

# Computational Chemistry Laboratory

Sree Ganesh Balasubramani, Filipp Furche

Winter 2018  
Version 1.0

# Acknowledgments

Some chapters are based on the Chem 137 Lab manual, which was co-authored by Guo Chen, Matthew M. Agee, Erik P. Almaraz, Ruben Magaña, and Pei-Hsuan Yang.

This material is based upon work supported by the National Science Foundation under CHE-1464828.

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
<b>2</b>	<b>Structures and Spectroscopy of Water-Halide Clusters</b>	<b>7</b>
2.1	Introduction . . . . .	7
2.2	Statement of the Model . . . . .	7
2.3	Background . . . . .	8
2.3.1	Ensembles . . . . .	8
2.3.2	Global Minimum Search by AIMD . . . . .	8
2.3.3	Thermal Corrections . . . . .	9
2.3.4	Infrared Spectra . . . . .	10
<b>3</b>	<b>Electronic Spectroscopy of Luciferin Dyes</b>	<b>12</b>
3.1	Introduction . . . . .	12
3.2	Background . . . . .	12
3.2.1	The Born-Oppenheimer Approximation . . . . .	12
3.2.2	Franck-Condon Absorption . . . . .	13
3.2.3	Emission and Fluorescence the Einstein A and B coefficients . . . . .	13
3.2.4	Calculation of excitation energies using response theory . . . . .	13
<b>4</b>	<b>Bimetallic Lanthanide and Actinide Metallocene Hydrides</b>	<b>16</b>
4.1	Introduction . . . . .	16
4.2	Background . . . . .	16
4.2.1	Localized orbitals . . . . .	16
4.2.2	Traditional Lanthanide chemistry . . . . .	17
4.2.3	Cyclopentadienyl ligands and isolation of non-traditional oxidation states of Ln . . . . .	17
4.2.4	Electron paramagnetic resonance (EPR) spectroscopy . . . . .	18
<b>A</b>	<b>Good Scientific Computing Practices</b>	<b>19</b>
A.1	Principles . . . . .	19
A.2	Project Design . . . . .	19
A.3	Data Generation . . . . .	20

---

A.4	Data Analysis . . . . .	20
A.5	Publication . . . . .	21
A.6	Version Control . . . . .	21
A.6.1	Basic concepts . . . . .	21
<b>B</b>	<b>The Computational Workplace</b>	<b>23</b>
B.1	Keeping Records . . . . .	23
B.2	Your UNIX User Account . . . . .	23
B.3	Basic UNIX Shell Commands . . . . .	24
B.4	The Vi and Emacs Editors . . . . .	24
B.5	UNIX File Systems . . . . .	27

# List of Figures

3.1	Vertical excitation from ground electronic level to excited electronic level within the clamped nuclei approximation . . . . .	14
-----	--	----

# List of Tables

4.1	Localization . . . . .	16
B.1	Basic shell commands . . . . .	25
B.2	Command-line features of bash . . . . .	26
B.3	Table: Basic Vi commands. . . . .	26
B.4	Useful key-combinations for the emacs editor . . . . .	27

## **Chapter 1**

# **Introduction**

## Chapter 2

# Structures and Spectroscopy of Water-Halide Clusters

### 2.1 Introduction

The physical and chemical properties of water-halide solutions depend strongly on dimensionality: Compared to the bulk solution, certain ions may have higher concentrations at or close to the surface at interfaces and in other confined environments [1]. This has important consequences for the structure and properties of low-dimensional ion solutions. For example, the selective enrichment of  $\text{Br}^-$  at the surface of sea-salt aerosol particles explains the disproportionately large role of bromide ions in tropospheric halogen chemistry [2].

Cluster models of water-halide solutions have a well-defined chemical composition and can be studied by high-resolution spectroscopy and electronic structure calculations. Charged clusters may be generated and separated using mass spectrometry, and characterized by photoelectron spectroscopy as well as UV- and IR-spectroscopy. The key to highly resolved UV- and IR spectra of cluster ions in the gas phase is action spectroscopy: One or several weakly bound inert gas atoms such as Ar are attached to the cluster ions, and the absorption cross section is determined by monitoring dissociation of the weakly bound complex [3].

The goal of this experiment is the determination of minimum structures and relative energies of  $\text{X}^-(\text{H}_2\text{O})_3$ ,  $\text{X}=\text{F}, \text{Cl}$ , clusters. The thus obtained structures will be validated by comparison of computed IR spectra to experimental spectra.

Additional background reading: Schwabl [4]; Allen and Tildesley [5]; Wilson, Decius, Cross [6]

### 2.2 Statement of the Model

Our experiment is based on the following key assumptions:

- The spectroscopic properties of the water-halide clusters may be described by a classical grand canonical ensemble.



- All low-lying cluster structures are sampled by *Ab initio* molecular dynamics simulations (AIMD) with sudden quenching.
- The nonrelativistic Born-Oppenheimer ground state provides an adequate approximation for the electronic properties.
- The minimum structures, their relative energies, and nuclear vibrational spectra are well described by hybrid DFT.
- The harmonic oscillator approximation is accurate enough to distinguish IR spectra of different cluster structures.

## 2.3 Background

### 2.3.1 Ensembles

Within the canonical ensemble, the high-temperature limit of the statistical operator is

$$\hat{W} = \frac{e^{-\beta\hat{H}}}{Z}, \quad (2.1)$$

where  $\beta = 1/(kT)$ ,  $k$  is the Boltzmann constant,  $T$  is the temperature,  $\hat{H}$  is the molecular Hamiltonian, and

$$Z = \langle e^{-\beta\hat{H}} \rangle \quad (2.2)$$

is the canonical partition function; brackets are used to denote the trace operation.

#### Assignment P 2.1. Boltzmann Ensemble

1. Define the terms microcanonical and canonical ensemble.
2. Describe qualitatively how Eq. (2.1) for the statistical operator may be derived.
3. Why is it reasonable to use the high temperature limit for water-halide clusters?

### 2.3.2 Global Minimum Search by AIMD

In AIMD, the nuclei are approximated by classical particles moving on the (ground state) Born-Oppenheimer potential energy surface obtained from *ab initio* calculations. The classical equations of motion are integrated using time propagation methods, e.g., the leapfrog Verlet algorithm: The nuclear momenta  $\mathbf{P}$  and positions  $\mathbf{Q}$  at the  $i + 1/2$ -th and  $i$ -th time step are computed from

$$\mathbf{P}(t_{i+1/2}) = \mathbf{P}(t_{i-1/2}) + \mathbf{F}(t_i)\delta t, \quad (2.3a)$$

$$\mathbf{Q}(t_{i+1}) = \mathbf{Q}(t_i) + \mathbf{M}^{-1}\mathbf{P}(t_{i+1/2}), \quad (2.3b)$$

for fixed time-step  $\delta t$ ,  $t_{i+1} = t_i + \delta t$ ,  $i = 0, 1, 2, \dots$ . The initial positions of the nuclei at  $t_0$  and their initial momenta are specified as input.  $\mathbf{M}$  is the mass tensor, which is diagonal in Cartesian

coordinates. The force on the classical nuclei is obtained from the nuclear gradient of the electronic ground state potential energy surface  $E(\mathbf{Q})$ ,

$$\mathbf{F}(\mathbf{Q}) = -\nabla_{\mathbf{Q}}E(\mathbf{Q}) = -\frac{dE(\mathbf{Q})}{d\mathbf{Q}}. \quad (2.4)$$

The nuclear potential energy surface and its gradient are only required locally at a given position  $\mathbf{Q}$ .

For systems with relatively few degrees of freedom that relax on similar time-scales, free dynamics with sudden quenching (FDSQ) is an effective technique to search for the global potential energy minimum. In FDSQ, several nuclear trajectories are started from an initial structure with randomly assigned initial momenta. In free dynamics, no external fields are present, and the total energy remains constant. The initial kinetic energy of the nuclei is often expressed in terms of a pseudo-temperature  $T = 2/(3N_n)E^{\text{kin}}$ , where  $N_n$  is the number of nuclei and  $E^{\text{kin}}$  denotes the nuclear kinetic energy, and should be chosen large enough to overcome isomerization barriers quickly, but small enough to prevent undesirable side-reactions such as fragmentation. The nuclear potential energy of an FDSQ trajectory is used to detect local minima, which are located by “sudden quenching”, i.e., setting the nuclear kinetic energy to zero. In practice, it is common to perform geometry optimization starting from the local minima determined by the free dynamics simulation.

There is no guarantee that finite-time FDSQ simulations will detect the global minimum and/or other low-energy minima, but the likelihood increases with simulation time.

#### Assignment P 2.2. Global Minimum Search by AIMD

1. Derive the Verlet-leapfrog equations from Newton’s second law.
2. (\*) Derive the Verlet-leapfrog equations from the Hamiltonian equations of motion.
3. The leapfrog algorithm requires knowledge of the nuclear momenta at  $t_{-1/2}$ , which are generated from the initial momenta by a half-timestep backwards. Write down the equation for this half-timestep.
4. Why is it not strictly correct to refer to a nuclear “temperature” in free dynamics simulations?
5. Why does geometry optimization mimic “sudden quenching”, and why is it preferable?

### 2.3.3 Thermal Corrections

Under experimental conditions, different isomers of water-halide clusters are in chemical equilibrium. At finite temperature and constant pressure  $p$ , the composition of a mixture of different cluster isomers is determined by the Gibbs free energy (chemical potential) of each species. The Gibbs free energy is related to the canonical partition function by

$$G = -kT \ln Z + pV, \quad (2.5)$$

where  $V$  denotes the volume. Within the rigid rotor – harmonic oscillator (RRHO) approximation, the canonical partition function separates into a translational (trans), rotational (rot), and a vibrational (vib) part,

$$Z = Z^{\text{trans}} Z^{\text{rot}} Z^{\text{vib}}. \quad (2.6)$$

The translational partition function is

$$Z^{\text{trans}} = V(2\pi M kT)^{3/2}, \quad (2.7)$$

where  $M$  is the molecular mass. Within the rigid rotor approximation, the rotational partition function in the high-temperature approximation for a non-linear molecule is

$$Z^{\text{rot}} = \frac{1}{\sigma} \sqrt{\frac{\pi(kT)^3}{|\mathbf{I}|}}, \quad (2.8)$$

where  $\mathbf{I}$  is the inertial tensor and  $\sigma$  is a symmetry factor which is 2 for molecules with an inversion center and 1 for molecules without. The vibrational partition function within the Harmonic oscillator approximation is

$$Z^{\text{vib}} = \frac{1}{2} \prod_{n=1}^{N^{\text{vib}}} \text{csch}(\beta\Omega_n/2), \quad (2.9)$$

where  $\Omega_n$  denotes the harmonic vibrational frequencies, and  $N^{\text{vib}}$  is the number of vibrational degrees of freedom.

#### Assignment P 2.3. Thermal Corrections

1. What state variables does  $G$  depend on?
2. What is  $N^{\text{vib}}$  for a nonlinear molecule?
3. Explain why electronic contributions to the partition function, Eq. (2.6), may be neglected for this experiment.
4. Derive Eq. (2.9).
5. For vibrational modes with frequencies below  $100 \text{ cm}^{-1}$  the RRHO approximation becomes unreliable. Explain why.

#### 2.3.4 Infrared Spectra

The harmonic vibrational frequencies  $\Omega_n$  and the normal modes  $\mathbf{X}$  of a polyatomic molecule are the solutions of the generalized eigenvalue problem

$$\mathbf{V}\mathbf{X} = \mathbf{M}\mathbf{X}\Omega^2, \quad (2.10)$$

where  $\Omega$  is diagonal with diagonal elements  $\Omega_n$ .

$$V_{ij} = \frac{\partial^2 E(\mathbf{Q})}{\partial \mathbf{Q}_i \partial \mathbf{Q}_j} \quad (2.11)$$

is the force or spring constant matrix, also called potential energy Hessian. It requires calculations of all second derivatives of the potential energy with respect to nuclear displacements. The normal modes  $\mathbf{X}$  are normalized according to

$$\mathbf{X}^\dagger \mathbf{M} \mathbf{X} = \mathbf{1}. \quad (2.12)$$

The infrared spectrum of a molecular cluster can be crudely interpreted using the harmonic approximation: The main peaks are given by the fundamental vibrational frequencies, and the peak intensity is proportional to the square of the vibrational transition dipole moment

$$\frac{1}{\mu_n \Omega_n} \left| \frac{d\mathbf{D}}{d\mathbf{X}_n} \right|^2, \quad (2.13)$$

which involves the normal mode derivative of the dipole moment  $\mathbf{D}$  and the reduced mass  $\mu_n$ .

**Assignment P 2.4.** Infrared Spectra

1. What main features do you expect to see in the infrared spectra of a small water-halide cluster? Why?
2. The main computational cost in computing IR spectra goes into the calculation of the force constant matrix  $\mathbf{V}$ . With this information, how would you go about investigating isotope substitution effects computationally?
3. (\*) Derive Eq. (2.10) using Lagrangian mechanics.

## Chapter 3

# Electronic Spectroscopy of Luciferin Dyes

### 3.1 Introduction

Luciferin is the molecule responsible for the chemiluminescence observed in fireflies. The firefly produces the luciferase enzyme that cleaves a carboxyl group to generate an oxo species in the excited state in the presence of oxygen [7]. Luciferase is known to not only work on the wild type luciferin, but also a large number of chemically modified luciferins, making it a very versatile biosensor/marker. It has been noted that there is a strong correlation between the oscillator strengths of these molecules and their emission and absorption strengths.

While synthetic organic chemists are equipped with a large number of tools for all kinds of chemical modifications on luciferin, the process of testing the effectiveness of a chemically modified luciferin dye is expensive and time consuming. Since there is a correlation between the spectral strengths and the oscillator strengths, it is possible to make predictions about suitable candidates for experimental verifications from theoretical calculations. Time-dependent density functional theory (TDDFT) is an adequately accurate method for computing the transition properties between ground and excited states [8]. While there are various drawbacks of TDDFT in calculating excited state properties, it can usually be verified with other, more expensive methods, such as coupled cluster response based methods.

The aim of this experiment is to find out which chemically modified luciferin molecules are potentially better candidates for better emitters.

### 3.2 Background

#### 3.2.1 The Born-Oppenheimer Approximation

Within non-relativistic quantum mechanics the molecular Hamiltonian can be written as

$$\hat{H}(\mathbf{r}, \mathbf{R}) = \hat{T}_N(\mathbf{R}) + \hat{T}_e(\mathbf{r}) + \hat{V}_{eN}(\mathbf{r}, \mathbf{R}) + \hat{V}_{NN}(\mathbf{R}) + \hat{V}_{ee}(\mathbf{r}) \quad (3.1)$$

where  $N$  denotes nuclear terms,  $e$  denotes electronic terms,  $\mathbf{r}$  denote electronic coordinates and  $\mathbf{R}$  denote nuclear coordinates. The Schrödinger equation is

$$\hat{H}(\mathbf{r}, \mathbf{R}) |\Psi(\mathbf{r}, \mathbf{R})\rangle = E |\Psi(\mathbf{r}, \mathbf{R})\rangle \quad (3.2)$$

Within the Born-Oppenheimer (BO) approximation the wavefunction of the molecular system can be written as

$$|\Psi(\mathbf{r}, \mathbf{R})\rangle = \chi(\mathbf{R})\phi(\mathbf{r}) \quad (3.3)$$

where the separation of the nuclear and electronic coordinates results in the electronic Schrodinger equation with the nuclear coordinates as parameters.

### Assignment P 3.1. BO Approximation

1. State the key assumptions underlying the BO approximation.
2. Why is the BO approximation not capable of describing internal conversion?

### 3.2.2 Franck-Condon Absorption

Within the BO approximation since the nuclear coordinates are assumed to be fixed, there is no change in the geometry of the molecule during the process of electronic excitation. The excitation strengths are determined by the vibrational states with maximum overlap and this is called the Franck-Condon principle in theoretical spectroscopy.

### Assignment P 3.2. Franck-Condon Principle

Give a rationale for the Franck-Condon principle.

### 3.2.3 Emission and Fluorescence the Einstein A and B coefficients

### 3.2.4 Calculation of excitation energies using response theory

Consider a time-independent Hamiltonian whose Eigenstates ( $\{\Psi_m\}$ ) satisfy the time-independent Schrödinger equation

$$H_0 |\Psi_m\rangle = E_m |\Psi_m\rangle \quad (3.4)$$

Suppose if a time-dependent perturbation  $F(t)$  is applied to this system at time  $t = 0$ , the Hamiltonian can be written as

$$\hat{H}(t) = \hat{H}_0 + \theta(t)F(t)\hat{\rho} \quad (3.5)$$

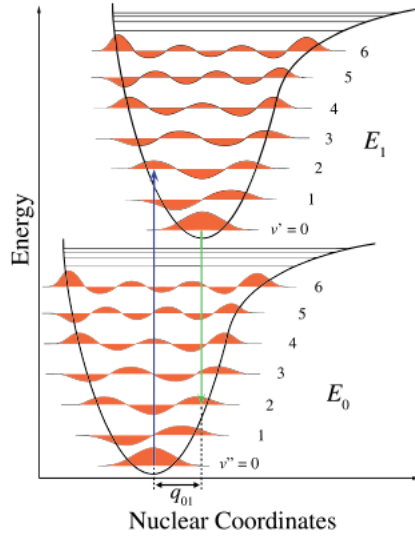


Figure 3.1: Vertical excitation from ground electronic level to excited electronic level within the clamped nuclei approximation

where the perturbation couples to the unperturbed Hamiltonian through the density operator  $\hat{\rho}$ . The time-dependent density can be expanded in a power series in the external perturbation as

$$\langle \Psi(t) | \hat{\rho} | \Psi(t) \rangle = \rho(t) = \rho_0 + \rho^{(1)}(t) + \rho^{(2)}(t) + \dots \quad (3.6)$$

$$\rho^{(1)}(t) = \int_0^t dt' \chi(t-t') F(t') \quad (3.7)$$

where  $\rho^{(1)}(t)$  is the linear density response and  $\frac{\delta \rho(t)}{\delta F(t')} = \chi(t-t')$  is the density-density response function in the time domain. The Lehmann representation of the density-density linear response function in complex frequency plane can be written as

$$\chi(\omega) = \lim_{\eta \rightarrow 0^+} \sum_{n=1}^{\infty} \left[ \frac{\langle \Psi_0 | \hat{\rho} | \Psi_n \rangle \langle \Psi_n | \hat{\rho} | \Psi_0 \rangle}{\omega - \Omega_n + i\eta} - \frac{\langle \Psi_0 | \hat{\rho} | \Psi_n \rangle \langle \Psi_n | \hat{\rho} | \Psi_0 \rangle}{\omega + \Omega_n + i\eta} \right] \quad (3.8)$$

where  $\{|\Psi_m\rangle\}$  is the set of Eigenstates of  $H_0$ ,  $\Omega_m = E_m - E_0$  are the excitation energies between the ground state and the  $m^{th}$  excited state.

**Assignment P 3.3.** Analytic properties of the linear response function

1. What do the poles of the function  $\chi(\omega)$  represent?
2. (\*) Show that the linear response function is analytic in the upper half of the complex  $\omega$  plane.
3. Show that the linear response function satisfies  $\chi(\omega) = \chi^*(-\omega)$  where  $*$  denotes the complex conjugate.

- 
4. How can you obtain the transition density  $\rho_{m0} = \langle \Psi_m | \hat{\rho} | \Psi_0 \rangle$  corresponding to a transition from the ground to  $m^{th}$  excited state from the linear response function?
  5. (\*) Derive Eq. 3.8 using the equation of motion for the density operator.



## Chapter 4

# Bimetallic Lanthanide and Actinide Metallocene Hydrides

### 4.1 Introduction

Metal-metal multiple bonds through the overlap of d orbitals in transition metal complexes was discovered in the 1960s[9]. Some well-known examples include  $[\text{Re}_2\text{Cl}_8]^{2-}$  [10]. Unlike transition metals lanthanide and actinide (Ln-Ln) multiple bonds were considered unlikely because of the limited radial extension of the valence f orbitals. Recently Ln complexes with stable  $4f^n 5d^1$  ground state electronic configuration were synthesized suggesting the possibility of formation of bimetallic complexes with d-d bonding. A more recent study using DFT predicts the possibility that the reduction of  $[\text{Cp}_2'\text{Ln}(\mu - \text{H})]_2$  complex can lead to stable Ln-Ln bonding[11, 12].

### 4.2 Background

#### 4.2.1 Localized orbitals

- Localization is a unitary transformation of occupied orbitals. Slater determinant, total energy and all properties are invariant.
- Edmiston-Ruedenberg localization[13]: minimize interorbital Coulomb energy. Other methods include Boys[14], Pipek-Mezey[15].

Advantages	Disadvantages
<ul style="list-style-type: none"><li>• chemically intuitive (Lewis structures) (?)</li><li>• transferable bond properties</li><li>• can speed up calculations</li></ul>	<ul style="list-style-type: none"><li>• not general, fails for many compounds</li><li>• orbital energy not well-defined</li><li>• assumptions on orbital extent</li></ul>

Table 4.1: Localization

Hund's condition for perfect localization: For all atoms, the following must hold:

$$N_{AO} = N_B = N_V \quad (4.1)$$

$N_{AO}$ : Number of AOs contributing to bonds  $N_B$ : Number of bonds  $N_V$ : Number of valence electrons contributing to bonds

**Assignment P 4.1.** Hund's Localization Condition

Check if the following molecules satisfy the Hund's condition for perfect localization:

1.  $\text{CH}_4$
2.  $\text{H}_2\text{O}$
3.  $\text{B}_2\text{H}_6$
4.  $\text{H}_2\text{O}^+$

#### 4.2.2 Traditional Lanthanide chemistry

The number and type of accessible formal oxidation states of an element in the periodic table is fundamental for understanding its reactivity. For the lanthanide series starting from lanthanum (La) to lutetium (Lu) it was known that the most stable oxidation state was always +3 regardless of the number of valence f electrons in the metal ion. The +2 oxidation state was initially known to exist in molecular complexes (in solution) only for the half filled  $4f^7$   $\text{Eu}^{2+}$ , filled shell  $4f^{14}$   $\text{Yb}^{2+}$  and the almost half-filled  $4f^6$   $\text{Sm}^{2+}$ .

**Assignment P 4.2.** Explain the trend in the ionic radius of the elements in the lanthanide series from La to Lu.

#### 4.2.3 Cyclopentadienyl ligands and isolation of non-traditional oxidation states of Ln

However, recent experiments and theoretical calculations have confirmed the existence of stable +2 oxidation states for the rare earth metals Y, Ho, Er, Pr, Gd, Tb, Lu, and the actinides U and Th under the special environment provided by  $\{\text{C}_5\text{H}_4\text{Si}(\text{CH}_3)_3\}^{-1}$  ligands[16].

**Assignment P 4.3.** Non-traditional oxidation states

1. Explain the 5d orbitals splitting in the ligand environment of the tris-cyclopentadienyl ligands using simple ligand field theory and group theory.
2. Using this explain the possibility of a  $\text{Ln}^{3+}/\text{Ln}^{2+}$  reduction following an electronic configuration change of  $4f^n \rightarrow 4f^n 5d^1$  rather than  $4f^n \rightarrow 4f^{n+1}$  (which is the traditional expectation).

#### 4.2.4 Electron paramagnetic resonance (EPR) spectroscopy

EPR spectroscopy is used to study molecules with unpaired electrons. It is similar to NMR spectroscopy

**Assignment P 4.4.** Predict the EPR splittings for this complex based on its electronic configuration

**Assignment 4.1.** Run a DFT calculation on the  $B_2H_6$  molecule using the def2-TZVP basis set and PBE functional and other standard settings in TURBOMOLE. Generate a cube format file to visualize the molecular orbitals using VMD or other visualization programs. Furthermore calculate the localized molecular orbitals and visualize them.

**Assignment 4.2.** Now do the same steps with Cp complexes

# Appendix A

## Good Scientific Computing Practices

### A.1 Principles

- Organization: Clarity about goals, scientific priorities
- Accountability: Keeping records
- Transparency: Methods and approach must be scientifically motivated and justified
- Efficiency: Achieving the scientific goals with the smallest necessary effort of (human and computational) resources
- Replicability and reproducibility: Results must be numerically replicable and repeatable by someone else under different conditions
- Critical awareness: Pro-actively checking for systematic errors, bias, human error; willingness to submit results to audit and criticism by others
- Reusability: Making results and code available to others

### A.2 Project Design

- Scientific goals need to be as clearly and specifically stated as possible before computation and/or coding starts: What, why, how
- Scientific goals must inform the methodology and project set-up, not vice versa
- Hypotheses should be stated as crisply as possible
- Before the start of code development projects, working equations and/or pseudocode should be completely written down
- The literature should be carefully reviewed to avoid duplication and giving due credit to others
- The project should be broken down in tasks (to-do list) with a rough time-line

### A.3 Data Generation

- Create a project directory in an easily accessible location in your backed-up home directory, e.g., `projects/ln_compounds` and add the following files/directories:
  - A README file briefly describing the purpose of the project (text format)
  - `/data` directory for storage of raw data
  - `/doc` directory for notes, reports, and manuscript files
  - `/results` directory for spreadsheets and data extraction
  - `/scr` directory for scripts, programs, source code
- Calculations are always performed on local scratch or work disks.
- Choose a directory structure that reflects the methodology. For example: `/work/phi/ln_compounds/thcpp3/tpss/svp`
- Don't overwrite raw data. Instead, make new directories for revised calculations. Use descriptive names, e.g., `"grid_m3, grid_3, ...` rather than `"run1, run2, ..."`.
- Once the calculations have finished, copy or sync the results to your `/data` directory:
  - At the minimum, all input and the main output files need to be copied.
  - The output must contain the version of the code. Additional information on the platform/compiler options used is desirable. Commands used to run applications, run scripts (e.g. for batch jobs, queueing systems) need to be saved as well.
  - Intermediates should not be saved, particularly when they are in binary format and consume lots of disk space.
- Everything (except for raw data) should be version controlled. Simplest version control method is using a CHANGELOG file.

### A.4 Data Analysis

- Transfer data from output to a spreadsheet by
  - copy and paste,
  - existing scripts,
  - custom scripts. If you use the latter, scripts must have a version and need to be tested and archived.
- Spreadsheets need to be well commented. Use tabs/sections to structure data. Each table needs a caption defining all data shown, column and row names must be descriptive and unique.

- Spreadsheet computing is error prone. Each formula needs to be accompanied by a descriptive comment and tested. Avoid doing complicated computations using a spreadsheet.
- Spreadsheet formats: .csv, xls (Google Sheets, OpenOffice Calc), Emacs org-mode
- Use specialized software for further analysis and/or plotting if necessary. Examples: gnuplot, R, jupyter notebook.
- Separate data generation and analysis work cycles
- The data need to be analyzed in a meaningful way: Give answers to the scientific questions, test hypotheses

## A.5 Publication

- Timely, effective publication of results is critical for accountability and reuseability
- Computational manuscripts need to include a “computational details” section describing computational methods
- Coordinates of optimized structures, computed spectra and other key observables should be supplied as supporting information
- Additional data (e.g. MD trajectories) should be deposited in electronic repositories such as UC’s escholarship and linked to the original publication.
- Published code repositories should be citeable by DOI
- Style [Strunk and White]: Concise and clear, determined by logical structure of the paper
- Graphical excellence [Tufte]:
  - Well-designed presentation of relevant data: Substance, statistics, design
  - Complex ideas communicated with clarity, precision, efficiency (least ink, smallest space, shortest time, maximum information)
  - Nearly always multivariate
  - Requires telling the truth about the data

## A.6 Version Control

### A.6.1 Basic concepts

- All important data of a project are gathered in a repository (repo)
- A single version of project is defined at all times by the master branch or trunk

- Contributors may locally pull or check out versions of the master branch, make changes and additions, and then push or check in to the master branch
- If the version of the master branch has changed, the new changes need to be merged into the local version before pushing

## Appendix B

# The Computational Workplace

### B.1 Keeping Records

Keeping records in a computational project is as important and necessary as keeping a journal in a real laboratory. Organization is crucial in computational chemistry because files and results can easily get lost. You should organize your data in a clear systematic fashion with subdirectories that can be easily understood and accessed. For example, do not put the calculations of two different molecules in the same directory. Do not place all the calculations with the same basis set in the same directory. Remember to never type results into documents by hand. Copy and paste results into your documents to avoid typos. Make it a habit to copy the results of your finished calculations to a location under your home directory where it is saved and will automatically be backed up.

One method of organizing your results is by using an OpenOffice.org Spreadsheet document to document your work. The OpenOffice.org Spreadsheet program can be accessed by clicking on the "Applications" button on the top left corner of the screen and is located under the Office tab. An example of a way to keep your data organized is demonstrated in Figure ???. For every computational project created, the spreadsheet document should contain the following information:

1. What has been done, e.g., what method, basis set, which isomer?
2. Where is the data located?
3. What are the main results?

### B.2 Your UNIX User Account

You will receive a username (typically your UCINetID) and a password at the beginning of this course. The first thing to do in the lab is to log on a workstation and to change your password. Your new password should have at least 8 characters, include at least two numbers and two special characters. Your password should not contain your user name or any other words or personal information that can be easily found out (e.g. doing an internet search on your name), but you should



be able to remember it. UNIX passwords are case-sensitive. To change your password, open up a shell terminal (clicking on the terminal icon on your application panel) and type:

```
passwd
```

You are personally responsible for your account including damage caused by abuse, so make sure your password is safe and do not share it with anyone. Please note that while you are free to use all installed software on the Modeling Facility workstations, it is illegal to copy or download installed proprietary software, even for your personal use.

Any computational project should be run on a local disc to avoid excessive network traffic. The default location is the so-called work disk. Create this directory and access it by doing the following:

```
mkdir work  
cd work
```

### B.3 Basic UNIX Shell Commands

Table B.1 lists some basic shell commands. For a more complete introduction to Linux see Ref. [Siever09a].

The default shell you will be using is the Bourne-again shell (bash). bash has a number of very useful command-line features listed in Table B.2 you should familiarize yourself with.

### B.4 The Vi and Emacs Editors

Vi is a basic UNIX text editor that is available in almost any UNIX environment. There are two basic modes in Vi, the command mode and the append mode. Table B.3 summarizes useful vi commands.

A more user-friendly editor is Emacs (and its X-windows version, Xemacs). Emacs has many options and modes, and Table B.4 displays only a small number of commands you need to get started with Emacs. You can access Emacs tutorials by clicking on "Help".

There are a number of other text editors available. If you are used to Microsoft Word, try OpenOffice, which has word compatibility.

Another useful package is Open Babel [**babel**] which is able to read, write and convert over ninety chemical file formats. This is particularly useful if the user wishes to open a certain file with different packages. For example, users could use OPEN BABEL when they wish convert a Protein Data Bank file (.pdb, standard format used by RCSB) that can normally be viewed with the VMD package to a XMOL molecule model file (XYZ, graphical coordinates) that can be opened with AVOGADRO. The users could use Open Babel to convert the PDB file to an XYZ file through using the terminal ?? and then opening the files with the respective packages. A sample sequence of the result of this is shown in ?? and ??.

Command	Description
pwd	Print working directory
ls	List all files in the current directory
ls -l	List all files in current directory with more details
cp file1 file2	Copy "file1" to "file2"
cp -r dir1 dir2	Recursively copy directory "dir1" to directory "dir2"
mkdir chemistry	Create directory "chemistry"
rmdir chemistry	Remove directory "chemistry"
cd chemistry/	Change working directory to "chemistry"
cd ..	Go to parent directory (..)
rm file1	Remove file1. Warning: once it's gone, it's gone
mv file1 file2	Rename "file1" to "file2". Warning: if file 2 already exists it will be lost
mv * chemistry/	Move all files into the directory chemistry
find * -name *.pdf -print	Find all .pdf files in all directories and their subdirectories
cat file1	Print contents of file1 on screen
less file1	Pager that allows you to scroll through file1 pressing <space>
grep "string" file1	Search file1 for "string" and print matching lines
which openoffice	Shows location of executable "openoffice"
man (command)	Shows manual page for "command"
bzip2 file, bunzip2 file.bz2	Compresses/Uncompresses 'file'
tar -vcjf dir.tar.bz2 *	Copy all files into compressed tape archive "dir.tar.bz2"
ps auwx	Print list of all running processes and their process IDs (PIDs)
kill <PID>	Kill (terminate) process
vi file1	Creates or opens file1 in the vi text editor
emacs file1	Creates or opens file1 in the emacs editor

Table B.1: Basic shell commands

Command	Description
<UP>, <DOWN>	Scroll through command history
<CTRL>-r	Search command history
<TAB>	Automatic completion
<CTRL>-c	Abort present command
<CTRL>-z	Stop present command
<CTRL>-a, <CTRL>-e	Jump to beginning/end of line
<CTRL>-d, <ALT>-d	Delete character/word
ls -l   less	Combine or pipe two commands (pipe is " ")
dscf > dscf.out	run program dscf and directs the output to "dscf.out"
dscf >> dscf.out	run program dscf and appends the output to "dscf.out"
define < define.input	run program define and redirects interactive input from file "define.input"
dscf &	Run program dscf in the background
nohup jobex &	Run program jobex in the background and continue running it even if the present shell is terminated
bg <PID>	Run (stopped) foreground process in the background
fg <PID>	Run (stopped) background process in the foreground

Table B.2: Command-line features of bash

Command	Description
a	Enter append mode
i, I	Insert text before cursor position (goes into append mode)
<Esc>	Leave append mode
o,O	Open new line below/before the current line in append mode
x	Delete character under cursor
dd	Delete whole line
D	Delete line from cursor position to the end of the line
r	Replace character under cursor
R	Enter replace mode
:w (file1)	Save current contents (in file1)
:wq	Save file and exit vi
:q!	Exit without saving changes

Table B.3: Table: Basic Vi commands.

Command	Description
<CTRL>-x <CTRL>-a	Save file and exit emacs
<CTRL>-x <CTRL>-c	Exit emacs
<CTRL>-u	Undo
<CTRL>-k	Delete line from cursor on to the end of line
<CTRL>-<space>	Mark beginning of a selected region
<CTRL>-y	Paste region

Table B.4: Useful key-combinations for the emacs editor

## B.5 UNIX File Systems

Unix keeps all files (including directories) using what is called a file system. After logging into your account, you will be placed in your home directory. Here you can create files, directories, and subdirectories. Your home directory is "mounted" on all Modeling Facility workstations and physically resides on the Modeling Facility Server. Your home directory (and all subdirectories) are backed-up in regular intervals.

Disk allocation is very important especially when dealing with I/O intensive calculations. The command "df -h" will print out the system hard disk space usage. This shows how much space is available and how much is being used in your main locations. The command "du -h" will list the disk usage in the current directory.

Any computational project should be run on a local disc to avoid excessive network traffic. The default location is the so-called work disk. Create this directory and access it by doing the following:

```
mkdir work  
cd work
```

Note that scratch disks are not backed up and may be lost in case of a hardware problem.

Organization is key in computational chemistry since files and results can easily get lost. You should organize your data using plenty and systematically clearly named subdirectories that can be easily understood and accessed. For example, don't put the calculations of two different molecules in the same directory and place all calculations with the same basis set in the same directory, etc.

# Bibliography

- [1] P. Jungwirth and D. J. Tobias. “Specific Ion Effects at the Air/Water Interface”. In: *Chemical Reviews* 106.4 (2006), pp. 1259–1281. doi: 10.1021/cr0403741.
- [2] B. J. Finlayson-Pitts. “The Tropospheric Chemistry of Sea Salt: A Molecular-Level View of the Chemistry of NaCl and NaBr”. In: *Chemical Reviews* 103.12 (2003), pp. 4801–4822. doi: 10.1021/cr020653t.
- [3] W. H. Robertson and M. A. Johnson. “Molecular aspects of halide ion hydration: the cluster approach”. In: *Annual Review of Physical Chemistry* 54.1 (2003), pp. 173–213. doi: 10.1146/annurev.physchem.54.011002.103801.
- [4] *Statistical Mechanics*. 2nd ed. Springer, 2006. ISBN: 978-3540323433.
- [5] *Computer Simulation of Liquids*. 2nd ed. Oxford, 2017.
- [6] *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. Dover, 1980. ISBN: 987-0-486-63941-3.
- [7] R. C. Steinhardt et al. “Brominated Luciferins Are Versatile Bioluminescent Probes”. In: *ChemBioChem* 18 (2017), pp. 96–100. doi: 10.1002/cbic.201600564.
- [8] F. Filipp and R. Ahlrichs. “Adiabatic time-dependent density functional methods for excited state properties”. In: *J. Chem. Phys.* 117.16 (2002), pp. 7433–7447. doi: 10.1063/1.1508368.
- [9] M. J. Bennett, F. A. Cotton, and R. A. Walton. “A Rhenium-to-Rhenium Triple Bond”. In: *J. Am. Chem. Soc.* 88.16 (1966), pp. 3866–3867. doi: 10.1021/ja00968a039.
- [10] F. A. Cotton. “Metal-Metal Bonding in  $[\text{Re}_2\text{X}_8]^{2-}$  Ions and Other Metal Atom Clusters”. In: *Inorganic Chemistry* 4.3 (1965), pp. 334–336. doi: 10.1021/ic50025a016.
- [11] Ryan R. Langeslay et al. “Expanding Thorium Hydride Chemistry Through  $\text{Th}_2^+$ , Including the Synthesis of a Mixed-Valent  $\text{Th}_4^+/\text{Th}_3^+$  Hydride Complex”. In: *J. Am. Chem. Soc.* 138.12 (2016), pp. 4036–4045. doi: 10.1021/jacs.5b11508.
- [12] Megan T. Dumas et al. “Synthesis and reductive chemistry of bimetallic and trimetallic rare-earth metallocene hydrides with  $(\text{C}_5\text{H}_4\text{SiMe}_3)^{1-}$  ligands”. In: *J. Organomet. Chem.* 849-850 (2017), pp. 38–47. doi: <https://doi.org/10.1016/j.jorganchem.2017.05.057>.
- [13] C. Edmiston and K. Ruedenberg. “Localized Atomic and Molecular Orbitals”. In: *Rev. Mod. Phys.* 35 (3 1963), pp. 457–464. doi: 10.1103/RevModPhys.35.457.

- 
- [14] J. M. Foster and S. F. Boys. "Canonical Configurational Interaction Procedure". In: *Rev. Mod. Phys.* 32 (2 Apr. 1960), pp. 300–302. doi: 10.1103/RevModPhys.32.300.
- [15] J. Pipek and P. G. Mezey. "A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions". In: *J. Chem. Phys.* 90.9 (1989), pp. 4916–4926. doi: 10.1063/1.456588.
- [16] W. J. Evans. "Tutorial on the Role of Cyclopentadienyl Ligands in the Discovery of Molecular Complexes of the Rare-Earth and Actinide Metals in New Oxidation States". In: *Organometallics* 35.18 (2016), pp. 3088–3100. doi: 10.1021/acs.organomet.6b00466.