

All rights reserved. This document may be reproduced for use at one's own not-for-profit institution, with the following statement included in the reproduction: Chu-Lagraf Q. Bioinformatics III: Analysis of Human Neurological Diseases at the Genetic Level. Center for Bioengineering and Computational Science, Union College, Schenectady, NY (<http://bioengineering.union.edu>). Funding provided by the Howard Hughes Medical Institute.

Bioinformatics III: Analysis of Human Neurological Diseases at the Genetic Level

Quynh Chu-LaGraff, Ph.D.
Biology Department
Union College
Schenectady, NY 12308
chulagrq@union.edu

In consultation with:
James Hedrick
Dept. Electrical & Computer Engineering
Union College
Schenectady, NY 12308
hedrickj@union.edu

Summary of Activity

- Students should review the material in the lab modules Bioinformatics I: Introduction to Bioinformatics (http://www.union.edu/academic_depts/bioengineering/_docs/bioinfinst1.pdf) and Bioinformatics II: Introduction to Molecular Phylogenies (http://www.union.edu/academic_depts/bioengineering/_docs/phyloginstr.pdf)
The handout for this lab module can be found at http://www.union.edu/academic_depts/bioengineering/_docs/geneticdisstudrevised.pdf
- Students will use a number of different databases, such as National Center for Biotechnology Information (NCBI) Entrez, PubMed, On-line Mendelian Inheritance in Man (OMIM), Clustal W, BLAST, Swiss-Prot, and Protein Data Bank, to get information about the genetic basis of an identified inherited neurological disease.
- Students will use BLAST programs to access DNA and protein sequences to identify evolutionarily conserved homologs, orthologs, and paralogs, if applicable.
- Finally, students will identify and access known X-ray crystallographic information on an identified protein, display relevant data in a three-dimensional structure, and identify potentially important amino-acid residues within the active site.

Integration of Disciplines

At the undergraduate level, it is a challenge to integrate Biology and Computer Science at a level that is both meaningful and accessible. One way to accomplish this task is to introduce students to the basic Bioinformatics concepts by having them study an inherited human disease. By the time they are college sophomores, students should have been exposed to the fundamentals of Bioinformatics, including 1) the types and extent of molecular information that is stored in large databases; 2) the nature of the stored information and how it is transferred; and 3) some basic tools for searching, manipulating, and analyzing sequence information. (See Introduction to Bioinformatics (http://www.union.edu/academic_depts/bioengineering/_docs/bioinfinst1.pdf.)

In addition, students should be introduced to the GenBank database and learn how to navigate through different gene databases (See Bioinformatics II: Introduction to Molecular Phylogenies (http://www.union.edu/academic_depts/bioengineering/_docs/phyloginstr.pdf.)

In this laboratory module, students will be taught a few steps beyond those covered in the previous modules. The basis of a BLAST algorithm, the CLUSTALW algorithm, the Protein Data Bank, and three-dimensional rendering of proteins will be discussed. As a class, students will be introduced to all of the necessary concepts and tools by analyzing a well-characterized inherited neurological disorder, Tay-Sachs disease. To solidify these concepts and exercises, students will be divided into pairs and each pair will select a human disease from a standard list for Bioinformatics analysis using all tools learned previously. At the completion of these studies, student pairs will present their results in a PowerPoint presentation as well as a written report to be handed in.

Learning Objectives

Students will develop the following skills:

- A thorough understanding of the various databases offered by NCBI and how to access sequence information via PubMed, Entrez, OMIM, and Protein Data Bank
- How to identify the genetic basis of an identified inherited neurological disease from DNA to protein sequences and how to identify evolutionarily conserved homologs, orthologs, and paralogs, if applicable, using BLAST programs
- How to identify and access known X-ray crystallographic information on an identified protein and display relevant data in a three-dimensional structure
- The ability to examine a three-dimensional protein structure and identify potentially important amino acid residues with active site(s)
- Integrate all learned skills to investigate a different genetic disease

Target Level

This lab module is intended for use by upper-level undergraduates who have completed courses in Introductory Biology and Cell and Molecular Biology, and preferably Basic Bioinformatics laboratory modules. It is designed for students interested in Neuroscience, Computational Biology, and Biochemistry.

Tools and Materials

This exercise requires a computer lab equipped with one computer per student with Ethernet or wireless connection, a projector connected to an instructor's computer, and at least one printer accessible by all computers.

The class will view the NOVA program *Cracking the Code of Life*.

Computers are used to access the following web-based materials:

Student handout for lab module

http://www.union.edu/academic_depts/bioengineering/_docs/geneticdisstudrevise_d.pdf

National Center for Biotechnology Information (NCBI; Bethesda, MD, USA)

- Comparison of DNA or protein sequences to sequences in the databases (BLAST)
<http://www.ncbi.nlm.nih.gov/blast/>
- Human genome resources:
<http://www.ncbi.nlm.nih.gov/genome/guide/human/>

Swiss-Prot Database

- <http://www.expasy.org>

ClustalW website

- <http://www.ebi.ac.uk/clustalw/index.html>

Protein Data Bank (Note: must open in Internet Explorer, not Safari)

- <http://www.rcsb.pdb.org>

Theory and background

These exercises build upon students' previous exposure to introductory bioinformatics exercises (Bioinformatics I: Introduction to Bioinformatics (http://www.union.edu/academic_depts/bioengineering/_docs/bioinfinst1.pdf) and Bioinformatics II: Introduction to Molecular Phylogenies (http://www.union.edu/academic_depts/bioengineering/_docs/phyloginstr.pdf), and information taught in the courses Introductory Biology and Molecular Biology of the Cell. Students performing this module should be thoroughly familiar with the basic concepts of DNA, RNA, and protein (i.e., the Central Dogma of molecular biology).

Students will plunge directly into the workings of the PUBMED database and learn to navigate through OMIM and Entrez to find information on an identified human neurological disease.

Many of the ideas and instructions for this module are condensed from a 5-week (10-hour) bioinformatics tutorial course designed by faculty at Washington University (Bednarski et al., 2005). <http://www.nslc.wustl.edu/courses/Bio3055/bio3055.html>

Safety Precautions

No safety precautions.

Miscellaneous Advice to Instructors

These exercises contain a wealth of material and it may be difficult for students to complete them during one three-hour lab period. Dividing the exercises up over two or three laboratory periods will allow students more time to navigate through the different databases, investigate the background of the disease, and familiarize themselves further with the techniques involved in the exercises.

Extensions and Options

Students can also spend time outside of class investigating the biological basis of other human diseases. Students should investigate diseases for which the OMIM database contains extensive genetic, biochemical, and cellular literature and for which the Protein Data Bank contains X-ray crystallographic data for the protein involved in the disease.

References

1. Bednarski AE, Elgin SCR, Pakrasi HB. An inquiry into protein structure and genetic disease: Introducing undergraduates to bioinformatics in a large introductory course. *Cell Biology Education* 4:207–220; 2005.
2. Bednarski AE, Elgin SCR, Pakrasi, HB. Biology 3055 laboratory: Studying the genetic basis of disease using web-based bioinformatics tools. <http://www.nslc.wustl.edu/courses/Bio3055/bio3055.html>