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

Ron Dror
Stanford University

Sept 24, 2024

Image credit:
Ansgar Philippsen

Real-time class participation encouraged, but you can join in person or virtually

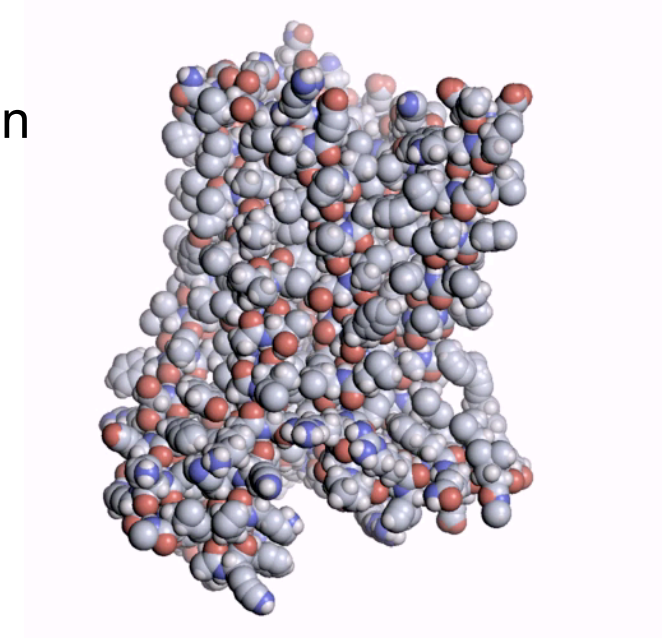
- Lecture live stream available to enrolled students on Canvas
  - Go to Canvas page for course: https://canvas.stanford.edu/courses/195742
  - Select “Panopto Course Videos” tab on the left-hand side

- If you’re feeling unwell or believe you have been exposed to COVID-19/flu, please attend class virtually

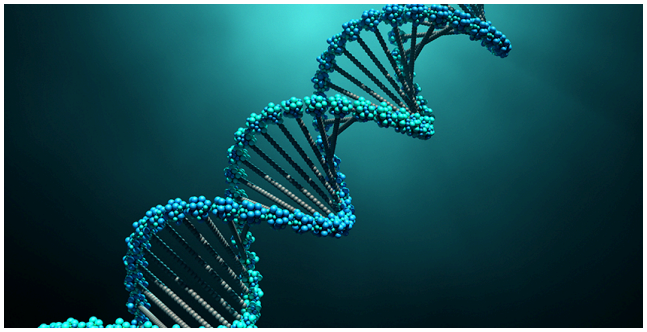
- If you’re not available during class time, please watch the recorded lecture before the next class

One-fifth of science Nobel Prizes relate to 3D structure/organization of biomolecules

Protein



DNA



RNA

- Biological structure is critical to:
  - Understanding how biology works
  - Diagnosing, preventing, and treating disease
  - Food and energy production (e.g., agriculture)

+ combatting climate change

Computation plays a critical and rapidly growing role in this field

Nobel Prize (2013): Computational models of biomolecules

AND THE WINNER OF THE NOBEL PRIZE IN SOFTWARE IS...

created foundation for much of today's work in computational biology

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"*.

nature

NEWS | 30 November 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Dramatic growth of research and commercial activity (startups, acquisitions, etc.) in both physical simulation and machine learning approaches for determining and exploiting biomolecular structure and dynamics

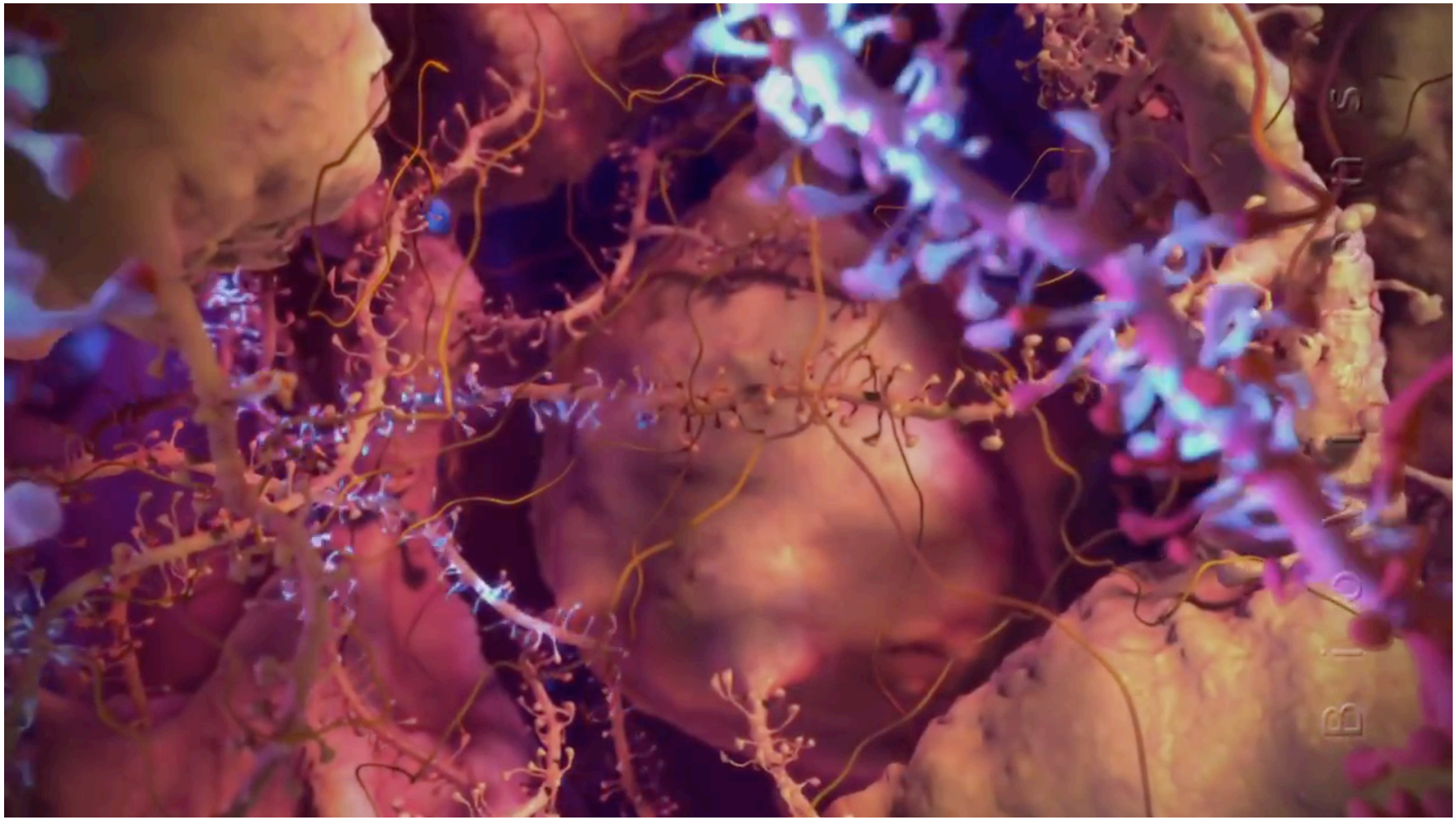
Outline for this lecture

- What is structure? dynamics = how a structure changes over time
  - Structure (and dynamics) at multiple spatial scales
- Why is structure important?
- Overview of topics we'll cover
- 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



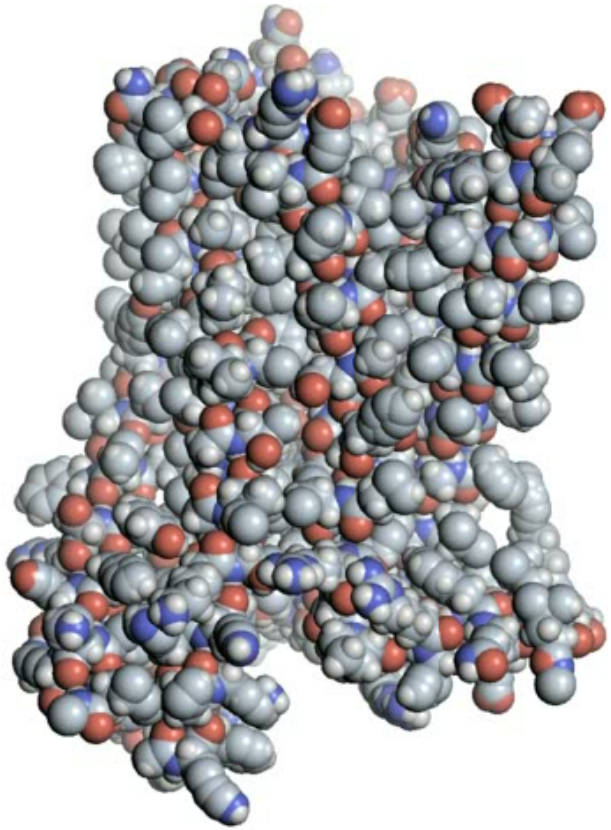
From *Inner Life of the Cell* | *Protein Packing*, XVIVO and Biovisions @ Harvard

What is structure?

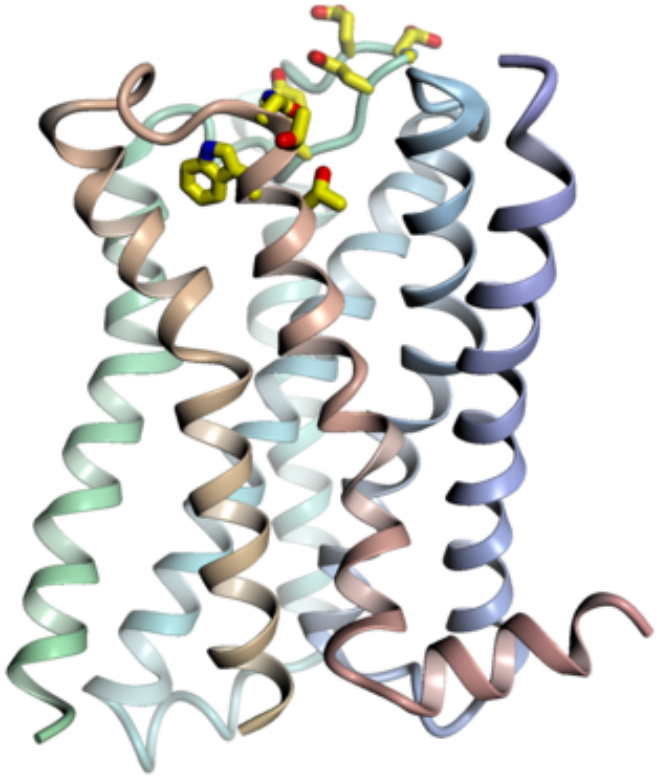
Structure (and dynamics)
at multiple spatial scales

Protein structure

all-atom rendering

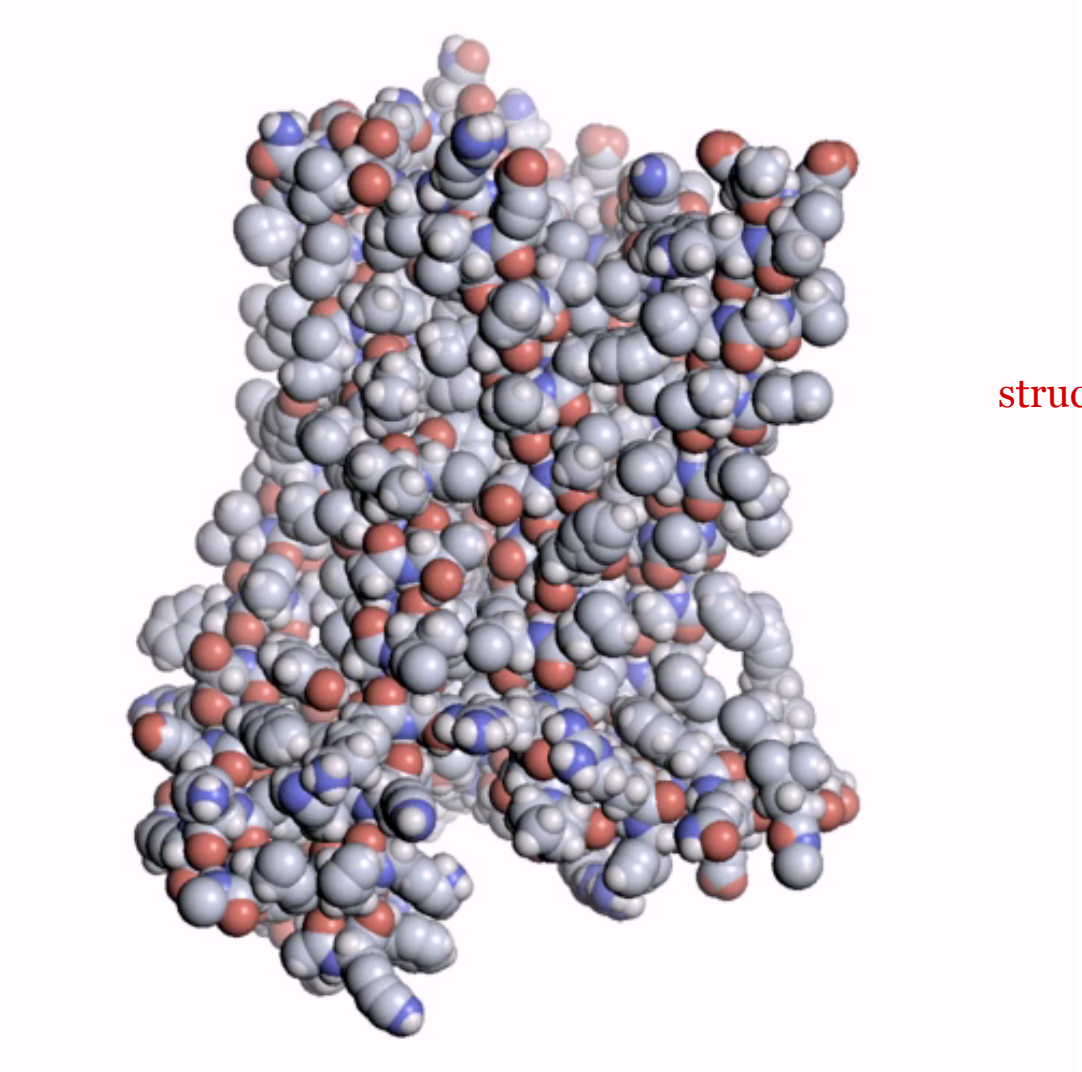


cartoon rendering of the same structure



An adrenaline receptor
(the β_2 adrenergic receptor)

Protein dynamics



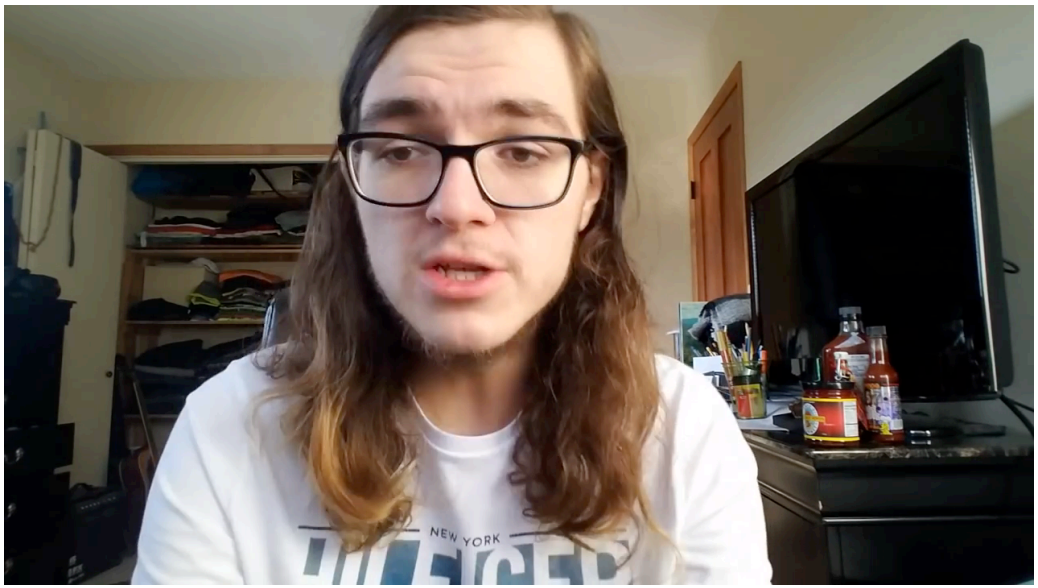
structure is not static!

β_2 adrenergic receptor

Example: how LSD binds to its target

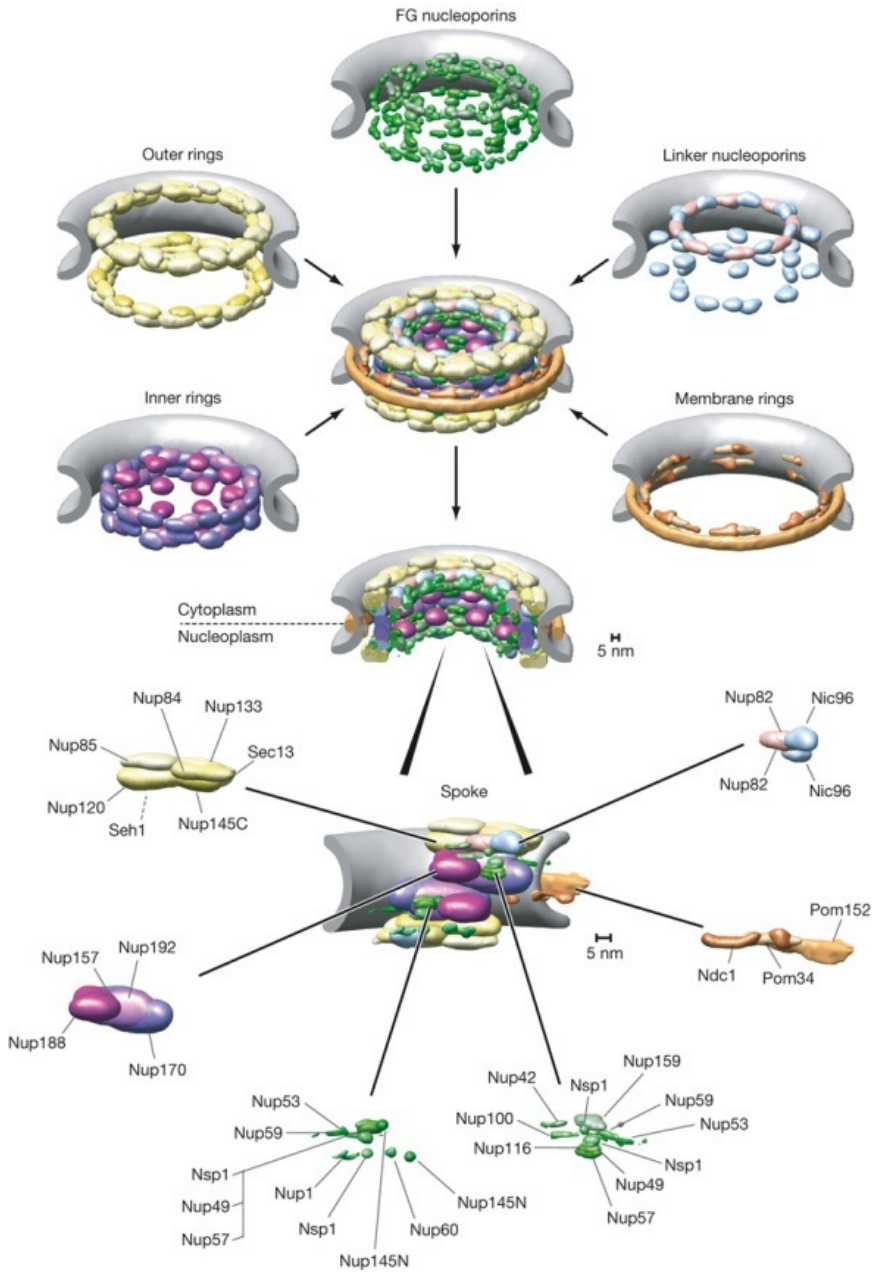
Wacker et al., *Cell* 168:377, 2017
Collaboration with Bryan Roth (UNC)

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



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

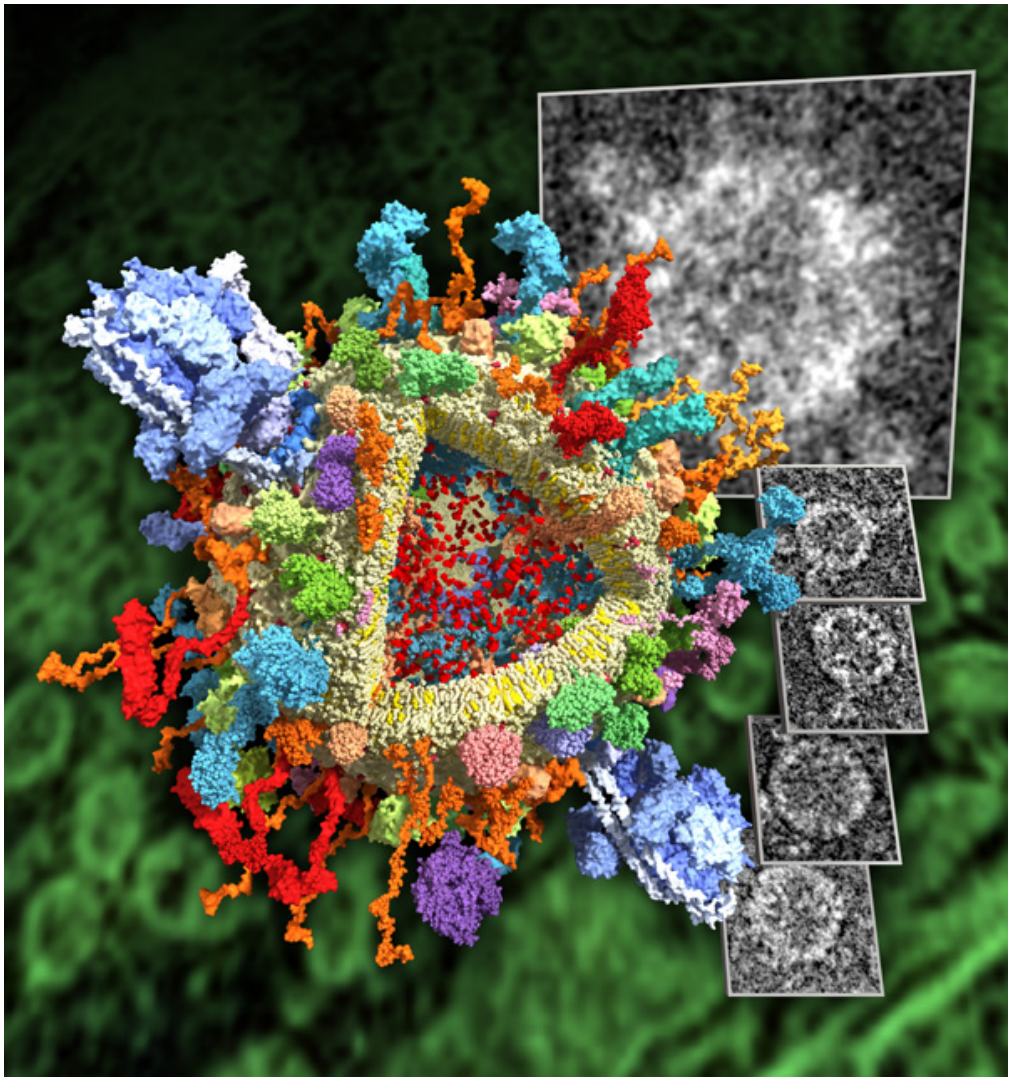
Proteins (and other molecules) often come together to form *macromolecular complexes*



Nuclear Pore Complex

Alber et al., *Nature* 2007

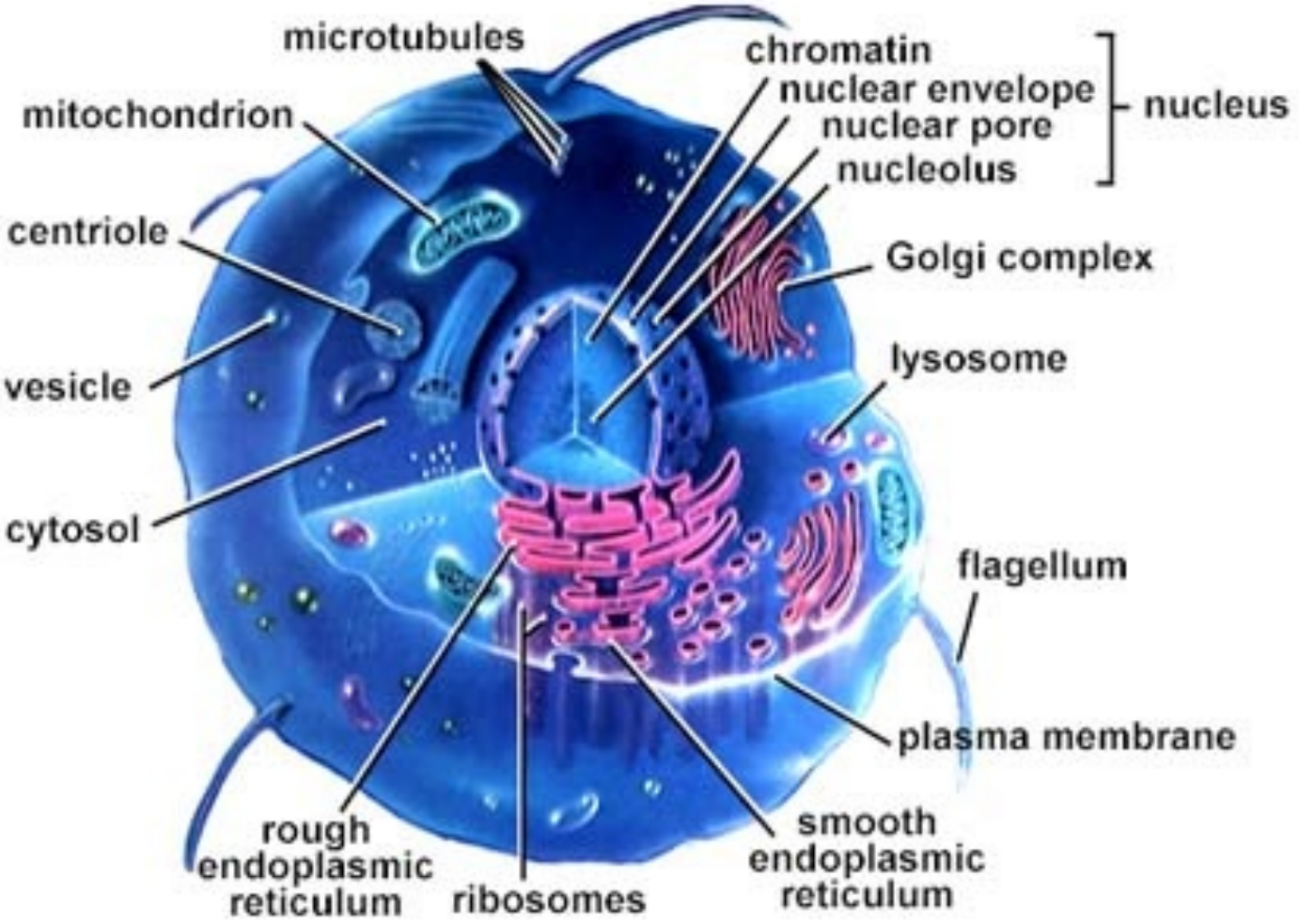
These come together to form organelles



Synaptic vesicle

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

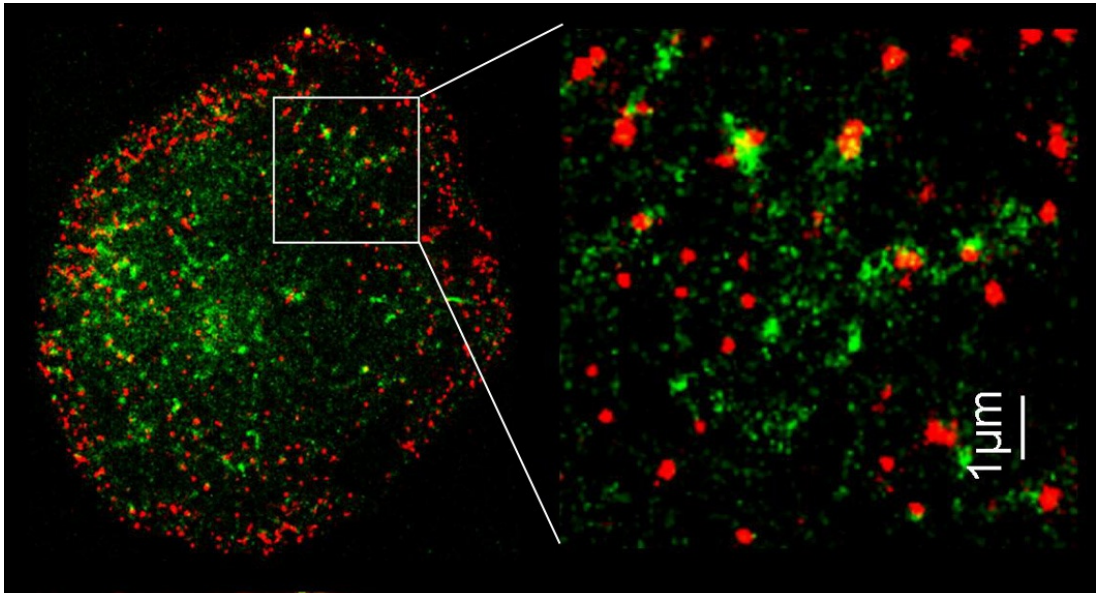
and cells



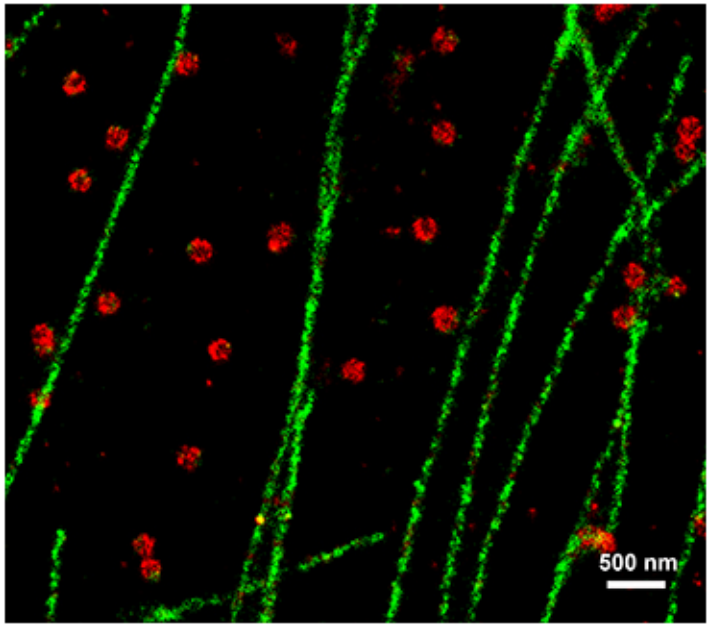
http://www.medfriendly.com/cell.html

Intracellular structure

modern microscopy allows us to observe locations of individual proteins over time



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)

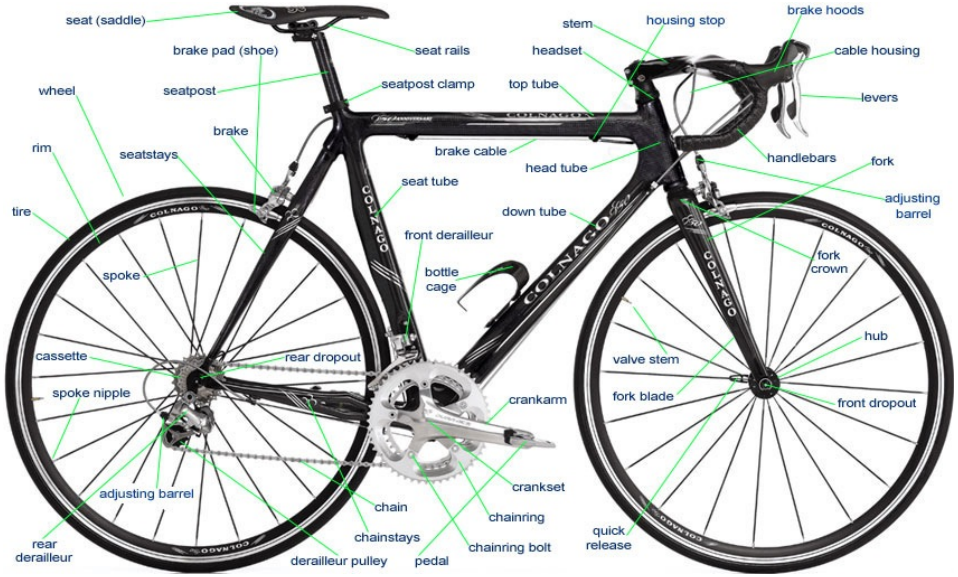
Janet Iwasa and Tomas Kirchhausen

Why is structure important?

To understand how a machine works, we need more than a list of its parts

Track Bike — DL 175

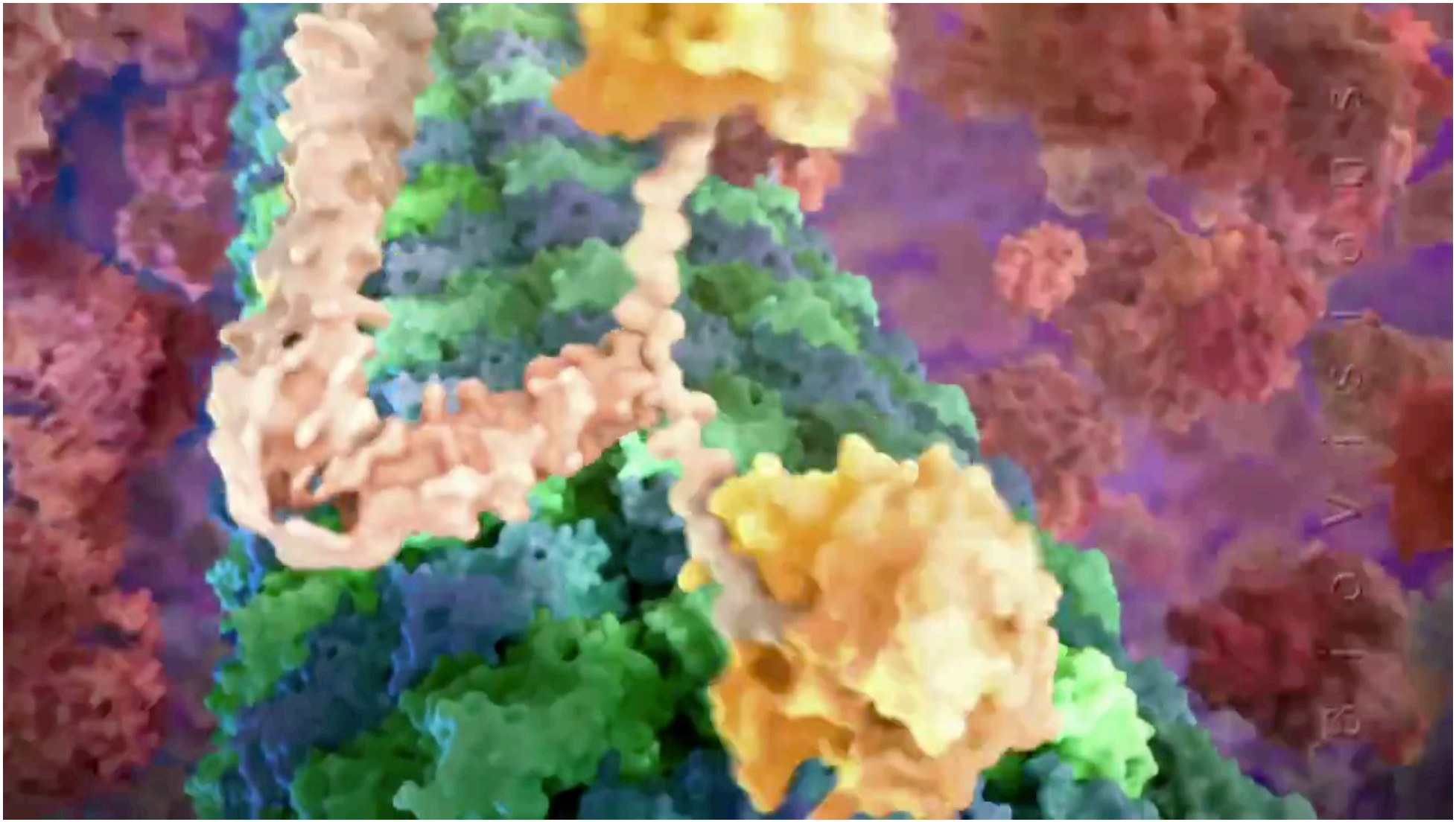
REF. NO.	IBM NO.	DESCRIPTION
1	156011	Track Frame 21", 22", 23", 24", Team Red
2	157040	Fork for 21" Frame
2	157039	Fork for 22" Frame
2	157038	Fork for 23" Frame
2	157037	Fork for 24" Frame
3	191202	Handlebar TTT Competition Track Alloy 15/16"
4		Handlebar Stem, TTT, Specify extension
5	191278	Expander 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



- We want to know the shapes of these parts, how they move, and how they affect each other

Structure determines function

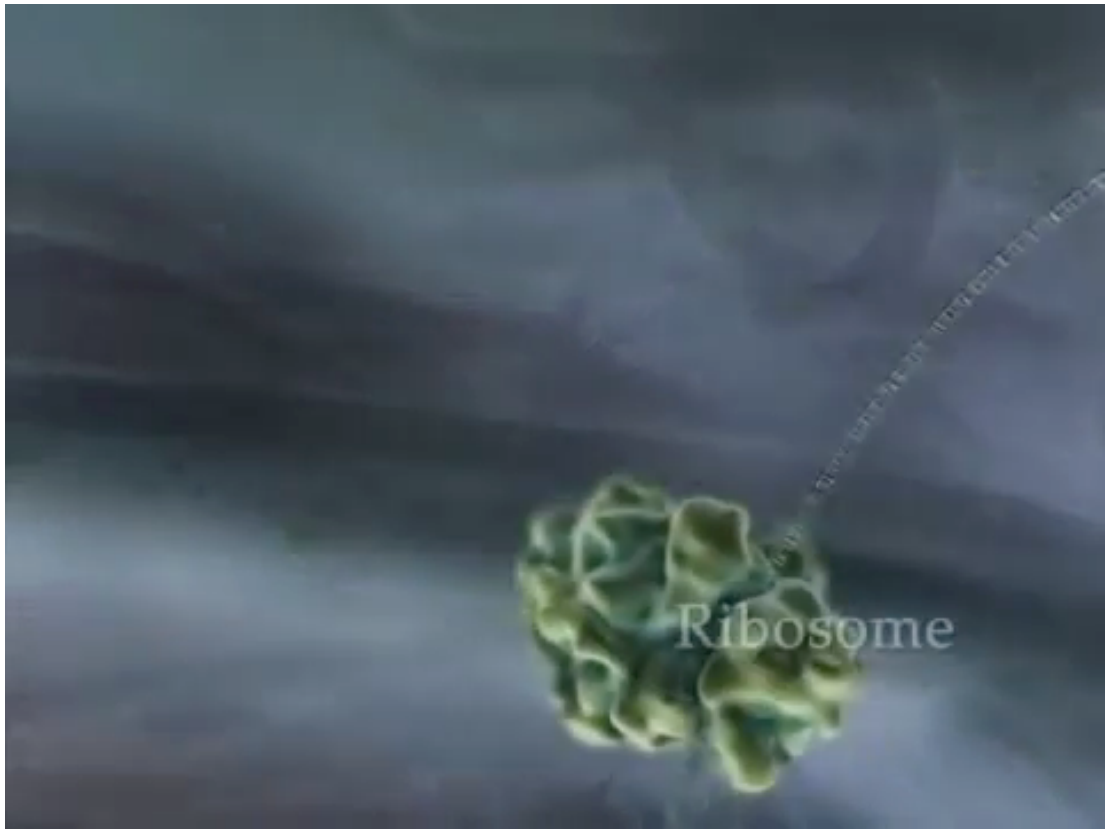
- Example: Motor protein (walks along microtubules, dragging load)



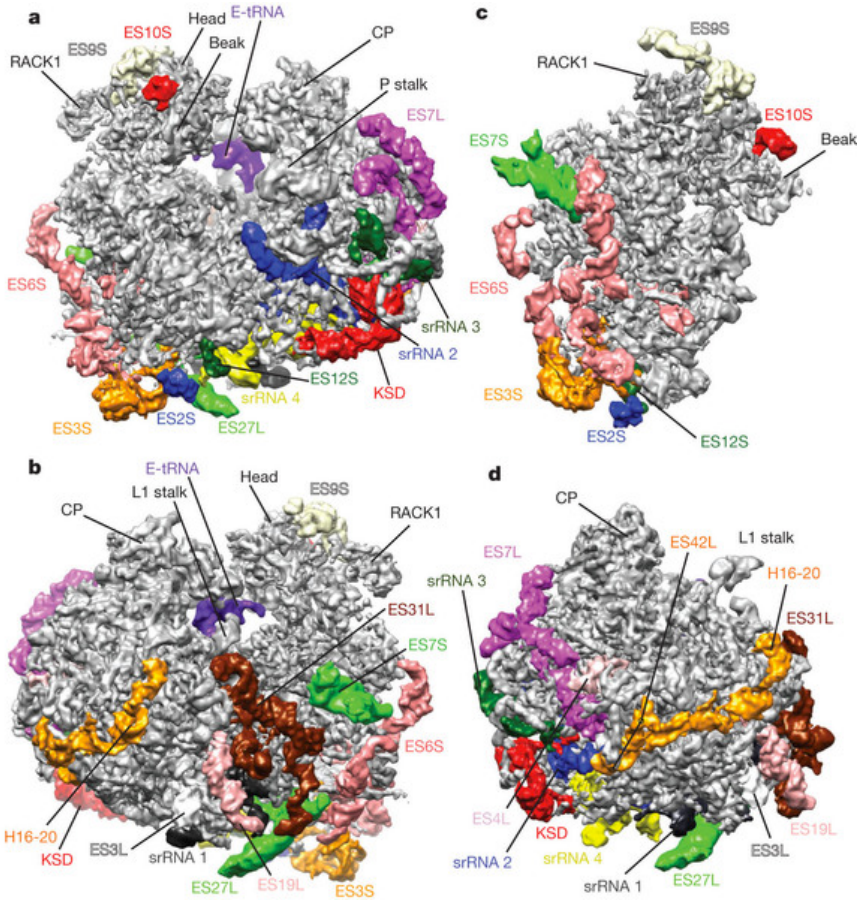
From *Inner Life of the Cell* | *Protein Packing*

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

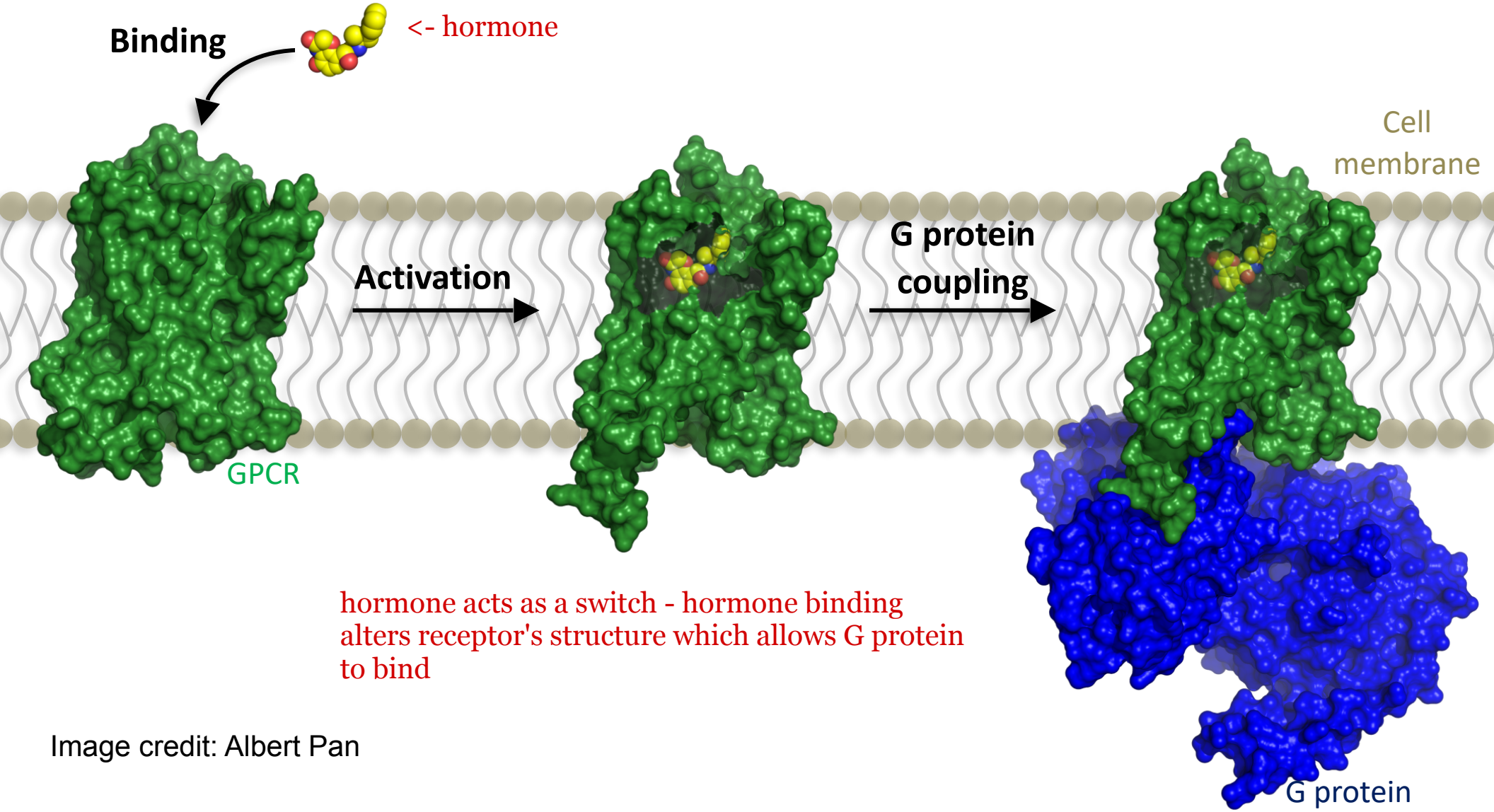


Hashem et al., Nature 494:385-9, 2013

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

> 1/3 of drugs on the market work by binding to these receptors

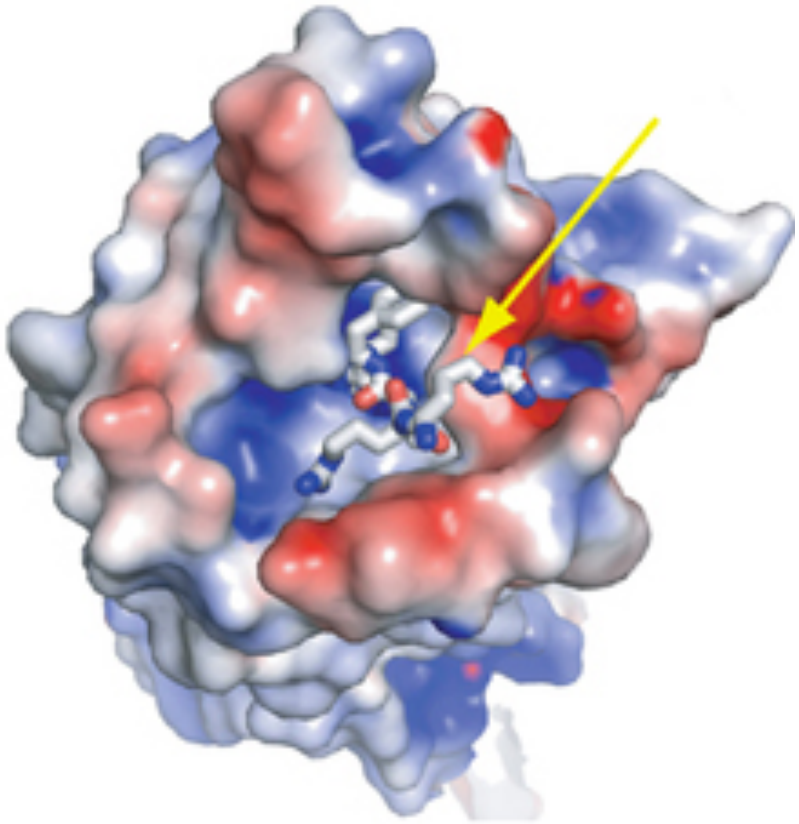


hormone acts as a switch - hormone binding alters receptor's structure which allows G protein to bind

Image credit: Albert Pan

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 tightly to the desired protein, alter behavior of the protein in a desired way, avoid binding to other proteins, etc.
- This requires solving challenging computational problems, even when a protein structure is already available

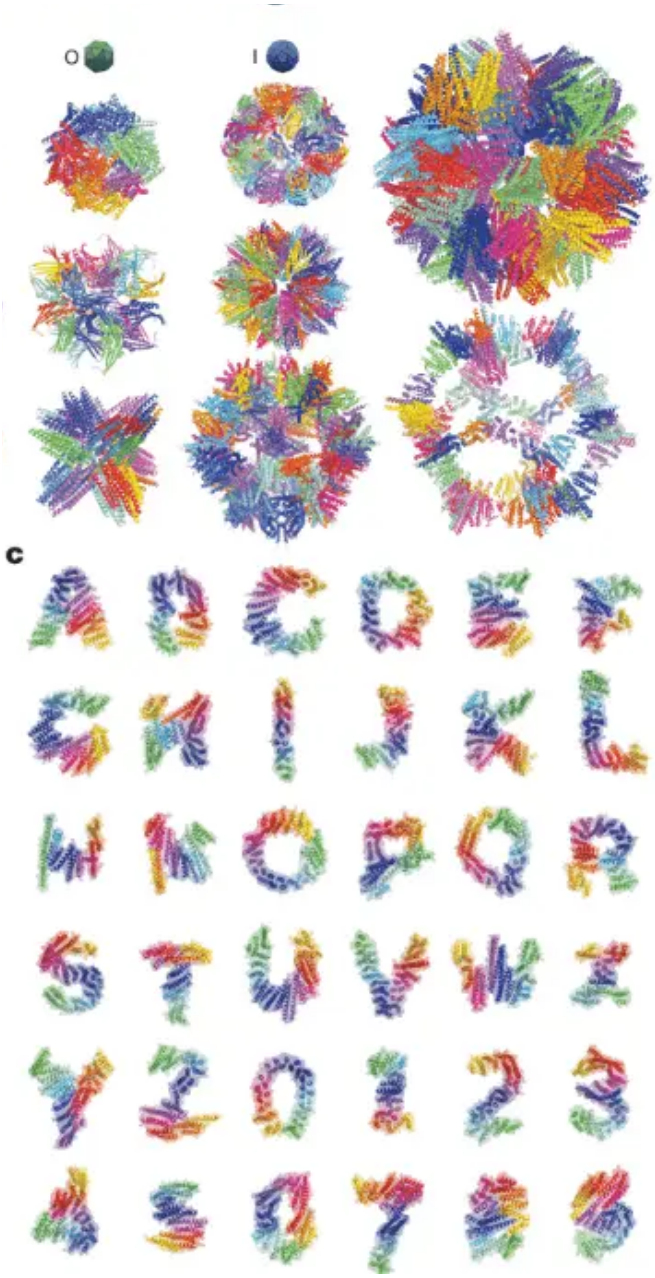


http://www.nih.gov/researchmatters/october2012/images/structure_l.jpg

Designing new biomolecular machines

- Protein design, RNA design, etc.
- Many applications within and beyond healthcare

goal of protein design is to create proteins with a particular shape so they perform a desired function



Ingraham et al., Illuminating protein space with a programmable generative model. *Nature* 623:1070–1078 (2023)

Overview of topics we'll cover

Biomolecular structure prediction

- Example: Protein structure prediction (“folding”)
  - Given the sequence of amino acids that make up a protein, predict its 3D structure

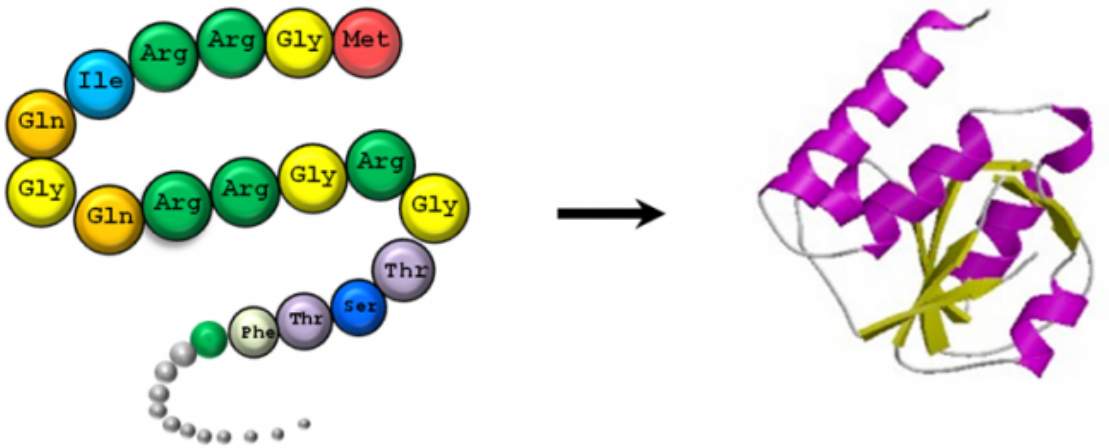


Image source: https://newenergyandfuel.com/wp-content/uploads/2014/09/Polypeptide-Chain.png

AlphaFold
August 2021

RoseTTAFold
August 2021

Biomolecular structure prediction

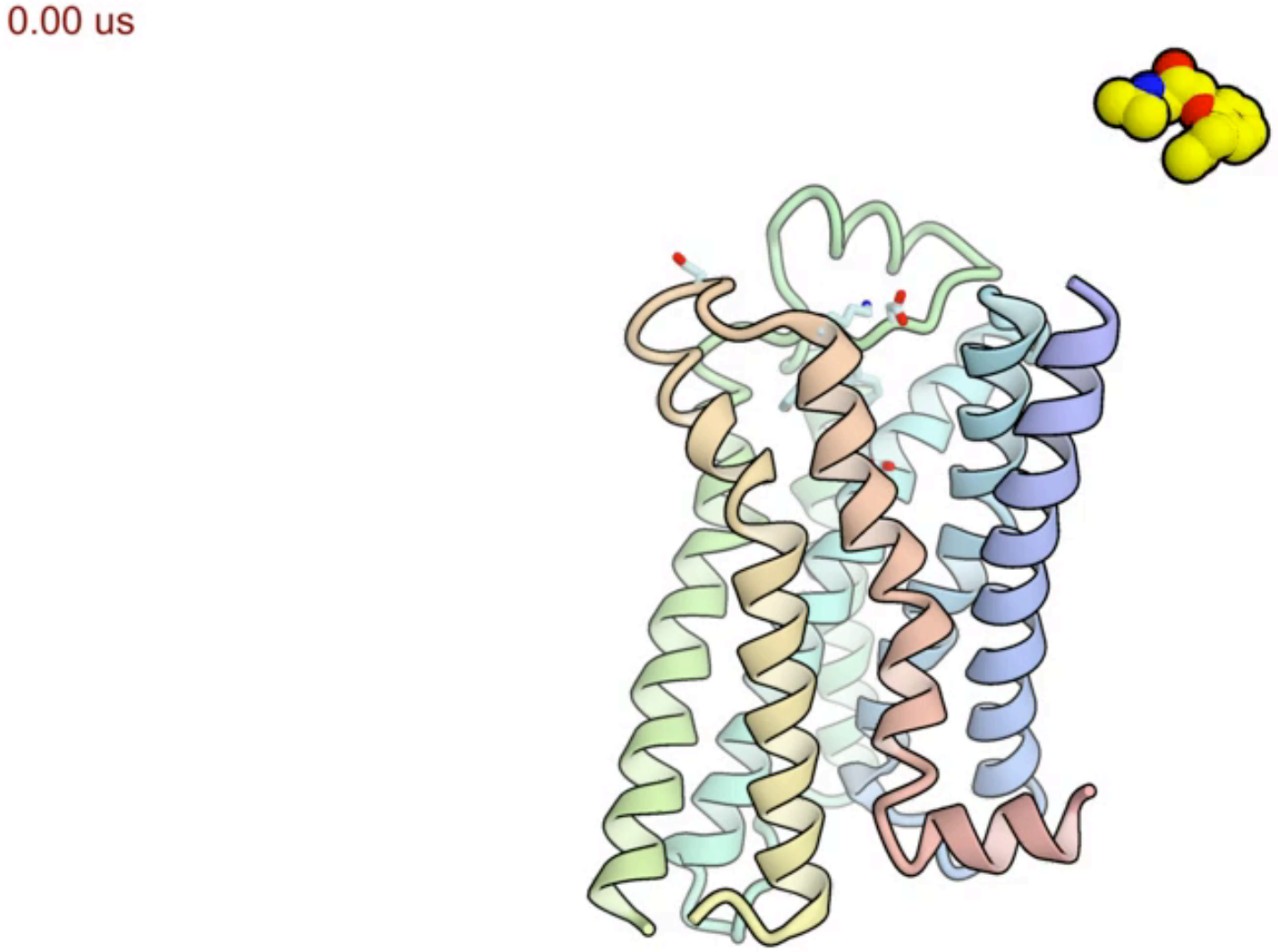
- Usually harder: predict structures of other biomolecules (e.g., RNA), or of multiple biomolecules bound to one another

Raphael Townshend, Stephan Eismann, Andrew Watkins, Ramya Rangan, Masha Karelina, Rhiju Das, and Ron Dror. Geometric deep learning of RNA structure. *Science* (August 2021)

Josh Abramson et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* (May 2024)

Molecular dynamics simulations

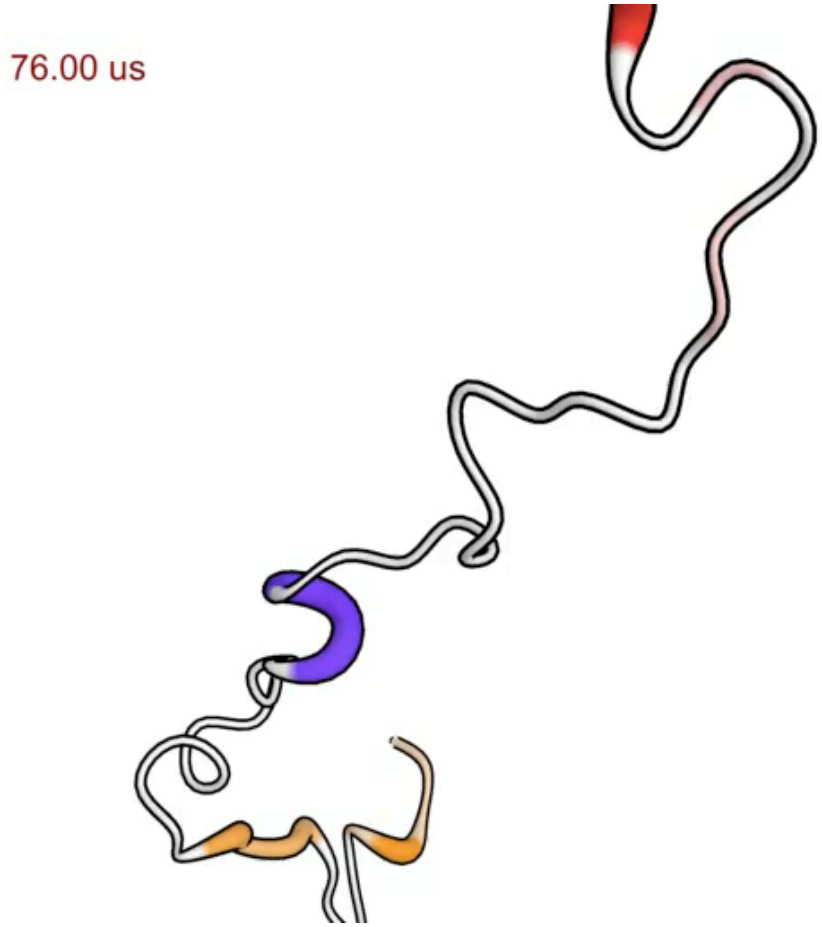
one way to capture how molecular structures change over time



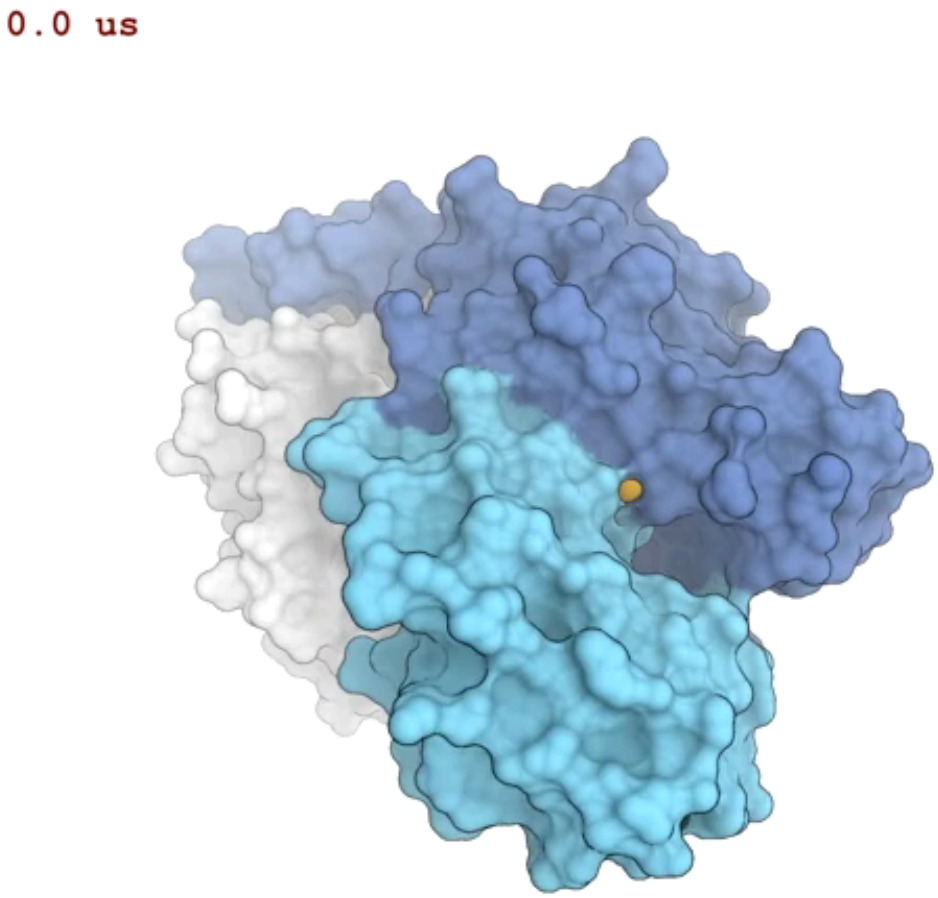
Beta-blocker binding to the β_2-adrenergic receptor

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)

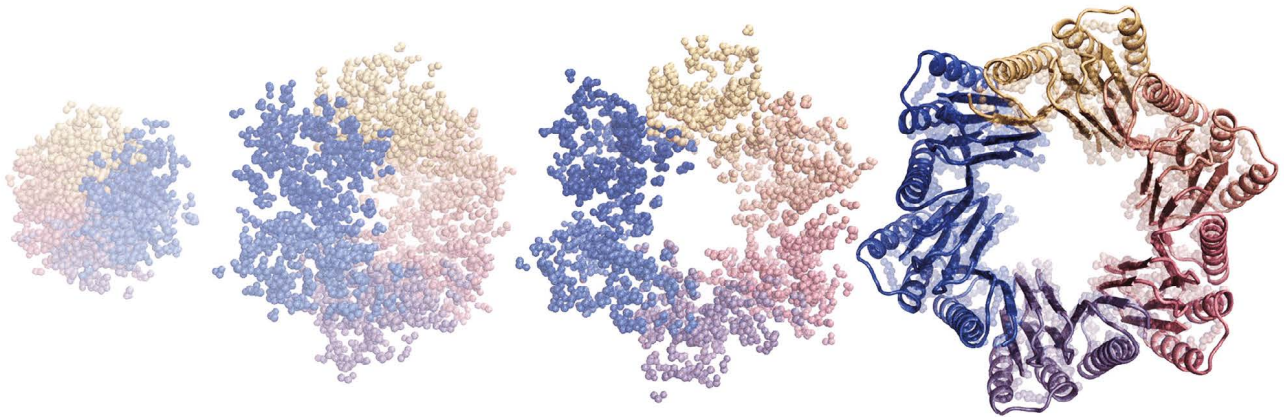
Protein design

- Given a desired protein structure (or function), design an amino acid sequence that achieves it

Divine *et al.*, Designed proteins assemble antibodies into modular nanocages. *Science* 372:eabd9994 (2021)

Protein design

- Given a desired protein structure (or function), design an amino acid sequence that achieves it



Two protein assemblies (right) were developed using an artificial-intelligence tool called RFdiffusion.

‘TRANSFORMATIVE’ AI DESIGNS CUSTOM PROTEINS ON DEMAND

Computer-devised biomolecules could form the basis of new vaccines or medicines. **By Ewen Callaway**

Nature | Vol 619 | 13 July 2023

AKA molecular docking

Ligand docking and virtual screening

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

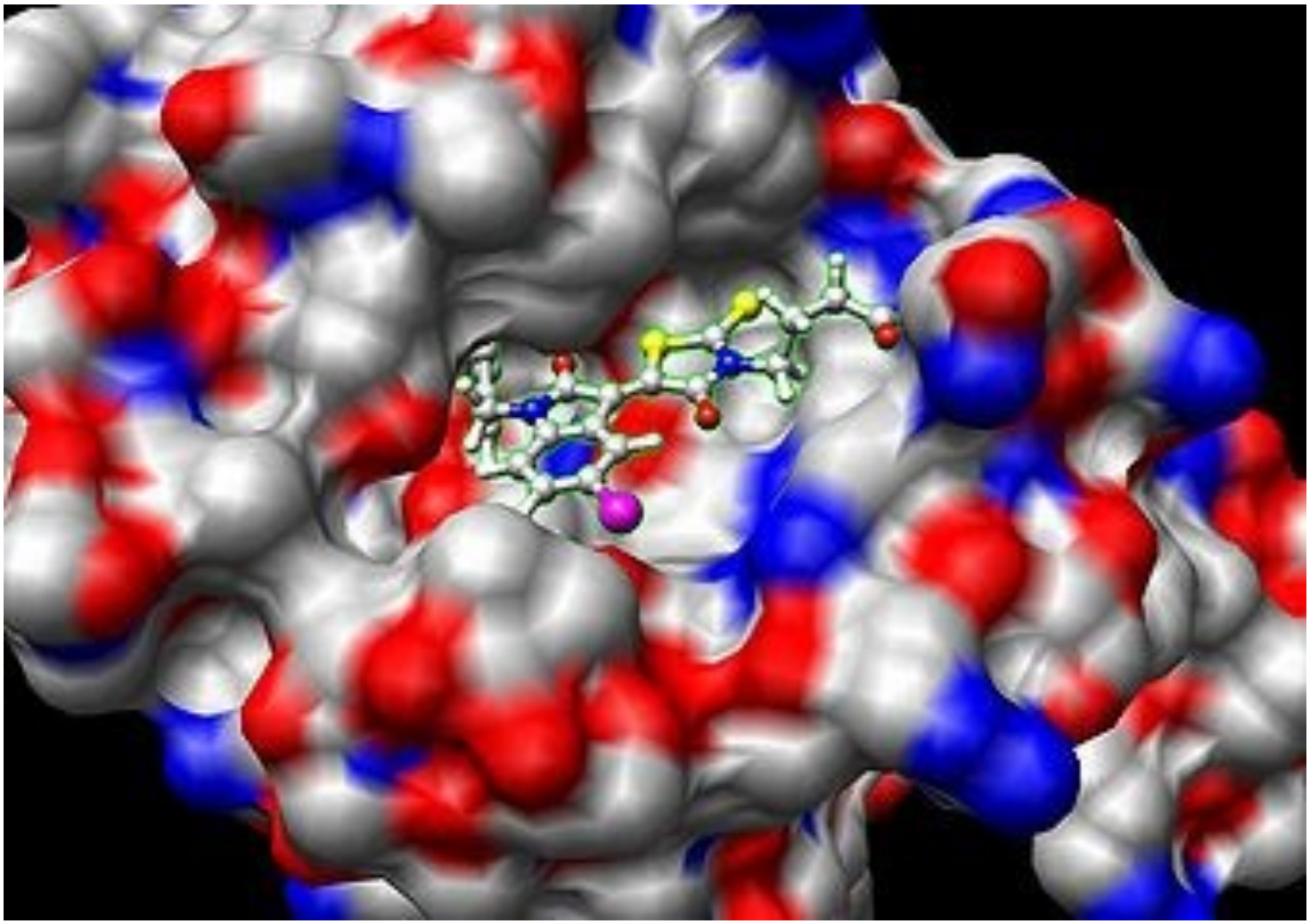
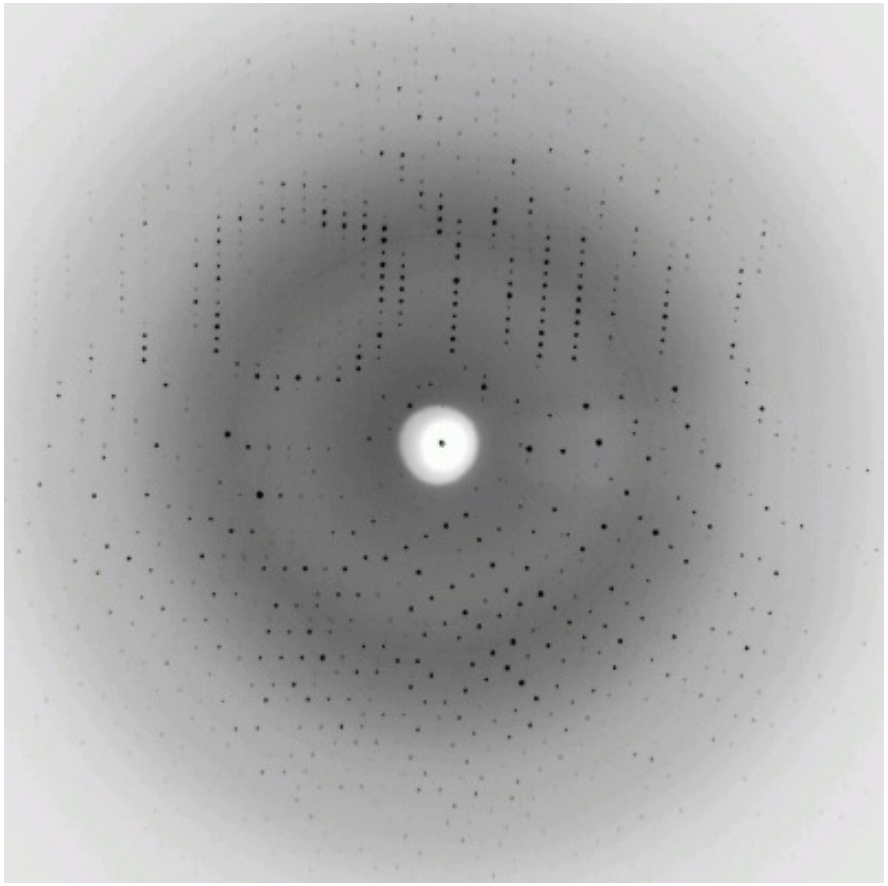


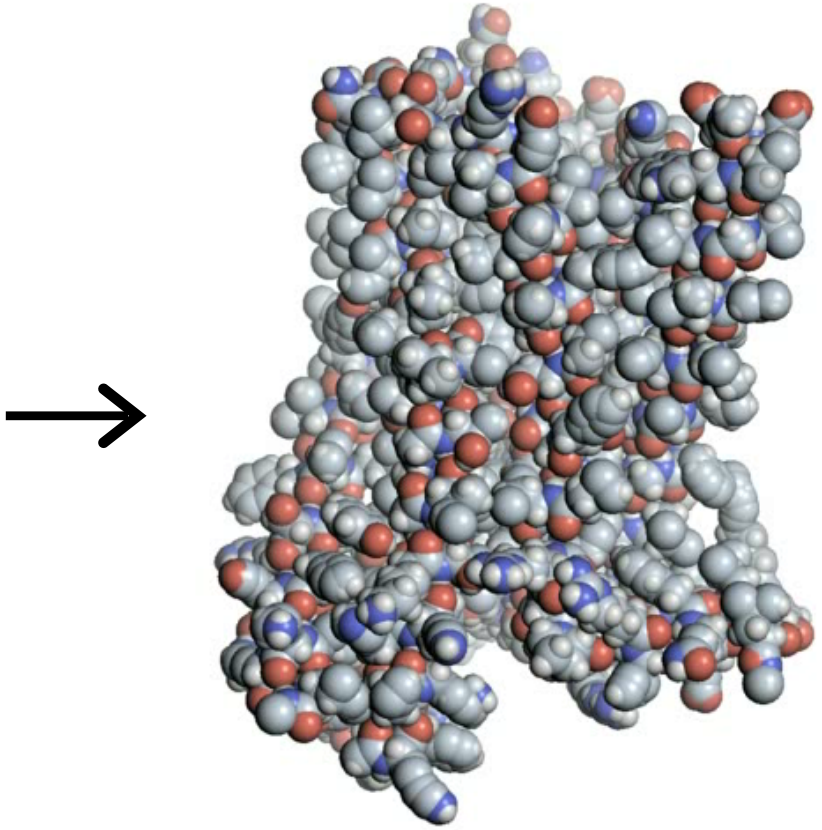
Image: Wikipedia

Determining molecular structures experimentally also requires solving challenging computational problems!

Determining molecular structures by crystallography

X-ray crystallography - most traditional way of experimentally determining molecular structure



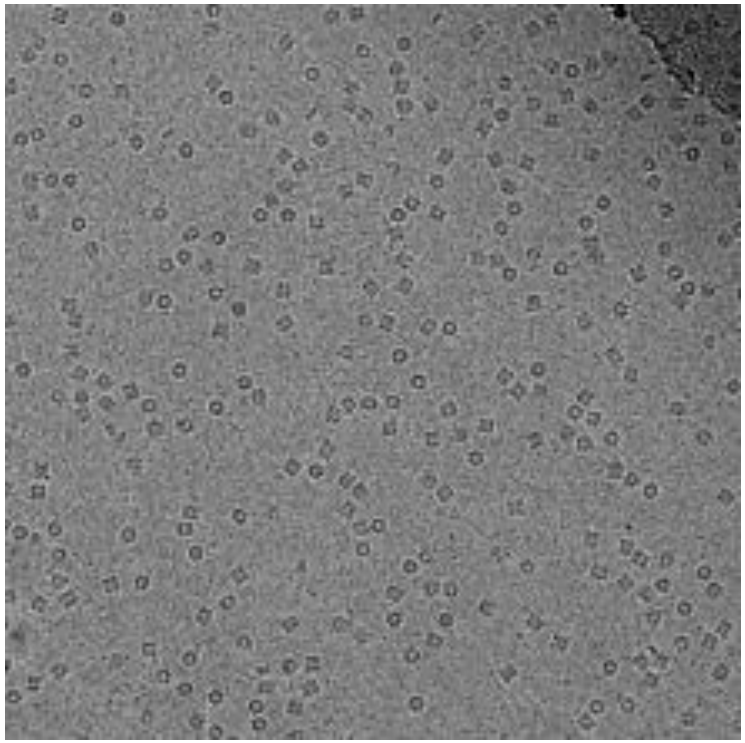


X-ray diffraction pattern

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

Protein structure

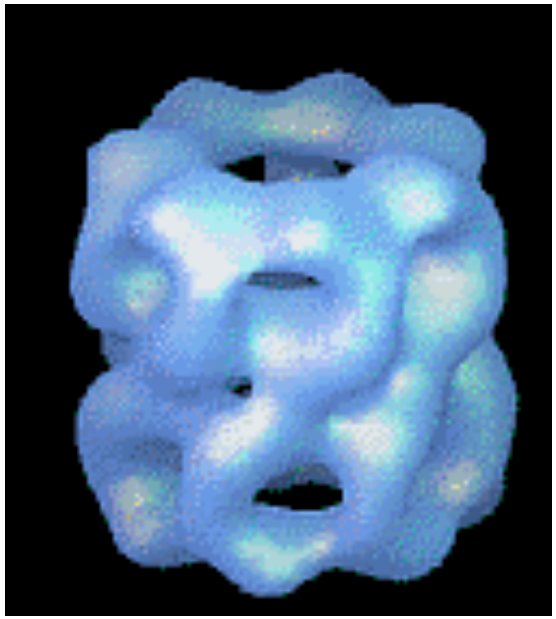
Determining molecular structures by cryogenic electron microscopy (CryoEM)



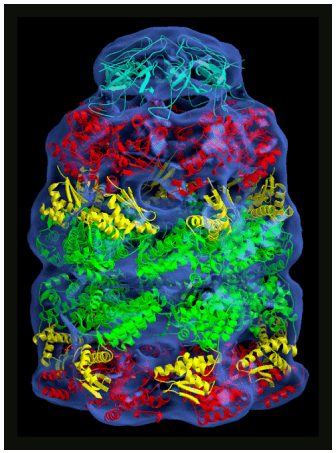
CryoEM image

raw CryoEM data is 10,000s of images like this

Image: Wikipedia



Reconstructed envelope

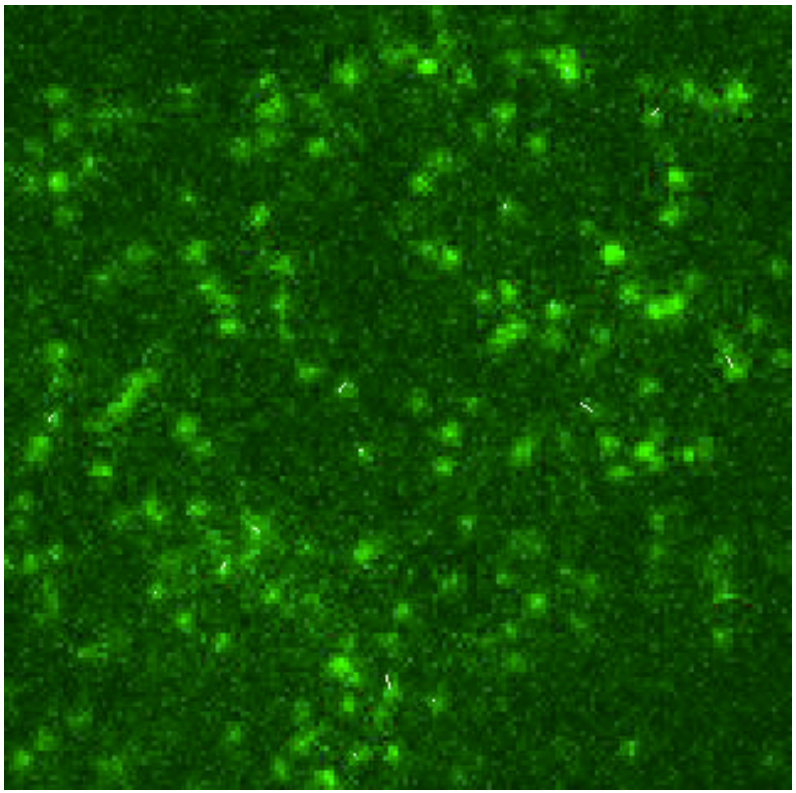


Structure

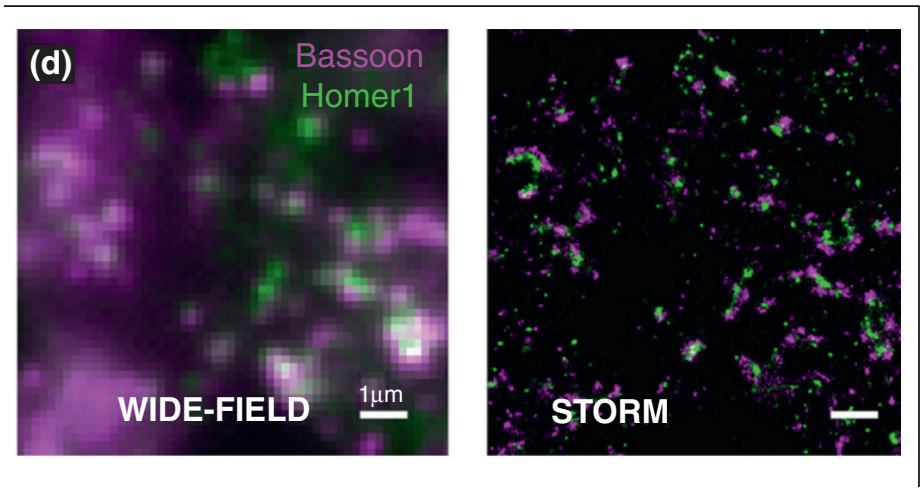
Image: http://people.cryst.bbk.ac.uk/~ubcg16z/chaperone.html

Fluorescence microscopy and cellular-level organization

track motions of individual molecules over time



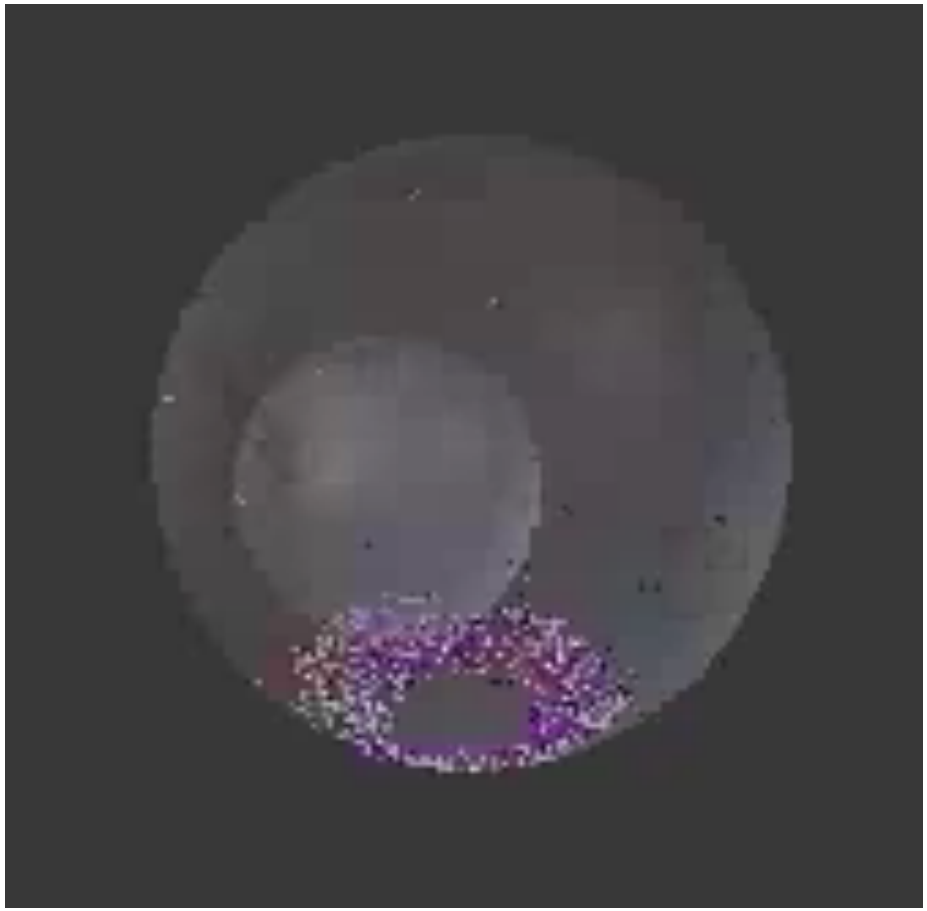
Data: Bettina van Lengerich, Natalia Jura
Tracking and movie: Robin Jia



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



can also predict molecular motions with simulation

Video: Naomi Latorraca

We'll also cover important underlying computational methods

- Machine learning
  - Supervised and unsupervised
- Image analysis
- Sampling from probability distributions

+ Fourier transforms

Previous familiarity with these concepts is not required!

Notes on course contents

- Course split roughly in two parts
  1. Atomic-level modeling of biomolecules
  2. Coarser-level modeling and imaging-based methods

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

- Some overlap in content with CS 274 (BIOE/BMI/GENE 214), but only about 20%.
  - This class (CS 279) is focused on structure. Much of CS 274 covers other bioinformatics topics.
  - Many students take both classes, in either order, or sometimes simultaneously.

Recurrent themes

Recurrent themes

- **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
- Similarities and differences in methods employed at **different spatial scales**
- **Energy functions** (which associate an energy or potential with each possible structure)
- Recurring math concepts: **Fourier transforms, convolution, Monte Carlo methods**

Course logistics

Course website

- https://cs279.stanford.edu
- See “Course policies and evaluation criteria” document on website
- To view last year’s lecture slides, follow “Fall 2023” link on website
  - This year’s content will be similar but not identical

Course announcements

- We will use Ed Discussion for announcements and for answering students' questions
  - https://edstem.org/us/courses/47160/discussion/
- If you can't access this page:
  - Create an Ed account using your Stanford (SUNet) email address
  - If you still can't access the page, email cs279-aut2425-staff@lists.stanford.edu

Expected background

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

- Assignments require coding in Python
  - You should have done some coding before.
  - You need not have used Python before, provided you're willing to learn.
    - Python and terminal tutorial: Friday at 10 am over Zoom. 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
  - I'll teach some additional relevant math concepts (e.g, Fourier transforms), with a focus on basic ideas/intuition rather than on equations.

Assignments and Exam

- Assignments
  - First three cover specific topics.
  - Fourth is a more open-ended “project.”
  - First assignment is shorter than second and third. For the project, we expect only a bit more work than the second and third assignments.
  - See collaboration and chatbot policy under “Course policies” on web page.

- Exam covering key concepts

Lectures

- Lecture live streams and recordings available to enrolled students on Canvas
  - Go to Canvas page for course and select “Panopto Course Videos” tab on the left-hand side
  - Or click here: https://canvas.stanford.edu/courses/195742

- Lecture slides will be available on course website, along with optional reading material

Participate in class

- I encourage you to join the class in real time (in person or virtually) and ask/answer questions
  - This makes the class better for everyone
  - We'll do small-group discussions in class

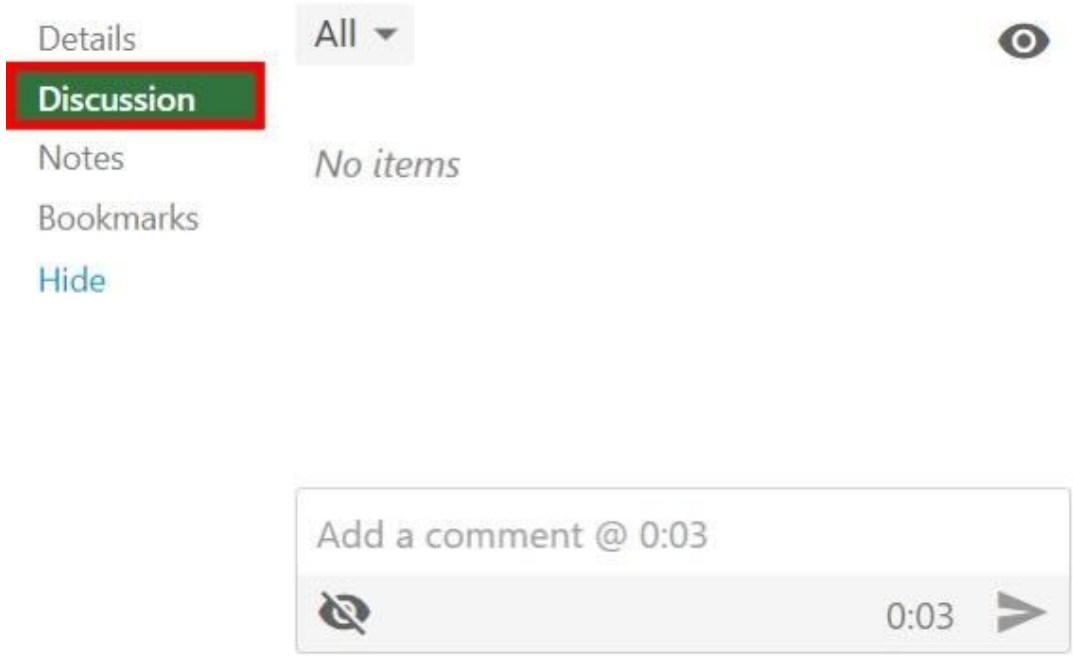
- You can earn extra credit for participation

- For those who are not available during class time:
  - Please watch recording before next class
  - You can also earn extra credit by answering other students' questions on Ed Discussion, or by completing optional "Challenge Questions" on assignments. Extra credit is not required to earn an A in the class.

Participate in class

- If joining in person, raise your hand to ask/answer questions
  - Please state your name

- If joining virtually, post questions/answers as comments through Panopto's Discussion feature, so that a TA can share them
  - Please post these as public comments. Do *not* select "moderator only."



Feedback welcome!

- I want to continue improving this course, and would appreciate your suggestions
- Please speak up when you don’t understand something

Course staff

- Prof. Ron Dror
  - http://drorlab.stanford.edu/rondror.html
  - Office hours: After every class, outside the classroom or at http://bit.ly/Ron-OH-2024

- TAs:
  - Ari Glenn
  - Briana Sobecks
  - Carrie Chen
  - Chiho Im
  - Hannah Park
  - Ishaan Singh
  - Masha Karelina
  - Luci Bresette

  - Office hours and contact info at cs279.stanford.edu in the “Office Hours” section, or via http://bit.ly/CS279-OH

- The best way to get most questions answered is by posting on Ed Discussion