

Fundamentals of Bioinformatics 2011 - Learning Objectives

Lecture 1: General: Modeling, optimality; Evolution: mutations, selection, sex orthologs

- The role of information → Bioinformatics
- The importance and role of a model
- The cell: global properties
 - Prokaryotes (bacteria and archaea) and eukaryotes
 - Phylogeny
- Speciation and cell differentiation
- Genomics

Lecture 2: Genomics (Genes, genomics data, codons, translation/ transcription, microarray & proteomics data, splicing, PSI-blast)

- DNA as an information carrying molecule
- DNA makes RNA makes protein
 - The molecular machinery as an information device
 - Information on DNA around (upstream and downstream) the protein coding region (all together called 'gene')
 - How much can DNA itself influence the encoding processes?
- What informational aspects are playing a role in these processes?

Lecture 3: Genomics (Sequence alignment / Bioinformatics for Systems Biology, Smith/Waterman (DP), sequence DBs)

- Genome size and number of genes
- Evolution, speciation, convergent vs. divergent evolution
- homology, orthology, paralogy
- Pairwise alignment: global, semi global, local
- Multiple sequence alignment (MSA)
- Blast, PSI-Blast
- PFAM database

Lecture 4: Ontologies & GO

- Biomedical data is complex and widely distributed
- Currently, data is difficult to integrate, but strategies exist:
 - Exchange formats, controlled vocabularies
- The Semantic Web offers new approaches to publishing and integrating data
 - An automated 'connection' between Bio(medical) Information and Bio(medical) Understanding

Lecture 5: Protein Bioinformatics: aa's, ss, folds, domains, disorder, TM, PDB

- Proteins Structure
- Structure → Function
- Structure more conserved than sequence
- PDB
- Structure alignment
- Structural Classification
 - SCOP/CATH
 - Domains
 - Convergence vs divergence

- Structure prediction (Part 5, 6)
 - Preview: homology vs 'ab initio'

Lecture 6: Protein Bioinformatics: protein interactions & thermodynamics

- Structure Prediction
- Simulation
 - Protein Folding
 - Solvation – Hydrophobic Effect
- Potential energy vs. Free energy (and Entropy)
 - How hard it is to get somewhere, vs. how likely it is.
- Protein-protein interactions
 - Hydrophobic Effect
 - Potentials of Mean Force

Lecture 7: Petri-net models

- Models:
 - All models are wrong, but some are useful.
 - A model can be useful if it brings understanding.
- Petri nets are graphs:
 - graphical (and formal) language to describe models.
 - Capture and simulate biological knowledge.
 - In → Interactions and logic.
 - Out ← state transitions ↔ state graph.
- State graph can be analysed.

Lecture 8: Repeat detection and/or Domain prediction

- Background: what is a domain?
- Predicting domains from structure
- Predicting domains from sequence alone
 - Linker prediction
 - Protein domain delineation based on:
 - consistency of multiple ab initio model tertiary structures (SnapDRAGON; Rosetta)
 - combining homology searching with domain prediction (Domaination)
 - sequence hydrophobicity patterns (SCOOBY-DOmain)

Lecture 9: Secondary Structure prediction

- Homology Modeling?
 - Why are helices and strands formed
 - Secondary structure features and constraints
- Secondary Structure Prediction
 - Inverse folding: Threading
 - Which sequence fits this structure? → Fold recognition
 - Assigning SS from tertiary structure → DSSP
 - Predicting SS from sequence
 - Early, Modern (MSA and machine learning), Consensus
 - Standard of Truth and Assessment (e.g. measures, jackknife and Cross Validation)

Lecture 10: Next Gen Sequencing

- Sanger sequencing
- NGS
- Applications
 - Reference genome
 - Methylation
 - Deep sequencing (RNA)
 - CHIP-Seq
- Algorithms
 - Alignment: Hashes
 - Assembly: De Bruijn Graphs