Multiscale Methods for the Simulation of Biological Structures and Processes

Marcus Elstner

Institute for Physical Chemistry

KIT

Biomolecules

Biopolymers

Nucleic acids: phosphate group, sugar, base

Proteins: amino acids

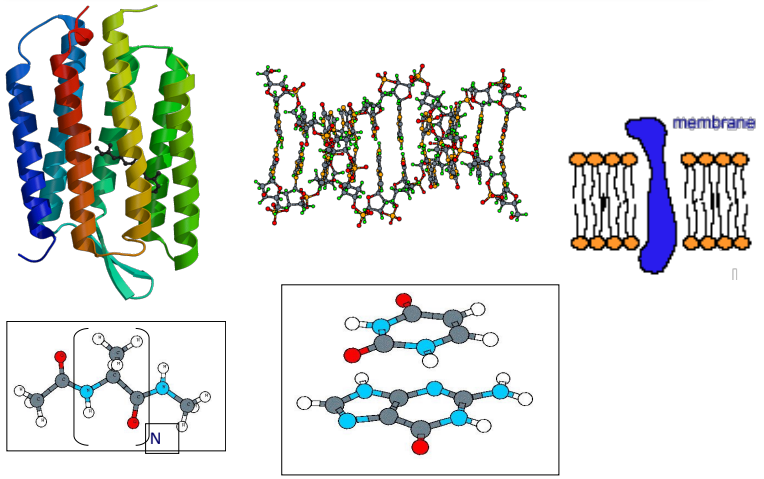
Carbohydrates: sugar units

Lipids: head group, alifatic chain

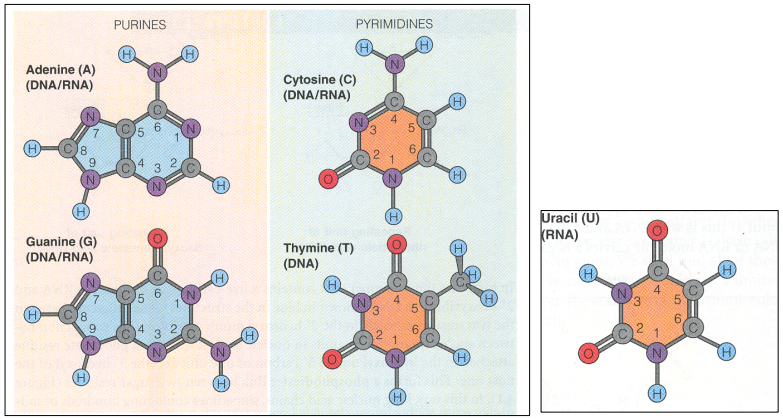
Outline

1) **Biological systems: structures**
2) **Interactions: role of entropy**
3) **Methods for different system sizes**
4) **Time scale problem**
5) **Multiscale**

Biological structures: proteins, DNA, lipids

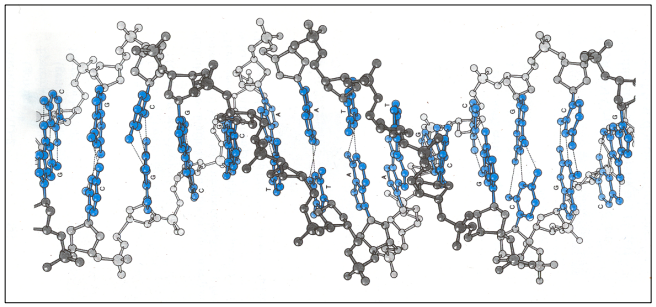


Nucleobases

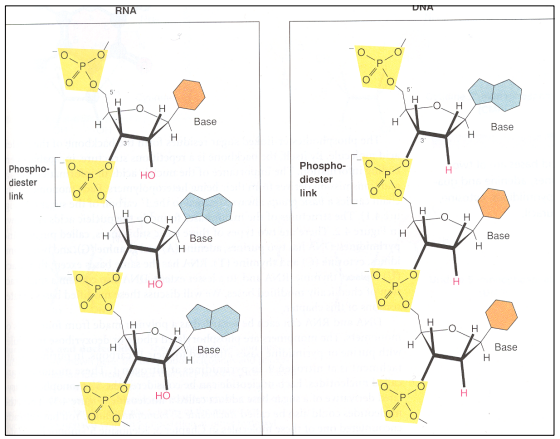


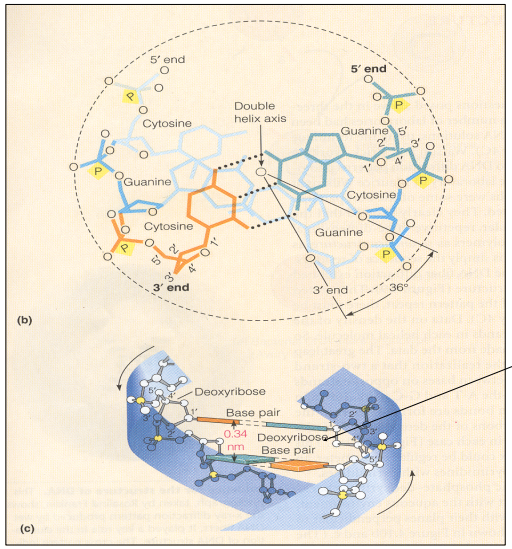
Secondary structure of DNA

- Double strand, base pairing (Watson-Crick)



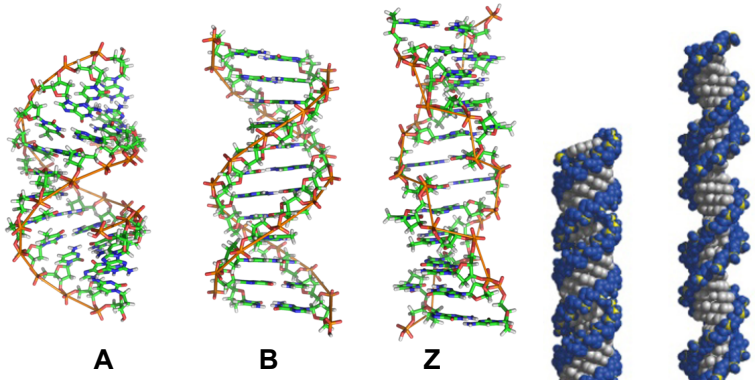
Polymers





stacking
interaction

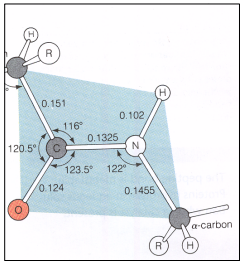
A- and B-DNA

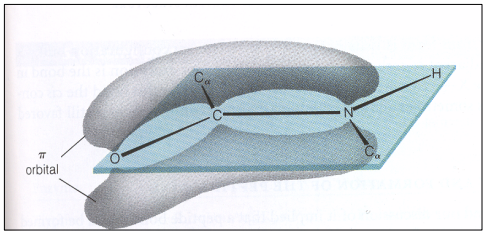


A-DNA: 11 bases per turn, 33° rotation

Z-DNA: left-handed, different steps

Peptide bond

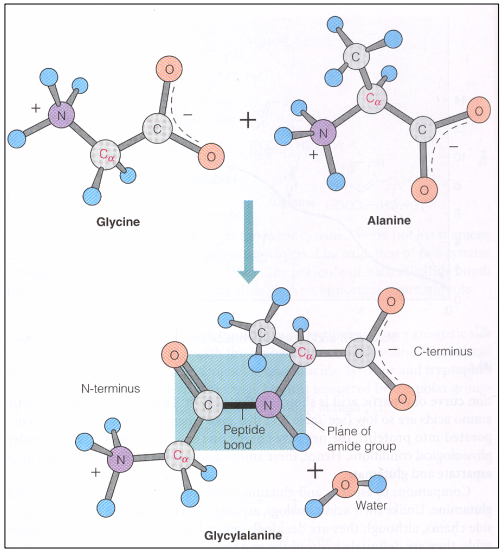




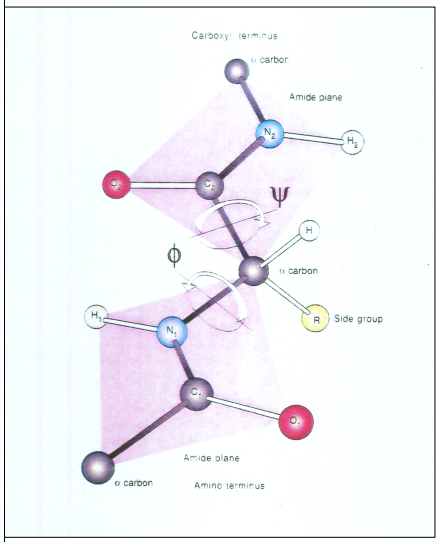
- C=N – partial double-bond character
  – barrier opposing rotation ~80 kJ/mol
- single bond ~10 kJ/mol, double bond ~200 kJ/mol

➔ only rotation around the single bonds free (C–Cα, N–Cα)

Condensation



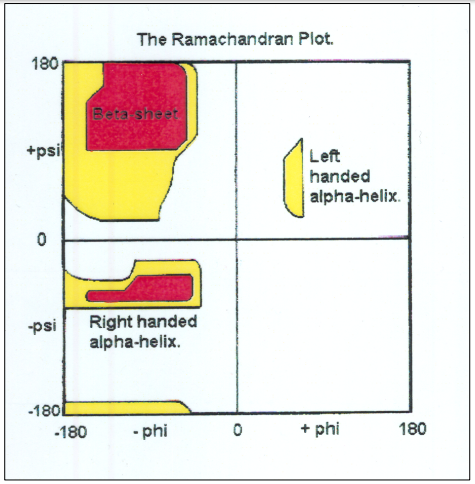
Conformation of peptides / proteins



single bonds C–Cα and N–Cα

Y und F – torsional angles around these bonds

Conformational space

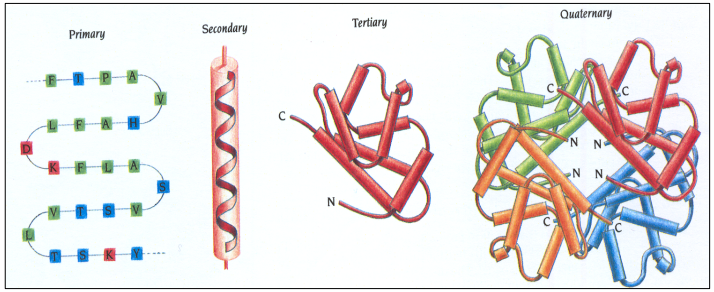


20 amino acids

size

polarity

Proteins

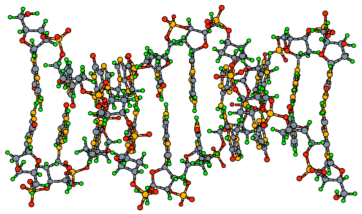


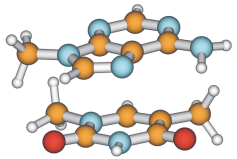
- Primary structure – sequence of amino acids
- Secondary structure – helix, sheet, turn
- Tertiary structure – 3D ordering of 2°-structure-elements

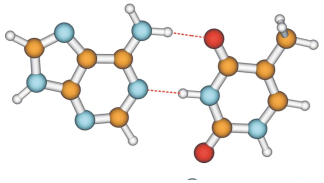
Interactions

- chemical bonds
- hydrogen bonds
- van der Waals interaction
- electrostatics
- hydrophobic (entropic) forces

DNA





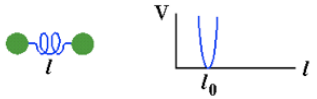


- hydrogen bonding: 0.2-0.5 eV
- stacking interactions: 0.2-1 eV

(VdW interactions: repulsion +dispersion)

(chemical bond: 4 eV)

Chemical bonds: harmonic or Morse potential



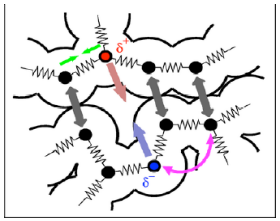
$E(l) = 0.5\ k*(l-l_0)^2$

parameter: k and l_0

equilibrium distance: l_0

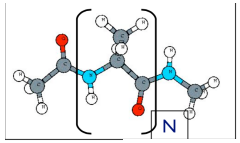
vib. frequency: $v = \sqrt{k/m}$

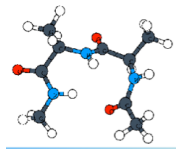
Problem: no bond breaking using harmonic oscillator model

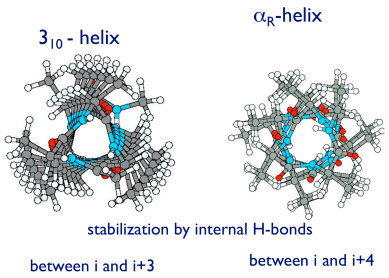


similar for angles and dihedrals

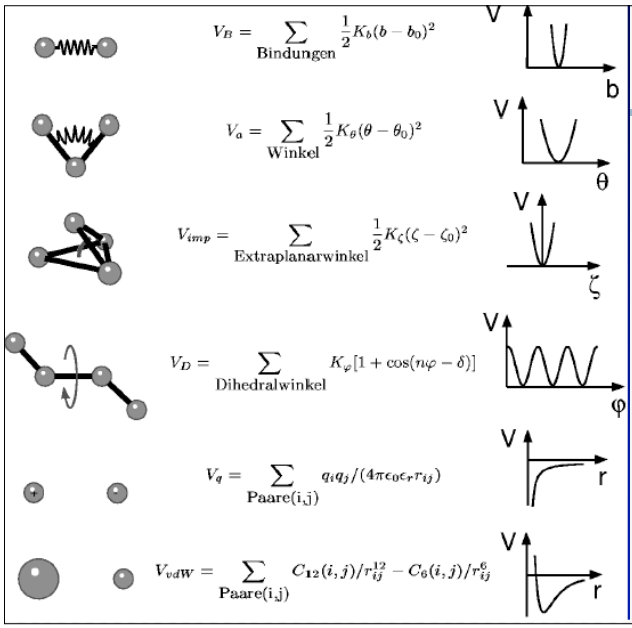
peptides





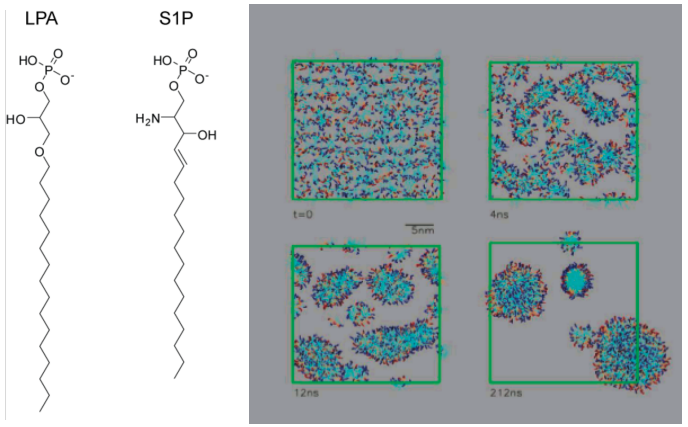


- **hydrogen bonding** important for structure: bonding motifes
- **van der Waals** interaction: energetic stabilization
- **solvation** (electrostatic): helix-dipole



Quelle: Grubmüller MPI Göttingen

lipids



what is driving force for that?

Formation of bilayers

Langmuir ***2006, 22, 998-1005***

Role of entropy

$$Q = \frac{1}{N!h^{3N}} \int exp\left(-\frac{\sum_i (p_i^2/2m_i)}{kT}\right) d\vec{p}_1, \ldots d\vec{p}_N$$
$$* \int exp\left(-\frac{V(\vec{x}_1, \ldots \vec{x}_N)}{kT}\right) d\vec{x}_1, \ldots d\vec{x}_N$$

total potential energy: sum of polymer energies

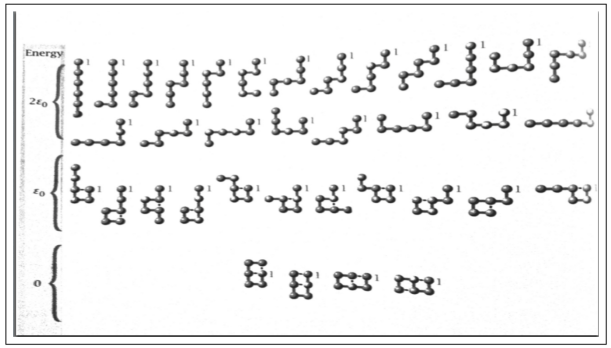
$$V(\vec{x}_1, \ldots \vec{x}_N) = \sum_i v_i.$$

for N non-interacting polymers:

$$Q = q^N = \left(\sum_i exp(-\beta v_i)\right)^N$$

$$q = 4e^{-v_0/kT} + 11e^{-v_1/kT} + 21e^{-v_2/kT}$$

Role of entropy



$v_2 = 2\epsilon$

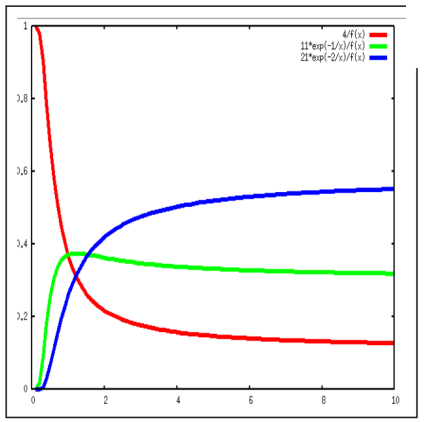
$v_1 = \epsilon$

$v_0 = 0$

number of states: $g_0 = 4$, $g_1 = 11$ und $g_2 = 21$

Role of entropy

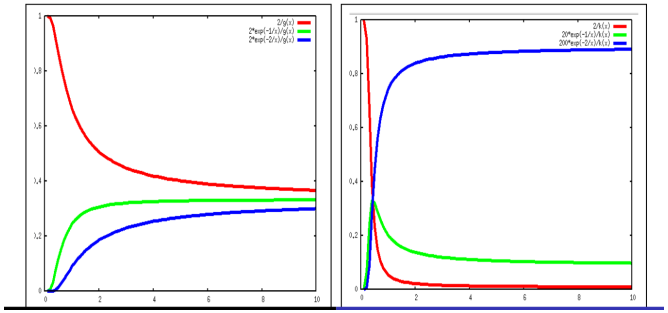
$$p_i(T) = \frac{g_i e^{-v_i/kT}}{q}$$



$$q = 4e^{-v_0/kT} + 11e^{-v_1/kT} + 21e^{-v_2/kT}$$

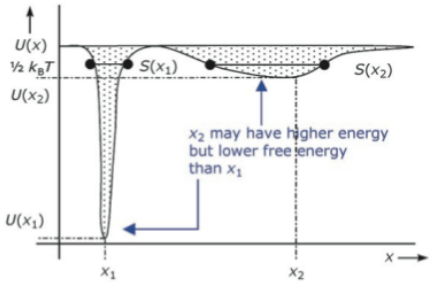
Role of entropy

p



T

$g_0 = 2,\ g_1 = 2,\ g_2 = 2$ $\qquad$ $g_0 = 4,\ g_1 = 20\ ,\ g_2 = 200$



van Gunsteren AC 2006

=> need long simulations to sample phase space

free energy difference => compute probabilities, to find a system in a certain state!

- have to overcome barrier
- definition of a state
- reaction coordinate for the transition

Role of entropy

$$lnp_i = lng_i - v_i/kT - lnq$$

$$p_i(T) = \frac{g_i e^{-v_i/kT}}{q}$$

$$-kTlnp_i - kTlnq = \mathbf{v_i - TkTlng_i = f_i}$$

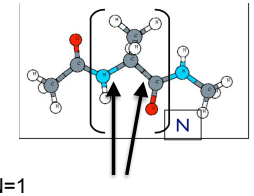
$v_i = u_i$ **energy of polymer** $\qquad$ $s_i = klng_i$ **number of states**

F = U - TS

$$\Delta f = f_i - f_j = -kTlnp_i + kTlnp_j = -kTln\frac{p_i}{p_j}$$

free energy difference of 2 states: log of populations!
from MD simulations!

Conformational space



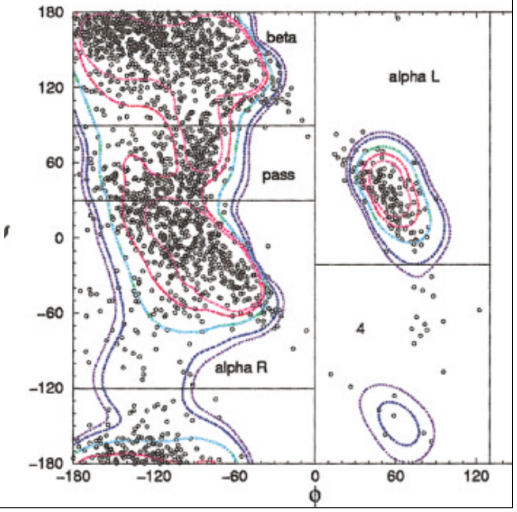
N=1

22 peptide atoms,
solvated in box with water
periodic boundary conditions

5-10 ns to converge

barrier < 0.1 eV

get free energy difference
from counting points in
the areas



Hydrophobic forces

free enthalpy of mixing: $\Delta G_{mix} = \Delta H_{mix} - T\Delta S_{mix}$

ideal mixtures: $\Delta H_{mix} = 0.$

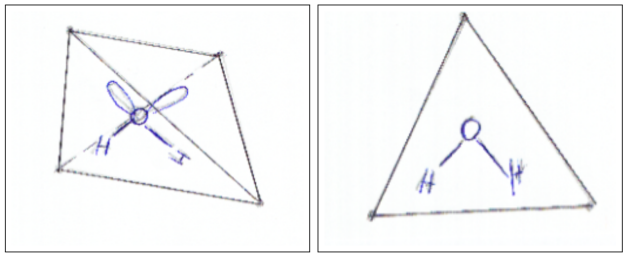
what about real mixtures of A and B?

consider interaction energies A-A, A-B and B-B:

enthalpic: if A-A and B-B are stronger bound than A-B
in this case, A and B would not mix

look at temperature dependence: entropic term is T-dependent

Hydrophobic forces: a simple model



$W_{Wasser} = 6$ $W_{hydrophob} = 3$

$T\Delta S_{hydro} = Tkln6 - Tklnln3 = kTln2$

per water molecule in contact with solute

Hydrophobic forces

consider 1 Mol benzene in solution at 300K:

assume, every benzene molecule is surrounded by 10 water molecules,
i.e. 10 Mol water molecules are conformationally hindered

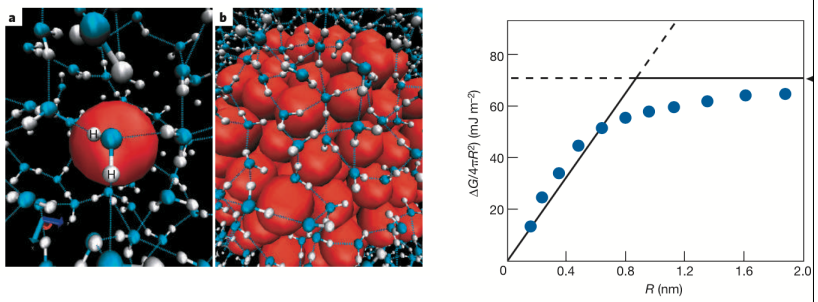
$10 * N_A T\Delta S_{hydro} = 10N_A kTln2 = 10RTln2 = 4.2kcal/Mol$

Benzol, Toluen, Ethybenzol

$\Delta H_{mix} \approx$ -0.5 – 0.5 kcal/mol,
$-T\Delta S_{mix} \approx +5kcal/mol$

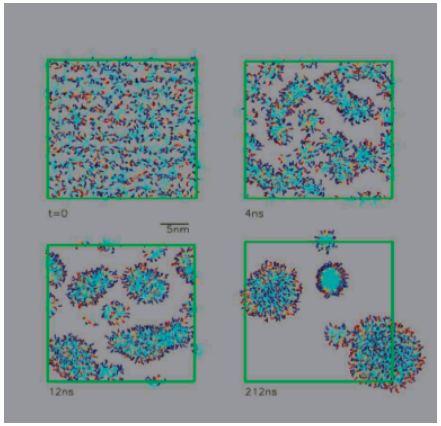
i.e. not mixing is entropic

Hydrophobic forces: size dependence



D. Chandler, Nature 2005

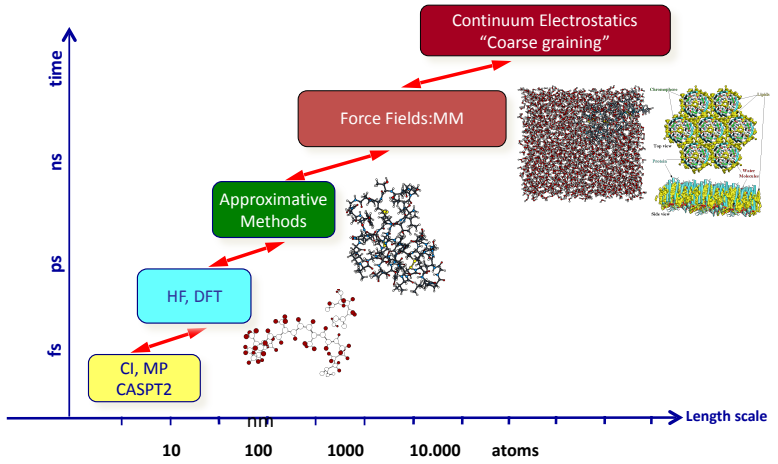
Simulation of lipids



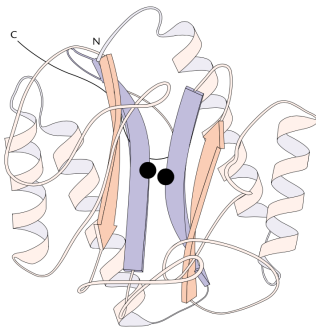
Formation of bilayers
is driven by hydrophobic forces

Langmuir ***2006, 22, 998-1005***

Multi-scale models in theoretical biophysics



Hydrophobic forces



hydrophobic residues
form a 'hydrophobic core'
inside the protein

Table 2 Hydrophobicities[a] of the 20 naturally occurring amino acids

Amino acid	Hydrophobicity (kJ mol^{-1})	(kcal mol^{-1})
Tryptophan	9.41	2.25
Phenylalanine	7.49	1.79
Isoleucine	7.53	1.80
Leucine	7.11	1.70
Cysteine	6.44	1.54
Methionine	5.14	1.23
Valine	5.10	1.22
Tyrosine	4.02	0.96
Proline	3.01	0.72
Alanine	1.30	0.31
Threonine	1.09	0.26
Glycine	0.00	0.00
Serine	–0.17	–0.04
Histidine	0.54	0.13
Glutamine	–0.92	–0.22
Asparagine	–2.51	–0.60
Glutamic acid	–2.68	–0.64
Aspartic acid	–3.22	–0.77
Lysine	–4.14	–0.99
Arginine	–4.23	–1.01

[a]The hydrophobicities are based on the solvent transfer free energies from octanol to water.

Semi-empirical methods

approximation, neglect and parametrization of integrals from HF and DFT

- integrals pre-computed, no evaluation during program runtime
- about 3 orders of magnitude faster than DFT-GGA (medium sized basis set)
- O(N^3) scaling: diagonalization of minimal basis Fock-matrix

-HF-based:

* CNDO, INDO, MNDO, AM1, PM3...PM7, MNDO/d, OM1,OM2 ...
* HF formalism, correlation effects implicit due to fitting of parameters
* CI, MRCI ... extension possible

-DFT-based:

* ***DFTB*** - models
* correlation effects explicit (GGA), problem of exchange
* TD-DFTB, GW, Bethe-Salpeter, LDA+U, SIC, range-separated functionals

DFTB is derived from DFT

inherits the problems of DFT-GGA:

- VdW interactions => empirical dispersion
- TD-DFT failures => limited use of TD-DFTB
- overpolarizability
- overbinding
- single reference method
- ...

but also the strenghts of DFT

- conceptual simplicity
- good geometries
- reasonable vib. frequencies
- ...

=> NDDO-type and DFTB methods complement each other

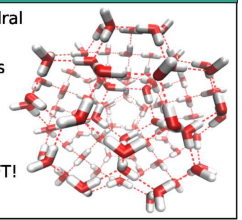
Timings: Performance of DFTB3

Theory	Basis	N_{bas}	t_{CPU}
B3LYP	def2-SVP	2400	18413.0
RI-PBE	def2-SVP	2400	2488.6
DFTB3	3OB	600	3.1

DFT calculations run with Turbomole 6.5
DFTB3 calculations run with DFTB+ 1.2
CPU: AMD Phenom™ II X6 1090T
RAM: 5 GB

Test: Icosahedral water cluster, 100 molecules

→ Up to 1000 times faster than DFT!



Approximations of DFTB2

$$E = \sum_i^{occ} \langle \Psi_i | \hat{H}^0 | \Psi_i \rangle + \frac{1}{2} \iint' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{xc}}{\delta\rho\, \delta\rho'} \right|_{n_0} \right) \Delta\rho\, \Delta\rho'.$$
$$- \frac{1}{2} \iint' \frac{\rho_0' \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0] n_0 + E_{cc}$$

- minimal basis
- neglect of crystal field and three-center terms
- initial density fixed

- second order expansion
- monopole approximation
- gamma

- two-body approximation
- fit procedure

$$E = \sum_i \sum_{\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + \sum_{\alpha\beta} U_{\alpha\beta}$$

Institute of Physical Chemistry

Density Functional Tight Binding Models

- DFTB1, DFTB2, DFTB3
- derived from a Taylor series expansion of DFT-GGA total energy functional
- reference density: sum of neutral atomic densities (spherical)

$$\rho_0 = \sum_\alpha \rho_\alpha$$

- main advantage:
  - compute Hamilton Matrix elements in DFTB1 from charge neutral dimers
  - corrections in DFTB2/DFTB3 due to charge transfer between atoms

DFTB1: Matrix elements

- Introduce LCAO **minimal basis set**:

$$\Psi_i = \sum_\mu c_\mu^i \phi_\mu$$

$$\sum_i n_i \left\langle \psi_i \left| -\frac{\nabla^2}{2} + v_{ext} + \int' \frac{\rho_0(\mathbf{r}')}{|r - r'|} d^3\mathbf{r}' + v_{xc}[\rho_0(\mathbf{r})] \right| \psi_i \right\rangle$$
$$= \sum_i \sum_{\mu\nu} c_\nu^{i*} c_\mu^i \left\langle \phi_\mu \left| \hat{H_{KS}}[\rho_0] \right| \phi_\nu \right\rangle$$
$$= \sum_i \sum_{\mu\nu} c_\nu^{i*} c_\mu^i H_{\mu\nu}[\rho_0]$$

- **2-center- approximation**:

$$\rho_0 = \rho_a + \rho_b$$

ϕ_μ located on atom a and ϕ_ν located on atom b.

Int. J. Quant Chem., **58**, 185 (1996), Phys. Rev. B **51**, 12947 (1995)

DFTB3: 3OB parameters

Table 5. Mean and Maximum Absolute Deviations for Our Modified G2/97 CHNO Test Set

property[a]	N[b]	DFTB2/MIO	MIO/calc	3OB	3OB-f	PBE[c]	B3LYP[c]
E^{at} (kcal/mol)	65	45.6	46.9	5.2	10.2	21.4	3.6
E^{at}_{max} (kcal/mol)		134.2	129.5	35.1	53.5	52.0	13.7
E^{at} (kcal/mol)[d]	65	7.4	7.1	4.8	8.3	4.2	2.4
E^{at}_{max} (kcal/mol)[d]		77.7	66.6	27.6	53.7	22.4	11.0
r (Å)	236	0.014	0.013	0.008	0.010	0.013	0.006
r_{max} (Å)		0.055	0.066	0.037	0.059	0.036	0.015
a (deg)	196	0.9	1.0	0.8	0.9	0.4	0.2
a_{max} (deg)		6.9	7.0	6.0	6.8	2.5	1.3

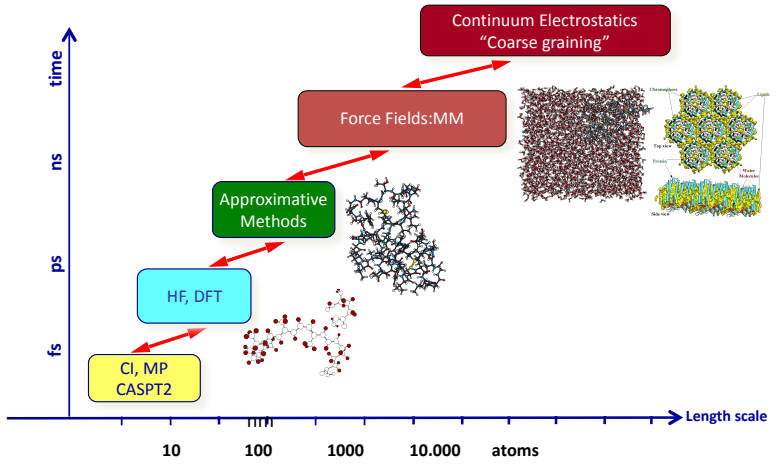
2 sets optimized:

3OB: focus on energies/geometries

3OB-f: focus on vib. freqencies/geometries

JCTC **2013**, *9*, 338

Multi-scale models in theoretical biophysics



DFTB3: 3OB parameters

reaction	G3B3	DFTB2/MIO	MIO/calc	MIO/fit	3OB	3OB-f	PBE[b]	B3LYP[b]	PBE[c]	B3LYP[c]
$H_3C{-}CH_3 + H_2 \rightarrow 2CH_4$	−18.2	+1.8	+1.0	+1.1	+1.5	−2.0	+0.3	−1.1	−0.4	−1.9
$H_2C{=}CH_2 + 2H_2 \rightarrow 2CH_4$	−56.8	−5.7	−6.6	−6.6	−1.6	−6.4	−5.5	−4.9	−2.8	−1.8
$HC{\equiv}CH + 3H_2 \rightarrow 2CH_4$	−105.1	−4.7	−5.7	−5.6	+2.7	+2.9	−13.4	−10.8	−6.8	−3.4
$C_6H_6 + 9H_2 \rightarrow 6CH_4$	−163.9	−7.0	−12.5	−12.4	−2.1	−31.7	−8.5	−13.2	−5.7	−9.3
$H_2N{-}NH_2 + H_2 \rightarrow 2NH_3$	−47.7	+4.0	−0.4	+0.3	−11.9	−11.9	+9.4	+7.6	+3.1	+1.5
$HN{=}NH + 2H_2 \rightarrow 2NH_3$	−78.6	−2.6	−10.1	−8.9	−13.8	−13.8	+12.5	+11.2	+1.2	+0.8
MAD		8.2	8.3	8.6	5.5	8.7	11.0	9.0	4.7	3.0
MAX		79.4	49.5	51.6	17.6	31.7	61.6	43.8	28.4	12.9
$H_3C{-}OH + H_2 \rightarrow CH_4 + H_2O$	−29.7	+0.4	−4.3	−4.1	+1.5	+1.6	+7.2	+6.2	+1.7	+1.0
$H_2C{=}O + 2H_2 \rightarrow CH_4 + H_2O$	−57.6	+0.3	−7.5	−7.4	+6.8	+5.5	+10.0	+10.0	+1.1	+2.0
$C{\equiv}O + 3H_2 \rightarrow CH_4 + H_2O$	−62.3	+10.3	+2.3	+2.4	+17.6	+18.0	+0.8	+4.9	−6.3	−1.0
$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$	−53.6	+3.4	−12.8	−13.2	+16.0	+13.8	+23.4	+20.1	+6.3	+5.1
$HC(=O)OH + 3H_2 \rightarrow CH_4 + 2H_2O$	−53.1	−7.1	−19.4	−19.2	+2.8	+1.1	+22.3	+19.0	+6.7	+4.7
$N_2O + 4H_2 \rightarrow 2NH_3 + H_2O$	−115.5	+45.9								
$HNO_3 + 3H_2 \rightarrow NH_3 + 2H_2O$	−119.9	+32.7								
$HNO_2 + 4H_2 \rightarrow NH_3 + 3H_2O$	−165.7	+79.4								
$H_2C{=}CH_2 + H_2 \rightarrow H_3C{-}CH_3$	−38.5	−7.5								
$HC{\equiv}CH + H_2 \rightarrow H_2C{=}CH_2$	−48.3	+1.0								
$C_6H_6 + 6H_2 \rightarrow 3H_2C{-}CH_2$	−109.2	−12.5								
$HN{=}NH + H_2 \rightarrow H_2N{-}NH_2$	−30.9	−6.6								
$N_2 + H_2 \rightarrow HN{=}NH$	42.3	+3.7								
$^1O_2 + H_2 \rightarrow HO{-}OH$	−66.3	+6.8								
$H_2C{=}NH + H_2 \rightarrow H_3C{-}NH_2$	−32.1	−1.2								
$HCN + H_2 \rightarrow H_2C{=}NH$	−15.3	+5.7								
$H_2C{=}O + H_2 \rightarrow H_3C{-}OH$	−27.9	−0.1	−3.3	−3.3	+5.3	+3.9	+2.8	+3.8	−0.6	+1.0
$HC(=O)OH + H_2 \rightarrow H_2C{=}O + H_2O$	4.5	−7.4	−11.9	−11.8	−4.0	−4.4	+12.3	+9.1	+5.5	+2.7
$HC(=O)OH + 2H_2 \rightarrow H_3C{-}OH + H_2O$	−23.4	−7.5	−15.1	−15.1	+1.3	−0.4	+15.2	+12.8	+5.0	+3.7
$HNO_3 + H_2 \rightarrow HNO_2 + H_2O$	−45.8	+46.6	+32.8	+34.2	+13.2	+13.2	+20.8	+13.4	+11.9	+5.1
$HN(CH_3)_2 + H_2 \rightarrow H_3C{-}NH_2 + CH_4$	−22.1	−1.5	−3.7	−3.5	−2.8	+6.4	+2.8	+1.2	−0.4	−1.9
$N(CH_3)_3 + H_2 \rightarrow HN(CH_3)_2 + CH_4$	−19.0	−3.2	−5.0	−4.9	−5.2	+4.1	+1.2	−0.7	−1.8	−3.6
$O(CH_3)_2 + H_2 \rightarrow H_3C{-}OH + CH_4$	−24.6	−1.4	−5.2	−5.1	−1.7	−1.5	+5.5	+4.4	+0.2	−0.9
$HC(=O)OCH_3 + H_2 \rightarrow HC(=O)OH + CH_4$	−24.9	−2.4	−5.2	−5.1	−1.2	−0.9	+3.4	+2.7	−1.3	−1.9
$HC(=O)OCH_3 + H_2 \rightarrow H_3C{=}O + H_2C{-}OH$	9.3	−10.2	−12.8	−12.7	−6.7	−6.9	+8.6	+5.6	+2.6	−0.2
$H_2C{=}CH_2 + CH_4 \rightarrow CH_3{-}CH_2{-}CH_3$	−22.6	−7.7	−7.0	−7.1	−2.3	−0.1	−5.2	−1.5	−1.1	+3.2
$HC{\equiv}CH + CH_4 \rightarrow CH_3{=}CH{-}CH_3$	−35.0	−0.9	−0.1	−0.2	+3.6	+12.5	−8.9	−4.9	−4.0	+0.4
$H_2C{-}NH_2 + CH_4 \rightarrow H_2C{-}CH_3 + NH_3$	−8.0	−0.7	−2.6	−2.4	−1.3	+11.3	+3.5	+3.7	+0.8	+1.2

DZP basis set: often used for large molecules or in QM/MM studies

JCTC **2013**, *9*, 338

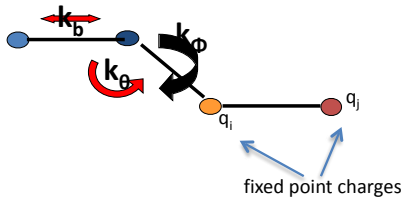
Molecular Mechanics (MM)

$$V = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} \sum_{n=1}^{N} k_\phi^{(n)} [1 + \cos((n\phi - \delta))] + \sum_{impropers} k_\omega (\omega - \omega_0)^2$$

$$+ \sum_{i,j} 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right] + \sum_{i,j} \left(\frac{q_i q_j}{D r_{ij}} \right)$$

Shortcomings.:

- Polarization
- Charge transfer
- no reactions!



Molecular Mechanics (MM)

Systematic Validation of Protein Force Fields against Experimental Data

-PLosOne 2012

Kresten Lindorff-Larsen[1☯], Paul Maragakis[1☯], Stefano Piana[1☯], Michael P. Eastwood[1], Ron O. Dror[1], David E. Shaw[1,2*]

of folded proteins. Second, we quantified potential biases towards different secondary structure types by comparing experimental and simulation data for small peptides that preferentially populate either helical or sheet-like structures. Third, we tested the force fields' abilities to fold two small proteins—one α-helical, the other with β-sheet structure. The results suggest that force fields have improved over time, and that the most recent versions, while not perfect, provide an accurate description of many structural and dynamical properties of proteins.

Refinement of protein structure homology models via long, all-atom molecular dynamics simulations

Alpan Raval,[1] Stefano Piana,[1*] Michael P. Eastwood,[1] Ron O. Dror,[1] and David E. Shaw[1,2*]

diverse set of fast-folding proteins. In MD simulations of 24 proteins chosen from the refinement category of recent Critical Assessment of Structure Prediction (CASP) experiments, we find that in most cases, simulations initiated from homology models drift away from the native structure. Comparison with simulations initiated from the native structure suggests that force field accuracy is the primary factor limiting MD-based refinement. This problem can be mitigated to some extent by

-Proteins 2012

Molecular Mechanics (MM)

would ab initio be better?

DFT-GGA is missing VdW: has to be added

what is the accuracy needed?

- proteins are marginally stable (40 kJ/mole for 100 AA protein)
- ab initio error per H-bond/VdW contact: 1-2 kJ/mole
- assume 20-30 contacts

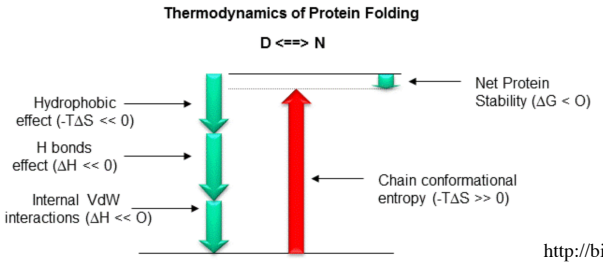
=> ????????

Molecular Mechanics (MM)

MM problems

- dihedral angles are an issue in MM (alpha-helix vs. beta sheets)
- hydrogen bonds: directionality
- interaction of charged residues
- but non-bonding interactions?

MM are fitted to thermodynamic data of organic liquids!



http://biowiki.ucdavis.edu/

Explicit Polarization Models

- fluctuating (point) charge models (FQ)
  - QM SCF → ρ/point charges/multipoles
  - Chemical hardness models (e.g. SCC-DFTB, CHARMM-FQ)

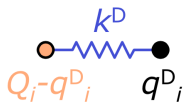
$$E^{\mathrm{ES}}(Q) = -\sum_i \mu_i Q_i + \tfrac{1}{2}\sum_{ij} \eta_{ij}(R_{ij},\eta_i,\eta_j) Q_i Q_j$$

- induced (atomic) dipole models
  - additive
  - interactive

$$\mu_i^{\mathrm{ind}} = \alpha_i \xi_i \left(M, \underline{\mu^{\mathrm{ind}}}\right)$$

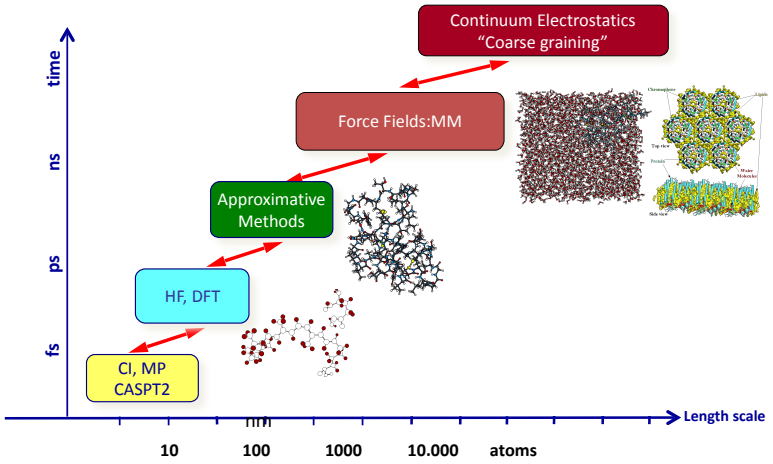
$$E^{\mathrm{ES}}(M,\mu^{\mathrm{ind}}) = -\tfrac{1}{2}\sum_{i\neq j} M_i T_{ij} M_j - \sum_{i\neq j} \mu_i^{\mathrm{ind}} T_{ij}\left(M_j + \tfrac{1}{2}\underline{\mu_j^{\mathrm{ind}}}\right) + \tfrac{1}{2}\sum_i \frac{1}{\alpha_i}\left|\mu_i^{\mathrm{ind}}\right|^2$$

- Drude oscillator model

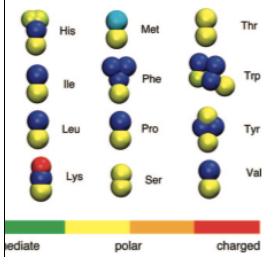


$$\alpha = \frac{q_D^2}{k_D}$$

Multi-scale models in theoretical biophysics



Coarse Graining (CG)



ained representation of all amino acids.
sent different particle types.

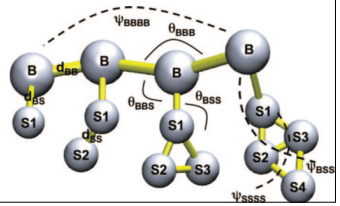
-bead with 4 heavy atoms (4 water model)

-bonding interactions

$$V_{\text{bond}}(R) = \frac{1}{2} K_{\text{bond}}(R - R_{\text{bond}})^2$$

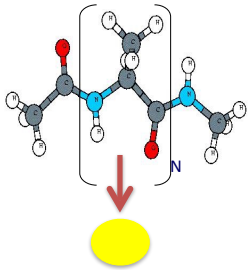
$$V_{\text{angle}}(\theta) = \frac{1}{2} K_{\text{angle}}\{\cos(\theta) - \cos(\theta_0)\}^2$$

-parametrize wrt crystal structural data



Martini force field, Monticelli et al. JCTC 4 (2008) 8

Coarse Graining (CG)



Martini force field, Monticelli et al. JCTC 4 (2008) 819

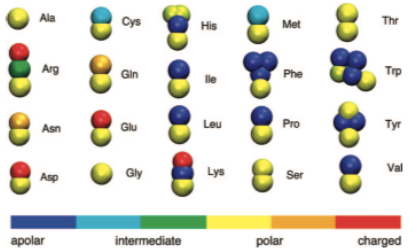
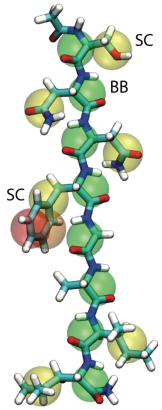
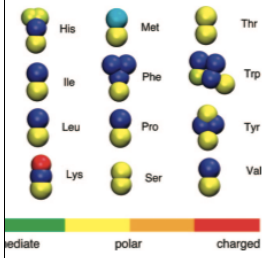


Figure 1. Coarse-grained representation of all amino acids. Different colors represent different particle types.



Coarse Graining (CG)



ained representation of all amino acids.
sent different particle types.

-bead with 4 heavy atoms (4 water model)

-non- bonding interactions

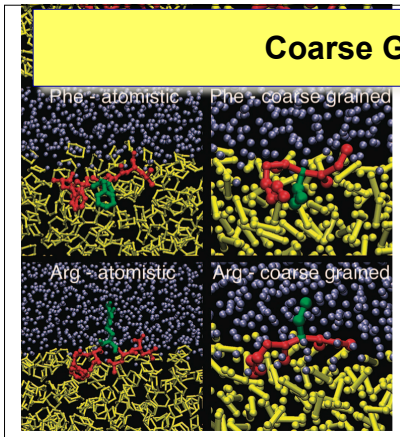
$$U_{\text{LJ}}(r) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right]$$

$$U_{\text{el}}(r) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r}$$

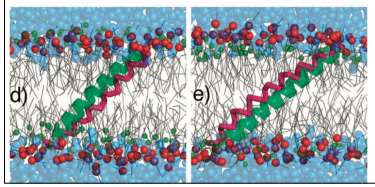
-effective dielectric (15)
-effective LJ potentials: include e.g. also hydrophobic effects, i.e. are effective free energy potentials

parametrize wrt. thermodynamic properties

Coarse Graining (CG)



- compare with all-atom simulations
- time step: 25 fs
- effective factor 4 (100fs)



Biophysics

Table 9: History and extrapolated future of computer simulations of molecular dynamics. The future is deduced from extrapolation based on an observed increase of computing speed of a factor 10 every 5 years over the past decades (see Figure 31).

Year	Molecular system (type, size)	Length of the simulation [s]
1957	first molecular dynamics simulation (hard discs)	
1964	atomic liquid (argon)	10^{-11}
1971	molecular liquid (water)	5×10^{-12}
1977	protein in a vacuum	2×10^{-11}
1983	protein in water	2×10^{-11}
1989	protein–DNA complex in water	10^{-10}
1997	polypeptide folding in solvent	10^{-7}
2001	micelle formation	10^{-7}
200x	folding of a small protein	10^{-3}

van Gunsteren AC 2006

Molecular Dynamics: MD

$$V = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} \sum_{n=1}^{N} k_\phi^{(n)} [1 + \cos((n\phi - \delta))] + \sum_{impropers} k_\omega (\omega - \omega_0)^2$$

$$+ \sum_{i,j} 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right] + \sum_{i,j} \left(\frac{q_i q_j}{D r_{ij}} \right)$$

Molecular Dynamics (MD):

Numerical integration of Newtons equation of motion F=m*a with

timestep: ~1fs

➔ trajectories

➔ 1ps dynamics: 1000 force evaluations

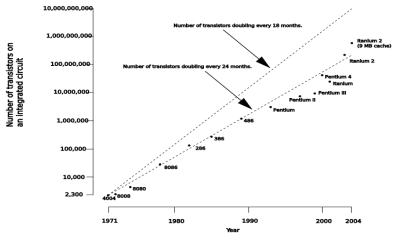
$$F_i = -\frac{\partial V}{\partial R_i}$$

And the future ...

van Gunsteren AC 2006

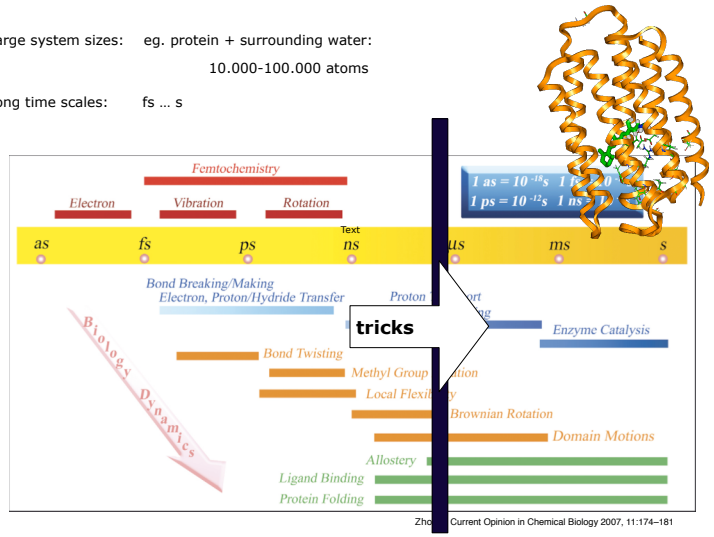
2001	biomolecules in water (ca. 10^4 atoms)	10^{-8}
2029	biomolecules in water (folding sooner?)	10^{-3}
2034	*E. coli* bacteria (ca. 10^{11} atoms)	10^{-9}
2056	mammalian cell (ca. 10^{15} atoms)	10^{-9}
2080	biomolecules in water (as fast as nature)	10^{6}
2172	human body (ca. 10^{27} atoms)	1

Moore's Law



The Challenge: size and time-scales

- large system sizes: eg. protein + surrounding water:
  10.000-100.000 atoms
- long time scales: fs … s



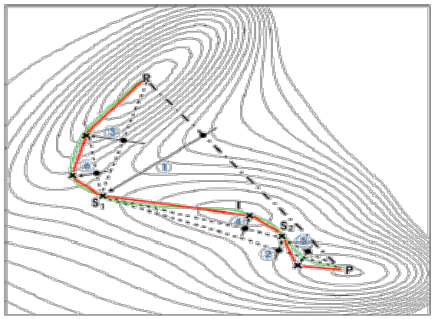
Zho… Current Opinion in Chemical Biology 2007, 11:174–181

How to study reactions and (rare) dynamical events

- reaction path methods
  - NEB (nudged elastic band, Jonsson)
  - CPR (conjugate peak refinement, Fischer, Karplus)
  - dimer method (Jonsson)
- free energy sampling techniques
  - umbrella sampling
  - free energy perturbation
  - transition path sampling
- accelerated MD
  - hyperdynamics (Voter)
  - chemical flooding (Grubmüller)
  - metadynamics (Parinello)
- replica exchange

Reaction path methods: e.g. CPR

- CPR: Conjugate peak refinement (Fischer & Karplus)
- NEB: nudged elastic band
- dimer method

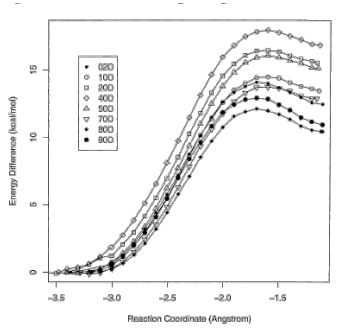


CPR: fig. from S. Fischer

These methods are very 'costly', i.e. they require the calculation of energy and forces several 1.000-10.000 times
=> prohibitive for DFT/ab initio methods

'Problem' of potential energy (MEP)

Different energy profiles for different protein conformations



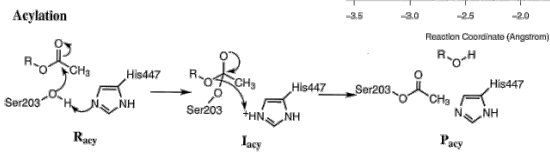


Figure 1. Acylation reaction mechanism of acetylcholine catalyzed by AChE.

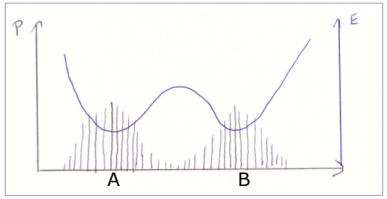
Zhang et al JPCB 107 (2003) 44459

Calculate potential of mean force

Sample the states A and B in MD:

Free energy is calculated from probabilities:

$$F_B - F_A = -kTln\frac{P(q_B)}{P(q_A)}$$

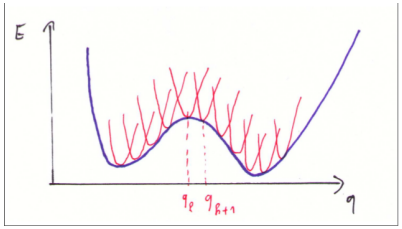


When the barrier is to high, force the system to cross it with additional potentials:

Subtract these afterwords:

'Umbrella sampling'

- need reaction coordinate



How to study reactions and (rare) dynamical events

- accelerated MD
  - hyperdynamics (Voter)
  - chemical flooding (Grubmüller)
  - metadynamics (Parinello)
  - replica exchange

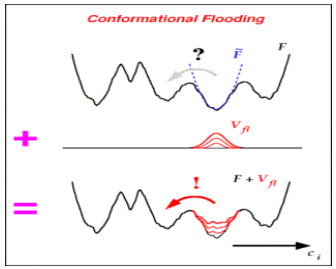
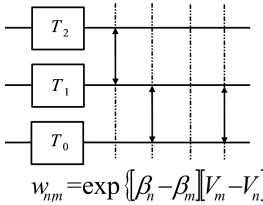


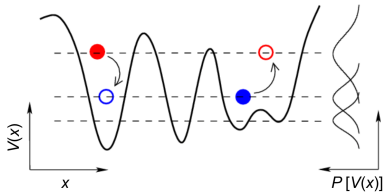
Fig. from H. Grubmüller

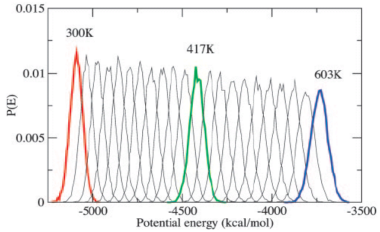
Sugita and Okamoto. *Chem. Phys. Lett.* 1999(314), pp. 141-151

replica exchange



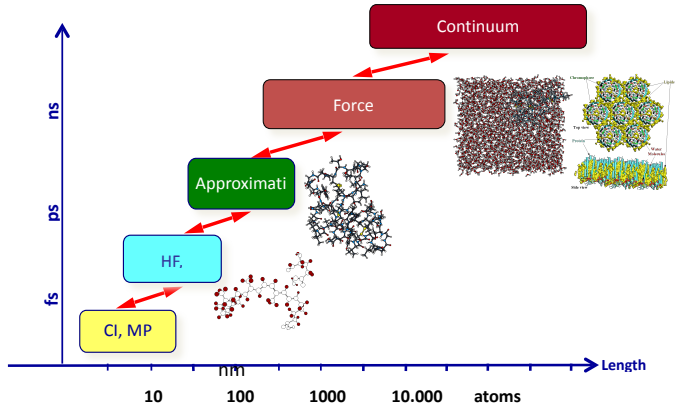
$$w_{nm} = \exp\{[\beta_n - \beta_m][V_m - V_n]\}$$





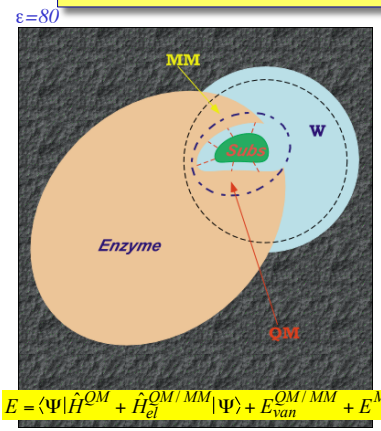
- preserves canonical ensemble
- higher T samples faster

Berne *et al.*, PNAS **102**, 13749 (2005)



- hierarchical modeling: process information (e.g. bottom up parametrization)
- hybride modeling: combine methods

Combined QM/MM



$$E = \langle \Psi | \hat{H}^{QM} + \hat{H}_{el}^{QM/MM} | \Psi \rangle + E_{van}^{QM/MM} + E^{MM}$$

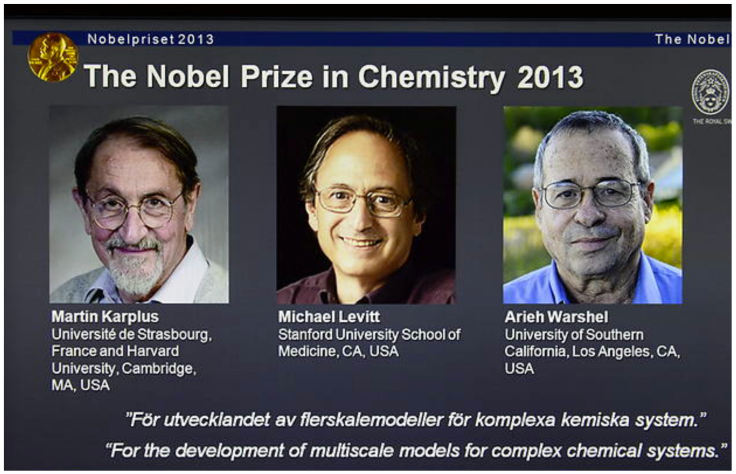
Quantum mechanical (QM)
- Bond breaking/formation
- Computationally demanding
  - DFT, AI: ~ 50 atoms
  - Semi-Empirical: ~$10^{2\text{-}3}$ atoms

Molecular mechanical (MM)
- Computationally efficient
  - ~$10^{3\text{-}5}$ atoms
- Generally for structural properties

Combined QM/MM
- Chemical Rx in macromolecules
- DFT (AI) /MM: Reaction path
- Semi-Empirical/MM: Potential of mean force, rate constants

- No polarization of MM region!
- No charge transfer between QM and MM

Combining methods



Combined QM/MM

1976 Warshel and Levitt

1986 Singh and Kollman

1990 Field, Bash and Karplus

QM
- Semi-empirical
- quantum chemistry packages: DFT, HF, MP2, LMP2
- DFT plane wave codes: CPMD

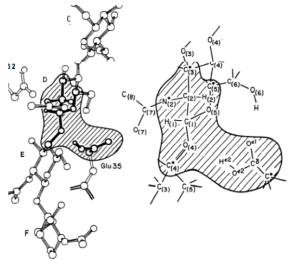
MM
- CHARMM, AMBER, GROMOS, SIGMA,TINKER, ...

Quantum Mechanics/Molecular Mechanics (QM/MM)

Theoretical Studies of Enzymic Reactions:
Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme

A. Warshel and M. Levitt
Medical Research Council Laboratory of Molecular Biology Hills Road, Cambridge CB2 2QH, England
and
Department of Chemical Physics The Weizmann Institute of Science Rehovot, Israel

J. Mol. Biol. (1976) **103**, 227–249

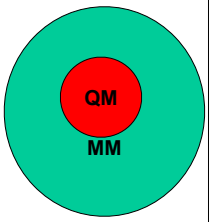


$$V = V_{\text{classical}} + V_{\text{quantum}} + V_{\text{quantum/classical}}$$

Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**.

QM/MM Methods

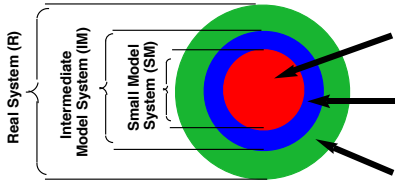
- Mechanical embedding: only steric effects
- Electrostatic embedding: polarization of QM due to MM
- Electrostatic embedding + polarizable MM



- **subtractive**: several layers: QM-MM
  doublecounting on the regions is subtracted
- **additive**: different methods in different regions +
  interaction between the regions

EMORY UNIVERSITY

The ONIOM Method (an ONION-like method)

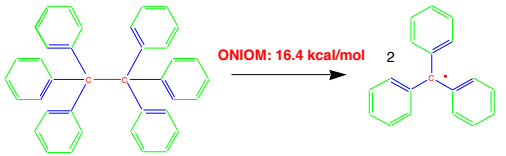


First Layer
Bond-formation/breaking takes place.
Use the "High level" (H) method.

Second Layer
Electronic effect on the first layer.
Use the "Medium level" (M) method.

Third Layer
Environmental effects on the first layer.
Use the "Low level" (L) method.

Example: The binding energy of ϕ_3C-C ϕ_3 (HPE)



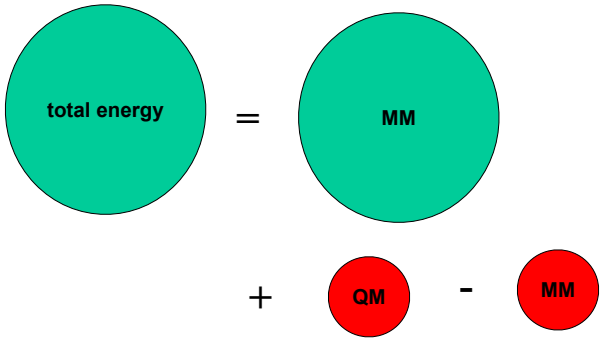
Hexaphenylethane
(**HPE**)

Triphenylmethyl radical
(**TPMR**)

from S. Irle

Subtractive QM/MM: ONIOM

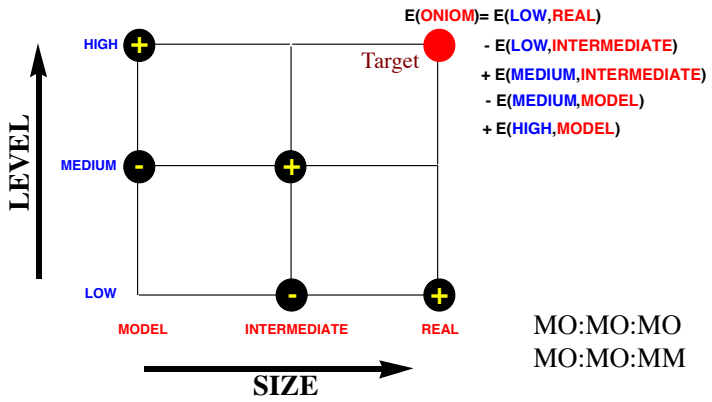
Morokuma and co.: GAUSSIAN



- Mechanical embedding: only steric effects
- Electrostatic embedding: polarization of QM due to MM

EMORY UNIVERSITY

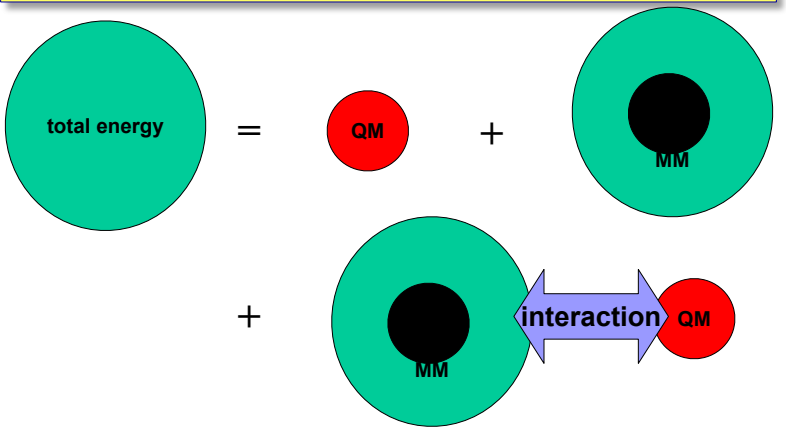
Three-layer ONIOM (ONIOM3)



MO:MO:MO
MO:MO:MM

from S. Irle

Additive QM/MM



- mostly electrostatic embedding: polarization of QM due to MM

Combined QM/MM

Bonds:

a) take force field terms

b) - link atom

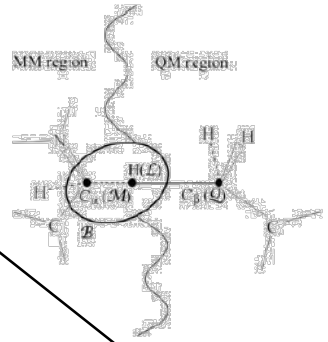
- pseudo atoms

- frontier bonds

Nonbonding:

- VdW
- electrostatics

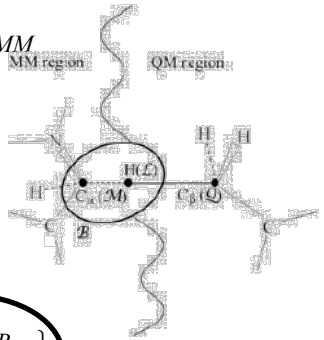
Amaro , Field , Chem Acc. 2003



$$\hat{H}_{QM/MM} = -\sum_{i,M}\frac{q_M}{r_{iM}} + \sum_{\alpha,M}\frac{Z_\alpha q_M}{R_{\alpha M}} - \sum_{\alpha,M}\left\{\frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^{6}}\right\} + \hat{H}^{\text{int.coor}}_{QM/MM}$$

Additive QM/MM:

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



$$\hat{H}_{QM/MM} = -\sum_{i,M}\frac{q_M}{r_{iM}} + \sum_{\alpha,M}\frac{Z_\alpha q_M}{R_{\alpha M}} - \sum_{\alpha,M}\left\{\frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^{6}}\right\} + \hat{H}^{\text{int.coor}}_{QM/MM}$$

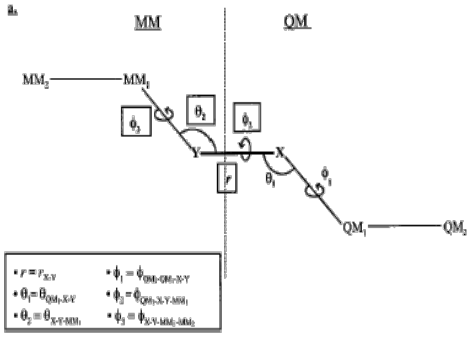
Elecrostatic mechanical embedding

Combined QM/MM

Bonds:

a)

from force field



Reuter et al, JPCA 2000

link atom

a) **constrain or not?**

(artificial forces)

relevant for MD

b) Electrostatics

- **LA included – excluded**

  (include!)

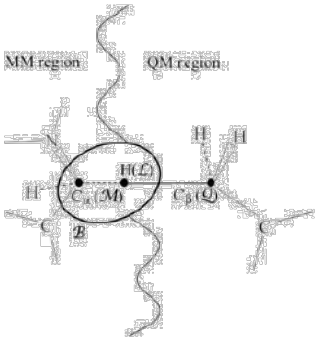
- **QM-MM:**

  exclude MM-host

  exclude MM-hostgroup

- **DFT, HF: gaussian broadening of MM point charges, pseudopotentails (e spill out)**

- *J. Chem. Phys.* ***2002, 117, 10534*** *J. Phys. Chem. B* ***2005, 109, 9082***



Amaro & Field, T. Chem Acc. 2003

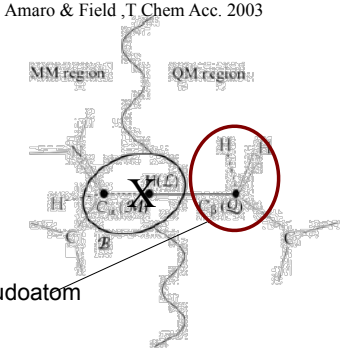
pseudoatoms

Amaro & Field ,T Chem Acc. 2003

Pseudobond- connection atom

Zhang, Lee, Yang, JCP **110**, 46

Antes&Thiel, JPCA **103** 9290



No link atom: parametrize $C_\beta H_2$ as pseudoatom

$$\hat{H}_{QM/MM} = -\sum_{i,M}\frac{q_M}{r_{iM}} + \sum_{\alpha,M}\frac{Z_\alpha q_M}{R_{\alpha M}} + \sum_{\alpha,M}\left\{\frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^{6}}\right\} + \hat{H}^{\text{int.coor}}_{QM/MM}$$

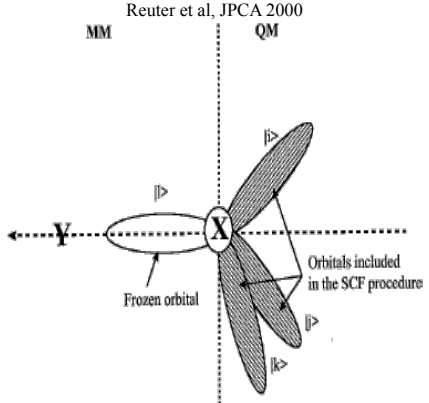
frozen orbitals

Warshel, Levitt 1976

Rivail + co. 1996-2002

Gao et al 1998

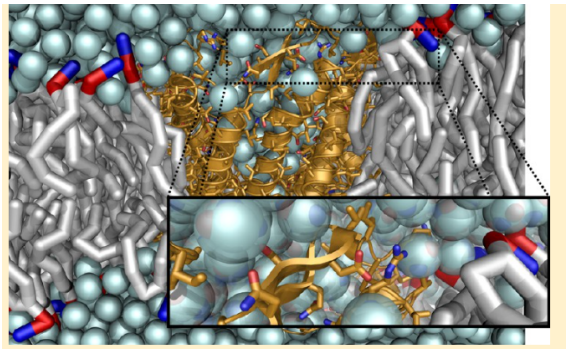
Reuter et al, JPCA 2000



$$\hat{H}_{QM/MM} = -\sum_{i,M}\frac{q_M}{r_{iM}} + \sum_{\alpha,M}\frac{Z_\alpha q_M}{R_{\alpha M}} + \sum_{\alpha,M}\left\{\frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^{6}}\right\} + \hat{H}^{\text{int.coor}}_{QM/MM}$$

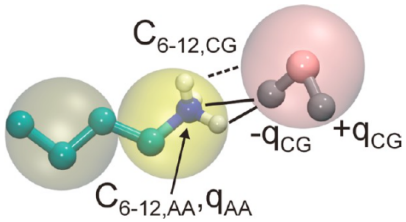
MM-CG

•very often, one needs more resolution, e.g. waters in protein, which is functionally important (CG: 4 water bead!)



JPC B 117, 3516.

MM-CG



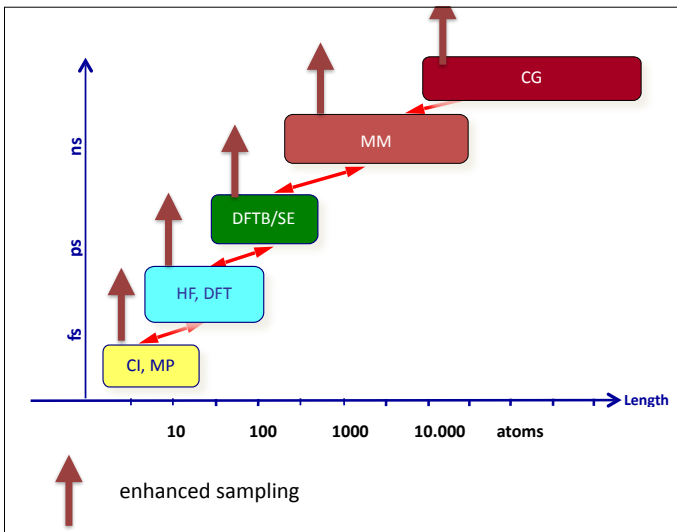
Mixed system: CG and AA (all atom)

CG-CG and AA-AA unchanged

place virtual CG sites at the center of the AA groups:
- CG description of the interaction
- distribute CG forces on atoms

introduce screening for Coulomb interactions
AA-AA
AA-CG
CG-CG

JPC B 117, 3516.



Adaptive schemes: change resolution on the fly

QM - MM

MM- CG