Project ideas

CS/BioE/CME/Biophys/BMI 279
Nov. 9, 2017
Ron Dror

Project Ideas

Project Ideas

- You are welcome (and encouraged) to pick something that is not on this list.

Project Ideas

- You are welcome (and encouraged) to pick something that is not on this list.
  - It should involve structure or spatial organization at a molecular or cellular level

Project Ideas

- You are welcome (and encouraged) to pick something that is not on this list.
  - It should involve structure or spatial organization at a molecular or cellular level
- It can be methods-focused or application-focused

Project Ideas

- You are welcome (and encouraged) to pick something that is not on this list.
  - It should involve structure or spatial organization at a molecular or cellular level
- It can be methods-focused or application-focused
  - I.e., you can code/modify software, or apply existing software to a biological problem

Project Ideas

- You are welcome (and encouraged) to pick something that is not on this list.
  - It should involve structure or spatial organization at a molecular or cellular level
- It can be methods-focused or application-focused
  - I.e., you can code/modify software, or apply existing software to a biological problem
- See project assignment sheet on website for details on project writeup and other information

Project Ideas

- You are welcome (and encouraged) to pick something that is not on this list.
  - It should involve structure or spatial organization at a molecular or cellular level
- It can be methods-focused or application-focused
  - I.e., you can code/modify software, or apply existing software to a biological problem
- See project assignment sheet on website for details on project writeup and other information
  - Joint projects, overlaps with projects for other classes, etc.

Protein structure prediction

- Sample codes:
  - Rosetta/PyRosetta (or Robetta webserver)
  - Phyre2 (web server)
  - I-Tasser (web server or download code)
  - Modeller (web server called ModWeb, or download code)
  - SWISS-MODEL (web server)
- Topics of interest include
  - Structure prediction methodology
  - Structures of proteins of interest
  - Effects of protein modification (e.g., mutation, phosphorylation)
- Related: RNA/DNA structure prediction

These web servers let you plug in your sequence and get a structure. But this shouldn't be your whole project! But you could use this tool to compare between structures, for example.

Ligand Docking

- Established, free codes and web servers:
  - Autodock Vina
  - SwissDock
- Rosetta Dock (newer; can use in PyRosetta framework)
- Note: But Stanford recently got a Stanford-wide license for GLIDE.
  - GLIDE and GOLD are powerful but not free.
  - UCSF DOCK is free and powerful, but is reputedly difficult to install and learn.

- ZINC ligand database: http://zinc.docking.org/

  A good database of millions of drug-like (i.e. reasonable-looking) ligands.

- Related: Protein–protein docking (e.g., with ZDock, Haddock) or protein–peptide docking (e.g., with FlexPepDock or Backrub servers)

Molecular dynamics simulation

- Write your own code
- Don't reuse BMI/BIOE/GENE 214/CS 274 code, but can build on it
  - Increase speed (fast electrostatics methods, parallelization), improve integrators, add restraints/constraints
- Or use existing software

GROMACS or the graphical user interface of Desmond are two good ways to get started.

  - GROMACS, Desmond, NAMD, AMBER (PMEMD module): designed for performance. GROMACS, Desmond, and NAMD are free (for academic use); AMBER is not
  - Tinker—slow, but designed to be easy to work with the code
  - Most of these are designed for Linux, but Windows and Mac executables are available for Tinker
- Either focus on simulation of a particular protein, or on methods (e.g., molecular dynamics vs. Monte Carlo)

Protein Design

Probably more difficult to get started with, but a very interesting area of current research.

- Rosetta software is free for academic use
- Rosetta Design server: http://rosettadesign.med.unc.edu/

Image analysis

- Useful software:
  - Matlab (general-purpose)
  - ImageJ (free, widely used for biological image processing)
  - CellProfiler (free, includes machine learning applications)
- Or write your own software (e.g., for segmentation, edge detection, cell counting)
  - Some students have implemented an FFT. Ideally you would also apply your implementation to images.
- Sample image sets:
  - https://data.broadinstitute.org/bbbc/
  - http://www.cellprofiler.org/
  - Please let me know of other good ones you find!

Reaction-diffusion simulation

- Use existing codes:
  - MCell, Smoldyn, Simmune
  - For MCell, consider using CellOrganizer or CellBlender to make models or renderings
- Write your own code
- Build a model of a cellular process, or consider methodological issues

Single-particle electron microscopy

- Software packages:
  - XMIPP: has a graphical user interface, somewhat easier to use
  - Relion: more mathematically sophisticated (Bayesian methods)
- Most use MPI, which complicates installation.
- Alternative: implement something yourself
  - Work in two dimensions for simplicity
  - Or tackle early stages in single-particle EM pipeline, such as particle picking

Crystallography

- Structure factors (i.e., primary crystallographic data) are often available in PDB.
  - See http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/structurefactors.html
- CNS software (http://cns-online.org/v1.3/)

Other topics

- CellPack (http://www.autopack.org/): packing proteins into a cell
- EVFold (http://evfold.org/): protein structure prediction based on covariation across sequences

  Basically, look for pairs of amino acid positions that have correlated evolution across different species. If two amino acids tend to change together, they might be in direct contact in the folded protein.
- Secondary structure prediction from sequence (e.g., using a neural network)