

Bioinformatics Syllabus

Course Description:

BINF 701/702 is the bioinformatics core course developed at the KU Center for Bioinformatics. The course is designed to introduce the most important and basic concepts, methods, and tools used in Bioinformatics. Topics include (but not limited to) bioinformatics databases, sequence and structure alignment, protein structure prediction, protein folding, protein-protein interaction, Monte Carlo simulation, and molecular dynamics. Emphasis will be put on the understanding and utilization of these concepts and algorithms. The objective is to help the students to reach rapidly the frontier of bioinformatics and be able to use the bioinformatics tools to solve the problems on their own research.

Instructors:

- Wonpil Im, Email: wonpil@ku.edu, Phone: 785-864-1993
- Ilya Vakser, Email: vakser@ku.edu, Phone: 785-864-1057
- Yang Zhang, Email: yzhang@ku.edu, Phone: 785-864-1948

Schedule:

2:00-2:50pm, MTWTF

Homework & Exams:

There will be several homework assignments that will constitute 50% of your grade. The final examination constitutes another 50% of your grade.

Student evaluation and grades:

The grading scale will be:

A = 90-100%

B = 80-89

C = 70-79

D = 60-69

F = below 60

Textbook:

A textbook is not required for this course. Assigned materials will either be handed out to the class or posted on the blackboard website.

Course policies:

Attendance and class participation are expected. Repeated absences without approval or valid justification will result in the reduction of the grade for the course.

Late assignments will be docked 10% for being late, and an additional 10% for each additional day they are late.

Cheating and Plagiarism will be considered academic misconduct and therefore subject to the University Senate Rules and Regulations Section 2.6.1-7. Briefly, if a student represents someone else's work for their own (either by cheating or plagiarism) in any classroom examination or assignment without making the proper acknowledgement of that work, it will be deemed academic misconduct. All work, either in class or on an assignment, is expected to be your own work. Homework: the penalty for the first offense will be a reduction in grade to a zero (F) for that specific work. The penalty for a second offense will be a reduction in grade assignment to F for the course. Exams: the penalty for the first offense will be a reduction in grade to F for the course. For more information on academic misconduct, please see

<http://www.clas.ku.edu/faculty/policies/misconduct.pdf>.

Table of content

PART I: BIOINFORMATICS BASICS

- 1. Introduction (Ilya Vakser)**
 - 1.1. What is bioinformatics
 - 1.2. Basic concepts
 - 1.2.1. Protein and amino acid
 - 1.2.2. DNA & RNA
 - 1.2.3. Sequence, structure and function
- 2. Bioinformatics databases (Yang Zhang)**
 - 2.1. Introduction
 - 2.1.1. Motivation
 - 2.1.2. Type of databases
 - 2.2. Nucleotide sequence databases
 - 2.2.1. Primary nucleotide sequence databases
 - 2.2.1.1.EMBL
 - 2.2.1.2.GeneBank
 - 2.2.1.3.DDBJ
 - 2.2.2. Secondary nucleotide sequence databases
 - 2.2.2.1.UniGene
 - 2.2.2.2.SGD
 - 2.2.2.3.EMI Genomes
 - 2.2.2.4.Genome Biology
 - 2.3. Protein sequence databases
 - 2.3.1. SwissProt/TrEMBL
 - 2.3.2. PIR
 - 2.4. Sequence motif databases
 - 2.4.1. Pfam
 - 2.4.2. PROSITE
 - 2.5. Protein structure databases
 - 2.5.1. Protein Data Bank
 - 2.5.2. SCOP
 - 2.5.3. CATH
 - 2.6. Other relevant databases
 - 2.6.1. KEGG
 - 2.6.2. PQS
 - 2.6.3. DockGround
- 3. Sequence alignment and database searching (Yang Zhang)**
 - 3.1. Single sequence alignments
 - 3.1.1. Biological motivation
 - 3.1.2. Pairwise alignments
 - 3.1.2.1.Scoring matrix
 - 3.1.2.1.1. PAM
 - 3.1.2.1.2. BLOSUM

- 3.1.2.2. Gap penalty
- 3.1.3. Dynamics programming
 - 3.1.3.1. Needleman-Wunsch
 - 3.1.3.2. Smith-Waterman
- 3.1.4. Heuristic methods
 - 3.1.4.1. FASTA
 - 3.1.4.2. BLAST
- 3.1.5. Statistics of sequence alignment score
 - 3.1.5.1. E-Value
 - 3.1.5.2. P-Value
- 3.2. Multiple sequence alignments
 - 3.2.1. ClustalW
 - 3.2.2. Profile
 - 3.2.2.1. Profile-sequence alignment
 - 3.2.2.2. Profile-profile alignment
 - 3.2.3. PSI-BLAST
 - 3.2.4. Hidden Markov Models
 - 3.2.4.1. Viterbi algorithm
 - 3.2.4.2. HMM based multiple-sequence alignment
 - 3.2.4.3. SAM

4. Protein structure alignments (Yang Zhang)

- 4.1. What is structure superposition?
 - 4.1.1. RMSD
 - 4.1.2. TM-score
- 4.2. What is structure alignment?
- 4.3. Different structure alignment algorithms
 - 4.3.1. DALI
 - 4.3.2. CE
 - 4.3.3. VAST
 - 4.3.4. TM-align
- 4.4. Number of protein folds in PDB

PART II: PROTEIN STRUCTURE

5. Protein secondary structure predictions (Yang Zhang)

- 5.1. What is protein secondary structure?
- 5.2. Hydrogen bond
- 5.3. How to define a secondary structure element?
- 5.4. Methods for predicting secondary structure
 - 5.4.1. Chou and Fasman method
 - 5.4.2. PHD
 - 5.4.3. PSIPRED
 - 5.4.4. SAM

6. Protein tertiary structure modeling (Yang Zhang)

- 6.1. Basic concepts
- 6.2. Protein folding and dynamic simulation

- 6.3. Comparative modeling
 - 6.3.1. Modeller
 - 6.3.2. Swiss-Modeller
- 6.4. Threading
 - 6.4.1. What is threading?
 - 6.4.2. Bowie-Luthy-Eisenberg
 - 6.4.3. Profile-profile alignment
 - 6.4.4. GenThreader
 - 6.4.5. PROSPECTOR
 - 6.4.6. FFAS03
 - 6.4.7. Meta-threading
 - 6.4.7.1.1. 3D-jury
 - 6.4.7.1.2. LOMETS
- 6.5. Ab initio modeling
 - 6.5.1. Anfinsen thermodynamic hypothesis
 - 6.5.2. UNRES
 - 6.5.3. ROSETTA
 - 6.5.4. TOUCHSTONE
- 6.6. Combined modeling approaches
 - 6.6.1. TASSER/I-TASSER
- 6.7. CASP: A blind protein structure prediction competition
- 7. Experimental methods for protein structure determination (Yang Zhang)**
 - 7.1. X-ray crystallography
 - 7.1.1. Diffraction theory
 - 7.1.2. Phase determination
 - 7.1.3. Calculating and interpreting electron density maps
 - 7.1.4. Model building and refinement
 - 7.1.5. Structure assessment
 - 7.1.6. Crystallization of macromolecules
 - 7.1.7. Dynamic crystallography
 - 7.2. Nuclear magnetic resonance (NMR)
 - 7.2.1. Classical NMR spectroscopy
 - 7.2.2. Theoretical description of NMR spectroscopy
 - 7.2.3. Experimental aspects of NMR spectroscopy
 - 7.2.4. Relaxation and dynamic processes
 - 7.2.5. Heteronuclear NMR experiments
 - 7.2.6. Sequential assignment and structure calculations

PART III: PROTEIN-PROTEIN INTERACTIONS

- 8. Experimental identification of protein-protein interactions (Ilya Vakser)**
 - 8.1.** Yeast two-hybrid assay
 - 8.2.** High-throughput mass spectrometry
 - 8.3.** Interaction networks and system biology
- 9. Protein quaternary structure modeling (Ilya Vakser)**
 - 9.1. Basic concepts

- 9.1.1. Degrees of freedom
- 9.1.2. Presentation of protein conformations
- 9.1.3. Hydrophobicity factor
- 9.1.4. Shape complementary
- 9.1.5. Docking Scoring function
- 9.2. Protein-protein docking algorithms
 - 9.2.1. Fast Fourier Transformation (FFT)
 - 9.2.1.1. GRAMM
 - 9.2.2. Semi-flexible docking: Side-chain refinement
 - 9.2.3. Clustering and refinement
- 9.3. Protein-ligand docking algorithms
- 9.4. Drug design
- 9.5. Multiple-threading algorithms
- 9.6. Homology modeling of protein-protein interactions
- 9.7. Protein and ligand binding
- 9.8. CAPRI

PART IV: BIOMOLECULAR SIMULATIONS

10. Basic concepts (Wonpil Im)

- 10.1. Units and derivatives
- 10.2. Force field and energy landscape
- 10.3. Truncation of nonbonded interactions

11. Conformational Sampling (Wonpil Im)

- 11.1. Introduction
- 11.2. Minimization and algorithms
- 11.3. Molecular dynamics
- 11.4. Ensembles (statistical mechanics)
- 11.5. Monte Carlo simulations

12. Solvation (Wonpil Im)

- 12.1. Introduction
- 12.2. Periodic boundary condition
- 12.3. Ewald summation
- 12.4. Implicit solvent model and continuum electrostatics
- 12.5. Monte Carlo simulation on parallel computers**

13. Advanced Techniques (Wonpil Im)

- 13.1. Introduction
- 13.2. Replica-exchange simulations
- 13.3. Restraint potentials
- 13.4. Free energy calculations
- 13.5. Membrane simulations

PART V: SELECTED TOPICS

14. Biological membranes (Wonpil Im)

- 14.1. Introductions

- 14.1.1. Biological roles
- 14.1.2. Structural features
- 14.2. Membrane lipids
 - 14.2.1. General structures
 - 14.2.2. Aggregation states
 - 14.2.3. Polymorphism
 - 14.2.4. Thermal transitions
 - 14.2.5. Electrostatic effects
 - 14.2.6. Molecular dynamics
- 14.3. Membrane proteins
 - 14.3.1. Crystallization
 - 14.3.2. Overview of structure features
 - 14.3.3. Structure/function relations
 - 14.3.4. Selected topics in membrane proteins
- 14.4. MD simulation of Membrane proteins
- 15. **Protein function (Ilya Vakser)**
 - 15.1. Sequence to function
 - 15.2. Structure to function
 - 15.3. Protein function identification methods and databases
- 16. **Phylogenetics (Yang Zhang)**
 - 16.1. Sequence-based taxonomy
 - 16.1.1. Why Phylogenetics?
 - 16.1.2. Models, assumptions, and interpretations
 - 16.2. From multiple alignment to phylogeny
 - 16.2.1. Neighbor joining
 - 16.2.2. Maximum likelihood and parsimony
 - 16.3. Computer tools for phylogenetic analysis
 - 16.3.1. DISTANCES
 - 16.3.2. GROWTREE
 - 16.3.3. PAUP
 - 16.3.4. PHYLIP
- 17. **Metabolism and networks (Yang Zhang)**