

Franca Fraternali Professor in Bioinformatics and Computational Biology



Contact details:

Telephone: +44 (0)20 7848 6843

E-mail: franca.fraternali@kcl.ac.uk

Department

Randall Division of Cell & Molecular Biophysics

Office hours: Tuesdays 14-16 pm

New Hunt's House, Guy's Campus
SE1 1UL
third floor room 3.12



Research Interests:

Computational Biology: Protein Dynamics and Stability

Structural Bioinformatics

Protein Structure Predictions

Protein-Protein Interactions

Molecular Simulations of Proteins and Nucleic Acids

Agenda 1st Lecture

Barriers in Biology: Protein Dynamics

Molecular Mechanics-MD

Equilibrium MD: Ensembles

How to maintain equilibrium conditions in MD

Liouville's Theorem-Ergodicity

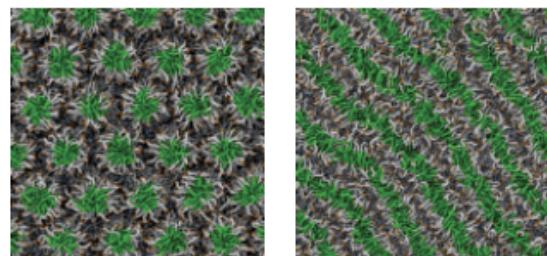
Does MD always reach equilibrium?

Molecular Transitions are rare events

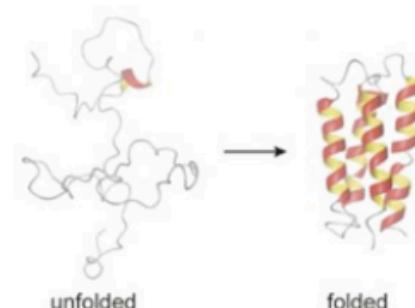
chemical reactions



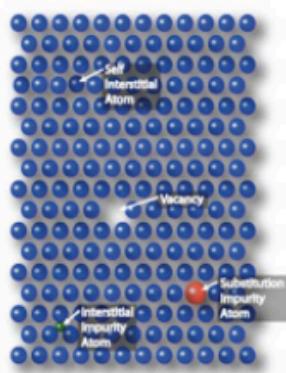
phase transitions



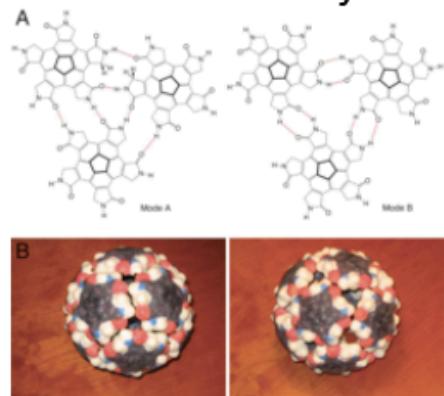
protein folding



defect diffusion



self-assembly

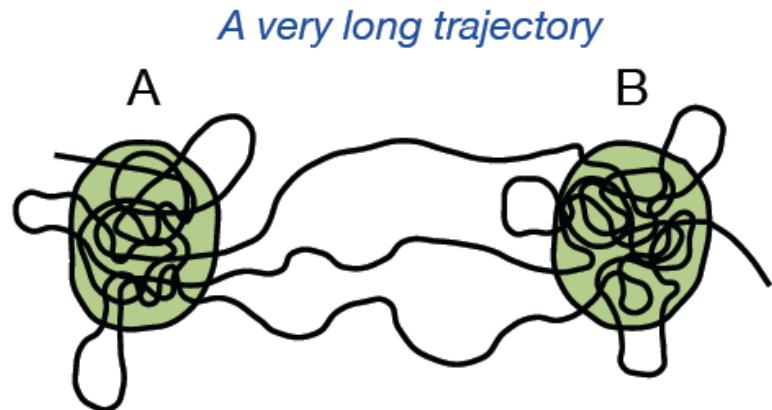


stock exchange crash



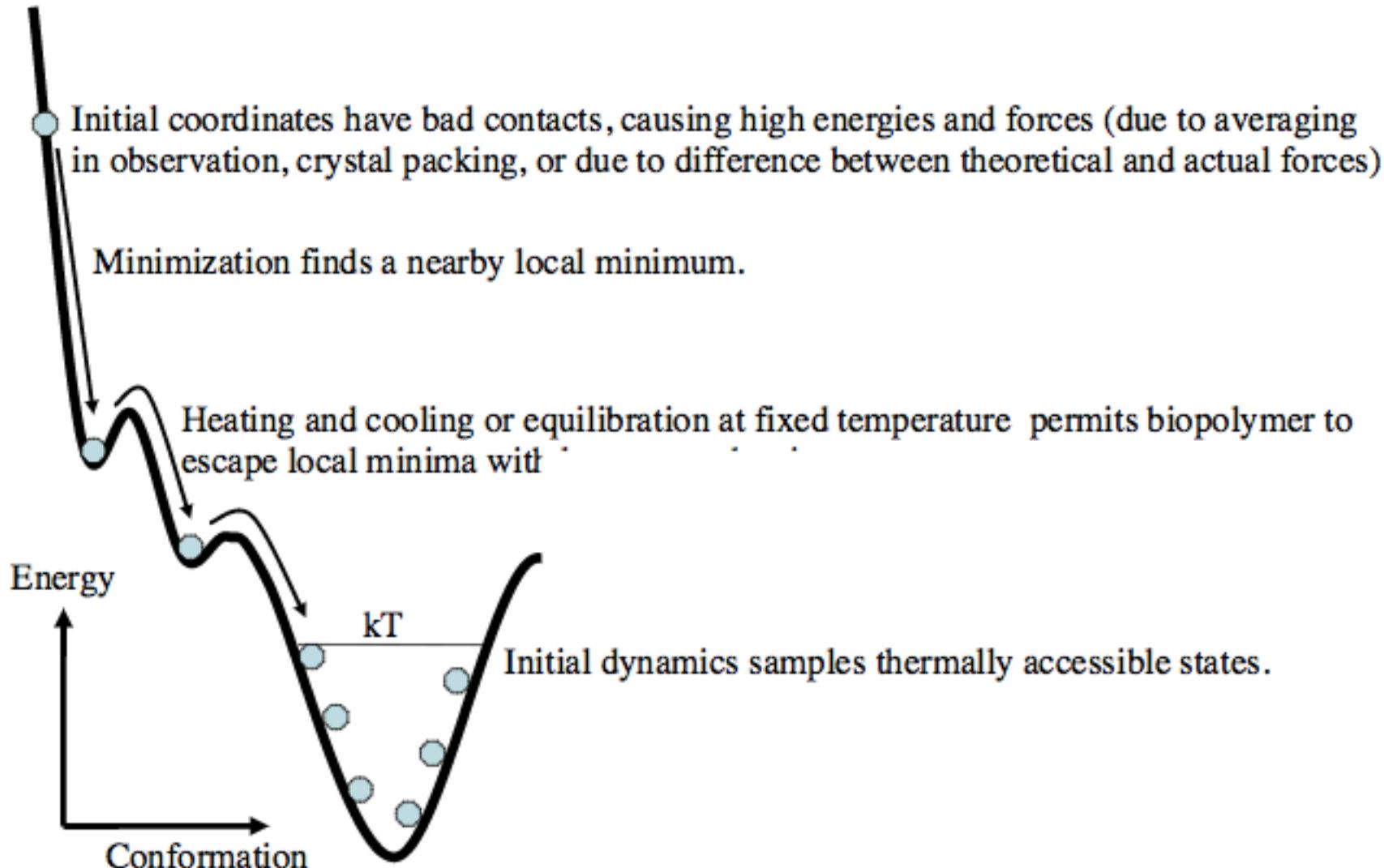
Molecular Transitions NOT SO rare....

- Molecular transitions are not rare in every day life
- They are only rare events with respect to the femto-second timescale of atomic motions
- Modeling activated transitions by straightforward simulation would be extremely costly (takes forever)
- Advanced methods: Transition state theory (TST), Bennet-Chandler (Reactive Flux) approach, Transition Path Sampling (TPS), Free energy methods, Parallel Tempering, String Method,....

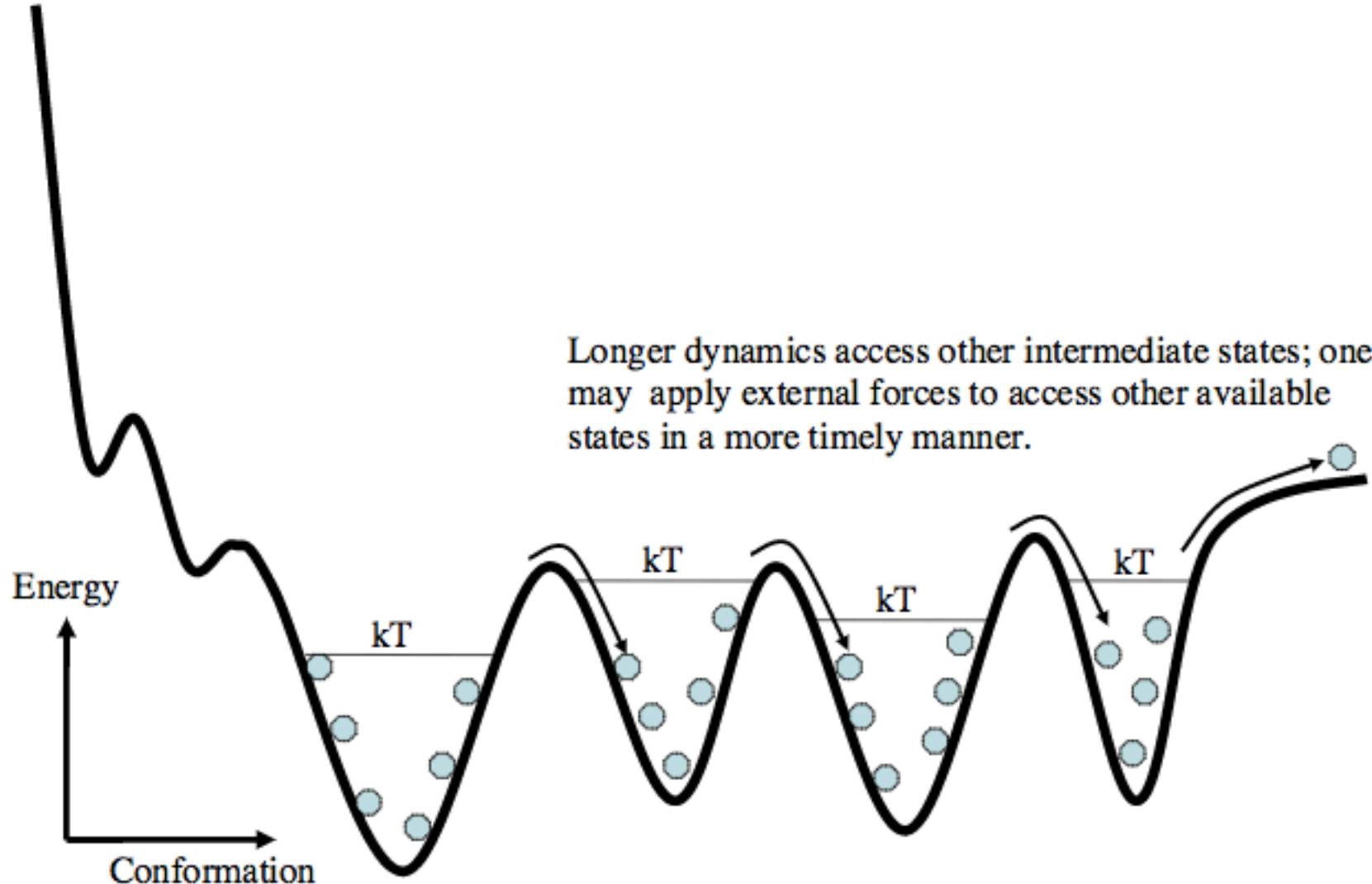


- A and B are (meta-)stable states, i.e. attractive basins
- transitions between A and B are rare
- transitions can happen fast
- system loses memory in A and B

From the Mountains to the Valleys



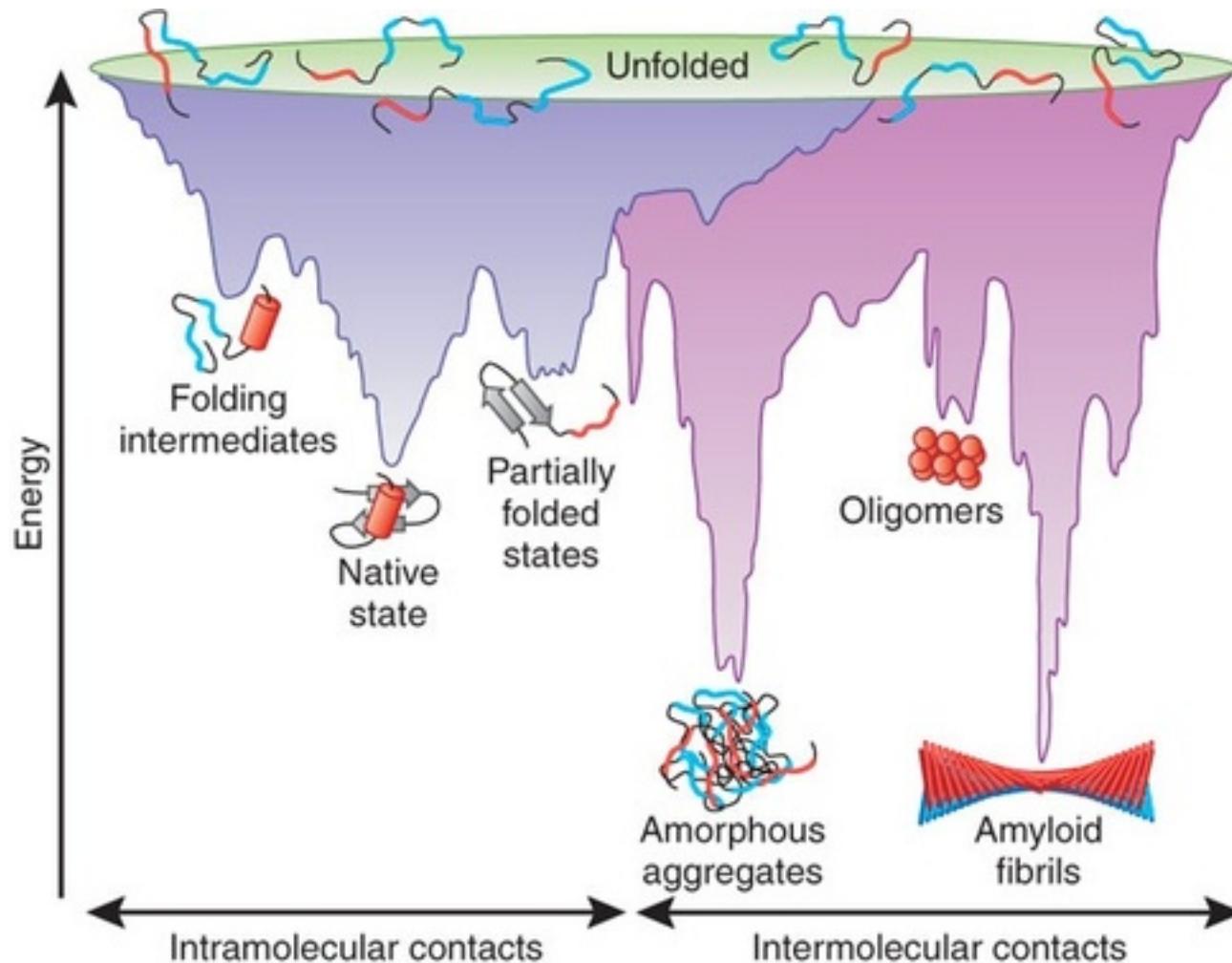
From the Mountains to the Valleys



The landscape is more complicated than this!

Folding - assembly

We aim at sampling native-like configurations and/or observable processes



A note of warning!



Simulation can replace or complement an experiment:

- | | |
|--------------------------------|---|
| 1. experiment
is impossible | collision of stars or galaxies
weather forecast |
| 2. experiment
is dangerous | flight simulation
explosion simulation |
| 3. experiment
is expensive | high pressure simulation
wind channel simulation |
| 4. experiment
is blind | many properties cannot be observed on very
short time scales and very small space scales |
-

Why do we study protein dynamics?

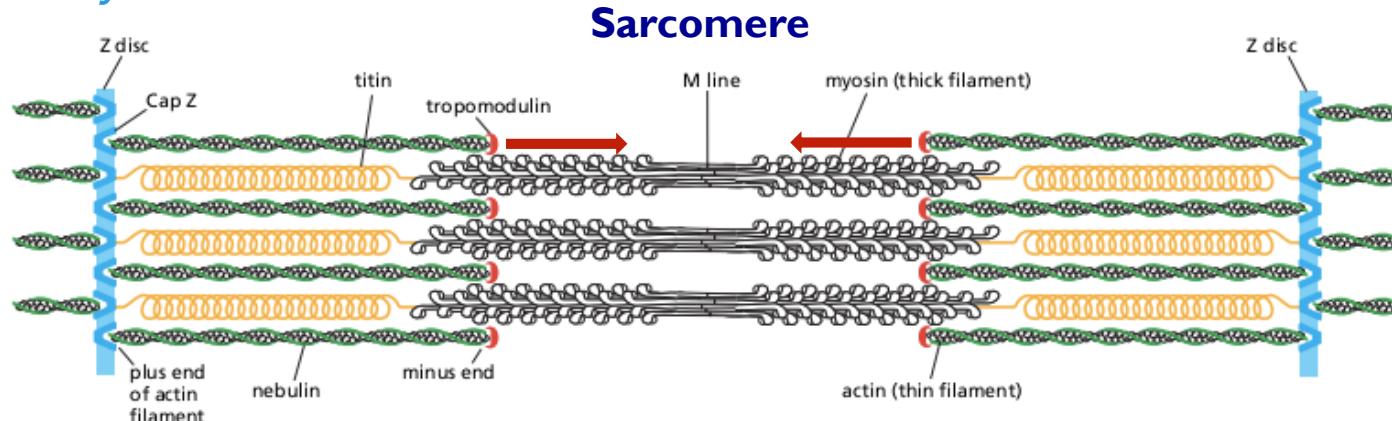
Protein dynamics is essential for protein function. In particular, the dynamical properties of proteins are a key element in:

- Conformational changes in enzymes
- Molecular motors
- Protein folding
- Protein-protein interactions and ligand binding
- Allosteric regulation of proteins
- Transport across membranes

Importance of Dynamics for Protein Function

Molecular Motors

Myosin II in muscle



Alberts et al., Molecular Biology of the Cell, 2008

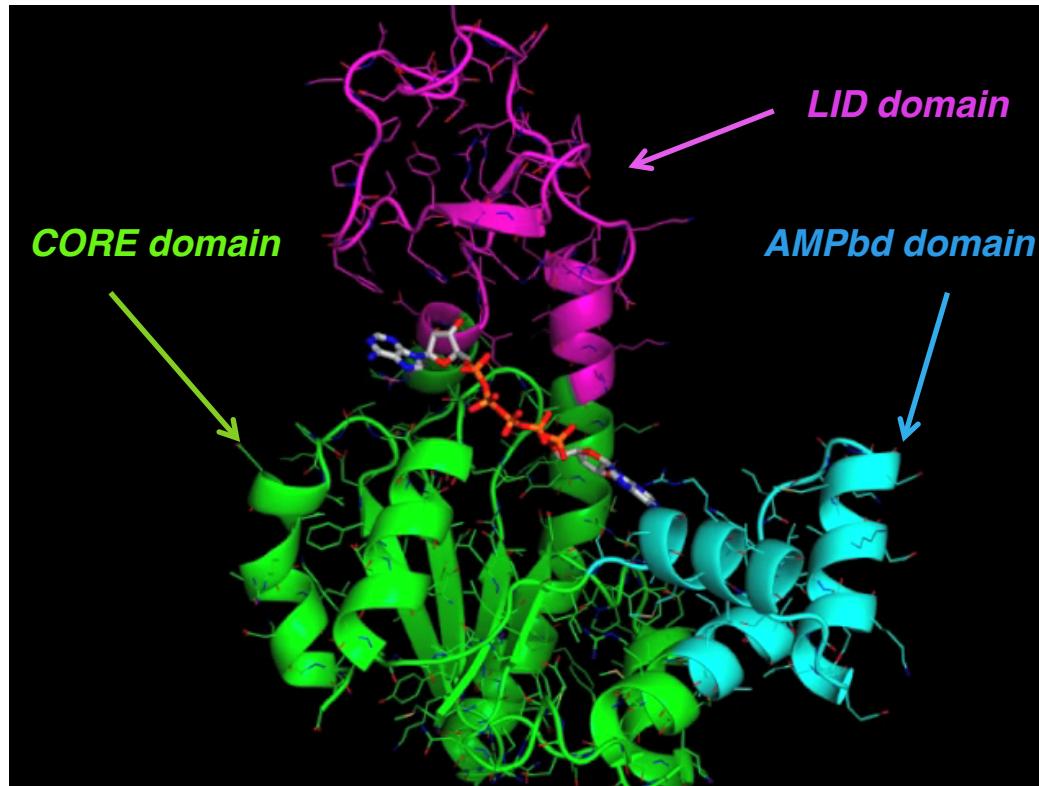


Actin and Myosin Action

<https://www.youtube.com/watch?v=NRzJjx3ANuE>

Importance of Dynamics for Protein Function

Conformational changes in enzymes



Matsunaga, Y. et al. *PLoS Comput Biol* 8, e1002555 (2012).

Importance of Dynamics for Protein Function

Conformational changes in enzymes

Triosephosphate Isomerase

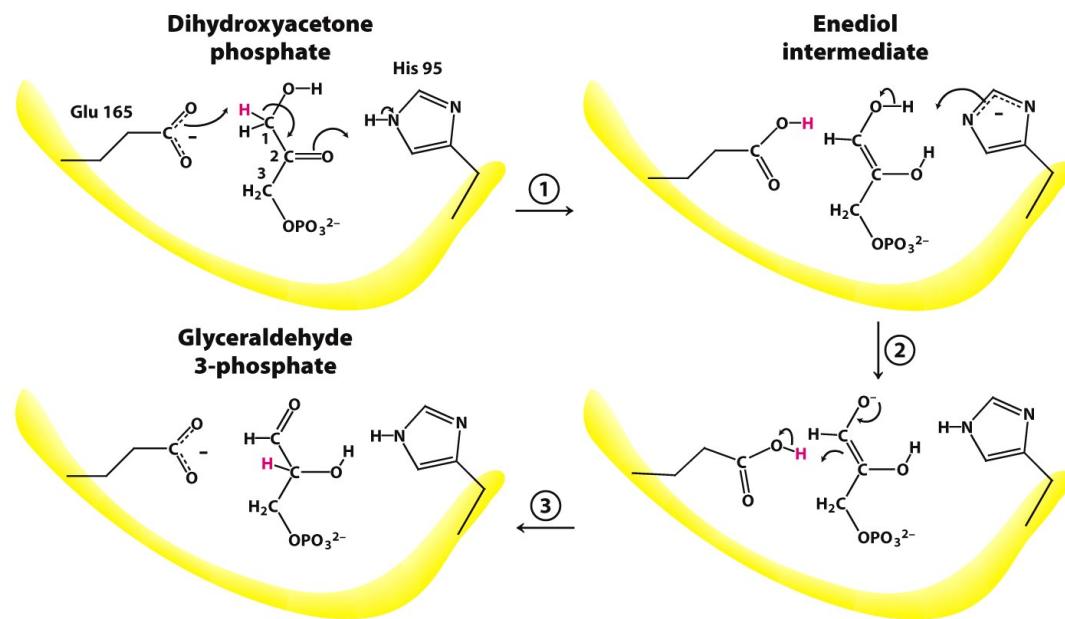
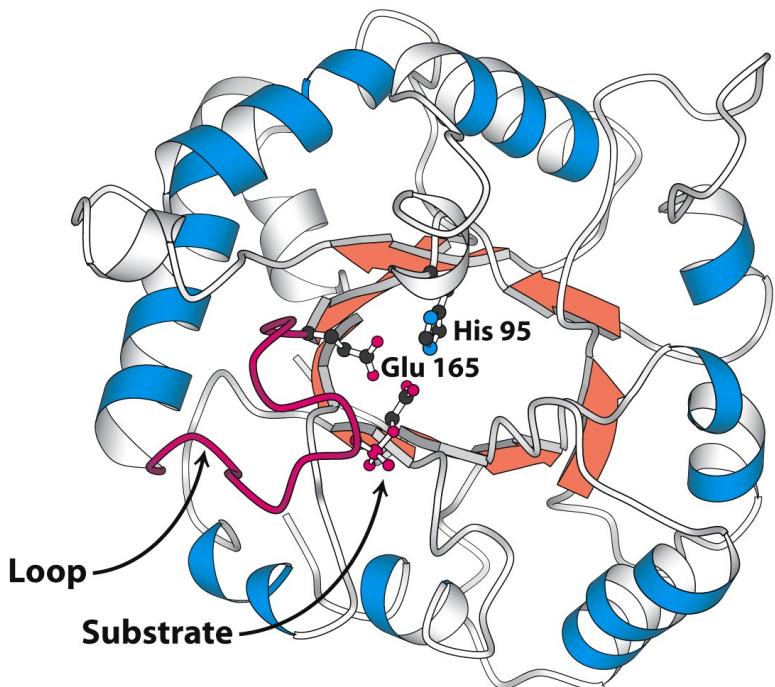
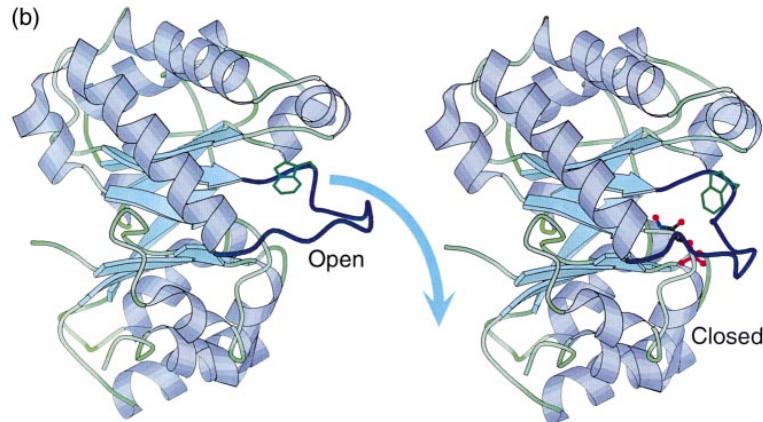


Figure 16.4
Biochemistry, Seventh Edition
© 2012 W.H. Freeman and Company

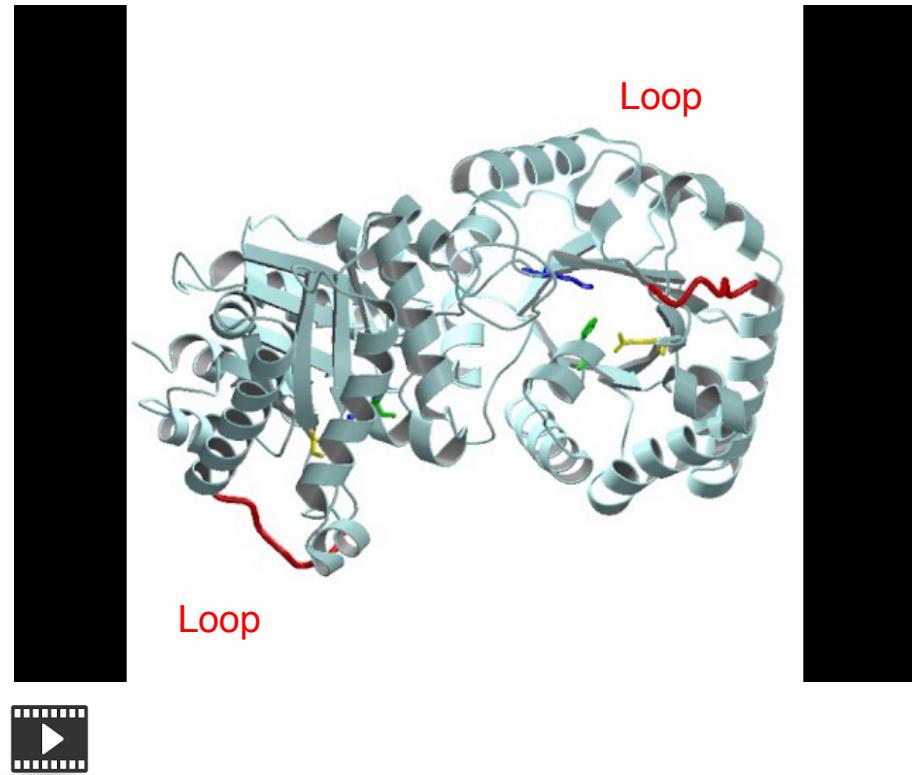
Importance of Dynamics for Protein Function

Conformational changes in enzymes

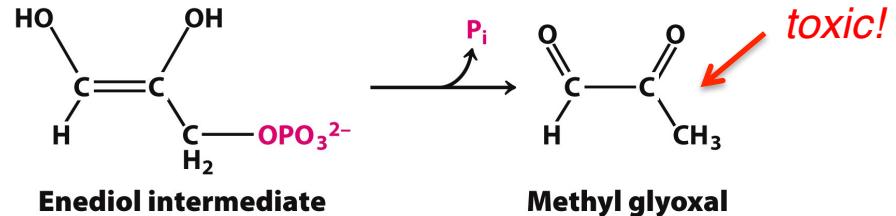
Triosephosphate Isomerase



Rozovsky, S. & McDermott, A. E. J. Mol. Biol.
310, 259 (2001).

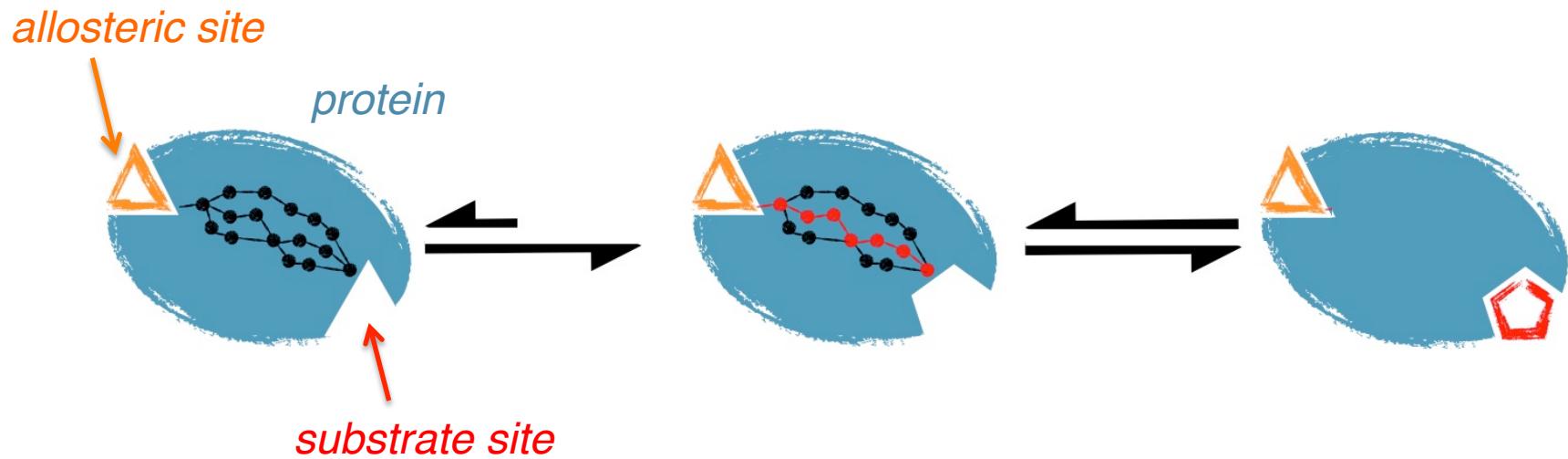


The loop closure over the active site upon substrate binding is necessary to stabilise the reaction intermediate and protect it from bulk water



Importance of Dynamics for Protein Function

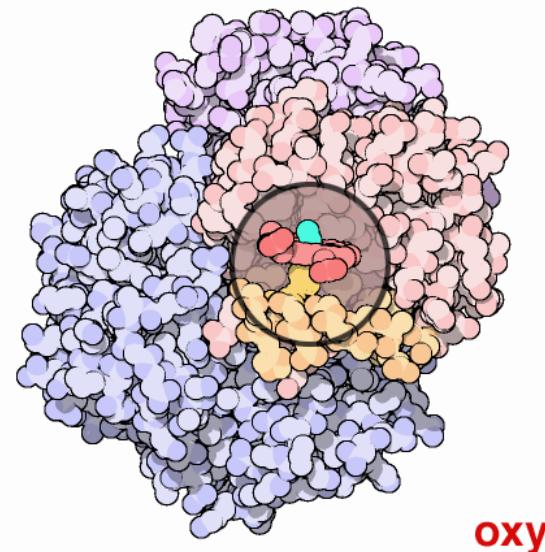
Allosteric regulation of proteins



Importance of Dynamics for Protein Function

Allosteric regulation of proteins

Hemoglobin



Cooperation Makes It Easier

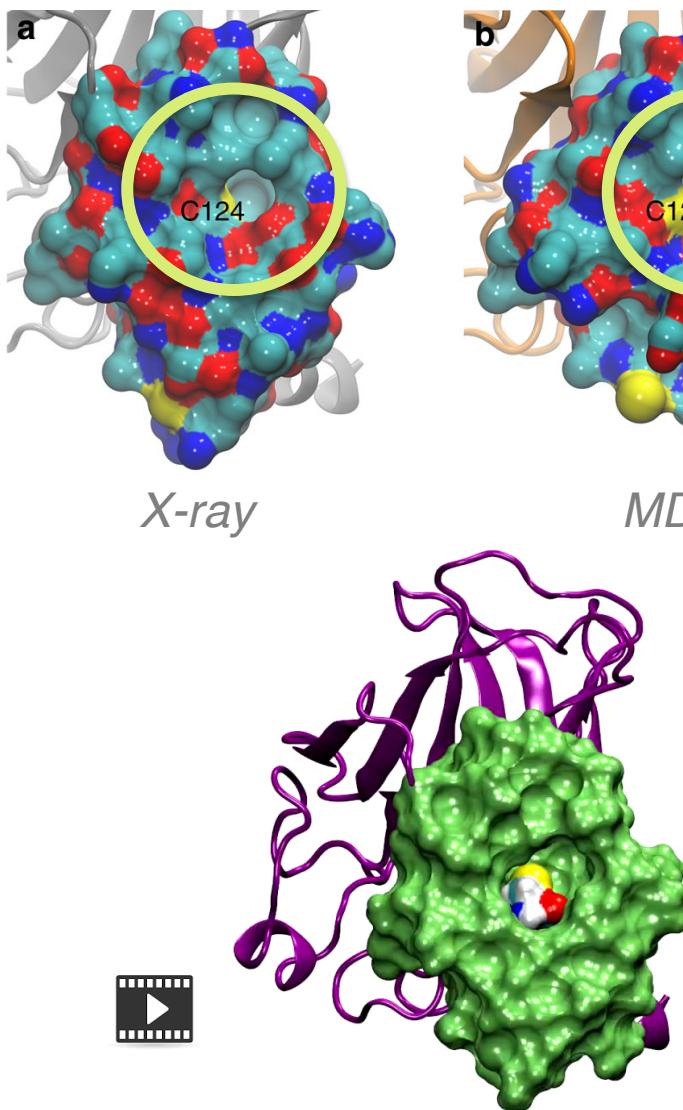
Hemoglobin is a remarkable molecular machine that uses motion and small structural changes to regulate its action. Oxygen binding at the four heme sites in hemoglobin does not happen simultaneously. Once the first heme binds oxygen, it introduces small changes in the structure of the corresponding protein chain. These changes nudge the neighboring chains into a different shape, making them bind oxygen more easily. Thus, it is difficult to add the first oxygen molecule, but binding the second, third and fourth oxygen molecules gets progressively easier and easier. This provides a great advantage in hemoglobin function. When blood is in the lungs, where oxygen is plentiful, oxygen easily binds to the first subunit and then quickly fills up the remaining ones. Then, as blood circulates through the body, the oxygen level drops while that of carbon dioxide increases. In this environment, hemoglobin releases its bound oxygen. As soon as the first oxygen molecule drops off, the protein starts changing its shape. This prompts the remaining three oxygens to be quickly released. In this way, hemoglobin picks up the largest possible load of oxygen in the lungs, and delivers all of it where and when needed.

In this animated figure, the heme group of one subunit, shown in the little circular window, is kept in one place so that you can see how the protein moves around it when oxygen binds. The oxygen molecule is shown in blue green. As it binds to the iron atom in the center of the heme, it pulls a histidine amino acid upwards on the bottom side of the heme. This shifts the position of an entire alpha helix, shown here in orange below the heme. This motion is propagated throughout the protein chain and on to the other chains, ultimately causing the large rocking motion of the two subunits shown in blue. The two structures shown are PDB entries 2hhb [\[link\]](#) and 1hho [\[link\]](#).

<http://www.rcsb.org/pdb/101/motm.do?momID=41>

Importance of Dynamics for Protein Function

Transient pockets for ligand binding



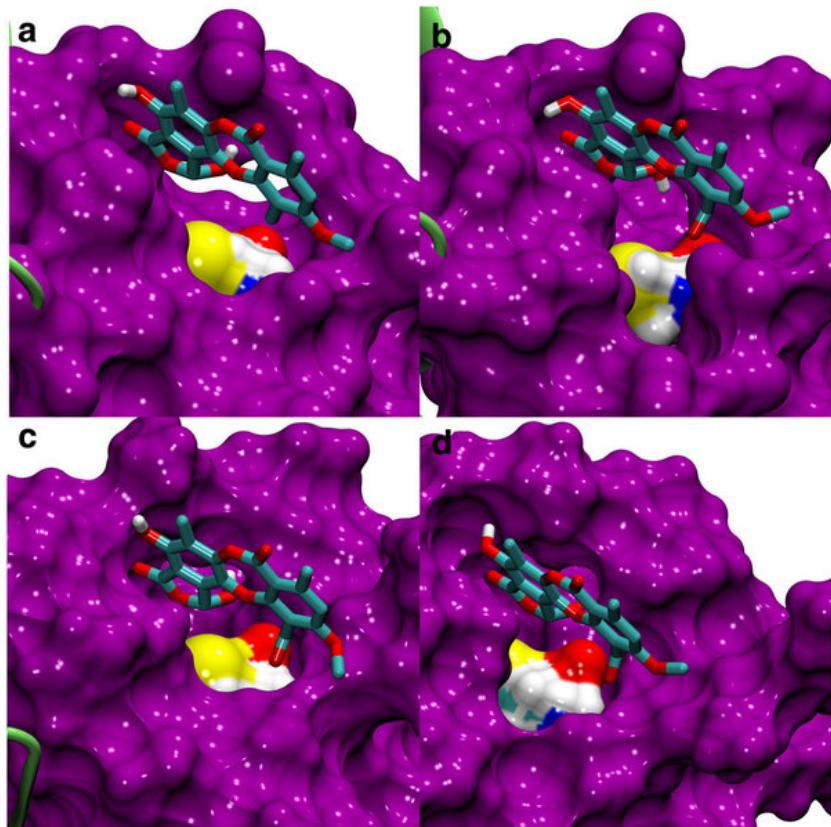
MD simulations are used to identify a dynamic binding pocket, not visible from experimental structures, on the surface of the tumour suppressor p53

The L1/S3 pocket (depicted in green surface representation) is generally occluded in the available crystal structures. In the MD simulations, the pocket transiently opens, exposing the Cys124 side chain (depicted inside the L1/S3 pocket).

Wassman, C. D. et al. Nat. Commun. 4, 140 (2013).

Importance of Dynamics for Protein Function

Transient pockets for ligand binding

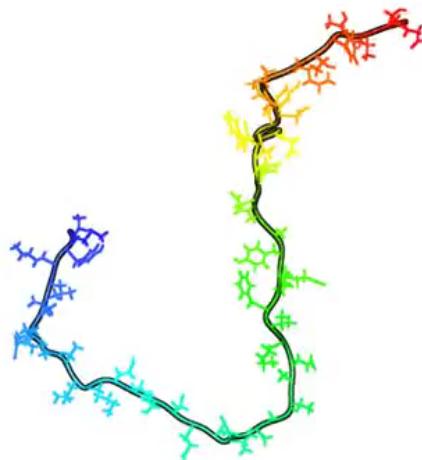


The transient binding pocket was used to identify a potential drug that can reactivate p53 cancer-related mutants.

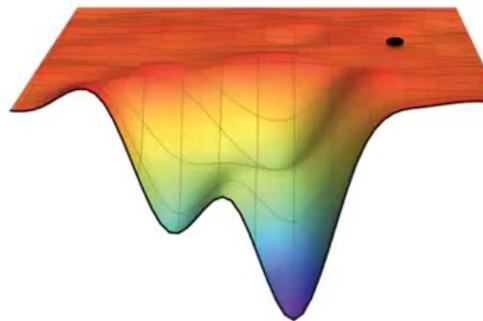
Molecular Mechanics and Force Fields

Energy landscapes of proteins

3D Structure



Energy landscape



<http://www.youtube.com/watch?v=YANAso8Jxrk>

*Molecular Mechanics Force Fields provide a
'recipe' (equations and parameters) to calculate the
potential energy of a protein from its atomic coordinates*

Molecular Mechanics and Force Fields

Levitt, M. et al *Computer Physics Communications* 91, 215(1995)

MM Force Fields

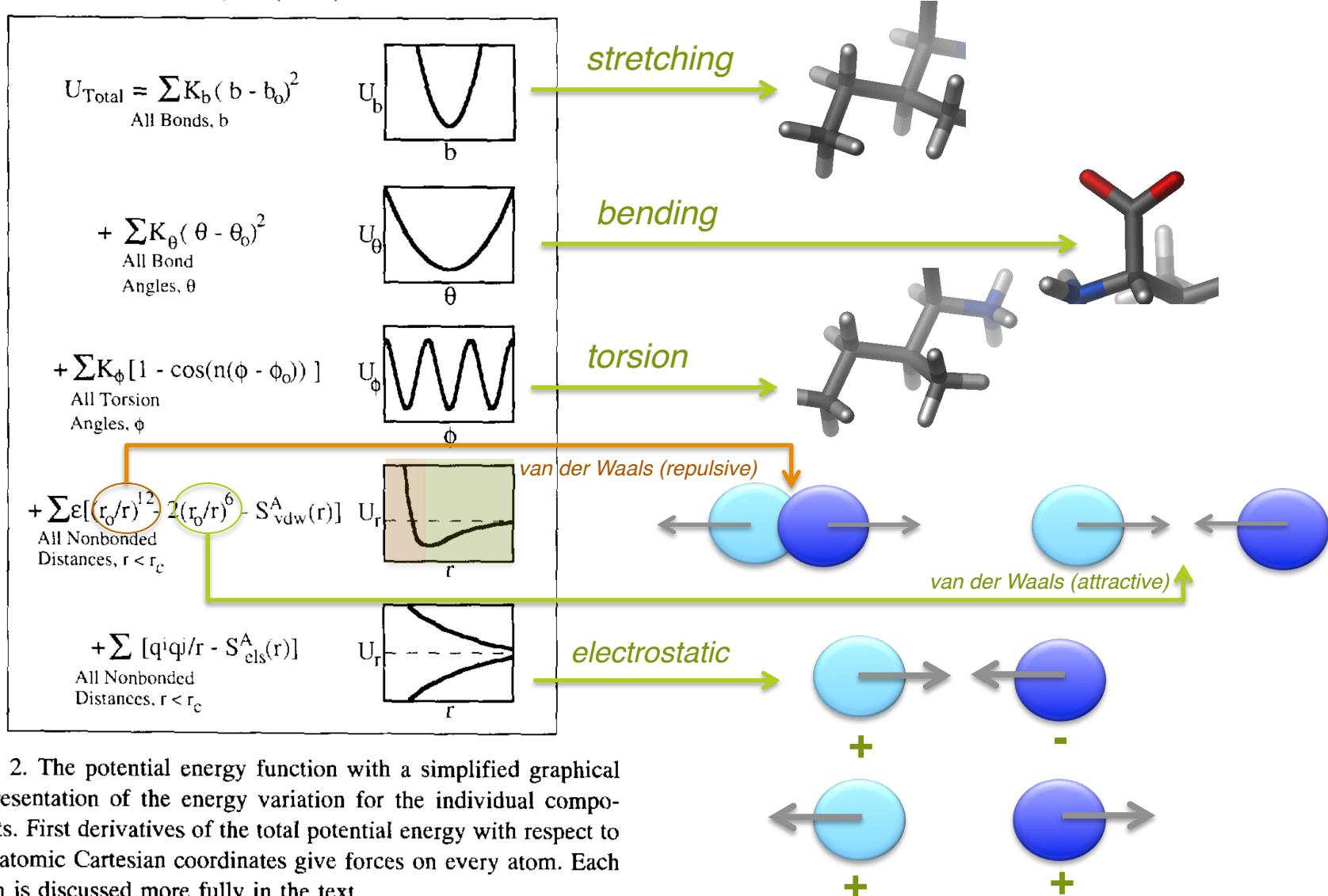


Fig. 2. The potential energy function with a simplified graphical representation of the energy variation for the individual components. First derivatives of the total potential energy with respect to the atomic Cartesian coordinates give forces on every atom. Each term is discussed more fully in the text.

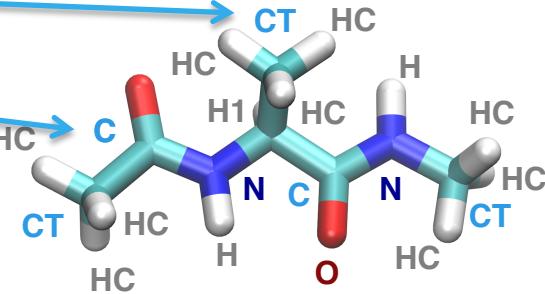
Molecular Mechanics and Force Fields

MM Force Fields

Assignment of Atom Types (based on the chemical environment)

Table 1. List of Atom Types^a

atom	type	description		
carbon	CT	any sp ³ carbon		
	C	any carbonyl, sp ² carbon		
	CA	any aromatic sp ² carbon and (C _{>} of Arg)		
	CM	any sp ² carbon, double bonded		
	CC	sp ² aromatic in 5-membered ring with one substituent + next to nitrogen (Cy in His)		
	CV	sp ² aromatic in 5-membered ring next to carbon and lone pair nitrogen (e.g. C _{>} in His (δ))		
	CW	sp ² aromatic in 5-membered ring next to carbon and NH (e.g. C _{>} in His (ϵ) and in Trp)		
	CR	sp ² aromatic in 5-membered ring next to two nitrogens (Cy and Ce in His)		
	CB	sp ² aromatic at junction of 5- and 6-membered rings (C _{>} in Trp) and both junction atoms in Ade and Gua		
	C*	sp ² aromatic in 5-membered ring next to two carbons (e.g. Cy in Trp)		
	CN	sp ² junction between 5- and 6-membered rings and bonded to CH and NH (Ce in Trp)		
	CK	sp ² carbon in 5-membered aromatic between N and N-R (C8 in purines)		
	CQ	sp ² carbon in 6-membered ring between lone pair nitrogens (e.g. C2 in purines)		
	N	sp ² nitrogen in amides		
	NA	sp ² nitrogen in aromatic rings with hydrogen attached (e.g. protonated His, Gua, Trp)		
	NB	sp ² nitrogen in 5-membered ring with lone pair (e.g. N7 in purines)		
	NC	sp ² nitrogen in 6-membered ring with lone pair (e.g. N3 in purines)		
	N*	sp ² nitrogen in 5-membered ring with carbon substituent (in purine nucleosides)		
	N2	sp ² nitrogen of aromatic amines and guanidinium ions		
	N3	sp ³ nitrogen		
nitrogen	OW	sp ³ oxygen in TIP3P water		
	OH	sp ³ oxygen in alcohols, tyrosine, and protonated carboxylic acids		
	OS	sp ³ oxygen in ethers		
	O	sp ² oxygen in amides		
	O2	sp ² oxygen in anionic acids		
	S	sulfur in methionine and cysteine		
	SH	sulfur in cysteine		
	P	phosphorus in phosphates		
	H	H attached to N		
	HW	H in TIP3P water		

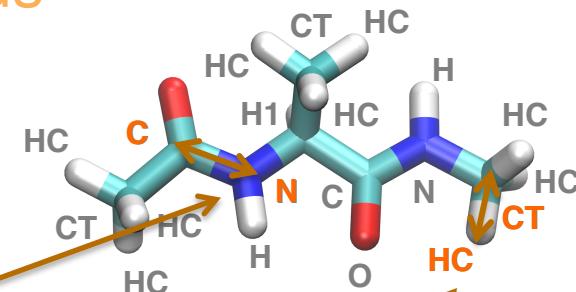


Molecular Mechanics and Force Fields

MM Force Fields

Assignment of parameters

$$U_{\text{stretch}}^{X-Y} = K_r^{X-Y} \times (r - r_{eq}^{X-Y})^2$$



Bond Parameters

bond	K_r^b	r_{eq}^c	bond	K_r^b	r_{eq}^c	bond	K_r^b	r_{eq}^c	bond	K_r^b	r_{eq}^c
C-CA	469.0	1.409	CA-HA	367.0	1.080	CM-HA	367.0	1.080	CT-S	227.0	1.810
C-CB	447.0	1.419	CA-N2	481.0	1.340	CM-N*	448.0	1.365	CT-SH	237.0	1.810
C-CM	410.0	1.444	CA-NA	427.0	1.381	CN-NA	428.0	1.380	CV-H4	367.0	1.080
C-CT	317.0	1.522	CA-NC	483.0	1.339	CQ-H5	367.0	1.080	CV-NB	410.0	1.394
C-N	490.0	1.335	CB-CB	520.0	1.370	CQ-NC	502.0	1.324	CW-H4	367.0	1.080
C-N*	424.0	1.383	CB-CN	447.0	1.419	CR-H5	367.0	1.080	CW-NA	427.0	1.381
C-NA	418.0	1.388	CB-N*	436.0	1.374	CR-NA	477.0	1.343	H-N	434.0	1.010
C-NC	457.0	1.358	CB-NB	414.0	1.391	CR-NB	488.0	1.335	H-N*	434.0	1.010
C-O	570.0	1.229	CB-NC	461.0	1.354	CT-CT	310.0	1.526	H-N2	434.0	1.010
C-O2	656.0	1.250	CC-CT	317.0	1.504	CT-F	367.0	1.380	H-N3	434.0	1.010
C-OH	450.0	1.364	CC-CV	512.0	1.375	CT-H1	340.0	1.090	H-NA	434.0	1.010
C*-CB	388.0	1.459	CC-CW	518.0	1.371	CT-H2	340.0	1.090	HO-OH	553.0	0.960
C*-CT	317.0	1.495	CC-NA	422.0	1.385	CT-H3	340.0	1.090	HO-OS	553.0	0.960
C*-CW	546.0	1.352	CC-NB	410.0	1.394	CT-HC	340.0	1.090	HS-SH	274.0	1.336
C*-HC	367.0	1.080	CK-H5	367.0	1.080	CT-HP	340.0	1.090	O2-P	525.0	1.480
CA-CA	469.0	1.400	CK-N*	440.0	1.371	CT-N	337.0	1.449	OH-P	230.0	1.610
CA-CB	469.0	1.404	CK-NB	529.0	1.304	CT-N*	337.0	1.475	OS-P	230.0	1.610
CA-CM	427.0	1.433	CM-CM	549.0	1.350	CT-N2	337.0	1.463	OW-HW	553.0	0.9572
CA-CN	469.0	1.400	CM-CT	317.0	1.510	CT-N3	367.0	1.471	S-S	166.0	2.038
CA-CT	317.0	1.510	CM-H4	367.0	1.080	CT-OH	320.0	1.410			
CA-H4	367.0	1.080	CM-H5	367.0	1.080	CT-OS	320.0	1.410			

Molecular Mechanics and Force Fields

Force Field Parametrisation

Force Fields parameters are derived from experimental data or calculations performed at a higher level of theory (Quantum Mechanics)

Equilibrium bond distances and angles: X-ray crystallography

Bond and angle force constants: vibrational spectra, normal mode calculations with Quantum Mechanics (QM)

Dihedral angle parameters: difficult to measure directly with experiments; fit to QM calculations for rotations around a bond with other motions fixed

Atom charges: fit to experimental liquid properties, ESP charge fitting to reproduce electrostatic potentials of high level QM, X-ray crystallographic electron density

van der Waals parameters: often most difficult to determine, fit to experimental liquid properties, intermolecular energy fitting

Molecular Mechanics and Force Fields

Force Fields for Biomolecules

FFs commonly used for biomolecules:

AMBER
CHARMM
GROMOS
OPLS

Improved over time and validated against experimental data, including:

- experimental structures
- secondary structure propensities
- NMR data (e.g. order parameters, chemical shifts, NOEs etc...)

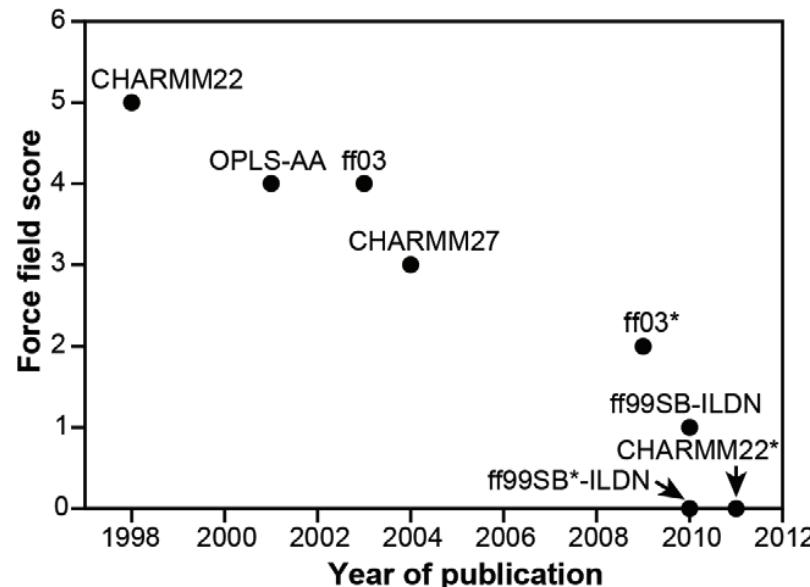


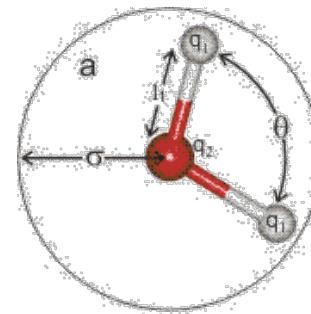
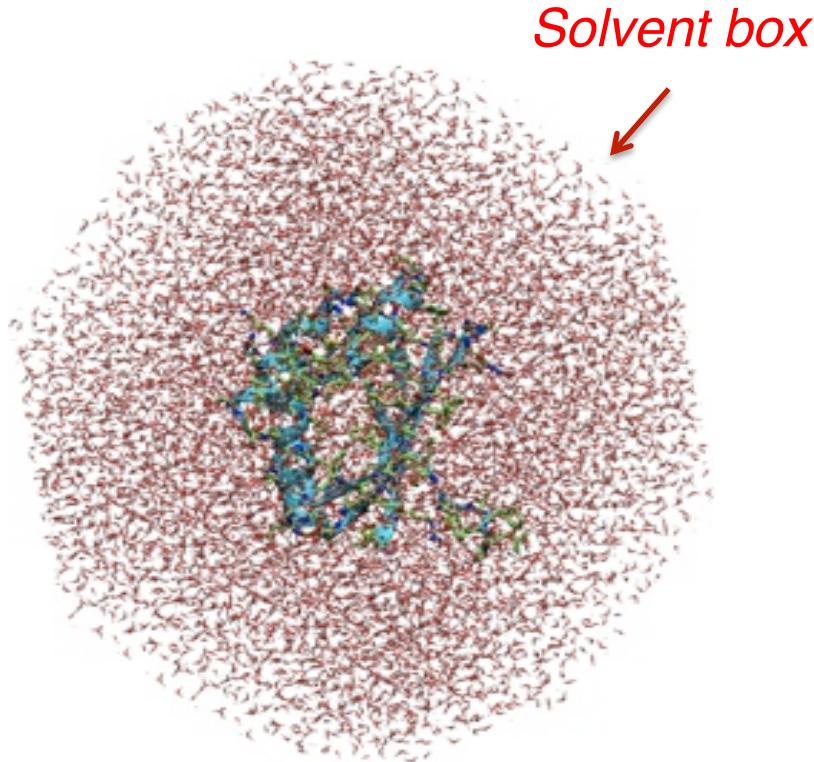
Figure 3. Improvement of force fields over time. For each force field, we assigned a score depending on the agreement with experiments in the tests presented here. Low scores indicate good agreement with experiments. These scores are plotted against the year in which the force field was published. For the force fields that involve multiple corrections (e.g., ff99SB*-ILDN), we use the year of the most recently published correction.

Lindorff-Larsen, K. et al. *PLoS ONE* 7, e32131 (2012).

Molecular Mechanics and Force Fields

Solvation Models: Explicit Solvation

Water molecules are explicitly included in the simulated system



*3-point water models
(TIP3P, SPC etc...)*

[http://www.science.oregonstate.edu/
~hetheriw/astro/rt/info/water/
water_models.html](http://www.science.oregonstate.edu/~hetheriw/astro/rt/info/water/water_models.html)

*Explicit solvation can accurately describe specific solvent-solute interactions.
However, it is very expensive, requiring $\sim N_{\text{wat}}^2$ calculations*

Molecular Mechanics and Force Fields

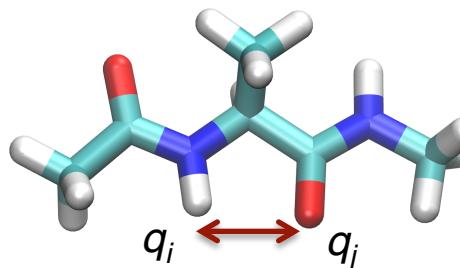
Solvation Models: Implicit Solvation

Water molecules are **not** explicitly included and the effect of water is only modelled.

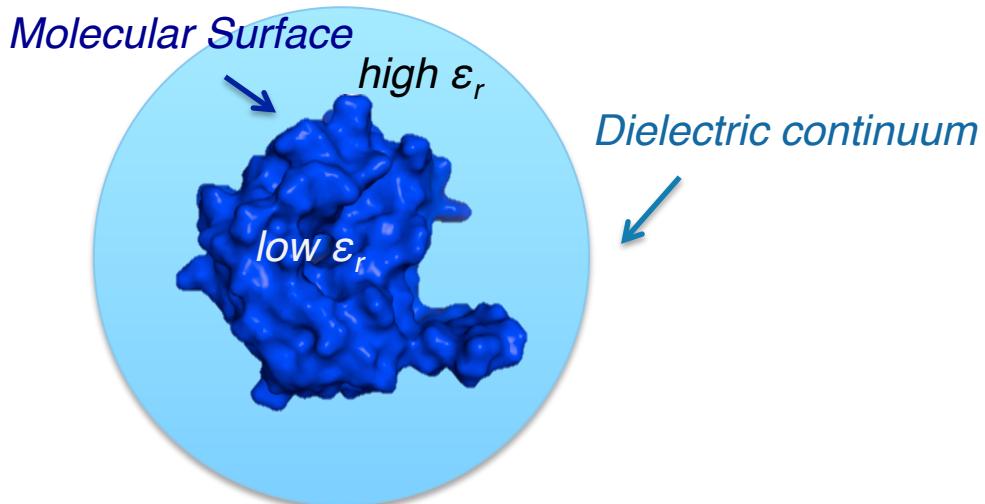
Distance-dependent dielectric constant

$$U_{coul} = \frac{q_i q_j}{\epsilon_r(r_{ij}) r_{ij}}$$

mimics screening effect of solvent



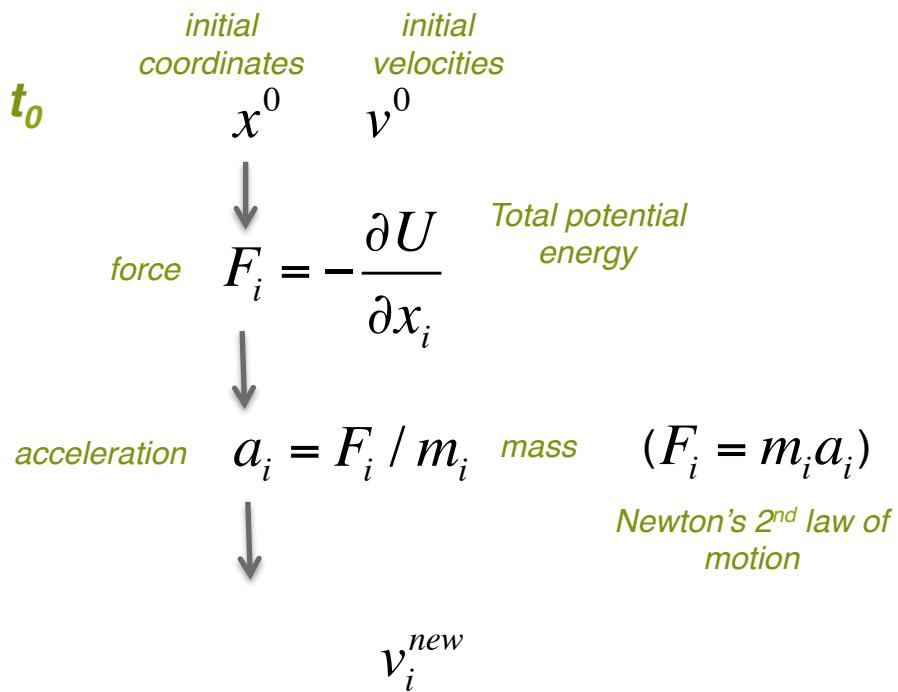
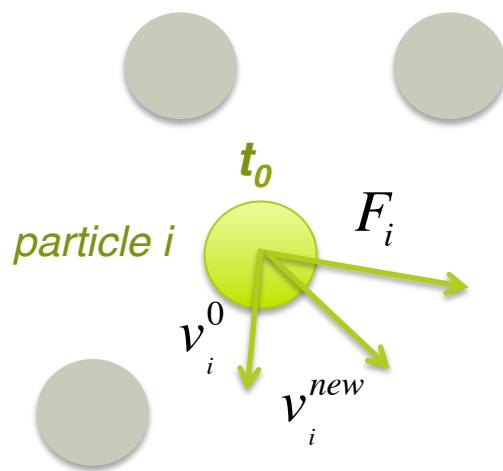
Methods based on the Poisson-Boltzmann equation



For a recent review of solvation models for biomolecular simulation:
Ren, P. et al. Quart. Rev. Biophys. 45, 427 (2012)

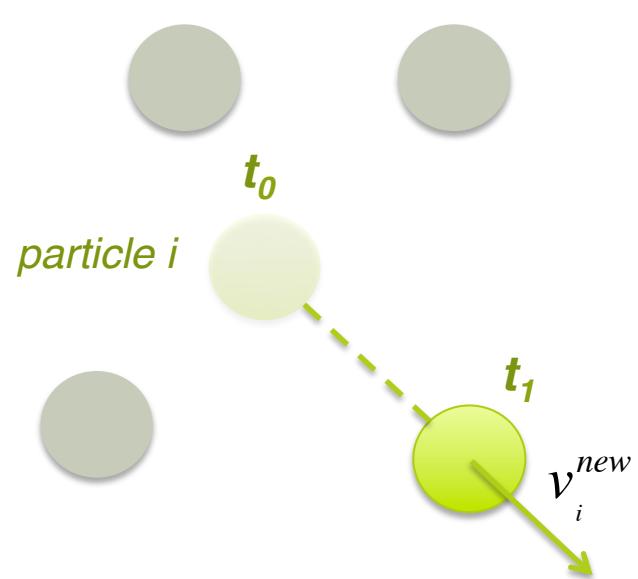
How to run a Molecular Dynamics simulation

Integration of equations of motion



How to run a Molecular Dynamics simulation

Integration of equations of motion

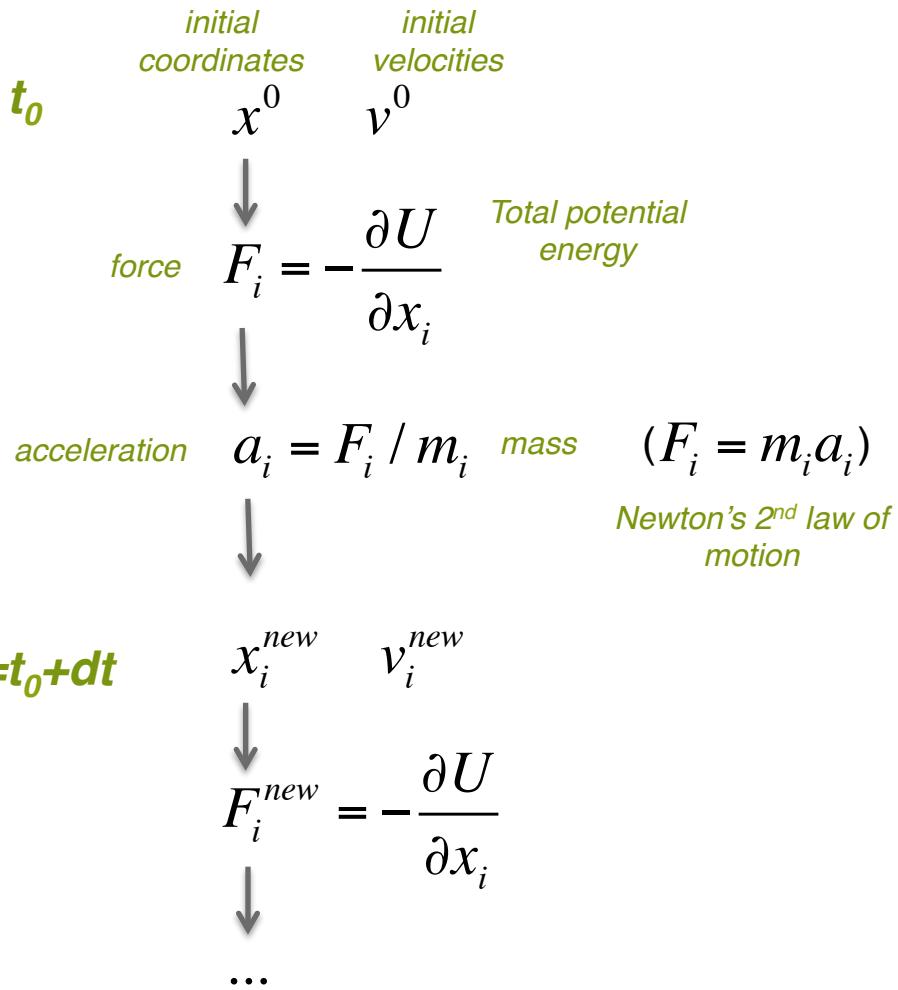
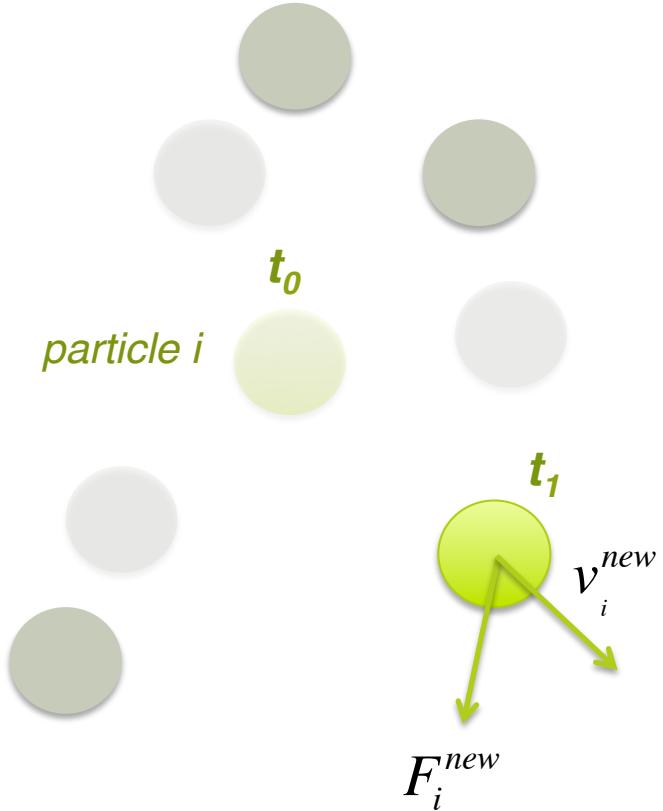


$$\begin{array}{c} \text{initial} \\ \text{coordinates} \\ \boldsymbol{x}^0 \\ \downarrow \\ \text{force} \\ \boldsymbol{F}_i = -\frac{\partial U}{\partial \boldsymbol{x}_i} \\ \downarrow \\ \text{acceleration} \\ \boldsymbol{a}_i = \boldsymbol{F}_i / m_i \\ \text{mass} \\ (\boldsymbol{F}_i = m_i \boldsymbol{a}_i) \\ \downarrow \\ \text{Newton's 2}^{\text{nd}} \text{ law of motion} \\ \boldsymbol{x}_i^{new} \quad \boldsymbol{v}_i^{new} \end{array}$$

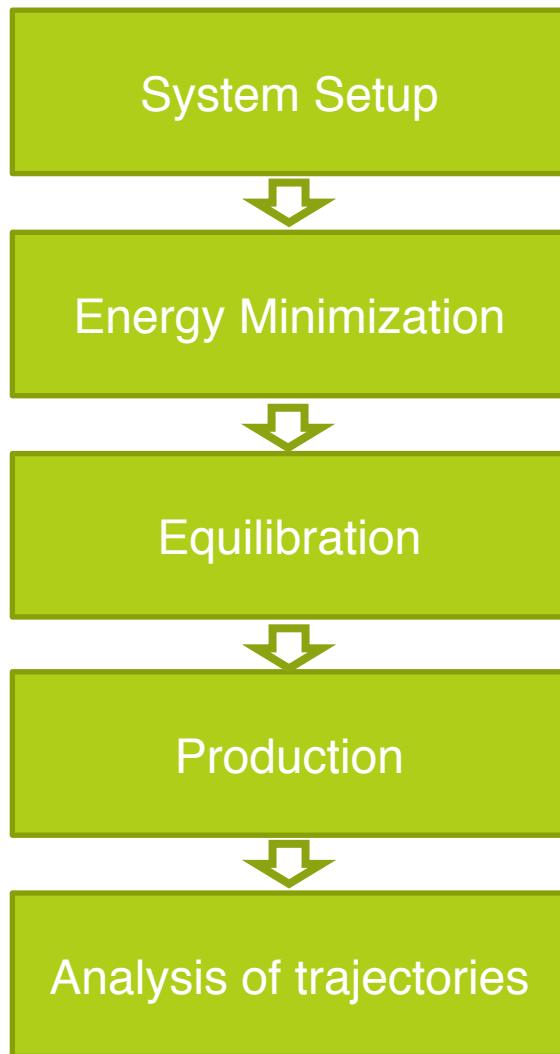
t_0 $t_1 = t_0 + dt$

How to run a Molecular Dynamics simulation

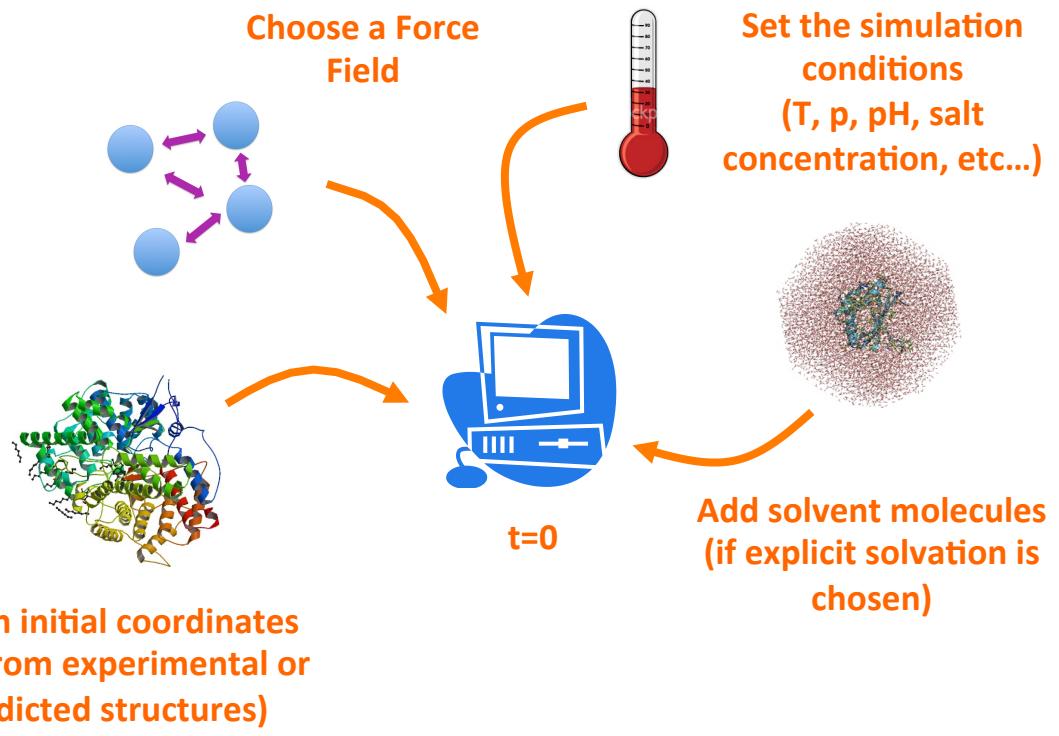
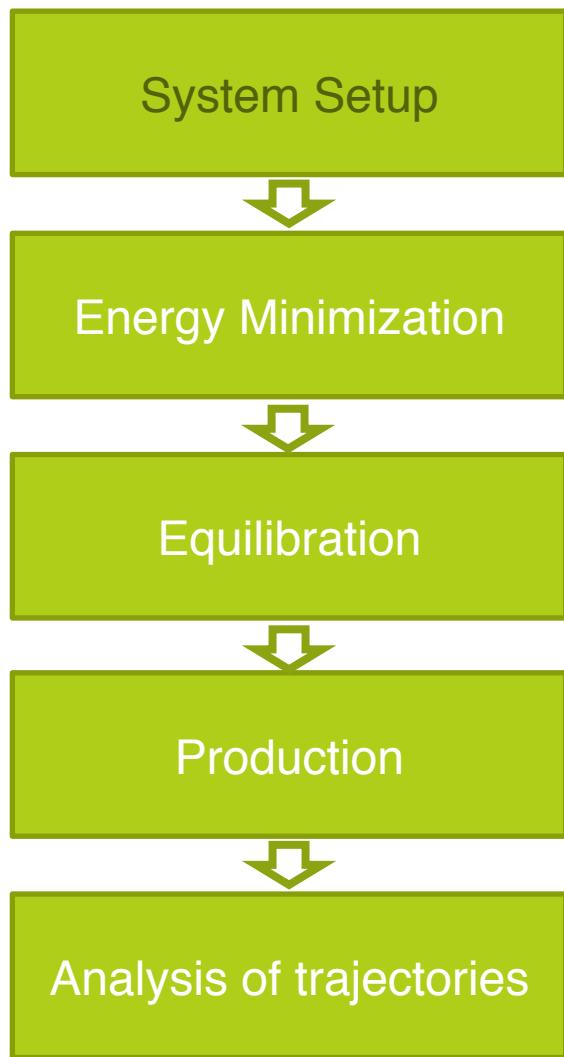
Integration of equations of motion



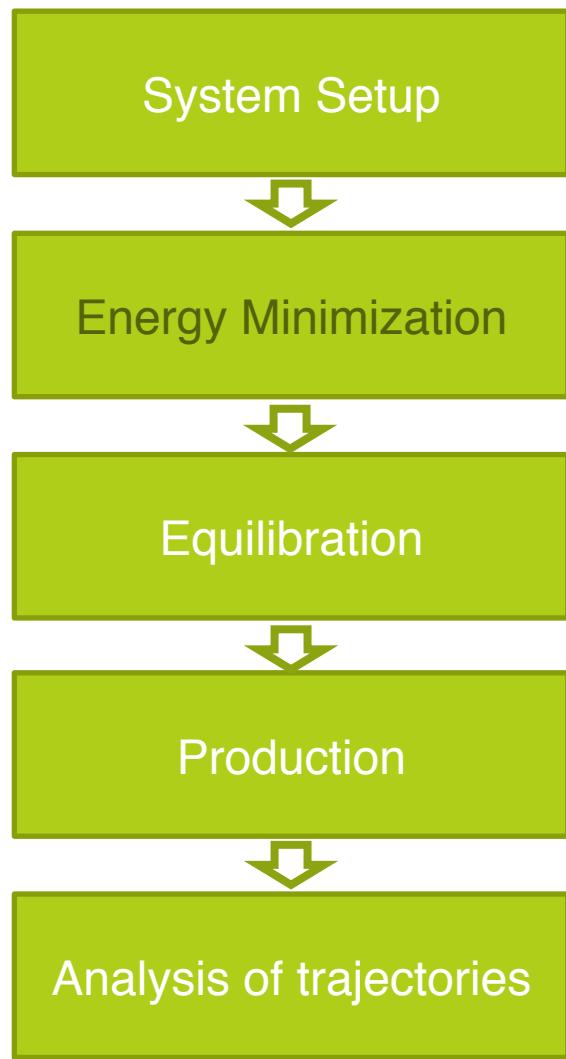
How to run a Molecular Dynamics simulation



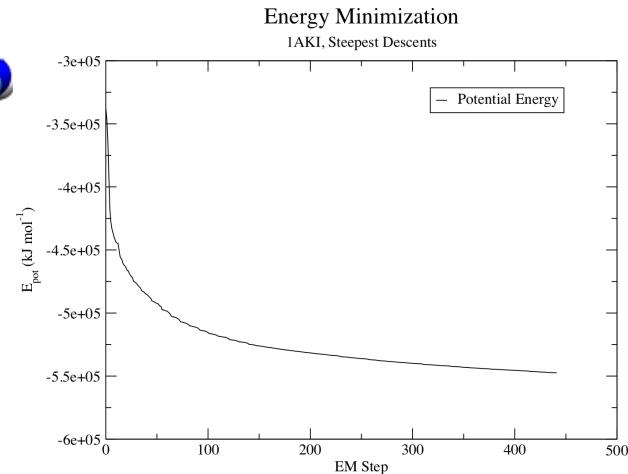
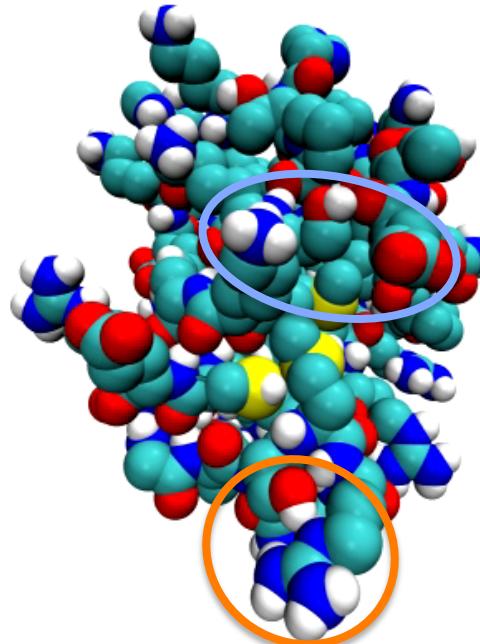
How to run a Molecular Dynamics simulation



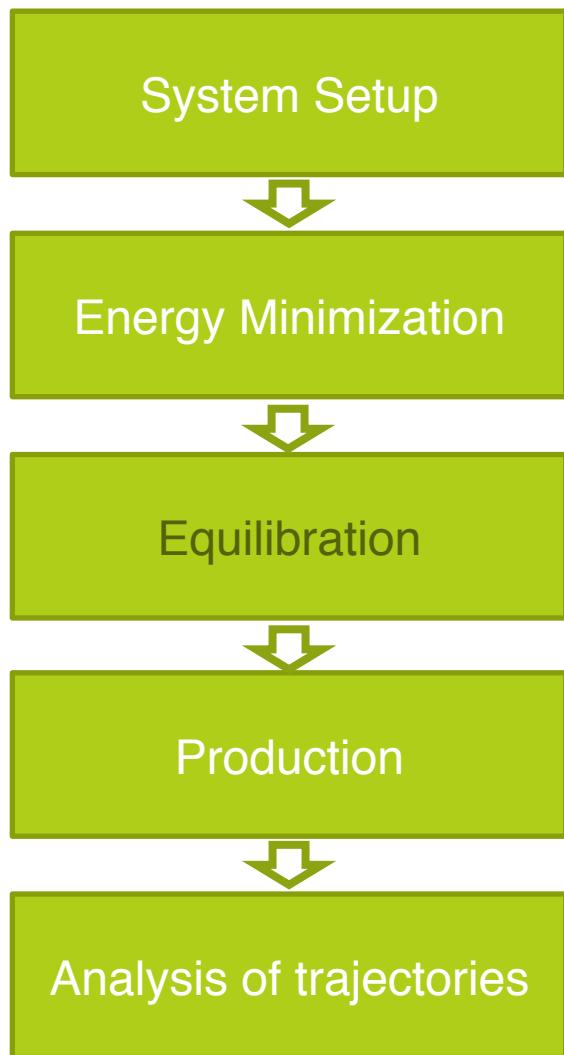
How to run a Molecular Dynamics simulation



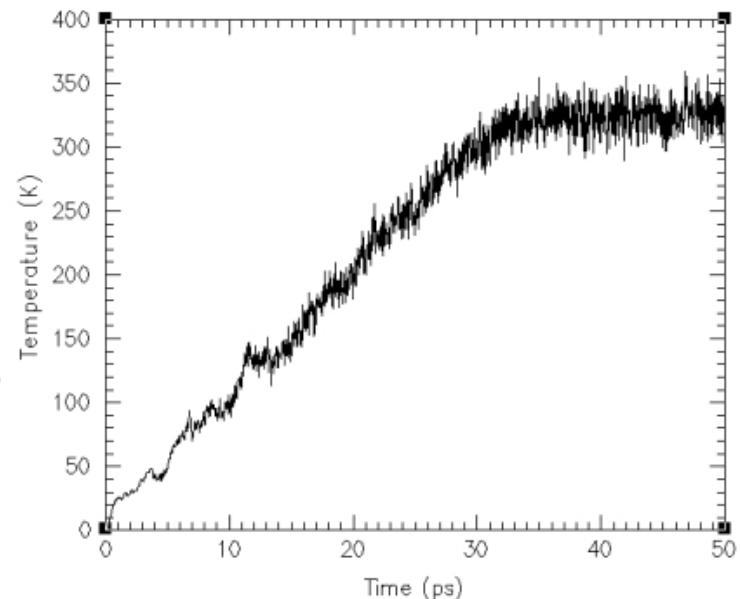
*Used mainly to relax the starting structure
(relieve unfavourable interactions like
steric clashes)*



How to run a Molecular Dynamics simulation



Heating-up of the system to the desired T. The density can be also adjusted to the desired value.

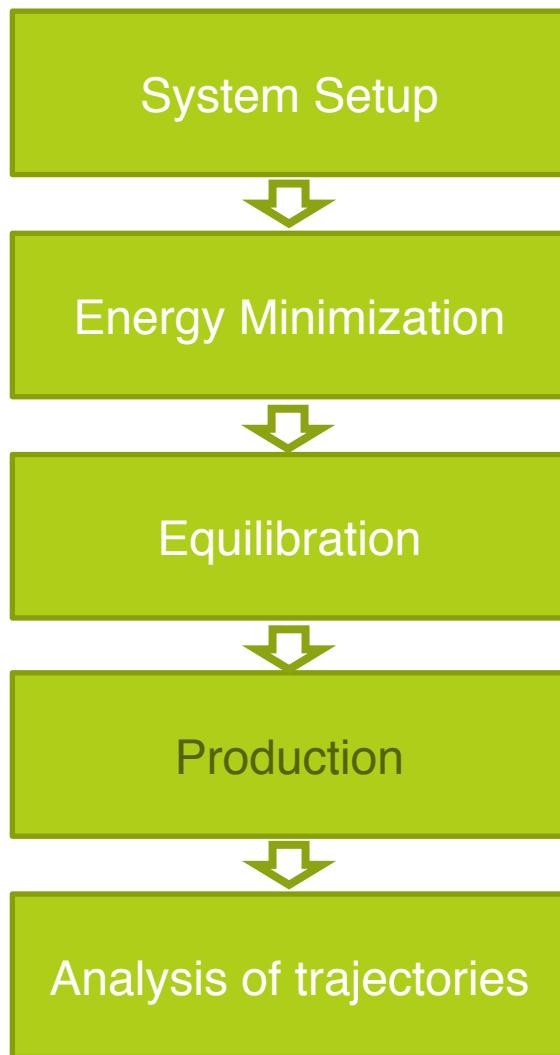


<http://ambermd.org/tutorials/basic/tutorial3/section6.htm>

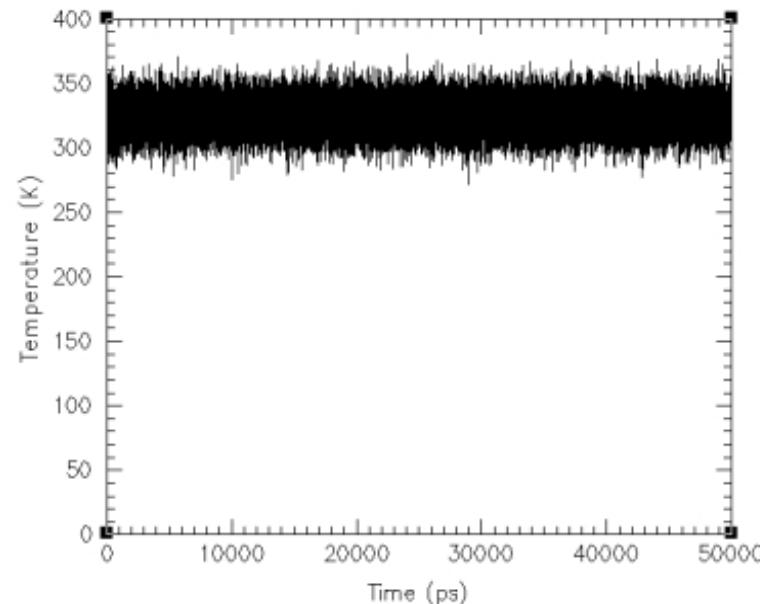
$$T = \frac{\sum_{i=1}^N m_i v_i^2}{3Nk_B}$$

mass
velocity
number of atoms
Boltzmann constant

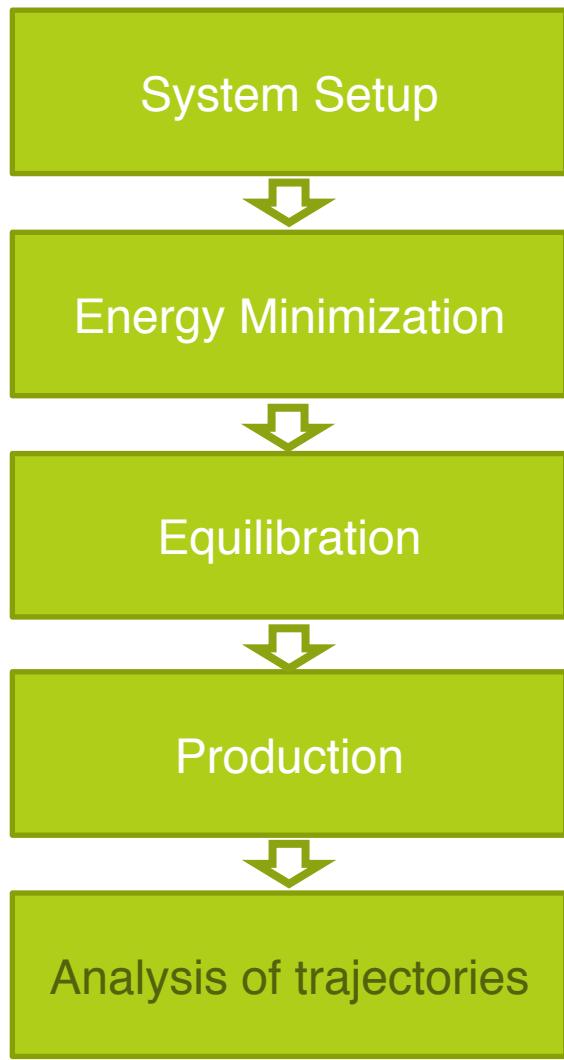
How to run a Molecular Dynamics simulation



After the average T and d values have been stabilised, production simulations can be run for data collection

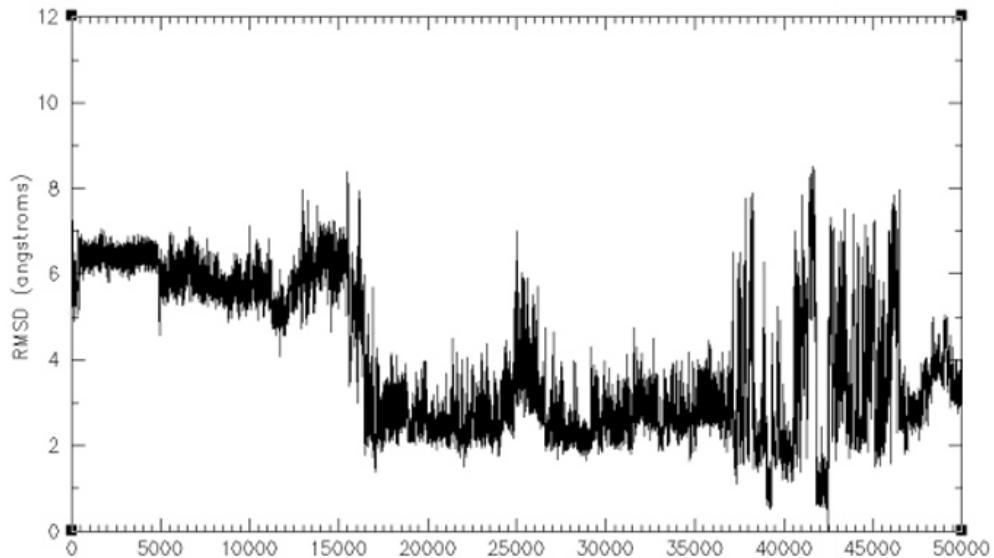


How to run a Molecular Dynamics simulation



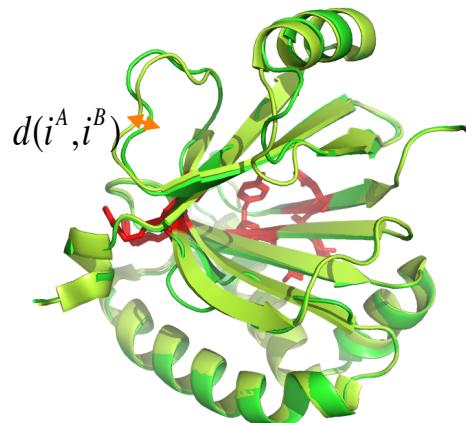
Time evolution of structural properties

*Root Mean Square Deviation (RMSD)
from a reference structure*

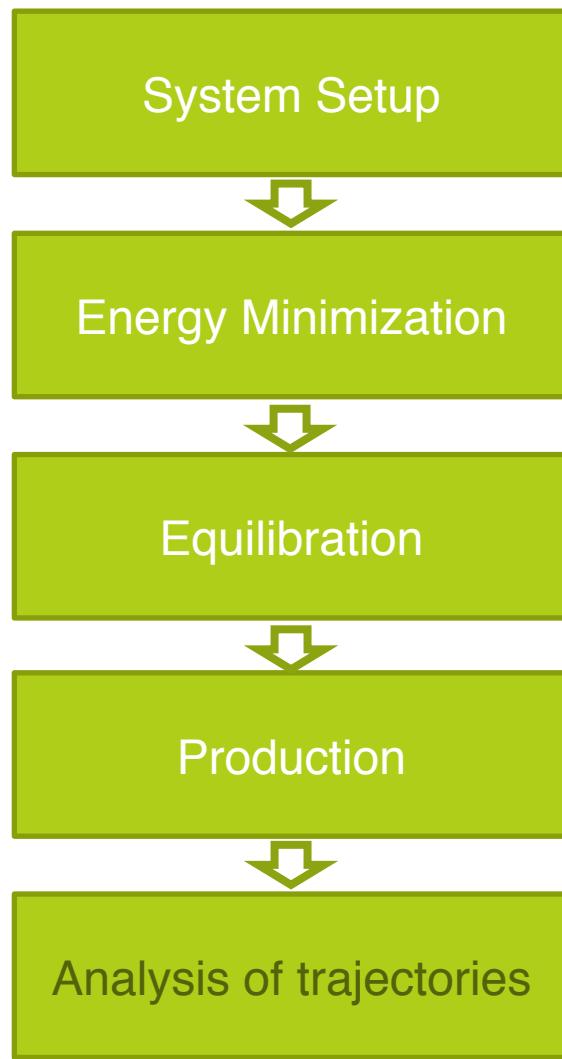


distance between equivalent
atoms in the two structures

$$RMSD(A, B) = \sqrt{\frac{\sum_{i=1}^{N_{at}} d^2(i^A, i^B)}{N_{at}}}$$



How to run a Molecular Dynamics simulation

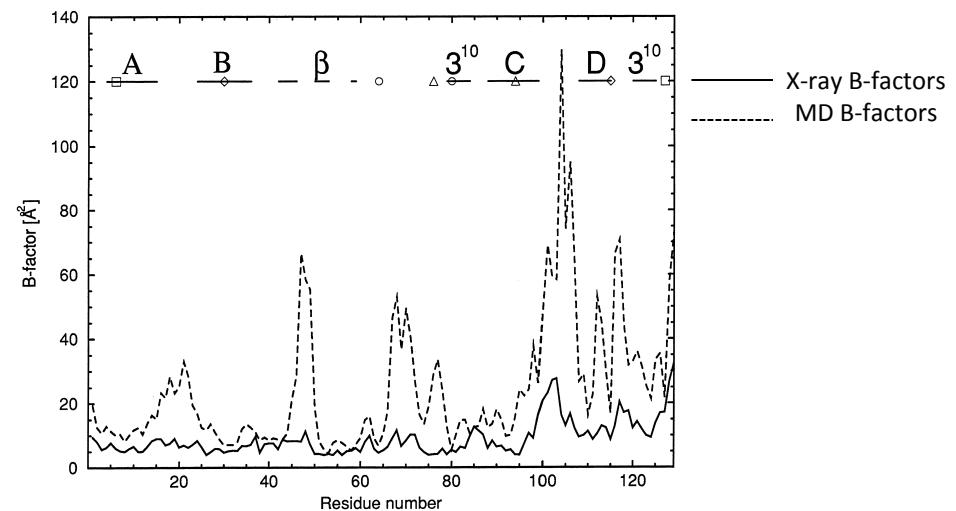
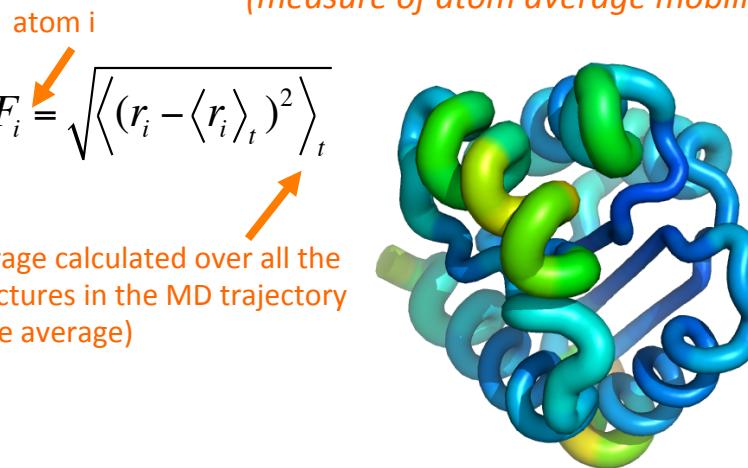


Time averages

Root Mean Square Fluctuations (RMSF) of atoms around their average positions (measure of atom average mobility)

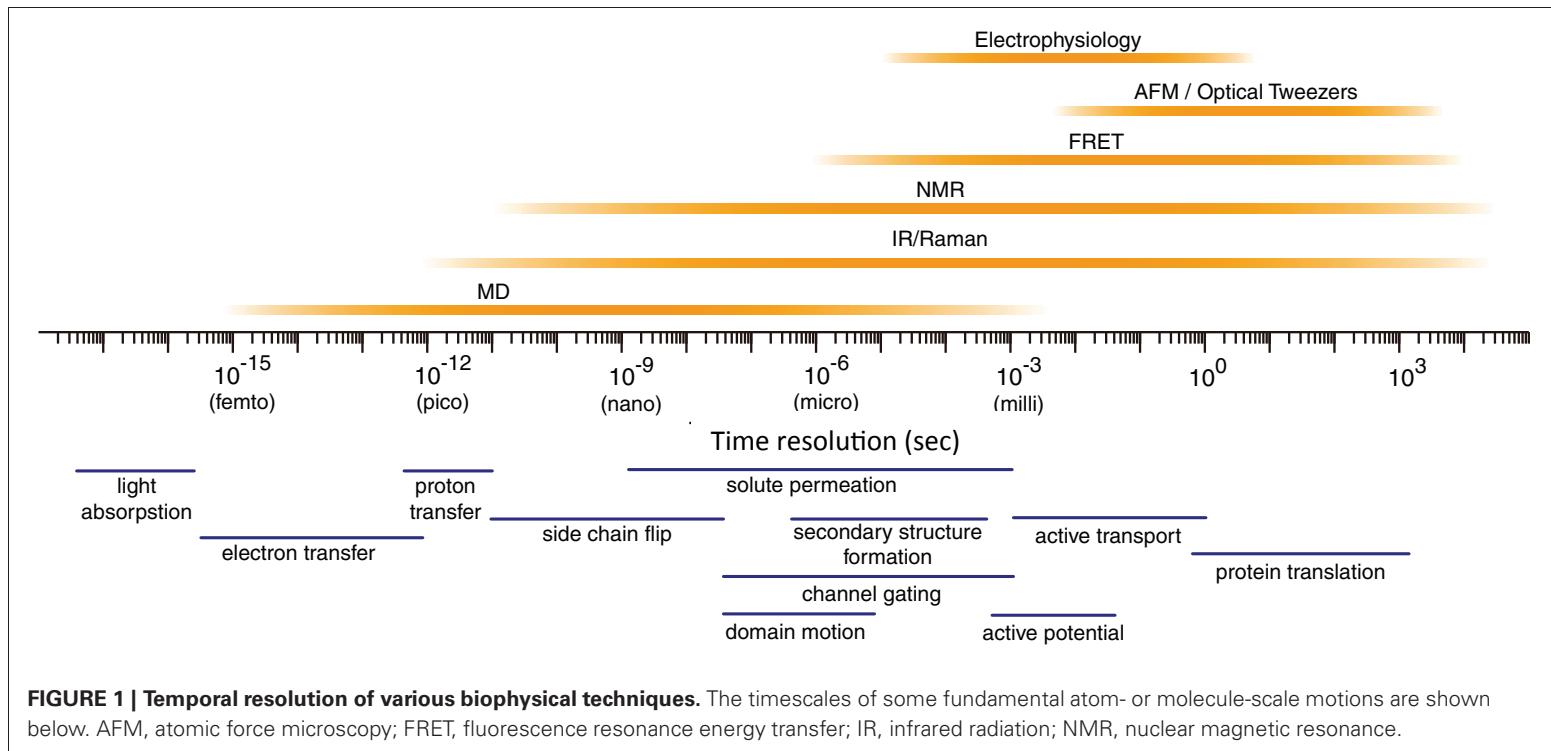
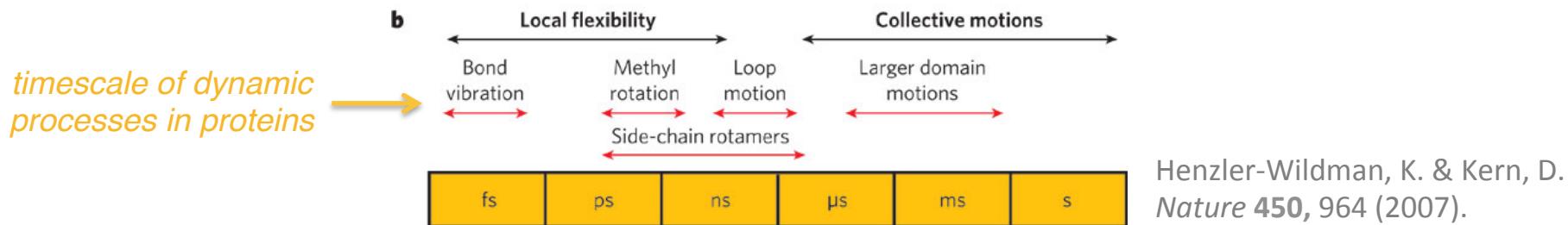
$$RMSF_i = \sqrt{\langle (r_i - \langle r_i \rangle_t)^2 \rangle_t}$$

average calculated over all the structures in the MD trajectory (time average)



What can be studied with MD

Accessible timescales



Ode, H., Nakashima, M., Kitamura, S., Sugiura, W. & Sato, H. *Front Microbiol* 3, 258 (2012).

Moore's law turns 50

Ever more from Moore

A microchip pioneer's prediction has a bit more life left in it

Apr 18th 2015 | From the print edition



Timekeeper



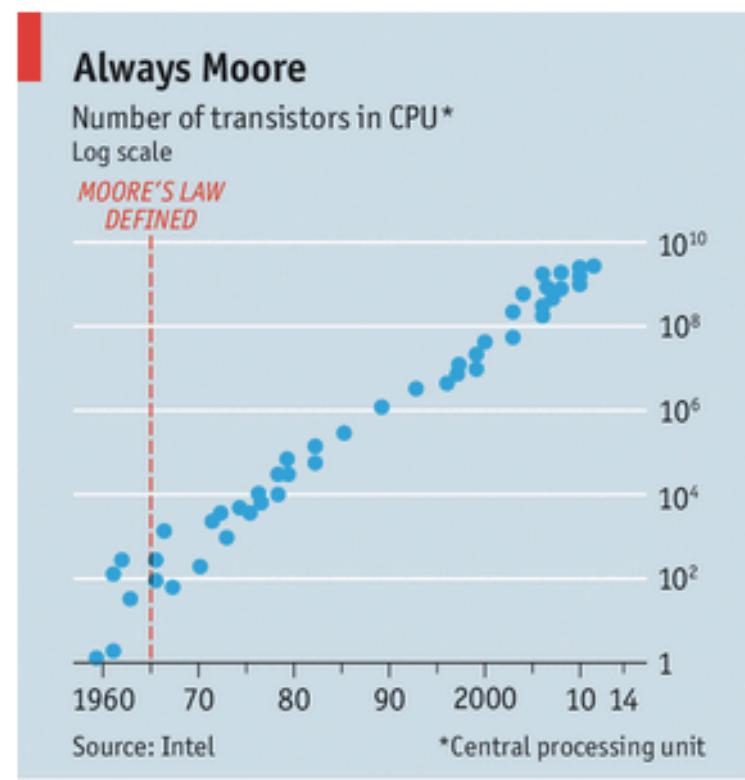
Like



Tweet



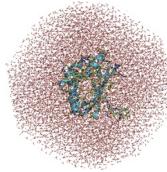
NEWS of the death of Moore's law has always been greatly exaggerated. People started to pronounce it deceased not long after Gordon Moore, co-founder of Intel, a chipmaker, published on April 19th 1965 a paper arguing that the number of transistors that can be etched on a given surface area of silicon would double every year. In a later paper he corrected his forecast to every two years, which has come to be stated as his "law". Regularly proving sceptics wrong, however, the exponential growth kept going (see chart), driving the digital revolution.



What can be studied with MD

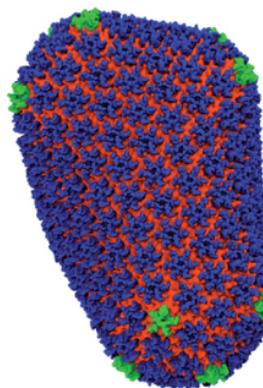
Size scale

single protein in water



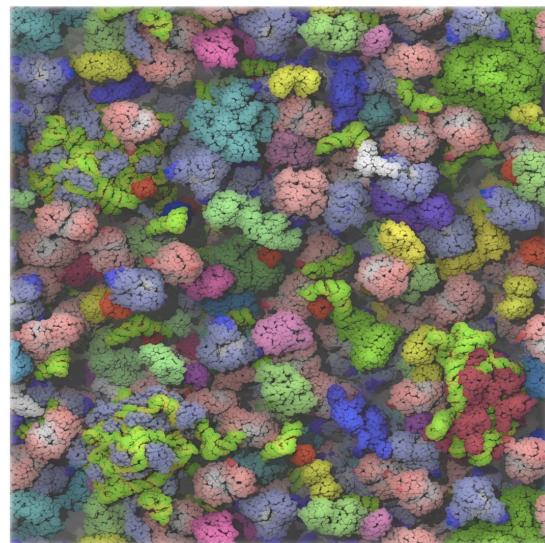
10-100k atoms

HIV-1 virus capsid



64 millions atoms

E. coli cytoplasm



1008 molecules

Zhao, G. et al. *Nature* **497**, 643 (2013).

McGuffee, S. R. & Elcock, A. H. *PLoS Comput Biol* **6**, e1000694 (2010).

system size

What can be studied with MD Equipment



**your
desktop/laptop**

~0.01TFlop/s



**small cluster
(lab equipment)**

~0.5-1TFlop/s



**supercomputer
(somewhere in the
world)**

$\sim 10^3\text{-}10^4$ TFlop/s

**GPU
(graphic cards)**



**$\sim 2\text{-}3$ TFlop/s
per GPU**

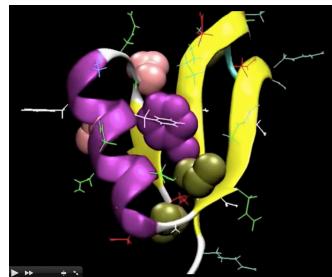
What can be studied with MD

Folding mechanisms



final structure

Simulation of a millisecond
folder: NTL9

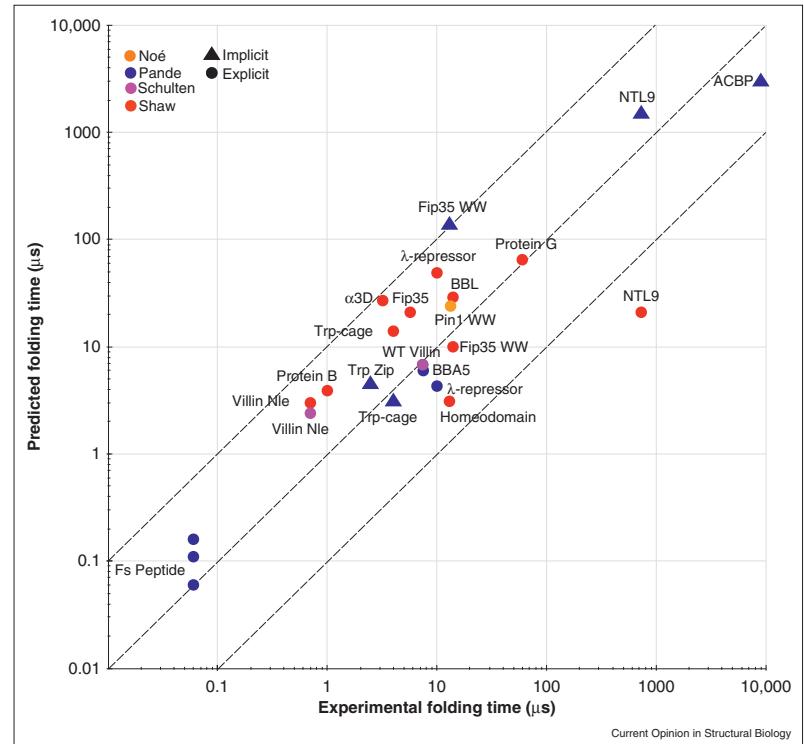
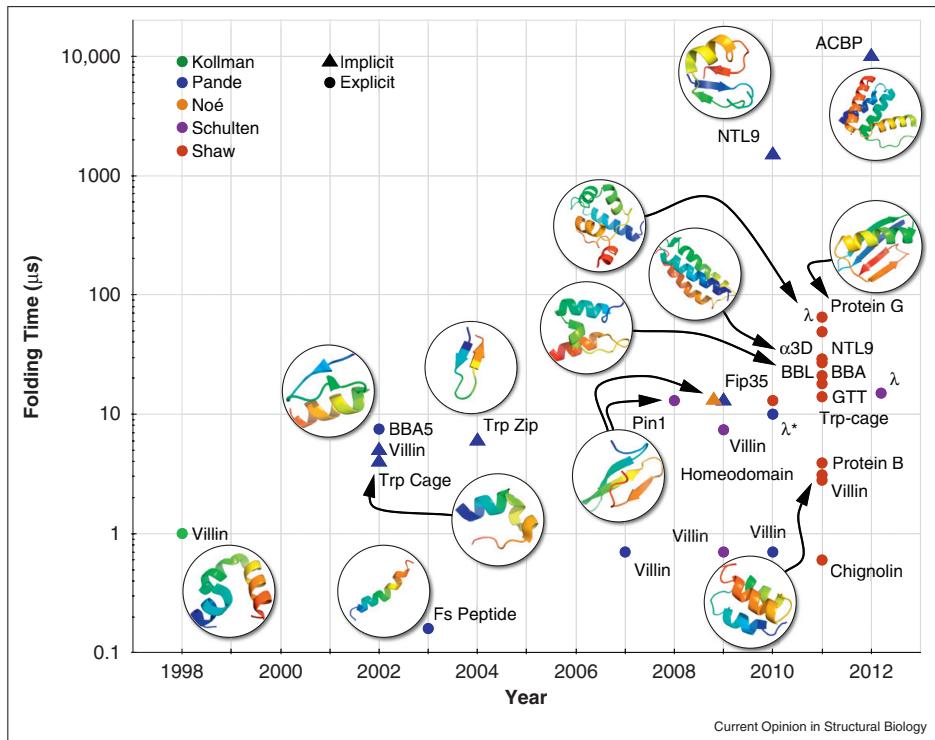


<http://www.youtube.com/watch?v=gFcp2Xpd29I>

Voelz, V. A. et al. *J. Am. Chem. Soc.* **132**, 1526–1528 (2010).

What can be studied with MD

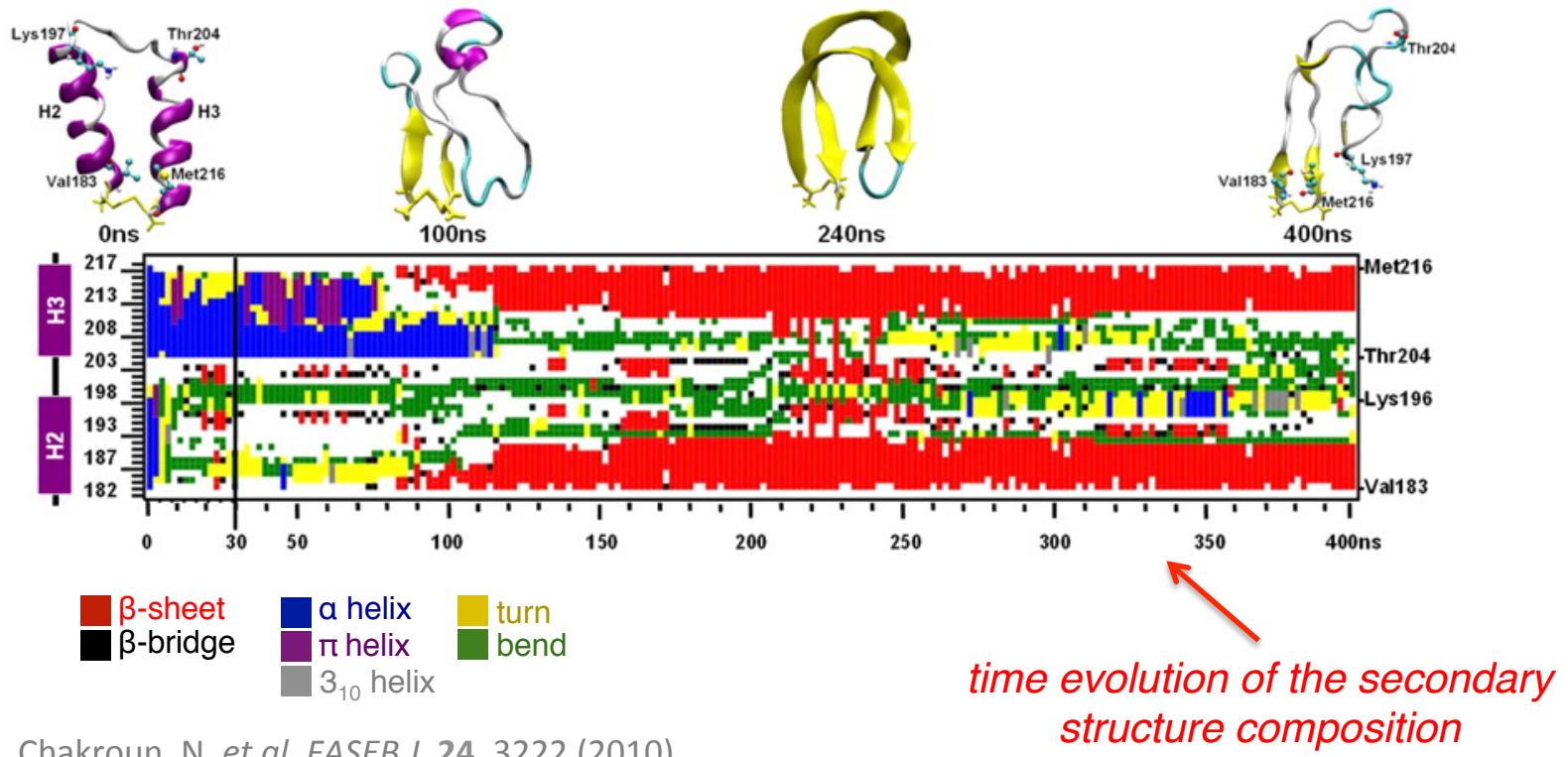
Folding rates



Lane, T. J., Shukla, D., Beauchamp, K. A. & Pande, *Curr. Op. Struct. Biol.* **23**, 58 (2012).

What can be studied with MD

Misfolding

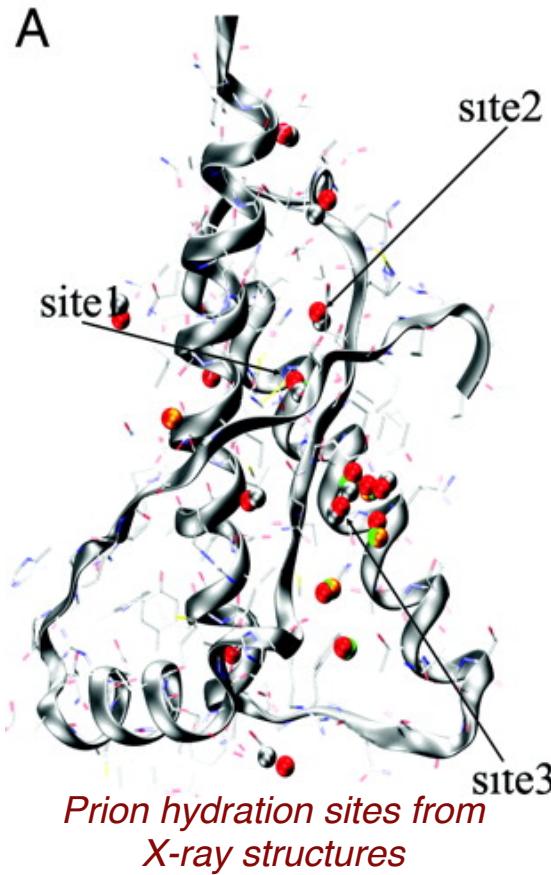
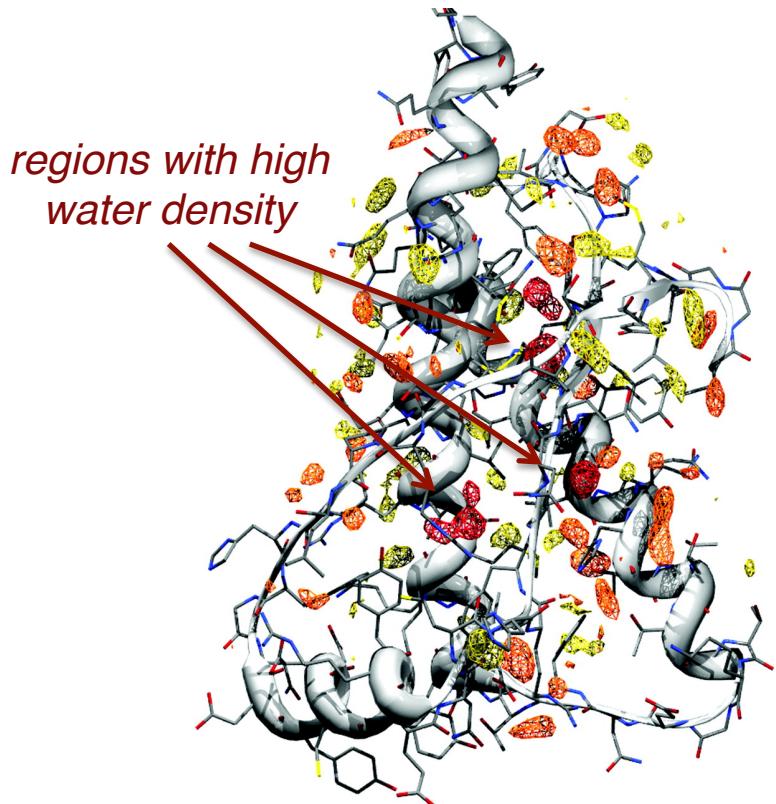


Chakroun, N. et al. FASEB J. 24, 3222 (2010).

The paper describes the transition of a fragment of the prion protein (H2H3) from an helical conformation to a β -strand rich conformation

What can be studied with MD

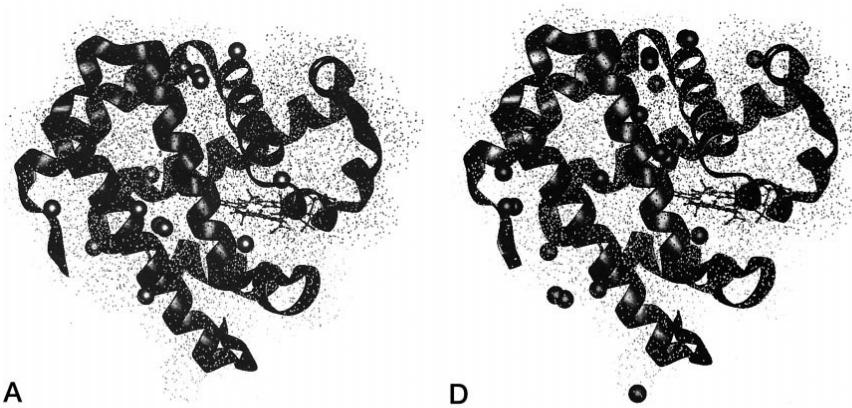
Solvent distribution around proteins



What can be studied with MD

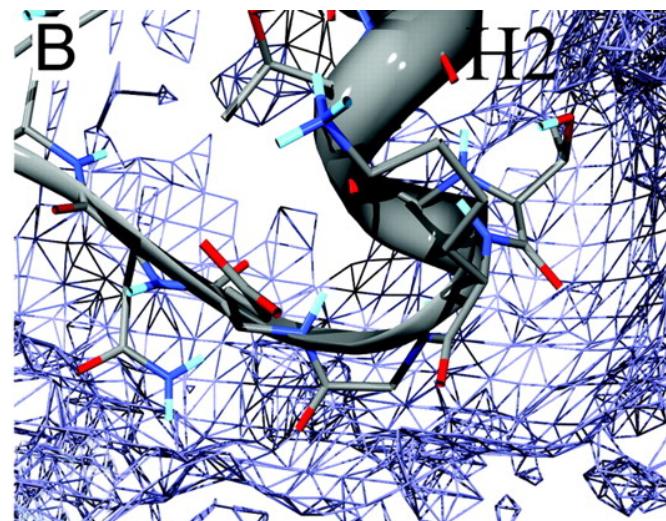
Solvent distribution around proteins

In addition to the location of hydration sites (regions with high water density), MD simulations can provide information on the dynamical properties of water



Hydration sites of myoglobin with high (right) and low (left) residence time

Makarov, V. A. *Biophys J.* **79**, 2966 (2000).



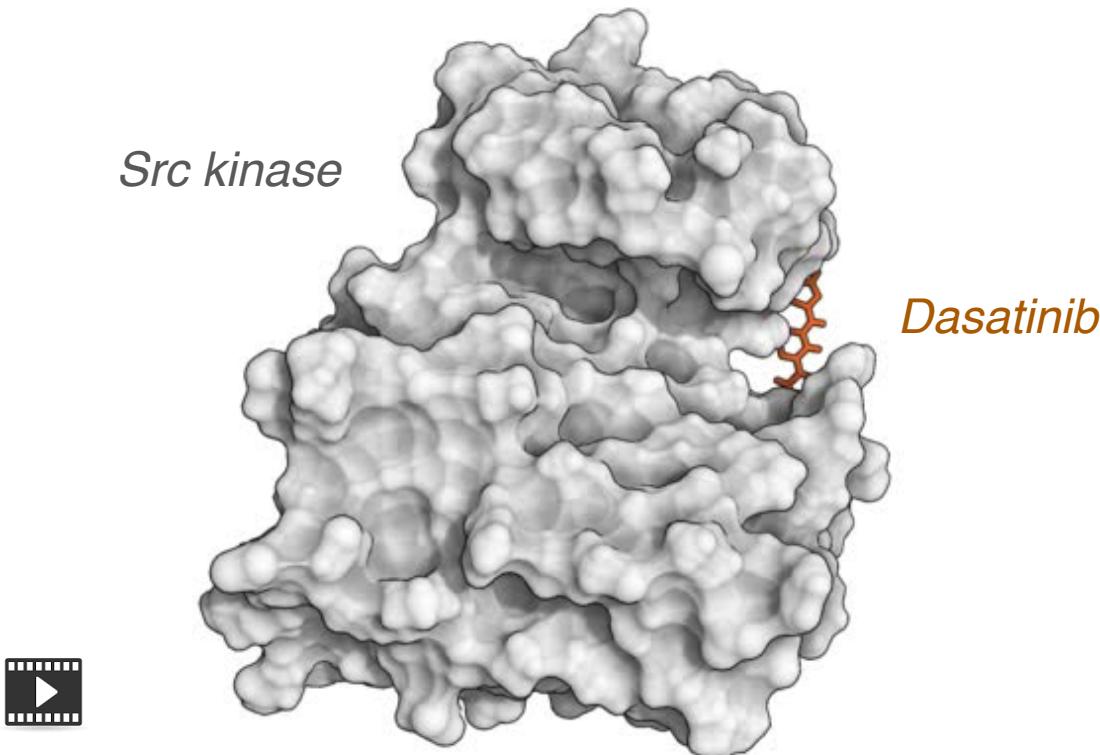
High-entropy water regions (blue) around the H2-H3 loop of prion

De Simone, A. et al. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7535 (2005).

$$S_{i,j,k} = -R \sum_{lmn} P_{l,m,n}^{i,j,k} \ln P_{l,m,n}^{i,j,k}$$

What can be studied with MD

Prediction of drug binding sites



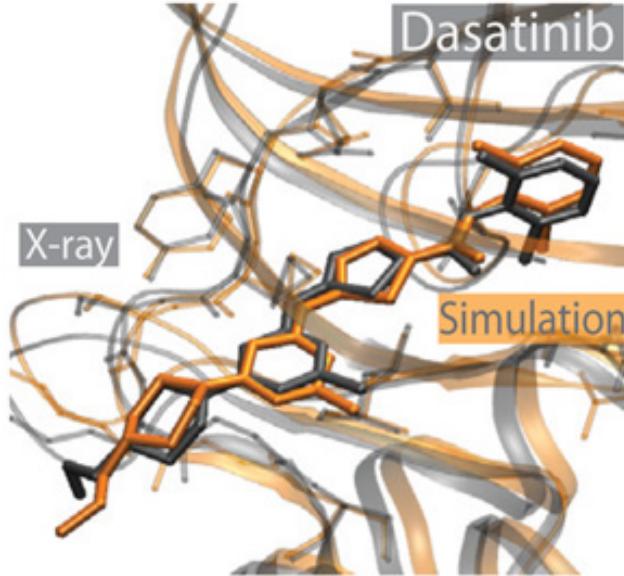
How Does a Drug Molecule Find Its Target Binding Site?

Shan, Y. et al *J. Am. Chem. Soc.* **133**, 9181 (2011)

The paper reports a long MD simulation (4 μ s) where a randomly placed drug finds its target binding site.

What can be studied with MD

Prediction of drug binding sites



How Does a Drug Molecule Find Its Target Binding Site?

Shan, Y. et al J. Am. Chem. Soc. **133**, 9181 (2011)

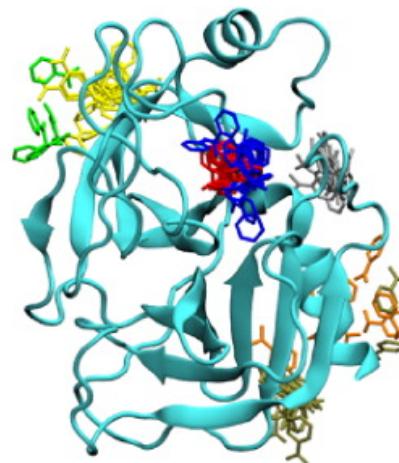
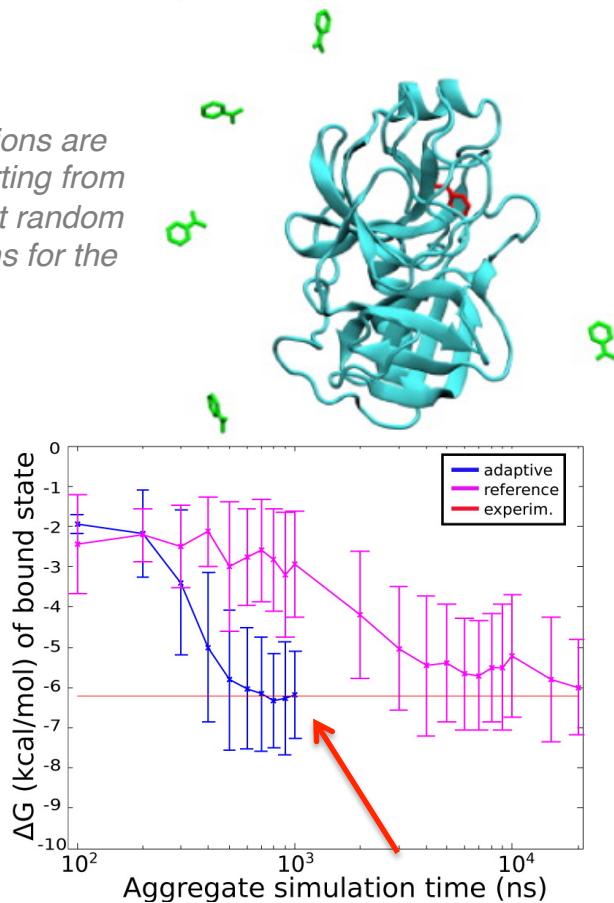
The simulated binding pose is in very good agreement with crystallographic data and was obtained without any prior knowledge of the binding site.

What can be studied with MD

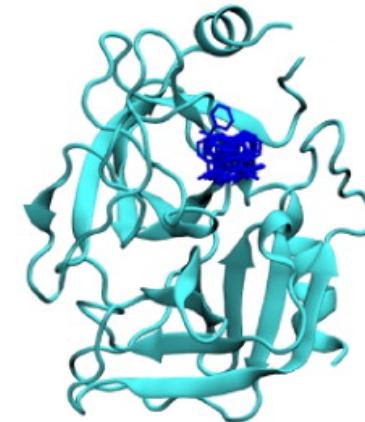
Prediction of binding affinities

Trypsin + benzamidine

parallel simulations are run starting from different random positions for the ligand



The different ligand poses sampled during the simulations are collected



The poses associated with the lowest ΔG_{bind} are identified

Doerr, S. & De Fabritiis, G. *J. Chem. Theory Comput.* **10**, 2064 (2014).

The calculations described in the paper make an efficient use of high-throughput parallel computing infrastructures (e.g. GPUGRID, http://www_gpugrid.net/). The final binding free energy ΔG_{bind} results from the analysis of many (100) short (10-ns) MD simulations run in parallel, for a total aggregate simulation time of 1 μ s.

We seek most of the time to
maintain equilibrium conditions!

Verlet integrator

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla} U(\vec{R})$$

Use positions and accelerations at time t and the positions from time $t-\delta t$ to calculate new positions at time $t+\delta t$.

$$\mathbf{r}(t + \delta t) \approx \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2$$

+

$$\mathbf{r}(t - \delta t) \approx \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2$$

“Verlet algorithm”



$$-\vec{\nabla} U(\vec{R})/m_i$$

$$\boxed{\mathbf{r}(t + \delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \mathbf{a}(t)\delta t^2}$$

Molecular chaos

Dynamics of “well-behaved” classical many-body system is chaotic.
Consequence: Trajectories that differ very slightly in their initial conditions diverge exponentially (“Lyapunov instability”)



Why should anyone believe Molecular Dynamics simulations ???

Answers:

1. Good MD algorithms (e.g. Verlet) can also be considered as good (*NVE!*) Monte Carlo algorithm – they therefore yield reliable STATIC properties (“Hybrid Monte Carlo”)
2. What is the point of simulating dynamics, if we cannot trust the resulting time-evolution???
3. All is well (probably), because of... The Shadow Trajectory-hypothesis

Shadow theorem (hypothesis)

- For any realistic many-body system, the shadow theorem is merely a hypothesis.
- It basically states that good algorithms generate numerical trajectories that are “close to” a REAL trajectory of the many-body system.
- Question: Does the Verlet algorithm indeed generate “shadow” trajectories?
- In practice, it follows an Hamiltonian, depending on the timestep, $\tilde{\mathcal{H}}(\mathbf{x}, \Delta t)$ which is close to the real Hamiltonian $\mathcal{H}(\mathbf{x})$, in the sense that for $\Delta t \rightarrow 0$ $\tilde{\mathcal{H}}(\mathbf{x}, \Delta t)$ converges to $\mathcal{H}(\mathbf{x})$
- Take a different look at the problem.
 - Do not discretize NEWTON's equation of motion...
 - ...but discretize the ACTION

Lagrangian Classical mechanics

- Newton:

$$F(x, t) = m\ddot{x}$$

- Lagrange (variational formulation of classical mechanics):

- Consider a system that is at a point r_0 at time 0 and at point r_t at time t , then the system follows a trajectory $r(t)$ such that:

$$S = \int_{t_b}^{t_e} dt [\mathcal{K} - \mathcal{U}]$$

is an extremum.

$$\mathcal{L}(\dot{r}, r) = K(\dot{r}) - U(r) = \frac{m\dot{r}^2}{2} - U(r)$$

$$\frac{\partial \mathcal{L}}{\partial \dot{r}} = \frac{\partial K}{\partial \dot{r}} = p$$

$$p = \frac{\partial \mathcal{L}(\dot{r}, r)}{\partial \dot{r}}$$

$$\frac{\partial \mathcal{L}}{\partial r} = -\frac{\partial U}{\partial r} = F$$

$$\dot{p} = \frac{\partial \mathcal{L}(\dot{r}, r)}{\partial r}$$

Lagrangian

For example, if we use Cartesian coordinates:

$$\mathcal{L}(r(t)) = \sum_{i=1}^N \frac{1}{2} m_i \dot{r}_i^2 - U(r_1, r_2, \dots, r_N)$$

Consider the “true” path $R(t)$, with $R(0) = r_0$ and $R(t) = r_t$.
Now, consider a path close to the true path:

$$r(t') = R(t') + \delta r(t')$$

Then the action S is an extremum if

$$\frac{\partial S}{\partial r(t')} = 0 \text{ for all } t$$

what does this mean?

Discretized action

$$S_{cont} = \int_{t_0}^{t_1} dt \mathcal{L}(t)$$
$$S_{disc} = \Delta t \sum_{i=0}^{i_{max}} \mathcal{L}(t_i) \quad \mathcal{L}(t_i) = K(t_i) - U(t_i)$$

For a one dimensional system this becomes

$$\mathcal{L}(t_i) \Delta t = \frac{1}{2} m \Delta t \frac{(x_{i+1} - x_i)^2}{\Delta t^2} - U(x_i) \Delta t$$

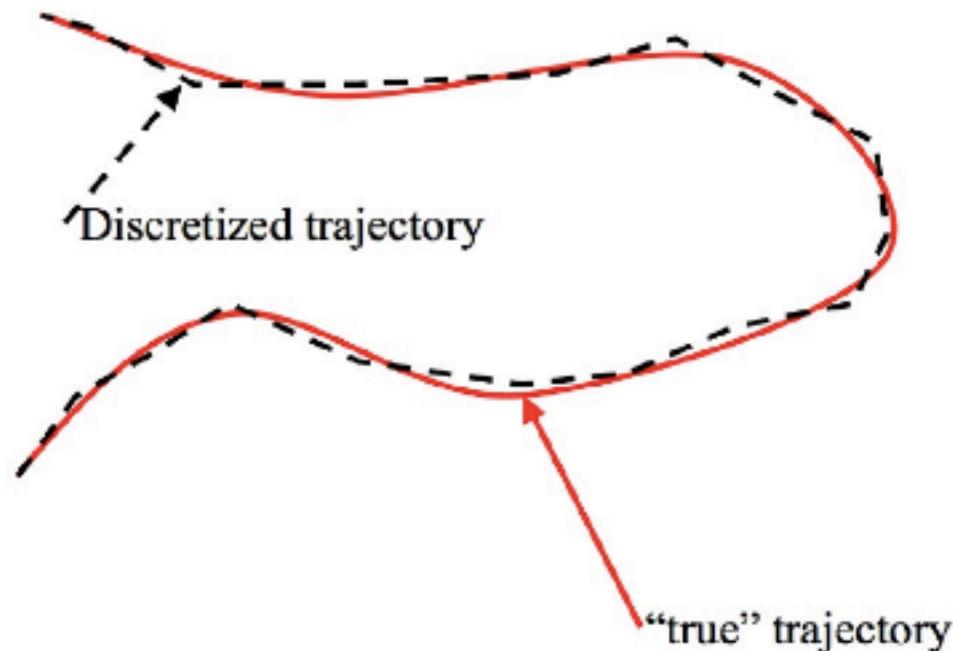
$$S_{disc} = \sum_{i=1}^{i_{max}} \left[\frac{m(x_{i+1} - x_i)^2}{2 \Delta t} - U(x_i) \Delta t \right]$$

Minimize the action

Now do the standard thing: Find the extremum for small variations in the path, i.e. for small variations in all x_i .

$$\frac{\partial S_{disc}}{\partial x_i} = 0 \text{ for all } i$$

This will generate a discretized trajectory that starts at time t_0 at X_0 , and ends at time t at X_t .



$$0 = 2x_i - x_{i+1} - x_{i-1} - \frac{\Delta t^2}{m} \frac{\partial U(x_i)}{\partial x_i}$$

$$x_{i+1} = 2x_i - x_{i-1} + \frac{\Delta t^2}{m} F(x_i)$$

- which is the Verlet algorithm!
- The Verlet algorithm generates a trajectory that satisfies the boundary conditions of a REAL trajectory –both at the beginning and at the endpoint.
- Hence, if we are interested in statistical information about the dynamics
(e.g. time-correlation functions, transport coefficients, power spectra...) ...then a “good” MD algorithm (e.g. Verlet) is fine.

How do we ensure we are in equilibrium conditions?

Ergodic system

It was introduced by Boltzmann regarding his hypothesis:
for large systems of interacting particles in equilibrium, the time
average along a single trajectory equals the space average

The hypothesis as it was stated was false, and the investigation
for the conditions under which these two quantities are equal
lead to the birth of ergodic theory.

A modern description of what ergodic theory is would be:
the study of the long term average behavior of systems evolving
in time.

How do we ensure we are in equilibrium conditions?

Liouville' theorem

The volume in phase space is conserved when the system evolves in time.

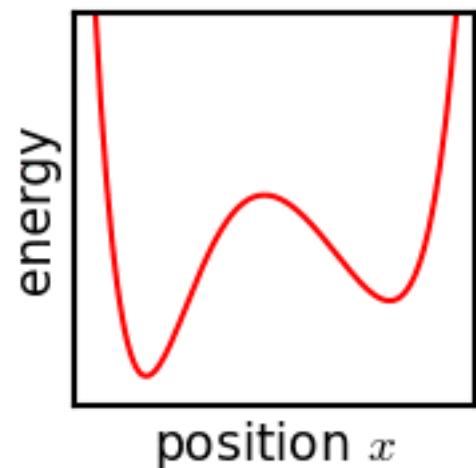
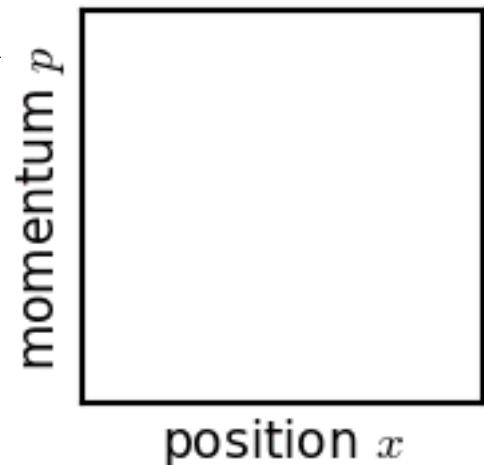
https://en.wikipedia.org/wiki/Liouville's_theorem_%28Hamiltonian%29#/media/File:Hamiltonian_flow_classical.gif

For stationary systems this implies that we have a density function $f(\mathbf{q}, \mathbf{p})$ that allows us to write an ensemble average of a phase space function A as follows:

$$1) \quad \langle A \rangle = \int f(\mathbf{q}, \mathbf{p}) A(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p}$$

the time average of this function is then defined by:

$$2) \quad \langle A \rangle_{\text{time}} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t A(t) dt$$



How do we ensure we are in equilibrium conditions?

Ergodicity

The most important, physically relevant, statement of the ergodic theorem is that 1) and 2) are equal, i.e. the ensemble average and time average of phase space functions are the same.

This leads to an important interpretation regarding the time evolution of an ergodic system, which is that all the regions of the accessible part of phase space are visited by the system regardless of the initial condition at $t=0$ and that the system spends an equal amount of time in all of them.

This also intuitively explains why the averages 1) and 2) are meant to be equal.

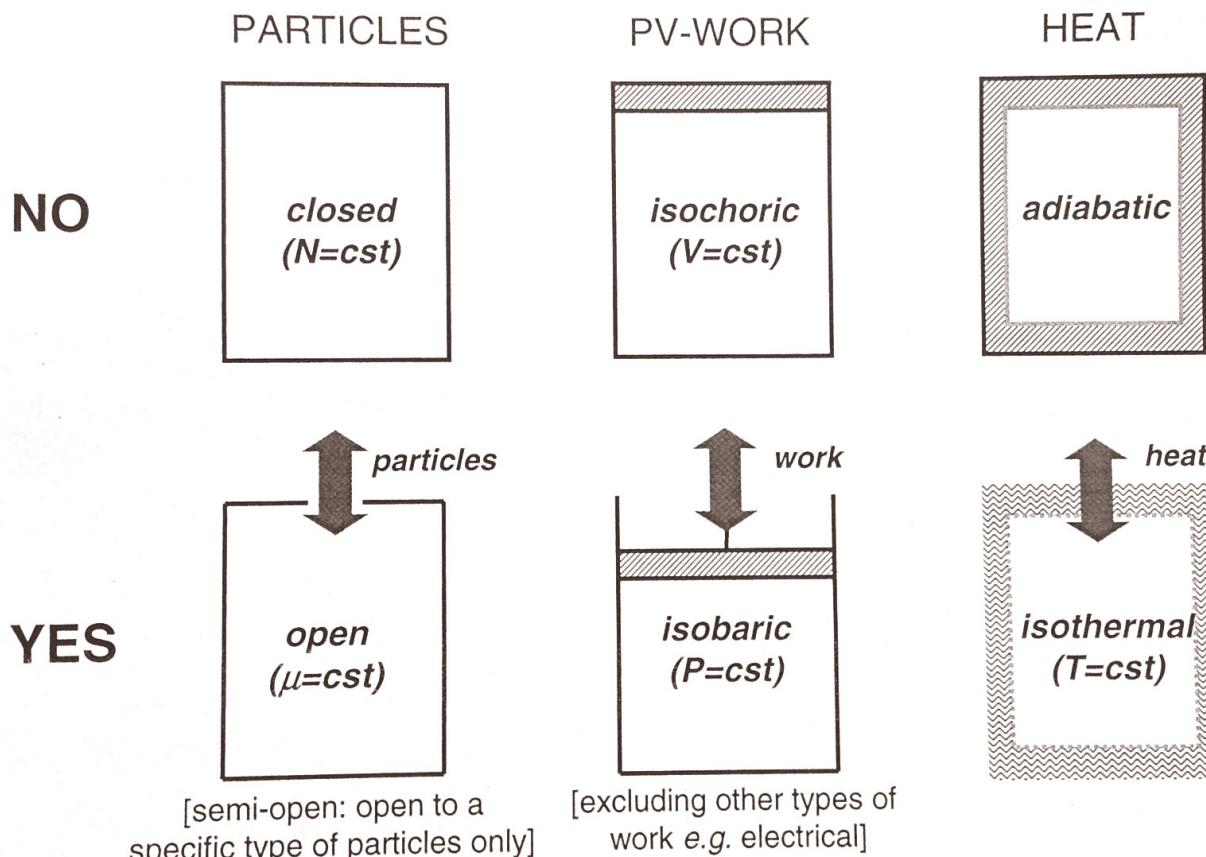
How do we maintain equilibrium conditions?

- Atomistically we require a 'detailed equilibrium', meaning that each microscopic step is reversible.
- Macroscopically we keep a set of system parameters constant, for example the number of particles, temperature and volume (see following slides).
- The thermodynamic definition of equilibrium is (deceptively) simple:
$$\Delta G = 0$$
.
This implies the absence of any driving force for spontaneous processes and therefore the absence of any net flows of matter or energy.

Ensembles

THERMODYNAMICAL BOUNDARY CONDITIONS

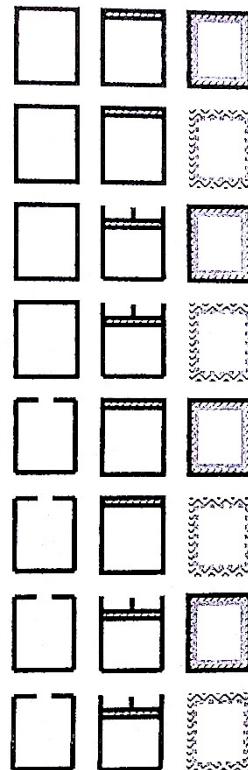
Thermodynamical systems (and the associated statistical-mechanical ensembles) may be classified according to what they are allowed to exchange with their surroundings



How do we maintain EQUILIBRIUM?

THERMODYNAMICAL BOUNDARY CONDITIONS (2)

This leads to eight types of thermodynamical systems (and associated statistical-mechanical ensembles)



microcanonical

canonical

isoenthalpic-isobaric

isothermal-isobaric (Gibbs)

grand-microcanonical*

grand-canonical

grand-isoenthalpic-isobaric*

generalized

independent variables

NVE

NVT

NPH

NPT

μVL

μVT

μPR

μPT

fully determine the thermodynamical state of the given system
MD by default!

MC, SD by default !

$H = \text{enthalpy}$
 $= E + PV$

*: unofficial names

$L = \text{Hill energy}$
 $= E - \mu N$

$R = \text{Ray enthalpy}^*$
 $= E + PV - \mu N$

Size undefined !
only intensive variables

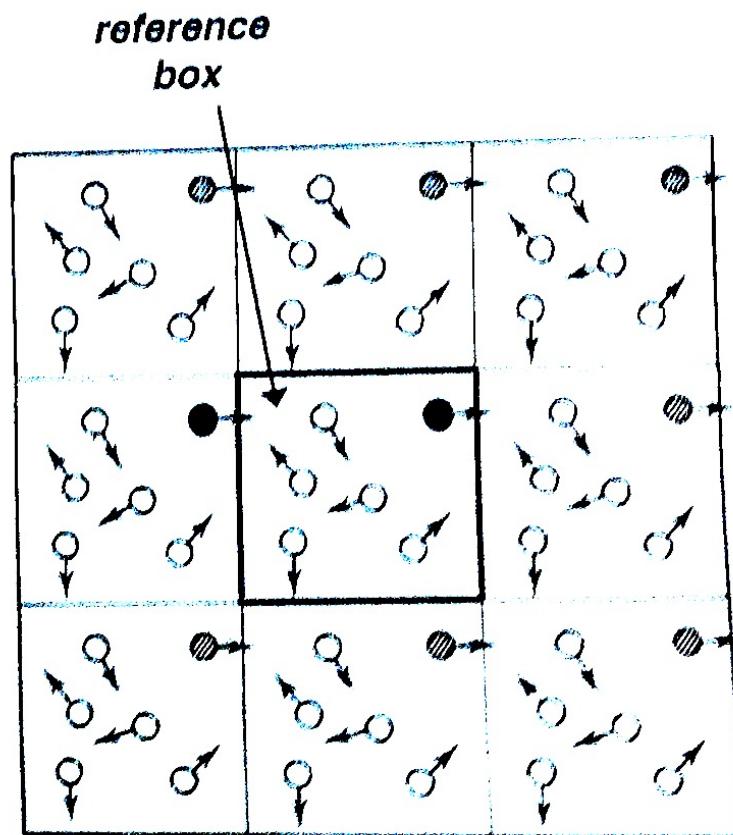
- **extensive** → additive over subsystems, exactly conserved in the system (i.e. instantaneous value is constant)
- **intensive** → determined by the surroundings (i.e. equilibrium average matches value)

[for a n -component system, N and μ are n -dimensional vectors, and semi-grand-ensembles can be defined; for non-confined systems in vacuum, $P=0$ and V is undefined]

How do we maintain CONCENTRATION?

PERIODIC BOUNDARY CONDITIONS (1)

- The simulated system (solute + solvent) consists of particles within a reference computational box of space-filling shape (e.g. cube)
- At each simulation step, particles exiting the box through one face are translated so that they reenter the box through the opposing face
- This procedure mimics a system consisting of an infinite lattice of periodic copies of the reference box (\rightarrow no interface to vacuum !)
- Only the coordinates of particles in the central box are actually stored in the computer

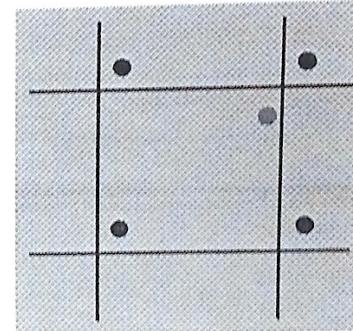


How do we maintain CONCENTRATION?

PERIODIC BOUNDARY CONDITIONS (2)

In principle, each particle in the reference box has non-bonded interactions with all other particles in the reference box and all their replicas in the periodic system

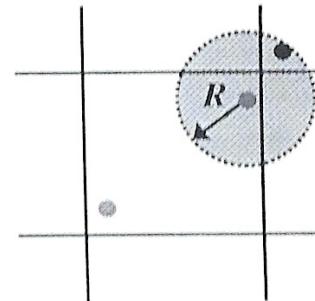
→ interaction may be evaluated using *lattice-sum methods* (Fourier series)



In practice, the non-bonded interaction is often truncated at a certain *cutoff distance* R

→ interaction are evaluated using a *double sum* over particles with a minimum-image distance smaller than R (in general smaller than the half box edge)

Covalent interactions are short-ranged and only act between minimum images



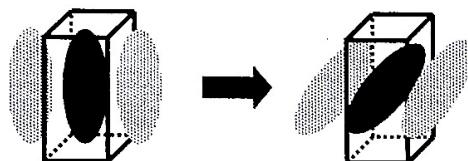
(minimum-image pair: atom and the closest periodic replica of another one)

How do we maintain CONCENTRATION?

PERIODIC BOUNDARY CONDITIONS (3)

Common box shapes (space-filling):

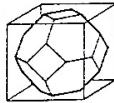
- Rectangle: for elongated macromolecules, but watch out if the molecule rotates



- Hexagonal prism: idem (may be used for DNA)
- Cube: isotropic, but requires a lot of solvent
- Truncated octahedron: almost isotropic, and requires less solvent for spherical molecules



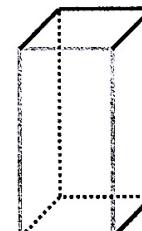
$$\begin{aligned} V &= L^3 \\ R_i &= L/2 \\ \Rightarrow V / V_i &\approx 1.9 \end{aligned}$$



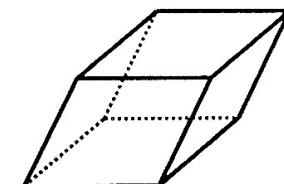
$$\begin{aligned} V &= L^3/2 \\ R_i &= \sqrt{3}L/4 \\ \Rightarrow V / V_i &\approx 1.5 \end{aligned}$$

- Triclinic: for crystal simulations and, more recently, for implementing any (optimal) box shape into a single simulation code

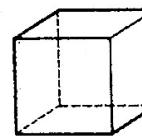
[as shown by Bekker, JCC **18** 1930 (1997), a simulation in any box shape can equivalently be carried out in a triclinic box]



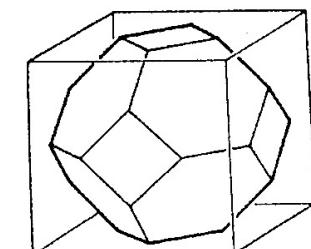
Rectangular



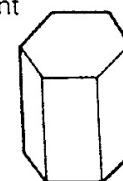
Triclinic



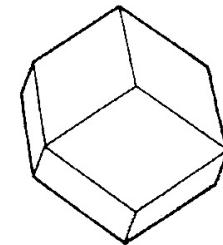
Cube



Truncated octahedron



Hexagonal prism

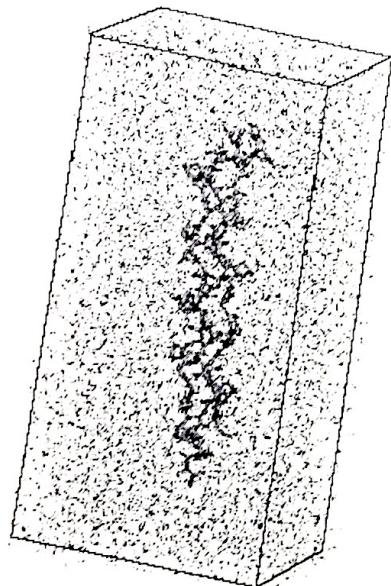


Rhombic dodecahedron

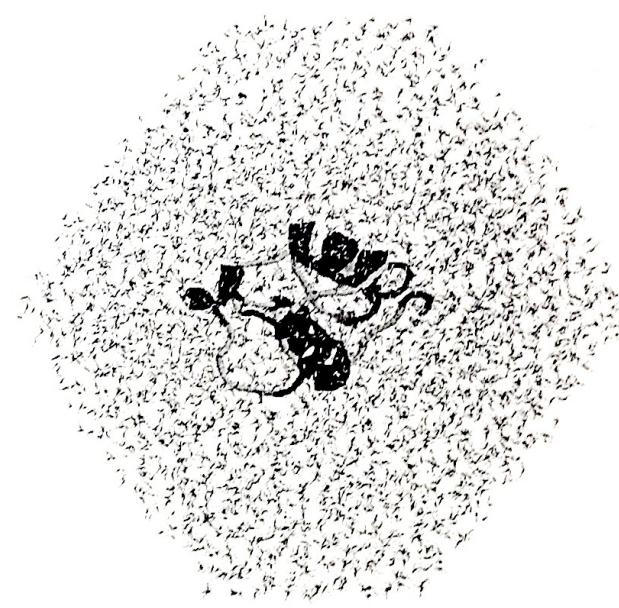
How do we maintain CONCENTRATION?

PERIODIC BOUNDARY CONDITIONS (4)

Examples:



*collagen peptide
in a rectangular box
[possibly not so clever]*



*prion protein in a
truncated octahedron*

Extension Liouville's theorem to non-equilibrium

The question is, how is the ergodic theorem generalized to also account for non-equilibrium systems, i.e. systems with dissipative dynamics and with sources/sinks of particles?

For one thing we no longer have the volume preserving transformation, thus Liouville's theorem is violated and the phase space volume is no longer incompressible under time transformations. Thus evaluating the convergence of (1) and (2) becomes non-trivial.

Web resources

Introductory MD simulation tutorials and How-to

GROMACS (freely available)

http://www.gromacs.org/Documentation/Tutorials#General_GROMACS_Use

<http://www.gromacs.org/Documentation/How-tos>

NAMD (freely available)

<http://www.ks.uiuc.edu/Training/Tutorials/index-all.html#namd>

AMBER

http://ambermd.org/tutorials/#basic_tut

VMD (visualisation of MD trajectories, freely available)

<http://www.ks.uiuc.edu/Training/Tutorials/vmd/tutorial-html/>

YASARA

<http://www.yasara.org/movies.htm>

Servers

MDWeb

<http://mmb.irbbarcelona.org/MDWeb/> (System set-up)

We-NMR

<https://www.wenmr.eu/wenmr/molecular-dynamics-software> (run simulations on the GRID, requires registration)

Databases

MolMovDB (database of macromolecular motions, contains morphs between different conformations)

<http://www.molmovdb.org/>

References

Andrew Leach, *Molecular Modelling: Principles and Applications* (2nd ed.), Prentice Hall, 2001.

Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).

Deupi, X. Molecular dynamics: A stitch in time. *Nat Chem* **6**, 7–8 (2014).

van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew Chem Int Ed Engl* **45**, 4064–4092 (2006).

Nussinov, R. & Tsai, C.-J. Allostery in disease and in drug discovery. *Cell* **153**, 293–305 (2013).

Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6679–6685 (2005).

Genheden S, & Ryde U. Will molecular dynamics simulations of proteins ever reach equilibrium? *Phys Chem Chem Phys.* 14(24):8662-77 (2012)

Hünenerger P. *Adv. Polym. Sci.* Thermostats Algorithms for Molecular Dynamics Simulations **173**:105–149 (2005)