

Computational Investigations of Biomolecules

Talk for the Hauptseminar
„Konformationsdynamik in Biomolekülen:
Experiment und Theorie“

Physics Department
Karlsruhe Institute for Technology (KIT)

Matthias Ernst
Supervisor W. Wenzel

2010/7/8-9



Outline

I. Introduction: Transport of small molecules in proteins

II. QM/MM – Theory and Methods

- 1) Classical Movement in Potential and Structure Optimisation
- 2) Schrödinger equation and Born-Oppenheimer-Approx.
- 3) Methods of Quantum Chemistry
 - a) Illustration of chemical bond: LCAO-MO-Theory
 - b) Hartree-Fock
 - c) Electron configuration methods
 - d) Density Functional Theory
- 4) Force Field Methods (Molecular Mechanics)
- 5) Overview: Capabilities, Scales
- 6) Hybrid Methods (QM/MM)

III. Dynamics

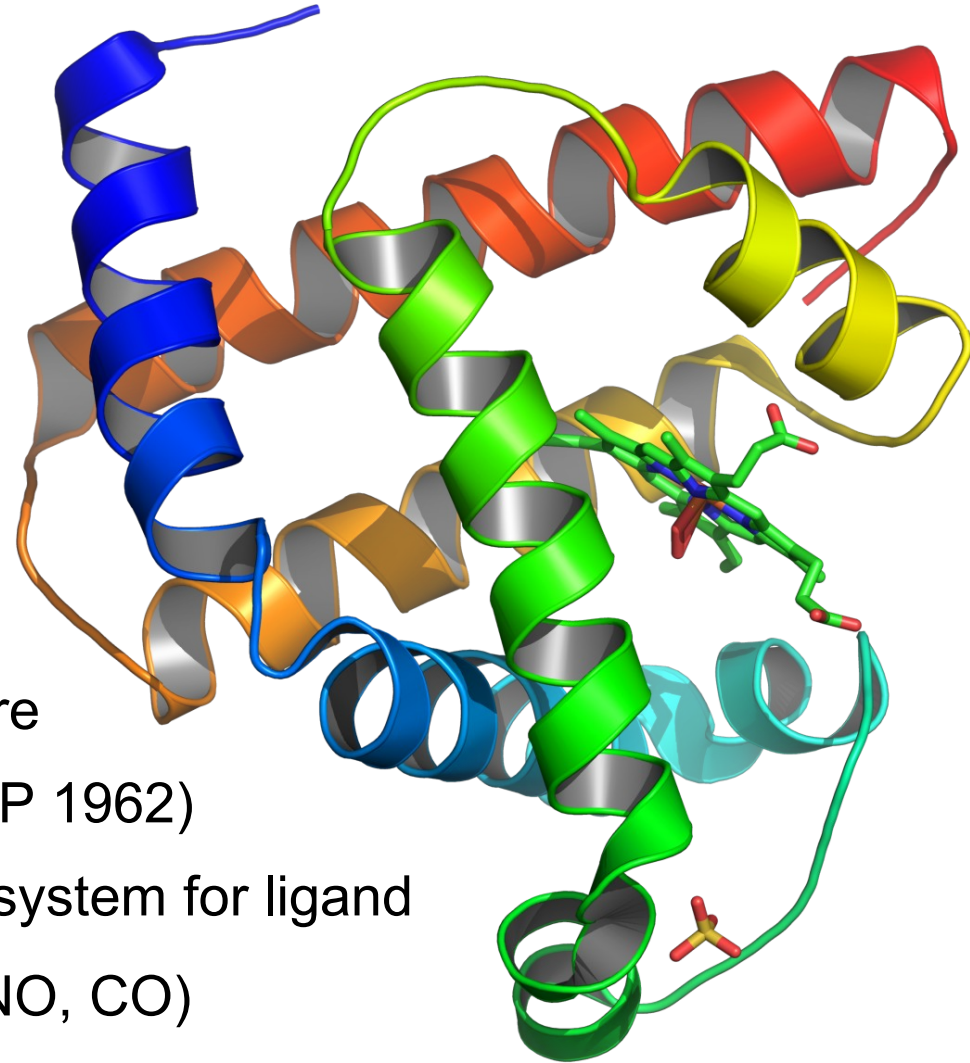
- 1) Molecular vibrations
- 2) Transition states and reaction rate constants
- 3) Car-Parrinello Molecular Dynamics

IV. Example: Ligand migration in Myoglobin

Transport of small Molecules in proteins

I. Introduction

- Myoglobin (Mb):
 - small heme-based protein
 - found in muscle tissue
 - capable of binding oxygen
 - => Oxygen storage with whales („Sperm whale“ = Pottwal)
- First protein of which 3D structure was revealed (Kendrew 1958, NP 1962)
- Very intensively studied, model system for ligand migration in proteins (e.g. O₂, NO, CO)
- Experimental and theoretical studies have been carried out
- No direct „channel“ for ligand from outside to Fe-Ion in heme
 - => How does ligand migration takes place?



1) Classical Movement in Potential

→ Movement given by Newton's second law (non-relativistic):

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = m \cdot \mathbf{a} = -\nabla V$$

→ Movement is determined by potential V

→ Movement without external forces:

minimize total energy

$$E = \mathcal{H} = T + V$$

→ If potential for molecular system is known:

- determination of molecular structure via energy minimisation
- motion of atoms by Newton's second law

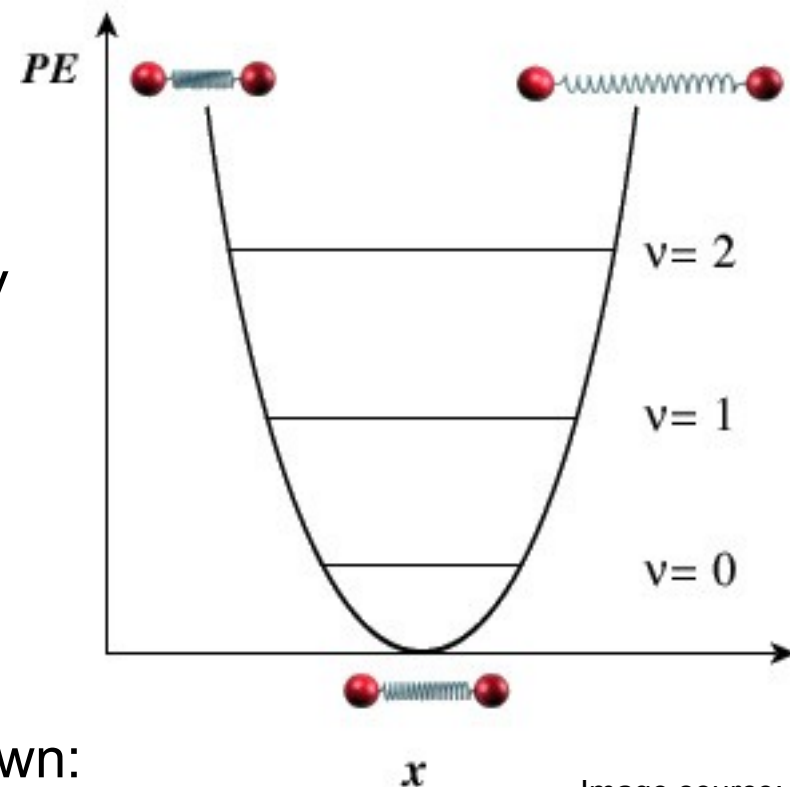


Image source:
http://neon.otago.ac.nz/chemlect/chem306/pca/IR_Raman/index.html

2) Schrödinger equation

- Atoms are small particles => quantum mechanics
- Schrödinger equation instead of Newton or Lagrange or Hamilton eq.
- Time-dependent SE:

$$\mathcal{H} \Psi = i \hbar \frac{\partial}{\partial t} \Psi$$

For bound systems, separation results in time independent SE:

$$\mathcal{H}(\mathbf{r}) \Psi(\mathbf{r}) = E \Psi(\mathbf{r})$$

- Hamiltonian for molecule of N nuclei and n electrons (a.u.):

$$\mathcal{H} = \hat{T} + \hat{V} = \underbrace{-\frac{1}{2} \sum_{i=1}^n \nabla_i^2}_{T_e} - \underbrace{\sum_{A=1}^N \frac{1}{2M_A} \nabla_A^2}_{T_N} - \underbrace{\sum_{i=1}^n \sum_{A=1}^N \frac{Z_A}{r_{iA}}}_{V_{eN}} + \underbrace{\sum_{i=1}^n \sum_{j>i=1}^n \frac{1}{r_{ij}}}_{V_{ee}} + \underbrace{\sum_{B>A}^N \sum_{A=1}^N \frac{Z_A Z_B}{R_{AB}}}_{V_{NN}}$$

2) Born-Oppenheimer-Approximation

→ Nuclei are much heavier than electrons

=> electrons are moving in field of „fixed“ nuclei

$$\mathcal{H}_{\text{elec}} = \underbrace{-\frac{1}{2} \sum_{i=1}^n \nabla_i^2}_{T_e} - \underbrace{\sum_{i=1}^n \sum_{A=1}^N \frac{Z_A}{r_{iA}}}_{V_{eN}} + \underbrace{\sum_{i=1}^n \sum_{j>i=1}^n \frac{1}{r_{ij}}}_{V_{ee}} + \underbrace{\sum_{B>A}^N \sum_{A=1}^N \frac{Z_A Z_B}{R_{AB}}}_{V_{NN}(\text{additional constant})}$$

electronic wave function depends parametrically on positions of nuclei

→ Nuclei move in average field of electrons:

$$\mathcal{H}_{\text{nuc}} = \underbrace{-\sum_{A=1}^N \frac{1}{2M_A} \nabla_A^2}_{T_N} + \underbrace{E_{\text{elec}}}_{\mathcal{H}_{\text{elec}} \Phi_{\text{elec}} = E_{\text{elec}} \Phi_{\text{elec}}}$$

→ Solutions of electronic SE lead to potential energy hypersurface (PES), nuclei movement is treated classically on PES (e.g. by minimisation)

2) Born-Oppenheimer-Approximation

Illustration of Potential energy surface

→ Nuclei are moving classically on the PES like „in the mountains“

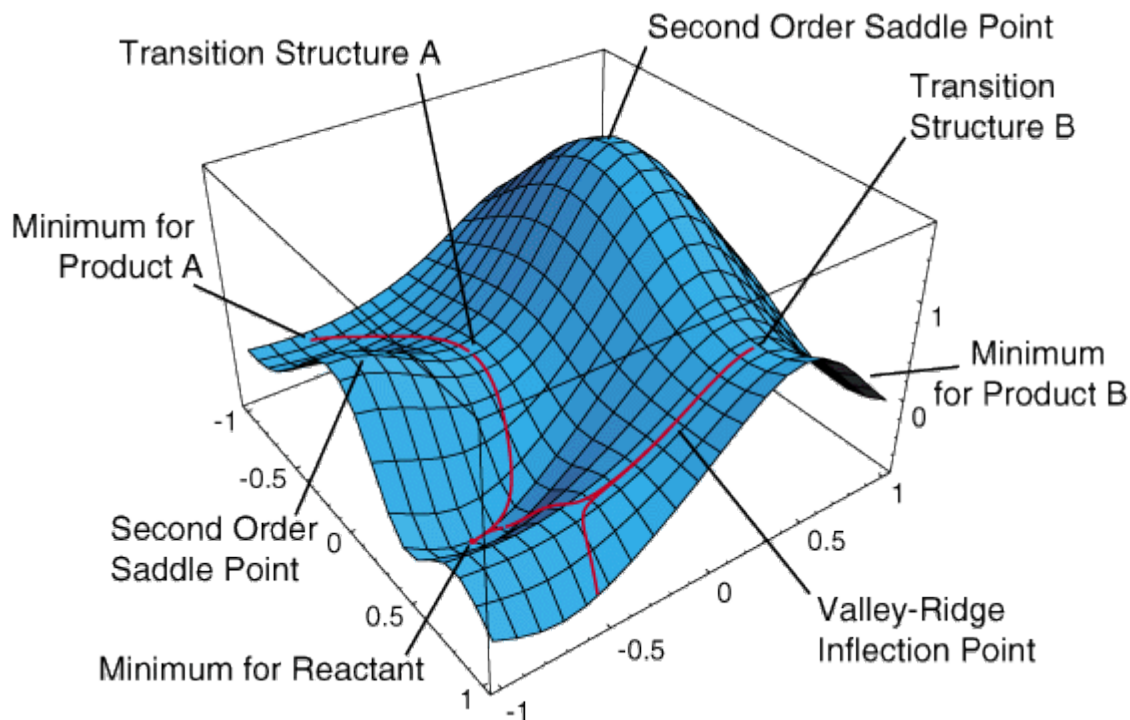


Image source:
<http://www.chem.wayne.edu/~hbs/chm6440/PES.html>

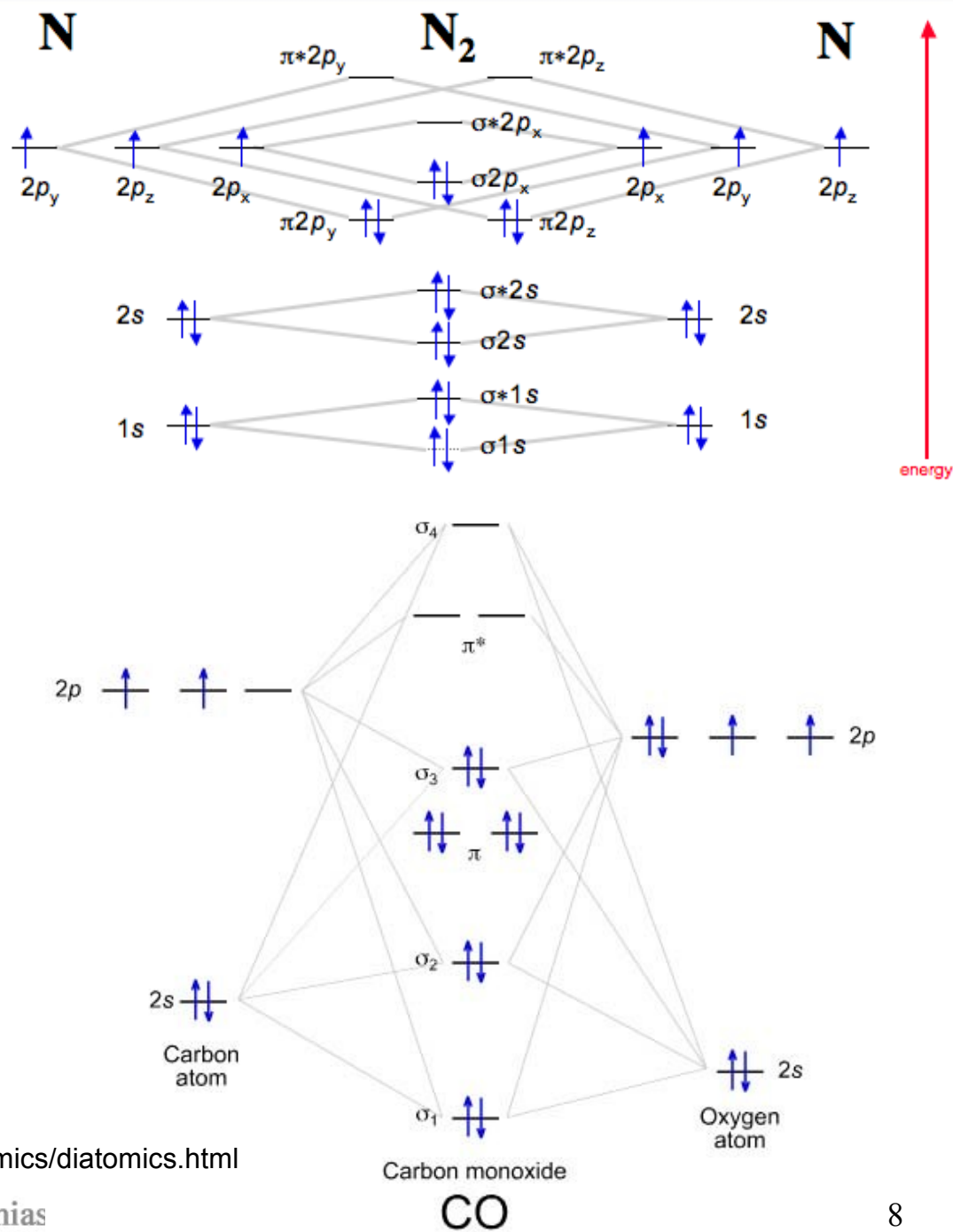
- Reaction can be seen as trajectory on PES
- minima: (meta-)stable configurations (Reactants, Products)
- 1st order saddle points: Transition states

3) Methods of Quantum Chemistry

a) LCAO-MO-Theory

- Molecular orbitals:
Linear Combination
of *Atomic Orbitals*
- Overlap between atomic orbitals forms bond
- #MOs = #AOs
- Antibonding stronger than bonding
- Qualitative description of chemical bond
- Exact energies cannot be predicted

Image source:
http://www.meta-synthesis.com/webbook/39_diatomics/diatomics.html



3) Methods of Quantum Chemistry

b) Hartree-Fock-Theory

- Each electron is moving independently in mean field of all the others
- *Ansatz* for wave-function: antisymmetrized product of single-particle functions (spin-orbitals) $\chi(\mathbf{x}) = \psi(\mathbf{r}) \cdot \sigma(\omega) \Rightarrow$ Slater determinant

$$\Psi \approx \Phi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(\mathbf{x}_1) & \chi_j(\mathbf{x}_1) & \dots & \chi_k(\mathbf{x}_1) \\ \chi_i(\mathbf{x}_2) & \chi_j(\mathbf{x}_2) & \dots & \chi_k(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(\mathbf{x}_N) & \chi_j(\mathbf{x}_N) & \dots & \chi_k(\mathbf{x}_N) \end{vmatrix} := |ij\dots k\rangle$$

Normalization implicit!

→ Energy:
$$E = \sum_{i=1}^n \langle i | \hat{h} | i \rangle + \frac{1}{2} \sum_{j=1}^n \left(\langle j | \hat{J}_j | j \rangle - \langle j | \hat{K}_j | j \rangle \right) + V_{nn}$$

$$\hat{h}_i = -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^N \frac{Z_A}{r_{iA}}$$

with Coulomb-Operator $\hat{J}_i | \chi_j(2) \rangle = \langle \chi_i(1) | \frac{1}{r_{12}} | \chi_i(1) \rangle | \chi_j(2) \rangle$

and Exchange Operator $\hat{K}_i | \chi_j(2) \rangle = \langle \chi_i(1) | \frac{1}{r_{12}} | \chi_j(1) \rangle | \chi_i(2) \rangle$

→ Variation of SD \Rightarrow Hartree-Fock-equations $\hat{f}_i | \chi_i \rangle = \varepsilon_i | \chi_i \rangle$

with Fock-Operator
$$\hat{f}_i = \hat{h}_i + \sum_{j=1}^n (\hat{J}_j - \hat{K}_j)$$

3) Methods of Quantum Chemistry

b) Hartree-Fock-Theory

- Fock-Operator depends on all orbitals => non-linear equations, solved iteratively until self-consistency is reached (SCF)
- *In praxi*: use of a finite basis (LCAO) => Matrix-Eigenvalue-Equations
- Result: self-consistent energy eigenvalues and orbital eigenvectors (HF-energy expression is NOT sum of orbital eigenvalues!)

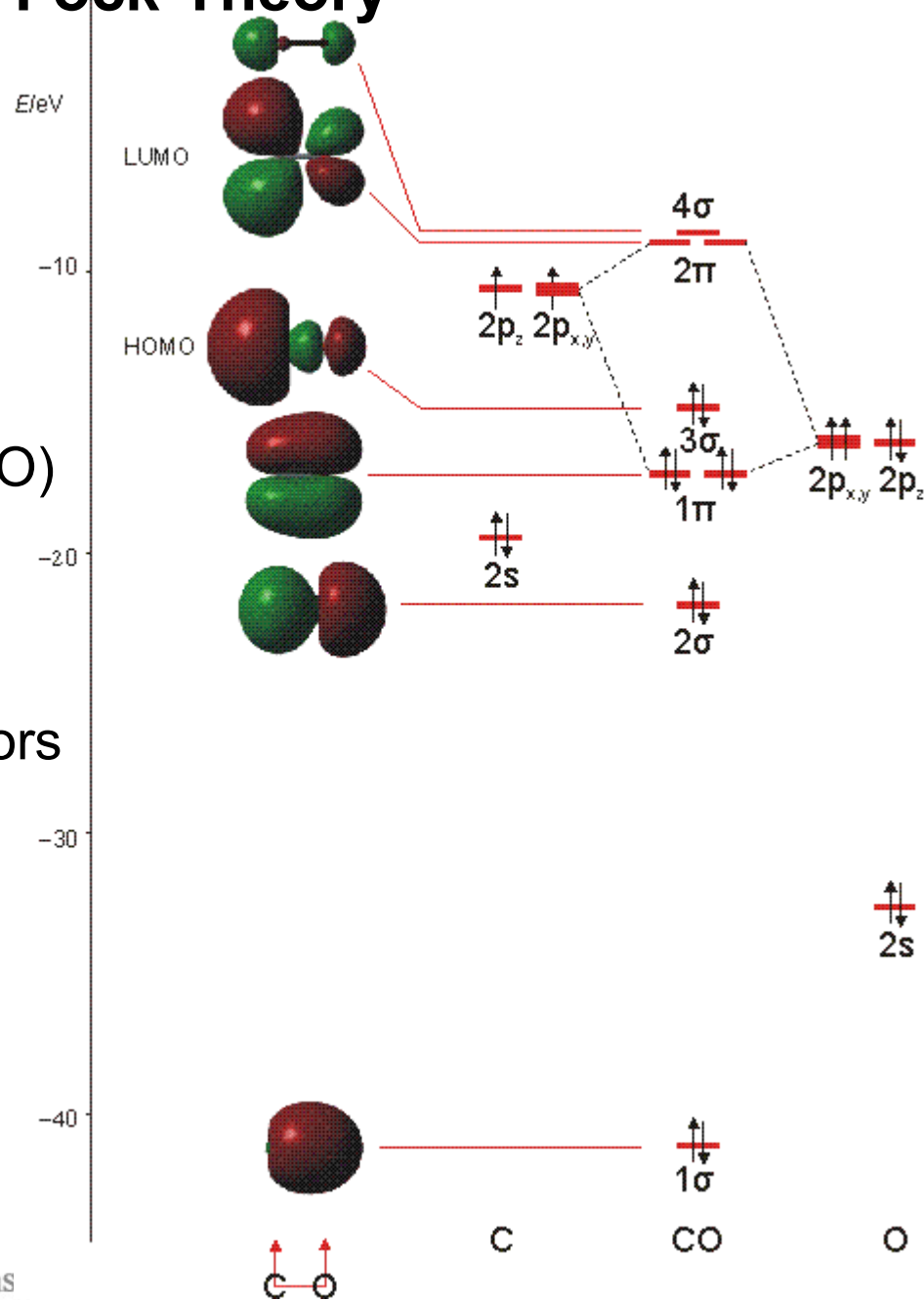
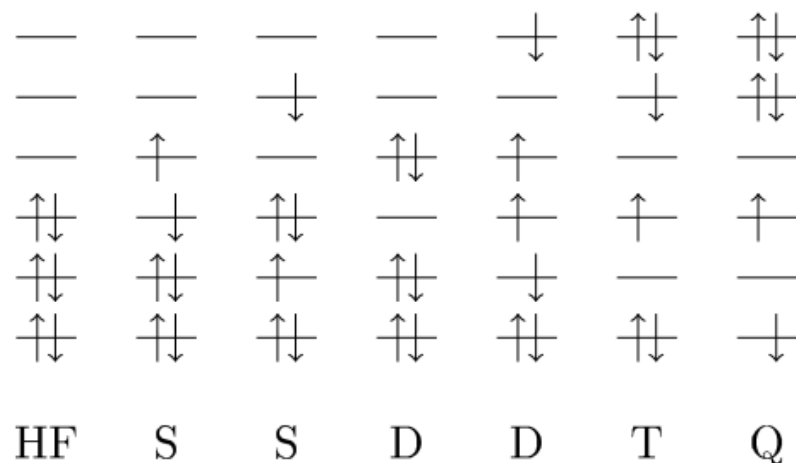


Image source: http://www.cup.uni-muenchen.de/ac/kluefers/homepage/L/kc1/co_valence_shell.html

3) Methods of Quantum Chemistry

c) Electron-Correlation-Methods

- Explicit electron correlation is missing in HF: $E_{\text{exact}} = E_{\text{HF}} + E_{\text{corr}}$
- HF-energy is optimum for a single SD => multi-determinant-ansatz
- Use unoccupied („virtual“) orbitals and excited determinants
- Expand wave-function in excited SD: $\Psi = a_0 \Phi_{\text{HF}} + \sum_{i=1} a_i \Phi_i$
- For i=single, double, etc. excitations => Configuration Interaction
- Problem: constructing all excited SD is unfeasible for all but the smallest systems, therefore: Truncated CI (e.g. CISD)
- Truncating CI expansion e.g. at doubles, other problems occur
- Other Methods: Møller-Plesset-PT, Coupled Cluster
- Accurate methods need bigger basis sets



3) Methods of Quantum Chemistry

d) Density Functional Theory (NP 1998, Kohn)

- Electron density instead of wave-function: 3 instead of 4N coordinates, but contains the entire information of the system under consideration

$$\rho(\vec{r}) = N \int \dots \int |\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)|^2 ds_1 d\vec{x}_2 \dots d\vec{x}_N$$

$$\rho(\vec{r} \rightarrow \infty) = 0 \quad \int \rho(\vec{r}) d\vec{r}_1 = N$$

- *In praxi*: use non-interacting reference system => eq. similar to HF

→ Energy:
$$E[\rho] = T[\rho] + E_{\text{Ne}}[\rho] + E_{\text{ee}}[\rho] = \underbrace{T_{\text{S}}[\rho] + E_{\text{Ne}}[\rho]}_{\text{reference system}} + \underbrace{J[\rho]}_{\text{known}} + \underbrace{E_{\text{xc}}[\rho]}_{?}$$

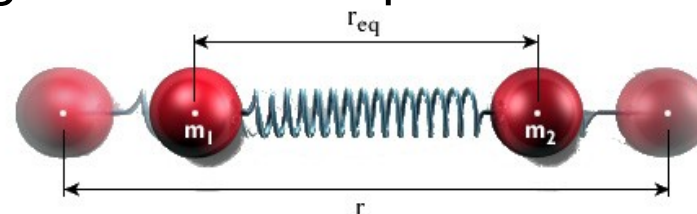
- Different implementations of exchange-correlation-functionals for E_{xc}

for different tasks (LDA, PBE, B3LYP etc.)

- No way of systematically improving functional
- Computations „cheap“, quality of result depends on specific functional
=> B3LYP optimized for organic chemistry, very good results

4) Force Field Methods (MM)

- Potential V is set up as atomic parametric function (bonding) and fitted to experimental (or ab-initio) data
- Avoid Schrödinger equation, bonding information is parametrized
- Simplified model of bonding:



„ball-and-spring“ instead of QM

- Parameters: size of atom, bond lengths and „stiffness“ etc.
- Energy mainly depends on configuration of the molecule:

$$E_{\text{FF}} = E_{\text{stretch}} + E_{\text{bend}} + E_{\text{torsion}} + E_{\text{vdWaals}} + E_{\text{el}} + E_{\text{cross}}$$

Image source:
http://neon.otago.ac.nz/chemlect/chem306/pca/IR_Raman/index.html

- Different types of force fields exist
- Absolute energy value has no meaning, only differences
- Predictions for unparametrized atoms and/or bond types difficult
- Quality depends on number of parameters and exp. values used for fits
- Significantly faster than QM methods => great use for biomolecules!

5) Overview: Capabilities & Scales

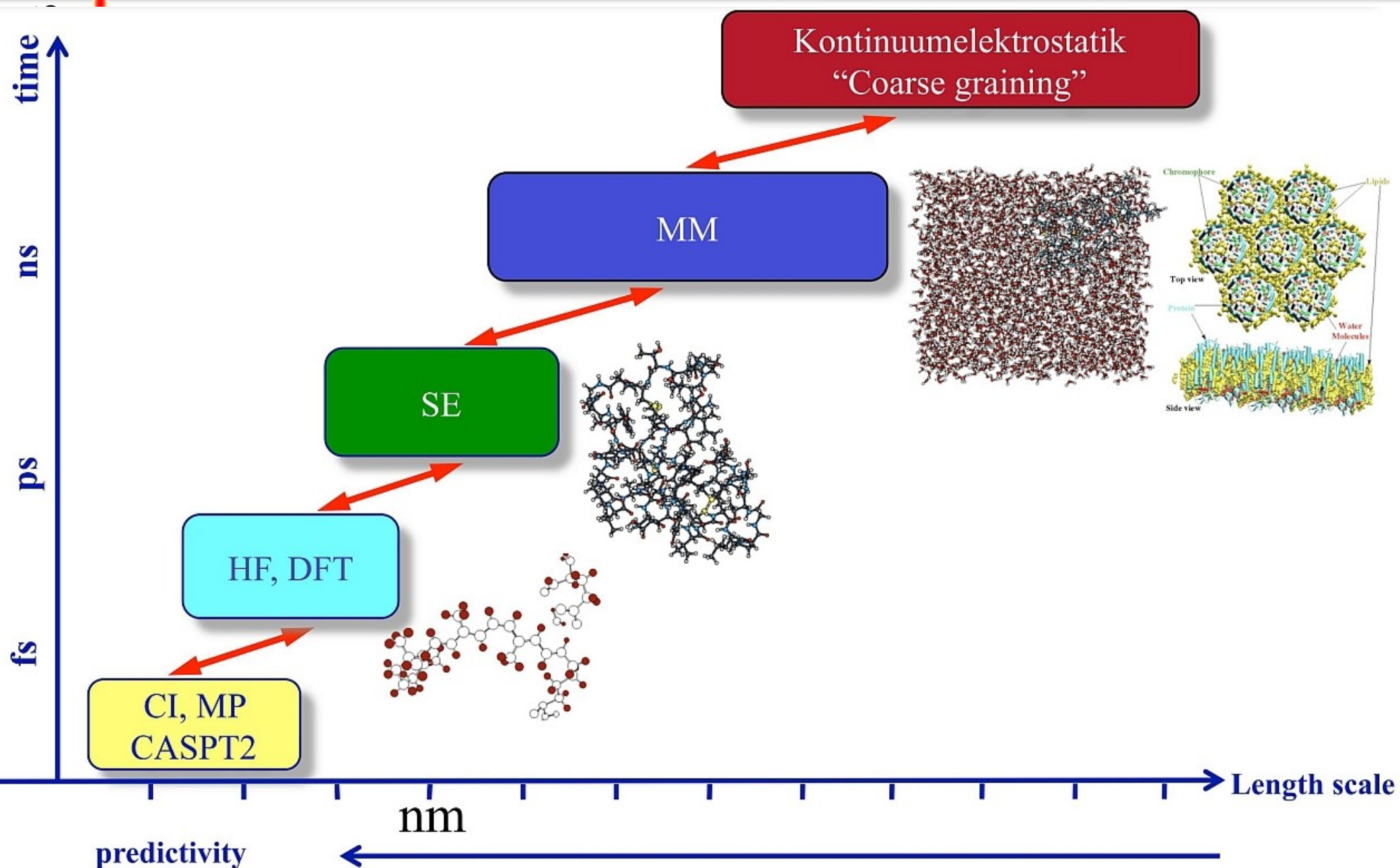
→ Ab-Initio: HF << MP2 < CISD < MP4(SDQ) ~ CCSD < MP4 < CCSD(T)

<i>Method</i>	<i>Accuracy</i>	<i>Max. number of atoms</i>	<i>Scaling</i>	<i>Simulation of mol. dynamics</i>
Molecular mechanics		->1000000	N^2	~10 μ s
Semi-empirical methods	Low	~2000	N^3	~10 ns
HF	Medium	~500	N^3 - N^4	~10 ps
DFT	Medium-Very High	~500	N^3	~10 ps
Perturbation & Variation Methods	High	~50	N^5 - N^{10}	~100 fs
Coupled Cluster	Very High	~20	N^5 - N^{10}	~100 fs

<http://www.pci.tu-bs.de/agelstner/lehre/thc1-0809/Einfuehrung.pdf>

http://vergil.chemistry.gatech.edu/courses/chem6485/pdf/Electronic_Structure_Theory.pdf

5) Overview: Capabilities & Scales



6) Hybrid Methods (QM/MM)

- Pure MM is unable to properly describe changes in chemical bonds (e.g. in electron-transfer-reactions)
- For big molecules, QM and MM methods can be combined: backbone is calculated using MM, active site using QM methods

$$\mathcal{H} = \mathcal{H}_{\text{QM}} + \mathcal{H}_{\text{MM}} + \mathcal{H}_{\text{QM/MM}}$$

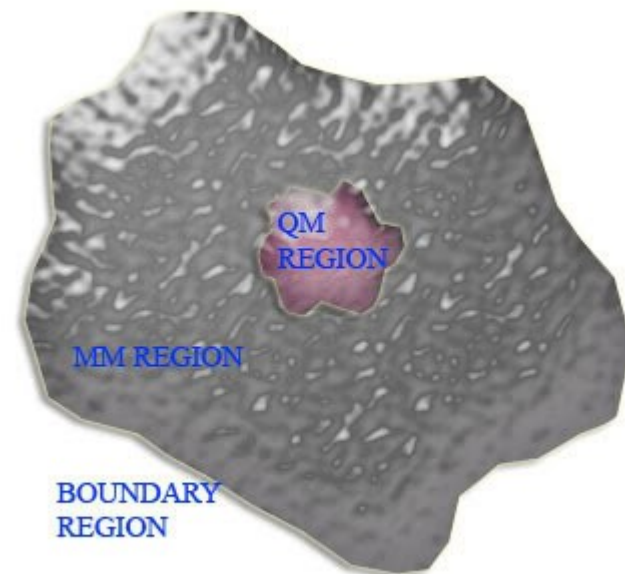


Image source: <http://unitedboinc.com/en/news/1-latest-news/88-what-is-hydrogenhome-brief>

- Challenge: describing connection between QM and MM (bond partitioning)
 - => link atom for QM but ignored in MM part or pseudopotentials
- ONIOM: high-level method for small model system, low-level for real s.
- Main Problem: Partition in QM and MM part not unique, thus no „black box“ method

1) Molecular Vibrations

→ Taylor-expansion of potential, truncated at 2nd order (mass-weighted coordinates)

$$V = V_0 + \sum_{i=1}^{3N} \left(\frac{\partial V}{\partial q_i} \right)_0 q_i + \frac{1}{2} \sum_{i,j=1}^{3N} \underbrace{\left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0}_{f_{ij}} + \dots \quad \frac{\partial^2 q}{\partial t^2} + \underbrace{k}_{F} q = 0$$

→ Stationary Points: gradient vanishes => undamped harmonic Oscillator

→ 3N cartesian coordinates => 3N-6(5, linear) vibrational normal modes

→ Often, only parts of molecule vibrate
=> typical group vibrations (Stretching, deformation, scissoring, rocking, breathing etc)

→ Experimentally:
IR-spectroscopy,
Raman spectroscopy

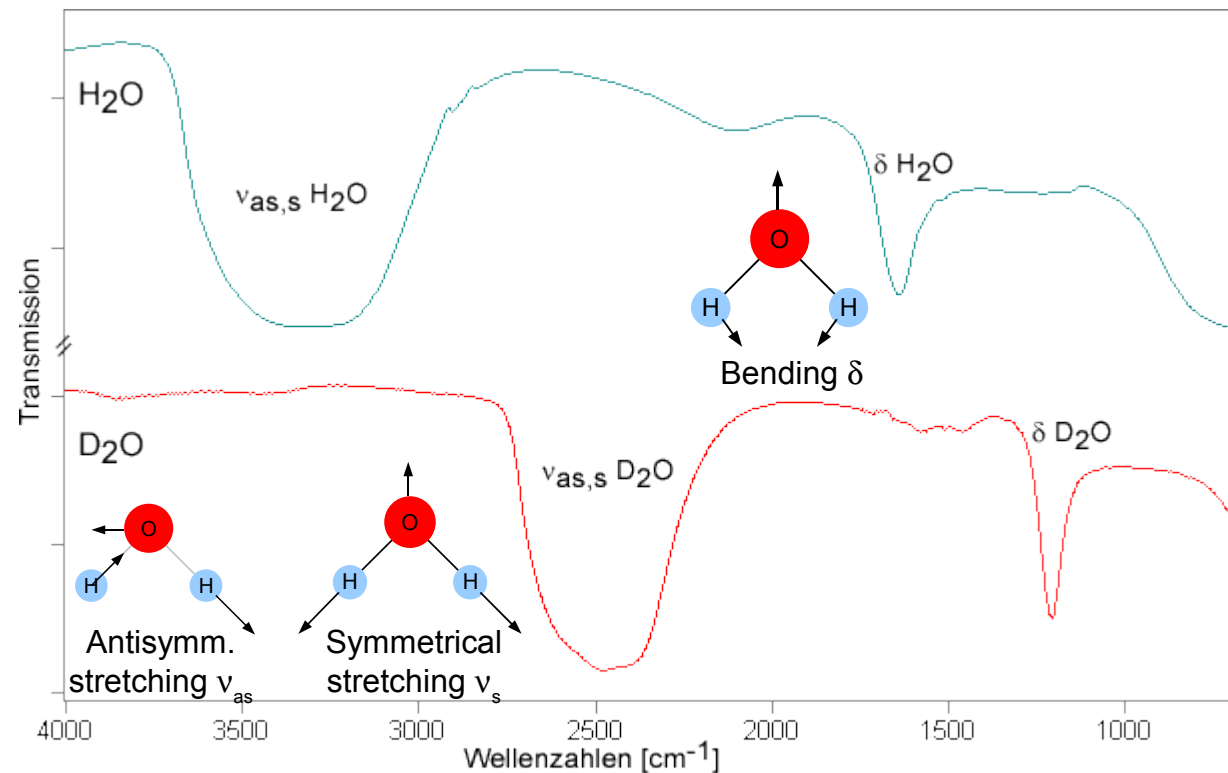
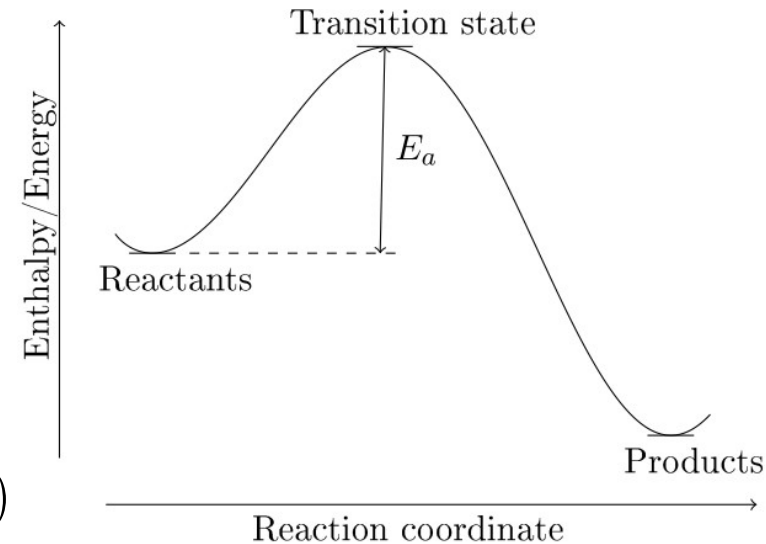


Image source: www.chemgapedia.de

2) Transition States, Reaction Rate Constants

- Reaction from reactants (R) to products (P) passes transition state (TS)
- Transition state „connects“ R to P
- On PES: 1st order saddle-point,
1 imaginary normal mode
- Reaction velocity can be expressed by
reaction rate and concentrations:

$$v_{\text{react}} = \frac{-dc(A)}{dt} = \frac{-dc(B)}{dt} = \frac{dc(P)}{dt} = k \cdot c(A)c(B)$$



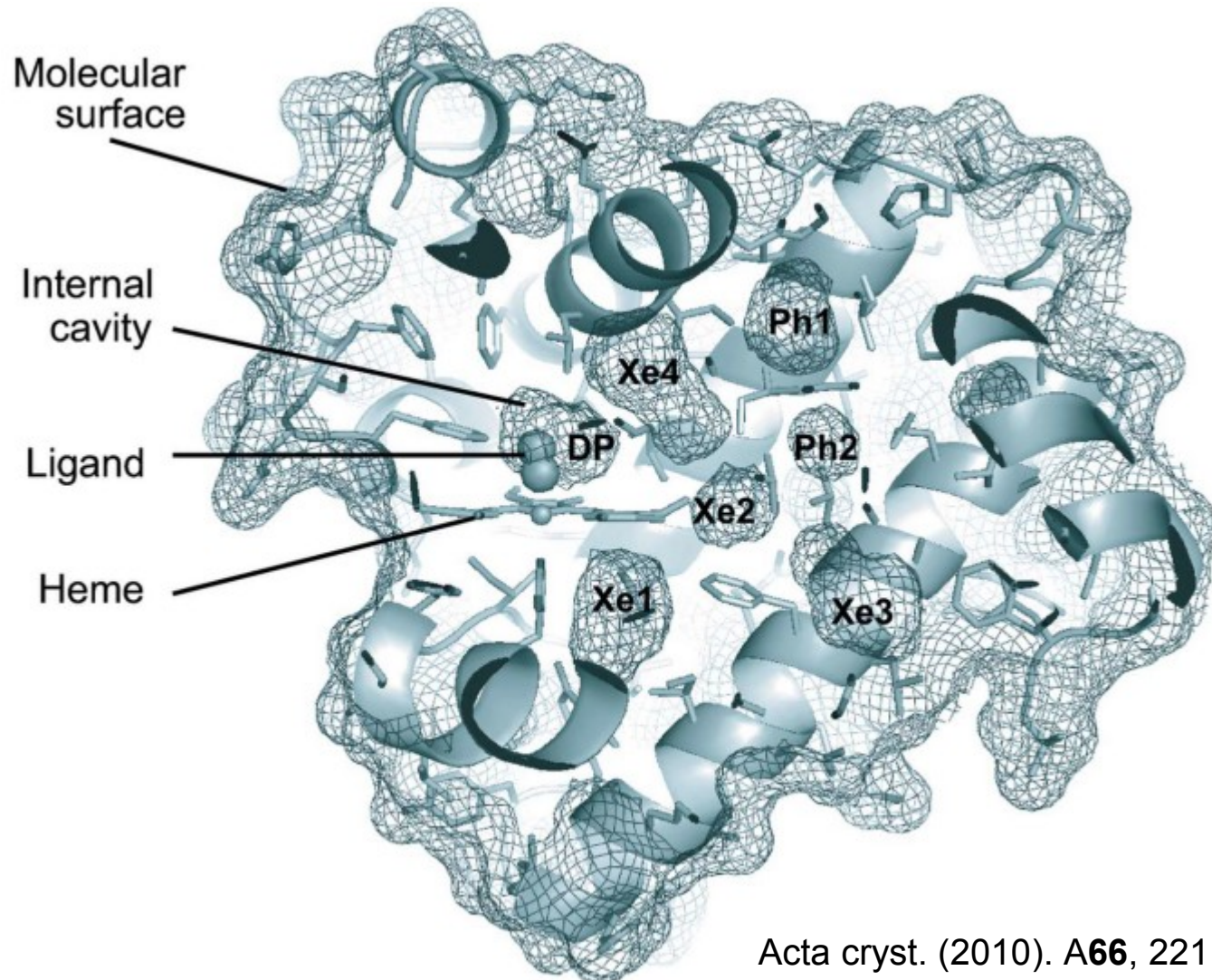
- Reaction rate determined by Energy barrier between R and TS and
temperature. Arrhenius equation (empirical): $k = A e^{\frac{-E_a}{RT}}$
- More refined model: Transition state theory
- Finding the (correct) TS is very difficult!
Experimentally: femtosecond spectroscopy
Computationally: „chemical intuition“ or algorithms (CPR, NEB,...)

3) Car-Parrinello Molecular Dynamics

- Use Lagrange mechanics, explicitly include wave-function parameters (orbitals) as additional degrees of freedom (with fictitious masses)
- Treat the entire system classically
- Nuclei and wave function evolve simultaneously in time
- Explicit wave function convergence is only needed at beginning, then iteratively updated
- Extra term in total energy arises (T_{orb}) which must be kept small
- First implemented with DFT for orbitals and plane-waves-basis, later with HF, MP2... and Gaussian basis functions
- Simulation of dynamical processes possible
- Good results due to error cancelation
- Can describe bond partitioning (MM can not)
- Small steps in time => many steps needed => only short intervals (ps)

Ligand migration in Myoglobin

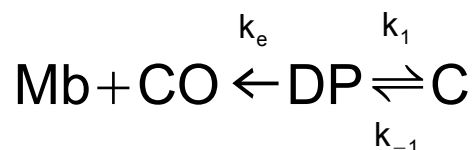
- Consider ligand migration in Myoglobin (Mb)
- No „permanent“ channel for gaseous molecules to enter or leave
- Ligand migration due to structure fluctuations
- Several „cavities“ which ligand can reside in, first identified under high Xenon pressure (Xe1-4)
- Two additional cavities (Ph1,2) found later via MD simulation



Ligand migration in Myoglobin

III. Example

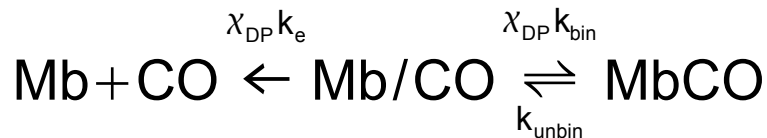
- Calculation of Heme-CO-Binding/Unbinding and migration via QM/MM: QM/MM calculation (using DFT) for binding, MD simulation for migration
- Initial structure from crystallographic data (hydrated Mb)
- 32 MD trajectories with different initial velocities
- Time step: 2fs, 15ns per trajectory, 500ns total simulation time
- Try to simplify model and to compare kinetics with experimental studies
- First result: Xe1&Xe2 only accessed from Xe3 => single proximal site Xe_p
- CO migration via $\text{Mb} + \text{CO} \leftarrow \text{DP} \rightleftharpoons \text{Xe4} \rightleftharpoons \text{Xe3} \rightleftharpoons \text{Xe}_p$
- Slowest step: $\text{DP} \rightleftharpoons \text{Xe4}$ (ns range), other steps faster
- Grouping Xe_p (Xe1&Xe2), Xe3 and Xe4 together as single site C which is in pre-equilibrium with respect to the other kinetic steps:



D'Abramo et. al, J. Phys. Chem. B., **2009**, 113, 51.

Ligand migration in Myoglobin

- Small fraction of CO escapes from protein, only from distal pocket
- CO migration: picosecond relaxation for transition along Mb cavities except DP-Xe4 exchange (nanosecond) => rate-limiting step
- Activation free energy for binding/unbinding is temperature independent
- CO distribution inside Mb is virtually instantaneously equilibrated during binding-unbinding step
- Overall scheme with χ_{DP} : probability of CO being located in DP (~0,35)



- Calculated rate constants match well the experimental data
- But: calculated results for solvated Mb are in contrast to results for crystallized Mb

D'Abramo et. al, J. Phys. Chem. B., **2009**, 113, 51.

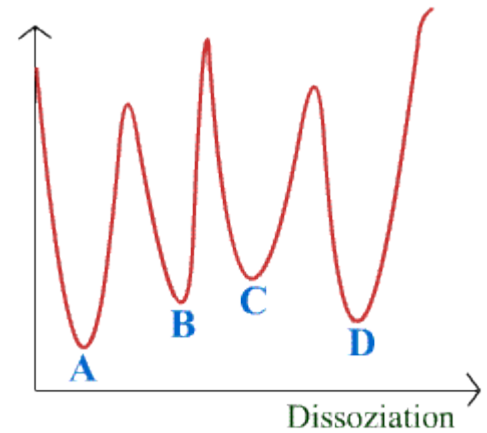
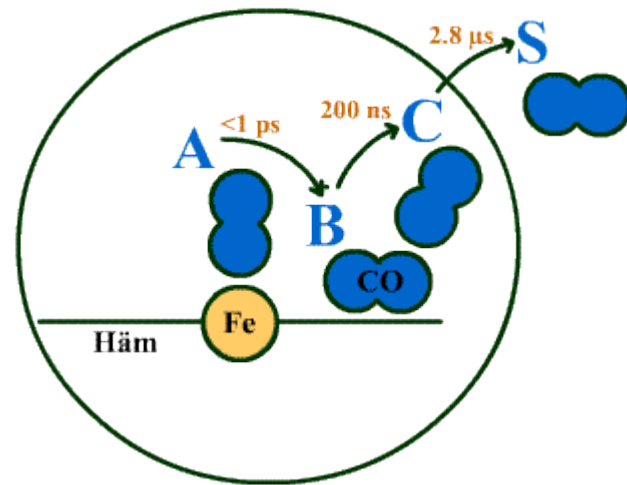
References

- (1) F. Jensen, *Introduction to Computational Chemistry*, 2nd ed., John Wiley & Sons, **2006**.
- (2) A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry*, Dover Publications, Mineaola, **1996**.
- (3) G. Adam, P. Lauger, G. Stark, *Physikalische Chemie und Biophysik*, 4. Aufl., Springer-Verlag, Berlin, **2003**.
- (4) A. Amadei, M. D'Abramo, I. Daidone, M. D'Alessandro, A. Di Nola, M. Aschi, *Theor. Chem. Acc.* **2007**, 117.
- (5) M. Anselmi, A. Di Nola, A. Amadei, *Biophys. J.* **2008**, 94, 4277.
- (6) M. D'Abramo, A. Di Nola, A. Amadei, *J. Phys. Chem. B* **2009**, 113, 16346.
- (7) A. Ostermann, R. Waschipky, F.G. Parak, G.U. Nienhaus, *Nature* **2000**, 404, 205.
- (8) H. Frauenfelder et. al, *PNAS.*, **2000**, 98, 5, 2370.
- (9) www.wikipedia.de
- (10) www.chemgapedia.de

This talk is available online: www.matthias-ernst.info/downloads_physik.html

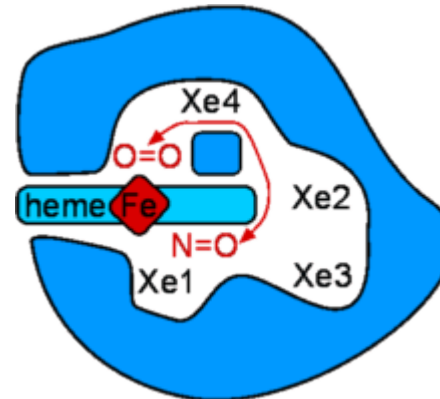
Myoglobin as catalyst

→ Pathway of CO in Mb:



→ Myoglobin may also catalyze reaction between small molecules

O₂ is bound to Heme,
fast reaction with NO
due to optimal orientation
on „entrance“



→ Reaction is pH-dependent

H. Frauenfelder et. al, PNAS., 2000, 98, 5.

Methods of Quantum Chemistry

c) Electron-Correlation-Methods: Møller-Plesset Perturbation Theory and Coupled Cluster Theory

- Perturbation theory: $\mathcal{H} = \mathcal{H}_0 + \lambda \mathcal{V}$, unperturbed solution: $\mathcal{H}_0 |\Phi_i\rangle = E_i^{(0)} |\Phi_i\rangle$
- Møller-Plesset-PT: $\mathcal{H}_0 = \sum_{i=1}^n \hat{f}_i$ (unperturbed Hamiltonian=sum over fock-operators)
- 0th order energy: sum of orbital energies, 1st order energy: HF-energy
corrections start at 2nd order (MP2, MP3, ...) but don't systematically improve results; „good“ HF reference is crucial; rather „cheap“
- Coupled Cluster: exponential ansatz => corrections to infinite order
 $|\Phi_i\rangle = e^{\hat{T}} |\Phi_0\rangle$, with cluster operator $\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_n$ is
$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2} \hat{T}^2 + \frac{1}{6} \hat{T}^3 + \dots = 1 + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2} \hat{T}_1^2) + (\hat{T}_3 + \hat{T}_2 \hat{T}_1 + \frac{1}{6} \hat{T}_1^3)$$
- Cluster Operator is truncated at certain level of excitation (CCSD etc),
additional corrections via PT: CCSD(T)
- Performs better than CI (size-extensive)
- Very accurate results, but also very expensive calculations