

# CHM 673

## Lecture 23: Fragmentation methods

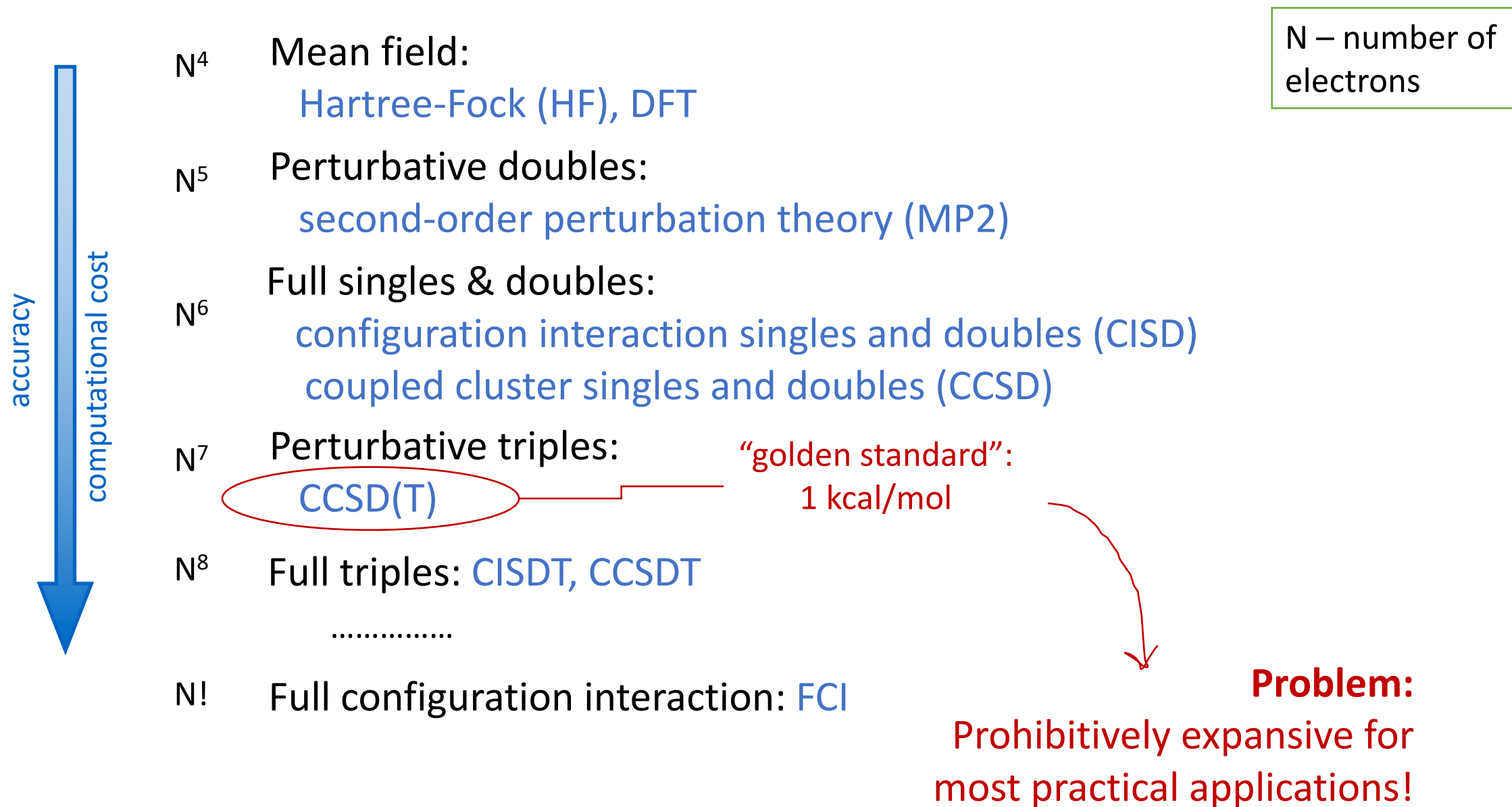
---

Suggested reading:

DOI [10.1021/cr200093j](https://doi.org/10.1021/cr200093j)

DOI [10.1007/s00214-006-0143-z](https://doi.org/10.1007/s00214-006-0143-z)

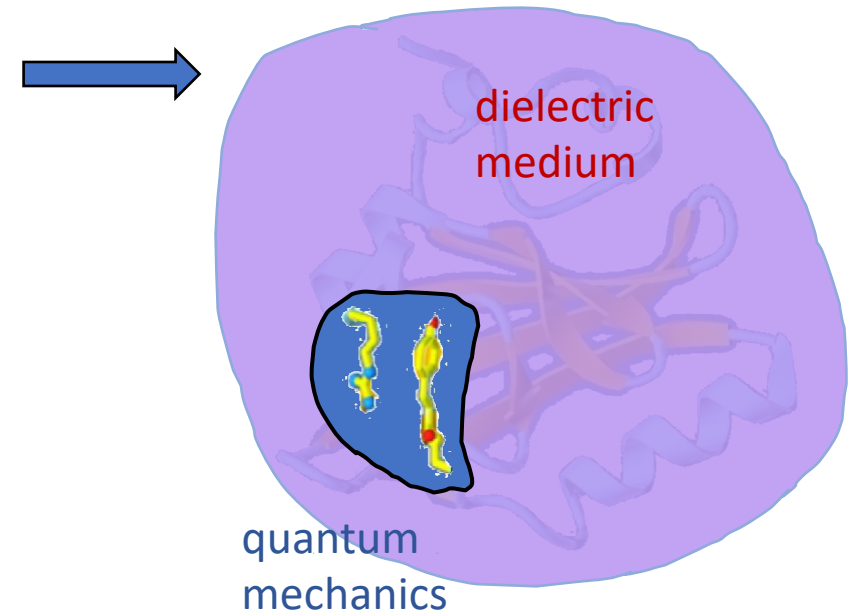
# Hierarchy of electronic structure methods



# Smart models for big problems

- Implicit solvent
  - Self-consistent reaction field
  - Polarizable continuum model
  - Langevin dipoles (Warshel)
- Explicit solvent: QM/MM
  - Force-fields, EFP
  - ONIOM (Morokuma)
- Fragmentation techniques
  - Fragment molecular orbitals (FMO) (Kitaura)
  - Divide-and-Conquer (Merz)
- Semiempirical and DFT

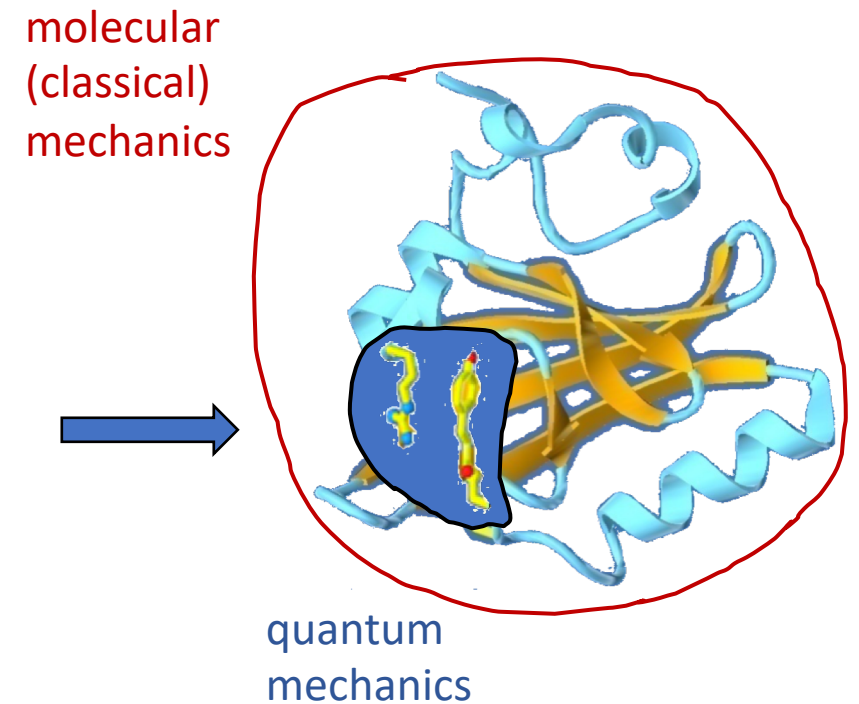
....



# Smart models for big problems

- Implicit solvent
  - Self-consistent reaction field
  - Polarizable continuum model
  - Langevin dipoles (Warshel)
- Explicit solvent: QM/MM
  - Force-fields, EFP
  - ONIOM (Morokuma)
- Separation techniques
  - Fragment molecular orbitals (FMO) (Kitaura)
  - Divide-and-Conquer (Merz)
- Semiempirical and DFT

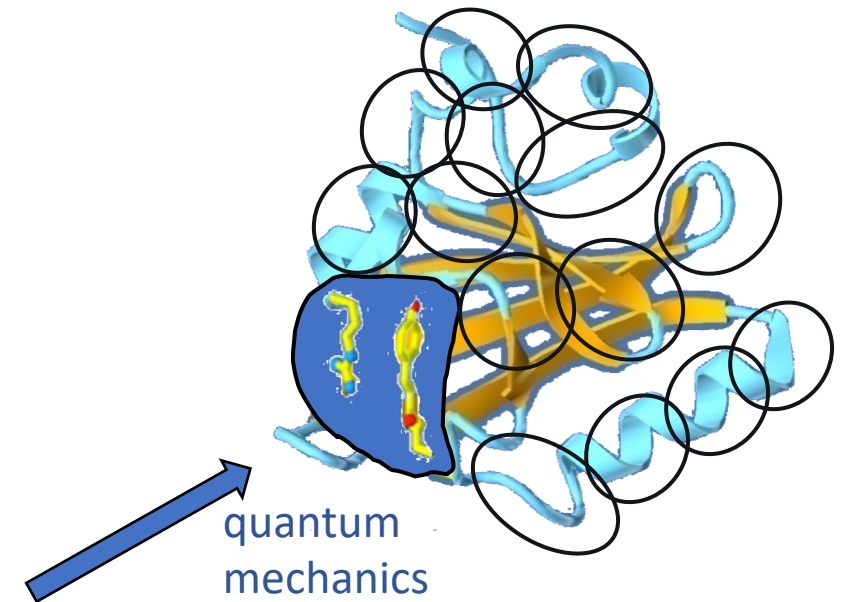
....



# Smart models for big problems

- Implicit solvent
  - Self-consistent reaction field
  - Polarizable continuum model
  - Langevin dipoles (Warshel)
- Explicit solvent: QM/MM
  - Force-fields
  - ONIOM (Morokuma)
- Fragmentation techniques
  - Fragment molecular orbitals (FMO) (Kitaura)
  - Divide-and-Conquer (Merz)
- Semiempirical and DFT

....



# Fragmentation methods

---

## Fragmentation methods



```
graph TD; A[Fragmentation methods] --> B[Density-based]; A --> C[Energy-based];
```

### Density-based

Density of the whole system is computed  
Total system energy is computed from  
the system density

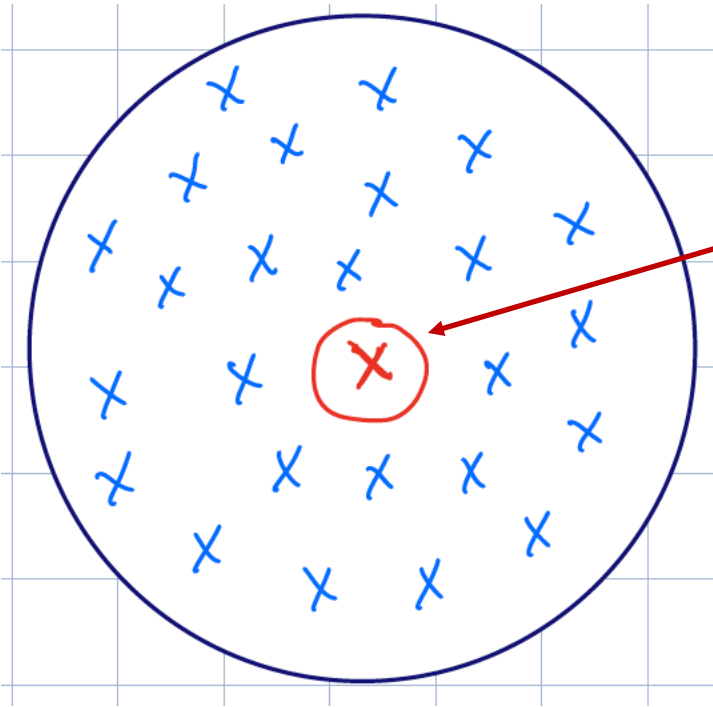
Fragment Molecular Orbital (FMO)  
Divide-and-Conquer (DC)

### Energy-based

Total system energy is computed directly  
from fragment energies

Systematic Fragmentation Method (SFM)  
Molecular Fractionation with Conjugate  
Caps (MFCC)

# Fragment Molecular Orbital method (FMO)



Molecular system is separated into fragments (monomers)  $x$   
 Each monomer density is computed in electrostatic bath of other fragments

For fragment  $x$ :

$$\tilde{F}^x C^x = S^x C^x \varepsilon^x$$

$$\tilde{F}^x = \tilde{h}^x + G^x$$

$$\tilde{h}_{\mu\nu}^x = h_{\mu\nu}^x + \sum_{A \in k} \left\langle \mu \left| \frac{-Z_A}{|\mathbf{r} - \mathbf{r}_A|} \right| \nu \right\rangle + \sum_{\lambda, \sigma \in k} D_{\lambda, \sigma}^k \langle \mu \lambda | \nu \sigma \rangle$$

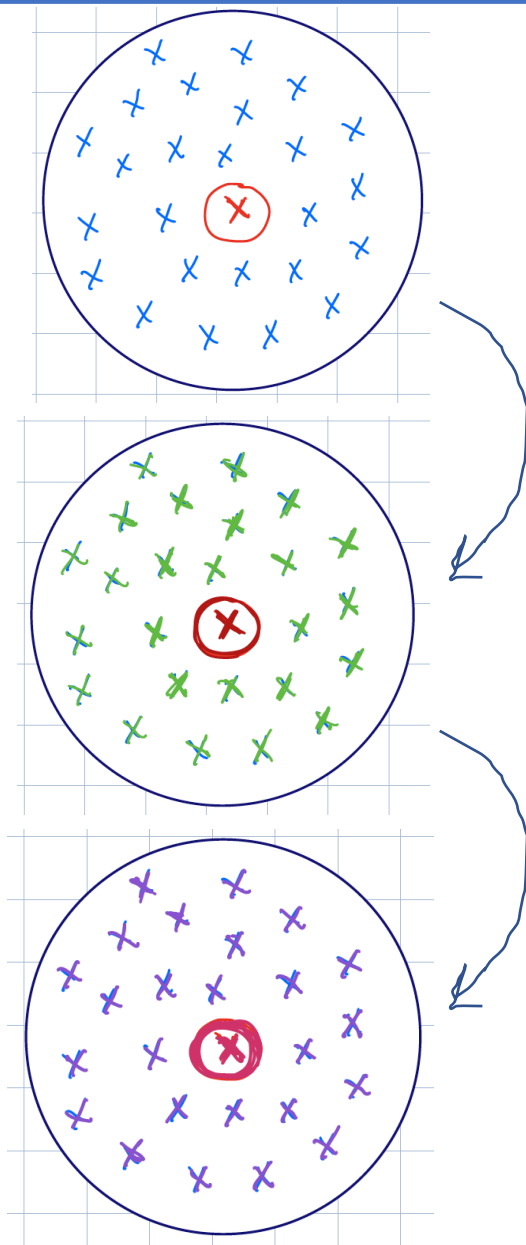
$k$  – other fragments

For large fragment-fragment distances

$$\sum_{\lambda, \sigma \in k} D_{\lambda, \sigma}^k \langle \mu \lambda | \nu \sigma \rangle \approx \sum_{A \in k} \left\langle \mu \left| \frac{P_A}{|\mathbf{r} - \mathbf{r}_A|} \right| \nu \right\rangle$$

$P_A$  – partial atomic charges

# Fragment Molecular Orbital method (FMO)



**Step 1.** Self-consistent loop over electronic densities of all fragments

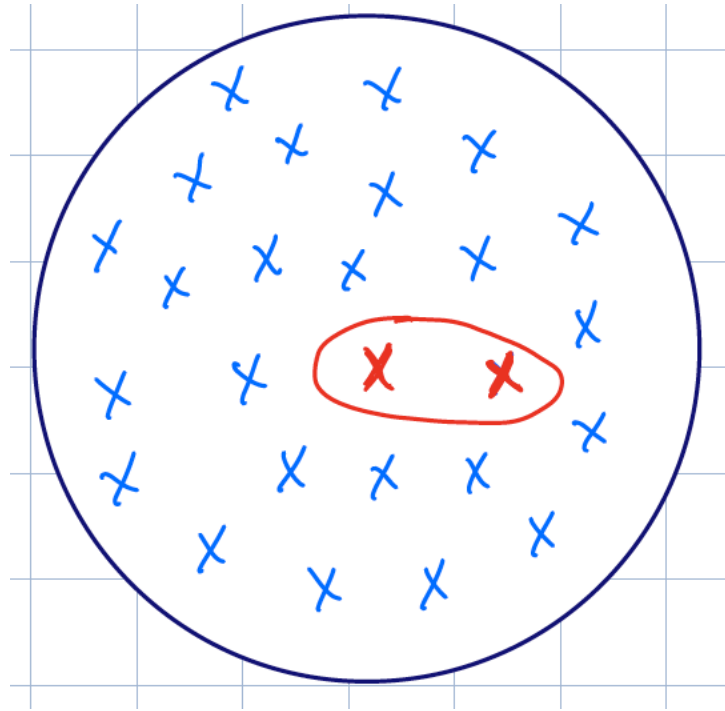
Electronic density of a fragment depends on electronic densities (molecular orbitals, partial charges) of other fragments

Effectively, double self-consistency should be reached:

- Each fragment density is converged till self-consistency (i.e., HF or DFT with modified Fock/Kohn-Sham matrix)
- Such HF or DFT calculations for all fragments are repeated until no change in densities of all fragments is observed



# Fragment Molecular Orbital method (FMO)



**Step 2.** Dimer calculations in presence of electrostatic bath of other fragments are conducted (to account for long-range interactions and inter-fragment correlation)

The total energy is:

$$E^{FMO2} = \sum_i^{nfrag} E_i + \sum_{i>j}^{nfrag} (E_{ij} - E_i - E_j)$$

- Can be extended to inclusion of trimer energies (FMO3)
- Only dimers composed of neighboring monomers can be considered (to save computational cost)


# Fragment Molecular Orbital method (FMO)

Main advantage: dramatic savings in computational cost!

N electrons

“standard” QM calculation  $\rightarrow cost \sim N^3$

FMO with  $n$  fragments  $\rightarrow cost \approx n f \left( \frac{N}{n} \right)^3 = f \frac{N^3}{n^2}$

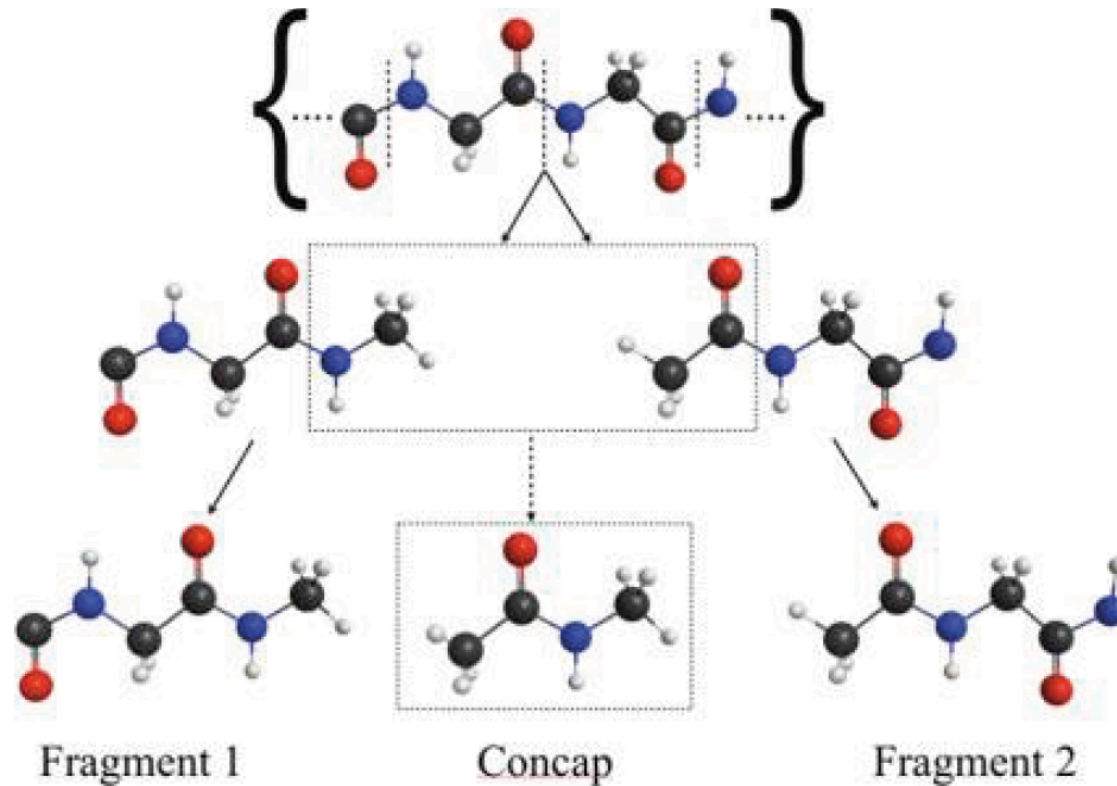


f - prefactor

Problems:

1. Hard to describe systems with delocalized electrons (graphenes, metals, etc)
2. Hard to achieve basis set convergence: larger basis sets result in large charge-transfer between fragments  $\rightarrow$  hard to converge monomer SCFs
3. Splitting of covalent systems into fragments is not easy (developed for C-C, C-N bonds, not general)

# Molecular fragmentation with conjugated caps (MFCC)



For a protein P composed of N amino acids:

$$P = nA_1A_2A_3...A_N$$

$$n = \text{NH}_3^+(\text{NH}_2)$$

$$A_N = \text{R}_N\text{CHCOO}^-$$

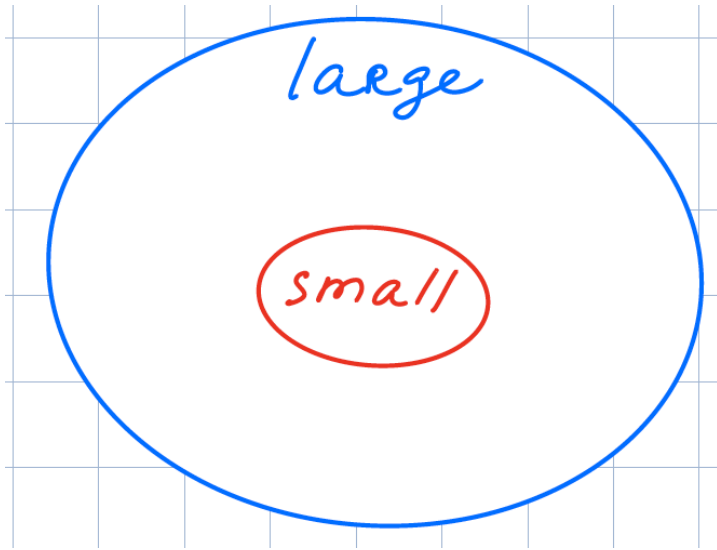
Density of the system is computed from density of monomers as:

$$\rho = \sum_i^N \rho_{A_i} - \sum_i^{N-1} \rho_{CC_i}$$

Protein – ligand interaction energy:

$$E(P - L) = \sum_i^N E(L - \text{Cap}^{i-1*} A_i \text{Cap}^i) - \sum_i^{N-1} E(L - \text{Cap}^i \text{Cap}^i)$$

# ONIOM (Morokuma)



In the simplest form:

$$E = E_{large}^{low\ level} + (E_{small}^{high\ level} - E_{small}^{low\ level})$$

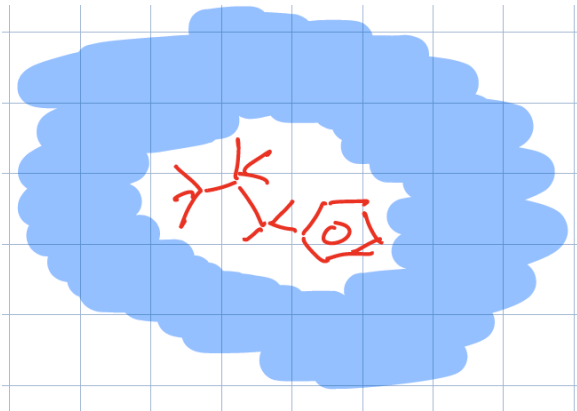
large: full system

small: "solute"

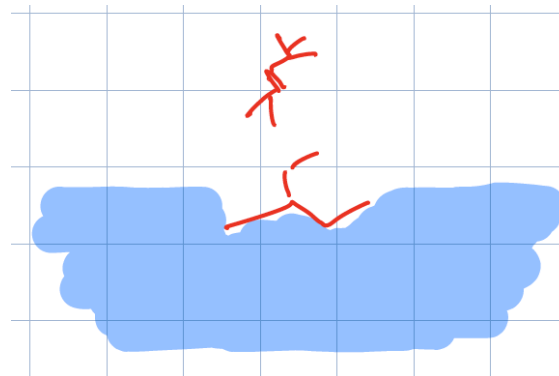
(large - small): "solvent"

High level: accurate QM method

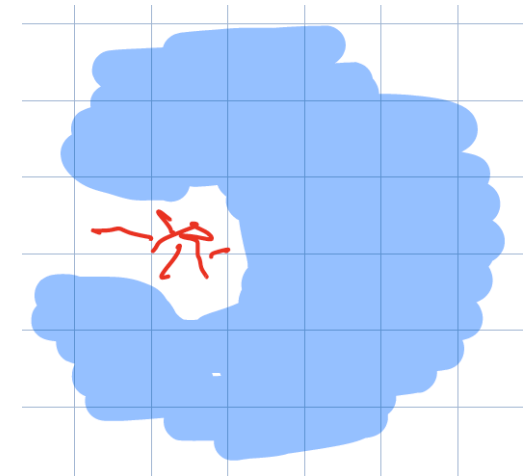
Low level: less accurate QM method or force field



solvation



surface reactions



protein-ligand interactions

# QM/MM models

---

## Advantage:

huge computational savings; often the only practical solution

## Challenges:

- How to describe interactions between QM and MM regions
- How to model covalent boundaries between QM and MM regions

# Molecular mechanics force fields

$$E = \underbrace{E_{str} + E_{bend} + E_{tors}}_{\text{bonded}} + \underbrace{E_{vdw} + E_{el}}_{\text{non-bonded}}$$

bonded

non-bonded

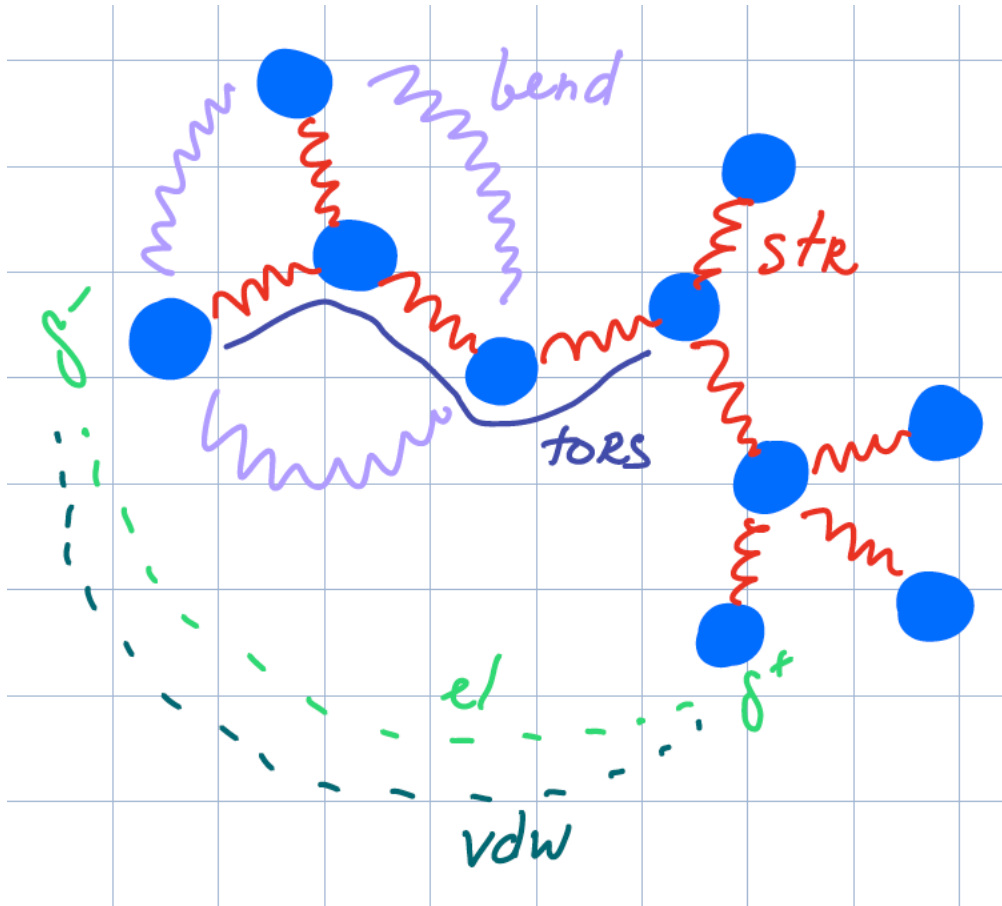
$$E_{str} = k^{AB} (R^{AB} - R_0^{AB})^2$$

$$E_{bend} = k^{ABC} (\Theta^{ABC} - \Theta_0^{ABC})^2$$

$$E_{tors} = \sum_{n=1} V_n^{ABCD} \cos(n\omega^{ABCD})$$

$$E_{vdw} = \varepsilon \left[ \left( \frac{R_0}{R_{AB}} \right)^{12} - 2 \left( \frac{R_0}{R_{AB}} \right)^6 \right]$$

$$E_{el} = \frac{Q^A Q^B}{\epsilon R_{AB}}$$



# Molecular mechanics force fields

---

N atoms

Bonded terms: computational cost  $\sim N$

Non-bonded terms: computational cost  $\sim N^2$

Parameterization cost of the force field:

bonded terms:  $\sim N^2$  (stretch),  $N^3$  (bend),  $N^4$  (torsion)

Non-bonded terms:  $\sim N$  (electrostatic),  $N$  or  $N^2$  (vdw)

# How to compute electronic charges on atoms?

No unique way!

1. AIM (atom-in-molecule) approach: numerically sum density around each atom; the boundaries between atoms are determined based on density gradient surfaces

2. Mulliken and Lowdin population analysis:

$$N = 2 \sum_i^{N/2} \int dr \psi_i^*(r) \psi_i(r) = \sum_{\mu\nu} [2 \sum_i^{N/2} C_{\mu i} C_{\nu i}] \int dr \phi_\nu^*(r) \phi_\mu(r) =$$

$$\sum_{\mu\nu} \underset{\substack{\uparrow \\ \text{density matrix}}}{D_{\mu\nu}} \underset{\substack{\uparrow \\ \text{overlap integral}}}{S_{\nu\mu}} = \sum_{\mu} (DS)_{\mu\mu} \quad \leftarrow \text{number of electrons associated with basis function } \mu$$

$$q_A = Z_A - \sum_{\mu \in A} (DS)_{\mu\mu} \quad \text{Mulliken}$$

$$q_A = Z_A - \sum_{\mu \in A} (S^{1/2} D S^{1/2})_{\mu\mu} \quad \text{Lowdin}$$

Mulliken and Lowdin analysis are strongly basis-set dependent!



# How to compute electronic charges on atoms?

## 3. Restrained electrostatic potential charges (RESP)

Compute QM electrostatic potentials at grid points  $r$  around a molecule:

$$\phi_{esp}(\mathbf{r}) = \sum_A^{nuc} \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\Psi^*(\mathbf{r}')\Psi(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}'$$

Find electronic charges on atoms by minimizing the error between QM-based potentials and the ones produced by electronic charges

Note:

RESP charges are dependent on selected grid

RESP charges might be conformation dependent

Charges inside large molecules (“buried charges”) might be not uniquely defined

# Electrostatic interactions

The point-charge model for electrostatics has serious deficiencies:

- Electrostatic potentials are not accurately reproduced
- Charges on atoms are not allowed to change when molecular geometry changes
- Only pairwise interactions are considered (many body polarization effects are neglected)

To improve overall accuracy:

- Allow non-atom-centered charges
- Add point dipoles, quadrupoles etc

To allow charges to change and account for many-body interactions:

- fluctuating charge model

$$E_{el} = \sum_A \underset{\substack{\uparrow \\ \text{el. potential}}}{\phi_A} Q_A + \sum_A \underset{\substack{\uparrow \\ \text{electronegativity}}}{\chi_A} Q_A + \frac{1}{2} \underset{\substack{\uparrow \\ \text{chemical hardness}}}{\eta_{AB}} Q_A Q_B$$

- polarizable models

$$\mu_{ind} = \alpha F \quad \leftarrow \text{polarizability, electric field}$$

$$E_{el}^{pol} = \frac{1}{2} \mu_{ind} \mathbf{F} = \frac{1}{2} \alpha \mathbf{F}^2$$

# QM/MM models

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM}$$

QM/MM models differ in the way  $H_{QM/MM}$  term is represented

1. Mechanics embedding: QM-MM interactions = electrostatic + vdw, described at MM level (this is analogous to a simple ONIOM method where low level = MM)
  - need partial charges and vdw parameters for the QM atoms
  - electronic density of the QM region is not affected by the MM environment
2. Electronic embedding: electronic charges on MM atoms affect the QM Hamiltonian:

$$\langle \Psi | \hat{H}_{QM/MM} | \Psi \rangle = \Delta E_{vdw} + \sum_k^{QM} \sum_m^{MM} \frac{Z_k q_m}{r_{km}} + \left\langle \Psi \left| \sum_{i \text{ elec}}^{QM} \sum_m^{MM} \frac{q_m}{r_{im}} \right| \Psi \right\rangle$$

QM region is aware of changes in MM, but MM does not know what happens to QM

3. Polarizable embedding: QM atoms polarize MM region
  - need polarizable force field; computationally more expensive
  - feedback between electrostatic fields in QM and MM regions