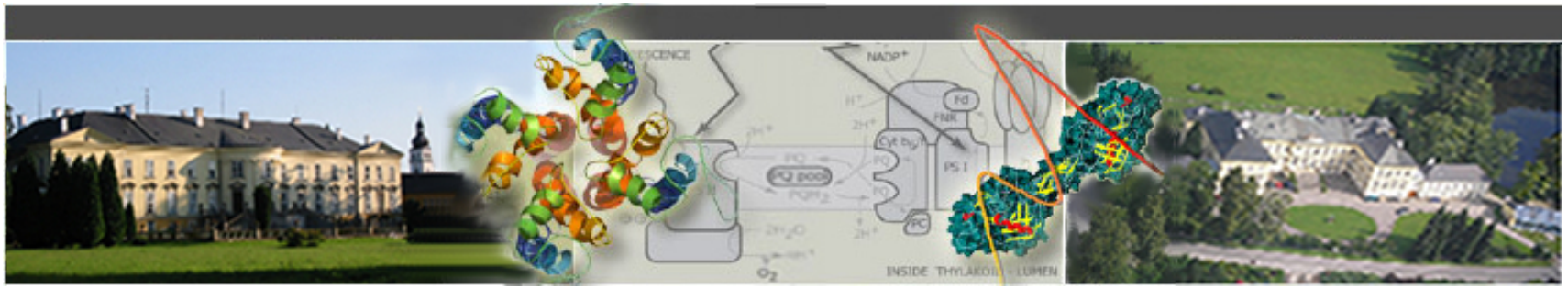
Biological Applications of QM/MM Calculations

David Řeha



Institute of Nanobiology and Structural Biology, ASCR,
Zámek 136, 373 33 Nové Hrady, Czech Republic

Molecular Modeling

Molecular Mechanics

Geometry optimizations

Molecular Dynamics Simulations

Monte Carlo Simulations

Molecular docking

Quantum Mechanics

Geometry optimizations

Localization of TS

Calculation of the spectra (UV-VIS,IR,NMR)

Calculation of electronic properties (charge derivation)

Calculation of the binding energy

Need to evaluate energy for given geometry

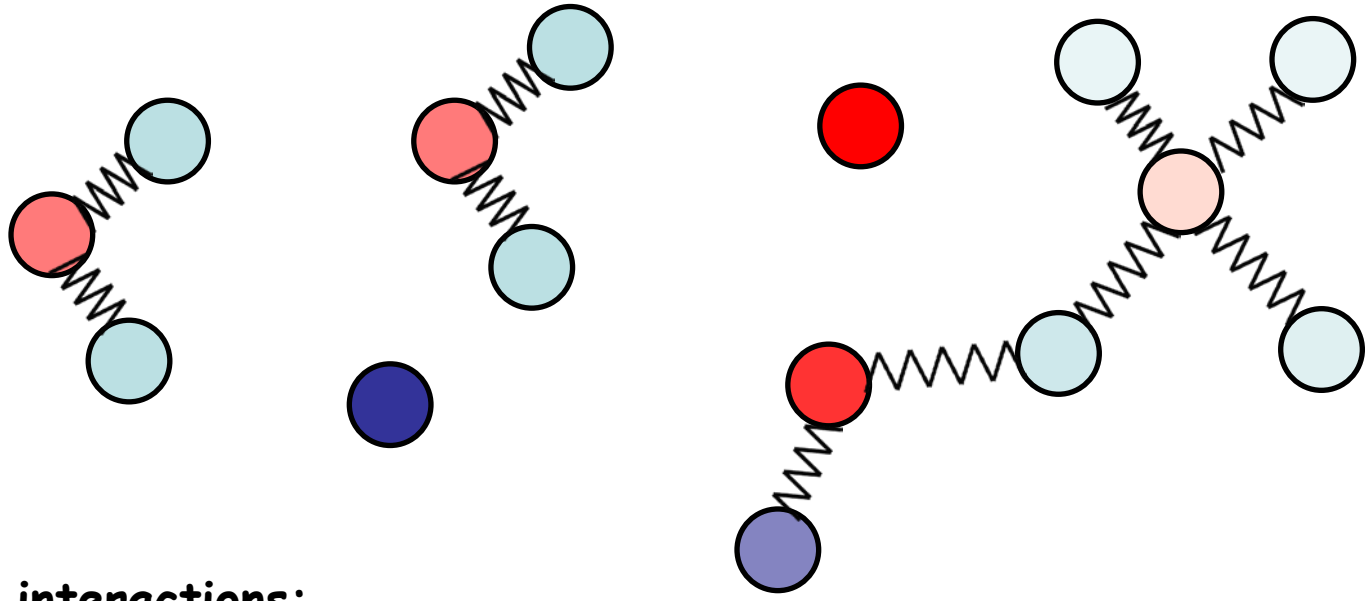
Force fields

Molecular Mechanics

Mechanical model – charged spheres connected by springs

Charged spheres – represent atoms

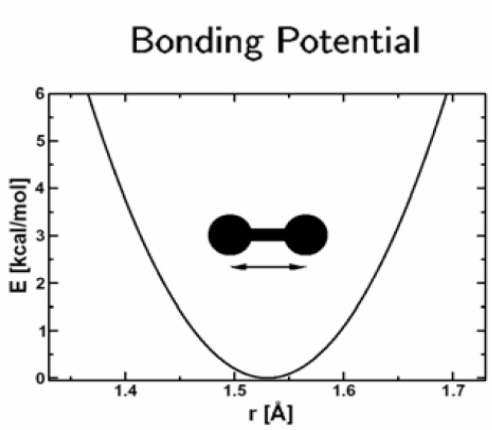
Harmonic springs – represent chemical covalent bonds

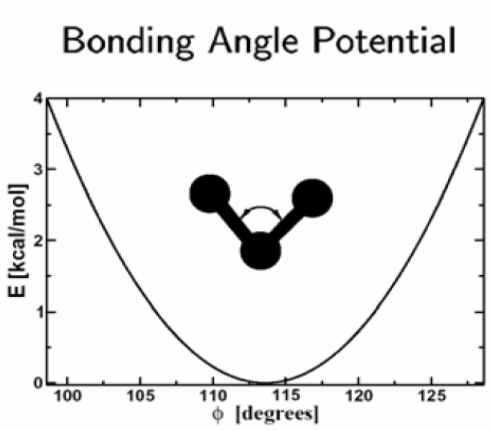


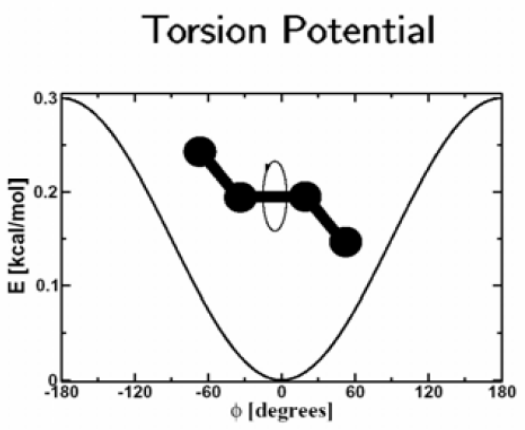
Two types of interactions:

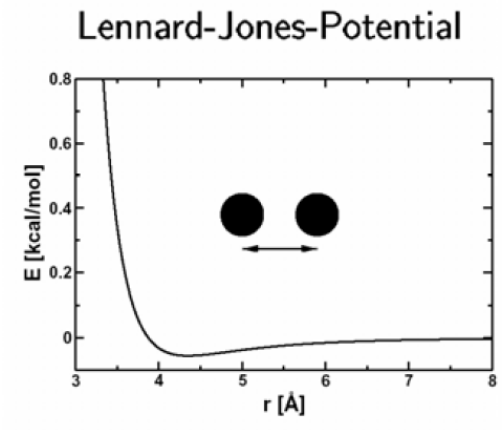
Bonding: all bonds, all bond angles, all bond dihedrals

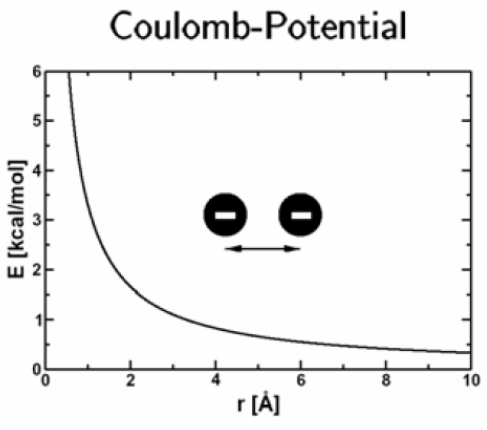
Nonbonding: electrostatic and Van der Waals interactions between all atom pairs











$$E = \sum_{b}^{nbonds} k_b \left(r_b - r_{ob}\right)^2 + \sum_{\theta}^{nangles} k_\theta \left(\theta - \theta_o\right)^2 + \sum_{\phi}^{ntorsions} k_\phi \left(1 - \cos\left(n\phi - \delta\right)\right)$$

$$+ \sum_i \sum_j \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right\}$$

Molecular Mechanics

Pros: relatively fast calculations of energy -> atomistic simulations of relatively large biological systems (up to 100000 atoms)

Geometry optimization

Localization of local geometrical minima where $\nabla E(X) = 0$

Molecular Dynamics Simulations

Solving Newton's equations of motion for given time period

$$F(X) = -\nabla E(X) = M \cdot \dot{V}(t) \qquad V(t) = \dot{X}\,(t)$$

Time evaluation of the system

Monte Carlo Simulations

Search of the PES (E(X)) using biased random sampling

No time dependence

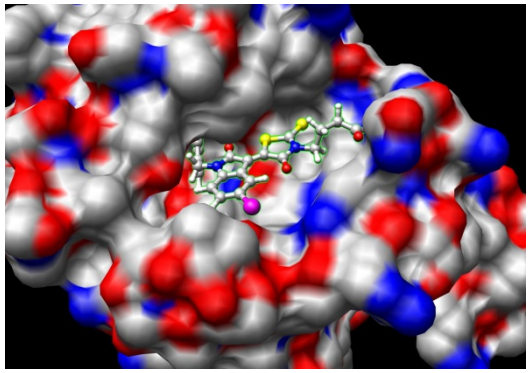
Molecular Mechanics

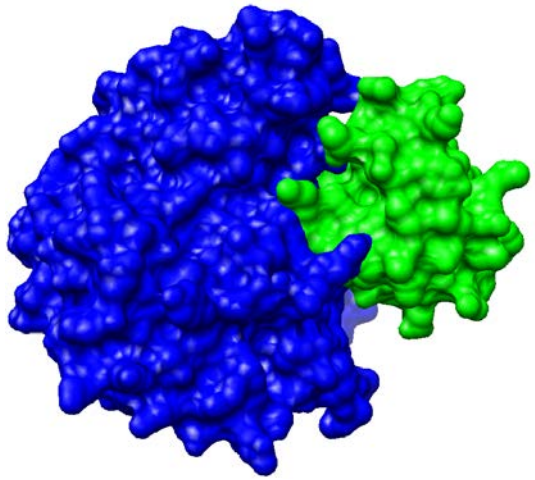
Molecular Docking

Predicts preferred position and orientation of the ligand within receptor (protein)

Selection of the best fitting ligand (from the database of ~100000 ligands) to the particular receptor (protein)

Predicts preferred position and orientation of two proteins





Scoring function -> based on the interaction energy

Molecular Mechanics

Cons:

Separation between bonding and nonbonding interactions

⇩

Covalent bonds in the system need to be defined before simulation

⇩

No forming or breaking of covalent bonds

⇩

No simulation of chemical reactions

Fixed atomic charges defined in the forcefiled

⇩

atomic charges don't respond to the change of environment -> **no polarization**

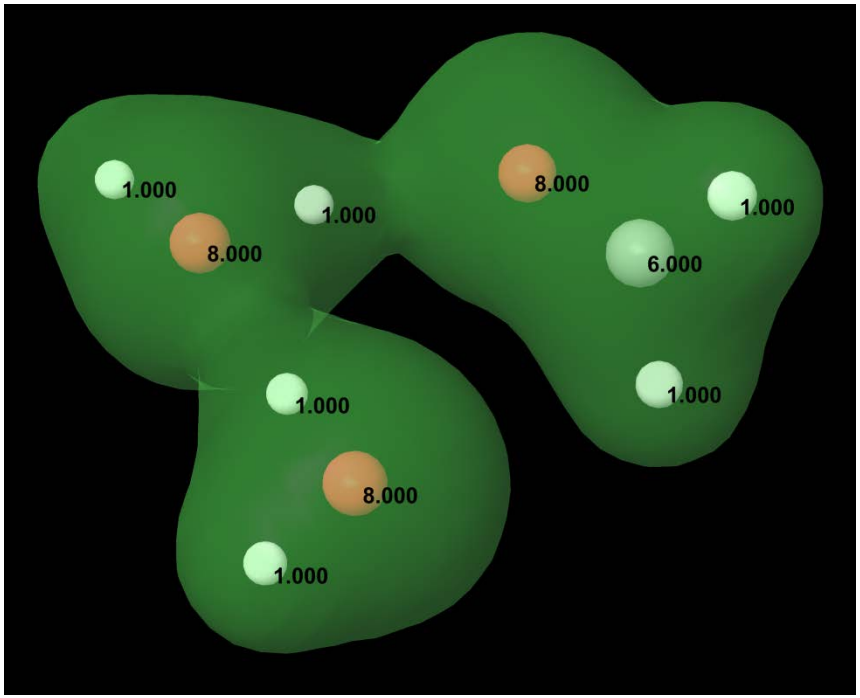
No empirical parameters for nonstandard systems
(heavy metals, ions, transition states, exited states..)

Quantum Mechanics

Electronic level

Model: Positively charged atomic nuclei (treated classically) and negatively charges electrons (treated as quantum wave functions)

Electrons described by wavefunctions Ψ , physical meaning is the density of electronic cloud



Total energy:
solving the Schrödinger equation

$$\hat{H}\Psi = E \cdot \Psi$$

Very complicated

⇩

need of simplifying approximations

One electron:

$$\rho(x,y,z) = \left|\Psi(x,y,z)\right|^2$$

N electrons:

$$\rho(x,y,z) = N\int\left|\Psi(x,y,z,x_2,y_2,z_2,\ldots,x_n,y_n,z_n)\right|^2 dx_2dy_2dz_2\ldots.dx_ndy_ndz_n$$

Quantum mechanics

Approximations

1. Born-Oppenheimer approximation

   System wavefunction separated for nuclei and electrons – nuclei treated classically, electrons treated QM

2. Expressing many-electron wavefunction as a product of one-electron wavefuctions

$$\Psi\left(x_1, y_1, z_1, x_2, y_2, z_2, ..., x_n, y_n, z_n\right) = \psi_1\left(x, y, z\right) \cdot \psi_2\left(x, y, z\right) \cdot ... \cdot \psi_n\left(x, y, z\right)$$

One-electron form of Schrödinger equation $\hat{F}\psi = \varepsilon \cdot \psi$

3. Expressing one-electron wavefuction (molecular orbital) as vector in the basis set of atomic orbitals *MOLCAO*

$$\psi_k = \sum_{i}^{nbasiset} C_{k,i}\chi_i$$

then operators become matrixes -> Schrödinger equation (differential) is transferred to algebraic form

$$\overline{F} \cdot \vec{C} = \varepsilon \cdot \vec{C}$$

Quantum mechanics

Levels of theory

Semiempirical – uses empirical parameters to compensate for neglected terms (Extended Huckel, NDO,INDO,AM1,PM3)

Ab inito HF level – no empirical parameters, electron-electron interaction is treated as one-electron interaction with averaged electronic clouds of remaining electrons -> does not include correlation energy

Density Functional – DFT based on electron density rather then wavefunction, minimization of density functional, however real density functional unknown -> used approximate functionals (like B3LYP)
Correlation energy partially included, however not pure *ab initio*

Ab inito correlated – based on the HF level, goes beyond HF in order to treat correlation energy. The most computationally demanding (CI, CC, MP2)

Quantum mechanics

Basis set

The wavefunction is described as a vector in the basis set – atomic orbitals

The larger basis set more accurate wavefunction (exact one -> infinity basis set)

The atomic orbitals used as basis set – Slater functions (orbitals)

$$\chi(r,\theta,\varphi) = Nr^{n-1}Y_l^m(\theta,\varphi)\cdot e^{-\zeta r}$$

In order to simplified calculations, the Gaussian orbital functions used instead

$$\chi(x,y,z) = Nx^n y^l z^m e^{-\alpha\left((x-x_0)^2+(y-y_0)^2+(z-z_0)^2\right)}$$

L=m+n+l equivalent to angular momentum quantum number (L=0 -> s-orbital; L=1 -> p-orbital; L=2 -> d-orbital). [x_0,y_0,z_0] –center of the orbital function; α - exponent

Quantum mechanics

The most common software:

Gaussian, Games, Turbomol, Molpro, Jaguar

As input, the software requires just:

- coordinates of atoms and their proton numbers
- the total number of electrons (usually as a net charge)
- total spin multiplicity
- level of the theory (ex. HF,MP2,…)
- basis set (ex. 6-31G*, aug-cc-pVDZ…).

⇩

No need of empirical parameters

It can perform single point energy calculations, geometrical minimizations, frequency calculations, calculation of electronic properties(dipole moment, polarizabilities, spectral parameters (UV, IR, NMR)), (MD simulations)

Quantum mechanics

Pros:

Calculate **exact electron distribution** (density) around atomic nuclei

There is no recognition of covalent bonds – covalent bonds are treated naturally (there is no strict border what is still covalent bond and what is not)

Program doesn't need information about the covalent bonds -> they form or break naturally

Electron density naturally respond to the environment -> **polarization is included**

It can capture **electron transfer** between molecules

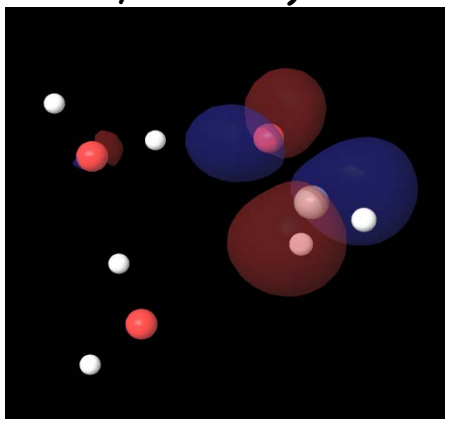
It is *ab inito* calculation, no need of empirical parameters

It can treat nonstandard systems (ions, metals)

Quantum mechanics - Applications

Calculation of the molecular orbitals

Can help predict part of the molecule where chemical reaction happens
Can calculate ionization potential and elect. affinity (HOMO, LUMO)



Calculation of the electron density

Derivation of the **atomic charges** by fitting them to electrostatic potential (ESP)

Calculation of dipole moment

Calculation of the interaction energy

Can calculate interaction energy between molecules **A** and **B**

$\Delta E_{int} = E_{AB} - E_{A} - E_{B}$

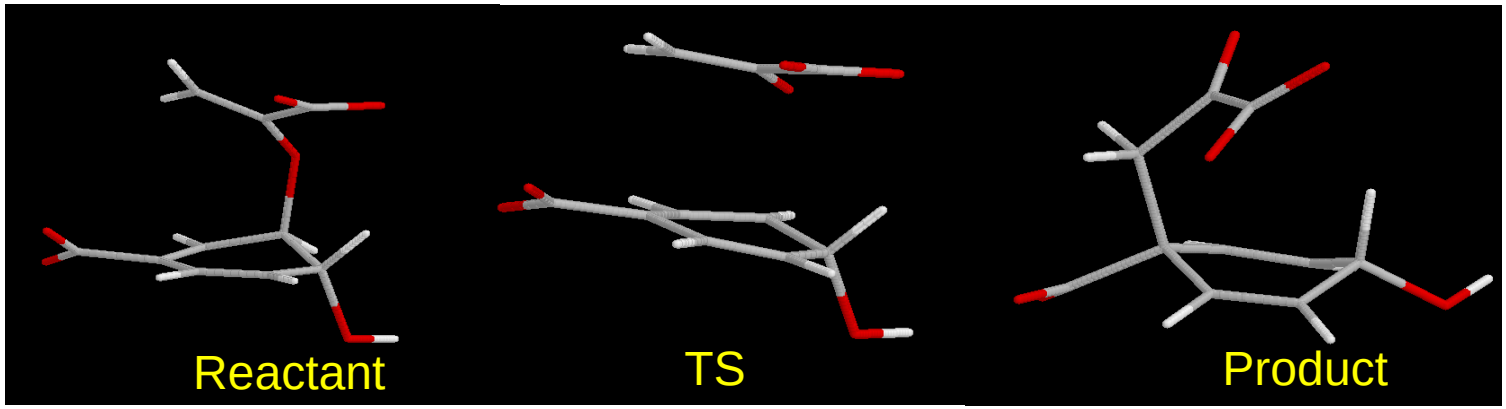
Quantum mechanics - Applications

Geometry optimization

Localization of local geometrical minima where $\nabla E(X) = 0$

Localization of the transition state

Calculation of the reaction barrier ΔE_A -> kinetics of the chemical reaction



ΔE_A = E(TS) - E(Reactant)

Calculation of the vibrational modes->IR spectra

Calculation of the other spectra (NMR, UV-VIS)

Quantum mechanics

Cons:

Very complicated calculation – computationally demanding

QM calculation can take 10^{10} times more time then MM

Unfavorable scaling – the computational time of some advance QM calculations is scaled N^6 with the size of the system N

With current computers it is possible to treat max 100-150 atoms

MD by semiempirical QM methods max up to 100 atoms

Impractical for biological systems

Hybrid QM/MM

Very practical for biological systems

System split in two regions:

QM region: Small part of the system (max 100 atoms), region of interest (like substrate in the active site of protein)

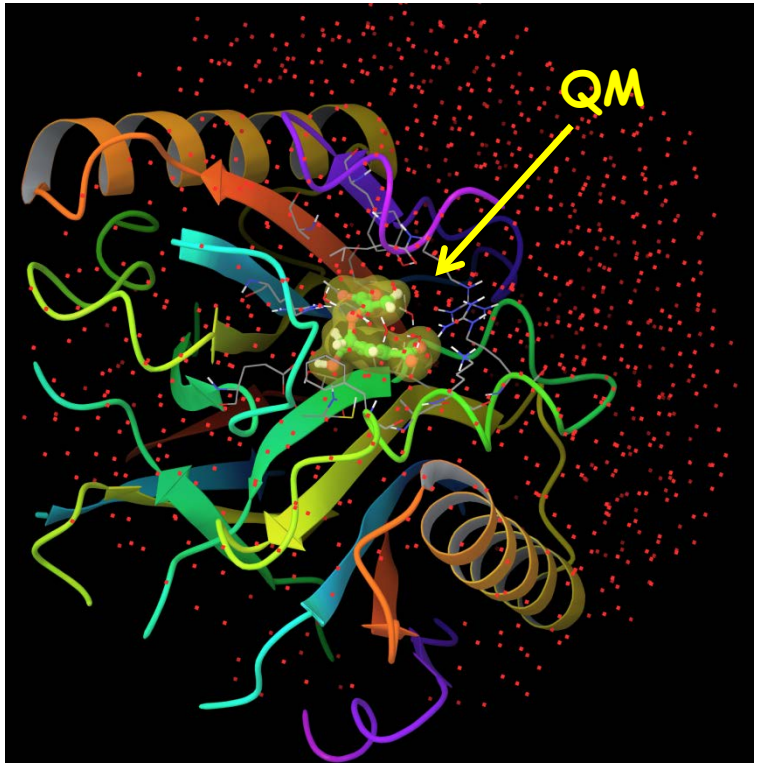
⇩

Treated by quantum mechanics

MM region: The rest of the system (rest of the protein), explicit solvent molecules

⇩

Treated by classical molecular mechanics



Advantages of both approaches, it can calculate chemical reaction in QM region
But it is fast enough to calculate large biological systems

Hybrid QM/MM

Additive approach

Total energy as the sum of the contributions:

$$E = E_{QM}(QS) + E_{MM}(MS) + E_{QM/MM}(QS/MS)$$

QS – QM region, MS – MM region

$$E_{QM/MM}(QS/MS) = E_{elst}(QS/MS) + E_{vdw}(QS/MS)$$

E_{elst}(QS/MS): interactions between point atomic charges of MM forcefield of MS with electronic density (wavefunction) of QS

E_{vdw}(QS/MS): vdw interactions between MS and QS calculated by MM forecefield

Hybrid QM/MM

Types of the embedding

Mechanical embedding – QS is calculated by QM method as isolated system
E_{elst}(QS/MS): is calculated by MM as the interactions between atomic charges (from MM forcefield) of QS and MS.

Electrostatic embedding – QS is calculated by QM in the presence of the forecefield atomic charges of MS. Wavefunction of QS is polarized by MS. E_{elst}(QS/MS): interactions between point atomic charges of MM forcefield of MS with electronic density (wavefunction) of QS

Polarizable embedding –similar like ***electrostatic embedding***, but MM forecefield atomic charges of MS are also polarized by QS using MM polarizable approach

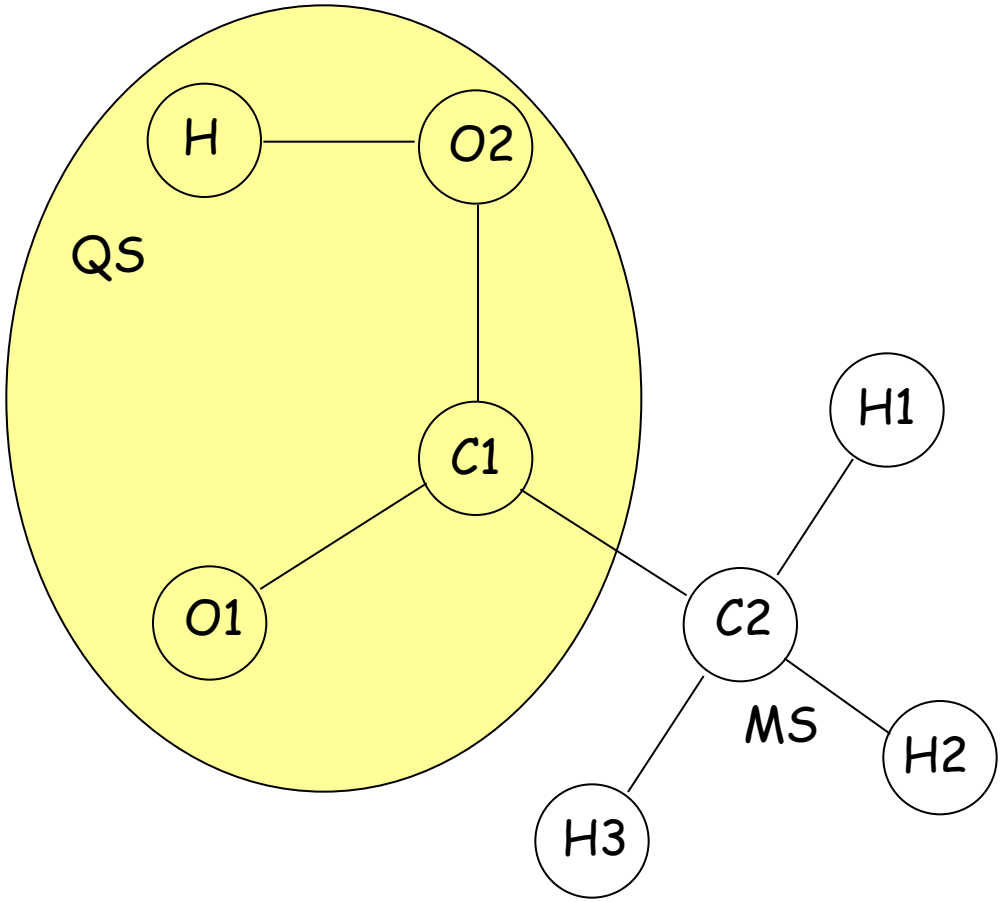
-> Method of induced dipoles

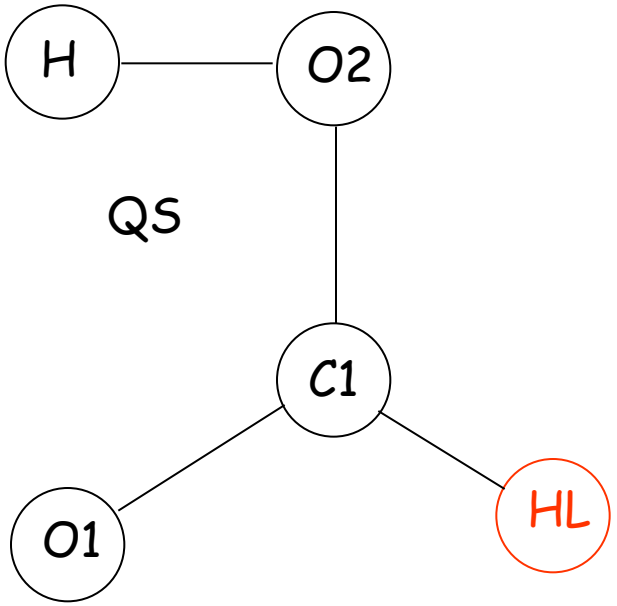
Hybrid QM/MM

Covalent bonds between QS and MS

Link atom approach:

Covalent bond between QS and MS replaced by hydrogen HL for QM calculation





Hybrid QM/MM

Subtractive approach

$$E = E_{QM}(QS) + E_{MM}(QS + MS) - E_{MM}(QS)$$

Advantage – no need to calculate difficult term of $E_{QM/MM}$(QS/MS)
easy when link atom used

Disadvantage – needed MM forcefield parameters for QS, even when they are subtracted, more difficulties with polarized embedding

Hybrid QM/MM

Many applications of QM/MM for biological systems

Detailed calculation of enzymatic reactions, calculations of ΔE, calculations of TS

Calculation of nonstandard ligands: charged, excited states ->better description using explicit polarization

MD simulations of enzymatic reactions (breaking/forming covalent bonds)

Refining the results of the molecular docking

- deriving atomic charges of ligand from QM calculations
- deriving (polarized) atomic charges of ligand from QM/MM calculations
- scoring function based on QM/MM energies

Hybrid QM/MM

The most common software:

QM/MM software – harness that connects QM software and MM software

Chemshell, Qsite - professional

Qomma, Comcum,…. – created by academic reserachers

Harness software requires the coordinates of the system and the definition of QS and MS, also the names of QM and MM software that should be connect Furthermore those QM and MM software requires their standard input files

Hybrid QM/MM - Applications

Calculation of Binding Energies

Similar to calculation of the interaction energies

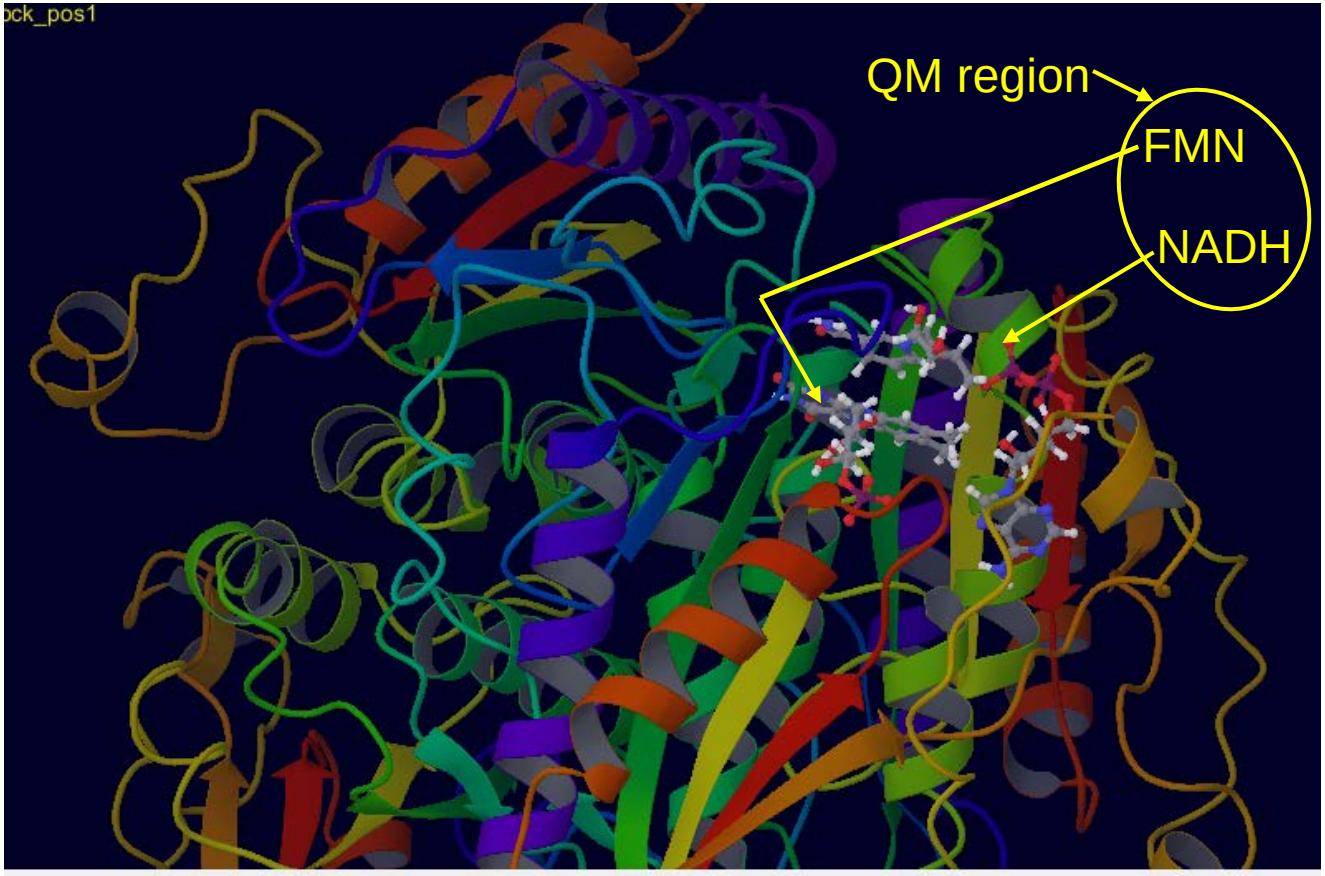
$$\Delta G_{bind} = E_{L+P}^{QM/MM} - E_{L}^{QM} - E_{P}^{QM/MM} + \Delta G_{solv}$$

L: Ligand (fully in QM region)
P: Protein (QM/MM)
ΔG_{solv} solvation energy: can be calculated using implicit solvent

Calculation of Binding Energies: NADH to WrbA

NADH (reduced) + FMN (oxidized) + H+ ->NAD^+(oxidized) + $FMNH_2$(reduced)

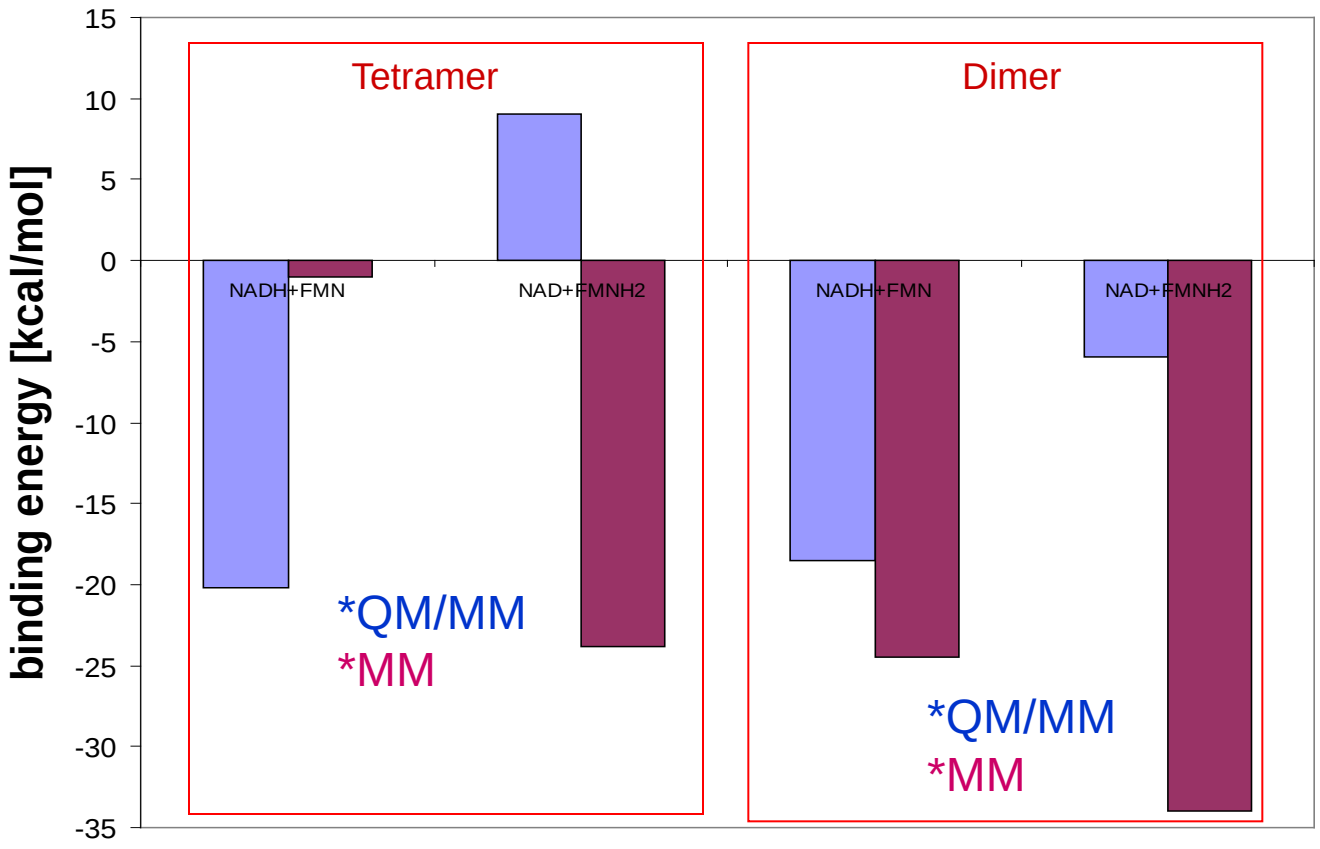
FMN H_2(reduced) + Q(oxidized) ->FMN(oxidized) + QH_2(reduced)



QM: DFT-D (B3LYP)
6-31G*

MM: OPLS 2005
PBS solvent

Calculation of Binding Energies: NADH to WrbA

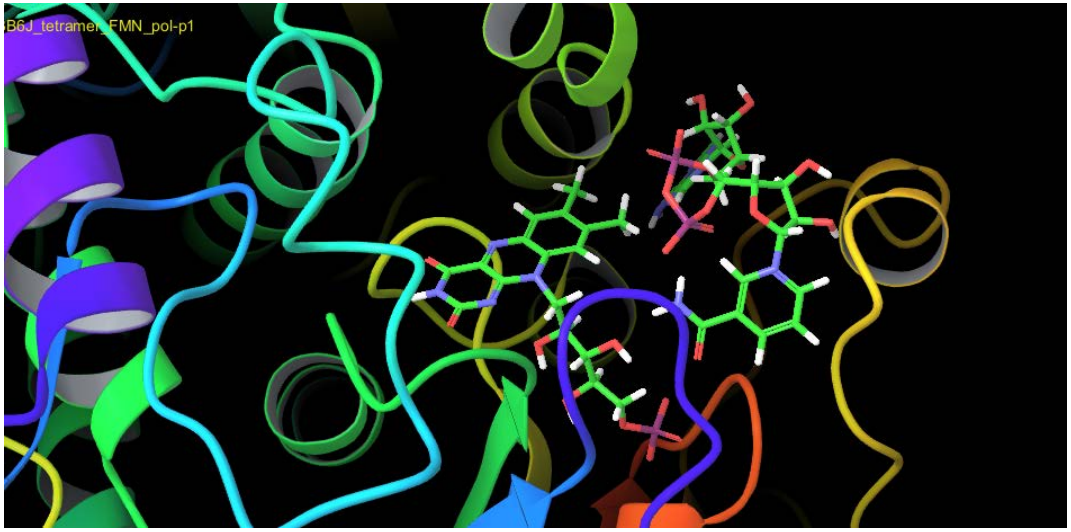


QM/MM Calculation supports the hopping mechanism

QM/MM Calculation supports experimental results. Tetramer is more active then dimer.

Calculation of Binding Energies: NADH to WrbA

Crystal (3B6J) position of NAD



[WrbA+FMN] – NADH 6.9kcal/mol

[WrbA+$FMNH_2$] – NAD^+ $^-$22.9 kcal/mol

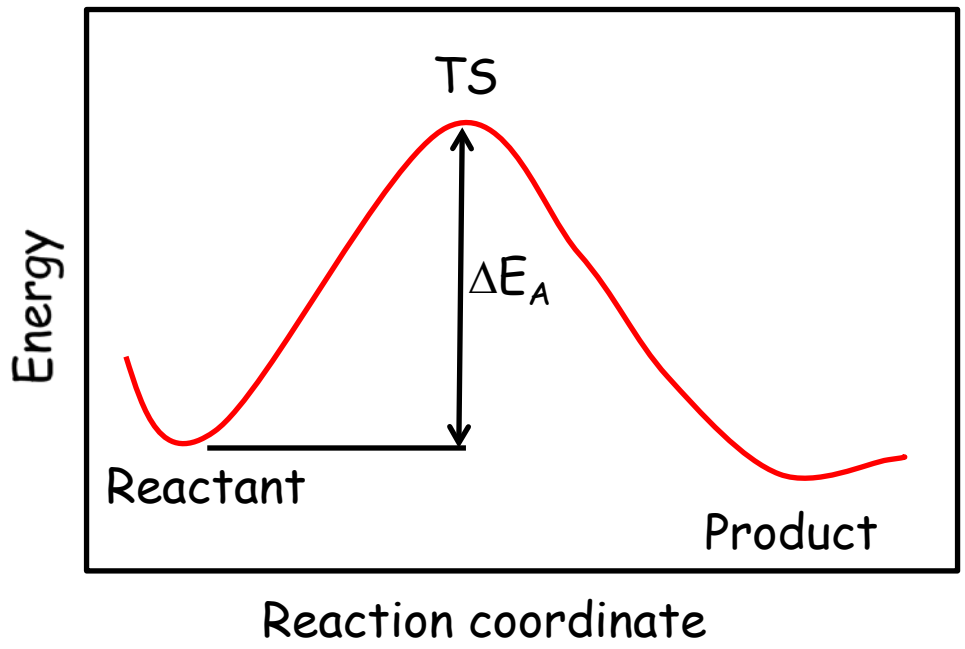
Crystal depicts leaving of NAD

Hybrid QM/MM - Applications

Calculation of kinetics of the enzymatic reaction

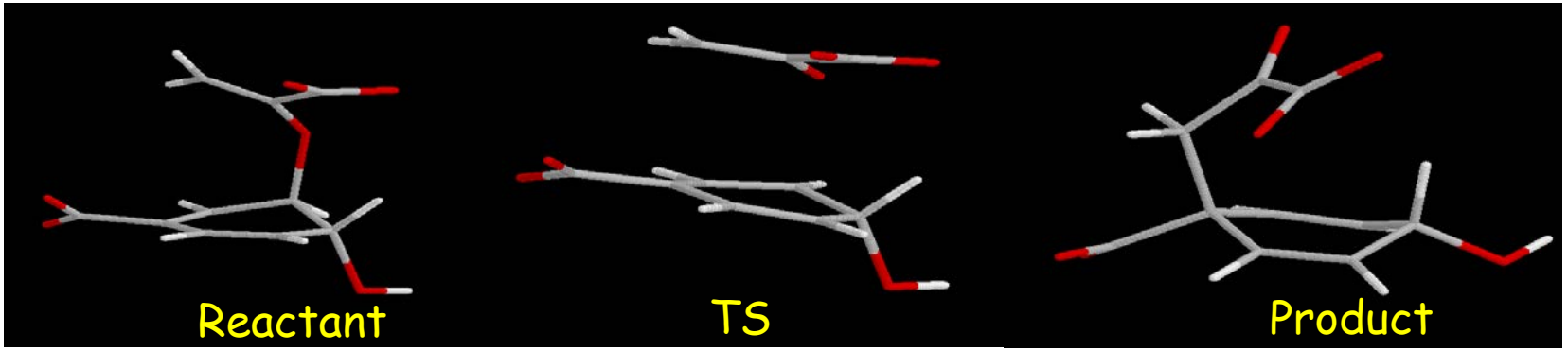
Calculation of the activation energy barrier within enzyme

$$\Delta E_A = E(TS) - E(Reactant)$$

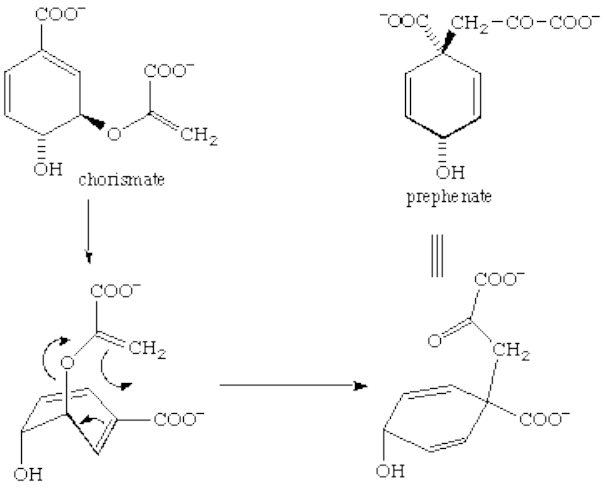


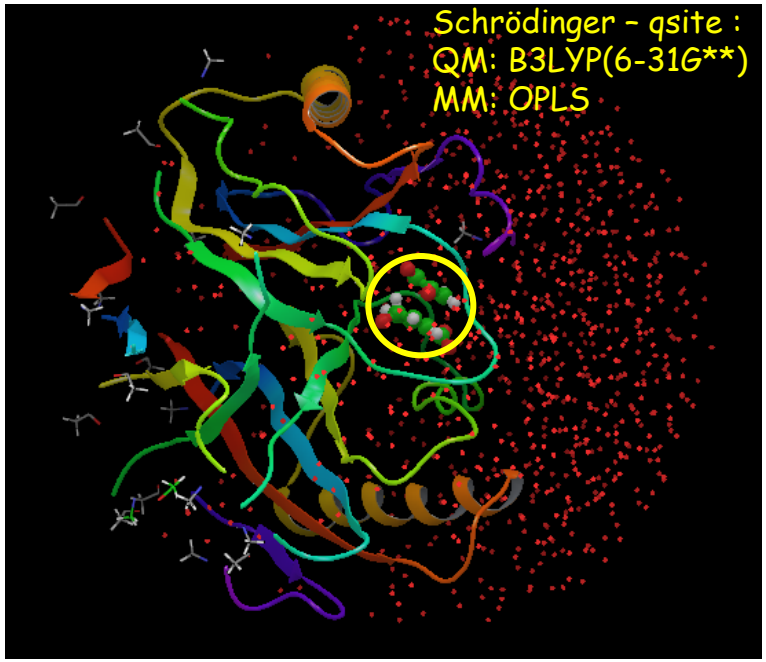
Enzymatic reaction: Chorismate to Prephanate

Example of typical enzymatic reaction
helps us to understand how enzymes works in general

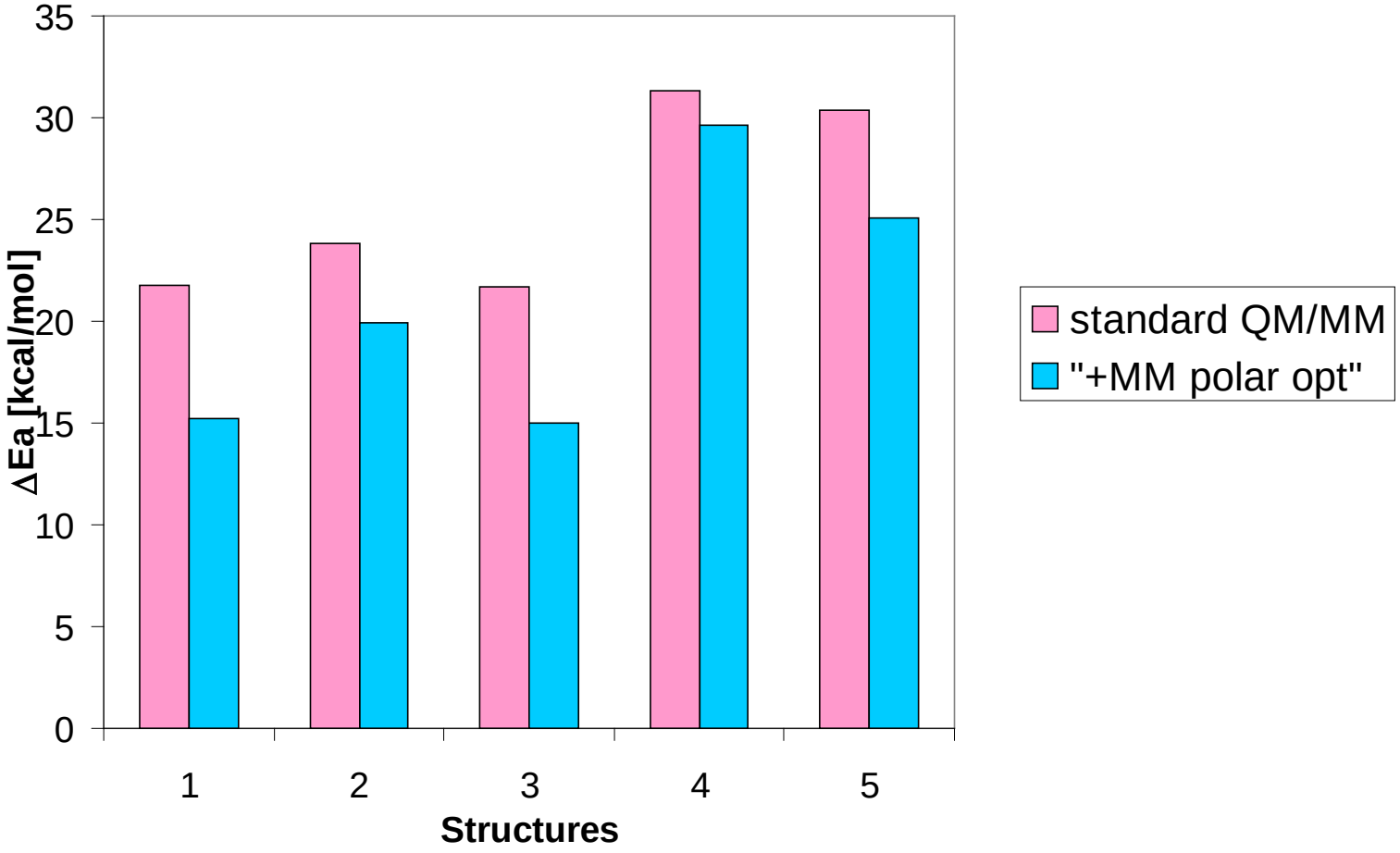


DEa = E(TS) - E(Reactant)





Enzymatic reaction: Chorismate to Prephanate



Inclusion of MM polarization decreases ΔEa by 3-7 kcal/mol

MM polarization important contribution in enzymatic reaction

Hybrid QM/MM - Applications

Molecular docking using QM/MM

Docking with inclusion of the **polarization**

Charges of ligand derived from QM/MM calculations
Ligand: QM part
Receptor(Protein) : MM part

Ligand wavefunction (electron density) polarized by MM charges of receptor (protein)
Ligand atomic charges derived from this electron density

⇩

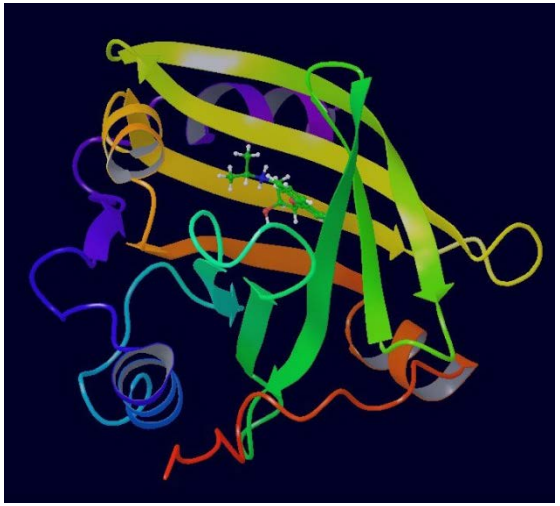
Charges depends on the position of the ligand within receptor

⇩

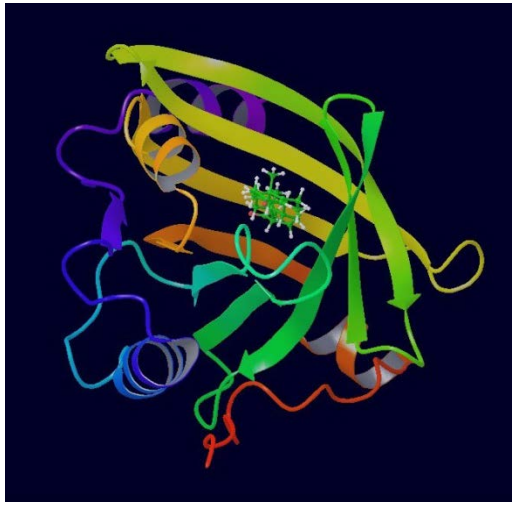
Iterative approach

Receptor charges polarized by method of **induced charges**

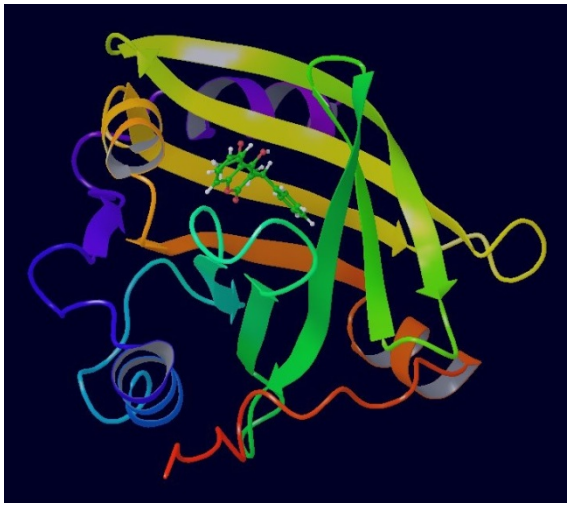
Docking of the drugs (progesterone, propranolol and warfarin) to human a_1-acid glycoprotein



Propranolol



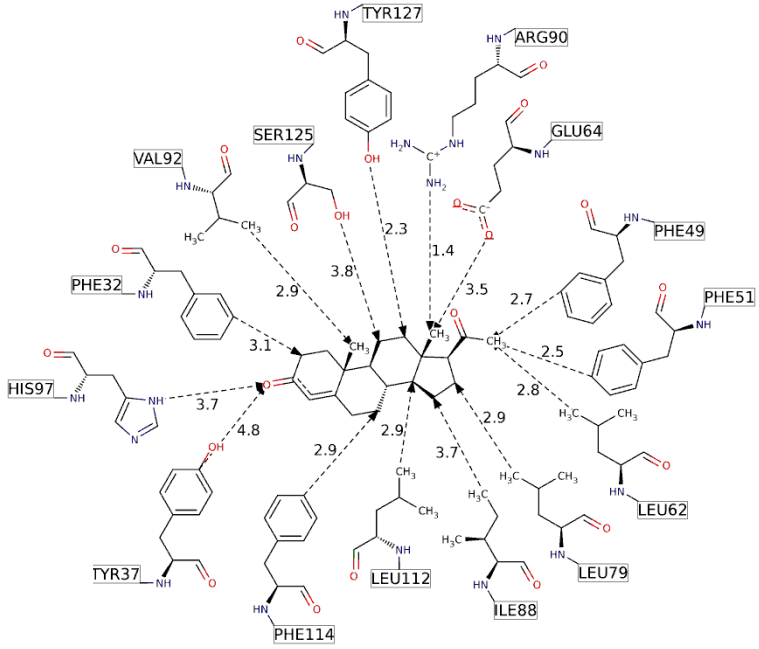
Progesteron



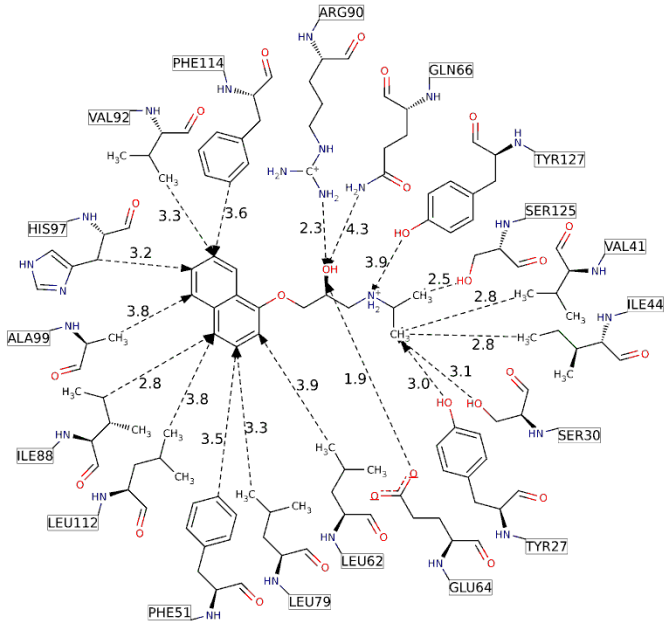
Warfarin

Molecular Docking

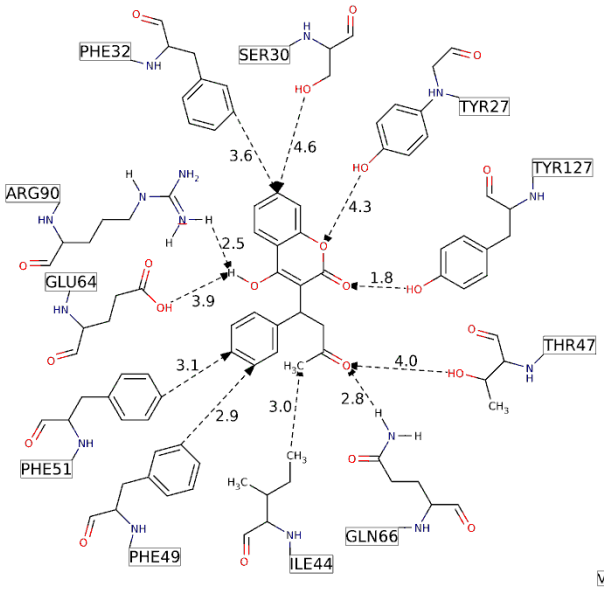
Progesteron



Propranolol



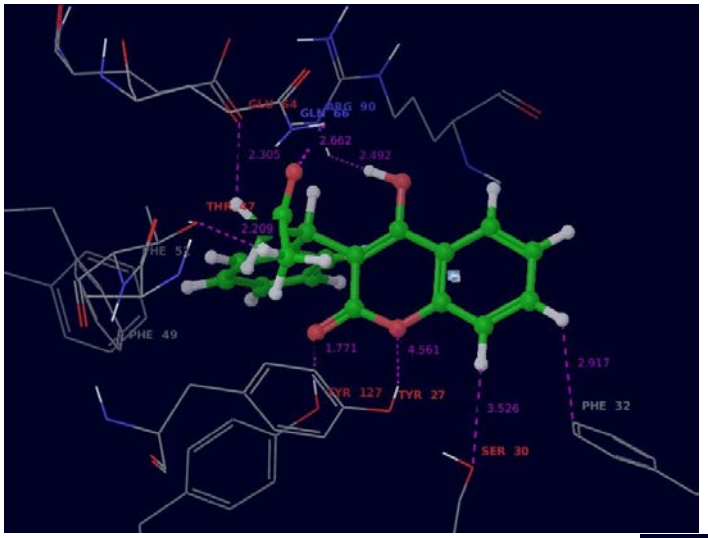
Warfarin



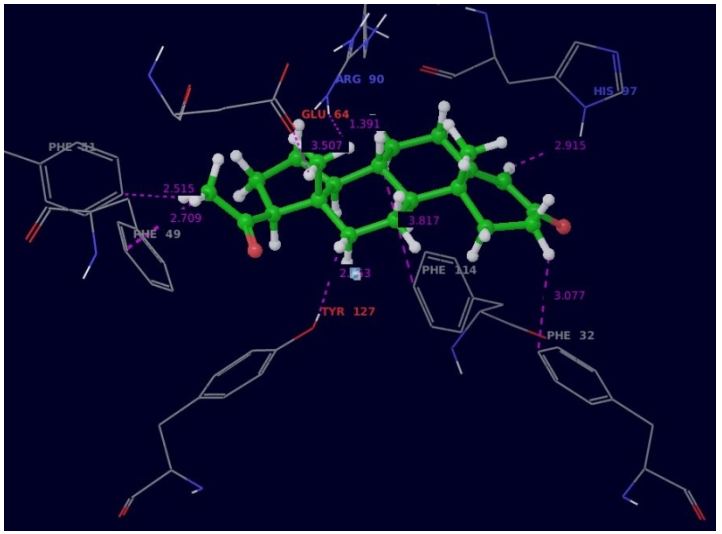
Molecular Docking

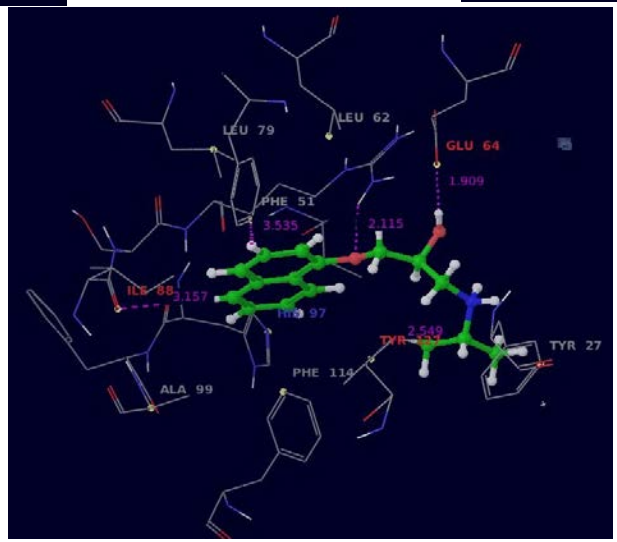
Calculation of the binding Energy

Warfarin -22.6kcal/mol



Progesteron -11.5kcal/mol





Propranolol-14.5kcal/mol

Take home message

MM: Fast (possible large biological systems) but not accurate for nonstandard systems (metal, ions, exited states), no chemical reactions, no explicit polarization

QM: No need of parameterization (good for nonstandard systems), modeling chemical reactions, explicit polarization, calculation of electronic properties but very computationally demanding (max 100 atoms)

QM/MM: Combining advantages of both QM and MM. The active site treated by QM the rest of the system treated by MM

Possible MD simulations and other methods of molecular modeling for the large biological systems with high accuracy in the site of interest

Acknowledgement

Prof. Christopher Reynolds – University of Essex, UK

Dr. Gyury Ferenczy – University of Technology and Economics, Budapest

Vasilina Zayats – INSB Nove Hrady, Czech Republic

David Roesel