Introduction to Computational Biophysics

Course Code: 04710262

Spring Semester, 2018

What will you learn

- Basic ideas and concepts in computational biophysics
- Basic techniques and tools and THEORY behind
  (no “black box”)
- hands-on experience of application of computational methods on simple systems
- Think computationally and quantitatively

Why you learn here is important

- Explain better your experimental results and understand the mechanisms
- Prediction and guide experiments
- Experimental and computation collaborative works required more often than before
- Industry: computer-aided drug discovery

- 18.03.02 Postdoc and PhD student openings in Computer-Aided Drug Design center
- 18.02.28 Molecular Modeling Scientist, Servier Monde, France, Croissy sur Seine
- 18.02.28 Principal Bioinformatician - MI/18/15 - Drug Discovery Unit
- 18.02.28 Senior Computational Chemist MI/18/10 - Drug Discovery Unit
- 18.02.28 Principal Computational Chemist - Drug Discovery Unit
- 18.02.27 Two computational roles at AstraZeneca, Sweden or UK
- 18.02.26 Postdoctoral Position Computational Materials Design
- 18.02.26 Research Leader, Drug Discovery Biologist, NYC, Schrödinger
- 18.02.26 Senior Applications Scientist, Mannheim, Germany or Cambridge, UK, Schrödinger
- 18.02.26 Force Field Development at D. E. Shaw Research
- 18.02.25 POSTDOCTORAL RESEARCH OPPORTUNITY IN COMPUTATIONAL BIOPHYSICS/CHEMISTRY IN NEW YORK, NEW YORK, USA
- 18.02.24 PhD position in Computational Materials Chemistry (University of Liverpool, UK)
- 18.02.22 PhD Studentship in Computational Physical Organic Chemistry (School of Chemistry, University of Edinburgh, UK)
- 18.02.22 Postdoctoral Position in Computational Biophysics/Chemistry, UCR, Los Angeles area
- 18.02.22 PhD Student position in Computational Biochemistry in the Department of Chemistry at Michigan Technological University
- 18.02.22 Three month fellowship in Computational Biophysics
- 18.02.21 Senior Computational Chemist (CADD), Atomwise Inc, San-Francisco
- 18.02.21 Postdoc in Simulation of Optical Properties of Organic Condensed Phase Systems at the University of Calfornia Merced
- 18.02.21 Positions for a post doc and a masters stundent in Computational Chemistry
- 18.02.20 Postdoc opening in computational biophysics in Paris
- 18.02.19 Applications Scientist, NYC, Schrödinger
- 18.02.19 Scientific Software Developer, NYC, Schrödinger
- 18.02.19 Scientific Software Developer - Indianapolis, IN
- 18.02.18 Research/PostDoctoral Associate position in Computational Biophysics and Chemistry
- 18.02.15 Postdoc in Gray Lab (Johns Hopkins; Rosetta Commons) - Protein Interactions
- 18.02.15 Postdoctoral position in Computational Biophysics/Molecular Dynamics of multi domain proteins - University of Lyon,France
- 18.02.13 Principal Scientist in Computational Chemistry - 181247, Boehringer Ingelheim, Germany
- 18.02.13 Postdoc in Simulations of Organic Materials for Battery Applications, DTU Energy, Denmark
- 18.02.12 Postdoc Opportunity in Computational Biology and Computer-Aided Drug Design
- 18.02.08 Senior Scientist, Computational Chemistry, Repare Therapeutics, Montreal, Quebec, Canada
- 18.02.08 Postdoctoral position in Computational Biophysics/Biochemistry at the Biomedical Research Foundation, Academy of Athens, Greece
- 18.02.05 Research Fellow, Queen's College, Belfast
- 18.02.02 Post Doc in Computational Chemisty - 181093, Boehringer Ingelheim Pharma, Biberach, Germany

Prerequisite

- Freshman calculus, phys, and chem completed
- A basic knowledge of programming
- Willing to think quantitatively in biology.
- Willing to use computer in learning/working.

Important information

- Monday 14:00-17:00, E104
- A laptop may be needed in classes
- VMD and SSHshell may be needed to installed
- Office, F409, Hours by appointments
  hanw@pkusz.edu.cn
- TA: Ms. Qinsi Xiong
  Phone: 13662675174
  Email:kingcy1992@163.com

Textbook :

- Lecture notes, posted after of class (usually one day after)
- Frenkel, and Smit, 《Understanding Molecular Simulations》 2nd edition，2002

Optional:

- Thomas Creighton 《Proteins: Structures and Molecular properties》
- D. C. Rapaport《The Art of Molecular Dynamics Simulation》 2nd Edition
- https://learnpythonthehardway.org/book/

Grading

Assignments: 40% (Monthly)

- Late return penalty:
- deduction by 10% for each day to the assignment to be returned

Final exam: 40% (take-home 20% + in-class 20%)

In-class assessment: 20%

Outline

- What is Biophysics?
- What is Computational Biophysics?
- What does it seek to address and what can it do?

Physics and Biology

- Biology studies life in its variety and complexity:
  structure, function, growth, evolution

- Physics studies fundamental laws and principles of Nature:
  structure, interaction, motion of matter

- Life is matter and physical laws also apply

Physics and Biology

Erwin Schrödinger, austrian

12 Aug 1887 – 4 Jan 1961

Time-dependent Schrödinger equation *(general)*

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = \hat{H}\Psi(\mathbf{r},t)$$

Nobel in physics

"What is Life?"

--"that the subject-matter was a difficult one and that the lectures could not be termed popular, even though the physicist's most dreaded weapon, mathematical deduction, would hardly be utilized."

1943, Trinity College, Dublin

Physics and Biology

Erwin Schrödinger, austrian

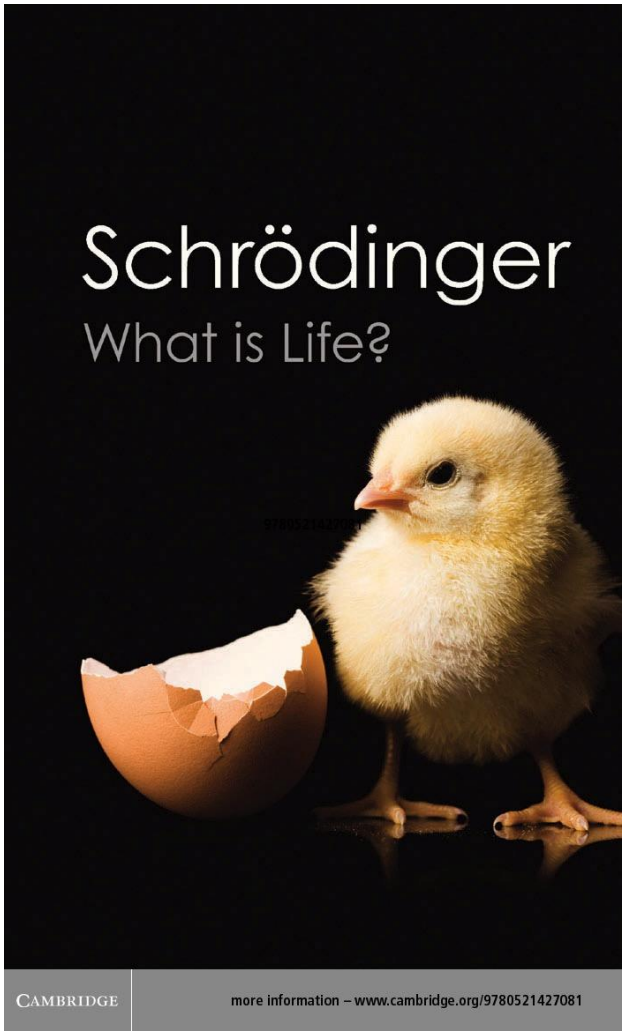
12 Aug 1887 – 4 Jan 1961

Time-dependent Schrödinger equation *(general)*

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = \hat{H}\Psi(\mathbf{r},t)$$

Nobel in physics

1944, Cambridge



What Is Life? The Physical Aspect of the Living Matters

- Laws on large scale due to chaos on small scale, “order-from-disorder”
  Diffusion

- Function of systems depend on temperature
  “If heated, a clock ceases to work, because it melts”

- Genetic materials must not be like crystal to store information
  Inspiring discovery of DNA double helix by Watson and Crick

Biophysics studies biology with physics

- Biophysics looks for principles that describe patterns in life with math and physics to gain insights. If the principles are powerful, they make detailed predictions that can be tested.

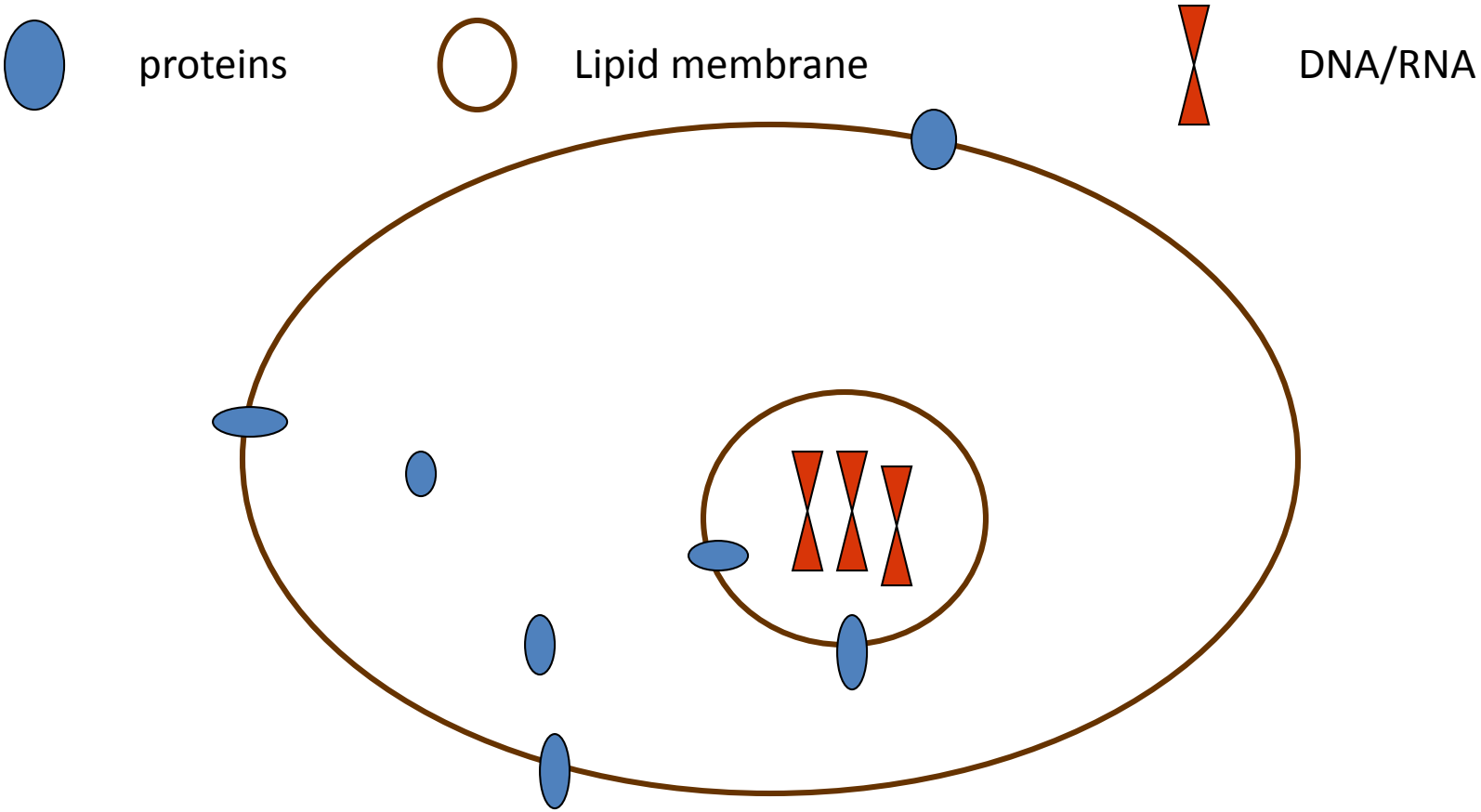
- Biophysicists study life at every level, from atoms and molecules to cells, organisms, and environments.

From Biophysical Society

An overview of biomolecules

- Living organisms are more ordered than their surroundings.
- So the first task is to maintain a separation between inside and outside.
- The second task is to spend energy to keep things in order.
- *The functions of life are to facilitate the acquisition and expenditure of energy.*

A simple model of a cell



Tasks of a living cell

- Gather energy from surroundings.
- Use energy to maintain inside/outside distinction.
- Use extra energy to reproduce.
- Develop strategies for being efficient at their tasks: developing ways to move around; developing signaling capabilities; developing ways for energy capture; developing ways of reproduction.

E. coli: a model system for chemotaxis

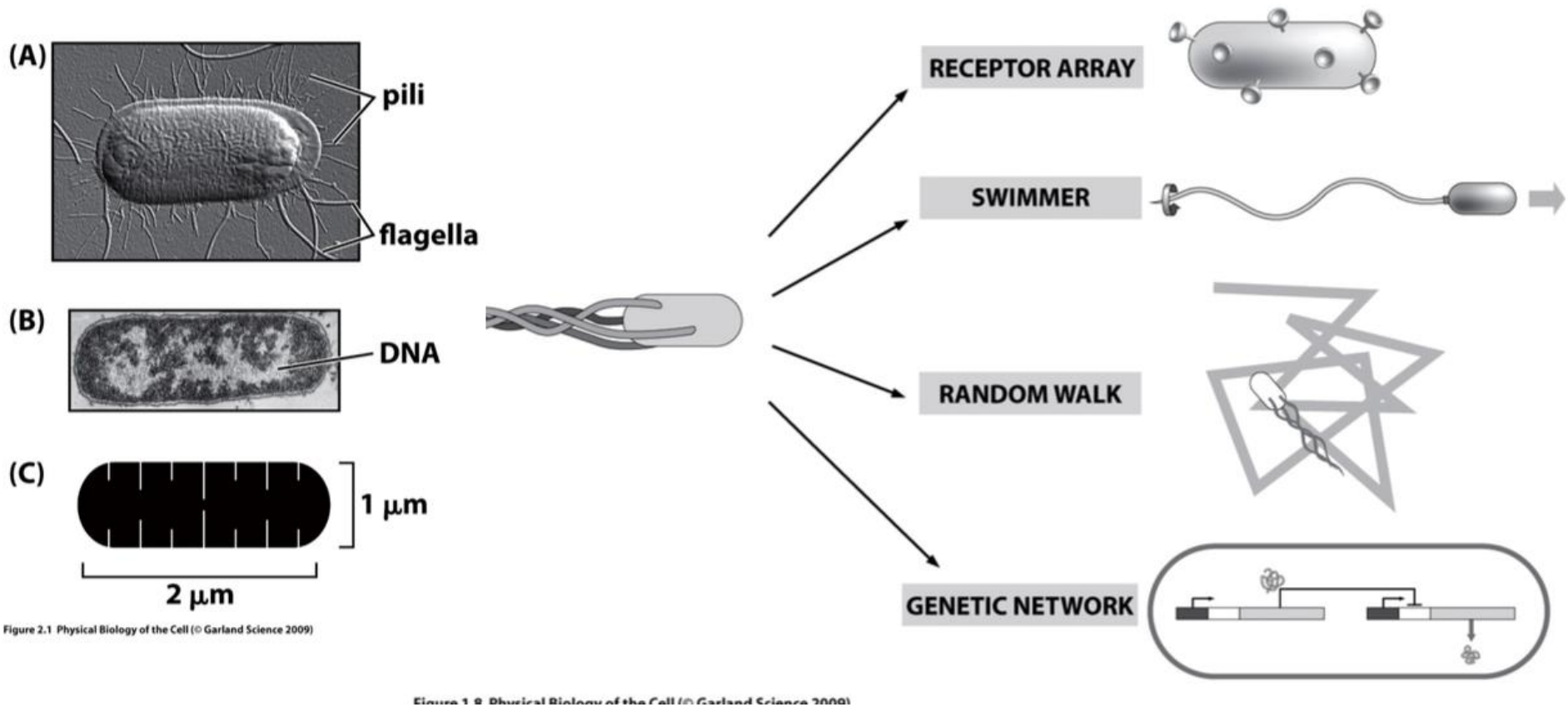
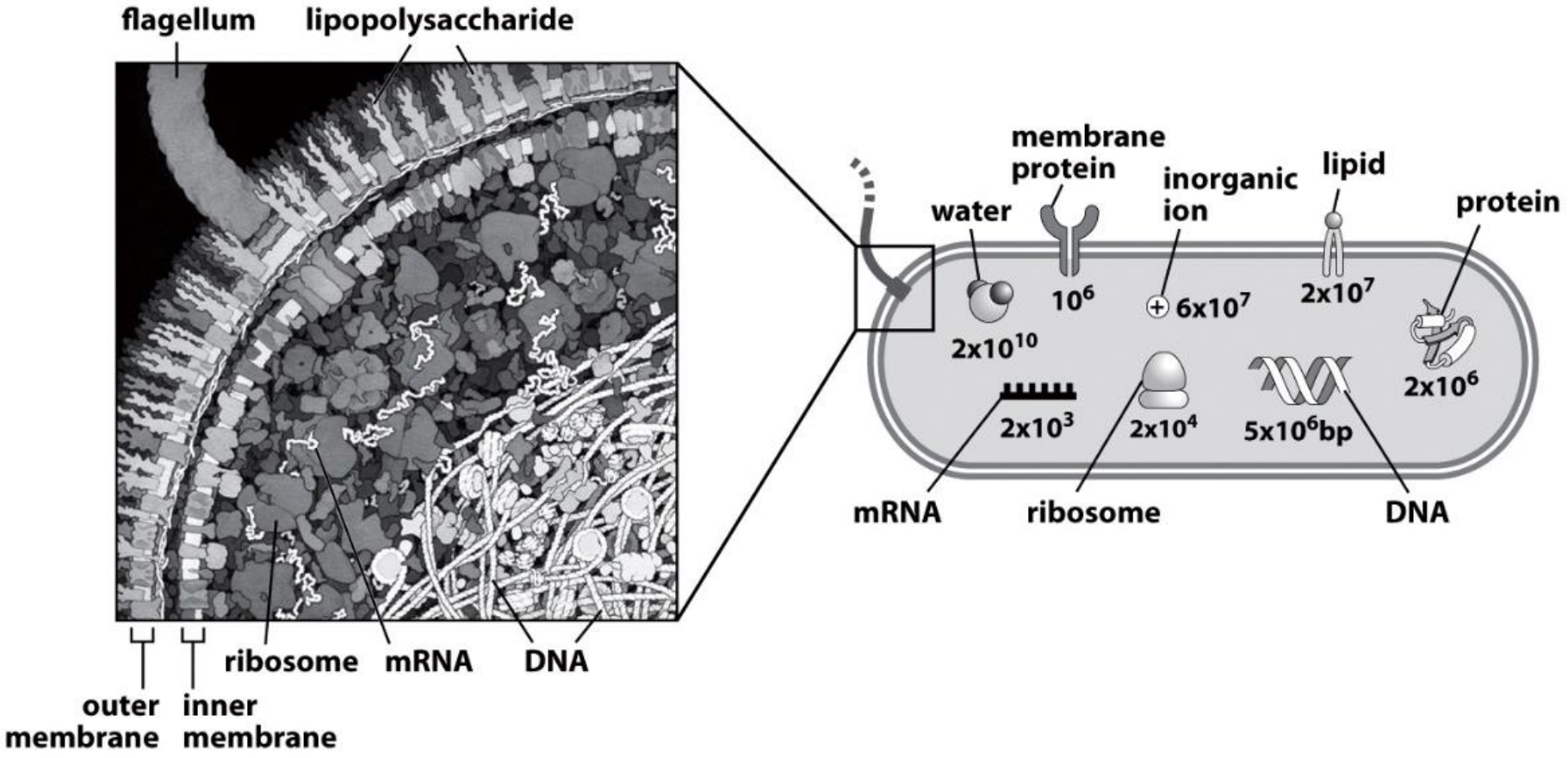


Figure 2.1 Physical Biology of the Cell (© Garland Science 2009)

Figure 1.8 Physical Biology of the Cell (© Garland Science 2009)

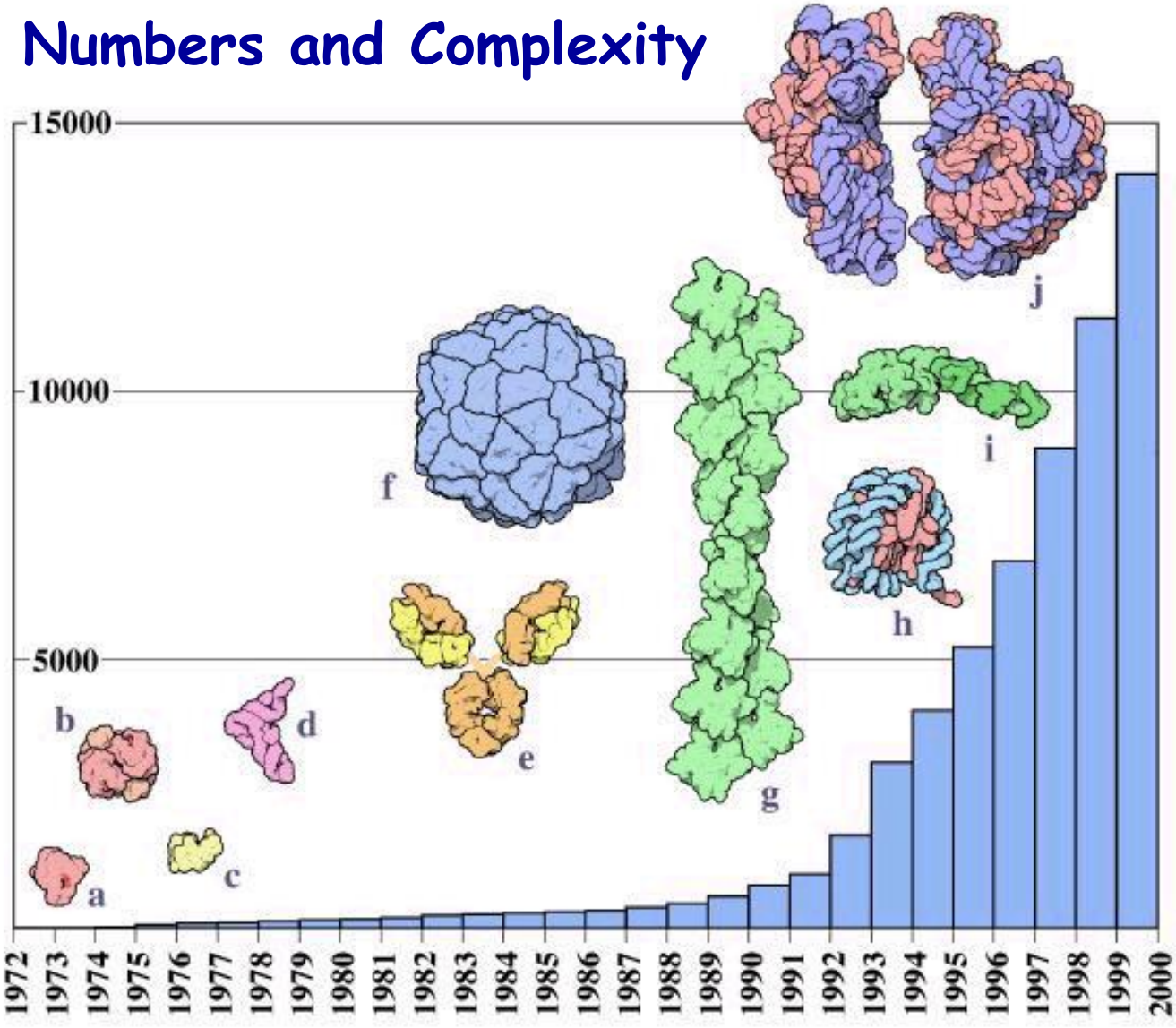
Rob Phillips et al. Physical Biology of the Cell 2008

Life is made of enormous large number of biomolecules organized in an orchestrated manner



Rob Phillips et al. Physical Biology of the Cell 2008

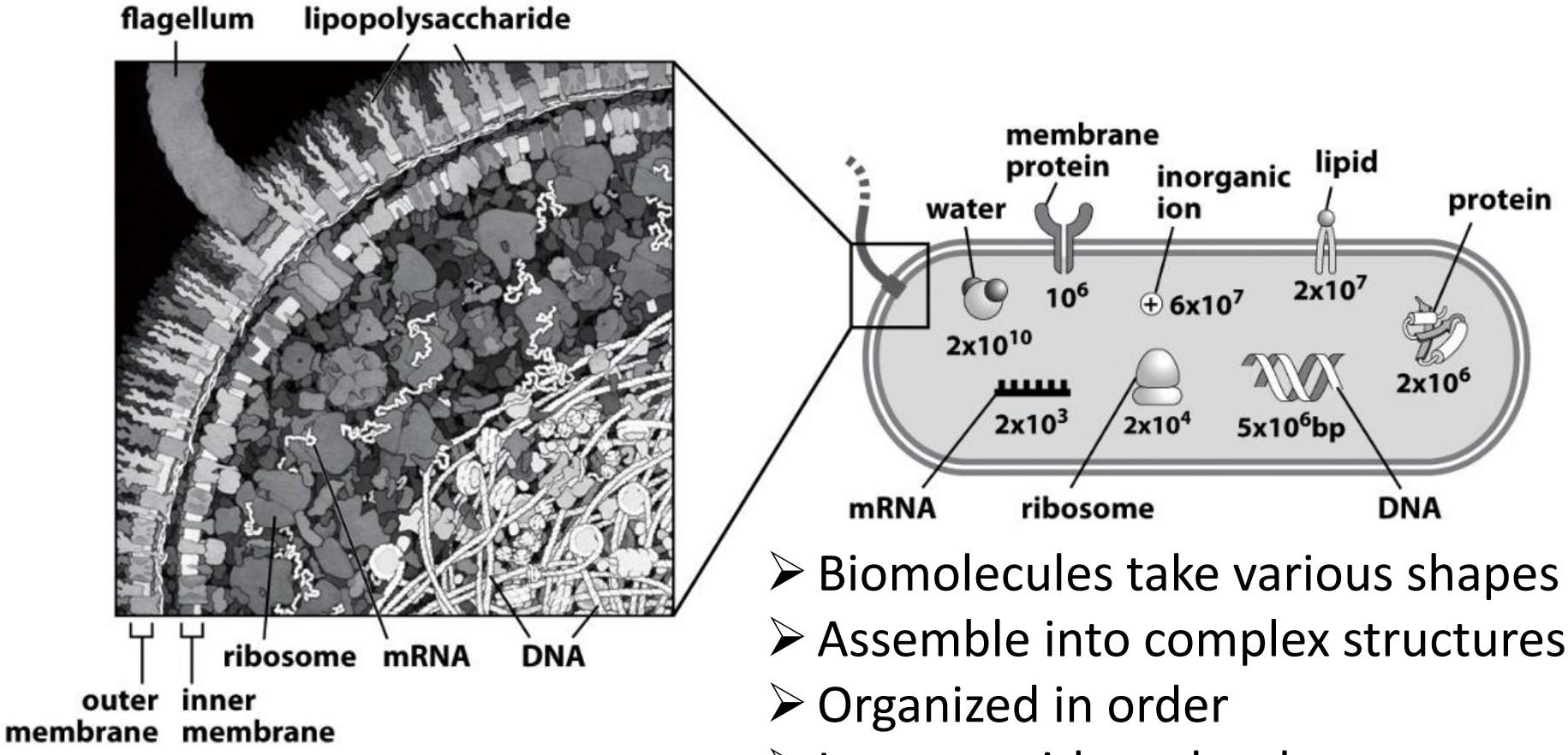
Status - Numbers and Complexity



(a) myoglobin (b) hemoglobin (c) lysozyme (d) transfer RNA
(e) antibodies (f) viruses (g) actin (h) the nucleosome
(i) myosin (j) ribosome

Courtesy of David Goodsell, TSRI

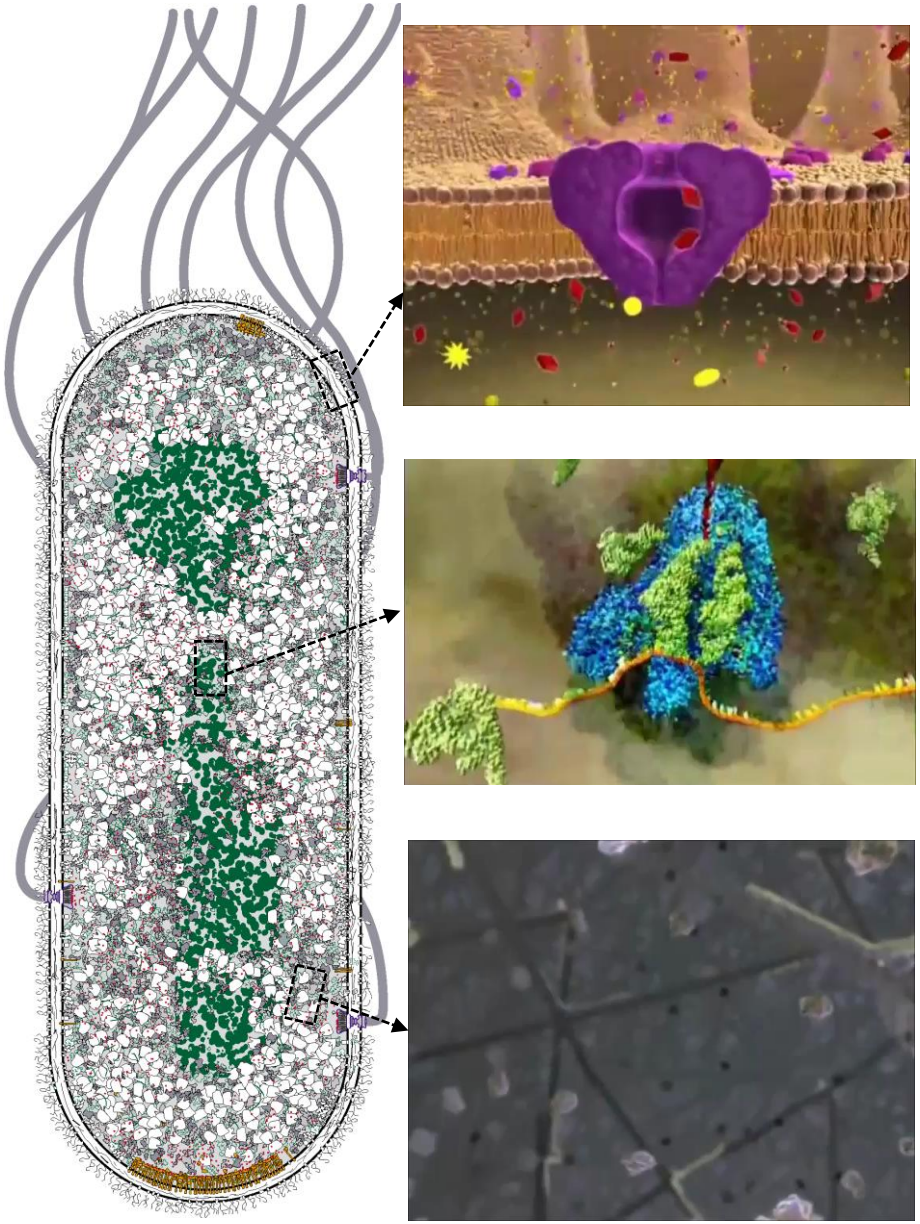
Life is made of enormous large number of biomolecules organized in an orchestrated manner



- Biomolecules take various shapes
- Assemble into complex structures
- Organized in order
- Interact with each other

Rob Phillips et al. Physical Biology of the Cell 2008

Biomolecules are dynamic



Structure

Dynamics

Function

Typical Questions We ask...

- **What drives large biomolecules to achieve their 3D structures? (Sequence-structure)?**
- **How do biomolecules transit, vibrate and change structures ?**
- **How are their functions realized? (Structure-function)**
- **How do they interact with each other, ligands, and environment and get influenced?**

What is Computational Biophysics?

- "Computational biophysics uses numerical simulation algorithms to study the physical principles underlying biological molecules and their processes."

- "It provides approximating solutions for theoretical biophysical problems lacking analytical solutions, simulates systems and make observations for which experiments are deemed infeasible."

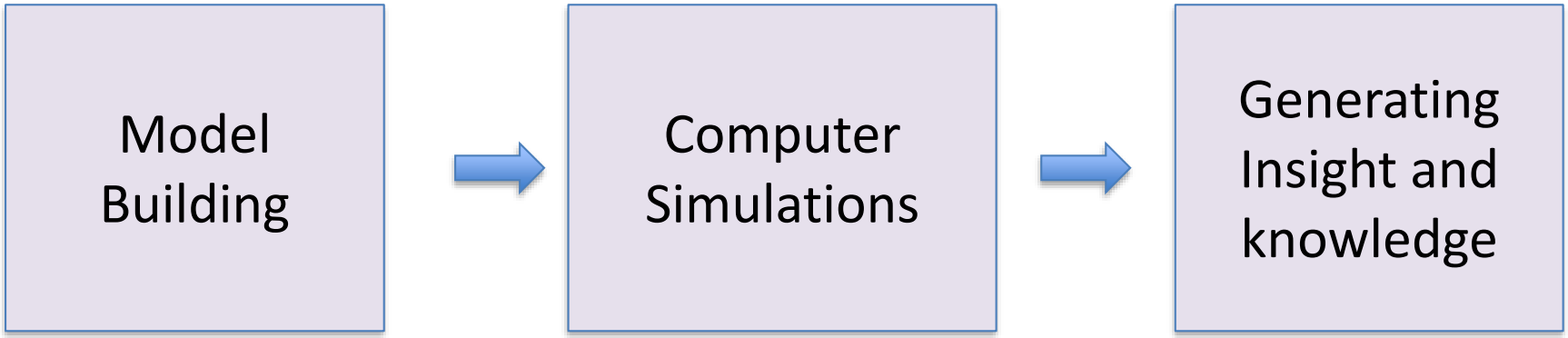
Source: www.nature.com

WHY DO SIMULATIONS?

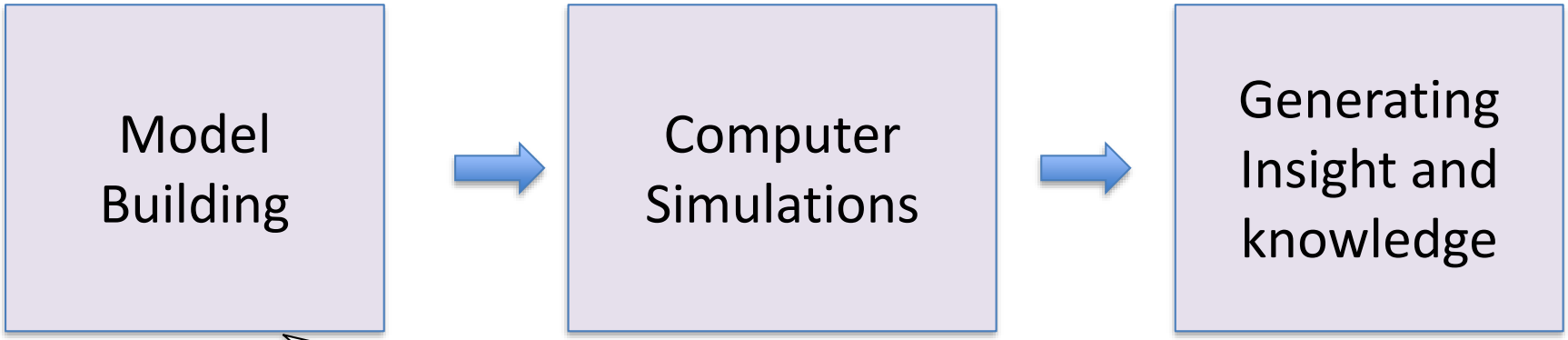
Numerical simulations fall between experiments and theoretical methods

- Where it is difficult or impossible to get exptl data
- Where there are no available experimental data
- Add atomic insight/prediction

Elements of Molecular simulations



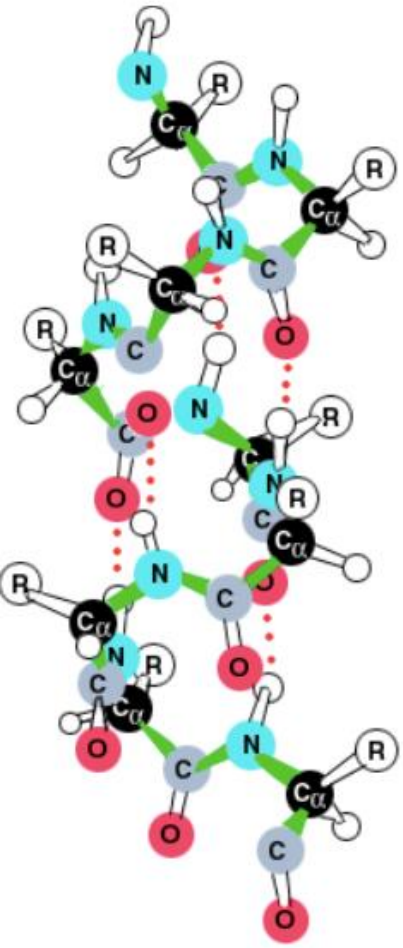
Elements of Molecular simulations

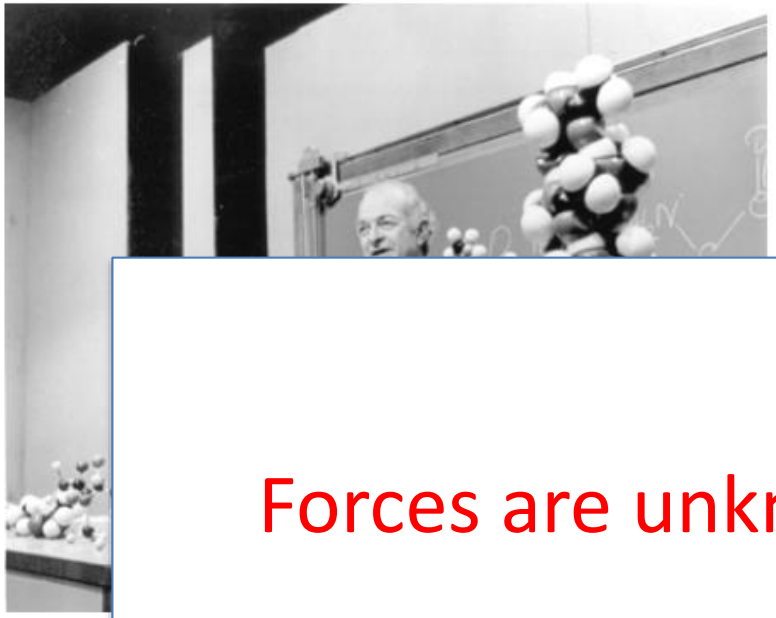


Describe molecular systems with (mathematical) models:

- Structures
- Forces

Ball-and-Sticks Models of Molecules





Forces are unknown!

1951

The alpha-helix

1901–1994

Physical Description of Molecular System

Time-dependent Schrödinger equation *(general)*

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = \hat{H}\Psi(\mathbf{r},t)$$

Paul Dirac, after his completion of formulism of quantum mechanics, comment: "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, …." *-1929, Dirac (Nobel in Physics)*

Paul Dirac

Full QM Description of Molecules

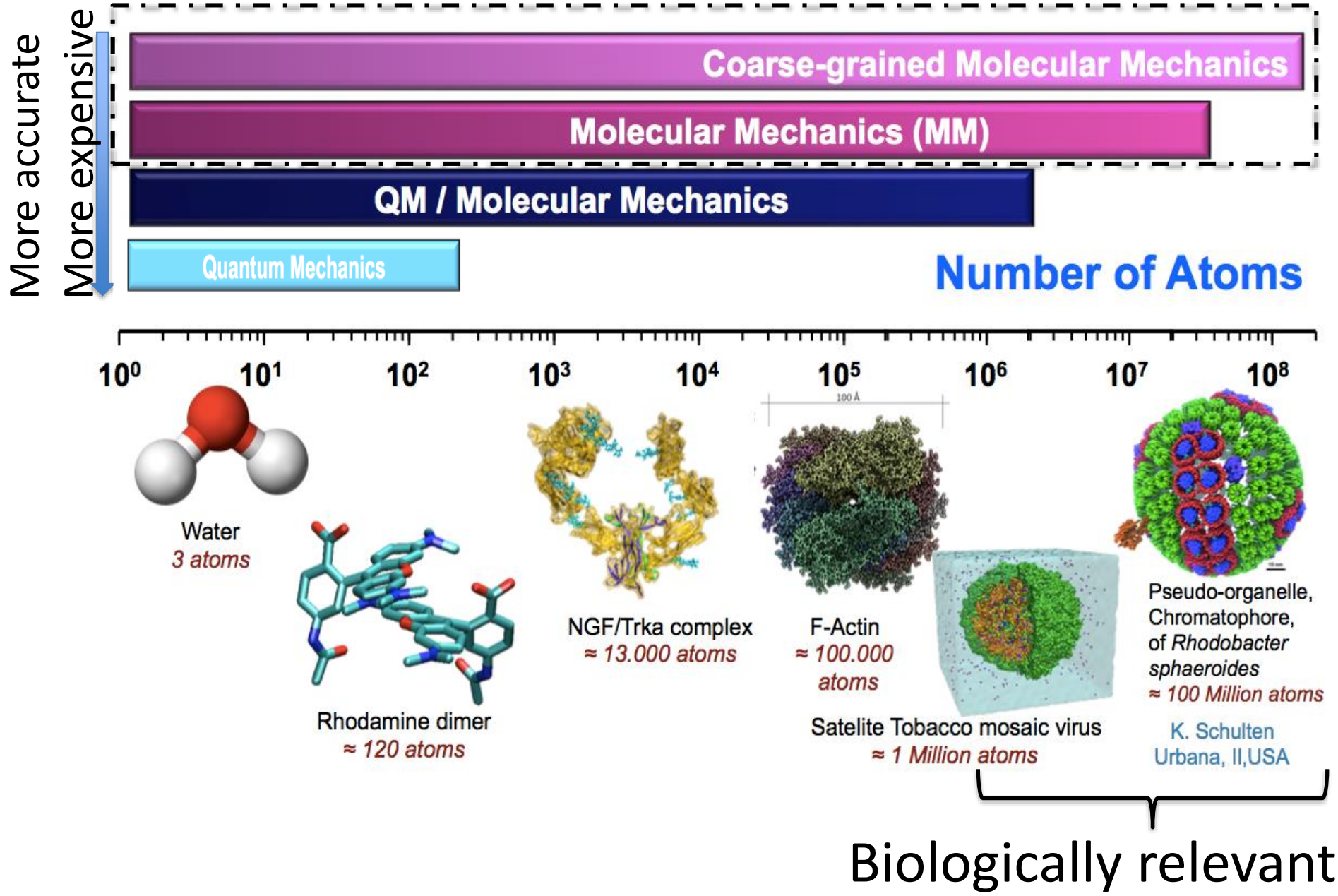
$$H = -\sum_k \frac{\hbar^2}{2M_k}\nabla^2 - \sum_i \frac{\hbar^2}{2m_i}\nabla^2 + \sum_k \sum_{l>k} \frac{Z_k Z_l e^2}{R_{kl}} + \sum_i \sum_{j>i} \frac{e^2}{r_{ij}} - \sum_{kj} \frac{Z_k e^2}{r_{kj}}$$

- H is function of coordinates of all nuclei and electrons
- Form of H is unknown and is almost impossible to be solved
- Numerical solution is expensive to obtain

"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and **the difficulty is only that the exact application of these laws leads to equations that are much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to explanation of the main features of complex atomic systems without too much computation**."

-1929 Dirac

Molecular Modeling at different levels



Classical Molecular Mechanics (Force field)

$$E(b,\vartheta,\varphi,r) = \sum \tfrac{1}{2} K_b (b-b_0)^2 + \sum \tfrac{1}{2} K_\theta (\theta-\theta_0)^2$$

All Bonds — All Angles

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

All Torsion Angles

Covalent/bonded Forces

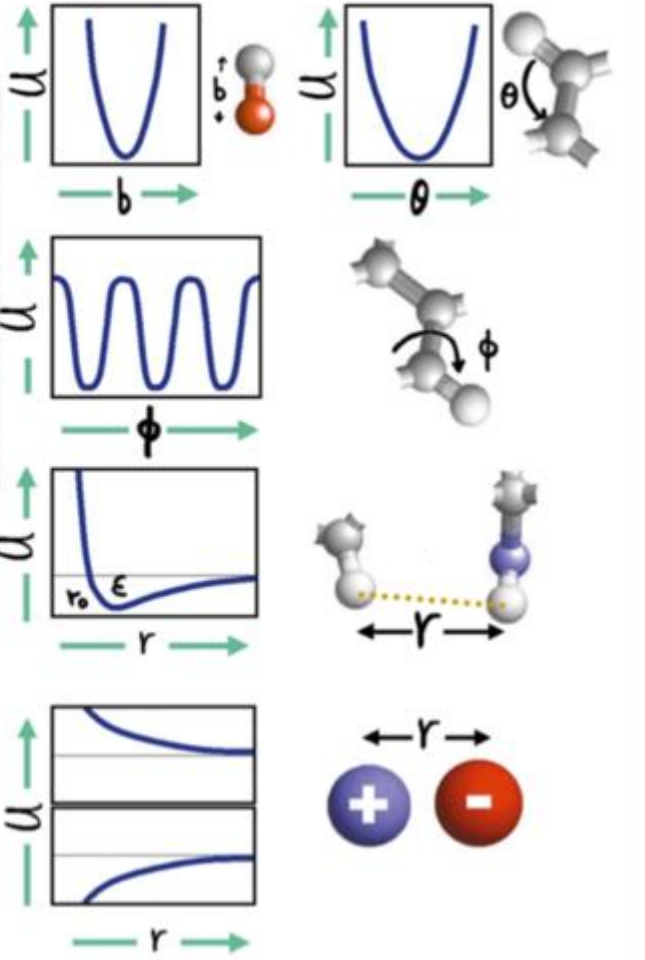
$$+ \sum \varepsilon \left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6} \right]$$

All nonbonded pairs

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

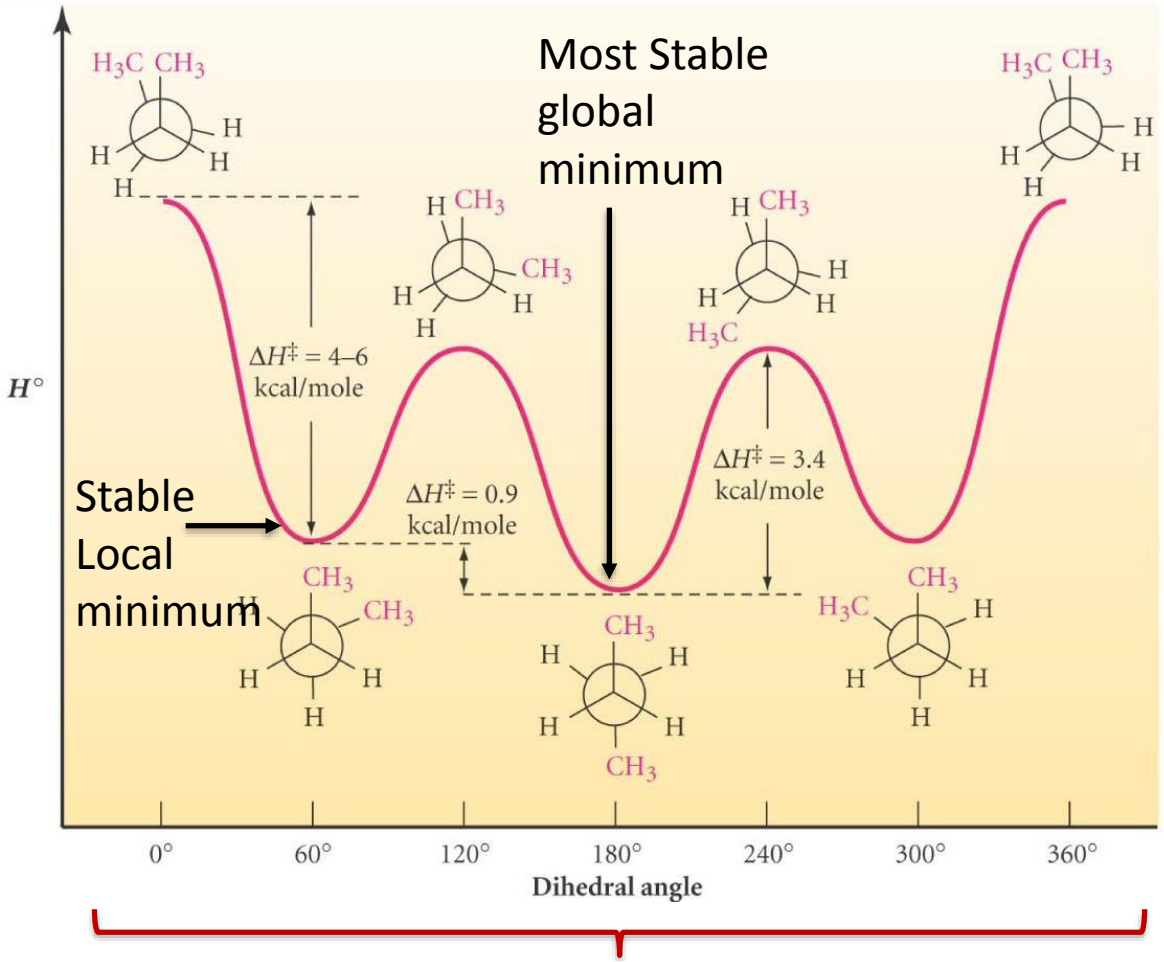
All partial charges

Non-Covalent/Non-bonded Forces



Limitations: point-charge approximation; unbreakable bonds

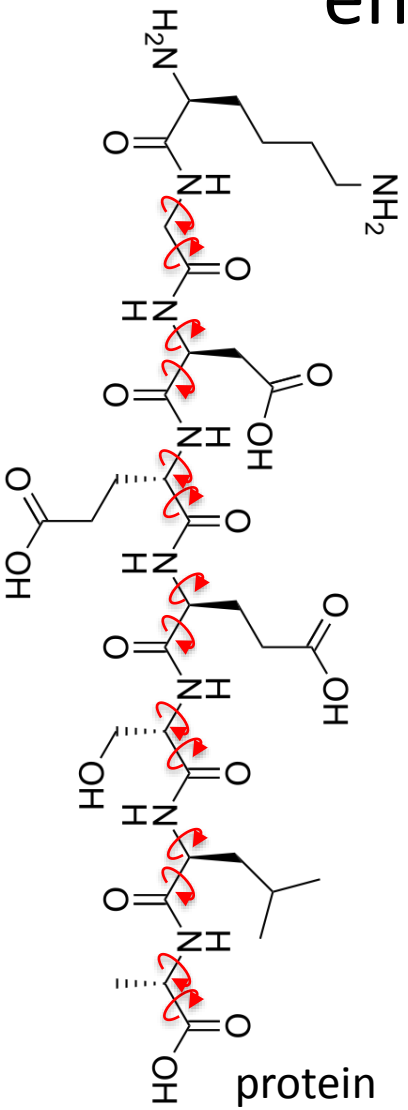
Conformations and conformational space
An example: n-butane



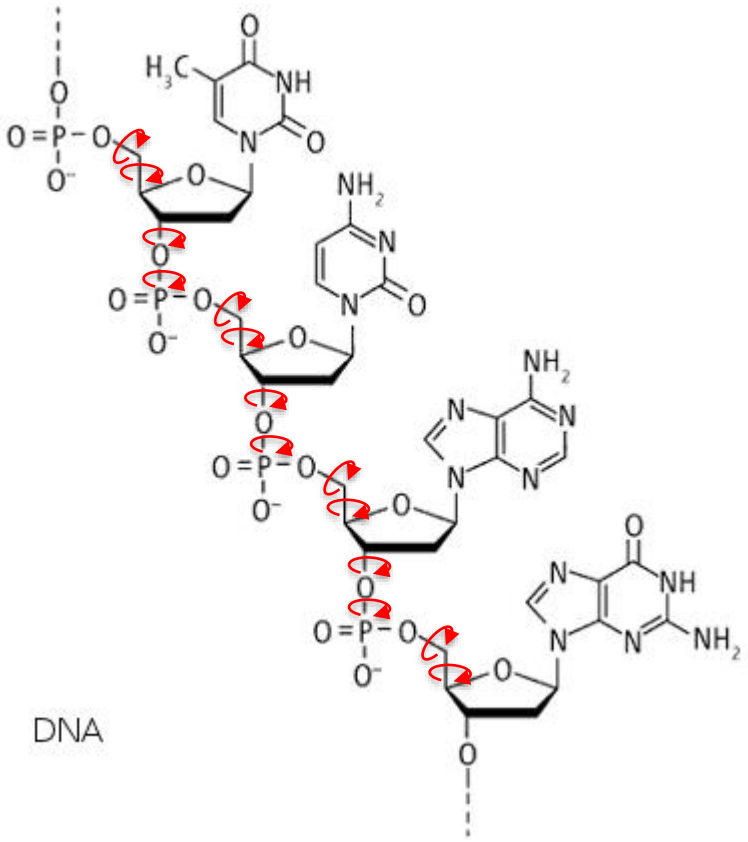
Conformations:

- Result from bond rotation
- Same chemical connectivity
- Different atomic arrangement
- Different energy, stability and probability
- Different properties

Conformational Space of Biomolecules are enormously large



protein

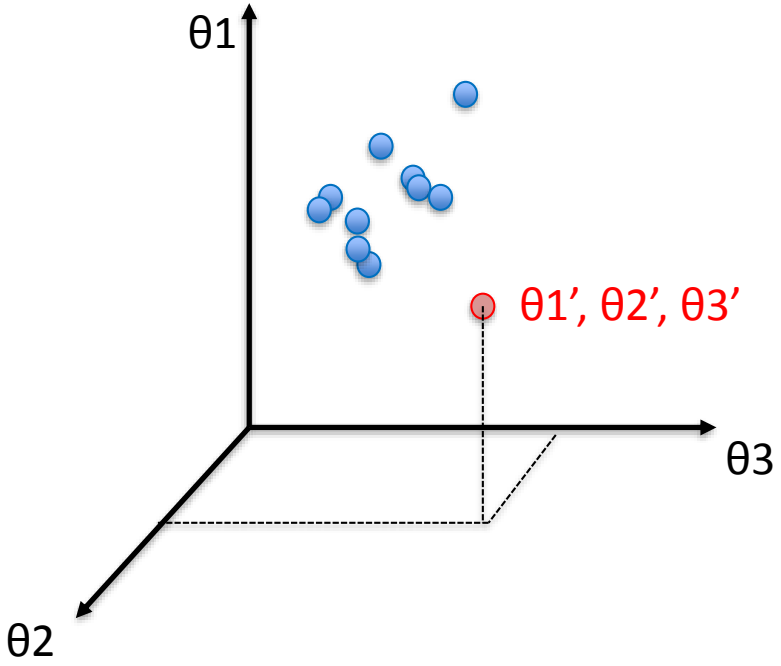


DNA

Each conformation has a unique set of rotation angles for all rotatable bonds

Conformations and Conformational Space

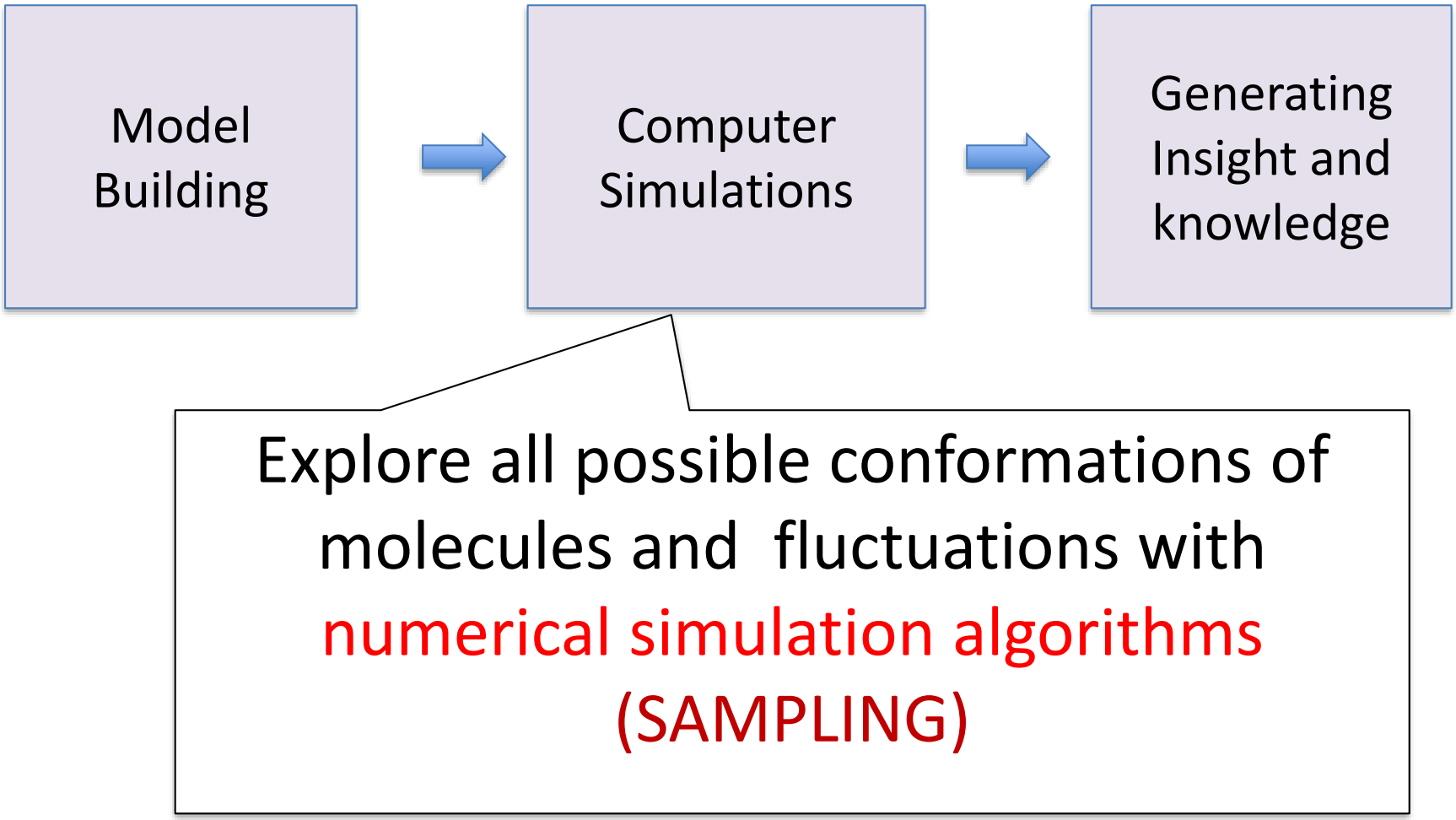
E.g.
A molecule with three rotatable bonds



Question:

- With given forces/potentials, what are the distribution of conformations?
- Most stable conformations?
- How system evolves in conformational space?

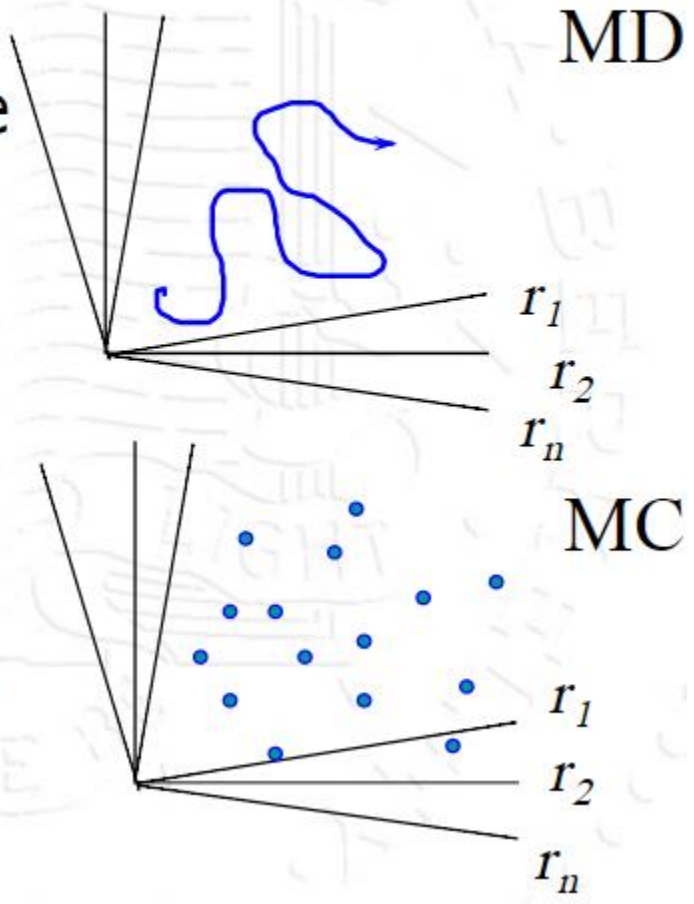
Elements of Molecular simulations



Molecular Simulations

- Molecular dynamics: solve equations of motion
- Monte Carlo: importance sampling

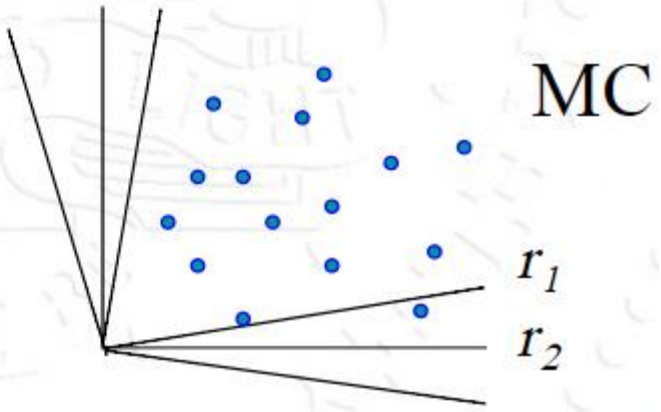
→ calculate thermodynamic and transport properties for a given intermolecular potential



Monte Carlo

- Generate a set of configurations with the *correct* probability
- Compute the thermodynamic and transport properties as averages over all configurations

(θ1, θ2, … θn, …)
(θ1', θ2', … θn', …)
(θ1", θ2", … θn", …)
:
:

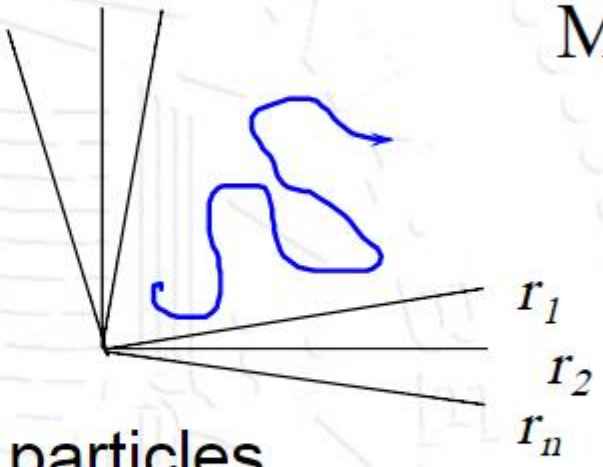


Molecular Dynamics

MD

- Theory:

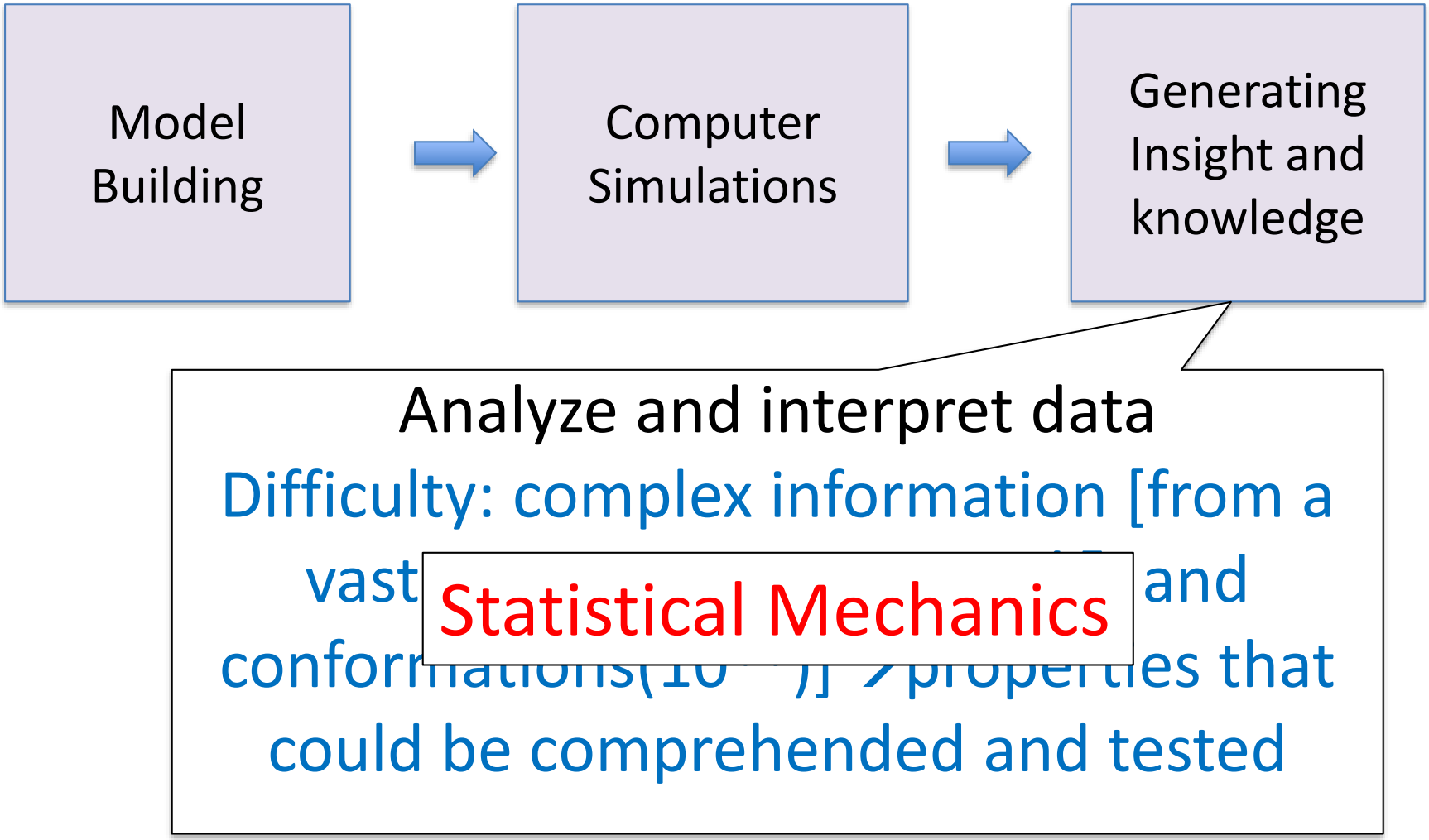
$$\mathbf{F} = m\frac{d^2\mathbf{r}}{dt^2}$$



- Compute the forces on the particles
- Solve the equations of motion
- Sample after some timesteps

(t, θ1, θ2, … θn, …)
(t’, θ1’, θ2’, … θn’, …)
(t”, θ1”, θ2”, … θn”, …)
:
:

Elements of Molecular simulations



2013 Nobel Prize in Chemistry

Photo: A. Mahmoud
Martin Karplus
Prize share: 1/3

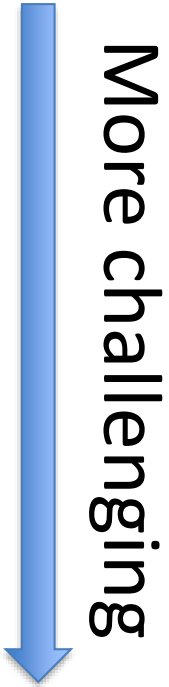
Photo: A. Mahmoud
Michael Levitt
Prize share: 1/3

Photo: A. Mahmoud
Arieh Warshel
Prize share: 1/3

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"*.

Application: CONFORMATION

- EXPERIMENTAL ANALYSIS
  (1) X-RAY refinement
  (2) NMR - structure determination.
- HOMOLOGY MODELLING
  Optimization of models
- Structural Prediction (ab inito)
  conformation in solution
  conformation in membrane



Structural Refinement

Structural Refinement

- Atomic positions cannot be measured directly

  X-ray: diffractions pattern

  NMR: J-coupling, nuclear overhauls effect …

  Cryo-EM: Electron density (low resolution)

- Molecular simulations provide atomic *positions but may not be sufficiently accurate; sampling issue*

Structural Refinement

Find conformations that minimize

$$E_{TOTAL} = E_{MM} + w_{exp}E_{exp}$$

E_{exp} : put energy penalty to structures that disagree experimental observations:

- E.g. NMR structure calculations modify the standard potential energy function to include NMR experimental constraints
  - distance constraints (NOEs)
  - dihedral constraints (NOEs, coupling constants, chemical shifts)
  - chemical shifts (^{1}H, ^{13}C)
  - residual dipolar coupling constants (RDCs)
- Controversial
  - not true experimental data
    - *but* similar to other parameterized geometric functions (bond length, angles etc)
  - bias structures to structures in PDB
    - *but* this is the criteria used to determine the quality of a protein structure

Convergence of NMR structure

Example: Molecular Dynamics Flexible Fitting

- Large complexes (ribosome, virus) are difficult to resolve for X-ray/NMR
- Cryo-EM: density map with medium-low resolution

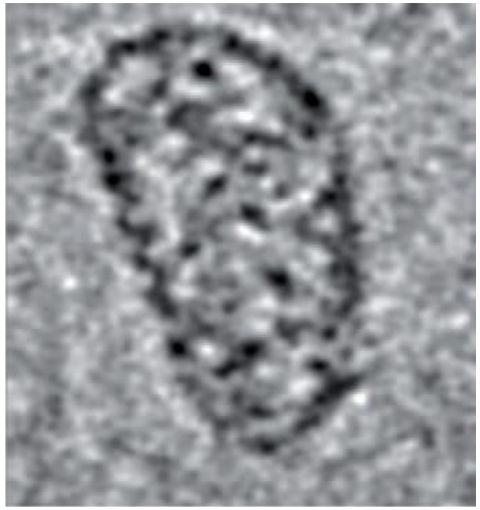
$$E_{TOTAL} = E_{MM} + wE_{xray/NMR} + w_{EM}E_{EM}$$

$$\mathbf{f}_i^{EM} = -\frac{\partial}{\partial \mathbf{r}_i} U_{EM}(\mathbf{R}) = -w_i \frac{\partial}{\partial \mathbf{r}_i} V_{EM}(\mathbf{r}_i)$$

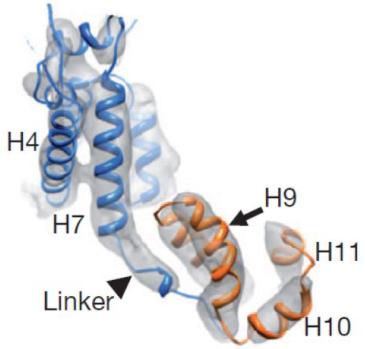
First Atomic Structure of Mature HIV Capsid

~13M Da/130K aa

0.8 nm

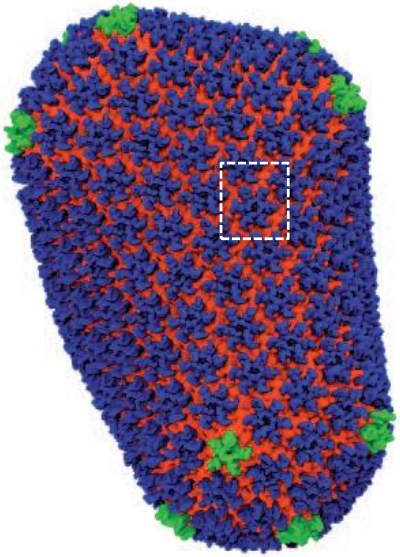


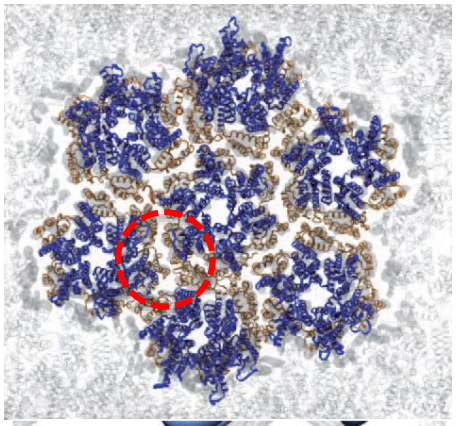
+

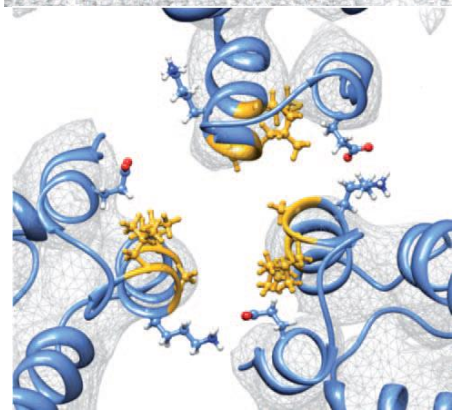


X-ray structure of subunit

MDFF

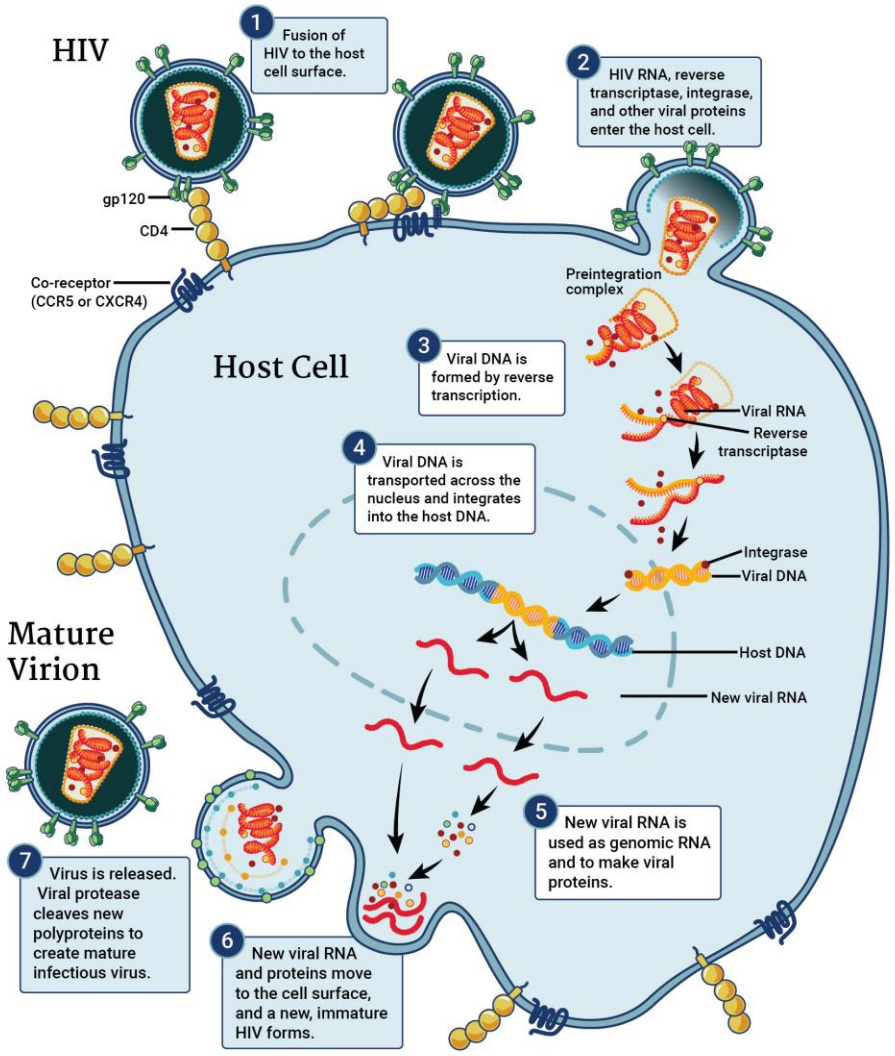






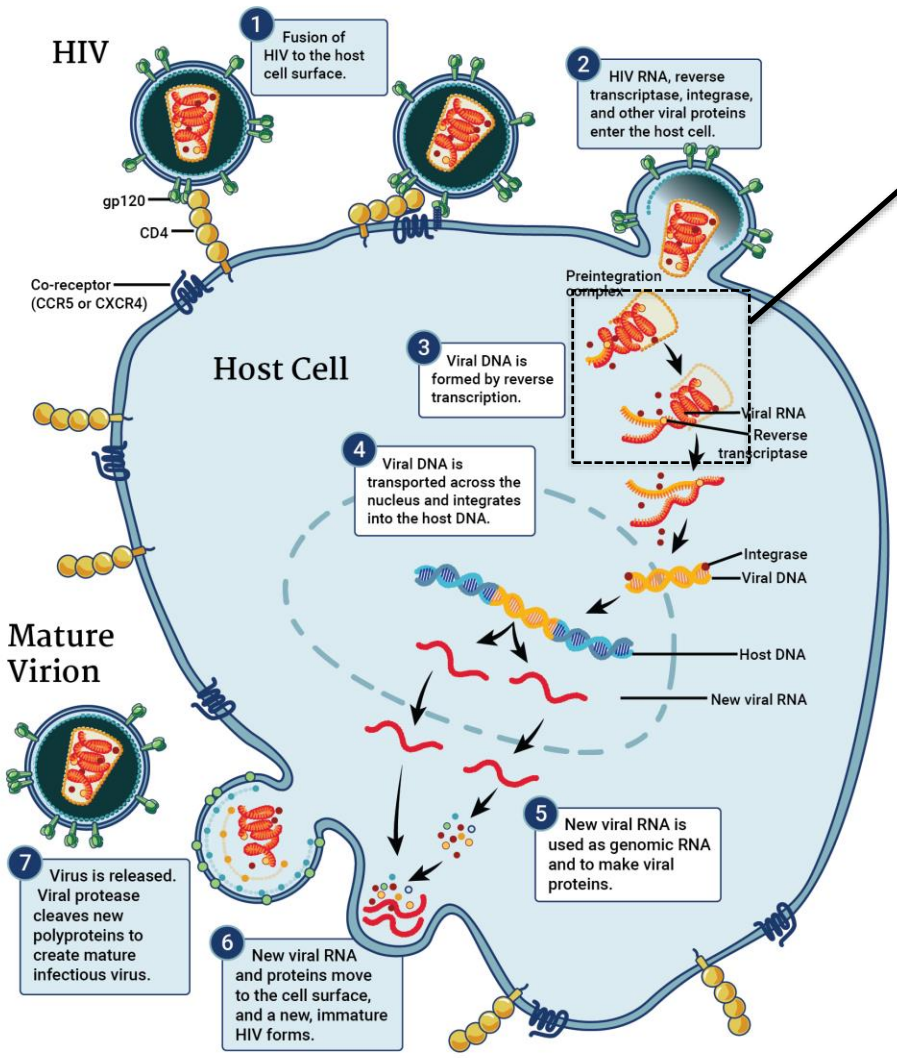
Nature 497:643-646, 2013

Capsid as a target to HIV



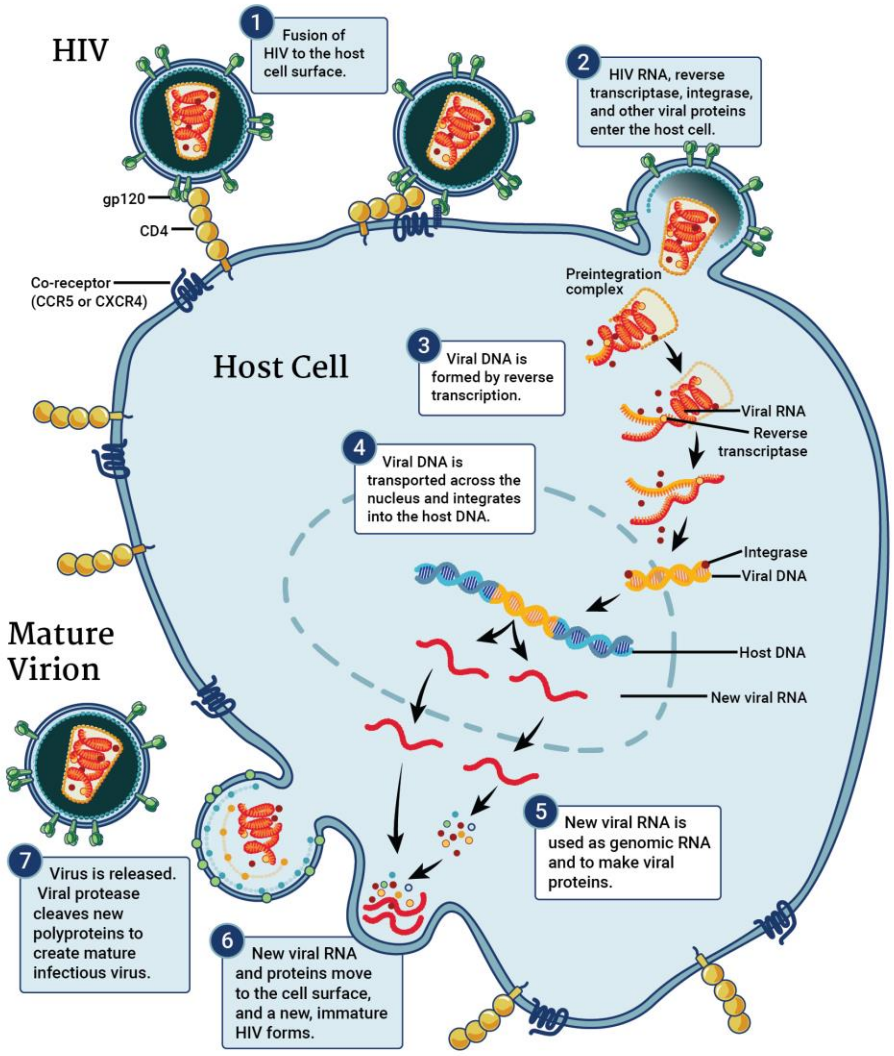
- HIV medicine: *stopping HIV at various stages of its life cycle*
- *Using capsid as target is hampered by lack of structural model*

Capsid as a target to HIV

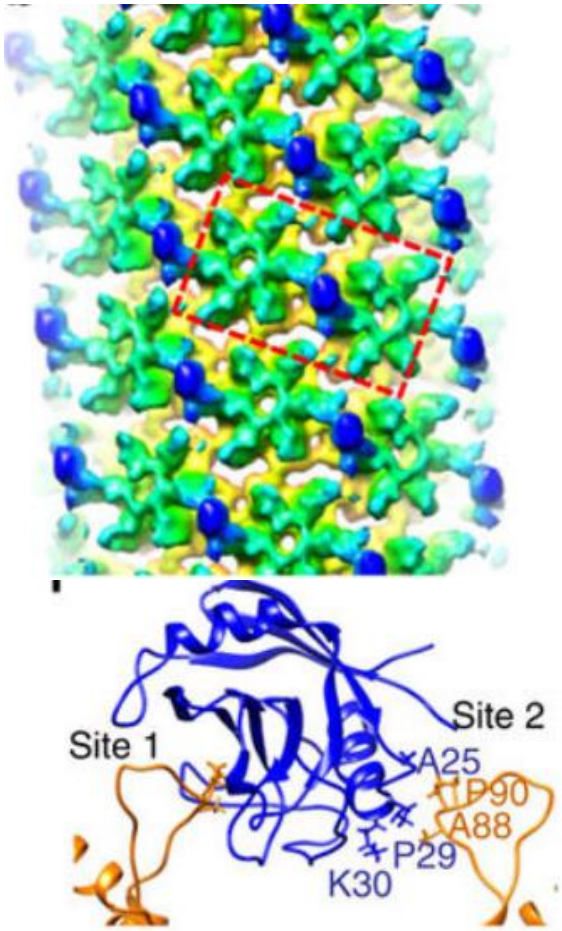


- Uncoating of capsid require binding of cyclophilin-A from host cells

Binding Site of Cyclophilin-A on HIV Capsid



MDFF study of capsid/Cyclophilin-A co-complex



Nature comm 2016, 7, 10714

THERMODYNAMICS

- FREE ENERGY CHANGE :

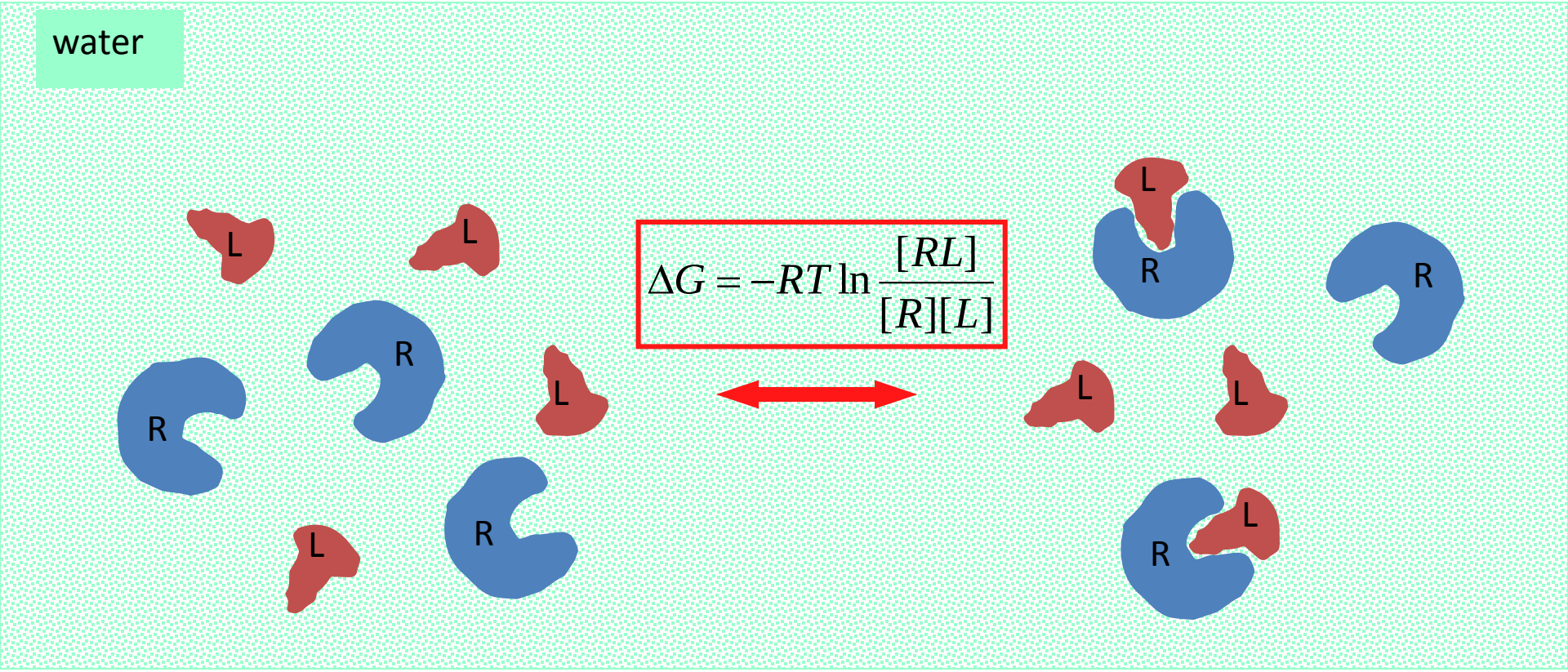
  enthalpy, entropy, heat capacity, equation of state (expt characterizable)

- BINDING AFFINITY:

  protein-protein/ligand/inhibitor interactions

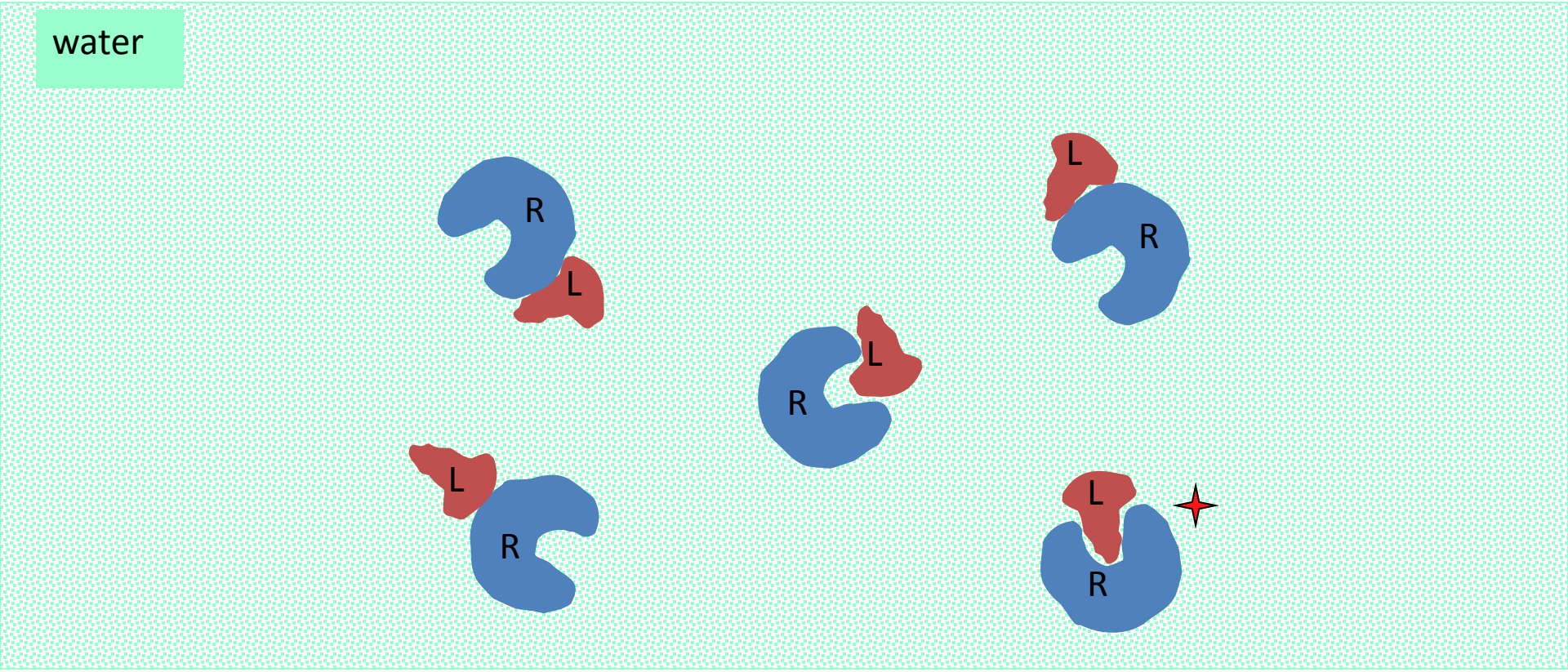
- STABILITY OF CHEMICAL MODIFICATION AND MUTATION

Protein-Receptor Interaction Thermodynamics



ΔG : Binding Free Energy

The Lowest Binding Free Energy ΔG



Why do we need to know binding Affinity?

- Protein-protein interaction is the cornerstone of cellular activity

- Computer-aided drug discovery

  - Known target and binding sites: *Ranking affinities*
  - Known target with unknown sites: *binding sites identification*
  - Unknown targets: *target identification*

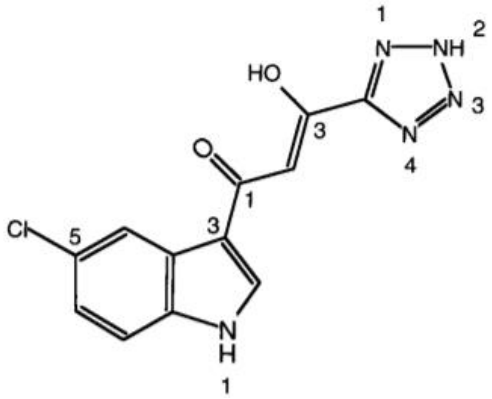
Methods of Affinity Calculation

	Ligand Flexible?	Protein Flexible?	Consider Free Energy?	Solvent Environment?
• Molecular Docking	No	No	No	No
• Flexible Docking	Yes	No	No	No
• Relaxed Complex Method	Yes	Yes	No	Partly
• End-point Binding Calculation	Yes	Yes	Partly	Partly
• Free Energy Perturbation	Yes	Yes	Yes	Yes

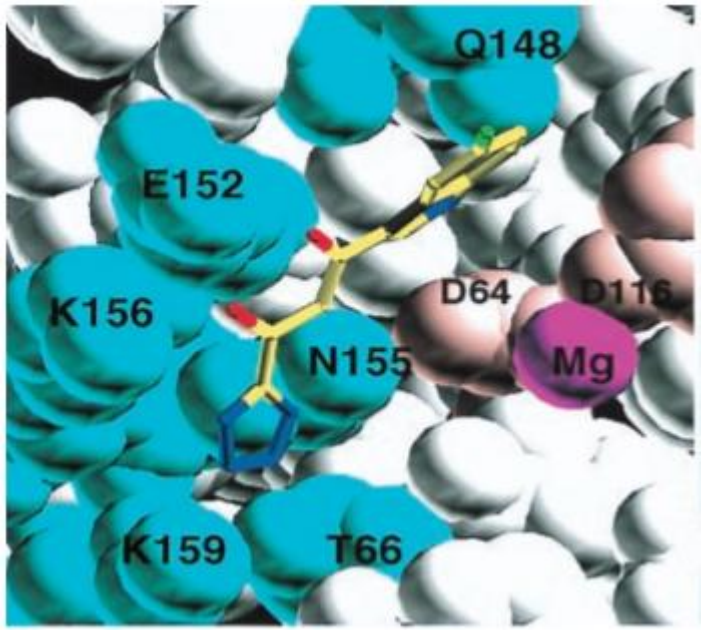
- For early stages, choose fast methods
- For a handful of top hits, choose accurate ones

Example: Computation-Aided Discovery of a HIV drug

HIV integrase (IN): integration of virus DNA into host genome



5CITEP, IC50: ~2uM



First IN-inhibitor structure

A critical loop near active site is missing

PNAS 1999 96, 13040

JA McCommon
UCSD

Use IN w/o inhibitors to start MD simulations

↓

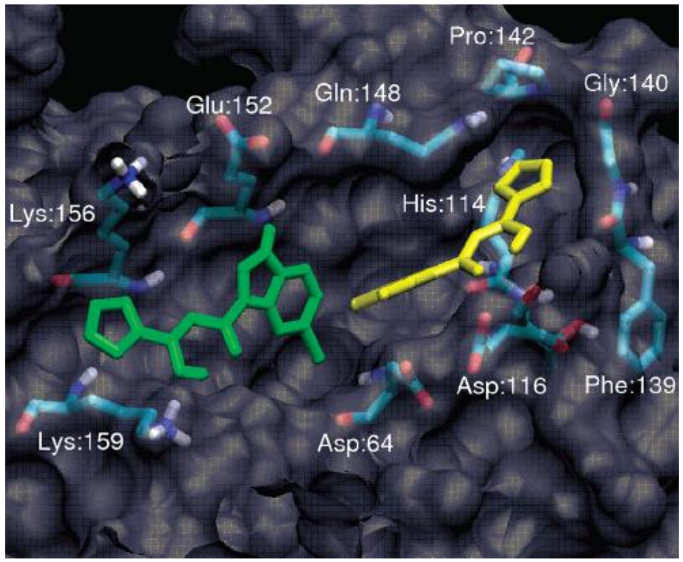
Select representative conformations

↓

Dock 5CITEP to IN in each conformation

J. Med. Chem. **2004,** *47,* 1879-1881

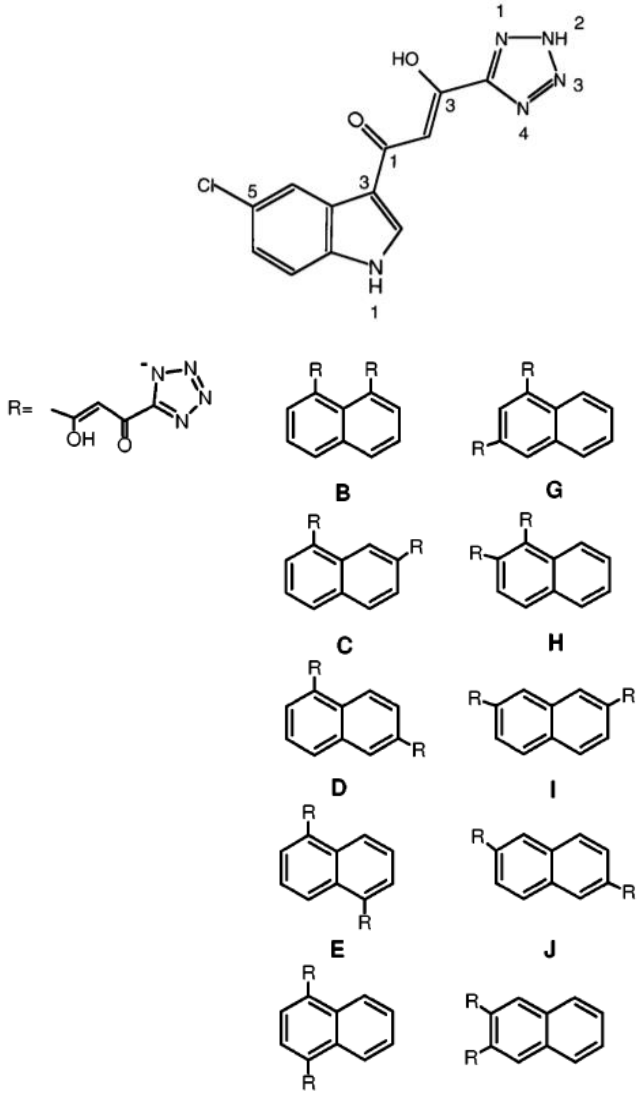
JA McCommon
UCSD



Crystal Site

New Site: open conformation

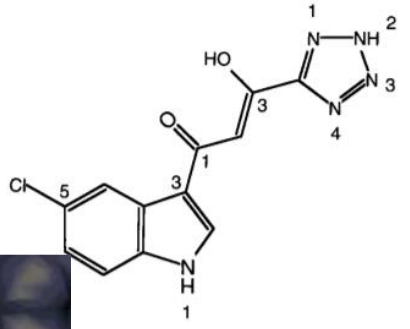
Butterfly version of

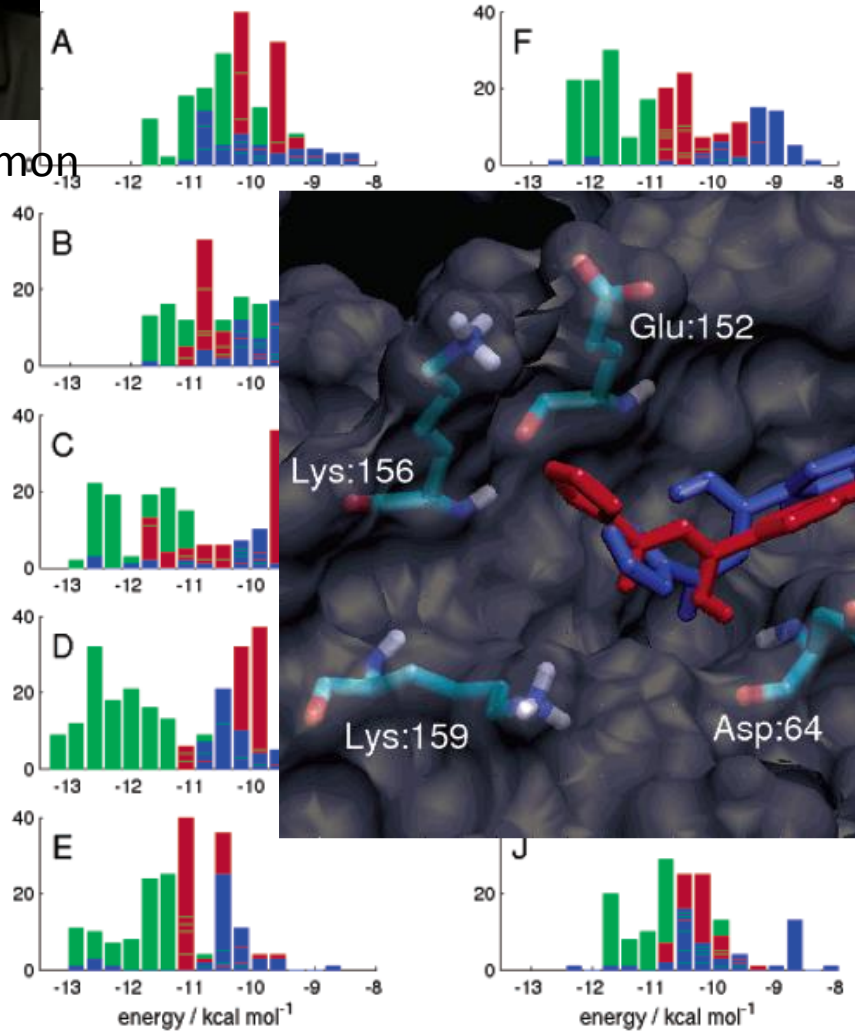


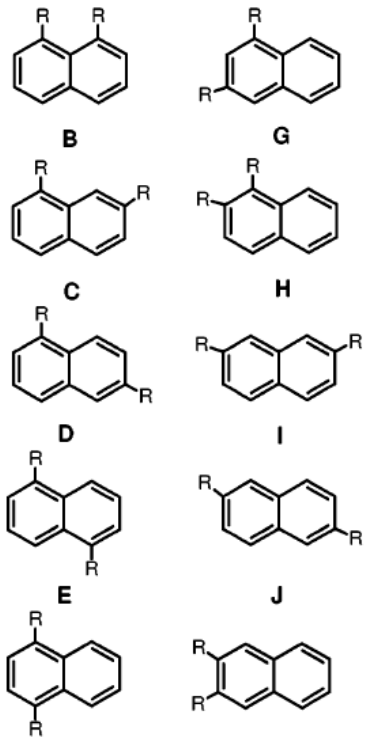
J. Med. Chem. **2004,** *47,* 1879-1881

JA McCommon
UCSD

Butterfly version of



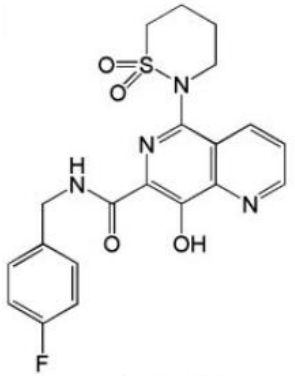




J. Med. Chem. **2004,** *47,* 1879-1881

Example: Computation-Aided Discovery of a HIV drug

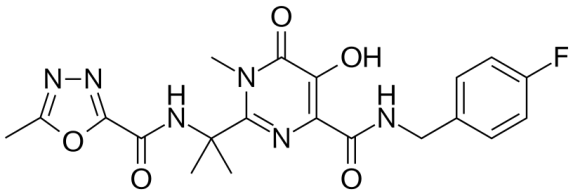
- Computation revealed a novel binding site
- Scientists from Merck developed high affinity compound to this site



L-870,810

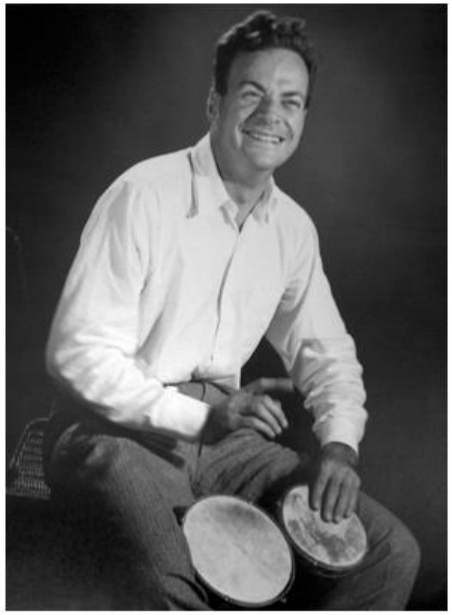
IC50: 4 nM, PNAS 2004, 101, 11233

- Raltegravir (Merck), Approved by FDA in 2007



Application: Dynamics of Biomolecules

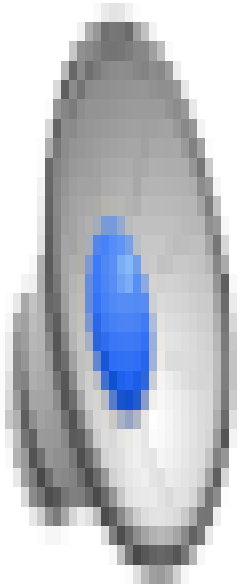
“...everything that living things do can be understood in terms of the jigglings and wigglings of atoms”



The Feynman Lectures in 1963

The protein behaves like liquid rather than solid

Fluid-like internal motions occur in proteins at room temperature



Martin Karplus
2013 Nobel

Nature 1977, 267, 585

Application: Dynamics of Biomolecules

What could be addressed?

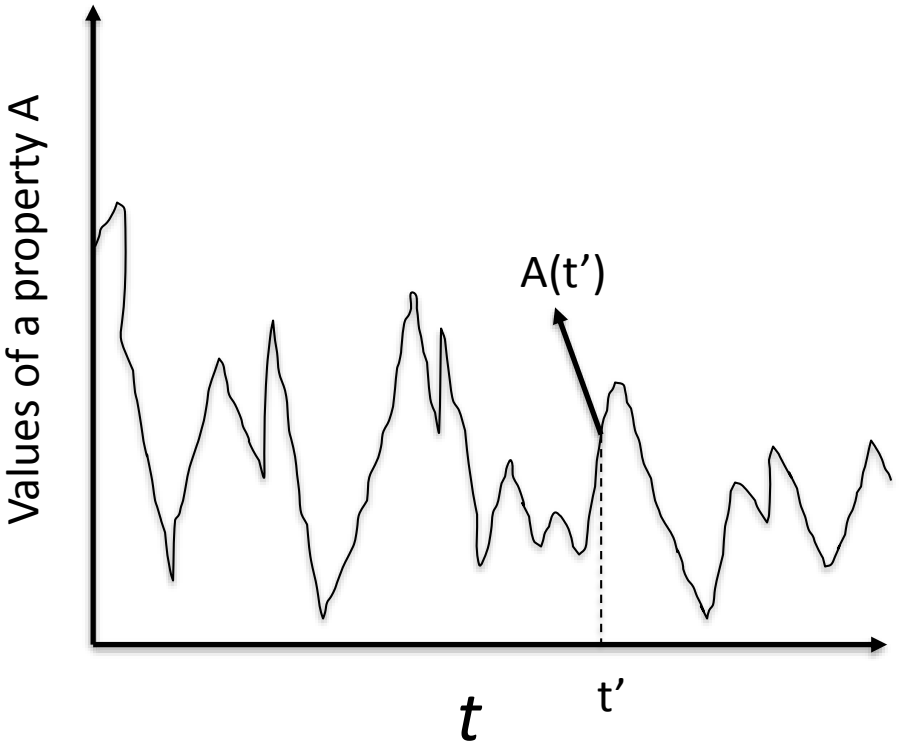
- How can a protein fold into native structure?

- How could a protein move its parts to carry out a function?
  (such as moving an ion from one side of cellular membrane to the other side)

- How could a protein/inhibitor get into/leave the binding site of a protein?

- Timescale/pathway/barrier/intermediate

A MD simulation will give you:



A could be:

- Secondary structure
- Shape of molecule
- Substrate-protein distance

:

E.g.

Root-mean-square deviation of atomic positions (RMSD)

$$\mathrm{RMSD}(\mathbf{v},\mathbf{w}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|v_i - w_i\|^2}$$

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}((v_{ix}-w_{ix})^2 + (v_{iy}-w_{iy})^2 + (v_{iz}-w_{iz})^2)}$$

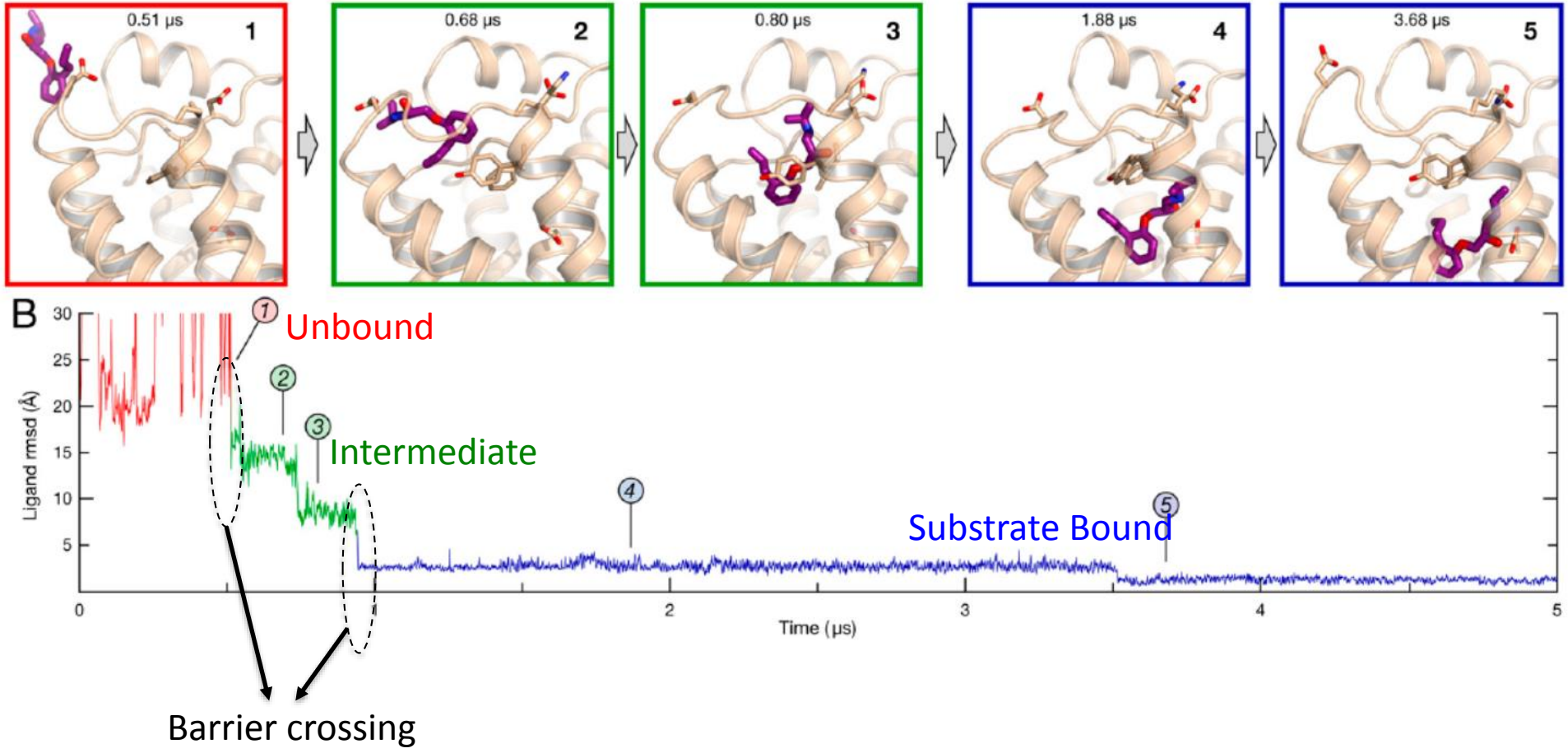
Pathway and mechanism of drug binding to G-protein-coupled receptors

Ron O. Dror[a,1,2], Albert C. Pan[a,1], Daniel H. Arlow[a,1], David W. Borhani[a], Paul Maragakis[a], Yibing Shan[a], Huafeng Xu[a], and David E. Shaw[a,b,2]

PNAS 2011, 108, 13118

Pathway and mechanism of drug binding to G-protein-coupled receptors

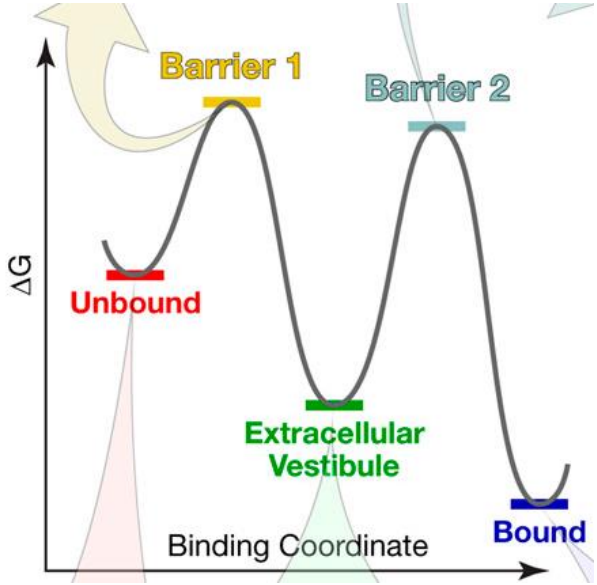
Ron O. Dror[a,1,2], Albert C. Pan[a,1], Daniel H. Arlow[a,1], David W. Borhani[a], Paul Maragakis[a], Yibing Shan[a], Huafeng Xu[a], and David E. Shaw[a,b,2]



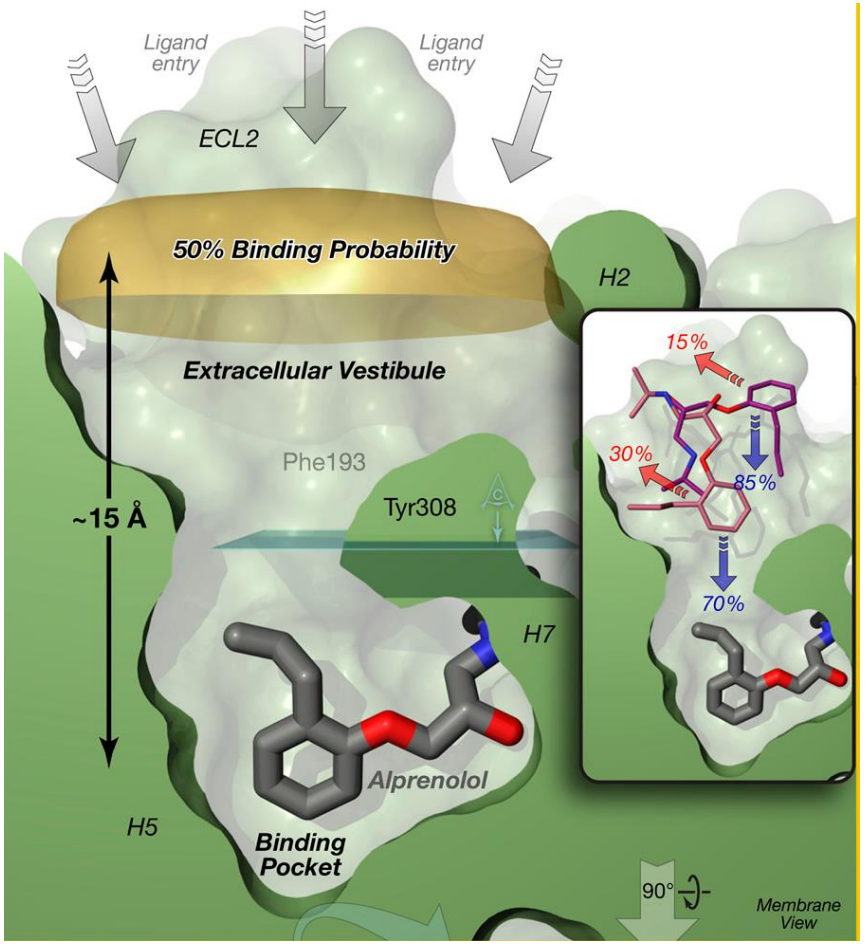
PNAS 2011, 108, 13118

Pathway and mechanism of drug binding to G-protein-coupled receptors

Ron O. Dror[a,1,2], Albert C. Pan[a,1], Daniel H. Arlow[a,1], David W. Borhani[a], Paul Maragakis[a], Yibing Shan[a], Huafeng Xu[a], and David E. Shaw[a,b,2]



The region that impacts rate most is far away from the binding site



PNAS 2011, 108, 13118

Patience is required to observe molecular events

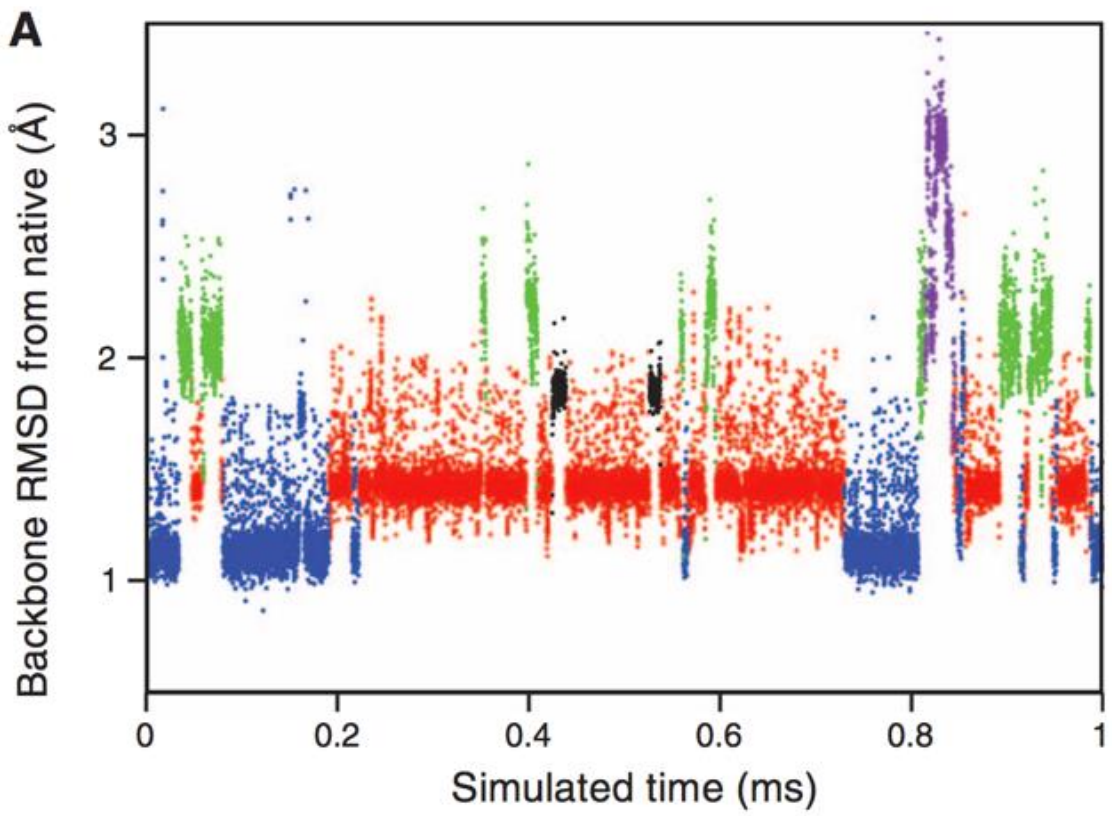
RESEARCH ARTICLE

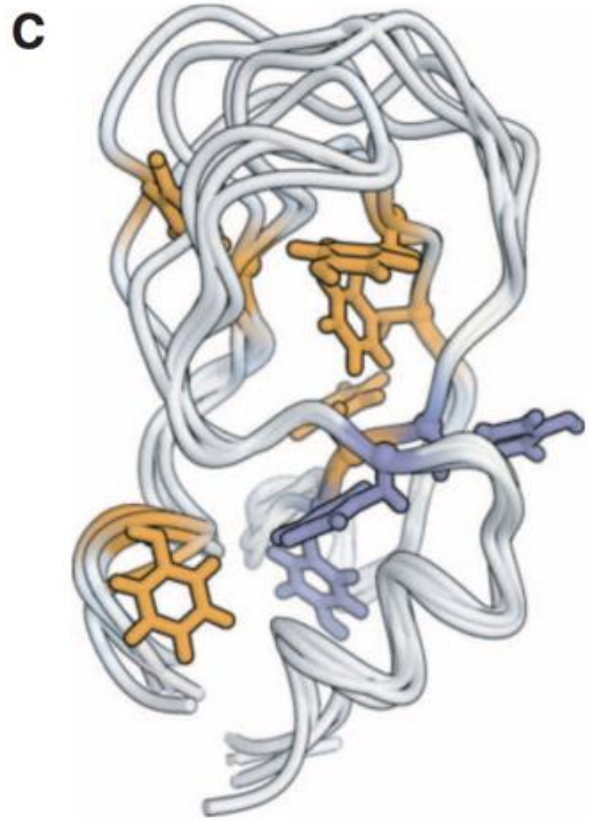
Atomic-Level Characterization of the Structural Dynamics of Proteins

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]

[1]D. E. Shaw Research, 120 West 45th Street, New York, NY 10036, USA.

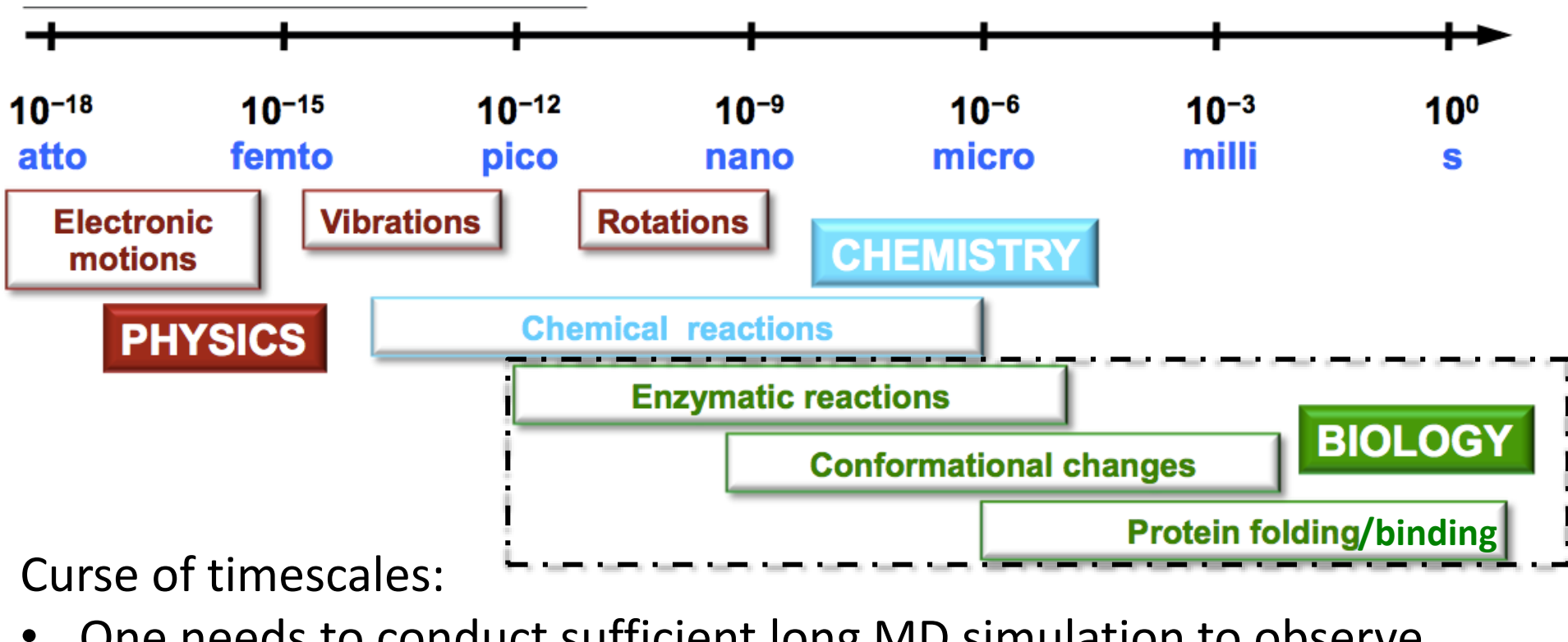
And there will always be a scale of dynamics you don't observe!





DE Shaw et al. (2010) *Science* 330:341-346.

Time scales of biomolecules



Curse of timescales:

- One needs to conduct sufficient long MD simulation to observe events

"sufficient long": order of magnitude longer than timescale of events

- MD simulation still have to be accumulated by integrating equation of motion femtosecond by femtosecond

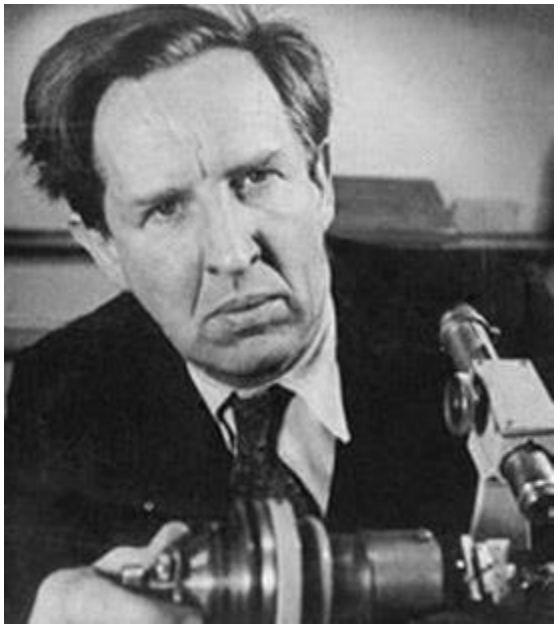
Moorse' law on simulation time scales

- Simulated BPTI for **9 ps** in vacuum (Karplus et al 1976)
- Simulated BPTI for **210ps** in a box of 2,607 water molecules (Levitt & Sharon, '88)
- Simulated Villin Headpiece for **1µs** in water (Kollman et al 1998)
- **One millisecond** simulation of BPTI in water (Shaw *et al.* 2010)
- But we are far from done for the time being

Seeing is believing: Computer Graphics

Graphic model at early time

Figure 7.1. Sir John Kendrew with the model of insulin, one of the first protein structures to be determined by X-ray crystallography. Components of the actual model are just visible through the forest of vertical support rods.



John Desmond Bernal (1901-1971)

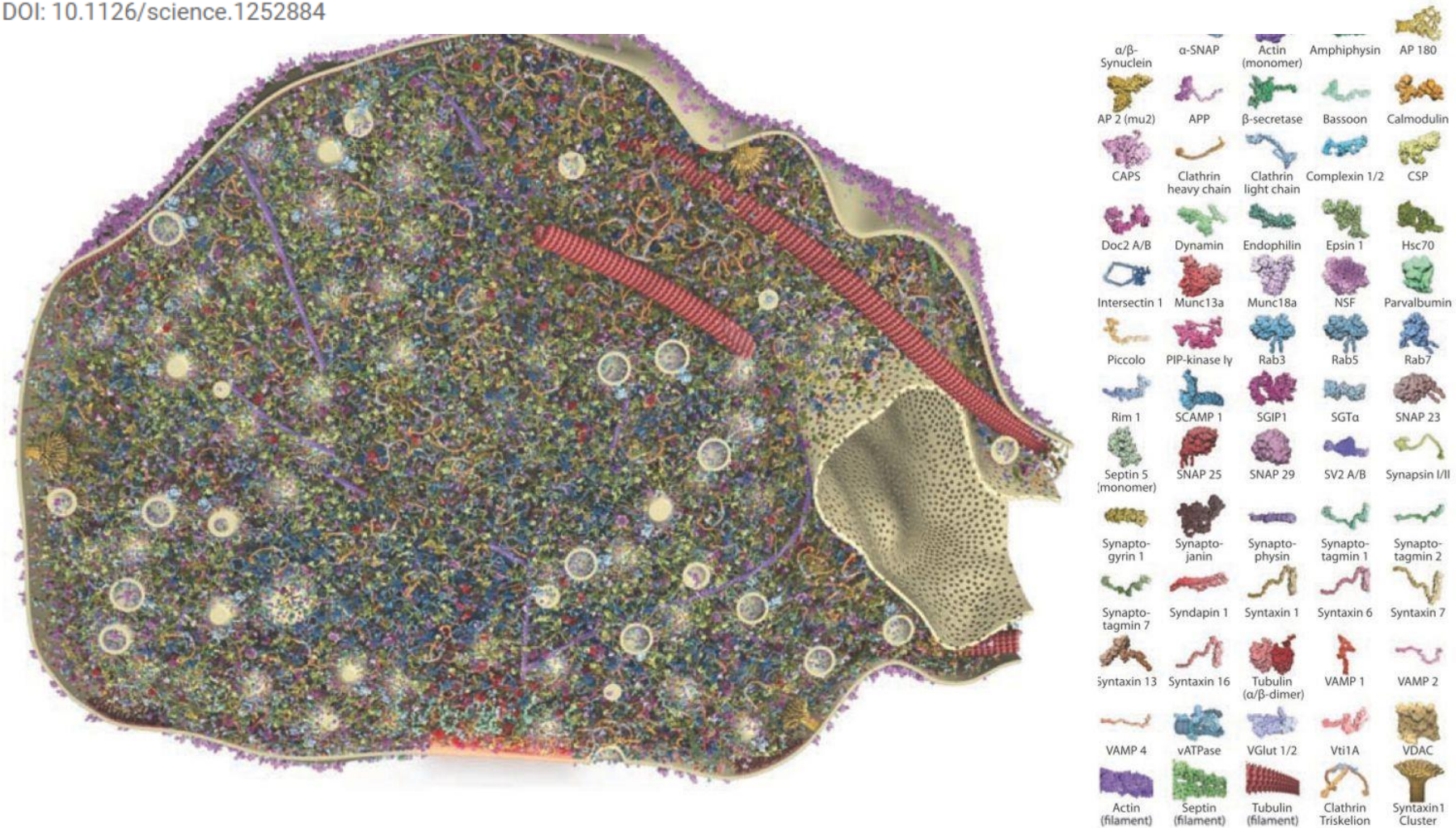
in 1962: "... I took a number of rubber balls and stuck them together with rods of a selection of different lengths ranging from 2.75 to 4 inches. I tried to do this in the first place as casually as possible, working in my own office, being interrupted every five minutes or so and not remembering what I had done before the interruption."

Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins

Benjamin G. Wilhelm[1,2], Sunit Mandad[3,*], Sven Truckenbrodt[1,5,*], Katharina Kröhnert[1], Christina Schäfer[1], Burkhard Rammn...

+ See all authors and affiliations

Science 30 May 2014:
Vol. 344, Issue 6187, pp. 1023-1028
DOI: 10.1126/science.1252884



Visual Molecular Dynamics (VDM) Can make pretty pictures of you beloved molecules and can do more…