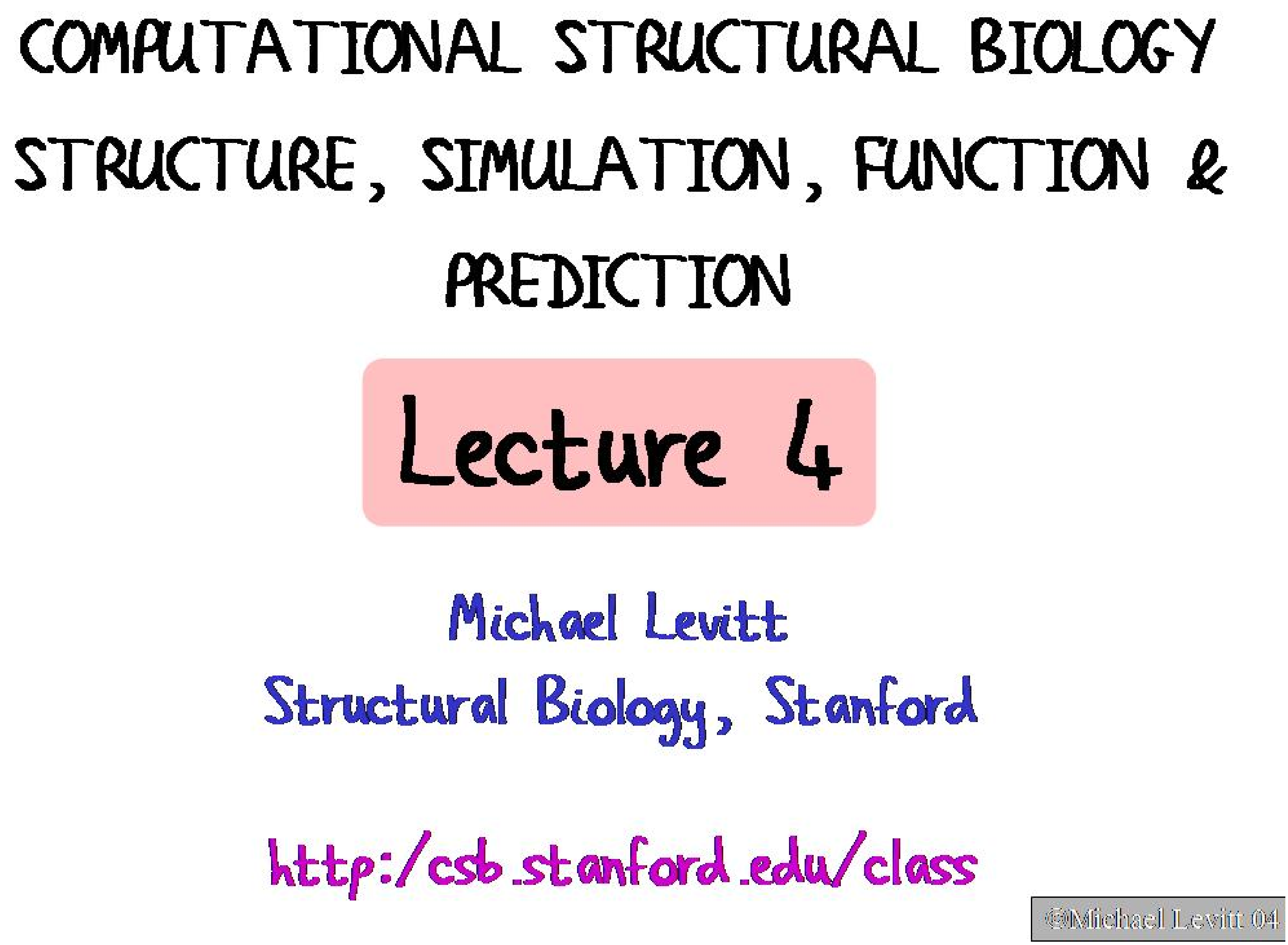


COMPUTATIONAL STRUCTURAL BIOLOGY STRUCTURE, SIMULATION, FUNCTION & PREDICTION

Lecture 4

Michael Levitt
Structural Biology, Stanford

http:/csb.stanford.edu/class

LECTURE 4 CONTENTS

Underlying Principles.

Simulation Methods.

Molecular Dynamics.

Simulating Solutions.

Simulating Alpha-Helix.

Simulating Folded Protein.

History of Simulation.

Underlying Principles
Concept 4.1

UNDERLYING PRINCIPLES

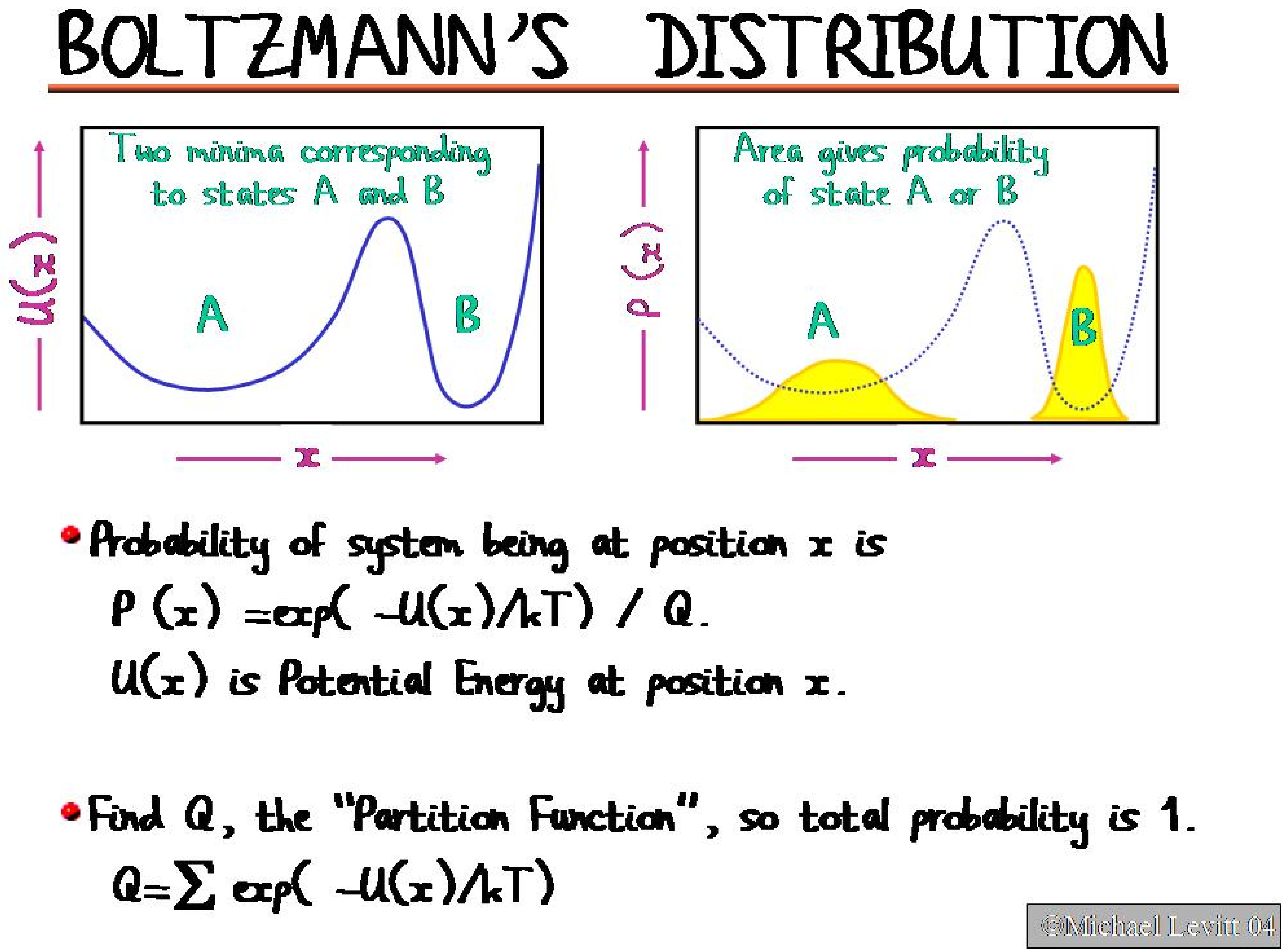
Boltzmann's Distribution.

Entropy & Free Energy.

Dynamic Properties.

Crossing Energy Barriers.

Rates of Motion.



BOLTZMANN'S DISTRIBUTION

- Probability of system being at position x is

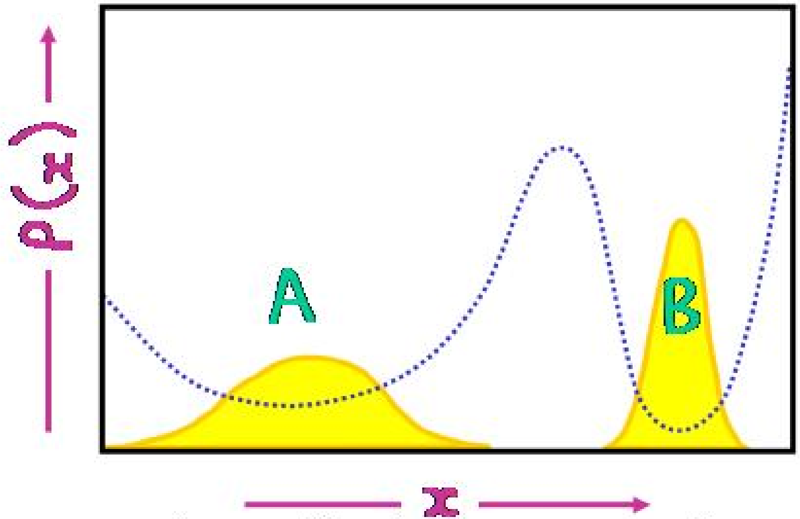
  $P(x) = \exp(-U(x)/kT) / Q.$

  U(x) is Potential Energy at position x.

- Find Q, the "Partition Function", so total probability is 1.

  $Q = \sum \exp(-U(x)/kT)$

ENTROPY & FREE ENERGY



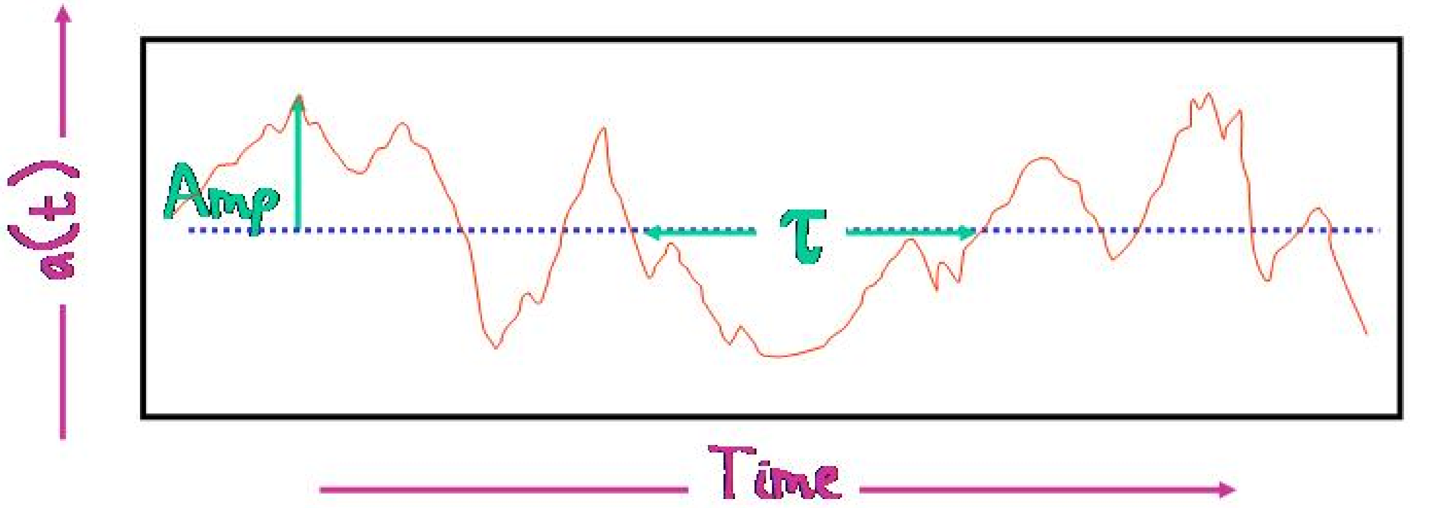
- At each position x, potential energy, $U(x)$ and probability, $P(x)$, are well defined.
- The Free-Energy and Entropy are only defined for states A and B.
- Enthalpy of state A, U_A, is the mean value of the potential energy. $U_A = \sum P(x)U(x)$.

Boltzmann's Constant

- Entropy of State A, S_A, is $S_A = -k_B \sum P(x) \log_e P(x)$, where sum is over points in State A.
- Free energy of State A is $G_A = U_A - TS_A$, T is temperature.
- The probability of State A is $\exp(-G_A / k_B T)$. At some temperature States A and B are equally likely.

DYNAMIC PROPERTIES

Imagine a(t) is some value that changes with time.



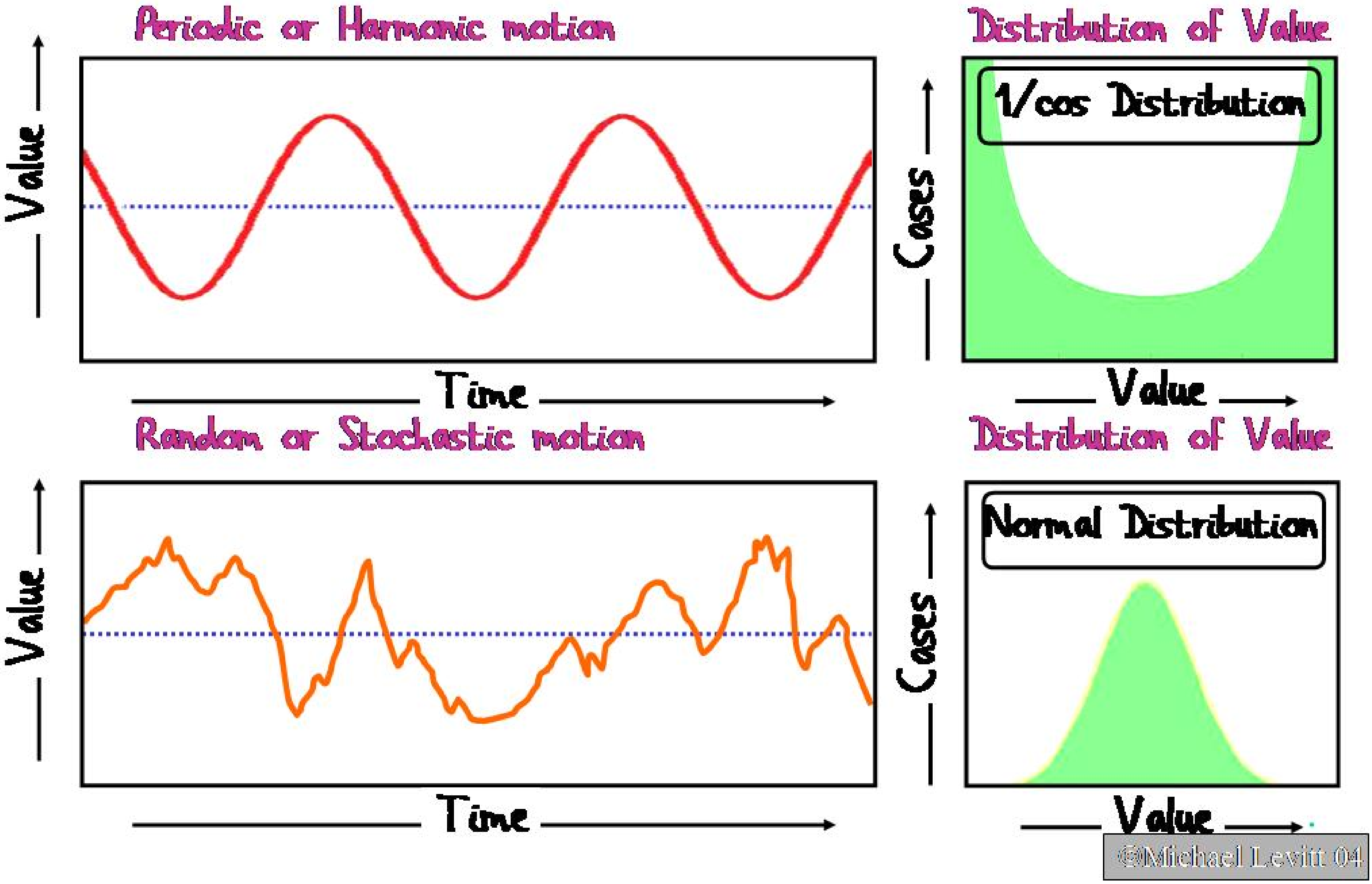
This time-course is characterized by:

- Amplitude calculated as (just like Standard Deviation)

$$\text{Amp} = \text{sqrt}\ \{1/n \sum a^2(t) - [1/n \sum a(t)]^2\ \}$$

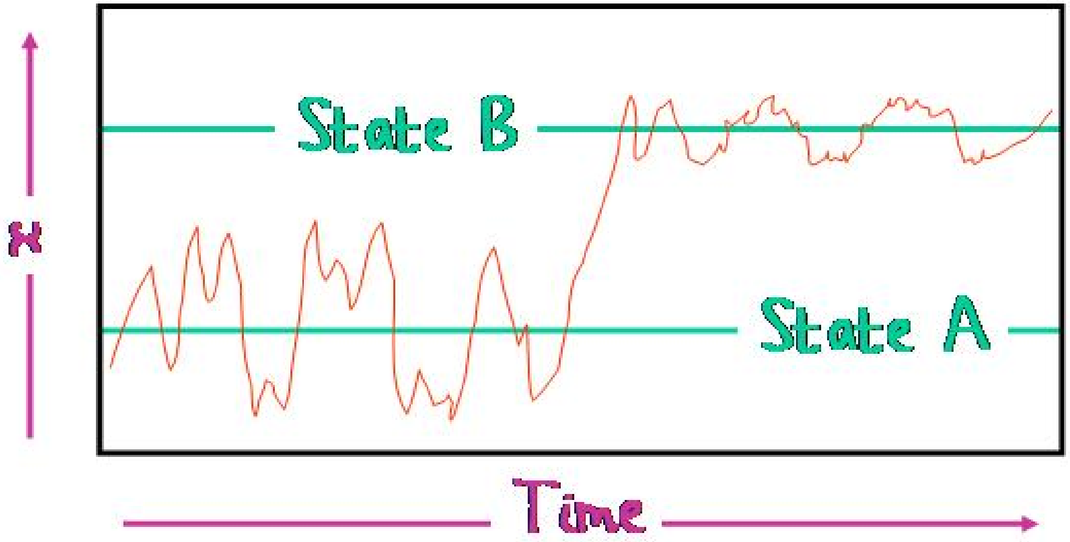
- Rate or time constant, τ (this is a "typical time").

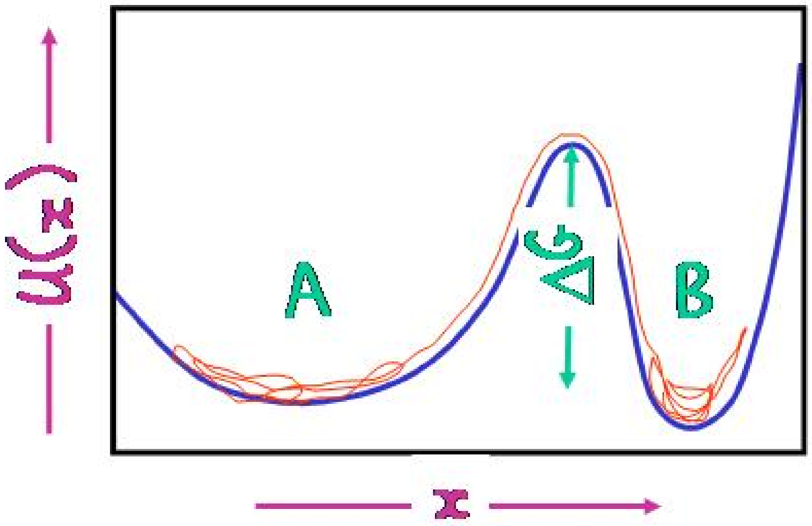
TYPES OF MOTION



©Michael Levitt 04

CROSSING ENERGY BARRIERS



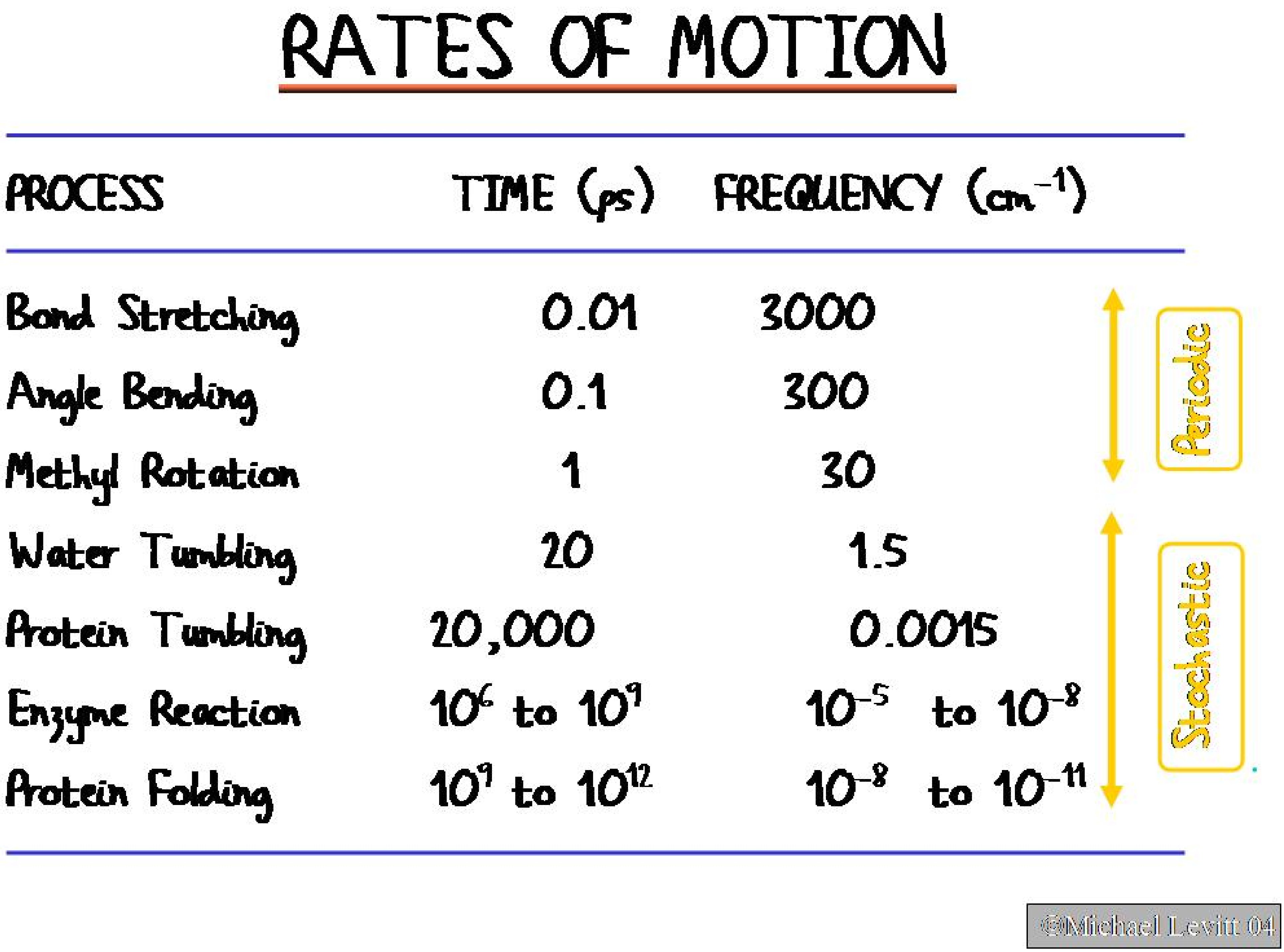


- The actual transition from State A to B is very quick (a few picoseconds).
- What takes time is the waiting. The average wait before going from A to B is:

$$\tau_{A\to B} = (h/k_bT)\exp[+\Delta G/k_bT],\ \text{where}\ \Delta G = (G_T - G_A)$$

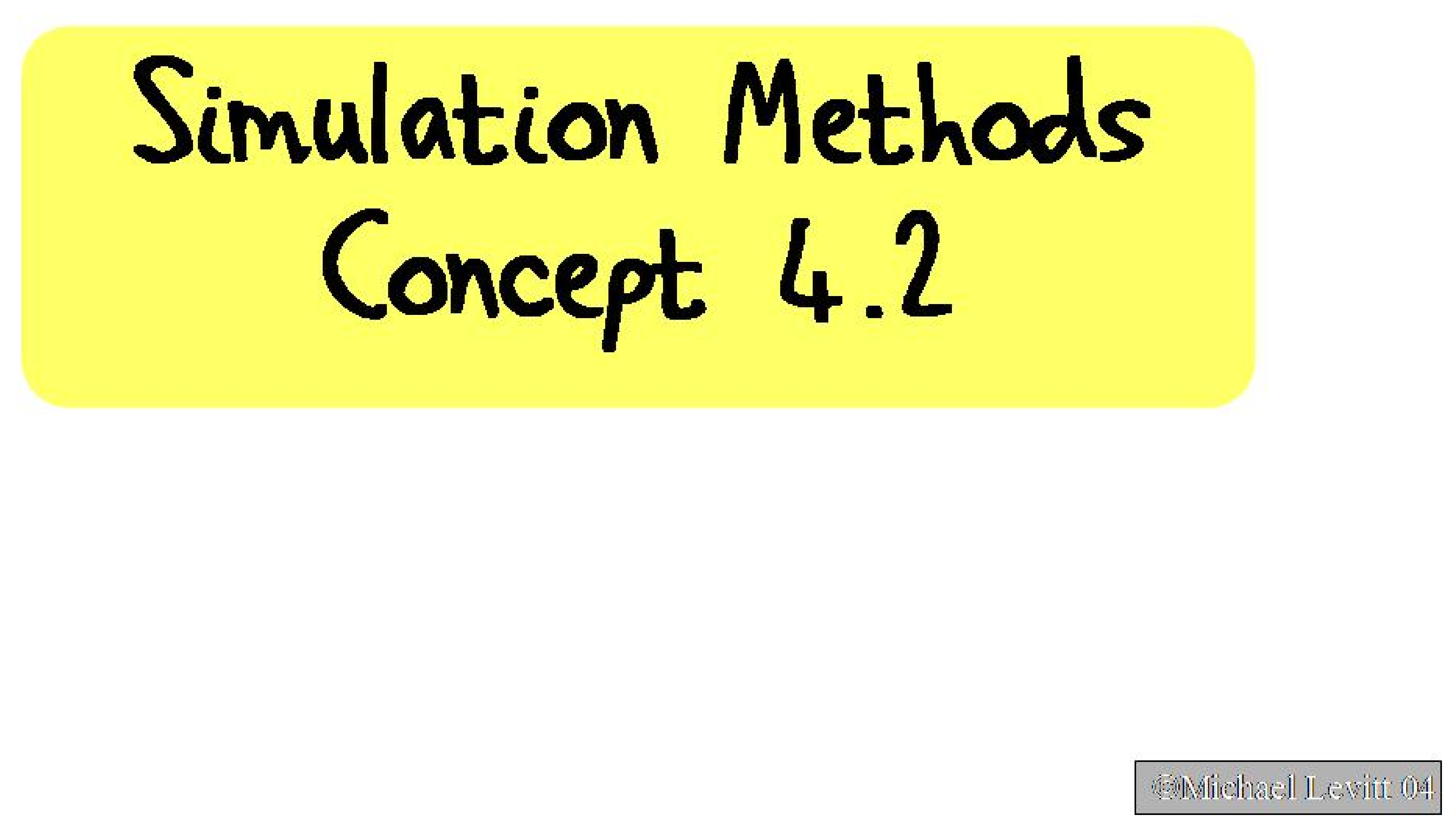
$(h/k_bT) \sim 0.016$ picoseconds at $T = 300^\circ K$ ($27^\circ C$).

h is Planck's constant, k_b is Boltzmann's constant



RATES OF MOTION

PROCESS	TIME (ps)	FREQUENCY (cm^{-1})	
Bond Stretching	0.01	3000	Periodic
Angle Bending	0.1	300	
Methyl Rotation	1	30	
Water Tumbling	20	1.5	Stochastic
Protein Tumbling	20,000	0.0015	
Enzyme Reaction	10^6 to 10^9	10^{-5} to 10^{-8}	
Protein Folding	10^9 to 10^{12}	10^{-8} to 10^{-11}	



Simulation Methods
Concept 4.2

SIMULATION METHODS

Moving Over Energy Surface.

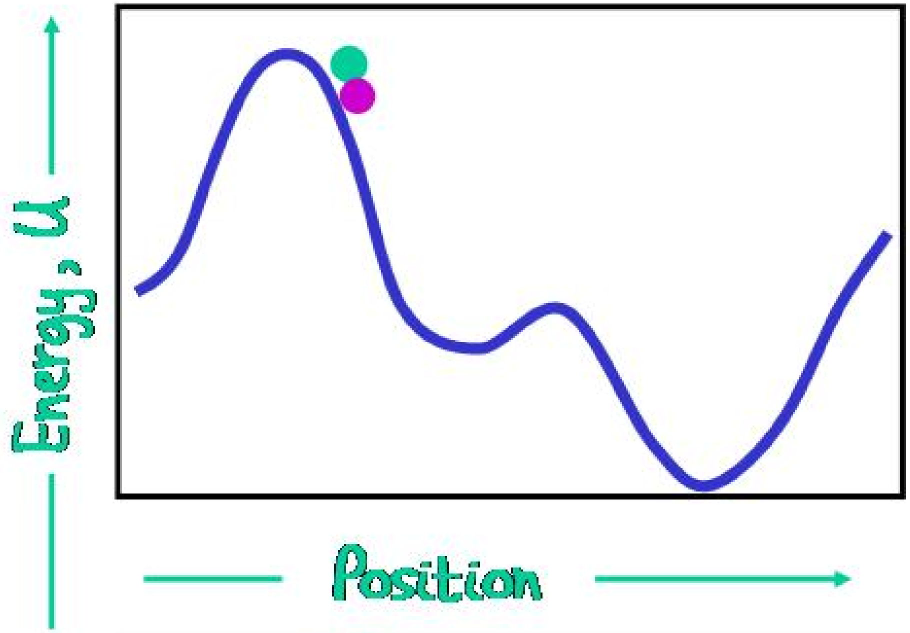
Different Energy Surfaces.

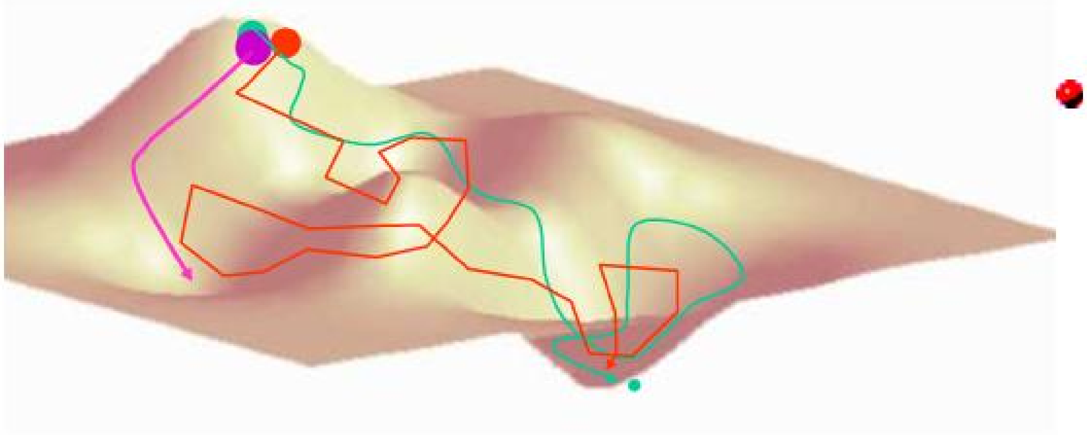
Levinthal's Paradox.

Energy Minimization.

Monte Carlo Method.

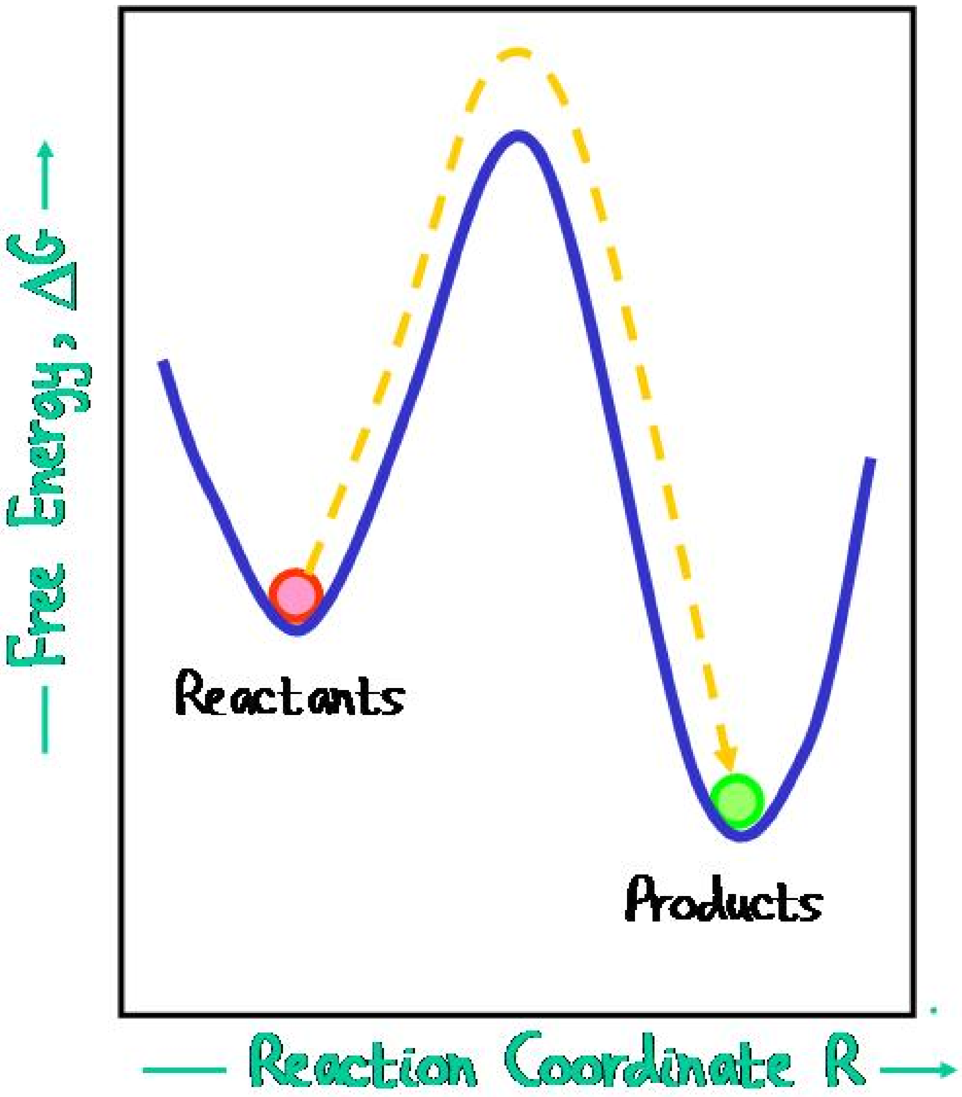
MOVING OVER ENERGY SURFACE





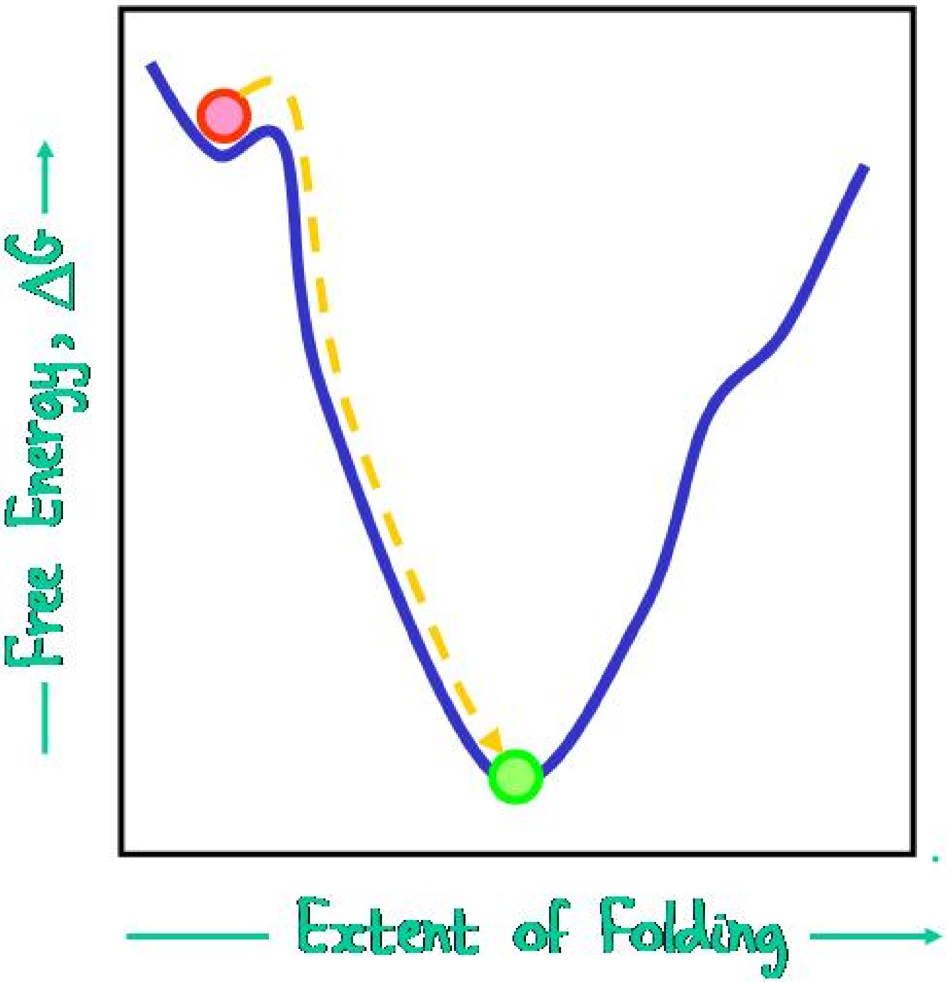
- Energy Minimization drops into local minimum.
- Molecular Dynamics uses thermal energy to move smoothly over surface.
- Monte Carlo Moves are random. Accept with probability $\exp(-\Delta U/kT)$.

REACTION COORDINATES



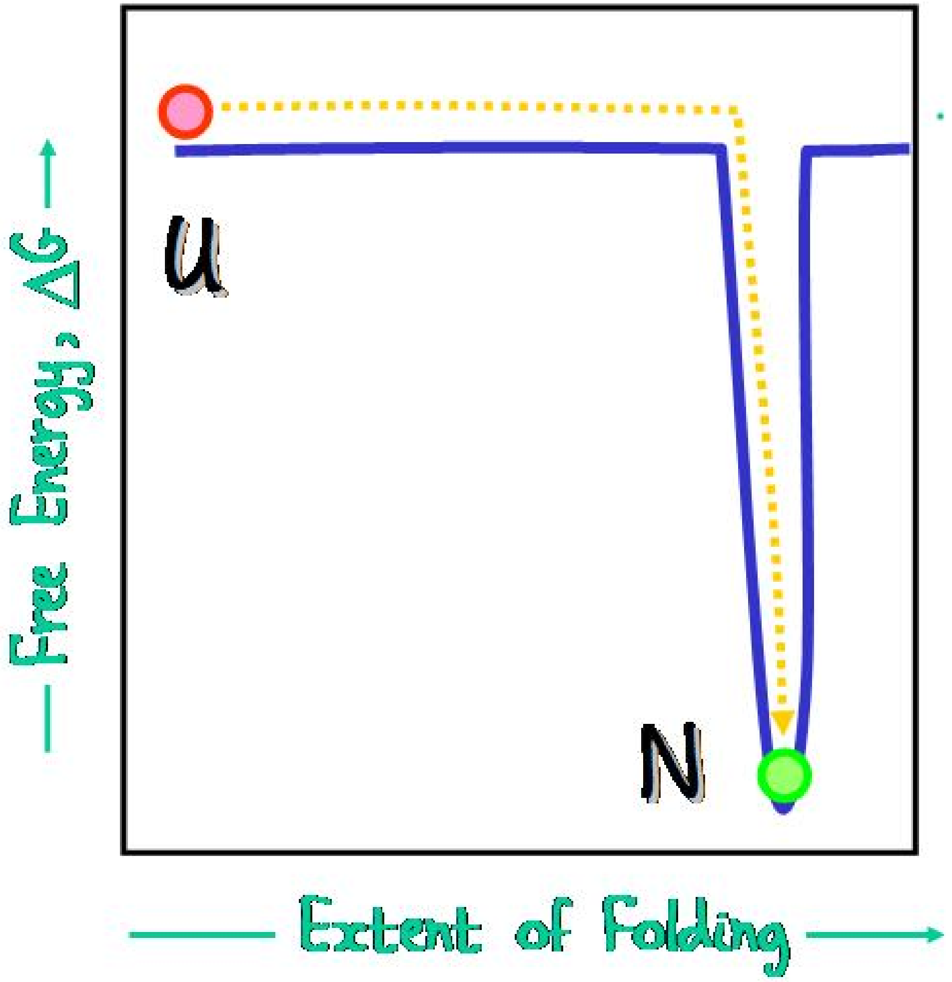
- Simple chemical reactions.
- Height of barrier between the reactant state and the product state makes the reaction slower (high barrier) or faster (low barrier).
- Just like jumping a fence.

PROTEIN FOLDING LANDSCAPE



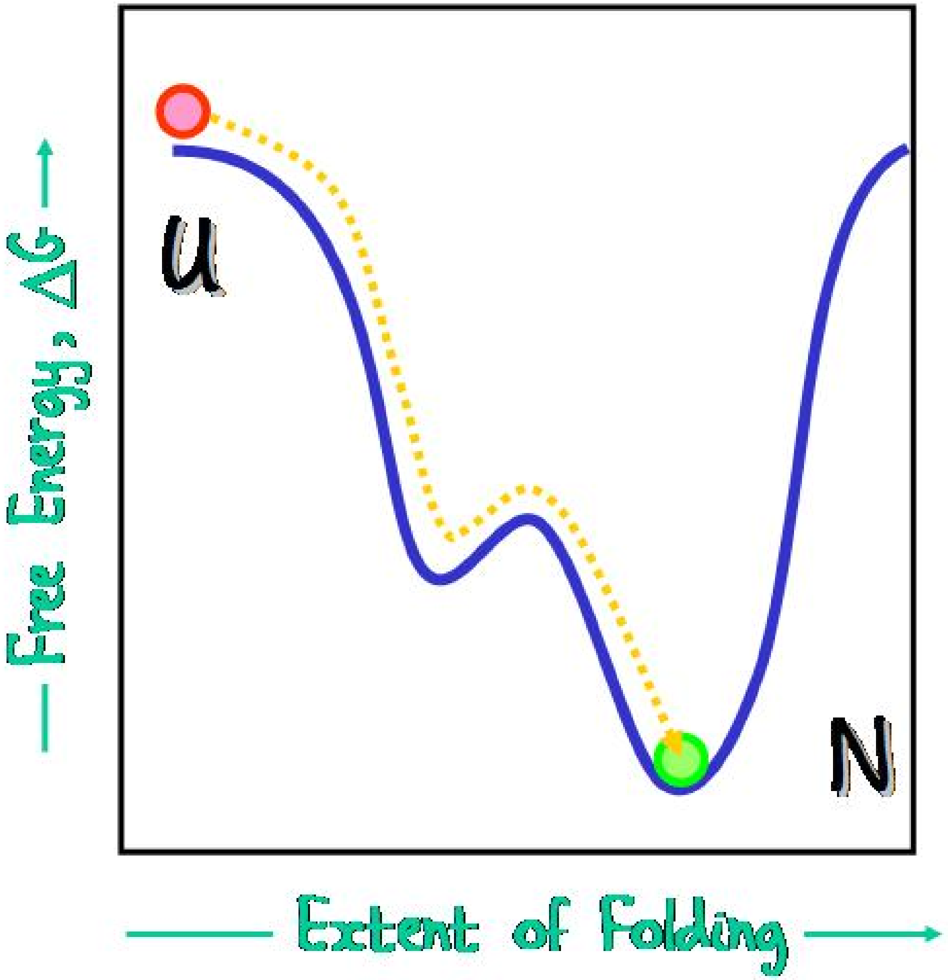
- The native state is believed to have the lowest free-energy Run down to it.
- Run down to it.
- Sounds Easy.

LEVINTHALS PARADOX



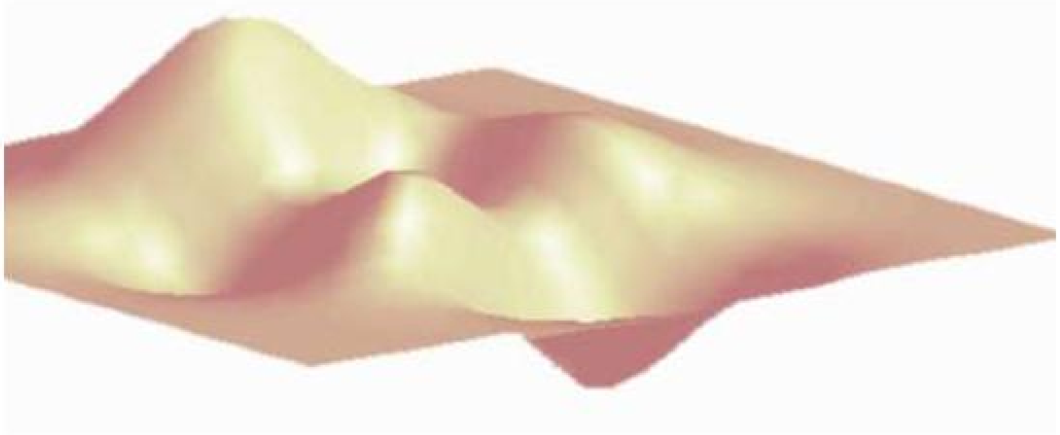
- If the energy surface is flat except for one deep minimum (golf-course) landscape, then proteins would never fold.
  Too many possible states.
- If each of 100 amino acids has 5 states, then the entire chain has 5^{100} states.
  Too big!

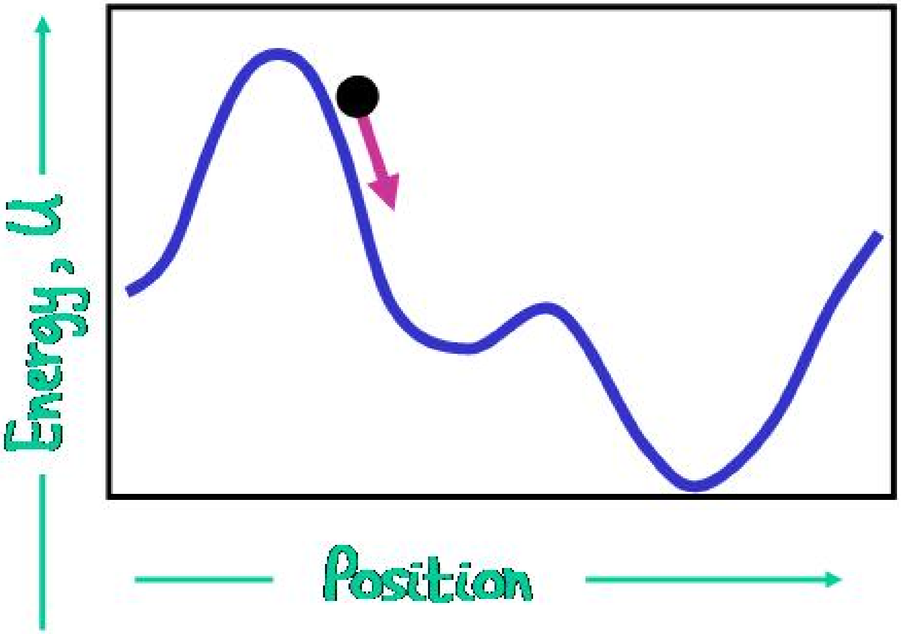
RESOLVE PARADOX



- Why should the free-energy surface be flat? The same energy terms that stabilize the native state should stabilize progressively near-native states.
- This is a folding funnel.
- With it, it does not matter how large the space of possible states is.

SIMPLE ENERGY MINIMIZATION



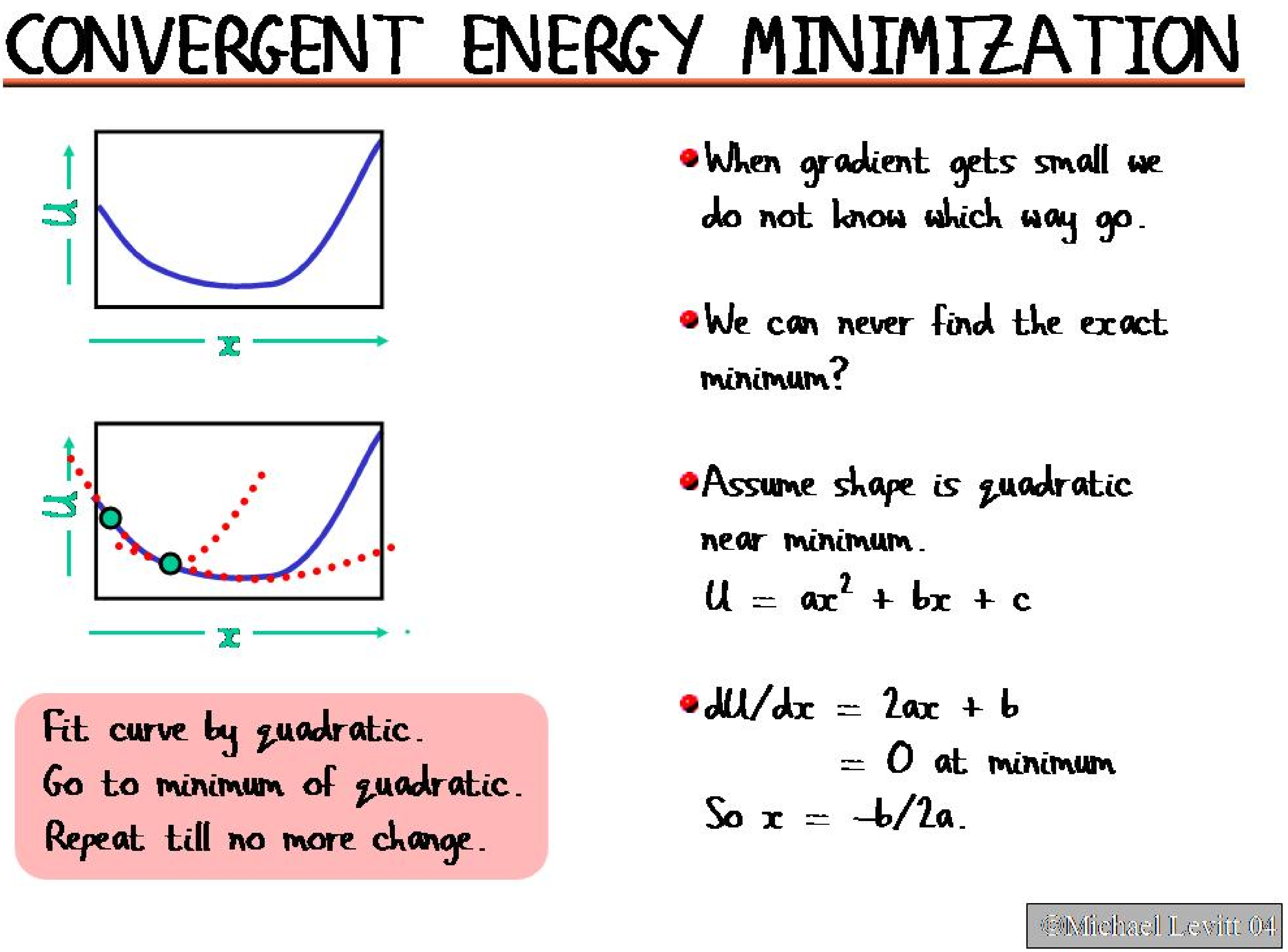


- The total potential energy, U, fully defines the system.
- Energy minima are stable as the net forces are zero.
- The forces are the gradients of the energy.

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

- Move downhill to reach a minimum. $\Delta x = -C\,dU/dx$

Steepest Descent Minimization.



CONVERGENT ENERGY MINIMIZATION

- When gradient gets small we do not know which way go.
- We can never find the exact minimum?
- Assume shape is quadratic near minimum.

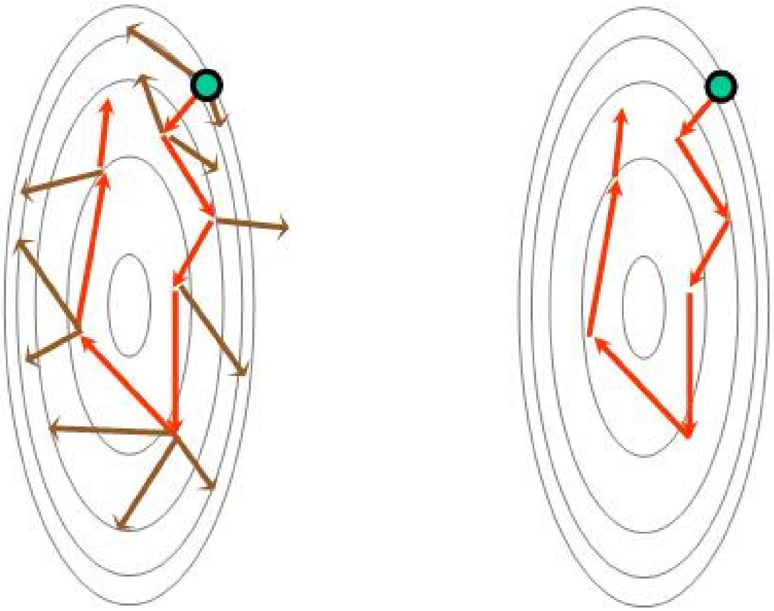
  $U = ax^2 + bx + c$
- $dU/dx = 2ax + b$

  $= 0$ at minimum

  So $x = -b/2a$.

Fit curve by quadratic.
Go to minimum of quadratic.
Repeat till no more change.

MONTE CARLO METHODS



- Normal Monte Carlo: Make random moves and accept some of them according to a special criterion discovered by Metropolis et al. 1955.

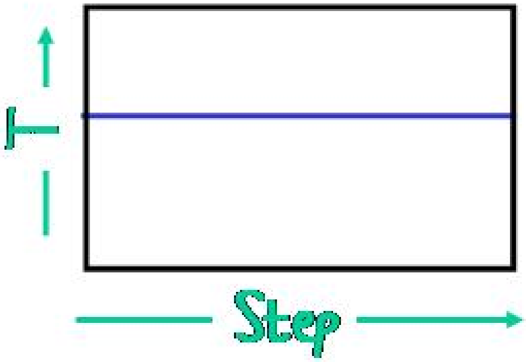
At current position, attempt many moves.

Accept the first move that obeys:

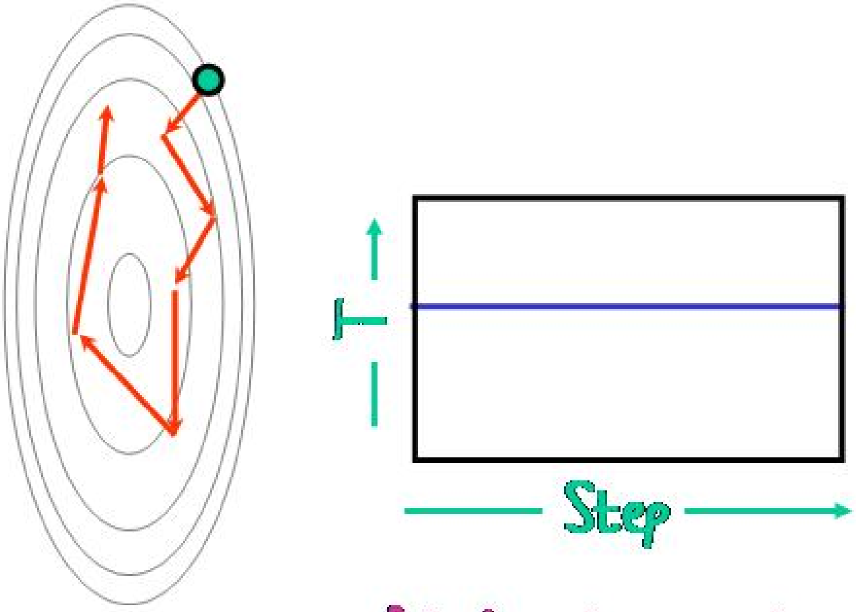
Random number, $R_N < \exp(-\Delta U/kT)$

Always accepts if move it to lower energy

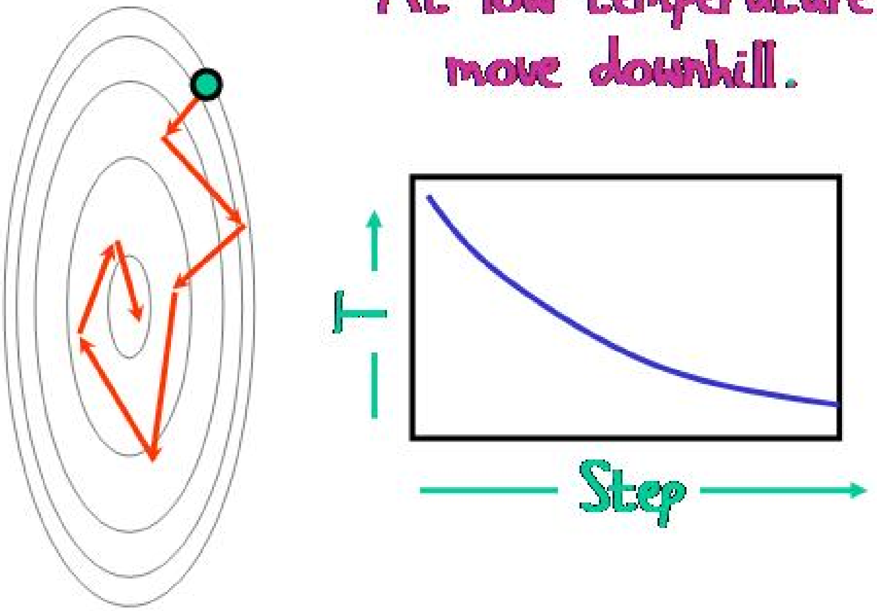
Accept if move is to higher energy with probability $\exp(-\Delta U/kT)$.



SIMULATED ANNEALING



At low temperature move downhill.



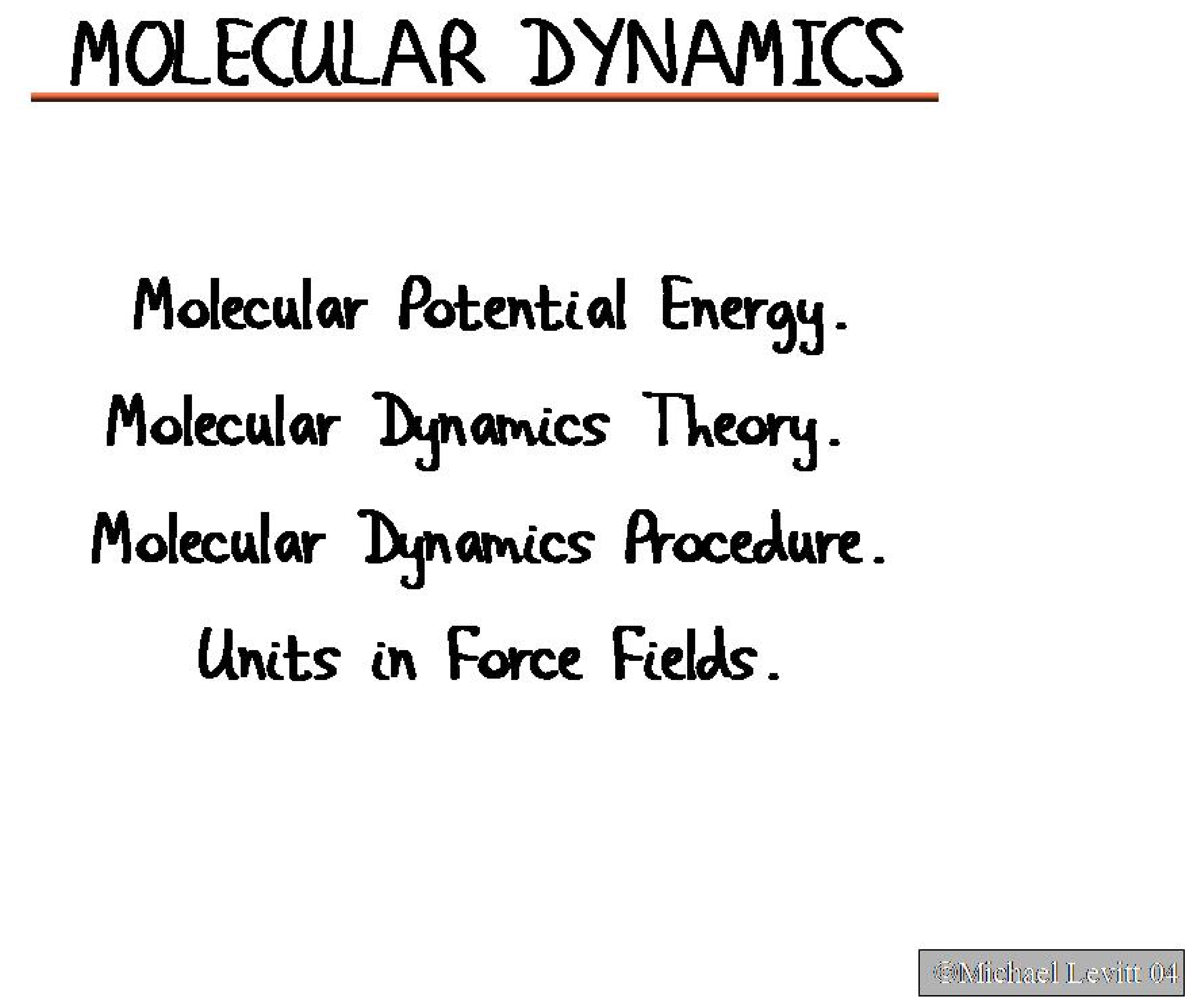
- **Normal Monte Carlo:**

  Make random moves and accept some of them.

- **Simulated Annealing:**

  Reduce T, temperature, as the run proceeds.

Molecular Dynamics
Concept 4.3



MOLECULAR DYNAMICS

Molecular Potential Energy.

Molecular Dynamics Theory.

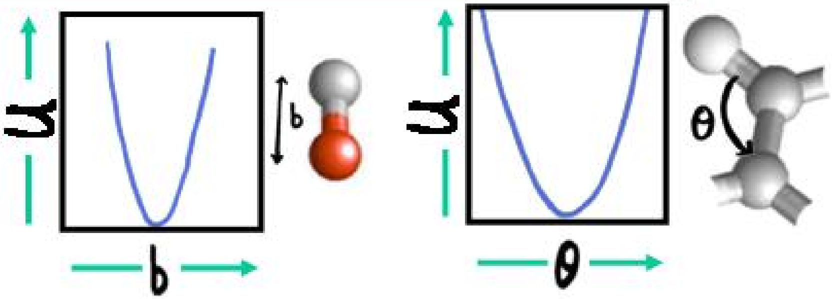
Molecular Dynamics Procedure.

Units in Force Fields.

MOLECULAR POTENTIAL ENERGY

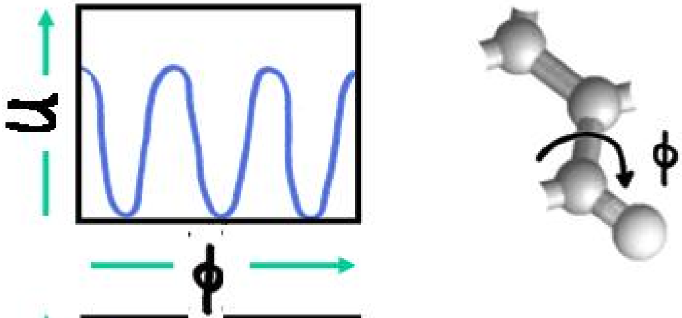
$$U = \sum \tfrac{1}{2} K_b (b - b_0)^2 + \sum \tfrac{1}{2} K_\theta (\theta - \theta_0)^2$$

All Bonds   All Angles



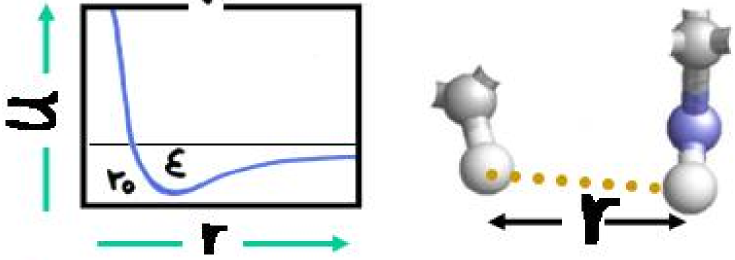
$$+ \sum K_\phi [1 - \cos(n\phi + \delta)]$$

All Torsion Angles



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

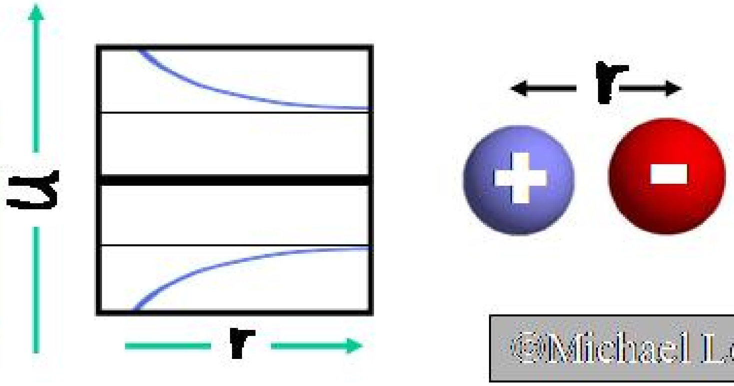
All nonbonded pairs



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

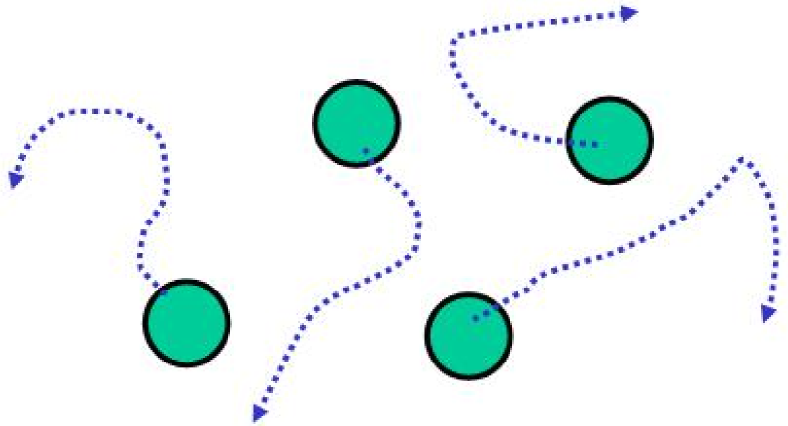
All partial charges

Simple sum
over many
terms



©Michael Levitt 04

MOLECULAR DYNAMICS THEORY



- All atoms move together.
- Forces between atoms change with time.
- Analytical solution to give $x(t)$ and $v(t)$ is impossible.
- Numerical solution is trivial.

$$x(t+\Delta t) = x(t) + v(t)\Delta t + [4a(t) - a(t-\Delta t)]\,\Delta t^2/6$$

New position Old position Old velocity Acceleration

$$v(t+\Delta t) = v(t) + [2a(t+\Delta t) + 5a(t) - a(t-\Delta t)]\,\Delta t/6$$

New velocity Old velocity Acceleration

Number of coordinates (not atoms) Temperature

$$U_{kinetic} = \frac{1}{2}\sum m_i v_i(t)^2 = \frac{1}{2} n k_B T$$

Kinetic energy Atomic masses, velocities Boltzmann's Constant

Total energy ($U_{potential} + U_{kinetic}$) must not change with time

Time step, Δt, must be very small at 10^{-15} seconds or 0.001 ps..

MOLECULAR DYNAMICS PROCEDURE

(1) Starting values for the atomic positions.

(2) Starting values for the atomic velocities.

(3) Iterate: position → forces → new position.

(4) Equilibration.

(5) Running the simulation.

MOLECULAR DYNAMICS PROCEDURE 1

(1) Starting values for the atomic positions.

- Take atomic positions from the pdb coordinates or from a model or just from random positions.
- Add water molecules to fill a rectangular box.
- Relax the entire structure by energy minimization.

(2) Starting values for the atomic velocities.

- Set all initial velocities to zero, corresponding to 0 K.
- Now heat slowly by choosing a random pair of atoms and given them equal but opposite momentum impulses.

MOLECULAR DYNAMICS PROCEDURE 2

(3) Iterate: position → forces → new position.

- Get forces from the total potential energy function.
- Use molecular dynamics theory to get new positions and new velocities for each atom.
- Calculate the total kinetic energy at each step and check the conservation of the total energy.

MOLECULAR DYNAMICS PROCEDURE 3

(4) Equilibration.

- Keep on increasing the temperature until it reaches the desired value (say room temperature).
  This should take tens of ps.

(5) Running the simulation.

- Continue for as many time steps as you can afford.
- Save coordinates every 0.5 ps or so for further analysis (keep every 250-th set of coordinates generated).

UNITS IN FORCE FIELDS

QUANTITY	UNIT
Energy	kcal/mol
Distance	Å
Mass	A.M.U. (H atom is 1 AMU)
Time	0.5×10^{-13} secs.
Frequency	cm^{-1}
Charge	e (electron is 1 e)

This is definitely a strange combination!

Simulating Solutions
Concept 4.4

SIMULATING LIQUIDS.

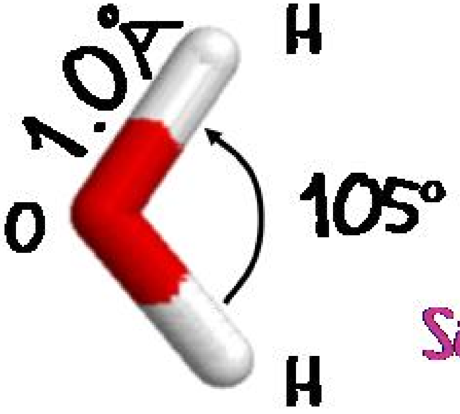
Simulating Pure Liquid Water.

Simulate Solutions.

Simulate the Hydrophobic Effect.

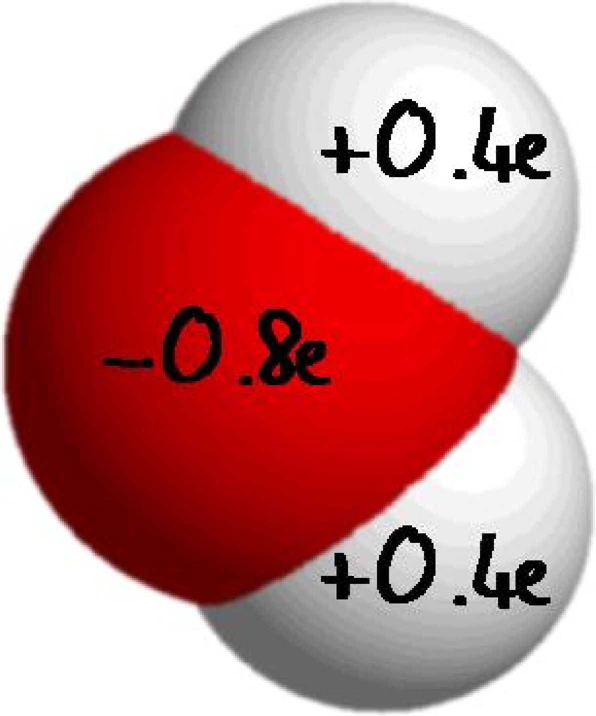
Look at Distribution Details.

WATER IS A VERY SIMPLE MOLECULE



H_2O

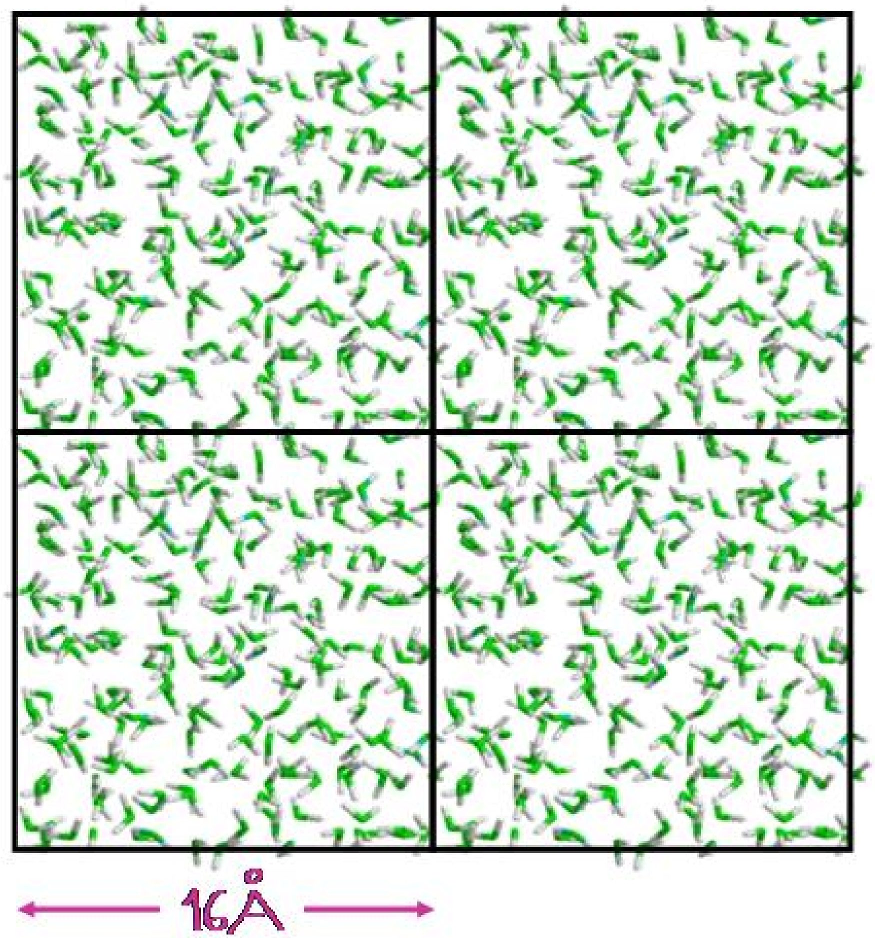
Simple geometry



Simple electrostatics

- The water molecule is one of the simplest.
- The properties of liquid water are not simple.
- Water and solutions have emergent properties.
- Many simple objects lead to complexity.

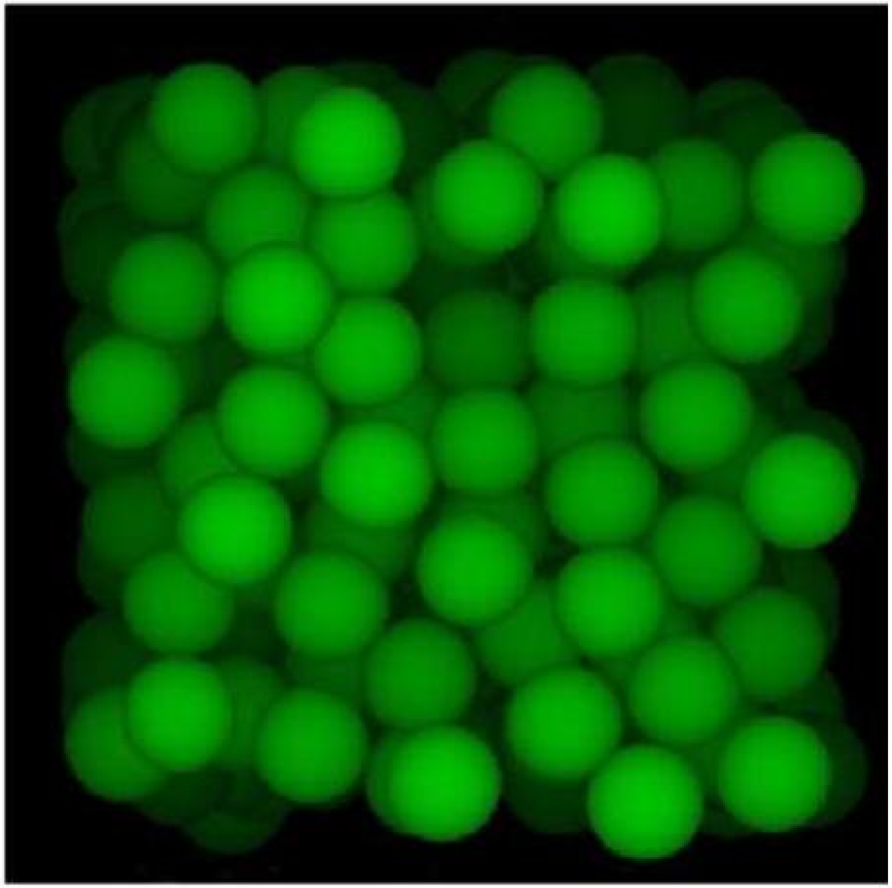
SIMULATING LIQUIDS



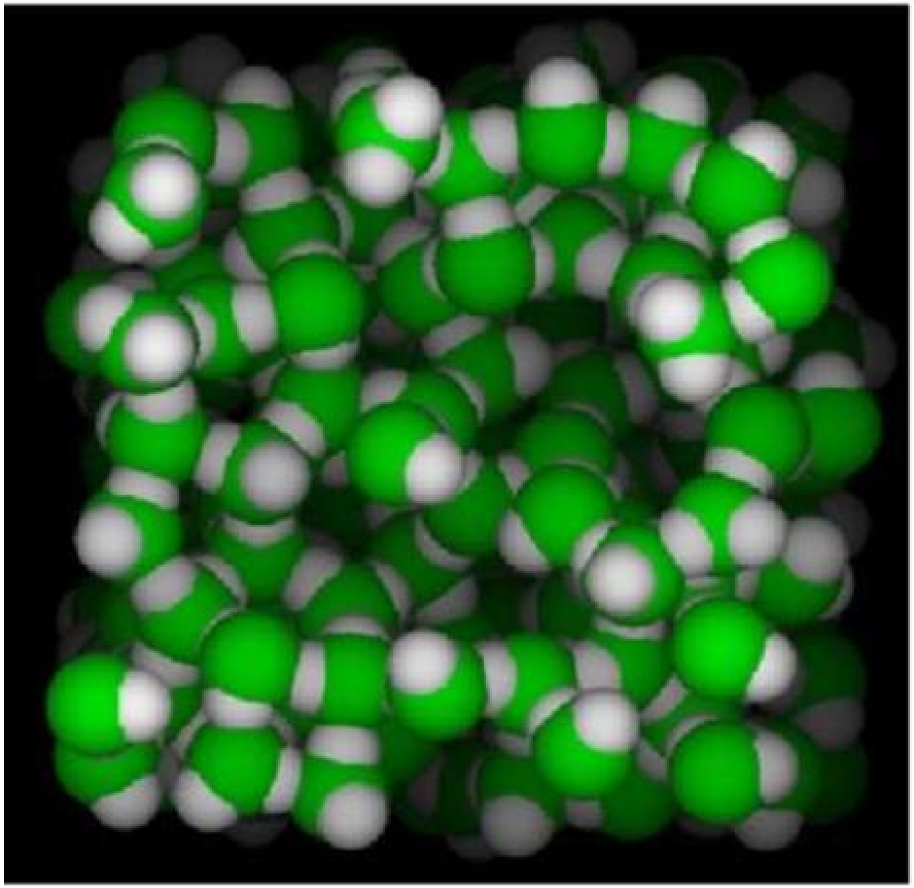
- Periodic box with 216 water molecules. Simulate motion for 100 ps.
- Calculate key experimental properties:
  Heat of vaporization
  Structure.
  Internal pressure
  Diffusion constant
- Compare with experiment and calibrate.

Rahman & Stillinger 1970

LIQUIDS: ARGON AND WATER

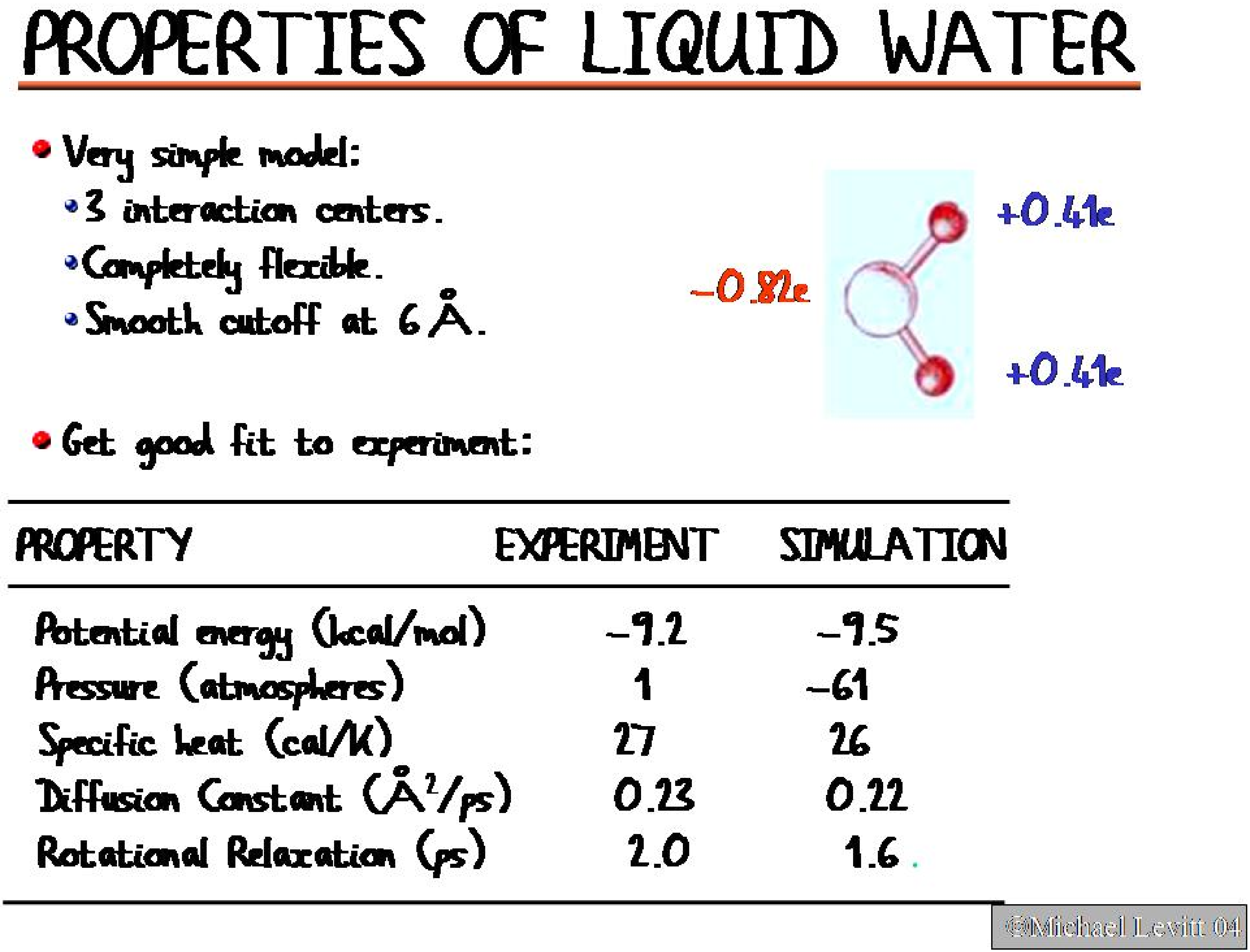


Argon is like a collection of hard spheres. Each Argon has 12 to 14 neighbors.



Water has an open structure. Due to tetrahedral geometry, each water has 4 to 5 neighbors.

PURE WATER DYNAMICS IN WATER AT ROOM TEMPERATURE



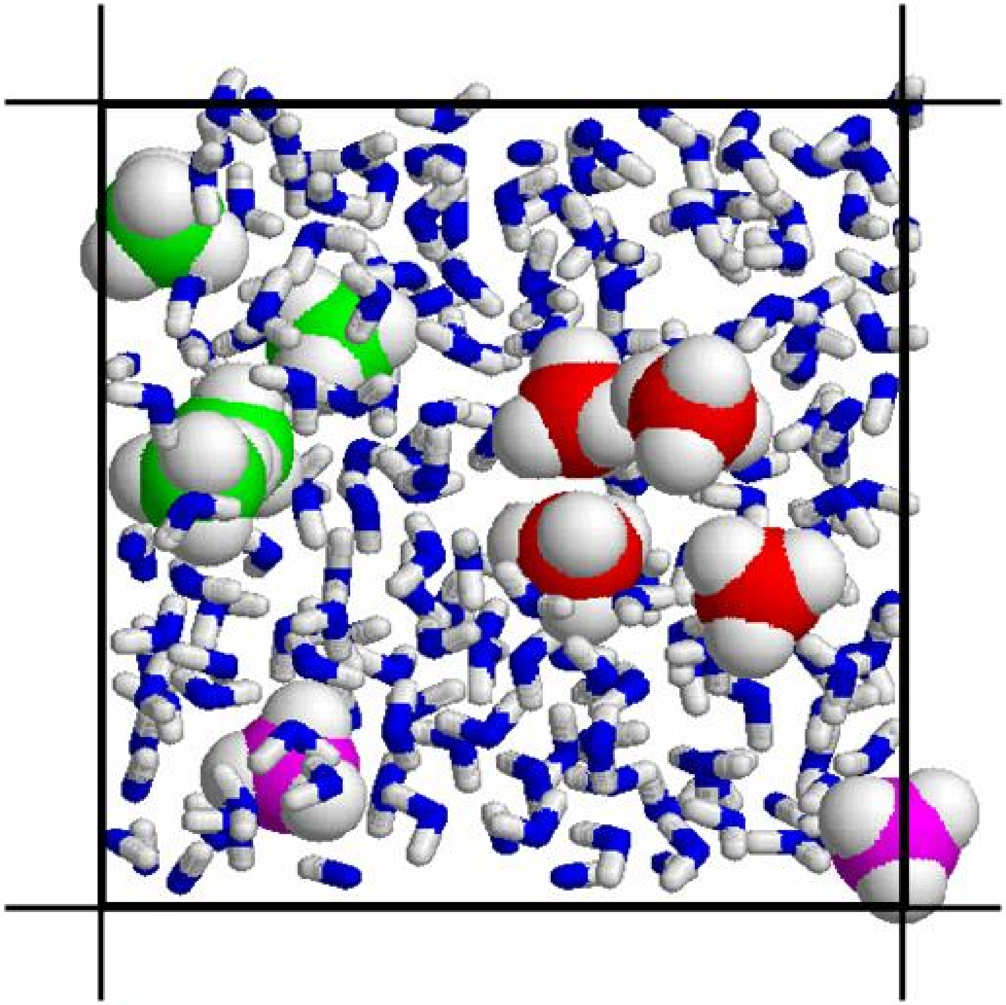
PROPERTIES OF LIQUID WATER

- Very simple model:
  - 3 interaction centers.
  - Completely flexible.
  - Smooth cutoff at 6 Å.

- Get good fit to experiment:

PROPERTY	EXPERIMENT	SIMULATION
Potential energy (kcal/mol)	−9.2	−9.5
Pressure (atmospheres)	1	−61
Specific heat (cal/K)	27	26
Diffusion Constant (Å^2/ps)	0.23	0.22
Rotational Relaxation (ps)	2.0	1.6

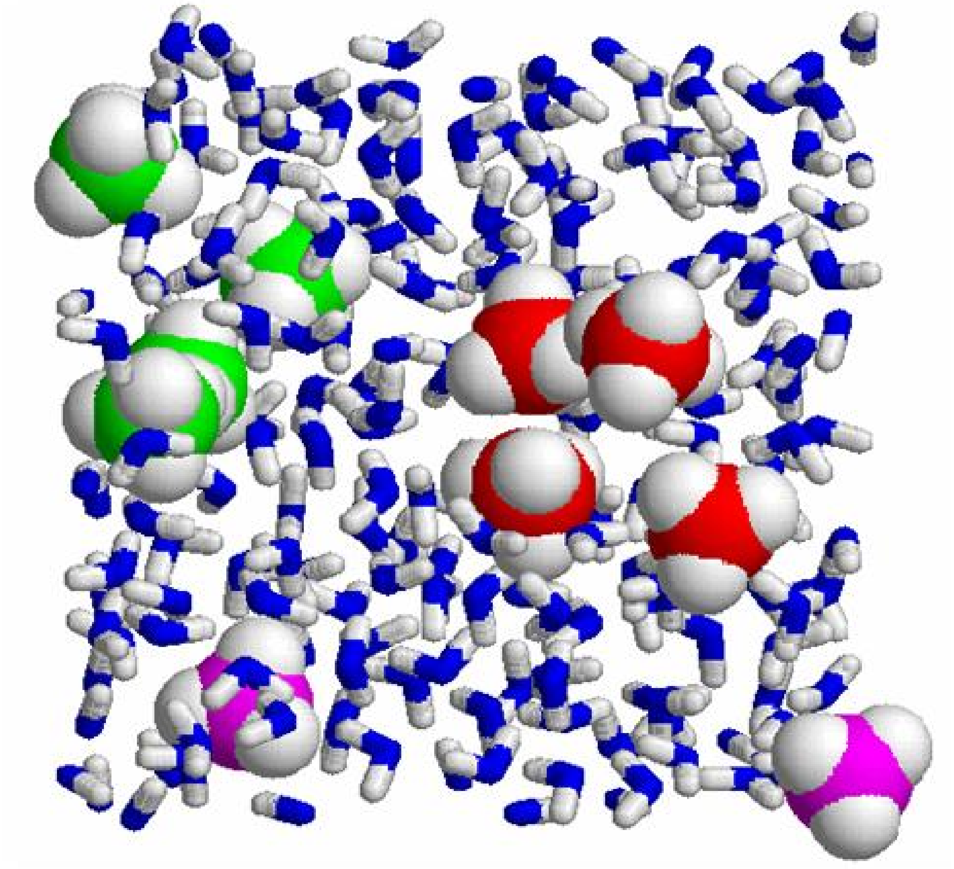
HYDROPHOBIC EFFECT



Box with periodic boundaries.

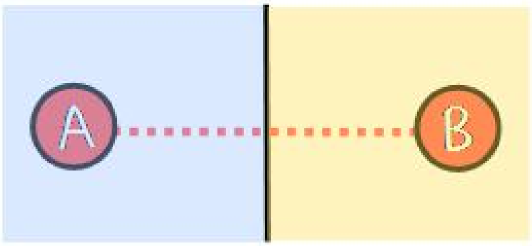
- Hydrophobic molecules cluster in water (not soluble).
- The energy is proportional to the surface area buried in cluster.
- This is not a pairwise additive force.

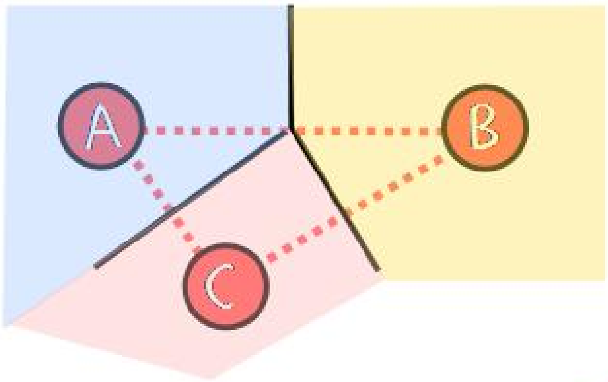
SIMULATING HYDROPHOBIC EFFECT

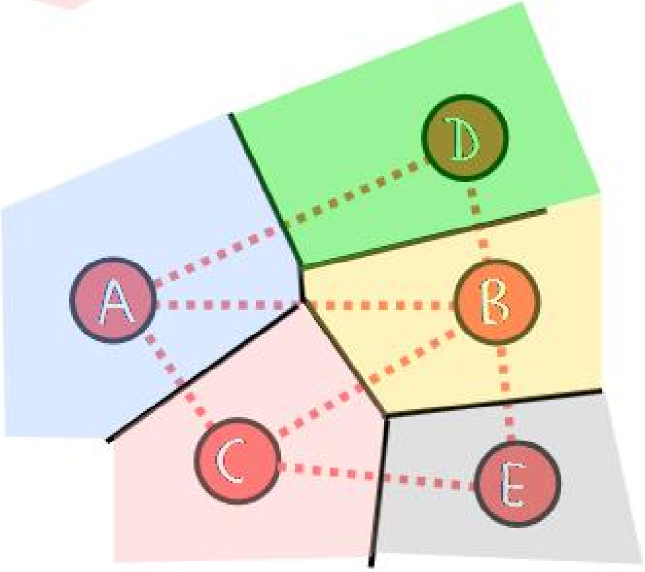


- 1 nanosecond MD simulations.
- Periodic water boxes.
- 30 mM to 3 Molar concentration solution.

VORONOI DECOMPOSITION

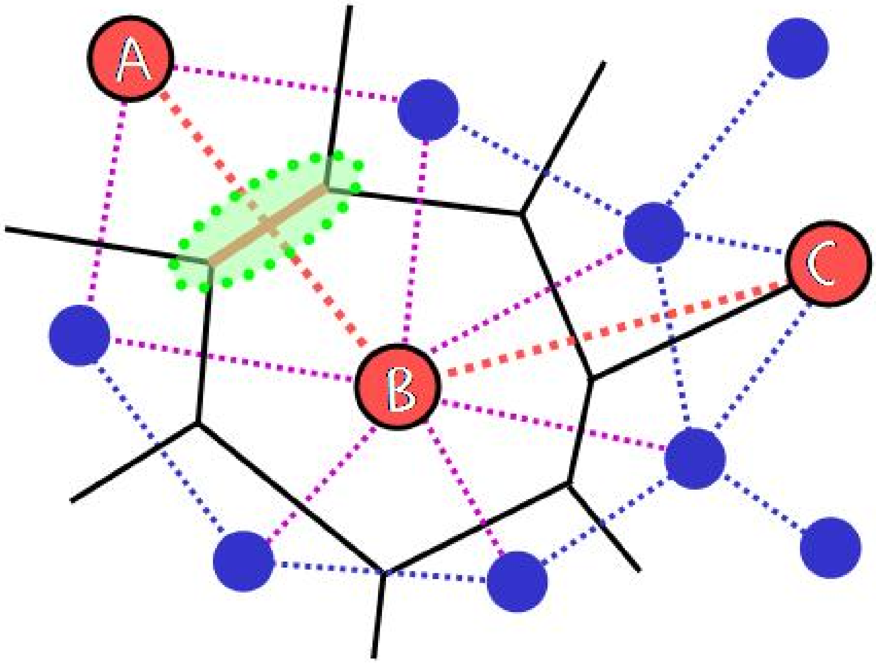






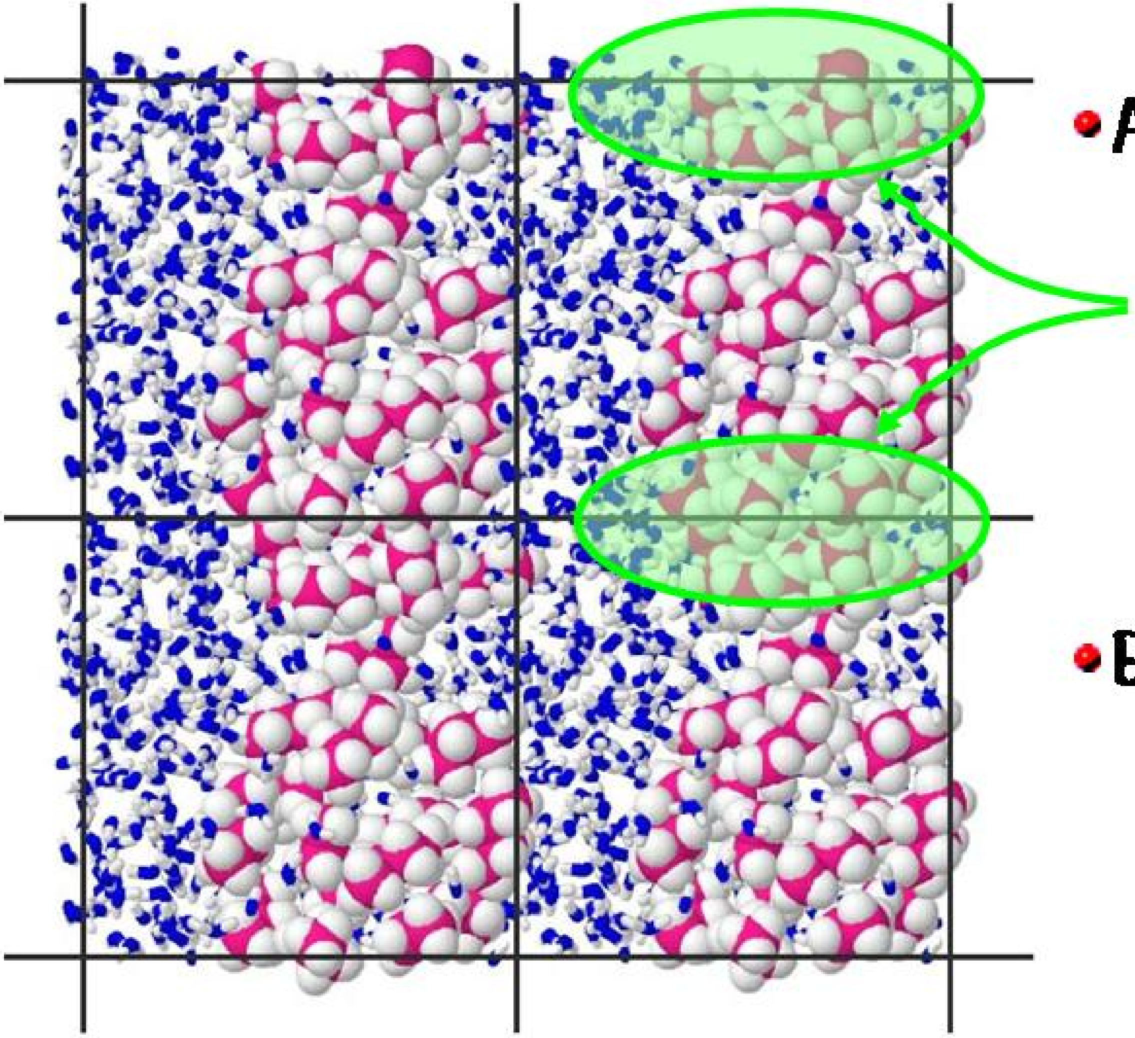
- Make a fence between A and B so that all points closer to A are on left, all points closer to B on right. Easy.
- Now add third point C.
- How about adding points D and E?.

VORONOI ANALYSIS OF CONTACTS

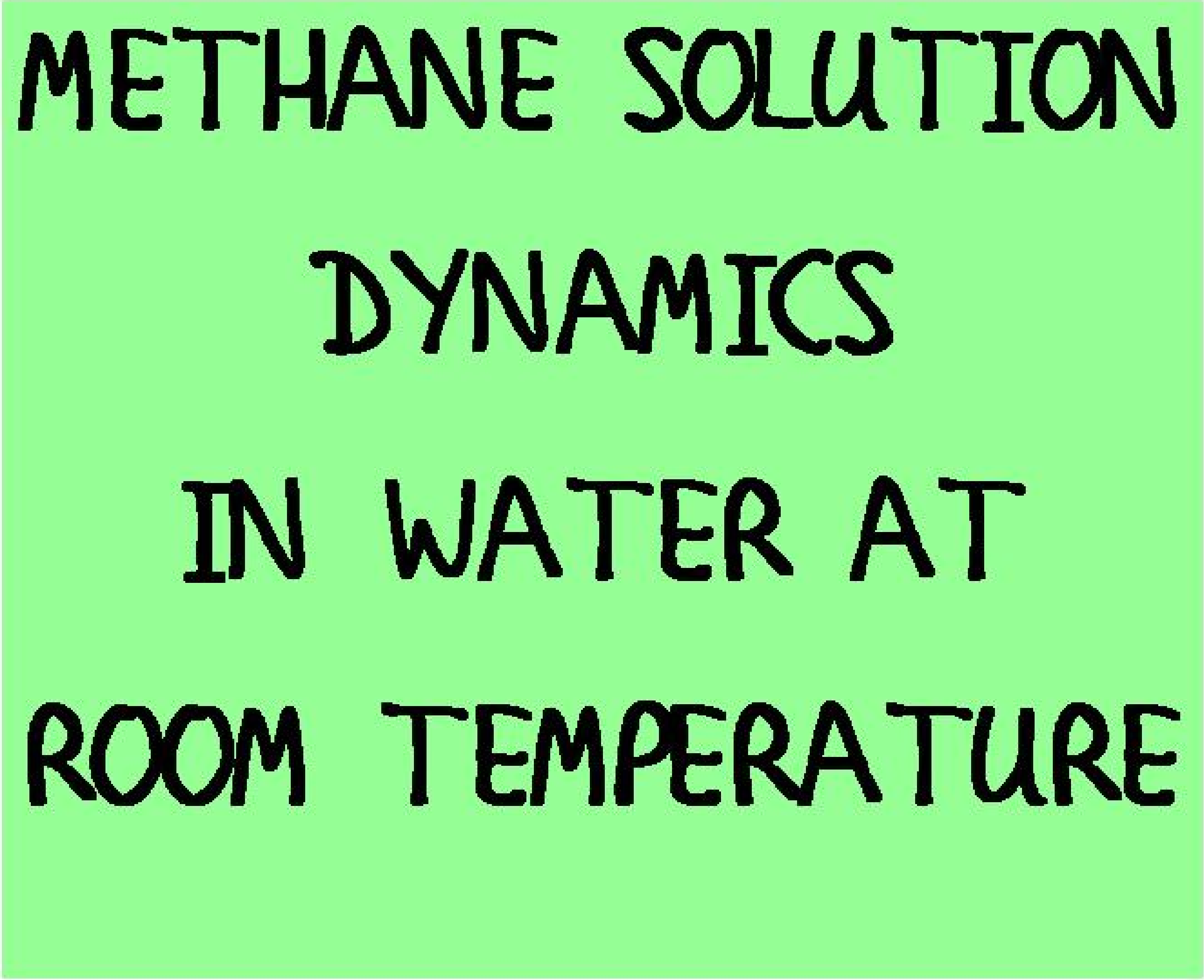


- Measure cluster formation by Voronoi construction.
- Contact measured by area of face common to A and B.
- Distance (AB) same as distance (BC), but only A and B touch though a shared face.

HYDROPHOBIC CLUSTERS CROSS BOX

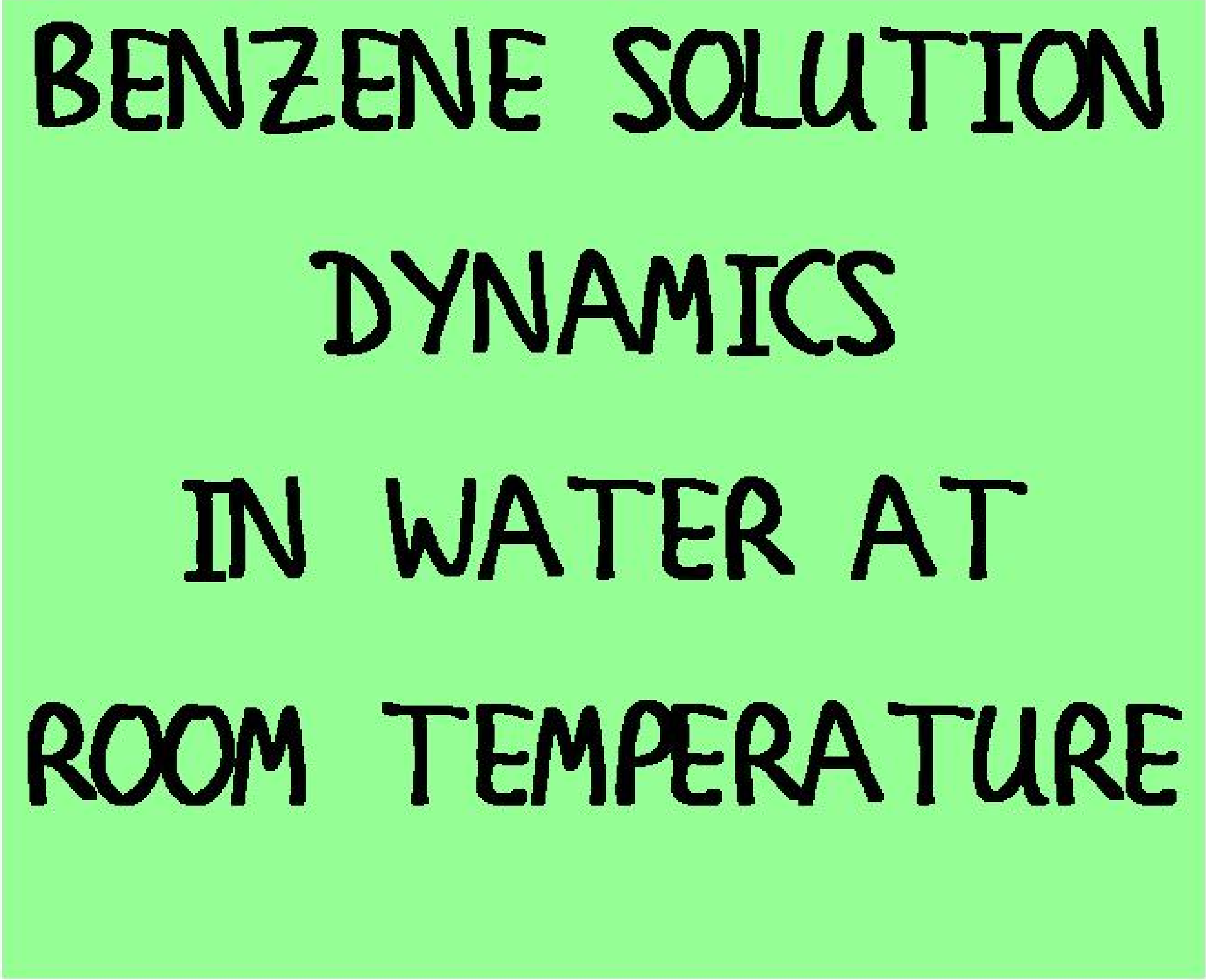


- At high concentration, the clusters extend throughout the periodic system.
- Excluded from the analysis as they are not isolated clusters.



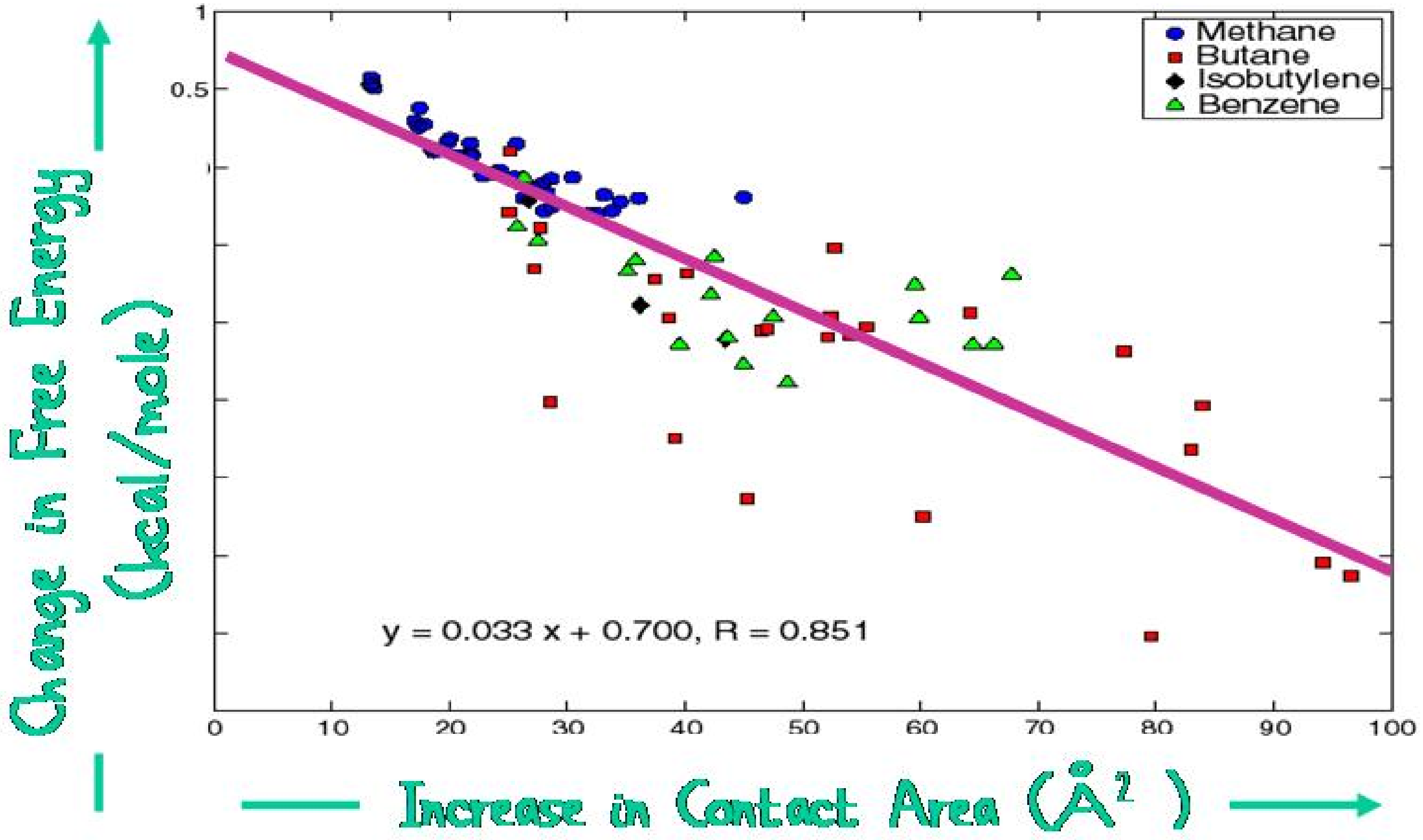
METHANE SOLUTION DYNAMICS IN WATER AT ROOM TEMPERATURE

BUTANE SOLUTION DYNAMICS IN WATER AT ROOM TEMPERATURE



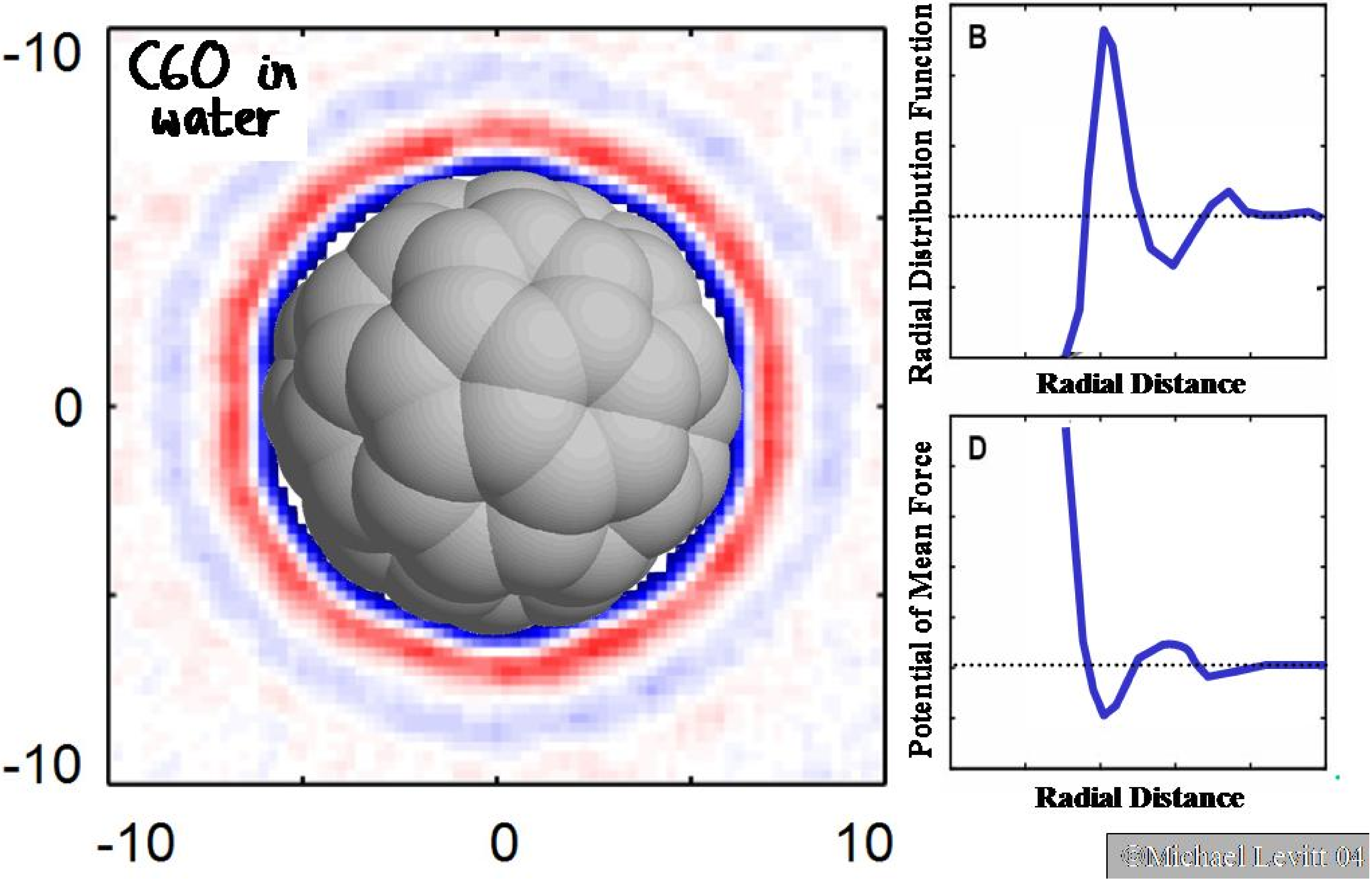
BENZENE SOLUTION
DYNAMICS
IN WATER AT
ROOM TEMPERATURE

HYDROPHOBIC ENERGY AND CONTACT SURFACE

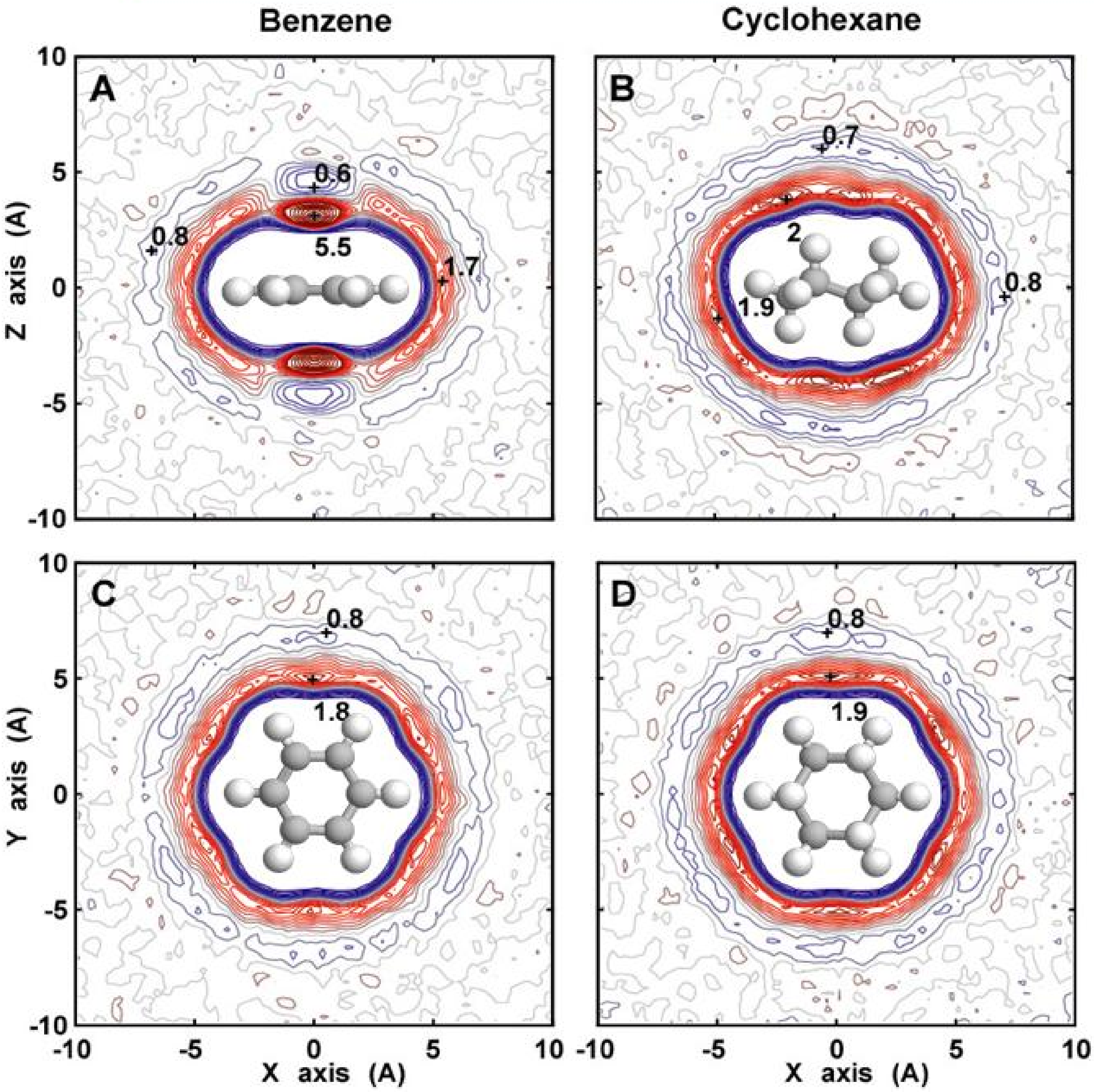


Slope of line is about 40 cal/mol per Å² which is close to the experimental value.

HYDROPHOBIC SOLUTES PERTURB WATER DENSITY



LARGE DENSITY CHANGES



Water density around hydrophobic solutes.

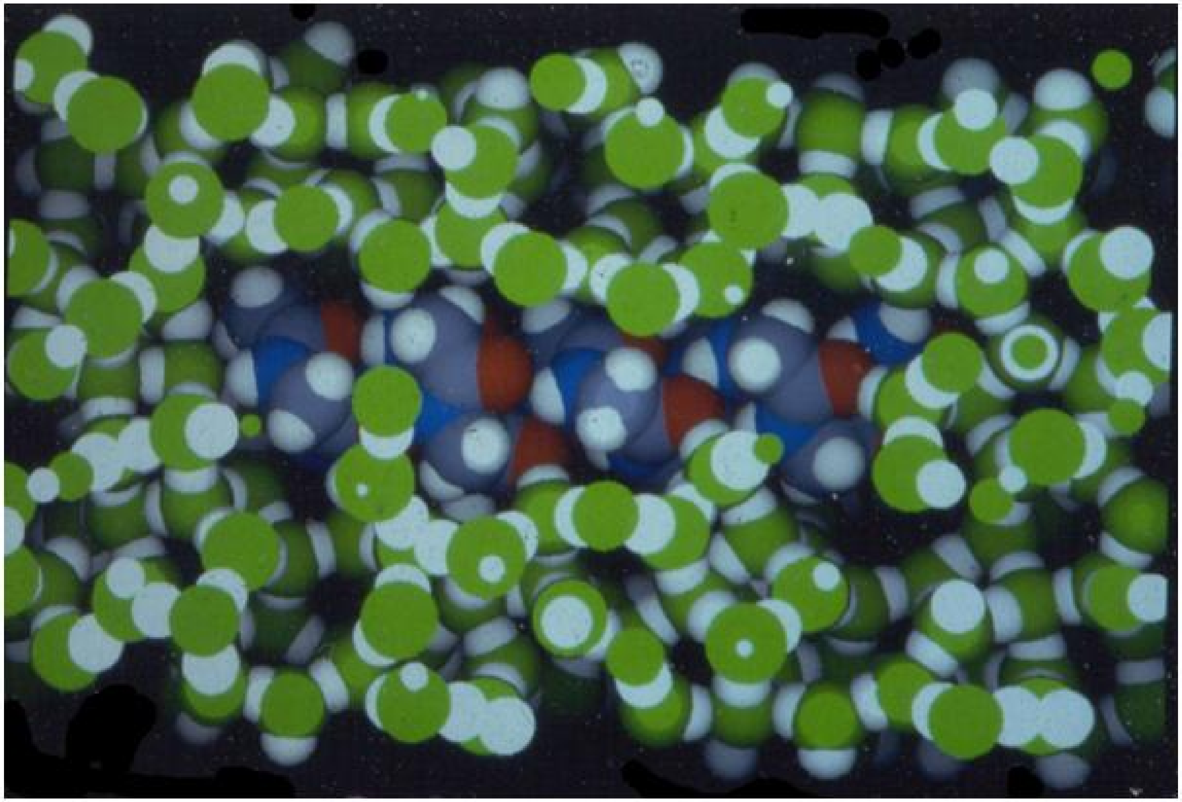
Lots of detailed structure.

Red contours are for high density.

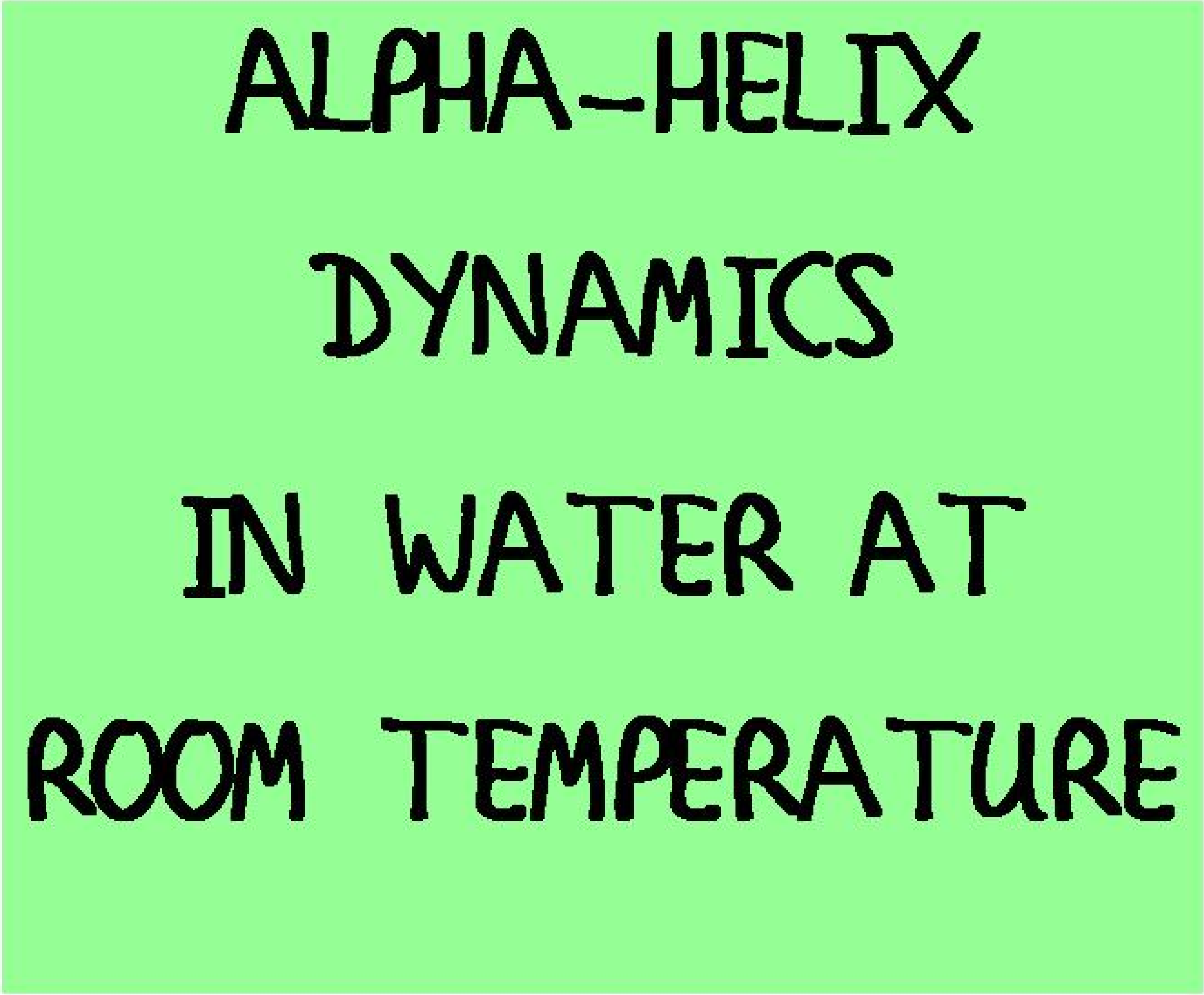
Blue contours are for low density.

Simulating Alpha-Helix
Concept 4.5

DYNAMICS OF THE ALPHA-HELIX



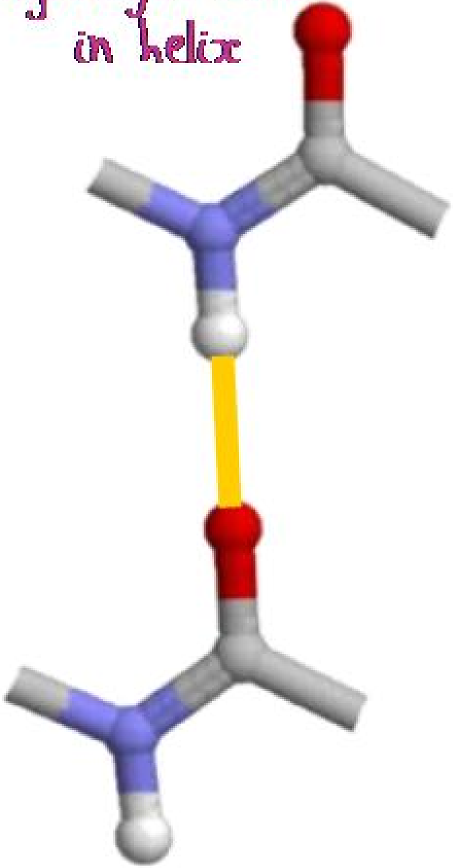
- Put α-helix in a box or water.
- Run molecular dynamics.
- Focus attention on hydrogen bonds.



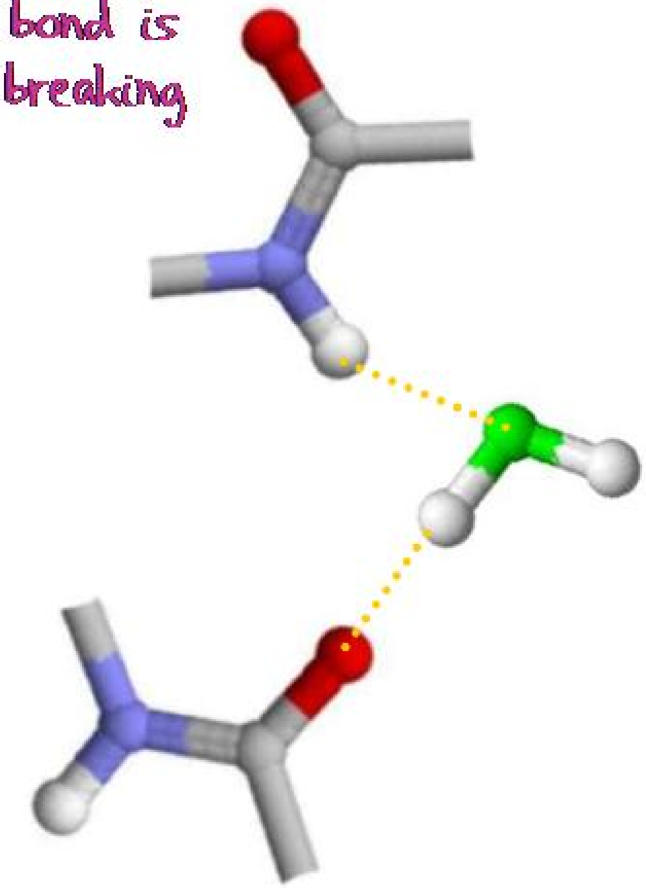
ALPHA-HELIX DYNAMICS IN WATER AT ROOM TEMPERATURE

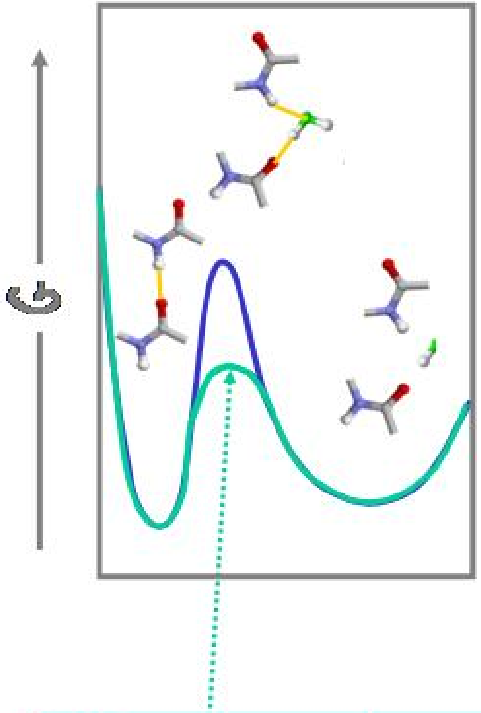
WATER ALLOWS HYDROGEN BONDS TO BREAK

Intact hydrogen bond in helix



Hydrogen bond is breaking





Free Energy barrier between states is much lower in water.

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

Simulating Folded Protein
Concept 4.6

DYNAMICS OF FOLDED PROTEIN

Movie Visualizing Motion.

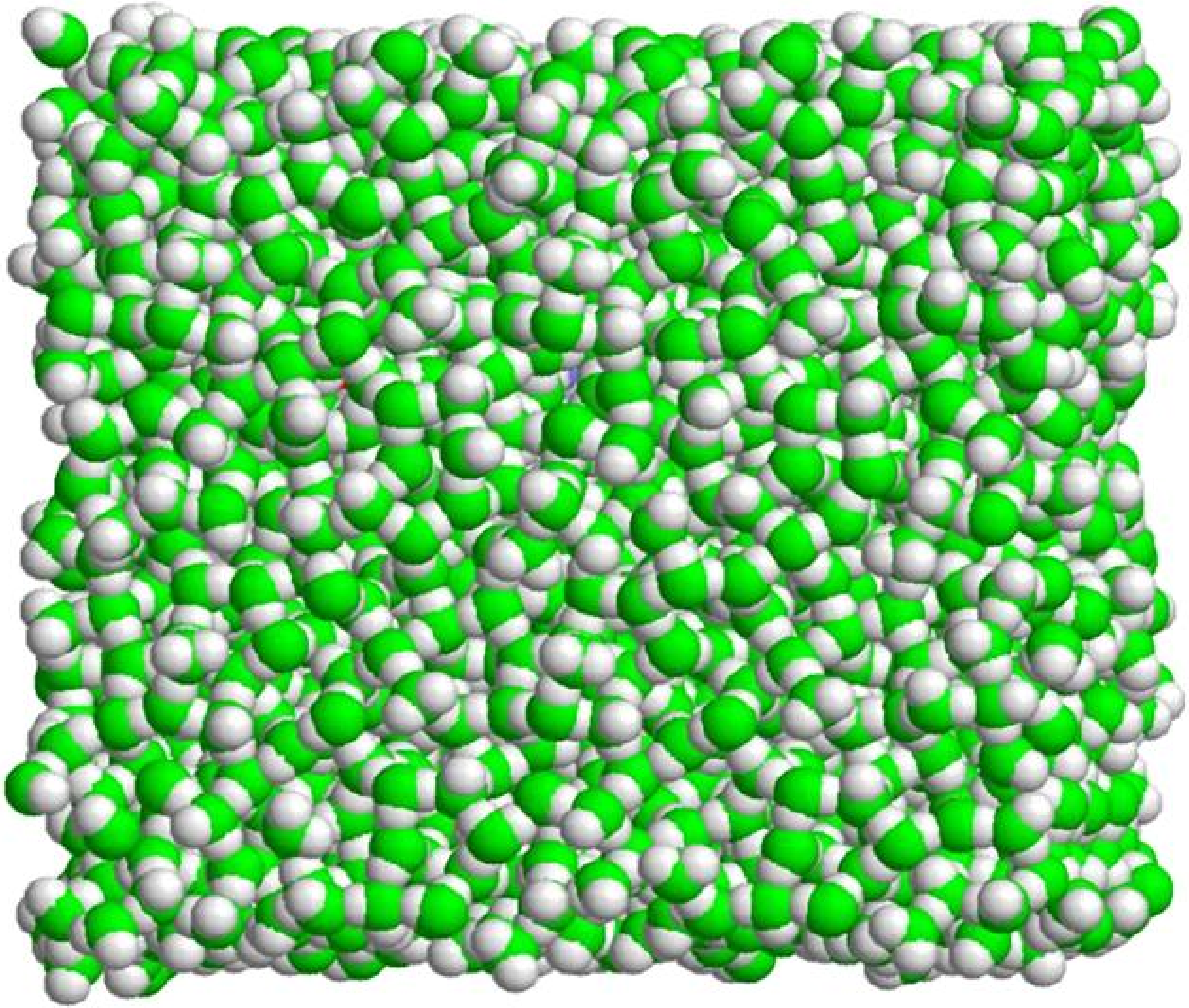
RMS Deviation from X-ray.

Hydrogen Bond Stability.

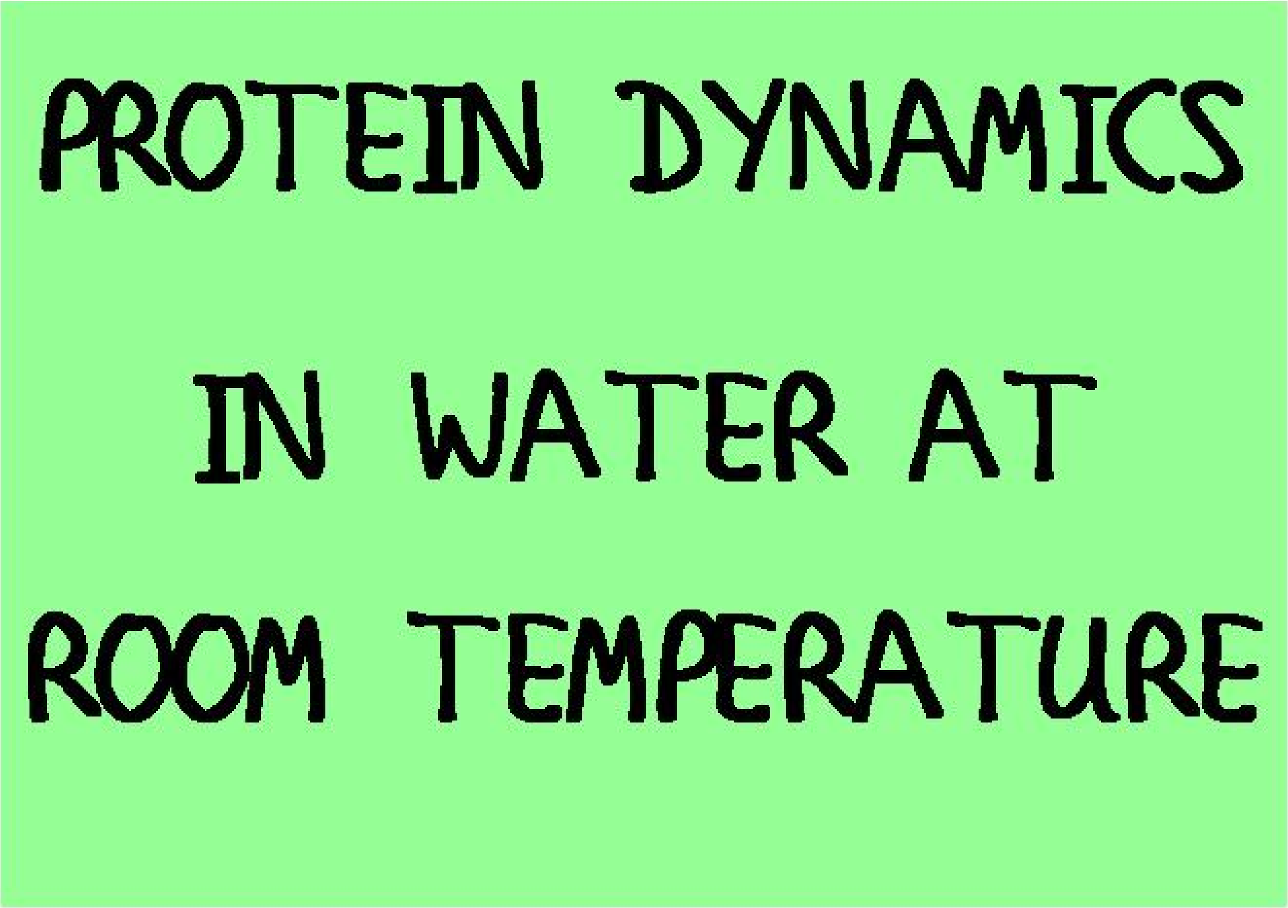
Density of Surface Water.

Orientation of Surface Water.

PROTEIN IN WATER

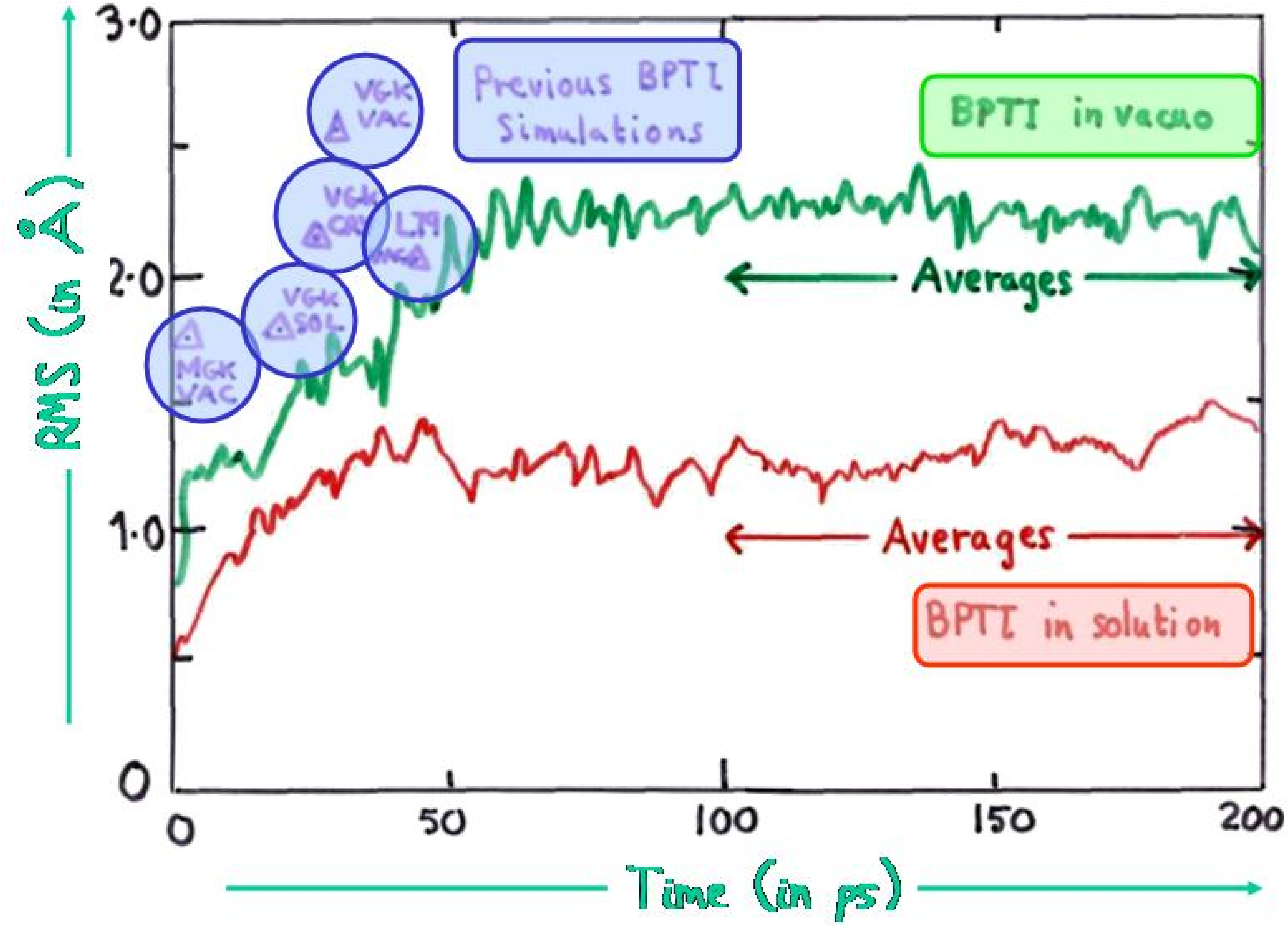


Here we have 2,607 water molecules per one protein molecule (19 Millimoles/liter). 90% of the atoms are water atoms.



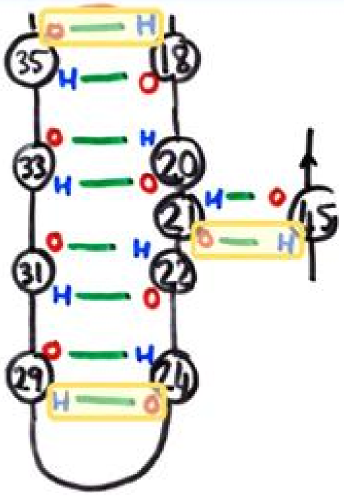
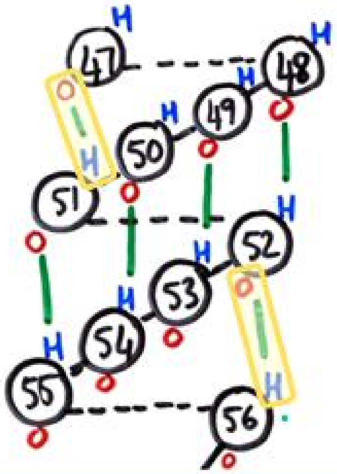
PROTEIN DYNAMICS IN WATER AT ROOM TEMPERATURE

RMS DEVIATION FROM X-RAY



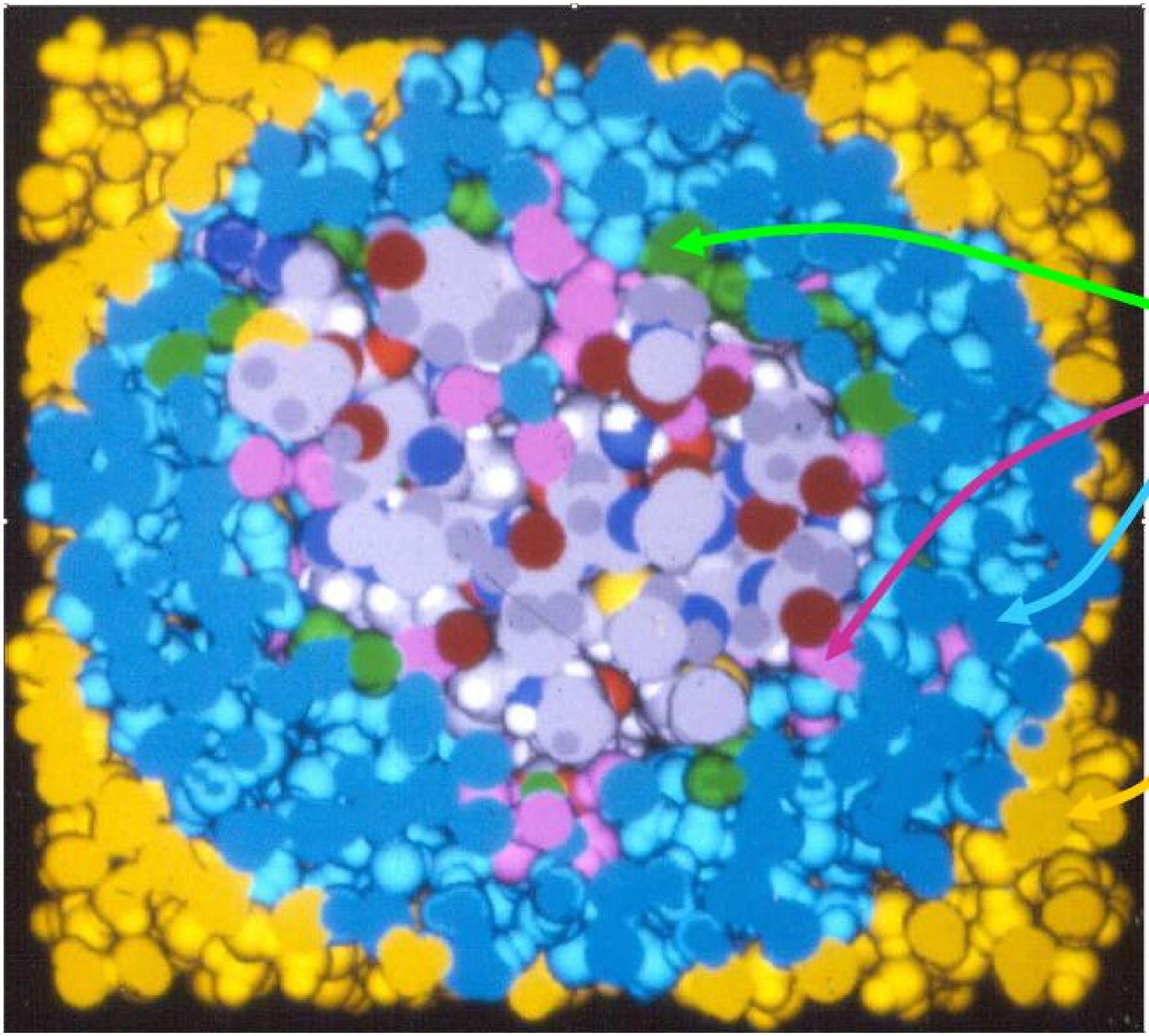
- In solution, trajectory remains much closer to the native X-ray structure than *in vacuo*.
- Much better than other runs.

HYDROGEN BOND STABILITY

Secondary Structure	O.. H Pair	Stability (%) in vacuo	Stability (%) in soln.
	35.. 18	12	57
	18.. 35	85	63
	33.. 20	71	76
	20.. 33	80	86
	31.. 22	53	93
	22.. 31	82	87
	29.. 24	72	67
	24.. 29	37	34
	45.. 21	63	86
	21.. 45	14	42
	47.. 51	76	66
	48.. 52	93	90
	49.. 53	90	98
	50.. 54	78	90
	51.. 55	73	93
	52.. 56	-	42

- Stability is percent of time bond is formed.
- Hydrogen bonds in solution are as strong as those in vacuo.
- The relative strength depends on the position of the hydrogen bond in the secondary structure

FOUR CLASSES OF WATER MOLECULES



Four classes of water molecules:

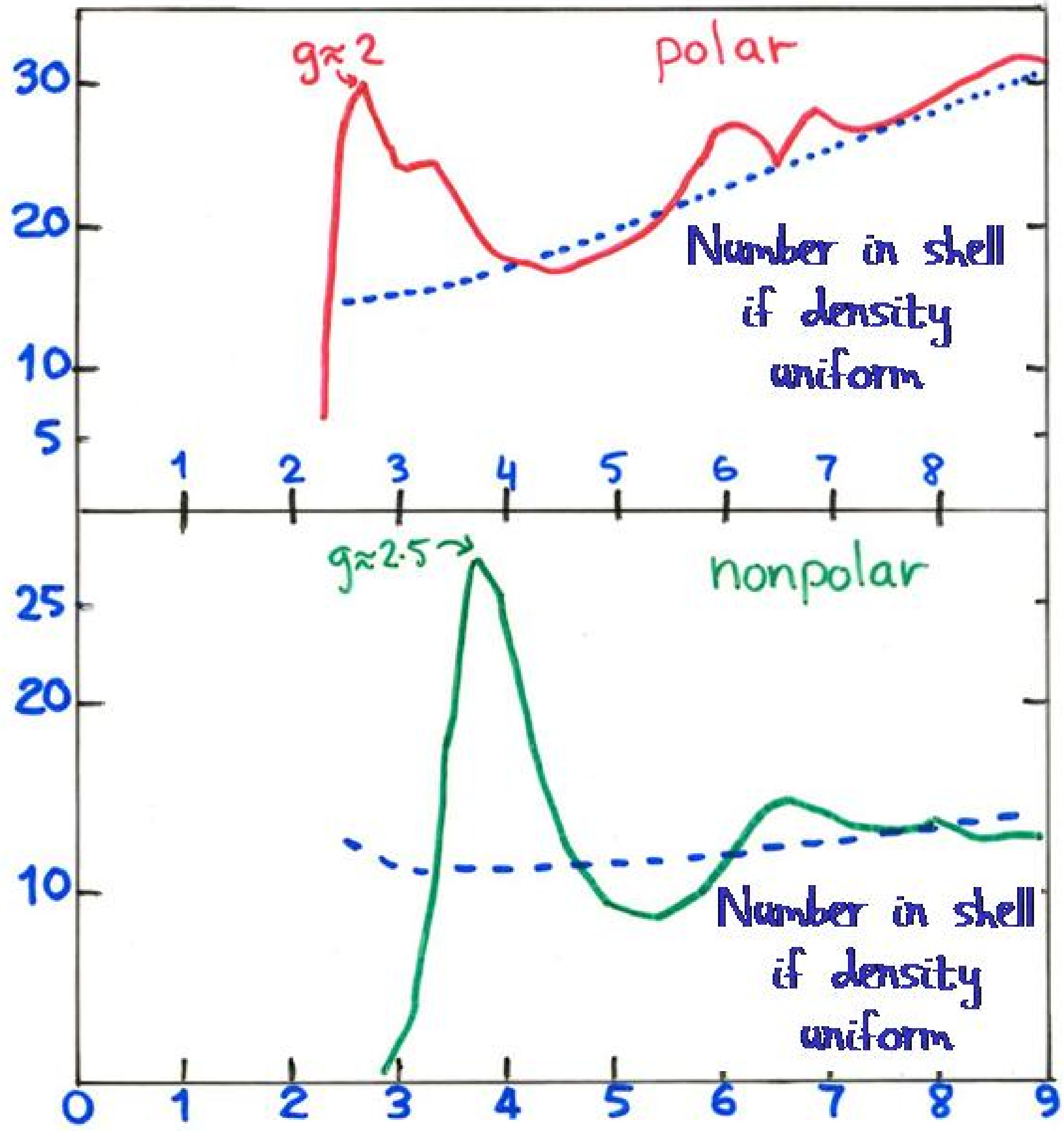
(I) Polar Surface

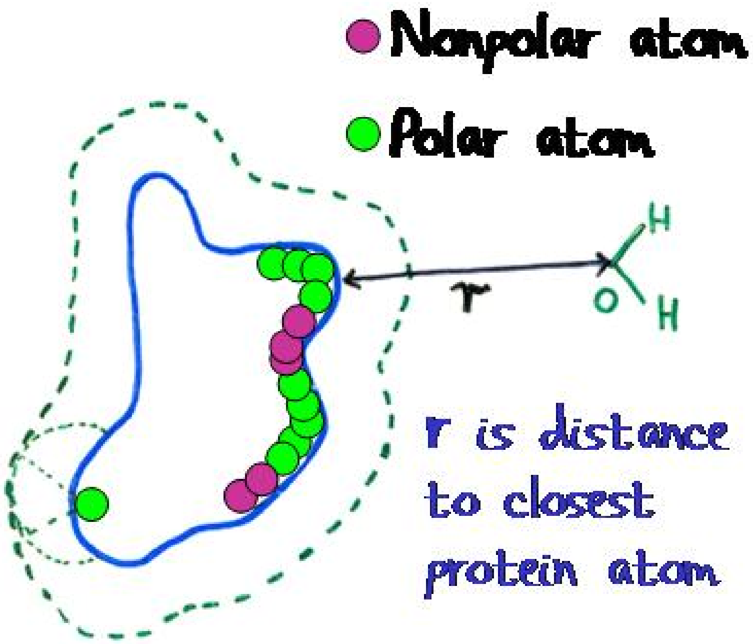
(II) Nonpolar Surface

(III) Layered

(IV) Bulk.

DENSITY OF SURFACE WATER





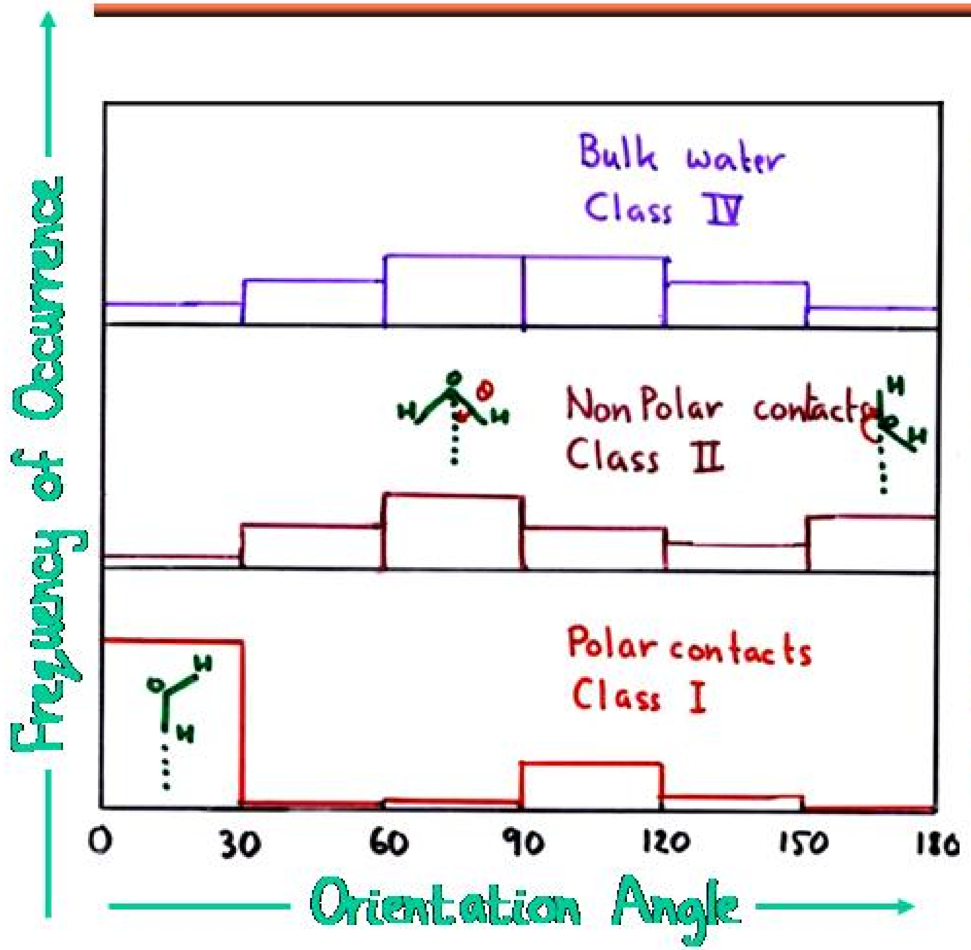
● Nonpolar atom

● Polar atom

r is distance to closest protein atom

- The water has a higher density next to the protein surface.
- This is particularly clear for the nonpolar surface.

ORIENTATION OF SURFACE WATER



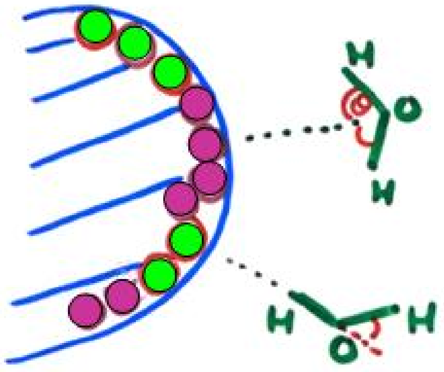
Bulk water is random orientation

Water never points hydrogen to nonpolar atom

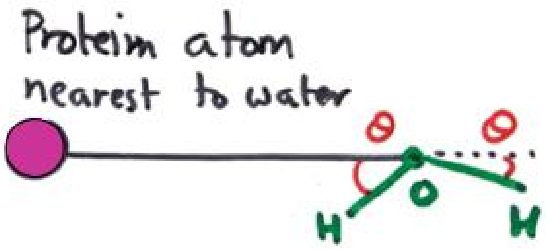
Water points hydrogen to polar atom

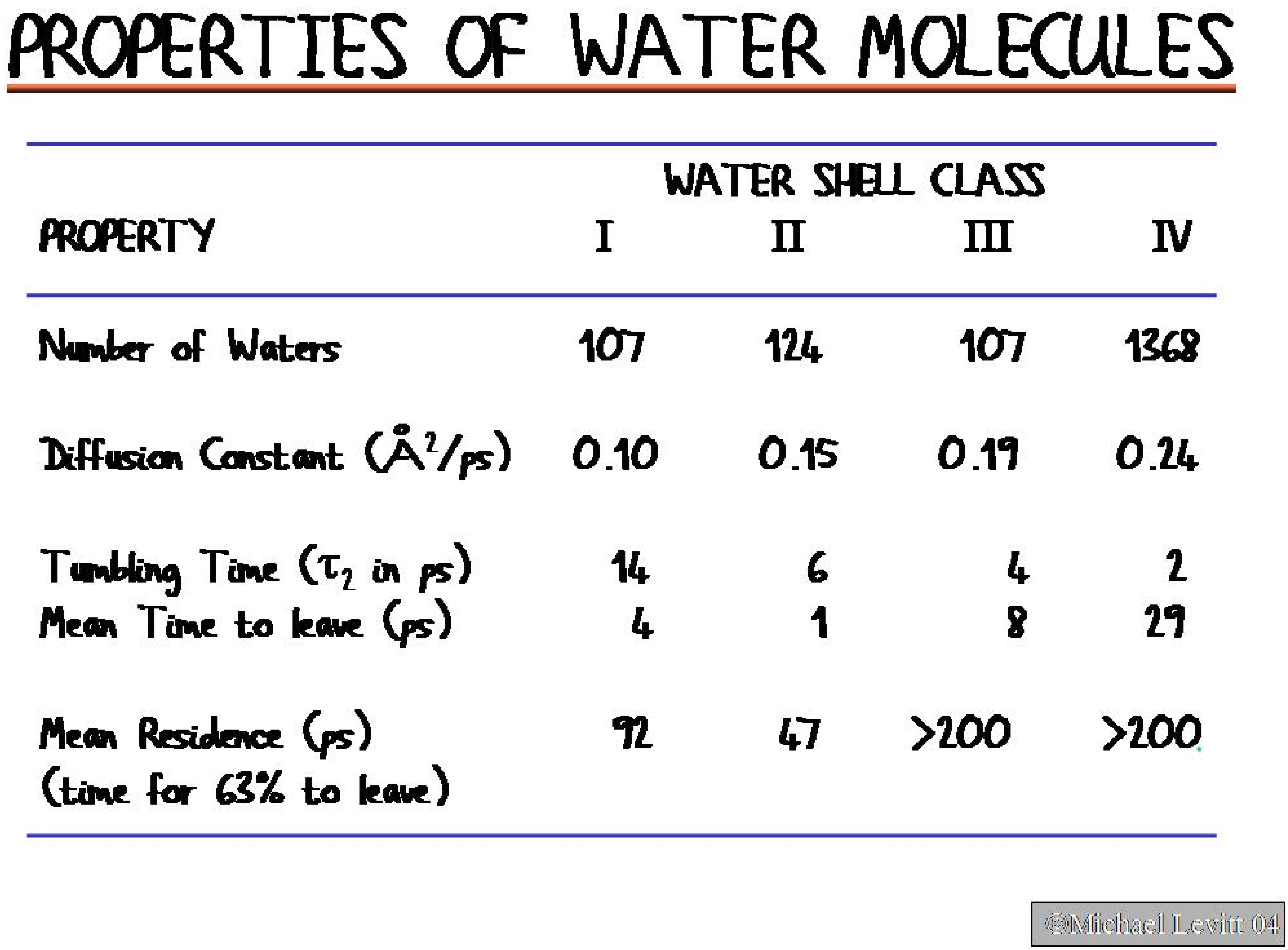
● Nonpolar atom

● Polar atom



- The water next to the polar surface is oriented to make hydrogen bonds.
- The water next to the nonpolar surface is also oriented.



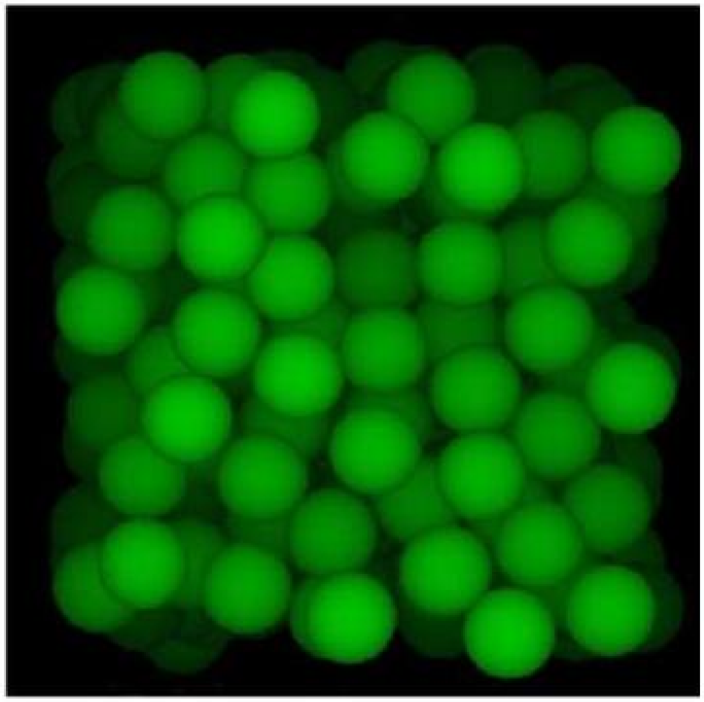


PROPERTIES OF WATER MOLECULES

	WATER SHELL CLASS			
PROPERTY	I	II	III	IV
Number of Waters	107	124	107	1368
Diffusion Constant ($Å^2/ps$)	0.10	0.15	0.19	0.24
Tumbling Time (τ_2 in ps)	14	6	4	2
Mean Time to leave (ps)	4	1	8	29
Mean Residence (ps) (time for 63% to leave)	92	47	>200	>200

History of Simulation
Concept 4.7

LOS ALAMOS 1943-45



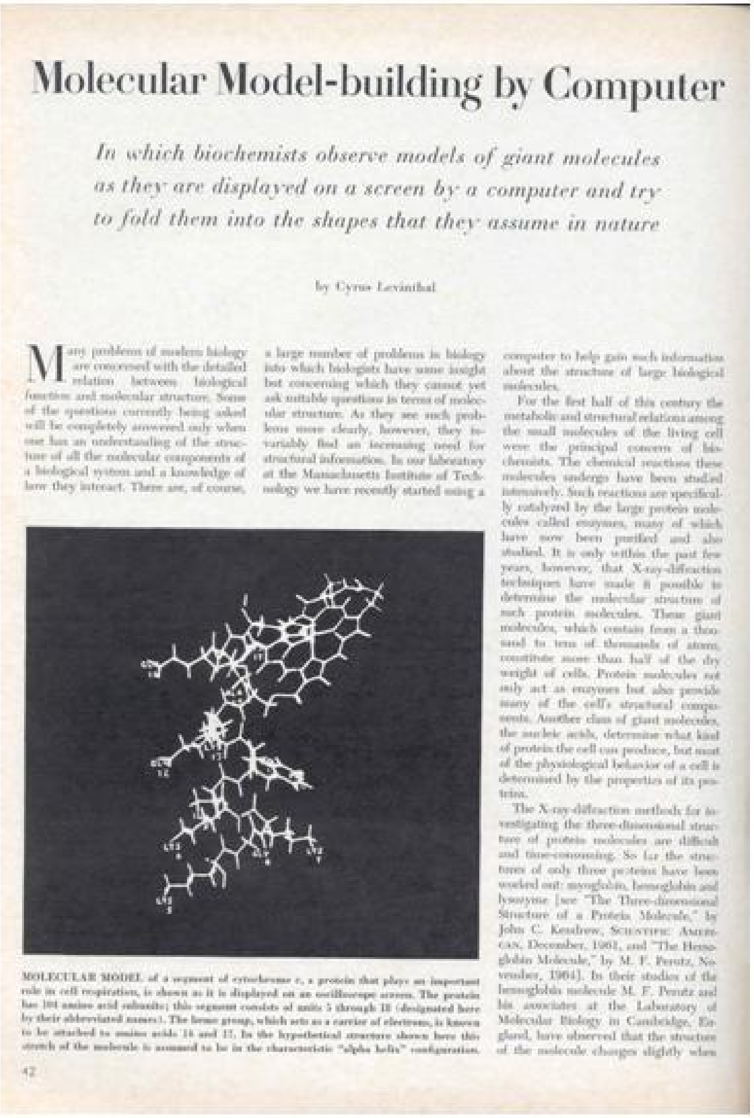
The Birth of the Monte Carlo Method.

When any sufficiently large nuclear explosion occurs within a container, unless the radioactive material is properly contained and the timing of triggering explosions perfect, neutrons stream out of one side of the container. This leak causes an asymmetrical, much weaker, and more unpredictable blast. In order to make the most potent blast possible, a series of complex events must be modeled so that the radioactive material explodes symmetrically. This research appears under the hygienic guise of solving the "neutron diffusion problem." Until 1943, when von Neumann and Stanley Ulam worked on the neutron diffusion problem, there were essentially only two sorts of modeling employed by scientists and mathematicians to describe complex events: deterministic methods (which are essentially applied mathematics) and variations on stochastic techniques (which were known simply as simulation).

To get around the apparently inevitable incorporation of the random, von Neumann devised a third kind of simulation called the "Monte Carlo" in homage to the games of luck he enjoyed in the gambling capital of Europe. He held that random elements in simulations were unacceptable, a form of contamination tantamount to cheating at cards. Indeed, his aversion to stochastic modeling and his appreciation of rule-based games is at the heart of his epistemology. In the Monte Carlo simulation, Von Neumann devised a non-stochastic formula for approximating the stochastic operators in non-trivial simulations. Essentially, he had found a deterministic way to model random events. At the same time, he had rigged the game in the house's favor. When the Monte Carlo simulation worked, it suggested not only that we could describe nature without relying on randomness or chance, but that nature itself was deterministic.

http://trace.ntu.ac.uk/frame2/articles/borg/JvN.html

LEVINTHAL 1966



Molecular Model-building by Computer

In which biochemists observe models of giant molecules as they are displayed on a screen by a computer and try to fold them into the shapes that they assume in nature

by Cyrus Levinthal

Many problems of modern biology are concerned with the detailed relation between biological function and molecular structure. Some of the questions currently being asked will be completely answered only when one has an understanding of the structure of all the molecular components of a biological system and a knowledge of how they interact. There are, of course, a large number of problems in biology into which biologists have some insight but concerning which they cannot yet ask suitable questions in terms of molecular structure. As they see such problems more clearly, however, they invariably find an increasing need for structural information. In our laboratory at the Massachusetts Institute of Technology we have recently started using a computer to help gain such information about the structure of large biological molecules.

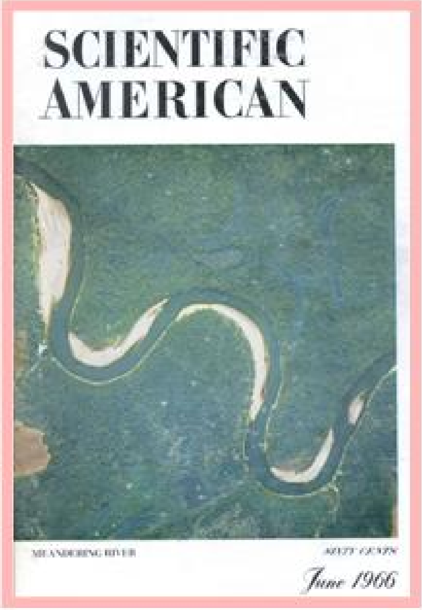
For the first half of this century the metabolic and structural relations among the small molecules of the living cell were the principal concern of biochemists. The chemical reactions these molecules undergo have been studied intensively. Such reactions are specifically catalyzed by the large protein molecules called enzymes, many of which have now been purified and also studied. It is only within the past few years, however, that X-ray-diffraction techniques have made it possible to determine the molecular structure of such protein molecules. These giant molecules, which contain from a thousand to tens of thousands of atoms, constitute more than half of the dry weight of cells. Protein molecules not only act as enzymes but also provide many of the cell's structural components. Another class of giant molecules, the nucleic acids, determine what kind of protein the cell can produce, but most of the physiological behavior of a cell is determined by the properties of its proteins.

The X-ray-diffraction methods for investigating the three-dimensional structure of protein molecules are difficult and time-consuming. So far the structures of only three proteins have been worked out: myoglobin, hemoglobin and lysozyme [see "The Three-dimensional Structure of a Protein Molecule," by John C. Kendrew, SCIENTIFIC AMERICAN, December, 1961, and "The Hemoglobin Molecule," by M. F. Perutz, November, 1964]. In their studies of the hemoglobin molecule M. F. Perutz and his associates at the Laboratory of Molecular Biology in Cambridge, England, have observed that the structure of the molecule changes slightly when

MOLECULAR MODEL of a segment of cytochrome c, a protein that plays an important role in cell respiration, is shown as it is displayed on an oscilloscope screen. The protein has 104 amino acid subunits; this segment consists of units 5 through 18 (designated here by their abbreviated names). The heme group, which acts as a carrier of electrons, is known to be attached to amino acids 14 and 17. In the hypothetical structure shown here this stretch of the molecule is assumed to be in the characteristic "alpha helix" configuration.

42

Earliest molecular graphics.
Cyrus Levinthal.
Scientific American 1966.

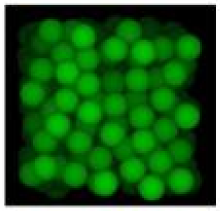


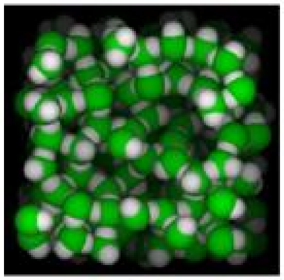
LEVINTHAL MOVIES

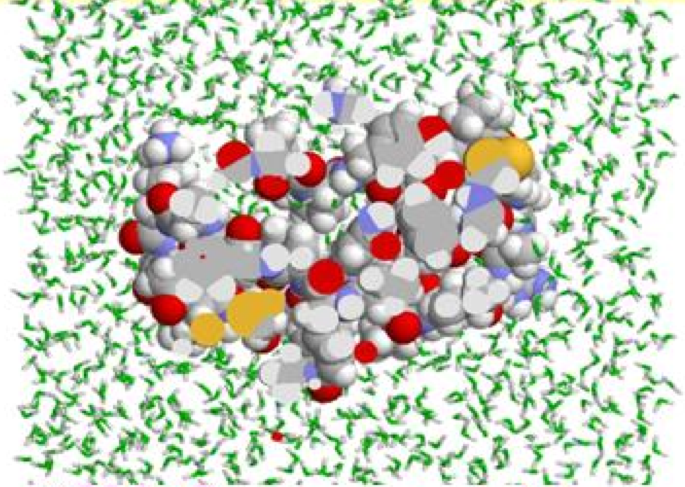
CHAIN, ALPHA-HELIX, MYOGLOBIN, LYSOZYME

50 YEARS OF SIMULATION

- We have 10,000,000 times more resources.
- Systems have become larger (100 times).
- Runs have become longer (100,000 times).
- Energy functions have become simpler.
- Fit reality well. Nothing bad has happened!


1955 Argon


1970 Water


1988 Protein in Water

?

2005.