



Applied Computational Chemistry

Lecture 1

Fernanda Duarte

(fernanda.duartegonzalez@chem.ox.ac.uk)



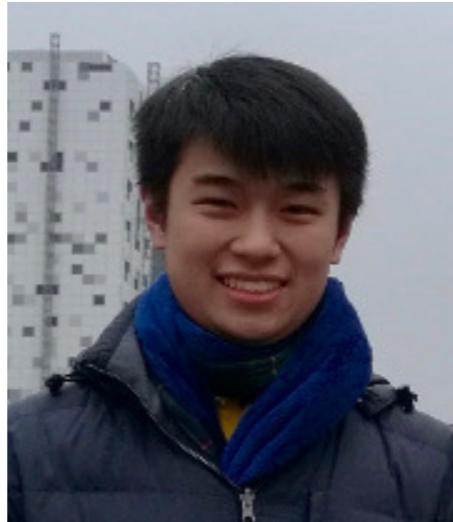
Twitter @duarte_group

duartegroupchem.org



Github: duartegroup

Demonstrators



Henry Chan
Cohort 2018



Aleksy Kwiatkowski
Cohort 2021

Synthesis for Biology and Medicine CDT

Timetable

	Tuesday	Wednesday	Thursday	Friday
10:00-11:00	Lecture 1	Lecture 3		Preparation for presentation
11:15-12:15	Lecture 2	Project Work	Project Work	Presentations
14:00- 15:00	Lab session	Project Work	Project Work	
15:30- 17:00	Lab session			

Workplan

Lab session

Work through them as you get familiar with the softwares

- Warm up: calculation of water
- Conformational analysis
- Transition state optimisation

Project Work

Work on the projects you have chosen (3 people)

- Read the literature (exp. & comp data) associated to it
- Reproduce some of the calculations
- Explore any novel aspect that you may find relevant

Presentations

- 15 min per group + 5 min questions
- Background/ rationale behind the technique used
- Main findings /conclusions
- Any novel aspect explored will be a “+”

- Familiarise yourself with computer modelling and theoretical methods
- Understanding chemistry literature that combine computation and experiments
- Learn the basic aspects of computational chemistry to collaborate effectively with computational chemistry colleagues
- Carry out basic computational chemistry projects with a good understanding of the methodologies, pros and cons

Outline

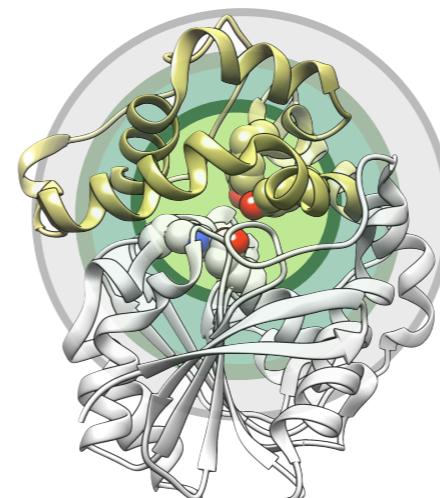
Lecture 1

- What is - and why is it relevant?
- A bit of history
- DFT/WFT
- Basis sets



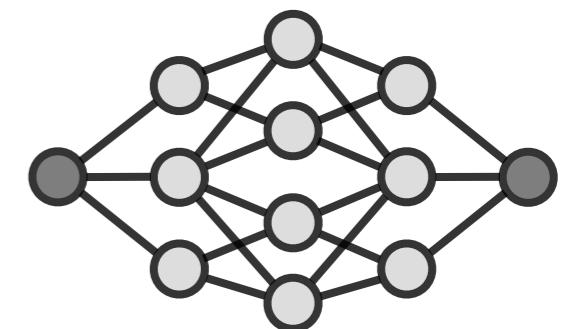
Lecture 2

- Recap of Day 1
- The zoo of DFT functionals
- MD and QM/MM



Lecture 3

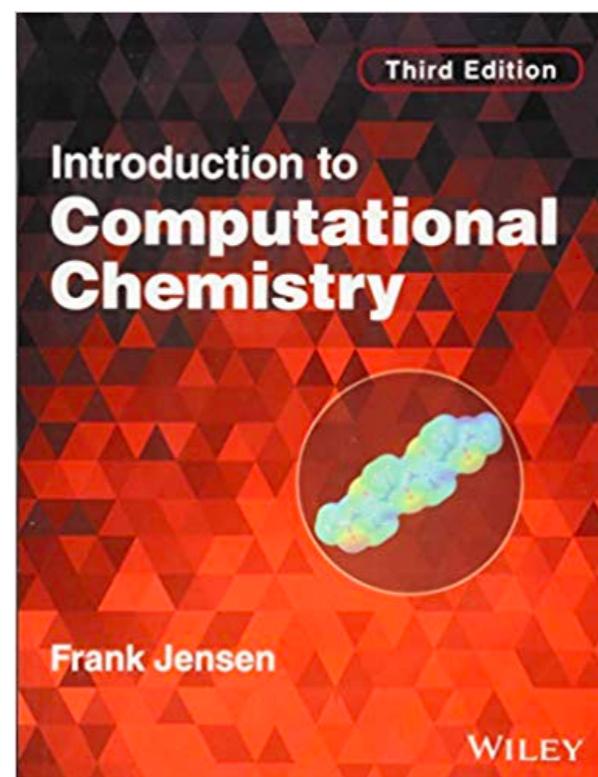
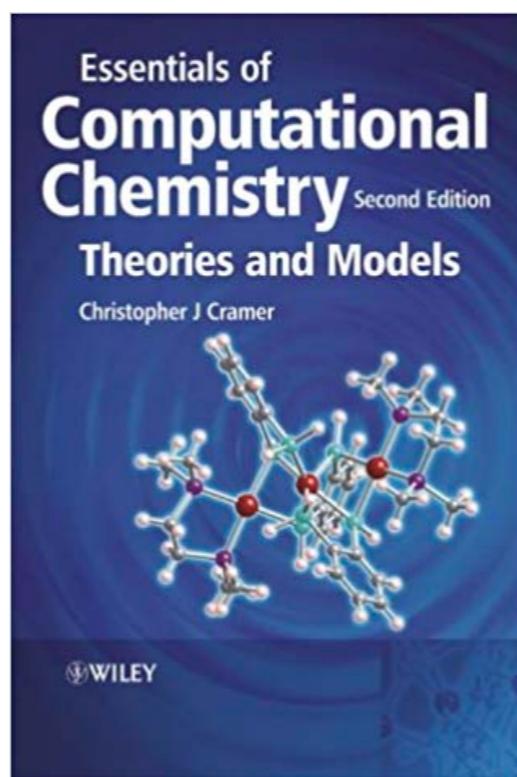
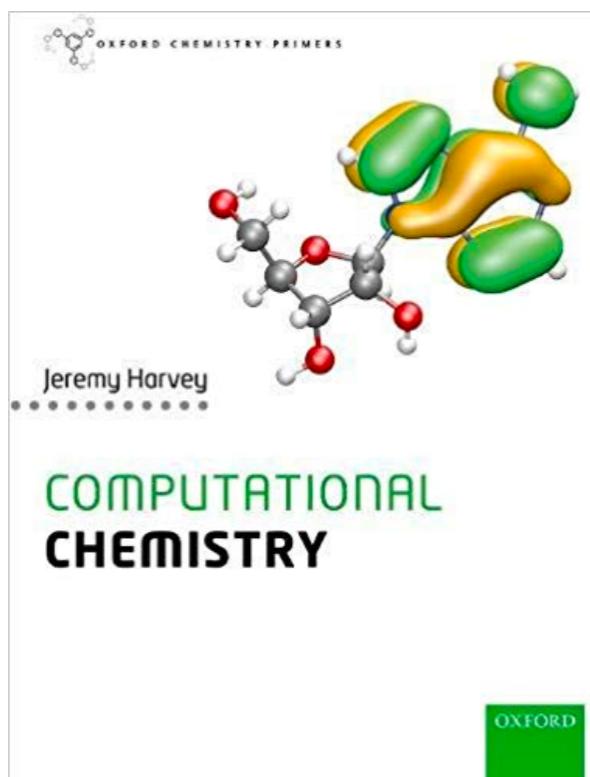
- Recap of Day 2
- Applications & Challenges
- Machine Learning



Aims

- Familiarise yourself with computer modelling and theoretical methods
- Understanding chemistry literature that combine computation and experiments
- Learn the basic aspects of computational chemistry to collaborate effectively with computational chemistry colleagues
- Carry out basic computational chemistry projects with a good understanding of the methodologies, pros and cons

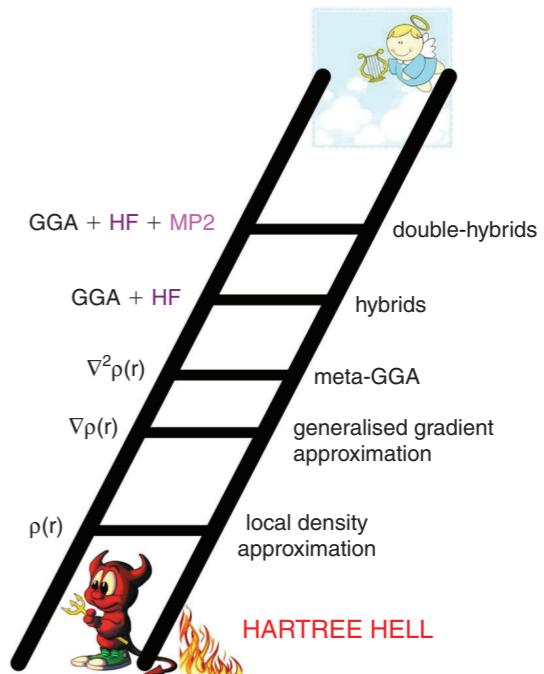
Useful References



Useful References

A Trip to the Density Functional Theory Zoo: Warnings and Recommendations for the User
Aust. J. Chem. **2019**, *72*, 563
<https://doi.org/10.1071/CH19023>

HEAVEN OF CHEMICAL ACCURACY



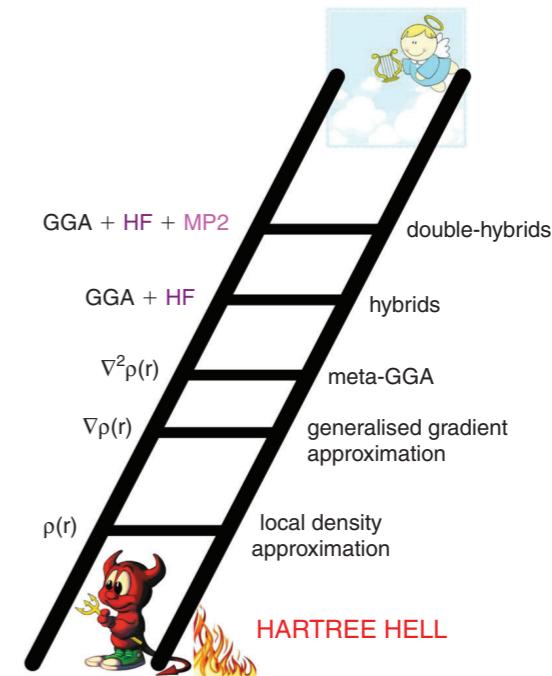
Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals
Mol. Phys. **2017**, *115*, 2315
<https://doi.org/10.1080/00268976.2017.1333644>



Wednesday Quiz

A Trip to the Density Functional Theory Zoo: Warnings and Recommendations for the User
Aust. J. Chem. **2019**, *72*, 563
<https://doi.org/10.1071/CH19023>

HEAVEN OF CHEMICAL ACCURACY



1. How are DFAs classified?
2. What is understood by chemical accuracy?
3. Describe with examples known inaccuracies in DFT
4. How can dispersion be accounted for?
5. What is BSSE and what is its effect on structural and chemical properties?
6. Best phrase of the paper?

Computational Chemistry

Computational chemistry is a branch of chemistry that uses computer simulation to assist in solving chemical problems (Wikipedia)

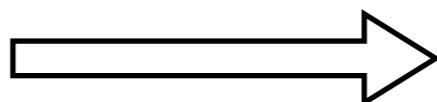
Computational Chemistry

Computational chemistry is a branch of chemistry that uses computer simulation to assist in solving chemical problems (Wikipedia)

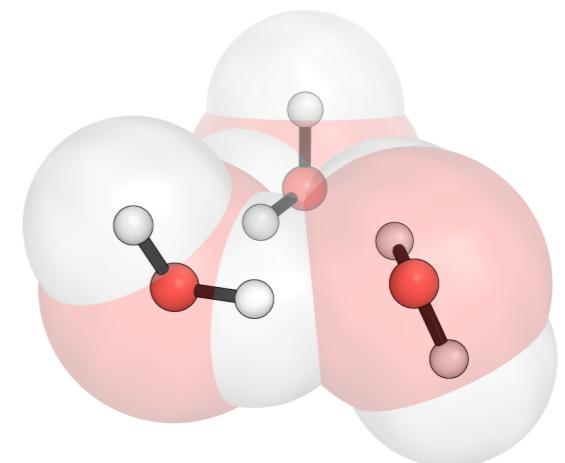
Macroscopic Observables



Interpret



Electronic/Molecular Structure



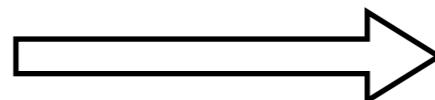
Computational Chemistry

Computational chemistry is a branch of chemistry that uses computer simulation to assist in solving chemical problems (Wikipedia)

Macroscopic Observables

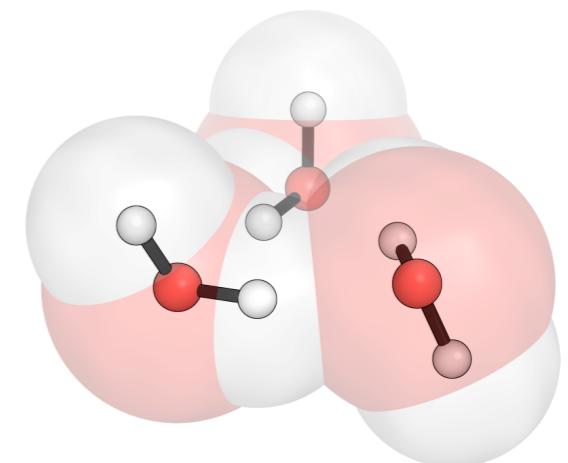
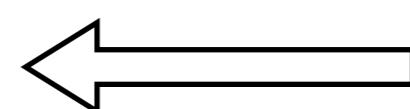


Interpret

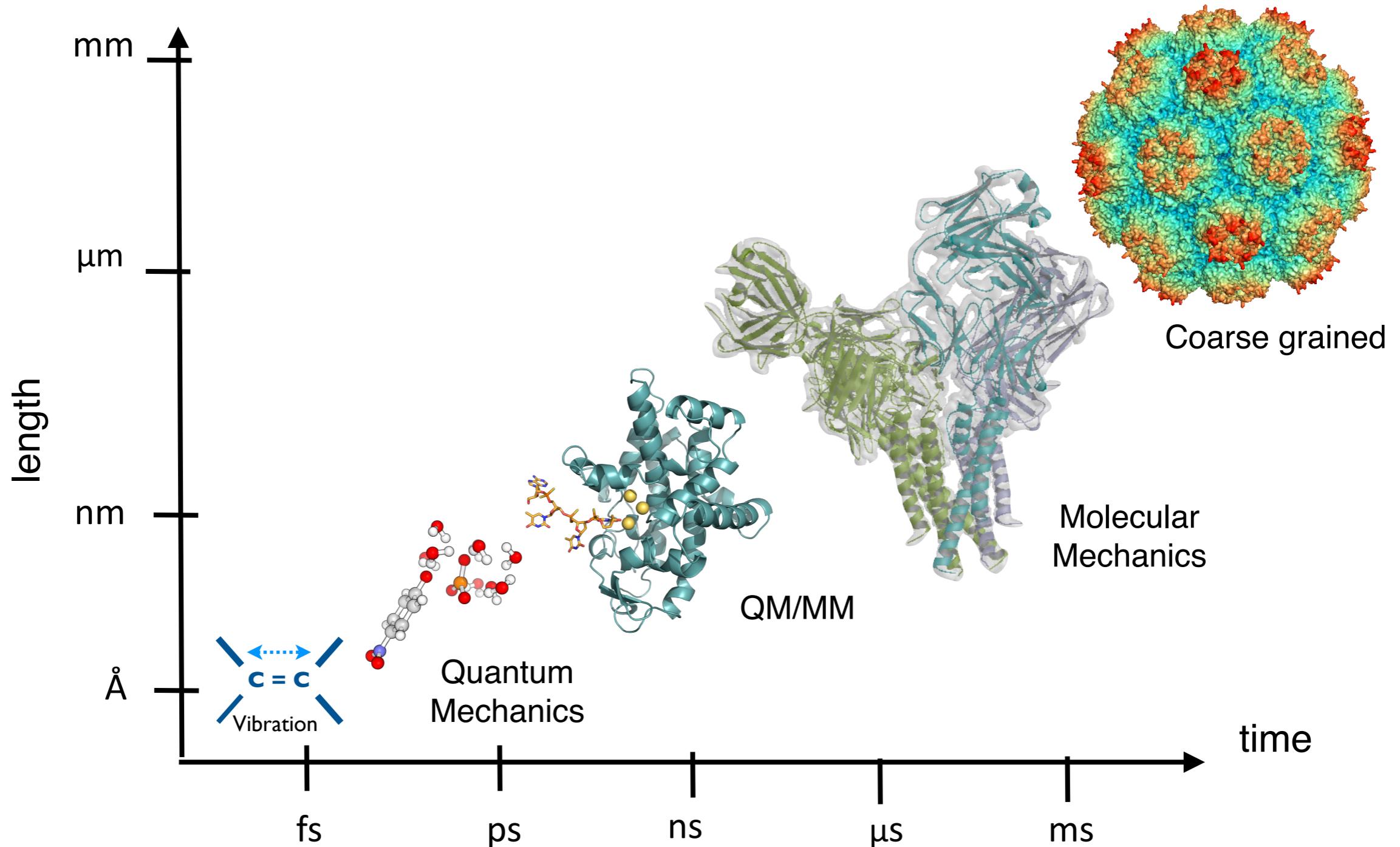


Electronic/Molecular Structure

Predict



Two schools

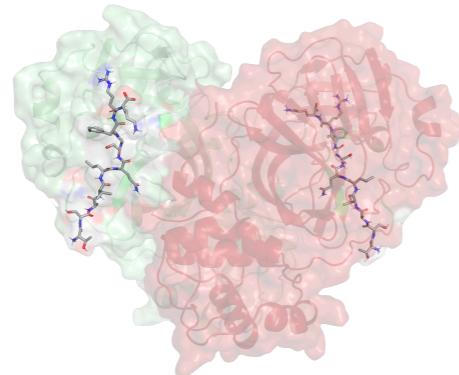


“Everything should be made as simple as possible, but not simpler.”

Two schools

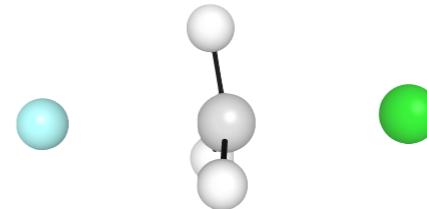
Interpretability | Accuracy | Fast predictions

Molecular Mechanics/EVB



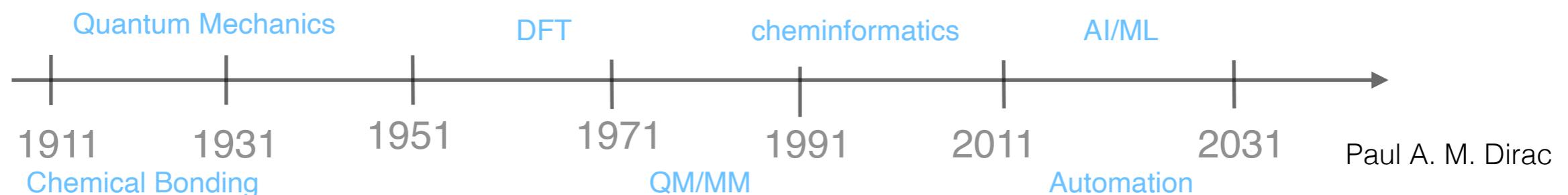
- Springs and balls
- Solving Newton's equation
- Non-reactive (exceptions EVB, ReaxFF)
- Accurate for small displacements
- Computationally efficient

DFT/WFT Approaches



- Electrons
- Solving Schrödinger's equation
- Electronic structure description
- Limited sampling
- Computationally costly

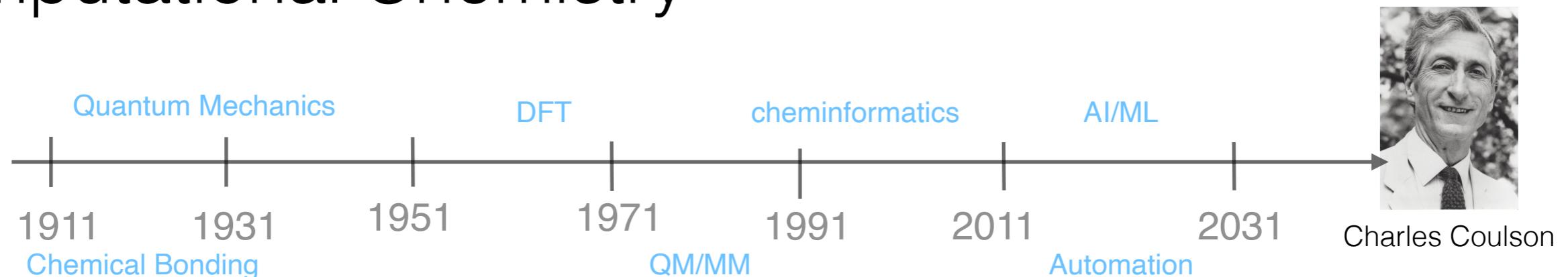
Computational Chemistry



The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that the exact application of these laws leads to equations much too complicated to be soluble.

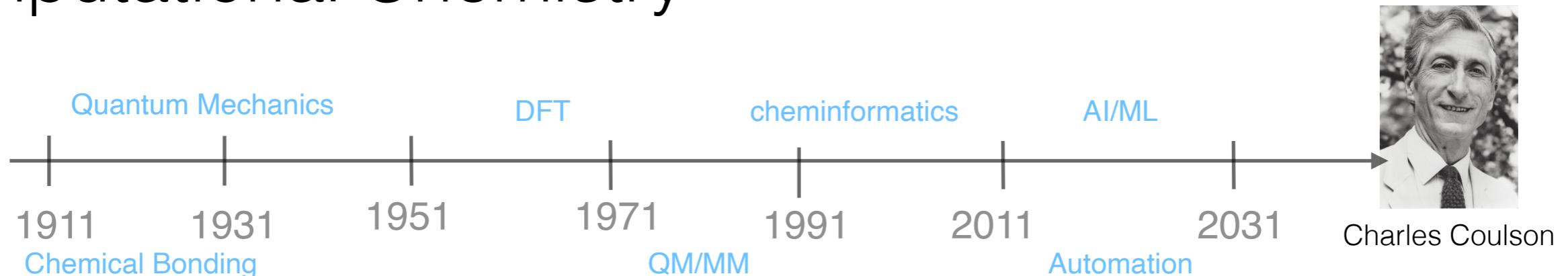
It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation

Computational Chemistry



15

Computational Chemistry



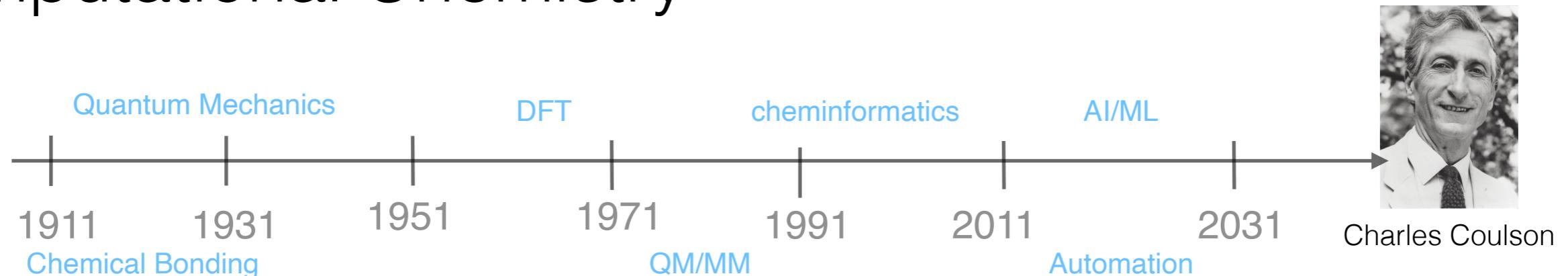
“Give us insight, not numbers”
(1959, Colorado Boulder)

“*ab-initio-ists*”

Those who care for the most **accurate** approximate solution to the Schrödinger eq.

$$\hat{H}_e \psi_e = E_e \psi_e$$
$$\hat{H}_e = \frac{-\hbar^2}{2m} \sum_i^{electrons} \nabla_i^2 - \sum_i^{electrons} \sum_A^{nuclei} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i < j}^{electrons} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

Computational Chemistry



“Give us insight, not numbers”
(1959, Colorado Boulder)

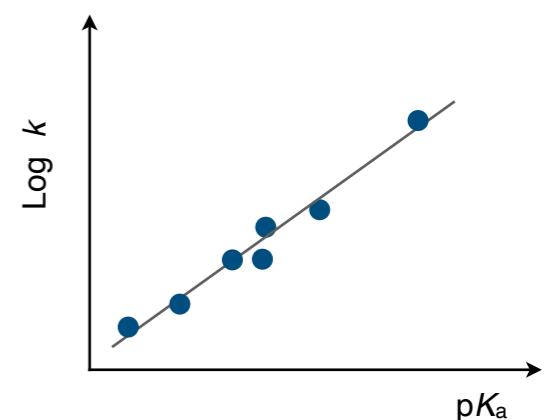
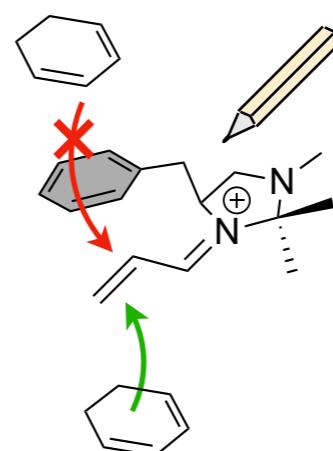
“*ab-initio-ists*”

Those who care for the most **accurate** approximate solution to the Schrödinger eq.

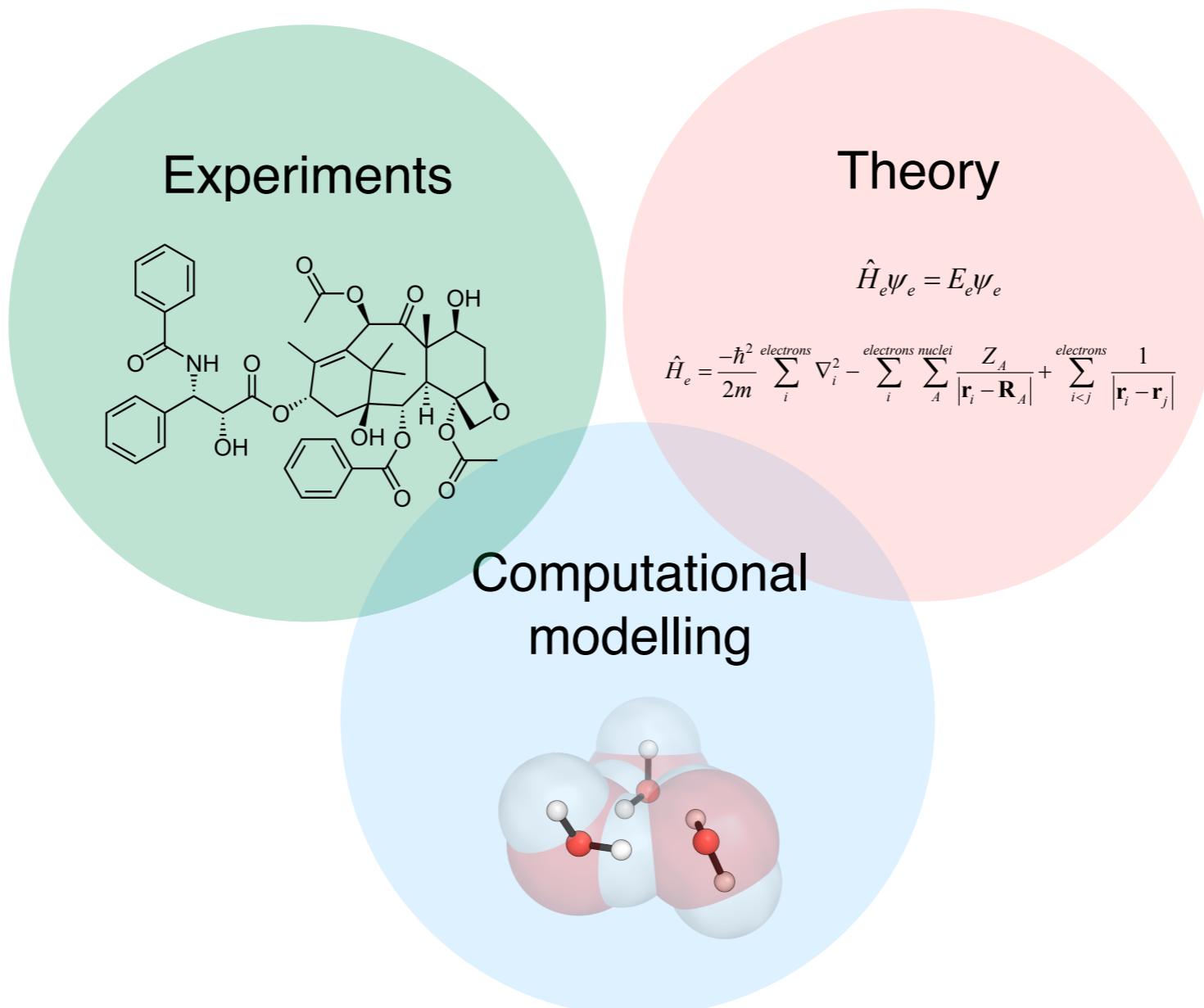
$$\hat{H}_e \psi_e = E_e \psi_e$$
$$\hat{H}_e = \frac{-\hbar^2}{2m} \sum_i^{electrons} \nabla_i^2 - \sum_i^{electrons} \sum_A^{nuclei} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i < j}^{electrons} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

“*posteriori-ists*”

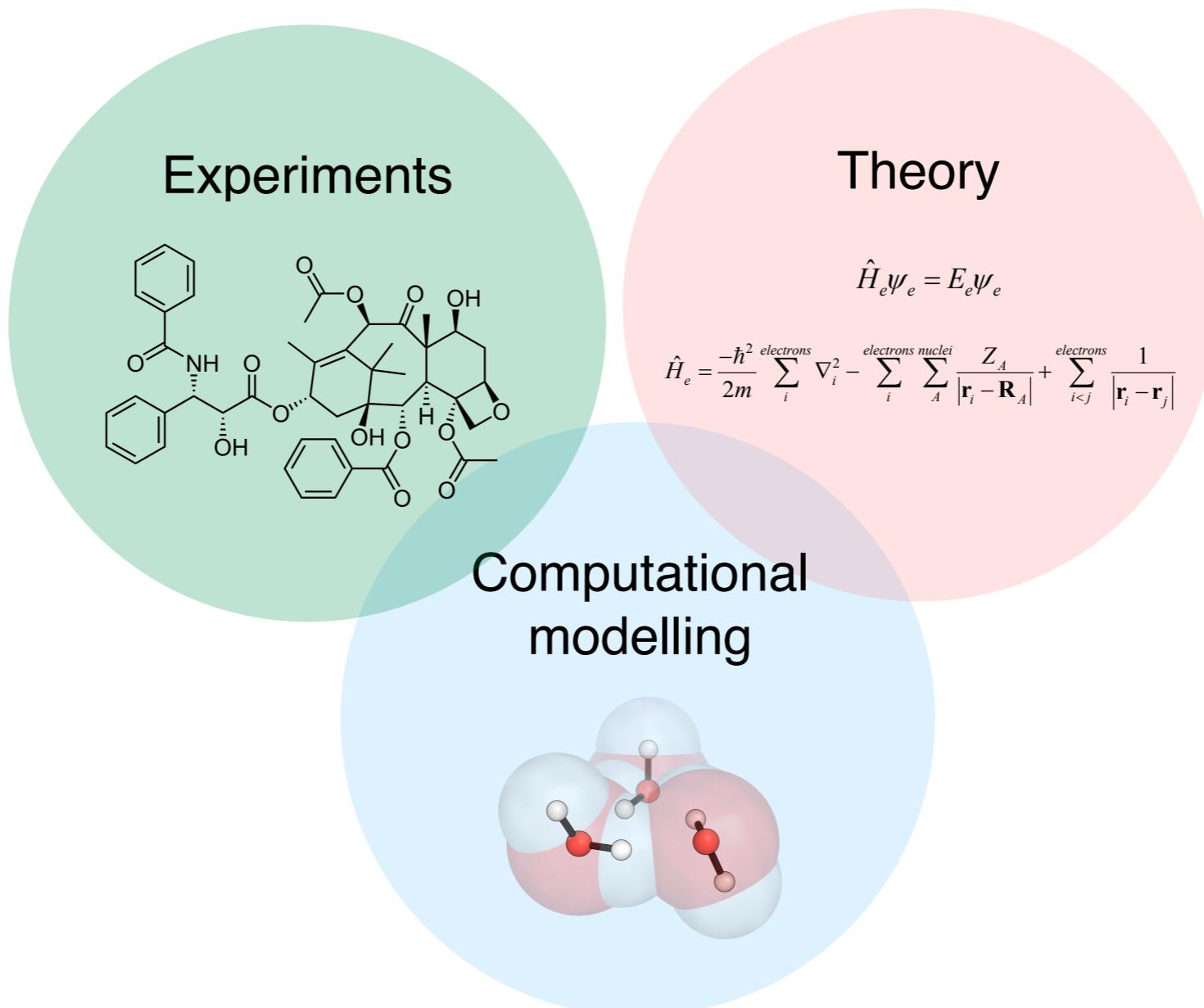
Those who are interested in **chemical concepts** and trends



Computational Chemistry



Computational Chemistry

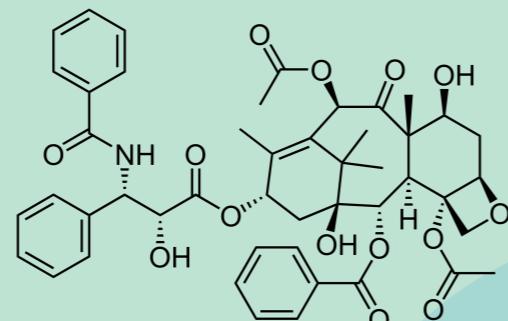


Computational Chemistry

and

“Give us insight, ~~not~~ numbers”

experimentalists



ab-initio-ists

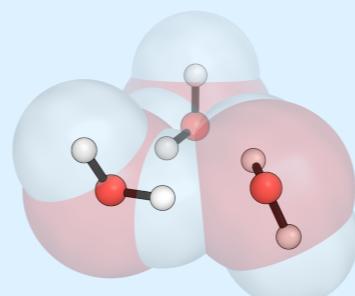
$$\hat{H}_e \psi_e = E_e \psi_e$$

$$\hat{H}_e = -\frac{\hbar^2}{2m} \sum_i^{\text{electrons}} \nabla_i^2 - \sum_i^{\text{electrons}} \sum_A^{\text{nuclei}} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i < j}^{\text{electrons}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

posteriori-ists

Machine Learning

Virtual Reality



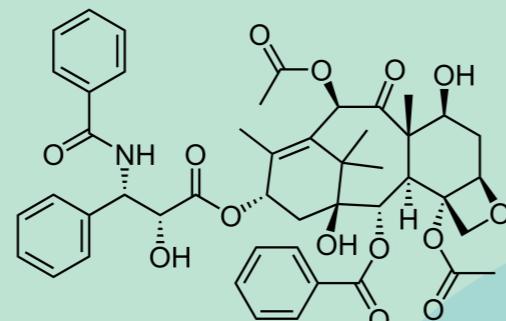
HTE

Computational Chemistry

and

“Give us insight, ~~not~~ numbers”

experimentalists



ab-initio-ists

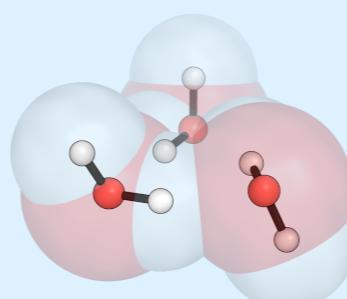
$$\hat{H}_e \psi_e = E_e \psi_e$$

$$\hat{H}_e = -\frac{\hbar^2}{2m} \sum_i^{\text{electrons}} \nabla_i^2 - \sum_i^{\text{electrons}} \sum_A^{\text{nuclei}} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i < j}^{\text{electrons}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

posteriori-ists

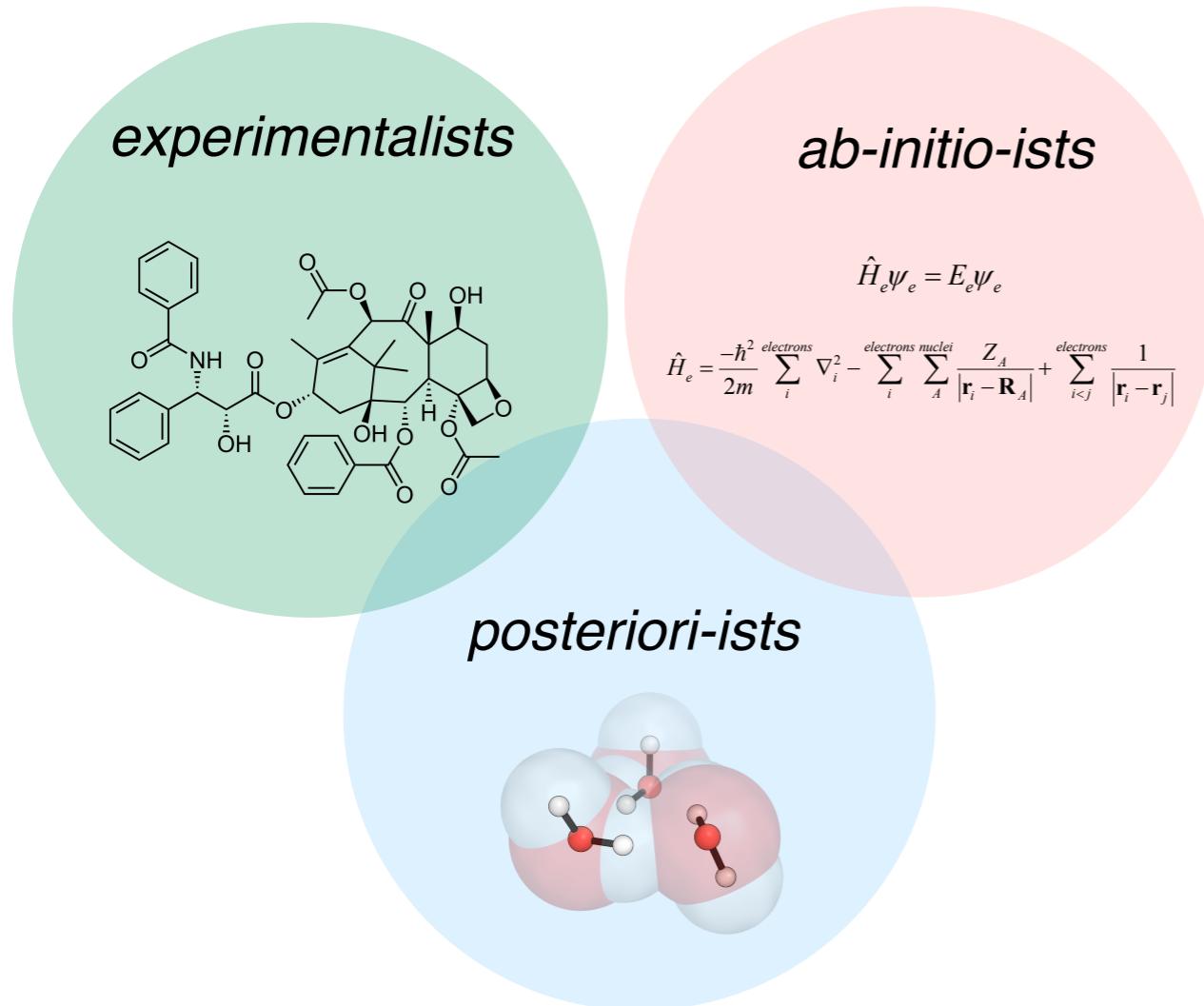
Machine Learning

Virtual Reality



HTE

Computational Chemistry

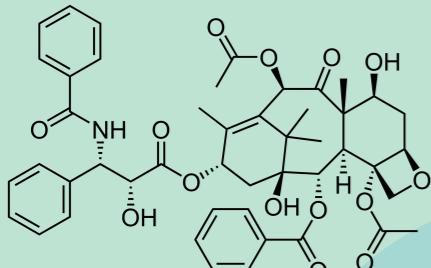


- Shorter lead-time to new drugs or materials
- Help in the characterisation of novel molecules
- Enhanced knowledge of molecular properties
- Scoring and refinement of molecular libraries
- Optimisation of existing processes
- Rationalisation of experimental outcomes
- Prediction of new properties

Grand Challenges in Computational Chemistry

“Give us insight, not numbers” **and** error bars

experimentalists

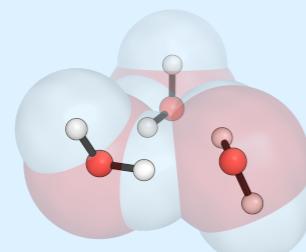


ab-initio-ists

$$\hat{H}_e \psi_e = E_e \psi_e$$

$$\hat{H}_e = -\frac{\hbar^2}{2m} \sum_i^{\text{electrons}} \nabla_i^2 - \sum_i^{\text{electrons}} \sum_A^{\text{nuclei}} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i < j}^{\text{electrons}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

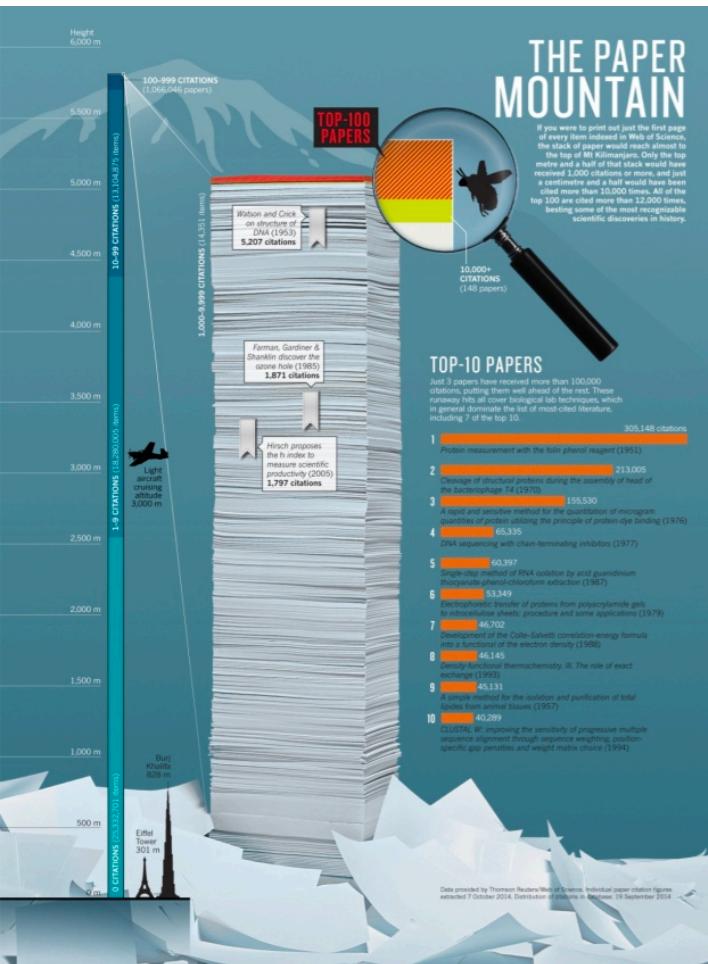
posteriori-ists



- Robust and fast electronic structure methods
- Automated approaches for finding new reactions
- Inclusion of solvation effects
- Accurate treatments of entropy
- Prediction of molecular and macroscopic properties

The impact of DFT/WFT approaches

The 100 most-cited research of all time



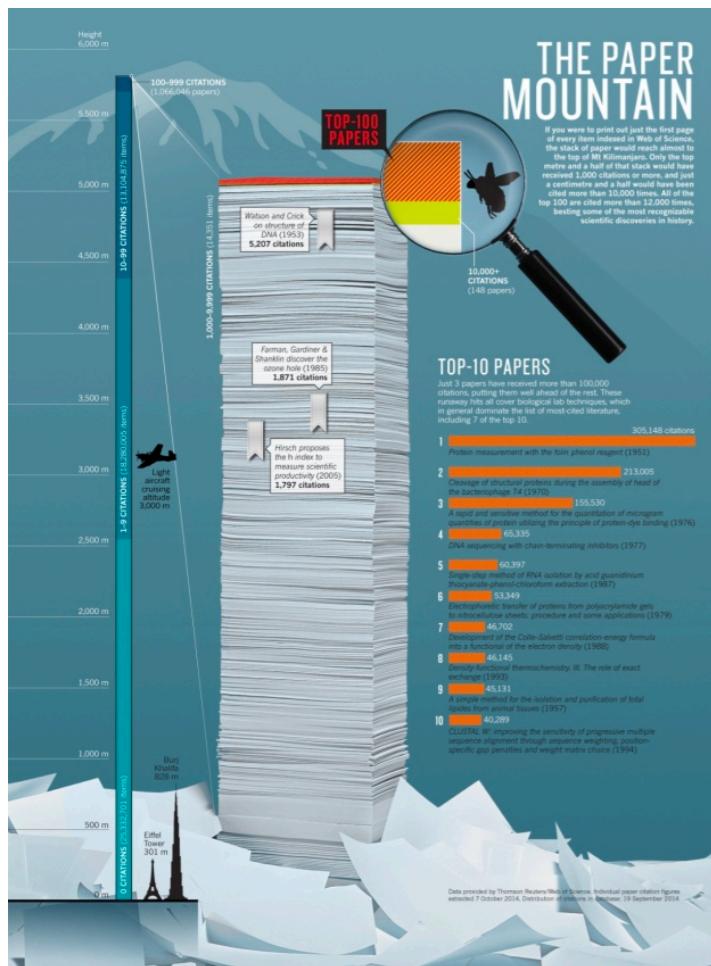
1. Protein measurement with the folin phenol reagent.

J. Biol. Chem. **1951**, *193*, 265.

Lowry O. H., Rosebrough N. J. Farr A. L. & Randall, R. J.

The impact of DFT/WFT approaches

The 100 most-cited research of all time



DFT/Computational Chemistry

1. *Protein measurement with the folin phenol reagent.*
J. Biol. Chem. **1951**, *193*, 265.
Lowry O. H., Rosebrough N. J. Farr A. L. & Randall, R. J.
7. *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.*
Phys. Rev. B **1988**, *37*, 785.
Lee C., Yang W.* & Parr R. G. (*also 93)
8. *Density-functional thermochemistry. III. The role of exact exchange*
J. Chem. Phys. **1993**, *98*, 5648.
Becke, A. D. (also 25)
35. *Self-consistent equations including exchange and correlation effects.*
Phys. Rev. **1965**, *140*, A1133.
Kohn, W. & Sham, L. J. (also 39)

B3LYP

How do we use it?

Which System Do I Have?



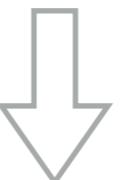
What Do I Want to Compute (and Why)?



Which Model /Method Should I Choose?



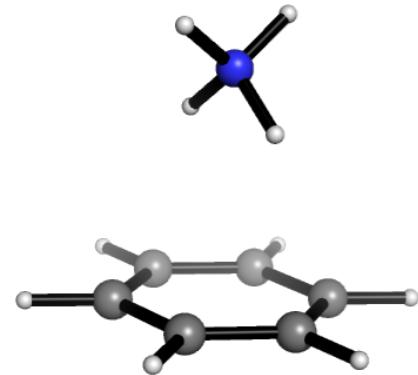
Verify Approach (vs. Experiment)



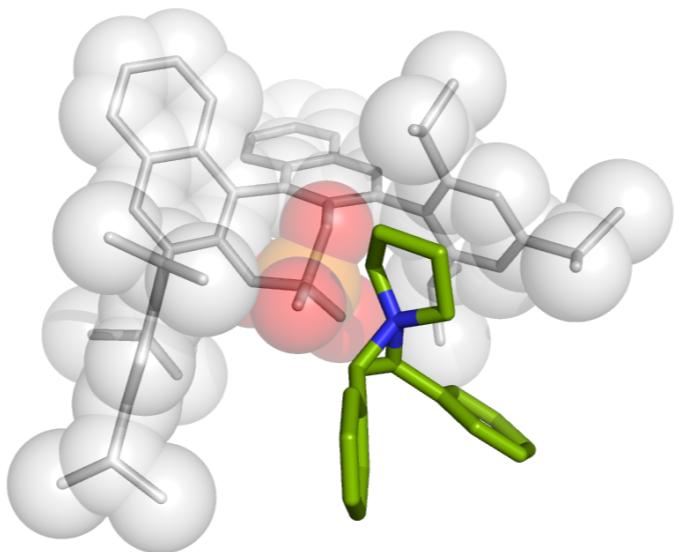
Interpret/Analyse

Which system do I have?

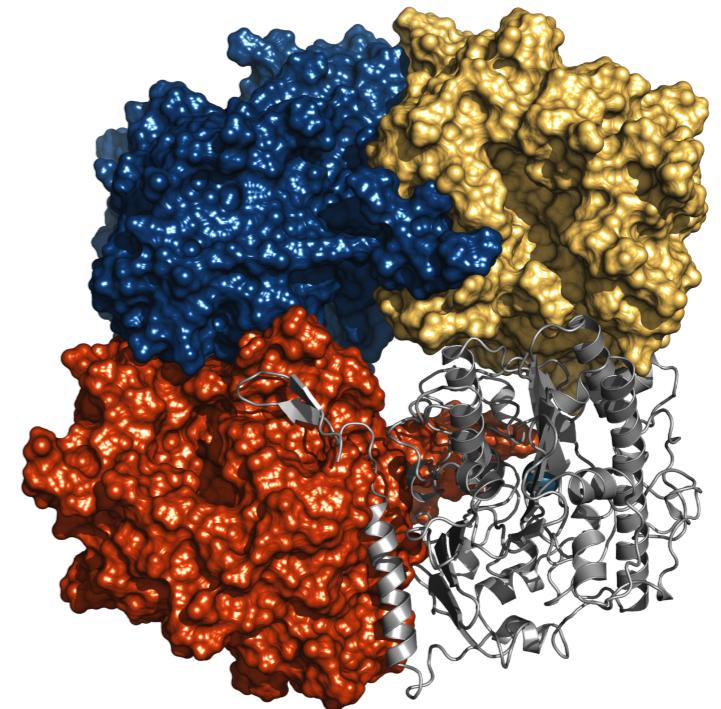
Size



17 atoms



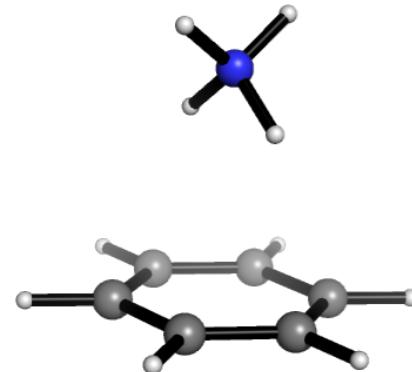
100 atoms



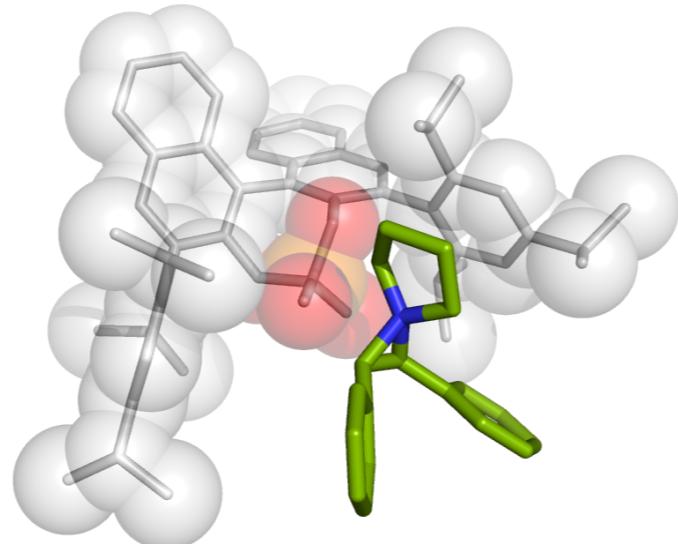
32000 atoms

Which system do I have?

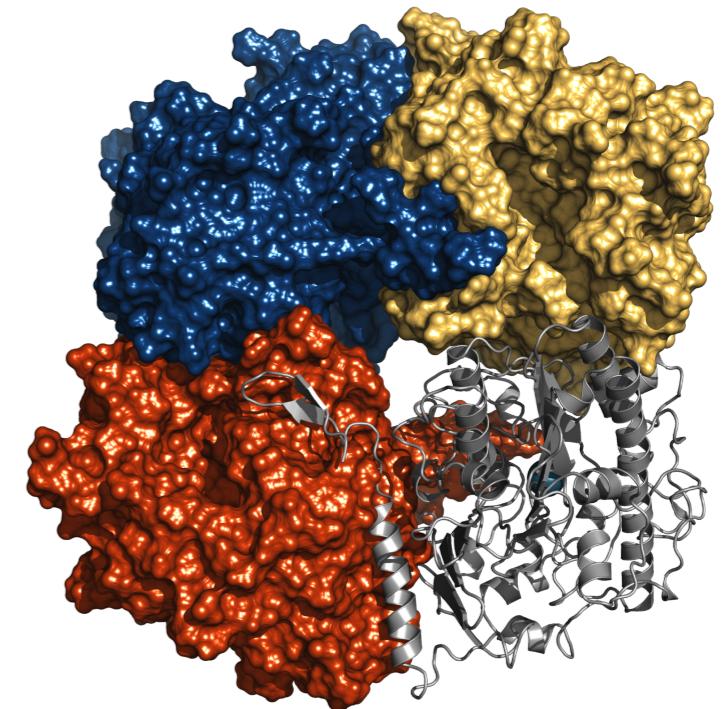
Size



17 atoms

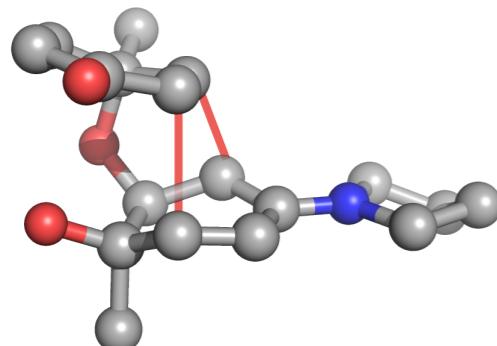


100 atoms

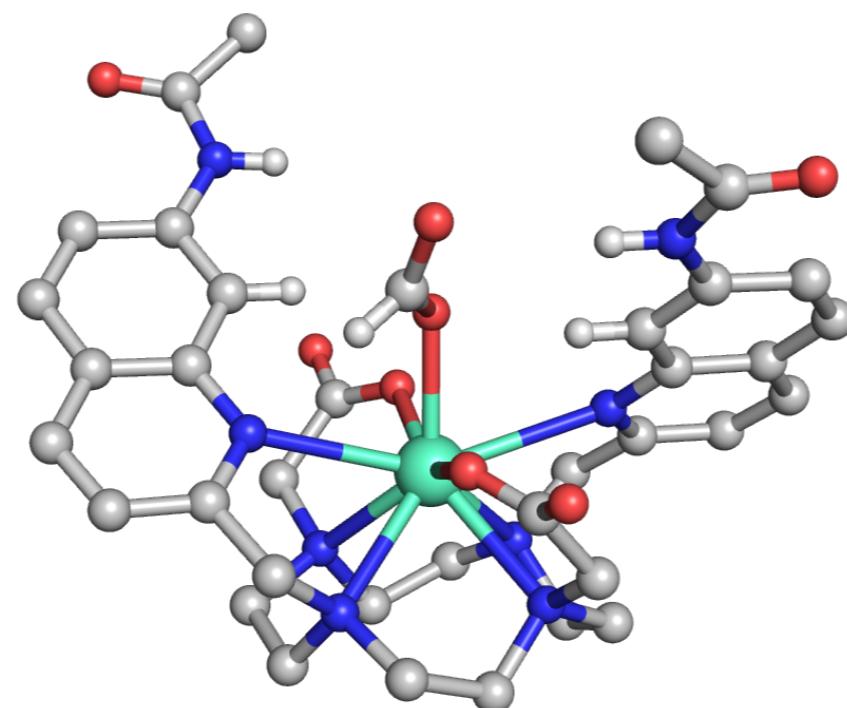


32000 atoms

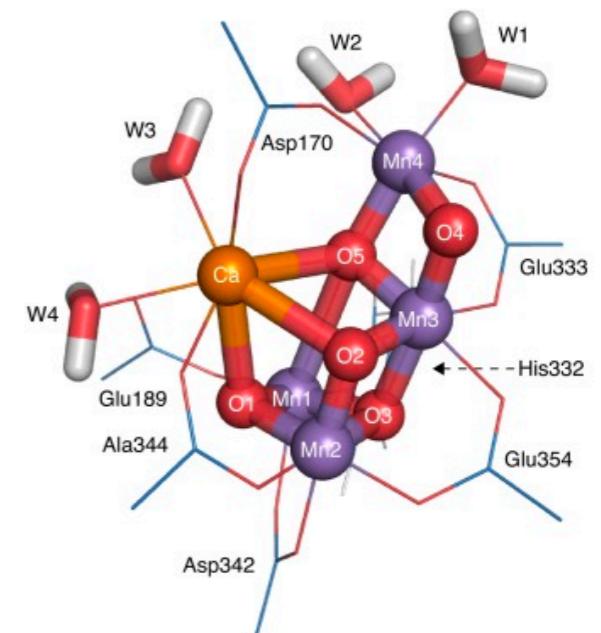
Complexity



Organic molecule



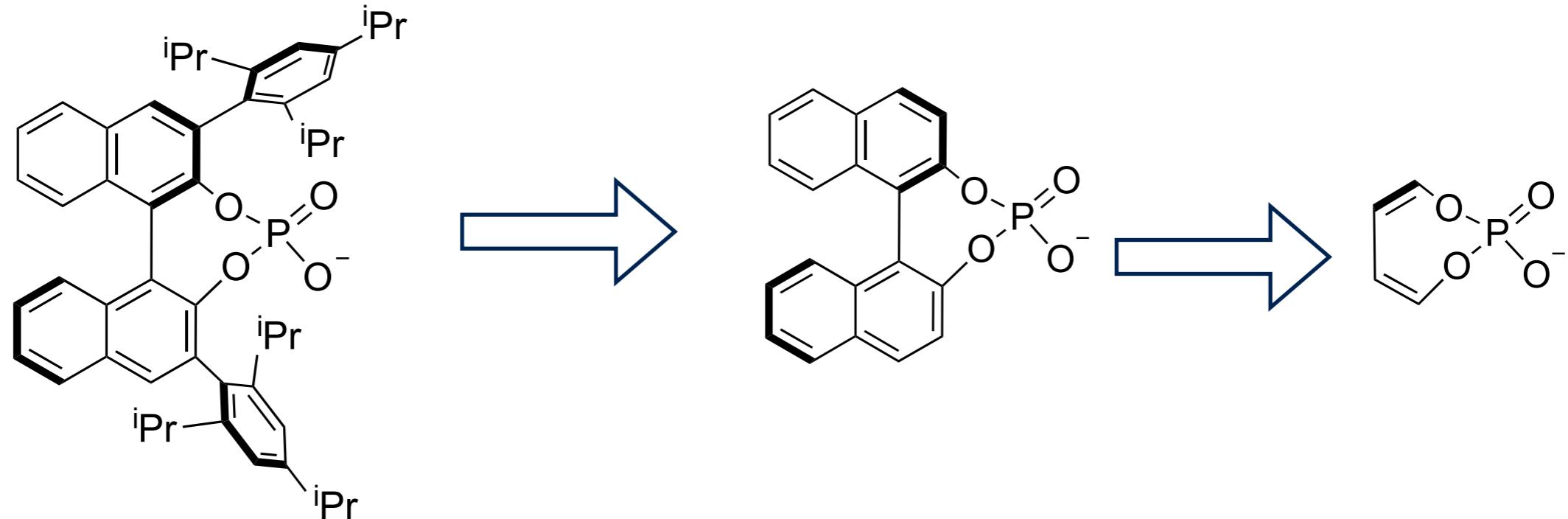
Eu-complex



Photosystem II
Multiple spin states

Which system do I have?

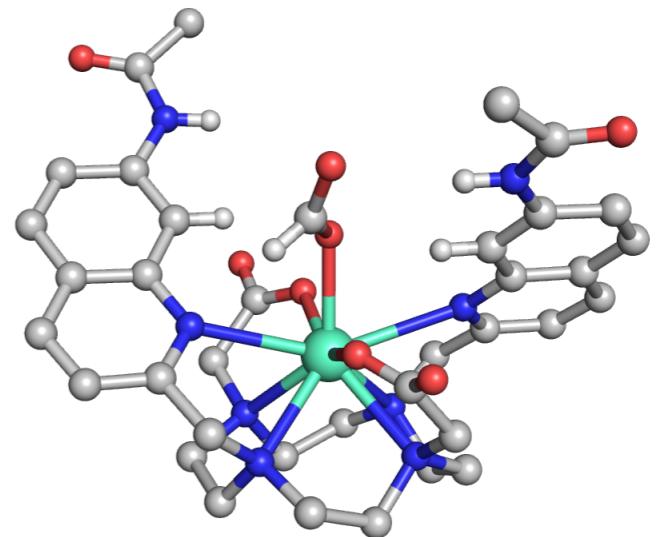
Model System



- Choice of model will depend on the problem and the available resources.
- Are the substituents required to understand the problem?
- A small model may be a good starting point.
- Factors out steric vs. electronic effects.

Which system do I have?

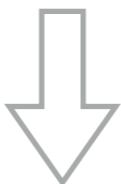
- Transition metals
 - Size, electronic structure, spin-orbit relativistic effects..
 - Low amount and often poor quality of data ...
- Radicals
 - Spin contamination, multideterminantal character.
 - Result usually very sensitive to the level of theory chosen.



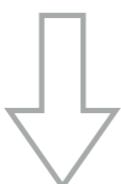
Which System Do I Have?



What Do You Want to Compute (and Why)?



Which Model /Method Should I Choose?



Verify Approach (vs. Experiment)



Interpret/Analyse

Which System Do I Have?



What Do You Want to Compute (and Why)?



Which Model /Method Should I Choose?



Verify Approach (vs. Experiment)



Interpret/Analyse

Which System Do I Have?



What Do You Want to Compute (and Why)?



Which Model /Method Should I Choose?



Verify Approach (vs. Experiment)



Interpret/Analyse

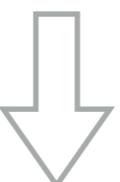
Which System Do I Have?



What Do You Want to Compute (and Why)?



Which Model /Method Should I Choose?



Verify Approach (vs. Experiment)



Interpret/Analyse



Often the most interesting
result is when the
“calculation gets it wrong”

Interpret/Analyse

Understand what you are comparing to ...

- Have ready your model/hypothesis
- Do you understand the experiment (theory, setup, sources of error)?
- Numerical precision (algorithm) vs Physical precision (comparable with experiment)
- Are there other interpretations I should consider?
- Test something that work and something that does not (false positive)

If the exp. method has a known error of 0.01 Å for a certain bond length, then there is no point in reporting geometries to 0.0000001 Å)

Which System Do I Have?



What Do You Want to Compute (and Why)?

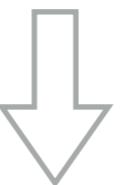


Which Model /Method Should I Choose?



Not easy to Answer

Verify Approach (vs. Experiment)



Interpret/Analyse

Which Method Should I Choose?

Electronic Schrödinger Equation

$$\hat{H}_e \psi_e = E_e \psi_e$$

$$\hat{H}_e = \frac{-\hbar^2}{2m} \sum_i^{electrons} \nabla_i^2 - \sum_i^{electrons} \sum_A^{nuclei} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i < j}^{electrons} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

Kinetic energy of the electrons

External potential

Schrödinger equation involves many-body interactions

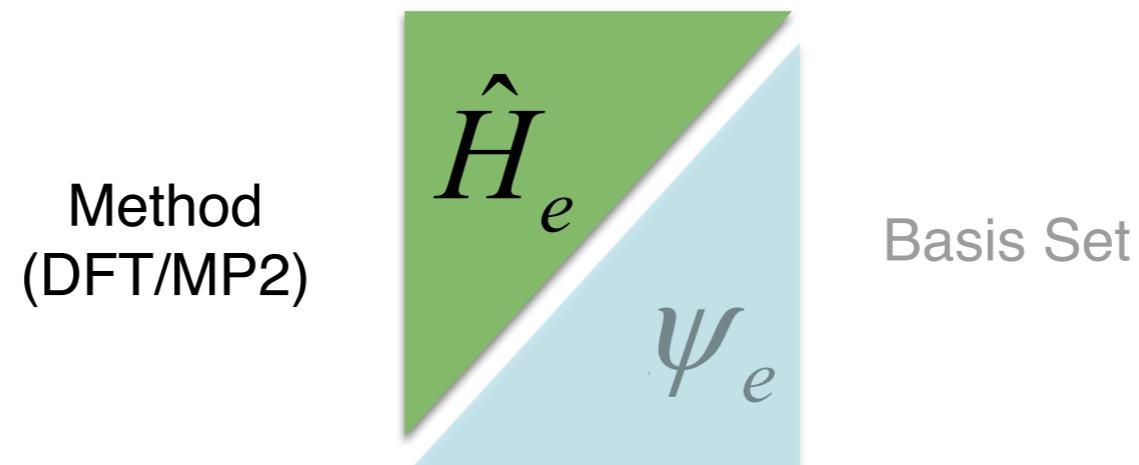
The Schrödinger equation can be solved analytically ('exactly') only for the simplest systems
(H, He⁺)

Which Method Should I Choose?

Electronic Schrödinger Equation

$$\hat{H}_e \psi_e = E_e \psi_e$$

$$\hat{H}_e = \frac{-\hbar^2}{2m} \sum_i^{electrons} \nabla_i^2 - \sum_i^{electrons} \sum_A^{nuclei} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i < j}^{electrons} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$



Often different approaches/levels of theory are required for different properties

ab initio and DFT approaches

Closed vs Open Shell

Single Reference	Multireference
HF (SCF)	CASCF
CISD	MRCISD
CCSD	MRCCSD
MP2	CASPT2
DFT	... works surprisingly well

ab initio and DFT approaches

Closed vs Open Shell

Single Reference

HF (SCF)

CISD

CCSD

MP2

DFT

Multireference

CASCF

MRCISD

MRCCSD

CASPT2

... works surprisingly well

Assume Ψ_e as a single **antisymmetric** product of one-electron functions (molecular orbitals)

$$\psi_e(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} [\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) - \phi_2(\mathbf{r}_1)\phi_1(\mathbf{r}_2)] = -\psi_e(\mathbf{r}_2, \mathbf{r}_1)$$

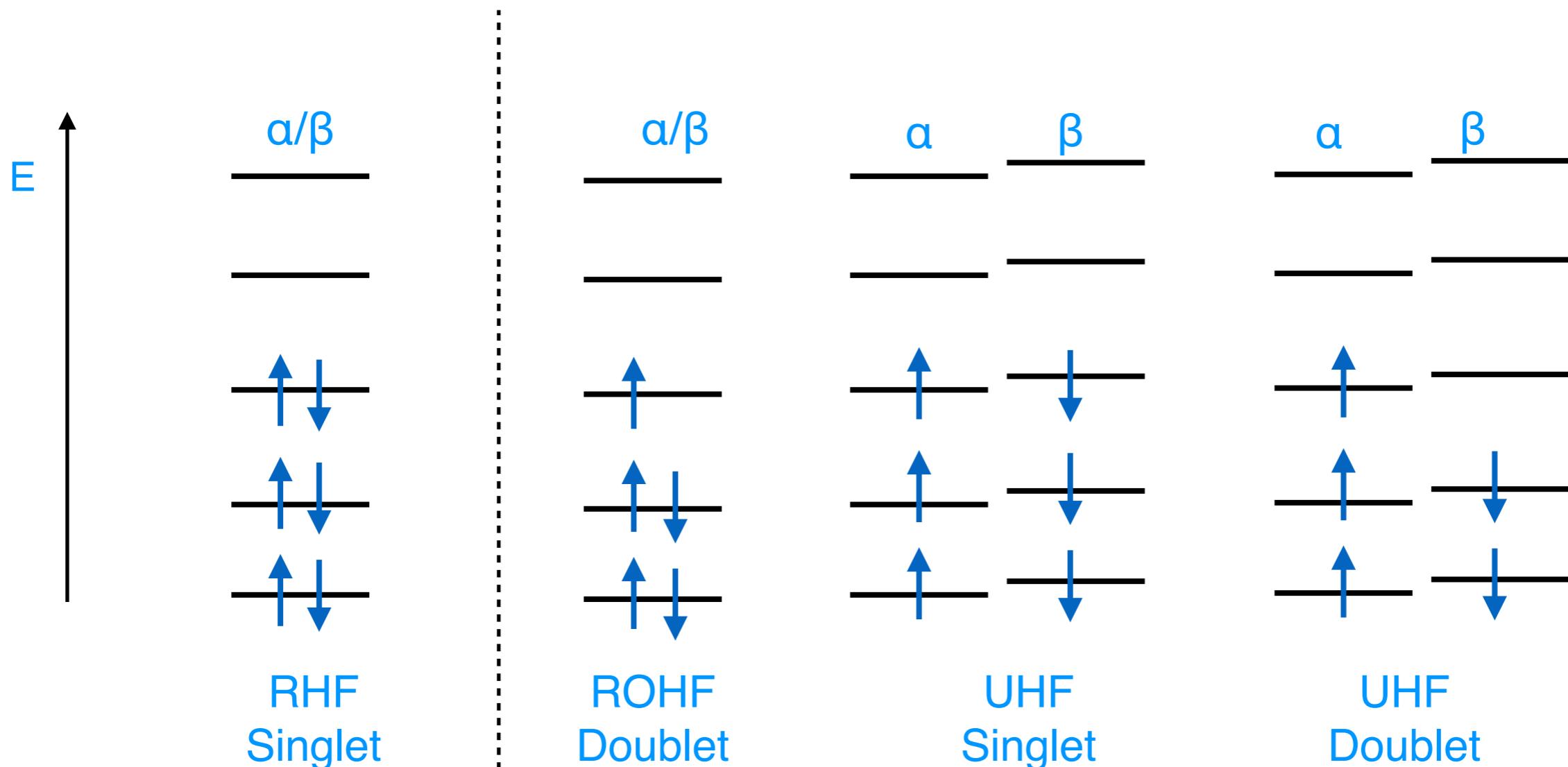
For a general N -electron system, we can write this product as a **Slater Determinant**

$$\psi_e(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) & \dots & \phi_N(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) & \dots & \phi_N(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{r}_N) & \phi_2(\mathbf{r}_N) & \dots & \phi_N(\mathbf{r}_N) \end{vmatrix} \equiv \psi$$

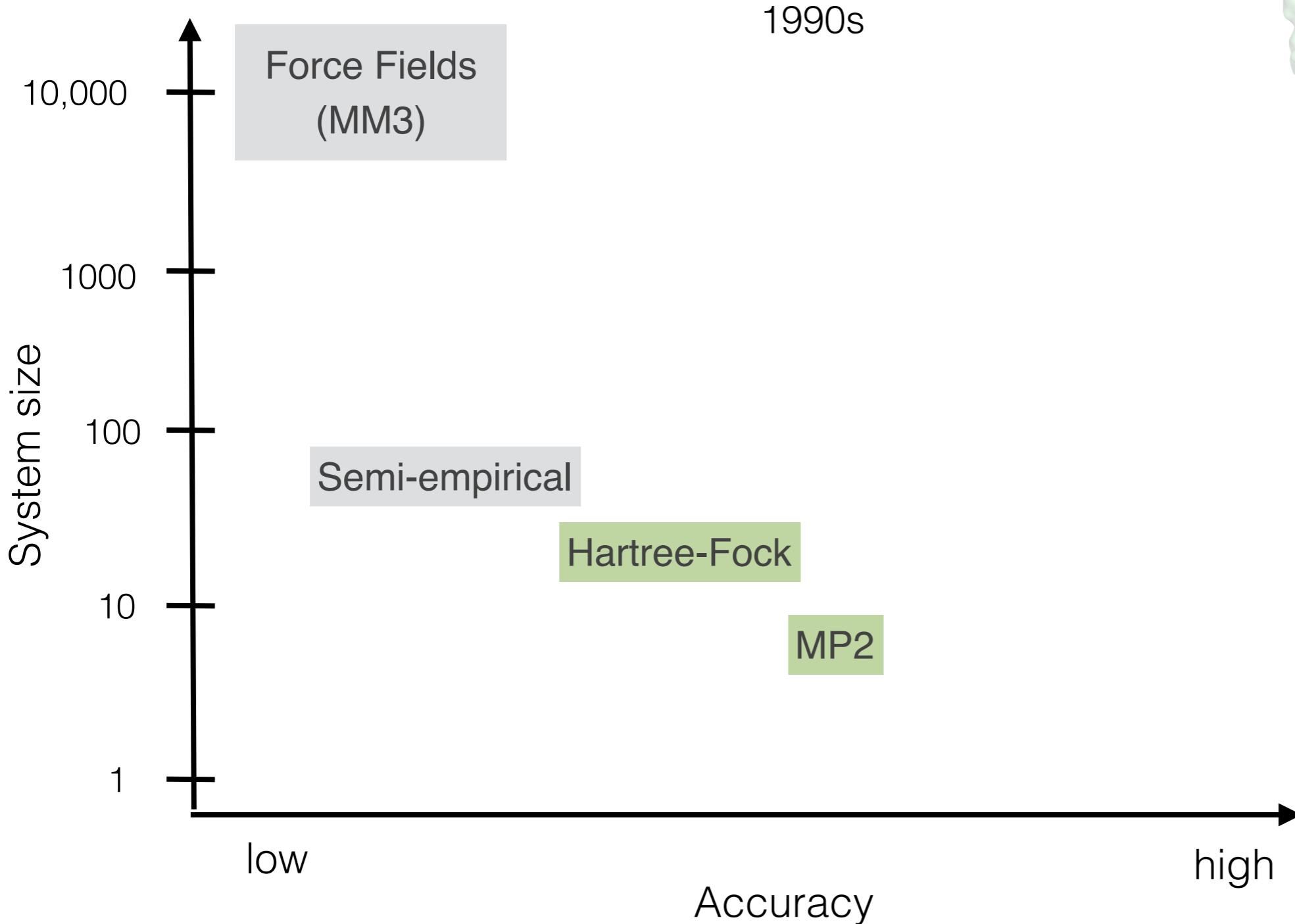
ab initio and DFT approaches

Single Reference (SR) vs Multireference (MR)

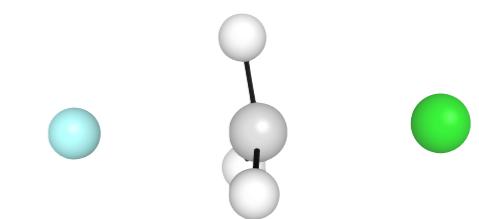
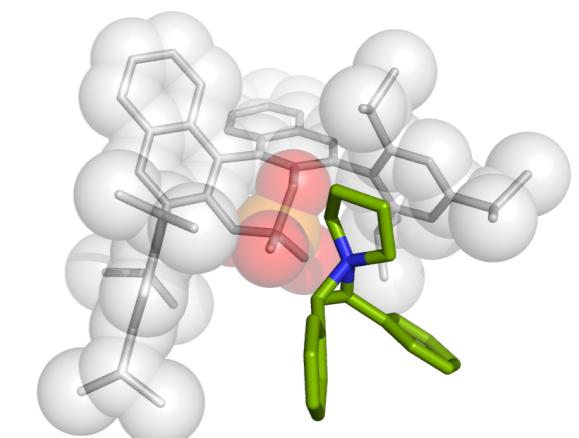
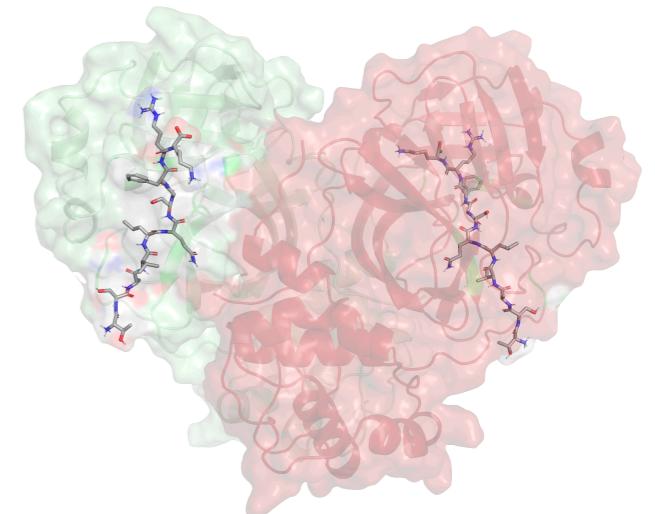
The Hartree-Fock approximation can be applied with or without restrictions on the spins of the MOs



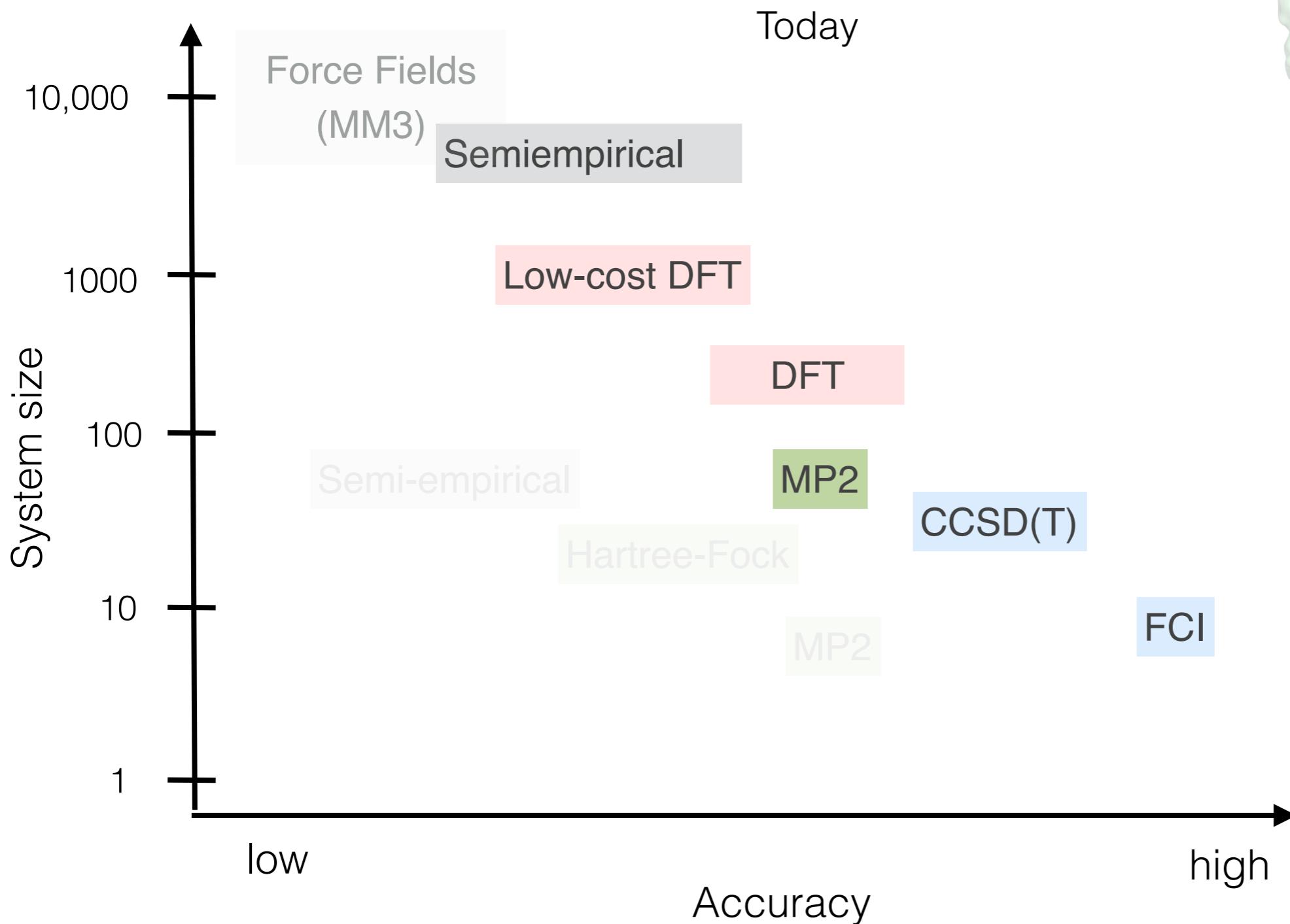
DFT/WFT Approaches



Trend applies for a 'good' basis set in each method



DFT/WFT Approaches



Trend applies for a 'good' basis set in each method



UNIVERSITY OF
OXFORD

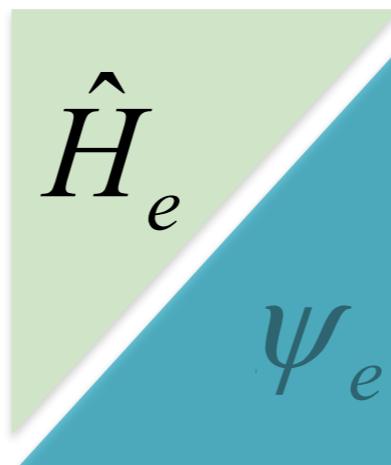
Which Method Should I Choose?

Electronic Schrödinger Equation

$$\hat{H}_e \psi_e = E_e \psi_e$$

$$\hat{H}_e = \frac{-\hbar^2}{2m} \sum_i^{electrons} \nabla_i^2 - \sum_i^{electrons} \sum_A^{nuclei} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i < j}^{electrons} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

Method
(DFT/MP2)



Often different approaches/levels of theory are required for different properties

Basis Set

Collection of vectors which defines a space in which a problem is solved.

Ψ_e

Requirements:

- It must be as large as possible, so that the MO constructed from different basis functions can be as close as possible to the true MO orbital
- It must be as small as possible, so that the calculation remains computationally tractable

Physicists

like plane wave basis sets

Chemists

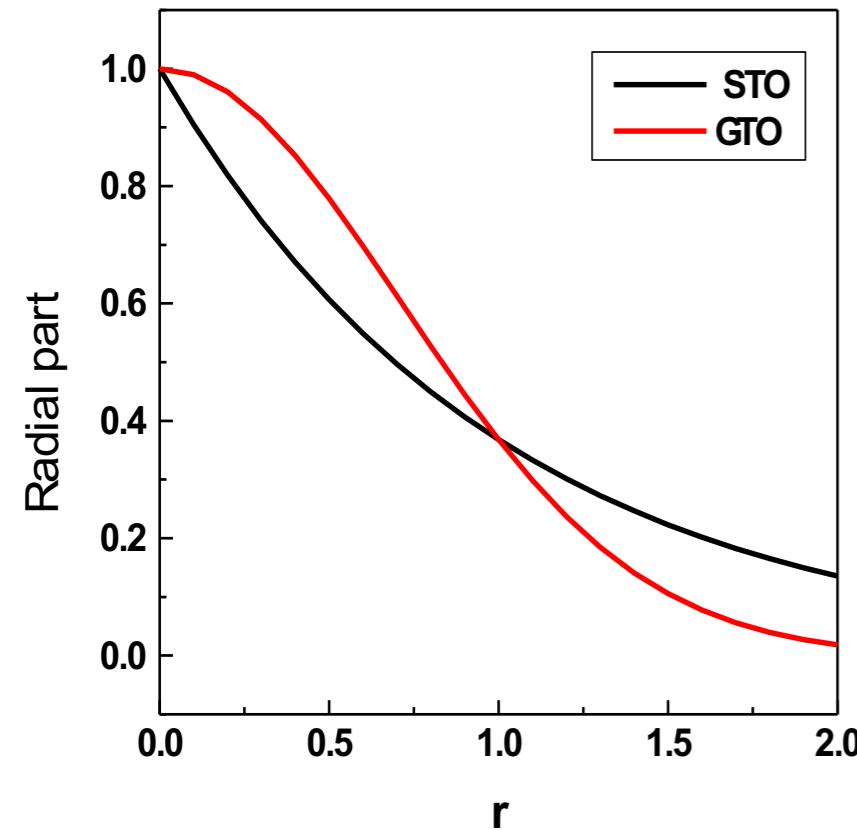
Atomic orbital like functions

LCAO-MO approximation

Basis Set

Slater type orbitals (STOs)

ψ_e



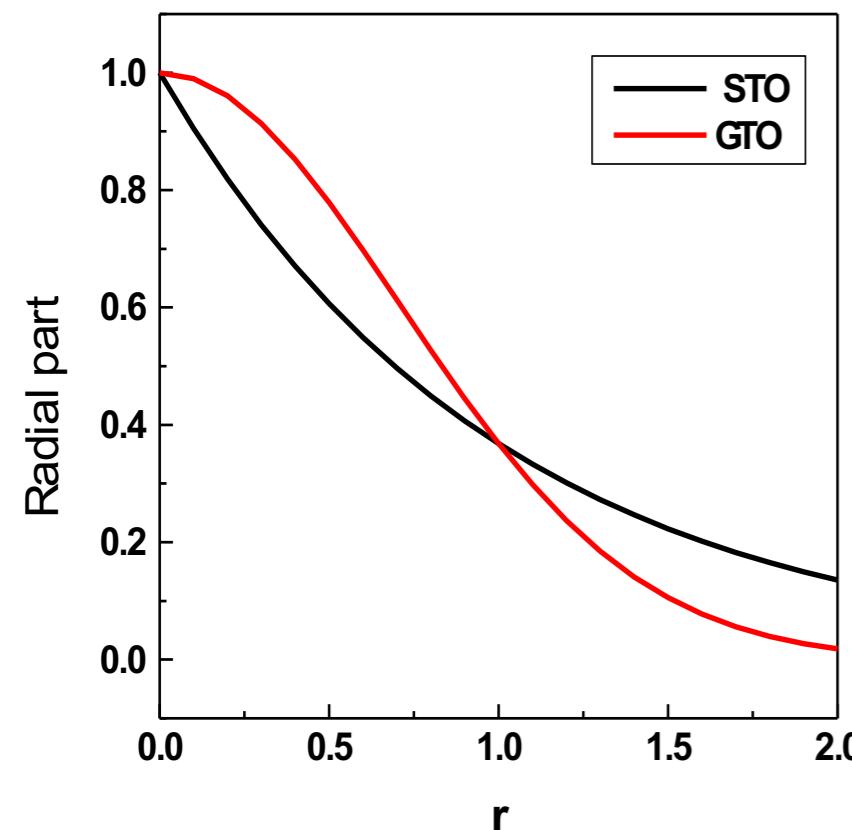
$$\phi_{nlm_l}^{STO} = N_\xi r^{n-1} e^{-\xi r} Y_{lm_l}$$

Hydrogen atom orbitals

3/4-centre integrals can't be performed analytically.
Numerically **expensive**

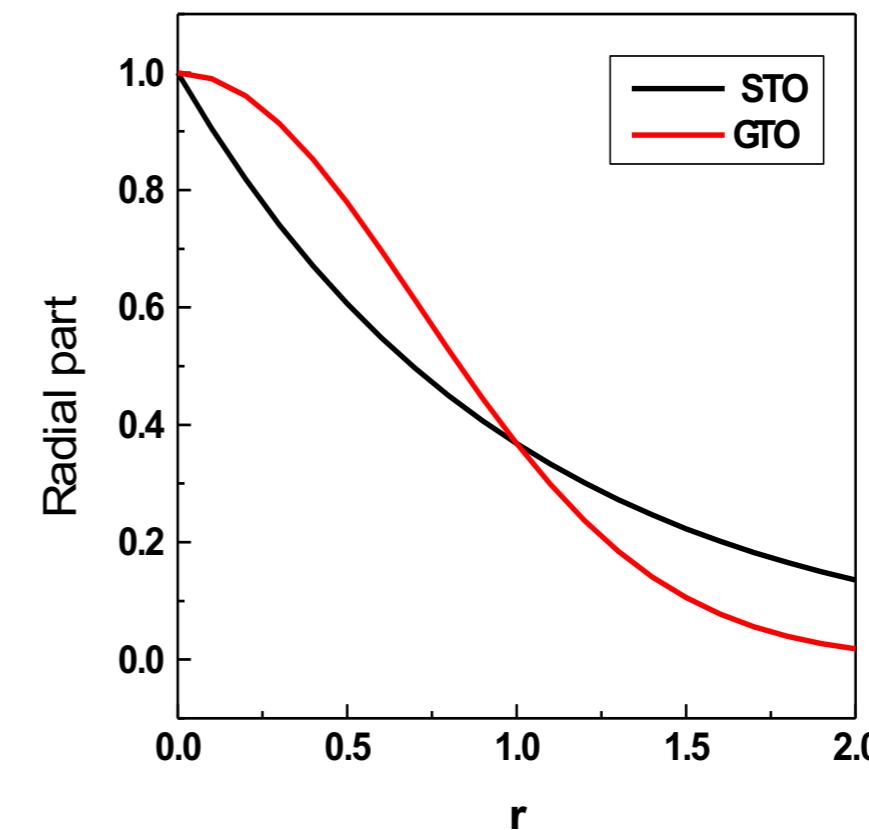
Basis Set

Slater type orbitals (STOs)



$$\phi_{nlm_l}^{STO} = N_\xi r^{n-1} e^{-\xi r} Y_{lm_l}$$

Gaussian type orbitals (GTOs)



$$\phi_{nlm_l}^{GTO} = N_\xi r^{n-1} e^{-\xi r^2} Y_{lm_l}$$

Hydrogen atom orbitals

3/4-centre integrals can't be performed analytically.
Numerically **expensive**

Known analytic form, more **efficient**

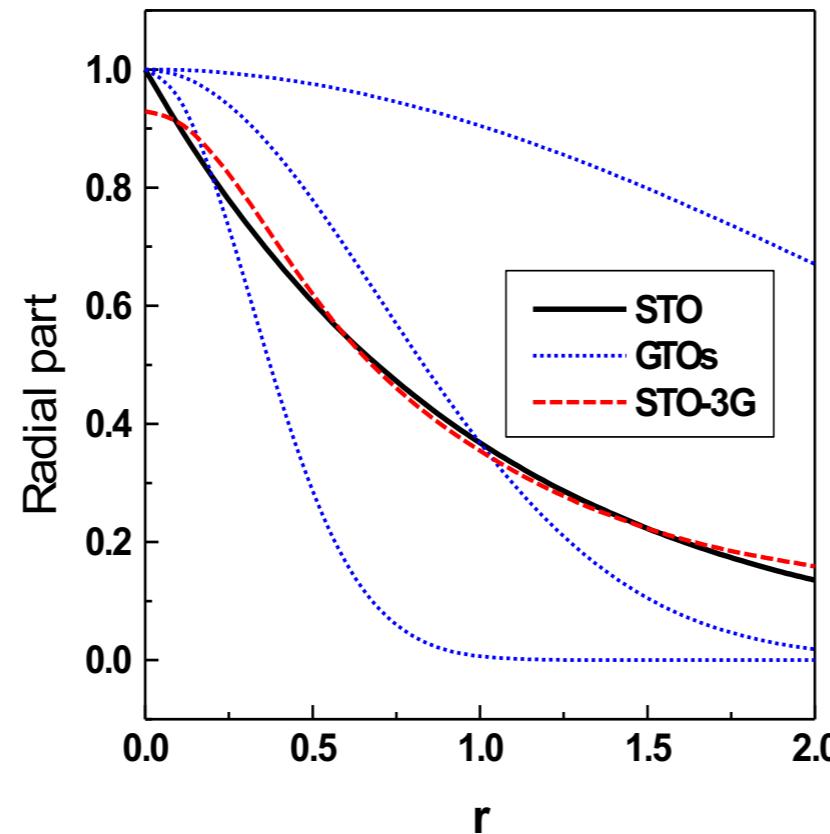
ψ_e

Basis Set

ψ_e

STOs vs GTOs

At the nucleus a GTO has zero slope, in contrast to a STO which has a cusp



Linear combination of gaussians to model one STO.

This gave rise to the **STO-nG** family of basis sets (n = gaussians used to model STO's)

Classification of basis sets

ψ_e

Type	Description	Example
Minimal Basis Sets	1 STO for each atomic shell, core or valence	C: 1s, 2s, 2px, 2py, 2pz (5 AO's -> 5 basis functions)
Double Zeta (DZ)	2 STOs per AO	C: 10 basis functions
Triple Zeta (TZ)	3 STOs for each AO	C: 15 basis functions
Split-valence basis	1 STO for each core AO at least 2 STOs for valence AO's	C atom: 9 basis functions

Classification of basis sets

ψ_e

Type	Description	Example
Minimal Basis Sets	1 STO for each atomic shell, core or valence	C: 1s, 2s, 2px, 2py, 2pz (5 AO's -> 5 basis functions)
Double Zeta (DZ)	2 STOs per AO	C: 10 basis functions
Triple Zeta (TZ)	3 STOs for each AO	C: 15 basis functions
Split-valence basis	1 STO for each core AO at least 2 STOs for valence AO's	C atom: 9 basis functions

...Having different-sized functions allows the orbital to get bigger or smaller when other atoms approach it

Classification of basis sets

split-valence basis

ψ_e

Pople split valence	Dunning (correlation consistent)	Karlsruhe
3-21G	cc-pvdz	aug-cc-pvdz
6-31+G*	cc-pvtz	aug-cc-pvtz
6-311++G**	cc-pvqz	aug-cc-pvqz

3 primitives assigned
for each AO

3-21G

valence AO described by 2
basis, one made of 2 primitives
and the other of 1

Classification of basis sets

split-valence basis

ψ_e

Pople split valence	Dunning (correlation consistent)	Karlsruhe
3-21G	cc-pvdz	aug-cc-pvdz
6-31+G*	cc-pvtz	aug-cc-pvtz
6-311++G**	cc-pvqz	aug-cc-pvqz

3 primitives assigned
for each AO

valence AO described by 2
basis, one made of 2 primitives
and the other of 1

3-21G

13 basis functions, 21 primitive gaussians, 13 cartesian basis functions
5 alpha electrons 5 beta electrons

Classification of basis sets

split-valence basis

ψ_e

Pople split valence	Dunning (correlation consistent)	Karlsruhe
3-21G	cc-pvdz	aug-cc-pvdz
6-31+G*	cc-pvtz	aug-cc-pvtz
6-311++G**	cc-pvqz	aug-cc-pvqz

3 primitives assigned
for each AO

3-21G

valence AO described by 2
basis, one made of 2 primitives
and the other of 1

13 basis functions, 21 primitive gaussians, 13 cartesian basis functions
5 alpha electrons 5 beta electrons

O 1s² 2s² 2p² 2p¹2p¹ (3 x 5 = 15P) ; 1 + 2 x 4 = 9 basis

H 1s¹ (2 x 3 = 6P) ; 2 x 2 = 4 basis

Classification of basis sets

Ψ_e

Type	Description	Example
Diffuse functions	Very small ζ exponents	Improve the description of the density at large distances anions, H-bonds, polarizability
Polarisation functions	Flexibility in angular $(\ell \rightarrow \ell+1)$	Improve the description of bonding

Functional/Basis Sets

h2o_bad_geometry.inp

```
! BP86 def2-SVP SP

*xyz 0 1
H -1.0 0.0 0.0
O 0.0 0.0 0.0
H 1.5 0.0 0.0
*
```

BP86 is a GGA functional composed of:
Becke 1988 exchange functional & Perdew 1986 correlation functional

def2-svp: Split valence polarization basis set