CS/BioE/Biophys/BMI/CME 279
Computational biology: Structure and organization of biomolecules and cells

Image credit:
Ansgar Philippsen

Sept 15, 2020

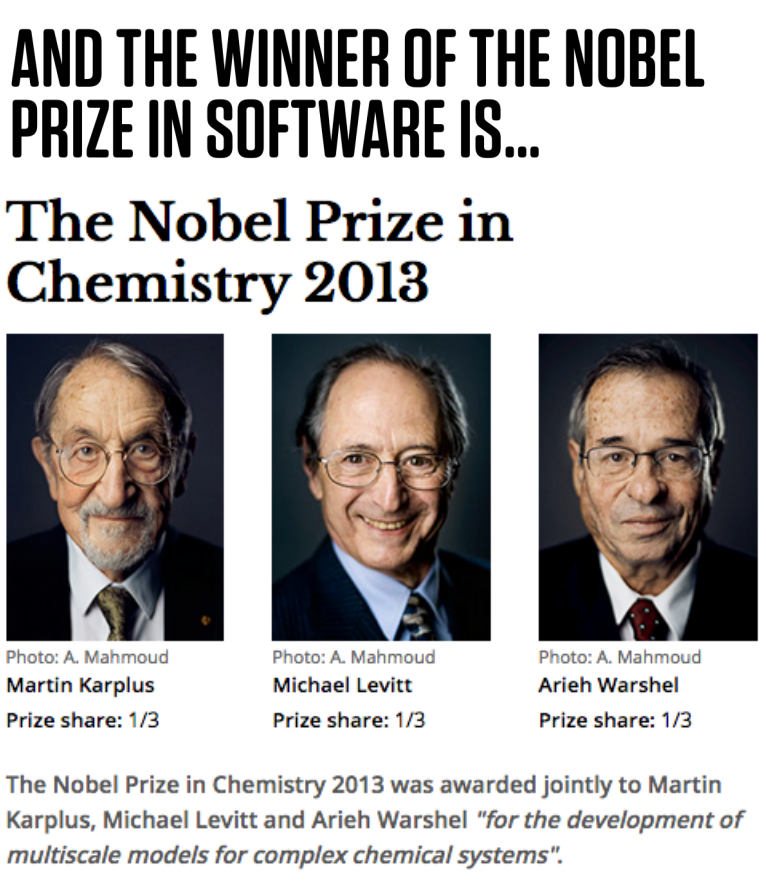
Ron Dror

20% of all science Nobel Prizes relate to 3D structure/organization of biomolecules

2017 Chemistry Nobel Prize: Cryoelectron microscopy

Experimental technique to figure out molecular structures

2013 Chemistry Nobel Prize: Computational models of biomolecules



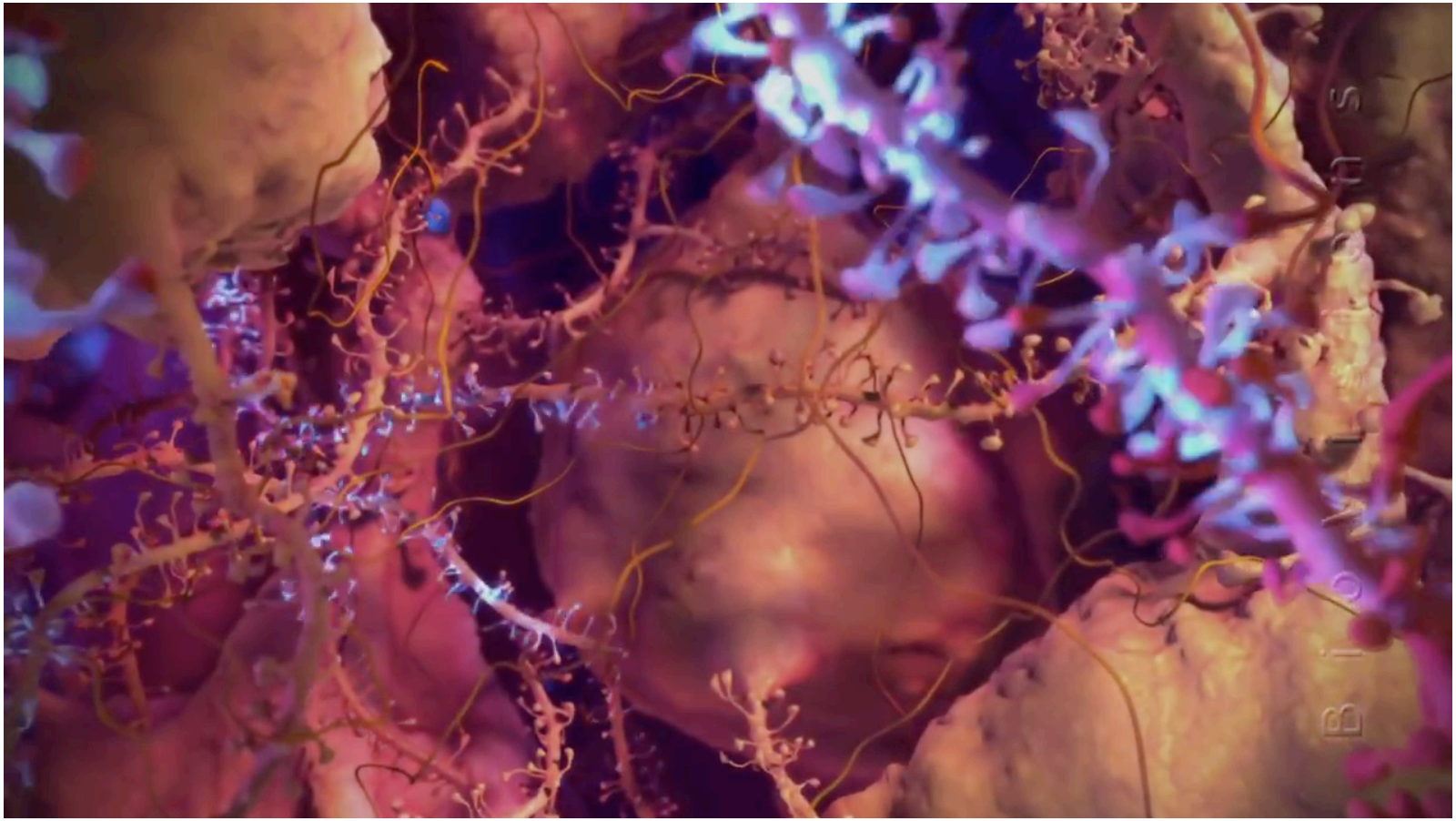
Outline for lecture 1 (course overview)

- What is structure?
  - Structure (and dynamics) at multiple spatial scales
- Why is structure important?
- How computation helps: An overview of course topics
- Recurrent themes
- Course logistics

What is structure?

In daily life, we use machines with functional *structure* and *moving parts*

Cells and biomolecules (e.g., proteins) are also machines whose function depends on structure and moving parts



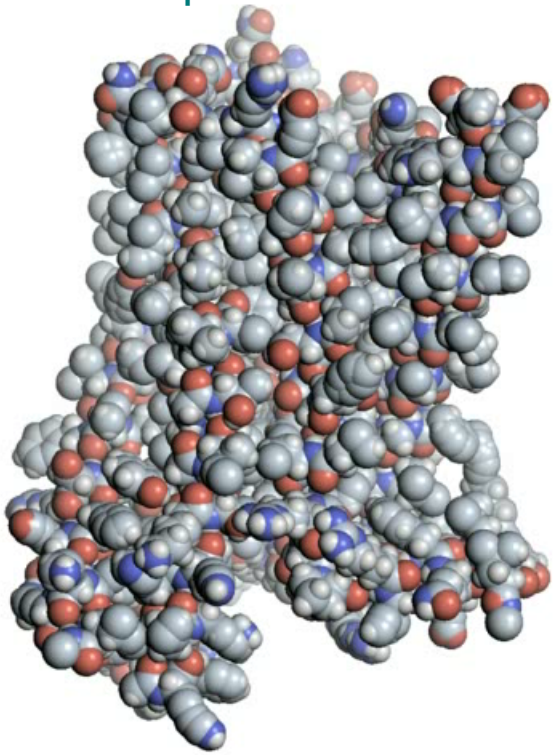
From *Inner Life of the Cell* | *Protein Packing*, XVIVO and Biovisions @ Harvard

What is structure?

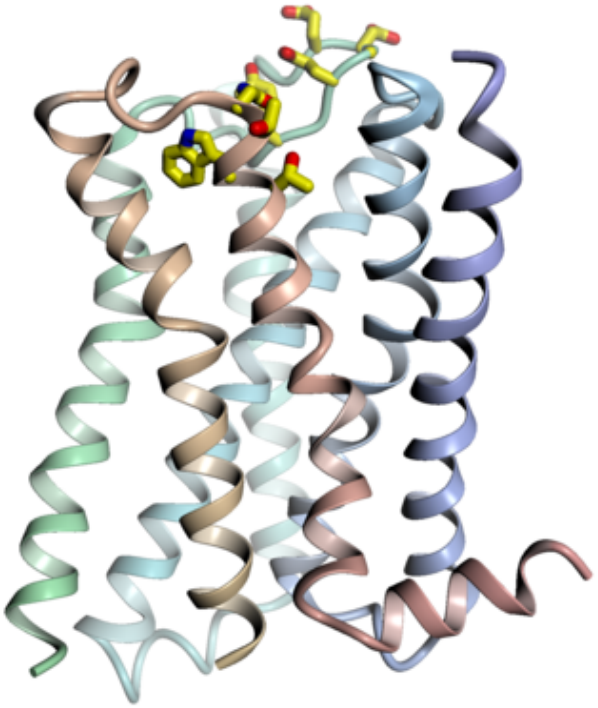
Structure (and dynamics)
at multiple spatial scales

Protein structure

All-atom representation



Cartoon representation

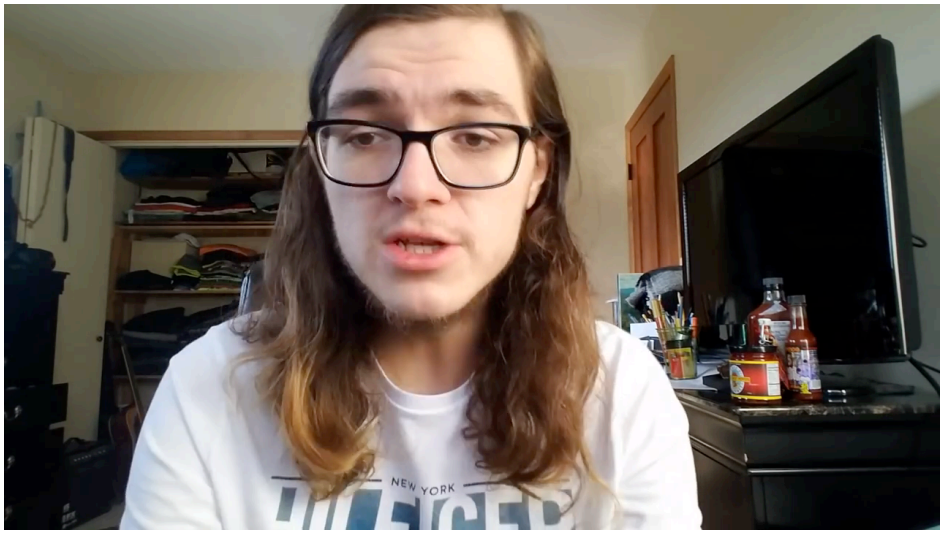


An adrenaline receptor
(the β_2 adrenergic receptor)

Example: how LSD binds to its target

Wacker et al., *Cell* 168:377, 2017
Collaboration with Bryan Roth (UNC)

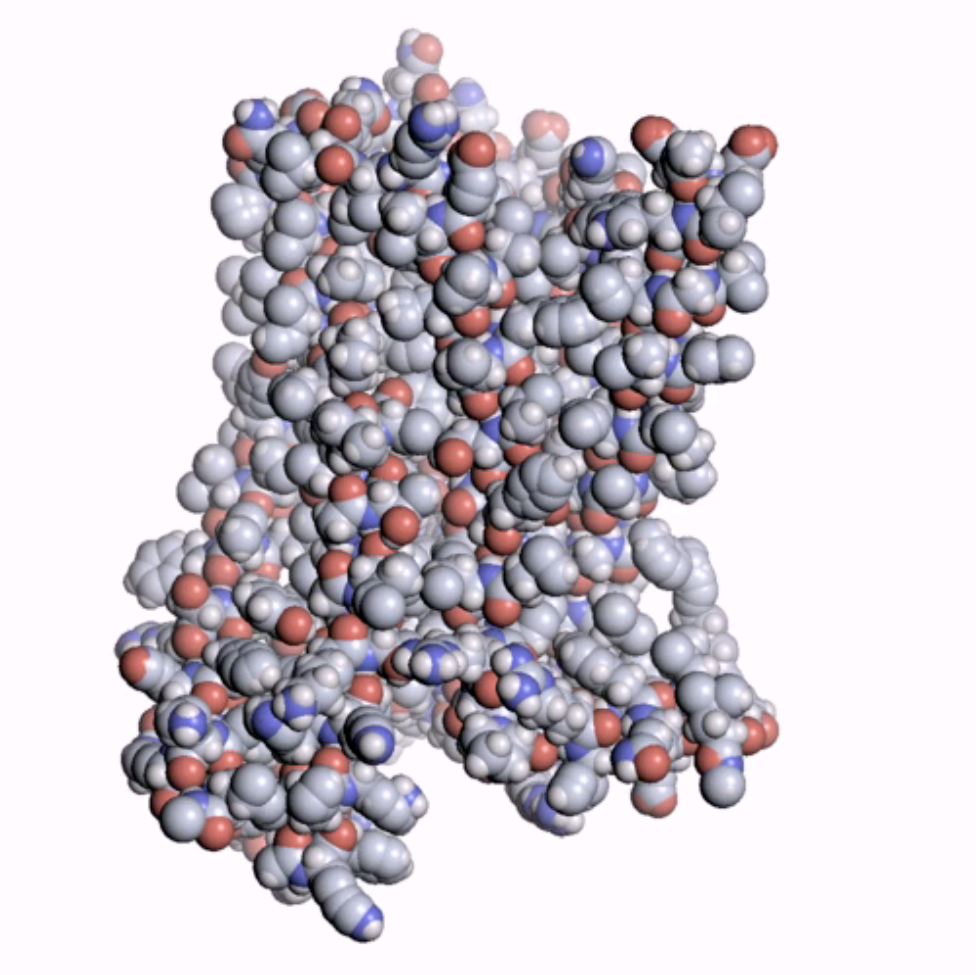
"Revealed: Why LSD Lasts So Long!"
AVI LSD YouTube Channel



https://www.youtube.com/watch?v=LjumHvnl-ME&feature=youtu.be

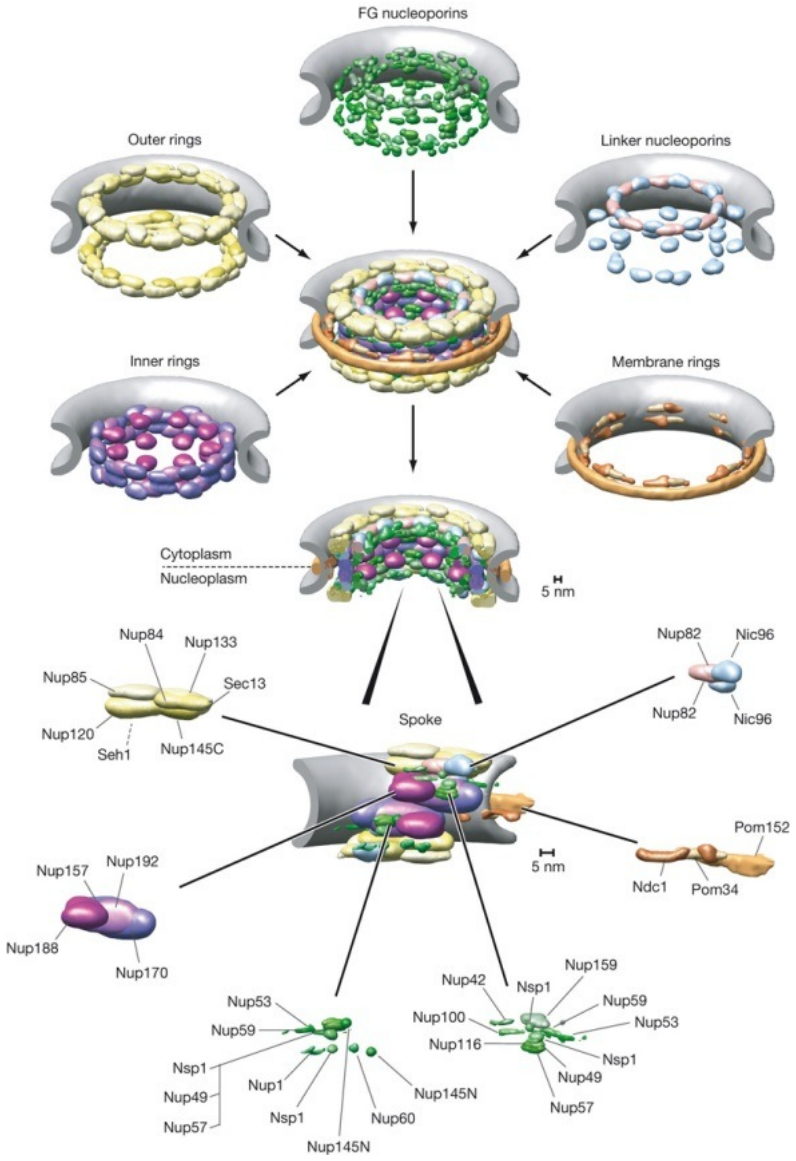
Atoms in protein move (NOT static)

Protein dynamics



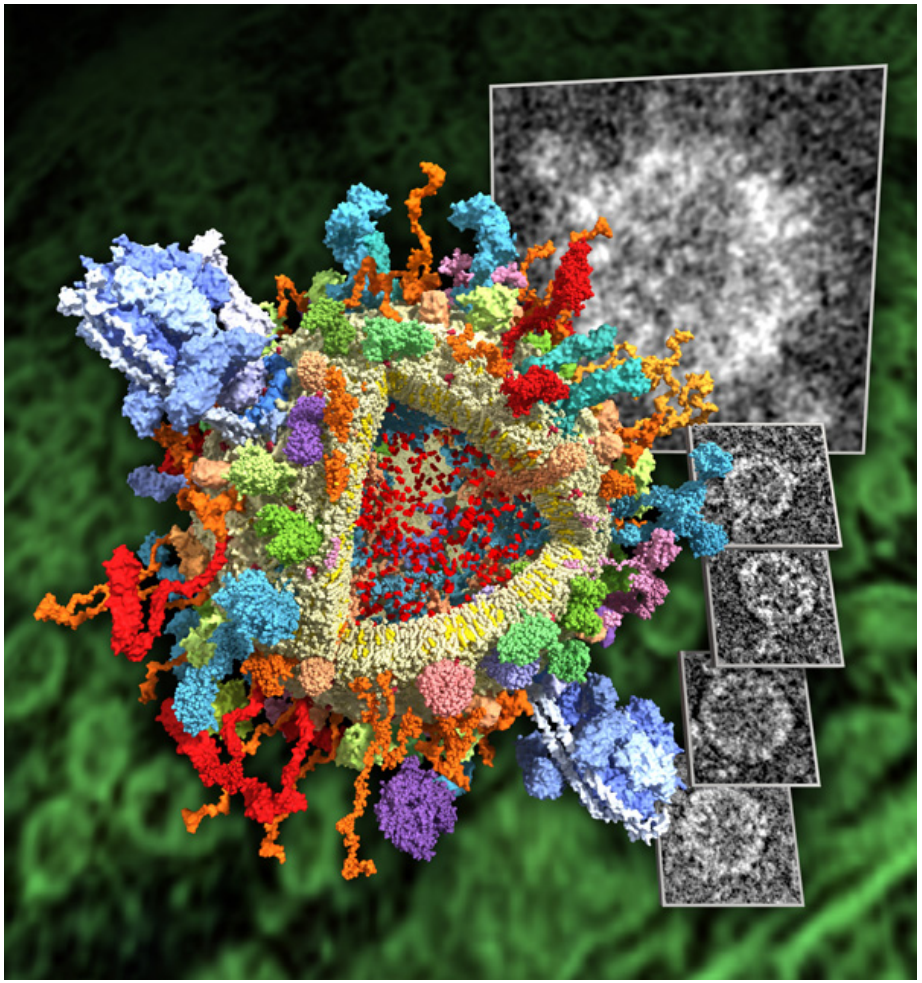
β_2 adrenergic receptor

Proteins (and other molecules) often come together to form *macromolecular complexes*



Nuclear Pore Complex
Alber et al., *Nature* 2007

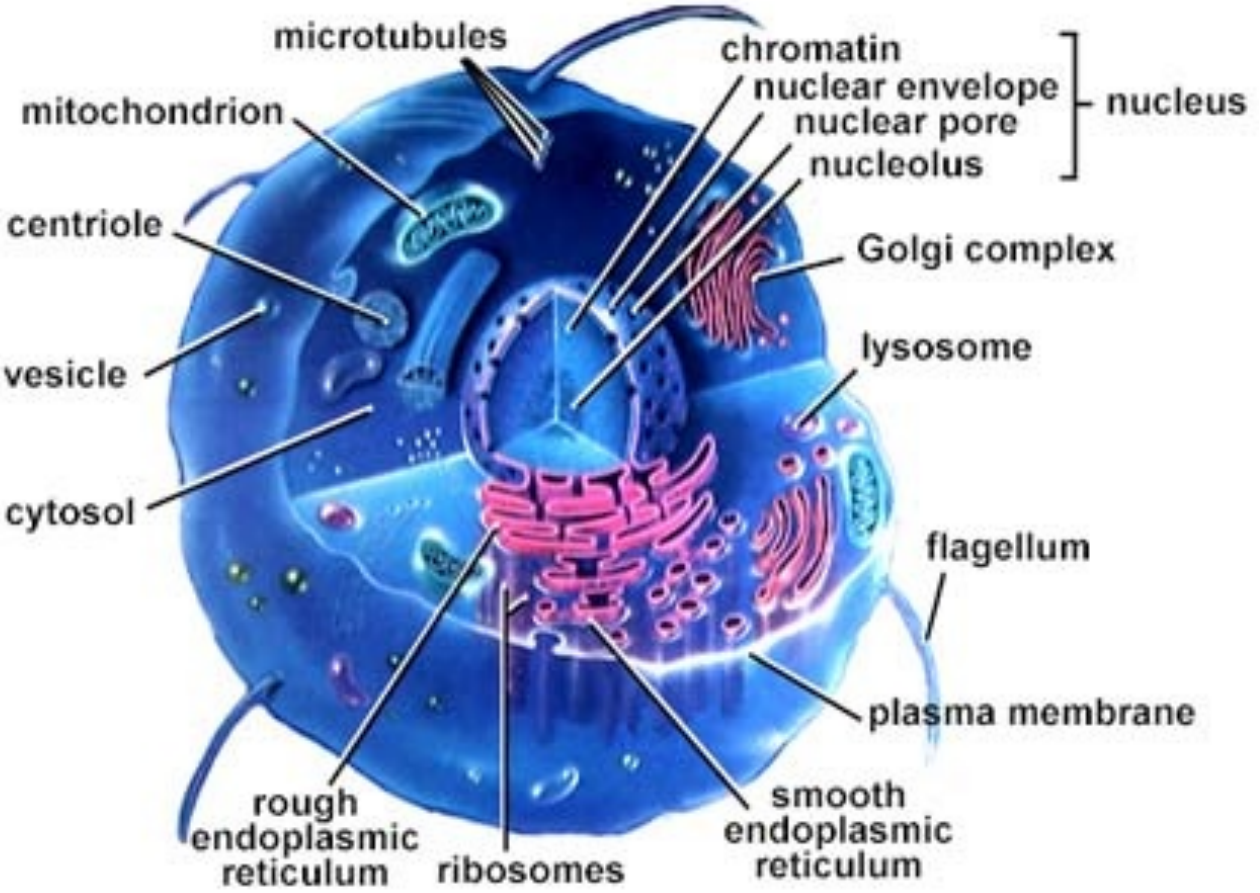
These come together to form organelles



Synaptic vesicle

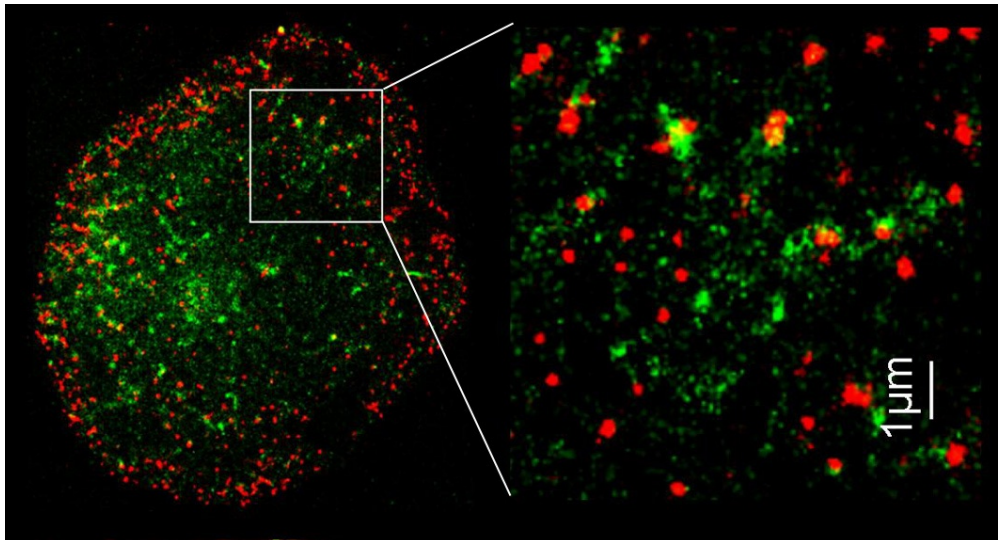
http://www.mpibpc.mpg.de/9547480/vesicle600.jpg

and cells

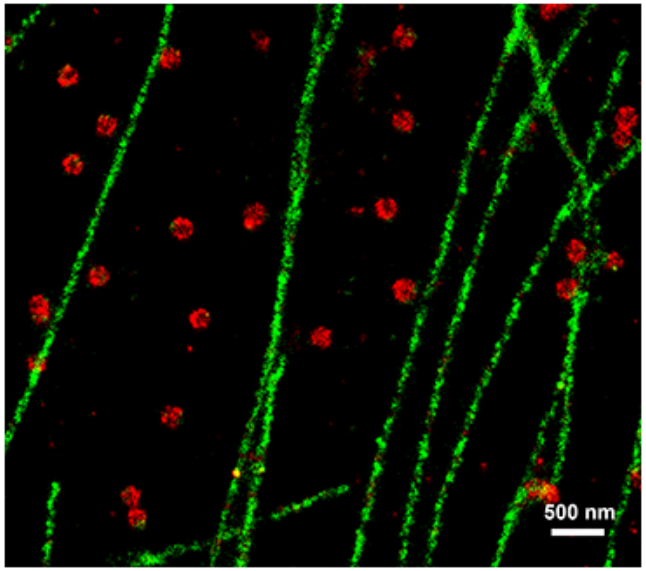


http://www.medfriendly.com/cell.html

Intracellular structure



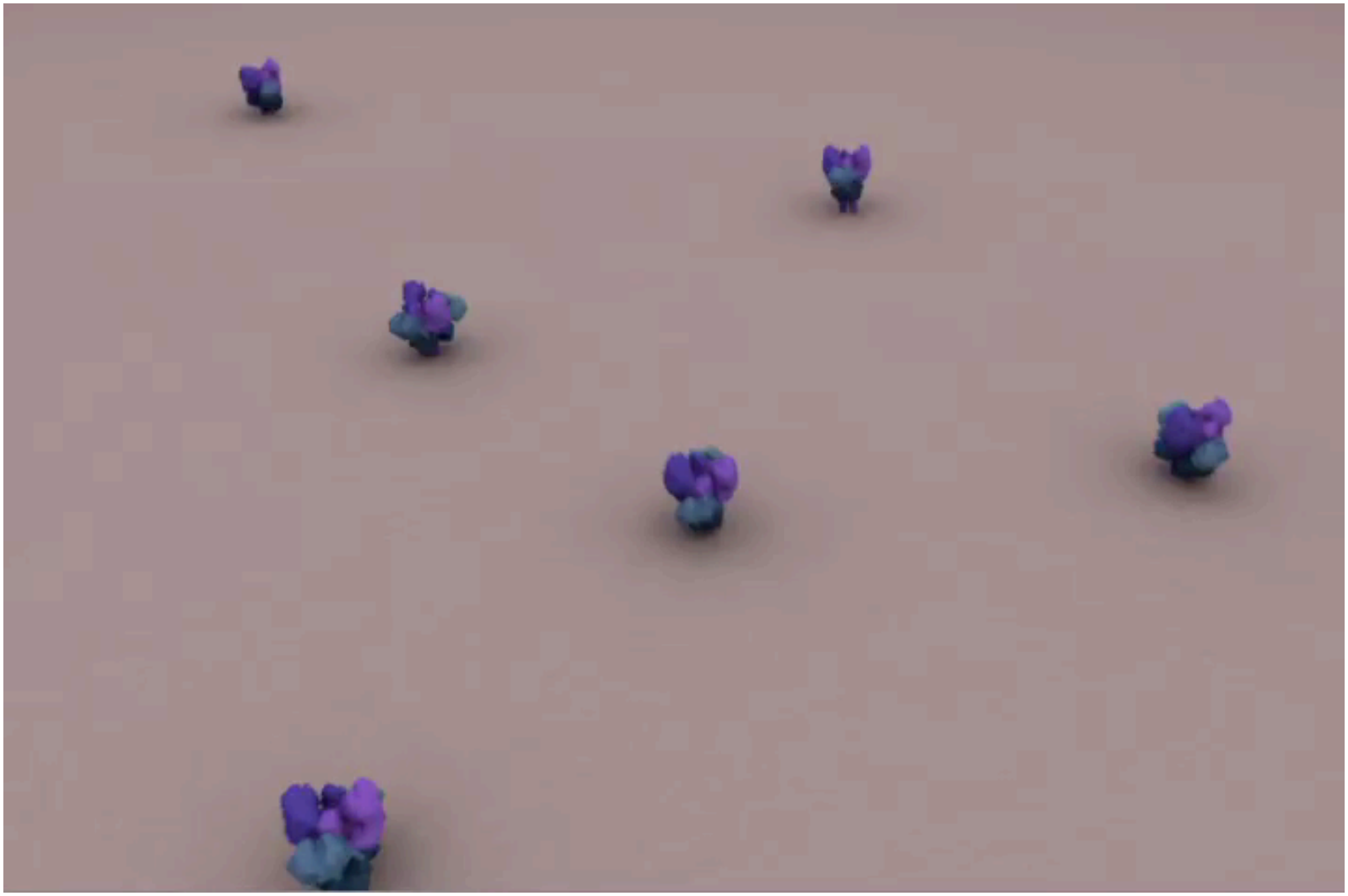
Chih-Jung Hsu, Janis Burkhardt and Tobias Baumgart



http://www.nikoninstruments.com/Products/Microscope-Systems/Inverted-Microscopes/N-STORM-Super-Resolution/(gallery); Zhuang group

David Goodsell

Intracellular dynamics (artist's rendition)



Janet Iwasa and Tomas Kirchhausen

Why is structure important?

Genomics is a great start ….

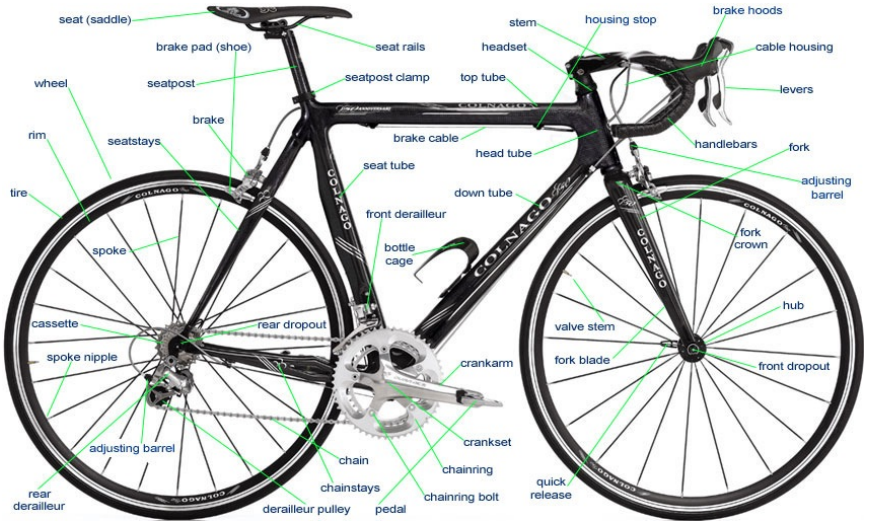
Track Bike — DL 175

REF. NO.	IBM NO.	DESCRIPTION
1	156011	Track Frame 21", 22", 23", 24", Team Red
2	157040	Fork for 21" Frame
2	157039	Fork for 22" Frame
2	157038	Fork for 23" Frame
2	157037	Fork for 24" Frame
3	191202	Handlebar TTT Competition Track Alloy 15/16"
4		Handlebar Stem, TTT, Specify extension
5	191278	Expander Bolt
6	191272	Clamp Bolt
7	145841	Headset Complete 1 x 24 BSC
8	145842	Ball Bearings
9	190420	175 Raleigh Pistard Seta Tubular Prestavalve 27"
10	190233	Rim, 27" AVA Competition (36H) Alloy Prestavalve
11	145973	Hub, Large Flange Campagnolo Pista Track Alloy (pairs)
12	190014	Spokes, 11 5/8"
13	145837	Sleeve
14	145636	Ball Bearings
15	145170	Bottom Bracket Axle
16	145838	Cone for Sleeve
17	146473	L.H. Adjustable Cup
18	145833	Lockring
19	145239	Straps for Toe Clips
20	145834	Fixing Bolt
21	145835	Fixing Washer
22	145822	Dustcap
23	145823	R.H. and L.H. Crankset with Chainwheel
24	146472	Fixed Cup
25	145235	Toe Clips, Christophe, Chrome (Medium)
26	145684	Pedals, Extra Light, Pairs
27	123021	Chain
28	145980	Seat Post
29		Seat Post Bolt and Nut
30	167002	Saddle, Brooks
31	145933	Track Sprocket, Specify 12, 13, 14, 15, or 16 T.

- But a parts list is not enough to understand how a bicycle works

Knowing the DNA sequence that encodes a protein is helpful, but is not enough to figure out everything about the protein's function

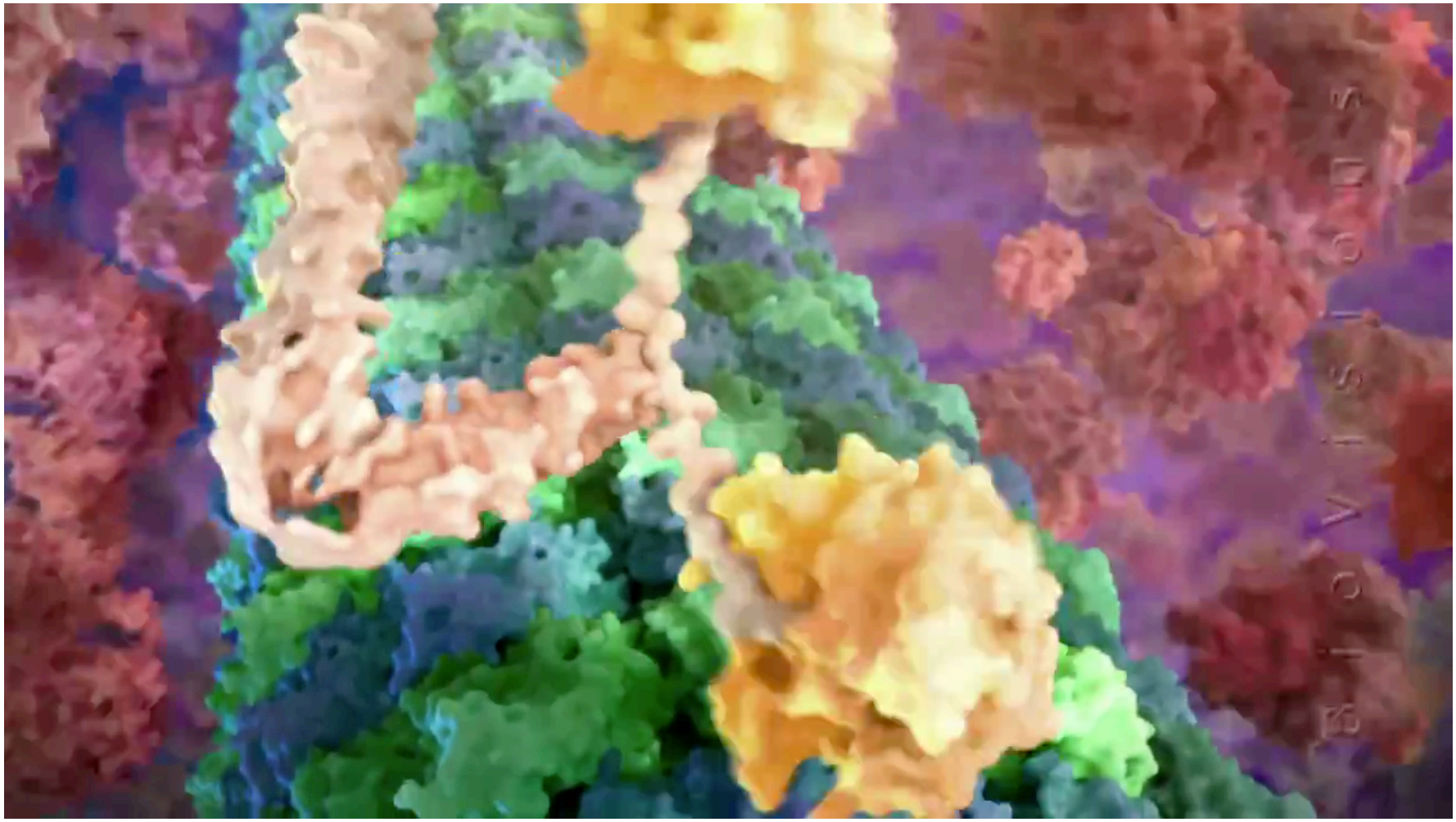
… but not the end



- We want the full spatiotemporal picture, and an ability to control it
- Broad applications, including drug design, medical diagnostics, chemical manufacturing, food production, and energy

Structure determines function

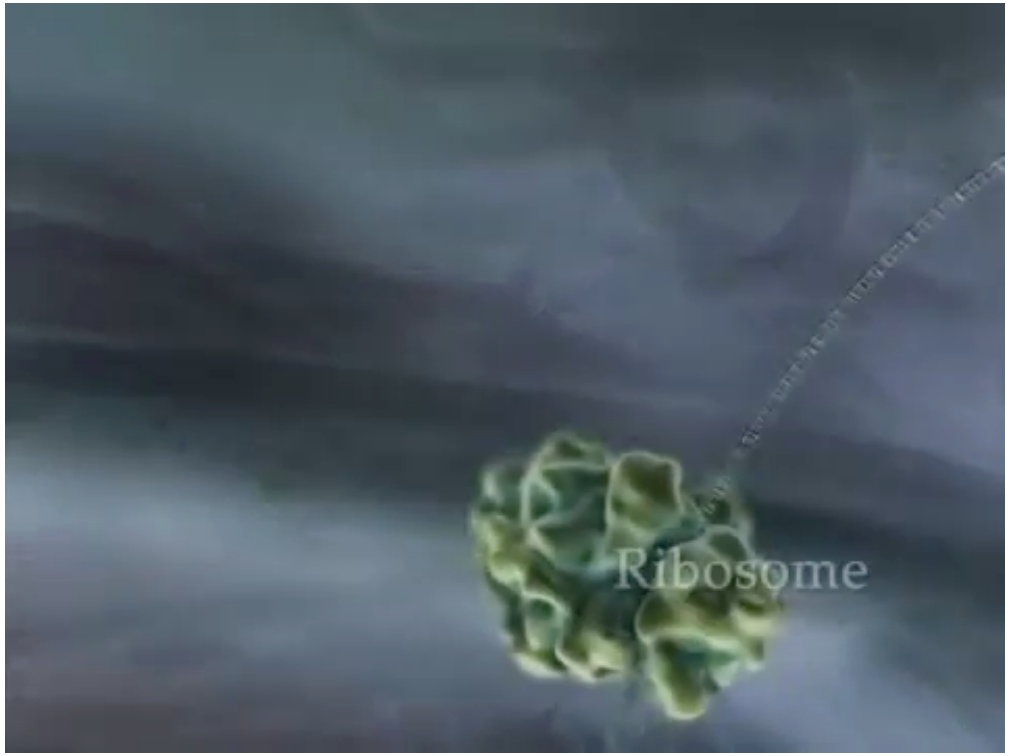
- Example: Motor protein (walks along microtubules, dragging load)



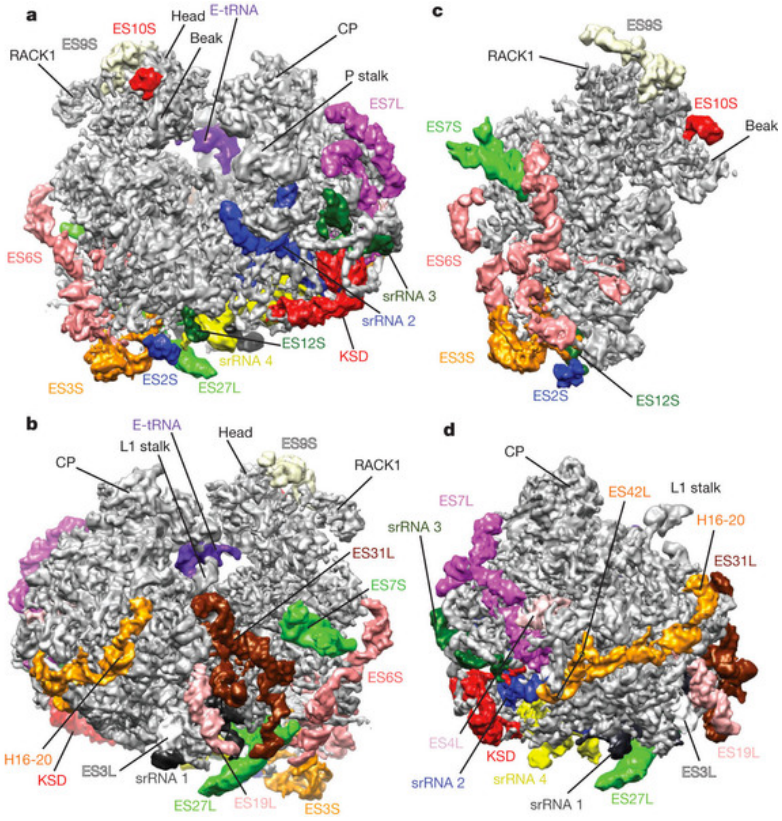
From *Inner Life of the Cell* | *Protein Packing*

Structure determines function

- Example: Ribosome
  - Complex of many proteins and RNAs that together makes new proteins (by reading the genetic code and combining amino acids)



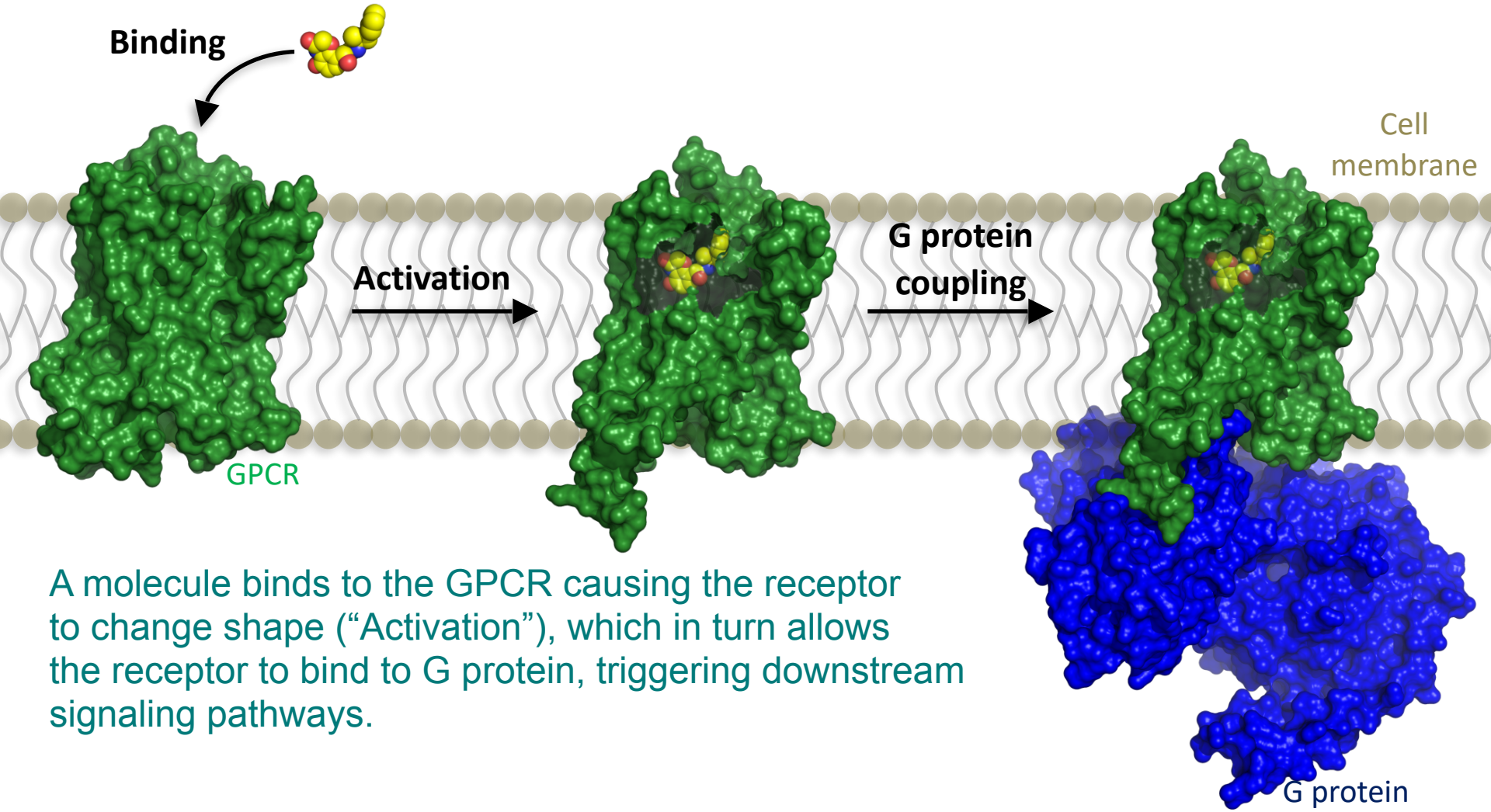
From *Inner Life of the Cell*, XVIVO and Biovisions @ Harvard



Hashem et al., Nature 494:385-9, 2013

Structure determines function

- Example: G protein–coupled receptors (GPCRs)
  - Largest class of human drug targets
  - Function: allow the cell to sense and respond to molecules outside it

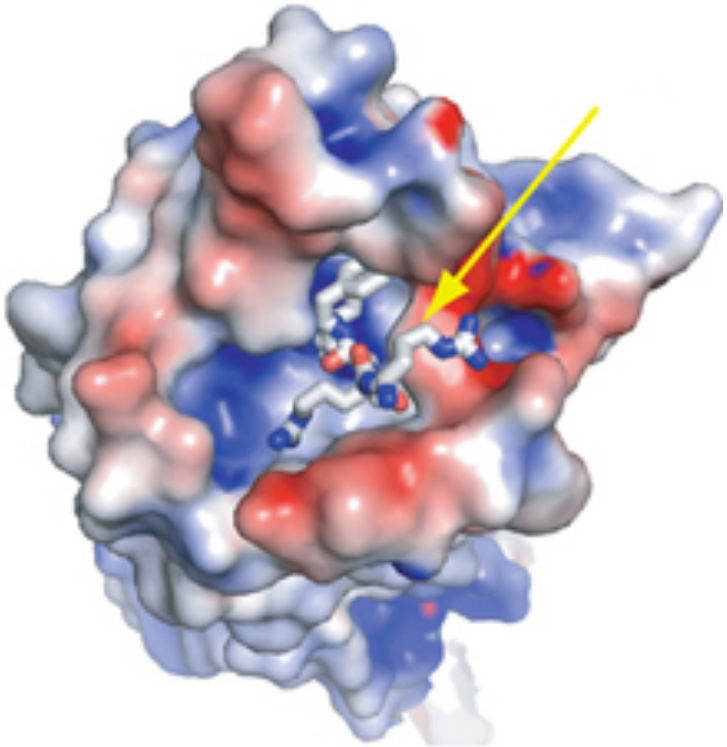


A molecule binds to the GPCR causing the receptor to change shape (“Activation”), which in turn allows the receptor to bind to G protein, triggering downstream signaling pathways.

Structure-based drug design

- Almost all drugs act by binding to proteins and altering their function
- Using knowledge of structures, we can design drugs that bind more tightly or more selectively, bind in different positions, alter behavior of protein in different ways, etc.

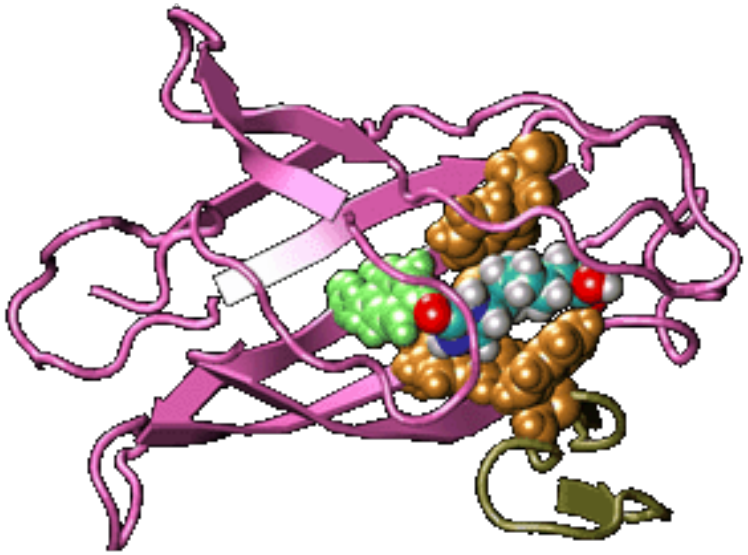
"selective binding": the drug binds to a certain protein but not to a different protein.



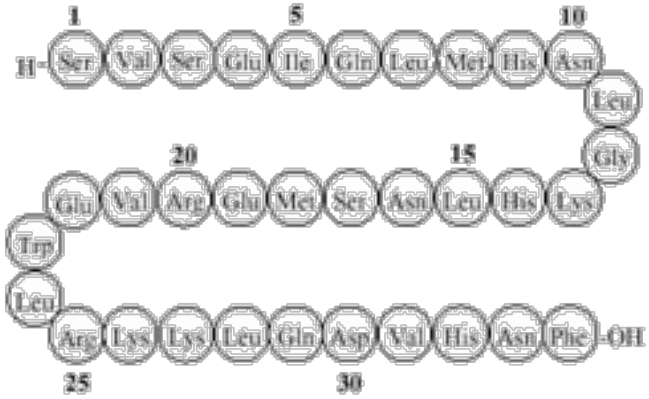
http://www.nih.gov/researchmatters/october2012/images/structure_l.jpg

Designing new biomolecular machines

- Protein design (for health or industrial applications)
- Cell design?



How?



http://zhanglab.ccmb.med.umich.edu/image/Protein_design.gif

How computation helps:
An overview of course topics

Protein structure prediction

- Sequence of amino acids → 3D coordinates
- Two basic approaches:
  - Homology modeling (infer structure from similar protein of known structure)
  - Ab initio prediction (using physics-based models)

Homology/template-based modeling is the most used one in practice -> use information from protein of known structure with similar sequence

Ab initio: start from the amino acid sequence, use physics to predict how the protein folds

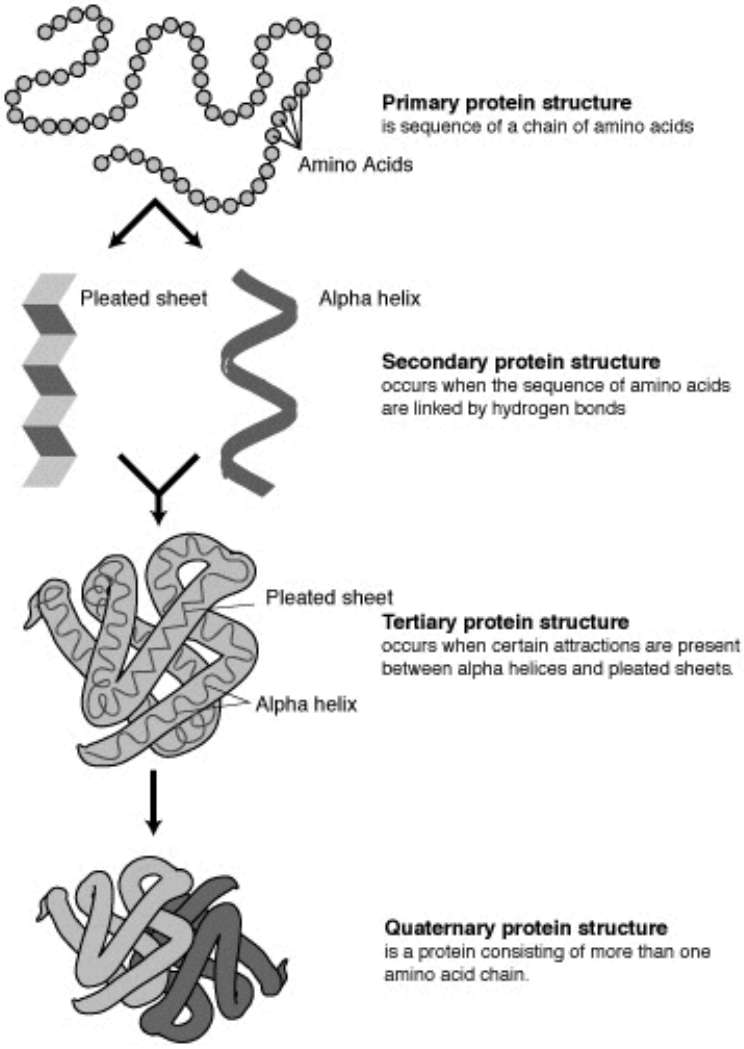
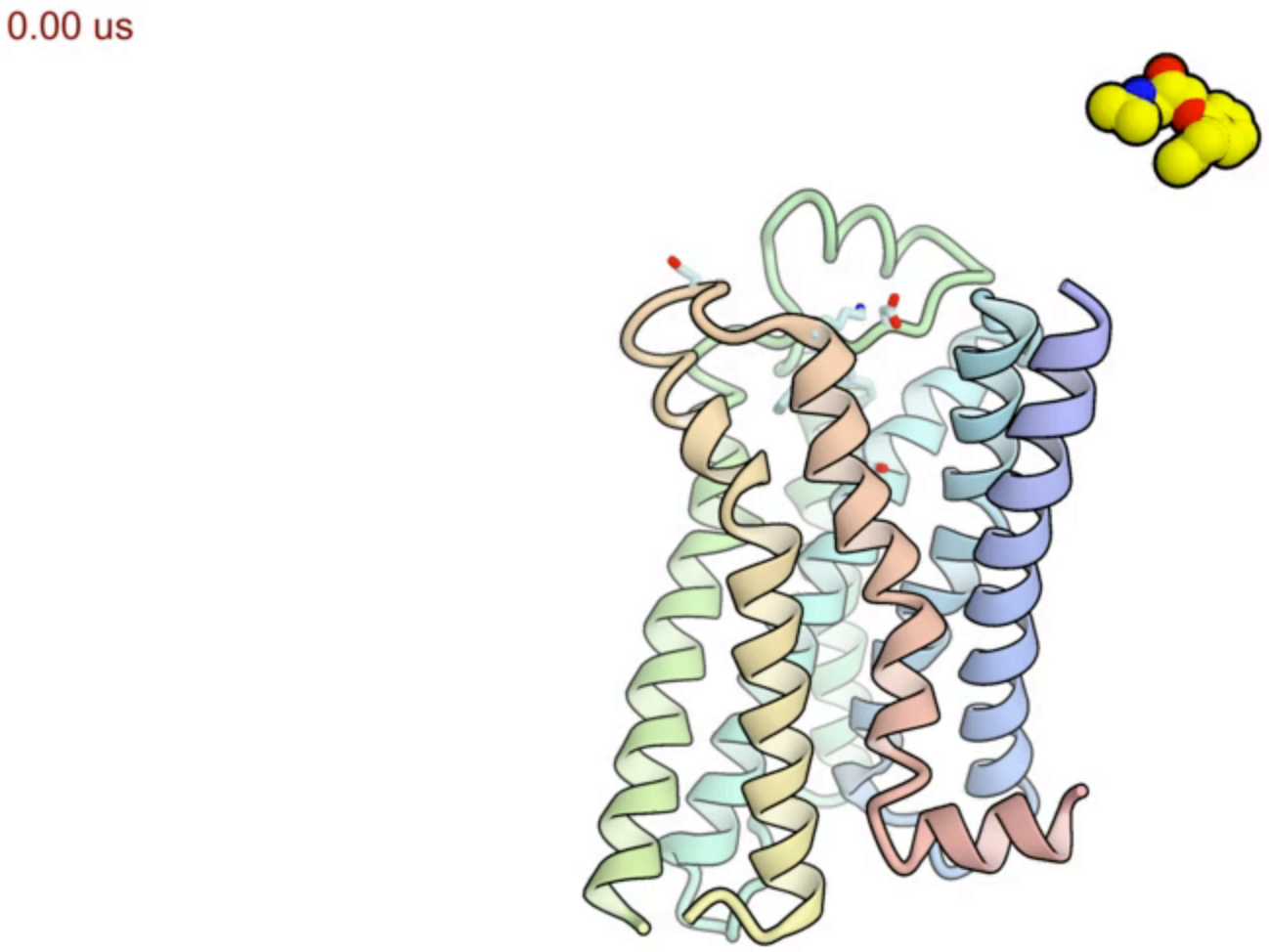


Image from Wikipedia

Protein structure is not static. It moves all the time.

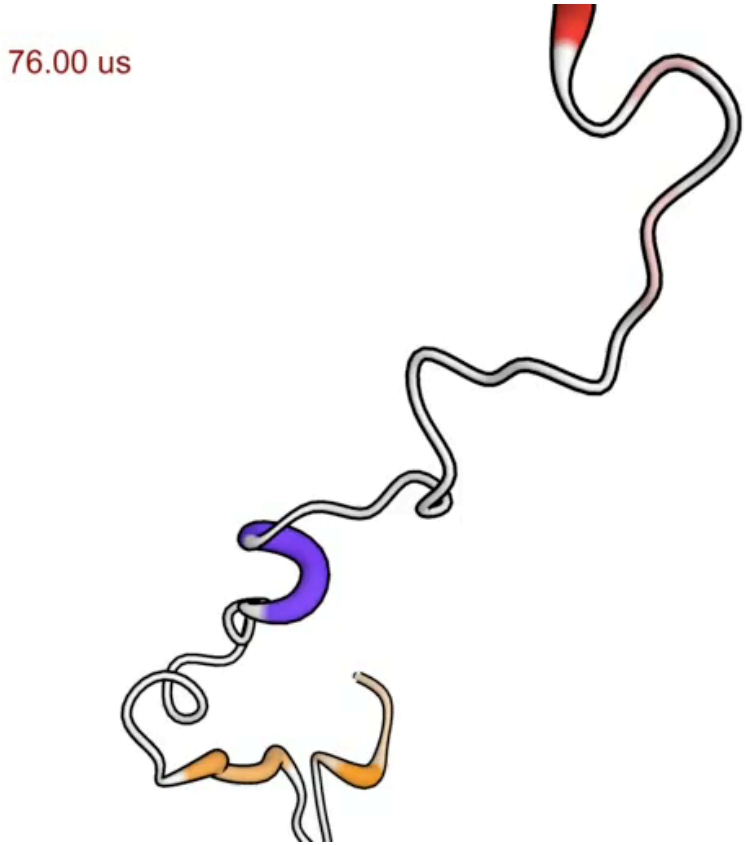
Molecular dynamics simulations



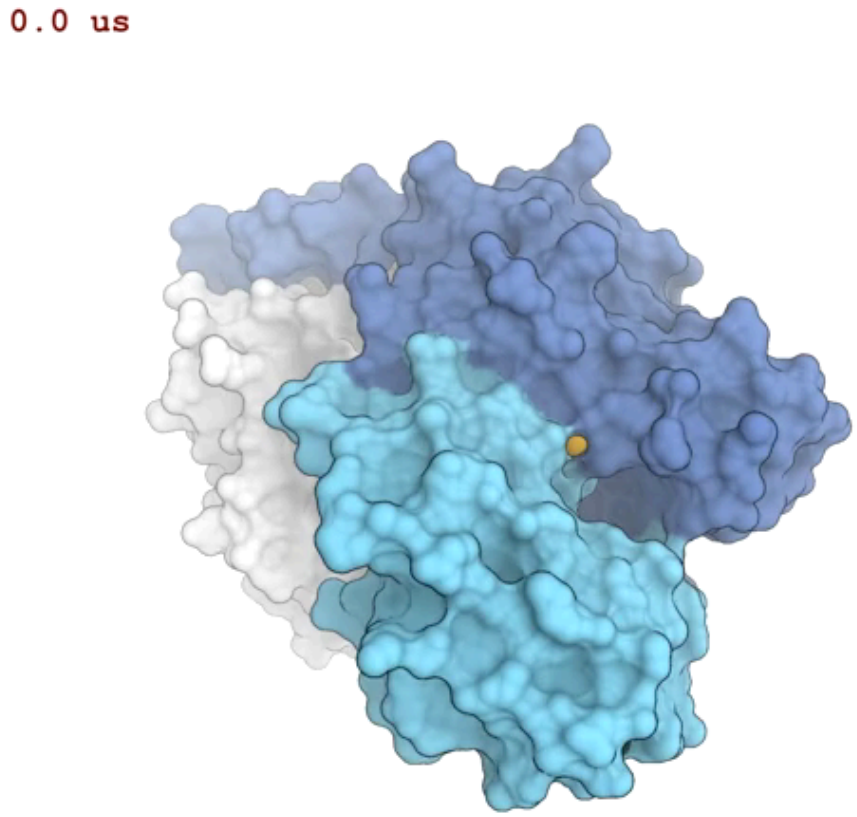
Beta-blocker binding to the β_2-adrenergic receptor

Dror et al., *PNAS* 2011

Molecular dynamics simulations



Folding of protein G
(Lindorff-Larsen et al., *Science*, 2011)

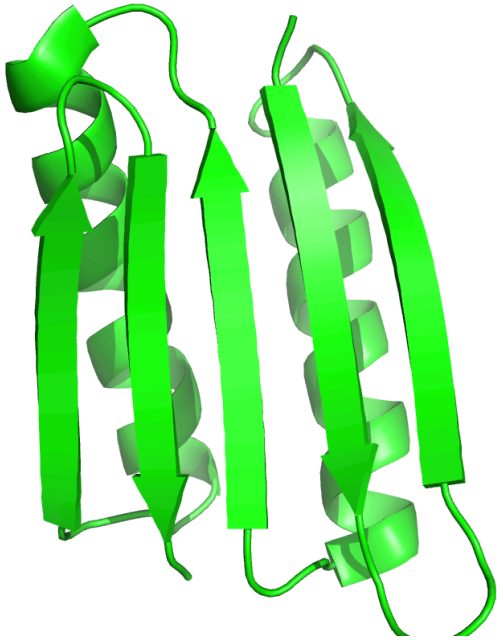


G-protein opens up, letting out the small molecule inside

Structural change in a
G protein (Dror et al., *Science* 2015)

Protein design

- Given a desired protein structure (or, in some cases, function), design the amino acid sequence that produces it



Top7, a protein with a designed fold
Kuhlman, Science 302:1364-8 (2003)

Ligand docking

Doing this experimentally is really difficult, nowadays, drug companies use “virtual screening” to screen for potential drug molecules computationally

Searching for potential drug molecules that bind to a target (usually a protein), and determine how they bind

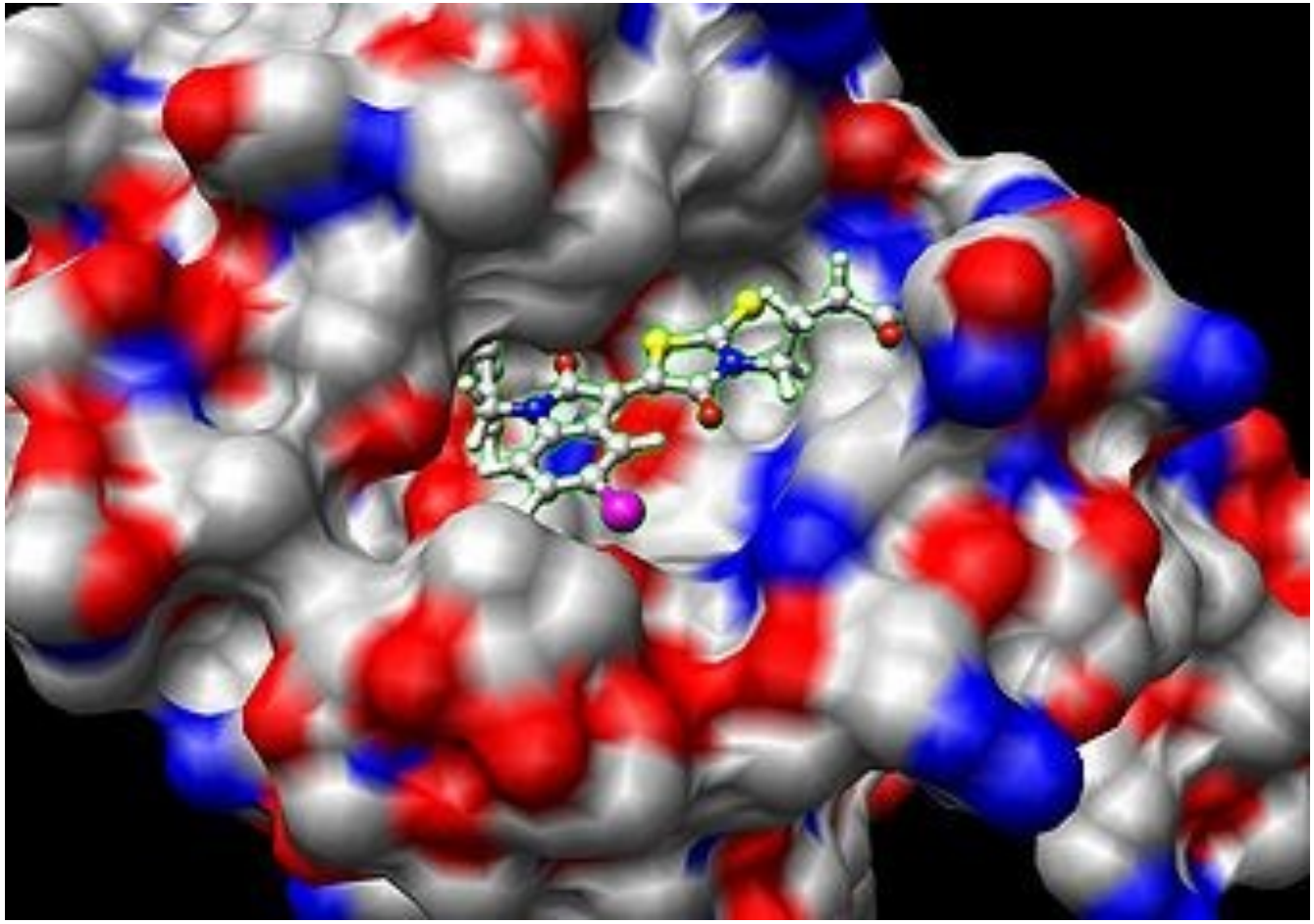
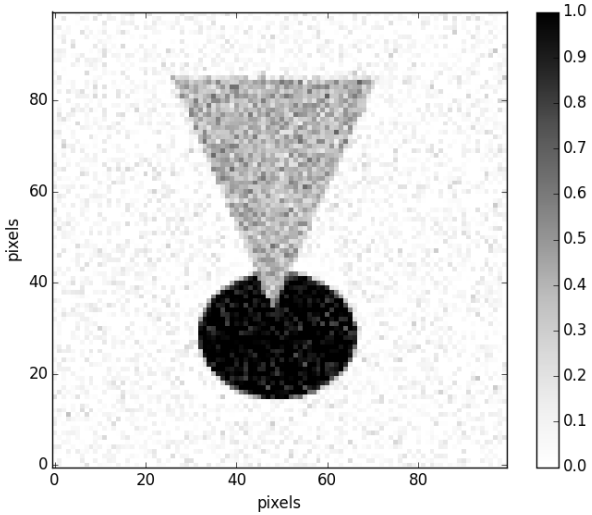


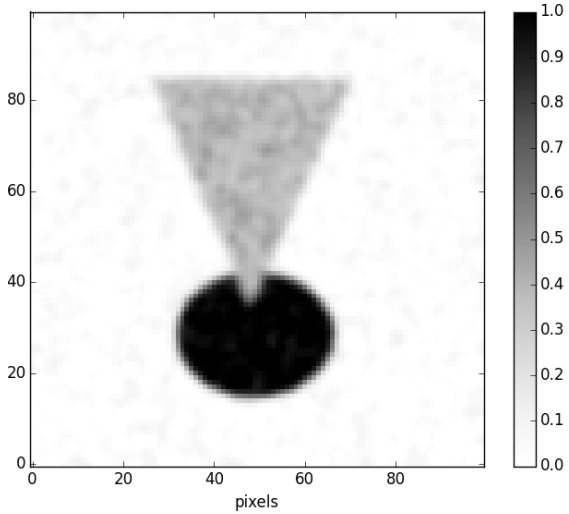
Image: Wikipedia

Image analysis

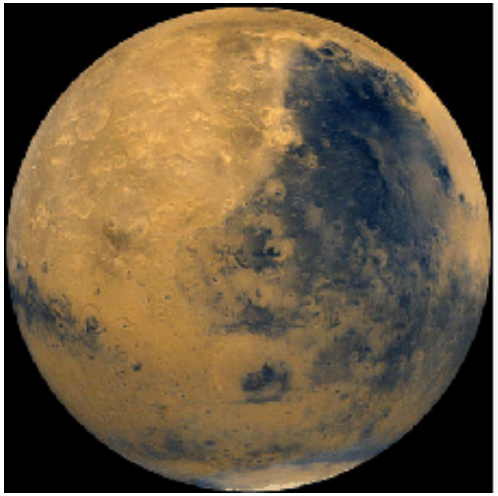
Original image



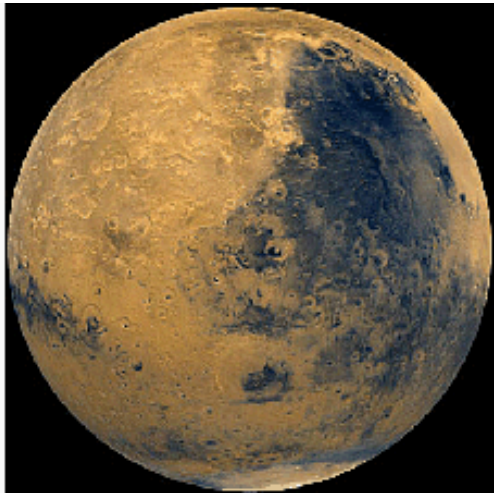
Denoised image



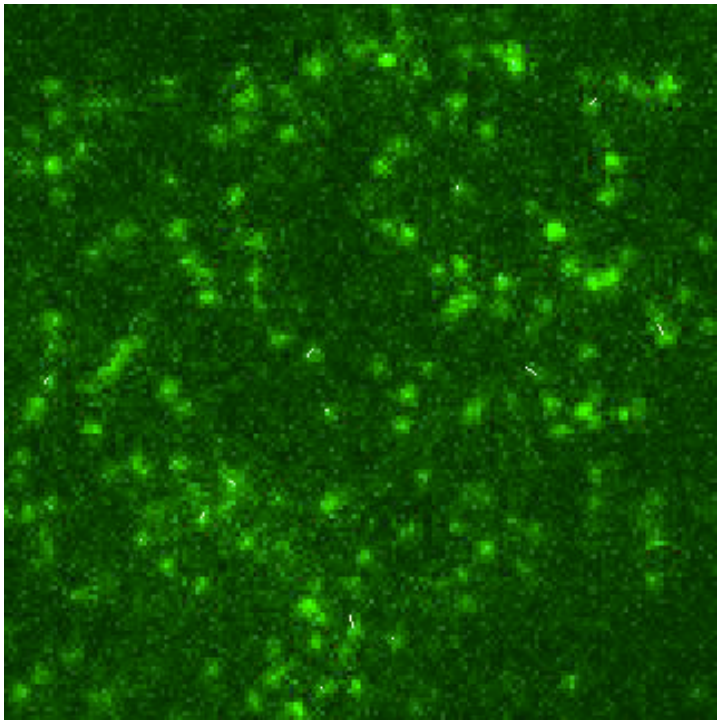
Original image



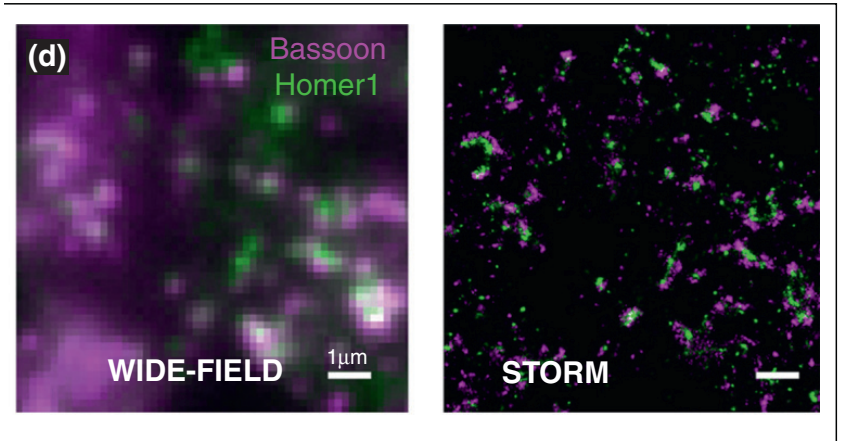
Sharpened image



Fluorescence microscopy and cellular-level organization



Data: Bettina van Lengerich, Natalia Jura
Tracking and movie: Robin Jia

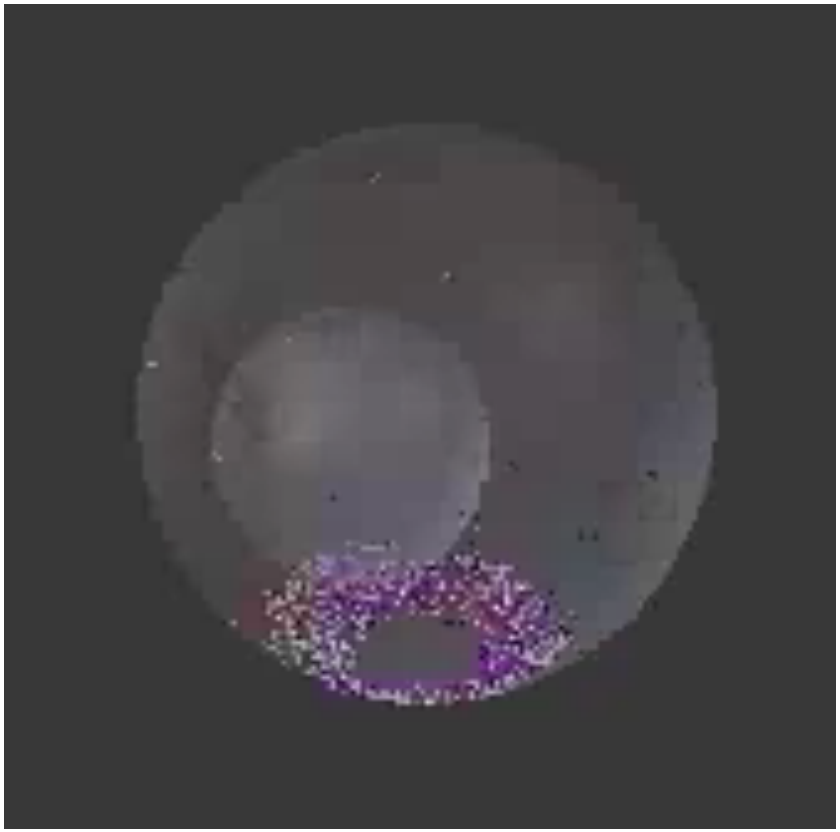


Sigrist & Sabatini, Current Opinion in Neurobiology 22:1-8, 2011

These types of microscopies take a bunch of images and use some clever computational techniques to produce the final image

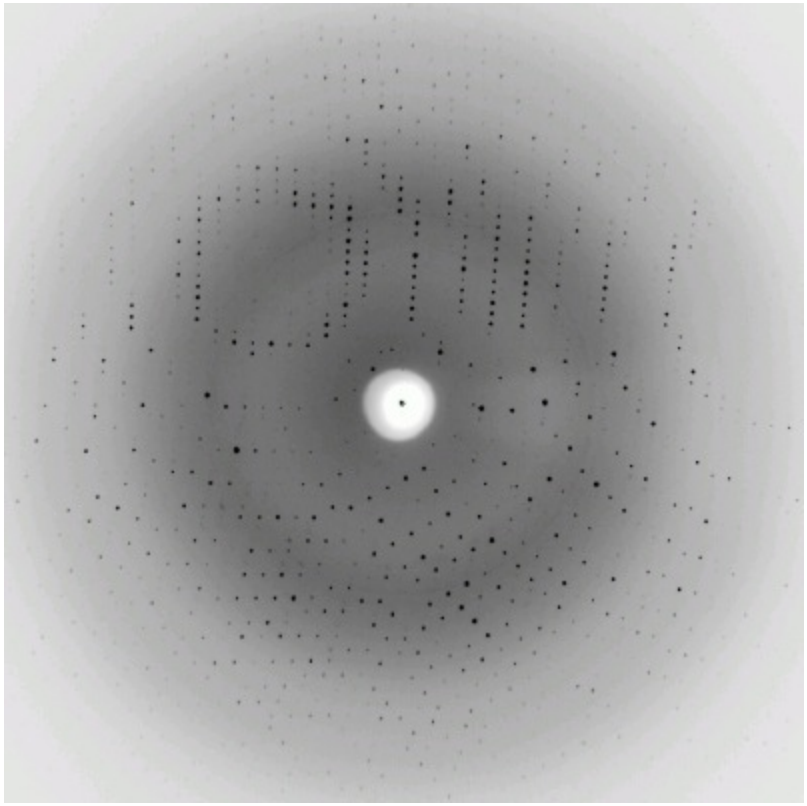
- Including super-resolution microscopy

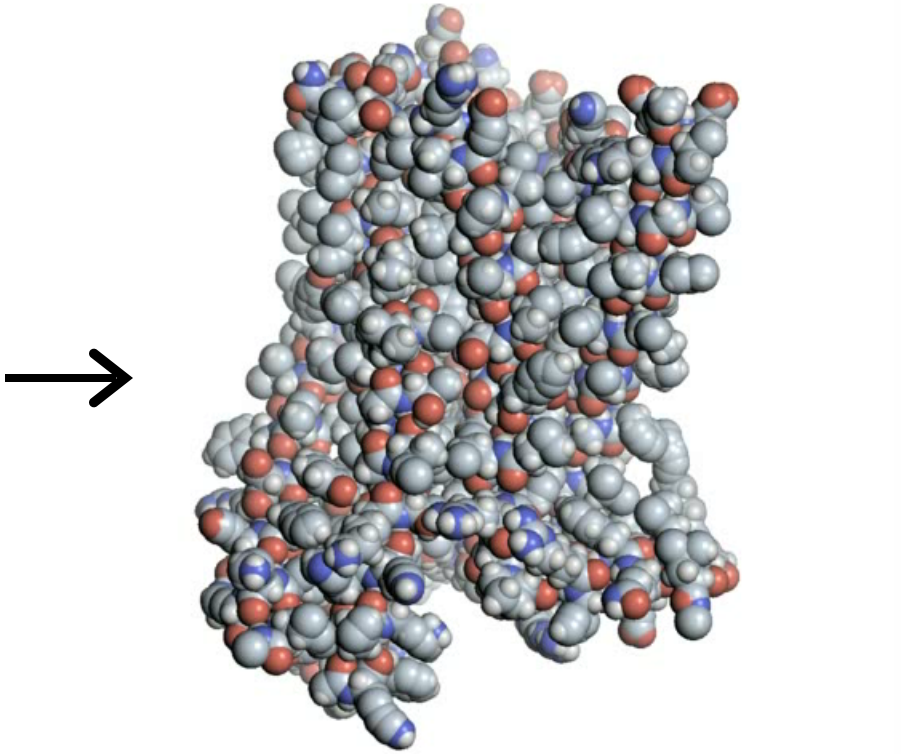
How molecules move about a cell: diffusion and cellular-level simulation



Video: Naomi Latorraca

Solving structures by x-ray crystallography





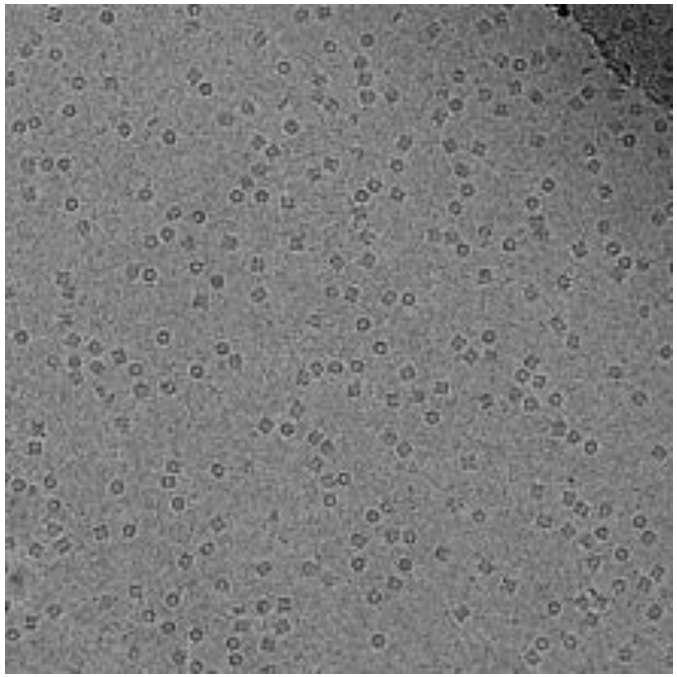
X-ray diffraction pattern

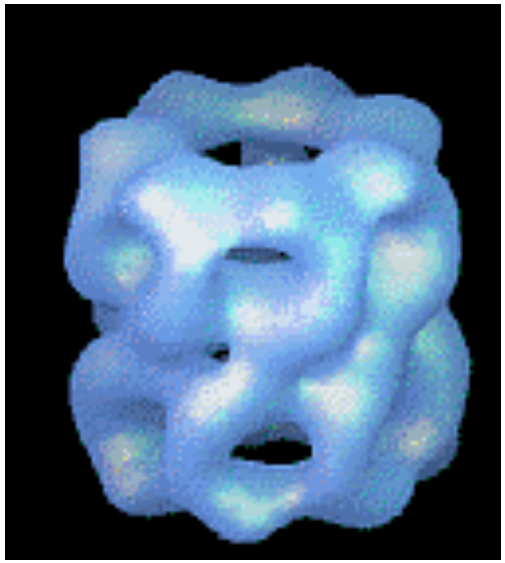
Image: http://www.chem.ucla.edu/harding/IGOC/X/x_ray_crystallography.html

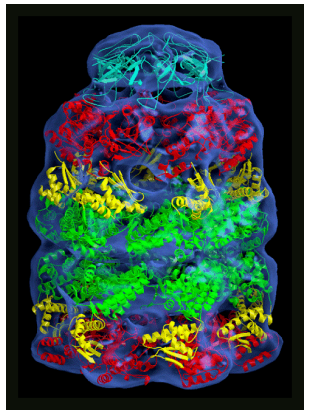
Protein structure

Solving structures by single-particle electron microscopy (cryoelectron microscopy)

Combining a bunch of 2D low resolution images viewed from different angles to produce the 3D rendering







CryoEM image

Reconstructed envelope

Image from Wikipedia

http://people.cryst.bbk.ac.uk/~ubcg16z/chaperone.html

Deducing genomic structure (i.e., the structure of chromosomes)

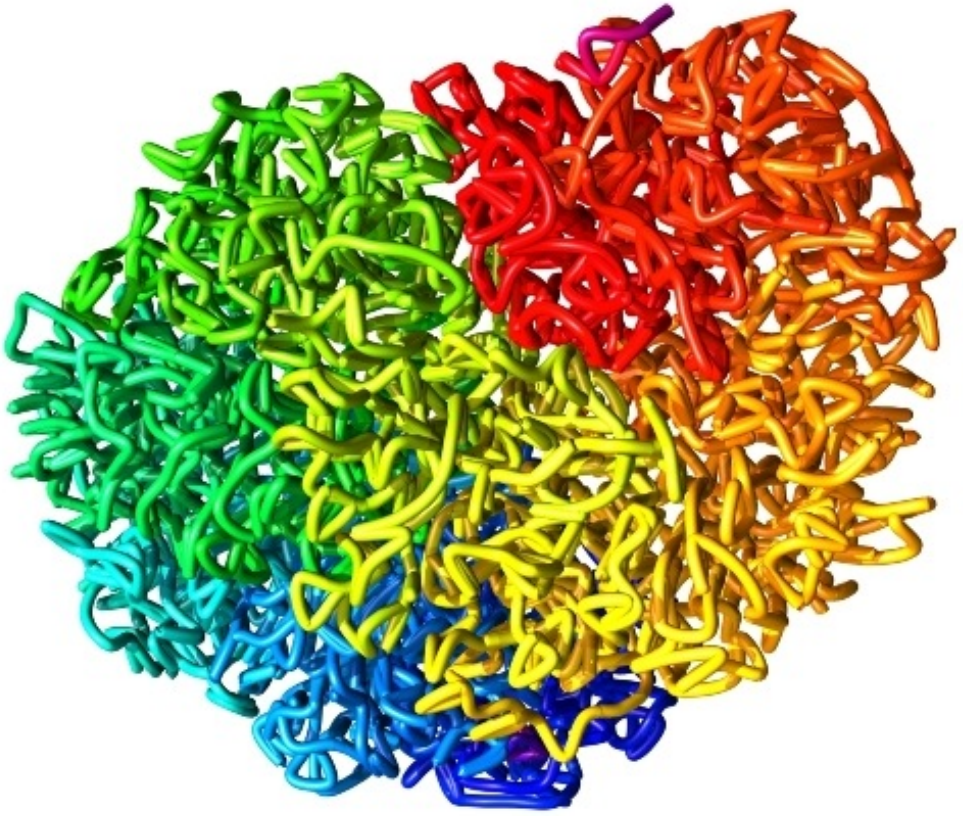
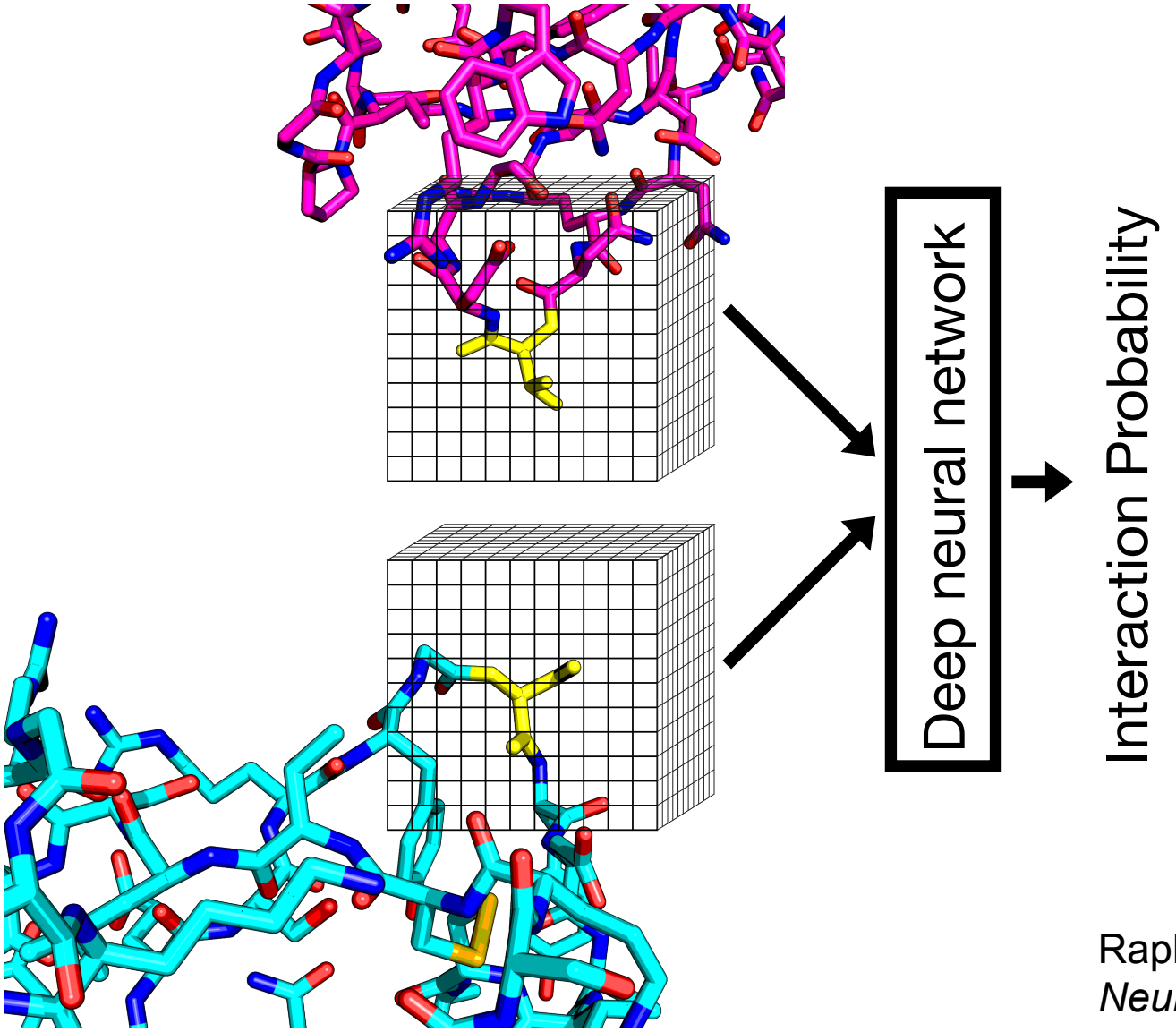


Image:http://www.biotechniques.com/news/Bringing-genome-structure-into-focus/biotechniques-312407.html

Machine learning in structural biology



Raphael Townshend et al.,
NeurIPS 2019

Recurrent themes

Recurrent themes

- Similarities and differences in methods employed at **different spatial scales**
- **Physics-based approaches** (modeling based on first-principles physics) vs. **data-driven approaches** (machine learning based on experimental data)
- Computation plays important role both in **structural interpretation of experimental data** and in **structural predictions in the absence** of such data
- **Energy functions** (which associate an energy or potential with each possible structure)
- Recurring math concepts: **Fourier transforms, convolution, Monte Carlo methods**

Course organization

Fine-scale ➔ Coarse-scale (roughly)

- Atomic-level modeling of proteins (and other macromolecules)
  - Biomolecular structure (including proteins)
  - Energy functions and their relationship to protein conformation
  - Molecular dynamics simulation
  - Protein structure prediction
  - Protein design
  - Ligand docking
- Coarser-level modeling and imaging-based methods
  - Fourier transforms and convolution
  - Image analysis
  - Microscopy
  - Diffusion and cellular-level simulation
  - X-ray crystallography
  - Single-particle electron microscopy
  - Genome structure

Focus will be on fundamentals, but most lectures will also cover topics of current research

Course logistics

Course website

- http://cs279.stanford.edu/
- See “Evaluation criteria and course policies” document on website
- Link to website from last year’s course, which includes all lecture slides
  - This year’s content will be similar but not identical
- **Please sign up on Piazza (via link on webpage) so that you get announcements**
  - Complete polls (office hours and operating systems)

Expected background

- Course is intended to be broadly accessible to students with *either* computational or biological backgrounds

- Assignments involve basic programming in Python.
  - You need not have used Python before. You should have done some programming (in any language) before.
  - Python tutorial: see website for time. It will be recorded so that you'll be able to view it later as well.

- You should have some previous exposure to biology, chemistry, and physics (at least in high school)

- You should have studied math through elementary calculus
  - We will teach some additional relevant math concepts (e.g, Fourier transforms), with a focus on basic ideas/intuition rather than on equations

Assignments, Project, Final assessment

- Three assignments
  - First one is substantially shorter than second and third.
- Project: More open-ended. A bit more work than second and third assignments.
- Final assessment covering key concepts

Lectures and reading

- Lectures will be recorded and available on Canvas
- No textbook. Slides available (with annotation), along with additional notes for some lectures and pointers to optional reading material
- Class participation encouraged!
  - Use the “raise hand” feature in Zoom
  - Extra credit for asking questions in class or answering questions on Piazza

Feedback welcome!

- I want to continue improving this course, and would appreciate your suggestions
- Please speak up when you don't understand something
  - Or ask on Piazza

Course staff

- Prof. Ron Dror
  - http://drorlab.stanford.edu/rondror.html
  - Office hours: Right after each class (until 4:30)
- TAs:
  - Raphael Townshend
  - Cynthia Hao
  - Daniel Tang
  - Patricia Suriana

  - Office hours and contact info at cs279.stanford.edu