CMSC423: Bioinformatic Algorithms, Databases and Tools

What you missed
Protein folding
Protein folding

• Note: mis-folded proteins may cause disease (e.g. Creutzfeld-Jakob a.k.a. mad cow)
• Drugs (e.g. antibiotics) often inhibit protein function – knowing structure can help design drugs

• Folding@home – lend your computer’s unused cycles to help fold proteins (like SETI@home) (do you believe in evolution or aliens ?)
Protein structure
(primary structure = sequence)

http://www.tulane.edu/~biochem/med/second.htm
Secondary structure (motifs)

helix

sheet

turn

http://alpha2.bmc.uu.se/~kenth/bioinfo/structure/secondary/01.html
Tertiary structure (3D shape)

Phage CRO Repressor on DNA. Andrew Coulson & Roger Sayle with RasMol, University of Edinburgh, 1993

HIV Protease + Glaxo Wellcome Inhibitor
Roger Sayle with RasMol, 1995

HIV Protease + Glaxo Wellcome Inhibitor
Roger Sayle with RasMol, 1995

http://www.umass.edu/microbio/rasmol/sayle1.htm
Folded shape: lowest free energy

• Energy components
  – electrostatic ($\sim 1/D^2$) (n² terms)
  – van der Waals (n² terms)
  – hydrogen bonding (n terms)
  – “bending” (n terms)
  – solvent (water/salt) (?? terms)
  – exclusion principle (no two atoms share same volume)

• Energy minimization
  – small perturbations & computation: hill climbing, simulated annealing, etc.

• Molecular dynamics
How do we know the truth?

- **X-ray crystallography**
  - crystallize protein
  - shine X-rays
  - examine diffraction patterns

- **Nuclear Magnetic Resonance (NMR)**
  - no crystallization necessary
  - magnetic field “vibrates” hydrogen atoms
  - Nobel prize: Kurt Wuethrich

http://www.cryst.bbk.ac.uk/BBS/whatis/cryst_an.html

http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/2dnmr.htm
Simpler problems

- Secondary structure prediction
- Side-chain conformation (assuming fixed backbone)
- Protein docking (how do proteins interact)
- Database searches (protein threading)
- Simpler energy functions
- Folding on a lattice (theoretical approximation)
- Critical Assessment of Fully Automated Structure Prediction – competition on proteins with unpublished 3D structure
Proteomics
Proteomics

• Large-scale analysis of proteins
  – protein-protein interactions (e.g. yeast 2-hybrid)
  – 2D gels (mass vs. isoelectric point)
  – Mass-spectrometry
  – Protein microarrays
  – etc.
Prolteomics

- Why proteomics? Are DNA/RNA microarrays not sufficient?
- RNA abundance is not necessarily related to protein abundance
- Many proteins are modified post-translation
  - addition of additional molecules (phosphate, sugars, etc.)
  - creation of complexes (hemoglobin is actually 4 molecules)
Mass spectrometry

• Technique for measuring the mass-to-charge ratio of ions

• Basic idea
  – shoot ions into a magnetic field
  – deflection depends on mass

• Output of a mass-spectrometer
  – ions “sorted” by mass
  – for each mass bucket - number of ions with that specific mass
Mass-spectrometry

http://www.cem.msu.edu/~reusch/VirtualText/Spectrpy/MassSpec/masspec1.htm
Tandem Mass Spectrometry

- First mass-spectrometer “focuses” on a specific protein
- Second mass-spectrometer breaks the protein into smaller chunks
- Problem: given the chunks, what was the original protein?
Peptide sequencing

- Peptide - a chunk of a protein, usually obtained by enzymatic cleavage of the protein (using trypsin)

Problem: Given an MS spectrum (weights of fragments), what was the sequence of the peptide?

Or: find the peptide (of mass m) that best matches the experimental data
Biological networks
Biological networks

• Genes/proteins do not exist in isolation
• Interactions between genes or proteins can be represented as graphs
• Examples:
  – metabolic pathways
  – regulatory networks
  – protein-protein interactions (e.g. yeast 2-hybrid)
  – genetic interactions (synthetic lethality)
Metagenomics
Why do we care?

- Bacteria are everywhere in the environment
- They are not all evil
- Bacteria can be quite useful

Bio-energy
Bio-remediation
Drug development antibiotics anti-cancer
Human microbiome

- **Human** = 1 order of magnitude more bacterial cells than human cells
  - critical to infant development (immune system, GI-tract)
  - provide essential nutrients (vitamin K, B12, essential amino-acids, ...)
  - help digest complex molecules – starches, plant material
  - imbalances in normal bacterial populations correlate with disease (IBD, colon cancer, ...)

Human microbiome project
nihroadmap.nih.gov/hmp/
Some challenges on real data

- Verrucomicrobiae
- Spirochaetes
- Gammaproteobacteria
- Fusobacteria
- Flavobacteria
- Erysipelotrichi
- Epsilonproteobacteria
- Deltaproteobacteria
- Clostridia
- Betaproteobacteria
- Bacteroidetes
- Bacilli
- Alphaproteobacteria
- Actinobacteria

Labels:
- P1
- P2
- P3
- P4
Spatial genomics
  

  

• Gene expression information in a spatial context
• Combines microarray analysis with computer graphics
Figure 2  Voxelation scheme

- Mouse brain cut up into voxels
- Run a separate microarray experiment on each voxel
Figure 4  Spatial gene expression patterns for the subset of correlated genes

Figure 7 SVD delineates anatomical regions of the brain

Figure 5  Putative regulatory elements shared between groups of correlated and anticorrelated genes