

Molecular Biology at the quantum level:  
Can modern density functional theory (DFT) forge the path?

**Classical methods**

For many purposes, the best methods are classical force field models:

- We can study large-scale dynamics (even millions of atoms)
- At biologically-relevant timescales

How do they work ?

We try to represent the energy and forces of a collection of atoms using algebraic expressions. Those algebraic expressions are based on macroscopic physics. Example : we treat bonds and angles as classical harmonic oscillators that obey Hooke's law.

$$E_{\text{ff}} = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihedrals}} + E_{\text{non-bonded}}$$

$E_{\text{bonds}}$  (bond between two atoms) and  $E_{\text{angles}}$  (angle formed by three atoms) represent harmonic oscillators.

$E_{\text{dihedrals}}$  are treated as shallow periodic potentials.

$E_{\text{non-bonded}}$  interactions between non-bound atoms (Lennard-Jones potential).

Some force fields have additional terms.

Disadvantages :

- We approximate microscopic phenomena using macroscopic theory
- The only physics present in the simulation is what has been explicitly included. We cannot gain any true atomic-level insight.
- The approach is very simple but it lacks chemical accuracy → not useful for, say, drug design

Some improvements to the force field method :

- Originally, force fields included the partial charge of an atom as a fitting parameter. The charge was kept constant during the whole simulation.
- Next generation : charges can rearrange during the simulation (polarizable force fields)
- Newest : include cross-terms that account for how changes in one internal coordinate can affect other energy terms.

All these improvements can improve accuracy, but they cannot correct for the lack of an explicit quantum mechanical treatment.

**Incorporating quantum mechanics**

We have to solve the time-independent Schrodinger equation in order to include quantum mechanics in the calculations. The problem is that the unknown function  $\psi$  depends on the coordinates of all the electrons within the system, so it is impossible to solve this for all but the most trivial systems.

Example : a water molecule has 10 electrons → its wave function is a function of 30 variables (10 electron positions in three dimensions) → impossible to solve analytically and it requires too much storage to solve numerically.

How do we solve this ? There are two approaches : wave function theories and density functional theory.

Particular properties of biological systems :

- They are large
- Their structure is dominated by Van der Waals weak interactions

In the review, they describe methods that are well-suited for biological systems.

**Wave function theory**

We assume that the total electron wave function can be written in as a product of single-electron states (orbitals). An example of this approach is Hartree-Fock (HF).

Hartree-Fock :

- Relatively fast
- Based on sound quantum mechanics
- BUT : the approximations it uses miss some important physics : correlation
- Anyway, good starting point for corrections that incorporate electron correlation into the total wave function.

Post-HF methods : they use HF results and incorporate electron correlation explicitly.

One of the best things of wave function methods is that they are organised in levels of theory.

- Full configuration interaction (CI) : we start with orbitals found by HF. We produce all possible excitations to a given set of orbitals (combinatorial problem). Not practical in all but the smallest systems.
- CCSD (coupled-cluster theory) : best thing after CI. It mimicks CI but only uses small numbers of electron excitations. Useful only for relatively small systems.
- Moller-Plesset perturbation theory : MP2, MP3, MP4, etc.

### Density functional theory (DFT)

Much more amenable to calculation of large systems. The quantum-mechanical energy of a set of atoms can be written uniquely as a functional of the electron charge density. The charge density that minimizes this functional is the ground-state charge density for the system, and all measurable properties of the system can be written in terms of this optimal charge density → this avoids the dimensionality problem. Density in real space is a function of only 3 variables. Energy is expressed as a density functional :

$$E_{\text{dft}}[n(r)] = E_{\text{k}}[n(r)] + E_{\text{n-e}}[n(r)] + E_{\text{e-e}}[n(r)] + E_{\text{xc}}[n(r)] + E_{\text{n-n}}$$

$E_{\text{k}}$  is the total kinetic energy of the system.

$E_{\text{n-e}}$  is the nucleus-electron, effective one-body term.

$E_{\text{e-e}}$  is the Hartree term giving the average electron-electron interaction.

$E_{\text{n-n}}$  nucleus-nucleus repulsion term, can be treated as a simple additive constant since it is uniquely determined by the positions of the nuclei.

$E_{\text{xc}}$  the sole unknown object in the equation. It must be approximated in some way. This approximation is what determines the applicability of the DFT method to a particular system.

How do we approximate this ?

Local density approximation (LDA) :

We assume that the exchange-correlation energy is a local functional of the density : one that depends on  $n(r)$  in a point-wise fashion. This approximation works when the density is slowly varying. LDA accounts for the fact that each electron experiences a highly dynamic environment rather than a mean field of the other electrons.

### Hybrid functionals

Very accurate functionals for covalently-bound molecules. Early success of those hybrid functionals led some people to believe that they could describe Van der Waals interactions. But, hybrid functionals are local or semi-local, while Van der Waals interactions are non-local. That is why hybrid functions cannot reliably describe VDW interactions.

### Van der Waals interactions in DFT

Historically, DFT has performed poorly when applied to systems with VDW interactions, but recent developments have overcome this issue.

VDW interactions occur between physically separated regions of charge, generally with little overlap of their density functions. They occur when electron motions in one atom correlate with electron motions in a nearby atom, setting up transient but interacting multipoles within each. Correlation between electrons lowers their energy relative to uncorrelated electrons, so the VDW force is always attractive.

Some approaches to include VDW interactions in DFT :

#### DFT-D

If a point of interest is located at distance  $r$ , the pairwise dispersion energy can be expanded in powers of  $1/r$ . For large distances  $r$ , the dipole term dominates and dispersion interactions go as  $1/r^6$ .

This method works by adding to the total energy a pairwise atomic correction, that includes a damping function (that ranges from 0 to 1). Since  $1/r^6$  has an asymptotic form, in small distances the energy would be overestimated.

#### DFT+vdW

To improve DFT-D, they calculate C6 on-the-fly from the charge density. Writing the c6 coefficients as density functionals allows for the polarizability of atoms to be a dynamic, environment-dependent quantity.

#### VdW-DF

An alternative is to express the total energy of a system directly as a non-local functional of the density.

### **QM/MM Methods for Biomolecular Systems**

Quantum mechanics/Molecular mechanics approaches :

- We use QM to describe chemical reactions and other electronic processes (charge transfer, electronic excitation, etc). But we can only use it for small systems (up to a few hundred atoms).
- We use force-field based molecular mechanics to treat bigger systems (several 100 000 atoms) that need to be studied for a longer time (tens of nanoseconds).

If we combine both methods, we can study the chemically active region with QM and the surroundings with MM. These combinations are called hybrid QM/MM methods. They enable us to model reactive biomolecular systems at a reasonable computational effort.

#### Introduction

The QM/MM approach is used nowadays not only for biomolecular systems, but also with inorganic/organometallic and solid-state systems. Topics :

- QM/MM partitioning
- Choice of suitable QM and MM methods
- Treatment of the QM/MM interactions
- Treatment of the QM/MM boundary regions
- QM/MM geometry optimization
- Molecular dynamics
- Free-energy simulation techniques

#### **QM-MM partitioning**

The entire system is partitioned into the inner region (I) that is treated quantum-mechanically, and the outer region (O) that is described by a force field. The total energy of the system cannot be written as the sum of energies of the subsystems, because there are strong interactions between the QM and the MM regions. Thus, we need to consider coupling terms, and take precautions at the boundary region, especially if the boundary cuts through covalent bonds.

Boundary region : the region where the standard QM and MM procedures are modified or augmented in any way.

Fixed QM-MM partitioning : the boundary is defined once-and-for-all at the beginning. Some QM/MM schemes exist where the boundaries can change during the course of the simulation (adaptive partitioning, « hot spot » methods). Those are useful to model processes with shifting active regions.

#### **Choice of QM/MM methods**

The QM/MM hybrid can accommodate almost any combination of QM and MM methods. The QM code must be able to perform the self-consistent field (SCF) treatment in the presence of the external point-charge field that represents the MM charge model in the case of electronic or polarized embedding.

QM methods :

- Many hybrids use DFT as the QM method (good computational-effort/accuracy ratio)
- Post-HF methods : CCSD (coupled cluster theory), Moller-Plesset perturbation theory, etc.
- Empirical valence-bond method (EVB)

Popular MM methods :

- AMBER
- CHARMM
- GROMOS
- OPLS-AA

### The MM energy-expression

What does the classical force field of the MM region contain ?

- Bonded terms : bond stretching, angle bending, torsions, out-of-plane deformations/improper torsions.
- Van der Waals interactions (Lennard-Jones type)
- Coulomb interaction terms between point charges

### The QM-MM Energy Expression

Subtractive QM/MM schemes require :

- 1) MM calculation on the entire system  $\rightarrow E_{MM}(S)$
- 2) QM calculation on the inner subsystem  $\rightarrow E_{QM}(I+L)$
- 3) MM calculation on the inner subsystem  $\rightarrow E_{MM}(I+L)$

The total energy,  $E_{QM/MM}(S)$ , is calculated as :  $E_{QM/MM}(S) = E_{MM}(S) + E_{QM}(I+L) - E_{MM}(I+L)$

Where L are the link atoms.

Conceptually, the subtractive model is like a MM approach in which a certain region of space has been cut out and is treated at the QM level.

Advantages :

- Very simple scheme, straightforward implementation
- No QM-MM coupling terms are needed.
- The standard QM and MM procedures can be used without any modifications

Drawbacks :

- Requires a complete set of MM parameters for the inner subsystem  $\rightarrow$  can be difficult to obtain
- The coupling between subsystems is handled at the MM level. This is problematic for the electrostatic interactions, that is typically represented by the Coulomb interaction between fixed atomic charges in the QM and MM.

Additive QM/MM scheme :

$$E_{QM/MM}(S) = E_{MM}(O) + E_{QM}(I+L) + E_{QM-MM}(I,O)$$

Where  $E_{QM-MM}(I,O)$  collects the interaction terms between the two subsystems.

This way, the MM calculation is done in the outer subsystem only. Most QM/MM schemes used nowadays are additive.

### Individual contributions in $E_{QM/MM}$ :

1. The electrostatic QM-MM interaction : can be handled in different ways. It is the electrostatic coupling between the QM charge density and the charge model used in the MM region.

- **Mechanical embedding** : the charge model of the MM method (typically rigid atomic point charges) is simply applied to the QM region as well. Conceptually easy, computationally efficient. BUT :

- The charges in the outer region do not interact with the QM density → the QM density is not influenced or polarized by the electrostatic environment.
- When the charge distribution of the QM region changes, for example, during a reaction, it would be sensible to update the charges. However, this would cause discontinuities in the potential energy surface.
- It is not easy to derive appropriate MM point charges for the inner region.
- These MM charges do not necessarily reproduce the true charge distribution in the inner region. Individual MM charges do not need to be physically meaningful, as long as the force field provides an overall balanced description.

The major disadvantages of mechanical embedding can be avoided by performing the QM calculation in presence of the MM charge model → electrostatic embedding. For example, we can include the MM point charges as one-electron terms in the QM Hamiltonian.

- **Electrostatic embedding** : the electronic structure of the inner region can adapt to changes in the charge distribution of the environment and is automatically polarized by it. The QM-MM electrostatic interaction is treated at the QM level → more advanced and more accurate description than a mechanical embedding scheme. We have to be careful at the QM-MM boundary not to cause overpolarization, especially when the boundary runs through a covalent bond. Most popular embedding scheme. As electrostatic embedding accounts for the interaction of the polarizable QM density with rigid MM charges, the next logical step is to introduce a flexible MM charge model that is polarized by the QM charge distribution → polarized embedding.
- **Polarized embedding** : many types.
  - The polarizable charge model in the MM region is polarized by the QM electric field but does not itself act back on the QM density.
  - Mutual polarization : the polarizable MM model is included in the QM Hamiltonian and therefore QM can polarize the MM region and MM can polarize QM.

## 2. Other nonbonded and bonded QM-MM interactions

In addition to the electrostatic interactions, there are also Van der Waals and bonded contributions to the QM-MM coupling term. Those are much easier to handle, they are handled purely at the MM level.

- **Van der Waals** : typically described as a Lennard-Jones potential. We need suitable parameters for the QM atoms in the inner regions. There is a discussion on how should we estimate the values for those parameters.
  - Additive scheme : only pairs consisting of one atom from the inner and one atom from the outer subsystem are considered.
  - Subtractive scheme : yields the exact same value as the Van der Waals additive term, because QM-QM Van der Waals pairs are canceled out.
- **Bonded terms** : bond stretching, angle bending, torsional, etc. The same reservations against using standard MM parameters to describe QM-MM interactions apply to the bonded interactions. Solution : the standard MM parameter set is retained and is complemented as necessary with additional bonded terms not covered by the default assignment rules of the force field.

## Covalent bonds that cross the QM-MM boundary

### Overview of boundary schemes

How to treat the boundary between QM and MM?