CSE 350/450: Structural Bioinformatics

Location: Packard Lab 416, Tuesdays and Thursdays, 1:10 pm - 2:25 pm
Professor: Brian Y. Chen, Department of Computer Science and Engineering, Lehigh University

1. Course Description

Solving problems at the leading edge of medical and industrial technologies depends, in many cases, on improving our understanding of protein function. For example, the debilitating side effects of cancer treatment could be reduced by developing drug molecules that selectively fit the unique structures of cancer proteins. In this case, and in many others, protein shape can yield many deep insights into how proteins function. This course is an exploration, through collaborative and interdisciplinary projects, of the biological, computational, and statistical ideas developed for protein structure alignment, finding functional sites, structure-function inference, molecular surfaces, and evolution in protein structure.

2. Textbook

Structural Bioinformatics, 2nd Edition
Edited by Jenny Gu, and Philip E. Bourne.
Publisher: Wiley-Blackwell, 2009.

3. Lecture Topics

Week 1: Introduction
 Lecture 1: Introduction to Structural Bioinformatics
 Lecture 2: Introduction to Proteins

Week 2: Motif Design and Geometric Matching
 Lecture 3: Motif Design
 Lecture 4: Geometric Matching

Week 3: Modeling Matching Catalytic Sites in Protein Structure Data
 Lecture 5: Match Scoring

Week 4: Algorithmic Motif Design
 Lecture 6: Algorithmic Motif Design
 Lecture 7: Composite motifs

Week 5: Protein Structure Alignment
 Lecture 8: Whole Structure Alignment
 Lecture 9: Geometric Hashing and GRATH

Week 6: Multiple Structure Alignment
 Lecture 10: Multiple Structure Alignment
 Lecture 11: Applications of Multiple Structure Alignment

Week 7: Protein Electrostatics
 Lecture 12: Protein Electrostatics

Week 8: Analyzing Molecular Surfaces
    Lecture 13: Alpha Shapes and Protein Cavities
    Lecture 14: Molecular Surfaces and Surface Analysis

Week 9: Analyzing Molecular Volumes
    Lecture 15: Volumetric Analysis of Protein Surfaces
    Lecture 16: Using Machine Learning to classify functional sites

Week 10: Protein-Protein Interactions
    Lecture 17: Protein-Protein Interactions
    Lecture 18: Predicting Protein Protein Interactions and Hotspots

Week 11: Protein-DNA Interactions
    Lecture 19: Protein-DNA Interactions
    Lecture 20: Predicting Protein-DNA Interactions

Week 12: Molecular Simulation and Docking
    Lecture 23: Molecular Simulation

Week 13: Computational Drug Design
    Lecture 21: Structure Based Drug Design

Week 14: Quantative Structure-Activity Relationships
    Lecture 22: From Hits to leads

Week 15: Protein Structure Prediction
    Lecture 24: Knowledge-based Protein Structure Prediction
    Lecture 25: Ab Initio Protein Structure Prediction

4. Course Structure

Semester Project (CSE350: 90%) (CSE450: 70%)

The semester project will be based on a prototype bioinformatics pipeline with biological, algorithmic and statistical modules. For their semester projects, students, working alone or in interdisciplinary groups, will retrofit the prototype with one or more modules. See the project wiki for details.

Class Participation (CSE350/450: 10%)

a) **Actively starting and/or participating in discussions** during the lecture periods that connect the topic of the lecture to a major field in the course can significantly improve your participation grade. Example: Suppose the lecture is covering protein structure alignment. Questions like "is protein structure alignment used in current biological research?", "It seems like there are an infinite number of possible structural alignments, how is an optimal alignment even possible?", "Is an optimal structural alignment even a biologically valid piece of information?" are discussion-starters that will be noted for class participation. Participating in these discussions will, likewise, be noted. There is no distinction in score between starting and participating in a discussion. Do not be concerned about interrupting the lecture, until the lecturer says it is time to move on from the discussion. For full credit, students are expected to ask one question or participate in one discussion per lecture.

b) **Attendance.** In a collaborative course like CSE 350, participation is key for both your own benefit and that of those around you. If you do not attend the course, then you deny yourself access to the knowledge of those around you, and you deny those around you a chance to learn from your experience. For this reason, attendance is very important. This is an advanced course, and, for efficiency, attendance will not be taken. If the lecturer notices that you are not in class, you will lose participation points. Obtain approval from the lecturer in advance for planned absences, which will not count against you.

c) **Cell phones and laptops** are to be set on silent mode. Sounds from any electronic device, including the vibration of a device against hard surfaces, are considered interruptions and will thus affect participation grades. However, lectures will pause for questions to be answered, and those questions may turn into discussions. Typing during lecture is considered an interruption, but during discussions, it is permissible, presumably because you are looking up information to add to the discussion and boost your participation grade. Tactful use of electronic devices respects the interests of your peers.

Review Paper (CSE450: 20%)
You will write a review of the current research relating to one of the topics covered in class. Inform the instructor of what topic you have chosen to do. See further descriptions below.

5. Prerequisites

This course draws from three primary subjects: biochemistry, algorithms, and statistics. Knowledge of all subjects is unnecessary, but a familiarity with at least one of these subjects is required. Having taken courses in one of the following groups is highly recommended, though only instructor permission is necessary to enroll. Students will be asked to state which set(s) of prerequisites they fulfill, and be thus associated with one of the primary subjects.

Group 1: BIOS 371+372 or CHM 371+372
Group 2: MATH 205, CSE 109, CSE 340
Group 3: MATH 312, MATH 334.

A background in programming is not required for the biological module and only light programming is necessary for the statistical module. Thus, programming is not required to complete the course with full marks. Likewise, a background in biology is not required for the computational or statistical modules, and a background in statistics is not required for the biological and computational modules. However, integrating the knowledge gained about other fields, from other students, the lectures, and the textbook, can provide significant extra credit.

6. Assessment

For the project, full credit is possible for completing, alone, a module of the pipeline (see course structure) in the student's declared background. Students working alone on a module outside of their declared background are eligible for up to 15% extra credit, to recognize the difficulty of working on material outside the student's background.

Groups can have at most three students, and must complete one module for each student in the group. Modules completed by a group of students cannot be of the same type - i.e. two biological modules for a group of two students is disallowed. Students integrating multiple modules as part of a group will be eligible for up to 5% extra credit for each integrated module, and thus a maximum of 15%. The instructor will divide dysfunctional groups into individuals, with individual modules, if necessary.

Additional Requirements for CSE450

Students taking CSE450 must prepare a 7-10 page (single spaced, 12 point font, 1 inch margins, including citations.) review paper on the current research relating to one of the fields discussed in class. CSE450 Reports are individual work, and not to be discussed with other students.

7. Outcomes
By completing this course, students will:
1. Understand the basic design and purpose of several major computational technologies in the field of structural bioinformatics
2. Be aware of how biological, algorithmic, and statistical concepts can be integrated to draw meaningful conclusions from multi-faceted biological data.
3. Have experience in the implementation challenges relating to these major technologies.
4. Have experience in technical communication with collaborators with technical expertise outside of their own field.

This course supports program missions to educate students that will:
1. Apply their education in computer science to the analysis and solution of scientific, business, and industrial problems.
3. Function effectively in a collaborative team and effectively communicate with members of the team.
4. Engage in continued education in their field of expertise

Students with Disabilities:
If you have a documented learning disability, and will be requesting academic accommodation for this class, please contact Dean Cheryl Ashcroft in the Office of the Dean of Students, UC 212, at x84152, or by email at caa4@lehigh.edu. She will establish the appropriate accommodations for your case.

Lehigh Student Senate Academic Integrity Statement:
We, the Lehigh University Student Senate, as the standing representative body of all undergraduates, reaffirm the duty and obligation of students to meet and uphold the highest principles and values of personal, moral and ethical conduct. As partners in our educational community, both students and faculty share the responsibility for promoting and helping ensure an environment of academic integrity. As such, each student is expected to complete all academic course work in accordance to the standards set forth by the faculty and in compliance with the university's Code of Conduct.