Computational biology in four dimensions

Ron Dror

Jan. 9, 2017

Image credit: Sam Hertig
Outline for lecture 1 (course overview)

- Course format and objectives
- What is structure?
  - Structure (and dynamics) at multiple spatial scales
- Why is structure important?
- Overview of course topics
  - Atomic-level modeling of biomolecules
  - Structures of macromolecular complexes
  - Cellular-level organization
- Course logistics
- Guidelines for presentations and critiques
- Immediate next steps
Course format and objectives
Focus is on presentation, discussion, and critique of cutting-edge research

- The majority of class time will be spent on presentations and discussion of important recent research papers
- Most presentations will be by students
  - Each student will present on a paper of interest to them
  - Professor & TAs will meet with students to review presentations
  - Professor will introduce topics in class
  - Professor and guest speaker will give initial presentations
- Professor will facilitate discussion. Students are strongly encouraged to share comments, questions
Course requirements

• One presentation
  – 20-30 minutes each
• Three written critiques of papers being presented
• Read the remaining papers, attend class, and participate in discussions
What do I want students to learn from this course?

1. Gain exposure to cutting-edge computational research in structural biology, broadly defined (a rapidly growing, interdisciplinary area).
2. Learn to critique and evaluate research, and practice critical reading of research papers.
3. Refine the skill of presenting deep technical material to a non-expert audience.

The latter two are broadly important skills. They’re especially important to practice if they make you a bit nervous.
Relationship to CS/CME/BioE/Biophys/BMI 279

• 371 and 279 cover the same general field
• 279 is a traditional lecture/homework course covering basics of the field
• 371 focuses on current research topics
• Neither is a prerequisite for the other
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.
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
Proteins (and other molecules) often come together to form macromolecular complexes.
These come together to form organelles

Synaptic vesicle
http://www.mpiibpc.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?
The cycle of life

From Michael Levitt
Genomics is a great start ....

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

<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>3</td>
<td>157039</td>
<td>Fork for 22&quot; Frame</td>
</tr>
<tr>
<td>4</td>
<td>157038</td>
<td>Fork for 23&quot; Frame</td>
</tr>
<tr>
<td>5</td>
<td>157037</td>
<td>Fork for 24&quot; Frame</td>
</tr>
<tr>
<td>6</td>
<td>191202</td>
<td>Handlebar TTT Competition Track Alloy 15/16&quot;</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Handlebar Stem, TTT, Specify extension</td>
</tr>
<tr>
<td>8</td>
<td>191278</td>
<td>Expander Bolt</td>
</tr>
<tr>
<td>9</td>
<td>191272</td>
<td>Clamp Bolt</td>
</tr>
<tr>
<td>10</td>
<td>145841</td>
<td>Headset Complete 1 x 24 BSC</td>
</tr>
<tr>
<td>11</td>
<td>145842</td>
<td>Ball Bearings</td>
</tr>
<tr>
<td>12</td>
<td>190420</td>
<td>175 Raleigh Pistard Seta Tubular Prestavalve 27&quot;</td>
</tr>
<tr>
<td>13</td>
<td>190233</td>
<td>Rim, 27&quot; AVA Competition (36H) Alloy Prestavalve</td>
</tr>
<tr>
<td>14</td>
<td>145973</td>
<td>Hub, Large Flange Campagnolo Pista Track Alloy (pairs)</td>
</tr>
<tr>
<td>15</td>
<td>190014</td>
<td>Spokes, 11 5/8&quot;</td>
</tr>
<tr>
<td>16</td>
<td>145837</td>
<td>Sleeve</td>
</tr>
<tr>
<td>17</td>
<td>145836</td>
<td>Ball Bearings</td>
</tr>
<tr>
<td>18</td>
<td>145170</td>
<td>Bottom Bracket Axle</td>
</tr>
<tr>
<td>19</td>
<td>145838</td>
<td>Cone for Sleeve</td>
</tr>
<tr>
<td>20</td>
<td>146473</td>
<td>L.H. Adjustable Cup</td>
</tr>
<tr>
<td>21</td>
<td>145833</td>
<td>Lockring</td>
</tr>
<tr>
<td>22</td>
<td>145239</td>
<td>Straps for Toe Clips</td>
</tr>
<tr>
<td>23</td>
<td>145834</td>
<td>Fixing Bolt</td>
</tr>
<tr>
<td>24</td>
<td>145835</td>
<td>Fixing Washer</td>
</tr>
<tr>
<td>25</td>
<td>145822</td>
<td>Dustcap</td>
</tr>
<tr>
<td>26</td>
<td>145823</td>
<td>R.H. and L.H. Crankset with Chainwheel</td>
</tr>
<tr>
<td>27</td>
<td>146472</td>
<td>Fixed Cup</td>
</tr>
<tr>
<td>28</td>
<td>145235</td>
<td>Toe Clips, Christophe, Chrome (Medium)</td>
</tr>
<tr>
<td>29</td>
<td>145684</td>
<td>Pedals, Extra Light, Pairs</td>
</tr>
<tr>
<td>30</td>
<td>123021</td>
<td>Chain</td>
</tr>
<tr>
<td>31</td>
<td>145980</td>
<td>Seat Post</td>
</tr>
<tr>
<td></td>
<td>167002</td>
<td>Seat Post Bolt and Nut</td>
</tr>
<tr>
<td></td>
<td>145933</td>
<td>Saddle, Brooks</td>
</tr>
<tr>
<td></td>
<td></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, 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
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.

Designing new biomolecular machines

- Protein design (for health or industrial applications)
- Cell design?
A strikingly large share of Nobel Prizes have recognized work on molecular structure.
2013 Nobel Prize recognized early developments underlying modern biomolecular computation

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".
2014 Nobel Prize recognized microscopy techniques used to study cellular structure (which also rely on computation)

The Nobel Prize in Chemistry 2014

Eric Betzig
Prize share: 1/3

Stefan W. Hell
Prize share: 1/3

William E. Moerner
Prize share: 1/3

The Nobel Prize in Chemistry 2014 was awarded jointly to Eric Betzig, Stefan W. Hell and William E. Moerner "for the development of super-resolved fluorescence microscopy".
Overview of course topics
Overview of course topics

Atomic-level modeling of biomolecules
Protein structure prediction and RNA/protein design by video game

screen shot from Foldit
Machine learning on structures

- Applying machine learning techniques to 3D structures of proteins or drug-like molecules — for example, for virtual drug screening

Gonczarek et al., NIPS 2016
Drug screening: computing accurate binding strengths for drug candidates

• “Alchemical” methods for computing binding free energies: simulate the drug as it gradually “disappears” from the binding pocket or “changes” into a different drug
• Sounds like magic, but it’s actually much more faithful to the physics than traditional docking methods

Beta-blocker alprenolol binding to an adrenaline receptor (Dror et al., PNAS 2011)
Coevolution methods for predicting structure from large numbers of genetic sequences

Key idea: amino acids in direct physical contact within a protein tend to mutate in a correlated fashion. Given enough sequence data, one can use this fact to predict structure.
Modern protein design

Design of a protein that selectively transports zinc ions through a cell membrane

**Fig. 1. Computational design and molecular dynamics simulations of Rocker.** (A) Schematic of the goal of obtaining conformational exchange between two oppositely oriented symmetry-frustrated states without being trapped in a symmetric state with both sites simultaneously occupied. (B) Metal site consists of a set of Exo-H motifs and a single Glu residue from each of the tight dimers. (C) Helical-wheel diagram of Rocker peptide. (D) The repacking algorithm placed Ala residues at the tight interface and Phe residues at the loose interface. Empty metal site on the left is omitted for clarity. (E) MD simulation of the design model with two Zn\(^{2+}\) ions placed at one metal site show stable interhelical distances for both tight and loose interfaces. Continuing the simulation after removing the Zn\(^{2+}\) ions maintained the tight interfaces, but resulted in an increased interhelical distance at the loose interface.

Molecular dynamics simulations and Markov State Models

Dror et al., Science 2015

https://folding.stanford.edu/home/faq/faq-simulation/
New methods for solving tough crystal structures

- Computational problem in crystallography: go from diffraction pattern to structure
  - Mathematically: invert a Fourier transform without phase information
- New methods for:
  - Solving low-resolution crystal structures using computational protein modeling
  - Solving structures using data from the new x-ray free electron lasers, which vaporize the crystal as they image it

Nucleic acid (i.e., DNA and RNA) structure and design

Computational design of co-assembling protein–DNA nanowires (Mou et al., *Nature* 2015)
Overview of course topics

Structures of macromolecular complexes
Single-particle electron microscopy

- The current revolution in structure determination through electron microscopy is due in part to new computational methods — e.g., Bayesian approaches.

*Image from Joachim Frank*

http://biomachina.org/courses/structures/091.pdf

Predicting protein–protein interactions

• Which proteins bind to one another?
• What is the structure of the bound complex?
• What types of networks do they form?

Calderwood et al., PNAS 104:7606-11, 2007

http://bmm.crick.ac.uk/~bmmadmin/Affinity/
Integrative modeling: combining diverse experimental data to deduce structures of large complexes

Alber et al., Nature 2007
Genome architecture

Techniques for mapping out the three-dimensional organization (and dynamics) of your DNA using a combination of computation and experiment
Overview of course topics

Cellular-level organization
Super-resolution microscopy methods

- New super-resolution microscopy methods
- Analysis methods that exploit compressed sensing

Tracking the motion of single molecules

- Fluorescence microscopy makes it possible to follow the motion of individual molecules in live cells, but actually doing the tracking is a challenging image analysis/computer vision problem.

Data: Bettina van Lengerich, Natalia Jura
Tracking and movie: Robin Jia
Cellular-level simulations

http://www.readdy-project.org/
Learning on cell shape and structure

Johnson et al., PLOS Computational Biology 11:e1004614 (2015)
Course logistics
Course web page


- Evaluation criteria on web page

- Please sign up on Piazza (via link on webpage) so that you get announcements
Course staff

• Prof. Ron Dror
  – Office hours: Monday and Wednesday after class (2:50-3:45), in classroom and then in Gates 204, or by appointment

• TA: Anthony Ma
  – Office hours: Tuesdays 1:30-2:30, Lathrop Tech Lounge

• TA: Osama El-Gabalawy
  – Office hours TBA
Expected background

• Course is intended to be broadly accessible to students with computational or biological backgrounds

• You should have:
  – Introductory biology background, at the level of a first-year course, and ideally also introductory chemistry and physics background
  – Introductory computer science background (e.g., CS 106A)
  – Math through calculus, and some previous exposure to mathematical modeling or probability.

• You’re not expected to be an expert in any of the areas we cover, but you are expected to be willing to read and present papers, ask questions, and share your thoughts
Honor code

• I’m not a tough grader, but I take the honor code seriously.

• In the context of this course, this means:
  – Avoid verbatim or near-verbatim copying of text from any source (unless it’s enclosed in quotation marks and clearly attributed to the source). You need to explain ideas in your own words.
  – Make sure that all materials (including figures) in presentations and critiques are properly cited.
Guidelines for presentations and critiques
Giving a good presentation

• Why should you care?
  – To develop a critical skill
  – To help make this a good course!

• See “Tips for Giving Clear Talks” by Kayvon Fatahalian (Stanford CS PhD 2010)
  – This focuses on giving talks on your own research but is applicable to the presentations for this class as well. Possible exception: I’d like you to point out both strengths and weaknesses in the work you present.
Giving a good presentation

• Make it understandable to a diverse audience
  – Students in class are from: CS, CME, EE, BioE, ChemE, chemistry, chemical & systems biology, biophysics, and more
  – Include appropriate background material. You may need to read other, related papers as well.

• Describe the main idea intuitively
  – You need to figure out what the main idea is. (Ask course staff for help if necessary.)
  – Do not rely on equations or code to convey intuition
  – You do not need to cover everything in the paper

• Strive for clarity
  – See Kayvon’s presentation

• Make it interesting and exciting!
Preparing for presentations

• Each presenter should meet with a TA, and then with me (Ron), before their presentation
  – By default, meet with me right after the class preceding your presentation. If one of the presenters has a conflict then, schedule a different time.

• You should prepare a complete draft of your presentation before these meetings

• You need to coordinate with the other students presenting on the same day.
  – For example, you may want shared introductory material
  – I would like to meet with all of you together
Writing a critique

- Although you should include a brief summary of the paper you’re critiquing, that shouldn’t be the focus.
- Instead, the focus should be on:
  - Strengths and weaknesses of the approach or results.
  - Other approaches that might have been applied instead.
  - Potential extensions or follow-on work.
- You’re encouraged to read related papers to help you write this.
  - Please indicate in the critique which other papers you read.
Immediate next steps
Choosing topics for presentation and critique

• By next Tuesday, Jan. 17, email cs371-win1617-staff@lists.stanford.edu a list of six topics (class sessions) you’d be interested in presenting on and twelve papers (or six topics) you’d be interested in critiquing
  – My default list of papers is at http://cs371.stanford.edu/schedule.html and will be complete by Tuesday afternoon. The “Main Papers” are the ones I’m expecting students to present.
  – You’re also welcome to suggest your own paper(s) of interest to present. If you do, specify any dates on which you can’t present.
  – You may rank order your lists if you’d like (you should specify that you’ve done so)

• Also include:
  – Brief description of your background
Next couple class periods

- Wednesday, Jan. 11: Ron to present on simulation of drug targets and simulation analysis
  - Reading materials on website
- Monday, Jan. 16: Guest lecture by Eli Groban (Autodesk) on virtual reality for biomolecules
  - Completely optional (MLK day).
Volunteers for initial presentations & critiques

- Email cs371-win1617-staff@lists.stanford.edu ASAP to volunteer to present or critique during the following four class periods:
  - Jan. 18: Using multiplayer online video games for structure prediction and design
  - Jan. 23: Coevolution methods for predicting structure from large numbers of genetic sequences
  - Jan. 25: Modern protein design
  - Jan. 27: Machine learning for structure-based virtual screening