UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR.961 - 13

WORKSHOP ON:
PROTEINS, MEMBRANES and their INTERACTIONS

22 JULY - 2 AUGUST 1996

"Advanced topics: reaction path calculations, simulations of chemical reactions and quantum computations"

PART III

Ron ELBER
The Hebrew University of Jerusalem
Department of Physical Chemistry
91904 Jerusalem
ISRAEL

MD simulations:

Equilibrium properties

Compute partition function or related quantities

$$Q = \int \exp[-\beta H]\, dP\, dX$$

$$Z = \int \exp[-\beta V(X)]\, dx$$

$$\beta \equiv 1/k_B T$$

$$\langle V \rangle = \frac{\int V(x)\, e^{-\beta V(x)}\, dx}{\int e^{-\beta V(x)}\, dx} = -\frac{\partial}{\partial \beta} \log(Z)$$

or

$$\langle A \rangle = \frac{\int A(X)\, e^{-\beta V(x)}\, dX}{\int e^{-\beta V(x)}\, dx}$$

Averaging $A(X)$ with a weight $e^{-\beta V(x)}$

Ergodic hypothesis

$$\langle A(t) \rangle_t = \frac{1}{\tau} \int_0^\tau A(t)\, dt = \langle A \rangle$$

Time average assumed equivalent to space average

How to obtain time average

?

start from a plausible structure $X(0)$ (X ray struct.)

sample velocities from Boltzmann distribution

Solve the Newton's eq. of motion $m\ddot{x} = -\partial U / \partial X$

Verlet algorithm

$$X(t+\Delta t) = X(t) + V(t)\Delta t + \frac{\Delta t^2}{2m}\left(-\frac{\partial U}{\partial x}(t)\right)$$

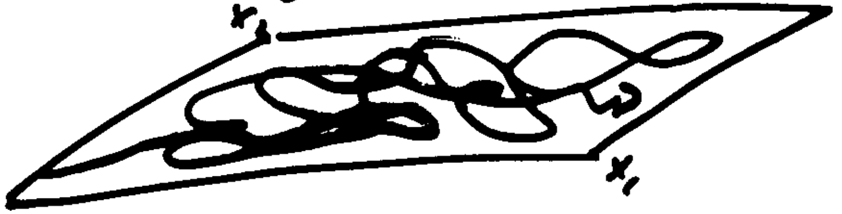
$$v(t+\Delta t) = V(t) + \Delta t \left(-\partial U(t) - \partial U(t+\Delta t)\right)$$

Pictorially

$$\langle A\rangle = \int e^{-\beta U} A \, dx \Big/ \int e^{-\beta U} dx$$

$$\langle A\rangle \simeq \frac{\sum_{i,j} \Delta x_{i1} \Delta y_{j2} \, e^{-\beta U(x_{i}, y_{j})} A(x_i, x_j)}{\sum \Delta x_{i} \Delta x_{j} \exp(-\beta U)}$$

$$\langle A\rangle = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0 A(t)\, dt$$



$$\langle A\rangle \simeq \frac{1}{\tau} \sum_i A(t_i)\, \Delta t$$

Relatively easy to compute averages. Difficult to make absolute sums.

$$Q = \int e^{-\beta U} dx - \langle ? \rangle$$

$$F = -k_B T \log(Q)$$

F - The free energy

Normally interested in

$$\Delta F_{AB} = F_A - F_B$$

$$K_{AB} = e^{-\beta \Delta F_{AB}} = \frac{[A]}{[B]}$$

$$F_A - F_B = -k_B T \log\left[\int e^{-\beta U_A} dx\right] + k_B T \log\left[\int e^{-\beta U_B} dx\right]$$

$$\Delta F_{AB} = -k_B T \log\left[\frac{\int e^{-\beta U_A} dx}{\int e^{-\beta U_B} dx}\right]$$

$$= -k_B T \log\left[\frac{\int e^{-\beta(U_A - U_B)} e^{-\beta U_B} dx}{\int e^{-\beta U_B} dx}\right]$$

$$\boxed{= -k_B T \log\left[\left\langle e^{-\beta(U_A - U_B)} \right\rangle_{U_B}\right]}$$

$$\cong -\frac{1}{\beta} \log\left[\left\langle 1 - \beta(U_A - U_B) \right\rangle_{U_B}\right]$$

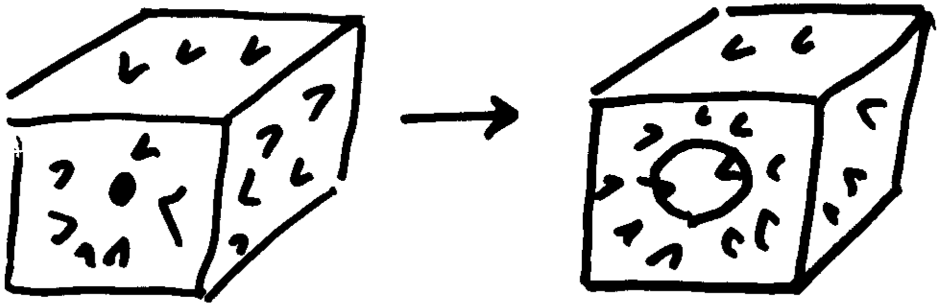
$$\cong \langle U_A - U_B \rangle_{U_B} = \langle \Delta U \rangle_{U_B}$$

If A and B are very different....

e.g. solvation of H_e / X_e in water

Compute $\langle e^{-\beta(U_{Xe}-U_{He})}\rangle_{U_{He}}$

Run He trajectory in water



Each H_e/water configuration replace H_e by X_e & compute U_{Xe}

ΔF [illegible] $\langle e^{-\beta(U_{Xe}-U_{He})}\rangle$ [illegible]

The solution to the sampling problems - intermediate states

Instead

$$He \longrightarrow Xe$$

We have

$$H_e \rightarrow 0.1 X_e + 0.9 H_e \rightarrow \dots 0.9 X_e + 0.1 H_e \rightarrow X_e$$

Compute a series of ΔF_q-s

$$\Delta F_q = -\frac{1}{\beta}\log\left[\langle e^{-\beta\Delta U_q}\rangle_q\right]$$

$$U_q = (1-q)U_{He} + qU_{Xe} \qquad 1 > q > 0$$

$$\Delta F = \int_0^1 dq\, \Delta F(q)$$

q is an arbitrary coordinate in the above case it is unphysical changing continuously He to Xe.

However in thermodynamics only the end point are relevant. The choice of q is therefore irrelevant as long as the end points are fixed.

There are examples in which q is important...

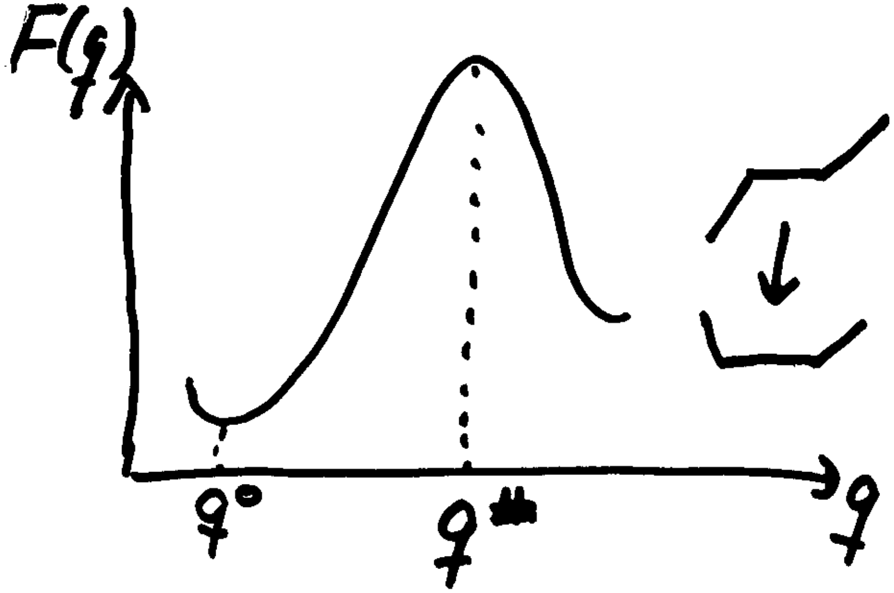
Transition state theory

$$\text{rate} = \langle V \rangle_+ \frac{e^{-\beta(F(q^*)-F(q^0))}}{\int e^{-\beta F(q)}\, dq}$$

$\langle V \rangle_+$ - thermally average velocity in the positive direction.

$$\langle V \rangle_+ = \frac{\int_0^\infty v\, e^{-\frac{1}{2}\beta m v^2}\, dv}{\int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta m v^2}\, dv}$$

A Reaction Coordinate



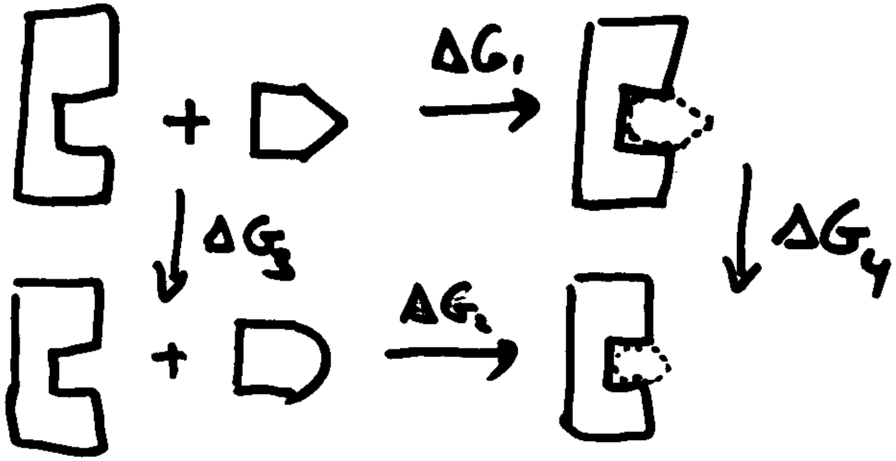
q^* - transition state

$$F(q^*) - F(q^\circ) = -\frac{1}{\beta}\log\left[\frac{\int e^{-\beta U(R,q^*)}\,dR}{\int e^{-\beta U(R,q^\circ)}\,dR}\right]$$

Thermodynamics cycle

Example:

Proteins bind different ligands



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

Exp: ΔG_1 ; ΔG_2

Calc: ΔG_3 ; ΔG_4

upine leghemoglobin

lobal fold similar to oxygen torage & transport proteins e.g. myoglobin, hemoglobin)

But

- found in plants
- serves as an oxygen trap to reduce oxygen concentration
- binding constant much larger than myoglobin & hemoglobin
- rate of binding much faster han myoglobin & hemoglobin
- much faster penetration to the active site and exit after photolysis

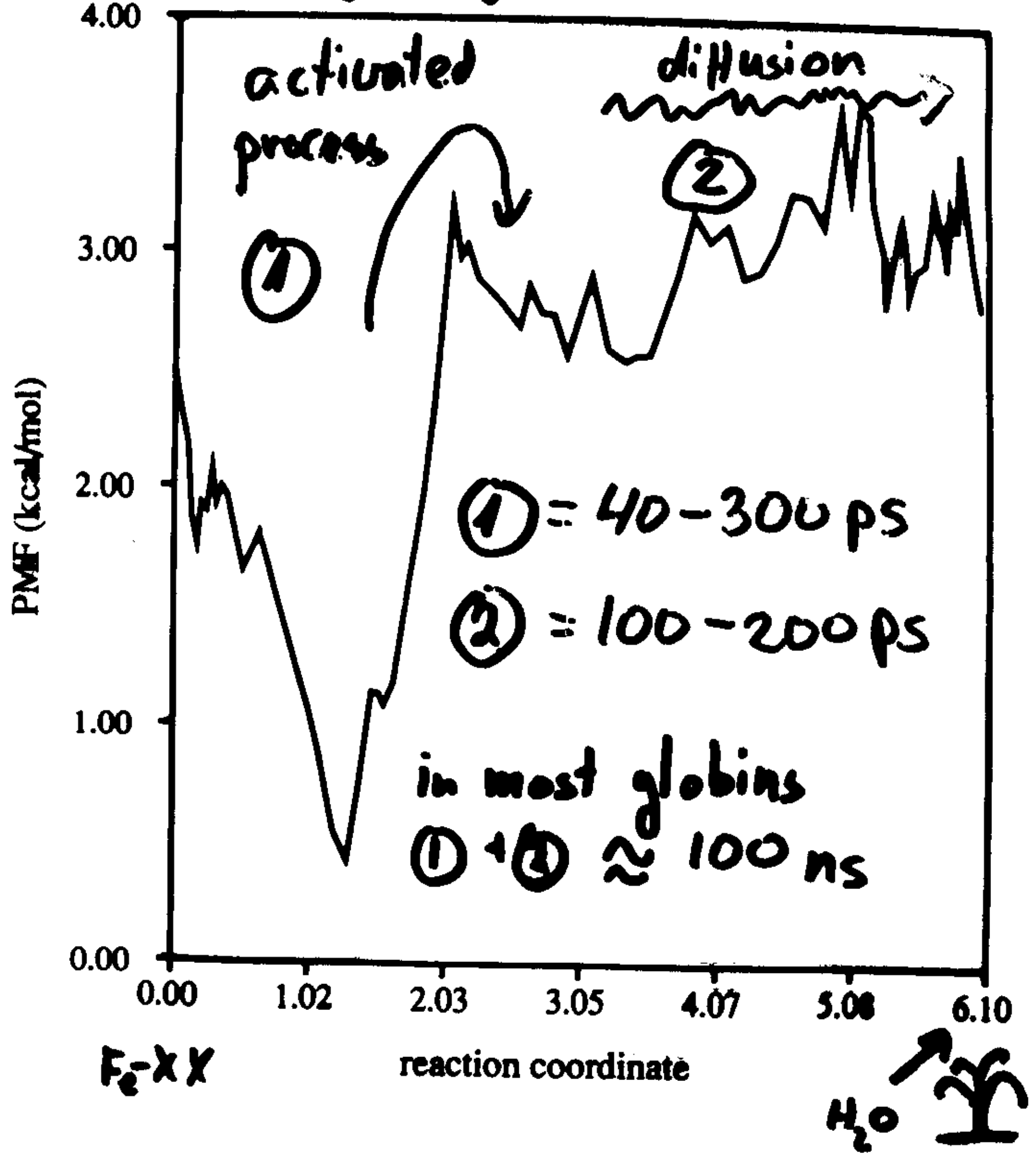
Diffusion of a ligand in a globin

Transient structure

Leghemoglobin: ligand buried in the [illegible] in matrix according to X-ray

experiment (motivated by theory): diffusion time scale ≈ free diffusion

otential of mean force for a ligand iffusion through leghemoglobin



① = 40-300 ps

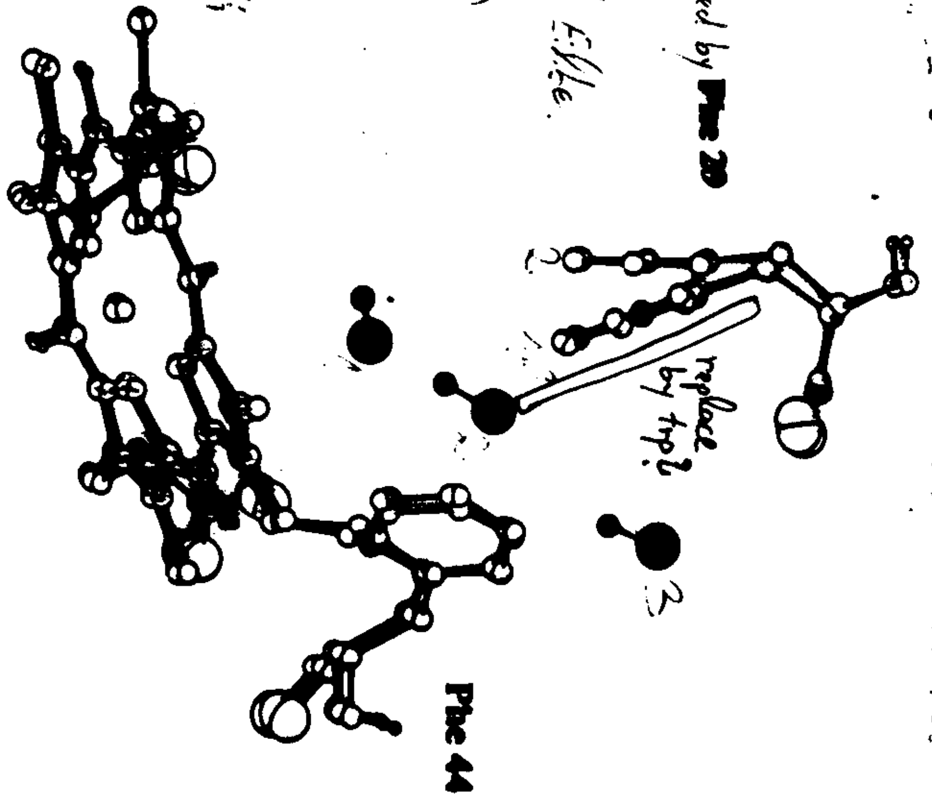
② = 100-200 ps

in most globins

① + ② ≈ 100 ns

Ligand diffusion through leghemoglobin

(a)



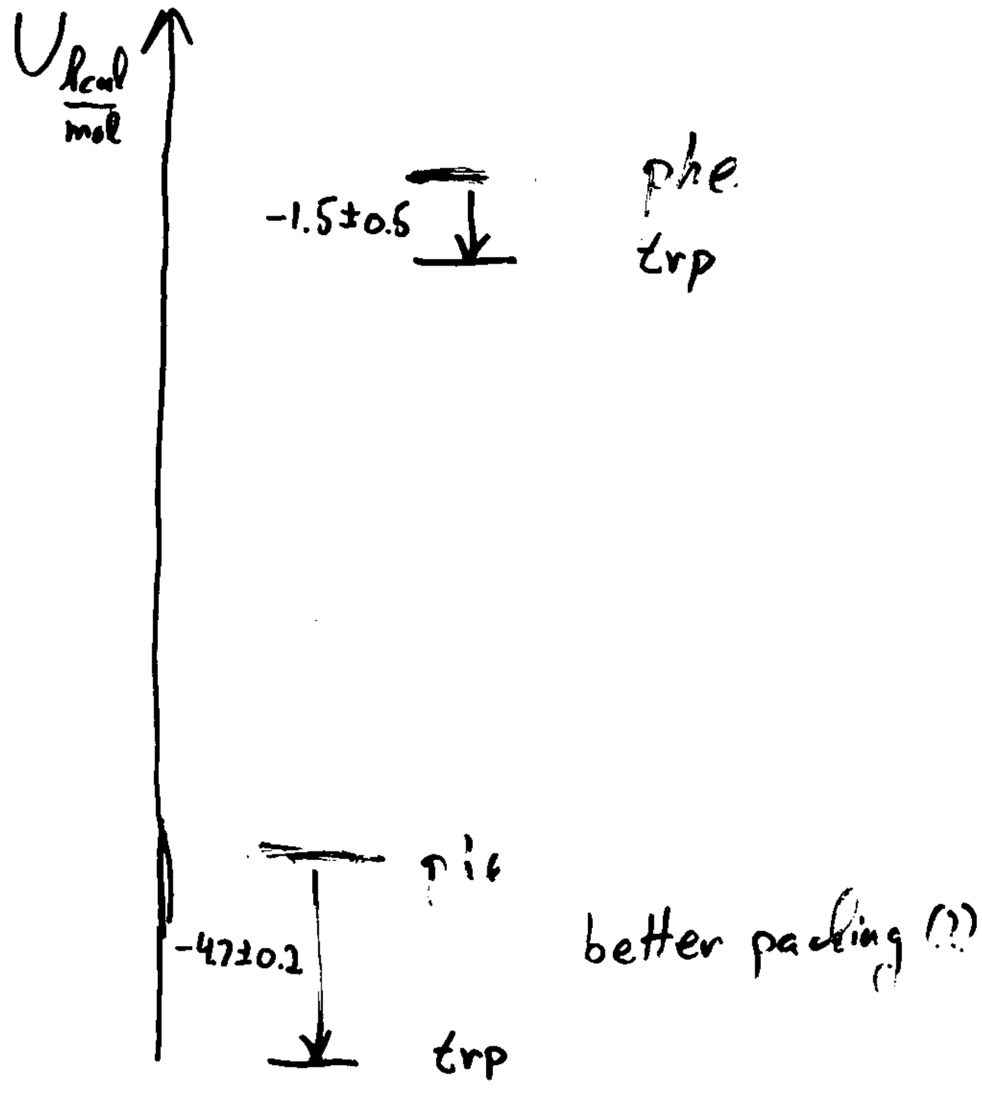
barrier dominated by Phe 20

Nowak, Czerminski & Elber.
JACS 113, 5627 (1991)

$\tau_{diffusion} \approx 250 ps$

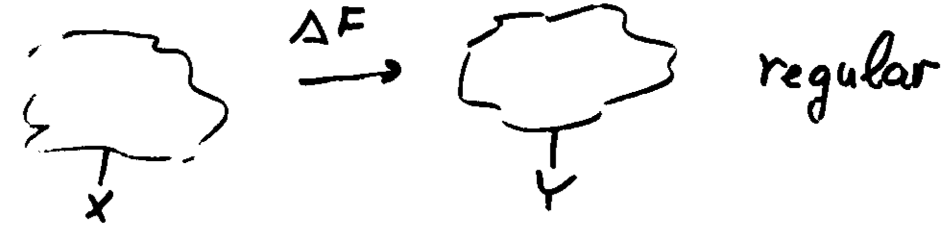
verified experimentally by Gibson

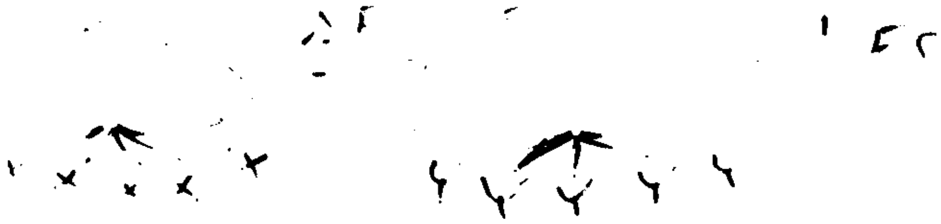
Mutating phe 29 to trp 29



Barrier increases by 3 kcal/mol

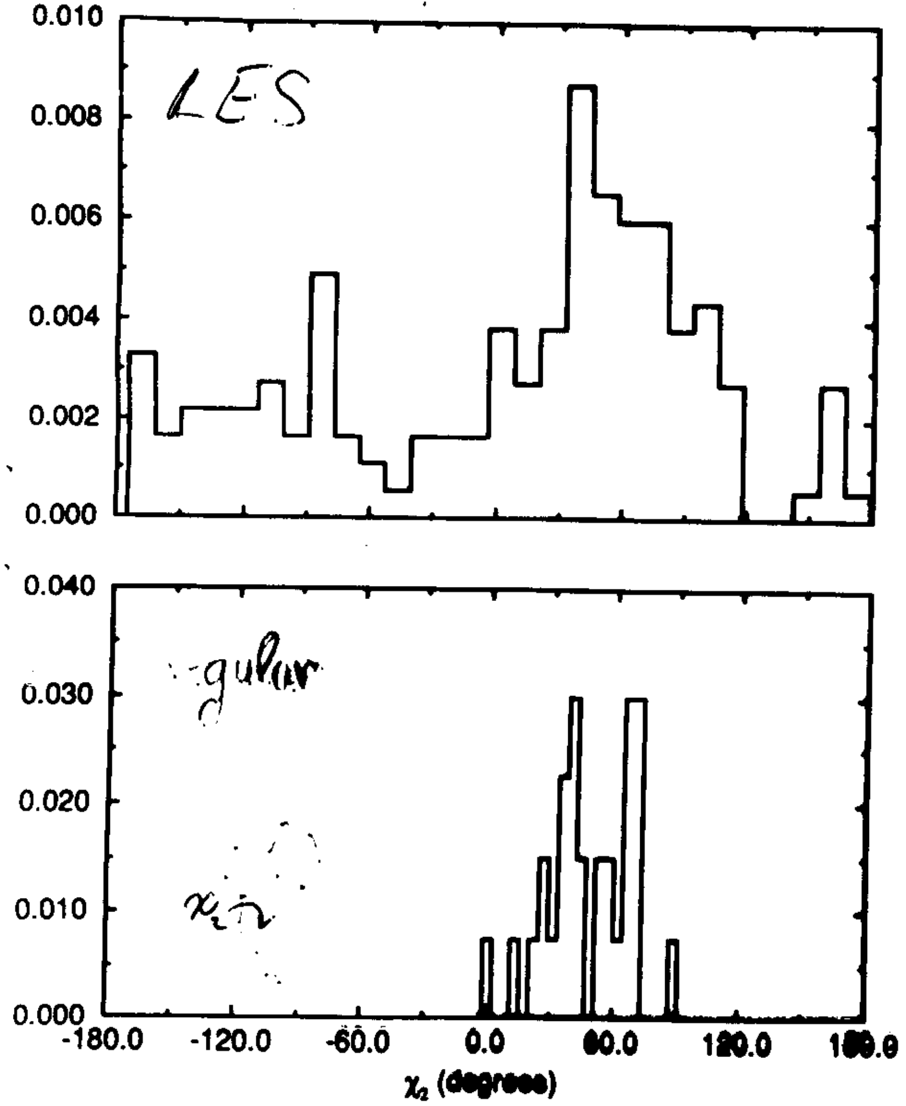
LES in calculation of free energy changes due to local structural change





Sampling dihedral angle conformational space (χ_2 for trp)

windows, ...



recent calculation: Is the trp mutation stable?

$$\Delta F = F_{trp} - F_{phe} = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{H_\lambda} d\lambda$$

$$\left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{H_\lambda} = \frac{1}{N} \sum_i \left(\partial H / \partial \lambda \right)_i$$

phe $$H_\lambda = (1-\lambda) H_{phe} + \lambda H_{trp} + H_{prot}$$ trp

(20 windows: -1.95 ± 3.8)

(35 windows: 3.9 ± 0.8 kcal/mol)

(1 window: 1.1 ± 0.1 kcal/mol)

(1 window: -1.6 ± 0.2 kcal/mol)

$\Delta F^{\lambda} = 3.9 \pm 0.8$

$\Delta F^{LES\lambda} = 4.9 \pm 1.5$

$$H_\lambda^{LES} = \frac{(1-\lambda)}{5} \sum_{i=1}^{5} H_{phe}^i + \frac{\lambda}{5} \sum_{i=1}^{5} H_{trp}^i + H_{prot}$$

phe
phe
phe
phe
phe

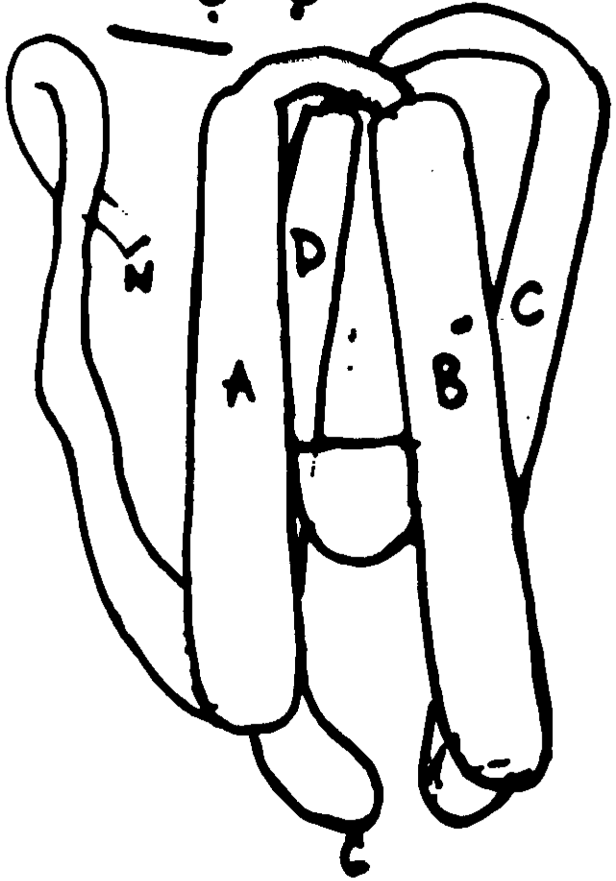
trp
trp
trp
trp
trp

20 windows: 5.4 ± 1.2 kcal/mol

Potential of mean force
to reduce system
dimensionality

$$W(q_1,\dots,q_n) = -\frac{1}{\beta}\log\left[\int e^{-\beta U(R,q)}\, dR\right]$$

re Protein Fluctuations
ominated by Rigid Helices Motions
??



f true: simulation of secondary
tructure motions
helices = rods

Secondary Structure (Helices) Motions in Proteins

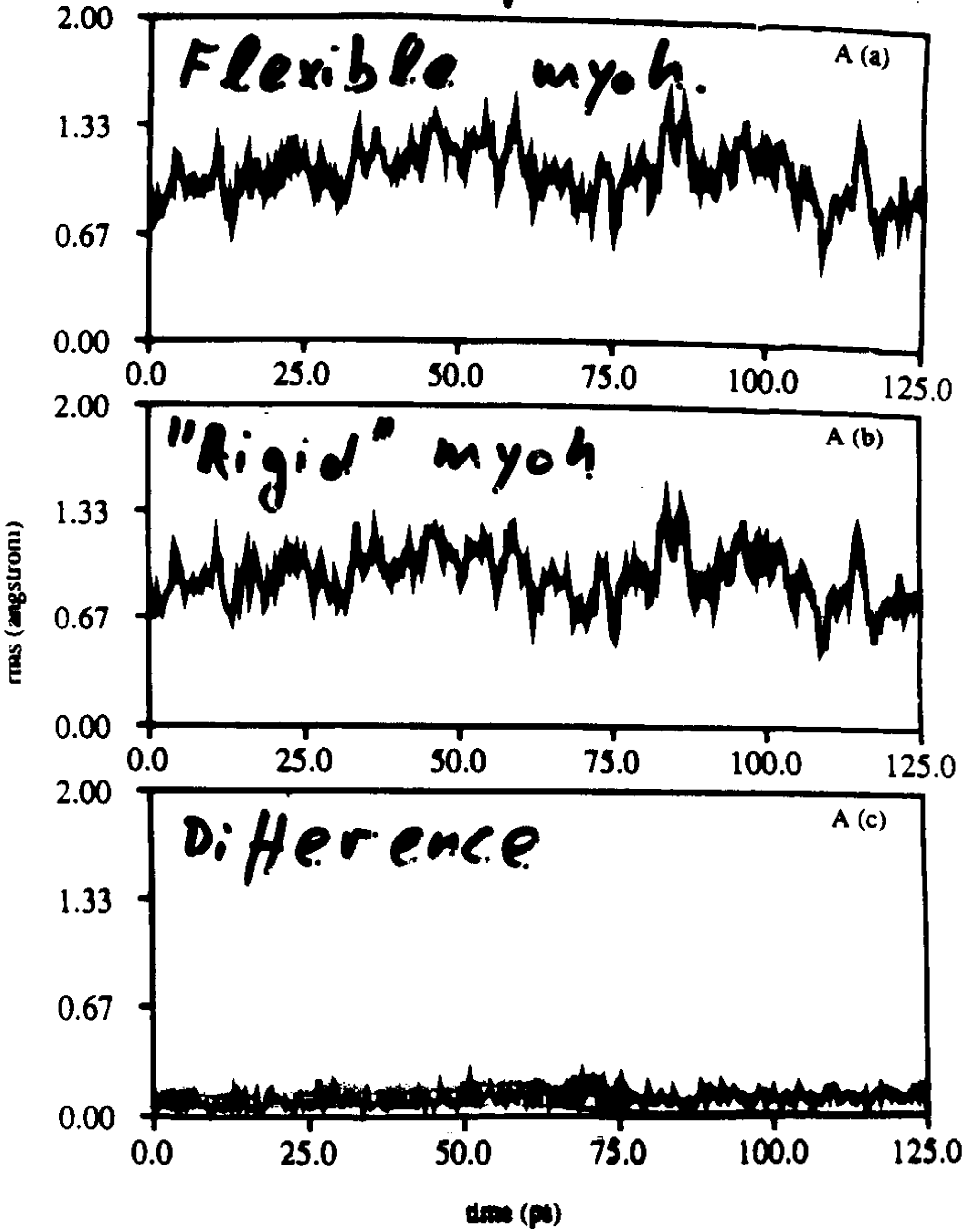
Danuta Rojewska

- "Rigid helices" valid approximation ??
- Get effective potential

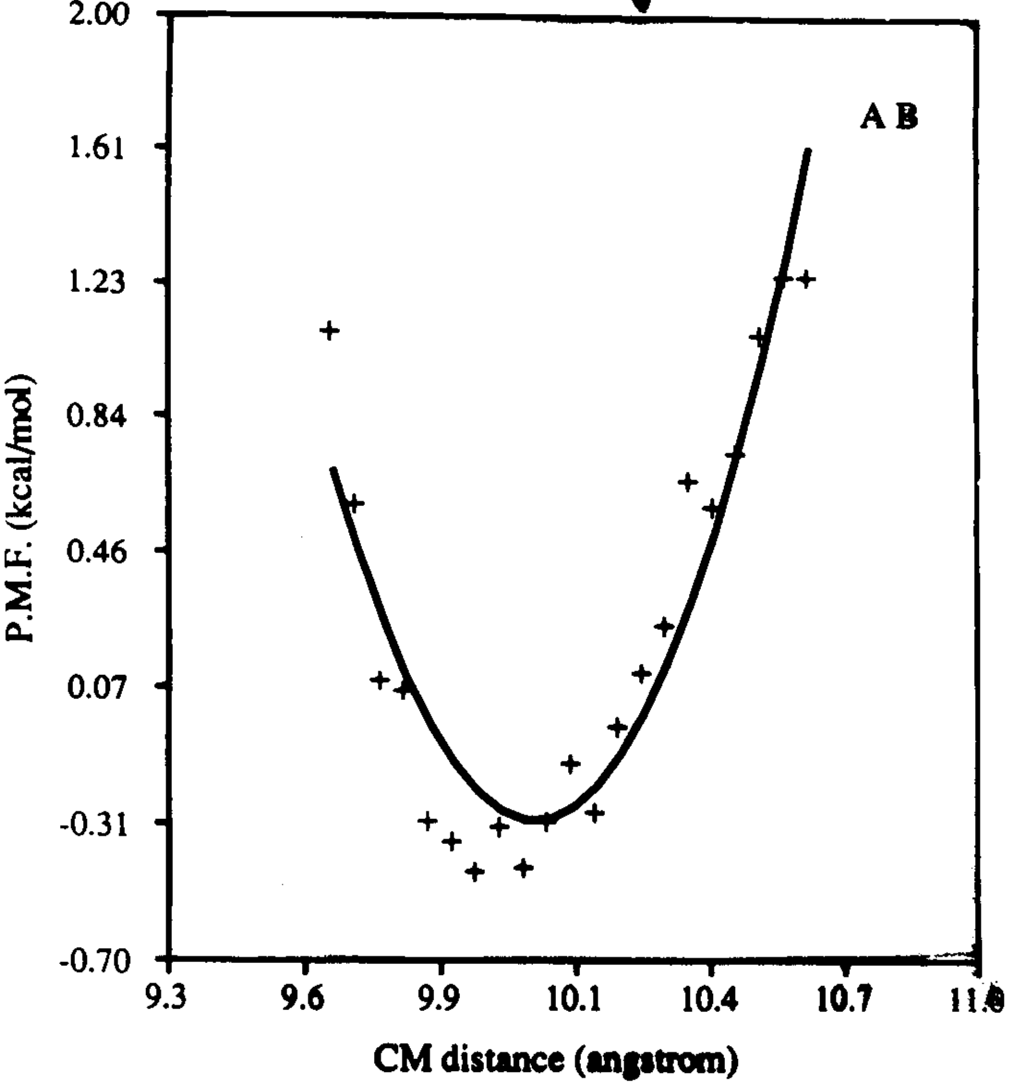
Study of myohemrythrin

PRF

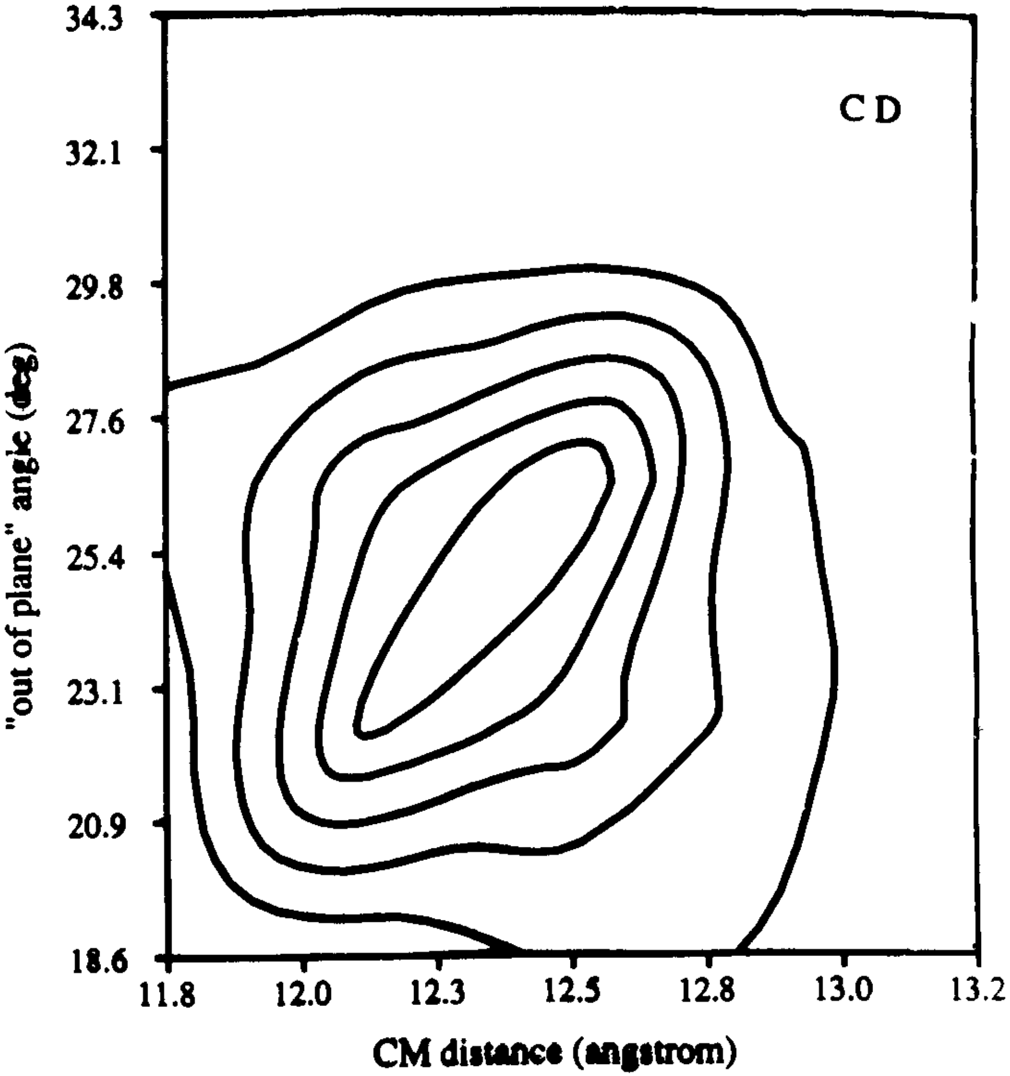
Fig 7,a Test rigid helices assumption



'otential of mean force for
elices A&B Danuta Rojews



Potential of mean force for
helices C & D



Extending the time scale of molecular dynamics simulations

Roberto Olender
and
Ron Elber

Hebrew University

Shekels: Israel Science Foundation

Molecular Dynamics:

Propagation in time - Given $X(t)$, $V(t)$ find $X(t+\Delta t)$, $V(t+\Delta t)$

Verlet algorithm:

$$X(t+\Delta t) = X(t) + V(t)\Delta t + 1/2M^{-1}\Delta t^2 F(t)$$

$$V(t+\Delta t) = V(t) + 1/2M^{-1}\Delta t[F(t)+F(t+\Delta t)]$$

To maintain <u>stability</u>
Δt must be small
$\Delta t \approx 10^{-15}-10^{-14}$ seconds
$t \approx 10^{-9}-10^{-8}$ seconds

NOT ACCESSIBLE TO DIRECT MOLCULAR DYNAMICS

(*) Ion transport through channels in membranes.

(*)Large conformational transitions and enzyme activations

(*) Protein folding

.
.
.

Statistical approximations (?!)
Reaction coordinates
Quasi-equilibrium

Molecular Dynamics Simulations

(*) No systematic approximations, NO approximations at all(!)

(*)Chaotic trajectories, high accuracy impossible.

(*) Selection of relevant trajectories

In contrast to....

Electronic structure calculations

(*) Many approximate techniques

(*) Accuracy: As desired

(*) Ground state and energy minima

The Onsager Machlup action:

Enables computations of trajectories between two coordinate sets fixed in space, and with a fixed length of time.

(*) The solution is stable for almost an arbitrary time step.

(*) High frequency motions are automatically eliminated, adding to the stability of the solution.

(*) A weight for indiviudal trajectories enables selection of more important trajectories and to bias the sampling in computations of averages.

LONG TIME DYNAMICS VIA PATH INTEGRAL CALCULATION

THE ONSAGER-MACHLUP ACTION

Program:

1. Theory and assumptions

2. Why large time step possible

3. Consequences of large Δt

5. Simple examples: HO, dipeptide

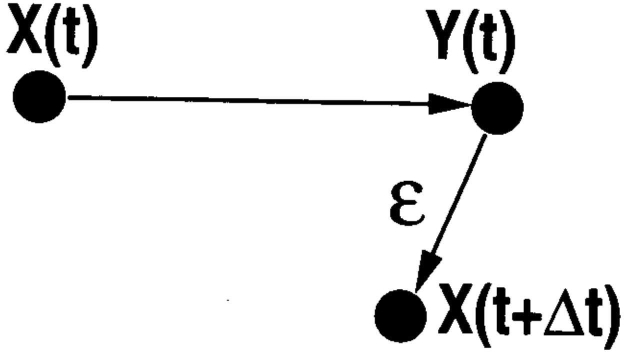
6. A more complex example: $10\mu s$ hemoglobin trajectory

PROPAGATION WITH ERRORS

$$Y(t) = \text{Propagator}(X(t))$$

$$X(t+\Delta t) = Y(t+\Delta t) + \varepsilon$$

(ε = error)



ASSUMPTION: ε a Gaussian random vector of errors.
E.g. Langevin equation

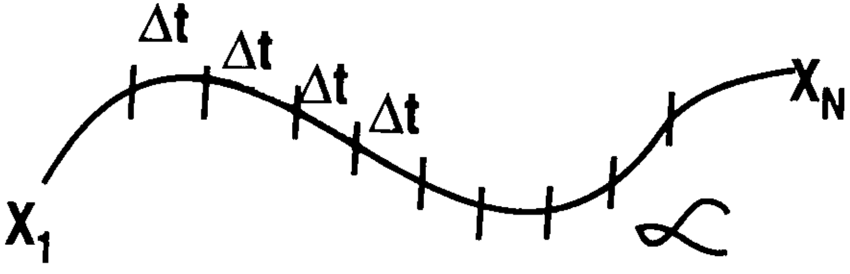
$$P(X(t+\Delta t)|X(t)) \propto \exp[\,-(X(t+\Delta t)-Y(t))^2 / 2\langle\varepsilon^2\rangle]$$

EXAMPLE: The Langevin eq.

$$X(t+\Delta t) = X(t) + V(t)\Delta t + 1/2M^{-1}\Delta t^2[-dU/dX(t)-\gamma V(t)+R]$$

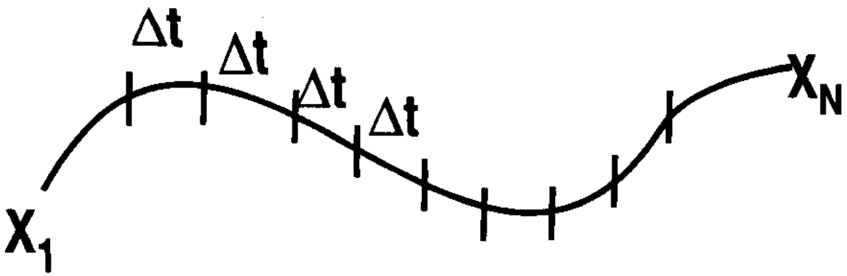
$$Y(t) = X(t) + V(t)\Delta t + 1/2M^{-1}\Delta t^2[-dU/dX(t)-\gamma V(t)]$$

$$P[X(t+\Delta t)|X(t);\Delta t] \propto \exp[-((X(t+\Delta t)-Y(t))^2[1/2\langle 2M/\Delta t^2 R\rangle^2]$$



$\propto$

Probability of a whole trajectory



$P(X_1|X_N;X(t))$

$\Pi_i P[X(t_i+\Delta t)|X(t_i);\Delta t] =$

$$\exp(-\Sigma_i(X(t_i+\Delta t)-Y(t_i))^2/2\sigma^2) =$$
$$\exp(-\Sigma_i(\varepsilon^2/2\sigma^2)$$

Error also equal to the local value of the differential equation:

$$Md^2X/dt^2+\gamma dX/dt + dU/dX = \varepsilon$$

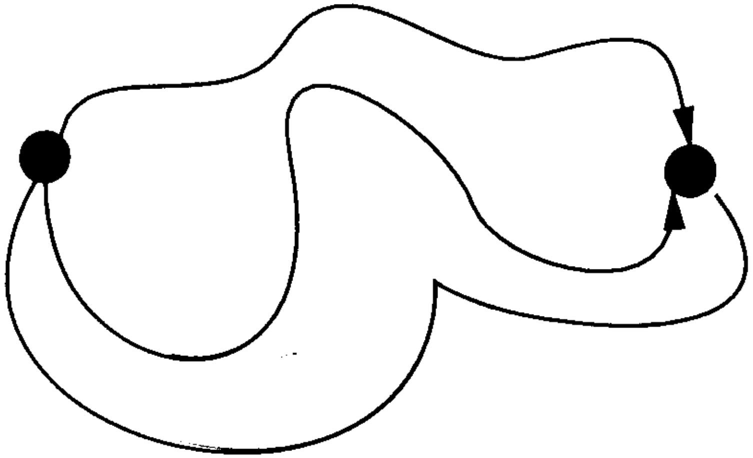
A weight for a trajectory

$\rightarrow P(X_1|X_N;X(t)) \propto$

$\exp[-1//2\sigma^2\int(Md^2X/dt^2+\gamma dX/dt+dU/dX)^2dt]$

$\exp[-1//2\sigma^2 S]$

This is the weight of a specific trajectory X(t)



THE RELATION BETWEEN THE PROBABILITY AND THE ACTION – S

$$P(X_1|X_N;X(t)) \propto \exp[-1//2\sigma^2 S]$$

$$S = \int (Md^2X/dt^2 + \gamma dX/dt + dU/dX)^2 dt$$

S is always positive, easy to optimize.

A minimum of S provides the most probable path.

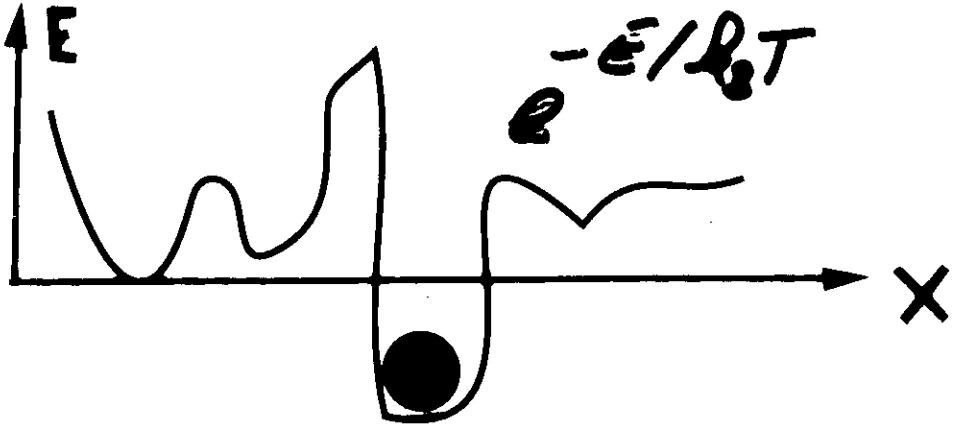
There is an analogy between the weight of a trajectory – $\exp[-1//2\sigma^2 S]$ and the energy for a single structure – $\exp(-\beta E)$.

$$Q = \int \exp(-\beta E)\, dX$$

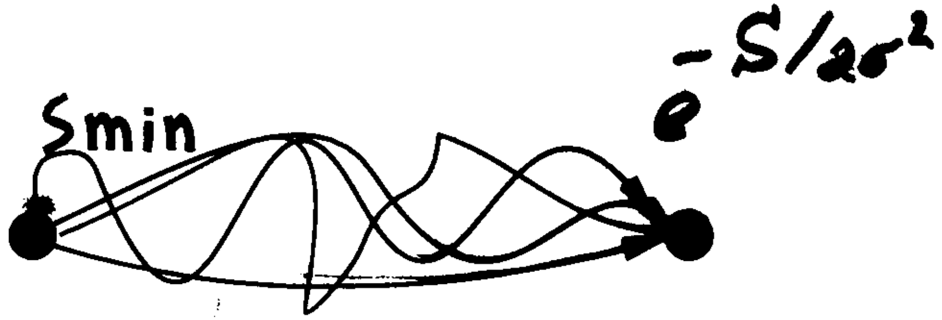
$$P(X_1|X_N;t) = \int \exp[-1//2\sigma^2 S]\, DX(t)$$

Analogy between the energy E and the action S:

E(X) – at a minimum the most probable structure



S({X(t)}) – at a minimum most probable trajectory

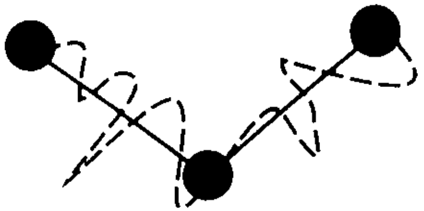


SHORT TIME APPROXIMATION TO S

ELIMINATES HIGH FREQUENCIES

$$\omega > 1/\Delta t$$

$$S \cong \{ M[X(t+\Delta t)+X(t-\Delta t)-2X(t)]/\Delta t^2 \\ + dU/dX(t) \}^2 \Delta t$$



Use exact expansion of $X(\tau)$:

$$X(\tau) = X(t-\Delta t) + [X(t+\Delta t)-X(t-\Delta t)]/2\Delta t\ \tau + \\ [X(t+\Delta t)+X(t-\Delta t)-2X(t)]/\Delta t^2\ \tau^2/2 \\ + \Sigma_k a_k \sin(k\pi\tau/\Delta t)$$

TO BE CONTINUED ...

Elimination of high frequencies, continuation...

with $dU/dX \approx dU/dX(t) = \text{constant}$

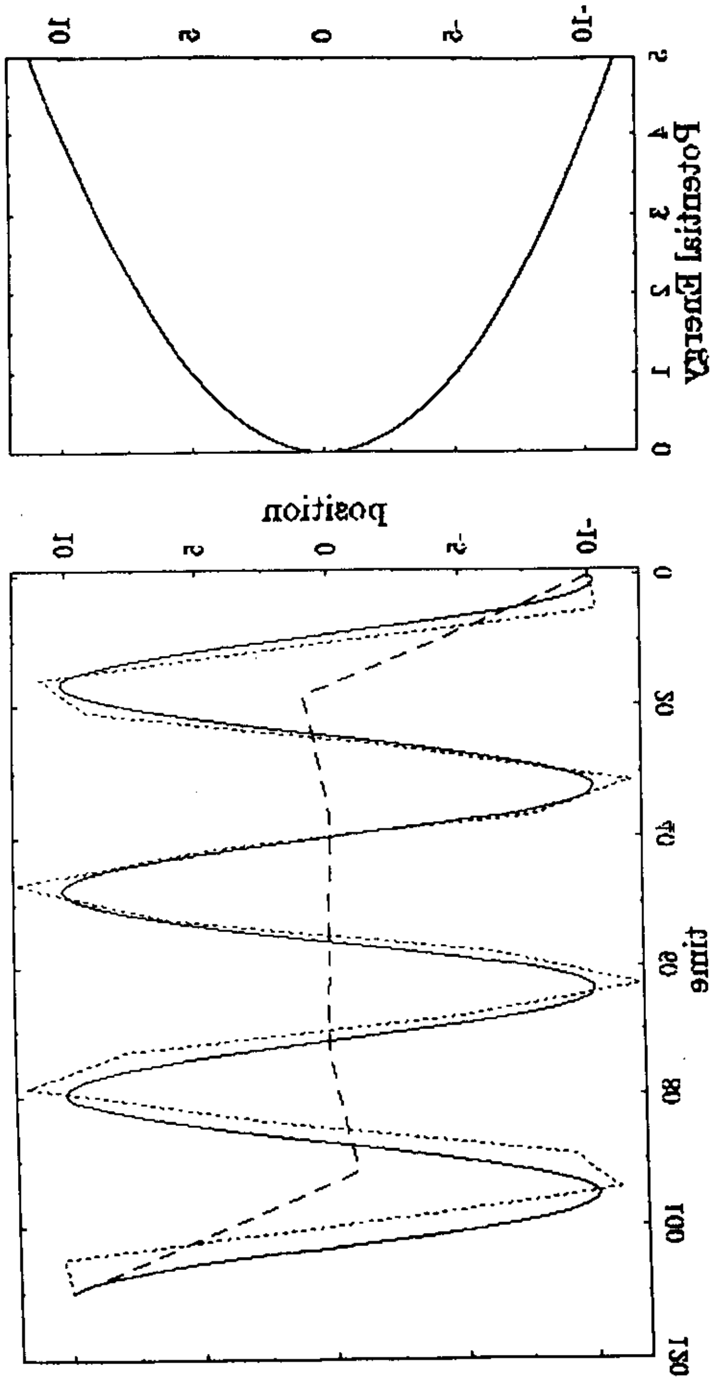
$$S = \{ M[X(t+\Delta t)+X(t-\Delta t)-2X(t)]/\Delta t^2 \\ + dU/dX(t) \}^2 \Delta t \\ + \pi^4/\Delta t^3 \Sigma_k k^4 a_k{}^2$$

$$\partial S/\partial a_k = 2\pi^4/\Delta t^3 k^4 a_k$$

$$\longrightarrow \quad a_k(\min) = 0$$

MAINTAIN STABILITY

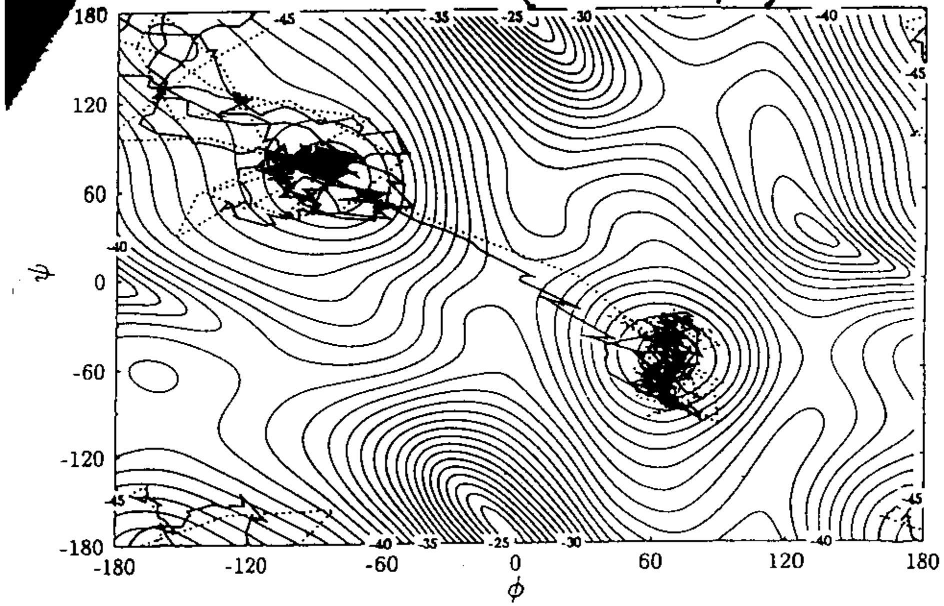
OM solution for the harmonic oscilator

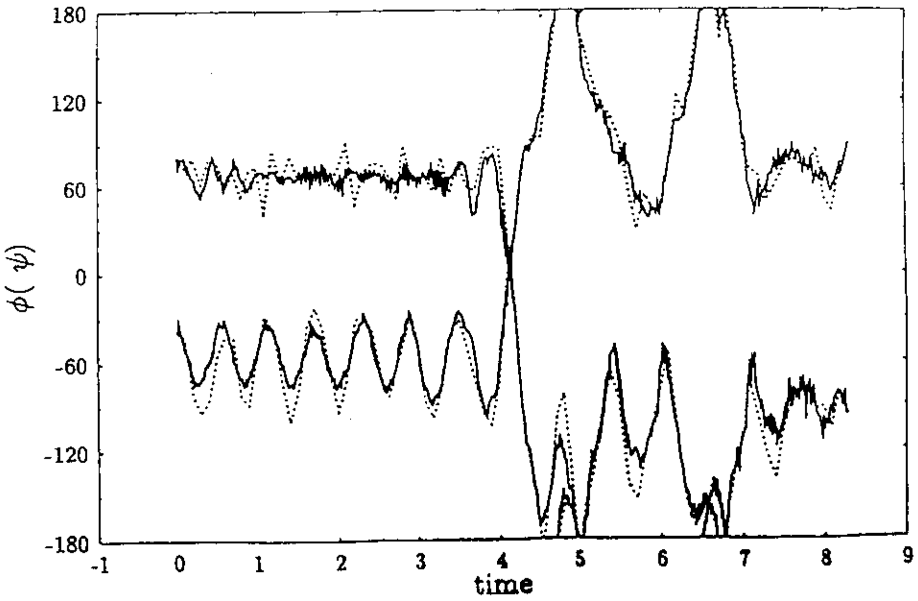


Time step of --- same order as period

ment of the dynamics togheter with the OM minimization

Optimization of an MD trajectory with OM ($\Delta t = 0.1$ PS)

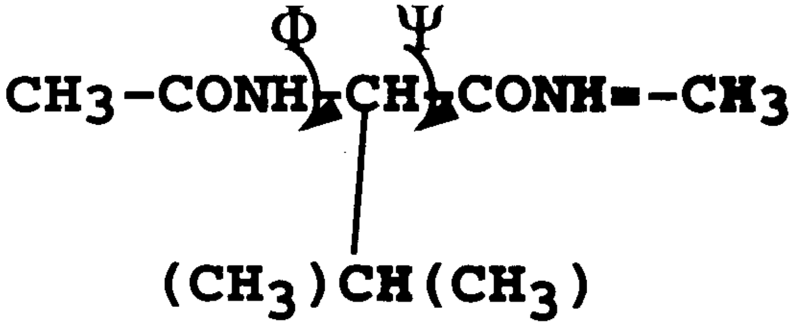




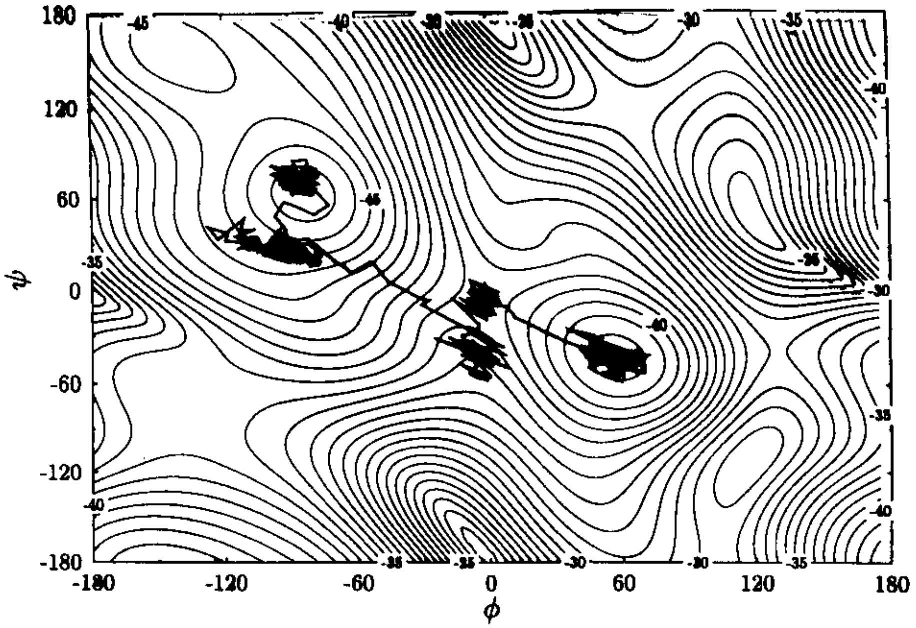
Onsager-Machlup action

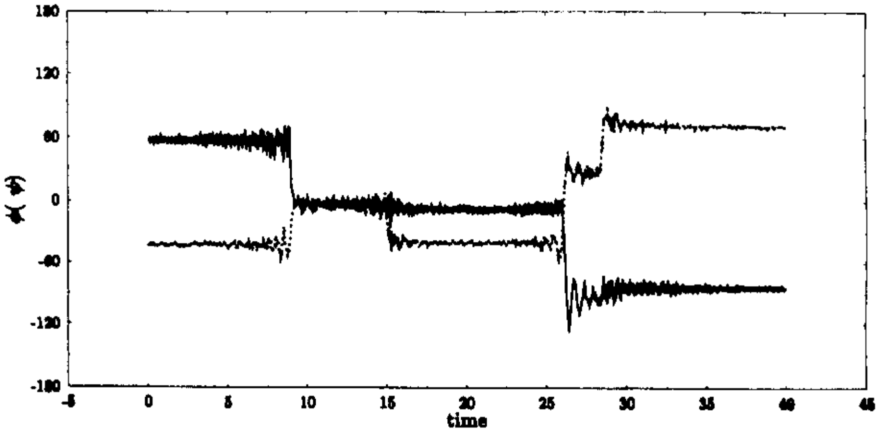
Roberto Olender

Example: Conformational transition in a dipeptide



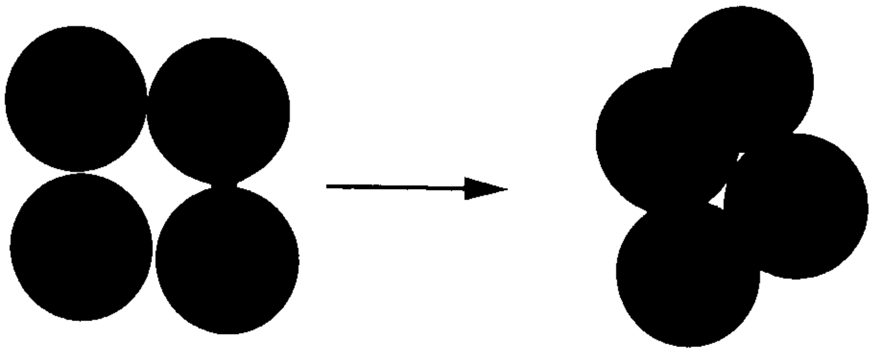
$$\begin{array}{c} CH_3-CONH-\overset{}{CH}-CONH-CH_3 \\ | \\ (CH_3)CH(CH_3) \end{array}$$



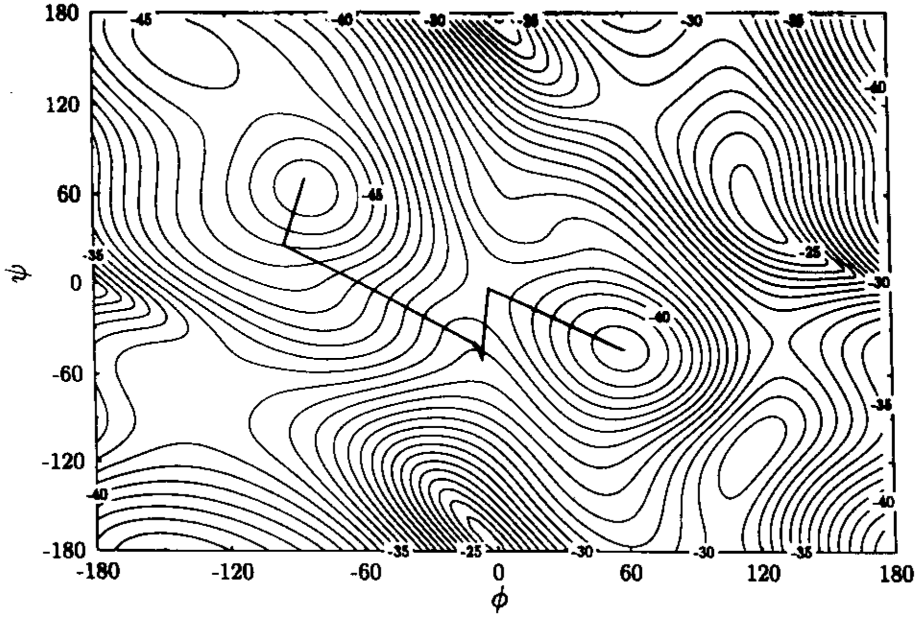


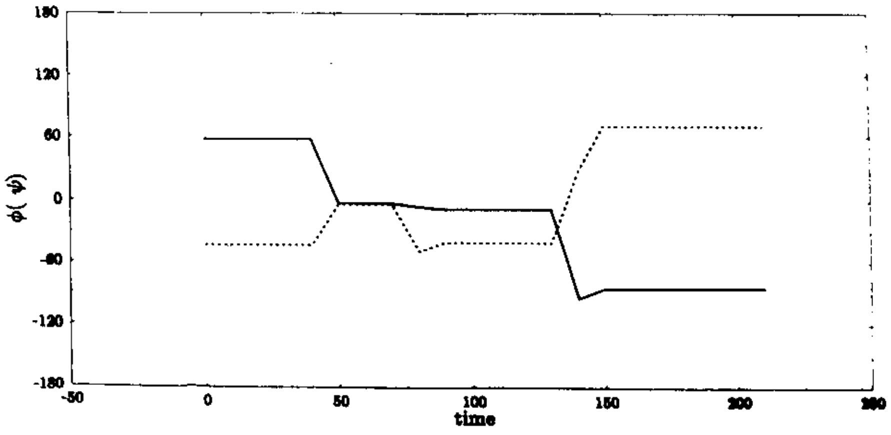
$\Delta t = 20\,fs$

The R->T transition in HEMOGLOBIN



A conformational transition from an oxygen binding structure to a structure with significantly lower oxygen affinity

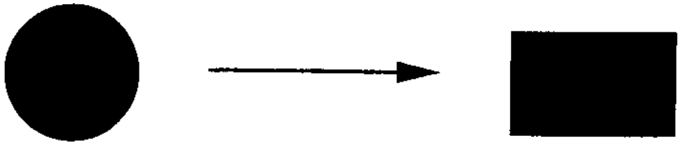




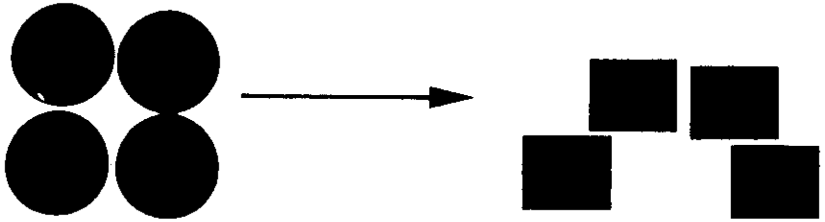
$\Delta t = 20$ ps

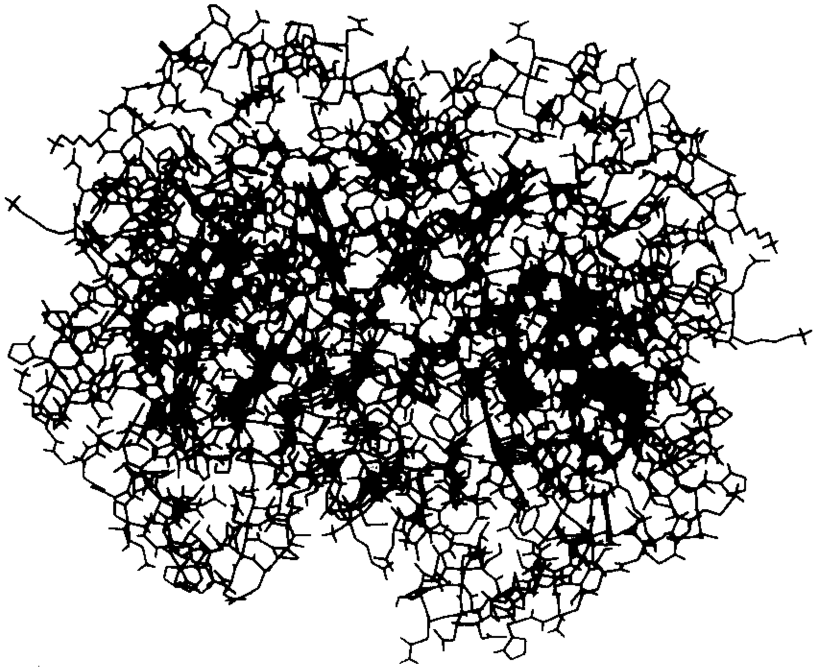
R -----> T

There are teriary changes (within the subunits)



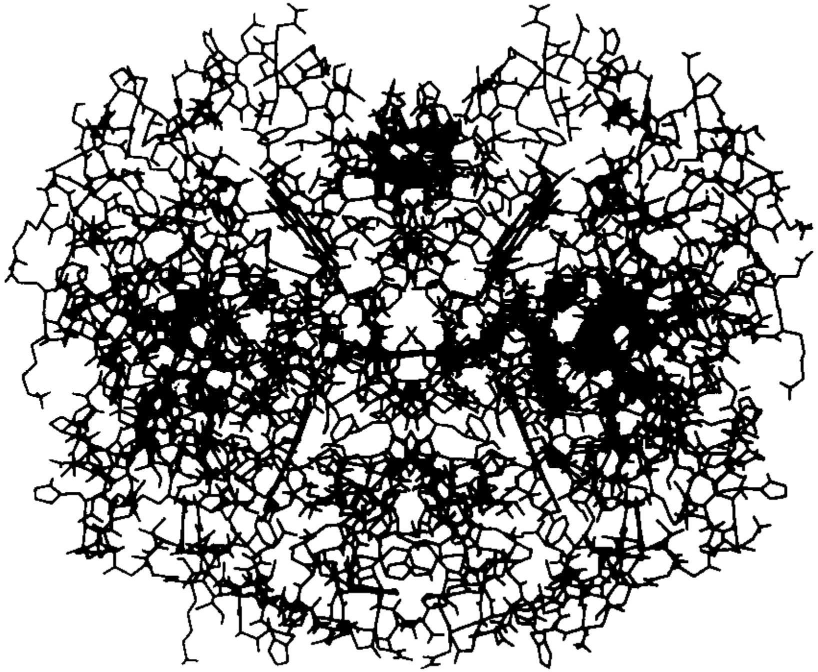
and quaternary changes (in the packing of the different subunits).





WHICH ONE IS FIRST?

Trajectory history



100 structures used to approximate the path integral.

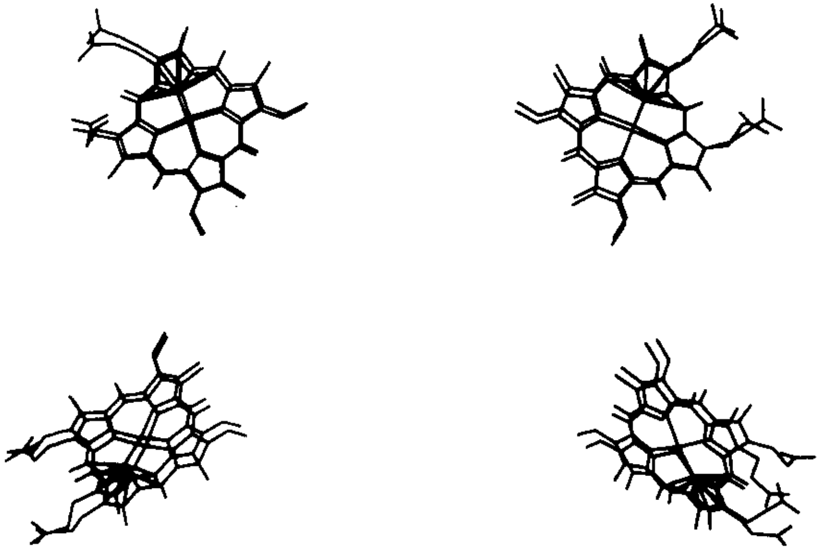
100 nanosecond is the time step.

10 μs total trajectory.

Initial point R.
Final point T.

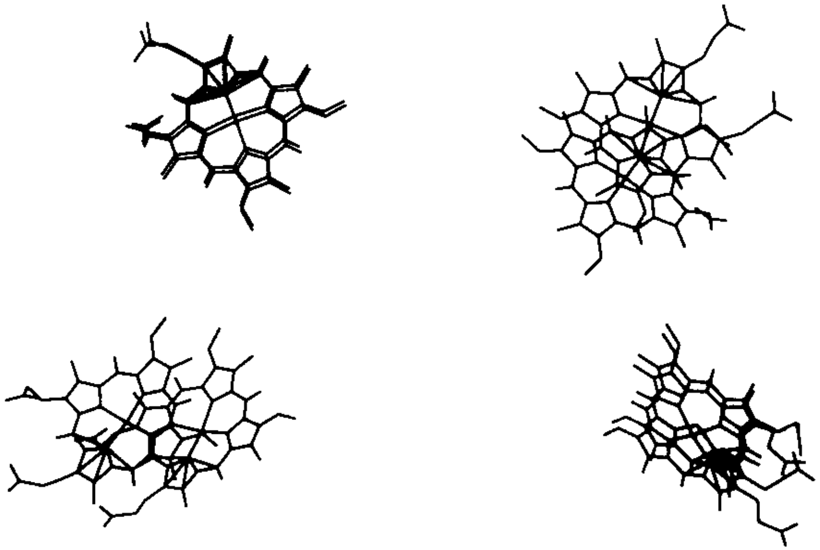
No oxygens.

HEM.1.CRD
HEM.20.CRD



MOIL-View by Carlos Simmerling

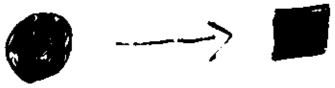
HEM.1.CRD
HEM.60.CRD



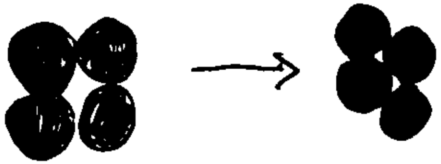
MOIL-View by Carlos Simmerling

Hemoglobin:

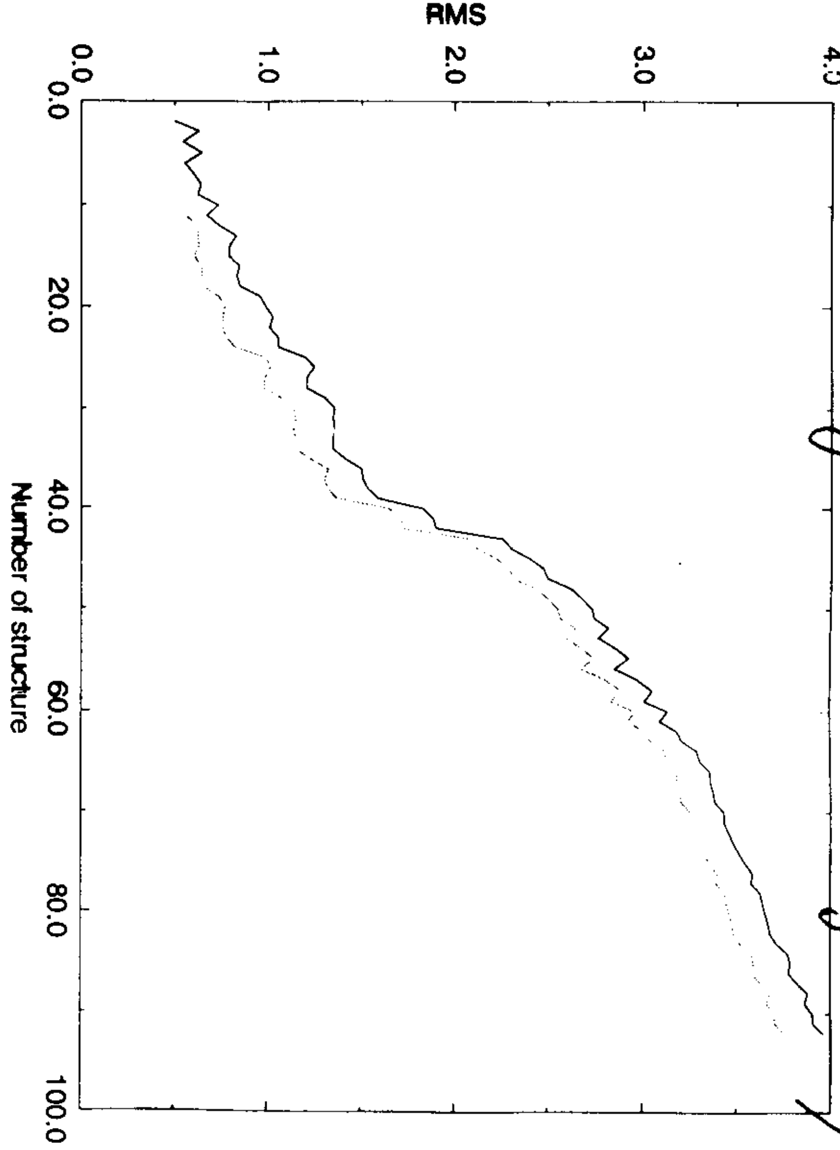
(*) Tertiary changes slow and continuous at all times.



(*) Changes in packing of the subunits abrupt.



Comparing the R state to structures along the R→T trajectory



OM path optimization
suggests itself as a tool for
computing very long
atomically detailed
trajectories,
even for very large systems.

Other things we know:

(*) How to formulate rate theory in terms of the OM.

(*) Run in parallel.

(*) Computations of relative rates (e.g. enzymatic reaction of mutant enzymes)

(*) Re–formulation and a new algorithm to compute the steepest descent path.

Things we work on

(*) Exact estimation of errors

(*) Simple approximations for the rate via path linearization.