I thought I needed to work in a Lab to do Experiments?

Ali A. Hassanali

Email: ahassana@ictp.it

The Abdus Salam International Centre for Theoretical Physics

In-Silico Experiments: The Past, Present and the Future

Ali A. Hassanali

Email: ahassana@ictp.it

The Abdus Salam International Centre for Theoretical Physics

Outline of Computational Lectures

Brief Introduction to Bio-molecular Modeling
(some history, where we are now and the future of modeling: Ali)

MD Tutorial: Simulation of a Biomolecule (Ali & Daniel)

Data Science/Data Analysis: Making Sense of Biological Data:
(Alessandro Laio, Alex Rodriguez and Pilar Cossio)

Warm-Up Questions

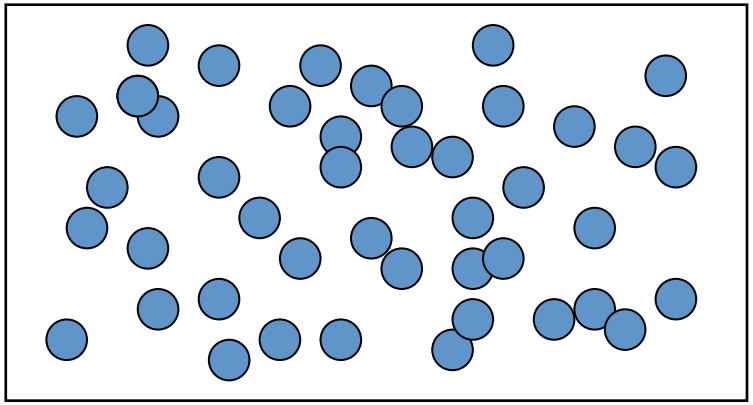
What is your favorite (bio) molecule?

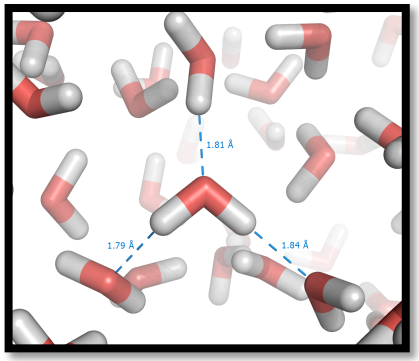
How does your favorite molecule look like?

At room temperature?

What type of motions will the molecule have?

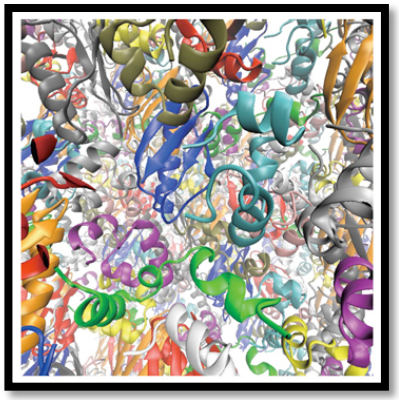
Molecular Dynamics in a NutShell





Water

(x, v)



Protein

Trajectory of Particles: Positions and Momenta

Taking the Experiment to Cyberspace

The **Nobel Prize in Chemistry 2013** was awarded jointly to **Martin Karplus**, **Michael Levitt** and **Arieh Warshel** *"for the development of multiscale models for complex chemical systems".*

Some Land Mark Papers

B. J. Alder and T. E. Wainwright (1957). "Phase Transition for a Hard Sphere System". *J. Chem. Phys.* **27** (5): 1208.

A. Rahman (1964). "Correlations in the Motion of Atoms in Liquid Argon". *Physical Review* **136**: A405-A411.

Molecular Dynamics: The Past, Present and Future

Martin Karplus
Department of Chemistry and Chemical Biology, Cambridge, MA 02138, USA and Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg, France

Received 30 May 2002; accepted 30 July 2002

Molecular Dynamics of Biological Macromolecules: A Brief History and Perspective

Abstract: *A description of the origin of my interest in and the development of molecular dynamics simulations of biomolecules is presented with a historical overview, including the role of my interactions with Shneior Lifson and his group in Israel. Some early applications of the methodology by members of my group are summarized, followed by a description of examples of recent applications and some discussion of possible future directions.* © 2003 Wiley Periodicals, Inc. Biopolymers 68: 350–358, 2003

Keywords: *NMR; molecular dynamics simulations; trajectory calculations; proteins; conformational change; channels*

Bringing Together Biomolecular Simulation and Experimental Studies

A Biochemical Society Focused Meeting in conjunction with the Molecular Graphics and Modelling Society held at Manchester Interdisciplinary Biocentre, Manchester, U.K., 10–11 September 2007. Organized and Edited by Mike Sutcliffe (Manchester, U.K.).

Biomolecular simulation: historical picture and future perspectives

Wilfred F. van Gunsteren*[1] and Jožica Dolenc*†
*Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH-Hönggerberg, CH-8093 Zurich, Switzerland, and †Faculty of Chemistry and Chemical Technology, University of Ljubljana, SI-1000 Ljubljana, Slovenia

Abstract

Over the last 30 years, computation based on molecular models is playing an increasingly important role in biology, biological chemistry and biophysics. Since only a very limited number of properties of biomolecular systems are actually accessible to measurement by experimental means, computer simulation complements experiments by providing not only averages, but also distributions and time series of any definable, observable or non-observable, quantity. Biomolecular simulation may be used (i) to interpret experimental data, (ii) to provoke new experiments, (iii) to replace experiments and (iv) to protect intellectual property. Progress over the last 30 years is sketched and perspectives are outlined for the future.

Biomolecules Actually Dance: 1977

The First Bio-molecular MD Simulations

Nature Vol. 267 16 June 1977 585

articles

Dynamics of folded proteins

J. Andrew McCammon, Bruce R. Gelin & Martin Karplus

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

It is now 25 years since the first molecular dynamics simulation of a macromolecule of biological interest was published.[1] The simulation concerned the bovine pancreatic trypsin inhibitor (BPTI), which has served as the "hydrogen molecule" of protein dynamics because of its small size, high stability, and relatively accurate X-ray structure, available in 1975;[2] interestingly, its physiological functions remain unknown. Although this simulation was done in vacuum with a crude molecular mechanics potential and lasted for only 9.2 ps, the results were instrumental in replacing our view of proteins as relatively rigid structures (In 1981, Sir D. L. Phillips commented: "Brass models of DNA and a variety of proteins dominated the scene and much of the thinking".[3]) with the realization that they were dynamic systems, whose internal motions play a functional role. Of course, there were already experimental data (such as the pioneering hydrogen exchange studies of Linderstrom-Lang and his co-worker),[4,5] pointing in this direction. It is now recognized that the X-ray structure of a protein provides the average atomic positions, but the atoms exhibit fluid-like motions

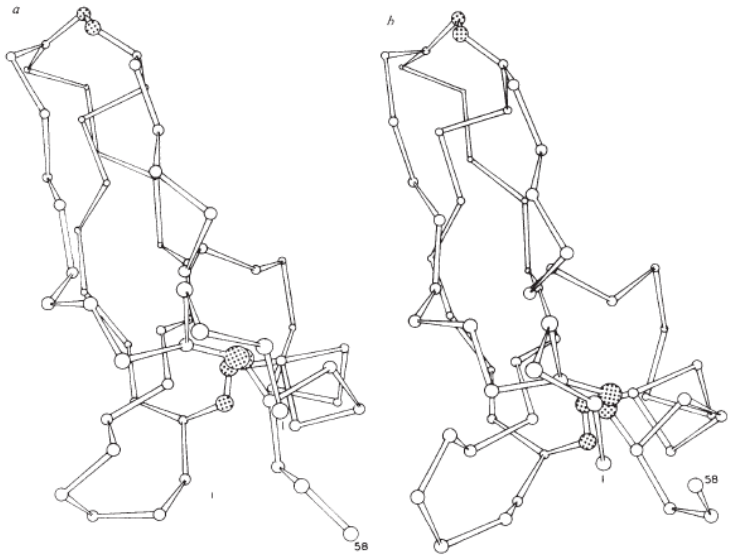


Fig. 1 The peptide backbone (α carbons) and disulphide bonds of PTI. *a*, X-ray structure[21]. *b*, Time evolved structure after 3.2 ps of dynamical simulation.

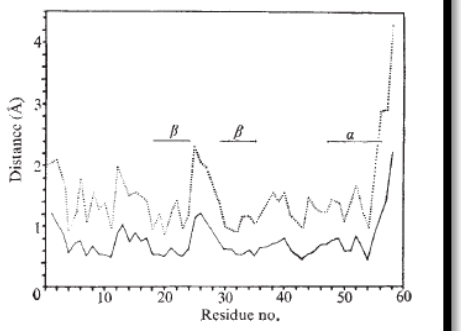


Fig. 2 R.m.s. displacements (———) and maximum displacements (. . . . .) of the α carbons, relative to the dynamical average structure.

Wait A Second … Where is the Water?

Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7557–7561, October 1988
Biophysics

Accurate simulation of protein dynamics in solution

(deviation from x-ray crystal structure/fluctuation amplitudes/hydrogen-bond stability)

MICHAEL LEVITT* AND RUTH SHARON

Department of Chemical Physics, Weizmann Institute of Science, Rehovot 76100, Israel

Communicated by A. Klug, June 23, 1988

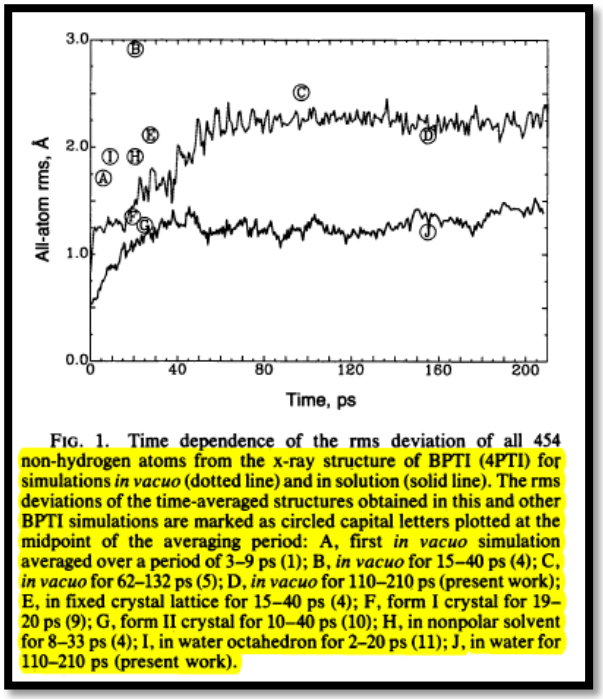
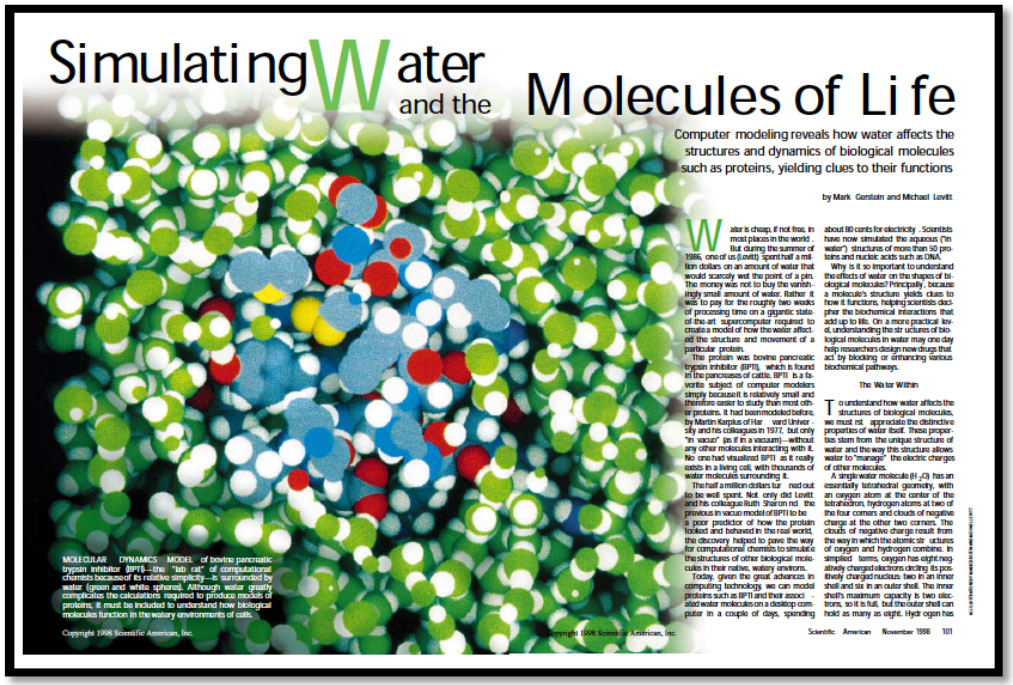


FIG. 1. Time dependence of the rms deviation of all 454 non-hydrogen atoms from the x-ray structure of BPTI (4PTI) for simulations *in vacuo* (dotted line) and in solution (solid line). The rms deviations of the time-averaged structures obtained in this and other BPTI simulations are marked as circled capital letters plotted at the midpoint of the averaging period: A, first *in vacuo* simulation averaged over a period of 3–9 ps (1); B, *in vacuo* for 15–40 ps (4); C, *in vacuo* for 62–132 ps (5); D, *in vacuo* for 110–210 ps (present work); E, in fixed crystal lattice for 15–40 ps (4); F, form I crystal for 19–20 ps (9); G, form II crystal for 10–40 ps (10); H, in nonpolar solvent for 8–33 ps (4); I, in water octahedron for 2–20 ps (11); J, in water for 110–210 ps (present work).



Simulating Water and the Molecules of Life

Computer modeling reveals how water affects the structures and dynamics of biological molecules such as proteins, yielding clues to their functions

by Mark Gerstein and Michael Levitt

MOLECULAR DYNAMICS MODEL of bovine pancreatic trypsin inhibitor (BPTI)—the "lab rat" of computational chemists because of its relative simplicity—is surrounded by water (green and white spheres). Although water greatly complicates the calculations required to produce models of proteins, it must be included to understand how biological molecules function in the watery environments of cells.

Copyright 1998 Scientific American, Inc.

Water is cheap, if not free, in most places in the world. But during the summer of 1986, one of us (Levitt) spent half a million dollars on an amount of water that would scarcely wet the point of a pin. The money was not to buy the vanishingly small amount of water. Rather it was to pay for the roughly two weeks of processing time on a gigantic state-of-the-art supercomputer required to create a model of how the water affected the structure and movement of a particular protein.

The protein was bovine pancreatic trypsin inhibitor (BPTI), which is found in the pancreases of cattle. BPTI is a favorite subject of computer modelers simply because it is relatively small and therefore easier to study than most other proteins. It had been modeled before, by Martin Karplus of Harvard University and his colleagues in 1977, but only "in vacuo" (as if in a vacuum)—without any other molecules interacting with it. No one had visualized BPTI as it really exists in a living cell, with thousands of water molecules surrounding it.

The half a million dollars turned out to be well spent. Not only did Levitt and his colleague Ruth Sharon red the previous in vacuo model of BPTI to be a poor predictor of how the protein looked and behaved in the real world, the discovery helped to pave the way for computational chemists to simulate the structures of other biological molecules in their native, watery environs.

Today, given the great advances in computing technology, we can model proteins such as BPTI and their associated water molecules on a desktop computer in a couple of days, spending about 80 cents for electricity. Scientists have now simulated the aqueous ("in water") structures of more than 50 proteins and nucleic acids such as DNA.

Why is it so important to understand the effects of water on the shapes of biological molecules? Principally, because a molecule's structure yields clues to how it functions, helping scientists decipher the biochemical interactions that add up to life. On a more practical level, understanding the structures of biological molecules in water may one day help researchers design new drugs that act by blocking or enhancing various biochemical pathways.

The Water Within

To understand how water affects the structures of biological molecules, we must first appreciate the distinctive properties of water itself. These properties stem from the unique structure of water and the way this structure allows water to "manage" the electric charges of other molecules.

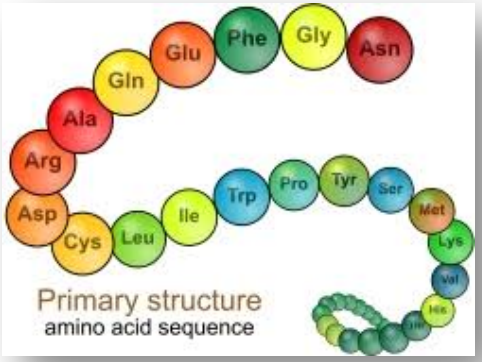
A single water molecule (H_2O) has an essentially tetrahedral geometry, with an oxygen atom at the center of the tetrahedron, hydrogen atoms at two of the four corners and clouds of negative charge at the other two corners. The clouds of negative charge result from the way in which the atomic structures of oxygen and hydrogen combine. In simplified terms, oxygen has eight negatively charged electrons circling its positively charged nucleus: two in an inner shell and six in an outer shell. The inner shell's maximum capacity is two electrons, so it is full, but the outer shell can hold as many as eight. Hydrogen has

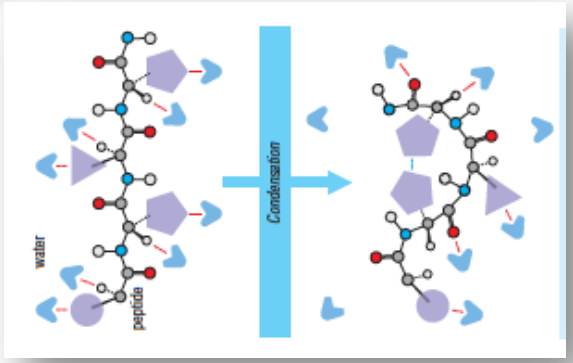
Scientific American November 1998 101

10-100 Picoseconds

Coming a Long Way from 100's of Picoseconds!

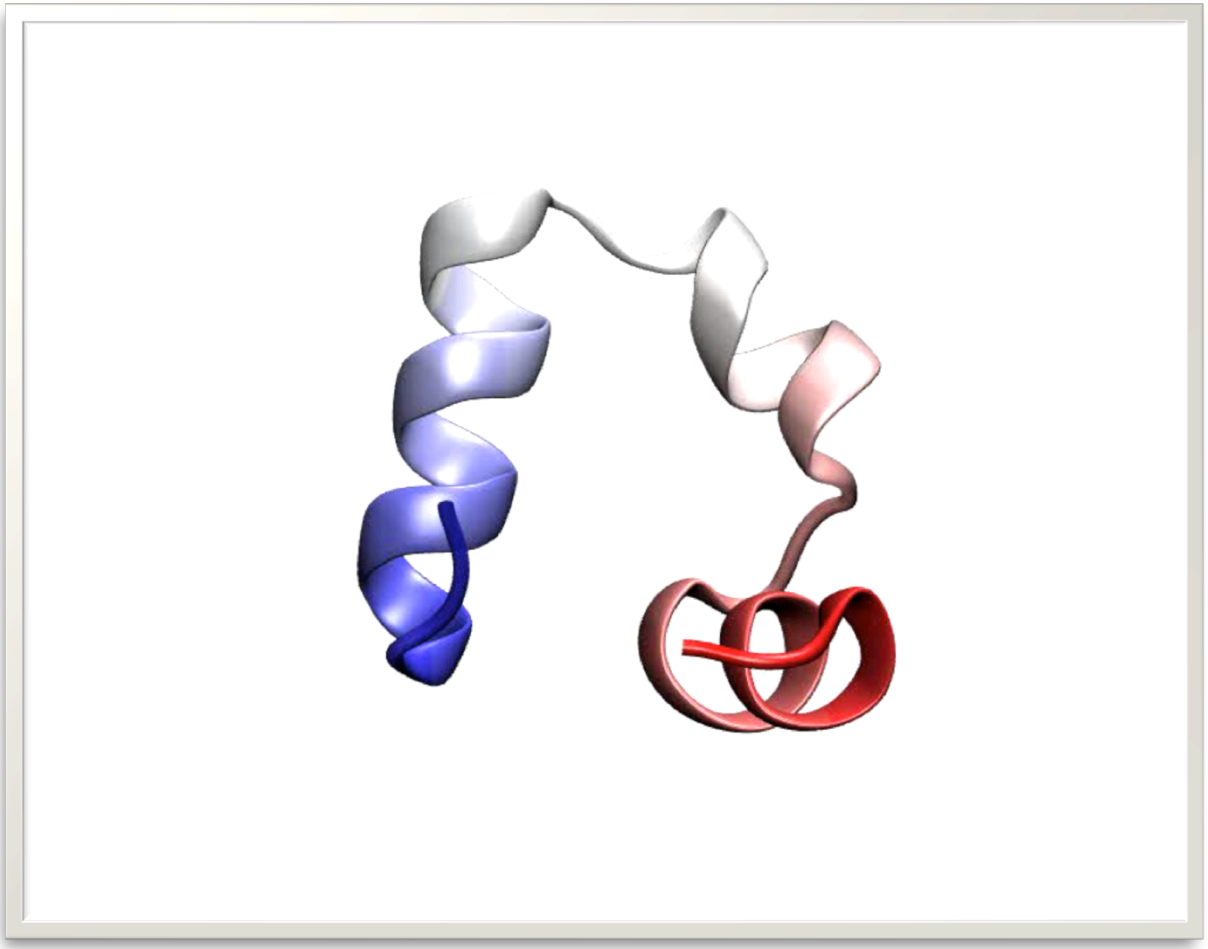
How does a protein fold?





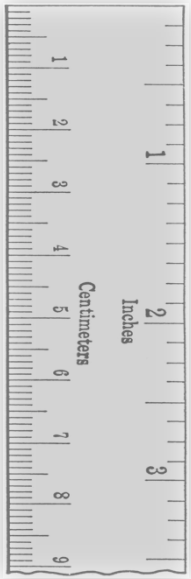
What pathways does the protein take?

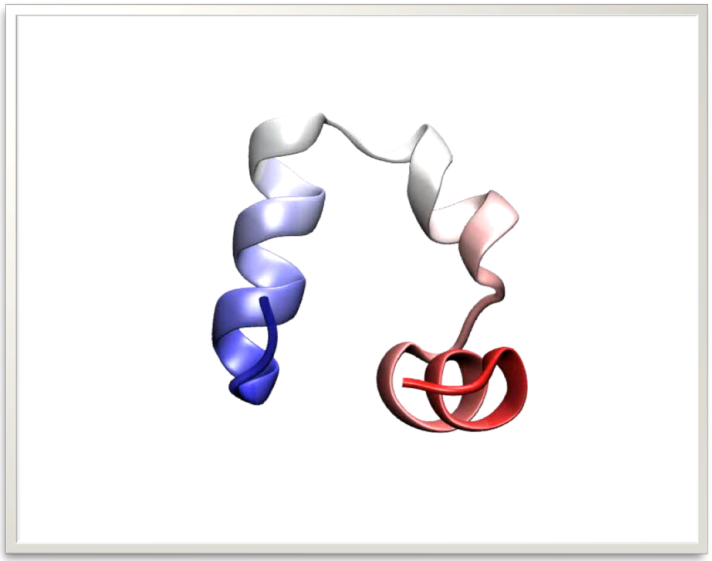
Watching a Protein Fold



μs

What Are The Length/Time/Energy Scales?



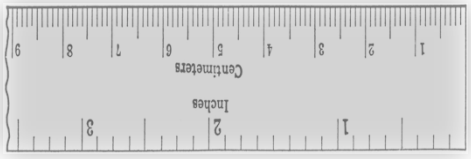


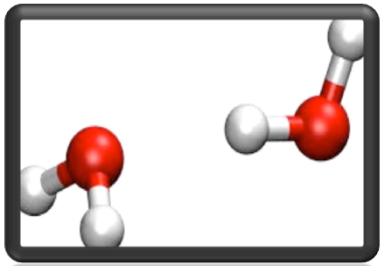
How large/heavy is the protein?

On what time-scales does the protein move?

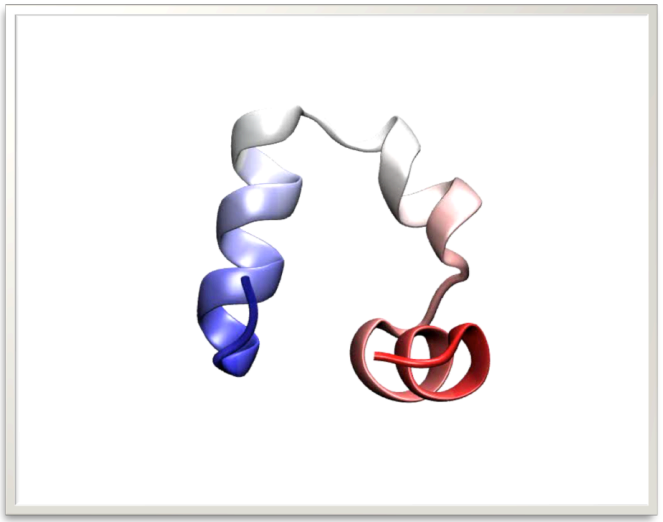
What energies are relevant to the motions?

Getting a Sense of Length-Scales





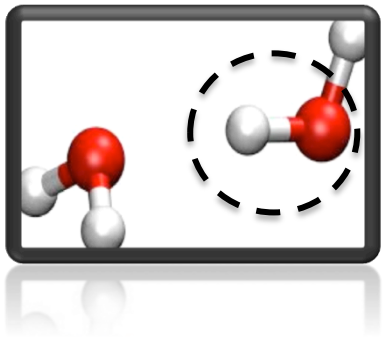
$\sim 3\,\mathring{A}$



?

?

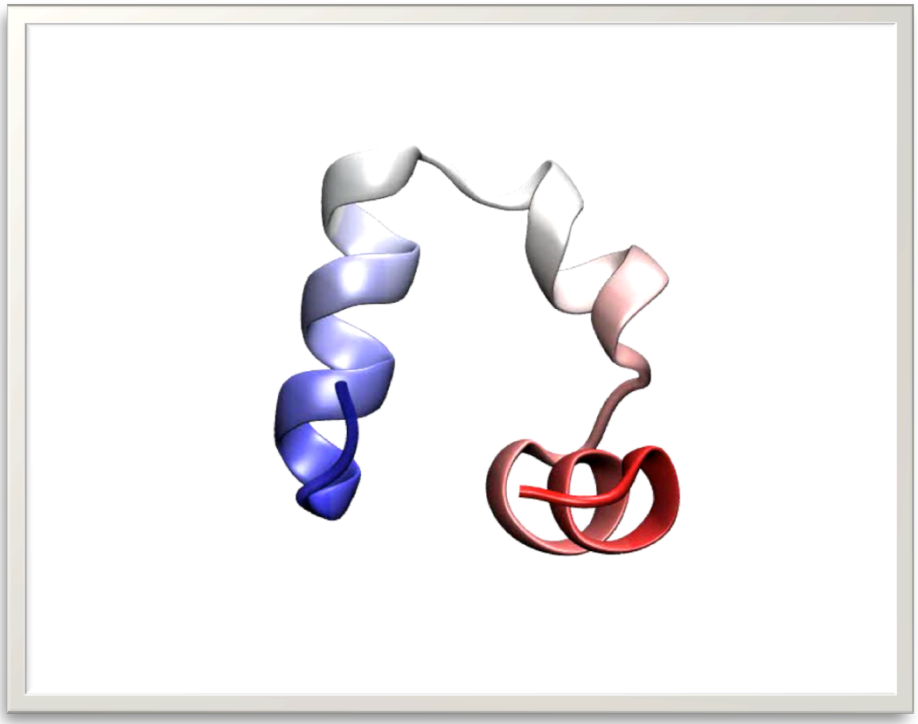
Getting a Sense of Time-Scales



How fast are bond vibrations?

How fast do hydrogen bonds break and form in water?

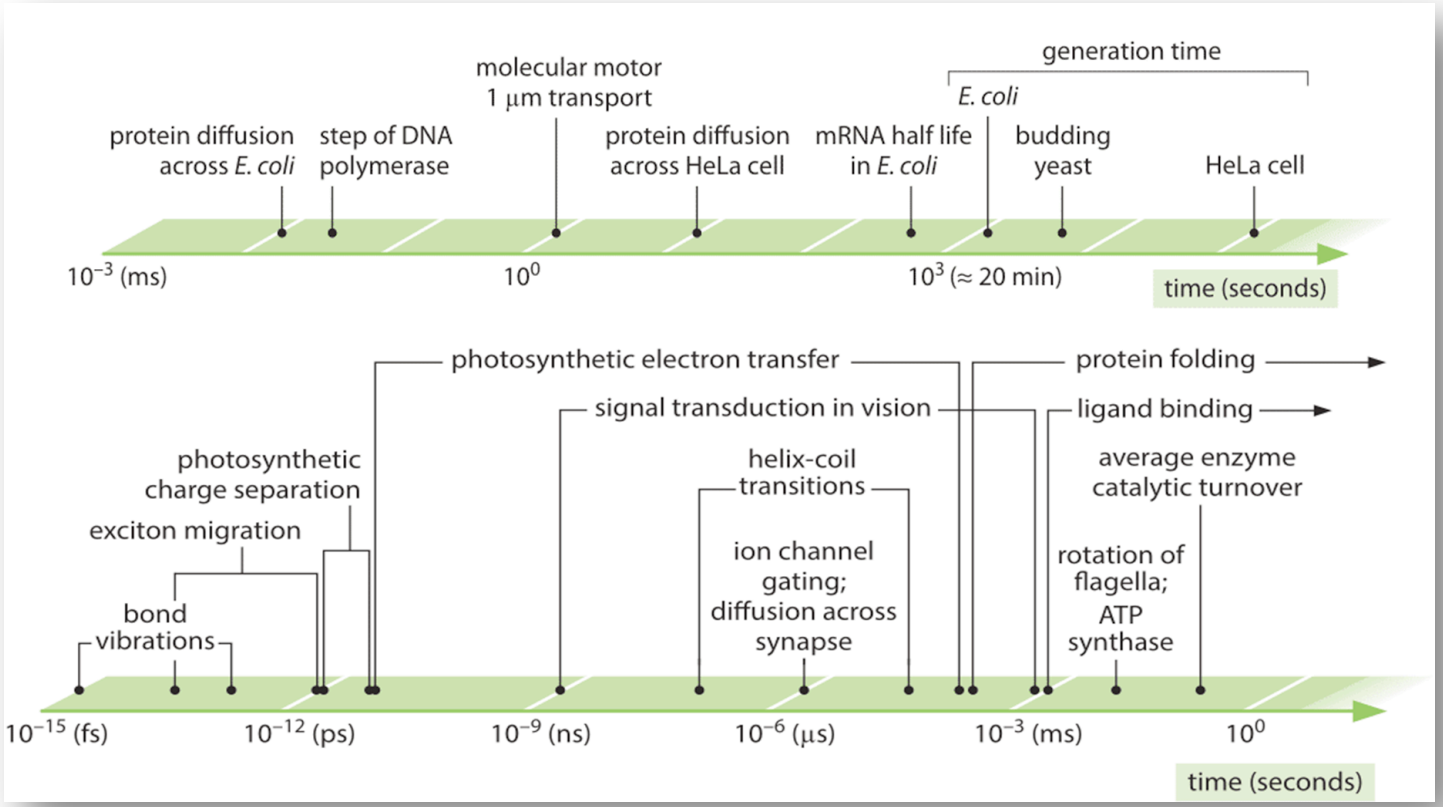
Timescales in Protein Folding Process



Side Chain Fluctuations: picoseconds

Protein Folding: microseconds

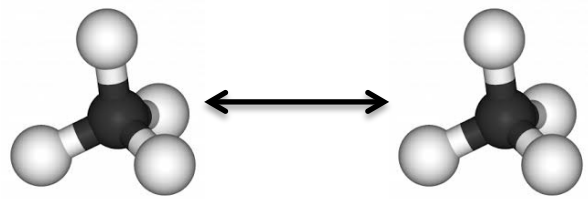
Hierarchy of different timescales in biology



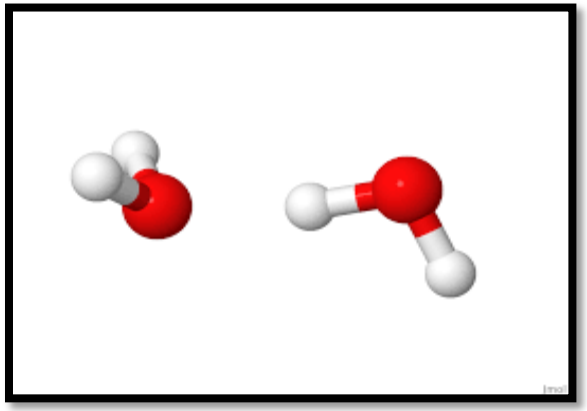
Physical Biology of the Cell: Rob Phillips

Sampling *Important* Degrees of Freedom

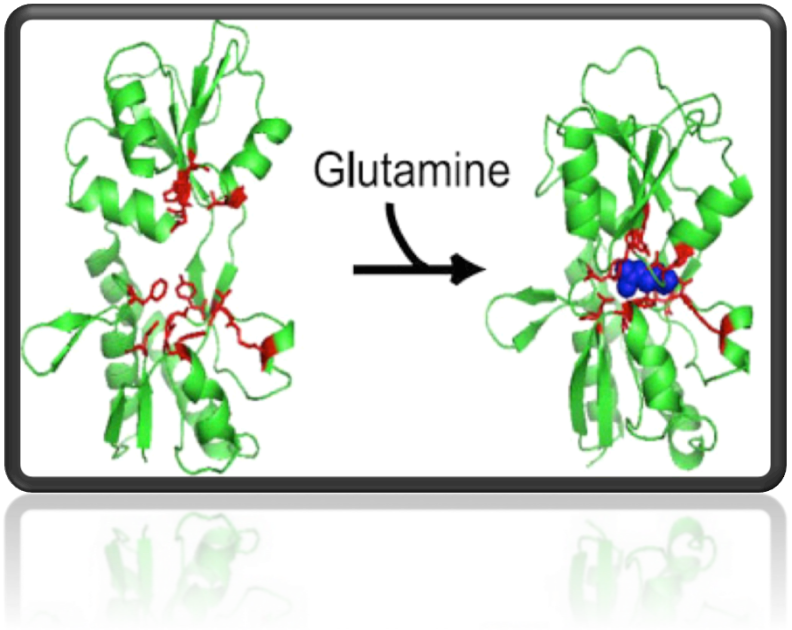
Methane-Methane



Water-Water

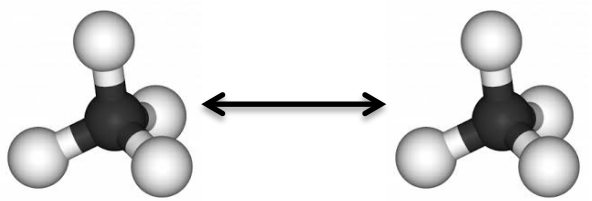


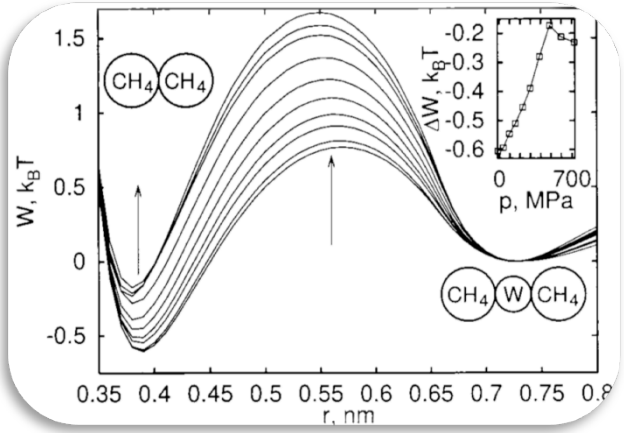
Ligand-Protein



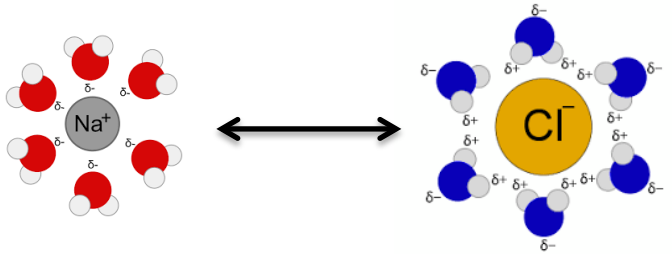
Potential of Mean Force: 2 Molecules in Solution

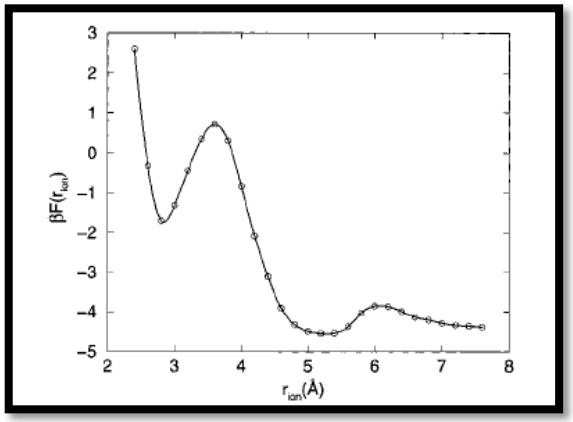
Methane-Methane





Sodium-Chloride



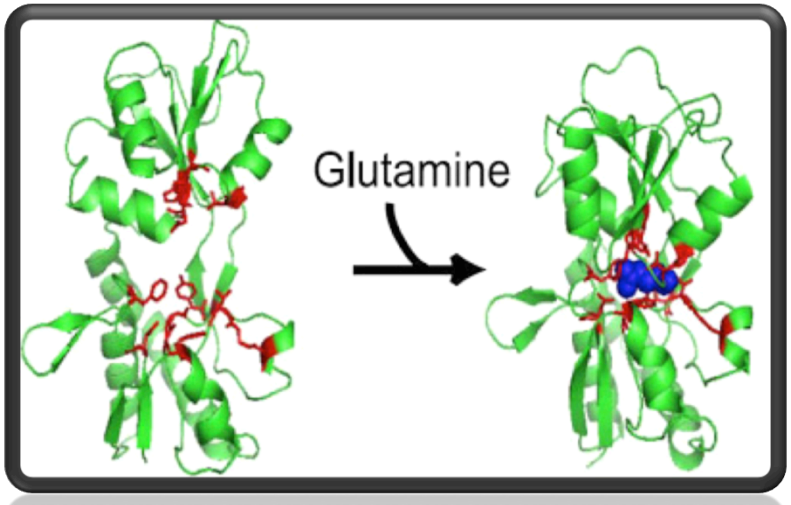


Connecting to Experiments: Thermodynamics, Kinetics, Pathways

Two Biomolecules Interacting in Solution: Energies

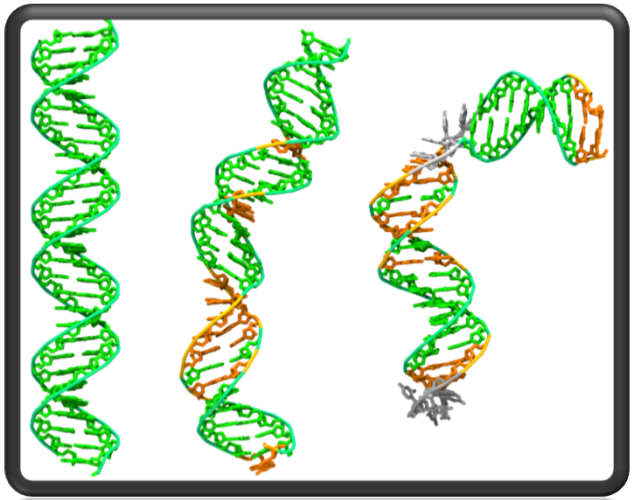
What kinds/forms of energy?

Ligand-Protein



Electrostatic/van der Waals

DNA Bending



Elastic

Home Work: Reading Material for this week

Physics Today: Rob Phillips

The Biological Frontier of Physics

Problems at the interface between biology and physics offer unique opportunities for physicists to make quantitative contributions to biology. Equally important, they enrich the discipline of physics by challenging its practitioners to think in new ways.

Rob Phillips and Stephen R. Quake

In the introduction to his classic magnetic-monopole paper of 1931, Paul Dirac remarked,

> There are at present fundamental problems in theoretical physics awaiting solution, e.g. the relativistic formulation of quantum mechanics and the nature of atomic nuclei (to be followed by more difficult ones such as the problem of life), the solution of which problems will presumably require a more drastic revision of our fundamental concepts than any that have gone before.[1]

structural biology, biochemistry, and genetics have yielded an explosion of biological data that are increasingly quantitative in character. For example, gene expression is routinely characterized in terms of how much, when, and where. Similarly, data on some machinery of the cell are reported graphically in terms of force–velocity curves. As a result, despite the field's reputation as a soft science, nearly all of biology is now ripe for quantitative analysis of the sort that physicists are used to. The opportunities are analogous to those that came to astrophysics once astronomical observations were coupled to spectrometry.

Life presents many interesting questions for physicists. As illustrations, we discuss three problems at the interface between physics and biology—steppingstones to more general thinking that will enrich physics. First, we describe the molecular machines that form the basis of life. The energies and length scales at which those machines operate are intriguing because they are in the regime

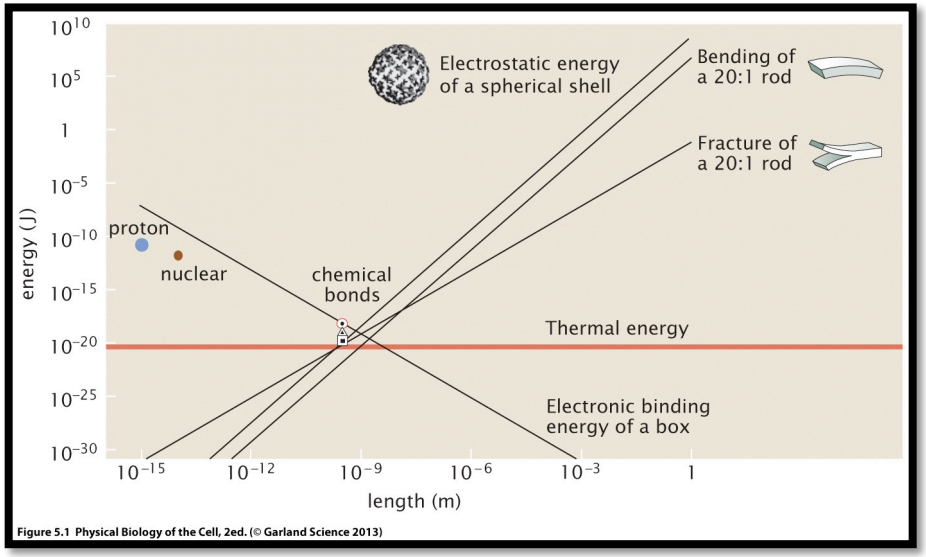
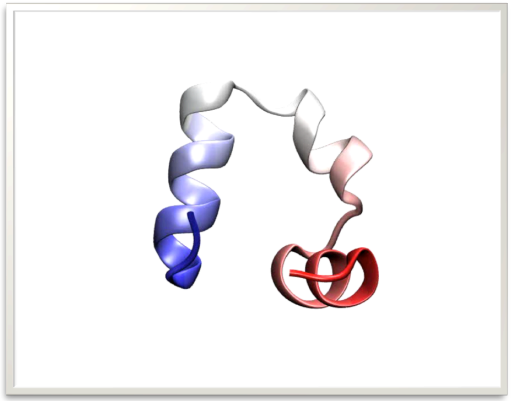
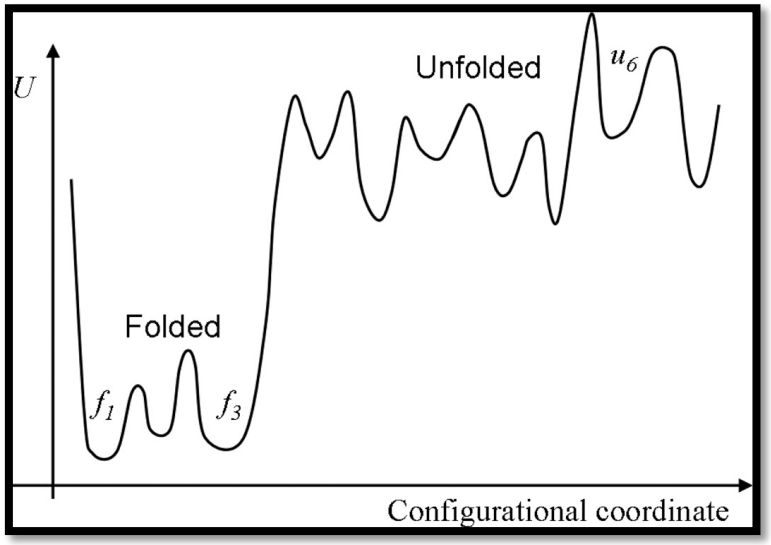


Figure 5.1 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

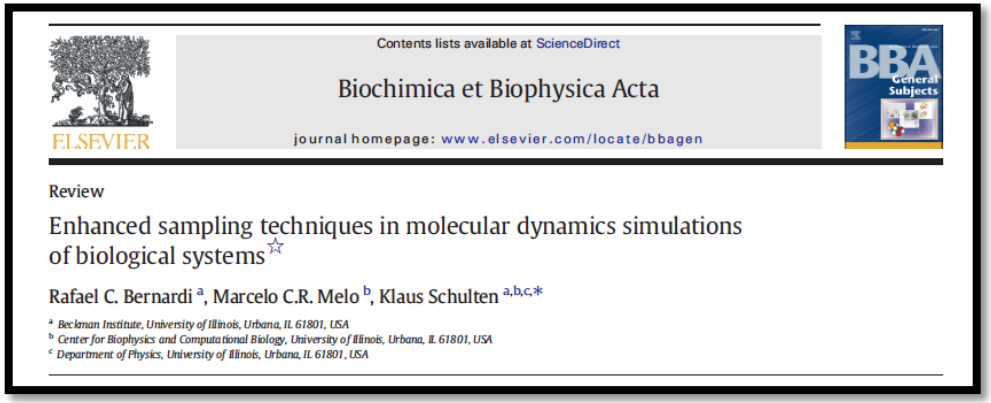
$$k_B T = 0.6 kcal/mol = 2.5 kJ/mol = 25 meV$$

Overcoming the Sampling Problem





Smart Computer Algorithms and Statistical Mechanics



Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

ELSEVIER

BBA General Subjects

Review

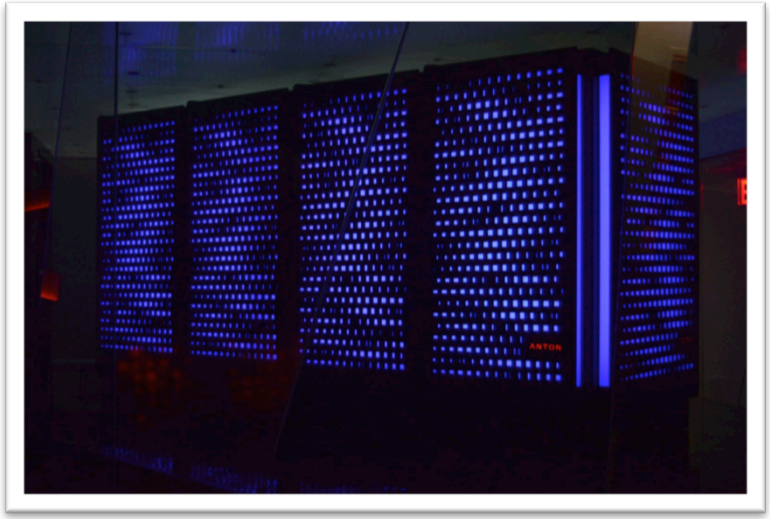
Enhanced sampling techniques in molecular dynamics simulations of biological systems☆

Rafael C. Bernardi [a], Marcelo C.R. Melo [b], Klaus Schulten [a,b,c,*]

[a] Beckman Institute, University of Illinois, Urbana, IL 61801, USA
[b] Center for Biophysics and Computational Biology, University of Illinois, Urbana, IL 61801, USA
[c] Department of Physics, University of Illinois, Urbana, IL 61801, USA

Rage Against the Machines: Millisecond Dynamics

D.E. Shaw

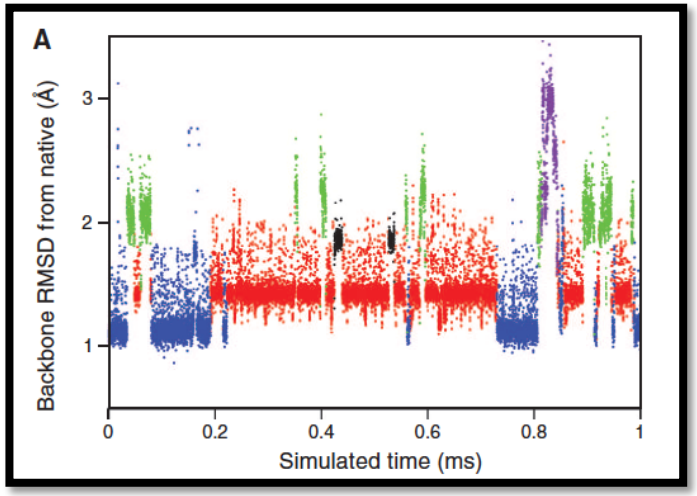


Anton Machine

Atomic-Level Characterization of the Structural Dynamics of Proteins

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

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



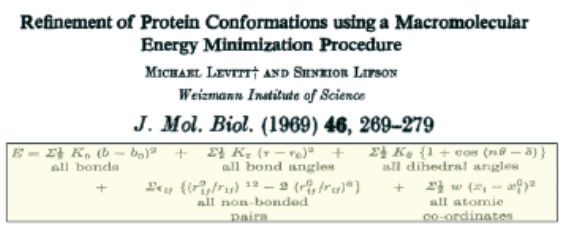
What About Modeling Chemistry?

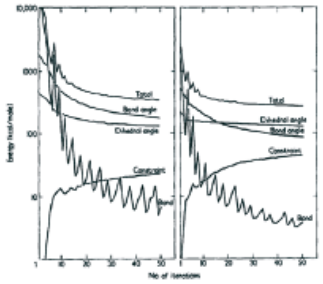
MACROMOLECULAR ENERGY MINIMIZATION 1969

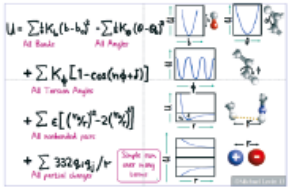
Refinement of Protein Conformations using a Macromolecular Energy Minimization Procedure

MICHAEL LEVITT† AND SHNEIOR LIFSON
Weizmann Institute of Science

J. Mol. Biol. (1969) **46**, 269–279

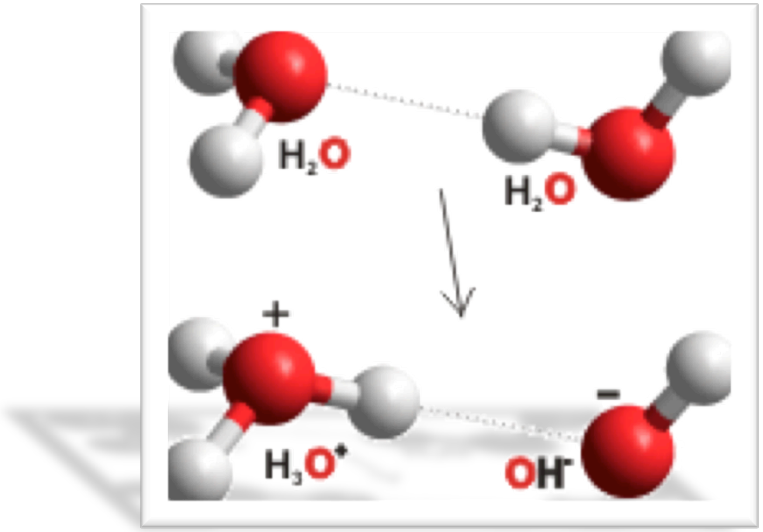






First protein structure refinement

FIGURE 10. Steepest descent energy minimization was used to move all non-hydrogen atoms by changing the Cartesian coordinates of the two proteins, myoglobin and lysozyme. This reduced the net forces and moved the structure towards an equilibrium. Note how a restraint on atom positions was used to correct for limitations of the energy function, principally the omission of the coulombic electrostatic term. Our paper [11] reports 50 steps of minimization, which is totally trivial by today's standards; these 50 steps took about 1000 secs. on the Golem A computer. The same calculation of forces used for energy minimization could also be used to simulate molecular dynamics (Fig. 8), which had previously been applied by Annesur Rahman to liquid argon [12] and then together with Frank Stillinger to more complicated liquid water [13].



Bond Breaking & Formation

Quantum Mechanics/Molecular Mechanics

Theoretical Studies of Enzymic Reactions:

Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme

A. Warshel and M. Levitt

Medical Research Council Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, England

and

Department of Chemical Physics
The Weixmann Institute of Science
Rehovot, Israel

(Received 12 September 1975, and in revised form 10 February 1976)

QUANTUM MECHANICS OF ENZYMIC REACTIONS 1976

J. Mol. Biol. (1976) **103**, 227–249

A. Warshel and M. Levitt

Medical Research Council Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, England

and

Department of Chemical Physics
The Weizmann Institute of Science
Rehovot, Israel

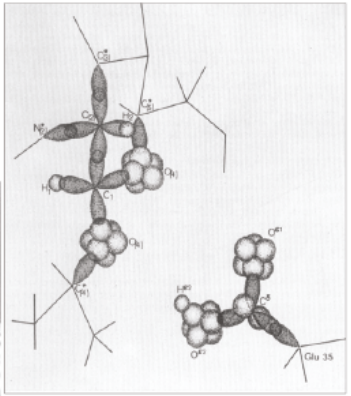
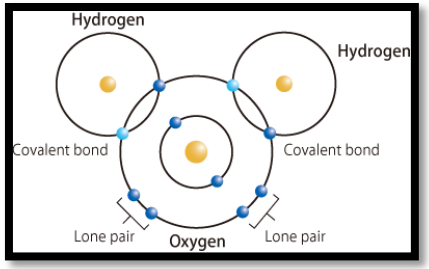


FIGURE 13. When Philips solved the x-ray structure of lysozyme, he proposed that its catalytic action is due to using binding energy to distort the substrate. Specifically, the six member sugar ring adjacent to the bond to be cleaved was thought to be deformed from a chair to a half-boat. Calculations done in my thesis [X:17] and published in a conference proceedings volume [X:18] showed that the enzyme was too soft to cause such a deformation and led us to propose electrostatic rather than steric strain. With Arieh Warshel, we added quantum mechanical orbitals to a small part of the system, while the rest was still treated classically in what has become known as QM/MM. The calculations now possible showed that the substrate is indeed electrostatically strained [X:19].

Numerous Applications in Enzyme Catalysis

Car-Parrinello/Ab Initio Molecular Dynamics



VOLUME 55, NUMBER 22 PHYSICAL REVIEW LETTERS 25 NOVEMBER 1985

Unified Approach for Molecular Dynamics and Density-Functional Theory

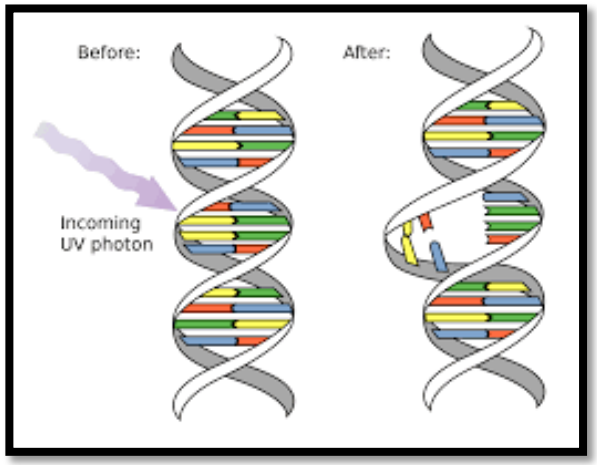
R. Car
International School for Advanced Studies, Trieste, Italy
and
M. Parrinello
Dipartimento di Fisica Teorica, Università di Trieste, Trieste, Italy, and International School for Advanced Studies, Trieste, Italy
(Received 5 August 1985)

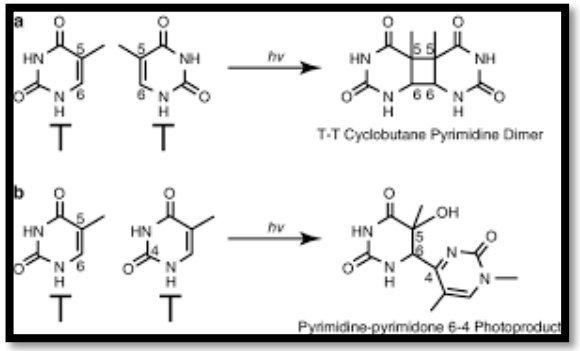
We present a unified scheme that, by combining molecular dynamics and density-functional theory, profoundly extends the range of both concepts. Our approach extends molecular dynamics beyond the usual pair-potential approximation, thereby making possible the simulation of both covalently bonded and metallic systems. In addition it permits the application of density-functional theory to much larger systems than previously feasible. The new technique is demonstrated by the calculation of some static and dynamic properties of crystalline silicon within a self-consistent pseudopotential framework.

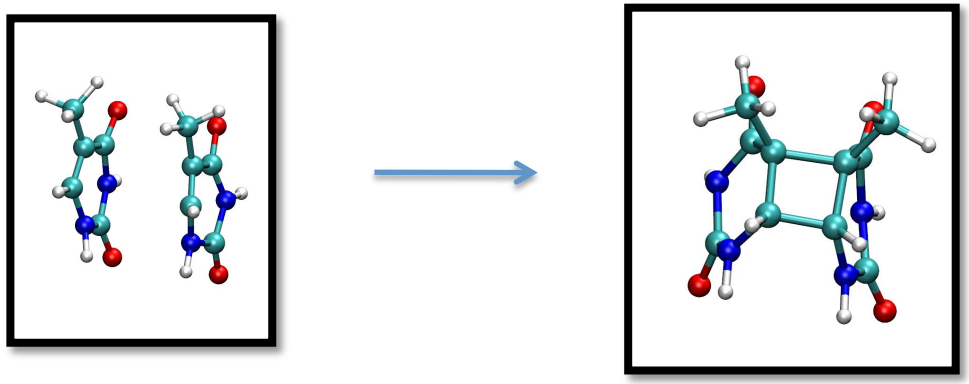
PACS numbers: 71.10.+x, 65.50.+m, 71.45.Gm

Fictitious Dynamics of Electrons

Watching DNA Repair On the Computer



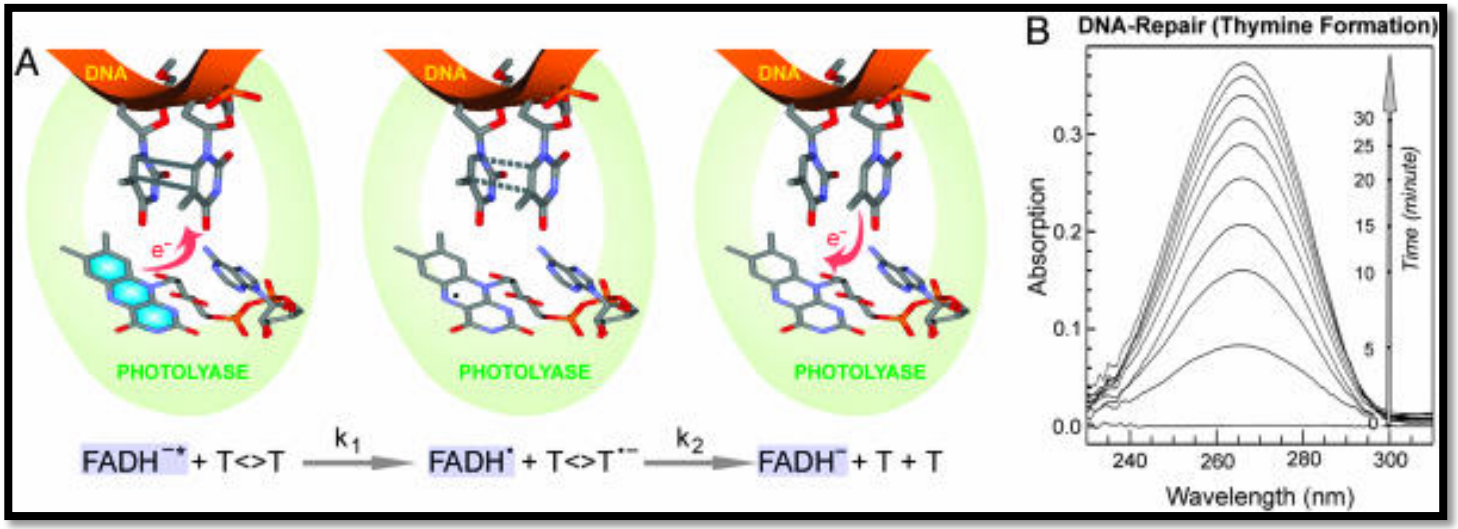




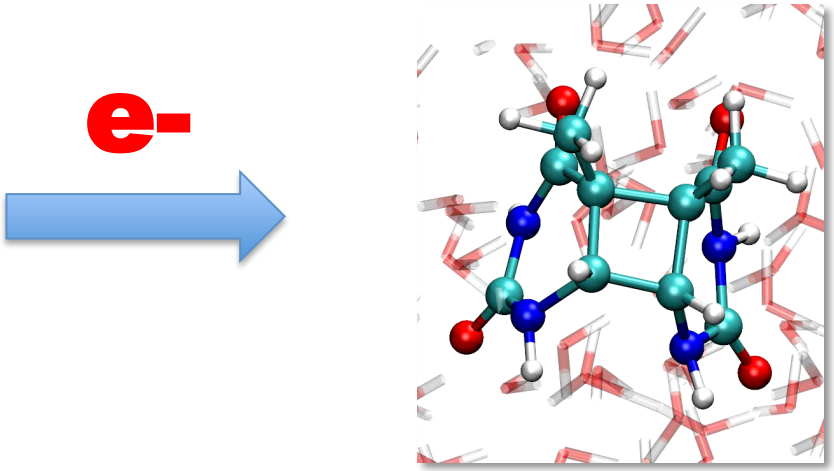
UV Light Can Damage the Skin

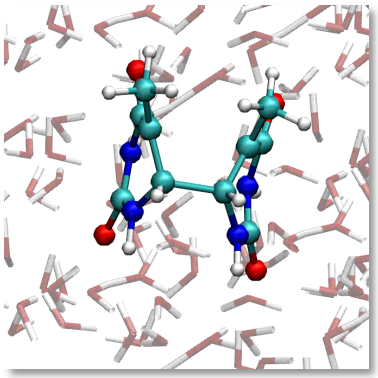
How is DNA Fixed by Nature: DNA Photolyase

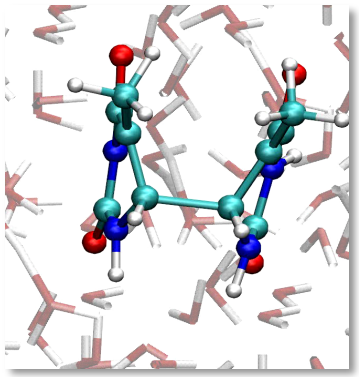
Aziz Sancar



Short-Circuiting DNA Repair on the Computer







J. Phys. Chem. B 115 (14), 3860, 2011

J. Phys. Chem. B 115 (14), 3848, 2011

Bigger and More Complex Systems: Can one simulate the cell?

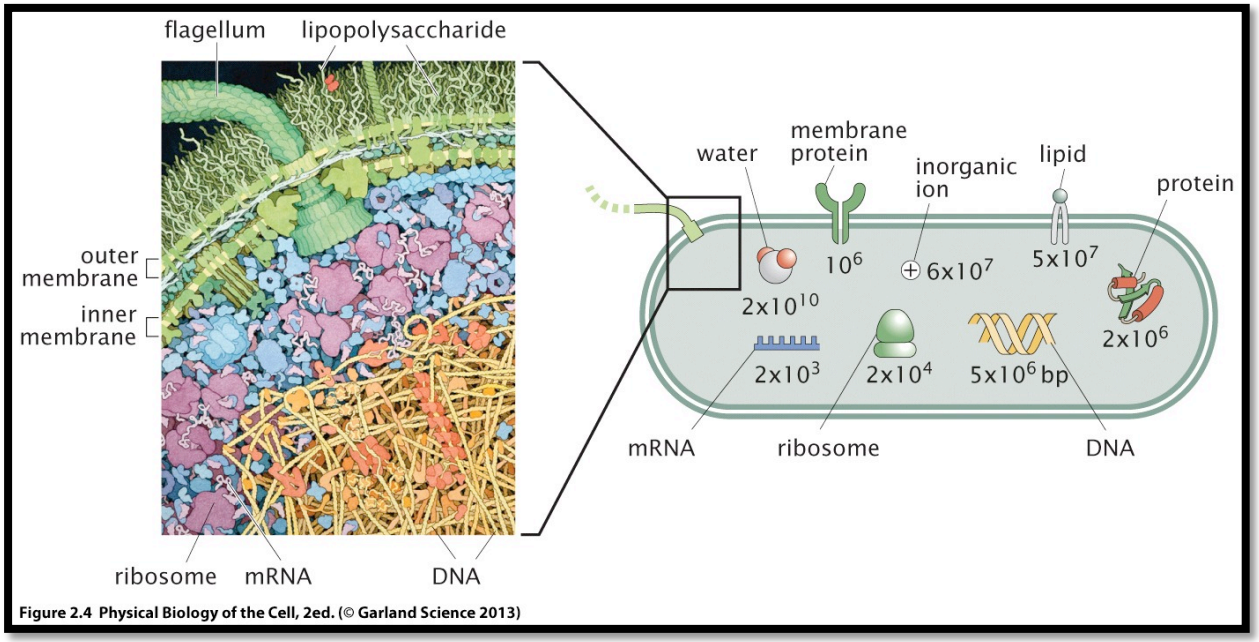
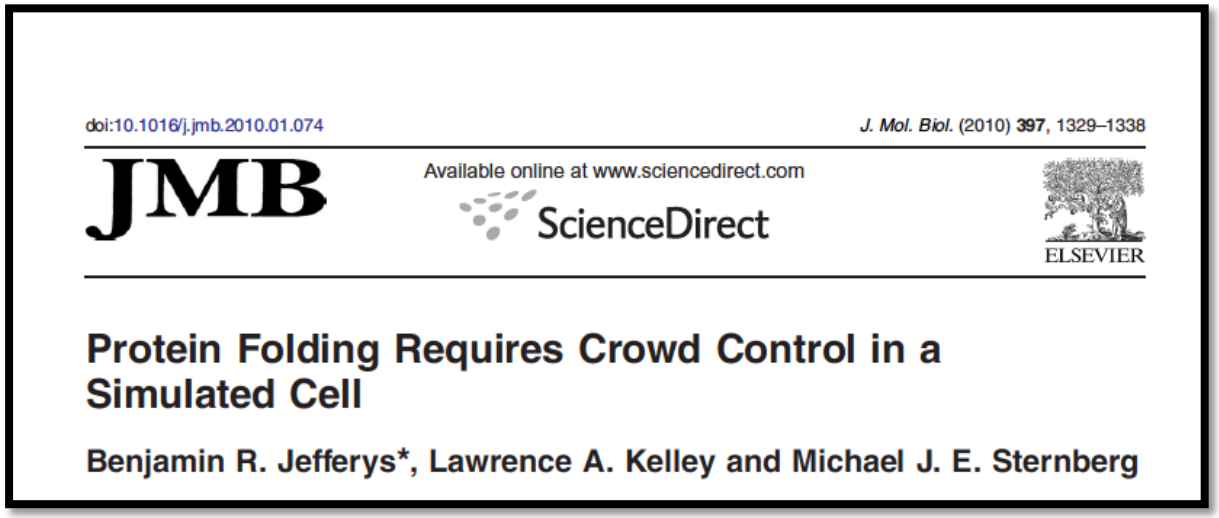


Figure 2.4 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Cells are much more crowded than our simple simulation models

Modeling Protein Folding in the Cell



doi:10.1016/j.jmb.2010.01.074 *J. Mol. Biol.* (2010) **397**, 1329–1338

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Protein Folding Requires Crowd Control in a Simulated Cell

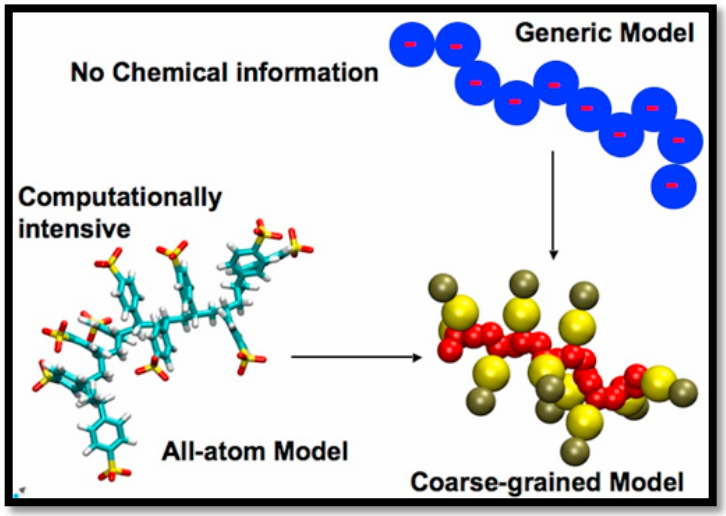
Benjamin R. Jefferys*, Lawrence A. Kelley and Michael J. E. Sternberg

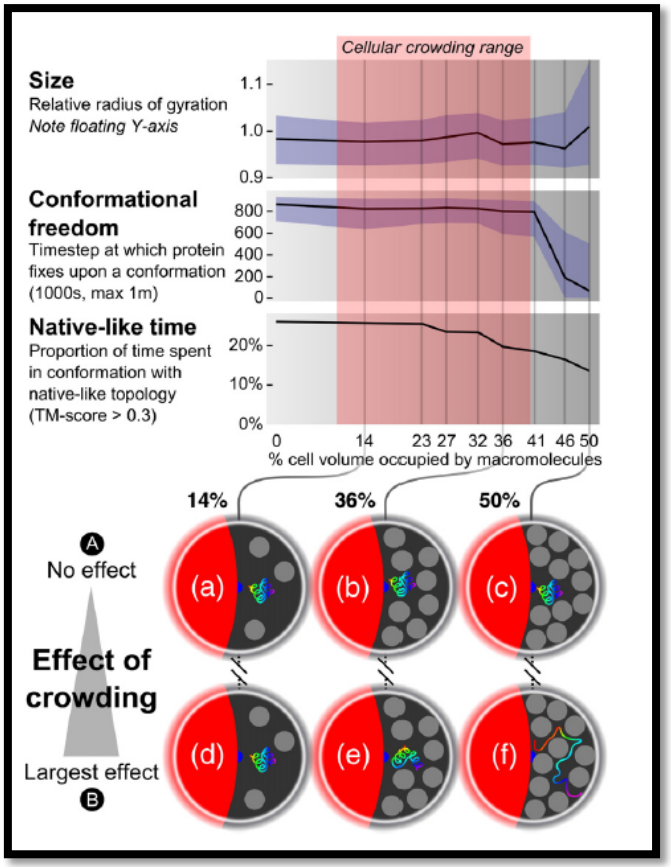
What is the biologically active conformation in the cell?

How does the protein find its state in the cell?

Modeling Protein Folding in the Cell

Coarse-Grained Protein and Water





Protein Folding Changes in the Cell

Modeling the Huge Ribosome

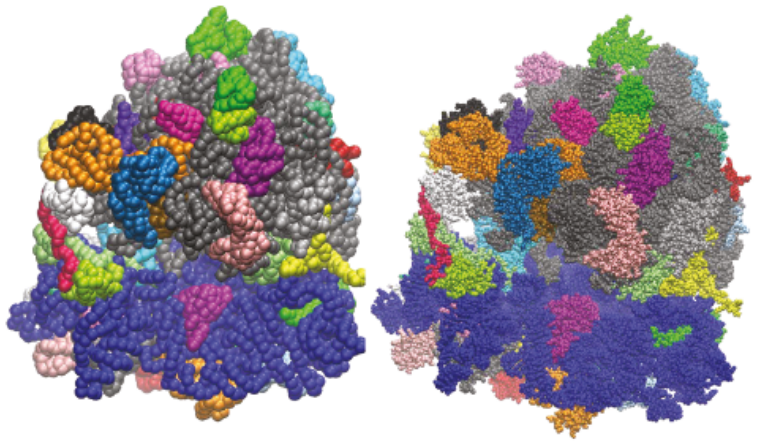
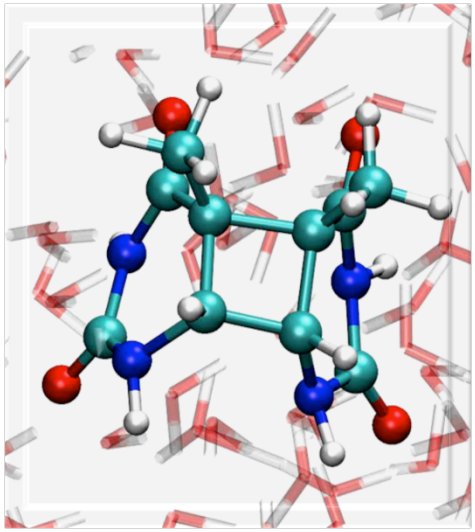


FIGURE 17. COARSE-GRAINED & ALL-ATOM NORMAL MODE DYNAMICS OF ENTIRE RIBOSOME. Together with Jenelle Bray & Junjie Zhang, the torsional angle normal mode method we developed in 1985 [29] has been improved so that it can handle any number of independent bodies each with its own rotational and translational degrees of freedom. Although the entire ribosome is large, with 4,500 nucleotides in 7 RNA chains and 6,000 amino acids in 49 protein chains [30–32], we can represent its low-frequency motion by just 538 degrees of freedom, 6 for each of 56 chains and an additional 202 for internal degrees of freedom. The motion is simulated with all 167,000 atoms as well as with 11,062 interaction centers in a coarse-grained representation like that we introduced [14]. The motions of the four lowest frequency modes are very similar for the two models. The video of these modes shows functionally suggestive relative motion of the heavy (30S) and light (16S) particles that include jaw closing, rotational grinding and rocking.

Michael Levitt:
Nobel Lecture

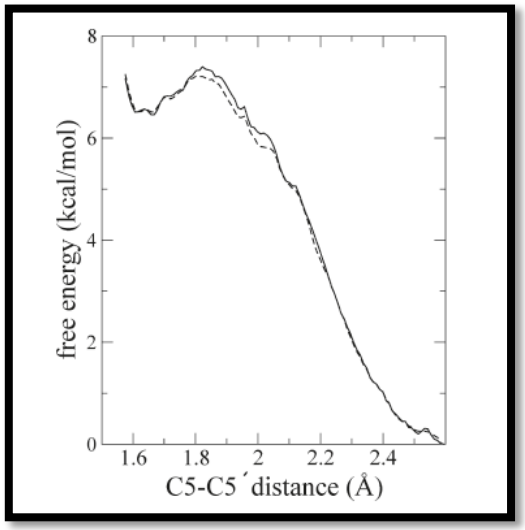
The Age of Big Biological Data & Compexity

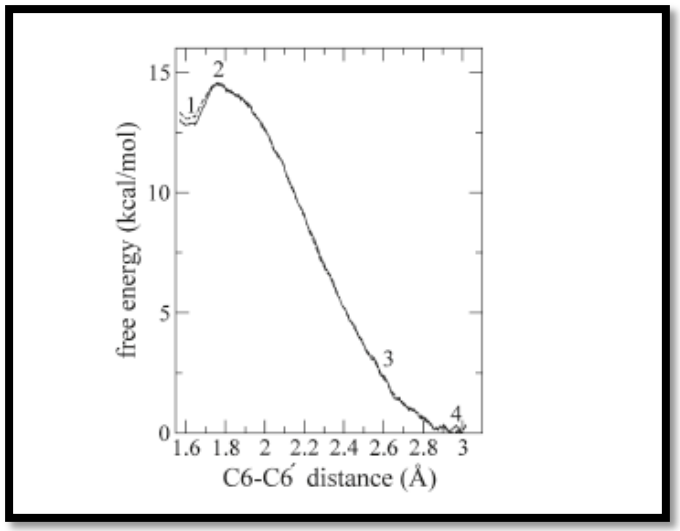


Chemical Intuition

Visualize Simulations

Dimensionality Reduction





Two Dimensional Free Energy Surfaces

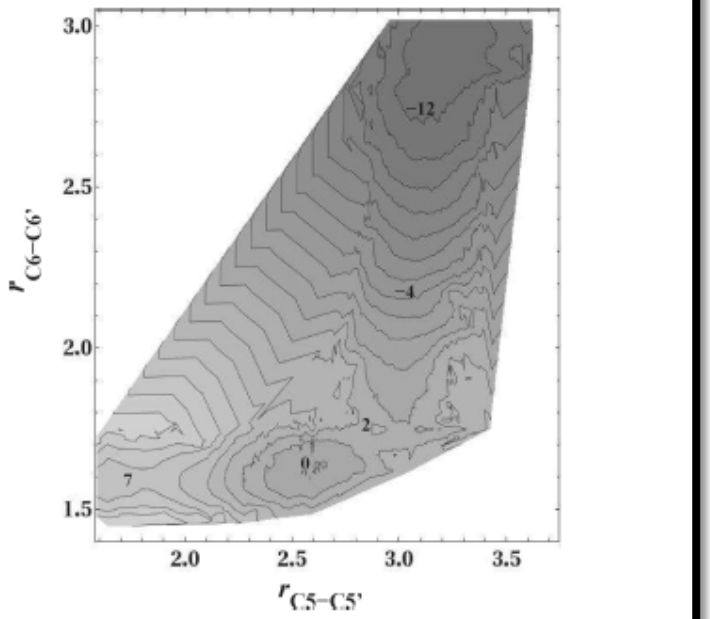
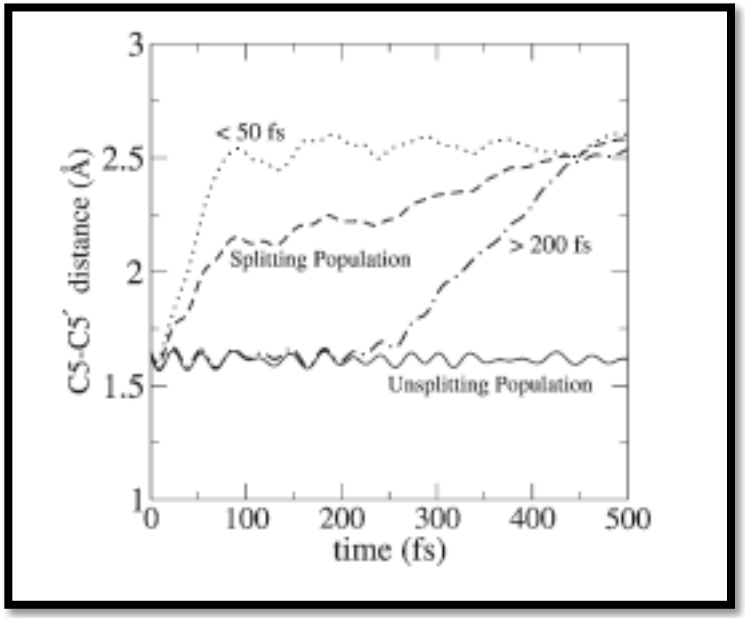


Figure 5. Overall reaction free energy surface for thymine CPD anion repair (kcal mol^{-1}). The free energy well in which the C5—C5′ bond is broken and the C6—C6′ bond is intact ($r_{C5-C5'}$ = 2.57 Å, $r_{C6-C6'}$ = 1.61 Å) is assigned as the zero of energy. The energy value on several contours is labeled.

What About Other Coordinates?

The Role of Water In Biology

Solvent Fluctuations Drive Chemistry



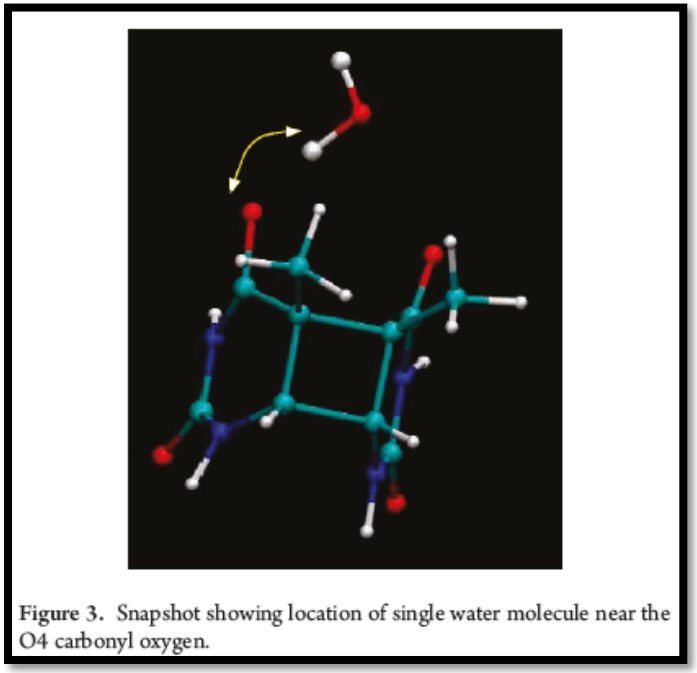
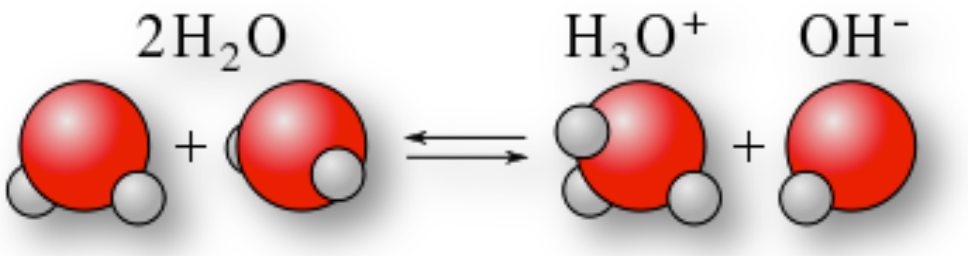


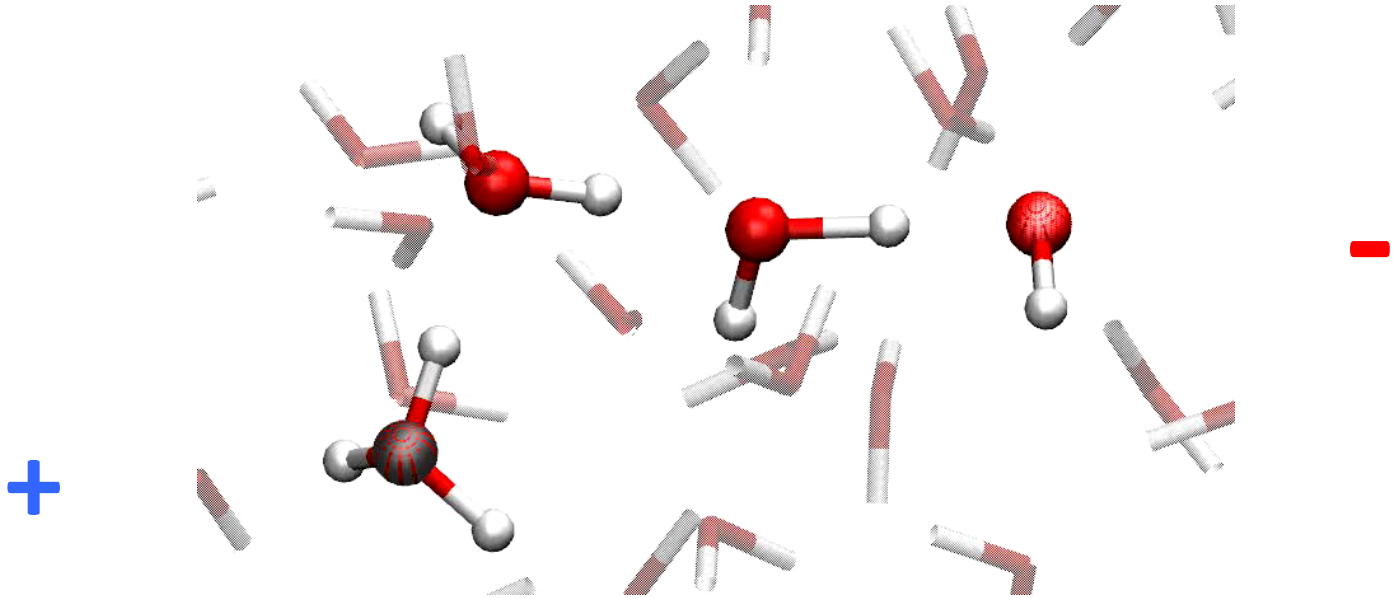
Figure 3. Snapshot showing location of single water molecule near the O4 carbonyl oxygen.

Smart Ways to Discover Important Coordinates Automatically

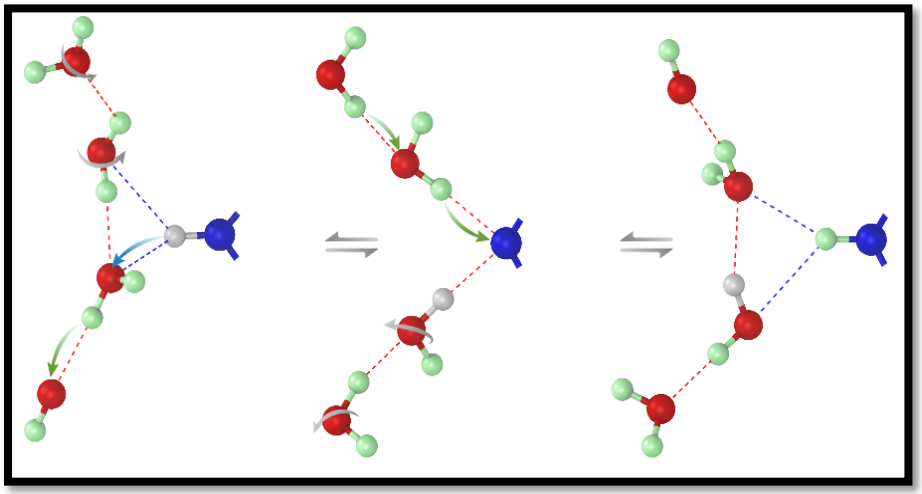
Water is Extremely Important in Biology



PNAS ***108(51):20410 (2011).***



Proton Exchange in Biological Systems

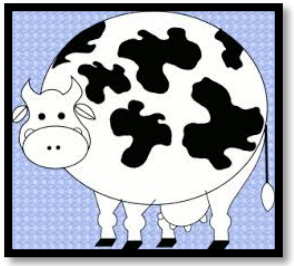


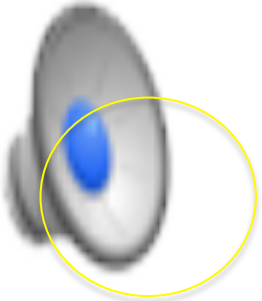
Base-Catalyzed: Hydroxide Ion

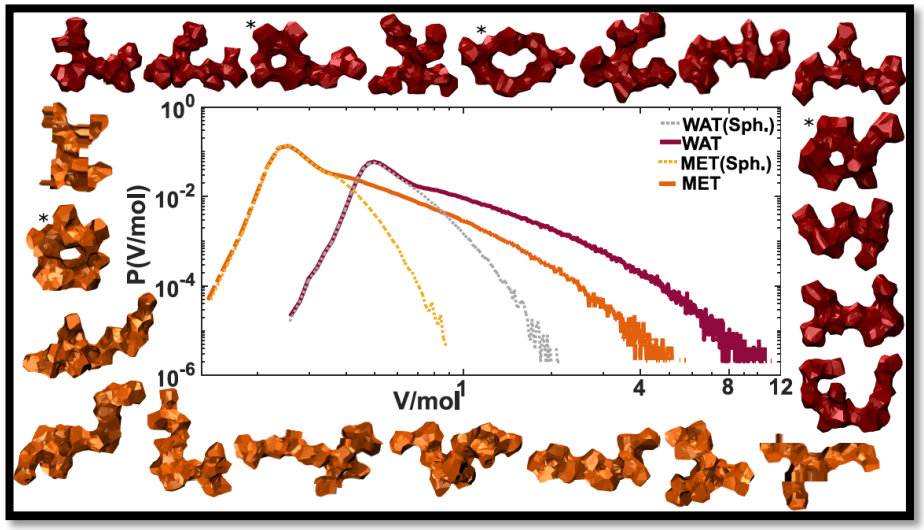
What is the mechanism by which this happens?

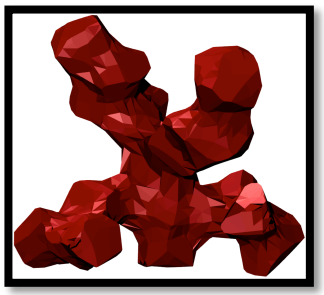
Going Beyond the Spherical Cow

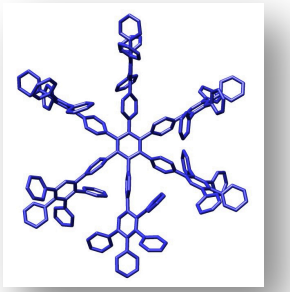
Simplify Models, but not that much!

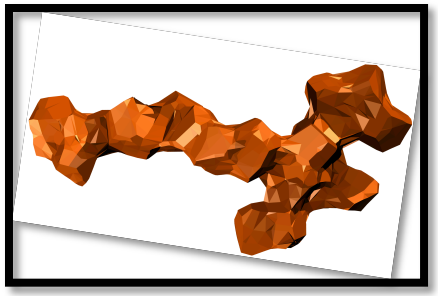


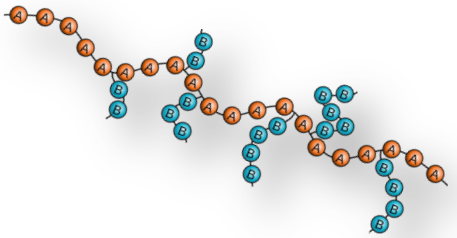




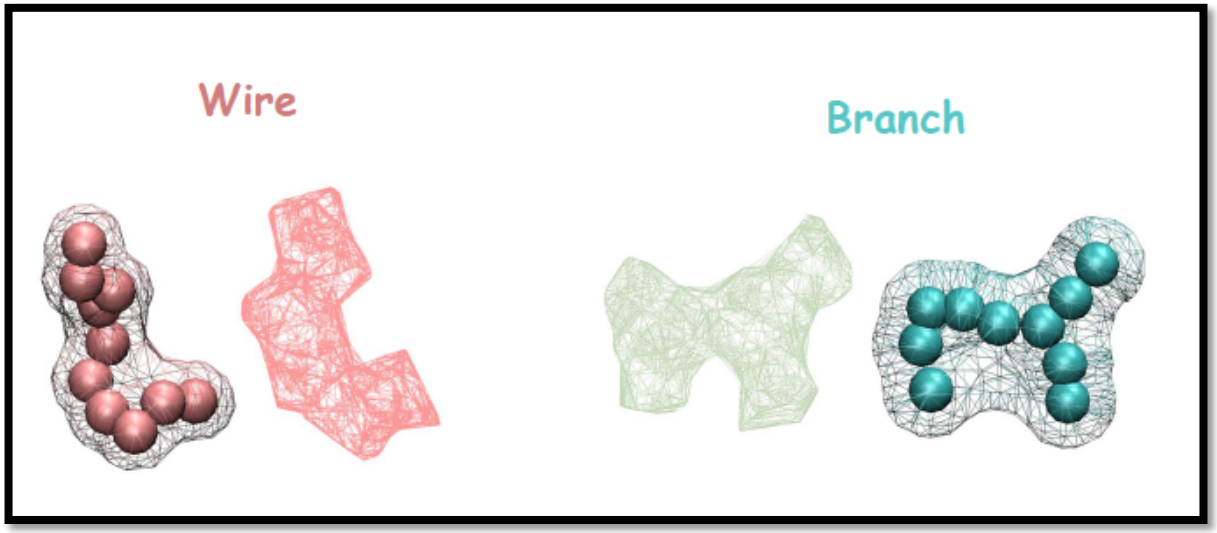








Shapes in Water Resemble Small Polymers



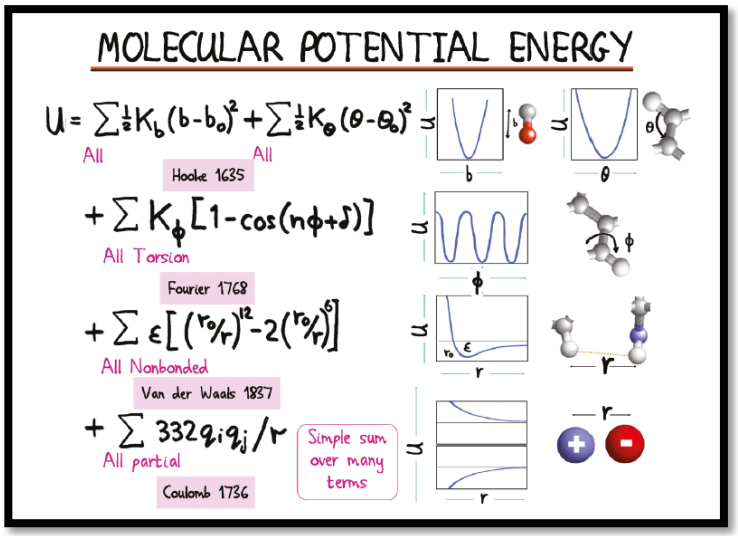
$$\Delta G_{solv} \sim 17 kJmol^{-1}$$

$$\Delta G_{solv} \sim 20.73 kJmol^{-1}$$

$$\Delta G_{void} \sim 17 kJmol^{-1}$$

J. Phys. Chem. Letters. Accepted

The Future of Biological Simulations



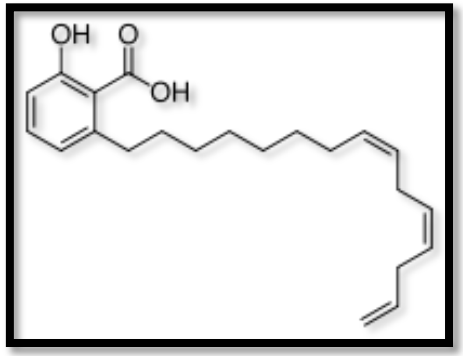
Beyond Classical Force fields: Polarization and Charge Transfer

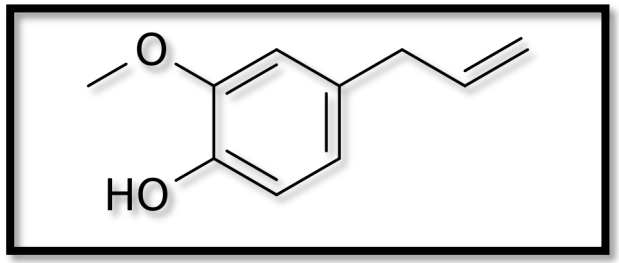
Quantum Mechanics: Electrons and Nuclei

Going Beyond Chemical Intuition

High dimensional Protein-Water Systems

Bio-Chemical Physics of Natural Products

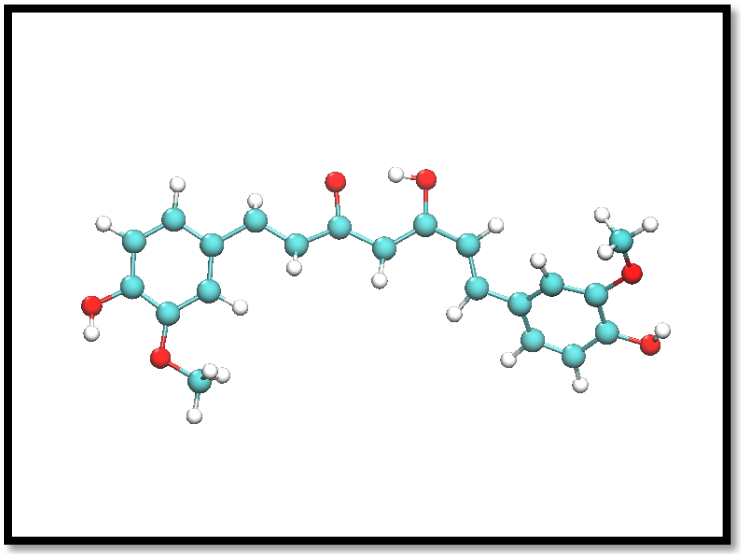




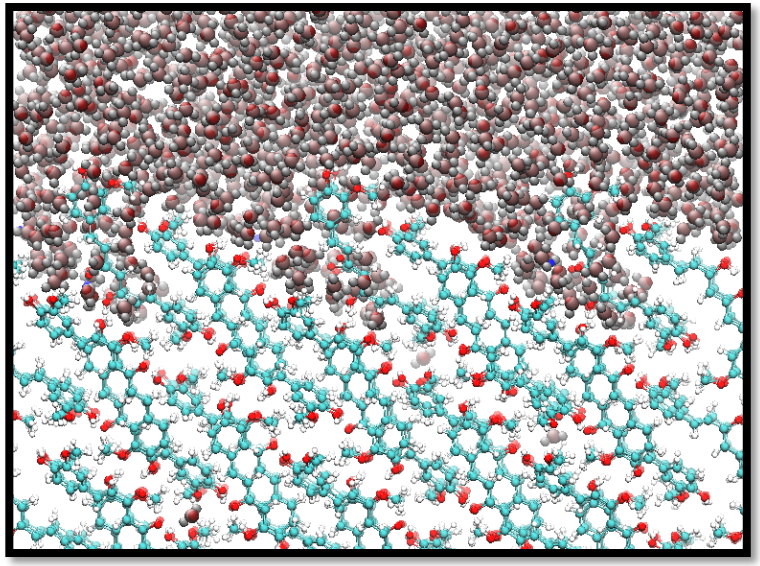
Theoretical studies of natural product compounds –

understanding the role of hydration

The Case of Curcumin in Water



MD Tutorial:
Curcumin in Liquid Water



Curcumin Crystal-Water Interfaces

Last Parting Thoughts

- Before Computation a Solid Background in Theory is Needed
- Beware of Simplifying Models in Computational Biophysics
- Lots of Interesting Basic Science in the Bio-Chemical Physics of Model Systems
- Making stronger connections to experiments

Computational vs Experimental Biologist