



UNIVERSITÁ DEGLI STUDI DI ROMA "SAPIENZA"

Tesi di Dottorato di Ricerca in Biofisica, XXII ciclo

**Molecular Dynamics, Quantum
Mechanics and Statistical Mechanics
study of two Gramicidin S analogs and
Syringomycin E**



Laura Zanetti Polzi

Supervisor:

Prof. Alfredo Di Nola

Dr. Andrea Amadei

June 2010

UNIVERSITÁ DEGLI STUDI DI ROMA "SAPIENZA"

Facoltá di Scienze Matematiche, Fisiche e Naturali

Corso di Dottorato in Biofisica

PhD Thesis

**Molecular Dynamics, Quantum Mechanics and
Statistical Mechanics study of two Gramicidin S
analogs and Syringomycin E**

Laura Zanetti Polzi

Supervisor:

Prof. Alfredo Di Nola

Dr. Andrea Amadei

June 2010

Front cover illustration: Julian Voss-Andreae, Green Fluorescent Protein, 2004,
Steel with process marks.

To my little girl...the best researcher I ever met

CONTENTS

1	Introduction	1
2	Methods for Molecular Simulations	7
2.1	Classical Molecular Dynamics	7
2.1.1	Force Field Models	8
2.1.2	Integration of Motion Equations	9
2.1.3	Long-range Interactions	10
2.1.4	The Boundary Conditions	12
2.1.5	Enhanced efficiency methods	12
2.1.6	Constant Temperature/Constant Pressure Molecular Dynamics	14
2.2	Essential Dynamics	15
2.3	Free energy calculations	17
2.3.1	Probability ratio method	18
2.3.2	Thermodynamic Integration	19
2.3.3	Perturbation method	20
2.3.4	Potential of Mean Force	20
2.4	The Perturbed Matrix Method (PMM)	21
2.4.1	Basic Derivations	22
2.4.2	Application of the PMM to the reconstruction of infrared spectra	24
3	The quasi-Gaussian entropy theory	27
3.1	The quasi-Gaussian entropy theory in canonical ensemble	28
3.1.1	Definition of the system	28
3.1.2	Definition of the reference states and excess properties	30
3.1.3	The potential energy distribution	31
3.1.4	The thermodynamic master equation	33
3.1.5	The conjugated pressure equation	34
3.1.6	Phase-space confinement	35
3.2	Description of different statistical states	36
3.2.1	Gaussian state	36
3.2.2	Gamma state	37

3.3	Extension of the QGE theory to threat conformational equilibria . . .	41
3.3.1	Conformational thermodynamics of peptides	41
4	Structural, Thermodynamic, and Kinetic Properties of Gramicidin Analogue GS6	47
4.1	Methods	48
4.2	Theory	48
4.2.1	Thermodynamic Characterization	48
4.2.2	Kinetic characterization	50
4.3	Results and Discussion	52
4.4	Conclusions	55
5	Reconstruction of infrared spectra by means of a QM/MD procedure	57
5.1	Methods	59
5.1.1	Unperturbed quantum chemical calculations	59
5.1.2	Molecular dynamics simulations	60
5.1.3	Theory	60
5.2	Results and Discussion	61
5.2.1	Infrared spectra of trans-NMA	61
5.2.2	Infrared spectra of β -hairpins	62
5.3	Conclusions	69
6	Analysis of IR Spectra of Gramicidin S analogs	71
6.1	Methods	72
6.2	Results and Discussion	74
6.3	Conclusions	79
7	Structure of the Lipodepsipeptide Syringomycin E	83
7.1	Methods	84
7.2	Results	86
7.3	Discussion	89
8	Concluding Remarks	93
	Bibliography	95
	Acknowledgments	118

This Thesis is based on the following publications:

1. L. Zanetti Polzi, M. Anselmi, M. D'Alessandro, A. Amadei and A. Di Nola
Structural, thermodynamic, and kinetic properties of Gramicidin analogue GS6 studied by molecular dynamics simulations and statistical mechanics
Biopolymers, 91(12), 1154-1160 (2009).
2. I. Daidone, M. Aschi, L. Zanetti Polzi, A. Di Nola and A. Amadei
On the origin of IR spectral changes upon unfolding
Chem. Phys. Lett., 488(4-6), 213-218 (2010).
3. L. Zanetti Polzi, I. Daidone, M. Anselmi, M. Aschi, G. Carchini, A. Di Nola and A. Amadei
Analysis of infrared spectra of β -hairpin peptides as derived from molecular dynamics simulations
Submitted (2010).
4. M. Anselmi and T. Eliseo and L. Zanetti Polzi, M. R. Fullone, V. Fogliano, A. Di Nola, M. Paci and I. Grgurina
Structure of the lipodepsipeptide syringomycin E in phospholipids and sodium dodecylsulphate micelle: a circular dichroism and NMR spectroscopy combined with molecular dynamics study in octane
Submitted (2010).

INTRODUCTION

For half a century, the three-dimensional structures of proteins were studied indirectly. Details of the protein activity mechanism and its relationship to structure could only be inferred from the properties of the protein. In 1958 a landmark paper¹ on the structure of myoglobin by John Kendrew and co-workers showed the first 3D structure of a protein. Its resolution was too low to show the atomic details of what seemed to be a multiply bent sausage, and the crude map gave only a hint of the revolution to come. The subsequent low-resolution structure of hemoglobin, which resembled four myoglobin molecules stacked together, was reported by Max Perutz and colleagues in 1960² and gave the first intimation of protein families.

Solving the structures of proteins at high resolution uncovered a new problem and initiated a novel field of research: that of protein folding. The “protein folding problem”, that is the question of how a protein’s amino acid sequence dictates its three-dimensional atomic structure, is regarded as three different problems: (a) the folding code: the thermodynamic question of what balance of interatomic forces dictates the structure of the protein, for a given amino acid sequence; (b) protein structure prediction: the computational problem of how to predict a protein’s native structure from its amino acid sequence; and (c) the folding process: the kinetics question of what routes or pathways some proteins use to fold so quickly.

A major milestone in protein science was the thermodynamic hypothesis of Christian Anfinsen and colleagues.^{3,4} From his now-famous experiments on ribonuclease, Anfinsen postulated that the native structure of a protein is the thermodynamically stable structure; it depends only on the amino acid sequence and on the conditions of solution, and not on the kinetic folding route.

Cyrus Levinthal and others, such as Michael Levitt and Oleg Ptitsyn, wanted to predict tertiary structures by predicting their folding pathways. Levinthal famously pointed out that it seemed impossible that an unfolded protein could fold spontaneously by a random process on a biological time scale. Mechanisms were proposed that could overcome the “Levinthal paradox” by simplifying the folding process and

breaking it down into subprocesses that could occur stepwise.

The search for folding mechanisms has driven major advances in experimental protein science. These include Fluorescence Resonance Energy Transfer (FRET), Small-angle X-ray Scattering (SAXS), circular dichroism and real-time NMR;^{5,6} mutational methods that give quantities called ϕ -values⁷⁻⁹ (now also used for ion-channel kinetics and other rate processes¹⁰) or ψ -values,¹¹ which can identify those residues most important for folding speed; hydrogen exchange methods that give monomer-level information about folding events;^{12,13} and the extensive exploration of protein model systems, including cytochrome c, CI2, barnase, apomyoglobin, the src, α -spectrin, and fyn SH3 domains, proteins L and G, WW domains, trpzip, and trp cage.¹⁴ In addition, peptide model experimental test systems provide insights into the fast early-folding events.¹⁵⁻¹⁷ Among the other experimental techniques, fast laser temperature-jump methods¹⁸ combined to time-resolved infrared (TRIR) spectroscopy^{19,20} have been recently developed and widely used to follow unfolding or refolding transitions in peptides. This method, in spite of its good time resolution, has the limitation that the band shapes for different secondary structures may be similar to one another and highly congested,²¹ thus leading to spectra which are difficult to interpret. Furthermore, single-molecule methods are beginning to explore the conformational heterogeneity of folding.²²⁻²⁵

However, modern mechanistic studies have to achieve spatial atomic-resolution information and so it was necessary to develop methods that could analyze folding at the level of individual residues and atoms. Experimental X-ray or NMR structures are typically the starting point for studies of the relationship between protein structure at the atomic resolution and function. Such structures, however, do not generally describe the dynamical properties that may be as important to protein function and regulation as the detailed positioning of backbone and side-chain atoms.

Molecular Dynamics (MD) simulations provide a unique ability to describe directly the dynamical aspects of protein structure by tracking the time-dependent positions of all atoms in the system. The first attempts at simulation were severely limited by computational power and by approximations in energy functions, which are still imperfect today. Initially disparaged, these methods are now indispensable for understanding the mechanisms of folding of proteins as well as their intricate details because atomistic simulation, benchmarked by experimentation, is the only way of analyzing a complete folding pathway and calculating the folding energetics. The first all-atom MD simulation of a small protein in vacuo, performed more than three decades ago, covered less than 10 ps.²⁶ Over the years, improvements in molecular dynamics algorithms, software, and computer hardware have allowed MD simulations to access longer timescales. A decade ago, Duan and Kollman performed the first microsecond-length all-atom MD simulation of a small protein.²⁷ This achievement required a Herculean effort, involving four months of supercom-

puter time despite the small size of the simulated system (slightly less than 10000 atoms) and the use of many approximations. In the last few years, the generation of accurate microsecond-timescale simulations of systems with tens of thousands of atoms has become more practical, and a number of studies with individual trajectories longer than one microsecond have been reported.²⁸⁻³⁵

Despite the huge advances in computational power, it is still not possible, in general, to predict the structures of proteins *de novo*. The pathway of protein unfolding can be calculated by molecular dynamics simulation³⁶ from the known 3D structure. Folding simulations are more difficult, but have been aided by the discovery of ultra-fast-folding proteins, which fold a million times faster than those that prompted the Levinthal paradox 40 years ago: these proteins fold within a few microseconds on the time scale that is accessible to full atomistic simulation. Such pathways are benchmarked by experimentation, with ϕ -value analysis of transition states and NMR spectroscopic structural determination of intermediates and analysis of denatured states. Simulations using ϕ values as constraints can also be used to construct transition states.³⁷ The pathway of folding and unfolding of the Engrailed homeodomain from *Drosophila melanogaster*, for example, has been solved at atomic resolution using this whole gamut of techniques.^{38,39} In addition, structurally homologous proteins of vastly different amino-acid sequence can be identified from databases to study the change of folding mechanism with structure and derive more general principles about the mechanism.⁴⁰ Several protein folding pathways are now known in detail at atomic resolution⁴¹ thanks to the application of a combination of technologies, all of which initially met with some skepticism.

A key issue has been whether semiempirical atomic physical force fields are good enough to fold up a protein in a computer. Physics-based methods are currently limited by large computational requirements owing to the formidable conformational search problem and, to a lesser extent, by weaknesses in force fields. Nevertheless, there have been notable successes in the past decade enabled by the development of large supercomputer resources and distributed computing systems. The first milestone was a supercomputer simulation by Duan and Kollman in 1998 of the folding of the 36-residue villin headpiece in explicit solvent, for nearly a microsecond of computed time, reaching a collapsed state 4.5 Å from the NMR structure.²⁷ In addition, three groups have folded the 20-residue Trp-cage peptide to ~ 1 Å.⁴²⁻⁴⁴ More recently, Lei and Duan⁴⁵ folded the albumin-binding domain, a 47-residue, three-helix bundle, to 2.0 Å. Physics-based approaches are also folding small helices and β -hairpin peptides of up to ~ 20 residues that have stable secondary structures.⁴⁶⁻⁵⁰

Molecular dynamics (MD) simulation is nowadays a well-established method for modeling the dynamics of proteins, the characterization of which provides insight into the workings of biomolecular systems at spatial and temporal scales that are difficult to access experimentally. Moreover, a recent confluence of progress in both simulation and laboratory techniques has presented more opportunities to compare

computational results to experimental data gathered on similar timescales. These developments allow not only for a better understanding of the biomolecular systems under study, but also for systematic validation of the models and methods underlying MD simulations,⁵¹ promising to increase the utility of MD simulations in the study of proteins and other biomolecular systems.

Thanks to the combined use of experimental and theoretical methods, the knowledge of the folding process has considerably evolved during the last decades. Prior to the mid-1980s, the protein folding code was seen a sum of many different small interactions, such as hydrogen bonds, ion pairs, van der Waals attractions, and water-mediated hydrophobic interactions. A key idea was that the primary sequence encoded secondary structures, which then encoded tertiary structures.⁵² However, through statistical mechanical modeling, a different view emerged in the 1980s, namely, that there is a dominant component to the folding code, that it is the hydrophobic interaction, that the folding code is distributed both locally and non-locally in the sequence, and that a protein's secondary structure is as much a consequence of the tertiary structure as a cause of it.^{53,54} There is considerable evidence that hydrophobic interactions must play a major role in protein folding. (a) Proteins have hydrophobic cores, implying nonpolar amino acids are driven to be sequestered from water. (b) Model compound studies show 1-2 kcal/mol for transferring a hydrophobic side chain from water into oil-like media.⁵⁵ (c) Proteins are readily denatured in nonpolar solvents. (d) Sequences that are jumbled and retain only their correct hydrophobic and polar patterning fold to their expected native states,⁵⁶⁻⁵⁹ in the absence of efforts to design packing, charges, or hydrogen bonding. In this scenario, the driving force of folding is the minimization of the nonpolar surface that is exposed to water, while simultaneously providing hydrogen-bonding interactions for buried backbone groups, usually in the form of secondary structures such as α -helices, β -sheets, and reverse turns.⁶⁰

Reverse turns constitute one of the most common structural features in globular proteins.⁶¹ Despite their prevalence, they have always been more challenging to categorize than α -helices or β -sheets because of their nonperiodic nature and the heterogeneity of their structures. They are broadly defined as those regions of the polypeptide where a change of chain direction occurs.⁶² Because chain reversals allow a protein to fold onto itself, forming a compact globular state, they may play a key role in protein folding. Turns have been proposed to be important in folding also because they are capable of initiating productive structure formation without a large loss in chain entropy since the interactions involved in turn formation are largely local.⁶³⁻⁶⁷ This view is consistent with a hierarchical folding model in which certain turns containing residues with high turn propensities (such as glycine and aspartic acid) serve as active nucleation sites for structure formation, originating from the corner residues and propagating toward the flanking β -strands. The turn as the site for chain reversal becomes a nucleation point that facilitates cooperative formation

of neighboring interactions

Reverse turns comprise the widely distributed β -turns, as well as the less prevalent γ -turns and α -turns, and may include well-defined loops, such as ω -loops. The short, self-contained nature of reverse turns has led to their extensive study using peptide models. In an approach pioneered by the Blout laboratory, cyclic peptides have served as constrained models for reverse turns.⁶⁸ Numerous crystal structures,^{69–73} as well as NMR studies of cyclic peptides in solution,^{74–78,78–81} confirmed the original characterization of β -turn types and provided illustrations of γ -turns and α -turns. The similarity of these local structures in cyclic peptides to those observed in proteins gave early credence to the importance of local energetic contributions to turn conformations. Because of the conformationally constrained nature of cyclic peptides, they have also been useful in defining spectroscopic signatures of turns.^{73,74,82,83}

Over the past two decades, considerable research has been focused on the next stage in a hierarchical build-up of structure from a β -turn: β -hairpins. These structural features are stabilized by virtue of turn propensities of amino acid residues as well cross-strand interactions between the sequences flanking the turns. Favorable energetic contributions to β -hairpin stability have been demonstrated to include cross-strand aromatic-aromatic, aromatic-polar, hydrogen-bonding, hydrophobic, and salt-bridge interactions,^{84–88} in combination with loop conformational propensity and entropy terms dependent on loop length.^{87,89}

In this thesis, three cyclic peptides are studied, all of them having a biological relevance, as they show a broad spectrum antibiotic activity. Moreover, they all show a hemolytic activity which limits their usage as drugs: it is thus of fundamental importance to clarify the relationship between their structure and activity in order to design novel peptides with the same antibiotic activity but without any hemolytic effect.

Two of the three peptides studied here are β -hairpins and all show hydrogen-bonded turns. These peptides, for which experimental data are available, are used here as model systems to analyze the structure and the dynamics of formation of turns and in particular of β -hairpins. The folding/unfolding transitions were observed and some experimental data were excellently reproduced, thus validating the models used but also helping the understanding of the experimental data.

In chapter 4 a complete thermodynamic and kinetic characterization has been achieved for the folding/unfolding transition of a small, fast-folding cyclic β -hairpin peptide, a synthetic analog of the well-known antibiotic Gramicidin S. In chapters 5 and 6 a new theoretical methodology to reconstruct experimental infrared spectra is presented and applied to two synthetic analogs of Gramicidin S, shedding light on the interpretation of vibrational spectra. In chapter 7 a combined NMR and MD study is presented of the cyclic peptide Syringomycin E, in order to clarify its structure in the membrane environment.

METHODS FOR MOLECULAR SIMULATIONS

In this chapter some basic concepts and methodologies of molecular simulations are introduced with a particular attention devoted to the methods relevant to this thesis. Several books on these subjects can be found with a deeper insight into these problems.^{90–92} As the method used in the present thesis to study the properties of macromolecules in solution is classical Molecular Dynamics (MD), a very brief description of its basic principles is presented in section 2.1. In sections 2.2 and 2.3 some methods to obtain relevant physical properties of the studied system are briefly presented.

The classical approach has the disadvantage to not include electrons explicitly in the calculations and hence molecular properties strictly related to electronic degrees of freedom cannot be investigated. For these purposes, quantum mechanical techniques should be used in order to properly evaluate the internuclear forces, but at a very high computational cost. A compromise can be reached by using first-principle methods for a small reactive region of the whole system and force field methods for the remaining part. This approach allows to study electronic properties of systems, fully including their dynamics. One of these hybrid methods, the Perturbed Matrix Method (PMM), developed in our laboratories, will be introduced in section 2.4

2.1 Classical Molecular Dynamics

Aim of computer simulations of molecular systems is to compute macroscopic behavior from microscopic interactions. A model of the real world is constructed, both measurable and unmeasurable properties are computed and the former are compared with experimentally determined properties. If the model used is validated by the comparison, it could even be used to predict unknown or unmeasurable quantities. A theoretical treatment of the motions and interactions of molecules should be founded, rigorously speaking, on quantum mechanics principles, due to the microscopic nature of these objects. Unfortunately, first-principle approaches are often

unpractical because they require very large computational facilities and they are definitely prohibitive for systems containing thousands of atoms. Hence, a certain level of approximation becomes necessary and it should be chosen in such a way that those degrees of freedom that are essential to a proper evaluation of the quantity or property of interest can be sufficiently sampled. When excluding chemical reactions, low temperatures or details of hydrogen atoms motion, it is relatively safe to assume that the system is governed by the laws of classical mechanics.

In classical MD, a trajectory (configurations as a function of time) of the molecular system is generated by simultaneous integration of Newton's equations of motions for all atoms in the system:

$$\frac{d^2 r_i}{dt^2} = m_i^{-1} F_i \quad (2.1)$$

$$F_i = - \frac{\partial V(r_1, \dots, r_N)}{\partial r_i} \quad (2.2)$$

The force acting on atom i is denoted by F_i , the mass by m_i and time is denoted by t . MD simulations require calculation of the gradient of the potential energy $V(r_1, \dots, r_N)$, which therefore must be a differentiable function of the atomic coordinates r_i . This potential energy function, or *force field*, is called an *effective interaction* function since the average effect of the omitted (electronic) degrees of freedom has been incorporated in the interaction between the (atomic) degrees of freedom explicitly present in the model.

The choice of molecular model and force field is essential to a proper prediction of the properties of a system. Therefore, it is of great importance to be aware of the fundamental assumptions, simplifications and approximations that are implicit in the various types of models used in the literature.

2.1.1 Force Field Models

A huge variety of force fields is currently used in the molecular dynamics community, sometimes differing for minor changes, e.g. CHARMM,⁹³ AMBER,⁹⁴ GROMOS.⁹⁵ A typical molecular force field, or effective potential, for a system of N atoms with masses m_i ($i = 1, 2, \dots, N$) and Cartesian position vectors r_i has the following form:

$$\begin{aligned} V(r_1, r_2, \dots, r_N) &= \sum_{bonds} \frac{1}{2} K_b (b - b_{eq})^2 + \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_{eq})^2 \\ &+ \sum_{dihedrals} K_\phi [1 + \cos(n\phi - \delta)] + \sum_{imp.dihedrals} \frac{1}{2} K_\xi (\xi - \xi_{eq})^2 \\ &+ \sum_{pairs} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{pairs} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \end{aligned} \quad (2.3)$$

The first term represents the covalent bond stretching interaction between two atoms linked by a harmonic potential where b_{eq} is the minimum energy bond length and K_b is the force constant changing with the particular bond type. The second term is a three-body interaction corresponding to the valence angle, θ , deformation expressed as a harmonic potential where θ_{eq} is the equilibrium valence angle and K_θ the force constant. The third and fourth terms are used for the (four-body) dihedral angle interactions: a harmonic term for improper dihedral angles, ξ , that are not allowed to make transitions, i.e. to keep the aromatic rings planar, and a sinusoidal term for all the other dihedral angles, ϕ . The last two terms are sums over the pairs of non-bonded atoms and represent the effective non-bonded interactions expressed in terms of van der Waals and Coulombic interactions between atoms i and j at a distance r_{ij} . The parameters ϵ_{ij} and σ_{ij} are the typical constants defining the Lennard-Jones potential, q_i and q_j are the atom charges and ϵ_0 is the dielectric constant in vacuum.

The parameters used in the force field (Eq. 2.3) can be determined in different ways. Generally two main approaches are followed. The first one is to fit them with results obtained from *ab initio* calculations on small molecular clusters. The alternative way is to fit the force field parameters to experimental data, like crystal structures, energy and lattice dynamics, infrared or X-ray data on small molecules, liquid properties like density and enthalpy of vaporization, free energy of solvation, nuclear magnetic resonance data, etc. Whatever method is used, the resulting model is far to be universal. It is worth to note that every force field is usually well suited for specific general conditions, i.e. particular thermodynamic conditions (temperature, density, pressure, etc.) and also boundary conditions. Moreover, they are optimized for specific classes of molecules, such as inorganic molecules, organic molecules, biomolecules (DNA, proteins, lipids), etc.

2.1.2 Integration of Motion Equations

Newton's equations of motion, a second-order differential equation, can be written as two first-order differential equations for the particle positions $r_i(t)$ and velocities $v_i(t)$ respectively:

$$\frac{dv_i(t)}{dt} = m_i^{-1}F_i \quad (2.4)$$

$$\frac{dr_i(t)}{dt} = v_i(t) \quad (2.5)$$

A standard method for solution of the previous ordinary differential equations is the finite difference approach. The general idea is the following. Given the molecular positions, velocities and forces at time t , we attempt to obtain the positions, velocities and forces at a later time $t + \delta t$, to a sufficient degree of accuracy. The

equations are solved on a step-by-step basis; the choice of the time interval δt will depend somewhat on the method of solution, but δt will be significantly smaller than the typical time taken for a molecule to travel its own length.

Many different algorithms fall into the general finite difference pattern, like Verlet, and its computationally efficient variant *leap-frog*,^{96,97} Beeman⁹⁸ or the Gear predictor-corrector.⁹⁹

2.1.3 Long-range Interactions

One of the most challenging problems in molecular dynamics simulations is the treatment of long-range interactions, which usually correspond to the electrostatic forces between molecules. To reduce the computational cost, the size of the simulated system is generally small and, as a consequence, a correct evaluation of the intermolecular interactions is not trivial. Many different methods were developed to reproduce reasonably the thermodynamics of bulk liquids. Here we consider two of the most used techniques: the use of a cut-off radius and the Ewald sum.

The *cut-off method* is based on the truncation of the forces when the distance between the interacting particles is greater than a specified value, called the cut-off radius, r_c . In this way, the only interactions felt by the i -th particle are those due to the particles contained in a sphere of radius r_c and centered at r_i . This method is doable only if the intermolecular forces decay rapidly with the distance. In fact, when the forces are negligible at a distance $\geq r_c$, the main structural and dynamical properties are correctly reproduced. Otherwise deviations from the correct bulk behavior are expected.

Another methodology in MD simulations is the use of a *periodic lattice method* in which all the interactions between the molecular system in the central cubic cell and its virtual replica are included. The Coulomb interaction energy in a periodic system of N charged particles is obtained by a sum over all pairs of which one atom lies in the central box and the other is its periodic image:

$$E = \frac{1}{8\pi\epsilon_0} \sum_{|n|=0}^{\infty} \left(\sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|r_{ij} + n|} \right) \quad (2.6)$$

The sum over n is a summation over all simple cubic lattice points $n = (n_x L, n_y L, n_z L)$, with L the side length of the cubic cell and n_x, n_y, n_z integers. The case $i = j$ is omitted for $n = 0$. It was shown that the sum over n for such kind of potential (r^{-1}) is only conditionally convergent, then its limit may vary or even diverge if the order of terms in the sum is changed. A solution to this problem was developed following a physical idea:⁹⁰ each point charge is surrounded by a charge distribution of equal magnitude and opposite sign, which spreads out radially from the charge, $\rho^G(r)$. This distribution has the effect to screen the interactions between the neighboring point charges and hence the interaction energy becomes

short-ranged. Commonly, the screening charges have a Gaussian form. The total charge distribution is given by:

$$\rho_i(r) = \rho_i^q(r) + \rho_i^G(r) \quad (2.7)$$

where $\rho_i^q(r)$ is the distribution of the point charge of the i -th particle and $\rho_i^G(r)$ is the corresponding Gaussian distribution.

First, the interaction energy due to the distribution 2.7 is calculated in the real space, then, in order to recover the original charge distribution, a canceling function is added in the reciprocal space, which is equal to $-\rho_i^G(r)$, realized by means of a Fourier transform. Hence the final form of the total interaction energy is given by:

$$\begin{aligned} E = & \frac{1}{8\pi\epsilon_0} \sum_{i=1}^N \sum_{j=1}^N \left(\sum_{|n|=0}^{\infty} \frac{q_i q_j \operatorname{erfc}(\alpha|r_{ij} + n|)}{|r_{ij} + n|} \right. \\ & + \left. \frac{1}{\pi L^3} \sum_{k \neq 0} \frac{4\pi^2 q_i q_j}{k^2} \exp(-k^2/4\alpha^2) \cos(k \cdot r_{ij}) \right) \\ & - \frac{\alpha}{4\pi^{3/2}\epsilon_0} \sum_{i=1}^N q_i^2 + \frac{|\sum_{i=1}^N q_i r_i|^2}{2\epsilon_0 L^3 (2\epsilon' + 1)} \end{aligned} \quad (2.8)$$

Here $\operatorname{erfc}(x)$ is the complementary error function, which falls to zero with increasing its argument. Thus, if the parameter α is large enough, the sum over n in the first term reduces to the only term $n = 0$. The second term is a sum over the reciprocal vectors $k = 2\pi n/L$. Again, if α is large, a lot of terms in the k -space sum are needed to get a convergence of the energy. The last two terms are, respectively, a correction function, due to the fact that a self-interaction of the canceling distribution is included in the recipe, and the energy contribution of the depolarizing field, which is compensated by the effect of the external dielectrics. Note that in the Ewald sum the virtual cubic cells are ordered as concentric spherical layers starting from the central box. Clearly the infinite sum is truncated at a certain point and the resulting spherical system is immersed in a continuum dielectrics with dielectric constant ϵ' . The last term in equation 2.8 is the sum of the contributions of the depolarizing field and the reaction field due to the external dielectrics. If the sphere is embedded in a medium with an infinite dielectric constant, this term vanishes because of a perfect compensation of the two effects.

Other *periodic lattice methods* are often used in computer simulations for their computational stability and efficiency. These methods, like the Particle Mesh Ewald (PME)¹⁰⁰ method, can be considered of the same family of the method shown here.

2.1.4 The Boundary Conditions

An important characteristic of the molecular dynamics simulations is the way in which the boundaries are treated. Due to computational limits, a typical simulated system contains $10^4 - 10^5$ atoms, and then is quite small compared to macroscopic matter. This means that, if the molecules are arranged in a cubic box, a relatively great part of them will lie on the surface and will experience quite different forces from molecules in the bulk. The consequence of the finite size of the system is that the boundary conditions may affect seriously the results of the simulations, especially when the system of interest is a homogeneous liquid or a solution. Usually, periodic boundary conditions (PBC)⁹⁰ are adopted to reduce the surface effects. This technique consists on simulating the system in a central cubic box surrounded by an infinite number of copies of itself. During the simulation, the molecules in the original box and their periodic images move exactly in the same way. Hence, when a molecule leaves the central box one of its images will enter through the opposite side. As a result, there are no physical boundaries neither surface molecules. Note that other shapes of the box can be used as the truncated octahedron or the rhombic dodecahedron.

2.1.5 Enhanced efficiency methods

A clear gap exists between time scales that can be obtained by computer simulation techniques applied to biological macromolecules and the times required for most biological processes. Therefore, several techniques have been developed to increase the time scales of MD simulations.

When Newton's equations of motion are integrated, the limiting factor that determines the time step that can be taken is the highest frequency that occurs in the system. In solvated biological macromolecules, the vibrations of bonds involving hydrogen atoms form the highest frequency vibrations. The bond stretching frequency of an O-H bond is typically about 10^4 Hz, so the average period would be of the order of 10 fs.¹⁰¹ This limits the time-step to be taken in MD simulations to about 0.5 fs (a rule of thumb exists that states that for a reasonable sampling of a periodic function, samples should be taken at least twenty times per period). The introduction of a method to constrain these bonds (or, in fact, all covalent bonds) allowed to increase the time step to a typical value of 2 fs. The most widely used constraint algorithm for large molecules is *SHAKE*¹⁰² and, more recently, *LINCS*¹⁰³. Since these bond vibrations are practically uncoupled from all other vibrations in the system, constraining them does not notably alter the rest of the dynamics of the system. This is not true, however, for bond-angle fluctuations, which form the second-highest frequency vibrations. Constraining bond-angles has a severe effect on many other fluctuations in the system, including even global, collective fluctuations, limiting the use of methods that use bond-angle constraints to only a few

specific cases.¹⁰¹

The notion that a number of discrete classes of frequencies of fluctuations in simulations of biomolecules can be distinguished, however, can be utilized to design more efficient algorithms. Forces that fluctuate rapidly need to be recalculated at a higher frequency than those that fluctuate on a much longer time scale. Although not trivial to implement, a number of successful applications of so-called *multiple time-step* algorithms have been reported in the literature (for a review, see Schlick *et al.*¹⁰¹). Speed up factors of 4-5 have been claimed for such methods with respect to unconstrained dynamics, making them only slightly more efficient than simulations with covalent bond-length constraints.

Dummy atoms

For the remaining degrees of freedom, the shortest oscillation period as measured from a simulation is that of bond-angle vibrations involving hydrogen atoms. The obvious solution would be to constrain all bond angles involving hydrogen atoms in all molecules, in addition to all bond lengths. With the constraint algorithms this can already be done, but SHAKE tends to break down with time steps beyond 2 fs, whilst LINCS cannot handle the highly connected constraints that arise from constraining both bonds and angles. Thus, the most elegant solution would be to eliminate these high-frequency degrees of freedom from the system altogether.

For hydrogen atoms in large molecules (e.g. proteins), this can be implemented in a rather straightforward manner. Instead of connecting a hydrogen atom with bonds, angles, and dihedrals to the molecule, the position of the hydrogen will be generated every MD step based on the position of three nearby heavy atoms. All forces acting on the hydrogen atom will be redistributed over these heavy atoms and, to keep the total mass in the system constant, the mass of each hydrogen atom that is treated in this way should be added to the bonded heavy atom. A particle treated in this manner is referred to as a *dummy atom*¹⁰⁴. Introducing *dummy atoms*, a net gain in simulation efficiency of a factor of 3 to 3.5 can be achieved.

Roto-translational constraint

Another approach to reach equilibrium conformational properties at an enhanced rate is by constraining the rotational and translational motions in polyatomic systems.¹⁰⁵ This method is generally used to study biomolecules in solution. In such a system, the internal motions of the solute are often more interesting than its rotational and translational motions. This algorithm is implemented in a *leap-frog* integration scheme coupled with SHAKE. The use of the *roto-translational constraint* presents several advantages, like a reduction of the molecular relaxation time and the possibility of reducing the amount of solvent molecules to be used.¹⁰⁶

2.1.6 Constant Temperature/Constant Pressure Molecular Dynamics

When Newton's equations of motion are integrated the total energy is conserved (adiabatic system) and if the volume is held constant the simulation will generate a microcanonical ensemble (NVE). However, this is not always very convenient. Other statistical ensembles, such as canonical (NVT) and isothermal-isobaric (NPT) ensembles, better represent the conditions under which experiments are performed than the standard microcanonical ensemble. Moreover, with the automatic control of temperature and/or pressure, slow temperature drifts that are an unavoidable result of force truncation errors are corrected and also rapid transitions to new desired conditions of temperature and pressure are more easily accomplished.

Several methods for performing MD at constant temperature have been proposed, ranging from *ad hoc* rescaling of atomic velocities in order to adjust the temperature, to consistent formulation in terms of modified equations of motion that force the dynamics to follow the desired temperature constraint. The three most utilized methods are described next.

The *thermal bath coupling* method, or Berendsen bath,¹⁰⁷ has the great advantage of being simple. This algorithm simulates a coupling of the system with an external thermal bath at the temperature T_0 and the interaction between this bath and the system is modulated by a time constant τ . The coupling is obtained multiplying for a constant λ the velocities. The temperature T is scaled to the reference temperature T_0 via an exponential law.

The *isothermal*, or isogaussian, method¹⁰⁸ allows to fix the temperature exactly constant. Using this algorithm, a variable is added to the motion equations, acting as a friction coefficient changing in time in order to keep the kinetic energy constant. This method correctly generates the configurational properties of the canonical ensemble, while the momenta distribution is not canonical.¹⁰⁵

Nosé-Hoover thermostat is based on the use of an extended Lagrangian, i.e. a Lagrangian that contains additional, artificial coordinates and velocities.^{109,110} The conventional Nosé-Hoover algorithm only generates the correct distribution if there is a single constant of motion. Normally, the total energy, that includes the artificial variables, is always conserved. This implies that one should not have any other conserved quantity. If we have more than one conservation law, we have to use the Nosé-Hoover chains to obtain correct canonical distribution.¹¹¹

The various methods for carrying out MD at constant pressure are based on the same principles as the constant temperature scheme with the role of the temperature played by the pressure and the role of the atomic velocities played by the atomic positions.

2.2 Essential Dynamics

The Essential Dynamics (ED) analysis is a method to seek those collective degrees of freedom that best approximate the total amount of fluctuation of a dynamical system.^{112,113} A brief description will be given here. ED is based on a principal component analysis (PCA) of (MD generated) structures. A PCA is a multidimensional linear least squares fit procedure. To understand how this is applicable to protein dynamics, the usual three-dimensional (3D) Cartesian space to represent protein coordinates (which is e.g. used to represent protein conformations in the Brookhaven Protein Data Bank or PDB) needs to be replaced by another, multidimensional space. A molecule of N particles can be represented by N points in 3D space. With 3 coordinates per point, this adds up to $3N$ coordinates. In a $3N$ -dimensional space, however, such a structure can be represented by a single point. In this space, this point is characterized by $3N$ coordinates. This representation is convenient since a collection or trajectory of structures can now be regarded as a cloud of points. Like in the case of a two-dimensional cloud of points, also in more dimensions, always one line exists that best fits all points. As illustrated for a two-dimensional example (Figure 2.1), if such a line fits the data well, the data can be approximated by only the position along that line, neglecting the position in the other direction. If this line is chosen as coordinate axis, then the position of a point can be represented by a single coordinate. In more dimensions the procedure works similarly, with the only difference that one is not just interested in the line that fits the data best, but also in the line that fits the data second-best, third best, and so on (the principal components). These directions together span a plane, or space, and the subspace responsible for the majority of the fluctuations has been referred to as the 'essential subspace'. Applications of such a multidimensional fit procedure on protein configurations from MD simulations of several proteins has proven that typically the ten to twenty principal components are responsible for 90% of the fluctuations of a protein.¹¹²⁻¹¹⁴ These principal components correspond to collective coordinates, containing contributions from every atom of the (protein) molecule. Summarized, a limited number of collective motions is responsible for a large percentage of a protein conformational fluctuations.

If all atoms in a protein were able to move uncorrelated from each other, an approximation of the total fluctuation by only a few collective coordinates would not be possible. The fact that such an approximation is successful is the result of the presence of a large number of internal constraints and restrictions ('near-constraints') defined by the interactions present in a given protein structure. Atomic interactions, ranging from covalent bonds (the tightest interactions) to weak non-bonded interactions, together with the dense packing of atoms in native-state protein structures form the basis of these restrictions.

In the study of protein dynamics, only internal fluctuations are usually of interest.

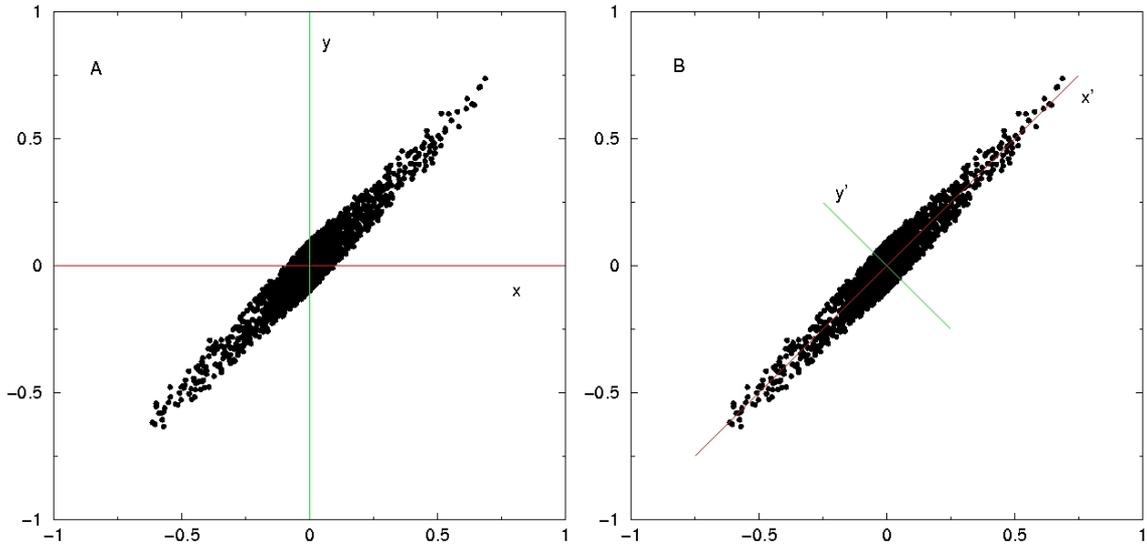


Figure 2.1: Example of Essential Dynamics in two dimensions. With a distribution of points as depicted here, two coordinates (x,y) are required to identify a point in the cluster in panel A, whereas one coordinate (x') approximately identifies a point in panel B

Therefore, the first step in an Essential Dynamics analysis is to remove overall rotation and translation. This is done by translation of the center of mass of every configuration to the origin after which a least squares rotational fit of the atoms is performed onto to a reference structure. The actual principal component analysis is based on construction and diagonalization of the covariance matrix of positional fluctuations. Defining the $3N$ dimension column vector $X(t)$ representing the atomic coordinates of the system at time t , the covariance matrix is built up according to:

$$C = \langle \Delta X \Delta X^T \rangle \quad (2.9)$$

where $\Delta X = X(t) - \langle X \rangle$ and the angle brackets represent a time or ensemble average. Particles moving in a correlated fashion correspond to positive matrix elements (positive correlation) or negative elements (negative correlation) and those that move independently to small matrix elements. The orthogonal transformation T that diagonalizes this (symmetric) matrix contains the eigenvectors or principal components of C as columns and the resulting diagonal matrix Λ contains the corresponding eigenvalues:

$$\Lambda = T^T C T \quad (2.10)$$

The eigenvalues are the positional mean square fluctuations along the corresponding eigenvectors. When the eigenvectors are sorted to decreasing eigenvalues,

the first eigenvectors are those collective motions that best approximate the sum of fluctuations and the last eigenvectors correspond to the most constrained degrees of freedom. The characteristics of these collective fluctuations can be studied by projecting the ensemble of structures onto single eigenvectors and by translation of these projections to 3D space to visualize the atomic displacements connected with that eigenvector. As stated above, analysis of MD trajectories of several proteins have shown that few collective coordinates dominate the dynamics of native proteins (together often referred to as the 'essential subspace'). In a number of cases these main modes of collective fluctuation were shown to be involved in the functional dynamics of the studied proteins.^{112,115,116}

ED analysis can be applied to any subset of atoms of the ensemble of structures and are not restricted to ensembles generated by MD simulation. Applications to collections of X-ray structures,^{116,117} NMR structures¹¹⁸ and structures derived from distance constraints¹¹⁹ have been reported. Since collective (backbone) fluctuations dominate the dynamics of proteins, usually only backbone or C_α coordinates are used to save computation time and to prevent problems with apparent correlation of side chain motions with backbone motions which are merely the result of poor statistics. However, even when the method is applied to only C_α atoms, the diagonalization of the covariance matrix can still be an enormous computational task.

2.3 Free energy calculations

In general terms, a microscopic description of a particular molecular system can be given in the form of a Hamilton operator or function. This is often simply expressed as the Hamiltonian $H(p, q)$ of the generalized coordinates q and their conjugate momenta p . For example, the Hamiltonian for a classical system of N atoms, expressed in terms of the Cartesian coordinates r and momenta p of each of the atoms, has the form $H(p, r) = K(p) + U(r)$, where K is the kinetic and U the potential energy. In the canonical ensemble the fundamental formula for the Helmholtz free energy, A , is:¹²⁰

$$A(N, V, T) = -k_B T \ln Q(N, V, T) \quad (2.11)$$

where the partition function Q is:

$$Q(N, V, T) = h^{-3N} \int \int e^{-H(p,r)/k_B T} dp dr \quad (2.12)$$

where V is the volume of the system, T the absolute temperature, k_B Boltzmann's constant, h Planck's constant, and it is assumed that the N atoms are distinguishable. The factor before the integral actually comes from quantum mechanics. The essential difficulty in calculating the free energy of a system is evident from Eq.

2.12, which is dependent on a $6N$ -dimensional integral to be carried out over phase space.

By means of statistical mechanics, free energy differences may also be expressed in terms of averages over ensembles of atomic configurations for the molecular system of interest. Such an ensemble can be generated by MC or MD simulation techniques. If the *ergodic hypothesis* is verified, that is the simulated trajectory will visit all possible microstates available to it, given an infinite amount of time the following equivalence holds:

$$\langle \mathcal{A}(q(t), p(t)) \rangle_{ensemble} = \langle \mathcal{A}(q(t), p(t)) \rangle_{time} \quad (2.13)$$

that is the ensemble average of a generic physical observable, $\mathcal{A}(t)$, is equivalent to its time average. In principle this equivalence offers a valid method, the time average, to obtain physical properties from our "virtual" experiment, namely computer simulations. However, despite its inherent simplicity, the computation of thermodynamic properties from molecular simulations remains far from trivial due to the limit of infinite sampling of phase space and to unavoidable numerical errors.

Within the framework of statistical mechanics, a variety of formulae for determining the difference in free energy between two states of a system, or the projection of such a difference in free energy along a spatial (reaction) coordinate, have been derived. The different formulations available are all equivalent within the limit of infinite sampling of phase space. In practice, as only a part of the total phase space accessible to a realistic system can ever be sampled by molecular simulations techniques, there are often significant differences in accuracy between the free energy estimates obtained from different formulae. Below a list of the most useful statistical mechanical formulae and computational methods to obtain the difference in free energy $\Delta A_{A \rightarrow B} = A_B - A_A$ between a state B and a state A of a molecular system in a canonical ensemble is provided.

2.3.1 Probability ratio method

In equilibrium thermodynamics, free energy changes are related to the populations (or probabilities) of states. Hence, the most straightforward way to determine the difference in free energy between two states of a system is simply to count the number of configurations in the two corresponding states. For example, in the case of folding, this involves counting the number of folded conformations N_F and the number of unfolded conformations N_U in an ensemble generated during a MD or MC simulation, with the difference in free energy being given by

$$\Delta A_{U \rightarrow F} = -k_B T \ln \frac{Q_F}{Q_U} = -k_B T \ln \frac{p_F}{p_U} = -k_B T \ln \frac{N_F}{N_U} \quad (2.14)$$

where k_B is the Boltzmann constant, T is the temperature, Q_F and Q_U are the partition functions of the folded and unfolded states, respectively, and p_F and p_U are the probability densities of finding the system in the folded or unfolded states, respectively. This technique is only appropriate when folded and unfolded conformations occur with sufficient frequency in the ensemble to obtain reliable statistics. Direct counting has the advantage that it does not depend on the definition of a reaction coordinate and it is particularly well-suited to situations in which the end states are themselves ensembles of structures, such as in the study of protein/peptide folding.

2.3.2 Thermodynamic Integration

Integrations methods determine the change in free energy between two states of a system from the integral of the work required to go from an initial state to a final state *via* a reversible path. In *Thermodynamic Integration* (TI) method an arbitrary coupling parameter, λ , is introduced in the the Hamiltonian $H(p, q, \lambda)$. The coupling parameter is chosen such that when $\lambda = \lambda_A$ the Hamiltonian of the molecular system corresponds to that of state A, i.e. $H(p, q, \lambda_A) = H(p, q)$ and when $\lambda = \lambda_B$ the Hamiltonian of the system corresponds to that of state B, i.e. $H(p, q, \lambda_B) = H(p, q)$. If the Hamiltonian is a function of λ the free energy in Eq. 2.11 will also be a function of λ , and the derivative of the free energy with respect to λ will be given by

$$\frac{dA(\lambda)}{d\lambda} = \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad (2.15)$$

From this, it follows directly that the free energy difference between state A and state B of a molecular system is given by

$$A(\lambda_B) - A(\lambda_A) = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (2.16)$$

which is the so-called thermodynamic integration formula.¹²¹ The ensemble average $\langle \partial H / \partial \lambda \rangle$ is most commonly determined from simulations at a series of λ values between λ_A and λ_B and the integral in Eq. 2.16 evaluated numerically. The choice of λ is arbitrary and λ may equally refer to a spatial coordinate or to a non-physical coordinate in parameter space. In either case, the functional dependence of the system on λ effectively describes the pathway from the initial to the final state.

2.3.3 Perturbation method

An alternative to the TI method is to adopt a perturbation approach. In the *perturbation method* (PM) the free energy change is expressed by the following relation:¹²²

$$A_B - A_A = -k_B T \ln \frac{Q_B}{Q_A} = -k_B T \ln \langle e^{\Delta H/k_B T} \rangle_B \quad (2.17)$$

where Q_B and Q_A are the partition functions of state B and A respectively, $\Delta H = H_B - H_A$ is the energy difference, k_B is the Boltzmann constant and T the absolute temperature. The subscript on the brackets $\langle \dots \rangle$ indicates that the ensemble average is performed with respect to the probability function representative of the final state, B , of the system. Thus, the free energy change is calculated directly from one MD simulation of the state B averaging the quantity $e^{\Delta H/k_B T}$. Usually, due to the known insufficient sampling of the tails of the distribution, this method gives accurate results when the energies of the initial and final states of the system differ by a relatively small amount ($\leq 2k_B T$). Otherwise, it is possible to decompose the total free energy change by defining intermediate states along a given path between the initial and final states, hence computing as a sum of partial free energy changes.

2.3.4 Potential of Mean Force

The difference in free energy between two states of a molecular system is a single number. Often we would like to know how the free energy of a system, or the *potential of mean force* (PMF), changes as a function of a particular coordinate within the system, most commonly a spatial coordinate. Chosen this coordinate, r , and considering the partial derivative of the free energy with respect to this coordinate, we obtain:

$$\frac{\partial A}{\partial r} = -k_B T \frac{1}{Q} \frac{\partial Q}{\partial r} = -k_B T \frac{1}{Q} \int \int -\frac{\partial U(q)}{\partial r} \frac{1}{k_B T} e^{-H(p,q)/k_B T} dpdq \quad (2.18)$$

Considering that $-\partial U(q)/\partial r$ is the force acting along r , $F(r)$, and that the average value of a generic function, $f(p, q)$, is given by:

$$\langle f(p, q) \rangle = \frac{1}{Q} \int \int f(p, q) e^{-H(p,q)/k_B T} dpdq, \quad (2.19)$$

Eq. 2.18 becomes

$$\frac{\partial A}{\partial r} = -\langle F(r) \rangle \quad (2.20)$$

Hence, if we are interested in the free energy change between two positions r_A and r_B , we get

$$A_B - A_A = \int_{r_A}^{r_B} -\langle F(r) \rangle dr \quad (2.21)$$

Usually the ensemble average $-\langle F(r) \rangle$ is most commonly determined from simulations at a series of r values between r_A and r_B and the integral in Eq. 2.21 evaluated numerically.

2.4 The Perturbed Matrix Method (PMM)

As it was stated in the previous sections, classical MD methods do not explicitly treat the electrons in a molecular system. Instead, they perform computations considering only interactions among nuclei. This assumption makes classical calculations quite inexpensive computationally, and allows them to be used for very large systems containing many thousands of atoms. Although this approximation is excellent for a wide range of materials, it also carries several limitations. Among them the fact that classical methods cannot treat chemical problems where electronic effects predominate. On the other hand, methods based on the electronic structure use the laws of quantum mechanics rather than classical physics as the basis for their computations. *Ab initio* techniques seek to calculate the properties of a system from first principles with no parametrization. The goal is to solve the Schrödinger equation as accurately as possible, using a series of rigorous mathematical approximations. Electronic structure methods provide high quality quantitative predictions for a broad range of systems, not limited to any specific class. Several molecular properties can be calculated with these methods, such as structures,^{123,124} spectra,^{125,126} electron scattering properties,¹²⁷ etc. Unfortunately, solving the Schrödinger equation accurately for systems with many electrons is a computationally intensive process and thus sophisticated *Ab initio* methods are limited in size of system they could handle.

The combination of quantum mechanics (QM) and molecular mechanics (MM), or more in general classical methods, offer an ideal method that combines the accuracy of QM and the power of MD necessary for biological systems which are intrinsically dynamical. The concepts behind these hybrid QM/MM methods were introduced already in the late 1970s in computational biology.¹²⁸ The QM/MM approach consists in partitioning the system of interest into two subsystems: one (QM) contains a small number of atoms and is described by quantum mechanics and the other (MM), where events such as bond breaking/formation or electronic transitions do not take place, is described by a suitable MM force field. There are several open questions and problems about the use of QM/MM approaches, as the choice of the QM part, the limited size of the system that can be treated quantum-mechanically, the modeling of the MM part and the interactions between the QM and MM regions. Nevertheless, this strategy has been demonstrated to be useful in several fields of study, such as materials science (e.g. for the study of chemically

induced stress in real materials¹²⁹), for the study of chemical processes in condensed phases (e.g. chemical reactions in solution¹³⁰) or molecular spectroscopic properties in solution.¹³¹

One of the goals of QM/MM methods is the understanding of electronic properties of a chromophore in a protein environment. This kind of problem in fact requires to mix electronic structure calculations with structural and dynamical information of the protein. In general, classical MD is used to provide a statistical ensemble of configurations representing the state of the system, and QM calculations are performed only for a limited number of configurations after the simulation is finished.^{132,133} These methods provide a very detailed description of the solvent and/or protein structure but, because the number of QM calculations is usually very large, their calculation level has to be restricted. On the other hand, if high level QM calculations are needed, the number of configurations used have to be dramatically reduced, with a consequent sampling deficiency.

A few years ago, a theoretical method, the Perturbed Matrix Method (PMM), has been developed for modeling perturbed molecular eigenstates.^{134,135} In the PMM, instead of including directly in the Hamiltonian operator the perturbation term, as usual for Hartree-Fock based calculations, the effect of the perturbation is obtained diagonalizing the perturbed Hamiltonian matrix constructed in the basis set of the unperturbed Hamiltonian eigenstates.

The above approach is in principle a rigorous procedure within the only practical approximation arising from truncating the dimension of the perturbed Hamiltonian matrix, that is to say the size of the unperturbed basis set used. However, it should be remarked that the present approach could suffer from a not rigorous derivation of the unperturbed eigenstates due to the well known limitations which could arise when highly excited states are concerned. On the other hand, it proved to be as an investigating tool very promising for large and computationally demanding molecular systems,¹³⁶⁻¹³⁸ since the unperturbed eigenfunctions have to be evaluated only once at a given geometry, although the external perturbation is varied. Finally, it should be noted that the perturbed Hamiltonian matrix can be constructed and diagonalized nowadays, at a relatively low computational cost.

2.4.1 Basic Derivations

The time independent Schrödinger equation, in matrix notation, for a perturbed system is:

$$\tilde{H}c_i = U_i c_i \quad (2.22)$$

where $\tilde{H} = \tilde{H}^0 + \tilde{V}$, c_i is the i -th eigenvector of the perturbed Hamiltonian matrix \tilde{H} , U_i the corresponding Hamiltonian eigenvalue, \tilde{H}^0 is the unperturbed Hamiltonian matrix and \tilde{V} is the perturbation energy matrix. The Hamiltonian

matrix and its eigenvectors can be expressed in the basis set defined by the unperturbed Hamiltonian matrix eigenvectors, and hence the element of the Hamiltonian matrix is

$$H_{l,l'} = \langle \Phi_l^0 | \hat{H} | \Phi_{l'}^{0'} \rangle = U_l^0 \delta_{l,l'} + \langle \Phi_l^0 | \hat{V} | \Phi_{l'}^{0'} \rangle \quad (2.23)$$

where Φ_l^0 is the l -th eigenfunction of the unperturbed Hamiltonian operator, U_l^0 the corresponding energy eigenvalue, $\delta_{l,l'}$ the Kroenecker's delta and \hat{V} the perturbation energy operator. From the above equations it is evident that for obtaining the eigenvectors and eigenvalues, and hence every property, of the perturbed Hamiltonian eigenstates, we only have to diagonalize the matrix \tilde{H} , as given by Eq. 2.23. For a system interacting with an external field, we can express in general the perturbation operator in Eq. 2.23 in terms of the electric potential \mathcal{V} as

$$\hat{V} = \sum_j q_j \mathcal{V}(\mathbf{r}_j) \quad (2.24)$$

with \mathbf{r}_j the coordinates of the j -th charged particle and q_j the corresponding charge. Expanding at the second order \mathcal{V} around a given position \mathbf{r}_0 we have

$$\begin{aligned} \mathcal{V}(\mathbf{r}_j) \cong \mathcal{V}(\mathbf{r}_0) - \sum_{k=1}^3 E_k (r_{j,k} - r_{0,k}) \\ - \frac{1}{2} \sum_{k'=1}^3 \sum_{k=1}^3 \left(\frac{\partial E_k}{\partial r_{k'}} \right)_{\mathbf{r}=\mathbf{r}_0} (r_{j,k} - r_{0,k})(r_{j,k'} - r_{0,k'}) \end{aligned} \quad (2.25)$$

where k and k_0 define the three components of a vector in space and \mathbf{r} is the generic position vector. From these equations, defining with q_T the total charge, we readily obtain

$$\langle \Phi_l^0 | \hat{V} | \Phi_{l'}^{0'} \rangle \cong q_T \mathcal{V}(\mathbf{r}_0) \delta_{l,l'} - \mathbf{E} \cdot \langle \Phi_l^0 | \hat{\mu} | \Phi_{l'}^{0'} \rangle + \frac{1}{2} T_r \left[\tilde{\Theta} \tilde{Q}_{l,l'} \right] \quad (2.26)$$

$$Q_{k,k'}^{l,l'} = \left[\tilde{Q}_{l,l'} \right]_{k,k'} = \sum_j q_j \langle \Phi_l^0 | (r_{j,k} - r_{0,k})(r_{j,k'} - r_{0,k'}) | \Phi_{l'}^{0'} \rangle \quad (2.27)$$

where

$$\Theta_{k,k'} = - \left(\frac{\partial E_k}{\partial r_{k'}} \right)_{\mathbf{r}=\mathbf{r}_0} \quad (2.28)$$

$$\hat{\mu} = \sum_j q_j (r_j - r_0) \quad (2.29)$$

Hence the complete perturbed Hamiltonian matrix is

$$\tilde{H} = \tilde{H}^0 + \tilde{I}_{qT}\mathcal{V}(\mathbf{r}_0) + \tilde{Z}_1(\mathbf{E}) + \tilde{Z}_2(\tilde{\Theta}) \quad (2.30)$$

with

$$[\tilde{Z}_1]_{l,\nu} = -\mathbf{E} \cdot \langle \Phi_l^0 | \hat{\mu} | \Phi_l^{0'} \rangle \quad (2.31)$$

$$[\tilde{Z}_2]_{l,\nu} = \frac{1}{2}T_r[\tilde{\Theta} \tilde{Q}_{l,\nu}] \quad (2.32)$$

From the last equations it is evident that a second order expansion of the electric potential, able to describe electric fields up to linear behavior over the molecular size, requires the knowledge of the total charge and the unperturbed dipoles and quadrupoles. Higher order expansions can be in principle worked out in the same way but would require information on higher order multipoles which are typically very difficult to obtain. Moreover, it is rather unusual that an applied electric field is beyond the linear approximation over a molecular size. Note that for uncharged systems and homogeneous applied fields, i.e. $\tilde{\Theta}$ is a zero elements matrix, Eq. 2.30 reduces to

$$\tilde{H} = \tilde{H}^0 + \tilde{Z}_1(\mathbf{E}) \quad (2.33)$$

The above equations could be used to describe either a perturbed Born-Oppenheimer (BO) surface, i.e. Φ_l^0 corresponding to the l -th unperturbed electronic eigenstate for a fixed nuclear position, as well as perturbed complete quantum mechanical eigenstates including nuclear degrees of freedom. However, it has been shown¹³⁵ that it is very convenient to use the above equations to evaluate the perturbed energy on the BO surfaces, and then to include the possible nuclear quantum degrees of freedom. The previous equations are completely general, but involve in principle infinite dimensional matrices and vectors and hence in their exact form are practically undoable. However, if we are interested in the perturbed ground state and the first excited states, we can approximate Eq. 2.22 using a finite dimensional Hamiltonian matrix constructed, via Eq. 2.23, by a limited subspace of the unperturbed eigenstates, ranging from the ground state to a given excited one.

2.4.2 Application of the PMM to the reconstruction of infrared spectra

In this section, the procedure applied in chapters 5.1 and 6 to study the amide I mode of peptides, is briefly presented.

Once the unperturbed quantum center (QC) mass-weighted Hessian eigenvectors are defined by standard quantum mechanics (QM) calculations on the isolated quantum center (see section 2.4.1), we may obtain, for a given j^{th} mode, the per-

turbed ω_j^2 value at each environment configuration by the PMM as follows. Once the mode is selected, it is possible to generate a number of QC configurations along the mode coordinate, q . For each of these QC configurations an orthonormal set of unperturbed electronic eigenfunctions can be obtained by standard QM procedures and, using equation 2.30, a number of perturbed electronic Hamiltonian eigenstates and eigenvalues may be evaluated at each molecular dynamics (MD) frame. Then, the perturbed electronic ground state energy, ϵ_0 (*i.e.*, the ground state eigenvalue of the matrix \tilde{H}), along the mode coordinate q can be expressed at each time frame as follows:

$$\epsilon_0(q, t) \cong \epsilon'_0(q, t) + q_T \mathcal{V}(t) + \Delta V(q, t) \quad (2.34)$$

with ϵ'_0 the ground state eigenvalue of the matrix $\tilde{H}^0 + \tilde{Z}_1$ (equations 2.30 and 2.31). Finally, the perturbed harmonic frequency is evaluated at each time frame via a polynomial fit of such an energy curve, providing the second order energy derivative at the minimum position.

As in previous papers,^{136–138} ΔV may be considered as essentially constant along the mode coordinate, at least within the harmonic range, and hence the perturbed frequency at each time frame can be properly evaluated by using only ϵ'_0 . To support this assumption, a normal mode analysis was performed on trans-N-methylacetamide/D₂O clusters and on a tri-glycine peptide, clearly showing that the amide I mode in such systems is virtually identical to that of the isolated trans-NMA. These calculations thus show that the higher order energy term, ΔV , may be neglected in PMM calculations on liquid state systems for which significantly lower shifts than in the clusters are to be expected.

THE QUASI-GAUSSIAN ENTROPY THEORY

Accurate methods to obtain the statistical mechanics and thermodynamics of simulated condensed systems are clearly of great importance as they can provide essential information to describe and predict the behavior of a molecular complex system. Despite of the great development of the simulation methods, the evaluation of essential thermodynamic properties such as free energy and entropy, and of many related observables, are very difficult and typically the methodologies used can only provide limited "local" information, i.e., a few thermodynamic properties at a given temperature and density, requiring a rather heavy computational effort. Moreover, the basic theoretical principles underlining these methods, i.e., thermodynamic integration (TI) and perturbation method (PM), can be affected by severe problems due to the slow convergence (see sections 2.3.2 and 2.3.3). It is therefore a challenge in theoretical physical chemistry to develop and optimize more analytical methods providing the thermodynamics of a simulated system at relatively low computational costs. From a theoretical point of view, the key point is the evaluation of the (configurational) partition function, which is in general a high-dimensional integral over all coordinates. For complex systems with a realistic Hamiltonian, it is virtually impossible to derive in this way rigorous but easy to handle expressions for the various thermodynamic properties. Only for very special Hamiltonians analytical solutions are available, e.g. for the (monoatomic) ideal gas, a set of quantum or classical harmonic oscillators or one and two-dimensional Ising spin systems. However, in this chapter we will show that it is possible to switch from the description in terms of this high-dimensional partition function in the canonical ensemble to the one-dimensional internal energy distribution function of the system, which is a special projection of the Hamiltonian. The advantage of this approach is the fact that these distributions, because of the macroscopic character of usual systems, are almost Gaussian ("quasi-Gaussian"). Hence, mathematically speaking, relatively simple functions can be used to model the real distributions, yielding very compact expressions for the corresponding thermodynamic functions. The theory based on

this, which will be presented in this chapter, will be denoted as the “quasi-Gaussian entropy theory” (QGE).

3.1 The quasi-Gaussian entropy theory in canonical ensemble

In this section the derivation of the basic theory for the temperature dependence of thermodynamics properties in the canonical ensemble is presented. Choosing a proper reference state, the excess entropy S^* can be expressed in terms of the moment generating function of the excess internal energy distribution function $\rho(\Delta\mathcal{U}')$. For a macroscopic system the application of the central limit theorem demonstrates that this distribution function can be modeled as a unimodal function, close to a Gaussian (“quasi-Gaussian”). Hence, from the statistical mechanical definition of $\rho(\Delta\mathcal{U}')$, the generalized Pearson system of distributions can be used to generate and classify unimodal distribution of increasing complexity. The parameters of these distributions and hence of the excess entropy can be expressed in terms of the isochoric heat capacity and a limited set of its temperature derivatives.

3.1.1 Definition of the system

The Helmholtz free energy of a system at fixed volume, temperature and number of molecules is

$$A = -k_B T \ln Q \quad (3.1)$$

where Q is the overall partition function. For a system of n identical molecules in the classical limit

$$Q = \frac{(2\pi k_B T)^{d/2}}{n! h^d (1 + \gamma)^n} \sum_l \int e^{-\beta(\Phi + \psi + \varepsilon_l)} \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x} \quad (3.2)$$

where \mathbf{x} are the semiclassical atomic coordinates, Φ is the (classical) intermolecular potential energy, ψ is the (classical) intramolecular potential energy, ε_l is the overall l th quantum energy (in general a function of the coordinates which typically only refers to vibrational states), $\beta = 1/k_B T$ and the sum runs over all accessible quantum states $\{l\}$ of the system in the temperature range of interest. Moreover d is the total number of semiclassical degrees of freedom of the system, $1 + \gamma$ is the symmetry coefficient of the molecule, h is the Planck constant and \tilde{M} is the (classical) mass tensor of the molecule. We can simplify Eq. 3.2, considering that in general for systems where the (quantum) vibrational energies have a significant dependence on the coordinates, such a dependence is typically observed only up to temperatures where the molecules are largely confined in the vibrational ground state. In such systems, as the temperature is increased to values where the first excited states

become significantly populated, the vibrational energies converge to a coordinates independent value, usually close to the ideal gas ones. For these kind of systems we can rewrite Eq. 3.2 as

$$Q \propto \sum_l \int e^{-\beta(\Phi+\psi+\varepsilon_l)} + \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x} \quad (3.3)$$

$$= \sum_l e^{-\beta E_l} \int e^{-\beta(\Phi+\psi)} e^{-\beta(\varepsilon_l-E_l)} \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x} \quad (3.4)$$

with E_l a reference quantum energy of the l th state. We assume, at least for all the terms in Eq. 3.4 significantly different from zero (i.e., βE_l not too large), that

$$\int e^{-\beta(\Phi+\psi)} e^{-\beta(\varepsilon_l-E_l)} d\mathbf{x} \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} \cong \int e^{-\beta(\Phi+\psi)} e^{-\beta(\varepsilon_0-E_0)} \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x} \quad (3.5)$$

Hence

$$Q \cong \int e^{-\beta U'} \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x} \quad (3.6)$$

$$Q_{qm}^{ref} = \sum_l e^{-\beta E_l} \quad (3.7)$$

$$\Theta = \frac{(2\pi kT)^{1/2} Q_{qm}^{ref}}{n! h^d (1 + \gamma)^n} \quad (3.8)$$

with $U' = \Phi + \psi + \varepsilon_0 - E_0$ and where E_0 and Q_{qm}^{ref} can be typically obtained by quantum calculations of the isolated molecule, i.e. in ideal gas conditions. It is easy to see that Eq. 3.6 is always exact if the system is completely confined in the vibrational ground state, or if no vibrations are present (e.g. monoatomic molecules) or when the vibrational energies can be considered independent of the coordinates, and therefore from the temperature, hence being identical to the ideal gas ones. If moreover, at least in the whole temperature range of interest, only a part of the configurational space is energetically accessible (i.e., the system is confined within a part of the configurational space) we can rewrite Eq. 3.6 as

$$Q \cong Q_{ref}^{qm} \int^* e^{-\beta U'} \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x} \quad (3.9)$$

where now the star denotes an integration in the accessible part of the configurational space only. It should be noted that the inaccessible configurations not necessarily correspond only to the ones which are forbidden by a simple excluded volume concept, where it is assumed that the pair potential energy has an infinite barrier as in a hard sphere liquid. In fact, according to the *total* intermolecular potential energy, even configurations with non penetrating molecules might be ener-

getically excluded up to very high temperatures. Clearly in the infinite temperature limit every finite energy configuration will be accessible implying that only the confinement due to the infinite energy configurations can be considered exact, i.e., really temperature independent. Eq. 3.9, which reduces to Eq. 3.6 in case the whole configurational space is available, is a very general expression that can be used for many different types of molecules.

3.1.2 Definition of the reference states and excess properties

For all the systems where Eq. 3.9 can be used, we can define a reference state at the same temperature and density, but without inter (Φ) and intramolecular potential energy (ψ). We have

$$A_{ref} = -kT \ln Q_{ref} \quad (3.10)$$

$$Q_{ref} = \Theta \int \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x} \quad (3.11)$$

Therefore the excess Helmholtz free energy is

$$A' = A - A_{ref} \quad (3.12)$$

$$A' = A^* - kT \ln \epsilon \quad (3.13)$$

$$A^* = -kT \ln \langle e^{-\beta U'} \rangle = -kT \ln \langle e^{-\beta U'} \rangle_{ref} \quad (3.14)$$

$$\langle e^{\beta U'} \rangle = \frac{\int^* e^{-\beta U'} \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} e^{\beta U'} d\mathbf{x}}{\int^* e^{-\beta U'} \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} e^{\beta U'} d\mathbf{x}} \quad (3.15)$$

$$\langle e^{-\beta U'} \rangle_{ref} = \frac{\int^* e^{-\beta U'} \prod_{j=1}^n (\delta \tilde{M}_j)^{1/2} d\mathbf{x}}{\int^* \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x}} \quad (3.16)$$

$$\epsilon = \frac{\int^* \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x}}{\int \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x}} \quad (3.17)$$

Then we can write the internal energy, heat capacity, pressure and entropy as

$$U' = - \left(\frac{\partial}{\partial \beta} \ln \frac{Q}{Q_{ref}} \right)_V = \langle U' \rangle \quad (3.18)$$

$$C'_V = \left(\frac{\partial U'}{\partial T} \right)_V = \left(\frac{\partial \langle U' \rangle}{\partial T} \right)_V \quad (3.19)$$

$$S' = - \frac{(A' - U')}{T} = S^* + k \ln \epsilon \quad (3.20)$$

$$S^* = -k \ln \langle e^{\beta(U' - \langle U' \rangle)} \rangle \quad (3.21)$$

and

$$p' = - \left(\frac{\partial A'}{\partial V} \right)_T = p^* + T\xi \quad (3.22)$$

$$p^* = - \left(\frac{\partial A^*}{\partial V} \right)_T \quad (3.23)$$

$$\xi = k \left(\frac{\partial \ln \epsilon}{\partial V} \right) \quad (3.24)$$

Finally, from these equations follow the excess enthalpy and the Gibbs free energy

$$H' = U' + p'V \quad (3.25)$$

$$G' = A' + p'V \quad (3.26)$$

3.1.3 The potential energy distribution

Eq. 3.21 can be explicitly expressed as

$$S^* = -k \ln G_{\Delta U'}(\beta) = -k \ln \int e^{\beta \Delta U'} \rho(\Delta U') d\Delta U' \quad (3.27)$$

$$\Delta U' = U' - \langle U' \rangle \quad (3.28)$$

where $G_{\Delta U'}(\beta)$ is the moment generating function^{139,140} of the potential energy distribution function $\rho(\Delta U')$ (note that ρ is in general temperature dependent). From the fact that a macroscopic system can be considered as a very large collection of identical, independent subsystems (elementary systems), we can conclude that because of the central limit theorem^{139,141} the corresponding potential energy distribution function can be regarded as uninormal. In fact the possible deviations from the uninormal shape in the far tail of the distribution are in general negligible because of the very sharply peaked behavior of the curve around its mode (maximum of probability), also due to the macroscopic nature of the system^{120,142}, and therefore the integrand in Eq. 3.27 can be modeled considering ρ as a quasi-Gaussian distribution. In general we can express the potential energy distribution as^{142,143}

$$\rho(\Delta U') = \frac{\Omega(\Delta U')}{\int^* e^{-\beta \Delta U'} \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x}} e^{-\beta \Delta U'} \quad (3.29)$$

with

$$\Omega(\Delta U') = \int^* \delta(\Delta U'(\mathbf{x}) - \Delta U') \prod_{j=1}^n (\det \tilde{M}_j)^{1/2} d\mathbf{x} \quad (3.30)$$

and hence

$$\frac{d\rho}{d\Delta U'} = -\rho(\Delta U') \left[\beta - \frac{d \ln \Omega}{d\Delta U'} \right] \quad (3.31)$$

Using a Padé approximant^{144,145}, instead of a usual Taylor series, to expand the function between brackets in Eq. 3.31 around the mode of the distribution, we finally obtain a general differential equation, equivalent to the one of the generalized Pearson system of curves¹⁴⁶⁻¹⁴⁸, which can be used to obtain the possible potential energy distribution functions:

$$\frac{d\rho}{d\Delta U'} = -(\Delta U' - \Delta U'_m) \rho \frac{P^m(\Delta U')}{G^n(\Delta U')} \quad (3.32)$$

where $\Delta U'_m$ is the value of $\Delta U'$ where ρ has its mode and $P^m(\Delta U')$ and $G^n(\Delta U')$ are some arbitrary polynomials of order m and n :

$$P^m(\Delta U') = \sum_{i=0}^m a_i (\Delta U')^i \quad (3.33)$$

$$G^n(\Delta U') = \sum_{j=0}^n b_j (\Delta U')^j \quad (3.34)$$

where without loss of generality $a_0 = 1$. The solutions of Eq. 3.32 are therefore fully defined by the parameters $\Delta U'_m$, $\{a_i\}$ and $\{b_j\}$ which can be expressed¹⁴⁶ in terms of the central potential energy moments M_n . Hence with the use of physical-mathematical restrictions¹⁴⁶ we can select the physically acceptable distributions, fully defined by a limited set of central moments, and then order them according to their increasing complexity. We also showed that these potential energy moments can be expressed as a function of the isochoric heat capacity and a limited number of its temperature derivatives¹⁴⁶

$$M_2(T) = kT^2 C'_V \quad (3.35)$$

$$M_3(T) = (kT^2)^2 \left(\frac{\partial C'_V}{\partial T} \right)_V + 2(kT)^2 T C'_V \quad (3.36)$$

...

$$M_k(T) = M_k \left(T, C'_V, \left(\frac{\partial C'_V}{\partial T} \right)_V, \left(\frac{\partial^2 C'_V}{\partial T^2} \right)_V, \dots, \left(\frac{\partial^{k-2} C'_V}{\partial T^{k-2}} \right)_V \right) \quad (3.37)$$

where $M_k = \langle (\Delta U')^k \rangle$ is the k th central potential energy moment and we used the fact that $C_V^* = C'_V$. If the functional shape of ρ is defined by the first n central moments, from Eqs. 3.27 and 3.37 it follows that

$$S^* = S^* \left(T, C'_V, \left(\frac{\partial C'_V}{\partial T} \right)_V, \left(\frac{\partial^2 C'_V}{\partial T^2} \right)_V, \dots, \left(\frac{\partial^{n-2} C'_V}{\partial T^{n-2}} \right)_V \right) \quad (3.38)$$

3.1.4 The thermodynamic master equation

We define the *intrinsic entropy function* as ¹⁴⁶

$$\begin{aligned}\alpha &= \frac{S'}{C'_V} = \frac{S^*}{C'_V} + \frac{k \ln \epsilon}{C'_V} \\ &= \alpha^* \left(T, C'_V, \left(\frac{\partial C'_V}{\partial T} \right)_V, \left(\frac{\partial^2 C'_V}{\partial T^2} \right)_V, \dots, \left(\frac{\partial^{n-2} C'_V}{\partial T^{n-2}} \right)_V \right) + \frac{k \ln \epsilon}{C'_V}\end{aligned}\quad (3.39)$$

with $\alpha^* = S^*/C'_V$ the *confined* intrinsic entropy function, and using the general thermodynamic relation $(\partial S'/\partial T)_V = (\partial S^*/\partial T)_V = C'_V/T$, we obtain the *thermodynamic master equation* ¹⁴⁶(TME)

$$\frac{C'_V}{T} = C'_V \left(\frac{\partial \alpha^*}{\partial T} \right)_V + \alpha^* \left(\frac{\partial C'_V}{\partial T} \right)_V \quad (3.40)$$

This is a completely defined differential equation where its unique, always existing solution gives the temperature dependence of the ideal reduced isochoric heat capacity C'_V , once the values of $C'_V, \dots, (\partial^{n-2} C'_V)/(\partial T^{n-2})_V$ at one arbitrary temperature T_0 are known. Note that in Eqs. 3.39 and 3.40 we used a generalized expression of the intrinsic entropy function, treating explicitly the possible confinement of the system, which reduces to the usual one when $\epsilon = 1$ and therefore $\alpha = \alpha^*$. From the solution of the master equation we can obtain the confined entropy S^* via Eq. 3.38, and then the excess internal energy and free energy at fixed density via

$$U'(T) = U'(T_0) + \int_{T_0}^T C'_V(T) dT \quad (3.41)$$

$$A^*(T) = U'(T) - TS^*(T) \quad (3.42)$$

It is interesting to note that, since

$$\alpha^* = \frac{S^*}{C'_V} = -\frac{S^*/\beta}{(\partial S^*/\partial \beta)_V} \quad (3.43)$$

the confined intrinsic entropy function is the ratio between the average slope of S^* versus β (remembering that $S^* = 0$ when $\beta = 0$) and the instantaneous slope $(\partial S^*/\partial \beta)_V$. One can moreover prove that ^{146,149}

$$\lim_{T \rightarrow \infty} \alpha^*(T) = -\frac{1}{2} \quad (3.44)$$

which implies that the (confined) thermodynamics of every system in the infinite temperature limit tends to a Gaussian one, see section 3.2.1.

3.1.5 The conjugated pressure equation

From Eq. 3.38 it is moreover possible to obtain the excess pressure of the system p' as a function of the temperature. In fact from the basic thermodynamic relation

$$\left(\frac{\partial S^*}{\partial V}\right)_T = \left(\frac{\partial p^*}{\partial T}\right)_V \quad (3.45)$$

we obtain

$$\frac{1}{T} \left(\frac{\partial C'_V}{\partial V}\right)_T = \left(\frac{\partial^2 p^*}{\partial T^2}\right)_V \quad (3.46)$$

and hence in general

$$\left(\frac{\partial f_0}{\partial V}\right)_T = T \left(\frac{\partial^2 p^*}{\partial T^2}\right)_V \quad (3.47)$$

$$\left(\frac{\partial f_1}{\partial V}\right)_T = \left(\frac{\partial^2 p^*}{\partial T^2}\right)_V + T \left(\frac{\partial^3 p^*}{\partial T^3}\right)_V \quad (3.48)$$

$$\dots \quad (3.49)$$

$$\left(\frac{\partial f_l}{\partial V}\right)_T = l \left(\frac{\partial^{l+1} p^*}{\partial T^{l+1}}\right)_T + T \left(\frac{\partial^{l+2} p^*}{\partial T^{l+2}}\right)_V \quad (3.50)$$

where

$$f_0 = C'_V ; \quad f_1 = \left(\frac{\partial C'_V}{\partial T}\right)_V ; \quad f_l = \left(\frac{\partial^l C'_V}{\partial T^l}\right)_V \quad (3.51)$$

Therefore from Eqs. 3.38, 3.39 and 3.45 we finally obtain another closed differential equation at fixed density for the temperature dependence of the confined pressure p^* , the *conjugated pressure equation* (CPE)

$$\left(\frac{\partial p^*}{\partial T}\right)_V = T \alpha^*(T) \left(\frac{\partial^2 p^*}{\partial T^2}\right)_V + C'_V \left(\frac{\partial \alpha^*}{\partial V}\right)_T \quad (3.52)$$

or, equivalently, using Eqs. 3.38 and 3.46-3.51

$$\begin{aligned} \left(\frac{\partial p^*}{\partial T}\right)_V &= \sum_l \left[\left(\frac{\partial S^*}{\partial f_l}\right)_{T, f_l}, \left(\frac{\partial f_l}{\partial V}\right)_T \right] \\ &= \sum_l \left[\left(\frac{\partial S^*}{\partial f_l}\right)_{T, f_l}, \left\{ l \left(\frac{\partial^{l+1} p^*}{\partial T^{l+1}}\right)_V + T \left(\frac{\partial^{l+2} p^*}{\partial T^{l+2}}\right)_V \right\} \right] \end{aligned} \quad (3.53)$$

with $l' \neq l$ and $l = 0, \dots, (n-2)$. Eq. 3.52 or 3.53 can be solved once the first $n-1$ temperature derivatives of the confined pressure are known at one arbitrary temperature T_0 . Its solution provides the temperature dependence of $(\partial p^*/\partial T)_V$ for a system at fixed density and, if the excess pressure is known at one temperature as well, also the temperature dependence of the excess pressure, enthalpy and Gibbs

free energy via

$$p'(T) = p'(T_0) + \int_{T_0}^T \left(\frac{\partial p^*}{\partial T} \right)_V dT + \xi(T - T_0) \quad (3.54)$$

$$H'(T) = U'(T) + p'(T)V \quad (3.55)$$

$$G'(T) = A'(T) + p'(T)V \quad (3.56)$$

Note that it is not necessary to explicitly solve the conjugated pressure differential equation if the temperature dependence of A^* is already known from the thermodynamic master equation, since $p^*(T) = -(\partial A^*(T)/\partial V)_T$. As $A^*(T)$ is an explicit function of T only, with U'_0 , C'_{V0} , $(\partial C'_{V0}/\partial T)_V$ (i.e., the values of U' etc. at the reference temperature T_0), as parameters depending only on the volume, the resulting volume derivatives at T_0 can be related to p_0^* , $(\partial p_0^*/\partial T)_V$, $(\partial^2 p_0^*/\partial T^2)_V$, \dots , according to Eqs. 3.47-3.51 (see also the description of the Gamma and Inverse Gaussian states, sections 3.2.1 and 3.2.2).

Hence the knowledge of the potential energy distribution function at one temperature, as well as the knowledge of ϵ and ξ fully defines the whole excess thermodynamics of a system at every temperature at fixed density. Every potential energy distribution function therefore defines a different statistical state of the system with a thermodynamical complexity given by the number of the heat capacity temperature derivatives, necessary to define the corresponding type of potential energy distribution function.

3.1.6 Phase-space confinement

A very simple model for the phase-space confinement is excluded volume due to “hard body” interactions. If, because of the strong Pauli repulsions at close contact, it is possible to define for (almost) spherical molecules like water an interparticle distance $\sigma_{HS} = 2r_{HS}$ at which the two-particle interaction energy is virtually infinite in the temperature range of interest, the confined ideal reference state can be described as a hard sphere (HS) system^{120,150}. Using the Carnahan-Starling equation of state (EOS)¹⁵¹,

$$p_{HS} = \rho_N kT \left[\frac{1 + \eta + \eta^2 - \eta^3}{(1 - \eta)^3} \right] \quad (3.57)$$

with p_{HS} the pressure, $\eta = v\rho_N$, ρ_N the number density, $v = \pi\sigma_{HS}^3/6$ the hard-sphere volume per molecule and σ_{HS} the hard-sphere diameter, we obtain¹⁵⁰ for the excess Helmholtz free energy

$$A'_{ref} = -NkT \left[\frac{3\eta^2 - 4\eta}{(1 - \eta)^2} \right] = -NkT \ln \bar{\epsilon}_{HS} \quad (3.58)$$

where $\bar{\epsilon}_{HS} = \epsilon_{HS}^{1/N}$ is the hard sphere phase-space fraction *per molecule*. Hence

$$\bar{\epsilon}_{HS} = \exp \left\{ \frac{3\eta^2 - 4\eta}{(1 - \eta)^2} \right\} \quad (3.59)$$

and

$$\xi_{HS} = Nk \frac{d \ln \bar{\epsilon}_{HS}}{dV} = -\frac{Nk}{V} \left[\frac{2\eta^2 - 4\eta}{(1 - \eta)^3} \right] \quad (3.60)$$

which is always positive since $\eta < 1$. Within this model, p^* is the pressure of the system with respect to a HS fluid at the same temperature and density. For more complex molecules deviating from an almost spherical shape a simple HS description may not be sufficient. For non spherical molecules several equation of state have been proposed.

3.2 Description of different statistical states

In this section the temperature dependence of thermodynamic properties for different statistical states are presented. In particular, the simplest solutions of the Generalized Pearson systems, the Gaussian and Gamma states, are discussed in details.

3.2.1 Gaussian state

The symmetric Gaussian distribution corresponds to $\{m = 0, n = 0\}$ in the generalized Pearson system, Eq. 3.32. The distribution is given by

$$\rho(\Delta U') = \frac{1}{\sqrt{2\pi b_0}} \exp \left\{ -\frac{\Delta U'^2}{2b_0} \right\} \quad (3.61)$$

with $b_0 = M_2$. Using the fact that the moment generating function of a Gaussian is simply ^{139, 146, 152}

$$G_{\Delta U'}(\beta) = \exp \left\{ \frac{b_0 \beta^2}{2} \right\} \quad (3.62)$$

and expressing M_2 in terms of C'_V (Eq. 3.35), we obtain for the confined intrinsic entropy function

$$\alpha^* = -\frac{1}{2} \quad (3.63)$$

The thermodynamic master equation (TME), Eq. 3.40, therefore reduces to

$$\left(\frac{\partial C'_V}{\partial T} \right)_V = -2 \frac{C'_V}{T} \quad (3.64)$$

the solution of which provides for a confined Gaussian state ¹⁴⁶

$$\alpha^*(T) = -\frac{1}{2} \quad (3.65)$$

$$C'_V(T) = C'_{V0} \left(\frac{T_0}{T} \right)^2 \quad (3.66)$$

$$S'(T) = -\frac{1}{2} C'_{V0} \left(\frac{T_0}{T} \right)^2 + k \ln \epsilon \quad (3.67)$$

$$U'(T) = U'_0 + T_0 C'_{V0} \left(1 - \frac{T_0}{T} \right) \quad (3.68)$$

and

$$A'(T) = U'_0 + T_0 C'_{V0} \left(1 - \frac{T_0}{2T} \right) - kT \ln \epsilon \quad (3.69)$$

where U'_0 and C'_{V0} are the values of U' and C'_V at an arbitrary reference temperature T_0 . Since $(\partial\alpha^*/\partial V)_T = 0$, the conjugated pressure equation (CPE), Eq. 3.52 is in this case a simple first order differential equation

$$\left(\frac{\partial p^*}{\partial T} \right)_V = -\frac{1}{2} T \left(\frac{\partial^2 p^*}{\partial T^2} \right)_V \quad (3.70)$$

The solution is

$$p^*(T) = p^*_0 + T_0 \left(\frac{\partial p^*_0}{\partial T} \right)_V \left(1 - \frac{T_0}{T} \right) \quad (3.71)$$

where p^*_0 and $(\partial p^*_0/\partial T)_V$ are the values at the reference temperature T_0 . The excess pressure $p'(T)$ follows from Eq. 3.71, using $p' = p^* + T\xi$ and $(\partial p'/\partial T)_V = (\partial p^*/\partial T)_V + \xi$ (Eq. 3.22):

$$p'(T) = p'_0 + T_0 \left\{ \left(\frac{\partial p'_0}{\partial T} \right)_V - \xi \right\} \left(1 - \frac{T_0}{T} \right) + \xi(T - T_0) \quad (3.72)$$

It must be noted that only for the confined properties the infinite temperature limits are always physically meaningful.

3.2.2 Gamma state

The Gamma distribution, shown in Figure 3.1 and given by: ^{139,146,149}

$$\rho(\Delta U') = \frac{b_1(1/b_1^2)^{b_0/b_1^2}}{\Gamma(b_0/b_1^2)} (b_0 + b_1 \Delta U')^{b_0/b_1^2 - 1} \exp \left\{ -\frac{b_0 + b_1 \Delta U'}{b_1^2} \right\} \quad (3.73)$$

with $\Gamma(\cdot)$ the Gamma function ¹⁵³, $b_0 = M_2$ and $b_1 = M_3/(2M_2)$, corresponds to the $\{m = 0, n = 1\}$ solution of the generalized Pearson system, Eq. 3.32.

The moment generating function of this Gamma distribution is ^{139,146,152}

$$G_{\Delta U'}(\beta) = \exp \left\{ -\beta \frac{b_0}{b_1} \right\} (1 - \beta b_1)^{-b_0/b_1^2} \quad (3.74)$$

and expressing the central moments M_2 and M_3 , appearing in b_0 and b_1 , in terms of C'_V and $(\partial C'_V/\partial T)_V$ (Eqs. 3.35-3.36), the confined intrinsic entropy function is in this case given by ¹⁴⁶

$$\alpha^* = \frac{1}{\delta} + \frac{1}{\delta^2} \ln(1 - \delta) \quad (3.75)$$

with

$$\delta = \frac{M_3}{2kTM_2} = \frac{T(\partial C'_V/\partial T)_V}{2C'_V} + 1 \quad (3.76)$$

The expression of α^* , combined with the general thermodynamic master equation (Eq. 3.40), yields

$$T \left(\frac{\partial \alpha^*}{\partial T} \right)_V + 2(\delta - 1)\alpha^* - 1 = 0 \quad (3.77)$$

where $(\partial \alpha^*/\partial T)_V = (d\alpha^*/d\delta)(\partial \delta/\partial T)_V$ and $(d\alpha^*/d\delta)$ follows from Eq. 3.75. After straightforward algebra we obtain as a simple form of the TME

$$\left(\frac{\partial \delta}{\partial T} \right)_V = -\frac{\delta(1 - \delta)}{T} \quad (3.78)$$

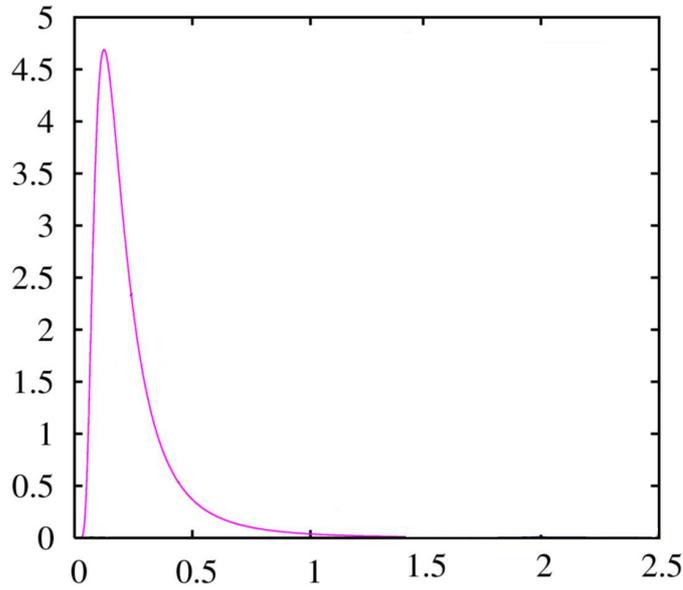


Figure 3.1: The Gamma distribution

The solution provides $\delta(T)$, which can be expressed in terms of C'_V and $(\partial C'_V/\partial T)_V$. Hence this gives rise to a new differential equation in C'_V , the solution of which yields

for a confined Gamma state ^{146,149}

$$\delta(T) = \frac{T_0 \delta_0}{T(1 - \delta_0) + T_0 \delta_0} \quad (3.79)$$

$$\alpha^*(T) = \frac{1}{\delta(T)} + \frac{1}{\delta^2(T)} \ln \{1 - \delta(T)\} \quad (3.80)$$

$$C'_V(T) = C'_{V_0} \left(\frac{\delta(T)}{\delta_0} \right)^2 \quad (3.81)$$

$$S'(T) = \frac{C'_{V_0}}{\delta_0^2} [\delta(T) + \ln \{1 - \delta(T)\}] + k \ln \epsilon \quad (3.82)$$

$$U'(T) = U'_0 + (T - T_0) C'_{V_0} \frac{\delta(T)}{\delta_0} \quad (3.83)$$

$$A'(T) = U'_0 - \frac{T_0 C'_{V_0}}{\delta_0} - \frac{T C'_{V_0}}{\delta_0^2} \ln \{1 - \delta(T)\} - kT \ln \epsilon \quad (3.84)$$

with δ_0 the value of δ at T_0 . Note that $(\partial C'_V / \partial T)_V$ is always negative, like in the Gaussian state. Since for a Gamma state the pressure differential equation (Eq. 3.52) is very complicated, it is more convenient to obtain the p^* via the volume derivative of $A^*(T)$, using Eq. 3.84. Thus, with the usual notation where $(\partial U'_0 / \partial V)_T$ etc. are the derivatives evaluated at T_0

$$p^*(T) = - \left[\frac{\partial A^*}{\partial U'_0} \left(\frac{\partial U'_0}{\partial V} \right)_T + \frac{\partial A^*}{\partial C'_{V_0}} \left(\frac{\partial C'_{V_0}}{\partial V} \right)_T + \frac{\partial A^*}{\partial \delta_0} \left(\frac{\partial \delta_0}{\partial V} \right)_T \right] \quad (3.85)$$

where from general thermodynamic relations we have

$$\left(\frac{\partial U'_0}{\partial V} \right)_T = T_0 \left(\frac{\partial p_0^*}{\partial T} \right)_V - p_0^* \quad (3.86)$$

$$\left(\frac{\partial C'_{V_0}}{\partial V} \right)_T = T_0 \left(\frac{\partial^2 p_0^*}{\partial T^2} \right)_V \quad (3.87)$$

and

$$\left(\frac{\partial \delta_0}{\partial V} \right)_T = \frac{1}{C'_{V_0}} \left(\frac{d\alpha_0^*}{d\delta_0} \right)^{-1} \left[\left(\frac{\partial p_0^*}{\partial T} \right)_V - \alpha_0^* T_0 \left(\frac{\partial^2 p_0^*}{\partial T^2} \right)_V \right] \quad (3.88)$$

as follows from the CPE at T_0 , with $(d\alpha_0^* / d\delta_0)$ following from Eq. 3.75. The derivatives $\partial A^* / \partial U'_0$, $\partial A^* / \partial C'_{V_0}$ and $\partial A^* / \partial \delta_0$ follow from Eq. 3.84. After tedious but straightforward algebra we finally obtain

$$p^*(T) = p_0^* + B_0^* + B_1^* \frac{T}{T(1 - \delta_0) + T_0 \delta_0} + B_2^* \frac{T}{T_0} \ln \left\{ \frac{T(1 - \delta_0)}{T(1 - \delta_0) + T_0 \delta_0} \right\} \quad (3.89)$$

where

$$B_i^* = A_{i1} T_0 \left(\frac{\partial p_0^*}{\partial T} \right)_V + A_{i2} T_0^2 \left(\frac{\partial^2 p_0^*}{\partial T^2} \right)_V \quad i = 0, 1, 2 \quad (3.90)$$

with

$$2A_{01} = -\frac{2(1-\delta_0)\ln(1-\delta_0) + \delta_0}{D} \quad (3.91)$$

$$A_{02} = \frac{1}{\delta_0} \frac{(1-\delta_0)\ln(1-\delta_0) + \delta_0}{D}$$

$$A_{11} = \frac{\delta_0}{D} \quad (3.92)$$

$$A_{12} = -\frac{1}{\delta_0} \frac{\ln(1-\delta_0) + \delta_0}{D} \quad (3.93)$$

$$A_{21} = \frac{2(1-\delta_0)}{D} \quad (3.94)$$

$$A_{22} = \frac{1}{D}$$

and

$$D = 2(1-\delta_0)\ln(1-\delta_0) + \delta_0(2-\delta_0) \quad (3.95)$$

The excess pressure $p'(T)$ follows from Eq. 3.89, using $p' = p^* + T\xi$, $(\partial p'/\partial T)_V = (\partial p^*/\partial T)_V + \xi$ and $(\partial^2 p'/\partial T^2)_V = (\partial^2 p^*/\partial T^2)_V$ (Eq. 3.22):

$$p'(T) = p'_0 + B_0 + B_1 \frac{T}{T(1-\delta_0) + T_0\delta_0} + B_2 \left(\frac{T}{T_0}\right) \ln \left\{ \frac{T(1-\delta_0)}{T(1-\delta_0) + T_0\delta_0} \right\} + \xi T \quad (3.96)$$

where

$$B_0 = A_{01}T_0 \left(\frac{\partial p'_0}{\partial T}\right)_V + A_{02}T_0^2 \left(\frac{\partial^2 p'_0}{\partial T^2}\right)_V + A_{03}T_0\xi \quad (3.97)$$

$$B_i = A_{i1}T_0 \left\{ \left(\frac{\partial p'_0}{\partial T}\right)_V - \xi \right\} + A_{i2}T_0^2 \left(\frac{\partial^2 p'_0}{\partial T^2}\right)_V \quad i = 1, 2 \quad (3.98)$$

and

$$A_{03} = -\frac{\delta_0(1-\delta_0)}{D} \quad (3.99)$$

As pointed out previously¹⁴⁹, since the Gamma distribution has a limited domain, it can be defined either from $-\infty$ to some upper limit or from some under limit to $+\infty$. The first case corresponds to a distribution with an asymmetric tail on the left (defining the negative Gamma state Γ_- , with $\delta < 0$ and $-\frac{1}{2} < \alpha^* < 0$), the second case corresponds to an asymmetric tail on the right (defining the positive Gamma state Γ_+ , with $0 < \delta < 1$ and $\alpha^* < -\frac{1}{2}$). Note that for $\delta = 0$ the distribution is a Gaussian. The properties of and differences between the two Gamma states have been described in detail¹⁴⁹.

Basically, the Γ_+ state is physically acceptable in the whole semi-classical temperature range, whereas the Γ_- state must be considered as an approximation to a more complicated statistical state within some temperature interval. Because of the finite upper energy limit in a Γ_- distribution, the approximation will be worse for increasing temperature. One other difference is the low temperature limit: while a

Γ_+ state may be extrapolated to $T \rightarrow 0$ (although of course in that limit the semi-classical description is not valid any more), for a Γ_- state there exist a temperature

$$T_* = -T_0\delta_0/(1 - \delta_0) > 0 \quad (3.100)$$

at which the solution encounters a singularity.

3.3 Extension of the QGE theory to threat conformational equilibria

The QGE theory presented in section 3.1 is basically an extension of the fluctuation theory.¹⁵⁴ In QGE theory, the fundamental expressions of statistical mechanics are reformulated in terms of the distribution function of the fluctuations of a macroscopic property, such as the potential energy of the system. By modeling the distribution of this property, an analytical solution for the thermodynamics of the system can be obtained. It has been shown that this theory reproduces experimental fluidliquid state thermodynamics with high accuracy for a variety of physicalchemical systems over very large temperature and density ranges.^{149,155-158} However, attempts to describe the statistical mechanics of conformational equilibria based on simulation data have been hitherto limited to simple molecules in the (ideal) gas phase. Hereby, an extension of the QGE theory is presented to describe the complete thermodynamics of the conformational states of macromolecular systems (e.g. peptides) over a large temperature range.

3.3.1 Conformational thermodynamics of peptides

In the QGE theory, the thermodynamic quantities of the system are expressed in terms of an excess with respect to a theoretical thermodynamic reference condition (QGE reference condition). Such a reference state is identical to the actual system conditions for chemical composition, number of molecules, volume, and temperature, but its Hamiltonian does not include any intra- and intermolecular potential energy, that is, molecules do not interact, and only the kinetic energy, the reference (quantum) vibrational energy and the reference electronic ground-state energy are considered. Therefore, for such a virtual ideal gas state, the semiclassical degrees of freedom of each molecule move freely, and the (quantum) vibrational modes within the molecule, classically equivalent to holonomic constraints and typically associated with bond length and angle coordinates, ensure the topological stability of the molecules. The thermodynamics of this reference condition can be typically obtained by simple statistical mechanical calculations, while the excess thermodynamics can be expressed by the distribution of the systems potential energy fluctuations. Hence, for a given model distribution considered, a corresponding analytical model for the

excess thermodynamics is obtained. It has been shown that the QGE Gamma state (i.e., the QGE solution based on modeling the potential energy fluctuations by a Gamma distribution) provides an accurate quantitative description of the excess thermodynamics of liquid systems made of rigid or semirigid molecules.^{149, 155–161} Here, the solute is considered to be infinitely diluted, and therefore, the complete solution thermodynamics can be obtained considering only a single solute embedded in a large number of solvent molecules,¹⁶⁰ that is, the simulation box termed “whole system” here.

To express the solute conformational state thermodynamics in the QGE framework, the chemical potential of the solute’s i th conformational state, μ_i , is related to the corresponding excess chemical potential μ'_i ^{160,162} (the chemical potential shift between the actual and QGE reference conditions)

$$\Delta\mu_i = \Delta\mu'_i + \Delta\mu_{i,ref} = \mu_i - \mu_r = A_i - A_r = -k_B T \ln \frac{P_i}{P_r} \quad (3.101)$$

The Δ always refers to the change of a property in a given conformation with respect to its value in the reference conformational state r , for example, $\Delta\mu'_i = \mu'_i - \mu'_r$. A_i and A_r are the Helmholtz free energies of the whole solute-solvent system with the solute in the i th and reference conformations, respectively, the subscript *ref* indicates properties in the QGE reference condition, and P is the equilibrium probability of a given conformational state. The reference chemical potential change in Eq. 3.101 is readily obtained from the definition of the QGE reference condition,^{160,162} assuming no significant variation of the solute quantum vibrational partition function for the different conformational states

$$\Delta\mu_{i,ref} = -k_B T \ln \frac{\int_i \sqrt{\det \tilde{M}} dx_{in}}{\int_r \sqrt{\det \tilde{M}} dx_{in}} = -k_B T \ln \frac{\sqrt{\det \tilde{M}_i}}{\sqrt{\det \tilde{M}_r}} - k_B T \ln \frac{\Omega_i}{\Omega_r} \quad (3.102)$$

with

$$\sqrt{\det \tilde{M}_i} = \frac{\int_i \sqrt{\det \tilde{M}} dx_{in}}{\int_i dx_{in}} \quad (3.103)$$

$$\Omega_i = \int_i dx_{in} \quad (3.104)$$

and likewise for state r . Here, x_{in} are the solute (classical) internal coordinates, the subscript of the integral sign indicates that integration is taken only over the intramolecular configurational subspace associated with the corresponding conformational state and \tilde{M} is the (classical) solute mass tensor. Note that if only the bond length stretching vibrational modes are considered in the peptide (quantum) vibrational partition function, the assumption that vibrational energies are independent of the conformational states is an excellent approximation. In fact, for the high-

frequency stretching modes, the atomic environment modifications induced by the conformational transitions, which are unaccompanied by covalent rearrangement, typically provide frequency variations within a few cm^{-1} .

The excess chemical potential in Eq. 3.101 is provided by the Gamma state solution of the QGE theory¹⁶⁰

$$\Delta\mu'_i = \Delta u'_{0,i} - \Delta c'_{V0,i} T_0 \Lambda(T) + p'(T) \Delta v_i - k_B T \ln \frac{\bar{\epsilon}_i}{\bar{\epsilon}_r} \quad (3.105)$$

Substituting Eqs. 3.102 and 3.105 into Eq. 3.101 provides along an isochore

$$\Delta\mu_i = \Delta u'_{0,i} - \Delta c'_{V0,i} T_0 \Lambda(T) + p'(T) \Delta v_i - k_B T \ln \gamma_i \quad (3.106)$$

with

$$\begin{aligned} \Delta u'_{0,i} &= u'_i(T_0) - u'_r(T_0) \\ \Delta c'_{V0,i} &= c'_{V,i}(T_0) - c'_{V,r}(T_0) \\ \Delta v_i &= v_i - v_r \\ \gamma_i &= \frac{\Omega_{i\bar{\epsilon}_i} \sqrt{\det \tilde{M}_i}}{\Omega_{r\bar{\epsilon}_r} \sqrt{\det \tilde{M}_r}} \end{aligned} \quad (3.107)$$

Here, u' is the excess partial molecular internal energy, c'_V the excess isochore partial molecular heat capacity, and v the partial molecular volume, which is temperature independent along an isochore for a Gamma state. The $\bar{\epsilon}$ is the confinement fraction providing a pure entropic partial molecular term that is typically associated with the hard-body-excluded volume; $p' = p - p_{ref}$ is the excess pressure, and T_0 is the reference temperature

$$\Lambda(T) = \frac{1}{\delta_0} + \frac{T}{T_0 \delta_0^2} \ln \frac{T(1 - \delta_0)}{T(1 - \delta_0) + T_0 \delta_0} \quad (3.108)$$

where δ_0 is a dimensionless constant. Note that within solute infinite dilution conditions, p' and Λ , being intensive properties, are fully determined by the solvent, that is, they are identical to the corresponding pure solvent Gamma state functions.^{160,162}

Using general thermodynamic relations for excess partial molecular properties,¹⁶⁰ we may relate the partial molecular Helmholtz free energy a and the chemical potential with the corresponding excess properties via

$$\mu' = a' + p'v = \mu - mu_{ref} = a - a_{ref} + pv - p_{ref}v_{ref} \quad (3.109)$$

with $a_{ref} = \mu_{ref} - p_{ref}v_{ref}$; $p_{ref} = \rho_s k_B T$ and $p_{ref}v_{ref} = k_B T$. This provides:

$$a' = a - a_{ref} + \rho_s k_B T v - k_B T \quad (3.110)$$

where ρ_s is the solvent molecular density. In the QGE reference condition, the partial molecular internal energy is independent of the system volume, that is, $u' = u - u_{ref}$, where, within our approximations, u_{ref} is a constant identical for all of the conformational states. Using the general expressions of the Gamma state excess thermodynamic properties,¹⁶⁰ we obtain along an isochore for the partial molecular properties

$$\Delta a_i(T) = \Delta u'_{0,i} -'_{V_0,i} T_0 \Lambda(T) - k_B T \ln \gamma_i - \rho_s k_B T \Delta v_i \quad (3.111)$$

$$\Delta s_i(T) = \frac{T_0 \Delta c'_{V_0,i}}{T} \left[\frac{T - T_0}{T(1 - \delta_0) + T_0 \delta_0} + \Lambda(T) \right] + k_B \ln \gamma_i + \rho_s k_B \Delta v_i \quad (3.112)$$

$$\Delta u_i(T) = \Delta u'_i(T) = \Delta u'_{0,i}(T) + \frac{(T - T_0) \Delta c'_{V_0,i} T_0}{T(1 - \delta_0) + T_0 \delta_0} \quad (3.113)$$

$$\Delta c_{V,i}(T) = \Delta c'_{V,i} = \Delta c'_{V_0,i} \left(\frac{T_0}{T(1 - \delta_0) + T_0 \delta_0} \right)^2 \quad (3.114)$$

$$\Delta u_{V,i}(T) = \left(\frac{\partial(\beta \Delta \mu_i)}{\partial \beta} \right)_V = U_i(T) - U_r(T) \quad (3.115)$$

with $\beta = (k_B T)^{-1}$, U_i and U_r being the internal energies of the whole system with the solute in the i th and reference conformations, respectively, s the partial molecular entropy, and use having been made of $\Delta a_{i,ref} = \Delta \mu_{i,ref}$. Note that the entropy change (Eq. 3.112) is expressed by three terms. The first one, proportional to $\Delta c'_{V_0,i}$, reflects the changes of the potential energy fluctuations due to the conformational change, the second one ($k_B \ln \gamma_i$) provides the effect of the different accessible configurational volumes for the rototranslationally fixed solute, and the third one ($\rho_s k_B \Delta v$) is due to the change of the available three-dimensional volume resulting from different solute partial molecular volumes.

The properties $\Delta u'_{0,i}$, $\Delta c'_{V_0,i}$, Δv_i and $\ln \gamma_i$ are regarded as parameters and can be estimated if simulation data are available for at least four different thermodynamic conditions (e.g., temperatures). A fitting procedure is performed in two stages. First, $\Delta u'_{0,i}$, $\Delta c'_{V_0,i}$ and Δv_i may be obtained by fitting the mean potential energy of the whole simulation box with the corresponding Gamma state expression, for each solute conformation. Second, by fitting the chemical potential change as obtained by MD simulations via $\Delta \mu = A_i - A_r = -k_B T \ln(P_i/P_r)$ with Eq. 3.106, γ may be evaluated for each conformation.

The previous equations readily provide the solute thermodynamics considering

all of its conformational states. Of particular interest, when dealing with peptides, is to distinguish between the species experimentally characterized as “unfolded” (U) and “folded” (F). In particular, the probability of the unfolded condition, P_U , and the unfolding free energy, $\Delta\mu_U$, are given by

$$P_U = \frac{\sum_{i \in U} \exp(-\beta\Delta\mu_i)}{\sum_j \exp(-\beta\Delta\mu_j)} \quad (3.116)$$

$$\Delta\mu_U = -k_B T \ln \frac{P_U}{1 - P_U} = -k_B T \ln \frac{\sum_{i \in U} \exp(-\beta\Delta\mu_i)}{\sum_{j \notin U} \exp(-\beta\Delta\mu_j)} \quad (3.117)$$

It is worthwhile noting that this QGE model, describing the peptide thermodynamics along an isochore, might be also used to reconstruct the peptide thermodynamics along an isobar by using QGE models for different solution isochores.

STRUCTURAL, THERMODYNAMIC, AND KINETIC PROPERTIES OF GRAMICIDIN ANALOGUE GS6

Gramicidin S (GS) analogues belong to an important class of cyclic peptides, characterized by an antiparallel double-stranded β -sheet structure with Type II' β -turns. Such compounds can be used as model systems to understand the folding/unfolding process of β -hairpins and more in general of β -structures. In the present study, we specifically investigate the folding/unfolding behavior of the hexameric Gramicidin S analogue GS6 by using all-atoms molecular dynamics (MD) simulations at different temperatures, coupled to a statistical mechanical model based on the Quasi Gaussian Entropy theory. In previous articles,^{163,164} we quantitatively characterized the folding/unfolding kinetics and thermodynamics of simple peptides by using MD simulations and advanced theoretical models based on statistical mechanics. Such an approach permits to describe the structural, thermodynamic, and kinetic properties of the peptide and to quantitatively characterize its folding/unfolding transitions.

GS analogues containing 6, 8, 10, 12, 14, and 16 residues were largely studied with circular dichroism spectroscopy, showing, in the series composed by GS6, GS10, G14, a higher propensity to form ordered β -sheet with respect to the others.¹⁶⁵ The solution structures of GS6, GS10, GS14 have been solved by NMR spectroscopy,^{165,166} revealing that they form stable antiparallel β -hairpin structures, bordered by two Type II' β -turns. Moreover, the dynamics of β -turn formation and the folding/unfolding rates of the same peptides were investigated using equilibrium Fourier transform infrared spectroscopy and T-jump relaxation probed by time-resolved infrared spectroscopy.¹⁶⁷

The aim of the present study is to quantitatively reproduce the experimentally observed properties to test the accuracy of the theoretical-computational approach employed, providing a deeper understanding of the thermodynamic and kinetic mechanism involved in the folding/unfolding process of GS6 and, more in general,

of small peptides.

4.1 Methods

The initial structure of cyclo[(Lys-dTyr-Pro)₂] was created using Pymol version 0.99 (DeLano Scientific) on the basis of the experimental chemical structure.^{165,166} Geometry optimization of that structure was done by using the empirical potential energy function of the GROMOS96 43a1 force field.¹⁶⁸ The peptide, in its starting conformation, was solvated with water and placed in a periodic cubic box large enough to contain the peptide and 0.5 nm of solvent on all sides. The two lysine side chains were protonated as to reproduce a pH of about 7: two negative counterions (Cl⁻) were then added by replacing two water molecules to achieve a neutral condition. Molecular Dynamics simulations, in the NVT ensemble, with fixed bond lengths¹⁶⁹ were performed with the GROMACS software package¹⁷⁰ and with the GROMOS96 43a1 force field. Water was modeled by the simple point charge (SPC) model.¹⁷¹ A nonbond pairlist cutoff of 9.0 Å was used, and the long-range electrostatic interactions were treated with the particle mesh Ewald method.¹⁷² The isokinetic temperature coupling¹⁷³ was used to keep the temperature constant at the desired value. After various equilibration MD runs, six all-atom MD simulations in explicit water at six different temperatures and with different time lengths were carried out: 400 ns at 280 K, 300 ns at 310 K, 300 ns at 360 K, 180 ns at 400 K, 60 ns at 500 K, and 60 ns at 600 K. For the last two simulations, a time step of 1 fs was used, while for the others the time step was 2 fs.

4.2 Theory

4.2.1 Thermodynamic Characterization

Given a system in thermodynamic equilibrium, the change in free energy on going from a reference state, *ref*, of the system to a generic state, *i* (e.g., from unfolded to folded), at constant temperature and constant volume can be evaluated as:

$$\Delta A_{ref \rightarrow i} = -RT \ln \frac{p_i}{p_{ref}} \quad (4.1)$$

where R is the ideal gas constant, T is the absolute temperature and p_i and p_{ref} are the equilibrium probabilities of finding the system in state i and state ref , respectively. The conformational space and relative free energy as defined by the reaction coordinates given by the two distances between the oxygen and nitrogen atoms involved in the two hydrogen bonds (H-bonds). By projecting the MD trajectories onto the plane defined by these conformational coordinates (see Figure 4.1), three regions corresponding to three secondary structure states (conformational

states) were identified: the folded state (F) in which both H-bonds are formed; the intermediate state (I) in which only one of the two H-bonds is formed; and the unfolded state (U) in which none of the two H-bonds is formed.

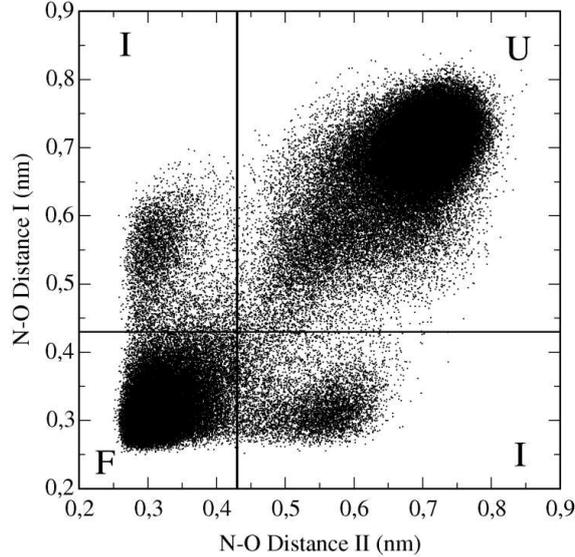


Figure 4.1: Projection of the trajectory at 310 K on the plane of the two NO (H-bond) distances between Lys residues. Three conformational states can be defined: the folded state (F) in which both H-bonds are formed, the intermediate state (I) in which only one of the two H-bonds is formed, and the unfolded state (U) in which none of the H-bonds is formed.

For every region, the corresponding MD frames were counted providing the equilibrium probabilities and hence, according to Eq. 4.1, the free energy changes $\Delta A_{ref \rightarrow i}$. Note that the variation of the Helmholtz free energy, due to the conformational transition of the peptide in the simulation box, exactly corresponds to the peptide chemical potential change (i.e., $\Delta A_{ref \rightarrow i} = \Delta \mu_{ref \rightarrow i}$). Such peptide chemical potential variations and the mean potential energies of the whole simulation box, taken for each conformational state for a set of temperatures (see section 4.1), were utilized to obtain, by a fitting procedure, a detailed model of the peptide partial molar thermodynamics according to the QGE theory as described in details in section 3.3. Such a QGE model provides the temperature dependence of the peptide chemical potential change and related partial molar properties according to:

$$\Delta \mu = \Delta u'_0 - \Delta c'_{v_0} T_0 \Lambda(T) + p'(T) \Delta v - RT \ln \gamma \quad (4.2)$$

where $\Delta u'_0$ and $\Delta c'_{v_0}$ are the variations of the partial molar excess internal energy

and isochoric heat capacity with respect to the reference state at the reference temperature T_0 (in the present study $T_0 = 310$ K), Δv is the partial molar volume change with respect to the reference state and $R \ln \gamma$ corresponds to a partial molar entropic term due to hard body effects. Moreover, p_0 is the pure solvent excess pressure and

$$\Lambda(T) = \frac{1}{\delta_0} + \frac{T}{T_0 \delta_0^2} \ln \left(1 - \frac{\delta_0 t_0}{T(1 - \delta_0) + \delta_0 T_0} \right) \quad (4.3)$$

with δ_0 a temperature independent dimensionless constant obtained by the pure solvent simulations. Note that $\Delta u'_0$, $\Delta c'_{v0}$, Δv and $R \ln \gamma$ are the parameters, corresponding to the physical properties fully defining the QGE model, which are obtained via the fitting procedure.

4.2.2 Kinetic characterization

To describe in a simple way the kinetics of the folding-unfolding process, the single conformational degree of freedom q can be considered as defined by the bisector of the plane identified by the two hydrogen bonds coordinates in Figure 4.1 by using the free energy profile along q , as obtained via:

$$\Delta A(q) = -RT \ln \frac{\rho(q)}{\rho(q_{ref})} \quad (4.4)$$

with $\rho(q)$ the equilibrium probability density, it can be readily obtained the complete kinetics of the folding/unfolding transitions by solving a Fokker-Plank type equation:¹⁷⁴

$$\frac{\partial \rho(q, t)}{\partial t} = \frac{D}{KT} \left[\rho \left(\frac{d^2 \Delta A}{dq^2} \right) + \left(\frac{\partial \rho(q, t)}{\partial q} \right) \left(\frac{d \Delta A}{dq} \right) \right] + D \left(\frac{\partial^2 \rho(q, t)}{\partial q^2} \right) \quad (4.5)$$

where $\rho(q, t)$ is the time-dependent probability density and D is the diffusion coefficient along q (in the present case at 310 K, $D = 1.66 \cdot 10^{-4} \text{ nm}^2/\text{ps}$). It must be remarked that such a model is sufficiently accurate only within the approximation of an instantaneous relaxation of all the other degrees of freedom during the diffusion along q . When some degrees of freedom other than q relax at a comparable rate of the diffusion along the chosen reaction coordinate, a proper kinetic model must include such degrees of freedom and their coupling. This can be accomplished either explicitly including all the relevant reaction coordinates in the Fokker-Plank type equation or implicitly considering the effects of the other degrees of freedom by monitoring the fluctuation time behavior of q as provided by MD simulations, hence still describing the kinetics via a single reaction coordinate.

In the present case, the trajectory of the reaction coordinate q (see Figure 4.2) reveals the presence of two distinct fluctuation regimes corresponding to two free

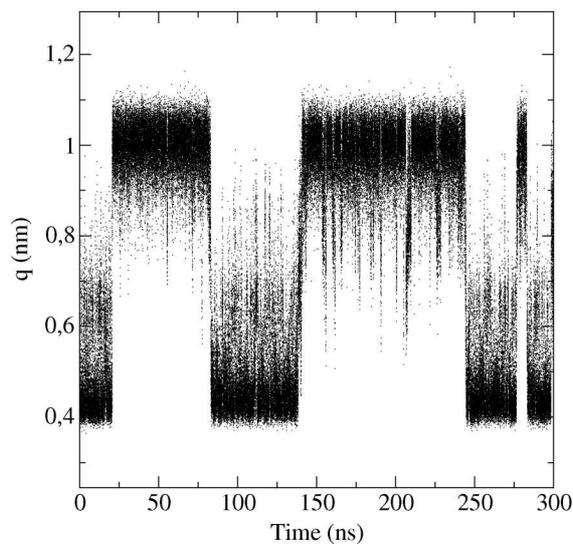


Figure 4.2: Time evolution of the reaction coordinate q during the MD run at 310 K.

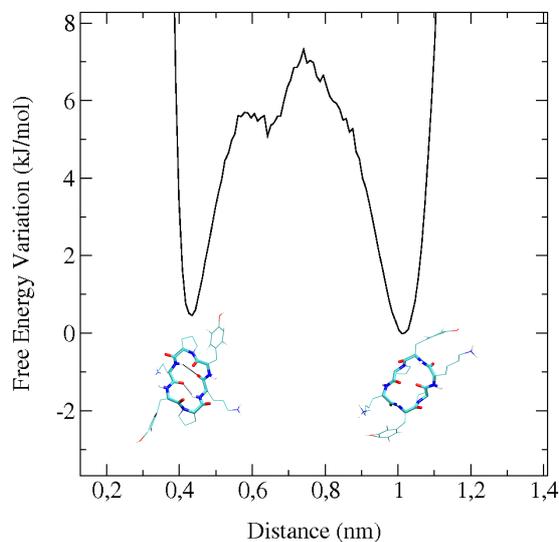


Figure 4.3: Free energy variation (at 310 K) along the reaction coordinate q , defined by the bisector of the plane given by the two NO (H-bond) distances between Lys residues. Two local minima can be observed corresponding to the folded (left) and the unfolded (right) states. Two representative snapshots of the folded and unfolded states extracted from the MD run at 310 K are also reported below the corresponding minimum. Hydrogen bonds are represented with dotted lines.

energy basins (see Figure 4.3) and characterized by different fluctuation distributions: one peaked at about 0.4 nm (corresponding to the folded state basin) and the other peaked at about 1.0 nm (corresponding to the unfolded state basin). It must be noted that the two identified distributions significantly overlap in the 0.7-0.9 nm range, hence indicating that GS6 folding/unfolding kinetics is characterized by the relaxation of a set of coupled degrees of freedom corresponding to the interconversion kinetics of the two q -fluctuation regimes.

In this study, the transition from one fluctuation regime to the other is defined as occurring when the reaction coordinate reaches the value corresponding to the peak of the other regime (e.g., starting from the folded state, when q reaches 1.0). In this way, the q trajectory can be easily subdivide into subparts corresponding to a single fluctuation regime and hence the mean lifetime for the folded to unfolded and reverse transition can be evaluated. In practice, the use of such a criterion for the MD simulation data of GS6 at 310 K, provides 4 and 3 blocks of the q trajectory for the folded and unfolded state, respectively (the lower and upper fluctuation blocks in Figure 4.2). The mean time interval of the folded/unfolded blocks (i.e., the average of the corresponding time intervals), gives hence the estimate of the unfolding/folding mean lifetime.

4.3 Results and Discussion

As mentioned in section 4.2, the plane defined by the two (H-bonds) N-O distances between Lys residues as conformational space was used to describe the relevant conformational state of GS6.

Using 0.43 nm as cut-off distance for the formation of each hydrogen bond (0.43 nm corresponds to the probability minimum of the N-O distance distribution, see Figure 4.4), the chosen conformational space may be divided into three regions representing different conformational states: the folded state (F), the intermediate state (I), and the unfolded state (U).

It is worth to note that the density peaks of the folded and unfolded states in Figure 4.4, corresponding to the free energy minima and relative structures along the bisector of the plane in Figure 4.1, clearly indicate that the β -hairpin structure is relatively stable in GS6 in agreement with experimental data.^{165,166} The Type II' β -turns are stabilized by the intraturn hydrogen bond, involving the Lys residues. Moreover, our MD simulations provide for the folded state the correct (i.e., experimentally observed) arrangement of the residues in each β -turn with D-Tyr and Pro residues in positions $i+1$ and $i+2$, respectively, and their side chains relative arrangement in line with the so-called equatorial-axial rule. In the unfolded state, the β -hairpin structure is completely disrupted with the main chain of the peptide distorted and proline and tyrosine residues no more in the proper arrangement of the β -turn. In Figure 4.5, the temperature dependence of the unfolding

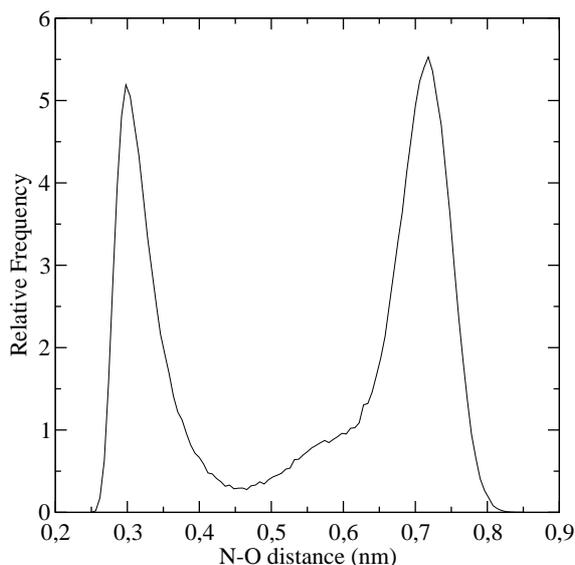


Figure 4.4: Distribution function (at 310 K) of the distance between the nitrogen and the oxygen atoms involved in the hydrogen bonds. The left peak corresponds to the folded structure while the right peak corresponds to the unfolded one.

chemical potential change and the isochoric internal energy change are reported, $\Delta u_v = (\partial\beta\Delta\mu/\partial\beta)_v$, as provided by the QGE model. In the Figure, the corresponding values as obtained by the conformational probabilities and mean potential energies given by the MD simulations are also reported.

To test severely the accuracy of the model prediction, in Figure 4.6 the isochoric excess heat capacity change due to the peptide insertion into the solvent as obtained by the QGE model with the corresponding values as provided by MD data were compared (the latter were evaluated by using the mean square fluctuations of the simulation box potential energy). Note that these MD-based values were not used within the parametrization procedure of the model. From these figures, it is clear the accuracy (within the noise) of the QGE model in describing and predicting the peptide thermodynamics as provided by the MD simulations.

Interestingly, the unfolding chemical potential change shows a negative broad maximum centered at 450 K indicating that the unfolded state is thermodynamically more stable in the whole temperature range considered with an increased stability for lower and higher temperatures. When considering the energetic and entropic contributions to the unfolding chemical potential change, as provided by the QGE model (see Figure 4.7), it clearly emerges that the unfolded state is characterized by a lower entropy and energy than the folded state, hence indicating a reduction of the peptide partial molar entropy coupled to an energetic optimization due to the unfolding process. Such a counterintuitive thermodynamic behavior, observed and

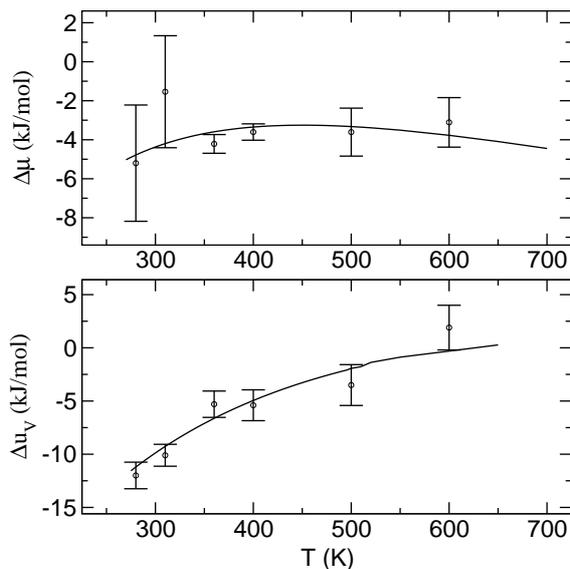


Figure 4.5: Plot of the unfolding chemical potential change (upper panel) and isochoric internal energy change (lower panel) as a function of temperature along the isochore. In the Figure, the circles represent the values obtained by the direct use of MD data with their relative error bar, while the solid line represents the corresponding QGE model prediction.

described in details in a recent article for a completely different peptide,¹⁶⁴ is driven by the charged and/or polar groups solvent exposure occurring upon unfolding and causing a relevant solvent rearrangement in the peptide first solvation shells leading to the peptide partial molar volume reduction (electrostriction).

To characterize in a simple way the kinetics of the folding/unfolding transitions, a single reaction coordinate q , defined by the bisector of the plane utilized to define the conformational space (see Figure 1), was considered. The use of a Fokker-Plank type equation to obtain the kinetic rate constants is appropriate only within the approximation that all the other degrees of freedom may be considered as instantaneously relaxed along the reaction coordinate transition. As evidenced in the theory section the reaction coordinate fluctuation as provided by the MD simulations (see Figure 2) are incompatible with such an assumption, as clearly shown by the significant overlapping of the distributions for the folded and unfolded q -fluctuation regimes.

According to the method briefly outlined in section 4.2, the transition from one fluctuation regime to the other (within our approximation, the folding/unfolding transition) is defined as occurring when the reaction coordinate reaches the value corresponding to the peak of the other regime. At 310 K, the unfolding mean lifetime results of about 31 ns and the folding mean lifetime of about 59 ns (10 ns and 27 ns

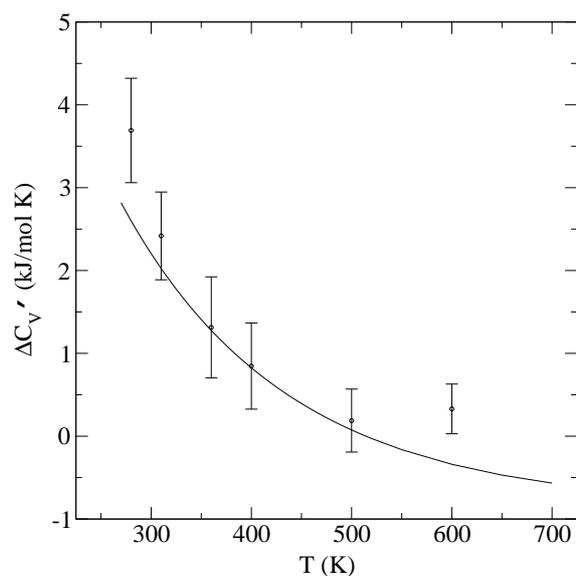


Figure 4.6: Plot of the excess isochoric heat capacity change due to the peptide insertion into the solvent. In the Figure, the circles represent the values obtained by using the mean square potential energy fluctuation of the simulation box with the relative error bars, while the solid line represents the corresponding QGE model prediction.

are the corresponding standard errors), in good agreement with the experimentally measured values at 324 K (71 ns and 100 ns for the unfolding and folding mean lifetime, respectively).¹⁶⁷ Interestingly, the use of the Fokker-Plank type equation for the single reaction coordinate considered, provides a significantly faster kinetics (1.2 ns and 1.4 ns for the unfolding and folding lifetimes) indeed confirming its inaccuracy due to the presence of kinetically coupled degrees of freedom involved in the folding/unfolding transitions.

The nature of the hidden degrees of freedom coupled to the reaction coordinate chosen is elusive, probably involving the peptide side chains and solvent molecules, and beyond the scope of the present study. However, given the interest on the folding/unfolding kinetic mechanism, the characterization of the coupled degrees of freedom involved in the folding/ unfolding transitions will be addressed in future works.

4.4 Conclusions

In the present study, the folding/unfolding process for the Gramicidin analogue GS6 has been extensively investigated by the use of MD simulations coupled to statistical mechanical models to characterize the thermodynamics and kinetics of the process.

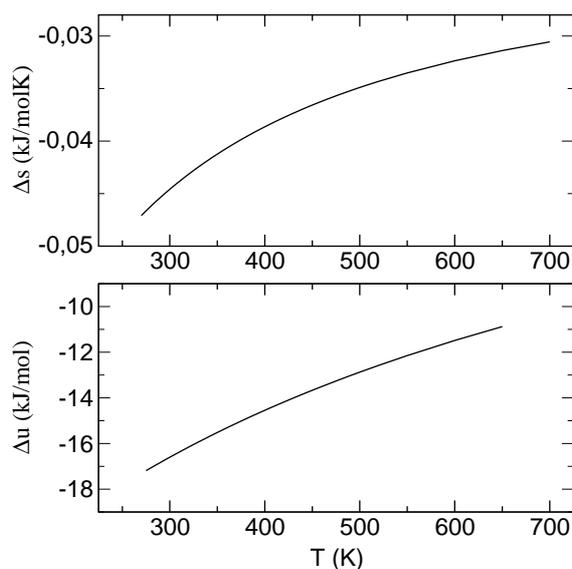


Figure 4.7: Plot of the unfolding partial molar entropy change (upper panel) and internal energy change (lower panel) of the peptide as a function of temperature along the isochore as provided by the QGE model.

Such an approach allowed to quantitatively obtain the peptide folding/unfolding partial molar thermodynamics and the corresponding kinetic rate constants. Results show that GS6 unfolded state is characterized by a lower entropy and energy than the folded one, as a consequence of the peptide partial molar volume decrease. Such a thermodynamic behavior driven by the charged and/or polar groups solvent exposure (electrostriction), is in line with previous data on a completely different small peptide.¹⁶⁴ It is worth to note that the emerging electrostriction-driven folding/unfolding thermodynamics as evidenced by the small peptides we studied, is likely to be a specific feature of solvated small peptides where the effects of intramolecular interactions and configurational freedom are relatively small compared to the peptide-solvent thermodynamic coupling. However, the present data confirm that the solvent exposure of charged and/or polar chemical groups typically results in electrostriction effects, hence possibly implying that electrostriction may play a significant role also in larger peptides and even in proteins. Characterization of the folding/unfolding kinetics provided clear indications that a set of coupled degrees of freedom is involved in the relaxation process, hence implying that a simple one-dimensional Fokker-Plank type equation cannot be used to model accurately the kinetics. The analysis of the reaction coordinate trajectory in terms of fluctuation regimes (i.e., fluctuation distributions) allowed a simple identification of folding and unfolding transitions, providing estimates of the corresponding mean lifetimes which match rather well the experimentally determined values.

RECONSTRUCTION OF INFRARED SPECTRA BY MEANS OF A QM/MD PROCEDURE

The understanding of protein-mediated processes at the atomic level requires the use of techniques monitoring protein structural changes. Infrared (IR) absorption spectra of amide modes have long provided a tool for determining the secondary structure of peptides and proteins due to the high sensitivity of amide vibrational frequencies and intensities (particularly for the amide I mode, mostly corresponding to the C=O stretching) on local atomic organization (*e.g.*, hydrogen bonds, solvation effects, hydrophobic interactions, etc.)^{175–177}. However, the complexity of IR signals of solvated peptides and proteins does not allow a detailed frequency-structure assignment to be determined experimentally.

In the last years many different theoretical-computational methods have been proposed to model protein and peptide IR spectroscopic behaviors, the development and parametrization of which is at present an active area of research. Rigorous *ab initio* methods for the determination of the vibrational frequencies do not allow to scale up to biologically relevant biomolecules and to properly include the complexity of the solute-solvent interplay, which would require the construction and diagonalization of the solute-solvent Hessian matrix of the quantum vibrational degrees of freedom at each representative liquid-state configuration. Therefore, hybrid approaches are commonly used to reproduce the band positions and line shapes of structurally well-defined molecules^{21, 178–191}.

Recent methods^{178, 179} combining electronic structure/molecular dynamics (MD) calculations make use of an empirical relation providing the instantaneous frequency of a solute as obtained by fitting *ab initio* vibrational frequencies of a solute-solvent cluster either to a linear combination of the electrostatic potentials¹⁷⁸ or to the components of the electric fields¹⁷⁹ acting on the solute atoms. Another approach^{180, 181} is based on reconstructing the vibrational behavior via purely classical atomic mo-

tions making use of the time autocorrelation function of the electric dipole. These two classes of methods are limited to the size of the solute and have been applied to rather small molecules such as trans-N-methylacetamide (trans-NMA),¹⁸⁰ a model system of the peptide group, or a tri-alanine peptide.¹⁸¹

For the study of peptides and small proteins, other methods have appeared in the literature^{21, 182–191}. Many of these^{182–189} employ quantum mechanical (QM) calculations to determine the vibrational frequencies and eigenstates for single amides which are, then, transferred to the full peptides and/or proteins and coupled empirically. An alternative approach^{21, 190, 191} makes use of Hessian calculations on the whole isolated peptide/protein in a given configuration which are, then, used to reconstruct the local, single-residue, vibrational frequencies via the Hessian matrix reconstruction method.¹⁹⁰ In these approaches coupling effects are included by adding a simplified electrostatic interaction, typically based on dipole-dipole interactions and commonly termed transition dipole coupling - TDC (coupling through space) and an empirical term providing the frequency variations due to first-neighbors relative rotations (coupling through chemical bonds). The solvatochromic effect, when included, is modeled via an empirical term relating the single-residue frequency to the perturbing electric potential exerted by the solvent, as obtained by calculations on solute/solvent clusters and MD simulation.

One of the most striking results emerging from a number of the above mentioned studies^{183, 186, 191} is that the often used empirical secondary-structure/frequency correlations are at best approximated and at worst misleading. The sub-bands from helices, sheets, turns, and loops are much broader and more overlapped than is commonly assumed.

A rather recent challenge, that goes beyond the application of the IR spectroscopy to structurally well-defined polypeptides, is its use to follow protein folding kinetics. Time-resolved IR spectroscopies have been applied to a number of α -helical^{192, 193} and β -hairpin peptides,^{167, 194} the basic secondary structural elements of proteins. A common spectroscopic feature has been observed for both α -helices and β -hairpins, *i.e.*, the amide I band typically shifts to lower frequency in the folded state with respect to the unfolded state, and therefore the absorption difference spectrum (unfolded–folded) shows a negative signal at ≈ 1620 – 1630 cm^{-1} and a positive signal at ≈ 1660 – 1670 cm^{-1} .

The observation that both α -helices and β -hairpins show the same negative-positive spectral feature in the difference spectrum suggests that this characteristic is not dependent on the type of secondary structure but, rather, more generally on differences in the folded and unfolded conformations. Which is, then, its origin? Can this spectral feature be safely used as a signature of the folding/unfolding process? These are crucial, still open, questions that we address in the present work.

In this study the amide I bands of the folded and unfolded states of two β -hairpin peptides, a 15-residue designed peptide termed peptide 1 and a 10-mer cyclic analog

of Gramicidin S, GS10, are evaluated using a theoretical-computational approach based on the Perturbed Matrix Method (PMM),^{134–138,195,196} a mixed QM/MD method whose main aim is to keep the configurational complexity of the system (peptide+solvent molecules) with a proper treatment of the quantum degrees of freedom of a portion of the system to be explicitly treated at electronic level (the backbone peptide group). Such a method has been already successfully applied to reproduce the IR spectrum of aqueous carbon monoxide,¹³⁶ of carbon monoxide within Myoglobin¹³⁷ and of liquid water.¹³⁸ The good agreement observed here between the theoretical and experimental spectra of the two peptides shows the basic correctness of the calculations. Therefore, the conclusions on the contribution of individual, or subsets of, peptide groups to the spectra are considered to be meaningful, allowing the structure-spectrum correlations in unfolded–folded amide I difference spectra to be understood at atomic detail.

5.1 Methods

5.1.1 Unperturbed quantum chemical calculations

As a model of the peptide group, *i.e.*, the quantum center to be explicitly treated at electronic level, trans-NMA was chosen (Figure 5.1). Quantum chemical calculations were carried out on the isolated trans-NMA molecule at the Time Dependent Density Functional Theory (TD-DFT) with the 6-31+G(d) basis set.

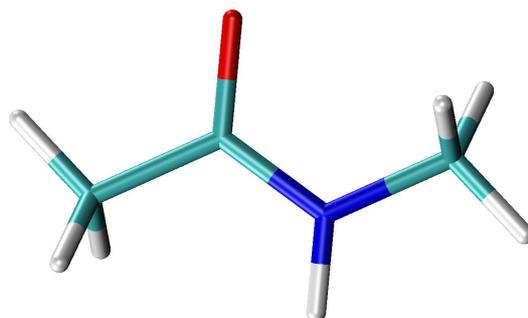


Figure 5.1: Structure of trans-N-methylacetamide

This level of theory was selected because it represents a good compromise between computational costs and accuracy. The mass-weighted Hessian matrix was calculated on the optimized geometry at the B3LYP/6-31+G(d) level of theory and subsequently diagonalized for obtaining the unperturbed eigenvectors and related

eigenvalues. The eigenvector corresponding in vacuo to the amide I mode was, then, used to generate a grid of points (*i.e.*, configurations) as follows: a step of 0.05 a.u. was adopted and the number of points was set to span an energy range of 20 kJ/mol (in the present case 31 points). For each point, six unperturbed electronic states were then evaluated at the same level of theory providing the basis set for the PMM calculations, *i.e.*, the Φ_k^0 eigenfunctions (see section 2.4.1).

5.1.2 Molecular dynamics simulations

A series of 50 ns-long atomistic MD simulations of peptide 1 and GS10 were performed in explicit solvent. For each peptide, three starting structures representing the unfolded state were extracted randomly from a simulation of 10 ns that was started from a fully extended configuration, for peptide 1, and from a high-Temperature generated structure for the cyclic GS10 peptide; three starting structures representing the folded state were taken from a simulation of 10 ns that was started from the NMR structure. The MD simulations were performed with the program GROMACS¹⁹⁷ and the GROMOS96 force field⁹⁵ was used for the peptide. The solvent used was D₂O, as to reproduce the experimental conditions, and was modeled using the deuterated spc water model¹⁹⁸. Each of the twelve starting configurations was placed in a dodecahedral water box large enough to contain the peptide and at least 1.0 nm of solvent on all sides at a water density of 55.32 mol/l. Periodic boundary conditions were used and the long range electrostatic interactions were treated with the Particle Mesh Ewald method¹⁹⁹. Coordinates were saved at every 1 ps. Simulations were performed in the NVT ensemble with the isokinetic temperature coupling²⁰⁰ to keeping the temperature constant at 300 K. Three positive (Na+) and two negative (Cl-) counter ions for peptide 1 and GS10, respectively, were added by replacing the corresponding number of water molecules so as to achieve a neutral simulation box. A 20 ns-long atomistic MD simulation of trans-NMA in aqueous solution was performed in explicit solvent (deuterated spc water model¹⁹⁸) using the same conditions and simulation protocol described above. For the trans-NMA equilibrium structure, that was kept fixed during the simulation, the DFT-based optimization as described above was used. Atomic charges were calculated using standard fitting procedures²⁰¹ on the optimized geometry at the B3LYP/6-31+G(d) level of theory.

5.1.3 Theory

In PMM calculations,^{134–136, 138, 195, 196} similarly to other QM/MM procedures,^{202–204} it is essential to pre-define a portion of the system to be treated at electronic level, hereafter termed as quantum center (QC), with the rest of the system described at a classical atomistic level exerting an electrostatic perturbation on the QC electronic states (see section 2.4). An orthonormal set of unperturbed electronic Hamiltonian

(\tilde{H}^0) eigenfunctions (Φ_k^0) are initially evaluated on the QC structure of interest. It could be then constructed for each QC-environment configuration (as generated by MD simulation) the perturbed electronic Hamiltonian (\tilde{H}) (see Eq. 2.30): the diagonalization of \tilde{H} provides a set of eigenvectors and eigenvalues representing the QC perturbed electronic eigenstates and energies.

Such a procedure, explained in details in section 2.4.2 and already used for determining perturbed vibrational states for biatomic QCs,^{136,137} has been recently shown to efficiently work for evaluating perturbed vibrational states also for polyatomic QCs.¹³⁸ The basic approximation of the method is that for typical quantum vibrational degrees of freedom the environment perturbation does not significantly alter the vibrational modes (*i.e.*, the mass-weighted QC Hessian eigenvectors) but only the related eigenvalues. In the present study, this procedure has been applied to study the amide I mode of trans-NMA, a model system for the peptide group, in D₂O solution. Subsequently, the same protocol has been adopted also for evaluating perturbed amide I frequencies for the peptide groups (residues) belonging to the N -residues β -hairpin peptides in D₂O solution. Trans-NMA fitted to the peptide group of each residue was used as the corresponding QC and hence its (unperturbed) mass-weighted Hessian eigenvectors provide the vibrational modes of each peptide group. The sidechain of the considered peptide group, the $N-1$ residues and the solvent define the perturbing environment at each configuration generated by MD simulation and the distribution of the oscillators perturbed-frequencies make up the total amide I band.

5.2 Results and Discussion

5.2.1 Infrared spectra of trans-NMA

The amide I experimental vibrational frequency of the isolated trans-NMA ranges from 1707 cm⁻¹ (in an argon matrix at 20 K²⁰⁵) to 1714-1731 cm⁻¹ (in the gas phase at ≈ 100 K²⁰⁶) depending on the experimental conditions. The computed unperturbed amide I frequency, evaluated here using standard geometry optimizations and frequency calculations at the harmonic approximation (see Methods section), is 1754 cm⁻¹, 47 cm⁻¹ higher than the experimental one (we take as the reference experimental value the one recorded in the argon matrix at 20 K.²⁰⁵) This shift is rather common for this kind of calculations^{205,207} and is usually attributed to various slight inaccuracies of the quantum calculations including the harmonic approximation, as indicated by a recent work.²⁰⁸

The experimental IR spectrum of trans-NMA in the amide I region in D₂O solvent at room temperature is presented in Figure 5.2 (the experimental data are taken from Song et al.²⁰⁹) The maximum of the amide I peak in solution is at about 1622-1623 cm⁻¹, *i.e.*, downshifted by $\approx 80-85$ cm⁻¹ with respect to the frequency

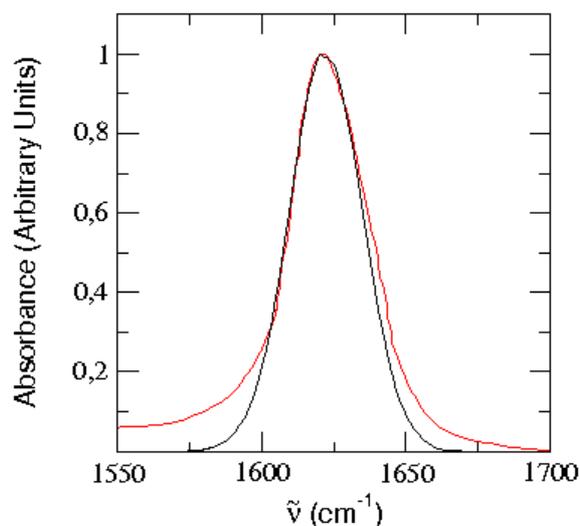


Figure 5.2: Comparison of the computed (black line) and experimental (red line) infrared spectra in the amide I region of trans-NMA in D₂O solution. For the sake of comparison the computed frequencies have been uniformly shifted to lower frequencies by 76 cm⁻¹ in order to align the computed to the experimental peak centered at 1622 cm⁻¹

of the isolated trans-NMA^{178,206} (unperturbed frequency), with a full width at half maximum (fwhm) of ≈ 30 cm⁻¹. The amide I band of trans-NMA in D₂O solution, computed using the PMM/MD procedure (see section 5.1.3 and 2.4.2), provides a spectrum with the maximum at 1698 cm⁻¹ corresponding to a downshift of 56 cm⁻¹ with respect to the unperturbed frequency and hence reproducing most of the large negative frequency shift experimentally observed. Further, the spectrum shape and width is very well reproduced by PMM/MD calculations (fwhm=30 cm⁻¹) as clearly shown in Figure 5.2.

It is worth noting that the missing 25-30 cm⁻¹ of the PMM/MD-derived shift with respect to the experimental one are due to a combination of higher order effects possibly including the slight inaccuracies of the calculated dipoles involved in the definition of the \tilde{Z}_1 matrix (Eqs. 2.30 and 2.31 in section 2.4.1) and the fact that we disregard any excitonic coupling and/or anharmonic behavior.

5.2.2 Infrared spectra of β -hairpins

The structures of the two β -hairpins studied here, peptide 1 (SESYINPDGTWTVTE) and GS10 (VKLYPVKLYP), are presented in Figure 5.3. The peculiar feature of these two peptides is that they show an opposite pattern of the aminoacids of the β -sheet with the hydrogen-bonded C=O groups pointing inwards (in) or with the C=O groups pointing towards the solvent (out), *i.e.*, for GS10 the “in” residues are

hydrophobic and the “out” residues are hydrophilic, while for peptide 1 the “in” residues are hydrophilic and the “out” residues are mainly hydrophobic.

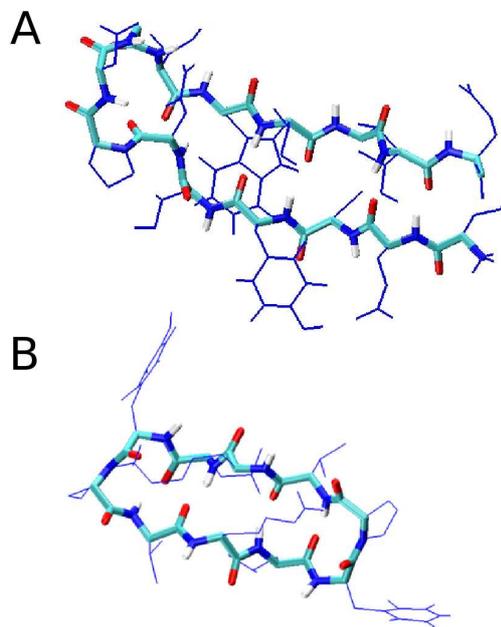


Figure 5.3: NMR-derived β -hairpin structures of peptide 1²¹⁰ (A) and GS10¹⁶⁵ (B).

In the experiments spectra at different Temperatures are collected to monitor folded and unfolded populations (within the usual assumption that the higher the Temperature, the higher the unfolded population). As the Temperature decreases (*i.e.*, folded state population increases), the amide I band shifts to lower frequency and therefore the unfolded–folded absorption difference spectrum shows a negative signal at $\approx 1632\text{ cm}^{-1}$ and a positive signal at $\approx 1664\text{ cm}^{-1}$, crossing zero at $\approx 1645\text{ cm}^{-1}$ (see Figure 5.4, panel A). Commonly, the contribution to these signals from different structural elements is assigned on the basis of empirical rules, *e.g.* the ≈ 1665 band component is typically associated with turn or loop structures and the negative-going feature centered at ≈ 1630 is considered characteristic of β -sheet structures^{167,210,211}. Nevertheless, the assignment is often not unique, *e.g.* several α -helical peptides show the same negative-positive feature in the difference spectrum centered at the same positions.^{192,193,212}

To shed light into the structure/spectrum correlations, the spectra of the folded and unfolded states are here calculated from simulation by applying the PMM/MD procedure to the folded and unfolded state ensembles, respectively.

The experimentally-derived and computed (via the PMM/MD procedure described in sections 2.4.2 and 5.1.3) difference spectra (unfolded–folded) of peptide 1

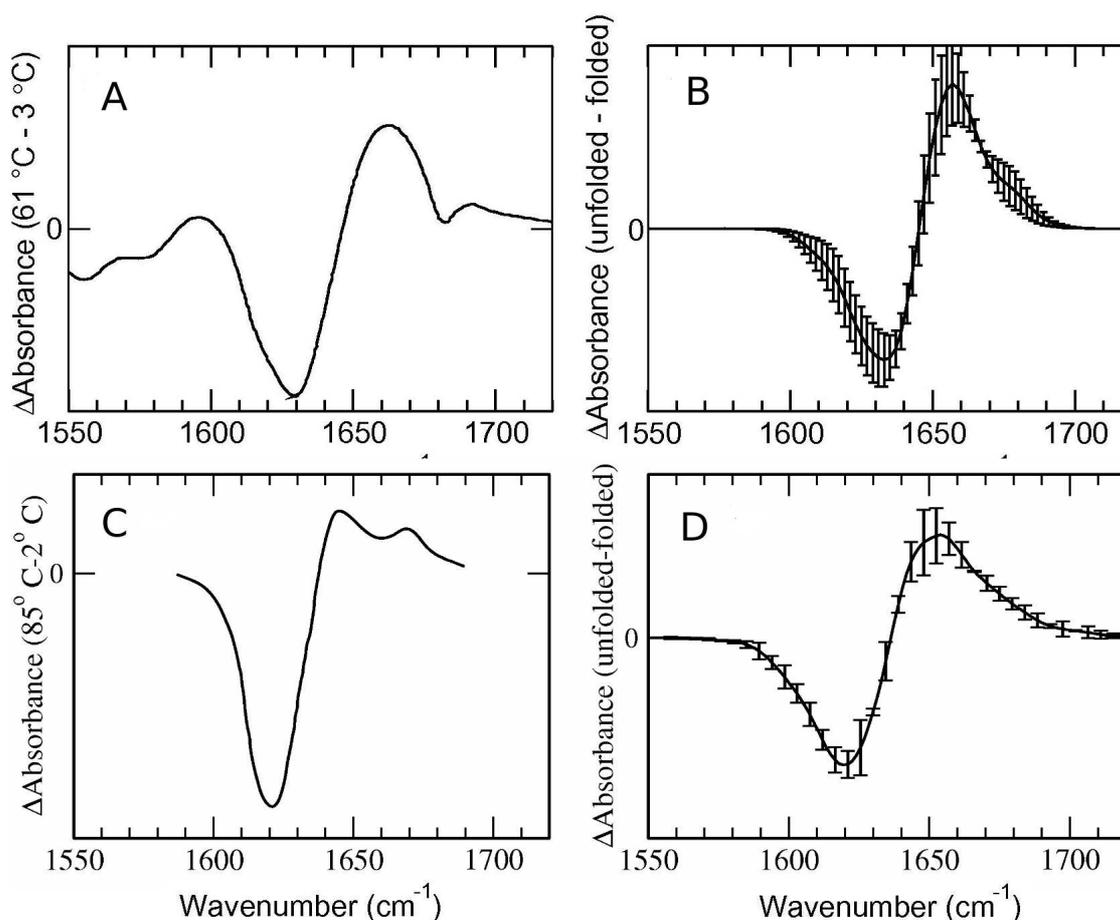


Figure 5.4: Comparison between experimental (panels A and C) and computed (panels B and D) difference spectra for peptide 1 (panels A and B) and GS10 (panels C and D). The experimental difference spectra were generated by subtracting the spectrum collected at 3.0 °C from the one collected at 61.0 °C (A) and by subtracting the spectrum collected at 2.0 °C from the one collected at 85.0 °C (C). The computed difference spectra (B and D) were generated by subtracting the spectrum of the folded state from the spectrum of the unfolded state. For the sake of comparison, the computed frequencies (panels B and D) have been uniformly shifted to lower frequencies by 73 cm^{-1} (B) and 87 cm^{-1} (D) in order to align the computed amide I peak to the experimental maximum. The absorbances are given in arbitrary units. The error bars correspond to a standard error of the corresponding property estimated over three independent sets of trajectories.

and GS10 are presented in Figure 5.4 and are in very good agreement with the experimental ones, showing the characteristic negative-positive signature. The computed frequencies of peptide 1 are upshifted by 73 cm^{-1} with respect to the experimental ones. This shift is consistent with the corresponding difference observed for trans-NMA in D_2O solution (76 cm^{-1} - see caption of Figure 5.2) and is, hence,

largely determined by the slight inaccuracies of the estimated unperturbed properties, as discussed in the previous subsection. The computed frequencies of GS10 are upshifted by 87 cm^{-1} with respect to the experimental ones.

One useful characteristic of our approach is that the amide I signal arising from any desired peptide group of the polypeptide chain may be isolated. In Figure 5.5, panel A the contributions from each peptide group (aminoacid) to the difference spectrum are shown together, along with the total signal. Interestingly, it can be seen that part of the residues shows the negative-positive signal, while others show the opposite trend.

The questions arise to which is the origin of the two different trends for different aminoacids and to why the total IR difference spectrum of peptides, both α -helices and β -hairpins, appeared to date^{167,192–194,212} shows the negative-positive signature. In what follows we investigate the role played by two factors: the position of the residue in the folded β -hairpin, *i.e.*, “in”, “out” or “turn” residue, and the nature of the side chain, *i.e.*, hydrophobic (phobic) or hydrophilic (philic).

The contributions to the difference spectrum arising from these different factors were evaluated by summing up the signals of the individual residues belonging to the given category. These are shown in Figure 5.5 panel B and Figure 5.5 panel C. It can be seen that the role played by the position in the folded structure is not unique in the two peptides, *i.e.*, for peptide 1 the “in” residues show a positive-negative signal and the “out” residues a negative-positive signal, while the opposite is true in the GS10 peptide. Instead, the nature of the side chain, regardless of its position in the folded structure, shows the same trend in both peptides, *i.e.*, all residues with a hydrophobic sidechain show a negative-positive signal giving rise to a pronounced negative-positive contribution to the total signal, while residues with a hydrophilic sidechain show both trends giving rise to a weak positive-negative overall contribution (Figure 5.5 panel C). Hence, the global negative-positive spectroscopic feature, typical of peptides, arises from the overcompensating contribution of the hydrophobic sidechains.

Why do the hydrophobic sidechains provide a negative-positive signal in the unfolded–folded difference spectrum, *i.e.*, why is the unfolded amide I peak shifted to higher frequencies with respect to the folded band? To answer this question, we study the effect of the electric field, \mathbf{E} , exerted by the environment on the different vibrating C=O dipoles. The component of the electric field along the vibrating C=O dipole, $E_{C=O}$, for a representative hydrophobic aminoacid in the unfolded and folded states is reported in Figure 5.6, panel A. It can be seen that for the hydrophobic sidechains $E_{C=O}$ is lower in the unfolded state, giving rise to the observed shift to higher frequencies of the amide I band with respect to the folded state (negative-positive signal in the unfolded–folded difference spectrum) (see scheme in Figure 5.6 panel B).

Such electric field variations mainly arise from the fact that the sidechain of

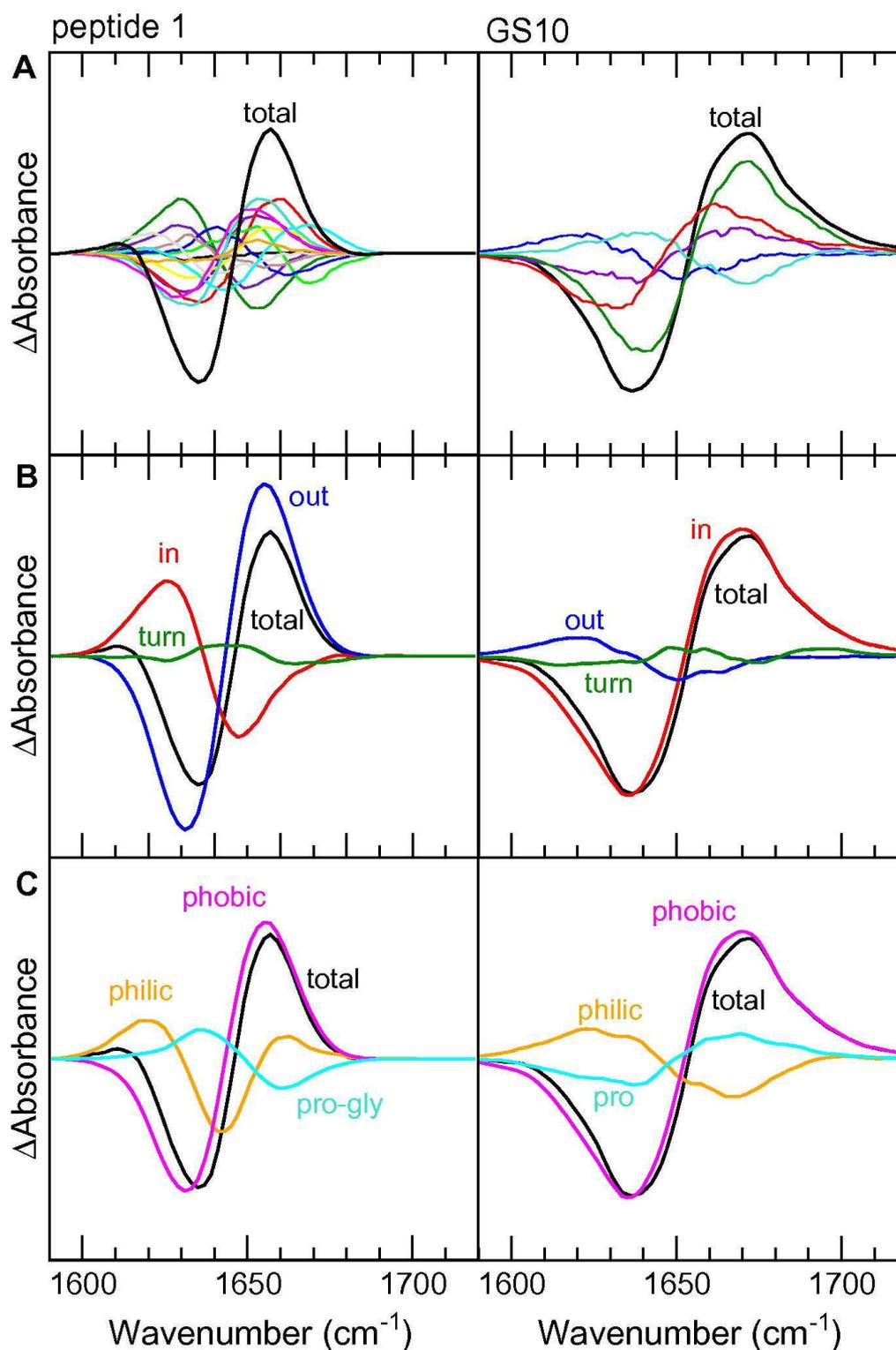


Figure 5.5: Contribution of the single peptide groups to the computed unfolded–folded amide I difference spectrum for peptide 1 (left) and GS10 (right). A) All the peptide groups are shown separately. B) The signals arising from “in”, “out” and “turn” peptide groups are grouped into three separated groups. C) The signals from “phobic”, “philic” and aminoacids not included into any category (pro and gly) are grouped into three separated groups. The absorbances are given in arbitrary units.

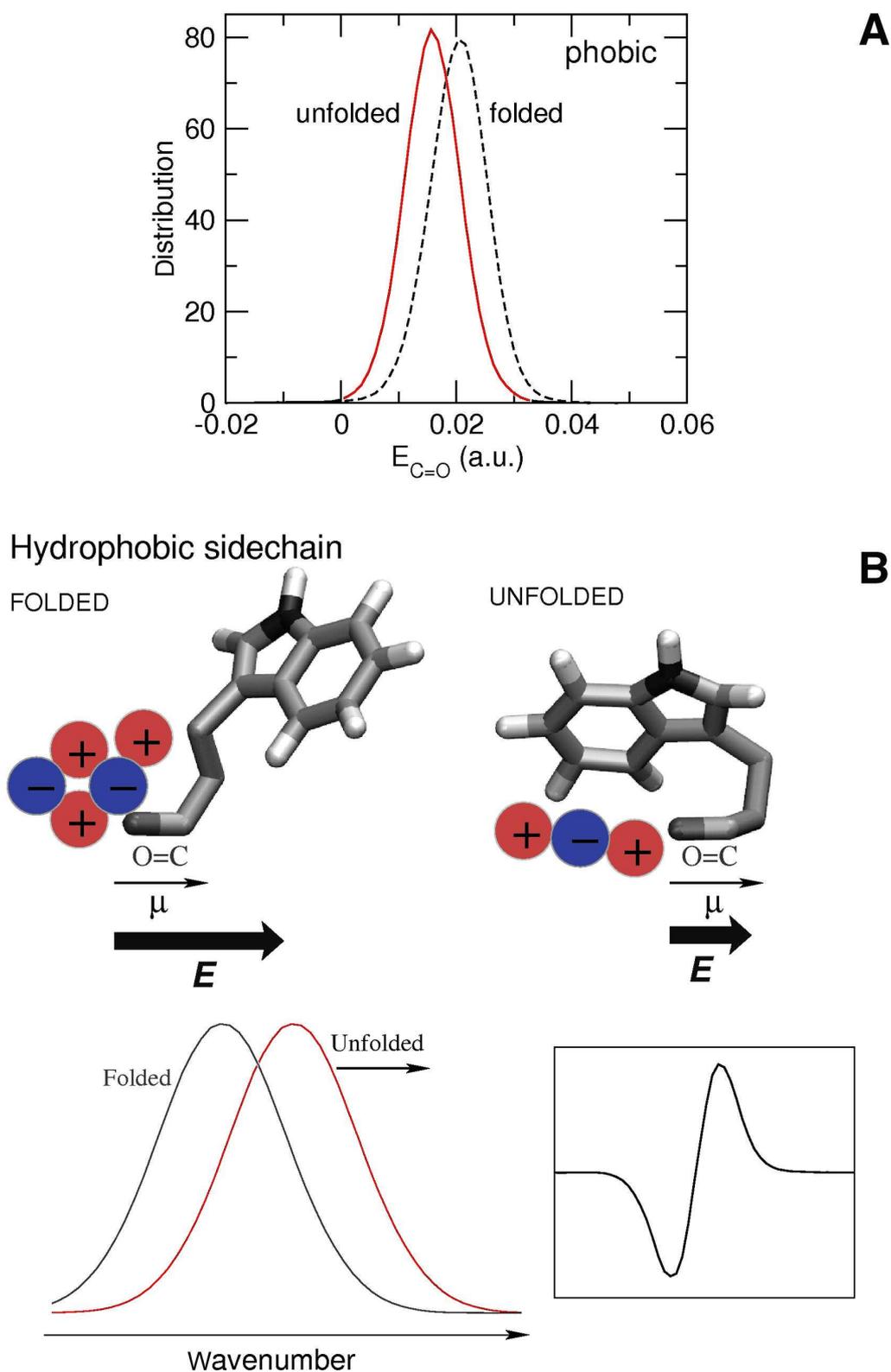


Figure 5.6: Effect of the hydrophobic sidechains on the amide I band. A) Component of the electric field along the vibrating C=O dipole of the peptide group, $E_{C=O}$, for a representative hydrophobic residue in the unfolded and folded states. B) Scheme summarizing the effect of the hydrophobic sidechain on the amide I mode vibrational frequency of the peptide group, to which the sidechain is attached, in the folded and unfolded states.

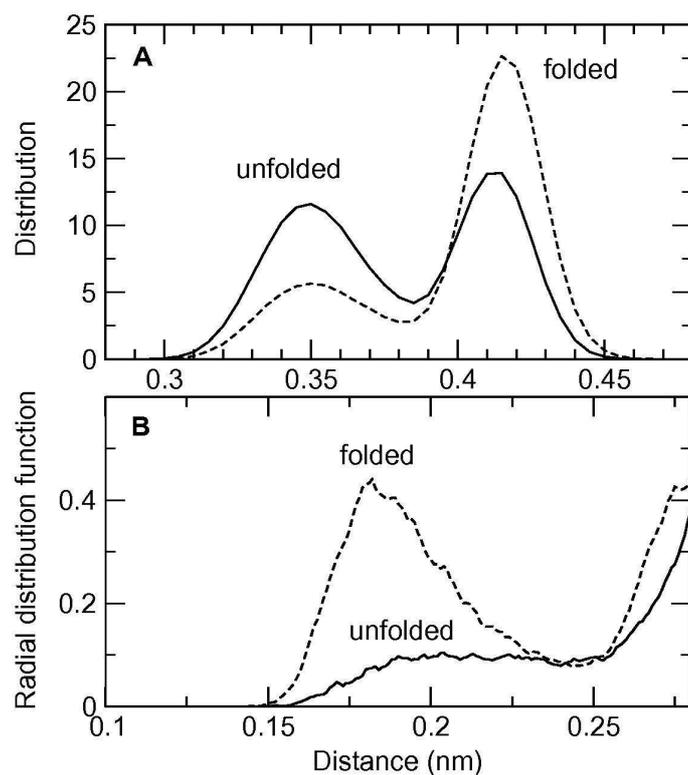


Figure 5.7: Distributions of the sidechain-carbonyl distance (distance between the corresponding centers of mass) (A) and radial distribution functions of the polar atoms around the peptide carbonyl group (B) for the unfolded (solid line) and folded (dashed line) states of a representative hydrophobic residue. It can be noted that in the unfolded state the population at shorter distances is increased, thus implying that the sidechain is closer to its peptide group in the unfolded state than in the folded one.

the hydrophobic aminoacids of the studied peptides is on average closer to its own peptide group in the unfolded state with respect to the folded configurations (see Figure 5.7).

In the unfolded configurations, the hydrophobic sidechains provide a less polar environment to the vibrating amide I mode (mainly C=O stretching) (Figure 5.6, panel B). Therefore, such decreased local electrostatic interactions lead to a shift of the unfolded peak toward higher frequencies, *i.e.*, towards the amide I band of the isolated peptide group (the trans-NMA in vacuo). The global negative-positive signal in the unfolded–folded amide I difference spectrum, being made up of all the peptide groups, is hence dominated by the effect of the hydrophobic sidechains (Figure 5.6, panel B).

5.3 Conclusions

In recent folding-kinetics experiments, changes in the IR difference spectrum (unfolded–folded) is being used to monitor time changes in unfolded/folded populations. Given the increasing amount of these experiments, an understanding at the atomic level of the structural features underlying the spectral differences is due.

Here, by means of theoretical-computational methods, the IR spectra in the amide I region of two β -hairpin peptides, peptide 1 and GS10, for which the experimental spectra are available were studied. The peptides show the common pattern observed in the unfolded–folded difference spectrum of peptides studied to date, *i.e.*, the amide I peak of the unfolded state is shifted toward higher frequencies with respect to the folded peak, leading to the characteristic negative-positive signal. The amide I bands are computed using a QM/MD methodology, the Perturbed Matrix Method (PMM), which provides results in good agreement with the experiments. Particular attention is given to the effect arising from the position of the different residues, *i.e.*, forming or not inter-strand hydrogen bonds in the folded state, and from the nature of the sidechains, *i.e.*, hydrophobic or hydrophilic.

The results show that the main determinant to the negative-positive signal in the unfolded–folded difference spectrum arises from the hydrophobic sidechains. In the unfolded configurations the hydrophobic sidechains are on average closer to their own peptide group with respect to the case of folded configurations, hence providing a less polar environment to the vibrating amide I mode (mainly C=O stretching). A less polar environment gives rise to a shift of the unfolded peak toward higher frequencies, *i.e.*, toward the amide I band of the isolated peptide group, with respect to the folded peak. This effect is seen for hydrophobic residues both involved and not in inter-strand hydrogen bonds, thus overwhelming possible effects arising from the different position of the aminoacids in the secondary structure of folded configurations.

The good quantitative reproduction of the experimental IR spectral changes upon folding of the studied peptides, as obtained by the PMM/MD procedure used in this paper disregarding any excitonic coupling (see the SI), indicates that the main features of the unfolded–folded difference spectrum may arise from peptide-solvent interactions without really requiring higher order effects. Therefore, the amide I excitonic coupling might be less relevant to IR spectral differences upon folding than it is often assumed^{176,213–215}.

These results show that a physically coherent procedure, not involving the use of empirical, adjustable parameters, may provide an efficient tool to characterize and interpret IR spectra of peptides and, possibly, proteins in solution in terms of their atomistic behavior. One possible outlook is a further development of the method to include higher order effects, such as anharmonic correction and excitonic coupling, in order to study and interpret 2D IR spectra which are recent tools used to probe

structures and dynamics in complex systems^{177,214,215}.

ANALYSIS OF IR SPECTRA OF GRAMICIDIN S ANALOGS

In this chapter, the study of IR spectra by means of theoretical-computational techniques started in chapter 5.1 is expanded by reconstructing the IR temperature dependent spectra of GS6 and GS10. GS6 has been already studied from a thermodynamic and kinetic point of view (see chapter 4), and the unfolded-folded spectrum of GS10 were presented in the previous chapter. Here, the analysis of IR spectra of β -hairpin peptides is deepened in order to shed light in the interpretation of experimental IR spectra.

β -hairpin peptides are particularly interesting model systems because, despite their small size (typically less than 20 residues), they exhibit various properties that are typical in globular proteins, e.g. they may contain a hydrophobic core and/or exhibit a cooperative thermal folding/unfolding transition.^{216,217}

Although experimental measurements of the folding kinetics of β -hairpins are scarce,^{87,89,167,210,216,218} in the past decade a remarkable number of theoretical and computational studies have been conducted to investigate hairpin structure, stability and folding transitions.^{217,219–227} Molecular dynamics (MD) simulations have proven to be a powerful tool for the investigation of the atomistic behavior of solvated β -hairpin peptides, including structural/conformational transitions^{228–230} and, more recently, even folding/unfolding kinetics and thermodynamics.^{163,231,232}

Fourier transform infrared (FTIR) spectroscopy has been widely used to study peptides and proteins secondary structures under equilibrium conditions^{19,233} (see chapter 5.1). Despite its diffusion to obtain structural information of numerous peptides and proteins in solution, this method has the limitation that the band shapes for different secondary structures may be similar to one another and highly congested,²¹ thus leading to spectra which are difficult to understand in terms of structures and conformational changes. IR spectra interpretation is typically based on empirical rules and widely accepted assumptions which general validity cannot be always demonstrated (e.g., IR spectra at different temperatures are collected in

order to monitor the unfolding process within the assumption that the higher the temperature the higher the unfolded population).

Here, the IR spectra of two β -hairpin analogs of Gramicidin S, for which experimental spectra were available (see chapter 4, are computationally reconstructed by means of a theoretical-computational approach based on the Perturbed Matrix Method (PMM),^{134,135,195,196} in order to clarify the relationship between conformational and spectroscopic features and to shed light on folding/unfolding transitions. The structure of the 6-meric (GS6) and a 10-meric (GS10) analogs of Gramicidin S are presented in Figure 6.1. The results obtained, in combination with previous theoretical-computational data on folding/unfolding thermodynamics and kinetics of GS6 (see chapter 4),²³² are finally compared to the available experimental data obtained by means of temperature dependent FTIR spectroscopy.¹⁶⁷

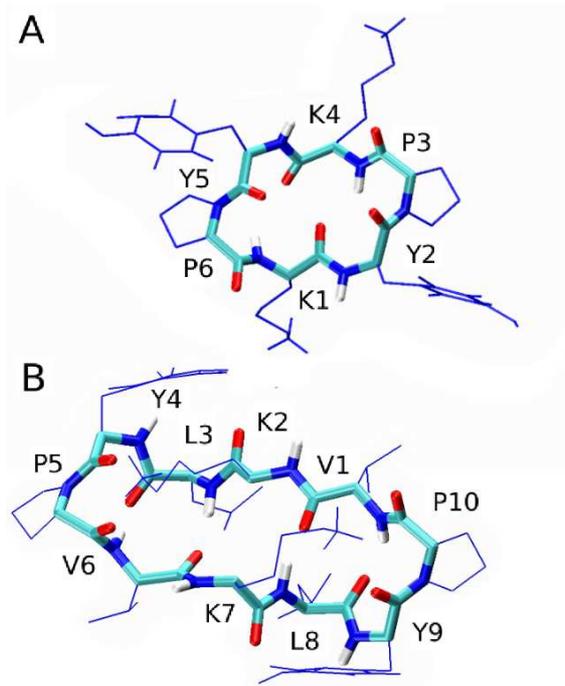


Figure 6.1: Snapshots of the β -hairpin structure of GS6 (panel A) and GS10 (panel B) extracted from MD runs. Hydrogen bonds are represented with dotted lines.

6.1 Methods

For the following analysis, the same trajectories used in chapter 4, obtained as explained in section 4.1, are used for GS6. The same simulation protocol was used to obtain for GS10 the following trajectories:

200 ns at 280 K, 900 ns at 330 K, 980 ns at 360 K, 490 ns at 400 K.

The folded and unfolded conformation for the two peptides are defined via the hydrogen bonds (H-bonds) that characterize the hairpin. In the case of GS6 there are two H-bonds and, as previously described in details (see section 4.2), the secondary structure states are defined by projecting the MD trajectories onto the plane defined by the two distances between the oxygen and nitrogen atoms involved in the two H-bonds. The folded and unfolded conformational states are identified: in the folded state both H-bonds are formed and in the unfolded state none of the two H-bonds is formed.

In the case of GS10 there are four hydrogen bonds. Coherently with the definition used for the smaller peptide, the folded and unfolded conformations are defined by the projection of the MD trajectories along the eigenvector corresponding to the largest eigenvalue obtained by diagonalizing the covariance matrix of the distances between the oxygen and nitrogen atoms involved in the four H-bonds. Such an eigenvector defines a conformational coordinate (q) providing the largest concerted H-bonds fluctuations and hence well describing the peptide folding/unfolding transitions. In Figure 6.2 it is shown the time course of the q conformational coordinate as obtained from the MD trajectory at 330 K, clearly indicating the folded and unfolded states.

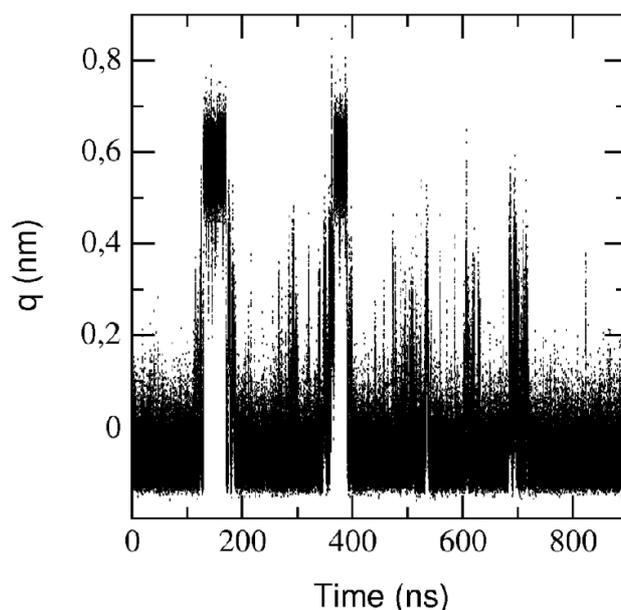


Figure 6.2: Time evolution of the projection of the MD trajectory at 330 K on the first eigenvector of the covariance matrix of the distances between the oxygen and nitrogen atoms involved in the four H-bonds of GS10.

6.2 Results and Discussion

In chapter 4 the thermodynamics and kinetics of GS6 has been quantitatively characterized by means of MD simulations and a statistical mechanical model based on the Quasi Gaussian Entropy (QGE) theory.^{162,234–236} The results obtained can be summarized as follows: *i*) the unfolded state is thermodynamically more stable than the folded one in the whole temperature range considered (280 K-600 K) with an increased stability at lower and higher temperatures; *ii*) the unfolded state is characterized by a lower entropy and energy than the folded one; *iii*) the unfolding and folding mean life-times were of the order of tens of nanoseconds.

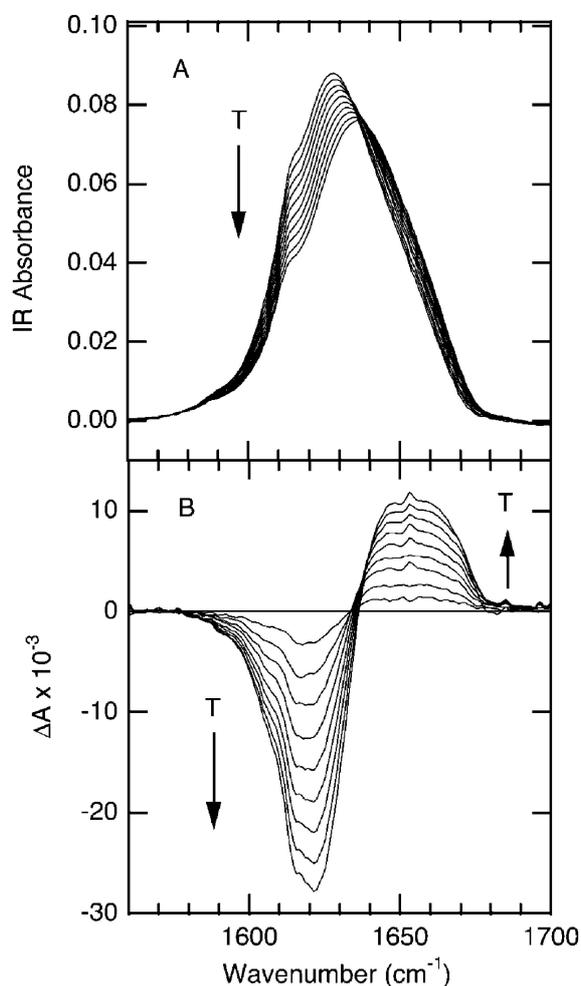


Figure 6.3: Temperature-dependent FTIR spectra for the 6-meric cyclic peptide GS6 obtained by Maness et al.¹⁶⁷ (A) Absorbance spectra in the amide I region; the temperature of the individual traces range from 2°C to 85°C. (B) Difference spectra obtained by subtracting the spectrum at 2°C from the spectra at higher temperatures.

The results concerning the structural features of the folded conformation of the peptide, the order of magnitude of the folding/unfolding chemical potential change and the folding/unfolding kinetic rates were in good agreement with the available experimental data.^{165,167} However, the calculated folding/unfolding entropy and energy change, according to *ii*), showed an opposite sign with respect to the corresponding experimental estimates as derived by IR temperature dependent spectra, thus leading to increasing folded state population as temperature increases within the experimental temperature range (280 K-360 K). Starting from these observations, the IR temperature dependent spectra are here reconstructed from the simulated trajectories and compared with the experimental ones, in order to clarify both the thermodynamic data and the experimental spectra.

The IR spectrum of GS6 obtained by Maness et al.¹⁶⁷ shows a peak centered at $\approx 1630 \text{ cm}^{-1}$, corresponding to the amide I band. This peak is experimentally studied at different temperatures (from 2 °C to 85 °C) by means of equilibrium FTIR spectroscopy in order to estimate the folded and unfolded state thermodynamics, assuming that at the extreme temperatures the population is either fully folded (2 °C) or fully unfolded (85 °C), in contrast to the theoretical-computational results. As the temperature increases, the peak shifts to higher frequencies and its intensity decreases: consequently, the difference spectra obtained by subtracting the spectrum at the lowest temperature from the ones at higher temperatures exhibit a negative-positive trend along the frequency axis. The experimental spectra obtained by Maness et al. are reported in Figure 6.3.

The corresponding computed spectra obtained at 6 different temperatures (280 K, 310 K, 360 K, 400 K, 500 K and 600 K) are presented in Figure 6.4 (panel A). In the same Figure (panel B) the temperature difference spectra obtained by subtracting from the spectra at 360 K and 310 K the spectrum at 280 K is also shown. Comparison of this Figure with the corresponding experimental data,¹⁶⁷ clearly indicates that the PMM/MD procedure well reproduces the experimental results, properly providing the spectrum shape (fwhm, $\approx 35 \text{ cm}^{-1}$ in the computed spectrum vs $\approx 40 \text{ cm}^{-1}$ in the experimental one) and the temperature trend (see panel B). Note that the spectra in Figure 6.4 have been uniformly shifted to lower frequencies by $\approx 95 \text{ cm}^{-1}$ in order to align the computed amide I band with the experimental maximum. Such a shift, largely due to the inaccuracies of the *ab-initio* calculations, has been applied to all the computed GS6 IR signals reported in this paper.

The spectra at different temperatures are experimentally studied to monitor folded and unfolded populations in order to elucidate the folding/unfolding kinetics and thermodynamics of β -hairpins, within the usual assumption that the unfolded population raises with the temperature. The PMM/MD procedure is then applied to the folded and unfolded ensembles as derived from the analysis of the MD trajectories (see Methods) and the spectra of the folded and unfolded states are re-

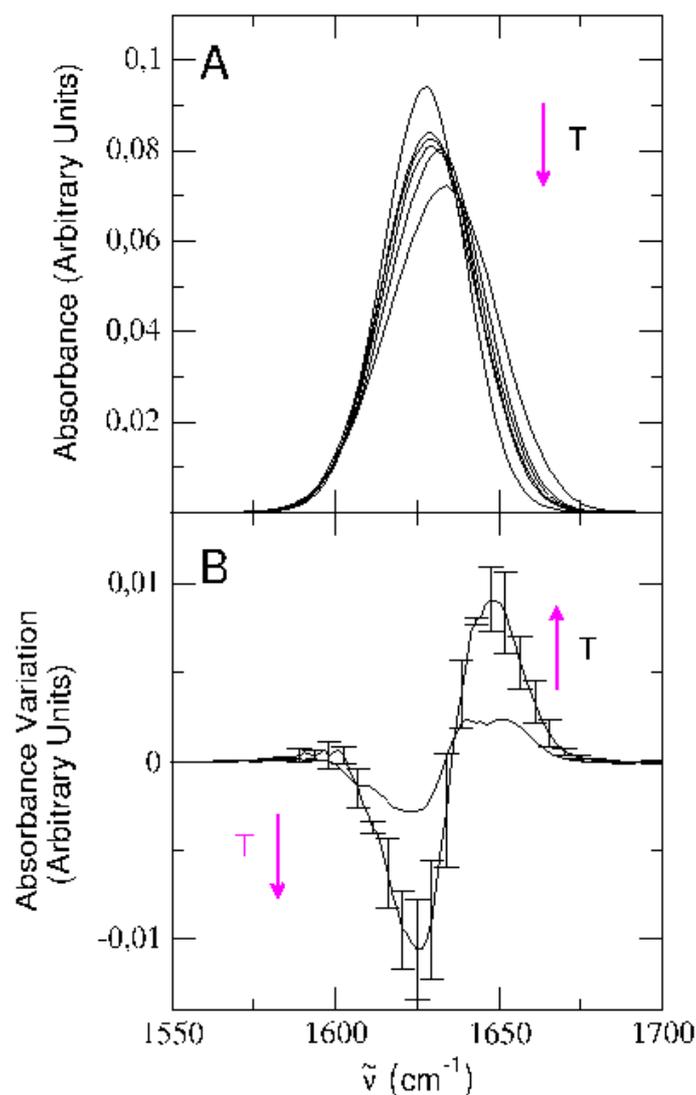


Figure 6.4: Panel A: Computed IR spectra in the amide I region of GS6 in D₂O solution at 280 K, 310 K, 360 K, 400 K, 500 K and 600 K. Panel B: Difference spectra obtained by subtracting the spectrum calculated at 280 K from the spectra calculated at 310 K and 360 K. For the last one error bars are reported corresponding to the standard error evaluated over three independent trajectories.

constructed. In Figure 6.5 (panel A) the computed GS6 unfolded–folded difference spectra evaluated at 280 K, 310 K and 360 K are reported (it is worth to note that such difference spectra are, within the noise, essentially indistinguishable). These spectra show a positive-negative trend as the frequency increases, that is an opposite trend with respect to the experimental and computed temperature difference spectra. This is in agreement with the thermodynamic results²³² which showed that the folded state population increases as the temperature raises in the range between 280 K and 360 K (according to *i*) and consequently the temperature difference spectra

should provide an opposite trend with respect to the unfolded-folded difference one. The spectroscopic effect of the increasing folded state population in temperature is shown in panel B of Figure 6.5, where it is reported the folded and unfolded state contribution to the IR spectrum at each of the three considered temperatures.

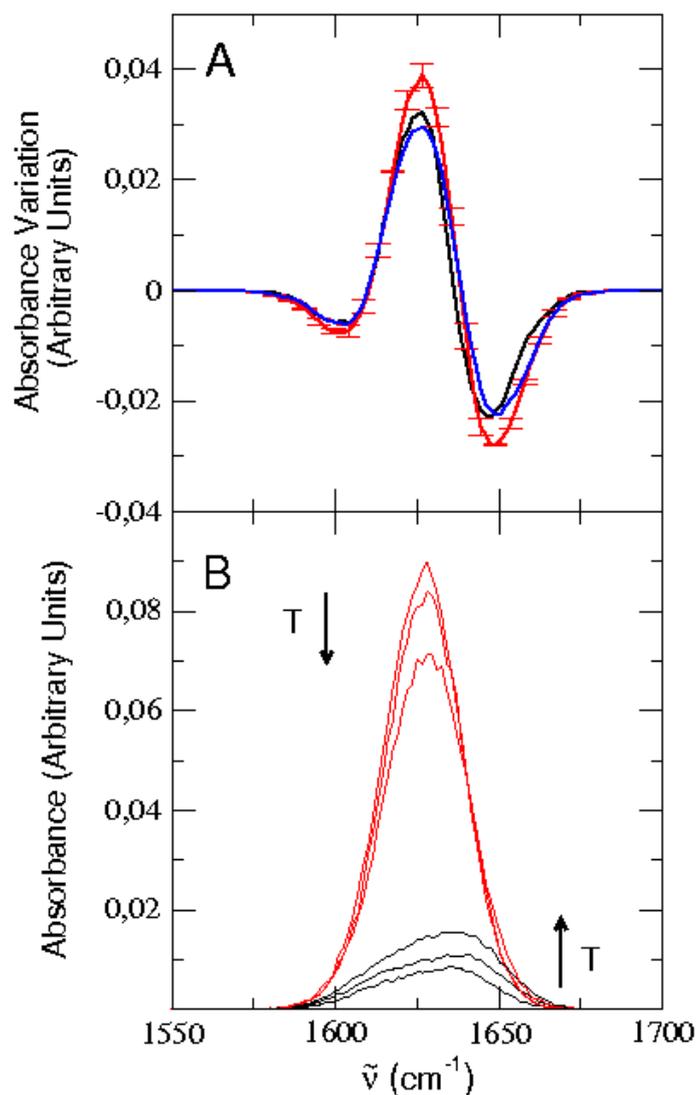


Figure 6.5: Panel A: Difference spectra of GS6 obtained by subtracting the spectrum of the folded state from the spectrum of the unfolded state at 280 K (black line); 310 K (red line) and 360 K (blue line). Error bars are reported corresponding to the standard error evaluated over three independent trajectories for the spectra at 310 K. Panel B: Contribution to the IR spectra of the folded state (black lines) and unfolded state (red lines) at 280 K, 310 K, and 360 K.

The temperature dependent spectrum of GS10 is also reconstructed: in Figure 6.6 (panel A) the spectra obtained at 280 K, 330 K, 360 K, 400 K are reported. There is a good agreement between the experimental and the calculated spectra, both in the temperature trend (see panel B and the corresponding experimental data¹⁶⁷) reported in Figure 5.4 and in the full width at half maximum ($\approx 40 \text{ cm}^{-1}$ in the experimental spectrum and $\approx 44 \text{ cm}^{-1}$ in the calculated one).

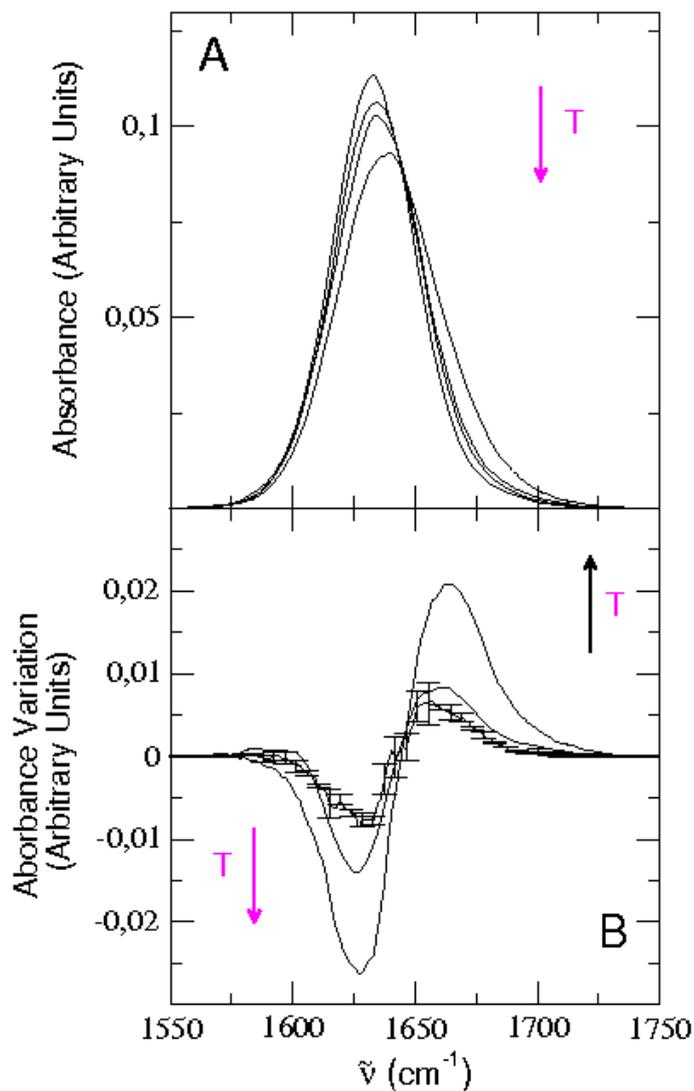


Figure 6.6: Panel A: IR spectra in the amide I region of GS10 in D_2O solution at 280 K, 330 K, 360 K and 400 K. Panel B: Difference spectra obtained by subtracting the spectrum calculated at 280 K from the spectra calculated at 330 K, 360 K and 400 K. For the first one error bars are reported corresponding to the standard error evaluated over three independent trajectories.

In this case too, the computed spectra have been shifted to lower frequencies by $\approx 87 \text{ cm}^{-1}$ in order to align the obtained amide I band with the experimental maximum.

From the Figure, it can be seen that in this case both the temperature difference spectrum (panel B) and the unfolded–folded difference spectrum (inset, already shown in Figure 5.4), show a common negative-positive signal, thus implying that for this peptide an increasing unfolded state population is present as temperature increases.

In order to better understand the results concerning GS6, the contribution to the total spectrum of each residue belonging to that peptide is examined. The total trend results from the contribution of both negative-positive and positive-negative signals. In Figure 6.7, the contribution to the total difference spectrum of hydrophobic, hydrophilic and neutral residues is reported for GS6 and, for the sake of comparison, also for GS10 (data already shown in Figure 5.5). In the case of GS10 (Figure 6.7, panel A), as previously discussed, the hydrophobic side chains are characterized by a negative-positive signal that is much more intense with respect to the one of neutral and hydrophilic residues.

In the case of GS6, no hydrophobic residues are present and the positive-negative trend of the unfolded–folded difference spectrum results from the contribution of both positive-negative and negative-positive signals (Figure 6.7, panel B). The absence of hydrophobic residues in GS6 peptide seems hence essential to determine the inversion of the spectral trend. Furthermore, the unusual thermodynamic behavior of GS6, being driven by electrostriction (see section 4.3, *i.e.* the charged and/or polar groups solvent exposure during unfolding leads to the rearrangement of the solvent molecules, is likely to be determined by the presence of hydrophilic side chains only in the peptide.

The apparent discrepancy between some calculated and experimental data for GS6 has been clarified, residing in the peculiar thermodynamic behavior of the peptide which doesn't allow nor the common assumption that the higher the temperature, the higher the unfolded population, neither the usual empirical rules that assign a spectral contribution to a particular structural element.

6.3 Conclusions

IR temperature-dependent spectra have been widely used in order to characterize folding/unfolding transitions in peptides and proteins and, more recently, also to characterize folding/unfolding kinetics. In this paper, the experimental temperature dependent spectra of two small β -hairpin peptides, GS6 and GS10, are reconstructed by means of all-atoms MD simulations and a theoretical-computational approach based on the PMM method. The temperature difference spectra are analyzed as well as the unfolded–folded difference spectra. The temperature difference spectra

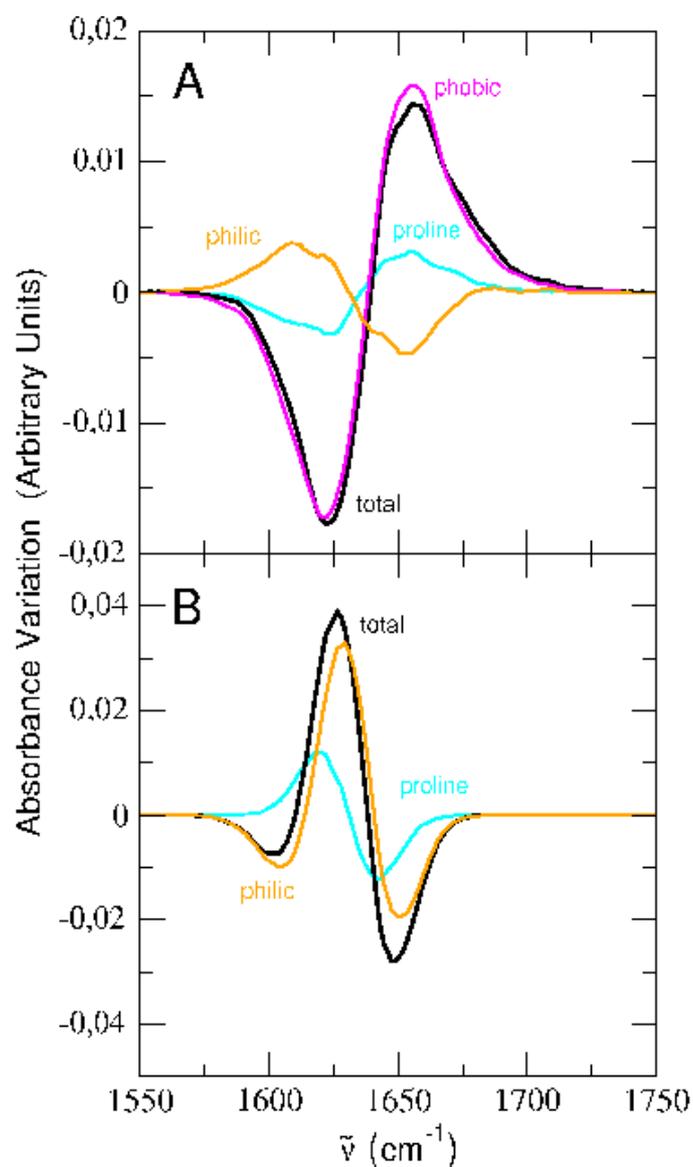


Figure 6.7: Contribution of hydrophilic, hydrophobic and proline to the total unfolded-folded difference spectrum of GS10 (panel A) and GS6 (panel B).

are in good agreement with the experimental ones, showing the usual negative-positive trend. The unfolded-folded difference spectra, on the contrary, result to be different for the two peptides. For GS10 the unfolded-folded difference spectrum shows the same features of the temperature difference spectra, hence indicating the usual unfolded population increase as temperature raises. In the case of GS6 an opposite trend, that is a positive-negative signal, is observed for the unfolded-folded difference spectrum. This can be attributed to the absence of hydrophobic side chains which constitute the main contribution to the IR unfolded-folded difference spectrum, showing an intense characteristic negative-positive signal. The inversion

between the temperature difference spectrum and the unfolded-folded one, indicates that in the considered temperature range the folded state population increases as the temperature is raising from 280 K to 360 K, confirming the previous data on GS6 thermodynamics²³² as obtained by means of the statistical mechanical model based on the QGE theory (see chapter 4). Consequently, for GS6 peptide it is not possible to interpret the experimental temperature difference spectra obtained by subtracting the spectrum at the lowest temperature from the spectra at higher temperatures, as representative of the unfolding process.

The use of PMM/MD procedures in order to reconstruct IR spectra, providing the possibility to separate the contribution of each side chain to the total spectrum, highlights the importance of the hydrophilic or hydrophobic nature of residues in determining the IR signal of peptides. It can be then a useful tool to interpret experimental IR spectra and also to better understand the folding/unfolding process.

STRUCTURE OF THE LIPODEPSIPEPTIDE SYRINGOMYCIN E

Syringomycin is the most studied member of the family Lipodepsipeptide compounds that characterize the secondary metabolism of the plant associated bacterial species *Pseudomonas syringae* pv. *syringae*. It is a nine amino acids containing lactone ring, with the N-terminal residue N-acylated by a 3-hydroxyacyl moiety with 10, 12 or 14 carbon atoms in the three homologous forms, called SRA1, SRE and SRG respectively^{237–239}. The lipopeptide is cyclised through an ester linkage between the side chain hydroxyl of 1-Serine and the backbone carboxyl group of C-terminal residue, 4-chlorothreonine (Figure 7.1). These structural features are shared by other lipodepsinonapeptides (LDNPs) produced by various strains of *P. s. pv. syringae*: syringotoxin (ST)²⁴⁰, syringostatins²³⁸, pseudomycin²⁴¹, where the N-terminal residue, Ser, and the C-terminal tripeptide 3-(OH)Asp-Dhb-4 -(Cl)Thr, are conserved.

Syringomycin was shown to play a role in plant-microbe interactions by enhancing the bacterial virulence²⁴² and it also displays a phytotoxic activity in vitro and a prominent fungicidal activity.²⁴³ Moreover, the activity of LDNPs against human pathogens, as well as against causal agents of life threatening infections, stimulated the interest for possible medical applications of these compounds.^{244,245} However, the primary target of SRE in plant, microbial and animal cells is the plasma membrane and the toxic activity is due to the formation of pores acting as non selective ion channels; the resulting membrane depolarization and non controlled ion flux leads to cell death.²⁴⁶

The interesting and not fully exploited activities of SRE, led us to investigate its conformation in the membrane-mimicking environments, in order to gain an insight into the structural determinants of the membrane interaction. In perspective, such information could be useful in designing structural analogues with the aim to enhance the selectivity towards membranes of different compositions. Previously a structural investigation of SRE based on NMR spectroscopy has been reported.²⁴⁷

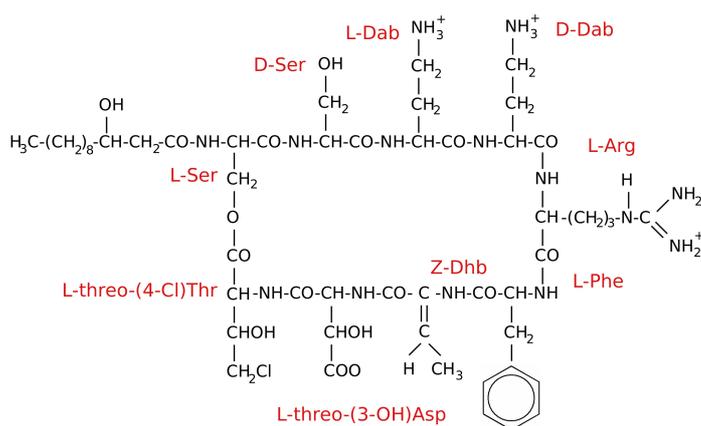


Figure 7.1: The chemical formula of SRE. Chiralities are as follows: 1-(L)Ser; 2-(D)Ser; 3(D)Dab; 4(L)Dab; 5(L)Arg; 6(L)Phe; 7Dhb, 8OH-(L)Asp, 9Cl-(L)Thr

The authors determined the structure in water solution by NMR and that structure was then translated into an octane medium by computer molecular dynamics (MD) simulations (octane is used to mimic the membrane environment).

In this study, ^1NMR spectra of SRE bound to negatively charged SDS micelles have been recorded and MD simulations have been performed in octane solution. A subset of NOEs were used and the three-dimensional structure of SRE in the membrane-like environment was obtained. The following discussion will focus on the computational part of the study, citing the relevant experimental results in the results and discussion sections (7.2 and 7.3).

7.1 Methods

The initial structure of SR-E was created using Pymol version 0.99 (DeLano Scientific), on the basis of the experimental chemical structure and stereochemistry.^{238, 239} Geometry optimization of that structure was done by using the empirical potential energy function of the GROMOS96 force field¹⁶⁸ that contains terms representing covalent bond stretching, bond angle bending, harmonic dihedral angle bending, sinusoidal dihedral torsion, van der Waals and electrostatic interactions. As SRE contains some residues that were undefined in the GROMOS96 force field, the corresponding parameters of the potential energy function were taken from similar functional groups. To obtain the structure in an apolar solvent and in analogy with previously reported simulations,²⁴⁷ the molecule was solvated in a periodic cubic box of dimensions 3.41x3.41x3.41 nm with 147 molecules of octane in order to reproduce the density of liquid octane at 20°C (702.52 g/l). The system was then equilibrated

Atom Pair	Intensity	$r_{i,j}$ (nm)
NH(Ser 2)-CH2(Tail)	weak	0.55
NH(Dab 3)-CH(OH-Asp 8)	strong	0.37
CH(Dab 3)-NH(OH-Asp 8)	weak	0.55
NH(Dab 4)-NH ₃ ⁺ (Dab 4)	weak	0.50
NH(Arg 5)-NH ₃ ⁺ (Dab 4)	weak	0.50

Table 7.1: Distance restraints (r_{ij}) applied between each atom pair during the equilibration procedure with the corresponding intensity of the NOE signal

and a first MD simulation at a temperature of 300 K was performed for 500 ps. All MD simulations, in the NVT ensemble, with fixed bond lengths¹⁰³ and a time step of 2 fs, were performed with the GROMACS software package.²⁴⁸ A non bond pairlist cutoff of 9 Å was used and the long range electrostatic interactions were treated with the particle mesh Ewald method.¹⁹⁹ The isokinetic temperature coupling²⁰⁰ was used to keep the temperature at the desired value. Six snapshots of the system were extracted from the trajectory of the first simulation at regular time intervals of 100 ps. Each configuration was simulated at 600 K with the gradual addition of the interproton distance restraints estimated from two-dimensional NOESY spectra obtained in SDS micelles. Only five of the experimentally detected NOEs were applied during the simulation as distance restraints (see Table 7.1). All NOEs were classified into three groups as strong, medium or weak and given upper limits $r_{i,j} = 0.32, 0.40$ and 0.50 nm, respectively,²⁴⁹ according to previous calculations on other lipodepsipeptides.²⁵⁰⁻²⁵²

As the united-atom representation was used, in which protons attached to aliphatic groups were treated implicitly, an additional distance term of 0.05 nm was added to the upper distance bounds when the distance restraint involved an implicit proton. An attractive half-harmonic restraining potential was applied to force the molecule to satisfy selected NOE distances. The gradual addition of the NOEs restraints was achieved by performing a series of 200 ps length MD simulations at 600 K of each configuration with the following values of the half-harmonic force constant $k : 0.05k_0, 0.3k_0, 0.5k_0, 0.8k_0, k_0$, with $k_0 = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$.

The temperature was then lowered again to 300 K and a 200 ps length run was performed with force constant k_0 for the distance restraints. The distance restraints were then set to zero and the productive MD run with no distance restraint was performed for 15 ns at 300 K for each of the six configurations. The ${}^3J_{HN-C^\alpha H}$ coupling constants were obtained from the simulation, by the Altona equation:²⁵³ ${}^3J_{HN-C^\alpha H} = 6.4\cos^2\theta - 1.4\cos\theta + 1.9$ where θ is the dihedral angle between HN-C $^\alpha$ H.

7.2 Results

The MD simulations after a first equilibration step were run in the presence of a number of NOE restraints (see Table 7.1) and a stepwise procedure of simulated annealing, as reported in section 7.1. The distance restraints were then removed and a productive run of 15 ns was performed. It resulted only one conformation of SRE in octane, a membrane-like environment, in agreement with the experimental NOEs. This structure, reported in Figure 7.2, retains the NOE distances also in the absence of the restraining potential during the 15 ns productive simulation, and is also consistent with the NOE distances obtained by NMR and measured on the model but not included as restraints.

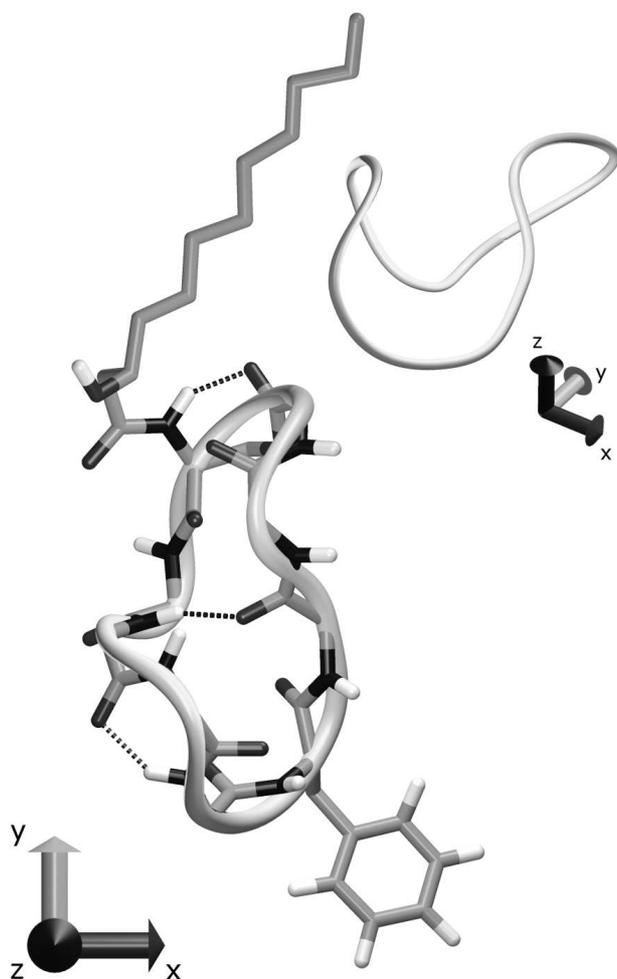


Figure 7.2: Representative snapshot of the structure of SR-E in octane. Black dots represent hydrogen bonds. The insert shows the backbone conformation that resembles the seam of a tennis ball.

The analysis of the trajectories reveals that the conformation of SRE in the membrane-like environment (Figure 7.2) is stabilized by three hydrogen bonds: the first one, involving residues NH-Dab3 and CO-z-Dhb7, induces the bending of the chain, while the other two, formed respectively between NH-Arg5 and CO-Dab3 and between NH-Ser1 and CO-4(Cl)-Thr9, define the structure of the loop. Figure 7.3 (panels A-C) shows the trajectories of the nitrogen-oxygen distances, involved in each of the three H-bonds. The three H-bonds are stable during the whole simulation, thus confirming the stability of the structure. The bending of the chain, induced by the hydrogen bond between the Dab3 and zDhb7 residues, involves five residues (Figure 7.2). It is worth noting that this bending is coherent with the strong NOE signal experimentally observed between the NH-Dab3 and C_{α} H-3(OH)-Asp8 groups that is reproduced during the entire simulation and monitored through the distance between the Dab3-NH hydrogen atom and the H- C_{α} atom of 3(OH)-Asp8, as shown in Figure 7.3 (panel D). No stable hydrogen bonds between the acyl chain and the peptide ring were detected from the analysis of the trajectories, suggesting that the lipid tail remains elongated and does not fold above the peptide ring, as shown in Figure 7.2.

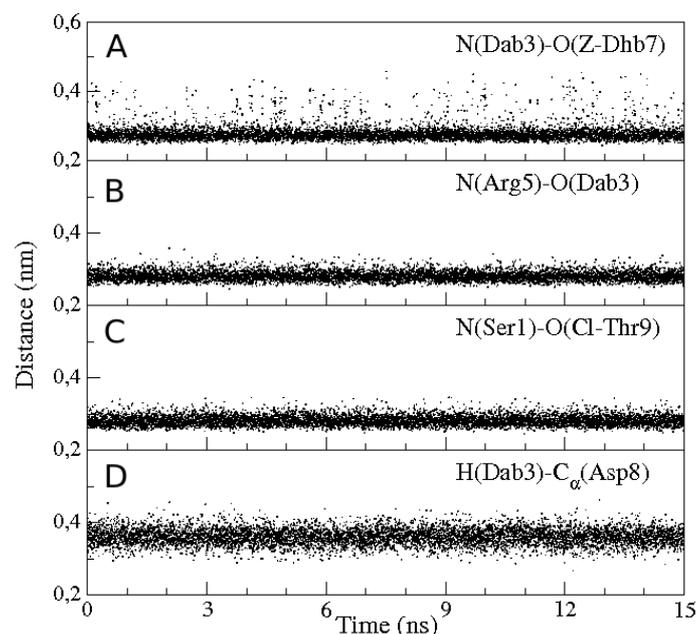


Figure 7.3: Time evolution of the distance between the nitrogen and the oxygen atoms involved in the three backbone-backbone hydrogen bonds (first three panels) and of the distance between the hydrogen and the α -carbon atom involved in the strong NOE signal (last panel).

It can be observed that the conformation of the backbone resembles the seam of a tennis ball, a pattern found for the first time in the bioactive lipodepsipep-

Residue	Averaged dihedral angles	
	Φ (degrees)	Ψ (degrees)
1Ser	-49 ± 12	133 ± 15
2Ser	70 ± 12	-100 ± 11
3Dab	121 ± 11	73 ± 9
4Dab	65 ± 8	-51 ± 14
5Arg	-70 ± 12	87 ± 15
6Phe	58 ± 11	105 ± 17
7Z-Dhb		
8OH-Asp	-103 ± 17	-33 ± 11
9Cl-Thr	-63 ± 11	

Table 7.2: Averaged Φ/Ψ dihedral angles. The errors correspond to the standard deviation obtained by the dihedral angle distribution.

tide WLIP, produced by *Pseudomonas reactans*^{254,255} and subsequently in many of lipodepsipeptide molecules in a membrane-like environment (TFE from 20 to 40% v/v) determined by NMR data and MD simulations like pseudomycin, syringopeptin 25-A, syringotoxin and cormycin.^{241,250-252} In particular, the structure of SRE is characterized by the presence of four consecutive turns, involving residues Ser2, Dab4, Phe6 and the lactone bond region, respectively. In Table 7.2 the mean values of the dihedral angles ϕ and ψ with their standard deviations are reported.

The sequences Ser1-Ser2-Dab3 and Dab3-Dab4-Arg5 form two γ -turns although ϕ/ψ torsion angles slightly deviate from the usual values, being ϕ_2 and ψ_2 70° and -100° (with standard deviations of 12° and 11°) and ϕ_4 and ψ_4 65° and -51° (with standard deviations of 8° and 14°). The third turn, corresponding to the sequence Arg5-Phe6-zDhb7, cannot be classified since it is characterized by unusual torsion angles: ϕ_6 and ψ_6 values of 58° and 105° , respectively; this effect might be ascribed to the proximity of the unsaturated residue zDhb7. Finally, the last turn involves the region connecting residues Ser1 and 4(Cl)-Thr9. The length of the tail is about 15 Å and the distance between N(Phe6) and C_β (Ser1), roughly representing the length of the ring, is about 10 Å.

In Table 7.3 are reported the mean value of the $^3J_{HN-C^\alpha H}$ coupling constants obtained by Altona equation from the dihedral angles as indicated in section 7.1^{251,253} together with their standard deviations, σ , and compared with the experimental values in the NMR spectra. Calculated and experimental values are in agreement, with the exception of Ser1 and (Cl)-Thr9 $^3J_{HN-C^\alpha H}$. This suggests that only the Ser1-(Cl)-Thr9 conformation is not well reproduced. It is worth noting that this discrepancy is in the molecular region around the residue (Cl)-Thr9 involved in the lactone linkage with the N-terminal Ser1, which is, in turn, acylated by the fatty acid residue.

Previous studies²⁴⁷ reported that SRE adopts in octane two different conforma-

Residue	$^3J_{HN-H\alpha}$ (Hz)	
	NMR (SDS)	MD
1Ser	8.4	3.2±0.9
2Ser	6.4	5.5±1.4
3Dab	7.9	9.4±0.4
4Dab	8.1	6.7±0.2
5Arg	5.3	5.4±1.4
6Phe		
7Z-Dhb		
8OH-Asp	8.8	8.7±1.4
9Cl-Thr	8.0	4.7±1.3

Table 7.3: $^3J_{HN-H\alpha}$ coupling constants as measured by NMR spectroscopy and calculated from the simulated trajectories. The errors on the calculated constants correspond to the standard deviation obtained by the $^3J_{HN-H\alpha}$ distribution.

tions referred to as “open” in which the acyl chain is elongated and “closed” in which the acyl chain is bent toward the peptide backbone. Considering the position of the acyl chain, the position of the Phe6 residue in the loop and the orientation of the peptide ring plane, our SRE structure reported in Figure 7.2 resembles the open conformation. However, a rather different pattern of hydrogen bonds and salt bridges was observed in the two structures. Thus, the three-dimensional structures of the lipopeptide obtained by MD simulations using the NMR data obtained in water solution or in phospholipid bilayer are different. In addition, an approximate charge distribution of the SRE molecules in the membrane-like environment, and then, likely, in membranes, was determined as shown in Figure 7.4. It can be observed that there is a small negatively charged region and a somewhat larger positively charged one. Both are linked to a neutral region and a lateral fatty acid chain. This arrangement can be considered as a building block that would account for the interactions of the syringomycin molecules with one another and with the charged and hydrophobic regions of the micelles in the self assembling in the pore.

7.3 Discussion

The results of this study indicate that the interaction of syringomycin with phospholipids is accompanied by a significant conformational modification of the lipopeptide molecule and provide an insight into the structural determinants of the membrane interaction. The model of the three-dimensional structure of SRE was obtained by MD simulations in octane solution, using NMR data measured in SDS micelle which showed that SDS micelles are suitable as a mimic of the phospholipid bilayer. In fact, CD spectra of SRE in phospholipid liposomes and in SDS micelles were nearly superimposable, indicating a conformational equivalence of SRE in these two me-

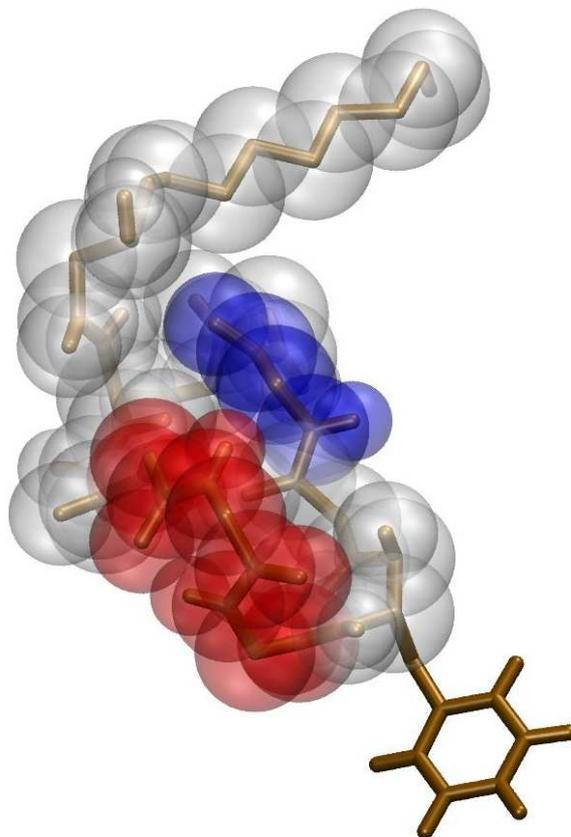


Figure 7.4: van der Waals surface of SR-E in octane as determined by molecular dynamics simulation. Neutral residues are in white, positive residues in red, negative residues in blue.

dia, at difference with the conformation displayed in water.²⁴⁷ The SDS micelle, in fact, captures the essential features which modulate peptide-membrane interaction: the presence of a strongly hydrophobic core, and a flexible polar interface capable of forming H-bonds and salt bridges with solvent and the peptide. Moreover, the advantage of the SDS micelles is that they tumble faster in solution than the large phospholipid vesicles, resulting in narrower NMR signals and, consequently, higher spectral resolution which enables the determination of detailed three-dimensional structures by conventional proton NMR techniques.

The conformation of the membrane-bound syringomycin obtained in this study is in line with several experimental data on SRE induced ion channels deriving from biophysical investigations. For example, the extended conformation comprising the 15 Å long acyl moiety and a macrocycle of about 10 Å, considering the distance between N(Phe6) and C_β(Ser1), is coherent with a model in which two trimers are

aligned along the membrane.²⁵⁶ Moreover, the surface charge distribution conceivably plays an important role in the formation of ion channels. The assembly of these supramolecular structures, which requires the interactions between the lipopeptide molecules as well as interactions with the membrane components, whose involvement in the SRE channel structure has been demonstrated, is guided by molecular recognition. The structural model of SRE presented in this study shows three distinct electron density regions: neutral, a smaller negatively charged and a larger positively charged region. It could be hypothesized that, depending on the lipid species, different macromolecular arrangements could be formed. For example, it could be expected that with negatively charged lipids SRE molecules would arrange in a way to minimize the exposure of the negatively charged parts of the molecule. Conversely, with zwitterionic lipids, a more extensive interaction between SRE molecules and the membrane components could be envisaged. The correlation between the type of the membrane lipid and the channel structure and functions has been proven by different approaches. EPR experiments²⁵⁷ indicated an active role of the lipid molecules, aligned in ordered, motionally constrained structure, in the architecture of the pores created in the presence of syringomycin. Moreover, the number of lipid molecules involved in one pore, estimated as at least 40, can depend on the lipid composition of the membrane. These evidences on the lipid involvement in the pore structure complement those obtained by conductivity measurements.^{256,257}

The results of the present study can help to gain an insight into the complex network of molecular recognition events that is important for the structure and functions of the SRE ion channels. In fact, knowing how a peptide folds and interacts with target molecules is a unique requirement for the interpretation of its mode of action on the molecular level. Furthermore, it is also a starting point for the planning of structural modifications aimed at the improvement of selectivity.

Modeling studies are under course to envisage these intermolecular interactions and propose a molecular model for the pore. Further studies are planned in the future and, in particular, the effects of sterol or sphingolipid components both on the membrane on the SRE behavior - i.e. conformational modifications - such to facilitate their aggregation and therefore eventually the pore formation will be studied.

CONCLUDING REMARKS

As summarized in the methods section of the present thesis, recent developments in computer simulations of biological macromolecules have enhanced the range of applicability of these techniques in the study of folding processes. The methods used in this thesis form another contribution to this field and the applications to peptide model systems have yielded interesting results. The major problem with molecular dynamics (MD) simulations of the folding process of macromolecular systems, such as proteins, is due to the conformational sampling efficiency. This difficulty is also present in systems with a lower complexity, such as peptides, but is more tractable than for proteins. Experimentally, peptides fold at very fast rates, requiring probing on the nanosecond-microsecond time resolution, hence offering a unique opportunity to bridge the gap between theoretical and experimental understanding of protein folding. Thus MD simulations become an extremely powerful tool not only to understand and interpret the experiments at the microscopic level, but also to study regions which are not accessible experimentally.

In the present thesis, three cyclic peptides served as model systems for the study of the structure and dynamics of turns, a very common and important feature of the secondary structure of proteins. In particular, two β -hairpin peptides have been deeply studied, and their folding/unfolding transitions have been characterized.

In Chapter 4 the folding process for the Gramicidin analogue GS6 has been extensively investigated by the use of MD simulations coupled to statistical mechanical models to characterize the thermodynamics and kinetics of the process. Such an approach allowed to quantitatively obtain the peptide folding/unfolding partial molar thermodynamics and the corresponding kinetic rate constants. Results showed a very interesting and peculiar thermodynamic behavior of GS6, characterized by an unfolded state with lower entropy and energy than the folded one. This behavior, already observed for another peptide, is likely to be a specific feature of solvated small peptides where the effects of intramolecular interactions and configurational freedom are relatively small compared to the peptide-solvent thermodynamic coupling.

In Chapter 5, by means of a novel theoretical-computational methods, the IR spectra in the amide I region of trans-N-methylacetamide and of the folded and unfolded states of two β -hairpin peptides, peptide 1 and GS10, for which the experimental spectra are available, were reconstructed. The method used, based on mixed QM/MM calculations, proved to be able to reproduce rather well the experimental data and the analysis of the computed spectra revealed interesting features unachievable by means of experimental techniques. The results showed indeed that the main determinant to the characteristic signal of the unfolded–folded difference spectrum, experimentally observed for a wide variety of peptides and proteins, arises from the hydrophobic sidechains. In the unfolded configurations the hydrophobic sidechains are on average closer to their own peptide group with respect to the case of folded configurations, hence providing a less polar environment to the vibrating amide I mode (mainly C=O stretching). A less polar environment gives rise to a shift of the unfolded peak toward higher frequencies, *i.e.*, toward the amide I band of the isolated peptide group, with respect to the folded peak, thus leading to a negative-positive difference spectrum.

This methodology to reconstruct IR spectra is thus very promising, as it provides the opportunity to analyze the contribution of single secondary structures and even of single residues to the total infrared spectra of peptides and proteins, shedding light in the interpretation of experimental spectra which is mostly based on empirical rules and widely accepted assumptions which general validity cannot be always demonstrated.

In Chapter 6 the same methodology explained in Chapter 5 was used to reproduce the experimental temperature dependent spectra of two small β -hairpin peptides, GS6 and GS10, giving results in good agreement with the experimental ones. The unfolded–folded difference spectrum of GS6 has been also reproduced, showing a peculiar trend, in agreement with its thermodynamic behavior: the unfolded–folded difference spectrum showed an opposite trend with respect to the temperature difference one. This can be attributed to the absence, in the primary sequence of this peptide, of hydrophobic side chains which constitute the main contribution to the IR unfolded–folded difference spectrum, showing an intense characteristic negative-positive signal. The inversion between the temperature difference spectrum and the unfolded-folded one, indicates that in the considered temperature range the folded state population increases as the temperature is raising and so that it is not possible to interpret the temperature difference spectra obtained by subtracting the spectrum at the lowest temperature from the spectra at higher temperatures, as representative of the unfolding process as it has been done in experimental studies. Molecular Dynamics simulations, together with the other theoretical methods used, have thus shown to be very helpful in the interpretation of experimental infrared spectra.

In Chapter 7 a MD simulation of the Lipodepsipeptide Syringomycin E was per-

formed and the results were compared with NMR experimental spectra in order to understand the structure of the peptide in the membrane environment. The results of this study indicate that the interaction of syringomycin with phospholipids is accompanied by a significant conformational modification of the lipopeptide molecule and provide an insight into the structural determinants of the membrane interaction. Moreover, The conformation of the membrane-bound syringomycin obtained by means of the MD simulation is in line with several experimental data on SRE induced ion channels deriving from biophysical investigations.

BIBLIOGRAPHY

- [1] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, and H. Wyckoff. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181:662–666, 1958.
- [2] M. F. Perutz, M. G. Rossmann; A. F. Cullis, H. Muirhead, and G. Will. Structure of hemoglobin: A three-dimensional fourier synthesis at 5.5 \AA resolution, obtained by x-ray analysis. *Nature*, 185:416–422, 1960.
- [3] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–30, 1973.
- [4] E. Haber and C. B. Anfinsen. Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.*, 237:1839–44, 1962.
- [5] K. W. Plaxco and C. M. Dobson. Time-resolved biophysical methods in the study of protein folding. *Curr. Opin. Struct. Biol.*, 6(5):630–636, 1996.
- [6] D. J. Brockwell, D. A. Smith, and S. E. Radford. Protein folding mechanisms: new methods and emerging ideas. *Curr. Opin. Struct. Biol.*, 10:16–25, 2000.
- [7] A. R. Fersht and S. Sato. ϕ -value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. USA*, 101:7976–81, 2004.
- [8] D. P. Goldenberg. Genetic studies of protein stability and mechanisms of folding. *Annu. Rev. Biophys. Biomol. Struct.*, 17:481–507, 1988.
- [9] L. S. Itzhaki, D. E. Otzen, and A. R. Fersht. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.*, 254:260–88, 1995.
- [10] G. D. Cymes, C. Grosman, and A. Auerbach. Structure of the transition state of gating in the acetylcholine receptor channel pore: a ϕ -value analysis. *Biochemistry*, 41:5548–55, 2002.

- [11] T. R. Sosnick, R. S. Dothager, and B. A. Krantz. Differences in the folding transition state of ubiquitin indicated by ϕ and ψ analyses. *Proc. Natl. Acad. Sci. USA*, 101:17377–82, 2004.
- [12] M. M. Krishna, L. Hoang, Y. Lin, and S. W. Englander. Hydrogen exchange methods to study protein folding. *Methods*, 34:51–64, 2004.
- [13] H. Maity, M. Maity, M. M. Krishna, L. Mayne, and S. W. Englander. Protein folding: the stepwise assembly of foldon units. *Proc. Natl. Acad. Sci. USA*, 102:4741–46, 2005.
- [14] K. L. Maxwell, D. Wildes, A. Zarrine-Afsar, M. A. De Los Rios, and A. G. Brown. Protein folding: defining a standard set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.*, 14:602–16, 2005.
- [15] O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber. The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci. USA*, 96:9597–601, 1999.
- [16] G. S. Jas, W. A. Eaton, and J. Hofrichter. Effect of viscosity on the kinetics of α -helix and β -hairpin formation. *J. Phys. Chem. B*, 105:261–72, 2001.
- [17] F. Krieger, B. Fierz, O. Bieri, M. Drewello, and T. Kiefhaber. Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.*, 332:265–74, 2003.
- [18] R. H. Callender, R. B. Dyer, R. Gilmanshin, and W. H. Woodruff. Fast events in protein folding: the time evolution of primary processes. *Annu. Rev. Phys. Chem.*, 49:173–202, 1998.
- [19] A. Barth. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta*, 1767:1073–1101, 2007.
- [20] E. Chen, R. A. Goldbeck, and D. S. Kliger. Nanosecond time-resolved spectroscopy of biomolecular processes. *Annu. Rev. Biophys. Biomol. Struct.*, 26:327–355, 1997.
- [21] S. Hahn, S. Ham, and M. Cho. Simulation studies of amide i ir absorption and two-dimensional ir spectra of beta-hairpins in liquid water. *J. Phys. Chem. B*, 109:11789–11801, 2005.
- [22] C. Cecconi, E. A. Shank, C. Bustamante, and S. Marqusee. Direct observation of the three-state folding of a single protein molecule. *Science*, 309:2057–60, 2005.

- [23] T. A. Laurence, X. Kong, M. Jager, and S. Weiss. Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl. Acad. Sci. USA*, 102:17348–53, 2005.
- [24] S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, and A. A. Deniz. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc. Natl. Acad. Sci. USA*, 104:2649–54, 2007.
- [25] B. Schuler, E. A. Lipman, and W. A. Eaton. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419:743–47, 2002.
- [26] J.A. McCammon, B.R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.
- [27] Y. Duan and P.A. Kollman. Pathways to protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282:740–744, 1998.
- [28] P.L. Freddolino, F. Liu, M. Gruebele, and K. Schulten. Ten-microsecond molecular dynamics simulation of a fast-folding ww domain. *Biophys J*, 94:75–77, 2008.
- [29] D.L. Ensign, P.M. Kasson, and V.S. Pande. Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. *J Mol Biol*, 374:806–816, 2007.
- [30] A. Perez, F.J. Luque, and M. Orozco. Dynamics of b-dna on the microsecond time scale. *J Am Chem Soc*, 129:14739–45, 2007.
- [31] M.M. Seibert, A. Patriksson, B. Hess, and D. van der Spoel. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *J Mol Biol*, 354:173–183, 2005.
- [32] P. Maragakis, K. Lindorff-Larsen, M.P. Eastwood, R.O. Dror, J.L. Klepeis, I.T. Arkin, M.O. Jensen, H. Xu, N. Trbovic, and R.A. Friesner. Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *J Phys Chem B*, 112:6155–58, 2008.
- [33] K. Martinez-Mayorga, M.C. Pitman, A. Grossfield, S.E. Feller, and M.F. Brown. Retinal counterion switch mechanism in vision evaluated by molecular simulations. *J Am Chem Soc*, 128:16502–03, 2006.
- [34] A. Grossfield, M.C. Pitman, S.E. Feller, O. Soubias, and K. Gawrisch. Internal hydration increases during activation of the g-protein-coupled receptor rhodopsin. *J Mol Biol*, 381:478–486, 2008.

- [35] R.O. Dror, D.H. Arlow, D.W. Borhani, M. O. Jensen, S. Piana, and D.E. Shaw. Identification of two distinct inactive conformations of the β -adrenergic receptor reconciles structural and biochemical observations. *Proc Natl Acad Sci U S A*, 106:4689–4694, 2009.
- [36] V. Daggett and M. Levitt. Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.*, 232:600–619, 1993.
- [37] M. Vendruscolo, E. Paci, C. M. Dobson, and M. Karplus. Three key residues form a critical contact network in a protein folding transition state. *Nature*, 409:641–45, 2001.
- [38] T. L. Religa, J. S. Markson, U. Mayor, S. M. Freund, and A. R. Fersht. Solution structure of a protein denatured state and folding intermediate. *Nature*, 437:1053–56, 2005.
- [39] U. Mayor, N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, S. G. Jas, S. M. V. Freund, O. V. D. O. V. Alonso, V. Daggett, and A. R. Fersht. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature*, 421:863–867, 2003.
- [40] S. Gianni, N. R. Guydosh, F. Khan, T. D. Caldas, U. Mayor, G. W. N. White, M. L. DeMarco, V. Daggett, and A. R. Fersht. Unifying features in protein-folding mechanisms. *Proc. Natl Acad. Sci. USA*, 100:13286–91, 2003.
- [41] V. Daggett and A. Fersht. The present view of the mechanism of protein folding. *Nature Rev. Mol. Cell Biol.*, 4:497–502, 2003.
- [42] C. Simmerling, B. Strockbine, and A. E. Roitberg. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.*, 124:1258–59, 2002.
- [43] J. W. Pitera and W. Swope. Understanding folding and design: replica-exchange simulations of "trp-cage" miniproteins. *Proc. Natl. Acad. Sci. USA*, 100:7587–92, 2003.
- [44] S. Chowdhury, M. C. Lee, G. Xiong, and Y. Duan. Ab initio folding simulation of the trp-cage mini-protein approaches nmr resolution. *J. Mol. Biol.*, 327:711–17, 2003.
- [45] H. Lei and Y. Duan. Ab initio folding of albumin binding domain from all-atom molecular dynamics simulation. *J. Phys. Chem. B*, 111:5458–63, 2007.
- [46] A. K. Felts, Y. Harano, E. Gallicchio, and R. M. Levy. Free energy surfaces of β -hairpin and α -helical peptides generated by replica exchange molecular dynamics with the agbnp implicit solvent model. *Proteins*, 56:310–21, 2004.

- [47] S. Gnanakaran and A. E. Garcia. Validation of an all-atom protein force field: from dipeptides to larger peptides. *J. Phys. Chem. B*, 107:12555–57, 2003.
- [48] S. Jang, E. Kim, and Y. Pak. Direct folding simulation of α -helices and β -hairpins based on a single all-atom force field with an implicit solvation model. *Proteins*, 66:53–60, 2007.
- [49] T. Yoda, Y. Sugita, and Y. Okamoto. Secondary structure preferences of force fields for proteins evaluated by generalized-ensemble simulations. *Chem. Phys.*, 307:269–83, 2004.
- [50] R. Zhou. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins*, 53:148–61, 2003.
- [51] R.D. Schaeffer, A. Fersht, and V. Daggett. Combining experiment and simulation in protein folding: closing the gap for small model systems,. *Curr Opin Struct Biol*, 18:4–9, 2008.
- [52] C. B. Anfinsen and H. A. Scheraga. Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.*, 29:205–300, 1975.
- [53] K. A. Dill. Dominant forces in protein folding. *Biochemistry*, 29:7133–55, 1990.
- [54] K. A. Dill. Polymer principles and protein folding. *Protein Sci.*, 8:1166–80, 1999.
- [55] R. Wolfenden. Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins. *J. Gen. Physiol.*, 129:357–62, 2007.
- [56] M. H. J. Cordes, A. R. Davidson, and R. T. Sauer. Sequence space, folding and protein design. *Curr. Opin. Struct. Biol*, 6:3–10, 1996.
- [57] M. H. Hecht, A. Das, A. Go, L. H. Bradley, and Y. Wei. De novo proteins from designed combinatorial libraries. *Protein Sci.*, 13:1711–23, 2004.
- [58] S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, and M. H. Hecht. Protein design by binary patterning of polar and nonpolar amino acids. *Science*, 262:1680–85, 1993.
- [59] D. E. Kim, H. Gu, and D. Baker. The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc. Natl. Acad. Sci. USA*, 95:4982–86, 1998.

- [60] H. J. Dyson, P. E. Wright, and H. A. Scheraga. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl. Acad. Sci. USA*, 103:13057–61, 2006.
- [61] J. S. Richardson. The anatomy and taxonomy of protein structure. *Adv Protein Chem.*, 34:167–339, 1981.
- [62] G. D. Rose, L. M. Gierasch, and J. A. P. Smith. Turns in peptides and proteins. *Adv Protein Chem.*, 37:1–109, 1985.
- [63] H. J. Dyson and P. E. Wright. Peptide conformation and protein folding. *Curr. Opin. Struct. Biol*, 3:60–65, 1993.
- [64] P. E. Wright, H. J. Dyson, and R. A. D. Lerner. Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding. *Biochemistry*, 27:7167–75, 1988.
- [65] P. N. Lewis, F. A. Momany, and H. A. Scheraga. 1971 sep;68(9):2293-7. folding of polypeptide chains in proteins: a proposed mechanism for folding. *Proc Natl Acad Sci U S A.*, 68:2293–97, 1971.
- [66] S. S. Zimmerman and H. A. Scheraga. Local interactions in bends of proteins. *Proc Natl Acad Sci U S A.*, 74:4126–29, 1977.
- [67] D. E. Kim, C. Fisher, and D. Baker. A breakdown of symmetry in the folding transition state of protein l. *J Mol Biol.*, 298:971–84, 2000.
- [68] J. Rizo and L. M. Gierasch. Constrained peptides: models of bioactive peptides and protein substructures. *Annu Rev Biochem.*, 61:387–418, 1992.
- [69] I. L. Karle. Variability in the backbone conformation of cyclic pentapeptides. crystal structure of cyclic(gly-l-pro-d-phe-gly-l-ala). *Int J Pept Protein Res.*, 28:420–27, 1986.
- [70] I. L. Karle, J. W. Gibson, and J. Karle. The conformation and crystal structure of the cyclic polypeptide -gly-gly-d-ala-d-ala-gly-gly .3h2o. *J Am Chem Soc.*, 92:3755–60, 1970.
- [71] I. L. Karle and J. Karle. An application of a new phase determination procedure to the structure of cyclo(hexaglycyl)demihydrate. *Acta Cryst.*, 16:969–975, 1963.
- [72] I. L. Karle and D. W. Urry. Crystal structure of cyclic (apgvv)₂, an analog of elastin, and a suggested mechanism for elongation/contraction of the molecule. *Biopolymers*, 77:198–204, 2005.

- [73] C. Ramakrishnan, P. K. C. Paul, and K. Ramnarayan. Cyclic peptides-small and big and their conformational aspects. *Proc Int Symp Biomol Struct Interact Suppl J Biosci*, 8:239–251, 1985.
- [74] L. M. Gierasch, C. M. Deber, V. Madison, C. H. Niu, and E. R. Blout. Conformations of (x-l-pro-y)₂ cyclic hexapeptides. preferred beta-turn conformers and implications for beta turns in proteins. *Biochemistry*, 20:4730–38, 1981.
- [75] L. G. Pease, C. M. Deber, and E. R. Blout. Cyclic peptides. v. 1 h and 13 c nuclear magnetic resonance determination of the preferred beta conformation for proline-containing cyclic hexapeptides. *J Am Chem Soc.*, 95:258–60, 1973.
- [76] D. A. Torchia, A. Di Corato, S. C. Wong, C. M. Deber, and E. R. Blout. Cyclic peptides. ii. solution conformations of cyclo(prolylserylglycylprolylserylglycyl) from nuclear magnetic resonance. *J Am Chem Soc.*, 94:609–15, 1972.
- [77] D. A. Torchia, S. C. Wong, C. M. Deber, and E. R. Blout. Cyclic peptides. 3. solution conformations of cyclo(serylprolylglycylserylprolylglycyl) from nuclear magnetic resonance. *J Am Chem Soc.*, 94:616–20, 1972.
- [78] K. D. Kopple, M. Ohnishi, and A. Go. Conformations of cyclic peptides. iv. nuclear magnetic resonance studies of cyclo-pentaglycyl-l-leucyl and cyclo-diglycyl-l-histidyl-diglycyl-l-tyrosyl. *Biochemistry*, 8:4087–95, 1969.
- [79] A. E. Tonelli and A. I. Brewster. The conformational characteristics in solution of the cyclic hexapeptide gly-gly-d-ala-d-ala-gly-gly. *J Am Chem Soc.*, 94:2851–54, 1972.
- [80] A. E. Tonelli and A. I. R. Brewster. The conformational characteristics in solution of the synthetic cyclic hexapeptide. *Biopolymers*, 12:193–200, 1973.
- [81] Y. A. Ovchinnikov and V. T. Ivanov. Conformational states and biological activity of cyclic peptides. *Tetrahedron*, 31:2177–209, 1975.
- [82] V. Madison. Conformational energy and circular dichroism computed for cyclo-(pro-gly)₃. *Biopolymers*, 12:1837–1852, 1973.
- [83] C. A. Bush, S. K. Sarkar, and K. D. Kopple. Circular dichroism of beta turns in peptides and proteins. *Biochemistry*, 17:4951–54, 1978.
- [84] S. E. Kiehna and M. L. Waters. Sequence dependence of beta-hairpin structure: comparison of a salt bridge and an aromatic interaction. *Protein Sci.*, 12:2657–67, 2003.
- [85] R. M. Hughes and M. L. Waters. Model systems for beta-hairpins and beta-sheets. *Curr Opin Struct Biol.*, 16:514–24, 2006.

- [86] B. Ciani, M. Jourdan, and M. S. Searle. Stabilization of beta-hairpin peptides by salt bridges: role of preorganization in the energetic contribution of weak interactions. *J Am Chem Soc.*, 125:9038–47, 2003.
- [87] R. B. Dyer, S. J. Maness, E. S. Peterson, S. Franzen, R.M.Fesinmeyer, and N. H. Andersen. The mechanism of beta-hairpin formation. *Biochemistry*, 43:11560–11566, 2004.
- [88] F. Blanco, M. Ramirez-Alvarado, and L. Serrano. Formation and stability of beta-hairpin structures in polypeptides. *Curr Opin Struct Biol.*, 8:107–11, 1998.
- [89] D. Du, Y. Zhu, C. Huang, and F. Ga. Understanding the key factors that control the rate of beta-hairpin folding. *Proc.Natl.Acad.Sci.U.S.A.*, 101(45):15915–15920, 2004.
- [90] M. P. Allen and D. J. Tildesly. *Computer simulation of liquids*. Oxford University Press, Oxford, 1989.
- [91] D. Frenkel and B. Smit. *Understanding Molecular Simulation. From Algorithms to Applications*. Academic Press, Boston, 1996.
- [92] W. F. van Gunsteren and P. K. Weiner. *Computer simulation of biomolecular systems*. Escom Science, Leiden (NL), 1989.
- [93] B. R. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187, 1983.
- [94] S. Weiner, P. Kollman, D. Nguyen, and D. Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.*, 7:230, 1986.
- [95] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Hochschulverlag AG an der ETH Zürich, Zürich, 1996.
- [96] L. Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159:98, 1967.
- [97] L. Verlet. Computer "experiments" on classical fluids. II. Equilibrium correlation functions. *Phys. Rev.*, 165:201–23, 1968.
- [98] D. Beeman. Some multistep methods for use in molecular dynamics calculations. *J. Comput. Phys.*, 20:130, 1976.

- [99] C. Gear. *Numerical Initial Value Problems in Ordinary Differential Equations*. Englewood Cliffs, NJ, USA, 1971.
- [100] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577–8593, 1995.
- [101] T. Schlick, E. Barth, and M. Mandziuk. Biomolecular dynamics at long timesteps:bridging the timescale gap between simulation and experimentation. *Annu. Rev. Biomol. Struct.*, 26:181–222, 1997.
- [102] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion in a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.*, 23:327–341, 1977.
- [103] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463–1472, 1997.
- [104] K. A. Feenstra, B. Hess, and H. J. C. Berendsen. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comp. Chem*, 20:786–798, 1999.
- [105] A. Amadei, G. Chillemi, M. A. Ceruso, A. Grottesi, and A. Di Nola. Molecular dynamics simulations with 0 roto-traslational motions: Theoretical basis and statistical mechanical consistency. *J. Chem. Phys.*, 112:9–23, 2000.
- [106] H. Bekker, J. P. van den Berg, and T. A. Wassenaar. A method to obtain a near-minimal-volume molecular simulation of a macromolecule, using periodic boundary conditions and rotational constraints. *J. Comput. Chem.*, 25:1037–1046, 2004.
- [107] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [108] D. Brown and J. Clarke. Molecular dynamics simulations of polymer fiber microstructure. *J. Chem. Phys*, 84:2858, 1986.
- [109] S. Nosè. Constant temperature molecular dynamics methods. *Prog. Theoret. Phys. Supplement*, 103:1–46, 1991.
- [110] W. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev.*, A31:1695–1697, 1985.

- [111] G. J. Martyna, M. L. Klein, and M. Tuckerman. Nosé hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97:2635–2645, 1992.
- [112] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins: Struct. Funct. Genet.*, 17:412–425, 1993.
- [113] A.E. García. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.*, 66:2696–2699, 1992.
- [114] A. Kitao, F. Hirata, and N. Gō. The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *J. Chem. Phys.*, 158:447–472, 1991.
- [115] D. M. van Aalten, A. Amadei, A. B. Linssen, V. G. Eijssink, G. Vriend, and H. J. C. Berendsen. The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *PROTEINS: Struct. Funct. Gen.*, 22:45–54, 1995.
- [116] B. L. de Groot, D. M. F. van Aalten, A. Amadei, and H. J. C. Berendsen. Domain motions in Bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. *Proteins: Struct. Funct. Genet.*, 31:116–127, 1998.
- [117] D. M. van Aalten, D. A. Conn, B. L. de Groot, H. J. C. Berendsen, J. B. Findlay, and A. Amadei. Protein dynamics derived from clusters of crystal structures. *Biophys. J.*, 73:2891–2896, 1997.
- [118] R. Abseher, L. Horstink, C. W. Hilbers, and M. Nilges. Essential spaces defined by nmr structure ensembles and molecular dynamics simulation show significant overlap. *PROTEINS: Struct. Funct. Gen.*, 31:370–382, 1998.
- [119] B. L. de Groot, D. M. van Aalten, R. M. Scheek, A. Amadei, G. Vriend, and H. J. C. Berendsen. Prediction of protein conformational freedom from distance constraints. *PROTEINS: Struct. Funct. Gen.*, 29:240–251, 1997.
- [120] D.A. McQuarrie. *Statistical mechanics*. Harper & Row, New York, 1976.
- [121] J. K. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3:300–313, 1935.
- [122] M. Mezei and D. L. Beveridge. *Computer Simulations and Biomolecular Systems*. Ann. New York Academy of Sciences, New York, d.L. Beveridge and W.L. Jorgensen edition, 1986.

- [123] F. Furche and R. Ahlrichs. Absolute configuration of d_2 -symmetric fullerene c_{64} . *J. Am. Chem. Soc.*, 124:3804–3805, 2002.
- [124] R. Ahlrichs, J. Besinger, A. Eichhofer, D. Fenske, and A. Gbureck. Synthesis, crystal structure and binding properties of the mixed valence clusters $cu_{32}as_{30}(dppm)_8$ and $cu_{26}te_{12}(pet_2ph)_{12}$. *Angew. Chem. Int. Ed.*, 39:3929–3933, 2000.
- [125] B. O. Roos, P. A. Malmqvist, V. Molina, L. Serrano-Andres, and M. Merchán. On the theoretical characterization of the low-energy absorption band in pyrrole. *J. Chem. Phys.*, 116:7526–7536, 2002.
- [126] R. Improta, A. di Matteo, and V. Barone. Effective modeling of intrinsic and environmental effects on the structure and epr parameters of nitroxides by an integrated qm/mm/pcm approach. *Theor. Chem. Acc.*, 104:273–279, 2000.
- [127] P. Carsky, R. Curik, F. A. Gianturco, R. R. Lucchese, , and M. Polasek. Computing the exchange interaction in electron scattering from polyatomic molecules. *Phys.Rev. A*, 65:052713, 2002.
- [128] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103:227–249, 1976.
- [129] F. F. Abraham, J. Q. Broughton, N. Bernstein, and E. Kaxiras. Spanning the continuum to quantum length scales in dynamic simulation of brittle fracture. *Europhys. Lett.*, 44:783–792, 1998.
- [130] J. Gao and M. Freindorf. Hybrid ab initio qm/mm simulation of n-methyl acetamide in aqueous solution. *J. Phys. Chem. A*, 101:3182–3194, 1997.
- [131] F. Aquilante, M. Cossi, O. Crescenzi, G. Scalmani, and V. Barone. Computation of acetone uv spectrum in gas phase and in aqueous solution by a mixed discrete/continuum model. *Mol. Phys*, 101:1945–1953, 2003.
- [132] B. G. Schulze and J. D. Evanseck. Cooperative role of arg45 and his64 in the spectroscopic a3 state of carbonmonoxy myoglobin: molecular dynamics simulations, multivariate analysis and quantum mechanical computations. *J. Am.Chem. Soc.*, 121:6444–6454, 1999.
- [133] E. S. Manas, J. M. Vanderkooi, and K. A. Sharp. The effect of protein environment on the low temperature electronic spectroscopy of cytochrome c and microperoxidase-11. *J. Phys. Chem. B*, 103:6334–6348, 1999.
- [134] M. Aschi, R. Spezia, A. Di Nola, and A. Amadei. A first principles method to model perturbed electronic wave-functions: the effect of an external electric field. *Chem.Phys.Lett*, 344:374–380, 2001.

- [135] R. Spezia, M. Aschi, A. Di Nola, and A. Amadei. Extension of the perturbed matrix method: application to a water molecule. *Chem.Phys.Lett.*, 365:450–456, 2002.
- [136] A. Amadei, F. Marinelli, M. D’Abramo M., M. D’Alessandro, M. Anselmi, A. Di Nola, and M. Aschi. Theoretical modeling of vibroelectronic quantum states in complex molecular systems: Solvated carbon monoxide: a test case. *J.Chem.Phys.*, 122:124506, 2005.
- [137] M. Anselmi, M. Aschi, A. Di Nola, and A. Amadei. Theoretical characterization of carbon monoxide vibrational spectrum in sperm whale myoglobin distal pocket. *Biophys.J.*, 92, 2007.
- [138] M. Aschi and A. Amadei. Theoretical-computational modeling of quantum vibrational states of polyatomic systems in complex environment. *IJLSS*, In Press, 2009.
- [139] J.K. Patel, C.H. Kapadia, and D.B. Owen. *Handbook of statistical distributions*. Marcel Dekker, New York, 1976.
- [140] N.L. Johnson, S. Kotz, and A.W. Kemp. *Univariate discrete distributions*. John Wiley Son, New York, 1992.
- [141] J.K. Patel and C.B. Read. *Handbook of the normal distribution*. Marcel Dekker, New York, 1982.
- [142] T.L. Hill. *Statistical Mechanics*. McGraw-Hill Book Company, New York, 1956.
- [143] J.E. Mayer and M.G. Mayer. *Statistical Mechanics*. Wiley, New York, 1940.
- [144] G.A. Baker Jr. *Essentials of Padé approximants*. Academic Press, New York, 1975.
- [145] G.A. Baker Jr. and J.L. Gammel. *The Padé approximant in theoretical physics*. Academic Press, New York, 1970.
- [146] A. Amadei, M.E.F. Apol, A. Di Nola, and H. J.C. Berendsen. The quasi-gaussian entropy theory: Free energy calculations based on the potential energy distribution. *J. Chem. Phys.*, 104:1560–1574, 1996.
- [147] J.K. Ord. *Families of frequency distributions*. Griffin, London, 1972.
- [148] K.A. Dunning and J.N. Hanson. Generalized pearson distributions and non-linear programing. *J. Statist. Comp. and Simul.*, 6:115–128, 1978.

- [149] M.E.F. Apol, A. Amadei, and H. J.C. Berendsen. Application of the quasi-gaussian entropy theory to the calculation of thermodynamic properties of water and methane in the liquid and gas phase. *J. Chem. Phys.*, 104:6665–6678, 1996.
- [150] J.P. Hansen and I.R. McDonald. *Theory of simple liquids*. Aca, New York, 1986.
- [151] N. Carnahan and K. Starling. Equation of state for nonattracting rigid spheres. *J. Chem. Phys.*, 51:635–636, 1969.
- [152] N.I. Johnson and S. Kotz. *Continuous Univariate Distributions*. Houghton Mifflin, New York, 1970.
- [153] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1964.
- [154] L. D. Landau and E. M. Lifshitz. *Course of Theoretical Physics, Part 1 (Statistical Physics)*. Pergamon Press, Oxford, U.K, 1980.
- [155] M. E. F. Apol, A. Amadei, and H. J. C. Berendsen. Derivation of a thermodynamic closure relation in the isothermal-isobaric ensemble using the quasi-gaussian entropy theory. *Chem. Phys. Lett.*, 256:172–178, 1996.
- [156] A. Amadei, D. Roccatano, M. E. F. Apol, H. J. C. Berendsen, and A. Di Nola. Prediction of the liquid-vapor equilibrium pressure using the quasi-gaussian entropy theory. *J. Chem. Phys.*, 105:7022–7025, 1996.
- [157] D. Roccatano, A. Amadei, M. E. F. Apol, A. Di Nola, and H.J. C. Berendsen. Application of the quasi-gaussian entropy theory to molecular dynamics simulations of lennard-jones fluids. *J. Chem Phys.*, 109:6358–6363, 1998.
- [158] A. Amadei, M. E. F. Apol, G. Chillemi, H.J. C. Berendsen, and A. Di Nola. Derivation of a generalfluid equation of state based on the quasi-gaussian entropy theory: application to the lennard-jones fluid. *Molec. Phys.*, 96(10):1469–1490, 1999.
- [159] M. D’Alessandro, M. D’Abramo, G. Brancato, A. Di Nola, and A. Amadei. Statistical mechanics and thermodynamics of simulated ionic solutions. *J. Phys.Chem. B*, 106:11843–11848, 2002.
- [160] M. D’Abramo, M. D’Alessandro, and A. Amadei. On the use of the quasi gaussian entropy theory in the study of simulated dilute solutions. *J. Chem. Phys.*, 120:5526–5531, 2004.

- [161] M. D'Alessandro, M. D'Abramo, M. Paci, and A. Amadei. Theoretical study of the thermodynamics of a solvated peptide: Contryphan vn. *Physica Scripta*, T118,:196,, 2005.
- [162] A. Amadei, M. E. F. Apol, and H. J. C. Berendsen. Extensions of the quasi-gaussian entropy theory. *J.Chem.Phys.*, 106:1893–1912, 1997.
- [163] I. Daidone, M. D'Abramo, A. Di Nola, and A. Amadei. Theoretical characterization of alpha-helix and beta-hairpin folding kinetics. *J.Am.Chem.Soc.*, 127:14825–14832, 2005.
- [164] F. Noe, I. Daidone, J. C. Smith, A. Di Nola, and A. Amadei. Solvent electrostriction-driven peptide folding revealed by quasi-gaussian entropy theory and molecular dynamics simulation. *J. Phys. Chem. B*, 112(35):11155–11163, 2008.
- [165] A. C. Gibbs, L. H. Kondejewski, W. Gronwald, A.M. Nip, R.S. Hodges, B. D. Sykes, and D. S. Wishart. Unusual beta-sheet periodicity in small cyclic peptides. *Nat.Struct.Biol*, 5:284–288, 1998.
- [166] A. C. Gibbs, T. C. Bjorndahl, R. S. Hodges, and D. S. Wishart. Probing the structural determinants of type ii' beta-turn formation in peptides and proteins. *J.Am.Chem.Soc.*, 124(7):1203–1213, 2002.
- [167] S. J. Maness, S. Franzen, A. C. Gibbs, T.P. Causgrove, and R.B. Dyer. Nanosecond temperature jump relaxation dynamics of cyclic beta-hairpin peptides. *Biophys.J.*, 84:3874–3882, 2003.
- [168] W. F. van Gunsteren, S. Billeter, A. Eising, P. Hunenberg, P. Kruger, A. E. Mark, W. Scott, and I. Tironi. *Biomolecular simulations: the GROMOS96 manual and user guide*. Hochschulverlag an der ETH ZN—rich, ZN—rich, 1996.
- [169] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J.Comp.Chem.*, 18:1463–1472, 1997.
- [170] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Comput.Phys.Commun.*, 91:43–56, 1995.
- [171] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. *Interaction models for water in relation to protein hydration in "Intermolecular Forces"*. D. Reidel Publishing Company, Dordrecht, The Netherlands, 1981.

- [172] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An $n^2 \log(n)$ method for ewald sums in large systems. *J.Chem.Phys.*, 98:10089–10092, 1993.
- [173] D. Brown and J. H. R. Clarke. A comparison of constant energy, constant temperature and constant pressure ensembles in molecular dynamics simulations of atomic liquids. *Mol.Phys.*, 51:1243–1252, 1984.
- [174] A. Amadei, M. D' Abramo, I. Daidone, A. Di Nola M. D' Alessandro, and M. Aschi. Statistical mechanical modelling of chemical reactions in complex systems: the kinetics of the haem carbon monoxide binding-unbinding reaction in myoglobin. *Theor. Chem. Acc.*, 117:637–647, 2007.
- [175] M. Jackson and H. H. Mantsch. The use and misuse of ftir spectroscopy in the determination of protein structure. *Crit. Rev. Biochem. and Mol. Biol.*, 30:95 – 120, 1995.
- [176] A. Barth and C. Zscherp. What vibrations tell us about proteins. *Q. Rev. Biophys.*, 35:369–430, 2002.
- [177] C. Fang, A. Senes, L. Cristian, W. F. DeGrado, and R. M. Hochstrasser. Amide vibrations are delocalized across the hydrophobic interface of a transmembrane helix dimer. *Proc Natl Acad Sci USA*, 103:16740–16745, 2006.
- [178] K. Kwac and M. Cho. Molecular dynamics simulation study of n-methylacetamide in water. i. amide i mode frequency fluctuation. *J. Chem. Phys.*, 119:2247–2256, 2003.
- [179] J. R. Schmidt, S. A. Corcelli, and J. L. Skinner. Ultrafast vibrational spectroscopy of water and aqueous n-methylacetamide: Comparison of different electronic structure/molecular dynamics approaches. *J. Chem. Phys.*, 121:8887–8897, 2004.
- [180] M. Gageot, R. Vuilleumier, M. Sprik, and D. Borgis. Infrared spectroscopy of n-methylacetamide revisited by ab initio calculations. *J. Chem. Theory Comput.*, 1:772, 2005.
- [181] R. D. Gorbunov, P. H. Nguyen, M. Kobus, and G. Stock. Quantum-classical description of the amide i vibrational spectrum of trialanine. *J. Chem. Phys.*, 126:054509, 2007.
- [182] S. Krimm and J. Bandekar. Vibrational spectroscopy and conformation of peptides, polypeptides and proteins. *Adv.Protein Chem.*, 38:181–364, 1986.
- [183] H. Torii and M. Tasumi. Model calculations on the amide i infrared bands of globular proteins. *J.Chem.Phys.*, 96:3379–3387, 1992.

- [184] S. H. Lee and S. Krimm. Ab initio based vibrational analysis of alpha-ply(l-alanine). *Biopolymers*, 46:283–317, 1998.
- [185] A. Moran and S. Mukamel. The origin of vibrational mode couplings in various secondary structural motifs of polypeptides. *Proc.Natl.Acad.Sci. USA*, 101:506–510, 2004.
- [186] J. W. Brauner, C. R. Flach, and R. Mendelsohn. A quantitative reconstruction of the amide i contour in the ir spectra of globular proteins: from structure to spectrum. *J.Am.Chem.Soc.*, 127:100–109, 2005.
- [187] S. Yang and M. Cho. Ir spectra of n-methylacetamide in water predicted by combined quantum mechanical/molecular mechanical molecular dynamics simulations. *J.Chem.Phys.*, 123:134503, 2005.
- [188] P. Bour and T. A. Keiderling. Vibrational spectral simulation for peptides of mixed secondary structure: method comparisons with the trpzip model hairpin. *J.Phys.Chem.B*, 109:23687–23697, 2005.
- [189] A. E. Smith and A. Tokmakoff. Amide i two dimensional infrared spectroscopy of beta-hairpin peptides. *J.Chem.Phys*, 126:045109, 2007.
- [190] S. Ham, S. Cha, J. H. Choi, and M. Cho. Amide i modes of tripeptides: hessian matrix reconstruction and isotope effects. *J Chem.Phys.*, 119:1451, 2003.
- [191] J. H. Choi, H. Lee, K. K. Lee, S. Hahn, and M. Cho. Computational spectroscopy of ubiquitin: comparison between theory and experiments. *J.Chem.Phys.*, 126:045102, 2007.
- [192] C.Y. Huang, Z. Getahun, Y. Zhu, J.W. Klemke, W.F. DeGrado, and F. Gai. Helix formation via conformation diffusion search. *Proc. Natl. Acad. Sci. USA*, 99:2788–2793, 2002.
- [193] J. A. Ihalainen, J. Bredenbeck, R. Pfister, J. Helbing, L. Chi, I. H. M. van Stokkum, G. A. Woolley, and P. Folding and unfolding of a photoswitchable peptide from picoseconds to microseconds. *Proc Natl Acad Sci USA*, 104:5383–5388, 2007.
- [194] Y. Xu, R. Oyola, and F. Gai. Infrared study of the stability and folding kinetics of a 15-residue β -hairpin. *J. Am. Chem. Soc.*, 125:15388–15394, 2003.
- [195] A. Amadei, M. D’Abramo, C. Zazza, and M. Aschi. Electronic properties of formaldehyde: a theoretical study. *Chem.Phys.Lett.*, 381:187–193, 2003.

- [196] A. Amadei, M. D'Alessandro, M. D'Abramo, and M. Aschi. Theoretical characterization of electronic states in interacting chemical systems. *J. Chem. Phys.*, 130(8):084109, 2009.
- [197] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, 95:43–56, 1995.
- [198] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.
- [199] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.
- [200] D. Brown and J. H. R. Clarke. A comparison of constant energy, constant temperature, and constant pressure ensembles in molecular dynamics simulations of atomic liquids. *Mol. Phys.*, 51:1243–1252, 1984.
- [201] B. H. Besler, K. M. Merz, and P. A. Kollman. Atomic energies derived from semiempirical methods. *J Comp Chem*, 11:431–439, 1990.
- [202] J. Gao and D. G. Truhlar. Quantum mechanical methods for enzyme kinetics. *Ann.Rev.Phys.Chem*, 53:467–505, 2002.
- [203] T. Vreven and K. Morokuma. Hybrid methods: Oni-omqm:mm and qm/mm. *Ann.Rep.Comput.Chem.*, 2:35–52, 2006.
- [204] H. M. Senn and W. Thiel. Qm/mm studies of enzymes. *Curr.Opin.Chem.Biol.*, 11:182–187, 2007.
- [205] S. Ataka, H. Takeuchi, and M. Tasumi. Infrared studies of the less stable cis form of n-methylformamide and n-methylacetamide in low-temperature nitrogen matrices and vibrational analyses of the trans and cis forms of these molecules. *J. Mol. Struct.*, 113:147–160, 1984.
- [206] J. Kubelka and T. A. Keiderling. Ab initio calculation of amide carbonyl stretch vibrational frequencies in solution with modified basis sets. 1. n-methylacetamide. *J. Phys. Chem. A*, 105:10922–10928, 2001.
- [207] N. G. Mirkin and S. Krimm. Ab initio vibrational analysis of hydrogen-bonded trans- and cis-n-methylacetamide. *J. Am. Chem. Soc.*, 113:9742–9747, 1991.
- [208] S. K. Gregurick, G. M. Chaban, and R. B. Gerber. Ab initio and improved empirical potentials for the calculation of the anharmonic vibrational states and intramolecular mode coupling of n-methylacetamide. *J. Phys. Chem. A*, 106:8696–8707, 2002.

- [209] S. Song, S. A. Asher, S. Krimm, and K. D. Shaw. Ultraviolet resonance raman studies of trans and cis peptides: photochemical consequences of the twisted pi. *J. Am. Chem. Soc.*, 113:1155–1163, 1991.
- [210] Y. Xu, R. Oyola, , and F. Gai. Infrared study of the stability and folding kinetics of a 15-residue beta-hairpin. *J. Am. Chem. Soc.*, 125:15388–15394, 2003.
- [211] J. L. R. Arrondo, F. J. Blanco, L. Serrano, and F. M. Goni. Infrared evidence of a β -hairpin peptide structure in solution. *FEBS Lett.*, 384:35–37, 1996.
- [212] J. Bredenbeck, J. Helbing, J. R. Kumita, G. A. Woolley, and P. Hamm. α -helix formation in a photoswitchable peptide tracked from picoseconds to microseconds by time-resolved ir spectroscopy. *Proc. Natl. Acad. Sci. USA*, 102:2379–2384, 2005.
- [213] P. Hamm, M. Lim, and R. M. Hochstrasser. Structure of the amide i band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem B*, 102:6123–6138, 1998.
- [214] T. Hayashi and S. Mukamel. Vibrational-exciton couplings for the amide i, ii, iii and a. modes of peptides. *J. Phys. Chem. B*, 111:11032–11046, 2007.
- [215] L. P. DeFlores, Z. Ganim, R. A. Nicodemus, and A. Tokmakoff. Amide i'-ii' 2d ir spectroscopy provides enhanced protein secondary structural sensitivity. *J. Am. Chem. Soc.*, 131:3385–3391, 2009.
- [216] V. Munoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton. Folding dynamics and mechanism of beta-hairpin formation. *Nature*, 390:196–199, 1997.
- [217] V. Munoz, E. R. Henry, J. Hofrichter, and W. A. Eaton. A statistical mechanical model for beta-hairpin kinetics. *Proc. Natl. Acad. Sci. U.S.A.*, 95:5872–5879, 1998.
- [218] R. P. Chen, J. J. Huang, H. L. Chen, H. Jan, M. Velusamy, C. T. Lee, W. Fann, R. W. Larsen, and S. I. Chan. Measuring the refolding of beta-sheets with different turn sequences on a nanosecond time scale. *Proc. Natl. Acad. Sci. U.S.A.*, 101:7305–7310, 2004.
- [219] A. R. Dinner, T. Lazaridis, and M. Karplus. Understanding beta-hairpin formation. *Proc. Natl. Acad. Sci. U.S.A.*, 96:9068–9073, 1999.
- [220] D. K. Klimov and D. Thirumalai. Mechanisms and kinetics of beta-hairpin formation. *Proc. Natl. Acad. Sci. U.S.A.*, 97:2544–2549, 2000.
- [221] A. E. Garcia and K. Y. Sanbonmatsu. Exploring the energy landscape of a beta-hairpin in explicit solvent. *Proteins: Struct., Funct., Genet.*, 42:345–354, 2001.

- [222] R. Zhou, B. J. Berne, and R. Germain. The free energy landscape for beta-hairpin folding in explicit water. *Proc.Natl.Acad.Sci.U.S.A.*, 98:14931–14936, 2001.
- [223] D.K. Klimov and D. Thirumalai. Stiffness of the distal loop restricts the structural heterogeneity of the transition state ensemble in sh3 domains. *J. Mol. Biol.*, 315:721–737, 2002.
- [224] Y. Zhou and A. Linhananta. Role of hydrophilic and hydrophobic contacts in folding of the second beta-hairpin fragment of protein g: Molecular dynamics simulation studies of an all-atom model. *Proteins:Struct.,Funct.,Genet.*, 47:154–162, 2002.
- [225] B. Ma and R. Nussinov. Energy landscape and dynamics of the beta-hairpin g peptide and its isomers: Topology and sequences. *Protein Sci.*, 12:1882–1893, 2003.
- [226] P. G. Bolhuis. Transition-path sampling of beta-hairpin folding. *Proc.Natl.Acad.Sci.U.S.A.*, 100:12129–12134, 2003.
- [227] C. D. Snow, L. Qiu, D. Du, F. Gai, S. J. Hagen, and V. S. Pande. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc.Natl.Acad.Sci.U.S.A.*, 101:4077–4082, 2004.
- [228] D. Roccatano, A. Amadei, A. Di Nola, and H. J. C. Berendsen. A molecular dynamics study of the 41-56 beta-hairpin from b1 domain of protein g. *Protein Sci.*, 10:2130–2143, 1999.
- [229] V. S. Pande and D. S. Rokhsar. Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein g. *Proc.Natl.Acad.Sci.U.S.A.*, 96:9062–9067, 1999.
- [230] B. Zagrovic, E. J. Sorin, and V. Pande. beta-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol.Biol.*, 313:151–169, 2001.
- [231] L. Thukral, J. C. Smith, and I. Daidone. Common folding mechanism of a beta-hairpin peptide via non-native turn formation revealed by unbiased molecular dynamics simulations. *J.Am.Chem.Soc.*, In Press, 2009.
- [232] L. Zanetti-Polzi, M. Anselmi, M. D’Alessandro, A. Amadei, and A. Di Nola. Structural, thermodynamic and kinetic properties of gramicidin analogue gs6 studied bt molecular dynamics simulations and statistical mechanics. *Biopolymers*, 91(12):1154–1160, 2009.

- [233] R. Schweitzer-Stenner. Advances in vibrational spectroscopy as a sensitive probe of peptide and protein structure: A critical review. *Vibr. Spectrosc.*, 42:98–117, 2006.
- [234] A. Amadei, M. E. F. Apol, A. Di Nola, and H. J. C. Berendsen. The quasi-gaussian entropy theory: Free energy calculations based on the potential energy distribution function. *J.Chem.Phys.*, 104:1560–1574, 1996.
- [235] A. Amadei, M. E. F. Apol, and H. J. C. Berendsen. On the use of the quasi-gaussian entropy theory in noncanonical ensembles. i. prediction of temperature dependence of thermodynamic properties. *J.Chem.Phys.*, 109:3004–3016, 1998.
- [236] M. E. F. Apol, A. Amadei, and H. J. C. Berendsen. On the use of the quasi-gaussian entropy theory in noncanonical ensembles. ii. prediction of density dependence of thermodynamic properties. *J.Chem.Phys.*, 109:3017–3027, 1998.
- [237] A. Segre, R.C. Bachmann, A. Ballio, F. Bossa, I. Grgurina, N.S. Iacobellis, G. Marino, P. Pucci, M. Simmaco, , and J.Y. Takemoto. The structure of syringomycins a1, e and g. *FEBS Lett.*, 255:27–31, 1989.
- [238] N. Fukuchi, A. Isogai, J. Nakayama, S. Takayama, S. Yamashita, K. Suyama, J.Y. Takemoto, and A. Suzuki. Structure and stereochemistry of three phyto-toxins syringomycin, syringostatin and syringotoxin produced by pseudomonas syringae pv. syringae. *J. Chem. Perkin. Trans.*, 1:1149–1157, 1992.
- [239] A. Scaloni, R.C. Bachmann, J.Y. Takemoto, D. Barra, M. Simmaco, and A. Ballio. Stereochemical structure of syringomycin, a phytotoxic metabolite of pseudomonas syringae pv. syringae. *Nat. Prod. Lett.*, 4:159–164, 1994.
- [240] A. Ballio, F. Bossa, A. Collina, M. Gallo, N.S. Iacobellis, M. Paci, P. Pucci, A. Scaloni, A. Segre, and M. Simmaco. Structure of syringotoxin, a bioactive metabolite of pseudomonas syringae pv. syringae. *FEBS Lett.*, 269:377–380, 1990.
- [241] A. Ballio, F. Bossa, D. Di Giorgio, P. Ferranti, M. Paci, P. Pucci, A. Scaloni, A.L. Segre, and G.A. Strobel. Novel bioactive lipodepsipeptides from pseudomonas syringae: the pseudomycins. *FEBS Lett.*, 355:96–100, 1994.
- [242] B.K. Scholz-Schroeder, M.L. Hutchison, I. Grgurina, and D.C. Gross. The contribution of syringopeptin and syringomycin to virulence of pseudomonas syringae pv. syringae strain b 301 d on the basis of sypa and syrB1 biosynthesis mutant analysis. *Mol. Plant Microbe Interact.*, 14:336–348, 2001.
- [243] I. Grgurina, A. Barca, S. Cervigni, M. Gallo, A. Scaloni, and P. Pucci. Relevance of chlorine-substituent for the antifungal activity of syringomycin and

- syringotoxin, metabolites of the phytopathogenic bacterium *Pseudomonas syringae* pv. *syringae*. *Experientia*, 50:130–133, 1994.
- [244] A.J. De Lucca, T.J. Jacks, J.Y. Takemoto, B. Vinyard, J. Peter, E. Navarro, , and T.J. Walsh. Fungal lethality, binding, and cytotoxicity of syringomycin-e. *Antimicrob. Agents Chemother.*, 43:371– 373, 1999.
- [245] A.J. De Lucca and T.J. Walsh. Antifungal peptides: novel therapeutic compounds against emerging pathogens. *Antimicrob. Agents Chemother.*, 43:1–11, 1999.
- [246] M.L. Hutchison and D.C. Gross. Lipopeptide phytotoxin produced by *Pseudomonas syringae* pv. *syringae*: comparison of the biosurfactant and ion channel-forming activities of syringopeptin and syringomycin. *Mol. Plant Microbe Interact.*, 10:347–354, 1997.
- [247] E. Matyus, L. Monticelli, K.E. Kover, Z. Xu, K. Blasko, J. Fidy, and D.P. Tieleman. Structural investigation of syringomycin-e using molecular dynamics simulation and nmr. *Eur. Biophys. J.*, 2006.
- [248] D. van der Spoel, R. van Drunen, and H. J. C. Berendsen. *GRONINGEN MACHINE FOR CHEMICAL SIMULATION*. Department of Biophysical Chemistry, BIOSON Research Institute, Nijenborgh 4 NL-9717 AG Groningen, 1994.
- [249] R. Kaptein, R. Boelens, R.M. Scheek, and W.F. van Gunsteren. Protein structures from nmr. *Biochemistry*, 27:5389–5395, 1988.
- [250] F. Bossa A. Ballio, D. Di Giorgio, A. Di Nola, C. Manetti, M. Paci, A. Scaloni, and A.L. Segre. Solution conformation of the *Pseudomonas syringae* pv. *syringae* phytotoxic lipodepsipeptide sp25a: 2d nmr , distance geometry and molecular dynamics. *Eur. J. Biochem.*, 234:747–758, 1995.
- [251] V.M. Coiro, A.L. Segre, A. Di Nola, M. Paci, A. Grottesi, G. Veglia, and A. Ballio. Solution conformation of the *Pseudomonas syringae* msu 16h phytotoxic lipodepsipeptide pseudomycin a determined by computer simulations using distance geometry and molecular dynamics from nmr data. *Eur. J. Biochem.*, 257:449–456, 1998.
- [252] S. Bare, V. Coiro, A. Scaloni, A. Di Nola, M. Paci, A. Segre, and A. Ballio. Conformations in solution of the fuscopeptins, phytotoxic metabolites of *Pseudomonas fuscovaginae*. *Eur. J. Biochem.*, 266:1–10, 1999.
- [253] C.A.G. Hassnoot, F.A.A.M. Leeuw, and A. Altona. The relationship between proton-proton nmr coupling constants and substituent electronegativities: An empirical generalization of the Karplus equation. *Tetrahedron*, 36:2783–2792, 1998.

-
- [254] E. Vaillo, A. Ballio, P.L. Luisi, and R.M. Thomas. The spectroscopic properties of the lipodepsipeptide syringomycin e. *Biopolymers*, 32:1317–1326, 1992.
- [255] R.J. Mortishire-Smith, J.C. Nutkins, L.C. Packman, C.L. Brodey, P.B. Rainey, K. Johnstone, and D.H. Williams. Determination of the structure of an extracellular peptide produced by mushroom saprotroph *Pseudomonas reactivans*. *Tetrahedron*, 47:3647–3654, 1991.
- [256] V.V. Malev, L.V. Schagina, P.A. Gurnev, J.Y. Takemoto, E.M. Nestorovich, and S. M. Bezrukov. Syringomycin e channel: a lipidic pore stabilized by lipopeptide? *Biophys. J.*, 82:1985–1994, 2002.
- [257] O.S. Ostroumova, YA. Kaulin, and L.V. Schagina. Effect of agents modifying the membrane dipole potential on properties of syringomycin e channels. *Langmuir*, 23:6889–6892, 2007.

Acknowledgments

I am grateful to all the people who supported me during these years: above all to Prof. Di Nola, for his contribution to my formation, and Andrea Amadei, for his irreplaceable and precious guidance during my Ph.D. I would also like to thank Massimiliano Aschi, for his invaluable collaboration, and Isabella Daidone, for sharing with me her experience and scientific passion, and the other members of the group.

A special acknowledgment to the coordinator of the Ph. D. course in Biophysics, Prof. Colosimo.

Thanks to my friends and my family.