Università degli Studi di Milano
Department of Chemistry

PhD Course in Chemistry
XXXVI cycle

Semiclassical vibrational spectroscopy of biomolecular systems

Candidate:
Giacomo Botti

Supervisor:
Dr. Riccardo Conte

12th of July, 2024



UNIVERSITÀ DEGLI STUDI DI MILANO

Non senza fatiga si giunge al fine

Toccata Nona
Girolamo Frescobaldi

La suite lui prouva que non

Le Gorille
George Brassens

Heartbreak and toil and suffering gone

D-Day Dodgers
Harry Pynn

Because it's there

George Mallory

And you, to whom adversity has dealt the final blow \|\| With smiling bastards lying to you everywhere you go \|\| Turn to, and put out all your strength of arm and heart and brain \|\| And like the Mary Ellen Carter, rise again.\|\|

The Mary Ellen Carter
Stan Rogers

Contents

1 Introduction **1**

2 Theory **5**
2.1 Harmonic Frequencies 5
2.2 Classical Spectra 8
2.3 Semiclassical Initial Value Representation 13
2.4 MC and DCSCIVR 17
2.5 Action-Angle Variables 18
2.6 Einstein-Brillouin-Keller quantization 21
2.7 Adiabatic switching 22

3 Adiabatic Switching, on-the-fly **27**
3.1 Increasing the trajectory stability 27
3.2 The on-the-fly implementation 28
3.3 Preliminary tests 32
3.4 Small Molecules 34
3.5 Glycine 36
3.6 Conclusions: improving the sampling 44

4 The spectrum of proline **45**
4.1 The cornerstone of proteins 45
4.2 Computational details 49
4.3 A spectrum for two conformers 51
4.4 Conclusions: the role of anharmonicity 58

5 The spectrum of guanine-cytosine pair **61**
5.1 Tunneling between DNA bases 61
5.2 Simulation setup for QCT and DCSCIVR 65
5.3 Discussion of the simulated spectra 67
5.4 Conclusions: the absence of quantum tunneling 78

6 The need for quantum effects **81**
6.1 Quantum methods for quantum effects 81
6.2 Adiabatic quantization 82
6.3 Conclusions: the quantum method touchstone 85

7 Conclusions **87**

Acronyms **101**

List of Figures

1.1 Introduction: the theoretical spectroscopist's triangle 2

3.1 OTF AS: flow chart of an adiabatic switching step 31
3.2 OTF AS: switching trajectory energy profile (H_2) 33
3.3 OTF AS: MC and ASSCIVR spectra (H_2) 34
3.4 OTF AS: MC and ASSCIVR spectra (H_2CO) 35
3.5 OTF AS: glycine equilibrium geometry 37
3.6 OTF AS: peak comparison for Gly 40
3.7 OTF AS: glycine ZPE comparison 41
3.8 OTF AS: comparison 42
3.9 OTF AS: refined spectra of mode 16 42

4.1 Proline: experimental spectrum 47
4.2 Proline: equilibrium geometries of the main conformers 48
4.3 Proline: spectra of OH stretching 53
4.4 Proline: IR spectrum of *trans*-Proline 54
4.5 Proline: spectra of CH stretching 56
4.6 Proline: spectra of CO stretching 56
4.7 Proline: spectra of COH and CNH bnd 57

5.1 Guanine-cytosine: structural formulæ of the main tautomers . 64
5.2 Guanine-cytosime: double proton hopping reaction 64
5.3 Guanine-cytosine: experimental spectrum 65
5.4 Guanine-cytosine: equilibrium geometries 68
5.5 Guanine-cytosine: harmonic, QCT and semiclassical spectra (1) 70
5.6 Guanine-cytosine: harmonic, QCT and semiclassical spectra (2) 71
5.7 Guanine-cytosine: low frequency analysis 72
5.8 Guanine-cytosine: experimental spectrum (alkylated bases) . . 73
5.9 Guanine-cytosine: experimental and semiclassical (K9K-1) . . . 75
5.10 Guanine-cytosine: experimental and semiclassical (K7E-1) . . . 77

6.1 NQE: schematic representation of the AS procedure 83

6.2 NQE: final energy distributions of AS H_2CO (50 000 au) 84
6.3 NQE: final energy distributions of AS H_2CO (50 000 au) 84
6.4 NQE: classical spectra of H_2CO mode 6 85
6.5 NQE: ASSCIVR and MD spectra of H_2CO 86

List of Tables

2.1 Theory: trajectory rejection using AS 25

3.1 OTF AS: computational times for finite differences AS 29
3.2 OTF AS: trajectory time comparison 32
3.3 OTF AS: energy standard deviation 33
3.4 OTF AS: frequency values for H_2CO 36
3.5 OTF AS: frequency analysis for Gly 39
3.6 OTF AS: glycine frequencies using DC-ASSCIVR 43
3.7 OTF AS: refined frequencies of mode 16 43

4.1 Proline: comparison between different levels of theory 49
4.2 Proline: conformer energies 52
4.3 Proline: ZPE and adiabatic electronic energies 52
4.4 Proline: harmonic analysis of the conformers 53
4.5 Proline: harmonic and ASSCIVR CH stretching frequencies . . 55
4.6 Proline: summary of the assignment 58

5.1 Guanine-cytosine: basis set comparison 66
5.2 Guanine-cytosine: conformational landscape 69
5.3 Guanine-cytosine: comparison between QCT and harmonic . . 69
5.4 Guanine-cytosine: frequency comparison for NH oop *bnd* . . . 72
5.5 Guanine-cytosine: basis set comparison for Gua NH@NH_2 *str* . 74
5.6 Guanine-cytosine: frequency comparison for K9K-1 and K7E-1 76

6.1 NQE: H_2CO AS vibrational frequencies 83

Listings

3.1 BASH script for handling the *ab initio* tasks with ORCA. 28
4.1 Sample input for proline optimization 49
4.2 Sample input for proline classical evolution 50
5.1 Sample input for guanine-cytosine optimization 65
5.2 Sample input for guanine-cytosine evolution 67

Chapter 1

Introduction

All the work of this thesis revolves around the application of semiclassical (SC) vibrational spectroscopy to biomolecular systems. This was done by both implementing new methods and formalizing the best practices required to treat biomolecular systems with the most accurate level of electronic theory compatible with semiclassical dynamics. Vibrational spectroscopy is a priceless technique in an uncountable number of fields, ranging from astrophysics to archaeology. As professor emeritus J.S.Siegel (Tianjin University) said during a class, vibrational spectroscopy "is the eye we have on structure", since its signals are heavily influenced not only by the structure of the target compound, but also by its environment. Indeed, even interactions that are weak if compared with a chemical bond can have a visible effect on the vibrational spectrum. As an experimental field, vibrational spectroscopy is constantly improved *via* new instruments and new techniques, making it possible to obtain selectively the spectrum of weakly bonded species such as carbonyl monohydrates and nucleobase pairs.[1–6] But with improved experiments come new challenges in the experimental assignment.

For instance, the experimental spectra of biomolecular systems are particularly difficult to assign, because they require methods precise enough to distinguish between conformers of the same species, yet affordable enough to treat medium-to-large size system, with the possibility of including the solvent. In this case, one has to balance the three sides of a particular "engineering triangle" (Figure 1.1) formed by the accuracy of the potential, the accuracy of the vibrational model and the system size. A large-size biomolecular system, such as a peptide or a DNA strand, could be treated *ab initio* only by adopting the harmonic approximation as model for the vibrations. On the other hand, molecular dynamics (MD) can be performed on the same system using a force field potential, thus

adopting a classical dynamics model for the vibrational problem. [7–15]

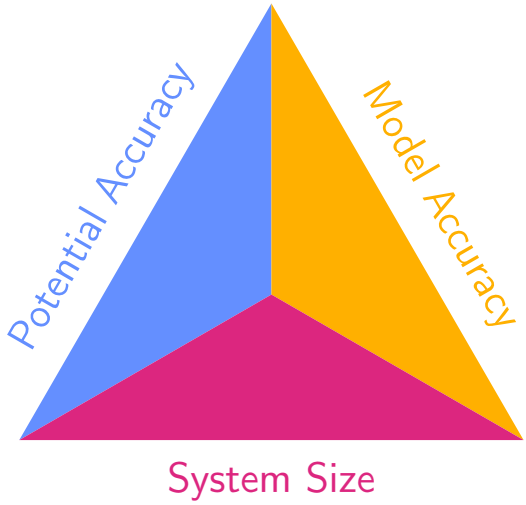


FIGURE 1.1: *The theoretical spectroscopist has their own triangle, composed by the accuracy of the potential, the accuracy of the model and the size of the target system.*

On top of that, the proper description of biomolecular systems should include anharmonicity and nuclear quantum effects (NQE). The former is commonly described as a deviation from the harmonic estimate and comes from the fact that the actual molecular potential is harmonic only in neighbourhood of the minimum. The latter include phenomena such as tunneling split, overtones, combination bands, quantum resonances and even the plain but fundamental zero point energy (ZPE).[1] Indeed, biomolecular systems are quite anharmonic and the inclusion of quantum effects can be pivotal in understanding their conformational landscape and *in vivo* activity.[16, 17] Anharmonicity can be included *via* MD, allowing the system to properly explore its potential energy surface (PES). Unfortunately, a treatment based on classical mechanics cannot include NQE.

Still, many quantum methods, such as discrete variable representation (DVR) and vibrational configuration interaction (VCI), scale unfavorably with the number of degrees of freedom of the systems, making the application to biomolecules rather prohibitive.[18, 19] Vibrational perturbation theory (VPT) is a popular family of time-independent methods to compute vibrational frequencies, but the case of nearly-degenerate vibrational states needs an *ad hoc* treatment, hindering the application to large systems, which usually have very crowded vibrational states.[20, 21] Moving to a time-dependent description, multiconfiguration time-dependent Hartree (MCTDH) requires *ad hoc* PES and dealing with high dimen-

sional systems is still non trivial, even if new techniques are being developed to overcome this issue.[22–25] On the other hand, path integral methods like centroid molecular dynamics (CMD) and ring polymer molecular dynamics (RMPD) are hindered by resonances between the vibrational modes and the ring polymer modes.[26]

Finally, a practical way to add quantum effects to classical molecular dynamics simulations is represented by the semiclassical initial value representation (SCIVR) molecular dynamics.[27–37] In this approach, the quantum mechanical propagator is approximated using classical dynamical quantities collected along a trajectory, giving a reliable description of both anharmonicity and NQE. However, when applying SCIVR on large dimensional systems, one still has to deal with the computational times, the difficulty to get a resolved spectroscopic signal and the necessity of maintaining the stability of the semiclassical simulation, in spite of the chaoticity of classical trajectories. For these reasons, several semiclassical techniques have been developed over the years, including multiple coherent semiclassical initial value representation (MCSCIVR) and a divide-and-conquer (DC) strategy to deal with the spectroscopy of high dimensional systems.[38–47] Recently, the combination of adiabatic switching (AS) and SCIVR methods (ASSCIVR) was use to overcome the issue of chaotic trajectories.[48–53]

This last method was only implemented on PES. Since PES of biomolecular systems are not readily available, during my Ph.D. I worked on adapting AS to on-the-fly dynamics so that any electronic structure software could be used to compute the semiclassical spectrum, as in the case of MCSCIVR. The implementation and benchmark of on-the-fly ASSCIVR are described in Chapter 3. During the work on on-the-fly ASSCIVR, I discovered that the experimental spectrum of proline by Stepanian *et al.* presented some features incompatible with the lowest energy conformer,[54] so I proceeded with a systematic assignment using on-the-fly ASSCIVR, whose results are included in Chapter 4. In Chapter 5 I discuss the absence of quantum tunneling marks in laser-desorbed guanine-cytosine (GC) pair IR-UV hole burning spectrum.[6] Indeed, GC pair can undergo a tautomerization *via* double proton hopping that may be the origin of punctual DNA mutations. A tunneling mechanism has been proposed for this tautomerization, yet it remained unconfirmed.[55–57] The 29-atom GC pair was simulated using DCSCIVR on DFT-D/B3LYP def2-TZVP as to have the best possible description of the hydrogen bond network. I then go back to anharmonicities and quantum effects in Chapter 6, where I describe a comparison between classical and semiclassical methods on H_2CO. Finally, Chapter 7 contains a summary of the results and some future perspectives.

Chapter 2

Theory

2.1 Harmonic Frequencies

In a Born-Oppenheimer (BO) framework, the vibrational motion of a molecule can be described using classical mechanics. Under this assumption, a molecule can be identified as a natural Lagrangian system (NLS), *i.e.* a system whose Lagrangian function assumes the form $L = T - V$, with T (kinetic energy) quadratic form of $\dot{q}$ and V (potential energy) uniquely function of the positions q. It can be demonstrated that the Lagrangian equations of a NLS take the form

$$\mathbb{A}\ddot{\mathbf{q}} = G(\mathbf{q}, \dot{\mathbf{q}}) - \frac{\partial V}{\partial \mathbf{q}} \tag{2.1}$$

where $\mathbb{A}$ is the kinetic matrix of the quadratic form T and G is quadratic function in $\dot{\mathbf{q}}$; in particular, $\forall \lambda \geq 0$:

$$G(\mathbf{q}, \lambda\dot{\mathbf{q}}) = \lambda^2 G(\mathbf{q}, \dot{\mathbf{q}}) \tag{2.2}$$

From multiplying Equation (2.1) by $\mathbb{A}^{-1}$, the Lagrangian equation are expressed in normal form:

$$\ddot{\mathbf{q}} = \mathbb{A}^{-1}\left(G(\mathbf{q}, \dot{\mathbf{q}}) - \frac{\partial V}{\partial \mathbf{q}}\right) \tag{2.3}$$

In this form, it is easier to demonstrate that the equilibrium points of a NLS are given by the critical points of the potential. Indeed, we have to turn our equations of motion to a system of first order differential equations to have critical points. To do so, we introduce the variable $\mathbf{v} = \dot{\mathbf{q}}$, so that the critical points will have the

form $(\mathbf{q}(t), \mathbf{v}(t)) = (\mathbf{q}_0, 0), \forall t$. By substituting this expression into Equation (2.3), we get

$$0 = \dot{\mathbf{v}} = \mathbb{A}^{-1} \left(G(\mathbf{q}_0, 0) - \frac{\partial V}{\partial \mathbf{q}}(\mathbf{q_0}) \right) \tag{2.4}$$

By Equation (2.2), we have $G(\mathbf{q}_0, 0) = 0$, so that by multiplying everything by $\mathbb{A}$ we get the thesis:

$$\frac{\partial V}{\partial \mathbf{q}}(\mathbf{q}_0) = 0 \tag{2.5}$$

Around a generic equilibrium point $\mathbf{q}_0$ of a NLS, the Lagrangian equations can be linearized as

$$\mathbb{A}_0 \ddot{\mathbf{x}} + \mathbb{H}\mathbf{x} = 0 \tag{2.6}$$

where $\mathbf{x} = \mathbf{q} - \mathbf{q}_0$ is the displacement from equilibrium, $\mathbb{A}_0 := \mathbb{A}(\mathbf{q}_0)$ is the kinetic matrix in $\mathbf{q}_0$, and $\mathbb{H}$ is the Hessian matrix of the potential with respect to the positions, evaluated in $\mathbf{q}_0$. To obtain such linearization, we take Equation (2.3) with $\mathbf{q} = \mathbf{q}_0 + \mathbf{x}$:

$$\ddot{\mathbf{x}} = \left[\mathbb{A}(\mathbf{q}_0 + \mathbf{x})\right]^{-1} \left\{ G(\mathbf{q}_0 + \mathbf{x}, \dot{\mathbf{x}}) - \frac{\partial V}{\partial \mathbf{q}}(\mathbf{q}_0 + \mathbf{x}) \right\} \tag{2.7}$$

and we expand to the first order in $(\mathbf{x}, \dot{\mathbf{x}})$. In the parenthesis, we have that $G(\mathbf{q}_0 + \mathbf{x}, \dot{\mathbf{x}}) = \mathcal{O}(\dot{x}^2)$, so that:

$$G(\mathbf{q}_0 + \mathbf{x}, \dot{\mathbf{x}}) - \frac{\partial V}{\partial \mathbf{q}}(\mathbf{q}_0 + \mathbf{x}) = -\frac{\partial V}{\partial \mathbf{q}}(\mathbf{q}_0) - \mathbb{H}\mathbf{x} + \mathcal{O}(x^2 + \dot{x}^2) \tag{2.8}$$

The inverse kinetic matrix is expanded as follows:

$$\left[\mathbb{A}(\mathbf{q}_0 + \mathbf{x})\right]^{-1} = \left[\mathbb{A}(\mathbf{q}_0)\right]^{-1} + \mathcal{O}(x) \tag{2.9}$$

In this way, the equation of motion is

$$\ddot{\mathbf{x}} = -\left[\mathbb{A}(\mathbf{q}_0)\right]^{-1} \mathbb{H}\mathbf{x} + \mathcal{O}(x^2) \tag{2.10}$$

$$\mathbb{A}_0 \ddot{\mathbf{x}} = -\mathbb{H}\mathbf{x} + \mathcal{O}(x^2) \tag{2.11}$$

We now move into finding a general solution for the system shown in Equation (2.6). We begin by observing that the equilibrium kinetic matrix $\mathbb{A}_0$ is symmetrical and positive-definite, whereas the equilibrium Hessian of the potential $\mathbb{H}$ is symmetrical. We also take the hypothesis that $\mathbf{q}_0$ is a stable equilibrium point, so $\mathbb{H}$ is semipositive-definite. We start by looking for autoparallel solutions, that is solutions with the form

$$\mathbf{x}(t) = s(t)\mathbf{u} \tag{2.12}$$

with $\mathbf{x}(t)$ general solution of the system, $\mathbf{u}$ time-independent vector and $s(t)$ time-dependent scalar. We substitute in Equation (2.6) to obtain

$$\ddot{s}\mathbb{A}\mathbf{u} = -s\mathbb{H}\mathbf{u} \tag{2.13}$$

For this equality to be verified, the vectors $\mathbb{A}\mathbf{u}$ and $\mathbb{H}\mathbf{u}$ must be at least parallel. That is, there must be a scalar λ so that

$$\lambda\mathbb{A}\mathbf{u} = \mathbb{H}\mathbf{u} \tag{2.14}$$

This is the eigenvalue equation of $\mathbb{H}$ with respect to $\mathbb{A}$. If $(\lambda, \mathbf{u})$ solves Equation (2.14), then it is a non-trivial solution of the system:

$$(\lambda\mathbb{A} - \mathbb{H})\mathbf{u} = 0 \tag{2.15}$$

This allows us to find the eigenvectors λ as

$$\det(\lambda\mathbb{A} - \mathbb{H}) = 0 \tag{2.16}$$

and, by substitution, the vectors $\mathbf{u}$. To find the particular solutions, we can substitute the eigenvalue and eigenvector couple $(\lambda, \mathbf{u})$ in Equation (2.13) and use the relationship of Equation (2.14), resulting in

$$(\ddot{s} + \lambda s)\mathbb{A}\mathbf{u} = 0 \tag{2.17}$$

Since $\mathbb{A}$ in non-degenerate, we have

$$\ddot{s} + \lambda s = 0 \tag{2.18}$$

In the case of $\lambda > 0$, we can define $\omega^2 := \lambda$ and recognize that the equation above is the equation of a harmonic oscillator with solution

$$s(t) = s_0 \cos(\omega t) + \frac{\dot{s}_0}{\omega} \sin(\omega t) \tag{2.19}$$

Thus, if a system is described by Equation (2.6), with $(\lambda, \mathbf{u})$ respectively eigenvalue and eigenvector of $\mathbb{H}$ with respect to $\mathbb{A}$; let it be $0 < \lambda =: \omega^2$, then

$$\mathbf{x}(t) = \left(s_0 \cos(\omega t) + \frac{\dot{s}_0}{\omega} \sin(\omega t) \right) \mathbf{u} \tag{2.20}$$

is a solution of the system.

The $\mathbf{u}$ vectors are generally known as normal modes of vibration. This result tells us that a molecule — like any NLS — behaves like a set of uncoupled harmonic oscillators in the neighborhood of the equilibrium points of its potential.

We can exploit this result to compute the harmonic frequencies and vibrational normal modes of a molecule. Indeed, we can turn Equation (2.14) into a proper eigenvalues equation:

$$\lambda \mathbf{u} = \mathbb{A}^{-1}\mathbb{H}\mathbf{u} \tag{2.21}$$

It is possible to demonstrate that for a molecule, the matrix $\mathbb{A}^{-1}\mathbb{H}$ corresponds to the mass-weighted Cartesian Hessian of the potential. By numerical diagonalization of this matrix, we can obtain the harmonic frequencies as square roots of the eigenvalues, and the normal modes of vibration as the eigenvectors.

Harmonic frequencies are easy to compute, even for large systems. However, they usually overestimate the exact vibrational frequency, since they cannot account for well anharmonicity and vibrational mode coupling. Moreover, harmonic frequency cannot include NQE, since the molecule is effectively treated as a set of uncoupled classical harmonic oscillators.

2.2 Classical Spectra

The major shortcomings of the harmonic approximation come from its local framework: in the minimum of the potential the system cannot experience any relevant anharmonicity of the potential and any coupling between vibrational modes. The system has to get out from the bottom of the well and sample a greater portion of the phase space. One way to do all of this is following the time evolution of system instead of just linearizing the equations of motion with the harmonic approximation. Therefore, we want a model capable of simulating the vibrational experiment. Any spectroscopic experiment can be modelled as a response process: the system at equilibrium is perturbed by the impinging light and by following its relaxation dynamics we can extract the vibrational frequencies. Thus, given the equilibrium Hamiltonian H_0 and the observable A, we are looking for an expression for the change of A if the system is perturbed, that can be later related to the vibrational spectroscopy experiment.

In a canonical ensemble, the ensemble average of the observable A is given by

$$\langle A \rangle = \frac{1}{Z} \int e^{-\beta H_0} A \, d\Gamma \tag{2.22}$$

where Z is the canonical partition function, $d\Gamma$ is the infinitesimal phase space hypervolume and β is the thermal constant $1/k_B T$. If the system is subject to an arbitrary perturbation λB, with λ arbitrarily small, then it will be described by the perturbed Hamiltonian

$$H' = H_0 + \lambda B \tag{2.23}$$

and the ensemble average of our observable will become

$$\langle A\rangle' = \frac{1}{Z'}\int e^{-\beta H_0-\beta\lambda B}A\,d\Gamma \tag{2.24}$$

Since for small λ, we can approximate

$$Z' = \int e^{-\beta(H_0+\beta\lambda B)}\,d\Gamma \simeq \int e^{-\beta H_0}\,d\Gamma = Z \tag{2.25}$$

and the integrand of the phase space integral, $e^{-\beta\lambda B}\sim 1-\beta\lambda B$

$$\langle A\rangle' = \langle A\rangle - \frac{\beta\lambda}{Z}\int e^{-\beta H_0}BA\,d\Gamma = \langle A\rangle + \langle \Delta A\rangle \tag{2.26}$$

where $\langle\Delta A\rangle$ is the system response in linear approximation. We can write it in terms of the perturbation B as

$$\langle \Delta A\rangle = -\frac{\beta\lambda}{Z}\int e^{-\beta H_0}BA\,d\Gamma = -\beta\lambda\,\langle BA\rangle \tag{2.27}$$

thus concluding that the change of the observable A when the equilibrium Hamiltonian H_0 is perturbed by an arbitrarily small perturbation can be described with the linear response approximation.

In an ergodic system, the time average of any observable A, $\overline{A}$, converges to the ensemble average and *vice versa*, thus

$$\langle \Delta A\rangle = \overline{\Delta A} = -\beta\lambda\overline{B(0)A(t)} \tag{2.28}$$

since the perturbation B is taken instantaneous at time $t=0$. The time average of B and A can be expressed as the cross-correlation function

$$C(t) = \overline{B(0)A(t)} = \lim_{T\to\infty}\frac{1}{T}\int_0^T \langle B(t_0)A(t_0+t)\rangle\,dt_0 \tag{2.29}$$

where $\langle B(t_0)A(t_0+t)\rangle$ is the average over all components. If $B=A$, $C(t)$ is more properly referred as an autocorrelation function. In conclusion, the Fourier transform of $C(t)$ informs us on the system oscillations after the perturbation.

Usually, the vibrational spectrum of a molecule is registered as an IR spectrum. The perturbation that allows to register an IR spectrum is given by the product between the dipole moment $\boldsymbol{\mu}$ and the electric field $\mathbf{E}$ carried by the excitation wave:

$$B = -\boldsymbol{\mu}\cdot\mathbf{E} \tag{2.30}$$

The related observable is the dipole moment itself, so

$$\langle\Delta A\rangle = \langle\Delta\mu\rangle \propto \langle\boldsymbol{\mu}_0\cdot\boldsymbol{\mu}(t)\rangle = C_{\mu\mu}(t) \tag{2.31}$$

In this way, we have related the vibrational spectrum to the response of the dipole moment to the perturbing electromagnetic wave, that can be computed as the Fourier transform of the autocorrelation function $C_{\mu\mu}(t)$:

$$S(\omega) \propto \omega \mathscr{F}\left[C_{\mu\mu}(t)\right](\omega) \tag{2.32}$$

In first approximation, $\boldsymbol{\mu}$ is linear with the normal mode position $\mathbf{q}$, so that the C_{qq} autocorrelation function can be employed instead of the more complicated $C_{\mu\mu}$. However, historically the velocity-velocity autocorrelation function — C_{vv} or C_{pp} — was used. In this case, the spectrum yields reliable vibrational frequencies as the peak maxima, but has no information regarding the absorption intensity:

$$I(\omega) \propto \omega \mathscr{F}\left[C_{pp}(t)\right](\omega) \tag{2.33}$$

This can be easily grasped in the case of a harmonic oscillator, with

$$p(t) = \dot{q}(t) = -\omega_0 q(0) \sin(\omega_0 t) \tag{2.34}$$

The Fourier transform of the C_{pp} of such oscillator features peaks at $\pm\omega_0$, *i.e.*

$$I(\omega) = C_1 \delta(\omega - \omega_0) + C_2 \delta(\omega + \omega_0) \tag{2.35}$$

where C_1 and C_2 are coefficients independent of ω, and $\delta(\omega)$ is the Dirac Delta centered on $\omega = 0$.

Proof. We start by inserting Equation (2.34) into

$$I(\omega) = \lim_{T\to+\infty} \frac{1}{T} \int_{-\infty}^{+\infty} e^{i\omega t} \left[\int_0^T p(t_0) p(t_0 + t)\, dt_0 \right] dt \tag{2.36}$$

$$= \lim_{T\to+\infty} \frac{1}{T} \int_{-\infty}^{+\infty} e^{i\omega t} \left[\int_0^T q_0^2 \omega_0^2 \sin(\omega_0 t_0) \sin(\omega_0 t_0 + \omega_0 t)\, dt_0 \right] dt \tag{2.37}$$

Using the identity

$$\sin(\omega_0 t_0) \sin(\omega_0 t_0 + \omega_0 t) = \frac{1}{2} \left[\cos(\omega_0 t) - \cos(2\omega_0 t_0 + \omega_0 t)\right] \tag{2.38}$$

we reach

$$I(\omega) = \lim_{T\to+\infty} \frac{q_0^2 \omega_0^2}{2T} \int_{-\infty}^{+\infty} e^{i\omega t} \left\{ \int_0^T \left[\cos(\omega_0 t) - \cos(2\omega_0 t_0 + \omega_0 t)\right] dt_0 \right\} dt \tag{2.39}$$

$$= \lim_{T\to+\infty} \frac{q_0^2 \omega_0^2}{2T} \left(\mathfrak{A} - \mathfrak{B}\right) \tag{2.40}$$

where

$$\mathfrak{A} = \int_{-\infty}^{+\infty} e^{i\omega t} \int_0^T \cos(\omega_0 t)\, dt_0 dt = \int_{-\infty}^{+\infty} e^{i\omega t} \cos(\omega_0 t) T\, dt \tag{2.41}$$

$$= \frac{T}{2} \left[\delta(\omega + \omega_0) + \delta(\omega - \omega_0)\right] \tag{2.42}$$

and

$$\mathfrak{B} = \int_{-\infty}^{+\infty} e^{i\omega t} \int_0^T \cos(2\omega_0 t_0 + \omega_0 t)\, dt_0 dt \tag{2.43}$$

$$= -\int_{-\infty}^{+\infty} \frac{e^{i\omega t}}{2\omega_0} \left\{\sin(2\omega_0 T + \omega_0 t) - \sin(\omega_0 t)\right\}\, dt \tag{2.44}$$

By means of Euler's relation, we can split $\mathfrak{B}$ into four integrals:

$$\mathfrak{B} = -\frac{1}{4i\omega_0} \left[\int_{-\infty}^{+\infty} e^{i\omega t} e^{2i\omega_0 T} e^{i\omega_0 t}\, dt - \int_{-\infty}^{+\infty} e^{i\omega t} e^{-2i\omega_0 T} e^{-i\omega_0 t}\, dt + \right.$$
$$\left. - \int_{-\infty}^{+\infty} e^{i\omega t} e^{i\omega_0 t}\, dt + \int_{-\infty}^{+\infty} e^{i\omega t} e^{-i\omega_0 t}\, dt \right] \tag{2.45}$$

$$= -\frac{1}{4i\omega_0} \left[e^{2i\omega_0 T}\delta(\omega_0 + \omega) - e^{-2i\omega_0 T}\delta(\omega_0 - \omega) - \delta(\omega_0 + \omega) + \delta(\omega_0 - \omega)\right] \tag{2.46}$$

In the end,

$$C_1 = \lim_{T\to+\infty} \left\{ \frac{q_0^2\omega_0^2}{4} - \frac{q_0^2\omega_0^2(e^{-2i\omega_0 T} + 1)}{8i\omega_0 T} \right\} \tag{2.47}$$

$$C_2 = \lim_{T\to+\infty} \left\{ \frac{q_0^2\omega_0^2}{4} + \frac{q_0^2\omega_0^2(e^{+2i\omega_0 T} - 1)}{8i\omega_0 T} \right\} \tag{2.48}$$

□

To simplify the numerical computation, a time-averaging (TA) procedure can be employed. As before, we start with the spectrum expressed as the Fourier transform of the time average of the C_{pp} function:

$$I(\omega) = \lim_{T\to+\infty} \frac{1}{T} \int_{-\infty}^{+\infty} e^{i\omega t} \int_0^T p(t_0) p(t_0 + t)\, dt_0 dt \tag{2.49}$$

The conjugate momentum $p(t_0)$ can be expressed as the antitransform of its Fourier transform $\tilde{p}(\omega)$:

$$p(t_0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\nu t_0} \tilde{p}(\nu)\, d\nu \tag{2.50}$$

the same can be done with $p(t+t_0)$:

$$p(t+t_0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\sigma(t+t_0)}\tilde{p}(\sigma)\,d\sigma \tag{2.51}$$

As *per* Fourier theorem:

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{ixt}\,dt \tag{2.52}$$

With these relations, with can rearrange $I(\omega)$ into:

$$I(\omega) = \lim_{T\to+\infty}\frac{1}{T}\int_{-\infty}^{+\infty} dt\, e^{i\omega t}\int_0^T \frac{dt_0}{4\pi^2}\int_{-\infty}^{+\infty} d\nu\, e^{-i\nu t_0}\tilde{p}(\nu)\int_{-\infty}^{+\infty} d\sigma\, e^{-i\sigma(t_0+t)}\tilde{p}(\sigma) \tag{2.53}$$

$$= \lim_{T\to+\infty}\frac{1}{4\pi^2 T}\int_0^T dt_0\int_{-\infty}^{+\infty} d\nu\, e^{-i\nu t_0}\tilde{p}(\nu)\int_{-\infty}^{+\infty} d\sigma\, e^{-i\sigma t_0}\tilde{p}(\sigma)2\pi\delta(\omega-\sigma) \tag{2.54}$$

$$= \lim_{T\to+\infty}\frac{1}{2\pi T}\int_0^T dt_0\int_{-\infty}^{+\infty} d\nu\, e^{-i\nu t_0}\tilde{p}(\nu)e^{-i\omega t_0}\tilde{p}(\omega) \tag{2.55}$$

$$= \lim_{T\to+\infty}\frac{1}{2\pi T}\int_{-\infty}^{+\infty} d\nu\,\tilde{p}(\nu)\tilde{p}(\omega)\frac{1}{2}\int_{-T}^{+T} dt_0\, e^{i(-\nu-\omega)t_0} \tag{2.56}$$

$$= \lim_{T\to+\infty}\frac{1}{2\pi T}\int_{-\infty}^{+\infty} d\nu\,\tilde{p}(\nu)\tilde{p}(\omega)\frac{2\pi}{2}\delta(\omega+\nu) \tag{2.57}$$

$$= \lim_{T\to+\infty}\frac{1}{2T}\tilde{p}(-\omega)\tilde{p}(\omega) \tag{2.58}$$

Since $p\in\mathbb{R}$, $\tilde{p}(-\omega)=\tilde{p}^*$. This allows us to reach the final form of the momentum-momentum power spectrum:[58]

$$I(\omega) = \lim_{T\to+\infty}\frac{1}{2T}|\tilde{p}(\omega)|^2 = \lim_{T\to+\infty}\frac{1}{2T}\left|\int_0^T dt\, p(t)e^{i\omega t}\right|^2 \tag{2.59}$$

To deal with chemical systems of even large dimensionality one can treat the molecule nuclei classically, thus treating the molecule as a set of anharmonic, coupled oscillators. In this case, the C_{pp} or sometimes C_{qq} functions are preferred, because both positions and momenta are already computed during the trajectory.

However, classical trajectories lack any ability of including quantum effects, even the most basic one such as ZPE, which is the energy a quantum system possesses even in its ground state, and that a classical system at rest does not have. For this reason, it is common practice to run the classical trajectories at

an energy close to that of the system ZPE. Since the exact quantum ZPE is not readily available, the trajectories are initialized in the torus corresponding to the harmonic ZPE of the system. This method is known as quasiclassical trajectory (QCT) or quasiclassical molecular dynamics (QCMD) and it is a reliable and affordable way to estimate vibrational fundamentals when no NQE are involved.[19, 59] Indeed, QCT is limited by its classical formulation and cannot be used to extract actual anharmonic ZPE values or to investigate system heavily affected by quantum effects.[2, 43, 60, 61]

2.3 Semiclassical Initial Value Representation

To include NQE, we can compute the time-dependent vibrational density of states (VDOS) $I(E)$, as the Fourier transform of the survival amplitude $\langle\Psi_0|\Psi_t\rangle$:[29]

$$I(E) = \frac{1}{2\pi\hbar}\int_{-\infty}^{+\infty} e^{iEt/\hbar}\langle\Psi_0|\Psi_t\rangle = \sum_j |c_j|^2\delta(E-E_j) \tag{2.60}$$

where E is the absolute energy, E_j is the j-th eigenvalue of the system Hamiltonian and c_j is the projection of Ψ_t on the j-th eigenfunction of system Hamiltonian. If $|\Psi_t\rangle$ were evolved exactly, $I(E)$ would be a collection of delta functions centered at the exact vibrational eigenvalues.

A way to propagate $|\Psi_t\rangle$ is the path integral formulation of quantum dynamics. $|\Psi_t\rangle$ is projected on the positions $\mathbf{q}$ and then — by introducing an identity in the initial positions $\mathbf{q}_0$ — the path integral kernel $K(\mathbf{q},\mathbf{q}_0,t)$ can be isolated:

$$\langle\mathbf{q}|\Psi_t\rangle = \int_{-\infty}^{+\infty} d\mathbf{q}_0\,\langle\mathbf{q}|e^{-\frac{i}{\hbar}\hat{H}t}|\mathbf{q}_0\rangle\langle\mathbf{q}_0|\Psi_0\rangle = \int_{-\infty}^{+\infty} d\mathbf{q}_0\,K(\mathbf{q},\mathbf{q}_0,t)\langle\mathbf{q}_0|\Psi_0\rangle \tag{2.61}$$

where the kernel is expressed with the usual path integral formulation

$$\langle\mathbf{q}|e^{-\frac{i}{\hbar}\hat{H}t}|\mathbf{q}_0\rangle = K(\mathbf{q},\mathbf{q}_0,t) = \int \mathcal{D}\left[\mathbf{q}_t\right] e^{\frac{i}{\hbar}S_t} \tag{2.62}$$

where S_t is the action of the path and the differential $\mathcal{D}\left[\mathbf{q}_t\right]$ means that the integral is evaluated on all possible path between $\mathbf{q}_0$ and $\mathbf{q}_t$. The path integral evolution is based on trajectories, opening the way to approximations based on classical mechanics. Indeed, it can be demonstrated that a path integral of a system with quadratic potential is exact even if only classical trajectories (and not *all* trajectories) are considered.[62] However, the complex exponential is not a good distribution for Monte Carlo importance sampling, and its oscillations hinders the numerical convergence of the integral.

These numerical issues are addressed by introducing a stationary phase approximation. The oscillating terms of the complex exponential tend to cancel out, but when the derivative of the phase S_t is near zero, the oscillations are much more sparse and constructive interference is much more frequent. This justify expanding S_t around its stationary points and truncating the expansion at the second order. By adopting the stationary phase approximation on S_t, we impose the principle of least action on the paths. This effectively limits the integration on classical trajectories, making the kernel more manageable:

$$K(\mathbf{q}, \mathbf{q}_0, t) = \sum_{\{n|S'_n=0\}}^{\infty} \sqrt{\left(\frac{m}{\Delta t}\right)^n \frac{1}{(2\pi i\hbar)^F \det(\mathbb{H}_S)}} e^{\frac{i}{\hbar}S_n} \tag{2.63}$$

where m is the mass, Δt is the evolution time-step, F is the number of degrees of freedom and $\mathbb{H}_S$ is the second derivative matrix of the action S with respect to the positions $\mathbf{q}$.

Since now the integration is limited to classical trajectories, the kernel can be manipulated using classical identities to obtain Van Vleck's propagator:[63]

$$K_{\text{vv}}(\mathbf{q}, \mathbf{q}_0, t) = \sum_{S'_t=0} \sqrt{\frac{1}{(2\pi i\hbar)^F} \left|\frac{\partial \mathbf{q}}{\partial \mathbf{p}_0}\right|^{-1}} e^{\frac{i}{\hbar}S_t} \tag{2.64}$$

where $\mathbf{p}_0$ are the initial momenta. Van Vleck's propagator was obtained from Schrödinger equation before the path integral formulation was established. The propagator adds quantum effects on classical trajectories, but diverges every time $\partial\mathbf{q}/\partial\mathbf{p_0} = 0$, and it is formulated as a double boundary problem. This means that each trajectory must start at $\mathbf{q}_0$ and end at $\mathbf{q}$ without crossing any caustic point in which the propagator diverges. This make Van Vleck's propagator extremely difficult to evaluate, even if a modification by M. C. Gutzwiller using Maslov's indexes allows the Van Vleck's propagator to cross caustic points.[64]

However, to best tackle both issues, the initial value representation (IVR) was introduced by W.H. Miller.[28, 65, 66] Indeed, the Liouville theorem allows the substitution:

$$\sum_{S'_t=0} \int d\mathbf{q} \;\rightarrow\; \int d\mathbf{p}_0 \left|\frac{\partial \mathbf{q}}{\partial \mathbf{p}_0}\right| \tag{2.65}$$

that brings us to the IVR formulation of the survival amplitude:

$$\langle\Psi_0|\Psi_t\rangle = \int d\mathbf{q}\,\langle\Psi_0|\mathbf{q}\rangle\langle\mathbf{q}|\Psi_t\rangle = \int d\mathbf{q}\,\langle\Psi_0|\mathbf{q}\rangle \int d\mathbf{q}_0\,K_{\mathrm{vv}}(\mathbf{q},\mathbf{q}_0,t)\langle\mathbf{q}_0|\Psi_0\rangle \quad (2.66)$$

$$= \sum_{S_t'=0} \iint d\mathbf{q}d\mathbf{q}_0\,\langle\Psi_0|\mathbf{q}\rangle \sqrt{\frac{1}{(2\pi i\hbar)^F}\left|\frac{\partial\mathbf{q}}{\partial\mathbf{p}_0}\right|^{-1}} e^{\frac{i}{\hbar}S_t}\langle\mathbf{q}_0|\Psi_0\rangle \quad (2.67)$$

$$= \iint d\mathbf{p}_0 d\mathbf{q}_0\,\langle\Psi_0|\mathbf{q}\rangle \sqrt{\frac{1}{(2\pi i\hbar)^F}\left|\frac{\partial\mathbf{q}}{\partial\mathbf{p}_0}\right|} e^{\frac{i}{\hbar}S_t}\langle\mathbf{q}_0|\Psi_0\rangle \quad (2.68)$$

There are several ways to obtain the Herman-Heller-Kluk-Kay (HHKK) propagator starting from Van Vleck's.[67, 68] Following Miller,[69, 70] we can introduce a Filinov filter, thus obtaining the HHKK propagator from the Van Vleck one, expressed in IVR:[30, 71–73]

$$\langle\Psi_0|\Psi_t\rangle = \frac{1}{(2\pi\hbar)^F}\iint d\mathbf{p}_0 d\mathbf{q}_0\,C_{\mathrm{hhkk}} e^{\frac{i}{\hbar}S_t}\langle\Psi_0|\mathbf{p}_t\mathbf{q}_t\rangle\langle\mathbf{p}_0\mathbf{q}_0|\Psi_0\rangle \quad (2.69)$$

$$C_{\mathrm{hhkk}} = \sqrt{\left|\frac{1}{2}\left(\frac{\partial\mathbf{p}_t}{\partial\mathbf{p}_0} + \frac{\partial\mathbf{q}_t}{\partial\mathbf{q}_0} - i\Gamma\hbar\frac{\partial\mathbf{q}_t}{\partial\mathbf{p}_0} + \frac{i}{\Gamma\hbar}\frac{\partial\mathbf{p}_t}{\partial\mathbf{q}_0}\right)\right|} \quad (2.70)$$

The C_{hhkk} prefactor includes the approximate quantum effects upon the classical dynamics (for instance it is known to be responsible for the estimate of the ZPE), but it is also the main bottleneck of semiclassical calculations. Indeed, the partial derivatives present in C_{hhkk} are the elements of the stability matrix of the trajectory, $\mathbb{M}$. The stability matrix measures the chaotic behaviour of the system: *per* Liouville's theorem its determinant should be unitary, but due to small inaccuracies in the numerical propagation, in chaotic systems one gets a value of the determinant of $\mathbb{M}$ which diverges with time. When the prefactor diverges above a chosen and usually small threshold, the trajectory should be discarded or taken into consideration only until the stability test is passed, since the Fourier transform is not able to collect any meaningful signal.[42, 74] Thus, the evolution of $\mathbb{M}$ is both necessary and critical for semiclassical dynamics, and it requires the calculation of the Hessian matrix of the potential at each step of the trajectory. This is the most time-consuming task in computing the semiclassical VDOS, even if the overall effort can be reduced by exploiting the periodicity of vibrational trajectory and building a database of Hessian matrices. An early implementation built the database by following the trajectory in chronological order, but a new version uses a neural gas machine learning approach to select the most representative point in which compute the Hessian. This method goes under the name of hessian database (HDB).[75, 76] For certain cases, it has been demonstrated that even using a single Hessian does not hinder the accuracy of

the results.[77] The main issue with HHKK is the slow numerical convergence of the phase-space integral using Monte Carlo. This means that a lot of trajectories must be evolved to obtain an accurate VDOS, not including the ones discarded due to chaotic behaviour. Some techniques were devised to enhance the phase-space sampling, such as weighting the contribution of the discarded trajectories, approximating the prefactor or improving the initial conditions distribution.[42, 53, 74, 78–82]

The convergence of the HHKK phase-space integral is greatly improved by introducing a TA formulation like that shown in Equations (2.53) to (2.58). It is now convenient to introduce the separability approximation on the prefactor, that is exact for harmonic systems.[83, 84] This allows us to rearrange everything in a square module:

$$I(E) = \frac{1}{(2\pi\hbar)^F} \iint d\mathbf{p}_0 d\mathbf{q}_0 \, \frac{1}{2\pi T\hbar} \left| \int_0^T dt \, e^{\frac{i}{\hbar}(S_t+\phi_t+Et)} \langle \Psi | \mathbf{p}_t \mathbf{q}_t \rangle \right|^2 \tag{2.71}$$

with F number of vibrational degrees of freedom and ϕ_t phase of the prefactor C_{hhkk}. This formulation — known as time-averaged semiclassical initial value representation (TASCIVR) — has two main advantages: first, the phase-space integrand is always positive, greatly facilitating Monte Carlo integration; second, it is integrable *via* Monte Carlo Importance Sampling, using as variable distribution the Husimi distribution given by:

$$|\langle \Psi | \mathbf{p}_t \mathbf{q}_t \rangle|^2 = \prod_{j=1}^{F} \exp\left\{ -\frac{\Gamma_{jj}(q_{0,j} - q_j)^2}{2} - \frac{(p_{0,j} - p_j)^2}{2\Gamma_{jj}\hbar^2} \right\} \tag{2.72}$$

where Γ is generally a diagonal matrix with each of its non-zero elements chosen equal to ω_j, harmonic frequency of the j-th mode. The use of the Husimi distribution is justified by the fact that the vibrational power spectrum is quadratic in the wavefunction.[82] The distribution is centered around

$$\begin{cases} p_{0,j} = \sqrt{\hbar\omega_j(2n+1)} \\ q_{0,j} = q_{\mathrm{eq},j} \end{cases} \tag{2.73}$$

in which n is the vibrational quantum number of the desired state. These initial conditions correspond to a harmonic oscillator of frequency ω_j and energy equivalent to the harmonic quantum level.

2.4 Multiple Coherent and Divide-and-Conquer Semiclassical Initial Value Representation

As is, TASCIVR power spectrum is still limited to systems with an available PES, since thousands of trajectories are still needed for the integration to converge. This makes TASCIVR unfeasible for systems of chemical — or even biological — interest. However, Heller and De Leon demonstrated that a single semiclassical trajectory can reproduce the quantum system, as long as its energy is the exact quantum one.[85] Since the exact quantum energy is unknown, in first approximation the single trajectory can be run starting from the conditions shown in Equation (2.73), so that the system would run on the harmonic energy of the desired state. In addition, a combination of coherent states of the form

$$|\Psi_{\text{ref}}\rangle = \prod_{j=1}^{F} \left(|\mathbf{p}_0\mathbf{q}_0\rangle + \epsilon_j|-\mathbf{p}_0\mathbf{q}_0\rangle\right) \tag{2.74}$$

is used as reference state. The choice of the parameter ϵ_j allows the selection of the different vibrational states based on their symmetry, by enhancing the relative signal. For example, setting all the $\epsilon_j = 1$ will select the state with an even total quantum number, such as the ZPE state and the even overtones. If instead one $\epsilon_j = -1$, the states with an odd number of vibrational quanta in the j-th mode will be enhanced, for instance the fundamental frequencies of vibration. This gives us the multiple coherent semiclassical initial value representation (MCSCIVR) final formula:

$$I(E) = \left(\frac{1}{2\pi\hbar}\right)^F \frac{1}{2\pi T\hbar} \left| \int_0^T dt\, e^{\frac{i}{\hbar}[S_t(\mathbf{p}_0,\mathbf{q}_0)+\phi_t(\mathbf{p}_0,\mathbf{q}_0)+Et]} \langle \Psi_{\text{ref}}|\mathbf{p}_t,\mathbf{q}_t\rangle \right|^2 \tag{2.75}$$

This formula made on-the-fly semiclassical dynamics possible, opening the way to the simulation of the vibrational spectra of small molecules of biological interest, such as amino acids.[38, 86, 87]

However, as the system grows in dimensions, the curse of dimensionality sets in. This is a well known phenomenon in statistics, that affects the probability distributions with a large number of stochastic variables. In the case of semiclassical dynamics, as the system dimensionality grows it is more and more likely that the full dimensional overlap $\langle\Psi|\mathbf{p}_t\mathbf{q}_t\rangle$ will be close to zero, because it is less and less likely that during a trajectory the system will go back *exactly* to the reference state. To exorcise the curse of dimensionality, it is possible to obtain the semiclassical spectrum of a subspace of the full vibrational space, using a technique known

as divide-and-conquer semiclassical initial value representation (DCSCIVR).[39, 41] In DCSCIVR, the vibrational modes are grouped together based on their coupling. A quick grouping criterion is using the off-diagonal elements of the average Hessian matrix along the trajectory,[39] while more advanced criteria are based on the subspace Jacobian matrix.[40, 43] When the subspaces have been identified, all the elements of Equation (2.75) can be straightforwardly evaluated in each subspace, with the exception of the potential V included in the classical action S. In this case, an effective potential $\tilde{V}$ is employed:

$$\tilde{V}(\tilde{\mathbf{q}}_M) = V(\tilde{\mathbf{q}}_M; \mathbf{q}^{\mathrm{eq}}_{F-M}) + \lambda \tag{2.76}$$

where M is the subspace dimension, $V(\tilde{\mathbf{q}}_M; \mathbf{q}^{\mathrm{eq}}_{F-M})$ is the potential of the system with all but the M modes at equilibrium, and λ is an artificial external field for the other terms:

$$\lambda = V(\tilde{\mathbf{q}}_M; \mathbf{q}_{F-M}) - \left[V(\tilde{\mathbf{q}}_M; \mathbf{q}^{\mathrm{eq}}_{F-M}) + V(\tilde{\mathbf{q}}^{\mathrm{eq}}_M; \mathbf{q}_{F-M})\right] \tag{2.77}$$

The subspace is now effectively treated as a normal system, but with a reduced dimensionality, so less affected by the curse. The global spectrum can be reconstructed by collecting the spectra of all subspaces. When coupled with MCSCIVR, DCSCIVR is capable of pushing the semiclassical method to large biological systems such as peptides, nucleosides and supramolecular clusters.[41, 61, 75, 88, 89]

2.5 Action-Angle Variables

Let us consider a harmonic oscillator, described by the Hamiltonian H:

$$H = \frac{p^2}{2m} + \frac{k}{2}q^2 \tag{2.78}$$

where p is the conjugate momentum of the position q, m is the oscillator mass and $k = m\omega^2$ is the spring stiffness, or the strength of the oscillator. We want to find a canonical transformation to map this Hamiltonian into spherical coordinates.

Definition (Canonical Transformation)**.** Let χ and γ be two coordinate systems. A coordinate transformation from χ to $\chi = \chi(\gamma)$ is canonical if for all Hamiltonians $H = H(\chi)$, it maps the solutions $s(\chi, t)$ of H Hamilton's equations in the solutions $v(\gamma, t)$ of the Hamilton's equations of $K(\gamma) := H(\chi(\gamma))$, so that $v(\gamma, t) = s(\chi(\gamma), t)$.

In practice, a transformation is canonical if the fact that $v(\gamma, t)$ solves the Hamilton's equations of K implies that $s(\chi(\gamma), t)$ solves the Hamilton's equations

of H, and *viceversa*. For ease of calculation, we group together p and q variables using the Hamiltonian vector field formalism. We define:

$$\mathbf{x} := (p_1, \dots, p_F, q_1, \dots, q_F)^T \tag{2.79}$$

$$\frac{\partial H}{\partial \mathbf{x}} := \left(\frac{\partial H}{\partial x_1}, \dots, \frac{\partial H}{\partial x_{2F}} \right)^T \tag{2.80}$$

By defining the $(2F, 2F)$ matrix $\mathbb{E}$

$$\mathbb{E} = \begin{bmatrix} 0 & -\mathbb{I} \\ \mathbb{I} & 0 \end{bmatrix} \tag{2.81}$$

(where $\mathbb{I}$ is the identity matrix), Hamilton's equations take the form:

$$\dot{\mathbf{x}} = \mathbb{E} \frac{\partial H}{\partial \mathbf{x}}(\mathbf{x}) = \mathbb{X}_H(\mathbf{x}) \tag{2.82}$$

where $\mathbb{X}_H$ is the Hamiltonian vector field. Then, we can employ the following theorem to construct a canonical transformation.

Theorem. *A coordinate transformation $x = x(y)$ is canonical if and only if it satisfies the condition:*

$$\sum_{lk} \frac{\partial y_a}{\partial x_l} E_{lk} \frac{\partial y_b}{\partial x_k} = E_{ab} \tag{2.83}$$

where E_{lk} and E_{ab} are elements of the $\mathbb{E}$ matrix.

We can furtherly simplify Equation (2.83) by introducing Poisson bracket

Definition (Poisson braket)**.** The quantity

$$\{F, H\} := \sum_k \left(\frac{\partial F}{\partial q_k} \frac{\partial H}{\partial p_k} - \frac{\partial F}{\partial p_k} \frac{\partial H}{\partial q_k} \right) \tag{2.84}$$

is the Poisson bracket of F and H.

This allows us to express the F time-derivative in a more concise manner:

$$\dot{F} = \{F; H\} \tag{2.85}$$

In a Hamiltonian vector field,

$$\{F, H\} = \dot{F}(\mathbf{x}(t)) = \sum_h \frac{\partial F}{\partial x_h} \dot{x}_h = \sum_h \frac{\partial F}{\partial x_h} \sum_k E_{hk} \frac{\partial H}{\partial x_k} \tag{2.86}$$

$$= \sum_{h,k} \frac{\partial F}{\partial x_h} E_{hk} \frac{\partial H}{\partial x_k} \tag{2.87}$$

so that Equation (2.83) can be written as:

$$\{y_a, y_b\} = E_{ab} \tag{2.88}$$

This form is called conservation of fundamental Poisson bracket.

Corollary. *A coordinate transformation is canonical if and only if it conserves the fundamental Poisson bracket.*

Using this corollary, we can demonstrate that the following transformations are canonical: for $\alpha_i > 0$,

$$Q_i = \alpha_i q_i \qquad\qquad P_i = \frac{1}{\alpha_i} p_i \tag{2.89}$$

and

$$q = \sqrt{2P}\sin Q \qquad\qquad p = \sqrt{2P}\cos Q \tag{2.90}$$

The harmonic oscillator Hamiltonian of Equation (2.13) suggests the canonical transformation

$$\tilde{p} = \frac{p}{\sqrt[4]{mk}} \qquad\qquad \tilde{q} = \sqrt[4]{mk}q \tag{2.91}$$

that reduces the Hamiltonian to

$$H = \frac{\omega}{2}(\tilde{p}^2 + \tilde{q}^2) \tag{2.92}$$

At this point, we can get to polar-like coordinates by using the canonical transformation shown in Equation (2.90). In these variables, the Hamiltonian reduces to

$$H = \omega P \tag{2.93}$$

Since the harmonic oscillator phase space has been mapped into an circle, the Hamilton's equations assume the trivial form of

$$\dot{Q} = \omega \qquad\qquad \dot{P} = 0 \tag{2.94}$$

These coordinates are known as action-angle variables for the harmonic oscillator and are very useful in treating system close to their equilibrium points. Every integrable system has access to its own action-angle variables. Finally, action-angle variables are fundamental in the theory of the adiabatic invariant and in the quantization of simple quantum systems, since it can be demonstrated that the action variable P is the adiabatic invariant $I = P \propto \Delta S$, where S is the classical action.[90]

2.6 Einstein-Brillouin-Keller quantization

In quantum theory, the motion of a system is described *via* classical mechanics, with the restriction that certain constants of motion have to be integers. The restrictions are called quantum conditions and the integers quantum numbers. Quantum theory was replaced by quantum mechanics, but it remains crucial in the semiclassical treatment of the quantum problem, since it provides a bridge between classical and quantum mechanics. The core problem in quantum theory was finding a reliable method to decide between integer and half-integer quantum numbers.[91] The solution of this problem rested in the Einstein-Brillouin-Keller (EBK) quantization formula.

Following Brillouin,[92] we can obtain the quantum condition for an F dimensional system of Hamiltonian $H(\mathbf{q}, \mathbf{p}, t)$ by introducing the approximate eigenfunction

$$\Psi_0 = A(\mathbf{q}, t) \exp\left\{\frac{i}{\hbar} S(\mathbf{q}, t)\right\} \tag{2.95}$$

Where A is the wave amplitude and S is the action. Since the exact eigenfunction Ψ is single valued, so must be Ψ_0. If S is multiple valued and ΔS is the difference between two of its values, it is necessary that

$$\Delta S = nh \qquad n \in \mathbb{N} \tag{2.96}$$

so that Ψ_0 is single valued even if S is not. Brillouin's conclusion is equivalent to Einstein's generalization of Bohr-Sommerfeld-Wilson conditions to nonseparable systems.[91, 93, 94] This condition only yields integer quantum numbers, but it is based on the assumption that the amplitude A is single valued. If this is not the case, the quantum condition becomes

$$\Delta S = h\left[n + \frac{i}{2\pi}\Delta \log(A)\right] \tag{2.97}$$

This can be easily demonstrated if we take ΔA as the difference between two values of A. Then:

$$\Delta S = hn + \frac{hi}{2\pi}\left(\Delta \log A\right) = hn + \frac{hi}{2\pi}\left[\log(A + \Delta A) - \log(A)\right] \tag{2.98}$$

$$= hn + \frac{hi}{2\pi} \log\left(\frac{A + \Delta A}{A}\right) = hn + \frac{hi}{2\pi} \log\left(1 + \frac{\Delta A}{A}\right) \tag{2.99}$$

We can now consider Ψ_0', another approximate eigenfunction with amplitude $A + \Delta A$ and action $S + \Delta S$, with ΔS given by Equation (2.97) and subject to Brillouin's quantum condition. By substitution, we find that it is now equivalent

to Ψ_0:

$$\Psi_0' = (A + \Delta A) \exp\left\{\frac{i}{\hbar} S + 2\pi i n - \log\left(1 + \frac{\Delta A}{A}\right)\right\} \tag{2.100}$$

$$= (A + \Delta A) \exp\left\{\frac{i}{\hbar} S\right\} \exp\left\{\log\left(\frac{A}{A + \Delta A}\right)\right\} \tag{2.101}$$

$$= (A + \Delta A) \exp\left\{\frac{i}{\hbar} S\right\} \frac{A}{A + \Delta A} = \Psi_0 \tag{2.102}$$

If, for example, two values of A only differ in sign, we would have

$$\Delta \log A = \log(-A) - \log(A) \tag{2.103}$$

$$= \log|-A| + i \arg(-A) - \log|A| - i \arg(A) \tag{2.104}$$

$$= i \arg(-A) - i \arg(-A) - i\pi = -i\pi \tag{2.105}$$

that yields

$$\Delta S = h\left[n + \frac{1}{2}\right] \tag{2.106}$$

that includes half-integer quantum numbers.

By using the action-angle Hamiltonian of Equation (2.93), we can obtain the analytical expression of the harmonic oscillator quantum energy E_{ho}:

$$E_{\text{ho}} = \omega P = \omega \Delta S = h\omega\left[n + \frac{1}{2}\right] \tag{2.107}$$

In conclusion, obtaining the quantized energy of a system using EBK is quite simple, when we have an action-angle Hamiltonian $H(S)$ and some insight on the topology of the potential. We start by defining a vector of quantum numbers $\mathbf{n}$, to be put into the lattice

$$\Delta \mathbf{S} = h\left(\mathbf{n} + \frac{\mathbf{m}}{4}\right) \tag{2.108}$$

where $\mathbf{m}$ is the vector made by the Maslov's indexes. The Maslov index is the total phase loss during one period, in $\pi/2$ units.[95] This formula is Keller's generalization of Brillouin's Equation (2.106). Keller's formulation considers also the phase delays of S and $\log A$ when crossing a phase space caustic point, hence the inclusion of the Maslov index in the $\mathbf{m}/4$ term.[91] We can the express each energy level by computing $E_{\mathbf{m}} = H(\Delta \mathbf{S})$.[93]

2.7 Adiabatic switching

Semiclassical dynamics of chaotic systems is quite cumbersome, since its foundations lie in classical mechanics and the quantum effects depend upon the mon-

odromy matrix $\mathbb{M}$. When using TASCIVR, this increases the number of trajectories required for the phase-space integration to converge, whereas for MCSCIVR it will decrease the effective length of the single trajectory.

The chaoticity issue may be alleviated by improving the sampling of initial conditions. As illustrated in Section 2.4, the best sampling would be giving each classical trajectory the exact (unknown) quantum energy.[85] Usually, a harmonic estimate is used instead, since it is readily available even for systems of considerable dimensionality. Indeed, a set of F harmonic oscillators constitutes an F-dimensional integrable systems, that is quantizable, as shown in Section 2.6. However, a molecule is not an integrable system and the harmonic energy typically overestimate the exact energy and the initial harmonic state is not an eigenstate of the molecular Hamiltonian. The harmonically-sampled trajectories may still be numerically unstable, hindering the accuracy and precision of the semiclassical results.

Adiabatic switching is a technique that was developed to obtain non-harmonic quantization for the initial conditions of QCT by exploiting the classical adiabatic theorem, that establishes the conditions to turn an integrable Hamiltonian into a non-integrable one using a classical trajectory.[49, 90, 96–99] AS has also been used to extract Wigner distributions and vibrational energies.[48, 52, 100, 101] An improved final sampling has been obtained by performing AS in Eckart frame, in order to obtain the ZPE and some vibrational modes of methane.[50] Furthermore, internal coordinates where used to reduce roto-vibrational couplings.[51]. However, AS cannot provide quantum eigenfuctions and its efficiency is reduced when treating systems with a very high vibrational density of states.[102–104]

The foundation of AS is the adiabatic theorem:[90]

Theorem (Adiabatic Theorem)**.** *In a Hamiltonian system described by $H(q,p;\lambda)$, with λ slowing varying with time in a period T,*

$$T\frac{d\lambda}{dt} \ll \lambda \tag{2.109}$$

the dependence of the energy E from λ can be expressed as the invariance of a combination of E and λ. This quantity is constant during the motion of a system with slowly-varying parameters, and it is called the adiabatic invariant *I:*

$$\overline{\frac{dI}{dt}} = 0 \qquad\qquad I \equiv \frac{1}{2\pi}\oint p\,dq = \frac{1}{2\pi}\Delta S \tag{2.110}$$

where q are the system positions and p are the conjugate momenta.

The canonical transformation shown in Section 2.5 is available also for the systems ruled by the adiabatic theorem. In this case, the new equations of motions

are

$$\dot{\zeta} = \frac{dE(I)}{dI} \qquad \dot{I} = 0 \tag{2.111}$$

It can be demonstrated that the oscillations of the adiabatic invariant during a trajectory are

$$\Delta I \sim \Re \int \frac{\dot{\lambda} i e^{i\zeta}}{\omega(I,\lambda)}\, d\zeta \tag{2.112}$$

so as long as $\lambda(t)$ and $1/\omega(t)$ have no singularities (as required by the adiabatic theorem) the oscillation of the invariant are exponentially small when the integral is evaluated using analytical continuation. By comparing Equation (2.110) with Equation (2.107), we can transpose the classical adiabatic theorem to the quantum realm, realizing that the quantum equivalent of I is none other then the quantum number n.

To apply AS to a molecular system, we have to write the switching Hamiltonian $H_{\text{as}}(q,p;\lambda)$ in a way that the slow-varying perturbation λ turns the initial harmonic Hamiltonian (H_{ho}) into the final exact one (H_{ex}), in a way that would allow us to insert the quantum Hamiltonian we can afford. A trivial form for H_{as} can be mutuated from the theory of thermodynamic integration:

$$H_{\text{as}} = [1-\lambda(t)]\, H_{\text{ho}} + \lambda(t) H_{\text{ex}} \tag{2.113}$$

with

$$\lambda(t) = \frac{t}{T_{\text{as}}} - \frac{1}{2\pi} \sin\left(\frac{2\pi t}{T_{\text{as}}}\right) \tag{2.114}$$

where T_{as} is the total length of the AS trajectory. This form of λ guarantees that the harmonic Hamiltonian will be turned into the exact one, the more *adiabatically* the longer the total time T_{as}.

Ideally, the AS evolution could be performed directly with action-angle variable, using the equations of motion:

$$\dot{\zeta} = \frac{dE}{dn} \qquad \dot{n} = 0 \tag{2.115}$$

However, it is more convenient to use coordinates and momenta built using action-angle variables:

$$p_0 = \sqrt{2I\omega}\sin(2\pi\zeta) = \sqrt{(2n+1)\hbar\omega}\sin(2\pi\zeta) \tag{2.116}$$

$$q_0 = \sqrt{\frac{2I}{\omega}}\cos(2\pi\zeta) = \sqrt{\frac{(2n+1)\hbar}{\omega}}\cos(2\pi\zeta) \tag{2.117}$$

where ω is the harmonic vibrational frequency and ζ the angle variable. In this way, AS is inserted seamlessly into the TASCIVR workflow.[53]

In the framework of semiclassical dynamics, AS is able to improve the energy estimate of the initial conditions, ideally reaching Heller and de Leon's conditions.[85] AS is used to obtain points of the phase space that reflect an approximate quantization of the actual molecular Hamiltonian, preserving the vibrational quanta of the starting harmonic state. These points can be then used as initial conditions for classical trajectories that are now quasi-periodic, at least in the short time-frame of the semiclassical methods.[53] As an example of AS capabilities, I reproduce Table 2.1 from Ref.[53], that includes the percentage of trajectories discarded during either a TASCIVR or an adiabatic switching semiclassical initial value representation (ASSCIVR) run for CH_4. Trajectories are discarded when the monodromy matrix determinant varies from one more than a chosen threshold σ.

TABLE 2.1: *Percentage of trajectory rejection in TASCIVR (TA) and ASSCIVR (AS) simulations of CH_4 (with total time* 0.60 ps *and* 1.21 ps*) for several rejection thresholds (σ). From Ref.[53]*

	0.60 ps		1.21 ps	
σ	TA (%)	AS (%)	TA (%)	AS (%)
1×10^{-2}	54.6	0.0	90.4	1.4
1×10^{-3}	61.7	0.0	94.0	5.3
1×10^{-4}	70.7	0.0	97.3	16.6
1×10^{-5}	81.1	0.1	99.0	42.2
1×10^{-6}	90.2	1.1	99.7	79.4

Chapter 3

Adiabatic Switching, on-the-fly

The content of this chapter is based on the work reported in Ref.[105]

3.1 Increasing the trajectory stability

Semiclassical (SC) spectroscopy has been employed on numerous biomolecular systems, including protonated and neutral species, supramolecular complexes and nucleosides.[39, 61, 88, 106–108] For this purpose the DCSCIVR technique has been adopted with satisfactory and accurate results allowing for an easier assignment of spectral features. However, DCSCIVR requires particular care in the partition of the normal modes into subspaces. This process increases the simulation cost, since it requires the computing of the projected potential of each subspace. This — and the loss of any full-dimensional insight — makes DCSCIVR a very expensive post-processing tool, the lesser of two evils.

To assign the spectra of such systems, narrower signals are desirable, so an application of ASSCIVR to very high dimensional systems is with no doubt a path worth being pursued. When performing semiclassical spectroscopy with multiple trajectories, AS strongly reduces the number of trajectories to be discarded, thus increasing the Monte Carlo integration efficiency, as shown in Table 2.1 of Section 2.7.[53] On the other hand, when running a single on-the-fly trajectory, the increased stability afforded by AS should lead to the collection of more significant signal. In both cases, ASSCIVR reduces the signal width and noise, thus increasing the spectrum accuracy. This is why I decided to interface ASSCIVR with *ab initio* on-the-fly molecular dynamics, thus proving the advantages that

such an approach brings in when dealing with biomolecular systems.

3.2 The on-the-fly implementation

I wrote a Fortran90 code to evolve a preparatory adiabatic switching trajectory. The code can now interface with either NWChem or Orca electronic structure suites, for the computation of *ab initio* energies and gradients. Initially, a finite difference method was implemented to compute the gradient, but it turned out to be too slow for the practical application, even to medium sized systems, as shown in Table 3.1. This happens because the number of single point calculations required to compute the finite difference gradient is proportional to the number of vibrational modes. Therefore, the code now requires the computation of the *ab initio* gradients directly to the electronic structure software. This allows to have directly the analytical gradient when available, at little more than the single point calculation time. If the analytical gradient is not available, the finite difference gradient is automatically selected by the electronic structure software. In addition, the code can be easily adapted to other electronic structure suites that have the analytical gradient of the desired *ab initio* method. This is done by adapting the Bash script shown in Listing 3.1. The mass-weighted switching gradient is deduced from the switching Hamiltonian H_{as}:

$$\dot{\mathbf{p}}_{\mathrm{as}} = -\frac{\partial H_{\mathrm{as}}}{\partial \mathbf{q}} = -\frac{\partial}{\partial \mathbf{q}}\left[(1-\lambda)H_{\mathrm{harm}} + \lambda H_{\mathrm{ai}}\right] \tag{3.1}$$

$$= (\lambda - 1)\frac{\partial H_{\mathrm{harm}}}{\partial \mathbf{q}} - \lambda \frac{\partial H_{\mathrm{ai}}}{\partial \mathbf{q}} \tag{3.2}$$

$$= (\lambda - 1)\frac{dV_{\mathrm{harm}}}{d\mathbf{q}} - \lambda \frac{d\mathbf{x}}{d\mathbf{q}}\left(\frac{\partial V_{\mathrm{ai}}}{\partial \mathbf{x}}\right) \tag{3.3}$$

where H_{harm} and V_{harm} are the harmonic Hamiltonian and potential, H_{ai} and V_{ai} are the *ab initio* Hamiltonian and potential, $\mathbf{q}$ are the mass-weighted normal mode positions and $\mathbf{x}$ are the Cartesian coordinates. Given the normal mode rotation matrix $\mathbb{C}$ and the vector of atomic masses $\mathbf{m}$,

$$\frac{d\mathbf{x}}{d\mathbf{q}} = \frac{\mathbb{C}^T}{\sqrt{\mathbf{m}}} \tag{3.4}$$

```
1 #File per generare i set di traiettorie
2
3 #!/bin/bash
4 export EXE=/home/rconte01/orca_5_0_3_linux_x86-64_shared_openmpi411/
    ↪ orca
5
```

```
FILE=‘head -n 2 input_walk | tail -n 1‘
NAT=‘head -n 1 $FILE | awk ’{print $1}’‘

#echo $FILE
#echo $NAT

echo ’! B3LYP D3BJ Def2-TZVPD ENGRAD VeryTightSCF Bohrs
%PAL NPROCS 20 END
%SCF
MAXITER 500
END
* xyz 0 1’ > step.inp

cat step.xyz >> step.inp

echo ’*’ >> step.inp

$EXE step.inp > step.out


LC_ALL=C fgrep "FINAL SINGLE POINT ENERGY" step.out | awk ’{print $5 }’ >> energie.dat
LC_ALL=C fgrep -A$((NAT+2))  "CARTESIAN GRADIENT" step.out | tail -n $NAT | awk ’{print $4, $5, $6}’ >> energie.dat
```

LISTING 3.1: *BASH script for handling the* ab initio *tasks with ORCA.*

TABLE 3.1: *Computational times (t_{as}, in* min*) for a hundred adiabatic switching steps performed with the finite difference gradients, compared with those (t_{qmd},* min*) of a corresponding molecular dynamics trajectory evolved using NWCHEM* `dft qmd` *task. Both methods use DFT/B3LYP level of theory with aVDZ basis set. The column N_ν accounts for the number of vibrational degrees of freedom.*

Molecule	N_ν	t_{as}	t_{qmd}	t_{as}/t_{qmd}
H_2	1	3.56	0.48	7.42
H_2CO	6	254.60	10.20	24.96
CH_4	9	333.25	5.88	56.67

In SC molecular dynamics symplectic integrators are adopted to ensure high precision in the dynamics runs, and their careful application can increase the efficiency of the simulation.[109] One can commonly choose between two levels of complexity according to the usual trade-off between accuracy and computational costs. At a computationally cheaper level, the very popular velocity-Verlet algorithm is a second-order symplectic algorithm, while more refined 4-th order symplectic algorithms have also been developed. Since on-the-fly dynamics is performed using the built-in modules of the electronic structure software, it is limited

to the more available velocity-Verlet, whereas 4-th order is preferred when a PES is available. In the applications here presented, for the adiabatic switching dynamics I have employed the 4-th order symplectic algorithm by Brewer *et al* for the smaller molecules,[110] and the velocity-Verlet one for the larger systems. The SC dynamics is always based on the velocity-Verlet algorithm, being exclusively on-the-fly.

As stated in Section 2.7, the switching dynamics is performed by evolving the set of coordinates shown in Equation (2.117), where n reflects the desired vibrational state and ζ can be either randomly generated or set equal to $-1/4$ to start a single trajectory from the equilibrium geometry. In the case of a single trajectory, starting from the equilibrium geometry reduces mode coupling during the switching trajectory, as the normal modes variables are well-defined in the neighbourhood of the equilibrium condition. Harmonic action variables are a straightforward way to start the trajectory from an energy corresponding to the harmonic quantization. Furthermore, increasing the vibrational quantum number n_i allows one to excite selectively the i-th normal mode, thus allowing for refinement of the associated semiclassical spectroscopic signal according to the MCSCIVR approach.

At each step of the dynamics the system is evolved as follows:

1. The harmonic potential V_{harm} and gradient $\mathbf{g}_{\text{harm}}$ are evaluated according to the initial phase-space coordinates;

2. Mass-scaled normal modes are mapped to Cartesian coordinates to write the electronic structure software input;

3. The gradient task is launched;

4. The *ab initio* energy and gradient are read from the electronic structure software output file using the script shown in Listing 3.1. The analytical gradient is transformed from Cartesian to mass-scaled normal modes;

5. The switching Hamiltonian H_{as} and switching force $\dot{\mathbf{p}}_{\text{as}}$ are evaluated using Equations (2.113) and (3.3);

6. The system is evolved with the chosen integration algorithm.

This iterative procedure is represented in Figure 3.1. Due to the constant back and forth from normal modes to Cartesian coordinates, the switching trajectory takes a little more time than an equal-length standard trajectory performed with the chosen electronic structure software, as summarized in Table 3.2.

The code was written so that it can be seamlessly incorporated in the established semiclassical procedure in place of the harmonic equilibrium conditions of MCSCIVR. For clarity sake, the on-the-fly ASSCIVR work consists in

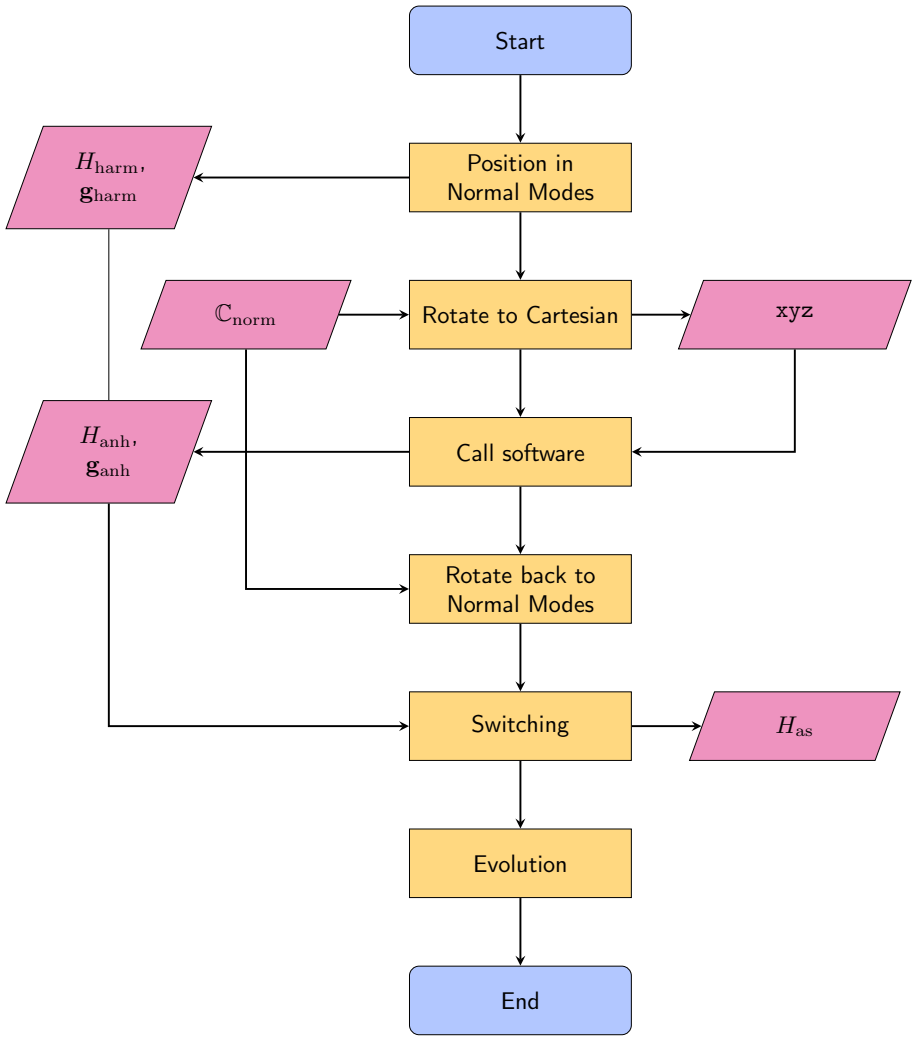


FIGURE 3.1: *Flow chart of a single step of the on-the-fly AS trajectory.*

1. *Preparation*: the system is prepared in harmonic state $(\mathbf{p}_{\text{harm}}; \mathbf{q}_{\text{harm}})$;
2. *Adiabatic Switching*: the harmonic state is evolved under the switching Hamiltonian Equation (2.113), reaching the anharmonic state $(\mathbf{p}_{\text{as}}; \mathbf{q}_{\text{as}})$;
3. *Classical Dynamics*: the anharmonic state $(\mathbf{p}_{\text{as}}; \mathbf{q}_{\text{as}})$ is evolved with classical molecular dynamics, with an electronic structure software;
4. *SCIVR*: the classical action and Hessian matrices are computed from the classical trajectory, and the SCIVR power spectrum initialized at anharmonic conditions is obtained via Equation (2.75). Apart from the trajectory initial conditions, this step has not changed from the previous semiclassical applications, including using a reference state centered at the equilibrium conditions.

3.3 Preliminary tests

I attested the efficiency and accuracy of different trajectory integration methods on several molecules. The efficiency was evaluated by estimating the ratio of CPU times between a 2500-step AS trajectory and an equal length, velocity-Verlet-integrated classical dynamics trajectory, for both the symplectic second (velocity Verlet) and fourth order integrators. Trajectories were evolved on-the-fly with NWChem using DFT/B3LYP with aVDZ basis set. Results are reported in Table 3.2, from which it was clear that, not surprisingly, the fourth order symplectic algorithm is substantially slower than the second order one. It was promising for the application of on-the-fly ASSCIVR to notice that if the same velocity-Verlet integrator is used, then the AS procedure requires just a fraction of additional computational effort.

Table 3.2: *Ratio between the CPU time t for an AS trajectory (as) using analytical gradient with second (superscript* (2)*) and fourth order (superscript* (4)*) symplectic algorithm, and an equal length NWChem molecular dynamics (md) trajectory. In the last column, the ratio of AS CPU times is shown.*

Molecule	$t_{\text{as}}^{(2)}/t_{\text{md}}$	$t_{\text{as}}^{(4)}/t_{\text{md}}$	$t_{\text{as}}^{(4)}/t_{\text{as}}^{(2)}$
H_2	1.20	5.41	4.51
CH_4	1.14	5.71	5.00
H_2CO	1.12	5.90	5.00
Gly	1.41	6.26	4.44

I then evaluated the accuracy of the integrator in terms of conservation of energy during an on-the-fly classical dynamics run at H_{harm} energy. For the AS

code, this was accomplished by using Equation (2.113) with $\lambda = 0$ for all the trajectory. In this way, I estimated the loss in accuracy by adopting the faster velocity-Verlet integrator, without any influence from the AS procedure. The results are reported in Table 3.3. From the standard deviations it was evident that the fourth order is much more accurate than velocity Verlet, whose accuracy was comparable only by reducing the time step by one order of magnitude. However, the amplitude of velocity-Verlet oscillations do not scale with the system dimensions, so that they become negligible if compared to the total energy of a biomolecular system such as glycine.

TABLE 3.3: *Standard deviation σ (in cm^{-1}) of the trajectory total energy, for a classical trajectory ($\lambda = 0$ at each time step) run at the harmonic zpe energy $E_{\mathrm{zpe}}^{\mathrm{harm}}$. The subscript stands for the time step length in atomic units.*

		2-nd order integrator		4-th order integrator	
Molecule	$E_{\mathrm{zpe}}^{\mathrm{harm}}$	σ_1	σ_{10}	σ_1	σ_{10}
H_2	2178	0.075	7.649	1.109×10^{-7}	1.125×10^{-3}
CH_4	9740	0.058	7.518	4.491×10^{-8}	5.750×10^{-4}
H_2CO	5780	0.011	3.180	6.915×10^{-9}	2.070×10^{-4}
Gly	17336	0.078	8.650	7.109×10^{-8}	8.260×10^{-4}

The amplitude of the total energy oscillations during a switching trajectory is related to the different energy conservation properties of the chosen integrator. This is evident from the energy profile for the on-the-fly AS dynamics performed for H_2 and reported in Figure 3.2. The fourth order integrator is clearly more accurate and for this reason we employed it in our on-the-fly ASSCIVR tests on small molecules described further below.

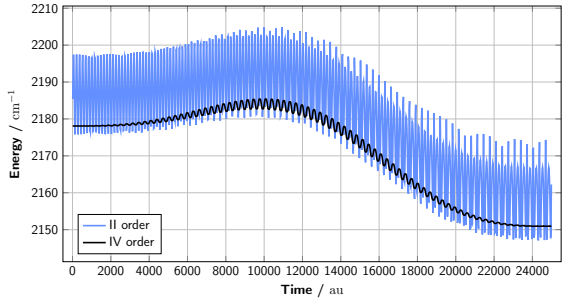


FIGURE 3.2: *Energy profile for the AS trajectory of H_2. The trajectories were integrated for a total of 2500 steps of* 10 au *at* DFT/B3LYP *aVDZ level of theory.*

3.4 Small Molecules

The accuracy of the on-the-fly ASSCIVR method was initially tested on small molecules, *i.e.* H_2 and H_2CO, for which MCSCIVR is already known to provide satisfactory results. For each molecule, I ran a preliminary 25 000 au-long AS trajectory initiated from harmonic ZPE conditions. For the SCIVR part of the simulation, this was followed by a 25 000 au-long classical trajectory started from the final phase space coordinates of the AS trajectory. The number of Hessian matrices to be computed was cut down using HDB. Both simulations were carried out at DFT/B3LYP level of theory with aVDZ basis set. In Figure 3.3, H_2 ASSCIVR spectra are compared with MCSCIVR ones, obtained with an equal length trajectory at same level of electronic theory but started from harmonic conditions. The two spectra are quite similar, as confirmed by the extracted frequency values: the ZPE values are 2136 and 2127 cm^{-1} (the two most intense signals in the two power spectra); the fundamental transition frequencies are 4158 and 4143 cm^{-1}; the overtone frequencies are 8429 and 8395 cm^{-1} for ASSCIVR and MCSCIVR, respectively. The spectra of H_2CO are compared between the two methods and the available experiments in Figure 3.4. Detailed frequencies are given in Table 3.4.[111]

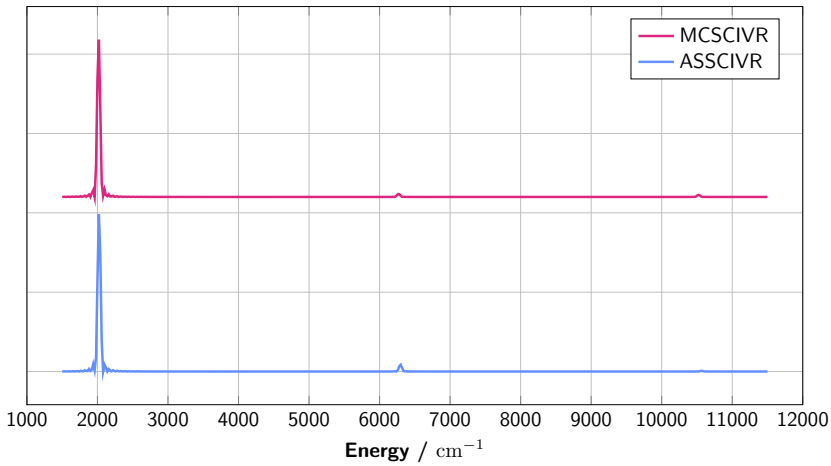


Figure 3.3: *MCSCIVR (magenta, top) and ASSCIVR (blue, bottom) power spectra for H_2 based on single trajectories at DFT/B3LYP level of theory.*

The theoretical results are in excellent agreement between each other and the mean absolute error (MAE) with respect to the experiment is below 20 cm^{-1}, in spite of the DFT level of theory employed. The spectroscopic signals for H_2CO

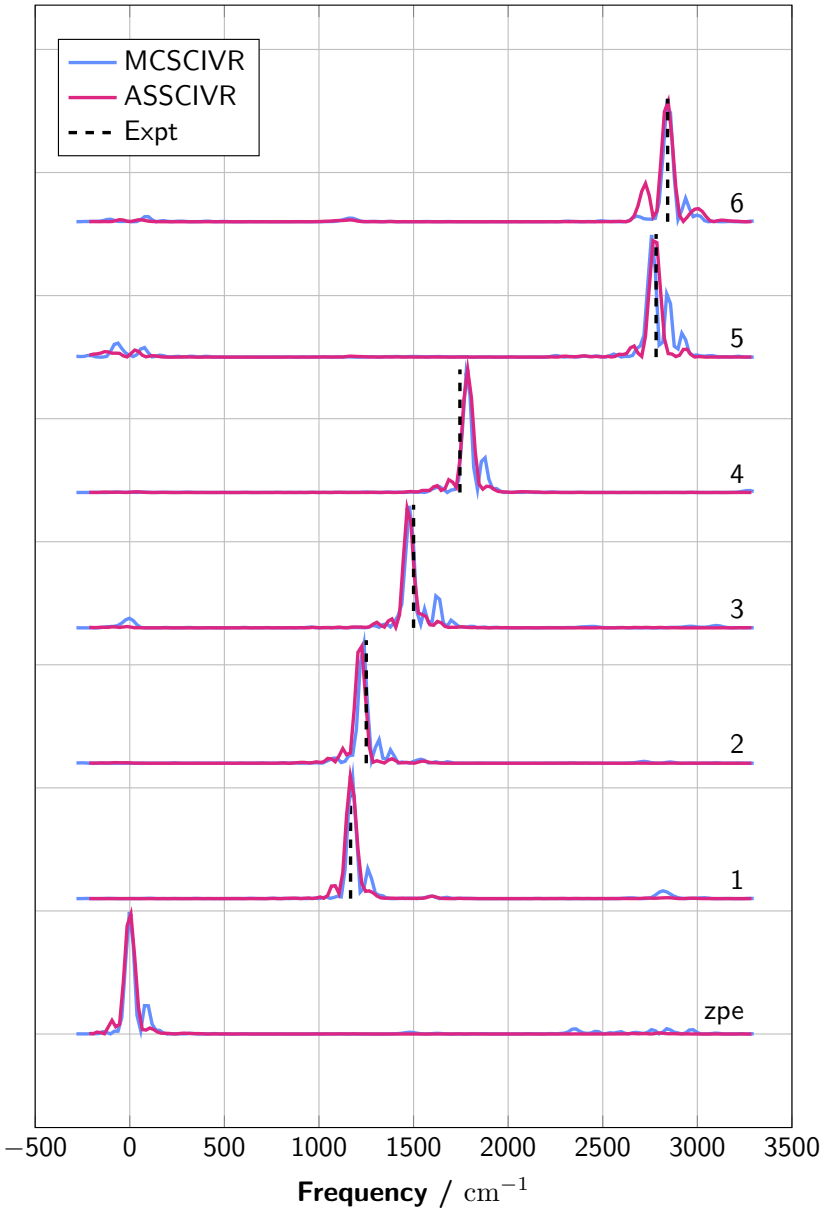


FIGURE 3.4: *MCSCIVR (blue) and ASSCIVR (magenta) spectra of formaldehyde from* 25 000 au *long trajectories. Experimental values (black) are taken from Ref.[111].*

TABLE 3.4: *ZPE and fundamental frequencies for H_2CO; the FWHM values are reported in parenthesis. Experimental values (Expt) are from Ref.[111]. MAE stands for mean absolute error. All values are in* cm^{-1}.

Mode	Expt	MCSCIVR	ASSCIVR
ZPE	—	5862 (51)	5714 (62)
1	1167	1174 (51)	1170 (63)
2	1250	1232 (49)	1218 (59)
3	1500	1482 (50)	1472 (58)
4	1745	1784 (50)	1786 (62)
5	2782	2760 (50)	2776 (64)
6	2843	2846 (57)	2840 (67)
MAE	—	18	19

have comparable widths, with MCSCIVR peaks slightly narrower. This is unexpected because ASSCIVR has been demonstrated to provide narrower peaks, a feature that is clear moving to larger molecules. However, this turns out to be a fictitious feature. In fact, the MCSCIVR signals are characterized by a small side peak, which is blueshifted by about $50\,\mathrm{cm}^{-1}$ with respect to the main one. Since H_2CO does not have any vibrational mode below $100\,\mathrm{cm}^{-1}$, the side peaks cannot represent the fingerprint of any combination of vibrational levels. They are instead a simulation artifact due to a slow and almost unperceivable rotation of the molecule during the trajectory, which splits the vibrational signal into two components. The effect is twofold: first, the vibrational estimate provided by the MCSCIVR method is less reliable; second, the two signals are narrower than the expected single one. As for the ASSCIVR simulations of H_2CO, we point out that they provide better defined spectroscopic signals.

3.5 GLYCINE

The spectroscopic study of glycine (Figure 3.5) saw the application of recent advances in both PES description and semiclassical molecular dynamics.[16, 61, 104, 112] A first semiclassical study of glycine by means of MCSCIVR was based on *ab initio* on-the-fly molecular dynamics.[87] It allowed to investigate dynamically the very complex energy landscape of the four lowest-lying conformers of glycine, corroborating previous findings of other static quantum approaches like vibrational self-consistent field (VSCF) and second-order VPT. A permutationally invariant PES was recently employed to identify eight conformers of glycine.[16] Both diffusion monte carlo (DMC) and ASSCIVR calculations have

been performed using this surface to elucidate the link between the conformers through zero-point energy calculations. The estimates of the two methods were in excellent agreement (about $5\,\mathrm{cm}^{-1}$).

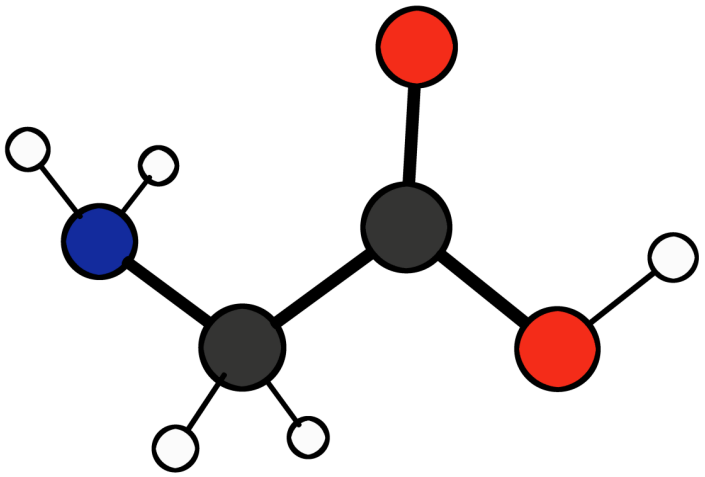


FIGURE 3.5: *Equilibrium geometry for the 10-atoms amino acid glycine*

For small molecules, single-trajectory ASSCIVR simulations provide just a slight improvement over the MCSCIVR results, as expected. However, the complexity of the spectrum increases as the system gets larger. This is the case of biomolecular systems, that range from the small organic compounds to functional aggregates made of thousands of atoms. For such systems any improvement in precision of the simulations may be crucial for the correct interpretation of the spectroscopic insights as it is shown in both Chapters 4 and 5. To properly assess on-the-fly ASSCIVR capabilities on biomolecular systems, it was applied to the well-known glycine too.

Glycine is a floppy molecule with low interconversion barriers between conformers. Perturbing the actual Hamiltonian with an AS procedure may result in an undesired change of conformer. Therefore, I adopted a longer switching trajectory (35 000 au) to get closer to the adiabatic hypothesis, thus preventing undesired interconversions. In addition to that, the switching trajectory was initialized using the harmonic ZPE equilibrium conditions, corresponding to those shown in Equation (2.117), in which $n_i = 0$, $\zeta_i = -1/4\ \forall i$ normal modes. Finally,

previous experience with this amino acid suggested to assign zero energy to the first vibrational normal mode. This mode is a low-frequency internal rotation that sits along the conformer change coordinate, playing a pivotal role in the conformational change of the molecule. However, for conservation of the action variables during the AS process, the momentum of the first vibrational mode was not set to zero, so as not to lose the initial harmonic quantization. The first vibrational mode was "switched off" only at the end of the AS run, by setting to zero the relative component of $\mathbf{p}_{\text{as}}$ before the semiclassical dynamics.

With all these precautions in place, I obtained the ASSCIVR spectra at DFT/B3LYP aVDZ level of theory. A selection of spectra thusly obtained is compared with analogous MCSCIVR spectra in Figure 3.6. Experimental and harmonic DFT/B3LYP aVDZ frequencies are included. The spectral analysis is reported in Table 3.5.[113, 114] The ASSCIVR spectra are sharper than the MCSCIVR ones and much cleaner, as it can be appreciated from the ZPE spectra in Figure 3.7. This is a consequence of the improved initial conditions given by AS, with respect to the harmonic-based MCSCIVR ones.

ASSCIVR glycine signals generally feature a much less intense side peak at lower frequency. This is present in the spectra of the ZPE and of several fundamentals (Figure 3.7). Given the low intensity and its position, this is likely due to the coupling with a rotational state. This coupling is an effect of having evolved the switching trajectory in normal modes, which are rigorously defined only at the equilibrium geometry. The back-and-forth between normal modes and Cartesian coordinates (although necessary) introduces some degree of vibro-rotational coupling. The side peak is expected to disappear when employing thousands of trajectories, a task not affordable with an on-the-fly simulation. As expected, there are some minor discrepancies between MCSCIVR and ASSCIVR results, due to the different initial conditions. Overall, ASSCIVR simulations are more precise than MCSCIVR ones while their accuracy is comparable, as demonstrated by the similar mean absolute error values of about $35\,\text{cm}^{-1}$ (Table 3.5). Finally, the ZPE values are within just $5\,\text{cm}^{-1}$ with DMC calculations recently performed on a high-level analytical PES for glycine, as shown in Figure 3.7.[16]

The DCSCIVR method was applied on the classical trajectory that followed the switching one, without any change in the on-the-fly AS method, but at an increased computational cost. The resulting spectra are shown in Figure 3.8, while the peaks frequencies and FWHM are reported in Table 3.6. It can be appreciated that applying DCSCIVR after the AS method can improve the precision of some spectra (as in mode 12, Figure 3.8), but can also add spurious signals (as in mode 24, Figure 3.8). These are mostly due to the choice of the wrong subspace. This comparison shows that for small-to-medium systems, ASSCIVR can be adopted to improve the semiclassical simulation at a lower cost that DCSCIVR. Moreover, contrary to ASSCIVR, DCSCIVR cannot collect the full-dimensional ZPE signal,

TABLE 3.5: *ZPE energies and transition frequencies (FWHM values in parenthesis) for glycine spectra obtained with MCSCIVR and ASSCIVR, compared with harmonic (DFT/B3LYP aVDZ) and experimental values.[113, 114] The "Mean" row reports the average FWHM value. All the data are reported in* cm^{-1}.

Mode	MCSCIVR	ASSCIVR	Harm	Expt.
ZPE	17158 (145)	17152 (96)	17363	—
1	99 (107)	72 (92)	59	—
2	181 (126)	203 (96)	208	204
3	229 (131)	261 (93)	249	250
4	450 (150)	461 (99)	457	458
5	499 (146)	495 (93)	510	500
6	569 (120)	665 (134)	628	615
7	667 (137)	596 (93)	647	619
8	780 (133)	738 (99)	815	801
9	776 (69)	732 (95)	906	907
10	914 (150)	904 (115)	911	883
11	1103 (156)	1071 (103)	1120	1101
12	1106 (130)	1115 (100)	1158	1136
13	1159 (153)	1129 (99)	1175	1166
14	1243 (150)	1266 (98)	1294	1297
15	1299 (147)	1329 (97)	1371	1340
16	1339 (151)	1358 (100)	1384	1405
17	1375 (117)	1414 (102)	1439	1429
18	1628 (86)	1606 (108)	1656	1608
19	1776 (154)	1778 (92)	1804	1779
20	2945 (142)	3005 (98)	3049	2943
21	2934 (149)	3006 (101)	3086	2969
22	3316 (161)	3422 (105)	3495	3359
23	3327 (92)	3463 (81)	3569	3410
24	3640 (142)	3638 (96)	3736	3585
Mean	— (133.76)	— (99.40)	—	—
MAE	35.87	35.43	41.61	—

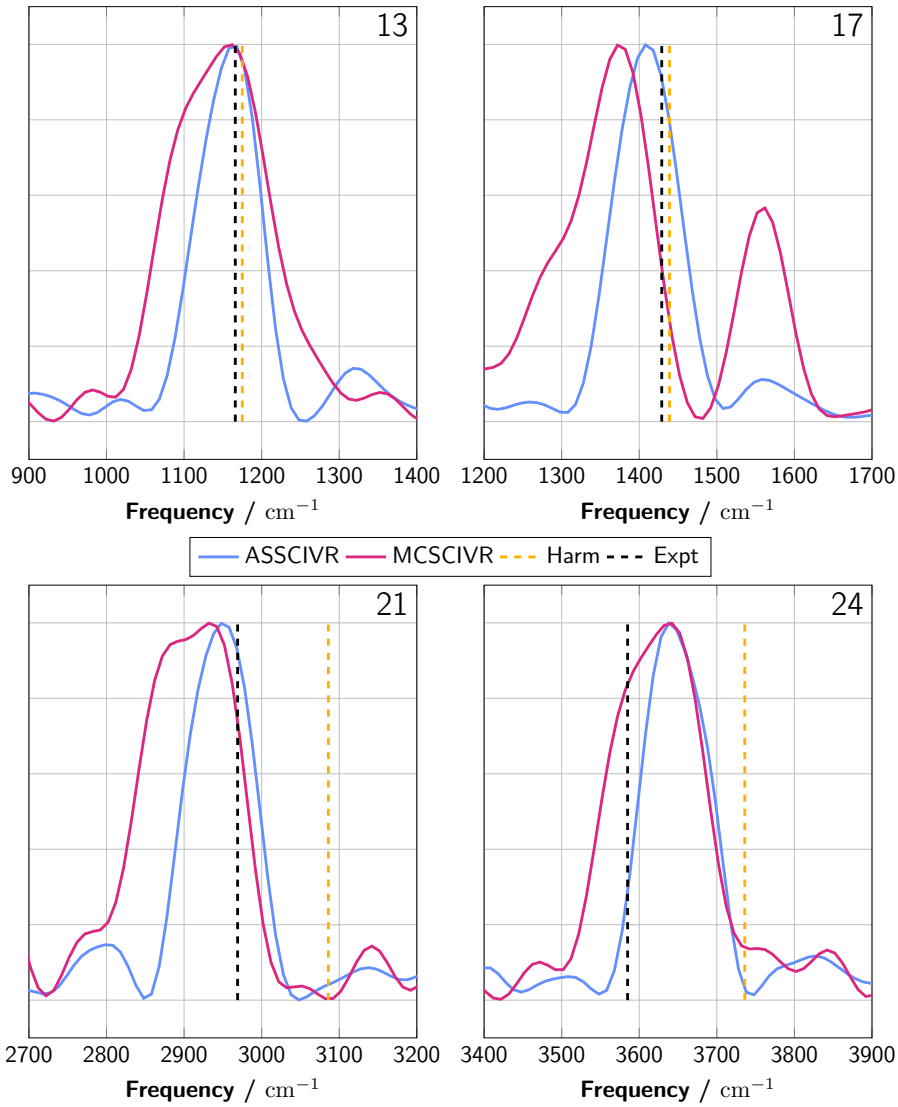


FIGURE 3.6: *Comparison between the glycine peaks obtained through the different methods, at* DFT/B3LYP *aVDZ level of theory. In blue the ones obtained with ASSCIVR; in magenta the MCSCIVR ones; in black the experimental frequencies [113, 114] and in yellow the harmonic frequencies.*

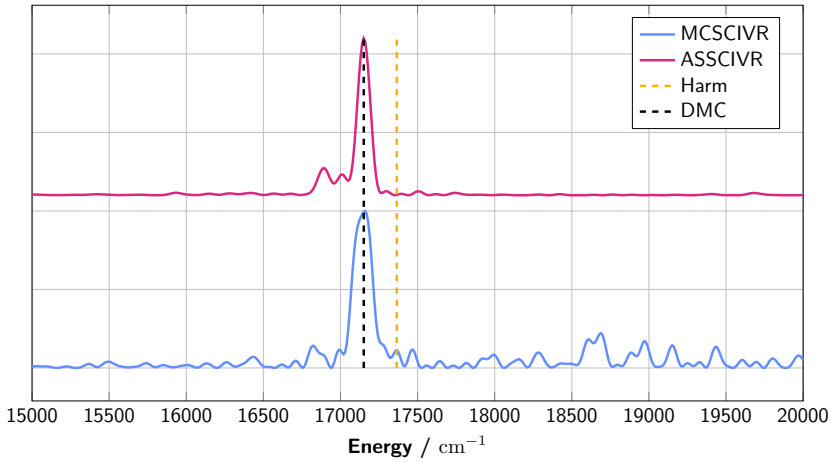


FIGURE 3.7: *Glycine ZPE power spectra. In magenta: ASSCIVR zpe spectrum, obtained with a* 25 000 au *classical trajectory at* DFT/B3LYP *aVDZ level of theory; in blue: the analogous MCSCIVR spectrum. The harmonic ZPE is in dashed yellow, while the DMC one is in dashed black.[16]*

that is a fundamental feature in analysing the behaviour of biomolecular systems, and a prerequisite to extract the semiclassical wavefunctions.[16, 103, 106, 108, 112]

Nonetheless, some semiclassical vibrational frequencies are quite far from the experimental value, requiring further refinement. This is usually performed by following the evolution of the molecule in a particular excited vibrational state. In harmonic approximation, this is done by increasing the vibrational quantum number of the desired normal mode. The same holds for ASSCIVR, since the adiabatic theorem guarantees that the vibrational quantum number is conserved through the switching trajectory, as shown in Section 2.7. The refinement was employed to improve mode 16, that was poorly described with both MCSCIVR and ASSCIVR. The results of the refinement procedure are illustrated in Table 3.7 and Figure 3.9: the ASSCIVR spectrum is clearly less noisy than the one obtained with MCSCIVR, but is further away from the experimental value. A possible reason is that mode 16 was mixed with mode 15 by the adiabatic trajectory, resulting in a structured peak that SCIVR cannot resolve. The mixing of vibrational states with similar energies is now a well-known drawback of AS, as described in Chapter 6. Even so, this test proves that the adiabatic initial conditions can be employed in the refinement procedure as much as the harmonic ones, when accounting for the inherited limitations of AS.

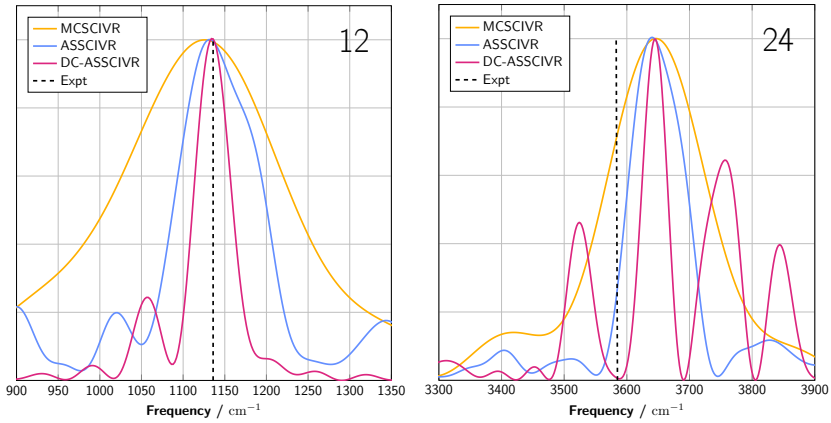


FIGURE 3.8: *Comparison between the glycine peaks obtained through the different methods. On the left, those relative to mode 12; on the right, those relative to mode 24. In blue, the ones obtained with ASSCIVR; in yellow the MCSCIVR ones; in magenta the DC-ASSCIVR ones; in black the experimental frequencies from [113].*

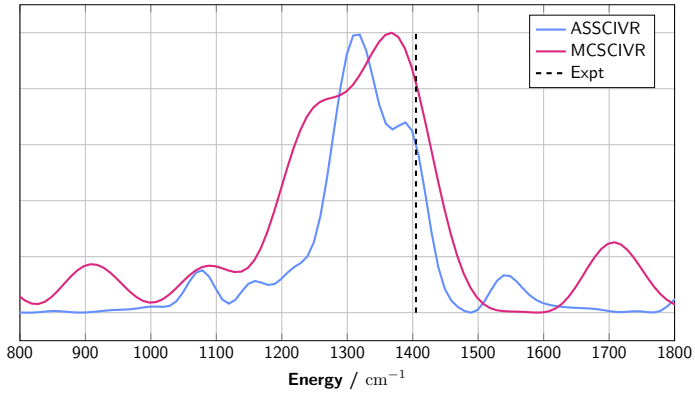


FIGURE 3.9: *Comparison between the glycine semiclassical spectra of mode 16, obtained after the refinement procedure. In blue, the one obtained with ASSCIVR; in magenta, the MCSCIVR one; in black, the experimental frequency from [113].*

TABLE 3.6: *Vibrational frequencies and FWHM (in* cm^{-1}*) of glycine DC-ASSCIVR spectra; in the AE column, the absolute error from the experiment.[113]*

Mode	Freq. (FWHM)	Expt.	AE
1	86 (86)	—	—
2	191 (126)	204	13
3	312 (48)	250	62
4	453 (47)	458	5
5	506 (105)	500	6
6	667 (93)	615	52
7	630 (46)	619	11
8	797 (102)	801	4
9	876 (101)	907	31
10	917 (99)	883	34
11	1096 (40)	1101	5
12	1134 (44)	1136	2
13	1132 (49)	1166	34
14	1285 (38)	1297	12
15	1320 (167)	1340	20
16	1300 (46)	1405	105
17	1411 (49)	1429	18
18	1620 (141)	1608	12
19	1785 (53)	1779	6
20	2939 (105)	2943	4
21	2935 (54)	2969	34
22	3377 (150)	3359	18
23	3431 (203)	3410	21
24	3644 (42)	3585	59
Mean	(84.75)	—	24.66

TABLE 3.7: *Frequency, FWHM and absolute error from the experiment (in* cm^{-1}*) of glycine mode 16 signal, with and without refinement (+), for both MCSCIVR and ASSCIVR.*

Method	Freq. (FWHM)	AE
MC	1320 (180)	85
MC+	1367 (226)	38
AS	1288 (95)	117
AS+	1314 (145)	91

3.6 Conclusions: improving the sampling

On-the-fly ASSCIVR confirmed the main advantages of the method already demonstrated for small molecules on analytical PES.[53] Specifically, the adiabatic switching initialization of the trajectory generally leads to narrower peaks and more accurate results.

There are two potential drawbacks of the technique which anyway did not hamper any of the shown simulations. One is the necessity to generate a single trajectory when solving the electronic problem on-the-fly. It may happen that at the end of the AS procedure the energy content of the trajectory is slightly below the ZPE of the system, in which case the final frequency estimate could be slightly shifted towards the harmonic estimate. This had a very minor impact in our simulations. The second issue is the possible, accidental introduction of a small amount of roto-vibrational coupling during the preliminary AS run. This has produced some minor side peaks that were not a threat to the correct assignment of spectral features. This made almost impossible to obtain a resolved spectrum of CH_4, a molecule notoriously prone to spinning. The problem could be partially eased by implementing the adiabatic switching procedure in internal coordinates,[51] which is yet to be implemented in semiclassical dynamics. Finally, the trajectories with a quantum of excitation in mode 16 were probably affected by vibrational state mixing. This drawback of AS was investigated in a later work, described in Chapter 6.[115]

The promising results on glycine prompted me to use ASSCIVR to assign experimental spectrum of proline.[54] Still, this proved to be much more than a straightforward application of an established procedure. What emerged from a thorough analysis of proline experimental spectrum is described in Chapter 4, in which ASSCIVR pushes the limit for a full-dimensional calculation, allowing the treatment of 17-atom proline in full dimensionality.

Chapter 4

The spectrum of proline

The content of this chapter is based on work reported in Ref.[116]

4.1 The cornerstone of proteins

While developing on-the-fly adiabatic switching semiclassical initial value representation, I employed the global minimum of proline to test the method capabilities in treating systems of considerate dimensionality.[105] However, there was no agreement between ASSCIVR, scaled harmonic and previously assigned OH stretching (str) frequencies. Due to this controversial results, I performed a thorough spectral analysis of proline experimental spectrum using ASSCIVR.

Proline is the only natural amino acid whose side chain forms a covalent bond with the amidic nitrogen, creating a pyrrolidinic ring and making proline formally an imino acid. The ring imposes strict constraints on the conformational space,[117] thus reducing the Ramachandran plot of both proline and the preceding amino acid along the primary structure of the protein.[118] The conformational rigidity plays a fundamental role in protein stability and folding. This is the reason for the common presence of proline in specific positions of β sheets and γ-turns. Furthermore, proline gets organized in peculiar structures, like prolinic left helices, and takes part in the stress-response metabolism.[117, 119–123]

The particular structure of proline has also influenced the biochemical nomenclature and a peptidic bond involving proline is commonly dubbed as "prolyl bond". It is well known that a generic peptidic bond is planar, and the two resulting *trans-* and *cis-* isomers are generally well separated in energy. The prolyl bond is different, because the *trans-* and *cis-* isomers are rather close in energy, due to steric hindrance of the pirrolidinic ring.[124] However, the presence of

the "wrong" proline isomer in the tertiary structure may hinder protein folding, making the prolinic *cis/trans* isomerism a rate determining step in certain folding kinetics.[125, 126] Many enzymes that catalyze such isomerism are present in nature and employ a variety of mechanisms. One involves partial stabilization of the nitrogen doublet through a hydrogen bond from the enzymatic pocket or a neighbouring residue.[125–127] This prompts N pyramidalization that reduces the isomerization barrier. In certain cases, the enzyme stabilizes the peptide in a conformation in which the H atom comes from the peptide itself: this mechanism is a form of autocatalysis.[128–130]

Proline is also a key player in asymmetric catalysis due to its coordinating properties. For instance, in aldol condensation, proline is faster in forming the enamine than any other secondary amine, thanks to the carboxylic H atom which stabilizes the transition state.[131] The mechanism of proton relay further facilitates enamine formation.[132] Then, in a second step of the process, a new transition state involving a single proline is formed, and the carboxylic H guides the electrophile through an H bond. This proton transfer is fundamental in aldol and Mannich reactions.[133–135].

Both the biological and catalytic behaviour of proline revolves around the role of hydrogen bond acceptor of the pyrrolidinic nitrogen. Vibrational spectroscopy is a priceless tool to investigate hydrogen bond dynamics, since its signals are heavily influenced by the bond strength. The vibrational spectrum of proline was obtained by Adamowicz and coworkers in an Ar matrix at 14 K (Figure 4.1). Their assignment was mostly based on two features: the OH str peak at 3559 cm^{-1} and the split C=O str peak at 1766 and 1789 cm^{-1}. Annealing analysis excluded any matrix effect in the splitting, suggesting the presence of two conformers in the spectrum. Initially, the split in the C=O str seemed to be due to two conformers, both with a non-hydrogen-bonded OH.

The experimental assumption was based on the OH str frequencies of the other amino acids conformers. Following a common classification of amino acids conformers, the main proline conformers were classified by the experimentalists as

type I with the N−H···O=C hydrogen bond (see Figure 4.2, panel B)

type II with the N···H−O hydrogen bond (see Figure 4.2, panel A)

type III with the N−H···O−H hydrogen bond (see Figure 4.2, panel C)

It is also possible to classify the conformers by the relative position of the NH and COOH groups with respect to the ring: *trans-* conformers have them pointing in opposite directions, while *cis-* conformers have them pointing in the same direction.[119] The type of hydrogen bond formed by the conformer is then used to distinguish within *trans-* and *cis-* conformers. This standard classification labels

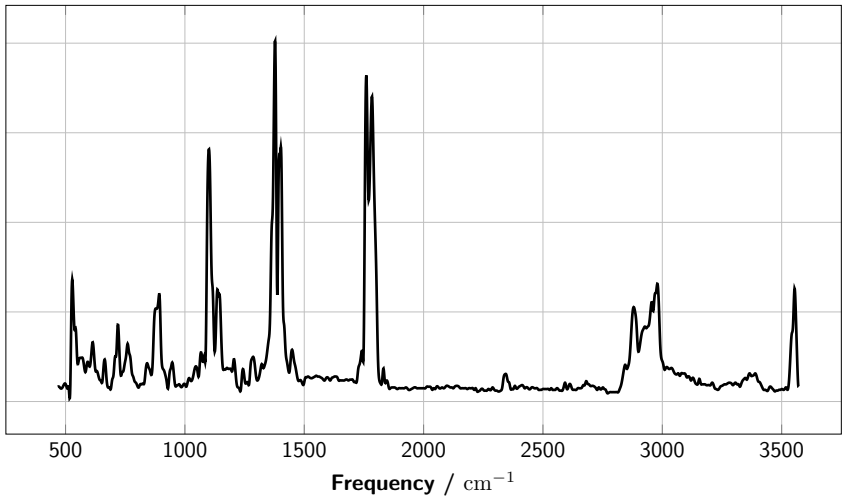


FIGURE 4.1: *Experimental spectrum of proline in Ar matrix, obtained at* 14 K *by Stepanian* et al.*[54] (Copyright © 2001 American Chemical Society)*

three main conformations: a *cis*- with the N−H···O=C hydrogen bond (Figure 4.2, panel B), a *trans*- with the N···H−O hydrogen bond (Figure 4.2, panel A) and another *cis*- with the N−H···O−H hydrogen bond (Figure 4.2, panel C).[136] These correspond, respectively, to Adamowicz's type I, II, and III conformers.[54]

In the case of proline, the OH str frequency was used to distinguish between the first two types of conformers: for type I the frequency should be around 3560 cm^{-1}, whereas for type II it is typically around 3200 −3300 cm^{-1}. For this reason, the 3559 cm^{-1} signal was assigned to the OH str of a type I conformer and the lack of any signal around 3200 −3300 cm^{-1} initially excluded the presence of any type II conformer. The presence of a type III conformer was excluded due to its high energy.[54]

Even so, this qualitative picture was at odds with theoretical energy calculations which predicted a type II conformer as the most stable one. Those calculations were performed by Stepanian *et al.* with single point calculations at MP4(SDTQ) and CCSD(T) level with different basis sets, using the equilibrium geometry found using the MP2 aVDZ level of theory and basis set. The harmonic frequencies and ZPE were obtained at DFT/B3LYP aVDZ level. The predicted scaled harmonic OH str frequency returned a value close to that of the CH band, suggesting that the OH str could be identified by subtracting the CH contribution from the spectrum. This was done by means of the deuterated proline spectrum obtained at the same temperature conditions. In spite of a very weak spectral

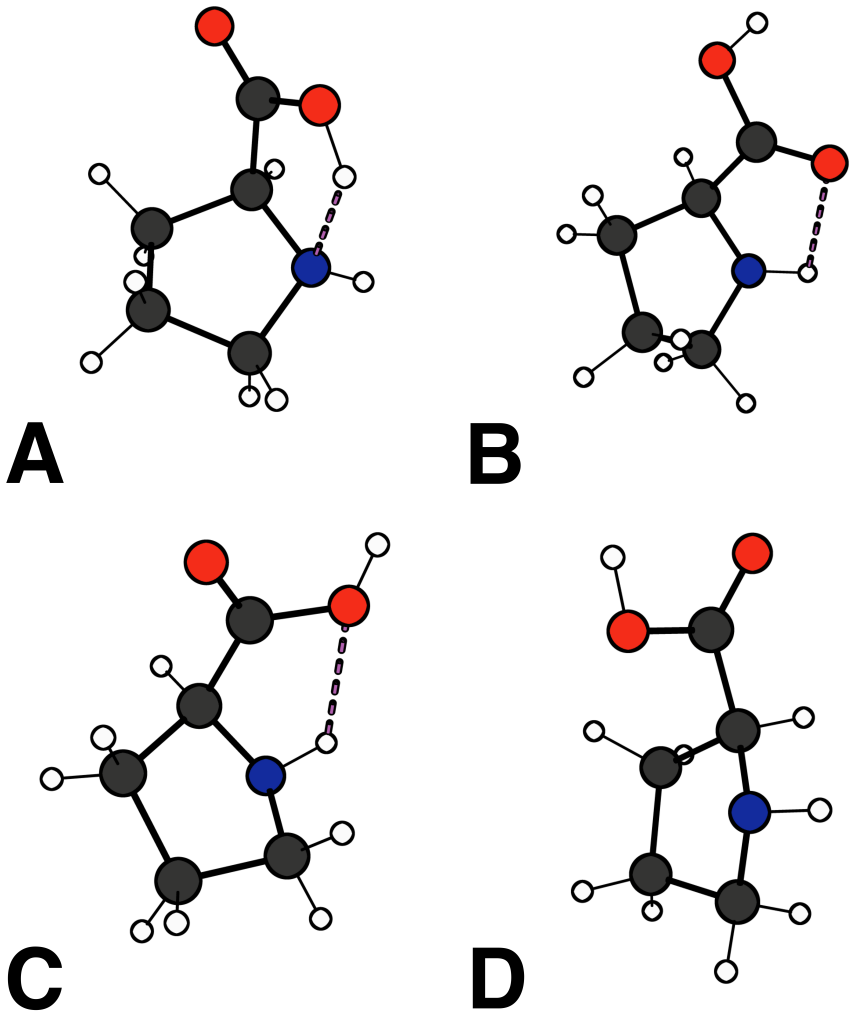


FIGURE 4.2: *A:* trans*-proline (type II), B:* cis*2-proline (type I), C:* cis*1-proline (type III) and D:* trans*2-proline (no hydrogen bond). The dashed line indicates the intramolecular hydrogen bond.*

difference signal, the authors concluded that the OH str of a type II conformer was actually present at 3025 cm^{-1} and that both type I and type II conformers contributed to the spectrum.

4.2 Computational details

I computed the vibrational spectra of proline conformers using the on-the-fly ASSCIVR method described in Chapter 3. All the *ab initio* tasks were performed with NWChem, using Grimme's DFT-D with B3LYP functional and aVDZ basis set.[137, 138] The inclusion of the most recent dispersion correction improved the description of the hydrogen bond interaction without increasing the basis set dimensions. This choice of level of theory was verified by means of energy calculations performed for *trans*- and *cis*1- proline at MP2 and CCSD(T) level of theory, demonstrating the good accuracy of DFT-D. The comparison is summarized in Table 4.1. Furthermore, the previous semiclassical work on glycine undertaken at the same level of theory and same basis set provided very accurate results in comparison to both MP2-based calculations and experimental data.[87]

Table 4.1: *Comparison between different levels of theory. The DFT-D column indicates the energy of the minimum geometry (in Hartree) at DFT-D/B3LYP level of theory; the MP2 column reports the same for the MP2 level of theory; the CCSD(T) column present the CCSD(T) single point energy on the MP2 minimum. Finally, the* Δ *row presents the difference in electronic energy between the two conformers for all levels of theory explored. All calculations were performed with aVDZ basis set.*

Conformer	DFT-D	MP2	CCSD(T)
trans-Pro	-401.238985	-400.128763	-400.289997
*cis*1-Pro	-401.233645	-400.123508	-400.283791
Δ	0.005340	0.005255	0.006207

The optimized geometries were obtained with NWChem `dft optimize` task as shown in the sample input (Listing 4.1). The harmonic frequencies where computed on the optimized geometry using NWChem `dft frequencies` task.

```
start Pro_aVDZ_opt
echo

set geom:dont_verify .true.

geometry units au noautosym noautoz
<xyz geometry>
end
```

```
9
10 basis
11 * library aug-cc-pvdz
12 end
13
14 dft
15   maxiter 100
16   xc b3lyp
17   disp vdw 3
18 end
19
20 driver
21   maxiter 150
22   tight
23 end
24
25 task dft optimize
```

LISTING 4.1: *Sample input for proline optimization*

The adiabatic switching run was initialized at the harmonic ZPE quantization shown in Equation (2.73) and evolved for 2500 steps of 10 au each. The final positions and velocities were then used as initial conditions for the semiclassical dynamics. The classical trajectory was evolved for 25 000 au using NWCHEM `dft qmd` task, with the initial Cartesian velocities given to NWCHEM with an in-house modification. Following a well-consolidated procedure in semiclassical dynamics,[38, 86] to determine more accurately the OH str frequency of *cis*1-proline a simulation was run with one quantum of energy in the normal mode corresponding to the OH str of *cis*1-proline. A sample input for the classical evolution is shown in Listing 4.2

```
1  start Pro_qmd_2500_b3lyp_aVDZ
2  echo
3  print low
4
5  geometry noautosym noautoz
6  <xyz geometry>
7  end
8
9  basis
10 * library aug-cc-pvdz
11 end
12
13 dft
14    xc b3lyp
15 end
16
17 qmd
18   nstep_nucl  2501
19   dt_nucl     10.0
```

```
  targ_temp   0.01
  com_step    1
  rand_seed   12345
  thermostat  none
  print_xyz   1
end
task dft qmd
```

Listing 4.2: *Sample input for proline classical evolution*

The number of Hessian matrices to be computed was reduced to around 150 for each conformer using HDB. Some test were performed to use DCSCIVR on top of ASSCIVR, but the results showed no substantial improvement in signal-to-noise ratio, because the improved stability of the trajectory from adiabatic initial conditions was more than enough.

Finally, the Fourier spectrum of the dipole-dipole autocorrelation function was computed from NWChem output trajectory, using the contributor code `qmd_analysis.f90`, available on the NWChem repository.

4.3 A spectrum for two conformers

From the search of the main minima of proline PES of proline, I identified *trans*-proline as a type II conformer, *cis*1-proline as a type III conformer, and *cis*2-proline as a type I conformer (Figure 4.2). When performing the semiclassical part of the dynamics (at anharmonic ZPE) for *cis*2-proline, a quick interconversion to a higher electronic energy conformer took place. I labeled this additional conformer, characterized by the absence of the internal hydrogen bond, as *trans*2-proline (Figure 4.2, panel D). Owing to this quick interconversion of *cis*2-proline, I employed *cis*1-proline in my simulations for the *cis*-type conformer. To strengthen this choice I probed the *cis*2-/*trans*2-proline system by means of an additional ASSCIVR simulation. The resulting simulation provided more elaborated peaks, as expected on the basis of the more complex dynamics, but frequency values for the main modes of vibration were in very close agreement with those found for *cis*1-proline. Table 4.2 reports the absolute and relative electronic energy values of the four conformers. Adiabatic electronic energies including harmonic and anharmonic ASSCIVR ZPE estimates provide a similar picture and can be found in Table 4.3.

In the high frequency region, the spectrum of proline is characterized by the OH and NH stretches. Harmonic estimates in Table 4.4 show that, considering the expected red shift due to anharmonicity, the OH str experimental signal at $3559\,\mathrm{cm}^{-1}$ cannot be assigned to *trans*-proline. Therefore, a second conformer must contribute to the spectrum and I focused on *cis*1-proline, given the fast interconversion of *cis*2-proline. Figure 4.3 reports the ASSCIVR density of states

simulations for the OH stretch. These provide signals at 3329 cm^{-1} and 3522 cm^{-1} for the *trans*- and *cis*1- conformer respectively, confirming that the experimental peak at 3559 cm^{-1} is representative of the OH stretch of the *cis*1- conformer. Surprisingly, there is no evidence in the experimental spectrum of the OH signal relative to the *trans*-conformer. This is very unexpected since looking at Table 4.4 the most intense OH signal should be the *trans*-proline one. My estimate of the OH str of the *trans*- conformer (3329 cm^{-1}) does not agree with the previous assignment in the literature (3025 cm^{-1}).[54]

TABLE 4.2: *Absolute (in E_h) and relative (in cm^{-1}) electronic energies of the four main conformers of proline at* DFT-D-B3LYP *aVDZ level of theory. GM stands for global minimum.*

Conformer	Energy	Energy from GM
trans-proline (GM)	-401.23899	0
*cis*2-proline	-401.23589	679
*cis*1-proline	-401.23364	1172
*trans*2-proline	-401.23125	1697

TABLE 4.3: *Harmonic (Harm) and semiclassical (ASSCIVR) zero point energies (in cm^{-1}) and relative electronic energies including harmonic (Harm) and semiclassical (ASSCIVR) zero-point energies of the four conformers (in cm^{-1}).* Cis*2-proline semiclassical ZPE wasn't obtained due to the rapid interconversion to* trans*2-proline*

	Energy with ZPE		Relative Energy	
Conformer	Harm	ASSCIVR	Harm	ASSCIVR
trans-proline	31843	31362	0	0
*cis*2-proline	31688	—	524	—
*cis*1-proline	31708	31242	1037	1052
*trans*2-proline	31716	31207	1570	1542

To clarify this point and the weakness of OH str signal of *trans*-proline, I performed an additional calculation of the absorption intensity, to compare with the double harmonic estimate provided by the NWCHEM software.[137] The additional simulation used the Fourier transform of the autocorrelation dipole-dipole function calculated on the semiclassical part of the same ZPE trajectory employed for the ASSCIVR simulation of *trans*-proline. Therefore, I focused on the relative intensity of CO str and OH str of *trans*-proline. The CO str was chosen as a reference because it is one of the most intense signals in the experimental

TABLE 4.4: *Double harmonic approximation (frequency ω, and intensity S) for the OH and NH stretches of the four main conformers of proline. Calculations performed at* DFT-D/B3LYP *aVDZ level of theory. Frequencies are in* cm^{-1} *and intensities in* $\mathrm{km\,mol}^{-1}$.

Conformer	ω_{OH}	ω_{NH}	S_{OH}	S_{NH}
trans-proline	3355	3560	346.09	6.64
*cis*1-proline	3740	3548	69.06	23.02
*cis*2-proline	3742	3520	65.94	1.28
*trans*2-proline	3744	3529	66.99	6.04

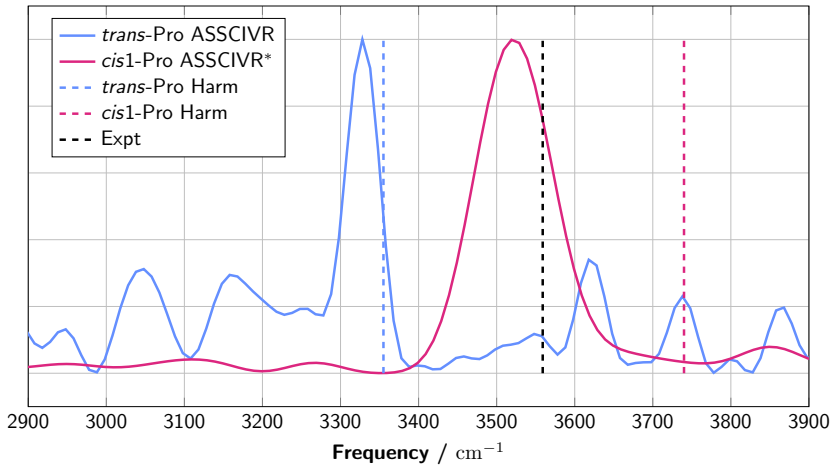


FIGURE 4.3: *ASSCIVR density of states simulations for the OH stretches of* trans*-proline (blue) and* cis*1-proline (magenta). The corresponding harmonic frequencies are indicated by same-color vertical dashed lines. The experimental value for the* cis*1-proline OH stretch (from Ref.[54]) is represented by the dashed black line. No experimental value for the* trans*-proline OH stretch is shown due to controversial assignment (see text).* * *excited trajectory employed (see text).*

spectrum and its double harmonic intensity was comparable with the OH str one. Figure 4.4 shows a comparison between the experimental IR spectrum and the simulations. From the figure it is clear that at the harmonic level the OH str is almost as intense as the CO one. However, once anharmonicities are taken into account, the intensity of the OH str decreases dramatically and the reason for the "missing" signal is eventually clarified. Furthermore, the classical anharmonic calculation is in agreement with the ASSCIVR estimate for the frequency of vibration of the OH str.

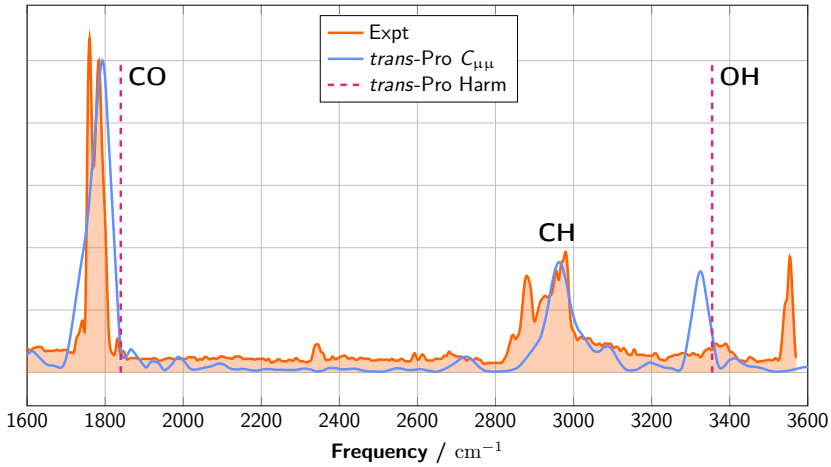


FIGURE 4.4: Trans-*proline: experimental IR spectrum from Ref.[54] (orange area, Copyright © 2001 American Chemical Society), double harmonic approximation for the CO and OH (magenta dashed lines) str, and Fourier transform of the classical dipole-dipole autocorrelation function (blue). Harmonic and classical intensity estimates have been scaled to match the experimental CO str.*

Other signals of low intensity well above $3000\,cm^{-1}$ in the experimental proline spectrum are related to the NH str. The two investigated conformers are characterized by ASSCIVR frequencies for the NH str close to each other at around $3460\,cm^{-1}$. This overestimates by about 70-90 cm^{-1} the experimentally assigned frequencies. I could have obtained a better agreement by running excited trajectories as in the case of the OH str of *cis*1-proline. However, I decided not to pursue this strategy because I could distinguish the two conformers from the OH str and confirm the experimental assignment, without the increase in computational effort required to run the excited trajectories.

Moving to lower frequencies, the experimental band of CH str extends for about $150\,cm^{-1}$ in the frequency domain. Seven normal modes for each con-

former fall in this range that are compatible with CH str. Furthermore, there is an interesting double peak at 2885 and 2984 cm^{-1} (see Figure 4.1). To better understand the origin of this spectroscopic feature I studied the CH str modes of both *trans*- and *cis*1-proline. Calculated harmonic and ASSCIVR frequency values are reported in Table 4.5, confirming the complexity of this band. The *trans*-proline frequencies are generally at higher values than the ones of the *cis*1 conformer. So, while the CH str signals of the *trans*- conformer seem to be responsible for the higher-frequency part of the band, the lower-frequency part is better described by the *cis*1- conformer. This is even more evident from Figure 4.5, where mode 41 is plotted for both conformers. A couple of modes (modes 42 and 43) are estimated at frequencies somewhat higher than the experimentally assigned range, which extends between 2846 and 2984 cm^{-1}. Also in this case, agreement could have been improved by running additional excited trajectories, but this was not necessary since the description of the band given by the two basic ASSCIVR simulations was already satisfactory.

TABLE 4.5: *Harmonic (Harm) and ASSCIVR (AS) frequencies for the seven CH str modes (modes 37-43) of* trans*-proline (column 2 and 3) and* cis*1-proline (column 4 and 5). Calculations were performed at* DFT-D/B3LYP *aVDZ level of theory. All frequencies are in* cm^{-1}.

	trans-proline		*cis*1-proline	
CH str	Harm	AS	Harm	AS
mode 43	3135	3071	3138	3078
mode 42	3117	3060	3109	2953
mode 41	3086	2970	3090	2887
mode 40	3074	2949	3085	2982
mode 39	3057	2938	3065	2920
mode 38	3050	3017	3034	2946
mode 37	3003	2911	2941	2883

The next relevant feature in the vibrational spectrum of proline is represented by the CO str. The experimental spectrum shows again a double peak signal at 1766 and 1789 cm^{-1} (Figures 4.1 and 4.4). However, this time the gap (about 20 cm^{-1}) is much narrower than in the case of the OH str and CH str. In spite of this, my simulations were able to resolve the split. Results are plotted in Figure 4.6 and the ASSCIVR estimates at 1766 and 1799 cm^{-1} are in excellent agreement with the experimental findings at 1766 and 1789 cm^{-1}, respectively.

Finally, I focused on the region around 1400 cm^{-1}, where a double peak structure with frequencies at 1384 and 1405 cm^{-1} is found in the experimental IR spectrum. My ASSCIVR calculations are reported in Figure 4.7. The signals

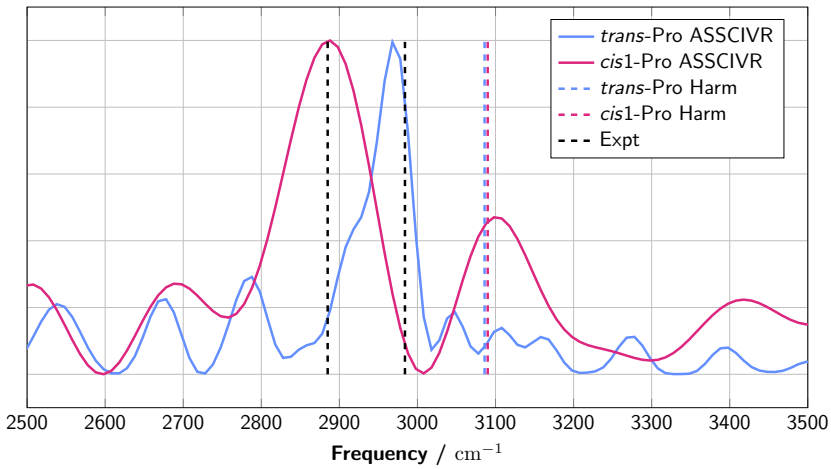


FIGURE 4.5: *ASSCIVR density of states simulations for a representative CH str mode of* trans*-proline (blue) and* cis*1-proline (magenta). The corresponding harmonic frequencies are indicated by same-color vertical dashed lines. Experimental values (from Ref.[54]) are reported with black dashed lines.*

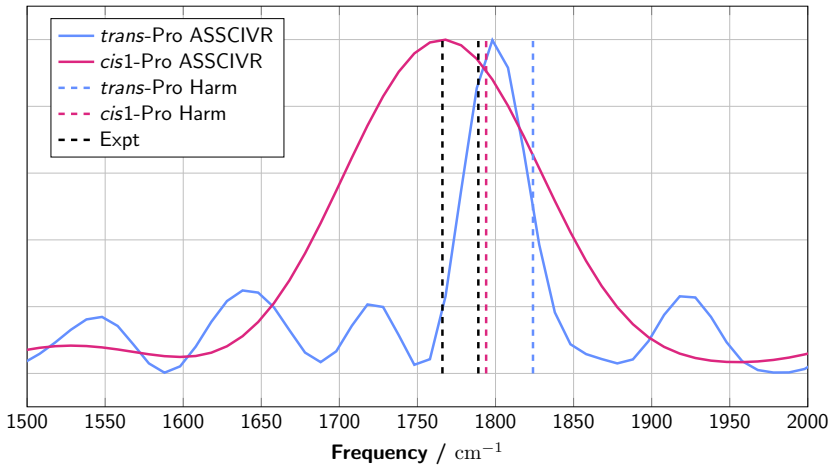


FIGURE 4.6: *ASSCIVR density of states simulations for the CO str of* trans*-proline (blue) and* cis*1-proline (magenta). The corresponding harmonic frequencies are indicated by same-color vertical dashed lines. Experimental values (from Ref. [54]) are reported with black dashed lines.*

correspond to two very mixed motions that I could associate in normal mode approximation to the COH and CNH bending (bnd). My assignment is that the two CNH bnd, one of *trans*-proline and one of *cis*1-proline, are responsible for these two experimental peaks. I assigned the other two low-intensity experimental signals at 1284 and 1412 cm^{-1} to the two COH bnd, with the lower-frequency one belonging to *cis*1-proline. I summarize the main results on *trans*-proline and *cis*1-proline in Table 4.6.

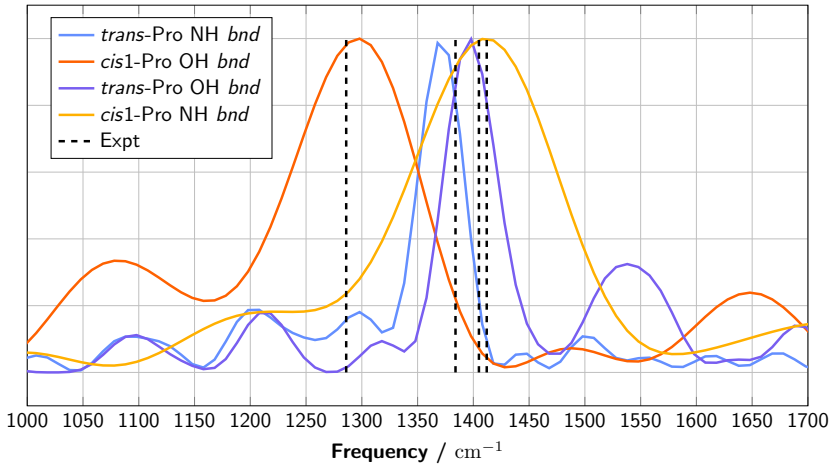


FIGURE 4.7: *ASSCIVR density of states simulations for the COH and CNH bnd of* trans*-proline (violet and blue) and* cis *1-proline (orange and yellow). Experimental values (from Ref.[54]) are reported with black dashed lines.*

TABLE 4.6: *Calculated ASSCIVR frequencies of vibration for* trans*-proline and* cis*1-proline at* DFT-D/B3LYP *aVDZ level of theory. Experimental values from Ref.[54] are given in the last column; [2846,2984] indicates the frequency range of the experimental CH str band. All values are in* cm^{-1}.

Mode	*trans*-proline	*cis*1-proline	Expt.
OH str	3329	3522	3559 (*cis*)
NH str	3461	3467	3393,3369
CH str	3071	3078	
CH str	3060	2953	
CH str	2970	2887	
CH str	2949	2982	[2846,2984]
CH str	2938	2920	
CH str	3017	2946	
CH str	2911	2883	
CO str	1799	1766	1789,1766
NH bnd	1371	1411	1384,1405
OH bnd	1397	1297	1412,1286

4.4 CONCLUSIONS: THE ROLE OF ANHARMONICITY

The only OH str easily assignable was the one referred to a *cis*- type conformer at $3559\,cm^{-1}$. In the literature the OH str of *trans*-proline was assigned at $3025\,cm^{-1}$ as the result of a difference spectrum upon selective deuteration.[54] Instead, I assigned the signal at $3329\,cm^{-1}$ and the dynamical calculation of the absorption intensity hinted at a reduced intensity due to anharmonicity. This explains the lack of any peak at this frequency and further stresses the need of including anharmonic effects in vibrational simulation.

By performing simulations for both *trans*- and *cis*1-proline, I was also able to describe the set of split bands and peaks. For instance, in the case of the CO str a split signal is a clear clue of the presence of at least two conformers, since a single CO bond is present in proline. Likewise, the double-peaked signal associated to the COH/CNH bnd was accurately reproduced.

In general, my on-the-fly ASSCIVR simulations were able to reproduce the experimental results with good accuracy. I stress that my results are fully (45-) dimensional and are mainly based on just two short-time semiclassical trajectories, as established by on-the-fly semiclassical theories. I employed a third trajectory with a quantum of excitation in the OH str mode of *cis*1-proline only to refine the estimate of that specific frequency. Additional ASSCIVR trajectories with one quantum of excitation in the NH and CH stretches would have probably led

to a better agreement with the experiment, but it was not necessary since assignment of those signals was not controversial. The semi-quantitative picture I got for NH and CH stretches is fully satisfactory and running several additional trajectories would have come at a high computational cost. For lower energy signals (CO str and COH/CNH bnd) my frequency estimates are in excellent agreement with the experiment. Finally, I also notice that my simulations provide narrower and more precise peaks for *trans*-proline and wider ones for *cis*1-proline. This is mainly an effect of the more energetically isolated well for the global minimum *trans*-proline.

Chapter 5

The spectrum of guanine-cytosine pair

The content of this chapter is based on the work reported in Ref.[139]

5.1 Tunneling between DNA bases

The duplication process of DNA heavily relies on the complementarity of nucleobases in forming hydrogen bonds. On the one hand, the hydrogen bond allows the DNA molecule to open and close the double helix at occurrence, so that the genetic information can be duplicated by DNA polymerase. On the other hand, it allows for the formation of other tertiary structures, such as B-DNA, A-DNA or even triple-stranded DNA.[140] Different base clusters are also possible, such as the Hoogsteen pairs and G-quadruplexes.[89, 141] Given the presence of keto and amino groups, nucleobases present a great number of tautomers, a characteristic which can induce mismatches during the replication processes.[142] Even if RNA is the nucleic acid most affected by non-canonical pairing,[143] the nucleobase pair with the highest number of stable tautomers is guanine-cytosine, which is in common with DNA. Since the conformational landscape of the GC pair is quite variegated, I adopted the nomenclature established in the experimental literature.[3, 4] According to this nomenclature, the GC pair tautomer is identified by indicating:

1. the guanine tautomer, which can be either enol- (e) or keto- (k)
2. the position of the imidazolic hydrogen of guanine, which is either on nitrogen 7 or 9

3. the tautomer of cytosine which is keto- (K), enol- (E) or imino- (I)

4. a number related to the relative energy of the hydrogen-bond network

In this nomenclature, the GC canonical tautomer — the Watson and Crick (WC) one — is identified as K9K-1 (Figure 5.1, right).

One of the proposed mutagenic mechanisms is initiated by a proton hopping event, in which the two hydrogen atoms involved in the supramolecular bonding hop from one base to the other, as illustrated in Figure 5.2. As a consequence, new tautomers are generated then mistakenly paired during DNA reproduction. However, the debate around the mechanism of this proton hopping is still open.[17, 55, 57, 144–146] On one side, the double proton hopping process is thermally improbable and its products are too short-lived to have any biological impact, even if this aspect does not exclude that the double proton hopping could happen as a defense mechanism, in a deactivation pathway.[57, 147, 148] Indeed, in the case of excited states or ionized base pairs, the proton hopping barrier drops to the point that tautomerization can happen spontaneously.[57, 149, 150] On the other side, the process could occur *via* tunneling, becoming fast enough to make the product short life irrelevant.[145] If this is the case, the hopping tautomer could enter the polymerase λ pocket and be incorporated erroneously.[144] This hypothesis has not been confirmed yet, leaving the discussion about tunneling in DNA bases is open.[55, 151]

Understanding the dynamical properties of the nucleobase pairs is fundamental to unravel the DNA structural and mechanical peculiarities, in particular those related to the hydrogen bond network. The discernment the dynamics of this network is pivotal to develop an accurate model for *in vivo* DNA stability, base tautomerization and mismatch, even more so if quantum effects are to be included. Vibrational spectroscopy is an invaluable tool to investigate soft matter and especially hydrogen-bonded systems.[1, 2, 42, 61, 105, 116, 152–159] Since in aqueous solution the bases prefer the stacked disposition, the hydrogen-bonded structure of the isolated dimer is observed only in gas phase.[6, 160] This means that in the experimental gas-phase spectra, the hydrogen bond contribution to the pair formation can be isolated from the π-stacking and hydrophobic one. The vibrational spectra could also help to clarify if any proton hopping is taking place, by inspecting if specific features of the proton-hopped tautomer are present in the spectra.[3–6, 161–163]

The main limitation of these experiments is the absence of a grounded assignment of the main spectral features, which would allow the clear identification of the corresponding tautomer. Current simulations of DNA bases vibrations are limited to static approximations or low-dimensional models, due to the system intrinsic complexity.[3, 5, 21, 164] These approximations fall short when dealing with hydrogen-bonded systems,[2] and a higher accuracy is necessary to assign the

nucleobase pair experimental spectra, where there is a plethora of tautomers.[3, 4, 161] Specifically, the experimental spectrum of the GC dimer was previously assigned to a high energy tautomer, instead of the expected global minimum one, which is the WC structure.[6, 165] The WC tautomer presents ketonic guanine and cytosine, whereas the high energy conformer has enolic cytosine and a different hydrogen disposition of guanine imidazolic ring (respectively, K9K-1 and K7E-1, shown in Figure 5.1).

The experimental spectrum was obtained by IR-UV hole burning spectroscopy on a mixture of laser-desorbed guanine and cytosine and it is shown in Figure 5.3.[3, 4, 6] IR-UV hole burning consists of irradiating the mixture with an IR "burn" laser pulse followed by a UV "probe" pulse after 100 ns. The UV laser is tuned to an electronic transition that leads to ionization after absorption of another UV photon. If the IR excitation energy coincides with a vibrational excitation, it depletes the ground state, decreasing the ion signal in a mass spectrometer. In this way, one can record sharp IR spectra of mass and optically selected clusters.[6]

The experimental spectra were assigned using the those of the isolated bases and scaled harmonic frequencies at the RI-MP2 TZVPP level of theory.[3, 4, 6] The spectrum obtained by laser ablation of guanine and cytosine was initially assigned as either K7E-1 or K9E-1.[3, 4] Upon investigation of the low frequency region, it was concluded that the compound responsible for the spectrum was K7E-1, given the position of the N^7H bnd signal. In fact, above $500\,cm^{-1}$ K7E-1 presents only the COH in-plane bending, with an harmonic frequency of $508\,cm^{-1}$, whereas N^7H bnd has a scaled harmonic frequency of $477\,cm^{-1}$. Instead, K9E-1 scaled harmonic frequencies predict the COH in-plane bnd at $506\,cm^{-1}$ and the N^9H at $515\,cm^{-1}$. In first approximation, this makes K9E-1 incompatible with the low frequency portion of the desorbed pair spectrum, that has only one signal around $500\,cm^{-1}$.[5]

The first goal of my simulations was to check if the WC structure was present in the experimental spectrum of Figure 5.3. Nir *et al.* based their assignment on a scaled-harmonic approach, which is often a non-reliable technique with no description of quantum effects,[2] and on the isolated bases computational setup spectra, which do not account for inter-molecular effects. Therefore, it was possible that a precise quantum spectrum of the WC tautomer spectrum matches the experimental one, demonstrating the presence of relevant quantum effects. Nonetheless, confirming the absence of the WC tautomer in the spectrum would support the hypothesis that the WC excited state lifetime is unexpectedly short.[3, 6, 166]

Then, I wanted to verify if the vibrational spectrum revealed the fingerprints of quantum effects related to the double proton hopping mechanism. There are two main ways whereby I could point out this phenomenon: I could simulate the double proton hopping product (*i.e.* the enolic form labeled as E9K-1) and

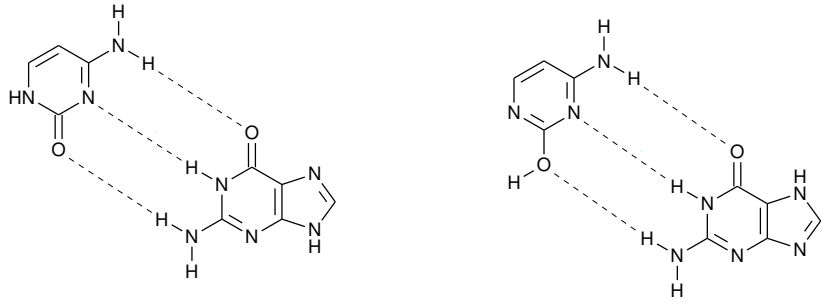


FIGURE 5.1: *On the right: the WC tautomer of GC pair,* κ9κ*-1; on the left: the tautomer present in the IR-UV hole burning spectrum following Abo-Riziq* et al. *assignment.[6] The second tautomer presents a ketonic guanine with the imidazolic hydrogen on* N^7 *and an enolic cytosine, and it is therefore labelled as* κ7ε*-1.*

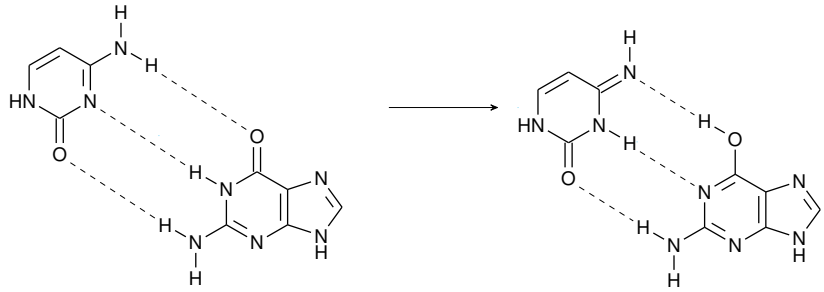


FIGURE 5.2: *Double proton hopping reaction of GC WC tautomer (*κ9κ*-1); the hopping product presents an enolic guanine with the imidazolic hydrogen on* N^9 *and an immino cytosine, and it is therefore labelled as* ε9ι*-1.*

check if it was present in the experimental spectrum; furthermore, I could look for differences in the spectral features between WC tautomer spectra computed with a theoretical method able to point out quantum effects and another method unable to do that.

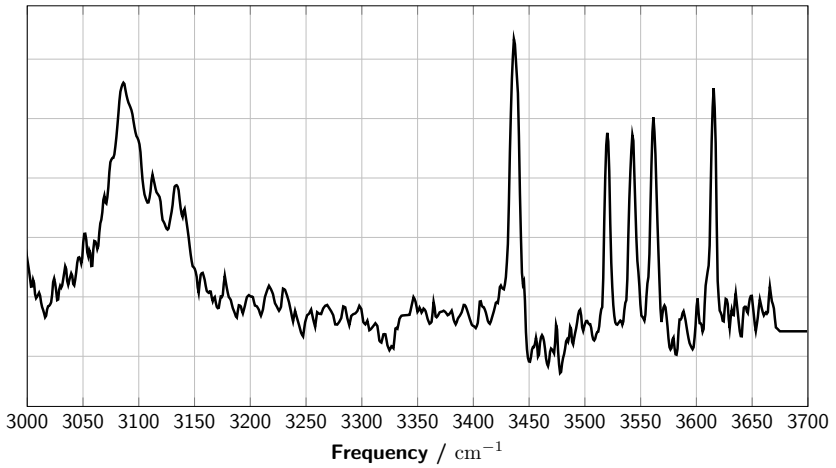


FIGURE 5.3: *Experimental spectrum obtained with IR-UV hole burning on a mixture of laser-desorbed guanine and cytosine, from Ref.[6] (Copyright (2004) National Academy of Sciences, U.S.A.).*

5.2 Simulation setup for QCT and DCSCIVR

To compute the vibrational power spectra of the GC dimer, I employed both QCT and DCSCIVR, described in Sections 2.2 and 2.4. I applied these methods *on-the-fly* using NWChem,[137] avoiding the cumbersome task of building a PES. The choice of these two methods allowed me to limit the computational effort, even when investigating the 29-atom GC pair with a relative expensive level of theory, such as Grimme's DFT-D with B3LYP functional.[138] After a brief test with both 6-31G and aVDZ basis set, I selected the def2-TZVP basis set,[167] being large enough to be comparable to aVDZ, but much faster, as shown in Table 5.1.

The geometry optimization were performed with NWChem `dft optimize` task with parameters selected to obtain a good minimum geometry.[137] A sample optimization input is shown in Listing 5.1. The harmonic frequencies where computed on the optimized geometry using NWChem `dft frequencies` task.

```
1 start k7e-1_aVDZ_D_opt
```

TABLE 5.1: *Comparison between the time required for an optimization step (`dft optimize`) and a harmonic frequency calculation (`dft frequencies`) for GC pair using* DFT-D/B3LYP *with either def2-TZVP (def2) or aVDZ; the aVDZ time is put to one.*

	K9K-1		K7E-1	
	def2	aVDZ	def2	aVDZ
`dft optimize`	1.09	1.00	1.23	1.00
`dft frequencies`	0.33	1.00	0.35	1.00

```
2  echo
3
4  set geom:dont_verify .true.
5
6  geometry noautosym noautoz
7  <xyz geometry>
8  end
9
10 basis spherical
11 * library Def2-TZVP
12 end
13
14 dft
15  grid fine
16  convergence energy 1.00E-07
17  convergence density 1.000000E-05
18  convergence gradient 1E-06
19  tolerances tol_rho 1.00D-11
20  tolerances accCoul 12
21  maxiter 300
22  xc b3lyp
23  disp vdw 3
24 end
25
26 driver
27   maxiter 200
28 end
29
30 task dft optimize
```

LISTING 5.1: *Sample input for guanine-cytosine optimization*

The QCT spectra were obtained with a 25 000 aut trajectory at DFT-D/B3LYP def2-TZVP level of theory, propagated using NWCHEM `dft qmd` task. The trajectory was initialized with the harmonic quantization initial conditions shown in Equation (2.73), so that both QCT and DCSCIVR could be performed on the same trajectory. A sample input for the trajectory evolution in shown in List-

ing 5.2. The initial velocities were fed to NWChem by an in-house modification. Due to trajectory instability, the high frequency modes of K7E-1 where obtained from a trajectory initialized with the normal modes under 800 cm^{-1} "switched off", *i.e.* with initial momentum zero.

```
start k7e-1_def2_D
echo
print low

geometry noautosym noautoz
<xyz geometry>
end

basis
* library Def2-TZVP
end

dft
   maxiter 100
   xc b3lyp
   disp vdw 3
end

qmd
  nstep_nucl  2500
  dt_nucl     10.0
  targ_temp   0.01
  com_step    1
  rand_seed   12345
  thermostat  none
  print_xyz   1
end
task dft qmd
```

Listing 5.2: *Sample input for guanine-cytosine evolution*

The DCSCIVR spectra were obtained only for K9K-1 and K7E-1. To reduce the machine time, I used the HDB method, resulting in only 139 computed Hessian matrices for K9K-1 and 196 for K7E-1. The relevant modes of K9K-1 fell into a 17D and a 7D subspaces, whereas the relevant modes of K7E-1 were grouped into a 5D. a 3D and two 2D subspaces.

5.3 Discussion of the simulated spectra

Based on the literature, I selected four tautomers for my investigation.[3–5] K9K-1 is the canonical and most stable tautomer, *i.e.* the WC tautomer. Then, I considered the K7E-1 tautomer, because its presence was reported in the experimental

spectra. Furthermore, I focused on the K9E-1 tautomer since it is indistinguishable from K7E-1 when looking at the high harmonic frequencies, and it is the most similar to K9K-1. Finally, I considered also the E9I-1 tautomer because it is formed upon double proton hopping starting from K9K-1, and the presence of E9I-1 — even in traces — would suggest the presence of a fast-proceeding double proton hopping mechanism resulting in a stable tautomeric form.[3, 55, 57, 144, 145, 147]

The optimized geometries with relative energies are shown in Figure 5.4. The complete energy analysis is reported in Table 5.2. The relative energies are in good agreement with those reported in the literature, confirming the WC tautomer (*i.e.* K9K-1) as the global minimum of the hydrogen-bonded structure of the GC pair. These four conformers present very different energies, even if the H bond structure is quite similar.

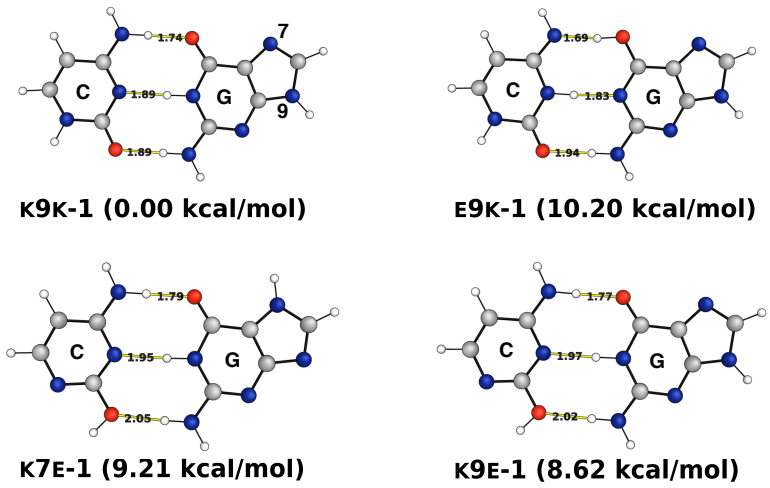


FIGURE 5.4: *Equilibrium geometries and energies (in* kcal mol^{-1}*) of the four lowest-lying GC tautomers, at* DFT-D/B3LYP *def2-TZVP level of theory. The hydrogen bonds are indicated by the yellow lines, with distances in* Å.

The QCT values are compared with the harmonic frequencies in Table 5.3 to appreciate the level of anharmonicity of the system. In the high frequency region, the QCT power spectra of K9K-1 and E9I-1 reported in Figure 5.5 are different, mainly in the guanine NH_2 str. This suggests that the main difference in this frequency region is due to the O···HNH hydrogen bond effects on the guanine NH_2 group. Given that in absence of quantum effects QCT is generally a rather accurate method, this feature provides a way to confirm or deny the presence of E9I-1

TABLE 5.2: *Comparison between the minimum energies (in Hartree) of the four examined tautomers, at theory level* DFT-D/B3LYP *def2-TZVP.* ΔE *is the energy gap from* K9K-1

Tautomer	K9K-1	K7E-1	E9I-1	K9E-1
E / E_h	-937.9355465	-937.9209256	-937.9193547	-937.9218620
ΔE / E_h	0	0.01462088	0.01619172	0.0136845
ΔE / cm^{-1}	0	3209	3554	3003

in K9K-1 gas-phase spectra. Conversely, in Figure 5.6, the QCT power spectra in the high frequencies region of K7E-1 and K9E-1 are basically indistinguishable from each other, within the unavoidable peak width originated by the finite-time Fourier transform. To confirm de Vries and coworkers assignment, based on the low frequency region, I compared the QCT spectra of both K9E-1 and K7E-1 imidazolic NH bnd with the experimentalist scaled harmonic frequencies.[5] This comparison is presented in Table 5.5. Even when properly accounting for mode anharmonicity, K7E-1 presents only one peak around $500\,cm^{-1}$, in accord with the experiment (Figure 5.7).

TABLE 5.3: *Comparison between the QCT frequencies (QCT, in* cm^{-1}*) of the normal modes above* $3300\,cm^{-1}$ *for the four tautomers, and the harmonic frequencies at theory level* DFT-D/B3LYP *def2-TZVP (Harm., in* cm^{-1}*); the QCT frequencies where obtained from a single* 25 000 aut *trajectory at theory level* DFT-D/B3LYP *def2-TZVP. Double harmonic intensities (Int., in* $km\,mol^{-1}$*) are included. G stands for guanine and C for cytosine.*

K9K-1

Motion	Harm.	Int.	QCT
G NH@NH_2 str	3686	117	3621
C NH@NH_2 str	3667	86	3574
G N^9H str	3641	68	3530
C NH str	3613	91	3504
G NH@NH_2 str	3378	1319	3377

E9I-1

Motion	Harm.	Int.	QCT
G NH_2 a.str	3702	136	3550
GN^9H str	3646	69	3545
C NH str	3631	114	3552
C im. NH str	3533	17	3441
G NH_2 s.str	3493	741	3435

K7E-1

Motion	Harm.	Int.	QCT
C OH str	3757	110	3658
G NH_2 a.str	3683	163	3533
C NH_2 a.str	3675	86	3558
G N^7H str	3647	84	3533
G NH_2 s.str	3544	373	3444

K9E-1

Motion	Harm.	Int.	QCT
C OH str	3757	108	3660
G NH_2 a.str	3693	161	3549
C NH@NH_2 str	3674	85	3581
G N^9H str	3640	72	3551
G NH_2 s.str	3539	485	3427

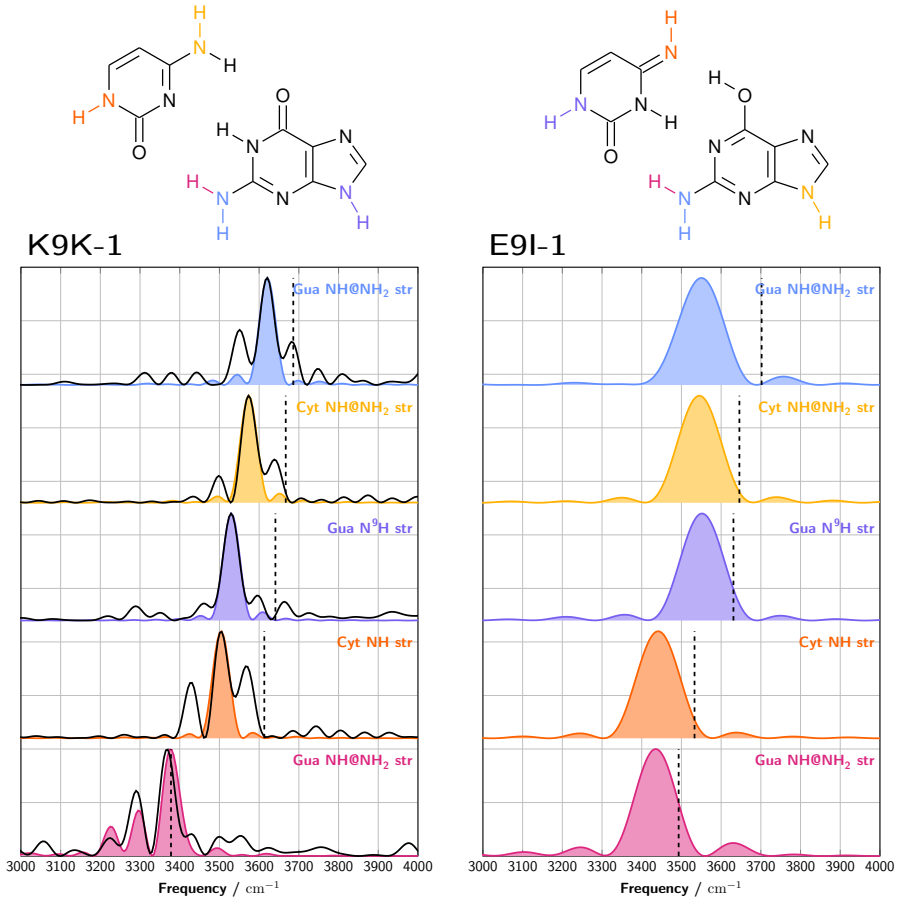


FIGURE 5.5: *Power spectra in the frequency region above* $3300\,\mathrm{cm}^{-1}$ *for the* K9K-1 *and* E9I-1 *tautomers, obtained with QCT on a* 25 000 aut *long trajectory, starting from harmonic conditions at* DFT-D/B3LYP *def2-TZVP level of theory. The harmonic frequencies are reported as black dashed lines. For* K9K-1 *the DCSCIVR spectra are reported in black solid lines. See Tables 5.3 and 5.6 for the numerical values.*

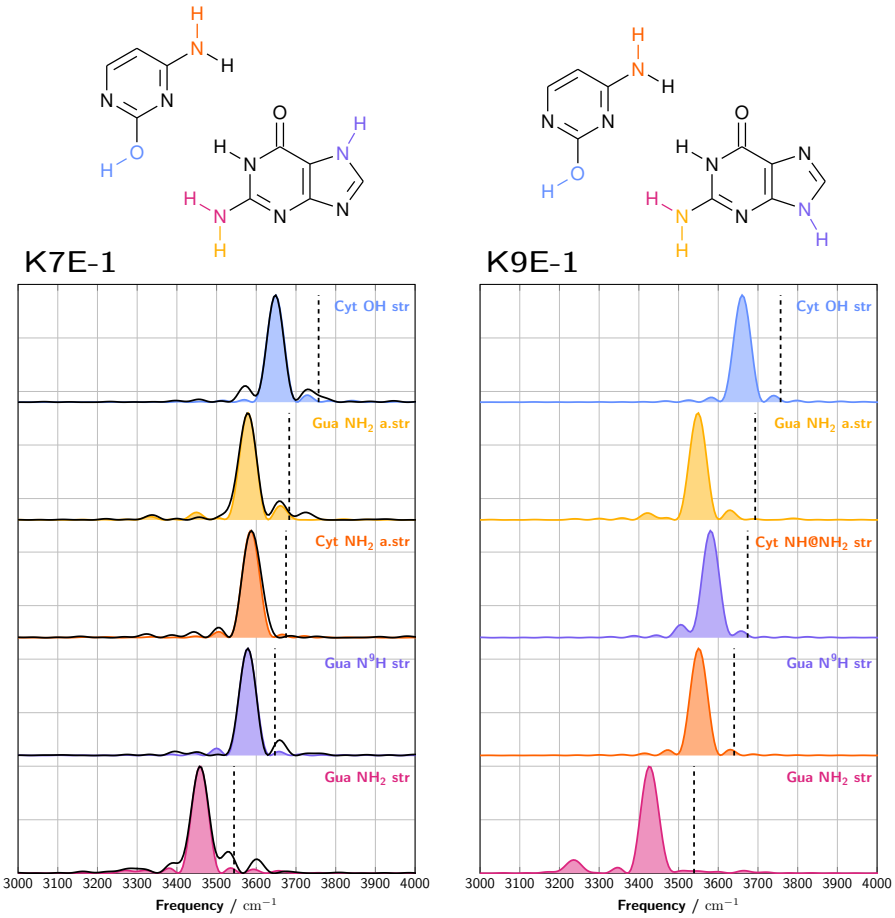


FIGURE 5.6: *Power spectra in the frequency region above* $3300\,\mathrm{cm}^{-1}$ *for the* K7E-1 *and* K9E-1 *tautomers, obtained with QCT on a* 25 000 aut *long trajectory, starting from harmonic conditions at* DFT-D/B3LYP *def2-TZVP level of theory. The harmonic frequencies are reported as black dashed lines. For* K7E-1 *the DCSCIVR spectra are reported in black solid lines. See Tables 5.3 and 5.6 for the numerical values.*

TABLE 5.4: *Comparison between the QCT and scaled harmonic frequencies (in* cm^{-1}*) for the* N^7H *and* N^9H *out-of-plane bnd of* K7E-1 *and* K9K-1 *respectively; scaled harmonic frequencies by Bakker* et al. *[5]*

Mode	QCT	Scal. Harm.
N^7H oop bnd	426	477
N^9H oop bnd	506	515

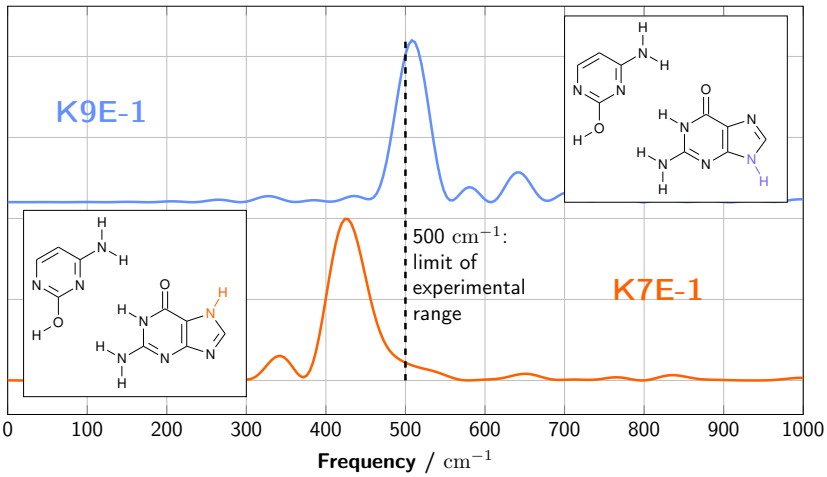


FIGURE 5.7: *QCT spectra and frequencies for the* N^7H *(orange) and* N^9H *(blue) out-of-plane bndof* K7E-1 *and* K9K-1 *respectively; the limit of the experimental range is shown as dashed black line.[5]*

I did not consider for DCSCIVR simulations the E9I-1 tautomer since it is populated only at high energies. Furthermore, the QCT power spectra of E9I-1 are characterized by a much larger FWHM (about $125\,\mathrm{cm}^{-1}$ against about $50\,\mathrm{cm}^{-1}$ for the other tautomers). This is an evidence that E9I-1 is a metastable state and it is about to convert to the WC tautomer in spite of the very short time of the simulation (about 600 fs).

Even if K9K-1 and K9E-1 are lower in energy, I still considered the K7E-1 tautomer because my QCT analysis confirmed the previous assignment in the low energy portion of the spectrum.[3–6, 165] The DCSCIVR spectra are shown in Figures 5.5 and 5.6 as black solid lines and they confirm the QCT spectra values.

The only way the experimentalists had to obtain a vibrational spectrum related to K9K-1 was to lock the tautomer by alkylation, thus collecting the spectrum of *ethyl*-K9-*methyl*-K-1, so I began the assignment from the *ethyl*-K9-*methyl*-K-1 spectrum (Figure 5.8).[3] I made the assumption that alkylation had no other effect in the high-frequency range than removing the corresponding NH stretches. All the spectral features were assigned with good agreement between the experimental and computed frequencies, as shown in the graphical comparison in Figure 5.9 and the analysis is reported in Table 5.6.

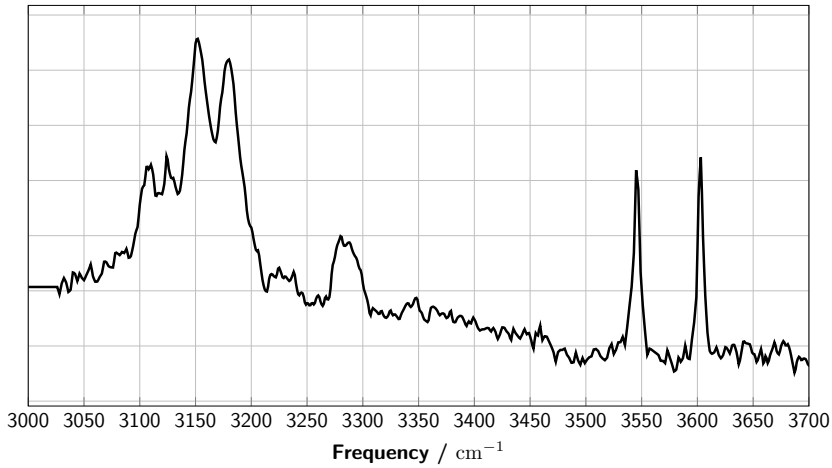


FIGURE 5.8: *Experimental spectrum obtained with IR-UV hole burning on a mixture of laser-desorbed ethyl-guanine and methyl-cytosine, from Ref.[6] (Copyright (2004) National Academy of Sciences, U.S.A.).*

All the computed frequencies are blue-shifted with respect to the experiment. In particular, the guanine NH *str* in NH_2 (Gua NH@NH_2) is associated with the

broad signal around 3280 cm^{-1}. The broad signal is an effect of the strong hydrogen bond, while the larger disagreement between the experimental and DCSCIVR frequency of this vibrational mode (compared to the other modes) is probably due to the chosen DFT-D/B3LYP level of theory. Indeed, this vibrational frequency is very similar for the harmonic approximation, QCT and DCSCIVR estimates, even for different basis sets (Table 5.5). This is quite an unusual behavior and it suggests that either there is an improper description of the potential energy surface or my assumption breaks down and the experimental alkylation makes a difference for this mode. Nonetheless, I was able to assign this feature, since this is the only signal compatible with the guanine NH str in the NH_2 DCSCIVR frequency. In conclusion, I can exclude the presence of the E9I-1-like tautomer, since there is no experimental signal that I can compare to E9I-1 guanine NH_2 stretches. This means that I find no spectral evidence of a double proton hopping mechanism.

TABLE 5.5: *Harmonic (Harm), QCT and DCSCIVR frequencies (in* cm^{-1}*) for K9K-1 guanine NH str in* NH_2*, obtained with 6-31G**, aVDZ and def2-TZVP basis set, using DFT-D/B3LYP level of theory. QCT and DCSCIVR frequencies were obtained with a single* 25 000 au *trajectory at the indicated level of theory.*

Basis set	Harm	QCT	DCSCIVR
6-31G**	3389	3374	—
aVDZ	3356	3321	—
def2-TZVP	3378	3377	3368

I then proceeded with the assignment of the experimental spectrum obtained by laser desorption of guanine and cytosine, starting with the K7E-1 DCSCIVR spectra, as suggested in the literature.[3, 5, 6] The graphical comparison is shown in Figure 5.10 and the analysis is reported in Table 5.6. My assignment of the experimental signals differs from the one previously reported by de Vries and coworkers, and the main reason is that DCSCIVR fully accounts for mode anharmonicity while a rough scaled harmonic approximation has been previously employed by the same authors for their assignment. Specifically, the modes responsible for the three-pronged structure of the experimental spectrum are in different order than the previous experimental assignment: the peak at 3561 cm^{-1} is assigned to cytosine NH_2 str instead of guanine NH_2 asymmetric str; the peak at 3543 cm^{-1} is assigned to guanine N^7H str instead of cytosine NH_2 str; the peak at 3520 cm^{-1} is assigned to guanine NH_2 asymmetric str instead of guanine N^7H str. However, my assignment confirms the experimental hypothesis that K7E-1 is responsible for the IR-UV hole burning spectrum of laser-desorbed GC pairs. The main argument is that the K7E-1 experimental spectrum lacks the broad signal

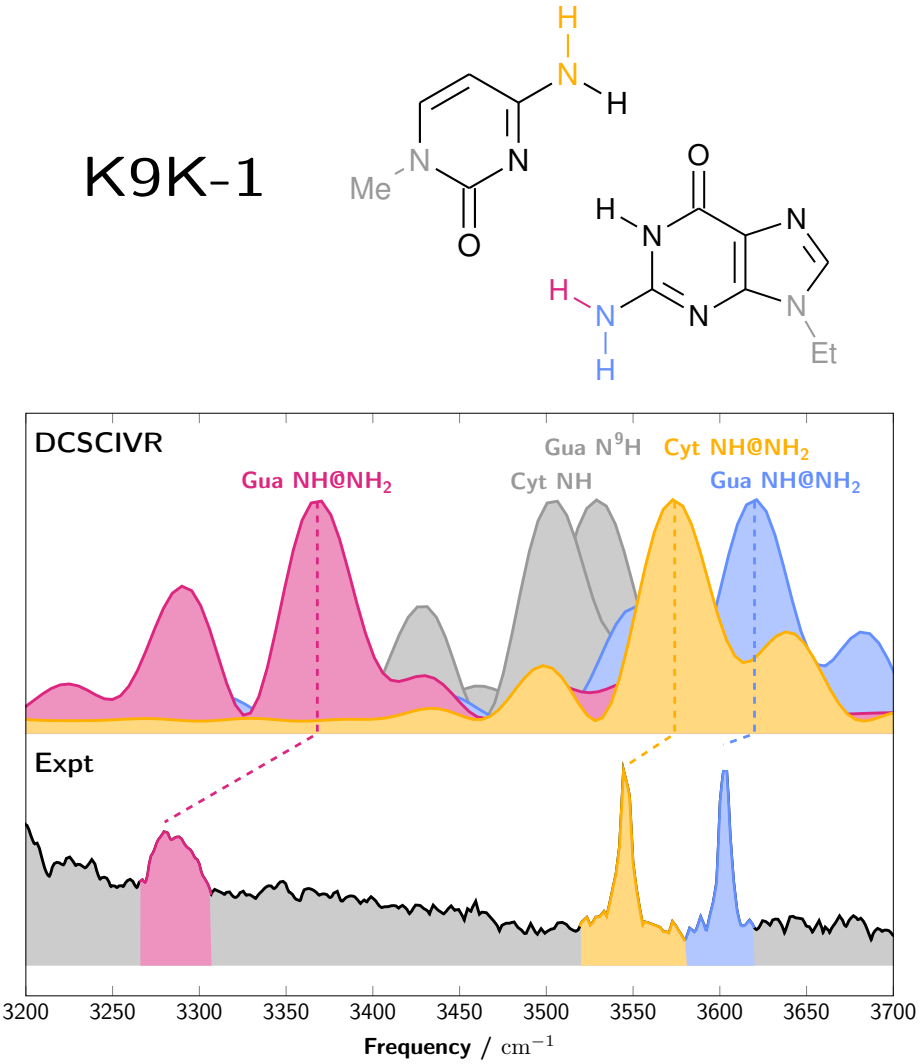


FIGURE 5.9: *Comparison between the experimental spectrum of ethyl-K9-methyl-K-1 and the DCSCIVR spectra of K9K-1, obtained from a single* 25 000 aut *trajectory, starting from harmonic conditions at DFT-D/B3LYP def2-TZVP level of theory. The grey peaks are the NH stretches suppressed by the experimental alkylation. The experimental spectrum is reproduced from Ref.[6]. (Copyright (2004) National Academy of Sciences, U.S.A.)*

around 3280 cm^{-1} present in the *ethyl*-K9-*methyl*-K-1 spectrum and assigned to the K9K-1 guanine NH str in NH_2, which is heavily influenced by the hydrogen bond with the cytosine ketonic function. The experimental spectrum shown in Figures 5.3 and 5.10 is instead characterized by a sharp peak at 3436 cm^{-1}, compatible with the K7E-1 guanine symmetric NH_2 str. We can also exclude E9I-1 because its highest frequency mode, the cytosine NH str at 3552 cm^{-1}, is too low in frequency to match the experimental signal at 3615 cm^{-1}, and our QCT and DCSCIVR frequencies are even blue shifted compared to the experiment.

TABLE 5.6: *Comparison between the DCSCIVR vibrational frequencies and the experimental ones (in* cm^{-1}*) for the main guanine-cytosine pair tautomers. The DCSCIVR frequencies are obtained from a* 25 000 aut *classical trajectory at* DFT-D/B3LYP *def2-TZVP level of theory for each tautomer, whereas the experimental frequencies are taken from the ethyl-K9-methyl-K-1 spectrum and the guanine-cytosine pair spectrum (K7E-1).[6] Mean Absolute Error (MAE) is calculated on only 3 data against the experiment for K9K-1 and on 5 data for K7E-1.*

K9K-1			K7E-1		
Mode	DCSCIVR	Expt.	Mode	DCSCIVR	Expt
Gua NH str in NH_2	3620	3603	Cyt OH str	3656	3615
Cyt NH str in NH_2	3574	3545	Gua NH_2 a.str	3534	3520
Gua N^9H str	3529	alkylated	Cyt NH_2 a.str	3559	3561
Cyt NH str	3504	alkylated	Gua N^7H str	3536	3543
Gua NH str in NH_2	3368	3283	Gua NH_2 s.str	3448	3436
MAE	43		MAE	15	

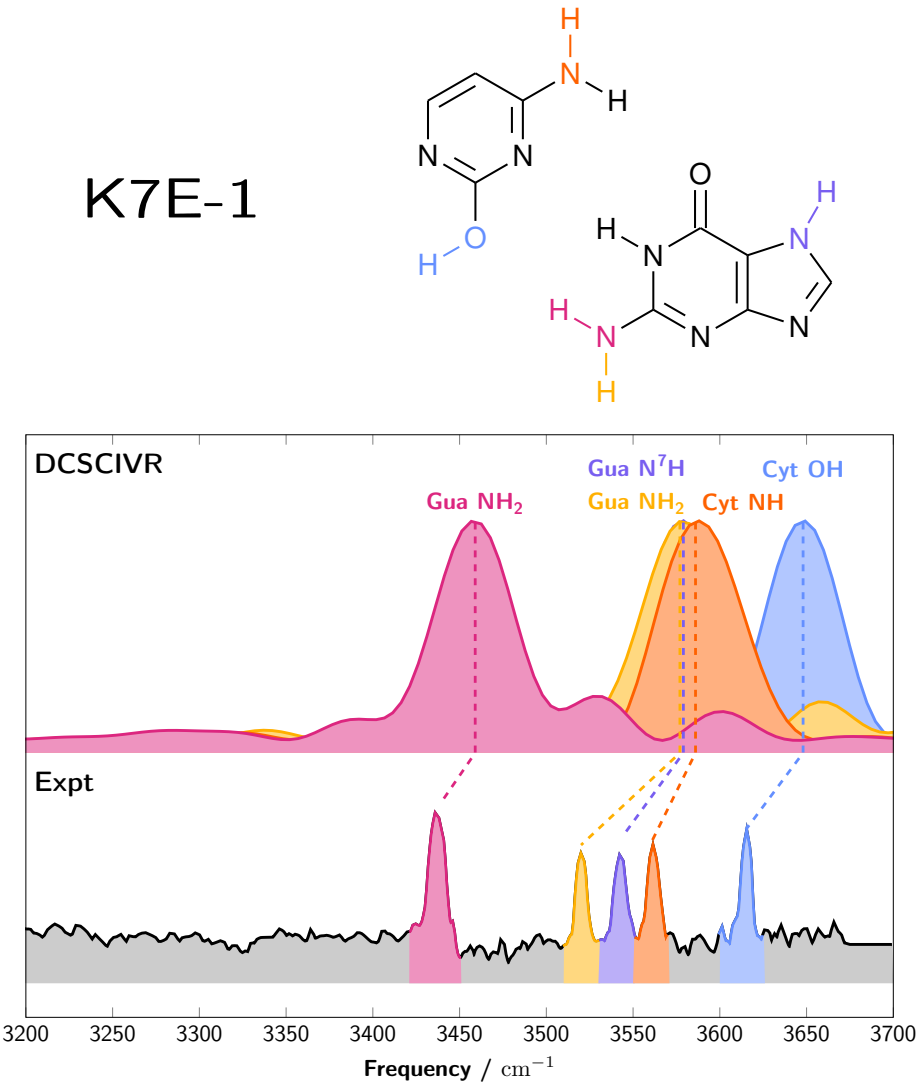


FIGURE 5.10: *Comparison between the experimental spectrum of the guanine-cytosine dimer and the DCSCIVR spectra of* K7E-1*, obtained from a single* 25 000 aut *trajectory, starting from harmonic conditions at* DFT-D/B3LYP *def2-TZVP level of theory. The experimental spectrum is reproduced from Ref.[6]. (Copyright (2004) National Academy of Sciences, U.S.A.).*

5.4 Conclusions: the absence of quantum tunneling

By comparing my results with the experimental spectra found in the literature,[3, 4, 6, 165] I managed to assign the relevant spectral features in the high frequency region of two experimental spectra, one for the isolated guanine-cytosine dimer and the other one for its alkylated form, in agreement with the literature. Indeed, the presence of a peak at 3436 cm^{-1} in the spectrum of the guanine-cytosine dimer is the signature feature of a high energy keto-enol tautomer (K7E-1), since the NH peak in the experimental spectrum of the alkylated Watson-and-Crick tautomer (K9K-1) is at lower frequency. This peak corresponds to the vibration of the guanine NH_2 group, which points towards either the enolic (for K7E-1) or ketonic (for K9K-1) form of cytosine. In K7E-1 the peak has been anticipated at higher frequencies (3448 cm^{-1}), with a longer O···HN distance (2.05 Å), whereas K9K-1 shows the corresponding peak at lower frequencies (3368 cm^{-1}) and a shorter O···HN distance (1.89 Å). In terms of normal modes, K7E-1 is characterized by the usual symmetric/asymmetric motions for the NH_2 str, while in K9K-1 the two NH of the guanine NH_2 are uncoupled and vibrate separately. This means that in K7E-1 the hydrogen bond is weaker than the one in K9K-1, resulting in a smaller perturbation of the original guanine free NH_2 vibrational motion. The fact that the corresponding E9I-1 mode has a QCT frequency of 3435 cm^{-1} and a O···HN distance of 1.94 Å clarifies that the frequency of the guanine NH stretch does not depend only on the hydrogen bond acceptor, but also on the relative position of the two bases.

The absence of the WC form is probably due to the short lifetime of its excited state in the UV pump phase.[3, 6] Theoretical calculations by Domcke *et al.* suggest that an internal conversion coadiuvated by the crossing with two doorway states may occur.[166] In the high frequency region I cannot distinguish between K7E-1 and K9E-1, given that their QCT signals lie well within the peak width. However, this does not jeopardize my conclusions, since K7E-1 and K9E-1 only differ for the position of a hydrogen not involved in the hydrogen bonds.

Eventually, in both experimental spectra I could rule out the presence of the tautomeric form (E9I-1) originated via double proton hopping from the WC form. Indeed, the distinctive signals of this enol-tautomer are absent in the experiments. Specifically, the highest frequency transition — attributed to the cytosine free NH str — is too low in frequency, and the guanine NH_2 asymmetric str is absent. This suggests that no hopping mechanism is detectable under the IR-UV hole burning experiment conditions.

My conclusions are strengthened by the lack of clear differences between the QCT spectra and the DCSCIVR ones, that show equivalent fundamental frequen-

cies of vibrations. This agrees with recent calculations of the kinetic constant for the conversion from E9I-1 to K9K-1 by Angiolari *et al.*,[56] which illustrate that this process is very fast and influenced by the environment. If the mutagenesis of base coupling was mediated by the double proton hopping tunneling, it would happen much faster than the conversion to the WC form. Quantum effects are usually more evident in the gas phase than in solvated systems, so my investigation hints at the possibility that quantum effects do not play a major role for DNA in solution, as they do not in gas phase. However, for a better characterization of the mechanism, future spectroscopic investigations (both at experimental and theoretical level) should include the role of the environment. In case of simulations, this opens the question of how many DNA bases should be stacked above and below a GC pair, for it to behave like in a real strand.

Chapter 6

The need for quantum effects

The content of this chapter is based on the work reported in Ref.[115]

6.1 Quantum methods for quantum effects

In the previous chapters, I employed methods capable of introducing quantum effects into vibrational simulations, solving some open issues in spectroscopy, but I found no evidence of them, as far as the fundamental vibrational frequencies were concerned. This gives a rather partial picture of classical methods shortcomings and semiclassical methods capabilities: one could think that classical methods are good enough for the vibrational spectroscopy of biomolecular systems.

However, this not the case: semiclassical spectroscopy detected quantum effects even in fundamental frequency of vibration, as in the case of protonated glycine.[61] Moreover, the inclusion of quantum effects is fundamental to go beyond the reproduction of the main spectral features. For this reason, I collaborated with other members of the group to demonstrate that to have a quantum spectrum, one must use a method that not only is derived from Schrödinger's equation, but is also tested against explicit quantum features (such as ZPE) or experimental spectra of system with conclamated quantum effects.

To do so, we investigated the vibrational spectrum of H_2CO with classical molecular dynamics, quasiclassical trajectory and adiabatic switching, comparing the results with the adiabatic switching semiclassical initial value representation and the quantum mechanical frequencies. All the methods were run on the quartic force field by Martin *et al.*,[168] to remove any influence and time constrain

regarding the potential, including the variational method employed by Carter *et al.* to obtain the quantum mechanical frequencies.[169] Both MD and QCT rely on the velocity autocorrelation function formalism illustrated in Section 2.2, but they differ in the way their trajectories are initialized: for the former, in a Maxwell-Boltzmann distribution that undergoes thermalization using a stochastic velocity rescaling algorithm;[170] for the latter, in a Husimi distribution centered on the target harmonic state energy.[171] Each MD trajectory consisted in a 200 000 au long thermalization, followed by a production run 30 000 au, whereas the QCT were evolved for 25 000 au. Since it was my task in this work, the AS procedure is detailed in Section 6.2. The theory behind AS is included in Section 2.7.

On the other hand, ASSCIVR trajectories are initialized using the phase-space distribution centered on the anharmonic quantum energies obtained *via* AS. The semiclassical power spectrum was then computed using the TASCIVR formulation introduced in Section 2.3. The switching trajectories were evolved for 50 000 au and were followed by 50 000 au long semiclassical trajectories. Due to the monodromy matrix instability against a strict rejection criterion, only around 500 trajectories were used to collect the semiclassical power spectrum.

Finally, the results on H_2-tagged protonated glycine were used as a demonstration that even single-trajectory methods such as on-the-fly divide-and-conquer semiclassical initial value representation are capable of describing quantum effects in system of biological interest.[61]

6.2 Adiabatic quantization

I evolved the AS trajectories following the procedure described in Section 2.7. An initial run was performed by evolving 9000 switching trajectories for 50 000 au on an analytical PES. The switching procedure is illustrated in Figure 6.1. Each set of trajectories was initialized at the harmonic energy of either the ZPE or the first excited state of every normal mode. The final energies were then enveloped using Gaussian functions with $\sigma = 7.3\,\mathrm{cm}^{-1}$, obtaining the distributions shown in Figures 6.2 and 6.3. Finally, the envelope maxima are taken as the AS quantized energies. From the comparison presented in Table 6.1, one can appreciate that AS is very close to the experiment and it is the only classical method that is capable of giving a ZPE estimate, starting from quantized initial conditions. However, the short switching trajectories mixed together mode 5 and 6 (symmetric and asymmetric CH str). This mixing is a known shortcoming of the AS procedure, coming from a non-adiabatic (*i.e.* too quick) switching. This makes AS very cumbersome to use on systems with a high vibrational states density, because discriminating between really close states requires longer and longer trajectories.

To improve the description of the two CH str, I repeated the AS procedure, evolving the 9000 trajectories for 200 000 au. The increase in computational time

is negligible when compared with that required for the ASSCIVR simulation, that is weighted down by the computation of the instantaneous Hessian matrix. The new results demonstrate that the longer the switching time, the better the mode resolution (Table 6.1 and Figure 6.3).

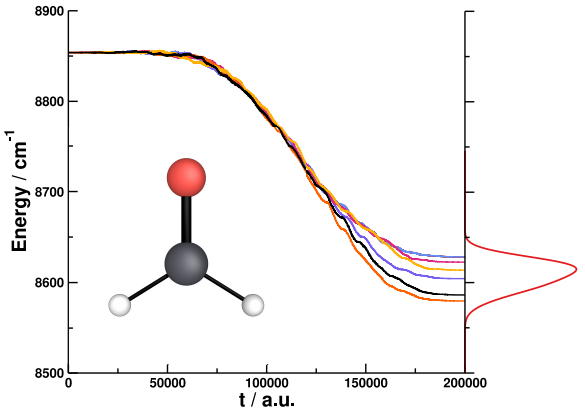


FIGURE 6.1: *Schematic representation of the AS procedure: a set of trajectories is initialized using harmonic conditions and each trajectory is evolved under Equation* (2.113)*; the final energies are enveloped in a Gaussian distribution. From Ref.[115].*

TABLE 6.1: *H_2CO energy distribution maxima (AS, in* cm^{-1}*) obtained after 9000 switching trajectories of* 50 000 au *and* 200 000 au*; each trajectory set was initialized at the harmonic quantum state of the target normal mode; to compare with the harmonic (Harm. in* cm^{-1}*) and quantum mechanical (QM in* cm^{-1}*, from [169]) frequencies, the ZPE value was subtracted from the vibrational energies; between parenthesis, the second maximum value.*

50 000 au			
Mode	QM	Harm.	AS
1	1171	1192	1159
2	1253	1275	1247
3	1509	1543	1505
4	1750	1781	1749
5	2783	2929	2736 (2828)
6	2842	2996	2805 (2886)
zpe	—	5858	5764

200 000 au			
Mode	QM	Harm.	AS
1	1171	1192	1160
2	1253	1275	1247
3	1509	1543	1507
4	1750	1781	1749
5	2783	2929	2792
6	2842	2996	2851
zpe	—	5858	5764

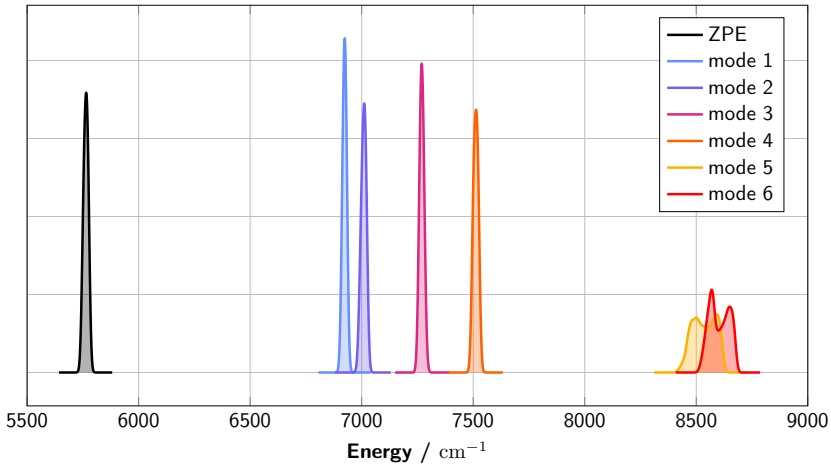


FIGURE 6.2: *H_2CO energy distributions obtained after 9000 switching trajectories of* 50 000 au*; each trajectory set was initialized at the harmonic quantum state of the target normal mode; the distributions were obtained as a Gaussian envelope with* $\sigma = 7.3\,\mathrm{cm}^{-1}$.

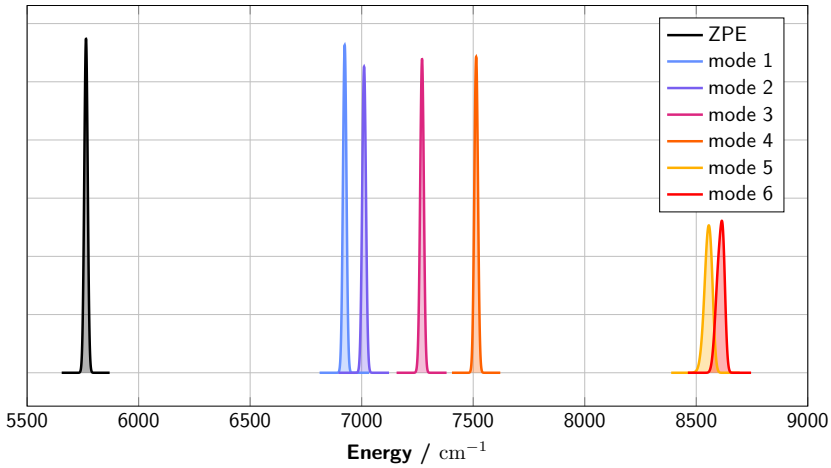


FIGURE 6.3: *H_2CO energy distributions obtained after 9000 switching trajectories of* 200 000 au*; each trajectory set was initialized at the harmonic quantum state of the target normal mode; the distributions were obtained as a Gaussian envelope with* $\sigma = 7.3\,\mathrm{cm}^{-1}$.

6.3 Conclusions: the quantum method touchstone

The comparison between the classical spectra of H_2CO mode 6, obtained by running QCT at different energies and MD at different temperatures is reported in Figure 6.4. It is clear that the classical simulation returns a harmonic estimate of the vibrational frequencies when is performed at low temperature or low energy. Indeed, a system with very little energy can only explore the bottom of the well, in which the harmonic approximation is still valid (Section 2.1). Running the MD at the experiment temperature is not enough, since classical mechanics does not consider the ZPE and the system is still confined in the harmonic portion of the potential. It is interesting that the QCT spectrum with the excited trajectory (ν_6 exc. in Figure 6.4) is closer to the quantum one than the one obtained with energy equal to 1.5 the ZPE, even if the amount of energy in the system is grossly the same. This suggests that the outcome of a classical method is not only influenced by the energy level, but also by the energy distribution.

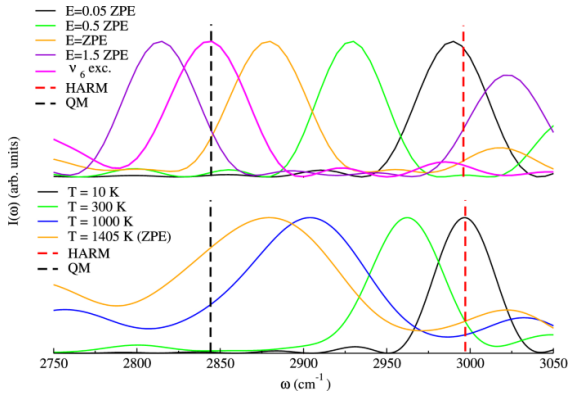


Figure 6.4: *Comparison between the classical spectra of H_2CO mode 6 (obtained under different conditions), the harmonic and the quantum mechanical frequencies;* above: *the QCT spectra;* below: *the MD spectra. From Ref.[115].*

AS proved a formidable quantization method, capable of obtaining very reliable estimates (including that for ZPE) at a fraction of other methods cost. Even so, the state mixing hinders its application on more complex systems. Moreover, it heavily relies on a large number of trajectories to obtain a significant distribution of final energies. This relegates AS to systems for which an analytical PES is available.

The test on H_2CO clearly shows that anharmonicity can be detected with

classical methods, when they are used correctly. However, when looking at the comparison in Figure 6.5, one can notice that the ASSCIVR has many more spectral features, corresponding to the ZPE signal (set to zero), overtones and combinations bands. These are quantum features which do not fall at the sum and difference of the fundamental frequencies, so they cannot be mistaken for the *higher harmonics* that are sometimes present in the classical spectra. The semi-classical treatment, descending from Schrödinger equation *via* the Van Vleck's propagator, does not depend on temperature in its frequency estimates and goes beyond the simple AS quantization. Indeed, ASSCIVR is able to catch the spectral features of regions where AS suffers more from state mixing, because of the high density of vibrational states. Finally, AS cannot compute eigenfunctions, while any full-dimensional method based on SCIVR can.[104, 106, 108, 112] Nonetheless, AS low computational effort makes it a perfect initialization procedure for SCIVR. In this fashion, AS can be even employed on-the-fly as shown in Chapters 3 and 4.

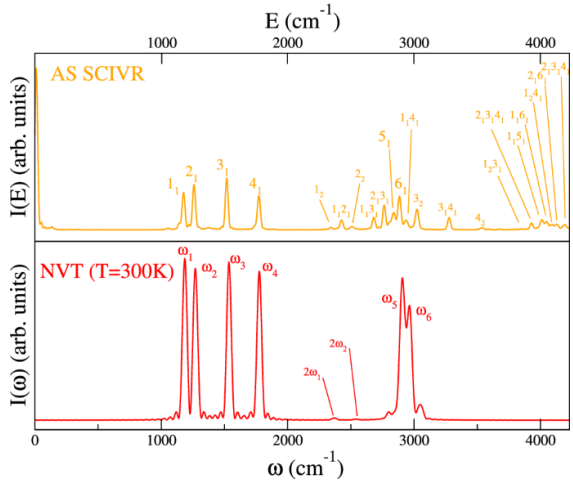


FIGURE 6.5: *Comparison between the ASSCIVR spectra (*above*) of H_2CO and the MD one at* 300 K *(*below*); the ASSCIVR has been translated by the ZPE value to be compared with the MD one. From Ref.[115].*

To conclude, this work plainly shows the necessity to use quantum methods to include quantum effects in vibrational spectroscopy. Even if classical methods like QCT produce reliable fundamental frequencies by accounting for anharmonicity,[2] it is only by including nuclear quantum effects that one can hope to unravel the most interesting features of IR spectra and — hopefully — NIR spectra.

Chapter 7

Conclusions: the past, the present and the future

My Ph.D. had three objectives: improving the semiclassical spectroscopy toolset; applying semiclassical spectroscopy to high dimensional systems with the best level of electronic theory allowed; proving the benefits of treating the vibrational problem with a semiclassical method.

In Chapter 3 I described the implementation of AS to generate the initial conditions for an on-the-fly semiclassical trajectory. The use of AS as preparation step for semiclassical dynamics finds its roots in Heller and de Leon's work assessing semiclassical exactness.[85] This led to successful test on PES.[75] With my implementation, the switching trajectory can now be evolved using an *ab initio* software with a computational cost comparable with that of a plain one. In the case of glycine, on-the-fly ASSCIVR gives cleaner spectra than the semiclassical state-of-the-art method, MCSCIVR, which is initialized using the harmonic quantization. The implementation details were published in Ref.[105].

While working on ASSCIVR, I could not assign the experimental spectrum of proline using only the absolute minimum conformer.[54] This prompted a conformational analysis, followed by the power spectra computation of two proline conformers, using ASSCIVR. I assigned the spectrum to two conformer, one being the absolute minimum. What carried the previous assignment astray was the OH str signal of the absolute minimum conformer, that was believed hidden in the CH str band in disaccord with the ASSCIVR prediction. Using the classical dipole-dipole autocorrelation spectrum, I discovered that this OH str does not appear in the spectrum because of a decrease in intensity due to anharmonic effects and probably quantum effects. Proline OH str is heavily affected by an intramolecular H-bond, which is also what makes proline such an interesting catalyst for aldolic

reactions.[131, 135, 172] The work on proline is described in Chapter 4 and a summary was published in Ref.[116]. By using ASSCIVR, I was able to compute the full-dimensional spectrum of 17-atom proline without any DC method and without any loss in signal-to-noise ratio. This shows that on-the-fly ASSCIVR improves upon MCSCIVR, allowing the treatment of larger systems without the additional cost of DCSCIVR, and without compromising of the full-dimensional description.

When treating even larger systems, the DCSCIVR is still unavoidable, due to some drawbacks of AS such as the vibrational state mixing and the introduction of spurious rotational coupling. This is why I used DCSCIVR to investigate the presence of tunneling in the vibrational spectrum of guanine-cytosine pair.[6] In Chapter 5, I describe the conformational analysis and subsequent spectral assignment. Using a combination of QCT and DCSCIVR, I was able to exclude the presence of both the quantum effects and the tunneling reaction product in the experimental spectrum. By comparison with the spectrum of the alkylated Watson and Crick tautomer, I confirmed the previous assignment, that attributed the experimental spectrum to a higher energy conformer. The 29-atom GC pair was treated using DFT-D/B3LYP def2-TZVP, thus being the larger system on which semiclassical spectroscopy was carried out with a triple valence basis set. The results where composed into Ref.[139].

At the end of my Ph.D. I collaborated on a comparison between classical and semiclassical methods applied to vibrational spectroscopy. Even for a system such as H_2CO, the classical methods proved incapable of reproducing quantum features such as ZPE, overtones and combination bands. Moreover, even if classical methods are fully able to account for anharmonicity, their results heavily depended on the energy (for NVE) or on the temperature (for NVT) of the simulation. Indeed, if the energy of the system is too low, the phase-space sampling is limited to the harmonic portion of the potential and the spectrum will provide a harmonic estimate. The same goes for the temperature. The best initial conditions to obtain a classical spectrum are those of QCT, based on the harmonic quantization. The only classical method able to reproduce the quantum ZPE was AS quantization, described in Section 2.7. Nonetheless, AS cannot easily resolve vibrational states that are very close together, and it is best employed as a sampling method for semiclassical dynamics. This comparison was published in Ref.[115], establishing a possible method to discriminate classical and quantum methods in vibrational spectroscopy.

Nonetheless, the low computational cost, easiness of application and good accuracy make QCT a valuable method to compute the anharmonic frequencies of fundamental vibrations.[2] This is why I opted for QCT in investigating the H-bond effects in $DMSO(H_2O)_n$ cluster, in a work-in-progress with University of Perugia. DMSO and H_2O mixtures present a very low eutectic point of freezing,

which still has no solid explanation. With the tools provided by experimental and theoretical spectroscopy, I hope to shine a light on the role of H-bond in the chemico-physical properties of these mixtures.

The vibrational spectroscopy of organic compound hydrates is quite a vibrant field. Suhm and coworkers recently reported a rather universal Darling-Dennison resonance in the IR spectra of carbonyl monohydrates.[1, 173] The resonance causes an intensity borrowing between the H-bonded OH str and the combination band between a water libration and the water bending overtone. Since this resonance involves a combination band that includes an overtone, it cannot be reproduced using a classical method. I believe that the description of this resonance using SCIVR could be a significant stepping stone for this family of methods and for theoretical spectroscopy in general. Indeed, this work can establish a method to move SCIVR beyond the IR and into the NIR region, where only overtones and combination bands can be detected.

Finally, I would like to continue my work on the study of anharmonic effects in quantum and classical systems. Indeed, new motion properties emerge when going from the second order expansion of kinetic and potential energy to successive approximations. As shown in Section 2.1, in the first case the equations of motion are linear, so we speak of *linear* oscillations. By improving the approximation, we are bound to observe the *non linearity* or *anharmonicity* of the oscillations. A common way to solve the equations of motion for a set of anharmonic oscillators uses a perturbative approach.[90, 174] In this case, the solution for each oscillator assumes the shape:

$$Q_\alpha = Q_\alpha^0 + Q'_\alpha \tag{7.1}$$

where the positions of the α-th oscillator, Q_α are expanded so that $Q_\alpha^0 \gg Q'_\alpha$ and

$$\ddot{Q}_\alpha^0 + \omega_\alpha^2 Q_\alpha^0 = 0 \tag{7.2}$$

So that Q_α^0 represent an ordinary harmonic oscillation:

$$Q_\alpha^0 = a_\alpha \cos(\omega_\alpha t + \alpha_\alpha) \tag{7.3}$$

This solution can be substituted into the equation of motion, leaving only the terms in $\mathbf{Q}^0 = \left\{Q_\alpha^0\right\}$ inside of the non-homogeneous term f_α:

$$\ddot{Q}'_\alpha + \omega_\alpha Q'_\alpha = f_\alpha(\mathbf{Q}^0, \dot{\mathbf{Q}}^0, \ddot{\mathbf{Q}}^0) \tag{7.4}$$

This is a non-homogeneous ordinary equation, whose right-hand side can be written as the sum of simple periodic functions, as follows:

$$Q_\alpha^0 Q_\beta^0 = a_\alpha a_\beta \cos(\omega_\alpha t + \alpha_\alpha)\cos(\omega_\beta t + \alpha_\beta) \tag{7.5}$$

$$= \frac{a_\alpha a_\beta}{2}\left\{\cos\left[(\omega_\alpha + \omega_\beta)t + \alpha_\alpha + \alpha_\beta\right] + \cos\left[(\omega_\alpha - \omega_\beta)t + \alpha_\alpha - \alpha_\beta\right]\right\} \tag{7.6}$$

Thus, the second member contains terms that correspond to oscillations with frequencies equal to the algebraic sum of the natural oscillations of the system. These frequencies are known as combination frequencies. By progressing with the approximation or by adding more oscillators, more combination frequencies appear and beyond the third order these anharmonic oscillations can have the same frequency of a fundamental (for example: $\omega_\alpha + \omega_\beta - \omega_\beta$), inserting resonance terms that are not physical. These resonances are removed by expanding in series the fundamental frequencies.[90]

In vibrational spectroscopy, these combination frequencies appear in spectra obtained using a classical dynamics method and are usually mistaken with *quantum* combination bands and overtones. However, a classical system only has as many frequencies as oscillators, and the presence of combinations is exclusively due to the use of a harmonic base (the Fourier transform) to describe an anharmonic motion.

In the future, I would like to build upon this theoretical insight. In particular, I want to find the combination frequencies in the velocity-velocity autocorrelation spectra of a set of analytical and molecular systems, and then compare these spectra with those obtained with TASCIVR and DVR. I expect to find classical combination frequencies in the classical spectra that have no counterpart in the semiclassical and DVR spectra, that would instead represent the quantum state corresponding to combination bands and overtones.

Bibliography

(1) Fischer, T. L.; Wagner, T.; Gottschalk, H. C.; Nejad, A.; Suhm, M. A. *J. Phys. Chem. Lett.* **2021**, *12*, 138–144.

(2) Fischer, T. L. et al. *Phys. Chem. Chem. Phys.* **2023**, *25*, 22089–22102.

(3) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. *Phys. Chem. Chem. Phys.* **2002**, *4*, 732–739.

(4) Nir, E.; Plützer, C.; Kleinermanns, K.; de Vries, M. *Eur. Phys. J. D* **2002**, *20*, 317–329.

(5) Bakker, J. M.; Compagnon, I.; Meijer, G.; von Helden, G.; Kabeláč, M.; Hobza, P.; de Vries, M. S. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2810–2815.

(6) Abo-Riziq, A.; Grace, L.; Nir, E.; Kabelac, M.; Hobza, P.; de Vries, M. S. *Proc. Natl. Acad. Sci.* **2005**, *102*, 20–23.

(7) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(8) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(9) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(10) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

(11) Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497–1506.

(12) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(13) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063.

(14) Esser, A.; Belsare, S.; Marx, D.; Head-Gordon, T. *Phys. Chem. Chem. Phys.* **2017**, *19*, 5579–5590.

(15) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. *J. Chem. Theory Comput.* **2018**, *14*, 2084–2108.

(16) Conte, R.; Houston, P. L.; Qu, C.; Li, J.; Bowman, J. M. *J. Chem. Phys.* **2020**, *153*, 244301.

(17) Löwdin, P.-O. *Rev. Mod. Phys.* **1963**, *35*, 724–732.

(18) Colbert, D. T.; Miller, W. H. *J. Chem. Phys.* **1992**, *96*, 1982–1991.

(19) Qu, C.; Bowman, J. M. *J. Phys. Chem. Lett.* **2018**, *9*, 2604–2610.

(20) Barone, V.; Biczysko, M.; Bloino, J.; Borkowska-Panek, M.; Carnimeo, I.; Panek, P. *Int. J. Quantum Chem.* **2012**, *112*, 2185–2200.

(21) Fornaro, T.; Biczysko, M.; Monti, S.; Barone, V. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10112–10128.

(22) Jäckle, A.; Meyer, H.-D. *J. Chem. Phys.* **1998**, *109*, 3772–3779.

(23) Wang, H.; Thoss, M. *J. Chem. Phys.* **2003**, *119*, 1289–1299.

(24) Meng, Q.; Meyer, H.-D. *J. Chem. Phys.* **2013**, *138*, 014313.

(25) Sasmal, S.; Schröder, M.; Vendrell, O. *J. Chem. Phys.* **2024**, *160*, 064109.

(26) Witt, A.; Ivanov, S. D.; Shiga, M.; Forbert, H.; Marx, D. *J. Chem. Phys.* **2009**, *130*, 194510.

(27) Miller, W. H. *Proc. Natl. Acad. Sci.* **2005**, *102*, 6660–6664.

(28) Miller, W. H. *J. Phys. Chem. A* **2001**, *105*, 2942–2955.

(29) Heller, E. J. *Acc. Chem. Res.* **1981**, *14*, 368–375.

(30) Kay, K. G. *Chem. Phys.* **2006**, *322*, 3–12.

(31) Pollak, E. In *Quantum Dynamics of Complex Molecular Systems*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007, pp 259–271.

(32) Shalashilin, D. V.; Child, M. S. *J. Chem. Phys.* **2001**, *115*, 5367–5375.

(33) Grossmann, F. *J. Chem. Phys.* **2006**, *125*, 014111.

(34) Conte, R.; Ceotto, M. In *Quantum Chemistry and Dynamics of Excited States*; Wiley: 2020, pp 595–628.

(35) Ma, X.; Di Liberto, G.; Conte, R.; Hase, W. L.; Ceotto, M. *J. Chem. Phys.* **2018**, *149*, 164113.

(36) Begušić, T.; Vaníček, J. *J. Chem. Phys.* **2020**, *153*, 184110.

(37) Vaníček, J. *J. Chem. Phys.* **2023**, *159*, 014114.

(38) Ceotto, M.; Atahan, S.; Tantardini, G. F.; Aspuru-Guzik, A. *J. Chem. Phys.* **2009**, *130*, 234113.

(39) Ceotto, M.; Di Liberto, G.; Conte, R. *Phys. Rev. Lett.* **2017**, *119*, 010401.

(40) Gandolfi, M.; Rognoni, A.; Aieta, C.; Conte, R.; Ceotto, M. *J. Chem. Phys.* **2020**, *153*, 204104.

(41) Cazzaniga, M.; Micciarelli, M.; Moriggi, F.; Mahmoud, A.; Gabas, F.; Ceotto, M. *J. Chem. Phys.* **2020**, *152*, 104104.

(42) Bertaina, G.; Di Liberto, G.; Ceotto, M. *J. Chem. Phys.* **2019**, *151*, 114307.

(43) Di Liberto, G.; Conte, R.; Ceotto, M. *J. Chem. Phys.* **2018**, *148*, 014307.

(44) Di Liberto, G.; Conte, R.; Ceotto, M. *J. Chem. Phys.* **2018**, *148*, 104302.

(45) Buchholz, M.; Grossmann, F.; Ceotto, M. *J. Chem. Phys.* **2017**, *147*, 164110.

(46) Buchholz, M.; Fallacara, E.; Gottwald, F.; Ceotto, M.; Grossmann, F.; Ivanov, S. D. *Chem. Phys.* **2018**, *515*, 231–235.

(47) Buchholz, M.; Grossmann, F.; Ceotto, M. *J. Chem. Phys.* **2018**, *148*, 114107.

(48) Sun, Q.; Bowman, J. M.; Gazdy, B. *J. Chem. Phys.* **1988**, *89*, 3124–3130.

(49) Huang, J.; Valentini, J. J.; Muckerman, J. T. *J. Chem. Phys.* **1995**, *102*, 5695–5707.

(50) Qu, C.; Bowman, J. M. *J. Phys. Chem. A* **2016**, *120*, 4988–4993.

(51) Nagy, T.; Lendvay, G. *J. Phys. Chem. Lett.* **2017**, *8*, 4621–4626.

(52) Bose, A.; Makri, N. *J. Chem. Theory Comput.* **2018**, *14*, 5446–5458.

(53) Conte, R.; Parma, L.; Aieta, C.; Rognoni, A.; Ceotto, M. *J. Chem. Phys.* **2019**, *151*, 214107.

(54) Stepanian, S. G.; Reva, I. D.; Radchenko, E. D.; Adamowicz, L. *J. Phys. Chem. A* **2001**, *105*, 10664–10672.

(55) Slocombe, L.; Winokan, M.; Al-Khalili, J.; Sacchi, M. *J. Phys. Chem. Lett.* **2023**, *14*, 9–15.

(56) Angiolari, F.; Huppert, S.; Pietrucci, F.; Spezia, R. *J. Phys. Chem. Lett.* **2023**, *14*, 5102–5108.

(57) Soler-Polo, D.; Mendieta-Moreno, J. I.; Trabada, D. G.; Mendieta, J.; Ortega, J. *J. Chem. Theory Comput.* **2019**, *15*, 6984–6991.

(58) Koszykowski, M. L.; Noid, D. W.; Marcus, R. A. *J. Phys. Chem.* **1982**, *86*, 2113–2117.

(59) Esser, T. K.; Knorke, H.; Asmis, K. R.; Schöllkopf, W.; Yu, Q.; Qu, C.; Bowman, J. M.; Kaledin, M. *J. Phys. Chem. Lett.* **2018**, *9*, 798–803.

(60) Rognoni, A.; Conte, R.; Ceotto, M. *J. Chem. Phys.* **2021**, *154*, 094106.

(61) Gabas, F.; Di Liberto, G.; Conte, R.; Ceotto, M. *Chem. Sci.* **2018**, *9*, 7894–7901.

(62) Feynman, R. P.; Hibbs, A. R., *Quantum Mechanics and Path Integrals*; McGraw-Hill: 1965.

(63) Van Vleck, J. H. *Proc. Natl. Acad. Sci.* **1928**, *14*, 178–188.

(64) Gutzwiller, M. C. *J. Math. Phys.* **1967**, *8*, 1979–2000.

(65) Sun, X.; Miller, W. H. *J. Chem. Phys.* **1997**, *106*, 6346–6353.

(66) Bonnet, L. *J. Chem. Phys.* **2020**, *153*, 174102.

(67) Grossmann, F.; Xavier, A. L. *Phys. Lett. A* **1998**, *243*, 243–248.

(68) Deshpande, S. A.; Ezra, G. S. *J. Phys. Math. Gen.* **2006**, *39*, 5067–5078.

(69) Wang, H.; Manolopoulos, D. E.; Miller, W. H. *J. Chem. Phys.* **2001**, *115*, 6317–6326.

(70) Miller, W. H. *Mol. Phys.* **2002**, *100*, 397–400.

(71) Herman, M. F.; Kluk, E. *Chem. Phys.* **1984**, *91*, 27–34.

(72) Kay, K. G. *J. Chem. Phys.* **1994**, *100*, 4377–4392.

(73) Kay, K. G. *J. Chem. Phys.* **1994**, *100*, 4432–4445.

(74) Kay, K. G. *J. Chem. Phys.* **1994**, *101*, 2250–2260.

(75) Conte, R.; Gabas, F.; **Botti**, **G.**; Zhuang, Y.; Ceotto, M. *J. Chem. Phys.* **2019**, *150*, 244118.

(76) Gandolfi, M.; Ceotto, M. *J. Chem. Theory Comput.* **2021**, *17*, 6733–6746.

(77) Begušić, T.; Cordova, M.; Vaníček, J. *J. Chem. Phys.* **2019**, *150*, 154117.

(78) Di Liberto, G.; Ceotto, M. *J. Chem. Phys.* **2016**, *145*, 144107.

(79) Shao, J.; Makri, N. *J. Phys. Chem. A* **1999**, *103*, 7753–7756.

(80) Petersen, J.; Pollak, E. *J. Chem. Phys.* **2015**, *143*, 224114.

(81) Walton, A. R.; Manolopoulos, D. E. *Mol. Phys.* **1996**, *87*, 961–978.

(82) Kröninger, F.; Lasser, C.; Vaníček, J. *Front. Phys.* **2023**, *11*, 1106324.

(83) Kaledin, A. L.; Miller, W. H. *J. Chem. Phys.* **2003**, *119*, 3078–3084.

(84) Kaledin, A. L.; Miller, W. H. *J. Chem. Phys.* **2003**, *118*, 7174–7182.

(85) De Leon, N.; Heller, E. J. *J. Chem. Phys.* **1983**, *78*, 4005–4017.

(86) Ceotto, M.; Atahan, S.; Shim, S.; Tantardini, G. F.; Aspuru-Guzik, A. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3861.

(87) Gabas, F.; Conte, R.; Ceotto, M. *J. Chem. Theory Comput.* **2017**, *13*, 2378–2388.

(88) Gabas, F.; Di Liberto, G.; Ceotto, M. *J. Chem. Phys.* **2019**, *150*, 224107.

(89) Moscato, D.; Gabas, F.; Conte, R.; Ceotto, M. *J. Biomol. Struct. Dyn.* **2023**, 1–11.

(90) Landau, L. D.; Lifshitz, E. M., *Mechanics*; Butterworth–Heinemann, Oxford: 1976.

(91) Keller, J. B. *Annals of Physics* **1958**, *4*, 180–188.

(92) Brillouin, L. *J. Phys. Radium* **1926**, *7*, 353–368.

(93) Berry, M V; Tabor, M *Proc. R. Soc. Lond. A* **1976**, *349*, 101–123.

(94) Stone, A. D. *Physics Today* **2005**, *58*, 37–43.

(95) Curtis, L. J.; Ellis, D. G. *Am. J. Phys.* **2004**, *72*, 1521–1523.

(96) Solovev, E. A. *J. Exp. Theor. Phys.* **1978**, *75*, 1261–1268.

(97) Skodje, R. T.; Borondo, F.; Reinhardt, W. P. *J. Chem. Phys.* **1985**, *82*, 4611–4632.

(98) Johnson, B. R. *J. Chem. Phys.* **1985**, *83*, 1204–1217.

(99) Zakrzewski, J.; Saini, S.; Taylor, H. S. *Phys. Rev. A* **1988**, *38*, 3877–3899.

(100) Bose, A.; Makri, N. *J. Chem. Phys.* **2015**, *143*, 114114.

(101) Johnson, B. R. *J. Chem. Phys.* **1987**, *86*, 1445–1450.

(102) Aieta, C.; Gabas, F.; Ceotto, M. *J. Phys. Chem. A* **2016**, *120*, 4853–4862.

(103) Aieta, C.; Gabas, F.; Ceotto, M. *J. Chem. Theory Comput.* **2019**, *15*, 2142–2153.

(104) Aieta, C.; Micciarelli, M.; Bertaina, G.; Ceotto, M. *Nat. Commun.* **2020**, *11*, 4348.

(105) **Botti**, **G.**; Ceotto, M.; Conte, R. *J. Chem. Phys.* **2021**, *155*, 234102.

(106) Micciarelli, M.; Gabas, F.; Conte, R.; Ceotto, M. *J. Chem. Phys.* **2019**, *150*, 184113.

(107) Gabas, F.; Conte, R.; Ceotto, M. *J. Chem. Theory Comput.* **2020**, *16*, 3476–3485.

(108) Micciarelli, M.; Conte, R.; Suarez, J.; Ceotto, M. *J. Chem. Phys.* **2018**, *149*, 064115.

(109) Moghaddasi Fereidani, R.; Vaníček, J. *J. Chem. Phys.* **2023**, *159*, 094114.

(110) Brewer, M. L.; Hulme, J. S.; Manolopoulos, D. E. *J. Chem. Phys.* **1997**, *106*, 4832–4839.

(111) Nakanaga, T.; Kondo, S.; Saëki, S. *J. Chem. Phys.* **1982**, *76*, 3860–3865.

(112) Aieta, C.; Bertaina, G.; Micciarelli, M.; Ceotto, M. *J. Chem. Phys.* **2020**, *153*, 214117.

(113) Stepanian, S. G.; Reva, I. D.; Radchenko, E. D.; Rosado, M. T. S.; Duarte, M. L. T. S.; Fausto, R.; Adamowicz, L. *J. Phys. Chem. A* **1998**, *102*, 1041–1054.

(114) Bazsó, G.; Magyarfalvi, G.; Tarczay, G. *J. Phys. Chem. A* **2012**, *116*, 10539–10547.

(115) Conte, R.; Aieta, C.; **Botti**, **G.**; Cazzaniga, M.; Gandolfi, M.; Lanzi, C.; Mandelli, G.; Moscato, D.; Ceotto, M. *Theor Chem Acc* **2023**, *142*, 53.

(116) **Botti**, **G.**; Aieta, C.; Conte, R. *J. Chem. Phys.* **2022**, *156*, 164303.

(117) MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397–412.

(118) Summers, N. L.; Karplus, M. *J. Mol. Biol.* **1990**, *216*, 991–1016.

(119) Reva, I.; Stepanian, S.; Plokhotnichenko, A.; Radchenko, E.; Sheina, G.; Blagoi, Y. *J. Mol. Struct.* **1994**, *318*, 1–13.

(120) Czinki, E.; Császár, A. G. *Chem. - Eur. J.* **2003**, *9*, 1008–1019.

(121) Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648–1652.

(122) Adzhubei, A. A.; Sternberg, M. J. E. *J. Mol. Biol.* **1993**, *229*, 472–493.

(123) Schmidt, J.; Kass, S. R. *J. Phys. Chem. A* **2013**, *117*, 4863–4869.

(124) Silzel, J. W.; Murphree, T. A.; Paranji, R. K.; Guttman, M. M.; Julian, R. R. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 1974–1980.

(125) Fischer, G. *Chem. Soc. Rev.* **2000**, *29*, 119–127.

(126) Reimer, U.; Mokdad, N. E.; Schutkowski, M.; Fischer, G. *Biochemistry* **1997**, *36*, 13802–13808.

(127) Fanghanel, J.; Fischer, G. *Front. Biosci.* **2004**, *9*, 3453.

(128) Fischer, S.; Dunbrack, R. L.; Karplus, M. *J. Am. Chem. Soc.* **1994**, *116*, 11931–11937.

(129) Wiberg, K. B.; Laidig, K. E. *J. Am. Chem. Soc.* **1987**, *109*, 5935–5943.

(130) Fischer, S.; Michnick, S.; Karplus, M. *Biochemistry* **1993**, *32*, 13830–13837.

(131) Bahmanyar, S.; Houk, K. N. *J. Am. Chem. Soc.* **2001**, *123*, 12911–12912.

(132) Allemann, C.; Gordillo, R.; Clemente, F. R.; Cheong, P. H.-Y.; Houk, K. N. *Acc. Chem. Res.* **2004**, *37*, 558–569.

(133) Hoang, L.; Bahmanyar, S.; Houk, K. N.; List, B. *J. Am. Chem. Soc.* **2003**, *125*, 16–17.

(134) Nielsen, M.; Worgull, D.; Zweifel, T.; Gschwend, B.; Bertelsen, S.; Jørgensen, K. A. *Chem. Commun.* **2011**, *47*, 632–649.

(135) Bahmanyar, S.; Houk, K. N.; Martin, H. J.; List, B. *J. Am. Chem. Soc.* **2003**, *125*, 2475–2479.

(136) Lesarri, A.; Mata, S.; Cocinero, E. J.; Blanco, S.; López, J. C.; Alonso, J. L. *Angew. Chem. Int. Ed.* **2002**, *41*, 4673–4676.

(137) Valiev, M.; Bylaska, E.; Govind, N.; Kowalski, K.; Straatsma, T.; Van Dam, H.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T.; de Jong, W. *Comput. Phys. Commun.* **2010**, *181*, 1477–1489.

(138) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(139) **Botti**, **G.**; Ceotto, M.; Conte, R. *J. Phys. Chem. Lett.* **2023**, *14*, 8940–8947.

(140) Kim, S. K.; Takahashi, M.; Nordén, B. *Biochim. Biophys. Acta - Gene Structure and Expression* **1995**, *1264*, 129–133.

(141) Hoogsteen, K. *Acta Cryst* **1963**, *16*, 907–916.

(142) Watson, J. D.; Crick, F. H. C. *Cold Spring Harb. Symp. Quant. Biol.* **1953**, *18*, 123–131.

(143) Lemieux, S. *Nucleic Acids Res.* **2002**, *30*, 4250–4263.

(144) Slocombe, L.; Al-Khalili, J. S.; Sacchi, M. *Phys. Chem. Chem. Phys.* **2021**, *23*, 4141–4150.

(145) Slocombe, L.; Sacchi, M.; Al-Khalili, J. *Commun. Phys.* **2022**, *5*, 109.

(146) Umesaki, K.; Odai, K. *J. Phys. Chem. B* **2020**, *124*, 1715–1722.

(147) Gheorghiu, A.; Coveney, P. V.; Arabi, A. A. *Interface Focus* **2020**, *10*, 20190120.

(148) Hartweg, S.; Hochlaf, M.; Garcia, G. A.; Nahon, L. *J. Phys. Chem. Lett.* **2023**, *14*, 3698–3705.

(149) Groenhof, G.; Schäfer, L. V.; Boggio-Pasqua, M.; Goette, M.; Grubmüller, H.; Robb, M. A. *J. Am. Chem. Soc.* **2007**, *129*, 6812–6819.

(150) Chen, H.-Y.; Kao, C.-L.; Hsu, S. C. N. *J. Am. Chem. Soc.* **2009**, *131*, 15930–15938.

(151) Rangadurai, A.; Szymanski, E. S.; Kimsey, I.; Shi, H.; Al-Hashimi, H. M. *J Biomol NMR* **2020**, *74*, 457–471.

(152) Kabeláč, M.; Ryjáček, F.; Hobza, P. *Phys. Chem. Chem. Phys.* **2000**, *2*, 4906–4909.

(153) Seki, T.; Chiang, K.-Y.; Yu, C.-C.; Yu, X.; Okuno, M.; Hunger, J.; Nagata, Y.; Bonn, M. *J. Phys. Chem. Lett.* **2020**, *11*, 8459–8469.

(154) Fecko, C. J.; Eaves, J. D.; Loparo, J. J.; Tokmakoff, A.; Geissler, P. L. *Science* **2003**, *301*, 1698–1702.

(155) Dean, J. L. S.; Fournier, J. A. *J. Phys. Chem. B* **2022**, *11*, 4.

(156) Biswal, H. S.; Gloaguen, E.; Loquais, Y.; Tardivel, B.; Mons, M. *J. Phys. Chem. Lett.* **2012**, *3*, 755–759.

(157) Morawietz, T.; Urbina, A. S.; Wise, P. K.; Wu, X.; Lu, W.; Ben-Amotz, D.; Markland, T. E. *J. Phys. Chem. Lett.* **2019**, *10*, 6067–6073.

(158) Gabas, F.; Conte, R.; Ceotto, M. *J. Phys. Chem. Lett.* **2022**, *13*, 1350–1355.

(159) Fallacara, E.; Depondt, P.; Huppert, S.; Ceotto, M.; Finocchi, F. *J. Phys. Chem. C* **2021**, *125*, 22328–22334.

(160) Kabelac, M.; Hobza, P. *Chemistry A European J* **2001**, *7*, 2067–2074.

(161) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. *J. Chem. Phys.* **2001**, *115*, 4604–4611.

(162) Plützer, C.; Hünig, I.; Kleinermanns, K.; Nir, E.; de Vries, M. S. *ChemPhysChem* **2003**, *4*, 838–842.

(163) Plützer, C.; Hünig, I.; Kleinermanns, K. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1158–1163.

(164) Jiang, Y.; Wang, L. *J. Phys. Chem. B* **2019**, *123*, 5791–5804.

(165) Brauer, B.; Gerber, R. B.; Kabeláč, M.; Hobza, P.; Bakker, J. M.; Abo Riziq, A. G.; de Vries, M. S. *J. Phys. Chem. A* **2005**, *109*, 6974–6984.

(166) Sobolewski, A. L.; Domcke, W. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2763–2771.

(167) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

(168) Lee, T. J.; Martin, J. M. L.; Taylor, P. R. *J. Chem. Phys.* **1995**, *102*, 254–261.

(169) Carter, S.; Pinnavaia, N.; Handy, N. C. *Chem. Phys. Lett.* **1995**, *240*, 400–408.

(170) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(171) Bonnet, L.; Rayez, J. *Chem. Phys. Lett.* **1997**, *277*, 183–190.

(172) Bahmanyar, S.; Houk, K. N. *J. Am. Chem. Soc.* **2001**, *123*, 11273–11283.

(173) Darling, B. T.; Dennison, D. M. *Phys. Rev.* **1940**, *57*, 128–139.

(174) Moghimi-Araghi, S.; Loran, F. *Eur. J. Phys.* **2021**, *42*, 055802.

Acronyms

A

AS adiabatic switching. 3, 23–25, 27, 31–34, 37, 38, 41, 44, 81–83, 85–88

ASSCIVR adiabatic switching semiclassical initial value representation. 3, 25, 27, 30, 32–45, 49, 51–58, 81–83, 86–88

aVDZ augmented correlation-consistent polarized valence-only double-zeta basis set. 29, 32–34, 38–41, 47, 49, 52, 53, 55, 58, 65, 66, 74

B

bnd bending. 57–59, 63, 69, 72

BO Born-Oppenheimer. 5

C

CMD centroid molecular dynamics. 3

D

DC divide-and-conquer. 3, 42, 43, 88

DCSCIVR divide-and-conquer semiclassical initial value representation. 3, 18, 27, 38, 51, 65–67, 70, 71, 73–78, 82, 88

def2-TZVP Karlsruhe basis sets with valence triple-zeta and polarization. 65, 66, 68–71, 74–77

DMC diffusion monte carlo. 36, 38, 41

DNA deoxyribonucleic acid. 1, 3, 61, 62, 79

DVR discrete variable representation. 2, 90

E

EBK Einstein-Brillouin-Keller. 21, 22

F

FWHM full width half maximum. 36, 38, 39, 43, 73

G

GC guanine-cytosine. 3, 61–66, 68, 74, 79, 88

H

HDB hessian database. 15, 34, 51, 67

HHKK Herman-Heller-Kluk-Kay. 15, 16

I

IR infrared. 3, 9, 54, 55, 63, 74, 78, 86, 89

IVR initial value representation. 14, 15

M

MCSCIVR multiple coherent semiclassical initial value representation. 3, 17, 18, 23, 30, 34–43, 87, 88

MCTDH multiconfiguration time-dependent Hartree. 2

MD molecular dynamics. 1, 2, 81, 82, 85, 86

N

NIR near infrared. 86, 89

NLS natural Lagrangian system. 5–7

NQE nuclear quantum effects. 2, 3, 8, 13

P

PES potential energy surface. 2, 3, 17, 30, 36, 38, 44, 51, 65, 82, 85, 87

Q

QCMD quasiclassical molecular dynamics. 13

QCT quasiclassical trajectory. 13, 23, 65, 66, 68–74, 76, 78, 81, 82, 85, 86, 88

R

RMPD ring polymer molecular dynamics. 3

RNA ribonucleic acid. 61

S

SC semiclassical. 1, 27, 29, 30

SCIVR semiclassical initial value representation. 3, 32, 34, 41, 86, 89

str stretching. 45–47, 49–52, 54–56, 58, 59, 68, 74, 76, 78, 82, 87, 89

T

TA time-averaging. 11, 16

TASCIVR time-averaged semiclassical initial value representation. 16, 17, 23–25, 82, 90

U

UV ultraviolet. 3, 63, 74, 78

V

VCI vibrational configuration interaction. 2

VDOS vibrational density of states. 13, 15, 16

VPT vibrational perturbation theory. 2, 36

VSCF vibrational self-consistent field. 36

W

WC Watson and Crick. 62–65, 67, 68, 73, 78, 79, 88

Z

ZPE zero point energy. 2, 12, 13, 15, 17, 23, 34, 36–39, 41, 44, 47, 50–52, 81–83, 85, 86, 88

I thank *Instituto Nazionale di Previdenza Sociale* (INPS) for financing my scholarship.

My most sincere gratitude goes to Riccardo Conte and Michele Ceotto, that steered me from my Bachelor degree up to here: I owe you everything I know and I am proud of being one of your students.

I will never forget the support and friendship of Giacomo, and of Chiara, Marco, Michele, Davide, Cecilia, Daniele and Marco, and of all the youngsters that came and went, and of all the colleagues of the XXXVI cycle: it has been an honor to share this Ph.D. with you. I should say I would gladly repeat this experience with you, but we all know a Ph.D. is more than enough.

I thank prof. Maria Noelia Faginas Lago, prof. Andrea Lombardi, prof. Marco Paolantoni and dr. Leonardo Pacifici, and dr. Fabio Finocchi, dr. Philippe Depondt, dr. Simon Huppert, dr. Erika Fallacara and dr. Niccolò Avallone. Thanks to you, I felt at home in Perugia and Parigi as much as I do in Castell'Arquato.

To Alessandro, Giacomo & Giacomo, Marcello, Martina, Nadia, Luca and Ilaria, that were with me even if not always near me, thank you! And by the way: what are you doing for the next holidays?

I am grateful to my family, that carried my both literally and metaphorically through this endeavour: it is the safety of the harbour that makes the danger of the tempest acceptable.

To the wee little devil, whose only task was embellishing this thesis with spelling mistakes and grammatical errors, *etaoin shrdlu* to you, you little bastard. And thank you, spelling checker: I discovered you too late, but your addition was very welcomed.

Castell'Arquato,
June 12, 2024