

EOMT

# **Óscar's wonderful end of master's thesis**

Óscar Iglesias González

June 7, 2020

CC Ediciones, Oscar's Room HQ

### **Disclaimer**

You can edit this page to suit your needs. For instance, here we have a no copyright statement, a colophon and some other information. This page is based on the corresponding page of Ken Arroyo Ohori's thesis, with minimal changes.

### **No copyright**

©② This book is released into the public domain using the CC0 code. To the extent possible under law, I waive all copyright and related or neighbouring rights to this work. To view a copy of the CC0 code, visit:

<http://creativecommons.org/publicdomain/zero/1.0/>

### **Colophon**

This document was typeset with the help of KOMA-Script and L<sup>A</sup>T<sub>E</sub>X using the kaobook class.

### **Publisher**

First printed in May 2019 by CC Ediciones, Oscar's Room HQ

What, that's impossible? Wait, how? How is this sunflower so chill. How, that's impossible, he can't be this chill. He is so chill, I can't believe it.

– Unknown



# Preface

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

# Contents

<b>Contents</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Marine toxins and saxitoxin . . . . .	1
1.2 The work up until now . . . . .	2
1.3 Pushing onwards . . . . .	3
<b>2 Computational methods and specifications</b>	<b>5</b>
2.1 Methods and techniques . . . . .	5
Calculation level . . . . .	5
Geometry optimization . . . . .	6
Magnetic shielding computation . . . . .	6
Electronic transition study . . . . .	6
Vibrational analysis . . . . .	6
Raman spectroscopy . . . . .	7
Spectra envelope calculation . . . . .	8
2.2 Software . . . . .	10
2.3 Hardware . . . . .	10
<b>3 The sunflower molecules</b>	<b>13</b>
3.1 Introduction and origin . . . . .	13
3.2 Sunflower design . . . . .	13
3.3 Study of geometry . . . . .	14
3.4 Study of stability . . . . .	16
Ring strain . . . . .	16
3.5 Study of aromaticity . . . . .	17
Nucleus-Independent Chemical Shift (NICS) . . . . .	18
3.6 Spectroscopic characterization . . . . .	22
Vibrational spectroscopy . . . . .	22
Electronic spectroscopy . . . . .	22
<b>4 Sunflower-saxitoxin complexes</b>	<b>25</b>
4.1 Spectroscopic study of lone STX . . . . .	25
Vibrational spectroscopy . . . . .	25
Electronic spectroscopy . . . . .	26
4.2 Study of adsorption . . . . .	27
Sampling and optimization . . . . .	27
Maxwell-Boltzmann statistics . . . . .	28
Basis Set Superposition Error correction . . . . .	28
4.3 Study of non-covalent interactions . . . . .	29
4.4 Study of UV-vis behavior . . . . .	30
General UV-vis spectroscopy . . . . .	30
Charge transfer analysis . . . . .	31
4.5 Resonance Raman . . . . .	32
Generation and comparison of spectra . . . . .	32
Final selection . . . . .	35

<b>APPENDIX</b>	<b>37</b>
<b>A Additional graphs and spectra</b>	<b>39</b>
A.1 NICS of all sunflowers . . . . .	40
A.2 Raman spectra of all sunflowers . . . . .	43
A.3 UV-vis spectra of all sunflowers . . . . .	44
A.4 Combined enhancement factor graphs . . . . .	47
A.5 Combined resonance Raman spectra . . . . .	49
<b>Bibliography</b>	<b>51</b>
<b>Notation</b>	<b>53</b>

# List of Figures

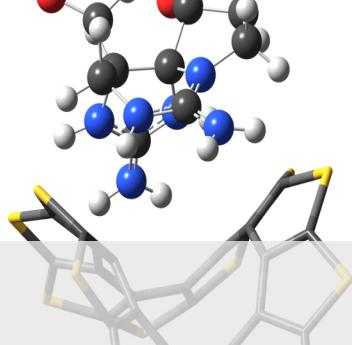
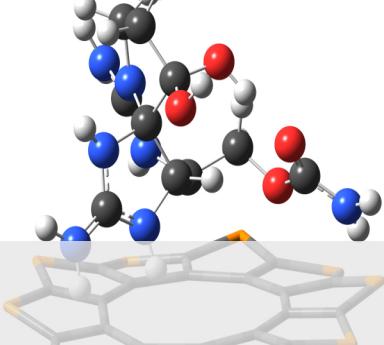
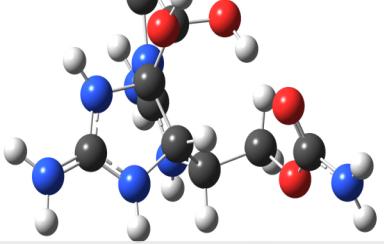
2.1 Raman spectrum as simple peaks . . . . .	9
2.2 Raman spectrum with gaussian envelope . . . . .	9
2.3 Electronic spectrum with gaussian envelopes . . . . .	10
3.1 Structure of sunflower . . . . .	13
3.2 Strain of thiophenic circulenanes . . . . .	13
3.3 General structures of the sunflower family . . . . .	14
3.4 Thiophene-like structure template . . . . .	14
3.5 Shape of the sunflowers . . . . .	14
3.6 Homodesmotic reaction used to calculate strain energies . . . . .	17
3.7 Strain energies of sunflower groups . . . . .	17
3.8 NICS applied to sunflower . . . . .	18
3.9 Linear combination of base 3D functions . . . . .	21
3.11 NICS applied to Se10 . . . . .	21
3.10 NICS applied to test systems . . . . .	21
3.12 UV-vis spectrum of S08 . . . . .	23
4.1 Structure of STX . . . . .	25
4.2 Raman spectrum of lone STX . . . . .	26
4.3 UV-vis spectrum of lone STX . . . . .	27
4.4 Conformers of STX-S08 . . . . .	27
4.5 Counterpoise corrected interaction energies . . . . .	29
4.6 UV-vis spectrum of S08-STX . . . . .	30
4.7 UV-vis absorption spectra of all complexes . . . . .	31
4.8 Combined resonance graph for As12-STX . . . . .	35
A.1 Part 1 of NICS 2D projections . . . . .	41
A.2 Part 2 of NICS 2D projections . . . . .	42
A.3 Part 1 of flower UV-vis spectra . . . . .	44
A.4 Part 2 of flower UV-vis spectra . . . . .	45
A.5 Part 3 of flower UV-vis spectra . . . . .	46
A.6 Part 1 of combined EF RR graphs . . . . .	47
A.7 Part 2 of combined EF RR graphs . . . . .	48

# List of Tables

2.1 Raman activity of STX . . . . .	8
2.2 Overview of techniques, level and software . . . . .	10
3.1 Sunflowers in this study . . . . .	14
3.2 Bond length study . . . . .	16
3.3 xyz coordinates . . . . .	19
3.4 Calculated predictors . . . . .	20

3.5 UV-vis absorption range of isolated flowers . . . . .	23
4.1 Raman modes of STX . . . . .	26
4.2 Energies of STX-S08 conformers . . . . .	27
4.3 Maxwell-Boltzmann populations of STX-S08 . . . . .	28
4.4 Maxwell-Boltzmann populations for all sets . . . . .	29
4.5 UV absorption range of selected complexes . . . . .	31
4.6 Classification of individual vibrations . . . . .	33





# 1 Introduction

This work is about molecules.

## 1.1 Marine toxins and saxitoxin

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language.

the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 1.2 The work up until now

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift –

not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

### 1.3 Pushing onwards

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

```

137     self.bonds = []
138     for i in range(len(self.atoms)):
139         for j in range(i + 1, len(self.atoms)):
140             if i != j:
141                 d = np.linalg.norm(self.coords[i, :] - self.coords[j, :])
142                 if MIN_BOND_LENGTH <= d <= MAX_BOND_LENGTH:
143                     self.bonds.append({'atoms': (i, j),
144                                       'elements': (self.atoms[i],
145                                                    self.atoms[j]),
146                                       'distance': d})
147
148     def find_axis(self):
149         self.main_axis = best_fitted_plane(self.coords)

```

## 2 Computational methods and specifications

### 2.1 Methods and techniques

Throughout this work, many different computational methods and techniques were employed in order to carry out a diverse set of studies. Some of them, the ones that are the most relevant to the flow of the work, are specifically explained and developed in detail in their own sections. Others, however, are deemed to be more of a general character, already well known within the field, or unimportant for the understanding of the main ideas of this thesis. This section is meant to serve as an overview of the methodology, specifications and theory behind the whole study, and also to provide explanations for that less special set of techniques. This part may be skimmed and consulted at a later time, or even skipped entirely, as the rest of the work is presented in such a way that can be followed without deep knowledge of the subjects that are explained here.

#### Calculation level

All of the electronic structure calculations were performed using Density Functional Theory (DFT) in the Kohn-Sham formulation. Specifically, the functional of choice was Minnesota's M06-2X.<sup>[1]</sup> M06-2X is a highly parametrized hybrid meta-GGA functional that features a 54 % of Hartree-Fock exchange. It was chosen because it has been extensively trained to perform well in a variety of contexts that include thermochemistry, the study of non covalent interactions, and vibrational and electronic spectroscopy. While the molecules discussed in the work are novel and have not been previously characterized, it's expected that their chemistry and their spectroscopic behavior aren't out of the ordinary. Therefore, it should be possible to describe them properly and efficiently with such a functional.

As for the basis functions, the set of choice was Weigend and Ahlrichs' def2-SVP.<sup>[2]</sup> def2-SVP is of split valence set and includes polarization functions, but is overall a fairly small set. However, it has been deemed sufficiently extense to obtain accurate geometry optimizations, and qualitatively good energies and spectra.

Considering the size of the systems of the study, the large amount of planned computations, the available computational resources, and the actual accuracy needs of the project, the M06-2X and def2-SVP calculation level was considered appropriate.

[1]: Zhao et al. (2008), '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'

[2]: Weigend et al. (2005), 'Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy'

## Geometry optimization

The optimization of the geometry of a molecule is a crucial part of any computational modeling. At its core, it's a process where the geometry of a system gets iteratively modified (and its energy gets calculated at each step) with the aim of reaching a stationary point on its potential energy surface. In the case of this work, these optimizations are always minimizations as we just look for energy minima. In Gaussian, our electronic structure computation software of choice, these calculations are carried out using the Berny algorithm in its GEDIIS[3] implementation.

[3]: Li et al. (2006), 'Energy-represented DIIS within a hybrid geometry optimization method'

[4]: Magyarfalvi et al. (2011), 'Shielding Calculations: GIAO Methods'

[5]: Yüksel et al. (2008), 'Preparation, GIAO NMR Calculations and Acidic Properties of Some Novel 4,5-dihydro-1H-1,2,4-triazol-5-one Derivatives with Their Antioxidant Activities'

[6]: Runge et al. (1984), 'Density-Functional Theory for Time-Dependent Systems'

[7]: Adamo et al. (2013), 'The calculations of excited-state properties with Time-Dependent Density Functional Theory'

[8]: Ochterski (1999), *Vibrational Analysis in Gaussian*

## Magnetic shielding computation

Nuclear magnetic resonance type calculations, namely the computation of the magnetic shielding in the Nucleus-Independent Chemical Shift study in Section 3.5, were carried out using the GIAO method. GIAO stands for gauge-including atomic orbital. It solves what is often referred to as "the gauge problem", which can be defined as an error that arises when doing calculations with a magnetic perturbation while having the atomic orbital basis functions depend on position.[4] Magnetic perturbations usually affect the atomic orbital set of a molecule as rotations. Atomic orbitals located near to the axis of the rotation won't suffer from this error, as their basis sets should still be able to properly describe the perturbed wave function. However, as the distance to the axis of rotation increases, the linear translation due to a rotation gets larger, and the description of the atomic orbitals gets progressively worse. The GIAO method solves this problem by using sets of atomic basis functions that depend explicitly on the magnetic field.

GIAO calculations have been successful in describing the magnetic shielding of a variety of large nuclei in optimized and isolated large molecules.[5] Therefore, it has been considered as an appropriate way of computing the absolute magnetic shieldings needed in this work.

## Electronic transition study

The prediction of ultraviolet-visible (UV-vis) spectra, a result of electronic transitions within a molecule, requires the computation of the energies of its electronically excited excited states. As the electromagnetic waves responsible for such excitations have a time-dependent nature, in this work this is achieved through the application of time-dependent DFT.[6, 7]

## Vibrational analysis

Any given non linear molecule with  $N$  atoms has  $6$  translational and rotational normal modes, and  $3N - 6$  vibrational ones. Vibrational normal modes are orthogonal, that is, they're vibrational motions that are independent and don't cause movement to the other normal modes.

The frequencies of these modes are calculated through the following procedure.[8] First, a Hello, here is some text without a meaning. This text

should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

### Normal mode calculation and decomposition

By translating the cartesian displacements that result from the frequency calculations into redundant internal coordinates, it's possible to identify which of the atoms of the molecule or system are the most involved in a particular vibrational mode. Having the vibrations expressed in such a way is useful in automating the analysis of modes. In this work, this is used to classify the vibrational modes of a dimer. Each individual vibration within the mode is classified by whether it belongs to molecule A, to molecule B, or is a mixture of the two. Thanks to this calculation, the modes can be filtered by their percentage of A, B or mixed vibrations. This idea is further explored and applied in Section 2.

### Raman spectroscopy

Raman spectroscopy is a technique that allows for the detection of those vibrational normal modes that affect the polarizability of a molecule (i.e. the ease with which its electron cloud can be distorted). This kind of vibrational transitions are tied to a physical phenomenon known as Raman or inelastic scattering. For an inelastic scattering event to happen, a photon has to excite a molecule and get it to a virtual energy state before being emitted. After this, the photon has either a lower or a higher energy, and the molecule ends up in a different vibrational or rotational state due to the energy exchange between them. **Explain how Raman modes are identified and how are their intensities calculated.**

#### Resonance Raman

The regular Raman effect is usually very weak, as the probability of Raman scattering occurring is extremely low.<sup>1</sup>

Resonance Raman (RR) spectroscopy is a variation of Raman in which the intensities are greatly amplified. This technique relies in the usage of a carefully selected laser wavelength to perform the measurement:

1: Only about 1 in every  $10 \times 10^7$  photons results in inelastic scattering

if it's close enough to an electronic transition, the scattering effect, and therefore the intensity of the measured Raman peaks, can increase by several orders of magnitude. In regular Raman scattering, the energy that the molecule absorbs from the photon just makes it reach a virtual energy level before returning to a different rotational or vibrational state. By using photons that have the same energy as electronic transitions, however, the molecule reaches a real excited electronic state. This excited state has a different geometry, and that affects the polarizability of the molecule and increases the magnitude of the subsequent Raman signal.[9]

[9]: Hirakawa et al. (1975), 'Molecular Geometry in an Excited Electronic State and a Preresonance Raman Effect'

**Explain how it's calculated using CPHF!**

### Surface-enhanced Raman spectroscopy and surface interactions

Surface-enhanced Raman spectroscopy (SERS) is a variation of Raman spectroscopy that explores the large enhancements in intensity that the Raman signals of a molecule may experience when interacting with a surface. While such effect was originally observed in roughened silver and is usually studied using metal surfaces and clusters, it has also been found in non-metal nanostructures and surface-like molecules, which will be referred to as "SERS substrates". The nature of this effect has been related to resonance. Specifically, it's been said that it draws upon three different kinds: surface plasmon resonance, charge-transfer resonance, and molecular resonance.[10] The first of these effects, surface plasmon resonance, is defined as the coherence of the oscillation of the conduction electrons of the substrate with an external exciting electromagnetic radiation. Charge-transfer resonance may be present when there is a significant transfer of electron density between the substrate and the molecule. Finally, molecular resonance may appear too as a property almost exclusive to the molecule, but can still have a significant contribution to the total effect.

The main focus of this work, however, is Resonance Raman. The adsorption of the molecule to a substrate makes it benefit from SERS effects, but it shares importance with its other purpose: shifting the UV-vis absorption range. Adhering a molecule that absorbs short wavelengths to a substrate that absorbs long ones may result in a complex that has a range of absorption that's higher than that of the molecule alone. Since real-life laser devices for Raman spectroscopy can get rare, expensive and energetic to the point of being destructive as the wavelength gets shorter, it's desired to achieve such effects. For these reasons, while SERS is an interesting technique that certainly plays a part in the results achieved in this work, we don't delve too deep into the explanation of its nature. **Also it's really hard to actually separate and interpret the contributions. :(**

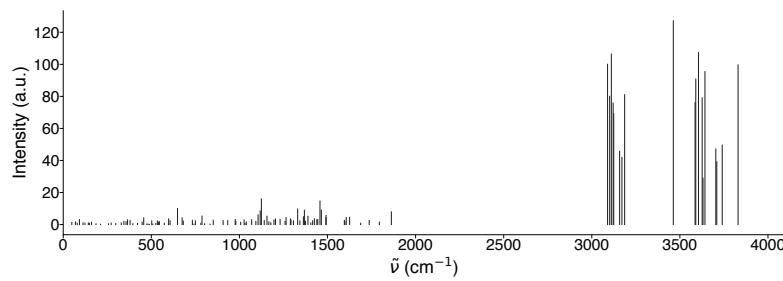
[10]: Lombardi et al. (2008), 'A Unified Approach to Surface-Enhanced Raman Spectroscopy'. Such as the free flowing electrons in a metal cluster, or the electron cloud of a large conjugated covalent system such as graphene

**Table 2.1:** Raman activity for each vibrational mode of STX

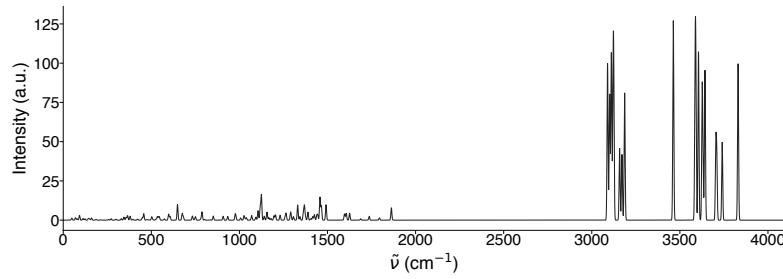
Mode	$\tilde{\nu}$ cm <sup>-1</sup>	Ram. act.
1	0.0917	0.0236
2	0.9542	1.9687
3	0.8552	1.2691
4	1.7897	3.3270
	...	

### Spectra envelope calculation

This work deals with two types of spectra: electronic and vibrational. Computational chemistry software is able to predict spectroscopic values such as electronic transition energies, vibrational mode wave numbers, Raman activities and oscillator strengths. However, to get from those numbers to the familiar bands and peaks that are characteristic of experimental spectra, a few extra steps are involved. This additional



**Figure 2.1:** Raman spectrum of STX using a simple peak representation



**Figure 2.2:** Raman spectrum of STX using an envelope of gaussian functions

representation procedure will be exemplified by plotting both kinds of spectra for the saxitoxin (STX) molecule.

We start from the list of vibrational normal modes that results from a Raman calculation. Then, their vibrational frequencies in the form of wave numbers as well as their corresponding Raman activity values are extracted using a custom script. Plotting these values directly as vertical lines along the wave number range results in Figure 2.1.

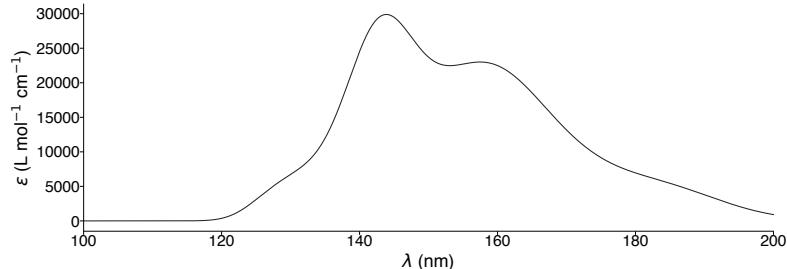
While such a graph could still be useful to display and compare the intensities of the vibrational modes, it could be harder to compare to real experimental spectra. In order to make our theoretical spectra look more realistic, then, an envelope line is added. This is done by replacing each of the simple vertical lines by gaussian functions. These functions are designed so that they are as tall as the intensity of the peak that they're representing. They are evaluated at the full range of frequencies, and linearly combined forming a continuous curve that smoothly wraps all of the peaks.

In the case of Raman spectra, the value of each of the gaussians at a certain wave number value is calculated following Equation 2.1.

$$I_i(\tilde{\nu}) = I_i^{\max} e^{-\left(\frac{\tilde{\nu}-\tilde{\nu}_i}{\sigma}\right)^2} \quad (2.1)$$

The value of  $\sigma$ , which is the full width at half maximum of each curve, is set arbitrarily at  $4\text{ cm}^{-1}$ . The same STX Raman spectrum, plotted after computing this set of equations, is displayed in Figure 2.2.

As for the electronic spectra, the procedure is quite similar. In this case, the values to be plotted are the energies of the electronic transitions in the form of wavelengths, as well as their corresponding oscillator strengths. Similarly to before, an envelope curve is created by calculating as many



**Figure 2.3:** Electronic spectrum of STX using an envelope of gaussian functions

gaussian functions as there are transitions, but with a slightly different formula as displayed in Equation 2.2.

$$\varepsilon_i(\lambda) = 1.306\,297 \times 10^8 \frac{f_i}{\sigma} e^{-\left(\frac{1/\lambda - 1/\lambda_i}{\sigma}\right)^2} \quad (2.2)$$

Here,  $\lambda_i$  and  $f_i$  are the wavelength and oscillator strength of each transition. The constant before the exponential (see Equation 2.3) is a conversion factor that is added in order to achieve the right units for the intensity  $\varepsilon$ , which should be L mol<sup>-1</sup> cm<sup>-1</sup>. A value of 0.4 eV was chosen for  $\sigma$ . For the case of STX, the spectrum that results from combining the gaussian band shapes for the 50 most intense electronic transitions is shown in Figure 2.3

$$\frac{\sqrt{\pi}e^2N}{1000\ln(10)c^2m_e} = 1.306\,297 \times 10^8 \quad (2.3)$$

Where  $e$  and  $m_e$  are the charge and mass of the electron,  $N$  is Avogadro's number, and  $c$  is the speed of light

## 2.2 Software

The software used to perform all of the calculations is summarized in Table 2.2

**Table 2.2:** Overview of the computational techniques, calculation level and software that were used in this work

Calculation	Technique	Spec.	Functional	Basis set	Software
Geometry optimization	DFT	GEDIIS	M06-2X	def2SVP	Gaussian
Vibrational analysis	DFT		M06-2X	def2SVP	Gaussian
Raman activity	DFT		M06-2X	def2SVP	Gaussian
Resonance Raman activity	DFT	CPHF	M06-2X	def2SVP	Gaussian
Electronic transition calculation	TD-DFT		M06-2X	def2SVP	Gaussian
Magnetic shielding calculation	DFT	GIAO	b3lyp	6-31G*	Gaussian
Surface generation					nics.py

## 2.3 Hardware

All of the electronic structure calculations were performed using either the Centro de Supercomputación de Galicia's (CESGA) infrastructures, or the proprietary cluster of the S3 research group.

CESGA's supercomputer, FinisTerra-II (FTII), is a Bull ATOS bullx machines that features 320 computation nodes, 7712 cores, 44 544 GB of RAM, and 750 000 GB of storage capacity. All of the calculations carried out at FTII were performed in standard nodes, utilizing 12 cores at a time,

and 60 GB of RAM. These nodes include each 2 Intel®Xeon®E5-2630 v3 2.50 GHz processors with 24 cores total, and 128 GB of RAM.

S3's cluster is composed of nodes that feature 16 cores running on Intel®Xeon®E5-2630 v3 2.40 GHz processors, and 64 GB of RAM.

Smaller calculations such as input file generation, output parsing, surface estimation, graph representation... were carried out in the author's personal computer.





## 3 The sunflower molecules

### 3.1 Introduction and origin

The search for effective and interesting SERS substrates ended up leading us to a novel class of molecules researched and presented in the year 2006 by Chernichenko and his colleagues.[11] The first representative of this family, nicknamed as "sulflower", is the ocatathio[8]circulene. This highly symmetric structure, which may be described as a form of carbon sulfide and as a belt of annulated thiophene cycles, is claimed to have great stability, high symmetry and unusual electronic properties.

From a synthetic point of view, sulflower also proved to be simple and straightforward to develop despite its complex appearance: starting from tetra thiophene, sulphurizing its free sites and acidifying to get polythiol, and removing the excess sulfur by vacuum pyrolysis. This process allowed the team to achieve yields of 56% starting from commercially available reagents.

Interestingly, the team proposes that it could be possible to prepare materials with diverse electronic properties by using different types of heteroatoms and varying on the basic structure of the molecule. Such a statement made apparent the potential of this family of molecules: highly symmetrical, stable, surface-like structures with variable electronic behavior could act as suitable SERS substrates. This chapter is entirely dedicated to that premise: the study and characterization of sulflower and sulflower-like molecules, which from now on I will collectively refer to as "sunflowers". By designing, generating and studying our own family of sunflowers, we will be contributing to characterize a novel and interesting group of molecules, and we may be able to identify an ideal SERS environment for STX.

[11]: Chernichenko et al. (2006), "Sulflower": A New Form of Carbon Sulfide'

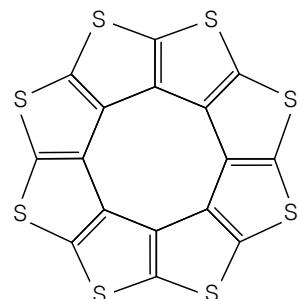


Figure 3.1: Structure of sulflower

### 3.2 Sunflower design

To start designing a family of molecules, first we must clearly state their defining pattern. For this purpose, we adopt Chernichenko et al.'s own proposal: "a novel class of heterolytic circulenes". We start expanding the model by answering the question: are thiophene based circulenes with other than 8 rings stable enough to be worth considering? The answer is in the original paper itself. Figure 3.2, which was recreated to use our calculation level and adapt to the style of the document, shows that 8 ring structures are the most stable alongside 9. The details about this calculation are further explained in Section 3.4. However, considering their low relative energies and the fact that they have an even number of electrons (which would greatly simplify later calculations), 10 and 12 ring sunflowers were also chosen as part of the study.

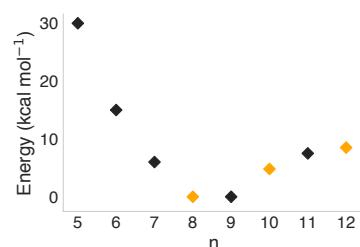
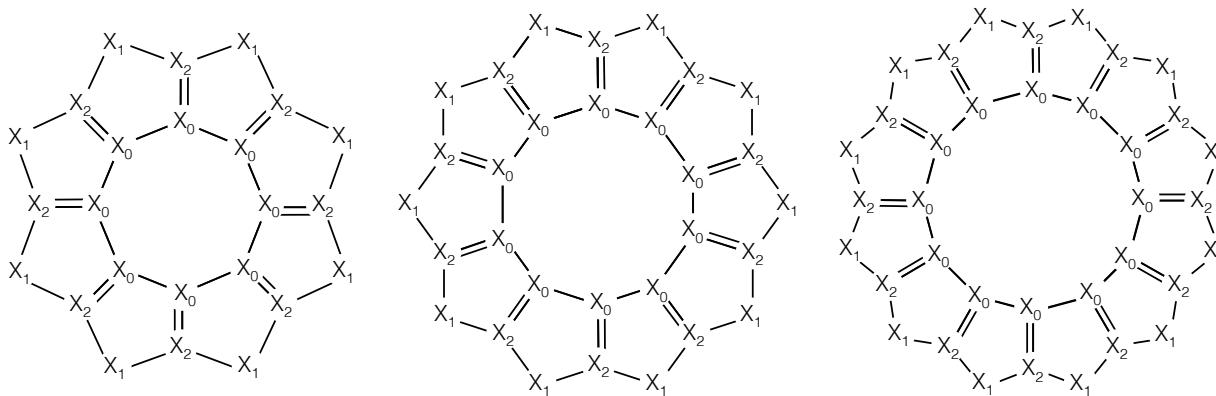


Figure 3.2: Strain of thiophenic circulenes with n rings



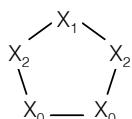
**Figure 3.3:** From left to right, general structures of the 8, 10 and 12 ring sunflowers

**Table 3.1:** Subset of the sunflower family that is going to be studied

$X_0$	$X_1$	$X_2$	number of rings		
			8	10	12
C	S	C	S08	S10	S12
C	Se	C	Se08	Se10	Se12
C	As	C	As08	As10	As12
C	As	N	AsN08	AsN10	AsN12
C	P	C	P08	P10	P12
C	P	N	PN08	PN10	PN12

Expanding upon this idea to allow for further heteroatom substitution, we ended up with the templates in Figure 3.3. The possibilities were numerous, but we settled for S, Se, As and P substitutions on  $X_1$  sites, and N substitutions in some cases in  $X_2$  sites. A full table detailing all of the structures that were generated and will comprise this study can be found in Table 3.1, as well as the short names or IDs that were given to each species based on its composition and number of petals for the sake of abbreviation.

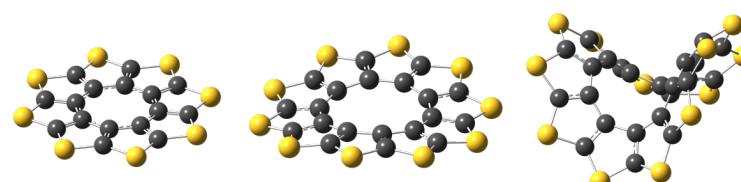
### 3.3 Study of geometry



**Figure 3.4:** Thiophene-like structure template (hydrogen is added to adjust for neutrality as needed)

Coordinate files for all of the designed sunflowers were created using 3D molecule modeling software. Then, they were optimized at the M06-2X/def2SVP calculation level.

Using the output files for these calculations, all of the bond distances in all of the systems were extracted using Python. Then, the bonds were grouped by type: inner (I, bonds between  $X_0$  atoms, which were carbon in all cases), middle (M, between  $X_0$  and  $X_2$  atoms, which were either C-C or C-N bonds), and outer (O, between  $X_2$  and  $X_1$  atoms). The mean and standard deviation (SD) of these groups was computed for all of the



**Figure 3.5:** Shape of optimized S08, S10 and S12, from left to right (note the progressive increase in their deviation from planarity)

flowers. Additionally, thiophene-like pentagonal structures of the form displayed in Figure 3.4 were modeled and optimized, and their bond length data was compared with the rest and presented in Table 3.2.

How is this information useful? In the first place, it allows us to easily compare the average length of each bond type as the number of units increases. More interestingly, it serves as a way to assess the nature of each bond group. Take the inner C-C bonds. In all cases, the average length lies between the usual values for single and double C-C bonds,<sup>1</sup> an information that immediately suggests us that the flowers might be conjugated systems with  $\pi$  electron delocalization. However, it could also be possible that there were equal amounts of single and double bonds, and that these apparently conjugated bond lengths were just a result of a lousy statistical approximation. That's why the SD metric was also computed. Groups with a relatively high SD such as the I bonds of As08, As12, P08 and P12 are actually composed of longer and shorter bonds, and it's likely that their  $\pi$  electrons aren't delocalized. On the other hand, groups with small SD values are more likely to correspond to conjugated systems, which may be stabilized by electron delocalization. This concept of bond length equalization and conjugation may be tied to that of aromaticity, where equalized C-C bonds could be a sign of it, and uneven ones could indicate non aromaticity or antiaromaticity.<sup>[12]</sup> This topic will be covered in greater detail in Section 3.5.

1: 1.54 Å for single and 1.34 Å for double bonds [add reference](#)

[12]: Chen et al. (2005), 'Nucleus-Independent Chemical Shifts (NICS) as an Aromaticity Criterion'

**Table 3.2:** Bond length statistics for each studied species, units are Å, and  $\tilde{x}$  and  $\sigma$  are the mean and the standard deviation

			I ( $X_0-X_0$ )		M ( $X_0-X_2$ )		O ( $X_1-X_2$ )	
			$\tilde{x}$	$\sigma$	$\tilde{x}$	$\sigma$	$\tilde{x}$	$\sigma$
S	08	ring	1.429	-	1.367	0.000	1.719	0.000
		08	1.423	0.000	1.377	0.000	1.754	0.000
		10	1.482	0.000	1.399	0.000	1.712	0.000
		12	1.466	0.004	1.389	0.000	1.732	0.004
Se	08	ring	1.434	-	1.363	0.000	1.857	0.000
		08	1.454	0.000	1.385	0.000	1.865	0.000
		10	1.478	0.003	1.389	0.000	1.861	0.004
		12	1.466	0.007	1.381	0.000	1.877	0.005
As	08	ring	1.468	-	1.354	0.000	1.918	0.000
		08	1.434	0.050	1.448	0.000	1.859	0.008
		10	1.436	0.000	1.454	0.001	1.856	0.001
		12	1.435	0.042	1.446	0.002	1.867	0.007
AsN	08	ring	1.495	-	1.283	0.000	1.854	0.000
		08	1.434	0.000	1.361	0.000	1.858	0.000
		10	1.450	0.001	1.364	0.001	1.856	0.004
		12	1.440	0.004	1.357	0.001	1.873	0.006
P	08	ring	1.467	-	1.359	0.000	1.794	0.001
		08	1.410	0.048	1.445	0.000	1.764	0.006
		10	1.431	0.000	1.471	0.000	1.736	0.000
		12	1.429	0.045	1.462	0.002	1.747	0.005
PN	08	ring	1.485	-	1.293	0.000	1.707	0.000
		08	1.404	0.000	1.364	0.000	1.752	0.000
		10	1.448	0.000	1.385	0.000	1.715	0.001
		12	1.435	0.002	1.378	0.001	1.732	0.004

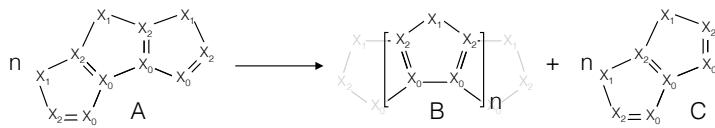
### 3.4 Study of stability

Using the output files from the previous step, specifically those from the initial optimization, calculations related to the energies of the flowers were performed in order to assess their stabilities.

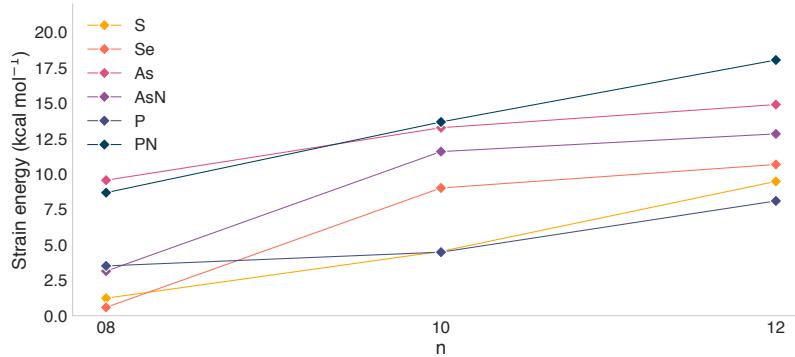
#### Ring strain

Ring strain may be defined as a kind of instability that arises when angles of the bonds in a cyclic molecule deviate from their optimal values, which they would be able to adopt if they weren't coiled in the shape of a ring. The energy associated to this phenomenon can serve as a way to compare the stabilities of flowers with different numbers of petals, as it was done in the original paper and reproduced in Figure 3.2. Its calculation relies on the design of an homodesmotic reaction. What does this mean? In a homodesmotic reaction, the number of atoms and the type of hybridizations in the products are the same as in the reactants. Our goal is to devise a chemical equation where the strained structure appears in only one of its sides and the rest of the chosen molecules don't present any stray effects.<sup>[13]</sup> If we manage to create such an equation, the reaction

[13]: Vidal-Vidal et al. (2005), 'Nitrogen Doped Nanohoops as Promising CO<sub>2</sub> Capturing Devices'



**Figure 3.6:** Homodesmotic reaction used to calculate strain energies



**Figure 3.7:** Strain energies of all of the studied sunflower groups

energy will correspond to the strain, as the rest of the components of the energy should nullify themselves between the two sides of the reaction.

In this case, apart from the sunflower structure B, the equations that were designed included molecules A and C, non-cyclic structures composed of 3 and 2 petals. By following this reaction, which is fully displayed in Figure 3.6, the strain energy was calculated as shown in Equation 3.1.

This calculation, which is the one that was used to generate the previous sunflower stability curve Figure 3.2, was applied to all of the sunflowers of the study, and the resulting strain energies are presented in Figure 3.7. As it can be seen, the strain energies always increase with the number of petals (as it's the case for the original sunflower series). However, it may be noted that they increase following slightly different tendencies. Notable examples are the P family, where the 8 and 10 petal species have very similar strains; and the Se, As and AsN families, where the energies of the 10 and 12 petal structures are very close. These energy values are an useful indication of the relative strains of the different species within a family, and an estimation of their stabilities with respect to the original sunflower. Nevertheless, it should be kept in mind that this calculation is an approximation, and that the values of the so called strain energy might be accounting for other effects such as conjugation stabilization or repulsion between the heteroatoms.

$$E_{\text{strain}} = \frac{E_B + nE_C - nE_A}{n} \quad (3.1)$$

## 3.5 Study of aromaticity

Aromaticity is a property related to molecules that contain a ring or a chain of resonance bonds that increases their stability. It's typically found in flat ring structures, although the definition has been discussed and expanded since its initial association to benzene,<sup>[14]</sup> and there are other varieties such as annulenic, azulenic, inorganic, homoaromatic, three-dimensional,  $\sigma$ -electron, and even metallic electron aromaticity.<sup>[12]</sup> Considering their ringed structure and possibly high conjugation, we wondered if the sunflowers could present any kind of aromatic tendencies. Having already studied their degree of bond length equalization in

[14]: Faraday (1825), 'On new compounds of carbon and hydrogen, and on certain other products obtained during the decomposition of oil by heat'

[12]: Chen, et al. (2005), 'Nucleus-Independent Chemical Shifts (NICS) as an Aromaticity Criterion'

Section 3.3, we set out to assess their aromaticity from the point of view of their magnetic properties.

## Nucleus-Independent Chemical Shift (NICS)

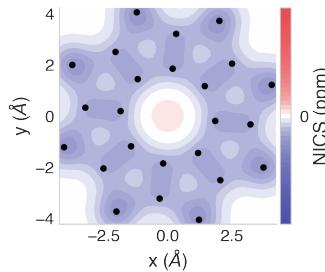
[15]: Ragué Schleyer et al. (1996), ‘Nucleus-Independent Chemical Shifts: A Simple and Efficient Aromaticity Probe’

Since its introduction in 1996 by Schleyer et al.[15], NICS has become the most popular and widely used technique to estimate aromaticity in a quantitative way. It’s based on the study of electronic ring currents. As the electrons in the possibly-aromatic systems have a certain degree of free circulation, an external magnetic field perpendicular to the main plane of the system is able to induce a ring current. Said ring currents generate their own magnetic field, which can weaken or strengthen the effect of the external field, resulting in decreased or increased NMR chemical shifts. Aromatic systems will experience shielding on the inside of the ring and deshielding on the outside. Antiaromatic systems will experience the opposite.

### The basic application of NICS

NICS is usually applied to single rings, and quantified by computing the absolute magnetic shielding at key locations. This key location is usually one of the following: the center of the ring (calculated as the non-weighted mean of the positions of the heavy atoms), several points along a central axis perpendicular to the plane of the ring (calculated trivially in planar rings as the plane that passes through any three ring atoms), at several points on a grid perpendicular to the plane of the ring, or as three dimensional isosurfaces using dense grids.

### A custom solution for non-planar molecules



**Figure 3.8:** Custom NICS technique applied to the original 8 petal sunflower molecule as a sample of the legacy approach

As seen in Figure 3.5 in Section 3.3, 8 petal sunflowers are perfectly planar. In previous studies, NICS calculations on flat grids parallel to their planes were applied with success. All of the points of the grid were equidistant to the plane of the molecule to be able to properly compare their values (the magnitude of the magnetic shielding decreases the farther away from the molecule that it is measured). The positive values of the magnetic shielding were plotted in red, and the negative ones in blue, which overlaid over the atoms of the flower, resulted in graphs like Figure 3.8. However, 10 and 12 petal sunflowers have bended and warped shapes, and the “general plane” of the molecule is no longer a plane. In order to translate the old method, a new approach was developed and applied. The idea of extending a grid of points over the surface of the system and plotting it as a 2D projection is maintained, but what is exactly the surface of a non planar molecule, and how can it be generalized for non planar molecules?

Defining the concept of “general surface” for a warped ring is a tricky task, but we’ve set on the following description: a 3D surface that is equidistant to all of the heavy atoms of a molecule and to all of the middle points of their bonds, and that follows a smooth and intuitive tendency in the places where are neither atoms nor bonds. A general surface could be loosely thought of as a blanket that was being draped over a physical

molecular model of the system, sitting on the places where there are balls and sticks, and covering the empty spaces smoothly without caving in.

To mathematically model such an idea, several approaches were evaluated and tested,<sup>2</sup> but in the end, the method was based on linearly combining a set of simple 3D cartesian functions.

This idea was born while thinking about the curvature of the largest sunflowers. "This Se10 molecule looks like it has the shape of a saddle."<sup>3</sup> What if we could find the precise parameters for a saddle that adjusted to the shape of this molecule?" This concept was expanded to include a wider variety of molecules and shapes: there's a limit to the angles and deformations that a ring in a molecule can display... so therefore, it should be possible to get a decent approximation to any molecular surfaces by combining and fine tuning a selection of common 3D shapes in the form of functions. The remaining question is... how can we find which 3D functions should be used, and what coefficients should be assigned to them? The functions were picked out by hand, and will later be enumerated and justified. The coefficients, however, had to be calculated somehow. Our solution, which will be fully explained hereunder, relies upon the core concepts of statistical learning.

First off, we start with just the coordinate file of the optimized structure of the studied molecule. We will use Se10 as an example. This file should contain the position of each of the atoms in the 3D cartesian space relative to the origin of coordinates (i.e., an xyz file). In order to stabilize the calculations and apply the method properly, the molecule has to undergo two transformations: it has to be translated so that its center coincides with the origin of coordinates (0, 0, 0); and it has to be rotated so that its main axis aligns with z, or (0, 0, 1).

The first one is trivial, the centroid of the ring is calculated as the mean of all of the positions of the atoms, and is then subtracted from them.

The second one, the rotation, is not as straightforward. First, the main axis of the molecule has to be calculated. In this case, the main axis is defined as the direction where there is the least amount of variance. Mathematically, it's equal to the eigenvector that corresponds to the smallest eigenvalue of the covariance matrix of the relative coordinates.<sup>4</sup> After having calculated such a vector, it's a matter of calculating the angle that it forms with the z axis using simple dot product, and designing and applying an adequate rotation matrix.

Se10, as well as all of the other sunflowers of the study, only contains heavy atoms that are relevant to the problem, so we can just ignore their chemical elements and treat every point equally. However, it should be noted that in other cases there might be irrelevant atoms such as hydrogens, or parts of the molecule that don't belong to the surface that we are trying to identify. In those situations, special care should be taken: the transformation operations have to be designed around the important atoms, but they should be applied to the whole system.

At this point, the relevant part of the molecule (i.e. the ring object of study), should be centered and aligned with the z axis. All of these preparatory steps ensure the optimal application of the actual method.

2: Such as using interpolation algorithms to calculate the grid points in the empty spaces, or generating compound surfaces by tiling all of the possible planes based on groups of three atoms.

3: The equation of a basic saddle is essentially  $z = x^2 - y^2$ .

**Table 3.3:** xyz coordinates of Se10

x	y	z
0.9154	0.0917	0.0236
-0.0971	0.9542	1.9687
1.0978	0.8552	1.2691
1.8126	1.7897	3.3270
		...

4: That is, the coordinates after having been translated to (0, 0, 0) in the previous step

$$f_i \approx \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad i = 1, \dots, n \quad (3.2)$$

The method itself consists in the optimization of the coefficients of a multiple linear regression model. A linear regression model with  $p$  predictors and  $n$  data points, as shown in Equation 3.2, consists of several elements.  $y$  is the target feature, the one that we wish to predict. The various  $x$  are the predictors, variables on which we base our predictions. The various  $\beta$  values are the coefficients that accompany the predictors and that we want to approximate.

In our case in particular, the bits of real data that we have -our ground truth- are the coordinates of the atoms of the molecule, which include  $x$ ,  $y$  and  $z$  values. Looking at it from the point of view of statistical learning, the problem could be defined as “designing a model that predicts the values of  $z$  by using the values of  $x$  and  $y$ ”. If such a model was developed, we could just input a square grid of  $(x, y)$  values and get the height (or  $z$  value) that they should be placed at to conform to our concept of general surface. There is just a problem with this approach: using just  $x$  and  $y$  by themselves as the predictors is not enough to generate the kind of warped, wavy surfaces that we’re looking for! Just by adding different proportions of  $x$  and  $y$ , the best one could get would be a tilted plane.

That is where the 3D surface selection approach gets mixed in. We can combine the  $x$  and  $y$  values in certain ways in order to generate a wider variety of basic shapes to linearly combine, resulting in a better set of predictors. For this problem, the following set of predictors was chosen:  $x^2$  and  $y^2$  (half cylinders, which result in a saddle when subtracted and a dome when added),  $xy$  (a saddle aligned with the diagonal  $x = y$ ), and  $\sqrt{x^2 + y^2}$  (a cone).<sup>5</sup>

Having defined the elements of the model, it’s a matter of transforming the  $x$  and  $y$  values into their function forms, initializing a linear model, training it with these predictors and the values of the target variable  $z$ , and retrieving the coefficients. This is all done using the convenient scikit-learn,[16] a machine learning framework for the Python language. Without delving deep in how their lineal model’s inner workings, it can be said that it relies on two key parts: a cost function that measures its deviation from the target, and an algorithm that minimizes it by iteratively modifying its coefficients. When the model has been trained, its accuracy is measured by computing the values of the  $R^2$  metric for two sets of points: the original coordinates of the atoms (i.e., the data points that the model is based on), and the middle points of each of the bonds (i.e. a collection of new data that the model hadn’t previously known about).

All of this operations have been combined and encapsulated inside a custom Python command line application available in the author’s GitHub account[17].

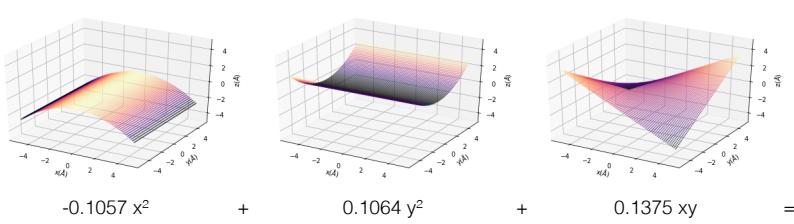
Continuing with the running example, the output of the adjustment of Se10 would be as follows.

```
$ python nics.py surface se10.xyz
```

```
Linear regression results:
x^2: -0.1056911
y^2: 0.1064271
xy: 0.1375246
sqrt(x^2 + y^2): -0.0081293
```

[17]: Iglesias (2020), *nics.py: a command line application to help perform and interpret non planar NICS calculations in Gaussian*

5: It should be noted that this is not proper statistical learning practice: in a real application the predictors shouldn’t have any correlation between them. Pedregosa et al. (2011), ‘Scikit-learn: Machine Learning in Python’



**Figure 3.9:** Linear combination of base 3D functions to model the surface of Se10

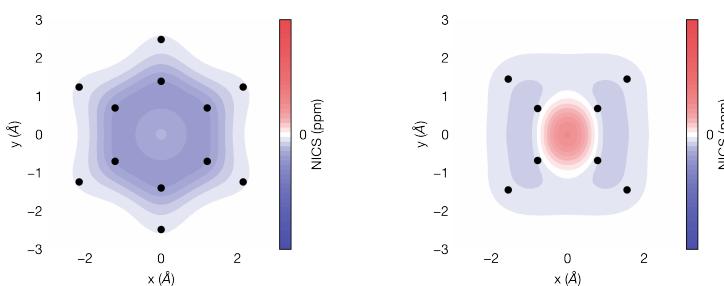
R2 score: 0.978916

R2 score of atom midpoints: 0.988095

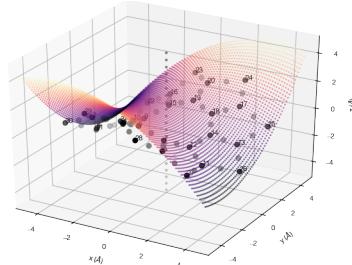
As a way of clearly displaying the result, the idea behind the method, and the elements of the linear combination, the functions have been plotted individually (omitting the last one due to its low contribution) in Figure 3.9.

After the surface has been defined, the results of the fitting are used to calculate a square grid of equally spaced points that are all equidistant to this molecule's surface. In this case, a distance of 1 Å is set by general agreement in the literature. Then, both the molecule atoms and the grid points are written into a Gaussian input file (the former as dummy atoms), where the values of NICS will be calculated. The calculation, specifically, estimates the values of the magnetic shielding are calculated with the Gauge-Independent Atomic Orbital (GIAO)[18] method.

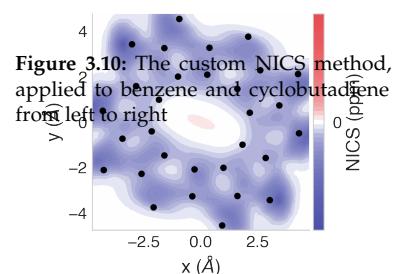
When the calculation is finished, the results can be visualized in 3D and plotted using the same program. As for the interpretations of these results, it was said that through this technique aromatic compounds will experience shielding -negative values, marked in blue- on the inside of the ring and deshielding -positive values, marked in red- on the outside. On the other hand, antiaromatic systems will experience the opposite effect. This is the expected behavior and should be the basis for the evaluation of the technique, so to further verify it and illustrate it, the method has been applied to the well known model systems of benzene and cyclobutadiene. Their 2D NICS graphs are included in Figure 3.10.



The outcome for the Se10 calculation, plotted as a 2D projection onto the  $xy$  plane, is displayed in Figure 3.11. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you



[18]: Keith et al. (1993), 'Calculation of magnetic response properties using a continuous set of gauge transformations'



**Figure 3.10:** The custom NICS method, applied to benzene and cyclobutadiene from left to right

**Figure 3.11:** Custom NICS technique applied to the Se10 system

information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 3.6 Spectroscopic characterization

After having obtained a general profile of the nature of this novel family of molecules, we start delving into the main topic of the work: spectroscopy.

### Vibrational spectroscopy

Any non linear molecule with  $N$  atoms has  $3N - 6$  vibrational normal modes

Sunflowers of 8, 10 and 12 petals have 66, 84 and 102 vibrational modes, respectively. Depending on whether these vibrations actively affect the dipole or the polarizability of the flowers, they can be detected using infrared spectroscopy, Raman spectroscopy, or both. In this work we will focus on the vibrational modes that are visible through Raman spectroscopy, as it's the technique of choice of the study. This preference for Raman isn't arbitrary: **Reasons: Raman can be used on water-based samples (such as seawater), and it's the base technique of the amplification methods that we want to study.**

#### Raman spectra

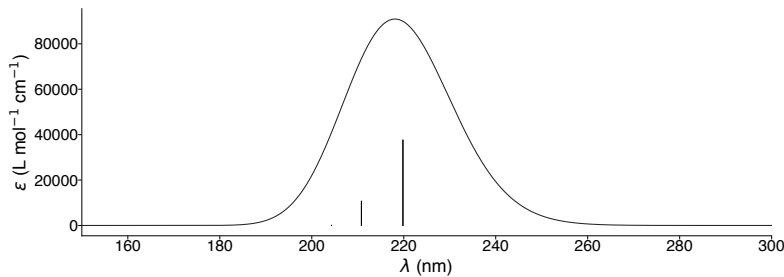
Raman spectra for all 18 sunflower systems were generated following the procedures detailed in Section 1.

### Electronic spectroscopy

Keeping in mind the final objectives of the work of applying resonance Raman, the electronic spectroscopic behavior of the flowers was deemed as an important point to study. The characterization of the electronic transitions of a molecule might give plenty of insight on how does it interact with light.

### UV-vis spectra

The most straightforward way to study electronic spectroscopy is by simulating ultraviolet-visible (UV-vis) spectra. In order to generate this, the energies of the 50 first transitions from the ground state to excited states were computed using TD-DFT in the Gaussian09 suite. Then, using their energy and oscillator strength values, Gaussian functions were calculated for each of the transitions and linearly combined to create continuous spectra similar to those obtained using real spectrophotometers. This procedure is further explained in Section 1. UV-vis spectra were generated for all of the sunflowers, but since they are fairly simple graphs, only the one for S08 will be explained in this section as an example. It's displayed in Figure 4.6.



**Figure 3.12:** UV-vis spectrum of S08

We can see that the spectrum for S08 goes from 190 nm to 260 nm, approximately. Plotted as straight vertical lines, the individual electronic transitions have been included, where their height is proportional to their oscillator strength. However, this is not an adequate representation in this case. What appear to be 2 electronic transitions located at 210 nm and 219 nm are, in fact, 4 different transitions (specifically, there are 2 groups of 2 each): it's just that due to the high symmetry of the molecule, transitions that occur in regions of the molecule that are very electronically similar might have coinciding wavelengths.

This was true for most of the flowers: many of the transitions with the higher intensities were grouped in pairs. The only cases where this effect didn't clearly happen were the sunflowers with the most warped and most bent geometries, such as As12, AsN12, P12 and PN12.

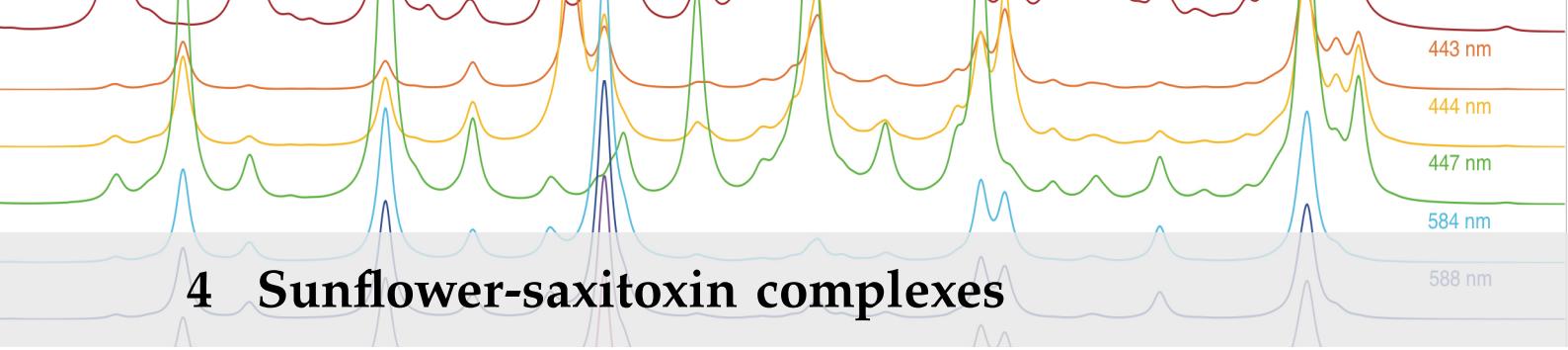
In any case, since the purpose of studying individual electronic transitions is to serve as a guide when applying resonance techniques and we won't study the resonance of the isolated flowers, they were left out of the spectra.

The rest of the UV-vis spectra of the flowers may be found in Section A.3. Their absorption ranges, nonetheless, have been summarized and presented in Table 3.5 as an overview.

**Table 3.5:** Approximate UV-vis absorption range of isolated flowers

System	$\lambda$ (nm)
S08	180-260
S10	180-280
S12	200-400
Se08	200-280
Se10	200-400
Se12	225-500
As08	250-800
As10	250-700
As12	300-900
AsN08	300-450
AsN10	300-650
AsN12	300-850
P08	225-700
P10	250-500
P12	250-700
PN08	250-500
PN10	250-550
PN12	300-700





## 4 Sunflower-saxitoxin complexes

Having obtained a general characterization of the sunflower-type molecules and their spectroscopy, it's time to get back to the problem at hand and start looking into how they can be applied.

Let's reintroduce the molecule that motivated this whole study: saxitoxin (STX). For the purposes of this study, the STX structure features two guanidinium moieties (which are easily susceptible to protonation), two hydroxyl, and one carbamate group as it can be seen in Figure 4.1.

The STX being doubly protonated in the figure is not an arbitrary choice. While studying its acid-base behavior in previous work we faced a certain issue: the STX molecule has many possible protonated variations, and at the pH of real life samples, there would be a coexistence of several of them. This was a problem, because having to apply the study to these different multiple forms in order to account for the situation would greatly increase the number of calculations. After some thought, it was decided that the simplest way to solve this issue would be to work at a pH where only one of the protonated species would be present at a significant amount. It was found that, at the moderate pH value of 6, the majority of the STX could be found as its doubly protonated form. Since in real life experimental conditions attaining a pH of 6 in an hypothetical water-based sample would only imply the addition of a few drops of dilute acid, it was decided that all further studies would be carried out using such diprotonated structure.

### 4.1 Spectroscopic study of lone STX

The goal of this work was presented as "finding suitable substrate to aid in the detection of STX" from the beginning, but this desire came from a place of previous study and understanding about the spectroscopic properties of the lone STX.

This section aims to display and share the most important of these previously known facts, which are about the behavior of saxitoxin in both vibrational and electronic spectroscopy contexts.

#### Vibrational spectroscopy

As a non linear molecule with 40 atoms, STX presents a total of 114 vibrational normal modes. Many of those modes cause changes in its polarizability, and are therefore visible in its Raman spectrum. In this section, we pinpoint and describe a selection of the most characteristic and noticeable vibrational modes that could allow for the identification of the isolated STX.

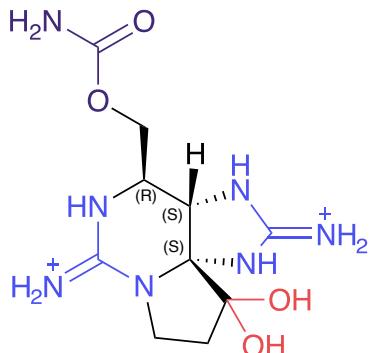
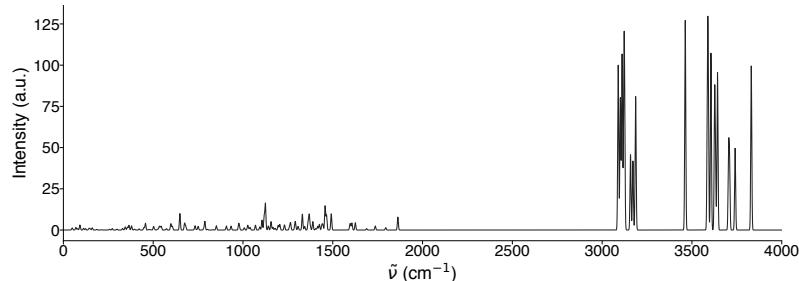


Figure 4.1: Structure of STX



**Figure 4.2:** Raman spectrum of lone STX

**Table 4.1:** Selected Raman active vibrational modes for STX. Letters  $\delta$  and  $\nu$  are scissoring and stretching vibrations, subscripts  $s$  and  $as$  mean symmetric and antisymmetric; and G5, G6 and Cb are the 5 atom guanidinium moiety, the 6 atom guanidinium moiety, and the carbamate group.

Mode	$\tilde{\nu}$	Description
90	1606.2	$\delta_s$ NH <sub>2</sub> G5
91	1625.5	$\delta_s$ NH <sub>2</sub> G5
93	1736.6	$\delta_s$ NH <sub>2</sub> G6
94	1794.6	$\delta_s$ NH <sub>2</sub> G5
95	1862.4	$\delta_s$ NH <sub>2</sub> , $\nu_{as}$ OCN Cb
105	3586.1	$\nu_s$ NH <sub>2</sub> G5
106	3590.2	$\nu_s$ NH <sub>2</sub> G6
107	3605.6	$\nu_s$ NH <sub>2</sub> Cb
111	3704.0	$\nu_{as}$ NH <sub>2</sub> G5
112	3709.4	$\nu_{as}$ NH <sub>2</sub> G6
113	3740.3	$\nu_{as}$ NH <sub>2</sub> Cb

### Raman spectrum

The Raman spectrum of STX was calculated and plotted in the same way as those of the sunflowers.

These modes were deemed as sufficient to identify the lone STX in a sample. This is a good exercise to start identifying characteristic vibrations that could be attributed to STX. However, it must be kept in mind that this analysis only serves as an initial approximation, as it's possible that these same vibrational modes have shifted wave numbers and/or are different when the STX gets adsorbed and forms complexes.

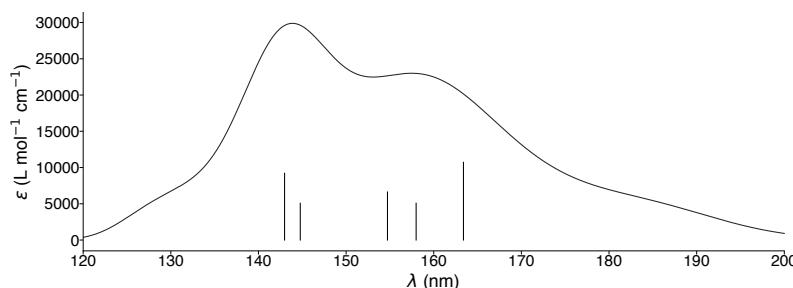
### Electronic spectroscopy

STX doesn't have any special chromophore groups, and its excited states are available at the UV range of electromagnetic radiation. Nevertheless, it's important to identify and study these transitions in order to facilitate the application of further techniques.

### UV-vis spectrum

As before, using TD-DFT as implemented in Gaussian09, the first 50 electronic transitions were calculated and used to generate the UV-vis spectrum of STX, displayed in Figure 4.3.

As it may be noted, the range of absorption (between 120 nm and 200 nm) is located within the UV region. **Introduce Resonance Raman here???**  
**I'm not sure.**



**Figure 4.3:** UV-vis spectrum of lone STX, the 5 electronic transitions with the largest contributions have been annotated

## 4.2 Study of adsorption

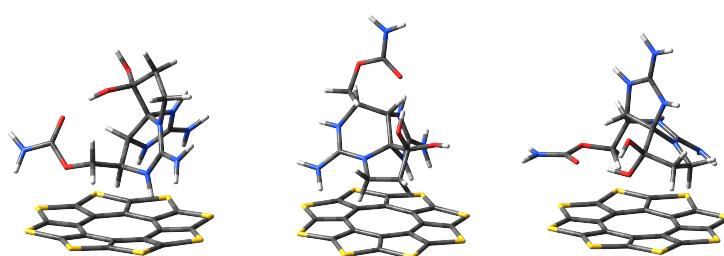
After having characterized the spectroscopic profile of STX, it's time to assess its interactions with the members of the sunflower family, keeping in mind that the main goal of this work is determining if any of them could be a substrate suitable for its detection. The first study that must be carried out, before any spectroscopic technique can be applied, is that of adsorption: how does STX adhere to the flowers, how stable is the resulting complex, what is the nature of that interaction... This is a crucial matter, as there's no point in calculating spectra for a system that isn't stable.

### Sampling and optimization

In order to take into account the possible conformational variability, the STX was manually given an array of different relative positions and angles with respect to the surface of the flowers. Following this idea, 10 different variations were modeled for each of the 18 STX-sunflower pairs, resulting in a total of 180 structures. These various orientations will be referred to as "conformers". All of the conformers for each pairing were optimized at the M06-2X/def2SVP calculation level, and their final energies were compared in order to identify the most stable ones and discard the unstable. Taking the STX-S08 system as an example, the relative energies of all of its generated conformers are displayed in Table 4.2. As it can be seen, the most stable conformer (MSC) is STX-S08-1. However, STX-S08-2, STX-S08-9 and STX-S08-7 are also close energy-wise. A question arises, how can we determine which conformers are stable enough to be important, and which are not?

**Table 4.2:** Relative energies of the STX-S08 conformers, with respect to the most stable one

System ID	Rel. E (kcal mol <sup>-1</sup> )
STX-S08-1	0.000
STX-S08-2	1.337
STX-S08-9	7.133
STX-S08-7	7.504
STX-S08-6	9.896
STX-S08-4	18.554
STX-S08-3	20.321
STX-S08-10	21.734
STX-S08-5	22.529
STX-S08-8	29.715



**Figure 4.4:** Examples of conformers of the STX-S08 system, from left to right, STX-S08-1, STX-S08-9, and STX-S08-3

**Table 4.3:** Maxwell-Boltzmann populations of the STX-S08 conformer set, expressed as percentages

System ID	Population (%)
STX-S08-1	58.56
STX-S08-2	34.16
STX-S08-9	3.30
STX-S08-7	2.83
STX-S08-6	1.08
STX-S08-4	0.03
STX-S08-3	0.02
STX-S08-10	0.01
STX-S08-5	0.01
STX-S08-8	0.00

## Maxwell-Boltzmann statistics

Our answer to this problem consisted in applying Maxwell-Boltzmann statistics to transform these energies into population fractions. This concept essentially translates as the fraction of each conformer that would be present in a macroscopic sample at a certain temperature. As modeled in Equation 4.1,  $p_i$  represents the population fraction of the conformer  $i$ , while  $p_{MSC}$  is the fraction of the MSC of that particular set of conformers. As for the rest of the elements of the equation,  $\varepsilon$  corresponds to the absolute energies of the systems,  $N$  is the total number of conformers in each set (which is 10 in our case),  $k$  is Boltzmann's constant, and  $T$  is the temperature of the system in K (which for the purposes of this study is set at 298 K, or is it 273??).

$$\begin{aligned} \frac{p_i}{p_{MSC}} &= e^{\varepsilon_{MSC} - \varepsilon_i / kT} \\ \sum_{i=1}^N \frac{p_i}{p_{MSC}} &= \frac{\sum_{i=1}^N p_i}{p_{MSC}} = \frac{1}{p_{MSC}} \\ \frac{p_i/p_{MSC}}{1/p_{MSC}} &= \frac{e^{\varepsilon_{MSC} - \varepsilon_i / kT}}{\sum_{i=1}^N \frac{p_i}{p_{MSC}}} = p_i \end{aligned} \quad (4.1)$$

Continuing with the example, the populations for STX-S08 were computed and are displayed in Table 4.3. As an arbitrary threshold, it was decided to filter out all of the conformers with populations lower than 1 %, and to just keep studying the remaining ones. That is, all further calculations that involve the computation of weighted mean values or spectra will only take into account conformers with populations higher than that value.

## Basis Set Superposition Error correction

These optimization calculations have served as a way to estimate the populations of the conformers and to identify the most stable and relevant ones. However, they cannot be used directly to obtain accurate values for the interaction energies due to the Basis Set Superposition Error (BSSE). In this case, this error arises when the STX molecule is close to the sunflower and their basis functions overlap. The part of STX basis functions that comes near the sunflower improves its part of the calculation, and vice versa. This is a problem because in order to get the interaction energy we have to subtract the energy of the complex from the energies of the isolated molecules, but the former has a better calculation level than the latter.

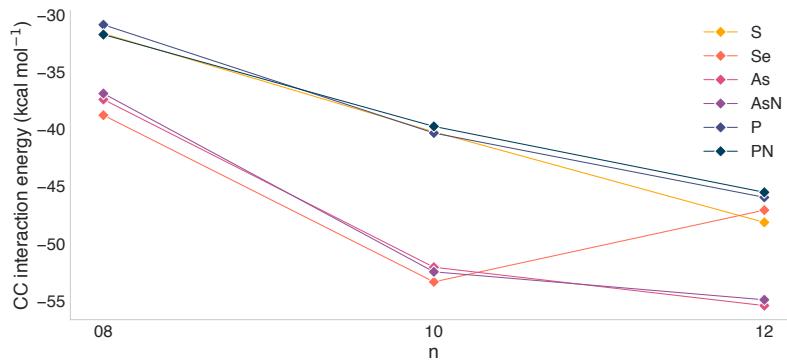
To solve this problem, we used the counterpoise method. For hypothetical molecules A and B, this technique estimates their BSSE by placing the basis functions of molecule A next to molecule B, right where molecule A would go in the optimized geometry of the complex.<sup>[19, 20]</sup> However, its nuclei are omitted, and just the energy of A is calculated using such an extended basis set. The same process is applied to get the corrected energy of isolated B. Finally, the same formula as in Equation 4.2 is

$$V_{STX\text{-}flower} = E_{STX\text{-}flower} - E_{STX} - E_{flower} \quad (4.2)$$

[19]: Duijneveldt et al. (1994), 'State of the Art in Counterpoise Theory'  
[20]: Rösch (2003), *Counterpoise Correction*

System n of petals	S %	Se %	As %	AsN %	P %	PN %
08	58.56	79.24	60.96	81.32	40.79	65.13
	34.16	15.57	35.22	15.98	37.48	29.71
	3.30	1.90	2.19	2.16	20.53	4.61
	2.84	1.15	1.38	0.25	0.44	0.24
	1.08	0.14	0.16	0.23	0.32	0.21
10	38.68	61.74	99.64	100.00	86.97	51.05
	36.07	38.26	0.30	0.00	10.12	46.11
	25.21	0.00	0.00	0.00	2.83	2.26
	0.02	0.00	0.00	0.00	0.06	0.56
	0.01	0.00	0.00	0.00	0.02	0.01
12	99.99	58.93	99.75	97.75	93.29	93.83
	0.01	18.98	0.10	2.23	3.26	5.97
	0.00	14.52	0.08	0.01	2.42	0.20
	0.00	6.03	0.04	0.00	0.74	0.00
	0.00	1.55	0.02	0.00	0.16	0.00

**Table 4.4:** Maxwell-Boltzmann populations for the 5 most stable conformers in all sets, as percentages, with non significant conformers marked in grey



**Figure 4.5:** Counterpoise corrected interaction energies as weighted averages for all of the sets

applied to the new values, which results in the counterpoise corrected (CC) interaction energy.

All such energies were calculated for all of the STX-flower system conformers. Their weighted averages using the population values of Table 4.4 were computed, and are displayed in Figure 4.5.

As it can be noted, all of the systems present negative interaction energies, that is, the complex has a lower energy than the sum of the energies of its constituents. This is a good indication that the systems are all stable and can be further studied. All of the families (except for Se) appear to have a common tendency: the higher the amount of petals, the higher the stabilization. This could be due to the flower having a larger area and bending in a way that maximizes the interactions with the STX molecule.

### 4.3 Study of non-covalent interactions

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference

between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

#### 4.4 Study of UV-vis behavior

One of the main points of adsorbing the STX to the flowers is the desire of increasing its range of absorption of UV-vis radiation. It's possible that the flower-STX complex absorbs light at longer wavelengths than the isolated STX, and this could be highly beneficial to apply further detection techniques based on resonance.

##### General UV-vis spectroscopy

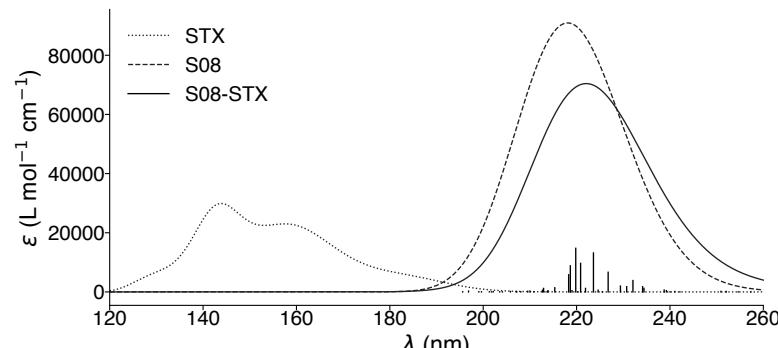
Electronic spectra were calculated as described in Section 2.1 and Section 1 for the most stable conformers of all of the sunflower-STX complexes.

First, they were plotted together with the UV-vis spectrum of STX and of their corresponding isolated flower. The individual transitions with significant weights were also plotted as straight vertical lines. This allowed us to compare the absorption ranges of the complex versus those of the isolated STX and flowers, and to judge if the shifting effect is good enough.

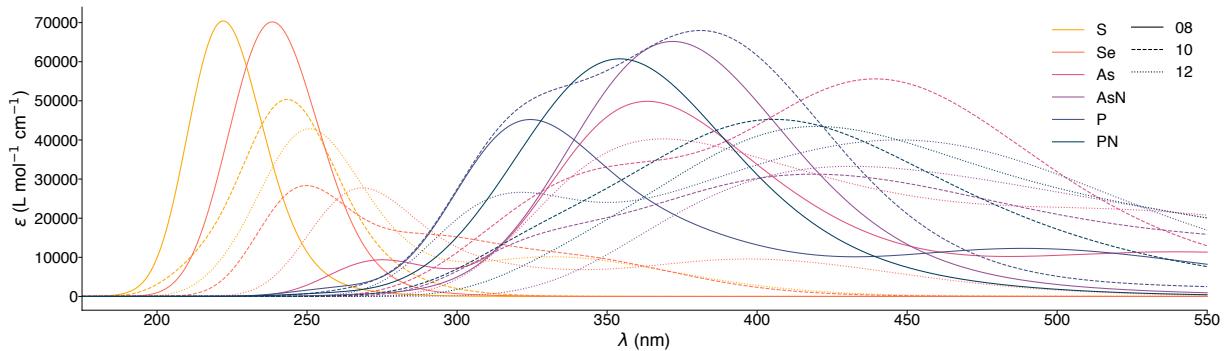
As an example of this, the UV-vis spectra of STX, S08 and S08-STX are displayed in Figure 4.6.

As it may be noted, the spectra of the complex has almost the same absorption range as that of the isolated S08, and in any case, it's definitely shifted towards longer wavelengths in contrast with that of STX. This effect was the same for all of the flower-STX complexes, and the full list of compared UV-vis spectra can be found in Section A.3.

After this, the UV-vis absorption spectra of all of the complexes were plotted together in order to visually compare just their absorption ranges. This plot is displayed in Figure 4.7.



**Figure 4.6:** UV-vis spectra of S08-STX and its components



**Figure 4.7:** UV-vis absorption spectra of all flower-STX complexes

Since this work aims to propose realistic detection techniques, it's important that the systems are actually detectable and identifiable in a real setting. With this purpose in mind, and keeping into account the usual wavelengths of commercially available lasers for Raman,<sup>1</sup> all of the complexes with absorption ranges starting below 250 nm were filtered out. From the remaining ones, a select few were hand-picked according to the upper limit of their ranges and to their specific individual electronic transitions. This left us with 8 systems out of the starting 18, which was deemed as a sufficient amount to continue with the last part of the study.

The UV absorption range for those remaining complexes was also displayed on Table 4.5 for convenience. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

1: Which is between 300 nm and 800 nm

**Table 4.5:** UV absorption range of selected complexes

System	λ (nm)
As10	300-600
As12	300-900
AsN08	300-450
AsN10	300-650
AsN12	350-850
P12	300-700
PN10	300-550
PN12	325-625

## Charge transfer analysis

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 4.5 Resonance Raman

At last, it was time to apply the previous parts of the study and test the performance of the selected sunflowers. A nicer explanation of Resonance raman here? Move the one from Methods? I'm not sure... Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. RR spectra were generated for each of the flower-STX systems by using incident laser wavelengths in their particular ranges of absorption. Specifically, the whole of their absorption interval was covered by selecting wavelengths with a step of 3 nm.<sup>2</sup> Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

2: So in the case of As10, for example, the Resonance raman calculations were performed with lasers of 300, 303, 306... all up until 600 nm

3: Between 186 and 222 depending on the size of the flower

### Generation and comparison of spectra

The RR calculations from the previous step amounted to a total of 2858, where the amplification, position and relevance of each vibrational mode<sup>3</sup> had to be evaluated. For this purpose, the two following metrics were developed.

### Individual molecule contribution to a complex vibrational mode

The first of these evaluations was differentiating which vibrational normal modes involve the vibration of the flower, which modes involve the vibration of the STX, and which modes are mixed. As introduced in Section 2.1, this was done by analyzing the displacements of the vibrations translated into redundant internal coordinates. To understand how this is done, we will use the As12-STX system as an example. The standard Gaussian09 output describes normal modes as cartesian displacements of normal coordinates, indicating vectors for each atom. For instance, the beginning of the output of vibrational normal mode number 42 looks like this:

	42
	A
Frequencies	-- 229.2846
Red. masses	-- 36.2305
Frc consts	-- 1.1222
IR Inten	-- 3.9161

Atom	AN	X	Y	Z
1	33	-0.15	-0.08	0.01
2	33	-0.03	-0.04	-0.04
3	6	0.13	0.06	0.06
4	6	0.06	0.05	0.01
5	33	0.10	-0.01	-0.03
...				

This contains all of the necessary information to study the movements of the atoms and understand the vibrations of the mode, but it's difficult to interpret. By specifying the keyword `intmodes` in the frequency calculation, this output gets translated into the much more readable redundant internal coordinates notation. This is As12-STX's mode number 42 in this new format.

! Normal Mode 42 !				
! Name	Definition	Value	Relative Weight (%)	!
! R1	R(1,34)	-0.0693	0.4	!
! R9	R(4,63)	-0.067	0.4	!
! R15	R(8,9)	0.1332	0.8	!
! A4	A(2,3,30)	-0.0755	0.4	!
! A19	A(10,7,63)	0.0567	0.3	!
! A21	A(5,9,8)	0.1548	0.9	!
! D1	D(36,1,34,24)	0.0965	0.5	!
! D3	D(34,1,36,27)	0.1767	1.0	!
! D5	D(6,2,3,4)	-0.0882	0.5	!
...				

Here the displacements are expressed as bond length extensions between two atoms (R), angle openings and closings between three (A), and dihedral angle torsions between four (D). Their magnitude is expressed as a single positive or negative value, whose absolute value is then weighted and displayed as the Relative Weight as a percent. By differentiating whether the atoms involved in a certain vibrational motion belong to the As12 flower or to the STX, said vibration can be classified as an exclusive flower vibration, as an exclusive STX vibration, or as a mixed one. Such classification, illustrated in Table 4.6, can be easily automated using a script.

By adding up the relative weights of the vibrations in each category, we can find out which of the elements of the flower-STX complex dominate a particular mode. In this case, mode 42 has a 92.54 % contribution from the STX, a 0.00 % contribution from the flower, and a 7.45 % contribution from mixed vibrations.

From a practical point of view, the fact that it's composed almost exclusively of isolated STX vibrations makes this a good choice of vibrational mode to look for in a RR spectra of this complex. In contrast to other mixed or flower-exclusive vibrational modes, a mode like this is a clear sign of the presence of the STX.

**Table 4.6:** Classification of the individual vibrations of normal mode 42 for As12-STX (atoms of the flower and the STX are marked in blue and red, respectively)

Vib. def.	Category
R(1,34)	STX
R(4,63)	Mixed
R(8,9)	STX
A(2,3,30)	STX
A(10,7,63)	Mixed
A(5,9,8)	STX
D(36,1,34,24)	STX
D(34,1,36,27)	STX
D(6,2,3,4)	STX

### Resonance Raman enhancement factor

Even if a mode is exclusive to our molecule of interest, it needs to have a sufficient intensity in the final spectrum: otherwise it cannot be detected and cannot be of any use in the identification. In the context of RR, this means that it has to benefit from a certain level of amplification during resonance experiments.

To evaluate this for all of the tested laser wavelengths, the enhancement factor (EF) metric was designed and applied. The EF for a certain vibrational mode  $i$  and laser wavelength  $\lambda$  is defined in Equation 4.3.

$$EF_{i,\lambda} = \log_{10} \left( \frac{I_{i,\lambda}}{I_{i,SL}} \right) \quad (4.3)$$

In this equation,  $I_{i,\lambda}$  corresponds to the Raman activity of the mode  $i$  in the amplified spectrum, and  $I_{i,SL}$  is the activity of the same mode in a regular Raman prediction. Using a logarithmic scale makes easier to visualize the very high amplifications that arise with this technique, which in the case of this work, have reached values near  $10 \times 10^{10}$  times than the standard Raman.

Similarly as before, this calculation is easy to automate and can serve as a way to filter out wavelengths that don't generate sufficient amplification, or discard modes that aren't sufficiently amplified.

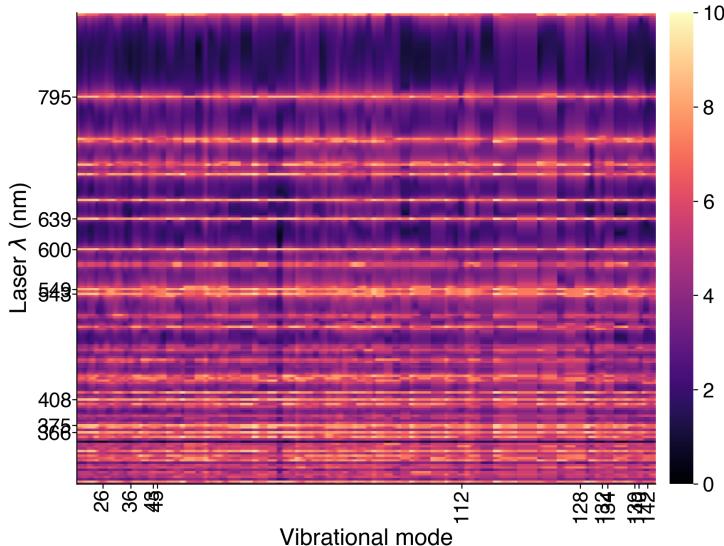
### Combined resonance graphs

In order to make use of the two newly defined metrics, an output file processing pipeline was designed. Large amounts of data from the internal mode decompositions, standard Raman spectra and many amplified RR spectra were processed, combined and analyzed using Python scripts.

This information was filtered and displayed in the following manner: First, Gaussian envelopes for all of the RR spectra as well as the standard Raman spectrum were calculated. These envelopes were equivalent to those used in previous Raman spectra: 10 000 points divided across the most relevant range for each system. Then, each of the points in the amplified RR curves was converted into EF values using Equation 4.3. These were plotted as a colored grid: all of the RR wavelengths on the vertical axis, and the 10 000 calculated wave number values on the horizontal axis. The squares of the grid, therefore, were colored according to the calculated EF values.

This resulted in quite colorful graphs where the brightest horizontal lines correspond to the laser wavelengths that generate larger amplifications, and bright vertical areas correspond to highly amplified vibrational modes.

Then, in order to make full use of the metrics, these graphs were annotated using a special procedure. In regard to the vibrational modes, two thresholds were set: a maximum percent of flower contribution, and a minimum percent of STX contribution. This ensured that the displayed modes were relevant. As for the laser wavelengths, only the ones that



**Figure 4.8:** Combined resonance graph for As12-STX

produced amplifications above a certain threshold value for the EF of the previously selected vibrational modes were displayed.

Applying this method to the running example of As12-STX, then, resulted in Figure 4.8

### Final selection

While the previous method serves as a convenient way to visualize the best amplifications, it's true that it deviates a little from the usual representations of Raman data. This final section aims to tie this insight together, and to display it in a more standard way in order to facilitate the comparison and selection of a definitive combination of flower, laser wavelength, and set of vibrational modes.

For this reason, a complementary way of plotting amplified Raman spectra was designed.

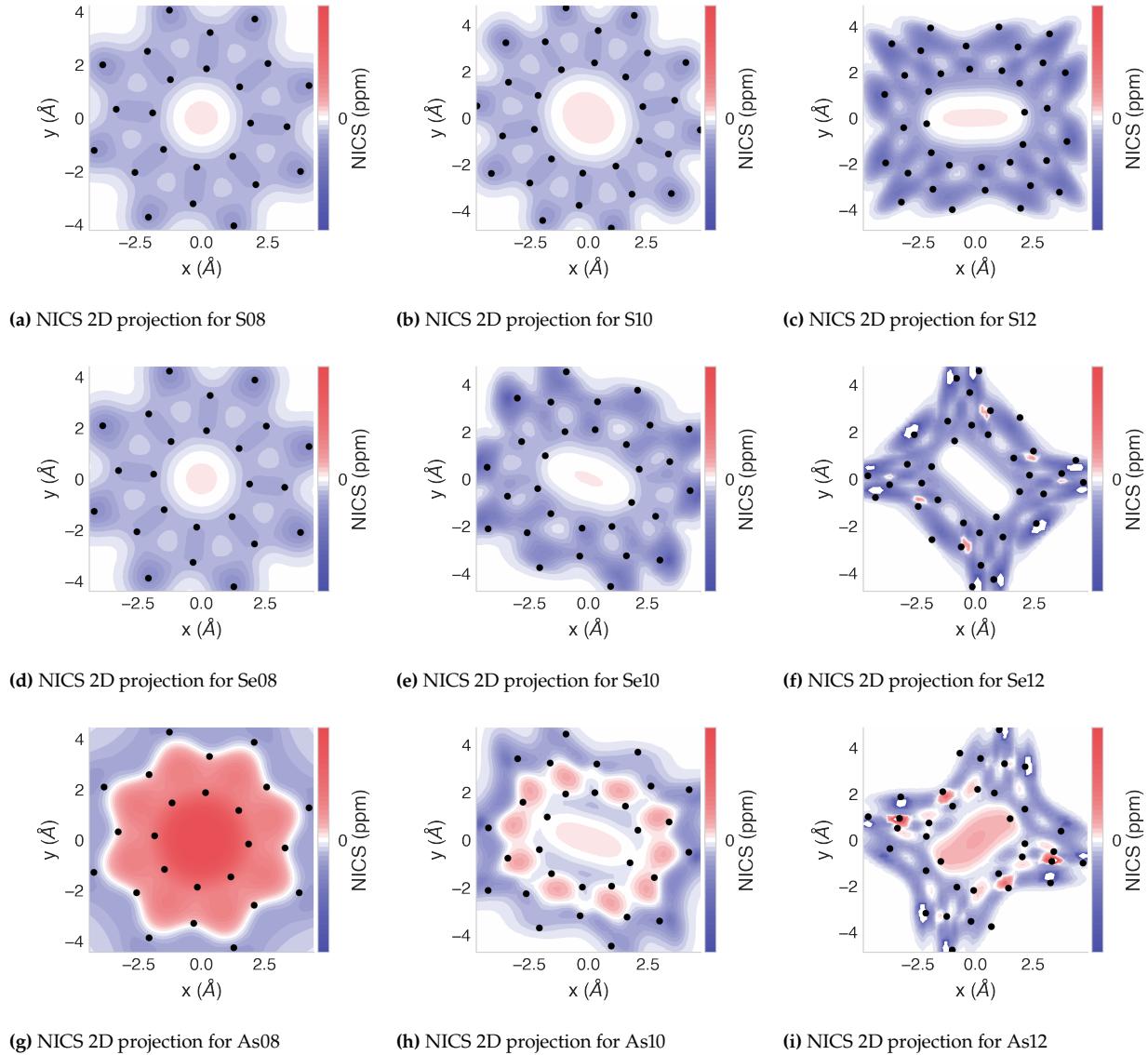


## **APPENDIX**

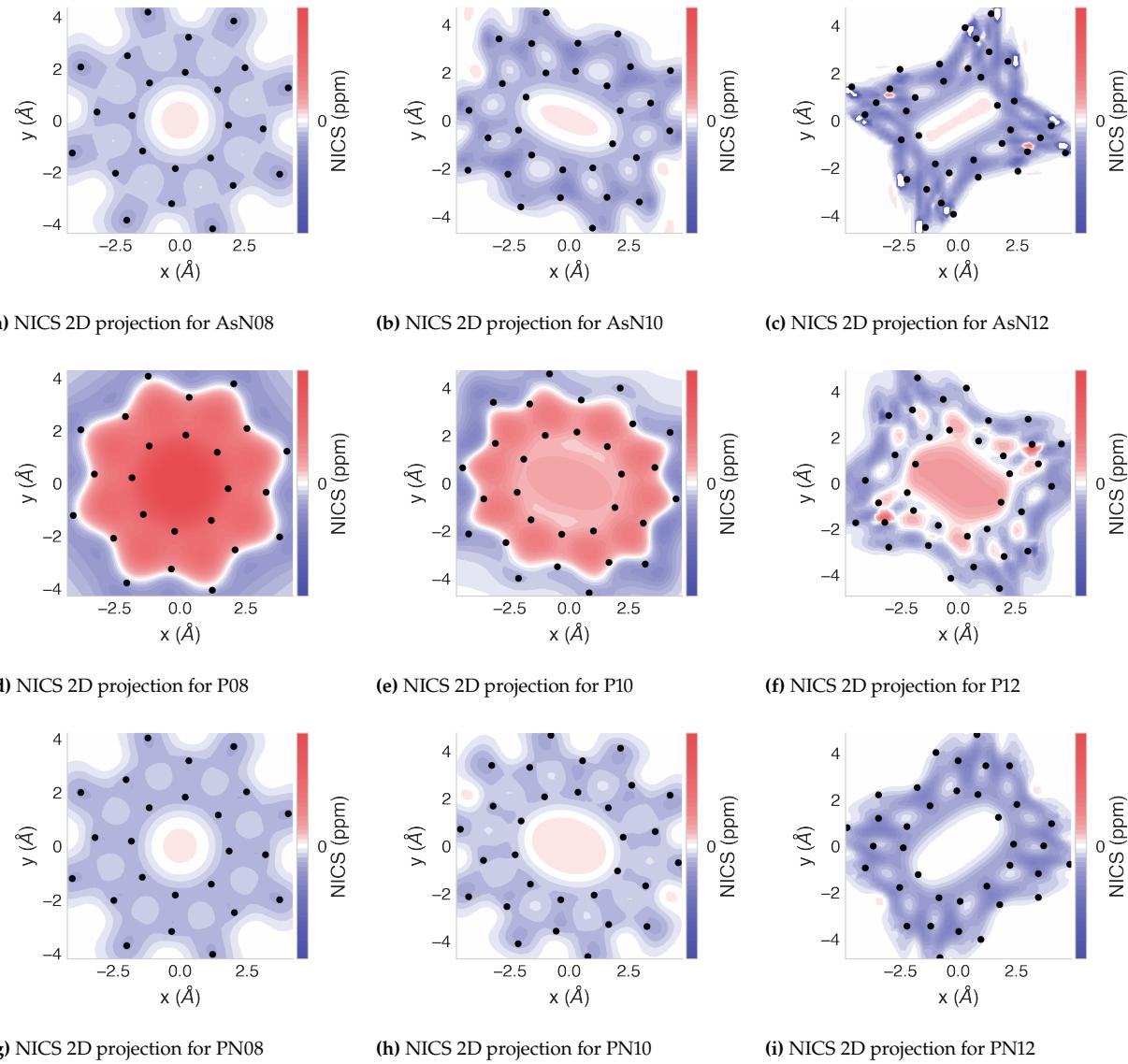


## A Additional graphs and spectra

**A.1 NICS of all sunflowers**



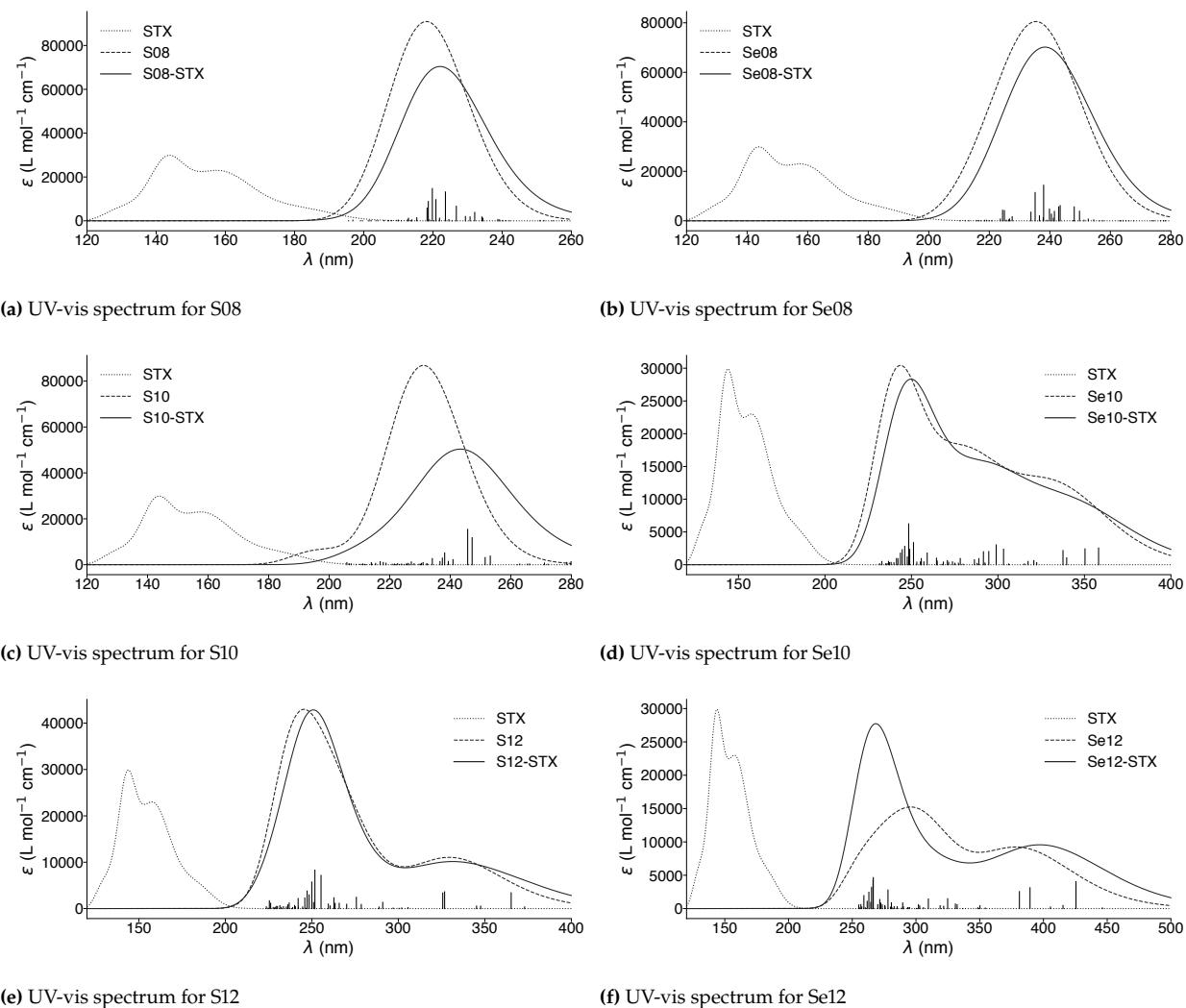
**Figure A.1:** Part 1 of NICS 2D projections



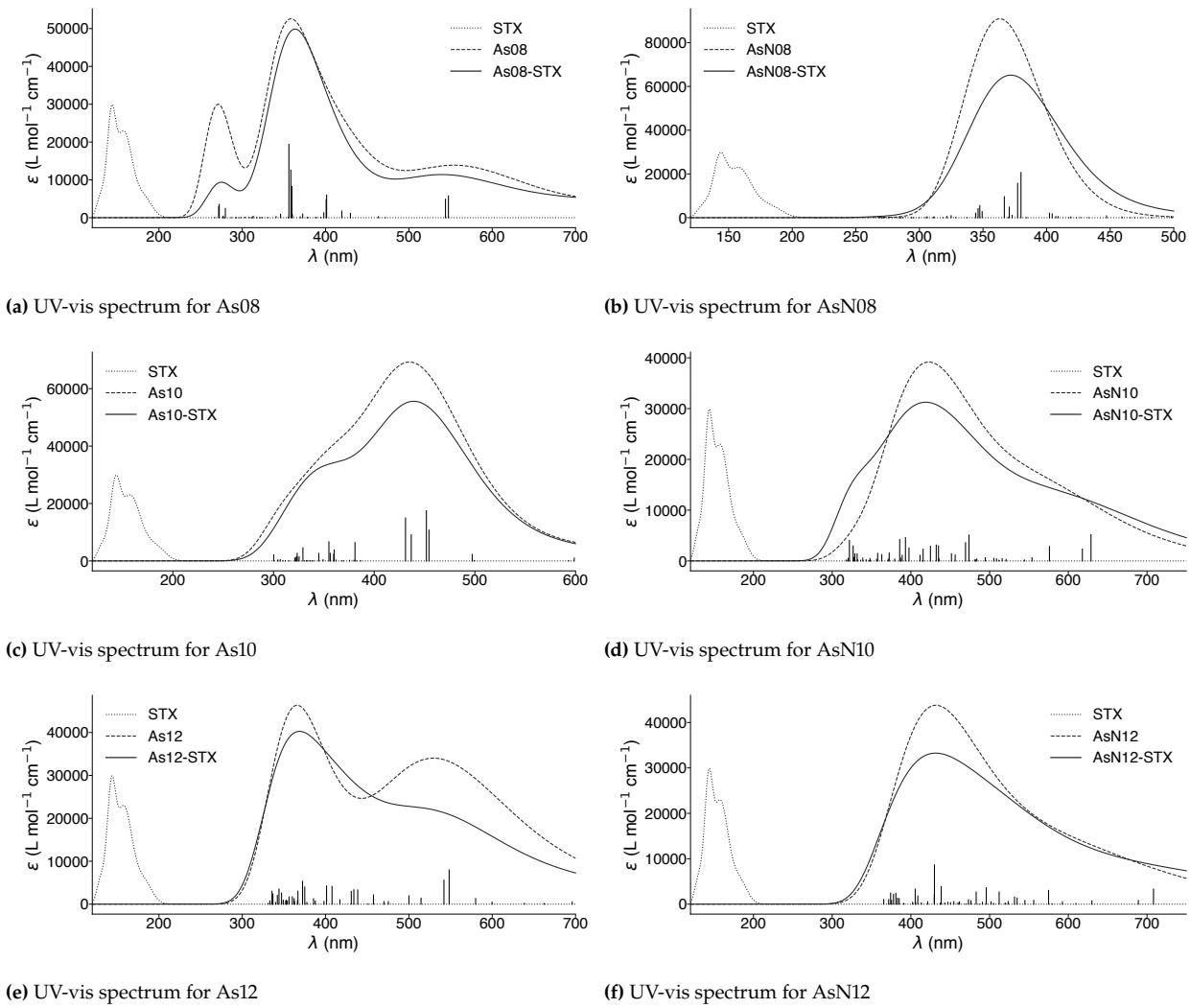
**Figure A.2:** Part 2 of NICS 2D projections

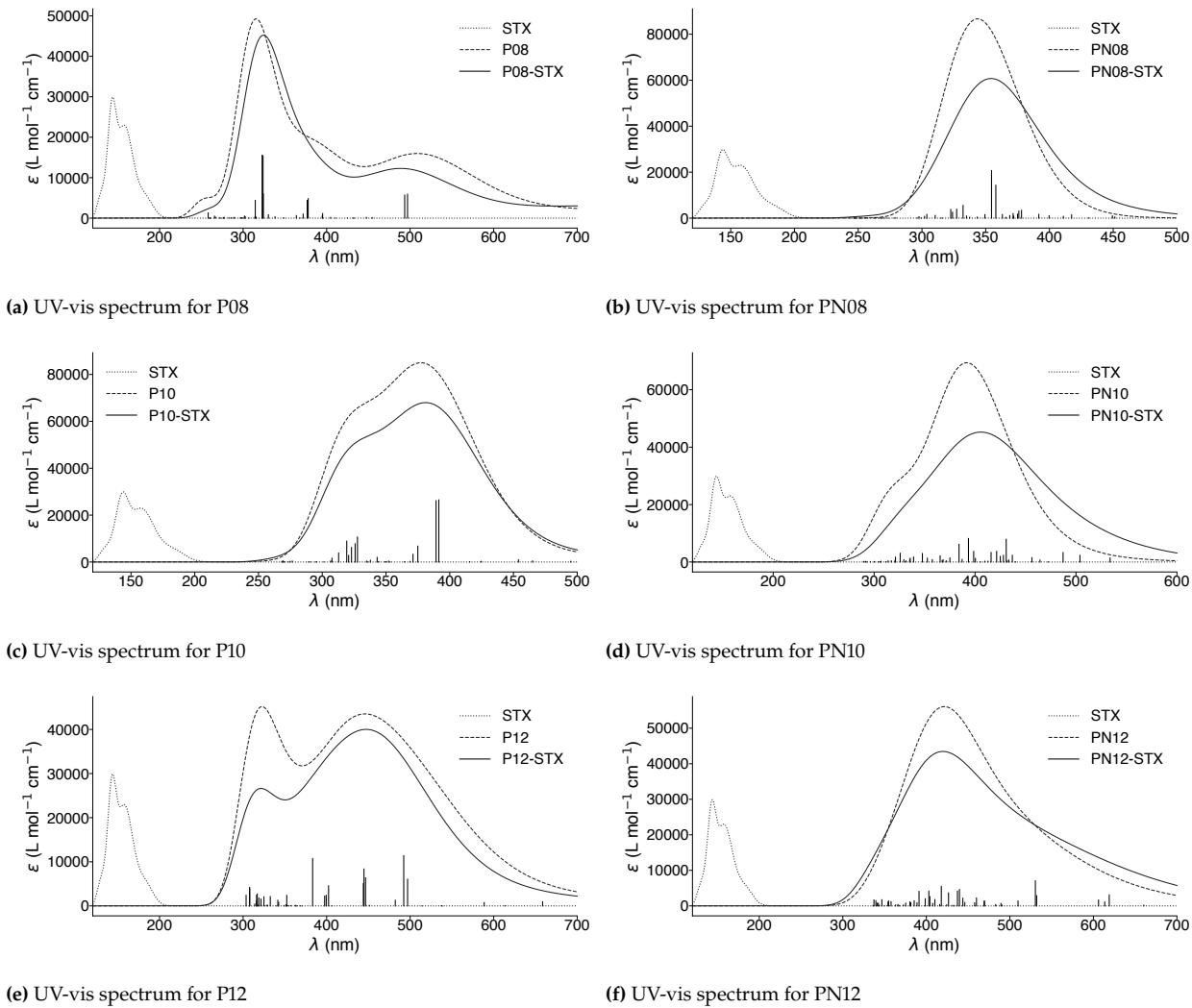
## **A.2 Raman spectra of all sunflowers**

### A.3 UV-vis spectra of all sunflowers



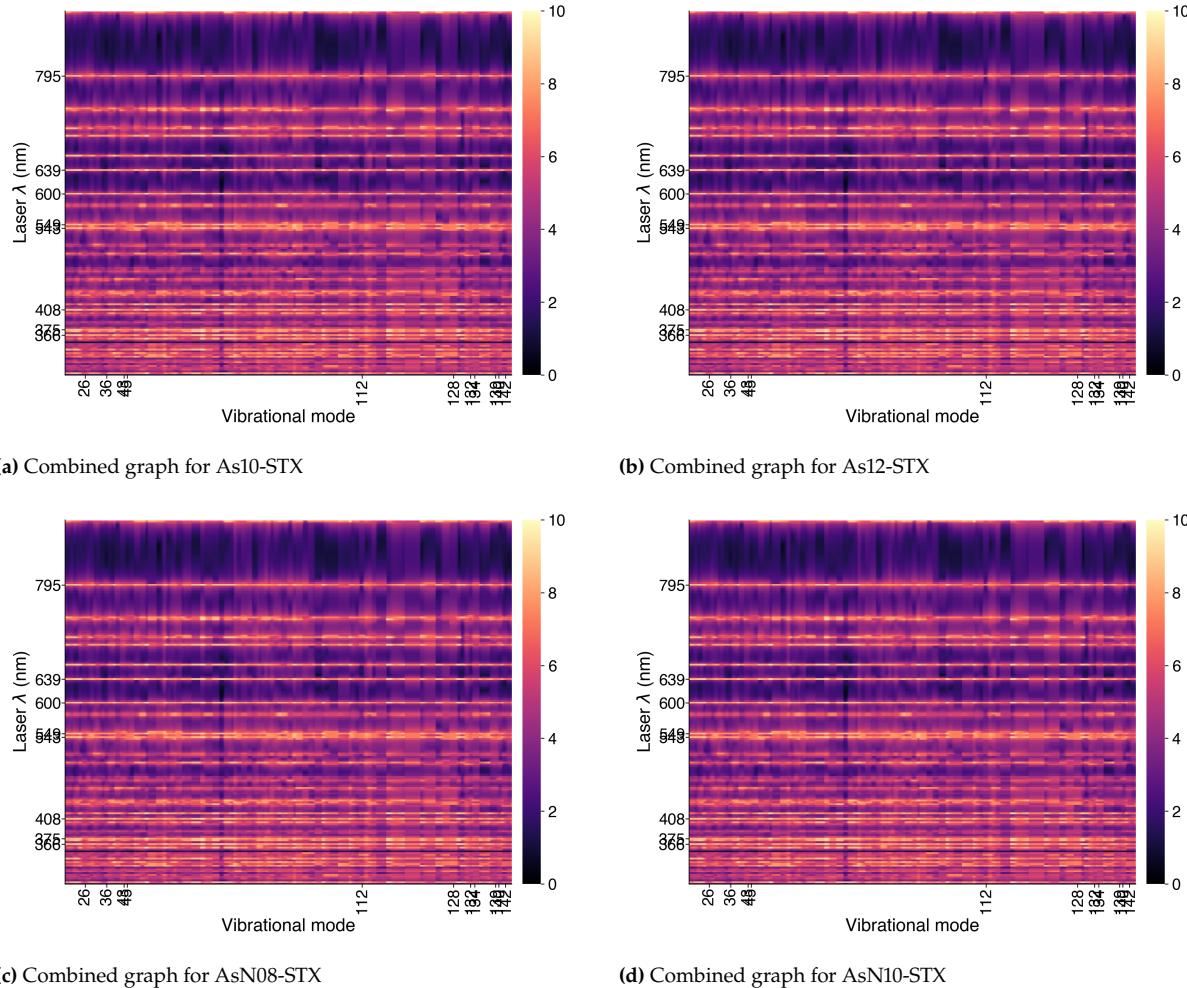
**Figure A.3:** Part 1 of flower UV-vis spectra

**Figure A.4:** Part 2 of flower UV-vis spectra

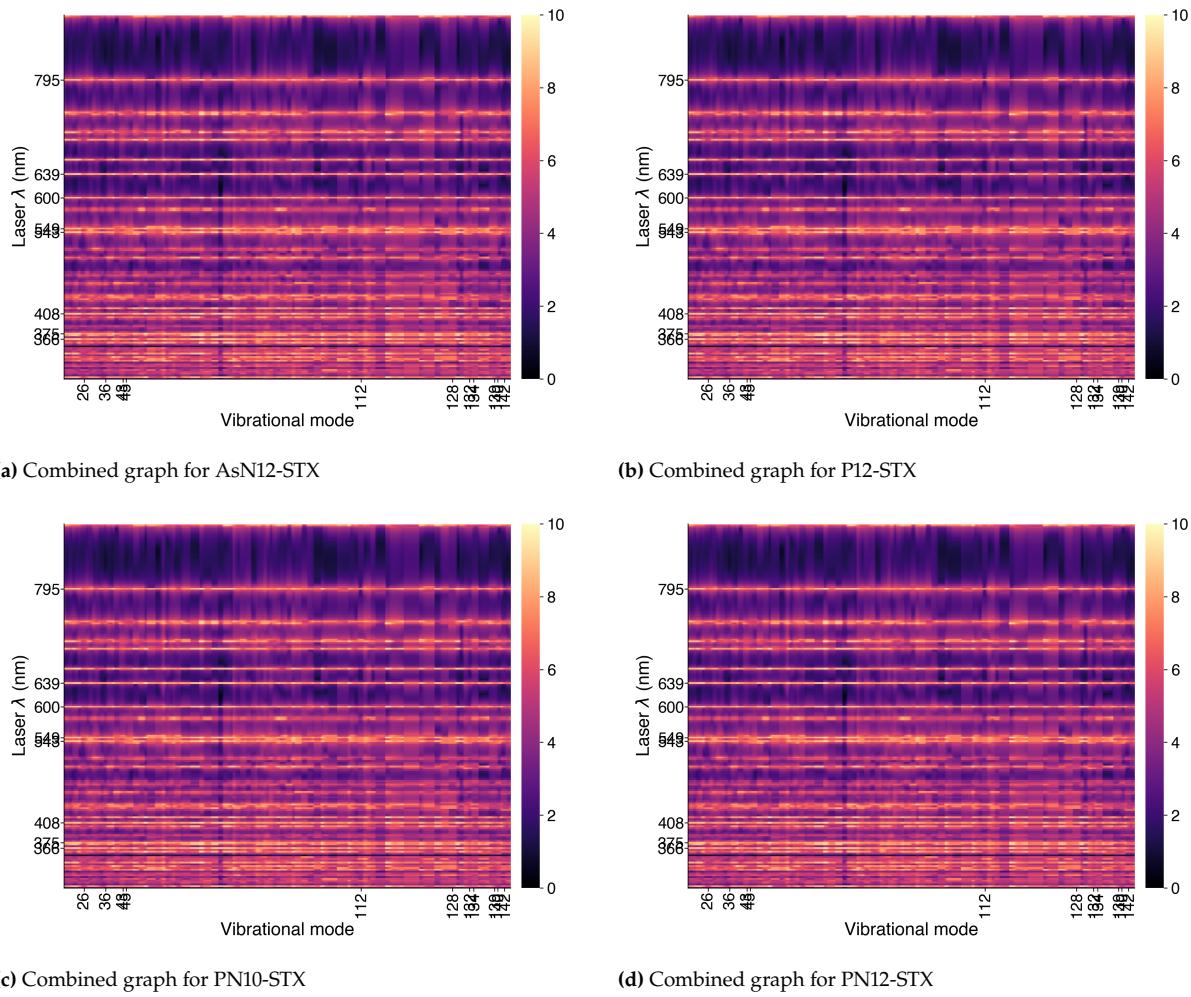


**Figure A.5:** Part 3 of flower UV-vis spectra

## A.4 Combined enhancement factor graphs



**Figure A.6:** Part 1 of combined EF RR graphs



**Figure A.7:** Part 2 of combined EF RR graphs

## A.5 Combined resonance Raman spectra



# Bibliography

Here are the references in citation order.

- [1] Yan Zhao and Donald G. Truhlar. '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'. In: *Theor. Chem. Acc.* 120 (2008), pp. 215–241 (cited on page 5).
- [2] Florian Weigend and Reinhart Ahlrichs. 'Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy'. In: *Phys. Chem. Chem. Phys.* 7 (2005), pp. 3297–3305 (cited on page 5).
- [3] X. Li and M. J. Frisch. 'Energy-represented DIIS within a hybrid geometry optimization method'. In: *J. Chem. Theory and Comput.* 2 (2006), pp. 835–839 (cited on page 6).
- [4] Gábor Magyarfalvi et al. 'Shielding Calculations: GIAO Methods'. In: *eMagRes* (2011) (cited on page 6).
- [5] Haydar Yüksek et al. 'Preparation, GIAO NMR Calculations and Acidic Properties of Some Novel 4,5-dihydro-1H-1,2,4-triazol-5-one Derivatives with Their Antioxidant Activities'. In: *Int. J. Mol. Sci.* 9(1) (2008), pp. 12–32 (cited on page 6).
- [6] Erich Runge and E. K. U. Gross. 'Density-Functional Theory for Time-Dependent Systems'. In: *Phys. Rev. Lett.* 52 (1984), pp. 997–1000 (cited on page 6).
- [7] Carlo Adamo and Denis Jacquemin. 'The calculations of excited-state properties with Time-Dependent Density Functional Theory'. In: *Chem. Soc. Rev.* 42 (2013), pp. 845–856 (cited on page 6).
- [8] Joseph W. Ochterski. *Vibrational Analysis in Gaussian*. Retrieved: 2020. 1999. URL: <http://gaussian.com/vib/> (cited on page 6).
- [9] A. Y. Hirakawa and M. Tsuboi. 'Molecular Geometry in an Excited Electronic State and a Preresonance Raman Effect'. In: *Science* 188 (1975), pp. 359–361 (cited on page 8).
- [10] John R. Lombardi and Ronald L. Birke. 'A Unified Approach to Surface-Enhanced Raman Spectroscopy'. In: *J. Phys. Chem. C* 112 (2008), pp. 5605–5617 (cited on page 8).
- [11] Konstantin Yu. Chernichenko et al. "Sulflower": A New Form of Carbon Sulfide'. In: *Angew. Chem. Int. Ed.* 45 (2006), pp. 7367–7370 (cited on page 13).
- [12] Zhongfang Chen et al. 'Nucleus-Independent Chemical Shifts (NICS) as an Aromaticity Criterion'. In: *Chem. Rev.* 118 (2005), pp. 3842–3888 (cited on pages 15, 17).
- [13] Ángel Vidal-Vidal, Carlos Silva López, and Olalla Nieto Faza. 'Nitrogen Doped NanoHoops as Promising CO<sub>2</sub> Capturing Devices'. In: *Phys. Chem. Chem. Phys.* 20 (2005), pp. 8607–8615 (cited on page 16).
- [14] M. Faraday. 'On new compounds of carbon and hydrogen, and on certain other products obtained during the decomposition of oil by heat'. In: *Philos. Trans. R. Soc. London* 115 (1825), pp. 440–466 (cited on page 17).
- [15] Paul von Ragué Schleyer et al. 'Nucleus-Independent Chemical Shifts: A Simple and Efficient Aromaticity Probe'. In: *J. Am. Chem. Soc.* 118 (1996), pp. 6317–6318 (cited on page 18).
- [16] F. Pedregosa et al. 'Scikit-learn: Machine Learning in Python'. In: *J. Mach. Learn. Res.* 12 (2011), pp. 2825–2830 (cited on page 20).
- [17] Ó. Iglesias. *nics.py: a command line application to help perform and interpret non planar NICS calculations in Gaussian*. Version 1.0. 2020. URL: <https://github.com/oscarigrexas/nics> (cited on page 20).
- [18] T. A. Keith and R. F. W. Bader. 'Calculation of magnetic response properties using a continuous set of gauge transformations'. In: *Chem. Phys. Lett.* 210 (1993), pp. 223–231 (cited on page 21).

- [19] Frans B. van Duijneveldt, Jeanne G. C. M. van Duijneveldt-van de Rijdt, and Joop H. van Lenthe. 'State of the Art in Counterpoise Theory'. In: *Chem. Rev.* 7 (1994), pp. 1873–1885 (cited on page 28).
- [20] N. Rösch. *Counterpoise Correction*. Retrieved: 2020. 2003. URL: <http://gaussian.com/uvvisplot/> (cited on page 28).

# Notation

The next list describes several symbols that will be later used within the body of the document.

$c$  Speed of light in a vacuum inertial frame

$h$  Planck constant

