



Computational and Mathematical Biology in HPC Science Gateways



SOX3 Award # 2231406

Revised Course Description

In some cases, it is too dangerous or impossible to do an experiment, so we can do numerical experiments through mathematical modeling and simulation. Besides learning mathematical modeling, the students in this course will learn basic commands, syntax, and fundamental programming in Python, and use them for solving problems in biology. The course targets students having majors/minors in mathematics, biology, and/or chemistry interested in learning computational and mathematical biology. The course consists of three (3) parts: 1) fundamental programming in Python, 2) computational biology, and 3) mathematical biology.

Use Cases

- 1) By using Python, determine the relative frequencies of constituent amino acids for each protein secondary structural class from the Protein Data Bank (PDB). See the Datasets column.
- 2) By using Python, research the best fit model for the data from in vitro experiments in which nanoparticles induced into cancer cells in biology laboratory at Jarvis Christian University, Hawkins, Texas.

Resource Needs/List

- 1) **GitHub**. The students need to collaborate in solving biological problems, so they share code through GitHub: <https://github.com/wsamyono/BulldogTeamFaceHackGW23>



GitHub

- 2) **Google Colab**. The students use Colab for collaborative programming without installing Python on their computers. <https://colab.research.google.com/>



- 3) **Jupyter Notebooks**. The students may use TACC and Jupyter 2 resources for running computationally larger Jupyter Notebook instances.



Your feedback is welcome!



Sample HPC/Gateways Exercise

Exercises 12 is from this journal paper: [An Introduction to Programming for Bioscientists: A Python-Based Primer | PLOS Computational Biology](#)

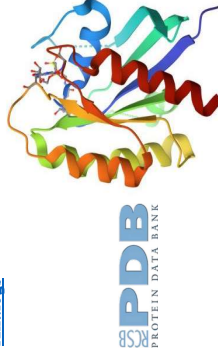


PLOS COMPUTATIONAL BIOLOGY

Exercise 12 (cumulative project):

First, obtain a set of several hundred protein structures from the PDB, as plaintext **.pdb** files (the exact number of entries is immaterial). Then, from this pool of data, determine the relative frequencies of the constituent amino acids for each protein secondary structural class; use only the three descriptors "helix," "sheet," and, for any AA not within a helix or sheet, "irregular." (Hint: In considering file parsing and potential data structures, search online for the PDB's file-format specifications.) Output your statistical data to a human-readable file format (e.g., comma-separated values, **.csv**) such that the results can be opened in a statistical or graphical software package for further processing and analysis. As a bonus exercise, use Python's **matplotlib** package to visualize the findings of your structural bioinformatics analysis.

The protein structure data 5KYK is from RCSB PDB [RCSB PDB - 5KYK: Covalent GTP-competitive inhibitors of KRAS G12C: Guanosine bisphosphonate Analogs](#)



Implementation Schedule

- Spring 2024 - GitHub and Google Colab
- Summer 2024 - DesignSafe-CI with Jupyter Notebooks
- Spring 2025 - Jetstream2 custom instances
- Summer 2025 - TACC computational resources possibly using TAPIS

Gateway Community Mentor Syllabus Suggestions



JARVIS CHRISTIAN UNIVERSITY HAWKINS, TEXAS Summer: Spring 2024 Course Number: CSCI 3300 Instructor: Dr. Wilfredo Samyono Contact: Dr. Wilfredo Samyono Email: wsamono@jarvis.edu Time of Class: Tue-Thur 1:00 - 1:20 PM CT Classroom Location: Zoom and HyperV 14	Office Location: Merrell Hall Jarvis Hall, Room 200 Office Hours: TBA
---	---

COURSE DESCRIPTION This course is designed to provide students with a solid foundation in computational biology and its applications. The course will cover the fundamentals of programming in Python, the basics of mathematical modeling, and the use of computational tools to analyze biological data. The course will also cover the basics of structural bioinformatics, including the use of the Protein Data Bank (PDB) and the RCSB PDB website. The course will conclude with a project in which students will apply their knowledge of computational biology to a real-world problem.	COURSE INSTRUCTORS Dr. Wilfredo Samyono Dr. Wilfredo Samyono, PhD Mathematics Department Jarvis Christian University wsamono@jarvis.edu
--	---

STUDENT LEARNING EXPECTED OUTCOMES/COMPETENCIES 1. Understanding the basics of programming in Python. 2. Understanding the basics of mathematical modeling and its applications. 3. Understanding the basics of structural bioinformatics and the use of the Protein Data Bank (PDB) and the RCSB PDB website.	STUDENT LEARNING EXPECTED OUTCOMES/COMPETENCIES 1. Understanding the basics of programming in Python. 2. Understanding the basics of mathematical modeling and its applications. 3. Understanding the basics of structural bioinformatics and the use of the Protein Data Bank (PDB) and the RCSB PDB website.
--	--

METHODS OF INSTRUCTION Lectures, discussions, and hands-on exercises.	METHODS OF INSTRUCTION Lectures, discussions, and hands-on exercises.
---	---

COURSE CONTENT 1. Introduction to Python programming. 2. Mathematical modeling and its applications. 3. Structural bioinformatics and the use of the Protein Data Bank (PDB) and the RCSB PDB website.	COURSE CONTENT 1. Introduction to Python programming. 2. Mathematical modeling and its applications. 3. Structural bioinformatics and the use of the Protein Data Bank (PDB) and the RCSB PDB website.
--	--

ASSESSMENT Exams, assignments, and projects.	ASSESSMENT Exams, assignments, and projects.
--	--

ADDITIONAL INFORMATION This course is a required course for students pursuing a degree in Computational Biology.	ADDITIONAL INFORMATION This course is a required course for students pursuing a degree in Computational Biology.
--	--

RECOMMENDED COURSEWORK None.	RECOMMENDED COURSEWORK None.
--	--

PREREQUISITES None.	PREREQUISITES None.
-------------------------------	-------------------------------

IMPORTANT DATES TO REMEMBER The course will run from June 1 to August 1, 2024.	IMPORTANT DATES TO REMEMBER The course will run from June 1 to August 1, 2024.
--	--

CLASSROOM POLICIES Students are expected to attend all classes and participate actively in the course.	CLASSROOM POLICIES Students are expected to attend all classes and participate actively in the course.
--	--

ADDITIONAL RESOURCES The course will use a variety of resources, including textbooks, online materials, and software.	ADDITIONAL RESOURCES The course will use a variety of resources, including textbooks, online materials, and software.
---	---

CONTACT INFORMATION Dr. Wilfredo Samyono wsamono@jarvis.edu	CONTACT INFORMATION Dr. Wilfredo Samyono wsamono@jarvis.edu
--	--

Resources / Science Gateways

1. **Research Collaboratory for Structural Bioinformatics Protein Data Bank**
Source for related computational tools

2. **DesignSafe-CI**
We may use some of the modules to expand our course.



3. **SGX3 Resource Catalog**
Other gateways may be found for future use in courses



Datasets

1. **Research Collaboratory for Structural Bioinformatic Protein Data Bank: RCSB PDB**



2. **Genomic Data Commons Data Portal: GDC Data Portal**



Possible Expansions

The course could be expanded to more advanced topics in the future by adding some more modules and new recent science gateways resources, and literature. We may collaborate with other faculty from different institutions to teach and conduct research in computational and mathematical biology.

Authors



Faculty Participant
Widodo Samyono, PhD
Mathematics Department,
Jarvis Christian University
wsamono@jarvis.edu



HPG/Gateways Mentor
Je'aimé Powell
Texas Advanced Computing Center,
University of Texas at Austin
jpowell@tacc.utexas.edu