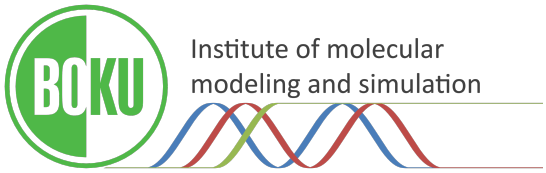
Molecular dynamics simulations



Chris Oostenbrink
Institute of Molecular Modeling and Simulation
Department of Material Science and Process Engineering
University of Natural Resources and Life Sciences, Vienna (BOKU)

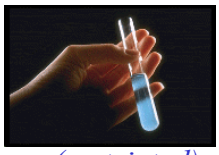
When is computational modeling useful ?

Simulation can replace or complement the experiment:

1. Experiment is impossible	*Inside of stars* *Weather forecast*
2. Experiment is too dangerous	*Flight simulation* *Explosion simulation*
3. Experiment is expensive	*High pressure simulation* *Windchannel simulation* *Trial and error drug design*
4. Experiment is blind	*Some properties cannot be observed on very short time-scales and very small space-scales*

Molecular simulation and experiment

experiment



(restricted)

simulation

(unrestricted)

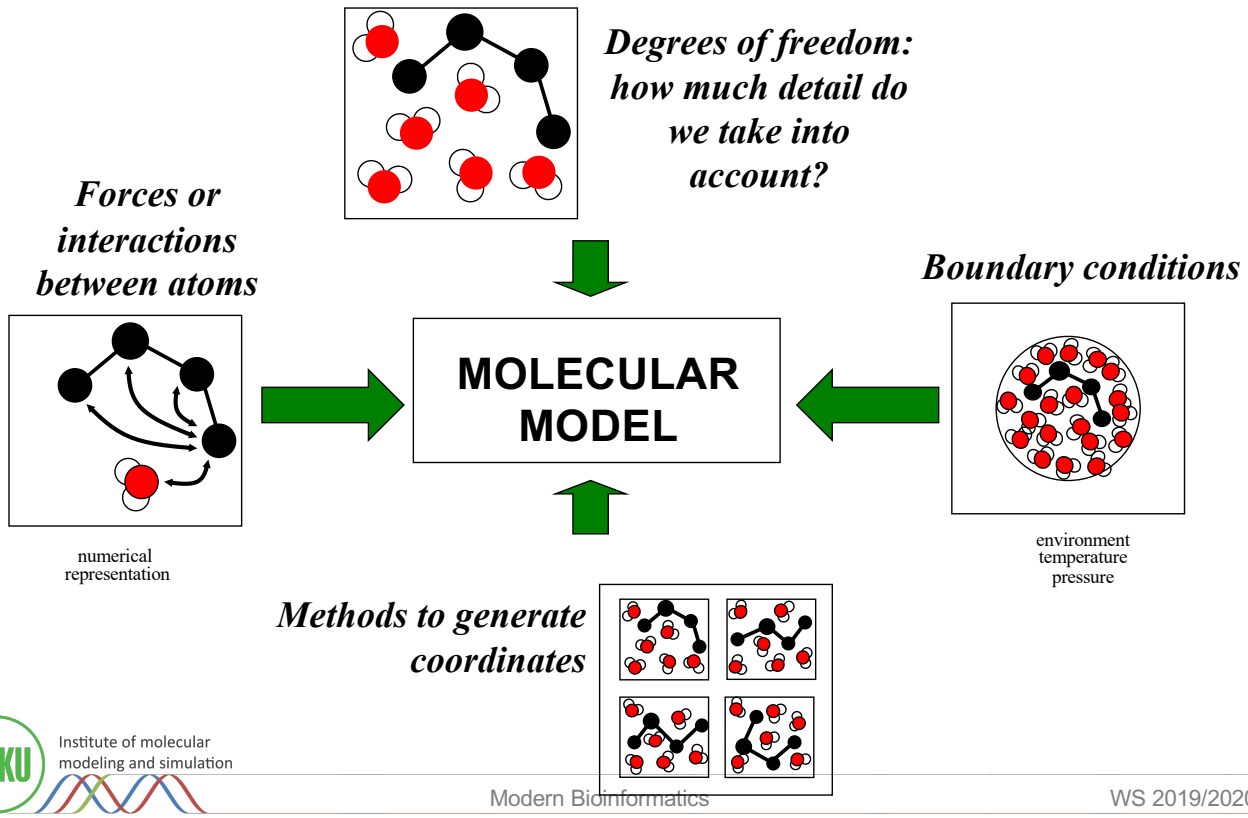
Resolution*

size :	10^{23} molecules	1 molecule
time :	1 second	10^{-15} seconds

*: Single molecules / 10^{-15} seconds possible
(but not both in the liquid phase)

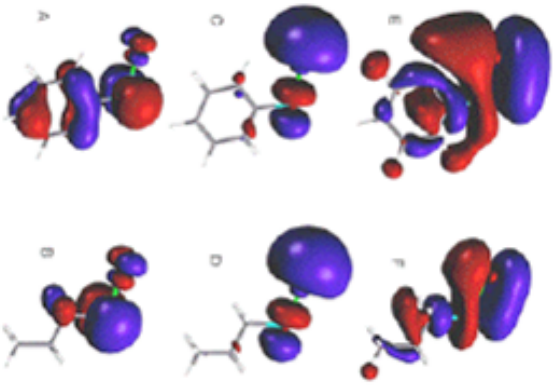
Simulation and experiment are complementing methods to study different aspects of nature

A model for molecular computations



A molecule has a certain energy

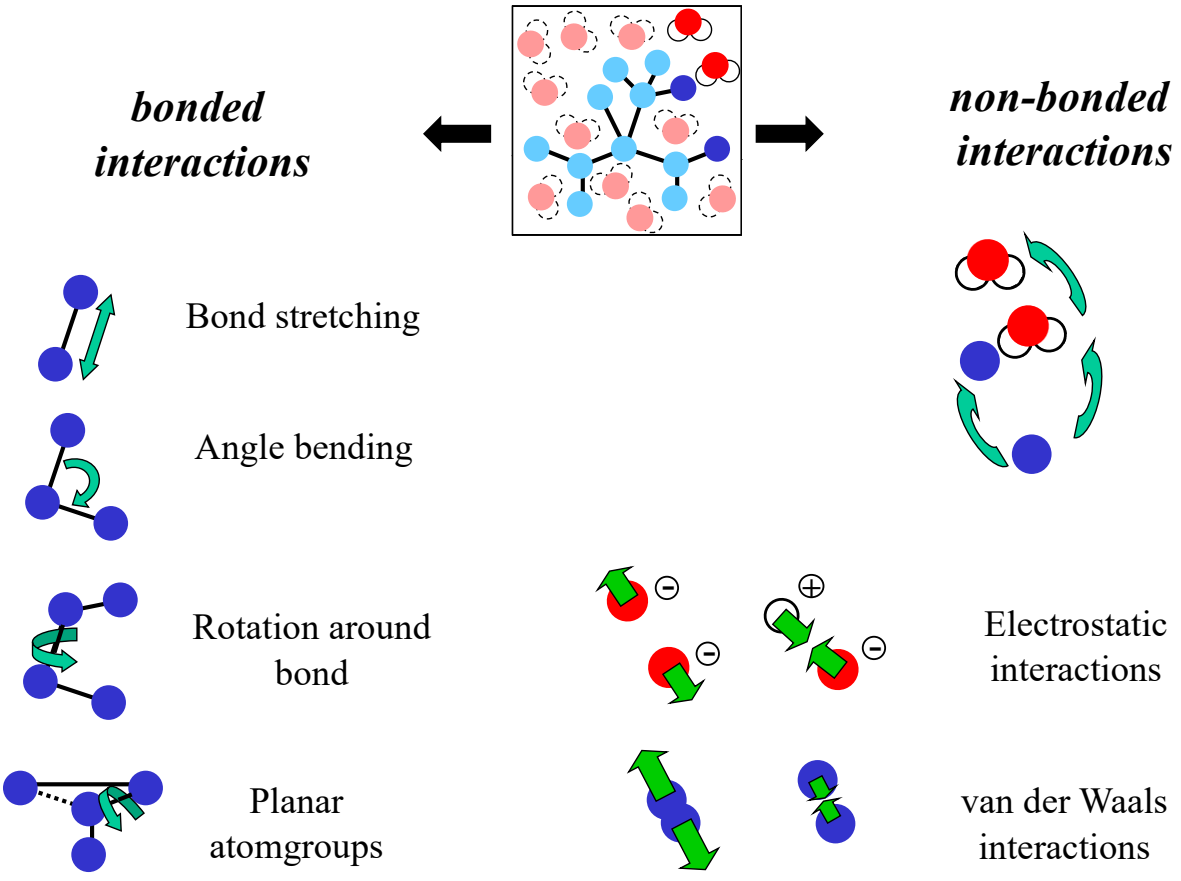
- Point charges with an electron cloud around it
  - Quantum mechanics, ab initio or semi-empirical



$$\hat{H}\psi(\mathbf{r}) = E\psi(\mathbf{r})$$

- Collection of balls and springs:
  - Molecular mechanics, force field representation

Molecular mechanical interactions



Interacting Particles

Physical Terms

$$V^{bond}\left(\mathbf{r}^N\right)=\sum_{bonds\ i}\frac{1}{2}K_i^b\left[b_i\left(\mathbf{r}^N\right)-b_i^0\right]^2$$

$$V^{v.d.Waals}\left(\vec{r}^{\,N}\right)=\sum_{pairs\ i<j}4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

$$V^{angle}\left(\mathbf{r}^N\right)=\sum_{angles\ i}\frac{1}{2}K_i^a\left[\theta_i\left(\mathbf{r}^N\right)-\theta_i^0\right]^2$$

$$V^{Coulomb}\left(\vec{r}^{\,N}\right)=\sum_{pairs\ i<j}\frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_iq_j}{r_{ij}}$$

$$V^{torsion}\left(\mathbf{r}^N\right)=\sum_{torsion\ i}K_i^{\varphi}\left[1+\cos\left(m_i\varphi_i\left(\mathbf{r}^N\right)+\delta_i\right)\right]$$

$V^{pol}\left(\vec{r}^{\,N}\right)=$ N-body polarization energy

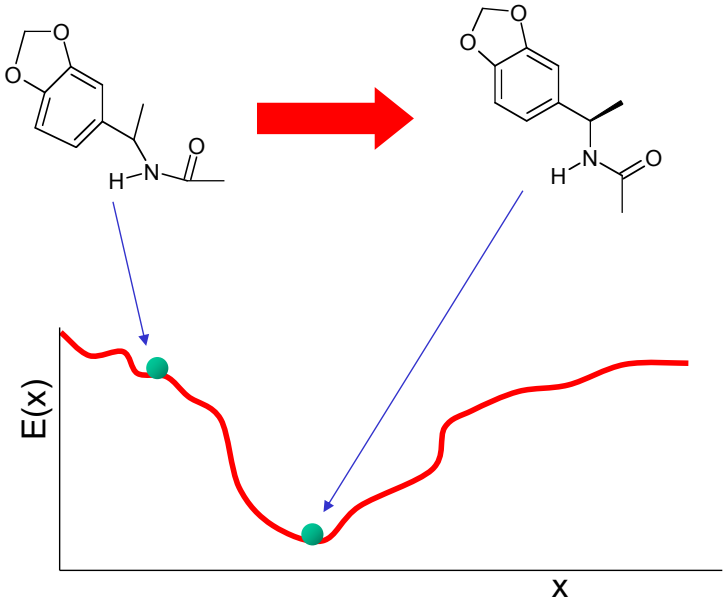
$V^{ext}\left(\vec{r}^{\,N}\right)=$ external fields energy

Special Interaction Terms examples

- restraints on the system:
  - from experimental data
  - to bias the sampling

Energy minimisation

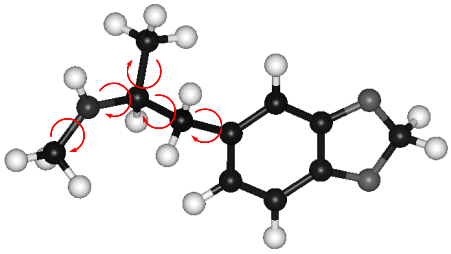
- Find the lowest-energy conformation of a molecule



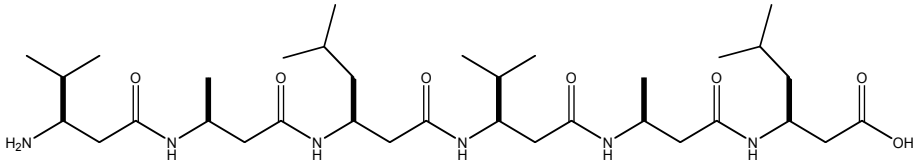
- Compare to a marble rolling down a slope

Different conformations

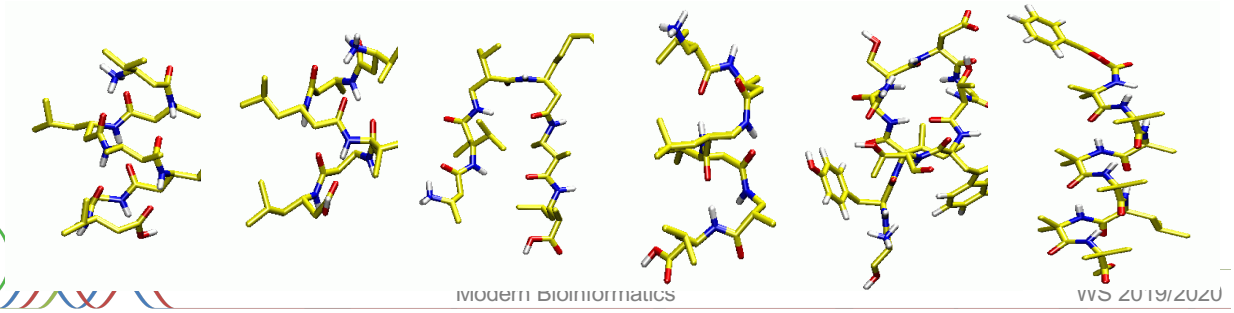
- Rotate around bonds



- One compound



- Many different conformations



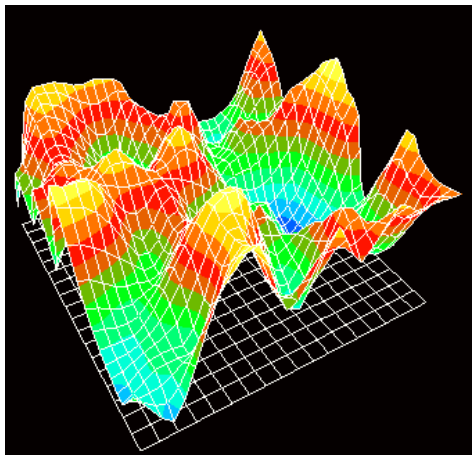
Different conformations

- Every conformation is associated with an energy, as a function of the positions of all particles, $\mathbf{q} = (x_1,y_1,z_1,x_2,y_2,z_2,\ldots)$

  $E = f(\mathbf{q}) = f(x_1,y_1,z_1,x_2,y_2,z_2,\ldots)$

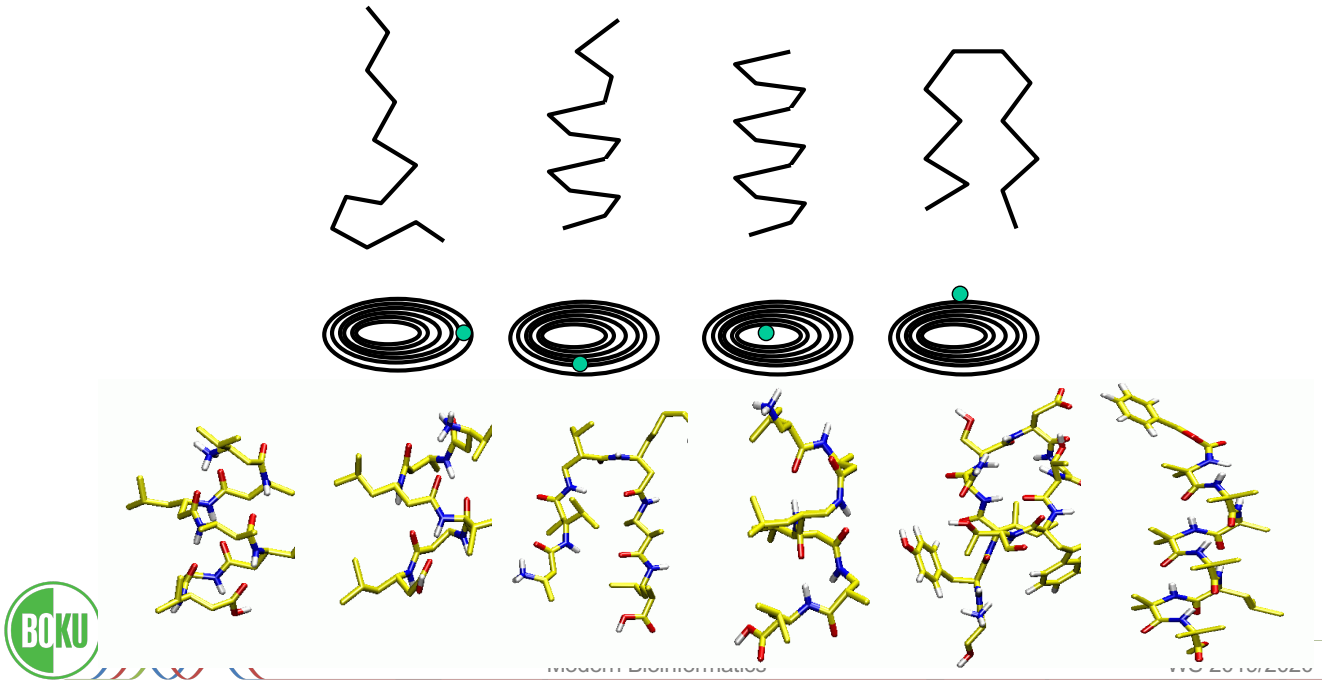
- Compare **q** to a point on a multi-dimensional energy surface (3N-6)-dimensional

- Minima are favourable conformations
- Saddel points are transition states

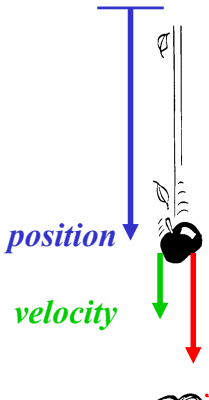


Conformations on the surface

- Every conformation is represented by a specific point on the 3N-6 dimensional surface



Classical laws of motion



Situation at time t

Force is determined by relative positions

acceleration = force / mass

Δ velocity = acceleration × Δ t

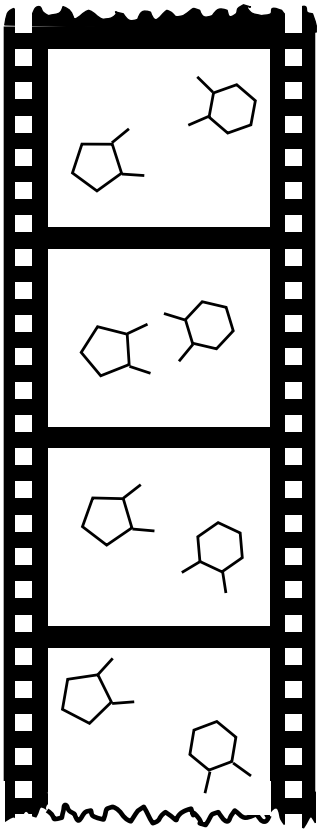
Δ position = velocity × Δ t

Situation at time t+Δt

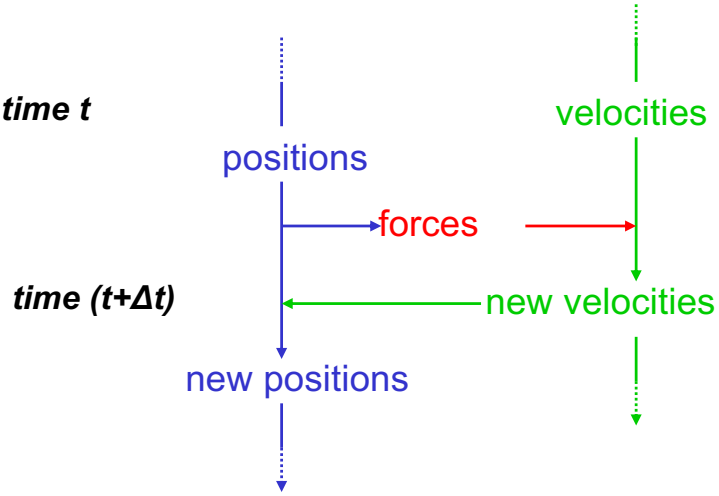
Sir Isaac Newton
1642 -1727

Determinism …

Molecular dynamics



... Comparable to shooting a movie of molecular motion...



Leap frog algorithm

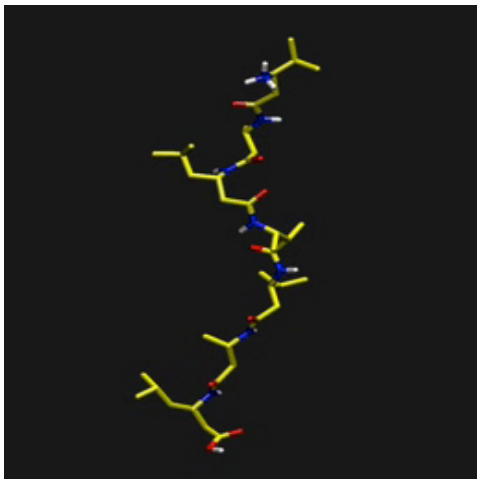
History

Year	molecular system: type, size	length of the simulation in seconds
1957	first molecular dynamics simulation (hard discs, two dimensions)	
1964	atomic liquid (argon)	10^{-11}
1971	molecular liquid (water)	$5 \cdot 10^{-12}$
1976	protein (no solvent)	$2 \cdot 10^{-11}$
1983	protein in water	$2 \cdot 10^{-11}$
1989	protein-DNA complex in water	10^{-10}
1997	polypeptide folding in solvent	10^{-7}
2001	micelle formation	10^{-7}
2010	folding of a small protein	10^{-6}

Folding simulation

- Proteins are too large systems to simulate the slow folding process.
- Smaller model compounds can be correctly folded on the computer.

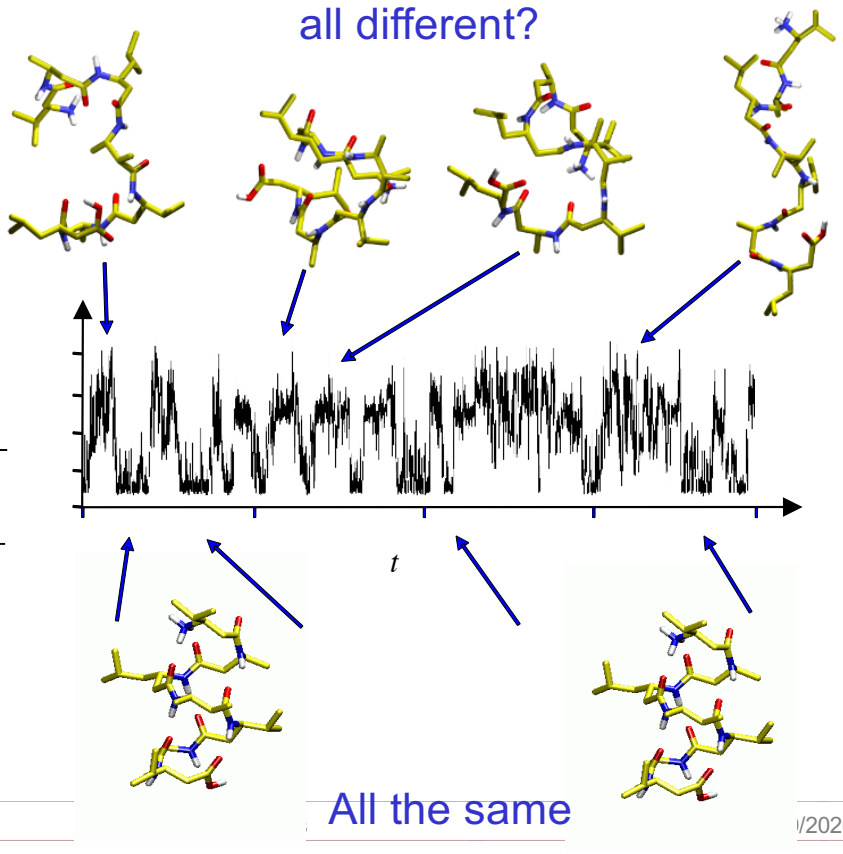
⇨ Information about folding mechanisms and the unfolded state



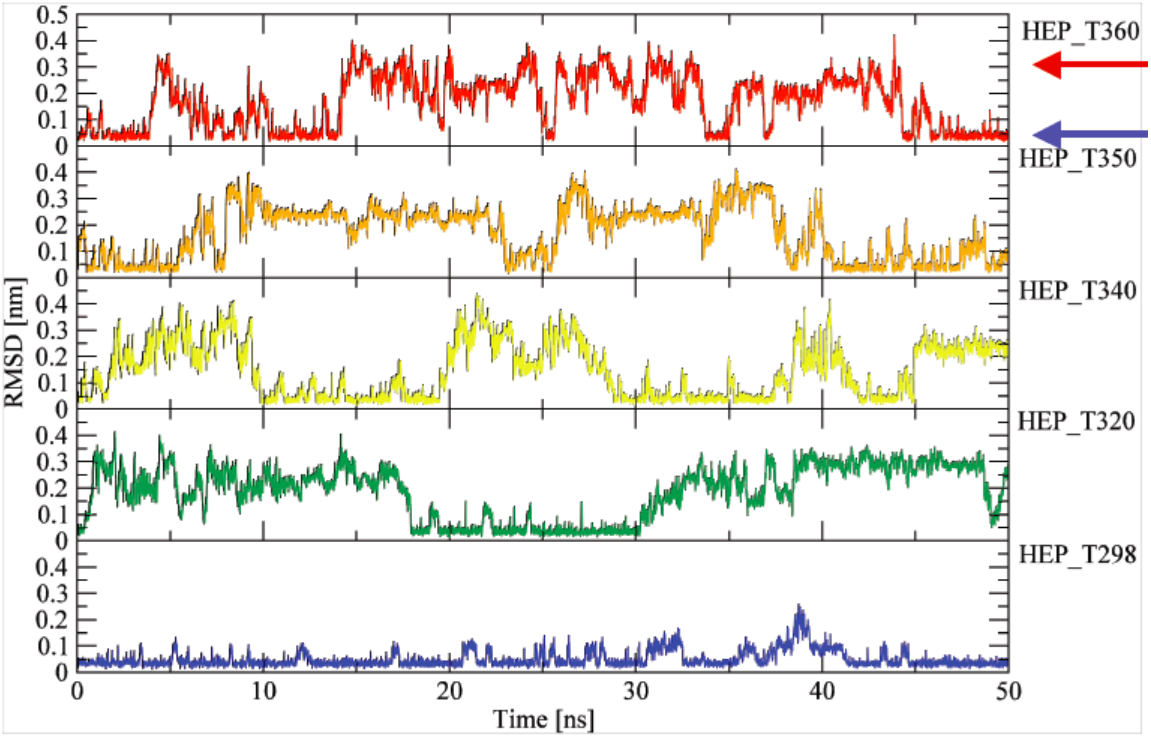
RMSD: Root mean square deviation

- A measure to compare two structures
- Here we compare the structures seen in the simulation to the experimentally determined 'folded' structure

$$D = \sqrt{\frac{\sum_i \left|\mathbf{q}_i - \mathbf{q}_i^{ref}\right|^2}{N}}$$

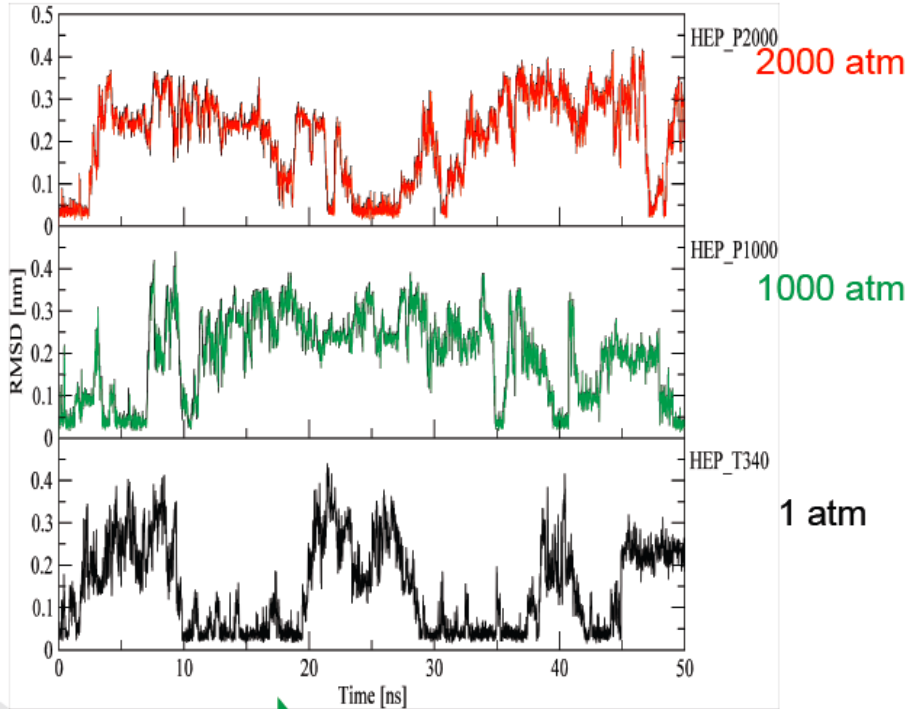


Temperature dependency



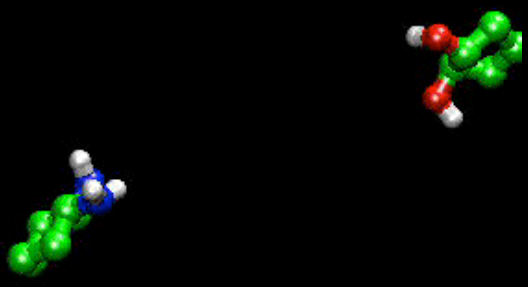
folding equilibrium depends on temperature

Pressure dependency



folding equilibrium depends on pressure

Diol + Diamine + 252 CCl_4 Molecules
$2.1 – 2.2 \cdot 10^{-9}$ seconds



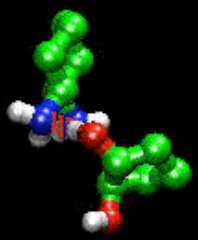
Complex formed

Diol + Diamine + 252 CCl_4 Molecules
$3.2 – 4.0 \cdot 10^{-9}$ seconds

Hydrogen bonds

O ➔ N

N ➔ O

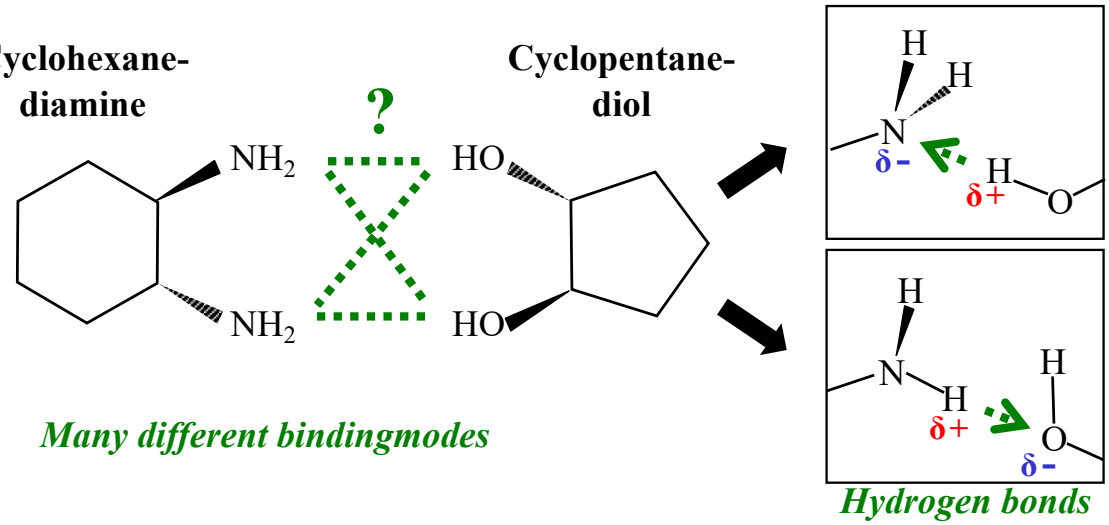


the molecules are free again…

Binding equilibrium of two small molecules

Complex :

Cyclohexane-diamine ? **Cyclopentane-diol**



Many different bindingmodes

Hydrogen bonds

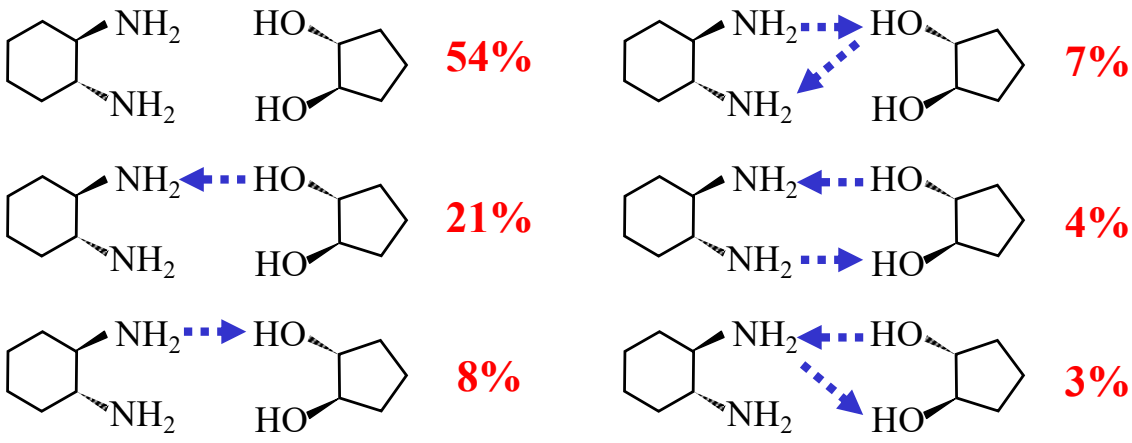
Average binding strength (free enthalpy) :

	Experimental		*MD simulation*
	Benzene	*CCl_4*	
ΔG_b [kJ/mol]	**-9.3**	**-11.5**	**-10.4**

Results of the simulation

Experimentally hardly (or not) possible !

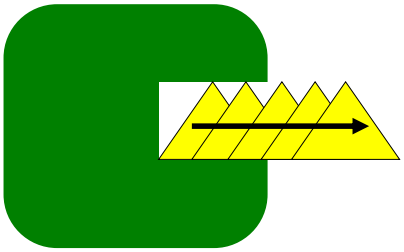
Occurrence of different binding modes :

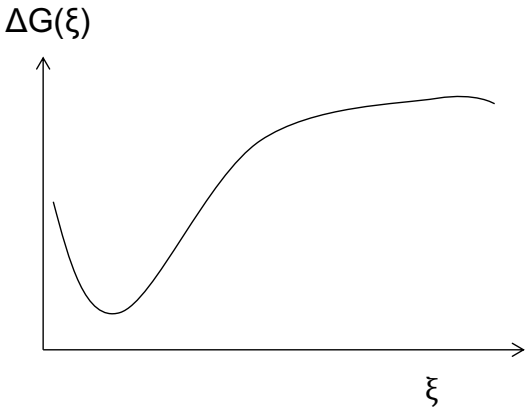


Life time :

- Average life time of the complex: $2 \cdot 10^{-10}$ sec (max. $3 \cdot 10^{-9}$ sec)
- Average life time of a hydrogen bond: $5 \cdot 10^{-12}$ sec

Ligand binding

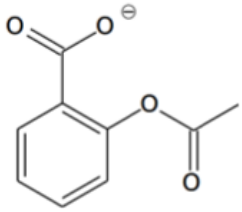




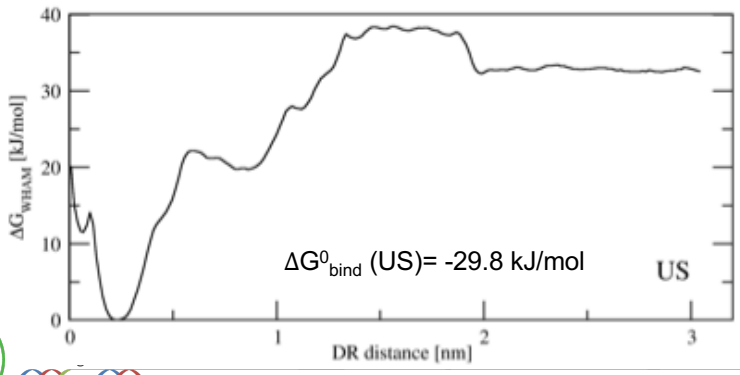
- We can try to calculate the (un)binding of a ligand
- Calculate the potential of mean force along the reaction coordinate
- Binding free energy is difference between bound and unbound values
- Information about the binding processes

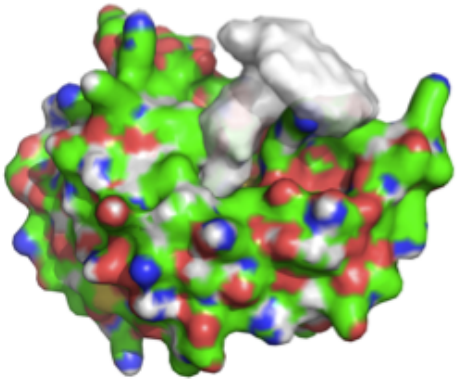
Binding processes

- Aspirin binding to cytosolic Phospholipase 2
- Umbrella sampling with distance restraints from the active site
  GROMOS11, GROMOS 54A7 force field
  31 x 10 ns, 300 K, 1 atm, SPC water



- Weighted histogram analysis (WHAM)
- Barriers along the way

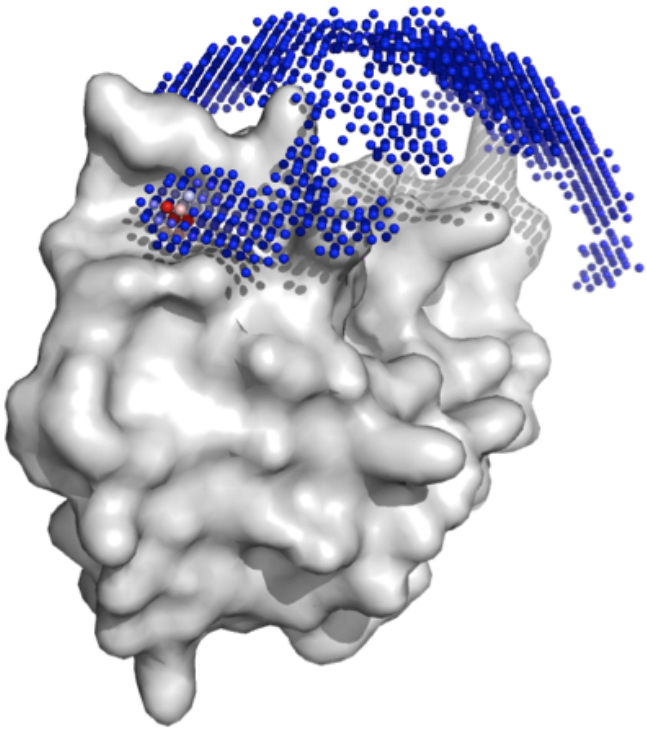




ΔG^0_{bind} (exp)= -29.6 kJ/mol

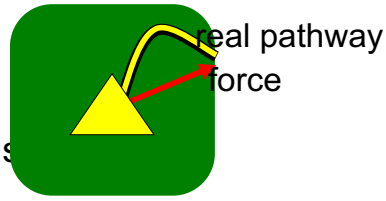
Barriers: real or artifact?

- Centre of mass of aspirin
- Seems to get stuck behind a part of the protein
- Resolve
  - Single path
  - Reversible binding

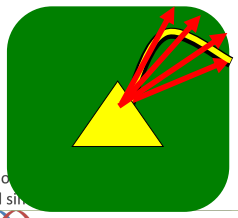


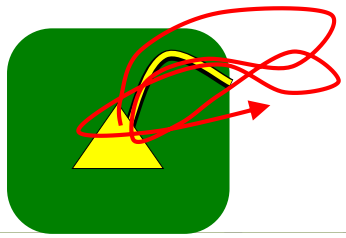
Potential of mean force

- Pulling along a ‘wrong’ path will give the correct free energy difference
  - In the limit of infinite sampling
  - In practice, the value is very path dependent
- Multiple paths and orientations play a role
- We want to simulate the ensemble of possible paths



- Possible solutions:
  - Pull the molecule out many times
  - Enhanced sampling (REMD, Local Elevation, …) to bind reversibly





Replica exchange MD

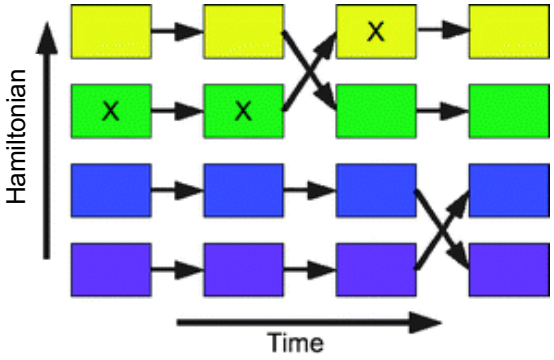
- Run simulations at different conditions
- Mix them using the Metropolis criterion (MC)
- For each of the simulations you get a correct ensemble

- Replicas differ in, λ-dependent, Hamiltonian

$$H(\mathbf{p},\mathbf{r},\lambda) = K(\mathbf{p}) + V^{phys}(\mathbf{r}) + V^{rest}(\mathbf{r},\lambda)$$

$$V^{rest}(\mathbf{r},\lambda) = \tfrac{1}{2} K\left[(1-\lambda) r_0^A + \lambda r_0^B - r_{ij}\right]^2$$

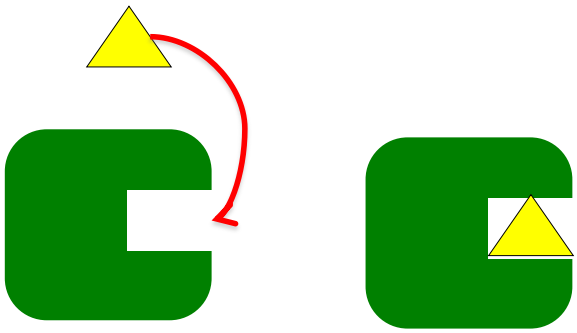
- At large distances, the ligand diffuses
- Returns via a different pathway
- Broad ensemble at every λ
- Round trips: reversible binding



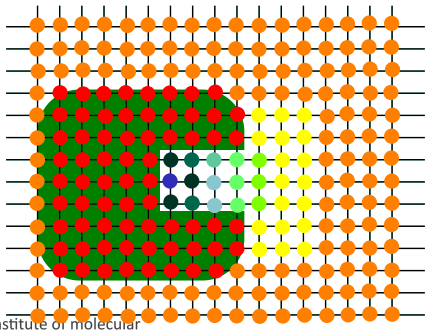
Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **113**, 6042–6051 (2000)
Figure: A. Patriksson, D. van der Spoel, Phys. Chem. Chem. Phys., **10**, 2073 (2008)

Distancefield coordinate

- Calculate the shortest route *not* through the protein



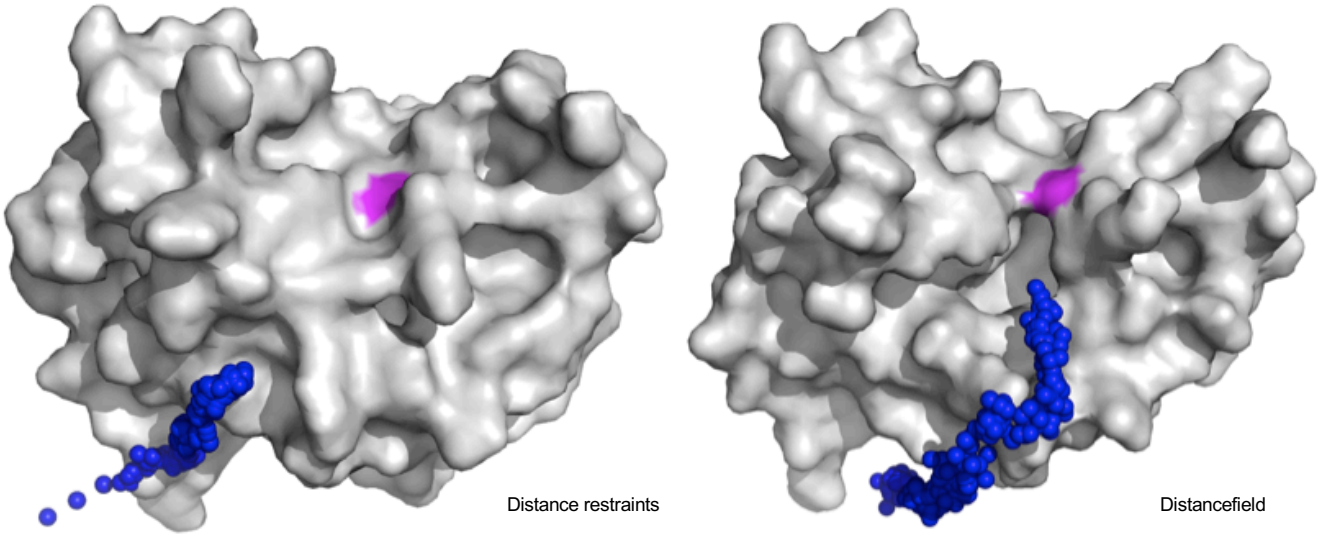
- Use a grid



Dijkstra's algorithm to find the shortest way
- Find nearest grid point to a reference position
- For all neighboring points assign distance
  - Penalty when moving into the protein
- Move to the next point with the smallest distance
- Periodicity is automatically taken care of

Back to the example of aspirin

- Distance restraints push into the protein and distort structure
- Distancefield restraints curve around the protein



Application of distancefield

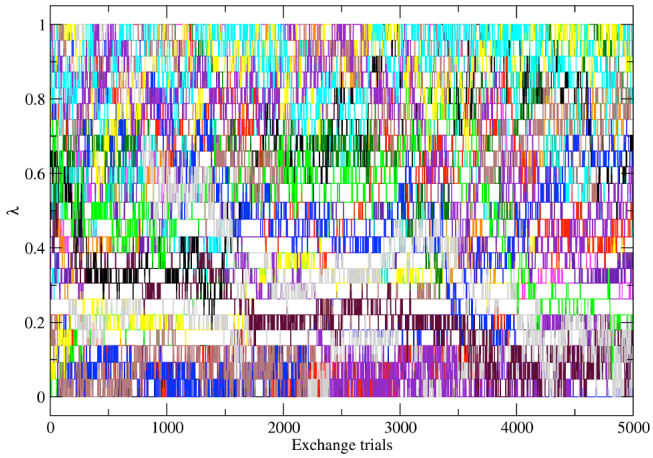
Distancefield coordinate allows for reversible binding / unbinding

Various applications implemented in GROMOS

Local elevation / Metadynamics

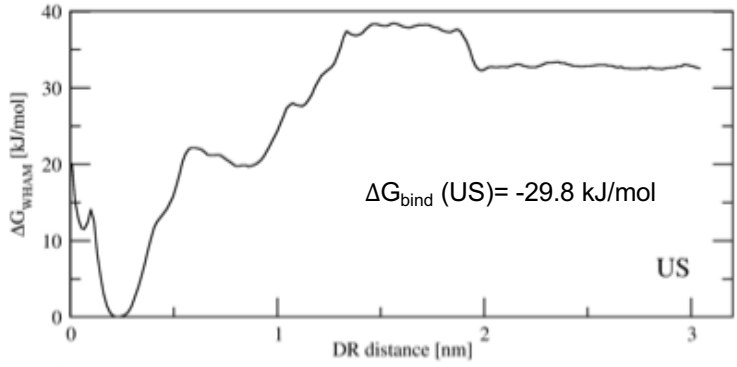
Hamiltonian replica exchange

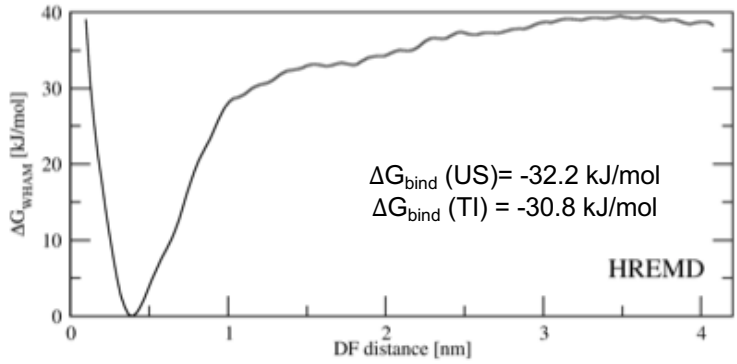
24 replicas
restraints at different distances
alternating switching time 2 ps
10 ns per replica

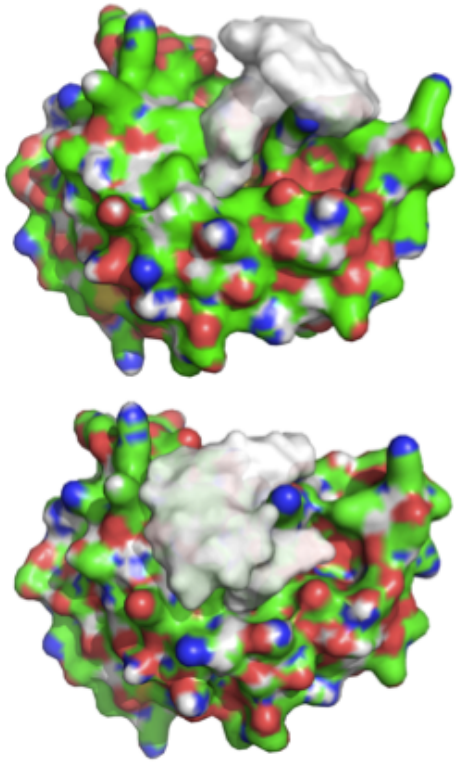


De Ruiter and Oostenbrink, *J. Chem. Theory Comp.* (2013) 9:883

Resulting PMF and routes



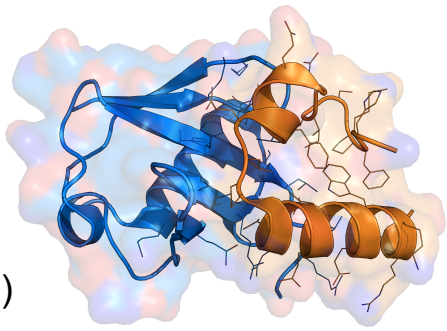




De Ruiter and Oostenbrink, *J. Chem. Theory Comp.* (2013) **9**:883

Protein-protein interactions

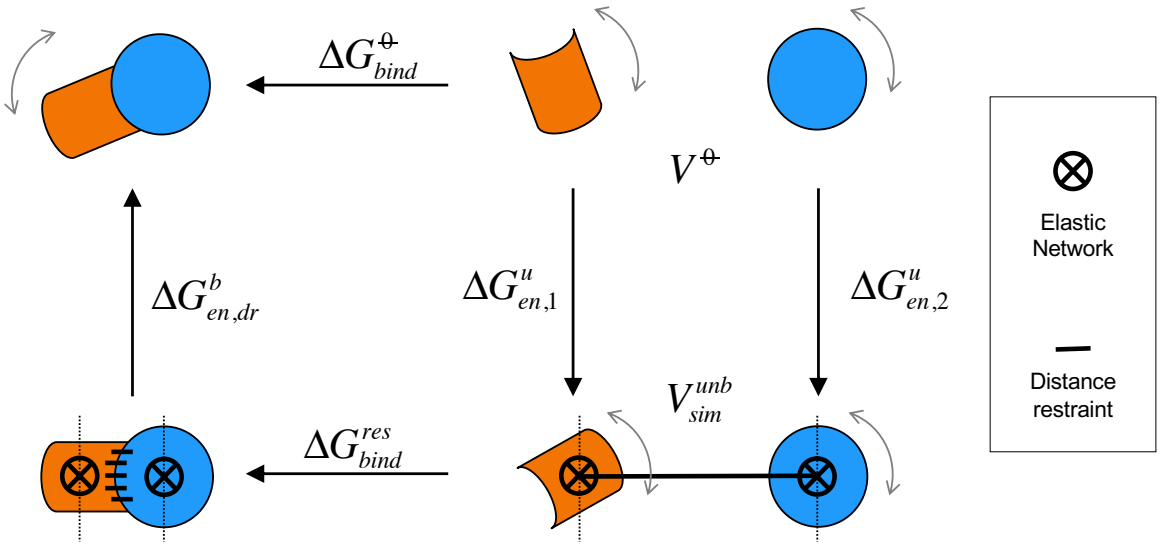
- Model system: Ubiquitin-UBM2
  - Experimental (NMR) structure available



- To achieve reversible binding:
- 3 sets of λ-dependent distance restraints
  - 12 between Cα at the binding site (“specific”)
  - 1 between Cα-COMs of binding partners
  - 2 elastic networks on each binding partner
    - corresponding to a snapshot from the bound complex
    - Cα-Cα distance restraints between 0.4 and 0.9 nm

- 54A8 ff, modified Gromacs 5.1.2, 1.4 nm cut-off, reaction-field, NPT, 300 K, 1 bar, SPC water, 150 mM NaCl

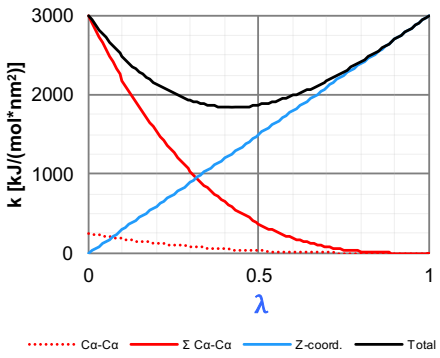
Thermodynamic cycle



$$\Delta G^{\ominus}_{bind} = \Delta G^{res}_{bind} + \Delta G^{b}_{en,dr} + \Delta G^{u}_{en,1} + \Delta G^{u}_{en,2} + \Delta G^{\ominus}_{corr}$$

Binding/unbinding

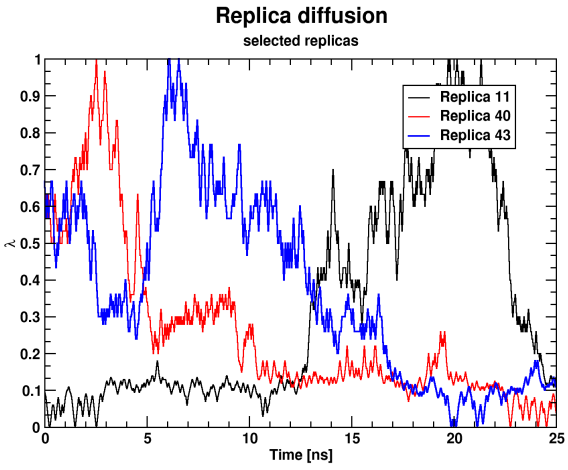
- Binding process (ΔG^{res}_{bind}) simulated in z-coordinate only or radially
- increase in distance of restraints from 0 to 2.5 nm (λ = 0 to λ = 1)
  - specific C-C distance restraints are turned off (n = 0, m = 2)
  - COM-COM distance restraint is turned on (linearly)

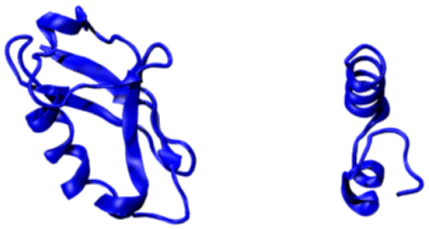


- HREMD with time between switching attempts of 20ps
- optimized λ-spacing
  - replica diffusion should give “round-trips”
  - 54 unequally spaced replicas

Binding/unbinding

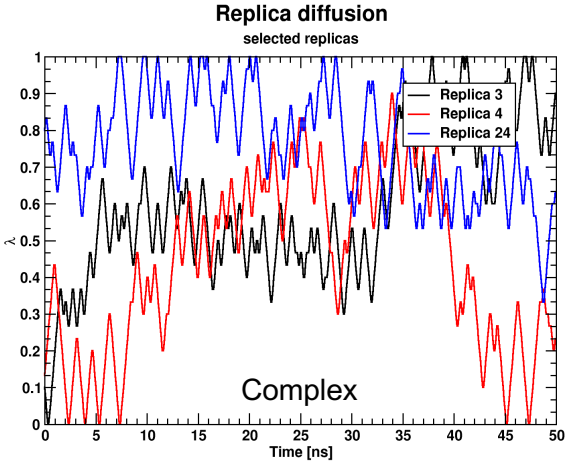
- Binding/unbinding in radial or z-coordinate (ΔG_{bind}^{res})
  - Similar results
- 50 ns of H-REMD in 54 replicas
  - Free energy:
    - thermodynamic integration over λ
    - Bennets acceptance ratio

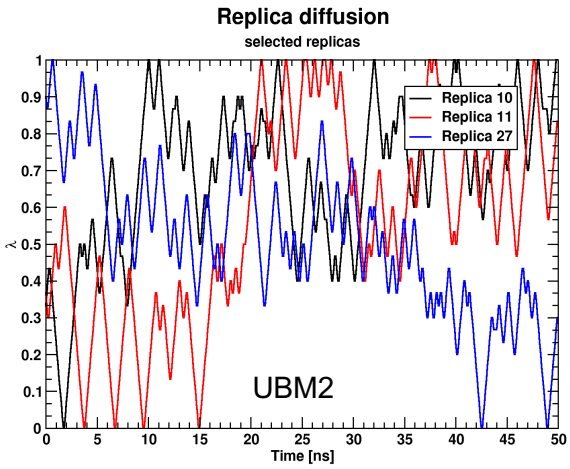




Turning on/off elastic network

- Turn on elastic network C-C restraints from λ = 0 to λ = 1
  - specific C-C distance restraints are also turned on in the complex, $\Delta G_{en,dr}^{b}$
  - all restraints are soft at λ < 1
- HREMD with time between switching attempts of 100 ps
  - 31 equally spaced replicas





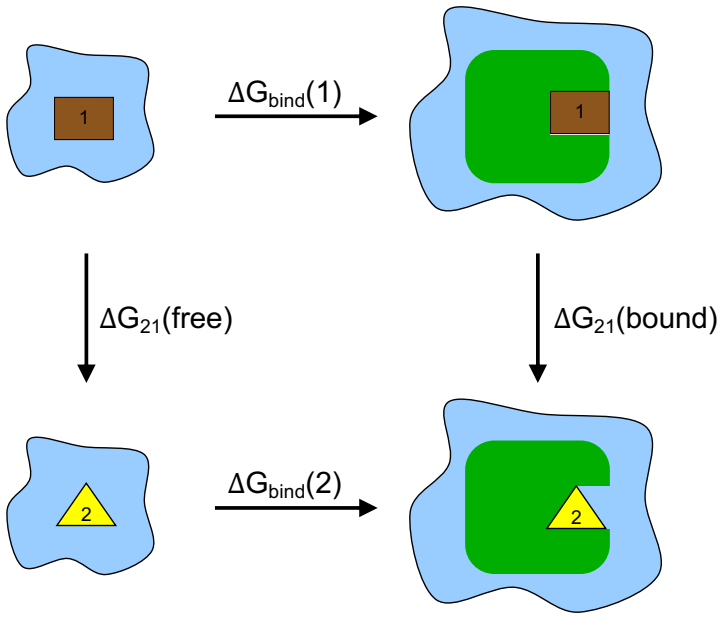
Summary ΔG^0_{bind} (kJ/mol)

System/Experiment	ΔG^{res}_{bind} (incl. cor.)	$\sum \Delta G^{b/u}_{en}$	$\Delta G^{\ominus}_{bind}$
Simulation WT RS	-36.2 ± 1.1	+10.1 ± 2.1	-26.1 ± 2.4
Simulation WT ZS	-32.6 ± 2.8	+10.1 ± 2.1	-22.5 ± 3.5
Simulation WT ZL	-35.5 ± 2.1	+10.1 ± 2.1	-25.4 ± 3.0
Experiment: WT ITC (*Cui et al. 2010*)			**-25.1**
Simulation P692A RS	-33.2 ± 0.6	+11.4 ± 2.3	-21.8 ± 2.3
Simulation P692A ZS	-31.6 ± 1.8	+11.4 ± 2.3	-20.2 ± 2.9
Simulation P692A ZL	-33.9 ± 1.9	+11.4 ± 2.3	-22.5 ± 3.0
Experiment: P692A ITC (*Cui et al. 2010*)			**-20.4**

Perthold and Oostenbrink, *J. Chem. Theory Comp.* **13** (2017) 5697 - 5708

Thermodynamic cycle for binding

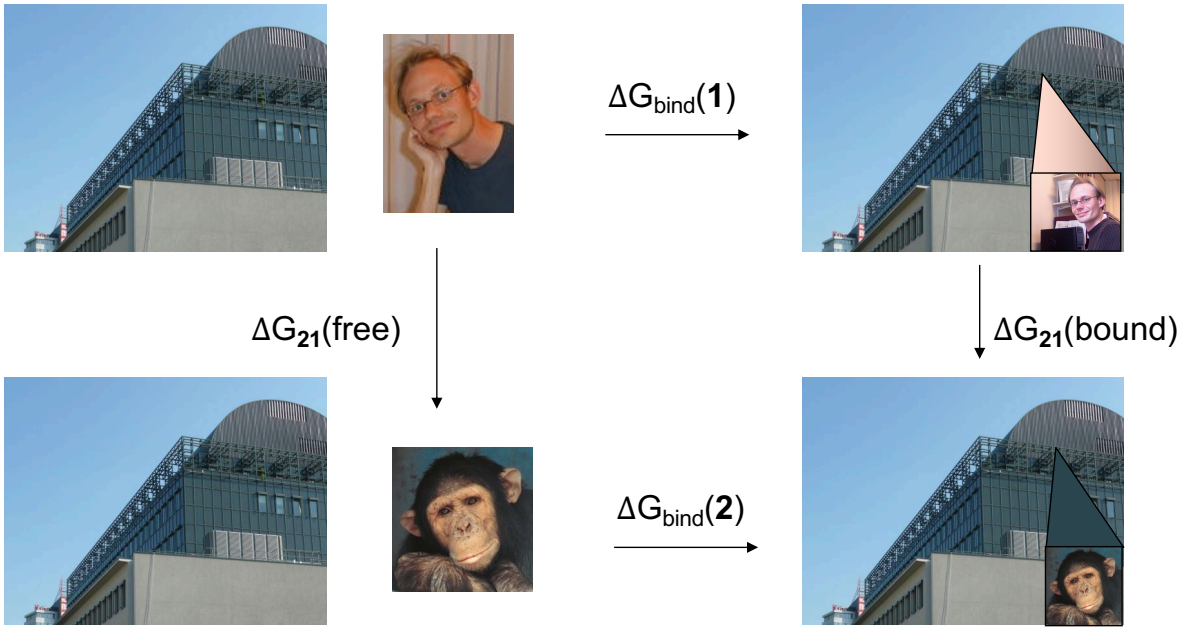
- Free energy is independent of the path (state function)
- Thermodynamic cycle
- Relative free energies
- Computational alchemy



$$\Delta\Delta G_{bind} = \Delta G_{bind}(2) - \Delta G_{bind}(1)$$
$$= \Delta G_{21}(bound) - \Delta G_{21}(free)$$

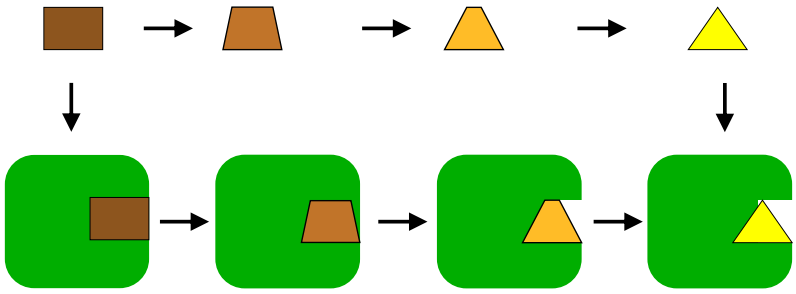
Who fits better at BOKU?

- Are there others that are more suitable?

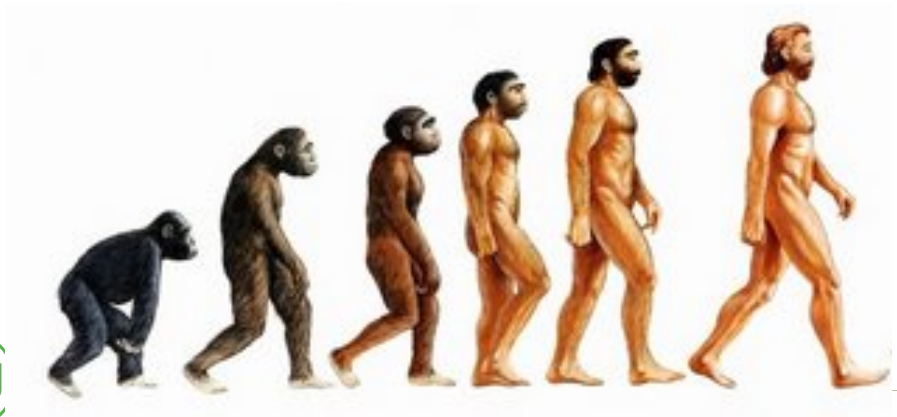


- Compare two employees when they are free and at BOKU

Gradually change one in the other



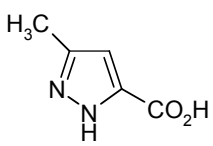
Change ligand 1 into ligand 2, in solution and when bound to the protein



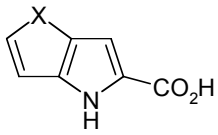
As long as the end-states are defined, the intermediates do not have to be physically possible

Example: DAAO inhibitors

- Three inhibitors of the enzyme D-amino acid oxidase were studied



1



3 X = O
4 X = S

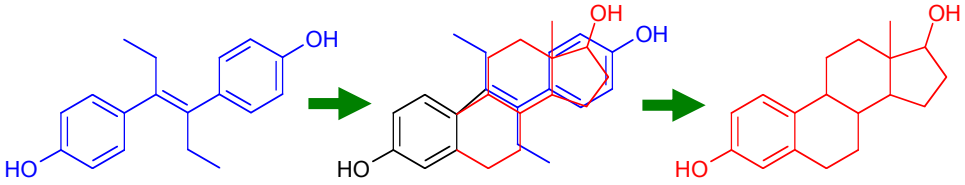
	3->1	**3->4**	**4->1**
Calculated values:			
ΔG_{free}	106.3 ±1.5	86.1 ±0.8	20.4 ±1.1
$\Delta G_{complex}$	113.8 ± 2.2	87.3 ±3.5	36.7 ±2.0
$\Delta\Delta G_{bind}$	7.5 ± 3.7	1.2 ± 4.3	16.3 ± 3.1
Experimental $\Delta\Delta G_{bind}$ based on:			
IC_{50}[a]	8.2	-0.9	9.1
IC_{50}[b]	4.6	0.1	4.6
ITC	9.4	0.8	8.6
SPR[c]	14.1	1.6	12.4

Overall, the relative binding free energies are very well reproduced

J.H.M. Lange, J. Venhorst, M.J.P. van Dongen, J. Frankena, F. Bassissi, N.M.W.J. de Bruin, C. den Besten, S.B.A. de Beer, C. Oostenbrink, N. Markova and C.G. Kruse, *Eur. J. Med. Chem.* (2011) **46**, 4808 - 4819

Computational alchemy

- Modify one compound into another one in small steps



$$E(\mathbf{q},\mathbf{p},\lambda) = (1-\lambda)E_A(\mathbf{q},\mathbf{p}) + \lambda E_B(\mathbf{q},\mathbf{p})$$

- In a formula:

$\lambda = 0 \rightarrow E = E_A \qquad \lambda = 1 \rightarrow E = E_B$

Along the way? The protein ‘sees’ a mixture of A and B

$$\Delta G_{AB} = \sum_{\lambda=0}^{1} -k_B T \ln \left\langle e^{-\Delta E(\lambda \rightarrow \lambda + d\lambda)/k_B T} \right\rangle$$

Example: ER

- Relative free energy of three compounds
- In three different media (vacuum, solution, protein)
- In 11 discrete steps, forward and backward TI

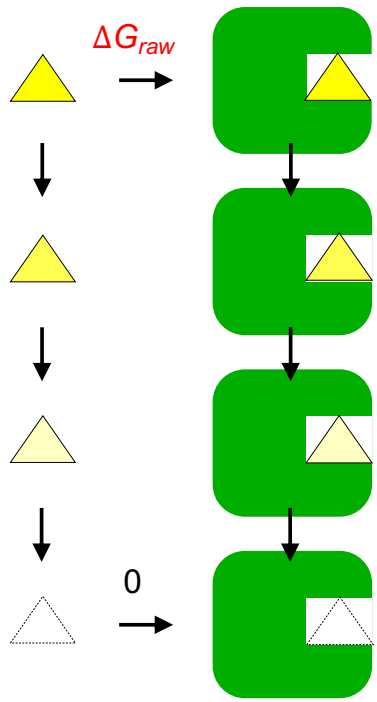
Table 4. TI Results (kJ mol^{-1})[a]

	DES ↔ E2			DES ↔ GEN		
TI	forward	backward	hysteresis	forward	backward	hysteresis
vacuum	76.3	76.1	0.2	187.1	186.9	0.2
solvent	79.0	81.6	−2.6	151.5	157.3	−5.8
protein	80.4	78.2	2.2	173.1	165.3	7.8
$\Delta\Delta G_{solv}$	2.8	5.5	−2.7	−35.6	−29.5	−6.0
$\Delta\Delta G_{bind}$	1.4	−3.4	4.8	21.6	8.0	13.6
$\Delta\Delta G_{bind}$	3.8[b]			11.3[b]		
(expt)	0.79[c]			21.69[c]		

Aspirin corrections

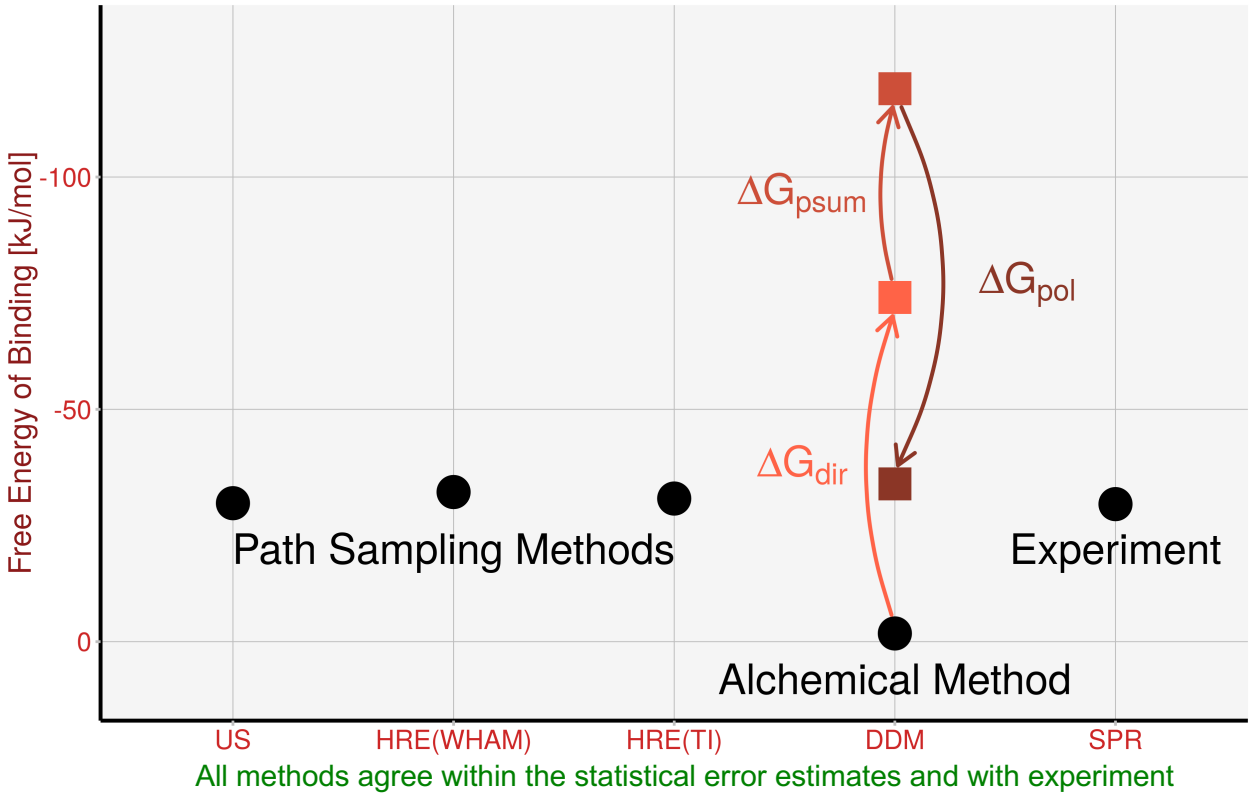
- Binding affinity of Aspirine to phospholipase A2
- Thermodynamic integration to remove the interactions with the surroundings
- Three independent sets of simulations
- Correcting for electrostatic artifacts

ΔG_{raw}	1.1 kJ/mol	
ΔG_{dir}	-70.8 kJ/mol	
ΔG_{psum}	-52.0 kJ/mol	
ΔG_{pol}	94.2 kJ/mol +	
ΔG_{bind}(calc)	**-27.5 kJ/mol**	(+/- 2.6 kJ/mol)
ΔG_{bind}(exp)	**-29.6 kJ/mol**	



- Excellent agreement with experiment!

Summary of aspirin binding

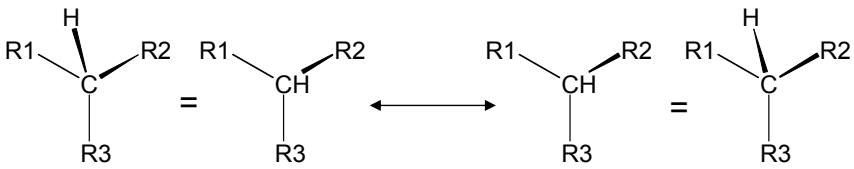


All methods agree within the statistical error estimates and with experiment

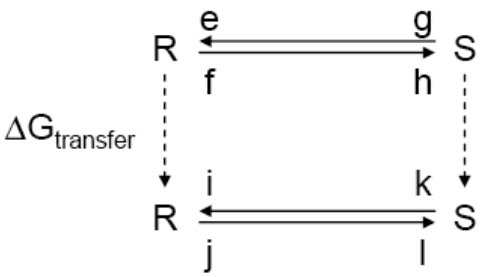
Stereospecific propranolol binding

- R- and S-Propranolol have similar affinity for CYP450 2D6
- 20 fold decrease of affinity of R-Propranolol to F483A mutant

- Free energy calculation to convert R-propranolol into S-propranolol



WT

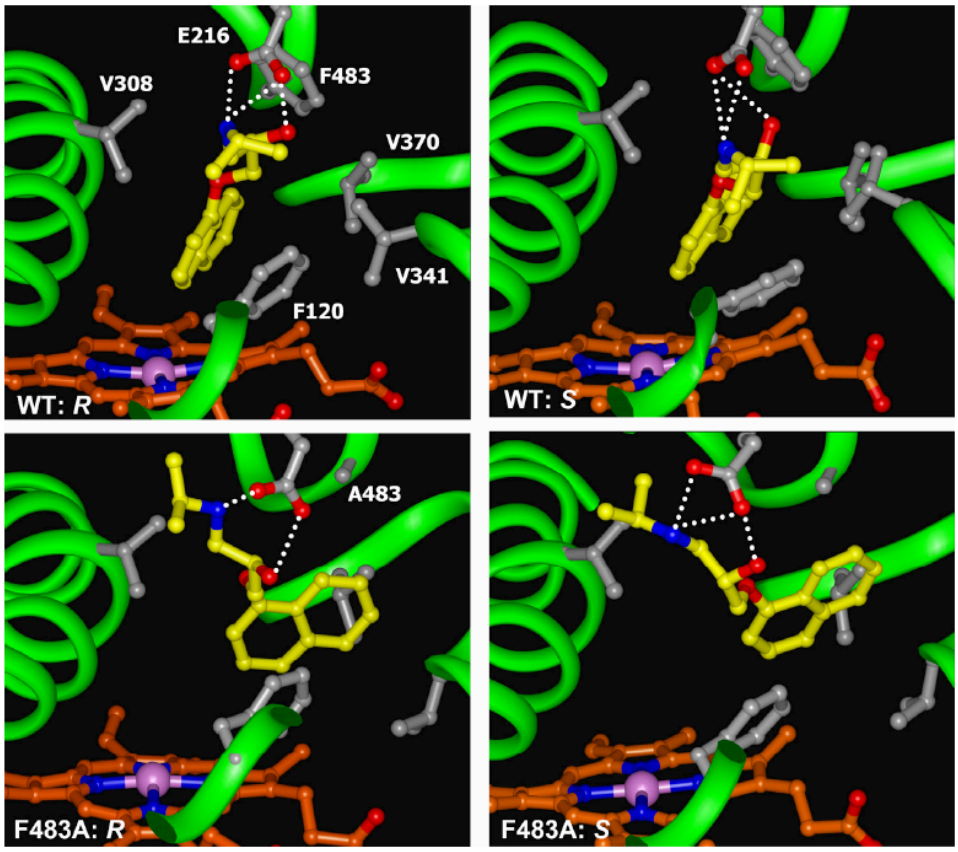


F483A

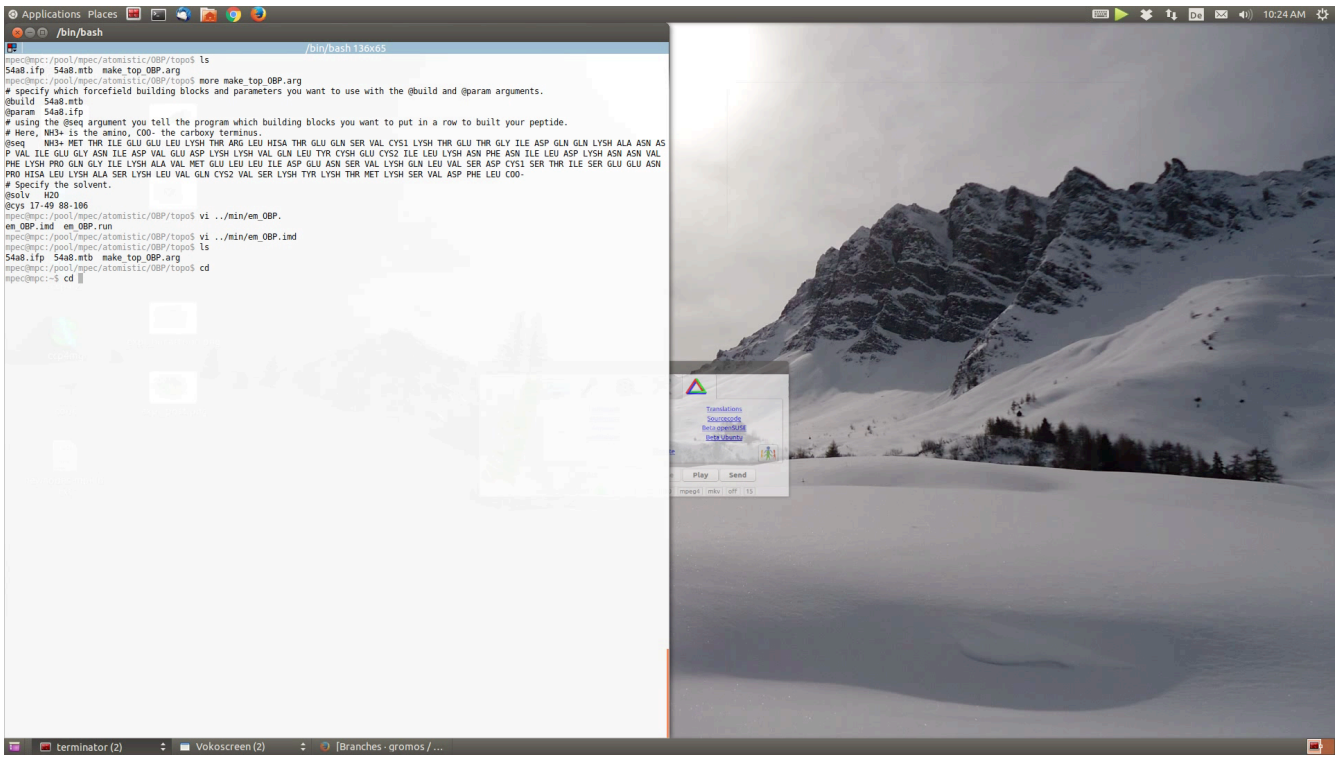
$\Delta\Delta G_{bind} = \Delta G_{bind}(R) - \Delta G_{bind}(S)$
= -7.7 kJ/mol (experimental)

$= \Delta G_{inv}(F483A) - \Delta G_{inv}(WT)$
= -7.5 kJ/mol (simulation)

Molecular picture

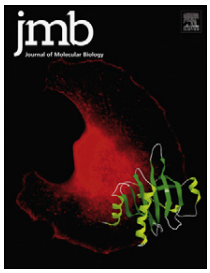


WHAT IT'S REALLY LIKE TO RUN A SIMULATION



Papers I

- J. Mol. Biol. (2017) 429, 930 - 947
- Free energy perturbation calculation of relative binding free energy between broadly neutralizing antibodies and the gp120 glycoprotein of HIV-1



Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the gp120 Glycoprotein of HIV-1

Anthony J. Clark[1], Tatyana Gindin[6], Baoshan Zhang[3], Lingle Wang[4], Robert Abel[4], Colleen S. Murret[1], Fang Xu[1], Amy Bao[3], Nina J. Lu[3], Tongqing Zhou[3], Peter D. Kwong[2,3], Lawrence Shapiro[2,3], Barry Honig[5] and Richard A. Friesner[1,*]

Papers II

- Chem. Res. Toxicol. (2019) 32, 1374 - 1383
- Binding modes and metabolism of caffeine in Cytochrome P450 1A2

ACS AuthorChoice
Article
Cite This: *Chem. Res. Toxicol.* 2019, 32, 1374–1383
pubs.acs.org/crt

Chemical Research in Toxicology

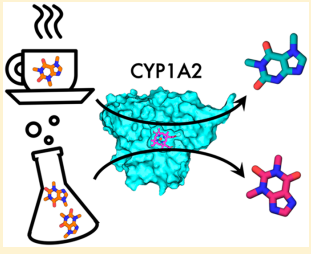
Binding Modes and Metabolism of Caffeine

Zuzana Jandova,[†] Samuel C. Gill,[‡] Nathan M. Lim,[§] David L. Mobley,[‡] and Chris Oostenbrink[*,†]

[†]Institute of Molecular Modeling and Simulation, University of Natural Resources and Life Sciences, Vienna, 1180 Vienna, Austria
[‡]Department of Chemistry and [§]Department of Pharmaceutical Sciences, University of California, Irvine, Irvine, California 92697, United States

Supporting Information

ABSTRACT: A correct estimate of ligand binding modes and a ratio of their occupancies is crucial for calculations of binding free energies. The newly developed method BLUES combines molecular dynamics with nonequilibrium candidate Monte Carlo. Nonequilibrium candidate Monte Carlo generates a plethora of possible binding modes and molecular dynamics enables the system to relax. We used BLUES to investigate binding modes of caffeine in the active site of its metabolizing enzyme Cytochrome P450 1A2 with the aim of elucidating metabolite-formation profiles at different concentrations. Because the activation energies of all sites of metabolism do not show a clear preference for one metabolite over the others, the orientations in the active site must play a key role. In simulations with caffeine located in a spacious pocket above the I-helix, it points N3 and N1 to the heme iron, whereas in simulations where caffeine is in close proximity to the heme N7 and C8 are preferably oriented toward the heme



Conclusions

- Molecular dynamics simulations form a powerful tool to study biomolecules
  - Insight into structure, dynamics and function at an atomic level
  - Complementary to experiment

- Free energy calculations for e.g. drug design / lead optimisation
  - Binding affinities via path-sampling methods
  - Binding affinities via alchemical methods

- Protein flexibility and conformational freedom is important

Institute of molecular modeling and simulation
Modern Bioinformatics WS 2019/2020