CEITEC
Central European Institute of Technology
Brno, Czech Republic

Structural Bioinformatics and Molecular Modeling

Jaroslav Koča
(jkoca@ceitec.cz)

Molecular Modeling

(with computational chemistry as a basis)

Jaroslav Koča

jkoca@ceitec.cz

Experiment vs Simulations

problem

hypothesis

experiments

simulations

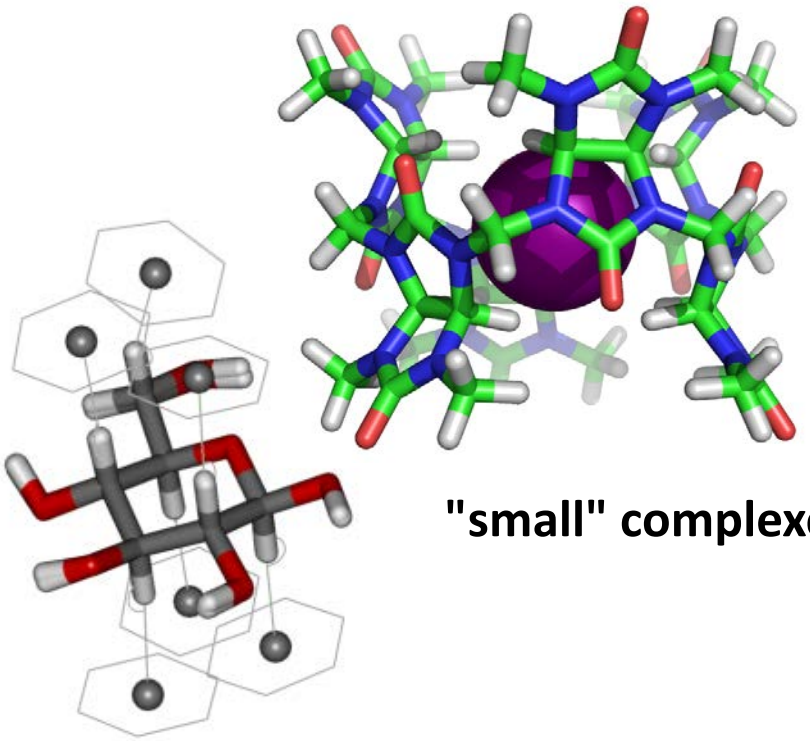
result

knowledge

- reality around us
- resolution
- expensive, dangerous, too difficult to carry out

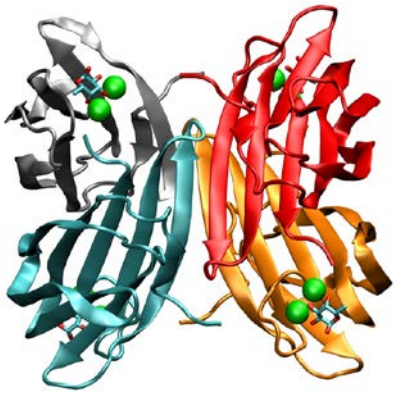
- resolution (most often atomic resolution)
- any thinkable setup
- model

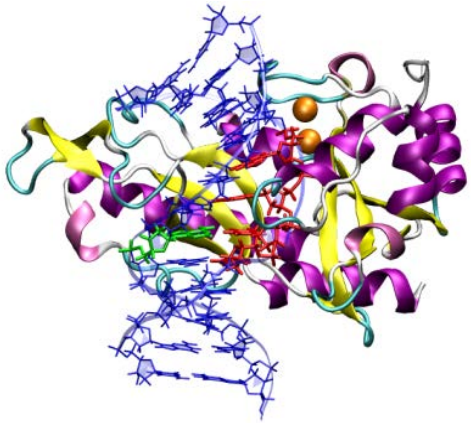
What can be studied?



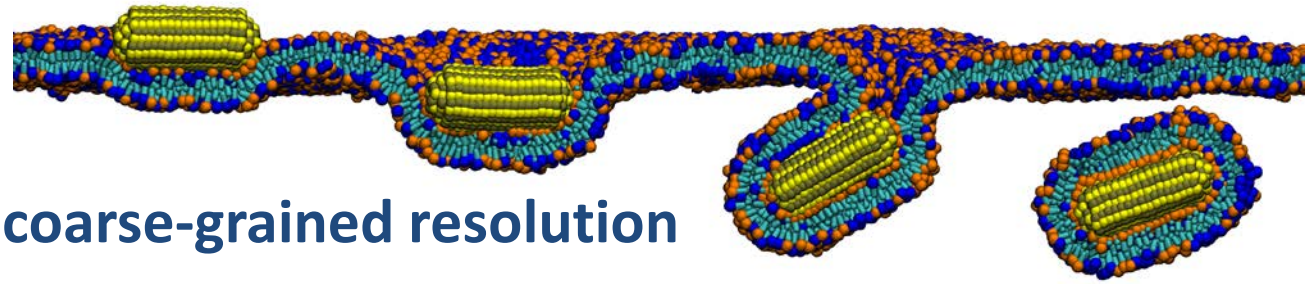
"small" complexes

atomic resolution



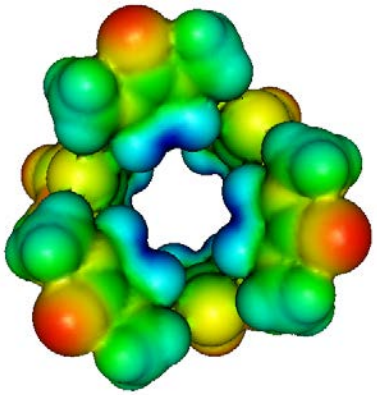


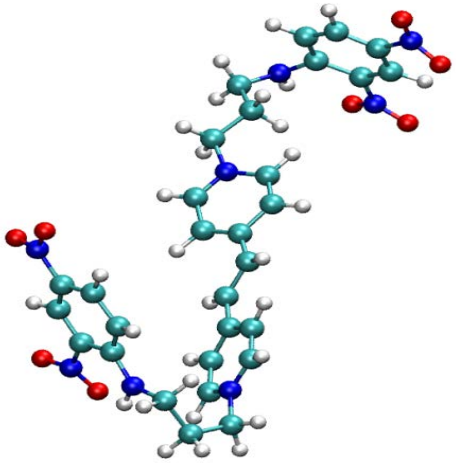
biomolecular systems

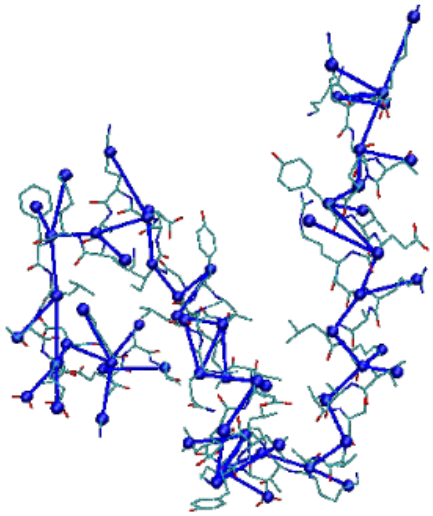


coarse-grained resolution

Levels of Theory







Quantum Mechanics	Molecular Mechanics	*Coarse-grained* Mechanics
Atomic Resolution		Bead Resolution
reactivity	conformational movement	domain movement, folding
up to 1'000 atoms	up to 1'000'000 atoms	up to 1'000'000 beads
up to 100 ps	up to µs	up to ms

Quantum Mechanics

Composition of Molecules

chemistry

physics

molecule

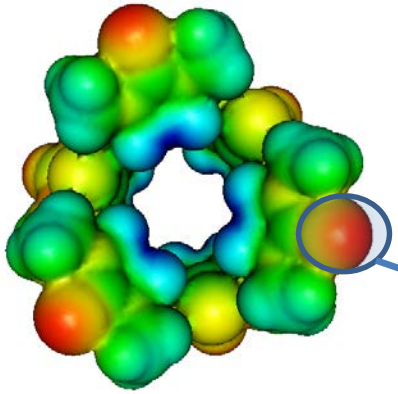
atoms

electrons

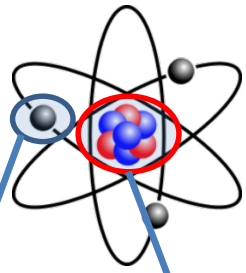
nucleus

protons
neutrons

wave character



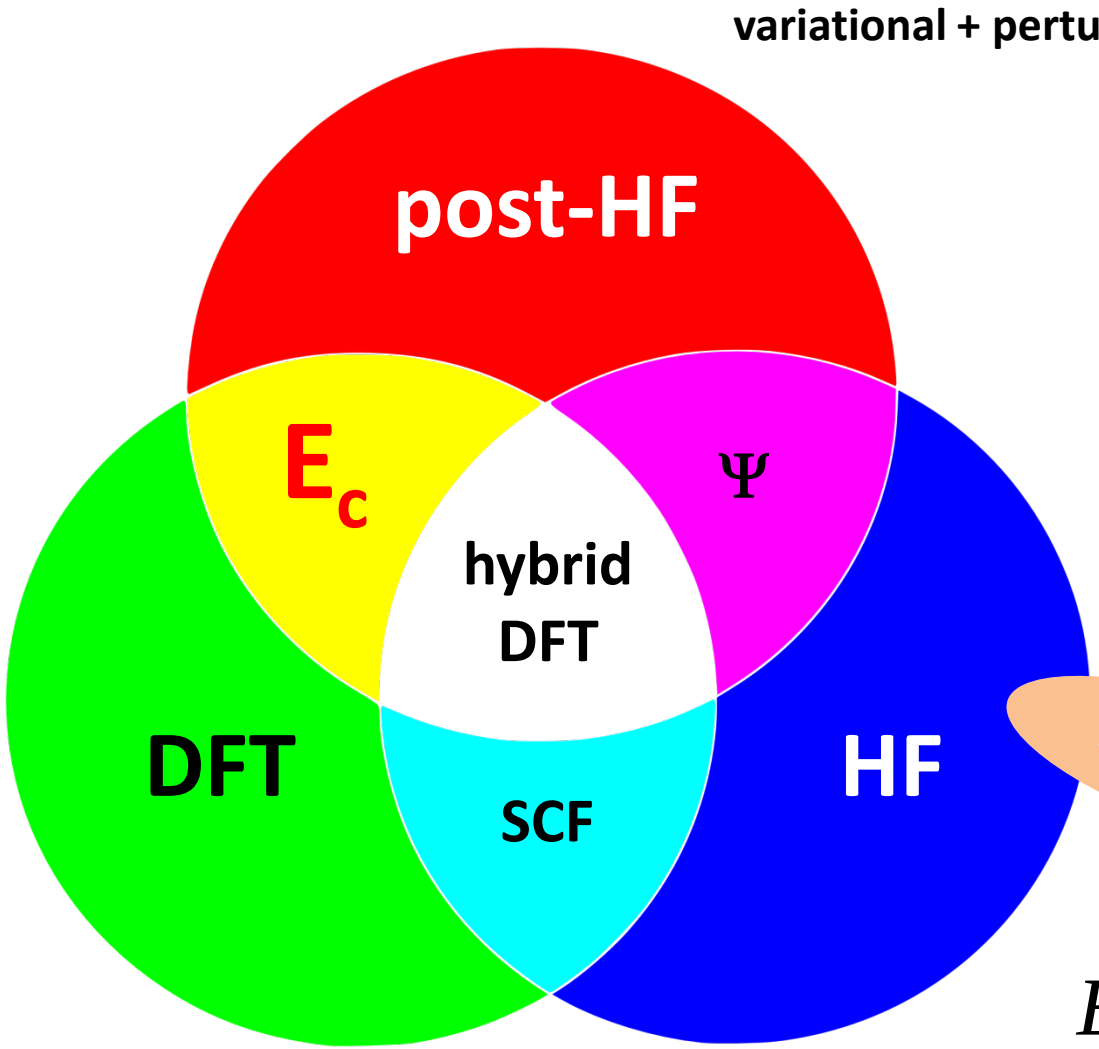
old fashion model



electron nucleus

In quantum chemistry, a nucleus is considered as a mass point with certain positive charge.

Classifications of Methods



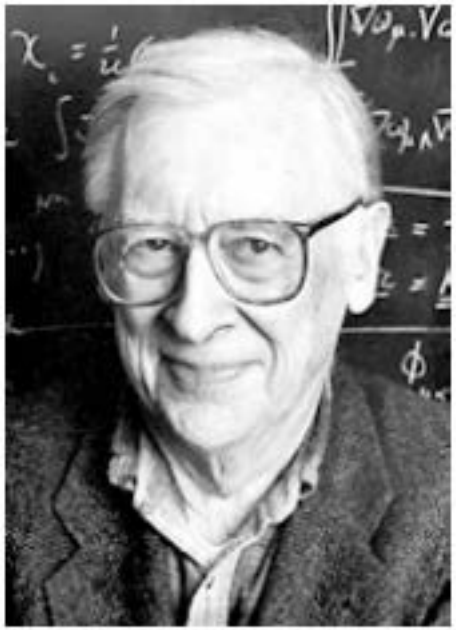
$$E_k = E[\rho_k]$$

$$E_k = E[\Psi_k]$$

Andrew Gilbert

The Nobel Prize in Chemistry 1998

Walter Kohn



John A. Pople

The Nobel Prize in Chemistry 1998 was divided equally between **Walter Kohn** "for his development of the **density-functional theory**" and **John A. Pople** "for his development of **computational methods in quantum chemistry**"

http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1998/

Computational Complexity

	Methods			
Formal Scaling	**HF**	**CI metods**	**MP metods**	**CC metods**
N^4 -> N^2 -> N^1	HF, DFT			
N^5			MP2	CC2 (iterative)
N^6		CISD	MP3, MP4(SDQ)	CCSD (iterative)
N^7			MP4	CCSD(T), CC3 (iterative)
N^8		CISDT	MP5	CCSDT
N^9			MP6	
N^{10}		CISDTQ	MP7	CCSDTQ (iterative)

Legend

HF – Hartree-Fock method

DFT – density functional theory methods
CI – configuration interaction methods
MP - Møller–Plesset perturbation methods
CC – coupled cluster methods

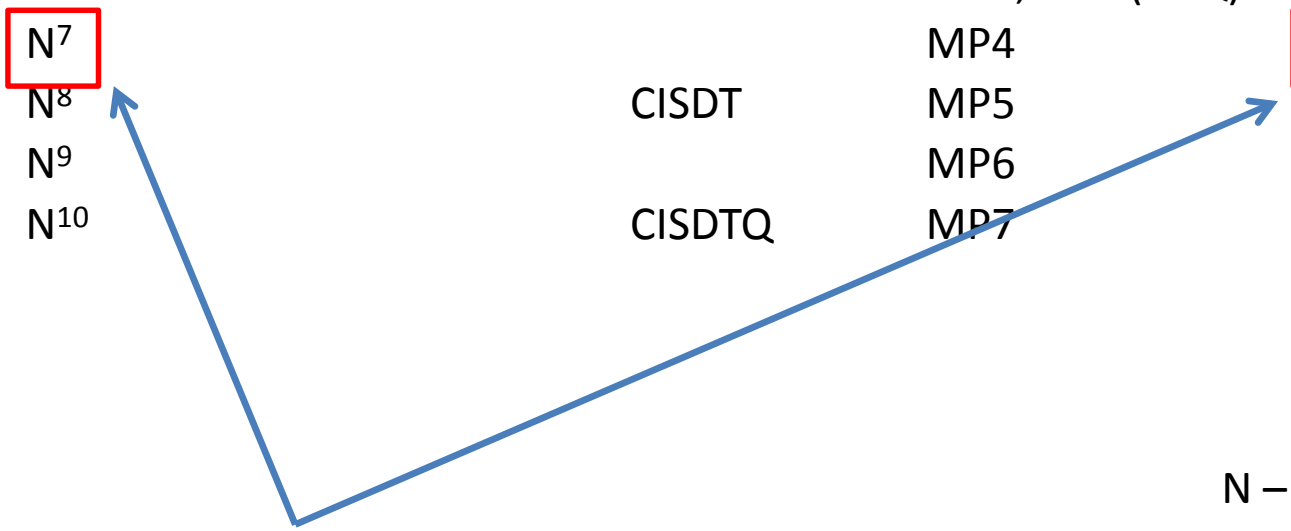
N – number of basis functions

$$N \approx N_A \overline{N_{BF}}$$

Methods highlighted in the box can provide results, which include part of omitted correlation energy.

Complexity & Accuracy

	Methods			
Formal Scaling	**HF**	**CI metods**	**MP metods**	**CC metods**
N^4 -> N^2 -> N^1	HF, DFT			
N^5			MP2	CC2 (iterative)
N^6		CISD	MP3, MP4(SDQ)	CCSD (iterative)
N^7			MP4	CCSD(T), CC3 (iterative)
N^8		CISDT	MP5	CCSDT
N^9			MP6	
N^{10}		CISDTQ	MP7	CCSDTQ (iterative)



Coupled cluster methods are able to reach **chemical accuracy** but only for small molecules.
On currently available hardware, it is possible to apply the CCSD(T) method to systems containing up to **50 atoms**.

N – number of basis functions

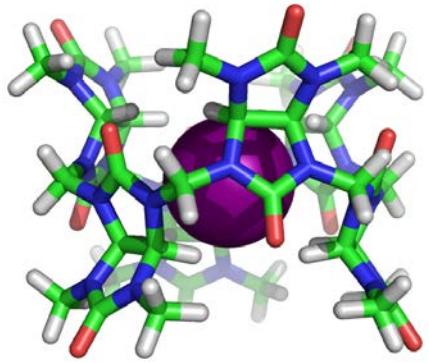
$$N \approx N_A \overline{N_{BF}}$$

Accuracy

Small Numbers from Big Numbers

RI-BLYP-d3/def2-TZVPP (vacuum)

```
BU6/I(-) -4152.181032604 Hartree
BU6      -3854.321084579 Hartree
I(-)      -297.740268591 Hartree
         ---------------
            -0.119679434 Hartree
                 ~ -75.1 kcal/mol
```



bambus[6]uril/anion interaction
(139 atoms)

"Chemical" Accuracy

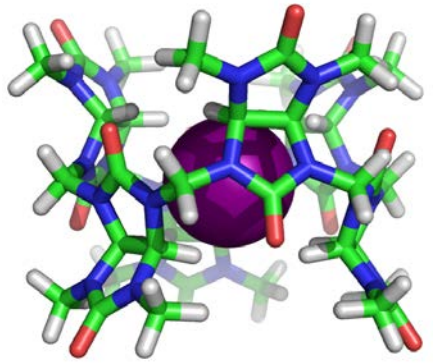
error < 1 kcal mol^{-1}

Accuracy

Small Numbers from Big Numbers

RI-BLYP-d3/def2-TZVPP (vacuum)

```
BU6/I(-) -4152.181032604 Hartree
BU6      -3854.321084579 Hartree
I(-)      -297.740268591 Hartree
          ---------------
             -0.119679434 Hartree
                   ~ -75.1 kcal/mol
```



bambus[6]uril/anion interaction
(139 atoms)

"Chemical" Accuracy

error < 1 kcal mol^{-1}

Take Home Message

Quantum Mechanics

- **due to dual character** (particles/waves) of electrons and nuclei, any chemical system has to be described by the **Schrödinger equation** (SE)
- albeit very simple notation of SE, its **analytical** solution is not impossible even for simple chemical systems (two and more electrons)
- **several approximations** were introduced that make **numerical** solution of SE possible but some of them introduces very **serious errors** (correlation energy)
- it is possible to fix these errors but procedures are very **computationally demanding** even for small systems (up to 50 atoms)
- once **approximate solution** of SE is known then **any property** (including energy) can be easily obtained from **wavefunction of given state**
- SE is naturally able to describe **chemical reactions**

Molecular Mechanics

Can we make calculations faster?

Molecular Mechanics

Schrodinger equation - quantum mechanical description

$$\hat{H}\psi_k^{\mathbf{R}}(\mathbf{r}_e) = E_k(\mathbf{R})\psi_k^{\mathbf{R}}(\mathbf{r}_e)$$

approximation
electron motions is omitted
(electron motions is implicitly included in empirical parameters)

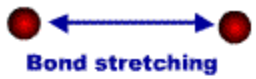
$$E_k(\mathbf{R}) = \underbrace{E_{bonds} + E_{angles} + E_{torsions}}_{\text{bonded contributions}} + \underbrace{E_{ele} + E_{vdw}}_{\text{non-bonded contributions}} + \ldots$$

Classical physics - mechanical description

Bonded Contributions

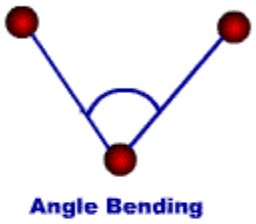
Bonds

$$E_{bonds} = \sum_{b=1}^{bonds} \frac{1}{2} K_b \left(d_b - d_{b0}\right)^2$$



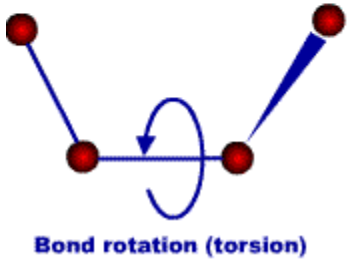
Angles

$$E_{angles} = \sum_{a=1}^{angles} \frac{1}{2} K_a \left(\theta_a - \theta_{a0}\right)^2$$



Torsion angles

$$E_{torsions} = \sum_{t=1}^{torsions} \sum_{n} \frac{V_{t,n}}{2} \left(1 + \cos\left[n\varphi_t - \delta_{t,n}\right]\right)$$

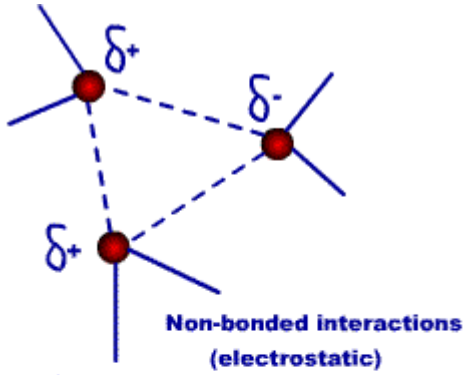


empirical parameters

Non-bonded Contributions

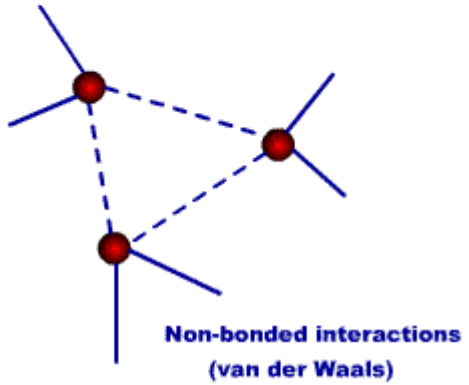
Electrostatic interactions

$$E_{ele} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{4\pi\varepsilon_o} \frac{q_i q_j}{r_{ij}}$$



van der Waals interactions

$$E_{ele} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]$$



empirical parameters

PME – particle meshed Ewald (N $\log_2$ N)

N – number of atoms

Take Home Message

Molecular Mechanics

- based on **classical mechanics**, **accuracy is compromised**
- **empirical parameters (parameter set, force field)** derived from experimental or high-level QM calculations are required
- **parameter transferability problems**, parameters are derived for limited set of systems (proteins, nucleic acids, lipids, saccharides, etc), no general parameter set yet
- **very fast calculations** of systems containing up to millions of atoms
- electrons are not explicitly accounted in the theory thus it is **not possible** to describe **chemical reaction** (exceptions: ReaxFF, EVB)
- suitable for study of **conformational changes and structure**

Potential Energy Surface

Configuration Space

$E(\mathbf{R})$

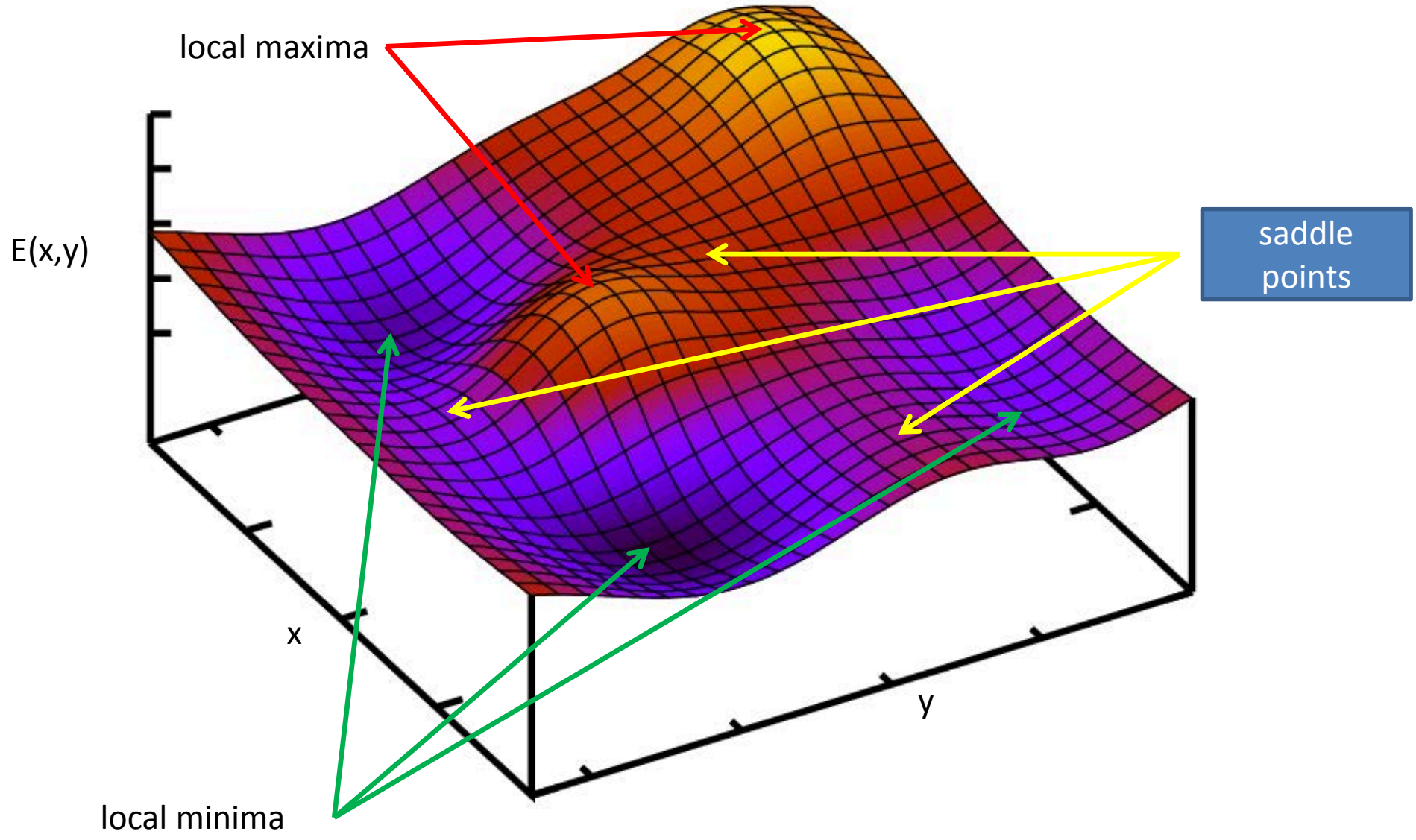
R = point in 3N-dimensional space (N is number of atoms)

$$\mathbf{R} = \{x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_N, y_N, z_N\}$$

Cartesian coordinates of the first atom

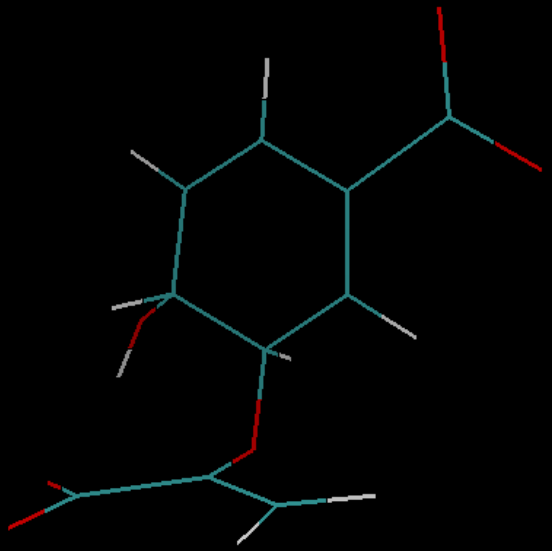
Individual points form the **configuration space**.
Every point in the configuration space represents a **unique structure** of studied system.

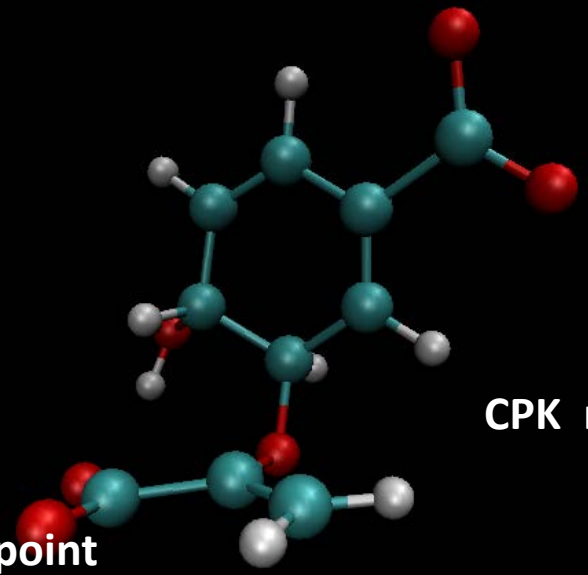
Two-dimensional Case



Models – small molecules

lines

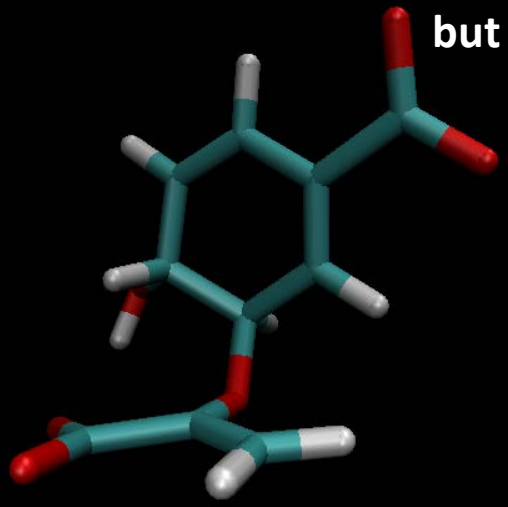


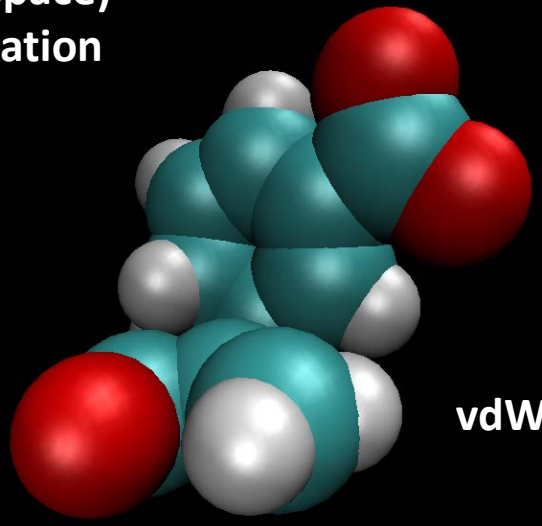


CPK model

the same structure (point in the configuration space) but different visualization

tubes

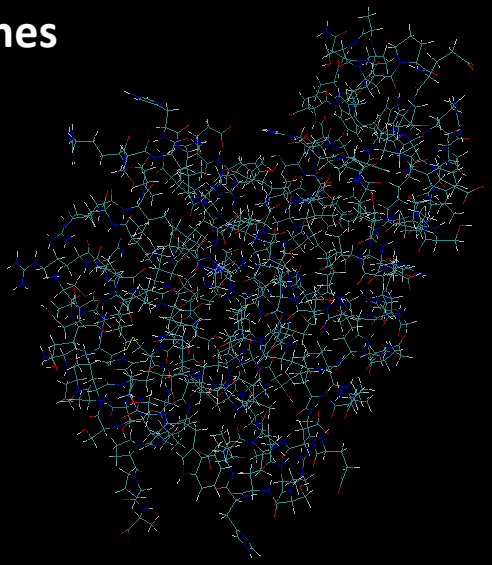




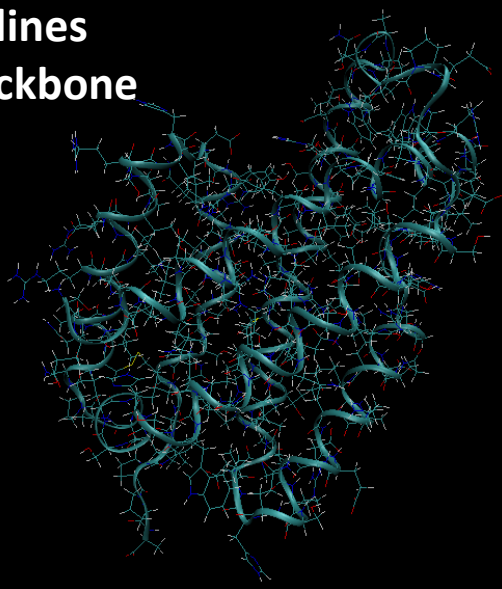
vdW model

Models – biomolecules

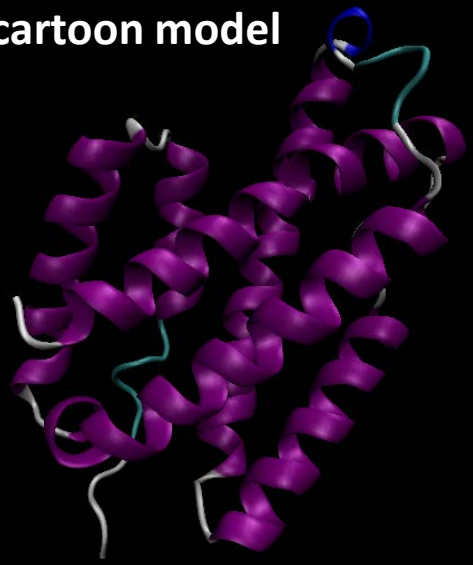
lines



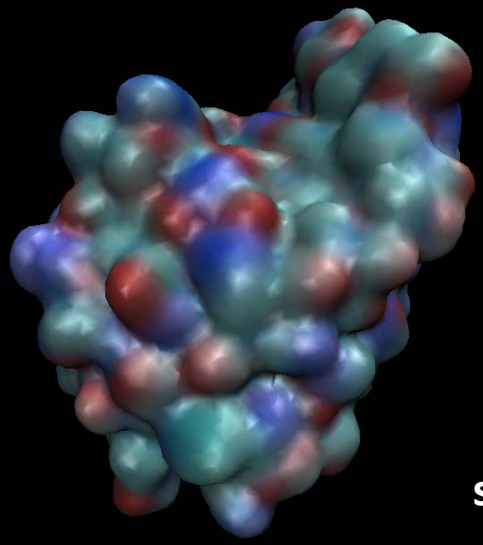
lines
backbone



cartoon model



the same structure (point in the configuration space) but different visualization



surface

Different visualization models are used to depict various structural features or internal properties of molecules or molecular assemblies, which then improve perceptions of studied systems/problems.

Take Home Message

Potential Energy Surface

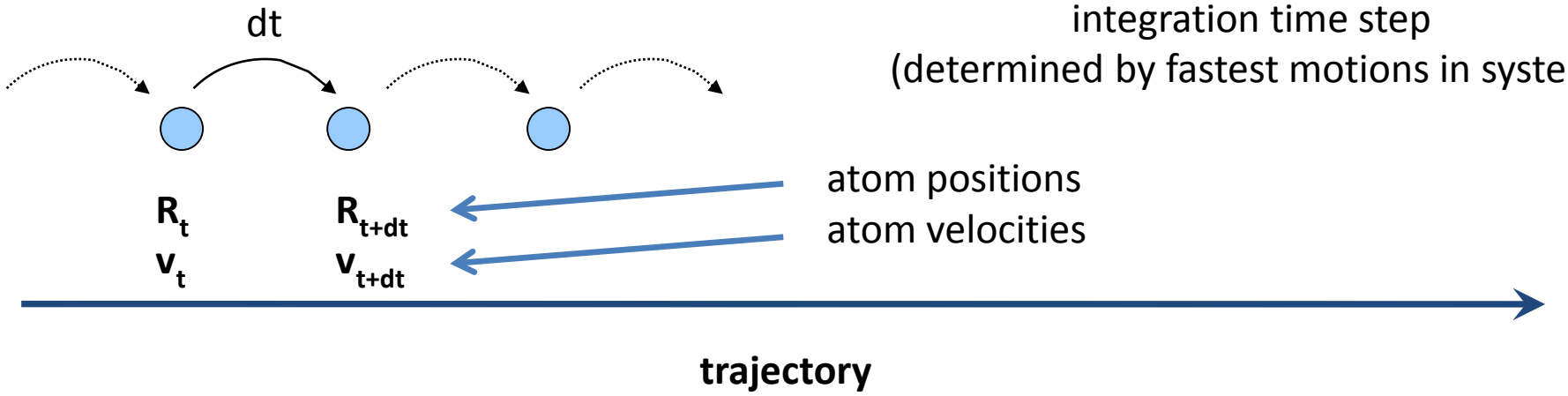
- stationary points (**local minima** and **saddle points**) are very important for description of reactivity and kinetics of small molecular systems
- stationary points represents **unique structures** of molecular system

Molecular Dynamics

How to get thermodynamical parameters of complex systems ...

Molecular Dynamics

Evolution of system in time



dt = 1fs

integration time step
(determined by fastest motions in system)

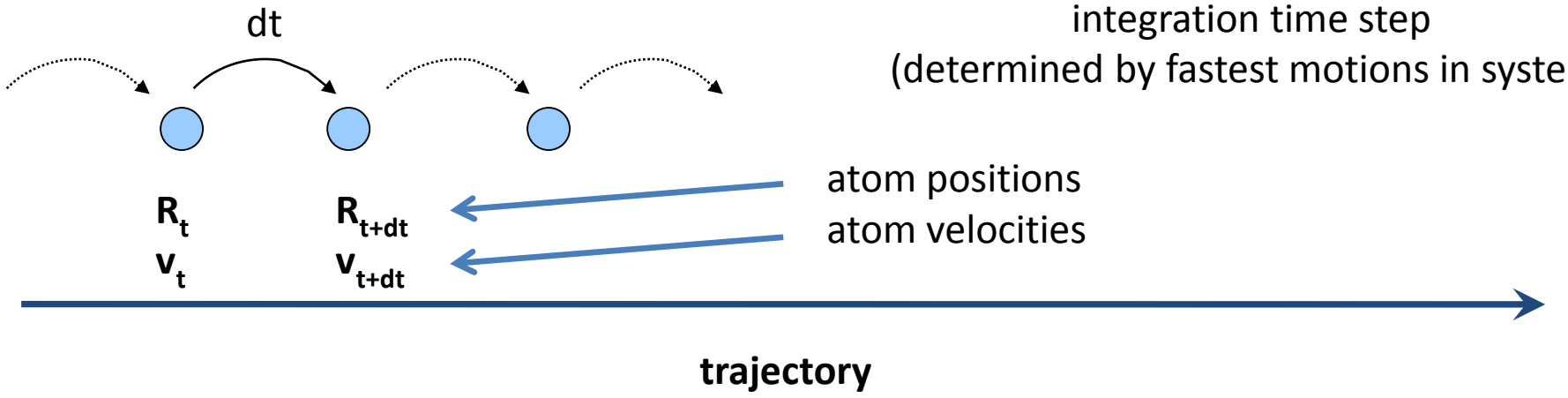
Equation of Motions

$$-\frac{\partial^2 E(\mathbf{R})}{\partial \mathbf{r}_i^2} = m_i \frac{d^2 \mathbf{r}_i}{dt^2}$$

$$\mathbf{F}_i = m_i \mathbf{a}_i$$

Molecular Dynamics

Evolution of system in time



dt = 1fs

integration time step
(determined by fastest motions in system)

Sampling problem

1 μs = 10^9 fs

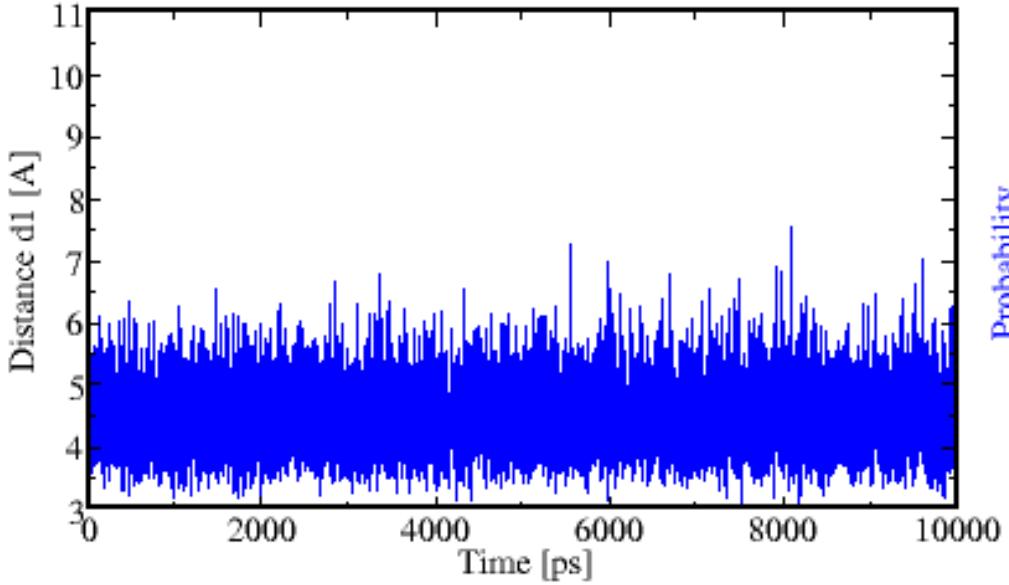
1631 atoms, AMBER, GPU acceleration

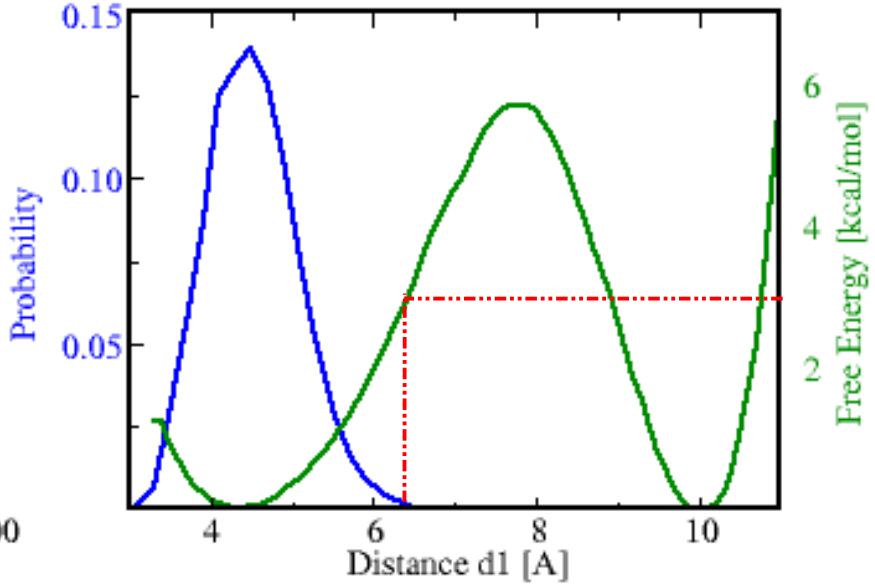
1 step ~ 1 ms CPU/GPU time
10^9 steps ~ 12 days CPU/GPU time

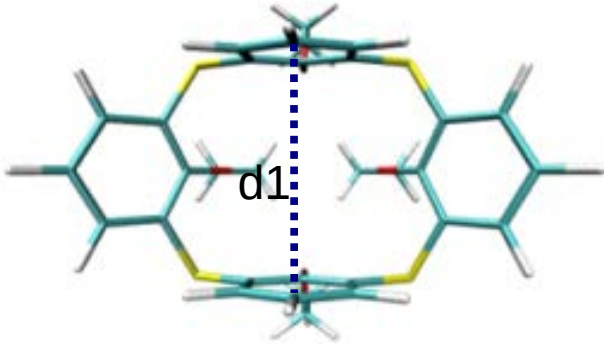
Characteristic timescales:

- Protein folding (ms)
- Chemical reactions (ms, s, h)

Sampling Problem







10 ns long simulation is able to discover free energy landscape with depth only about **3 kcal/mol**.

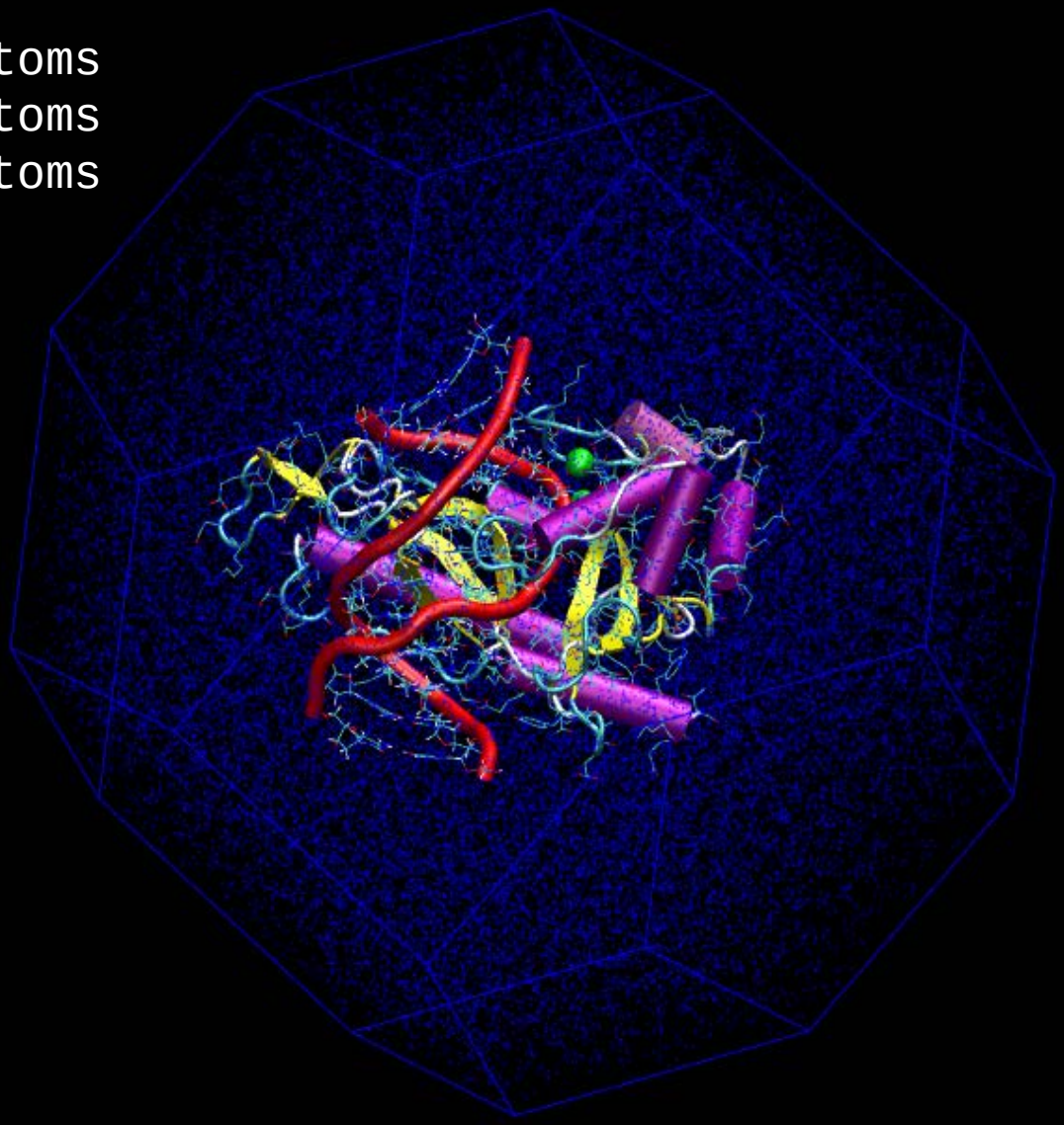
Multiscale Methods

QM/MM Methods

How to describe reactivity occurring in biomolecules (enzymes)?

Enzymatic Reactions

Enzyme: ~4,300 atoms
Water: ~42,000 atoms
Total: ~46,500 atoms

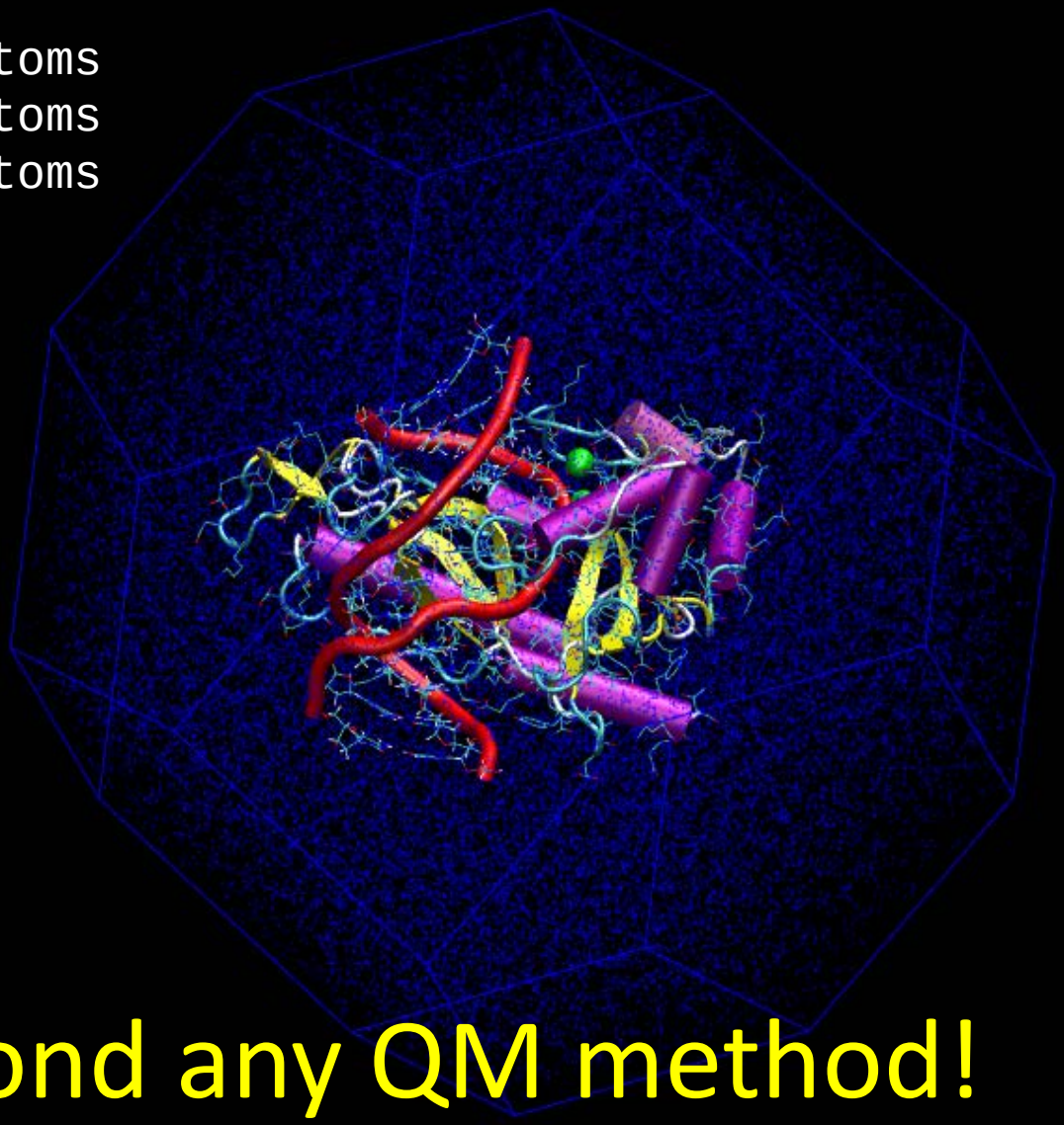


MutH

Enzymatic Reactions

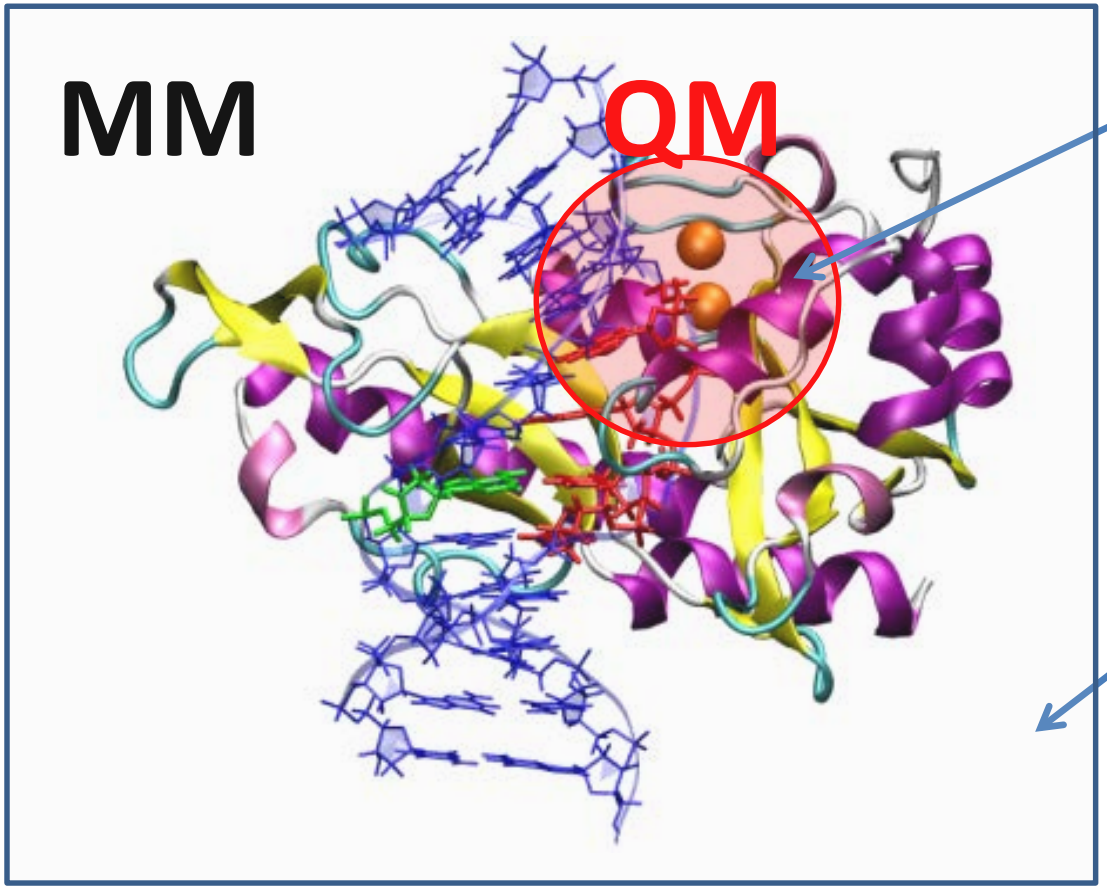
Enzyme: ~4,300 atoms
Water: ~42,000 atoms
Total: ~46,500 atoms

MutH



Beyond any QM method!

Enzymatic Reactions



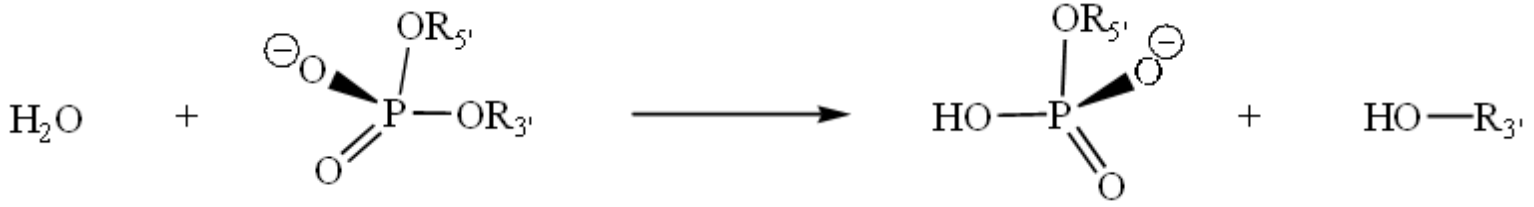
QM

active site only
tens to hundreds atoms

MM

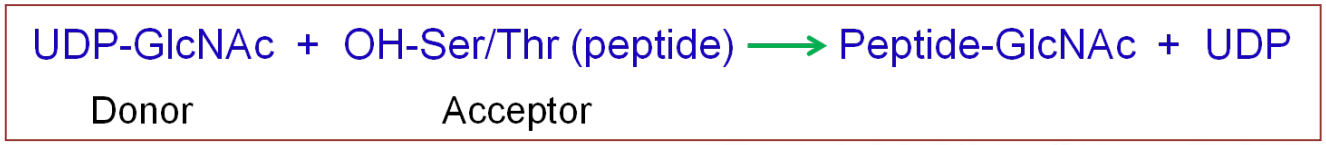
rest of enzyme
solvent (water)

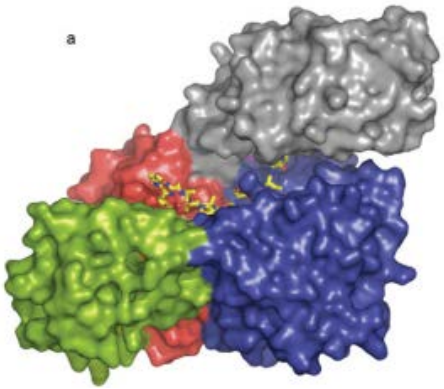
Catalyzed reaction – hydrolysis of phosphodiester bond



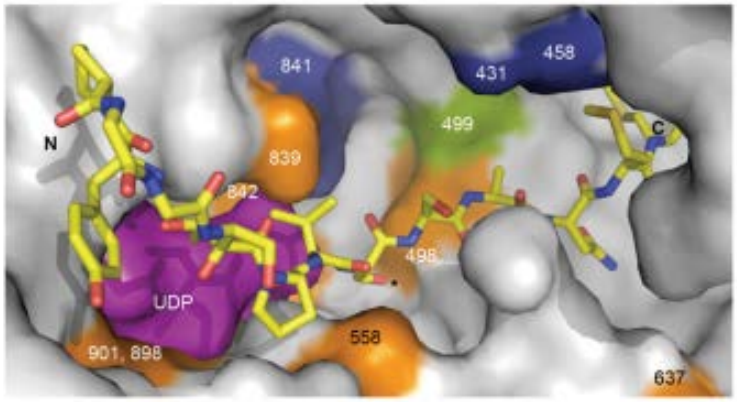
Introduction - OGT glycosyltransferase

- Uridine diphospho-*N*-acetylglucosamine: polypeptide β-N-acetylaminyltransferase; PDB ID: 3PE4
- Enzymatic transfer of N-acetylglucosamine molecule on Ser/Thr residue of protein
- Inverting glycosyltransferase of the GT-B family
- Post-translational modification: first reported in 1984





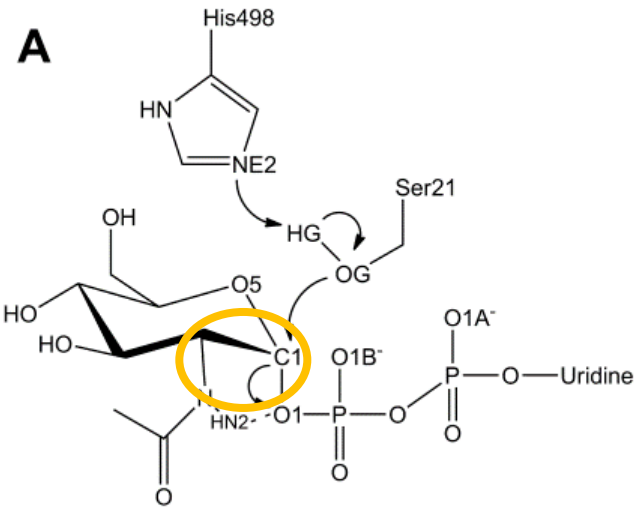
OGT Crystal Structure (Walker et al, 2011)



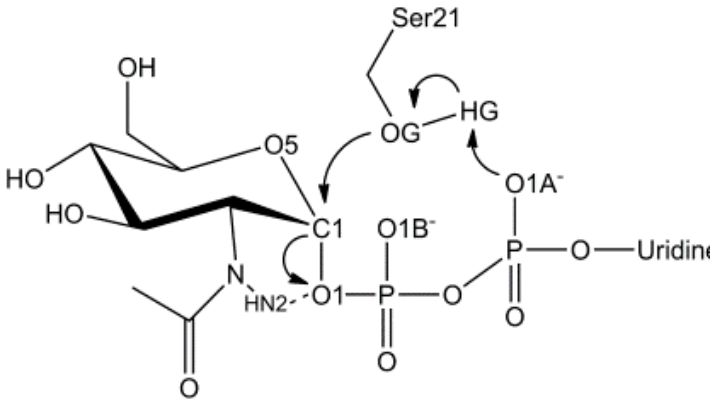
OGT Catalytic Site

Different Mechanisms Proposed for OGT

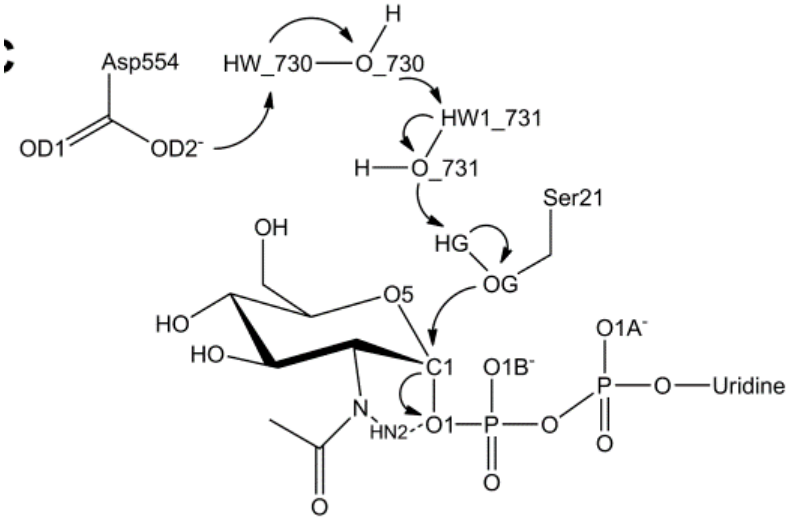
1. **His498 as catalytic base (M_{His})**
   (Lazarus et al. 2011
   Tvaroska et al. 2012)



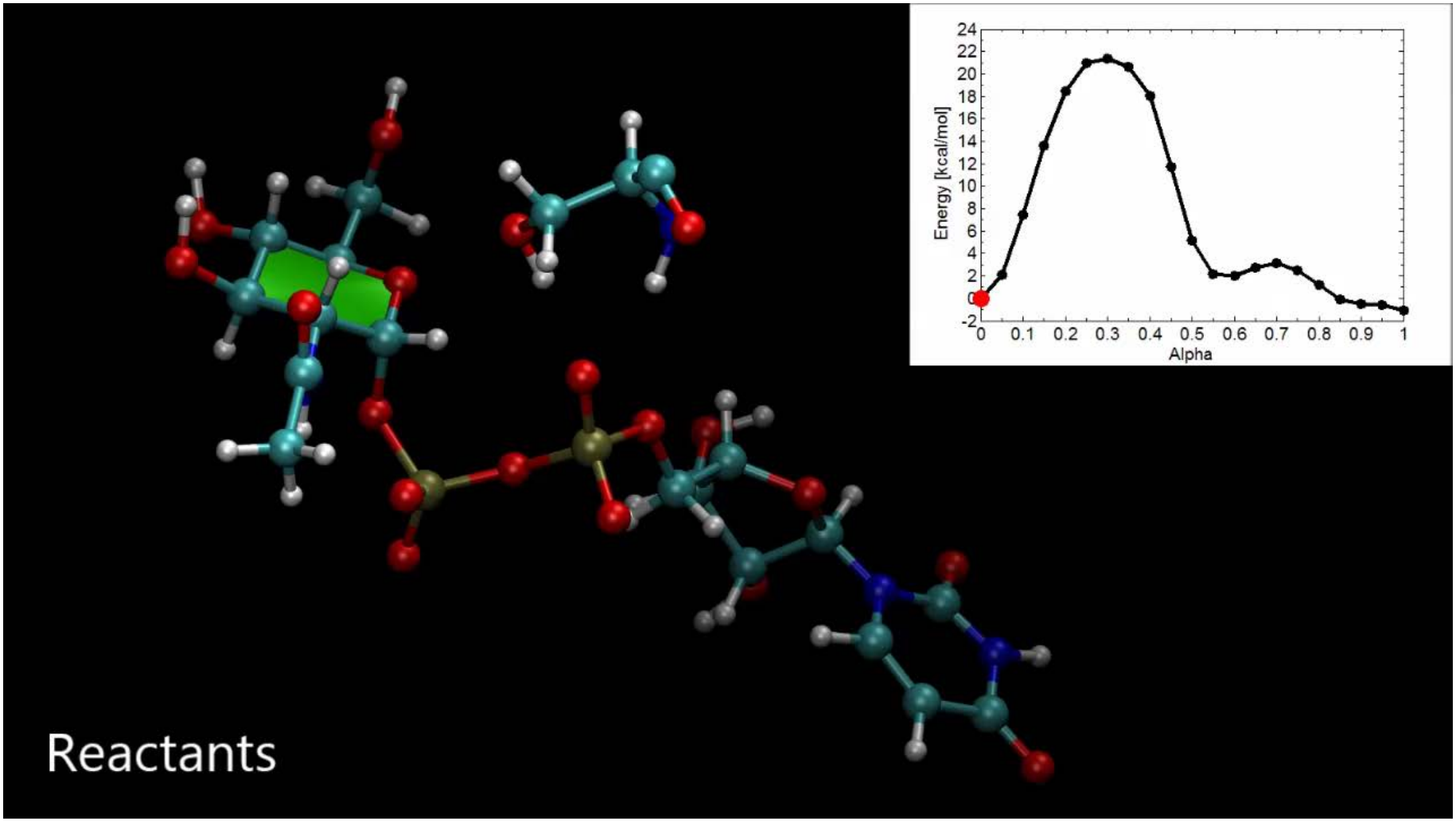
2. **α-phosphate as base (MPO4)**
   (Schimpl et al. 2012)



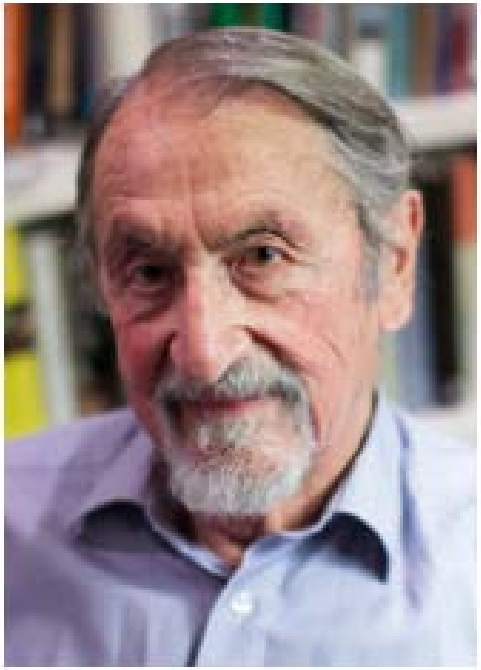
3. **Water molecule for shunting proton to ASP554 (MAsp)**
   (Lazarus et al. 2012)



Most Probable Reaction Path – M_{PO4}



Nobel Laureates in Chemistry 2013



© Harvard University

Martin Karplus

Université de Strasbourg, Strasbourg, France, Harvard University, Cambridge, MA, USA

Photo: © S. Fisch

Michael Levitt

Stanford University School of Medicine, Stanford, CA, USA

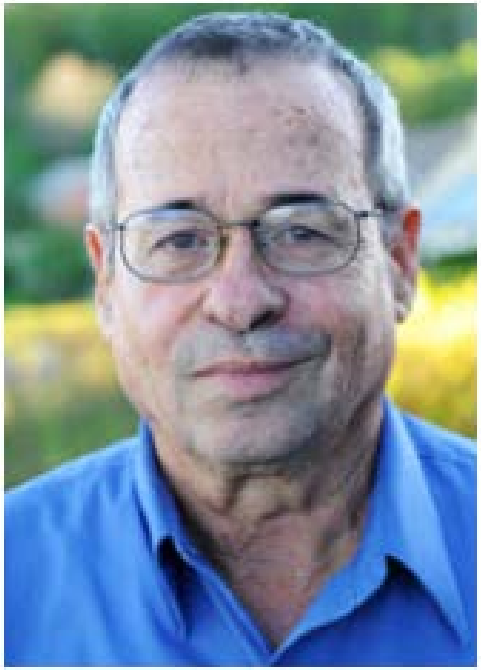


Photo: Wikimedia Commons

Arieh Warshel

University of Southern California, Los Angeles, CA, USA

http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/

Take Home Message

Figure 2. Newton and Schrödinger's cat. Previously, classical physics and quantum chemistry belonged to rivalling worlds. The Nobel Laureates in Chemistry 2013 have opened a gate between those worlds and have brought about a flourishing collaboration.

- study of chemical reactions occurring in very large molecular assemblies is possible by QM/MM methods ("routine" calculations)
  - enzymatic reactions (drug development)
  - catalysis in solid state (zeolites, industry)
- many unresolved issues (QM accuracy, MM polarization, boundary problems , ...)

http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/

Thank you for your attention !

Central European Institute of Technology
Masaryk University
Kamenice 753/5
625 00 Brno, Czech Republic

www.ceitec.muni.cz | info@ceitec.muni.cz