PYTHON SCRIPTING
for Biochemistry & Molecular Biology

PART 2

Thursday, April 20th 2023 • 12:00 PM EDT

Registration (required): go.rutgers.edu/l6tcjpsx

LEARNING OBJECTIVES

This IQB Crash Course is second in a series of planned workshops designed to introduce life scientists to the power and flexibility of solving problems with Python. In the first workshop (pdb101.rcsb.org > Train > Training Courses), we explored some Python syntax, loops, logic, file parsing, working with multiple files, and molecular visualization using Python in Jupyter notebooks. If you have not used Python or done coding before, we encourage you to watch the first workshop and practice coding there.

We are offering this second workshop virtually and we will continue to use a live-coding approach. It will not be just another lecture. You will be coding right along with us as you learn additional skills, using some new libraries that we will introduce.

Each live coding segment will be followed by a hands-on programming exercise to introduce you to the programming language and to help you identify the challenges associated with coding. Following a brief review of Python and the Jupyter Notebook environment, you will learn to build and extract data from data frames using the Pandas library. Then you will learn curve fitting with enzyme kinetics data. We will explore two different methods to extract data from mmCIF files (the default storage format for the RCSB PDB) using the Biopython library and the RCSB PDB Application Program Interface (API). You will be exposed to lot of new material!

If time allows, we are also going to learn about molecular visualization in Jupyter Notebooks using the iCN3D tool. One of our major goals is to impart sufficient confidence that you will move beyond this workshop to explore other Python libraries and approaches for data analysis and presentation that apply directly to your professional development.

PROGRAM

12:00–12:05 PM Introduction
Stephen K. Burley, M.D., D.Phil., Founding Director, Institute for Quantitative Biomedicine, Rutgers University

12:05–12:30 PM Python Primer
Jessica A. Nash, Ph.D., Software Scientist, Molecular Sciences Software Institute

12:30–1:00 PM Working with Pandas
Paul A. Craig, Ph.D., Professor of Biochemistry, Rochester Institute of Technology

1:00–1:30 PM Questions and Hands-On Exercise

1:30–1:45 PM Break

1:45–2:15 PM Nonlinear regression
Paul A. Craig

2:15–2:45 PM Questions and Hands-On Exercise

2:45–3:00 PM Break

3:00–3:30 PM Biopython and mmCIF files
Jessica Nash

3:30–4:00 PM Questions and Hands-On Exercise

4:00–4:15 PM Break

4:15–4:45 PM PDB API and mmCIF files
Jessica Nash

4:45–5:00 PM Break

5:00–5:55 PM Molecular Visualization with iCN3D in Jupyter Notebooks
Jessica Nash and Paul Craig

5:55 PM Closing
Stephen K. Burley

CO-ORGANIZERS:
Paul A. Craig, Rochester Institute of Technology
Jessica A. Nash, Molecular Sciences Software Institute

CO-SPONSORS:
RCSB Protein Data Bank
rcsb.org

Institute for Quantitative Biomedicine (IQB)
iqb.rutgers.edu