Fakulteta za informacijske študije
Faculty of information studies

Naložba v vašo prihodnost
Operacijo delno financira Evropska unija
Evropski sklad za regionalni razvoj

Kreativno jedro: Simulacije
Creative core: Simulations

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA IZOBRAŽEVANJE, ZNANOST IN ŠPORT

Operacijo delno financira Evropska unija in sicer iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa krepitve regionalnih razvojnih potencialov za obdobje 2007-2013, 1. razvojne prioritete: Konkurenčnost podjetij in raziskovalna odličnost, prednostne usmeritve 1.1: Izboljšanje konkurenčnih sposobnosti podjetij in raziskovalna odličnost.

Molekularne simulacije bioloških makromolekul

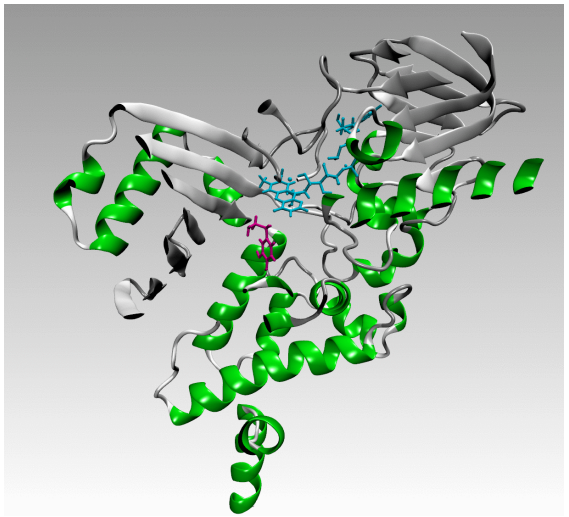
Janez Mavri

e-mail janez.mavri@ki.si

Kemijski inštitut

Slovenia

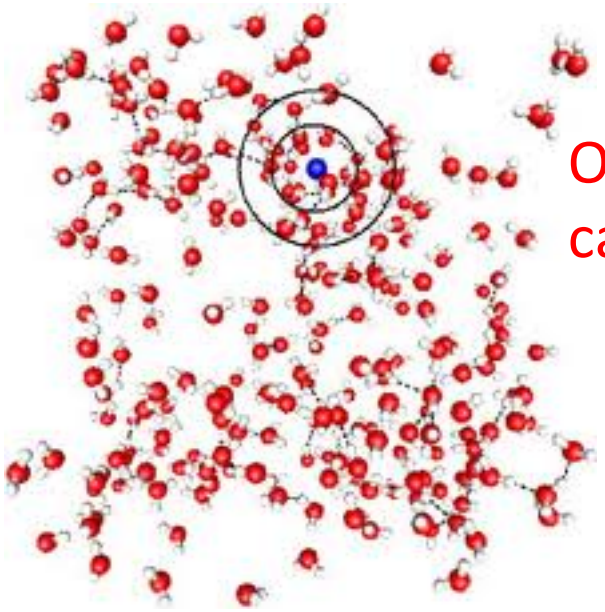
Laboratory for Biocomputing and Bioinformatics



Outline

- Need for thermal averaging
- Force field
- Elements of Molecular dynamics
  (initial condition, integration of equations of motion,
  long range electrostatics, boundary condition)
- NMR and X-ray refinement
- How to perform MD in practice using GROMACS
- Simulation of enzyme reactions
- Perspectives

Textbooks say that all hydrated proteins represent soft condensed matter. What does it mean:

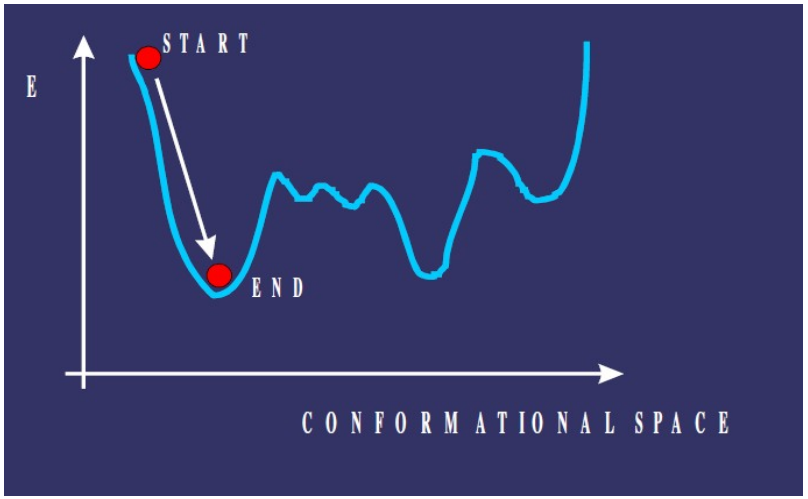


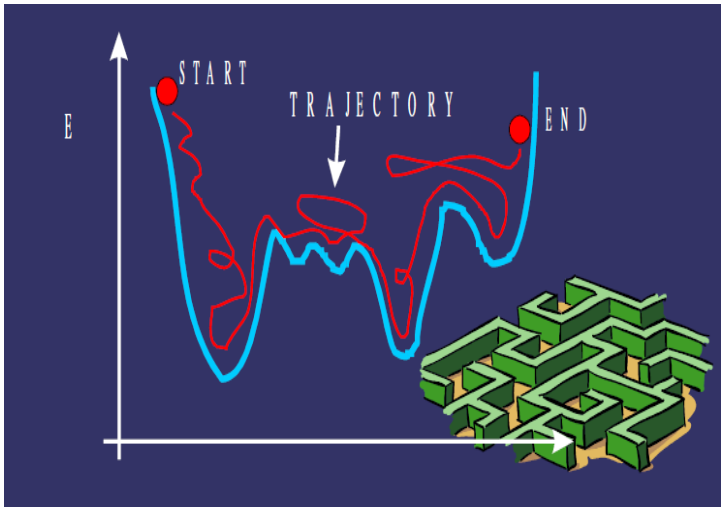
One single geometry of several water molecules can not describe aqueous solution !

Several milions of configuration are within k_BT =0.59 kcal/mol per degree of freedom at room temperature. The idea of thermal averaging is to search the phase space numerically and to calculate ensemble averages and dynamic quantities.

Energy minimization vs. thermal averaging (energy minimization==geometry optimization)

Energy minization means that we are looking
for such configuration of the nuclei that the force is zero and energy minimal.





Thermal averaging means that we are (smartly) exploring relevant configurations.

Molecular dynamics means numerical solving equations of motion for the nuclei (particles)

$$m_i \frac{d^2 r_i(t)}{dt^2} = F_i = -\nabla V$$

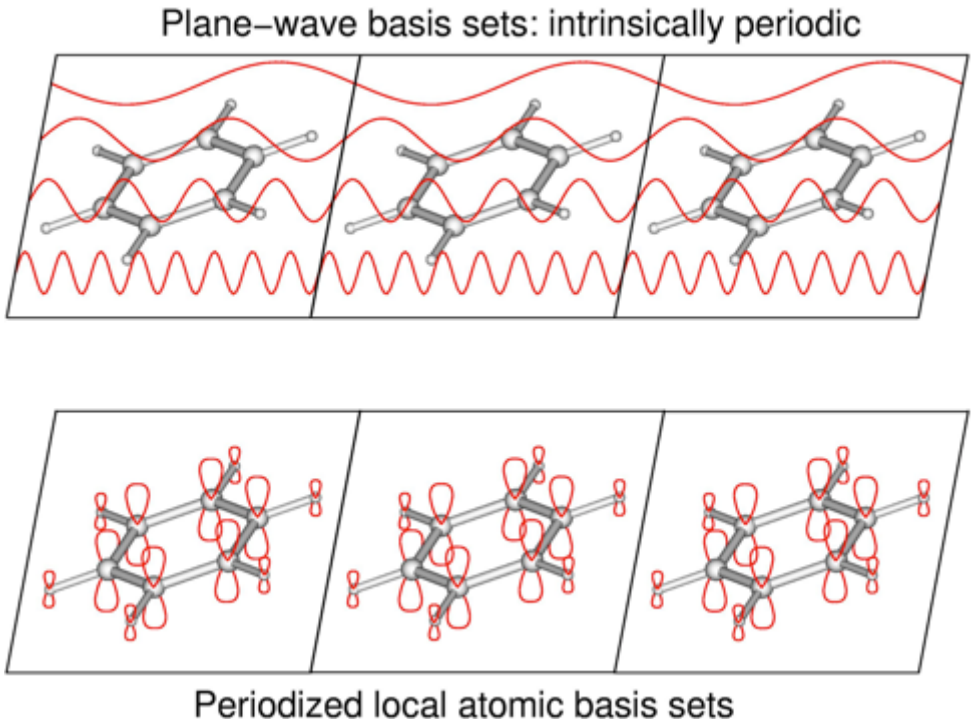
Few algorithms are in practical use. Accurate integrators of ordinary differental equations like Runge-Kutta 4-th order are not practical Leapfrog algorithm:

$$\begin{aligned} r(t+\Delta t) &= r(t) + v(t+\frac{1}{2}\Delta t)\Delta t \\ v(t+\frac{1}{2}\Delta t) &= v(t-\frac{1}{2}\Delta t) + a(t)\Delta t \end{aligned}$$

Car Parrinello

simulation is solving classical equations of motion for ab initio potential calculated on the DFT level. In order to avoid SCF calculations at each time step coefficients of the basis functions are carried with using the arbitrary small mass value for the electrons. Traditionally plane wave basis set is used.

Roberto Car and Michele Parrinello



Force field construction

Force field is a simple (empirical) description of the energy terms that can be calculated fast and several times for large systems in order to save computer time.

Functional form of a typical nonpolarizable force field

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{\text{eq}})^2 +$$

$$\sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right]$$

J. Am. Chem. Soc. **1995,** *117,* 5179−5197

A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules

Wendy D. Cornell,[†] Piotr Cieplak,[‡] Christopher I. Bayly,[§] Ian R. Gould,[⊥] Kenneth M. Merz, Jr.,[∥] David M. Ferguson,[&] David C. Spellmeyer,[#] Thomas Fox, James W. Caldwell, and Peter A. Kollman*

Force field construction is not easy

Biologically relevant systems are highly polar or ionic

Critical are nonbonding parameters, specially electrostatics and repulsive terms.

Terms associated with high frequency modes (bonds, angles, improper dihedrals) are less relevant.

Dispersion terms are less relevant.

Bulk water holds together electrostatics and is balanced by repulsive (hard core) terms.

Application of these two terms determines density and heat of evaporation.

Force fields for proteins and nucleic acid
are based on the Hagler-Lifson force field.

Hagler-Lifson force field was constructed
on the basis of thermodynamics and density of
peptide crystals.

Shneior Lifson

Other force fields were constructed from it

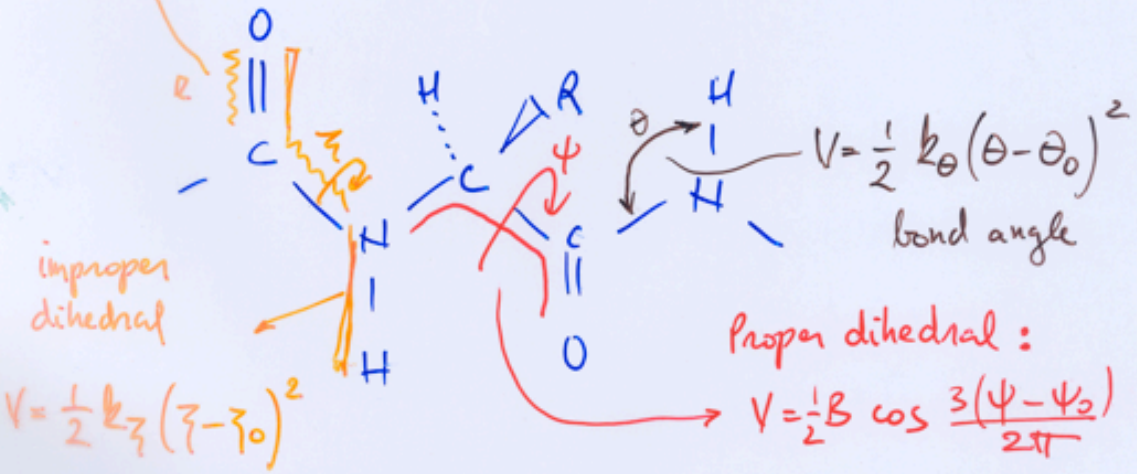
DISCOVER MOLARIS CHARMM GROMOS AMBER

OPLS force field of Jorgensen is based on liquid data

FORCE FIELD FOR A PROTEIN & WATER

$V = \frac{1}{2} k (r - r_0)^2$ bond

H_2O } SPC WATER MODEL



NONBONDING INTERACTIONS :

LJ AT ALL CENTERS BUT THE HYDROGENS

$$LJ = 4\varepsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right)$$

APPLY LORENZ-BERTHELOT MIXING RULES

$$\sigma_{AB} = \frac{1}{2} (\sigma_{AA} + \sigma_{BB})$$

$$\varepsilon_{AB} = \sqrt{\varepsilon_{AA} \varepsilon_{BB}}$$

CHARGES :

EMPIRICAL CHARGES

$$V = \frac{1}{4\pi\varepsilon_0} \cdot \frac{q_i q_j}{r_{ij}}$$

OR MERZ-KOLLMAN CHARGES HF/6-31G(d)

Warning!

Integration of classical equations of motion is at room temperature strictly valid only for frequencies below 200 cm^{-1}.

What is the applied time step:

Recipe: 1/10 -1/40 of the period of the fastest periodic motion.

In hydrated protein the fastest motion is NH stretching corresponding to 3400 cm^{-1}=10 fs

1 fs = 1.10^{-15} s

One can freeze fast motions by applications of holonomic constraints such as SHAKE: Ryckaert, J-P; Ciccotti G, Berendsen HJC (1977). Journal of Computational Physics 23, 327

or RATTLE: H.C. Andersen, (1983) J. Comp. Phys. 52,24-34

How long real time we can simulate?

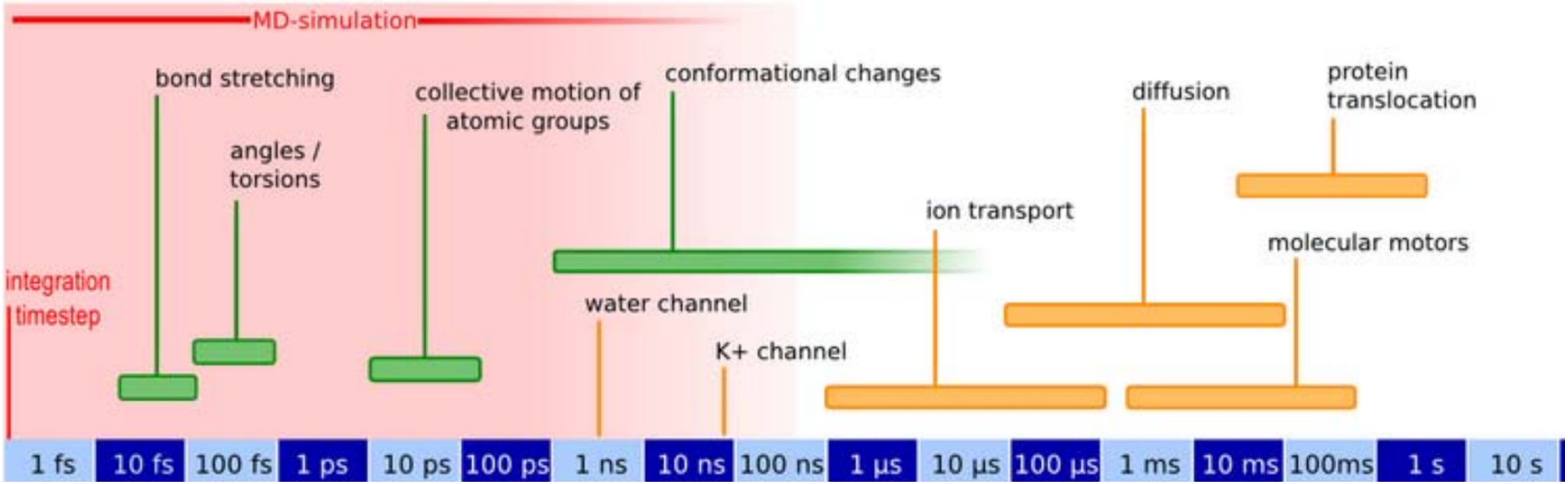


Figure from http://www.pasteur.edu.uy/sites/ipmont/files/MD.pdf

Temperature control

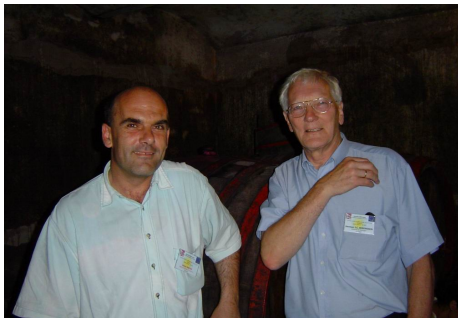
Temperature has something to do with kinetic energy

All thermostats scale the atomic velocities in this or that way

Berendsen thermostat: $\frac{dT}{dt} = \frac{T_0 - T}{\tau}$

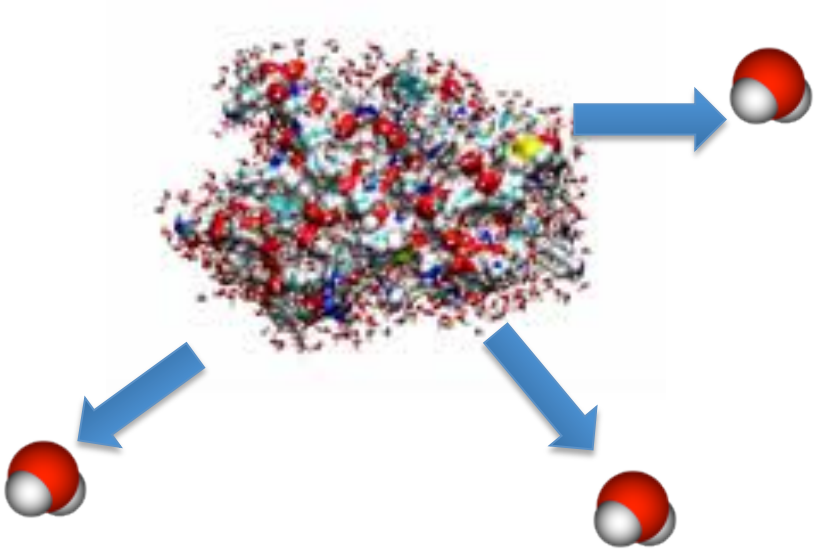
Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. (1984). "Molecular-Dynamics with Coupling to an External Bath". Journal of Chemical Physics 81 (8): 3684–3690

Beware of frozen ball syndrome when simulating dry proteins.

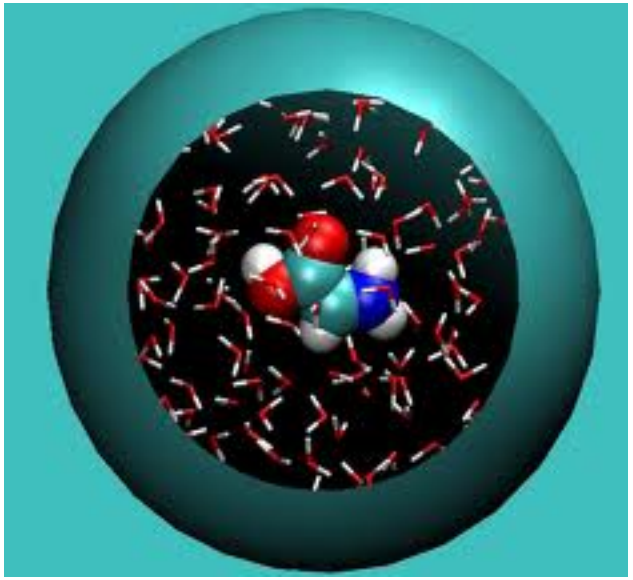


Molecular dynamics of an isolated protein is easy but is not a realistic system for molecular biology.

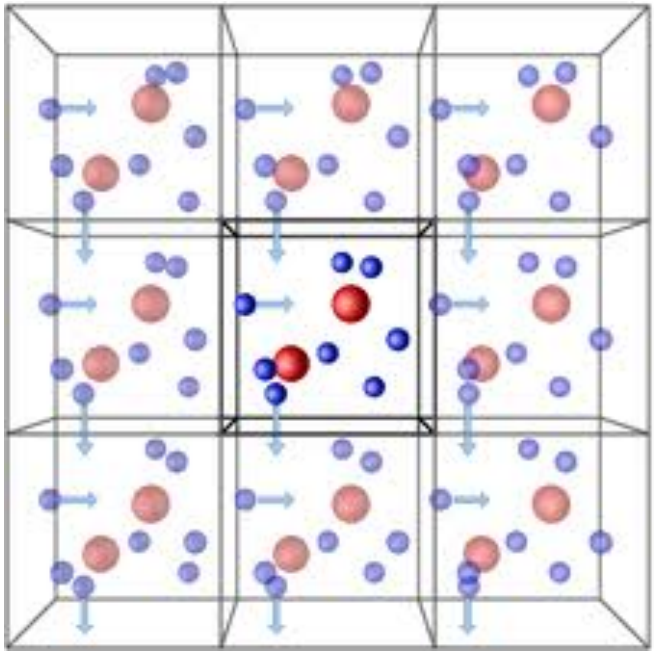
We need to add water molecules and (counter)ions



water molecules will slowly evaporate during the simulation

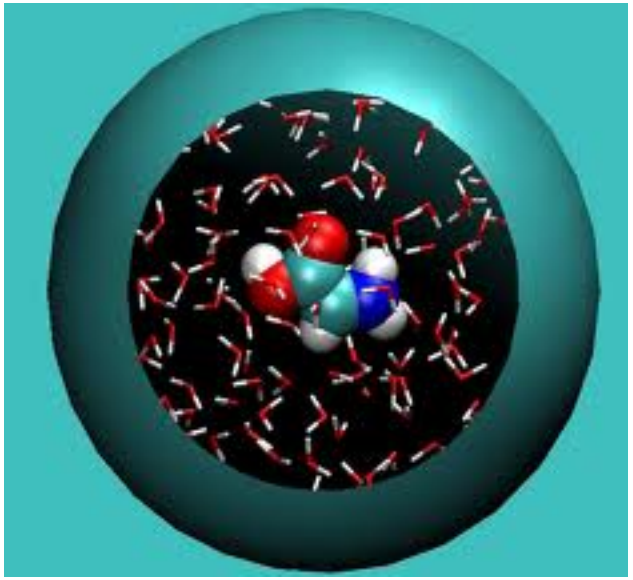


Surface constrained solvent along with proper treatment of dielectric properties at the boundary:
Warshel's school

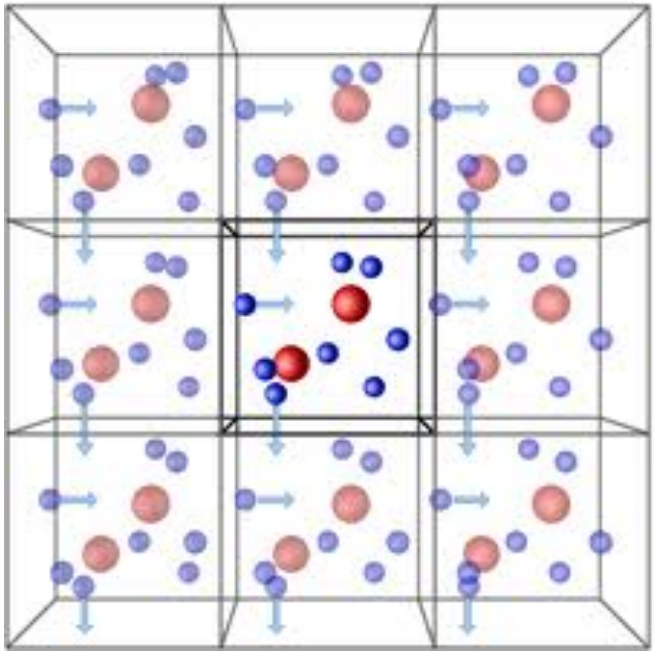


Periodic boundary condition.

The only long range interaction is electrostatics.
We can set it to zero for Cutoff distance (Cutoff< BOX/2)
Or we consider all the images
(Ewald summation, numerical implementation is PME Particle Mesh Ewald)



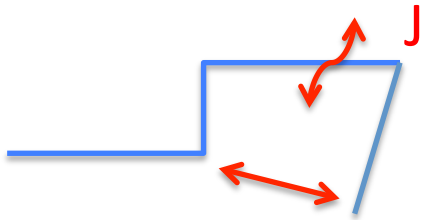
Warning !
Treatment of long range electrostatics is integral part of the force field.



Structure refinement by molecular simulation

The basic is idea is inclusion of the experimental data into the force field. Typically harmonic form is used.

Nuclear Overhauser Effect connectivities of NMR coupling constants.



NOE= 3-6A

For X-ray refinement harmonic position restraints are used. The applied force constants depend on the B factors.

Calculations of free energy differences

(required to estimate binding affinities and chemical reacitvity)
need special approaches. In general they can not be calculated from a long trajectory.

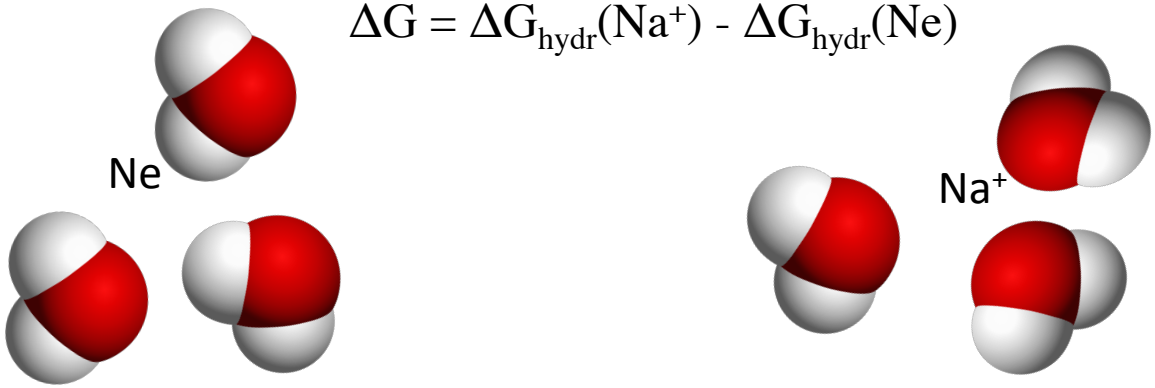
Methods of choice are thermodynamic integration and thermodynamic perturbation.

The basic idea is that one makes potential dependent on the coupling parameter λ that couples these two thermodynamic states. Alchemic changes are possible.

λ=0

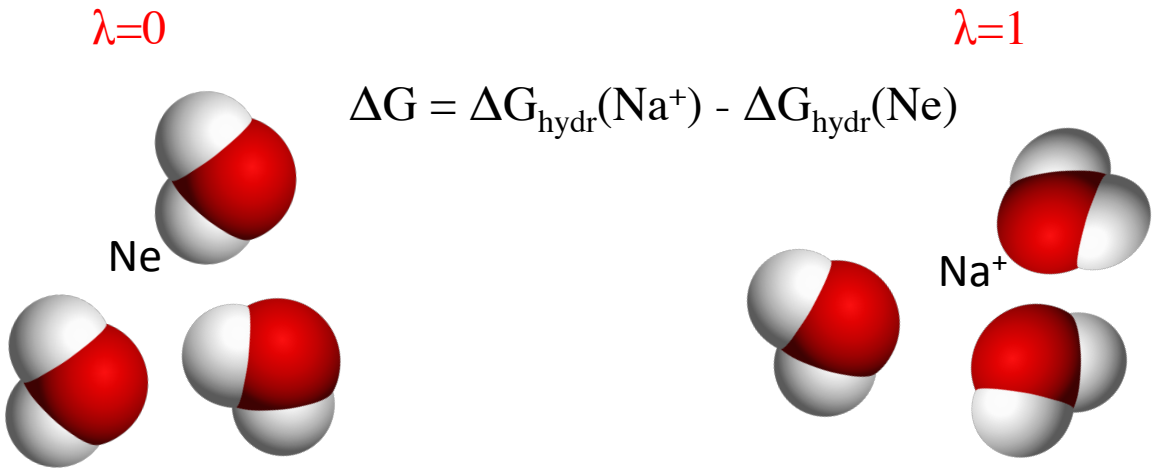
λ=1

$$\Delta G = \Delta G_{hydr}(Na^+) - \Delta G_{hydr}(Ne)$$



Beware of Born correction that originates from solvent contribution beyond spherical cutoff R_{cutoff}

$$\Delta G_{corr} = - \tfrac{1}{2}\, 332\, q^2 / R_{cutoff}\ (1-1/\varepsilon)$$



Protein Molecular Dynamics in Practice

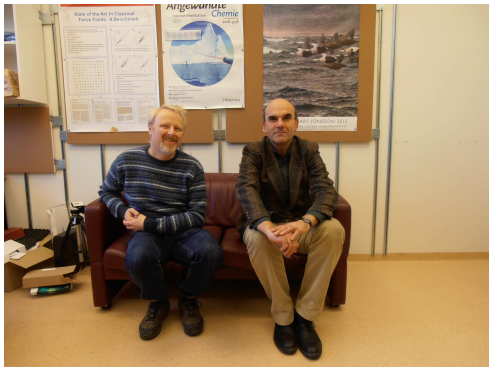
Preparing and running a Molecular Dynamics Simulation

Protein MD in practice

Software

What do we want ?

- Speed
- Good documentation and helpful community
- Lots of features (the useful ones)
- Ease of use
- Preferably cheap or freeware
- GPU support



David van der Spoel,
Uppsala University,
the original GROMACS developer

Free

GROMACS - faster on single CPU machines, open-source, most force fields included, good documentation, GPU support, good parallelization

NAMD - VMD integration, faster on large systems (100 cores with low latency network), closed-source, most FF, GPU support, very good parallelization

Tinker, Abalone, Adun, Cosmos, Desmond

Not free

GROMOS, CHARMM, AMBER, TerraChem, Discovery Studio (expensive!)

Wilfred van Gunesteren,
ETH Zuerich,
the original GROMOS developer

Hardware and $

A single 1000€ PC with a dedicated GPU:

CPU: Intel I7 2600	ca. 300 €
GPU: Geforce GTX 660 TI	ca. 250 €
RAM: 16 GB	ca. 150 €

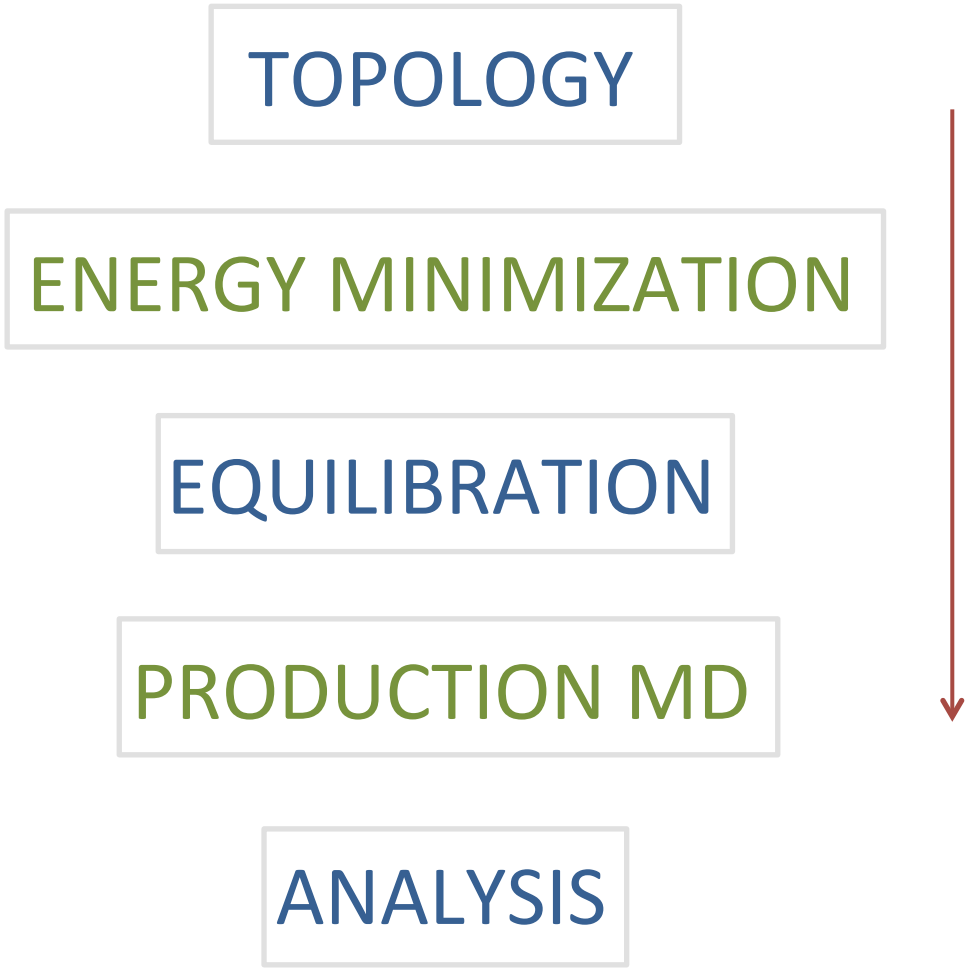
Performance on a 3000 atom system:

No GPU:	170 ns/day
With GPU:	380 ns/day

GPUs are the future

Protein MD in practice

THE WORKFLOW

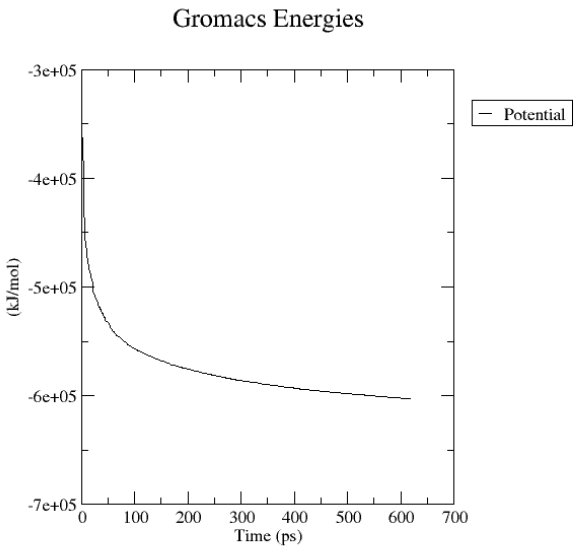


Protein MD in practice

ENERGY MINIMIZATION

Prior the simulation we remove potential energy.

Usually, the algorithm employed is the **steepest descents** minimizer.

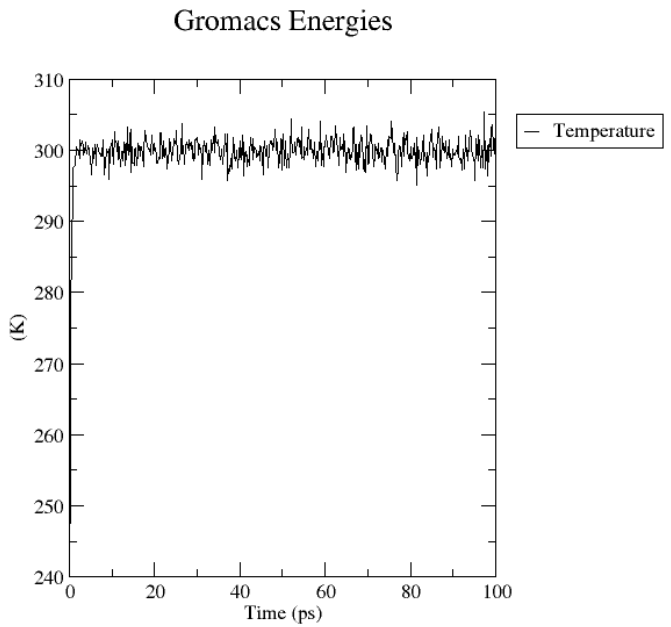


Protein MD in practice

EQUILIBRATION

The hydrated protein needs to be thermalized.

Equilibration time of 100 ps a good value for hydrated protein.
The temperature is held at the
desired value of **300 K** by using
a **thermostat**.



Protein MD in practice

PRODUCTION MD

Continue from NVT equilibrated configuration.

Change the simulation length.

Run MD.

Get some coffee…

Protein MD in practice

MyD88 TIR domain

Some point mutations in MyD88 TIR domain are responsible for cancerous cell survival.

Together with Laboratory for Biotechnology@NIC (Roman Jerala and Monika Avbelj) we are studying the possible causes for the **activity** of these mutants, both experimentally and with MD simulation.

The question is, what structural changes are present in the mutant protein, that would change the conformation or flexibility so much it would readily bind to another TIR while the wildtype would not.

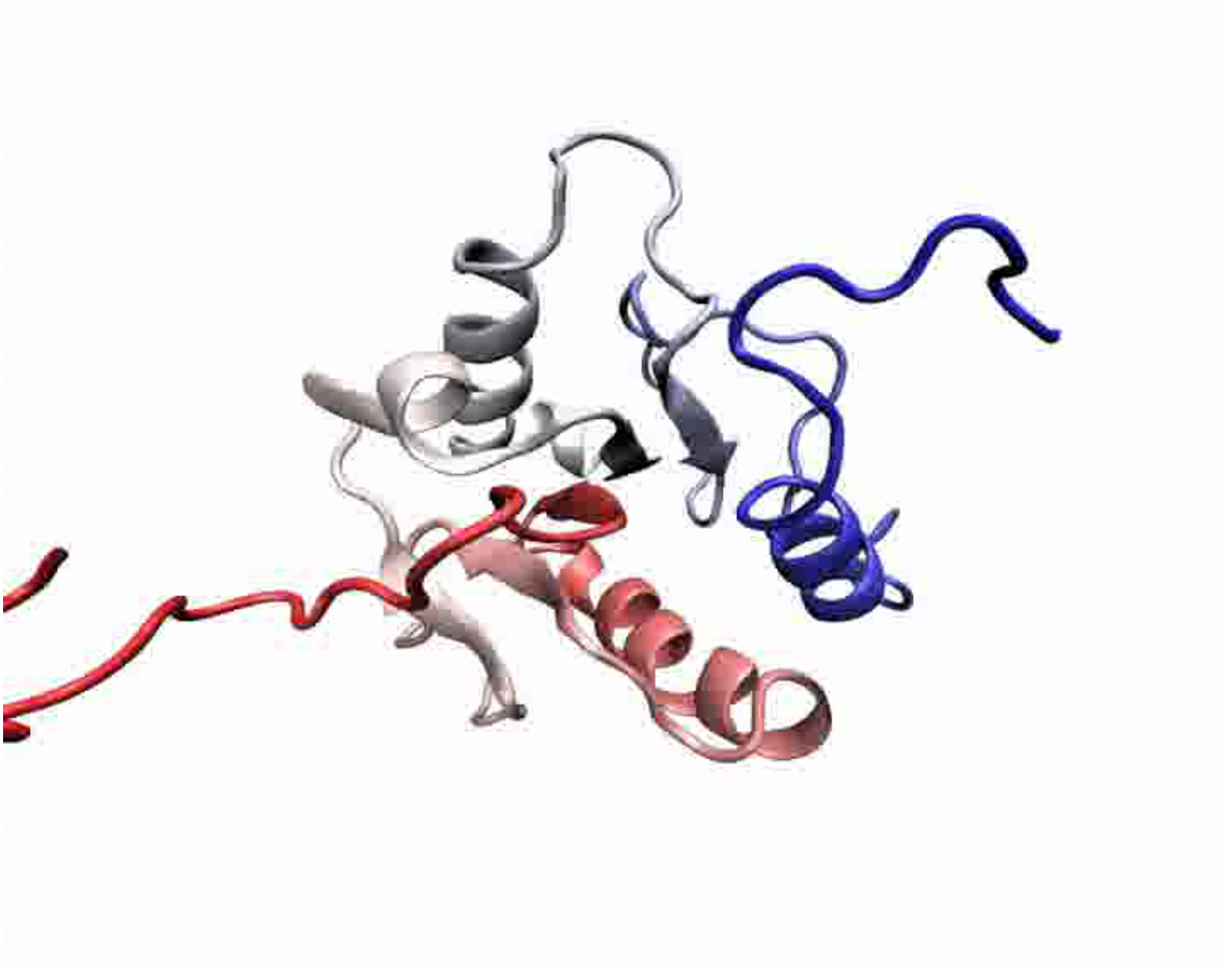
Protein MD in practice

ANALYSIS - visual

VMD

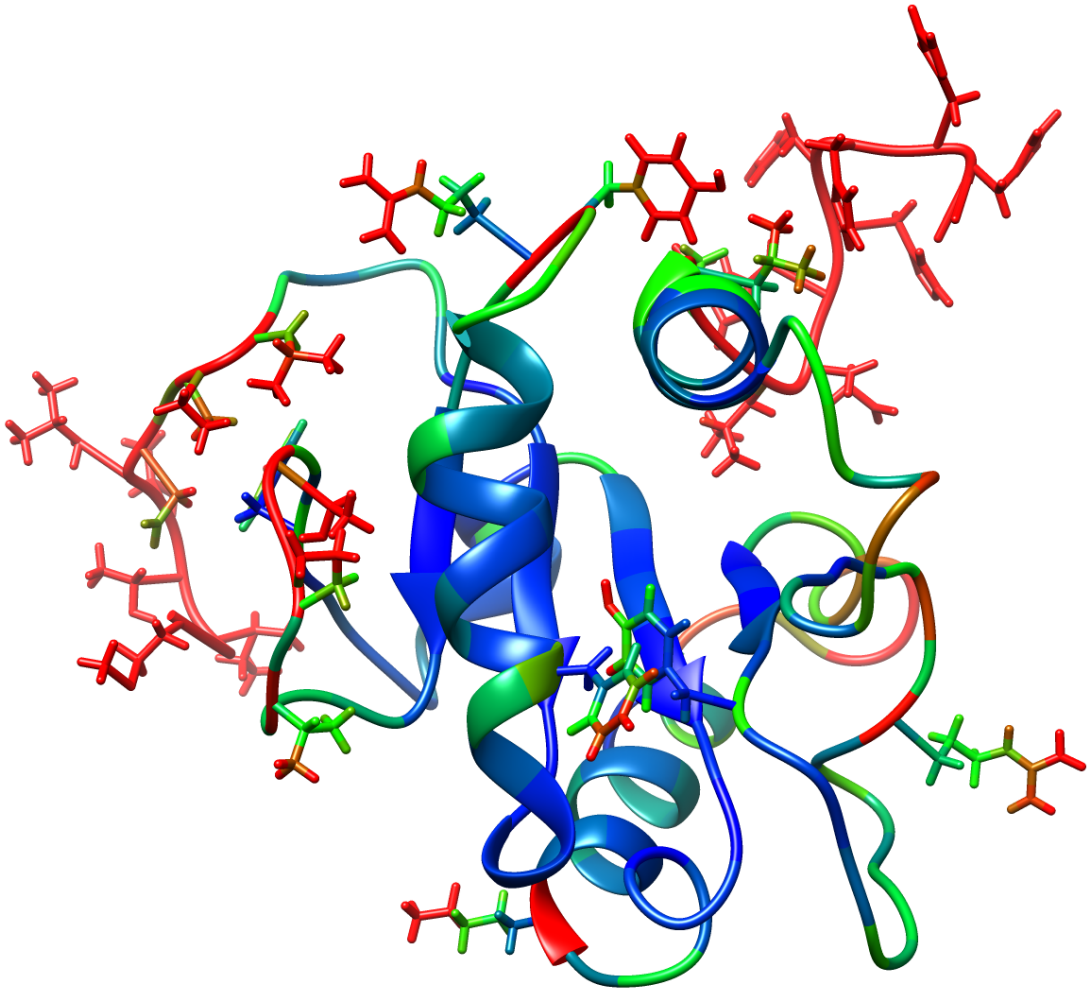
PyMOL

Chimera



Protein MD in practice

ANALYSIS – structure



Computed B-factor or Temperature factor – indicator of **thermal motion** about an atom.

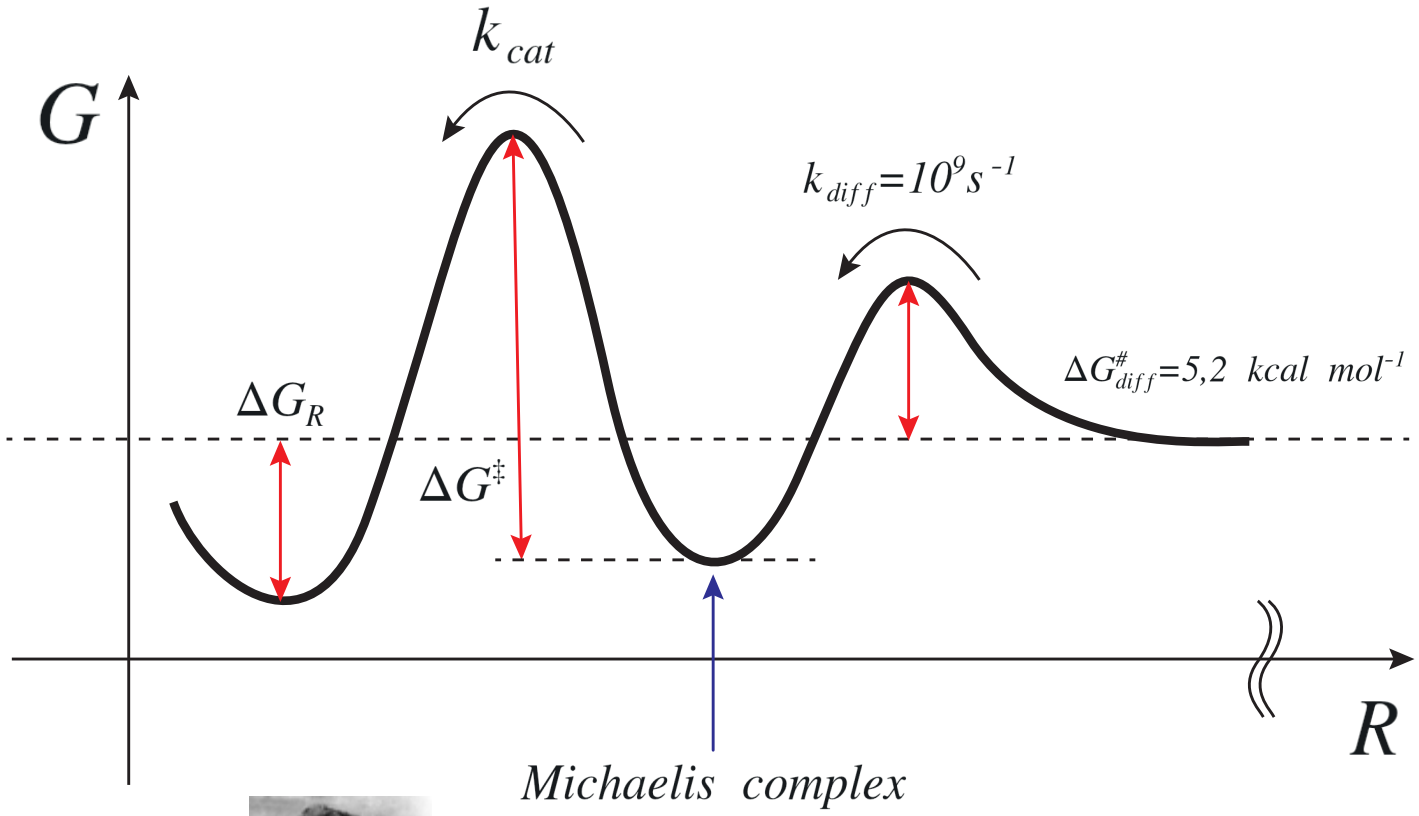
Simulation of enzymatic reactions is not easy and requires QM/MM approaches.

The hybrid QM/MM (quantum mechanics/molecular mechanics) approach is a molecular simulation method that combines the strength of both QM (accuracy) and MM (speed) calculations, thus allowing for the study of chemical processes in solution and in proteins.
The QM/MM approach was introduced in the 1976 paper of Warshel and Levitt.

Warshel, A; Levitt, M (1976). "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme". Journal of Molecular Biology 103 (2): 227–49.

http://en.wikipedia.org/wiki/QM/MM
June 21, 2013

$$k_{cat} = \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{k_B T}}$$

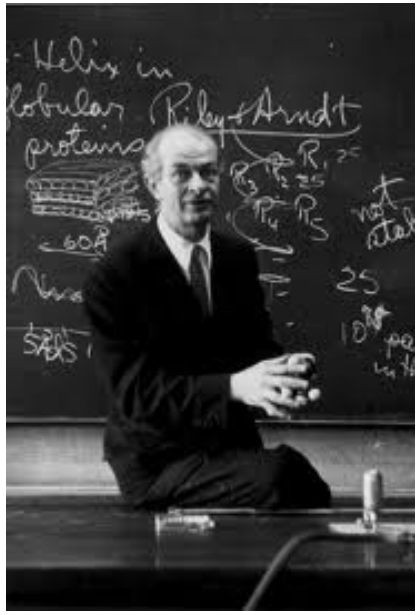


Leonor Michaelis

Maud Menten

L. Michaelis and Miss Maud L. Menten (1913), "Die Kinetik der Invertinwirkung", *Biochem Z* **49**: 333–369

What are the factors responsible for enzyme catalysis?

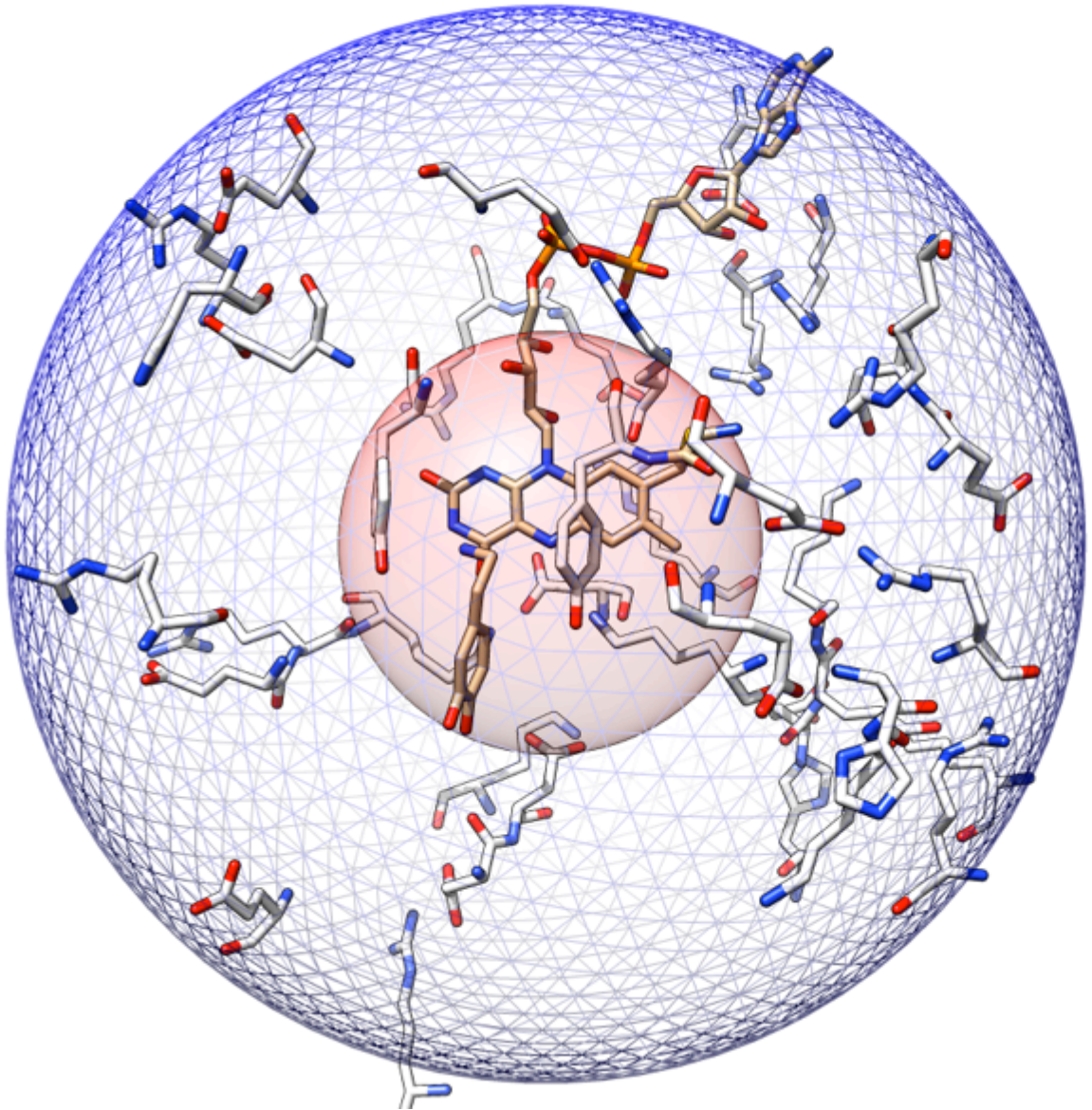


Linus Pauling rationalized the nature of enzyme catalysis by more favorable solvation of the transition state than the reactants, relative to the situation in aqueous solution. As solvent he considered enzyme environment.

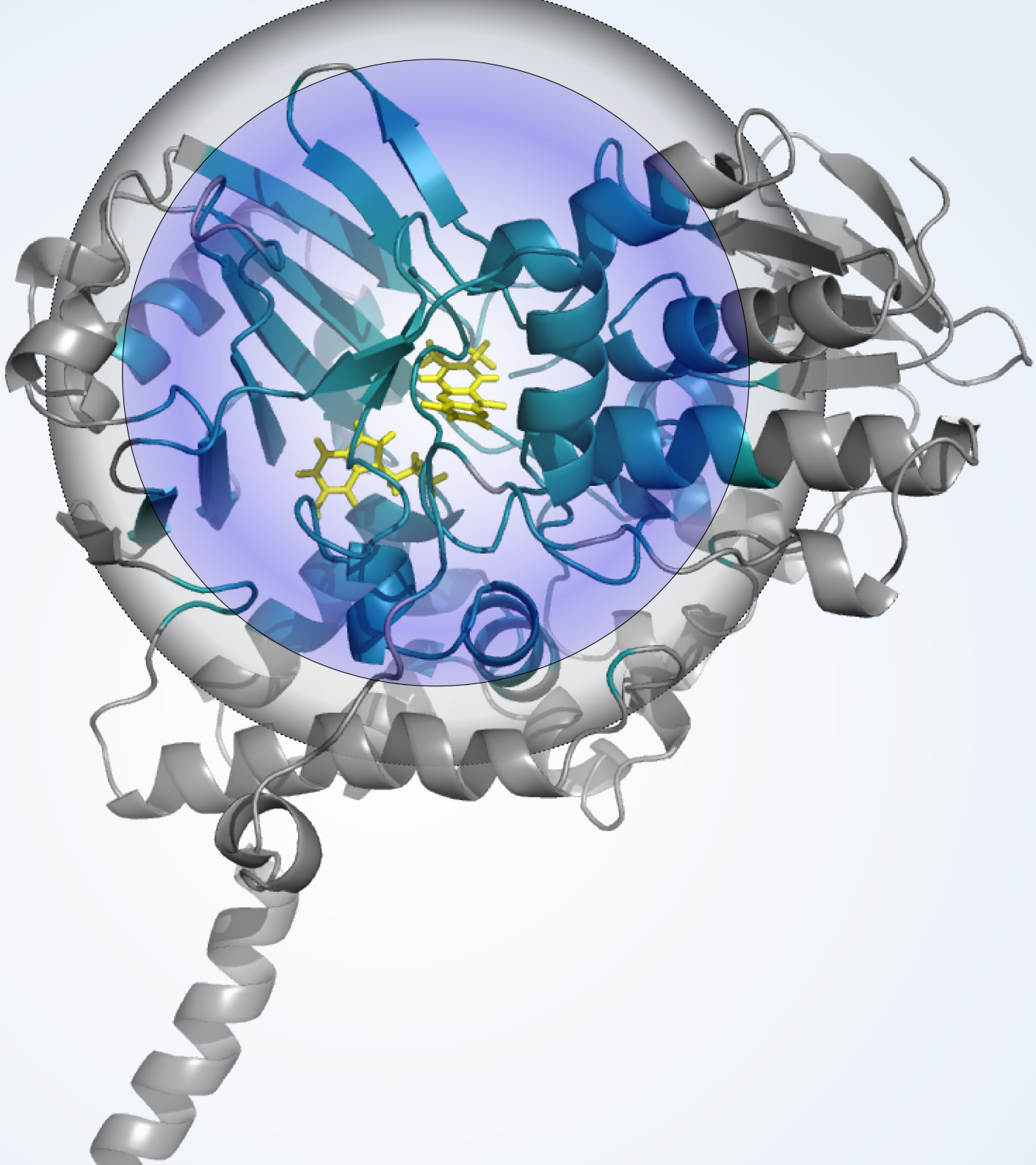
Arieh Warshel recognized the praeorganized electrostatics as the only relevant factor for enzyme catalysis.

Therefore transition state analogs should be excellent inhibitors.

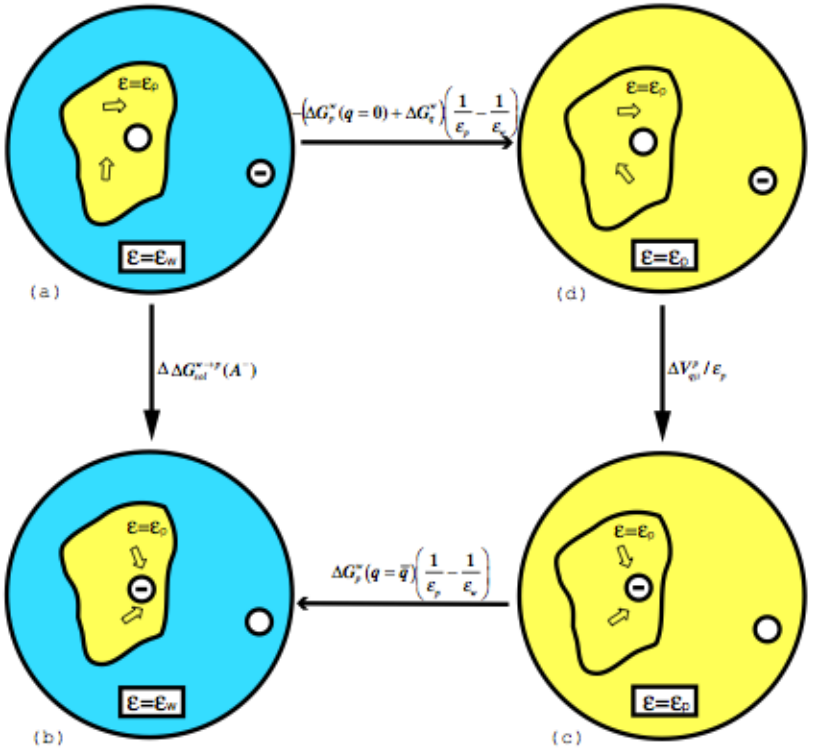
Praeorganized electrostatics means that polar and ionizable groups (some are ionized !) around the transition state are organized in a way that they stabilize (=solvate) transition state better than the reactants. Water molecules are not allowed to enter in significant numbers.



QM/MM



It is essential to determine the protonation states of the ionizable residues



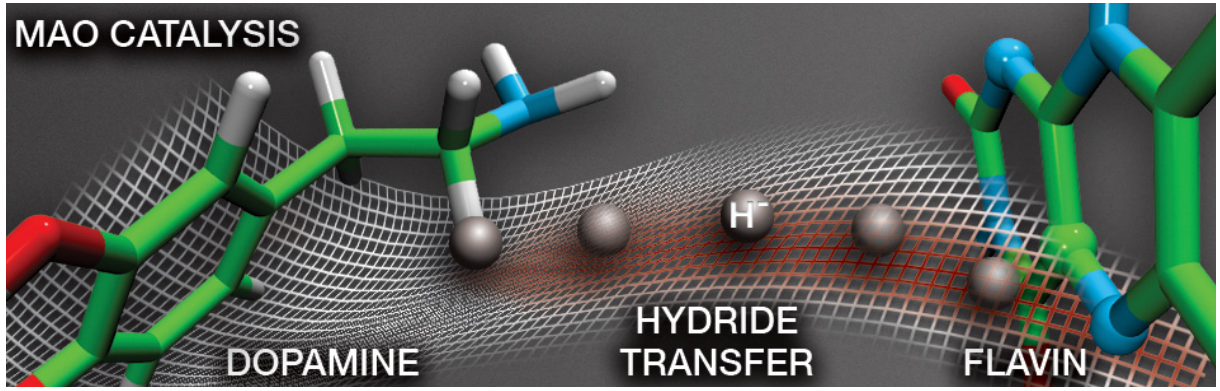
MOLARIS: Version 9.11

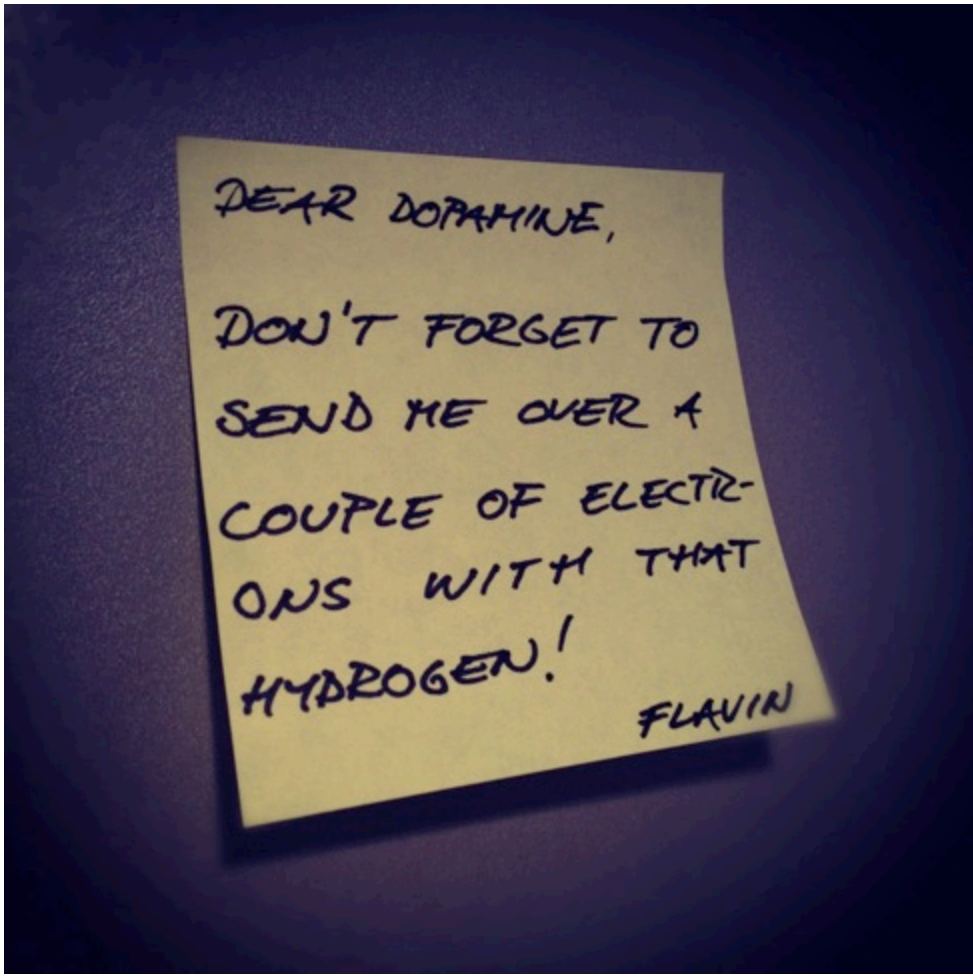
Reference Manual

Release date: July 15, 2011

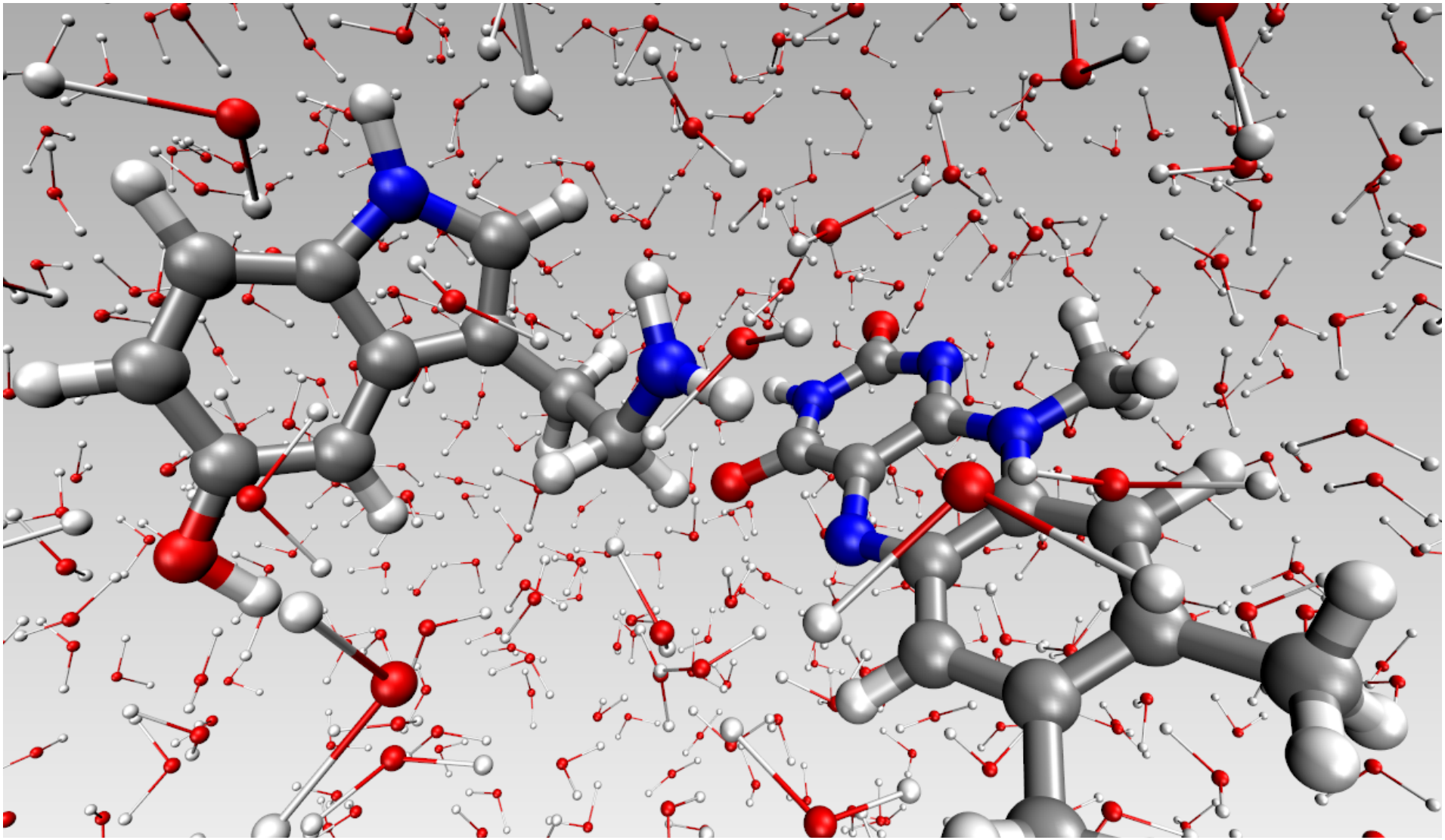
Arieh Warshel, University of Southern California

Monoamine oxidases metabolize neurotranmitters dopamine and serotonin.
Rate limiting step is abstraction of hydride from methylene group.

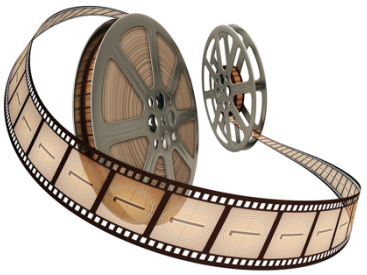


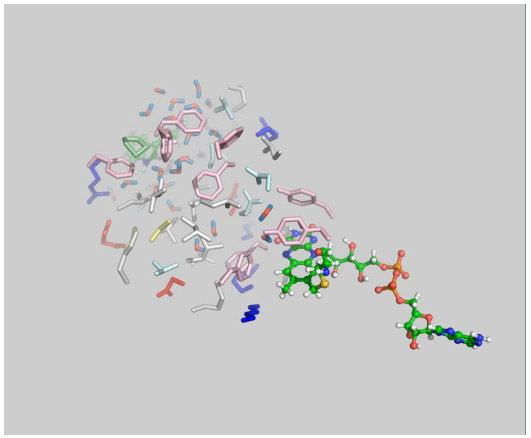


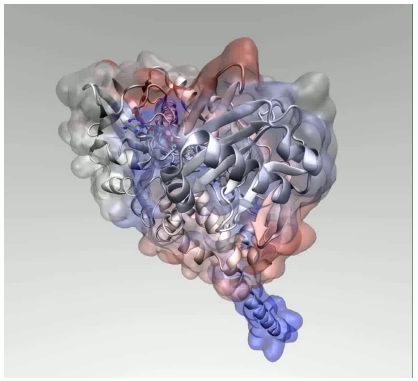
•Reference reaction in water



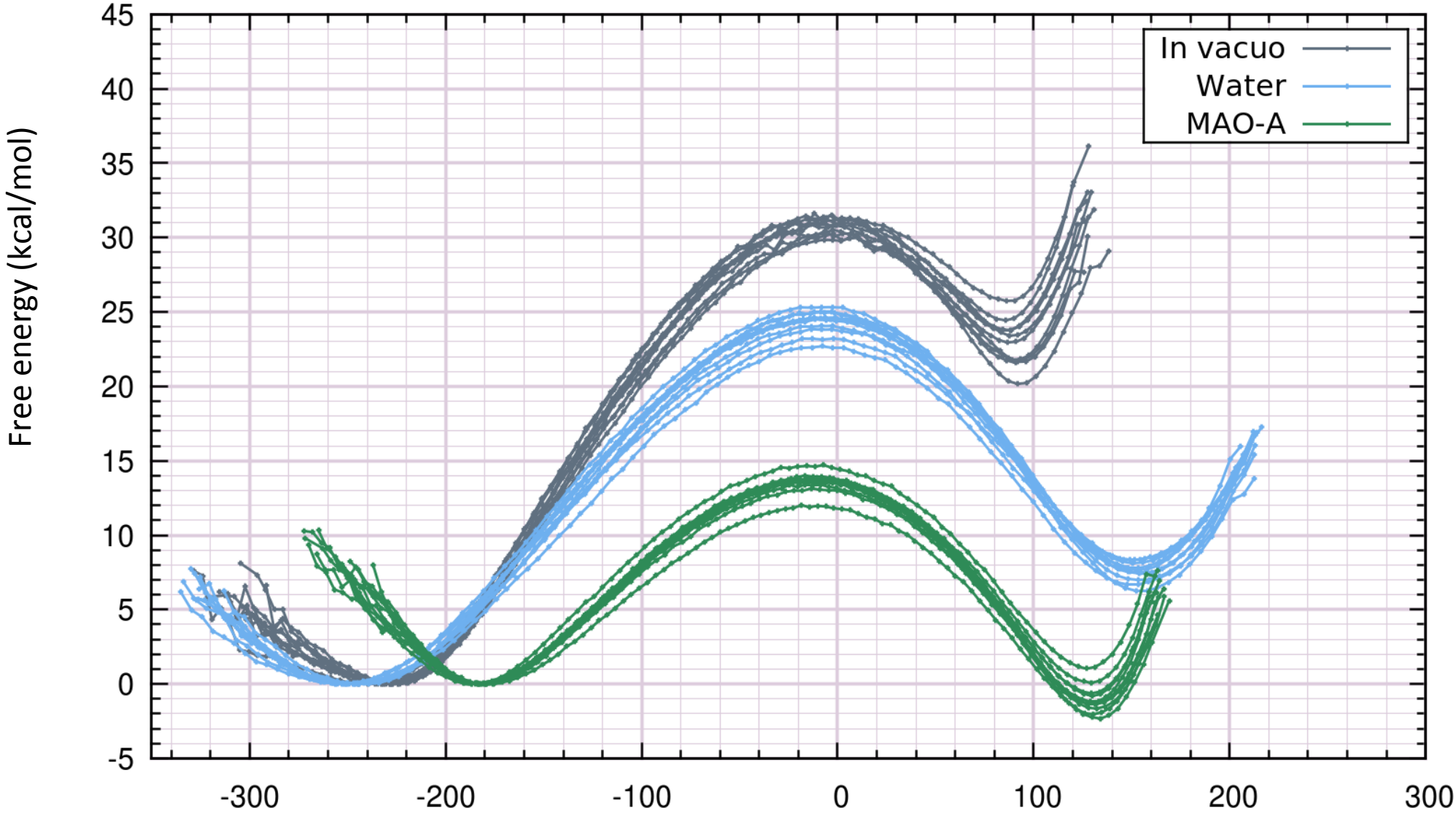
Miha’s movies





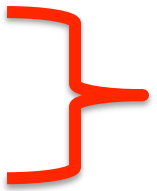


Serotonin Decomposition



Serotonin Decomposition

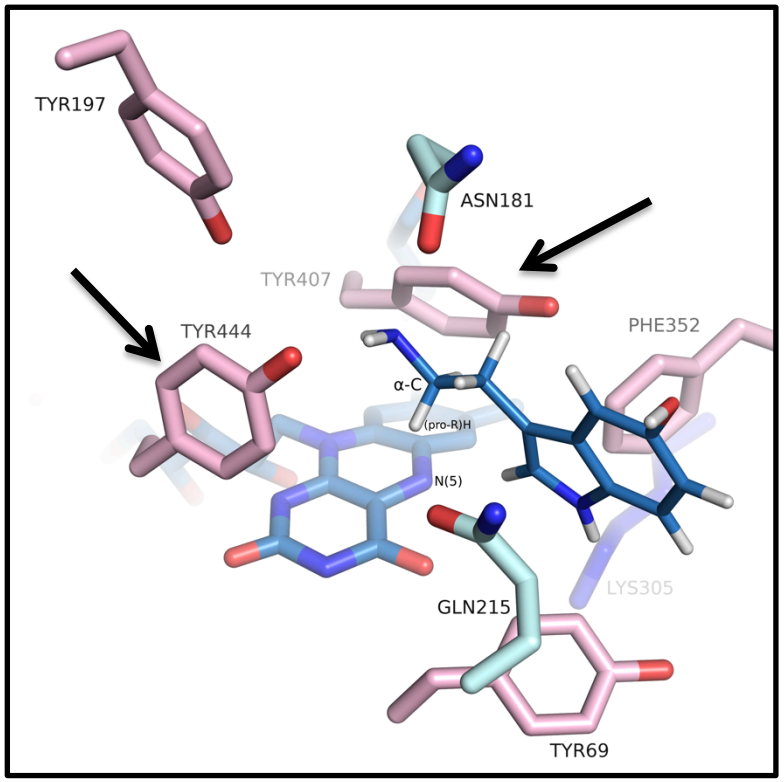
Environment	Activation free energy (kcal/mol)
Gas phase	30.9± 0.5
Water	24.3± 0.8
Enzyme	13.6± 0.7



MAO A provides 10.7 kcal/mol to catalysis

Matej Repič,[a#] Miha Purg,[a#] Robert Vianello[b] and Janez Mavri [a,c]*, to be submitted 2014

Effects of Point Mutations

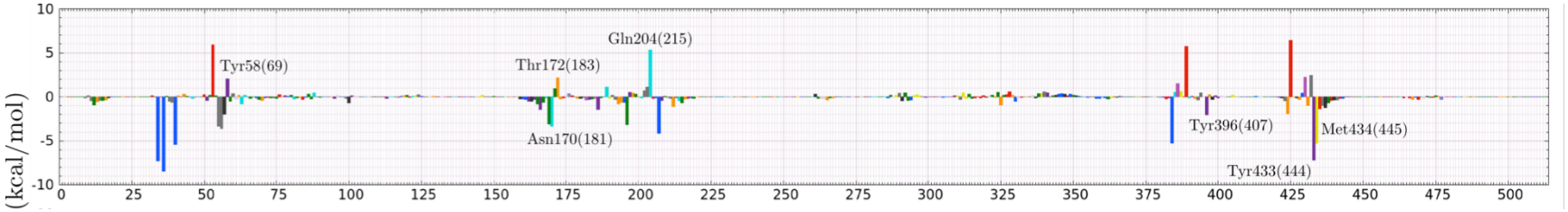


Aromatic cage is essential for catalysis

Double mutation of Tyr practically deactivates the enzyme .

Mutation Tyr444Phe lowers the rate for factor of 15 (exp. value is 120).

We can predict effects of point mutations on MAO A reactivity. This has clinical significance for understanding depression, manias e.g. warrior gene.



WT

Perspectives

Clinically relevant point mutations

Design of novel reversible inhibitors based on the transition state structure (Achieving MAO A/MAO B selectivity will be very demanding task !)

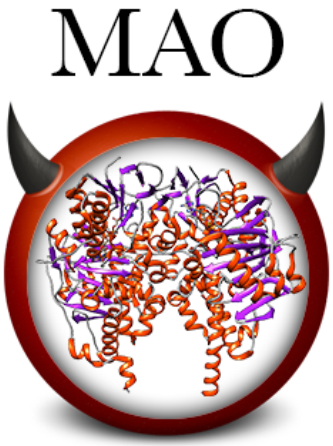
Calculation of H/D kinetic isotope effect for MAO B

Chemical step of MAO inhibition needs to be addressed by QM/MM methodology

Simulation of deamination step

Simulation of oxidative half-reaction

Work in progress: MAO A catalyzed decomposition of noradrenaline and substituted benzylamines



The diabolical enzyme.

Beware of improper models

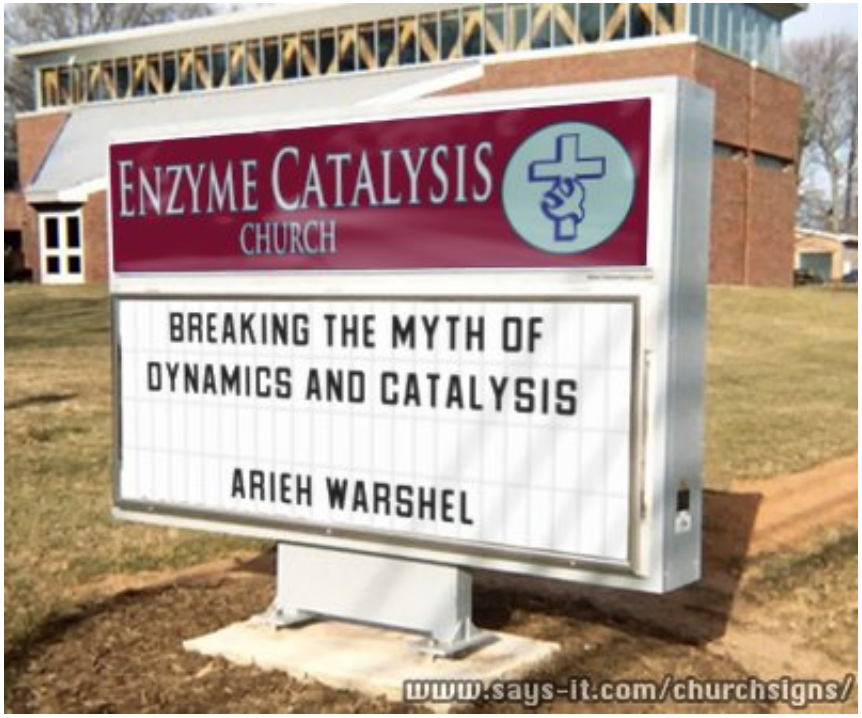
Dynamical and other esoteric effects do not contribute to catalysis.

http://rosettadesigngroup.com/blog/649/is-dynamics-the-missing-link-for-understanding-enzyme-catalysis/

2010

Is dynamics the missing link for understanding enzyme catalysis?

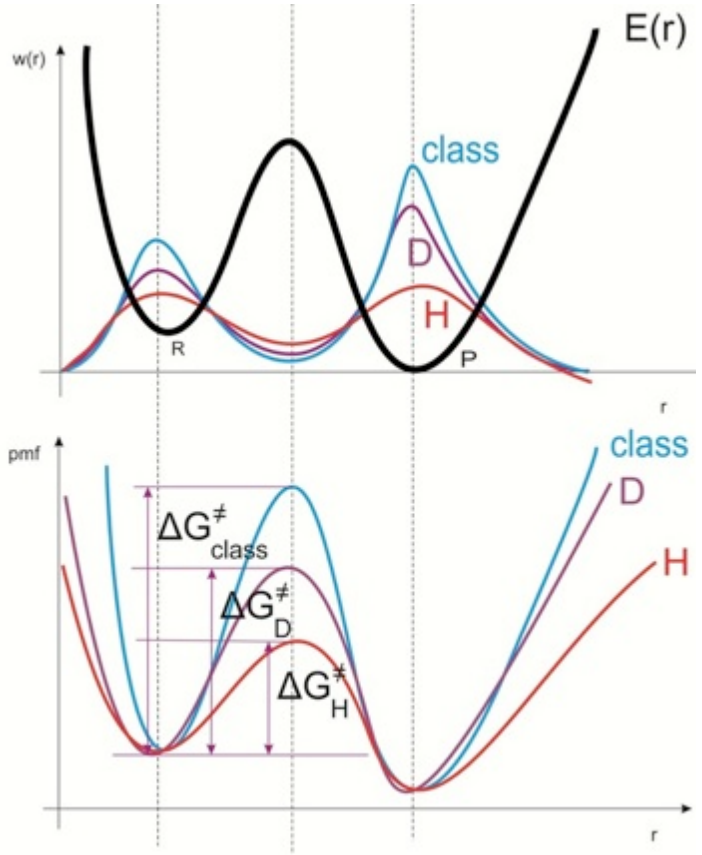
How do enzymes catalyze reactions? There are countless answers of course, but one answer that has gained much attention and popularity in recent years is – through intrinsic dynamics. Is that so? PNAS recently published a paper by Arieh Warshel entitled: “Enzyme millisecond conformational dynamics do not catalyze the chemical step”. Warshel, an avid assailant of the coupling between dynamics and catalysis was met by Martin Karplus, devoted advocate for catalytic dynamics, to engage in a public dispute over the letters section of PNAS. Who do you find more convincing?



Tunneling is change of the probability density at the region of the barrier relative to the classical treatment of the particles due to quantum nature of their motion.

For chemical reactions this means effective increase of the probability density at the barrier region giving rise to lowering of the barrier in terms of free energy and increased rate of crossing.

Tunneling is not a dynamical phenomenon and is perfectly compatible with the transition state theory. Technically speaking it can be calculated also by Monte Carlo methodology.



Dynamical effects are deviations from the transitions state theory due to barrier recrossing.

Take home messages:

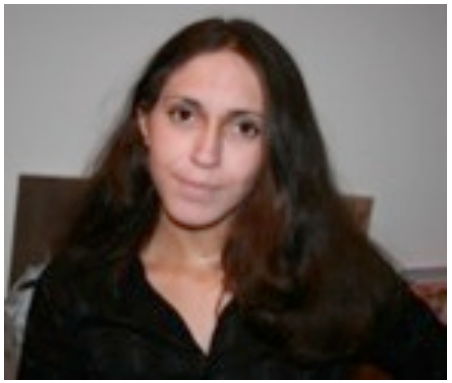
Currently, classical molecular dynamics of hydrated proteins is possible using regular PCs (using both CPU and GPU) and free software on a time scale of hundreds of nanoseconds.

One can gain atomic level insight into structure, stability and dynamics of hydrated proteins. It is possible to add additional environment complexity by including membrane, counterions and nucleic acid.

Ligand binding affinites can be simulated by using free energy calculations.

Studies of chemical reactivity needs application of QM/MM methods.

Arieh Warshel,
USC



Lynn S.C. Kamerlin
University of Uppsala, Sweden

Robert Vianello,
IRB, Zagreb

Matej Repič

Miha Purg

Without these fine coworkers computational enzymology in our laboratory would not be on the same level.

$ Ministry for Science and Higher Education of Slovenia,
grants P1-0012, J1-2014, P3-067
FP7 People-Marie Curie
COST CM1103


n.v. 2200 m
Triglavski dom
1h 15min
Triglav
2h 15min
Mercator in Procter&Gamble
PD Ljubljana-Matica
n.v. 2200 m
Aljažev dom
2h 45min
Mercator in Procter&Gamble
PD Ljubljana - Matica