Università degli Studi di Milano

Facoltà di Scienze e Tecnologie

Dipartimento di Chimica

PhD in Chemistry

XXXI Cycle

Implementation of Semiclassical Theories for Spectroscopy

Advisor:
Prof. Michele Ceotto

Ph.D. Candidate:

Fabio Gabas

Id. N. R11309

Academic year 2017/2018

Referees:

Prof. David Coker, from Boston University

Prof. Nandini Ananth, from Cornell University

Examination Board:

Prof. Chiara Castiglioni, from Politecnico di Milano

Prof. Dmitry Shalashilin, from University of Leeds

Prof. Leonardo Lo Presti, from Università degli Studi di Milano

Final exam:

Date: January, 22nd, 2019

Università degli Studi di Milano, Dipartimento di Chimica, Milano, Italy.

Contents

[1]This chapter is the reproduction with some minor changes of the paper **Fabio Gabas**, Riccardo Conte, and Michele Ceotto. "On-the-fly ab initio semiclassical calculation of glycine vibrational spectrum." *Journal of chemical theory and computation* **13.6** (2017): 2378-2388.

[2]This chapter is the reproduction (with some changes) of the paper "Protonated Glycine Supramolecular Systems: the need for quantum dynamics" *Chemical Science*, **9**, (2018): 7894-7901 by **Fabio Gabas**, Giovanni Di Liberto, Riccardo Conte, and Michele Ceotto. This paper has been selected as "Pick of the Week" by the editorial board of Chemical Science. Furthermore, I have prepared a cover art image that has been chosen as front image for an issue of Chemical Science.

Introduction

Vibrational spectroscopy constitutes a cornerstone of modern chemistry because it represents a fundamental tool for understanding and characterizing the chemical composition of a molecule. This is due to the fact that nuclear vibrations are influenced by their positions in the molecule, by the presence of certain functional groups and by the presence of a solvent. A complete map of the vibrations is useful because it may serve as a molecular fingerprint. [1, 2] Two main techniques have been developed to detect the presence and type of molecular vibrations: infrared (IR) and Raman spectroscopy. The IR exploits the absorption of infrared radiation, while Raman spectroscopy is usually based on the inelastic scattering of visible light. Experimentally a lot of advances have been achieved to overcome the limitations of early linear absorption infrared spectroscopy. Thanks to the pump–probe nonlinear 2-dimensional infrared spectroscopy (2D-IR), to the coupling of mass spectroscopy (MS) with the Infrared Multi Photon Dissociation (IR-MPD) and to various ion trapping strategies it is now possible to record vibrational spectra of systems and ions in real time and at room temperature, close to relevant physiological temperatures of biomolecules, either in gas or liquid phase.[3, 4, 5, 6, 7] A crucial point for the experiment is represented by the correct assignment of the vibrational modes, expecially for high dimensional systems where the spectrum presents a large number of signals. Here the theoretical calculations can play a determinant role.

From a theoretical point of view there are different ways to predict and assign vibrational spectra. The most direct one is the harmonic calculation, which approximates the potential energy surface in the neighborhood of a minimum with a harmonic potential. Frequencies are obtained from the eigenvalues of the Hessian matrix at the equilibrium geometry. In order to be compared with the experiment such calculation needs to be scaled using empirical scaling factors that help to recover the anharmonicity originally neglected.[8, 9] There are other methods that include anharmoncities, within variational, like for example the vibrational self-consistent field (VSCF) method, or perturbative frameworks, like the vibrational second-order perturbation theory (VPT2). [10, 11] Requiring an optimized minimum, all these static approaches can be easily applied to gas phase molecules but, on the other hand, in condensed phase a minimum structure is difficult to predict. Other kinds of systems, where a single optimized minimum is not easy to find nor very representative, are the small peptides, that present different conformers lying in a small energy range. Modern experimental setups that work at room temperature can readily obtain spectra of peptides exploring more than one conformation, and in these cases a comparison with static or harmonic theoretical predictions can lead to misleading results.

For a correct description of such systems the adoption of a dynamically based

approach seems to be mandatory. Here comes the contribution of molecular dynamics (MD) which permits exploration of more than one minimum on the potential energy surface and that can describe the presence of solvents. Usually the dynamics is evaluated using an ab-initio method to calculate the electronic energy and consequently an autocorrelation function is derived to obtain a vibrational spectrum. The advent of density functional theory (DFT) has permitted to investigate molecular systems at reasonable computational costs, expecially when associated to Car-Parrinello molecular dynamics (CPMD).[12] However, real-time dynamical quantum effects, which are relevant for certain systems or when the experiments present very low temperatures, are missing in all these strategies.

Semiclassical methods fill this gap brilliantly.[13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24] They are based on classical molecular dynamics but they employ such information to compute the autocorrelation function of a reference state through an approximation to the Feynman quantum propagator able to provide quantum effects. In addition to positions, momenta and energy along the trajectory, semiclassical methods require also to evaluate the Hessian matrix at each step, making application of semiclassical dynamics more computationally demanding with respect to the previously presented strategies. Also for this reason, until a decade ago, semiclassical applications were limited to small size molecules, like water, carbon dioxide and formaldehyde.[25, 26] Thanks to recent facilities, approximations and theory developments, it is now possible for the semiclassical methods to handle systems of increasing dimensionality.

Within this thesis work I will first describe these advances, and then I apply them to molecules with an increasing number of atoms, reaching a maximum of 46 for a single isolated molecule or 37 in a supramolecular systems. The aim of this work is twofold. The first is to demonstrate the validity of semiclassical approaches in accurately detecting the vibrational frequencies of systems of such kind. The second one instead concerns the importance of including quantum effects. Moving towards systems with many atoms, in fact, the most commonly employed methodology is classical molecular mechanics (MM), which neglects all quantum effects, considering their role in such big size systems as marginal. In some other cases, like for example in QM/MM methods, the totality of the degrees of freedom is partitioned into different areas of interest to be treated either with classical or quantum level of theory. In this thesis work I will apply the semiclassical method to high dimensional systems with the purpose to demonstrate that it is possible, and in some instances even necessary, to include quantum contributions. The semiclassical method can successfully fullfill this task considering the whole system homogeneously, differently to the previously

described QM/MM like methods.

In the next section of this thesis I will briefly describe the semiclassical theory foundation together with the last innovative formulations. Furthermore, in the same section I will also provide some elements of classical molecular dynamics, evaluated either through force fields or with ab initio molecular dynamics. Then I move to some applications, and finally I will conclude with some considerations and future developments.

Part I.

Theoretical basis

1. Theory of Semiclassical Spectroscopy

In this chapter the theoretical basis of the semiclassical method applied to vibrational spectroscopy will be illustrated. Starting from the first formulations of the van Vleck-Gutzwiller and Herman-Kluk propagators, this section will end up with the description of the most recently developed improvements and approximations. Exploiting these innovative formulations, the study of isolated molecules characterized by different local minima along with complex supramolecular and quasi-solvated systems has been possible. Outcomes are reported in the section called "Applications to molecular systems".

1.1. The quantum propagator

The dynamics of atomic-scale molecular systems is carefully described by the Time Dependent Schröedinger Equation (TDSE):

$$i\hbar \frac{\partial |\psi(t)\rangle}{\partial t} = \hat{H} |\psi(t)\rangle \tag{1.1}$$

where $\hat{H}$ is the Hamiltonian, a Hermitian operator composed by the sum of the kinetic energy operator $\hat{T}$ and the potential energy operator $\hat{V}$, while $|\psi(t)\rangle$ represents the wavefunction of the system. Starting from the TDSE, it is possible to derive the Feynman path integral formalism by means of the quantum propagator operator, which specifies how the system propagates from the initial to the final state.[27] In the Appendix 1.a the equivalence of these two formulations is reported.

The wavefunction in eq. 1.1 can be express through a coordinate representation

$$\psi(x,t) = \langle x|e^{-\frac{i}{\hbar}\widehat{H}t}|\psi(x,0)\rangle \tag{1.2}$$

where $|\psi(x,0)\rangle$ stands for the wavefunction in its initial conditions, defined for eq. 1.1. The quantum propagation is represented by the right hand side of equation 1.2. In the following lines the propagator will be written as the sum of the probability amplitude matrix $\langle x|e^{-\frac{i}{\hbar}\widehat{H}t}|x_0\rangle$ times the time vector $\langle x_0|\psi(x,0)\rangle$ over all the possible paths that go from $(x,0)$ to (x,t).

The start is the generic definition of the quantum propagator,

$$U(x_N,x_0,t) = \langle x_N|e^{-\frac{i}{\hbar}\widehat{H}t}|x_0\rangle \tag{1.3}$$

where x_0 and x_N are two states of the system and is valid the assumption that the Hamiltonian operator, $H = \frac{p^2}{2m} + V(x)$, is time independent. It is certainly true that

$$e^{-\frac{i}{\hbar}\widehat{H}t} = \left[e^{-\frac{i}{\hbar}\widehat{H}\frac{t}{N}}\right]^N \tag{1.4}$$

stating that the propagator $U(t)$ for a time t is equal to the product of N propagators of time $\frac{t}{N}$, $U(\frac{t}{N})$. Considering now the limit $N \to \infty$ and defining

$$\varepsilon = \frac{t}{N}, \tag{1.5}$$

it is possible to write

$$\left[e^{-\frac{i\varepsilon}{\hbar}(\frac{p^2}{2m}+V(x))}\right]^N \simeq \left[e^{-\frac{i\varepsilon}{2m\hbar}p^2}\cdot e^{-\frac{i\varepsilon}{\hbar}V(x)}\right]^N \tag{1.6}$$

by neglecting the commutator of the $\hat{p}$ and $\hat{V}$ operators. Therefore, there is the product of N terms to compute

$$\langle x_N|e^{-\frac{i\varepsilon}{2m\hbar}p^2}\cdot e^{-\frac{i\varepsilon}{\hbar}V(x)}e^{-\frac{i\varepsilon}{2m\hbar}p^2}\cdot e^{-\frac{i\varepsilon}{\hbar}V(x)}...|x_0\rangle\,. \tag{1.7}$$

For this purpose, inserting $N-1$ times the identity in the form

$$I = \int_{-\infty}^{+\infty} \mathrm{d}x \, |x\rangle \, \langle x| \tag{1.8}$$

the product becomes

$$\begin{aligned} U(x_N, x_0, t) = \int_{-\infty}^{+\infty} \prod_{n=1}^{N} \mathrm{d}x_n \, \langle x_N | e^{-\frac{i\varepsilon}{2m\hbar}P^2} \cdot e^{-\frac{i\varepsilon}{\hbar}V(x)} | x_{N-1} \rangle \times \\ \langle x_{N-1} | e^{-\frac{i\varepsilon}{2m\hbar}P^2} \cdot e^{-\frac{i\varepsilon}{\hbar}V(x)} | x_{N-2} \rangle \ldots \langle x_1 | e^{-\frac{i\varepsilon}{2m\hbar}P^2} \cdot e^{-\frac{i\varepsilon}{\hbar}V(x)} | x_0 \rangle . \end{aligned} \tag{1.9}$$

Looking now to the matrix element

$$\langle x_n | e^{-\frac{i\varepsilon}{2m\hbar}P^2} \cdot e^{-\frac{i\varepsilon}{\hbar}V(x)} | x_{n-1} \rangle \tag{1.10}$$

and let the operator $V(x)$ acts upon $|x_{n-1}\rangle$ to give

$$\langle x_n | e^{-\frac{i\varepsilon}{2m\hbar}P^2} | x_{n-1} \rangle \, e^{-\frac{i\varepsilon}{\hbar}V(x_{n-1})} \tag{1.11}$$

it is possible to notice that the remaining matrix element represents the free particle propagator from x_{n-1} to x_n in time ε, that has the following expression, (see Appendix 1.b for the mathematical derivation)

$$\langle x_n | e^{-\frac{i\varepsilon}{2m\hbar}P^2} | x_{n-1} \rangle = \left(\frac{m}{2\pi i\hbar\varepsilon} \right)^{\frac{1}{2}} e^{\frac{im(x_n - x_{n-1})^2}{2\hbar\varepsilon}}. \tag{1.12}$$

Substituting the eq 1.12 in the 1.11 it is possible to obtain this expression for the 1.10

$$\left(\frac{m}{2\pi i\hbar\varepsilon} \right)^{\frac{1}{2}} e^{\frac{im(x_n - x_{n-1})^2}{2\hbar\varepsilon}} e^{-\frac{i\varepsilon}{\hbar}V(x_{n-1})} \tag{1.13}$$

and taking into account all the N terms of the product the propagator assumes the following formulation

$$\begin{aligned} U(x_N, x_0, t) = \left(\frac{m}{2\pi i\hbar\varepsilon} \right)^{\frac{1}{2}} \left[\int_{-\infty}^{+\infty} \prod_{n=1}^{N-1} \left(\frac{m}{2\pi i\hbar\varepsilon} \right)^{\frac{1}{2}} \mathrm{d}x_n \right] \times \\ \sum_{e^{n=1}}^{N} \frac{im(x_n - x_{n-1})^2}{2\hbar\varepsilon} - \frac{i\varepsilon}{\hbar}V(x_{n-1}) \\ . \end{aligned} \tag{1.14}$$

In this expression is present in the integrand the discretized version of $e^{\frac{iS}{\hbar}}$, where S is the classical action of the system. Therefore moving on the continuum version the previous equation becomes

$$U(x_N, x_0, t) = \int [\mathcal{D}x]\, e^{\frac{i}{\hbar}S(x,x',t)} \tag{1.15}$$

where

$$\int [\mathcal{D}x] = \lim_{N\to\infty} \left(\frac{m}{2\pi i\hbar\varepsilon}\right)^{\frac{1}{2}} \left[\int_{-\infty}^{+\infty} \prod_{n=1}^{N-1} \left(\frac{m}{2\pi i\hbar\varepsilon}\right)^{\frac{1}{2}} \mathrm{d}x_n\right]. \tag{1.16}$$

The expression reported in eq. 1.15 is known as Configuration Space Path Integral, or simply Path Integral, and it is very important since it can provide us with a picture of the connection between the macroscopic world made of smooth paths and the molecular scenario, governed by quantum rules.

Another possible representation of the path integral is derived starting from eq. 1.7 and inserting the following identity relations between every exponential:

$$I = \int_{-\infty}^{+\infty} \mathrm{d}x\, |x\rangle \langle x| \tag{1.17}$$

$$I = \int_{-\infty}^{+\infty} \frac{\mathrm{d}p}{2\pi\hbar}\, |p\rangle \langle p| \tag{1.18}$$

where the relation

$$\langle x|p\rangle = e^{\frac{ipx}{\hbar}} \tag{1.19}$$

holds. Similarly it is possible to compute the same mathematical steps from the derivation above

$$\begin{aligned} U(x_N, x_0, t) = \int_{-\infty}^{+\infty} [\mathcal{D}p]\,[\mathcal{D}x]\, \langle x_n|e^{-\frac{i\varepsilon}{2m\hbar}P^2}|p_n\rangle \langle p_n|e^{-\frac{i\varepsilon}{\hbar}V(x)}|x_{n-1}\rangle \times \\ \langle x_{n-1}|e^{-\frac{i\varepsilon}{2m\hbar}P^2}|p_{n-1}\rangle \langle p_{n-1}|e^{-\frac{i\varepsilon}{\hbar}V(x)}|x_{n-2}\rangle \ldots \\ \ldots \langle x_1|e^{-\frac{i\varepsilon}{2m\hbar}P^2}|p_1\rangle \langle p_1|e^{-\frac{i\varepsilon}{\hbar}V(x)}|x_0\rangle \end{aligned} \tag{1.20}$$

where

$$\int_{-\infty}^{+\infty} [\mathcal{D}p]\,[\mathcal{D}x] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \prod_{n=1}^{N} \frac{dp_n}{2\pi\hbar} \prod_{n=1}^{N} dx_n. \tag{1.21}$$

In this formulation every exponential operator can be applied directly to the eigenstates to its own right. Collecting all the eigenvalues finally the propagator becomes

$$U(x_N, x_0, t) = \int_{-\infty}^{+\infty} [\mathcal{D}p]\,[\mathcal{D}x]\, e^{\sum_{n=1}^{N}\left[\frac{-i\epsilon}{2m\hbar}p_n^2 + \frac{i}{\hbar}p_n(x_n - x_{n-1}) - \frac{i\epsilon}{\hbar}V(x_{n-1})\right]} \tag{1.22}$$

and the following version moving instead on the continuous

$$U(x_N, x_0, t) = \int_{-\infty}^{+\infty} [\mathcal{D}p]\,[\mathcal{D}x]\, e^{\frac{i}{\hbar}S} \tag{1.23}$$

that is known as the Phase Space Path Integral formulation for the propagator and it will have a paramount importance as the foundation of the semiclassical approximation.

Indeed, equation 1.23 is exact in principle, but it requires an integration over all the possible paths that lead from the initial condition to the final one. Such condition is clearly difficult to fulfill, but calculations can be simplified without significant loss of accuracy by means of the semiclassical approximation to the propagator.

1.2. The semiclassical approximation to the propagator

One way to solve the Path Integral is to use the so-called "stationary phase approximation".[28] In order to familiarise with such a mathematical tool it is possible to start looking at the stationary phase approximation applied to an oscillatory one-dimension integration

$$\int_{-\infty}^{+\infty} \mathrm{d}x\, e^{if(x)} \approx \sum_{\{x_j | f'(x_j)=0\}} \sqrt{\frac{2\pi i}{f''(x_j)}}\, e^{if(x_j)} \tag{1.24}$$

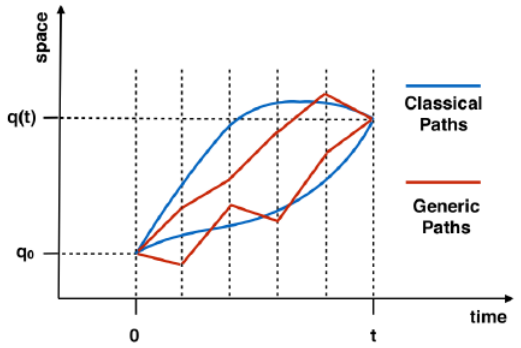

Figure 1.1.: Representation of Feynman path integral integration. Reproduced from ref [29]

that can be readily extended to the multidimensional case by substituting the second derivative with the determinant of the matrix of second derivatives. Comparing the left hand side of equation 1.24 with 1.15, it is possible to conclude that the stationary phase approximation to quantum propagator is applied by setting to zero the first derivative of the classical action. Because of Hamilton's principle, the first derivative of the action is zero for any classical path, so such a mathematical operation implies that the summation over all the possible paths, depicted in red in Figure 1.1 is now reduced only to classical paths, reported in blue in the same Figure.

Starting from

$$U(x_N, x_0, t) \simeq \sum_{classical\ paths} \int [\mathcal{D}x]\, e^{\frac{i}{\hbar} S(x_0, x_N, t)} \tag{1.25}$$

the action can be expressed through an expansion to the second order, $S \simeq S^0 + \delta S + \frac{\delta^2 S}{2}$, where the term S^0 is the action depending on x_0 and x_N , which are respectively the initial and final condition of each classical trajectory. Considering that for classical path is true that $\delta S = 0$, a Gaussian integral is obtained, whose calculation leads to the van Vleck version of the semiclassical propagator [30]

$$U_{VV} \simeq \sum_{classical\ paths} \left(\frac{1}{2\pi i\hbar}\right)^{\frac{N}{2}} \left| -\frac{\delta^2 S}{\delta x_0 \delta x_N} \right|^{\frac{1}{2}} e^{\frac{i}{\hbar} S^0(x_0, x_N, t)}. \tag{1.26}$$

Eq. 1.26 describes a scenario in which the zero-th order term of the classical action gives an oscillatory contribution, due to the exponential factor, while the

second order term, contained in the pre-exponential prefactor and known as van Vleck determinant, describes the quantum fluctuations around the classical path. In that formulation the prefactor is composed by a second derivative that depends on x_0 and x_N. The sign of that determinant depends on the number of times in which the classical trajectory goes through a caustic point. This is accounted for by the Maslow index ν. Starting from this consideration, Gutzwiller factored out the phase of the prefactor giving the following new implementation, known as the van Vleck-Gutzwiller propagator [31]

$$U_{VVG} \simeq \sum_{classical\ paths} \left(\frac{1}{2\pi i\hbar}\right)^{\frac{N}{2}} \left| -\frac{\delta^2 S}{\delta x_0 \delta x_N} \right|^{\frac{1}{2}} e^{\frac{i}{\hbar} S^0(x_0, x_N, t)} e^{-\frac{1}{2}\pi\nu}. \tag{1.27}$$

Certainly, this formulation would be easier to manage and calculate if the quantities depended only on initial conditions rather than having initial and final states simultaneously. For this reason, by employing Hamilton's equations $\dot{p}_t = -\frac{\delta H}{\delta x_t}$ and $\dot{x}_t = \frac{\delta H}{\delta p_t}$, the previous formula can be written moving from (x_0, x_N) to (x_0, p_0)

$$U_{VVG} \simeq \sum_{classical\ paths} \left[\frac{1}{2\pi i\hbar} \times \left(\frac{\delta x_N}{\delta p_0}\right)^{-1}\right]^{\frac{1}{2}} e^{\frac{i}{\hbar} S^0(x_0, p_0, t)} e^{-\frac{1}{2} i\pi\nu} \tag{1.28}$$

where x_0 and p_0 are the initial positions and momenta of the classical trajectories and the factor $e^{-\frac{1}{2}\pi\nu}$ accounts for the fact that the classical trajectory can pass through a caustic point that brings to a change in the sign of the determinant. The last formula represents the van Vleck-Gutzwiller semiclassical propagator, in its Initial Value Representation (IVR).[32, 33] Note that the prefactor now has the form $(\frac{\delta x_N}{\delta p_0})^{-1}$, which represents the inverse of the sensitivity of the final position with respect to the initial momentum. It is an element of the stability matrix, or monodromy matrix

$$\mathbf{M} = \begin{pmatrix} \partial p_t/\partial p_0 & \partial p_t/\partial x_0 \\ \partial x_t/\partial p_0 & \partial x_t/\partial x_0 \end{pmatrix}. \tag{1.29}$$

The monodromy matrix is a measure of how sensitive the trajectory is to the initial conditions. In particular, in Eq. 1.28, the larger the prefactor the smaller the weight of that trajectory. The importance of the prefactor and its criticality in the stability of the calculation will be successively discussed. A brief list of all the ways that have been address to solve its complexities and approximate its role is presented in Appendix 1.d.

Another possible expression for the propagator is formulated by employing the coherent states $|\mathbf{p_t}, \mathbf{q_t}\rangle$. In theoretical vibrational spectroscopy working in normal modes $(\mathbf{p}, \mathbf{q})$ is a convenient choice and using such formalism, in the coordinate representation coherent states have a Gaussian shape of this kind

$$\langle x|\mathbf{p}_t, \mathbf{q}_t\rangle = \left(\frac{\det\gamma}{\pi^N}\right)^{\frac{1}{4}} exp\left[-\frac{1}{2}(x-\mathbf{q}_t)^T\gamma(x-\mathbf{q}_t) + \frac{i}{\hbar}\mathbf{p}_t^T(x-\mathbf{q}_t)\right]. \quad (1.30)$$

This device was exploited by Heller, Herman and Kluk that formulated, in such a way, a new expression of the IVR propagator:[34, 35]

$$U_{HK} = \left(\frac{1}{2\pi\hbar}\right)^N \iint \mathrm{d}\mathbf{p}_0\mathrm{d}\mathbf{q}_0 C_t\left(\mathbf{p}_0, \mathbf{q}_0\right) e^{\frac{i}{\hbar}S_t(\mathbf{p}_0,\mathbf{q}_0)} \langle x|\mathbf{p}_t, \mathbf{q}_t\rangle \langle \mathbf{p}_0, \mathbf{q}_0|x'\rangle \quad (1.31)$$

where the integration is done on the starting $\mathbf{q}_0$ and $\mathbf{p}_0$ conditions, $\langle x|\mathbf{p}_t, \mathbf{q}_t\rangle \langle \mathbf{p}_0, \mathbf{q}_0|x'\rangle$ is the product of the coherent states, γ is the matrix of the width parameters of the Gaussian wavepackets and $C_t\left(\mathbf{p}_0, \mathbf{q}_0\right)$ is the pre-exponential factor that can be evaluated by means of the following expression:

$$C_t\left(\mathbf{p}, \mathbf{q}\right) = \sqrt{\det\left[\frac{1}{2}\left(\mathbf{M_{qq}} + \gamma^{-1}\mathbf{M_{pp}}\gamma + \frac{i\gamma^{-1}}{\hbar}\mathbf{M_{pq}} + \frac{\hbar\gamma}{i}\mathbf{M_{qp}}\right)\right]} \quad (1.32)$$

where the matrices $\mathbf{M}$ are the elements of the Monodromy Matrix.

The importance of the Herman-Kluk (HK) formulation is due to the fact that using Gaussian wavepackets the integration becomes amenable to numerical calculations. Therefore, by means of an integration algorithm like the Box Muller [36], it is possible to sample the phase-space initial conditions and to apply a Monte Carlo method to compute the integration.

1.3. Application to vibrational spectroscopy

The power spectrum of a wavepacket, for a bound state problem, is given by the expression

$$\sigma(\omega) = \sum_n |c_n|^2\delta(\omega-\omega_n) \quad (1.33)$$

where c_n are the coefficients of the wavepacket relative to the Hamiltonian eigenfunctions, $\Psi(x,t) = \sum_n c_n\psi_n(x)e^{-\frac{i}{\hbar}E_nt}$, and $\omega_n = \frac{E_n}{\hbar}$. Following this formulation,

the spectrum is simply the sum of the absolute squares of the energy components of the wavepacket, positioned at the energies of the eigenvalues. It can be easily demonstrated that it is possible to compute the spectrum also starting from the wavepacket autocorrelation function, (see Appendix 1.c for more details)

$$I(E) = \frac{1}{2\pi\hbar} \int_{-\infty}^{+\infty} \langle \Psi(0) | \Psi(t) \rangle \, e^{iEt} \mathrm{d}t. \tag{1.34}$$

Recalling the propagator definition (eq. 1.2), the equation 1.34 is re-written as

$$I(E) = \frac{1}{2\pi\hbar} \int_{-\infty}^{+\infty} \mathrm{d}t \, e^{\frac{i}{\hbar}Et} \langle \Psi | e^{-\frac{i}{\hbar}\hat{H}t} | \Psi \rangle \tag{1.35}$$

$$= \mathrm{Re} \left(\frac{1}{\pi\hbar} \int_{0}^{\infty} \mathrm{d}t \, e^{\frac{i}{\hbar}Et} \langle \Psi | e^{-\frac{i}{\hbar}\hat{H}t} | \Psi \rangle \right) \tag{1.36}$$

where Re indicates that, thanks to the unitarity of the quantum propagator, only the real part of the expression in brackets is taken into account. In eq. 1.36 it is possible to substitute the propagator in the integrand with one of the semiclassical approximations seen in the previous paragraph. For example, inserting the Herman-Kluk formulation of the propagator, the following expression for the spectrum is obtained:

ab

$$\begin{aligned} I(E) = & \frac{1}{(2\pi\hbar)^N} \frac{\mathrm{Re}}{\pi\hbar} \int_{0}^{\infty} \mathrm{d}t e^{\frac{i}{\hbar}Et} \\ & \times \left(\int \mathrm{d}\mathbf{p_0} \int \mathrm{d}\mathbf{q_0} C_t \left(\mathbf{p_0}, \mathbf{q_0} \right) e^{\frac{i}{\hbar}S_t(\mathbf{p_0},\mathbf{q_0})} \langle \Psi | \mathbf{p}_t, \mathbf{q}_t \rangle \langle \mathbf{p}_0, \mathbf{q}_0 | \Psi \rangle \right) \end{aligned} \tag{1.37}$$

where $(\mathbf{p}_t, \mathbf{q}_t)$ are the momenta and positions along the classical trajectory, $S(\mathbf{p}, \mathbf{q})$ is the classical action, $C_t(\mathbf{p}, \mathbf{q})$ represents the Herman-kluk prefactor described in equation 1.32 and $\langle x | \mathbf{p}_t, \mathbf{q}_t \rangle$ is the Gaussian wavepacket, centered at $(\mathbf{p}_t, \mathbf{q}_t)$. In this way the spectrum of a generic system, ranging from simple molecules to even complex supramolecular aggregates, can be recovered by using the information coming from molecular dynamics.

1.4. Overcoming the limits of the Herman-Kluk propagator: the time average filter

Despite the potentialities of the Herman-Kluk propagator, demonstrated by the noticeable number of works existing in the literature ,[14, 13] its application to molecular systems remains strongly limited. This is due essentially to the high number of trajectories requested for the convergence of the integral calculations. This number is proportional to the number of degrees of freedom (DOF) of the system to be analysed, in a way that application of the semiclassical method to molecules of biological interest with many DOF is completely forbidden at this level. In order to overcome this limitation, in 2003 Kaledin and Miller proposed a time-average filter approach that has represented a key step on the way to molecular vibrational investigation.[37, 38] The main ingredient of this filter consists in a time average version of the original phase-space integral, switching from

$$I_{PS} = \int \mathrm{d}\mathbf{p}_0 \int \mathrm{d}\mathbf{q}_0 A(\mathbf{p}_t, \mathbf{q}_t) \tag{1.38}$$

to

$$I_{PS,TA} = \int \mathrm{d}\mathbf{p}_0 \int \mathrm{d}\mathbf{q}_0 \frac{1}{T} \int \mathrm{d}t A(\mathbf{p}_t, \mathbf{q}_t) \tag{1.39}$$

that is, substantially, a time averaging of the integrand along the trajectory, for each initial condition. As illustrated below, by changing the order of integration and by invoking Liouville's theorem, for which the phase-space distribution function is constant along the trajectories and the change $\mathrm{d}\mathbf{p}_0, \mathrm{d}\mathbf{q}_0 \rightarrow \mathrm{d}\mathbf{p}_t, \mathrm{d}\mathbf{q}_t$ has a unitary Jacobian, the two formulations are fully equivalent.

$$I_{PS,TA} = \frac{1}{T} \int \mathrm{d}t \int \mathrm{d}\mathbf{p}_0 \int \mathrm{d}\mathbf{q}_0 A(\mathbf{p}_t, \mathbf{q}_t) \tag{1.40}$$

$$= \frac{1}{T} \int \mathrm{d}t \int \mathrm{d}\mathbf{p}_t \int \mathrm{d}\mathbf{q}_t A(\mathbf{p}_t, \mathbf{q}_t) \tag{1.41}$$

$$= \frac{1}{T} \int \mathrm{d}t I_{PS} = I_{PS}. \tag{1.42}$$

At first glance, equation 1.39 may seem more complicated than equation 1.38, since it presents an additional integral to solve, but actually the time average of A is now a smoother function of the phase space variables, and hence the statistical convergence of the calculations is improved. This means that a minor number of

initial conditions $\mathbf{p}_0, \mathbf{q}_0$ are needed to make the calculation converge, opening the way to the analysis of bigger molecules, otherwise unfeasible. Applying the filter strategy as it appears in Eq. 1.39 to the Herman Kluk formulation reported in 1.37 and switching the time integration with the phase space integration, it is possible to write

$$\begin{aligned} I(E) =& \frac{1}{(2\pi\hbar)^N} \int \mathrm{d}\mathbf{p} \int \mathrm{d}\mathbf{q} \frac{\mathrm{Re}}{\pi\hbar T} \int_0^T \mathrm{d}t_1 \int_{t_1}^{\infty} \mathrm{d}t_2 C_{t_2}\left(\mathbf{p}_{t_1}, \mathbf{q}_{t_1}\right) \\ & \times \left\langle \Psi | \mathbf{p}_{t_2}, \mathbf{q}_{t_2} \right\rangle e^{\frac{i}{\hbar}\left(S_{t_2}(\mathbf{p},\mathbf{q}) + Et_2\right)} \left[\left\langle \Psi | \mathbf{p}_{t_1}, \mathbf{q}_{t_1} \right\rangle e^{\frac{i}{\hbar}\left(S_{t_1}(\mathbf{p},\mathbf{q}) + Et_1\right)} \right]^*. \end{aligned} \tag{1.43}$$

This expression presents two different time integrals in two different time variables $t_1$and t_2 that make the calculation difficult to manage. A desirable improvement will consist in the collapse of the original integrand into an absolute squared integrand in a single time variable T. Unfortunately, the presence of the prefactor $C_{t_2}\left(\mathbf{p}_{t_1}, \mathbf{q}_{t_1}\right)$, which has a dependency on both the time variables, hampers a direct simplification of the method. For this reason, Kaledin and Miller proposed to approximate the complete HK prefactor to its phase only, according to the following equation

$$C_{t_2}\left(\mathbf{p}_{t_1}, \mathbf{q}_{t_1}\right) \approx e^{\frac{i}{\hbar}\phi_{t_2}} e^{-\frac{i}{\hbar}\phi_{t_1}}, \tag{1.44}$$

where ϕ indicates the phase of the HK prefactor C_t. Eq. 1.44 is called "separable approximation", because it assumes that the two time scales are exactly separable. This is definitely true only for the harmonic oscillator, but it has been demonstrated to be accurate also for a large number of more complicated systems.[39, 40, 41, 42, 25, 43] Substituting the separable approximation (1.44) into equation 1.43 the Time Average filtered version of the Herman Kluk propagator is obtained in its Initial Value Representation, (TA-SCIVR)

$$\begin{aligned} I(E) =& \frac{1}{(2\pi\hbar)^N} \frac{1}{2\pi\hbar T} \int \mathrm{d}\mathbf{p} \int \mathrm{d}\mathbf{q} \\ & \times \left| \int_0^T \mathrm{d}t \left\langle \Psi | \mathbf{p}_t, \mathbf{q}_t \right\rangle e^{\left\{\frac{i}{\hbar}\left[S_t(\mathbf{p},\mathbf{q}) + Et + \phi_t(\mathbf{p},\mathbf{q})\right]\right\}} \right|^2 . \end{aligned} \tag{1.45}$$

Furthermore the collapse of the two time scales $t_1$and t_2 into a single time t implies the evaluation of the integral in t_2 from 0 to T, leading to the dividing factor 2 that appears in equation 1.45.

Using such formulation of the semiclassical approach, Kaledin and Miller were able to recover the vibrational spectrum of H_2, H_2O and CH_2O with very high accuracy, comparable to exact quantum results. Most importantly they demonstrated with their work that the time average filter was able to drastically reduce the number of trajectories required for the convergence of the integral calculation, at the point that meaningful information could be recovered even from a single long trajectory.

1.5. Fighting the curse of dimensionality, the Multiple-Coherent approach

The introduction of TA-SCIVR allowed application of the semiclassical method to small molecular systems. For all these systems, the potential energy surfaces (PES) used for the calculation of the potential energy, were pre-computed or pre-constructed by means of fitting techniques. Such a strategy permits to exploit an analytical formula in order to obtain energy values in very short computational times. Unfortunately, construction of a full dimensional PES is generally not affordable for systems with many degrees of freedom, because of the high computational effort that the ab-initio calculations on which the fitting procedure is based may require, expecially if a high level of electronic theory is employed. Moreover, a hypothetical accurate PES for big size systems requires complex analytical expressions that will yield significant time consuming computations.

For this reason the use of a First Principles Molecular Dynamics (FPMD) approach, where the potential and its derivatives are calculated *on-the-fly* along the dynamics, represents a valid alternative to the use of a pre-existing PES, as highlighted by Ceotto et al. in 2009.[25, 26] In the same publication the authors brought to light another important issue. In fact, investigating the vibrational spectrum of CO_2 molecules, they observed that, when a single trajectory is employed for the semiclassical calculation, poor results for the higher vibrational levels are obtained. This event can be understood by looking at the correlation function itself (Eq. 1.34). In fact, using the semiclassical formulation, a well-defined peak arises in the spectrum if the reference state $|\Psi\rangle$ significally overlaps with each eigenstate of the system. It is possible to better understand this statement by looking at the numerical form of the correlation function,

$$\begin{aligned}
&\langle\Psi|\mathbf{p}_t,\mathbf{q}_t\rangle\,\langle\mathbf{p}_0,\mathbf{q}_0|\Psi\rangle = \\
&\exp[-\gamma(\mathbf{q}_t-\mathbf{q}^{\mathrm{i}}{}_{eq})^2/4-\gamma(\mathbf{q}_0-\mathbf{q}^{\mathrm{i}}{}_{eq})^2/4-\gamma(\mathbf{p}_t-\mathbf{p}^{\mathrm{i}}{}_{eq})^2/4-\gamma(\mathbf{p}_0-\mathbf{p}^{\mathrm{i}}{}_{eq})^2/4] \\
&\times\exp[-i(\mathbf{p}_t\mathbf{q}_t-\mathbf{p}_0\mathbf{q}_0)/2-i\mathbf{p}^{\mathrm{i}}{}_{eq}(\mathbf{q}_t-\mathbf{q}_0)/2+i\mathbf{q}^{\mathrm{i}}{}_{eq}(\mathbf{p}_t-\mathbf{p}_0)/2]
\end{aligned} \tag{1.46}$$

from which emerges that the signal in the Fourier transform is significally different from 0 when the momenta and positions explored from the trajectory, $(\mathbf{p}_t,\mathbf{q}_t)$, are close to $(\mathbf{p}^{\mathrm{i}}{}_{eq},\mathbf{q}^{\mathrm{i}}{}_{eq})$. Related to this idea, in a pioneering work published by De Leon and Heller,[44] it has been demonstrated that accurate semiclassical results can be obtained even by means of a single trajectory if it is run at the correct (quantum) energy. Following this insight, Ceotto and coworkers suggested the possibility to use a reference state specifically built in order to mimic all the spectral features with the proper eigenvalue spacings. One way to fulfill this request is to chose for the trajectories the equilibrium geometry $\mathbf{q}^{\mathrm{i}}{}_{eq}$ as initial positions and $\mathbf{p}^{\mathrm{i}}{}_{eq}=\sqrt{(n+1/2)\hbar\omega}$ as initial momenta, n being the harmonic oscillator quantum number and ω the harmonic frequency. Usually a single trajectory is run for each of the N_{st} states with the reference state appropriately set. Such strategy is pictorially represented in Figure 1.2

Inserting this idea in the previous formula is possible to re-write the semiclassical formulation of the vibrational spectrum, reaching the so-called Multiple Coherent (MC) TA-SCIVR[42, 40] :

$$\begin{aligned}
I(E)=&\frac{1}{(2\pi\hbar)^N}\frac{\mathrm{Re}}{\pi\hbar T}\sum_{j=1}^{N_{trajs}=N_{states}}\int_0^T\mathrm{d}t_1\int_0^T\mathrm{d}t_2 C_{t_2}(\mathbf{p}^j(t_1),\mathbf{q}^j(t_1)) \\
&\times\sum_{k=1}^{N_{states}}\langle\mathbf{p}^k_{eq},\mathbf{q}^k_{eq}|\mathbf{p}^j(t_2),\mathbf{q}^j(t_2)\rangle\, e^{\frac{i}{\hbar}(S_{t_2}(\mathbf{p}^j(0),\mathbf{q}^j(0))+Et_2)} \\
&\times\left[\sum_{k=1}^{N_{states}}\langle\mathbf{p}^k_{eq},\mathbf{q}^k_{eq}|\mathbf{p}^j(t_1),\mathbf{q}^j(t_1)\rangle\, e^{\frac{i}{\hbar}(S_{t_1}(\mathbf{p}^j(0),\mathbf{q}^j(0))+Et_1)}\right]^*
\end{aligned} \tag{1.47}$$

where the phase space integral has been replaced by a sum over the coherent states

$$|\Psi^{(K)}\rangle=\prod_{i=1}^{N_{vib}}|\mathbf{p}^{(K)}_{eq,i},\mathbf{q}^{(K)}_{eq,i}\rangle \qquad K=1,...,N_{st}. \tag{1.48}$$

In practice, instead of choosing the initial points for the trajectories by a random sampling of the phase space around the minimum, now it is possible to select

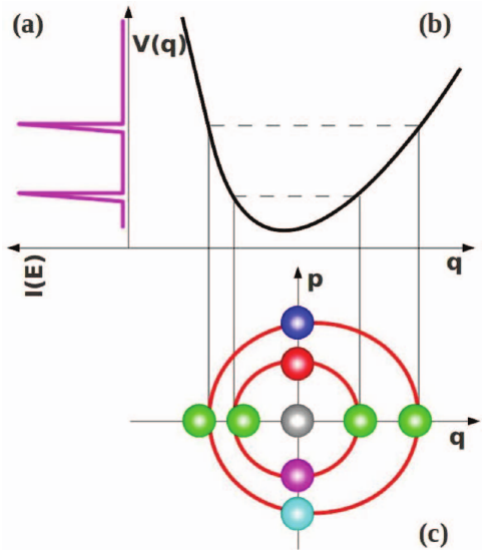

Figure 1.2.: Several possible reference state for the MC-SCVIR approach. In panel (a) is represented the power spectrum; in panel (b) the potential energy and in panel (c) the phase space in which classical trajectories (red closed lines) and coherent states (filled colored circles) are sketched. Reproduced from ref [40]

a very small set of trajectories (one for each of the N_{st} states) with the initial conditions selected to resemble the real eigenfunctions of the system. Figure 1.2 pictorially represents the Multiple Coherent strategy. In panel (c) is represented the phase space where several initial conditions are visualized by means of coloured circles. The corresponding trajectories explore the phase space associated to a region of the potential energy surface, depicted in panel (b), that permits to recover the exact quantum level in the vibrational spectrum, reported in panel (a).

In this thesis work all the MC-SCIVR analysis have been performed by choosing equilibrium positions and harmonic Zero Point Energy (ZPE) momenta as initial trajectory conditions for each normal mode of the system.

Using the MC-SCIVR it has been possible to employ a small set of 8 trajectories to recover with very similar accuracy the water vibrational eigenvalues obtained from 32 000 trajectories used for the Monte Carlo integration of the TA-SCIVR approach.[45] Most interestingly, the possibility to use a very limited number of trajectories allows to use FPMD, giving the opportunity to study complex or large systems, for which a pre-computed PES is not available.

Another important property of the reference state choice for the multiple coherent states formulation is the possibility to determine a priori which peaks will appear in the spectrum. A couple of years upon publication of the MC-SCIVR strategy, in fact, Ceotto and coworkers proposed to define the reference state through a combination of coherent states , written in the general form

$$|\Psi^{(K)}\rangle = \prod_{i=1}^{N_{vib}} \sum_{\alpha=1}^{N_\alpha} \varepsilon_{\alpha,i}^{(K)} \, |\mathbf{p}_{eq,\alpha,i}^{(K)}, \mathbf{q}_{eq,\alpha,i}^{(K)}\rangle \qquad K = 1, ..., N_{st} \tag{1.49}$$

where N_α is the number of coherent states employed in this "augmented" reference state and the value of the coefficient $\varepsilon_{\alpha,i}^{(K)}$ allows to enforce parity or molecular symmetry to favor detection of signals corresponding to specific mode excitations or symmetry species. This feature is particularly important in view of multidimensional systems analysis, because it is possible to use the antisymmetric choice ($\varepsilon = -1$) for each single mode in order to isolate its fundamental transition contribution from others. If the symmetric combination ($\varepsilon = 1$) is chosen instead for all the states, then it is possible to visualize the ZPE. Indeed in this thesis all the spectra have been produced using the Multiple-Coherent approach, together with the time-average filter, exploiting this type of combination of coherent states.

1.6. The "Divide and Conquer", a new formulation toward systems of high-dimensionality

The introduction of the multiple-coherent approach permitted the study of higher dimensional molecular systems, starting from important staple molecules, as benzene, up to biologically relevant systems, like the glycine molecule. The vibrational spectrum of glycine, the smallest between all the aminoacids, has been investigated within this thesis work, and it is illustrated in details in the first chapter of the Part II.

Despite this important advance, this system represents the size limit for the semiclassical machinery, as presented so far. The reason why this happens has already been understood and explained in previous sections. The semiclassical wavepacket, built as the direct product of coherent states $|\Psi(t)\rangle = |\Psi_1(t)\rangle ... |\Psi_N(t)\rangle$, is employed for the calculation of the Fourier Transform of the time-dependent overlap $\langle\Psi(0)|\Psi(t)\rangle$. In order to have a clear signal in the spectrum, this time-evolved wavepacket has to significally overlap with its initial state. For a multidimensional system this overlap has to happen for all the dimensions at the same time. In this particular aspect, a semiclassical spectrum is deeply different from the "classical" one, where the vibrational density is investigated through a three-dimensional dipole autocorrelation function. Within the literature, this multidimensionality issue is known as "the curse of dimensionality". A successful way to overcome this limitation has been presented by Ceotto and coworkers for the first time in 2017.[46] The key ingredient is the observation that being able to reduce the dimensionality of the system, the overlap of the time evolved wavepacket with its initial guess is more likely to occur. Therefore the idea is to partition all the degrees of freedom of the system in subdimensional groups, to apply the semiclassical method on these dimensions in order to obtain a partial spectrum, and finally to sum up over all the partial spectra to get the spectral density of the whole multidimensional system. In Figure 1.3 this strategy is pictorially sketched.

From a mathematical point of view, the "reduction" of the dimension is accomplished by means of a projection of the full dimensional system onto subdimensional subspaces. Starting from the expression of the TA-SCIVR (eq. 1.45), it is possible to write the spectrum of the single subspace, with all the quantities projected

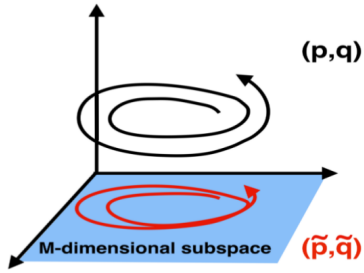

Figure 1.3.: Pictorial representation of the projection procedure. Reproduced from ref [46]

$$\tilde{I}(E) = \frac{1}{(2\pi\hbar)^M}\frac{1}{2\pi\hbar T}\int \mathrm{d}\tilde{\mathbf{p}}(0)\int \mathrm{d}\tilde{\mathbf{q}}(0) \times \left|\int_0^T \mathrm{d}t\, \langle\tilde{\Psi}|\tilde{\mathbf{p}}_t,\tilde{\mathbf{q}}_t\rangle\, e^{\{\frac{i}{\hbar}[\tilde{S}_t(\tilde{\mathbf{p}},\tilde{\mathbf{q}})+Et+\phi_t(\tilde{\mathbf{p}},\tilde{\mathbf{q}})]\}}\right|^2 \tag{1.50}$$

where M is the dimension of the subspace (with $M < N$, where N are the dimensions of the system), $(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ are the momenta and positions of the degrees of freedom of the subspace and $\langle\tilde{x}|\tilde{\mathbf{p}}_t, \tilde{\mathbf{q}}_t\rangle$ is the M-dimensional coherent state defined as

$$\langle\tilde{x}|\tilde{\mathbf{p}}_t,\tilde{\mathbf{q}}_t\rangle = \left(\frac{\det(\tilde{\mathbf{\Gamma}})}{\pi^M}\right)^{\frac{1}{4}} e^{-(\tilde{\mathbf{x}}-\tilde{\mathbf{q}}(t))^T\tilde{\mathbf{\Gamma}}(\tilde{\mathbf{x}}-\tilde{\mathbf{q}}(t))/2+i\tilde{\mathbf{p}}^T(t)(\tilde{\mathbf{x}}-\tilde{\mathbf{q}}(t))/\hbar} \tag{1.51}$$

where $\tilde{\mathbf{\Gamma}}$ is also the projection of the Gaussian width matrix $\mathbf{\Gamma}$. The expression reported in eq. 1.50 describes the Divide-and-Conquer semiclassical approach, also known with the abbreviation DC SCIVR. Among all the quantities involved in eq. 1.50 only the potential energy, which is one of the components of the action S, is not trivially separable, since it depends on all the degrees of freedom of the system simultaneously. To overcome this issue, one of the proposed potential energy expressions is

$$V_S(\tilde{\mathbf{q}}(t)) = V(\tilde{\mathbf{q}}(t); \mathbf{q}^{\mathrm{eq}}_{\mathrm{N_{vib}-M}}) + \lambda(t) \tag{1.52}$$

where

$$\lambda(t) = V(\tilde{\mathbf{q}}(t); \mathbf{q}_{\mathrm{N_{vib}-M}}(t)) - [V(\tilde{\mathbf{q}}(t); \mathbf{q}^{\mathrm{eq}}_{\mathrm{N_{vib}-M}}) + V(\mathbf{q}^{\mathrm{eq}}_{\mathrm{M}}; \mathbf{q}_{\mathrm{N_{vib}-M}}(t))]. \tag{1.53}$$

In other words, an external time-dependent field has been defined to obtain an expression that is in principle exact for a separable system. What remains to define is the criterion for mode partitioning. So far several methodologies have been proposed, each with different levels of accuracy and amounts of computational effort required. The most inexpensive one is based on the analysis of the Hessian matrix, averaged along the trajectory. After choosing a threshold ε value, comparing the off diagonal terms with ε, it is possible to determine if two modes are coupled or not. Applying this strategy recursively it is possible to subdivide the full dimensional space in smaller sets. This approach is almost instantaneous if you have already calculated the Hessians along the trajectory, but presents a strong arbitrary factor consisting in the choice of ε.

A similar strategy is based on the investigation of the stability matrix, see eq. 1.29, for the definition of the amount of coupling between modes.[47] It exploits dynamical information but relies on an arbitrary ε value choice.

A more rigorous criterion is the one called "Jacobi space decomposition method".[48] It takes advantage of Liouville's theorem, for which the determinant of the Jacobi Matrix, obtainable starting from the monodromy matrix (eq. 1.29), has to be equal to 1 for an isolated system. Employing this observation, one can calculate this determinant for each possible space subdivision along the dynamics and estabilish which one provides the subspaces with the determinant closer to 1. Despite the clear advantage of this approach in avoiding arbitrariness, this method requires a lot of calculation and the computational effort dramatically increases with the dimensionality of the system since the number of possible combinations blows up fast. Since all the systems studied in this thesis presents a high number of degrees of freedom, the criterion that has been used in this thesis work for the subspace partition is the one based on the Hessian matrix.

Together with advances on the theoretical side, other minor approximations have been applied in order to tackle molecular systems with reasonable computational effort. Such measures concern how to handle the prefactor and the reduction of the number of Hessians to calculate. All the details about these features are discussed in the Appendix 1.d.

The capabilities of DC-SCIVR has been demonstrated evaluating spectra of systems with a high number of degrees of freedom, like the fullerene molecule, and even supramolecular systems characterized by high complexity, like, for example, water clusters.[48, 46] In this thesis work the DC-SCIVR method has been applied together with FPMD, interfacing semiclassical software with NWChem, a free downloadable software used for quantum mechanical calculations and also with Gromacs, a suite of codes used for classical dynamics.

1.7. Appendix

Appendix 1.a: Equivalence of Feynman's formalism to Schrödinger's equation

The Feynman path integral formalism is equivalent to the Shrödinger one, the difference is that the first approach is global and deals with propagation over finite times, while the latter is local in time and deals with time evolution over infinitesimal time scales. Following the Schrödinger equation the change of the state vector $|\psi\rangle$ in infinitesimal time ε is given by

$$|\psi(\varepsilon)\rangle - |\psi(0)\rangle = -\frac{i\varepsilon}{\hbar}\hat{H}\,|\psi(0)\rangle \tag{1.54}$$

that in the coordinate representation becomes

$$|\psi(x,\varepsilon)\rangle - |\psi(x,0)\rangle = -\frac{i\varepsilon}{\hbar}\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x,0)\right]|\psi(x,0)\rangle\,. \tag{1.55}$$

The same change from the point of view of the path integral formalism, to first order in ε (meaning that there is only a slice between the start and end points of the path), is

$$\psi(x,\varepsilon) = \int_{-\infty}^{+\infty} U(x,\varepsilon;x')\psi(x',0)\mathrm{d}x' \tag{1.56}$$

being

$$U(x,\varepsilon;x') = \sqrt{\frac{m}{2\pi\hbar i\varepsilon}}e^{\frac{i}{\hbar}\frac{(x-x')^2 m}{2\varepsilon} - \frac{i}{\hbar}V(\frac{x+x'}{2},0)\varepsilon}. \tag{1.57}$$

(See Appendix 1.b). Here the argument of V has been set to 0 since there is already a factor ε in front of it, and other variations in time of V between 0 and ε will produce a second order term in ε. By substituting eq. 1.57 into eq. 1.56 the result is

$$\psi(x,\varepsilon) = \sqrt{\frac{m}{2\pi\hbar i\varepsilon}}\int_{-\infty}^{+\infty} e^{\frac{i}{\hbar}\frac{(x-x')^2 m}{2\varepsilon}} e^{-\frac{i}{\hbar}V(\frac{x+x'}{2},0)\varepsilon}\psi(x',0)\mathrm{d}x'. \tag{1.58}$$

Looking now at the exponential term $e^{\frac{i}{\hbar}\frac{(x-x')^2 m}{2\varepsilon}}$ it is possible to observe that, being ε infinitesimal and $\hbar$ small, it oscillates very rapidly as $(x-x')$ varies, making

the integral vanish except in the region where it is stationary. The only stationary point in this case is $x = x'$, where the phase is approximately zero. By defining η as

$$\eta = x - x' \tag{1.59}$$

the 1.58 becomes

$$\psi(x,\varepsilon) = \sqrt{\frac{m}{2\pi\hbar i\varepsilon}} \int_{-\infty}^{+\infty} e^{\frac{i}{\hbar}\frac{\eta^2 m}{2\varepsilon}} e^{-\frac{i}{\hbar}V(x+\frac{\eta}{2},0)\varepsilon} \psi(x+\eta,0)\mathrm{d}x'. \tag{1.60}$$

Working to first order in ε and therefore to second order in η it is possible to expand $\psi(x+\eta,0)$ and $e^{-\frac{i}{\hbar}V(x+\frac{\eta}{2},0)\varepsilon}$ to be

$$\psi(x+\eta,0) = \psi(x,0) + \eta\frac{\partial\psi}{\partial x} + \frac{\eta^2}{2}\frac{\partial^2\psi}{\partial x^2} + ... \tag{1.61}$$

$$e^{-\frac{i}{\hbar}V(x+\frac{\eta}{2},0)\varepsilon} = 1 - \frac{i\varepsilon}{\hbar}V(x+\frac{\eta}{2},0) + ... \tag{1.62}$$

$$= 1 - \frac{i\varepsilon}{\hbar}V(x,0) \tag{1.63}$$

since terms of order $\eta\varepsilon$ are to be neglected. In this way equation 1.60 becomes

$$\psi(x,\varepsilon) = \sqrt{\frac{m}{2\pi\hbar i\varepsilon}} \int_{-\infty}^{+\infty} e^{-\frac{\eta^2 m}{2i\hbar\varepsilon}} \left[\psi(x,0) - \frac{i\varepsilon}{\hbar}V(x,0)\psi(x,0) + \eta\frac{\partial\psi}{\partial x} + \frac{\eta^2}{2}\frac{\partial^2\psi}{\partial x^2}\right] d\eta \tag{1.64}$$

that can be solved as a Gaussian integral obtaining

$$\psi(x,\varepsilon) = \sqrt{\frac{m}{2\pi\hbar i\varepsilon}} \left[\psi(x,0)\sqrt{\frac{2\pi\hbar i\varepsilon}{m}} - \frac{\hbar\varepsilon}{2im}\sqrt{\frac{2\pi\hbar i\varepsilon}{m}}\frac{\partial^2\psi}{\partial x^2} - \frac{i\varepsilon}{\hbar}\sqrt{\frac{2\pi\hbar i\varepsilon}{m}}V(x,0)\psi(x,0)\right] \tag{1.65}$$

which can be rearranged as follows

$$\psi(x,\varepsilon) - \psi(x,0) = -\frac{i\varepsilon}{\hbar}\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x,0)\right]\psi(x,0) \tag{1.66}$$

to agree with the Schrödinger prediction in eq. 1.55.

Appendix 1.b: Free particle propagator

In this short section the propagator of a free particle will be evaluated. The system is characterized by the following Hamiltonian

$$\widehat{\mathrm{H}} = \frac{\hat{p^2}}{2m}. \tag{1.67}$$

By definition of quantum propagator is true that,

$$\langle q|e^{-\frac{i}{\hbar}\widehat{H}t}|q'\rangle = \langle q|e^{-\frac{i}{\hbar}\frac{\hat{p^2}}{2m}t}|q'\rangle\,. \tag{1.68}$$

Now, inserting in the 1.68 two identity relations

$$\int_{-\infty}^{+\infty} \mathrm{d}p\,|p\rangle\,\langle p| = 1, \qquad \int_{-\infty}^{+\infty} \mathrm{d}p'\,|p'\rangle\,\langle p'| = 1 \tag{1.69}$$

it is possible to obtain

$$\int_{-\infty}^{+\infty} \mathrm{d}p \int_{-\infty}^{+\infty} \mathrm{d}p'\,\langle q|p\rangle\,\langle p|e^{-\frac{i}{\hbar}\frac{\hat{p^2}}{2m}t}|p'\rangle\,\langle q'|p'\rangle \tag{1.70}$$

$$= \int_{-\infty}^{+\infty} \mathrm{d}p \int_{-\infty}^{+\infty} \mathrm{d}p'\,\langle q|p\rangle\,\langle p|p'\rangle\,e^{-\frac{i}{\hbar}\frac{p^2}{2m}t}\,\langle q'|p'\rangle \tag{1.71}$$

$$= \int_{-\infty}^{+\infty} \mathrm{d}p'\,\langle q|p'\rangle\,\langle p'|q'\rangle\,e^{-\frac{i}{\hbar}\frac{p^2}{2m}t} \tag{1.72}$$

$$= \int_{-\infty}^{+\infty} \mathrm{d}p'\,\frac{1}{\sqrt{2\pi\hbar}}e^{\frac{i}{\hbar}p'q}\,\frac{1}{\sqrt{2\pi\hbar}}e^{-\frac{i}{\hbar}p'q'}e^{-\frac{i}{\hbar}\frac{p^2}{2m}t} \tag{1.73}$$

$$= \frac{1}{2\pi\hbar}\int_{-\infty}^{+\infty} \mathrm{d}p' e^{-\frac{i}{\hbar}\frac{p^2}{2m}t}e^{\frac{i}{\hbar}p'(q-q')} \tag{1.74}$$

that is a Gaussian integral that, once solved, leads to

$$= \frac{1}{2\pi\hbar}\sqrt{\frac{2\pi\hbar m}{it}}e^{-\frac{m(q-q')^2}{2\hbar it}} \tag{1.75}$$

$$= \sqrt{\frac{m}{2\pi\hbar i t}} e^{\frac{i}{\hbar}\frac{(q-q')^2 m}{2t}} \tag{1.76}$$

that is the formulation of the free particle propagator.

Appendix 1.c: Spectrum definitions

As reported in this chapter, the definition of "spectrum" is given by

$$\sigma(\omega) = \sum_n |c_n|^2 \delta(\omega - \omega_n). \tag{1.77}$$

Expansion of the wavepacket of a bound system in the Hamiltonian's eigenfunctions leads to

$$\Psi(x,t) = \sum_n c_n \psi_n(x) e^{-\frac{i}{\hbar}E_n t} \tag{1.78}$$

where $\omega_n = \frac{E_n}{\hbar}$, E_n being the eigenvalue energies. Another definition of the spectrum is based on the Fourier Transform of the wavepacket autocorrelation function

$$\sigma(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \langle \Psi(0) | \Psi(t) \rangle \, e^{i\omega t} \mathrm{d}t. \tag{1.79}$$

Substituting eq. 1.78 into this last formulation it is possible to demonstrate, with a small number of steps and exploiting the ortonormality of the eigenfunctions of H, that it is fully equivalent to eq. 1.77

$$\sigma(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\sum_m c_m^* \psi_m^*(x))(\sum_n c_n \psi_n(x) e^{-\frac{i}{\hbar}E_n t}) \mathrm{d}x e^{i\omega t} \mathrm{d}t \tag{1.80}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \sum_{m,n} c_m^* c_n \delta_{mn} e^{-\frac{i}{\hbar}E_n t} e^{i\omega t} \mathrm{d}t \tag{1.81}$$

$$= \frac{1}{2\pi} \sum_n |c_n|^2 \int_{-\infty}^{+\infty} e^{-\frac{i}{\hbar}E_n t} e^{i\omega t} \mathrm{d}t \tag{1.82}$$

$$= \sum_n |c_n|^2 \delta(\omega - \frac{E_n}{\hbar}) \tag{1.83}$$

$$= \sum_n |c_n|^2 \delta(\omega - \omega_n) \tag{1.84}$$

that is exactly the equation reported in 1.77.

Appendix 1.d: Additional tools for semiclassical vibrational studies

Within this chapter I already presented the Monodromy Matrix **M**,

$$\mathbf{M} = \begin{pmatrix} \partial p_t/\partial p_0 & \partial p_t/\partial x_0 \\ \partial x_t/\partial p_0 & \partial x_t/\partial x_0 \end{pmatrix} \tag{1.85}$$

its role in the pre-exponential factor in semiclassical formulations

$$C_t(\mathbf{p},\mathbf{q}) = \sqrt{\det\left[\frac{1}{2}\left(\mathbf{M_{qq}} + \gamma^{-1}\mathbf{M_{pp}}\gamma + \frac{i\gamma^{-1}}{\hbar}\mathbf{M_{pq}} + \frac{\hbar\gamma}{i}\mathbf{M_{qp}}\right)\right]} \tag{1.86}$$

and its importance being involved in the account of quantum contributions to the propagator operator.

What it has been missing so far is a description of the numerical problems connected to such a prefactor. In fact the numerically-integrated monodromy matrix elements become exponentially large as time evolves when the dynamics is chaotic and the dimensionality of the system increases. This generates issues about both integral convergence and numerical stability of the codes. For these reasons several approximations to the prefactor have been proposed in the literature.[39, 49, 13, 50, 51] Some strategies deal with the removal of chaotic trajectories from the double phase-space integral evaluation, but such an approach is not of course available for formulations like MC-SCIVR or DC-SCIVR where a small number of trajectories, and often even a single one, are employed. Other approaches are based on "probable" estimations of the prefactor to keep fixed during all the dynamics, while in other strategies the prefactor is reformulated by using log-derivative quantities.

The most efficient route, at least for the studies here presented, turned to be the strategy called "Matrix Regularization". The idea is to monitor the monodromy matrix elements during the dynamics evolution. Thanks to the diagonalization of the matrix, the most "chaotic" degrees of freedom can be found by looking at their eigenvalues. Then a sudden zeroing of such eigenvector and eigenvalue followed by recalculation of the monodromy matrix is employed to make the trajectory gain numerical stability. The monodromy matrix regularization has been extensively used along this work, since complex, supramolecular or quasi-solvated systems are found to be characterized by a chaotic dynamics.

Another important ease to the overall computational effort required by semiclassical methods is represented by the reduction of the number of Hessian matrices to calculate. In agreement with the semiclassical formulation, the Hes-

sian matrix is requested step-by-step along the dynamics. The ab-initio Hessian calculation is however computationally very expensive and some methods to approximate such matrix are highly desirable. The principal tool that I employed for the production of the results of this thesis is the Compact Finite Difference scheme (CFD), developed by Ceotto and Hase in 2013.[39] It works by using a reduced number of ab-initio Hessian matrices and by estimating the others starting from the latest available calculated Hessian and the gradient of the potential energy. The gradient is infact always available because already calculated during the ab-initio molecular dynamics. As a rule of thumb I calculated one Hessian matrix in two steps, or three steps for very large systems.

An alternative approach is represented by the construction of a database of Hessian matrices. After choosing a threshold on the position coordinates, a limited number of Hessians is calculated because for "similar" sets of positions the same matrix is employed. This approach, in this basic version, may require a very small number of Hessian calculations, depending on the threshold value. Within this thesis work such an approach has been tested and employed for the calculation of the vibrational spectrum of the glycine molecule interacting with nine water molecules, a system composed by 105 degrees of freedom.

2. Treatment of Classical Molecular Dynamics

The aim of this chapter is to give a brief review of two different approaches for classical dynamics: Molecular Mechanics (MM) and Ab-Initio Molecular Dynamics (AIMD). The validity of these methodologies will be here discussed, with particular emphasis on theoretical vibrational spectroscopy. Within this thesis I largely used AIMD to reproduce vibrational spectra in a classical framework and also to calculate the trajectories required by the semiclassical method. I also tried to use classical molecular mechanics to recover information exploitable for spectroscopic purposes.

2.1. Molecular Modelling and Molecular Mechanics

In order to understand what Molecular Modelling is, one can simply refer to the definition of "molecular" and "modelling" words respectively. A model is an idealised description of a system that helps in calculations or predictions. The word "molecular" clearly identifies the object of a "model" in some way related to molecules. Even if some basic models can be computed using paper and pencil, nowadays molecular modelling is always associated to computer simulations. Evidently the computer plays the role of a tool that expresses distinct theories in the form of algorithms. Quantum mechanics, for example, from which the semiclassical methodologies have been derived in the previous Section, was born many years before the construction of the first computers and today it is exploited by a large number of available softwares. Unfortunately application of quantum mechanics is limited to small-size systems due to the computational efforts required. For this reason a large part of Molecular Modelling problems are tackled by means of more simplified approaches, like for instance Molecular Mechanics.

Molecular Mechanics methods, also known as Force Field methods, are based on two main assumptions. The first one is that the electronic degrees of freedom can be neglected in the description of the system dynamics. This assertion has its foundation on the Born-Oppenheimer approximation, in which the electrons

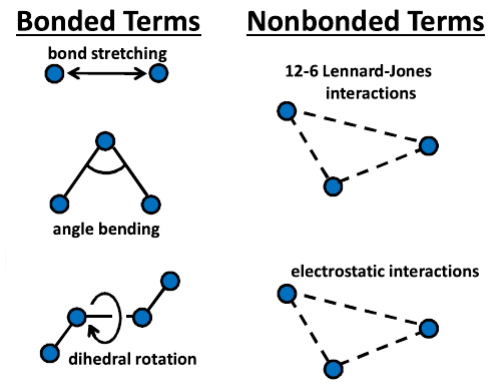

Figure 2.1.: Reproduced from ref [55]

are described by means of adiabatic states whose energy is just a function of the nuclear coordinates. The second assumption is that the energy function can be approximated using an analytical expression. An important step in the development of this class of methods is represented by the construction of such an energy function. It is generically built as a sum of simple contributions such as bond stretchings, opening and closing of atom angles, torsions and so on, as pictorially illustrated in Figure 2.1. All terms will be described in details in the following Section. Another characteristic aspect of Force Fields is represented by their transferibility. Force Fields work with a relatively small set of parameters that relate on specific atoms or groups of atoms.They can be applied to a wide range of systems and problems. [52, 53, 54, 55]

Thanks to this handful of ingredients, Force Fields methods are able to tackle very high dimensional systems of biological importance, like proteins or nucleobases, even including the solvent in its explicit form.

2.1.1. Building a simulation

Working with a Force Field is surprisingly similar to building a real experiment. In this latter case, the sample has to be prepared, than the instrument has to be connected to the sample and finally comes the measure of the property, taken for a period of time. In the same way the molecular mechanics analysis works. First of all the model has to be set up. The system is prepared as a collection of N atoms in a numerical coordinate representation. Then it is interfaced with the software that performs the dynamics and with the set of parameters that represents the Force Field. At this stage the dynamics starts. The system

is evolved for a certain period of time, and finally the property of interest is measured. As in a real experiment, so in the Force Field simulation, the longer is the evolution the more accurate is the measure of the property.

Also the code structure can be easily outlined, highlighting the most important steps that a molecular dynamics program performs. After the initialization stage, where the initial positions and velocities are set, the dynamical loop begins. For each step of the trajectory the calculation of electronic energies and forces is needed, so that the integration of the equations of the motion can be performed numerically. Each of this section is subdivided in more complex operations, depending also on the kind of simulation, the Force Field employed and the quantities that have to be measured, but in all cases the most time-consuming part of the program is certainly represented by the electronic energy and force calculations.

2.1.2. The energy function

A great number of molecular mechanics force fields can be seen as a collection of terms divided in intra-molecular and inter-molecular components. The simplest force fields have only four components: one for the description of the bond stretching, another one for the angle bending and the remaining two for the torsional terms and for non-bonded interactions. These are the four fundamental terms that are ubiquitous in force fields, while other more complex versions present a greater number of additional components. The idea is often related to the fact that the system is penalised as far as every term is away from the equilibrium configuration. For instance, the bond term can be obtained by the elementary Hooke's law formula

$$E_{bond} = \frac{k_b}{2}(r - r_e)^2, \tag{2.1}$$

where k_b is the bond stretching constant, r is the distance between the atoms pair in the considered frame and r_e represents the equilibrium distance. In the same fashion it is possible to recover the angle term

$$E_{angle} = \frac{k_\theta}{2}(\theta - \theta_e)^2, \tag{2.2}$$

in which the only differences are the replacement of the position r with the angle θ between a triplet of atoms , and k_θ that stands for the bending constant. These first two components are considered the "hard" degrees of freedom of a system because a relatively large amount of energy is required to cause significant

deformations. In general, most part of the variation in the energy comes from the contribution of the torsional and non-bonded terms and from their complex interplay. The torsional part is commonly expressed as a cosine series

$$E_{torsion} = \sum_{n=0}^{N} \frac{V_n}{2}[1 + cos(n\omega - \gamma)], \tag{2.3}$$

where V represents the barrier height for the rotation, ω is the angle between the two planes, n stands for the multiplicity that is the minimum number of points for a complete rotation of 360° and finally γ is the phase factor. The last staple contribution comes from the non-interaction term, that arises from the interactions through space. Usually it is composed of two main groups, the electrostatic interactions and the van der Waals family of interactions. For the first contribution the base formulation is given from the well-known Coulomb's law

$$E_{el} = \sum_i \sum_j \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}, \tag{2.4}$$

where it is clear the direct dependency on the atomic charges q and the inverse correlation between energy and the atomic distance r. Of course this simple formula can be refined using multipolar expansions, introducing polarization contributes or detailed terms for treating specific groups, as the aromatic-aromatic interactions. Finally there is the van der Waals component, that is generally depicted with the Lennard-Jones 12-6 function

$$E_{vdw} = 4\varepsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] = \frac{A}{r^{12}} - \frac{B}{r^{6}}. \tag{2.5}$$

This is comprehensive of the attractive r^{-6} and repulsive contributions r^{-12}, rewritten in the left part of the equations 2.5 setting $A = 4\varepsilon\sigma^{12}$ and $B = 4\varepsilon\sigma^{6}$, where σ is the collision diameter and ε is the well depth.

Including all these parts in a unique formula the typical force field energy function is finally obtained

$$\begin{aligned} E = &\sum_{i}^{bonds} \frac{k_{bi}}{2}(r_i - r_{i,e})^2 + \sum_{i}^{angles} \frac{k_{\theta i}}{2}(\theta_i - \theta_{i,e})^2 + \\ &+ \sum_{i}^{torsions} \frac{V_i}{2}[1 + cos(n_i\omega_i - \gamma_i)] + \sum_{i}^{atoms} \sum_{j>i}^{atoms} C_{ij}\left(\frac{A_i A_j}{r^{12}} - \frac{B_i B_j}{r^6} + \frac{q_i q_j}{r_{i,j}}\right) \end{aligned} \tag{2.6}$$

Beyond this simple picture, the scheme may be complicated with the addition of terms for cross couplings between internal coordinates, accounting of hydrogen bonds, and specific terms for delocalized π systems or many other aspects linked to the chemical environment.[52, 53, 54]

2.1.3. Amber Force Field

As it can be guessed from the equations described in the previous Section, the construction of a force field may require a lot of parameters. These can be derived from experimental data, or, more often, from quantum mechanical calculations. Furthermore, some force fields are constructed specifically to reproduce some thermodynamical data, as in the case of the OPLS (Optimized Parameters for Liquid Simulations).[56]

The most important feature that a force field must have is transferibility. For this reason it is crucial that the set of parameters is able to reproduce the result of an entire family of molecular systems, like for example proteins and nucleotides or inorganic molecules. In this work I chose to adopt Amber, in its 1994 version, as implemented in the Gromacs package of softwares.

The history of Amber dates back to the first half of 1980s, when enough experience on proteins and nucleic acids had been accumulated to write the first version of a force field that was able to consider both families of compounds.[57] Over the years, many versions of the Amber force field have been published, with the support of more reliable experimental techniques and powerful computational tools, that made more accurate theoretical data available. The *ff94* version has been implemented with some accurate features, like for example an extended fit over the charges of several small protein conformations in order to assign atomic charges that are simultaneously precise and transferable. Careful analysis has been dedicated to solvent simulation, in both explicit and implicit form. Indeed the Amber force field has been designed to provide good results for biomolecules in water.[58] This can be a great caveat to bear in mind using this kind of method for vibrational spectroscopy of molecular systems *in vacuum*, although solvent-free simulations are feasible and permitted by using this specific force field.

In this work I tried to apply the Amber *ff94* force field within the semiclassical framework to provide vibrational frequencies of a single DNA base and of a small dipeptide, comparing the result with a more standard approach in which ab-initio molecular dynamics is employed. Such a usage of molecular mechanic is certainly weird, but it can be useful to the semiclassical community to understand to what extent a force field can provide a rough spectral density estimate.

Within the molecular mechanics community, application of a force field to vibrational spectroscopy is quite immediate and straightforward. In fact, the only approach that is commonly employed is the harmonic oscillator approximation of the potential energy surface, in the normal mode coordinate representation. This standard approach has been considered as a comparison term in all the applications of this thesis work.

All the calculations have been here performed using Gromacs, one of the most common software for molecular dynamics simulations.[59] It includes different force fields, but it is primarily oriented to biomolecular system treatment. It has been implemented to give the possibility to model each term in the electronic energy equation described above, accordingly to the parameter set of each force field. For example it is possible to choose either the harmonic or the Morse description of the equation for the oscillator that mimics the atomic bond stretching.

Certainly many other force fields exist, which are more accurate and appropriate for a gas-phase investigation. But as a first step I decided to start from a standard setup that can represent the "zeroth order investigation" of classical molecular mechanics exploitability.

2.2. Ab Initio Molecular Dynamics

In the introduction some existing theoretical methodologies to study vibrational spectra of molecular systems have been briefly mentioned. Also the difference between static and dynamics-based approaches have been highlighted. Between all the methods that rely on dynamics, Molecular Mechanics, as described above, has become the most common approach for investigation of molecular dynamics, especially in condensed phase. Of course the neglect of electronic degrees of freedom makes force fields intrinsically limited. All the electronic polarization effects are extremely approximated and also the chemical reactivity is in general ignored. For these reasons an Ab-Initio Molecular Dynamics, in which the electronic structure is evaluated on-the-fly by means of a quantum mechanical method, is strongly desiderable.

In a schematic manner, like in the Molecular Mechanic approach also the AIMD treats the system as a collection of N nuclei and N electrons, where the Born-Oppenheimer approximation is valid and the dynamics of the nuclei is considered classically on the ground-state electronic surface. Despite this, AIMD considers the electronic degrees of freedom and therefore permits chemical bond breaking and accounts for electronic polarization effects. Unfortunately, some

extent of the anharmonicity remains hidden also to this method. Mode coupling leading to combination bands and some purely quantum effects are still not considered.

All vibrational spectra in the AIMD approach are based on the Fourier transform of the autocorrelation functions: power spectra need atomic velocities, IR spectra rely on molecular dipole moments and Raman spectra are obtained from molecular polarizabilities. The power spectrum is the simplest to obtain from a trajectory, since it requires only nuclear velocities. It is the only kind of spectrum that I took into consideration throughout this thesis. In practice the velocity autocorrelation is calculated and summed up for each atom and consequently it is Fourier transformed to get the power spectrum:

$$I(E) = \lim_{T\to\infty} \int_0^T dt\, e^{iEt} \int\int dq_0 dp_0\; p(q_0, p_0)\, v(t)v(0), \tag{2.7}$$

where $p(q_0, p_0)$ represents the phase space density for sampling. An equivalent version can be obtained by applying a time average filter

$$I(E) = \lim_{T\to\infty} \int\int dq_0 dp_0\; p(q_0, p_0)\; \frac{1}{2T} \left| \int_0^T dt\; e^{iEt}\; v(t) \right|^2, \tag{2.8}$$

that helps the integrations to converge. Although the conformational phase space is explored in the microcanonical ensemble, where the energy of the system is conserved (NVE), in standard applications the effect of temperature can be evaluated with a preliminary thermalization stage (NVT).

The literature is plenty of vibrational analysis performed through application of AIMD. It is used for the evaluation of power spectra and absorption spectra via the calculations of the molecular dipole moment. Some works on the computational setup, like for example the choice of the simulation time step and the energy distribution between normal modes have been published by Horníček.[60] Applications on biologically relevant molecules have been presented by the Gaigeot group and by Cho and coworkers, while specific investigation on the effect of water molecule interactions have been explored by Silvestrelli and from the group of Iftimie and Tuckerman.[61, 62, 63, 64, 65] The advances done on the ab-initio methods, especially with the advent of DFT theory, permit to treat systems of medium size like dipeptides and small peptides.[66, 67] Furthermore in some works the dynamics is performed through the Car Parrinello Molecular Dynamics (CPMD) where the Schröedinger equation is calculated only at the beginning of the simulation and then propagated adiabatically alongside the nuclei instead of being solved step by steps as in

the more common Born-Oppenheimer Molecular Dynamics (BOMD).[12] Using CPMD, vibrational spectra of biologically relevant systems have been calculated both in gas phase and in acqueous solution.[64, 68, 69] Additionally some recent investigations on small amminoacids interacting with metal surfaces in condensed phase have been carried out by means of DFT-based MD, leading AIMD into the wide field of surface studies that are very important for nanodevices and biomaterials.[70, 71, 72]

AIMD has been largely used in this thesis both to simulate trajectories required from the semiclassical approach and to recover the spectral density in a classical framework using equations 2.7 and 2.8. A lot of considerations have been drawn from the comparison between these two strategies and will be extensively debated in all the applications presented. In general the ab initio molecular dynamics appeared to be a powerful tool to detect the fundamental frequencies and to discriminate the nature of each signal in the vibrational spectrum.

Part II.

Applications to molecular systems

3. Outline

In the previous chapters two approaches for molecular dynamics have been reviewed, classical mechanics and ab initio calculations, together with different techniques to obtain the vibrational power spectrum. In the following the validity of semiclassical methods will be demonstrated, with a specific emphasis on MC-SCIVR and on the most recently developed DC-SCIVR, that have been specifically thought for high dimensional molecular systems.

The first application is represented by the glycine molecule. Besides its importance as the smallest among the amminoacids and hence as a building block of proteins, it represent the limit case for the MC-SCIVR methodology in terms of dimensionality. The relevance of the semiclassical method over other static approaches in detecting the vibrational spectrum of a molecule that presents different accessible minima on the potential energy surface will be discussed.

As a follow up of this first analysis the semiclassical theory will be tested to reproduce the vibrational spectrum of protonated glycine both untagged and tagged with a increasing number of hydrogen molecules, with the aim to correctly detect a specific peak for which harmonic estimates previously reported in literature largely overestimate experimental findings. In order to perform such an analysis DC-SCIVR in conjunction with ab initio molecular dynamics has been adopted for the first time. Before applying it to this new system the method has been benchmarked on the neutral glycine and the agreement of DC-SCIVR with MC-SCIVR vibrational energy levels estimates will be shown. At this stage DC-SCVIR analysis can be reasonably performed on the protonated glycine set of systems demonstrating the agreement with experimental positions of all the peaks in the spectrum. The same accord between predicted and experimental frequencies don't occur when harmonic or classic methodologies are employed, highlighting the presence of quantum effects and hence the importance of the semiclassic investigation for such kind of systems.

Following this line, in the next chapter another DC-SCIVR application will be shown, involving the glycine molecule in its neutral and zwitterionic conformation, interacting this time with a different number of water molecules. The aim is the comparison with high level ab initio harmonic calculation previously reported in literature, that predicts a blue shift in the vibrational spectrum de-

pending on which form of the solvated glycine is stabilized by the solvation. This application will demonstrate that the semiclassical method can be successfully applied on supramolecular systems of high dimensionality, finding the same frequencies trend of a previous work in the literature.

Alongside the applications to high dimensional molecular systems, alternative strategies for the reduction of the computational effort of the semiclassical machinery is constantly under investigation. The last application has the aim to explore the possibility to exploit empirical force field to perform classic dynamics and Hessian calculation with a highly reduced computational cost. The validity of this approach in comparison with ab initio molecular dynamics will be discussed, evaluating the spectra of the deoxyguanosine, a nucleoside that constitutes the DNA, and of the Ac-Phe-Met-NH_2, a biologically interesting dipeptide.

All these applications can effectively bring to light the validity of the new DC-SCIVR semiclassical approach associated to the ab initio molecular dynamics in detecting the power spectrum of medium-size systems in isolated, supramolecular and also quasi-solvated forms. Following this trend the semiclassical method appears to be set to move toward high dimensional molecules or condensed phase systems.

4. On-the-fly ab initio Semiclassical Calculation of Glycine Vibrational Spectrum[1]

In this chapter I present an on-the-fly ab initio semiclassical study of vibrational energy levels of glycine. It is based on the multiple coherent approach (MC-SCIVR) integrated with monodromy matrix regularization for chaotic dynamics. All four lowest-energy glycine conformers are investigated by means of single-trajectory semiclassical spectra obtained upon classical evolution of on-the-fly trajectories with harmonic zero-point energy. For the most stable conformer I, additional on-the-fly trajectories are run for each vibrational mode with energy equal to the first harmonic excitation, and refined vibrational energies are eventually obtained. An analysis of trajectories evolved up to 50000 atomic time units demonstrates that, in this time span, conformer II and III can be considered as isolated species, while conformer I and IV show a pretty facile interconversion. However, this I-IV interconversion does not seem to have a substantial influence on vibrational energy levels, thus validating previous perturbative studies based on the assumption of isolated conformers.

4.1. Introduction

Glycine, the simplest among aminoacids, in addition to his evident importance in biology and medical sciences, has since long played a prominent role in both experimental and theoretical chemistry. On one hand, it is the prototypical structural unit of proteins, and it has been detected in the interstellar medium[73, 74] with important implications on theories about the origin of life on Earth; on the other, the presence of multiple shallow minima in the glycine potential energy surface (PES) represents a challenge for any theoretical application. One aspect shared by theory and experiment when investigating this simple but quaint

[1]This chapter is the reproduction with some minor changes of the paper **Fabio Gabas**, Riccardo Conte, and Michele Ceotto. "On-the-fly ab initio semiclassical calculation of glycine vibrational spectrum." *Journal of chemical theory and computation* **13.6** (2017): 2378-2388.

aminoacid is the elusiveness of its conformers. In fact, experimental microwave spectra have not always come up with consistent conclusions about the relative stability of glycine conformers,[75, 76] while theoretical studies have pointed out the presence of several minima with their relative stability being dependent on the level of electronic theory employed.[77, 78, 79, 80] Eventually, at least for some conformers, equilibrium structural parameters and fundamental transition (0->1) frequencies have been determined experimentally and confirmed by theoretical geometry optimization and anharmonic frequency estimates.[81] Anyway, the interest in this molecule is still very vivid and it keeps on stimulating new investigations that sometimes may even involve more elusive conformers.[82]

In their pioneeristic study (based on simulations at the DFT/B3LYP and MP2 levels of theory with several types of Dunning's correlation-consistent double-ζ basis sets[83]) Adamowicz et al.[84] have successfully employed the harmonic approximation to validate the assignments of glycine IR spectra obtained for the three most stable conformers isolated in low-temperature Argon matrices. Anharmonic effects have been later included in other studies. For instance, Gerber et al.[85] performed an investigation of glycine conformer I based on the vibrational self-consistent field (VSCF) method. They employed a semi-empirical electronic structure method with a coordinate scaling procedure to match harmonic frequencies, and further approximated the potential as a sum of single-mode and coupled two-mode contributions to ease computational costs. In another work, Fernandez-Clavero and co-workers[86] focused their attention on the far-infrared (low frequency) modes of glycine conformer I. They were able to include anharmonicity by adopting a variational method based on reduced-dimensionality model hamiltonians[87] with the remaining degrees of freedom allowed to relax. Hobza et al.[88] provided calculations of all 24 quantum anharmonic frequencies of conformer I. They employed a perturbative second-order vibrational perturbation theory (VPT2) approach[89, 90] with electronic calculations performed at the MP2 level of theory with aVDZ basis set. Furthermore, they also got accurate estimates for six high-frequency modes (i.e. O-H, N-H_2, C-H_2, and C=O stretching modes) of the other three main conformers. Finally, a comprehensive set of studies concerning all four more stable glycine conformers has been recently presented by Barone and co-workers, ranging from purely spectroscopical studies[91, 81] to reaction dynamics.[92] In their spectra simulations, Barone et al. adopted the VPT2 approach with variational treatment of resonances (the so-called deperturbed second-order vibrational perturbation theory (DVPT2)[11, 93]) associated to a fourth-order representation of the PES computed at the DFT/B3LYP level of theory with medium sized basis sets. The calculated frequencies are in very good agreement

with experimental data from Ref. [84].

A potential drawback of previously outlined approaches is that the single conformers are treated as independent species in isolated energy wells. However, the shape of the glycine PES (with its already mentioned shallow wells) certainly warrants a deeper analysis of the influence of conformer inter-conversion on vibrational frequencies. For this purpose, the semiclassical methods, as described in details in the theoretical part of this thesis, are by construction capable to yield quantum features starting from classical trajectories evolved on a global, full-dimensional PES.[94, 95, 18, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108] In particular, of relevance for the present application is the multiple-coherent semiclassical initial value representation (MC-SCIVR) that can be associated to ab initio on-the-fly dynamics with excellent results,[26, 109, 110, 47, 41, 40] a key feature when dealing with molecules the size of glycine or more. In fact, even if a large variety of very precise PES fitting techniques is available (see, for instance, Refs. [111, 112, 113, 114, 115, 116]) the computational burden to generate the hundreds of thousands of ab initio energies needed to accurately fit the PESs of such molecules is generally not affordable. For all these reasons, a semiclassical approach appears to be a straightforward choice for an investigation of multi-well effects, as indeed recently demonstrated by two of the authors in studying the ammonia vibrational spectrum.[45]

In this application, I employ on-the-fly MC-TA-SCIVR to estimate the vibrational frequencies of the four principal glycine conformers, and to investigate the role of conformer inter-conversion. The research also demonstrates that semiclassical methods are suitable to treat larger molecular systems and that they are not limited to model systems or small molecules.

4.2. Computational details

All electronic energy calculations including on-the-fly dynamics have been performed by means of the free NWChem suite of codes[117] at the DFT/B3LYP level of theory with aVDZ basis set. This choice was encouraged by several past studies, which have demonstrated that density functional theory with hybrid functionals and medium-size basis sets is capable to account very well for anharmonic effects.[118, 81, 119, 120] Figure 4.1 reports a sketchy representation of the glycine PES with the four lowest-energy conformers obtained upon optimization. The corresponding energy data are presented in Table 4.1 for both the conformers and the transition states between the wells. In addition, Table 4.2 reports the main geometrical parameters (bond lengths, planar and dyhedral angles). Conformers

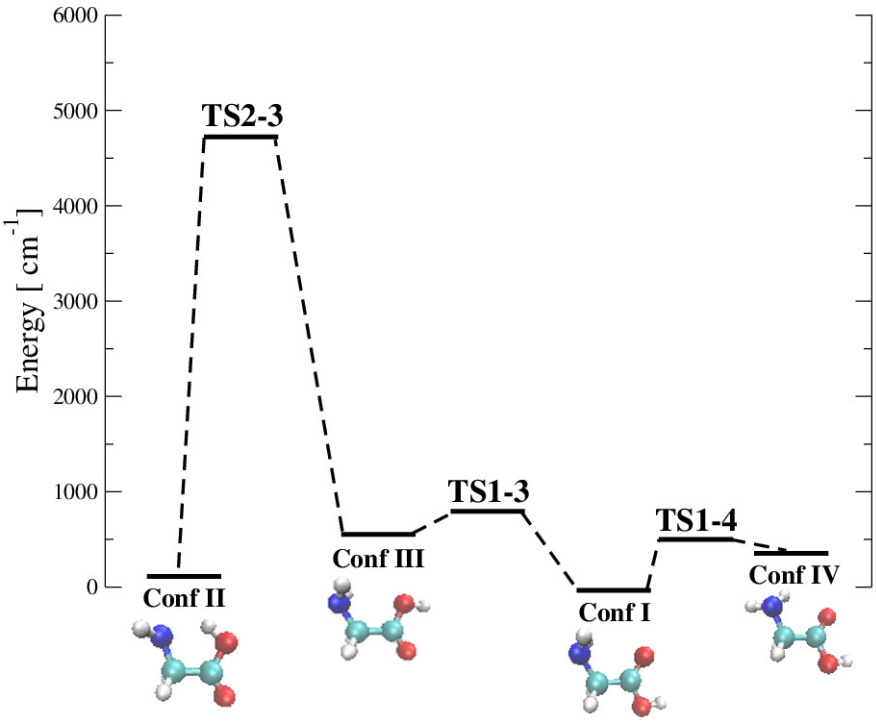

Figure 4.1.: Schematic representation of the glycine PES (cm^{-1}) with its four main conformers. The energy of Conf I is set to 0.

have planar, C_s symmetry (Conf I), or non-planar, C_1 symmetry (Conf II, III, IV). Harmonic frequencies have been obtained upon Hessian diagonalization at the four equilibrium geometries, and are reported in Table 4.3. Vibrational modes are labeled in decreasing order of frequency, according to Csaszar's notation.[78]

Full-dimensional on-the-fly trajectories have been generated starting from initial conditions (i.e. atomic equilibrium positions and atomic velocities) selected according to the specific conformer and the chosen internal energy. The molecule was given no rotation with all energy concentrated on vibrational modes. Trajectories have been evolved with time steps of 10 atomic time units each, for a total of 50000 a.u. (approximately 1.2 ps) for conformer I simulations and up to 25000 a.u. (or 0.6 ps) for the other conformers. For every conformer, a single ab initio trajectory was run with harmonic zero-point energy, i.e. with no quanta of excitation in any vibrational modes. For conformer I, the investigation has been

	Energy (cm^{-1}/ kcal/mol)
Conf I	0 / 0
Conf II	171 / 0.49
Conf III	570 / 1.63
Conf IV	457 / 1.31
TS 1-4	502 / 1.44
TS 1-3	772 / 2.21
TS 2-3	4765 / 13.62

Table 4.1.: Energetics of the four lowest-energy conformers of glycine.

refined by generating a new trajectory for each of the first 23 modes. Each of these trajectories was started with one quantum of harmonic excitation in the specific mode under examination. The lowest-frequency mode, mode #24, is an internal torsion that in past studies has often been neglected (see, for instance, Refs. [91, 81]). Following those investigations, and reckoning that when ν_{24} is singularly excited the total energy is anyway very close to the zero-point one, I have decided not to consider mode ν_{24} in the refinement procedure and report only its estimate as obtained from zero-point energy simulations.

MC-TA-SCIVR requires the evaluation of the Hessian matrix at each step along the trajectory, a computationally costly procedure that represents a bottleneck for the entire simulation. However, previous research has shown that the Hessian can be approximated by means of a compact finite-difference scheme (see also Appendix 1.d).[49, 39] I have tested and employed this approach for glycine and found out that to keep accuracy and avoid artificially shifted peaks in the spectrum, for the low-frequency excitations the Hessian can be calculated ab initio every three steps, while higher-frequency ones require exact Hessian evaluations every two steps.

Finally, I have faced the monodromy matrix instability issue typical of semiclassical methods. In classical dynamics, the monodromy matrix ($\mathcal{M}$) is defined as a 4-element matrix, with each element being itself a matrix made of the partial derivatives of instantenous positions or momenta with respect to initial positions or momenta. Eigenvalues of $\mathcal{M}$ yield an estimate of the trajectory stability, with large real eigenvalues associated to chaotic motion. In semiclassical dynamics, the elements of the monodromy matrix are used to evaluate the Herman Kluk prefactor, and their numerical stability is then strictly related to the reliability and feasibility of the whole semiclassical calculation. In particular, the determinant of $\mathcal{M}$ should be unitary at all times during the simulation, but finite-precision nu-

	Conf I	Conf II	Conf III	Conf IV
C-C	1.52	1.54	1.53	1.51
C-O*	1.36	1.34	1.36	1.36
C-O	1.21	1.21	1.21	1.21
C-N	1.45	1.47	1.45	1.46
H-O*	0.97	0.99	0.97	0.97
C-H	1.10	1.10	1.10	1.10
C-H	1.10	1.10	1.10	1.10
N-H	1.02	1.01	1.02	1.01
N-H	1.02	1.01	1.02	1.01
C-C-O	125.85	122.75	124.06	125.32
C-C-O*	111.35	113.93	113.41	111.67
O*-C-O	122.81	123.33	122.53	122.97
N-C-C	115.91	111.58	119.42	110.58
C-O*-H	107.18	104.89	106.50	106.99
C-C-H	107.51	107.00	105.90	108.47
N-C-H	109.84	111.85	109.56	110.68
C-C-H	107.51	107.00	105.93	105.44
N-C-H	109.84	111.85	109.60	114.93
H-C-H	105.70	106.70	105.52	106.40
C-N-H	110.32	112.56	111.09	111.15
C-N-H	110.32	112.56	111.09	110.11
H-N-H	105.73	107.86	106.55	108.39
O*-C-C-N	180.00	1.00	1.06	162.77
O-C-C-H	123.30	58.24	56.89	105.55
O*-C-C-H	-56.70	-121.65	-123.01	-72.41
O-C-C-N	0.00	-179.11	-179.04	-19.27
O-C-C-H	-123.30	-55.86	-54.87	-140.81
O*-C-C-H	56.70	124.24	125.23	41.23
H-O*-C-C	180.00	-0.24	179.87	176.78
H-O*-C-O	0.00	179.87	-0.03	-1.24
H-N-C-C	58.22	117.21	59.35	38.43
H-N-C-H	179.71	119.09	178.54	158.66
H-N-C-H	-63.86	-122.96	-63.03	-81.22
H-N-C-C	-58.22	-120.65	-59.08	158.55
H-N-C-H	63.86	-3.05	63.20	39.33
H-N-C-H	-179.71	-0.83	-178.37	-80.80

Table 4.2.: Geometrical parameters for the equilibrium configuration of the four conformers calculated at the DFT/B3LYP level of theory with aVDZ basis set. Bond distances are in Angstrom. Angles are in degrees. O* indicates the oxygen atom bonded to the hydrogen one.

Mode	Conf I	Conf II	Conf III	Conf IV
ν_1	3735	3612	3735	3740
ν_2	3568	3528	3583	3594
ν_3	3495	3448	3504	3501
ν_4	3089	3112	3091	3070
ν_5	3051	3061	3054	2957
ν_6	1804	1830	1796	1808
ν_7	1656	1646	1653	1618
ν_8	1438	1449	1437	1474
ν_9	1384	1416	1370	1435
ν_{10}	1371	1333	1344	1318
ν_{11}	1294	1328	1338	1252
ν_{12}	1175	1212	1180	1209
ν_{13}	1158	1159	1166	1137
ν_{14}	1120	1068	1133	1105
ν_{15}	911	915	904	1014
ν_{16}	908	886	876	840
ν_{17}	816	869	791	813
ν_{18}	647	809	678	658
ν_{19}	629	639	590	620
ν_{20}	510	547	514	519
ν_{21}	458	508	494	462
ν_{22}	249	312	255	276
ν_{23}	208	238	243	168
ν_{24}	56	27	14	95
Harmonic zpe	17364	17478	17372	17341

Table 4.3.: Harmonic frequencies and zero-point energies (cm^{-1}) for the four glycine conformers (DFT/B3LYP level of theory, aVDZ basis set). Mode labels are after Csaszar.[78]

merical integration is not accurate enough to keep it constant in the instance of a highly chaotic trajectory. A rigorous way to keep trace of the monodromy matrix stability is based on the evaluation of the determinant of the positive-definite matrix $\mathcal{D} = \mathcal{M}^T\mathcal{M}$.[50] In the glycine simulations, as soon as the determinant of $\mathcal{D}$ differed from unity more than 10^{-2}, the trajectory (and thus the simulation) was stopped. This generally happened between 15000 and 30000 a.u. depending on the internal excitation (usually the higher the excitation, the faster the rejection), type of conformer (non-planar conformers have been found to be more dynamically unstable), and the Hessian approximation adopted (which hinders preservation of determinant unitarity). To overcome the problem and be able to perform long-time simulations, I have employed a monodromy matrix regularization technique based on the discard of the largest eigenvalue of $\mathcal{M}$ whenever it becomes greater than a chosen threshold value (generally of the order of 10^3 - 10^4). Recently, this procedure has been tested on a set of model and molecular systems, demonstrating that it is able to preserve the accuracy of semiclassical results.[51] The technique prevents the semiclassical pre-exponential factor from becoming numerically unstable, thus allowing semiclassical spectra to be based on longer-time dynamics (see also Appendix 1.d).

4.3. Results and discussion

My first investigation concerns the calculation of zero-point energy and fundamental frequencies for the global minimum conformer. Figure 4.2 shows the peaks obtained with the multiple coherent approach by means of a single trajectory with internal energy equal to the harmonic zpe. The figure is divided into three parts, respectively a low-frequency or fingerprint region (zpe peak plus modes #24 - #17), a medium-frequency section (modes #16 - #9), and a high-frequency part (modes #8 - #1). The intensity of all peaks has been normalized to that of the zpe peak, and harmonic frequencies (previously listed in Table 4.3) are reported as dashed, vertical lines to help appreciate the anharmonicity of the quantum frequencies. A refinement of calculated frequencies is obtained by employing a different trajectory with specific harmonic excitation (one quantum) for each mode. Figure 4.3 reports the outcome of this second approach, and Table 4.4 shows a detailed numerical comparison between MC-TA-SCIVR, DVPT2, and experimental results.

Peaks in the two Figures are well resolved, and frequencies are generally in qualitative good agreement (around 35 cm^{-1} or less for a large majority of modes) with previous DVPT2[91] and experimental works.[84] Such an accu-

racy is remarkable if compared to mean absolute errors found when applying semiclassical methods to much smaller molecules like ammonia (38 cm^{-1})[45] or H_2O (about 20 cm^{-1}).[51] However, differently from ammonia or water, exact quantum mechanical results are not available for glycine and a straightforward quantitative comparison between the reported data is not possible, since DVPT2 studies employed a different basis set from the one here adopted, and experimental spectra always come with some uncertainties in their interpretation. For these reasons these comparisons are intended to be mainly qualitative. From inspection of the two Figures, I note that anharmonicity is often increased in the refined procedure (Figure 4.3), and that higher anharmonicity is usually found for the very low-frequency modes, associated to large amplitude motions, or, on the opposite, for the high-frequency modes involving the O-H (mode #1), and C-H_2 (mode #4 - #5) stretches, in agreement with previous studies.[88] The lower MAE value of MC-TA-SCIVR simulation coming from the single trajectory with harmonic zpe energy with respect to the one coming from the refinement procedure is here to be ascribed to an accidental compensation of errors.

I then move to a comparison of vibrational frequencies regarding conformers II, III, and IV. Experimental data are not available for the latter, and they are also scarse for the former two. As reported in Table 4.5, and with the previously illustrated caveat about data comparisons, MC-TA-SCIVR results (shorthanded as MC in Table 4.5) are in good agreement with DVPT2 and the experiment, especially for the mid-frequency modes. The two symmetry groups to which the four glycine conformers belong (C_s or C_1) have only monodimensional representations, but, in spite of this, a few modes are degenerate when calculated semiclassically. This is actually an accidental outcome of these calculations that involves modes which are very similar regarding both frequency and type of motion. A couple of comments does support this claim. On one hand, also other approaches are not totally immune from this accidental degeneracy drawback. For instance, the two C-H_2 stretches (mode #4 and #5) are nearly degenerate in DVPT2 simulations (see data for conformers II and III in Table 4.5). On the other, semiclassical approaches in general, and specifically MC-TA-SCIVR ones, are fully able to properly and correctly account for nuclear symmetry, as demonstrated in several previous works.[37, 38, 26, 45, 40]

A peculiar feature of MC-TA-SCIVR is the capability to detect overtones at no additional cost. In fact, the reference state depends on the choice of the $\varepsilon(j)$ parameters. As described in the theoretical part of this thesis, when a trajectory is given an initial single harmonic excitation in the j-th mode, then the reference state is built with a value $\varepsilon(j)$ set equal to -1. In this way, not only the peak relative to the fundamental transition of the j-th mode is selected, but it is also

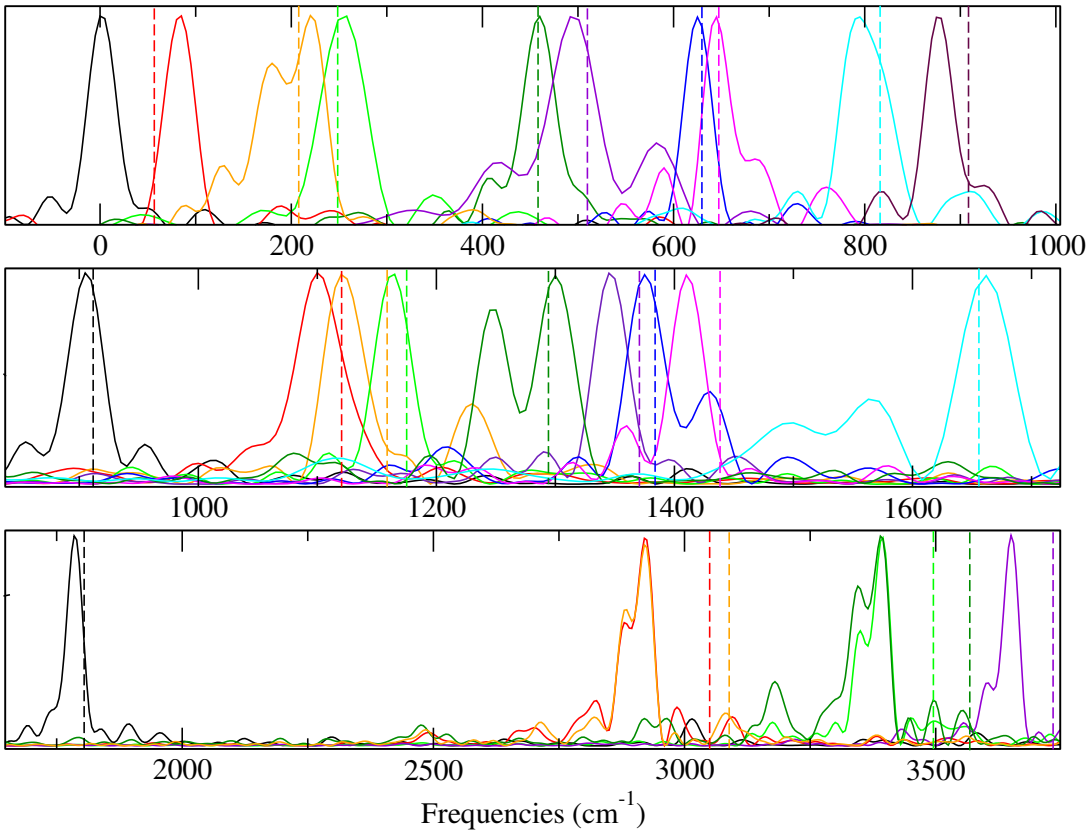

Figure 4.2.: Composition of glycine semiclassical power spectra calculated via MC-TA-SCIVR and a single classical trajectory with harmonic zero-point energy. Upper panel: zero-point energy and modes #24 - #16. Middle panel: modes #15 - #7. Bottom panel: modes #6 - #1. The trajectory was evolved on-the-fly for 50000 a.u. The anharmonic zero-point energy has been set to 0. Vertical dashed lines indicate the harmonic frequencies. Peak intensities have been individually normalized to that of the most intense peak.

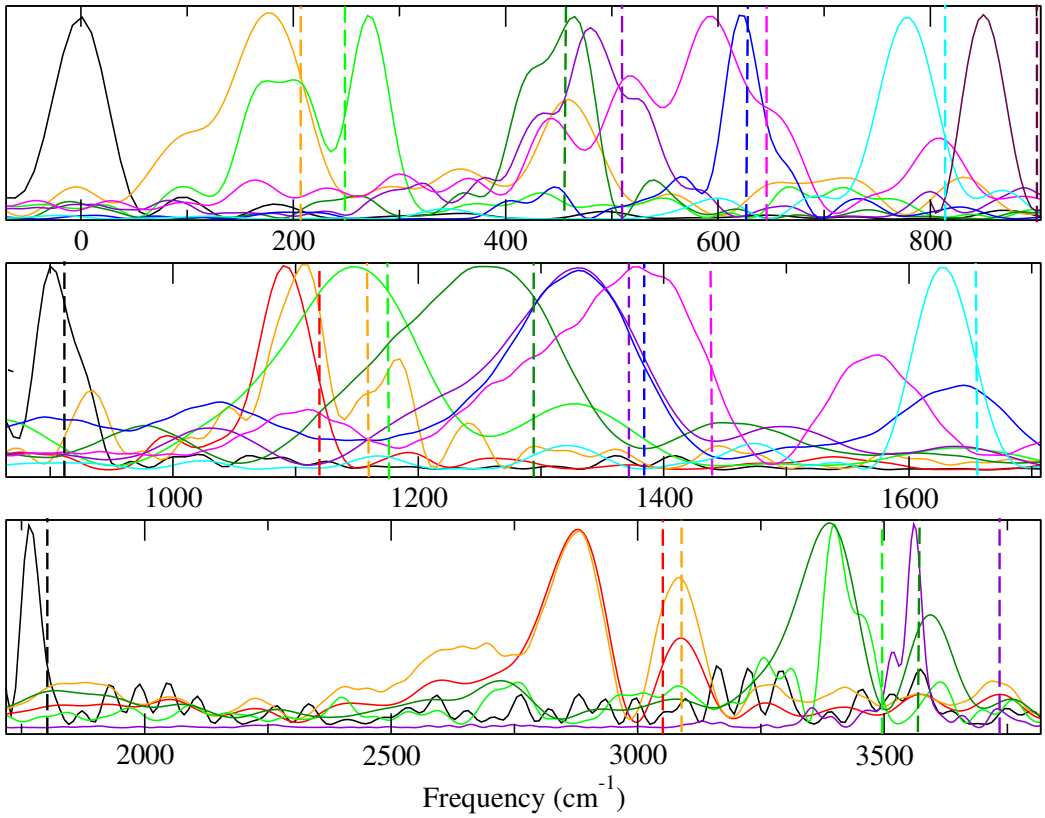

Figure 4.3.: Composition of refined MC-TA-SCIVR power spectra calculated from single trajectories with one quantum of harmonic excitation. Upper panel: zero-point energy and modes #23 - #16. Middle panel: modes #15 - #7. Bottom panel: modes #6 - #1. Trajectories were evolved on-the-fly for 50000 a.u. The semiclassical zero-point energy has been set to 0. Vertical dashed lines indicate the harmonic frequencies. Peak intensities have been individually normalized to that of the most intense peak.

Conf I	MC-TA-SCIVR (zpe)	MC-TA-SCIVR (1 exc)	DVPT2[a]	Experiment[b]
ν_1	3650	3565	3575	3585
ν_2	3390	3395	3418	3410
ν_3	3395	3405	3367	3359
ν_4	2920	2885	2961	2969
ν_5	2920	2885	2947	2943
ν_6	1785	1765	1774	1779
ν_7	1675	1625	1612	1608
ν_8	1410	1380	1435	1429
ν_9	1375	1330	1387	1405
ν_{10}	1345	1330	1353	1340
ν_{11}	1300	1250	1286	1297
ν_{12}	1165	1150	1164	1166
ν_{13}	1120	1105	1144	1136
ν_{14}	1100	1090	1103	1101
ν_{15}	905	900	863	883
ν_{16}	875	845	907	907
ν_{17}	795	785	802	801
ν_{18}	645	600	603	619
ν_{19}	625	625	633	615
ν_{20}	490	485	494	500
ν_{21}	460	470	461	458
ν_{22}	260	275	255	250
ν_{23}	220	180	203	204
ν_{24}	85	-	-	-
MAE	21	30	8	
zpe	17160	17160	-	-

Table 4.4.: Comparison between calculated anharmonic vibrational frequencies for conformer I and corresponding experimental values. Data and Mean Absolute Errors (MAE) are reported for MC-TA-SCIVR (1 trajectory with zero-point energy (zpe), or with one quantum of harmonic excitation (1 exc)), deperturbed second-order vibrational perturbation theory,[a] and experiment.[b]
[a]from ref. [91]
[b]from ref. [84].

	Conf II				Conf III				Conf IV		
	Harm	MC	DVPT2[a]	Exp[b]	Harm	MC	DVPT2[a]	Exp[b]	Harm	MC	DVPT2[c]
ν_1	3612	3390	3440	3410	3735	3645	3576	3560	3740	3655	3579
ν_2	3528	3420	3373		3583	3492	3425	3410	3594	3385	3441
ν_3	3448	3390	3235	~3275	3504	3411	3371		3501	3439	3361
ν_4	3112	2940	2955		3091	2835	2933		3070	2953	2956
ν_5	3061	2970	2953	2958	3054	2835	2932	2958	2957	2755	2866
ν_6	1830	1785	1821	1790	1796	1774	1779	1767	1808	1801	
ν_7	1646	1620	1618	1622	1653	1701	1641	1630	1618	1585	
ν_8	1449	1410	1431	1429	1437	1431	1417	1429	1474	1405	
ν_9	1416	1380	1377	1390	1370	1359	1339		1435	1270	
ν_{10}	1333	1320	1322		1344	1368	1318	1339	1318	1279	
ν_{11}	1328	1290	1294		1338	1305	1339		1252	1171	
ν_{12}	1212	1170	1169	1210	1180	1170	1153		1209	1180	
ν_{13}	1159	1140	1145		1166	1179	1128	1147	1137	1036	
ν_{14}	1068	1035	1044		1133	1134	1105	1101	1105	1090	
ν_{15}	915	900	899	911	904	909	895		1014	991	
ν_{16}	886	915	851	867	876	828	827	852	840	820	
ν_{17}	869	810	840		791	828	770	777	813	684	
ν_{18}	809	780	805	786	678	675	656	644	658	658	
ν_{19}	639	615	633		590	594	604		620	559	
ν_{20}	547	540	543		514	531	499		519	523	502
ν_{21}	508	510	502		494	522	488		462	485	467
ν_{22}	312	285	299	303	255	261	260		276	190	278
ν_{23}	238	240	229		243	225	224		168	225	156
MAE	54	26	18		53	48	15				
zpe	17478	17190			17372	17220			17341	17080	

Table 4.5.: Comparison between harmonic (Harm), MC-TA-SCIVR (MC), DVPT2, and experimental estimates of fundamental frequencies for glycine conformers II, III, and IV. Single trajectories with zero-point energy have been employed for the MC-TA-SCIVR calculations.
[a] from ref. [91] .
[b] from ref. [84].
[c] from ref. [81].

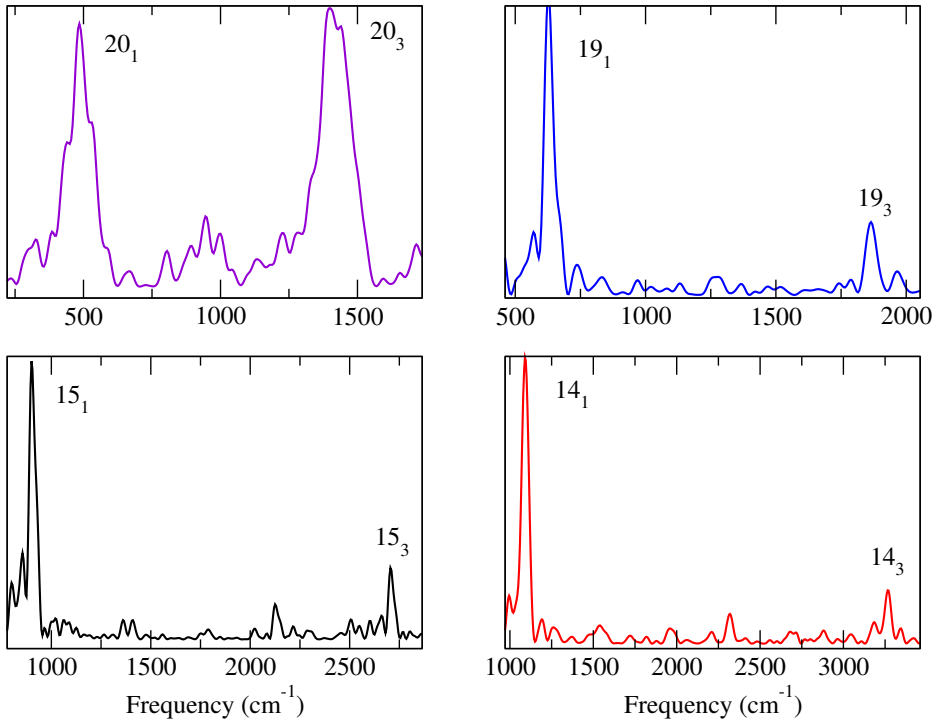

Figure 4.4.: Fundamental and second overtone frequencies obtained via MC-TA-SCIVR from single trajectories with one quantum of harmonic excitation. Peak frequencies are (cm^{-1}): $20_1 = 485$, $20_3 = 1400$; $19_1 = 625$, $19_3 = 1865$; $15_1 = 900$, $15_3 = 2705$; $14_1 = 1090$, $14_3 = 3265$.

possible to identify spectral signals associated to an odd number of excitations in that mode. As a result, for some modes I have been able to get a well-resolved peak for the second overtone (i.e. at the frequency corresponding to a triple excitation of the mode). Four examples are shown in Figure 4.4.

A potential drawback, which could lead to noisy or unreliable spectra, of a method based on classical trajectories started with given initial quantized conditions (like MC-TA-SCIVR) is what is commonly refered to as "zero-point energy leak". Zero-point energy leak is due to the fact that once the classical dynamics is started, then there is no warranty that each mode preserves its quantized energy during the trajectory evolution. As a consequence, some modes might end up having less energy than required by the zero-point motion, while others heat up even considerably. The issue has to be carefully investigated in systems, like glycine, where energy can flow between many coupled degrees of freedom, and especially in simulations based on a single trajectory that cannot rely on the

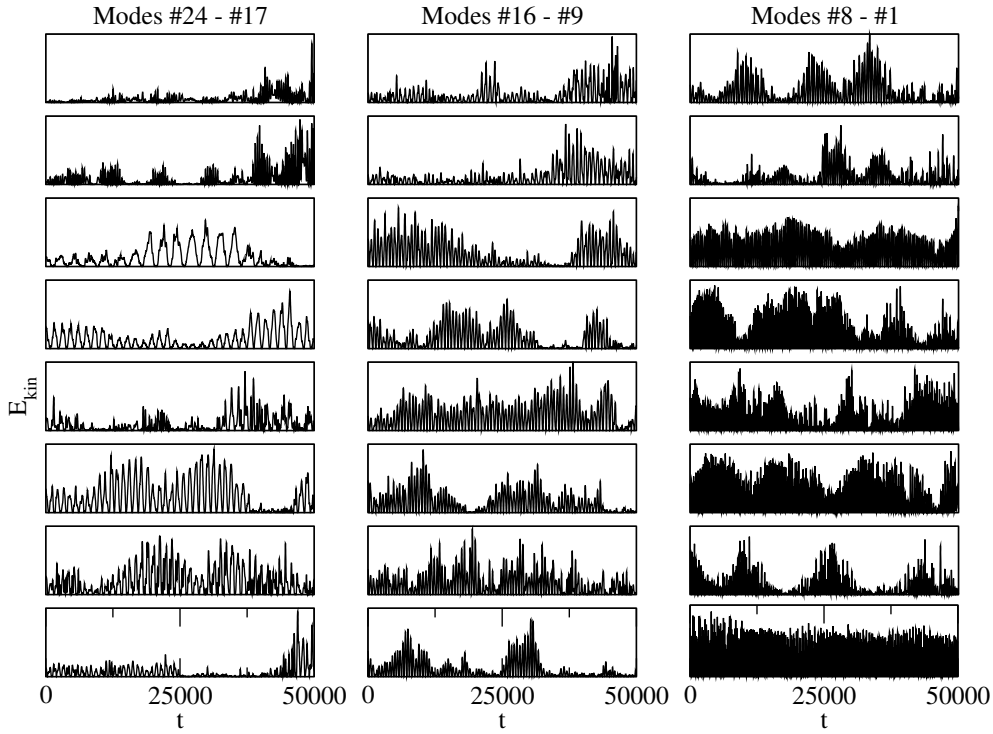

Figure 4.5.: Partition of the instantaneous kinetic energy among the 24 vibrational modes of glycine. The trajectory has been run starting from the equilibrium geometry of conformer I and with harmonic zero-point energy.

"wash-out" of an ensemble average. In Figure 4.5 I report the distribution of the classical kinetic energy among glycine vibrational modes as a function of time. The ideal behavior, typical of a set of completely uncoupled oscillators, would be a periodic motion in the kinetic energy plots demonstrating that the internal vibrational energy in the different modes is constant and transforms back and forth between potential and kinetic forms. This is with good approximation found in Figure 4.5, especially at short evolution times. Close to the end of the simulation it appears that some of the low-frequency modes (mode #24, #23, #21) start to gain energy at the expenses mainly of the O-H stretch (mode #1). However, this effect is only marginal and indeed does not affect the quality of the semiclassical spectra.

To complete this Section, I focus on the possibility to have a substantial "multi-well" effect on vibrational frequencies. For a molecule like glycine, the

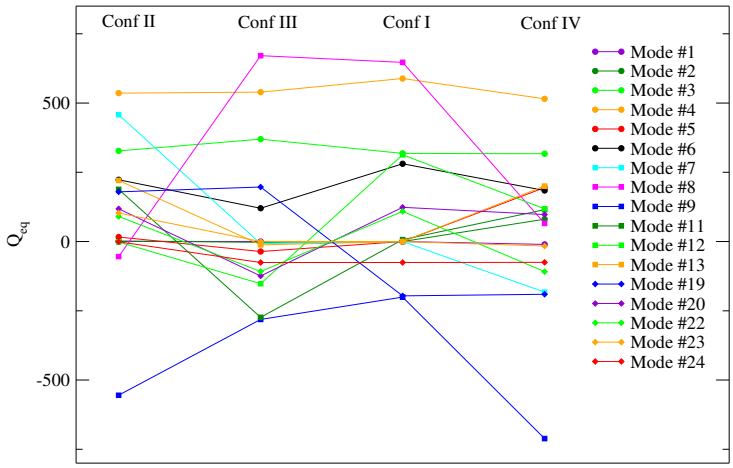

Figure 4.6.: Normal mode equilibrium coordinates for the four conformers. For clarity of the Figure, some modes are not reported since they are flat around $Q_{eq} = 0$.

conformers should not be considered as isolated ones due to their relatively low interconversion barriers. Modes related to large-amplitude motions along the inter-conversion path, in fact, could in principle be largely influenced by the presence of multiple wells. For this reason, precise perturbative treatments should rely on a global surface, which, however, is computationally prohibitive for a molecule the dimension of glycine. One advantage of MC-TA-SCIVR over perturbative methods is that it naturally includes this "multi-well" effect if the classical trajectories visit several wells. To assess the importance of the effect in the present study, I have developed a procedure based first on trajectory investigation, and then on a comparison between frequencies calculated from trajectories of different length. Figure 4.6 shows the equilibrium normal mode coordinates for most vibrations.

For some of the modes, the equilibrium coordinate does not change moving from a conformer to another. I interpret these modes as those not linked to any inter-conversion paths. On the contrary, there are modes that have substantially different equilibrium coordinates for different conformers. These are likely related to the inter-conversion path and, when excited, have a chance to lead to a change in glycine conformation. Following this idea, I have run and analysed trajectories with single harmonic excitation starting from all conformers up to 50000 atomic time units. No inter-conversion between conformer II and conformer III has been found, nor between conformer III and conformer I. The

inter-conversion between conformer I and conformer IV was pretty facile, instead, sometimes taking place even within an evolution time of just 25000 a.u. For some modes, the I-IV inter-conversion takes place after an evolution time larger than 25000 but shorter than 50000 a.u. Consequently, the "multi-well" effect (if present) should reveal itself with a substantial modification of the frequencies of these modes depending on whether they were calculated on the basis of a short-time (25000 a.u.) or long-time (50000 a.u.) dynamics. I have indeed performed these calculations and found out that with just one exception (i.e. mode #22 which shifts from 184 to 275 cm^{-1}) the frequencies of the involved modes are not modified by the "multi-well" effect.

4.4. Conclusions

The present work has its natural development in the investigation of vibrational frequencies of glycine in its protonated and zwitterionic forms. Results for neutral glycine are encouraging, because they have been obtained with very good accuracy through a computationally cheap approach. A study of vibrational motion of protonated glycine will be highly valuable. The species is found in solution,[121, 122, 123] and it is important in biological processes where it acts as an intermediate in many biochemical reactions. This kind of system will be discuss in next chapters in this thesis.

5. DC-SCIVR on neutral Glycine

The application presented in the previous chapter provides the size limit for standard MC-SCIVR calculations. As it has already been elucidated within the theoretical section, in order to have a clear signal in the spectrum it is mandatory that the evolved coherent state has a significant overlap with its initial state. In the first application I have shown a clear power spectrum of the 24 degrees of freedom neutral glycine molecule. Unfortunately even a single step forward, represented by the protonated glycine, did not provide reasonable results. To overcome this limitation, in my group, the Divide-and-Conquer approximation (DC-SCIVR) was developed (see theoretical section). By applying this new approach, partitions of smaller dimensional spaces substitute the totality of the degrees of freedom. In this way it is possible to handle not only protonated glycine, but also all the bigger systems that I am going to illustrate further in following chapters.

To prove the reliability of DC-SCIVR, I first employed it on conformer I of neutral glycine, in order to recover the same signals already evaluated with the multiple coherent approach. The simulation used the same set of information coming from the MC-SCIVR analysis, namely a trajectory 1.2 ps long with initial conditions in momenta that reflect the harmonic ZPE on each normal mode. The subdivision was done according to the Hessian criterion leading to six subspaces, the largest composed by 18 normal modes, followed by one subspace made of 2 modes and finally four monodimensional subspaces. In Table 5.1 all the fundamental frequencies obtained with both methodologies are reported, while in Figure 5.1 the Divide-and-Conquer spectral density is illustrated.

The results clearly show the validity of the Divide-and-Conquer approximation. With the exception of a small number of modes, the difference is within 10 cm^{-1}, that is a very good accuracy since the error is approximately equal to 20-25 cm^{-1}. The largest deviation is recovered for ν_{11}. In this specific case, both the full and the reduced dimensional analysis highlight a double peak. For the full dimensional case the highest intensity occurs for the peak nearest to the experimental value, while in the DC-SCIVR simulation to the farthest. Nevertheless the shape and the structure of the band is similar. Noticeably the deviation of these two approaches is minimal in the highest energy region of the spectrum. i.e. 1700-3600 cm^{-1},

mode	MC-SCIVR	DC-SCIVR	Δ
ν_1	3650	3650	0
ν_2	3390	3390	0
ν_3	3395	3390	5
ν_4	2920	2920	0
ν_5	2920	2920	0
ν_6	1785	1780	5
ν_7	1675	1675	0
ν_8	1410	1405	5
ν_9	1375	1360	15
ν_{10}	1345	1340	5
ν_{11}	1300	1240	60
ν_{12}	1165	1155	10
ν_{13}	1120	1115	5
ν_{14}	1100	1095	5
ν_{15}	905	900	5
ν_{16}	875	870	5
ν_{17}	795	820	25
ν_{18}	645	690	45
ν_{19}	625	620	5
ν_{20}	490	510	20
ν_{21}	460	455	5
ν_{22}	260	255	5
ν_{23}	220	215	5
ν_{24}	85	75	10
ZPE	17160	17100	60

Table 5.1.: Comparison between glycine vibrational levels calculated with the full-dimensional MC-SCIVR and with the DC-SCIVR approximation.

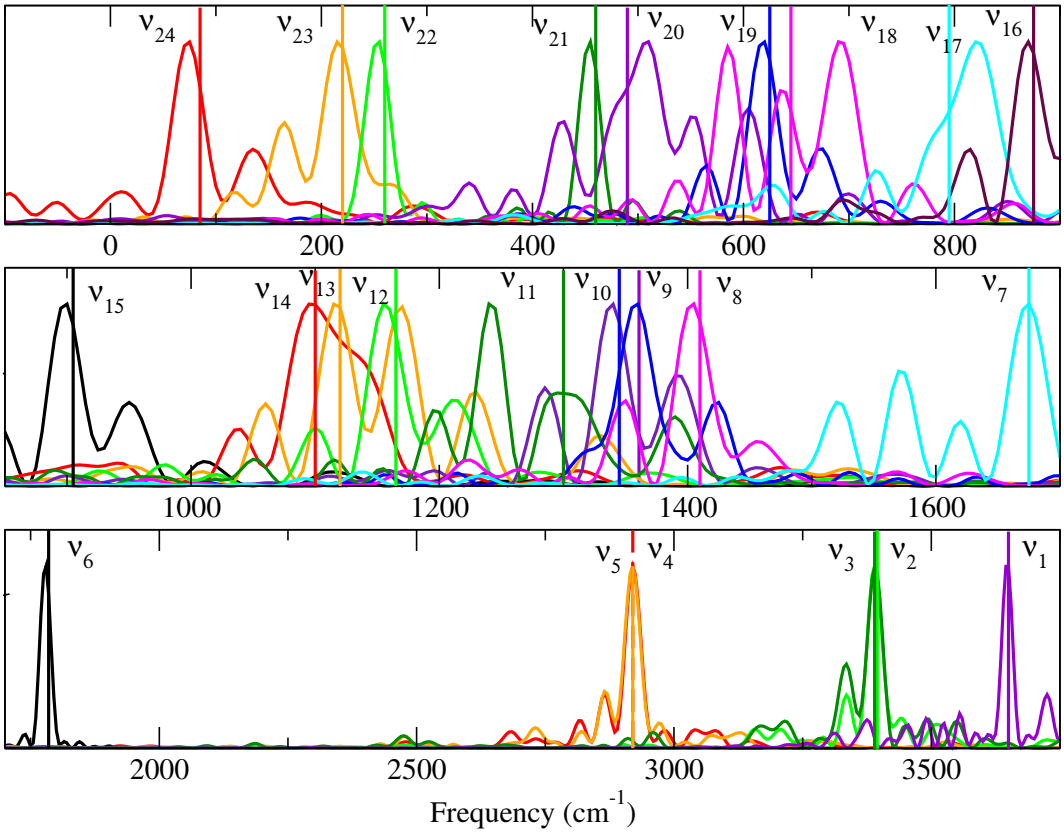

Figure 5.1.: Comparison of the power spectrum coming from a DC-SCIVR study with the results obtained with MC-SCIVR (solid lines)

that is the area that usually has been investigated. Unfortunately the Zero Point Energy is not clearly representable in the new DC-SCIVR method, because each subspace presents its own contribution. An estimate can be recovered by summing up the contributions from all subspaces, but it results quite off-the-mark with respect to the one calculated with MC-SCIVR. For this reason the peak of the ZPE is not reported in Fig. 5.1.

In light of these results I have adopted the Divide-and-Conquer methodology for all other applications I will illustrate in this thesis. As I will report in the next sections, I can confirm the validity and accuracy of DC-SCIVR for applications not only to medium and big size molecules, but also to more complex systems like supramolecular aggregates or molecular ensembles governed by weak interaction.

6. Protonated Glycine Supramolecular Systems: the need for quantum dynamics[1]

Although IR spectroscopy is one of the most commonly employed techniques to study molecular vibrations and interactions, characterization of the experimental IR spectra is not always straightforward. This is the case of protonated glycine supramolecular systems like $(GlyH + nH_2)^+$, whose IR spectra open questions which have still to find definitive answers even after theoretical spectroscopy investigations. Specifically, structural hypothesis formulated to explain the main experimental spectral features of $(GlyH + nH_2)^+$ systems have not been theoretically confirmed. In this work it has been demonstrated that simulations must account for quantum dynamical effects in order to solve this open issue. This is achieved by means the divide-and-conquer semiclassical initial value representation technique, which approximates the quantum dynamics of high dimensional systems with remarkable accuracy and overperforms not only the commonly employed but unfit scaled-harmonic approaches, but also pure classical dynamics simulations.

6.1. Introduction

Protonated systems are ubiquitous in chemistry.[124] They are involved in many different processes and instances, ranging from organic reactions and intermediates to biological and interstellar-medium events. Furthermore, protonation is determinant for the chemical properties of heteroatomic compounds, such as amino acids. For instance, hydrolysis of amides, peptides, and proteins at biological pH is initiated and driven by the process of protonation. In general,

[1]This chapter is the reproduction (with some changes) of the paper "Protonated Glycine Supramolecular Systems: the need for quantum dynamics" *Chemical Science*, **9**, (2018): 7894-7901 by **Fabio Gabas**, Giovanni Di Liberto, Riccardo Conte, and Michele Ceotto. This paper has been selected as "Pick of the Week" by the editorial board of Chemical Science. Furthermore, I have prepared a cover art image that has been chosen as front image for an issue of Chemical Science.

the electronic and conformational structure of proteins as well as their dynamics are strongly influenced by protonation with the resulting three-dimensional structure playing a key role in their biological activities.

Protonated glycine compounds are pivotal examples of protonated systems because they are the smallest building blocks of more complex biological entities, and a full comprehension of their dynamics is indeed essential for a correct understanding of the stability and reactivity of many other protonated systems. For this reason, in the past, protonated glycine compounds have been the subject of extensive experimental and computational studies.[125, 126, 127, 128, 67, 69, 129, 85] However, there are some fundamental questions about these systems which are still open. Specifically: to what extent is the proton shared between the amide and the carboxylic group? Is it a static or dynamical effect? Do nuclear quantum mechanical contributions play a major or a minor role for the properties of these protonated compounds? One should expect very peculiar quantum mechanical effects when the proton is shared between nucleophilic groups. The main reason for this expectation is that the proton is the only ion with basically zero ionic radius and it has the lightest mass. These peculiarities are at the origin of proton mobility and reactivity, and one would expect quantum mechanical contributions to be determinant. This investigation aims at providing the answers to the open questions illustrated above and at estimating the impact of quantum mechanical effects by comparing quantum and classical simulations versus available experimental results.

Frequently, in experimental IR works, the obtained spectra are interpreted and evaluated with the support of theoretical calculations. One of the most commonly used theoretical approach is the scaled-harmonic technique. In this method,[130] first the normal mode frequencies (i.e. the purely harmonic frequencies of vibration) at the minimum geometry are calculated by diagonalizing the equilibrium nuclear Hessian matrix and by taking the square roots of the Hessian eigenvalues. Then, they are scaled to account for anharmonicity and coupling between modes. Such an approach is widely employed since it is easily doable even for large size molecules as it only requires calculation of a single Hessian matrix. However, the approach remarkably neglects any dynamical and anharmonic effects that may become crucial when interactions such as hydrogen bonds dominate the interaction picture.[131] Even if several research groups have provided full sets of scaling constants for the different levels of theory and electronic basis sets employed[130, 132] as well as different scaling constants for calculations of different observables (frequency, zero point energy, enthalpy, entropy, etc.), the scaled harmonic approach is misleading for the interpretation of the glycine proton-bound dimer spectrum. Furthermore, it is generally classified

as an *ab initio* method in an improper way, since an empirical tuning parameter is enforced. A computational technique able to account for conformational and dynamical effects should be conveniently based on (quantum) molecular dynamics, since the dynamics allows to explore the actual Potential Energy Surface (PES) even far from the harmonic region.[133, 134, 135] Such a non-local approach may be crucial for a correct interpretation of hydrogen bonding.

The quantum dynamical way to spectroscopy and frequency computation is given by the Fourier transform of the autocorrelation of the time-evolved nuclear wavepacket averaged over the quantum density matrix of vibrational states (i.e. the power spectrum)[136]

$$\mathrm{I}_{qm}\left(E\right)=\frac{1}{2\pi\hbar}\int_{-\infty}^{+\infty}\mathrm{e}^{iEt/\hbar}\left\langle\Psi\left(0\right)|\Psi\left(t\right)\right\rangle\mathrm{d}t. \tag{6.1}$$

Eq.(6.1) includes all quantum mechanical spectroscopic information like zero point energy (ZPE), fundamental and overtone frequencies, anharmonicities and couplings, tunneling effects as well as quantum resonances between overtones and fundamentals. The classical equivalent of Eq.(6.1), reported also in the section 2.2 is the Fourier transform of the velocity autocorrelation function

$$\mathrm{I}_{cl}\left(\omega\right)=\frac{1}{2\pi\hbar}\int_{-\infty}^{+\infty}\mathrm{e}^{i\omega t}\left\langle\boldsymbol{v}\left(0\right)\cdot\boldsymbol{v}\left(t\right)\right\rangle\mathrm{d}t \tag{6.2}$$

Eq.(6.2) is limited to the calculation of classical fundamental frequencies, mode couplings and resonances. In other words, it accounts for the classical contribution to anharmonicity only. Anyway, both approaches are dynamical and represent a step forward with respect to single point harmonic calculations.

Unfortunately, when dealing with high dimensional systems, purely quantum mechanical simulations based on Eq.(6.1) are out of reach because of the so-called curse of dimensionality problem. Furthermore, accurate and fast-to-evaluate analytical PESs are usually not available and must be replaced by more computationally expensive *ab initio* "on-the-fly" calculations, whereby the dynamics can be performed (even if at a lower level of electronic theory), and which demand for a theoretical formalism that permits a convenient interface to them. As already discussed, semiclassical dynamics can be interfaced to *ab initio* "on-the-fly" calculations straightforwardly so I adopted it to calculate $\mathrm{I}_{qm}\left(E\right)$. Furthermore, in the previous part of this thesis, I have demonstrated the validity of the Divide-and-Conquer semiclassical strategy, both in tackle medium-size systems and in decrease the computational time demanding. Despite the success of the semiclassical simulations, one key methodological question remains open.

In fact, if on one hand quantum effects are hallmarks of spectroscopy, on the other hand for systems of high dimensionality it could be argued that a classical picture be enough to describe with sufficient accuracy the spectral features of at least fundamental transitions. This would require much less effort since a semiclassical simulation is significantly more computationally intense than a classical one. Eq.(6.2) requests to calculate at each time step the cartesian velocities $v(t)$ of the nuclei only. Instead I want to remark here that, in the DC-SCIVR, the calculation of $I_{qm}(E)$ implies to evaluate not only the position and the velocities of the nuclei at each time step, but also the nuclear Hessian (for evolution of the phase term). This is about an order of magnitude more expensive in terms of computational efforts.

6.2. Results and Discussion

To point out clearly the importance of quantum mechanical effects in vibrational spectra influenced by proton dynamics, I consider that, recently, Masson, Williams and Rizzo published a series of very interesting IR spectra, where protonated glycine, $GlyH^+$, was tagged by an increasing and controlled number of hydrogen molecules.[126] I focus particularly on two of these investigations. The first one regards protonated glycine solvated by a single hydrogen molecule $(GlyH + H_2)^+$, while in the other instance three H_2 molecules are involved $(GlyH + 3H_2)^+$. The minimum geometries of these systems are reported in Fig.(6.1).

This figure suggests that panel a) is characterized by a strong hydrogen bond interaction between one of the amide hydrogens and the carbonylic oxygen atom of the carboxylic acid group, while in panel b) the presence of the three hydrogen molecules may suppress the hydrogen bond interaction by inducing a reorientation of the amide group. In panel a) the $H \cdots O$ distance at the minimum geometry is about 1.90 Å, while in panel b) the distance is equal to 2.52 Å. This last distance is still shorter than the sum of hydrogen and oxygen van der Walls radii (2.72 Å), which is considered, as a rule of thumb, the limit for hydrogen bonding.

To check out whether the hydrogen bond is lifted or not by virtue of the H_2 tagging process and if quantum mechanical effects play any relevant role for this kind of interaction, I will first perform *ab initio* "on-the-fly" DC-SCIVR simulations using the DFT-B3LYP level of theory and employing the aVDZ basis set, and then compare with the experimental spectra. Fig.(6.2) reports the experimental and simulated spectra for $(GlyH + H_2)^+$.

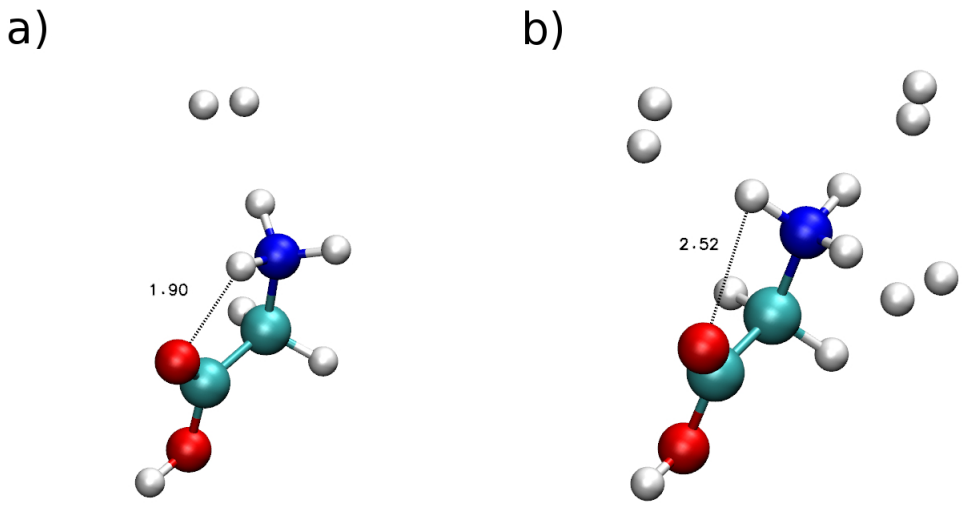

Figure 6.1.: Minimum configuration geometries obtained at B3LYP/aVDZ level of calculation. Panel a) for $(\mathrm{GlyH} + \mathrm{H_2})^+$ and b) for $(\mathrm{GlyH} + 3\mathrm{H_2})^+$. Distances are in Å.

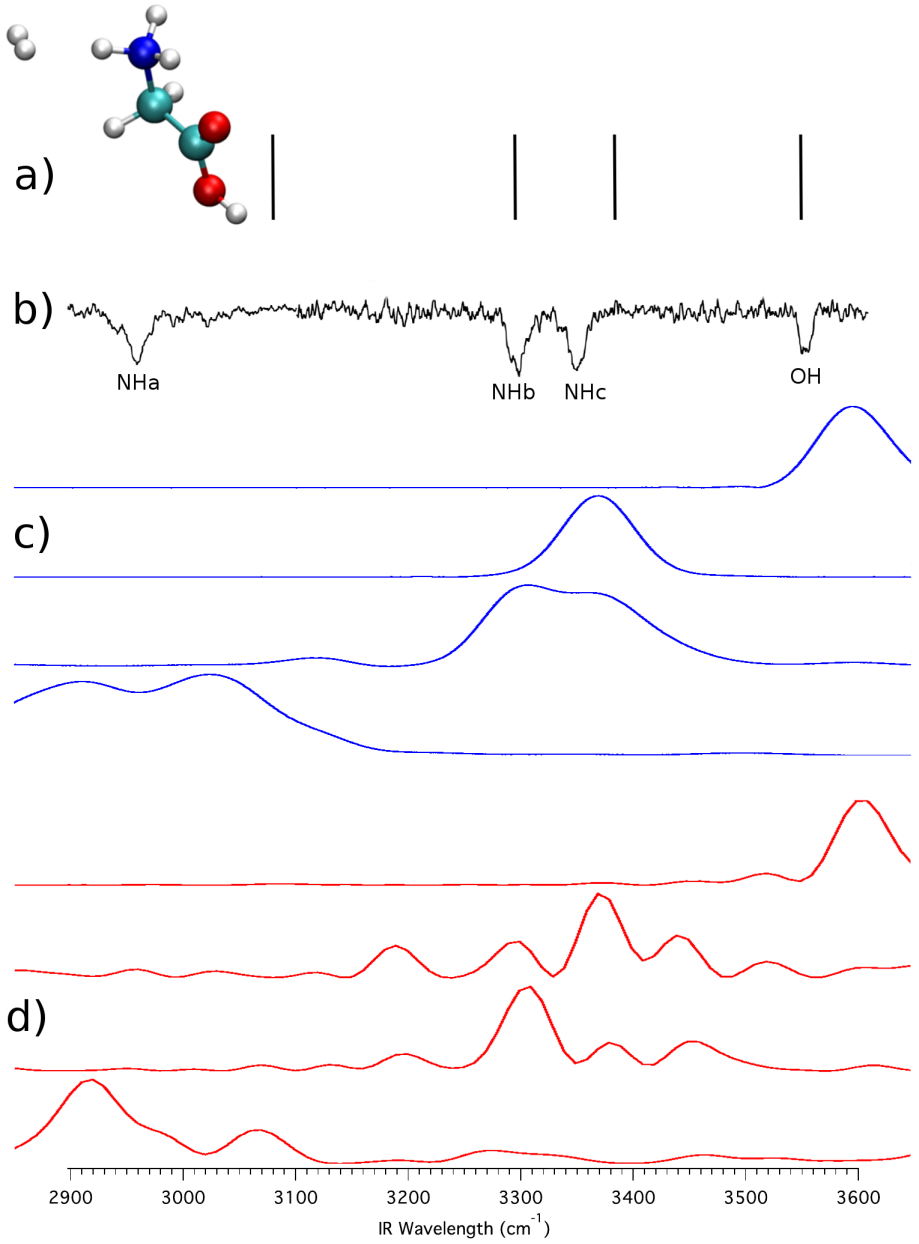

Figure 6.2.: $(GlyH + H_2)^+$ spectra. In a) the scaled-harmonic stick spectrum is presented; Panel b) reports the experimental spectrum;[126] c) is the I_{cl} classical spectrum according to Eq. (6.2), and d) is the I_{qm} semiclassical spectrum from Eq.(6.1). In the experimental spectrum, the label NHa is for the hydrogen bonded NH stretching frequencies, while NHb and NHc indicate the unbound ones. OH labels the homonymous stretching frequency.

The experimental results are reproduced in spectrum b), where the amide N-H stretch involved in the intramolecular hydrogen bond is located between 2950 and 3000 cm^{-1} and labeled as NHa, while NHb and NHc indicate the free NH stretching peaks. The signal at 3546 cm^{-1} corresponds to the free OH stretch. Upon adoption of a scaling coefficient equal to 0.96 to match the harmonic OH frequency (at MP2 level of theory with aVDZ basis set) with the experimental OH band, also the NHb and NHc peaks are reproduced quite well, while NHa is off by about 117 cm^{-1}, as shown by the stick spectrum on the top panel of Fig.(6.2).[126] Moving to classical simulations, I calculated the classical spectrum, I_{cl}, by means of Eq.(6.2) and report it in panel c) separately for each mode for a better comparison with the experiment. The main spectroscopic features are reproduced, even if the signal corresponding to the NHa and NHb bands is quite broad. Finally, in the semiclassical spectrum of panel d), calculated with semiclassical DC-SCIVR approach, the fundamental bands are faithfully reproduced, with the addition of combinations of bands and overtones that are too weak to be detected in the experiment and that are missing in the classical and scaled-harmonic simulations. In general, the simulated peaks are broader than the experimental ones because, on one hand, experiments are performed at very low temperature (a few K) and rotations are hindered or even blocked, while, on the other hand, in the simulations the dynamics is propagated only for a short time (less than 1 ps) before the Fourier transform is undertaken, and every mode (including internal rotators) is given an amount of energy according to its contribution to the ZPE and let free to evolve without any artificial constraints. Furthermore, the dynamics of the hydrogen bonding may contribute to the broadening of the NH stretching bands as shown by Gaigeot and coworkers by applying finite temperature classical molecular dynamics to small protonated peptides, such as Ala_2H^+ and Ala_3H^+.[67, 66, 69] I want to remark here that our calculations provide only the power spectrum of the system. The comparison with the experiment is therefore limited only to the peaks positions while no informations on intensity or width of the bands can be drawn. An accurate semiclassical method that includes peak intensities has been very recently developed and tested on small molecular systems.[137] It is based on a state-to-state approach which is however hard to employ for high dimensional systems with large density of states. To overcome this issue further research work is currently underway.

Passing now to the other system, i.e. $(GlyH + 3H_2)^+$, reported in panel b) of Fig.(6.1). Fig.(6.3) shows the spectra corresponding to those presented in Fig.(6.2) but this time for this bigger system. One may think that hydrogen molecules do not interact significantly with $GlyH^+$ and that it is possible to obtain in good

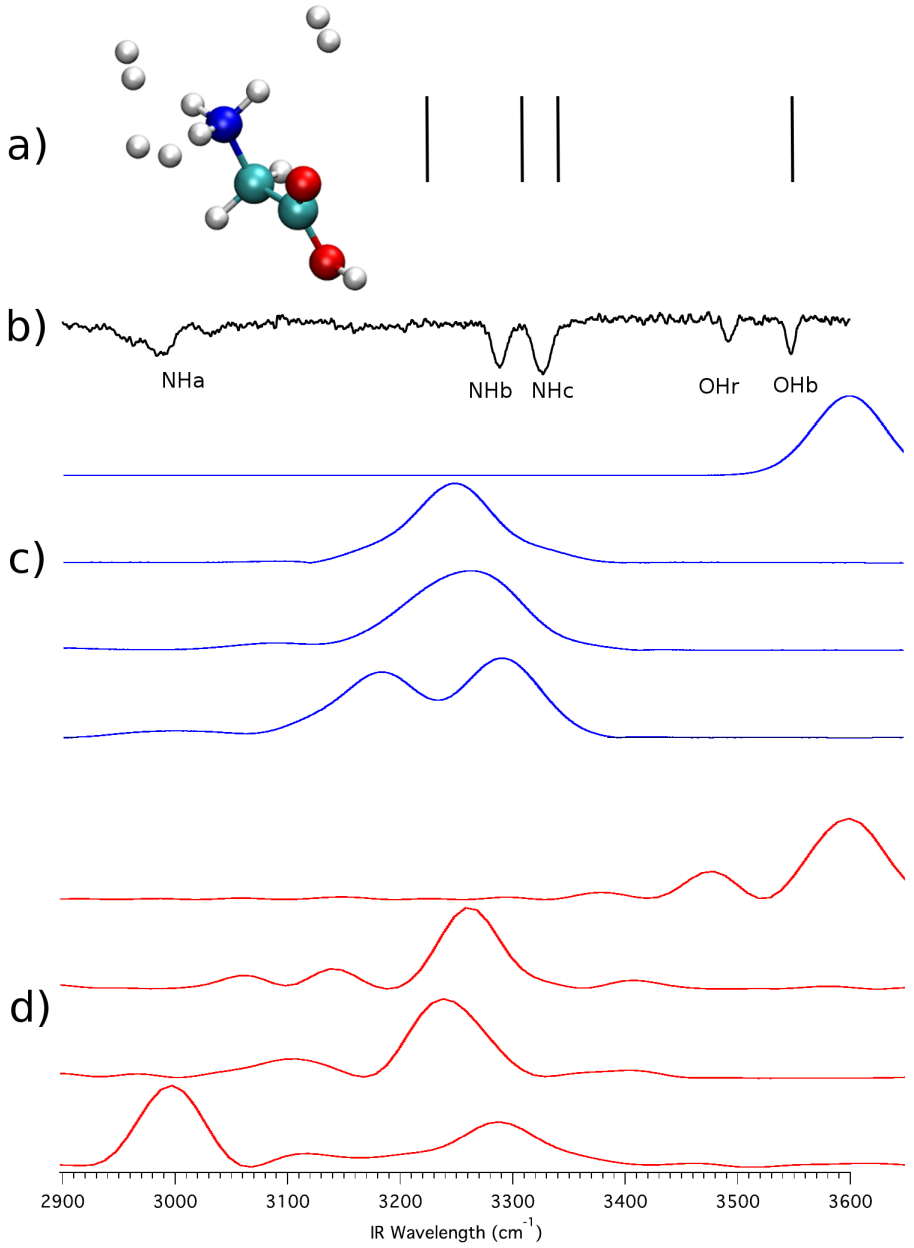

Figure 6.3.: The same as in Fig.(6.2) but for $(\mathrm{GlyH} + 3\mathrm{H_2})^+$.

approximation the IR signal of the isolated molecule. However, there are clear differences between the experimental spectra of Figures (6.2) and (6.3). One of them is represented by the blue shifted NHa peak. As usual, scaled-harmonic calculations are shown as a stick spectrum in panel (a) of Fig.(6.3) and they miss to account correctly for the anharmonicity of the NHa stretch motion. Once more, the scaling of harmonic frequencies brings us to a dead end. A classical approach based on Eq. (6.2) is not helpful in this circumstance as demonstrated by the set of spectra in panel (c) of Fig.(6.3) that clearly show classical mechanics overestimating the NHa stretch frequency. I believe that this is due to the fact that the intramolecular hydrogen bond and the dynamics of the involved proton have a prevalent quantum nature. In other words, a scaled-harmonic or classical dynamics approach lead to the wrong conclusion that the intramolecular hydrogen bond is broken in presence of $3H_2$ molecules interacting with $GlyH^+$. Conversely, a semiclassical simulation based on Eq.(6.1), reported in panel d), reproduces faithfully all the vibrational features of the experimental spectrum also in this case, including the strong anharmonicity of the NHa stretch and the consequent red shift, thus confirming that the hydrogen bond interaction is only weakened and not completely broken, even in presence of three H_2 molecules coordinated to the amide group.

Another key difference between Figures (6.2) and (6.3) lies on the appearance of a second OH stretch band, located at 3491 cm^{-1}, and labeled as OHr. Masson et al.[126] suggested that this band is given by a configuration where one of the three H_2 molecules interacts with the carboxylic group. Indeed, the peak is red-shifted by about 55 cm^{-1} with respect to the free OH stretch (the OHb band). This would mean that the experimental spectrum b) of Fig.(6.3) is actually originated by two different conformers. To validate the previous conformational hypothesis I consider a configuration with a single H_2 molecule tagging the carboxylic group. This geometry is not stable experimentally (in fact the OHr peak appears only when 3 or more H_2 molecules are involved) but it can be investigated theoretically. The system is reported in panel (a) of Fig.(6.4), and I focus on the OH stretch.

Still in panel a) the harmonic stick spectrum (at DFT-B3LYP level of theory with aVDZ basis set) for the OH stretch is presented after scaling by a factor of 0.96, which is the same coefficient employed in the previous simulations. This estimate is definitely off the mark. On the contrary, both the classical (panel c) and the semiclassical (panel d) peaks are quite accurate for the OH stretch, confirming that the OHr band is indeed due to the interaction between a H_2 molecule and the carboxylic group. The presence of the H_2 molecule weakens the OH bond leading to the observed red shift.

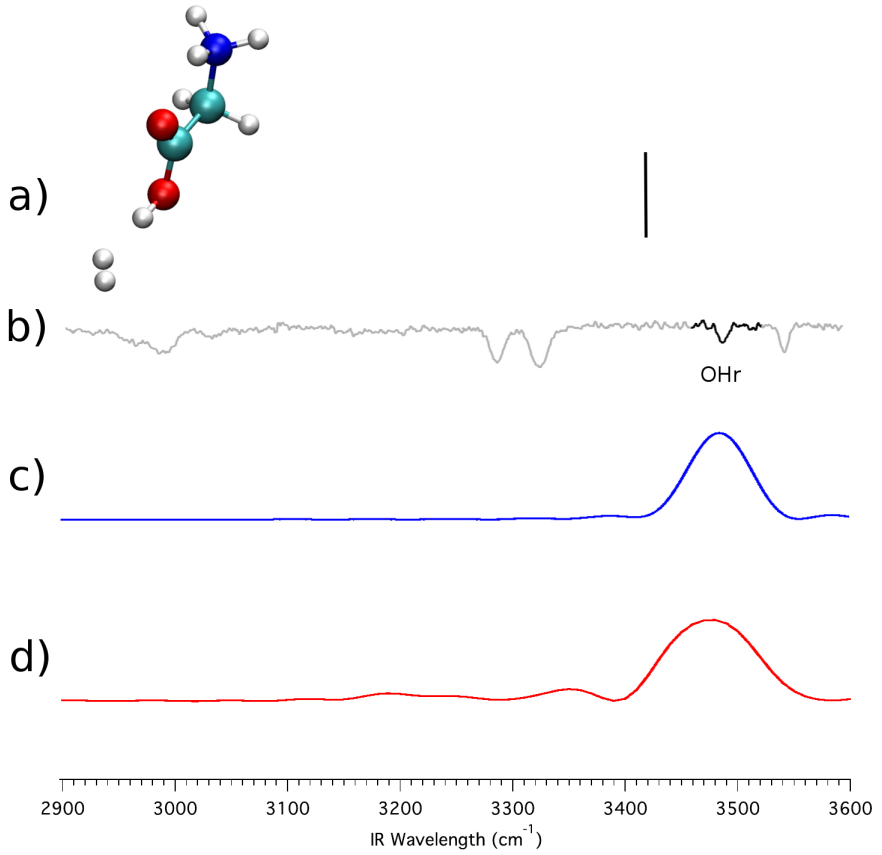

Figure 6.4.: $(GlyH + H_2)^+$ spectra, with the H_2 molecule coordinated to the carboxylic group. The labels (a)-(d) are as in Fig.(6.2).

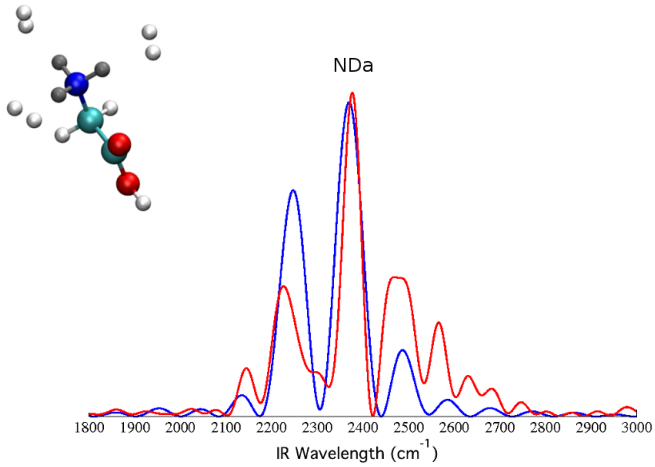

Figure 6.5.: Selectively deuterated $(\mathrm{GlyH} + 3\mathrm{H_2})^+$ system and the corresponding classical (blue continuous line) and semiclassical (red continuous line) spectra for the amide stretching mode. The deuterium atoms are colored in gray.

To further prove that the differences between the classical and the semiclassical spectra reported respectively in panel c) and d) of Fig.(6.3) are due to quantum mechanical effects only, I quenched them by deuterating all the three hydrogen atoms, pictorially represented by the gray atoms of the molecule in Fig.(6.5).

Then, I calculated the spectra for the deuterated $\mathrm{GlyH^+}$molecule tagged by the three H_2 molecules. I focus on the amide modes and selectively plot the NDa (previously NHa) band, both using the classical Eq.(6.2) and the quantum formulation of Eqs. 6.1. The results are reported in Fig.(6.5). The NDa band is centered around $2350\,\mathrm{cm}^{-1}$, which is significantly red-shifted with respect to the previous NHa band, because of the heavier deuterium mass. However, I note that the classical and semiclassical peaks are almost identical in this case. This proves that the previous discrepancy of about $150\,\mathrm{cm}^{-1}$ between the classical and the semiclassical NHa band location of Fig.(6.3) was exquisitely due to a quantum mechanical effect of the light hydrogen atom. It is quite surprising that this quantum anharmonic effect is so huge. However, when considering the strong anharmonicity of the NHa potential well and the consequent huge delocalization of the quantum mechanical vibrational eigenfunction, as pictorially represented Fig. (6.6), the prominent quantum mechanical nature of this hydrogen bond interaction appears fully justified. Furthermore in Table 6.1all the frequencies values calculated with the DC-SCIVR for all the four systems are reported

Table 6.1.: Semiclassical (DC-SCIVR) vibrational frequencies of tagged glycine systems reported in cm^{-1}. Values are compared with their harmonic (HO) and classical counterparts.

$(GlyH+H_2)^+$ (NH tagged)				$(GlyH+H_2)^+$ (OH tagged)			
Mode	DC-SCIVR	HO	Classical	Mode	DC-SCIVR	HO	Classical
NHa	2920	3139	3030	NHa	2950	3095	3035
NHb	3310	3377	3310	NHb	3340	3446	3340
NHc	3370	3485	3370	NHc	3370	3507	3350
OHb	3610	3696	3600	OHr	3480	3561	3485
$(GlyH+3H_2)^+$				(Deuterated $GlyH+3H_2)^+$			
Mode	DC-SCIVR	HO	Classical	Mode	DC-SCIVR	HO	Classical
NHa	3000	3312	3185	NDa	2376	2517	2370
NHb	3240	3367	3260	NDb		3099	
NHc	3260	3410	3250	NDc		3159	
OHb	3600	3701	3600	OHb		3701	

together with harmonic and classical estimates.

6.3. Conclusions

I conclude this chapter by remarking the importance to employ a quantum dynamical approach in calculating vibrational frequencies and to go beyond the scaled-harmonic level. This has been demonstrated by means of divide-and-conquer semiclassical dynamics, which has permitted to reproduce experimental anharmonicities quite well and to explain an open issues involving protonated glycine tagged with molecular hydrogen. In particular a peculiar spectral quantum feature due to hydrogen bonding and intermolecular interactions has been rigorously explained, a task that neither scaled-harmonic nor classical approaches were able to accomplish. On this point, I notice that the reference experiments were performed at very low temperatures, so I did not run standard thermalized classical simulations (which would have provided just harmonic estimates), but I estimated a classical analog of the quantum mechanical vibrational spectral density. Nevertheless, these classically-inspired calculations were not as satisfactory as semiclassical ones. Interestingly, by comparing the frequency value of the hydrogen stretching itself and when implied in glycine tagged systems, a red shift comparable to that of the OH stretch of glycine (~ 50 cm^{-1}) is displayed.

Remarkably, the DFT-B3LYP level of electronic theory adopted is not only suit-

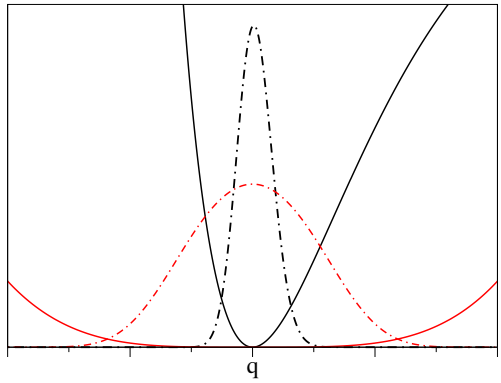

Figure 6.6.: Continuous black line and red lines show potential energy profiles of a stiff and a mild oscillators, while black and red dashed lines their respective ground state wavefunctions.

able for a realistic description of the entire supramolecular system but also able to provide frequency estimates in quantitative agreement with the experiments. Finally, I am able to answer the questions with which I introduced the chapter by stating that quantum effects play certainly a very important role in these protonated systems, the intramolecular hydrogen bond interaction has a strong impact on the NH stretch revealing an elevated degree of delocalization of the proton shared with the carboxylic group, and the very same interaction is influenced by the dynamics with the hydrogen bond being less and less directional as the number of tagging molecules increases. All these findings point out very clearly the crucial role that quantum dynamics may have, suggesting that it should not be neglected even when dealing with larger systems.

7. On the Zwitterionic Glycine Hydration mechanism

In previous chapters I applied the semiclassical method to study glycine based systems in gas phase, both bare and interacting with other small molecules. More than gas phase investigations, it is the study of the structures and properties of aminoacids in water solvent which is important, since that is their physiological environment. In this context, a lot of work arguing the real hydration mechanism of glycine in both its neutral and zwitterionic form has been published. A recent work by Perez de Tudela and Marx suggests a possible hydration mechanism of such molecular species. Moreover, in the same paper, they highlight some differences in the predicted IR vibrational spectra depending on the number of water molecules interacting with glycine. To obtain these results they used the harmonic approach employing an high ab-initio level of calculation. As a natural prosecution on supramolecular glycine based systems I study in this chapter the same set of molecular conformations described in Marx's work, aiming at both extending application of the semiclassical DC-SCIVR approach to quasi-solvated systems and confirming the spectroscopic findings employing an anharmonic, dynamics based theoretical method.

7.1. Introduction

Aminoacids are continuously under investigation, being the building block of biologically relevant molecular systems such as proteins. Even if their intrinsic gas-phase properties are important to be investigated, their principal role and functionality is played in physiological conditions, i.e. room temperature, ambient pressure and water-like solutions. In such an environment, their structures and behavior can be dramatically different from the gas-phase predictions. Stately the two terminal functional groups, carboxyl and amino, display the neutral form (COOH and NH_2 respectively) when the aminoacids are in vacuum, while when they are solvated in water the carboxyl is deprotonated as COO^- and the amino group is conversely protonated as $NH_3{}^+$. The importance of understanding the hydration and stabilization of aminoacids in water arises as a direct consequence.

The glycine molecule in its neutral and zwitterionic forms, being the smallest aminoacid, has obviously attracted the attention of the theoretical community. In fact, having only 10 atoms, this molecule permits the study of its hydration through the application of classical mechanical simulations as well as ab-initio calculations. If from the theoretical point of view a general consensus has been reached to establish toward eight or nine the number of water molecules necessary to stabilize the zwitterion, [138, 139, 140, 141, 142, 143, 144, 145, 146, 147] from the experimental side the debate is still open and such a number seems to assess itself around three or five. [148, 149, 121]

In a similar fashion, also the real mechanism through which the zwitterion is stabilized by the presence of a different number of water molecules is not clarified. What is broadly established is that a bridge constituted by water molecules interconnects the carboxyl and amino group, promoting a proton transfer that follows the well-known Grotthuss mechanism.[150, 146, 151, 152] The importance of such interconnecting water bridge goes beyond aminoacid stabilization. In fact it plays an important role also in the context of microsolvation and phototriggered intramolecular proton transfer in bulk solutions. [153, 154, 155, 156]

In this chapter I will refer to a recent work by Perez de Tudela and Dominik Marx that investigated the number of water molecules required for the stabilization of the zwitterionic glycine over the neutral form, trying to understand which mechanism and placement of the water molecules drives the change. [157] In particular, starting from a first regime, from one to four water molecules, in which the neutral form is clearly energetically favorite, they described two other regimes, one from five to eight water molecules, where the two different conformations are almost isoenergetic and finally the third and last regime, from nine water molecules upwards, where the zwitterionic glycine becomes the most stable form. The energetic trend calculated in that work can be visualized in Fig 7.1.

The physical reason for this stabilization is attributed by the authors to a bifurcated hydrogen-bonds wire that interconnects the two charged groups of the molecule. Such a double wire can be composed only starting from nine water molecules, while in smaller clusters there is a single molecular bridge that connects glycine extremities. In the same publication, they provided a spectroscopic insight to detect a clue of this stabilization mechanism by means of IR spectroscopy. In fact, through a MP2 harmonic frequencies analysis, they found that the lowest NH_3^+ stretching frequency stands almost isolated in the energy window between 2400 and 2800 cm^{-1} for all the zwitterionic glycines solvated with two up to eight water molecules, while, on the contrary, in the systems with nine and ten water molecules the same band shifts toward 3000

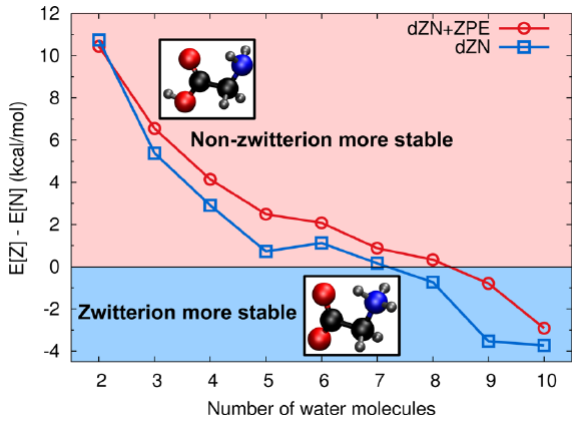

Figure 7.1.: Energy difference between the lowest-energy non-dissociated (N) and zwitterionic (Z) conformers of Gly·(H_2O) n clusters as a function of the number n of water molecules. Taken by ref. [157]

cm^{-1}. All the estimates provided are unscaled, hence these values are obviously to be considered in a qualitative way, but they are still able to give an important indication.

For this work I apply the semiclassical method to some of the structures found by Marx, with the aim of reproducing the qualitative pattern with a theoretical dynamically based approach. In particular I used the DC-SCIVR approach, described in details in the theoretical section of this thesis, which formulation is reported in equation 1.50 in the chapter 1. I here remind that semiclassical theory accounts for construction, not only for the anharmonicity of the system, but also for all the quantum effects, like zero point energy (ZPE), tunneling effects and couplings from overtones and fundamental frequencies. To underline such a resource, I calculate vibrational spectra also in a classical fashion starting from ab-initio calculated trajectories, by the use of the Fourier transform of the velocity autocorrelation function, as reported in equation 2.8 in chapter 2 and by the comparison of these two different approaches it can possible to appreciate the contribute of quantistic effects in such solvated systems. Both this methodologies are dynamical based and can certainly enrich the harmonic analysis done by Marx. In fact, as it will be reported in the results section, I confirm the spectroscopic trend detected by the Marx's harmonic study and I describe further insights on the stability of the zwitterionic glycine according to the number of water molecules, deriving from ab initio molecular dynamics.

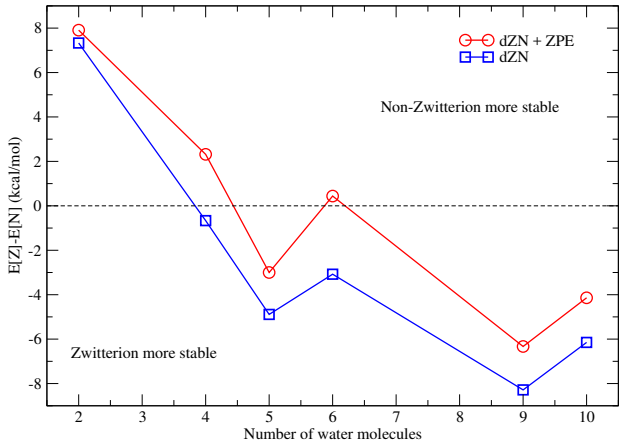

Figure 7.2.: Energy difference between the lowest-energy non-dissociated (N) and zwitterionic (Z) conformers of Gly·(H_2O)n clusters as a function of the number n of water molecules, calculated with DFT-B3LYP-D / 6-311g* level of theory.

7.2. Methods

All the optimizations, calculations of hessian matrices and ab-initio molecular dynamics, have been performed with the B3LYP-D DFT functional with 6-311G* basis set, as implemented in the NWChem software package. The energetic landscape obtained at this level of theory can be displayed in Figure 7.2.

Looking at Figure 7.2, the same qualitative trend shown in Figure 7.1 obtained with MP2 calculations, can be here recognized, even if the agreement is not quantitative. Therefore, for this preliminary study, I chose to adopt this kind of functional, while in other quantitative analysis is desirable to use a higher level of ab-initio calculation. All the trajectories have been evolved for 2500 steps of 10 a.u., (0.6 ps of total time). The dynamics initial conditions are chosen to be the equilibrium coordinates for the positions, while initial momenta have been set both giving the harmonic ZPE to each normal mode and assigning an additional quantum of energy to a selected normal mode, which represents a refinement procedure. All the classical spectrum reported in this chapter has been evaluated starting form the same trajectory run for the semiclassical analysis, in order to have a reliable and reasonable comparison.

Since semiclassical simulations require the evaluation of the Hessian matrix along the dynamics in addition to the potential energy and the gradient, in order to limit the computational work required by a complete analysis, I chose to

mode	Harmonic	DC-SCIVR	FT-vv
aNH2	3565	3440	3460
sNH2	3498	3450	3440
aCH2	3083	2990	2990
sCH2	3051	2960	2960
OH	2973	2830	3070

Table 7.1.: Harmonic, classical (FT-vv) and semiclassical frequencies values for Gly·2(H_2O) in its non-dissociated form.

analyze only one system per regime. Therefore vibrational spectra of glycine in its neutral and zwitterionic form interacting only with two, five and nine water molecules have been calculated.

In order to alleviate the computational effort, I used two different strategies to avoid the ab initio Hessian calculation at each trajectory step. The first one is the so-called "Hessian update", in which an intermediate Hessian is evaluated using the gradient in that point and the last ab initio calculated Hessian. The other method uses a database of a reduced number of ab-initio evaluated Hessians to approximate other Hessians at each point, see Appendix 1.d for more details. Hessian update has been used for glycine interacting with two and five molecules, while for the systems that includes glycine and nine water molecules the Hessian database has been employed.

7.3. Results

This section has been subdivided accordingly to the three regimes previously mentioned. As already stated I chose glycine interacting with two water molecules as a representative of the first regime, five water molecules for the second one, and nine for the third one.

Systems with two molecules see the neutral glycine as the energetically favorite structure. A signal of this stability has been obtained from the dynamical study of the zwitterionic form. In fact, after a short period of time (about 0.2 ps), the molecule interconverts into its neutral form. For this reason the vibrational spectrum of this system was not evaluated. However, in Table 7.1 and in Figure 7.3 the semiclassical and classical frequencies of the high-frequency region of the spectrum, calculated for non-dissociated neutral glycine, are listed and displayed.

All the shown modes present a non-negligible amount of anharmonicities. Classical fundamental frequencies are in agreement with semiclassical estimates,

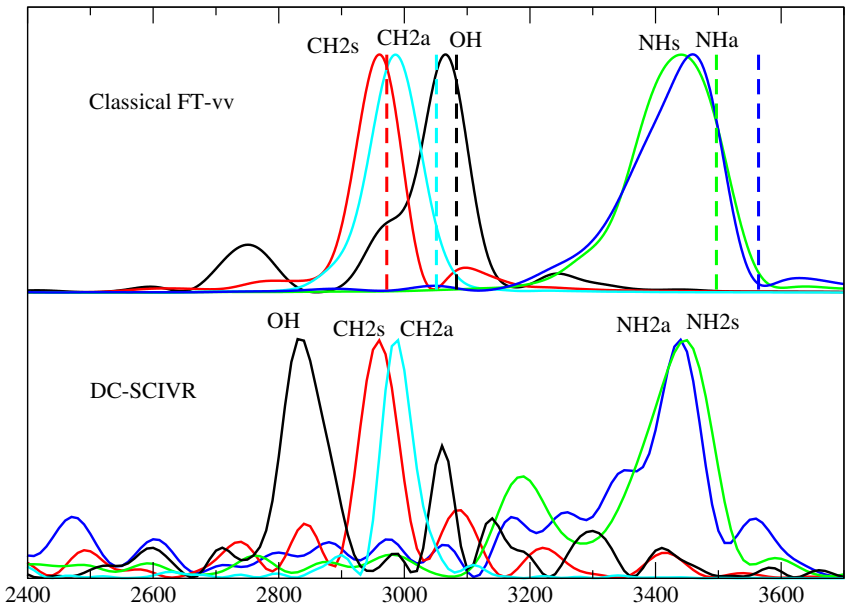

Figure 7.3.: Vibrational spectra of the non-dissociated Gly·2(H_2O). In the top panel the classical spectra together with the harmonic estimates (dashed lines) is reported. In the panel below there is the semiclassical spectral density.

mode	Harmonic	DC-SCIVR	FT-vv
aNH2	3600	3450	3450
sNH2	3518	3320	3340
aCH2	3112	2990	2990
sCH2	2985	2870	2850
OH	2959	3050	3030

mode	Harmonic	DC-SCIVR	FT-vv
NH	3536	3290	3180
aCH2	3161	3100	3100
sCH2	3091	3040	2870
sNH2	2825	2840	2660
aNH2	2756	2510	2630

Table 7.2.: Harmonic, classical (FT-vv) and semiclassical frequencies values for Gly·5(H_2O) in its non-dissociated form (on the left) and in the zwitterionic configuration (on the right)

except for the OH stretching mode, where the semiclassical peak appears largely red shifted.

Moving to the second regime, where I considered only the clusters with five water molecules, I found that both glycine forms are dynamically stable, during the employed evolving time (0.6 ps), when in the initial condition the momenta of each normal mode is equal to its harmonic ZPE contribution. Instead, the interconvertion from zwitterionic to non-dissociated form has been detected in other trajectories, in which the momenta initial conditions have an additional quantum of excitation. This change of conformation happens here with the help of water molecules (Grotthus mechanism). Giving quanta of excitation is a standard procedure in the semiclassical approach, used to accurately detect the desired frequency value, but represents here another signal of unstability of the zwitterion over non-dissociated form, as declared by Marx's analysis. In Table 7.2 all the computed frequencies are listed, while in Figure 7.4 the corresponding vibrational spectra are shown.

As already described above, these systems present a high amount of anharmonicities. In the non-dissociated glycine system, such deviation from the harmonic estimates can be seen by the classical spectra as much as the semiclassical ones. The same scenario is not replicated for the zwitterionic analysis. The NH_2 stretching bands, for example, are close in the classical spectra, while in the semiclassical simulations they appear splitted in two well-spaced bands. This can be important since one of this stretching is the peak described by Marx to identify the definitive stabilization of the zwitterion over the non-dissociated form.

Finally, in Table 7.3 and Figure 7.5 the results of the simulations of the nine glycine - water clusters are reported.

In this third regime, no interconversion has been observed, not even when an additional quantum of energy has been initially assigned. The same set of

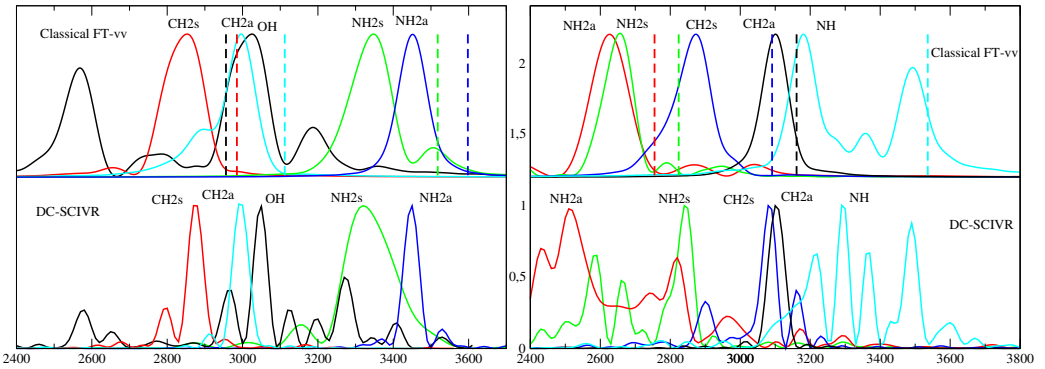

Figure 7.4.: Vibrational spectra of the non-dissociated Gly·5(H_2O) (on the left) and of the zwitterionic Gly·5(H_2O) (on the right). In the top panels the classical spectra together with the harmonic estimates (dashed lines) are reported. In the panel below there are the semiclassical spectral densities.

mode	Harmonic	DC-SCIVR	FT-vv
aNH2	3590	3510	3510
sNH2	3510	3395	3410
aCH2	3110	3100	3020
sCH2	2990	3000	3010
OH	2333	2610	2540

mode	Harmonic	DC-SCIVR	FT-vv
NH	3488	3430	3430
aCH2	3155	3030	3030
sCH2	3093	3040	3040
sNH2	3020	3020	2970
aNH2	2993	2870	2880

Table 7.3.: Harmonic, classical (FT-vv) and semiclassical frequencies values for Gly·9(H_2O) in its non-dissociated form (on the left) and in the zwitterionic configuration (on the right)

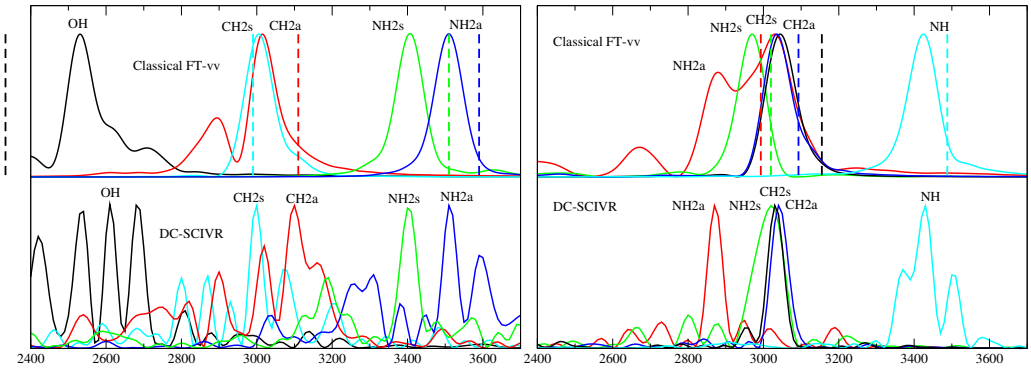

Figure 7.5.: Vibrational spectra of the non-dissociated Gly·9(H_2O) (on the left) and of the zwitterionic Gly·9(H_2O) (on the right). In the top panels the classical spectra together with the harmonic estimates (dashed lines) are reported. In the panel below there are the semiclassical spectral densities.

considerations presented for the cluster with five water molecules about the spectral density is appropriate also for systems with nine. In fact, for non-dissociated glycine, semiclassical and classical results are comparable, while for the zwitterionic form there are some discrepancies, especially in the detection of the NH_2 stretching band. This is noticeable looking at the peaks shape in the spectra, more than from the frequencies values. More importantly, the lowest NH2 stretching band in the zwitterion spectrum blue shifts toward 2870-2880 cm^{-1} here, moving from the 2510-2630 cm^{-1} region of the five water molecules system, confirming the same feature highlighted by Marx et al. as a proof of its stabilization. In Figure 7.6 this agreement is pictorially represented.

7.4. Conclusions

In this chapter semiclassical DC-SCIVR has been applied to supramolecular systems composed by the glycine in its neutral and zwitterionic conformations interacting with an increasing number of water molecules. The aim was to reproduce qualitatively the vibrational spectrum of such systems calculated with a harmonic approach by Marx and coworkers, implicitly demonstrating the validity of the semiclassical method in the spectrosopic field. Furthermore, a classical estimate of the spectrum for each system has been performed by means of the velocity autocorrelation function. The work done so far highlights some differences in the spectrum evaluated with a semiclassical approach or from a

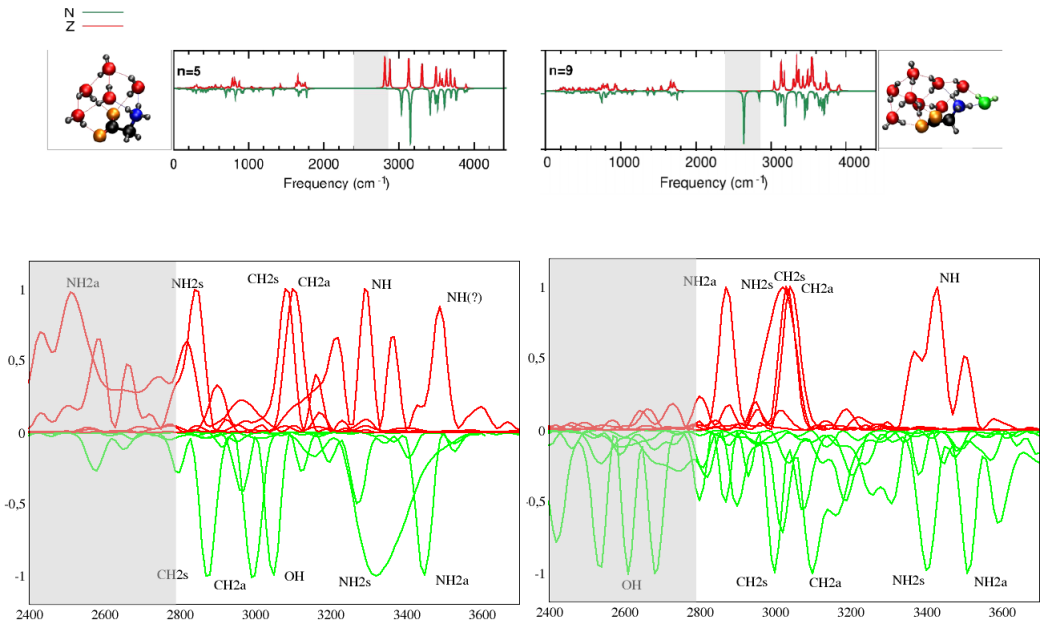

Figure 7.6.: Comparison between harmonic MP2 vibrational spectra of the Gly·(H_2O)n clusters (ref [157]) and semiclassical DC-SCIVR. On the left side the spectra of Gly·(5H_2O) are reported, and on the right spectra those of Gly·(9H_2O). In green the non-dissociated spectra are shown, while in red zwitterionic one is presented. The gray band highlights the spectral region where the the NH stretching appears only when the neutral glycine is the most stable form.

classical perspective. This observation points out the fact that quantum effects can affect the vibrational spectrum, playing a determinant role in this kind of systems, where hydrogen bond wires dominate the interaction picture. The dynamics itself has demonstrated its importance. For example, we were able to see that the zwitterionic form of glycine interconverts in its non-dissociated form when interacting with two or five molecules of water. This is an important issue, because it is a signal that the existence of a minimum geometry on the ab-initio potential energy surface is not sufficient to assess the effective stability of a structure.

8. Deoxyguanosine and Ac-Phe-Met-NH_2 vibrational spectra: a spectroscopical comparison between ab-initio and force field molecular dynamics

As described in the theoretical section, the semiclassical method is usually associated to ab-initio molecular dynamic (AIMD), especially when the system size does not allow for a construction of a pre-fitted PES. This combination showed very good accuracy [41, 26], but unfortunately it presents high computational costs. In contrast the Force Fields (FF) family of the molecular mechanics, described in chapter 2, is noticeably less demanding than ab initio approaches because, instead of solving the complex Schröedinger equation, they only have to solve an analytical formula to evaluate the energy of the system. In this chapter I discuss the accuracy of the Amber94 force field against the performance of DFT-based AIMD, presenting the vibrational power spectra of two biological systems, Deoxyguanosine and the Ac-Phe-Met-NH_2 dipeptide, in their very interesting high frequency region, by employing different approaches (i.e. at harmonic, semiclassical and classical). I will show that the semiclassical method associated to the FF potential gives the worst results, while better estimates are obtainable using FFs via harmonic frequency calculations. In particular, Amber94 is accurate for the normal modes associated to simple molecular motions, while it is poor for more complex normal modes. Conversely, AIMD always leads to accurate results. In light of these findings, I conclude that the Amber94 force field should be revised to permit a reliable semiclassical vibrational analysis, at least for in vacuum simulations.

8.1. Introduction

Throughout all the applications described in previous chapters I adopt the semiclassical method associated to AIMD, that evaluates the potential energy step by step using an ab-initio method that is secondarily employed also for Hessians computation. Despite all the existing strategies to alleviate the computational effort, see for example Appendix 1.d, ab-initio calculations remain a relevant time consuming factor and, most importantly, they become more and more hampering as the system dimensionality increases. Within this context, I wonder if a more efficient method for calculating trajectories and Hessians is viable. For instance, among all the available methods, exists the classical molecular dynamics implemented through force fields where the potential energy has an analytical formulation, as describe in the chapter 2. For this reason it is computationally relatively inexpensive and it is commonly used to tackle huge biological systems, like a solvated protein, nanotubes and DNA fragments.

Some works on the comparison between empirical and ab-initio force field were already published in literature. As an example, Gerber et al. already stated that classical force fields like Amber and OPLS performed worse than MP2, especially for the normal modes implied in anharmonic coupling.[158] Nevertheless, the possibility to exploit force fields for vibrational spectroscopy purposes is still an open question in the semiclassical community.

In this chapter I show the results coming from the application of Amber94, a commonly employed force field, for IR spectrum calculation of deoxyguanosine and Ac-Phe-Met-NH_2. The results obtained in this way will be compared with the same set of simulations coming from AIMD, performed with the DFT B3LYP functional. In particular, I adopted three different strategies to address the problem. First I evaluated the harmonic frequencies by diagonalizing the Hessian matrix at the minimum geometry. Then, I performed a classical estimate based on the Fourier Transform of the velocity autocorrelation function preceeded, whenever possible, by a thermalization buffer. Finally, I employed the recent semiclassical DC-SCIVR approach.

Both systems chosen for this work are fully parametrized in Amber94. This permits a triangular comparison between Amber, DFT based AIMD and experimental data. Moreover the two molecular systems investigated represent small building blocks of biological relevant systems: deoxyguanosine is one of the components of DNA; Ac-Phe-Met-NH_2 is a prototypical dipeptide for generic investigations on proteins. More precisely, deoxyguanosine is a nucleoside, a system composed by a nucleobase (guanine) and the sugar moiety (deoxyribose). Like all nucleobases, it has been lergely studied in the past, in particular to shed

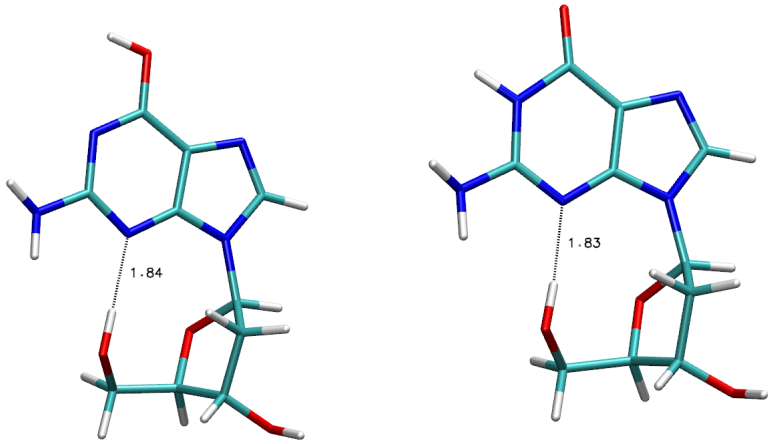

Figure 8.1.: Minimum structure of deoxyguanosine in its enolic (left) and ketonic form (right) evaluated at B3LYP/6-31G* level of theory.

light on the different properties of its tautomeric forms, implied, in particular, in the phenomenon known as mutagenesis. In Fig. 8.1 the minimum structures of the enolic and ketonic forms of deoxyguanosine are reported.

The vibrational spectrum of deoxyguanosine has been recently investigated for the isolated molecule and for the species interacting with water molecules, and some experimental features can be deduced by the spectra of all guanine conformers.[159] Another interesting system is represented by the dipeptide composed by L-Phenilalanine and L-Methionine. This biological molecule is important as a prototype system in which several hydrogen bonds govern the construction of the secondary and tertiary structure. In particular, there are three different H-bonds that play a key role in the stabilization of the folded structure: two occur internally at the lateral chain of the aminoacids, a NH---π in the Phe and a NH---S in the Met, while the last one is established between NH and C=O of the two terminal groups. (Fig 8.2) In a recent work Mons et al. studied conformation and vibrational spectrum of this dipeptide capped with acetyl and amide (Ac-Phe-Met-NH_2). Their investigation presented quantum DFT-D calculations together with experimental results in gas-phase, showing that the hydrogen bond that involves sulfur atom has a strength similar to the more classical intrabackbone hydrogen bond NH---O=C.[160]

What I found in this analysis is that the Amber results are accurate only

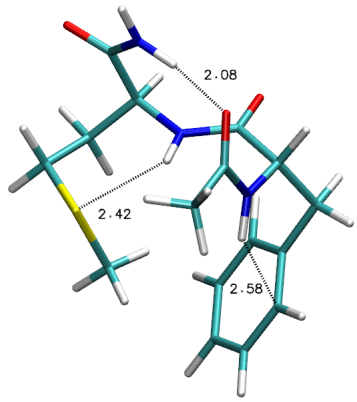

Figure 8.2.: Minimum structure of Acetyl-Phe-Met-NH_2 evaluated at B3LYP/6-31G* level of theory.

for specific normal modes, while for the majority of the frequencies values are definitely off the mark, whatever method is employed. On the other side, AIMD confirms its reliability, giving in all cases accurate results.

8.2. Method

All the ab-initio calculations have been performed by means of the program suite NWChem employing the DFT B3LYP functional in conjunction with the 6-31G* basis set for the deoxyguanosine, while the B3LYP-D with the same basis set has been employed for the dipeptide. Instead, for the Amber94 simulations, I used the Gromacs software package in its double precision installation version. In this latter case, all the classical molecular dynamics simulations were performed in vacuum, using the special keyword "Morse" to model the bond term in the potential energy formula via the Morse potential function. In particular, the optimization phase was performed with the L-BGFS algorithm [161, 162] (as indicated by the Gromacs manual) to have the calculation of the Hessian matrix accurate and reliable. The NVT phase was performed for 100 ps with 2 fs timestep, using a modified Berendsen thermostat implemented within Gromacs. The product phase was instead performed in the NVE ensemble using the velocity-Verlet integrator for trajectories 1 ps long for semiclassical analysis and

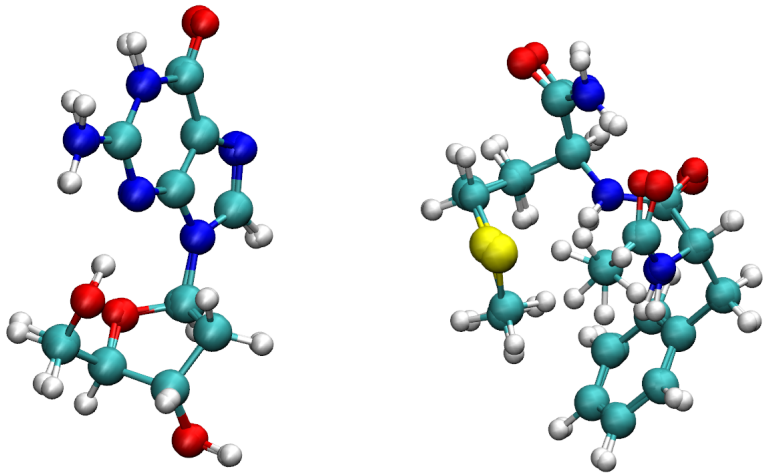

Figure 8.3.: Comparison between B3LYP and Amber geometry optimization results. Left panel: deoxyguanosine in its ketonic form. Right panel: Ace-Phe-Met-NH_2 dipeptide.

20 ps long for classical analysis.

The ab-initio trajectories were evolved for 0.6 ps, setting the kinetic energy due to the initial momentum of each mode to be equal to the harmonic ZPE of that mode. For some peaks in the spectrum a refined procedure was applied, assigning an additional quantum of energy to the associated specific normal mode. For vibrational spectra calculations I employed different methods. Together with the harmonic estimate of the frequencies, obtained through the diagonalization of the Hessian matrix at the equilibrium geometry, I used the DC-SCIVR method and the classical approach that exploits the velocity autocorrelation function, according to the equation 2.8 by using always the same set of trajectories.

8.3. Results

Minimization performed using the Amber94 force field leads to very similar structures to those found with the DFT B3LYP functional, for both systems. A pictorial view of this agreement can be visualized in Fig 8.3 where the structures coming from ab-initio DFT and Amber force field have been superimposed.

The RMSD calculated on the structure pairs is 0.3179 Å and 0.147 Å respectively for Ace-Phe-Met-NH_2 and the ketonic deoxyguanosine. Furthermore, with the

mode	Exp.	DC-SCIVR	FT-vv	$\Delta_{DC-SCIVR}$	Δ_{FT-vv}
3OH	3662	3650	3652	12	10
enol-OH	3588	3560	3548	28	40
aNH2	3577	3560	3577	17	0
sNH2	3458	3470	3487	12	29
5OH	3230	3270	3268	40	38

Table 8.1.: AIMD Semiclassical DC-SCIVR, classical from velocity-velocity correlation function (FT-vv) and experimental frequencies of enolic deoxyguanosine. The column $\Delta_{DC-SCIVR}$ indicates the absolute difference between experiment and semiclassical values, while Δ_{FT-vv} between experiment and classical values. The experimental frequencies come from ref. [159].

exclusion of the three hydrogen atoms of the methyl groups in the dipeptide, that are quite free to rotate and have a minor influence on the overall geometry similarity, I obtain an RMSD equal to 0.2225 Å. Despite the small size of the molecules under investigation, the RMSD values here obtained can be considered a "good one" to assess a strong similarity between the structures.

For this reason I can state that the Amber optimization leads to reliable equilibrium geometries for both systems.

8.3.1. Deoxyguanosine

The deoxyguanosine molecule exists in both enolic and ketonic forms, as illustrated in Figure 8.1. Experimentally only the enolic tautomer can be investigated but, unfortunately, the parameters in the Amber94 force field are tailored to the ketonic form of deoxyguanosine. For this reason, I studied both the tautomers applying the ab-initio class of methods, while I have been forced to limit the investigation to the ketonic form in the case of Amber94 analysis. I took the Saigusa experiment as a reference for the enolic form of deoxyguanosine,[159] while for the ketonic conformer I compared the corresponding modes of the ketonic form of guanine, taken from the experiment of Choi and Miller.[163] In doing this the insight by Nir et al. that highlighted the similarity of enolic deoxyguanosine and enolic guanine vibrational spectra has been exploited.[164]

In Table 8.1 and Fig. 8.4 the DC-SCIVR and classical frequencies of the high frequency region of the enolic form of deoxyguanosine evaluated through AIMD, together with the experimental results are reported for a semiclassic time evolution of 25.000 a.u. (0.6 ps).

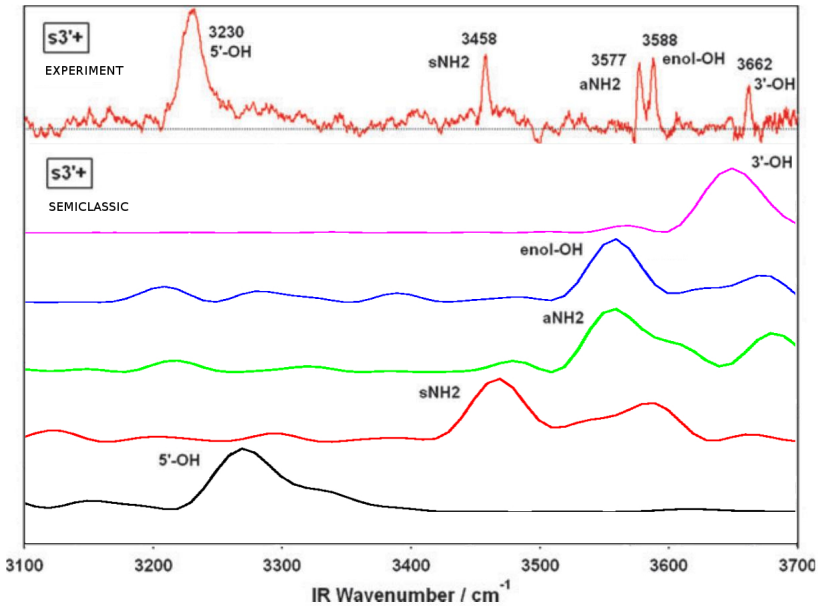

Figure 8.4.: AIMD Semiclassical DC-SCIVR and experimental spectra[159] of enolic deoxyguanosine.

From these results I can conclude that both semiclassical and classical frequencies are in agreement with the experiment. The accuracy is in agreement with the average error associated to the semiclassical method that is around 25-30 cm^{-1}. The similarity between classical and semiclassical spectra indicates the absence of relevant quantum effects in the system.

Switching now to the ketonic form of the molecule, I present results from the ab-initio family of methods and the one arising from Amber94. In Table 8.2 the DC-SCIVR and classical frequencies from the ab-initio calculations and from the force field are reported, together with the harmonic frequencies and the classical analysis performed with the thermalization phase. Furthermore, in Figure 8.5 the semiclassical spectra are reported.

The DC-SCIVR frequencies calculated ab-initio are in agreement with the experiment similarly to the enolic case reported above, i.e. accuracy is about 25 wavenumbers. Also in this case the ab-initio classical way to obtain spectra gives similar results with respect to the semiclassical one, confirming the limited influence of quantum effects.

Moving to the force field side, it is clear that all the methodology here employed fails in reproducing the experimental frequencies. Preceeding the NVE production phase with a termalization doesn't change the accuracy picture so

		DC-SCIVR	FT-vv	DC-SCIVR	FT-vv 300K	FT-vv 20K	FT-vv 1K	Harm.
mode	Exp.	*ab-initio*		*force field*				
3OH	-	3640	3662	3580	3656	3681	3681	3678
aNH2	3545	3560	3568	3400	3401	3426	3427	3425
NH	3438	3460	3473	3370	3398	3442	3443	3441
sNH2	3445	3460	3490	3190	3270	3289	3290	3288
5OH	-	3270	3252	3360	3382	3399	3397	3395

mode		Δ_{exp}		Δ_{exp}				
aNH2		15	23	145	144	119	118	120
NH		22	35	68	40	4	5	3
sNH2		15	45	255	175	156	155	157

Table 8.2.: Semiclassical DC-SCIVR, classical from velocity-velocity correlation function (FT-vv), and experimental frequencies of enolic deoxyguanosine obtained using ab-initio and force field methods. The column $\Delta_{DC-SCIVR}$ indicates the absolute difference between experimental and semiclassical values while Δ_{FT-vv} between experimental and classical values.Thermalization temperatures are specified in Kelvin. Δ_{Harm} stands for the absolute difference between experiment and the harmonic estimate. The experimental frequencies come from ref. [163]

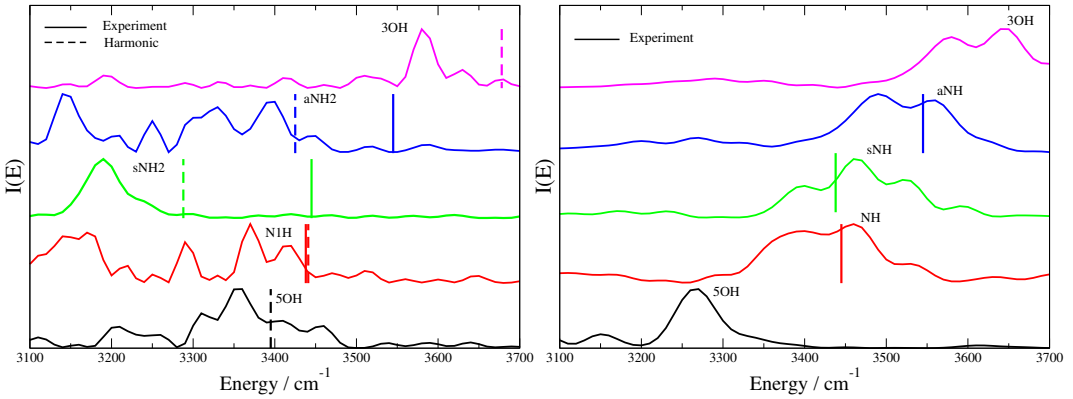

Figure 8.5.: Semiclassical DC-SCVIR results obtained using the force field (left panel), and the ab-initio molecular dynamics (right panel) for ketonic deoxyguanosine. The experimental values are reported with continous lines, while force field harmonic estimates with dashed lines. The experimental frequencies come from ref. [163]

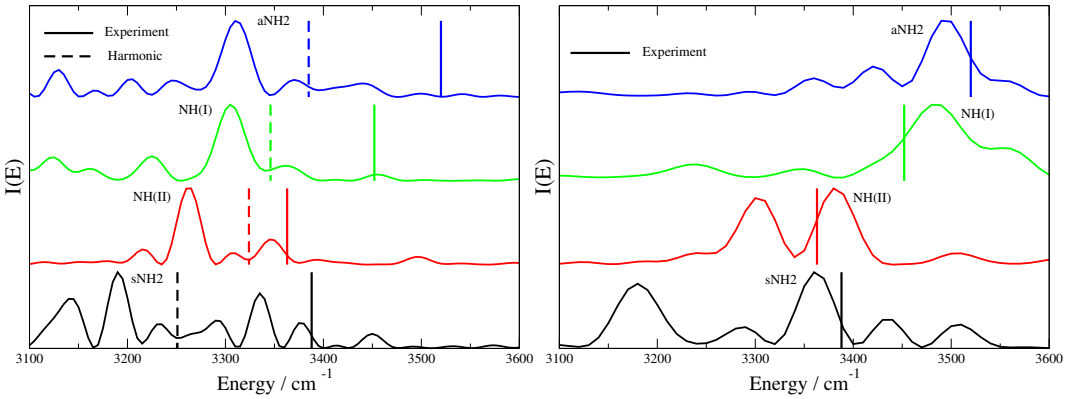

Figure 8.6.: High frequency region of the simulated spectra of the Ace-Phe-Met-NH_2dipeptide. On the left are shown the Amber94 results, while on the right the AIMD spectra. In each panel the experimental frequencies are reported with continuous lines, while the dashed lines in the Amber spectrum represent the harmonic estimates.

much, and, as expected, it leads to the harmonic set of frequencies as the temperature decreases, approaching 0 K. The only frequency value that is correctly predicted is the NH stretching when the harmonic or classical approaches at appropriate temperature are applied. This can be an indication that Amber can give a correct answer if the motion associated to the desired normal mode is free and uncoupled. The worst Amber estimate is provided when the semiclassical analysis is applied. These issues can be explained considering that the Amber minimum on the potential energy surface was specifically constructed to give harmonic frequencies as close as possible to the experiment frequencies. For this reason the actual potential well results more anharmonic in shape than the real one, leading all the methods that evaluate the vibrational power spectrum by exploring the phase space around the minimum to provide too low frequencies. In other words I can summarize this feature saying that the anharmonicity is overestimated.

8.3.2. Ace-Phe-Met-NH_2

Moving to the Ace-Phe-Met-NH_2 dipeptide I applied the same set of simulations on the most stable conformer, as reported in the work of Mons [160] and illustrated in Fig 8.2. The frequency values for each calculation is reported in Tab. 8.3 while images illustrating the spectra are shown in Fig 8.6.

All the considerations drawn for the deoxyguanosine molecules remain valid

		DC-SCIVR	FT-vv	DC-SCIVR	FT-vv 320K	FT-vv 20K	Harm.
mode	Exp.	*ab-initio*		*force field*			
aNH2	3520	3490	3490	3310	3376	3385	3385
NH(I)	3452	3480	3490	3305	3328	3347	3346
NH(II)	3363	3380	3396	3270	3317	3325	3324
sNH2	3388	3360	3356	3190	3253	3253	3251
mode		Δ_{exp}		Δ_{exp}			
aNH2		30	30	210	144	135	135
NH(I)		28	38	147	124	105	106
NH(II)		17	33	93	46	38	39
sNH2		28	32	198	135	135	137

Table 8.3.: Semiclassical DC-SCIVR, classical velocity-velocity correlation function (FT-vv) and experimental frequencies of Ace-Phe-Met-NH_2 dipeptide obtained with both ab-initio (left) and force field (right) methods. The column $\Delta_{DC-SCIVR}$ indicates the absolute difference between experimental and semiclassical values while Δ_{FT-vv} between experimental and classical values. Thermalization temperatures are specified in Kelvin. Δ_{Harm} stands for the absolute difference between the experiment and the harmonic estimate. The experimental frequencies come from ref. [160]

also in this case. Amber94 simulations in fact perform at their best when the harmonic approach is applied. They are reliable only for the detection of the NH(II) stretching band, that represents the NH involved in the hydrogen bond with the sulfur atom. The second NH stretching band, labelled as NH(I), forms a bond with the π electron system of the phenilalanine benzene ring. This interaction is certainly dependent on an accurate description of the chemical and electronic environment and for this reason the peak associated to such normal mode is not accurately predicted. All other methods are less effective, and in particular the semiclassical approach gives the worst set of frequencies. One more time classical simulations preceded by a termalization phase bring the results to the harmonic frequencies as the temperature goes to zero. On the other hand, the ab-initio calculations return more accurate results, in both the classical and semiclassical framework.

8.4. Potential energy surfaces analysis

In order to understand the origin of the excess of anharmonicity detected by the semiclassical investigation performed with the Amber94 potential, I looked at specific cuts of the potential energy surface. I evaluated with Amber94 and with the B3LYP functional the potential energy value of the system at its equilibrium configuration in all the normal modes except for the considered one, whose value was spanned on a grid. The cuts coming from both methods are illustrated in Figure 8.7.

For the majority of the normal modes I can observe that both level of calculations detect the correct anharmonic shape of the potential well, but that in general Amber94 presents wells that are broader than the DFT ones, in particular in the repulsive part of the curve, and around the minimum position the curves result more concave. The first observation can justify a lower set of frequency values obtained with analyses that exploit trajectories that run around the minimum, while the second one can be explained thinking that the Amber94 potential has been constructed to give harmonic estimates close to the real anharmonic ones.

8.5. Conclusions

In this chapter I have presented the vibrational spectra of the deoxyguanosine molecule, in both its ketonic and enolic conformations, and of the Ace-Phe-Met-NH_2dipeptide. When I employed semiclassical and classical methods associated with AIMD, I accurately reproduced the experimental high frequency region

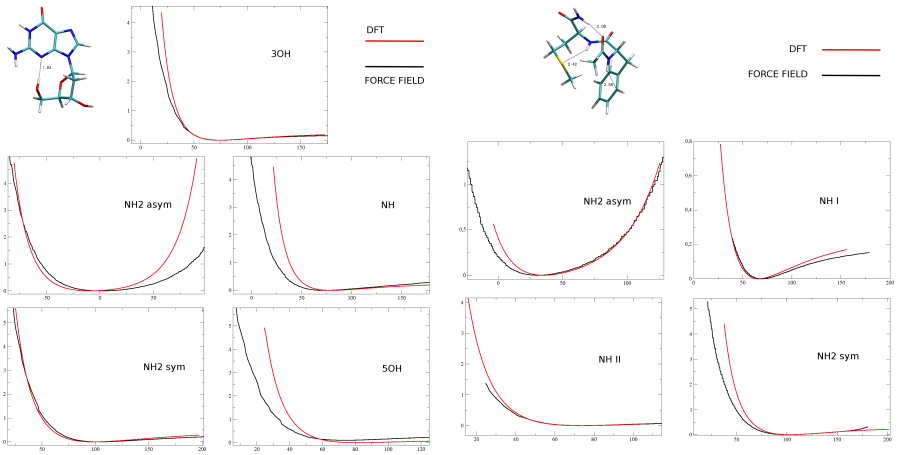

Figure 8.7.: Comparison between Amber94 and DFT B3LYP / 6-31G* potential energy surface sections for deoxyguanosine on the left and for the Ace-Phe-Met-NH_2dipeptide on the right.

in both applications. Using Amber94, on the contrary, I am not able to recover reasonable frequency values for the majority of the peaks in the spectrum. An exception is represented by simple NH stretching frequencies, denoting that the force field can give a valid estimate only when it deals with very simple motions, that are independent from the chemical environment. The best way to estimate vibrational frequencies using a force field seems to be the calculation of harmonic frequencies, while an evaluation through the velocity-velocity correlation function or the application of the semiclassical formalism largely worsen the results. Moreover I find that preceding the microcanonical phase of production with a thermalization leads the frequencies values to be equal to harmonic estimates, as the temperature decreases. The poor results coming from the applications of Amber94 can derive from the fact that the force field has been parametrized for being principally used in water solvent, while all the simulations here are performed in gas phase, although general observations on the PES exploration should be valid also in condensed phase. Performing a semiclassical analysis on a molecule in solvent is still an open field. There, the use of force fields could be still worth to be investigated.

Conclusions and Future Developments

In this thesis work the semiclassical method has been applied to the calculation of vibrational spectra of medium size molecular systems. The aim is to demonstrate that the semiclassical approach can be successfully employed for systems with dozens of degrees of freedom with the possibility to include the contribution of quantum effects, that can play a determinant role also in this kind of investigations. MC-SCIVR has been applied first to the vibrational spectrum calculation of the neutral glycine molecule. Here the problem of a multi-well potential energy surface has been treated, concluding that, with a dynamical-based method like MC-SCIVR, frequencies values are generally not influenced by the presence of more than one accessible minimum. The frequencies estimates are in agreement with other previously performed calculations and with the experiment, confirming the validity of the MC-SCIVR approach. Then the new DC-SCIVR has been benchmarked against MC-SCIVR on the same system, obtaining only slight differences in vibrational frequencies values, amounting to a few wavenumbers. This finding permitted the application of the DC-SCIVR semiclassical approach to systems bigger than the neutral glycine. In fact the spectrum of protonated glycine, isolated and also tagged with an increasing number of hydrogen molecules, has been evaluated with this new method. On such systems the importance of quantum effect inclusion arises clearly, since only the semiclassical method has been able to reproduce the experimental findings correctly. Moving to other applications, DC-SCIVR has proved its reliability to handle medium size systems dealing with the quasi solvated glycine problem and with biological systems, specifically a small dipeptide and a nucleoside.

Therefore the goals of this thesis work have been reached. By means of all the above mentioned applications, DC-SCIVR has demonstrated its capability to handle such kinds of systems and it has shown that quantum contributions should be not neglected because they can be relevant in some cases, for instance when the experimental temperature is low or light atoms like protons are involved.

DC-SCIVR associated to AIMD seems a promising tool for investigating even bigger systems of biological interest. Accurate results obtained for the Deoxyguanosine molecule indicate that a comprehensive investigation of vibrational spectra of all the DNA and RNA nucleobasis can be advisable, together with applications on the more interesting base-pairs systems. Alongside, studies on innovative strategies to ease the computational effort required are being constantly investigated. An example is the use of a novel and still under development Hessian database approach, described in Appendix 1.d, which has been pioneeringly applied in this thesis to the system made of glycine interacting with nine water molecules. Also the test of Amber94 Force Fields has represented an attempt in the direction of time consume reduction. The poor performance obtained

suggests that a different kind of force field, maybe specifically constructed for small organic molecules, should be better employed in future investigations.

In summary, all the results here reported promote the semiclassical method as a valuable tool in the theoretical vibrational spectroscopic field. In particular they point out the increase in dimensionality that MC-SCIVR employed in a Divide-and-Conquer fashion has made possible.

Bibliography

[1] Rohit Bhargava. Infrared spectroscopic imaging: the next generation. *Applied spectroscopy*, 66(10):1091–1120, 2012.

[2] Anastasia BS Elliott, Raphael Horvath, and Keith C Gordon. Vibrational spectroscopy as a probe of molecule-based devices. *Chemical Society Reviews*, 41(5):1929–1946, 2012.

[3] Hajime Torii and Mitsuo Tasumi. Model calculations on the amide-i infrared bands of globular proteins. *The Journal of chemical physics*, 96(5):3379–3387, 1992.

[4] Henry H Mantsch, Dennis Chapman, et al. *Infrared spectroscopy of biomolecules*. Wiley-Liss, 1996.

[5] Junrong Zheng and Michael D Fayer. Hydrogen bond lifetimes and energetics for solute/solvent complexes studied with 2d-ir vibrational echo spectroscopy. *Journal of the American Chemical Society*, 129(14):4328–4335, 2007.

[6] Jos Oomens, Gerard Meijer, and Gert von Helden. Gas phase infrared spectroscopy of cationic indane, acenaphthene, fluorene, and fluoranthene. *The Journal of Physical Chemistry A*, 105(36):8302–8309, 2001.

[7] G Gregoire, MP Gaigeot, DC Marinica, J Lemaire, JP Schermann, and C Desfrançois. Resonant infrared multiphoton dissociation spectroscopy of gas-phase protonated peptides. experiments and car–parrinello dynamics at 300 k. *Physical Chemistry Chemical Physics*, 9(24):3082–3097, 2007.

[8] Piotr Borowski. An evaluation of scaling factors for multiparameter scaling procedures based on dft force fields. *The Journal of Physical Chemistry A*, 116(15):3866–3880, 2012.

[9] Yuthana Tantirungrotechai, Ketthip Phanasant, Supacharee Roddecha, Panida Surawatanawong, Vallaya Sutthikhum, and Jumras Limtrakul. Scaling factors for vibrational frequencies and zero-point vibrational energies

of some recently developed exchange-correlation functionals. *Journal of Molecular Structure: THEOCHEM*, 760(1-3):189–192, 2006.

[10] Joel M Bowman, Tucker Carrington, and Hans-Dieter Meyer. Variational quantum approaches for computing vibrational energies of polyatomic molecules. *Molecular Physics*, 106(16-18):2145–2182, 2008.

[11] Vincenzo Barone. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.*, 122(1):014108, 2005.

[12] Roberto Car and Michele Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471–2474, Nov 1985.

[13] Kenneth G Kay. Semiclassical propagation for multidimensional systems by an initial value method. *J. Chem. Phys.*, 101(3):2250–2260, 1994.

[14] Kenneth G Kay. Numerical study of semiclassical initial value methods for dynamics. *J. Chem. Phys.*, 100(6):4432–4445, 1994.

[15] Eli Pollak. *The Semiclassical Initial Value Series Representation of the Quantum Propagator*, pages 259–271. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[16] Eric J Heller. *J. Chem. Phys.*, 94(4):2723–2729, 1991.

[17] Nandini Ananth, Charulatha Venkataraman, and William H Miller. Semiclassical description of electronically nonadiabatic dynamics via the initial value representation. *The Journal of chemical physics*, 127(8):084114, 2007.

[18] S. Bonella, D. Montemayor, and D. F. Coker. Linearized path integral approach for calculating nonadiabatic time correlation functions. *Proc. Natl. Ac. Sci.*, 102:6715–6719, 2005.

[19] S Bonella and DF Coker. A semiclassical limit for the mapping hamiltonian approach to electronically nonadiabatic dynamics. *The Journal of Chemical Physics*, 114(18):7778–7789, 2001.

[20] Sara Bonella and David F Coker. Semiclassical implementation of the mapping hamiltonian approach for nonadiabatic dynamics using focused initial distribution sampling. *The Journal of chemical physics*, 118(10):4370–4385, 2003.

[21] Pengfei Huo and David F Coker. Semi-classical path integral non-adiabatic dynamics: a partial linearized classical mapping hamiltonian approach. *Molecular Physics*, 110(9-10):1035–1052, 2012.

[22] Sergey V Antipov, Ziyu Ye, and Nandini Ananth. Dynamically consistent method for mixed quantum-classical simulations: A semiclassical approach. *The Journal of chemical physics*, 142(18):184102, 2015.

[23] Matthew S Church, Sergey V Antipov, and Nandini Ananth. Validating and implementing modified filinov phase filtration in semiclassical dynamics. *The Journal of chemical physics*, 146(23):234104, 2017.

[24] Matthew S Church, Timothy JH Hele, Gregory S Ezra, and Nandini Ananth. Nonadiabatic semiclassical dynamics in the mixed quantum-classical initial value representation. *The Journal of chemical physics*, 148(10):102326, 2018.

[25] Michele Ceotto, Sule Atahan, Sangwoo Shim, Gian Franco Tantardini, and Alan Aspuru-Guzik. First-principles semiclassical initial value representation molecular dynamics. *Phys. Chem. Chem. Phys.*, 11:3861–3867, 2009.

[26] Jorg Tatchen and Eli Pollak. Semiclassical on-the-fly computation of the s0 -> s1 absorption spectrum of formaldehyde. *J. Chem. Phys.*, 130(4):041103, 2009.

[27] Richard Phillips Feynman and A Ro Hibbs. *Quantum mechanics and path integrals*. McGraw-Hill, 1965.

[28] Michael V Berry and KE Mount. Semiclassical approximations in wave mechanics. *Rep. on Prog. Phys.*, 35(1):315, 1972.

[29] Riccardo Conte and Michele Ceotto. Semiclassical molecular dynamics for spectroscopic calculations. Under Review.

[30] John H Van Vleck. The correspondence principle in the statistical interpretation of quantum mechanics. *Proc. Natl. Acad. Sci.*, 14(2):178–188, 1928.

[31] Martin C Gutzwiller. Phase-Integral Approximation in Momentum Space and the Bound States of an Atom. *J. Math. Phys.*, 8(10):1979–2000, 1967.

[32] William H. Miller. Classical S Matrix: Numerical Application to Inelastic Collisions. *J. Chem. Phys.*, 53(9):3578–3587, 1970.

[33] William H Miller and Thomas F George. Semiclassical theory of electronic transitions in low energy atomic and molecular collisions involving several nuclear degrees of freedom. *J. Chem. Phys.*, 56(11):5637–5652, 1972.

[34] Michael F. Herman and Edward Kluk. A semiclasical justification for the use of non-spreading wavepackets in dynamics calculations. *Chem. Phys.*, 91(1):27–34, 1984.

[35] Kenneth G Kay. Integral expressions for the semiclassical time-dependent propagator. *J. Chem. Phys.*, 100(6):4377–4392, 1994.

[36] William J Thistleton, John A Marsh, Kenric Nelson, and Constantino Tsallis. Generalized box–müller method for generating q-gaussian random deviates. *IEEE transactions on information theory*, 53(12):4805–4810, 2007.

[37] Alexey L. Kaledin and William H. Miller. Time averaging the semiclassical initial value representation for the calculation of vibrational energy levels. *J. Chem. Phys.*, 118(16):7174–7182, 2003.

[38] Alexey L. Kaledin and William H. Miller. Time averaging the semiclassical initial value representation for the calculation of vibrational energy levels. ii. application to h2co, ch4, ch2d2. *J. Chem. Phys.*, 119(6):3078–3084, 2003.

[39] Yu Zhuang, Matthew R. Siebert, William L. Hase, Kenneth G. Kay, and Michele Ceotto. Evaluating the Accuracy of Hessian Approximations for Direct Dynamics Simulations. *J. Chem. Theory Comput.*, 9(1):54–64, 2012.

[40] Michele Ceotto, Gian Franco Tantardini, and Alan Aspuru-Guzik. Fighting the curse of dimensionality in first-principles semiclassical calculations: Non-local reference states for large number of dimensions. *J. Chem. Phys.*, 135(21):214108, 2011.

[41] Michele Ceotto, Stephanie Valleau, Gian Franco Tantardini, and Alan Aspuru-Guzik. First principles semiclassical calculations of vibrational eigenfunctions. *J. Chem. Phys.*, 134(23):234103, 2011.

[42] Michele Ceotto, Sule Atahan, Gian Franco Tantardini, and Alan Aspuru-Guzik. Multiple coherent states for first-principles semiclassical initial value representation molecular dynamics. *J. Chem. Phys.*, 130(23):234113, 2009.

[43] Dario Tamascelli, Francesco Saverio Dambrosio, Riccardo Conte, and Michele Ceotto. Graphics processing units accelerated semiclassical initial

value representation molecular dynamics. *J. Chem. Phys.*, 140(17):174109, 2014.

[44] N. De Leon and Eric J. Heller. Semiclassical quantization and extraction of eigenfunctions using arbitrary trajectories. *J. Chem. Phys.*, 78:4005–4017, 1983.

[45] Riccardo Conte, Alan Aspuru-Guzik, and Michele Ceotto. Reproducing deep tunneling splittings, resonances, and quantum frequencies in vibrational spectra from a handful of direct ab initio semiclassical trajectories. *J. Phys. Chem. Lett.*, 4(20):3407–3412, 2013.

[46] Michele Ceotto, Giovanni Di Liberto, and Riccardo Conte. Semiclassical "Divide-and-Conquer" Method for Spectroscopic Calculations of High Dimensional Molecular Systems. *Phys. Rev. Lett.*, 119(1):010401, 2017.

[47] Marius Wehrle, Miroslav Sulc, and Jiri Vanicek. On-the-fly ab initio semiclassical dynamics: Identifying degrees of freedom essential for emission spectra of oligothiophenes. *J. Chem. Phys.*, 140(24):244114, 2014.

[48] Giovanni Di Liberto, Riccardo Conte, and Michele Ceotto. Divide and conquer semiclassical molecular dynamics: A practical method for spectroscopic calculations of high dimensional molecular systems. *The Journal of Chemical Physics*, 148(1):014307, 2018.

[49] H Wu, M Rahman, J Wang, U Louderaj, WL Hase, and Y Zhuang. Higher-accuracy schemes for approximating the hessian from electronic structure calculations in chemical dynamics simulations. *J. Chem. Phys.*, 133(7):074101, 2010.

[50] Haobin Wang, David E. Manolopoulos, and William H. Miller. Generalized Filinov transformation of the semiclassical initial value representation. *J. Chem. Phys.*, 115(14):6317–6326, 2001.

[51] Giovanni Di Liberto and Michele Ceotto. The importance of the pre-exponential factor in semiclassical molecular dynamics. *J. Chem. Phys.*, 145:144107, 2016.

[52] Andrew R Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.

[53] Jan H Jensen. *Molecular modeling basics*. CRC Press, 2010.

[54] Jensen Frank. Introduction to computational chemistry. 1999.

[55] Robert A Latour. Perspectives on the simulation of protein–surface interactions using empirical force field methods. *Colloids and Surfaces B: Biointerfaces*, 124:25–37, 2014.

[56] William L Jorgensen. Quantum and statistical mechanical studies of liquids. 10. transferable intermolecular potential functions for water, alcohols, and ethers. application to liquid water. *Journal of the American Chemical Society*, 103(2):335–340, 1981.

[57] Scott J Weiner, Peter A Kollman, David A Case, U Chandra Singh, Caterina Ghio, Guliano Alagona, Salvatore Profeta, and Paul Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106(3):765–784, 1984.

[58] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[59] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[60] Jan Horníček, Petra Kaprálová, and Petr Bouř. Simulations of vibrational spectra from classical trajectories: Calibration with ab initio force fields. *The Journal of chemical physics*, 127(8):084502, 2007.

[61] Pier Luigi Silvestrelli, M Bernasconi, and Michele Parrinello. Ab initio infrared spectrum of liquid water. *Chemical Physics Letters*, 277(5-6):478–482, 1997.

[62] M Bernasconi, PL Silvestrelli, and M Parrinello. Ab initio infrared absorption study of the hydrogen-bond symmetrization in ice. *Physical review letters*, 81(6):1235, 1998.

[63] Radu Iftimie and Mark E Tuckerman. Decomposing total ir spectra of aqueous systems into solute and solvent contributions: A computational approach using maximally localized wannier orbitals. *The Journal of chemical physics*, 122(21):214508, 2005.

[64] Marie-Pierre Gaigeot and Michiel Sprik. Ab initio molecular dynamics computation of the infrared spectrum of aqueous uracil, 2003.

[65] Kijeong Kwac, Kyung-Koo Lee, Jae Bum Han, Kwang-Im Oh, and Minhaeng Cho. Classical and quantum mechanical/molecular mechanical molecular dynamics simulations of alanine dipeptide in water: Comparisons with ir and vibrational circular dichroism spectra. *The Journal of chemical physics*, 128(10):03B606, 2008.

[66] Marie-Pierre Gaigeot. Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective: gas and aqueous phase. *Phys. Chem. Chem. Phys.*, 12(14):3336–3359, 2010.

[67] A Cimas, TD Vaden, TSJA De Boer, LC Snoek, and M-P Gaigeot. Vibrational spectra of small protonated peptides from finite temperature md simulations and irmpd spectroscopy. *Journal of chemical theory and computation*, 5(4):1068–1078, 2009.

[68] Marie-Pierre Gaigeot, Rodolphe Vuilleumier, Michiel Sprik, and Daniel Borgis. Infrared spectroscopy of n-methylacetamide revisited by ab initio molecular dynamics simulations. *Journal of chemical theory and computation*, 1(5):772–789, 2005.

[69] DC Marinica, G Gregoire, C Desfrancois, JP Schermann, D Borgis, and MP Gaigeot. Ab initio molecular dynamics of protonated dialanine and comparison to infrared multiphoton dissociation experiments. *The Journal of Physical Chemistry A*, 110(28):8802–8810, 2006.

[70] Alessandro Motta, Marie-Pierre Gaigeot, and Dominique Costa. Aimd evidence of inner sphere adsorption of glycine on a stepped (101) boehmite alooh surface. *The Journal of Physical Chemistry C*, 116(44):23418–23427, 2012.

[71] YL Zhao, S Koppen, and T Frauenheim. An scc-dftb/md study of the adsorption of zwitterionic glycine on a geminal hydroxylated silica surface in an explicit water environment. *The Journal of Physical Chemistry C*, 115(19):9615–9621, 2011.

[72] Marco Nonella and Stefan Seeger. Monitoring peptide-surface interaction by means of molecular dynamics simulation. *Chemical Physics*, 378(1-3):73–81, 2010.

[73] Pascale Ehrenfreund, Daniel P. Glavin, Oliver Botta, George Cooper, and Jeffrey L. Bada. Extraterrestrial amino acids in orgueil and ivuna: Tracing the parent body of ci type carbonaceous chondrites. *Proc. Natl. Acad. Sci. USA*, 98(5):2138, 2001.

[74] Yi-Jehng Kuan, Steven B. Charnley, Hui-Chun Huang, Wei-Ling Tseng, and Zbigniew Kisiel. Interstellar glycine. *Astrophys. J.*, 593(2):848, 2003.

[75] Peter D. Godfrey and Ronald D. Brown. Shape of glycine. *J. Am. Chem. Soc.*, 117(7):2019–2023, 1995.

[76] Saraswathi Vishveshwara and John A. Pople. Molecular orbital theory of the electronic structures of organic compounds. 32. conformations of glycine and related systems. *J. Am. Chem. Soc.*, 99(8):2422–2426, 1977.

[77] Ching Han Hu, Mingzuo Shen, and Henry F. Schaefer III. Glycine conformational analysis. *J. Am. Chem. Soc.*, 115(7):2923–2929, 1993.

[78] Attila G. Csaszar. Conformers of gaseous glycine. *J. Am. Chem. Soc.*, 114(24):9568–9575, 1992.

[79] Jan H. Jensen and Mark S. Gordon. Conformational potential energy surface of glycine: a theoretical study. *J. Am. Chem. Soc.*, 113(21):7917–7924, 1991.

[80] B. Balta., M. Basma, V. Aviyente, C. Zhu, and C. Lifshitz. Structures and reactivity of gaseous glycine and its derivatives. *Int. J. Mass Spectrom.*, 201(1â3):69–85, 2000.

[81] Vincenzo Barone, Malgorzata Biczysko, Julien Bloino, and Cristina Puzzarini. Characterization of the elusive conformers of glycine from state-of-the-art structural, thermodynamic, and spectroscopic computations: Theory complements experiment. *J. Chem. Theory and Comput.*, 9(3):1533–1547, 2013.

[82] Vincenzo Barone, Malgorzata Biczysko, Julien Bloino, and Cristina Puzzarini. Glycine conformers: a never-ending story? *Phys. Chem. Chem. Phys.*, 15:1358–1363, 2013.

[83] David E. Woon and Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. iii. the atoms aluminum through argon. *J. Chem. Phys.*, 98(2):1358–1371, 1993.

[84] S. G. Stepanian, I. D. Reva, E. D. Radchenko, M. T. S. Rosado, M. L. T. S. Duarte, R. Fausto, and L. Adamowicz. Matrix-isolation infrared and theoretical studies of the glycine conformers. *J. Phys. Chem. A*, 102(6):1041–1054, 1998.

[85] Brina Brauer, Galina M. Chaban, and R. Benny Gerber. Spectroscopically-tested, improved, semi-empirical potentials for biological molecules: Calculations for glycine, alanine and proline. *Phys. Chem. Chem. Phys.*, 6:2543–2556, 2004.

[86] M. L. Senent, M. Villa, R. Dominguez-Gomez, and A. Fernandez-Clavero. Ab initio study of the far infrared spectrum of glycine. *Int. J. Quantum Chem.*, 104(4):551–561, 2005.

[87] M. L. Senent, S. Fernandez-Herrera, and Y. G. Smeyers. Ab initio determination of the roto-torsional energy levels of hydrogen peroxide. *Spectrochim. Acta Mol. Biomol. Spectrosc.*, 56(8):1457 – 1468, 2000.

[88] Ota Bludsky, Jana Chocholousova, Jaroslav Vacek, Friedrich Huisken, and Pavel Hobza. Anharmonic treatment of the lowest-energy conformers of glycine: A theoretical study. *J. Chem. Phys.*, 113(11):4629–4635, 2000.

[89] Paul E. Maslen, Nicholas C. Handy, Roger D. Amos, and Dylan Jayatilaka. Higher analytic derivatives. iv. anharmonic effects in the benzene spectrum. *J. Chem. Phys.*, 97(6):4233–4254, 1992.

[90] Nicholas C. Handy, Paul E. Maslen, Roger D. Amos, Jamie S. Andrews, Christopher W. Murray, and Gregory J. Laming. The harmonic frequencies of benzene. *Chem. Phys. Lett.*, 197(4):506 – 515, 1992.

[91] Malgorzata Biczysko, Julien Bloino, Ivan Carnimeo, Pawel Panek, and Vincenzo Barone. Fully ab initio ir spectra for complex molecular systems from perturbative vibrational approaches: Glycine as a test case. *J. Mol. Struct.*, 1009:74 – 82, 2012.

[92] Susanna Monti, Alessandro Corozzi, Peter Fristrup, Kaushik L. Joshi, Yun Kyung Shin, Peter Oelschlaeger, Adri C. T. van Duin, and Vincenzo Barone. Exploring the conformational and reactive dynamics of biomolecules in solution using an extended version of the glycine reactive force field. *Phys. Chem. Chem. Phys.*, 15:15062–15077, 2013.

[93] Jan M. L. Martin, Timothy J. Lee, Peter R. Taylor, and JeanâPierre Francois. The anharmonic force field of ethylene, c2h4, by means of accurate ab initio calculations. *J. Chem. Phys.*, 103(7):2589–2602, 1995.

[94] William H. Miller. Quantum dynamics of complex molecular systems. *Proc. Natl. Acad. Sci. USA*, 102(19):6660–6664, 2005.

[95] Kenneth G. Kay. The hermanâkluk approximation: Derivation and semiclassical corrections. *Chem. Phys.*, 322(1â2):3 – 12, 2006.

[96] C. Harabati, J. M. Rost, and F. Grossmann. Long-time and unitary properties of semiclassical initial value representations. *J. Chem. Phys.*, 120:26–30, 2004.

[97] F. Grossmann. A hierarchy of semiclassical approximations based on gaussian wavepackets. *Comments At. Mol. Phys.*, 34:141–160, 1999.

[98] Hiroki Nakamura, Shinkoh Nanbu, Yoshiaki Teranishi, and Ayumi Ohta. Development of semiclassical molecular dynamics simulation method. *Phys. Chem. Chem. Phys.*, 18:11972–11985, 2016.

[99] Alexey D. Kondorskiy and Shinkoh Nanbu. Electronically nonadiabatic wave packet propagation using frozen gaussian scattering. *J. Chem. Phys.*, 143:114103, 2015.

[100] Max Buchholz, Frank Grossmann, and Michele Ceotto. Mixed semiclassical initial value representation time-averaging propagator for spectroscopic calculations. *J. Chem. Phys.*, 144:094102, 2016.

[101] Jian Liu and William H. Miller. Using the thermal gaussian approximation for the boltzmann operator in semiclassical initial value time correlation functions. *J. Chem. Phys.*, 125:224104, 2006.

[102] Jian Liu and William H. Miller. Real time correlation function in a single phase space integral beyond the linearized semiclassical initial value representation. *J. Chem. Phys.*, 126:234110, 2007.

[103] Jian Liu and William H. Miller. Linearized semiclassical initial value time correlation functions using the thermal gaussian approximation: Applications to condensed phase systems. *J. Chem. Phys.*, 127:114506, 2007.

[104] Jian Liu and William H. Miller. Test of the consistency of various linearized semiclassical initial value time correlation functions in application

to inelastic neutron scattering from liquid para-hydrogen. *J. Chem. Phys.*, 128:144511, 2008.

[105] Jakob Petersen and Eli Pollak. Semiclassical initial value representation for the quantum propagator in the heisenberg interaction representation. *J. Chem. Phys.*, 143:224114, 2015.

[106] Hiroshi Ushiyama and Kazuo Takatsuka. Extended quantization condition for constructive and destructive interferences and trajectories dominating molecular vibrational eigenstates. *J. Chem. Phys.*, 122:224112, 2005.

[107] Satoshi Takahashi and Kazuo Takatsuka. Phase quantization of chaos in the semiclassical regime. *J. Chem. Phys.*, 127:084112, 2007.

[108] G. Tao. Efficient importance sampling in semiclassical initial value representation calculations for time correlation functions. *Theor. Chem. Acc.*, 133:1448, 2014.

[109] Stephanie Y. Y. Wong, David M. Benoit, Marius Lewerenz, Alex Brown, and Pierre-Nicholas Roy. Determination of molecular vibrational state energies using the ab initio semiclassical initial value representation: Application to formaldehyde. *J. Chem. Phys.*, 134(9):094110, 2011.

[110] Reuven Ianconescu, Joerg Tatchen, and Eli Pollak. On-the-fly semiclassical study of internal conversion rates of formaldehyde. *J. Chem. Phys.*, 139(15):154311, 2013.

[111] Bastiaan J. Braams and Joel M. Bowman. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.*, 28(4):577–606, 2009.

[112] Yuliya Paukku, Ke R Yang, Zoltan Varga, and Donald G Truhlar. Global ab initio ground-state potential energy surface of n_4. *J. Chem. Phys.*, 139(4):044309, 2013.

[113] Bin Jiang and Hua Guo. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. iii. molecule-surface interactions. *J. Chem. Phys.*, 141(3):034109, 2014.

[114] Riccardo Conte, Paul L Houston, and Joel M Bowman. Trajectory and model studies of collisions of highly excited methane with water using an ab initio potential. *J. Phys. Chem. A*, 119(50):12304–12317, 2015.

[115] R. Conte, C. Qu, and J. M. Bowman. Permutationally invariant fitting of many-body, non-covalent interactions with application to three-body methane-water-water. *J. Chem. Theory Comp.*, 11:1631–1638, 2015.

[116] Chen Qu, Qi Yu, and Joel M Bowman. Permutationally invariant potential energy surfaces. *Annual review of physical chemistry*, 69:151–175, 2018.

[117] M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T.L. Windus, and W.A. de Jong. Nwchem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.*, 181(9):1477 – 1489, 2010.

[118] Axel D. Becke. Densityâfunctional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98(7):5648–5652, 1993.

[119] Cristina Puzzarini, Malgorzata Biczysko, and Vincenzo Barone. Accurate harmonic/anharmonic vibrational frequencies for open-shell systems: Performances of the b3lyp/n07d model for semirigid free radicals benchmarked by ccsd(t) computations. *J. Chem. Theory Comput.*, 6(3):828–838, 2010.

[120] Cristina Puzzarini, Malgorzata Biczysko, and Vincenzo Barone. Accurate anharmonic vibrational frequencies for uracil: The performance of composite schemes and hybrid cc/dft model. *J. Chem. Theory Comput.*, 7(11):3702–3710, 2011.

[121] Eric G Diken, Jeffrey M Headrick, and Mark A Johnson. Photoelectron spectroscopy of the [glycineâ(h2o) 1, 2]- clusters: Sequential hydration shifts and observation of isomers. *J. Chem. Phys.*, 122(22):224317, 2005.

[122] Samy O Meroueh, Yanfei Wang, and William L Hase. Direct dynamics simulations of collision- and surface-induced dissociation of n-protonated glycine. shattering fragmentation. *J. Phys. Chem. A*, 106:9983–9992, 2002.

[123] Kyoyeon Park, Kihyung Song, and William L Hase. An ab initio direct dynamics simulation of protonated glycine surface-induced dissociation. *J. Mass. Spectrom.*, 265:326–336, 2007.

[124] Thomas H Lowry and Kathleen Schueller Richardson. *Mechanism and theory in organic chemistry*. Harper & Row New York, 1987.

[125] Ulrich J Lorenz and Thomas R Rizzo. Multiple isomers and protonation sites of the phenylalanine/serine dimer. *J. Am. Chem. Soc.*, 134(27):11053–11055, 2012.

[126] Antoine Masson, Evan R Williams, and Thomas R Rizzo. Molecular hydrogen messengers can lead to structural infidelity: A cautionary tale of protonated glycine. *J. Chem. Phys.*, 143(10):104313, 2015.

[127] Khadijeh Rajabi and Travis D Fridgen. Structures of aliphatic amino acid proton-bound dimers by infrared multiple photon dissociation spectroscopy in the 700- 2000 cm-1 region. *J. Phys. Chem A*, 112(1):23–30, 2008.

[128] Chad G Atkins, Khadijeh Rajabi, Elizabeth AL Gillis, and Travis D Fridgen. Infrared multiple photon dissociation spectra of proton-and sodium ion-bound glycine dimers in the n- h and o- h stretching region. *J. Phys. Chem A*, 112(41):10220–10225, 2008.

[129] Adeyemi A Adesokan and RB Gerber. Anharmonic vibrational spectroscopy calculations for proton-bound amino acid dimers. *The Journal of Physical Chemistry A*, 113(10):1905–1912, 2008.

[130] Anthony P Scott and Leo Radom. Harmonic vibrational frequencies: an evaluation of hartree- fock, møller- plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *J. Phys. Chem.*, 100(41):16502–16513, 1996.

[131] Udo Buck, Ingo Ettischer, Mario Melzer, Victoria Buch, and Joanna Sadlej. Structure and spectra of three-dimensional (H 2 O) n clusters, n= 8, 9, 10. *Phys. Rev. Lett.*, 80(12):2578, 1998.

[132] IM Alecu, Jingjing Zheng, Yan Zhao, and Donald G Truhlar. Computational thermochemistry: scale factor databases and scale factors for vibrational frequencies obtained from electronic model chemistries. *J. Chem. Theo. Comput.*, 6(9):2872–2887, 2010.

[133] Radu Iftimie, Peter Minary, and Mark E Tuckerman. Ab initio molecular dynamics: Concepts, recent developments, and future trends. *Proc. Natl. Acad. Sci.*, 102(19):6654–6659, 2005.

[134] Mark Tuckerman, Kari Laasonen, Michiel Sprik, and Michele Parrinello. Ab initio molecular dynamics simulation of the solvation and transport of hydronium and hydroxyl ions in water. *J. Chem. Phys.*, 103(1):150–161, 1995.

[135] Subha Pratihar, Xinyou Ma, Zahra Homayoon, George L Barnes, and William L Hase. Direct chemical dynamics simulations. *Journal of the American Chemical Society*, 139(10):3570–3590, 2017.

[136] Eric J Heller. The semiclassical way to molecular spectroscopy. *Acc. Chem. Res.*, 14(12):368–375, 1981.

[137] Marco Micciarelli, Riccardo Conte, Jaime Suarez, and Michele Ceotto. Anharmonic vibrational eigenfunctions and infrared spectra from semiclassical molecular dynamics. *The Journal of chemical physics*, 149(6):064115, 2018.

[138] Jan H Jensen and Mark S Gordon. On the number of water molecules necessary to stabilize the glycine zwitterion. *Journal of the American Chemical Society*, 117(31):8159–8170, 1995.

[139] Christine M Aikens and Mark S Gordon. Incremental solvation of nonionized and zwitterionic glycine. *Journal of the American Chemical Society*, 128(39):12835–12850, 2006.

[140] Sung-Woo Park, Suk Im, Sungyul Lee, and C Desfrancois. Structure and stability of glycine–(h2o) 3 cluster and anion: Zwitterion vs. canonical glycine. *International Journal of Quantum Chemistry*, 107(6):1316–1327, 2007.

[141] Steven M Bachrach. Microsolvation of glycine: a dft study. *The Journal of Physical Chemistry A*, 112(16):3722–3730, 2008.

[142] Ju-Young Kim, Suk Im, Bongsoo Kim, C Desfrancois, and Sungyul Lee. Structures and energetics of gly–(h2o) 5: Thermodynamic and kinetic stabilities. *Chemical Physics Letters*, 451(4-6):198–203, 2008.

[143] Toshiyuki Takayanagi, Takehiro Yoshikawa, Akira Kakizaki, Motoyuki Shiga, and Masanori Tachikawa. Molecular dynamics simulations of small glycine–(h2o) n (n= 2–7) clusters on semiempirical pm6 potential energy surfaces. *Journal of Molecular Structure: THEOCHEM*, 869(1-3):29–36, 2008.

[144] Kang Taek Lee, Kyu Young Han, Insun Oh, and Seong Keun Kim. Barrierless pathways in the neutral–zwitterion transition of amino acid: Glycine–(h2o) 9. *Chemical Physics Letters*, 495(1-3):14–16, 2010.

[145] Hakan Kayi, Ralf I Kaiser, and John D Head. A theoretical investigation of the relative stability of hydrated glycine and methylcarbamic acidâfrom

water clusters to interstellar ices. *Physical Chemistry Chemical Physics*, 14(14):4942–4958, 2012.

[146] Xiang-Jun Meng, Hong-Li Zhao, and Xing-Song Ju. Influences of n (2–5) water molecules on the proton transfer in hydrated glycine complexes. *Computational and Theoretical Chemistry*, 1001:26–32, 2012.

[147] Ju-Young Kim, Doo-Sik Ahn, Sung-Woo Park, and Sungyul Lee. Gas phase hydration of amino acids and dipeptides: effects on the relative stability of zwitterion vs. canonical conformers. *RSC Advances*, 4(31):16352–16361, 2014.

[148] Shoujun Xu, J Michael Nilles, and Kit H Bowen Jr. Zwitterion formation in hydrated amino acid, dipole bound anions: How many water molecules are required? *The Journal of chemical physics*, 119(20):10696–10701, 2003.

[149] Riet Ramaekers, Joanna Pajak, Bert Lambie, and Guido Maes. Neutral and zwitterionic glycine. h 2 o complexes: a theoretical and matrix-isolation fourier transform infrared study. *The Journal of chemical physics*, 120(9):4182–4193, 2004.

[150] Dominik Marx. Proton transfer 200 years after von grotthuss: Insights from ab initio simulations. *ChemPhysChem*, 7(9):1848–1870, 2006.

[151] Man YináYi et al. An am1 and pm3 molecular orbital and self-consistent reaction-field study of the aqueous solvation of glycine, alanine and proline in their neutral and zwitterionic forms. *Journal of the Chemical Society, Perkin Transactions 2*, (4):531–537, 1991.

[152] Shinichi Yamabe, Naoko Ono, and Noriko Tsuchida. Molecular interactions between glycine and h2o affording the zwitterion. *The Journal of Physical Chemistry A*, 107(39):7915–7922, 2003.

[153] Zorka Smedarchina, Willem Siebrand, Antonio Fernández-Ramos, Leonid Gorb, and Jerzy Leszczynski. A direct-dynamics study of proton transfer through water bridges in guanine and 7-azaindole. *The Journal of Chemical Physics*, 112(2):566–573, 2000.

[154] Terrence M Chang, Satrajit Chakrabarty, and Evan R Williams. Hydration of gaseous m-aminobenzoic acid: Ionic vs neutral hydrogen bonding and water bridges. *Journal of the American Chemical Society*, 136(29):10440–10449, 2014.

[155] Omar F Mohammed, Dina Pines, Jens Dreyer, Ehud Pines, and Erik TJ Nibbering. Sequential proton transfer through water bridges in acid-base reactions. *Science*, 310(5745):83–86, 2005.

[156] Omar F Mohammed, Dina Pines, Erik TJ Nibbering, and Ehud Pines. Base-induced solvent switches in acid–base reactions. *Angewandte Chemie International Edition*, 46(9):1458–1461, 2007.

[157] Ricardo Perez de Tudela and Dominik Marx. Water-induced zwitterionization of glycine: Stabilization mechanism and spectral signatures. *The journal of physical chemistry letters*, 7(24):5137–5142, 2016.

[158] Robert Benny Gerber, Galina M. Chaban, Susan K. Gregurick, and Brina Brauer. Vibrational spectroscopy and the development of new force fields for biological molecules. *Biopolymers*, 68(3):370–382, 2003.

[159] Hiroya Asami, Shu-hei Urashima, and Hiroyuki Saigusa. Hydration structures of 2'-deoxyguanosine studied by ir-uv double resonance spectroscopy comparison with guanosine. *Physical Chemistry Chemical Physics*, 11(44):10466–10472, 2009.

[160] Himansu S Biswal, Eric Gloaguen, Yohan Loquais, Benjamin Tardivel, and Michel Mons. Strength of nh··· s hydrogen bonds in methionine residues revealed by gas-phase ir/uv spectroscopy. *The journal of physical chemistry letters*, 3(6):755–759, 2012.

[161] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[162] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560, 1997.

[163] Myong Yong Choi and Roger E Miller. Four tautomers of isolated guanine from infrared laser spectroscopy in helium nanodroplets. *Journal of the American Chemical Society*, 128(22):7320–7328, 2006.

[164] E Nir, Ch Plützer, K Kleinermanns, and M De Vries. Properties of isolated dna bases, base pairs and nucleosides examined by laser spectroscopy. *The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics*, 20(3):317–329, 2002.