

Class 16: Structure Prediction (1)

Held: Tuesday, 25 October 2011

Summary: We begin to consider computational approaches to predicting the structure of proteins.

Related Pages:

- EBoard.
- Reading: Chapter 7.
- Due: Set up second PCR on Wednesday!.

Notes:

- RIP John McCarthy, inventor of LISP.
- We're playing with the schedule a bit. The following deadlines are subject to revision.
- Due November 1: Web Exploration from Chapter 7. (Further details forthcoming.)
- Due November 3: Programming Exploration from Chapter 7. (Further details forthcoming.)
- Due November 8: Response to Krings *et al.* (Distributed Thursday.)
- Due November 10: Project Proposals. (Further details forthcoming.)
- We plan to return outstanding work next week.
- Don't forget that we have lots of lab this week and next. (In particular, Thursday and next Tuesday are going to be long classes.)
- If you have not read Chapter 7 yet, you should read it asap.
- EC for any of the Grinnell College Young Innovator for Social Justice events (Awards ceremony Tuesday night; Coffee break on Wednesday at 2:30; Symposium on Wednesday at 4:15; Symposium Wednesday at 8:00 p.m.; Morris Dees Convo).
- EC for Wednesday's noontime Biology Seminar, Stuart Allison from Knox on "Ecological restoration and environmental change - a geographic comparison of attitudes towards restoration and the practice of restoration."
- EC for Wednesday's afternoon Biology Seminar at 4:15 on "One year Master's programmes in Environmental & Life Sciences at the only 100% graduate university in England" and "'Environmental Futures' A Summer program designed for students completing their junior year at Midstates Math & Science Consortium Colleges."
- EC for Friday's Biology Seminar.
- No EC for this week's Thursday extra (Discrete Structures and Math Requirements in CS).
- Next Tuesday at noon: Study abroad options in Budapest in CS.

Overview:

- Protein structure: What is it and why care?
- The Chou-Fasman Algorithm.
- Wet Lab.

Structure Prediction

- Lots of kinds of structures:
 - Primary, secondary, tertiary, quaternary
- Proteins have structure.
- RNA has structure.
- Even DNA sometimes has non-obvious structure
- Why do we care?
- Can we write programs that predict structures?
- Not perfectly (at least not yet)
- But can get some potentially useful info

The Chou-Fasman Algorithm

- Chou and Fasman provided an algorithm (well, perhaps more of a heuristic) for predicting some secondary structure:
 - alpha helices
 - beta sheets (aka beta strands)
 - hairpin turns (aka beta turns)
- The algorithm is mostly statistical
 - We look at the likelihood that each amino acid participates in each structure
- Preparation:
 - Take proteins with known structure
 - For each amino acid, count
 - Total number of appearances
 - Appearances in alpha helices
 - Appearances in the interior of alpha helices
 - Appearances in beta sheets
 - Appearances in beta turns (positions 0, 1, 2, 3)
 - Total each column
 - For each amino acid, find the frequency
 - ...
- Basic algorithm: We pass a “sliding window” along the sequence
 - The window selects a small number of amino acids (e.g., six)
 - When some heuristic is met (e.g., at least four amino acids out of six have a high probability of participating in an alpha helix), we expand out
 - We then apply a second heuristic to the expanded sequence

Copyright © 2009-2011 Vida Praitis and Samuel A. Rebelsky. This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.