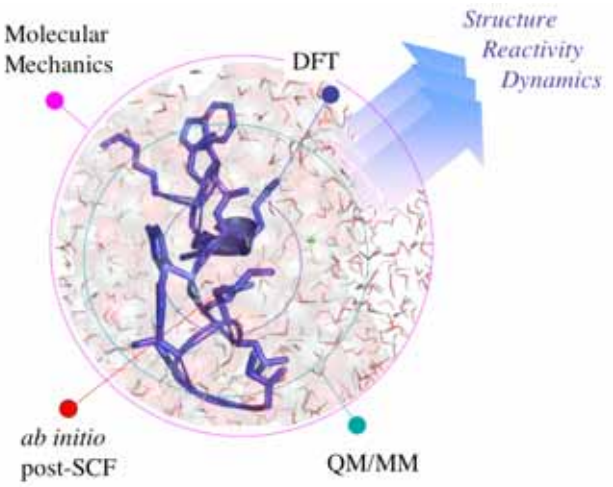
Molecular Dynamics Simulations of Biological Systems



Scuola di Chimica Computazionale
Introduzione, per Esercizi, all'Uso del Calcolatore
in Chimica Organica e Biologica.
Siena 25-29 Settembre 2006.

Giorgio Colombo
ICRM-CNR, Milano
g.colombo@icrm.cnr.it

Overview

1) Molecular Models:
   - Choice of the model
   - Sampling the configurational space
   - Studying problems with MD

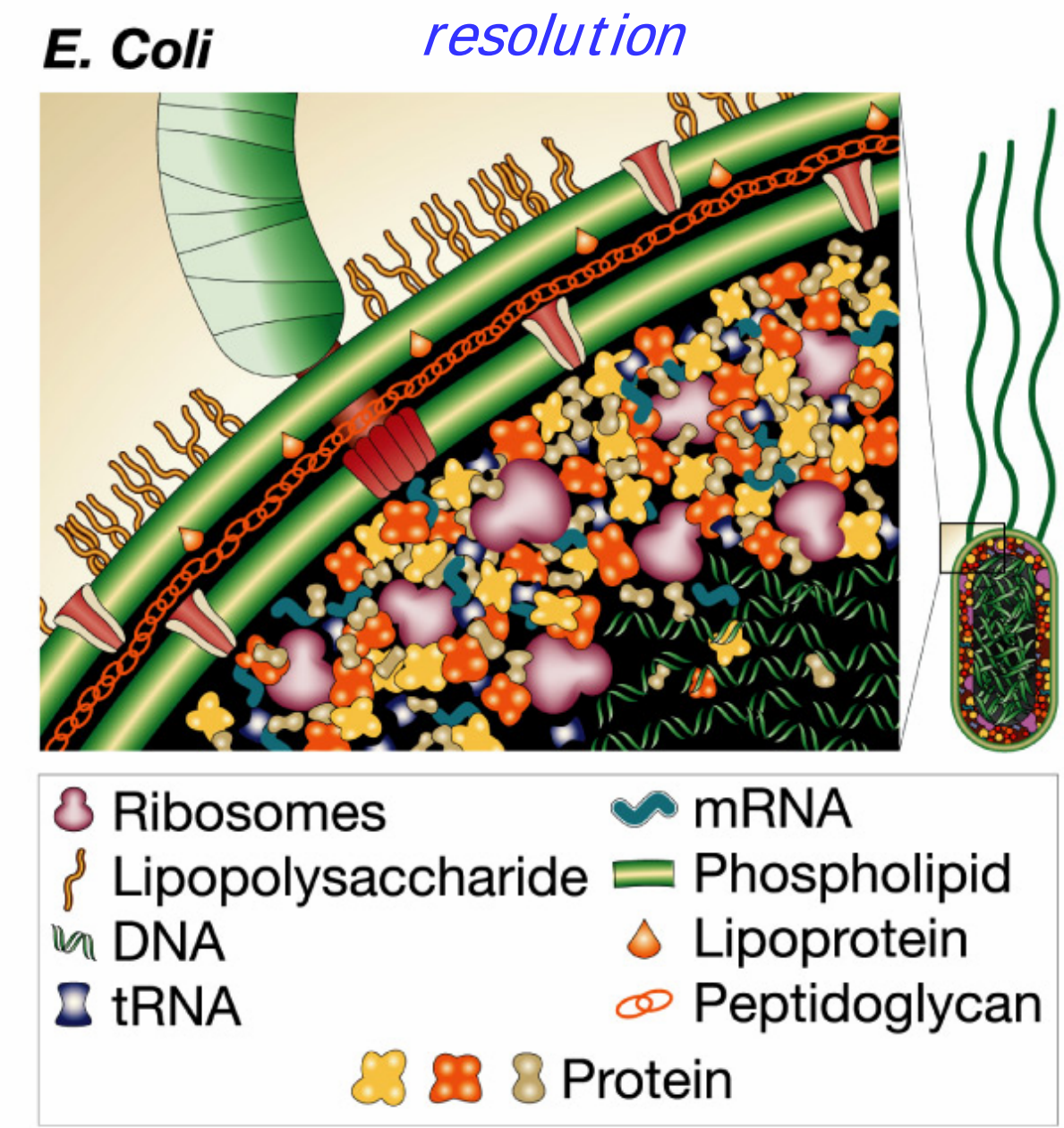
2) Molecular Dynamics:
   - Methodology
   - Treatment of long range forces; T and P control
   - Integration of the equations of motion
   - Convergence of the simulations
   - Real Case Studies

3) Application to a real chemical problem:
   NGR anticancer peptides

4) Conclusions and Outlook

Ultimate Goal:

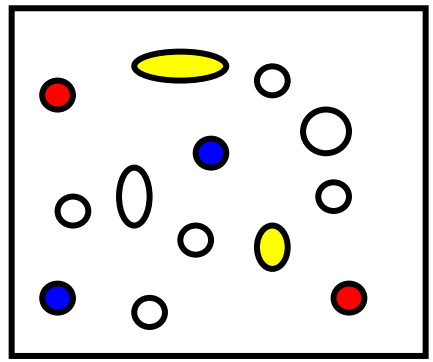
to understand cell mechanisms at atomic resolution



Molecular Simulations

Molecular simulation is a computational "experiment" conducted on a molecular model

10 to 100000 atoms are simulated



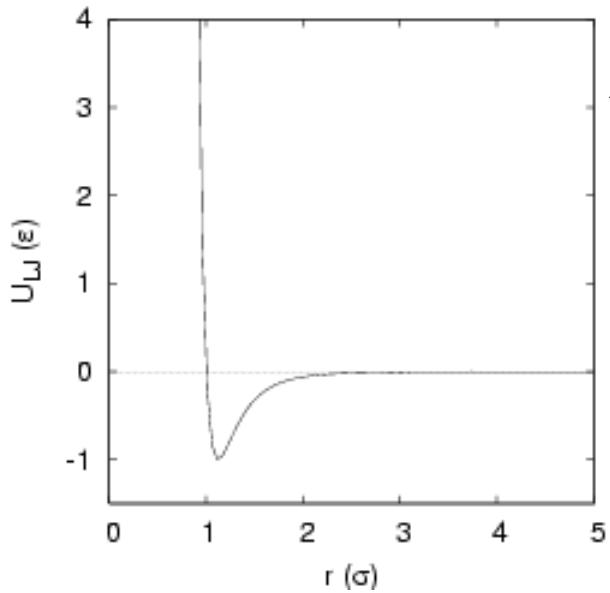
Many configurations are generated, and AVERAGES taken to yield the "measurement". One of these 2 methods is used:

Molecular Dynamics	Monte Carlo
Integration of equations of motion	Ensemble Average
Deterministic	Stochastic
Retains Time element	No element of time

Molecular simulation has the character of both theory and experiment. Applicable to molecules ranging from rare gases to polymers to metals.

Molecular Model

A Molecular model postulates the interactions between molecules.



A typical two body spherical potential Lennard-Jones

$$\phi_{LJ}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

More realistic models require other interatomic contributions

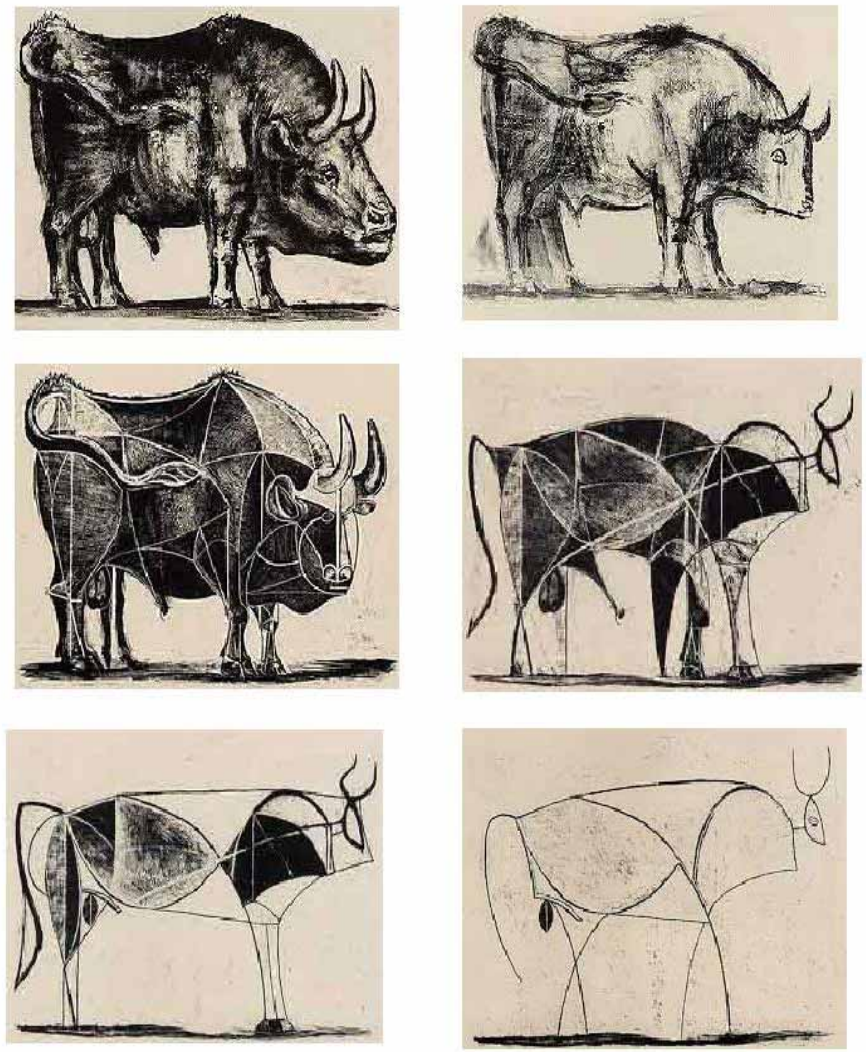
Intramolecular: Stretch, Bend, out of plane, torsion

Intermolecular: Van der Waals, electrostatic, multibody

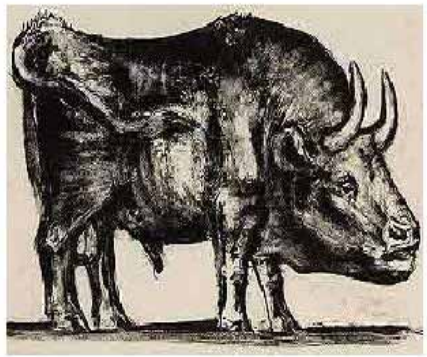
Quantum Mechanics

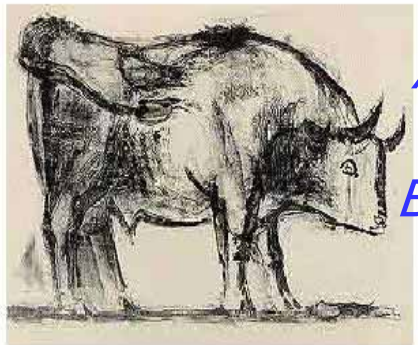
Importance of the Model

A good model should retain the basic ingredients for the description of the phenomenon



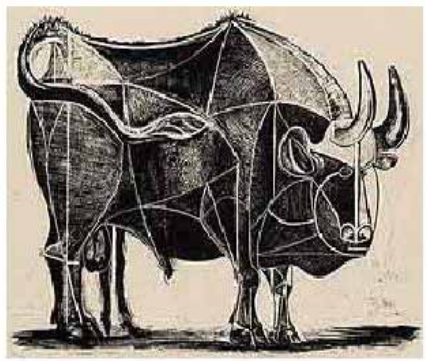
Quantum Mechanics
Nuclei, Electrons
Reactions

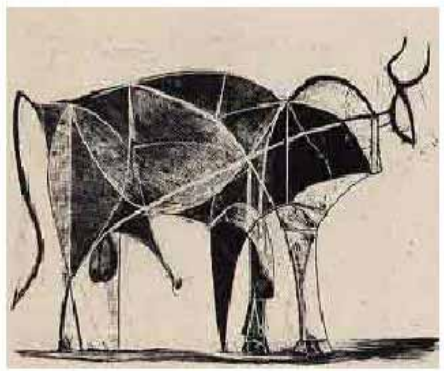




All Atoms, Polariz.
Atomic dipoles
Binding of charged ligands

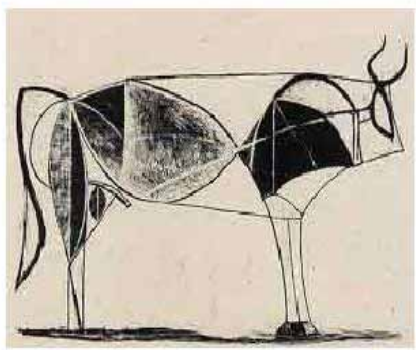
All Atoms
solute+solvent atoms
Hydration

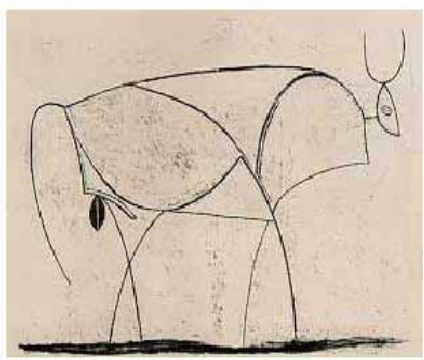




Solute Atoms
Solute atoms
Conformations

Group Atoms as Balls
Atom groups
Folding topologies





Residues as Balls
Residues
Folding thermodynamics

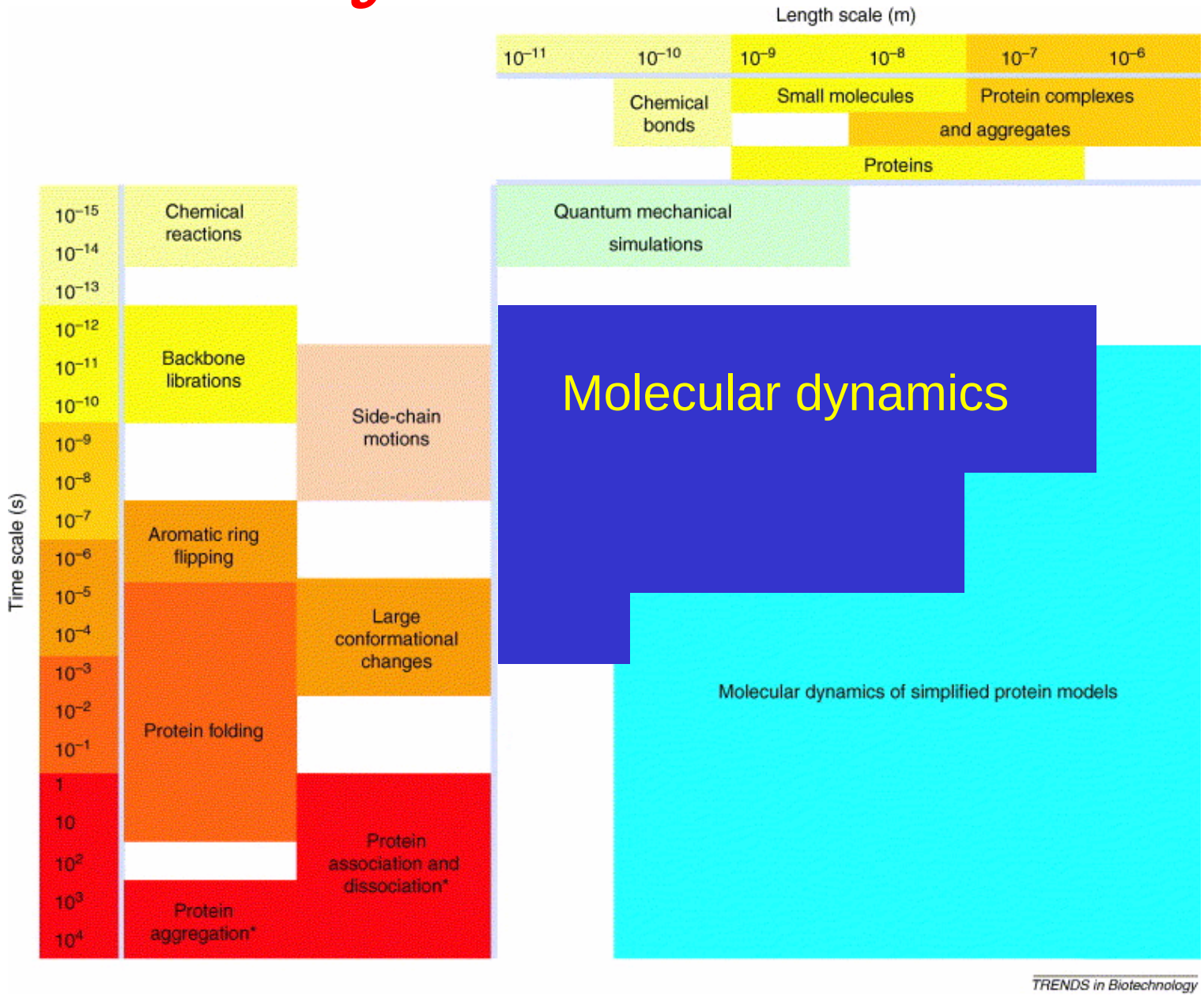
When choosing a model one should include only those degrees of freedom on which the property depends

Model	Degrees of freedom		Example of Property	
	Left	Removed	Predicted	Force Field
Quantum mechanical	Nuclei, electrons	nucleons	Reactions	Coulomb
All atoms, polariz	Atoms dipoles	electrons	Binding charged ligands	Ionic models
All atoms	Solute + solvent atoms	dipoles	hydration	GROMOS
All solute atoms	Solute atoms	solvent	Gas phase conformation	MM2
Groups of atoms as balls	Atom groups	Individual atoms	Folding topology of macromolecules	LW

Increase:
simplicity
speed
search power
timescale

Decrease:
complexity
accuracy

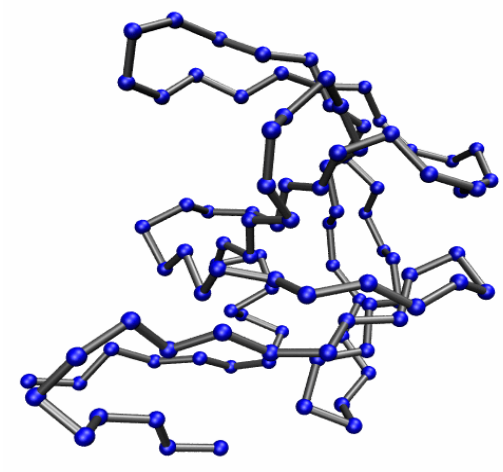
Biological Systems
Lengths and Time scales



Study of the folding of HIV-PR with a simplified model

Residues as Balls centered
On the Cα atom. Fixed bonds

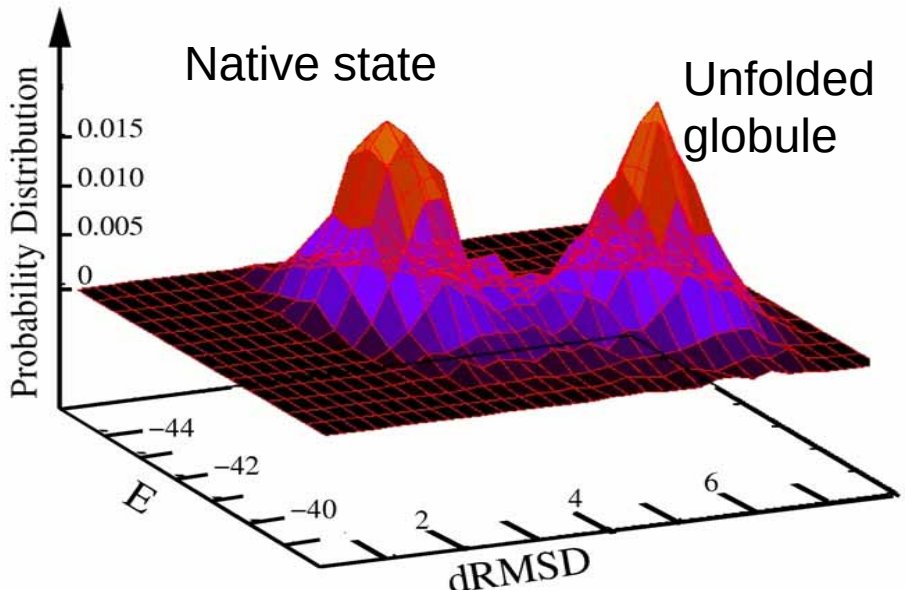
$$U(\{r_i\}) = \sum_{i<j} \varepsilon_{ij} \Delta(|r_i - r_j|) \Delta(|r_i^N - r_j^N|)$$



ε_{ij} Energy parameter

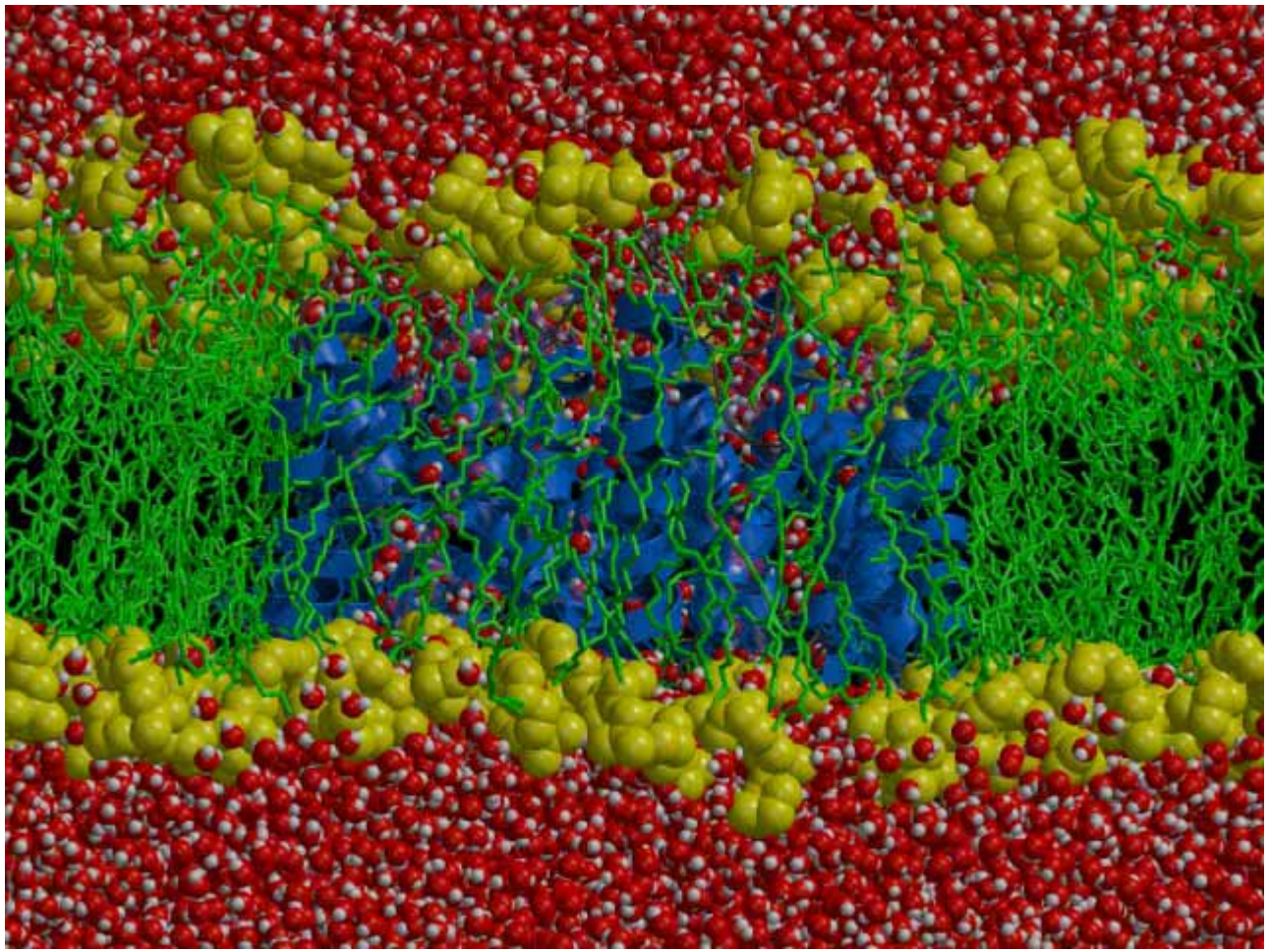
$\Delta(|r_i - r_j|) = 1$ if $|r_i - r_j| < 6.5$ Å
and $i < j + 2$
0 otherwise

$\{r^N\}$ are coordinates of the native conformation

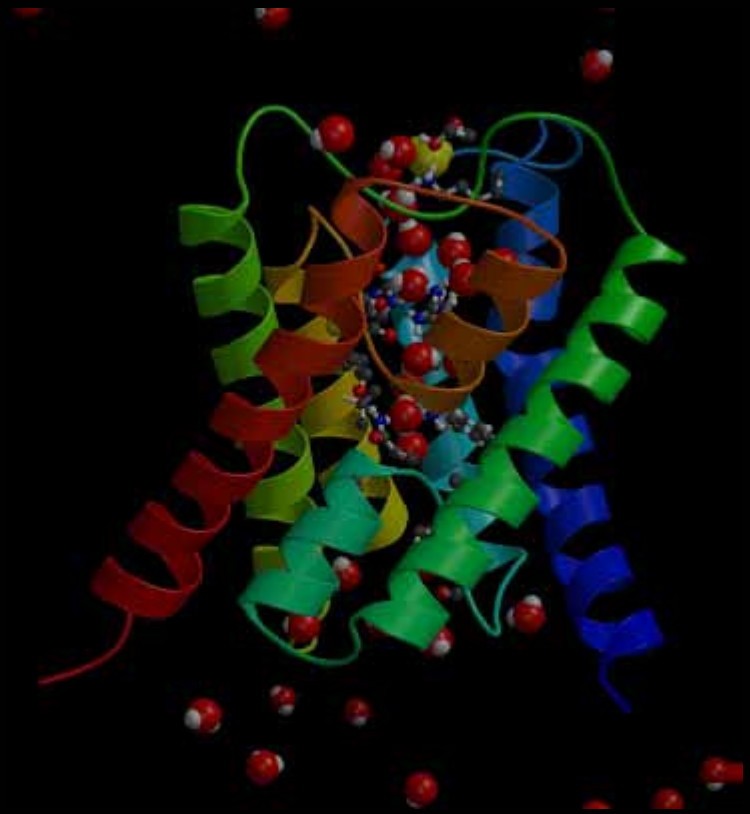


From G. Tiana

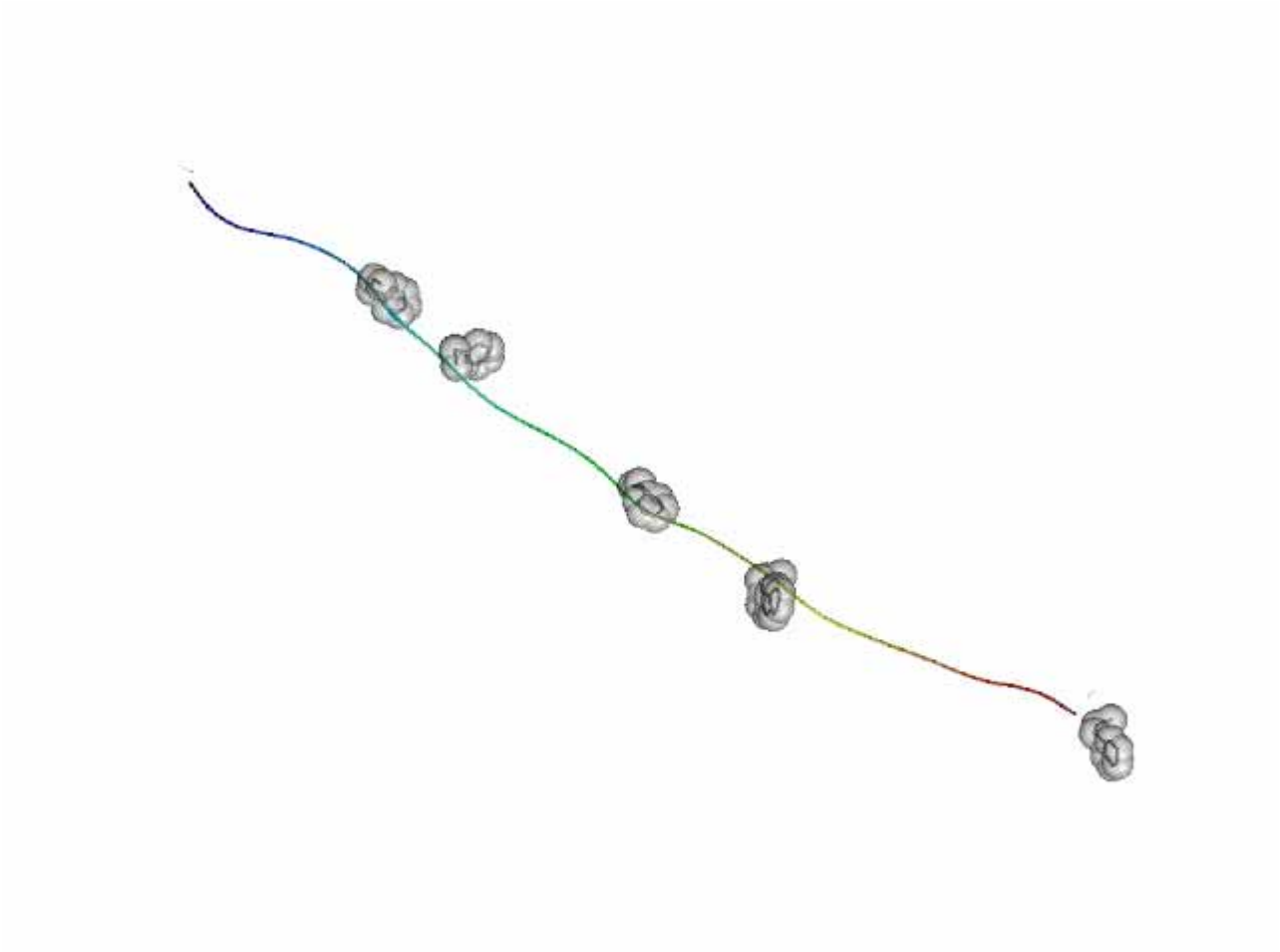
All-Atom MD Simulations of Complex Systems. Aquaporin translocation of water.



From B. De Groot; Science 2001



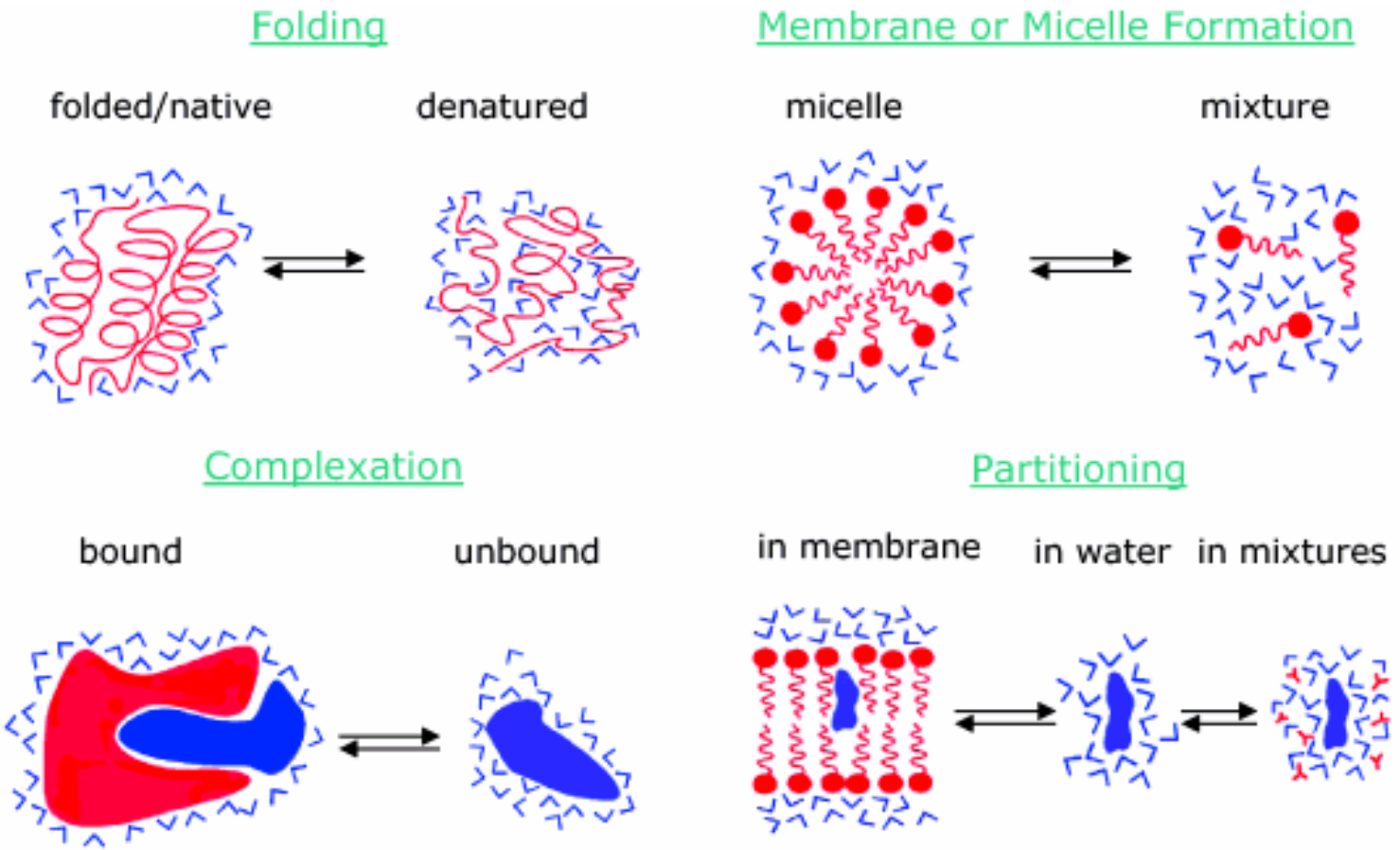
All-Atom MD Simulations of Complex Systems. Folding of a Small Protein.



From V. Pande

All-Atom MD Simulations of Complex Systems.

Phenomena Governed by Weak Intermolecular Forces



From WF van Gunsteren; Angew. Chem. 2006

Simulation of Complex Biological Systems

These processes are best described at the atomic or molecular level.
They occur close to room Temperature (~300K)
Energies involved are in the range of 1-10 $k_B T$

The processes are largely determined by the laws of Statistical Mechanics

A Number of experiments can be understood only in terms of ensembles of configurations.
Examples: NOE's, CD spectra….

Simulation of Complex Biological Systems

Chemical and Biological systems are generally too
inhomogeneous and complex
to be treated analytically

We need:

Numerical simulations of the behavior of the system to

Produce a statistical ensemble of configurations
representing the state of the system

Analyze the results using statistical mechanics

Statistical Mechanics

- Theoretical basis for derivation of macroscopic behaviors from microscopic origins
- Two fundamental postulates of equilibrium statistical mechanics
  - *microstates of equal energy are equally likely*
  - *time average is equivalent to ensemble average*
- Formalism extends postulates to more useful situations
  - *thermal, mechanical, and/or chemical equilibrium with reservoirs*
    - *systems at constant T, P, and/or m*
  - yields new formulas for probabilities of microstates
    - *derivation invokes thermodynamic limit of very large system*
- Macroscopic observables given as a weighted sum over microstates
  - *dynamic properties require additional formalism*

Ensembles

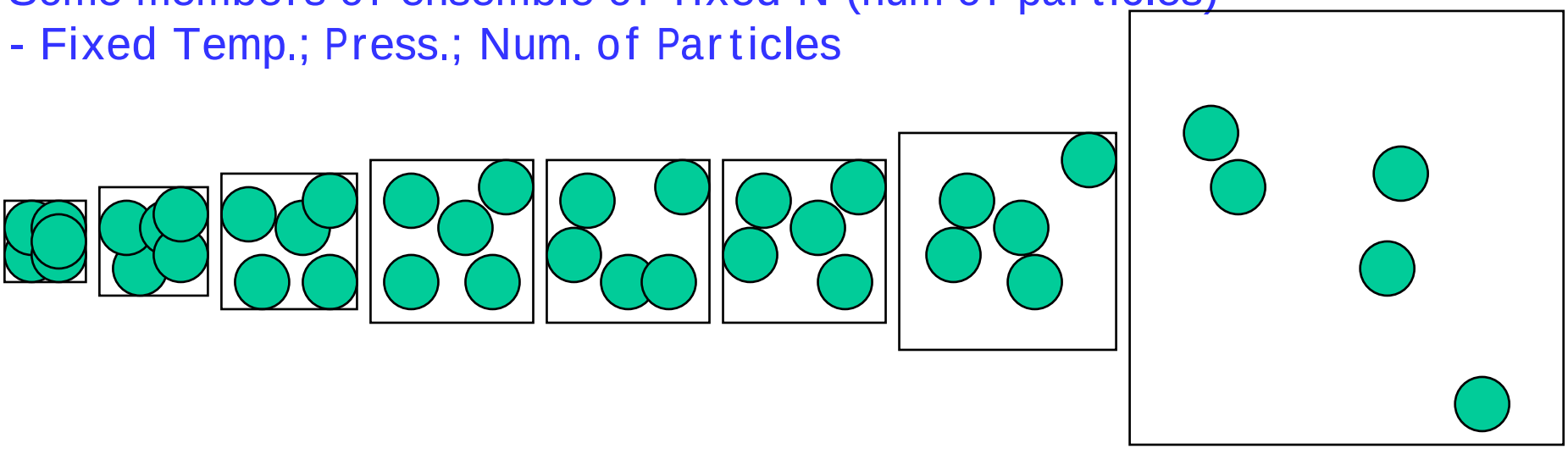
Definition of an ***ensemble***

- *collection of microstates subject to at least one extensive constraint*
  Microstate is specification of all atom positions and momenta
  Fixed total energy, total volume, and/or total number of molecules
  unconstrained extensive quantities are represented by full value range

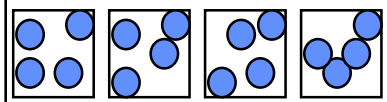
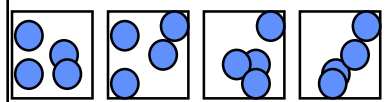
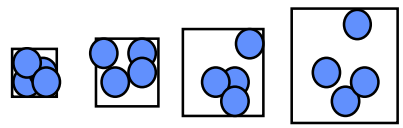
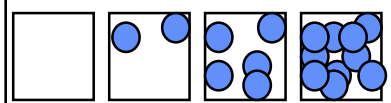
- *Probability distribution π describing the likelihood of observing each state, or the weight that each state has in ensemble average*

Example:

Some members of ensemble of fixed N (num of particles)
- Fixed Temp.; Press.; Num. of Particles



Commonly Encountered Ensembles

Name	All states of:	Probability distribution	Schematic
Microcanonical (EVN)	given EVN	$\pi_i = \frac{1}{\Omega}$	
Canonical (TVN)	all energies	$\pi(E_i) = \frac{1}{Q} e^{-\beta E_i}$	
Isothermal-isobaric (TPN)	all energies and volumes	$\pi(E_i, V_i) = \frac{1}{\Delta} e^{-\beta(E_i + PV_i)}$	
Grand-canonical (TVμ)	all energies and molecule numbers	$\pi(E_i, N_i) = \frac{1}{\Xi} e^{-\beta(E_i + \mu N_i)}$	

Note: $\beta \equiv 1/kT$

Ensemble and Time Averaging

- Configuration given by all positions and momenta
  - "phase space" $\Gamma = (p^N, r^N)$ r^N shorthand for "positions of all N atoms"
- Configuration variable $A(r^N, p^N)$
- Ensemble average
  - Weighted sum over all members of ensemble
  - In general $\langle A \rangle = \sum A_i \pi_i$
  - For example, canonical ensemble, classical mechanics:

$$\langle A \rangle = \frac{1}{Q} \frac{1}{h^{3N} N!} \int dp^N \int dr^N A(p^N, r^N) e^{-\beta E(p^N, r^N)}$$

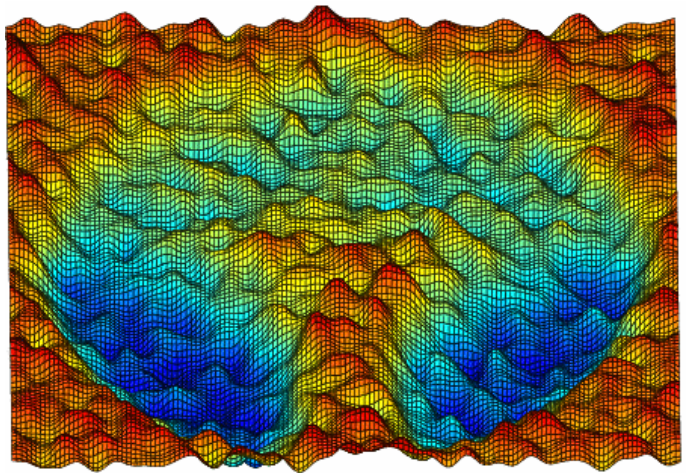
- Time average
  - Sum over all states encountered in dynamical trajectory of system

$$\bar{A} = \lim_{t \to \infty} \frac{1}{t} \int_0^t A\Big(\underbrace{p^N(t), r^N(t)}; \overbrace{p^N(0), r^N(0)} \Big) dt'$$

Should average over initial conditions

Given by equations of motion

Statistical Mechanics and Simulations



The state of a biomolecular system cannot be described in terms of one single minimum, but by a statistical mechanical ensemble of configurations.

This holds for most experiments, too. NOE's, CD spectra….

The weight of a configuration ***x*** *is given by the Boltzmann Factor*

$$P(x) \sim \exp(-V(x)/k_B T)$$

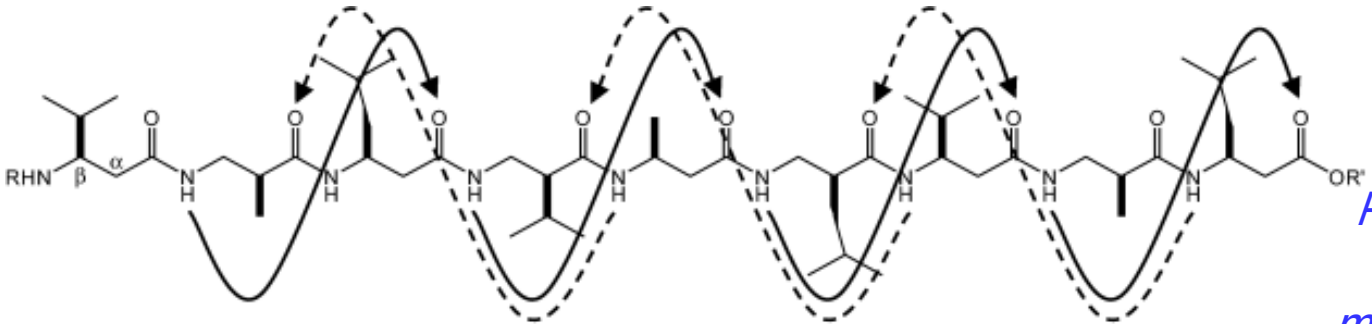
The exponential weighting implies that high energy regions will not contribute configurations that are significant to the state of the system, unless they are numerous (entropy).

Equilibrium properties are dominated by the parts of configurational space for which ***V(x)*** *is low.*

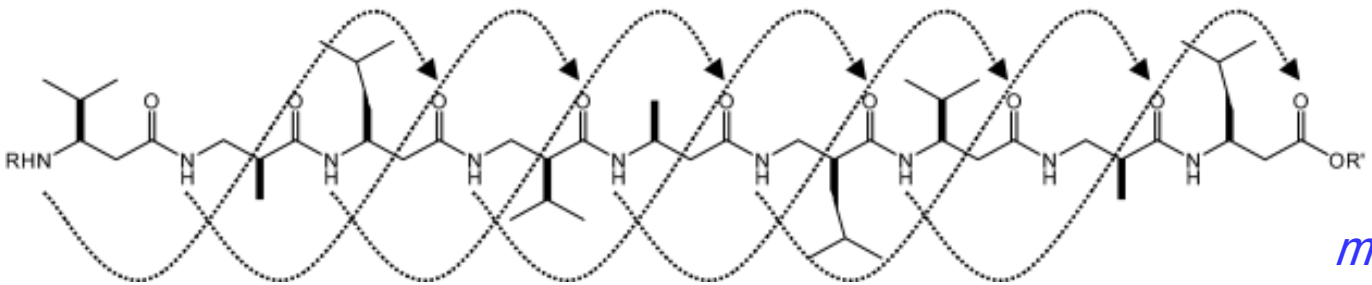
Therefore, one should search the low energy regions of the vast biomolecular energy surface.

Structural Ensembles for Peptides

One set of NOE constraints. Two structures needed to satisfy all of them. Found with MD simulations.



12/10 Helix
Presence of 10 (solid) and 12 (dashed) membered h-bond rings

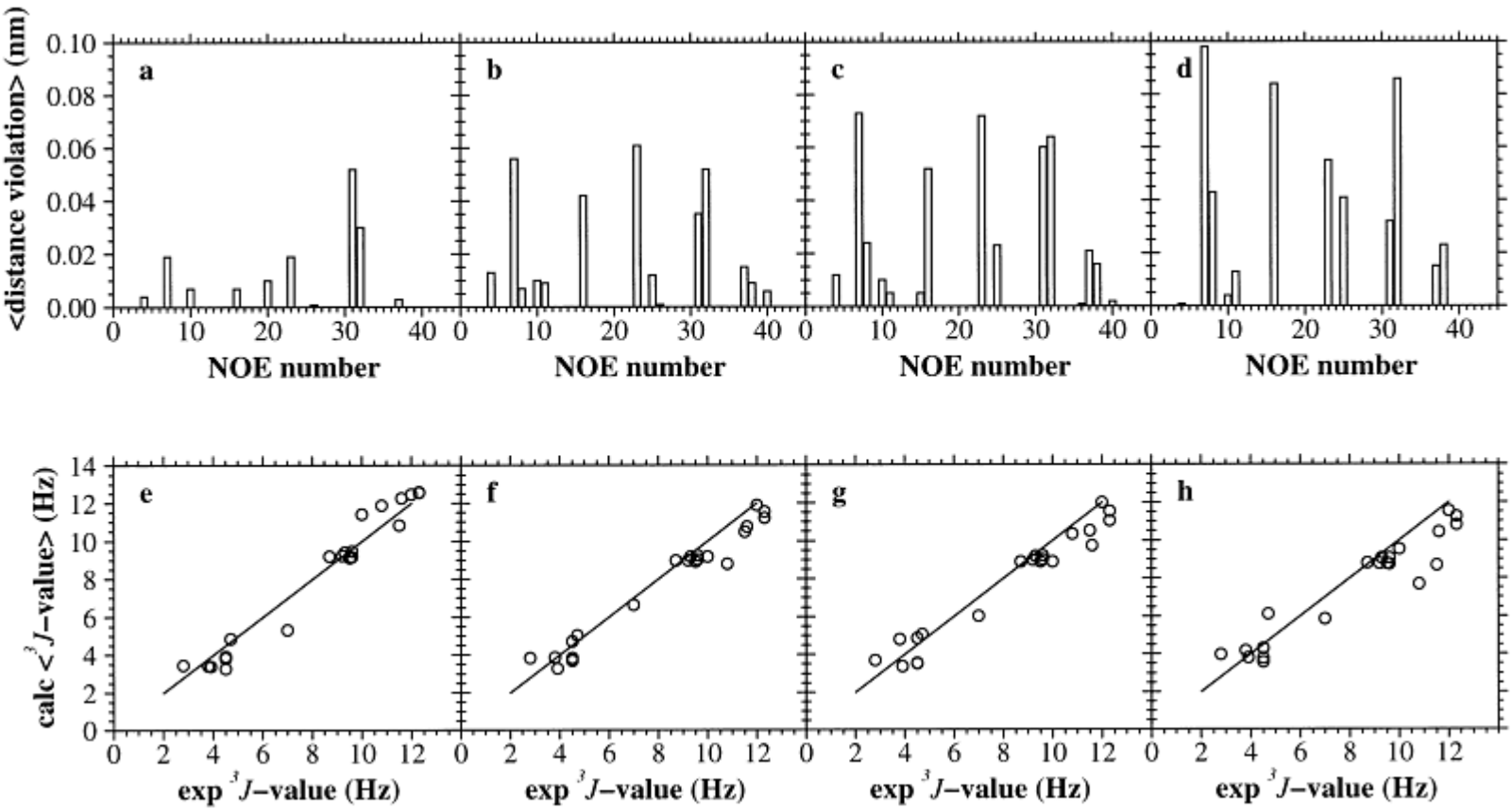


3/14 Helix
Presence of 14 membered h-bond rings

Daura; Proteins 1999

Violations of the 42 average inter-proton distances inferred from the NMR data at 298 K by interproton distances averaged from 50ns MD simulations at different temperatures

298 K *340 K* *350 K* *360K*

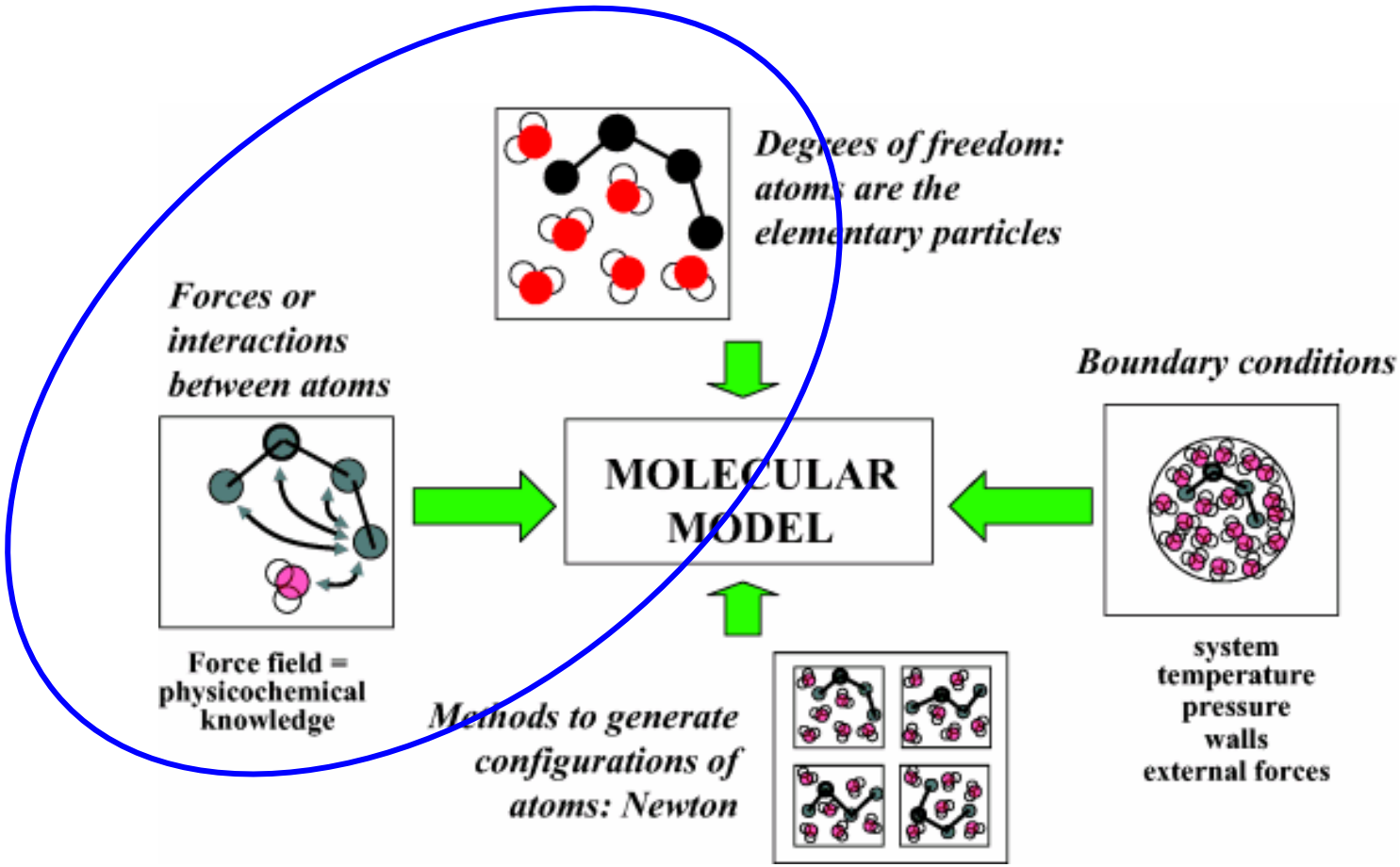


Not one single minimum.
Know your problem!!!!

Daura; Proteins 1999

Take Home Lesson

Basic Choices in model definition for molecular simulation



Generate an ensemble of configurations by computer simulations and average

Reasons for Using Simulations

Conditions inaccessible to experiments.

Examples:

Extreme Conditions
- Temperature
- Pressure
- Cost
- Time
- Morality

Or

Need more than the experiment can give.

Examples:

Atomic Detail (dynamics)
- NMR and X-ray average properties
- Pressure

Test models of real world

How We Use Computer Simulations

to study the properties of molecular systems in terms of atoms, to describe and predict:

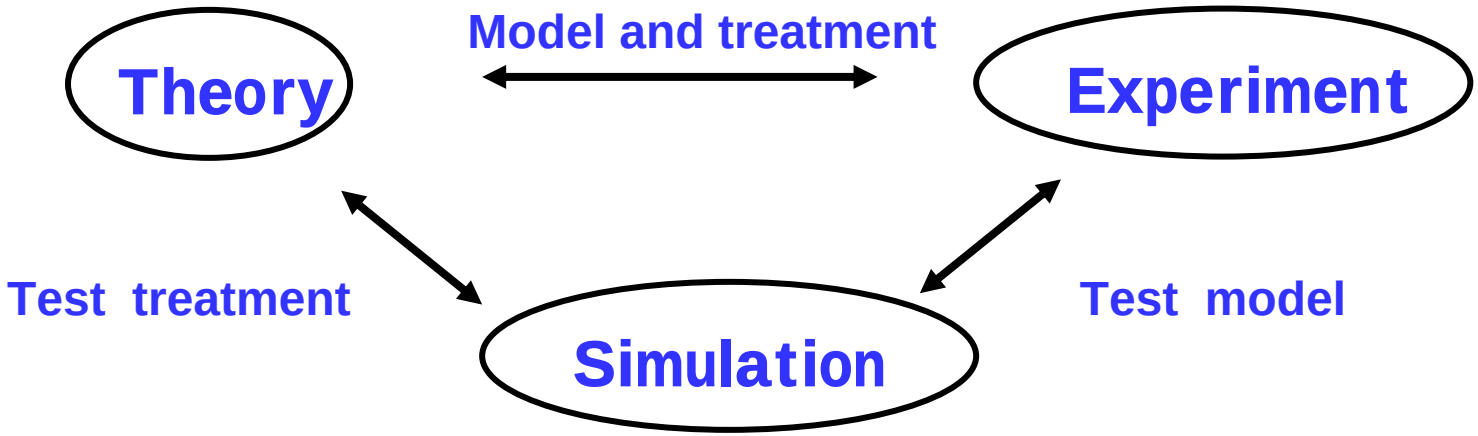
1) *Anti-tumor Peptide design. (JBC 2002; Biochemistry 2003; Cancer Cell 2005)*

2) *The structure-stability and solvent effects on biomolecular systems; identification of hot spots. (Prot. Sci. 2004; PNAS 2002; Proteins 2005)*

3) *Molecular Recognition. (Chemistry 2003; Carb. Res. 2004)*

4) *Processes inaccessible to experiments. Folding & Spontaneous aggregation of peptides. (Proteins 2004, 2005; JPCB 2004; J.Chem.Phys 2004; JMB 2005)*

Simulation of complex biological systems



The many particle problem

	Crystalline solid state	Liquid state macromolecules	Gas phase
Quantum N^4	possible	still impossible	possible
Classical $NlnN$	easy	computer simulations	trivial

Many particle system

Molecular Dynamics

Generating Atomic Resolution Information on the Dynamics of Biomolecules

Molecular Dynamics

Generates the ensemble of configurations via application of Newton's laws of motion to the atoms of the system

Advantage:
dynamical information about the system,
Distributions,
Time Series.

Disadvantage:
Not efficient at crossing high energy barriers

Methodology

A typical biomolecular force field consists of potential energy terms representing covalent and nonbonded interactions

Methodology

A typical force field or effective potential for a system of N atoms with masses m_i (i=1,2.....N) and cartesian position vectors r_i:

$$V(r_1, r_2, ...r_N) = \sum_{bonds} \frac{1}{2} K_b [b - b_0]^2 + \sum_{angles} \frac{1}{2} K_\theta [\theta - \theta_0]^2 + \sum_{\substack{improp \\ dihedrals}} \frac{1}{2} K_\zeta [\zeta - \zeta_0]^2 +$$

$$+ \sum_{dihedrals} K_\varphi [1 + \cos(n\varphi - \delta)] + \sum_{pairs(i,j)} \left[C_{12}(i,j) / r_{ij}^{12} - C_6(i,j) / r_{ij}^6 + q_i q_j / (4\pi\varepsilon_0 \varepsilon_r r_{ij}) \right]$$

$F_i = - dV(r_1, r_2, ...r_N) / dr_i$ $\qquad$ $d^2 r_i(t) / dt^2 = F_i / m_i$

$r_i(t)$ atom i position at time t as a function of other atoms

The integration is performed in small time-steps 1-10 fs

Easy Functional Form.

Transferable Parameters.

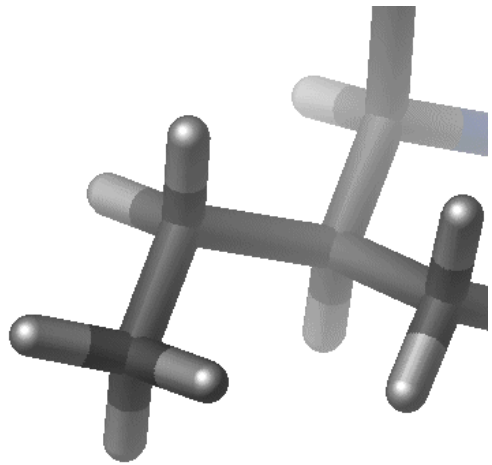
Millions of Steps - Computationally demanding!!!

Methodology:

Terms of the potential function

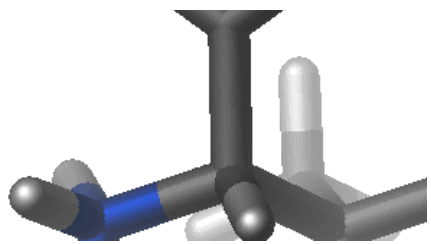
Bond term

$$\sum_{bonds} \frac{1}{2} K_b [b - b_0]^2$$



Angle term

$$\sum_{angles} \frac{1}{2} K_\theta [\theta - \theta_0]^2$$



Improper term

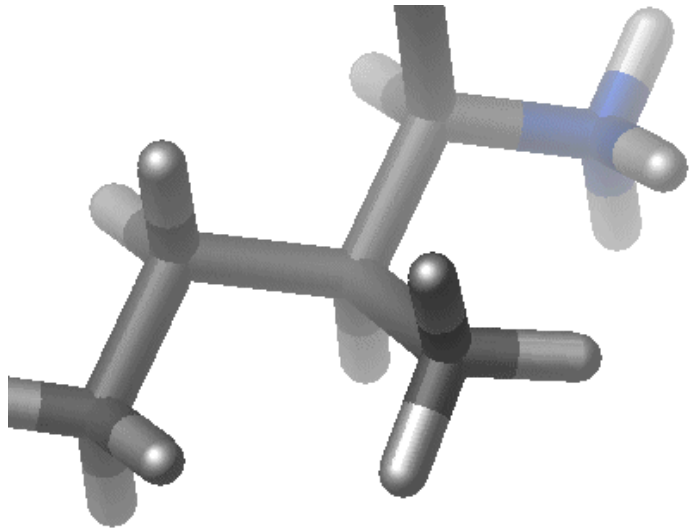
$$\sum_{\substack{improp \\ dihedrals}} \frac{1}{2} K_\zeta [\zeta - \zeta_0]^2$$

Methodology:

Terms of the potential function

Dihedral term

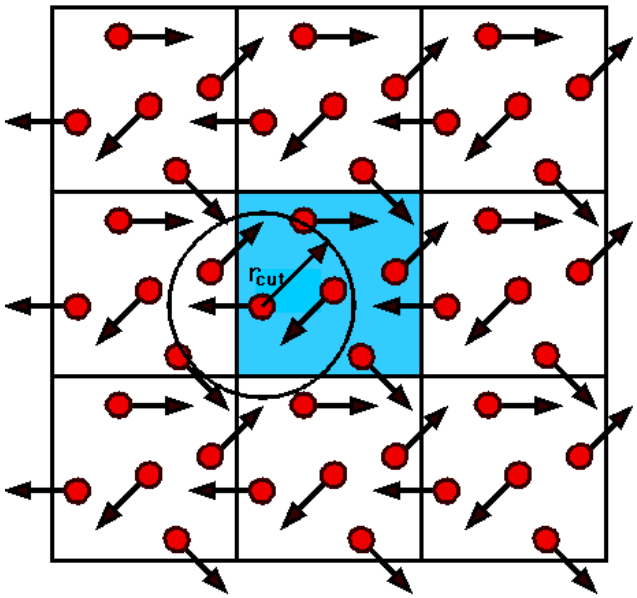
$$\sum_{dihedrals} K_{\varphi}\left[1+\cos(n\varphi-\delta)\right]$$



Non-Bonded term

$$\sum_{pairs(i,j)} \left[C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^{6} + q_i q_j/(4\pi\varepsilon_0\varepsilon_r r_{ij})\right]$$

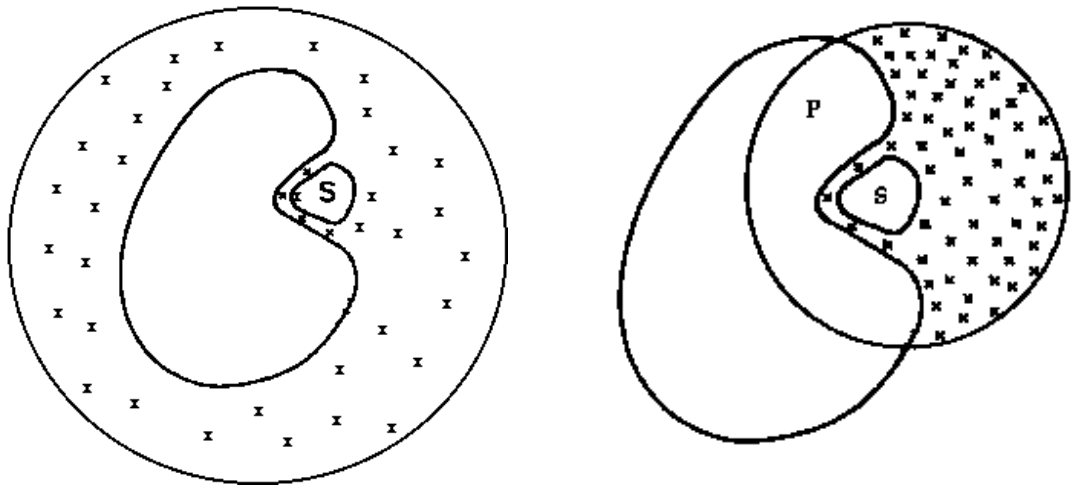
Molecular Dynamics: Explicit Solvation



Periodic Box

advantages: avoids boundary effects; mimics infinite environment; consistent treatment of long-range interactions possible.

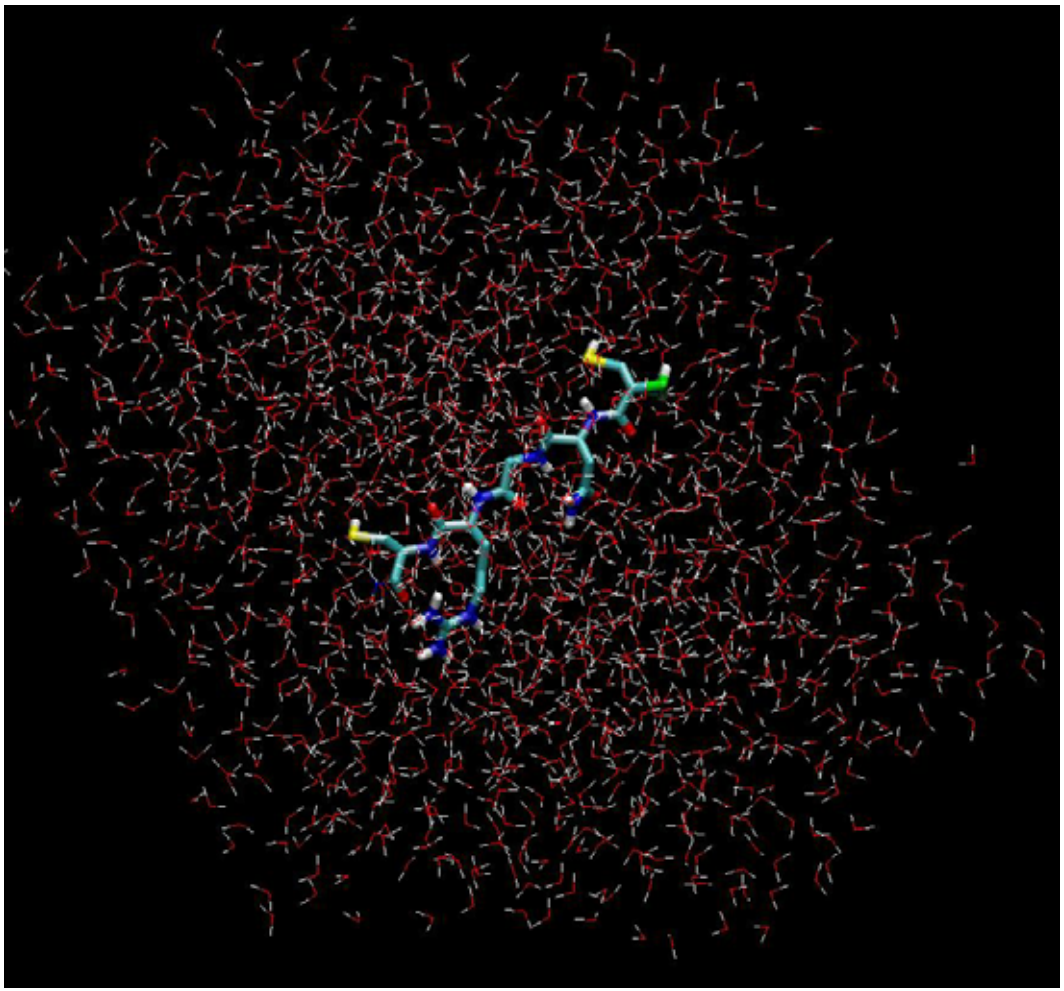
disadvantages: periodicity is artefact, effect must be evaluated (especially Coulombic artefacts).



Shell Solvation

advantages: simple; solvent can be limited to shell or
omitted and replaced by effective boundary potential.

disadvantages: strong boundary effects (outer layer must be discarded); very wrong electrostatic interactions unless reaction field is imposed.



Methodology: Nonbonded Forces

$$\sum_{pairs(i,j)} \left[C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^{6} + q_i q_j /(4\pi\varepsilon_0\varepsilon_r r_{ij}) \right]$$

Fast Decay **Slow Decay**

Nonbonded interactions govern the thermodynamics of processes described before.
Appropriate description of nonbonded forces is fundamental.

Energy difference driving the processes are of the order of tens kJ/mol.
Small amount derived from a summation over many atom pairs.

If N=1000, then we have ~1/2N(N-1) = 500 000 pairs of atomic interactions that yield a small difference. Need for accurate description of Nonbonded forces.
The bigger N becomes, the worse the situation.

Methodology: treatment of electrostatics

$$\sum_{pairs(i,j)} \left[C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^{6} + q_i q_j /(4\pi\varepsilon_0\varepsilon_r r_{ij}) \right]$$

Fast Decay

Slow Decay

The sums in this term run over all atom pairs in molecular systems, and it is proportional to N^2. All the other parts of the calculation are proportional to N.

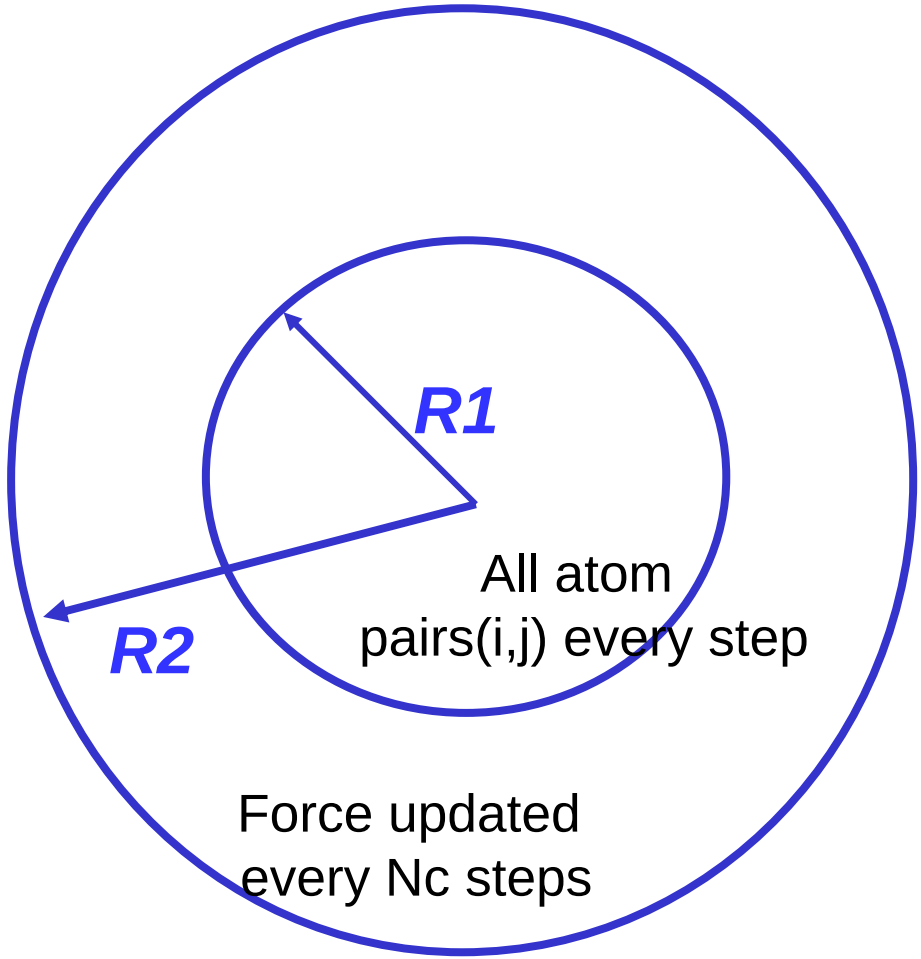
Several approximations-solutions:

1) cutoff methods

2) continuum methods

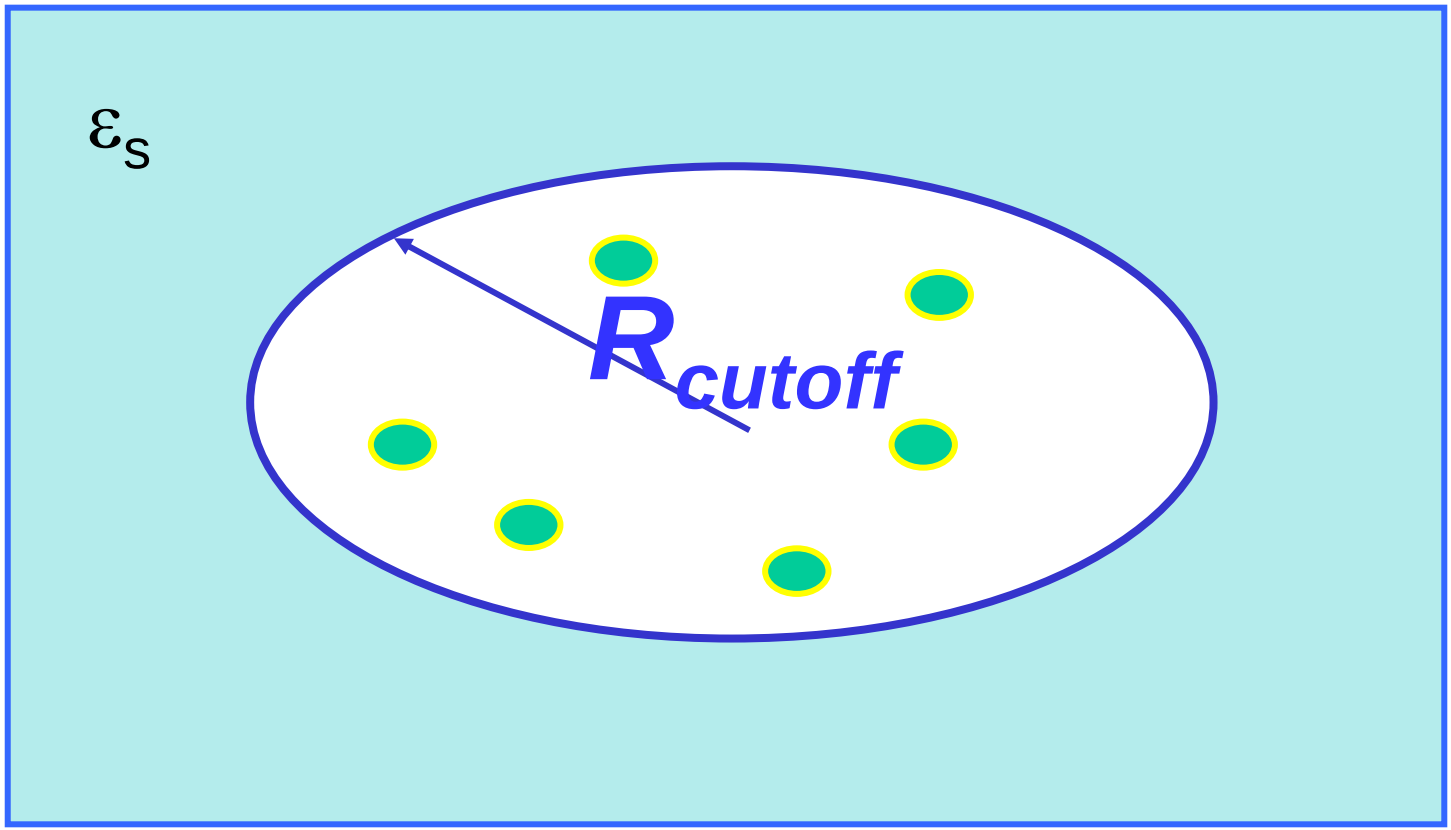
3) Periodic methods

Methodology: *Electrostatic Cut-offs*



Disadvantage: Force Truncation

Methodology:
Electrostatic Cut-offs + Reaction Field correction



$$E_i = \frac{2(\varepsilon - 1)}{\varepsilon_s + 1}\left(\frac{1}{r_c^3}\right)\sum_{j;r_{ij}\le r_c}\mu_j$$

μ_j are the dipoles of the neighboring molecules that are within the cutoff distance r_c of the molecule i. The interaction between the molecule i and the reaction field equals $E_i\mu_i$ which is added to the molecule-molecule interaction within the cutoff range.

Methodology: treatment of electrostatics - Continuum methods

If one part of the system is homogeneous, like the solvent around the solute, the homogeneous part can be considered a continuum.

The system is divided in two parts:

1) an inner region where charges q_i are explicititly treated

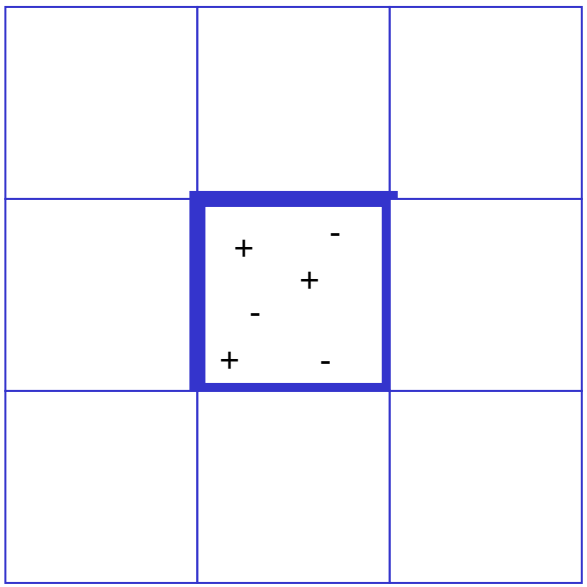
2) an outer region treated as a continuum with dielectric constant e

Poisson-Boltzmann Equation:

$$\nabla^2\varphi(r) = k^2\varphi(r)$$

Disadvantage: No explicit solvent molecule, which may be important for the phenomenon under exam

Methodology:
treatment of electrostatics-Periodic methods



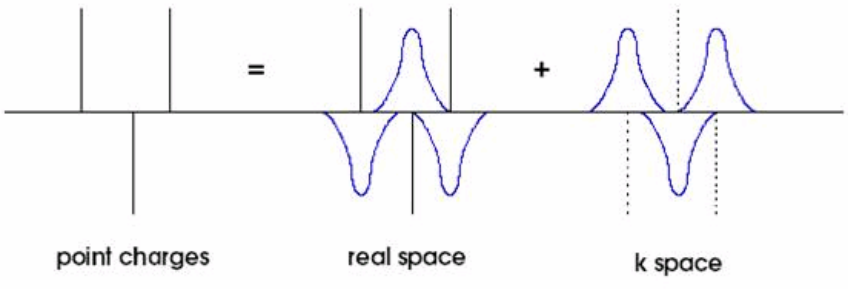
The system is replicated infinitely.
The charge distribution in the system is represented as delta functions

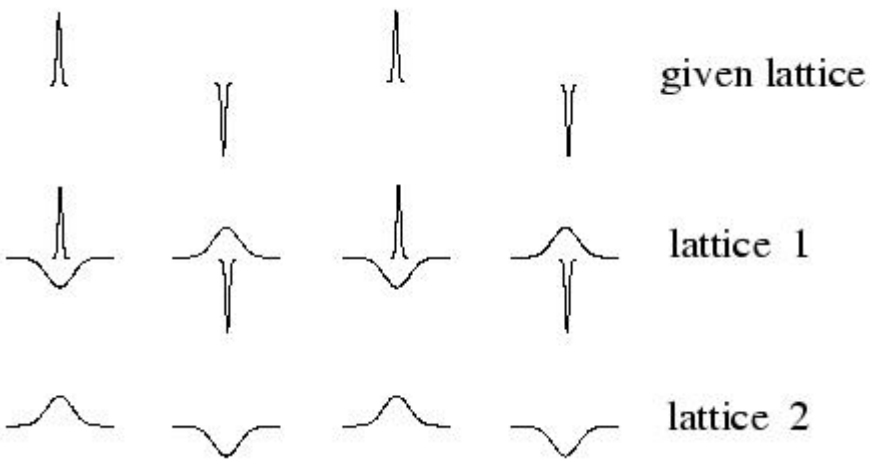
Each point charge is surrounded by a gaussian charge of opposite sign

The charge interactions become short-ranged.
An error function is used to recover the original distribution

Disadvantage: Artificial periodicity, computationally expensive

Methodology:
treatment of electrostatics-Periodic methods (PME)





The system is replicated infinitely. The charge distribution in the system is represented as delta functions

Each point charge is surrounded by a gaussian charge of opposite sign

The charge interactions become short-ranged. An error function is used to recover the original distribution

4.6 Long Range Electrostatics

4.6.1 Ewald summation

The total electrostatic energy of N particles and the periodic images are given by

$$V = \frac{f}{2} \sum_{n_x} \sum_{n_y} \sum_{n_z *} \sum_{i}^{N} \sum_{j}^{N} \frac{q_i q_j}{\mathbf{r}_{ij,\mathbf{n}}}. \tag{4.122}$$

$(n_x, n_y, n_z) = \mathbf{n}$ is the box index vector, and the star indicates that terms with $i = j$ should be omitted when $(n_x, n_y, n_z) = (0, 0, 0)$. The distance $\mathbf{r}_{ij,\mathbf{n}}$ is the real distance between the charges and not the minimum-image. This sum is conditionally convergent, but very slow.

Ewald summation was first introduced as a method to calculate long-range interactions of the periodic images in crystals [54]. The idea is to convert the single slowly-converging sum eqn. 4.122 into two quickly-converging terms and a constant term:

$$V = V_{dir} + V_{rec} + V_0 \tag{4.123}$$

$$V_{dir} = \frac{f}{2} \sum_{i,j}^{N} \sum_{n_x} \sum_{n_y} \sum_{n_z *} q_i q_j \frac{\operatorname{erfc}(\beta r_{ij,\mathbf{n}})}{r_{ij,\mathbf{n}}} \tag{4.124}$$

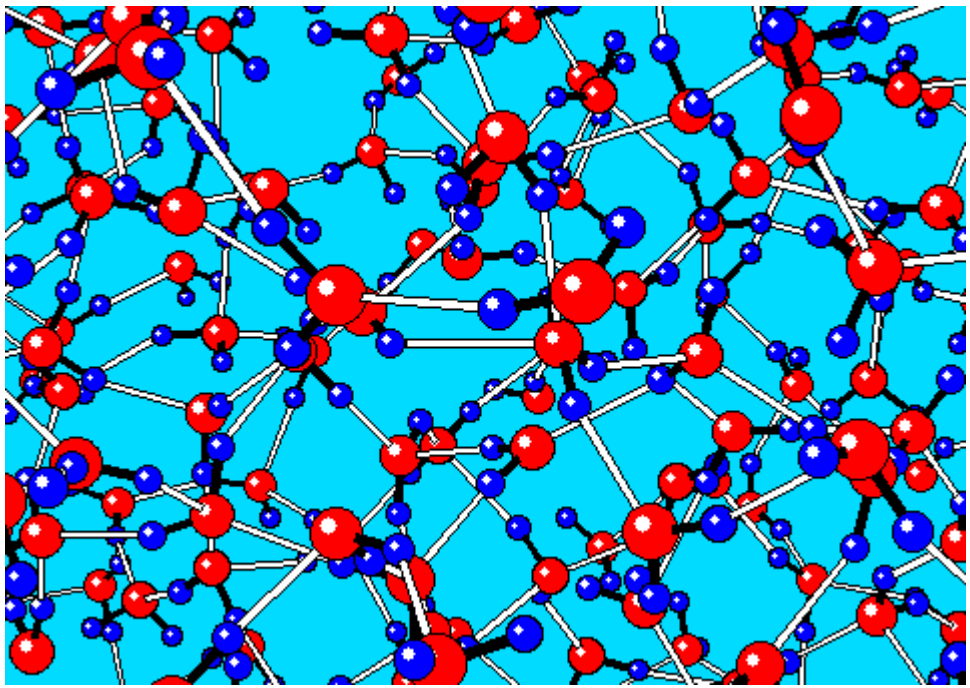
$$V_{rec} = \frac{f}{2\pi V} \sum_{i,j}^{N} q_i q_j \sum_{m_x} \sum_{m_y} \sum_{m_z *} \frac{\exp\left(-(\pi \mathbf{m}/\beta)^2 + 2\pi i \mathbf{m} \cdot (\mathbf{r}_i - \mathbf{r}_j)\right)}{\mathbf{m}^2} \tag{4.125}$$

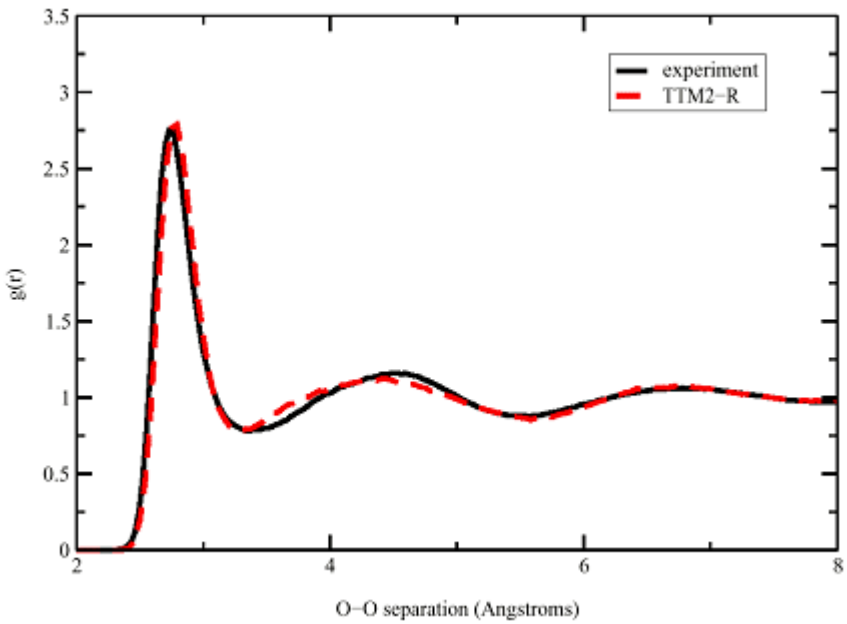
$$V_0 = -\frac{f\beta}{\sqrt{\pi}} \sum_{i}^{N} q_i^2, \tag{4.126}$$

where β is a parameter that determines the relative weight of the direct and reciprocal sums and $\mathbf{m} = (m_x, m_y, m_z)$. In this way we can use a short cutoff (of the order of 1 nm) in the direct space sum and a short cutoff in the reciprocal space sum (*e.g.* 10 wave vectors in each direction). Unfortunately, the computational cost of the reciprocal part of the sum increases as N^2 (or $N^{3/2}$ with a slightly better algorithm) and it is therefore not realistic for use in large systems.

Methodology: Nonbonded Forces

$$\sum_{pairs(i,j)} \left[C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^{6} + q_i q_j /(4\pi\varepsilon_0\varepsilon_r r_{ij}) \right]$$

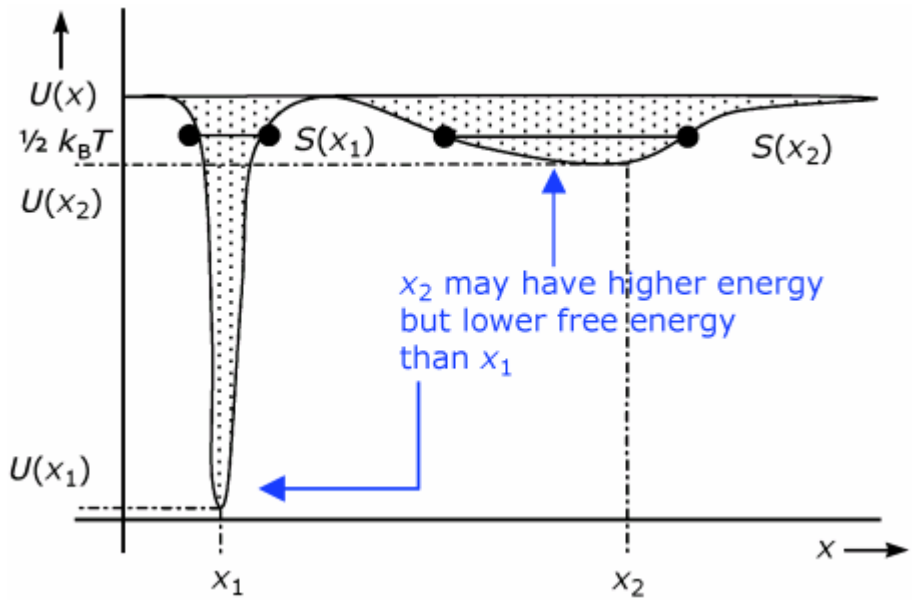




Example: The properties of liquid water are largely dependent on intermolecular interactions

Methodology: Nonbonded Forces

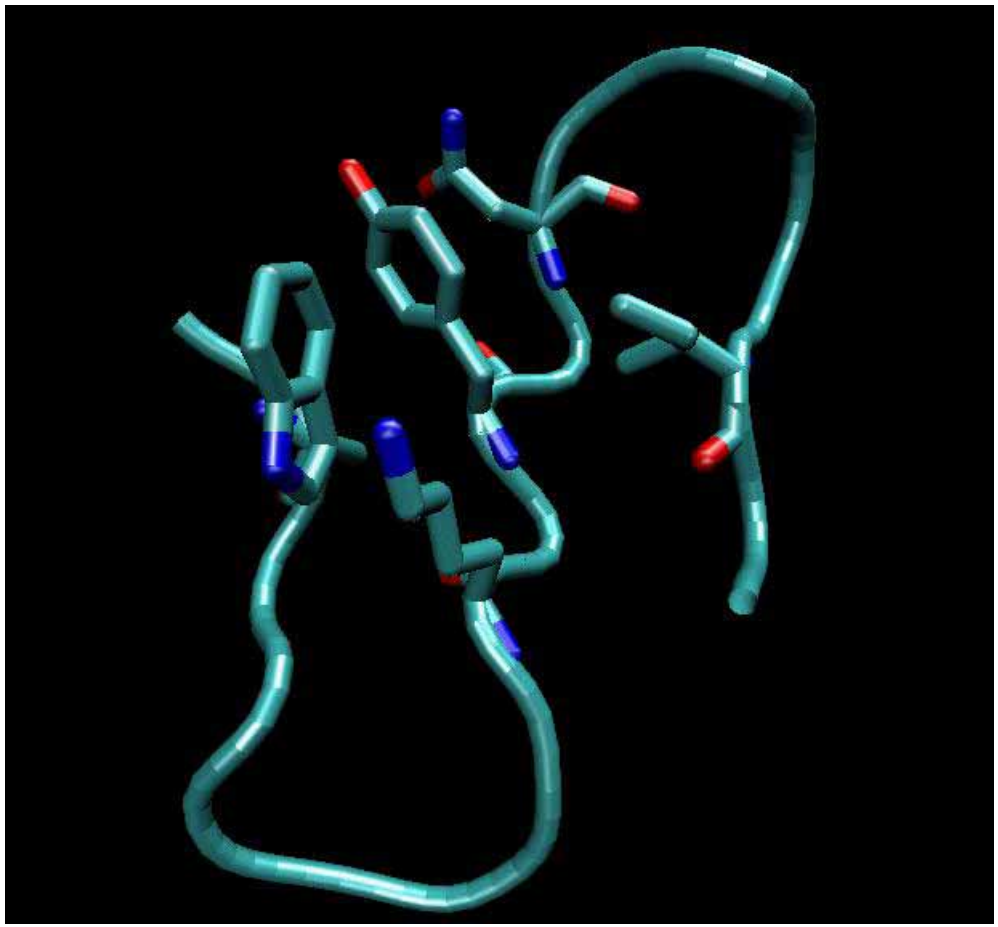
$$\sum_{pairs(i,j)} \left[C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^{6} + q_i q_j /(4\pi\varepsilon_0\varepsilon_r r_{ij}) \right]$$

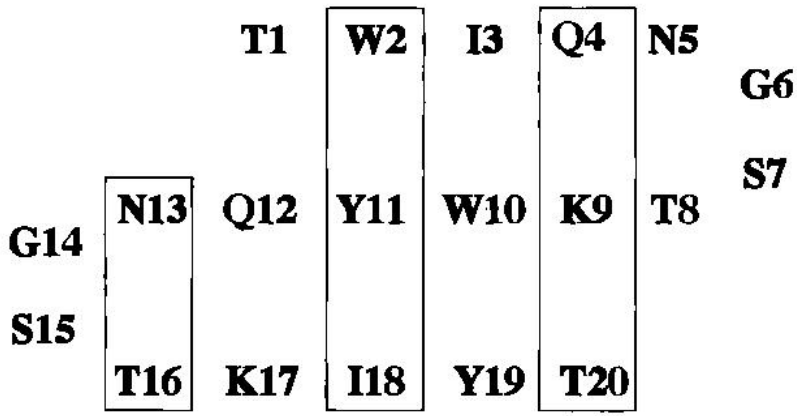


Example: Proper Evaluation of Entropy

Three Stranded Peptide model

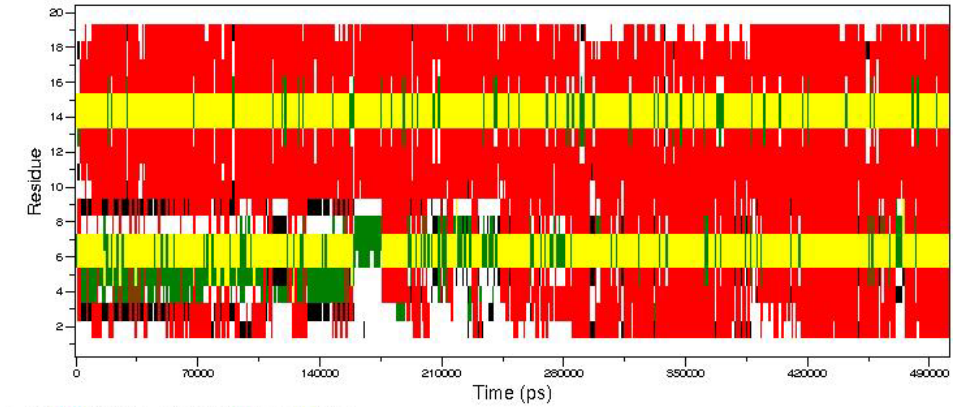
De Alba (Prot. Sci. 1999)



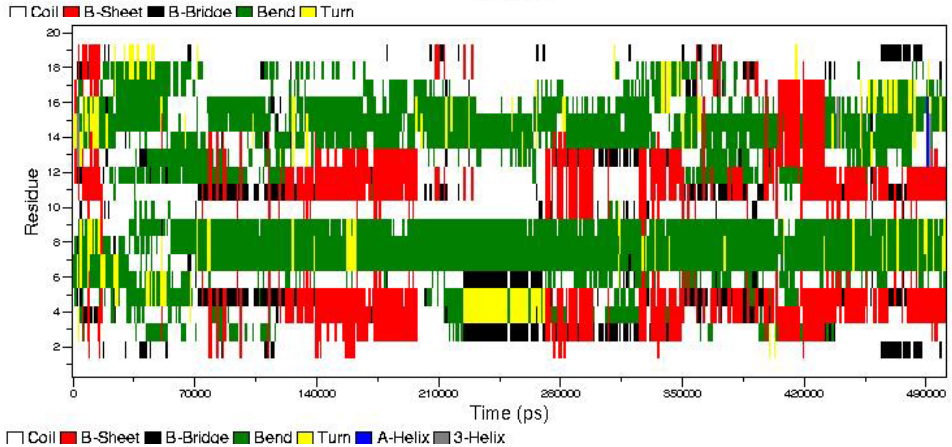


The "De Alba" peptide shows higher stability than betanova. 30-50 % three stranded b-sheet is present in solution (NMR det.)
Cooperative folding-unfolding.

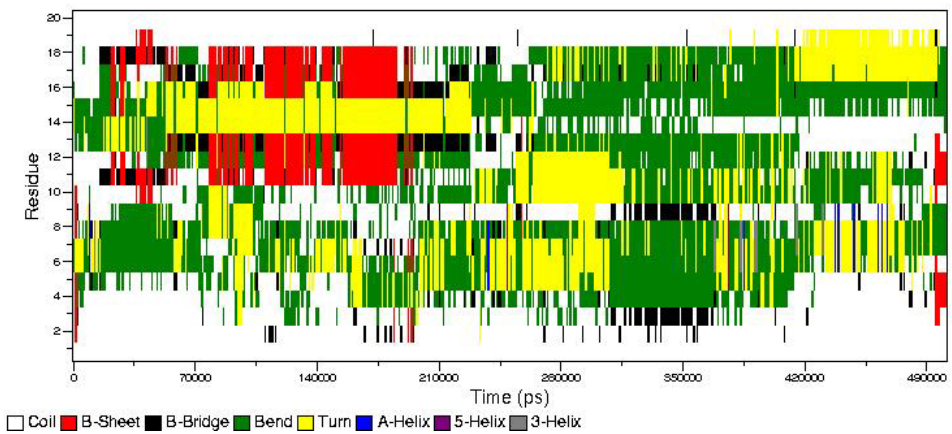
PME vs. Cutoff methods: Sec.-Struct. Evolution



Simulation 1
PME + Ions



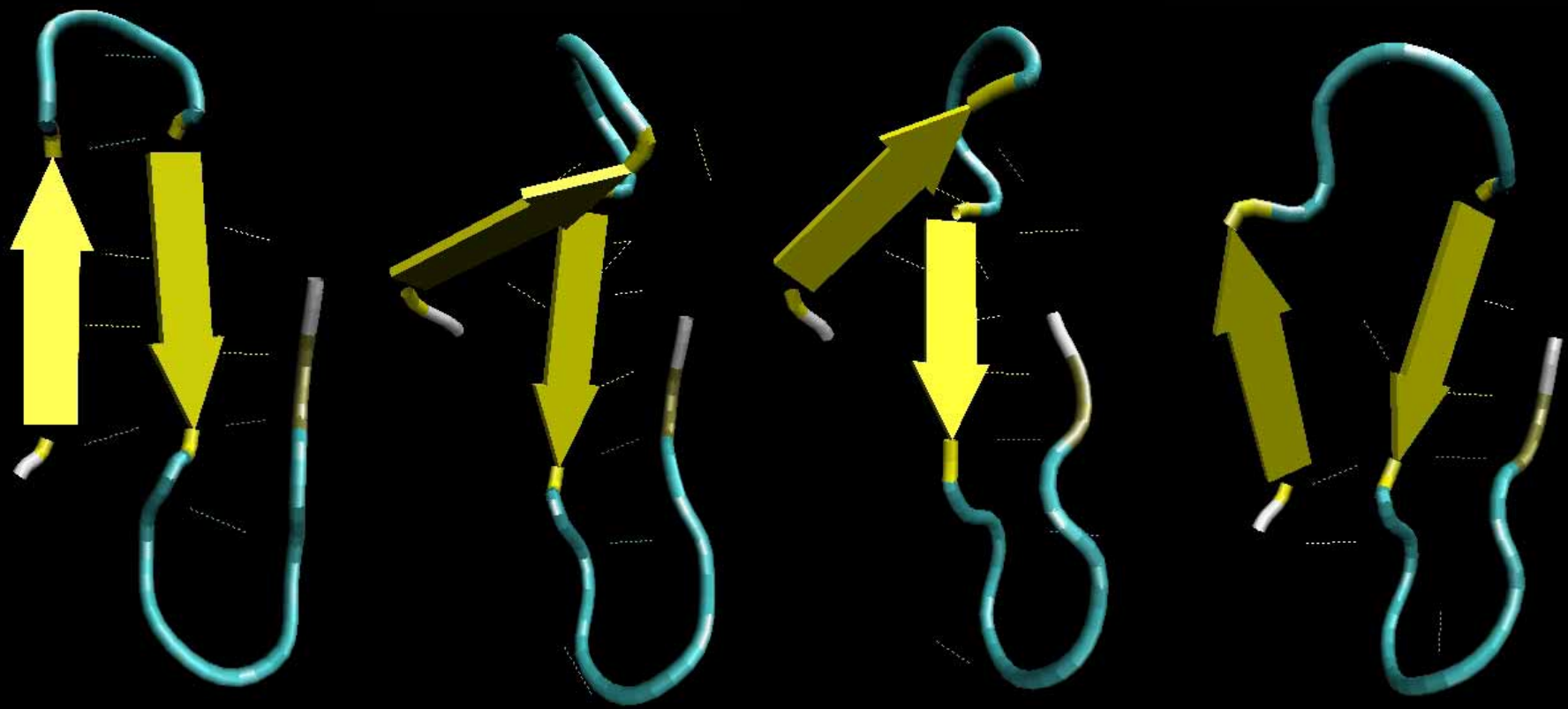
Simulation 2
Cutoff + NO Ions



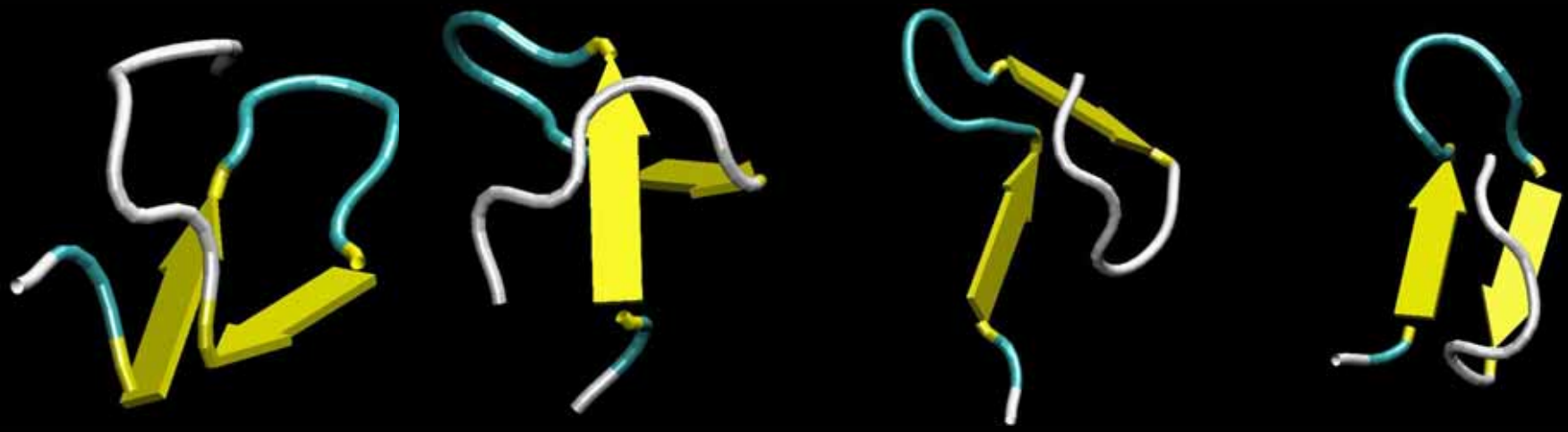
Simulation 3
Cutoff + Ions

Central structure of the 4 most populated clusters (over the whole simulation time). Simulation 1

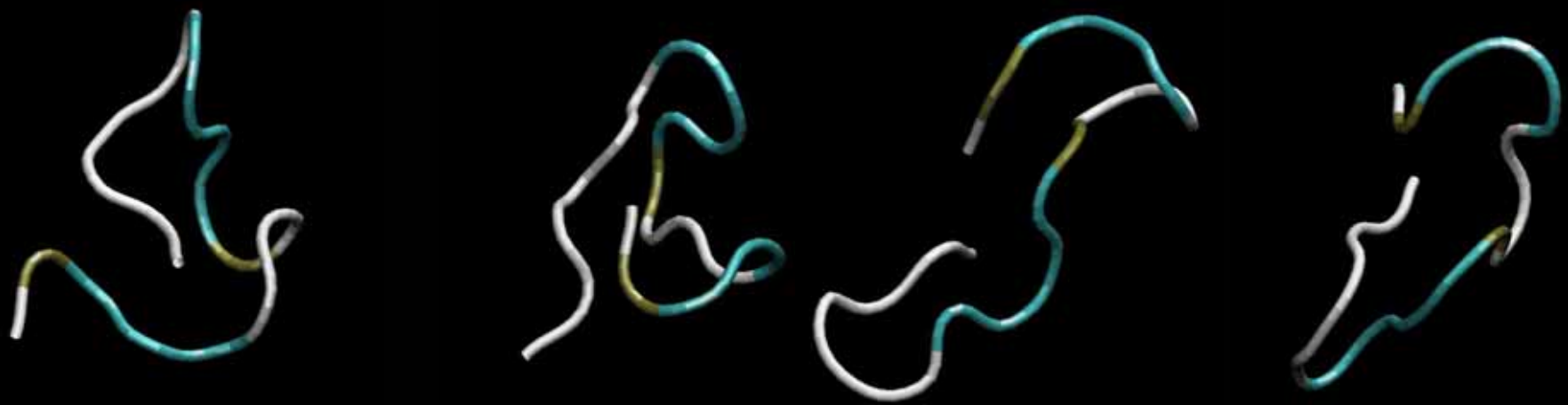
32% 26% 12% 10%



Central structure of the 4 most populated clusters (over the whole simulation time). Simulation 2

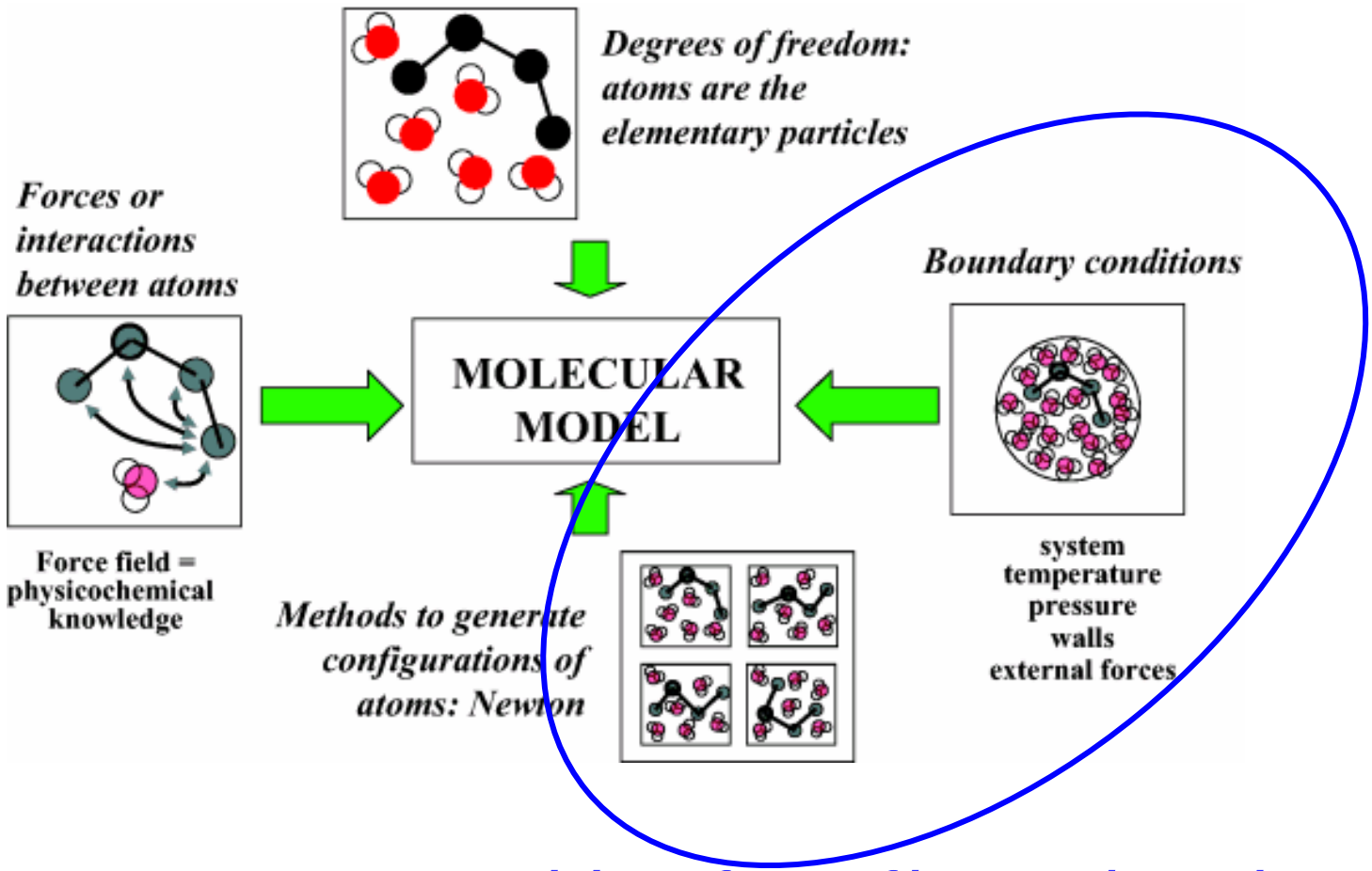


Central structure of the 4 most populated clusters (over the whole simulation time). Simulation 3



Take Home Lesson

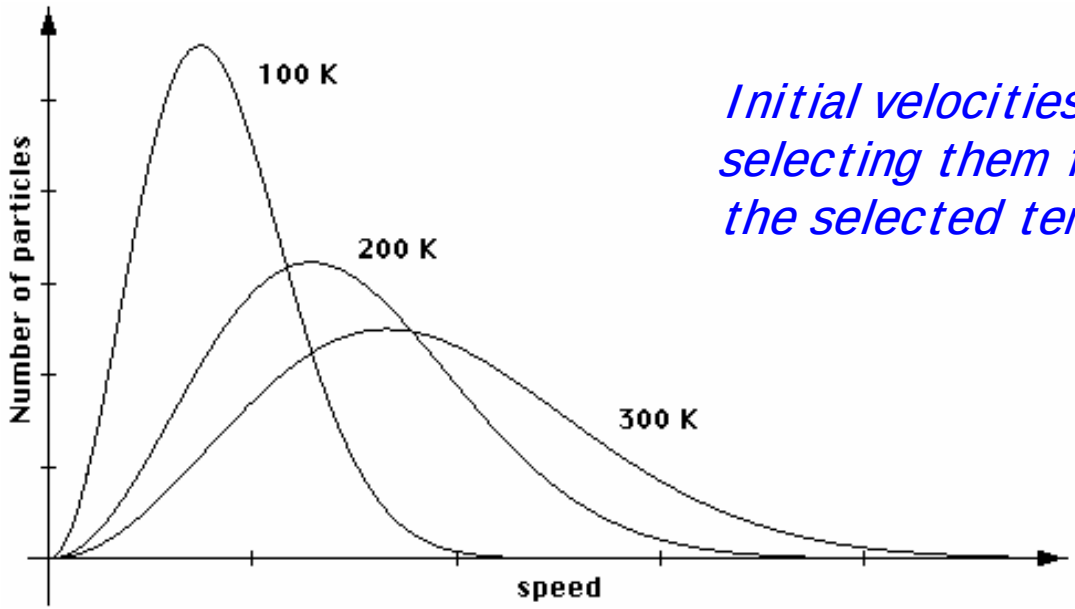
Basic Choices in model definition for Molecular Dynamics



Generate an ensemble of configurations by computer simulations and average

Molecular Dynamics: Starting and Running Simulations

Molecular Dynamics: Starting the simulation



Initial velocities are imposed on each atom by selecting them from a maxwellian distribution at the selected temperature.

Molecular Dynamics: Integrating the Equations of Motion

$F_i = - dV(r_1, r_2, ...r_N) / dr_i$ $\qquad$ $d^2 r_i(t) / dt^2 = F_i / m_i$

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\dddot{r} + O(\Delta t^4)$$

$$r(t-\Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}\dddot{r} + O(\Delta t^4)$$

Summing the two members

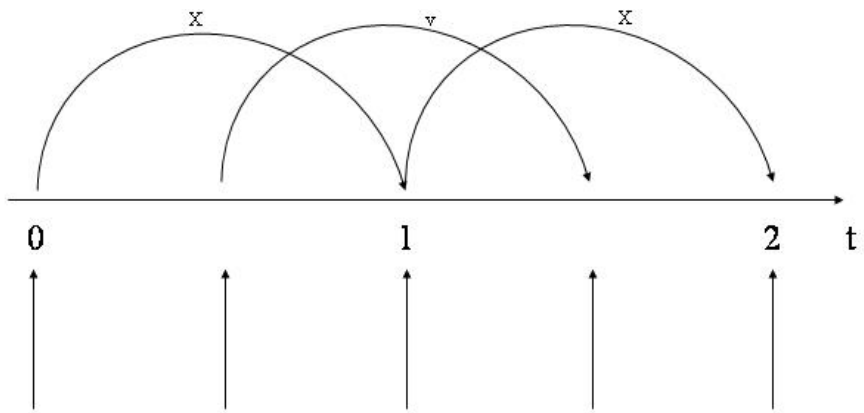
$$r(t+\Delta t) + r(t-\Delta t) = 2r(t) + \frac{f(t)}{m}\Delta t^2 + O(\Delta t^4)$$

$$r(t+\Delta t) \approx 2r(t) - r(t-\Delta t) + \frac{f(t)}{m}\Delta t^2$$

$$r(t+\Delta t) - r(t-\Delta t) = 2v(t)\Delta t + O(\Delta t^3) \qquad v(t) = \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t} + O(\Delta t^2)$$

Molecular Dynamics: The Leap Frog Algorithm

$$v(t+\frac{\Delta t}{2}) = v(t-\frac{\Delta t}{2}) + \frac{f(t)}{m}\Delta t \qquad r(t+\Delta t) = r(t) + v(t+\frac{\Delta t}{2})\Delta t$$



Biological processes:
order of tens of nanoseconds to milliseconds
MILLIONS OF INTEGRATION STEPS

Equilibrium quantities can be obtained by averaging over sufficiently-long trajectories. Dynamic information is extracted

Molecular Dynamics: Temperature and Pressure Control

No controls

Without controls, MD should generate a microcanonical ensemble (N; V;E constant). However, integration errors, force fluctuations, and inconsistencies in the forces (e.g. by using a cut-off radius) cause fluctuations and slow drifts in total energy.

Systems that are not in equilibrium will go to equilibrium while the temperature changes. We often prefer a N; V; T, or N; p; T ensemble or wish to control temperature in a prescribed, time-dependent scheme.

Weak coupling to a bath

Modify equations of motion such that system temperature or pressure approaches the required ('bath') temperature or pressure with a given time constant.

Advantages: the response of the system is first-order exponential; coupling is flexible, from strong (fast response) to weak (negligible influence on system behaviour),

Disadvantages: no known ensemble is generated with intermediate coupling constants; overall fluctuations cannot be used.

Alternative Methods: Extended Systems, Costraining to specific T and P

Molecular Dynamics: Temperature Control

$$\langle K \rangle_{NVT} = \frac{3}{2} N k_B T$$

Temperature is related to the average kinetic energy of particles. Control Temp. via Velocity rescaling

Berendsen Algorithm. The velocities are scaled at each time step, such that the rate at which the temperature changes is proportional to the difference in temperature between the bath and the system. If T_0 is defined as the reference temperature and T as the instantaneous one, we have:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}$$

Temperature deviation decays exponentially with a time constant τ. This method of coupling has the advantage that the strength of coupling can be varied and adapted to the different situations by just changing the scaling factor τ.

Disadvantage. It does not generate rigorous canonical averages: velocity rescaling artificially prolongs any temperature difference between components of the system, and in many cases the phenomenon of "hot" solvent-"cold" solute arises.

Molecular Dynamics: Pressure Control

Pressure control allows checking the behavior of the system as a function of pressure, enabling one to study, e.g. conformational changes induced by ultra-high pressure conditions.

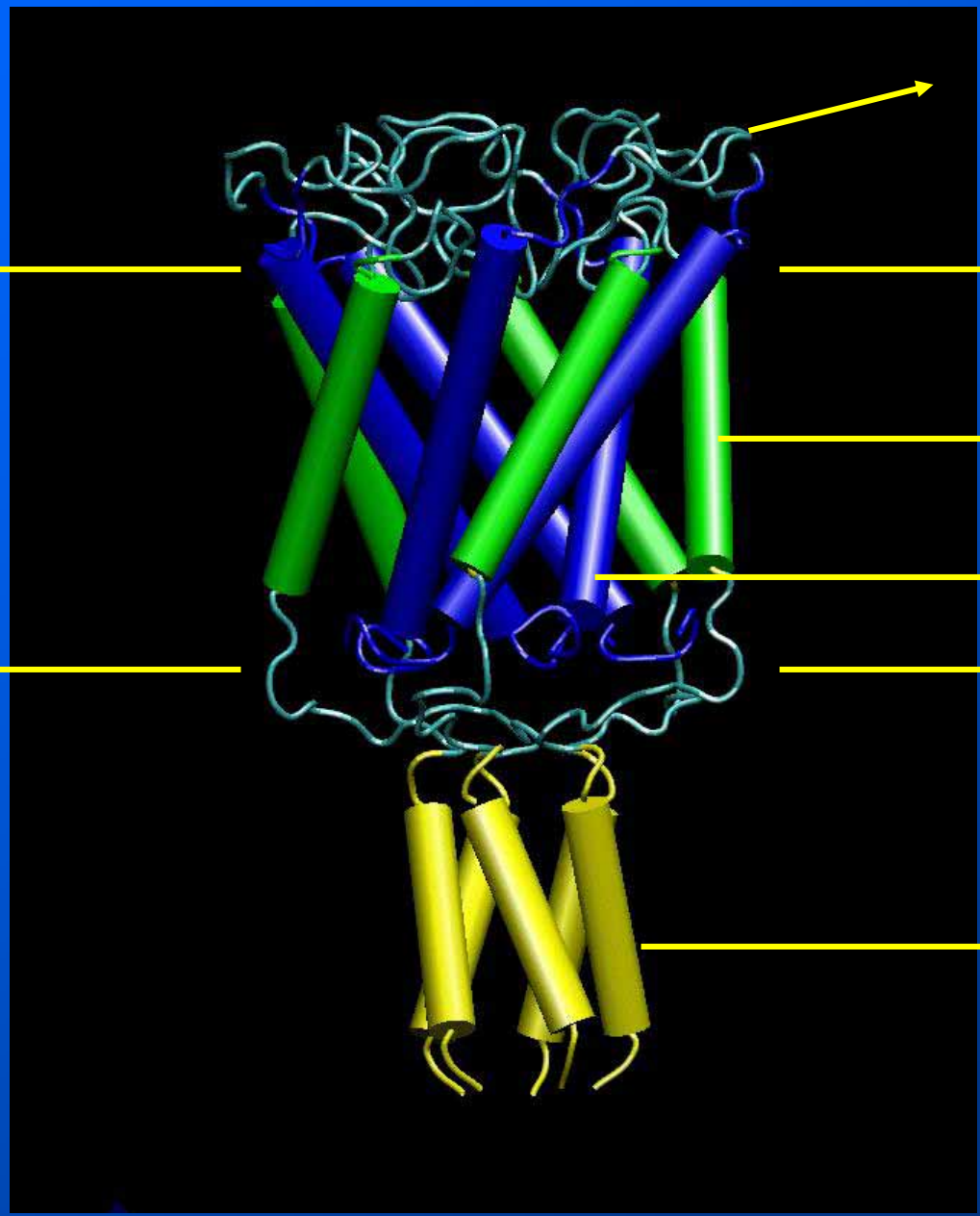
Berendsen Algorithm. It is based on the same philosophy as the temperature control. This algorithm rescales the coordinates and the box vectors every step with a matrix μ, which has the effect of a first-order relaxation of the pressure towards a given reference pressure P_0.

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P}$$

The scaling matrix μ is given by: $\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\tau_P}\beta_{ij}\left\{P_{0ij} - P_{ij}(t)\right\}$

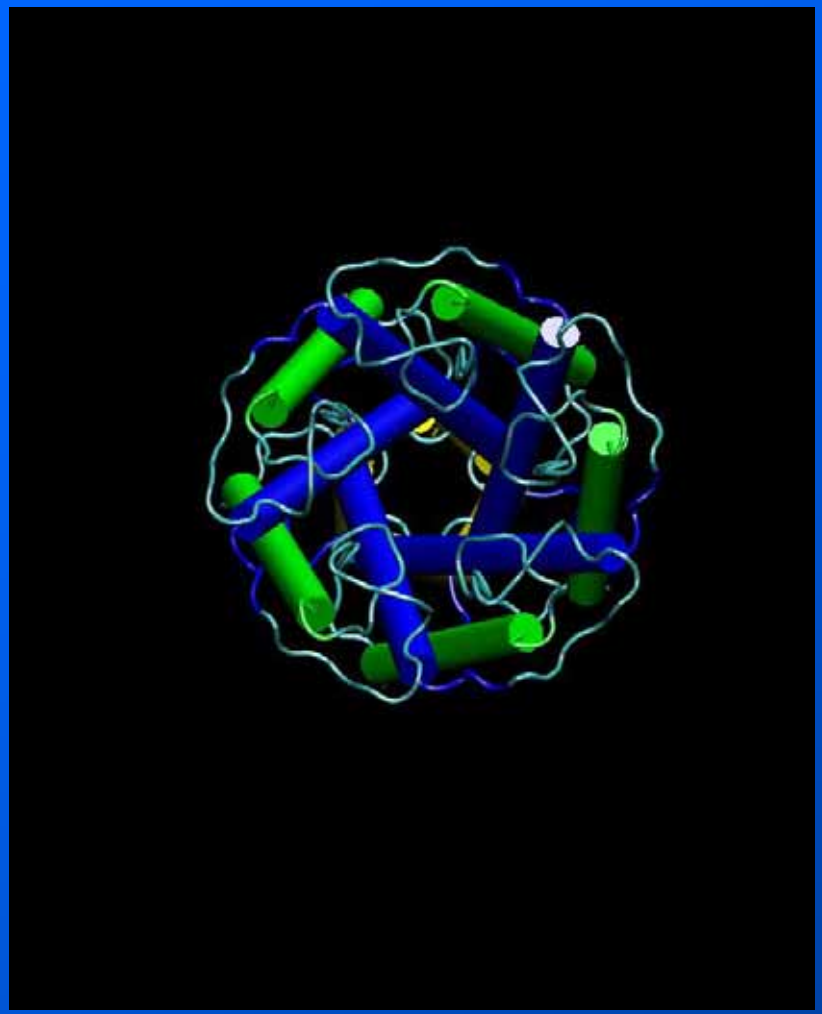
Here β is the isothermal compressibility of the system. For water at 1atm and 300K this value is: $\beta = 4.6 \times 10^{-10}$ Pa^{-1} = 4.6×10^{-5} Bar^{-1}. The scaling can be done isotropically or anisotropically depending on the type of system being simulated.

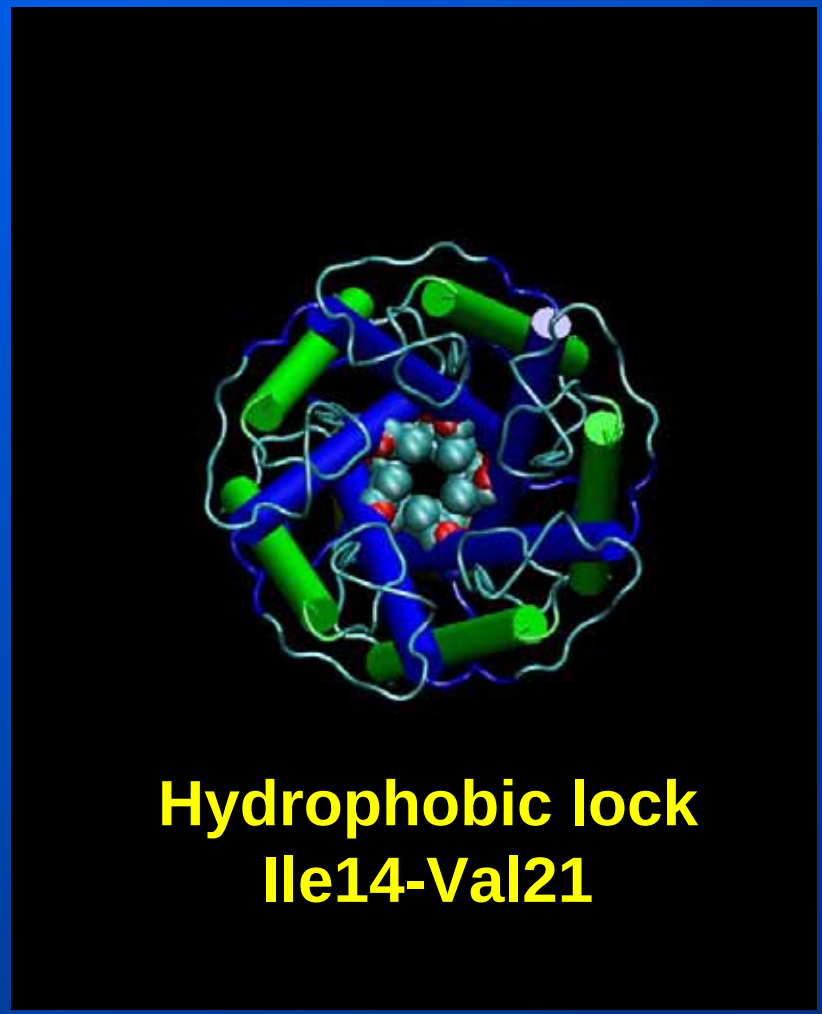
Mechanosensitive Channels of Large Conductance



Mechanosensitive Channels of Large Conductance

Top View



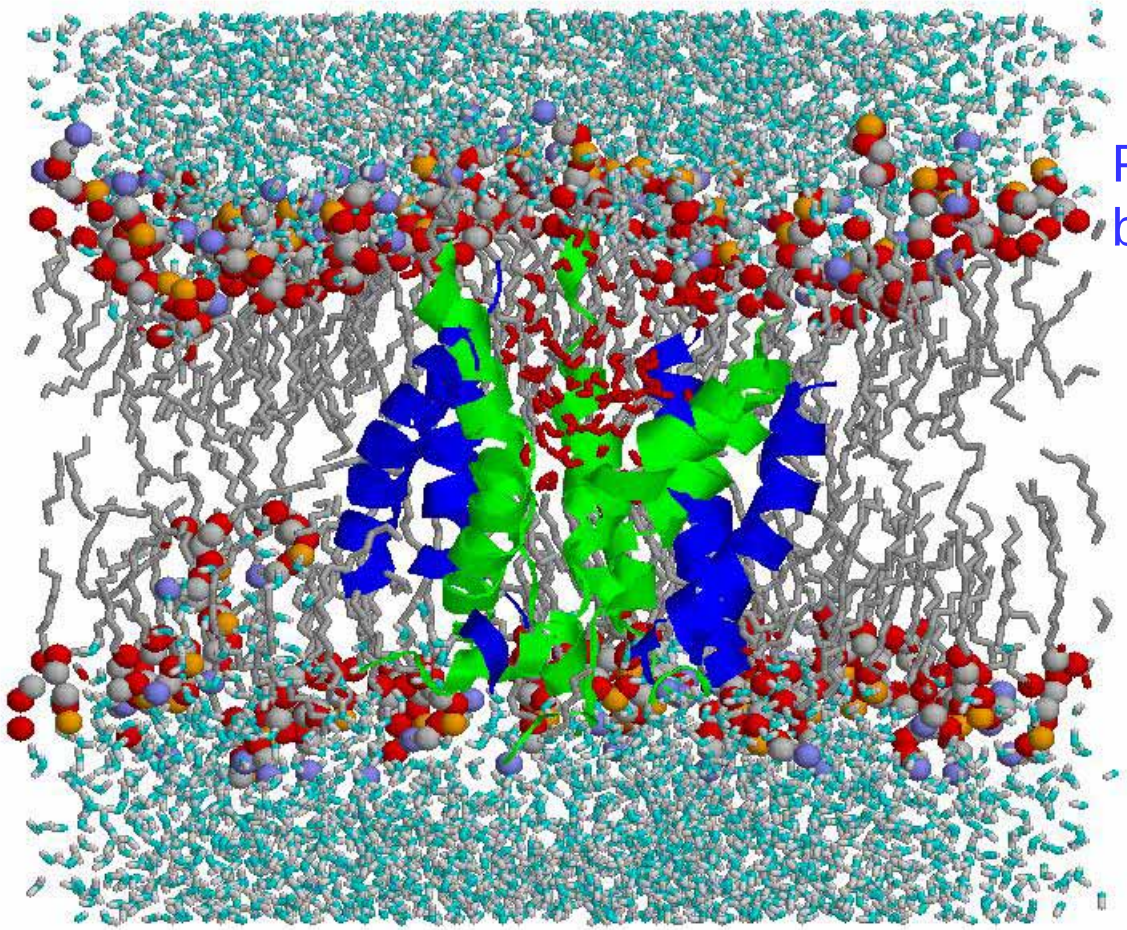


Hydrophobic lock
Ile14-Val21

What is the Opening Mechanism?

What is the Opening Mechanism?

Long timescale MD simulations in different conditions



Protein embedded in a POPC bilayer, solvated with SPC water

Four different pressure conditions

What is the Opening Mechanism?

Four different **30ns** simulations with different pressure conditions (Berendsen Pressure Coupl.) to mimick the osmotic differences

1) Simulation **NP** (Normal Pressure): P =1 bar in x, y, z

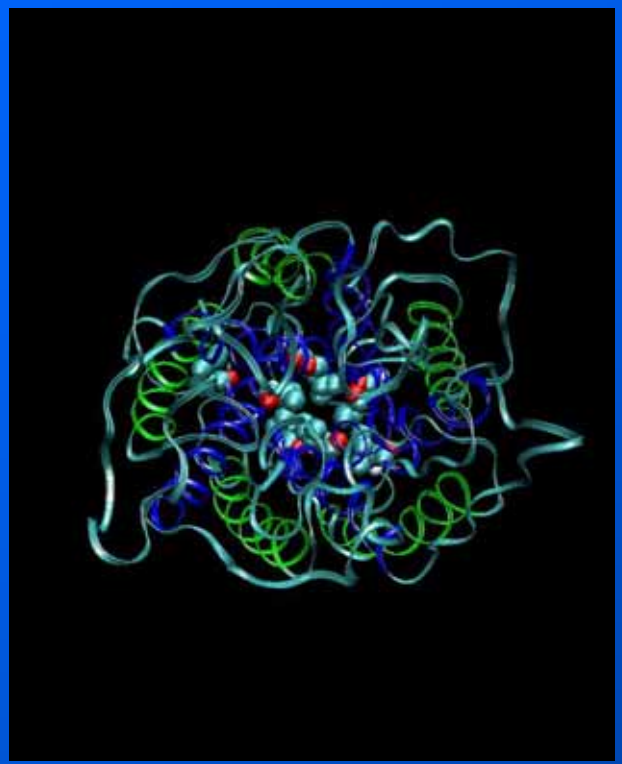
2) Simulation **P100** (Press100): P =1 bar in z, P=100 bar in xy

3) Simulation **S100** (Stretch100): P =1 bar in z, P=-100 in xy

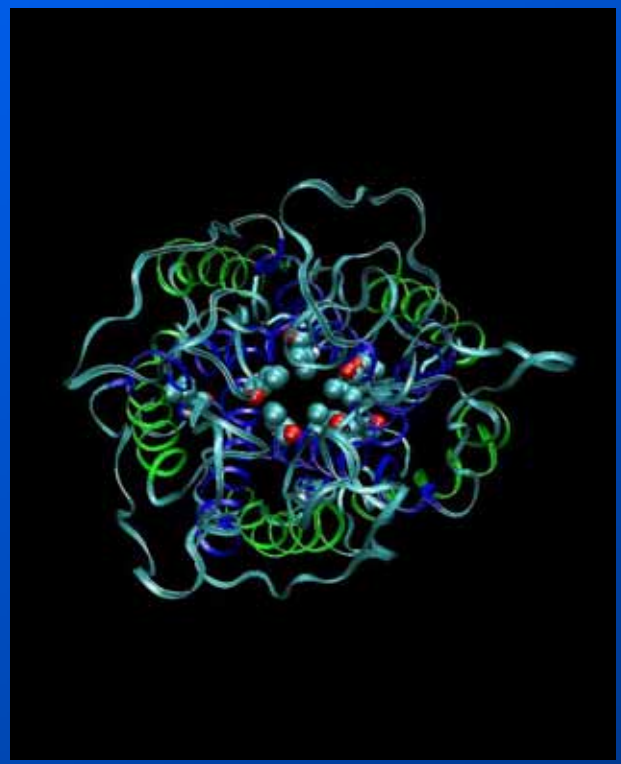
4) Simulation **S1000** (Stretch1000): P =1 bar in z, P=-1000 in xy

Molecular dynamics:
Structural Results

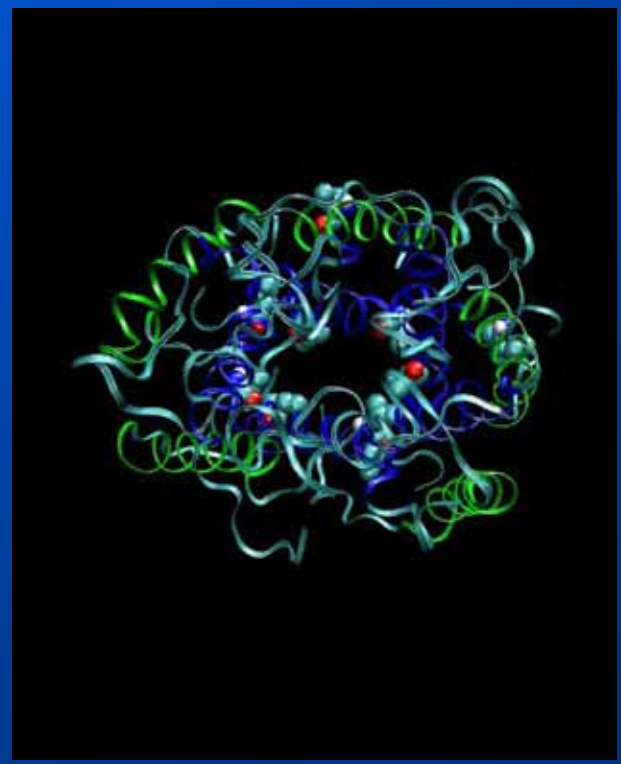
Hydrophobic Lock (Ile14-Val21) Opening



NP



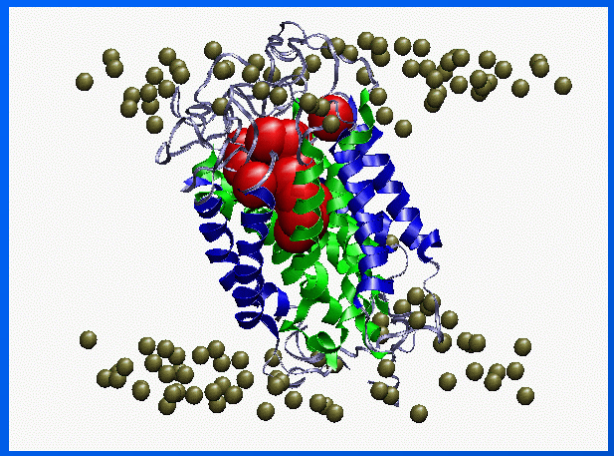
S100

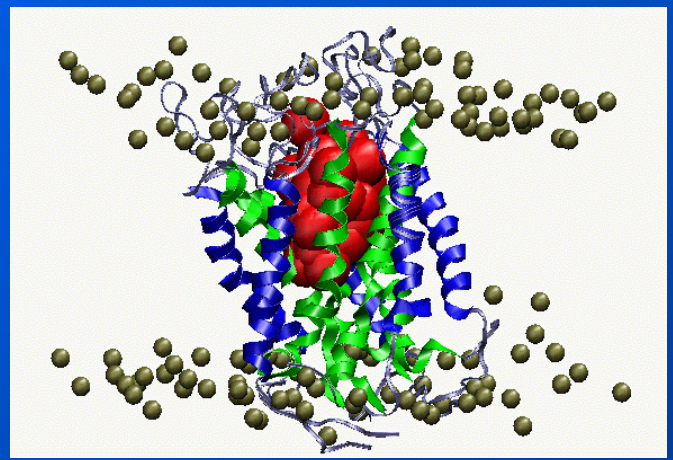


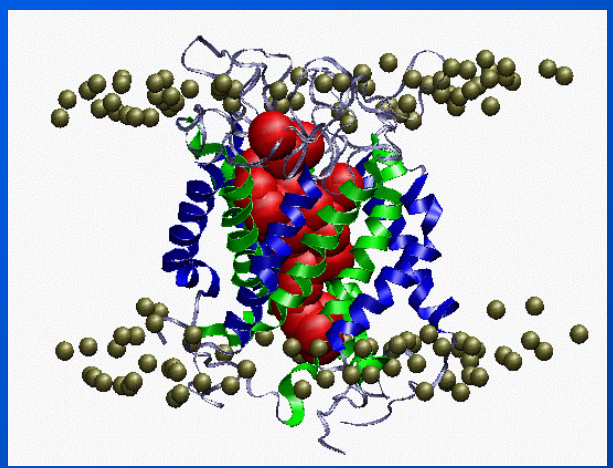
S1000

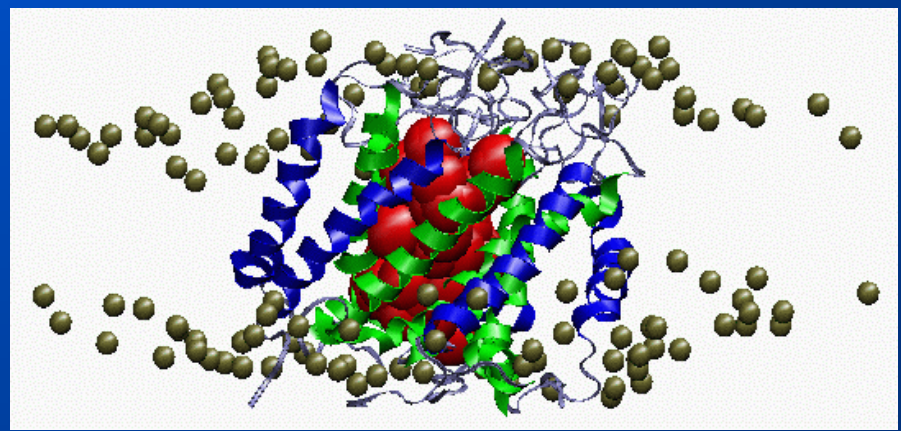
Water translocation

Increasing stretch

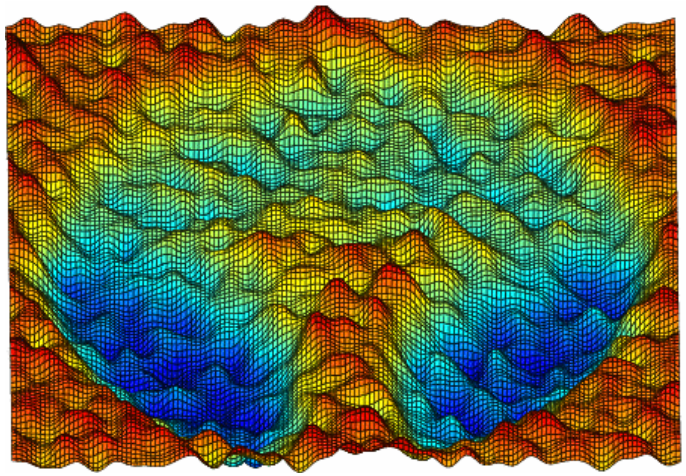








Statistical Mechanics and Simulations



The state of a biomolecular system cannot be described in terms of one single minimum, but by a statistical mechanical ensemble of configurations.

This is true for experiments, too! E.g. NOE's, CD spectra….

The weight of a configuration ***x*** *is given by the Boltzmann Factor*

$$P(x) \sim \exp(-V(x)/k_B T)$$

The exponential weighting implies that high energy regions will not contribute configurations that are significant to the state of the system, unless they are numerous (entropy).

Equilibrium properties are dominated by the parts of configurational space for which ***V(x)*** *is low.*

Therefore, one should search the low energy regions of the vast biomolecular energy surface.

Molecular Dynamics: Convergence of Simulated Properties

Biomolecular time scales range from femtoseconds to seconds. Computing power is limited.

MD simulations cover nanoseconds to microseconds.

Is this long enough to yield reliable averages?

A trajectory is representative if:

1) *the equilibration time of the simulation t_{equil} is longer than the relaxation time $t_{relax}(Q)$ of the property Q*

2) *the sampling period t_{sample} is much longer than $t_{relax}(Q)$*

$$t_{equil} > t_{relax}(Q)$$

$$t_{sample} >> t_{relax}(Q)$$

If conditions are not fulfilled then the average $\langle Q(t) \rangle$ will drift in time.

Molecular Dynamics: Convergence of Simulated Properties

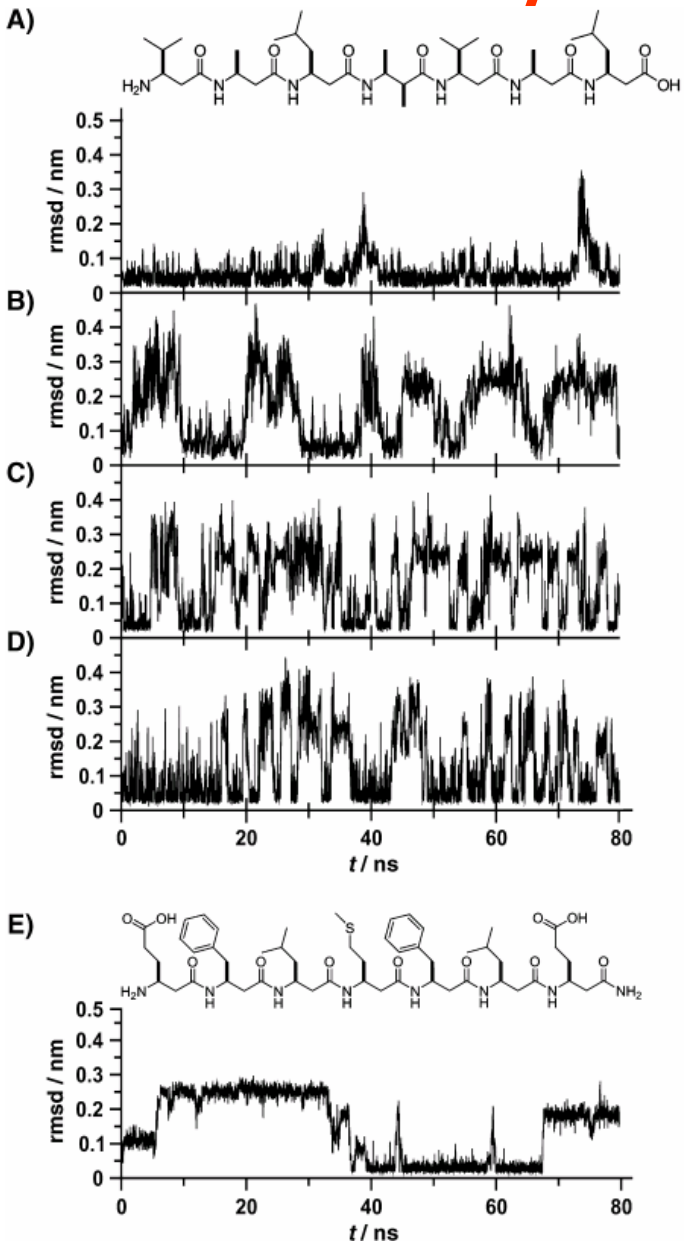
$t_{equil} > t_{relax}(Q)$

$t_{sample} >> t_{relax}(Q)$

Relaxation time depends on

1) *The type of system*
2) *The thermodynamic state point*
3) *The particular property*

Examples:
Potential energy relaxes faster than rmsd.
Free Energy converges slower than all other properties.



300K

340K

Reduced viscosity

Charge variation

Molecular Dynamics: Convergence of Simulated Properties

Control of convergence:

1) Check the avg. value $\langle Q(t) \rangle$; Fluctuations $\langle (Q(t)) - \langle Q(t) \rangle_t^2 \rangle_t^{1/2}$.

Calculate Autocorrelation function: $\langle Q(t')Q(t'+t) \rangle_{t'}$.

The decay time of the Autocorrelation function or the build-up of averages give an indication of $t_{relax}(Q)$.

2) Start several non equilibrium trajectories and measure the rate of relaxation of Q.

3) If different trajectories started from different initial states don't converge to the same average, then $t_{relax}(Q)$ longer than simulation time.

Molecular Dynamics: Problems with Convergence

*The Free Energy **F** of a system of N particles in a volume V at temperature T is a 6N-dimensional integral over all positions **r** and momenta **p** of the Boltzmann factor of the system Hamiltonian.*

$$F_{NVT} = -k_B T \ln\left(\left(N! h^{3N}\right)^{-1} \iint \exp\left(-H(p,r)/k_B T\right) dp dr\right)$$

The Integrand is always positive, and the omission of configurations leads to systematic errors.

It is still difficult to evaluate Free Energy properly

Molecular Dynamics

The General MD algorithm

1. Input initial conditions

Define Potential Interaction as a Function of atom positions

Positions *r* of all atoms in the system

Velocities *v* of all atoms in the system

Repeat 2, 3, 4 for the required number of steps

---

2. Calculate forces

Force on every atom

$$F_i = -\frac{\partial V}{\partial r_i}$$

Calculate potential and kinetic energies, viral and pressure

3. Update configuration

Solve Newton's equations of motion

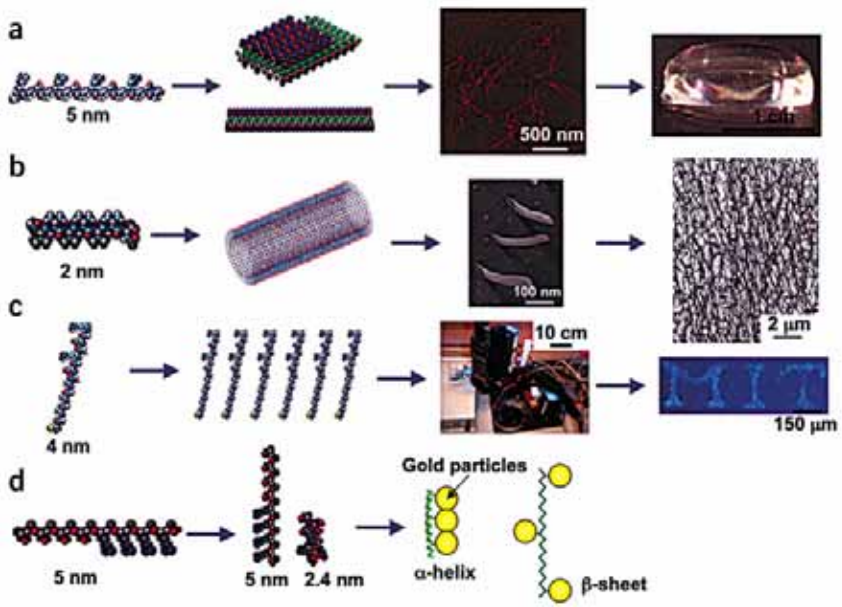
$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i}$$

4. Write Output

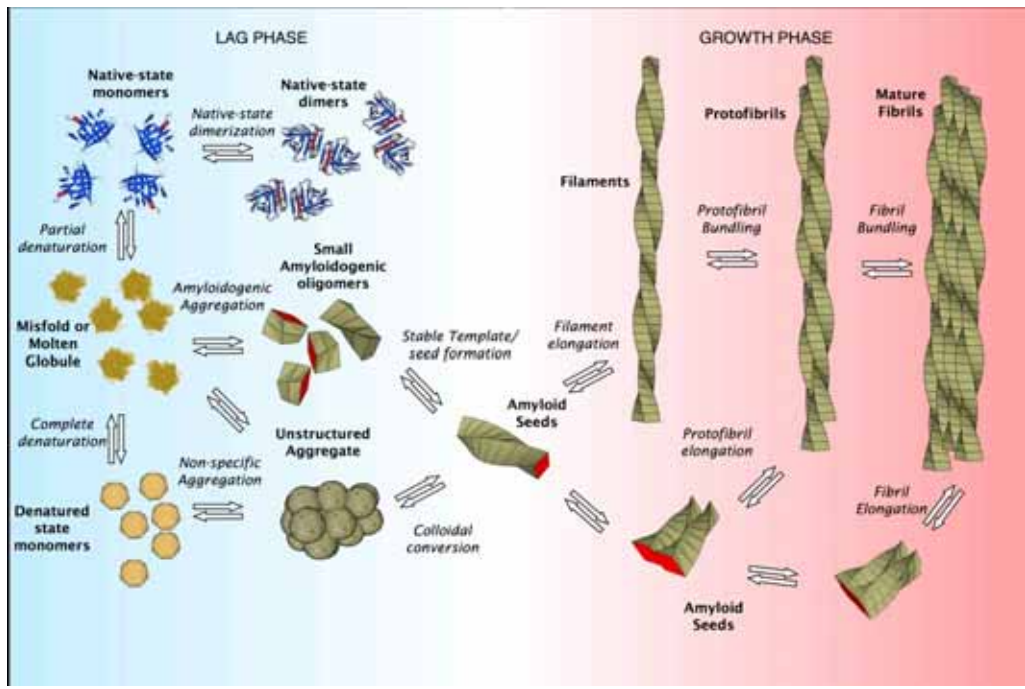
The role MD in new biological problems

Peptide/Protein self-assembly

New BionanoMaterials; devices



Understanding amyloid fibril formation. Alzheimer's; Parkinson's….

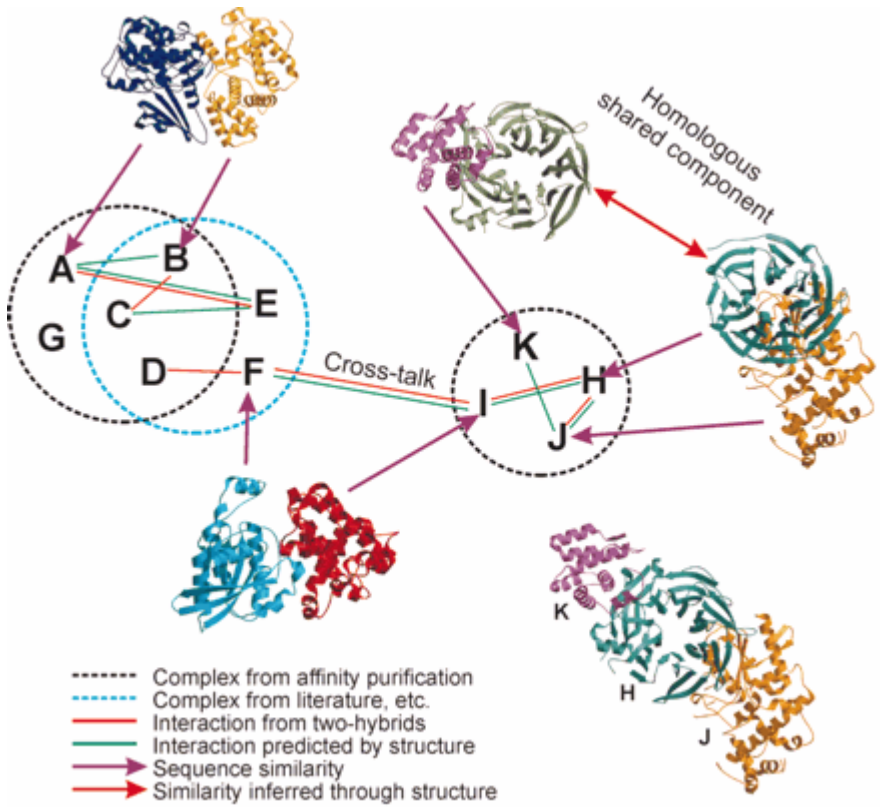


The role MD in new biological problems

Protein-Protein Interactions

Structural-Dynamic View of complex pathways

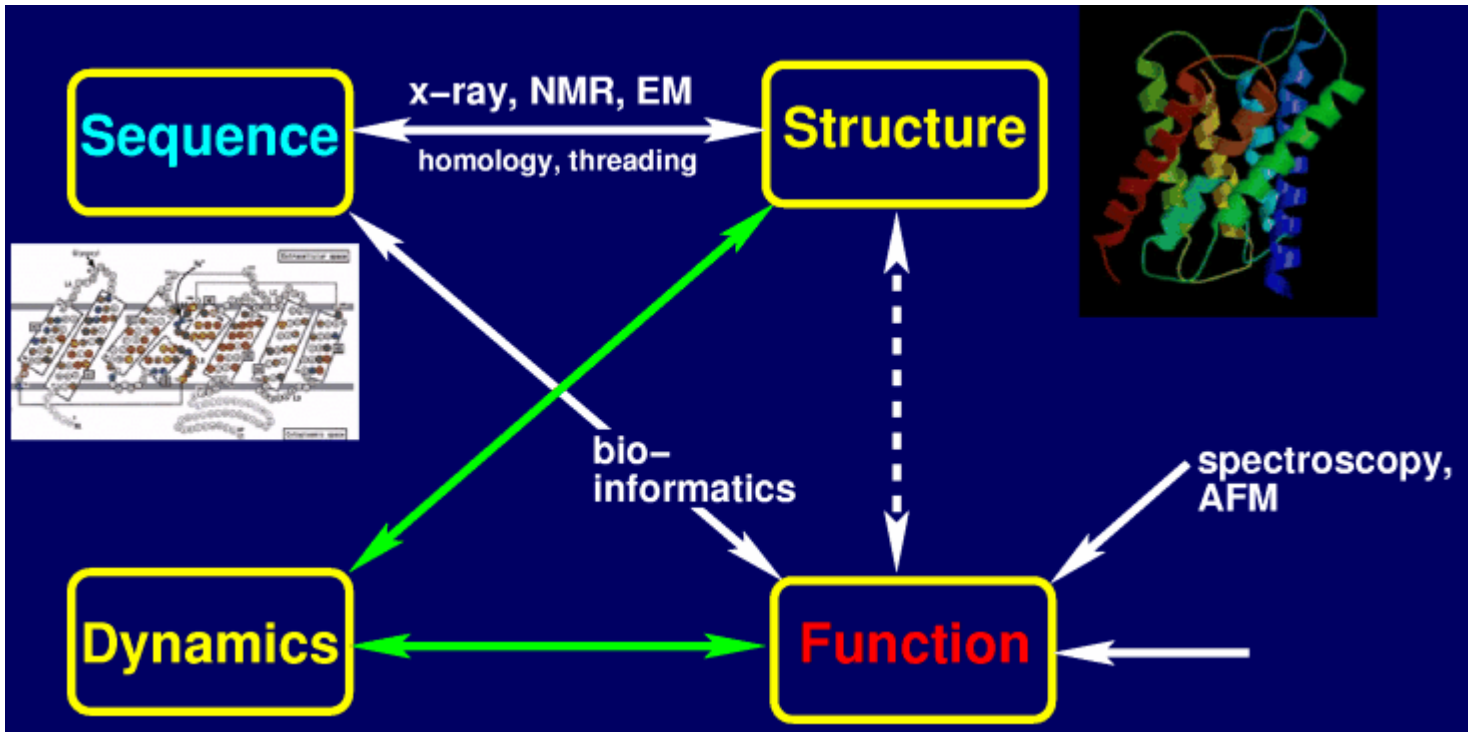
Synthetic Biology



Take Home Lesson

Know the system under exam, choose the right model and simulation level.

Many problems cannot be thought of in terms of 1 single structure



Advantages: High time and spatial resolution, Ensemble vision

Drawbacks: Limited time scales (hundreds ns) and system size (10^5 Atoms), no reactions

The Real World

Applications

Classical mechanics allows the study of systems of 100,000 particles over time spans in the 100 nanosecond range.

This includes much (but not all) of the biologically relevant motions of proteins and of lipids in membranes, even including spontaneous aggregation.

But can we distinguish the functionally relevant motions among the messy random fluctuations? Do simulations really help to understand and even predict function?

Real World Cases

Tumor Targeting Peptides

CNGRC (Cys - Asn - Gly - Arg - Cys):

Identified by in vivo panning of phage peptide library in tumor-bearing animals (Arap et al.,Science 1998)

useful for delivering anti-tumor compounds like chemotherapeutic drugs, apoptotic peptides, cytokines…

Corti and coworkers showed that targeted delivery of Tumor Necrosis Factor (TNF) to tumor vasculature can be obtained by coupling its N-terminus to the C-terminus of CNGRC

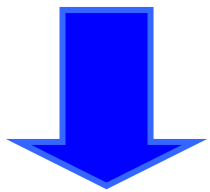
Aim: Design of more specific and more active sequences

GC et al, J. Biol. Chem. 2002. A. Corti et al, Mol. Immunol. 2006

Real World Cases

Tumor Targeting Peptides

Structure of the receptor on tumor is NOT KNOWN



We need to investigate the conformational characteristics of the ligand
AT ATOMIC DETAIL

Real World Cases

Tumor Targeting Peptides

We need to investigate the conformational characteristics of the ligand
AT ATOMIC DETAIL in solution
With Molecular Dynamics

NoSS => Cys - Asn - Gly - Arg - Cys

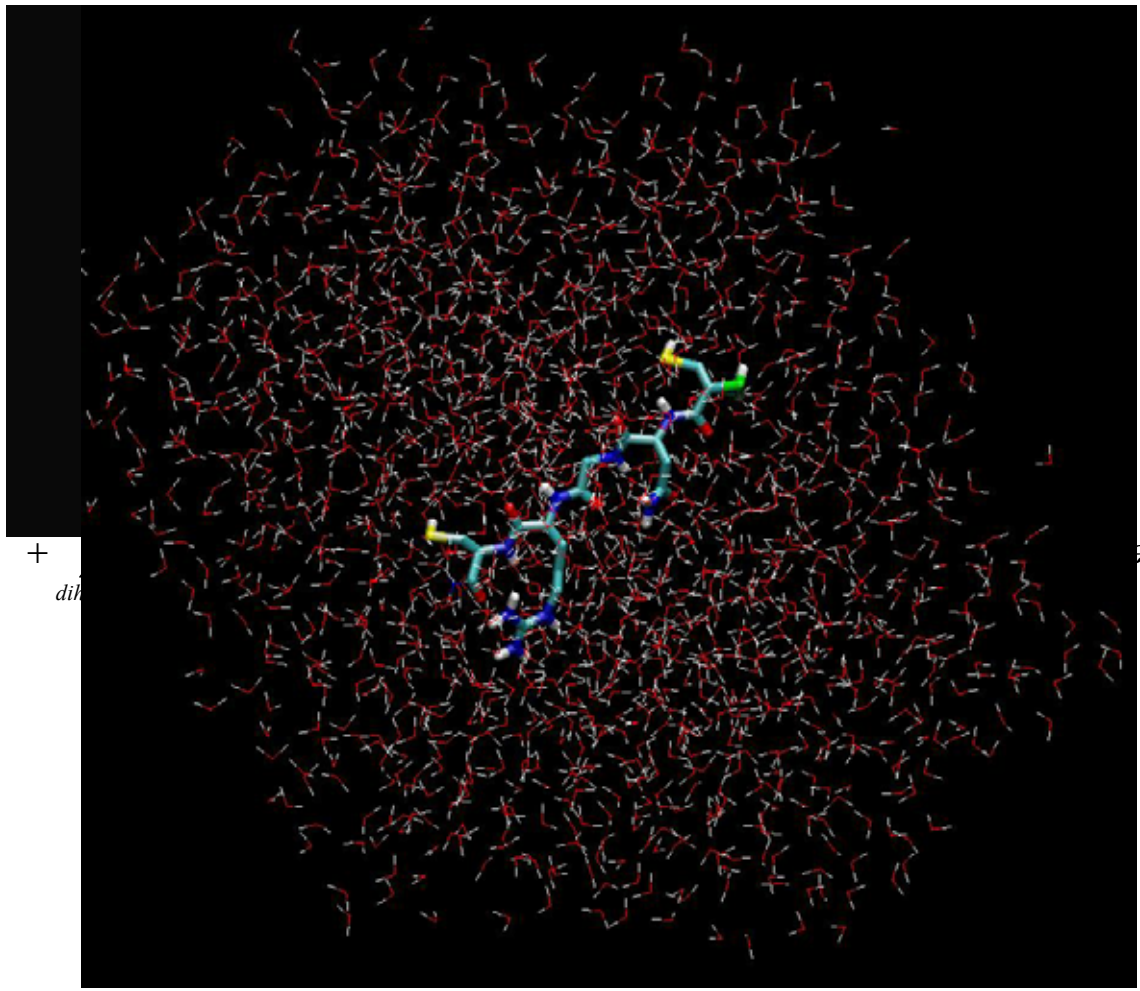
SS => Cys - Asn - Gly - Arg - Cys

Real World Cases

Tumor Targeting Peptides

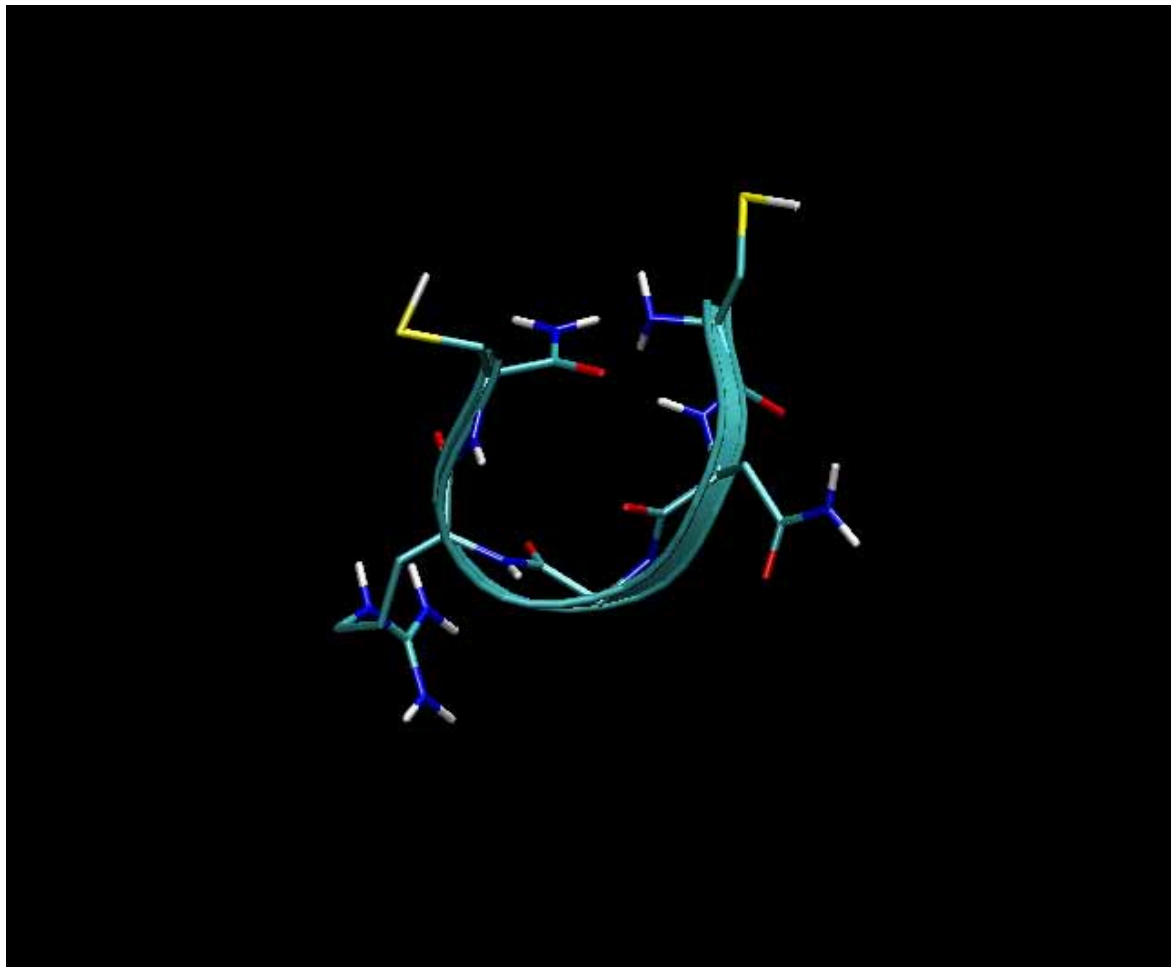
Solute

Equation Solvation



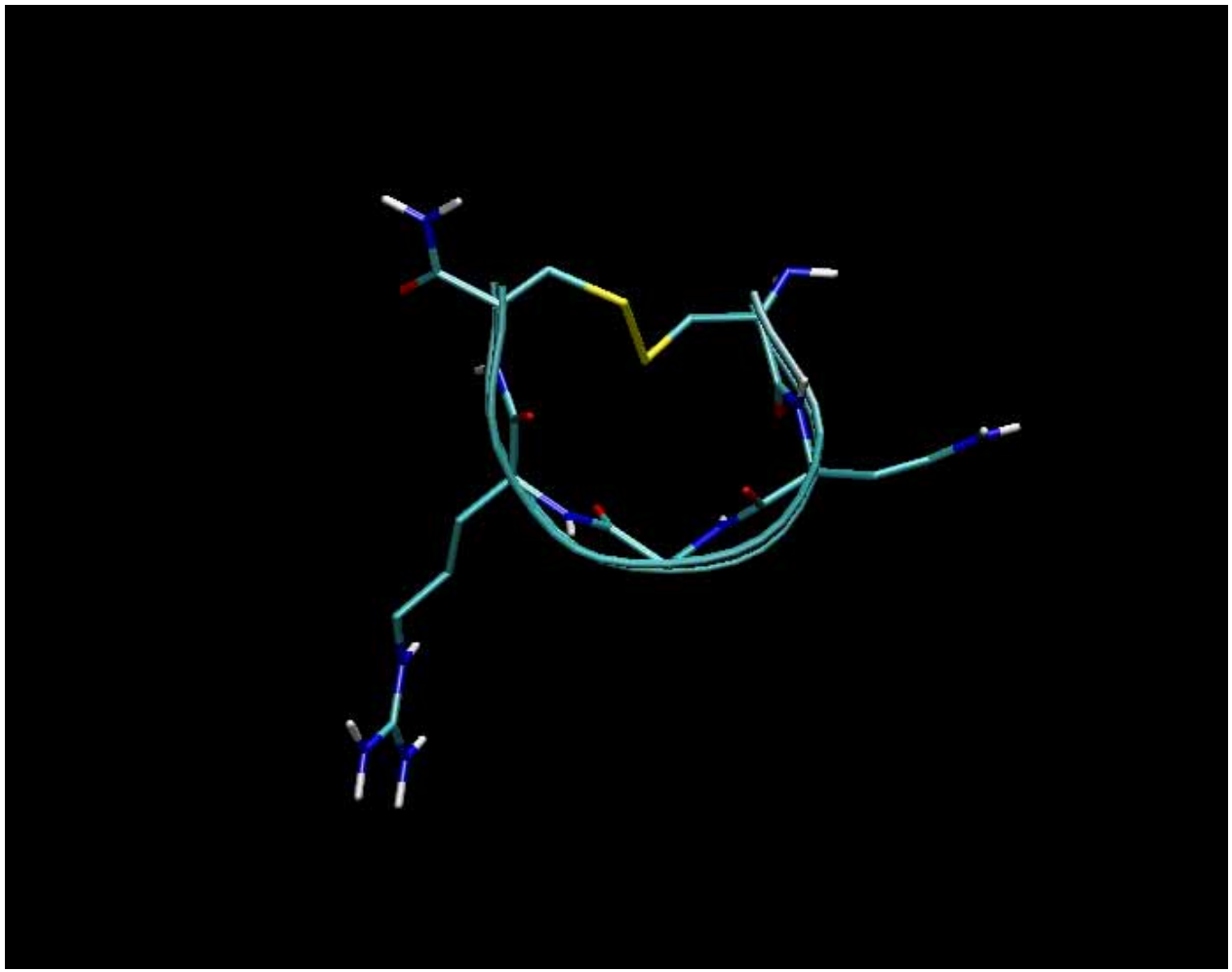
Real World Cases

NoSS System Evolution



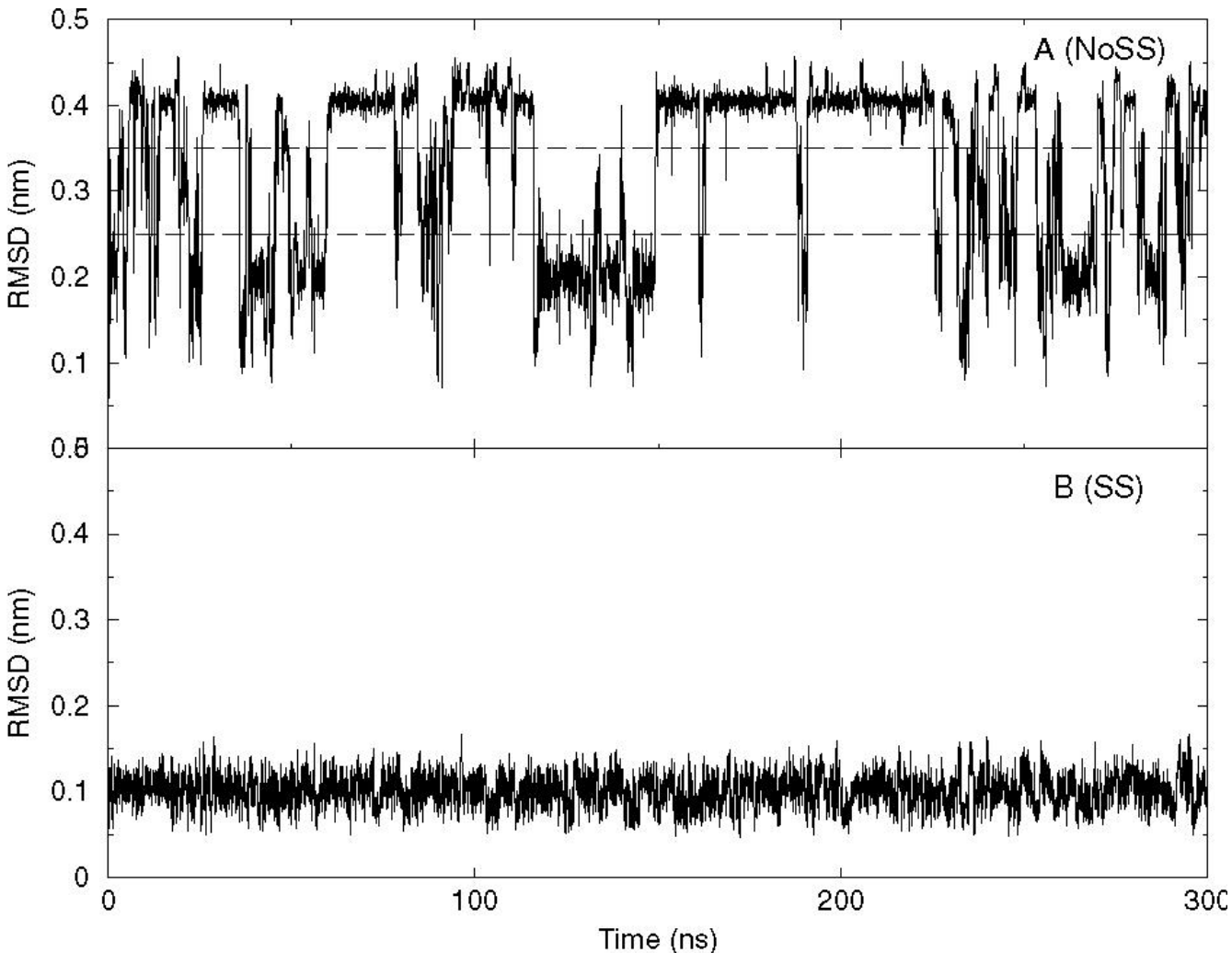
Real World Cases

SS System Evolution



Real World Cases

Tumor Targeting Peptides Analysis

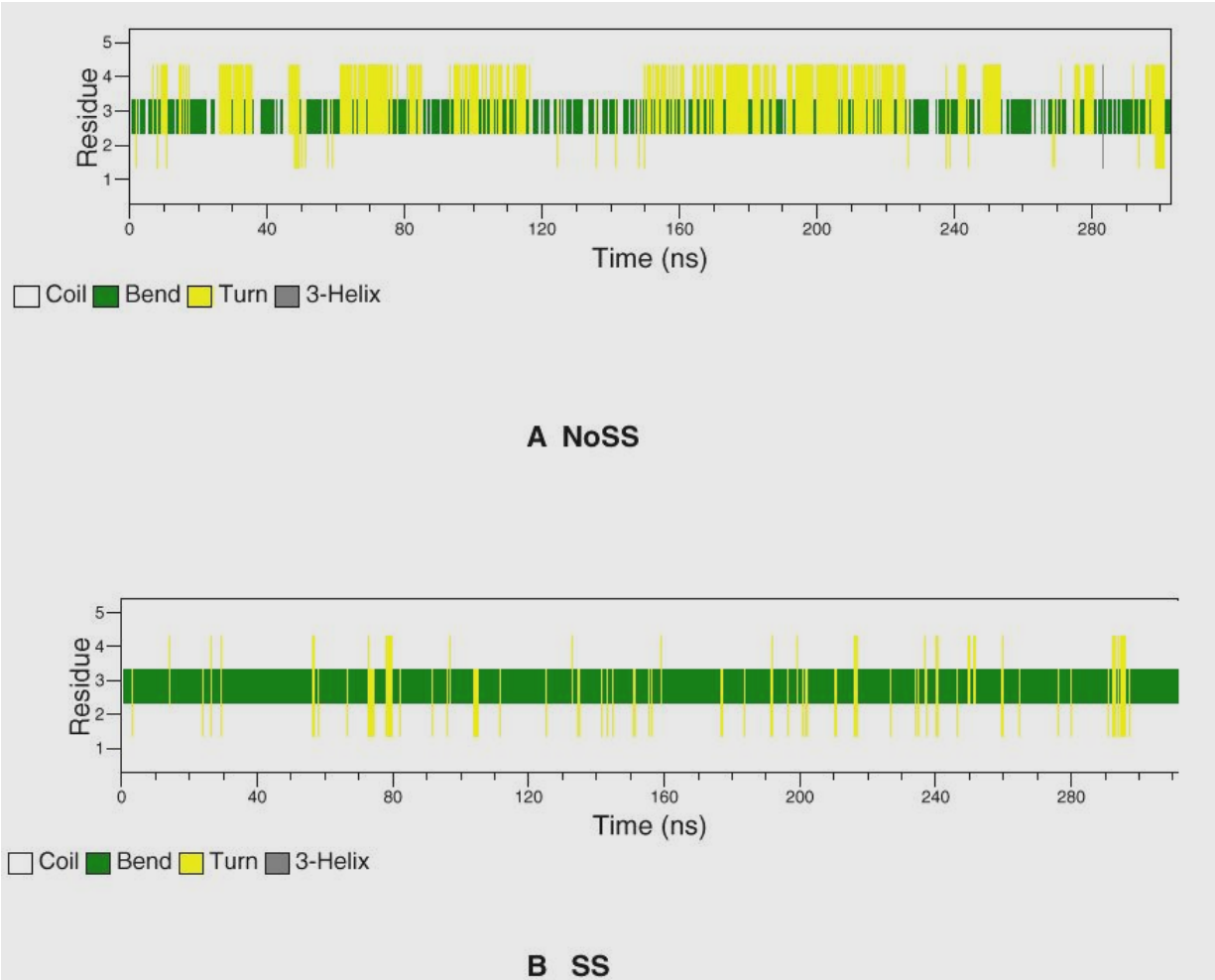


NoSS evolution

SS evolution

Real World Cases

Tumor Targeting Peptides
Secondary Structure Analysis



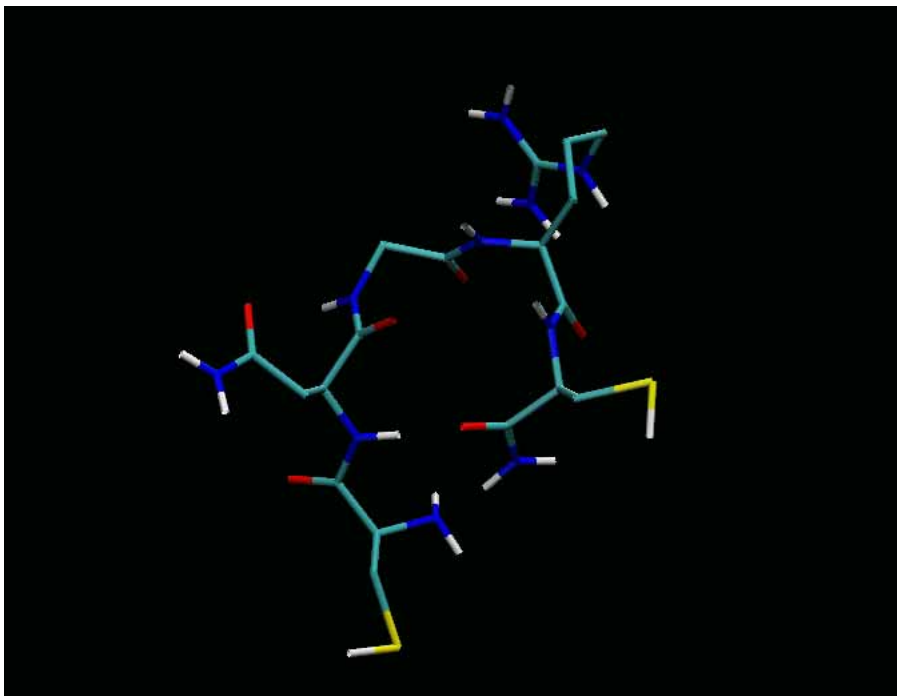
NoSS

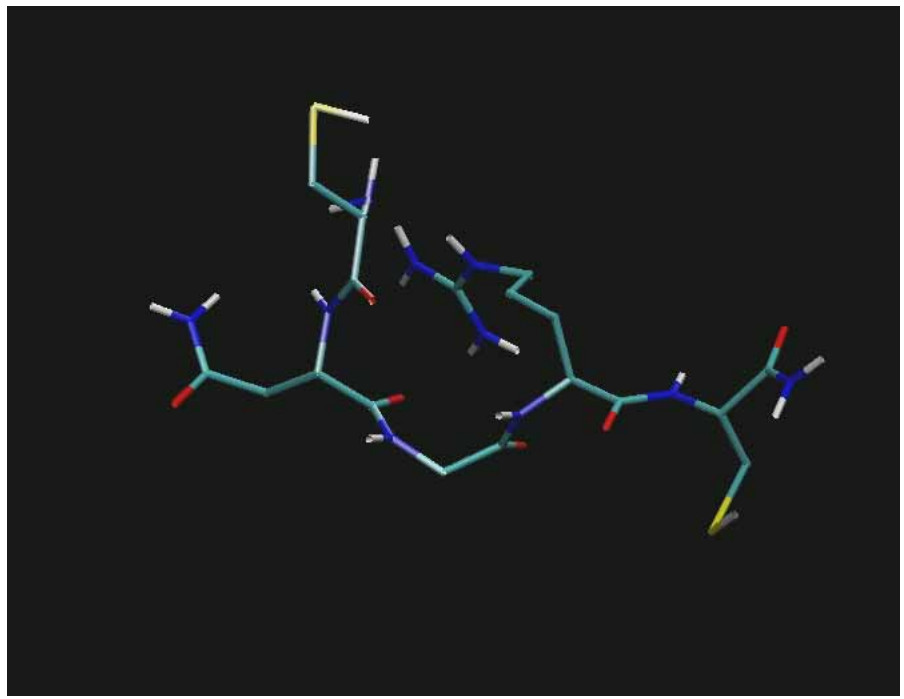
SS

Real World Cases

Tumor Targeting Peptides Analysis

NoSS System Conformations: Statistical Analysis
Representative Conformations

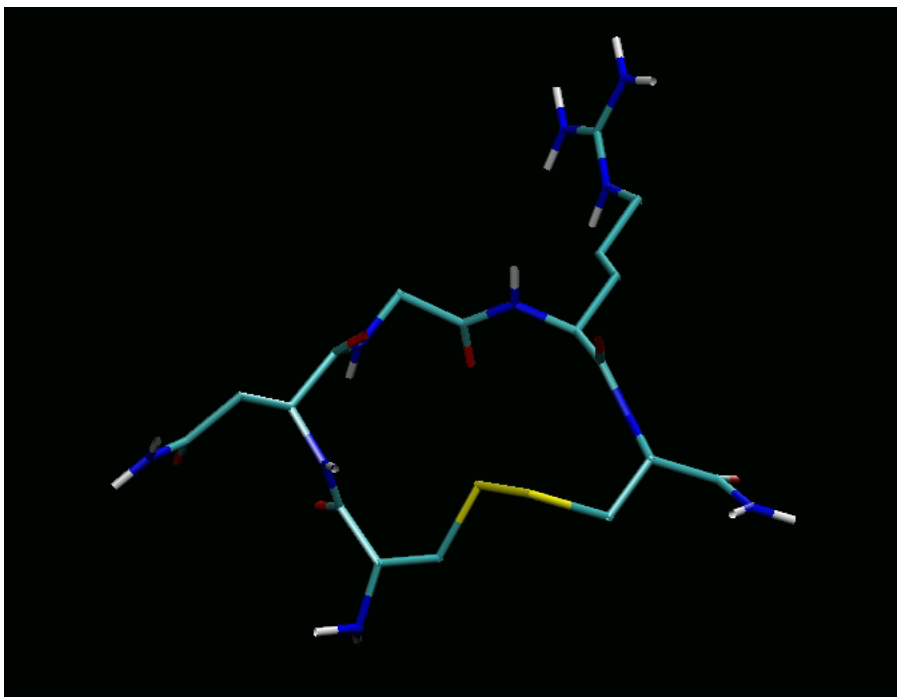


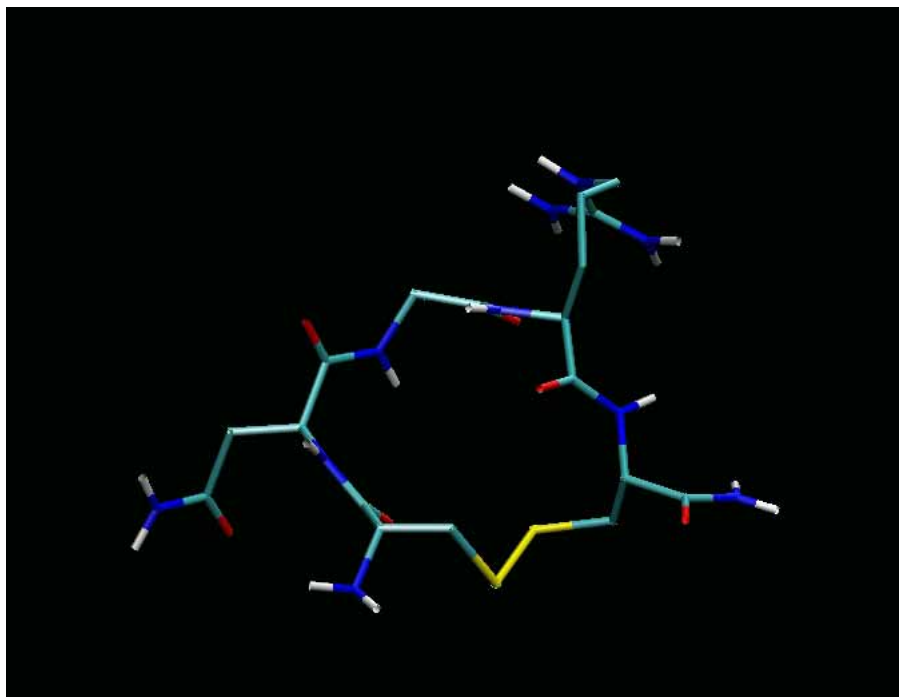


Real World Cases

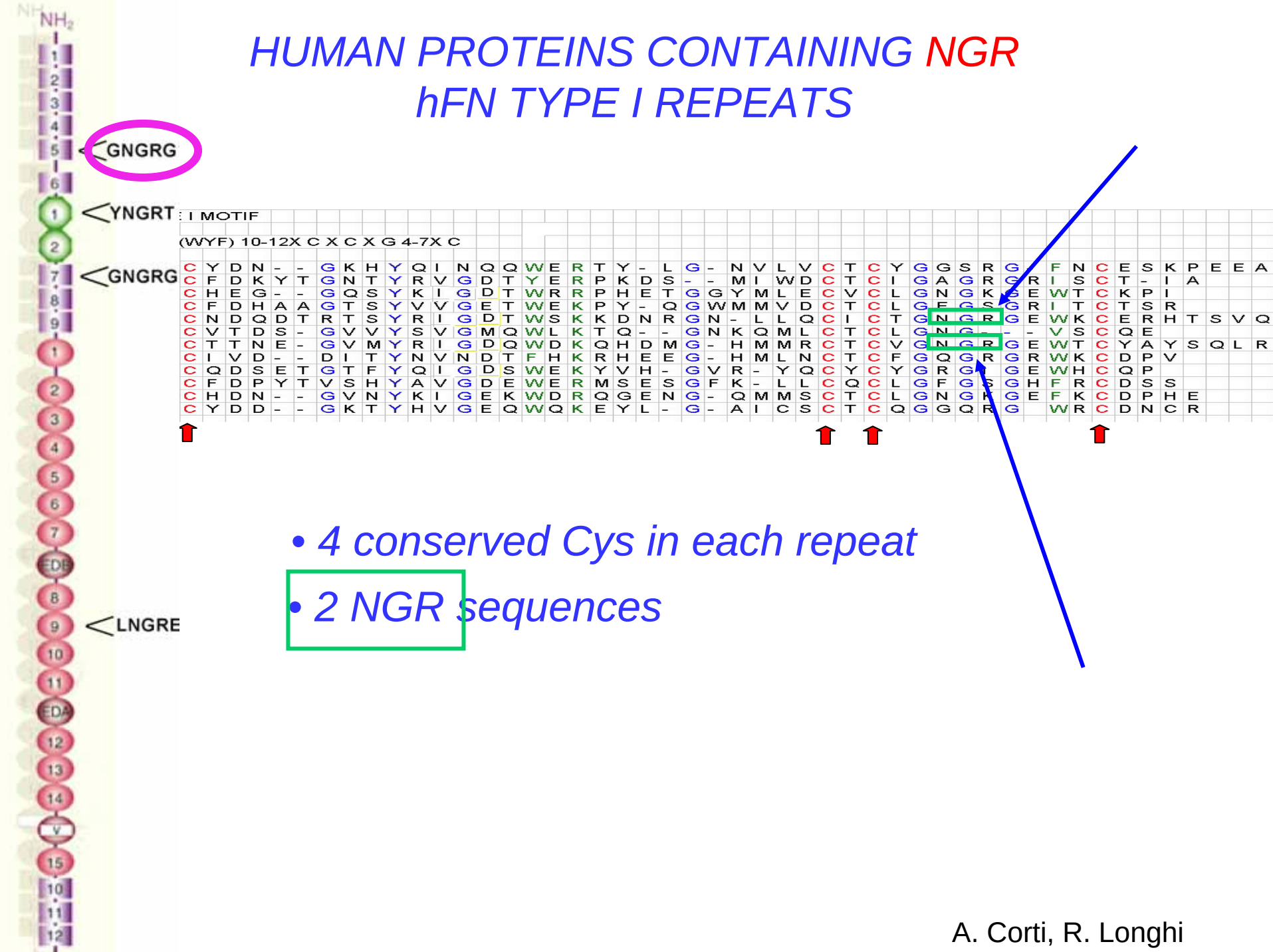
Tumor Targeting Peptides Analysis

SS System Conformations: Statistical Analysis
Representative Conformations





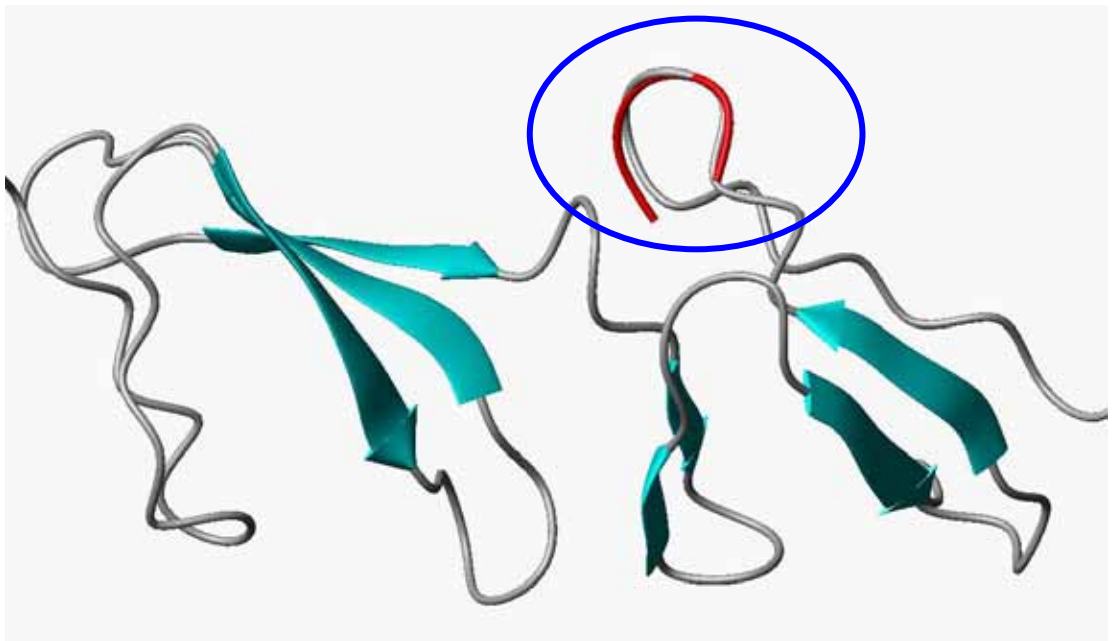
HUMAN PROTEINS CONTAINING NGR
hFN TYPE I REPEATS

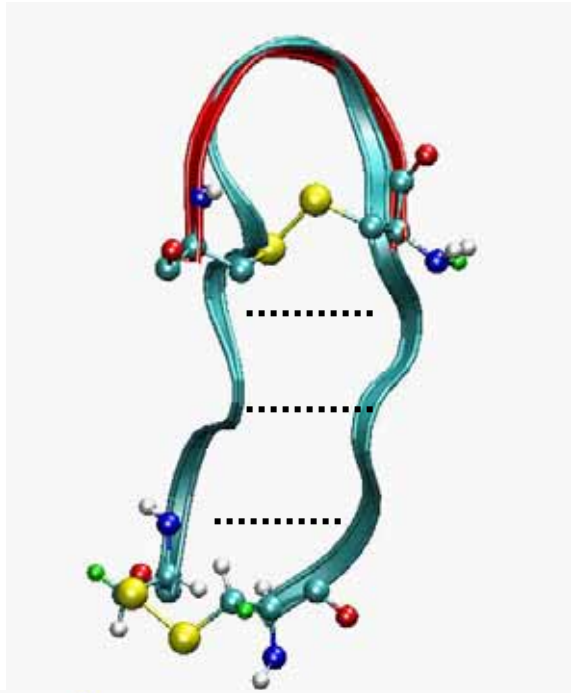


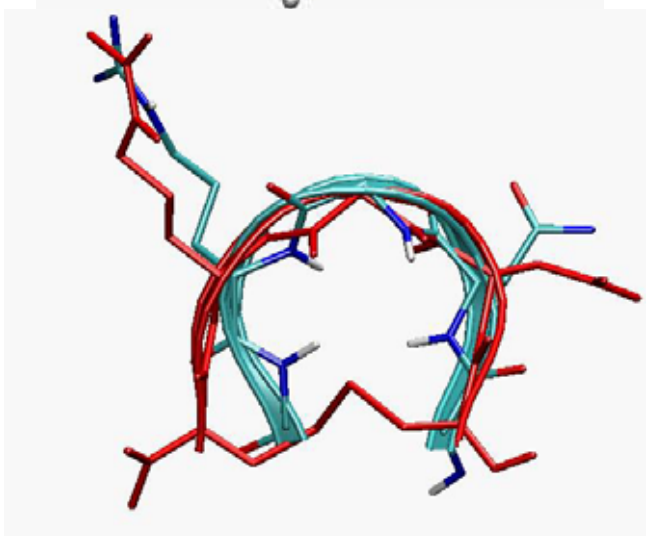
- *4 conserved Cys in each repeat*
- *2 NGR sequences*

A. Corti, R. Longhi

CNGRC PEPTIDE AND N-TERMINAL FN's GNGRG







The **CNGRC** peptide is superimposable with the **GNGRG** motif in N-term FN

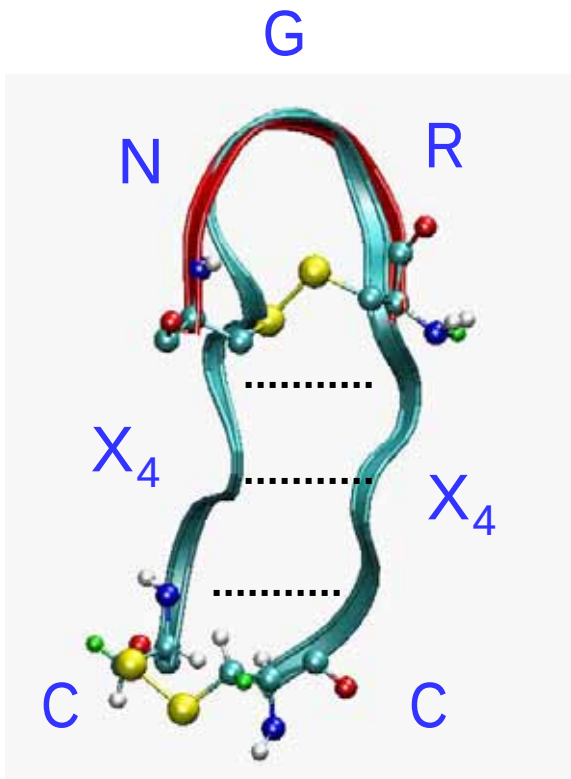
DESIGN OF NEW TUMOR HOMING PEPTIDES

Screening NGR containing peptides,
chemically synthesized

Cyclic peptides, based on hFN hairpin:

C -X_4-NGR-X_4-C

Motif Search in Prosite



Linear peptides:

peptides that naturally form β-turns

SCREENING OF NGR-CONTAINING PEPTIDES

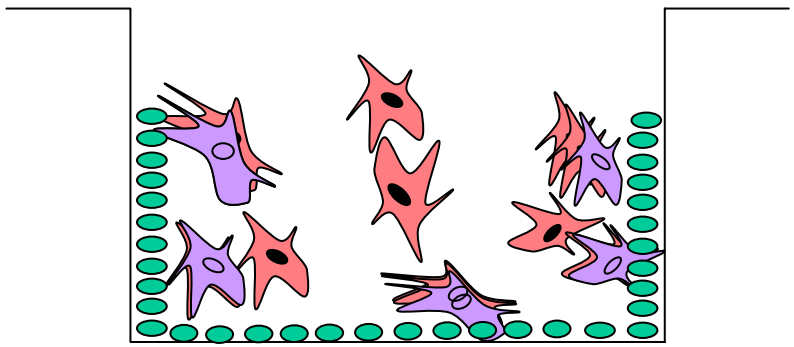
Peptide	Sequence
CNGRC	CNGRC
GNGRG	GNGRG
Scramble	RGGNG
QN	Ac -gCDLSQNGRNWKSC
TT	Ac –gCDLSTNGRTWKSC
CTS	Ac –gCLLTPNGRVNSSC
TRP	Ac -gSWTSENGRKWTWK
GB1	Ac –gRWQYVNGRKFTVQ
MAP4-gNGRg	(GNGRG-GSY)$_4$K$_2$K$_1$bA

hFN

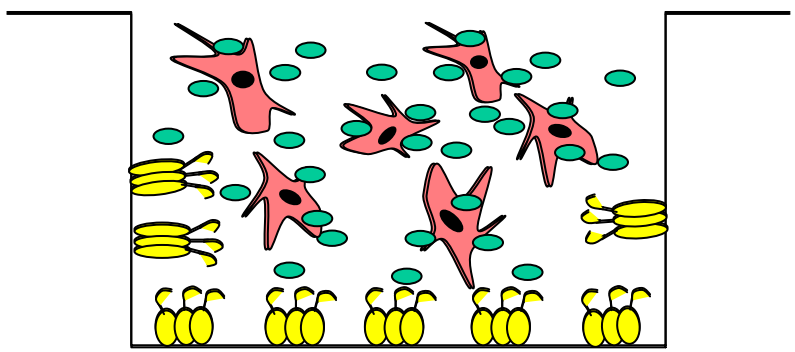
A. Corti, R. Longhi

EXPERIMENTS FOR THE SCREENING OF NGR-CONTAINING PEPTIDES

ADHESION ASSAY



COMPETITION ASSAY

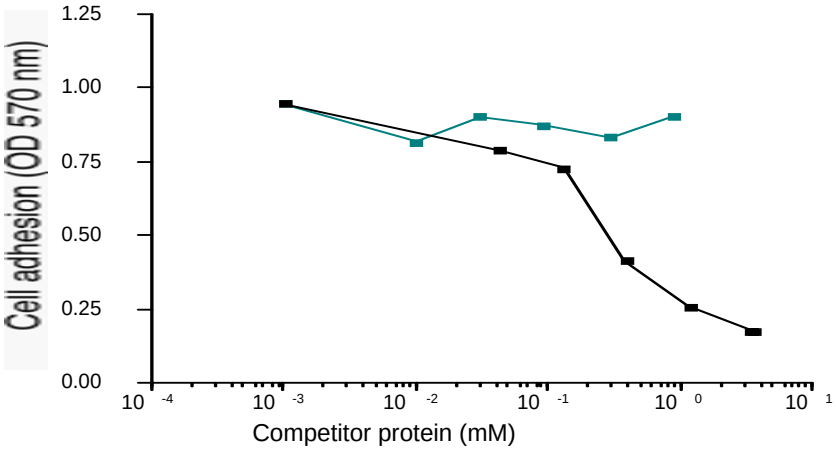


IS NGR A CELL ADHESION MOTIF??

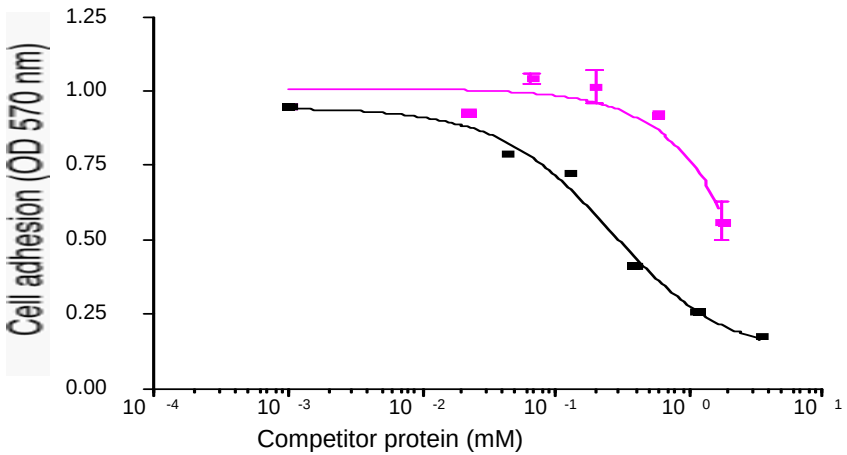
A. Corti

EFFECT OF **CYCLIC** NGR-CONTAINING PEPTIDES ON EA.hy926 CELLS ADHESION

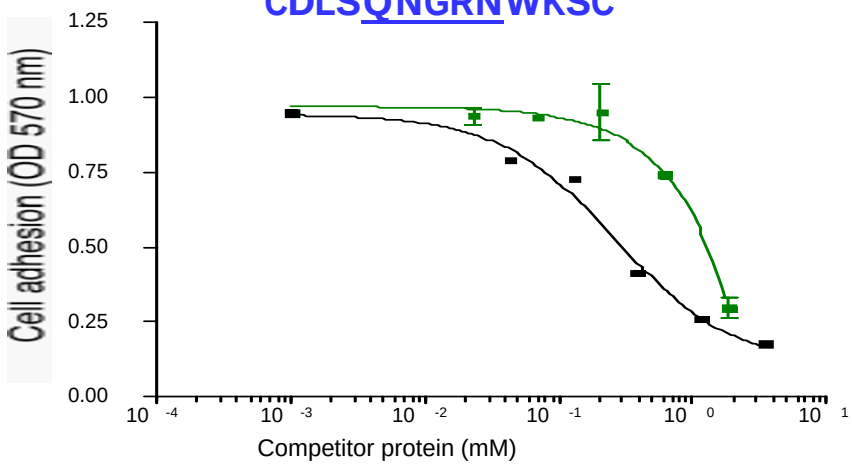
RGGNG



GCLLTPNGRVNSSC



CDLSQNGRNWKSC

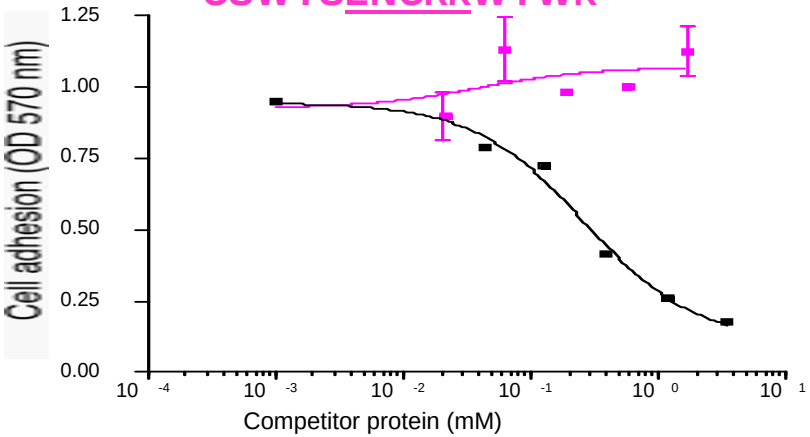


CONCLUSION:
Cyclic peptides are able to compete EA.hy926 binding to NGRhTNF.

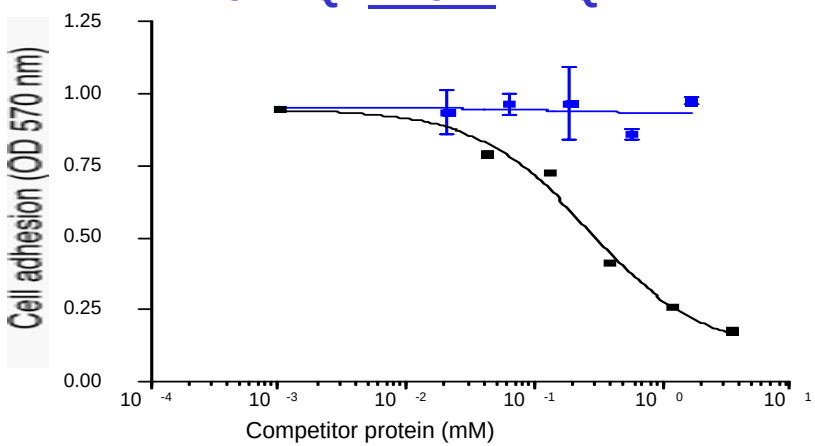
A. Corti, C. Longoni

EFFECT OF **LINEAR** NGR-CONTAINING PEPTIDES ON EA.hy926 CELLS ADHESION

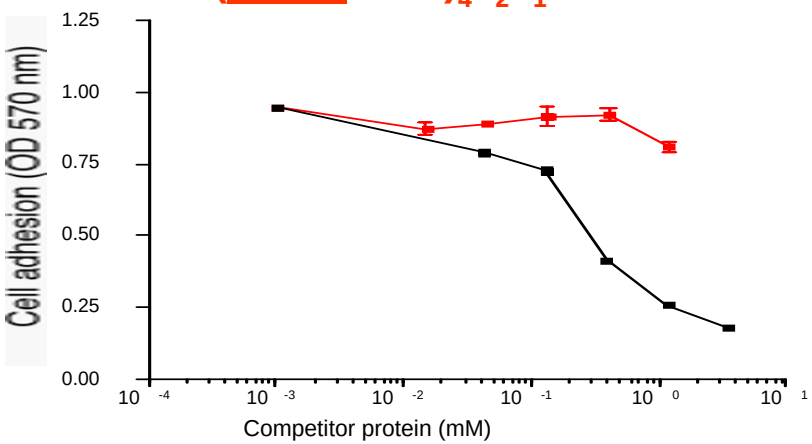
GSWTSENGRKWTWK



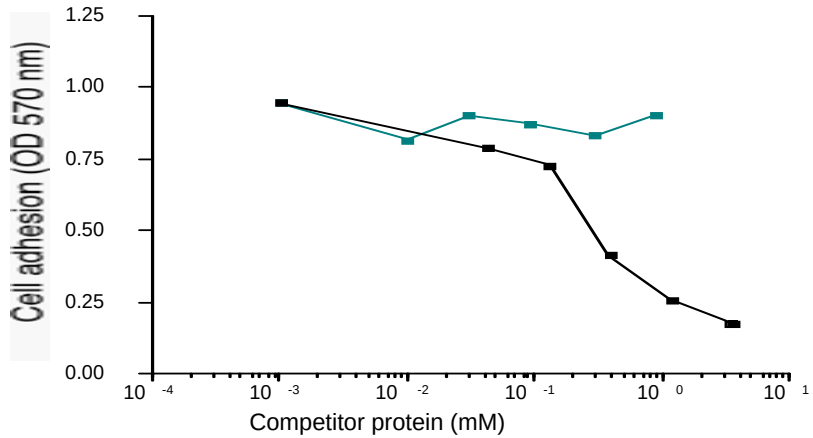
GRWQVVNGRKFTVQ



(gNGRg-GSY)$_4$K$_2$K$_1$bA



RGGNG

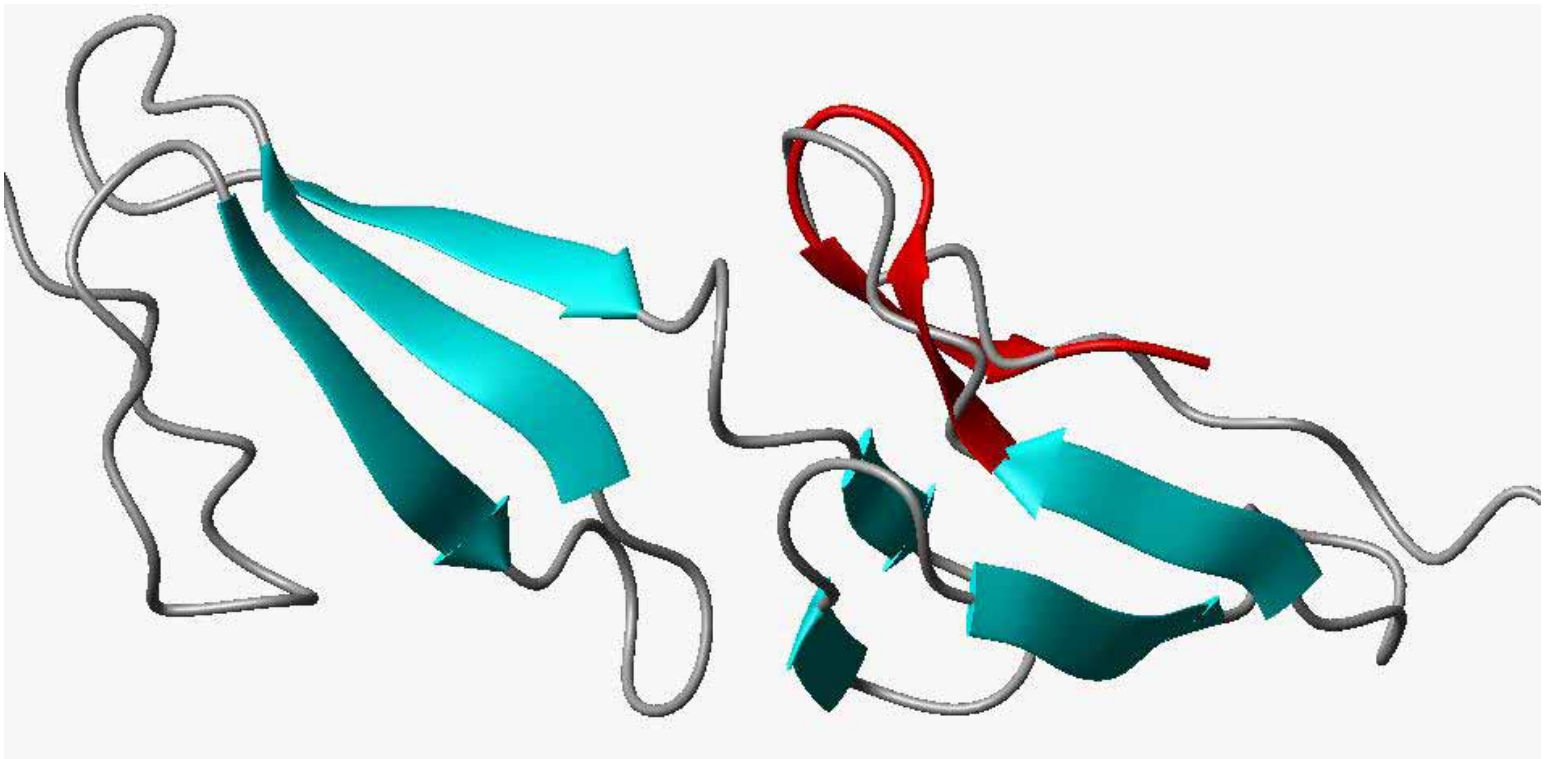


CONCLUSION:

Linear peptides do not compete EA.hy926 binding to NGRhTNF.

A. Corti, C. Longoni

Designed Active peptide Superimposed with hFN



Real World Cases

Tumor Targeting Peptides
Conclusion

Strong correlation between the Bent Conformation
and biological activity

The turn-like (Bent) conformation of the peptide is the Active one

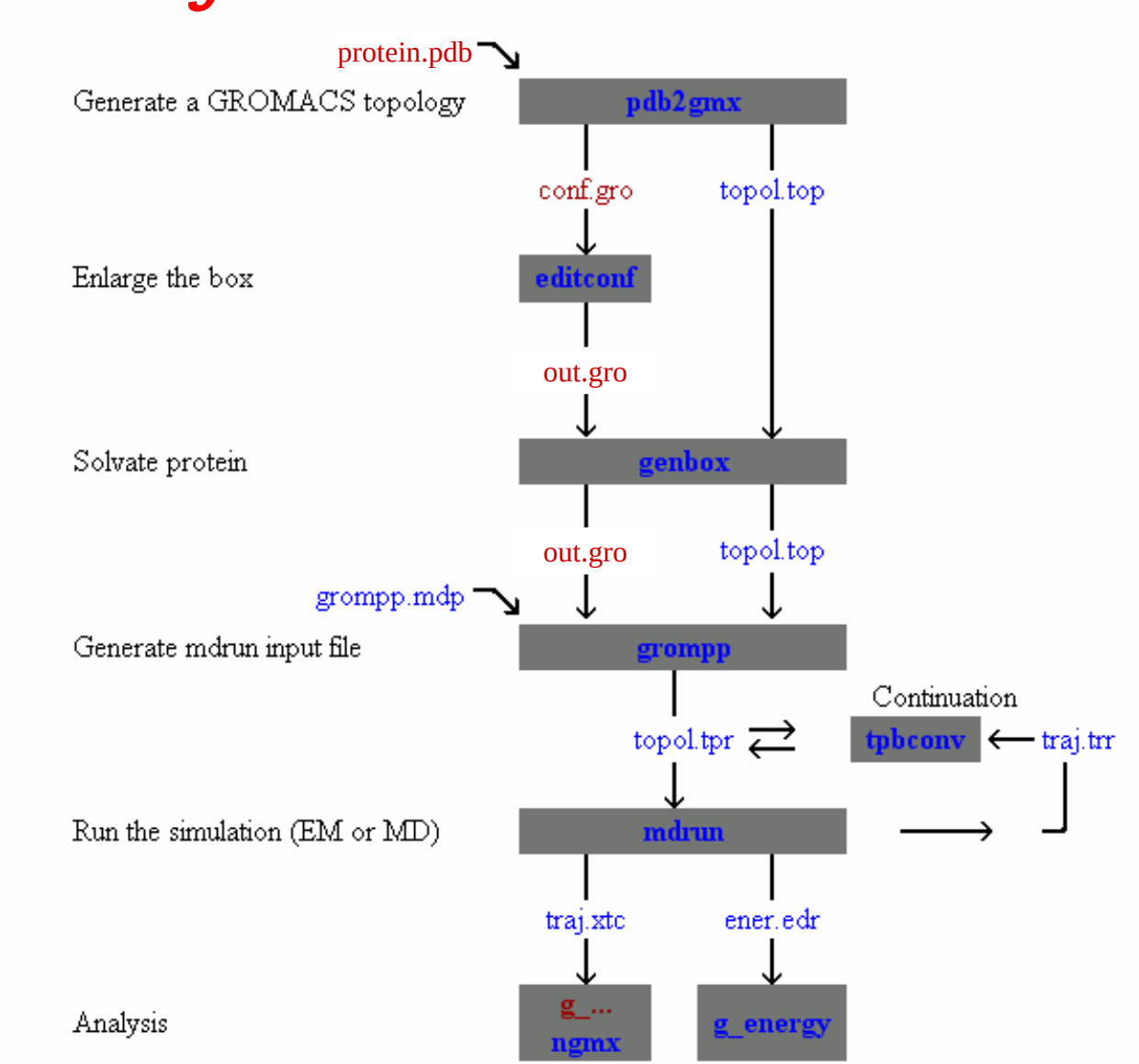
The turn propensity of the NGR sequence favors the formation
of intramolecular stabilizing interactions
Leading the linear peptide to populate an ensemble
of bent conformations recognized by the receptor

Importance for Drug-Design and mechanistic studies

GC et al, J. Biol. Chem. 2002. A. Corti et al, Mol. Immunol. 2006

Running Simulations with Gromacs. Flow Chart

www.gromacs.org



Visualize Structures on Your PC with VMD

Log on the Cluster

Change directory to your directory on the cluster.

```
% cd your_directory
```

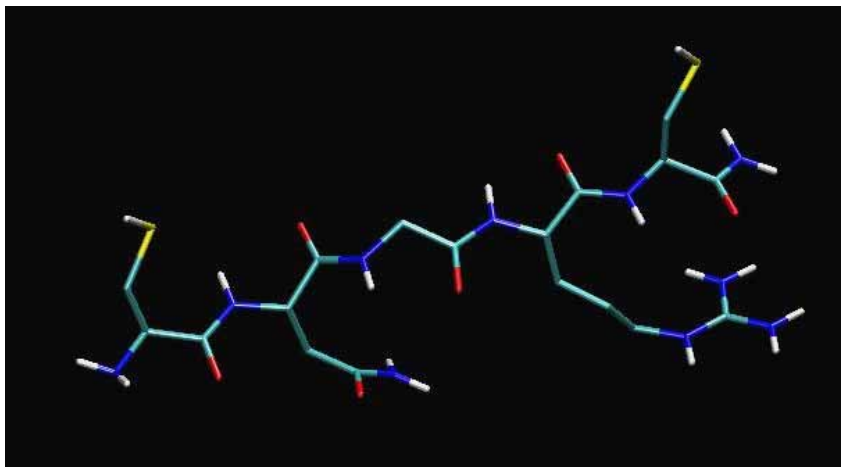
% cp $GRODATA/NO_SS/NOSS_start.pdb . Copies the starting file into your personal directory

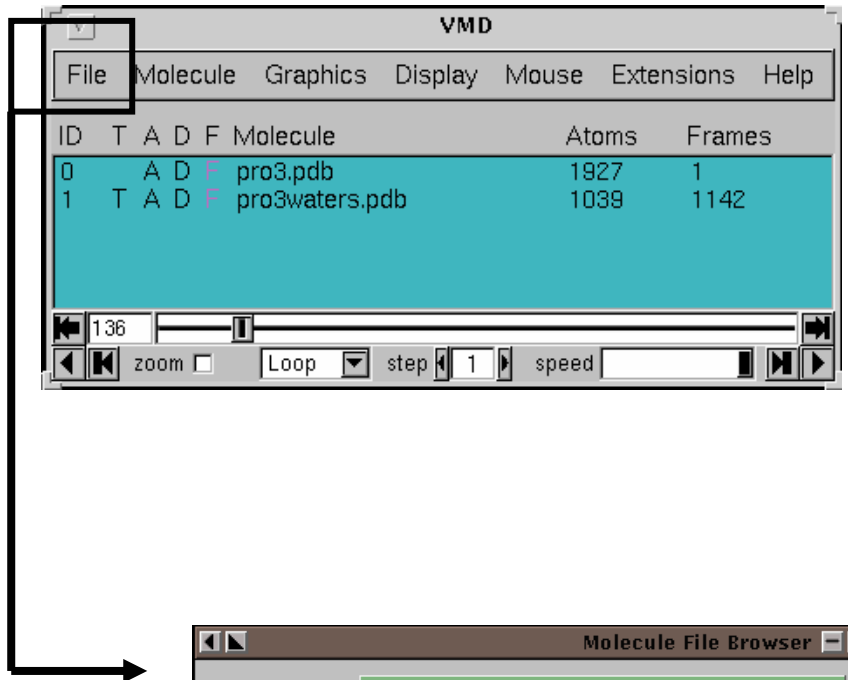
You can transfer the NOSS_start.pdb file back on your PC and check the structure with the VMD graphical program

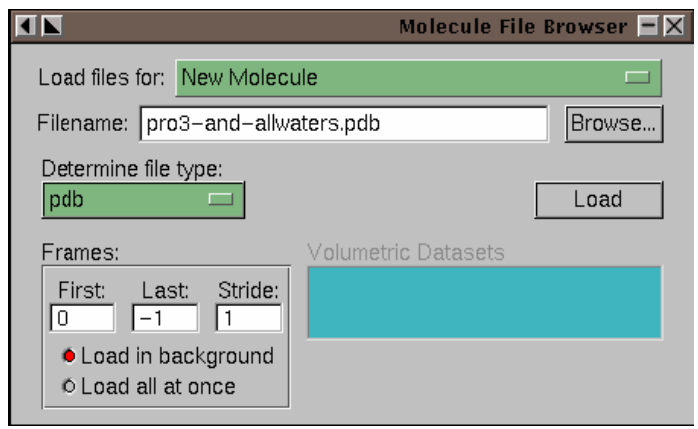
VMD program on PC in directory
C:\condivisioneNN\GROMACS con NN è il numero del PC.

Visualize Structures on Your PC with VMD

VMD. Visualization Tool
Load New Molecule NOSS_start.pdb



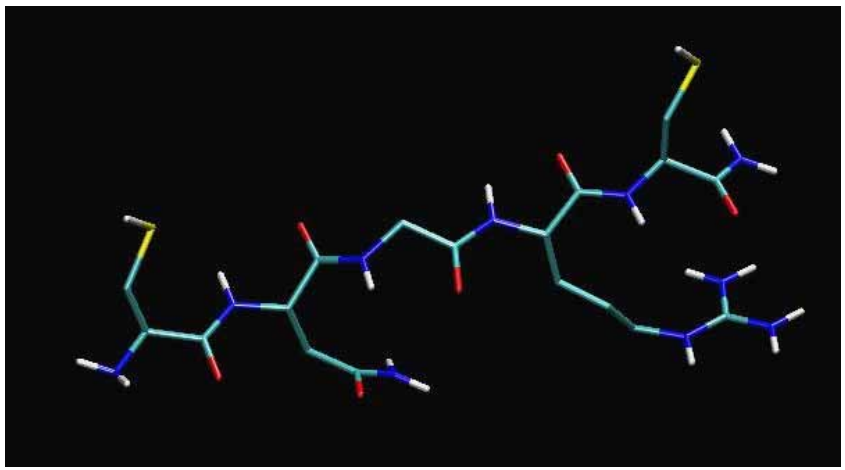


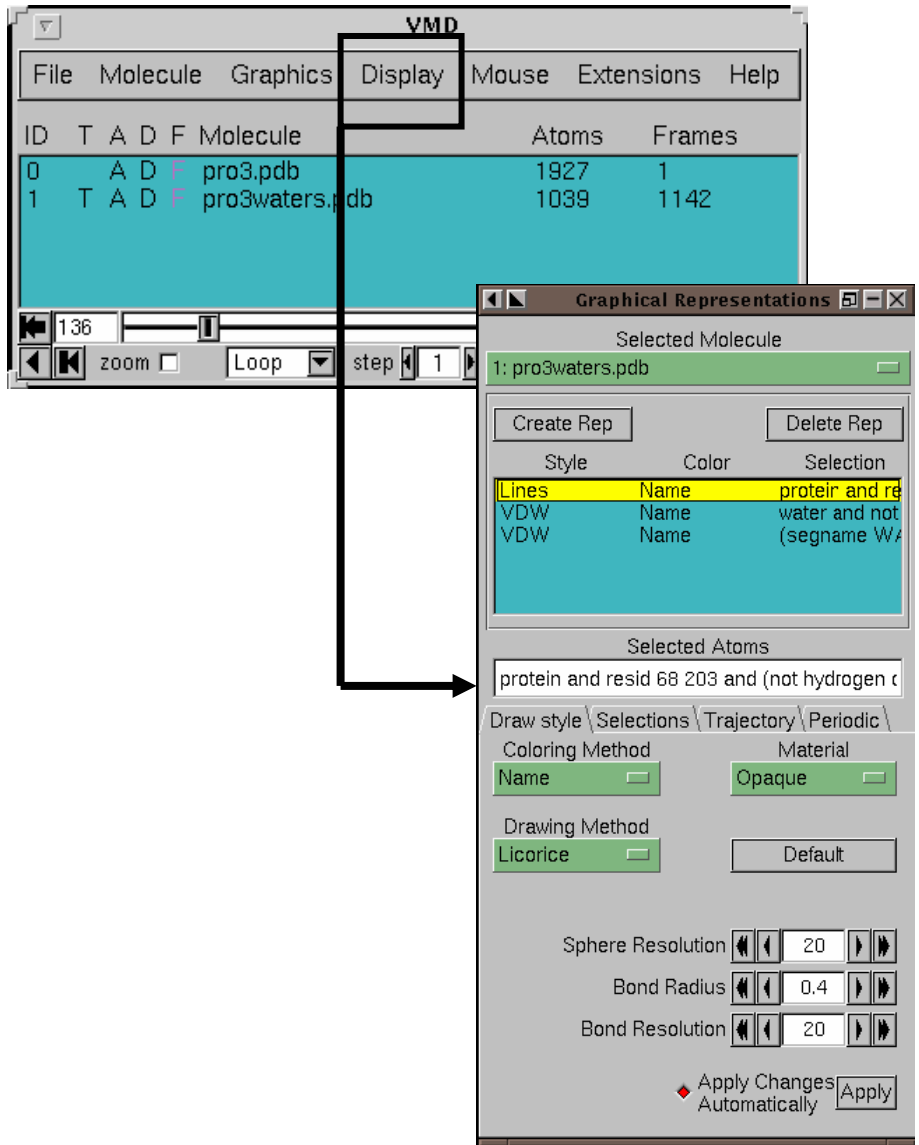


Visualize Structures on Your PC with VMD

VMD. Visualization Tool

Display New Molecule NOSS_start.pdb

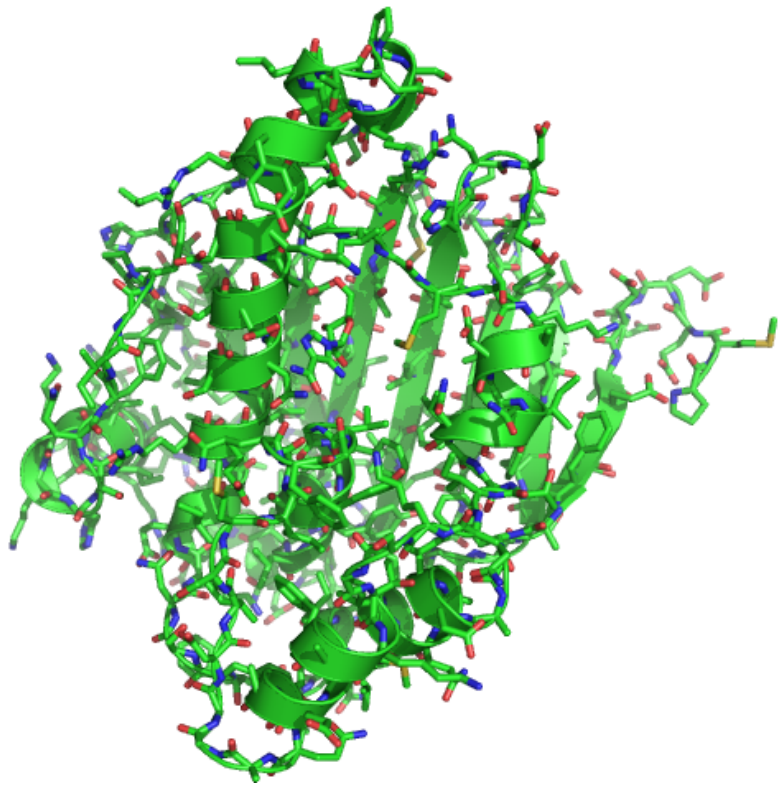




Running Simulations with Gromacs.

Protein.pdb

```
TITLE     Proteina Interessante
MODEL        1
ATOM      1  N   PRO     1      10.498 -12.747 -17.527  1.00  0.00
ATOM      2  H1  PRO     1      10.580 -12.033 -16.827  1.00  0.00
ATOM      3  H2  PRO     1      10.003 -12.317 -18.291  1.00  0.00
ATOM      4  CA  PRO     1      11.855 -13.149 -17.923  1.00  0.00
ATOM      5  CB  PRO     1      11.644 -14.588 -18.374  1.00  0.00
ATOM      6  CG  PRO     1      10.651 -15.141 -17.349  1.00  0.00
ATOM      7  CD  PRO     1       9.694 -13.951 -17.290  1.00  0.00
ATOM      8  C   PRO     1      12.548 -12.105 -18.809  1.00  0.00
ATOM      9  O   PRO     1      12.900 -11.010 -18.383  1.00  0.00
ATOM     10  N   MET     2      12.528 -12.380 -20.111  1.00  0.00
ATOM     11  H   MET     2      12.125 -13.237 -20.454  1.00  0.00
ATOM     12  CA  MET     2      13.171 -11.584 -21.162  1.00  0.00
ATOM     13  CB  MET     2      14.425 -12.313 -21.663  1.00  0.00
ATOM     14  CG  MET     2      15.625 -12.107 -20.748  1.00  0.00
ATOM     15  SD  MET     2      17.059 -13.053 -21.385  1.00  0.00
ATOM     16  CE  MET     2      18.172 -11.722 -21.786  1.00  0.00
ATOM     17  C   MET     2      12.272 -11.026 -22.271  1.00  0.00
ATOM     18  O   MET     2      12.439 -11.436 -23.413  1.00  0.00
ATOM     19  N   GLU     3      11.145 -10.420 -21.916  1.00  0.00
ATOM     20  H   GLU     3      10.950 -10.210 -20.958  1.00  0.00
ATOM     21  CA  GLU     3      10.049 -10.056 -22.827  1.00  0.00
ATOM     22  CB  GLU     3       8.724 -10.738 -22.497  1.00  0.00
ATOM     23  CG  GLU     3       8.077 -10.653 -21.113  1.00  0.00
ATOM     24  CD  GLU     3       8.790 -11.394 -19.977  1.00  0.00
ATOM     25  OE1 GLU     3       8.773 -12.641 -20.004  1.00  0.00
ATOM     26  OE2 GLU     3       9.449 -10.751 -19.144  1.00  0.00
```



Running Simulations with Gromacs.

Transform a PDB file in Gromacs Format.

Log on the Cluster

Change directory to your directory on the cluster node.

```
% cd your_directory
```

% G33 to start up the gromacs environment

Help for Gromacs

```
% gmx_command -h
```

Copy the NOSS_start.pdb file to your directory (the starting pdb structure)

```
% pdb2gmx -f NOSS_start.pdb -p NOSS.top -o NOSS.gro -inter
```

translates pdb into a gromacs file format and generates the topology file

You can transfer the NOSS_start.pdb file back on your PC and check the structure with the VMD graphical program

pdb2gmx will give you several options:

Select the Force Field:
0: GROMOS96 43a1 Forcefield (official distribution)
1: Gromacs Forcefield with all hydrogens (proteins only)
2: Gromacs Forcefield (see manual)
3: GROMOS96 43b1 Vacuum Forcefield (official distribution)
4: GROMOS96 43a2 Forcefield (development) (improved alkane dihedrals)

Type 0 to select the force field

Running Simulations with Gromacs.

Conf.gro

```
MD of 2 waters, t= 0.0
6
1WATER OW1 1 0.126 1.624 1.679 0.1227 -0.0580 0.0434
1WATER HW2 2 0.190 1.661 1.747 0.8085 0.3191 -0.7791
1WATER HW3 3 0.177 1.568 1.613 -0.9045 -2.6469 1.3180
2WATER OW1 4 1.275 0.053 0.622 0.2519 0.3140 -0.1734
2WATER HW2 5 1.337 0.002 0.680 -1.0641 -1.1349 0.0257
2WATER HW3 6 1.326 0.120 0.568 1.9427 -0.8216 -0.0244
1.82060 1.82060 1.82060
```

Lines contain the following information (top to bottom):
title string (free format string, optional time in ps after 't=')
number of atoms (free format integer)
one line for each atom (fixed format, see below)
box vectors (free format, space separated reals), values: v1(x) v2(y) v3(z) v1(y) v1(z) v2(x) v2(z) v3(x) v3(y), the last 6 values may be omitted (they will be set to zero).
Gromacs only supports boxes with v1(y)=v1(z)=v2(z)=0.

Running Simulations with Gromacs.

Topol.top

```
;
;           File 'topol.top' was generated
;           By user: giorgio (503)
;           On host: ufo.public
;           At date: Fri Jun  9 17:50:59 2006
;
;           This is your topology file
;           CYTOCHROME C
;
; Include forcefield parameters
;#include "ffG43a1.itp"
#include "mybon.itp"
;#include "mynb.itp"

[ moleculetype ]
; Name            nrexcl
Protein             3

[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge       mass  typeB    chargeB      massB
     1        CH3      1    ACE     CA      1          0     15.035   ; qtot 0
     2          C      1    ACE      C      2       0.38     12.011   ; qtot 0.38
     3          O      1    ACE      O      2      -0.38    15.9994   ; qtot 0
     4          N      2    GLY      N      3      -0.28    14.0067   ; qtot -0.28
     5          H      2    GLY      H      3       0.28      1.008   ; qtot 0
     6        CH2      2    GLY     CA      4          0     14.027   ; qtot 0
     7          C      2    GLY      C      5       0.38     12.011   ; qtot 0.38
     8          O      2    GLY      O      5      -0.38    15.9994   ; qtot 0
     9          N      3    ASP      N      6      -0.28    14.0067   ; qtot -0.28
    10          H      3    ASP      H      6       0.28      1.008   ; qtot 0
    11        CH1      3    ASP     CA      7          0     13.019   ; qtot 0
    12        CH2      3    ASP     CB      7          0     14.027   ; qtot 0
```

Running Simulations with Gromacs.

Topol.top

```
[ bonds ]
;  ai    aj funct            c0            c1            c2            c3
    1     2     2    gb_26
    2     3     2    gb_4
    2     4     2    gb_9
    4     5     2    gb_2
    4     6     2    gb_20
    6     7     2    gb_26
    7     8     2    gb_4
    7     9     2    gb_9
    9    10     2    gb_2
    9    11     2    gb_20

[ angles ]
;  ai    aj    ak funct            c0            c1            c2            c3
    1     2     3     2
    1     2     4     2    ga_18
    3     2     4     2    ga_32
    2     4     5     2    ga_31
    2     4     6     2    ga_30
    5     4     6     2    ga_17
    4     6     7     2    ga_12
    6     7     8     2    ga_29
    6     7     9     2    ga_18
    8     7     9     2    ga_32
    7     9    10     2    ga_31
    7     9    11     2    ga_30

[ dihedrals ]
;  ai    aj    ak    al funct            c0            c1            c2            c3            c4            c5
    1     2     4     6     1    gd_4
    2     4     6     7     1    gd_19
    4     6     7     9     1    gd_20
    6     7     9    11     1    gd_4
    7     9    11    16     1    gd_19
    9    11    12    13     1    gd_17
    9    11    16    18     1    gd_20
```

Running Simulations with Gromacs.

Remove strain and bad contacts.
Use *grompp* to create a run input file.

Copy file minim.mdp from $GRODATA to your working directory.

```
% grompp -f minim.mdp -c NOSS.gro -p NOSS.top -o NOSS_MIN.tpr
```

Use *mdrun* to actually minimize.

```
% mdrun -v -s NOSS_MIN.tpr -o minim_traj.trr -c minimized.gro -e
minim_ener.edr
```

To visualize the energy variation:

```
% g_energy -f minim_ener.edr -o minim_ener.xvg
```

To transform an .xvg file in a format readable by windows: .dat file

```
% sed "s/@/#/" minim_ener.xvg > minim_ener.dat
```

Transfer *minim_ener.dat* back to your PC and check it with *gnuplot*
Transfer *minimized.gro* back to your PC and check it with VMD

Running Simulations with Gromacs.
Command file

*.mdp

```
 ; VARIOUS PREPROCESSING OPTIONS =
title                    =
cpp                      = /lib/cpp
include                  =
define                   =

; RUN CONTROL PARAMETERS =
integrator               = md
; start time and timestep in ps =
tinit                    = 0.0
dt                       = 0.002
nsteps                   = 50000000
; number of steps for center of mass motion removal =
nstcomm                  = 1
comm-grps                =

; LANGEVIN DYNAMICS OPTIONS =
; Temperature, friction coefficient (amu/ps) and random seed =
bd-temp                  = 300
bd-fric                  = 0
ld-seed                  = 1993

; ENERGY MINIMIZATION OPTIONS =
; Force tolerance and initial step-size =
emtol                    = 100
emstep                   = 0.01
; Max number of iterations in relax_shells =
niter                    = 20
; Frequency of steepest descents steps when doing CG =
nstcgsteep               = 1000
```

Running Simulations with Gromacs.
Command file

*.mdp

```
; OUTPUT CONTROL OPTIONS =
; Output frequency for coords (x), velocities (v) and forces (f) =
nstxout                  = 10000
nstvout                  = 10000
nstfout                  = 0
; Output frequency for energies to log file and energy file =
nstlog                   = 1000
nstenergy                = 1000
; Output frequency and precision for xtc file =
nstxtcout                = 1000
xtc-precision            = 1000
; This selects the subset of atoms for the xtc file. You can =
; select multiple groups. By default all atoms will be written. =
xtc_grps                 = Protein
; Selection of energy groups =
energygrps               = Protein SOL

; NEIGHBORSEARCHING PARAMETERS =
; nblist update frequency =
nstlist                  = 5
; ns algorithm (simple or grid) =
ns_type                  = grid
; Periodic boundary conditions: xyz or none =
pbc                      = xyz
; nblist cut-off         =
rlist                    = 0.8
domain-decomposition     = no
```

Running Simulations with Gromacs.
Command file

*.mdp

```
; OPTIONS FOR ELECTROSTATICS AND VDW =
; Method for doing electrostatics =
coulombtype              = cut-off
rcoulomb-switch          = 0
rcoulomb                 = 1.4
; Dielectric constant (DC) for cut-off or DC of reaction field =
epsilon-r                = 1
; Method for doing Van der Waals =
vdw-type                 = Cut-off
; cut-off lengths        =
rvdw-switch              = 0
rvdw                     = 0.8
; Apply long range dispersion corrections for Energy and Pressure =
DispCorr                 = No
; Spacing for the PME/PPPM FFT grid =
fourierspacing           = 0.12
; FFT grid size, when a value is 0 fourierspacing will be used =
fourier_nx               = 0
fourier_ny               = 0
fourier_nz               = 0
; EWALD/PME/PPPM parameters =
pme_order                = 4
ewald_rtol               = 1e-05
epsilon_surface          = 0
optimize_fft             = no
```

Running Simulations with Gromacs.
Command file

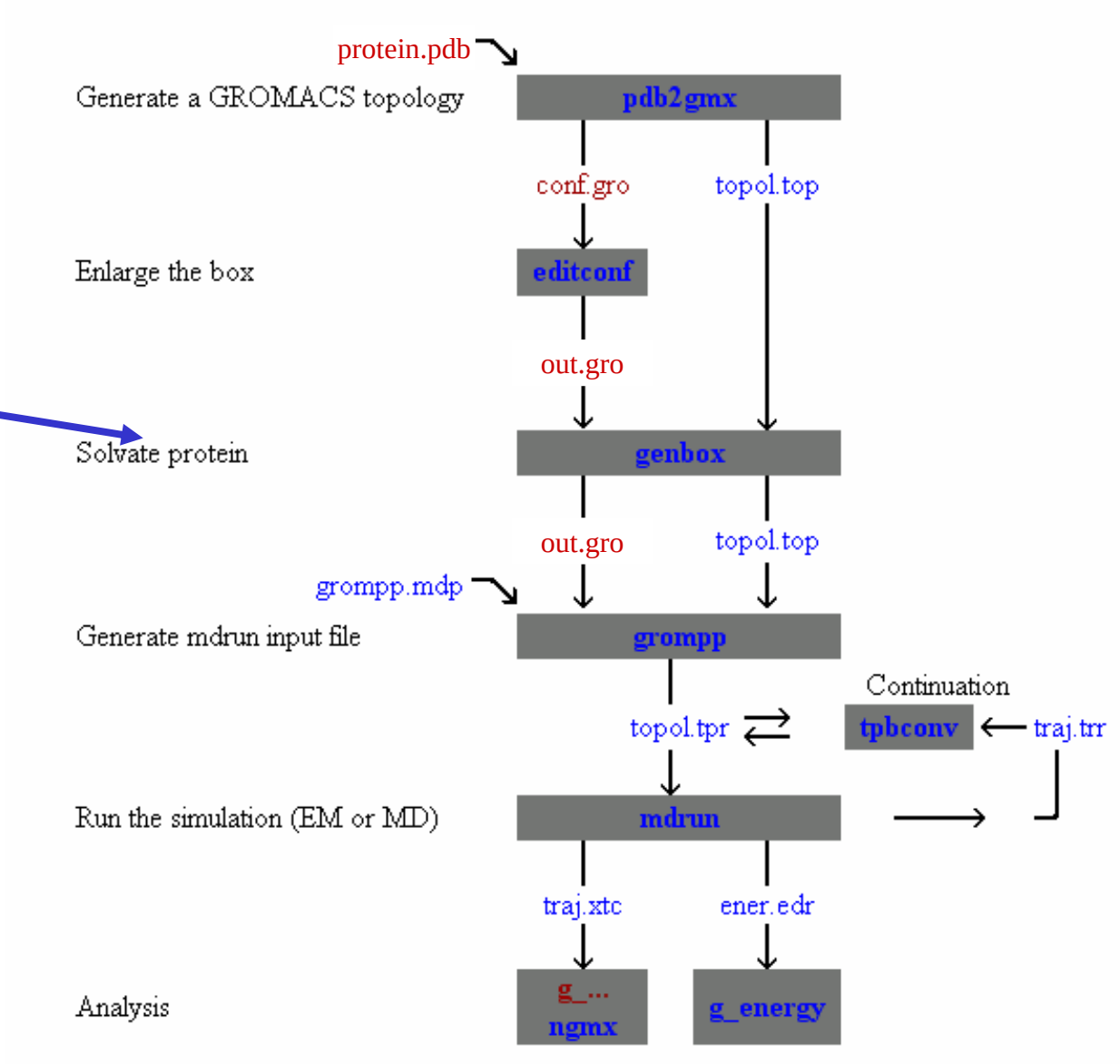
*.mdp

```
; OPTIONS FOR WEAK COUPLING ALGORITHMS =
; Temperature coupling   =
Tcoupl                   = yes
; Groups to couple separately =
tc-grps                  = protein  SOL
; Time constant (ps) and reference temperature (K) =
tau_t                    = 0.05     0.05
ref_t                    = 300      300
; Pressure coupling      =
Pcoupl                   = no
Pcoupltype               = Isotropic
; Time constant (ps), compressibility (1/bar) and reference P (bar) =
tau_p                    = 0.5
compressibility          = 4.5e-5
ref_p                    = 1.0

; SIMULATED ANNEALING CONTROL =
annealing                = no no
; Time at which temperature should be zero (ps) =
zero-temp_time           = 0

; GENERATE VELOCITIES FOR STARTUP RUN =
gen-vel                  = yes
gen-temp                 = 300
gen-seed                 = 250371
```

Running Simulations with Gromacs. Command file



Running Simulations with Gromacs.

Solvate the Peptide. Define the Box and add water.

Generate a cubic box.

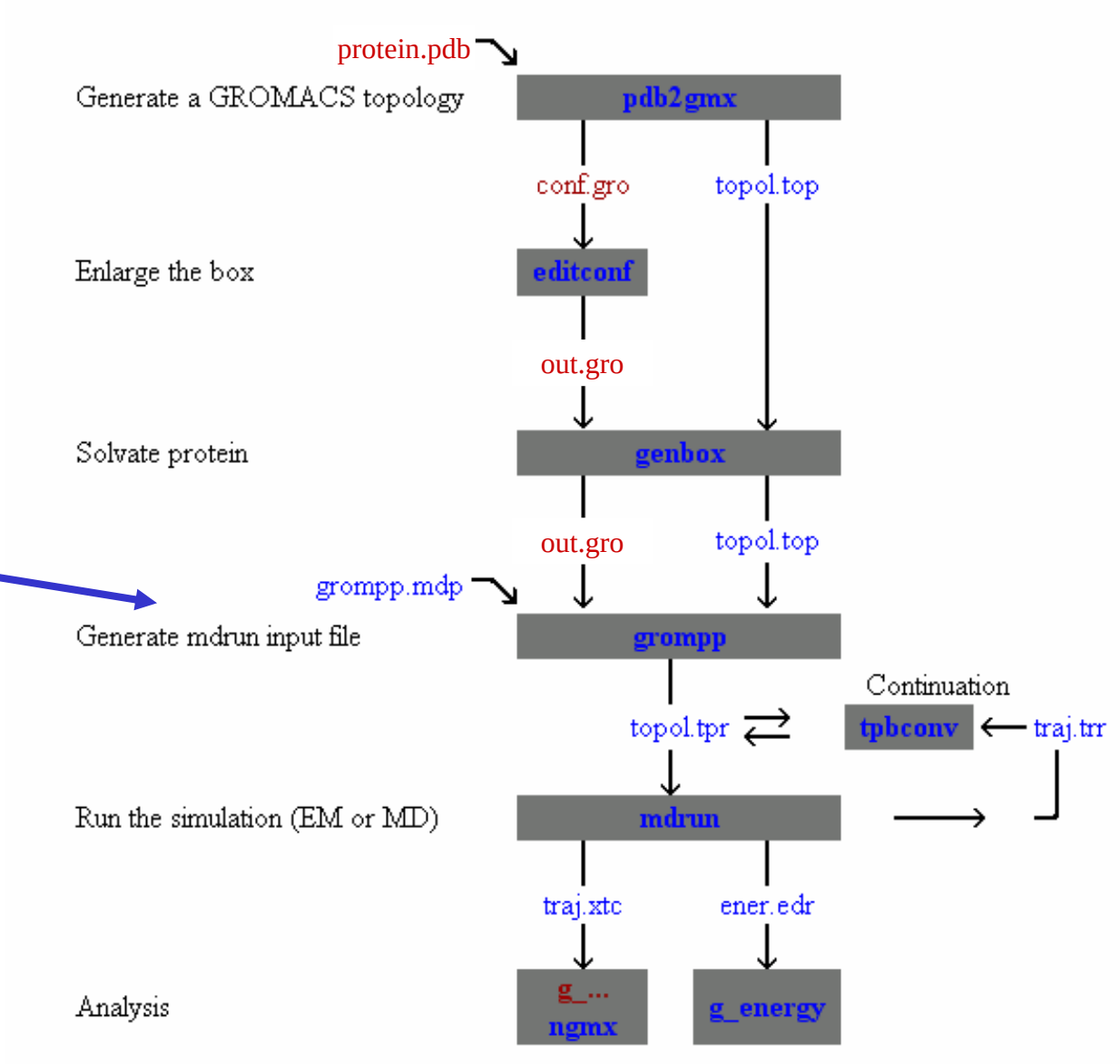
```
% editconf –f minimized.gro –o minimized_box.gro –d 0.7 –bt cubic
```

Fill it with water.

```
% genbox –cp minimized_box.gro –cs –o minimized_water.gro –p NOSS.top
```

The number of added water molecules is added automatically to the topology file.
Transfer *minimized_water.gro* back to your PC and check it with VMD.

Running Simulations with Gromacs. Analysis



Minimize the water solvated box and run the simulation

Remove strain and bad contacts.
Use *grompp* to create a run input file.

```
% grompp –f minim.mdp –c minimized_water.gro -p NOSS.top –o
NOSS_MIN.tpr
```

Use *mdrun* to actually minimize.

```
% mdrun –v –s NOSS_MIN.tpr –o minim_traj.trr –c
minimized_water_1.gro
```

Start the actual simulation.
Copy the file fullmd_sol.mdp to your working directory

```
% grompp –f fullmd_sol.mdp –c minimized_water_1.gro -p NOSS.top
–o NOSS_MD.tpr

% mdrun –v –s NOSS_MD.tpr
```

ANALYSIS. 1 rmsd

For the analysis we will use pre-run trajectories available in $GRODATA/NO_SS

In general the gromacs analysis tools work like:

```
%
g_rms –f $GRODATA/NOSS_300ns_100.xtc –s $GRODATA/topol_NOSS.tpr
-o rmsd_NOSS.xvg -other options
```

To transform an .xvg file in a format readable by windows: .dat file

```
% sed “s/@/#/” rmsd_NOSS.xvg > rmsd_NOSS.dat
```

Transfer *rmsd_NOSS.dat* back to your PC and check it with *gnuplot*

For help

```
% g_cluster –h
```

ANALYSIS. 2 structural clustering

For the analysis we will use pre-run trajectories available in $GRODATA/NO_SS

Cluster Analysis

```
%
g_cluster –f $GRODATA/NOSS_300ns_100.xtc –s
$GRODATA/topol_NOSS.tpr –g cluster.log –cl clusters_NOSS.pdb –
cutoff 0.15 –method gromos
```

To transform an .xvg file in a format readable by windows: .dat file

```
% sed “s/@/#/” rmsd_NOSS.xvg > rmsd_NOSS.dat
```

Transfer *clusters_NOSS.pdb* back to your PC and check it with *VMD*

For help

```
% g_rms –h
```

ANALYSIS. 2 structural clustering

For the analysis we will use pre-run trajectories available in $GRODATA/NO_SS

Secondary Structure Analysis

```
%
do_dssp -f $GRODATA/NOSS_300ns_100.xtc -s $GRODATA/topol_NOSS.tpr
-o ss_NOSS.xpm

xpm2ps -f ss_NOSS.xpm -di $GRODATA/ss.m2p -o plot_NOSS.eps
```

To transform an .xpm file in a format readable by windows: .eps file

Transfer *plot_NOSS.eps* back to your PC and check it with *ghostview. Double click on the plot_NOSS.eps icon*

For help

```
% do_dssp -h
```