University of Namur

MASTER IN CHEMISTRY RESEARCH FOCUS

Simulation of Raman optical activity signatures of the Cryptophane-PP-111

D'HAESE, Lou

Award date:
2022

Awarding institution:
University of Namur

Link to publication

Université de Namur
Faculté des Sciences

Simulation of Raman optical activity signatures of the Cryptophane-PP-111

Mémoire présenté pour l'obtention

du grade académique de Master Chimie «Chimie du Vivant» : Finalité Approfondie

Lou D’HAESE

Janvier 2022

Simulation of Raman optical activity signatures of the Cryptophane–PP–111

Master Thesis presented by:

Lou D'haese[1]

January 2022

Promoter: Vincent Liégeois[1]
Collaborator: Nicolas Daugey[2]

[1]Namur Institute of Structured Matter (NISM), Laboratoire de Chimie Théorique (LCT),
University of Namur, rue de Bruxelles, 61, B-5000 Namur, Belgium

[2]Institut des Sciences Moléculaires (ISM), Groupe Spectroscopie Moléculaire (GSM),
University of Bordeaux, cours de la Libération, 351, UMR-5255 Bordeaux, France

UNIVERSITE DE NAMUR
Faculté des Sciences
Secrétariat du Département de Chimie
Rue de Bruxelles 61 – 5000 NAMUR
Téléphone : +32(0)81 72.54.44 – Téléfax : +32(0)81 72.54.40
E-mail : enseignement.chimie@unamur.be - www.unamur.be/sciences

Simulation of Raman optical activity signatures of the Cryptophane–PP–111

Lou D'haese, Vincent Liégeois and Nicolas Daugey

Abstract

Cryptophane molecules consist in two hemispheres bound with three $-\mathrm{O(CH_2)}_n\mathrm{O}-$ linkers to form a quite rigid cage. The arrangement of these linkers makes them chiral, *anti* or *syn* for **PP** and **MM** enantiomers, which generates chiroptical properties. The cryptophane cavity allows to encapsulate some atoms or little molecules which impact their chiroptical properties. Cryptophane–111 possesses the smallest cavity and exhibits the highest binding constant for xenon encapsulation, (around $10\,000\,\mathrm{M}^{-1}$ at $293\,\mathrm{K}$). Xenon is hugely polarizable and is heavily used in medical applications, e.g. in MRI (magnetic resonance imaging), because xenon has the ability to enhance the signals. Thus, Cryptophane–111 might be interesting as a biological sensor.

Among the different vibrational techniques, we decide to focus on the Raman Optical Activity (ROA) spectroscopy, the optical activity counterpart of Raman spectroscopy, which is very sensitive to the conformations and surrounding effects.

The principal objective of this work is to properly simulate ROA spectra of our targeted system, Cryptophane–**PP**–111, in an organic solvent and with the xenon encapsulated in its cavity. In the first part of this work, the appropriate method has been determined on the **R**–methyloxirane system: **B3LYP** and ω**B97X–D** for the geometry optimization and the normal modes calculation; **TDHF/rDPS:3–21G** for the property derivatives calculation. In the second part, we have sampled the potential energy surface of the Cryptophane–PP–111 using two conformational search programs (GMMX and CREST) and only one conformer (**1**) has a substantial Boltzmann weight both in gas phase and in solution. Finally, our best final ROA spectra (**B3LYP/6–31G*//TDHF/rDPS:3–21G//SMD/**$\mathbf{CH_2Cl_2}$) with/without xenon have been compared with the experimental data [25]. Most of the ROA signatures were quite well reproduced except for the low frequency region certainly due to anharmonic effects.

In this work, we have constructed a simulation protocol that is able to reproduce most of the experimental ROA signatures. Some improvements are still possible especially at the low frequency region.

Thesis for the Master in Chemical Sciences focus on Research
January 2022
Promoter: Vincent Liégeois

UNIVERSITE DE NAMUR
Faculté des Sciences
Secrétariat du Département de Chimie
Rue de Bruxelles 61 – 5000 NAMUR
Téléphone : +32(0)81 72.54.44 – Téléfax : +32(0)81 72.54.40
E-mail : enseignement.chimie@unamur.be - www.unamur.be/sciences

Simulation des signatures vibrationnelles d'activité optique Raman du Cryptophane–PP–111

Lou D'haese, Vincent Liégeois et Nicolas Daugey

Résumé

Les cryptophanes sont des molécules cages qui peuvent encapsuler des atomes ou des petites molécules. Ils sont constitués de deux hémisphères, reliés entre–eux par trois liens $-\mathrm{O(CH_2)}_n\mathrm{O}-$ qui engendrent la chiralité du système en fonction de leur arrangement (créant des énantiomères *anti* et *syn* (**PP** et **MM**)). Le Cryptophane–111 possède la plus petite cavité et présente la constante de complexation la plus élevée pour l'encapsulation du xénon (environ $10\,000\,\mathrm{M}^{-1}$ à $293\,\mathrm{K}$). Le xénon est très polarisable et est largement utilisé dans le milieu médical, comme l'imagerie par résonance magnétique, pour sa capacité à augmenter les réponses moléculaires. Ainsi, le Cryptophane–111 pourrait être utilisé comme bio–senseur.

Parmi les différentes techniques vibrationnelles, nous nous sommes concentrés sur la spectroscopie d'activité optique Raman (AOR), le pendant optiquement actif de la spectroscopie Raman, qui est très sensible aux conformations et aux effets d'environnement.

Le principal objectif de ce travail est de simuler correctement les spectres AOR de notre système cible, le Cryptophane–**PP**–111, dans un solvant organique et avec le xénon encapsulé dans sa cavité.

Dans la première partie de ce travail, la méthode appropriée pour simuler les spectres AOR a été déterminée sur la molécule de **R**–méthyloxirane: les fonctionnelles d'échange–corrélation **B3LYP** et ω**B97X–D** pour l'optimisation de géométrie et le calcul des modes normaux; **TDHF/rDPS:3–21G** pour le calcul des dérivées des propriétés. Dans la seconde partie, la surface d'énergie potentielle du Cryptophane–**PP**–111 a été sondée avec deux programmes de recherche conformationnelle (GMMX et CREST) et un unique conformère (**1**), ayant un poids de Boltzmann important, a été trouvé. Finallement, nos meilleurs spectres AOR (**B3LYP/6–31G*//TDHF/rDPS:3–21G//SMD/CH**$_2$**Cl**$_2$), avec/sans xénon, ont été comparés aux spectres expérimentaux [25]. La plupart des signatures vibrationnelles ont été reproduites avec succès excepté celles de la zone de basse fréquence où des effets anharmoniques seraient présents.

Dans ce travail, nous avons construit un protocole de simulation qui est capable de reproduire les signatures vibrationnelles AOR expérimentales. Cependant, quelques améliorations restent possibles, surtout dans la zone de basse fréquence.

Mémoire de Master en Sciences Chimiques à Finalité Approfondie
Janvier 2022
Promoteur: Vincent Liégeois

Acronyms

ASC = Apparent Surface Charge

CPHF = Coupled–Perturbed Hartree–Fock

CPKS = Coupled–Perturbed Kohn–Sham

CRE = Conformer–Rotamer Ensemble

CREST = Conformer–Rotamer Ensemble Sampling Tool

CV = Collective Variable

DFT = Density Functional Theory

DFTB = Density Functional Tight–Binding

DOSD = Dipole Oscillator Strength Distributions

ECP = Effective Core Potential

EXX = Exact Hartree–Fock Exchange

GC = Genetic Structure Crossing

GGA = Generalized Gradient Approximation

GMMX = Global–MMX

HF = Hartree–Fock

HK = Hohenberg–Kohn

HWHM = Half–Width at Half–Maximum

IEF = Integral Equation Formalism

IR = Infrared

KS = Kohn–Sham

L.C.A.O. = Linear Combination of Atomic Orbitals

LDA = Local Density Approximation

MC = Monte Carlo

MD = Molecular Dynamics

mGGA = Meta–Generalized Gradient Approximation

MM = Molecular Mechanics

MMFF94 = Merck Molecular Force Field 94

MP = Model Potential

MPPT = Møller–Plesset Perturbation Theory

MTD = Meta–Dynamics

PCM = Polarizable Continuum Model

PES = Potential Energy Surface

RF = Reaction Field

RMSD = Root–Mean–Square Deviation

ROA = Raman Optical Activity

RSPT = Rayleigh–Schrödinger Perturbation Theory

SAS = Solvent Accessible Surface

SCF = Self–Consistent Field

SCRF = Self–Consistent Reaction Field

SES = Solvent Excluding Surface

SMD = Solvation Model Density

TDDFT = Time–Dependent Density Functional Theory

TDHF = Time–Dependent Hartree–Fock

UFF = Universal Force Field

VCD = Vibrational Circular Dichroism

vdW = van der Waals

VO = Valence–Only

XC = Exchange–Correlation

xTB = Extended Tight–Binding

ZDO = Zero–Differential Overlap

Contents

Chapter 1. Introduction

Cryptophane molecules consist in two hemispheres each made of three benzene rings connected together in ortho by three methylene groups. The two hemispheres are themselves bound with three $-\mathrm{O(CH_2)}_n\mathrm{O}-$ linkers to form a quite rigid cage. The arrangement of these linkers makes them chiral, *anti* or *syn* for **PP** and **MM** enantiomers (Fig. 1.1), which generates chiroptical properties.

Cryptophanes have been discovered in 1985 [1] and synthesized for the first time by André Collet [2] in 1987. From that moment, a plethora of derivatives has been synthesized [3–22]. The cryptophanes have been studied for their cavity which allows to encapsulate some atoms or little molecules. Indeed, their chiroptical properties are impacted by the encapsulation of guest molecules.

The smallest one, Cryptophane–111, was synthesized and characterized by Fogarty *et al.* [23] only in 2007. In fact, its cavity is the smallest one [24] and cannot encapsulate little molecules. However, it exhibits the highest binding constant for xenon encapsulation, (around $10\,000\,\mathrm{M}^{-1}$ at $293\,\mathrm{K}$). Xenon is hugely polarizable and is heavily used in medical applications, e.g. in MRI (magnetic resonance imaging), because xenon has the ability to enhance the signals. Thus, Cryptophane–111 might be interesting as a biological sensor.

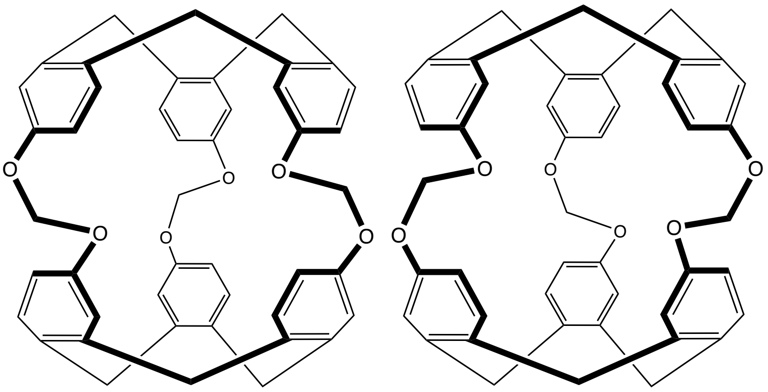


Figure 1.1: Sketches of Cryptophane–**PP**–111 (left) and Cryptophane–**MM**–111 (right).

Vibrational spectroscopies are powerful tools to unravel the structure of molecules, polymers, supramolecules or surfaces by probing their responses with respect to an external electromagnetic field. In addition, their signatures, one per vibrational degree of freedom, are sensitive towards the molecular environment such as: solvent or guest encapsulation.

Among the different techniques, the Raman Optical Activity (ROA) spectroscopy, the optical activity counterpart of Raman spectroscopy, and the Vibrational Circular Dichroism (VCD) can distinguish two enantiomers which both generate opposite spectra (as mirrors). As Infrared (IR) and Raman spectroscopies, ROA and VCD are complementary tools to study the structure of targeted systems. Here, we focused on ROA spectroscopy which is very sensitive to the conformations and surrounding effects. Moreover, experimentally, the low frequency region can be recorded while it is usually not the case for IR and VCD.

The principal objective of this work is to properly simulate ROA spectra of our targeted system, Cryptophane–**PP**–111, in an organic solvent and with the xenon encapsulated in its cavity. First, an appropriate method will be selected. This step is crucial since the vibrations have to be simulated properly, same for their respective intensity. This search will be performed on a reference chiral system, **R**–methyloxirane. This system has been heavily studied in the past thus experimental data are available and will be compared with our simulated spectra. Second, a conformational search will be performed. The flexibility of the linkers in the Cryptophane–**PP**–111 is a challenge. Indeed, we need to take into account all the stable conformations of our system. Their spectra will be mixed together, based on their respective Boltzmann weight, to create the final spectrum. The generation of the conformers is important thus two programs will be used: GMMX and CREST. The obtained conformers will be reoptimized using the appropriate method selected from the **R**–methyloxirane investigation. Third, the environment effects will be added to our gas phase system (solvent and xenon). Both are also important to simulate realistic spectra. The solvent effects will be added with both IEF–PCM and SMD approaches. Xenon possesses a huge number of electrons and their relativistic effects are considerable. Effective core potentials will be used to take into consideration these relativistic effects in our nonrelativistic calculations. Finally, the final ROA spectra will be compared with experimental data to evaluate the efficiency of our computational methods. The experimental spectra have been recorded by our collaborators; Nicolas Daugey is part of this group; at the University of Bordeaux and published in the article of Thierry Buffeteau *et al.* [25]. In this work, we will build a simulation protocol to properly simulate the ROA spectrum of a flexible system. If this objective is completed, larger cryptophane derivatives will be considered: Cryptophane–111(OMe)$_6$, Cryptophane–222, Cryptophane–223, Cryptophane–333, . . .

In Chapter 2, the different quantum chemistry methods to evaluate the energy and the selected properties of a system in gas phase and with environment effects are described. The classical mechanics method and tight–binding approaches used by the conformational search programs are also explained in this chapter. In Chapter 3, the description of the vibrational normal modes as well as the Raman and ROA intensities in the double–harmonic approximation is resumed. In Chapter 4, the global strategy to simulate ROA spectrum and the tools to analyze the signatures are presented. In Chapter 5, the results of this work are discussed. First, the appropriate method is determined with our chiral reference system, **R**–methyloxirane. Then, the conformational search is performed on the Cryptophane–**PP**–111 in gas phase with two programs (GMMX and CREST) to find the most stable conformations (reoptimized using the appropriate method). Subsequently, different solvents are added to our system and computed with two approaches (IEF–PCM and SMD). Afterwards, the xenon encapsulation is added to our system, preliminary in gas phase. Finally, the comparison of our final spectrum with experimental spectra is carried out. In Chapter 6, the conclusions of this work are presented as well as the outlook.

Chapter 2. Theoretical Methods

2.1 Quantum Chemistry

The wavefunction, Ψ, contains all the information of our system. Therefore, the wavefunction is essential in quantum chemistry. When an operator is applied on the wavefunction, it mathematically looks like an eigenvalue problem in which the wavefunction is an eigenvector. Here, the eigenvalues are the properties of our system and the operator is the time–independent Schrödinger [26–30] Hamiltonian, $\hat{H}$:

$$\hat{H}\Psi = E\Psi \,. \tag{2.1}$$

The Hamiltonian contains two operators: the kinetic energy operator $\hat{T}$, made up of the electronic and nuclear terms, and the potential energy operator $\hat{V}$, made up of the electronic, nuclear, and crossing terms.

$$\hat{H} = \hat{T} + \hat{V} = \hat{T}_e + \hat{T}_N + \hat{V}_{ee} + \hat{V}_{Ne} + \hat{V}_{NN} \,. \tag{2.2}$$

In the Born–Oppenheimer approximation [31], the $\hat{T}_N$ term is neglected with respect to the $\hat{T}_e$ term. Since the nuclei are much heavier than the electrons so their momentum operators are hugely smaller than the ones of the electrons. Then, the nuclei and electrons can be decoupled: for the electronic problem, the nuclei are considered immobile and the nuclei–nuclei repulsion is a constant term, $\hat{V}_{NN}$, which is added to the electronic energy.

2.1.1 Hartree–Fock

The Hartree–Fock (HF) [32–35] method consist in representing the N–electron wavefunction by a **unique** Slater determinant [36] built from one–electron wavefunctions, the spinorbitals $\Theta_i(\vec{x}) = \Phi_i(\vec{r})\sigma_i(\omega)$, products of spatial orbitals $\Phi_i(\vec{r})$ and spin functions $\sigma_i(\omega)$. The Slater determinant fulfills the antisymmetric principle of fermions (species with a 1/2 spin) so two electrons are not allowed to occupy the same spinorbital.

$$\Psi^{HF}(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \Theta_1(\vec{x}_1) & \Theta_2(\vec{x}_2) & \cdots & \Theta_N(\vec{x}_1) \\ \Theta_1(\vec{x}_2) & \Theta_2(\vec{x}_2) & \cdots & \Theta_N(\vec{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_1(\vec{x}_N) & \Theta_2(\vec{x}_N) & \cdots & \Theta_N(\vec{x}_N) \end{vmatrix} \,, \tag{2.3}$$

The spinorbitals, constituting the Slater determinant, are determined by invoking the variation principle to minimize the electronic energy system. If E_0 represents the exact energy of the ground state with the wavefunction Ψ_0, the approximated energy E_Φ of any wavefunction Φ can only be superior or equal to E_0:

$$E_\Phi - E_0 \geq 0 \,, \tag{2.4}$$

and the Hartree–Fock equation, with the Fock operator $f(\vec{r}_1)$, reads:

$$f(\vec{r}_1)\Theta_n(\vec{x}_1) = \left\{ h(\vec{r}_1) + \sum_{j=1}^{N} J_j(\vec{r}_1) - K_j(\vec{r}_1) \right\} \Theta_n(\vec{x}_1) = \epsilon_n \Theta_n(\vec{x}_1)\,, \tag{2.5}$$

where $J_j(\vec{r}_1)$ and $K_j(\vec{r}_1)$ are the Coulomb and exchange operators describing the averaged interactions between the electrons in Θ_n and Θ_j whereas $h(\vec{r}_1)$ groups the one–electron kinetic energy and electron–nuclei attraction operators. The j–sum runs over the N occupied spinorbitals. The Coulomb and exchange operators depend on the spinorbitals so the Hartree–Fock equation has to be solved iteratively, until convergence, a procedure known as the Self–Consistent Field (SCF).

2.1.1.1 L.C.A.O. Approximation

To solve the Hartree–Fock equation, a numerical approach has been tested where the space is decomposed into points network. Some difficulties appeared in the case of molecules. Then, to beat the problem, each molecular orbitals ϕ_n are represented by a linear combination of atomic orbitals (L.C.A.O.):

$$\phi_n(\vec{r}_1) = \sum_{q=1}^{K} C_{qn} \chi_q\,(\vec{r}_1)\,, \tag{2.6}$$

where K is the number of atomic orbitals, C_{qn} are the L.C.A.O. coefficients, and χ_q are the predefined atomic orbitals. By introducing this definition of molecular orbitals in the Hartree–Fock equation for a full occupied system, the eigenvalue problem can be written as (Roothaan–Hall equations [37, 38]):

$$\mathbf{FC} = \mathbf{SCE}\,, \tag{2.7}$$

where $\mathbf{F}$ is the Fock matrix, $\mathbf{C}$ is the matrix which contains the L.C.A.O. coefficients, $\mathbf{S}$ is the overlap matrix, and $\mathbf{E}$ is the diagonal matrix which contains the orbital energies. For Fock matrix, the elements read:

$$F_{pq} = H_{pq} + \sum_{r=1}^{K} \sum_{s=1}^{K} D_{sr} \big[2(pq|rs) - (ps|rq) \big]\,, \tag{2.8}$$

where $(pq|rs)$ and $(ps|rq)$ are bielectronic integrals between spatial orbitals. D_{sr} correspond to the density matrix elements. As soon as the basis set is chosen, the Hartree–Fock equation is solved iteratively, as explained earlier.

Moreover, the total Hartree–Fock energy reads:

$$\begin{aligned} E^{\mathrm{HF}} &= \sum_{A>B}^{M} \frac{Z_A Z_B}{R_{AB}} + \langle \Psi^{\mathrm{HF}} |\, \hat{H}\, | \Psi^{\mathrm{HF}} \rangle \\ &= E_{NN} + 2\sum_{i=1}^{N/2} h_{ii} + 2\sum_{i,j=1}^{N/2} \big(2J_{ij} + K_{ij}\big) \\ &= E_{NN} + 2\sum_{p,q}^{K} H_{pq} D_{qp} + \sum_{p,q}^{K} D_{qp} \sum_{r,s}^{K} D_{sr} \big[2(pq|rs) - (ps|rq) \big] \\ &= E_{NN} + \sum_{p,q}^{K} D_{qp} \big(H_{pq} + F_{pq} \big)\,. \end{aligned} \tag{2.9}$$

2.1.1.2 Basis sets

The construction of a basis set comes from the hydrogen–like solutions, products of a radial function $R_{n\ell}(\vec{r})$ by an angular function $Y_\ell^m(\theta,\phi)$. These functions are too complex for computational implementation and some approximations are needed: to remove the ℓ–dependence in radial part and to replace exponential functions by gaussian ones.

However, these approximations are really drastic. At $r \to 0$, the first derivative of gaussian functions is equal to 0, contrary to exponential functions. Moreover, the gaussian functions decrease more rapidly than the exponential ones. In order to tackle these problems, linear combinations of N gaussian functions have been proposed, named "contractions":

$$\chi_p^A(\vec{r}-\vec{R}_A) = \sum_{\mu=1}^{N} C_{\mu p}\,(x-x_A)^i(y-y_A)^j(z-z_A)^k N_{\mu p}\, e^{-\alpha_{\mu p}|\vec{r}-\vec{R}_A|^2}\,, \tag{2.10}$$

where the atomic orbital is centered around the atom A. $C_{\mu p}$ are the contracted coefficients of the gaussian functions, $\alpha_{\mu p}$ are the exponents of the gaussian functions, and $N_{\mu p}$ are the normalization factors. The sum $i+j+k=\ell$ defines the orbital form.

Different kind of basis sets have been proposed:

- **$\boldsymbol{n}$–ζ basis sets** [39]: each atomic orbital is described by $\boldsymbol{n}$ linear combinations of gaussian functions; **STO–3G** means atomic orbitals are described by **one** linear combination of **three** gaussian functions;

- **Valence $\boldsymbol{n}$–ζ basis sets** [39]: each valence atomic orbital is described by $\boldsymbol{n}$ linear combinations while the core orbitals are described by **one** linear combination; **6–31G** means valence atomic orbitals are described by **two** linear combinations; first with **three** gaussian functions and second with only **one** gaussian function; while the core orbitals are described by **one** linear combination of **six** gaussian functions;

- **Polarization functions** [39]: gaussian functions with higher ℓ value are added to better describe environment effects; **6–31G*** means d functions on carbon atoms are added to **6–31G**;

- **Diffuse functions** [39]: gaussian functions with low exponents are added to better describe the external part of orbitals and the electronic density; **6–31++G** means s and p diffuse functions (first +) on carbon atoms and s diffuse functions (second +) on hydrogen atoms are added to **6–31G**;

- (**aug**–)**cc–pV*n*Z basis sets** [40]: correlation–consistent basis sets constructed to beat the convergence problem of the previous basis sets, with/without diffuse functions (**aug**).

2.1.1.3 Effective Core Potentials

Effective Core Potentials (ECPs) are introduced to calculate the core electron contributions and remove them from the system. In fact, time calculation depends on the number of electrons in the system. Moreover, there are relativistic effect for core electrons that are not taken into account in nonrelativistic calculations. Thus, an effective valence–only (VO) Hamiltonian is used for n_v valence electrons treated explicitly in the calculation and N core electrons [41]:

$$\widehat{H}_v = \sum_{i}^{n_v} \widehat{h}_v(i) + \sum_{i<j}^{n_v} \widehat{g}_v(i,j) + \widehat{V}_{cpp} + \hat{V}_{cc}\,, \tag{2.11}$$

where v and c refer to valence and core electrons. $\widehat{h}_v(i)$ and $\widehat{g}_v(i,j)$ are the effective one–electron and two–electron operators, respectively. $\widehat{V}_{cpp}$ represents the core–polarization potential and $\hat{V}_{cc}$ represents the repulsion between all the core electrons and nuclei of the system.

Two types of ECPs can be distinguished: the model potentials (MPs), which try to model as accurately as possible the all electrons HF potential for the valence electrons and thus produce valence orbitals with a correct nodal structure, and pseudopotentials (PPs), which after a formal transformation from the valence orbitals to pseudovalence orbitals with a simplified radial nodal structure lead to further savings in the one–electron basis sets.

Nevertheless, this equation (Eq. (2.11)) is too heavy for computational implementation and some drastic approximations are needed. The core electrons system is separated from the valence electrons system and the atomic cores are inert and remain unchanged when they are transferred from one system to another ("frozen–core" approximation). The cores, and possibly the core orbitals contributions, in an all electrons FC Hamiltonian for the valence electrons, are replaced by an ECP modeling the real non–local HF potential to omit them from the calculation. The atomic cores do not interact with each other except for their mutual Coulomb repulsion.

There are different ECPs: **dhf** [42] and **def2** [42] which take 26 valence electrons into account; **CRENBL** [43] which takes 18 valence electrons into consideration; and the others, **CRENBS** [43], **LANL2DZ** [44], **SBKJC** [45], and **Stuttgart–RLC** [46] which take only 8 valence electrons into account in the calculations.

2.1.2 Møller–Plesset Perturbation Theory

Møller–Plesset perturbation theory (MPPT) [47] employs the Rayleigh–Schrödinger perturbation theory (RSPT) [28, 48] to the problem of electronic correlation. The wavefunction and the energy are expressed as a Taylor expansion where higher orders correspond to corrections to the zeroth order non–perturbed term. This method determines the wavefunctions and energies of a perturbed system provided that the wavefunctions and energies of the unperturbed system are known for. In MPPT, the unperturbed system corresponds to the Hartree–Fock solution. The unperturbed Hamiltonian reads as the sum of the one–electron Fock operators:

$$\hat{H}_0 = \sum_{i} f(i)\,. \tag{2.12}$$

The perturbed Hamiltonian is the full nonrelativistic Hamiltonian which contains the kinetic energy terms and the electron–electron repulsion terms. The perturbation $\lambda\hat{W}$, with the perturbation operator $\hat{W}$, is defined as the difference between the perturbed and unperturbed Hamiltonians:

$$\lambda\hat{W} = \hat{H} - \hat{H}_0 = \sum_{i<j} \frac{1}{r_{ij}} - \sum_i \nu^{\mathrm{HF}}(i)\,, \tag{2.13}$$

where $\nu^{\mathrm{HF}}(i)$ is the Fock potential and contains the Coulomb and exchange terms.

For example, MP2 corresponds to the second order correction [49–53]:

$$\begin{aligned} E_0^{\mathrm{MP2}} &= E_0^{(0)} + E_0^{(1)} + E_0^{(2)} \\ &= E_0^{\mathrm{HF}} + E_0^{(2)} \\ &= E_0^{\mathrm{HF}} + \sum_{i\neq 0} \frac{\left|\left\langle \Psi_i^{(0)} \middle| \hat{W} \middle| \Psi_0^{(0)} \right\rangle\right|^2}{E_0^{(0)} - E_i^{(0)}}\,, \end{aligned} \tag{2.14}$$

where $\Psi_i^{(0)}$ and $E_i^{(0)}$ are the molecular orbitals and their respective energies at the Hartree–Fock level.

2.1.3 Density Functional Theory

The complete N–electron wavefunction and the associated Schrödinger equation are replaced by the electron density and its associated calculation schemes in the density functional theory (DFT) [54–57]. This simplifies the quantum description of a system since the electron density depends only on the position of one electron.

"The external potential $V_{ext}(\vec{r}) = V(\vec{r})$ is determined, within a trivial additive constant, by the electron density $\rho(\vec{r})$." is the first Hohenberg–Kohn theorem [54] thus the energy is expressed under the form of a functional of the electron density:

$$E_\nu\big[\rho(\vec{r})\big] = T\big[\rho(\vec{r})\big] + V_{ne}\big[\rho(\vec{r})\big] + V_{ee}\big[\rho(\vec{r})\big]\,. \tag{2.15}$$

"For a trial density $\rho'(\vec{r})$ such as $\rho'(\vec{r}) \geq 0$ at any $\vec{r}$, the electron density obeys a variation principle:

$$E_\nu\big[\rho'(\vec{r})\big] \geq E_0\,, \tag{2.16}$$

where E_0 is the true ground state energy and $E_\nu\big[\rho'(r)\big]$ is the energy functional." is the second Hohenberg-Kohn theorem [54] and is very similar to the variation principle in the Hartree–Fock method.

The energy can be decomposed into two terms, an external potential–dependent term and a universal function, the Hohenberg and Kohn functional F_{HK} which contains the major part of the effects of exchange and correlation. That leads to another expression for the energy:

$$E_\nu\big[\rho(\vec{r})\big] = \int \rho(\vec{r})\nu(\vec{r})\mathrm{d}\vec{r} + F_{\mathrm{HK}}\big[\rho(\vec{r})\big]\,, \tag{2.17}$$

where $F_{\mathrm{HK}}\big[\rho(\vec{r})\big] = V_{ee}\big[\rho(\vec{r})\big] + G_{\mathrm{HK}}\big[\rho(\vec{r})\big]$ with $G_{\mathrm{HK}}\big[\rho(\vec{r})\big]$, another universal function, which contains the kinetic energy.

However, the electron density cannot determinate accurately the kinetic energy of the system. Then, Kohn and Sham [55] approximate $G_{\mathrm{HK}}\big[\rho(\vec{r}\,)\big]$ as:

$$G_{\mathrm{HK}}\big[\rho(\vec{r}\,)\big] = T_s\big[\rho(\vec{r}\,)\big] + E_{\mathrm{XC}}\big[\rho(\vec{r}\,)\big]\,, \tag{2.18}$$

where $T_s\big[\rho(\vec{r}\,)\big]$ is the kinetic energy functional of a system of non–interacting electrons and $E_{\mathrm{XC}}\big[\rho(\vec{r}\,)\big]$ is the exchange–correlation energy functional. Moreover, the exchange–correlation potential V_{XC} reads:

$$V_{\mathrm{XC}} = \frac{\partial E_{\mathrm{XC}}}{\partial \rho(\vec{r}\,)}\,, \tag{2.19}$$

and the eigenvalue problem can be expressed as:

$$\Big[-\frac{1}{2}\nabla^2 + V_{\mathrm{XC}}\Big]\Theta_n(\vec{x}\,) = \epsilon_n \Theta_n(\vec{x}\,)\,. \tag{2.20}$$

In order to solve the equation, the usage of Kohn–Sham spinorbitals $\Theta_{\mathrm{KS},i}(\vec{x}\,)$ is preferred and the procedure becomes iterative. However, the form of the exchange–correlation (XC) functional has not been defined and a myriad of exchange–correlation functionals appeared.

2.1.3.1 Exchange–Correlation Functionals

Starting from the Hartree world, the first rung corresponds to **local density approximation** (LDA) functionals. The XC energy of a non–uniform system can be obtained by applying uniform–electron–gas results for infinitesimal portions of the non–uniform electron distribution. The KS–LDA method describes reasonably well the molecular bonding and is clearly more reliable than HF method for atomization energies.

In **generalized gradient approximation** (GGA), the functionals depend on both the density and its gradient.

In **meta–GGA** (mGGA) functionals, an explicit dependence on the kinetic energy density is introduced.

At the beginning of the 90s, Becke proposed to introduce a percentage of exact Hartree–Fock exchange (EXX) in the functionals. The exchanges in LDA and GGA are local, so the potential at $\vec{r}$ depends only on the density with/without its gradient at $\vec{r}$, contrary to HF exchange which is non–local and presents the exact $-1/\vec{r}$ asymptotic behavior. These functionals are called "**hybrid functionals**" which incorporate explicitly a dependence of the XC potential on the occupied molecular orbitals. In fact, these functionals have been renamed "**global hybrids**" because the HF exchange is constant for any inter–electronic distance.

The list of global hybrid functionals, often combined with GGA or mGGA, is huge and only the ones under study are presented: B3LYP [58–61] with 20 % HF exchange, M06 [62] with 27 % HF exchange, and M06-2X [62] with 54 % HF exchange.

For large inter–electronic distance, long–range HF exchange is more important than short–range one so it has been proposed to vary the percentage of HF exchange as a function of the inter–electronic distance r_{12} to improve the electronic structure and density:

$$\frac{1}{r_{12}} = \frac{1-\big[\alpha+\beta\,\mathrm{erf}(\mu r_{12})\big]}{r_{12}} + \frac{\big[\alpha+\beta\,\mathrm{erf}(\mu r_{12})\big]}{r_{12}}\,. \tag{2.21}$$

At $r_{12} \to 0$, the amount of HF exchange attains α. At $r_{12} \to \infty$, it attains $\alpha + \beta$. Here, μ is the range–separating parameter which describes the speed of the transitions between short– and long–range. These functionals have been named "**range–separated hybrids**" and their efficiency depends on the α, β, and μ parameters. In ωB97X [63], $\alpha = 0.00$, $\beta = 1.00$ and $\mu = 0.33$.

However, the XC functionals do not evaluate properly the London dispersion forces which are required for describing van der Waals (vdW) interactions and the balance between electrostatic, exchange–repulsion, and dispersion effects. Grimme proposed to replace part of the non–local, long– and medium–range electron correlation effects by empirical dispersion corrections. These "**dispersion–including functionals**" describe better intermolecular and intramolecular interactions. In ωB97X-D [64], $\alpha = 0.16$, $\beta = 0.84$ and the D3 dispersion model is used.

2.1.3.2 Dispersion Model

Grimme has proposed several dispersion models [65–69] to properly evaluate the London dispersion forces. In fact, an empirical dispersion correction energy term is added to the energy of the system:

$$E_{\text{total}} = E_{\text{method}} + E_{\text{disp}} \, , \tag{2.22}$$

where E_{disp} is the dispersion correction energy, so the sum of dispersion energies between two atoms i and j at large distance R_{ij}, using the definition of Eisenschitz and London [70]:

$$E_{\text{disp}} = - \sum_{n=6,8,10...} s_n \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{C_n^{ij}}{R_{ij}^n} f_{\text{dmp}} \left(R_{ij} \right) . \tag{2.23}$$

Here, N is the number of atoms in the system, C_6^{ij} is the dispersion coefficient for atom pairs ij, s_6 is the global scaling factor that depends on the XC functional used, and R_{ij} is the interatomic distance. For small R_{ij} values, there are some singularities which appear so a damping function f_{dmp} is used to avoid these ones.

The first model [65], "D1", used the results of Wu *et. al* [71] for some atoms: H, C, N, O, F, and Ne. These results have been calculated by least–square fitting, assuming the additivity of C_6 coefficients [72] and by using the dipole oscillator strength distributions (DOSD) determined by Meath and coworkers [73–79].

The second model [66], "D2", has been proposed to describe heavier atoms.

The third model [67], "D3", used a different expression of the dispersion correction energy. This is the sum of two– and three–body energies where R^{-8} terms are taken into account for two–body energies. This model used a new damping function, proposed by Chai and Head–Gordon [64].

The D3 model has been improved [68], "D3BJ", by using a BJ–damping [80–82] which leads to a constant contribution of E_{disp} when $R_{ij} \to 0$.

Finally, the last fourth model [69], "D4", has been created for density functional theory approximations (as xTB) with a new charge–dependent scaling factor s_n. Indeed, the charge of the atom impacts its van der Waals radius but this effect was not taken into consideration in the previous versions.

2.1.4 Density Functional Tight–Binding

The density functional tight–binding (DFTB) methods are semiempirical methods which combine the efficiency of the old zero–differential overlap (ZDO) type minimal basis set methods with the accuracy of DFT. In ZDO methods, some two–electron repulsion integrals are ignored to decrease the calculation time (for N orbitals, there are N^4 two–electron repulsion integrals; after the approximation, there are only N^2 two–electron repulsion integrals). Unfortunately, the parameterization is really important in these methods and thousands of empirical values have to be determined in order to properly describe the system. It causes some problems to cover all the elements of the periodic table. To tackle these problems, the extended tight–binding (xTB) methods appeared. These methods were focused on the simulation of geometries, frequencies and non–covalent interactions, named "GFN–xTB" methods.

The total energy is reformulated in terms of a reference density ρ_0, ideally close to the final converged density ρ, and a density difference $\Delta\rho$, with $\rho = \rho_0 + \Delta\rho$. The total energy can be expressed as a Taylor expansion around $\Delta\rho = 0$:

$$E[\rho_0] = E^{(0)}[\rho_0] + E^{(1)}[\rho_0, \delta\rho] + E^{(2)}[\rho_0, (\delta\rho)^2] + E^{(3)}[\rho_0, (\delta\rho)^3] + \cdots \tag{2.24}$$

and is commonly truncated at the third–order term, leading to the acronym "DFTB3".

In the GFN2–xTB [83] method, the n^{th}–order terms are:

$$E^{(0)} = E_{\text{rep}} \tag{2.25}$$

$$E^{(1)} = E_{\text{EHT}} \tag{2.26}$$

$$E^{(2)} = E_{\gamma} + E_{\text{AES}} + E_{\text{AXC}} + E_{\text{disp}}^{\text{D4}'} \tag{2.27}$$

$$E^{(3)} = E_{\Gamma} \tag{2.28}$$

where E_{rep} corresponds to the repulsion energy, E_{EHT} represents the extended Hückel energy, E_{γ} is the isotropic second–order electrostatic energy, E_{AES} and E_{AXC} are the anisotropic electrostatic and the second–order anisotropic XC energies, respectively, and E_{Γ} is the third–order energy.

$E_{\text{disp}}^{\text{D4}'}$ corresponds to the dispersion energy, by using a modified D4 dispersion model where the atomic charges are taken from a Mulliken population in GFN2–xTB and are solved self–consistently.

2.2 Properties Calculation

Molecular properties are molecular responses to perturbations: geometrical distortions, electric fields, magnetic fields, or their combinations. The proper evaluation of these properties, coming from structure–property relationships, enables to design/discover specific compounds for targeted applications.

The coupled–perturbed Hartree–Fock (CPHF) method [84–94] is used to evaluate molecular responses with respect to a static perturbation.

2.2.1 First–Order Energy Derivative

From the total electronic energy of a closed–shell system, its gradient (first–order derivative) with respect to a general parameter α reads:

$$\begin{aligned}\frac{\partial E^{\mathrm{HF}}}{\partial\alpha} = E^{\mathrm{HF},\alpha} &= 2\sum_{p,q}^{K} H_{pq}^{\alpha}D_{qp} + 2\sum_{p,q}^{K} H_{pq}D_{qp}^{\alpha} + E_{NN}^{\alpha} \\ &+ 2\sum_{p,q}^{K} D_{qp}^{\alpha}\sum_{r,s}^{K} D_{sr}\big[2(pq|rs)-(ps|rq)\big] + \sum_{p,q}^{K} D_{qp}\sum_{r,s}^{K} D_{sr}\big[2(pq|rs)-(ps|rq)\big]^{\alpha}\end{aligned} \tag{2.29}$$

where H_{pq}^{α} are the first–order derivatives of one–electron integrals, $\big[2(pq|rs)-(ps|rq)\big]^{\alpha}$ are the first–order derivatives of two–electron integrals, and D_{qp}^{α} are the first–order derivatives of the density matrix elements. This later one is obtained by a CPHF iterative cycle which can be costly. This cycle can be completely avoided by removing the D_{qp}^{α} dependence:

$$E^{\mathrm{HF},\alpha} = 2\sum_{p,q}^{K} H_{pq}^{\alpha}D_{qp} + \sum_{p,q}^{K} D_{qp}\sum_{r,s}^{K} D_{sr}\big[2(pq|rs)-(ps|rq)\big]^{\alpha} - 2\sum_{p,q}^{K} S_{pq}^{\alpha}W_{qp} + E_{NN}^{\alpha} \tag{2.30}$$

where S_{pq}^{α} are the first–order derivatives of the overlap matrix elements and W_{qp} are the energy–weighted density matrix elements. This new equation satisfies the $\boldsymbol{2n+1}$ **rule** which states that, for variational wavefunctions, (2n+1)th–order perturbations to the energy can be evaluated if the nth–order perturbed wavefunction is known. Thanks to this rule, the second–order energy derivative can be calculated with the first–order energy derivative.

2.2.2 Second–Order Energy Derivative

From the total electronic energy of a closed–shell system, the second–order derivative with respect to α and β parameters, or the differentiation of the first–order derivative with respect to β, reads:

$$\begin{aligned}\frac{\partial^2 E^{\mathrm{HF}}}{\partial\alpha\partial\beta} = E^{\mathrm{HF},\alpha\beta} &= 2\sum_{p,q}^{K} H_{pq}^{\alpha\beta}D_{qp} + 2\sum_{p,q}^{K} H_{pq}^{\alpha}D_{qp}^{\beta} - 2\sum_{p,q}^{K} S_{pq}^{\alpha\beta}W_{qp} - 2\sum_{p,q}^{K} S_{pq}^{\alpha}W_{qp}^{\beta} + E_{NN}^{\alpha\beta} \\ &+ 2\sum_{p,q}^{K} D_{qp}^{\beta}\sum_{r,s}^{K} D_{sr}\big[2(pq|rs)-(ps|rq)\big]^{\alpha} + \sum_{p,q}^{K} D_{qp}\sum_{r,s}^{K} D_{sr}\big[2(pq|rs)-(ps|rq)\big]^{\alpha\beta}\end{aligned} \tag{2.31}$$

where $H_{pq}^{\alpha\beta}$ and $S_{pq}^{\alpha\beta}$ are the second–order derivatives of one–electron integrals, $\big[2(pq|rs)-(ps|rq)\big]^{\alpha\beta}$ are the second–order derivatives of two–electron integrals, D_{qp}^{β} are the first–order derivatives of the density matrix elements, and W_{qp}^{β} are the first–order derivatives of the energy–weighted density matrix elements.

The CPHF method occurs in the determination of $\mathbf{D}^{\beta}$ and $\mathbf{W}^{\beta}$ matrices with an iterative analog SCF procedure because these matrices depend on the spinorbitals that described the wavefunction. So CPHF cycle is introduced by differentiating the Hartree–Fock equation, the normalization condition, and the density matrix definition with respect to the perturbation(s).

When the perturbations are dynamic or a combination of static and dynamic ones, the method used is the time–dependent Hartree–Fock (TDHF) and an oscillating perturbation is introduced.

At the density functional theory level, there are analogs coupled–perturbed and time–dependent methods: coupled–perturbed Kohn–Sham (CPKS) and time–dependent density functional theory (TDDFT) methods [95, 96] where equivalent equations have to be solved. The difference is that HF exchange potential is replaced by an exchange–correlation potential where the evaluation of functional derivatives of the potential with respect to the electron density are needed too.

2.3 Environment Effects

Most applications of targeted systems are in solution and not in gas phase. The solvent interacts with our system approximated as a solute. The solute–solvent interactions have a double–impact on the calculation of molecular properties: they modify the equilibrium structure and they change the response of the solute to the external perturbations.

The explicit treatment of these solute–solvent interactions is too heavy for a computational implementation because all the solvent molecules are taken into account in the simulation. So an implicit treatment is done, often based on the Onsager reaction field (RF) cavity model [97], with an effective Hamiltonian where the solvent part is neglected. It only remains the solute–solvent interactions part:

$$\widehat{H}_{\text{eff}} = \widehat{H}_{\text{solute}} + \widehat{H}_{\text{interactions}} \tag{2.32}$$

In implicit models, the explicit solute–solvent interactions are replaced by a continuous electric field representing the statistical averaging over all solvent degrees of freedom at thermal equilibrium. This field is the reaction field $\vec{E}_{\text{RF}}(\vec{r})$ in the volume occupied by the solute, derived from the electrostatic potential $V(\vec{r})$.

However, the solute dipole moment (permanent and induced) polarizes the solvent which creates a reaction field that back polarizes the solute dipole until self–consistency. The reaction field model has to be replaced by a model which takes into consideration this set of polarizations, the self–consistent reaction field (SCRF) model. Afterwards, the approximation of the solute as a cavity is really important. A realistic description is a superposition of atomic spheres with radii close to the van der Waals values.

This definition leads to two surfaces: the positions assumed by the center of a solvent sphere rolling on the *vdW* surface of the solute define the solvent accessible surface (SAS), that is, the surface enclosing the volume in which the solvent center cannot enter; the same sphere used as a contact probe on the solute surface defines the solvent excluded surface (SES), that is, the surface enclosing the volume in which the whole solvent molecule cannot penetrate (see in Fig. 2.1).

In this SCRF model, the field can be simulated by a charge distribution on the surface of the cavity, the apparent surface charge (ASC) $\sigma(\vec{s})$, employing the polarizable continuum model (PCM) put in a convenient integral equation formalism (IEF) [98–123].

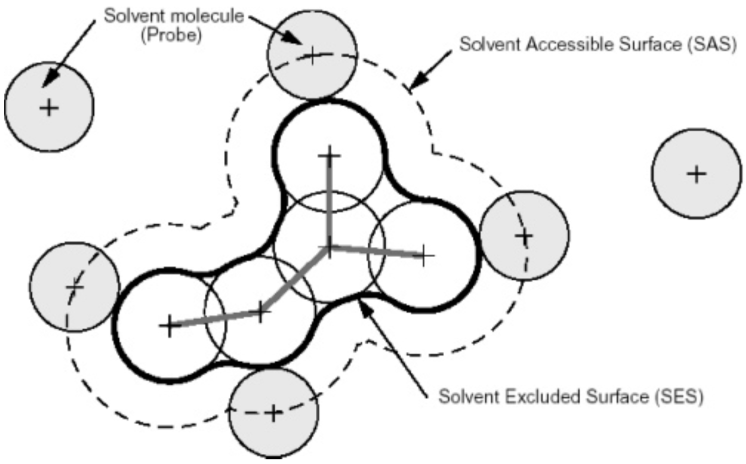


Figure 2.1: Solvent accessible surface (SAS) traced out by the center of the probe representing a solvent molecule. The solvent excluded surface (SES) is the topological boundary of the union of all possible probes that do not overlap with the molecule [98].

In all cases, the ASC defines a potential over the whole space. The cavity surface is approximated in terms of a set of finite elements, "tesserae", small enough to assume that $\sigma(\vec{s})$ is almost constant within each tesserae. Thus, it is possible to define a set of point charges, q_k, in terms of the local value of $\sigma(\vec{s})$ and the corresponding area A_k:

$$V_\sigma(\vec{r}) \simeq \sum_k \frac{\sigma\left(\vec{s}_k\right) A_k}{|\vec{r}-\vec{s}_k|} = \sum_k \frac{q_k}{|\vec{r}-\vec{s}_k|} \tag{2.33}$$

In a practical manner, within each SCF cycle, a second iterative procedure is introduced to evaluate $V_\sigma(\vec{r})$ and $\sigma(\vec{s})$. With the PCM approach, some parameters appeared: the radii type, the cavity type, and the electrostatic scaling factor. These parameters determine how to construct the solute molecular cavity. The default parameters are: Universal Force Field (UFF) atomic radii, van der Waals surface, and a scaling factor α of 1.100.

In the solvation model density (SMD) [124, 125], based on solvation models SMx [126–129], parametric functions are used instead of constant parameters (in PCM). The SMD calculates more accurately the Gibbs free energy of solvation ΔG_s^0. The SMD parameters are: intrinsic Coulomb radii, van der Waals surface, and a scaling factor α of 1.000. We note that the SMD cavity model is smaller than the PCM one.

2.4 Molecular Mechanics

Molecular mechanics (MM) [130] deals with the classical description of molecular and supramolecular systems (via Newtonian mechanics). The simplified assumptions and approximations enable one to use MM for wide applications to various systems, starting from simple low–molecular–weight molecules to large biomolecular complexes or material assemblies of many thousands of atoms. The modern MM can be considered as an extension of simple atom and bond representation of molecules and their complexes in classical chemistry.

In the beginning of the 1950s, scaled paper images and hand–made primitive models of atoms and bonds from wood and wire enabled subtle scientists to construct the first successful models of proteins and DNA [131, 132]. Stick, ball-and-stick, and space-filling plastic models as well as their computer images are widely used until now for both teaching and investigation. The MM approach is a generalization and quantitative representation of these models via mathematical expressions. A set of such expressions together with numerical coefficients is usually referred to as the force field; the MM method itself is frequently referred to as the force field method. Rigid balls of computer graphics or plastic models are replaced by “soft” ones, and hard sticks are replaced by springs. The systems consist of atoms, each atom being simulated as a single point particle. Molecular mechanics describe molecules in terms of “bonded atoms”, their positions are distorted from some idealized (equilibrium) geometries due to non–bonded interactions with the other atoms. MM does not include electrons explicitly, but different configurations of electron shells of the same chemical elements are implicitly considered in force field parameters.

2.4.1 Energy Calculation

The first principal approximation of MM method is the mathematical description of the total energy of the system as a sum of energy terms, responsible for the contributions of different physical nature, and contributions of the atoms. The simple representation of the potential energy of a molecule or molecular complex as a function of atom coordinates, R, can be represented by a sum of four main terms each term being a sum of many contributions [130]:

$$\Delta E(R) = \sum Eb + \sum Ea + \sum Et + \sum Enb\,. \tag{2.34}$$

The summation up in the terms is over all the chemical bonds Eb, valence angles Ea, torsion angles Et, and all pairs of atoms non–bonded to each other or to common third atom Enb. The energy terms depend on mutual positions of atoms and on the adjustable constants (parameters); the parameters are suggested to be transferable between atoms and molecules of the same type. The transferability of the force field parameters is the second important assumption of MM, and the number of atom types of the same chemical element depends on particular force field.

The Eb term refers to variations of bond lengths, so stretching, and the Ea refers to variations of bond angles, so bending. These terms are usually modeled as harmonic potentials centered on equilibrium values of bond lengths and bond angles, respectively [130]:

$$Eb_i = kb_i\big(r_i - r_i^0\big)^2 \qquad Ea_i = ka_i\big(\alpha_i - \alpha_i^0\big)^2\,, \tag{2.35}$$

where r_i and α_i are the values of bond lengths and bond angles, respectively with the ^{0} parameters for equilibrium values. kb and ka are the stretching force constant and the bending force constant, respectively. These adjustable parameters depend on the type of the atom. Some force fields may also contain cubic and higher-order contributions to these terms, or sometimes more flexible Morse potential can be used instead, as well as “cross terms” can be included to account for correlations between stretching and bending components.

The Et refers to the changes of torsion energy so interactions of electron shells of two atoms A and D which are connected through an intermediate chemical bond B–C [130]:

$$Et_i = kt_i\big(1 + \cos(n_i\phi_i - \delta_i)\big) \,, \tag{2.36}$$

where ϕ is the angle of rotation around the chemical bond B–C, kt is the torsion force constant, n is the multiplicity, and δ is the phase angle. Many force fields include terms responsible for "improper" torsions or out-of-plane bending, so terms related to four atoms not forming consecutive chemical bonds, which function as correction factors for out-of-plane deviations. These terms can be expressed via harmonic potentials like those for bond stretching and valence angle bending. Cross terms depending on both torsion angle and bond length or valence angle are added in some force fields.

The Enb refers to non–bonded interactions of all atom pairs not bonding to each other or to the same third atom. Each atom–atom term is usually represented by a sum of a Coulomb term and van der Waals interactions [130]:

$$Enb_{ij}(r_{ij}) = \frac{Kq_iq_j}{r_{ij}} - \frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} \;\;, \tag{2.37}$$

where r_{ij} is the distance between i and j atoms, q_i and q_j are their effective atom charge, A_{ij} is the adjustable parameter for dispersion (London) attraction interaction, and B_{ij} is the adjustable parameter for short–range repulsion interaction.

Most of the early force fields used for description of van der Waals interactions are Buckingham (6–exp) potential instead of Lennard–Jones (6–12) potential. The total expression for non–bonded interaction term are usually referred to as (1–6–12) or (1–6–exp) potential relating to the dependency of the terms on the interatomic distance. The Buckingham potential is more flexible but it is less convenient for computations. It has a maximum at short distance and then tend to negative infinite value. The majority of modern force fields utilize 6–12 expressions for description of van der Waals interactions, the total atom–atom potential being referred to as 1–6–12 one. Some force fields substitute 6–12 potential by 10–12 one for the interactions of hydrogen atoms of hydrogen bonds to describe more sharp distance dependence in the most important area of energy minimum corresponding to H–bond formation (referred to as 1–10–12 potential).

The calculations of potential energy are used to search for local energy minima (mutual atom positions corresponding to possible stable configurations), to construct and analyze multidimensional potential energy surfaces (PES), to follow trajectory of movement (in molecular dynamics simulations), or to study averaged thermodynamic and geometry characteristics (via Monte Carlo sampling) of the systems.

2.4.1.1 Monte Carlo versus Molecular Dynamics

The Monte Carlo (MC) and the molecular dynamics (MD) have the same goal: sampling the PES. In the MC method, the configurations are generated randomly with an energy criterion so the average is over the configurations. The new configurations are accepted or rejected thanks to the criterion and these new ones depend only on its predecessor and not the other previous configurations. Multiple MC simulations are done at different temperature and there is no time dependence. In the MD method, the particles have a velocity ν_i so $K_i = \frac{1}{2}m_i\nu_i^2$ and the configurations are generated thanks to the kinetic energy in a period of time with a time step parameter so the average is over the time. Some parameters can be fixed constant as the number of molecules, the volume, the temperature or the pressure. It involves to work with thermostat or barostat. The MD method is function of time so is “dynamic”, contrary to MC method.

2.4.2 Universal Force Field

The Universal Force Field (UFF) has been developed by A. K. Rappé *et al.* [133] in order to describe more systems and to use all the elements of the periodic table with simple relations.

The UFF energy expression can be written as:

$$E_{\text{UFF}} = E_R + E_\theta + E_\phi + E_\omega + E_{\text{vdW}} + E_{el}\,, \tag{2.38}$$

where the six constituent terms are defined below.

UFF describes the bond stretching as either an harmonic oscillator or a Morse function:

$$E_R = \frac{1}{2}k_{ij}\big(r - r_{ij}\big)^2 \tag{2.39}$$

$$E_R = D_{ij}\left(e^{-\alpha(r-r_{ij})} - 1\right)^2, \tag{2.40}$$

where k_{ij} is the force constant (kcal/molÅ^2), r_{ij} is the bond lengths and D_{ij} is the bond dissociation energy. The Morse function is a more accurate description since it implicitly includes anharmonic terms near equilibrium and leads to a finite energy for breaking bonds.

In UFF, the angle bending is described with a small cosine expansion:

$$E_\theta = k_{ijk}\sum_{n=0}^{m} C_n \cos(n\,\theta)\,, \tag{2.41}$$

where the coefficients C_n are chosen to satisfy appropriate boundary conditions including that the function have a minimum at the natural bond angle θ_0.

The torsional term for two bonds ij and kl connected via a common bond jk is described by:

$$E_\phi = \frac{1}{2}V_\phi\Big(1 - \cos(n\phi_0)\cos(n\phi_{ijkl})\Big)\,, \tag{2.42}$$

where V_ϕ is the rotational barrier.

For UFF, a one– or two–term cosine expansion is used for atoms i bonded exactly to three other atoms j, k, and l:

$$E_\omega = k_{ijkl}\Bigg(C_0 + C_1\cos(\omega_{ijkl}) + C_2\cos(2\omega_{ijkl})\Bigg)\,, \tag{2.43}$$

where k_{ijkl} is the force constant (in $\mathrm{kcal\,mol^{-1}}$) and ω_{ijkl} is the angle between the il axis and the ijk plane.

Non–bonded interactions (van der Waals forces) are included in the UFF as a Lennard–Jones 6–12 type expression:

$$E_{\mathrm{vdW}} = D_{ij}\Bigg[-2\bigg(\frac{x_{ij}}{x}\bigg)^{6} + \bigg(\frac{x_{ij}}{x}\bigg)^{12}\Bigg]\,, \tag{2.44}$$

where D_{ij} is the well depth (in $\mathrm{kcal\,mol^{-1}}$) and x_{ij} is the van der Waals bond length.

The electrostatic interactions are calculated by:

$$E_{el} = \frac{332.0637\,Q_i\,Q_j}{\epsilon\,R_{ij}}\,, \tag{2.45}$$

where Q_i and Q_j are charges, R_{ij} is the distance, and ϵ is the “dielectric constant” with $\epsilon = 1$.

2.4.3 Merck Molecular Force Field 94

The force field “Merck Molecular Force Field 94” (MMFF94) has been developed by Thomas A. Halgren [72, 134–140] in order that organic molecules are properly described.

This force field has primarily been derived from computational data: 500 molecular structures optimized at the HF/6–31G* level, 475 structures optimized at the MP2/6–31G* level, 380 MP2/6–31G* structures evaluated at a defined approximation to the MP4SDQ/TZP level, and 1450 structures partly derived from MP2/6–31G* geometries and evaluated at the MP2/TZP level.

The MMFF94 energy expression can be written as:

$$E_{\mathrm{MMFF94}} = \sum EB_{ij} + EA_{ijk} + EBA_{ijk} + EOOP_{ijk;l} + ET_{ijkl} + E_{\mathrm{vdW}_{ij}} + EQ_{ij}\,, \tag{2.46}$$

where the seven constituent terms are defined below.

MMFF94 employs the quartic function for the bond stretching:

$$EB_{ij} = 143.9325\,\frac{\mathrm{k}\,b_{\mathrm{IJ}}}{2}\,\Delta r_{ij}^2\,(1 + cs\,\Delta r_{ij} + \frac{7}{12}\,cs^2\,r_{ij}^2)\,, \tag{2.47}$$

where $\mathrm{k}\,b_{\mathrm{IJ}}$ is the force constant (md/Å), Δr_{ij} is the difference between actual and reference bond lengths, and cs is the cubic–stretching constant.

MMFF94 uses the cubic expansion for the angle bending:

$$EA_{ijk} = 0.043844\,\frac{\mathrm{k}\,a_{\mathrm{IJK}}}{2}\,\Delta\theta_{ijk}^2 \times (1 + cb\,\Delta\theta_{ijk})\,, \tag{2.48}$$

where $\mathrm{k}\,a_{\mathrm{IJK}}$ is the force constant (md Å/rad^2), $\Delta\theta_{ijk}$ is the difference between actual and reference bond angles, and cb is the cubic–bending constant.

MMFF94 employs the form for the stretching–bending interactions:

$$EBA_{ijk} = 2.51210\,(\mathrm{k}\,ba_{\mathrm{IJK}}\,\Delta r_{ij} + \mathrm{k}\,ba_{\mathrm{KJI}}\,\Delta r_{kj})\,\Delta\theta_{ijk}\,, \tag{2.49}$$

where $\mathrm{k}\,ba_{\mathrm{IJK}}$ and $\mathrm{k}\,ba_{\mathrm{KJI}}$ are the force constants (md/rad) that couple the i–j and k–j stretchings to the i–j–k bending. Δr and $\Delta\theta$ are defined as above.

MMFF94 uses the form for the out–of–plane bending, at tricoordinate centers:

$$EOOP_{ijk;l} = 0.043844\,\frac{\mathrm{k}\,oop_{\mathrm{IJK:L}}}{2}\,\Delta\chi^2_{ijk;l}\,, \tag{2.50}$$

where $\mathrm{k}\,oop_{\mathrm{IJK:L}}$ is the force constant (md Å/rad^2) and $\chi_{ijk;l}$ is the Wilson angle (degrees) between the bond j–l and the plane i–j–k.

MMFF94 uses the threefold representation for the torsion interactions, employed in MM2 and MM3, where Φ is the i–j–k–l torsion angle:

$$ET_{ijkl} = \frac{1}{2}\left[V_1(1+\cos\,\Phi) + V_2(1-\cos\,2\Phi) + V_3(1+\cos\,3\Phi)\right], \tag{2.51}$$

where V_1, V_2, and V_3 depend on the atom types I, J, K, and L for atoms i, j, k and l, where i–j, j–k, and k–l are bonded pairs.

MMFF94 employs the recently developed "Buffered–14–7" form [72] (terminology derives from the formal 14th and 7th power dependencies for the repulsive and attractive terms that would be obtained if the R^*_{IJ} "buffering constants" in the denominators were deleted) for the van der Waals interactions:

$$E_{\mathrm{vdW}_{ij}} = \epsilon_{IJ}\left(\frac{1.07\,R^*_{IJ}}{R_{IJ}+0.07\,R^*_{IJ}}\right)^7\left(\frac{1.12\,R^{*7}_{IJ}}{R^7_{IJ}+0.12\,R^{*7}_{IJ}}-2\right). \tag{2.52}$$

This form is used in conjunction with an expression that relates the minimum–energy separation R^*_{II} to the atomic polarizability α_I, with specially formulated combination rules, and with a Slater–Kirkwood expression for the well depth ϵ_{IJ}.

MMFF94 uses the buffered coulombic form for the electrostatic interactions:

$$EQ_{ij} = \frac{332.0716\,q_iq_j}{D\,(R_{ij}+\delta)^n}\,, \tag{2.53}$$

where q_i and q_j are partial atomic charges, R_{ij} is the internuclear separation (Å), $\delta = 0.05$ Å is the "electrostatic buffering" constant and D is the "dielectric constant". Normally, exponent $n = 1$ but $n = 2$ is also supported for distance–dependent dielectric constant.

2.4.4 Global–MMX

The Global–MMX (GMMX) is a steric energy minimization program which uses the currently selected force field for the global energy minimum and other low energy local minima. The conformational searching techniques are unique in their approach and are inspired from the methods described by the Still groups [141]. The search techniques in GMMX are based on the methods used in BAKMDL, developed by Professor Kosta Steliou of Boston University, and ported to the MMX force field by Mark Midland of UC Riverside and Joe Gajewski and Kevin Gilbert of Indiana University. These methods are unpublished work. GMMX is initially part of the PCMODEL program [142, 143] and has been implemented in GaussView [144].

There are two stages. The first stage randomly searches over the selected rings and rotatable bonds and keeps all the conformers minimized within 3.5 kcal of the lowest energy conformer found during the minimization. The second cycle reminimizes the structures found in the first cycle and keeps only those that are within 3.0 kcal of the lowest energy conformer.

2.4.5 Conformer–Rotamer Ensemble Sampling Tool

The conformer–rotamer ensemble sampling tool (CREST) [145] is a conformational search tool for the automatized exploration of the low–energy chemical structure space with no bond breaks/formations. The application of CREST is the generation of conformer ensembles which uses the iMTD–GC algorithm [146]. This tool is included in the xTB program and the GFN2–xTB is used.

2.4.5.1 Conformer Ensembles

A set of different conformers and their rotamers within a certain energy window around the same global covalent potential energy minimum is referred to as the conformer–rotamer ensemble (CRE). Conformers are stereoisomers of a molecule that differ only in their conformation but have the same covalent topology. They can be characterized by a distinct potential energy minimum on the PES. Rotamers have degenerate potential energy minima and thus are indistinguishable by any nuclear spin–independent quantum mechanical observable computed at the respective minima. A conformation is the Born–Oppenheimer–approximated equilibrium conformer including all its rotamers.

For the calculation of properties, it is necessary to include different molecular conformations by averaging the individually obtained Boltzmann weighted property of each constituent in the ensemble. To avoid multiple counting, which leads to incorrect Boltzmann averages so incorrect averaged properties. The distinction between identical structures is possible on the basis of three–dimensional structures and free energies of the isomers. Then, the energy is employed as a criterion.

In CREST, the root-mean-square deviation (RMSD) of atomic Cartesian coordinates and the difference between rotational constants (B_e) of two molecules are used for the comparison of their structure. However, these parameters do not tell us if the structures are equilibrium geometries or high–energetic intermediates. Then, these parameters are combined with the energy criterion. The distinction between two structures are represented in Fig. 2.2.

The conformers are different PES minima or, if the energy difference is small, have high RMSD and unequal rotational constant. The rotamers have similar energies and equal rotational constant but high RMSD. Identical structures have similar energies, equal rotational constant and very small RMSD. These ones are discarded as duplicates. The final CRE consists of all unique conformers and rotamers within a certain energy window. The choice for this window depends on the accuracy of the used QC and the type of application.

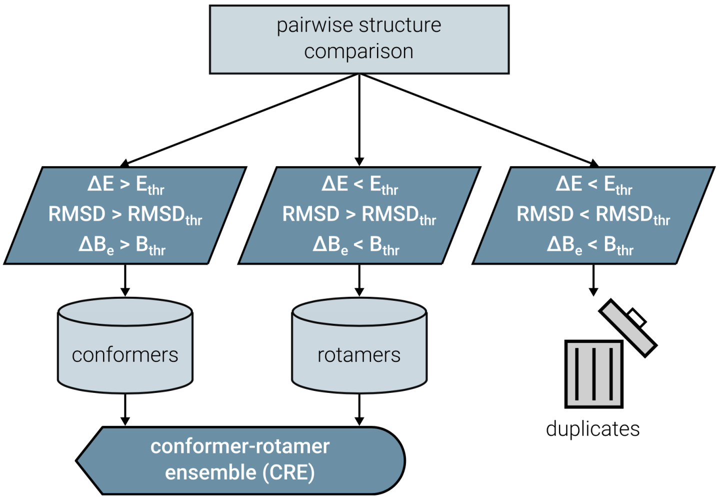


Figure 2.2: Schematic representation [145] of the sorting criteria to distinguish between identical structures, conformers and rotamers where E_{thr}, $RMSD_{thr}$, and B_{thr} are the respective predefined thresholds for the energy, atomic RMSD between the considered pair, and rotational constant.

The quality of an ensemble is related to its completeness which can be assessed by a maximized entropy S_{CR} according to the standard thermodynamic expressions:

$$S_{\mathrm{CR}} = R \sum_{i=1}^{\mathrm{CRE}} p_i \log(p_i)\,, \tag{2.54}$$

where the sum runs over all populations p_i of all species with energy ΔE_i at temperature T:

$$p_i = \frac{e^{-\Delta E_i/RT}}{\sum_{j=1}^{\mathrm{CRE}} e^{-\Delta E_j/RT}}\,. \tag{2.55}$$

The ensemble entropy S_{CR} is also linked to the ensemble free energy $G_{CR} = -T\,S_{CR}$, which is minimized for a complete CRE. This completeness criterion of a maximized S_{CR} only holds if the global minimum conformation is included in the ensemble and breaks down otherwise. A maximized ensemble entropy was used for determining technical parameter sets that are employed in the CREST program [146].

2.4.5.2 Algorithmic Details

Generating conformations is impractical for large and flexible molecules. This approach requires the manual *a priori* definition of the conformational coordinates. A meta-dynamics (MTD) simulation has been proposed to tackle up this problem. The MTD simulation is a based screening procedure that can be routinely used for the generation of molecular conformations in the gas–phase or in implicit solvation [146] in order to explore the PES. A history-dependent biasing potential is applied, where the collective variables (CVs) for the meta-dynamics are previous structures on the PES, expressed as atomic RMSD between them, which is calculated according to a quaternion algorithm [147]. This concept of MD simulations [148, 149] has already been used in the context conformational changes [150–154]. According to Pracht *et al.*, this is the first combination of MTD simulations with atomic RMSDs in order to generate conformers.

The biasing contribution is given in the form of a Gaussian potential by:

$$V_{\text{bias}} = \sum_{i=1}^{n} k_i \, e^{-\alpha_i \, \Delta_i{}^2} \,, \tag{2.56}$$

where the RMSDs enter as collective variables Δi, n is the number of reference structures, k_i is the pushing strength and the parameter α_i determines the potentials shape. From this energy expression, atomic forces are derived that enter as additional forces in the MTD simulations, which is also referred to as guiding forces. Since the addition of each bias Gaussian potential drives the structure further away from previous geometries, this allows otherwise unlikely high–barrier crossings where all atoms collectively explore huge regions of the PES (see Fig. 2.3).

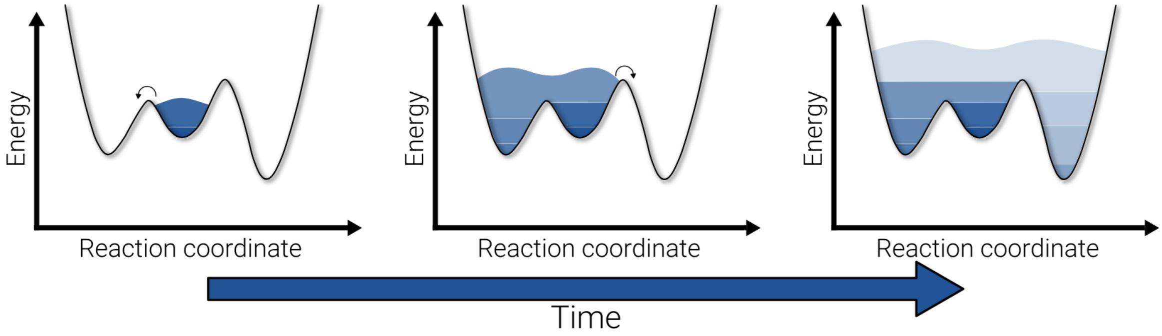


Figure 2.3: Schematic one dimensional PES [145] that is "filled" by several bias potentials over time, which allows larger barrier heights to be overcome.

However, instabilities appeared with the instantaneous addition of a biasing potential in the MD with large time steps, due to extensive local heating of the system. So the potential is made time–dependent and a damping function is used. The choice of the collective variables in MTDs is important and the standard root–mean–square deviation in Cartesian space as a metric is employed. The RMSD between the actual structure and a reference one reads:

$$\Delta_i = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left(r_j - r_j^{\text{ref},i} \right)^2} \,, \tag{2.57}$$

where r_j is a component of the Cartesian space vector of the actual molecule, $r_j^{\text{ref},i}$ is the corresponding element in reference structure i (with the same atom numbering), and N is the number of atoms.

For the automatized generation of conformers, Pracht *et al.* developed a composite approach consisting of MTD sampling, regular MD sampling and a procedure that is related to genetic structure crossing algorithms (GC) [155–157]. Hence, the procedure was termed iMTD–GC, where the lowercase i indicates an iterative strategy within the algorithm. The general workflow is outlined in Fig. 2.4.

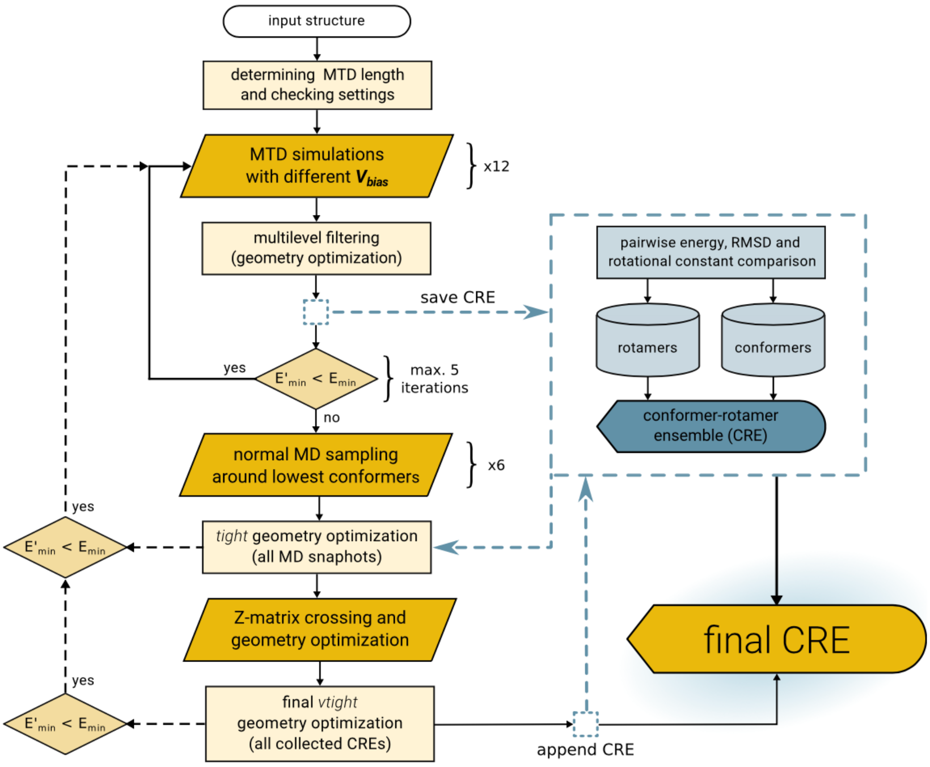


Figure 2.4: Outline of the iMTD-GC workflow [145].

First, the maximum MTD length is determined, which mainly depends on the molecular size and flexibility of the system. Then, technical general settings are evaluated, to check if the MTDs will run stable. The main step of each iMTD-GC conformational search is the MTD sampling. The automatization is the key step and different molecules require adjusted pairs of k_i and α_i to produce the best results. Then, a set of twelve MTDs is performed with different settings for the V_{bias} parameters. α_i has values between 0.1 and 1.3 $a_0{}^{-2}$, which can be seen as the “range” of the bias. k_i is scaled by the number of atoms N. The k_i/N has magnitudes of 0.75 to 3.00 mE_h.

All simulations are run with a standard algorithm in the NVT ensemble (number, volume and temperature are kept constant) using Berendsen thermostat [158] at a heat transfer time constant of 0.5 ps. The covalent bonds are constrained using the SHAKE algorithm [159] with an MD time step $d\tau$ of 5 fs to integrate the equations of motion.

Within the MTD simulation a new structure is added to the V_{bias} potential every 1.0 ps, which constantly drives the molecule into new conformations as time progresses. However, since the PES is constantly modified by the bias potential, the conformers form the MTD trajectory cannot directly compared to each other and have to be re–optimized without the biasing potential. This is done in a two–step filtering procedure, with very crude and then with tight convergence criteria.

Afterwards the re–optimized structure snapshots from the trajectory are sorted according to the procedure outlined in Fig. 2.2, which yields an initial CRE. If a new conformer is found that is lower in energy than the input structure, the entire procedure is restarted on this conformer, otherwise the workflow is continued. By default, the MTD iteration is restarted at least once, but not more than five times. All intermediate CREs are saved to be compared at a later stage.

In the second step, two unbiased MD simulations (at 400 K and 500 K) are run on the three lowest conformers. This is done to get conformations with low–energy barrier crossings, opposed to high–barrier conformational changes that can be obtained by the MTD simulations. The low–energy barrier crossings include simple torsional motions, such as group rotations, which are needed to complete the CRE regarding the rotamers. All structures from the MD simulations are sorted again and included in the intermediate CRE.

In the final step, the genetic structure crossing is performed with automatically generated Z–matrices [146]. Together with the regular MD simulations, this approach helps to further complete the CRE and is particularly useful for flexible systems. If in the MD or GC step a new lowest energy conformer is found, the entire procedure is restarted. However, unlike the MTD iterations, these iterations do not have a maximum number of cycles and will only terminate if the lowest conformer does not change any more. The advantage of these restarts is mainly observed for larger molecules whose global minimum structure can be way off the initial input geometry. All collected CREs are then optimized once more with very tight energy convergence criteria and the final CRE is created. Various energy thresholds and other MTD settings are employed within the workflow.

The optimizations are performed at the GFN2–xTB level. It seems to be consensus in the literature that geometries are often quite well reproduced by semiempirical quantum mechanics methods [160, 161]. This is particularly true for methods of the GFN–xTB family, which, as their name conveys, are parameterized to yield reasonable structures [162, 163]. Several recent studies also show that the GFN–xTB methods are among the best performing semiempirical methods for conformational energies and geometries [164, 165]. The good trade–off between accuracy and computational cost enables the use of GFN–xTB for the generation of conformers with the iMTD-GC workflow. Although it is technically possible, any higher level theoretical method, even low cost DFT with a small basis set, would be much to expensive for the vast amount of required geometry optimizations and evaluations.

Chapter 3. Vibrational Spectroscopy

The spectroscopy is the study of the interaction between the matter and a radiation, as a function of the wavelength or frequency of the radiation. In physics, the radiation is the emission or transmission of energy in the form of waves or particles through space or material medium. There are different methods, function of the type of the radiative energy and the nature of the interaction: electromagnetic radiation, electron or neutron particles, acoustic radiation, . . . ; absorption, emission, elastic scattering, inelastic scattering, . . .

The IR and Raman spectroscopies are spectroscopic techniques used to determine the vibrational modes of molecules. The type of the radiative energy is an electromagnetic radiation. The nature of the interaction is an absorption phenomenon for the IR spectroscopy and an inelastic scattering phenomenon for the Raman spectroscopy.

3.1 Scattering Phenomenon

When light encounters matter, the light is scattered in all directions of space (see left in Fig. 3.1). If the energy of the scattered light is the same as the energy of the incident light, so an elastic scattering, it is called “Rayleigh scattering” (see right in Fig. 3.1). If the energy of the scattered light is different than the energy of the incident light, so an inelastic scattering, it is called “Raman scattering” (see right in Fig. 3.1). In this case, there are two possibilities: “Stokes bands” when the energy of the scattered light is lower than the energy of the incident light; and “Anti–Stokes bands” when the energy of the scattered light is higher than the energy of the incident light. The energy difference is related to the vibrational transitions of the system.

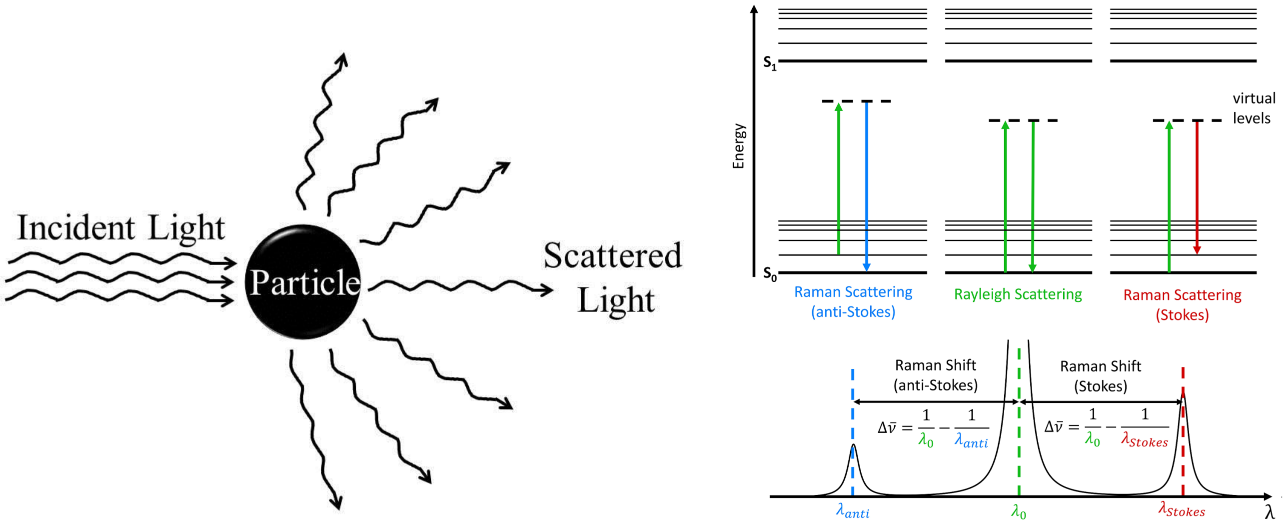


Figure 3.1: Sketch of scattering phenomenon (left) and Jablonski diagram (right).

3.2 Vibrations

Each system, composed of N particles ($\equiv$ atoms), possesses $3N$ degrees of freedom: 3 are associated to translations, 3 are associated to rigid rotations, and the rest $(3N-6)$ is associated to vibrations (except for linear systems: only 2 degrees of freedom associated to rotations so $3N-5$ vibrations).

Each atom coordinates $\vec{r}_{A\xi}$ can be decomposed into its three Cartesian components (x_A, y_A, z_A), where ξ denotes these components. The coordinates of atom A are given relative to their equilibrium position $\vec{r}_{A\xi,0}$ ($x_{A,0}$, $y_{A,0}$, $z_{A,0}$). Then, **mass–weighted Cartesian coordinates** q_i can be introduced: $q_1 = \sqrt{m_A}(x_A - x_{A,0})$, $q_2 = \sqrt{m_A}(y_A - y_{A,0})$, $q_3 = \sqrt{m_A}(z_A - z_{A,0})$, $q_4 = \sqrt{m_B}(x_B - x_{B,0})$, … This allows to express the kinetic energy in terms of time–derivatives of these coordinates:

$$T = \frac{1}{2}\sum_{i=1}^{3N} \dot{q}_i^2 \, . \tag{3.1}$$

The potential energy can be developed as a Taylor expansion at the equilibrium geometry:

$$V = V^0 + \sum_{i=1}^{3N}\left(\frac{\partial V}{\partial q_i}\right)_0 q_i + \frac{1}{2}\sum_{i,j}^{3N}\left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)_0 q_i q_j + \frac{1}{6}\sum_{i,j,k}^{3N}\left(\frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k}\right)_0 q_i q_j q_k + \cdots \ . \tag{3.2}$$

At the equilibrium geometry, the force constants $\left(\partial V/\partial q_i\right)_0 = f_i^{\mathrm{MW}}$ are equal to zero and the V^0 is chosen equal to zero. For small amplitudes of vibration, the **harmonic approximation** is used and the higher (anharmonic) terms are neglected:

$$V \approx \frac{1}{2}\sum_{i,j}^{3N}\left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)_0 q_i q_j = \frac{1}{2}\sum_{i,j}^{3N} f_{ij}^{\mathrm{MW}} q_i q_j \, , \tag{3.3}$$

where $f_{ij}^{\mathrm{MW}} = f_{ji}^{\mathrm{MW}}$ is the ij component of the **mass–weighted Hessian** denoted $\mathbf{H}^{\mathrm{MW}}$.

3.2.1 Normal Coordinates

At this point, **normal coordinates** are defined as **linear transformations** of the mass–weighted Cartesian coordinates:

$$|Q_k\rangle = \sum_{i}^{3N} |q_i\rangle \, U_{ik} \qquad k = 1, 2, \ldots, 3N \, , \tag{3.4}$$

and the inverse transformation reads:

$$|q_i\rangle = \sum_{i}^{3N} |Q_k\rangle \, (U^T)_{ki} \qquad i = 1, 2, \ldots, 3N \, , \tag{3.5}$$

where U_{ik} is a component of the unitary matrix transformation ($\mathbf{U}$) whose columns define the normal modes in terms of mass–weighted Cartesian coordinates and whose rows define the mass–weighted Cartesian coordinates in terms of normal modes.

The coefficients U_{ik} are chosen so that the kinetic and potential energies are separable and read, in normal coordinates:

$$T = \frac{1}{2}\sum_{i=1}^{3N} \dot{q}_i^2 = \frac{1}{2}\sum_{i=1}^{3N}\sum_{k,l}^{3N}(U^T)_{ki}U_{il}\dot{Q}_k\dot{Q}_l = \frac{1}{2}\sum_{k,l}^{3N}\delta_{kl}\dot{Q}_k\dot{Q}_l = \frac{1}{2}\sum_{k=1}^{3N}\dot{Q}_k^2\,; \tag{3.6}$$

$$V = \frac{1}{2}\sum_{i,j}^{3N} f_{ij}^{\mathrm{MW}} q_i q_j = \frac{1}{2}\sum_{i,j}^{3N}\sum_{k,l}^{3N}(U^T)_{ki}f_{ij}^{\mathrm{MW}}U_{jl}Q_kQ_l = \frac{1}{2}\sum_{k=1}^{3N}\lambda_k Q_k^2\,, \tag{3.7}$$

where δ_{kl} is the the Kronecker delta symbol which equals unity if $k = l$ and is zero otherwise.

These two equations imply that $\mathbf{U}$ is unitary and contains the eigenvectors of the **mass–weighted Hessian $\mathbf{H}^{\mathrm{MW}}$**. The eigenvalue problem reads:

$$\sum_{i,j}^{3N}(U^T)_{ki}f_{ij}^{\mathrm{MW}}U_{jl} = \delta_{kl}\lambda_k\,; \tag{3.8}$$

$$\mathbf{U^T}\mathbf{H}^{\mathrm{MW}}\mathbf{U} = \boldsymbol{\lambda} = \boldsymbol{\omega}^2\,, \tag{3.9}$$

where $\boldsymbol{\lambda}$ is a diagonal matrix containing the different eigenvalues λ_k.

The eigenvalues give the frequencies of the normal modes: $\omega_k^2 = 4\pi^2\nu_k^2 = 4\pi^2c^2\overline{\nu}_k^2$. The eigenvectors are the normal coordinate expression in term of the mass–weighted Cartesian coordinate ones: $U_{ik} = \langle q_i|Q_k\rangle = Q_{ik}$. The components of the k mode in terms of Cartesian coordinates are: $Q^c_{A\xi,k} = \big(1/\sqrt{m_A}\big)U_{ik} = \big(1/\sqrt{m_A}\big)Q_{ik}$.

3.2.2 Hamiltonian

In the **harmonic approximation**, the Hamiltonian for the nuclei is separable into $3N-6$ equation, one per normal mode of vibration:

$$\hat{H} = \sum_{k=1}^{3N-6}\left(-\frac{\hbar^2}{2}\frac{\partial^2}{\partial Q_k^2} + \frac{1}{2}\omega_k^2 Q_k^2\right) = \sum_{k=1}^{3N-6}\hat{H}_k\,. \tag{3.10}$$

Solving one of these equations corresponds to solving a particle in a 1D harmonic potential $V = \frac{1}{2}\,\omega_k^2\,Q_k^2$ called the **harmonic oscillator** model.

The total nuclear energy and nuclear wavefunction are the sum of the energy and the product of the wavefunction, respectively, of the different harmonic oscillators having each a given angular frequency ω_k.

$$\big|\nu^i\big\rangle = \prod_{k=1}^{3N-6}\big|\nu_k^i(Q_k)\big\rangle \qquad \text{and} \qquad E_{\nu^i} = \sum_{k=1}^{3N-6}E^k_{\nu_k^i}\,. \tag{3.11}$$

The solution of the harmonic oscillator reads:

$$\big|\nu_k^i(Q_k)\big\rangle = \sqrt{\frac{1}{2^{\nu_k^i}\,\nu_k^i!}}\left(\frac{\omega_k}{\pi\hbar}\right)^{1/4} e^{-\frac{\omega_k Q_k^2}{2\hbar}}H_{\nu_k^i}\left(\sqrt{\frac{\omega_k}{\hbar Q_k}}\right) \qquad \text{and} \qquad E^k_{\nu_k^i} = \hbar\omega_k\big(\nu_k^i + \frac{1}{2}\big)\,, \tag{3.12}$$

where $\nu_k^i = 0, 1, 2, \ldots$ is the quantum number associated to the k normal mode for the i electronic state and $H_{\nu_k^i}$ are the Hermite polynomials (Fig. 3.2).

The Hermite polynomials satisfy the equation:

$$H_\nu'' - 2xH_\nu' + 2\nu H_\nu = 0\,, \tag{3.13}$$

and the solutions are recursive:

$$H_{\nu+1} = 2xH_\nu - 2\nu H_{\nu-1}\,. \tag{3.14}$$

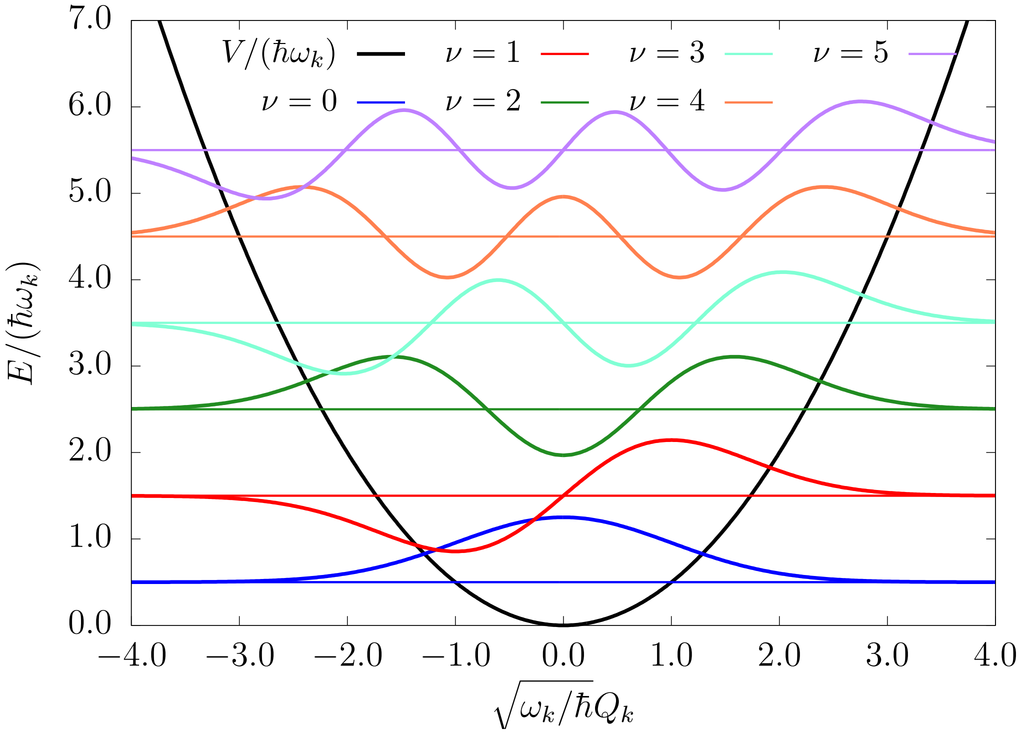


Figure 3.2: Sketch of the harmonic potential (in black) and the first energy levels with their associated wavefunction (in other colors).

Figure 3.3 represents the total energy levels for a model system with only 3 harmonic oscillators. The lowest energy level is called ground level or **zero–point energy** ($E = \frac{1}{2}\hbar\omega_k$) and corresponds to all quantum numbers being zero. The energy levels for which only one normal mode is excited, with a quantum number of one, are called fundamental levels. The energy levels for which only one normal is excited, but with a quantum number greater than one, are called overtone levels. The energy levels with multiple non–zero quantum numbers are called combination levels. The ground state is supposed to have the greatest population so fundamental transitions are the most important.

3.3 Double Harmonic Approximation

In the Born–Oppenheimer approximation, the total wavefunction can be decomposed as a product between the electronic and nuclear parts: $\left|\Psi^{\text{elec}}\right\rangle\left|\phi^{\text{nuc}}\right\rangle = \left|\Psi^{\text{elec}}\right\rangle\left|\phi^{\text{vib}}\right\rangle$. In fact, the nuclear part is a product of the translational, rotational, and vibrational parts. Here, the nuclear part is approximated to the vibrational part only.

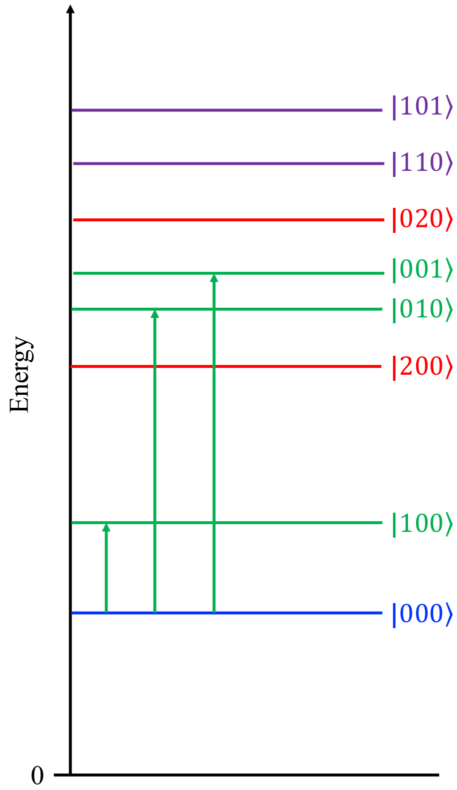


Figure 3.3: Sketch of first total energy levels for a model system with only 3 harmonic oscillators ($\equiv$ 3 normal modes). The fundamental transitions are represented in green.

For any vibrational spectroscopies, the intensity associated to the fundamental transitions are given by the following integral:

$$P_\xi^{\text{elec}} = \left\langle \phi_0^{\text{vib}} \right| \left\langle \Psi_g^{\text{elec}} \right| \hat{P} \left| \Psi_0^{\text{elec}} \right\rangle \left| \phi_{1_\xi}^{\text{vib}} \right\rangle = \left\langle \phi_0^{\text{vib}} \right| P^{\text{elec}}(Q) \left| \phi_{1_\xi}^{\text{vib}} \right\rangle , \tag{3.15}$$

where $P^{\text{elec}}(Q) = \left\langle \Psi_g^{\text{elec}} \right| \hat{P} \left| \Psi_0^{\text{elec}} \right\rangle$ is the expectation value of the property operator and is function of the nuclear/normal coordinates.

In fundamental transitions, only one vibrational mode ξ is excited at the time and with a quantum number of one ($\nu_\xi = 1$). The vibrational excited state is denoted $\left|\phi_{1_\xi}\right\rangle = |1_\xi\rangle$

The expectation value of the property P^{elec} can be expressed as a Taylor series in Q:

$$P^{\text{elec}}(Q) = P^0 + \sum_k \left(\frac{\partial P}{\partial Q_k}\right)_0 Q_k + \frac{1}{2}\sum_{k,l}\left(\frac{\partial^2 P}{\partial Q_k \partial Q_l}\right)_0 Q_k Q_l \cdots . \tag{3.16}$$

In the **electric harmonic approximation**, only the linear terms are included and the intensity, associated to a fundamental transition from $|0\rangle \rightarrow |1_\xi\rangle$, reads:

$$P_\xi^{\text{elec}} = \sqrt{\frac{\hbar}{2\omega_\xi}} \left(\frac{\partial P}{\partial Q_\xi}\right)_0 . \tag{3.17}$$

In the **double–harmonic approximation**, the **harmonic approximation** and the **electric harmonic approximation** are used. The intensity associated to a normal mode ξ is related to the derivative of a property with respect to the normal mode coordinates and only fundamental transitions are allowed.

3.4 Raman Optical Activity

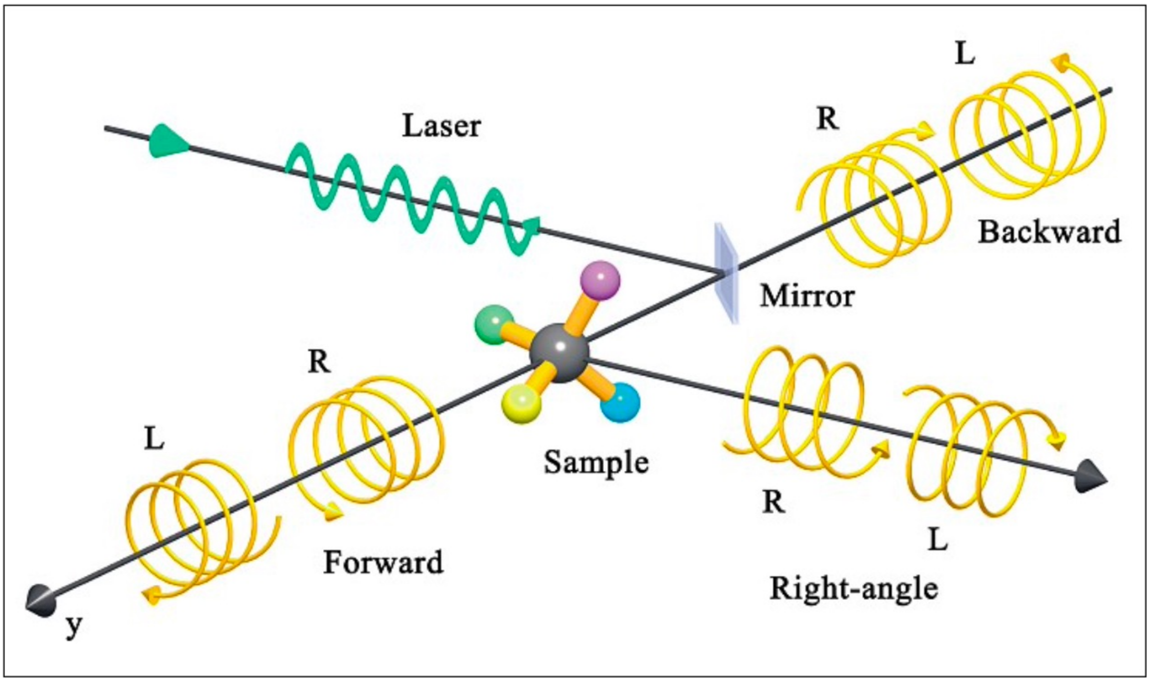


Fig. 1. Backward (π), forward (0), and right angle (π/2) scattered circular polarization (SCP) ROA arrangements (*R*: right circularly, *L*: left circularly polarized light).

Figure 3.4: Sketch of Raman Optical Activity experiment [172].

The Raman and ROA spectra have been simulated at the double–harmonic approximation [166–171]. Their intensities (in $\mathrm{m}^2\,\mathrm{sr}^{-1}$) associated to the p–th vibrational normal mode respectively read [169, 173–175]:

$$\mathrm{d}\sigma_p(\theta, p^s, p^i) = \frac{1}{90}\left(\frac{\mu_0}{4\pi}\right)^2 (\omega_0 - \omega_p)^3 \omega_0 \frac{\hbar}{2\omega_p}[\mathrm{A} \cdot a_p^2 + \mathrm{B} \cdot \beta_p^2]\mathrm{d}\Omega\,; \tag{3.18}$$

$$-\Delta\mathrm{d}\sigma_p(\theta, p^s, p^i) = \frac{1}{90}\left(\frac{\mu_0}{4\pi}\right)^2 (\omega_0 - \omega_p)^3 \omega_0 \frac{\hbar}{2\omega_p}\frac{1}{c}[\mathrm{C} \cdot aG_p^{'} + \mathrm{D} \cdot \beta_{Gp}^2 + \mathrm{E} \cdot \beta_{Ap}^2]\mathrm{d}\Omega\,, \tag{3.19}$$

where c is the speed of light in vacuum, $\omega_0 = 2\pi\nu_0$ is the angular frequency of the laser beam, μ_0 is the vacuum magnetic permeability, and $\hbar$ is the Planck constant divided by 2π.

σ measures the rate at which photons are removed from the incident beam by scattering over 4π steradians, relative to the rate at which they cross a unit area perpendicular to the direction of propagation of the incident beam.

$\mathrm{d}\sigma(\theta)/\mathrm{d}\Omega$ is the differential scattering cross–section, the scattering cross–section measured in a cone of solid angle $\mathrm{d}\Omega$ (in sr) around an angle θ between the incident and scattered lights.

A, B, C, D, E are multiplicative factors that are function of the scattering angle θ and the polarization of the incident (p^i) and scattered (p^s) radiations.

In this work, for the Raman and ROA spectra where $\theta = \pi$, so backward setup, they amount to 90, 14, 0, 48, 16, respectively. To obtain the spectrum at any abscise x (in cm^{-1}), the Lorentzian broadening is used:

$$I^{\mathrm{ROA}}(x) = \sum_{p}^{3N-6} \frac{-\Delta \mathrm{d}\sigma_p}{\mathrm{d}\Omega} \mathcal{L}(x; x_0 = \bar{\nu}_p, \Gamma) \,; \tag{3.20}$$

$$\mathcal{L}(x; x_0, \Gamma) = \frac{1}{\pi\Gamma} \frac{1}{1 + \frac{(x-x_0)^2}{\Gamma^2}} \,, \tag{3.21}$$

with the intensity in $\mathrm{m}^2\,\mathrm{cm}\,\mathrm{sr}^{-1}$ and Γ is the Half–Width at Half–Maximum (HWHM) in cm^{-1}.

a_p^2, β_p^2 in Eq. (3.18) are the two Raman invariants and require the evaluation of the first–order derivatives of the electric dipole — electric dipole polarizability tensor ($\partial\alpha/\partial Q_p$):

$$a_p^2 = \frac{1}{9} \sum_{\mu,\nu}^{x,y,z} \left(\frac{\partial \alpha_{\mu\mu}}{\partial Q_p} \right)_0 \left(\frac{\partial \alpha_{\nu\nu}}{\partial Q_p} \right)_0 \tag{3.22}$$

$$\beta_p^2 = \frac{1}{2} \sum_{\mu,\nu}^{x,y,z} \left[3 \left(\frac{\partial \alpha_{\mu\nu}}{\partial Q_p} \right)_0 \left(\frac{\partial \alpha_{\mu\nu}}{\partial Q_p} \right)_0 - \left(\frac{\partial \alpha_{\mu\mu}}{\partial Q_p} \right)_0 \left(\frac{\partial \alpha_{\nu\nu}}{\partial Q_p} \right)_0 \right] \tag{3.23}$$

In the above summations, the indices μ, and ν, are components of the electric fields (x, y, z). The subscript 0 indicates that the properties are evaluated at the equilibrium geometry.

$aG_p^{'}$, β_{Gp}^2, β_{Ap}^2 in Eq. (3.19) are the three ROA invariants:

$$aG_p^{'} = \frac{1}{9} \sum_{\mu,\nu}^{x,y,z} \left(\frac{\partial \alpha_{\mu\mu}}{\partial Q_p} \right)_0 \left(\frac{\partial G_{\nu\nu}^{'}}{\partial Q_p} \right)_0 \tag{3.24}$$

$$\beta_{Gp}^2 = \frac{1}{2} \sum_{\mu,\nu}^{x,y,z} \left[3 \left(\frac{\partial \alpha_{\mu\nu}}{\partial Q_p} \right)_0 \left(\frac{\partial G_{\mu\nu}^{'}}{\partial Q_p} \right)_0 - \left(\frac{\partial \alpha_{\mu\mu}}{\partial Q_p} \right)_0 \left(\frac{\partial G_{\nu\nu}^{'}}{\partial Q_p} \right)_0 \right] \tag{3.25}$$

$$\beta_{Ap}^2 = \frac{\omega_0}{2} \sum_{\mu,\nu}^{x,y,z} \sum_{\lambda,\kappa}^{x,y,z} \left[3 \left(\frac{\partial \alpha_{\mu\nu}}{\partial Q_p} \right)_0 \left(\frac{\epsilon_{\mu\lambda\kappa} \partial A_{\lambda\kappa\nu}}{\partial Q_p} \right)_0 \right] \tag{3.26}$$

where $\epsilon_{\mu\lambda\kappa}$ is the antisymmetric unit tensor of Levi–Civita. The ROA invariants involve two new first–order derivatives of polarizability tensors: the electric dipole — magnetic dipole ($\partial G'/\partial Q_p$) and the electric dipole — electric quadrupole ($\partial A/\partial Q_p$) tensors [176, 177].

According to Buckingham [176], the perturbed Hamiltonian of a molecule in an electromagnetic field with a pulsation ω can be written as the unperturbed Hamiltonian with the addition of some perturbation terms. These terms are function of the electric field F_σ and the magnetic field H_σ, their respective gradient $F_{\sigma\rho}$ and $H_{\sigma\rho}$, the dipolar electric moment μ_σ, the quadrupolar electric moment $\theta_{\sigma\rho}$, the dipolar magnetic moment m_σ, and the diamagnetic susceptibility $\chi_{\mathrm{d}\sigma\rho}$. The σ, ρ, and η are Cartesian components of a vector or a tensor.

Then, the perturbed Hamiltonian reads:

$$H_{per.} = H_{unper.}^0 - \sum_{\sigma} \mu_\sigma F_\sigma - \frac{1}{3} \sum_{\sigma,\rho} \theta_{\sigma\rho} F_{\sigma\rho} - \sum_{\sigma} m_\sigma H_\sigma - \frac{1}{2} \sum_{\sigma,\rho} \chi_{\mathrm{d}\sigma\rho} H_\sigma H_\rho - \cdots \,, \tag{3.27}$$

where σ and ρ are Cartesian components of a vector or a tensor. The multipoles can be evaluated with the time–dependent perturbation theory and their truncated expressions read:

$$\mu_\sigma = \mu_\sigma^0 + \sum_\rho \alpha_{\sigma\rho}\, F_\rho + \frac{1}{3}\sum_{\rho,\eta} A_{\sigma,\rho\eta}\, F_{\rho\eta} + \sum_\rho G^{'}_{\sigma\rho}\, \dot{H}_\rho\, \omega^{-1} + \cdots\ ; \tag{3.28}$$

$$\theta_{\sigma\rho} = \theta_{\sigma\rho}^0 + \sum_\eta A_{\eta,\sigma\rho}\, F_\eta + \cdots\ ; \tag{3.29}$$

$$m_\sigma = m_\sigma^0 - \sum_\rho G^{'}_{\rho\sigma}\, \dot{F}_\rho\, \omega^{-1} + \cdots\ , \tag{3.30}$$

where $\dot{H}_\rho$ and $\dot{F}_\rho$ are the derivatives of the components σ of the magnetic field and electric field, respectively, with respect to the time. These equations define the three types of polarizability implicated in the ROA intensity: α, A, and $G^{'}$.

Chapter 4. Methodology

4.1 ROA Spectrum Simulation

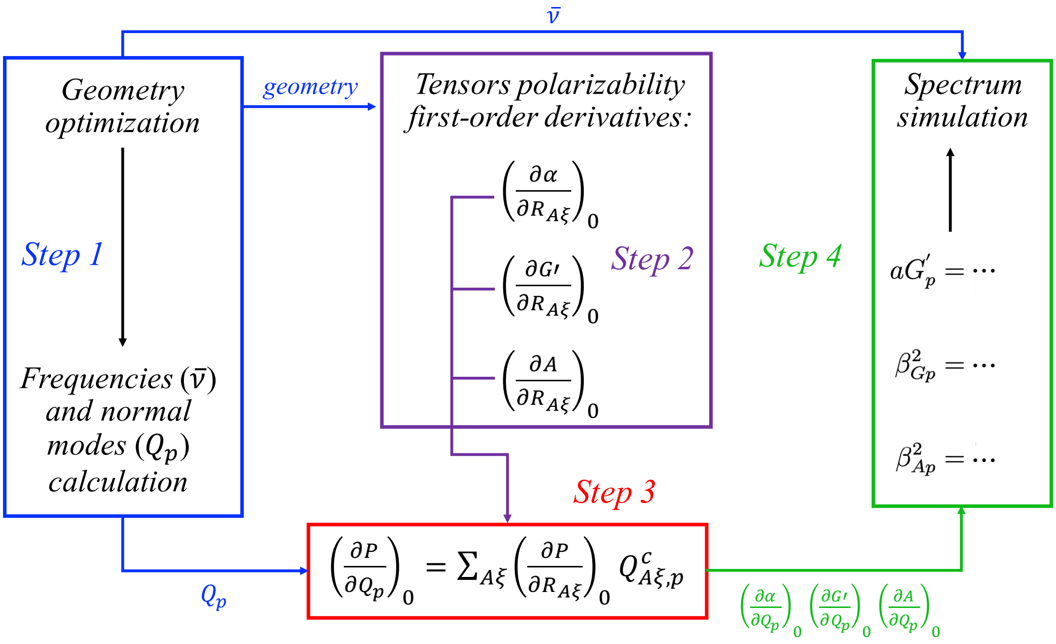


Figure 4.1: Sketch of the different steps to simulate ROA spectra.

The simulation of ROA spectra requires the evaluation of vibrational frequencies and normal modes as well as the calculation of Cartesian derivatives of polarizability tensors. Using Gaussian quantum chemistry program [178], we proceed in multi–steps (Fig. 4.1):

1. Geometry optimization followed by normal modes calculation using a given method.

2. Evaluation of the Cartesian derivatives of polarizability tensors using another method.

3. Conversion of all property derivatives from Cartesian to normal coordinates.

4. Calculation of the ROA invariants and simulation of the spectrum.

The reference method used MPPT at the **MP2** level with the basis set **cc–pVTZ** since it gives reliable geometries and normal modes of vibration. The reference method for the property derivatives calculation used **TDHF** method with the basis set **aug–cc–pVDZ**.

All the spectra have been simulated using DrawSpectrum [179] while DrawMol [180] is used to visualize the molecular structures and DrawVib [181] is used to visualize the normal modes. A visible wavelength of 532 nm is used to simulate our ROA spectra. The Maxwell–Boltzmann $(1/[1-\mathrm{e}^{(-\hbar\omega_p/(k_\mathrm{b}T))}])$ factor with $T = 298.15\,\mathrm{K}$ is used to account for the T–dependence of the populations of the vibrational levels in the spectra.

4.2 Benchmark Methods

4.2.1 First Step

Second step method is kept fixed at **TDHF/aug–cc–pVDZ** level while a list of XC functionals with a range of basis sets is tested for the geometry optimization and the normal modes calculation. XC functionals list: **B3LYP**, **M06**, **M06–2X**, ω**B97X** and ω**B97X–D**. Basis sets list: **6–31G**, **6–31G***, **6–31G****, **6–31++G****, their analog **6–311** ones, **cc–pVDZ**, **cc–pVTZ**, **cc–pVQZ**, **aug–cc–pVDZ**, and **aug–cc–pVTZ**.

4.2.2 Second Step

First step method is kept fixed at **MP2/cc–pVTZ** level while **TDHF** method, with the same list of basis sets, is tested for the evaluation of the Cartesian derivatives of polarizability tensors. The smaller **rDPS:3–21G** basis set, introduced by Zuber and Hug [182] especially for ROA intensity calculations, has also been tested.

4.3 Xenon Description in Calculations

In order to reduce calculation time and to take into account some relativistic effects induced by an heavy atom in our system, here xenon, the usage of an ECP is suggested. The search of ECPs available for xenon has been easily accomplished thanks to the site basissetexchange.org [183–185].

Table 4.1: Basis sets using the ECP **dhf** and **def2** for the description of xenon in the system.

ECP	Basis set	Nb of Atomic Orbitals
dhf	(aug–)cc–pVDZ–PP [186]	38 → 41
	(aug–)cc–pVTZ–PP [186]	57 → 61
	(aug–)cc–pVQZ–PP [186]	66 → 81
	(aug–)cc–pV5Z–PP [186]	91 → 97
	(aug–)cc–pwCVDZ–PP [186]	42 → 45
	(aug–)cc–pwCVTZ–PP [186]	66 → 70
	(aug–)cc–pwCVQZ–PP [186]	81 → 86
	(aug–)cc–pwCV5Z–PP [186]	86 → 92
	dhf–QZVPP [187]	31
	dhf–QZVPP [187]	47
def2	def2–TZVPP [188]	31
	def2–TZVPPD [188, 189]	34
	def2–QZVPP [188]	47
	def2–QZVPPD [188, 189]	50

Table 4.2: List of the basis sets using other ECPs for the description of xenon in the system.

ECP	Basis set	Nb of Atomic Orbitals
CRENBL	CRENBL	10
CRENBS	CRENBS	10
LANL2DZ	LANL2DZ	6
	LANL08 [190]	6
SBKJC	SBKJC–Polar.–LFK [191]	9
	SBKJC–VDZ	5
Stuttgart–RLC	Stuttgart–RLC	16

4.4 Conformational Search Method

For a flexible molecule, we need to explore the potential energy surface to obtain the different stable conformers. Each one is characterized by its Boltzmann weight. The overall spectrum is thus obtained as a Boltzmann averaged of the individual spectra. We proceed in multi–steps (Fig. 4.2):

1. Generation of the conformers with conformational search programs (GMMX and CREST).
2. Geometry optimization followed by normal modes calculation on the generated conformers using the appropriate method determined in Section 4.2.
3. Remove duplicates, based on the RMSDs for the structures and the energies.
4. Evaluation of the Cartesian derivatives of polarizability tensors for the non–redundant conformers using the appropriate method in Section 4.2.
5. Calculation of the populations for the non–redundant conformers, given by the Boltzmann weights based on the relative Gibbs free energy.
6. Simulation of the final spectrum as a Boltzmann averaged.

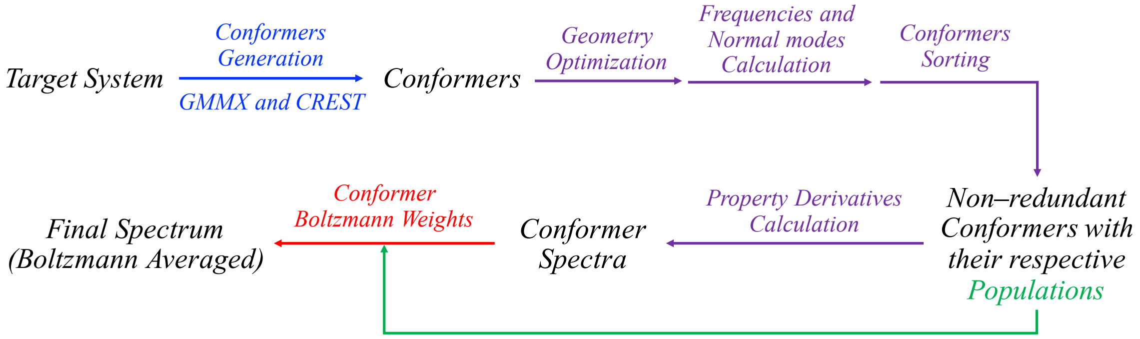


Figure 4.2: Sketch of the different steps to perform the conformational search on a system and to simulate the final ROA spectrum. Some steps are explained in Sections 4.1 and 4.2.

4.5 Interpretation Tools

In order to rationalize the relationship between the normal modes obtained with various methods, we define the overlap matrix as the square of the scalar product between two normal modes expressed in the mass–weighted coordinates. The overlap between the normal mode ℓ of method A and the normal mode p of method B reads:

$$O_{\ell,p}^{\mathrm{A,B}} = \left\langle Q_\ell^{\mathrm{A}} \middle| Q_p^{\mathrm{B}} \right\rangle^2 = \left(\sum_{i\alpha}^{3N} Q_{i\alpha,\ell}^{\mathrm{A}} Q_{i\alpha,p}^{\mathrm{B}} \right)^2 \tag{4.1}$$

To analyze even further this matrix, we introduce the concept of identical and non–equivalent modes. A normal mode is said to be identical between two methods when its overlap value is greater than, or equal to, a threshold value set at 0.95 throughout this work.

Since two identical modes of methods A and B involved similar atomic displacements, on can assume that their vibrational frequencies are related by a linear relationship:

$$\bar{\nu}^{\mathrm{A}} = k \bar{\nu}^{\mathrm{B}} \tag{4.2}$$

where k is the slope of the linear regression between the vibrational frequencies, fitted using a least–square approach. The linear regression is performed only for the fingerprint signatures which range from 600 to 1800 cm^{-1}.

To assess the agreement between both sets of vibrational frequencies (for identical modes), one can define the RMSD between the frequencies of method A (taken as reference) and the corrected frequencies of method B:

$$\mathrm{RMSD} = \sqrt{\frac{\sum_i^{N_{\text{ind modes}}} \left(k\bar{\nu}_i^{\mathrm{B}} - \bar{\nu}_i^{\mathrm{A}} \right)^2}{N}} \tag{4.3}$$

At last, to quantify the difference between the Raman and ROA spectra obtained with two different methods, the “S” quantity [192, 193] is calculated as the ratio between the (intensity) overlap of these spectra and the square root of their self–overlap:

$$S_{\mathrm{A,B}} = \frac{\langle I_{\mathrm{A}} | I_{\mathrm{B}} \rangle}{\sqrt{\langle I_{\mathrm{A}} | I_{\mathrm{A}} \rangle \langle I_{\mathrm{B}} | I_{\mathrm{B}} \rangle}} \tag{4.4}$$

The overlap is done in the frequency region of interest (600 — 1800 cm^{-1}). For Raman, the value of this quantity amounts to 0 or 1, when the spectra are different or identical, respectively. For ROA, the value ranges from -1 to 1 where -1 means a perfect overlap between the spectra of two enantiomers.

Chapter 5. Results and Analysis

5.1 R–methyloxirane

5.1.1 Comparison to Experiment

The reference method uses the (**MP2/cc–pVTZ**) method for the geometry optimization and normal modes calculation. For the property derivatives calculation, the reference method uses the (**TDHF/aug–cc–pVDZ**) method. It leads to **MP2/cc–pVTZ//TDHF/aug–cc–pVDZ**. This method gives reliable geometries and normal modes of vibration but is too expensive so the study of larger systems is compromised. Then, at a lower theory level (DFT), some XC functionals are tested and compared to the reference method.

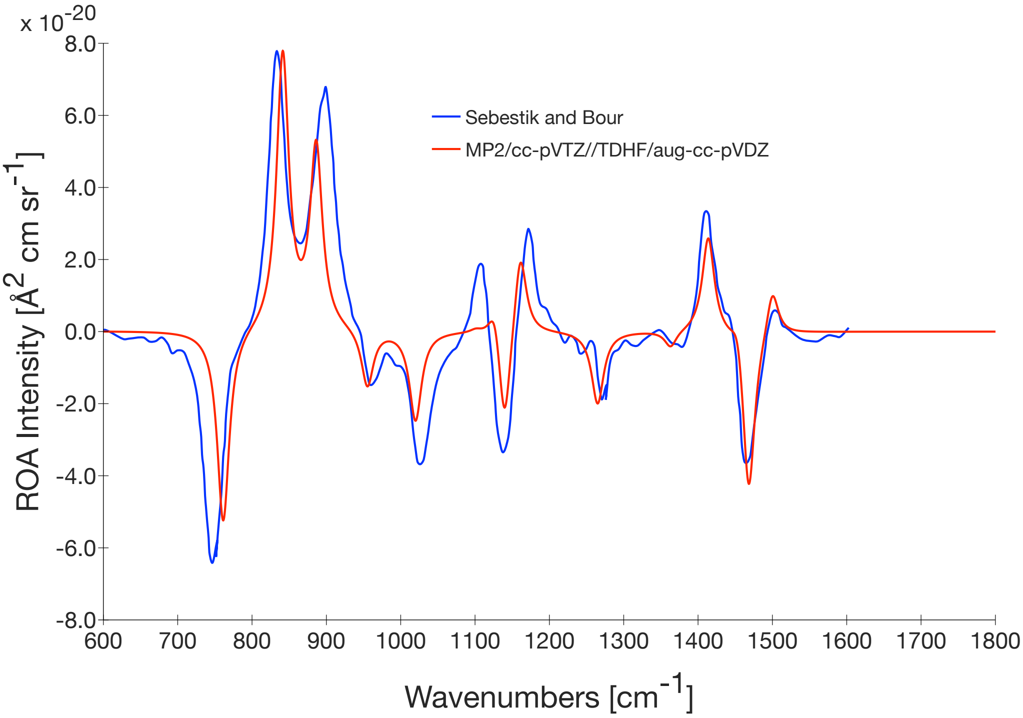


Figure 5.1: Experimental in neat liquid and simulated (**MP2/cc–pVTZ//TDHF/aug–cc–pVDZ** scaled by **0.97**, FWHM of 20 cm^{-1}) ROA spectra of **R**–methyloxirane.

First, the reference method is preliminary compared with an experimental measurement by Šebestík and Bouř [194] (see Fig. 5.1). Our reference calculation reproduces their experimental ROA spectrum provided that the calculated vibrational frequencies are scaled by a factor of **0.97**. This is consistent with the tabulated factor for **MP2** as obtained by Scott and Radom [195, 196]. The only discrepancy appears for the peak at around 1100 cm^{-1}. We can thus rely on this reference simulation method to compare the impact of the choice of the XC functional and basis set on Steps 1 and 2 (Fig. 4.1).

5.1.2 Impact of the method on Step 1

In order to rationalize the changes in the normal modes between different methods, the overlap matrix and the RMSD between the frequencies of identical modes are calculated thanks to Eqs. (4.1) and (4.3). Normal modes are considered identical with respect to our reference if their overlap value is superior to 0.95 (see Section 4.5). Their number for each XC functional/basis set is represented in Fig. 5.2.

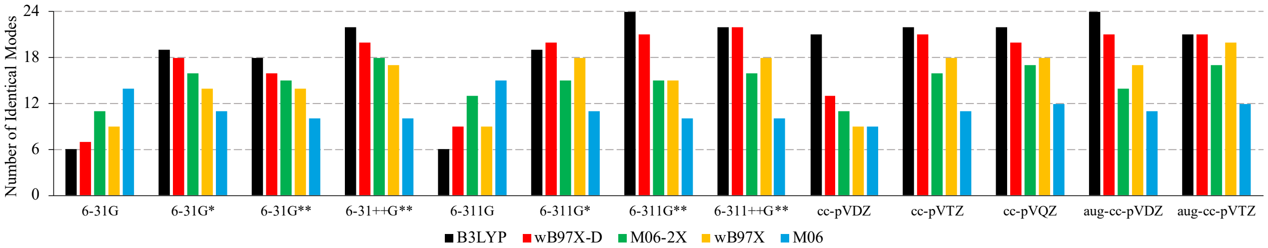


Figure 5.2: Number of identical modes for each **XC functional/basis sets** with respect to the reference **MP2/cc–pVTZ**.

We observe that **6–31G** and **6–311G** basis sets have the worst numbers. As there are no polarization and diffuse functions, these results were expected. The two best XC functionals are: first, **B3LYP**, and second, **ωB97X–D**. These methods have the biggest numbers of identical modes for most basis sets so they seem to properly calculate the normal modes of vibration.

Next, a linear regression in the fingerprint signature region (600 — 1800 cm^{-1}) between their vibrational frequencies and the reference ones, only for the identical modes, is performed. Then, their RMSD (between the frequencies), after applying the slope of the linear regression, is computed. All these information for **B3LYP** are presented in Table 5.1. The other XC functional information, Tables A.1 to A.4, are presented in Appendix A. No linear regression is calculated if there are less than 3 identical modes in the 600 — 1800 cm^{-1} region.

Small number of identical modes and large RMSD values were expected for **6–31G** and **6–311G** basis sets. It means that these basis sets are not appropriate for vibrational simulations. Going from **6–31(1)G** → **6–31(1)++G**** improves the agreement with the reference method (thus the number of identical modes increases and the RMSD value decreases). In fact, adding polarization and diffuse functions enhances the description of the system. For the **cc–pVXZ** basis sets, their number of identical modes is close to the maximum (**24**) and their RMSD are small and similar. These basis sets seem appropriate for vibrational simulations.

Finally, the impact of the method on ROA spectra is assessed with the overlap spectra values. The preliminary step is to "bring" the spectrum of a given method over the reference spectrum thanks to the slope of the linear regression. The vibrational frequencies are scaled by the slope, $k_{\textbf{Method/Ref}}$. However, the reference spectrum is already scaled by a factor of **0.97** thus the total scaling factor for the vibrational frequencies of a given method is the product of the slope by **0.97**. All the overlap spectra values for the different methods are given in Table 5.2.

Table 5.1: Number of identical modes, slope of the linear regression (LR), and RMSD for the vibrational frequencies between **B3LYP/basis sets** and the reference **MP2/cc–pVTZ**.

Basis set	Identical Modes	Modes in LR	Slope in LR	RMSD (cm^{-1})
6–31G	6	3	0.989579	20.1
6–31G*	19	11	0.994713	8.8
6–31G**	18	11	1.001723	6.0
6–31++G**	22	15	1.008709	6.1
6–311G	6	3	0.993686	20.4
6–311G*	19	12	1.002895	8.5
6–311G**	24	15	1.010103	5.7
6–311++G**	22	13	1.011361	7.1
cc–pVDZ	21	14	1.021249	7.0
cc–pVTZ	22	13	1.008915	6.4
cc–pVQZ	22	13	1.008784	6.9
aug–cc–pVDZ	24	15	1.021684	6.5
aug–cc–pVTZ	21	14	1.009986	7.4

Table 5.2: Overlap values between **XC functionals/basis sets** and the reference **MP2/cc–pVTZ**, with the intensities calculated at the **TDHF/aug–cc–pVDZ** level, in the 600—1800 cm^{-1} region. Bold values indicates overlaps greater than 0.700.

Basis set	B3LYP	M06	M06–2X	ωB97X	ωB97X–D
6–31G	0.017	0.507	0.353	0.306	0.263
6–31G*	0.557	0.406	0.459	0.454	0.605
6–31G**	**0.718**	0.360	0.479	0.413	0.538
6–31++G**	0.683	0.343	0.504	0.411	0.611
6–311G	0.000	0.509	0.383	0.323	0.241
6–311G*	0.585	0.418	0.542	0.467	**0.723**
6–311G**	**0.713**	0.354	0.499	0.394	0.569
6–311++G**	0.641	0.369	0.515	0.438	0.636
cc–pVDZ	**0.769**	-0.015	0.309	0.337	0.416
cc–pVTZ	0.682	0.356	0.485	0.400	0.562
cc–pVQZ	0.655	0.365	0.480	0.389	0.550
aug–cc–pVDZ	0.635	0.339	0.387	0.382	0.567
aug–cc–pVTZ	0.646	0.374	0.496	0.434	0.589

First, we observe that the largest values (in bold) are close to 0.800, which is still lower than the maximum value of 1.000. This can be explained by the fact that slight shifts in the position of the peaks decrease drastically the overlap spectra values even if the spectra have similar shapes.

Thus, it is difficult to judge the quality of the method solely based on these values and the spectra themselves have to be inspected. Second, **B3LYP** and **ωB97X–D** have the overall largest values so they best reproduce the reference spectrum. For example, some simulated B3LYP spectra are compared to the reference spectrum in Fig. 5.3. Their shapes are very similar thus the basis sets are quite equivalent but their overlap spectra values vary and amount to: **0.557**, **0.585**, **0.713**, **0.769**, and 0.682. A huge change in the overlap spectra value has been observed between **6–311G*** and **6–311G****, while only slight shifts are observed. Here, the spectra can be partitioned into different regions. In these spectra, the first three peaks (at low wavenumbers) and the peak around 1470 cm^{-1} have bigger impact on the overlap spectra values than the other peaks. Indeed, these peaks are the most intense ones and thus the global overlap is very sensitive to their relative position with respect to the reference peaks.

The two positive peaks at 840 and 890 cm^{-1}, of the **cc–pVDZ** spectrum, are close to each other as those of the reference ones. So its overlap spectra value is huge (**0.769**). If the peak at 1455 cm^{-1} was shifted up to 1470 cm^{-1}, the overlap spectra value would have been better. In the cc–pVTZ spectrum, the agreement with the reference spectrum is nearly perfect expect for two peaks: the one at 827 cm^{-1} which should be closer to 840 cm^{-1}; the other small negative peak at 1133 cm^{-1} which should be slightly positive. Its overlap spectra value (0.682) is hugely impacted by these two changes. Since **B3LYP** and **ωB97X–D** are our best methods, dispersion corrections (D2, D3, and D3BJ) were added to **B3LYP/cc–pVDZ** to test their influence on the ROA spectra. No huge variations, of the peaks position or intensity, have been observed. However, we keep in mind that, later, dispersion could be important for our cryptophanes.

For the step 1, the conclusions are: **B3LYP** and **ωB97X–D** are the best XC functional to simulate ROA spectra of **R**–methyloxirane; the overlap spectra values are too sensitive to the peaks position so we cannot rely solely on these numbers for a proof of quality.

5.1.3 Impact of the method on Step 2

Afterwards, the impact of the basis set on the property derivatives is discussed. Here, the position of the peaks is the same for all the methods and only their intensities vary so the overlap spectra values are a proof of quality. The different values are showed in Fig. 5.4. We observe that, except for **STO–3G**, all the values were huge (superior to 0.900). Adding diffuse functions increases the overlap spectra values. Indeed, diffuse functions are important to evaluate the polarizability. **rDPS:3–21G** basis set, introduced by Zuber and Hug [182], gives good results and performs as good as larger basis sets (**aug–cc–pVDZ** and **aug–cc–pVTZ**). **rDPS:3-21G** is small and contains tuned (for ROA calculation) diffuse functions so the calculation time is reduced with respect to larger basis sets.

For the step 2, the conclusions are: **rDPS:3-21G** performance is as good as larger basis sets for a reduced cpu time; the diffuse functions are important to evaluate the polarizability and the Cartesian derivatives of polarizability tensors; the overlap spectra values are a proof of quality since there is no shift of the peaks position.

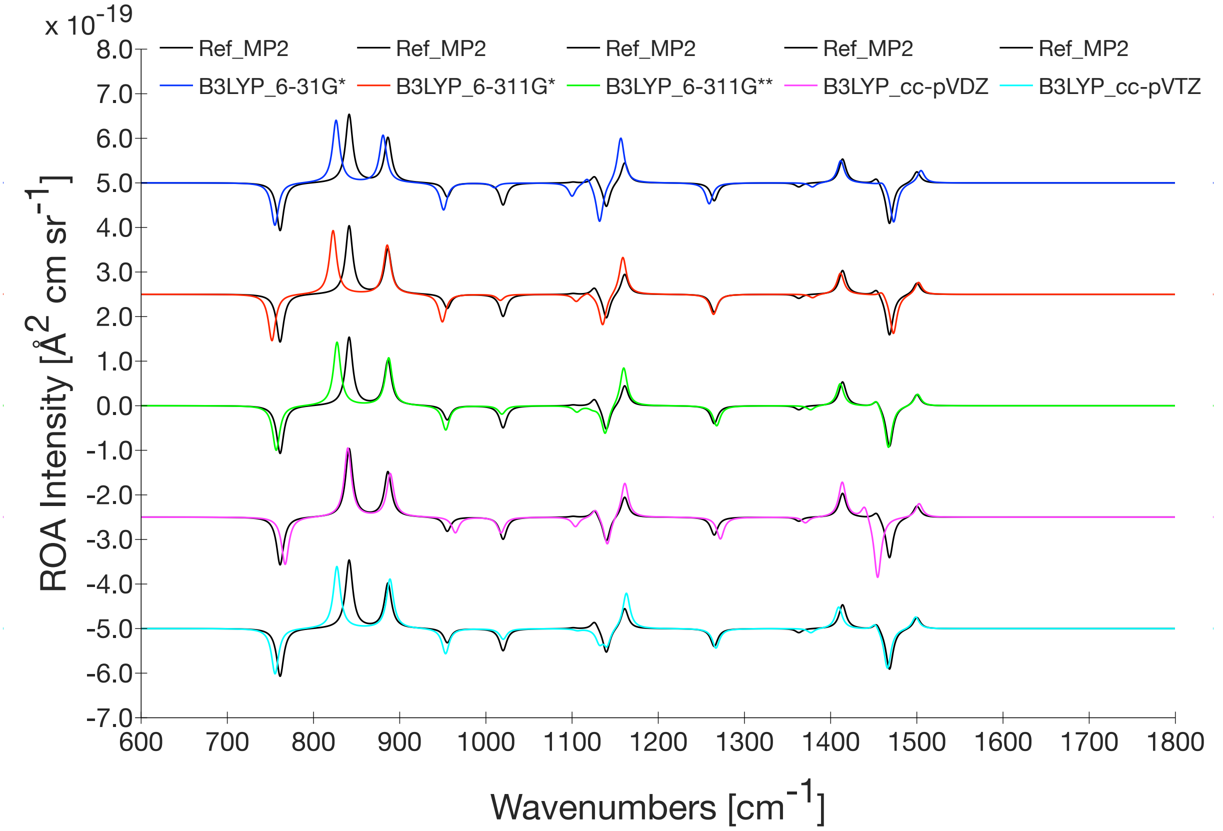


Figure 5.3: Simulated ROA spectra using **B3LYP/basis sets** and the reference **MP2/cc–pVTZ**, with the intensities calculated at the **TDHF/aug–cc–pVDZ** level, in the 600—1800 cm^{-1} region.

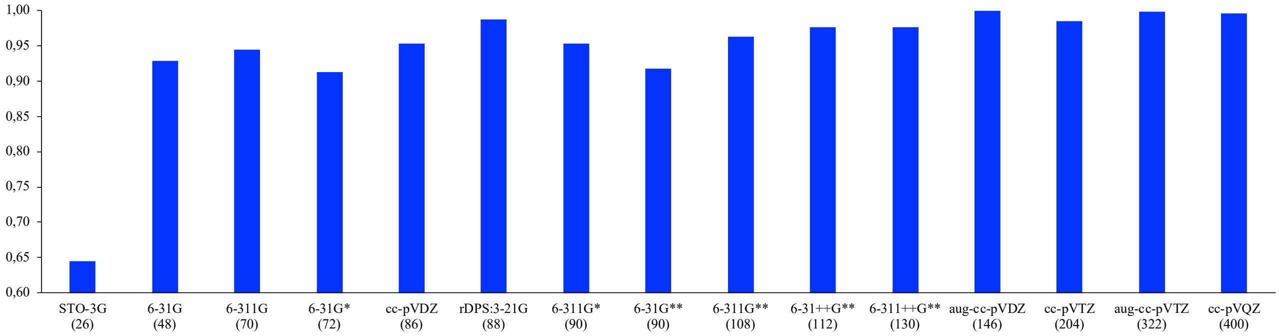


Figure 5.4: Overlap values between **TDHF/basis sets** and the reference **TDHF/aug–cc-pVTZ**, with the geometry optimized and the normal modes calculated at the **MP2/cc–pVTZ** level, in the 600—1800 cm^{-1} region. The basis sets have been ordered by their number of basis functions.

5.1.4 Investigation of ECPs for Xenon

At last, the search of a good ECP with its corresponding basis sets, for the xenon, is performed on XeF_2, XeF_4, and XeF_6 systems. They have linear, tetrahedral, and octahedral structure, respectively. Their reference Xe–F bond lengths [197–200] and frequencies [201] are compared to theoretical values in the Tables B.1 to B.6 in Appendix B. The method has to be determined with the biggest system so that the method can describe complex systems. Unfortunately, the XeF_6 reference values are given for a tetrameric agglomerate [200] thus they should be taken with care.

A simple difference ($|\Delta_i|$), with respect to the reference values, has been calculated for the Xe–F bond lengths and the frequencies. Since two normal modes are compared, a RMSD has been calculated in order to better compare the global deviation of the methods. The Tables B.1 to B.3 show small bond length deviations. Thus, we focus our interpretations on the frequencies (see Tables B.4 to B.6), which are more sensitive to the calculation method. The 8 valence electrons ECPs (**CRENBS**, **LANL2DZ**, **SBKJC**, and **Stuttgart–RLC**) have the largest deviations. It demonstrates the importance to not only take into account the last valence electrons. Based on the XeF_2 and XeF_4 systems, the results show that the triple–ζ basis sets (**(aug–)cc–pVTZ–PP**, **(aug–)cc–pwCVTZ–PP**, **def2–TZVPP**, **dhf–TZVPP**) give the smallest deviations.

The conclusions are: the best basis set is **aug–cc–pwCVTZ–PP**, which uses the ECP **dhf**; triple–ζ basis sets are necessary to properly describe the XeF_2 and XeF_4 systems.

5.2 Cryptophane–PP–111

In Section 5.1, the appropriate method for the ROA spectrum simulation of our target system, Cryptophane–**PP**–111 (Fig. 5.5), has been successfully determined: **B3LYP** and ω**B97X–D** for the geometry optimization and the normal modes calculation; **TDHF/rDPS:3–21G** for the property derivatives calculation.

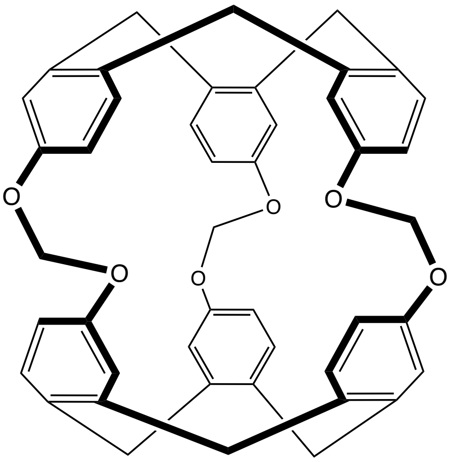


Figure 5.5: Sketch of Cryptophane–**PP**–111.

The two hemispheres of the Cryptophane–**PP**–111 are quite rigid and bound by three flexible linkers. Thus, a general conformational search has to be perform on this system (Section 4.4).

5.2.1 Conformational search

Two different programs have been used to generate the conformers: GMMX and CREST. Next, all the obtained conformers have been reoptimized using **B3LYP** and ω**B97X–D** XC functionals with the basis set **6–31G***. This basis set has been selected since the description of the normal modes is acceptable for a reduced time calculation. The relative energies of the different conformers (numbering from **1** to **16**) have been calculated, with respect to **1**, as well as their respective Boltzmann weight, based on the Gibbs free energy. These information are given in Table 5.3.

Table 5.3: Relative energies (in kcal/mol) with respect to **1** for the reoptimized conformers. The Boltzmann weights, based on the Gibbs free energy, are reported in parentheses. The values larger than 1.0% are highlighted in bold.

Numbering	GMMX	B3LYP	ωB97X–D
1	0.00	**0.00 (99.7%)**	**0.00 (13.6%)**
2	3.22	4.11 (0.3%)	4.75
3	3.34	6.56	7.52
4	4.18	8.24	3.97 (0.2%)
5	5.37	8.06	9.27
6	5.53	10.74	10.34
7	5.59	9.47	8.30
8	5.74	10.06	11.57
9	5.86	8.32	4.94 (0.3%)
10	6.67	7.82	**1.17 (16.1%)**
11	6.70	8.09	9.41
12	6.73	8.39	**-1.21 (69.6%)**
13	7.55	10.78	4.96 (0.3%)
14	7.56	10.78	11.13
15	7.63	10.79	11.14
16	7.66	10.78	11.14

We observe that only **1** is important for **B3LYP** contrary to ω**B97X–D** where **1**, **10**, and **12** have Boltzmann weight values larger than 1.0%. For such big systems with few aromatic rings, dispersion corrections might be necessary to properly describe the π–π interactions in our cryptophanes. Since **12** becomes more stable than **1** when we add dispersion corrections, the two structures have been compared to each other (Fig. 5.6). Their difference is a slight 20° rotation of an hemisphere with respect to the other one. The O atoms of **1** are nearly on top of each other contrary to **12** where they are more distant. In order to compare their cavity, a distance between the centroids of the 18 aromatic C atoms of each hemisphere has been calculated. The results are: 5.19 Å for **1** and 4.28 Å for **12**. There are steric effects for **1** and its hemispheres are further away.

Then, different dispersion corrections schemes (D2, D3, and D3BJ) have been tested on top of **B3LYP**. We notice that upon adding dispersion corrections, the conformer **12** becomes more stable than **1**, similarly to ω**B97X–D** results ((Table 5.4)). Moreover, other methods have been tested (**M06**, **M06–2X**, and **MP2**) and the relative energies are presented in Table 5.4. The most important value is the one of **MP2**, since MPPT is an higher level of theory than DFT, in which the stabilizing interactions are closer to **B3LYP** than the other methods. In conclusion, we only consider **1** obtained by **B3LYP** level for the simulation of the spectra. We will show later that **1** reproduces the experimental spectrum contrary to the spectrum of **12**.

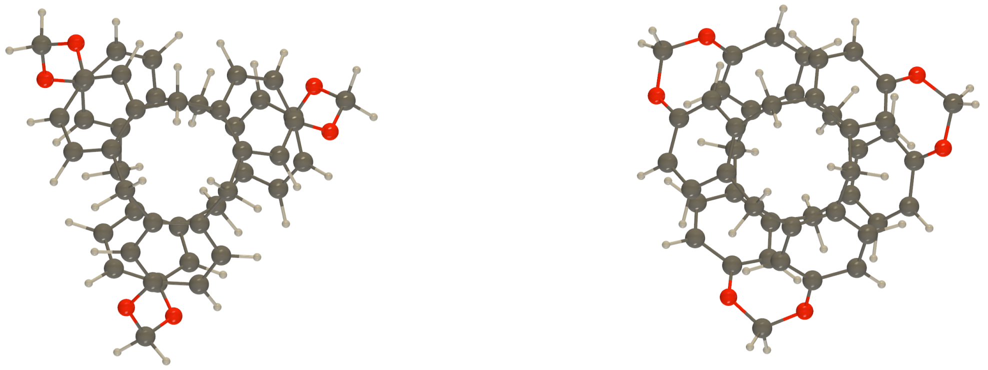


Figure 5.6: Structures of **1** (left) and **12** (right).

Table 5.4: Relative energies (in kcal/mol) of **12** with respect to **1** for the reoptimized conformers using **B3LYP** with different dispersion corrections (top table) and using different methods (bottom table).

Numbering	B3LYP	D2	D3	D3BJ
1	0.00	0.00	0.00	0.00
12	8.39	-1.93	-1.77	-1.97

Numbering	B3LYP	M06	M06–2X	ωB97X-D	MP2
1	0.00	0.00	0.00	0.00	0.00
12	8.39	0.72	1.90	-1.12	**13.53**

Even if only **1** is taken into account, RMSDs on all the structures have been calculated to eliminate the eventual duplicates. We observe that **13**, **14**, **15**, and **16** were duplicates. These results demonstrate the necessity to treat/analyze the generated reoptimized conformers.

Finally, the Cryptophane–**PP**–111 has a **D3** symmetry (Fig. 5.7) and its 255 normal modes are distributed into different irreducible representations: 43 **A1**; 42 **A2**; and 85 **E** (double–degenerated). With the xenon, there are 258 normal modes (+**A2** and +**E**).

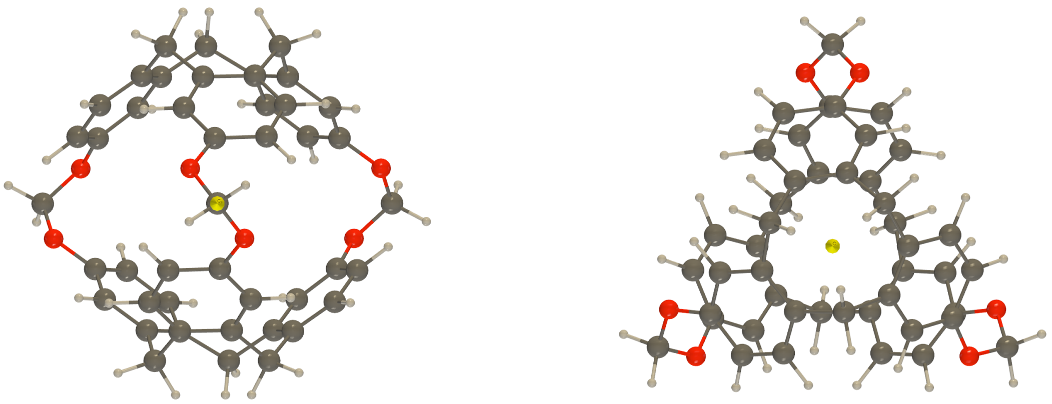


Figure 5.7: Sketches of the symmetry operations for a **D3** point group: $3C_2$ (left) and $2C_3$ (right).

5.2.2 Environment effects

In order to properly simulate realistic spectra, some environment effects have to be taken into consideration. For out target system, Cryptophane–**PP**–111, these environment effects are: the solvent effects and the xenon encapsulation. The description of the xenon atom is going to be performed with **aug–cc–pwCVTZ–PP** which uses the ECP **dhf**.

5.2.2.1 Solvent effects

The solvent effects are necessary to properly describe some of the normal modes of our system. The solute–solvent interactions, e.g. non–covalent interactions, electrostatic interactions, or hydrogen bonds, affect the molecular structures, so the normal modes are impacted. Here, the solvent is going to be added implicitly (with IEF–PCM approach) and thus the impact of hydrogen bonds cannot be investigated.

We focused on the chloromethane derivatives (CH_2Cl_2, $CHCl_3$, and CCl_4) since these molecules are hugely used as organic solvents. The spectra are represented in Figs. 5.8 and 5.9. The results show no huge changes in the global spectrum except in the low frequency region (Fig. 5.9). We observe a systematic decrease of the absolute intensity for the peak around 125 cm^{-1}. The implicit solvent parameters are the static and the dynamic/optical dielectric constants. In fact, the later one increases systematically (2.028346 for CH_2Cl_2, 2.090627 for $CHCl_3$, and 2.131892 for CCl_4) and the intensity of the peak might be influenced by this parameter.

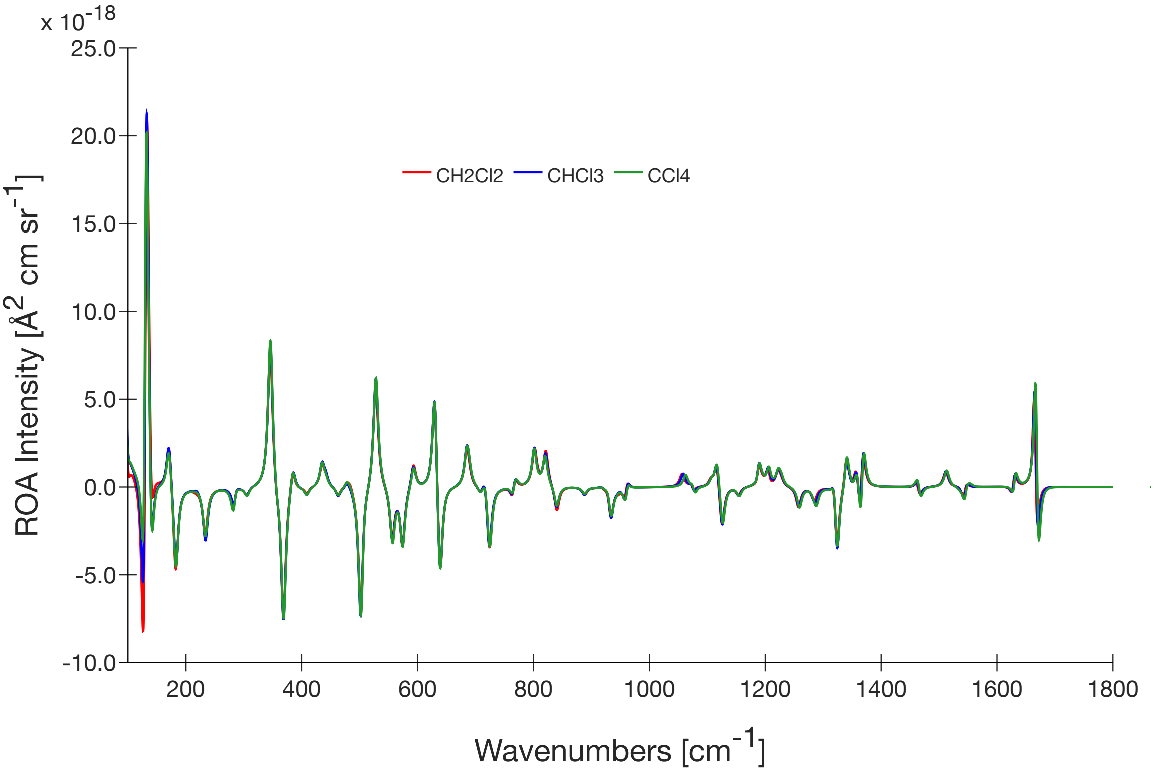


Figure 5.8: Simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G//IEF–PCM/solvents** ROA spectra of **1**, in the 100—1800 cm^{-1} region.

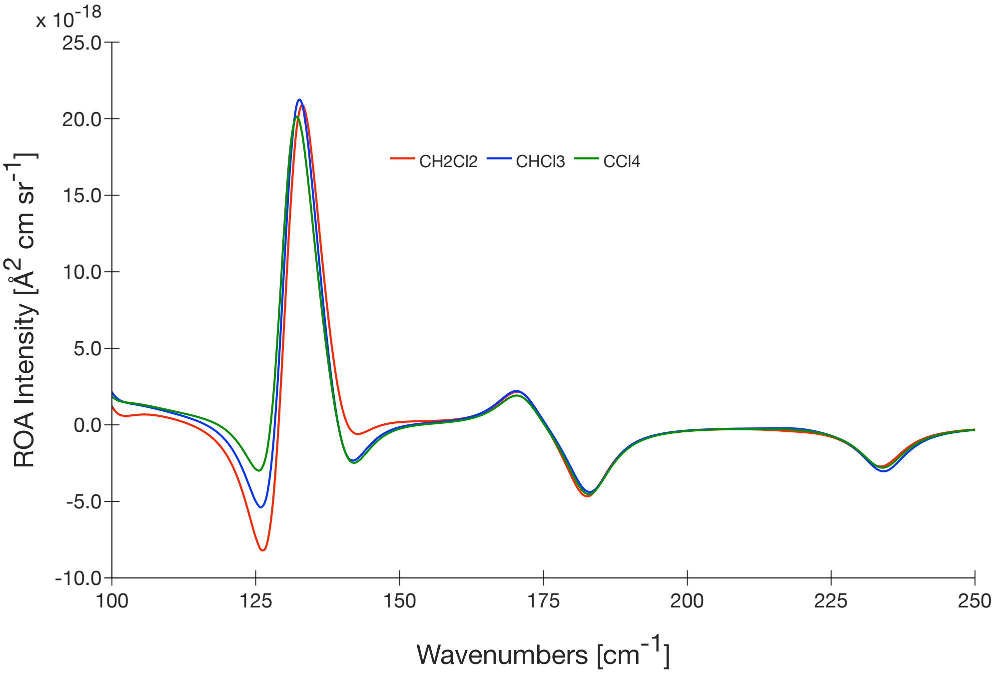


Figure 5.9: Simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G//IEF–PCM/solvents** ROA spectra of **1**, in the 100—250 cm^{-1} region.

Table 5.5: Relative energies (in kcal/mol) with respect to **1** for the reoptimized conformers using B3LYP/6–31G* with/without CH_2Cl_2. The Boltzmann weights, based on the Gibbs free energy, are reported in parentheses. The values larger than 1.0% are highlighted in bold.

Numbering	Vacuo	IEF–PCM	SMD
1	**0.00 (99.7%)**	**0.00 (99.4%)**	**0.00 (98.5%)**
2	4.11 (0.3%)	3.61 (0.4%)	3.56 (0.9%)
3	6.56	5.53 (0.1%)	5.19 (0.1%)
4	8.24	7.33	7.40
5	8.06	6.97	6.63
6	10.74	8.93	7.73
7	9.47	8.65	8.53
8	10.06	8.25	7.73
9	8.32	7.54	7.42
10	7.82	8.00	8.49
11	8.09	7.08	6.86
12	8.39	9.33	10.76
13	10.78	8.74	8.20
14	10.78	8.78	8.09
15	10.79	8.73	8.21
16	10.78	8.72	8.07

Moreover, the solvent effects have been computed with two approaches: IEF–PCM and SMD. The SMD approach is a special case of the IEF–PCM formalism where intrinsic Coulomb radii are used to describe the vdW surface of the cavity model instead of the vdW radii. This approach has been constructed to calculate more accurately the Gibbs free energy of solvation ΔG_s^0. The relative energies of the different conformers, with respect to **1**, and the Boltzmann weights, based on the Gibbs free energies, are represented in Table 5.5. The results show that only **1** is important in both approaches. We observe a general relative stabilization of the system, going from gas phase to IEF–PCM approach to SMD approach, for all the conformers, except for **10** and **12**. It has been determined earlier that **10** and **12** have structures in which their hemispheres are closer to each other. The solute–solvent interactions might be disfavorable for these structures.

The spectra of **1** obtained with both approaches are compared to each other in Figs. 5.10 and 5.11. There is no huge changes in the global spectrum. The impact is the strongest in the low frequency region (Fig. 5.9). For IEF–PCM, there is a negative peak around 125 cm^{-1}. For SMD, there is a negative peak around 140 cm^{-1}.

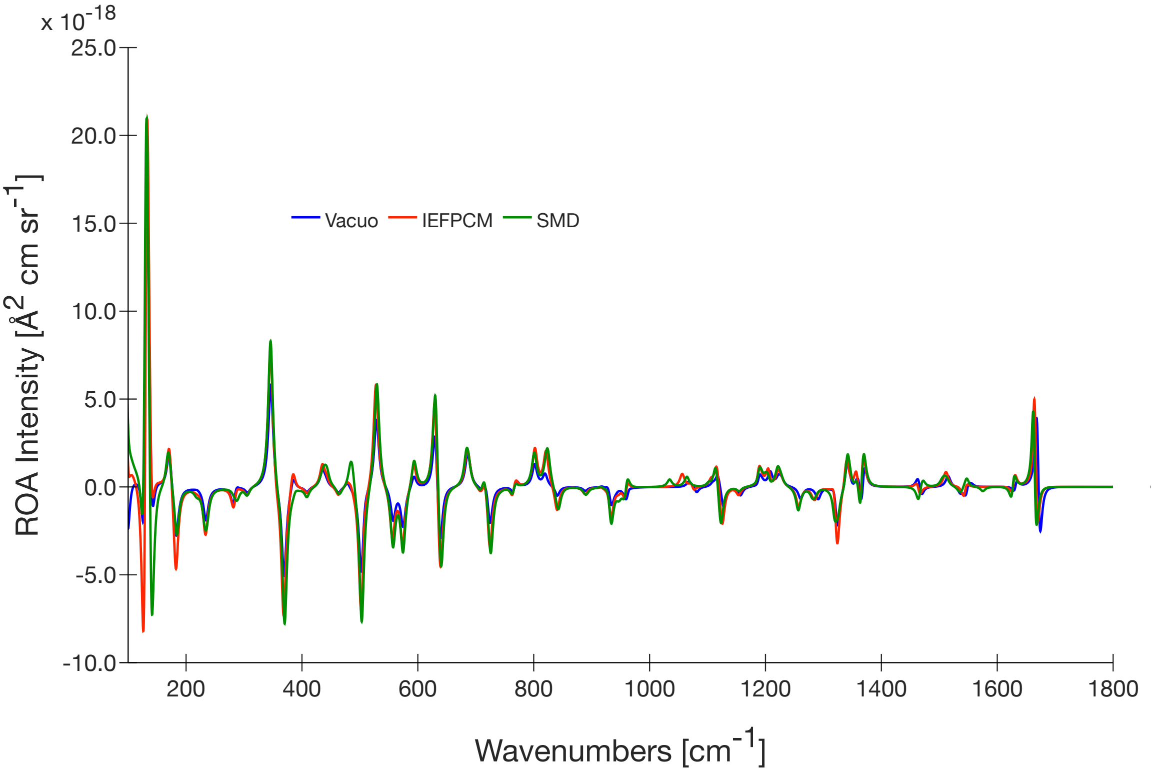


Figure 5.10: Simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G** ROA spectra of **1** with/without solvent (CH_2Cl_2), in the 100—1800 cm^{-1} region.

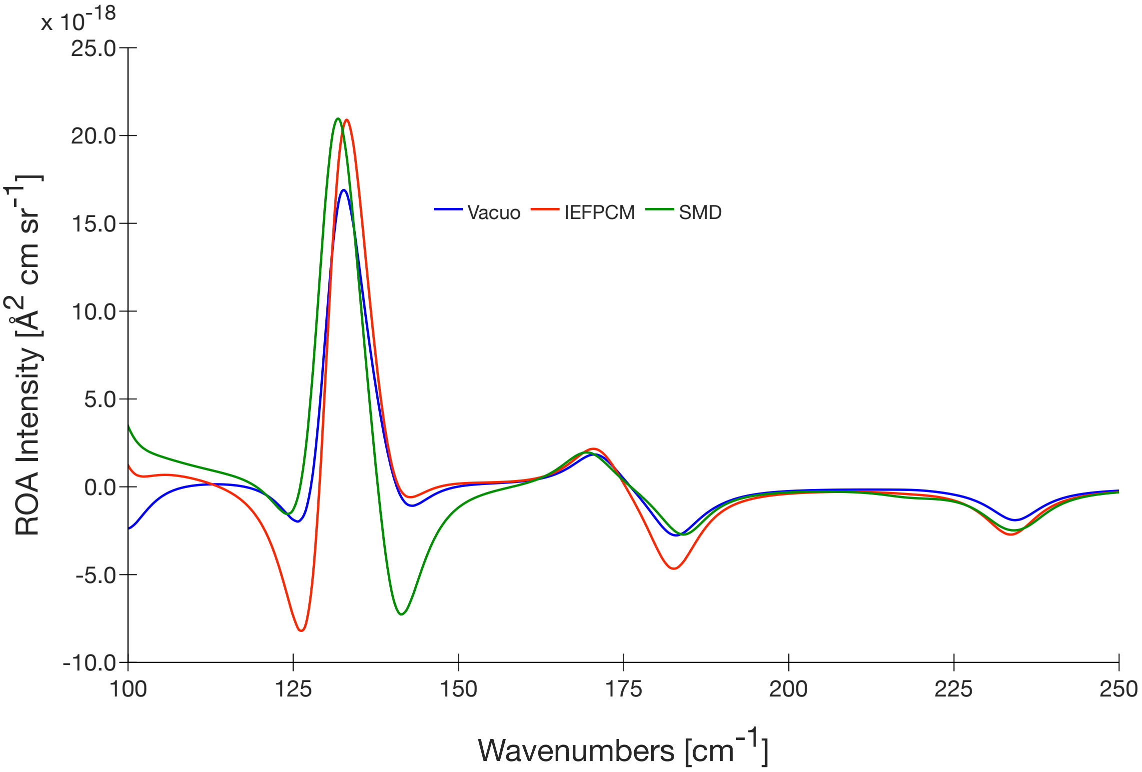


Figure 5.11: Simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G** ROA spectra of **1** with/without solvent (CH_2Cl_2), in the 100—250 cm^{-1} region.

5.2.2.2 Xenon Encapsulation

The xenon has been chosen since Cryptophane–111 exhibited the highest binding constant with xenon encapsulation, (around 10 000 M^{-1} at 293 K). Moreover, xenon is hugely polarizable and is heavily used in medical applications, e.g. in MRI (magnetic resonance imaging), because xenon has the ability to enhance the signals.

The ROA spectrum with a xenon atom encapsulated in the cavity has been computed in gas phase in Figs. 5.12 and 5.13. There is no huge changes in the global spectrum with the xenon encapsulation except in the low frequency region (Fig. 5.13): the absolute intensity of the peak around 142 cm^{-1} hugely increases; the peak around 126 cm^{-1} disappears; and the absolute intensity of the peak around 132 cm^{-1} slightly decreases. Thus, the normal modes involved in this region are analyzed and represented in Fig. 5.14. They are denoted with "**Xe**" if the xenon is encapsulated. The modes **10–11**, **14–15**, **14Xe–15Xe**, and **17Xe–18Xe** are **E** and degenerated. The modes **12** and **13Xe** are **A1**. The modes **13** and **16Xe** are **A2** and have no ROA intensity. The modes **10** and **11** are shifted (**14Xe** and **15Xe**) to higher frequencies, leading to the disappearance of the peak around 125 cm^{-1}. Thus, the peak around 132 cm^{-1} becomes a superposition of three normal modes (**13Xe**, **14Xe** and **15Xe**), while only the mode **12** contributes to the peak without xenon.

Next, both solvent effects and xenon encapsulation have been computed on our target system. The solvent is CH_2Cl_2 and has been computed in IEF–PCM and SMD approaches (Figs. 5.15 and 5.16). The global spectrum is again not hugely impacted even at the low frequency region.

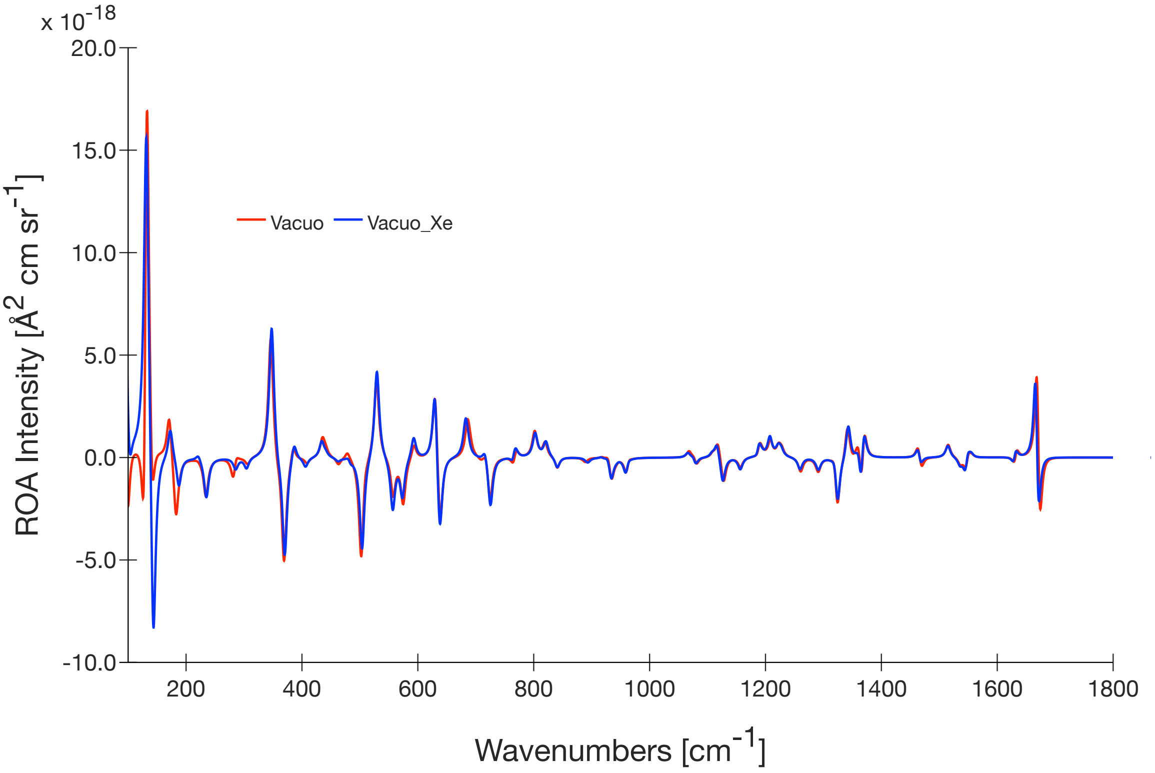


Figure 5.12: Simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G** ROA spectra of **1** with/without xenon in gas phase, in the 100—1800 cm^{-1} region.

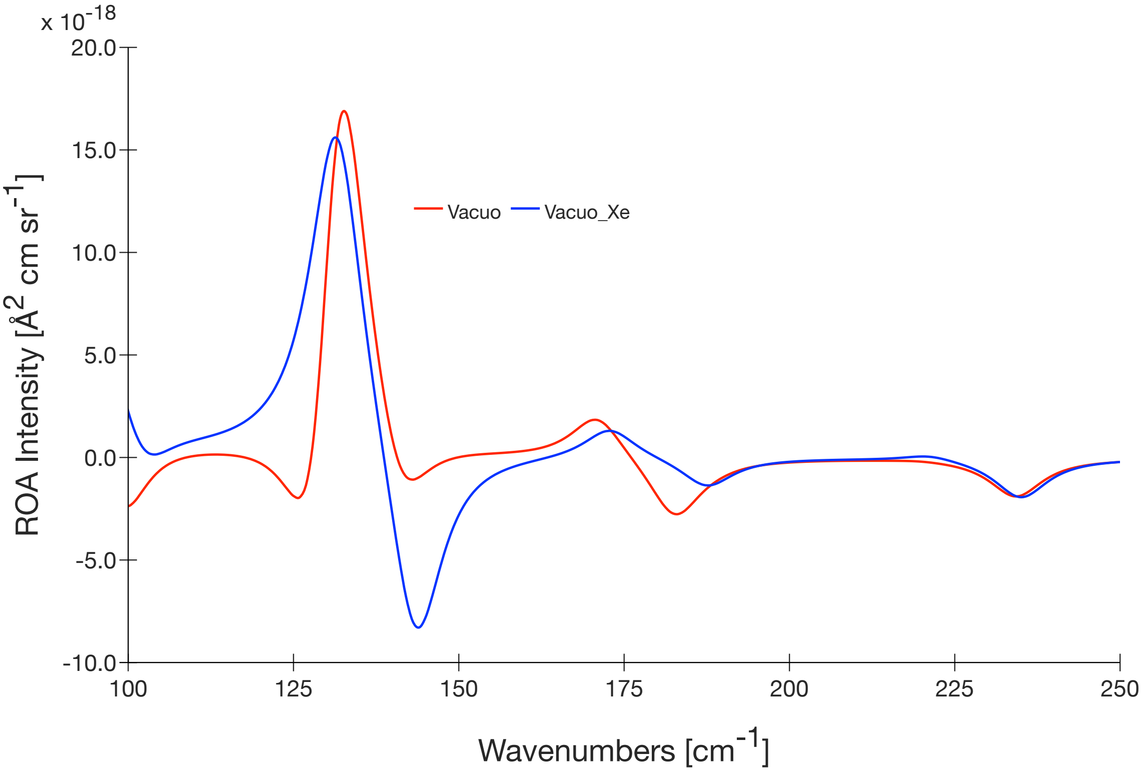
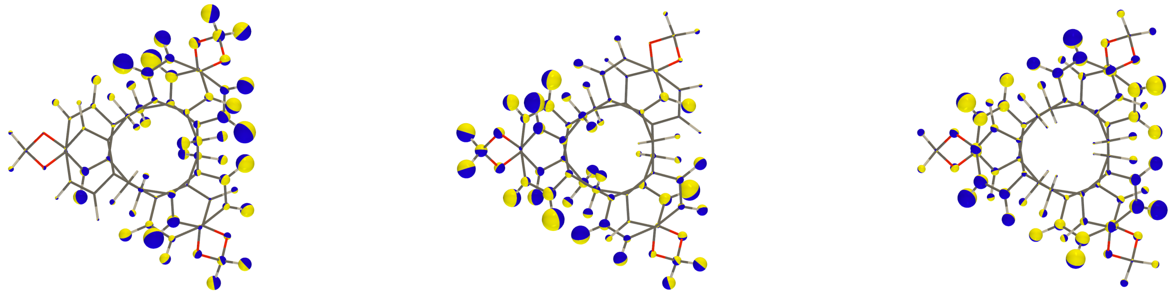
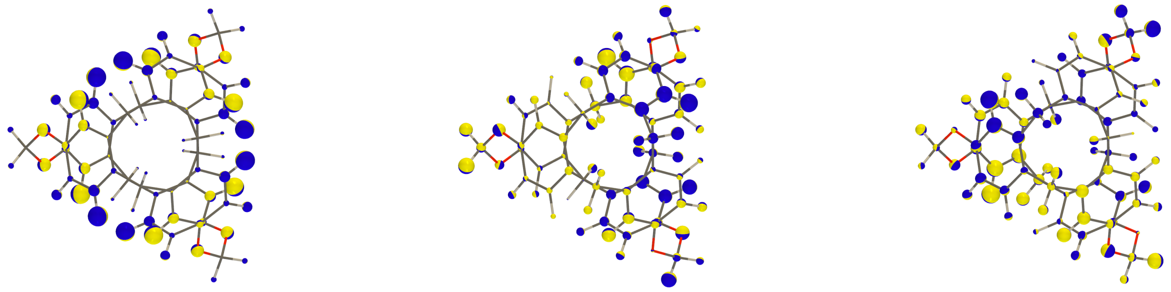


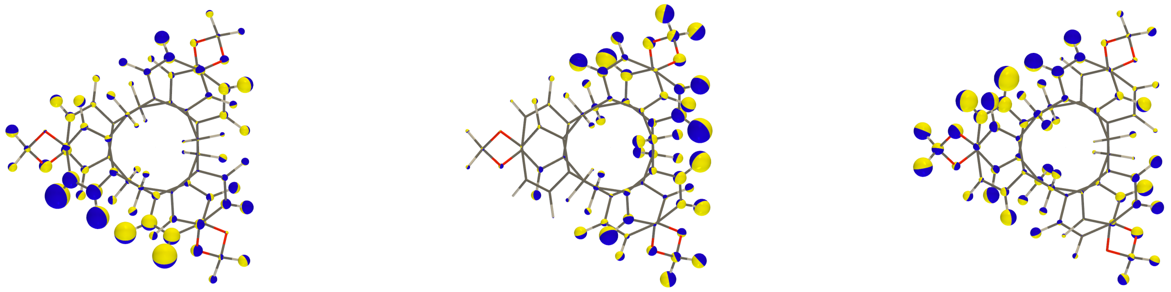
Figure 5.13: Simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G** ROA spectra of **1** with/without xenon in gas phase, in the 100—250 cm^{-1} region.



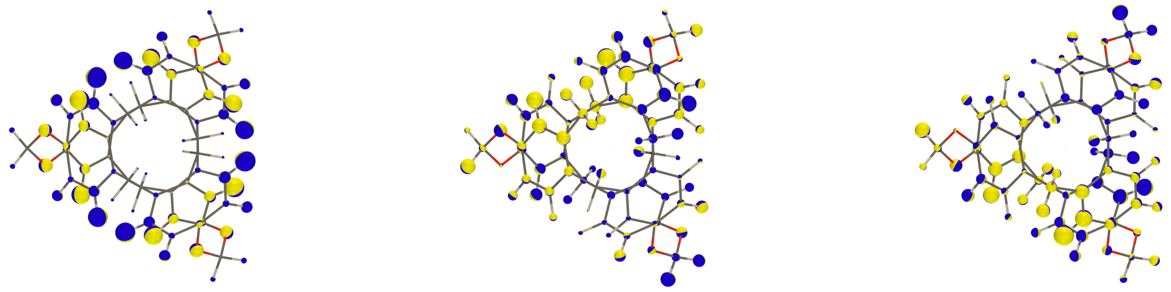
E – **10** (128.3 cm^{-1}) and **11** (128.4 cm^{-1}) **A1** – **12** (131.9 cm^{-1})



A2 – **13** (137.1 cm^{-1}) **E** – **14** (141.1 cm^{-1}) and **15** (141.3 cm^{-1})



A1 – **13Xe** (131.5 cm^{-1}) **E** – **14Xe** (131.7 cm^{-1}) and **15Xe** (131.8 cm^{-1})



A2 – **16Xe** (137.0 cm^{-1}) **E** – **17Xe** (143.4 cm^{-1}) and **18Xe** (143.5 cm^{-1})

Figure 5.14: Representation of the normal modes of **1** at the **B3LYP/6–31G*** level with/without xenon in gas phase for the low frequency region. The normal modes with the xenon are denoted "**Xe**". The irreducible representations of the normal modes are also given.

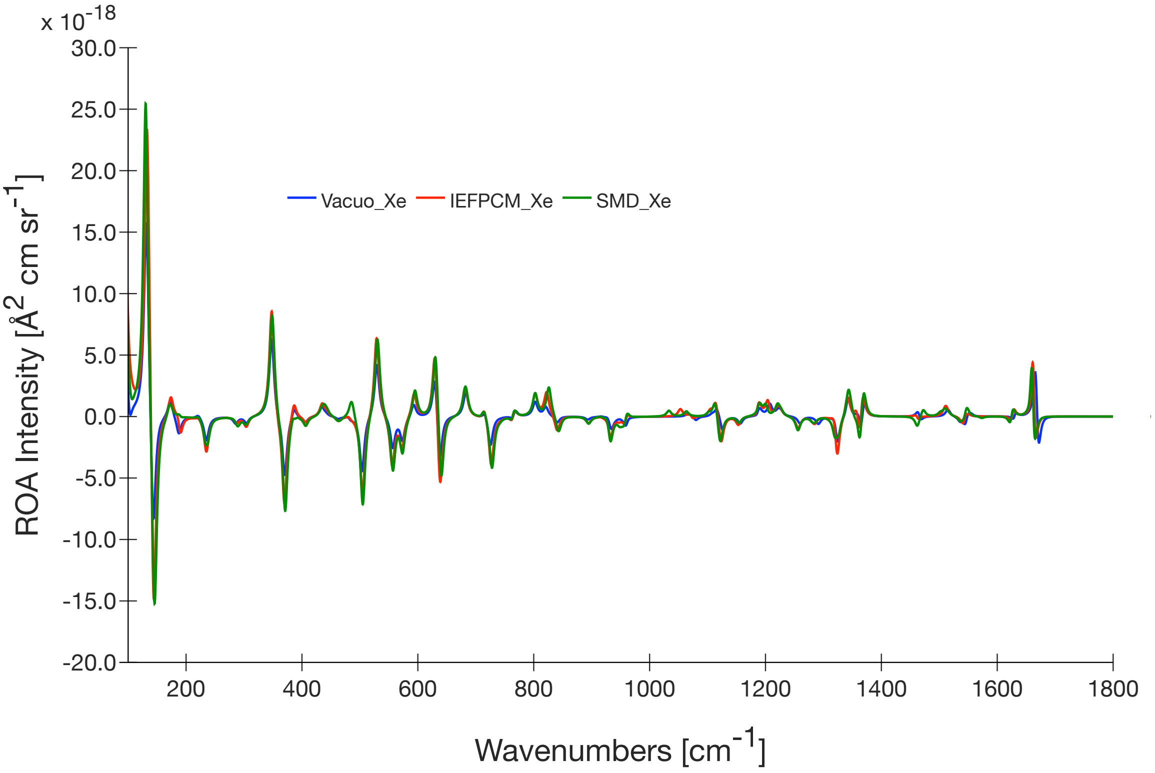


Figure 5.15: Simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G** ROA spectra of **1**, with xenon encapsulated in the cavity, with/without solvent (CH_2Cl_2), in the 100—1800 cm^{-1} region.

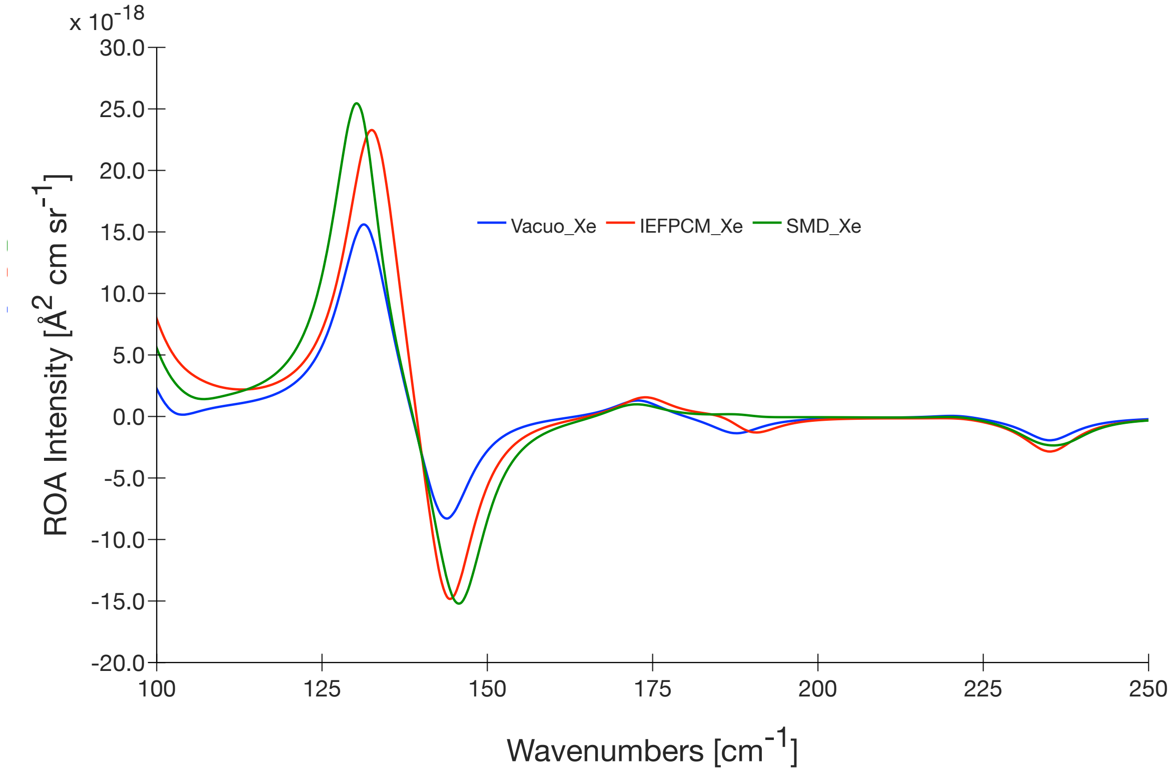


Figure 5.16: Simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G** ROA spectra of **1**, with xenon encapsulated in the cavity, with/without solvent (CH_2Cl_2), in the 100—250 cm^{-1} region.

5.2.2.3 Comparison to Experiment

Finally, our simulated method is compared to experimental data. The experiments have been recorded by our collaborators at the University of Bordeaux and published in the article of Thierry Buffeteau *et al.* [25]. They highlighted a decrease in the absolute intensity of a peak around 150 cm^{-1} with the xenon encapsulation. One should note that such signatures at very low wavenumbers are very difficult to record since they are often hidden by the Rayleigh band. Thanks to the upgrade of their equipment by Nicolas Daugey, they were able to measure the molecular responses in this region and it leads to these new results.

Since the experimental spectra, with/without xenon, have been recorded in CD_2Cl_2, they are compared with our simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G** spectra with/without xenon in CH_2Cl_2 (computed with the SMD approach), already presented earlier in Figs. 5.10 and 5.15. In Fig. 5.17, the global experimental spectra are well reproduced by our simulations, provided that the frequencies are scaled by a scaling factor of **0.98**. The most striking differences appear in the 1200—1500 cm^{-1} region for which our simulations do not reproduce perfectly the intensity of the peaks. Then, a zoom has been performed on the low frequency region (Fig. 5.18). For this region, another scaling factor (of **1.14**) has been used to better superpose the peaks. We observe that the positive peak around 150 cm^{-1} is quite well reproduced by our simulations but not its intensity. In addition, the decrease of the peak intensity around 150 cm^{-1}, observed experimentally with the xenon encapsulation, is not at all reproduced by our simulations. They present an increase of the intensity.

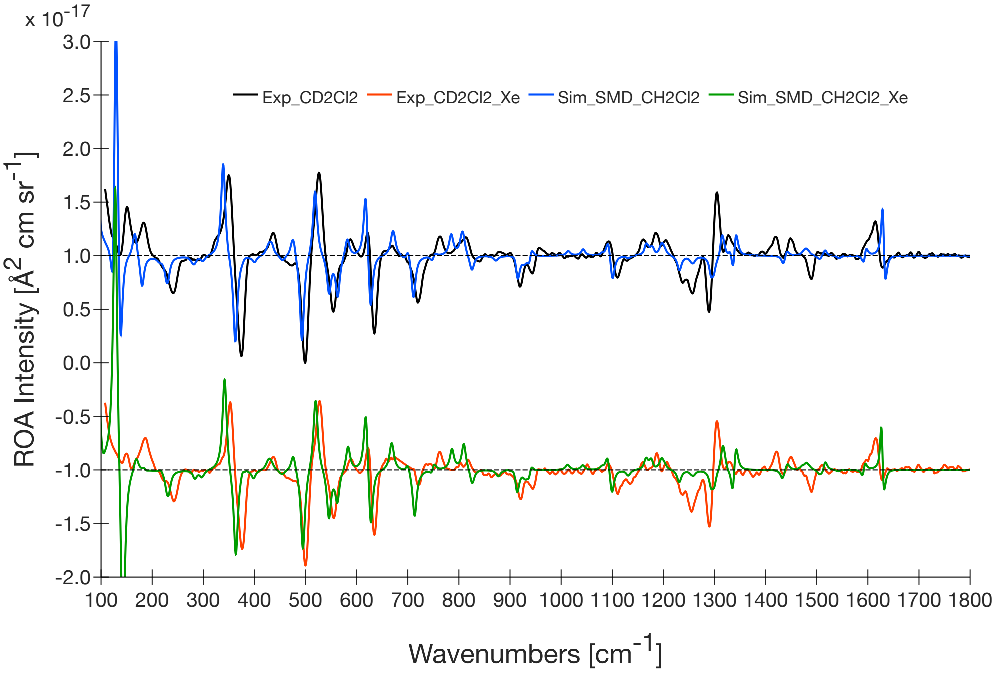


Figure 5.17: Experimental and simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G** in CH_2Cl_2 (computed with SMD) ROA spectra, with/without xenon encapsulated in the cryptophane cavity, in the 100—1800 cm^{-1} region.

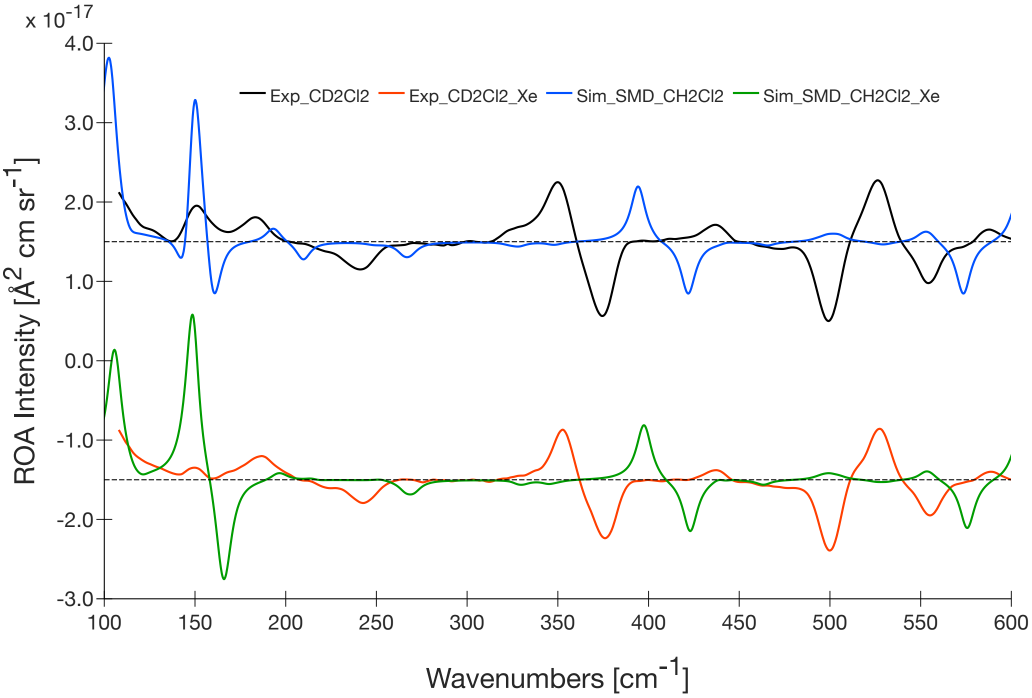


Figure 5.18: Experimental and simulated **B3LYP/6–31G*//TDHF/rDPS:3–21G** in CH_2Cl_2 (computed with SMD) ROA spectra, with/without xenon encapsulated in the cryptophane cavity, in the 100—600 cm^{-1} region.

The harmonic approximation has been used to simulate the Cryptophane–**PP**–111 ROA spectra, in solution and with xenon encapsulated in the cryptophane cavity. Therefore, the anharmonic contributions are not taken into account. However, they might be important to reproduce the experimental peaks at low frequencies. Unfortunately, it is very difficult to simulate the anharmonic ROA spectra. First, the evaluation of the quantities required is very cpu time–consuming. Second, the appropriate method might be different to treat the anharmonic effects than the one we have used to simulate the ROA spectra at the harmonic approximation.

Chapter 6. Conclusions and Outlook

6.1 Conclusions

In the first part of this work, we have investigated the impact of the method on the simulation of ROA spectra for a chiral reference system, **R**–methyloxirane. The simulated spectra using **B3LYP** and ω**B97X–D** XC functionals for the determination of the vibrational normal modes were in good agreement with the simulated spectrum obtained with our reference method **MP2/cc–pVTZ**. We have compared our reference spectrum with the experimental one recorded by Šebestík and Bouř [194] and we found a nearly perfect match between the two of them. For the evaluation of the Cartesian derivatives of polarizability tensors, the **rDPS:3–21G** basis set introduced by Zuber and Hug [182] gave impressive results, comparable to **aug–cc–pVDZ** but with a good reduction of cpu time (by a factor of around six).

In the second part, we have sampled the PES of the Cryptophane–PP–111 using the two conformational search programs (GMMX and CREST). After the reoptimization of the different conformers using the appropriate method (**B3LYP/6–31G***), only one conformer (**1**) has a substantial Boltzmann weight both in gas phase and in solution. The relative energies obtained when adding dispersion corrections on top of **B3LYP**, or already contained in ω**B97X–D**, are in contradiction with our **MP2** reference values. Subsequently, the ROA spectrum of **1** has been computed by including the effect of different solvents using the IEF–PCM approach. Moreover, for CH_2Cl_2, the IEF–PCM approach has been compared to the SMD one, which is supposed to better reproduce Gibbs free energy of solvation. The peaks at the low frequency region were more impacted by these approaches than the other ones. The spectrum using the SMD approach has been chosen for the comparison with the experiment. Afterwards, we added the xenon in the cryptophane cavity. The main observations were the shift of the peak around $125\,\mathrm{cm}^{-1}$ to higher frequencies ($132\,\mathrm{cm}^{-1}$) and the increase of the peak absolute intensity around $142\,\mathrm{cm}^{-1}$. The other peaks were not hugely impacted by the xenon encapsulation. Finally, we compared our final spectra (with/without xenon) and the experimental data [25]. Most of the ROA signatures were quite well reproduced except for the low frequency region certainly due to anharmonic effects.

At the end of this work, we have constructed a simulation protocol that is able to reproduce most of the experimental ROA signatures. Thus, for us, the global objective of this work has been completed. Some improvements are still possible especially at the low frequency region.

6.2 Outlook

As said above, the low frequency region could be better described if the anharmonic effects were taken into consideration. Moreover, in our simulations, the parameterization of the solvent effects in both IEF–PCM and SMD approaches can be studied further. The default surface is the vdW surface for the cavity model and it implies a hole at the center of the cryptophane. It means that the solvent could enter the cryptophane cavity. Experimentally, it has been proven that it is impossible for our system since its cavity is too small. Thus, other surfaces, to describe the cavity model, will be tested: SES and SAS. Preliminary calculations have been already performed and the representation of the different surfaces are showed in Fig. 6.1.

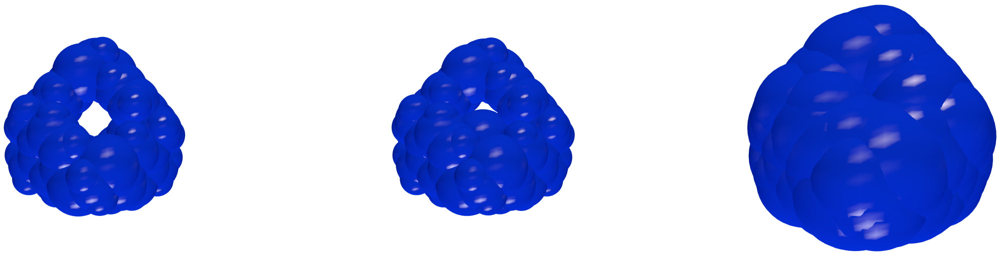


Figure 6.1: Sketches, for the Cryptophane–**PP**–111, of the vdW surface ≡ SES, without xenon (left), with xenon (middle), and the SAS with/without xenon (right) for CH_2Cl_2 in SMD.

At last, larger cryptophane derivatives systems will be considered: Cryptophane–111(OMe)$_6$, Cryptophane–222, Cryptophane–223, Cryptophane–333, . . . For these molecules, the solvent can enter in their cavity and it will require further and careful investigations.

6.3 Acknowledgements

The calculations were performed on the computers of the Consortium des Équipements de Calcul Intensif (CÉCI) and particularly those of the Technological Platform of High–Performance Computing, for which we gratefully acknowledge the financial support of the University of Namur (Convention Nos. 2.5020.11, GEQ U.G006.15, 1610468, and RW/GEQ2016).

I really want to thank all the people who helped me in this work, especially my promoter Vincent Liégeois and my direct collaborator Nicolas Daugey from th University of Bordeaux. I also want to thank my colleagues for their kindness and their advice. I want to thank the colleagues of Nicolas Daugey for their warm welcome.

Appendix A. R–methyloxirane

Table A.1: Number of identical modes, slope of the linear regression (LR), and RMSD for the vibrational frequencies between **M06/basis sets** and the reference **MP2/cc–pVTZ**.

Basis set	Identical Modes	Modes in LR	Slope in LR	RMSD (cm^{-1})
6–31G	14	10	1.003803	9.9
6–31G*	11	4	1.006968	9.7
6–31G**	10	3	1.013650	12.7
6–31++G**	10	3	1.019382	14.3
6–311G	15	9	1.005818	10.4
6–311G*	11	4	1.013529	11.0
6–311G**	10	3	1.019377	14.8
6–311++G**	10	3	1.020084	14.5
cc–pVDZ	9	/	1.000000	/
cc–pVTZ	11	4	1.017591	13.8
cc–pVQZ	12	5	1.015993	12.7
aug–cc–pVDZ	11	4	1.024234	9.4
aug–cc–pVTZ	12	5	1.016474	14.9

Table A.2: Number of identical modes, slope of the linear regression (LR), and RMSD for the vibrational frequencies between **M06–2X/basis sets** and the reference **MP2/cc–pVTZ**.

Basis set	Identical Modes	Modes in LR	Slope in LR	RMSD (cm^{-1})
6–31G	11	7	0.990549	16.7
6–31G*	16	7	0.987786	5.2
6–31G**	15	7	0.995169	6.2
6–31++G**	18	9	1.000449	5.5
6–311G	13	7	0.991421	16.7
6–311G*	15	8	0.992981	3.7
6–311G**	15	8	1.001021	5.8
6–311++G**	16	9	1.001397	5.5
cc–pVDZ	11	3	1.018213	13.5
cc–pVTZ	16	9	1.001071	6.4
cc–pVQZ	17	10	0.999850	6.9
aug–cc–pVDZ	14	7	1.011528	5.2
aug–cc–pVTZ	17	10	1.000934	6.2

Table A.3: Number of identical modes, slope of the linear regression (LR), and RMSD for the vibrational frequencies between ω**B97X/basis sets** and the reference **MP2/cc–pVTZ**.

Basis set	Identical Modes	Modes in LR	Slope in LR	RMSD (cm^{-1})
6–31G	9	6	0.984156	13.2
6–31G*	14	7	0.982717	7.5
6–31G**	14	7	0.991055	9.2
6–31++G**	17	8	0.996384	8.3
6–311G	9	6	0.986621	11.4
6–311G*	18	9	0.988752	9.1
6–311G**	15	6	0.999520	9.6
6–311++G**	18	9	0.997698	11.4
cc–pVDZ	9	/	1.000000	/
cc–pVTZ	18	9	0.997216	8.1
cc–pVQZ	18	9	0.997832	8.7
aug–cc–pVDZ	17	8	1.002233	13.2
aug–cc–pVTZ	20	11	0.995268	11.4

Table A.4: Number of identical modes, slope of the linear regression (LR), and RMSD for the vibrational frequencies between ω**B97X–D/basis sets** and the reference **MP2/cc–pVTZ**.

Basis set	Identical Modes	Modes in LR	Slope in LR	RMSD (cm^{-1})
6–31G	7	4	0.984572	13.4
6–31G*	18	9	0.983725	7.1
6–31G**	16	9	0.991548	8.3
6–31++G**	20	11	0.996050	7.6
6–311G	9	6	0.990883	19.7
6–311G*	20	11	0.990694	6.1
6–311G**	21	12	0.997989	9.4
6–311++G**	22	13	0.998964	8.4
cc–pVDZ	13	6	1.007962	15.2
cc–pVTZ	21	12	0.996791	9.7
cc–pVQZ	20	11	0.997904	9.0
aug–cc–pVDZ	21	12	1.004492	7.6
aug–cc–pVTZ	21	12	0.997496	9.4

Appendix B. $\mathbf{XeF_n}$ systems

Table B.1: Geometrical parameters of the XeF_2 structure optimized in ω**B97X–D** with different basis set for the Xe atom. The reference value for the Xe–F distance is 1.977 Å [197, 198].

Basis set	d_{Xe-F} (Å)	Δ
cc–pVDZ–PP	2.022	0.045
cc–pVTZ–PP	1.990	0.013
cc–pVQZ–PP	1.977	0.000
cc–pV5Z–PP	1.974	0.003
cc–pwCVDZ–PP	2.016	0.039
cc–pwCVTZ–PP	1.982	0.005
cc–pwCVQZ–PP	1.971	0.006
cc–pwCV5Z–PP	1.967	0.010
aug–cc–pVDZ–PP	2.022	0.045
aug–cc–pVTZ–PP	1.996	0.019
aug–cc–pVQZ–PP	1.981	0.004
aug–cc–pV5Z–PP	1.977	0.000
aug–cc–pwCVDZ–PP	2.018	0.041
aug–cc–pwCVTZ–PP	1.988	0.011
aug–cc–pwCVQZ–PP	1.975	0.002
aug–cc–pwCV5Z–PP	1.970	0.007
def2–TZVPP	1.991	0.014
def2–TZVPPD	1.994	0.017
def2–QZVPP	1.972	0.005
def2–TZVPPD	1.974	0.003
dhf–TZVPP	1.991	0.014
dhf–QZVPP	1.972	0.005
CRENBL	2.012	0.035
CRENBS	2.017	0.040
LANL2DZ	2.062	0.085
LANL08	2.057	0.080
SBKJC–Polar.–LFK	2.031	0.054
SBKJC–VDZ	2.065	0.088
Stuttgart-RLC	2.004	0.027

Table B.2: Geometrical parameters of the XeF_4 structure optimized in ω**B97X–D** with different basis set for the Xe atom. The reference value for the Xe–F distance is 1.953 Å [199].

Basis set	d_{Xe-F} (Å)	Δ
cc–pVDZ–PP	1.986	0.033
cc–pVTZ–PP	1.952	0.001
cc–pVQZ–PP	1.938	0.015
cc–pV5Z–PP	1.936	0.017
cc–pwCVDZ–PP	1.980	0.027
cc–pwCVTZ–PP	1.941	0.012
cc–pwCVQZ–PP	1.931	0.022
cc–pwCV5Z–PP	1.929	0.024
aug–cc–pVDZ–PP	1.990	0.037
aug–cc–pVTZ–PP	1.957	0.004
aug–cc–pVQZ–PP	1.942	0.011
aug–cc–pV5Z–PP	1.939	0.014
aug–cc–pwCVDZ–PP	1.984	0.031
aug–cc–pwCVTZ–PP	1.947	0.006
aug–cc–pwCVQZ–PP	1.934	0.019
aug–cc–pwCV5Z–PP	1.931	0.022
def2–TZVPP	1.949	0.004
def2–TZVPPD	1.952	0.001
def2–QZVPP	1.931	0.022
def2–TZVPPD	1.932	0.021
dhf–TZVPP	1.949	0.004
dhf–QZVPP	1.931	0.022
CRENBL	1.974	0.021
CRENBS	1.982	0.029
LANL2DZ	2.021	0.068
LANL08	2.020	0.067
SBKJC–Polar.–LFK	2.002	0.049
SBKJC–VDZ	2.034	0.081
Stuttgart-RLC	1.968	0.015

Table B.3: Geometrical parameters of the XeF_6 structure optimized in ω**B97X–D** with different basis set for the Xe atom. The reference value for the Xe–F distance is 1.890 Å [200].

Basis set	d_{Xe-F} (Å)	Δ
cc–pVDZ–PP	1.969	0.079
cc–pVTZ–PP	1.937	0.047
cc–pVQZ–PP	1.924	0.034
cc–pV5Z–PP	1.923	0.033
cc–pwCVDZ–PP	1.963	0.073
cc–pwCVTZ–PP	1.925	0.035
cc–pwCVQZ–PP	1.916	0.026
cc–pwCV5Z–PP	1.914	0.024
aug–cc–pVDZ–PP	1.975	0.085
aug–cc–pVTZ–PP	1.942	0.052
aug–cc–pVQZ–PP	1.927	0.037
aug–cc–pV5Z–PP	1.925	0.035
aug–cc–pwCVDZ–PP	1.968	0.078
aug–cc–pwCVTZ–PP	1.930	0.040
aug–cc–pwCVQZ–PP	1.918	0.028
aug–cc–pwCV5Z–PP	1.916	0.026
def2–TZVPP	1.932	0.042
def2–TZVPPD	1.936	0.046
def2–QZVPP	1.916	0.026
def2–TZVPPD	1.916	0.026
dhf–TZVPP	1.932	0.042
dhf–QZVPP	1.916	0.026
CRENBL	1.959	0.069
CRENBS	1.964	0.074
LANL2DZ	1.995	0.105
LANL08	1.997	0.107
SBKJC–Polar.–LFK	1.994	0.104
SBKJC–VDZ	2.020	0.130
Stuttgart-RLC	1.951	0.061

Table B.4: Frequency values for the normal modes of the XeF_2 with ω**B97X–D** and different basis set for the Xe atom. The reference values for $\nu_{1=2}$ and ν_4 are 213.2 cm^{-1} and 555.0 cm^{-1}, respectively [201]. Δ is the difference between experimental and reference values in cm^{-1}.

Basis set	$\nu_{1=2}$ (cm^{-1})	ν_4 (cm^{-1})	$\Delta_{1=2}$	Δ_4	RMSD
cc–pVDZ–PP	211.8	589.2	1.4	34.2	24.2
cc–pVTZ–PP	213.6	568.1	0.4	13.1	9.3
cc–pVQZ–PP	217.4	568.8	4.2	13.8	10.2
cc–pV5Z–PP	218.2	574.5	5.0	19.5	14.2
cc–pwCVDZ–PP	211.3	581.3	1.9	26.3	18.6
cc–pwCVTZ–PP	213.5	564.1	0.3	9.1	6.4
cc–pwCVQZ–PP	217.0	570.2	3.8	15.2	11.1
cc–pwCV5Z–PP	217.6	582.9	4.4	27.9	20.0
aug–cc–pVDZ–PP	201.6	587.6	11.6	32.6	24.4
aug–cc–pVTZ–PP	207.3	560.5	5.9	5.5	5.7
aug–cc–pVQZ–PP	217.5	561.2	4.3	6.2	5.3
aug–cc–pV5Z–PP	219.1	570.9	5.9	15.9	12.0
aug–cc–pwCVDZ–PP	200.8	579.8	12.4	24.8	19.6
aug–cc–pwCVTZ–PP	207.0	555.7	6.2	0.7	4.4
aug–cc–pwCVQZ–PP	216.6	564.1	3.4	9.1	6.9
aug–cc–pwCV5Z–PP	218.2	576.4	5.0	21.4	15.5
def2–TZVPP	211.1	558.8	2.1	3.8	3.1
def2–TZVPPD	207.7	554.0	5.5	1.0	4.0
def2–QZVPP	216.9	568.5	3.7	13.5	9.9
def2–TZVPPD	215.3	562.4	2.1	7.4	5.4
dhf–TZVPP	211.1	558.8	2.1	3.8	3.1
dhf–QZVPP	216.9	568.5	3.7	13.5	9.9
CRENBL	218.8	570.3	5.6	15.3	11.5
CRENBS	225.0	569.0	11.8	14.0	13.0
LANL2DZ	201.7	524.6	11.5	30.4	23.0
LANL08	202.0	519.3	11.2	35.7	26.5
SBKJC–Polar.–LFK	214.1	557.6	0.9	2.6	2.0
SBKJC–VDZ	200.8	557.0	12.4	2.0	8.9
Stuttgart-RLC	215.4	555.2	2.2	0.2	1.6

Table B.5: Frequency values for the normal modes of the XeF_4 with ω**B97X–D** and different basis set for the Xe atom. The reference values for ν_5 and $\nu_{8=9}$ are $291.0\,\mathrm{cm}^{-1}$ and $586.0\,\mathrm{cm}^{-1}$, respectively [201]. Δ is the difference between experimental and reference values in cm^{-1}.

Basis set	ν_5 (cm^{-1})	$\nu_{8=9}$ (cm^{-1})	Δ_5	$\Delta_{8=9}$	RMSD
cc–pVDZ–PP	283.9	608.1	7.1	22.1	16.4
cc–pVTZ–PP	294.5	596.1	3.5	10.1	7.5
cc–pVQZ–PP	301.0	604.5	10.0	18.5	14.9
cc–pV5Z–PP	301.4	606.0	10.4	20.0	15.9
cc–pwCVDZ–PP	284.4	598.3	6.6	12.3	9.9
cc–pwCVTZ–PP	295.5	598.1	4.5	12.1	9.1
cc–pwCVQZ–PP	301.3	604.6	10.3	18.6	15.0
cc–pwCV5Z–PP	301.6	608.4	10.6	22.4	17.5
aug–cc–pVDZ–PP	271.1	600.4	19.9	14.4	17.4
aug–cc–pVTZ–PP	287.7	583.0	3.3	3.0	3.2
aug–cc–pVQZ–PP	298.9	593.7	7.9	7.7	7.8
aug–cc–pV5Z–PP	299.9	599.6	8.9	13.6	11.5
aug–cc–pwCVDZ–PP	271.2	590.6	19.8	4.6	14.4
aug–cc–pwCVTZ–PP	288.6	584.5	2.4	1.5	2.0
aug–cc–pwCVQZ–PP	298.5	594.8	7.5	8.8	8.2
aug–cc–pwCV5Z–PP	299.9	600.7	8.9	14.7	12.1
def2–TZVPP	293.1	585.7	2.1	0.3	1.5
def2–TZVPPD	288.7	576.0	2.3	10.0	7.3
def2–QZVPP	301.1	605.3	10.1	19.3	15.4
def2–TZVPPD	299.2	599.9	8.2	13.9	11.4
dhf–TZVPP	293.1	585.7	2.1	0.3	1.5
dhf–QZVPP	301.1	605.3	10.1	19.3	15.4
CRENBL	293.7	604.7	2.7	18.7	13.4
CRENBS	291.5	605.7	0.5	19.7	13.9
LANL2DZ	265.8	554.8	25.2	31.2	28.3
LANL08	260.2	544.5	30.8	41.5	36.5
SBKJC–Polar.–LFK	283.4	589.6	7.6	3.6	6.0
SBKJC–VDZ	260.6	590.6	30.4	4.6	21.8
Stuttgart-RLC	293.6	599.3	2.6	13.3	9.6

Table B.6: Frequency values for the normal modes of the XeF_6 with ω**B97X–D** and different basis set for the Xe atom. The reference values for $\nu_{10=11}$ and $\nu_{12=13=14}$ are 520.0 cm^{-1} and 612.0 cm^{-1}, respectively [201]. Δ is the difference between experimental and reference values in cm^{-1}.

Basis set	$\nu_{10=11}$ (cm^{-1})	$\nu_{12=13=14}$ (cm^{-1})	$\Delta_{10=11}$	$\Delta_{12=13=14}$	RMSD
cc–pVDZ–PP	529.6	612.6	9.6	0.6	6.8
cc–pVTZ–PP	523.2	598.4	3.2	13.6	9.9
cc–pVQZ–PP	539.6	602.1	19.6	9.9	15.6
cc–pV5Z–PP	541.7	600.4	21.7	11.6	17.4
cc–pwCVDZ–PP	524.3	603.0	4.3	9.0	7.1
cc–pwCVTZ–PP	531.2	597.0	11.2	15.0	13.2
cc–pwCVQZ–PP	548.0	601.0	28.0	11.0	21.2
cc–pwCV5Z–PP	550.5	600.9	30.5	11.1	22.9
aug–cc–pVDZ–PP	522.7	601.5	2.7	10.5	7.6
aug–cc–pVTZ–PP	518.8	585.3	1.2	26.7	18.9
aug–cc–pVQZ–PP	533.0	591.4	13.0	20.6	17.2
aug–cc–pV5Z–PP	538.1	594.1	18.1	17.9	18.0
aug–cc–pwCVDZ–PP	517.1	591.2	2.9	20.8	14.9
aug–cc–pwCVTZ–PP	524.9	583.1	4.9	28.9	20.7
aug–cc–pwCVQZ–PP	541.7	590.4	21.7	21.6	21.7
aug–cc–pwCV5Z–PP	546.4	593.6	26.4	18.4	22.8
def2–TZVPP	512.1	586.6	7.9	25.4	18.8
def2–TZVPPD	505.9	576.1	14.1	35.9	27.3
def2–QZVPP	547.1	601.4	27.1	10.6	20.6
def2–TZVPPD	544.4	597.1	24.4	14.9	20.3
dhf–TZVPP	512.1	586.6	7.9	25.4	18.8
dhf–QZVPP	547.3	601.7	27.3	10.3	20.7
CRENBL	517.9	609.0	2.1	3.0	2.6
CRENBS	513.9	595.0	6.1	17.0	12.8
LANL2DZ	489.6	554.3	30.4	57.7	46.1
LANL08	477.6	554.4	42.4	57.6	50.6
SBKJC–Polar.–LFK	481.9	566.9	38.1	45.1	41.8
SBKJC–VDZ	503.3	588.1	16.7	23.9	20.6
Stuttgart-RLC	516.6	580.6	3.4	31.4	22.3

Bibliography

[1] Canceill, J.; Lacombe, L.; Collet, A. Analytical optical resolution of bromochlorofluoromethane by enantioselective inclusion into a tailor-made cryptophane and determination of its maximum rotation. *J. Am. Chem. Soc.* **1985**, *107*, 6993–6996.

[2] Collet, A. Cyclotriveratrylenes and cryptophanes. *Tetrahedron* **1987**, *43*, 5725–5759.

[3] Garel, L.; Lozach, B.; Dutasta, J. P.; Collet, A. Remarkable effect of the receptor size in the binding of acetylcholine and related ammonium ions to water-soluble cryptophanes. *J. Am. Chem. Soc.* **1993**, *115*, 11652–11653.

[4] Gambut, L.; Chauvet, J.-P.; Garcia, C.; Berge, B.; Renault, A.; Rivière, S.; Meunier, J.; Collet, A. Ellipsometry, Brewster Angle Microscopy, and Thermodynamic Studies of Monomolecular Films of Cryptophanes at the Air-Water Interface. *Langmuir* **1996**, *12*, 5407–5412.

[5] Kirchhoff, P. D.; Bass, M. B.; Hanks, B. A.; Briggs, J. M.; Collet, A.; McCammon, J. A. Structural Fluctuations of a Cryptophane Host: A Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **1996**, *118*, 3237–3246.

[6] Costante-Crassous, J.; Marrone, T. J.; Briggs, J. M.; McCammon, J. A.; Collet, A. Absolute Configuration of Bromochlorofluoromethane from Molecular Dynamics Simulation of Its Enantioselective Complexation by Cryptophane-C. *J. Am. Chem. Soc.* **1997**, *119*, 3818–3823.

[7] Kirchhoff, P. D.; Dutasta, J.-P.; Collet, A.; McCammon, J. A. Structural Fluctuations of a Cryptophane-Tetramethylammonium Host-Guest System: A Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **1997**, *119*, 8015–8022.

[8] Bartik, K.; Luhmer, M.; Dutasta, J.-P.; Collet, A.; Reisse, J. 129Xe and 1H NMR Study of the Reversible Trapping of Xenon by Cryptophane-A in Organic Solution. *J. Am. Chem. Soc.* **1998**, *120*, 784–791.

[9] Kirchhoff, P. D.; Dutasta, J.-P.; Collet, A.; McCammon, J. A. Dynamic and Rotational Analysis of Cryptophane Host-Guest Systems: Challenges of Describing Molecular Recognition. *J. Am. Chem. Soc.* **1999**, *121*, 381–390.

[10] Brotin, T.; Lesage, A.; Emsley, L.; Collet, A. 129Xe NMR Spectroscopy of Deuterium-Labeled Cryptophane-A Xenon Complexes: Investigation of Host-Guest Complexation Dynamics. *J. Am. Chem. Soc.* **2000**, *122*, 1171–1174.

[11] Huber, J. G.; Dubois, L.; Desvaux, H.; Dutasta, J.-P.; Brotin, T.; Berthault, P. NMR Study of Optically Active Monosubstituted Cryptophanes and Their Interaction with Xenon. *J. Phys. Chem. A* **2004**, *108*, 9608–9615.

[12] Brotin, T.; Roy, V.; Dutasta, J.-P. Improved Synthesis of Functional CTVs and Cryptophanes Using $Sc(OTf)_3$ as Catalyst. *J. Org. Chem.* **2005**, *70*, 6187–6195.

[13] Brotin, T.; Cavagnat, D.; Dutasta, J.-P.; Buffeteau, T. Vibrational Circular Dichroism Study of Optically Pure Cryptophane-A. *J. Am. Chem. Soc.* **2006**, *128*, 5533–5540.

[14] Huber, G.; Brotin, T.; Dubois, L.; Desvaux, H.; Dutasta, J.-P.; Berthault, P. Water Soluble Cryptophanes Showing Unprecedented Affinity for Xenon: Candidates as NMR-Based Biosensors. *J. Am. Chem. Soc.* **2006**, *128*, 6239–6246.

[15] Huber, G.; Beguin, L.; Desvaux, H.; Brotin, T.; Fogarty, H. A.; Dutasta, J.-P.; Berthault, P. Cryptophane-Xenon Complexes in Organic Solvents Observed through NMR Spectroscopy. *J. Phys. Chem. A* **2008**, *112*, 11363–11372.

[16] Bouchet, A.; Brotin, T.; Cavagnat, D.; Buffeteau, T. Induced Chiroptical Changes of a Water-Soluble Cryptophane by Encapsulation of Guest Molecules and Counterion Effects. *Chem. Eur. J.* **2010**, *16*, 4507–4518.

[17] Bouchet, A.; Brotin, T.; Linares, M.; Ågren, H.; Cavagnat, D.; Buffeteau, T. Enantioselective Complexation of Chiral Propylene Oxide by an Enantiopure Water-Soluble Cryptophane. *J. Org. Chem.* **2011**, *76*, 4178–4181.

[18] Brotin, T.; Montserret, R.; Bouchet, A.; Cavagnat, D.; Linares, M.; Buffeteau, T. High Affinity of Water-Soluble Cryptophanes for Cesium Cations. *J. Org. Chem.* **2012**, *77*, 1198–1201.

[19] Brotin, T.; Cavagnat, D.; Berthault, P.; Montserret, R.; Buffeteau, T. Water-Soluble Molecular Capsule for the Complexation of Cesium and Thallium Cations. *J. Phys. Chem. B* **2012**, *116*, 10905–10914.

[20] Brotin, T.; Goncalves, S.; Berthault, P.; Cavagnat, D.; Buffeteau, T. Influence of the Cavity Size of Water-Soluble Cryptophanes on Their Binding Properties for Cesium and Thallium Cations. *J. Phys. Chem. B* **2013**, *117*, 12593–12601.

[21] Daugey, N.; Brotin, T.; Vanthuyne, N.; Cavagnat, D.; Buffeteau, T. Raman Optical Activity of Enantiopure Cryptophanes. *J. Phys. Chem. B* **2014**, *118*, 5211–5217.

[22] Brotin, T.; Daugey, N.; Vanthuyne, N.; Jeanneau, E.; Ducasse, L.; Buffeteau, T. Chiroptical Properties of Cryptophane-223 and -233 Investigated by ECD, VCD, and ROA Spectroscopy. *J. Phys. Chem. B* **2015**, *119*, 8631–8639.

[23] Fogarty, H. A.; Berthault, P.; Brotin, T.; Huber, G.; Desvaux, H.; Dutasta, J.-P. A Cryptophane Core Optimized for Xenon Encapsulation. *J. Am. Chem. Soc.* **2007**, *129*, 10332–10333.

[24] Chaffee, K. E.; Fogarty, H. A.; Brotin, T.; Goodson, B. M.; Dutasta, J.-P. Encapsulation of Small Gas Molecules by Cryptophane-111 in Organic Solution. 1. Size- and Shape-Selective Complexation of Simple Hydrocarbons. *J. Phys. Chem. A* **2009**, *113*, 13675–13684.

[25] Buffeteau, T.; Pitrat, D.; Daugey, N.; Calin, N.; Jean, M.; Vanthuyne, N.; Ducasse, L.; Wien, F.; Brotin, T. Chiroptical properties of cryptophane-111. *Phys. Chem. Chem. Phys.* **2017**, *19*, 18303–18310.

[26] Schrödinger, E. Quantisierung als Eigenwertproblem. *Ann. Phys.* **1926**, *384*, 361–376.

[27] Schrödinger, E. Quantisierung als Eigenwertproblem. *Ann. Phys.* **1926**, *384*, 489–527.

[28] Schrödinger, E. Quantisierung als Eigenwertproblem. *Ann. Phys.* **1926**, *385*, 437–490.

[29] Schrödinger, E. Quantisierung als Eigenwertproblem. *Ann. Phys.* **1926**, *386*, 109–139.

[30] Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.* **1926**, *28*, 1049–1070.

[31] Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **1927**, *389*, 457–484.

[32] Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Math. Proc. Cambr. Phil. Soc.* **1928**, *24*, 89–110.

[33] Gaunt, J. A. A Theory of Hartree's Atomic Fields. *Math. Proc. Cambr. Phil. Soc.* **1928**, *24*, 328–342.

[34] Waller, I.; Hartree, D. R.; Fowler, R. H. On the intensity of total scattering of X-rays. *Proc. Roy. Soc. London* **1929**, *124*, 119–142.

[35] Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeit. Phys.* **1930**, *61*, 126–148.

[36] Slater, J. C. The Theory of Complex Spectra. *Phys. Rev.* **1929**, *34*, 1293–1322.

[37] Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69–89.

[38] Hall, G. G.; Lennard-Jones, J. E. The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials. *Proc. Roy. Soc. London* **1951**, *205*, 541–552.

[39] Hehre, W. J.; Ditchfield, K.; Pople, J. A. Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

[40] Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[41] Dolg, M.; Cao, X. Relativistic Pseudopotentials: Their Development and Scope of Applications. *Chem. Rev.* **2012**, *112*, 403–480.

[42] Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16–18 elements. *J. Chem. Phys.* **2003**, *119*, 11113–11123.

[43] LaJohn, L. A.; Christiansen, P. A.; Ross, R. B.; Atashroo, T.; Ermler, W. C. Ab initio relativistic effective potentials with spin–orbit operators. III. Rb through Xe. *J. Chem. Phys.* **1987**, *87*, 2812–2824.

[44] Wadt, W. R.; Hay, P. J. Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *J. Chem. Phys.* **1985**, *82*, 284–298.

[45] Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. Relativistic compact effective potentials and efficient, shared-exponent basis sets for the third-, fourth-, and fifth-row atoms. *Can. J. Chem.* **1992**, *70*, 612–630.

[46] Nicklass, A.; Dolg, M.; Stoll, H.; Preuss, H. Ab initio energy-adjusted pseudopotentials for the noble gases Ne through Xe: Calculation of atomic dipole and quadrupole polarizabilities. *J. Chem. Phys.* **1995**, *102*, 8942–8952.

[47] Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

[48] Strutt, J. W. *The Theory of Sound*, 2nd ed.; Dover Publications, New York, 1894.

[49] Frisch, M. J.; Head-Gordon, M.; Pople, J. A. A direct MP2 gradient method. *Chem. Phys. Lett.* **1990**, *166*, 275–280.

[50] Frisch, M. J.; Head-Gordon, M.; Pople, J. A. Semi-direct algorithms for the MP2 energy and gradient. *Chem. Phys. Lett.* **1990**, *166*, 281–289.

[51] Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, *153*, 503–506.

[52] Sæbø, S.; Almlöf, J. Avoiding the integral storage bottleneck in LCAO calculations of electron correlation. *Chem. Phys. Lett.* **1989**, *154*, 83–89.

[53] Head-Gordon, M.; Head-Gordon, T. Analytic MP2 frequencies without fifth-order storage. Theory and application to bifurcated hydrogen bonds in the water hexamer. *Chem. Phys. Lett.* **1994**, *220*, 122–128.

[54] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

[55] Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

[56] Parr, R. G.; Yang, W. *Density-functional theory of atoms and molecules*; Oxford Univ. Press, 1989.

[57] Salahub, D. R.; Zerner, M. C. *The Challenge of d and f Electrons*; ACS, Washington, D.C., 1989.

[58] Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

[59] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

[60] Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200–206.

[61] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[62] Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 1432–2234.

[63] Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 084106.

[64] Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

[65] Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

[66] Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

[67] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

[68] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[69] Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150*, 154122.

[70] Eisenschitz, R.; London, F. Über das Verhältnis der van der Waalsschen Kräfte zu den homöopolaren Bindungskräften. *ZS. f. Phys.* **1930**, *60*, 491–527.

[71] Wu, Q.; Yang, W. Empirical correction to density functional theory for van der Waals interactions. *J. Chem. Phys.* **2002**, *116*, 515–524.

[72] Halgren, T. A. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.

[73] Margoliash, D. J.; Meath, W. J. Pseudospectral dipole oscillator strength distributions and some related two body interaction coefficients for H, He, Li, N, O, H_2, N_2, O_2, NO, N_2O, H_2O, NH_3, and CH_4. *J. Chem. Phys.* **1978**, *68*, 1426–1431.

[74] Jhanwar, B. L.; Meath, W. J. Pseudo-spectral dipole oscillator strength distributions for the normal alkanes through octane and the evaluation of some related dipole-dipole and triple-dipole dispersion interaction energy coefficients. *Molec. Phys.* **1980**, *41*, 1061–1070.

[75] Jhanwar, B. L.; Meath, W. J. Dipole oscillator strength distributions, sums, and dispersion energy coefficients for CO and CO2. *Chem. Phys.* **1982**, *67*, 185–199.

[76] Kumar, A.; Meath, W. J. Pseudo-spectral dipole oscillator strengths and dipole-dipole and triple-dipole dispersion energy coefficients for HF, HCl, HBr, He, Ne, Ar, Kr and Xe. *Molec. Phys.* **1985**, *54*, 823–833.

[77] Kumar, A.; Meath, W. J. Dipole oscillator strength properties and dispersion energies for acetylene and benzene. *Molec. Phys.* **1992**, *75*, 311–324.

[78] Burton, G. R.; Chan, W. F.; Cooper, G.; Brion, C.; Kumar, A.; Meath, W. J. Valence shell absolute photoabsorption oscillator strengths, constrained dipole oscillator strength distributions, and dipole properties for CH_3NH_2, $(CH_3)_2NH$, and $(CH_3)_3N$. *Can. J. Chem.* **1994**, *72*, 529–546.

[79] Kumar, A.; Meath, W. J. Isotropic dipole properties for acetone, acetaldehyde and formaldehyde. *Molec. Phys.* **1997**, *90*, 389–398.

[80] Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

[81] Johnson, E. R.; Becke, A. D. A post-Hartree–Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 024101.

[82] Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.* **2006**, *124*, 174104.

[83] Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1493.

[84] McWeeny, R. Some Recent Advances in Density Matrix Theory. *Rev. Mod. Phys.* **1960**, *32*, 335–369.

[85] McWeeny, R. Perturbation Theory for the Fock-Dirac Density Matrix. *Phys. Rev.* **1962**, *126*, 1028–1034.

[86] Stevens, R. M.; Pitzer, R. M.; Lipscomb, W. N. Perturbed Hartree—Fock Calculations. I. Magnetic Susceptibility and Shielding in the LiH Molecule. *J. Chem. Phys.* **1963**, *38*, 550–560.

[87] Gerratt, J.; Mills, I. M. Force Constants and Dipole-Moment Derivatives of Molecules from Perturbed Hartree–Fock Calculations. I. *J. Chem. Phys.* **1968**, *49*, 1719–1729.

[88] Dodds, J. L.; McWeeny, R.; Raynes, W. T.; Riley, J. P. SCF theory for multiple perturbations. *Mol. Phys.* **1977**, *33*, 611–617.

[89] Dodds, J. L.; McWeeny, R.; Sadlej, A. J. Self-consistent perturbation theory. *Mol. Phys.* **1977**, *34*, 1779–1791.

[90] Woliński, K.; Sadlej, A. J. Self-consistent perturbation theory. *Mol. Phys.* **1980**, *41*, 1419–1430.

[91] Osamura, Y.; Yamaguchi, Y.; Schaefer, H. F. Analytic configuration interaction (CI) gradient techniques for potential energy hypersurfaces. A method for open-shell molecular wave functions. *J. Chem. Phys.* **1981**, *75*, 2919–2922.

[92] Osamura, Y.; Yamaguchi, Y.; Schaefer, H. F. Generalization of analytic configuration interaction (CI) gradient techniques for potential energy hypersurfaces, including a solution to the coupled perturbed Hartree–Fock equations for multiconfiguration SCF molecular wave functions. *J. Chem. Phys.* **1982**, *77*, 383–390.

[93] Pulay, P. Second and third derivatives of variational energy expressions: Application to multiconfigurational self-consistent field wave functions. *J. Chem. Phys.* **1983**, *78*, 5043–5051.

[94] Dykstra, C. E.; Jasien, P. G. Derivative Hartree—Fock theory to all orders. *Chem. Phys. Lett.* **1984**, *109*, 388–393.

[95] Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.

[96] Gross, E. K. U.; Kohn, W. In *Density Functional Theory of Many-Fermion Systems*; Löwdin, P.-O., Ed.; Advances in Quantum Chemistry; Academic Press, 1990; Vol. 21; pp 255–291.

[97] Onsager, L. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.

[98] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3093.

[99] Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117–129.

[100] Miertuš, S.; Tomasi, J. Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes. *Chem. Phys.* **1982**, *65*, 239–245.

[101] Pascual-Ahuir, J. L.; Silla, E.; Tunon, I. GEPOL: An improved description of molecular surfaces. 3. A new algorithm for the computation of solvent–excluding surface. *J. Comput. Chem.* **1994**, *15*, 1127–1138.

[102] Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027–2094.

[103] Cammi, R.; Tomasi, J. Remarks on the Use of the Apparent Surface Charges (ASC) Methods in Solvation Problems: Iterative versus Matrix-Inversion Procedures and the Renormalization of the Apparent Charges. *J. Comput. Chem.* **1995**, *16*, 1449–1458.

[104] Barone, V.; Cossi, M.; Tomasi, J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.* **1997**, *107*, 3210–3221.

[105] Cancès, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

[106] Mennucci, B.; Cancès, E.; Tomasi, J. Evaluation of Solvent Effects in Isotropic and Anisotropic Dielectrics and in Ionic Solutions with a Unified Integral Equation Method: Theoretical Bases, Computational Implementation, and Numerical Applications. *J. Phys. Chem. B* **1997**, *101*, 10506–10517.

[107] Mennucci, B.; Tomasi, J. Continuum solvation models: A new approach to the problem of solute's charge distribution and cavity boundaries. *J. Chem. Phys.* **1997**, *106*, 5151–5158.

[108] Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

[109] Barone, V.; Cossi, M.; Tomasi, J. Geometry optimization of molecular structures in solution by the polarizable continuum model. *J. Comput. Chem.* **1998**, *19*, 404–417.

[110] Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. Ab initio study of ionic solutions by a polarizable continuum dielectric model. *Chem. Phys. Lett.* **1998**, *286*, 253–260.

[111] Cossi, M.; Barone, V.; Robb, M. A. A direct procedure for the evaluation of solvent effects in MC-SCF calculations. *J. Chem. Phys.* **1999**, *111*, 5295–5302.

[112] Cammi, R.; Mennucci, B.; Tomasi, J. Second-Order Møller-Plesset Analytical Derivatives for the Polarizable Continuum Model Using the Relaxed Density Approach. *J. Phys. Chem. A* **1999**, *103*, 9100–9108.

[113] Tomasi, J.; Mennucci, B.; Cancès, E. The IEF version of the PCM solvation method: an overview of a new method addressed to study molecular solutes at the QM ab initio level. *J. Mol. Struct. THEOCHEM* **1999**, *464*, 211–226.

[114] Cammi, R.; Mennucci, B.; Tomasi, J. Fast Evaluation of Geometries and Properties of Excited Molecules in Solution: A Tamm-Dancoff Model with Application to 4-Dimethylaminobenzonitrile. *J. Phys. Chem. A* **2000**, *104*, 5631–5637.

[115] Cossi, M.; Barone, V. Solvent effect on vertical electronic transitions by the polarizable continuum model. *J. Chem. Phys.* **2000**, *112*, 2427–2435.

[116] Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Polarizable dielectric model of solvation with inclusion of charge penetration effects. *J. Chem. Phys.* **2001**, *114*, 5691–5701.

[117] Cossi, M.; Barone, V. Time-dependent density functional theory for molecules in liquid solutions. *J. Chem. Phys.* **2001**, *115*, 4708–4717.

[118] Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.* **2002**, *117*, 43–54.

[119] Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681.

[120] Cammi, R. Coupled-cluster theories for the polarizable continuum model. II. Analytical gradients for excited states of molecular solutes by the equation of motion coupled-cluster method. *Int. J. Quantum Chem.* **2010**, *110*, 3040–3052.

[121] Scalmani, G.; Frisch, M. J. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **2010**, *132*, 114110.

[122] Lipparini, F.; Scalmani, G.; Mennucci, B.; Cancès, E.; Caricato, M.; Frisch, M. J. A variational formulation of the polarizable continuum model. *J. Chem. Phys.* **2010**, *133*, 014106.

[123] Caricato, M. Absorption and Emission Spectra of Solvated Molecules with the EOM–CCSD–PCM Method. *J. Chem. Theory Comput.* **2012**, *8*, 4494–4502.

[124] Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges. *J. Chem. Theory Comput.* **2007**, *3*, 2011–2033.

[125] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

[126] Cramer, C. J.; Truhlar, D. G. General Parameterized SCF Model for Free Energies of Solvation in Aqueous Solution. *J. Am. Chem. Soc.* **1991**, *113*, 8305–8311.

[127] Giesen, D. J.; Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. In *Computational Thermochemistry*; Irikura, K., Frurip, D. J., Eds.; ACS, Washington, D.C., 1998; Vol. 677; pp 285–300.

[128] Hawkins, G. D.; Zhu, T.; Li, J.; Chambers, C. C.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. In *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M. A., Eds.; ACS, Washington, D.C., 1998; Vol. 712; pp 201–219.

[129] Cramer, C. J.; Truhlar, D. G. In *Trends and Perspectives in Modern Computational Science*; Maroulis, G., Simos, T. E., Eds.; Lecture Series on Computer and Computational Sciences, Leiden, 2006; Vol. 6; pp 112–140.

[130] Poltev, V. In *Handbook of Computational Chemistry*, 2nd ed.; Leszczynski, J., Kaczmarek-Kedziera, A., Puzyn, T., Papadopoulos, M. G., Reis, H., Shukla, M. K., Eds.; Springer Nature, Switzerland, 2017; pp 21–67.

[131] Pauling, L.; Corey, R. B.; Branson, H. R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Nat. Acad. Sci.* **1951**, *37*, 205–211.

[132] Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738.

[133] Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

[134] Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parametrization and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.

[135] Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520–552.

[136] Halgren, T. A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553–586.

[137] Halgren, T. A. Merck molecular force field. IV. Conformational energies and geometries for MMFF94. *J. Comput. Chem.* **1996**, *17*, 587–615.

[138] Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data and empirical rules. *J. Comput. Chem.* **1996**, *17*, 616–641.

[139] Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720–729.

[140] Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* **1999**, *20*, 730–748.

[141] Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. Conformations of cycloheptadecane. A comparison of methods for conformational searching. *J. Am. Chem. Soc.* **1990**, *112*, 1419–1427.

[142] Gilbert, K. E. PCMODEL (Version 10.0). Serena Software, www.serenasoft.com, 2020.

[143] Gilbert, K. E. GMMX. Serena Software, www.serenasoft.com, 2020.

[144] Dennington, R.; Keith, T. A.; Millam, J. M. GaussView Version 6. 2019; Semichem Inc. Shawnee Mission KS.

[145] Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.

[146] Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.

[147] Coutsias, E. A.; Seok, C.; Dill, K. A. Using quaternions to calculate RMSD. *J. Comput. Chem.* **2004**, *25*, 1849–1857.

[148] Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Nat. Acad. Sci.* **2002**, *99*, 12562–12566.

[149] Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826–843.

[150] Wen, E. Z.; Hsieh, M.-J.; Kollman, P. A.; Luo, R. Enhanced ab initio protein folding simulations in Poisson–Boltzmann molecular dynamics with self-guiding forces. *J. Mol. Graph. Model.* **2004**, *22*, 415–424.

[151] Spiwok, V.; Lipovová, P.; Králová, B. Metadynamics in Essential Coordinates: Free Energy Simulation of Conformational Changes. *J. Phys. Chem. B* **2007**, *111*, 3073–3076.

[152] Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148–154.

[153] Vymětal, J.; Vondrášek, J. Metadynamics As a Tool for Mapping the Conformational and Free-Energy Space of Peptides — The Alanine Dipeptide Case Study. *J. Phys. Chem. B* **2010**, *114*, 5632–5642.

[154] Deriu, M. A.; Grasso, G.; Tuszynski, J. A.; Gallo, D.; Morbiducci, U.; Danani, A. Josephin Domain Structural Conformations Explored by Metadynamics in Essential Coordinates. *PLOS Comput. Biol.* **2016**, *12*, 1–14.

[155] Hibbert, D. B. Generation and display of chemical structures by genetic algorithms. *Chemometrics Intell. Lab. Sys.* **1993**, *20*, 35–43.

[156] Leardi, R. Genetic algorithms in chemometrics and chemistry: a review. *J. Chemometrics* **2001**, *15*, 559–569.

[157] Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462–2474.

[158] Berendsen, H. C. *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to fluid Dynamics*; Cambridge University Press, Cambridge, 2007.

[159] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

[160] Valdes, H.; Pluháčková, K.; Pitonák, M.; Řezáč, J.; Hobza, P. Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747–2757.

[161] Bursch, M.; Neugebauer, H.; Grimme, S. Structure Optimisation of Large Transition-Metal Complexes with Extended Tight-Binding Methods. *Angew. Chem. Int. Ed.* **2019**, *58*, 11078–11087.

[162] Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86). *J. Chem. Theo. Comput.* **2017**, *13*, 1989–2009.

[163] Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theo. Comput.* **2019**, *15*, 1652–1671.

[164] Cavasin, A. T.; Hillisch, A.; Uellendahl, F.; Schneckener, S.; Göller, A. H. Reliable and Performant Identification of Low-Energy Conformers in the Gas Phase and Water. *J. Chem. Inf. Model.* **2018**, *58*, 1005–1020.

[165] Sharapa, D. I.; Genaev, A.; Cavallo, L.; Minenkov, Y. A Robust and Cost-Efficient Scheme for Accurate Conformational Energies of Organic Molecules. *ChemPhysChem* **2019**, *20*, 92–102.

[166] Helgaker, T.; Ruud, K.; Bak, K. L.; Jørgensen, P.; Olsen, J. Vibrational Raman optical activity calculations using London atomic orbitals. *Faraday Discuss.* **1994**, *99*, 165–180.

[167] Dukor, R. K.; Nafie, L. A. In *Encyclopedia of Analytical Chemistry: Instrumentation and Applications*; Meyers, R. A., Ed.; Wiley and Sons, Chichester, 2000; pp 662–676.

[168] Ruud, K.; Helgaker, T.; Bouř, P. Gauge-Origin Independent Density-Functional Theory Calculations of Vibrational Raman Optical Activity. *J. Phys. Chem. A* **2002**, *106*, 7448–7455.

[169] Barron, L. D. *Molecular Light Scattering and Optical Activity*, 2nd ed.; Cambridge University Press, 2004.

[170] Thorvaldsen, A. J.; Ruud, K.; Kristensen, K.; Jørgensen, P.; Coriani, S. A density matrix-based quasienergy formulation of the Kohn–Sham density functional response theory using perturbation- and time-dependent basis sets. *J. Chem. Phys.* **2008**, *129*, 214108.

[171] Cheeseman, J. R.; Frisch, M. J. Basis Set Dependence of Vibrational Raman and Raman Optical Activity Intensities. *J. Chem. Theory Comput.* **2011**, *7*, 3323–3334.

[172] Haesler, J.; Hug, W. Raman Optical Activity: A Reliable Chiroptical Technique. *Chimia* **2008**, *62*, 482–488.

[173] Long, D. A. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*; Wiley, 2001.

[174] Hug, W. Visualizing Raman and Raman optical activity generation in polyatomic molecules. *Chem. Phys.* **2001**, *264*, 53–69.

[175] Hug, W. *Handbook of vibrational spectroscopy*; John Wiley & Sons, 2006; pp 745–758.

[176] Buckingham, A. D. Permanent and Induced Molecular Moments and Long-Range Intermolecular Forces. *Adv. Chem. Phys.* **1967**, *12*, 107.

[177] Buckingham, A. D. The theoretical background to vibrational optical-activity - Introductory lecture. *Faraday Discuss.* **1994**, *99*, 1–12.

[178] Frisch, M. J. et al. Gaussian 16 Revision A.03. 2016; Gaussian Inc. Wallingford CT.

[179] Liégeois, V. DrawSpectrum. UNamur, www.unamur.be/drawspectrum, 2021.

[180] Liégeois, V. DrawMol. UNamur, www.unamur.be/drawmol, 2021.

[181] Liégeois, V. DrawVib. UNamur, www.unamur.be/drawvib, 2021.

[182] Zuber, G.; Hug, W. Rarefied basis sets for the calculation of optical tensors. 1. The importance of gradients on hydrogen atoms for the Raman scattering tensor. *J. Phys. Chem. A* **2004**, *108*, 2108–2118.

[183] Feller, D. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* **1996**, *17*, 1571–1586.

[184] Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis Set Exchange : A Community Database for Computational Sciences. *J. Chem. Inf. Model.* **2007**, *47*, 1045–1052.

[185] Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.* **2019**, *59*, 4814–4820.

[186] Peterson, K. A.; Yousaf, K. E. Molecular core-valence correlation effects involving the post-d elements Ga–Rn: Benchmarks and new pseudopotential-based correlation consistent basis sets. *J. Chem. Phys.* **2010**, *133*, 174116.

[187] Weigend, F.; Baldes, A. Segmented contracted basis sets for one- and two-component Dirac–Fock effective core potentials. *J. Chem. Phys.* **2010**, *133*, 174102.

[188] Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[189] Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, *133*, 134105.

[190] Roy, L. E.; Hay, P. J.; Martin, R. L. Revised Basis Sets for the LANL Effective Core Potentials. *J. Chem. Theory Comput.* **2008**, *4*, 1029–1031.

[191] Labello, N. P.; Ferreira, A. M.; Kurtz, H. A. Correlated, relativistic, and basis set limit molecular polarizability calculations to evaluate an augmented effective core potential basis set. *Int. J. Quantum Chem.* **2006**, *106*, 3140–3148.

[192] Lamparska, E.; Liégeois, V.; Quinet, O.; Champagne, B. Theoretical determination of the vibrational raman optical activity signatures of helical polypropylene chains. *ChemPhysChem* **2006**, *7*, 2366–2376.

[193] Liégeois, V.; Champagne, B. Vibrational Raman optical activity of pi-conjugated helical systems: hexahelicene and heterohelicenes. *J. Comput. Chem.* **2009**, *30*, 1261–1278.

[194] Šebestík, J.; Bouř, P. Raman Optical Activity of Methyloxirane Gas and Liquid. *J. Phys. Chem. Lett.* **2011**, *2*, 498–502.

[195] Scott, A. P.; Radom, L. Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Møller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *J. Phys. Chem.* **1996**, *100*, 16502–16513.

[196] Merrick, J.; Moran, D.; Radom, L. An evaluation of harmonic vibrational frequency scale factors. *J. Phys. Chem. A* **2007**, *111*, 11683–11700.

[197] Reichman, S.; Schreiner, F. Gas-Phase Structure of XeF_2. *J. Chem. Phys.* **1969**, *51*, 2355–2358.

[198] Levy, H. A.; Agron, P. A. The Crystal and Molecular Structure of Xenon Difluoride by Neutron Diffraction. *J. Am. Chem. Soc.* **1963**, *85*, 241–242.

[199] Burns, J. H.; Agron, P. A.; Levy, H. A. Xenon Tetrafluoride Molecule and Its Thermal Motion: A Neutron Diffraction Study. *Science* **1963**, *139*, 1208–1209.

[200] Burbank, R. D.; Jones, G. R. Structure of the cubic phase of xenon hexafluoride at 193.deg.K. *J. Am. Chem. Soc.* **1974**, *96*, 43–48.

[201] Malm, J. G.; Selig, H.; Jortner, J.; Rice, S. A. The Chemistry Of Xenon. *Chem. Rev.* **1965**, *65*, 199–236.