

Chemistry without a Lab: An Overview of Computational Quantum Chemistry

Greg Beran

Department of Chemistry
University of California, Riverside



Spring ACS 2023

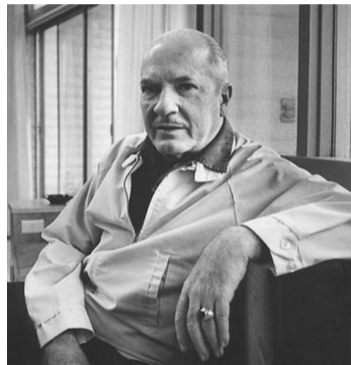
The chemistry jungle.

Chemistry is not a discipline today; it is a jungle....

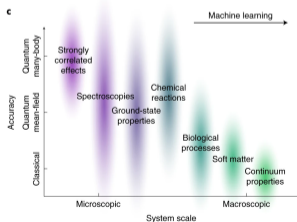
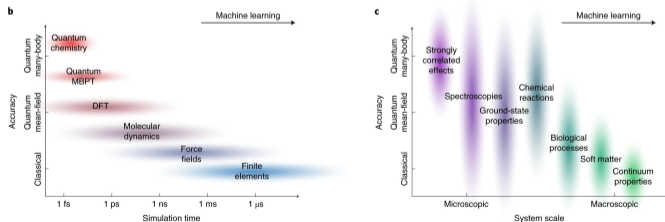
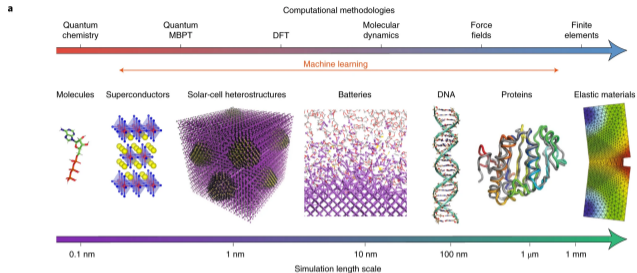
*When chemistry becomes a discipline, **mathematical chemists** will design new materials, predict their properties, and tell engineers how to make them—without entering the laboratory. We've got a long way to go on that one.*

— Robert Heinlein

from "Where To?", *Galaxy* magazine, Feb. 1952



Computation is rapidly transforming chemistry into a “discipline.”



Nature Materials, June 2021

PERSPECTIVE

Discovering and understanding materials through computation

Steven G. Louie, Yang-Hao Chan, Felipe H. da Jornada, Zhenglu Li and Diana Y. Qiu

REVIEW ARTICLES

Electronic-structure methods for materials design

Nicola Marzari, Andrea Ferretti and Chris Wolverton

Machine-learned potentials for next-generation matter simulations

Pascal Friederich, Florian Häse, Jonny Proppe and Alán Aspuru-Guzik

From predictive modelling to machine learning and reverse engineering of colloidal self-assembly

Marjolein Dijkstra and Erik Luijten

Mesoscopic and multiscale modelling in materials

Jacob Fish, Gregory J. Wagner and Sinan Keten

Louie, Chan, DAR Jornada, Li, and Qiu. *Nature Mater.* 20, 728 (2021).

Quantum chemistry plays a central role throughout computational chemistry.

- 1 Overview of Standard Quantum Chemistry Models
 - The molecular Schrödinger equation
 - Born-Oppenheimer approximation
 - Wave function methods
 - Density functional theory
 - Basis sets
 - Predictions versus reality
- 2 Selected Applications (biased by my own research)
 - Theory aiding experiment: NMR crystallography
 - Theory leading experiment: Photomechanical engines

1 Overview of Standard Quantum Chemistry Models

- The molecular Schrödinger equation
- Born-Oppenheimer approximation
- Wave function methods
- Density functional theory
- Basis sets
- Predictions versus reality

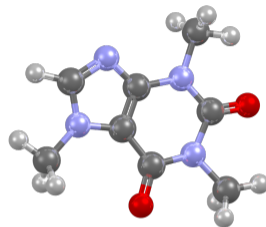
2 Selected Applications (biased by my own research)

- Theory aiding experiment: NMR crystallography
- Theory leading experiment: Photomechanical engines

The molecular Schrödinger equation is really complicated.

The molecular \hat{H} always contains:

- Kinetic energy for each nucleus
- Kinetic energy for each electron
- Attraction of each electron to each nucleus
- Repulsion between each pair of electrons
- Repulsion between each pair of nuclei



caffeine

$$\hat{H}\Psi(r, R) = E\Psi(r, R)$$

$$\hat{H} = -\sum_A^{\text{nuc.}} \frac{1}{2M_A} \nabla_A^2 - \sum_i^{\text{elec.}} \frac{1}{2} \nabla_i^2 - \sum_i^{\text{elec.}} \sum_A^{\text{nuc.}} \frac{Z_A}{r_{iA}} + \sum_i^{\text{elec.}} \sum_{j>i}^{\text{elec.}} \frac{1}{r_{ij}} + \sum_A^{\text{nuc.}} \sum_{B>A}^{\text{nuc.}} \frac{Z_A Z_B}{R_{AB}}$$

Such Hamiltonians are much too complicated to solve exactly. Need to approximate.

The Born-Oppenheimer approximation

Key approximation: the Born-Oppenheimer approximation approximately separates the electronic and nuclear variables.

- Only approximate since electron-nuclear attraction terms couple electrons and nuclei, preventing rigorous separation of variables.
- Nuclei are much more massive than electrons: $m_{nuc} \approx 1000m_e$
 - Means electrons move much faster than nuclei, respond quickly to changes in nuclear positions.
 - Like flies (electrons) buzzing around a rhino (nuclei).



The Born-Oppenheimer procedure

$$\text{Original Hamiltonian: } \hat{H} = -\sum_A^{\text{nuc.}} \frac{1}{2M_A} \nabla_A^2 - \sum_i^{\text{elec.}} \frac{1}{2} \nabla_i^2 - \sum_i^{\text{elec.}} \sum_A^{\text{nuc.}} \frac{Z_A}{r_{iA}} + \sum_i^{\text{elec.}} \sum_{j>i}^{\text{elec.}} \frac{1}{r_{ij}} + \sum_A^{\text{nuc.}} \sum_{B>A}^{\text{nuc.}} \frac{Z_A Z_B}{R_{AB}}$$

Strategy for using the Born-Oppenheimer approximation

- 1 Choose nuclear positions (i.e. molecular geometry), fix nuclei.

$$\Rightarrow \text{nuclear kinetic energy terms} \rightarrow 0 \text{ and } \frac{Z_A Z_B}{R_{AB}} \rightarrow \text{constant}$$

- 2 Solve resulting “electronic” Schrödinger equation, find electronic wavefunction and energy.

$$\hat{H}_{\text{elec}} = -\sum_i^{\text{elec.}} \frac{1}{2} \nabla_i^2 - \sum_i^{\text{elec.}} \sum_A^{\text{nuc.}} \frac{Z_A}{r_{iA}} + \sum_i^{\text{elec.}} \sum_{j>i}^{\text{elec.}} \frac{1}{r_{ij}} + \sum_A^{\text{nuc.}} \sum_{B>A}^{\text{nuc.}} \frac{Z_A Z_B}{R_{AB}}$$

- 3 Repeat Step 2 at multiple molecular geometries to map out a “potential energy surface”
- 4 If desired, solve “nuclear Schrödinger equation” that depends on the potential energy surface.
e.g. Computing harmonic vibrational frequencies

The potential energy surface maps out important species.

Potential Energy Surface (PES): The electronic energy as a function of nuclear positions.

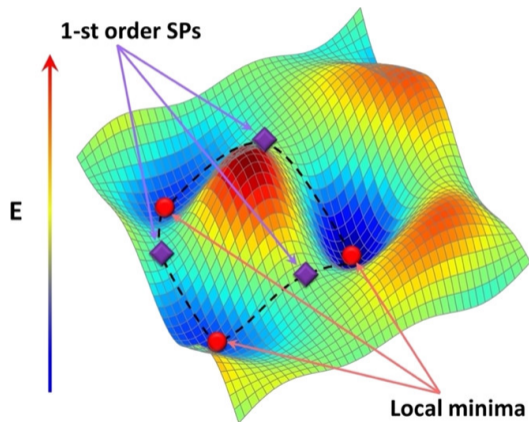
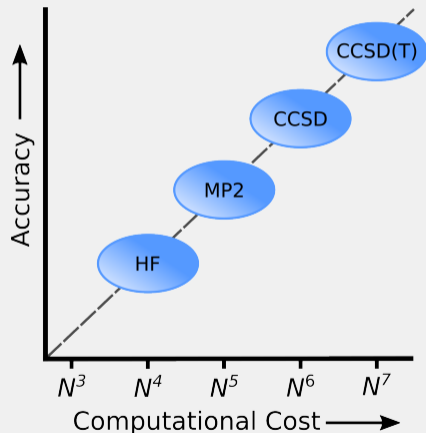


Image credit: Wang, Lv, Gao, & Ma. *Acc. Chem. Res.* 55 2068 (2022).

- **Optimize molecular geometry** to stationary points on the PES.
- **Stable species** \Rightarrow Minima
 - Reactants
 - Products
 - Reaction Intermediates
 - Conformers
 - **Have zero imaginary vibrational frequencies.**
- **Transition states** \Rightarrow 1st-order saddle points
 - **Have exactly 1 imaginary vibrational frequency.**

The hierarchy of popular quantum chemistry methods

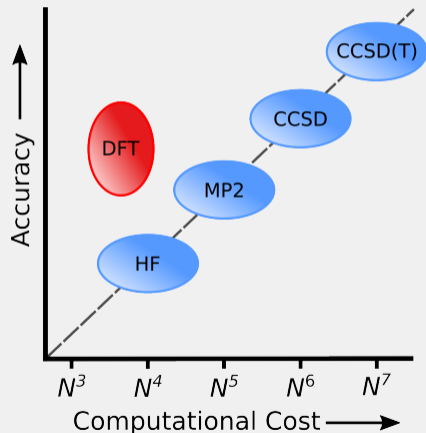
Always must seek to balance accuracy and computational efficiency.



- All of these methods seek to solve the electronic part of the Schrödinger equation.
- Hartree-Fock (HF) is the simplest method, but not very accurate.
- CCSD(T) is the “gold standard” of quantum chemistry. Practical upper limit of accuracy.
- Computational cost grows steeply for better methods.
 - e.g. CCSD(T): N^7 scaling means doubling system size increases cost $2^7 = 128$ times!

The hierarchy of popular quantum chemistry methods

Always must seek to balance accuracy and computational efficiency.



- All of these methods seek to solve the electronic part of the Schrödinger equation.
- Hartree-Fock (HF) is the simplest method, but not very accurate.
- CCSD(T) is the “gold standard” of quantum chemistry. Practical upper limit of accuracy.
- Computational cost grows steeply for better methods.
 - e.g. CCSD(T): N^7 scaling means doubling system size increases cost $2^7 = 128$ times!
- DFT has HF-like cost, but significantly better accuracy.

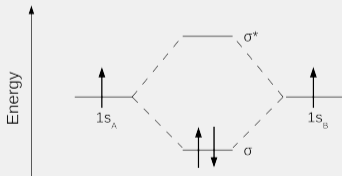
Hartree-Fock (HF) theory is quantitative molecular orbital (MO) theory.

HF forms molecular orbitals by taking linear combinations of atomic orbitals.

- Find the set of molecular orbitals (MOs) that minimizes the energy. (Variational principle)
- Pair up electrons in the lowest-energy orbitals. (Pauli Exclusion, Aufbau principles)

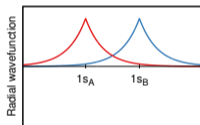
MO Diagram for H₂

$$\phi_{\sigma^*} = \frac{1}{\sqrt{2}} (\psi_{1s_A} - \psi_{1s_B})$$

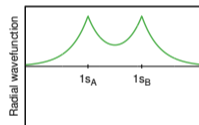


$$\phi_{\sigma} = \frac{1}{\sqrt{2}} (\psi_{1s_A} + \psi_{1s_B})$$

Form the σ bonding orbital by adding the 1s orbitals:

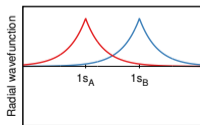


Interatomic Distance R

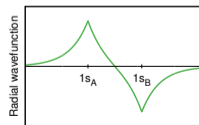


Interatomic Distance R

Form the σ^* anti-bonding orbital by subtracting the 1s orbitals:



Interatomic Distance R

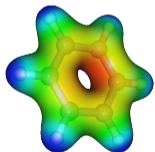
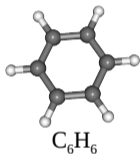


Interatomic Distance R

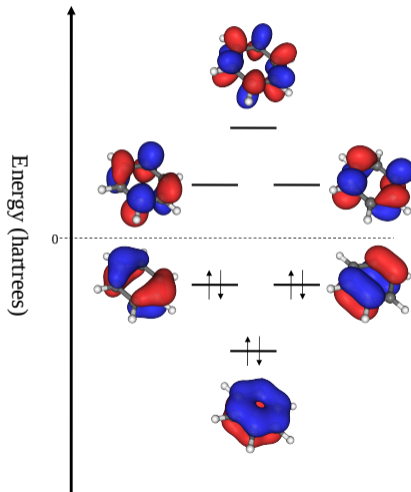
Sample Hartree-Fock calculation on Benzene

Look at only the π and π^* orbitals here.

**Benzene: HF/6-31G(d)
 π Molecular Orbitals**



Electrostatic Potential
Red = negative, blue = positive



Hartree-Fock is a mean field theory.

- Each electron feels only average “smear” of the other electrons.
 - HF lacks electron-electron correlations.
- HF captures $\sim 99\%$ of the total energy.
 - Gives decent molecular geometries.
 - But quantitatively poor reaction energies.
- Unfortunately, the missing 1%, called the **correlation energy**, is important for predicting chemistry accurately.

$$E_{corr} = E_{exact} - E_{HF}$$

Hartree-Fock is a mean field theory.

- Each electron feels only average “smear” of the other electrons.
 - HF lacks electron-electron correlations.
- HF captures $\sim 99\%$ of the total energy.
 - Gives decent molecular geometries.
 - But quantitatively poor reaction energies.
- Unfortunately, the missing 1%, called the **correlation energy**, is important for predicting chemistry accurately.

$$E_{corr} = E_{exact} - E_{HF}$$

Two potential strategies for improving upon Hartree-Fock theory

Both seek to capture that missing correlation energy.

- 1 Møller-Plesset perturbation theory
- 2 Configuration Interaction/Coupled Cluster theory

Second-Order Møller-Plesset Perturbation Theory (MP2)

Perturbation theory approximates the solution to a **hard problem** as the solution to an **easier problem plus corrections**.

Møller-Plesset Perturbation Theory uses perturbation theory to improve upon HF.

- **Hard problem:** Exact solution to electronic Schrödinger Eq. **Easier problem:** HF.

$$\hat{H}_{exact} = \hat{H}_{HF} + \hat{H}_{corr} \quad \Psi_{exact} \approx \Psi_{HF} + \Psi^{(1)} + \Psi^{(2)} + \dots \quad E_{exact} \approx E_{HF} + E^{(2)} + \dots$$

- **MP2:** First correction to the energy occurs at second-order ($E^{(2)}$).
 - Captures 99.8–99.9% of the total energy.
 - Improves significantly upon HF for energies and geometries.
 - Can be a good alternative to DFT in some cases.
 - Has issues with dispersion interactions, though corrections exist.
- **MP3, MP4**, etc: Higher-order corrections also can be computed, but coupled cluster methods usually perform better for similar computational effort.

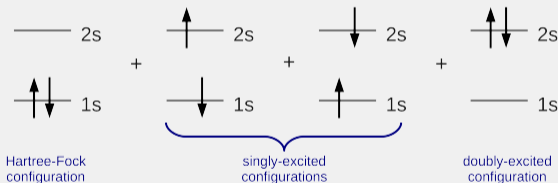
Phys. Chem. Chem. Phys. 24, 3695-3712 (2022).

Configuration Interaction

Configuration Interaction (CI): Optimize a wave function that is a linear combination of different possible ways of arranging the electrons in the Hartree-Fock orbitals.

Example: He atom with 2 orbitals

$$|\Phi_{CI}\rangle = c_1|1s \uparrow, 1s \downarrow\rangle + c_2|2s \uparrow, 1s \downarrow\rangle + c_3|1s \uparrow, 2s \downarrow\rangle + c_4|2s \uparrow, 2s \downarrow\rangle$$



Denote the configurations in terms of their “excitations” from the Hartree-Fock one.

- Including all possible configurations (Full CI) leads to “exact” result—impractical!
- Instead, truncate to e.g. only singly and doubly-excited configurations (CISD).

Coupled Cluster Theory

Truncated CI models like CISD have systematic size-consistency errors.

- Error grows with system size.
- Stems from neglect of important higher-order excitations.

Coupled cluster (CC) theory is similar to CI, but cures this problem.

$$|\Psi_{CC}\rangle = e^{\hat{T}}|\Phi_{HF}\rangle = \left(1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{6}\hat{T}^3 \dots\right)|\Phi_{HF}\rangle$$

- **CCSD**: Coupled cluster theory with singles and doubles— $O(N^6)$ cost.
- **CCSDT**: Coupled cluster theory with singles, doubles, and triples— $O(N^8)$ cost. **Too Expensive!**

Compromise by estimating the triples contribution via perturbation theory.

- **CCSD(T)**: Coupled cluster theory with singles, doubles, and perturbative triples— $O(N^7)$ cost.
- CCSD(T) gives benchmark-quality results, but at a high computational cost.

How well do these models perform?

Molecular Geometries:

Bond Lengths: Mean errors, 29 small molecules (pm)

| Basis | HF | MP2 | CCSD | CCSD(T) |
|---------|-------|-------|-------|---------|
| cc-pVDZ | -1.80 | 1.35 | 1.07 | 1.68 |
| cc-pVTZ | -2.63 | -0.12 | -0.63 | 0.01 |
| cc-pVQZ | -2.74 | -0.23 | -0.79 | -0.12 |

Molecular Energetics:

17 Reaction Energies: Mean Abs. Errors (kJ/mol)

| Basis | HF | MP2 | CCSD | CCSD(T) |
|----------|------|------|------|---------|
| cc-pCVDZ | 43.8 | 36.5 | 30.2 | 40.3 |
| cc-pCVTZ | 42.3 | 16.0 | 10.5 | 12.7 |
| cc-pCVQZ | 43.0 | 12.9 | 12.0 | 3.6 |

19 Atomization Energies: Mean Errors (kJ/mol)

| Basis | HF | MP2 | CCSD | CCSD(T) |
|----------|--------|-------|--------|---------|
| cc-pCVDZ | -450.1 | -76.2 | -125.4 | -103.3 |
| cc-pCVTZ | -426.1 | -4.2 | -65.1 | -34.9 |
| cc-pCVQZ | -423.7 | 17.7 | -46.1 | -14.3 |

CCSD(T) generally gives most uniform errors (tightest error distributions).

Density functional theory (DFT), in a nutshell.

Hohenberg-Kohn theorems

- 1 There exists a 1:1 mapping between density $\rho(r)$ and energy.

$$\rho(r) \leftrightarrow \hat{H} \leftrightarrow E \ \& \ \Psi(r_1, r_2, \dots, r_n) \quad \Rightarrow \quad E = E[\rho(r)]$$

- 2 Can prove a variational principle for finding the optimal density.

Density functional theory (DFT), in a nutshell.

Hohenberg-Kohn theorems

- 1 There exists a 1:1 mapping between density $\rho(r)$ and energy.

$$\rho(r) \leftrightarrow \hat{H} \leftrightarrow E \ \& \ \Psi(r_1, r_2, \dots, r_n) \quad \Rightarrow \quad E = E[\rho(r)]$$

- 2 Can prove a variational principle for finding the optimal density.

Problem: We don't know what the true functional looks like!

Density functional theory (DFT), in a nutshell.

Hohenberg-Kohn theorems

- 1 There exists a 1:1 mapping between density $\rho(r)$ and energy.

$$\rho(r) \leftrightarrow \hat{H} \leftrightarrow E \ \& \ \Psi(r_1, r_2, \dots, r_n) \quad \Rightarrow \quad E = E[\rho(r)]$$

- 2 Can prove a variational principle for finding the optimal density.

Problem: We don't know what the true functional looks like!

Kohn-Sham DFT provides a workable solution for approximate density functionals.

$$E[\rho(r)] = T_s[\{\phi_i(r)\}] + J_{ee}[\rho(r)] + J_{eN}[\rho(r)] + E_{xc}[\rho(r)]$$

- Exchange-correlation functional $E_{xc}[\rho(r)]$ contains the key approximations.
- Different functionals differ in how they approximate $E_{xc}[\rho(r)]$.

Density functionals have been likened to Jacob's ladder.



Jacob's Ladder from the original Luther Bibles of 1534 and 1545

Generally expect quality of results to improve on higher rungs.

But often have significant variation across functionals in a rung.

Rungs of the Ladder

- 1 **Local density approximation:** Depend on $\rho(r)$.
LDA, SVWN

Density functionals have been likened to Jacob's ladder.



Jacob's Ladder from the original Luther Bibles of 1534 and 1545

Generally expect quality of results to improve on higher rungs.

But often have significant variation across functionals in a rung.

Rungs of the Ladder

- 1 **Local density approximation:** Depend on $\rho(r)$.
LDA, SVWN
- 2 **Generalized gradient approximation (GGAs):**
Also depend on $\nabla\rho(r)$.
PBE, BLYP, etc.

Density functionals have been likened to Jacob's ladder.



Jacob's Ladder from the original Luther Bibles of 1534 and 1545

Generally expect quality of results to improve on higher rungs.

But often have significant variation across functionals in a rung.

Rungs of the Ladder

- 1 **Local density approximation:** Depend on $\rho(r)$.
LDA, SVWN
- 2 **Generalized gradient approximation (GGAs):**
Also depend on $\nabla\rho(r)$.
PBE, BLYP, etc.
- 3 **meta-GGAs:** Also depend on $\nabla^2\rho(r)$.
TPSS, M06L, SCAN, etc.

Density functionals have been likened to Jacob's ladder.



Jacob's Ladder from the original Luther Bibles of 1534 and 1545

Generally expect quality of results to improve on higher rungs.

But often have significant variation across functionals in a rung.

Rungs of the Ladder

- 1 **Local density approximation:** Depend on $\rho(r)$.
LDA, SVWN
- 2 **Generalized gradient approximation (GGAs):**
Also depend on $\nabla\rho(r)$.
PBE, BLYP, etc.
- 3 **meta-GGAs:** Also depend on $\nabla^2\rho(r)$.
TPSS, M06L, SCAN, etc.
- 4 **Hybrid functionals:**
Add a fraction of exact/HF exchange.
B3LYP, PBE0, TPSSh, etc.

Density functionals have been likened to Jacob's ladder.



Jacob's Ladder from the original Luther Bibles of 1534 and 1545

Generally expect quality of results to improve on higher rungs.

But often have significant variation across functionals in a rung.

Rungs of the Ladder

- 1 **Local density approximation:** Depend on $\rho(r)$.
LDA, SVWN
- 2 **Generalized gradient approximation (GGAs):**
Also depend on $\nabla\rho(r)$.
PBE, BLYP, etc.
- 3 **meta-GGAs:** Also depend on $\nabla^2\rho(r)$.
TPSS, M06L, SCAN, etc.
- 4 **Hybrid functionals:**
Add a fraction of exact/HF exchange.
B3LYP, PBE0, TPSSh, etc.
- 5 **Double-Hybrid functionals:** Add MP2-like correlation.
B2PLYP, DSD-P86PBE, etc.

Standard density functionals do not describe van der Waals dispersion.

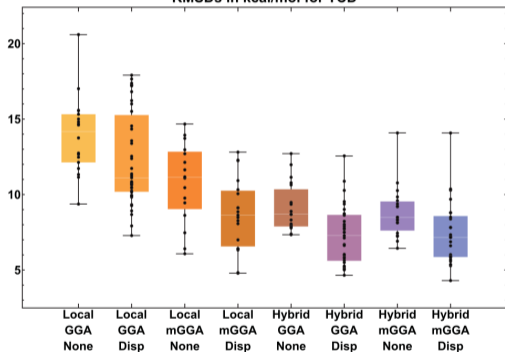
Augment DFT with a dispersion correction:

- Grimme's **D3** and **D4**.
 - Tkatchenko-Scheffler (**TS**) or Many-body Dispersion (**MBD**).
 - Becke and Johnson's Exchange-hole Dipole Moment (**XDM**).
 - A few functionals explicitly build in dispersion (ω **B97M-V**, etc).
-
- Current-generation corrections are largely non-empirical.
 - Most corrections are inexpensive and should generally be included.
 - They are crucial in systems where van der Waals interactions matter.
 - Example: Benzene crystal molar volume is **~45% too large** without dispersion correction!

Grimme, Hansen, Brandenburg, and Bannwarth. *Chem. Rev.* **116**, 5105 (2016).

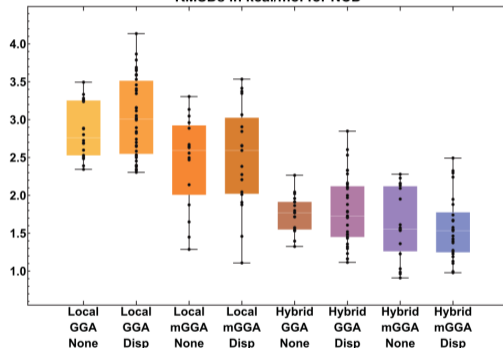
Reaction Thermochemistry

RMSDs in kcal/mol for TCD



Non-covalent interactions

RMSDs in kcal/mol for NCD



Mardirossian & Head-Gordon, *Mol. Phys.*, **115**, 2315 (2017).

Gaussian basis sets frequently used for molecular calculations.

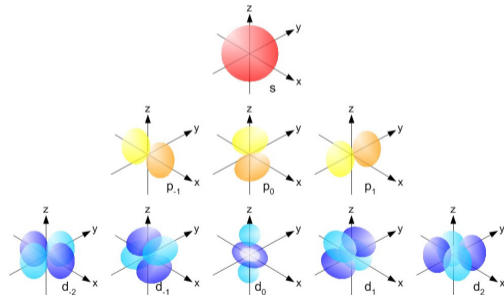
Express each molecular orbital as a linear combination of atomic orbitals (AOs).

$$|\phi\rangle = c_1|\chi_1\rangle + c_2|\chi_2\rangle + \cdots + c_n|\chi_n\rangle$$

- Each atomic orbital $\chi(r) = f(r)e^{-a|r|}$ represented via Gaussian functions $g(r)e^{-ar^2}$.

Gaussian basis set primer

- Minimal basis set:** Just the orbitals you'd expect based on periodic table. **Too small!**
- N-tuple zeta:** # of sets of AO functions. Double- ζ , triple- ζ , quadruple- ζ , etc.
- Polarization fns:** Higher angular momentum functions, allow density to polarize.
- Diffuse fns:** Large orbitals, saturate space between non-bonded atoms, or anions.
- Tight basis fns** can be important for special cases, e.g. some magnetic properties.



<http://chem.libretexts.org>

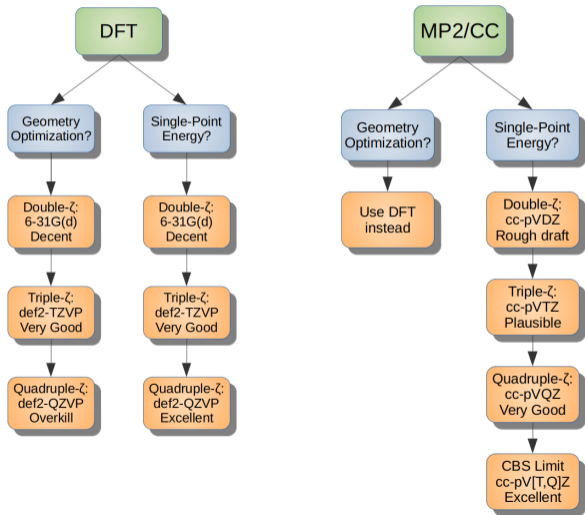
Here are some of the most common basis set families.

| Family | Basis Names | Comments |
|-----------------|---|--|
| Minimal | STO-3G, STO-6G | DON'T USE! |
| Pople | 6-31G(d), 6-31G(d,p) 6-311G(d), 6-311G(d,p) 6-311+G(d), 6-311++G(d,p) | Double- ζ , Smallest decent sets. Triple- ζ , moderately larger. The “+” adds diffuse basis functions. |
| Ahlrichs | def2-SVP, def2-TZVP, def2-QZVP | Double, triple, and quadruple- ζ . Good for DFT. |
| Jensen | pc- n (where $n = 1, 2, 3, 4$) | Another hierarchy that's good for DFT. |
| Dunning | cc-pVXZ, aug-cc-pVXZ (where $X = D, T, Q$) | Good hierarchies for MP2 and coupled cluster. The “aug-” adds diffuse basis functions. |

Can also find additional variants of these basis sets with further modifications.

My basic basis set recommendations.

The following rules of thumb should be helpful, but there can always be special cases.

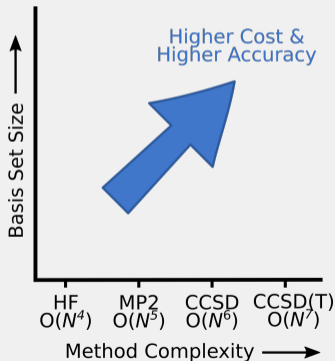


- Alternative basis sets of similar quality are also acceptable, especially for DFT.
- MP2 and coupled cluster need larger basis sets than DFT.
- DFT is usually “good enough” for geometry optimization.
- Add diffuse functions when needed: for anions, non-covalent interactions, etc.
e.g. 6-311+G(d,p), def2-XZVPD, or aug-cc-pVXZ
- Extra core basis functions are needed less often, but occasionally. e.g. cc-pwCVXZ, pcS-*n*

Perspective on the choice of a model chemistry

A **theoretical model chemistry** constitutes a chosen set of approximations used to describe your systems: Method, Basis Set, etc.

Model and Basis Set

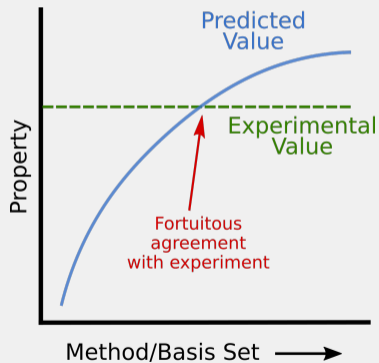


- **Challenge:** Best results come from high-quality methods **and** large basis sets—often cost-prohibitive.
- Chemical accuracy of ~ 1 kcal/mol is an aspirational goal.
 - Large-basis CCSD(T) can sometimes give this.
 - Most routine calculations will have larger errors.
- Composite approaches combine multiple calculations with different methods and basis sets to estimate high-quality method in the limit of large basis set.

Should your calculations agree with experiment?

Sometimes mediocre models match experiment better than higher-quality models.

Method/Basis Set vs. Expt.



Many approximations are involved beyond just method & basis set.

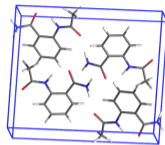
- Simplified molecular representation?
- Missing environment or solvent?
- Nuclear dynamics/vibrations?
- Finite temperature (H or G vs. E)?

Best practices:

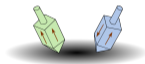
- Converge your calculations to the extent possible.
- Examine the sensitivity of your results to your modeling choices.
- Be aware of what approximations you're making and the role fortuitous error cancellation may play.

- 1 Overview of Standard Quantum Chemistry Models
 - The molecular Schrödinger equation
 - Born-Oppenheimer approximation
 - Wave function methods
 - Density functional theory
 - Basis sets
 - Predictions versus reality
- 2 Selected Applications (biased by my own research)
 - Theory aiding experiment: NMR crystallography
 - Theory leading experiment: Photomechanical engines

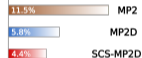
- 1 How do organic molecules pack in solids?
- 2 How do we build new models that predict these properties accurately and efficiently?
- 3 Can we predict crystalline properties to aid experiment?



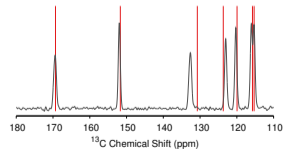
Scale MP2 spin components and correct dispersion



Relative Error

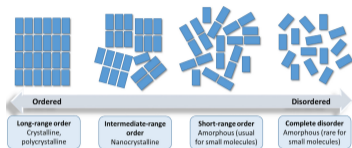


Candidate Structure 3



Solving crystal structures isn't always easy.

Not all materials can be fully characterized via traditional X-ray diffraction techniques.



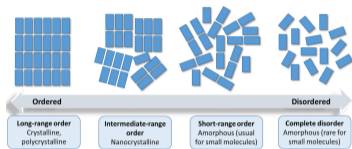
④ Some materials exhibit only short-range order.

e.g. Amorphous pharmaceuticals.

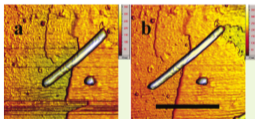
Peltonen & Strachan. *Int. J. Pharm.* **586**, 119492 (2020).

Solving crystal structures isn't always easy.

Not all materials can be fully characterized via traditional X-ray diffraction techniques.



Peltonen & Strachan. *Int. J. Pharm.* **586**, 119492 (2020).



Al-Kaysi et al. *J. Am. Chem. Soc.* **128**, 15938 (2006).

- 1 Some materials exhibit only short-range order.

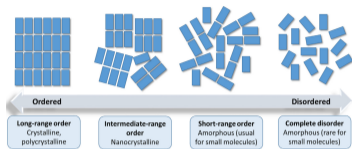
e.g. Amorphous pharmaceuticals.

- 2 Others can be hard to solve via powder X-ray diffraction.

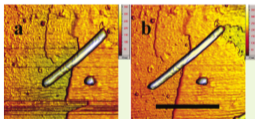
e.g. Dynamic organic crystal nanomaterials.

Solving crystal structures isn't always easy.

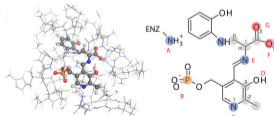
Not all materials can be fully characterized via traditional X-ray diffraction techniques.



Peltonen & Strachan. *Int. J. Pharm.* **586**, 119492 (2020).



Al-Kaysi et al. *J. Am. Chem. Soc.* **128**, 15938 (2006).



Caulkins et al, *J. Am. Chem. Soc.* **138**, 15214 (2016).

1 Some materials exhibit only short-range order.

e.g. Amorphous pharmaceuticals.

2 Others can be hard to solve via powder X-ray diffraction.

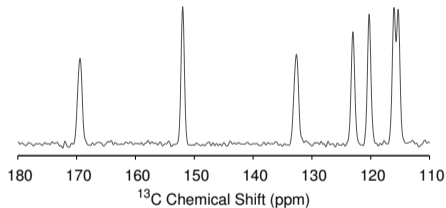
e.g. Dynamic organic crystal nanomaterials.

3 Diffraction cannot always reveal the salient features.

e.g. Catalytically important protons in the active site of tryptophan synthase.

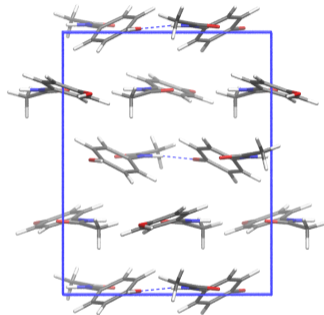
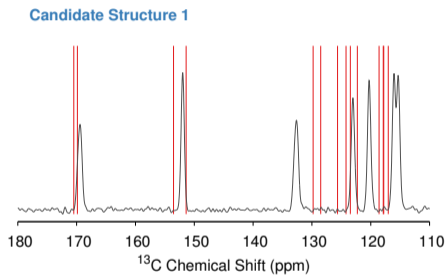
NMR crystallography can help in these situations.

Strategy: Identify candidate crystal structures from PXRD or crystal structure prediction whose simulated spectroscopic properties match experimentally observed ones.



NMR crystallography can help in these situations.

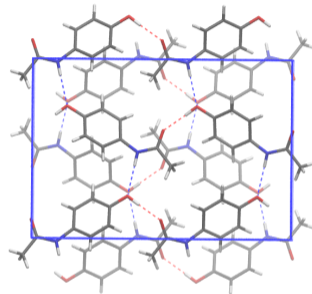
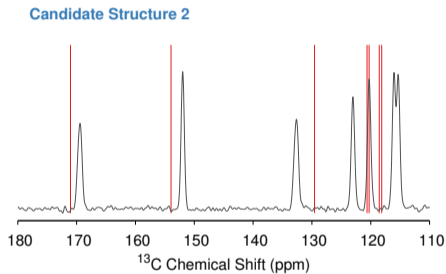
Strategy: Identify candidate crystal structures from PXRD or crystal structure prediction whose simulated spectroscopic properties match experimentally observed ones.



| Candidate 1 | |
|-----------------|-----|
| RMS Error (ppm) | 2.1 |

NMR crystallography can help in these situations.

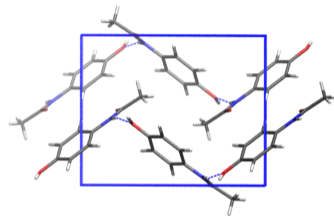
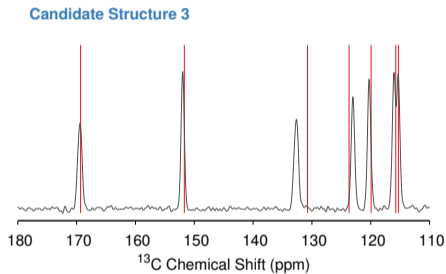
Strategy: Identify candidate crystal structures from PXRD or crystal structure prediction whose simulated spectroscopic properties match experimentally observed ones.



| | Candidate 1 | Candidate 2 |
|-----------------|-------------|-------------|
| RMS Error (ppm) | 2.1 | 2.6 |

NMR crystallography can help in these situations.

Strategy: Identify candidate crystal structures from PXRD or crystal structure prediction whose simulated spectroscopic properties match experimentally observed ones.

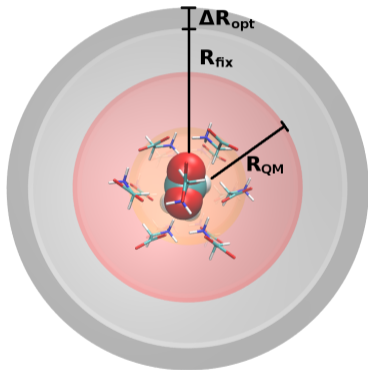


| | Candidate 1 | Candidate 2 | Candidate 3 |
|-----------------|-------------|-------------|-------------|
| RMS Error (ppm) | 2.1 | 2.6 | 0.8 |

Higher accuracy prediction = increased discrimination between candidates.

Our fragment methods enable higher-accuracy chemical shift predictions.

Replace a single large calculation on the full system with many smaller sub-system calculations.



2-Body Embedded Fragment Approach

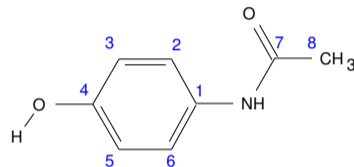
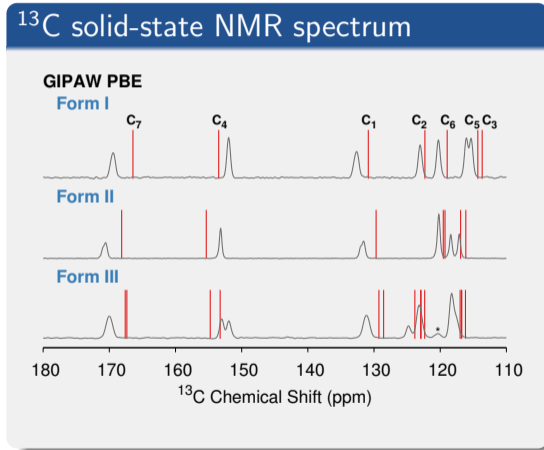
$$\tilde{\sigma}_i^A \approx \sigma_i^{A,emb.} + \sum_j \Delta^2 \sigma_{ij}^{A,emb.}$$

- Compute shieldings for central molecule
- Add series of corrections due to pairwise interactions with nearby molecules.
- Embed all monomer/dimer calculations in **self-consistently polarized point charges** or **polarizable continuum model (PCM)** to capture many-body polarization effects.

Enables use of more expensive hybrid DFT functionals that reduce errors vs. experiment by ~30–50% compared to standard GGAs.

Hybrid PBE0 improves discrimination for acetaminophen polymorphs.

Fragment PBE0 chemical shifts exhibit smaller errors than periodic GIPAW PBE ones.



RMS Errors (ppm)

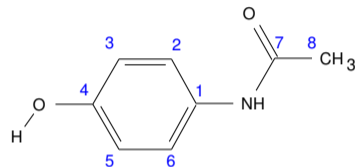
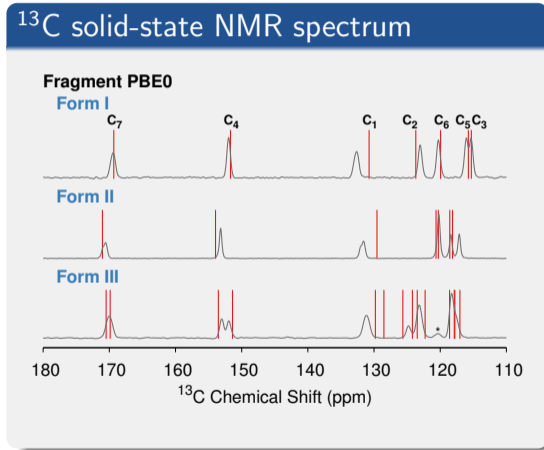
| | |
|---------------|-----|
| GIPAW PBE | 2.0 |
| Fragment PBE0 | 1.1 |

Experiment: Burley, Duer, Stein, and Vrclj. *Eur. J. Pharm. Sci.* **31**, 271–276 (2007).

Hartman, Day, and Beran. *Cryst. Growth Des.* **16**, 6479 (2016).

Hybrid PBE0 improves discrimination for acetaminophen polymorphs.

Fragment PBE0 chemical shifts exhibit smaller errors than periodic GIPAW PBE ones.



RMS Errors (ppm)

| | |
|---------------|-----|
| GIPAW PBE | 2.0 |
| Fragment PBE0 | 1.1 |

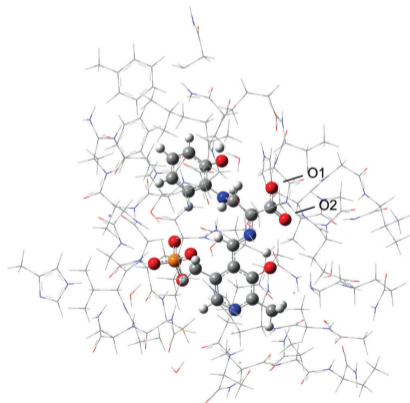
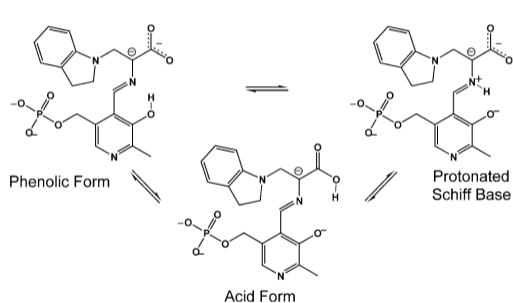
Experiment: Burley, Duer, Stein, and Vrclj. *Eur. J. Pharm. Sci.* **31**, 271–276 (2007).

Hartman, Day, and Beran. *Cryst. Growth Des.* **16**, 6479 (2016).

NMR crystallography is solving the enzyme mechanisms.

Protonation states at multiple sites play key role in the mechanism of tryptophan synthase and related proteins.

- ^{17}O shifts for O1 and O2 occur outside typical $\text{COOH} / \text{COO}^-$ shift ranges, hindering interpretation.

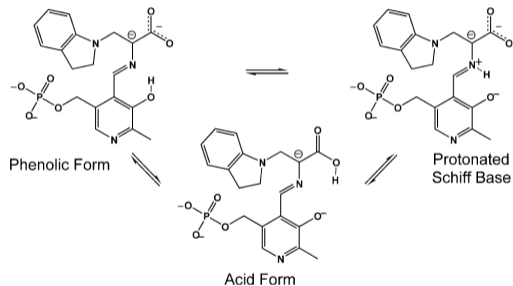


Young, Caulkins, Borchardt, Bulloch, Larive, Dunn, and Mueller. *Angew. Chem. Int. Ed.* **55**, 1350 (2016).

NMR crystallography is solving the enzyme mechanisms.

Protonation states at multiple sites play key role in the mechanism of tryptophan synthase and related proteins.

- ^{17}O shifts for O1 and O2 occur outside typical $\text{COOH} / \text{COO}^-$ shift ranges, hindering interpretation.



Predicted Isotropic Shifts (ppm)

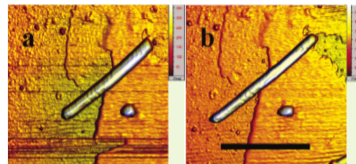
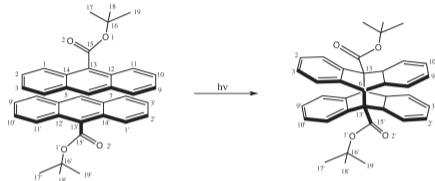
| Structure | $\delta(\text{O1})$ | $\delta(\text{O2})$ |
|----------------------|---------------------|---------------------|
| Experiment | 237.2 | 238.7 |
| Phenolic Form | 238.0 | 243.5 |
| Schiff Base Form | 228.5 | 233.6 |
| Acid Form | 171.3 | 113.1 |

^{17}O NMR experiments & chemical shift scaling parameters derived from our molecular crystal studies reveal fast-exchange dominated by the phenolic form.

Photomechanical materials transform light into mechanical work.

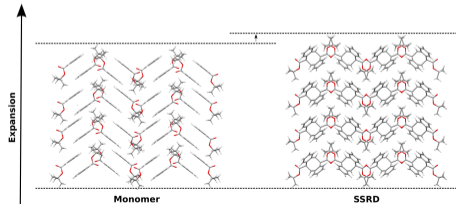
9-tert-butyl anthracene ester nanorods elongate $8 \pm 2\%$ upon irradiation.

Al-Kaysi, Müller, & Bardeen. *J. Am. Chem. Soc.* **128**, 15938 (2006).



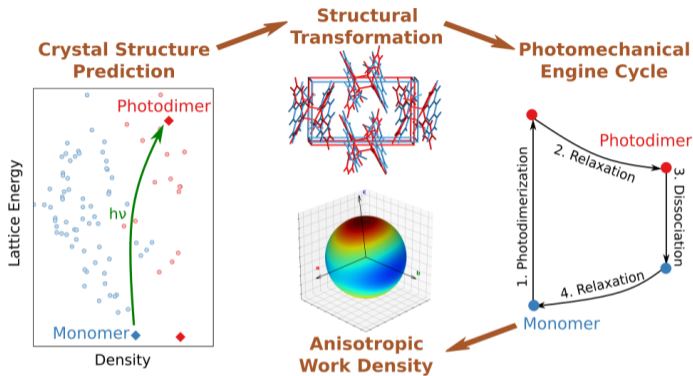
- Want to make light-powered nano-machines.
- Experimentally difficult to:
 - Determine photochemical structural changes.
 - Anticipate how altering the molecular will impact the photomechanical response.
- Can theory accelerate then slow trial-and-error experimental efforts?

NMR crystallography revealed the elongation mechanism.



Chalek et al. *Chem. Sci.* **12**, 453 (2021).

Can we design improved photomechanical materials *in silico*?



Use quantum chemistry to lead experiment and investigate:

- How will a molecule pack as a crystal?
- How will the crystal unit cell transform upon photochemical reaction?
- How much work can we extract from a photomechanical engine?

Cook, Li, Lui, Gately, Al-Kaysi, Mueller, Bardeen, & Beran. *Chem. Sci.* **14**, 937-949 (2023).

Cook, Perry, & Beran. *submitted* (2023).

Take-away messages.

- 1 **Quantum Chemistry is routinely making predictions that complement or lead experiment.**
 - Examples include organic chemistry, biochemistry, materials, etc.
- 2 **Many different quantum chemistry methods exist.**
 - DFT, MP2, CCSD(T), etc.
 - DFT offers a “sweet-spot” in terms of computational cost vs. accuracy.
 - But many functionals to choose from (Jacob’s Ladder).
- 3 **Best Practice: Investigate how your modeling decisions impact your predictions.**
 - Assess sensitivity/uncertainty.
 - Build confidence in your predictions.

U.C. Riverside Chemistry PhD Program

Our PhD Program at a glance:

- ~35 faculty, ~140 PhD students in Chemistry.
- >\$12 million in external grants & 330 papers published in 2021.
- Ranked 37th in U.S. in the (data-driven) Shanghai Rankings.
- 1 Nobel Laureate (Schrock), many ACS Award winners, etc.
- One of the most diverse research universities in the nation.
- Strong focus on research and individual mentoring.



Take advantage of everything Southern California has to offer:



Molecular systems

- Gaussian/GaussView (\$)
- Psi4
- Q-Chem (\$)
- Molpro (\$)
- Orca

Periodic systems

- VASP (\$)
- QuantumEspresso
- Crystal (\$)
- CASTEP
- FHI Aims (\$)

Further Reading:

- **Introduction to Computational Chemistry, 3rd Ed.**, F. Jensen, Wiley, 2017.
- **Essentials of Computational Chemistry, 2nd Ed.** C. Cramer, Wiley, 2004.
- **The ABC of DFT** K. Burke.
<https://dft.uci.edu/doc/g1.pdf>
- **Density Functional Theory: A Practical Introduction**, D. Sholl and J. Steckel, Wiley, 2009.