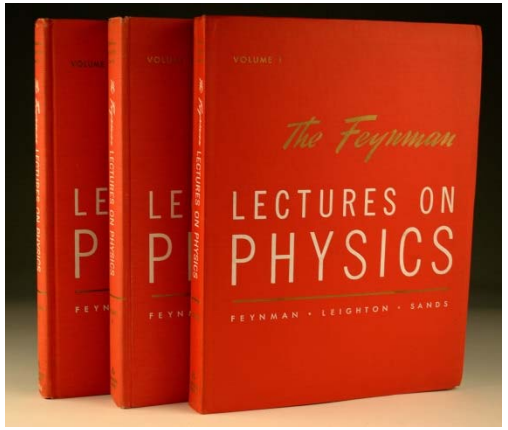
Molecular Dynamics Simulation, a household tool for future structural biology

Yibing Shan

What we want to achieve

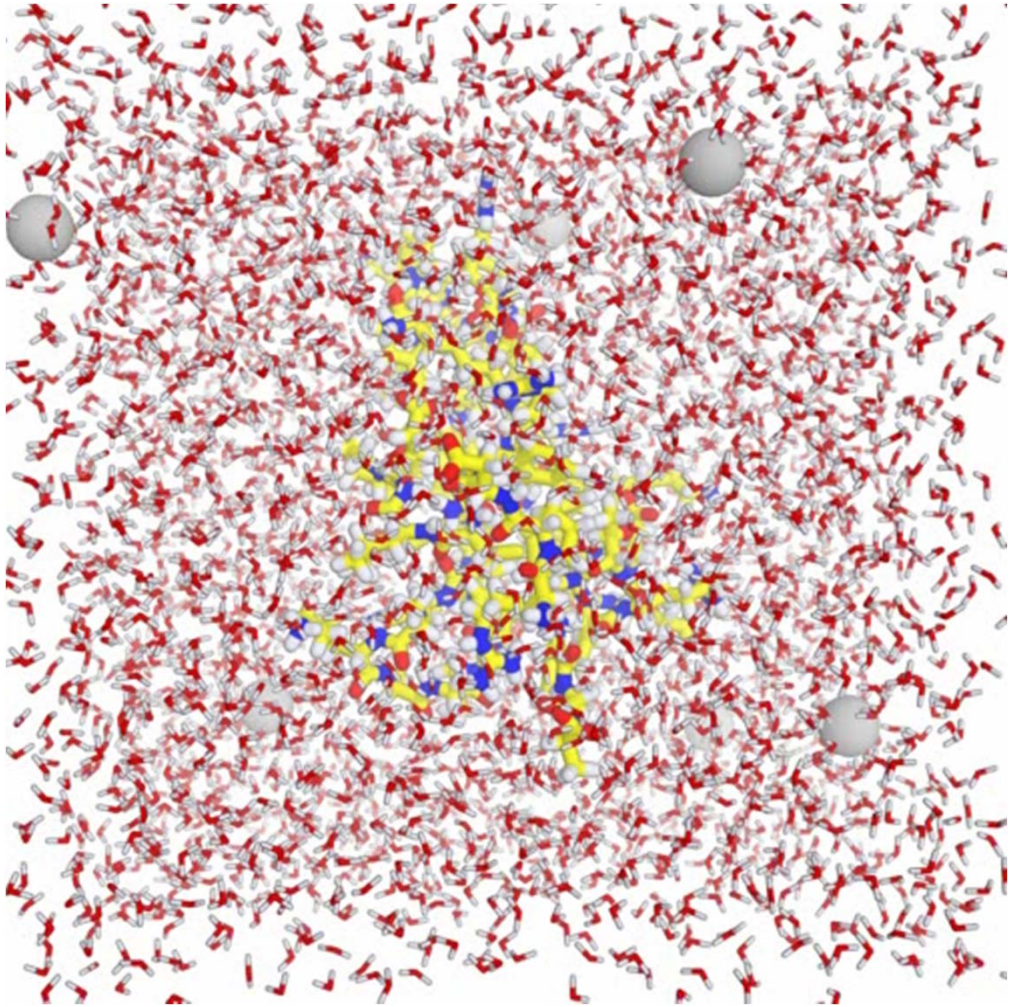
"…everything that is living can be understood in terms of the jiggling and wiggling of atoms."

--R. Feynman



What is molecular dynamics

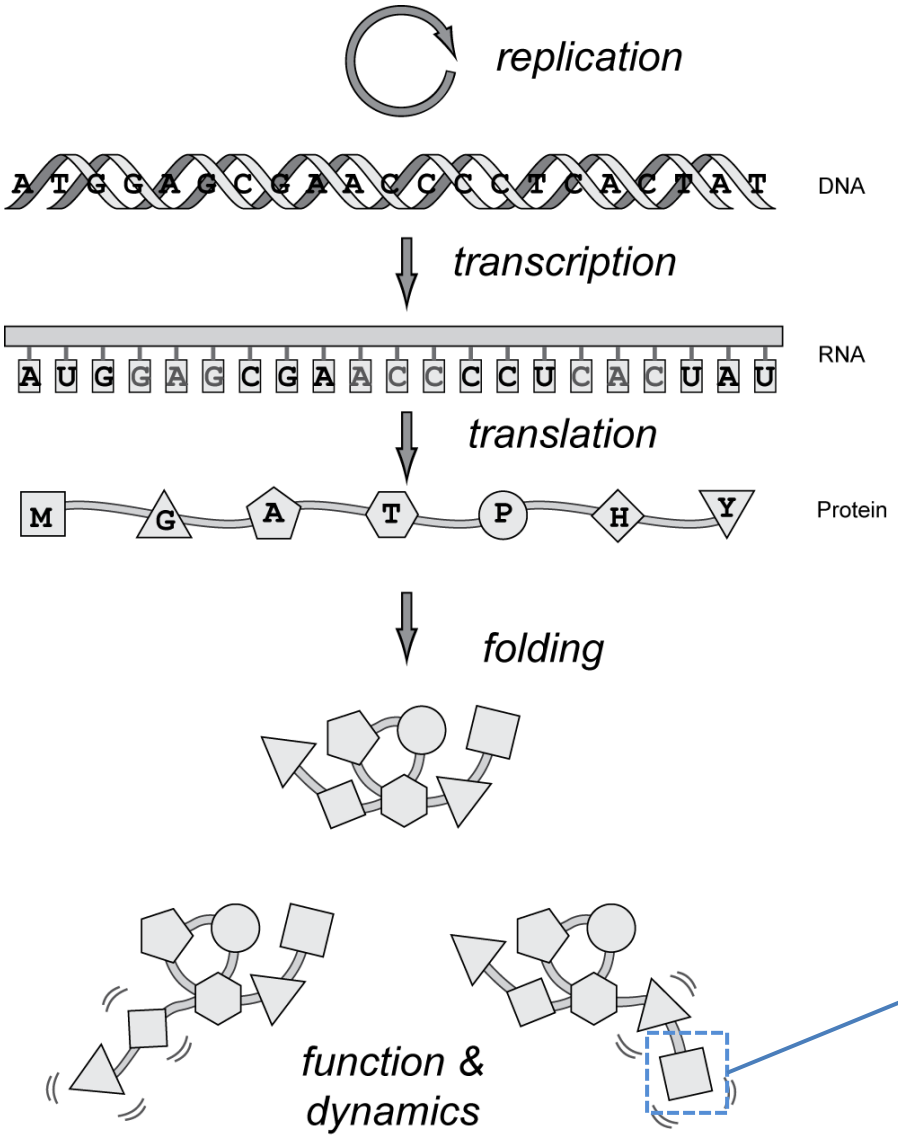
Typical MD trajectory output

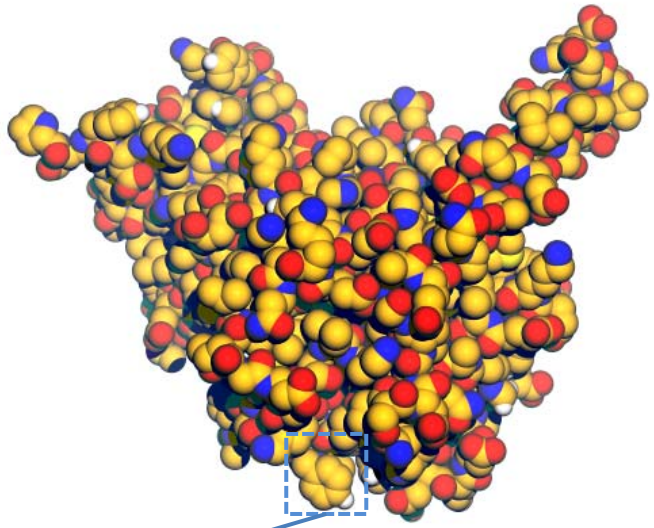


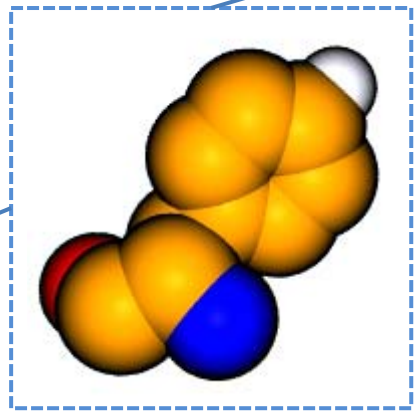
Everything Should Be Made as Simple as Possible, But Not Simpler

Einstein

Protein Structure and Function Paradigms

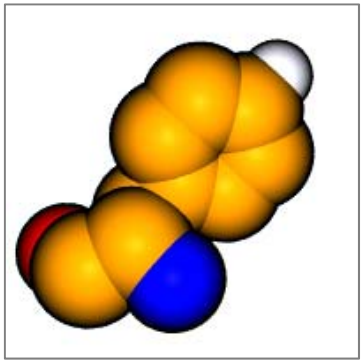




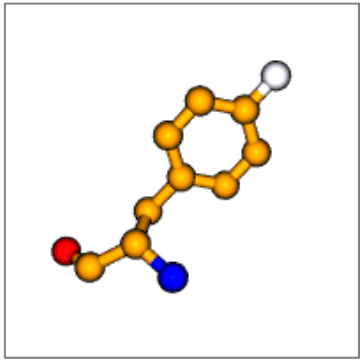


Molecular mechanics handles atoms connected by interactions

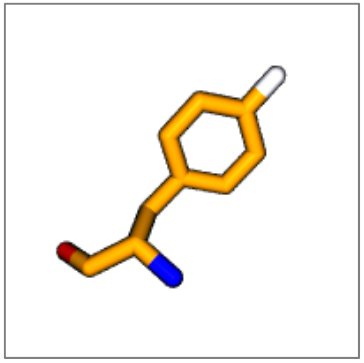
atoms

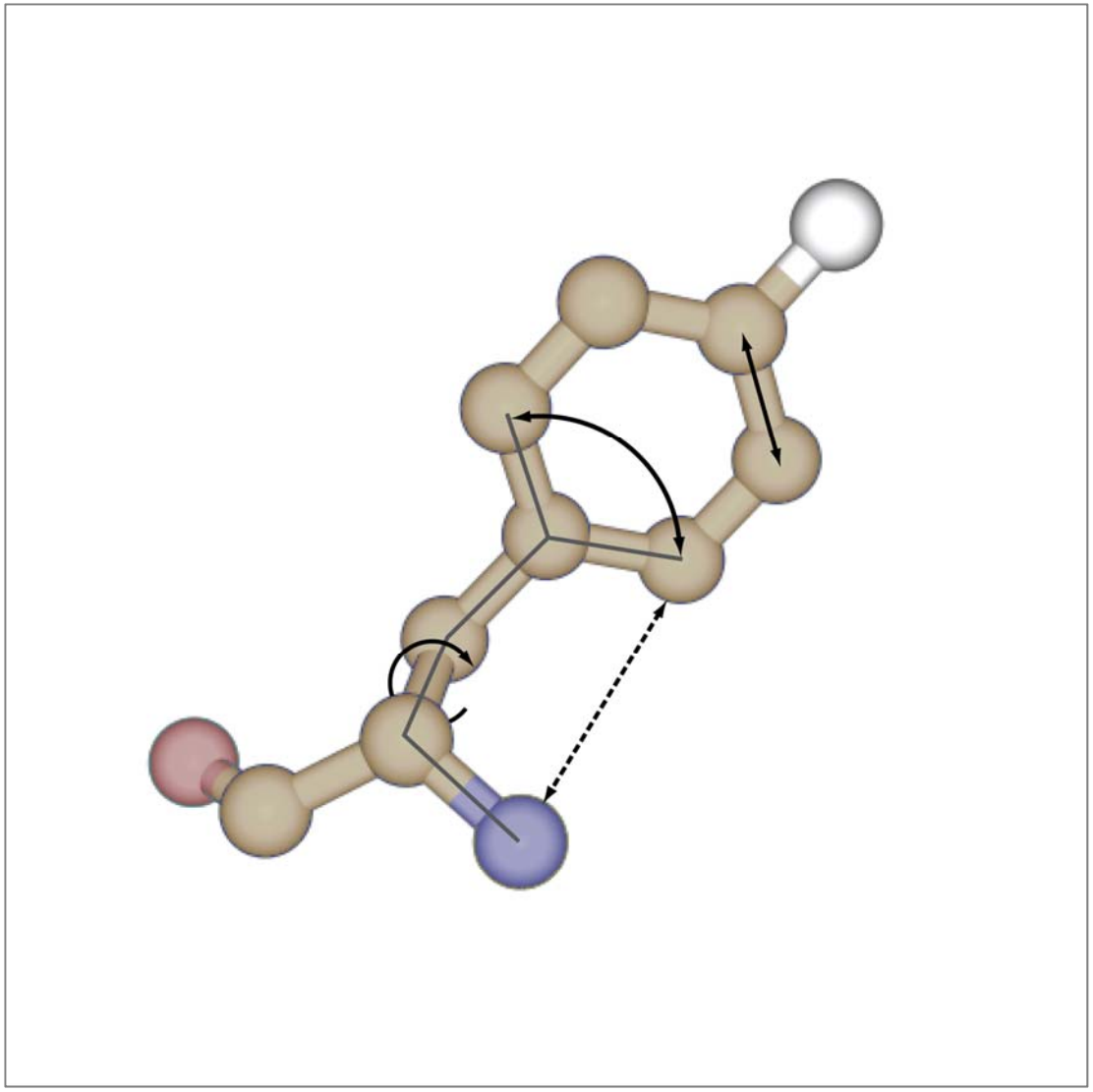


atoms & bonds



bonds



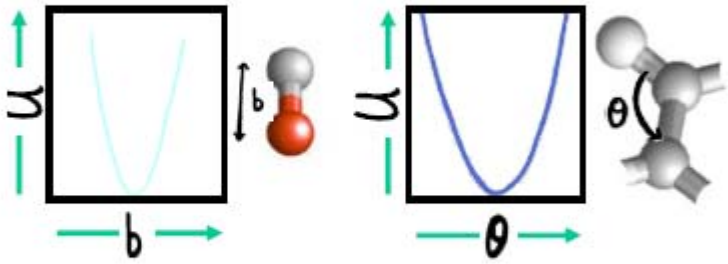


MOLECULAR POTENTIAL ENERGY

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

All Bonds — All Angles

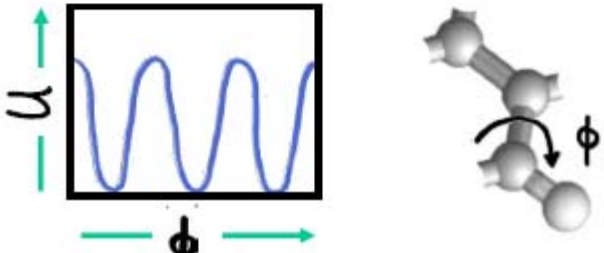
Hooke 1635



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

All Torsion Angles

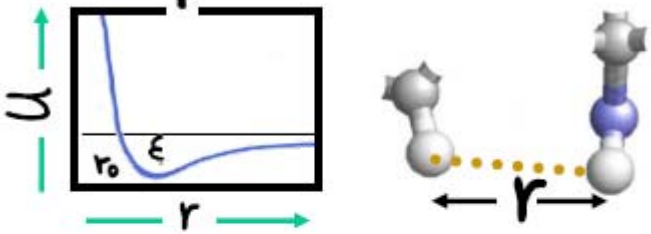
Fourier 1768



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

All Nonbonded pairs

Van der Waals 1837

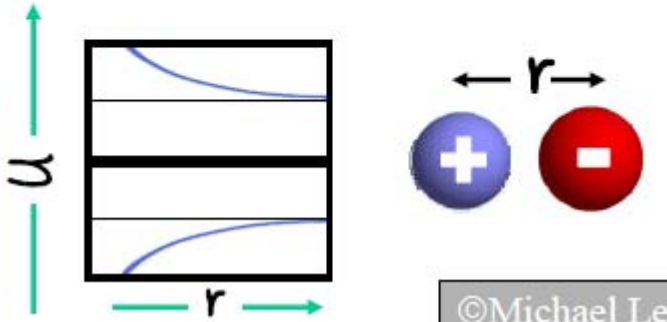


$$+ \sum 332 q_i q_j / r$$

All partial charges

Coulomb 1736

Simple sum over many terms

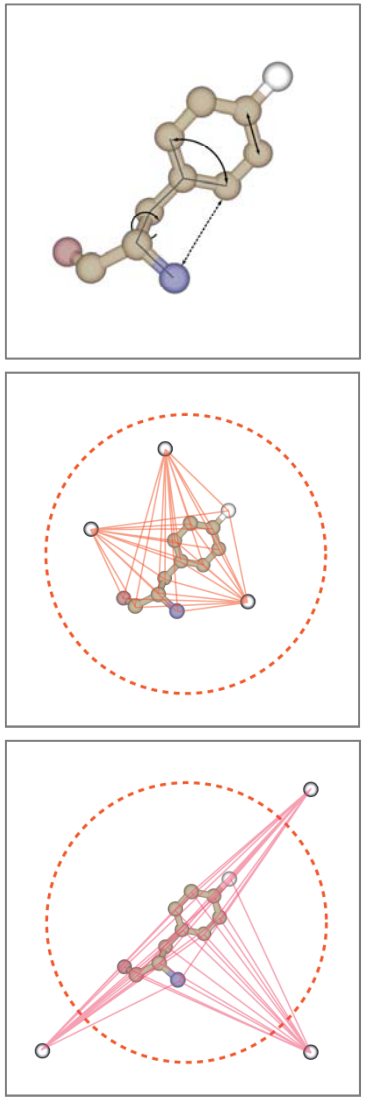


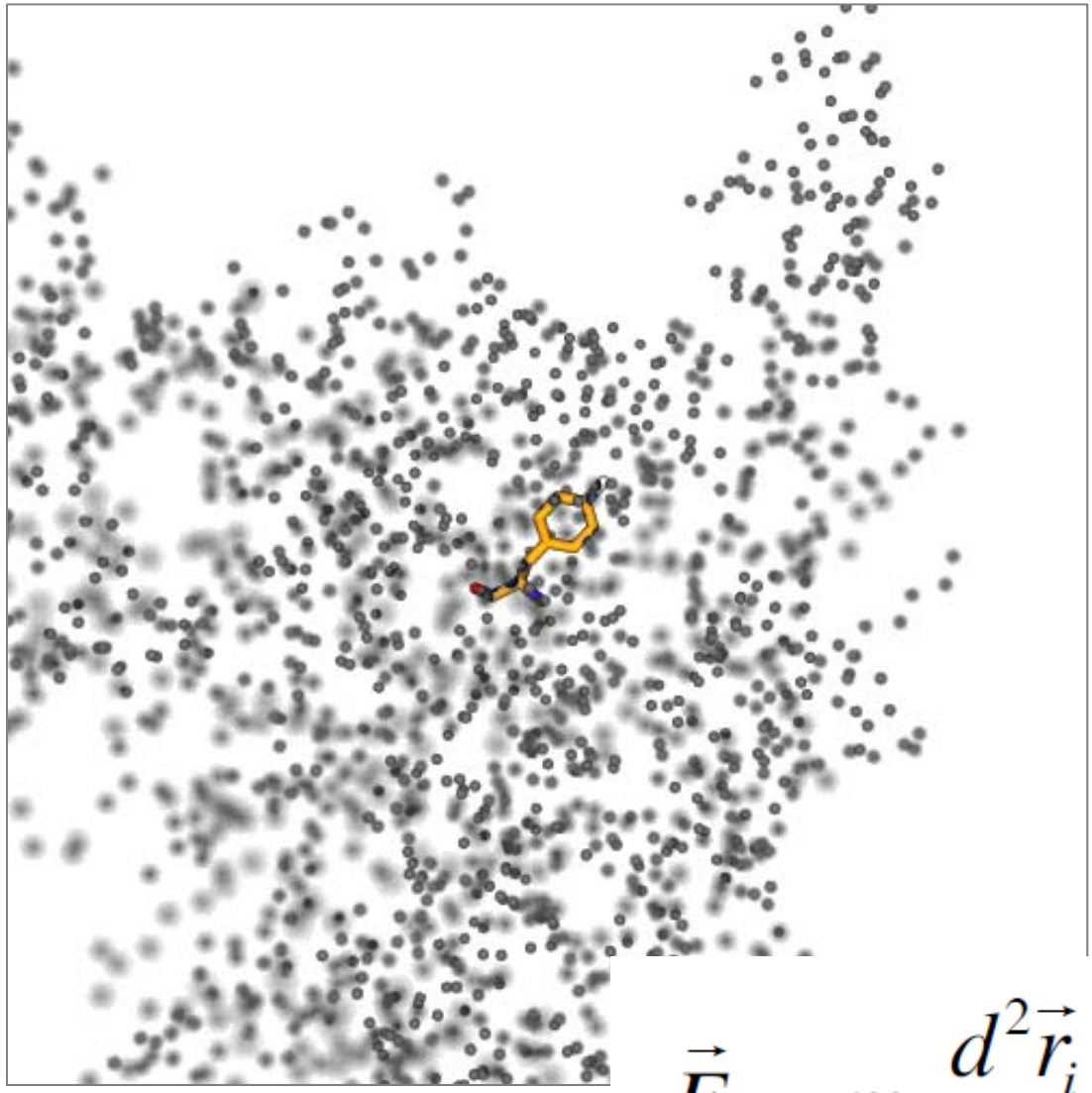
Integrate Newton's equation of motion

bonded

non-bonded
near

non-bonded
far





$$\vec{F}_i = m_i \frac{d^2 \vec{r}_i}{dt^2}$$

The theoretical foundation of MD

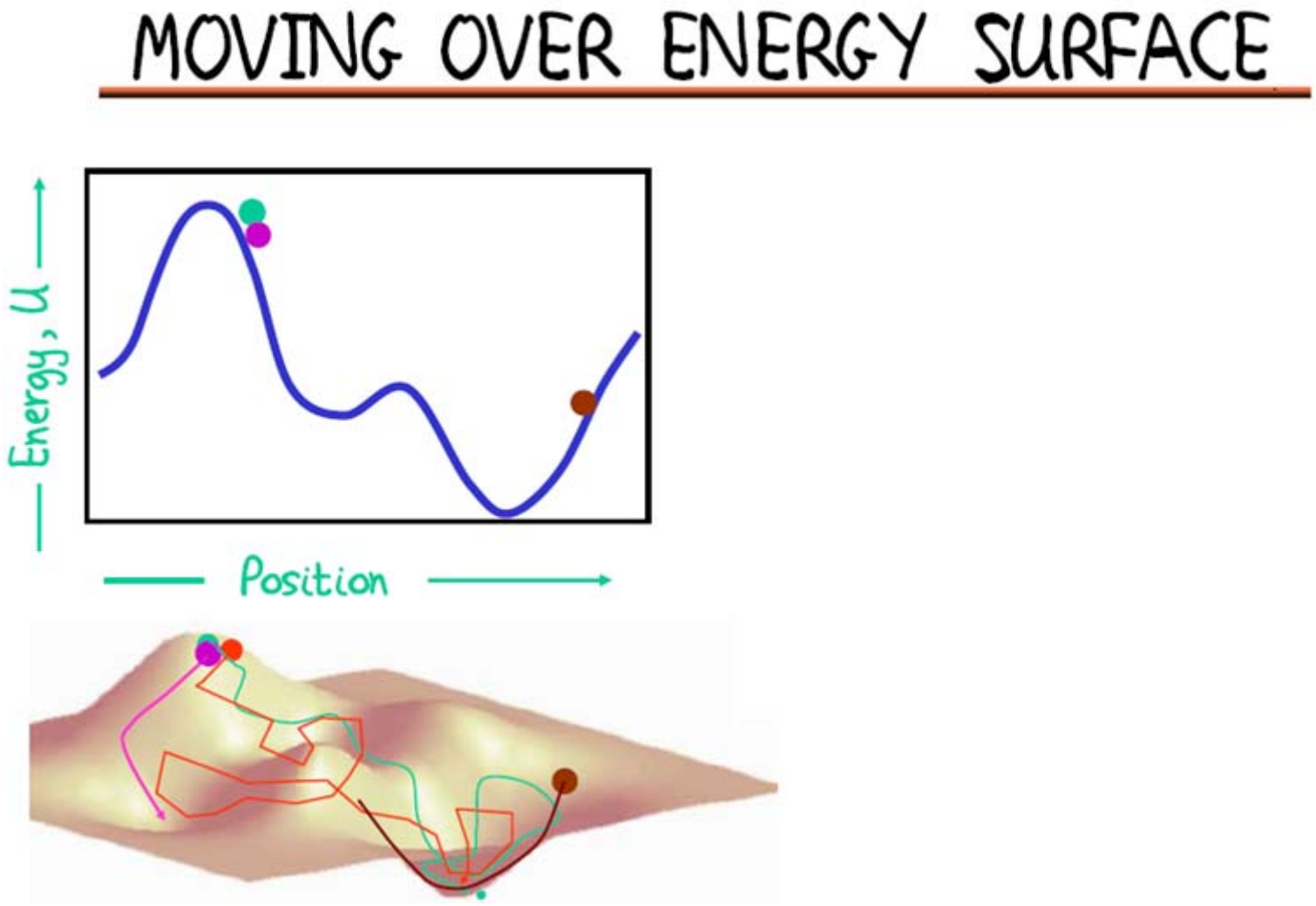
Why MD simulation of a molecular system is possible in theory?

Ergodic theorem (各态历经定理)

A system explores all possible states and can effectively attain thermal equilibrium

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time}$$

Free Energy landscape



The history of MD

Brief history of MD

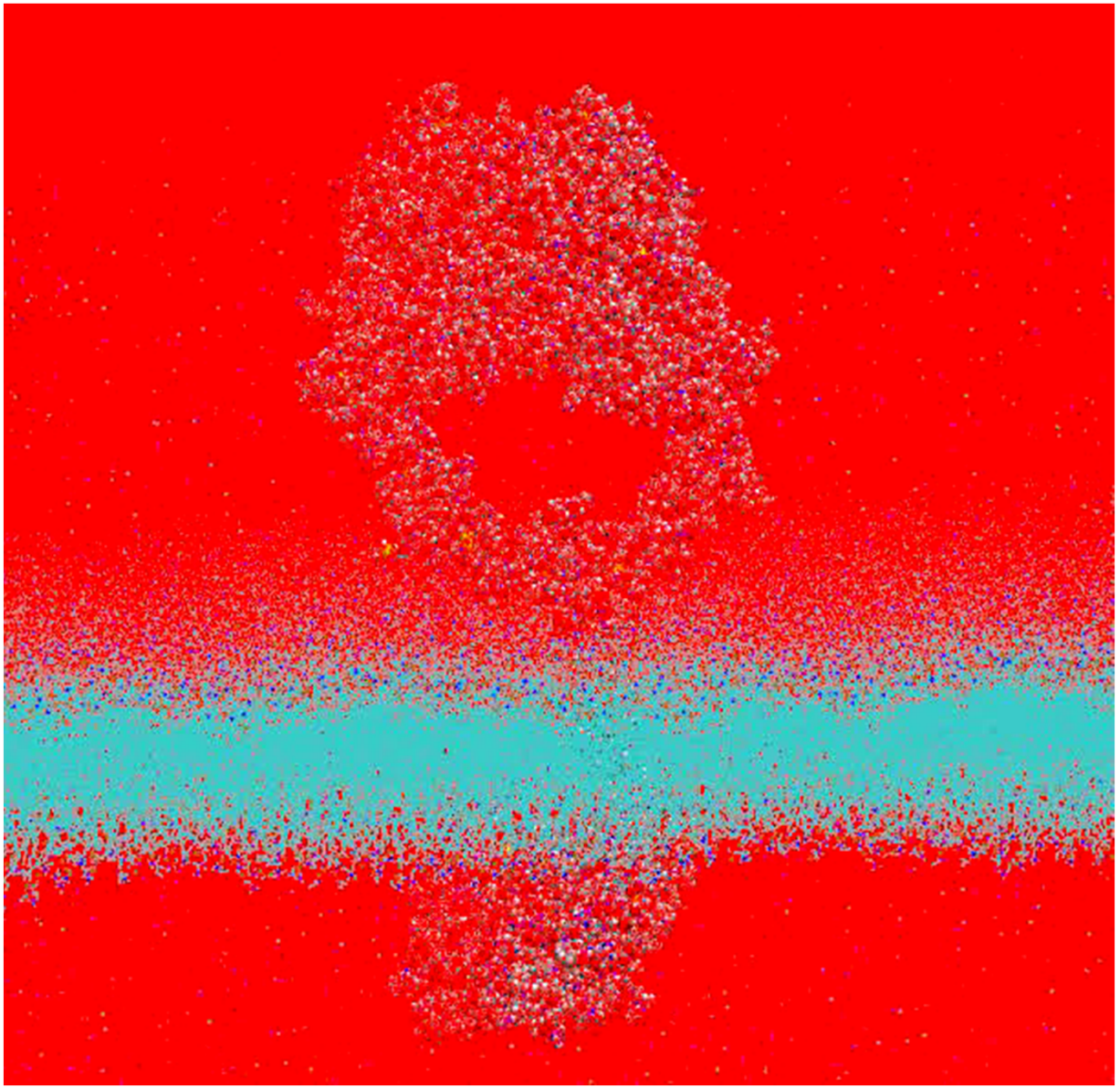
- 1953: Metropolis Monte Carlo (MC) by Metropolis, Rosenbluth, Rosenbluth, Teller & Teller
  - –simulation of a dense liquid of 2D spheres
- 1955: Fermi, Pasta, and Ulam
  - –simulation of anharmonic 1D crystal
- 1956: Alder and Wainwright
  - –molecular dynamics (MD) simulation of hard spheres
- (1958: First X-ray structure of a protein)
- 1960: Vineyard group
  - – Simulation of damaged Cu crystal

Brief history of MD

- 1964: Rahman
  - MD simulation of liquid Ar
- 1969: Barker and Watts
  - Monte Carlo simulation of water
- 1971: Rahman and Stillinger
  - MD simulation of water

Brief history of MD

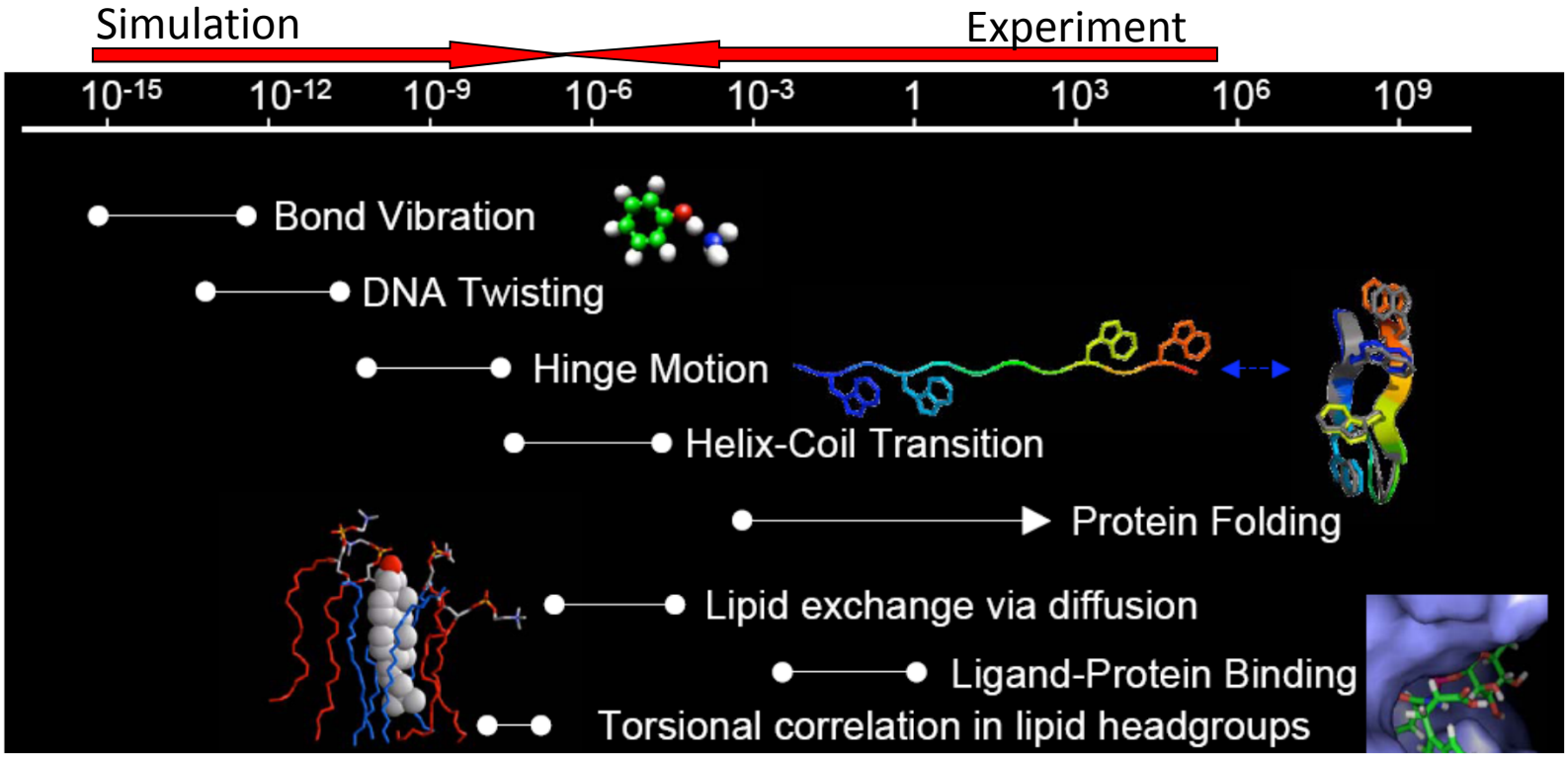
- 1970s: Simulations of small solutes and peptides
- 1977: McCammon, Gelin, Karplus
  - First MD simulation of proteins
- 1980s:
  - Free energy calculations
  - Protein-ligand docking calculations
- 1990s:
  - Continued force field development and sampling techniques
- 1998: Duan and Kollman: 1μs MD simulation of the folding of the Villin headpiece in explicit solvent
- 2009: Anton supercomputer specialized for MD
- 200: Karplus, Warshel, and Levitt--- Nobel Chemistry prize



ANTON

D E Shaw Research

Biomolecular Timescales



Adapted from Suits (IBM), originally from Chan & Dill (1993)

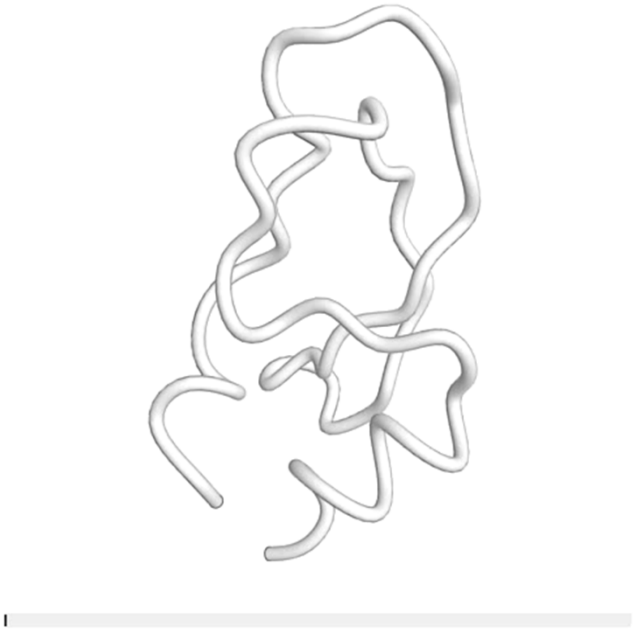
fs – time step during MD

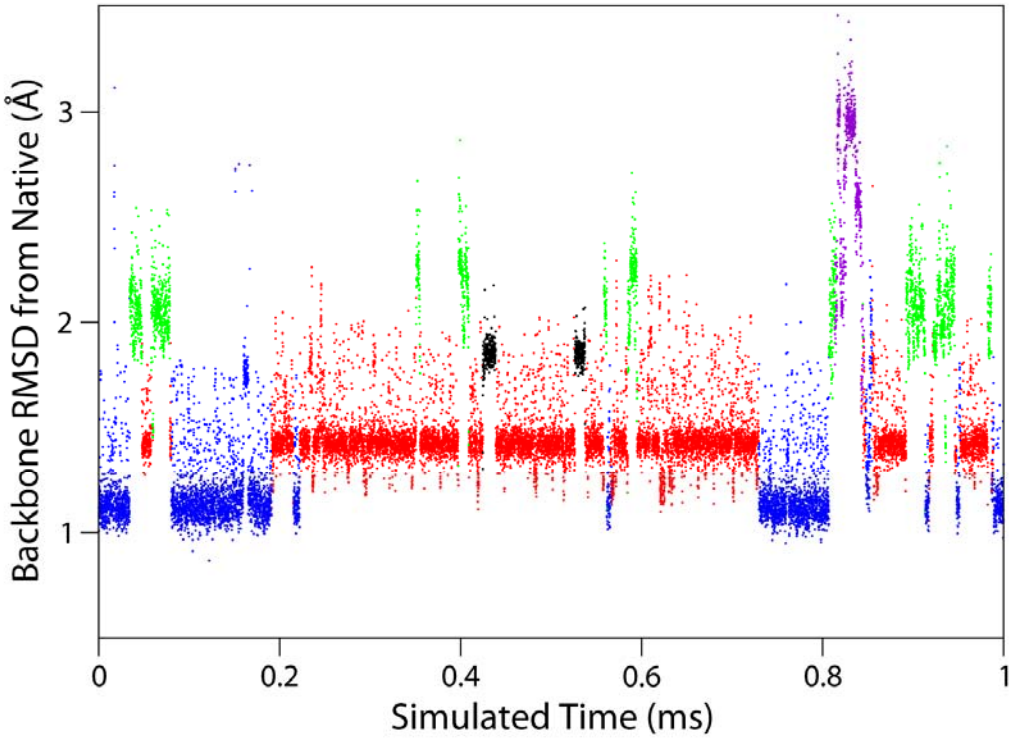
μs - of marginal biological interest (already a billion steps)

ms - this is where things get interesting (a trillion steps)

Understanding trajectories

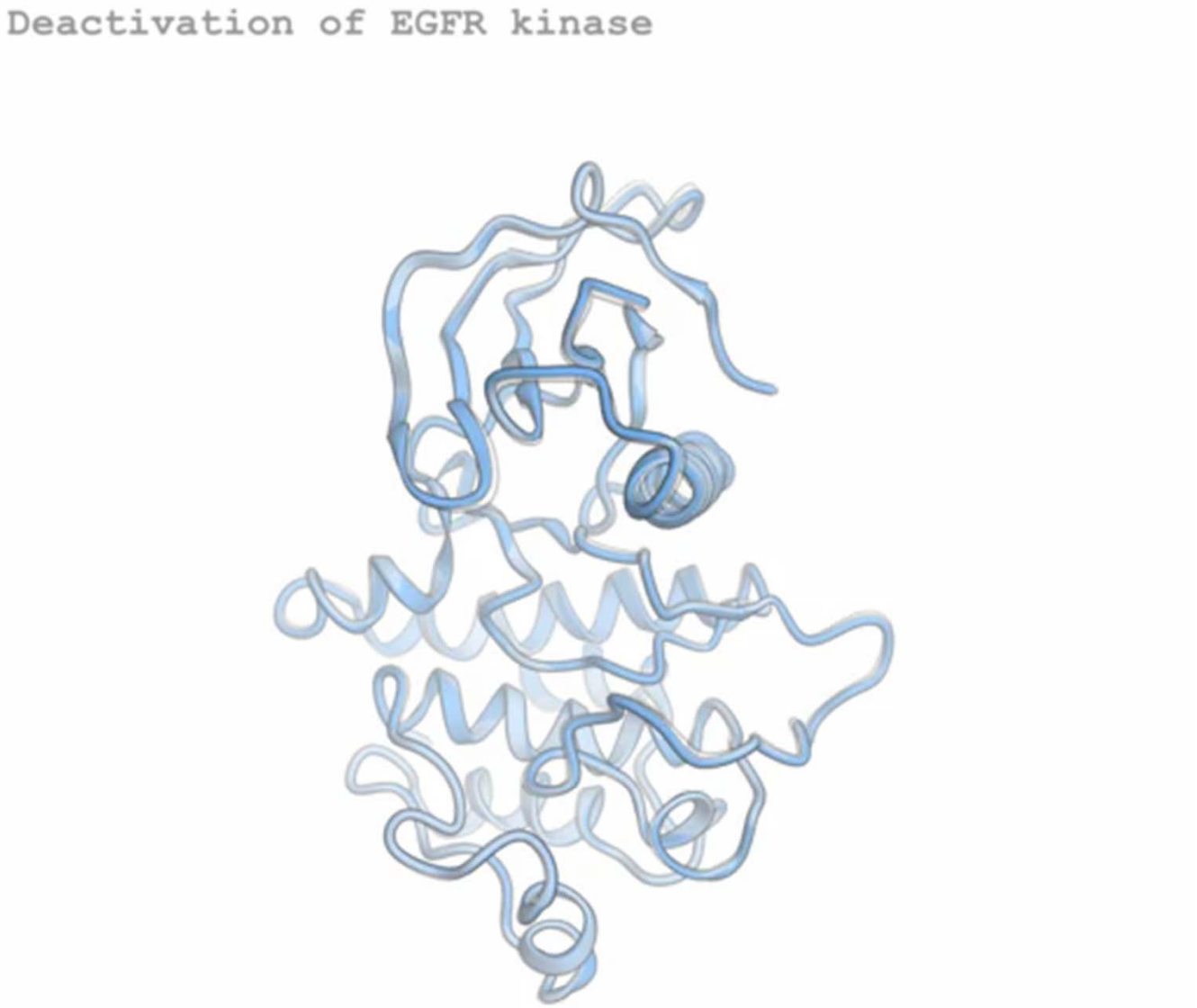
- Proteins tend to have a small number of important states
- Transitions between states are sudden events
- States often live for tens of microseconds to milliseconds





Shaw et al., Science 2010

Identification new conformation



Shan et al., PNAS 2013

Commonly uses MD softwares

CHARMM (Chemistry at HARvard Molecular Mechanics)

AMBER (Assisted Model Building with Energy Refinement)

NAMD (**N**ot (just) **A**nother **M**olecular **D**ynamics program)

GROMACS (GROningen MAchine for Chemical Simulations)

OpenMD

DESMOND

Commonly uses MD forcefields

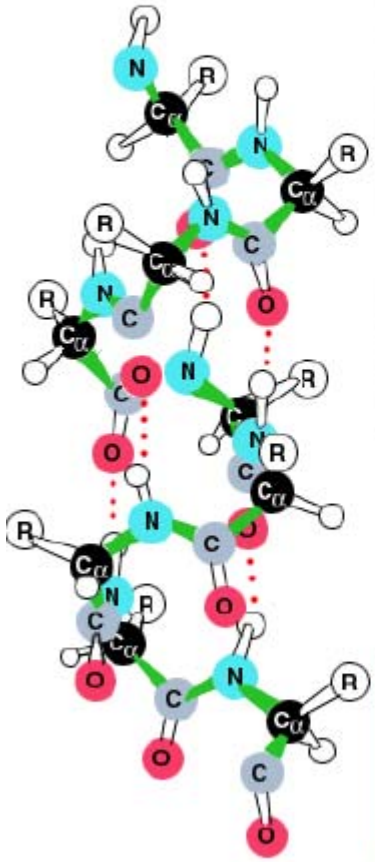
CHARMM (Chemistry at HARvard Molecular Mechanics)

AMBER (Assisted Model Building with Energy Refinement)

OPLS-AA

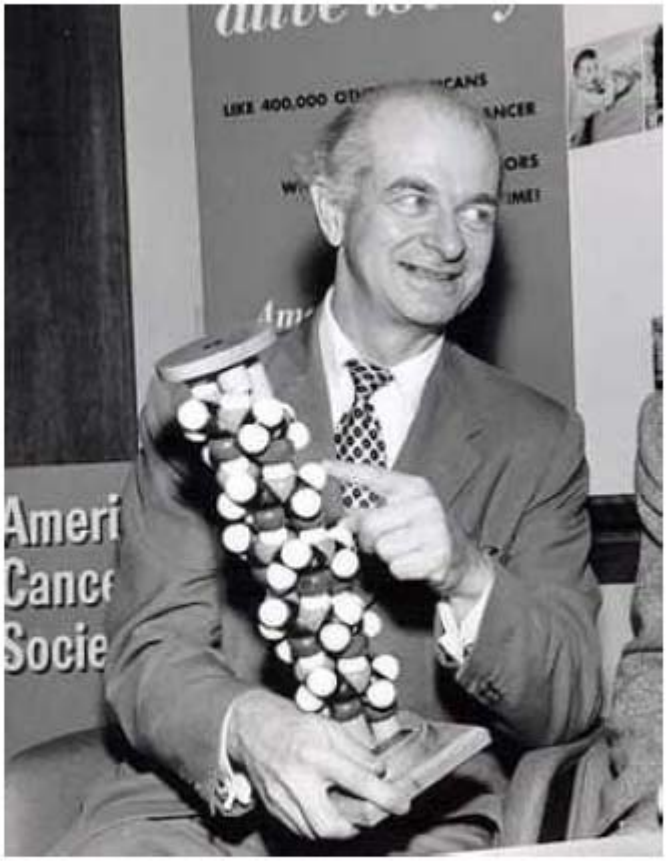
MD as a tool for structural biology

1951: PAULING THE GREAT CHEMIST



1951

The alpha-helix



1901–1994

1953: FRANCIS CRICK

No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

¹ Young, F. B., Gerard, R., and Jevons, W., *Phil. Mag.*, **40**, 169 (1930).

² Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1948).

³ The ATS, W. S., Wood's Hole Papers in Phys. Oceanog. Meteor., **11** (2) (1950).

⁴ Khuan, V. W., *Acta. Met. Astron. Polit.* (Stockholm), **2** (1) (1951).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

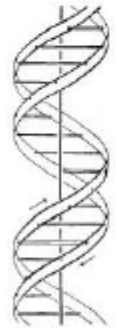
A Structure for Deoxyribose Nucleic A

We wish to suggest a structure for the of deoxyribose nucleic acid (D.N.A.). structure has novel features which are of conside biological interest.

A structure for nucleic acid has already proposed by Pauling and Corey[2]. They kindly their manuscript available to us in advanc publication. Their model consists of three twined chains, with the phosphates near the axis, and the bases on the outside. In our opi this structure is unsatisfactory for two reas (1) We believe that the material which gives X-ray diagrams is the salt, not the free acid. Wit the acidic hydrogen atoms it is not clear what f would hold the structure together, especially s negatively charged phosphates near the axis repel each other. (2) Some of the van der W distances appear to be too small.

Another three-chain structure has also been gested by Fraser (in the press). In his model phosphates are on the outside and the bases on inside, linked together by hydrogen bonds. structure as described is rather ill-defined, and this reason we shall not consi on it.

We wish to put forwa radically different structure the salt of deoxyribose n acid. This structure has helical chains each coiled r the same axis (see diagram). have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[3] model No. 1; that is, the bases are on the inside of the helix and the phosph the outside. The conn of the sugar and the near it is close to P 'standard configuratio sugar being roughly p cular to the attached base

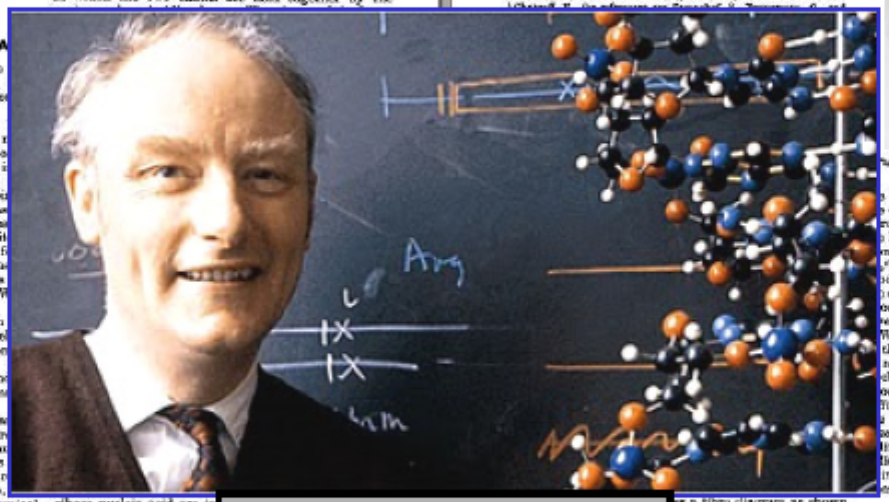


This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the



1916–2004

ribose nucleic acid are in of our structure. So far compatible with the exp be regarded as improve against more exact resul in the following comman of the details of the resi devised our structure, w entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published

738 NATURE April 25, 1953 Vol. 171

King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge. April 2.

¹ Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).

² Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).

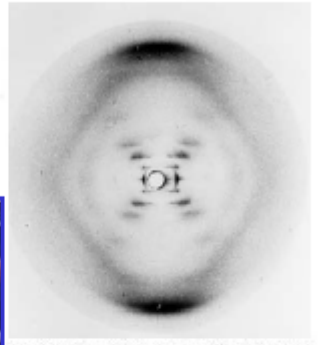
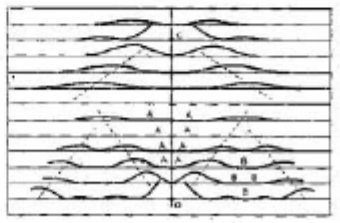


Fig. 1. Fibre diagram of deoxyribose nucleic acid from *B. sub.* Fibre axis vertical

innermost maxima of each Bessel function and origin. The angle this line makes with the equator roughly equal to the angle between an element of helix and the helix axis. If a unit repeats n times ng the helix there will be a meridional reflexion /*n* on the *n*th layer line. The helical configuration duces side-bands on this fundamental frequency, effect being to reproduce the intensity distribution out the origin around the new origin, on the *n*th er line, corresponding to *C* in Fig. 3.

We will now briefly analyse in physical terms some the effects of the shape and size of the repeat unit nucleotide on the diffraction pattern. First, if the cleotide consists of a unit having circular symmetry out an axis parallel to the helix axis, the whole fraction pattern is modified by the form factor of a nucleotide. Second, if the nucleotide consists of series of points on a radius at right angles to the lix axis, the phases of radiation scattered by the lices of different diameter passing through each nt are the same. Summation of the corresponding ssel functions gives reinforcement for the inner-



3. Diffraction pattern of system of radii corresponding to ture of deoxyribose nucleic acid. The crosses of Bessel tions are joined about 0 on the equator and on the first, ird, third and fifth layer lines for half of the nucleotide radii 3 A. diameter and remainder distributed along a radius, the s as a circle radius being proportional to the radius. About h the fourth layer line at 0·17 reflexions are plotted on an outer diameter of 12 A.

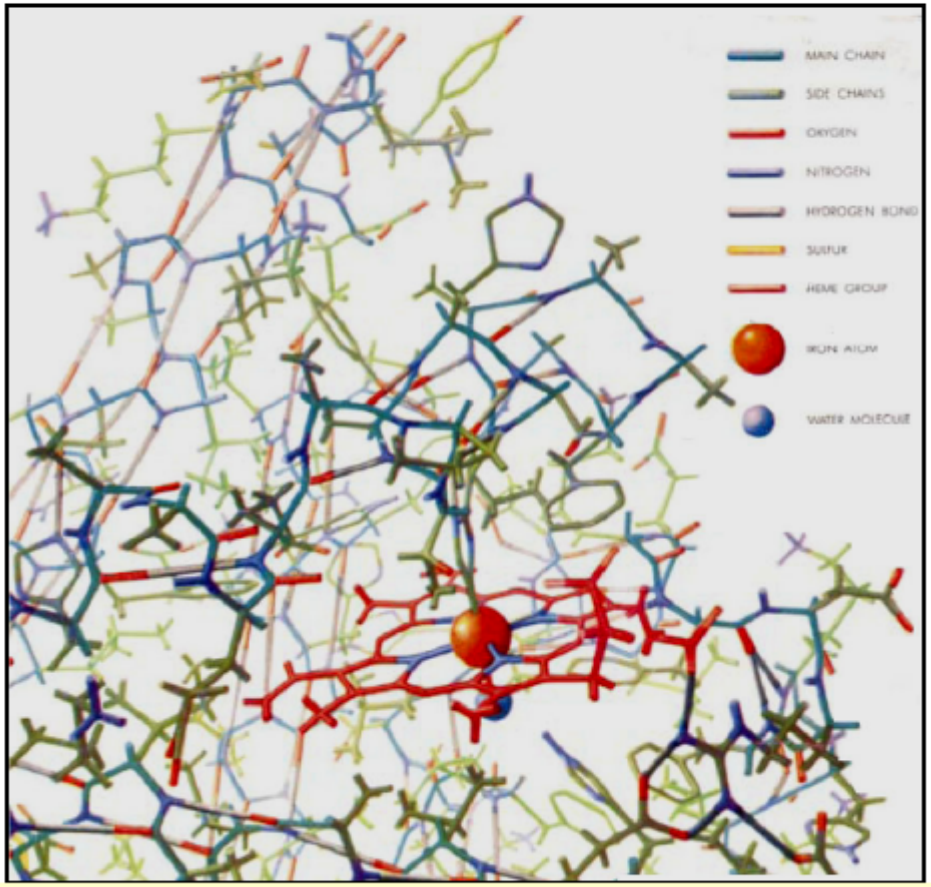
DNA Model and Experiment

1959: KENDREW AND MYOGLOBIN

1917–1997

First protein X-ray structure.

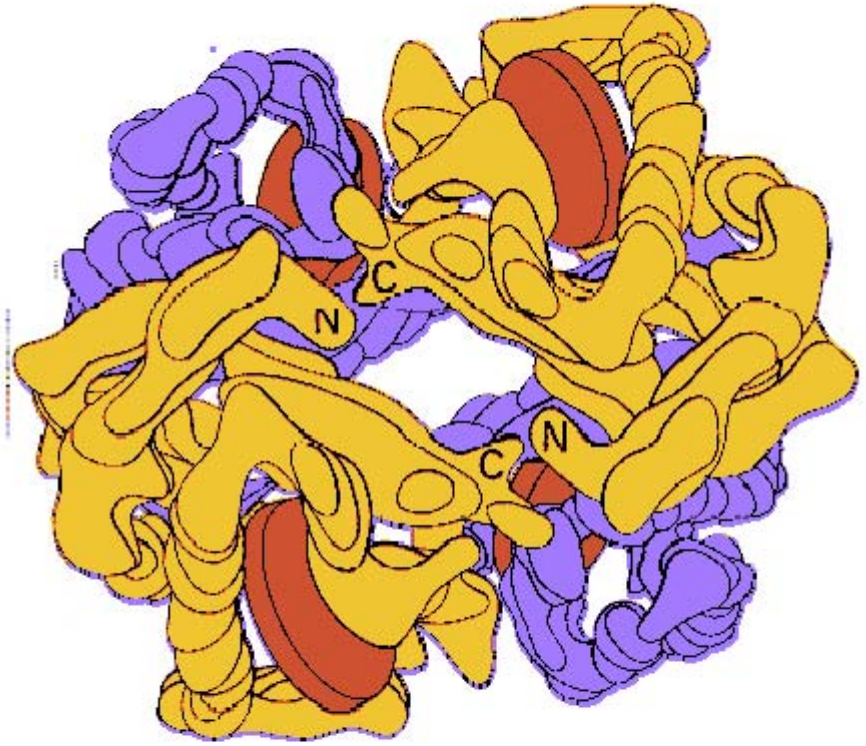
Scientific American 1961



Painted by artist Irving Geis

1962: PERUTZ AND HEMOGLOBIN

1914–2002

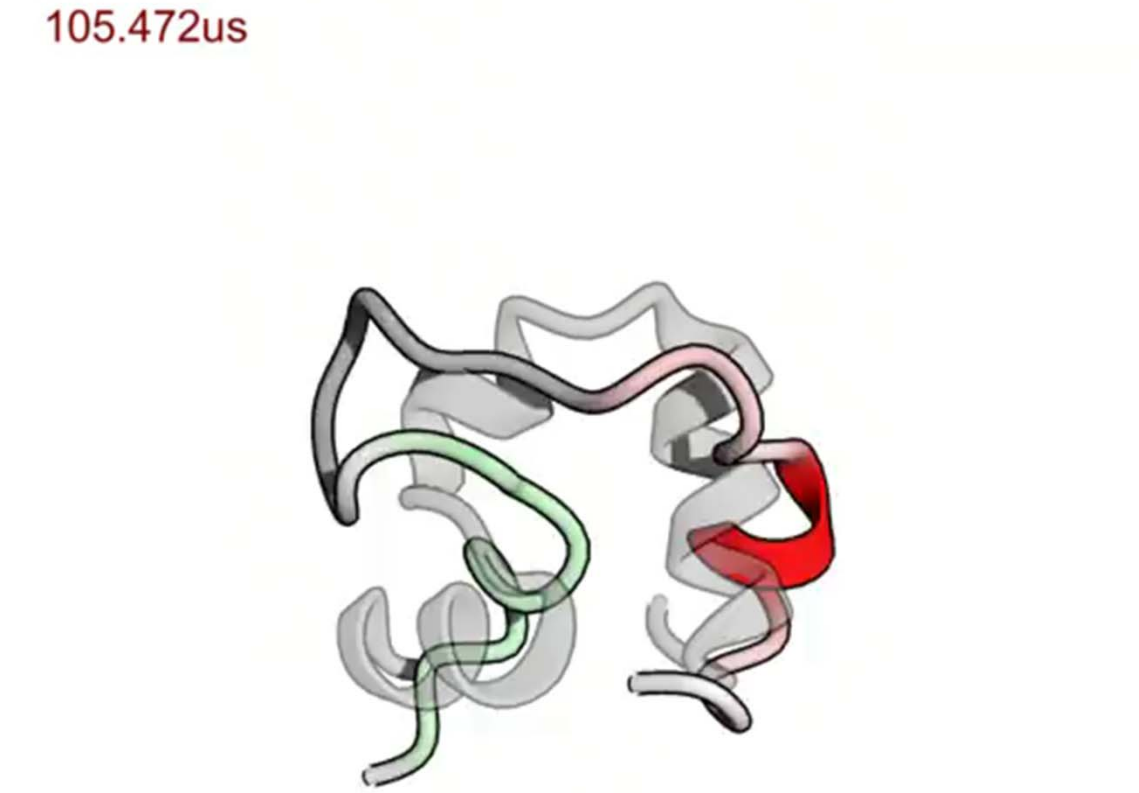


The REAL HERO of structural biology.

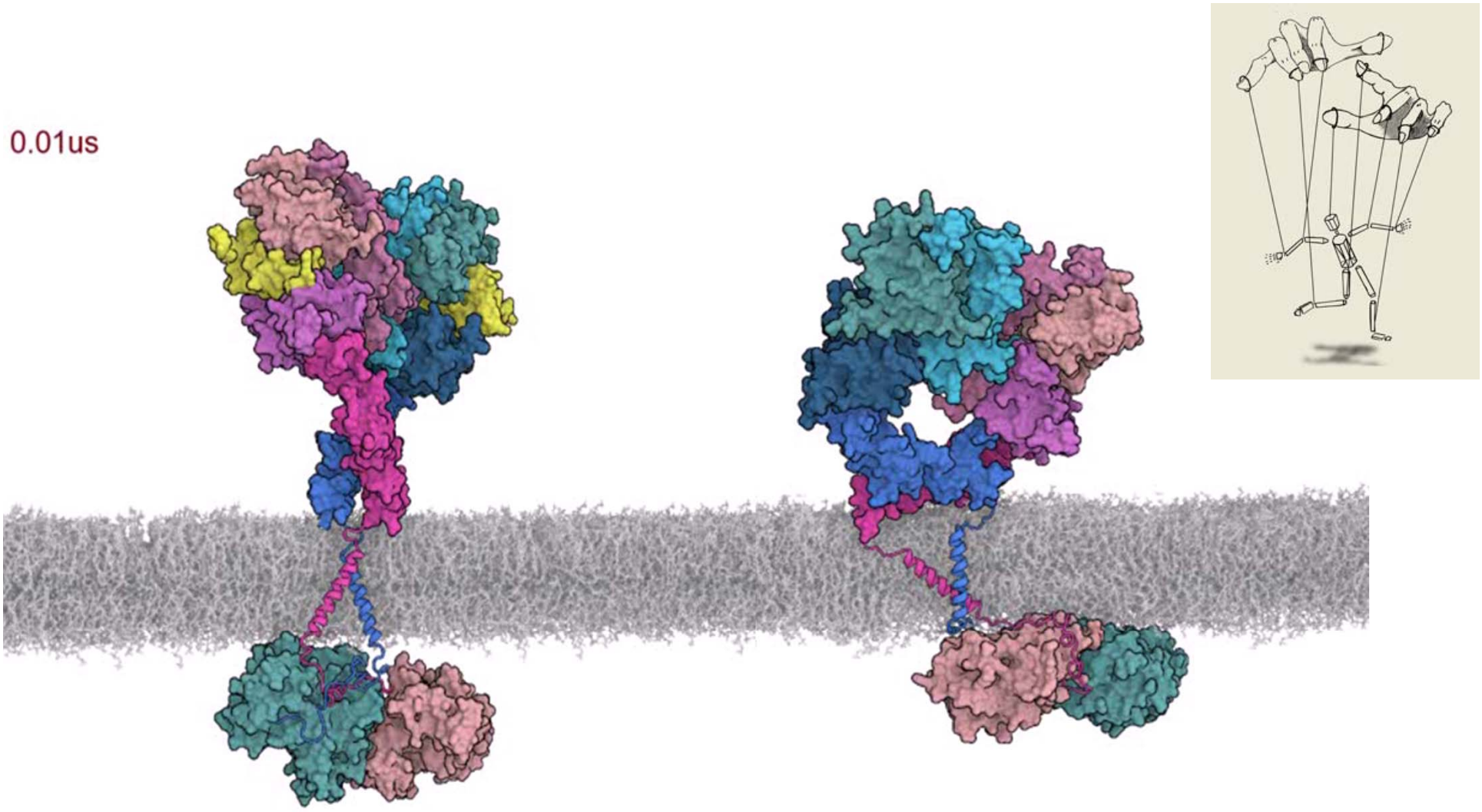
Why MD simulation

- MD simulations provide a molecular level picture of structure and dynamics of biological systems → property/structure relationships
- Experiments often do not provide the molecular level information available from simulations
- Simulators and experimentalists can have a synergistic relationship, leading to new insights into materials properties

Protein folding



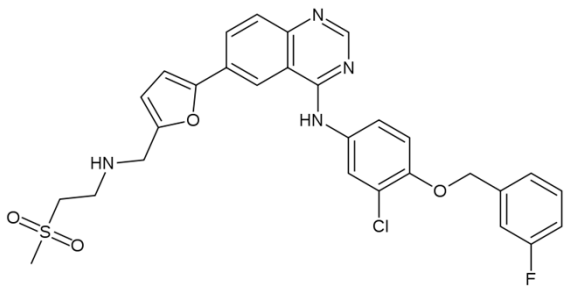
Piana et al., **PNAS** (2012)

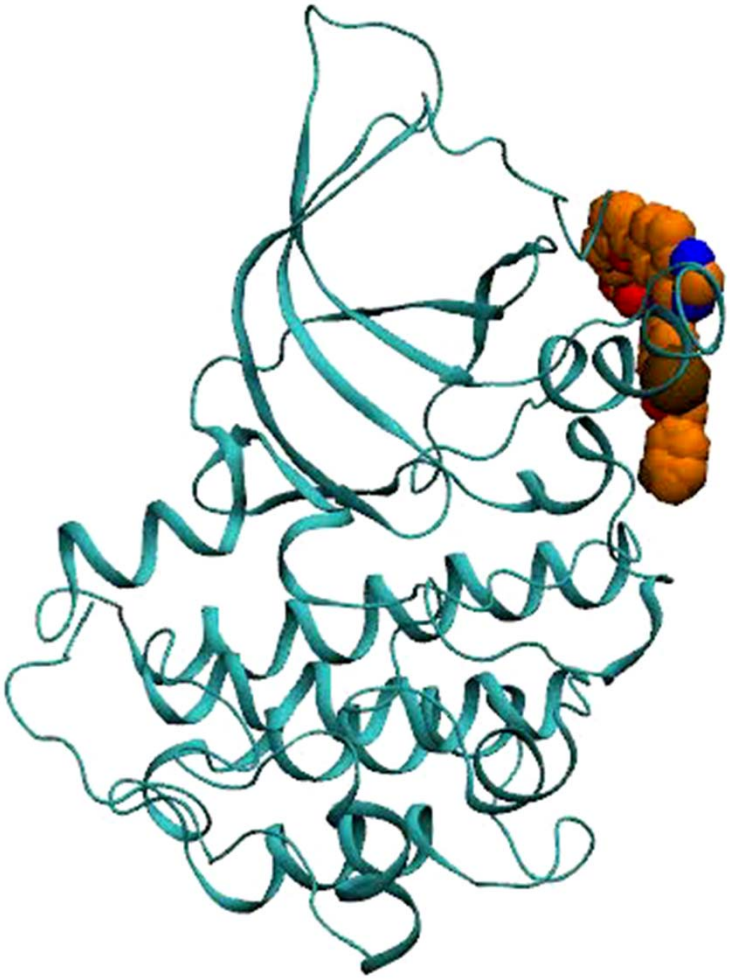


Active Inactive

Arkhipov et al., **Cell** (2013)
Endres et al., **Cell** (2103)

Lapatinib binding to EGFR kinase

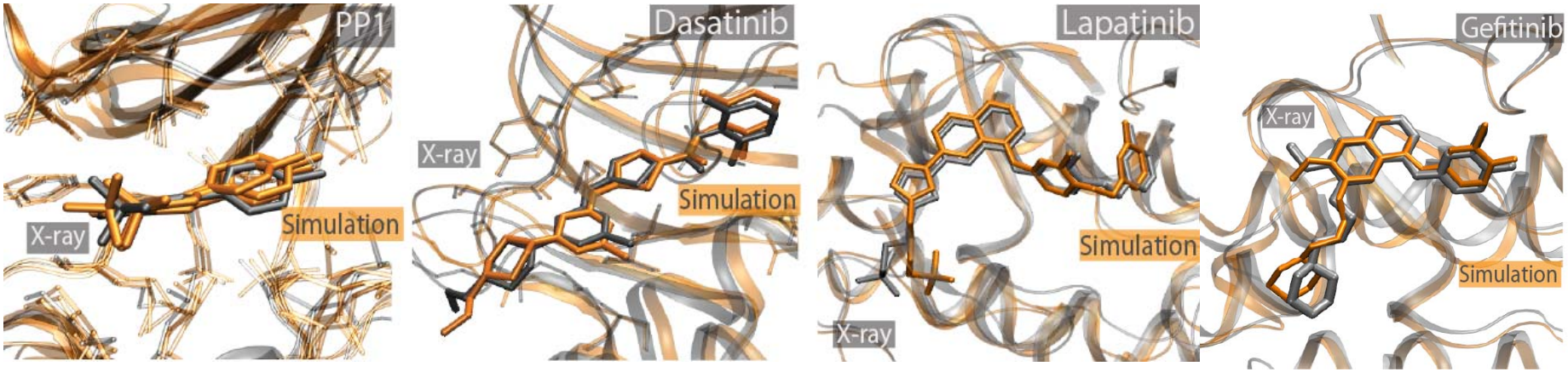




Shan* et. al., ***Cell*** 2012

Shan* et. al., ***JACS*** 2011

Simulation binding pose superimposed to the Xtal poses



Virtually identical to crystal structures

Order-of-magnitude correct kinetics

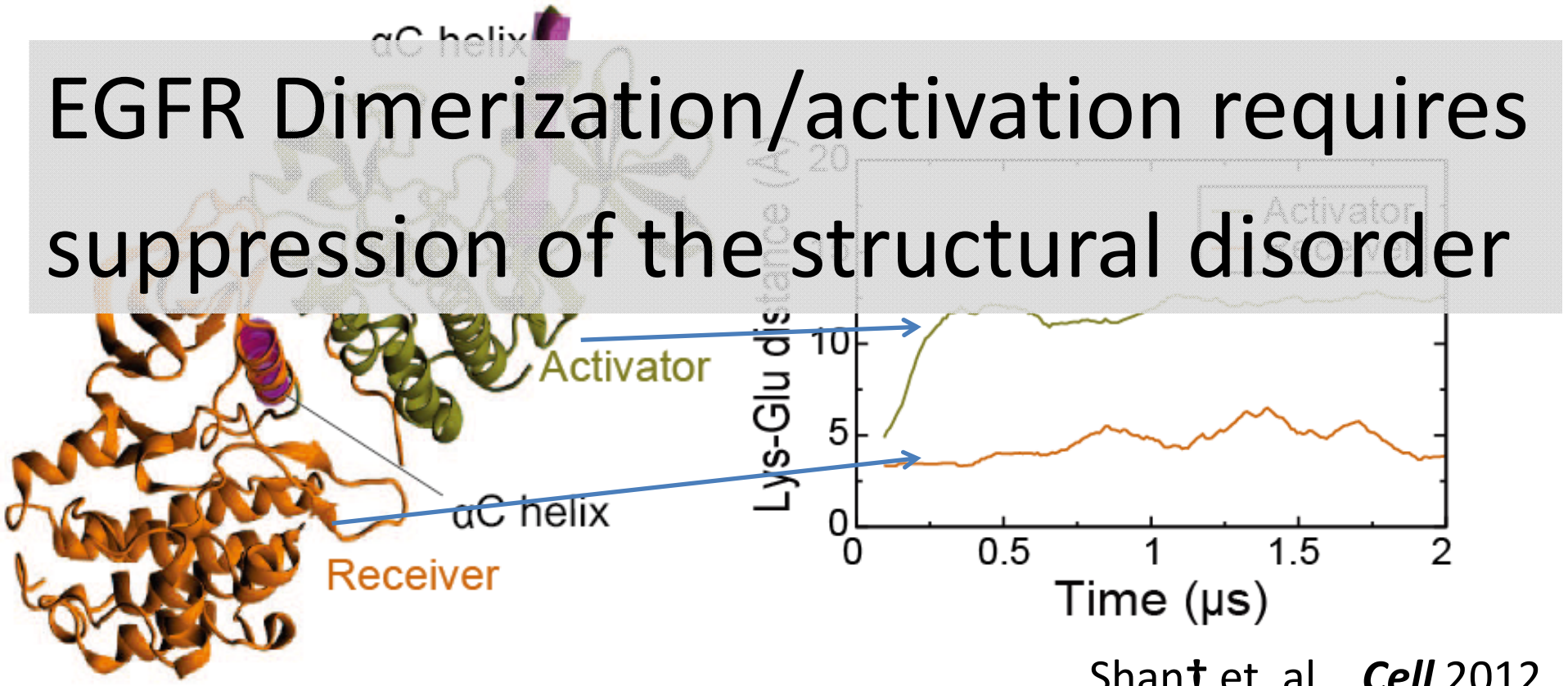
The most important thing to remember:

By using MD simulation we should make predictions and guide experiments, not only to explain what is already known.

Think like a biologist using simulation

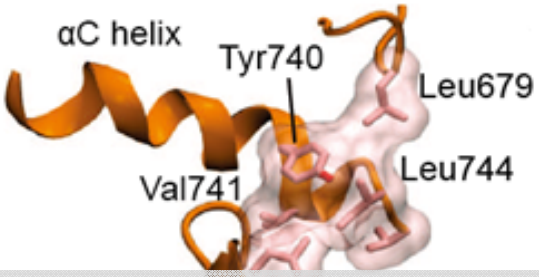
A Safety Measure Encoded in the Kinase Domain

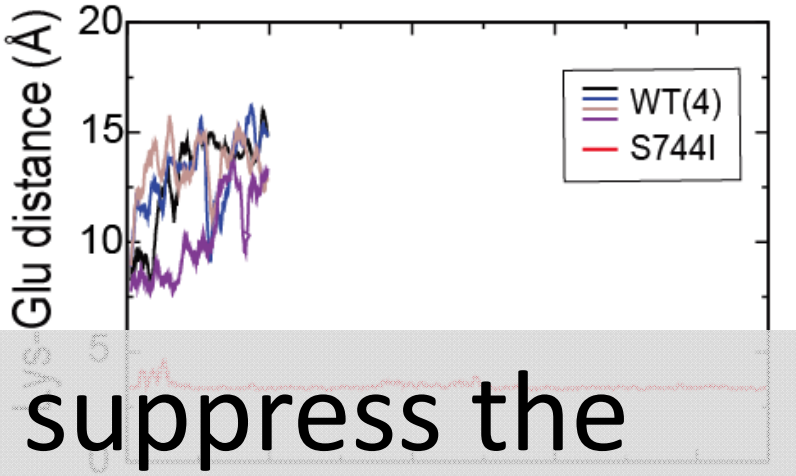
EGFR Dimerization/activation requires suppression of the structural disorder



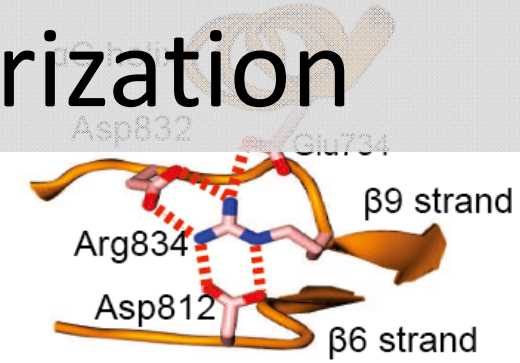
Shan† et. al., ***Cell*** 2012

Cancer Mutations Suppress Disorder

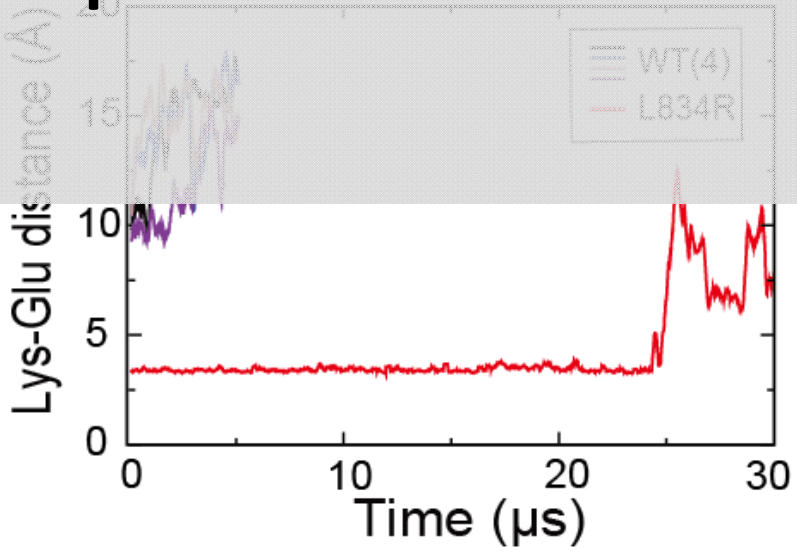




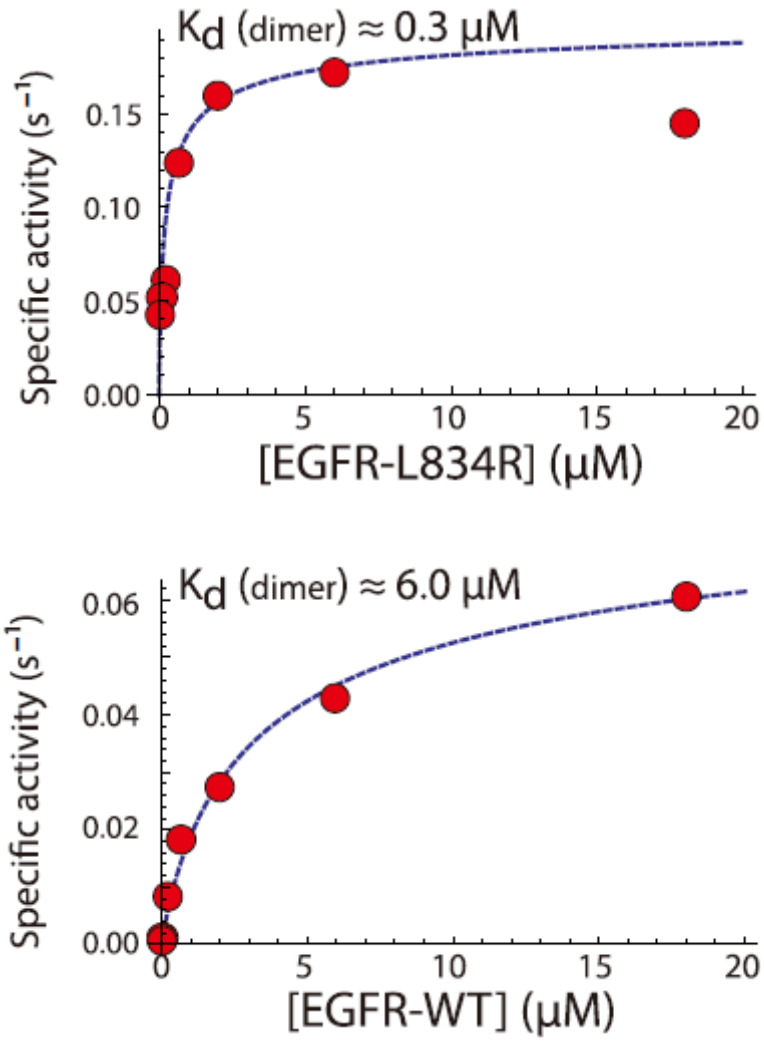
Cancer mutations suppress the disorder and predispose EGFR for dimerization



L834R mutation



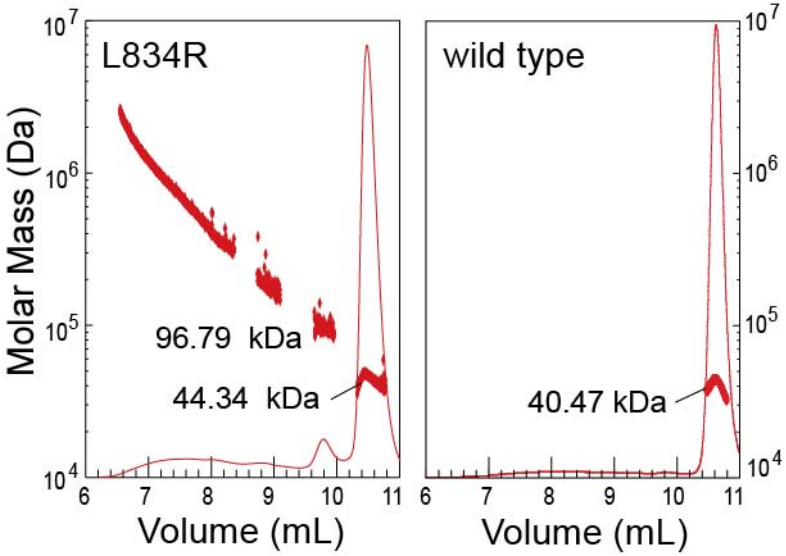
Higher Dimerization Rate and Activity

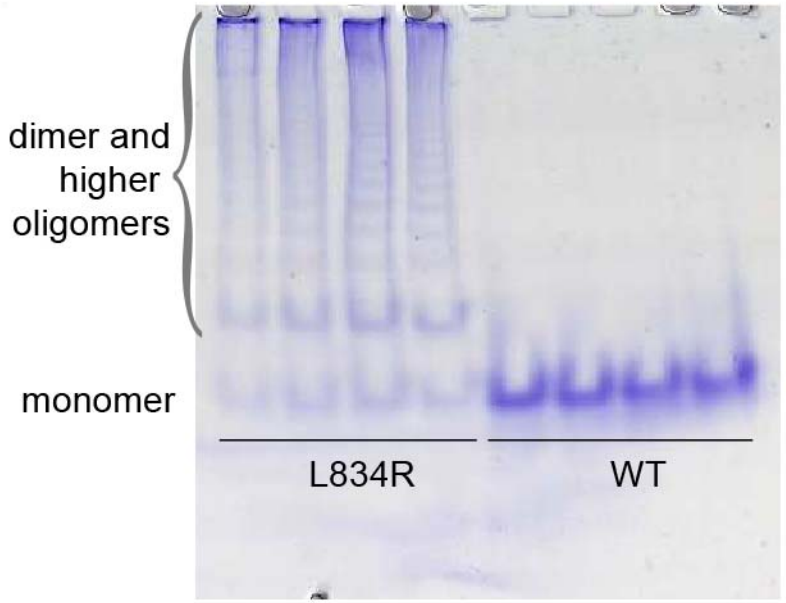


Greater difference at low density

Dimerization dependence remains

Indeed, the mutants are predisposed to dimerization…





Light scattering

Native Gel

Zhang & Kuriyan

Currently, MD is more useful as a qualitative tool than a quantitative one

Qualitative understanding can be powerful

In principle, MD can calculate free energy and kinetic rates

FEP—Free energy perturbation method
TI – Thermodynamic Integration
many other more empirical methods, such as MM-GBSA/PBSA

For the calculation of conformational energy
Umbrella sampling
Metadynamics

Many ideas to speed up MD

Replica exchange
Metadynamics
Accelerated molecular dynamics
Parallel MD simulations/Markov analysis

Molecular Dynamics is already an widely used tool in today's structural biology and a household tool for tomorrow's