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

AND THE WINNER OF THE NOBEL PRIZE IN SOFTWARE IS...
The Nobel Prize in Chemistry 2013

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”.

- Martin Karplus
- Michael Levitt
- Arieh Warshel

Prize share: 1/3
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 $\beta_2$ 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
Protein dynamics

Atoms in protein move (NOT static)

$\beta_2$ 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

David Goodsell

http://www.nikoninstruments.com/Products/Microscope-Systems/Inverted-Microscopes/N-STORM-Super-Resolution/(gallery); Zhuang group
Intracellular dynamics (artist’s rendition)
Why is structure important?
Genomics is a great start ....

- 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.

<table>
<thead>
<tr>
<th>REF. NO.</th>
<th>IBM NO.</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>156011</td>
<td>Track Frame 21&quot;, 22&quot;, 23&quot;, 24&quot;, Team Red</td>
</tr>
<tr>
<td>2</td>
<td>157040</td>
<td>Fork for 21&quot; Frame</td>
</tr>
<tr>
<td>2</td>
<td>157039</td>
<td>Fork for 22&quot; Frame</td>
</tr>
<tr>
<td>2</td>
<td>157038</td>
<td>Fork for 23&quot; Frame</td>
</tr>
<tr>
<td>2</td>
<td>157037</td>
<td>Fork for 24&quot; Frame</td>
</tr>
<tr>
<td>3</td>
<td>191202</td>
<td>Handlebar TTT Competition Track Alloy 15/16&quot;</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Handlebar Stem, TTT, Specify extension</td>
</tr>
<tr>
<td>5</td>
<td>191278</td>
<td>Expander Bolt</td>
</tr>
<tr>
<td>6</td>
<td>191272</td>
<td>Clamp Bolt</td>
</tr>
<tr>
<td>7</td>
<td>145841</td>
<td>Headset Complete 1 x 24 BSC</td>
</tr>
<tr>
<td>8</td>
<td>145842</td>
<td>Ball Bearings</td>
</tr>
<tr>
<td>9</td>
<td>190420</td>
<td>175 Raleigh Pistard Seta Tubular Prestavalve 27&quot;</td>
</tr>
<tr>
<td>10</td>
<td>190233</td>
<td>Rim, 27&quot; AVA Competition (36H) Alloy Prestavalve</td>
</tr>
<tr>
<td>11</td>
<td>145973</td>
<td>Hub, Large Flange Campagnolo Pista Track Alloy (pairs)</td>
</tr>
<tr>
<td>12</td>
<td>190014</td>
<td>Spokes, 11 5/8&quot;</td>
</tr>
<tr>
<td>13</td>
<td>145837</td>
<td>Sleeve</td>
</tr>
<tr>
<td>14</td>
<td>145636</td>
<td>Ball Bearings</td>
</tr>
<tr>
<td>15</td>
<td>145170</td>
<td>Bottom Bracket Axle</td>
</tr>
<tr>
<td>16</td>
<td>145838</td>
<td>Cone for Sleeve</td>
</tr>
<tr>
<td>17</td>
<td>144673</td>
<td>L.H. Adjustable Cup</td>
</tr>
<tr>
<td>18</td>
<td>145833</td>
<td>Lockring</td>
</tr>
<tr>
<td>19</td>
<td>145239</td>
<td>Straps for Toe Clips</td>
</tr>
<tr>
<td>20</td>
<td>145834</td>
<td>Fixing Bolt</td>
</tr>
<tr>
<td>21</td>
<td>145835</td>
<td>Fixing Washer</td>
</tr>
<tr>
<td>22</td>
<td>145822</td>
<td>Dustcap</td>
</tr>
<tr>
<td>23</td>
<td>145823</td>
<td>R.H. and L.H. Crankset with Chainwheel</td>
</tr>
<tr>
<td>24</td>
<td>146472</td>
<td>Fixed Cup</td>
</tr>
<tr>
<td>25</td>
<td>145235</td>
<td>Toe Clips, Christope, Chrome (Medium)</td>
</tr>
<tr>
<td>26</td>
<td>145684</td>
<td>Pedals, Extra Light, Pairs</td>
</tr>
<tr>
<td>27</td>
<td>123021</td>
<td>Chain</td>
</tr>
<tr>
<td>28</td>
<td>145980</td>
<td>Seat Post</td>
</tr>
<tr>
<td>29</td>
<td></td>
<td>Seat Post Bolt and Nut</td>
</tr>
<tr>
<td>30</td>
<td>167002</td>
<td>Saddle, Brooks</td>
</tr>
<tr>
<td>31</td>
<td>145933</td>
<td>Track Sprocket, Specify 12, 13, 14, 15, or 16 T.</td>
</tr>
</tbody>
</table>
... 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)
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.
Designing new biomolecular machines

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

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 $\beta_2$-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)

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

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

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

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)
Machine learning in structural biology

Deep neural network → Interaction Probability

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
  – 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](http://cs279.stanford.edu)