# Introduction

BIOMEDIN/BIOPHYS/CME/CS 279

Notes by Osama El-Gabalawy — 22 September 2015

## Lecture 1

In this class we will learn about the 3D Structure and Organization of Biomolecules and Cells. The intro lecture will cover the three primary questions and go over recurrent themes.

### What is structure?

- Structure is 3D spatial organization
- Carefully designed **structures** with **moving parts** carry out a **specific function** like a machine
- Structure and dynamics at multiple scales
  - Proteins (most of them) are relatively large molecules with defined structures. Protein can be thought of as a machine.
  - When the machine is moving in action, it is equivalent to **protein dynamics** (ie wiggling, shaking, and movement of the molecule).
  - The next level is macromolecular complexes: several proteins coming together to form a large biological complex.
  - Macromolecular complexes come together to form organelles, which together form a cell, which together form tissues which together form organs which together form the body.
  - In this class we will focus on the molecular through cellular scales.
- Dynamics on larger scales include intracellular dynamics which is all about motion and pieces coming together.

### Why structure is important?

- There's a very direct relationship between gene sequences and protein structure The DNA Sequence codes for amino acids that comprise the protein. The amino acid sequence determines the 3D-structure and function of the protein. The function of a protein acts on the original DNA sequence through evolution, and completes the cycle.

- Genomics is a list of parts. It's a good start but not enough info to understand the machine. We want a photograph of the machine, not just a list of parts. We want to go from genetic sequence data to how biological systems work. There are many applications of this knowledge in drug design, medical diagnostics, etc.
- Examples of how structure determines function
  - Kinesin motor protein "walking" (click for video link)
  - Ribosomes are macromolecular complexes that make proteins by putting amino acids together. A ribosome's parts come together and dissociate dynamically. (click for gif link)
  - G protein-coupled receptors (GPCRs) are membrane proteins that relay information to the inside of the cell by sensing presence of specific molecules outside of the cell. Activation of GPCRs allows it to bind to other proteins (most commonly G-proteins, hence the name). A third of the drugs in the market target these proteins by making them more or less active.
- Structure-based drug design
  - It's very expensive to just mix a bunch of chemicals with proteins and hope for an effective drug. Pharmaceutical companies are pouring money into rational drug design by structure-based analysis and rational design.
  - One limitation is that structures are static based photographs – we want an idea of the whole machine in motion.
- Protein Design
  - If you have a particular shape or function that you want the challenge is how to pick the protein sequence to get that function/shape.
  - There is a lot of methods to do it, and almost all of them employ computational methods.

## How computation plays a role?

- How does computation help us understand structure and spatial organization? There's a lot of ways to do it, and most of them employ computational methods.
- Protein Structure Prediction and the Protein Folding Problem
  - Given sequence of protein, predict the folded structure of the protein.
  - Turns out most proteins can spontaneously form their final structure from an unstructured startpoint, so the structure is, for the most part, inherent in the sequence

- Molecular Dynamics
  - Using physical models of atoms to predict how molecules/proteins interact with other molecules
  - Can use MD to simulate binding process of beta-blocker for example or folding of protein G (although this does not work for most molecules)
- Ligand Docking
  - This is a class of methods for screening for potential active drugs.
  - Given structure of a protein, you generally know where the drug needs to bind on the protein to induce an effect.
  - Thus the challenge is when given a small-molecule drug, you want to be able to predict with a heuristic method how well and how tight a candidate drug binds to a specific protein.
  - The method needs to be really quick in order to screen many drugs efficiently.
- Protein Design
  - The challenge here is to come up with sequence of amino acids that will yield a specific, desired protein structure
  - Top7 was the first protein to be designed to have a novel fold (although there is some controversy about this).
- Solving X-Ray Crystal Structures
  - The protocol to solve a protein structure by X-Ray Crystallography is to first crystallize proteins and then shoot a powerful x-ray beam yielding diffraction pattern. Repeat this many times to get different diffraction patterns.
  - Scientists use computational methods on the different X-ray diffraction patterns to solve the protein structure.
- Solving Structures by Cryo-electron Microscopy
  - This method is very useful for large protein complexes at very low temps (hence cryo)
  - Electron microscopy yields many 2D images of the complex, but you don't know what the orientations are. Moreover, the images tend to have low resolution.
  - The computational challenge is to put together 1,000's of images to get a structure
- There are other methods where researchers combine experimental data to deduce structures of huge complexes using data from Cryo-EM and many other experimental sources to get a structure (ie data fusion problem)

- Understanding cell-level organization makes use of microscopy images. The challenge is to get comprehensive info from a small set of images, which usually requires computational methods to analyze the images.
- Cell-level simulations also exist!

## Recurrent Themes

- Similarities and differences of techniques used at different spatial scales
- Physics-based approaches (based on principles of physics) vs data-driven approaches (based on experimental data and statistical inference). There's a continuum between these two extremes
- Computation can be used in two ways: structural interpretation of experimental data (CryoEM going from data to information) and structural predictions in the absence of such data
- Energy functions: looking at the structure of a protein and associating an energy. Structure of protein is generally not the minimum energy, but having energy function is very useful for different computational tasks
- Use of fourier transforms and convolutions (these mathematical concepts will be taught at an intuitional level)