

# CSE 307/407: Structural Bioinformatics

Location: TBD, Mondays, Wednesdays, Fridays, 2:10 pm – 3:00 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 and Readings

### Introduction

Lecture 1: Introduction to Structural Bioinformatics

Lecture 2: Introduction to Proteins

[Gu and Bourne](#), Chapter 1, 2, 4. Optional: Chapter 3.

### Volumetric Comparison and Statistical Modeling

Lecture 3: Volumetric Analysis of the Surfaces of Proteins

Lecture 4: Statistical Analysis of Volumetric Variation

[VASP: A Volumetric Analysis of Surface Properties Yields Insights into Protein-Ligand Binding Specificity file](#) [34]

### Molecular Simulation and Data Set Construction

Lecture 5: Molecular Dynamics Simulation

Lecture 6: The PDB and Data Set Construction

[Gu and Bourne](#), Chapter 27, 37

[Molecular Dynamics Simulations in Biology](#) [46]

[Energetics of Ion Conduction through the K<sup>+</sup> channel](#) [47]

### Geometric Matching and Match Scoring

Lecture 7: Geometric Matching

Lecture 8: Match Scoring

[Gu and Bourne](#), Chapter 10, 11, 21.

[Inferring Functional Relationships of Proteins from Local Sequence and Spatial Surface Patterns](#) [5].

[Algorithms for structural comparison and statistical analysis of 3D protein motifs](#) [1].

[A Model for Statistical Significance of Local Similarities in Structure](#) [4].

[A Statistical Model to Correct Systematic Bias Introduced by Algorithmic Thresholds in Protein Structural Comparison Algorithms](#) [2].

[An algorithm for constraint-based structural template matching: application to 3D templates with statistical analysis](#) [3].

### **Motif Refinement**

[Lecture 9: Algorithmic Motif Design](#)

[Lecture 10: Composite Motifs](#)

[Gu and Bourne](#), Chapter 10, 11, 21.

[The MASH pipeline for protein function prediction and an algorithm for the geometric refinement of 3D motifs](#) [7].

[Cavity-aware motifs reduce false positives in protein function prediction](#) [8].

[Composite motifs integrating multiple protein structures increase sensitivity for function prediction](#) [9].

[Analysis of substructural variation in families of enzymatic proteins with applications to protein function prediction](#) [10]

### **Protein Structure Alignment**

[Lecture 8: Whole Structure Alignment](#)

[Lecture 9: Geometric Hashing and GRATH](#)

[Gu and Bourne](#), Chapters 16, 17, 18.

[Mapping the Protein Universe](#) [11]

[Efficient Alignment of 3d structures with Geometric Hashing](#) [12]

[Protein structure comparison using iterated double dynamic programming](#) [13]

[Protein structure alignment by incremental combinatorial extension \(CE\) of the optimal path](#) [14]

[Quantifying the Similarities in Fold Space](#) [15]

### **Multiple Structure Alignment**

[Lecture 10: Multiple Structure Alignment](#)

[Lecture 11: Applications of Multiple Structure Alignment](#)

[Gu and Bourne](#), Chapter 23.

[Multiple Structural Alignment by Optimal RMSD Implies that the Average Structure is a Consensus](#) [16]

[Automated Multiple Structure Alignment and Detection of a Common Substructural Motif](#) [17]

[A Method for Simultaneous Alignment of Multiple Protein Structures](#) [18]

[MC-CE: Monte Carlo Combinatorial Extension](#) [19]

[Structural Evolution of the Protein Kinase-like Superfamily](#) [20]

[Multiple Flexible Structural Alignment with POSA](#) [21]

### **Analyzing Molecular Surfaces**

[Lecture 13: Alpha Shapes and Protein Cavities](#)

[Lecture 14: Molecular Surfaces and Surface Analysis](#)

[Analytical Shape Computation of Macromolecules I: alpha shapes](#) [26]

[Analytical Shape Computation of Macromolecules II: protein cavities](#) [27]

[Are Proteins Well packed?](#) [28]

[SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions](#) [29]

[A method for localizing ligand binding pockets in protein structures.](#) [30]

[Protein Clefts in Molecular Recognition and function](#) [31]

[Travel Depth: A new shape descriptor for macromolecules](#) [32]

[Identifying protein binding pockets with PASS](#) [33]

## Protein-Protein Interactions

[Lecture 17: Protein-Protein Interactions](#)

[Lecture 18: Predicting Protein Protein Interactions and Hotspots](#)

[Gu and Bourne](#), Chapter 26

[The Atomic Structure of Protein-Protein Recognition Sites](#) [37]

[Protein-Protein Interactions: Hot Spots and Structurally Conserved Residues Often Locate in Structurally Complemented Pockets that Pre-Organized in the Unbound State: Implications for Docking](#) [38]

[Protein interface conservation across structure space](#) [39]

[Segmenting Motifs in Protein-Protein Interface Surfaces](#) [40]

## Protein-DNA Interactions

[Lecture 19: Protein-DNA Interactions](#)

[Lecture 20: Predicting Protein-DNA Interactions](#)

[Gu and Bourne](#), Chapter 25.

[Recognition of Specific DNA sequences](#) [41]

[The Role of DNA shape in protein-DNA recognition](#) [42]

[Structure-based Prediction of C2H2 Zinc-Finger binding specificity: Sensitivity to docking geometry](#) [43]

[Structural alignment of protein-DNA interfaces: insights into the determinants of binding specificity](#) [44]

[Exploring the DNA-binding specificities of zinc fingers with DNA microarrays](#) [45]

## Computational Drug Design

[Lecture 21: Structure Based Drug Design](#)

[Gu and Bourne](#), Chapter 34

[Evaluating the Substrate-Envelope Hypothesis: Structural analysis of Novel HIV-1 Protease Inhibitors Designed to be Robust against Drug Resistance](#) [48]

## Protein Structure Prediction

[Lecture 24: Knowledge-based Protein Structure Prediction](#)

[Lecture 25: Ab Initio Protein Structure Prediction](#)

[Gu and Bourne](#), Chapters 28-32

[Comparative Protein Modelling by Satisfaction of Spatial Constraints](#) [54]

[On the Role of Structural Information in Remote Homology Detection and Sequence Alignment: New Methods Using Hybrid Sequence Profiles](#) [55]

[Solving and analyzing side-chain positioning problems using linear and integer programming](#) [56]

[Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences using Simulated Annealing and Bayesian Scoring Functions](#) [57]

[Improved Recognition of Native-Like Protein Structures Using a Combination of Sequence-Dependent and Sequence-Independent Features of Proteins](#) [58]

## Quantative Structure-Activity Relationships

[Lecture 22: From Hits to leads](#)

[Lecture 16: Using Machine Learning to classify functional sites](#)

[Hotpatch: A Statistical Approach to Finding Biologically Relevant Features on Protein Surfaces](#) [35]

[SCREEN: Finding Druggable Cavities with Random Forests](#) [36]

[The role of quantitative structure-activity relationships \(QSAR\) in biomolecular drug discovery](#) [49]

[Hit and Lead generation: beyond high throughput screening](#) [50]

[Rationalizing fragment based drug discovery for BACE1: insights from FB-QSAR, FB-QSSR, multi objective \(MO-QSPR\) and MIF studies](#) [51]

[Virtual screening for R-groups, including predicted pIC50 contributions, within large structural databases, using Topomer CoMFA](#) [52]

[Rethinking 3D-QSAR](#) [53]

## Protein Electrostatics

[Lecture 12: Protein Electrostatics](#)

[Gu and Bourne](#), Chapter 24.

[On the Role of Electrostatic Interactions in the Design of Protein-Protein Interactions](#) [22]

[Focusing of Electric Fields in the Active Site of Superoxide Dismutase](#) [23]

[On the impact of desolvation versus electrostatic attraction](#) [24]

[Rapid Electrostatically assisted association of proteins](#) [25]

## 4. Course Structure and Assessment

	<b>CSE397</b>	<b>CSE497</b>
Semester Project	75%	60%
CSE497 Report	0%	20%
Participation	25%	20%

### Semester Project (CSE397: 75%) (CSE497: 60%)

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 description for details.

### Class Participation (CSE397: 25%)(CSE497: 20%)

a) **Asking questions in lecture.** The participation score, evaluated at the end of the semester, is equal to the fraction of lectures in which the student has asked a question, except for the first lecture, which does not count. 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. Never be concerned about interrupting the lecture, until the lecturer says it is time to move on from the discussion – always ask the question immediately.

b) **Cell phones and laptops** are to be set on silent mode. Tactful use of electronic devices is permitted and respects the interests of your peers.

### Review Paper (CSE497: 20%)

Students taking CSE497 must prepare a 7-10 page 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. Formatting instructions: One page should be used for the title and your name. This title page does not count towards the page length, but bibliography pages do count towards page length. **Only full pages count towards page length.** Pages must be single spaced, and the first page must start at the top of the page. Use 12 point times fonts, 1 inch margins, and numbered bibliography references, on letter sized paper. Submit your document via email as a pdf.

## 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.

Alternative: permission of the instructor

## 6. 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

**Accommodations for Students with Disabilities:** If you have a disability for which you are or may be requesting accommodations, please contact both your instructor and the Office of Academic Support Services, University Center C212 (610-758-4152) as early as possible in the semester. You must have documentation from the Academic Support Services office before accommodations can be granted.

**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.

All work, unless explicitly stated in the problem definition, is to be an individual effort. You are encouraged to discuss assignments with one another, your friends, and with the instructors and graders of the course. Indeed, this may be the most effective method of learning. You may share concepts, approaches and strategies for producing a solution, **BUT YOU MAY NOT SHARE CODE UNDER ANY CIRCUMSTANCES**. All work submitted in your name must be your own. If necessary, violations will be considered as cases of academic dishonesty.