Conformational space and molecular dynamics

Explorant l'espai de moviments de la proteïna, i més...

Jordi Villà i Freixa

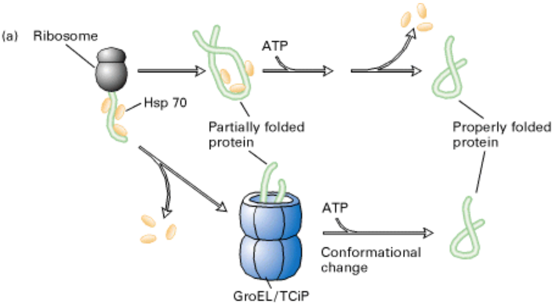
jordi.villa@upf.edu // cbbl.imim.es
Grup de Recerca en Informàtica Biomèdica
Universitat Pompeu Fabra
Barcelona

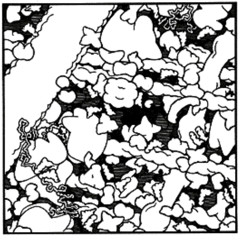
Barcelona, April 2010

What do we want to achieve?

- Use basic classical mechanics as a way to explore protein conformations (dynamics), to start with.
- Be sure we deal with a correct description of the interactions (energetics) (and if we do a good job we will be able to go beyond conformations)

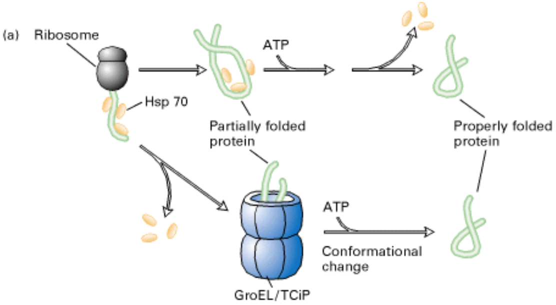
Are we sure?

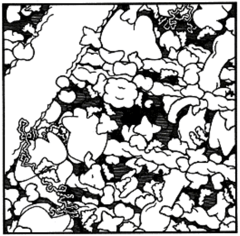




[Goodsell, 1998]

Are we sure?

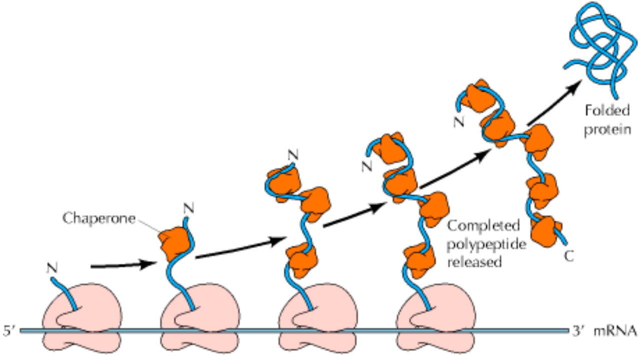




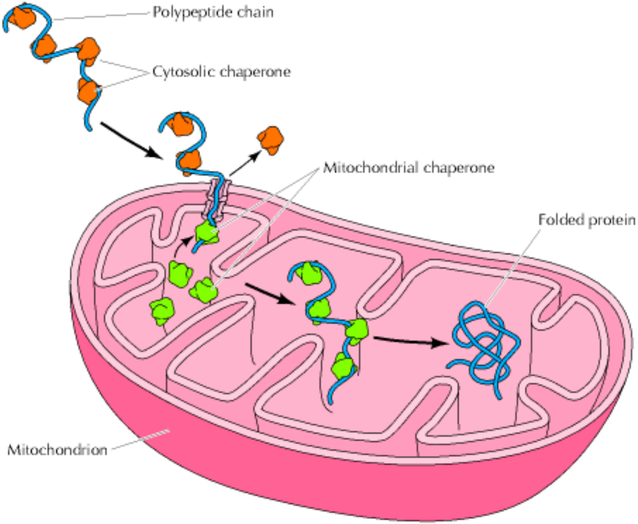
[Goodsell, 1998]

The crowded environment complicates folding *in vivo*

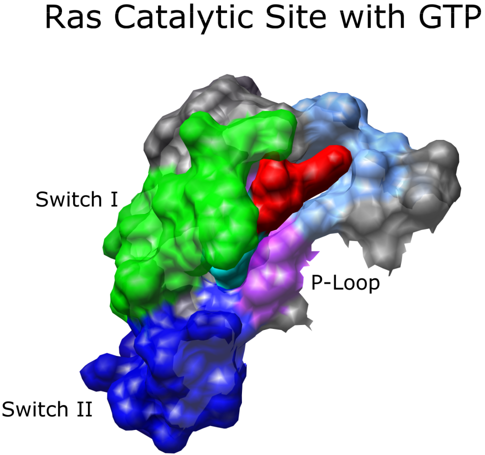
Are we sure?


Chaperone
N
N
N
N
N
Completed polypeptide released
C
Folded protein
5′
3′ mRNA

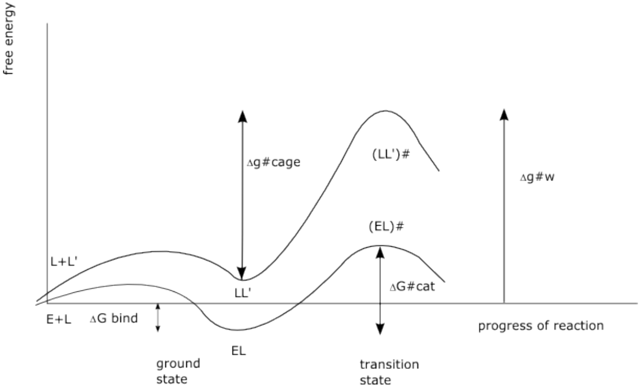
Are we sure?


Polypeptide chain
Cytosolic chaperone
Mitochondrial chaperone
Folded protein
Mitochondrion

Will we stop in structure?



Will we stop in structure?



(not-so-short) Basic background

(not-so-short) Basic background

- Essentials of thermodynamics

(not-so-short) Basic background

- Essentials of thermodynamics
- Statistical Mechanics principles

(not-so-short) Basic background

- Essentials of thermodynamics
- Statistical Mechanics principles
- Potential energy surfaces

(not-so-short) Basic background

- Essentials of thermodynamics
- Statistical Mechanics principles
- Potential energy surfaces
- Molecular simulations

Thermodynamics: some definitions

- *State variables* or *state functions*: variables that depend only on the state of the system: P, V, T, U... Other variables are dependent of the path followed to obtain them: w
- *Mechanical* and *non mechanical* variables: the former are additive, depending on the amount of particles in the system (P, V, U...) and the latter are independent of this value (T).

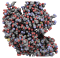
Path-dependant functions

Two important path-dependant functions:

work $w = \int_{\text{path}} \mathbf{Fdx}$ or, in terms of the change in volume of a gas expanding against an external pressure: $w = -\int_{V_1}^{V_2} p_{\text{ext}} dV$

heat the heat needed to increase the temperature of N moles of a given substance can be evaluated by: $q_V = \int_{T_1}^{T_2} nC_{V,m} dT$ for a constant volume process and $q_P = \int_{T_1}^{T_2} nC_{P,m} dT$ for constant pressure. C is called the heat capacity.

Thermodynamics laws



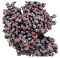
1st law $q + w = \Delta U$, where the internal energy U is a state function. For an ideal gas the internal energy is only dependent on the temperature. At constant pressure, $\Delta H = q_P$

2nd law Defining the entropy of a system as $\Delta S_{\mathrm{sys}} = \frac{q_{\mathrm{rev}}}{T}$, where q_{rev} is heat given reversivebly (without change in the temperature). All spontaneous (non-reversible) processes imply an increase of the universe entropy:

$$\Delta S_{\mathrm{universe}} = \Delta S_{\mathrm{sys}} + \Delta S_{\mathrm{surround}} > 0$$

3rd law The entropy of all pure substances at 0K is zero.

Obtaining Gibbs free energy expression



It is easy to see how, for a constant pressure chemical reaction:

$$\Delta S_{\text{universe}} = \Delta S_{\text{sys}} + \Delta S_{\text{surround}} = \Delta S_{\text{sys}} - \frac{\Delta H}{T}$$

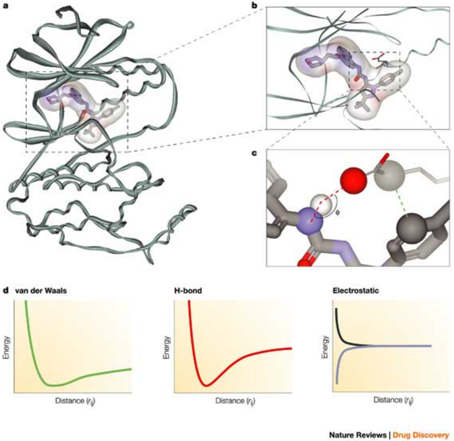
which, multiplied by T yields the usual expresion for the Gibbs free energy:

$$\Delta G = -T\Delta S_{\text{universe}} = \Delta H - T\Delta S$$

By using normal Gibbs free energies, we can relate thermodynamic quantities to equilibrium properties in a chemical reaction:

$$\Delta G^{\circ} = -RT \ln K^{\circ}$$

So, all this was macroscopic... what about the microscopic view?

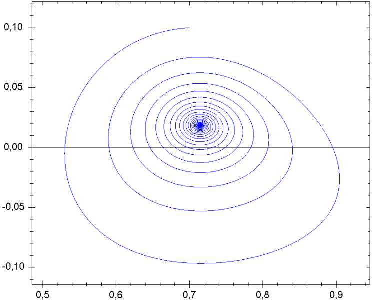


[Kitchen et al., 2004]

Statistical mechanics is the answer

Phase Space of a dynamical system

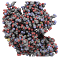
Phase space is the space of all possible states:



For N atoms, $6N$ values are required to define its state:

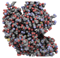
- $3N$ positions: $\mathbf{r}^N = x_1, y_1, z_1, x_2, \ldots, z_N$
- $3N$ momenta: $\mathbf{p}^N = m\frac{d\mathbf{r}^N}{dt} = p_{1x}, p_{1y}, \ldots, p_{Nz}$

Time average



All properties of a macroscopic system depend on the positions and momenta of all N particles that comprise the system. We can define a instantaneous value of a certain observable as $\mathcal{A}(\mathbf{p}^N(t), \mathbf{r}^N(t))$. However, this quantity $\mathcal{A}$ suffers fluctuations.

Time average

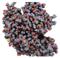


All properties of a macroscopic system depend on the positions and momenta of all N particles that comprise the system. We can define a instantaneous value of a certain observable as $\mathcal{A}(\mathbf{p}^N(t), \mathbf{r}^N(t))$. However, this quantity $\mathcal{A}$ suffers fluctuations. Over time, its average is found doing:

$$\mathcal{A}_{\mathrm{ave}} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} \mathcal{A}(\mathbf{p}^N(t), \mathbf{r}^N(t))\,dt$$

Which is directly comparable with the experimentally observable value of $\mathcal{A}$.

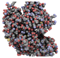
Ensemble average



Virtually impossible to achieve for the phase space in a macroscopic system.
Instead, by the ergodic hypothesis the *time* average is replaced by an *ensemble* average (or expectation value):

$$< \mathcal{A} > = \int\int d\mathbf{p}^N d\mathbf{r}^N \mathcal{A}(\mathbf{p}^N, \mathbf{r}^N) \underbrace{\rho(\mathbf{p}^N, \mathbf{r}^N)}_{\text{probability density}}$$

Probability density in *NVT*



Under conditions of constant N, V, and T, $\rho(\mathbf{p}^N, \mathbf{r}^N)$ takes the form of the Boltzmann distribution:

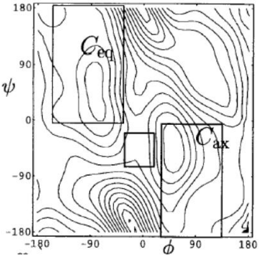
$$\rho(\mathbf{p}^N, \mathbf{r}^N) = \frac{\exp(-E(\mathbf{p}^N, \mathbf{r}^N)/k_{\mathrm{B}} T)}{Q}$$

So, finally we reach the key value: E.
Q is the partition function, E is the energy, k_{B} is the Boltzmann constant and T is the temperature. In the case of an *NVT* system of N identical particles, the classical partition function is:

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}} \int\int d\mathbf{p}^N d\mathbf{r}^N \exp\left[-\frac{H(\mathbf{p}^N, \mathbf{r}^N)}{k_{\mathrm{B}} T}\right]$$

with $H(\mathbf{p}^N, \mathbf{r}^N) = K(\mathbf{p}^N) + V(\mathbf{r}^N)$

How can we use Q?

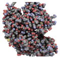


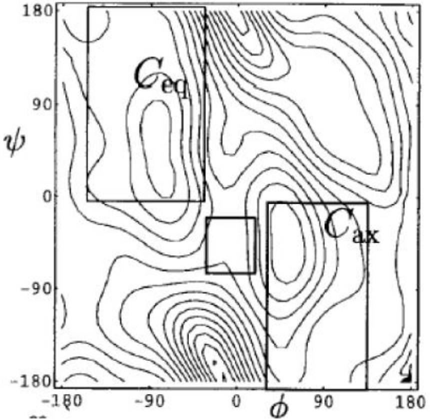
From the partition function we can evaluate several thermodynamic quantities. Among them, some mechanical properties (U) and other non-mechanical properties (like the Helmholtz free energy, A):

$$U = \frac{k_B T^2}{Q} \frac{\partial Q}{\partial T}$$

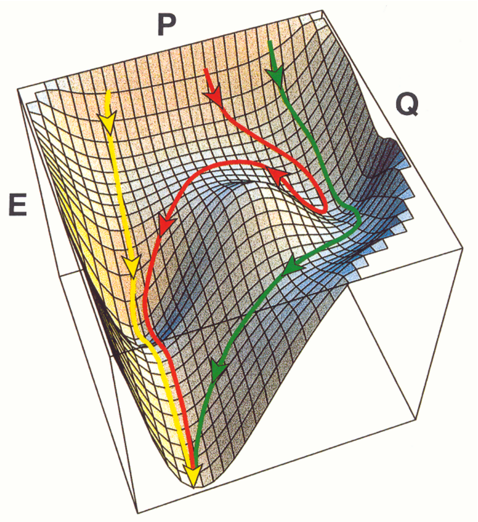
$$A = -k_B T \ln Q$$

Potential energy surface

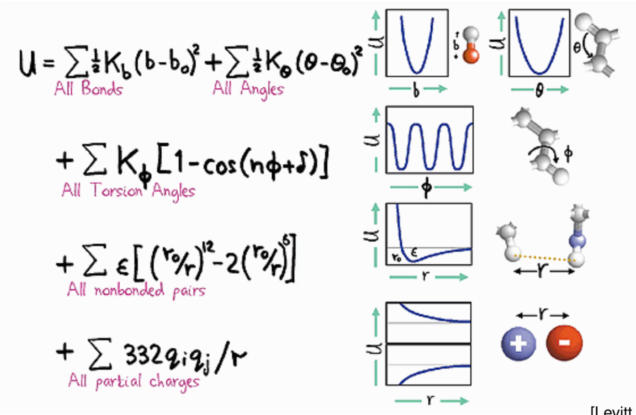




Potential energy surface



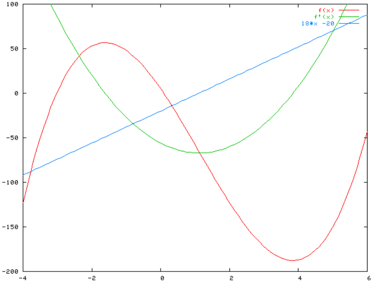
Potential energy surface



$$U = \sum \tfrac{1}{2}K_b(b-b_0)^2 + \sum \tfrac{1}{2}K_\theta(\theta-\theta_0)^2 + \sum K_\phi[1-\cos(n\phi+\delta)] + \sum \epsilon\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right] + \sum 332 q_i q_j / r$$

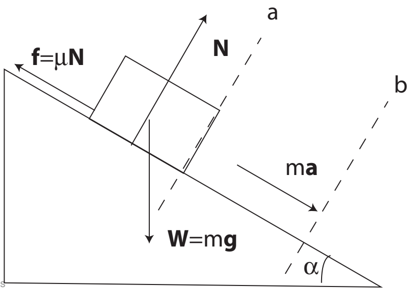
[Levitt, 2001]

On functions and derivatives

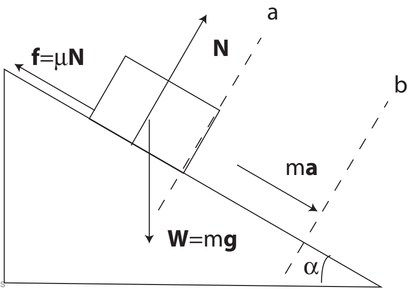


$$
\begin{aligned}
f(x) &= 3x^2 - 10x^2 - 56x + 5 \\
f'(x) &= 9x^2 - 20x - 56 \\
f''(x) &= 18x - 20
\end{aligned}
$$

$$\sum \mathbf{F} = m\mathbf{a}$$

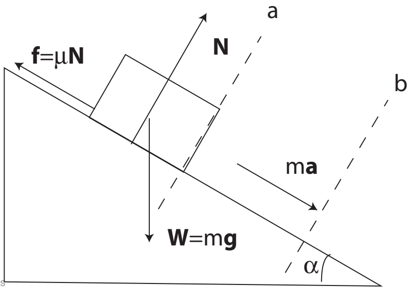


$$\sum \mathbf{F} = m\mathbf{a}$$



$$\sum \mathbf{F} = m\mathbf{a}$$

$$\begin{cases} mg\sin\alpha - \mu N = ma \\ -mg\cos\alpha + N = 0 \end{cases}$$

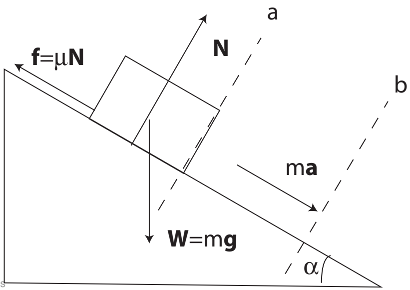


$$\sum \mathbf{F} = m\mathbf{a}$$

$$\begin{cases} mg\sin\alpha - \mu N = ma \\ -mg\cos\alpha + N = 0 \end{cases}$$

$$a = g(\sin\alpha - \mu\cos\alpha)$$

$$a = \frac{dv}{dt} = \frac{d^2x}{dt^2}$$



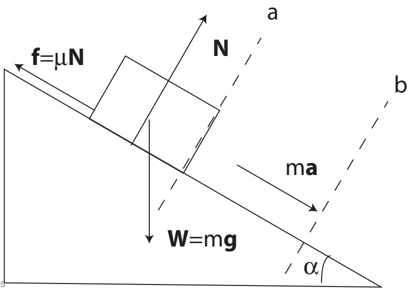
$$\sum \mathbf{F} = ma$$

$$\begin{cases} mg\sin\alpha - \mu N = ma \\ -mg\cos\alpha + N = 0 \end{cases}$$

$$a = g(\sin\alpha - \mu\cos\alpha)$$

$$a = \frac{dv}{dt} = \frac{d^2x}{dt^2}$$

$$\begin{cases} v - v_0 = \int_0^t adt = at \\ x - x_0 = \int_0^t vdt = \int_0^t (v_0 + at)dt = v_0 t + \frac{1}{2}at^2 \end{cases}$$



$$\sum \mathbf{F} = m\mathbf{a}$$

$$\begin{cases} mg\sin\alpha - \mu N = ma \\ -mg\cos\alpha + N = 0 \end{cases}$$

$$a = g(\sin\alpha - \mu\cos\alpha)$$

$$a = \frac{dv}{dt} = \frac{d^2x}{dt^2}$$

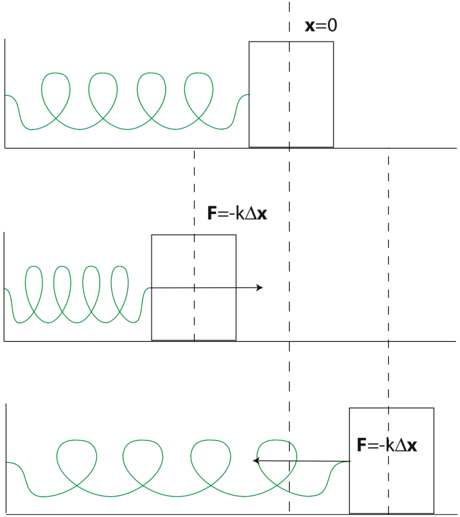
$$\begin{cases} v - v_0 = \int_0^t a dt = at \\ x - x_0 = \int_0^t v dt = \int_0^t (v_0 + at) dt = v_0 t + \frac{1}{2}at^2 \end{cases}$$

$$dW = \mathbf{F} \cdot \mathbf{dx}$$

$$W_{a\to b} = \int_a^b \mathbf{F} \cdot \mathbf{dx}$$

$$W_{a\to b} = \int_a^b mg(\sin\alpha - \mu\cos\alpha) dx = mg(\sin\alpha - \mu\cos\alpha)\Delta x$$

Hooke's law



Hooke's law

In one dimension:

$$m\frac{d^2x}{dt^2} = -kx$$

Hooke's law

In one dimension:

$$m\frac{d^2x}{dt^2} = -kx$$

If we define $\omega^2 = k/m$ we reduce the problem to:

$$\frac{d^2x}{dt^2} + \omega^2 x = 0$$

Hooke's law

In one dimension:

$$m\frac{d^2x}{dt^2} = -kx$$

If we define $\omega^2 = k/m$ we reduce the problem to:

$$\frac{d^2x}{dt^2} + \omega^2 x = 0$$

Which has the following solution:

$$\begin{aligned} x &= A\sin(\omega t + \phi) \\ v &= A\omega\cos(\omega t + \phi) \\ a &= -A\omega^2\sin(\omega t + \phi) = -\omega^2 x \end{aligned}$$

$$k = m\omega^2$$

And it is easy to see that the period can be evaluated as: $\tau = 2\pi\sqrt{m/k}$

Hooke's law and harmonic potential

$$
\begin{aligned}
\vec{F} &= -k\vec{x} = m\vec{a} = m\frac{d^2\vec{x}}{dt^2} \\
dW &= \vec{F}\cdot\vec{dx} \\
W_{0\to x} &= \int_0^x \vec{F}\cdot\vec{dx} \\
W_{0\to x} &= \int_0^x (-kx)dx = -\frac{1}{2}kx^2 = -\Delta U_{0\to x} \\
F &= -\frac{dU}{dx}
\end{aligned}
$$

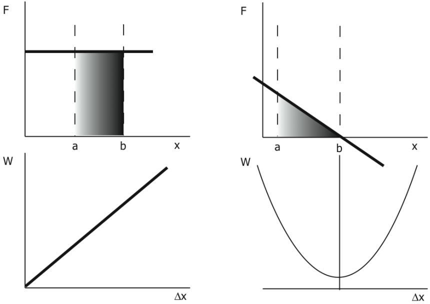
Hooke's law and harmonic potential

$$\begin{aligned}
\vec{F} &= -k\vec{x} = m\vec{a} = m\frac{d^2\vec{x}}{dt^2} \\
dW &= \vec{F}\cdot\vec{dx} \\
W_{0\to x} &= \int_0^x \vec{F}\cdot\vec{dx} \\
W_{0\to x} &= \int_0^x (-kx)\,dx = -\frac{1}{2}kx^2 = -\Delta U_{0\to x} \\
F &= -\frac{dU}{dx}
\end{aligned}$$

In general, for a system with N coordinates:

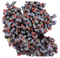
$$\boxed{\vec{F} = -\vec{\nabla}U = -\left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_N}\right) U}$$

Force and potential



- The work done corresponds to the area delimited in the $F = f(x)$ plot
- The force is the opposite of the derivative of the potential energy with respect to the displacement at each point of the $W = f(x)$ plot

Exploring PES



Molecular dynamics or Monte Carlo simulations can be used to generate an ensemble of systems, characterized by their energy. The canonical (*NVT*) ensemble partition function form can be evaluated from such simulation by:

$$Q = \sum_i \exp(-\varepsilon_i / k_\mathrm{B} T)$$

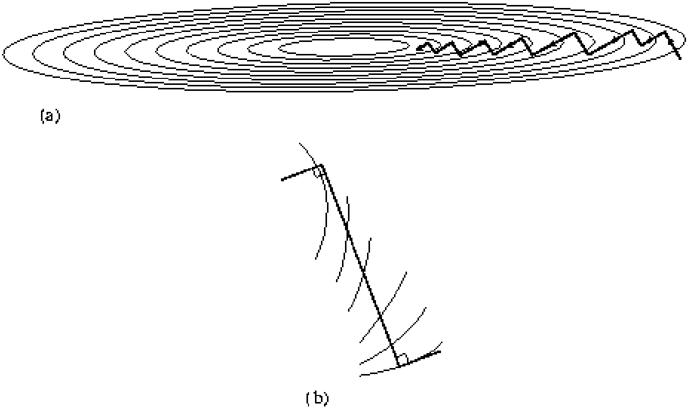
internal energy $U = k_\mathrm{B} T^2 \left(\frac{\partial \ln Q}{\partial T} \right)_V$

enthalpy $H = k_\mathrm{B} T^2 \left(\frac{\partial \ln Q}{\partial T} \right)_V + k_\mathrm{B} T V \left(\frac{\partial \ln Q}{\partial V} \right)_T$

Helmholtz free energy $A = -k_\mathrm{B} T \ln Q$

Gibbs free energy $G = -k_\mathrm{B} T \ln Q + k_\mathrm{B} T V \left(\frac{\partial \ln Q}{\partial V} \right)_T$

But, before, let us find minima on the PES



Minimization methods: Gradient search

Starting at point $\mathbf{x}_0$. As many times as needed, move from point $\mathbf{x}_i$ to $\mathbf{x}_{i+1}$ by minimizing along the line from $\mathbf{x}_i$ in the direction of the local downhill gradient $-\nabla f(\mathbf{x}_i)$.

Minimization methods: Newton-Raphson

From the Taylor expansion of the function:

$$f(\mathbf{x}) \approx f(\mathbf{x}_k) + \underbrace{(\mathbf{x}-\mathbf{x}_k)^T \cdot \mathbf{g}_k}_{\text{linear term}} + \underbrace{\frac{1}{2}(\mathbf{x}-\mathbf{x}_k)^T \cdot \mathbf{H}_k \cdot (\mathbf{x}-\mathbf{x}_k)}_{\text{quadratic term}}$$

we can take derivatives

$$\nabla f(\mathbf{x}) = \mathbf{g}_k + \mathbf{H}_k \cdot (\mathbf{x}-\mathbf{x}_k)$$

If we assume that $f(\mathbf{x})$ takes its minimum at $\mathbf{x} = \mathbf{x}^*$, the gradient is zero:

$$\mathbf{H}_k \cdot (\mathbf{x}^* - \mathbf{x}_k) + \mathbf{g}_k = 0$$

which is a simple linear system. The Newton-Raphson considers $\mathbf{x}^*$ to be the next point in the iterative formula:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}_k^{-1} \cdot \mathbf{g}_k$$

Minimization methods: Conjugated gradients I

- Let's come back to the gradient search
- Let's minimize $f(\mathbf{x})$ over the hyperplane that contains all previous search directions.

$$\mathbf{x}_0 + < \mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_i >$$

- If the vectors $\mathbf{p}_i$ are chosen to be L.I. we should ideally perform only N searches.

$$f(\mathbf{x}) \approx c - \mathbf{g} \cdot \mathbf{x} + \frac{1}{2}\mathbf{x} \cdot \mathbf{H} \cdot \mathbf{x}$$

- initial gradient $\mathbf{g}_0$ and an initial $\mathbf{h}_0 = \mathbf{g}_0$
- the CG method will construct $\mathbf{g}_{i+1} = \mathbf{g}_i - \lambda \mathbf{H} \cdot \mathbf{h}_i$ and $\mathbf{h}_{i+1} = \mathbf{g}_{i+1} + \gamma \mathbf{h}_i$

Minimization methods: Conjugated gradients II

- these vectors satisfy the orthogonality and conjugacy conditions:

$$\mathbf{g}_i \cdot \mathbf{g}_j = 0$$

$$\mathbf{h}_i \cdot \mathbf{H} \cdot \mathbf{h}_j = 0$$

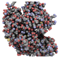
$$\mathbf{g}_i \cdot \mathbf{h}_j = 0$$

and the scalars are given by:

$$\lambda_i = \frac{\mathbf{g}_i \cdot \mathbf{g}_i}{\mathbf{h}_i \cdot \mathbf{H} \cdot \mathbf{h}_i}$$

$$\gamma_i = \frac{\mathbf{g}_{i+1} \cdot \mathbf{g}_{i+1}}{\mathbf{g}_i \cdot \mathbf{g}_i}$$

Simulated annealing (I)

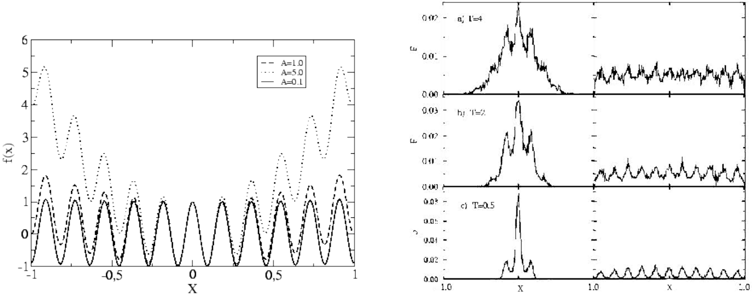


- SA simulates the finite T dynamics of the system
- Starting from $\mathbf{r}$ with energy $E(\mathbf{r})$ one generates a new $\mathbf{r}'$ with energy $E(\mathbf{r}')$ which replaces the original configuration with some probability function, e.g.:

$$P = \begin{cases} \exp\left(-\beta\left[E(\mathbf{r}') - E(\mathbf{r})\right]\right) & \text{if } E(\mathbf{r}') > E(\mathbf{r}) \\ 1 & \text{otherwise} \end{cases}$$

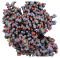
- At a given β SA samples the configurations $\mathbf{r}$ of the PES according to their thermodynamic probability.
- basic hopping technique, analogous to Monte Carlo/Metropolis algorithm when a Boltzmann distribution is used to decide the probability.

Simulated annealing (II)



(a) $f(x) = Ax^2 + cos(x/n)$ with different ruggedness (b) Distribution of 10000 SA processes started at random initial positions for the PES with A=1 (left) and A=0.1(right) at the given T

Monte Carlo/Metropolis algorithm



Each iteration is generated by randomly moving one atom or molecule. The energy of the new configuration is evaluated:

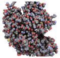
- if $V_i \leqslant V_{i-1}$ the new configuration is accepted
- if $V_i > V_{i-1}$ the new configuration is accepted with a probability evaluated from the Boltzmann factor of the differences in energy (when comparing with a random number in the interval $[0, 1]$

then:

$$< A >= \frac{1}{M} \sum_{i=1}^{M} A(\mathbf{r}^N)$$

It uses only the potential energy function ($H(\mathbf{r}^N) = V(\mathbf{r}^N)$)
Specially fitted for the canonical (NVT) ensemble; its equilibrium state is characterized by the minimum Helmholt free energy (A).

MC vs SA



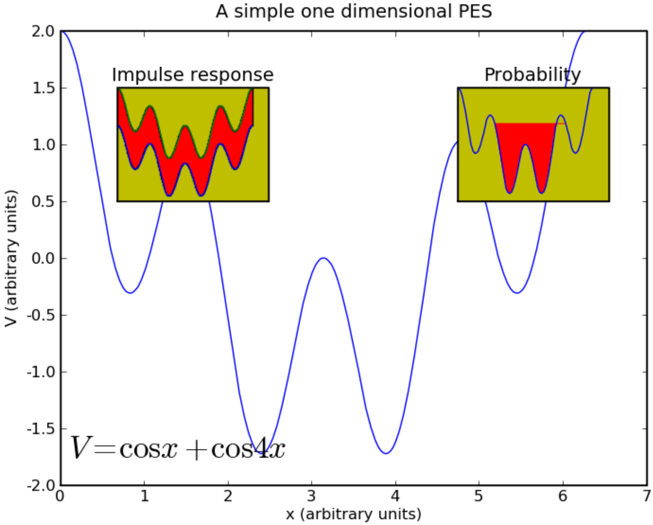
MC the probability function is Boltzmann distribution. The final distribution resembles the thermodynamic equilibrium distribution over physical states.

SA Any function that is useful to reach a minimum is ok. We just look for a minima, and no equilibrium distribution is sought!

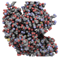
Other ensembles

- canonical: NVT: equilibrium state characterized by minimal A (Helmholtz free energy)
- microcanonical: NVE: equilibrium state characterized by maximum S
- isothermal-isobaric: fixed N, T and P; its equilibrium state is the minimum Gibbs function (G)
- grand canonical: fixed μ (chemical potential), V and T; its equilibrium state is characterized by the maximum value of PV.

NVE vs NVT


A simple one dimensional PES
Impulse response
Probability
$V = \cos x + \cos 4x$
V (arbitrary units)
x (arbitrary units)
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
0
1
2
3
4
5
6
7

Molecular dynamics

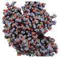


Thermodynamic averages are obtained from molecular dynamics as time averages using numerical integration of a given property A:

$$< A >= \frac{1}{M} \sum_{i=1}^{M} A(\mathbf{p}^N, \mathbf{r}^N)$$

MD provides kinetic energy contribution ($H(\mathbf{p}^N, \mathbf{r}^N) = K(\mathbf{p}^N) + V(\mathbf{r}^N)$) to the total energy. Specially fitted for the microcanonical (NVE) ensemble; its equlibrium state is characterized by the maximum entropy (S).

Molecular dynamics



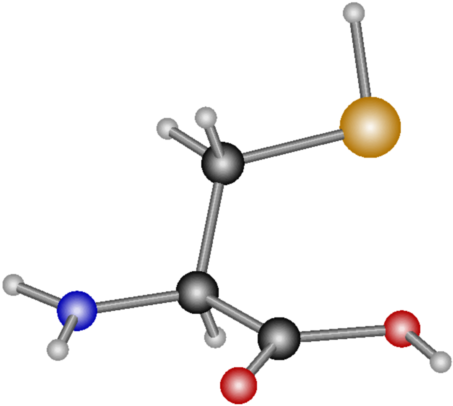
We need to integrate the following equation

$$\mathbf{F} = -\nabla U = m\frac{d^2\mathbf{x}}{dt^2}$$

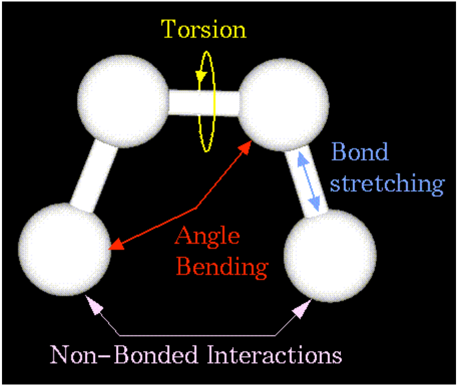
All algorithms for integration assume that the positions and dynamical properties (velocities, accelerations, etc.) can be approximated as Taylor series expansions. In the case of the Verlet algorithm:

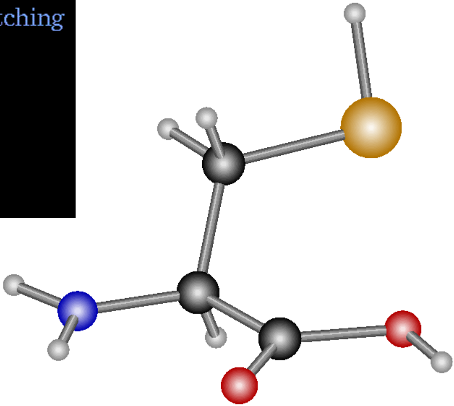
$$\begin{aligned}
\mathbf{x}(t+\delta t) &= \mathbf{x}(t) + \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) \\
\mathbf{v}(t+\delta t) &= \mathbf{v}(t) + \frac{1}{2}\delta t[\mathbf{a}(t) + \mathbf{a}(t+\delta t)]
\end{aligned}$$

A "real" system



A "real" system





Bond interactions

$$U_b = \sum_{i}^{N_b} k_b^i (r_i - r_0^i)^2$$

$$\mathbf{F} = -\nabla U$$

Bond force:

$$\mathbf{F}_b = -\left(\frac{\partial U_b}{\partial x_1}, \frac{\partial U_b}{\partial x_2}, \cdots \frac{\partial U_b}{\partial x_{3N}} \right)$$

Let's consider one single bond:

$$\begin{cases} \dfrac{\partial U_b^{ab}}{\partial x_a} = \dfrac{\partial U_b^{ab}}{\partial r_{ab}} \dfrac{\partial r_{ab}}{\partial x_a} \\ \dfrac{\partial U_b^{ab}}{\partial y_a} = \dfrac{\partial U_b^{ab}}{\partial r_{ab}} \dfrac{\partial r_{ab}}{\partial y_a} \end{cases}$$

Non-bonded interactions

$$U_{nb} = \underbrace{\sum_{i}^{N_{nb}} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]}_{\text{vanderWaals}} + \underbrace{\sum_{i}^{N_{nb}} \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_a q_b}{r}}_{\text{electrostatics}}$$

$$\mathbf{F} = -\nabla U$$

Non bonded force:

$$\mathbf{F}_{nb} = -\left(\frac{\partial U_b}{\partial x_1}, \frac{\partial U_b}{\partial x_2}, \cdots \frac{\partial U_b}{\partial x_{3N}}\right)$$

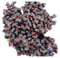
Let’s consider a single interaction:

$$\begin{cases} \dfrac{\partial U_{nb}^{ab}}{\partial x_a} = \dfrac{\partial U_{nb}^{ab}}{\partial r_{ab}} \dfrac{\partial r_{ab}}{\partial x_a} \\ \dfrac{\partial U_{nb}^{ab}}{\partial y_a} = \dfrac{\partial U_{nb}^{ab}}{\partial r_{ab}} \dfrac{\partial r_{ab}}{\partial y_a} \end{cases}$$

Constant T and/or P

Although mathematically more convenient, running MD on the *NVE* ensemble (where naturally the entropy S is obtained directly from the corresponding partition function) is less desirable than running in the *NVT* (to obtain the Helmholtz free energy, A) or even better in the *NPT* (to obtain the Gibbs free energy G).

Measuring the temperature



In a canonical ensemble (*NVT*) the temperature is constant.
In the microcanonical (*NVE*) it fluctuates.
T is directly linked to the kinetic energy:

$$K = \sum_{1=1}^{N} \frac{|\mathbf{p}|^2}{2m_i} = \frac{k_\mathrm{B} T}{2}(3N - N_c)$$

As due to the *theorem of the equipartition of energy* each degree of freedom contributes to K by $k_\mathrm{B} T/2$. N_c is the number of constraints in the system.

Constant T

Thus, in an unconstrained system:

$$K = \frac{3}{2} N k_{\mathrm{B}} T$$

let's assume that between two different times the velocity has increased by a factor λ like:

$$\Delta T = \frac{1}{2} \sum_{i=1}^{N} \frac{2}{3} \frac{m_i (\lambda v_i)^2}{N k_{\mathrm{B}}} - \frac{1}{2} \sum_{i=1}^{N} \frac{2}{3} \frac{m_i v_i^2}{N k_{\mathrm{B}}}$$

or

$$\Delta T = (\lambda^2 - 1) T(t)$$

which yields $\lambda = \sqrt{T_{\mathrm{new}} / T((t)}$. This involves that multiplying the velocities at each time step by the factor $\lambda = \sqrt{T_{\mathrm{target}} / T_{\mathrm{current}}}$ we ensure the velocities to remain constant through the simulation.

Constant T: Berendsen

Alternatively, the Berendsen thermostat couples the simulation to an external bath. In this case we scale the velocities in such a way that the rate of change of the temperature is proportional to the difference between the bath and the system:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_{\mathrm{bath}} - T(t))$$

and it yields the scaling factor for the velocities as $\lambda^2 = 1 + \frac{\delta t}{\tau}\left(\frac{T_{\mathrm{bath}}}{T(t)} - 1\right)$

Constant P

In this case we will want to control the volume of the system. In a way analogous to the control of the temperature, Berendsen proposed a “pressure bath”, yielding a rate of change of the pressure by:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P_{\mathrm{bath}} - P(t))$$

This yields a scaling factor for the volume:

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_P}(P(t) - P_{\mathrm{bath}})$$

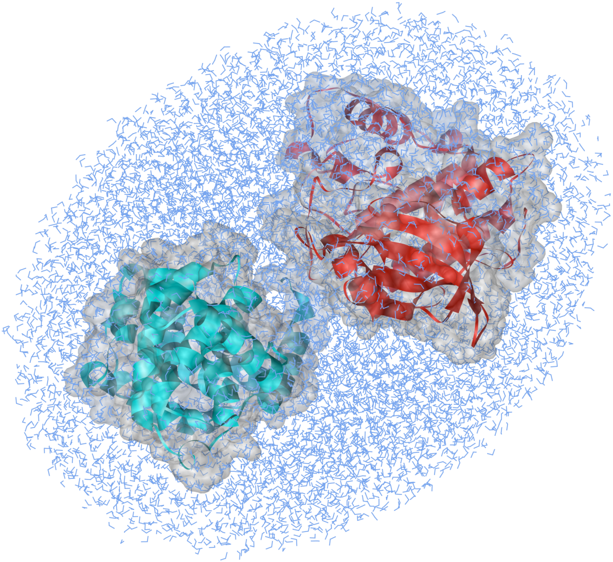
which is equivalent to scaling all atomic coordinates by $\lambda^{1/3}$:

$$\mathbf{r}_i' = \lambda^{1/3}\mathbf{r}_i$$

Microscopic simulations: additional considerations

- Need to describe a water potential: TIP3P, SPC...
- Do we include polarization?
- Convergence problems
- Boundary conditions
- Need to introduce cutoffs

We will need to restrict to a certain region



Boundary conditions

Why are they necessary?

- in 1 liter of water, we have around 3.3×10^{25} water molecules. Interactions with the walls can extend up to 10 molecules inside the fluid. If the diameter of the water molecule is approximately 2.8 , the number of water molecules close to the boundary is about 2×10^{19}. One in 1.5 milion water molecules is close to the wall!!!
- in a typical simulation we have the order of 10^3-10^4 molecules: most (if not all) will be close to the wall

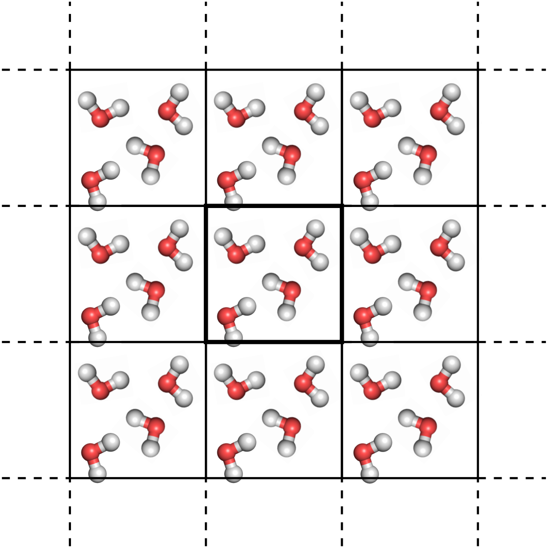
Bulk properties need to be considered properly.

Periodic boundary conditions

We may simply repeat the same cell unit adjacent to the previous one. In two dimensions a central box would be surrounded by 8 identical cells and in three dimensions this number would increase up to 26.
The coordinates of the particles in the image boxes can be computed simply by adding or substracting integral multiples of the box sides. If the replicated volume is a cube:

$$\begin{aligned} x_{\mathrm{right}} &= x_{\mathrm{central}} + 2a \\ y_{\mathrm{right}} &= y_{\mathrm{central}} + 2a \\ z_{\mathrm{right}} &= z_{\mathrm{central}} + 2a \end{aligned}$$

Periodic boundary conditions



Periodic boundary conditions

Possible geometries (see `http://mathworld.wolfram.com/Space-FillingPolyhedron.html`):

cube or general parallelepiped useful for its simplicity of implementation

hexagonal prism useful for DNA or elongated molecules simulations

truncated octahedron its spherical shape resembles well a water droplet

rhombic dodecahedron also spherical

elongated dodecahedron

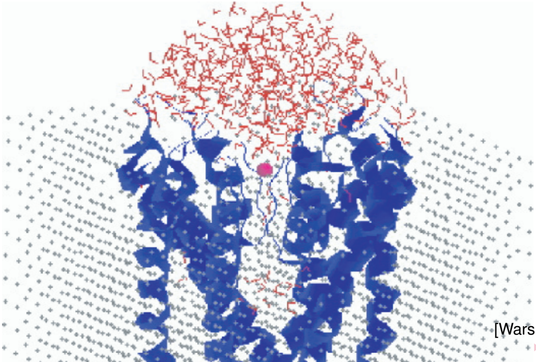
In some cases it is preferable to use non-standard periodic boundary conditions. For example, when simulating absorption on a solid surface, or when simulating an ion channel within a membrane, it is much more appropriate to use PBC in two dimensions.

Non-periodic boundary conditions

In some cases the system is better approximated by a spherical simulation. In addition the methods for PBC can introduce artifacts for long range electrostatic interactions.
In this cases, Non-PBC or *stochastic boundary conditions* are needed (SCAAS).

Non-periodic boundary conditions

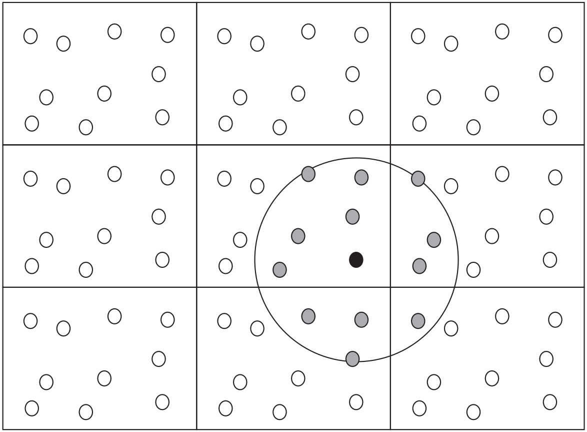
In some cases the system is better approximated by a spherical simulation. In addition the methods for PBC can introduce artifacts for long range electrostatic interactions.
In this cases, Non-PBC or *stochastic boundary conditions* are needed (SCAAS).



[Warshel et al., 2006]

Cutoffs and minimum image convention

Is it needed to include infinite number of interactions in the model? Would it be appropriate?



$g(\mathbf{r})$ I

Let us consider a spherical shell of thickness δr at a distance r from a chosen atom. The volume of such shell is

$$\begin{aligned} V &= \frac{4}{3}\pi(r+\delta r)^3 - \frac{4}{3}\pi r^3 \\ &= 4\pi r^2 \delta r + 4\pi r \delta r^2 + \frac{4}{3}\pi \delta r^3 \approx 4\pi r^2 \delta r \end{aligned}$$

Let us imagine that we are considering a perfect fluid with uncorrelated particles. If the number of particles per unit volume is ρ, then the total number in the shell is $4\pi\rho r^2\delta r$ and so the number of atoms in the volume varies as r^2.
The pair distribution function, $g(r)$, gives the probability of finding an atom (or molecule) at distance R from another atom (or molecule) compared to the ideal gas distribution.

$g(\mathbf{r})$ II

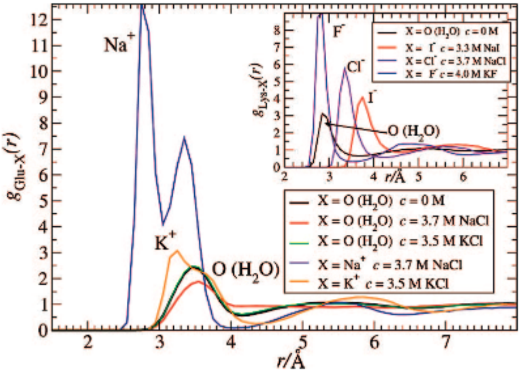
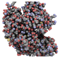


Figure 5. Radial distribution function (rdf) $g_{\mathrm{Glu-X}}(r)$ between the carboxylate carbon of the Glu side chains and X, where X = O (water oxygen), Na$^+$, or K$^+$, plotted for no salt ($c = 0$) and NaCl and KCl at concentrations of $c \simeq 3$–4 M. Note that $g_{\mathrm{Glu-O}}(r)$ at $c = 0$ (black line) and $g_{\mathrm{Glu-K^+}}(r)$ (green line) are nearly indistinguishable. (inset) rdf $g_{\mathrm{Lys-X}}(r)$ between the nitrogen of the Lys side chain and X, where X = O (water oxygen), F$^-$, Cl$^-$, or I$^-$.

Mean squared displacement

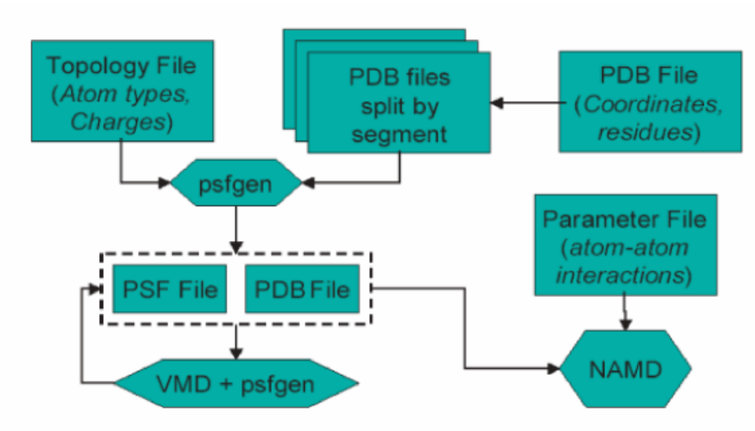


The *mean squared displacement* provides a means to stablish how the simulation is evolving:

$$\Delta r^2(t) = \frac{1}{N}\sum_{i=1}^{N}[\mathbf{r}_i(t) - \mathbf{r}_i(0)]^2$$

For a fluid without structure, the RMSD (square root of the MSD) increases with time. For a solid lattice the RMSD fluctuates close to zero with time. In this sense the $g(r)$ function can perform better to check if the structure of a fluid is statistically conserved.

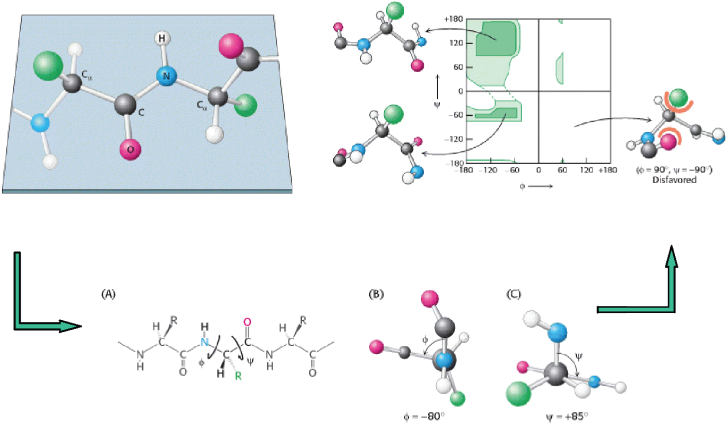
Lucky you!


Topology File
(*Atom types, Charges*)
PDB files split by segment
PDB File
(*Coordinates, residues*)
psfgen
Parameter File
(*atom-atom interactions*)
PSF File
PDB File
VMD + psfgen
NAMD

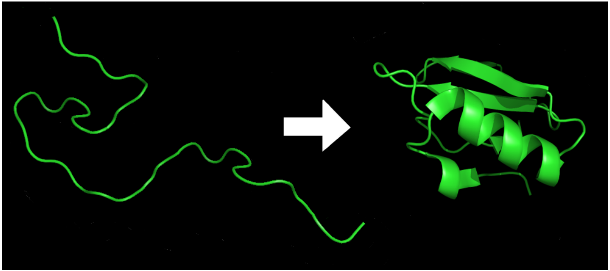
What have we learnt and what is next?

- Simulations are used to sample the configuration space.
- Statistical mechanics takes profit of this sampling to generate partition functions, leading to free energy values.
- How do we include the effects of the environment?
- How do we study differences in free energy? (processes!)
- What if we break bonds?

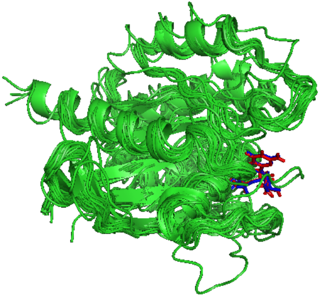
Levinthal's paradox

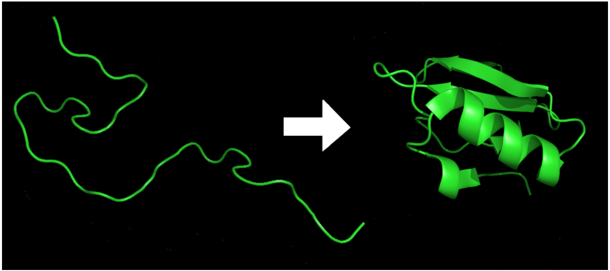


Folding

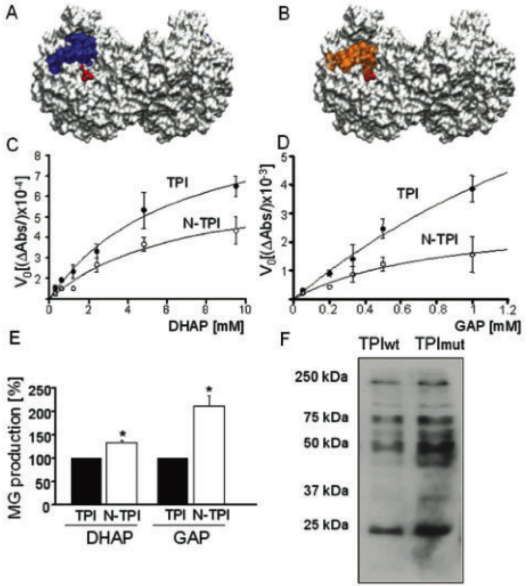


Folding and *simple* conformational changes





Subtile effects on conformational changes



[Guix et al., 2009]

Subtile effects on conformational changes

BE 2009
12307

Outline

Objectives

Basic background

Simulation

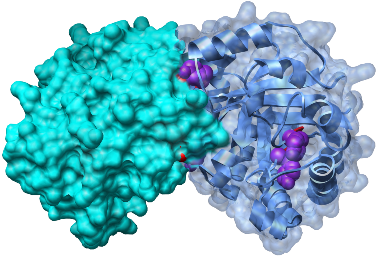
Folding

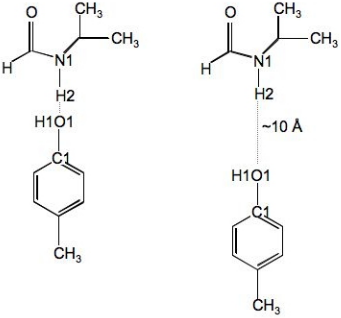
E vs dG

Mechanism

Molecular interactions

FE

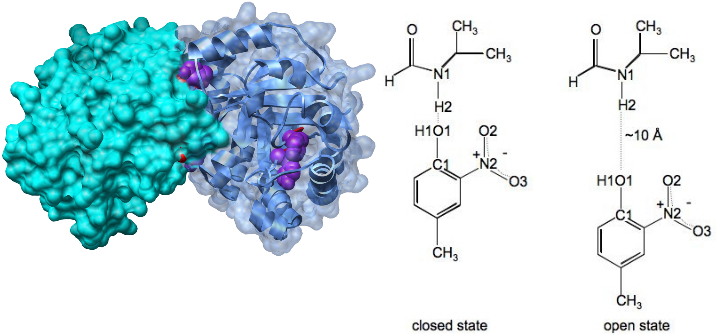




closed state

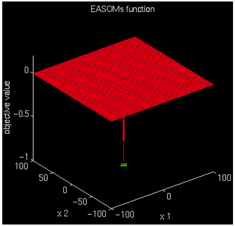
open state

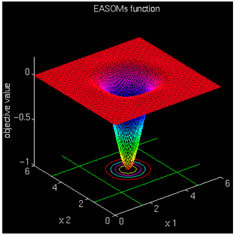
Subtile effects on conformational changes



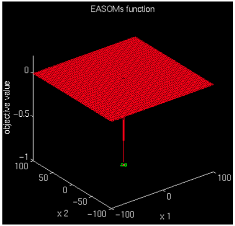
Miscione, Giupponi, Villà-Freixa, unsubmitted

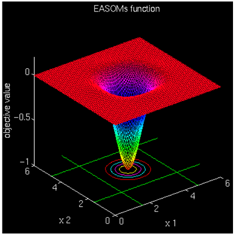
A simple abstract example





A simple abstract example

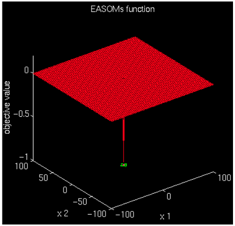


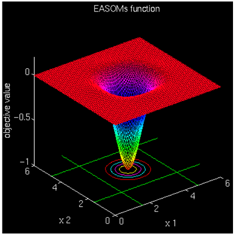


SO...

...in what case do we reach the minimum before?

A simple abstract example



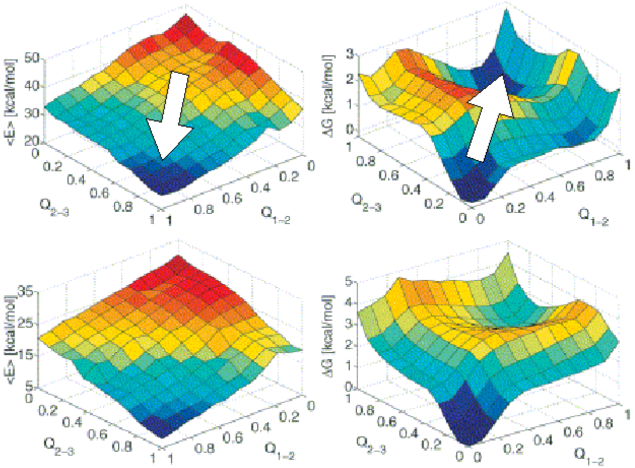


SO...

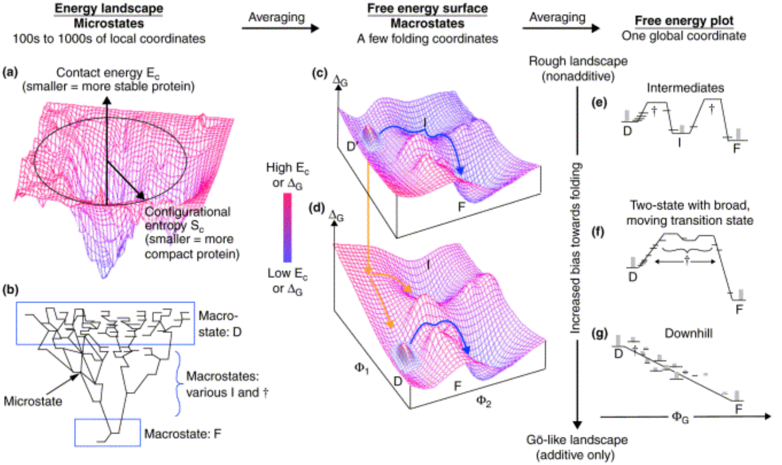
...in what case do we reach the minimum before?

$$\Delta G = \Delta H - T\Delta S$$

Energy and free energy can yield very different results → interpretations

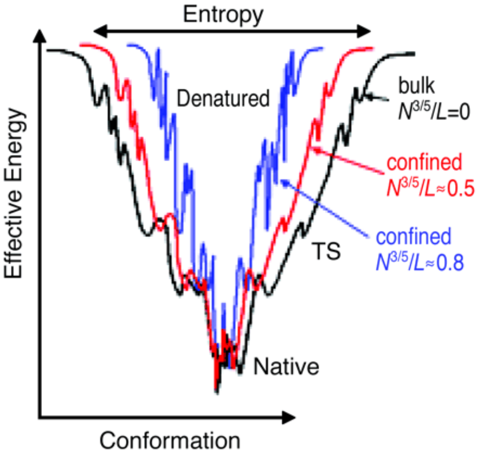


Energy and free energy can yield very different results → interpretations

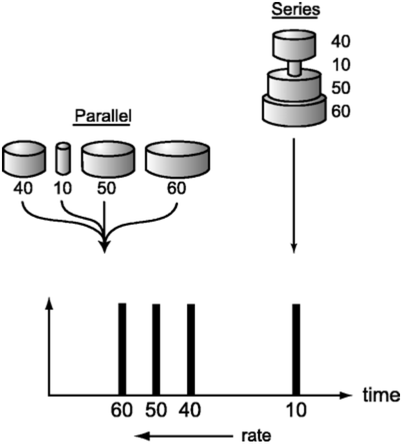


Current Opinion in Structural Biology

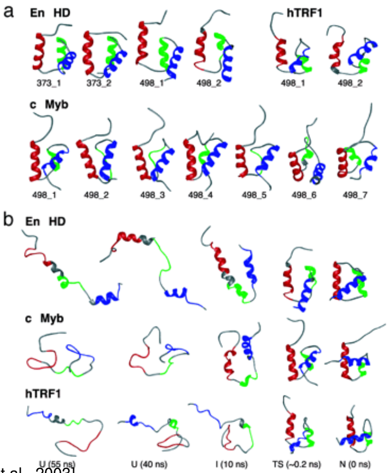
Energy and free energy can yield very different results → interpretations



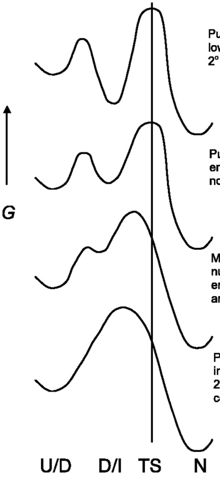
Different pathways to reach the same native structure



nucleation-condensation vs *diffusion-collision*



[Gianni et al., 2003]



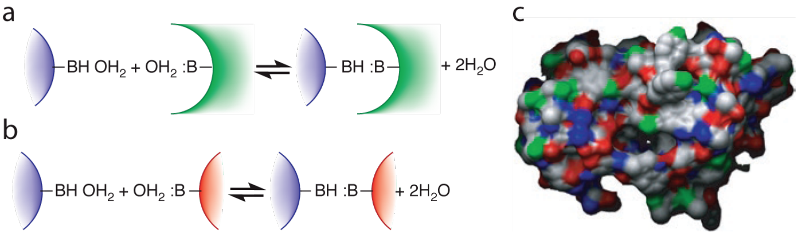
Pure diffusion-collision: low energy intermediate, 2° with no 3° interactions

Pure diffusion-collision: high energy intermediate, 2° with no 3° interactions

Mixed diffusion-collision/ nucleation-condensation: high energy intermediate with 2° and 3° interactions

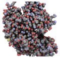
Pure nucleation-condensation: no intermediate, transition state with 2° and 3° interactions formed concurrently

Binding of cetuximab

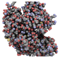


[Shoichet, 2007]

Electrostatics principles



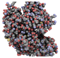
Electrostatics principles



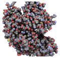
- Charge interactions obey Coulomb's law

Electrostatics principles

- Charge interactions obey Coulomb's law
- When more than two charges interact, the energies are sums of Coulombic interactions

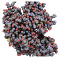


Electrostatics principles



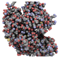
- Charge interactions obey Coulomb’s law
- When more than two charges interact, the energies are sums of Coulombic interactions
- Electric field, Gauss’s law and electrostatic potentials allow us to do such calculations

Coulomb's law I



- $$U(r) = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_1 q_2}{r} \tag{1}$$
- Clearly long ranged (r^{-1}) with respect to dispersion forces (r^{-6}) and strong
- Polarizable media (charges redistribute in response to an electric field) shield charges strongly (large ε_r)
- As ε_r depends on temperature, Eq. 1 resembles, actually, a free energy expression.
- Polarizability arises from: permament dipoles, atomic polarizabilities, hydrogen bonds

Coulomb’s law II

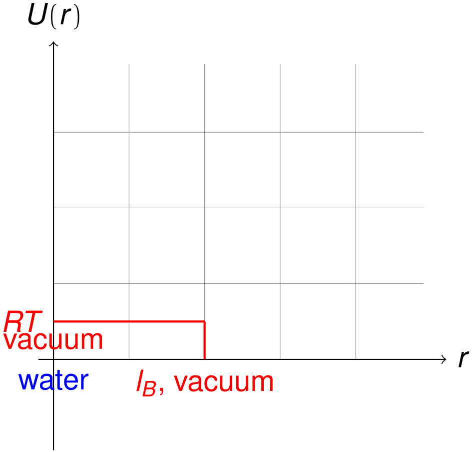


- The Bjerrum length l_B describes the charge separation at which the Coulomb energy $U(r)$ equals the thermal energy RT[?] For example, for $q_1 = q_2 = e$:

$$l_B = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{e^2N}{RT}$$

At vacuum this occurs around 560 (in water this needs to be divided by $\varepsilon_r = 80$). At bigger distances, the interactions are weaker than thermal energy RT, and particles are governed by Brownian motion.

Charges interact weaker in water



Electrostatic force and field

$$\mathbf{F} = -\nabla U = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_1 q_2}{r^2}\frac{\mathbf{r}}{r}$$

which, for a unit charge, becomes the electrostatic field if dependent on just one particle with charge q:

$$\mathbf{E}(\mathbf{r}) = \frac{\mathbf{F}(\mathbf{r})}{q_{\text{test}}} = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q}{r^2}\frac{\mathbf{r}}{r}$$

For more complex settings of charges, one may use Gauss' law, which equals the flux of the electrostatic field through any bounding surface to the sum of all charges enclosed:

$$\phi = \int_{\text{surface}} \varepsilon\mathbf{E}\cdot d\mathbf{s} = \frac{1}{\varepsilon_0}\sum_{i=1}^{n} q_i \qquad (2)$$

Electrostatic potentials

Electrostatic field describes forces (vectorial). Electrostatic potential describes energies (scalar). We start by:

$$\delta w = -\mathbf{F}d\mathbf{l} = -q\mathbf{E}d\mathbf{l}$$

which allows us to define the work done *against* the field when moving a charge q between two points as:

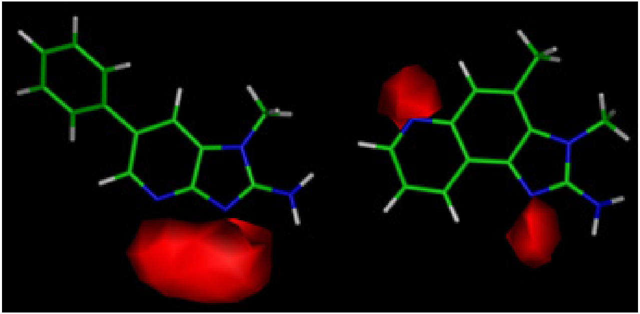
$$w_{A\to B} = -q\int_A^B \mathbf{E}d\mathbf{l}$$

and the corresponding difference in electrostatic potentials as:

$$\psi_B - \psi_A = \frac{w_{AB}}{q_{\text{test}}} = -\int_A^B \mathbf{E}d\mathbf{l}$$

(equivalent to $\mathbf{E} = -\nabla\psi$). Thus, the electrostatic potential around a point charge, $\psi_{\text{test}} = \frac{q_{\text{fixed}}}{4\pi\varepsilon_0\varepsilon_r r}$, and that produced by a charge density, $\psi_{\text{test}} = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\int_V \frac{\rho_{\text{fixed}}}{r_{12}}dV$ (r_{12}, distances between all the charges and the test charge).

Electrostatic potential surfaces



The work along equipotential curves is zero. In addition, electrostatic interactions are conservative forces.

Mixing thermodynamics with electrostatics

We shall investigate two situations:

- Moving charges from one point to another within a fixed electrostatic field (Session V on ion channels)
- Computing free energies for creating the electrostatic fields: "charging up" an assembly of originally uncharged particles:

$$\Delta G_{el} = w_{el} = \frac{1}{2}\sum_i q_i \psi_i = \frac{1}{8\pi\varepsilon_0\varepsilon_r}\sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}}$$

Ion solvation: Born Energy I

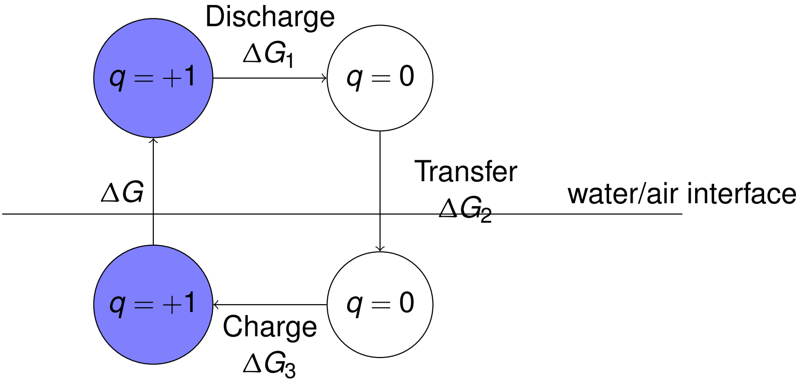
Charging up a continuous distribution resembles the above equation, but using a continuous description of the charge:

$$\Delta G_{el} = \frac{1}{2}\int_V \rho\psi_V dV$$

or, for a charged sphere or radius *a*:

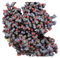
$$\Delta G_{el} = \frac{1}{2}\int_S \sigma\psi_S dS = \frac{1}{2}\left(\frac{q}{4\pi a^2}\right)\left(\frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q}{a}\right)4\pi a^2 = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q^2}{2a}$$

Ion solvation: Born Energy II



$$\Delta G_{el} = \Delta G_1 + \underbrace{\Delta G_2}_{\approx 0} + \Delta G_3 = \frac{q^2}{8\pi\varepsilon_0 a}\left(\frac{1}{\varepsilon_0} - \frac{1}{\varepsilon_w}\right)$$

Investigating processes: free energy perturbations (FEP)



Absolute value of *A* (Helmholtz free energy, the “natural” quantity in the canonical, *NVT*, ensemble) is difficult to get, but its relative value is easier:

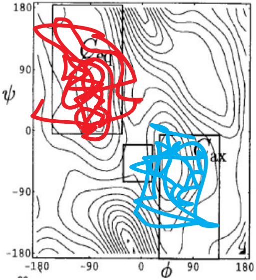
$$\Delta A = A_{\mathrm{Y}} - A_{\mathrm{X}} = -k_{\mathrm{B}} T \ln \frac{Q_{\mathrm{Y}}}{Q_{\mathrm{X}}}$$

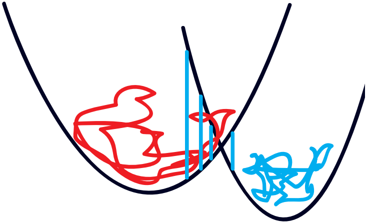
Zwanzig developed a better way to evaluate this quantity:

$$\Delta A = -k_{\mathrm{B}} T \ln < \exp[(H_{\mathrm{Y}} - H_{\mathrm{X}})/k_{\mathrm{B}} T] >_{\mathrm{X}}$$

Most times X and Y do not overlap in phase space and thus the evaluation of the above average is dificult.

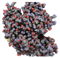
Free energy perturbations (FEP)





A is a state function, which implies we can use any path to go from one state to the other with identical result.

Free energy perturbations (FEP)



Thus, we can envisage an imaginary path that slowly leads from X to Y:

$$\begin{aligned}
\Delta A &= A_{\mathrm{Y}} - A_{\mathrm{X}} \\
&= (A_{\mathrm{Y}} - A_1) - (A_1 - A_{\mathrm{X}}) \\
&= -k_{\mathrm{B}} T \ln \left[\frac{Q_{\mathrm{Y}}}{Q_1} \frac{Q_1}{Q_{\mathrm{X}}} \right] \\
&= -k_{\mathrm{B}} T \ln < \exp[(H_1 - H_{\mathrm{X}})/k_{\mathrm{B}} T] >_{\mathrm{X}} \\
& \quad -k_{\mathrm{B}} T \ln < \exp[(H_{\mathrm{Y}} - H_1)/k_{\mathrm{B}} T] >_1
\end{aligned}$$

which can be extended to as many states as we need:

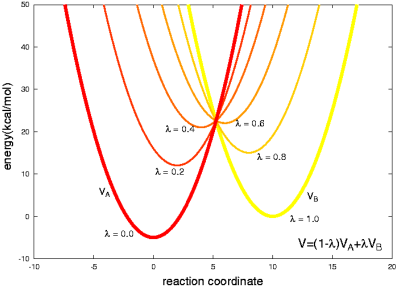
$$\Delta A = A_{\mathrm{Y}} - A_{\mathrm{X}} = -k_{\mathrm{B}} T \ln \left[\frac{Q_{\mathrm{Y}}}{Q_{\mathrm{Y}-1}} \cdots \frac{Q_3}{Q_2} \frac{Q_2}{Q_1} \frac{Q_1}{Q_{\mathrm{X}}} \right]$$

Free energy perturbations (FEP)

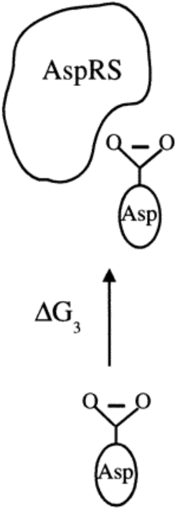
To efficiently implement a Free Energy Perturbation method, we need to move the systems from X to Y in small steps (even if they have no real physico-chemical meaning!) through a mapping potential:

$$E_m = \lambda_m E_{\mathrm{Y}} + (1 - \lambda_m) E_{\mathrm{X}}$$

where λ_m goes from 0 to 1.

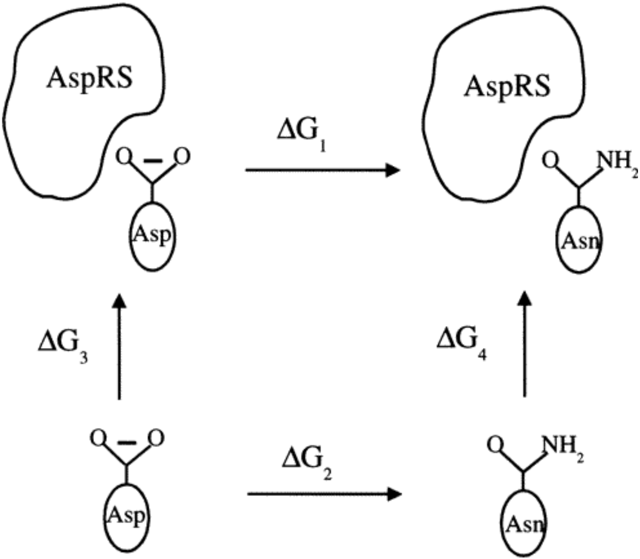


Alchemical transformations

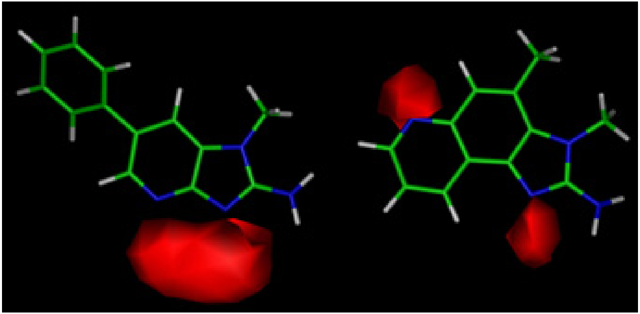


We can use FEP to do alchemical transformations

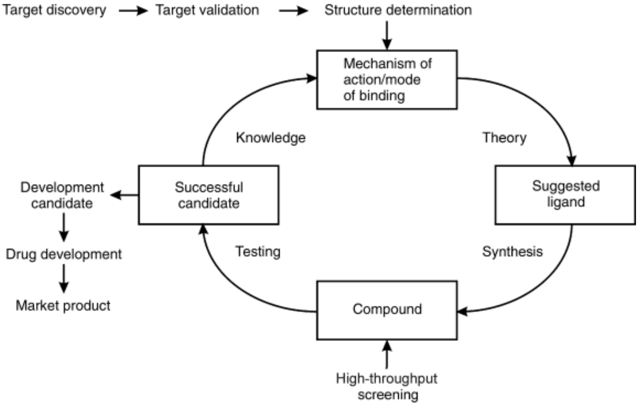
Alchemical transformations



Docking methods

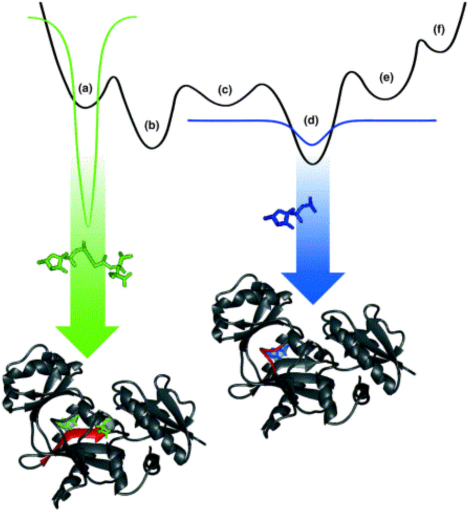


Drug discovery and drug design



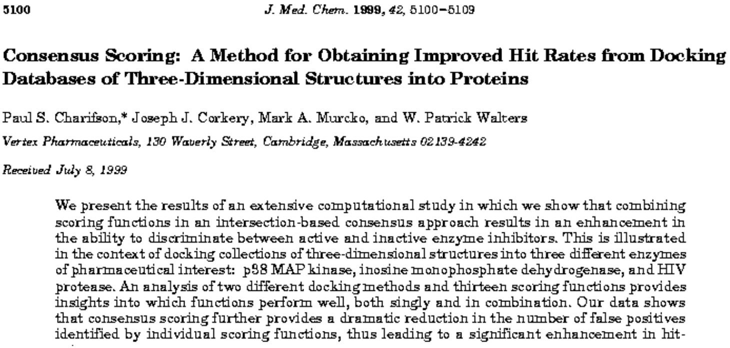
[Gohlke and Klebe, 2002]

Drug discovery and drug design



Current Opinion in Chemical Biology

Energy (docking)



5100 *J. Med. Chem.* **1999,** *42,* 5100–5109

Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins

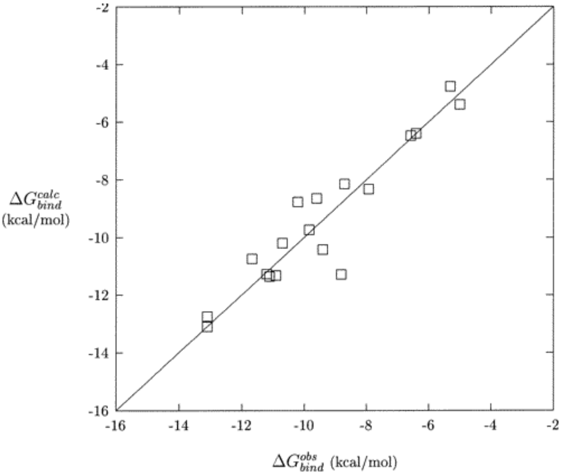
Paul S. Charifson,* Joseph J. Corkery, Mark A. Murcko, and W. Patrick Walters

Vertex Pharmaceuticals, 130 Waverly Street, Cambridge, Massachusetts 02139-4242

Received July 8, 1999

We present the results of an extensive computational study in which we show that combining scoring functions in an intersection-based consensus approach results in an enhancement in the ability to discriminate between active and inactive enzyme inhibitors. This is illustrated in the context of docking collections of three-dimensional structures into three different enzymes of pharmaceutical interest: p38 MAP kinase, inosine monophosphate dehydrogenase, and HIV protease. An analysis of two different docking methods and thirteen scoring functions provides insights into which functions perform well, both singly and in combination. Our data shows that consensus scoring further provides a dramatic reduction in the number of false positives identified by individual scoring functions, thus leading to a significant enhancement in hit-

Energy (docking) vs free energy (binding)



The LIE method [Aqvist et al., 2002]

Residue contributions to binding in Ras

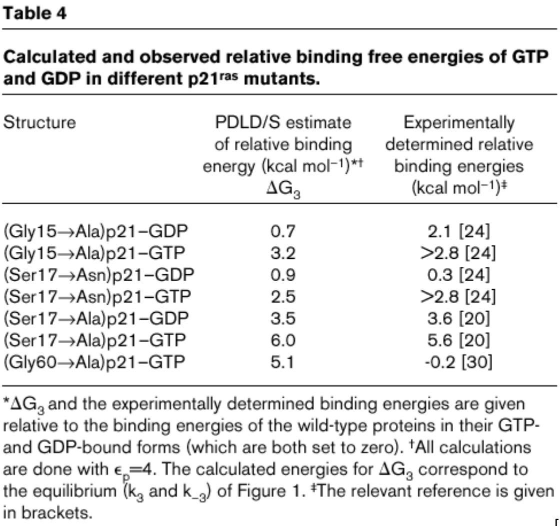


Table 4

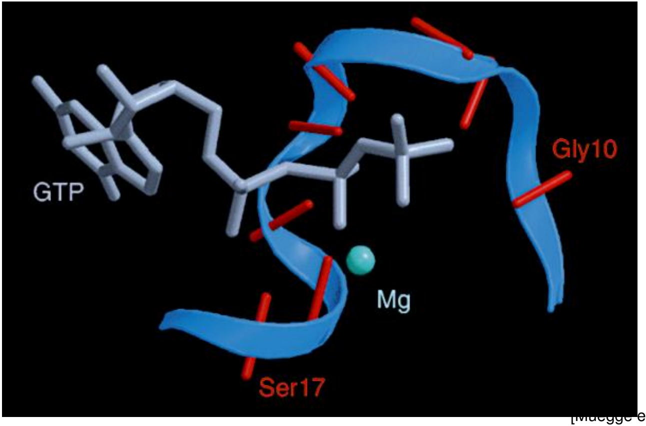
Calculated and observed relative binding free energies of GTP and GDP in different p21^{ras} mutants.

Structure	PDLD/S estimate of relative binding energy (kcal mol^{-1})*† ΔG_3	Experimentally determined relative binding energies (kcal mol^{-1})‡
(Gly15→Ala)p21–GDP	0.7	2.1 [24]
(Gly15→Ala)p21–GTP	3.2	>2.8 [24]
(Ser17→Asn)p21–GDP	0.9	0.3 [24]
(Ser17→Asn)p21–GTP	2.5	>2.8 [24]
(Ser17→Ala)p21–GDP	3.5	3.6 [20]
(Ser17→Ala)p21–GTP	6.0	5.6 [20]
(Gly60→Ala)p21–GTP	5.1	-0.2 [30]

*ΔG_3 and the experimentally determined binding energies are given relative to the binding energies of the wild-type proteins in their GTP- and GDP-bound forms (which are both set to zero). †All calculations are done with ϵ_p=4. The calculated energies for ΔG_3 correspond to the equilibrium (k_3 and k_{-3}) of Figure 1. ‡The relevant reference is given in brackets.

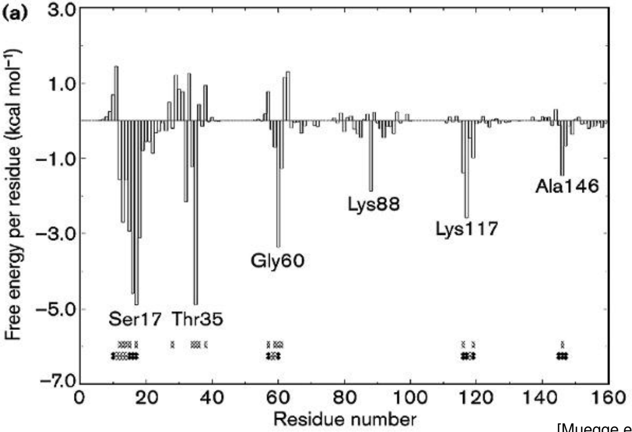
[Muegge et al., 1996]

Residue contributions to binding in Ras



[Muegge et al., 1996]

Residue contributions to binding in Ras



[Muegge et al., 1996]

Aqvist, J., Luzhkov, V., and Brandsdal, B. (2002).
Ligand binding affinities from md simulations.
Acc. Chem. Res., 35:358–365.

Carlson, H. A. (2002).
Protein flexibility and drug design: how to hit a moving target.
Curr Opin Chem Biol, 6(4):447–52.

Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. N., DeMarco, M. L., Daggett, V., and Fersht, A. R. (2003).
Unifying features in protein-folding mechanisms.
Proc Natl Acad Sci U S A, 100(23):13286–13291.

Gohlke, H. and Klebe, G. (2002).
Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors.
41:2644–2676.

Goodsell, D. (1998).
The Machinery of Life.
Springer-Verlag, New York.

Guix, F. X., Ill-Raga, G., Bravo, R., Nakaya, T., de Fabritiis, G., Coma, M., Miscione, G. P., Villà-Freixa, J., Suzuki, T., Fernàndez-Busquets, X., Valverde, M. A., de Strooper, B., and Muñoz, F. J. (2009).
Amyloid-dependent triosephosphate isomerase nitrotyrosination induces glycation and tau fibrillation.
Brain, 132(Pt 5):1335–45.

Kitchen, D., Decornez, H., Furr, J., and Bajorath, J. (2004).
Docking and scoring in virtual screening for drug discovery: Methods and applications.
Nat. Rev. Drug Disc., 3:935–949.

Levitt, M. (2001).
The birth of computational structural biology.
Nat. Struct. Biol., 8(5):392–393.

Muegge, I., Schweins, T., Langen, R., and Warshel, A. (1996).
Electrostatic Control of GTP and GDP Binding in the Oncoprotein p21 ras.
Structure, 4:475–489.

Shoichet, B. K. (2007).
No free energy lunch.
Nat Biotechnol, 25(10):1109–10.

Warshel, A., Sharma, P. K., Kato, M., and Parson, W. W. (2006).
Modeling electrostatic effects in proteins.
Biochim Biophys Acta, 1764(11):1647–1676.