

A large, semi-transparent white grid-based surface plot is centered on the slide. It has a light purple gradient and a wavy, undulating pattern, resembling a 3D landscape or a molecular surface.

Applied Computational Chemistry

Lecture 2

Fernanda Duarte

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



Twitter @duarte_group

duartegroupchem.org



Github: duartegroup

How do we use it?

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

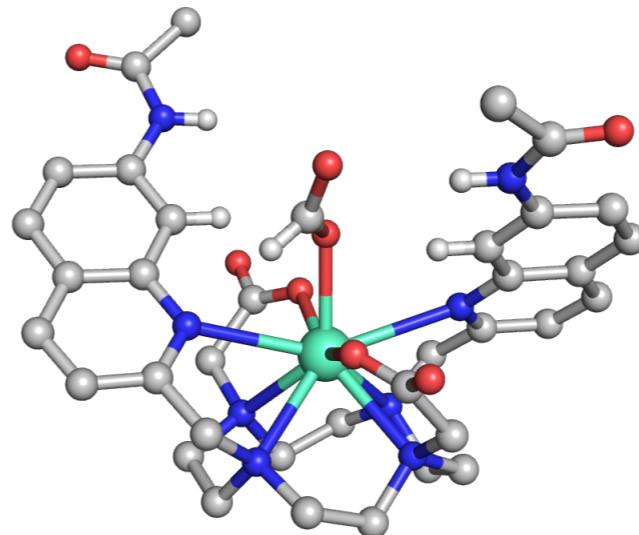
Classification of basis sets

ψ_e

- Electrostatic core potentials (ECP) e.g. lanl2dz or SDD

Cu $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s$

Pd $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^8$



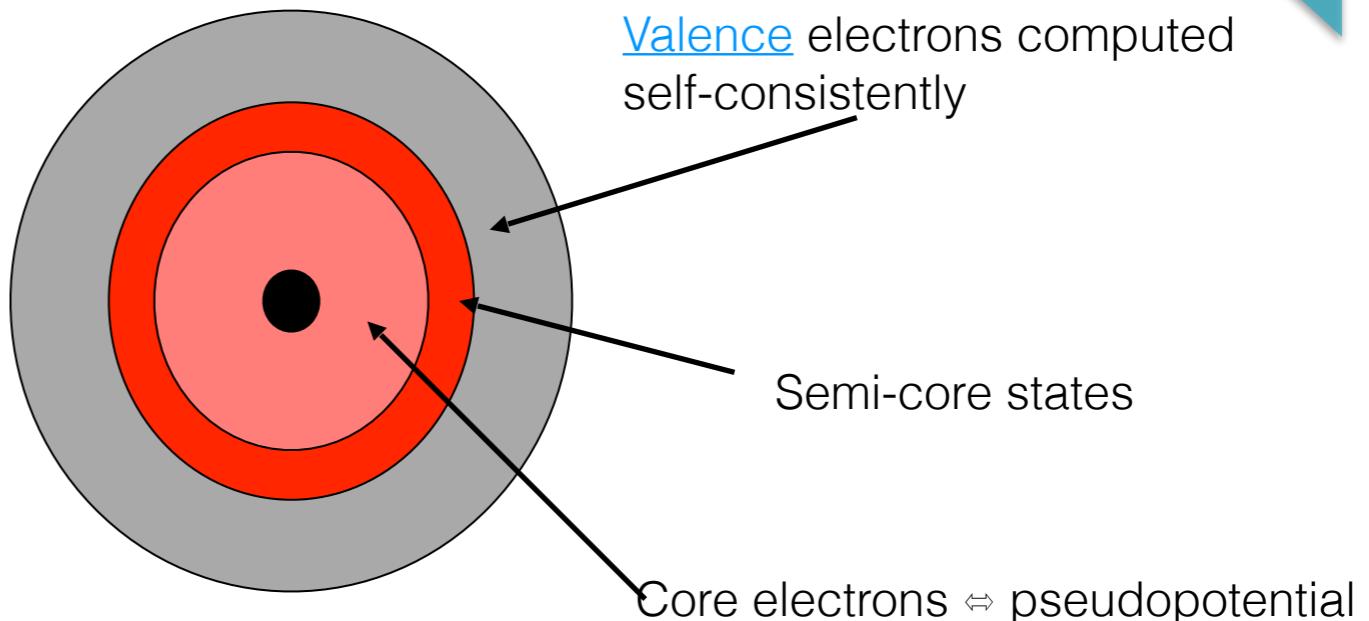
- Relativistic Effect:
 - For heavy elements, ECPs are chosen to reproduce all-e relativistic calculations
 - Other effects may (or may not) be included: e.g. spin-orbit coupling

Classification of basis sets

ψ_e

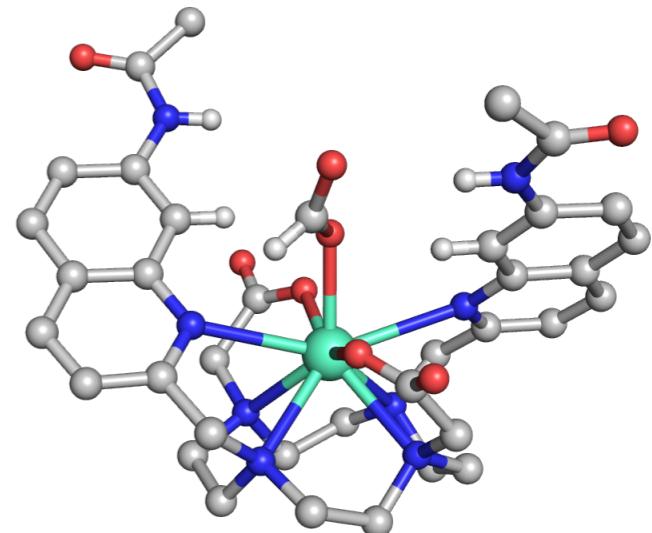
- Electrostatic core potentials (ECP) e.g. lanl2dz or SDD

Cu $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s$
Pd $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^8$



- Relativistic Effect:

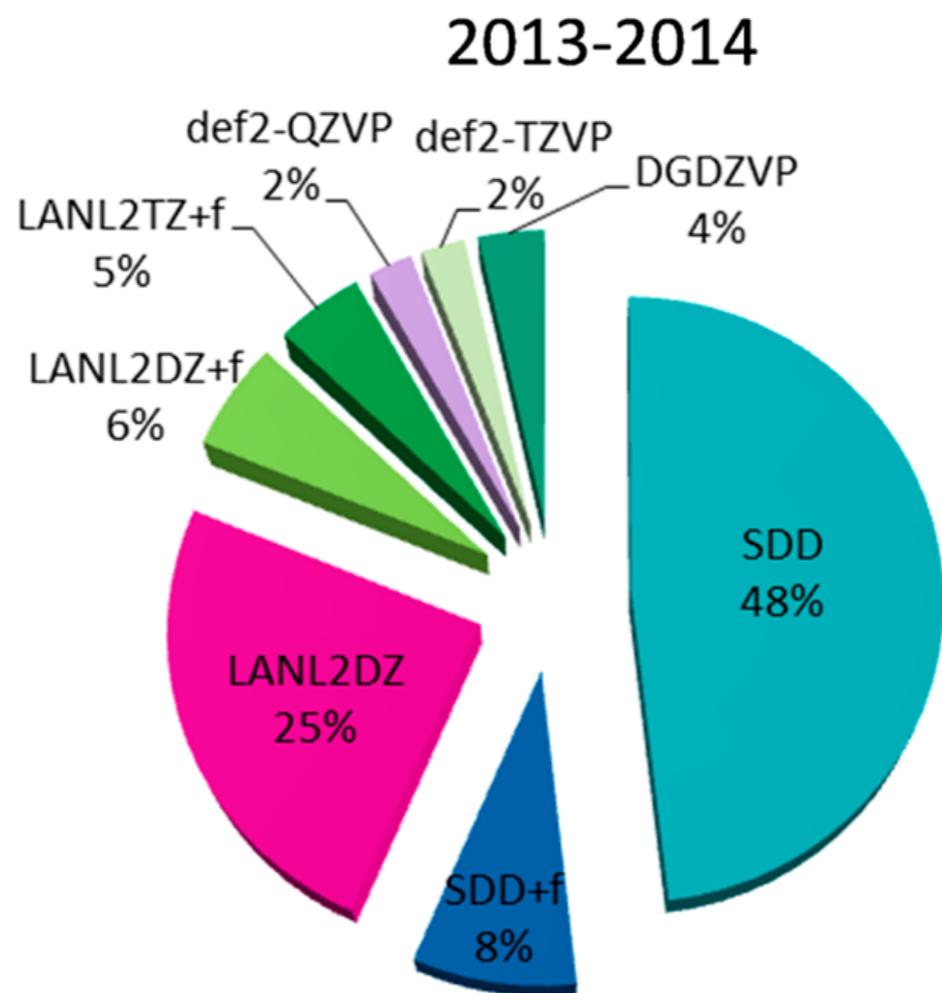
- For heavy elements, ECPs are chosen to reproduce all-e relativistic calculations
- Other effects may (or may not) be included: e.g. spin-orbit coupling



Classification of basis sets

ψ_e

Popular DFT methods used for Pd, Ni, Rh, and Ir



Basis set superposition error (BSSE)

ψ_e

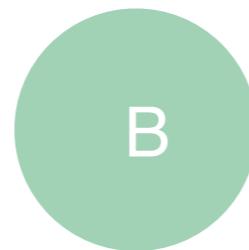
interaction energy of molecular complexes = $E(AB) - E(A) - E(B)$

Small basis

Poor description of A (anion)
Higher energy



$E(A)$



$E(B)$

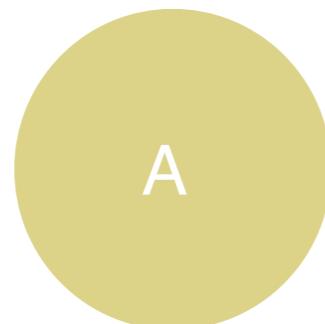
description of A is improved by the basis
from B (overstabilisation)



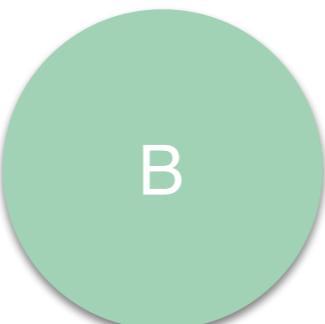
$E(AB)$

Large basis

Good description of A (anion)

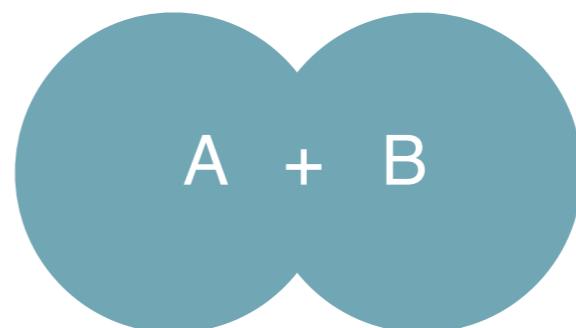


$E(A)$



$E(B)$

Good description of A in the complex



$E(AB)$

a spurious attractive contribution that arises from an improved description of the fragments in the combined basis rather than from a true physical interaction between the fragments

How do we use it?

Which System Do I Have?



What Do You Want to Compute (and Why)?

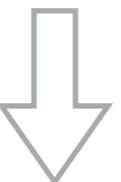


Which Model /Method Should I Choose?



Not easy to Answer

Verify Approach (vs. Experiment)



Interpret/Analyse

Density Functional Theory

Hohenberg–Kohn theorem

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

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

The **interacting system** is replaced by a **fictitious non-interacting system*** with same density

An easy Hamiltonian! A sum of one-electron operators

$$E[\rho] = T[\rho] + E_{Ne}[\rho] + E_H[\rho] + E_x[\rho] + E_c[\rho]$$

Kinetic
Energy

Nuc-e
interaction

Classic e-e
interaction

The rest is due to kinetic and e-e

*one electron moving in an “average” potential caused by the rest of the electrons

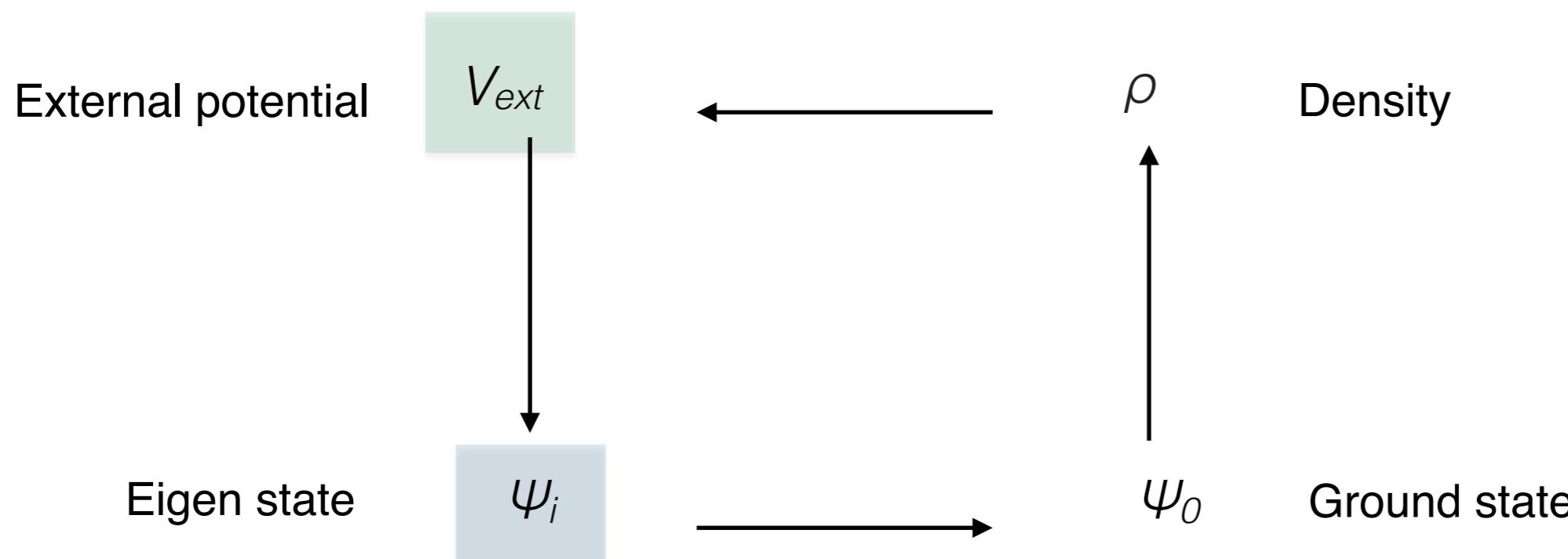
Density Functional Theory

Hohenberg–Kohn theorem

First H-K: V_{ext} is determined uniquely by the ground state density \rightarrow All properties of the system are determined given only the ground state density

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

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



Density Functional Theory

Hohenberg–Kohn theorem

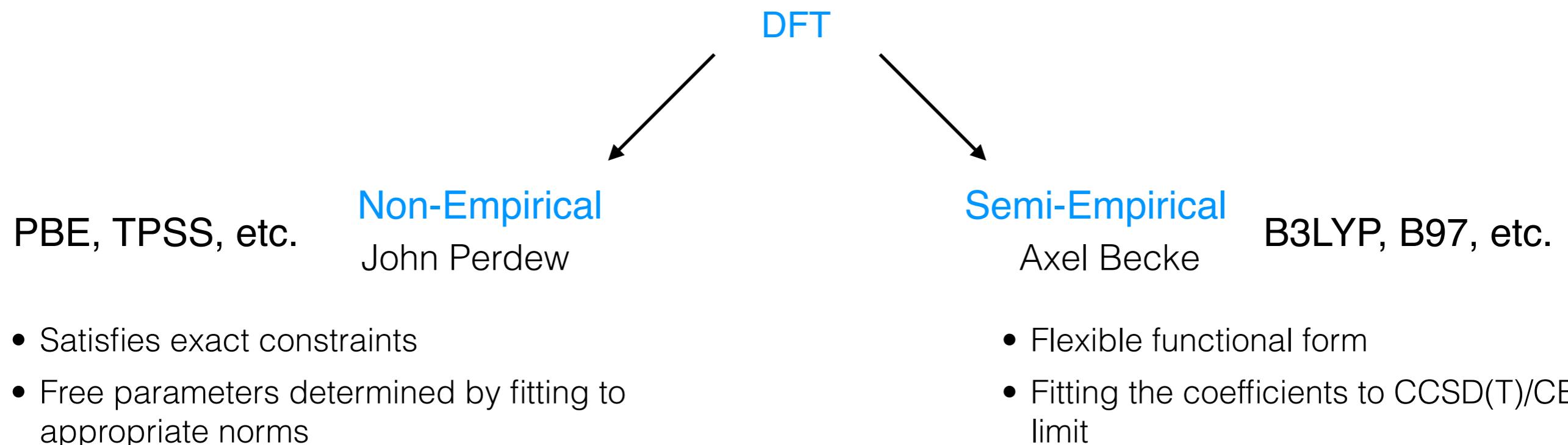
Second H-K: the electron density that minimises the energy of the functional is the true electron density corresponding to the solutions of the Schrödinger equation

Variational principle

$$E_0[\Psi_0] = \min_{\Psi} E[\Psi] = \min_{\Psi} \langle \Psi | T + V_{Ne} + V_{ee} | \Psi \rangle$$

$$E_0[\rho_0] = \min_{\rho} E[\rho] = \min_{\rho} T[\rho] + E_{Ne}[\rho] + E_{ee}[\rho]$$

Density Functional Theory



Hybrid DFT Functionals Contain a % of Hartree Fock

Density Functional Theory

The Jacob's Ladder(Perdew 2000)

Heaven of Chemical Accuracy

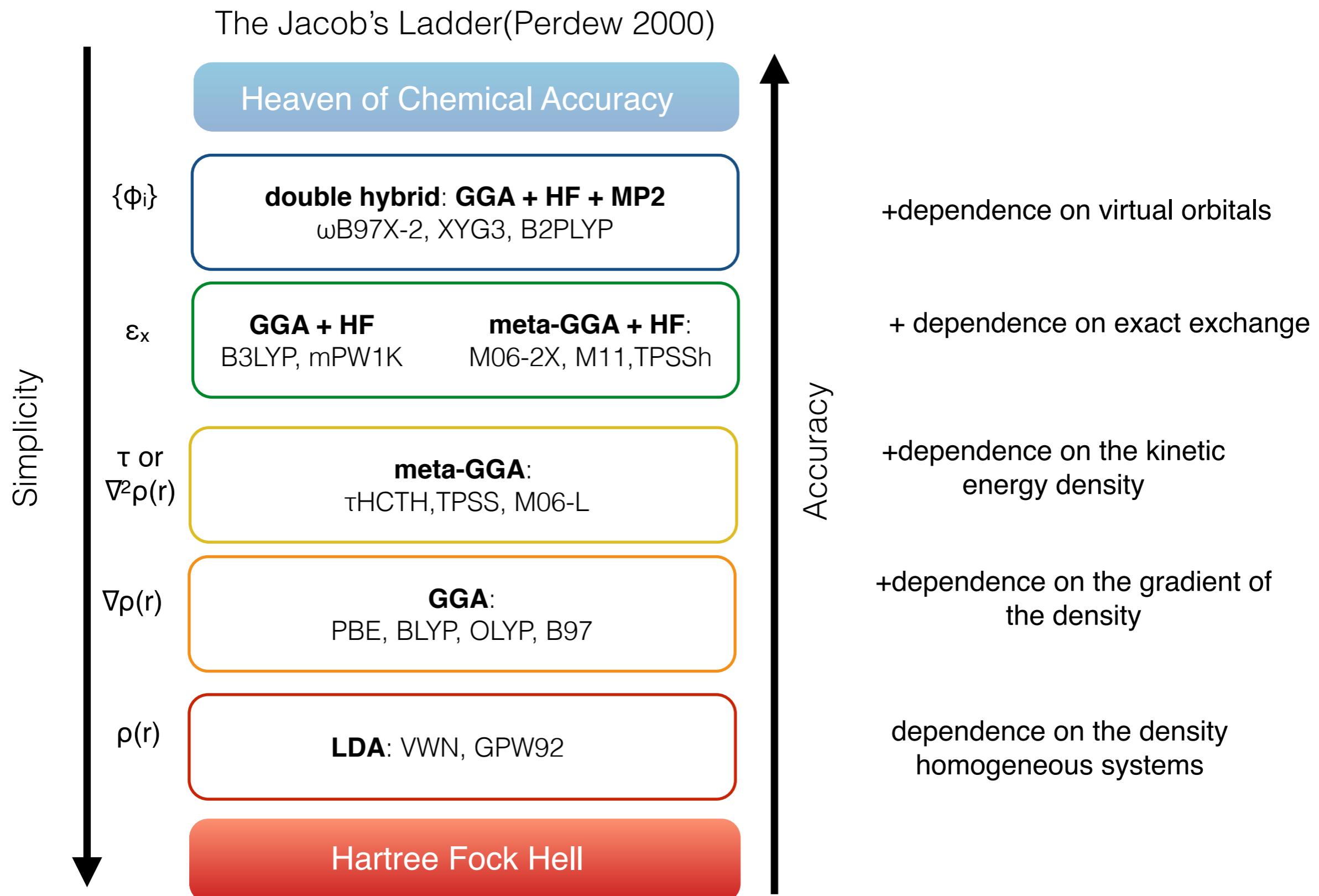
1 kcal/mol for reaction energies or barrier heights

0.1 kcal/mol for noncovalent interaction energies

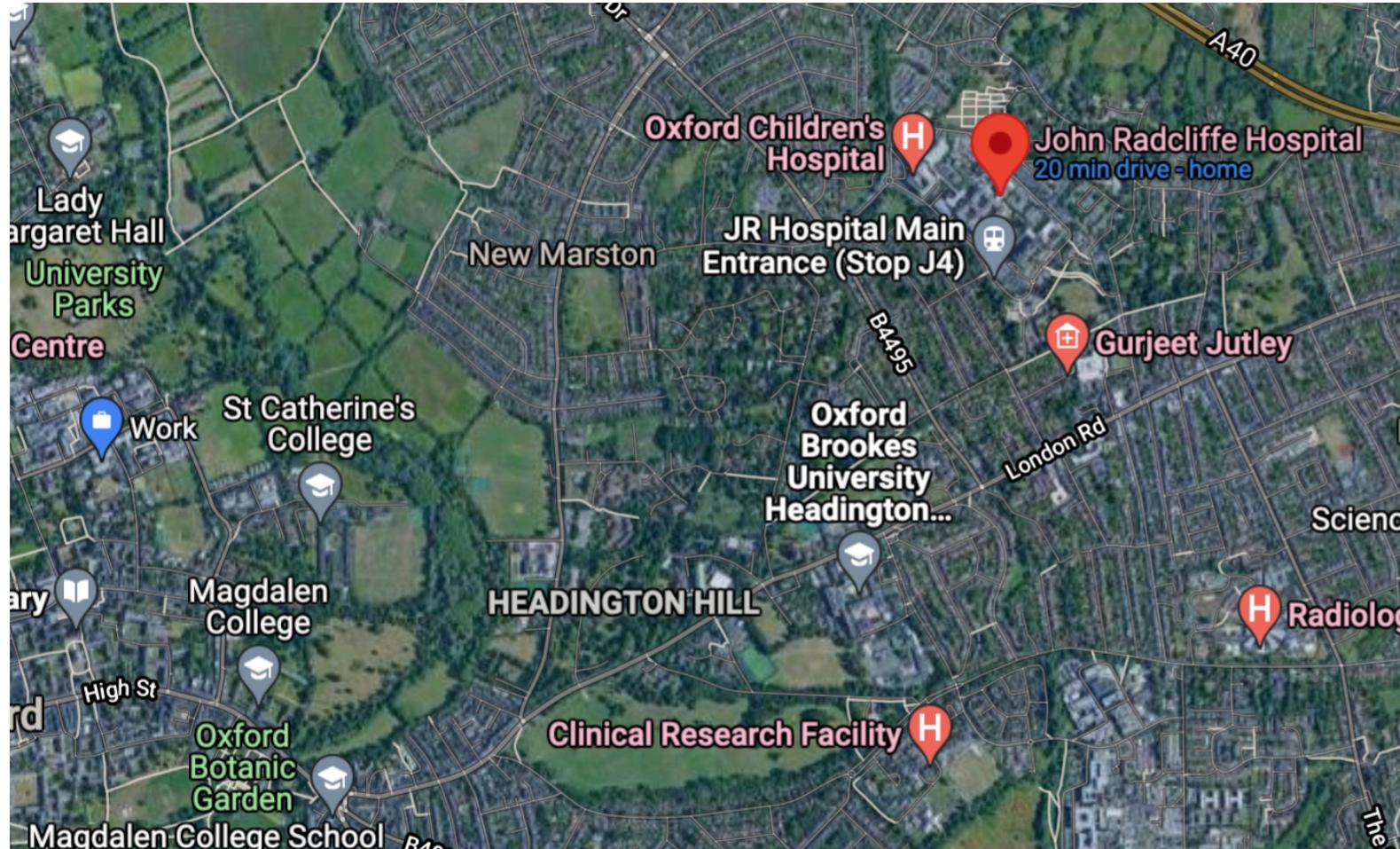
0.1 eV for electronic excitation energies.

Hartree Fock Hell

Density Functional Theory

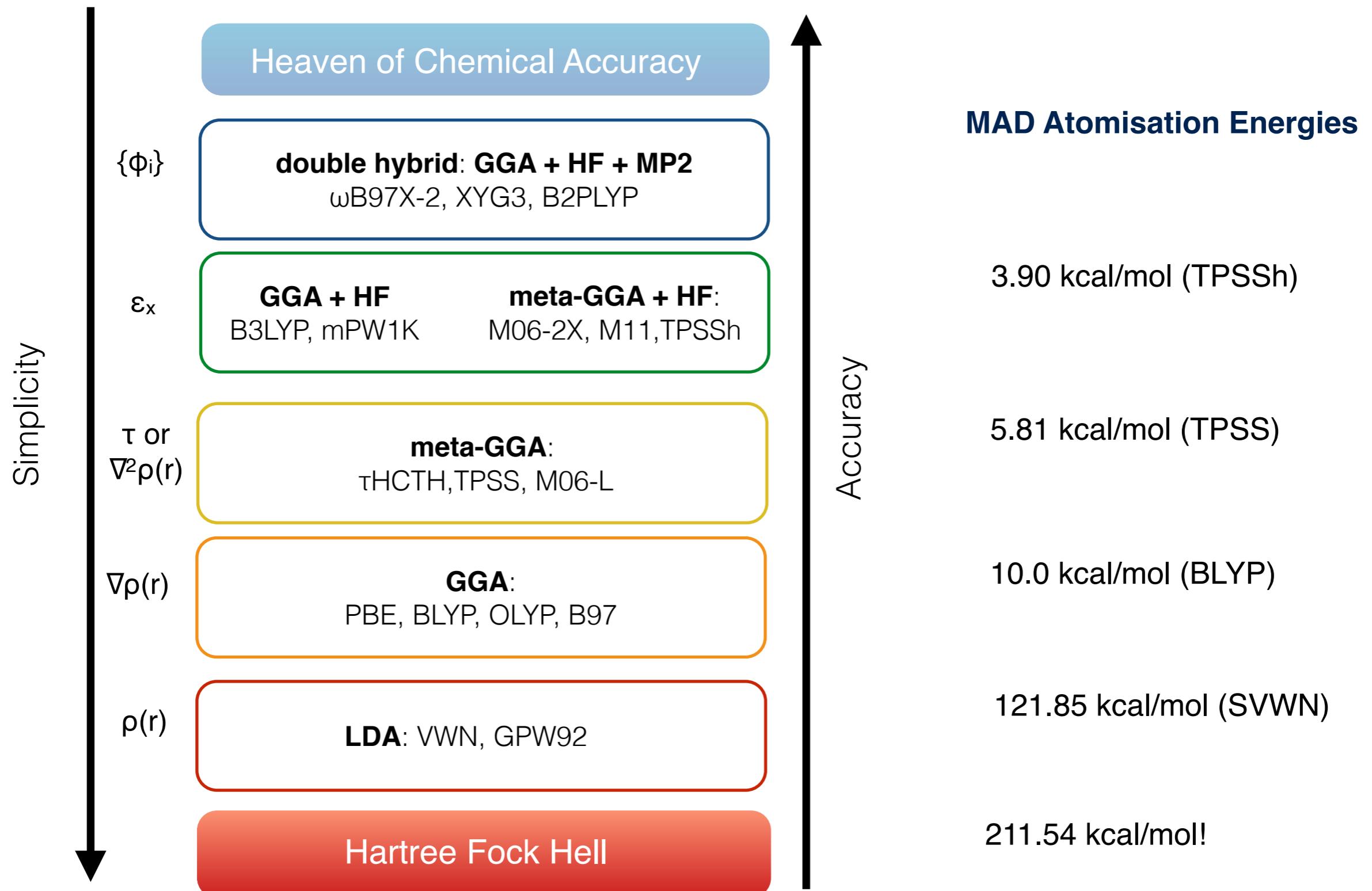


The devil is in the details



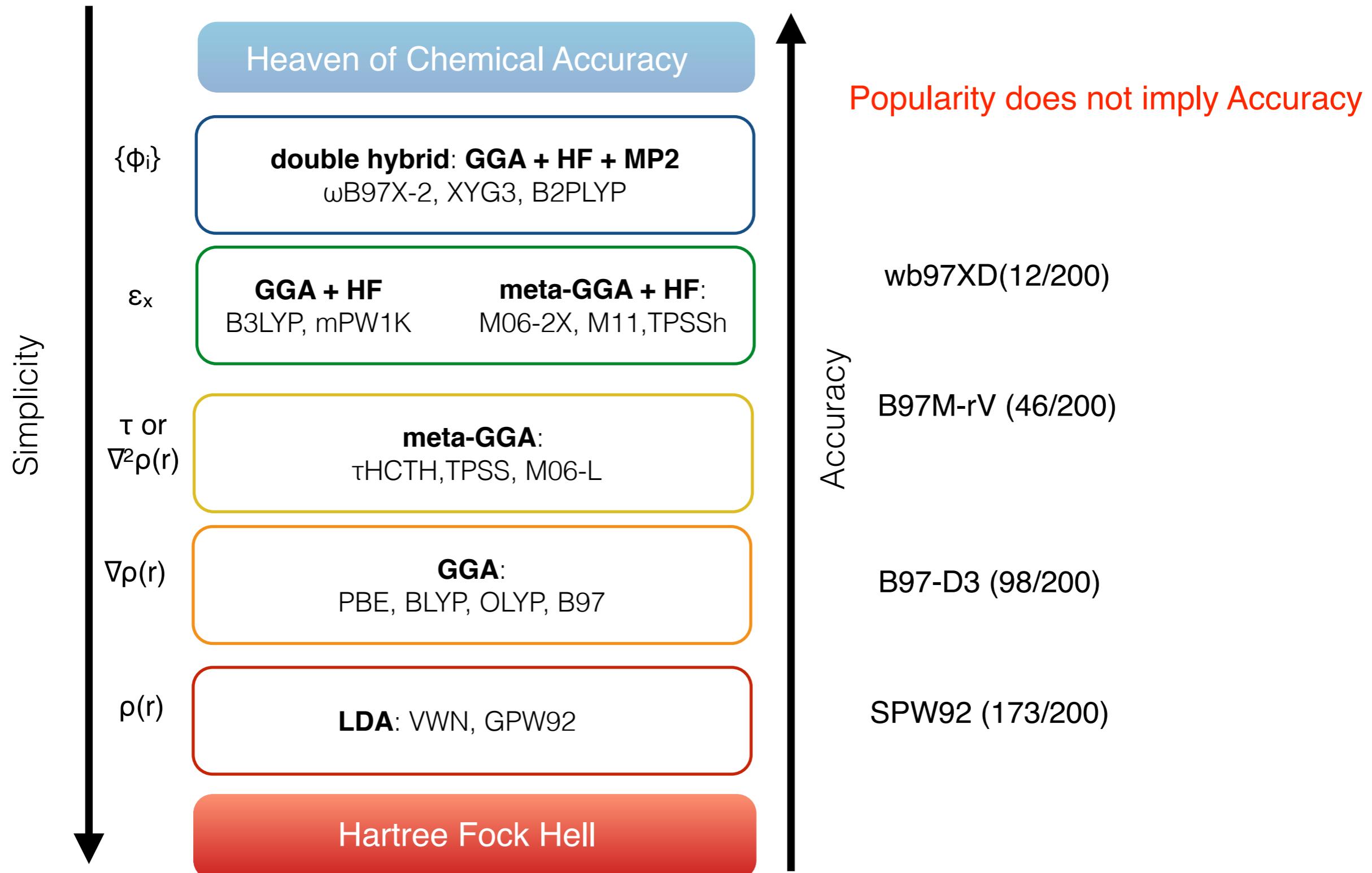
Density Functional Theory

The Jacob's Ladder(Perdew 2000)

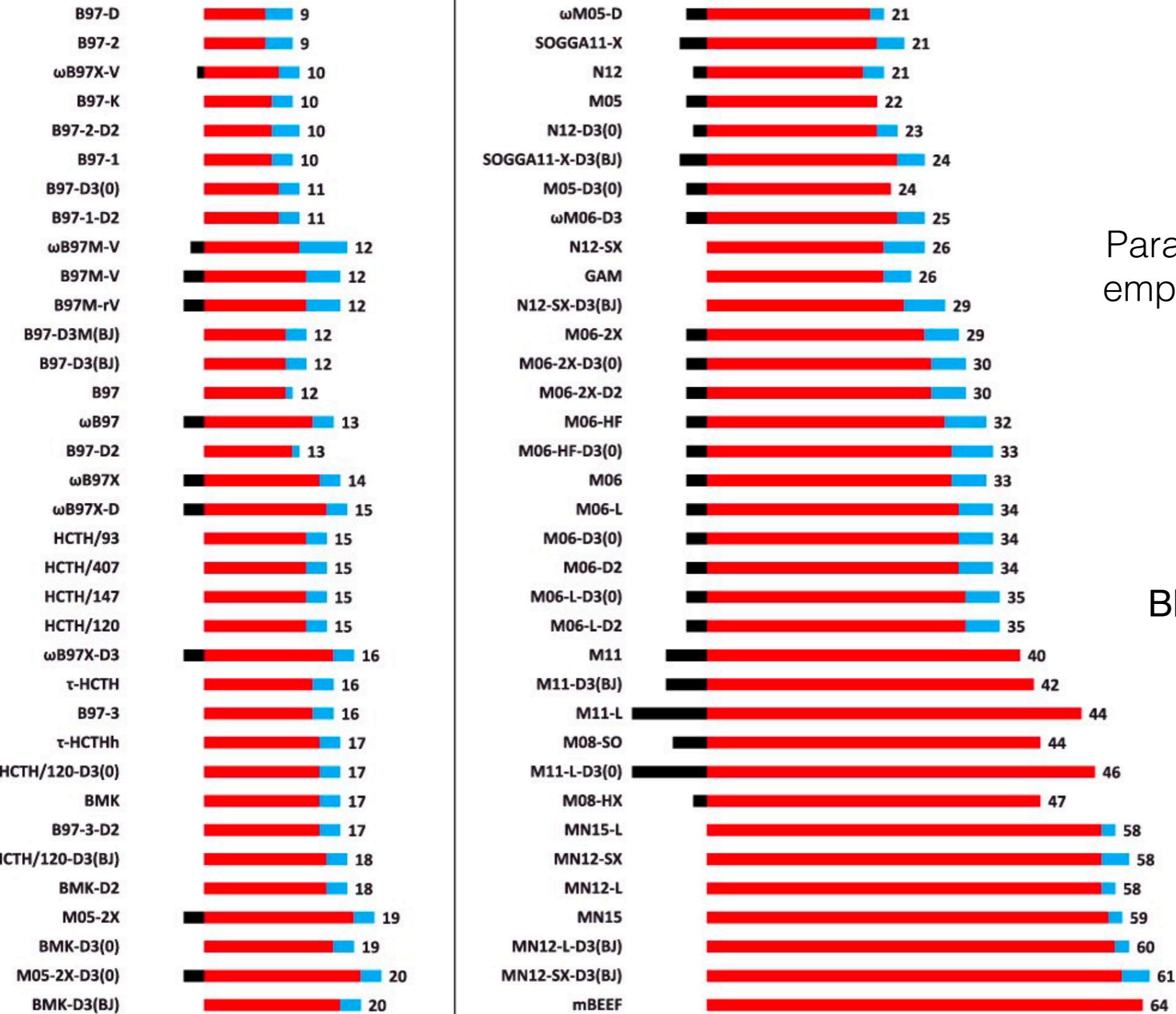


Density Functional Theory

The Jacob's Ladder(Perdew 2000)



Density Functional Theory



Parameter counting in semi-empirical density functionals

Red: fitted parameters
Blue:borrowed or preset parameters

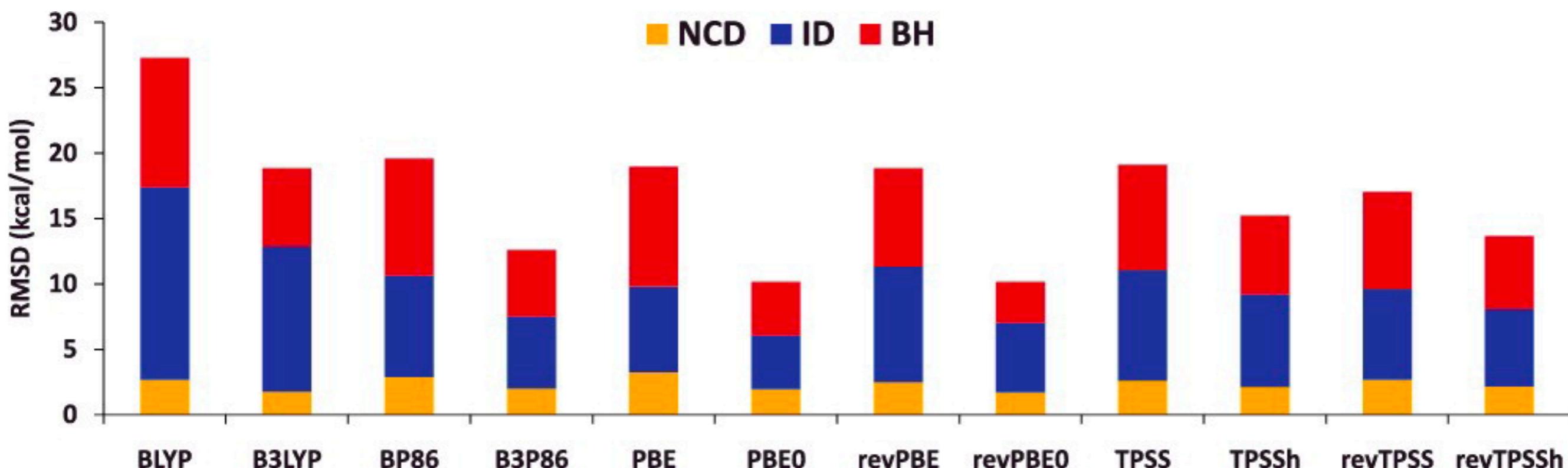
Black:number of constraints

Density Functional Theory

\hat{H}_e

Which Functional is Best for ...?

NCD 91 non-covalent inter. **ID** 155 isomerisation energies **BH** 206 barrier heights

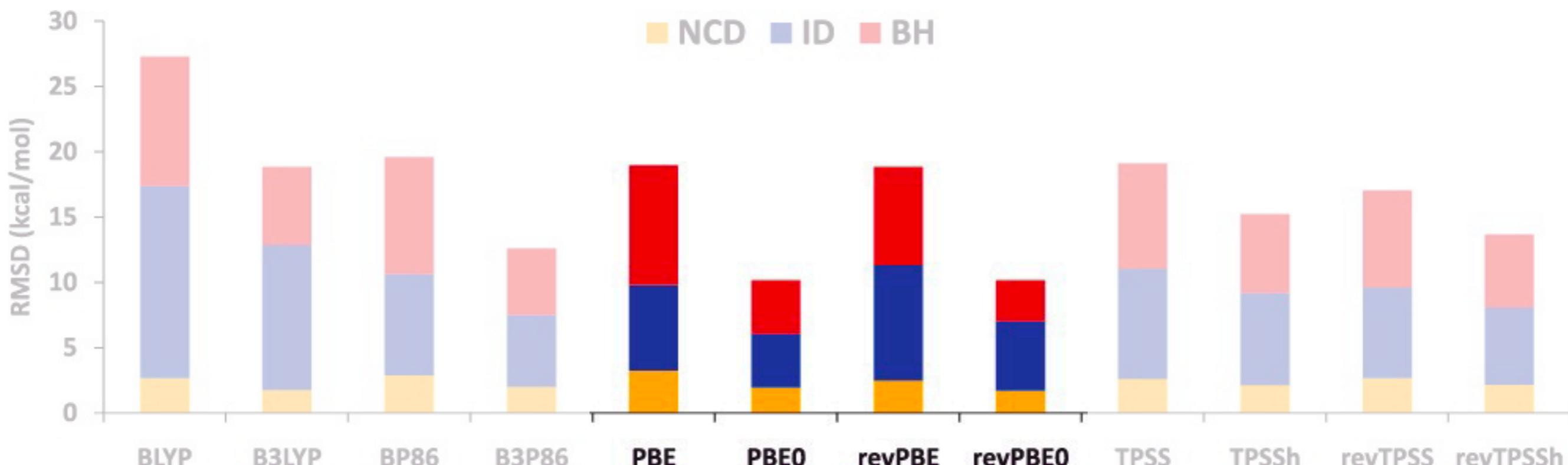


Density Functional Theory

\hat{H}_e

Which Functional is Best for ...?

NCD 91 non-covalent inter. **ID** 155 isomerisation energies **BH** 206 barrier heights



PBE: Perdew-Burke-Ernzerhof (GGA)

PBE0 (hybrid: GGA + HF)

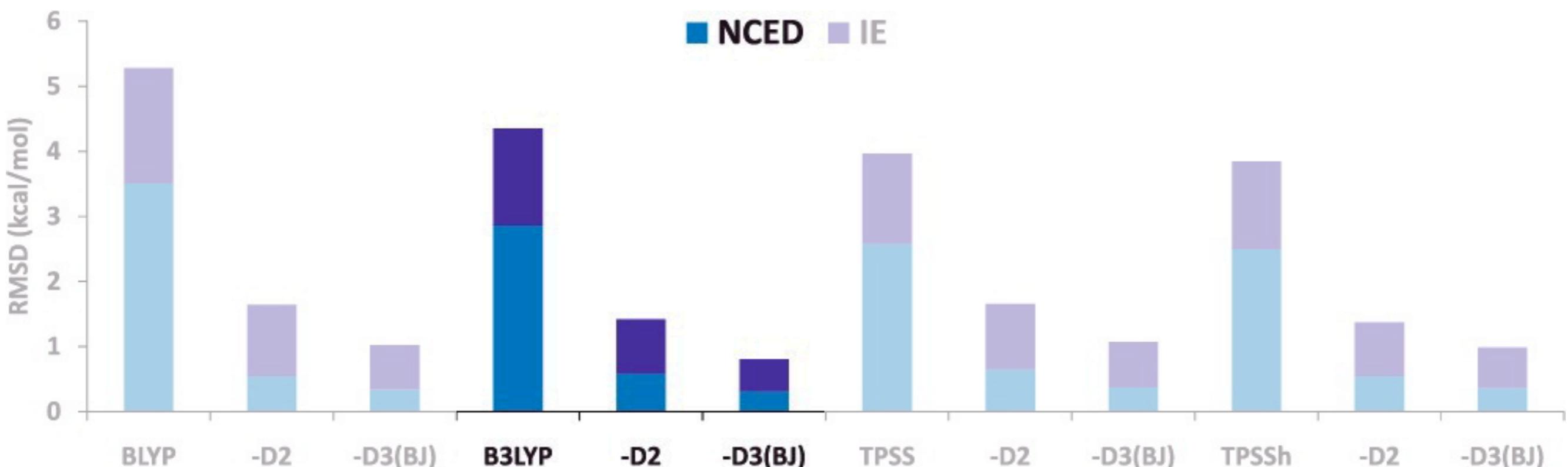
revPBE: increase of the gradient contribution to the exchange

Theoretical Method (Functional)

\hat{H}_e

Which Functional is Best for ...?

NCED 91 1744 non-covalent dimer interactions
IE 755 isomerisation energies

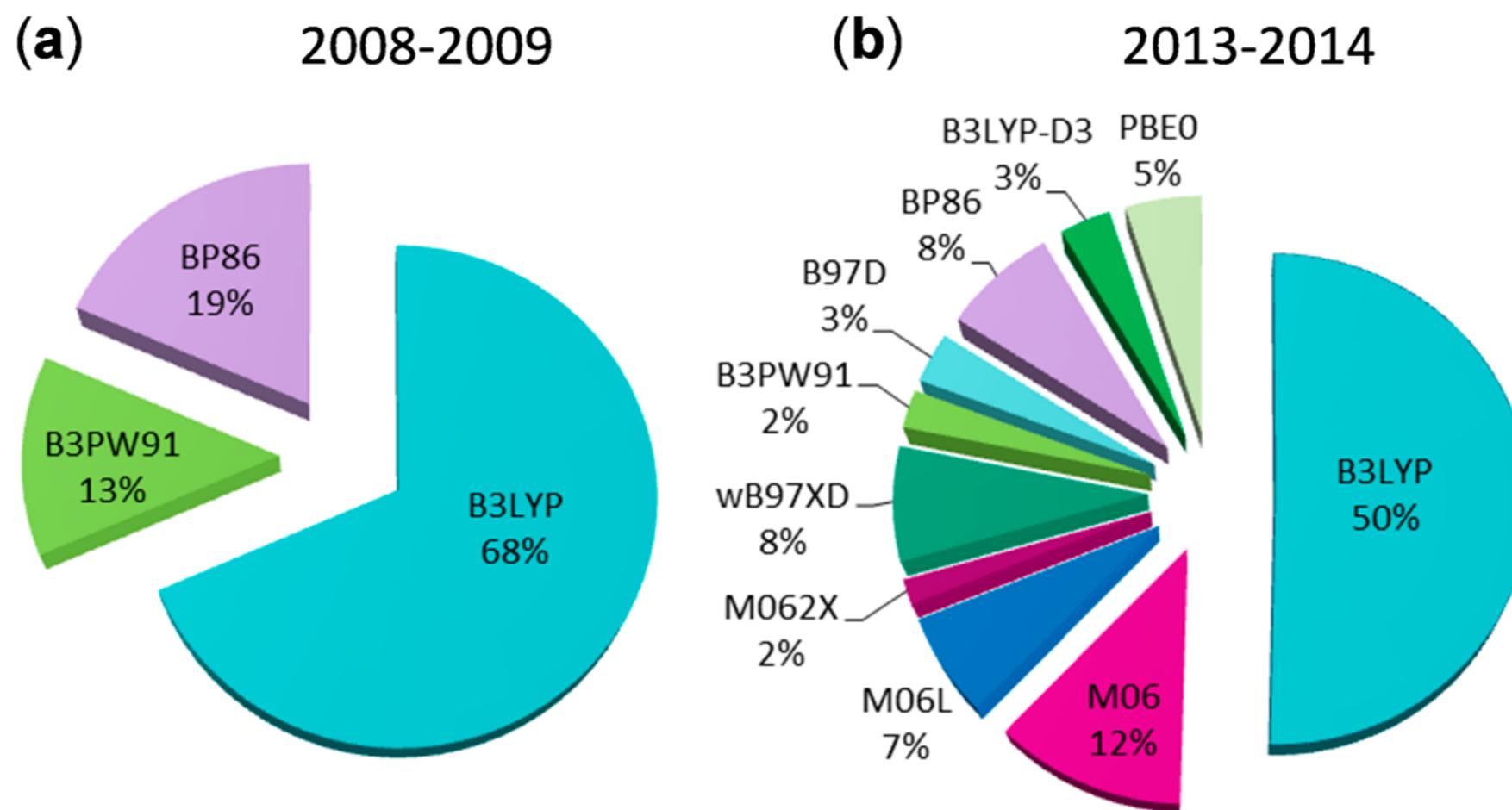


Density Functional Theory

\hat{H}_e

Which Functional is Best for TM Optimisation?

Popular DFT methods used for Pd, Ni, Rh, and Ir



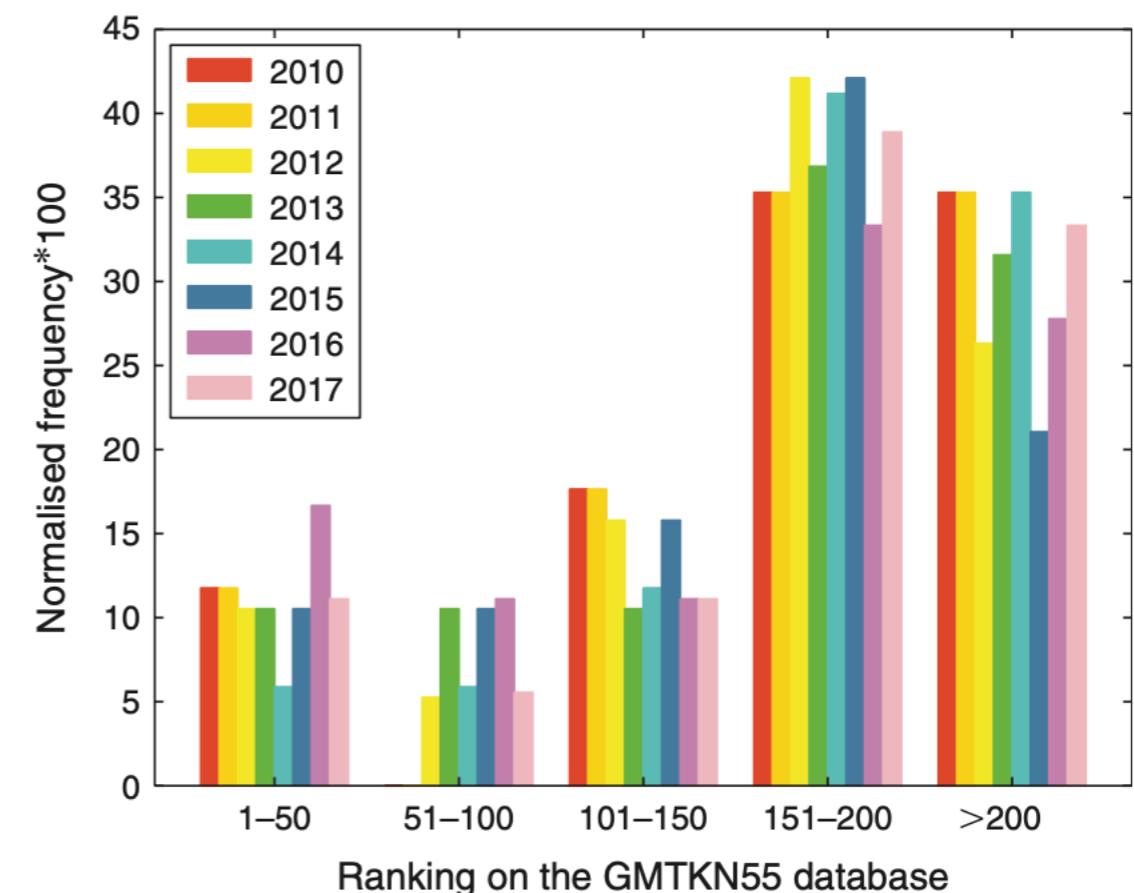
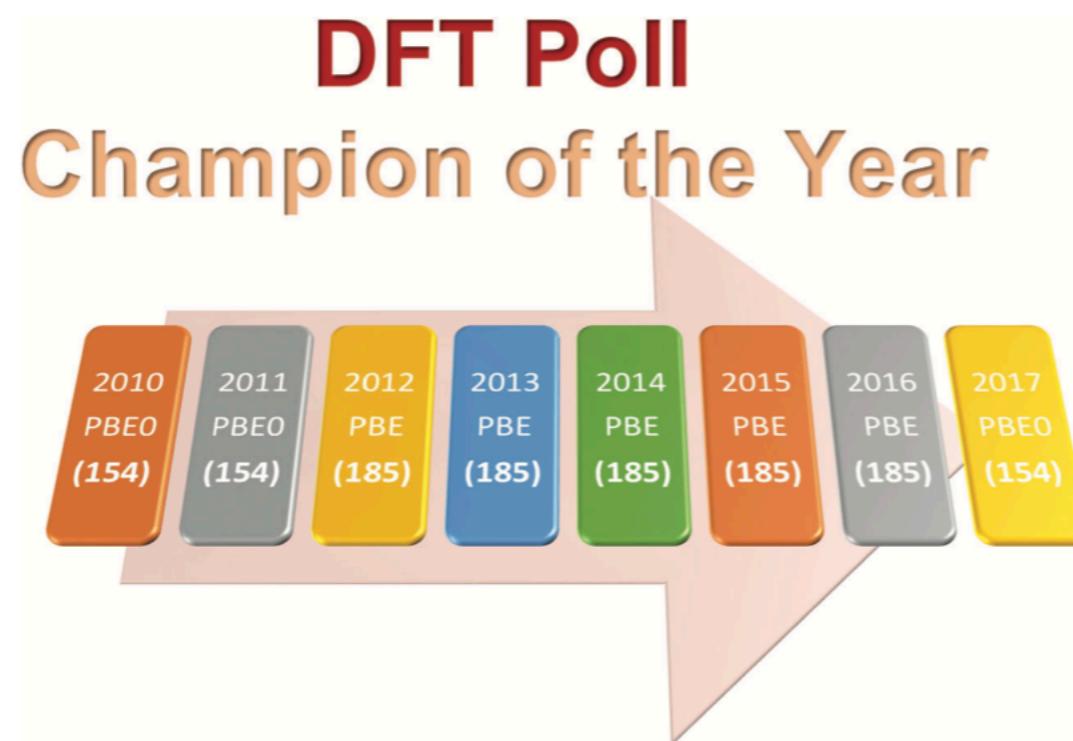
the dominance of B3LYP appears to be fading... (in GMTKN55, it only ranks 197th)

Density Functional Theory

\hat{H}_e

Popularity versus Accuracy and Reliability

(ranking for the GMTKN55 benchmark database)



Even with DFT-D3 correction, PBE and PBE0 rank 140th and 77th for GMTKN55. B3LYP-D3 appears 72nd

To keep in mind

- DFT is very useful for exploring many aspects of chemistry, but be careful!

Assess the methodology (functionals and basis sets)

Assess conformational flexibility

- DFT is not perfect:

Assumes the system is well described by a single electronic configuration

Does not describe long-range dispersion effects well

- The field is moving rapidly

New functionals and approaches being developed all the time

... modern functionals are ‘sacrificing physical rigor for the flexibility of empirical fitting’...as a result DFT is in need of new strategies for functional development’

Perdew et al. Science 2017, 6320, 49

Inaccuracies in DFT

... It all goes back to deficiencies in the treatment of exchange-correlation

$$E[\rho] = T[\rho] + E_{Ne}[\rho] + E_H[\rho] + E_x[\rho] + E_c[\rho]$$

- Dispersion Interactions
- Self-interaction Error
- Multi-reference problems

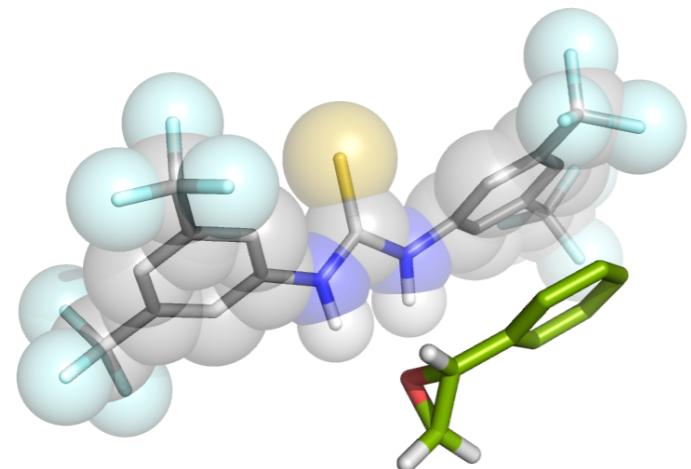
B3LYP seems to (luckily) cancel out the first two

Inaccuracies in DFT

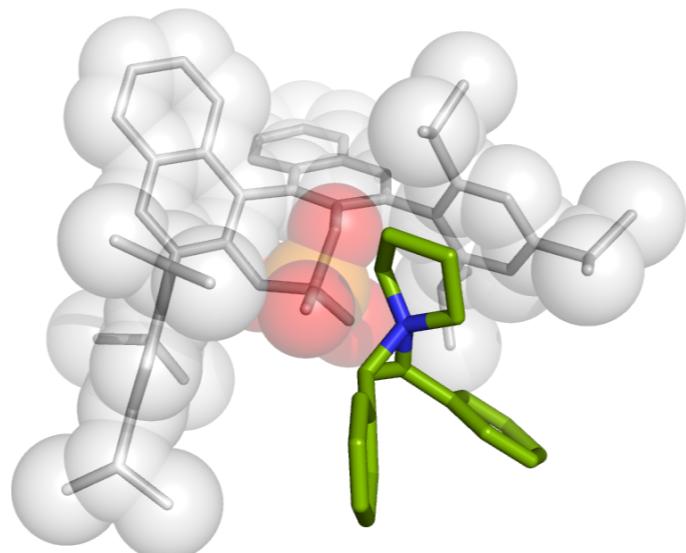
Non-covalent interactions

Dispersion arises from simultaneous excitation of electrons on different fragments – correlation problem

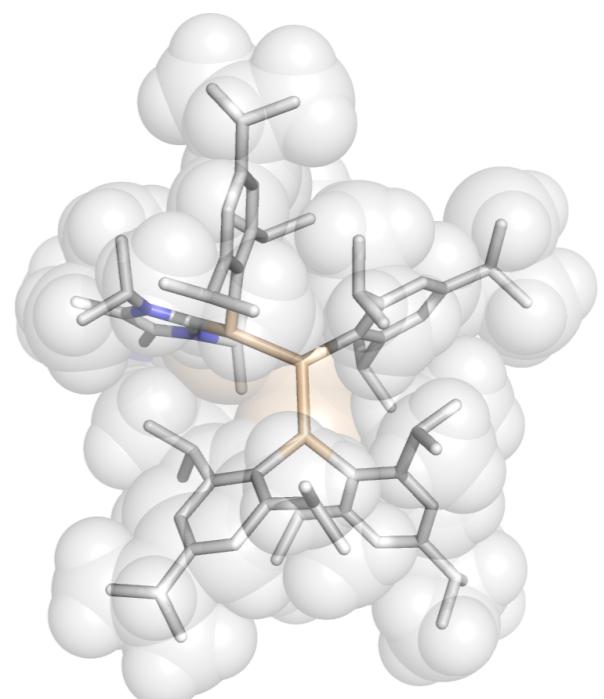
$$E_{\text{DFT-D}} = E_{\text{DFT}} + E_{\text{disp}}$$



Hydrogen Bonding & cation - π



Ion-pairing



Dispersion

Inaccuracies in DFT

D3 correction: includes both two- and three-body components

$$E_{\text{disp}} = E^{(2)} + E^{(3)}$$

$$E^{(2)} = -S_6 \sum_{A < B} V_6^{\text{PP}}(Z_A, Z_B, R_{AB}) - S_8 \sum_{A < B} V_8^{\text{PP}}(Z_A, Z_B, R_{AB})$$

$$E^{(3)} = \sum_{A < B < C} V^{\text{TBP}}(Z_A, Z_B, Z_C, R_{AB}, R_{AC}, R_{BC}, \cos \theta_a, \cos \theta_b, \cos \theta_c)$$

$$V^{\text{PP}} = \frac{\sqrt{C_6^A C_6^B}}{R_{AB}^6} f(R_{AB})$$

Inaccuracies in DFT

- Self-interaction Error

$$J[\rho] = \frac{1}{2} \int d\vec{r}_1 \int d\vec{r}_2 \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}}$$

It includes interaction of each electron with all – including itself!

In HF theory the self-interaction is exactly cancelled by exchange

Dissociation of radicals (H^{+2})

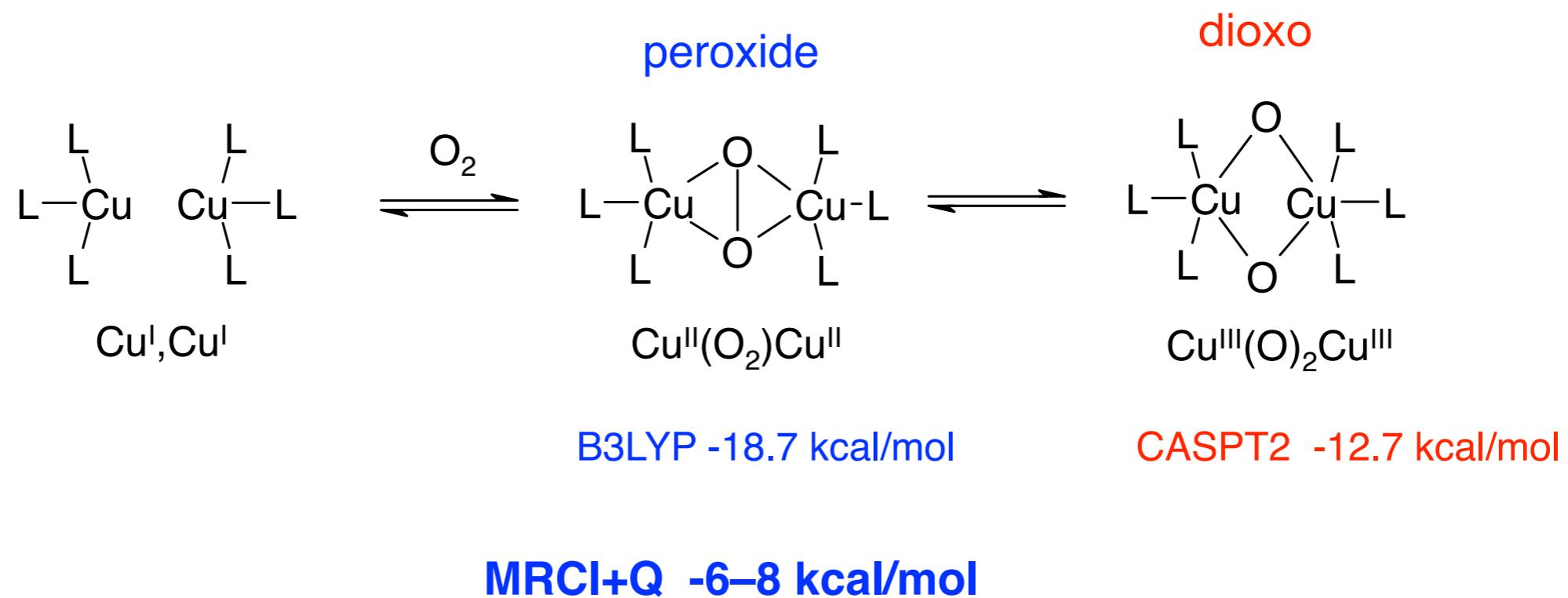
Activation Energies

Size-dependent

Inaccuracies in DFT

Multi-reference treatment

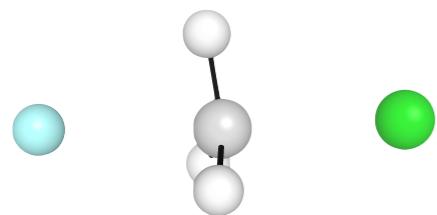
Using B3LYP two minima are found, the peroxide form is significantly more stable. This system has been reinvestigated using a more accurate multi-reference method and large basis sets



B3LYP gives qualitatively the correct answer

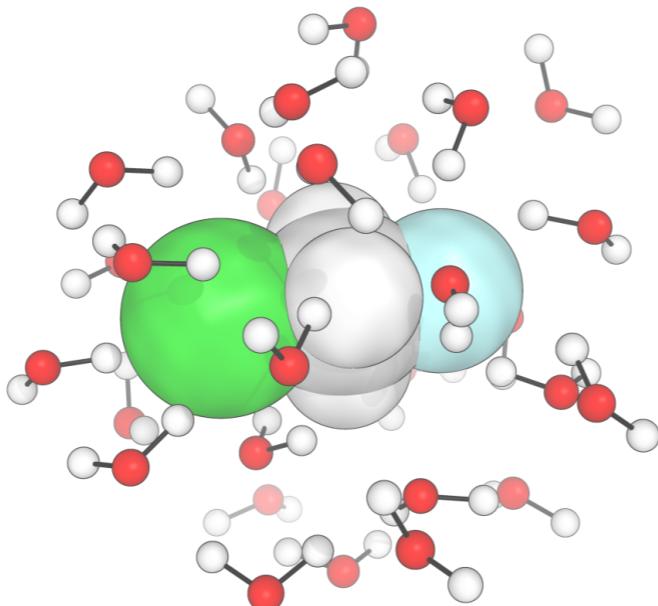
Inaccuracies in DFT

- Description of electronic structure ✓
- Dynamic processes ✗



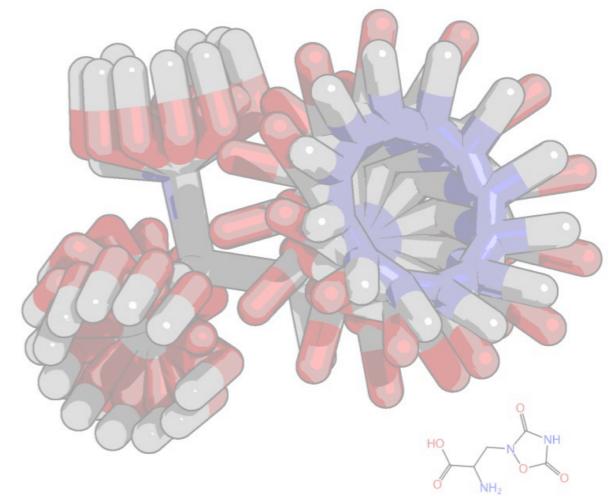
Possible but expensive

- Solvation ✗
- Entropy ✗



Only possible using MM
Implicit models using QM

- Conformational sampling ✗



Possible using MM if solvent is included

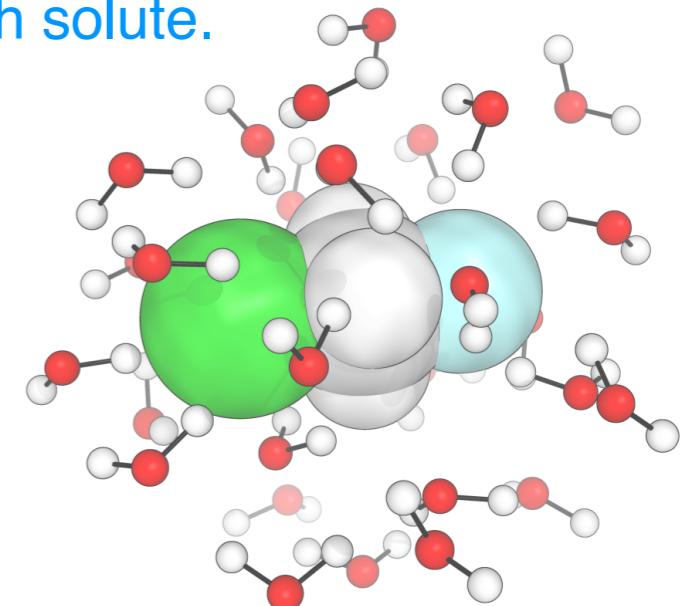
Inaccuracies in DFT

Specific Solvation: short range interactions of solvent molecules with solute.

Approach: adding solvent molecules into the calculation:

Where do you place the solvent molecules?

How many solvent molecules are needed?



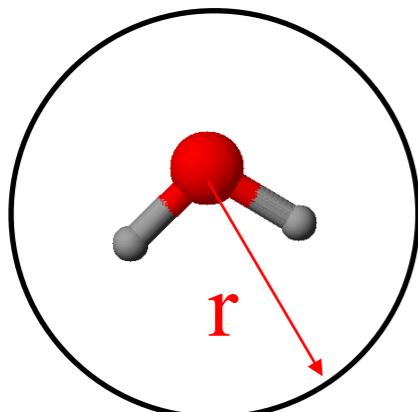
Long Range Effects: screening of charges by solvent polarisation.

Approach: Self-Consistent Reaction Field (SCRF) Methods. The solute is placed in a cavity defined by atomic van der Waals radii.

Different models (PCM, COSMO...) differ in e.g. how the charge distribution of the solute or the cavity surface is calculated, or in the way the cavity is defined.

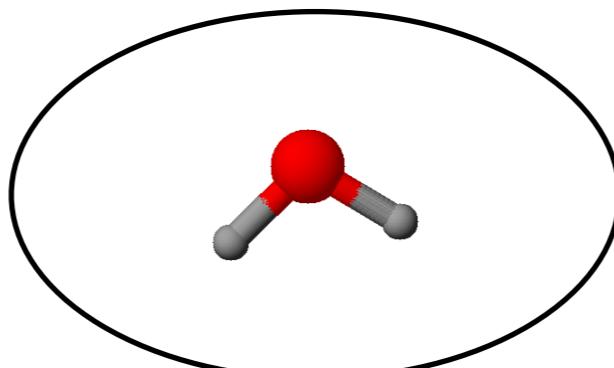
Implicit Solvation – Type of Solute cavity

Spherical
(Born)



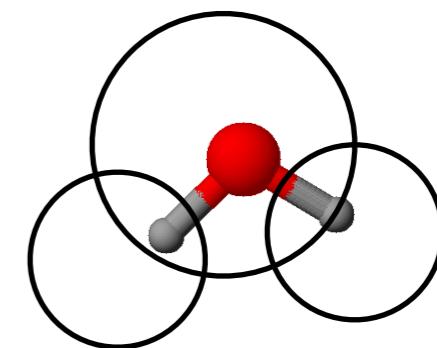
Sphere

Ellipsoidal
(Onsager)

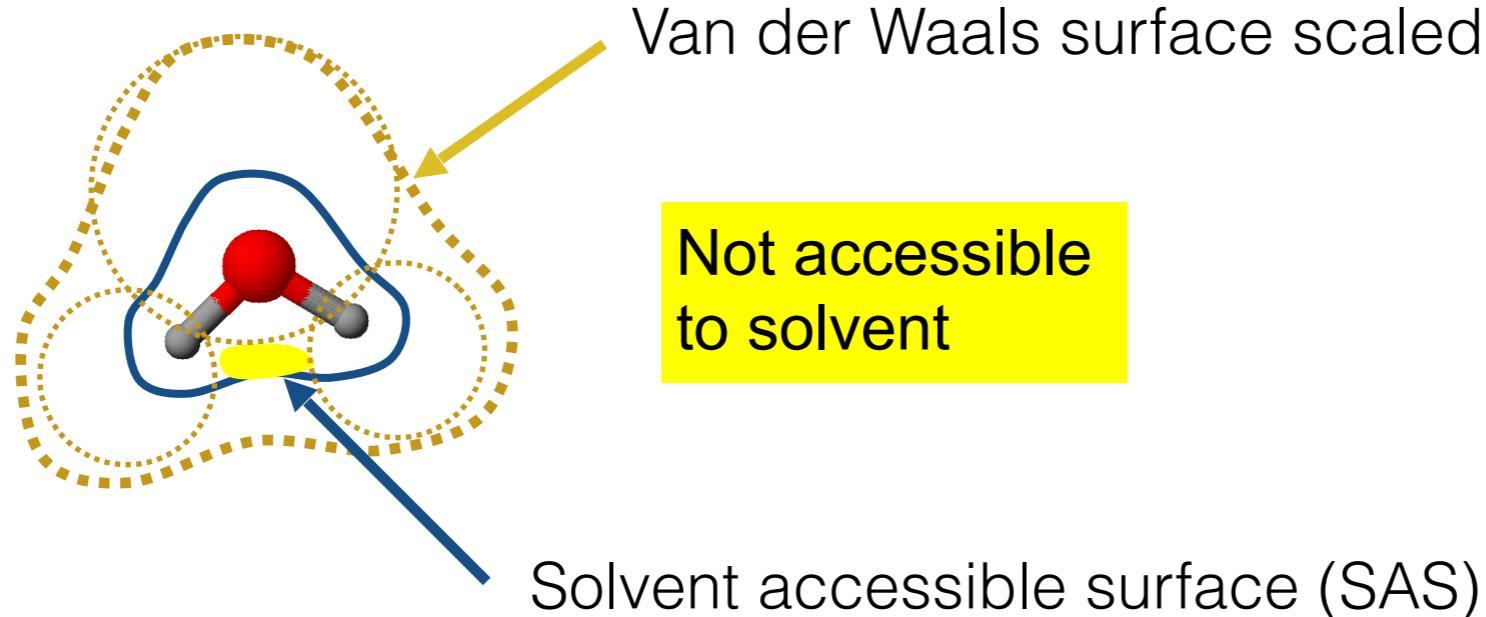


Ellipsoid

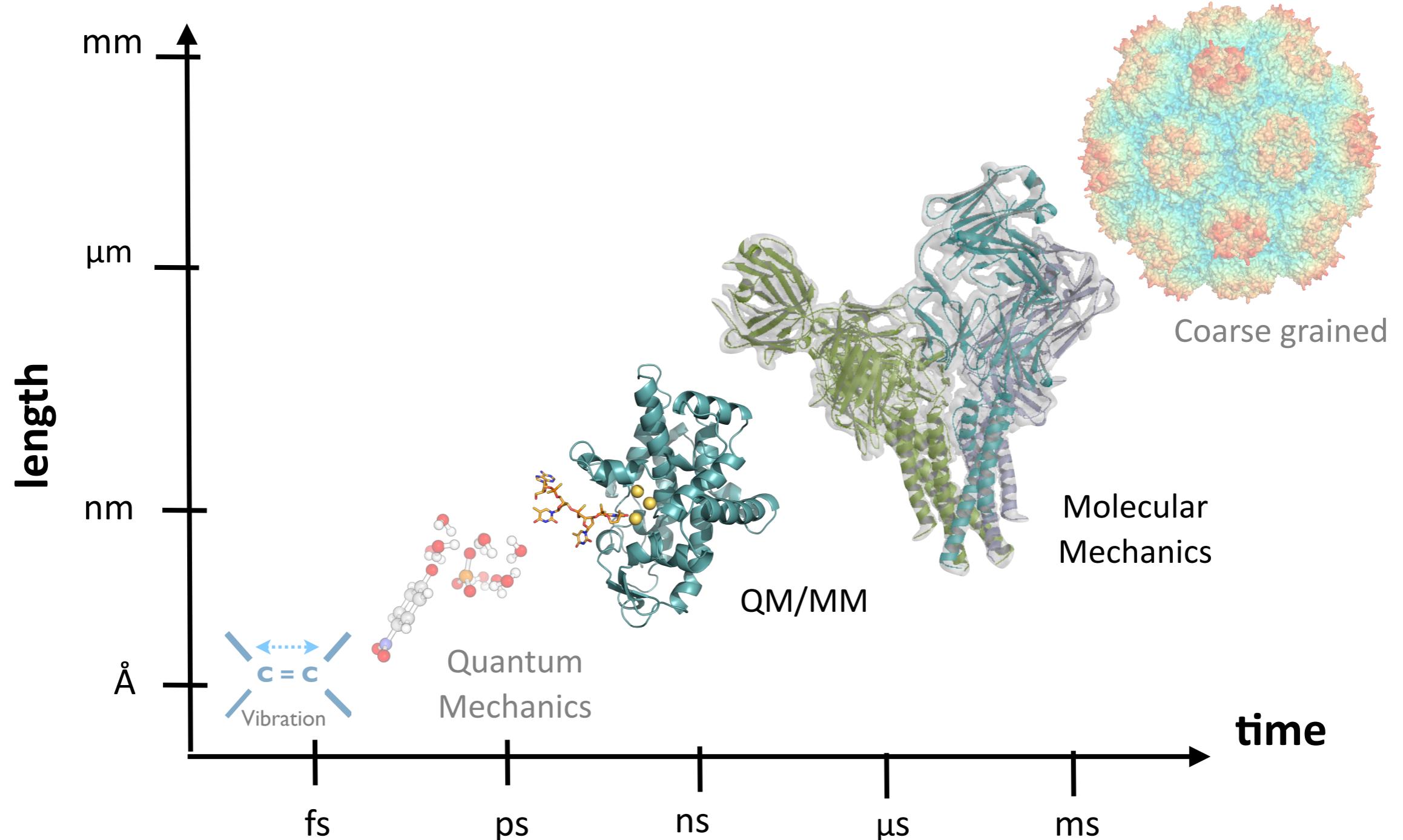
van de Waals
(Kirkwood)



Van der Waals surface

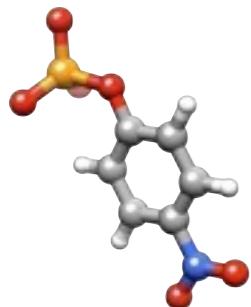


Classical Approaches



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

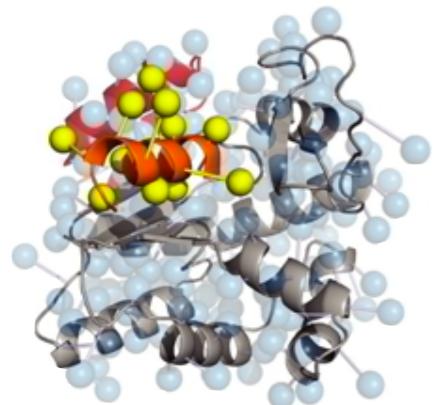
Classical Approaches



Quantum
Mechanics

Electronic Structure
(Schrödinger equation)

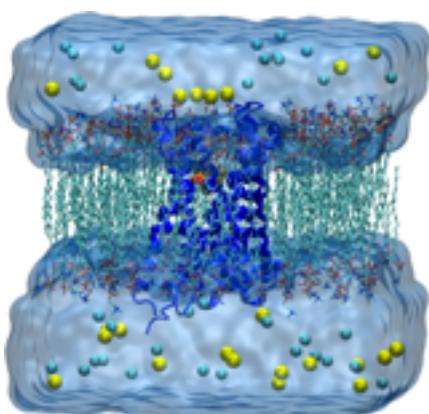
Accurate \Rightarrow Expensive (N^4)



Molecular
Dynamics

Empirical Forces
(Newton's equations/
Random move)

LESS accurate \Rightarrow Faster ($N \log N/N^2$)



Coarse grain

Reduction of number of
degrees of freedom
(Molecular/Langevin
Dynamics)

Interaction details eliminated

Classical Approaches

Which System Do I Have?

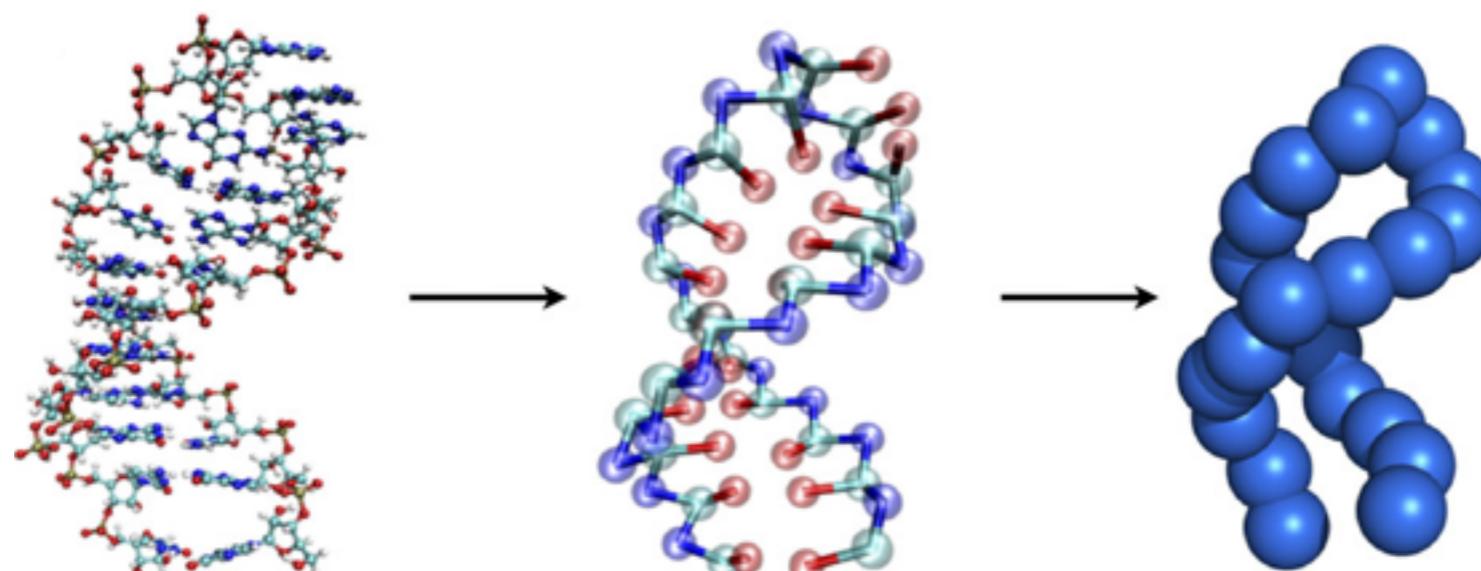


What Do You Want to Compute (and Why)?



How much detail we need to describe the phenomenon of interest?

What is the level of computational resources available?



Classical Approaches

- There is No Such Thing as A Structure!
- Is the crystal structure the biologically relevant conformer?
- Molecules are Dynamic, Not Static (Conformational Ensemble)



Molecular Dynamics

Position (r)



Crystal coordinates/homology models

Charge (q)

Bond information



Empirical energy function(CHARMM,OPLSAA,AMBER)

Momentum
($m + v$)



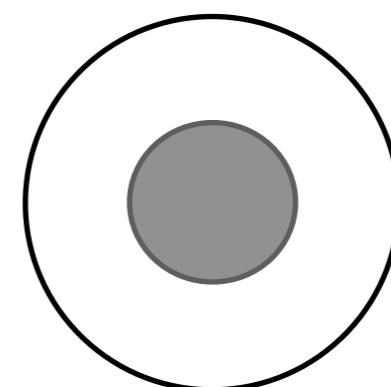
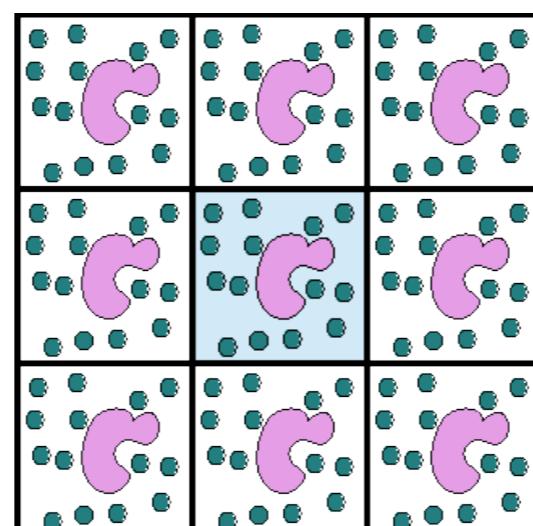
Newton's equations

Environment

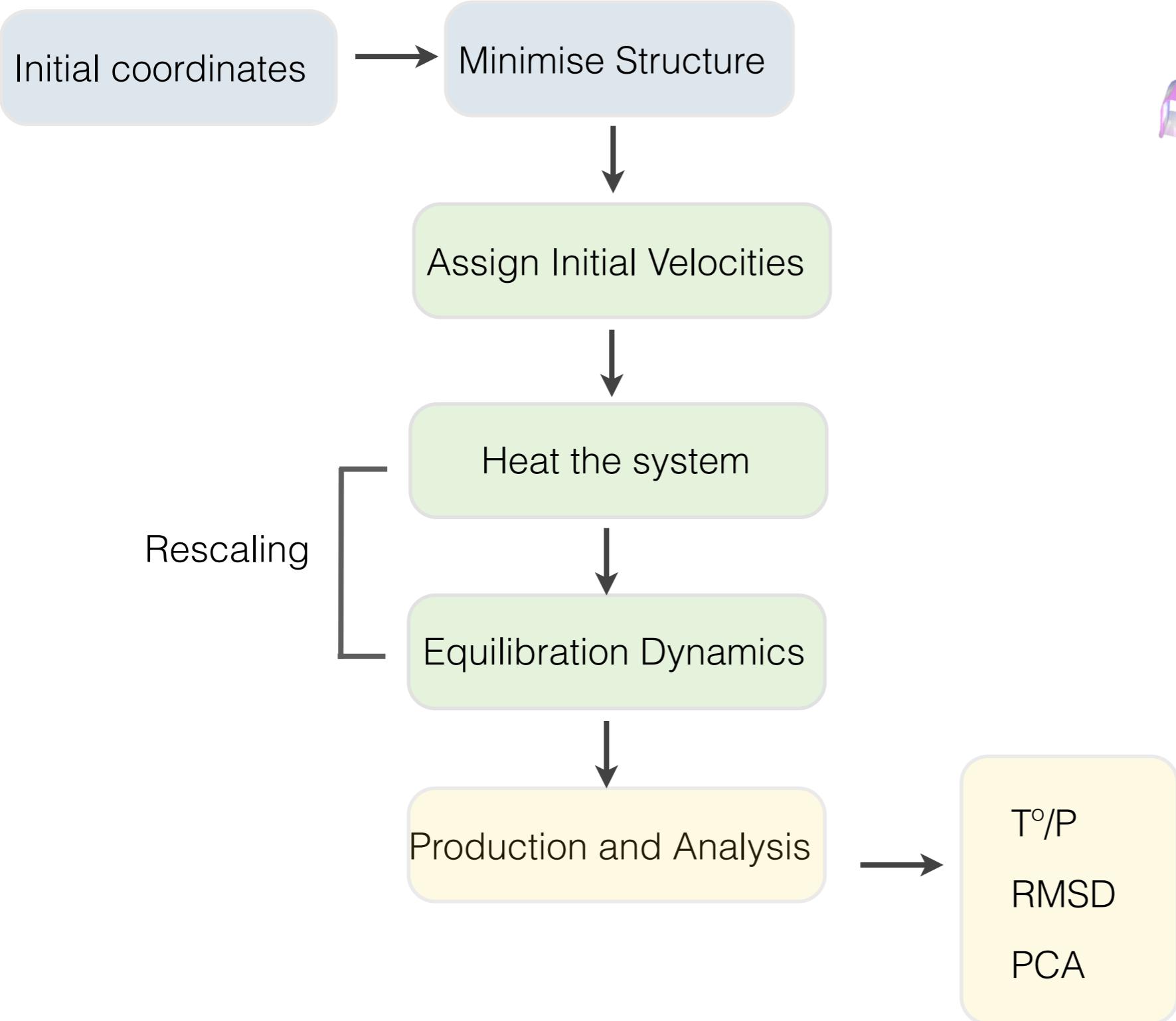


Solvent (Implicit/Explicit:TIP3P, SPC)

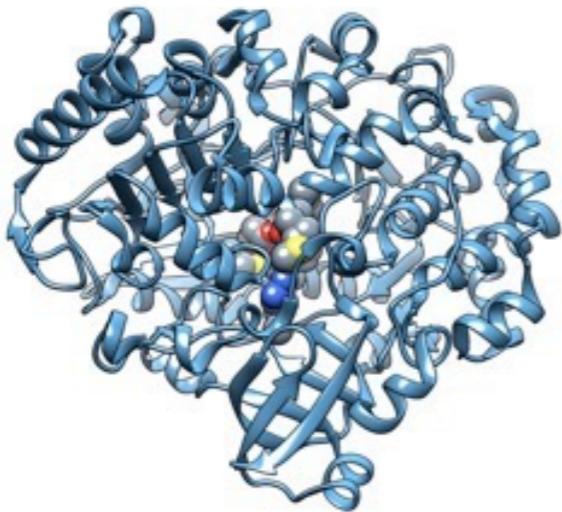
Spherical/Periodical boundaries



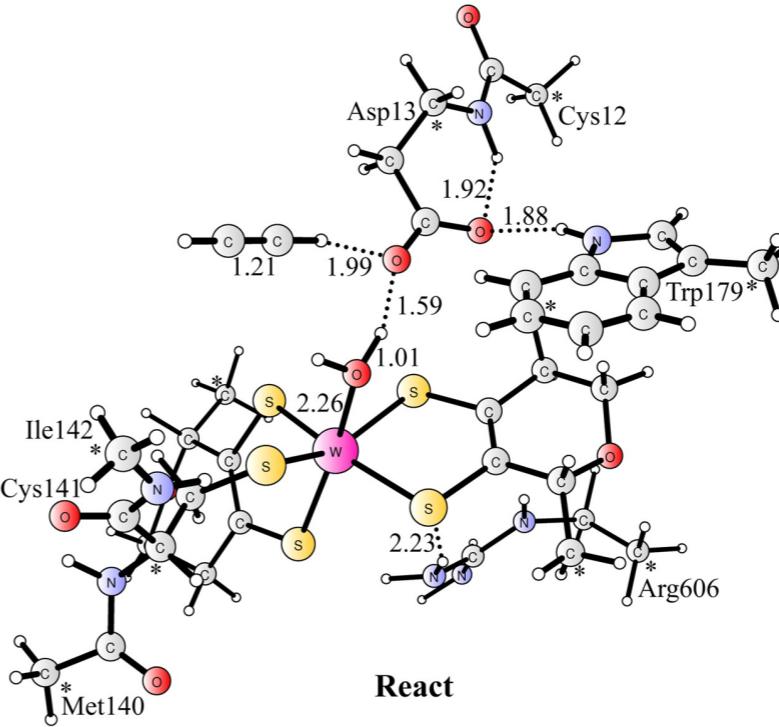
Molecular Dynamics



QM (Cluster Model)



Tungsten-dependent Acetylene Hydratase



Liao R et al. PNAS 2010;107:22523-22527

Position (r)



Crystal coordinates.

Model of the active site (100-200 atoms!). Scale N^4

Charge (q)
Bond information



Electronic structure

Energy/Geometry



Schrödinger eq.

Minimization Algorithms

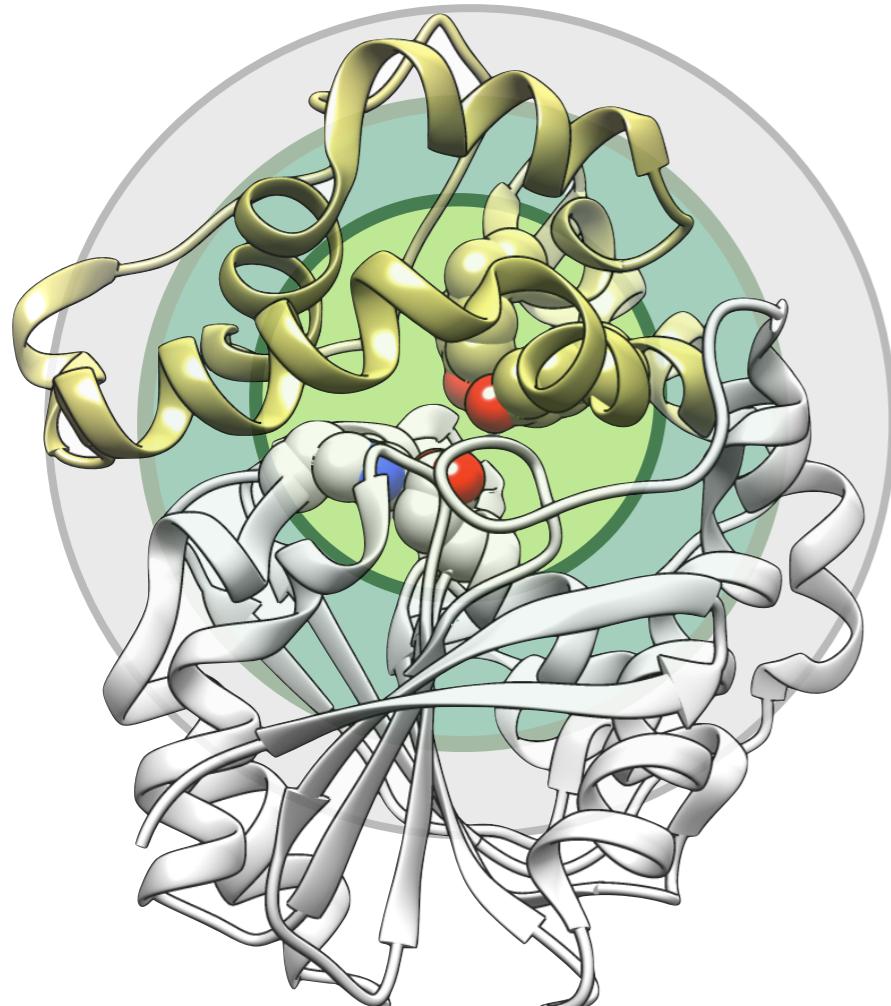
Environment



Steric / electrostatic effects from environment
are implicitly considered

QM/MM methods – Enzymatic reactions

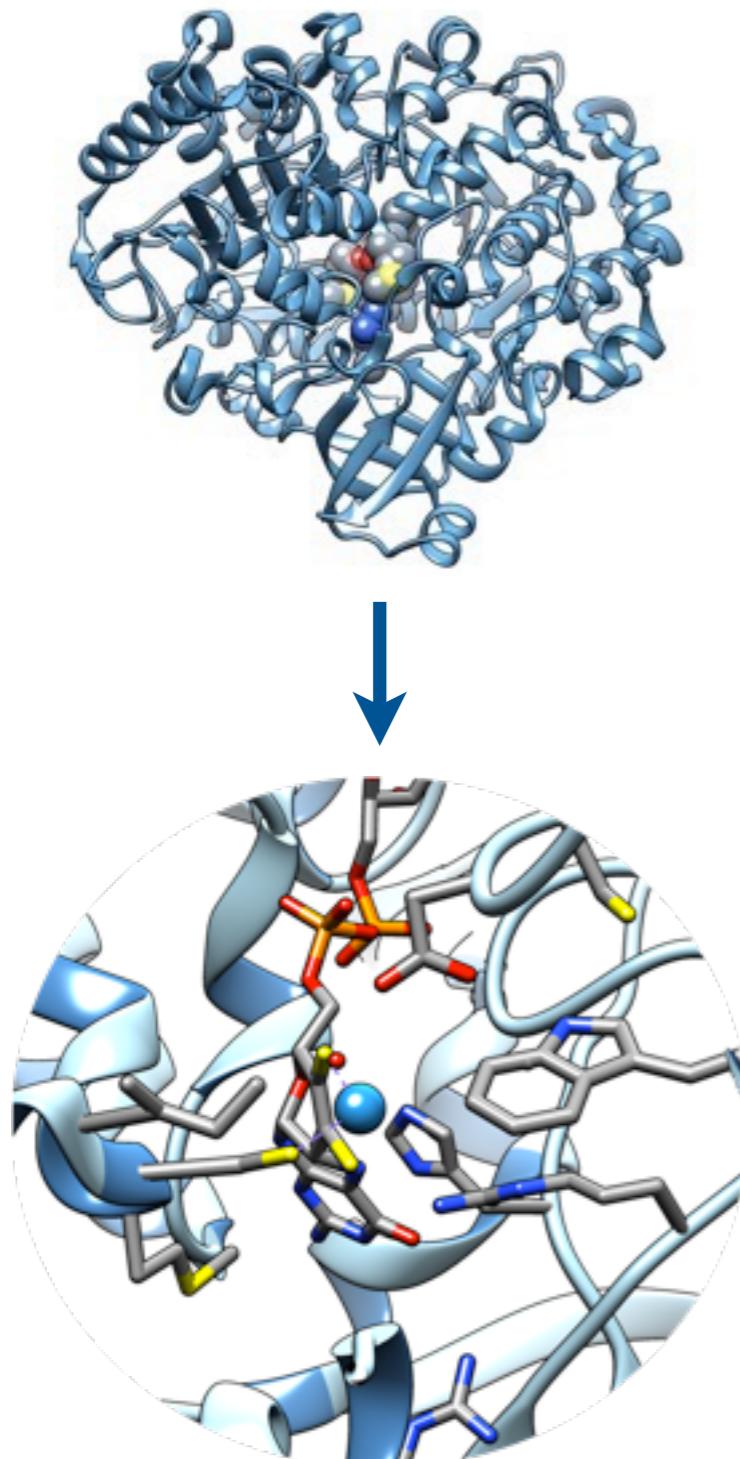
Combine the best of each world...



- ✓ The whole solvated protein is chosen as the model.
- ✓ Complete description of breaking/forming bond processes
- ✓ Explicitly incorporating effects of the environment
 - *How to handle QM/MM boundary?*
 - *Is the sampling problem solved?*

QM/MM methods – Enzymatic reactions

Case Study: Acetylene Hydratase



- QM-only (32 atoms)
Two-step reaction profile
- QM(**116 atoms**)/MM
One-step mechanism (very high barrier)
Influence of the key groups neglected in the QM-only
- QM(**157 atoms**)/MM
Two-step mechanism
Reasonable overall barrier
Change in the rate limiting step

QM/MM methods – Enzymatic reactions

Advantages & Challenges

- ✓ The region of most interest is treated with more **accurate methods**.
- ✓ **Bond broken/form** explicitly described
- ✓ The **environment** is considered at a lower level of theory.
- Large **computer time** in order to evaluate the QM energy.
- Limited **configurational sampling**. Reaction pathways will strongly depend on protein configurations.