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

Image credit:
Ansgar Philippsen

Sept 26, 2017

Ron Dror

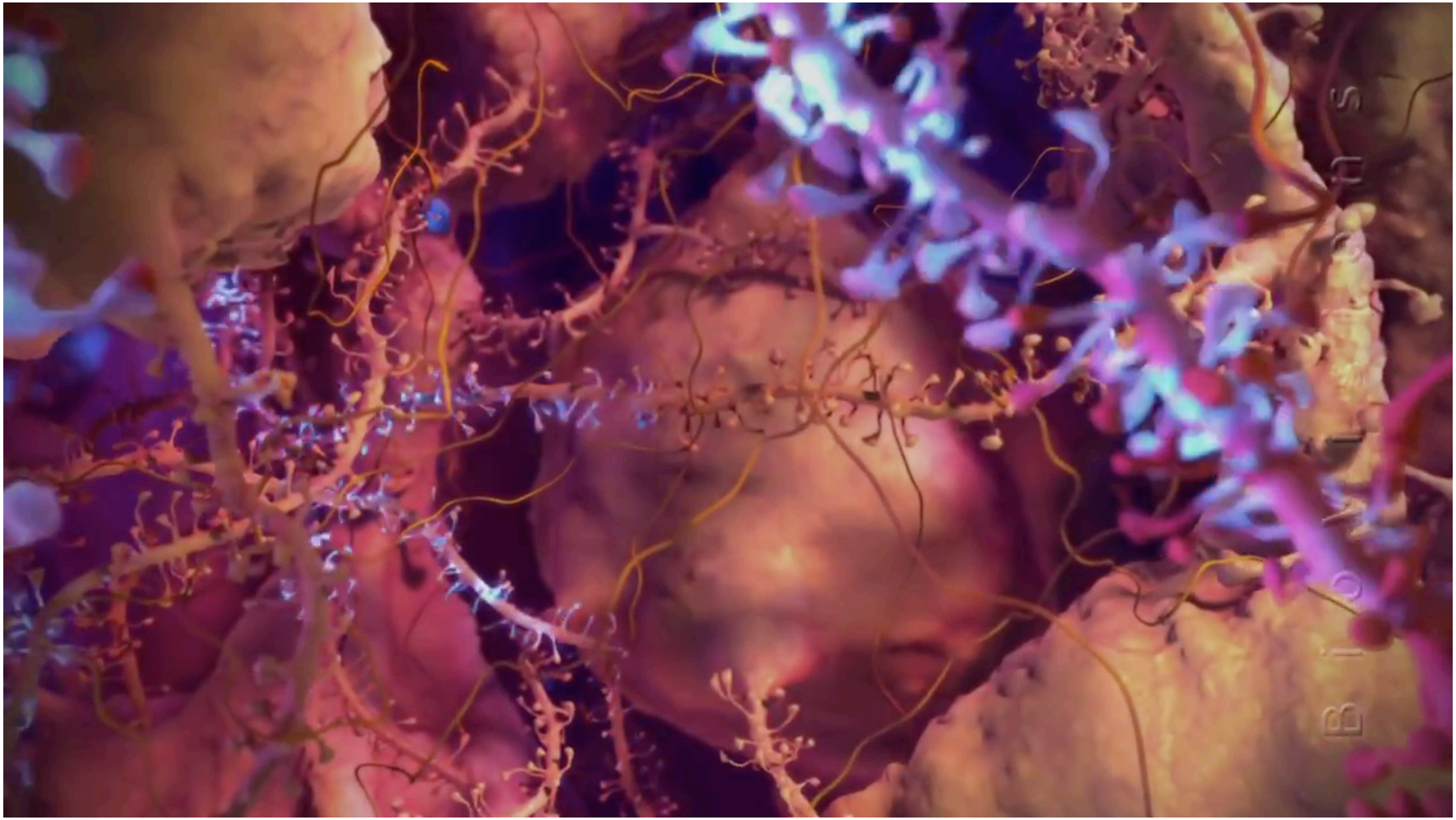
Outline for lecture 1 (course overview)

- 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

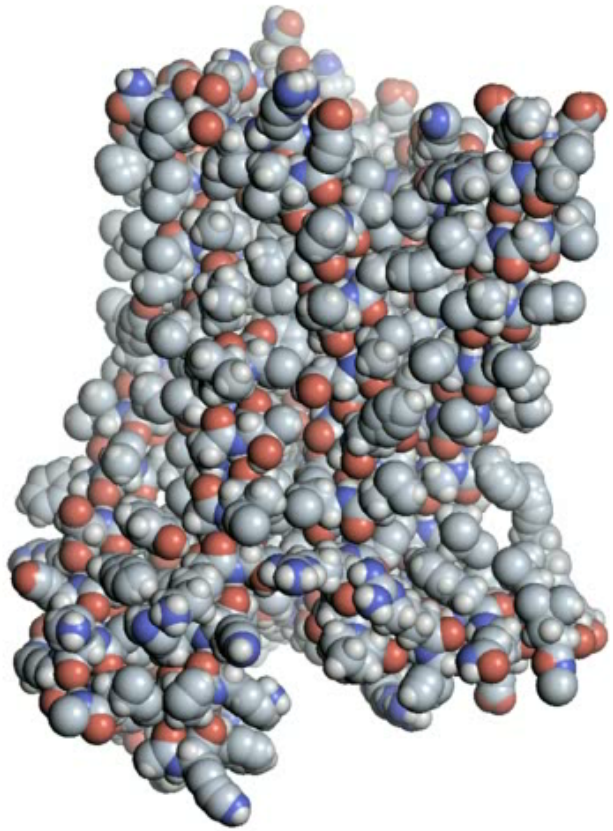


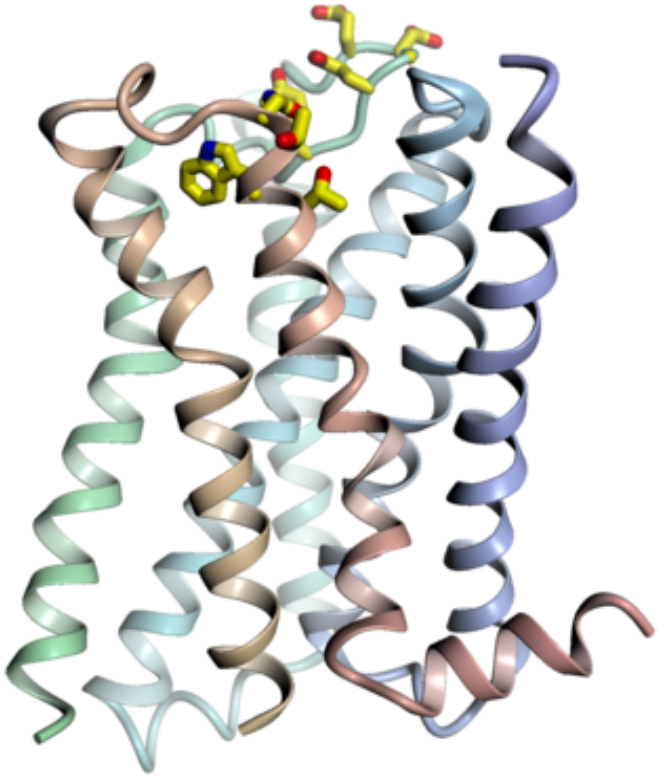
From *Inner Life of the Cell* | *Protein Packing*, XVIVO and Biovisions @ Harvard

What is structure?

Structure (and dynamics)
at multiple spatial scales

Protein structure



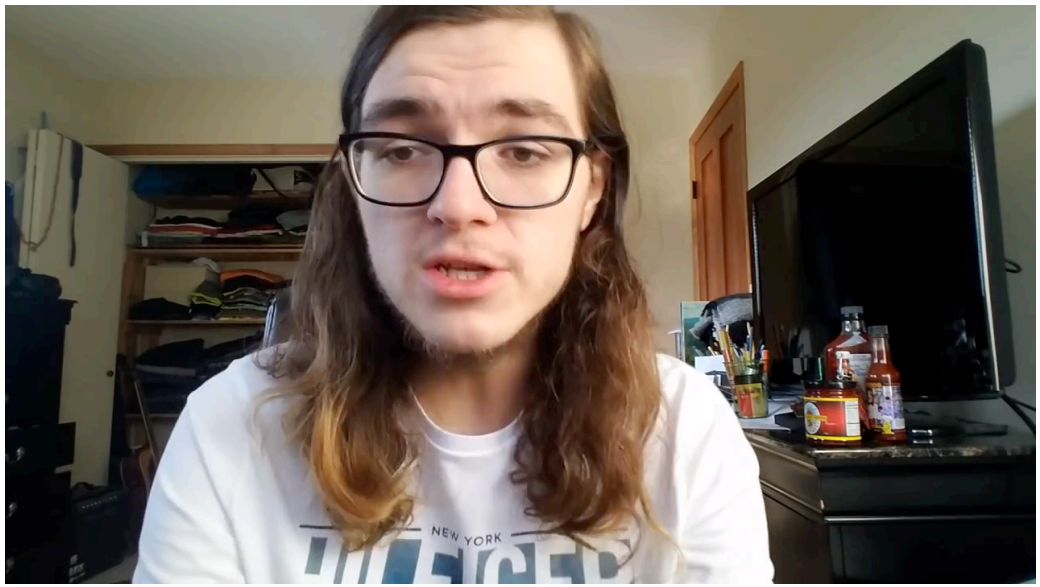


An adrenaline receptor
(the β_2 adrenergic receptor)

Example: how LSD binds to its target

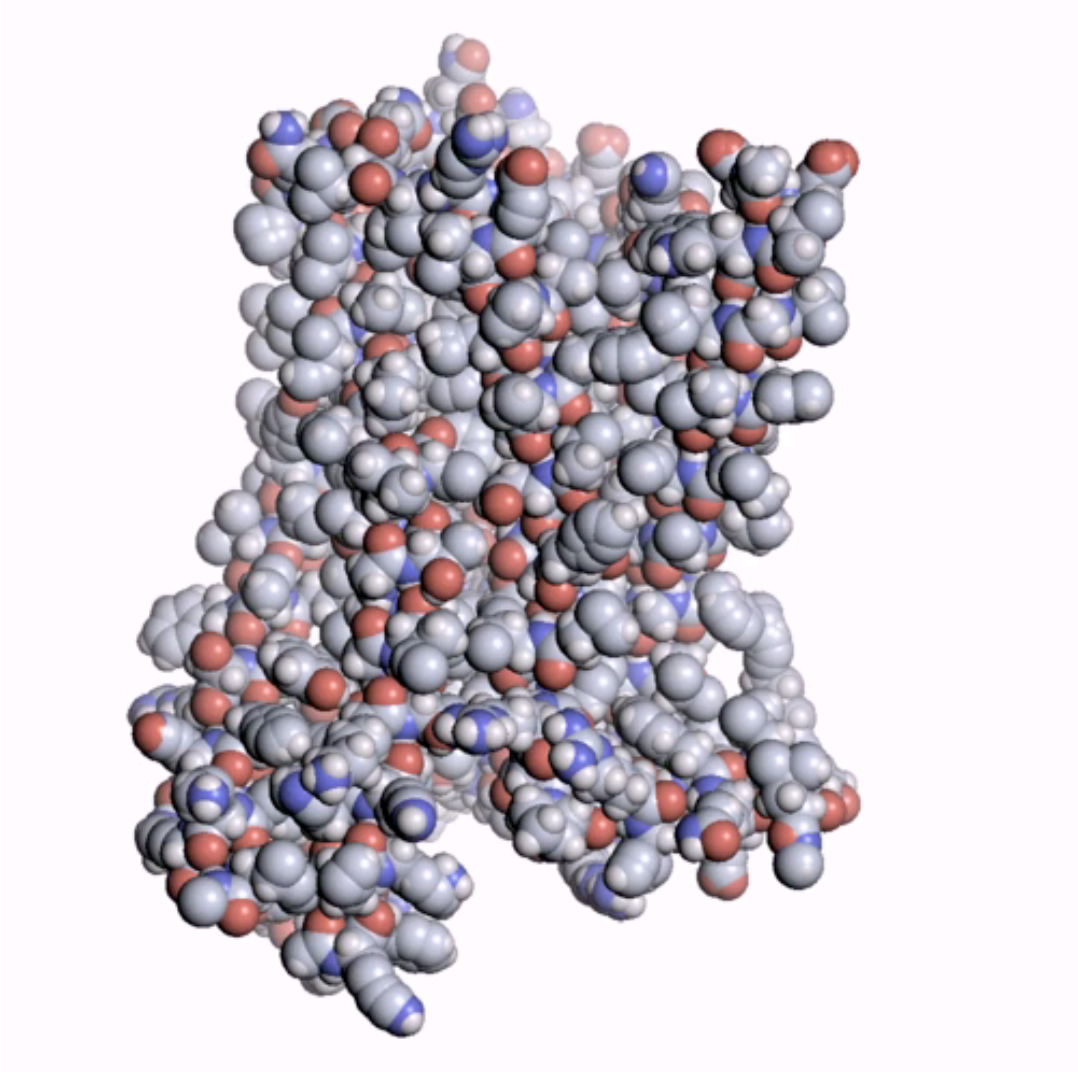
Wacker et al., *Cell* 168:377, 2017
Collaboration with Bryan Roth (UNC)

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



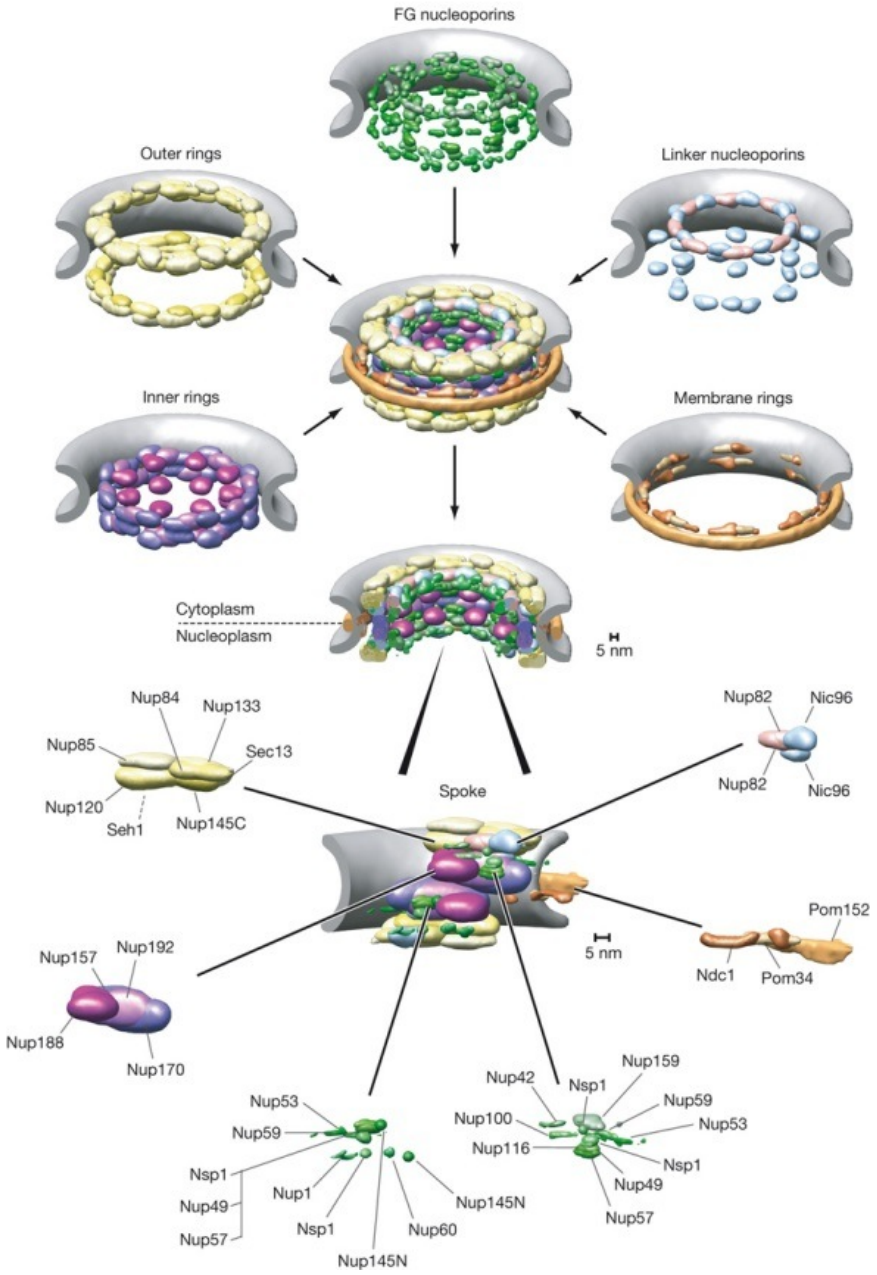
https://www.youtube.com/watch?v=LjumHvnl-ME&feature=youtu.be

Protein dynamics



β_2 adrenergic receptor

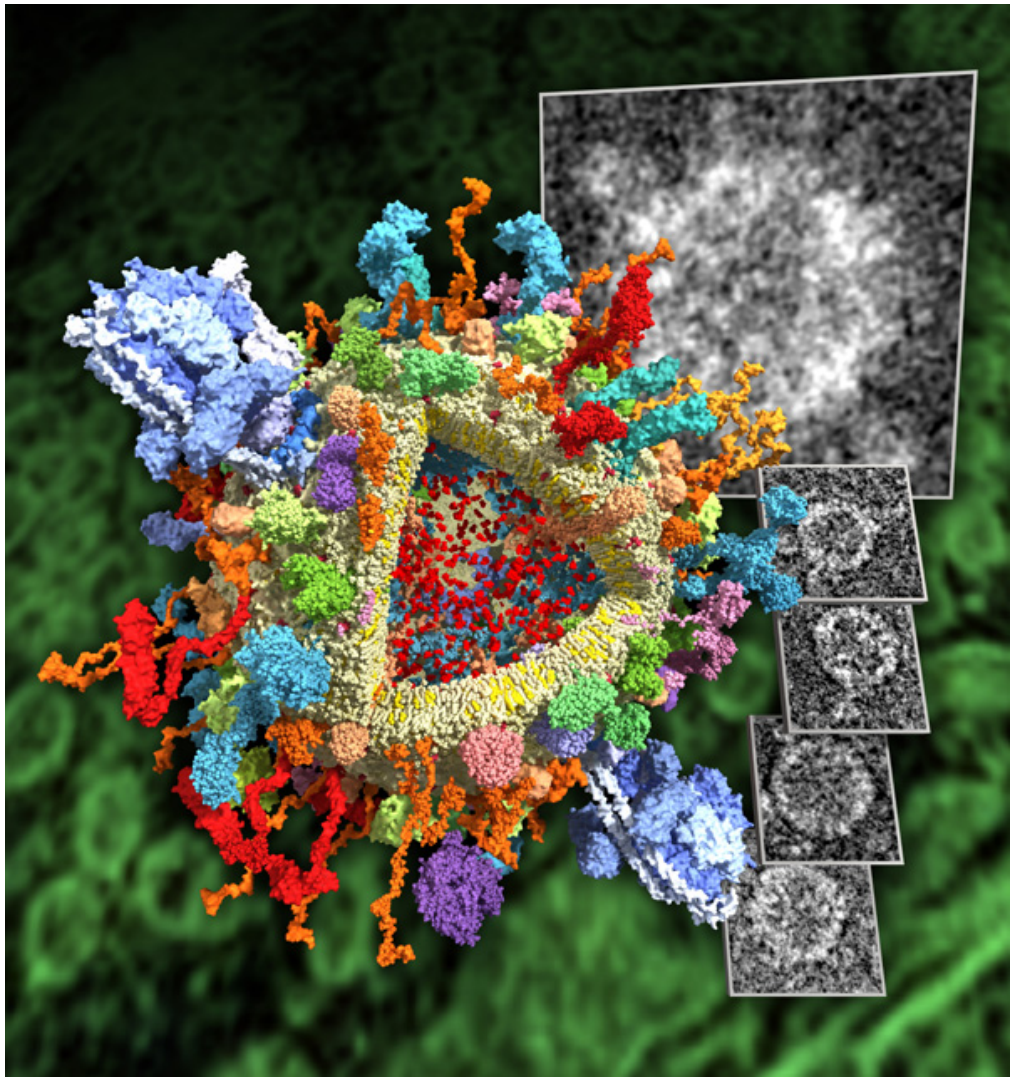
Proteins (and other molecules) often come together to form *macromolecular complexes*



Nuclear Pore Complex

Alber et al., *Nature* 2007

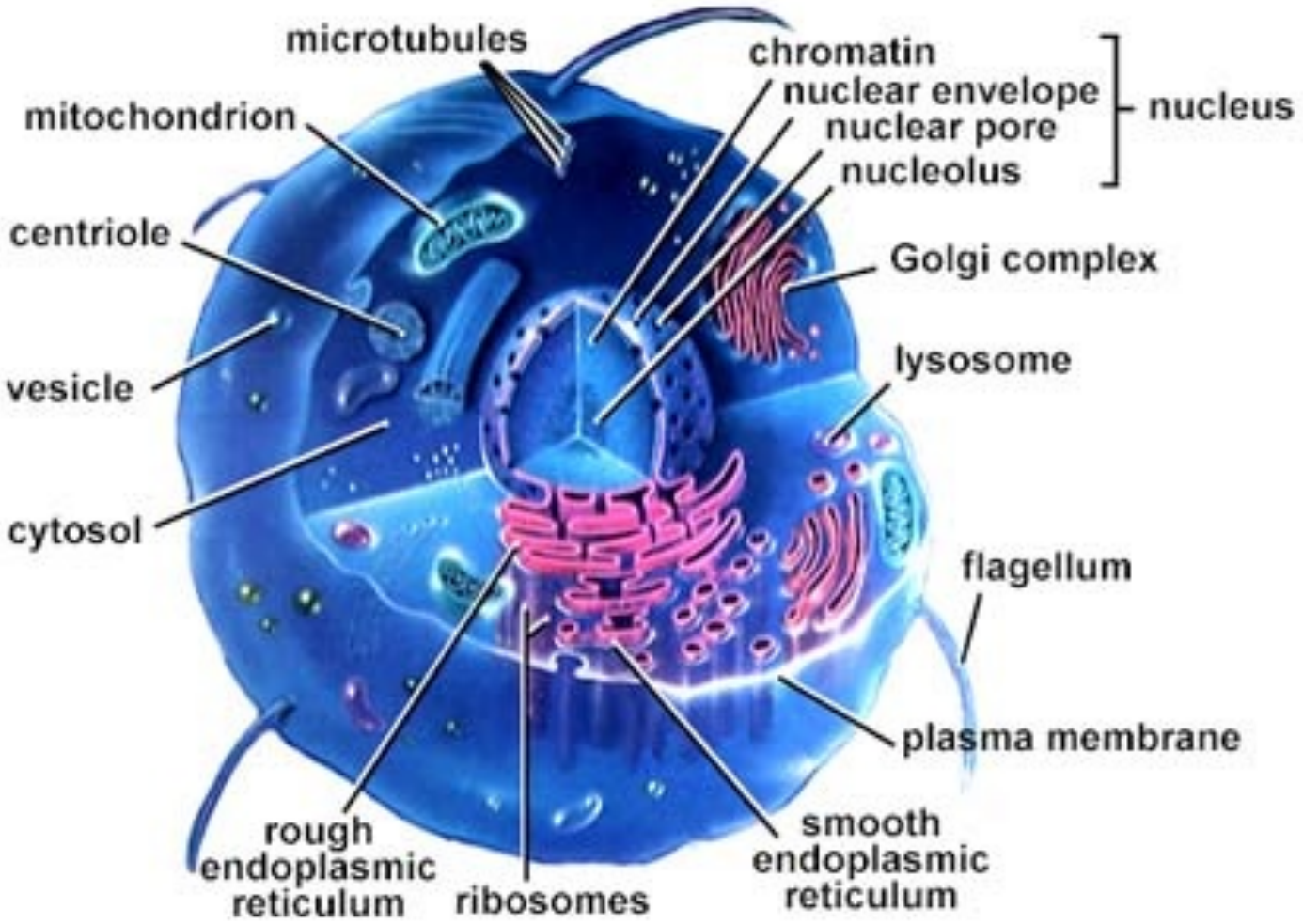
These come together to form organelles



Synaptic vesicle

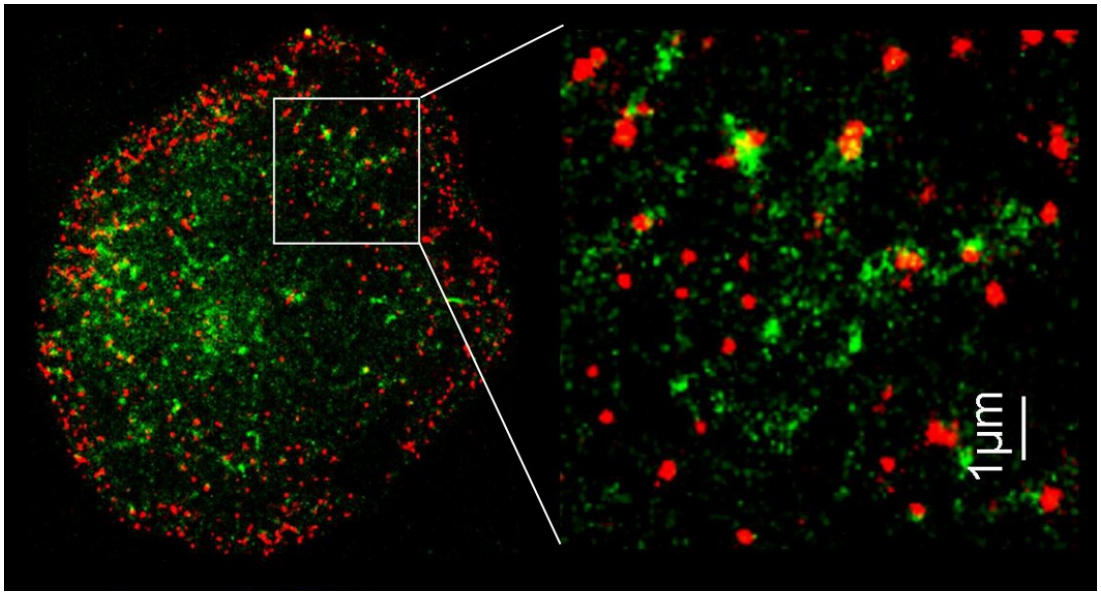
http://www.mpibpc.mpg.de/9547480/vesicle600.jpg

and cells

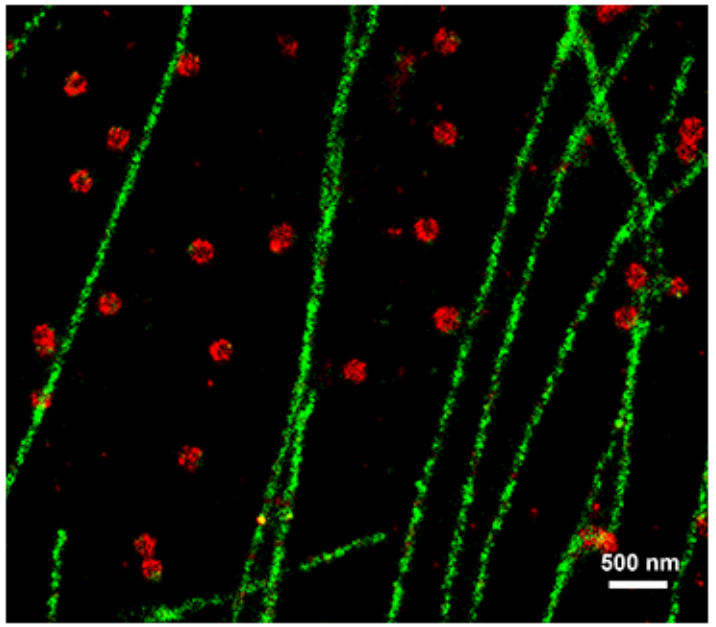


http://www.medfriendly.com/cell.html

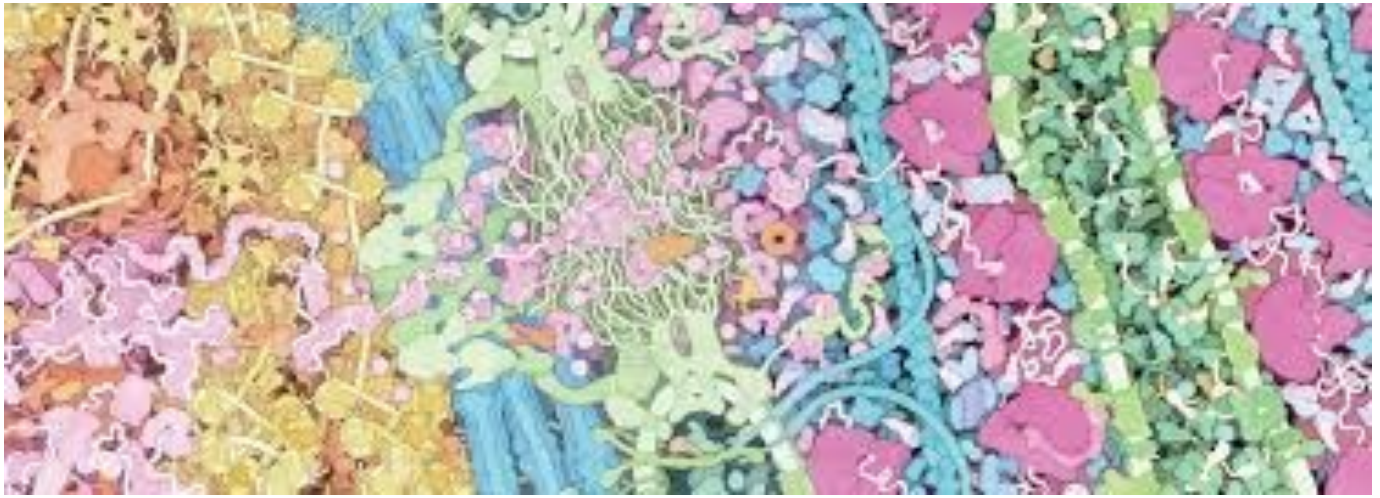
Intracellular structure



Chih-Jung Hsu, Janis Burkhardt and Tobias Baumgart



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



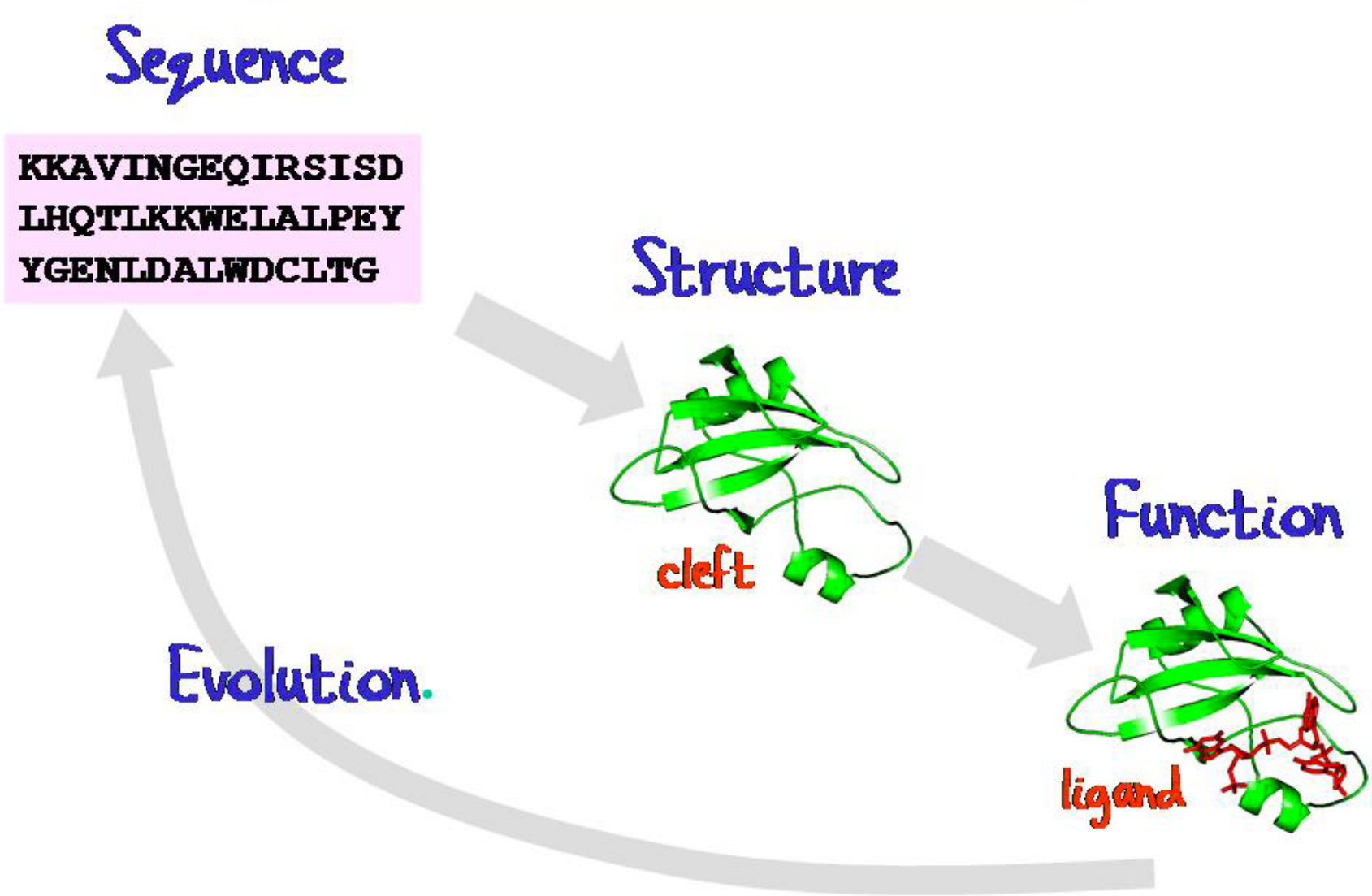
David Goodsell

Intracellular dynamics (artist's rendition)

Janet Iwasa and Tomas Kirchhausen

Why is structure important?

The cycle of life



From Michael Levitt

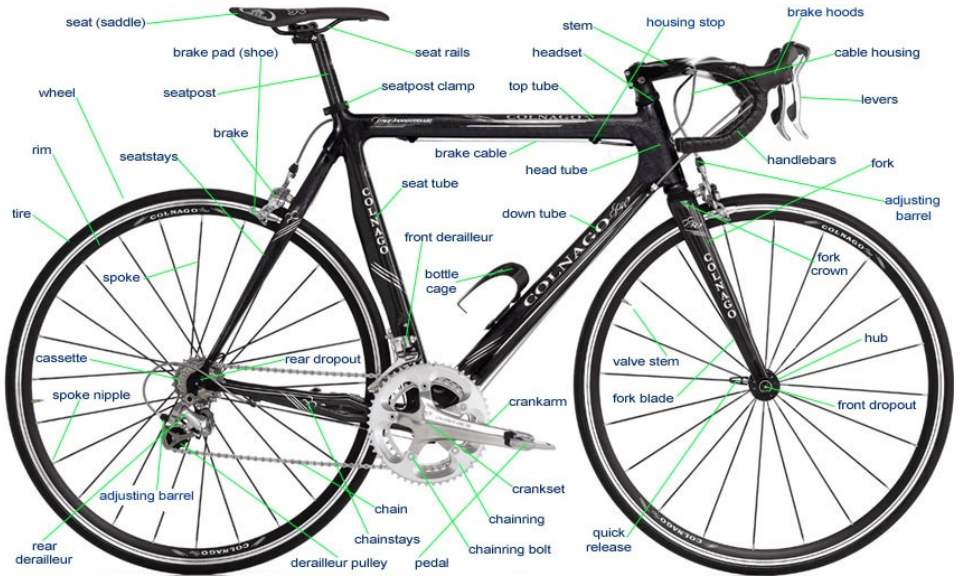
Genomics is a great start ….

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

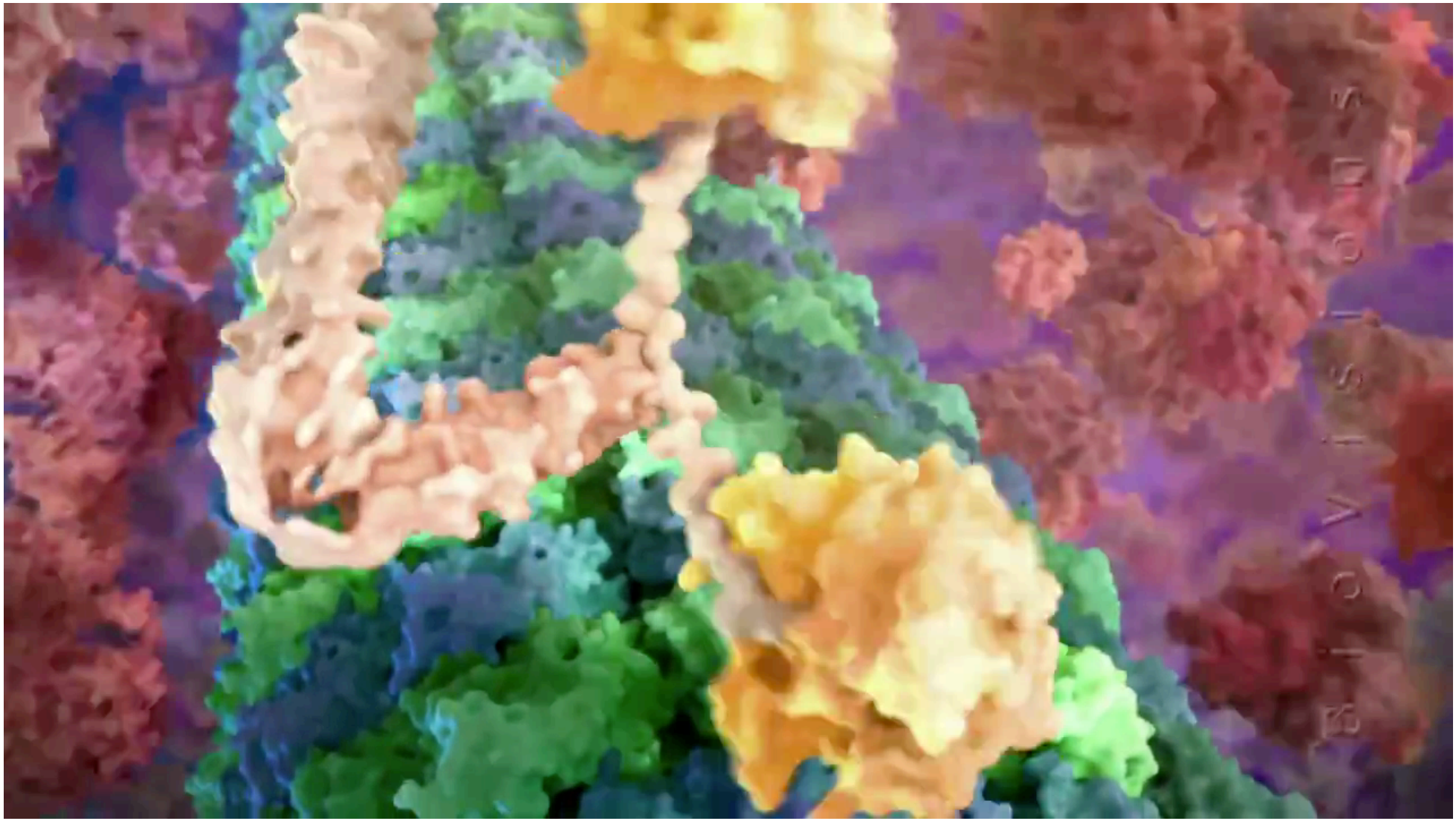
… but not the end



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

Structure determines function

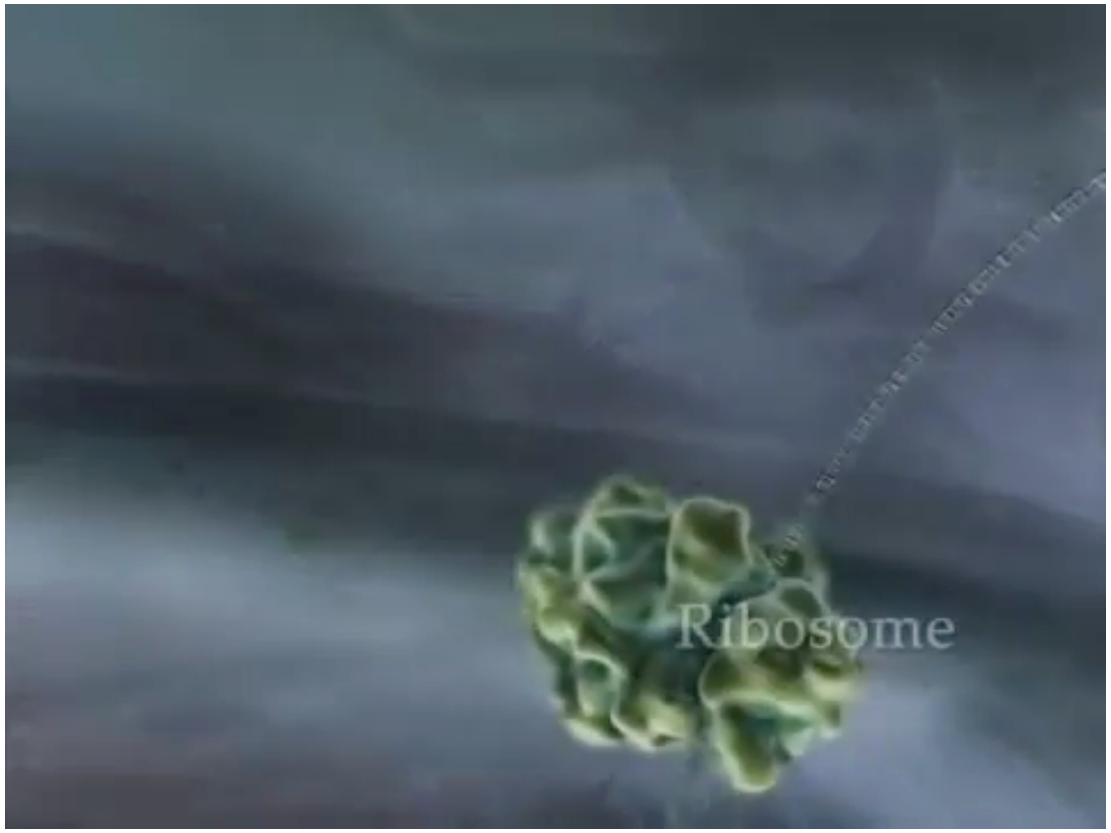
- Example: Motor protein (walks along microtubules, dragging load)



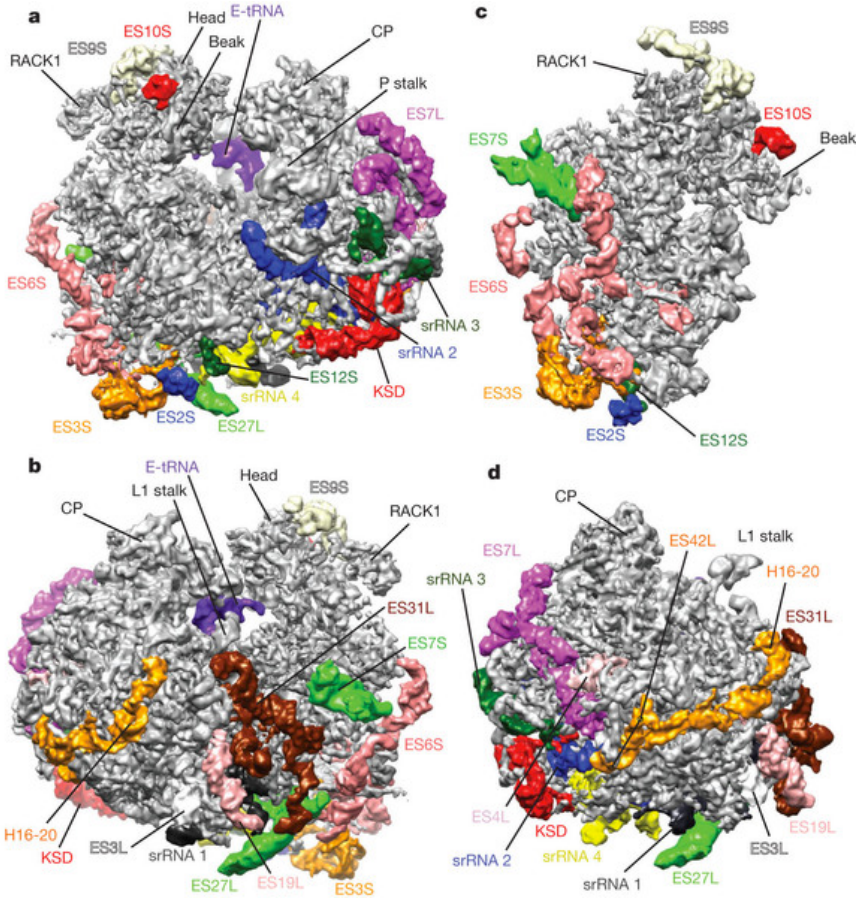
From *Inner Life of the Cell* | *Protein Packing*

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)



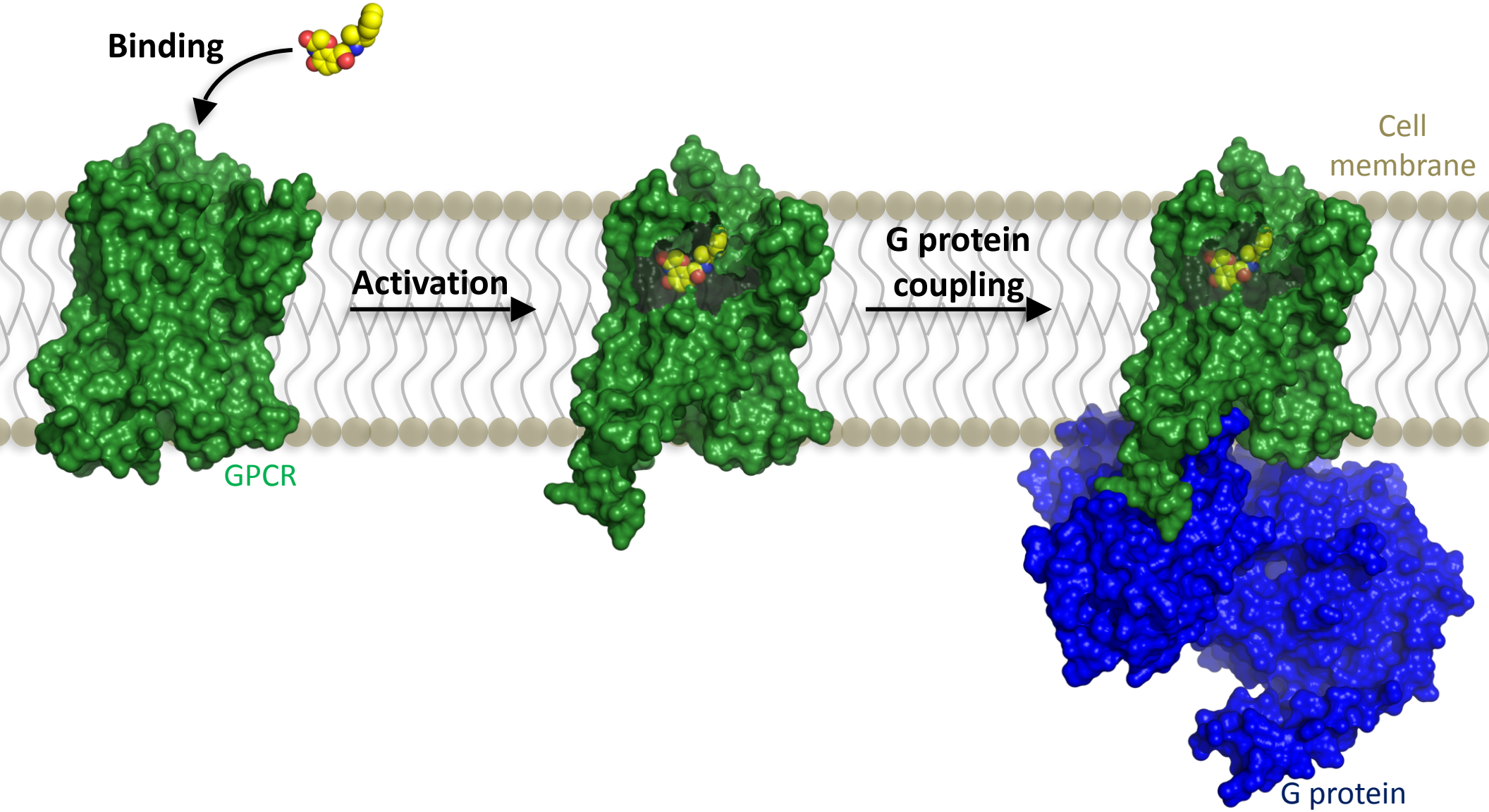
From *Inner Life of the Cell*, XVIVO and Biovisions @ Harvard



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

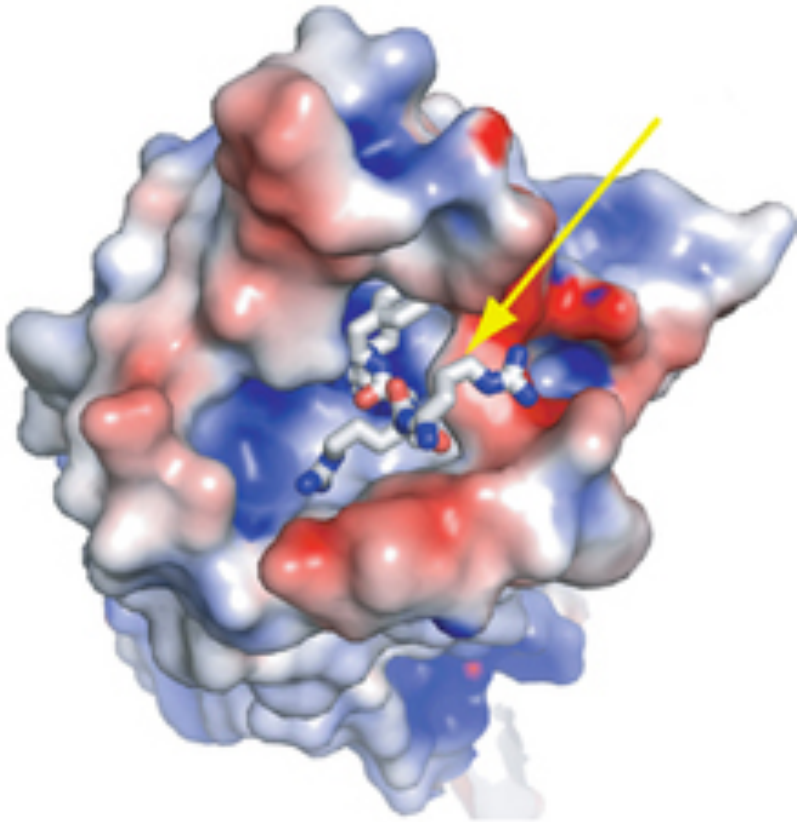
Structure determines function

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



Structure-based drug design

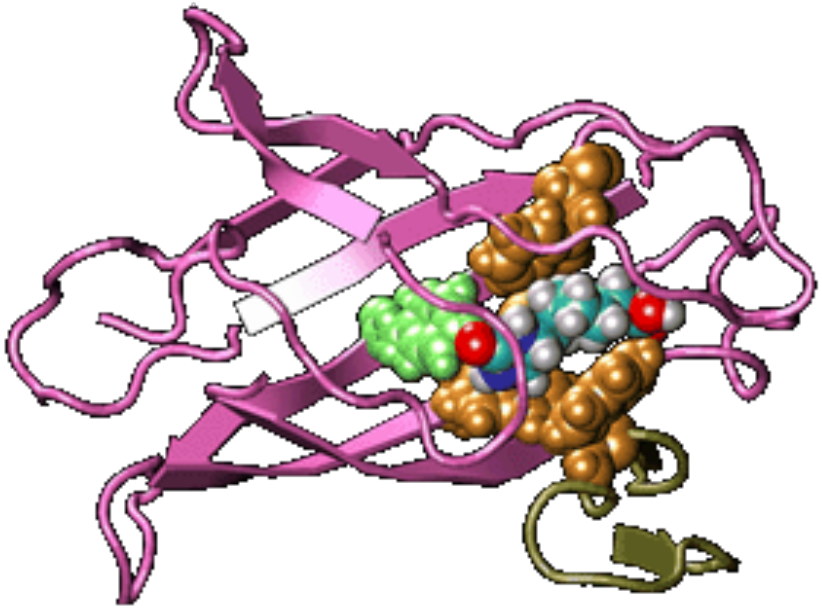
- Almost all drugs act by binding to proteins and altering their function
- Using knowledge of structures, we can design drugs that bind more tightly or more selectively, bind in different positions, alter behavior of protein in different ways, etc.



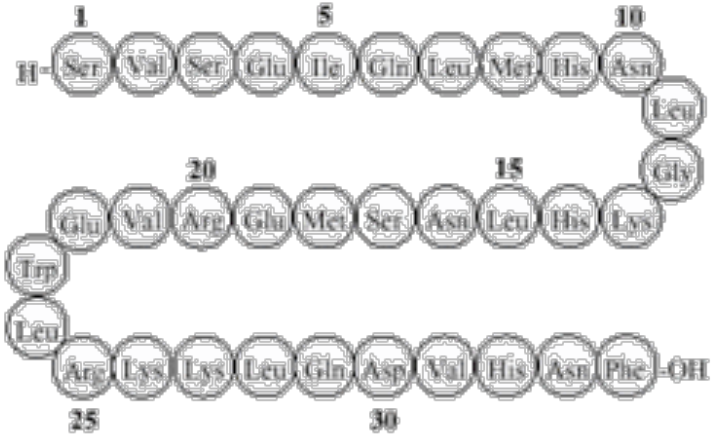
http://www.nih.gov/researchmatters/october2012/images/structure_l.jpg

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

2013 Nobel Prize recognized early developments underlying modern biomolecular computation

The Nobel Prize in Chemistry 2013

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"*.

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)

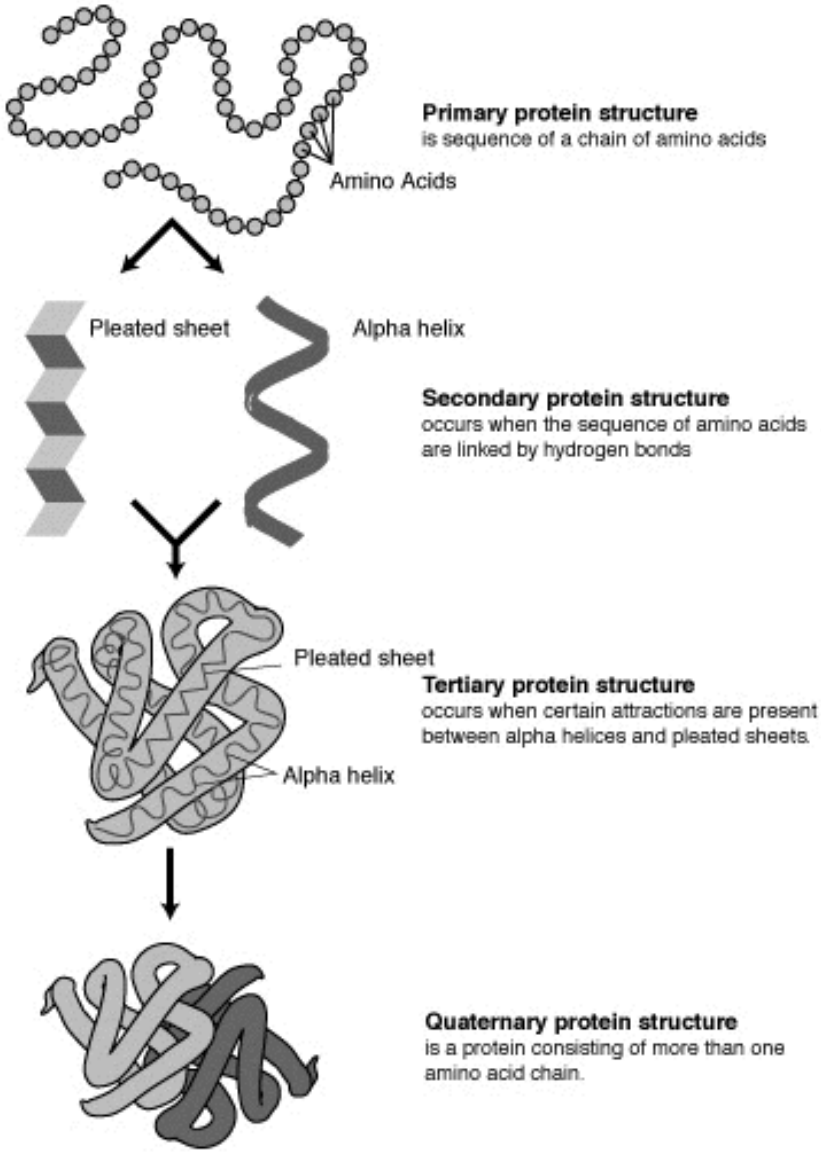
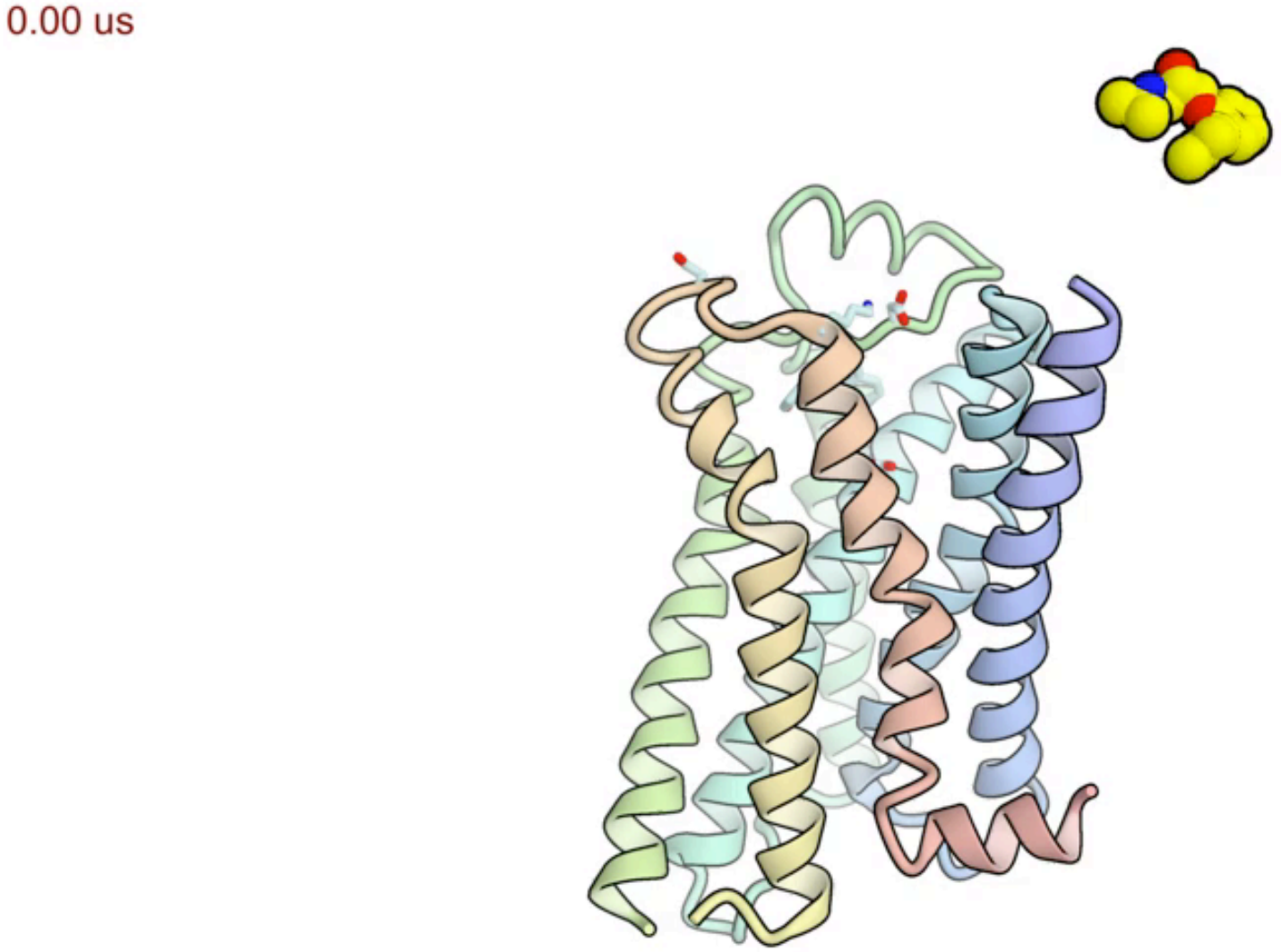


Image from Wikipedia

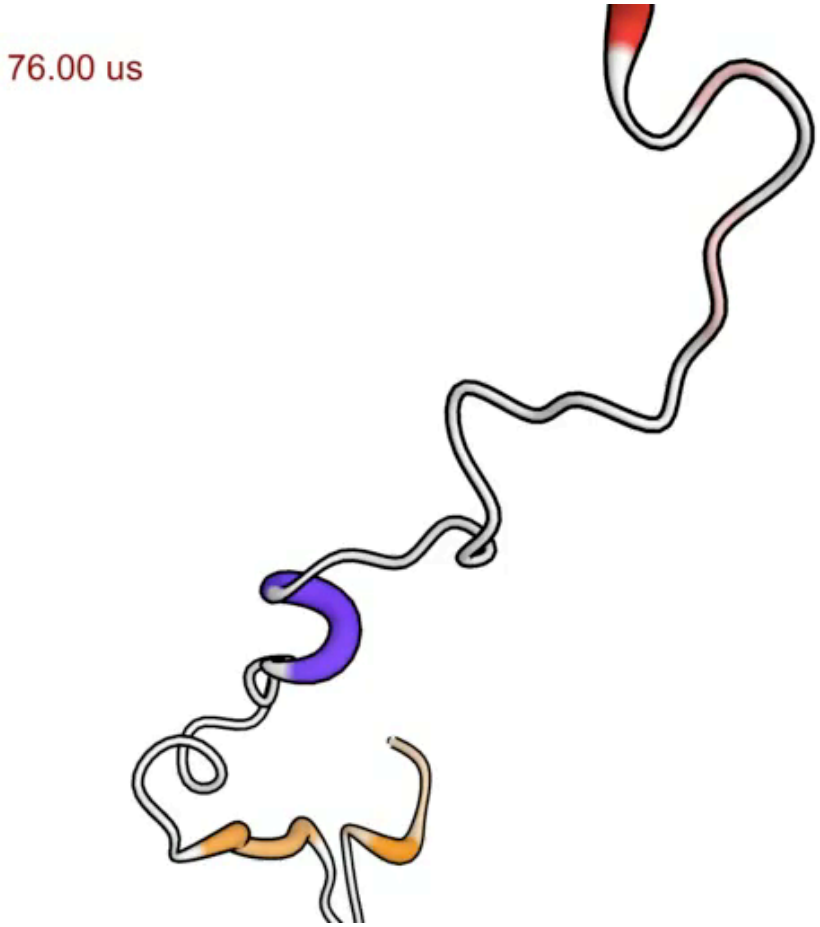
Molecular dynamics simulations



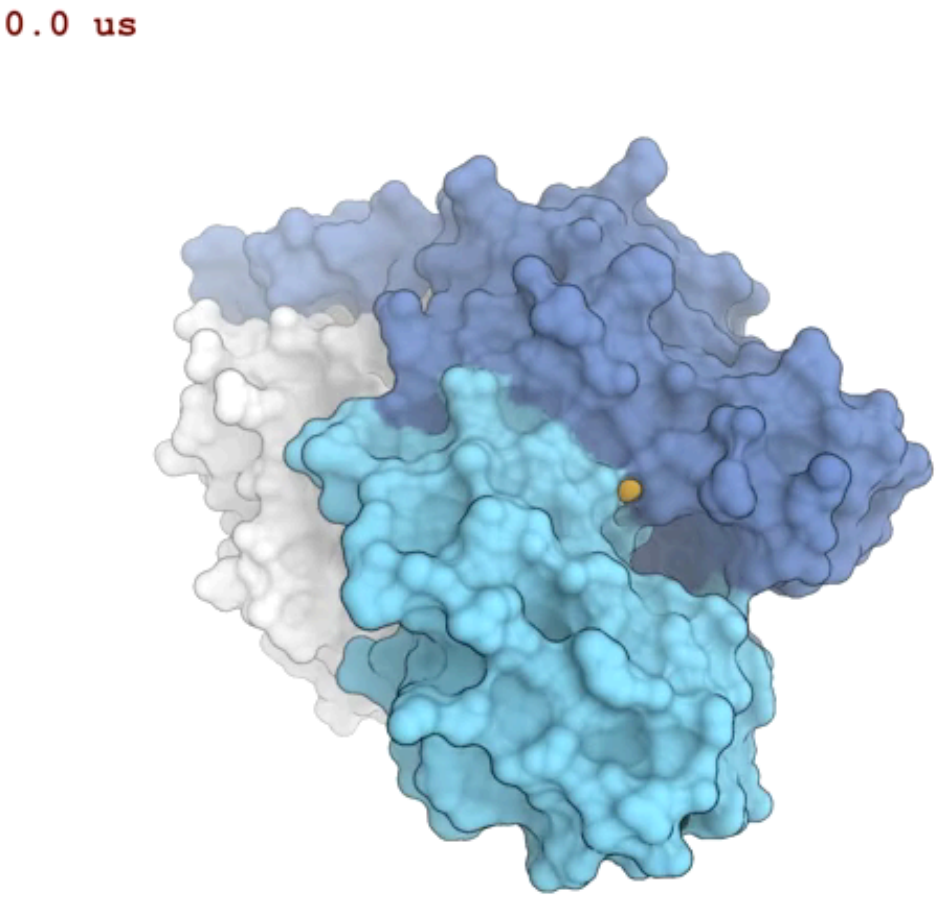
Beta-blocker binding to the β_2-adrenergic receptor

Dror et al., *PNAS* 2011

Molecular dynamics simulations



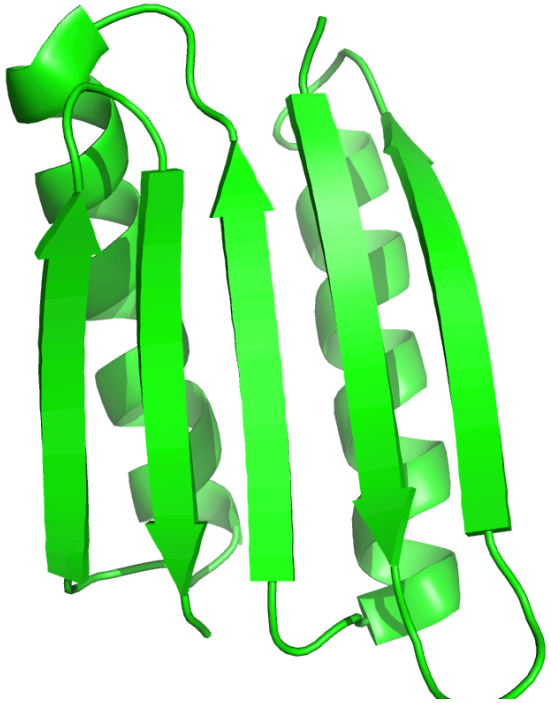
Folding of protein G
(Lindorff-Larsen et al., *Science*, 2011)



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

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

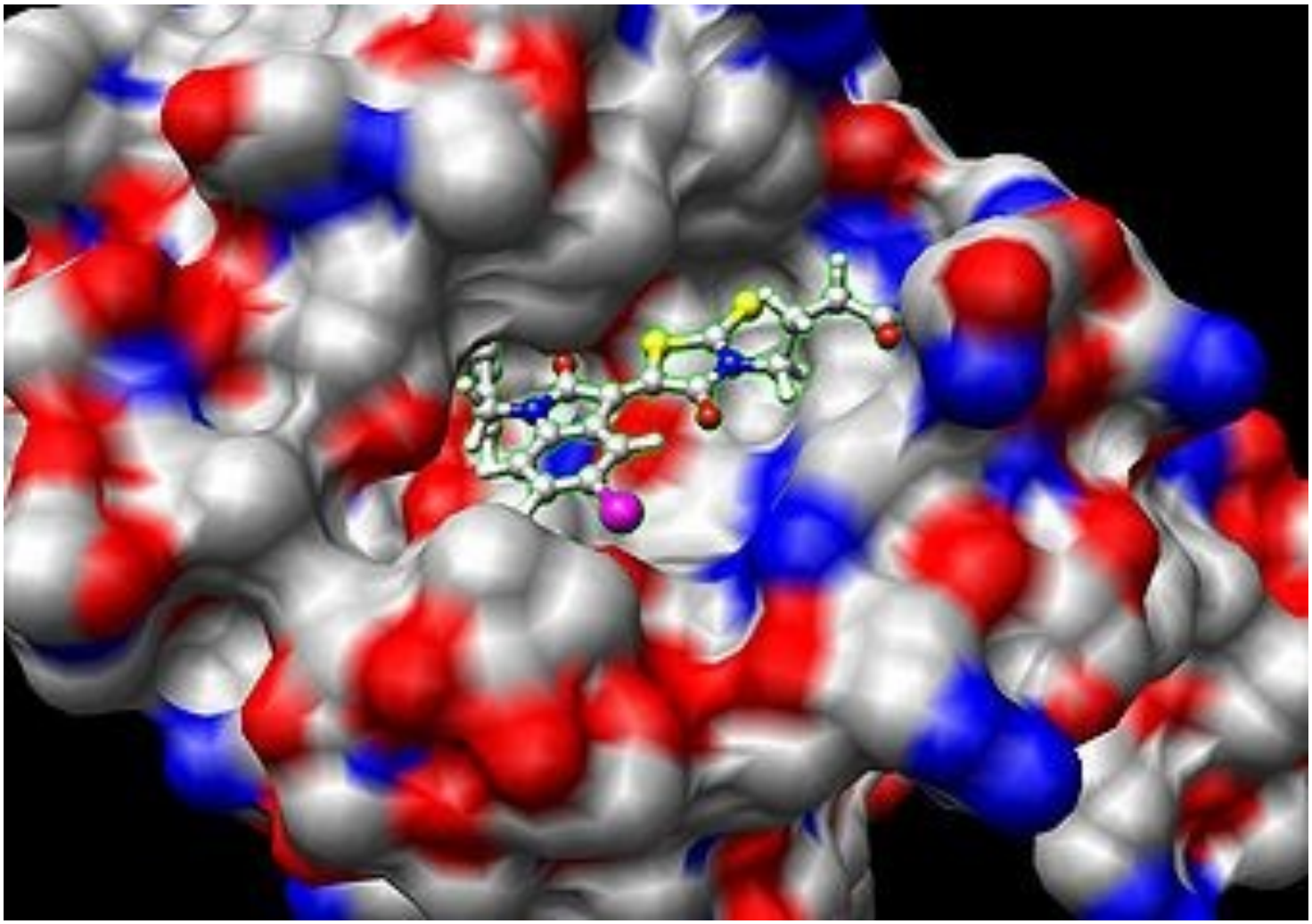
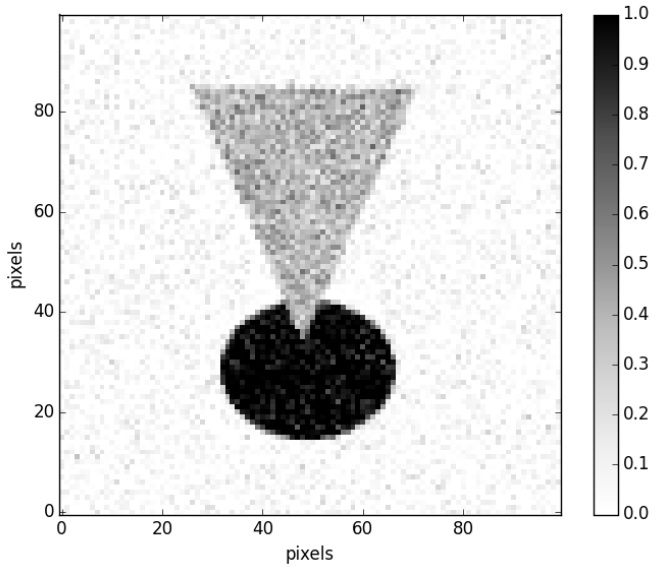


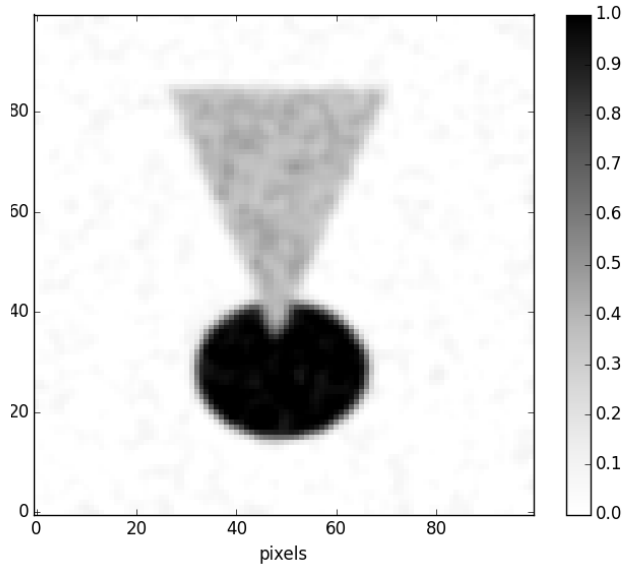
Image: Wikipedia

Image analysis

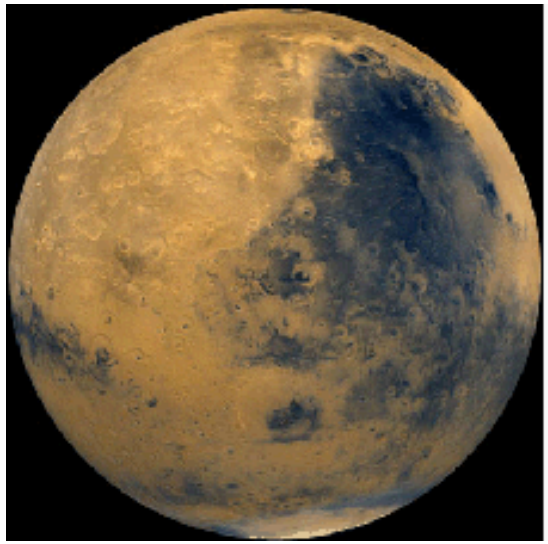
Original image



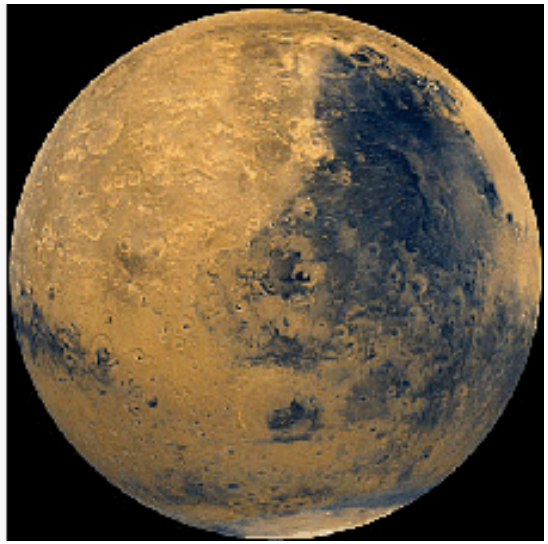
Denoised image



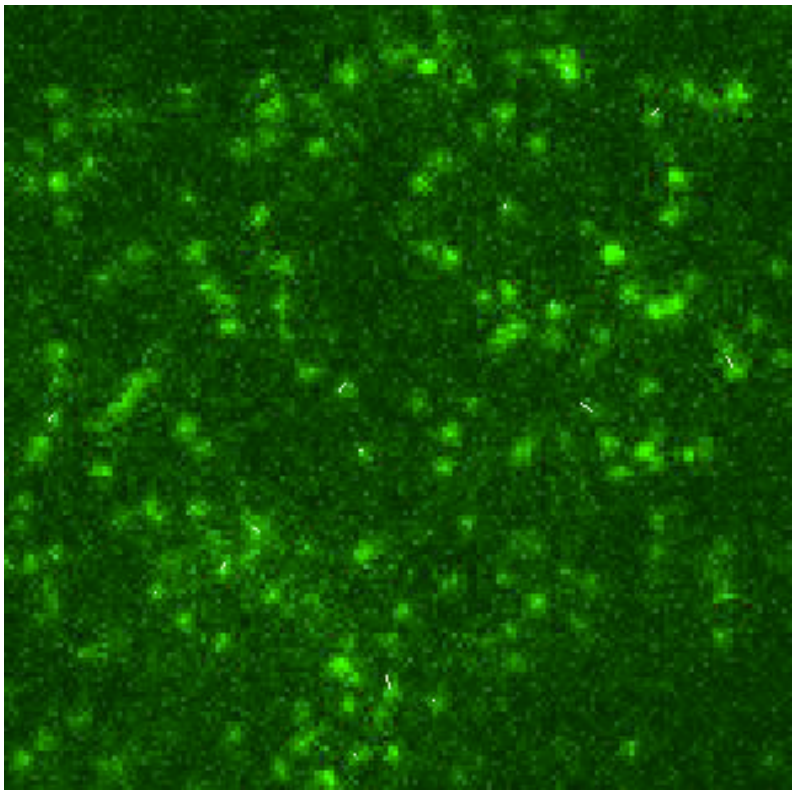
Original image



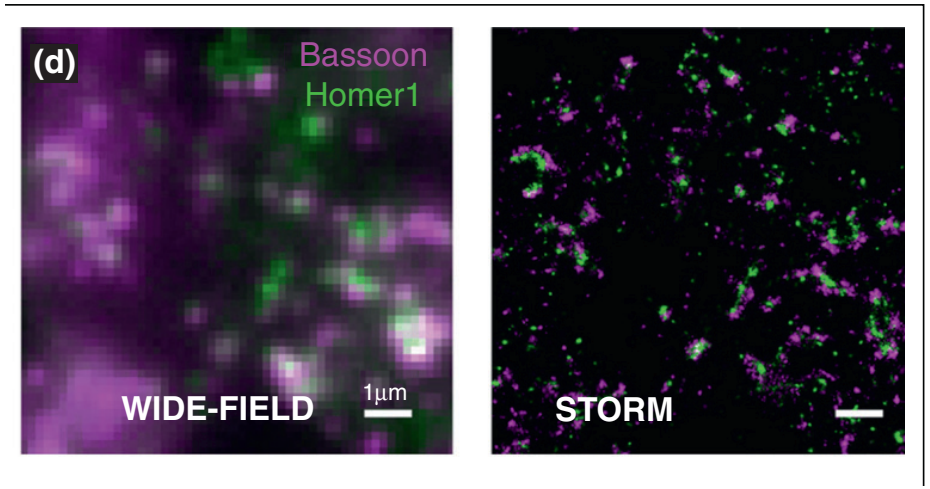
Sharpened image



Fluorescence microscopy and cellular-level organization



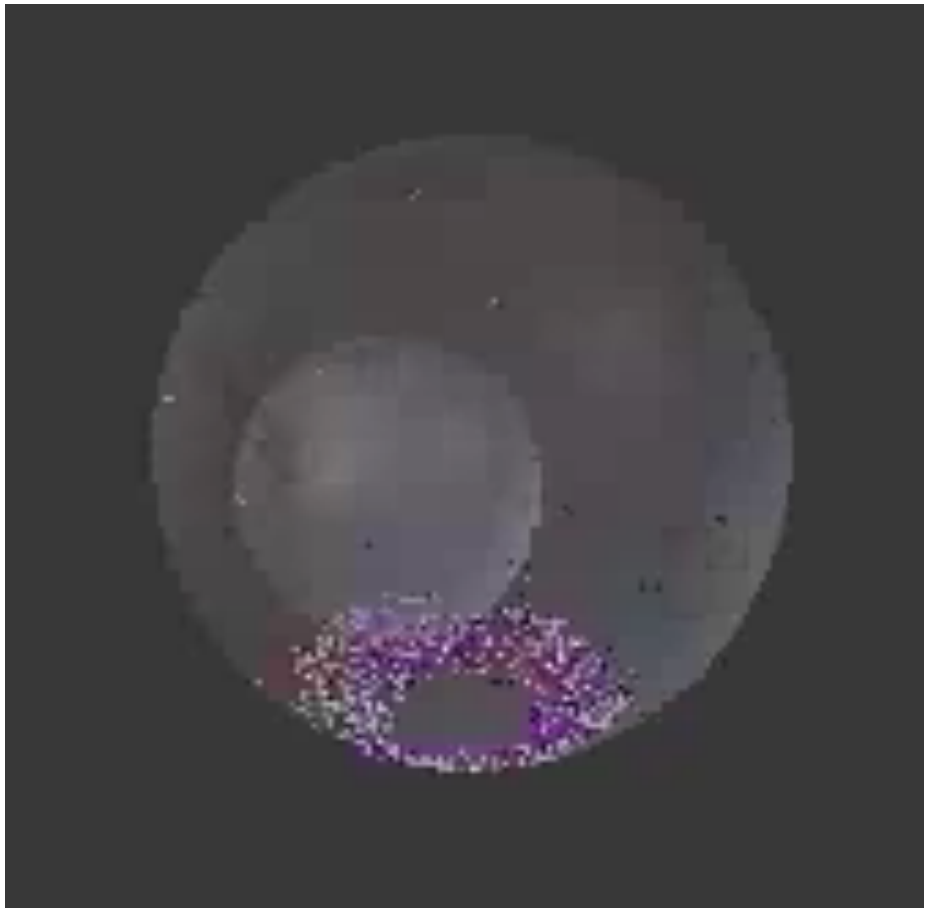
Data: Bettina van Lengerich, Natalia Jura
Tracking and movie: Robin Jia



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

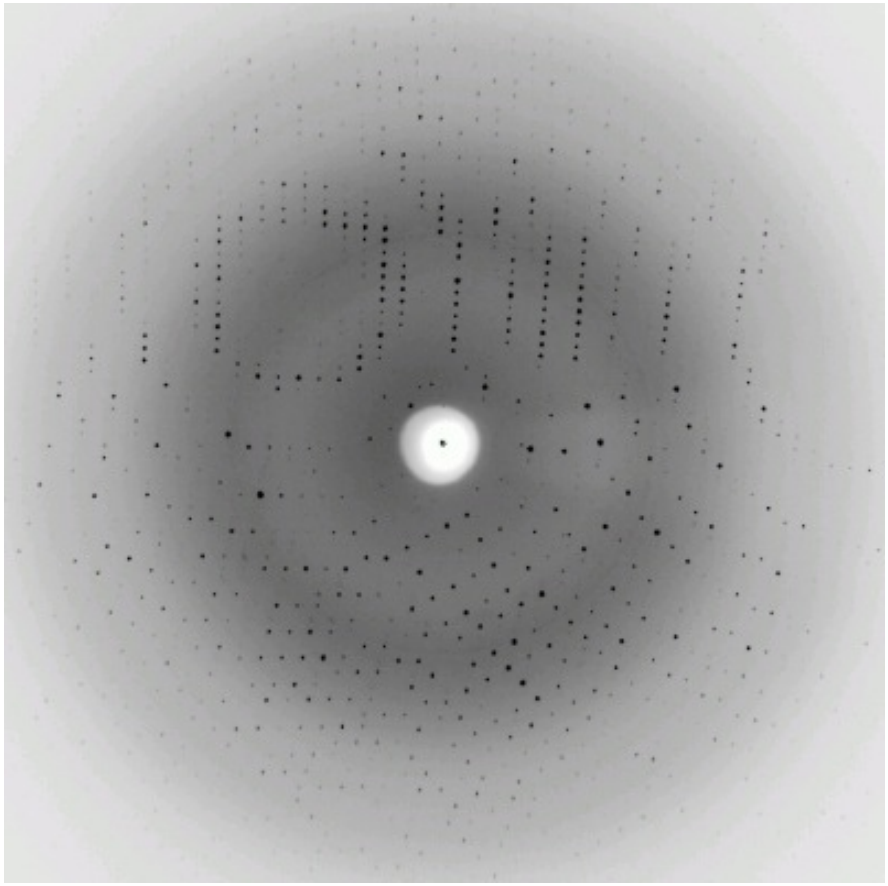
- Including super-resolution microscopy

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



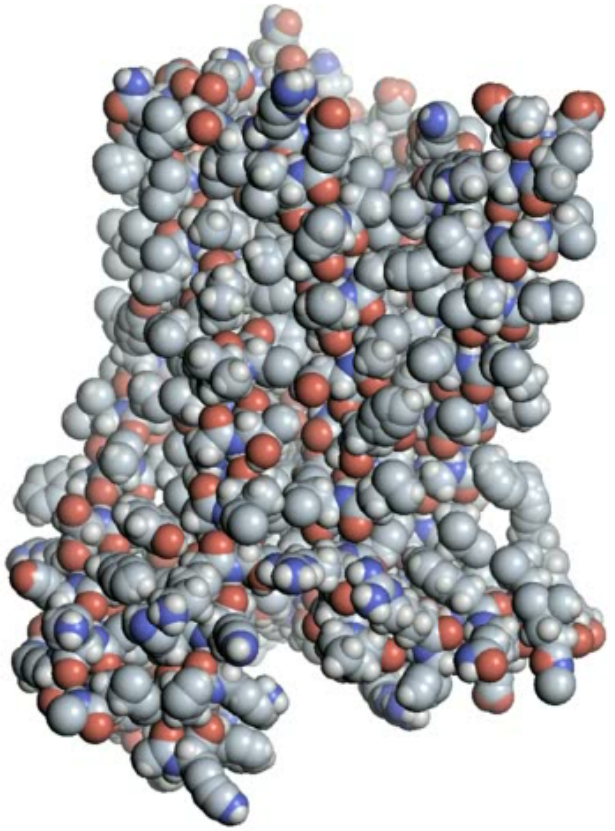
Video: Naomi Latorraca

Solving structures by x-ray crystallography



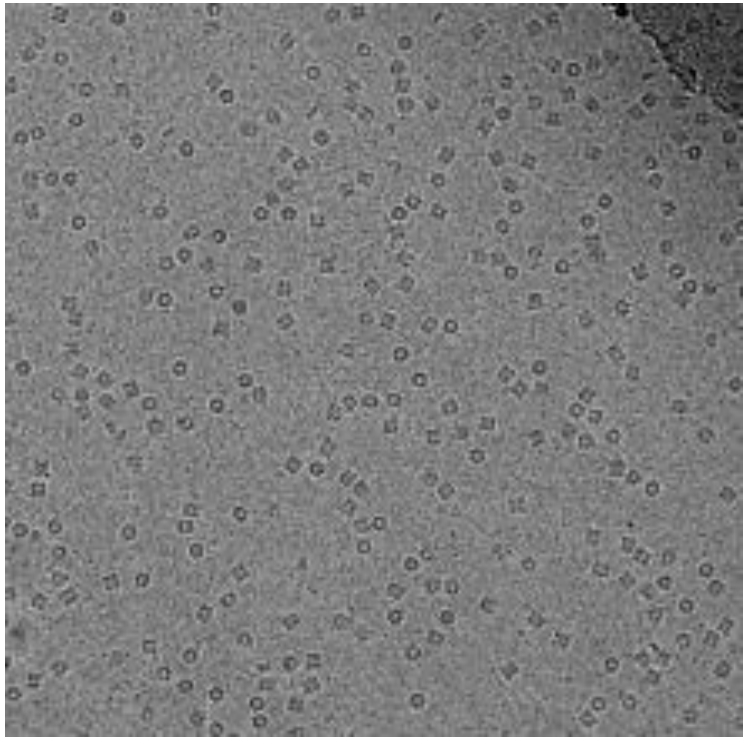
X-ray diffraction pattern

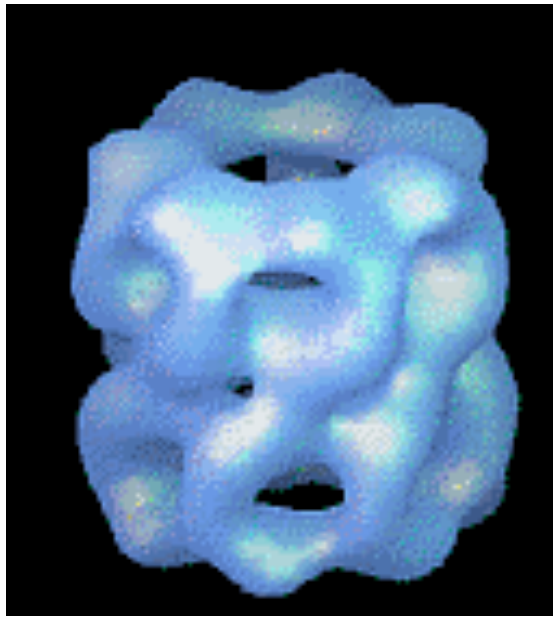
Image: http://www.chem.ucla.edu/harding/IGOC/X/x_ray_crystallography.html

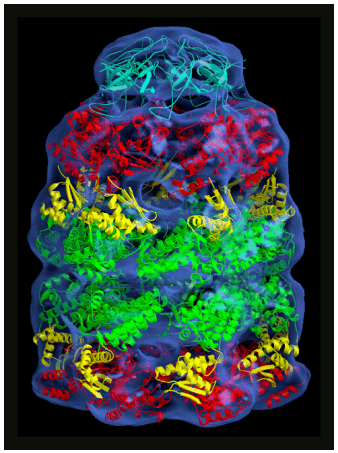


Protein structure

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







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)

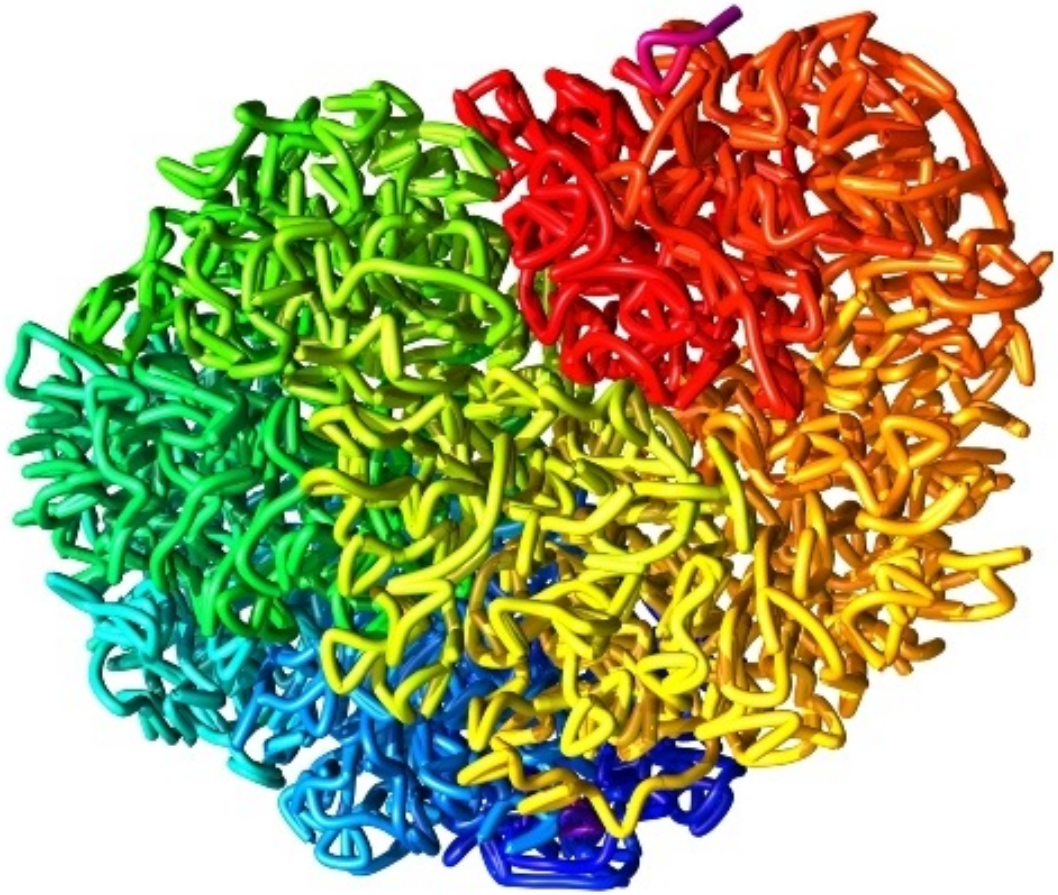


Image:http://www.biotechniques.com/news/Bringing-genome-structure-into-focus/biotechniques-312407.html

Emerging applications of machine learning and AI to structural biology

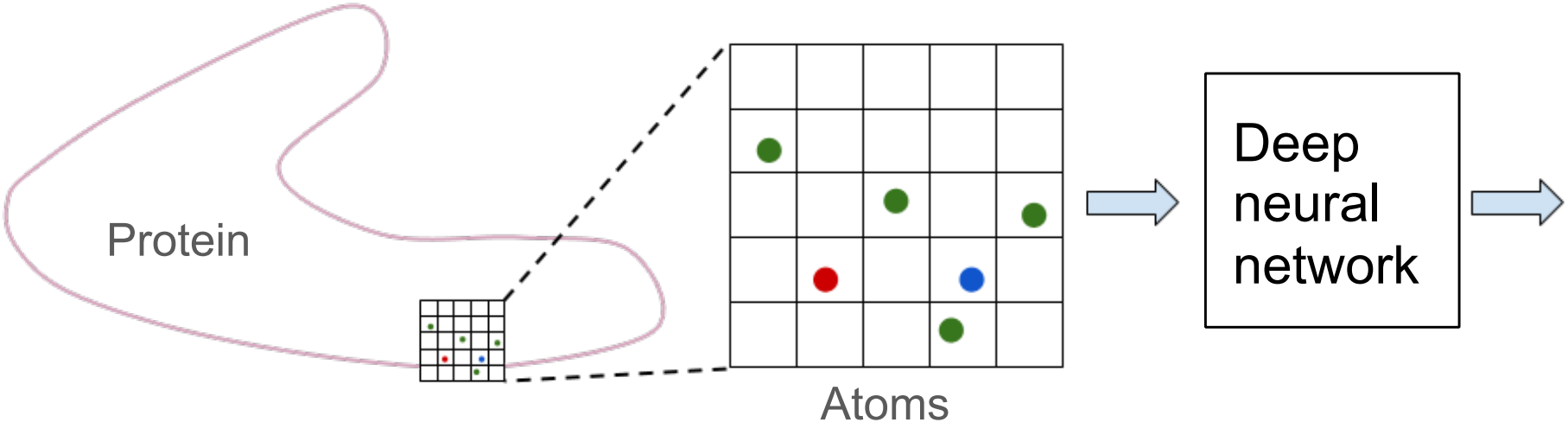


Figure adapted from Raphael Townshend

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** (inference/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)
  - Protein structure
  - 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

Focus will be on fundamentals, but most lectures will touch on current research topics

Course logistics

Course web page

- http://cs279.stanford.edu/
- Evaluation criteria and handouts on web page
- Link to last year's website, which includes list of lecture topics along with all lecture slides
  - This year's content will be similar
- **Please sign up on Piazza (via link on webpage) so that you get announcements**

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.
  - Basic tutorial Wed., Oct. 4, at 5 p.m. (location TBA); additional tutorial Oct. 17

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

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

Assignments, Project, Exam

- 3 assignments
  - First one is mini-assignment
- Project: More open-ended. About the same amount of work as second or third assignments.
- Final exam covering key concepts

Lectures and reading

- Lectures are not videotaped
- No textbook. Slides available, along with brief notes for some lectures and pointers to optional reading material
- Class participation encouraged!

Course staff

- Prof. Ron Dror
  - http://drorlab.stanford.edu/rondror.html
  - Office hours: Tuesdays 4:20–6:00 (Sriram 104 and then Gates 204), or by appointment
- TAs:
  - Rishi Bedi
  - Joe Paggi
  - Daniel Fernandes
  - Adrian Sanborn
  - Osama El-Gabalawy
  - Office hours and contact info at cs279.stanford.edu

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