstituted amino acids used. In particular, for the G2 mutant, the Aib residue (Gly2 substitute) falls exclusively in the α_D region. Accordingly, the substitution designed using NCAD successfully restricts the conformational flexibility of the Gly2 residue and, moreover, preserves the selected arrangement. Thus, the replacement of Gly2 by Aib results in the stabilization of the conformation adopted by the coded residue in the starting geometry.

Replacement of Gly3 by Aib also has a conformational confinement effect although less intense. Indeed, only the α_D , γ_D , and ε_D structures are adopted by the Aib residue in the G3 mutant (Figure 6.1.7.), their respective populations being 29.5%, 66.1% and 4.4% (Table 6.1.2.). In contrast, Gly3 in the unmodified peptide was found to visit all nine regions used by Perczel *et al.*¹⁴ to describe the conformational space. We also note that the two regions most populated by Gly3 in the wild-type system, namely α_D and γ_D , are kept by Aib in the G3 mutant. Thus, although the Gly-Aib exchange proved less restrictive in G3 than in G2, the results obtained for the G3 mutant illustrate again that properly selected nc-aa can be successfully used to reduce the conformational freedom of biological (macro)molecules through the stabilization of selected secondary structure motifs.

Table 6.1.2. Relative shares of populations (%) found for selected residues in wild-type Met-enkephalin (Gly2, Gly3, and Phe4) and for the α -tetrasubstituted α -amino acids included in the G2, G3, and F4 mutants.^a

	α_D	ε_D	γ_D	δ_D	β_{DL}	δ_L	γ_L	ε_L	α_L
Gly2	25.1	8.9	3.1	0.5	5.0	1.0	10.1	34.0	12.3
Gly3	29.2	7.3	30.3	3.6	7.0	1.9	5.7	7.3	7.7
Phe4	0.0	0.0	0.0	19.1	18.4	3.5	2.9	21.7	34.4
Aib2 (G2)	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Aib3 (G3)	29.5	4.4	66.1	0.0	0.0	0.0	0.0	0.0	0.0
(2 <i>R</i> ,3 <i>R</i>) ₃ diPhe4 (F4)	0.0	0.0	0.0	0.0	0.0	0.0	13.1	4.7	82.2

^a The nine types of backbone conformation defined in ref. 14 have been considered

Table 6.1.3.. Ratio between the populations of conformations^a found for the α -tetrasubstituted α -amino acids in the G2, G3, and F4 mutants and the corresponding proteinogenic residues in wild-type Met-enkephalin.

	α_D	ϵ_D	γ_D	δ_D	β_{DL}	δ_L	γ_L	ϵ_L	α_L
Aib2 (G2)	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Aib3 (G3)	1.0	0.6	2.2	0.0	0.0	0.0	0.0	0.0	0.0
(2 <i>R</i> ,3 <i>R</i>) _c 3diPhe4 (F4)	– ^b	– ^b	– ^b	0.0	0.0	0.0	4.5	0.2	2.4

^a The nine types of backbone conformation defined in ref. 14 have been considered. ^b This conformation was not adopted by the α -tetrasubstituted residue.

To further evidence the magnitude of the conformational restrictions introduced by the nc-aa in this study, Table 6.1.3. displays the ratio between the populations of the conformations found for the quaternary α -amino acids and those of the corresponding coded residues in the wild-type peptide. For instance, mutant G3 increases the population obtained for the γ_D conformation of Gly3 in Met-enkephalin by a factor of 2.2.

In the F4 mutant, the restricted geometry of (2*R*,3*R*)_c3diPhe affects mainly the ϕ angle, which remains fixed in the neighborhood of -60° during the whole MD trajectory (Figure 6.1.7.). In comparison, the Phe4 residue in Met-enkephalin was found to cover all the possible negative values of ϕ : $-180^\circ < \phi < -50^\circ$ (Figure 6.1.6.). As a result, the α_L conformation, which corresponds to that adopted by Phe4 in the original arrangement selected, significantly increases its population from the wild-type peptide (34.4%) to the F4 mutant (82.2%) and becomes the most visited structure during the simulation of the latter peptide. The selectivity ratio for the F4 mutant is 4.5 and 2.4 for the γ_L and α_L regions, respectively (Table 6.1.3.). Accordingly, the nc-aa used as a Phe4 substitute also stabilizes the desired conformation and further proves the utility and potential applicability of the new data base.

6.1.4. Conclusions

NACD (Non-Coded Amino acids Data base) is an easy-to-use research tool that integrates the intrinsic conformational preferences of non-coded amino acids to facilitate their use in the design of peptide- or protein-based compounds useful in different fields of the life and materials sciences. Our aim is to provide the outcome of structural studies of nc-aa, including their properties and chemical characteristics, obtained by scientists in the organic and physical chemistry disciplines to those focusing on biological, bioengineering and medicinal applications, bringing the fields together. The information provided in NCAD, which is not available in other data bases, is also expected to be a useful resource for protein designers, modelers and experts in bioinformatics. The present paper presents the structure and informatics architecture of the data base, the interface that connects the data base with the users, the description of the first family of nc-aa integrated into the data base (α -tetrasubstituted α -amino acids), and a test case showing the applicability of the data base. Researchers can now easily select the most appropriate residue from a collection of nc-aa with well established conformational properties and introduce it in a peptide sequence or any other (macro)molecule. Additional families of non-proteinogenic amino acids will be included in NCAD in due course.

6.1.5. Bibliography

- [1] (a) *Amino Acids, Peptides and Proteins in Organic Chemistry*; Volume 1: *Origins and Synthesis of Amino Acids*. Hughes, A. B., Ed.; Wiley-VCH: Weinheim, 2009. (b) Barrett, G. C.; Elmore, D. T. *Amino Acids and Peptides*. Cambridge University Press: Cambridge, 1998. (c) Mauger, A. B. *J. Nat. Prod.* 1996, 59, 1205. (d) Wagner, I.; Musso, H. *Angew. Chem. Int. Ed. Engl.* 1983, 22, 816.
- [2] (a) *Asymmetric Synthesis and Application of α -Amino Acids*. Soloshonok, V. A., Izawa, K. Eds.; American Chemical Society: Washington DC, 2009. (b) Cativiela, C.; Ordóñez, M. *Tetrahedron: Asymmetry* 2009, 20, 1. (c) Michaux, J.; Niel, G.; Campagne, J.-M. *Chem. Soc. Rev.* 2009, 38, 2093. (d) Ordóñez, M.; Rojas-Cabrera, H.; Cativiela, C. *Tetrahedron* 2009, 65, 17. (e) Cativiela, C.; Díaz-de-Villegas, M. D. *Tetrahedron: Asymmetry* 2007, 18, 569. (f) Ordóñez, M.; Cativiela, C. *Tetrahedron: Asymmetry* 2007, 18, 3. (g) Bonauer, C.; Walenzyk, T.; König, B. *Synthesis* 2006, 1. (h) Aurelio, L.; Brownlee, R. T. C.; Hughes, A. B. *Chem. Rev.* 2004, 104, 5823. (i) Lelais, G.; Seebach, D. *Biopolymers (Pept. Sci.)* 2004, 76, 206.
- [3] (a) *Peptaibiotics: Fungal Peptides Containing α -Dialkyl α -Amino Acids*. Toniolo, C., Brückner, H., Eds.; Wiley-VCH: Weinheim, 2009. (b) *Handbook of Biologically Active Peptides*. Kastin, A. J., Ed.; Academic Press: London, 2006. (c) Degenkolb, T.; Bruckner, H. *Chem. Biodiversity* 2008, 5, 1817. (d) Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* 2003, 20, 275.
- [4] (a) *Peptide and Protein Design for Biopharmaceutical Applications*. Jensen, K. J., Ed.; John Wiley & Sons: Chichester, 2009. (b) Nestor, J. J. *Curr. Med. Chem.* 2009, 16, 4399. (c) Horne, W. S.; Gellman, S. H. *Acc. Chem. Res.* 2008, 41, 1399. (d) Seebach, D.; Gardiner, J. *Acc. Chem. Res.* 2008, 41, 1366. (e) Chatterjee, J.; Gilon, C.; Hoffman, A.; Kessler, H. *Acc. Chem. Res.* 2008, 41, 1331. (f) Cowell, S. M.; Lee, Y. S.; Cain, J. P.; Hruby, V. J. *Curr. Med. Chem.* 2004, 11, 2785. (g) Sagan, S.; Karoyan, P.; Lequin, O.; Chassaing, G.; Lavielle, S. *Curr. Med. Chem.* 2004, 11, 2799. (h) Adessi, C.; Soto, C. *Curr. Med. Chem.* 2002, 9, 963.
- [5] (a) *Pharmaceutical Substances: Syntheses, Patents, Applications*. Kleemann, A., Engel, J., Kutscher, B., Reichert, D., Eds.; Thieme: Stuttgart, 2009. (b) Ersmark, K.; Del Valle, J. R.; Hanessian, S. *Angew. Chem. Int. Ed.* 2008, 47, 1202. (c) Menard, J.; Patchett, A. A. *Adv. Protein Chem.* 2001, 56, 13.

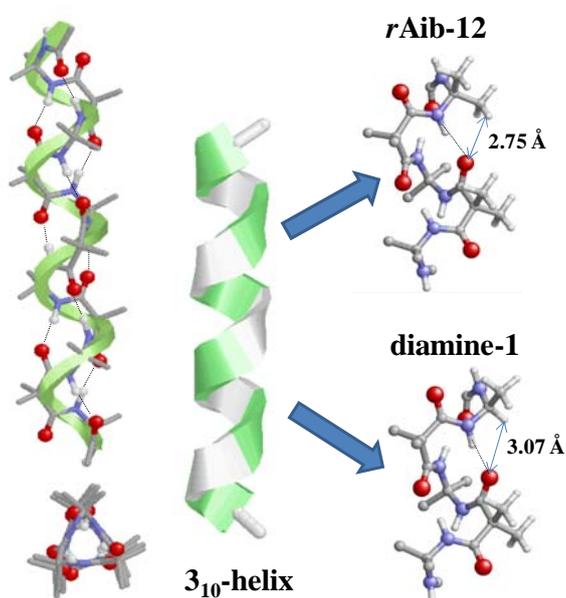
- [6] (a) Voloshchuk, N.; Montclare, J. K. *Mol. BioSyst.* 2010, 6, 65. (b) Wu, X.; Schultz, P. G. *J. Am. Chem. Soc.* 2009, 131, 12497. (c) Hartman, M. C. T.; Josephson, K.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 4356. (d) Wang, L.; Schultz, P. G. *Angew. Chem. Int. Ed.* 2005, 44, 34. (e) Hendrickson, T. L.; De Crécy-Lagard, V.; Schimmel, P. *Annu. Rev. Biochem.* 2004, 73, 147. (f) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* 2001, 292, 498.
- [7] (a) Lee, H. S.; Guo, J.; Lemke, E.; Dimla, R.; Schultz, P. G. *J. Am. Chem. Soc.* 2009, 131, 12921. (b) Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 9785. (c) Vázquez, M. E.; Nitz, M.; Stehn, J.; Yaffe, M. B.; Imperiali, B. *J. Am. Chem. Soc.* 2003, 125, 10150. (d) Murakami, H.; Hohsaka, T.; Ashizuka, Y.; Hashimoto, K.; Sisido, M. *Biomacromolecules* 2000, 1, 118.
- [8] (a) Alfonta, L.; Zhang, Z.; Uryu, S.; Loo, J. A.; Schultz, P. G. *J. Am. Chem. Soc.* 2003, 125, 14662. (b) Shinohara, H.; Kusaka, T.; Yokota, E.; Monden, R.; Sisido, M. *Sens. Actuators B* 2000, 65, 144.
- [9] (a) Lemke, E. A.; Summerer, D.; Geierstanger, B. H.; Brittain, S. M.; Schultz, P. G. *Nat. Chem. Biol.* 2007, 3, 769. (b) Bose, M.; Groff, D.; Xie, J.; Brustad, E.; Schultz, P. G. *J. Am. Chem. Soc.* 2006, 128, 388. (c) Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokohama, S. *Nat. Methods* 2005, 2, 201. (d) Rothman, D. M.; Vázquez, M. E.; Vogel, E. M.; Imperiali, B. *Org. Lett.* 2002, 4, 2865.
- [10] (a) Wang, W.; Takimoto, J. K.; Louie, G. V.; Baiga, T. J.; Noel, J. P.; Lee, K. F.; Slesinger, P. A.; Wang, L. *Nat. Neurosci.* 2007, 10, 1063. (b) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* 2002, 99, 19.
- [11] (a) Cellitti, S. E.; Jones, D. H.; Lagpacan, L.; Hao, X.; Zhang, Q.; Hu, H.; Brittain, S. M.; Brinker, A.; Caldwell, J.; Bursulaya, B.; Spraggon, G.; Brock, A.; Ryu, Y.; Uno, T.; Schultz, P. G.; Geierstanger, B. H. *J. Am. Chem. Soc.* 2008, 130, 9268. (b) Reid, P. J.; Loftus, C.; Beeson, C. C. *Biochemistry* 2003, 42, 2441.
- [12] (a) Zanuy, D.; Ballano, G.; Jiménez, A. I.; Casanovas, J.; Haspel, N.; Cativiela, C.; Curc6, D.; Nussinov, R.; Alemán, C. *J. Chem. Inf. Model.* 2009, 49, 1623. (b) Cejas, M. A.; Kinney, W. A.; Chen, C.; Vinter, J. G.; Almond Jr., H. R.; Balss, K. M.; Maryanoff, C. A.; Schmidt, U.; Breslav, M.; Mahan, A.; Lacy, E.; Maryanoff, B. E. *Proc. Natl. Acad. Sci. U.S.A.* 2008, 105, 8513. (c) Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* 2007, 111, 3236. (d) Yang,

- Z. M.; Liang, G. L.; Ma, M. L.; Gao, Y.; Xu, B. *Small* 2007, 3, 558. (e) Crisma, M.; Toniolo, C.; Royo, S.; Jiménez, A. I.; Cattivola, C. *Org. Lett.* 2006, 8, 6091. (f) Behanna, H. A.; Donners, J. J. J. M.; Gordon, A. C.; Stupp, S. I. *J. Am. Chem. Soc.* 2005, 127, 1193. (g) Brea, R. J.; Amorín, M.; Castedo, L.; Granja, J. R. *Angew. Chem. Int. Ed.* 2005, 44, 5710. (h) Hartgerink, J. D. *Curr. Opin. Chem. Biol.* 2004, 8, 604. (i) Hartgerink, J. D.; Beniash, E.; Stupp, S. I. *Science* 2001, 294, 1684. (j) Clark, T. D.; Buehler, L. K.; Ghadiri, M. R. *J. Am. Chem. Soc.* 1998, 120, 651.
- [13] (a) Lee, M.-R.; Stahl, S. S.; Gellman, S. H.; Masters, K. S. *J. Am. Chem. Soc.* 2009, 131, 16779. (b) Connor, R. E.; Tirrell, D. A. *Polym. Rev.* 2007, 47, 9. (c) Börner, H. G.; Schlaad, H. *Soft Matter* 2007, 3, 394. (d) Tu, R. S.; Tirrell, M. *Adv. Drug Delivery Rev.* 2004, 56, 1537. (e) Kirshenbaum, K.; Zuckermann, R. N.; Dill, K. A. *Curr. Opin. Struct. Biol.* 1999, 9, 530. (f) Kirshenbaum, K.; Barron, A. E.; Goldsmith, R. A.; Armand, P.; Bradley, E. K.; Truong, K. T. V.; Dill, K. A.; Cohen, F. E.; Zuckermann, R. N. *Proc. Natl. Acad. Sci. U.S.A.* 1998, 95, 4303. (g) Wang, X.; Huq, I.; Rana, T. M. *J. Am. Chem. Soc.* 1997, 119, 6444. (h) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* 1994, 27, 5471. (i) Cho, C. Y.; Moran, E. J.; Cherry, S. R.; Stephans, J. C.; Fodor, S. P. A.; Adams, C. L.; Sundaram, A.; Jacobs, J. W.; Schultz, P. G. *Science* 1993, 261, 1303.
- [14] (a) Perczel, A.; Angyán, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* 1991, 113, 6256. (b) Hudáky, I.; Kiss, R.; Perczel, A. *J. Mol. Struct.* 2004, 675, 177.
- [15] Allen, F. H. *Acta Crystallogr. B* 2002, 58, 380.
- [16] (a) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucl. Acids Res.* 2000, 28, 235. (b) Sussman, J. L.; Lin, D.; Jiang, J.; Manning, N. O.; Prilusky, J.; Ritter, O.; Abola, E. E. *Acta Crystallogr. D* 1998, 54, 1078.
- [17] Chen, P. P.-S. *ACM Trans. Data base Syst.* 1976, 1, 9.
- [18] Sayle, R. A.; Milner-White, E. J. *Trends Biochem. Sci.* 1995, 20, 374.
- [19] (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.
- [20] Alemán, C. *J. Phys. Chem. B* 1997, 101, 5046.

- [21] Alemán, C.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Casanovas, J. *J. Phys. Chem. B* 2002, *106*, 11849.
- [22] Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Alemán, C. *J. Phys. Chem. B* 2006, *110*, 5762.
- [23] Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Alemán, C. *J. Org. Chem.* 2003, *68*, 7088.
- [24] Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *J. Org. Chem.* 2007, *72*, 2174.
- [25] Casanovas, J.; Nussinov, R.; Alemán, C. *J. Org. Chem.* 2008, *73*, 4205.
- [26] Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *Chem. Phys. Lett.* 2006, *429*, 558.
- [27] Alemán, C.; Zanuy, D.; Casanovas, J.; Cativiela, C.; Nussinov, R. *J. Phys. Chem. B* 2006, *110*, 21264.
- [28] Aschi, M.; Lucente, G.; Mazza, F.; Mollica, A.; Morera, E.; Nalli, M.; Paradisi, M. P. *Org. Biomol. Chem.* 2003, *1*, 1980.
- [29] Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Org. Chem.* 2008, *73*, 644.
- [30] Revilla-López, G.; Torras, J.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Org. Chem.* 2009, *74*, 2403.
- [31] Rodríguez-Ropero, F.; Zanuy, D.; Casanovas, J.; Nussinov, R.; Alemán, C. *J. Chem. Inf. Model.* 2008, *48*, 333.
- [32] Alemán, C.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Casanovas, J. *J. Org. Chem.* 2009, *74*, 7834.
- [33] D'Amore, M.; Improta, R.; Barone, V. *J. Phys. Chem. A* 2003, *107*, 6264.
- [34] Cordoní, A.; Gomez-Catalan, J.; Jimenez, A. I.; Cativiela, C.; Perez, J. J. *J. Pept. Sci.* 2002, *8*, 253.
- [35] Flores-Ortega, A.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C.; Casanovas, J. *J. Org. Chem.* 2008, *73*, 3418.
- [36] Bisetty, K.; Gomez-Catalan, J.; Alemán, C.; Giralt, E.; Kruger, H. G.; Perez, J. J. *J. Pept. Sci.* 2004, *10*, 274.
- [37] Bisetty, K.; Govender, P.; Kruger, H. G. *Biopolymers* 2006, *81*, 339.
- [38] Cornell, W. D.; Gould, I. R.; Kollman, P. A. *J. Mol. Struct. (Theochem)* 1997, *392*, 101.
- [39] de Meijere, A. *Angew. Chem. Int. Ed. Engl.* 1979, *18*, 809.

- [40] Toniolo, C.; Crisma, M.; Formaggio, F.; Peggion, C.; Broxterman, Q.; Kaptein, B. *J. Inclusion Phenom. Macrocyclic Chem.* 2005, *51*, 121.
- [41] (a) Bodnar, R. J. *Peptides* 2009, *30*, 2432. (b) McLaughlin, P. J. *Proenkephalin-Derived Opioid Peptides*. In *Handbook of Biologically Active Peptides*; Kastin, A. J., Ed.; Academic Press: London, 2006, pp. 1313–1318. (c) Hughes, J.; Smith, T. W.; Kosterlitz, H. W.; Fothergill, L. A.; Morgan, B. A.; Morris, H. R. *Nature* 1975, *258*, 577.
- [42] (a) Cai, X.; Dass, C. *Rapid Commun. Mass Spectrom.* 2005, *19*, 1. (b) Graham, W. H.; Carter II, E. S.; Hicks, R. P. *Biopolymers* 1992, *32*, 1755. (c) Griffin, J. F.; Langs, D. A.; Smith, G. D.; Blundell, T. L.; Tickle, I. J.; Bedarkar, S. *Proc. Natl. Acad. Sci. U.S.A.* 1986, *83*, 3272. (d) Higashijima, T.; Kobayashi, J.; Nagai, U.; Miyazawa, T. *Eur. J. Biochem.* 1979, *97*, 43. (e) Spirtes, M. A.; Schwartz, R. W.; Mattice, W. L.; Coy, D. H. *Biochem. Biophys. Res. Commun.* 1978, *81*, 602. (f) Khaled, M. A.; Long, M. M.; Thompson, W. D.; Bradley, R. J.; Brown, G. B.; Urry, D. W. *Biochem. Biophys. Res. Commun.* 1977, *76*, 224.
- [43] (a) Koca, J.; Carlsen, H. J. *J. Mol. Struct.* 1995, *337*, 17. (b) Perez, J. J.; Loew, G. H.; Villar, H. O. *Int. J. Quantum Chem.* 1992, *44*, 263.
- [44] (a) Hansmann, U. H. E.; Okamoto, Y.; Onuchic, J. N. *Proteins* 1999, *34*, 472. (b) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* 1987, *84*, 6611.
- [45] Sanbonmatsu, K. Y.; García, A. E. *Proteins* 2002, *46*, 225.
- [46] Marcotte, I.; Separovic, F.; Auger, M.; Gagné, S. M. *Biophys. J.* 2004, *86*, 1587.
- [47] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* 1983, *79*, 926.
- [48] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* 2005, *26*, 1781.
- [49] (a) Crisma, M.; De Borggraeve, W. M.; Peggion, C.; Formaggio, F.; Royo, S.; Jiménez, A. I.; Cativiela, C.; Toniolo, C. *Chem. Eur. J.* 2006, *12*, 251. (b) Royo, S.; De Borggraeve, W. M.; Peggion, C.; Formaggio, F.; Crisma, M.; Jiménez, A. I.; Cativiela, C.; Toniolo, C. *J. Am. Chem. Soc.* 2005, *127*, 2036.
- [50] 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.

6.2. Integrating the intrinsic conformational preferences of non-coded α -amino acids modified at the peptide bond into the NCAD database



6.2.1 Introduction

Apart from the genetically coded α -amino acids found in proteins, there is a large number of amino acids grouped under the common name of *non-proteinogenic* or *non-coded* amino acids (nc-aa). Some of these nc-aa are present in nature, occurring in non-ribosomally produced peptides (*e.g.* antibiotics); however, most are synthesized by chemists. Among the latter, those used for protein engineering¹⁻⁶ and nanobiology⁷⁻¹² are particularly interesting, and their applications extend to many fields such as pharmacology (drug design),^{13,14} biotechnology (biosensors),^{15,16} and nanomedicine (systems for drug delivery and diagnosis through imaging).^{17,18}

In spite of the great significance and potential utility of nc-aa in modern biology, biomaterials engineering, and medicine, structural information is available only for a limited number of these. In the last two decades, the intrinsic conformational preferences of several tens of nc-aa have been investigated through sophisticated quantum mechanical calculations at the *ab initio* or Density Functional Theory (DFT) levels. Specifically, these reliable theoretical methods have been used to sample the

potential energy hypersurface of small peptide systems incorporating such nc-aa in the absence of external forces. At the same time, crystallographic and spectroscopic structural data on peptide sequences incorporating nc-aa are becoming available for an increasing number of cases. Comparison of such experimental information with that derived from high-level theoretical calculations helps establish the influence of packing forces, solvent effects, and the chemical environment on the conformational propensities of nc-aa.

In practice, the use of nc-aa with well-characterized conformational properties is frequently limited because the relevant information is highly dispersed. Quantum mechanical calculations describing the intrinsic conformational preferences of these compounds are typically reported in physical chemistry journals, while their synthesis and crystalline structures developed by organic and peptide chemists are published in journals specialized in these fields. In contrast, many applications of nc-aa are tested by researchers working on protein engineering, medicine, or materials science. In order to integrate the dispersed contributions about nc-aa, which is expected to facilitate their practical application in many research fields, a database of nc-aa was recently built.¹⁹

NCAD (Non-Coded Amino acids Database) is a database conceived and created to identify the nc-aa that are compatible with a given structural motif, which is the key requirement for the application of these compounds in the life and materials sciences. The database, which is publicly accessible through a web page, integrates structural and energetic descriptors of those nc-aa whose intrinsic conformational properties have been previously studied using *ab initio* or DFT quantum mechanical calculations and reported in the literature. In our recent work¹⁹ we described the technical aspects of the database (structure, contents, informatics features and relationships among the different descriptors) and presented the first family of nc-aa integrated into it. Specifically, we discussed the relevance of α -tetrasubstituted α -amino acids in the design of peptides with well-defined conformational properties and described the 29 α -tetrasubstituted α -amino acids whose intrinsic conformational propensities have been determined to date using *ab initio* or DFT quantum mechanical methods according to a literature search.¹⁹

In this work, we extend NCAD to three new families of compounds: (i) *N*-substituted α -amino acids; (ii) thio- α -amino acids; and (iii) diamines and diacids surrogates of α -amino acids that are used to build pseudopeptides containing one (or

more) retropeptide link(s), that is, a peptide bond in reversed direction to that of its immediate neighbors. The members of three such series can be used to engineer peptides and proteins incorporating modifications at the conventional –NHCO– peptide bond. Below, the work is organized as follows. First, a brief description of the contents and structure of NCAD is provided. After this, the three families of amino acids integrated in the database are presented. Next, the utility of the database is illustrated by engineering helical models using the information of the diamines and diacids included in NCAD. Finally, the conclusions of the work are outlined.

NCAD: Integrating the intrinsic conformational preferences of non-coded amino acids

NCAD integrates information about nc-aa whose intrinsic conformational preferences have been studied by applying *ab initio* or DFT quantum mechanical calculations to the smallest peptide systems which contains them (that is, the nc-aa bearing the amino and carboxylic acid termini as an amide, typically capped with the acetyl and methylamide groups). The information contained in NCAD is divided into three categories:

- 1) Minimum energy conformations. All the minima found for each compound contained in the database are described in detail. Dihedral angles, atomic coordinates and relative energies obtained at the highest level of theory among those used to study each nc-aa are stored through inter-connected descriptors. The backbone conformations of the different energy minima contained in NCAD have been identified using the nomenclature proposed by Perzcel *et al.*,²⁰ which distinguish nine different conformations in the potential energy surface $E=E(\varphi,\psi)$ of α -amino acids (*i.e.* γ_D , δ_D , α_D , ϵ_D , β_{DL} , ϵ_L , α_L , δ_L , and γ_L). This information is essential to check the compatibility of the different energy minima with the secondary structural motifs typically found in peptides and proteins.
- 2) Bibliographic information describing the available experimental information. Specifically, the following items are reported for each amino acid: (i) synthesis and characterization of the free form or with the amino group adequately protected for use in peptide synthesis; and (ii) spectroscopic and structural characterization when incorporated into peptides, the atomic coordinates extracted from the X-ray crystal structures being also included in selected cases.

3) Applications. The most relevant publications and patents describing the use of nc-aa in related fields along with the biological properties and relevance in materials science.

The labels identifying the information stored in the database are connected through a relational scheme. This facilitates the search process, which can be performed using a single or several descriptors. NCAD is accessed through a user-friendly interface, which integrates a simple window system, thus facilitating the use by non-experts in informatics. The interface is divided into six folders, the first one being used to define the criteria applied to the search process. The available search criteria are: (i) the chemical nature of the side chain; (ii) the structural feature which distinguishes a particular family of nc-aa; (iii) the molecular formula; (iv) the type of backbone conformation according to the nomenclature established by Perczel *et al.*;^{20,21} (v) the values of the (ϕ, ψ) dihedral angles, with the tolerance being specified by the user; (vi) the name of the researcher who performed the theoretical or experimental investigation; (vii) the experimental data reported in the literature relating to the synthesis and characterization; and (viii) the applications described in the literature. The amino acids fulfilling all the search criteria are displayed in a list. Once the user selects a given amino acid from this list, the information stored in the database is presented in the remaining five folders of the interface. The specific information provided in each folder was detailed in our previous work.¹⁹

6.2.2 Methods

Amino-acid surrogates leading to peptides with modified peptide bonds

In our previous work,¹⁹ we provided evidence for the usefulness of the α -tetrasubstituted α -amino acids stored in NCAD in the control of peptide conformation. Another means of modulating the conformational properties of peptides can come from the modification of the peptide bond itself. This change should perturb the intramolecular hydrogen bonding network, the electronic distribution, and the steric hindrance, thus affecting the conformational preferences of the peptide chain.

N-substituted α -amino acids. In *N*-substituted amino acids, the hydrogen atom attached to nitrogen is replaced by a functional group. The substituents most frequently used are: alkyl (*N*-alkylation), hydroxyl (*N*-hydroxylation) or amino (*N*-amination). To the best of our knowledge, reliable theoretical studies at the *ab initio* or DFT quantum mechanical levels on the intrinsic conformational preferences of *N*-

substituted amino acids have been performed only for *N*-methylalanine,²² *N*-hydroxyalanine²³ and *N*-aminoglycine²⁴ (Table 6.2.1.).

N-alkylation (in particular, *N*-methylation) has been largely used to change the pharmacological properties of peptides. Remarkable biological and pharmacological profiles have been observed for many *N*-methylated peptides.²⁵ The conformational preferences of *N*-methylated amino acids are dictated by increased steric hindrance and lack of hydrogen-bond donor ability associated with the replacement of the hydrogen atom by a methyl group.^{26,27} Figure 6.2.1.a presents the Ramachandran map corresponding to the (ϕ, ψ) backbone dihedral angles of all the energy minima reported for the *N*-acetyl-*N*'-methylamide derivatives of the *N*-methylated α -amino acid included in the database, namely *N*-methylalanine.²² The minima derived from *ab initio* quantum mechanical calculations for the analogous derivatives of glycine and alanine²⁸ (MeCO-Gly-NHMe and MeCO-Ala-NHMe, respectively) are also shown in Figure 6.2.1.a as a reference for proteinogenic amino acids. As it can be seen, the effect of the replacement of the amide hydrogen of Ala by a methyl group on the low-energy minimum (β_{DL}) is very significant, with the minimum being annihilated. The disappearance of this minimum should be attributed to the repulsive interactions produced by the confrontation of the *N*-methyl group and the oxygen atom of C=O moiety.

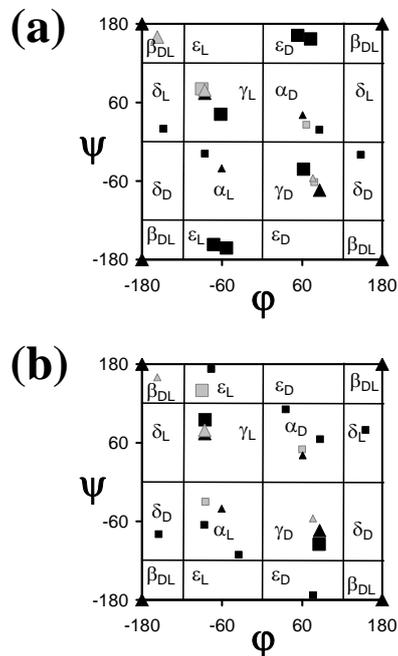


Figure 6.2.1. Ramachandran maps showing all the minimum energy conformations predicted by quantum mechanical calculations for *N*-modified α -amino acids included in NCAD (see Table 6.2.1.): (a) *N*-methylglycine (black squares) and *N*-methylalanine (gray squares); and (b) *N*-aminoglycine (black squares) and *N*-hydroxyalanine (gray squares). Minimum energy conformations of Gly (black triangles) and Ala (gray triangles) have been included for comparison. Minima with relative energies lower and higher than 2.0 kcal/mol are represented by large and small symbols, respectively.

The *N*-hydroxylation and *N*-amination of peptides have been used to impart conformational rigidity through the formation of intramolecular hydrogen bonds with neighboring C=O groups.^{29,30} Besides its hydrogen-bonding abilities, the *N*-hydroxyl group has been proven able to chelate metal ions, which may be of use for the design of peptides which can bind specifically to proteins containing metals in the active site.²⁹ The Ramachandran map in Figure 6.2.1.b shows all the energy minima reported for the *N*-acetyl-*N*'-methylamide derivatives of *N*-hydroxyalanine²³ and *N*-aminoglycine,²⁴ which are compared to those of Ala and Gly,²⁸ respectively. As can be seen, replacement of the amide hydrogen by a hydroxyl or amino group entails drastic conformational changes. The most remarkable ones can be summarized as follows: (i) the lowest energy minimum of Ala (γ_L) evolves towards ϵ_L upon *N*-hydroxylation; (ii) the extended β_{DL} conformation of Ala is not a minimum in *N*-hydroxyalanine; (iii) the steric interactions induced by the *N*-amino group produces a destabilization of the β_{DL} conformation in *N*-aminoglycine; and (iv) *N*-amination of Gly leads to the characterization of new local minima located at the ϵ_L , ϵ_D , δ_D , and δ_L regions of the Ramachandran map.

Thio- α -amino acids. Within the surrogates used to replace the amide bond of peptides, the thioamide functionality, in which the carbonyl oxygen is replaced by a sulfur atom, is amongst the most synthetically accessible.³¹⁻³³ This modification not only affects the proteolytic stability and bioactivity of peptides;³⁴⁻³⁶ it also produces significant changes in the secondary structure relative to the parent peptides.³⁷⁻³⁸ Both amides and thioamides are planar and exhibit high energy barriers for rotation about the C–N bond.^{40,41} However, the length of the C=S bond in thioamides (~0.45 Å larger than the C=O linkage in amides) and the distinct stereoelectronic features of the sulfur atom (its van der Waals radius is 32% larger than that of oxygen and sulfur is less electronegative) affect the hydrogen-bonding abilities of the thioamide bond and translate into a reduced proton-acceptor character.⁴²

The influence of this amide-thioamide replacement on the intrinsic conformational properties has been studied only for Gly and Ala (Table 6.2.1.).⁴³ It has been found that the sulfur atom reduces drastically the conformational flexibility of thioglycine when compared to the coded amino acid. Thus, the only minimum energy conformation identified for thioglycine,⁴³ which is two-fold degenerate, corresponds to a γ -turn motif (Figure 6.2.2.). Therefore, the minima found for Gly²⁸ at the β_{DL} and

α regions of the Ramachandran map are annihilated upon replacement of oxygen by sulfur. Figure 6.2.2. shows a Ramachandran map with the five minima found for thioalanine.⁴³ The three minima detected for Ala²⁸ were also identified as minimum energy conformations for thioalanine and exhibit similar relative energies. Thus, γ_L is the most stable conformation for the two amino acids while the relative energies of the γ_D and β_{DL} structures differ in 0.6 and 1.4 kcal/mol, respectively. The relative energies of the two additional minima characterized for thioalanine, α_L and δ_D , are above 4 kcal/mol. Accordingly, the changes produced in Ala by the presence of the sulfur atom are considerably less important than those observed for Gly.

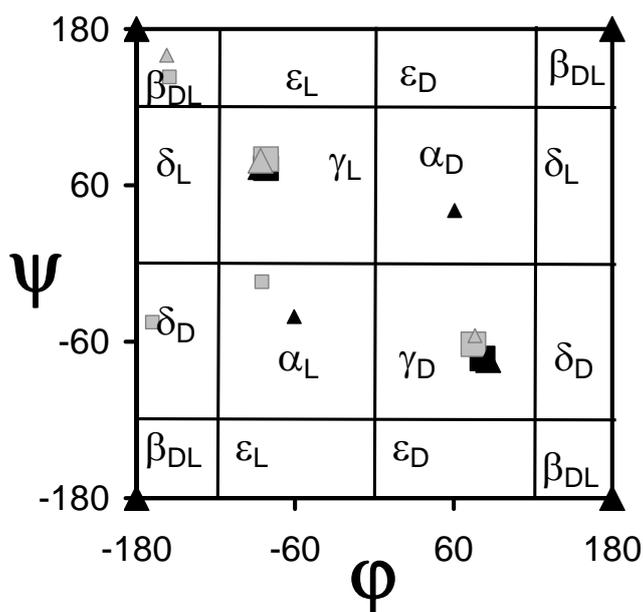
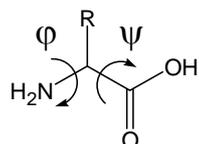


Figure 6.2.2. Ramachandran maps showing all the minimum energy conformations predicted by quantum mechanical calculations for Thio- α -amino acids included in NCAD (see Table 6.2.1.): thio-glycine (black squares) and thio-alanine (gray squares). Minimum energy conformations of Gly (black triangles) and Ala (grey triangles) have been included for comparison. Minima with relative energies lower and higher than 2.0 kcal/mol are represented by large and small symbols, respectively.

Table 6.2.1. *N*-substituted α -amino acids and thio- α -amino acids stored in NCAD. The systematic name (and the common name), the chemical structure, and the reference reporting the quantum mechanical study are given for each amino acid.

Systematic name (common name)	Structure	Ref.
(<i>S</i>)- <i>N</i> -methyl-2-aminopropionic acid (<i>N</i> -methyl-L-alanine)		22
(<i>S</i>)- <i>N</i> -hydroxy-2-aminopropionic acid (<i>N</i> -hydroxy-L-alanine)		23
hydrazinoacetic acid (<i>N</i> -aminoglycine)		24
aminothioacetic acid (thioglycine)		43
(<i>S</i>)-2-aminothiopropionic acid (L-thioalanine)		43

Diamides and diacids used to build retropeptide bonds. In the last decades, a great deal of effort has been devoted to the chemistry of the so-called retropeptides or partially modified retropeptides,⁴⁴⁻⁵⁰ in which the direction of all or some peptide bonds is reversed, *i.e.* NH-CO instead of CO-NH (Scheme 1). When the stereochemistry of one or more amino acids of the reversed segment is inverted, the



α -amino acid



diamine

diacid

Scheme 2

The Ramachandran map in Figure 6.2.3.a shows the location of the minimum energy conformations found for the N,N' -dimethylamide derivatives of the diacid surrogates of Gly, Ala, Aib, Val, and Δ Ala, hereafter denoted dc-Gly,⁶¹ dc-Ala, dc-Aib, dc-Val,⁶² and dc- Δ Ala,⁶³ respectively (Table 6.2.2.). The minimum energy conformations of the dc-Ala and dc-Aib derivatives were not reported in the literature and have been calculated in this work at the B3LYP/6-31G(d) level (the dihedral angles and relative energies are given in the Supporting Information). The resulting conformations have been used for engineering a 3_{10} -helical retropeptide (see below).

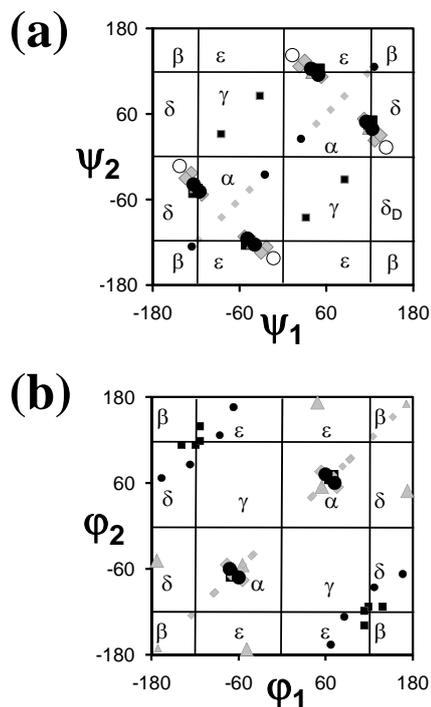
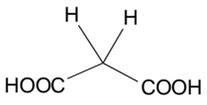
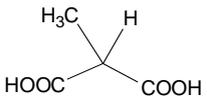
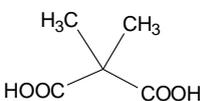
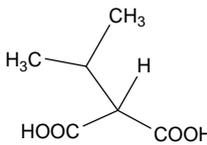
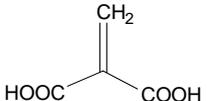
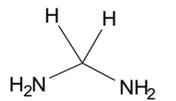
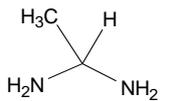
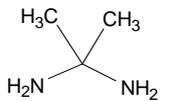
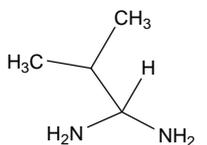


Figure 6.2.3. Ramachandran maps showing all the minimum energy conformations predicted by quantum mechanical calculations for diacids (a) and diamines (b) analogues of α -amino acids included in NCAD (see Table 6.2.1.). (a) Diacids: dc-Gly (black squares), dc-Ala (black circles), dc-Aib (gray triangles), dc-Val (gray diamonds) and dc- Δ Ala (empty circles); (b) Diamines: dm-Gly (black squares), dm-Ala (black circles), dm-Aib (gray triangles) and dm-Val (gray diamonds). The nomenclature used for the different diacids and diamines is discussed in the text. Minima with relative energies lower and higher than 2.0 kcal/mol are represented by large and small symbols, respectively.

Table 6.2.2.. Diacids and diamines analogues of α -amino acids stored in NCAD. The systematic name (and the abbreviated name used in this work), the chemical structure, and the reference reporting the quantum mechanical study are given for each amino acid.

Systematic name (abbreviated name)	Structure	Ref.
malonic acid (dc-Gly)		61
methylmalonic acid (dc-Ala)		This work
dimethylmalonic acid (dc-Aib)		This work
isopropylmalonic acid (dc-Val)		62
methylenemalonic acid (dc- Δ Ala)		64
methylenediamine (dm-Gly)		60
1,1-ethylenediamine (dm-Ala)		60
2,2-propylenediamine (dm-Aib)		This work
1,1-isobutylenediamine (dm-Val)		62

Due to molecular symmetry, structures characterized by dihedral angles ψ_1, ψ_2 ; $-\psi_1, -\psi_2$; ψ_2, ψ_1 ; and $-\psi_2, -\psi_1$ are energetically degenerate (Figure 6.2.3.a). The lowest energy conformation was found at $\psi_1, \psi_2 \approx 45^\circ, 120^\circ$ for all the diacids but dc- Δ Ala, in which case, the global minimum appears at $\psi_1, \psi_2 \approx 13^\circ, 142^\circ$. This small difference should be attributed to the hybridization state of the α carbon atom, namely sp^2 for dc- Δ Ala and sp^3 for dc-Gly, dc-Ala, dc-Aib, and dc-Val. These non-symmetric conformations are very similar to those found in the crystal state for malonamide derivatives.⁶⁴ On the other hand, Figure 6.2.3.a shows that the minima of dc- Δ Ala with relative energies larger than 2 kcal/mol (*i.e.* minima with $\psi_1, \psi_2 = \pm 25.0^\circ, \pm 25.0^\circ$ and $\pm 126.0^\circ, \pm 126.0^\circ$) are located in the map diagonals. However, the lowest energy conformation remains asymmetric ($\psi_1 \neq \psi_2$), even though this diacid is chemically symmetric.

Figure 6.2.3.b shows the energy minima found for the N, N' -diacetyl derivatives of the Gly, Ala, Aib and Val diamine surrogates, hereafter denoted dm-Gly,⁶⁰ dm-Ala,⁶⁰ dm-Aib, and dm-Val,⁶² respectively. The minimum energy conformations of the dm-Aib derivative have been calculated in this work at the B3LYP/6-31G(d) level (see the Supporting Information). For the reason mentioned above, structures with ϕ_1, ϕ_2 ; $-\phi_1, -\phi_2$; ϕ_2, ϕ_1 ; and $-\phi_2, -\phi_1$ are energetically equivalent, as reflected by the map symmetry. As can be seen, the lowest energy conformations of the diamines are located in the α region at $\phi_1, \phi_2 \approx 60^\circ, 60^\circ$. In addition, dm-Aib shows a low energy minima destabilized by 1.0 kcal/mol at $\phi_1, \phi_2 = 49^\circ, 173^\circ$. Minima with relative energies higher than 2 kcal/mol are essentially located in the map diagonal as well as in the regions labeled as ϵ and δ . As expected, the γ regions are scarcely populated for diamines and diacids derivatives (Figure 6.2.3.), whereas they are found to be preferred by the N -acetyl- N' -methylamide derivatives of proteinogenic amino acids. This is because in the latter compounds, the γ conformations are stabilized by an intramolecular hydrogen bond defining a seven-membered ring, whereas this interaction is not possible in diamines and diacids.

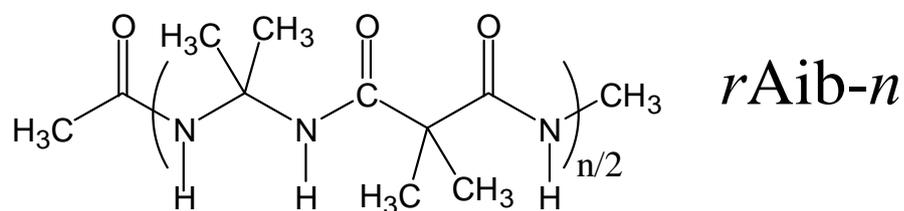
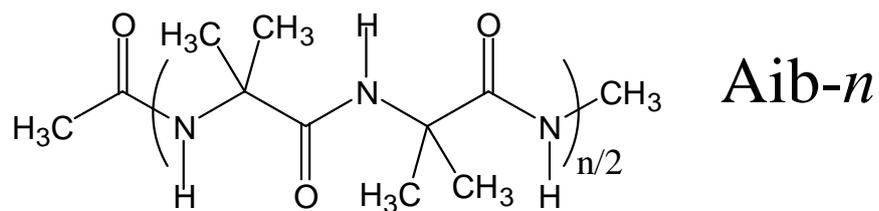
6.2.3 Results and discussion

***In silico* molecular engineering of a 3_{10} -helical motif using NCAD: A test case**

In this section, we present a computational design study aimed at constructing non-symmetric retropeptides exhibiting a 3_{10} -helix conformation by using information extracted from NCAD. This is not intended to provide a deep and rigorous investigation of the resulting retropeptides, but to illustrate the potential utility of the database in applications involving peptide and protein engineering through a test case.

It is well known that Aib-based peptides have a remarkable tendency to adopt helical conformations both in solution and in the solid state.⁶⁵⁻⁶⁹ Helices of the 3_{10} type are overwhelmingly formed by Aib homopeptides, independently of the environmental conditions, whereas small peptides containing a mixture of Aib and proteinogenic α -amino acids are found to accommodate either α - or 3_{10} -helices depending mainly on the peptide size and the Aib content.

In an earlier work, we used retropeptide bonds to enhance the stability of the 3_{10} -helix with respect to the α -helix in peptides containing Aib.⁷⁰ Specifically, Aib-containing homoretropeptides (hereafter denoted $rAib-n$, where n indicates the number of residues) cannot form the intramolecular hydrogen-bonding network that stabilizes the α -helix ($i \leftarrow i+4$) and this results in unfavorable interactions between the C=O groups of residues i and $i+4$. In contrast, the molecular architecture of $rAib-n$ (Scheme 3) in which diacid and diamine surrogates of Aib are alternated, (dm-Aib-dc-Aib) _{$n/2$} , is fully compatible with the hydrogen-bonding scheme of the 3_{10} -helix ($i \leftarrow i+3$). These features are schematically depicted in Figure 6.2.4., whereas Figure 6.2.5. shows the 3_{10} -helix adopted by $rAib-n$.



Scheme 3

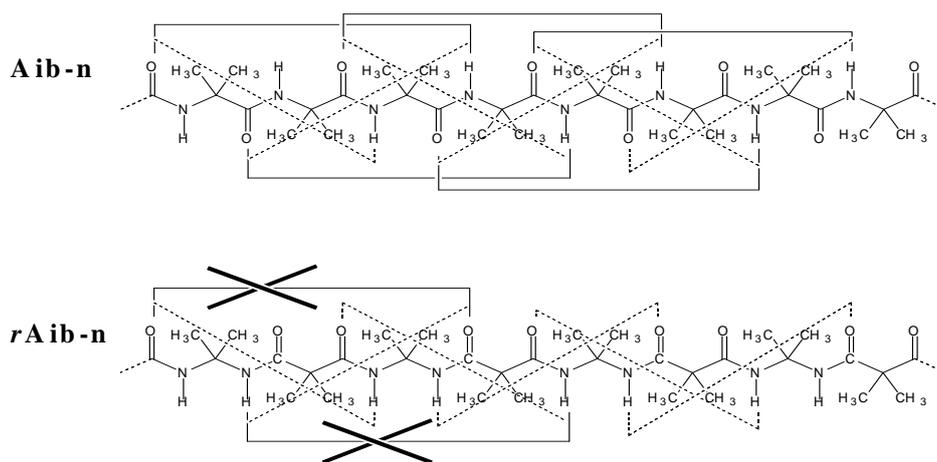


Figure 6.2.4. Hydrogen bonding network for the α - and 3_{10} -helices ($i \leftarrow i+4$ solid line and $i \leftarrow i+3$ dashed line, respectively) of Aib-*n* and *r*Aib-*n*.

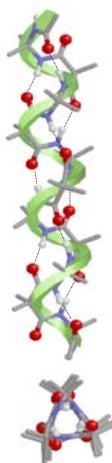


Figure 6.2.5. Axial (top) and equatorial (bottom) representation of the 3_{10} -helix adopted by *rAib-n*.

rAib-n are symmetric molecules, *i.e.* the two ends of the peptide are identical, which may be a limitation in some recognition processes. The information stored in NCAD for diamines and diacids was carefully examined considering the similarities between the backbone dihedral angles and relative stability of the minima found for dc-Aib and dm-Aib. Analysis of the conformational preferences of these residues indicates that two different strategies can be followed to design a non-symmetric retropeptide based on *rAib-n* through a single mutation without decreasing the stability of the 3_{10} -helix. The first strategy requires the replacement of dc-Aib or dm-Aib by dc-Ala or dm-Ala, respectively (*i.e.* removal of a methyl group in a diacid or diamide). The second involves the replacement of either dc-Aib or dm-Aib by the corresponding Val analogue. Although the helical conformation is energetically accessible for the diacids and, specially, the diamines of both Ala and Val, the larger conformational flexibility of the Val derivatives and the larger steric interactions expected for the isopropyl group in comparison to the methyl substituent suggest that the replacement of dc-Aib or dm-Aib by the corresponding Ala analogue is more appropriate. Accordingly, the first strategy was selected for further investigation.

We first examined the most appropriate position to carry out the targeted replacement. For this purpose, six mutants of a retropeptide containing 12 residues, *rAib-12* [*i.e.* (dm-Aib – dc-Aib)₆], were considered, their sequences being:

Diamine-1: dm-Ala – dc-Aib – (dm-Aib – dc-Aib)₅

Diacid-1: dm-Aib – dc-Ala – (dm-Aib – dc-Aib)₅

Diamine-2: (dm-Aib – dc-Aib) – dm-Ala – dc-Aib – (dm-Aib – dc-Aib)₄

Diacid-2: (dm-Aib – dc-Aib) – dm-Aib – dc-Ala – (dm-Aib – dc-Aib)₄

Diamine-3: (dm-Aib – dc-Aib)₂ – dm-Ala – dc-Aib – (dm-Aib – dc-Aib)₃

Diacid-3: (dm-Aib – dc-Aib)₂ – dm-Aib – dc-Ala – (dm-Aib – dc-Aib)₃

The six retropeptides were built up in a 3_{10} -helix conformation with initial backbone dihedral angles for diacids and diamines $\varphi_1, \varphi_2 = \psi_1, \psi_2 = -40.7^\circ, -30.3^\circ$. These values were selected in a previous work as being representative for this helical conformation.⁷¹ The molecular geometry of the six retropeptides was submitted to complete geometry optimizations using the B3LYP quantum mechanical method^{72,73} combined with the 6-31G(d) basis set. The relative energy of the six mutants under study is represented in Figure 6.2.6.. Interestingly, the three retropeptides mutated at a diacid position are systematically more stable than those mutated at the diamines, even though the energy difference is small (~ 1 kcal/mol). Furthermore, diacid-3 is slightly more stable than diacids-1 and -2. This result indicates that the most favorable mutation occurs at the center of the helix.

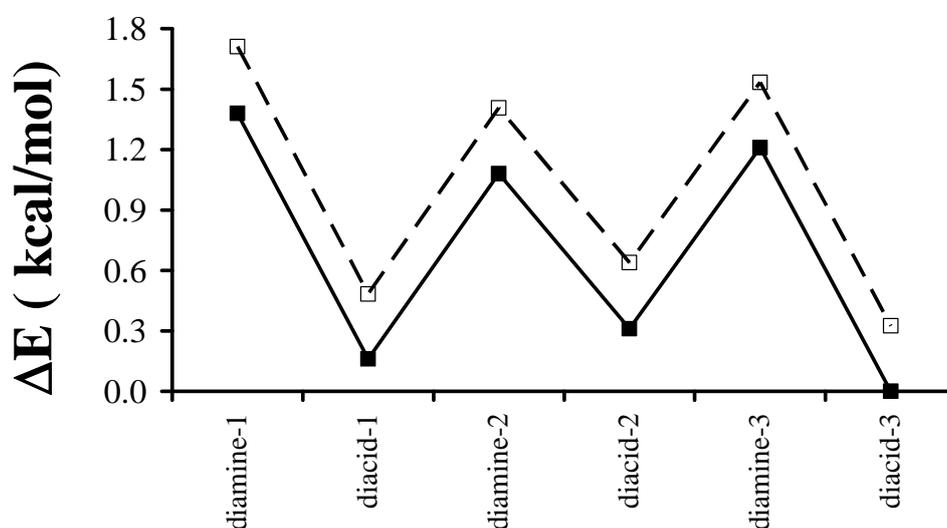


Figure 6.2.6. Relative energy (black squares and solid line) and Δ_h (empty squares and dashed line) of the six mutants of *rAib-12* studied in this work. The sequence of the mutants as well as the definition of Δ_h are provided in the text.

Comparison of the backbone dihedral angles obtained for the six optimized mutants to those derived from geometry optimization of *rAib-12* at the same theoretical level indicates that the conformational distortion introduced by dm-Ala and dc-Ala is almost negligible. Thus, the largest distortion was observed for diamine-1, where the φ_1 dihedral of dm-Ala differ in 16.2° from the corresponding angle in *rAib-12*. For the rest of the mutants, no deviation was found to exceed 10° . Similarly,

the end-to-end distance remained almost identical for the mutants and the parent peptide, the largest difference being 0.08 Å for diamine-1 and diacid-3. These results confirm that the helical stability of *rAib-12* is retained in the six mutants investigated. In order to quantitatively evaluate the effect of the mutation in the energetic stability of the 3_{10} -helix, the following isodesmic reaction has been considered:



Thus, the loss of stability of the helix (Δ_h) has been defined according to:

$$\Delta_h = E(\text{mutant}) + E(CH_3CH_3) - E(rAib-12) - E(CH_4) \quad (2)$$

where E refers to the energy of the different species. Figure 6.2.6. includes the values of Δ_h for the six mutants under study. As expected, the profile obtained for Δ_h fits that of the relative energies, even though the former is 0.3 kcal/mol higher. These results allow us to conclude that breaking the molecular symmetry of *rAib-12* by replacement of one Aib surrogate by the corresponding Ala analogue produces a destabilization of only 0.3 kcal/mol.

Inspection of the dipole moments calculated for the mutants indicates that incorporation of dm-Ala or dc-Ala produces a very small variation with respect to *rAib-12* ($\mu = 0.225$ Debyes), the largest variation being of only 0.092 Debyes (diamine-1). Accordingly, the energy preferences displayed in Figure 6.2.6. should not be attributed to electrostatic changes in the helix. However, a detailed analysis of the intramolecular distances may help explain such a relative energy profile. Specifically, inspection of *rAib-12* indicates that the methyl groups of the dm-Aib residues form a weak interaction with the oxygen atom of a proximal O=C moiety, which in turn, is involved in an intramolecular hydrogen bond with a neighboring amide hydrogen. This van der Waals interaction is illustrated in Figure 6.2.7., which also indicates that the average C-H...O=C distance is 2.75 ± 0.01 Å. Evaluation of such a distance for the six mutants under study shows that removal of the methyl group induces a variation that becomes significantly larger at the mutated positions (Figure 6.2.7.). Thus, the distance increases by 0.32 ± 0.02 and 0.25 ± 0.01 Å in the retropeptides mutated at the diamines and diacids, respectively, with respect to that found in *rAib-12*, and this may contribute to the stability differences observed (Figure 6.2.6.).

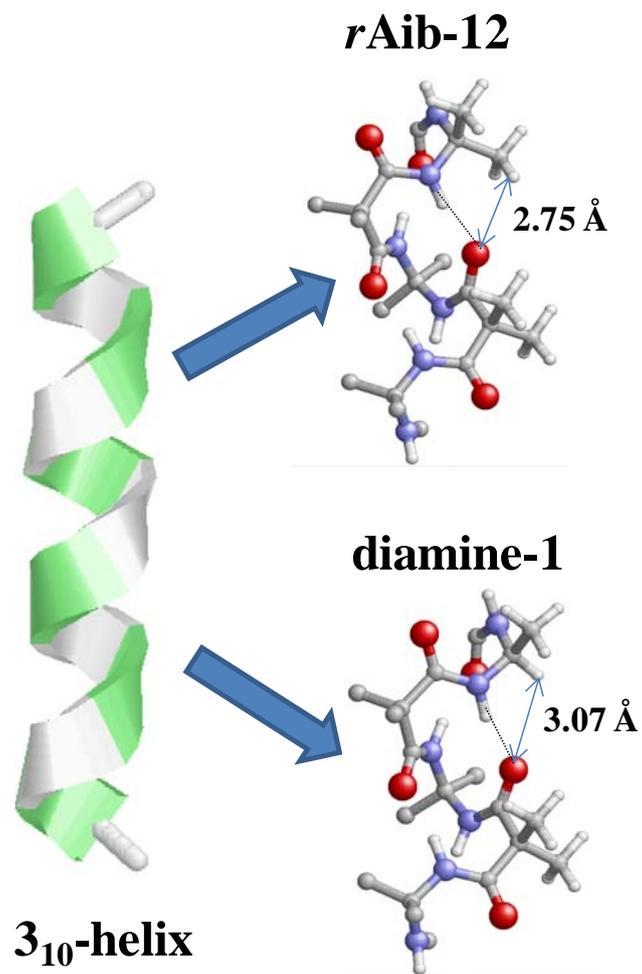


Figure 6.2.7. Detail about the C-H...O=C interaction (arrows) in the 3_{10} -helix of *rAib-12* and diamine-1. Averaged H...O distances are indicated.

6.2.4 Conclusions

In this work, NCAD has been extended through the incorporation of residues that may be used as building blocks for the construction of peptides with modified peptide bonds. The information related to *N*-substituted α -amino acids, thio- α -amino acids, and diamines and diacids used to build retropeptides described in the literature has been collected and integrated into the database. The conformational propensities of the new α -amino acid surrogates incorporated into NCAD have been compared to those of the corresponding coded residues. We also present an example illustrating the utility of NCAD and the usefulness of the new amino acid surrogates in peptide engineering. More specifically, the information contained in NCAD has been used to propose targeted replacements able to break the molecular symmetry of a retropeptide sequence designed to stabilize the 3_{10} -helix with respect to the α -helix.

6.2.5. References

- [1] Voloshchuk, N.; Montclare, J. K. *Mol. BioSyst.* 2010, 6, 65.
- [2] Wu, X.; Schultz, P. G. *J. Am. Chem. Soc.* 2009, 131, 12497.
- [3] Hartman, M. C. T.; Josephson, K.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 4356.
- [4] Wang, L.; Schultz, P. G. *Angew. Chem. Int. Ed.* 2005, 44, 34.
- [5] Hendrickson, T. L.; De Crécy-Lagard, V.; Schimmel, P. *Annu. Rev. Biochem.* 2004, 73, 147.
- [6] Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* 2001, 292, 498.
- [7] Zanuy, D.; Ballano, G.; Jiménez, A. I.; Casanovas, J.; Haspel, N.; Cativiela, C.; Curcó, D.; Nussinov, R.; Alemán, C. *J. Chem. Inf. Model.* 2009, 49, 1623.
- [8] Cejas, M. A.; Kinney, W. A.; Chen, C.; Vinter, J. G.; Almond Jr., H. R.; Balss, K. M.; Maryanoff, C. A.; Schmidt, U.; Breslav, M.; Mahan, A.; Lacy, E.; Maryanoff, B. E. *Proc. Natl. Acad. Sci. U.S.A.* 2008, 105, 8513.
- [9] Yang, Z. M.; Liang, G. L.; Ma, M. L.; Gao, Y.; Xu, B. *Small* 2007, 3, 558.
- [10] Crisma, M.; Toniolo, C.; Royo, S.; Jiménez, A. I.; Cativiela, C. *Org. Lett.* 2006, 8, 6091.
- [11] Brea, R. J.; Amorín, M.; Castedo, L.; Granja, J. R. *Angew. Chem. Int. Ed.* 2005, 44, 5710.
- [12] Hartgerink, J. D.; Beniash, E.; Stupp, S. I. *Science* 2001, 294, 1684.
- [13] Ersmark, K.; Del Valle, J. R.; Hanessian, S. *Angew. Chem. Int. Ed.* 2008, 47, 1202.
- [14] Menard, J.; Patchett, A. A. *Adv. Protein Chem.* 2001, 56, 13.
- [15] Lee, H. S.; Guo, J.; Lemke, E.; Dimla, R.; Schultz, P. G. *J. Am. Chem. Soc.* 2009, 131, 12921.
- [16] Shinohara, H.; Kusaka, T.; Yokota, E.; Monden, R.; Sisido, M. *Sens. Actuators B* 2000, 65, 144.
- [17] Cellitti, S. E.; Jones, D. H.; Lagpacan, L.; Hao, X.; Zhang, Q.; Hu, H.; Brittain, S. M.; Brinker, A.; Caldwell, J.; Bursulaya, B.; Spraggon, G.; Brock, A.; Ryu, Y.; Uno, T.; Schultz, P. G.; Geierstanger, B. H. *J. Am. Chem. Soc.* 2008, 130, 9268.
- [18] Reid, P. J.; Loftus, C.; Beeson, C. C. *Biochemistry* 2003, 42, 2441.
- [19] Revilla-López, G.; Torras, J.; Curcó, D.; Casanovas, J.; Calaza, M. I.; Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Grodzinski, P.; Alemán, C. *J. Phys. Chem. B* 2010, 114, 7413.

- [20] Perczel, A.; Angyán, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcocchia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* 1991, *113*, 6256.
- [21] Hudáky, I.; Kiss, R.; Perczel, A. *J. Mol. Struct.* 2004, *675*, 177.
- [22] Siodlak, D.; Gajewska, M.; Macedowska, A.; Rzeszotarska, B. *J. Mol. Struct. (Theochem)* 2006, *775*, 47.
- [23] Malde, A. K.; Khedkar, S. A.; Coutinho, E. C. *J. Chem. Theory Comput.* 2006, *2*, 312.
- [24] Alemán, C. *J. Phys. Chem. A* 2002, *106*, 144.
- [25] Chatterjee, J.; Gilon, C.; Hoffman, A.; Kessler, H. *Acc. Chem. Res.* 2008, *41*, 1331.
- [26] Teixido, M.; Albericio, F.; Giralt, E. *J. Pept. Res.* 2005, *65*, 153.
- [27] Rainaldi, M.; Moretto, V.; Crisma, M.; Peggion, E.; Mammi, S.; Toniolo, C.; Cavicchioni, G. *J. Peptide Sci.* 2002, *8*, 241.
- [28] Cornell, W. D.; Gould, I. R.; Kollman, P. A. *J. Mol. Struct. (Theochem)* 1997, *392*, 101.
- [29] Ye, Y.; Liu, M.; Kao, J. L. K.; Marshall, G. R. *Biopolymers* 2003, *71*, 489.
- [30] Dupont, V.; Lecoq, A.; Mangeot, J.-P.; Aubry, A.; Boussard, G.; Marraud, M. *J. Am. Chem. Soc.* 1993, *115*, 8898.
- [31] Clausen, K.; Thorsen, M.; Lawesson, S.-O. *Tetrahedron* 1981, *37*, 3635.
- [32] Ilankumaran, P.; Ramesha, A. R.; Chandrasekaran, S. *Tetrahedron Lett.* 1995, *36*, 8311.
- [33] Yokoyama, M.; Hasegawa, Y.; Hatanaka, H.; Kawazoe, Y.; Imamoto, T. *Synthesis* 1984, *84*, 827.
- [34] Bartlett, P. A.; Spear, K. L.; Jacobson, N. E. *Biochemistry* 1982, *21*, 1608.
- [35] Wildemann, D.; Schiene-Fischer, C.; Aumüller, T.; Bachmann, A.; Kiefhaber, T.; Lücke, C.; Fischer, G. *J. Am. Chem. Soc.* 2007, *129*, 4910.
- [36] Kessler, H.; Matter, H.; Geyer, A.; Diehl, H.; Kock, M.; Kurz, G.; Oppendoerf, F. R.; Callens, M.; Wierenga, R. K. *Angew. Chem., Int. Ed. Engl.* 1992, *31*, 328.
- [37] Guthrie, D. J. S.; Williams, C. H.; Elmore, D. T. *Int. J. Pept. Protein Res.* 1986, *28*, 208.
- [38] Hollósi, M.; Majer, M.; Zewdu, F.; Ruff, M.; Kajtár, M.; Kövér, K. E. *Tetrahedron* 1988, *44*, 195.
- [39] Doi, M.; Takehara, S.; Ishida, T.; Inoue, M. *Int. J. Pept. Protein Res.* 1989, *34*, 369.

- [40] Zhao, J.; Wildemann, D.; Jakob, M.; Vargas, C.; Schiene-Fischer, C. *Chem. Commun.* 2003, 2810.
- [41] Wiberg, K. B.; Rush, D. J. *J. Am. Chem. Soc.* 2001, *123*, 2038–2046
- [42] Alemán, C. *J. Phys. Chem. A* 2001, *105*, 6717.
- [43] Artis, D. R.; Lipton, M. A. *J. Am. Chem. Soc.* 1998, *120*, 12200.
- [44] Fletcher, M. D.; Campbell, M. M. *Chem. Rev.* 1998, *98*, 763.
- [45] Goodman, M.; Chorev, M. *Acc. Chem. Res.* 1979, *12*, 1.
- [46] Chorev, M.; Willson, C. G.; Goodman, M. *J. Am. Chem. Soc.* 1977, *99*, 8075.
- [47] Volonterio, A.; Bravo, P.; Zanda, M. *Org. Lett.* 2000, *2*, 1827.
- [48] Sani, M.; Bruche, L.; Chiva, G.; Fustero, S.; Piera, J.; Volonterio, A.; Zanda, M. *Angew. Chem. Int. Ed.* 2003, *42*, 2060.
- [49] Ceretti, S.; Luppi, G.; De Pol, S.; Fomaggio, F.; Crisma, M.; Toniolo, C.; Tomasini, C. *Eur. J. Org. Chem.* 2004, *20*, 4188.
- [50] Ando, S.; Mitsuyasu, K.; Soeda, Y.; Hidaka, M.; Ito, Y.; Matsubara, K.; Shindo, M.; Uchida, Y.; Aoyagi, H. *J. Pept. Sci.* 2010, *16*, 171.
- [51] Phan-Chan-Du, A.; Petit, M. C.; Guichard, G.; Briand, J. P.; Muller, S. *Biochemistry* 2001, *40*, 5720.
- [52] Pal-Bhowmick, I.; Pandey, R. P.; Jarori, G. K.; Kar, S.; Sadal, D. *Biochem. Biophys. Res. Commun.* 2007, *364*, 608.
- [53] Petit, M. C.; Benkirane, N.; Guichard, G.; Du A. P.; Marraud, M.; Cung, M. T.; Briand, J. P.; Muller, S. *J. Biol. Chem.* 1999, *274*, 3686.
- [54] Sakurai, K.; Chung, H. S.; Kahne, D. *J. Am. Chem.* 2004, *126*, 16288.
- [55] Guptasarma, P. *FEBS Lett.* 1992, *310*, 205.
- [56] Preissner, R.; Goede, A.; Michalski, E.; Frommel, C. *FEBS Lett.* 1997, *414*, 425.
- [57] Shukla, A.; Raje, M.; Guptasarma, P. *J. Biol. Chem.* 2003, *278*, 26505.
- [58] Alemán, C. *Proteins* 1997, *29*, 575.
- [59] Alemán, C.; Franco, L.; Puiggalí, J. *Macromolecules* 1994, *27*, 4298.
- [60] Alemán, C.; Puiggalí, J. *J. Org. Chem.* 1995, *60*, 910.
- [61] Sandrone, G.; Dixon, D.; Hay, B. P. *J. Phys. Chem. A* 1999, *103*, 3554.
- [62] Alemán, C. *J. Phys. Chem. B* 2001, *105*, 860.
- [63] Alemán, C. *J. Biomol. Struct. & Dyn.* 1996, *14*, 193.
- [64] Alemán, C.; Pérez, J. J. *J. Mol. Struct. (Theochem)* 1993, *285*, 221.
- [65] Di Blasio, B.; Pavone, V.; Lombardi, A.; Pedone, C.; Benedetti, E. *Biopolymers* 1993, *33*, 1037.

- [66] Balaram, P. *Curr. Opin. Struct. Biol.* 1992, 2, 845.
- [67] Toniolo, C.; Crisma, M.; Bonora, G. M.; Benedetti, E.; Di Blasio, B.; Pavone, V.; Pedone, C.; Santini, A. *Biopolymers* 1991, 31, 129.
- [68] Otda, K.; Kitagawa, Y.; Kimura, S.; Imanishi, Y. *Biopolymers* 1993, 33, 1337.
- [69] Toniolo, C.; Bonora, G. M.; Barone, V.; Bavoso, A.; Benedetti, E.; DiBlasio, B.; Grimaldi, P.; Lelj, F.; Pavone, V.; Pedone C. *Macromolecules* 1985, 18, 895.
- [70] Alemán, C. *Proteins* 1997, 29, 575.
- [71] Alemán, C.; Roca, R.; Luque, F. J.; Orozco, M. *Proteins* 1997, 28, 83.
- [72] Becke, A. D. *J. Chem. Phys.* 1993, 98, 1372.
- [73] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* 1988, 37, 785.
- [74] Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* 1973, 28, 213.

7. Summary and discussion of the results

The title of this thesis “A computational approach to the engineering of bioinspired nanosystems: design, applications and information management” clearly embeds the aim of this work. Computational methods in theoretical chemistry have proven to be useful and efficient tools for studying complex systems. This thesis has used several of these computational strategies to elucidate both structural features and physicochemical properties for new non-coded amino acids. This information has been used to modify bio-inspired systems. This subsequent work rationale is clearly shown in its chapter-based structure

- Characterization of non-coded amino acids
- Nanotechnological applications of designed peptides
- Development of a data base of available knowledge of non-coded amino acids.

The main structural and functional role in peptidic based molecules is played by amino acids. Control over the conformational preferences of amino acids holds the key for modulating structure and function of peptides and proteins. In this context non-coded amino acids arise as very promising modulators due to the huge palette of available molecules and the plasticity inherent to their design.

The use of conformationally restricted non-coded amino acids that act as surrogates for the natural-occurring ones is a powerful tool to restrain the conformational variability of proteins. These restraints imposed by non-coded amino acids may retain a protein in its bio-active conformation thus enhancing its activity and, at the same time, introduce new features such as resistance to proteolysis, bolstered ligand affinity and new spectroscopic properties among others. Regarding the variety of non-coded amino acids, three main types are highlighted in the first chapter: those conformationally restricted and derived from *i) cyclic side chain amino acids*, *ii) proline analogues* and *iii) optically active amino acids*. The first two types induce severe conformational restrictions due to steric hindrance of the side chain (*i*) or the inclusion of the peptide backbone in the cyclic side chain (*ii*). These two types have been used to develop arginine surrogates ($c_5\text{Arg}$ and ProArg , respectively) that show a restricted conformational profile and can be used in peptide engineering, thus reducing the conformational variability of the peptide while retaining some key features of arginine. Out of type *ii*, indoline-carboxylic acid (Inc) and its alpha methylated (αMeInc) derivatives are an even more conformationally restricted

subgroup thanks to the benzene ring fused with the proline ring, converting them in feasible restrained surrogates for proline. Besides, these group of amino acids have interesting structural properties regarding the *cis/trans* isomerization of their peptide bond because, alike most of amino acids, it has a significant preference for the *cis* arrangement. Amino acids in group *iii* are of outstanding interest not only for their conformational properties but also for their optical ones; a good example from this family being phenylazophenylalanine (PAP). Optical and structural properties of PAP make of it a candidate to not only modulate the structural properties of proteins and peptides thanks to the photoisomerization but also to be used as a spectroscopic probe.

Some relevant uses of non-coded amino acids in engineered systems are exposed in the second chapter of the present thesis. These examples clearly illustrate the usefulness of these amino acids as structure/activity modulators. The arginine surrogate c_5 Arg shows to be an excellent structural modulator of mutated Cc_5 REKA peptide; the lower conformational variability of the non-coded amino acid, when compared to the natural one, is expressed in Cc_5 REKA as a structural bias close to the conformations described as bio-active in CREKA wild type. This effect of conformationally restricted non-coded amino acids of groups *i* and *ii* over the conformational landscape of peptides and proteins is caused by the steric hindrance of the cyclic side chain of the new amino acid, the backbone reduced flexibility if the cyclic side chain includes the backbone and physicochemical interactions between the side chains of the amino acids. An example of peptide engineering can be observed when a β -amyloid fragment is modified to feature thienylalanines instead of phenylalanines. These replacements are expected to not only alter the conformational landscape of the peptide but to promote electron conductivity. These replacements lead to new interactions between the aromatic moieties of the molecules, thus enhancing new intra and inter-catenary interactions, the latter ones at relevant high concentration. This peptidic system is expected to show self-aggregation, which opens new fields for peptide engineering based on non-coded amino acids.

The molecular hyper-crowding of a system, which can be understood as an effect of the concentration, is a keystone for understanding the conformational preferences of peptides: as concentration increases, so does the probability of finding both the same intra-chain interactions and new ones between non-linked chains. This major effect has been studied implicitly for CREKA homing peptide when it is covalently

linked to a surface at a significant peptide density; these conditions allow seeing the effects of intra and inter-catenary interactions over the conformational profile of the peptide.

As it has been demonstrated, non-coded amino acids are extremely important in the engineering of bio-inspired systems. Theoretical chemistry offers powerful tools for their study at high resolution level and the information generated by these means has been, is and will be published through articles in indexed scientific publications. The number and diversity of those journals makes it an endeavor to find the calculated conformational characteristics of those molecules. Besides, theoretical research in the field started decades ago and this implies not only a spread of articles between publications but also a widespread distribution in time. All the mentioned facts become troublesome questions for scientists both interested in doing new research and looking for available knowledge to be used in applications. Furthermore, the need for a systematized and ordered data base of theoretical structural information of non-coded amino acids is expected to grow even more in the upcoming years. The third chapter of this thesis reports the design, implementation and running of NCADB: a data base devoted to store theoretically-obtained conformational information of non-coded amino acids and, if available, experimentally-acquired information of the same molecules. Yet, NCADB not only offers the compilation of all this knowledge but it also relates it to existing bibliography, authors and practical uses of the filed compounds. The interface of NCADB is user-friendly in order to ease the search for non-expert users. This new available tool for scientists is intended to eventually cover all main types of non-coded amino acid. At the moment though, it allows searching amino acids whose side chains-have been modified and those with modified carbonyl or amine groups. The combination of a user-friendly display of information, customized searches and potential uses information renders a unique data base.

In summary, all the presented work offers a taste of theoretical approaches to be used in designing new non-coded amino acids, studying their properties and assessing their potential uses, such as engineering new bio-inspired systems throughout specific replacements with non-coded amino acids.

8. Conclusions

- (1) The conformationally restricted arginine surrogate *c₅Arg* shows a clear preference for α -helical backbone conformations due to side chain-main chain interactions involving the guanidium group. The helical conformation is the most energetically accessible, independently of the polarity of the solvent. This conformational trend is featured without regard of the chosen isomer for the side chain arrangement and suggests a strong tendency to be found at $i+1$ position in β -turns in proteins.
- (2) The proline-derived arginine analogue (β *Pro*)*Arg* adopts an α -helical conformation for its two side chain isomers in polar and apolar solvents, this preference is maintained in spite of considering thermal agitation. These facts suggest that (β *Pro*)*Arg* has new conformational properties not exactly matching those of proline that allow it to replace arginine in proteins when present at $i+1$ position in β -turns.
- (3) Proline derived *Inc* and *α MeInc* show a significantly reduced conformational profile in comparison to proline due to the benzene moiety fused to the pyrrol ring. The α -helical conformation is preferred in water for the two compounds whereas γ -turn is adopted in gas phase and chlorinated solvents for both of them, although α -helical is also accessible for *α MeInc* in the latter solvents as a product of the steric stress induced by the methyl group. A dramatic restriction of the flexibility of the pyrrol ring is also found: no clear envelope forms are identified, just the α carbon shows a little deviation from the pyrrol plane.
- (4) The energetics of *cis/trans* isomerism of the peptidic bond involving pyrrol nitrogen shows a significant variability depending of the chosen solvation model. Single point calculations and optimizations with a continuum model of solvent point to *cis* isomer as the preferred arrangement whereas *trans* configuration is the most energetically favored in the QM/MM. Inner structural tension, water entropic contribution and hydrogen bonding ability must be taken into account to explain these divergences.
- (5) The full conformational landscape of *PAP* has been investigated for the two isomers of the azobenzenic side chain: both *cis* and *trans* tend to adopt semi-extended backbone conformation as the lowest energy one in gas phase. Differences arise as α -helical arrangements become unreachable for *trans* but not for *cis* in gas phase, though they are energetically less favored. The solvation with a continuum model of water modifies energetic order for *trans* that shifts its

most favored conformation to γ_L , while for *cis* the global minimum remains a semi-extended conformation and helical conformations increase their stability. The energy difference between *cis* and *trans* isomers remains favorable to the latter throughout the energy landscape ranging from 7 to 17 kcal·mol⁻¹.

- (6) Two optical properties of *PAP* have been studied and compared to those for azobenzene: $n \rightarrow \pi^*$ (np) and $\pi \rightarrow \pi^*$ (pp) wavelength transitions are almost identical for azobenzene and *PAP*. Moreover, variations for those values upon *trans* to *cis* isomerization are in clear agreement with data from azobenzene, a bathochromic shift (4 nm) for np and hypsochromic shift (-3 nm) for pp. The variation of $\lambda^{pp} = \lambda^{pp}(\phi, \psi)$ and $\lambda^{np} = \lambda^{np}(\phi, \psi)$ is never above 10 nm. All these facts suggest that the incorporation of azobenzene to a peptide scaffold does not alter significantly its optical properties, even if the peptide backbone changes its conformation.
- (7) Non-coded amino acid *c₅Arg* is an optimal replacement for arginine in the homing peptide *CREKA* to become the engineered peptide *Cc₅REKA*. The conformational landscape of *Cc₅REKA* shows that the minimal energy conformations are biased towards those reported as bioactive for *CREKA*. The β -turn centered in second and third residues that is usually depicted by *CREKA* is stabilized by the conformational preferences of *c₅Arg*. This non-coded amino acid also allows the saline bridge interaction between the charged side chains of the central fragments amino acids and should introduce resistance to proteolysis.
- (8) High density of *CREKA* peptide linked to a surface restricts conformational variability of the peptide in comparison to infinite dilution. This trend is featured without regard of the chemical activity of the surface. Despite these restraints, the lowest energy conformations found in each case are similar among them, especially for the three amino acids of the central fragment. Intra-molecular interactions of *CREKA* might play a more significant role than the inter-molecular in the structural preferences of the peptide.
- (9) The dynamic conformational behavior of (2-Thi)(2-Thi)VLKAA has been evaluated to ascertain the effect of concentration (i.e. inter-molecular interactions) on the conformational behaviour of the peptide. The results of MD simulations show significant differences in the backbone conformations adopted by the five central amino acids of the peptide depending on the fulfilment of NMR-derived restraints. Those alterations are not translated into major changes

in the radius of gyration of the peptide but they cause peptide end-to-end distance to be far less changing in comparison to the unrestrained MD simulation. NMR-derived restraints favours major proximity between the side chains of the two β -thienylalanine units.

(10) In order to rationally gather key information for potential uses of non-coded amino acid in biomedical and material science, a data base (NCADB) has been designed including conformational information obtained at high theoretical level for non-coded amino acid. The data base displays a user-friendly interface that enables searches by conformation, side chain chemical nature, amino acid biological type, publications author, experimental characterization and applied available knowledge. NCADB includes non-coded amino acids from different types: alpha-tetrasubstituted alpha amino acids, diamines and diacids for retropeptides, N-substituted amino acids and thio-amino acids. The usefulness of the NCADB has been demonstrated through two examples of molecular engineered molecules.

