Biomolecular modeling IIII

Marcus Elstner and Tomáš Kubař

2016, December 14

Enhanced sampling

How to save time, and time is money

Problem

with normal nanosecond length MD simulations:

It is difficult to overcome barriers to conformational transitions,
and only conformations in the neighborhood of the initial structure
may be sampled,
even if some other (different) conformations are more relevant,
i.e. have lower free energy

Special techniques are required to solve this problem.

Note – do not be afraid of Arrhenius

How often does something happen (in a simulation)?

$k = A \times \exp\left[-E_A/kT\right]$, let us have $A = 1 \times 10^9$ s^{-1}

E_A	k	$1/k$
kcal/mol	$1/$s	μs
1	0.19×10^9	0.005
3	6.7×10^6	0.15
5	0.24×10^6	4.2
7	8.6×10^3	120

So, if the process has to overcome a barrier of 5 kcal/mol,
we will have to simulate for 4 μs to see it happen once on average.

Replica-exchange molecular dynamics

REMD (or parallel tempering) – method to accelerate the sampling of configuration space, which can be applied
even if the configurations of interest are separated by high barriers.

Several (identical) replicas of the molecular system are simulated at the same time, with different temperatures.

The coordinates+velocities of the replicas may be switched (exchanged) between two temperatures.

Probability of replica exchange

The probability of the replica exchange between T_1 and T_2 is determined in (regular) time intervals from the instantaneous potential energies U_1 and U_2 in the corresponding simulations as

$$\mathcal{P}(1 \leftrightarrow 2) = \begin{cases} 1 & \text{if } U_2 < U_1, \\ \exp\left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2}\right) \cdot (U_1 - U_2)\right] & \text{otherwise.} \end{cases}$$

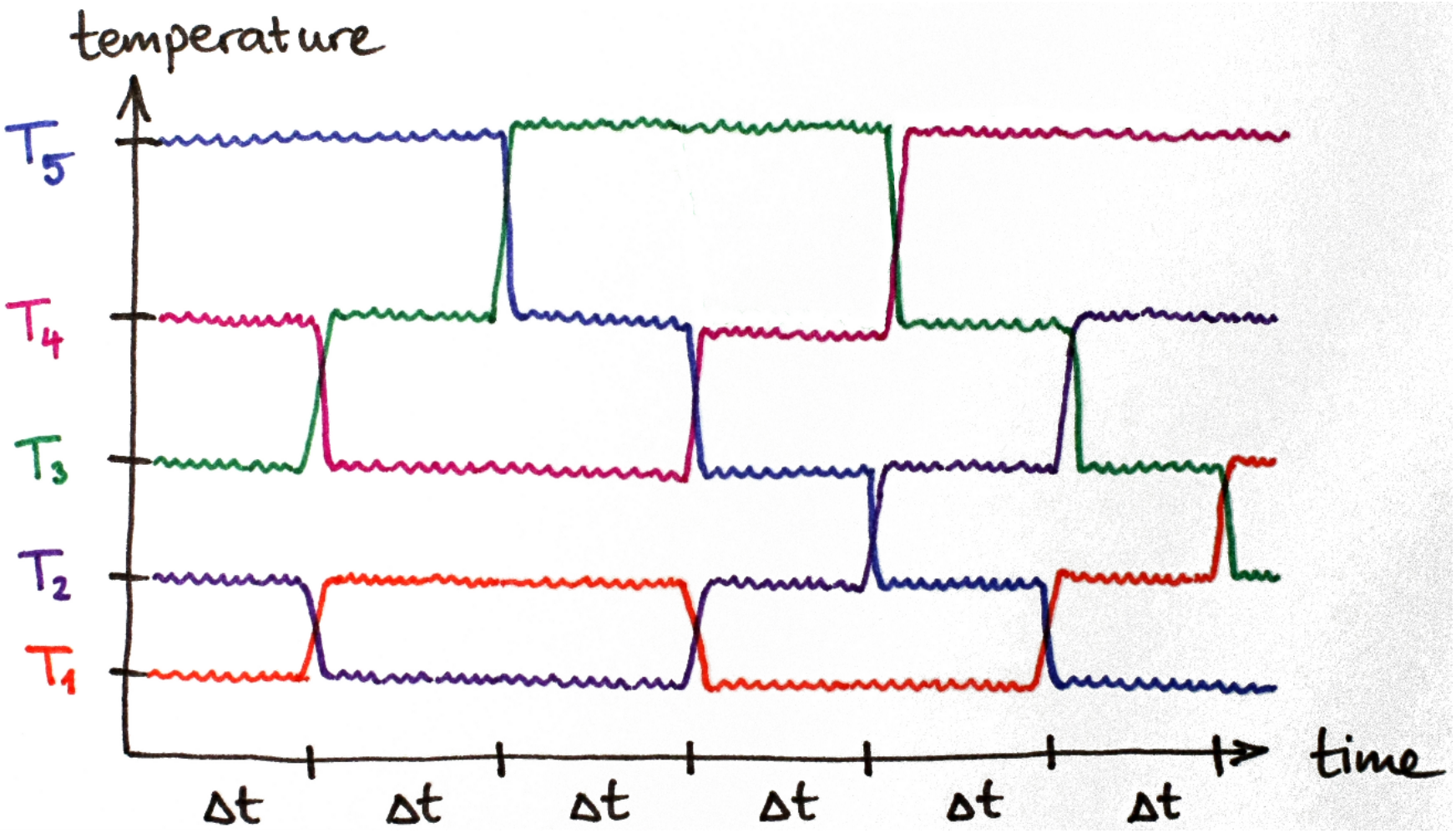
Then, if $\mathcal{P}(1 \leftrightarrow 2)$ is larger than a random number from $(0, 1)$, the replicas in simulations at T_1 and T_2 are exchanged.

Setup of the simulation of replicas

Simulated one replica at the temperature of interest ($T_1 = 300$ K) and several other replicas at higher temp. ($T_1 < T_2 < T_3 < \ldots$).

After (say) 1000 MD steps, attempt exchanges $1 \leftrightarrow 2$, $3 \leftrightarrow 4$ etc., and after next 1000 steps do the same for $2 \leftrightarrow 3$, $4 \leftrightarrow 5$ etc.
so only try to exchange replicas at "neighboring" temperatures

Setup of the simulation of replicas



Advantages of REMD

- due to the simulations at high temperatures
- faster sampling and more frequent crossing of energy barriers
- correct sampling at all temperatures obtained,
  above all at the (lowest) temperature of interest
- increased computational cost (multiple simulations)
  pays off with largely accelerated sampling
- simulations running at different temperatures are independent
  except at attempted exchanges $\rightarrow$ easy parallelization
- first application – protein folding (Sugita & Okamoto, Chem. Phys. Lett. 1999)

Choice of temperatures and disadvantages of REMD

For protein/water systems with all bond lengths constrained:

- $N_{\mathrm{dof}} \approx 2N$ (N – number of atoms)
- average probability is related to $T_2 - T_1 = \varepsilon T_1$ as

$$\overline{\mathcal{P}(1 \leftrightarrow 2)} \approx \exp\left[-2\varepsilon^2 N\right]$$

- set of temperatures may be designed to suit the problem

Disadvantages of parallel tempering REMD

- large number of atoms: low exchange probability
  $\rightarrow$ low efficiency
- high temperature – sensitive biostructures may not survive
  (membranes etc.)

how to apply the replica-exchange idea and avoid these issues?

Hamiltonian replica exchange – HREX

also called 'Replica exchange with solute tempering' (REST)

$$P = \exp\left[-\frac{U}{kT}\right] = \exp\left[-\beta U\right]$$

- note: $\frac{1}{2}U$ would be the same as $2T$
- force field energy U is combined from many individual terms
  - let us scale selected terms (not all of them!)
  - is not possible for temperature scaling (a single T)
  - 'heating' of a (small) part of the system
  - typically, a group of atoms – a ligand, or several AAs...

Hamiltonian replica exchange – HREX

Simulations 1 and 2 are performed
with different force fields U_1 and U_2

How to calculate the probability of exchange?
(q_1 and q_2 – coordinates of atoms in simulations 1 and 2)

$$\Delta = \frac{U_1(q_2) - U_1(q_1) - U_2(q_1) + U_2(q_2)}{kT}$$

$$\mathcal{P}(1 \leftrightarrow 2) = \begin{cases} 1 & \text{if } \Delta \leq 0, \\ \exp\left[-\Delta\right] & \text{otherwise.} \end{cases}$$

Then, if $\mathcal{P}(1 \leftrightarrow 2)$ is larger than a random number from $(0, 1)$, the replicas in simulations with U_1 and U_2 are exchanged.

HREX – a good variant

- divide the system into two parts:
- hot – small, will be subject to extended sampling
- cold – all of the rest

Generate replicas with different $\lambda_m < 1$, modify parameters in hot:

- scale the charges by $\sqrt{\lambda_m}$
- scale the LJ depths ε by λ_m
- additional scaling of dihedral angles

Then, the 'effective' temperatures are

- inside hot: $T/\lambda_m > T$
- interactions between hot and cold: $T/\sqrt{\lambda_m}$
- inside cold: T is retained

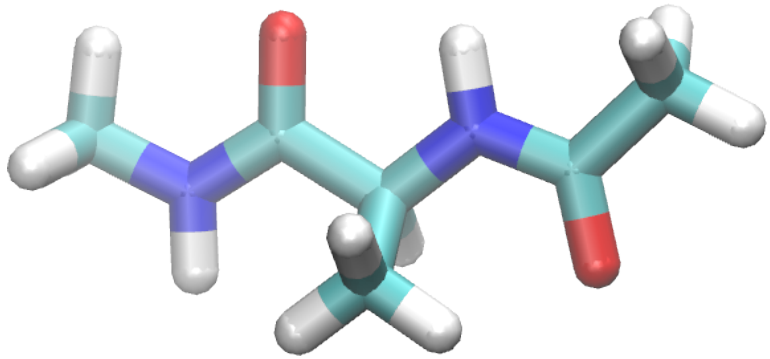
HREX

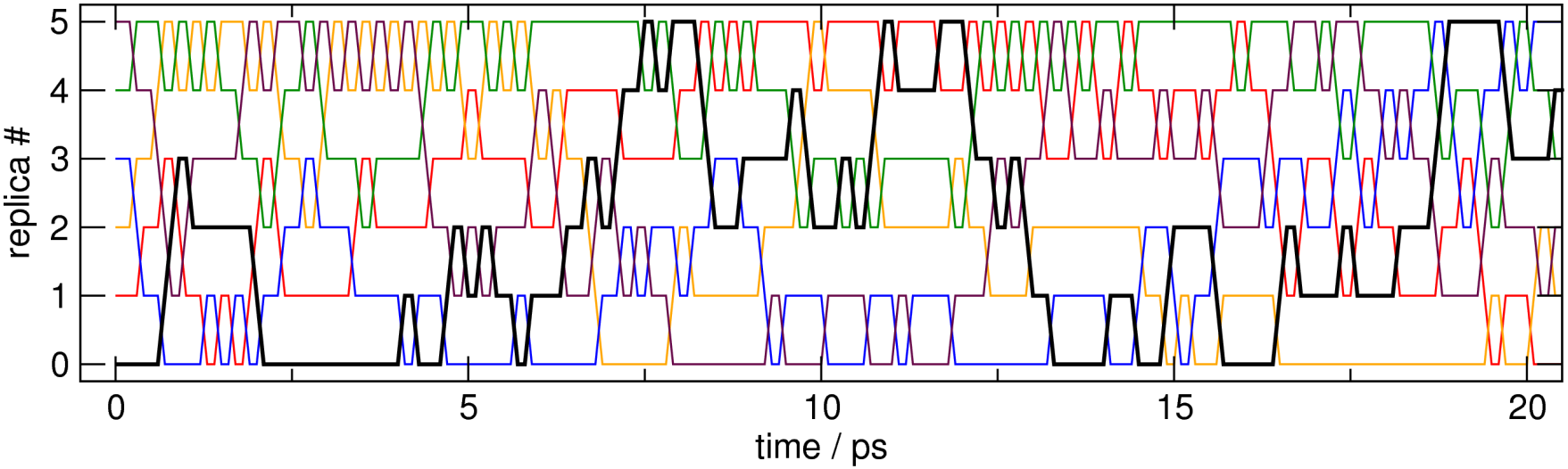
Meaning of temperature

- kinetic energy $\leftarrow$ velocities
  - does not change, is the same in hot and cold (300 K)
  - simulation settings need not be adjusted (time step!)
  - unlike in parallel tempering
- factor affecting the population of states
  - we play with this

HREX – example – solute tempering

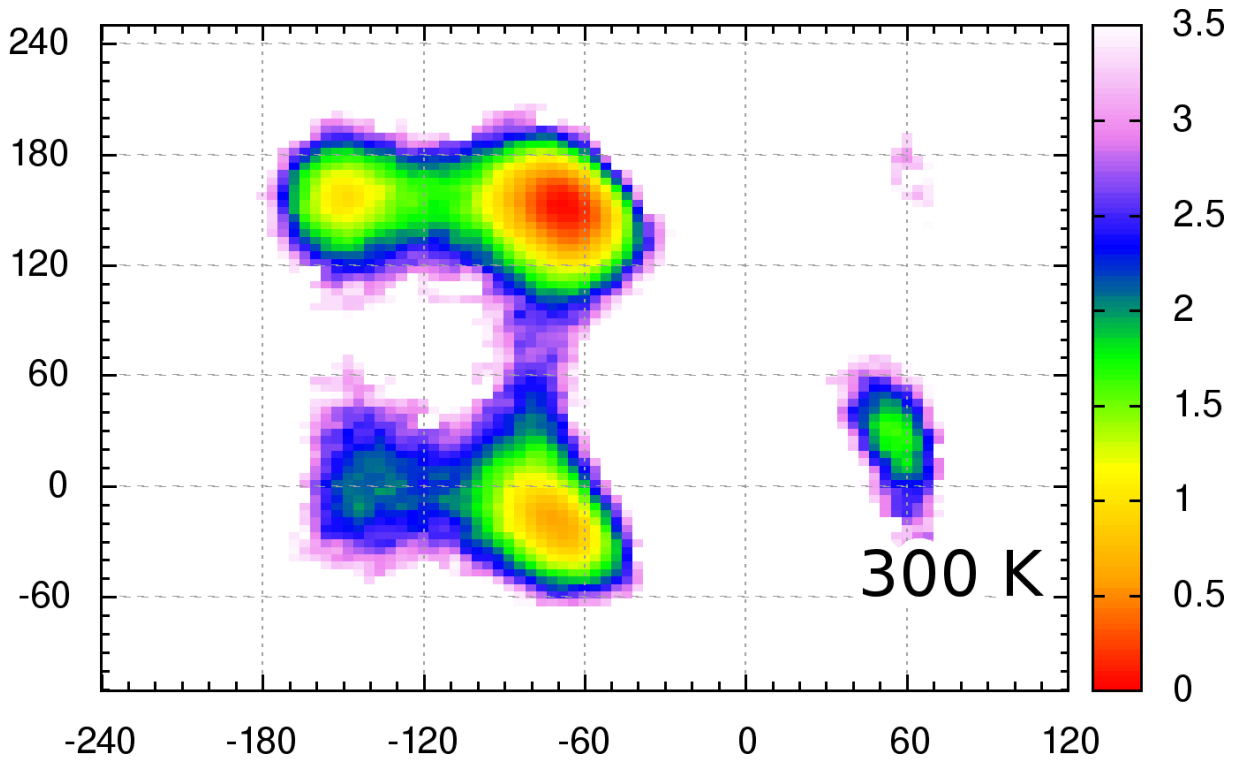
- alanine dipeptide – 22 atoms,
  1 pair of $\varphi - \psi$
- 5 replicas, $\lambda = 1 \ldots 0.18$
  i.e. $T_m = 300 \ldots 1700$ K
- exchange every 0.1 ps,
  observed $\overline{\mathcal{P}}$ =0.25–0.50





HREX – example

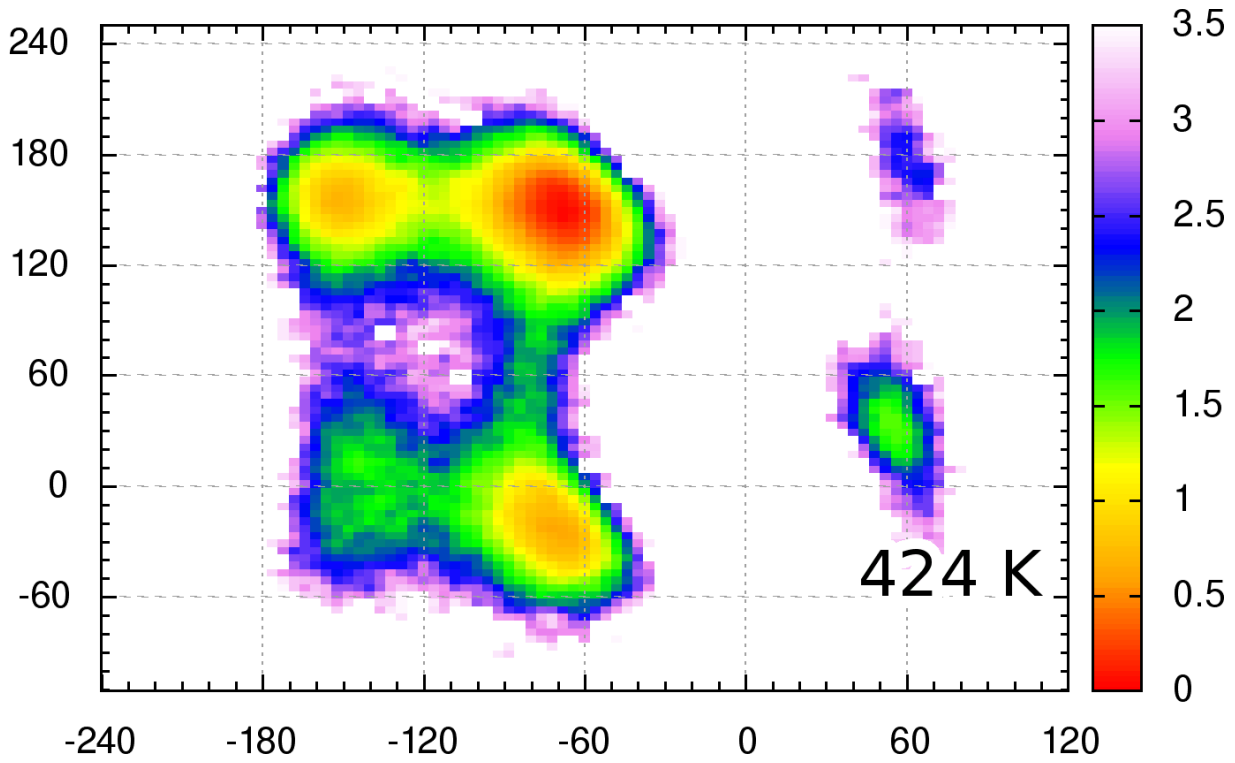
Solute tempering – dialanine – replica #0



$\varphi - \psi$ in degrees, ΔF in kcal/mol

HREX – example

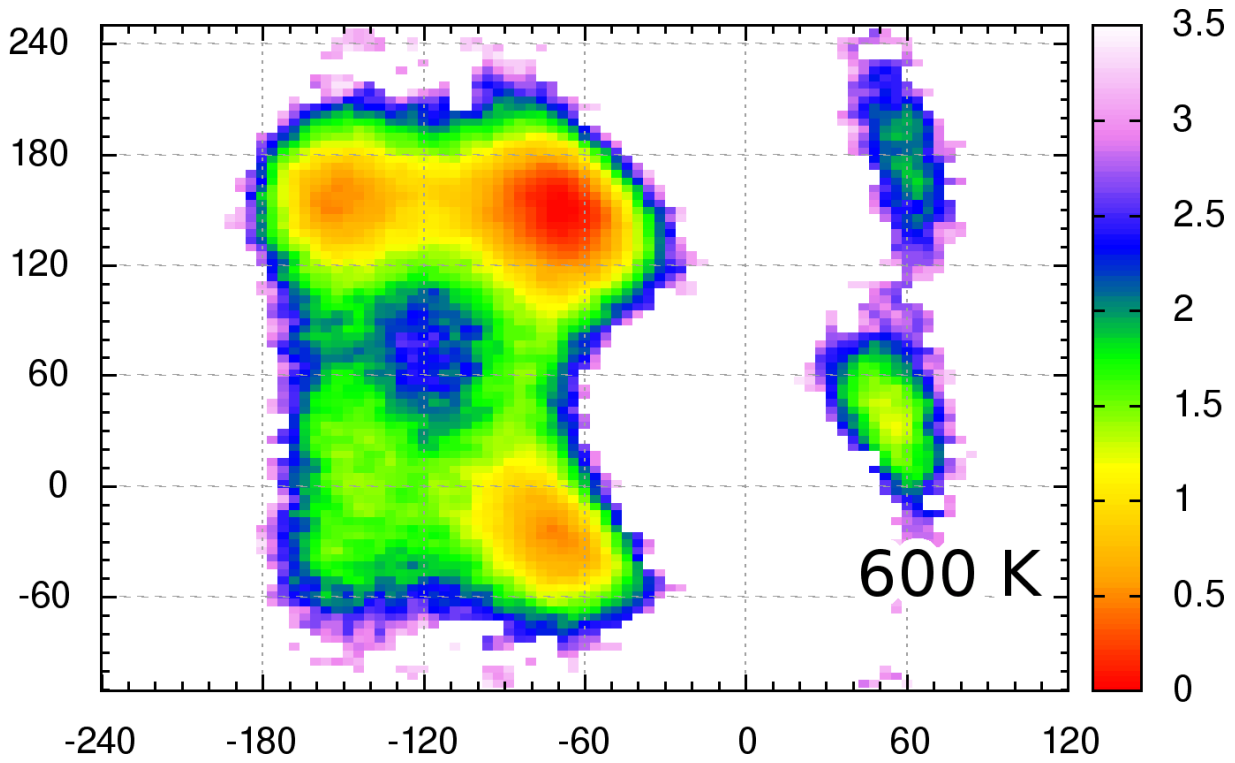
Solute tempering – dialanine – replica #1



$\varphi - \psi$ in degrees, ΔF in kcal/mol

HREX – example

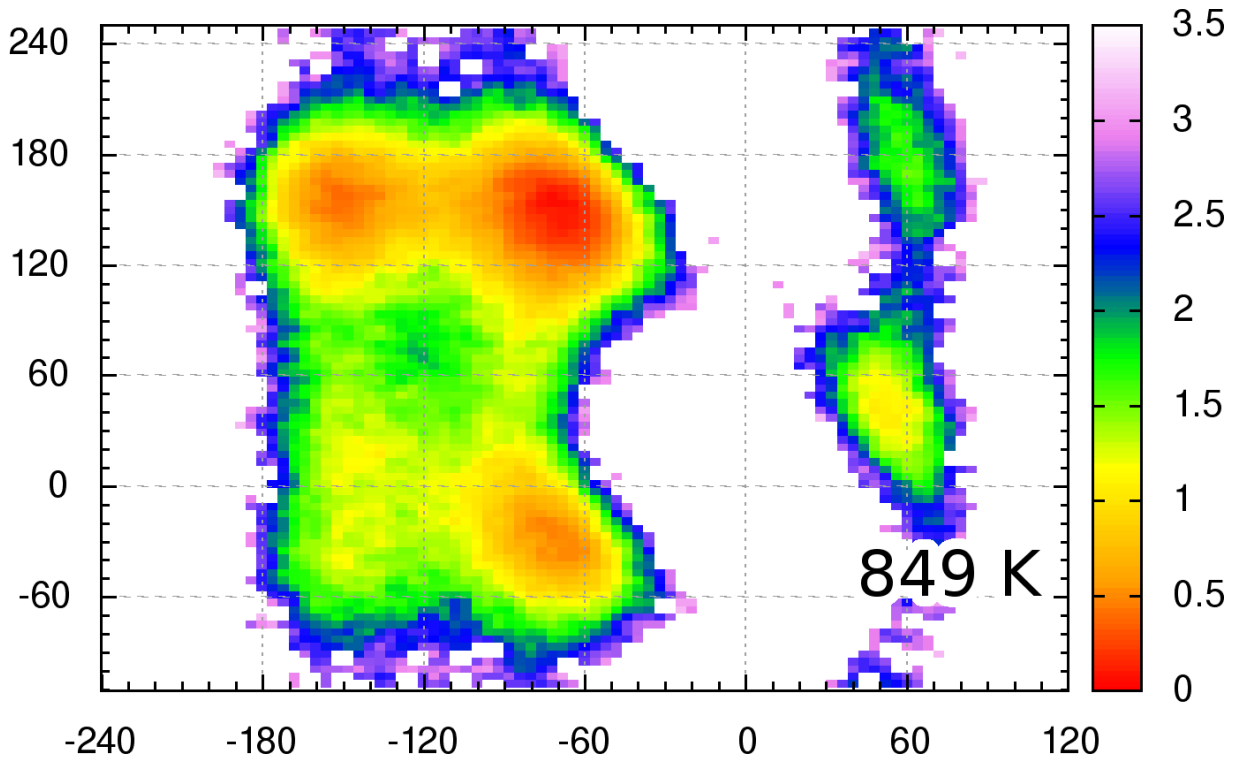
Solute tempering – dialanine – replica #2



$\varphi - \psi$ in degrees, ΔF in kcal/mol

HREX – example

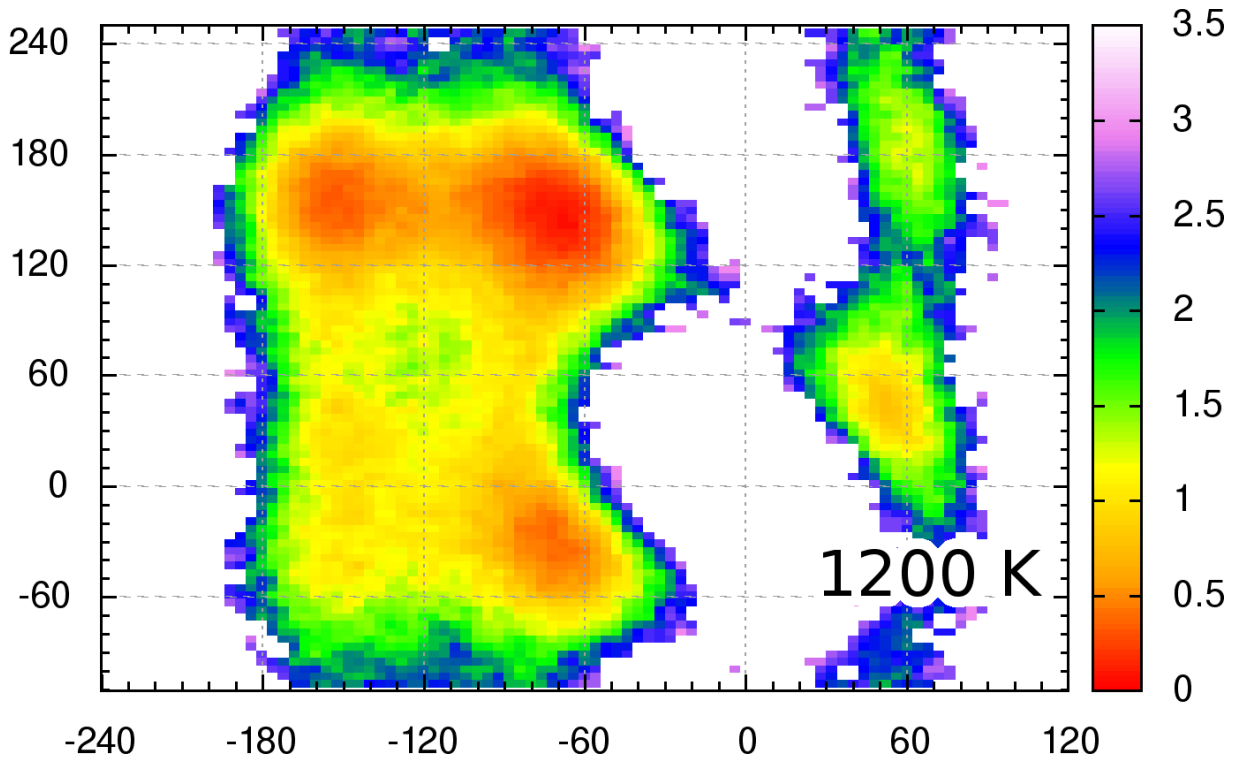
Solute tempering – dialanine – replica #3



$\varphi - \psi$ in degrees, ΔF in kcal/mol

HREX – example

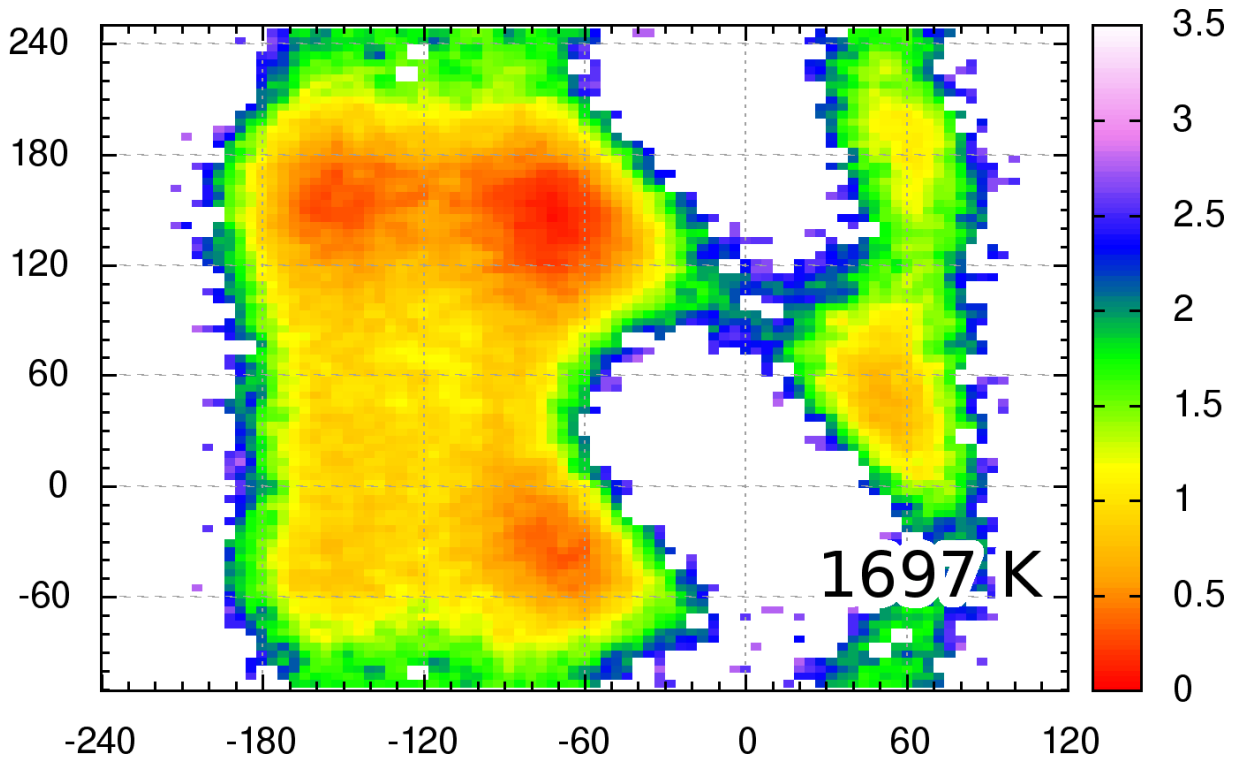
Solute tempering – dialanine – replica #4



$\varphi - \psi$ in degrees, ΔF in kcal/mol

HREX – example

Solute tempering – dialanine – replica #5



$\varphi - \psi$ in degrees, ΔF in kcal/mol

Methods using biasing potentials

Other approaches use a different idea:

It is easy to introduce an additional contribution
to the potential energy of the molecule

Example – the extra potential may force the molecule
over an energy barrier, to explore other conformations

It is 'unrealistic' – we do not simulate a real molecule
but this bias may be removed by a right post-processing

Note: use of NMR-based distance restrains in MD simulations
$\rightarrow$ 'NMR-refined' structure of the molecule (e.g. PDB ID 1AC9)

Metadynamics

– aimed at reconstructing the multidimensional free energy of complex systems (Laio & Parrinello 2002)

– based on an artificial dynamics (metadynamics) performed in the space of a few collective variables S (e.g. normal modes)

– at regular time intervals during the simulation,
an additional biasing energy function is added to the force field
    – a Gaussian that is centered on the current structure

using quotations by Alessandro Laio

Metadynamics – how it works

a new Gaussian is added at every time interval t_G,
and the biasing potential at time t is given by

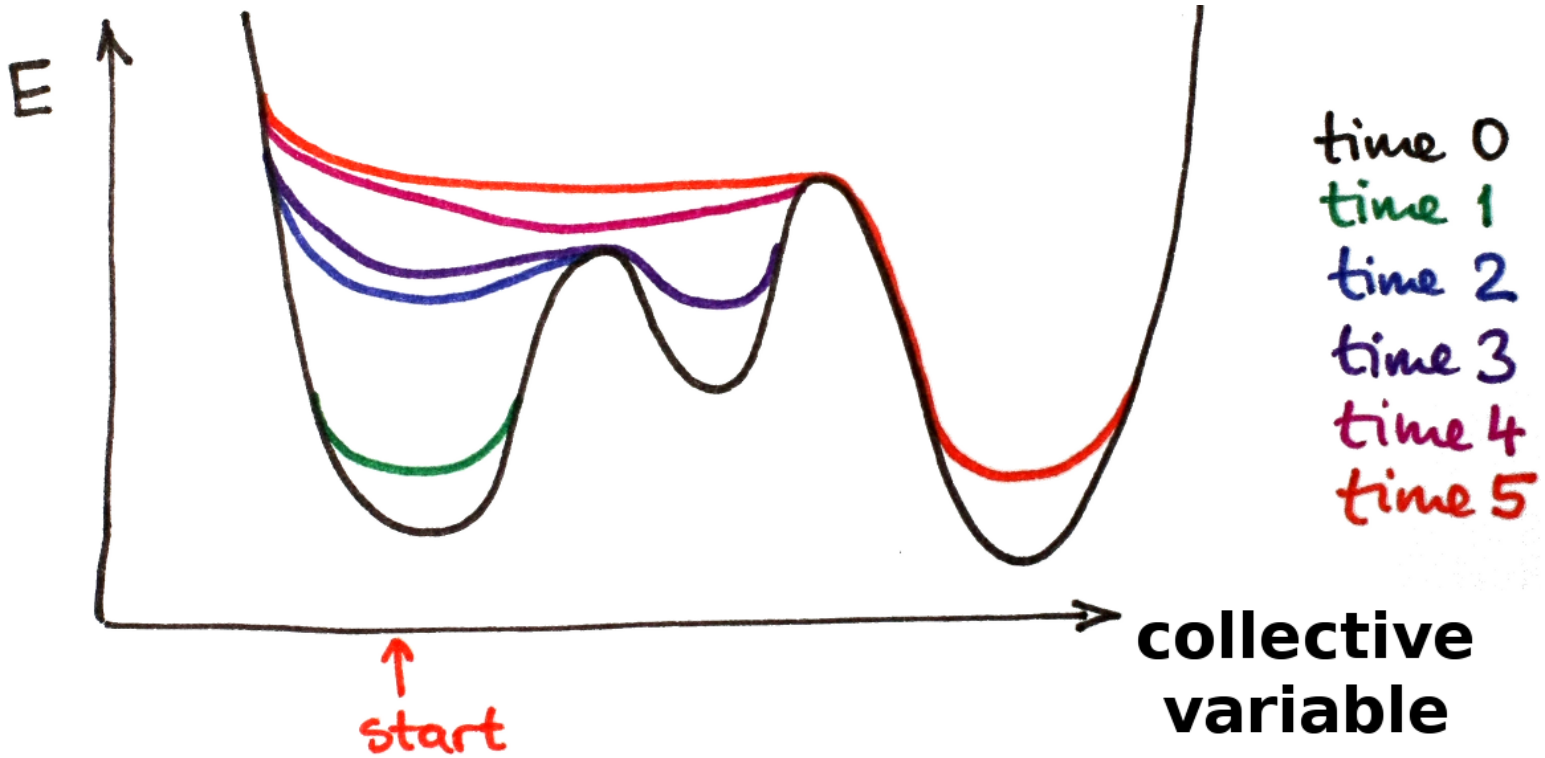
$$V_G(S(x), t) = \sum_{t'=t_G, 2t_G, 3t_G, \ldots} w \cdot \exp\left[-\frac{(S(x) - s_{t'})^2}{2 \cdot \delta s^2}\right]$$

w and δs – height and width of the Gaussians
$s_t = S(x(t))$ – value of the collective variable at time t

In the course of the simulation, this potential is filling the minima
on the free energy surface that the system is traveling through

So, the MD has a memory via the biasing potential

Metadynamics – what it looks like



https://www.youtube.com/watch?v=IzEBpQ0c8TA
https://www.youtube.com/watch?v=iu2GtQAyoj0

Properties of metadynamics

Metadynamics – to explore new reaction pathways,
accelerate rare events,
and also to estimate the free energies efficiently.

- The system escapes a local free energy minimum through the lowest free-energy saddle point.
- The dynamics continues, and all of the free-energy profile is filled with the biasing Gaussians.
- At the end, the sum of the Gaussians provides the negative of the free energy.

Properties of metadynamics

Crucial task – prior to simulation: identify the collective variables of interest

that are difficult to sample because of high barriers

These variables $S(x)$ are functions of the coordinates of the system;

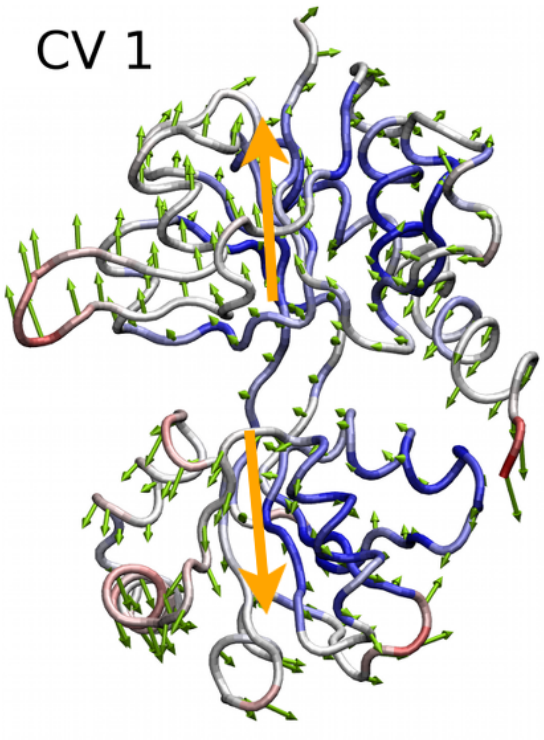
practical applications – up to 3 such variables,

and the choice depend on the process being studied.

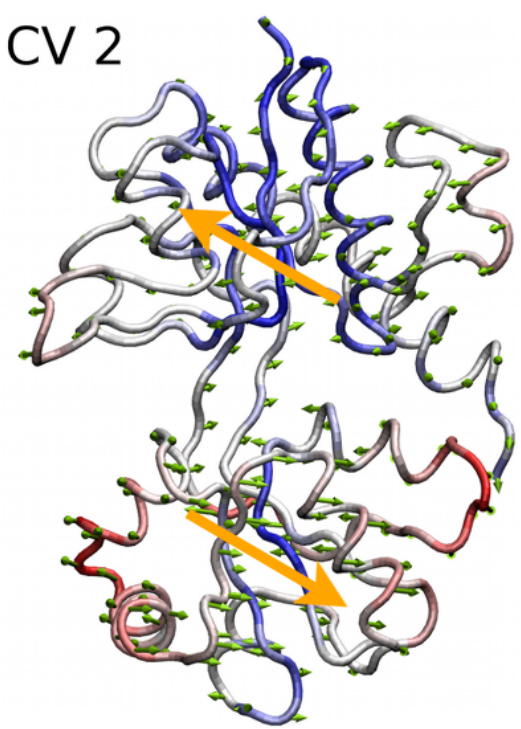
Typical choices – principal modes of motion obtained with PCA

Still, the choice of S may be difficult

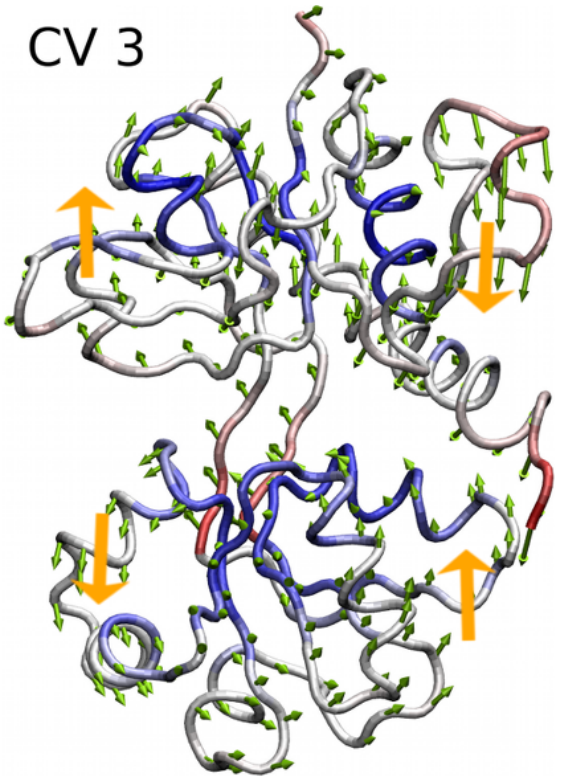
Example – opening of a protein binding pocket



clamshell



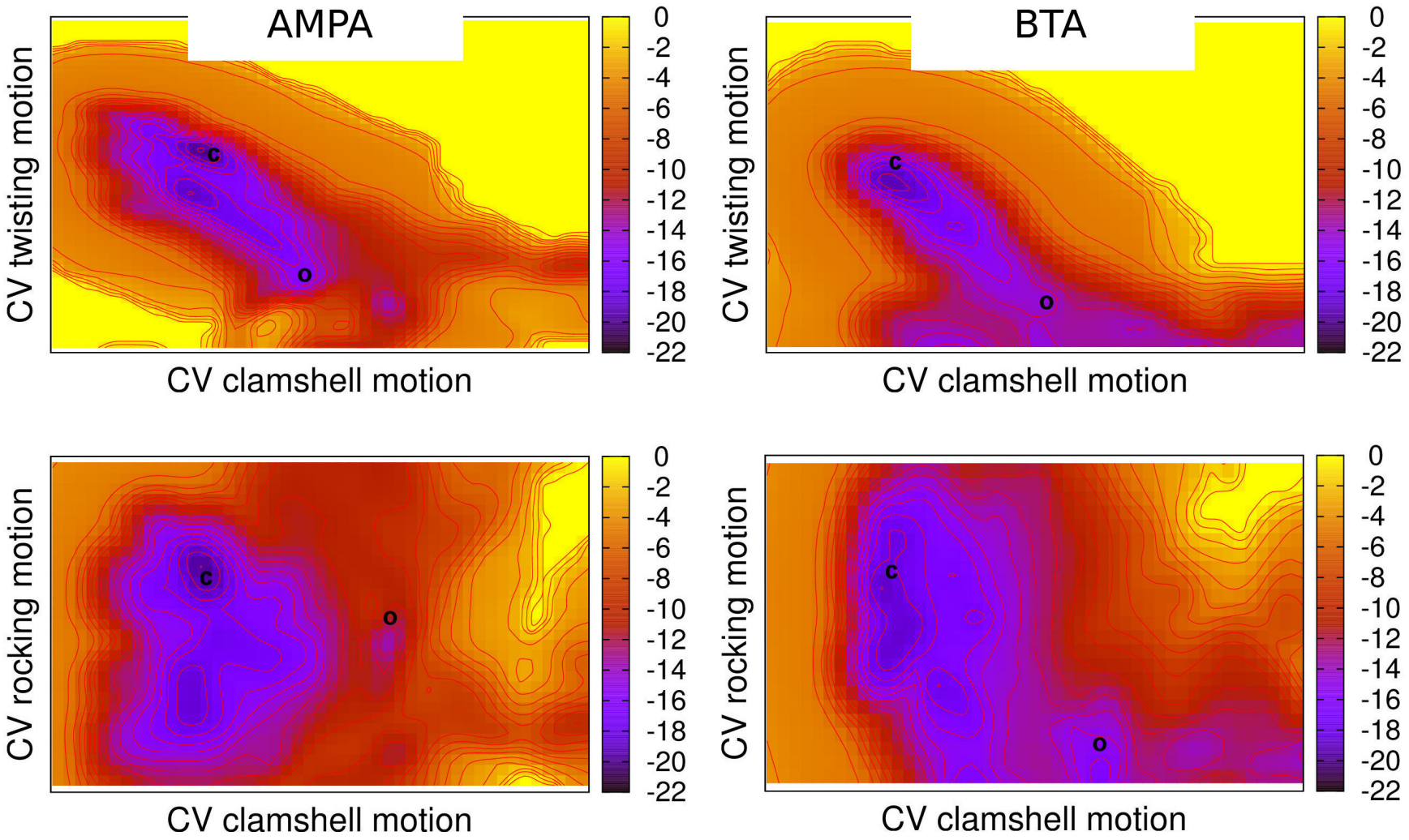
twisting



rocking

courtesy Tino Wolter

Example – opening of a protein binding pocket



courtesy Tino Wolter

Enhanced sampling methods – comparison

Biasing potential methods – metadynamics, umbrella sampling

- required: a priori choice of reaction coordinate(s) to be biased
- problem – success depends on that choice, possibly non-trivial

REMD with parallel tempering

- + no such required, can be used rather blindly
- – all of the system heated $\rightarrow$ may destroy something
- – no knowledge of the system may be embedded
- – poor efficiency for big systems: $\overline{\mathcal{P}(1 \leftrightarrow 2)} \approx \exp\left[-2\varepsilon^2 N\right]$
  $\rightarrow$ critical problem

Enhanced sampling methods – comparison

Hamiltonian replica exchange (HREX)

- in intermediate position
  between metadynamics/US and REMD-PT
- simpler to use than metadynamics/US
  – results depend not so strongly on the choices to be made
- efficiency does not depend on the overall system size

Free energy simulations

Motivation

physical quantities of prime interest in chemistry?

free energies – Helmholtz F or Gibbs G

– determine whether processes (reactions) run spontaneously or not

– holy grail of computational chemistry,
both for their importance
and because they are difficult to calculate

Convergence issue

(all of the formulas come from statistical thermodynamics)

– especially desperate for free energies:

$$F = k_B T \cdot \ln \left\langle \exp \left[\frac{E}{k_B T} \right] \right\rangle$$

serious issue – the large energy values enter an exponential,
and so the high-energy regions may contribute significantly!
$\rightarrow$ if these are undersampled, then free energies are wrong

– calculation of free energies impossible, special methods needed!

Tackling the issue

two fundamental approaches:
free energy perturbation and thermodynamic integration

several computational tricks for particular types of reactions:
alchemical simulations or umbrella sampling

important: not necessary to find the absolute value of free energy;
for a chemical reaction, we only need
the free energy difference (ΔF, ΔG) of reactant and product

"reaction" – not necessarily chemical bonds created or broken
– ligand binding a protein
– passage of a molecule/ion through membrane
– protein folding
...

Thermodynamic integration

Free energy as function of reaction coordinate λ: $F = F(\lambda)$,
with $\lambda = 0$ for reactant, $\lambda = 1$ for product

$$\Delta F = F(1) - F(0) = \int_0^1 \frac{\partial F(\lambda)}{\partial \lambda} \mathrm{d}\lambda$$

Free energy is a state function

$\rightarrow$ the result is independent of the chosen path $0 \rightarrow 1$

$\rightarrow$ reaction coordinate may be even an unphysical process

– change of chemical identity of atoms (alchemical simulations)

$$\Delta F = \int_0^1 \left\langle \frac{\partial E_\lambda}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda$$

Essence of TI – the average of derivative of total MM energy E
is evaluated in the simulation directly

How to do it practically

- perform a MD simulation for each chosen value of λ:
  – usually, equidistant values in the interval (0,1) are taken:
  0, 0.05, . . . , 0.95 and 1
- each of these simulations produces a value of $\left\langle \frac{\partial E}{\partial \lambda} \right\rangle_\lambda$
  – we obtain the derivative of F in discrete points for $\lambda \in (0, 1)$
- this function is integrated numerically,
  – the result is the desired free energy difference ΔF
- how many "windows", or λ values shall we choose?
  – we would like to have as few windows as possible,
  without compromising numerical precision
  – inaccuracy may be due to the numerical integration

Computational alchemy

TI looks complicated, but it is rather straightforward,
– common simulation programs run TI conveniently

Computational alchemy
– change of chemical identities of atoms or functional groups

Using a parameter λ, the force-field parameters of state 0 are changed to those of state 1 gradually:

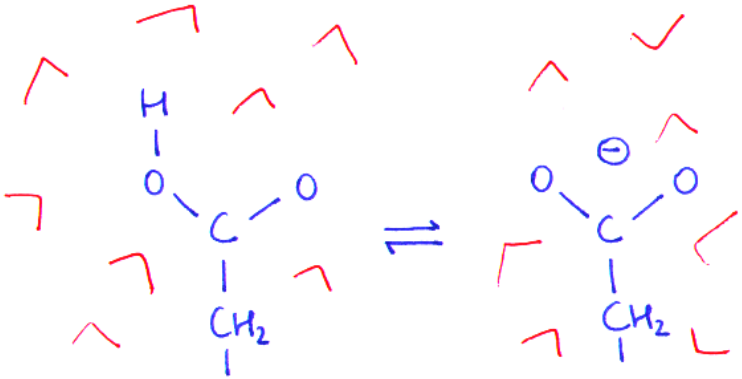
$$E_\lambda = (1 - \lambda) \cdot E_0 + \lambda \cdot E_1$$

Advantages of TI

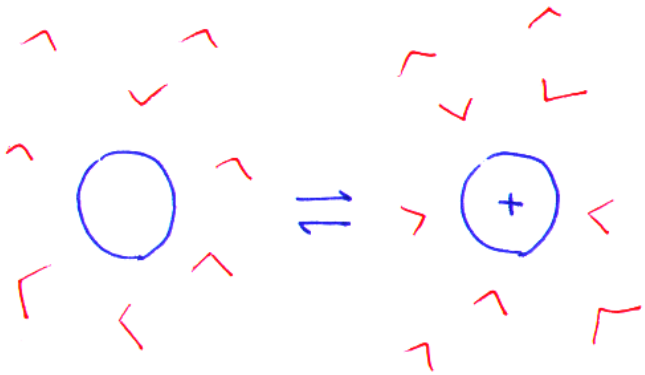
- evaluate the derivative of energies,
  no need to sample for the (large) total energies first
- it is not important what happens outside of the region
  where the reaction takes place (no contrib. to $E_1 - E_0$)
- the ensemble of structures that have to be sampled thoroughly
  is much smaller, and shorter simulation length is required

Examples

deprotonation of amino acid



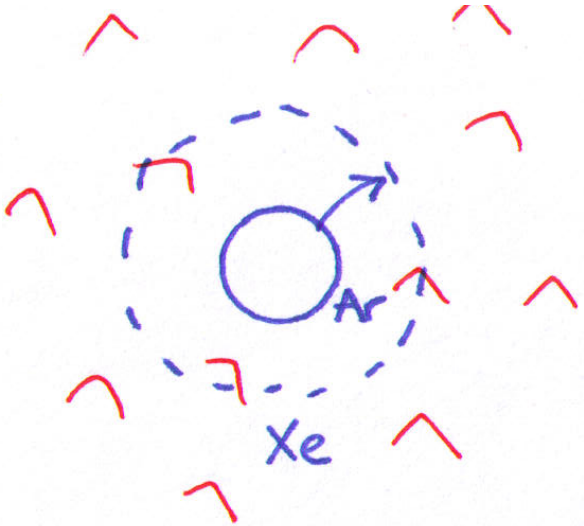
ionization of a molecule



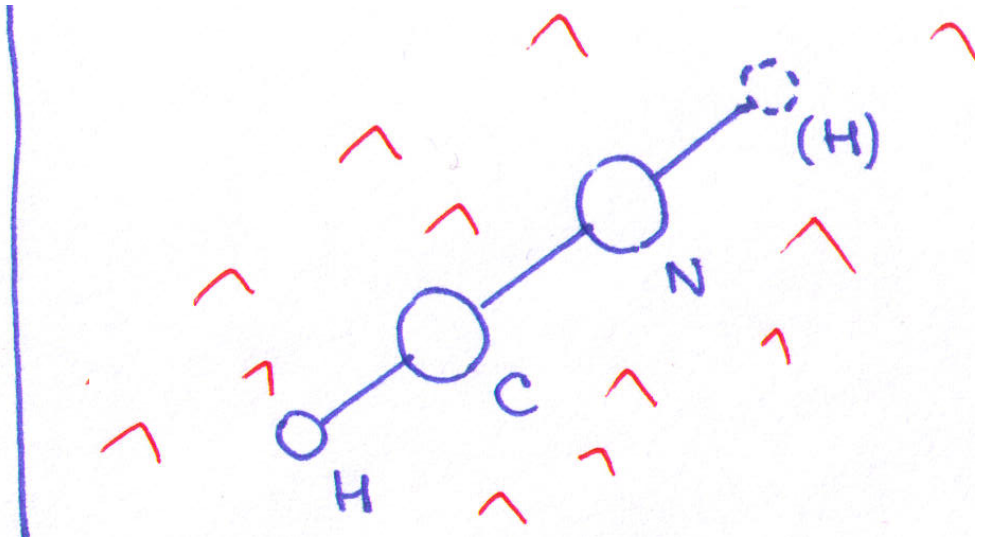
Examples

hydration free energy difference of:

argon and xenon



HCN and CNH



Example

The hydration free energy difference of argon and xenon

Let us interpolate between the parameters for the two elements:

$$\begin{aligned} \varepsilon_\lambda &= (1-\lambda)\cdot\varepsilon_0 + \lambda\cdot\varepsilon_1 \\ \sigma_\lambda &= (1-\lambda)\cdot\sigma_0 + \lambda\cdot\sigma_1 \end{aligned}$$

In the simulation, we start from $\lambda = 0$, i.e. an argon atom, and change it in subsequent steps to 1.

For each step (window), we perform an MD simulation with the corresponding values of the vdW parameters, and evaluate the free energy derivative

Example

Free energy of hydration of rare gas (neon)

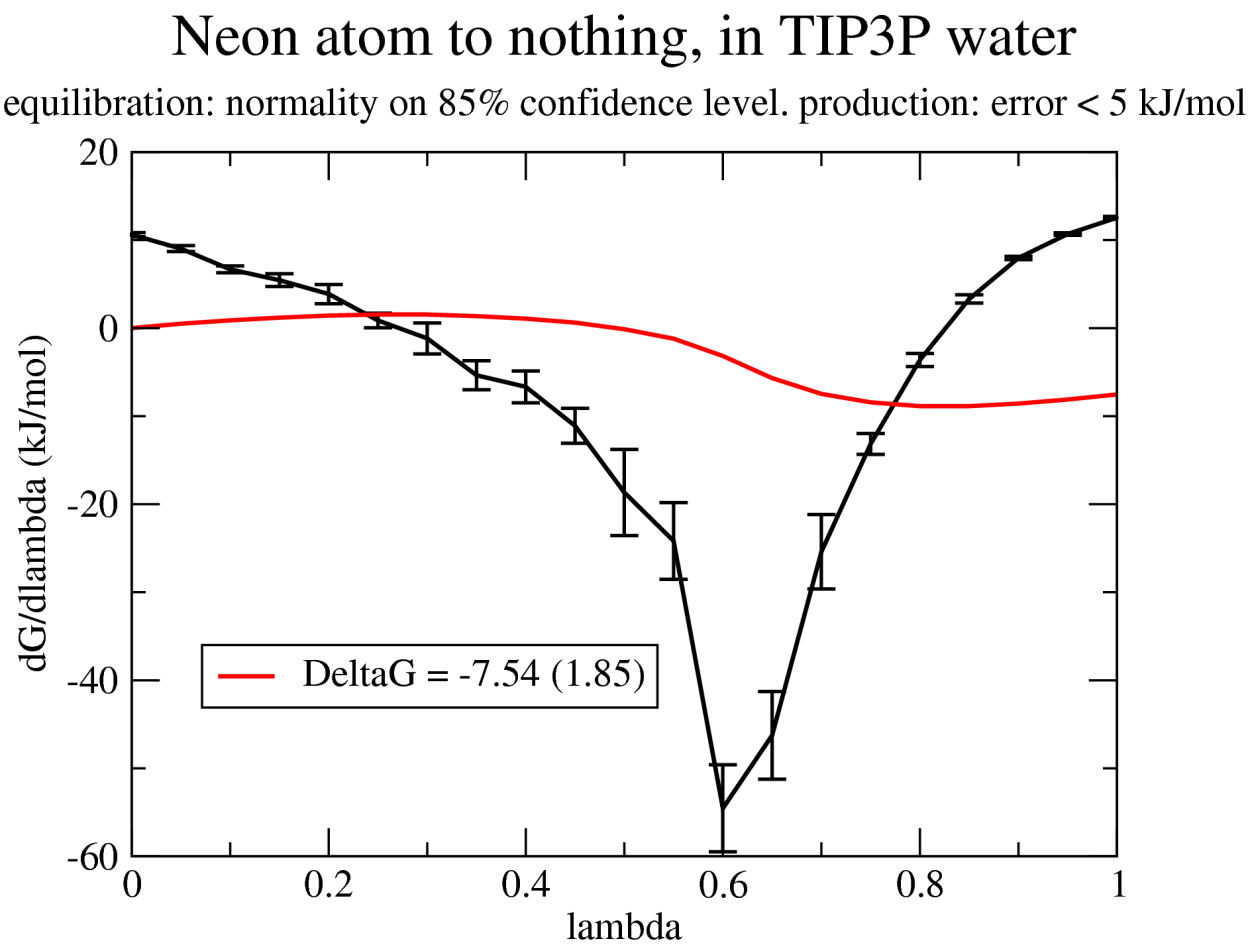
van der Waals parameters of the neon are gradually switched off by means of λ, so that the atom is effectively disappearing

The derivative of total energy with respect to λ is evaluated for 21 values of λ ranging from 0 to 1.

Then, TI gives the Gibbs energy difference of two states:

- a neon atom in water
- no neon atom in water $\equiv$
  $\equiv$ a neon atom outside of the solution, in vacuo

Example


Neon atom to nothing, in TIP3P water
equilibration: normality on 85% confidence level. production: error < 5 kJ/mol
dG/dlambda (kJ/mol)
lambda
DeltaG = -7.54 (1.85)
20
0
-20
-40
-60
0
0.2
0.4
0.6
0.8
1

Differences of differences

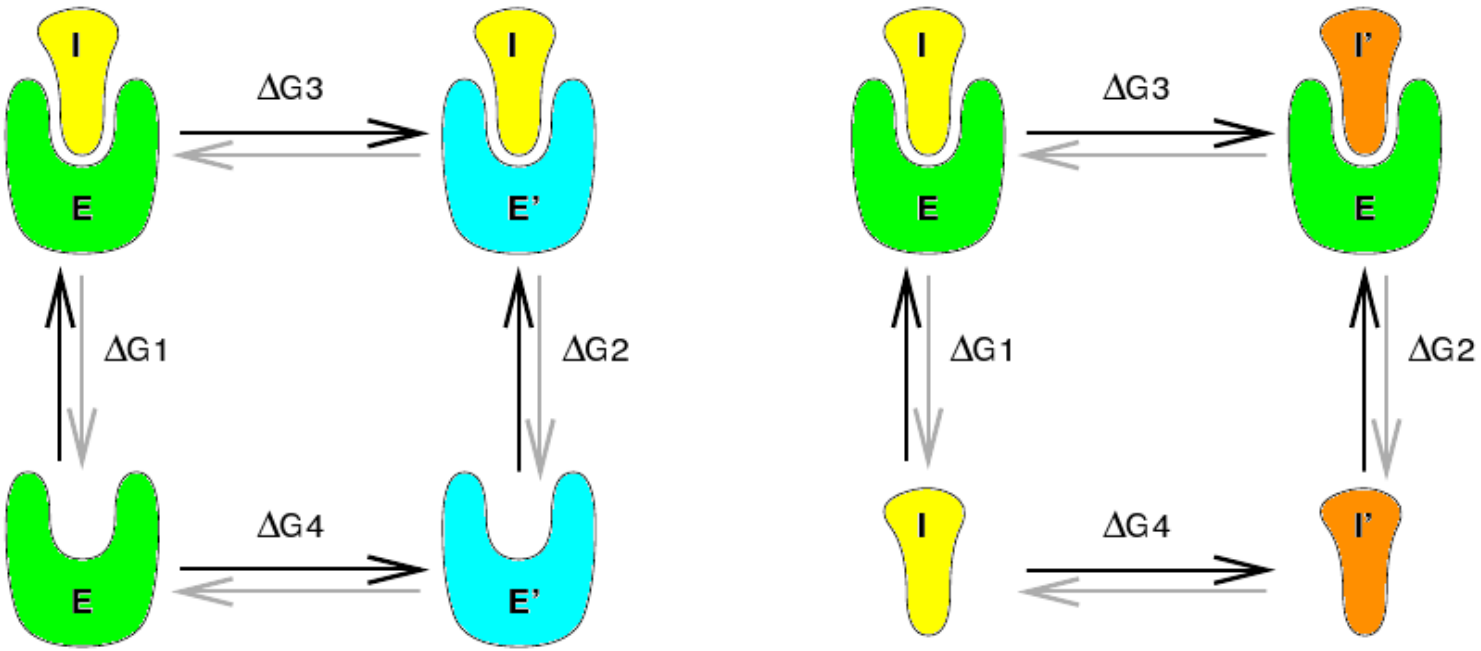
Often – we are interested not in the absolute free energies
and not even in the reaction free energies,
rather, in the difference (Δ) of reaction free energies (ΔF)
of two similar reactions:

$$\Delta\Delta F \text{ or } \Delta\Delta G$$

Reaction free energy difference

Example left: binding of an inhibitor molecule I to an enzyme E, difference of binding free energies to similar enzymes E and E′:

$$\begin{aligned} E + I &\rightleftharpoons EI & \Delta G_1 \\ E' + I &\rightleftharpoons E'I & \Delta G_2 \end{aligned}$$



Reaction free energy difference

The simulation of the ligand binding process itself – very difficult
(possibly large structural changes in the enzyme upon binding)

Solution of the problem – do not simulate the reaction of binding,
rather, the alchemical transmutation of enzyme E to E′.

E to E′ are very similar so this may be easy to do.
(example: mutation of a single AA, e.g. leucine to valine)

Then, the structure of complexes EI and E′I may be similar as well,
and the simulation may provide converged free energy.

Reaction free energy difference

Free energy is a state function $\rightarrow$ the sum of free energies around a thermodynamic cycle vanishes:
(e.g. clockwise in figure left):

$$\Delta G_1 + \Delta G_3 - \Delta G_2 - \Delta G_4 = 0$$

The difference of binding free energies equals
the difference of free energies calculated in alchemical simulations:

$$\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4$$

Geometric reaction coordinate

Sometimes, we need to know how the free energy changes along a geometric reaction coordinate q

The free energy is then a function of q

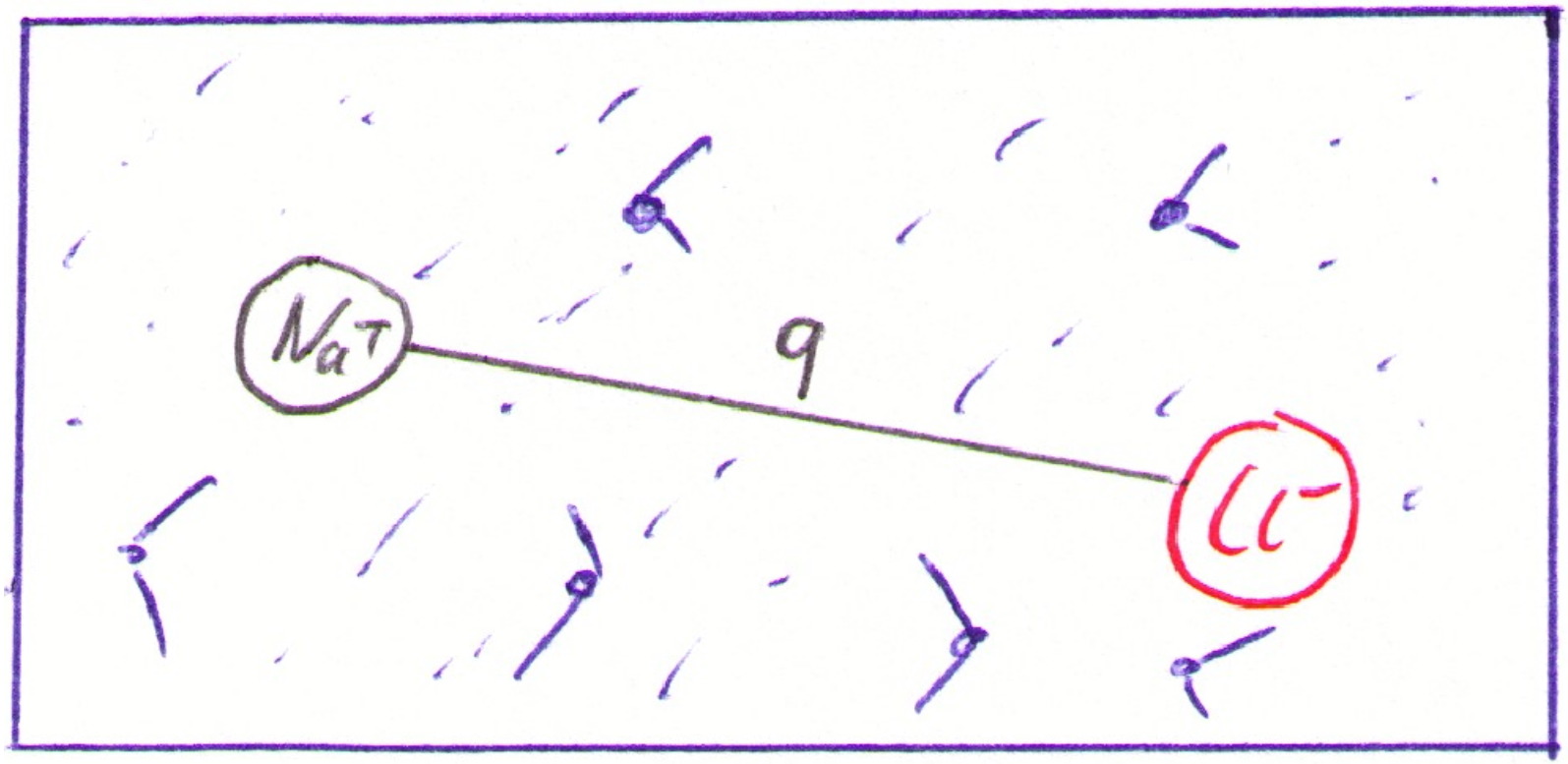
Such a function $F(q)$ is called the potential of mean force.

Examples:

- distance between two particles in a dissociating complex
- the position of a proton for a reaction of proton transfer
- the dihedral angle when dealing with conformational changes

Example

free energy of formation of an ion pair in solution:



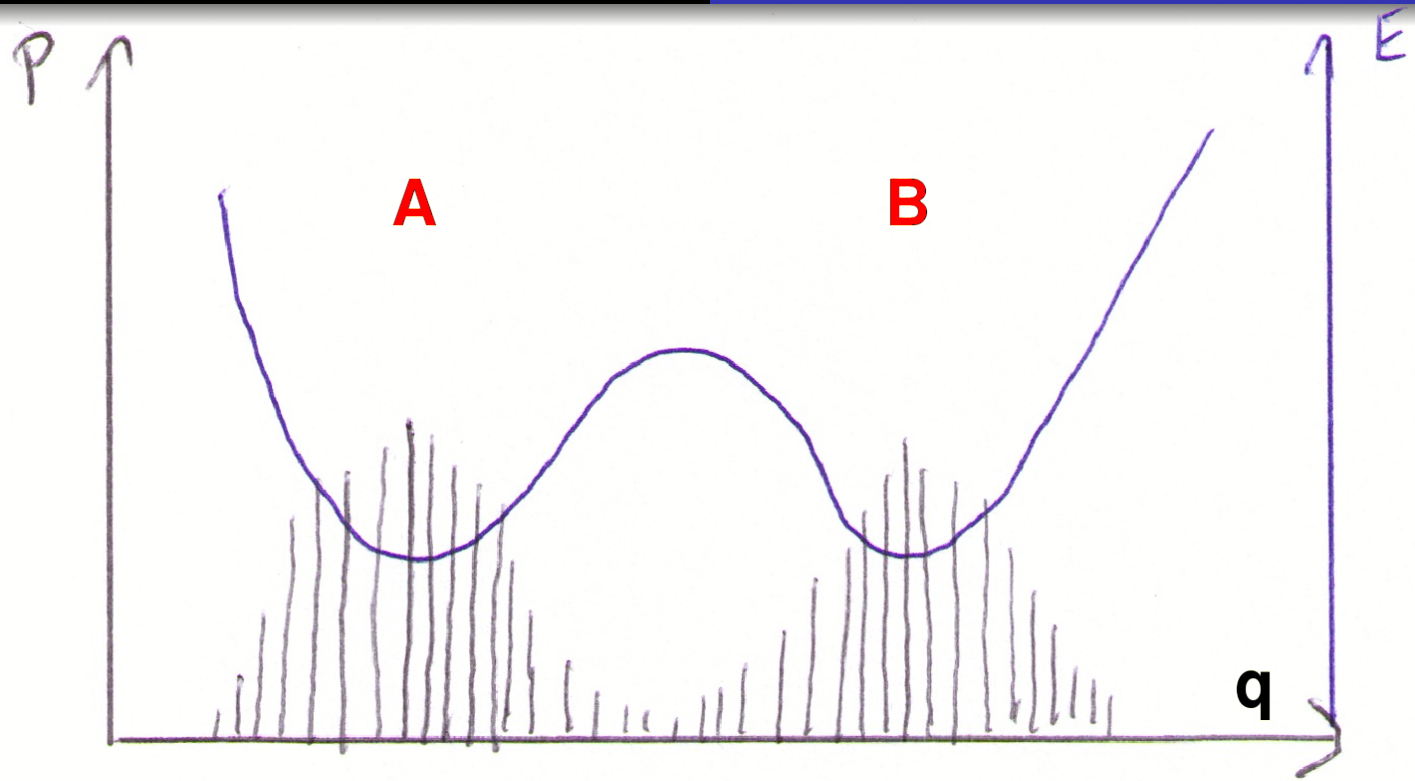
Straightforward approach

We perform an MD simulation for the system,
and then count how many times q takes the value q_1:
calculate the probability $P(q_1)$ of finding the system at q_1

Then, the free energy difference of two states with q_1 and q_2 is

$$F_2 - F_1 = -k_B T \ln \frac{P(q_2)}{P(q_1)}$$

which contains the equilibrium constant $P(2)/P(1)$



The problem:
If a high barrier has to be crossed to come from A to B,
a pure (unbiased) MD simulation will hardly make it
$\rightarrow$ the high-energy region (barrier) is described poorly (for sure)
$\rightarrow$ we may not obtain the product at all (possibly)

Working principle

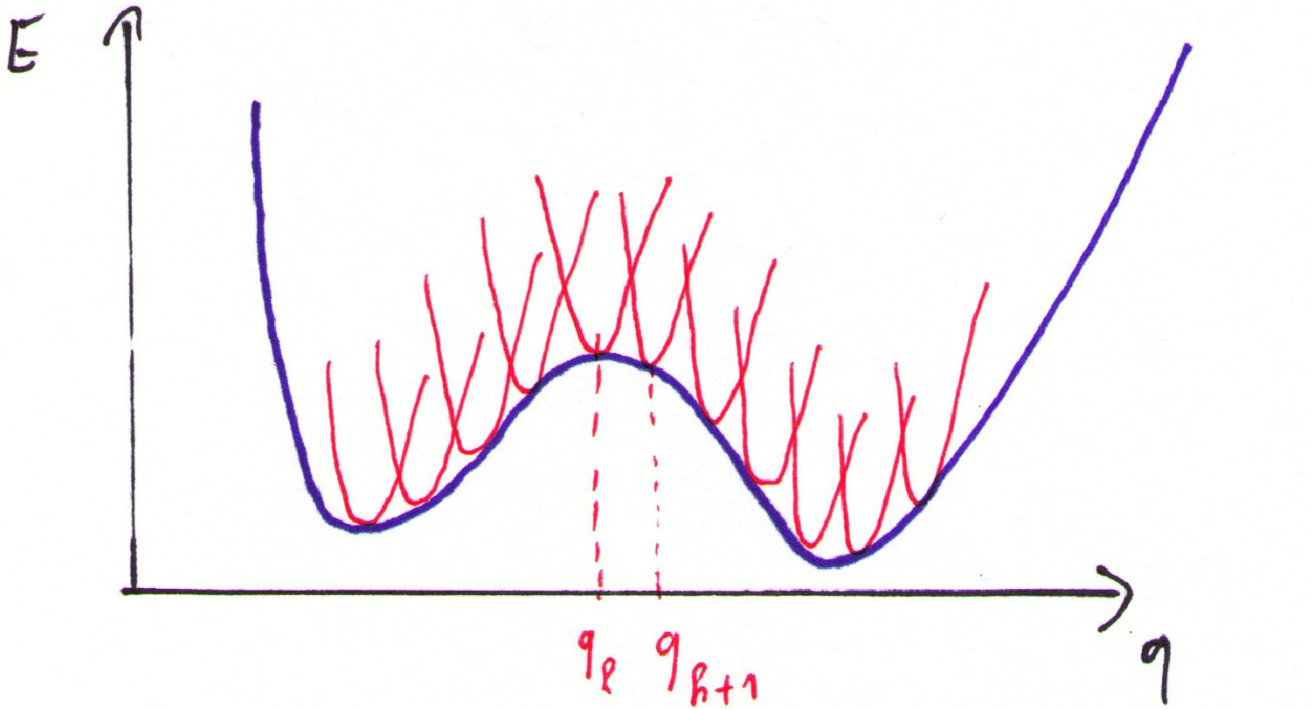
Apply an additional potential, also called biasing potential to restrain the system to values of reaction coordinate that would otherwise remain undersampled.

This is the principle of the umbrella sampling.

The additional potential will become a part of the force field, and it shall depend only on the reaction coordinate: $V = V(q)$

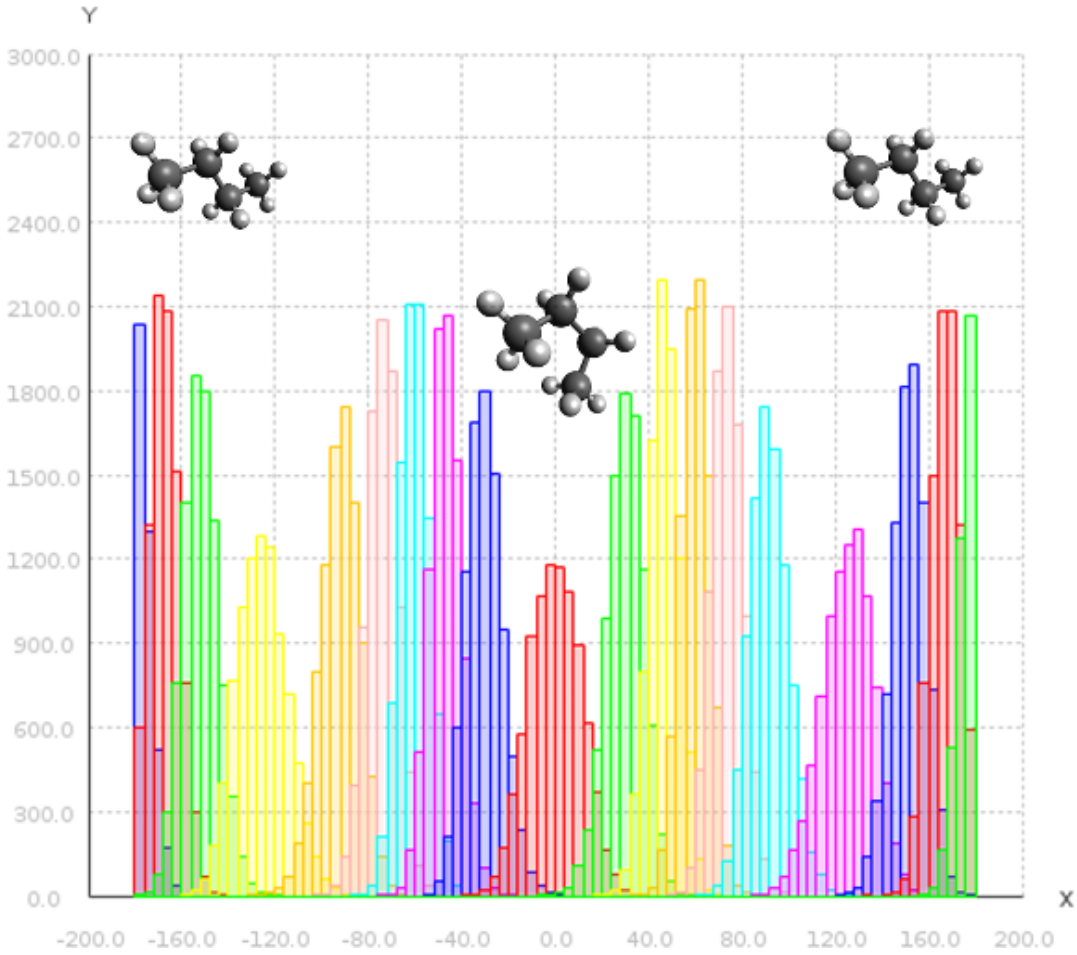
Practical umbrella sampling

We can use this scheme efficiently, by way of moving a biasing harmonic potential along the reaction coordinate:



Practical umbrella sampling

Example – probabilities from biased simulations – histograms



http://people.cs.uct.ac.za/~mkuttel/images/projectImages/WHAM.png

Potentials of mean force

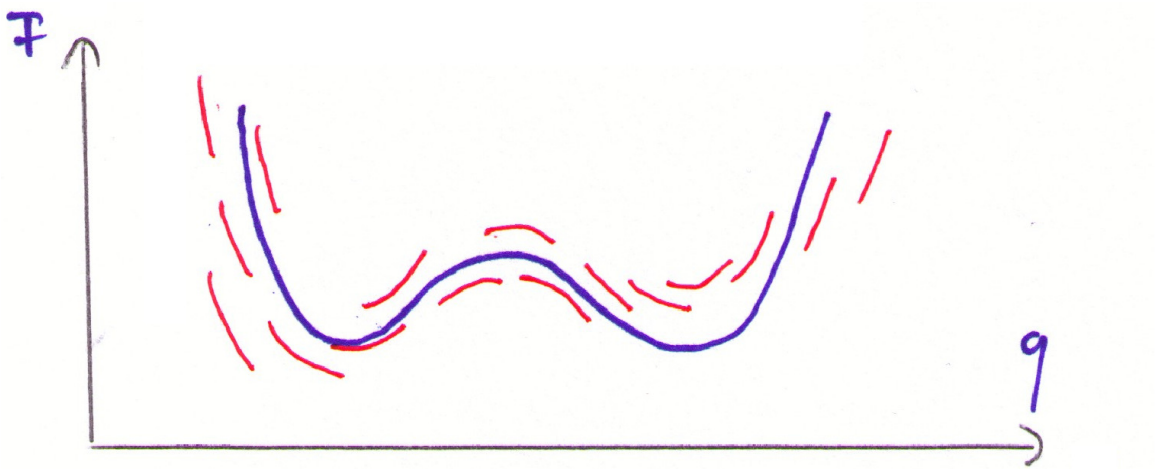
- We perform k simulations with biasing potentials V_k, which are designed to cover the interesting range of q
- From each of the k simulations, we extract the biased probability $P^*(q)$ of finding the system at the certain value of q – this differs from the real, unbiased probability $P(q)$
- We correct for bias to obtain the unbiased free energy $F(q)$

$$F_k(q) = -k_B T \ln P^*(q) - V_k(q) + K_k$$

where K_k is a constant shift that has to be determined

Potentials of mean force

- the curves $F_k(q)$ for simulations k and $k+1$ differ by a constant shift, difference of K
- main task – match the pieces to provide a continuous curve
- one way – fit the values K_k to obtain a resulting $F(q)$ curve that is as continuous and smooth as possible
- requirement – sufficient 'overlap' of the pieces F_k and F_{k+1}
- WHAM method – included in modern simulation programs



Molecular modeling in the drug design

Drug design

– to construct new chemical compounds interacting in a defined way with natural materials – proteins, NA, carbohydrates. . .
– typical example – find a potent inhibitor of an enzyme, which does not interact harmfully with other substances in the organism

– typical difficulties:

- the drug has to be a potent inhibitor
- it must not interact with other enzymes (might be lethal)
- it must not decompose too early (to reach destination)
- its metabolites must not be (too) toxic

hard and $$$ business

Molecular docking

"Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex." Wikipedia

Typical pharmacological problem – find a ligand molecule to bind to a protein as strongly and specifically as possible

Good news: the binding site (pocket) is usually known – often, the active or allosteric place of the protein

Bad news:

- many DoF – transl., rot. and internal flex. of the ligand
- only a small number of molecules can be docked manually, once the binding mode of a similar molecule is known (and, even similar molecules sometimes bind differently)

Molecular docking

a sequence of tasks:

1. Generate the pool of compounds to test – database of compounds, construction from a database of moieties,. . .
2. For each compound, find the binding mode. For this, try out several/many orientations and conformations (poses), and determine the most favorable
3. Evaluate the strength of the interaction. Accurate determination of ΔG_{bind} impossible; instead, a scoring function is employed

Molecular docking

Various levels of approximation may be employed

The simplest approach – exploit a database of molecules, and try to fit each molecule as a rigid body into the binding pocket
A natural expansion – consider the flexibility of the ligand

How to generate different configurations of the molecule?

- simple minimization or molecular dynamics
- Monte Carlo, perhaps combined with simulated annealing
- genetic algorithms

Efficient alternative – incremental construction of the ligand, which is partitioned into chemically reasonable fragments
– natural account for the conformational flexibility of the molecule

Molecular docking

problem of docking – it is all about sampling

No way to do molecular dynamics for every candidate molecule:

- MD takes much longer than what is affordable
  (would be OK for one ligand, but there are too many)
- MD would probably work only for quite rigid molecules moving relatively freely in the binding pocket (usually not the case)

Difference:

- If the goal is to dock a single molecule – a thorough search is affordable, involving MD, enhanced sampling. . .
- If we have to dock and assess many candidate ligands
  - simpler approaches have to be chosen
  - current state of the art – consider the flexibility of ligands
  - flexibility of protein (side chains) under development

Scoring function

needed: extremely efficient way to quantify the strength of binding

1. to find the right binding mode of each ligand
2. to compare the strength of binding of various ligands.

the quantity of interest – binding free energy
problem with free energy methods – too inefficient for docking
what we need here – a simple additive function to approximate ΔG_{bind}, which would give a result rapidly, in a single step

Scoring function

$$\Delta G_{\text{bind}} = \Delta G_{\text{solv}} + \Delta G_{\text{conf}} + \Delta G_{\text{int}} + \Delta G_{\text{rot}} + \Delta G_{\text{t/r}} + \Delta G_{\text{vib}}$$

ΔG_{solv} – change of hydration (ligand, protein) upon binding
ΔG_{conf} – deformation energy of the ligand (forced by the pocket)
ΔG_{int} – 'interaction energy' – a favorable contribution
due to the specific ligand–protein interactions
ΔG_{rot} – loss of entropy due to the frozen rotations
approx. $+RT \log 3 = 0.7$ kcal/mol per 3-state rotatable bond
$\Delta G_{\text{t/r}}$ – loss of trans. and rot. entropy upon association
– approx. the same for all ligands of similar size
ΔG_{vib} – change of vibrational modes – difficult, often ignored

Scoring function

- a 'force field' for the free energy of binding
- problem – although approximative, it is still too costly
- usually, very simple constructions, looking over-simplified in comparison with MM force fields; example (Böhm, 1994):

$$\Delta G = \Delta G_0 + \Delta G_{\text{Hbond}} \cdot \sum_{\text{Hbonds}} f(R,\alpha) + \Delta G_{\text{ionpair}} \cdot \sum_{\text{ionpairs}} f'(R,\alpha) + \Delta G_{\text{lipo}} \cdot A_{\text{lipo}} + \Delta G_{\text{rot}} \cdot N_{\text{rot}}$$

ΔG_{Hbond} – ideal hydrogen bond
$f(R,\alpha)$ – penalty function for a realistic hydrogen bond
$\Delta G_{\text{ionpair}}$ and $f'(R,\alpha)$ – dtto for ionic contacts
ΔG_{lipo} – due to hydrophobic interaction; non-polar SA A_{lipo}
ΔG_{rot} – due to a rotatable bond that freezes upon binding

Scoring function

Further concepts present in other scoring functions:

- partitioning of the surface areas of both the proteins and the ligand into polar and non-polar regions, and assigning different parameters to the interactions of different kinds of regions (polar-polar, polar-nonpolar, nonpolar-nonpolar)
- statistical techniques to parametrize the scoring function

Problem – such s.f. only describe well ligands that bind tightly
Modestly binding ligands
  – of increasing interest in docking studies
  – more poorly described by such functions

Possible solution – 'consensus scoring' – combining results from several scoring functions; performs better than any single s.f.

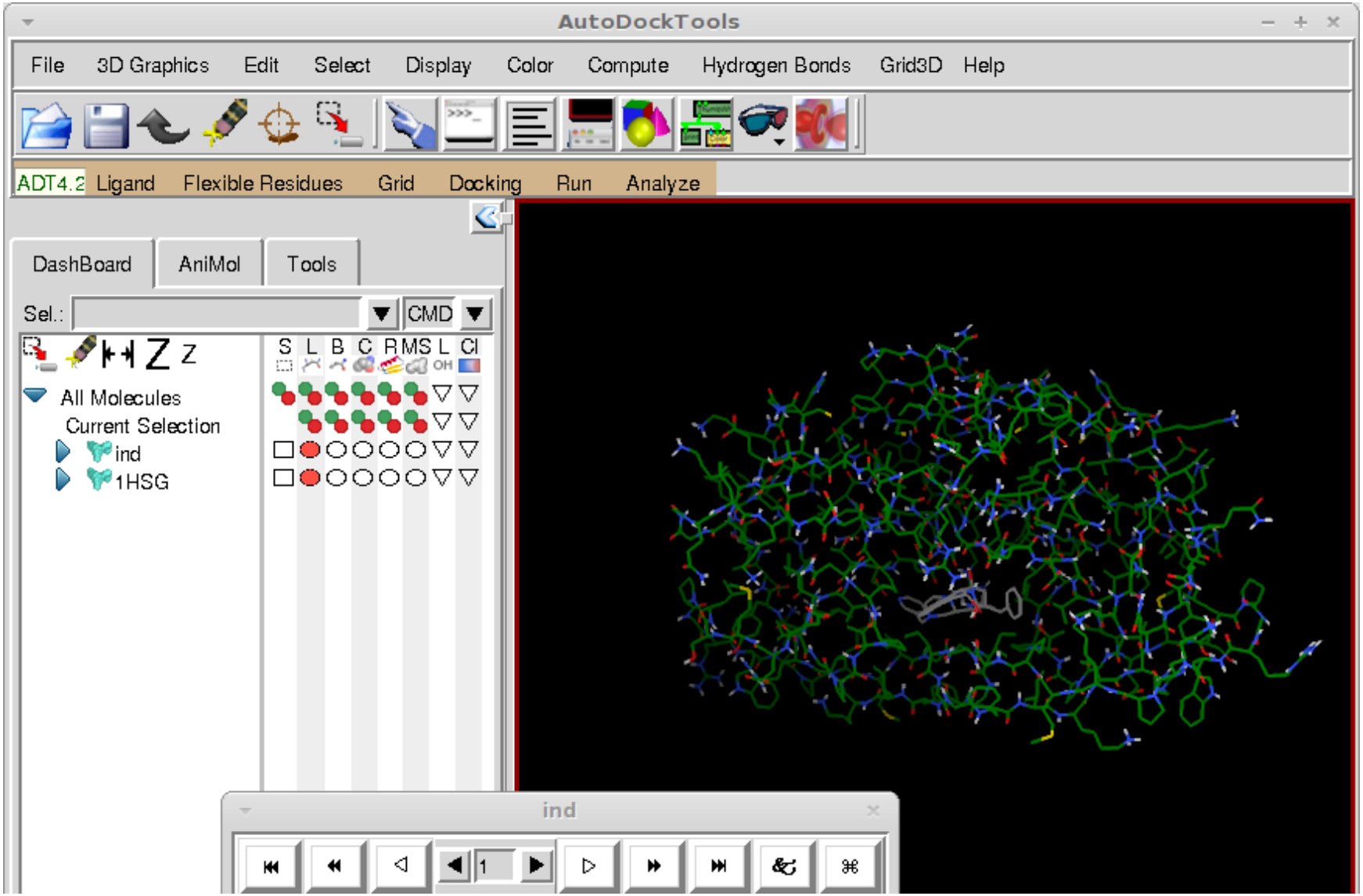
Scoring function

Comment on accuracy

an error of ΔG_{bind} of 1.4 kcal/mol
$\rightarrow$ ten-fold increase/decrease of the inhibition constant

or: as little as 4.2 kcal/mol of ΔG_{bind} lies
between a micro- and a nanomolar inhibitor

Therefore, the requirements on the accuracy of s.f.
are actually rather big


AutoDockTools
File
3D Graphics
Edit
Select
Display
Color
Compute
Hydrogen Bonds
Grid3D
Help
ADT4.2
Ligand
Flexible Residues
Grid
Docking
Run
Analyze
DashBoard
AniMol
Tools
Sel.:
CMD
S L B C R MS L Cl
All Molecules
Current Selection
ind
1HSG
ind
1

De novo design of ligands

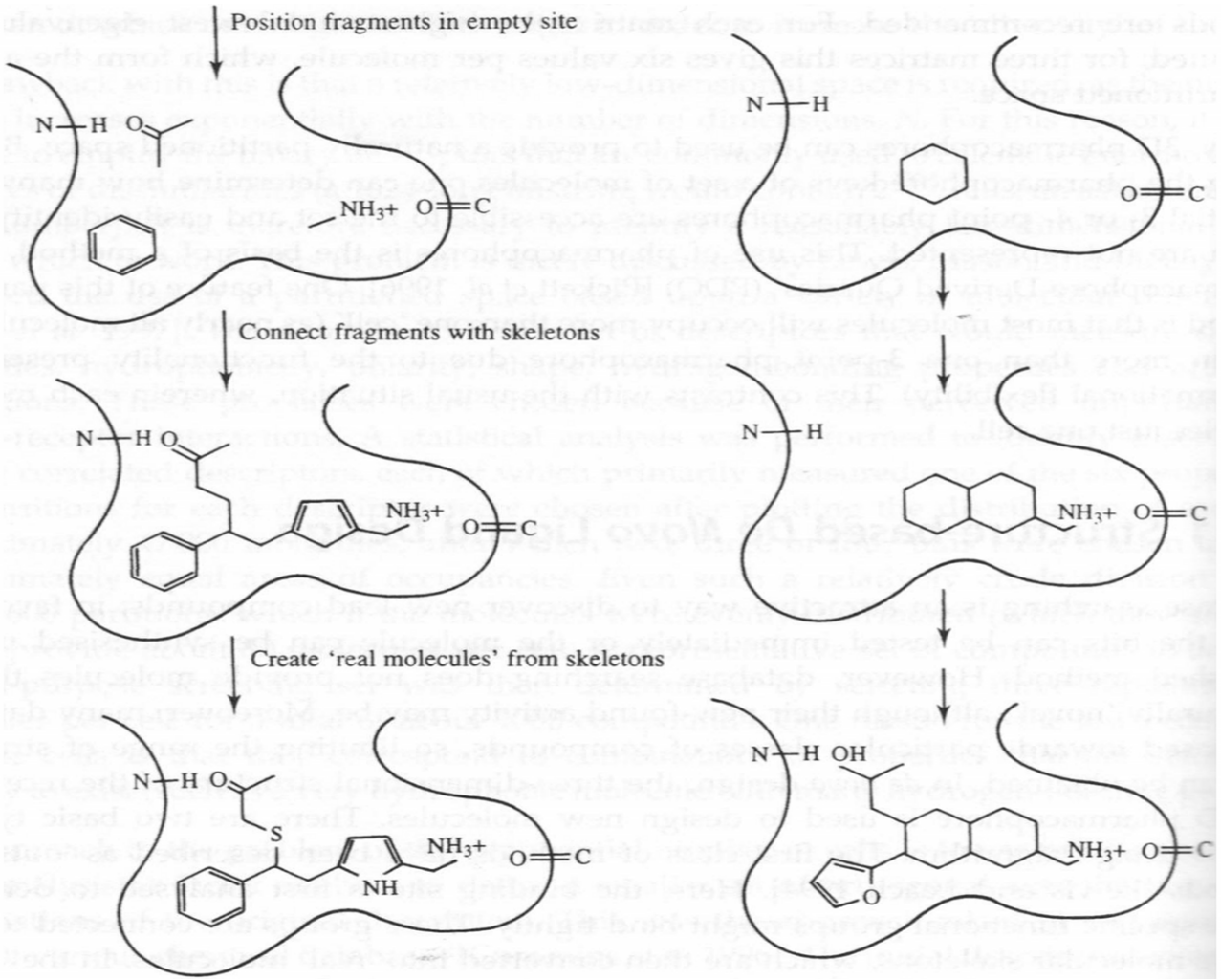
It may be a good idea to construct the ligand 'from scratch' – without relying on the content of a database.

2 basic types of *de novo* design:

- outside–in: Binding site is analyzed and tightly-binding ligand fragments are proposed. They are connected (db of linkers) $\rightarrow$ molecular skeleton of the ligand $\rightarrow$ actual molecule.
- inside–out: 'growing' the ligand in the binding pocket, driven by a search algorithm with a scoring function.

De novo design of ligands

OUTSIDE IN



INSIDE OUT

Molecular docking

Glossary of terms

- receptor / host / lock
- ligand / guest / key
- docking
- binding mode – position and orientation of ligand
- pose – a candidate for the binding mode
- scoring – determine how favorable a pose is
- ranking – of the poses to determine the binding mode