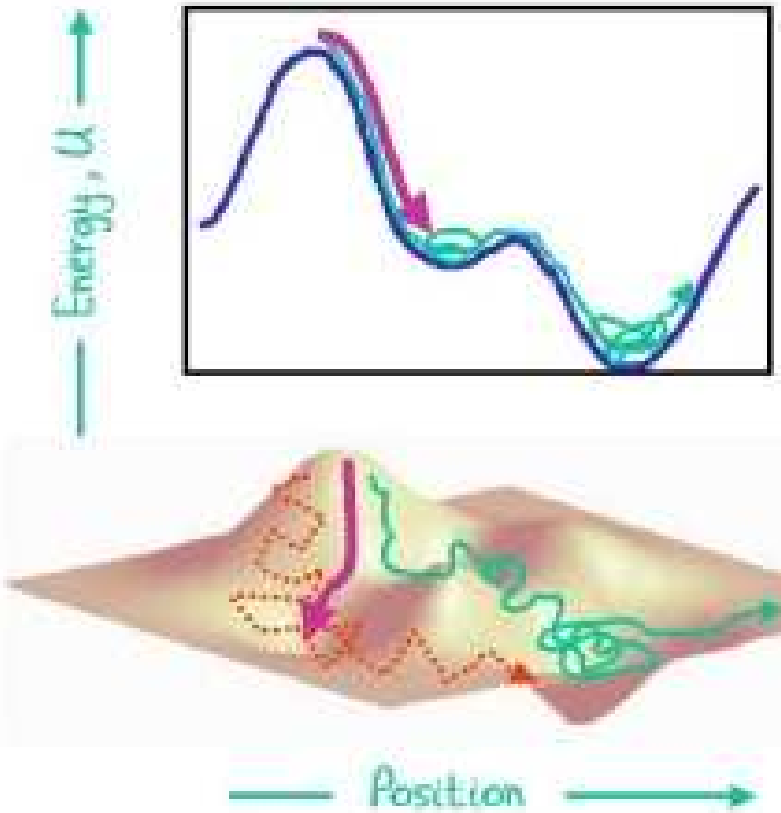
M1 Spécialité Bioinformatique Lecture 3

P. Derreumaux

Exploring Conformations

MOVING OVER THE ENERGY SURFACE

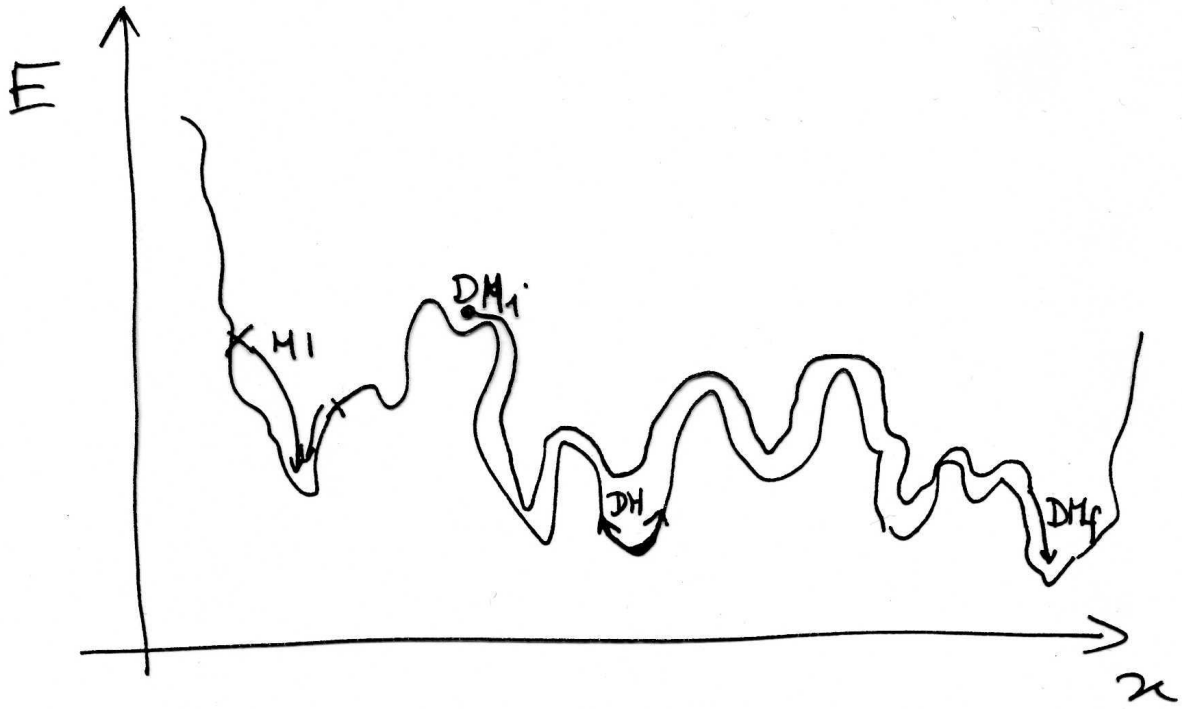


- Energy Minimization drops into local minimum.
- Molecular Dynamics uses thermal energy to move smoothly over surface.
- Monte Carlo Moves are random. Accept with probability $\exp(-\Delta U/kT)$.

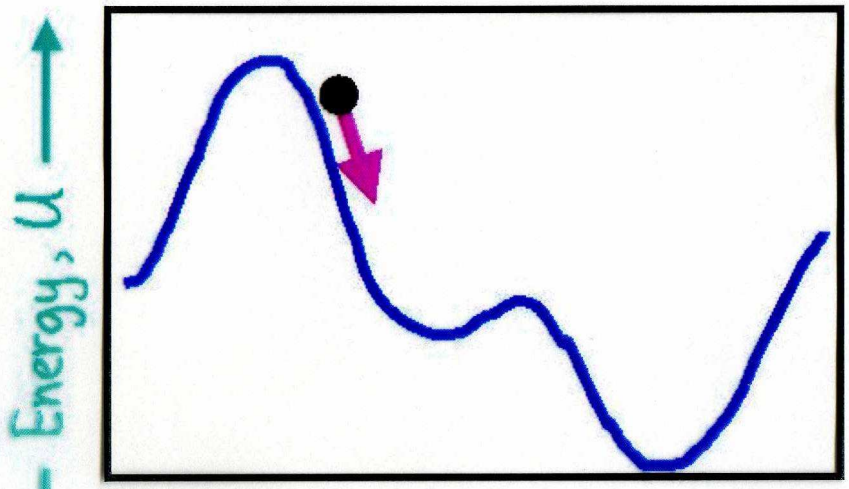
Minimization

Harmonic Dynamic

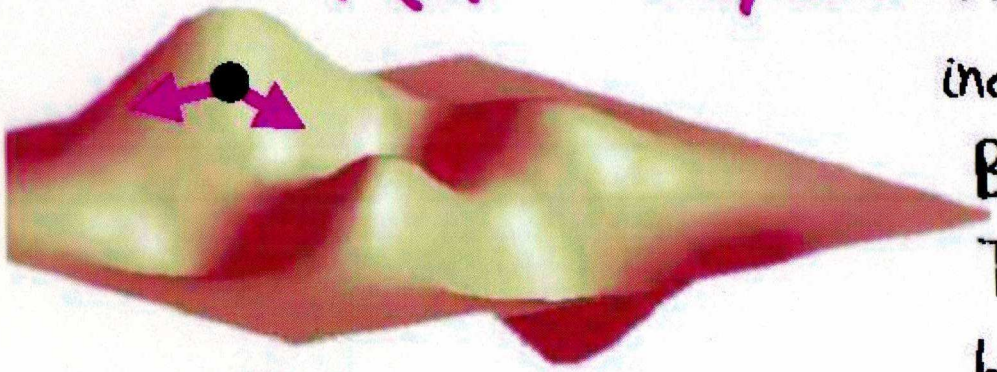
Molecular Dynamic



TOTAL POTENTIAL ENERGY . 2



$$F(x) = -dU/dx$$

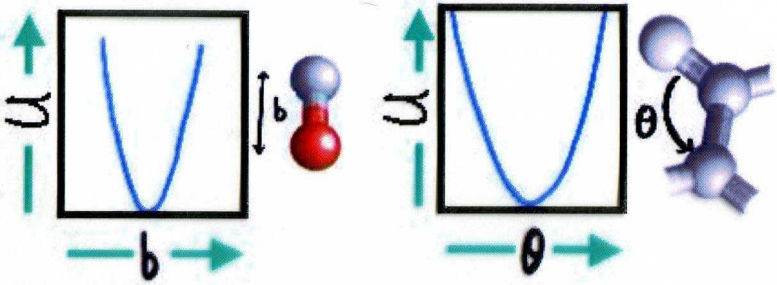


- The total potential energy or enthalpy fully defines the system, U.
- The forces are the gradients of the energy.
- The energy is a sum of independent terms for:
  Bonds, Bond angles, Torsion angles and non-bonded atom pairs.

TOTAL POTENTIAL ENERGY

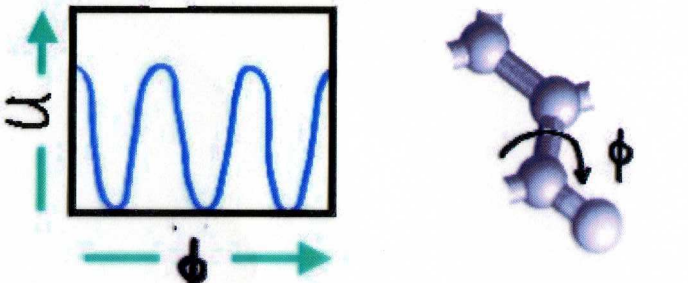
$$U = \sum \tfrac{1}{2} K_b (b - b_0)^2 + \sum \tfrac{1}{2} K_\theta (\theta - \theta_0)^2$$

All Bonds — All Angles



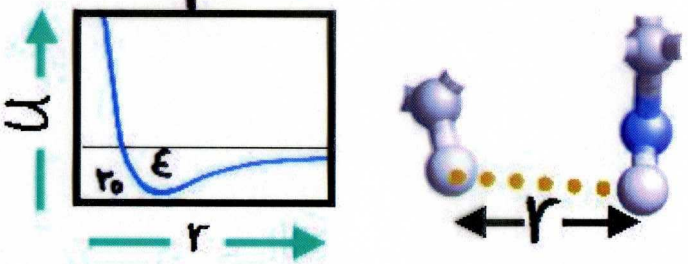
$$+ \sum K_\phi [1 - \cos(n\phi + \delta)]$$

All Torsion Angles



$$+ \sum \varepsilon \left[\left(r_0/r\right)^{12} - 2\left(r_0/r\right)^{6} \right]$$

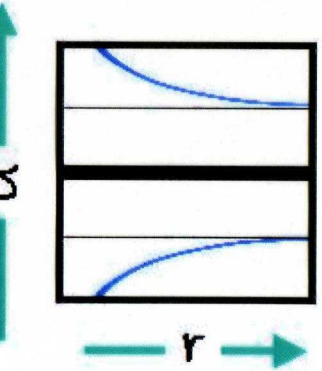
All nonbonded pairs

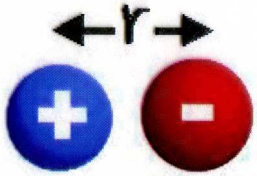


$$+ \sum 332 q_i q_j / r$$

All partial charges

ENCAD. Parameters from 1979 (Lifson)

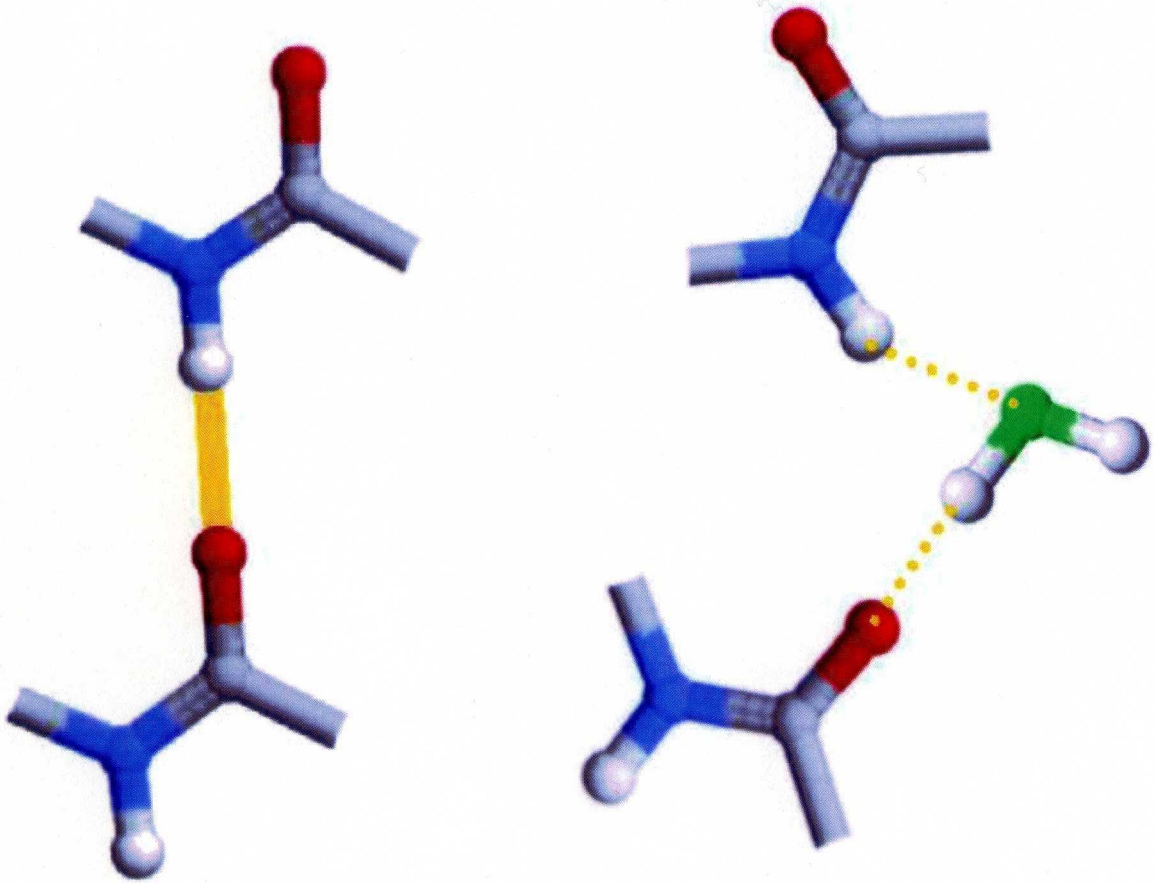


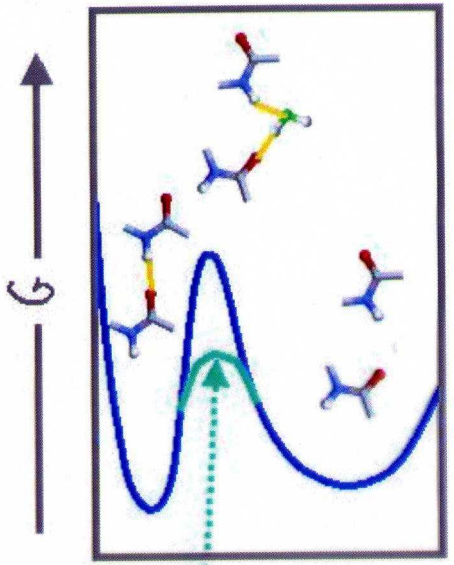


WATER ALLOWS HYDROGEN BONDS TO BREAK

Intact hydrogen bond in helix

Hydrogen bond is breaking





Free Energy barrier between states is much lower in water

- Water catalyzes the breakage of hydrogen bonds by stabilizing the transition state.

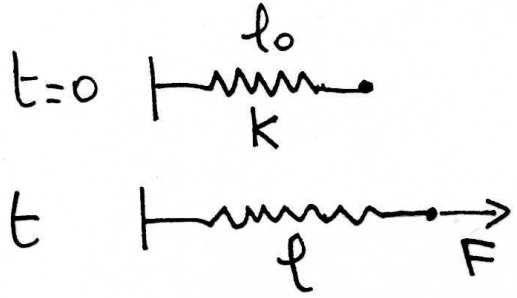
Various Energy models

- function of the level of chain representation
- implicit or explicit solvent effects
  - $\varepsilon = f(r)$
    - Poisson-Boltzmann
    - Born model
    - Gaussian model (EEF1)
  - SPC, TIP3P water models

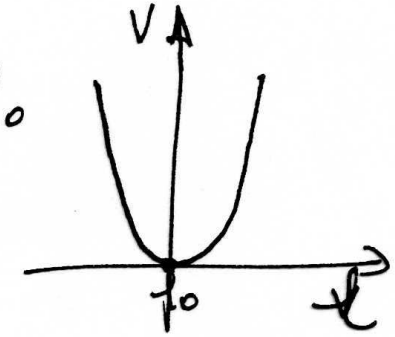
- Essentially two classes of energy models.
  - Physical Force Fields (based on structures, thermodynamic and vibrational data) → CHARMM, AMBER, GROMOS, ECEPP, ENCAD
  - Knowledge-based Force fields (based on structures, $\frac{\langle E_{native} \rangle - \langle E_{unfolded} \rangle}{\langle E^2_{unfolded} \rangle^{\frac{1}{2}}}$

    e.g Miyazawa-Jernigan, Shakhnovich, Skolnick, OPEP.

One-dimensional case



Minimization $V = \frac{1}{2} k x^2$ $x = l - l_0$



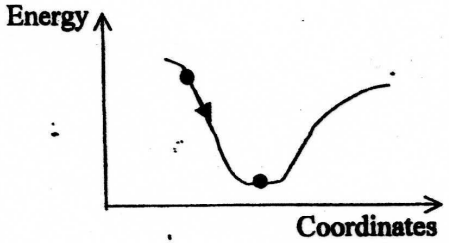
$$-\frac{dV}{dx} = -kx = 0 \rightarrow l = l_0$$

For N-dimensional case

Steepest Descent
Conjugate Gradient
Newton Raphson

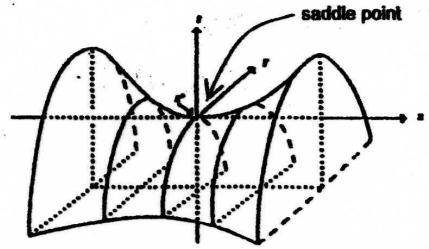
ENERGY MINIMIZATION

Find a minimum energy configuration



$$\frac{\partial V}{\partial x_i} = 0 \qquad \forall i$$

$$\frac{\partial^2 V}{\partial x_i^2} > 0 \qquad \forall i$$



$$\frac{\partial^2 V}{\partial x_i^2} > 0$$

$$\frac{\partial^2 V}{\partial x_i^2} < 0$$

Different methods

Direct methods	**simplex** **random search** **grid**	**not efficient**
First derivatives of potentials	**steepest descent** **conjugate gradient**	**moderately efficient** **efficient**
Second derviatives of potentials	**Newton-Raphson** **ÀBNR**	**efficient only near the minimum**

N - dimensional case (N particles, 3N coordinates)

- Minimization.

ordre 1 $\quad V(x+dx) = V(x) + \frac{\partial V}{\partial x}\,dx + \frac{1}{2}\frac{\partial^2 V}{\partial x^2}\,dx^2$

1 coordinate

$$\Rightarrow \frac{\partial V}{\partial x} = \frac{V(x+dx) - V(x)}{dx} = 0$$

ordre 1

N coordinates $\quad X_i, Y_i, Z_i$ with $i = 1$ to N

$$\vec{g} = -\overrightarrow{\text{grad}}\, V$$

$$\begin{pmatrix} \frac{\partial V}{\partial x_1} \\ \frac{\partial V}{\partial y_1} \\ \frac{\partial V}{\partial z_1} \\ \vdots \\ \frac{\partial V}{\partial z_n} \end{pmatrix}$$

Steepest descent : SD

k = iteration $\quad S_k = \frac{g_k}{|g_k|}$ $\quad g_k = -\text{grad}\, V_k$.

(direction de descente)

$$X_k = X_{k-1} + \lambda S_k$$

(λ : pas d'integration)

Rmq
- convergence lente
- elimination des mauvais contacts dans une structure.

Conjugate Gradient : CG.

$$s_k = -g_k + b_k s_{k-1} \qquad b_k = \frac{|g_k|^2}{|g_{k-1}|^2}$$

$g_k \equiv$ conjugate s_{k-1}

Rmq. more useful than SD

- order 2 } Newton - Raphson
  Truncated Newton - Raphson

  $\frac{\partial^2 V}{dx^2}$ required

- End of Minimization : — number max of steps.
  — $\|\vec{g}\| \leq 10^{-2} - 10^{-3}$

Schémas de Minimisation

Les différentes méthodes ont des avantages et des désavantages. Pour optimiser son utilisation il vaut mieux de combiner différents schémas de minimisation.

Il y naturellement différents schémas alternatifs. Mais, de façon générale, une bonne approche commencera par quelques pas de Steepest-Descent, continuera avec la méthode des Gradients-Conjugués et, dans les cas ou une plus grande qualité des résultats est recherchée, il est désirable de finir avec une méthode comme l'ABNR.

Convergence: Une question importante est de savoir quand on met fin à la minimisation, i. e., quel est le critère de convergence. La méthode la plus commune est celle du Gradient de la Racine de Minimes Carrés (Gradient Root-Mean-Square - GRMS), qui est définie comme la racine des minimes carrés de 3N gradients. Une bonne convergence est dans l'ordre de < 0.001. En certains cas une meilleure convergence pourra être exigée.

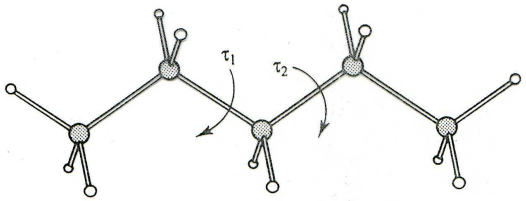
Applications de la Minimisation

Applications de la minimisation d'énergie en systèmes macromoléculaires pourront être:

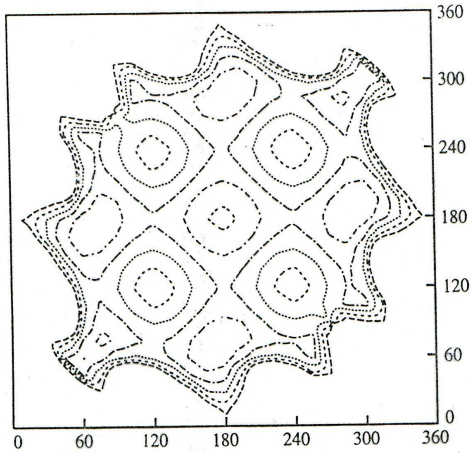
- Enlever les tensions dans des structures obtenues expérimentalement
- Refinement de modèles moléculaires
- Utilisation comme composante dans des méthodes de recherche du minime global d'un système.
- Analyse d'un système, par example dans des cartes adiabatiques (ex: Plot de Ramachandran).

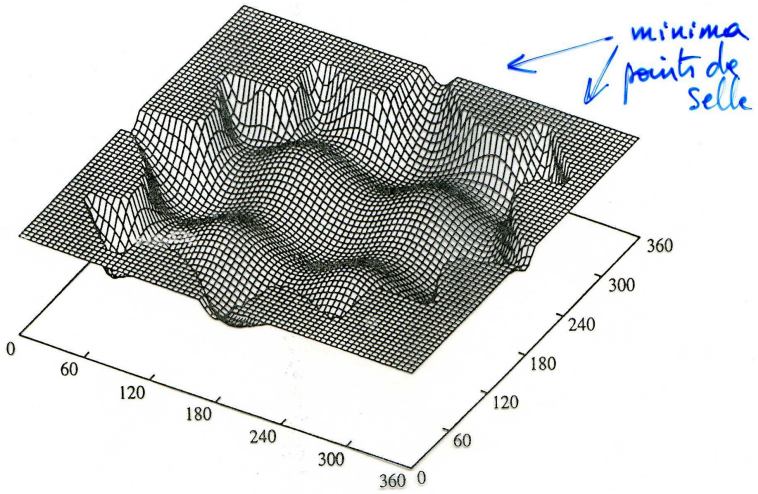
Grid Search

- An obvious solution to the multiple minimum problem is to generate a large number of starting conformations and map out the shape of the Potential Energy Surface (PES)
- Simplest and oldest approach is a grid search
  - Identify the rotable bonds of interest
  - Select the starting angles and increments
  - Loop through all the combinations holding the selected angles fixed and relaxing everything else
  - Done when all combinations are tried



Variation in the energy of pentane with the two torsion angles indicated and represented as a contour diagram and isometric plot. (Continued overleaf.)





Grid Search: (High) Number of Conformers to Consider

$$\#conf = \prod_{i=1}^{N} \frac{360}{\theta i}$$

Main problem is a combinatorial explosion

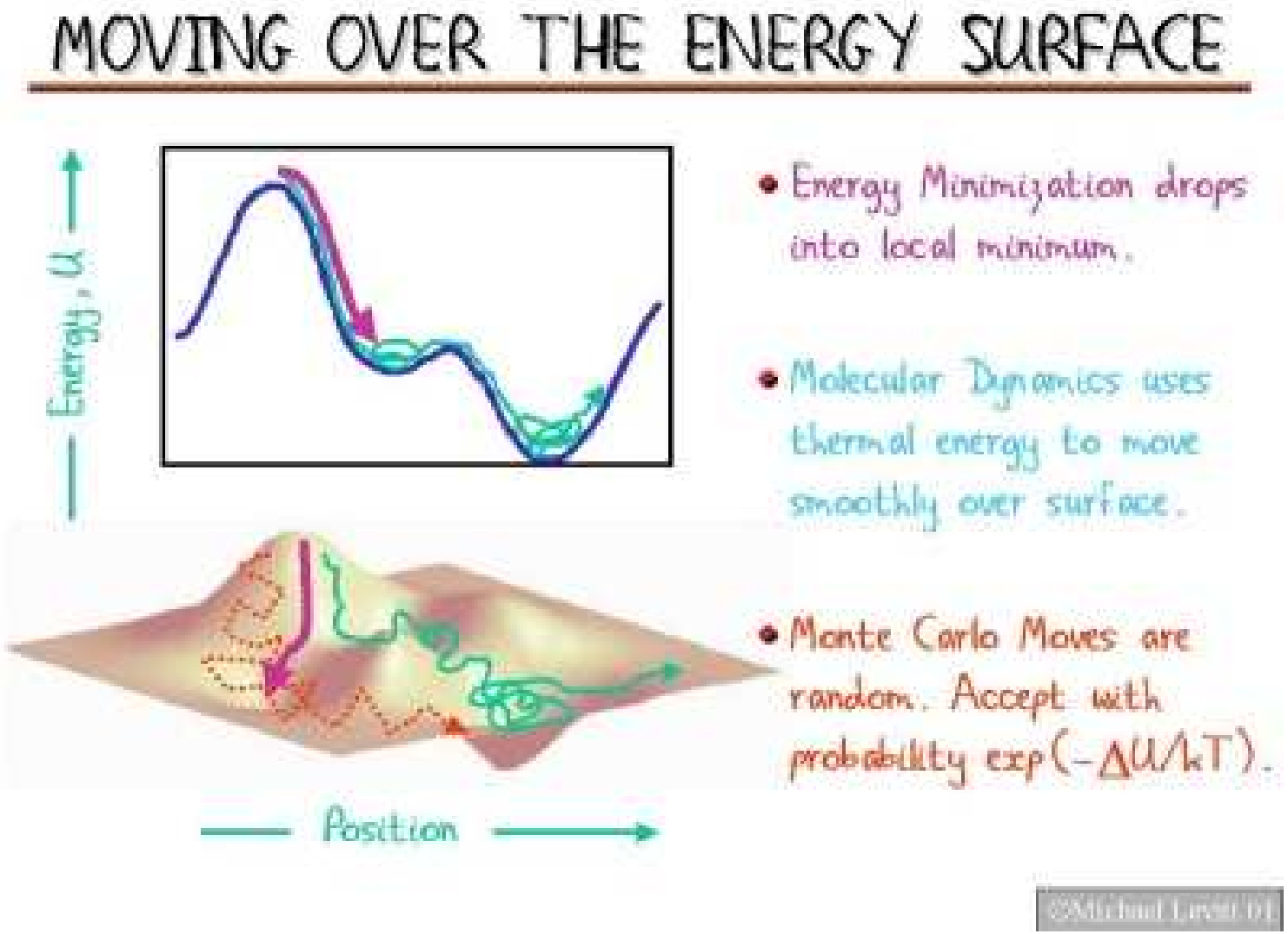
For 5 bonds, 30 degree increment: 248,832 structures

solution: pick preferred conformations (e.g. Pro, $\psi \simeq -60°$)

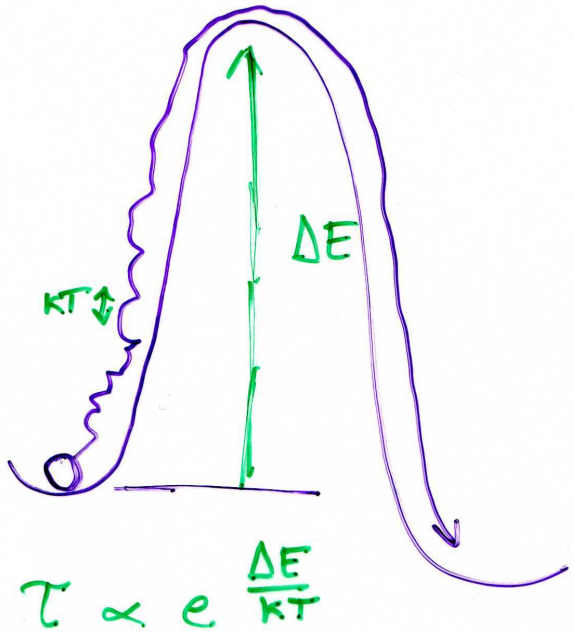
Second problem: don't get minima, get fixed points on surface

Basically laid a grid on the Potential Energy Surface (PES)

Method is slow, inefficient and limited for large models



MD and other accelerated methods



$\tau \propto e^{\frac{\Delta E}{KT}}$

méthodes

- DM (P. Kollman, 1998)
- Ensemble dynamics (Pande, 2001)
- replica exchange (Berne, 2000)
- smooth energy surfaces (Parinello, 2003)
- locating saddle points (ART, 2001)

ART : Activation Relaxation Technique

Molecular Dynamics Simulation

Molecule: (classical) N-particle system

Newtonian equations of motion:

$$m_i \frac{d^2}{dt^2} \vec{r}_i = \vec{F}_i(\bar{r})$$

$$\vec{F}_i(\bar{r}) = -\nabla_i V(\bar{r})$$

with

$$\bar{r} = (\vec{r}_1, \ldots, \vec{r}_N)$$

Integrate numerically via the „leapfrog“ scheme

$$\boldsymbol{v}(t + \frac{\Delta t}{2}) = \boldsymbol{v}(t - \frac{\Delta t}{2}) + \frac{\boldsymbol{F}(t)}{m}\Delta t$$

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t + \frac{\Delta t}{2})\Delta t$$

with

$\Delta t \approx 1fs!$

(equivalent to the Verlet algorithm)

MOLECULAR DYNAMICS STEPS

$\Delta t = \tau =$ time step.

Verlet :

$$r_i(t+\tau) = r_i(t) + \tau\, dr_i(t)/dt + \tau^2/2\, d^2r_i(t)/dt^2 + \tau^3/6\, d^3r_i(t)/dt^3 \quad (1)$$

$$r_i(t-\tau) = r_i(t) - \tau\, dr_i(t)/dt + \tau^2/2\, d^2r_i(t)/dt^2 - \tau^3/6\, d^3r_i(t)/dt^3 \quad (2)$$

$$r_i(t+\tau) = 2r_i(t) - r_i(t-\tau) + \tau^2\, d^2r_i(t)/dt^2 + O(\tau^4) \quad (1)+(2)$$

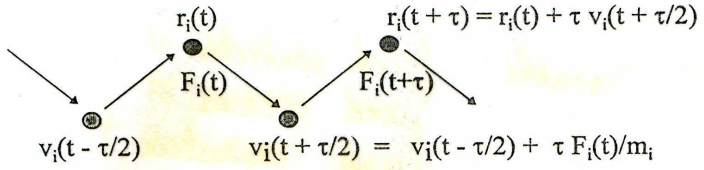
$$\ldots \quad v_i(t) = [r_i(t+\tau) - r_i(t-\tau)] / 2\delta t \quad (1)-(2)$$

Leap-frog :

$$v_i = [\, r_i(t+\tau) - r_i(t)] / \tau \approx v_i(t+\tau/2)$$

$$a_i = [\, v_i(t+\tau/2) - v_i(t-\tau/2)] / \tau \approx a_i(t)$$

$$r_i(t+\delta t) = r_i(t) - v_i(t+\tfrac{1}{2}\delta t)\, \delta t$$



Velocity Verlet :

$$r_i(t+\tau) = r_i(t) + \tau\, v_i(t) + \tau^2/2\, a_i(t)$$

$$v_i(t+\tau) = v_i(t) + \tau/2\, [\, a_i(t) + a_i(t+\tau)]$$

Others Beeman, Gear (higher orders)

INITIAL CONDITIONS

Positions - Crystallographic coordinates
- Model conformations

Velocities

The initial distribution of velocities are usually determined from a random distribution with the magnitudes conforming to the required temperature and corrected so there is no overall momentum, i.e.,

$$P = \sum_{i=1}^{N} m_i v_i = 0$$

The velocities, *vi,* are often chosen randomly from a Maxwell-Boltzmann or Gaussian distribution at a given temperature, which gives the probability that an atom i has a velocity v_x in the x direction at a temperature T.

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right]$$

The temperature can be calculated from the velocities using the relation

$$T = \frac{1}{(3N)^{*}} \sum_{i=1}^{N} \frac{|p_i|}{2m_i}$$

or 3N-6

with N: number of atoms

kinetic energy per atom $\frac{1}{2} m \langle v_i \rangle^2 = \frac{3}{2} k_B T$

Les vitesses initiales :

$$\frac{1}{2}N_{df}kT = E_{kin}$$

$$E_{kin} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2$$

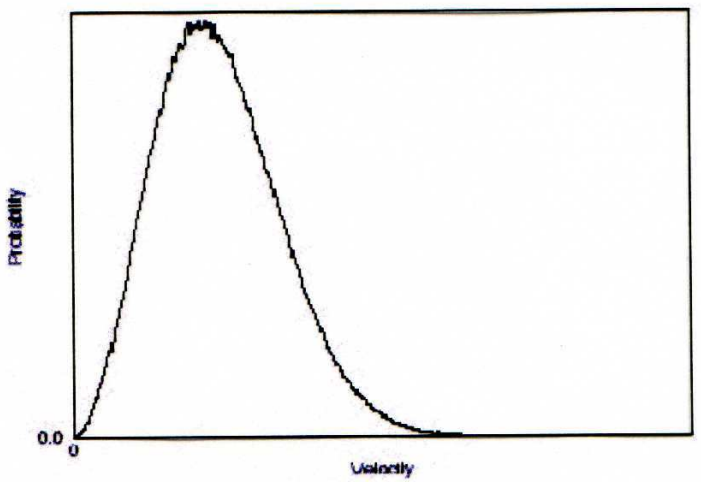


Figure 3.4: A Maxwellian distribution, generated from random numbers.

$$p(v_i) = \sqrt{\frac{m_i}{2\pi kT}}\exp(-\frac{m_i v_i^2}{2kT})$$

0.Visualisation

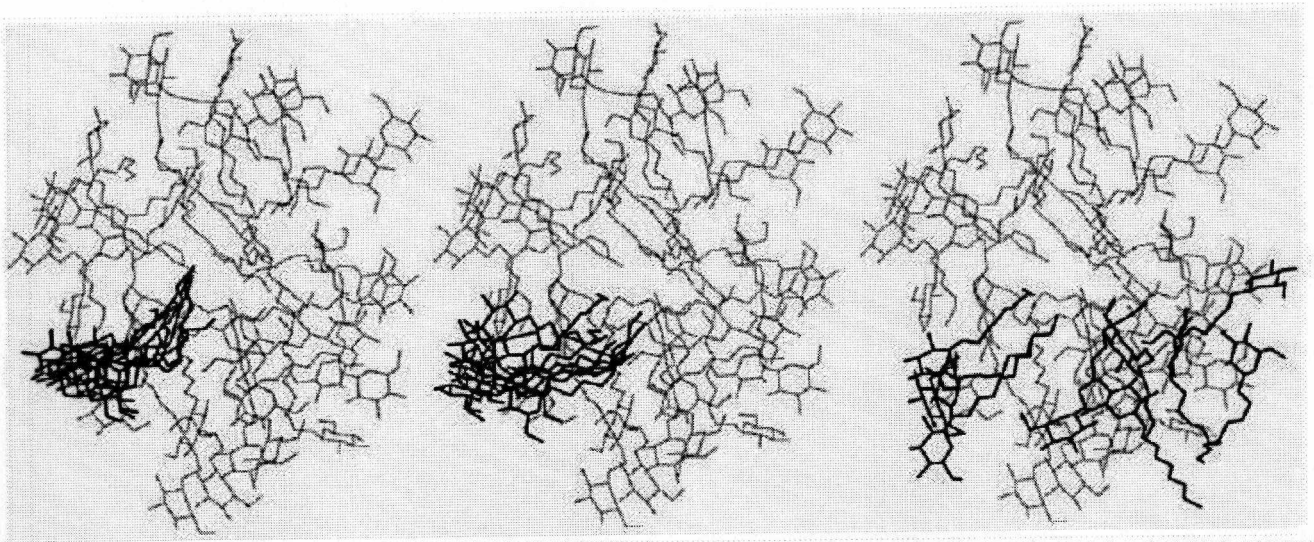


Figure 4. Trajectories of a single lipid from the 34mer simulation for 40 ps at 4 ps intervals (left), 400 ps at 40 ps intervals (center), and 4 ns at 400 ps intervals (right). Each trajectory starts at the beginning of the production run, and the center of mass motion of the micelle was removed from subsequent frames. Initial coordinates for the rest of the micelle are shown in gray to provide a sense of scale. The lipid was chosen for its mostly planar motion over the 4 ns; other lipids moved approximately the same distance but with more complex motion. In this and the remaining figures, water molecules are omitted for clarity.

1.Séries temporelles

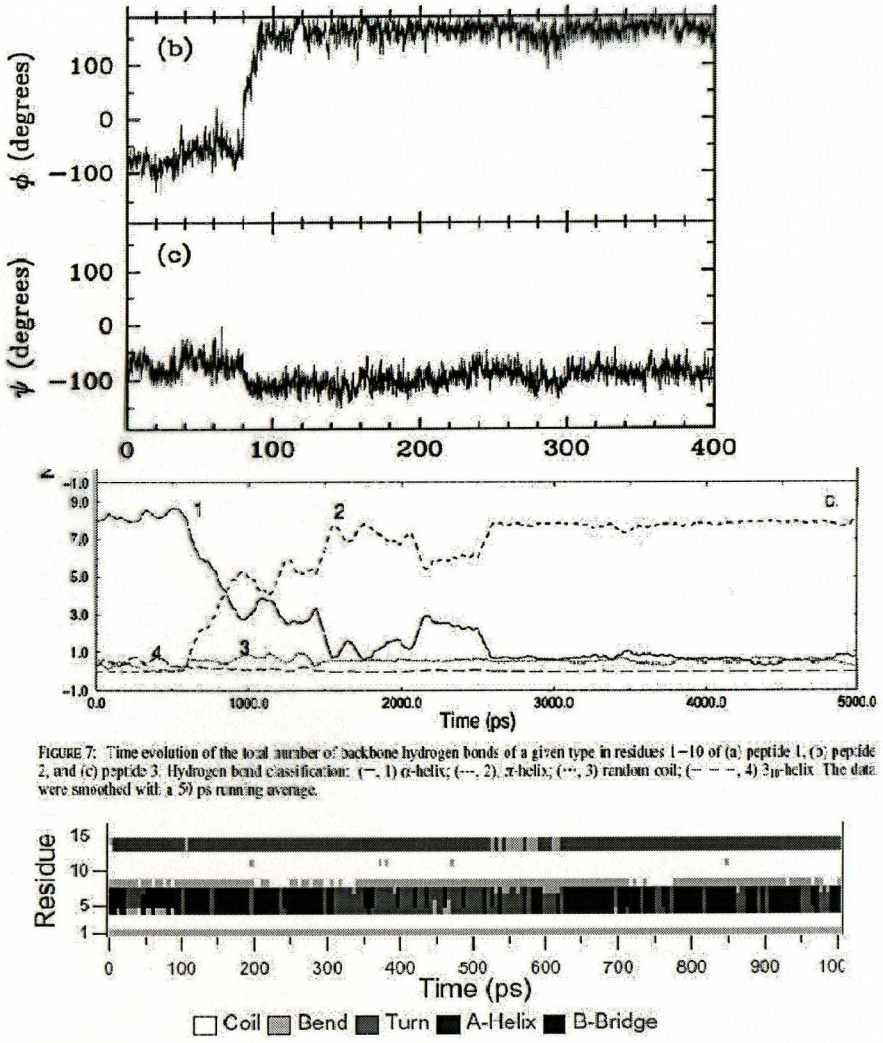


FIGURE 7: Time evolution of the total number of backbone hydrogen bonds of a given type in residues 1–10 of (a) peptide 1, (b) peptide 2, and (c) peptide 3. Hydrogen bond classification: (—, 1) α-helix; (---, 2) π-helix; (···, 3) random coil; (— — —, 4) 3_{10}-helix. The data were smoothed with a 50 ps running average.

Figure 8.10: Analysis of the secondary structure elements of a peptide in time.

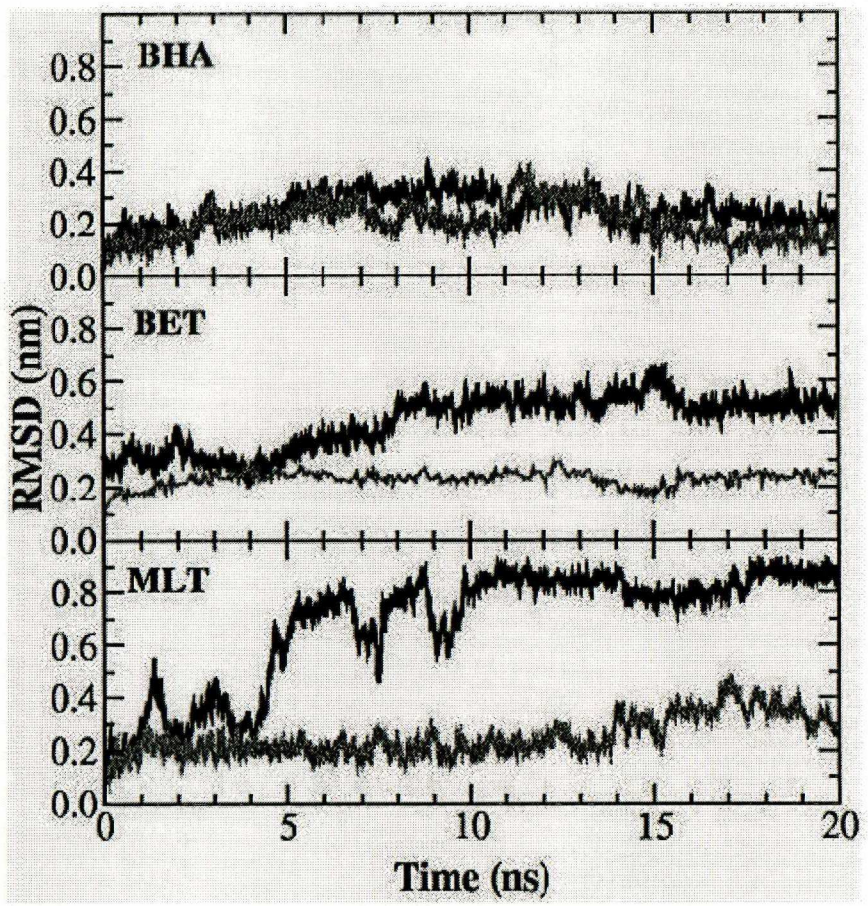
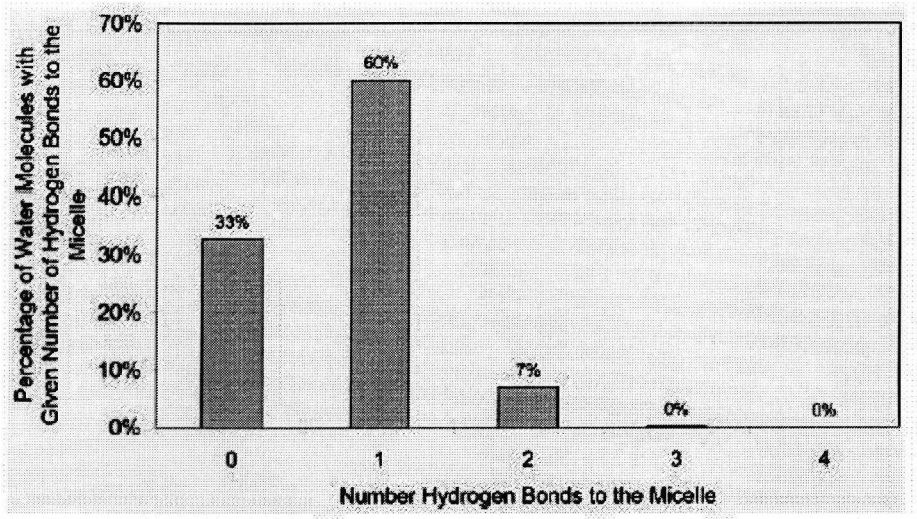
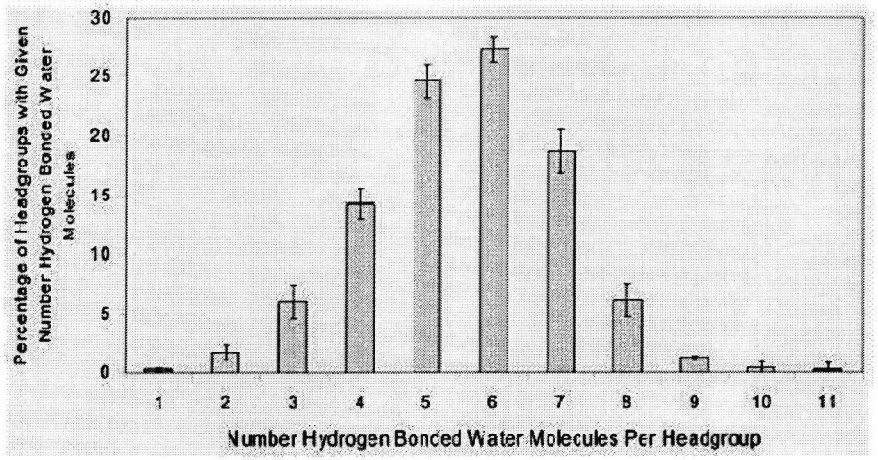


Fig. 1. Backbone RMSD with respect to the minimized initial structure of MLT, BET, and BHA simulations. Black curve, water simulation; gray curve, TFE/water simulation.

2.Statistiques





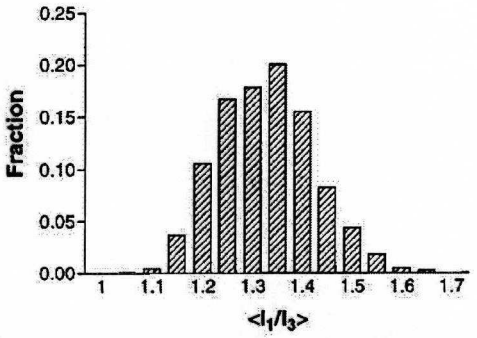
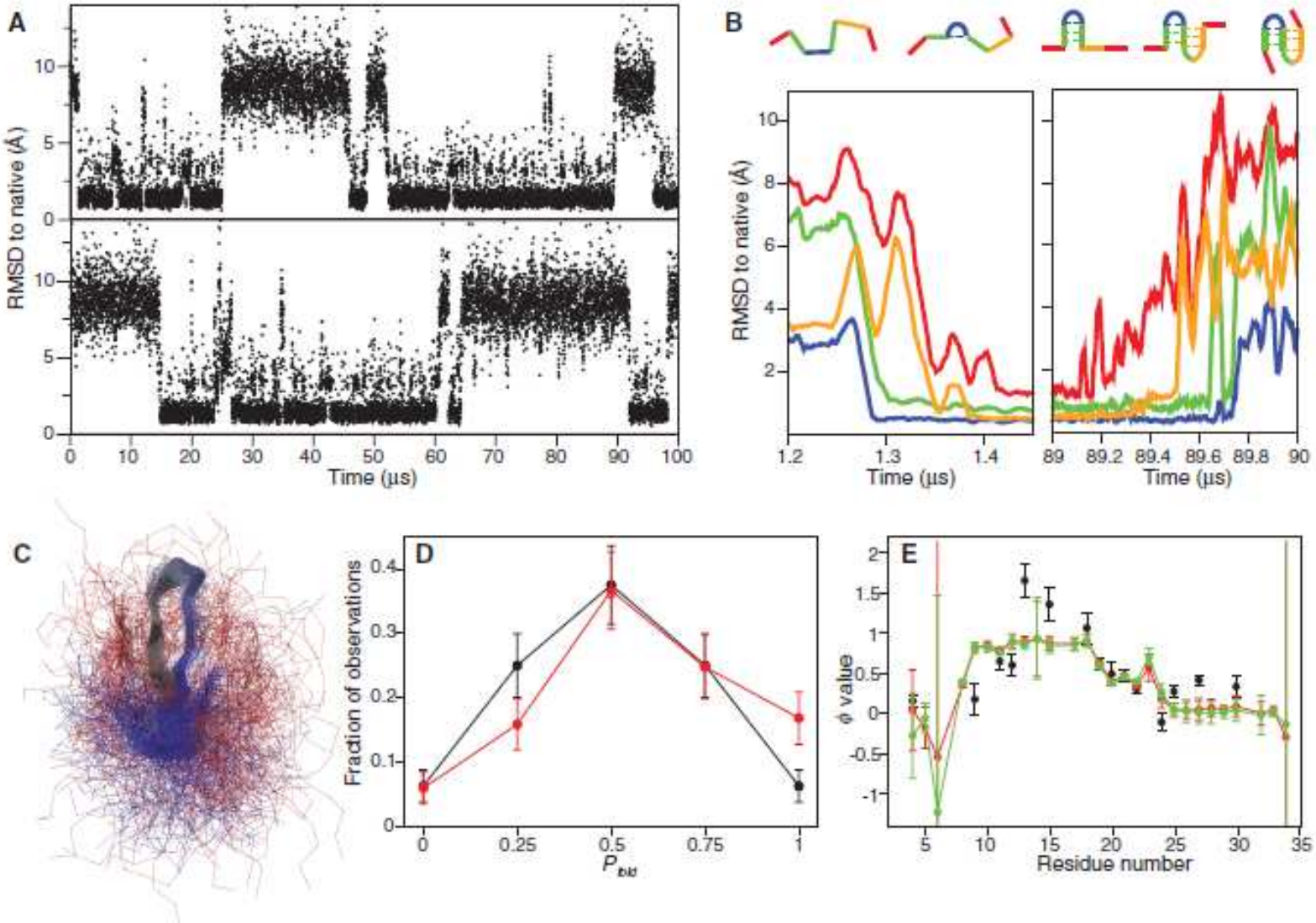


Figure 7. Histogram of the instantaneous ratios (calculated each ps) between the moments of inertia along the minor and major axes for the 34mer simulation.


A
RMSD to native (Å)
Time (μs)
B
RMSD to native (Å)
Time (μs)
Time (μs)
C
D
Fraction of observations
P_{fold}
E
ϕ value
Residue number

Langevin Dynamics (LD) Simulation

The Langevin equation is a *stochastic* differential equation in which two force terms have been added to Newton's second law to approximate the effects of neglected degrees of freedom. One term represents a frictional force, the other a *random* force $\vec{R}$. For example, the effects of solvent molecules not explicitly present in the system being simulated would be approximated in terms of a frictional drag on the solute as well as random kicks associated with the thermal motions of the solvent molecules. Since friction opposes motion, the first additional force is proportional to the particle's velocity and oppositely directed. Langevin's equation for the motion of atom i is:

$$\vec{F}_i - \gamma_i \vec{v}_i + \vec{R}_i(t) = m_i \vec{a}_i,$$

where $\vec{F}_i$ is still the sum of all forces exerted on atom i by other atoms explicitly present in the system.

This equation is often expressed in terms of the 'collision frequency' $\zeta = \gamma/m$.

The friction coefficient is related to the fluctuations of the random force by the *fluctuation-dissipation theorem*:

$$\langle \vec{R}_i(t) \rangle = 0,$$

$$\int \langle \vec{R}_i(0) \cdot \vec{R}_i(t) \rangle dt = 6 k_B T \gamma_i.$$

In simulations it is often assumed that the random force is completely uncorrelated at different times. That is, the above equation takes the form:

$$\langle \vec{R}_i(t) \cdot \vec{R}_i(t') \rangle = 6 k_B T \gamma_i \delta(t - t').$$

$\delta(t-t')$ Dirac delta functio

The temperature of the system being simulated is maintained via this relationship between $\vec{R}(t)$ and γ.

The jostling of a solute by solvent can expedite barrier crossing, and hence Langevin dynamics can search conformations better than Newtonian molecular dynamics ($\gamma = 0$).

Variants of Verlet, leap-frog, ... integration techniques are u

Worth Worrying About:

Ultimately, simulations are judged according to two basic criteria:

I. How well do the empirical energy surface and the chosen system composition approximate Nature?

$V(\vec{R})$:

How realistic are the chosen functional form and the associated numerical constants?

PSF Generation:

Which titratable groups should be protonated? Without employing quantum mechanics, protonations are assumed at the beginning and maintained throughout the simulation. Also, how much water is needed [15]? How many ions should be included?

II. How well is the energy surface (phase space) explored?

MD Simulation:

What length of simulation is sufficient? First, the system must be equilibrated such that system properties such as potential energy, temperature, and volume appear to have stopped drifting. Then the simulation must continue long enough to obtain reliable equilibrium averages.

MC Simulation:

Does the chosen 'move set' embody all motions relevant to the question being asked of the simulation? Have enough steps been taken?

Mistakes to Avoid:

Inconsistent $V(\vec{R})$:

The potential function (long-range cutoff keywords, distances, ...) should not be changed at different stages of a simulation study. All input scripts used in a research project that evaluate energies and forces (energy minimizations, annealings, dynamics simulations, ...) should *explicitly* (Don't trust the defaults!) do so in the same way.

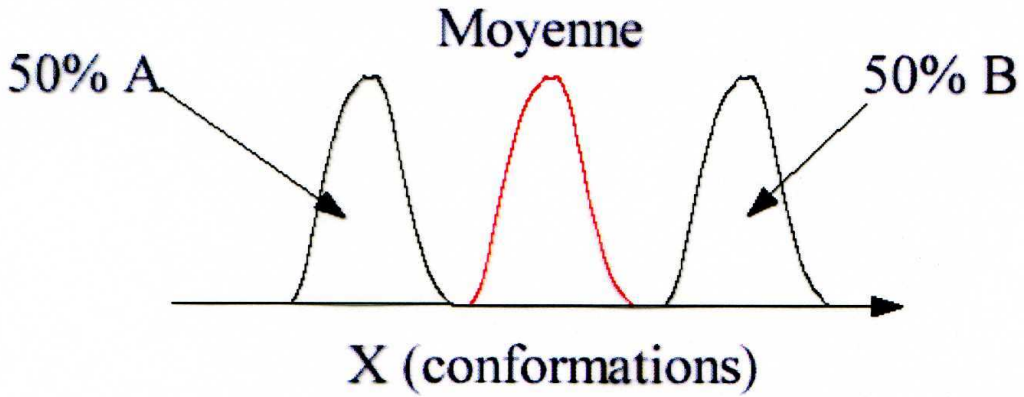
Submit and Forget:

Don't let a simulation run unmonitored. Check intermediate results daily. Plot the time dependences of the potential and total energies, the temperature, the pressure and volume (if applicable), and the root-mean-square deviation from a reference (crystallographic or t=0) structure.

Remember: Simulations are fiction aspiring to emulate reality. Pretty pictures and even a few good numbers do not guarantee good science.

RX; RMN:

- Structures recontruites à partir de données expérimentales moyennées (nbre de molécules x temps de la mesure)
  - Pb : Struct(<Exp>)≠<Struct(Exp)
  - Réalité physique de la structure moyenne ?



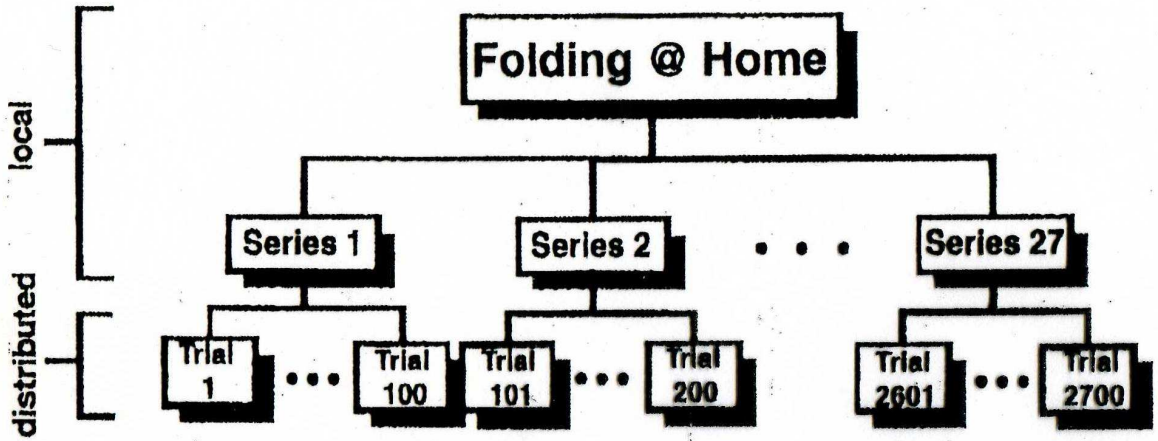
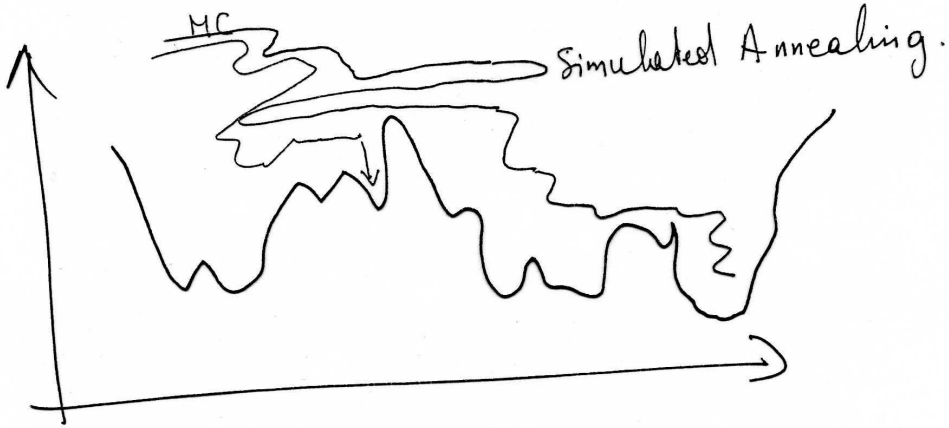


Figure 10. A schematic representation of the ensemble dynamics methodology. Locally, the Folding@Home servers initiate numerous Series of simulations. Each Series is independent of all others, and consists of 100 Trial simulations which are distributed to our thousands of users around the world. The data presented here were collected using a total of 27 Series offering 2700 Trials which resulted in a total of 38 μs of simulated time eight fully independent, successful folding trajectories. The Trials are coupled within a single Series as follows: when a given Trial crosses a free energy barrier (herein defined by an energy variance of 300 $kcal^2/mol^2$) all other Trials within that Series are restarted from that configuration, all 100 having different random number seeds (and thus different random force components) upon each restart. If no transitions above the required energy variance are detected, the method simplifies to a mass parallelization of fully independent simulations.

Simulated Annealing

(Recuit Simulé)

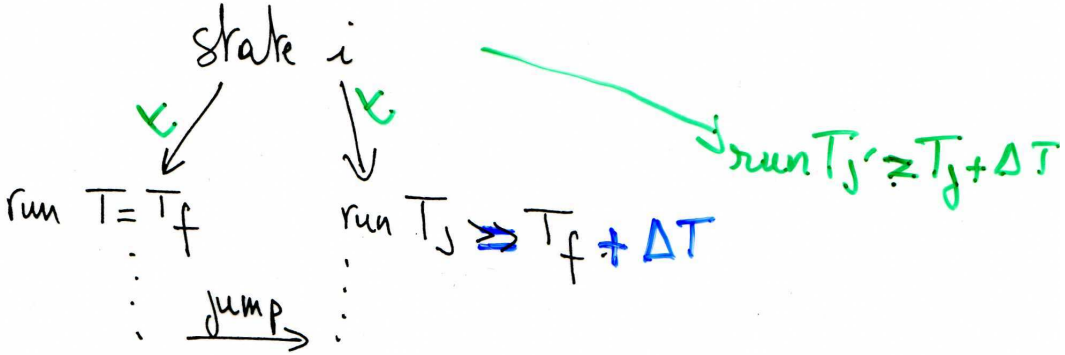
Kirkpatrick, Science 1983.



It is a MC variant where T is slowly decreased

Pb. • how to vary T
• if the variation of T is not adequate, the system does not converge to P_{eq}
• the protocol, which can be very complex, varies from one energy surface to another.

Simulated tempering (J-walking method)



$$\exp\left(-(\beta - \beta_J)\Delta E\right) \qquad \beta = \frac{1}{k_B T}, \quad \beta_J = \frac{1}{k_B T_J}.$$

also replica exchange MD simulations

- ΔT cannot be random
- number of jumps cannot be random.
  (exchange frequency $\simeq$ every 500 - 2000 MD steps

OTHER TECHNIQUES NOT SEEN HERE :

→Multicanonical simulations

→Genetic or Greedy Approaches

→Harmonic Dynamics

NEXT TIME: READING AND DISCUSSION

Atomic-Level Characterization of the Structural Dynamics of Proteins

David E. Shaw,[1,2]* Paul Maragakis,[1]† Kresten Lindorff-Larsen,[1]† Stefano Piana,[1]† Ron O. Dror,[1] Michael P. Eastwood,[1] Joseph A. Bank,[1] John M. Jumper,[1] John K. Salmon,[1] Yibing Shan,[1] Willy Wriggers[1]

Molecular dynamics (MD) simulations are widely used to study protein motions at an atomic level of detail, but they have been limited to time scales shorter than those of many biologically critical conformational changes. We examined two fundamental processes in protein dynamics—protein folding and conformational change within the folded state—by means of extremely long all-atom MD simulations conducted on a special-purpose machine. Equilibrium simulations of a WW protein domain captured multiple folding and unfolding events that consistently follow a well-defined folding pathway; separate simulations of the protein's constituent substructures shed light on possible determinants of this pathway. A 1-millisecond simulation of the folded protein BPTI reveals a small number of structurally distinct conformational states whose reversible interconversion is slower than local relaxations within those states by a factor of more than 1000.

Many biological processes involve functionally important changes in the three-dimensional structures of proteins. Conformational changes associated with protein folding (*1*), signal transduction (*2*), the catalytic cycles of enzymes (*3*), and the operation of molecular machines and motor proteins (*4*) often involve transitions among two or more structur-

ally distinct st
terized as "ba
"energy landsc

Substantial
both experime
techniques, in c
and the ways th
them. It has pr
ally characteriz
states and to el
anisms involve
states.

All-atom n
tions are desig
view of the n
cules (*9*), prod
the potential t
shots generate
tational constr

[1]D. E. Shaw Rese
10036, USA. [2]Cen
matics, Columbia
*To whom corres
David.Shaw@DES
†These authors c