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

Ron Dror
Stanford University

Sept 27, 2022

Image credit: Ansgar Philippsen
One-fifth of science Nobel Prizes relate to 3D structure/organization of biomolecules

- Biological structure is critical to:
  - Understanding how biology works
  - Diagnosing, preventing, and treating disease
  - Food and energy production (e.g., agriculture)
Computation plays a critical and rapidly growing role in this field

2013 Chemistry Nobel Prize: Computational models of biomolecules

‘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

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

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”.

Photo: A. Mahmoud

Martin Karplus
Prize share: 1/3

Michael Levitt
Prize share: 1/3

Arieh Warshel
Prize share: 1/3
Outline for this lecture

- What is structure?
  - 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

An adrenaline receptor
(the $\beta_2$ adrenergic receptor)
Protein dynamics

\( \beta_2 \) adrenergic receptor
Example: how LSD binds to its target

“Revealed: Why LSD Lasts So Long!”
AVI LSD YouTube Channel

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

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

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.

- 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)
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

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.

Designing new biomolecular machines

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

How?

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


AlphaFold
August 2021

RoseTTAFold
August 2021
Biomolecular structure prediction

- Usually harder: predict structures of other biomolecules (e.g., RNA), and of multi-molecule complexes

Molecular dynamics simulations

Beta-blocker binding to the $\beta_2$-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.

Ligand docking and virtual screening

Searching for potential drug molecules that bind to a target (usually a protein), and determine how they bind
Determining molecular structures experimentally also requires solving challenging computational problems!
Determining molecular structures by crystallography

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

Reconstructed envelope

Structure


Image: http://people.cryst.bbk.ac.uk/~ubcg16z/chaperone.html
Fluorescence microscopy and cellular-level organization

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

- Including super-resolution microscopy

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

Video: Naomi Latorraca
We’ll also cover important underlying computational methods

• Machine learning
  – Supervised and unsupervised
• Image analysis
• Sampling from probability distributions

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 2021” 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/29672/discussion

• If you can’t access this page:
  – Creat an Ed account using your SUNET email
  – If you still can’t access the page, email cs279staff@cs.stanford.edu
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
  – 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 policy under “Course policies” on web page.

• Exam covering key concepts
Lectures

• Lectures will not be recorded. Lecture slides will be available (with annotation) on website, along with additional notes for some lectures and pointers to optional reading material.

• Missing some lectures is fine.
  – For lectures you miss, read through the annotated lecture slides. TAs can explain anything that’s not clear to you during office hours.
  – If you’ll need to miss most or all lectures, I don’t recommend taking the course.

• Class participation encouraged!
  – We won’t take attendance.
  – You can earn extra credit for participation (in class, or by answering other students’ questions on Ed Discussion), but this isn’t required to get an A or even an A+.
Feedback welcome!

• I want to continue improving this course, and would appreciate your suggestions

• Please speak up when you don’t understand something
COVID-19 safety

• Per Stanford requirements, **masks are required**
  – You can take your mask off very briefly to sip water. Do not talk while it’s off.
  – No eating in class.

• COVID-19 vaccination required
Course staff

• Prof. Ron Dror
  – Office hours: After every class, outside the classroom

• TAs:
  – Armaun Sanayei
  – Grace Kim
  – Divya Nagaraj
  – Hyun Soo Jeon
  – Tristan Saucedo

– Office hours and contact info at [cs279.stanford.edu](http://cs279.stanford.edu)