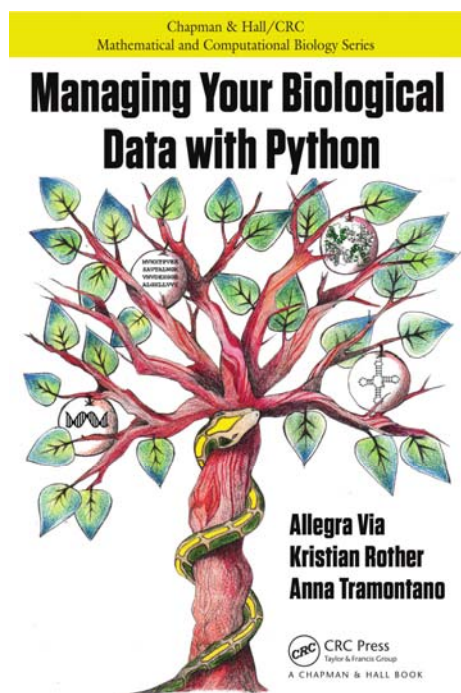


ALLEGRA VIA, KRISTIAN ROTHER, ANNA TRAMONTANO MANAGING YOUR BIOLOGICAL DATA WITH PYTHON



CRC Press
ISBN 978-1-4398-8093-7
Hard cover
560 pages
August 2014

Managing Your Biological Data with Python requires no prior programming experience and it provides biologists and other life scientists with the capability to work with biological data on their own using the powerful Python language. The book teaches programming techniques as well as ways to manage the data. It shows how to read data from files in different formats, analyze and manipulate the data, and write the results to a file or computer screen.

The first part of the text is an introduction to the Python language. The second part presents the basic elements of the language. The third part explains how to create bigger programs using techniques to write well-organized, efficient, and error-free code. The fourth part is focused on data visualization. The fifth part covers the Biopython programming library for reading and writing several biological file formats, querying the NCBI online databases, and retrieving biological records from the web. The last part is a cookbook of 20 specific programming “recipes”, ranging from secondary structure prediction and multiple sequence alignment analyses to superimposing protein three-dimensional structures.

There are numerous appendices to help the readers in their tasks and helps them easily analyze data and ultimately make better discoveries. Every piece of code presented in the book is aimed at solving real biological problems.

Table of Contents

Preface	xix
Part 1 Getting Started	1
Chapter 1 The Python Shell.....	5
Chapter 2 Your first Python Program.....	23
Part 2 Data Management	43
Chapter 3 Analyzing a Data Column.....	45
Chapter 4 Parsing Data Records.....	59
Chapter 5 Searching Data.....	77
Chapter 6 Filtering Data.....	93



Chapter 7 Managing Tabular Data.....	111
Chapter 8 Sorting Data	129
Chapter 9 Pattern Matching and Text Mining	143
Part 3 Modular Programming	163
Chapter 10 Divide a Program Into Functions	167
Chapter 11 Managing Complexity with Classes	189
Chapter 12 Debugging.....	205
Chapter 13 Using External Modules: The Python Interface to R	225
Chapter 14 Building Program Pipelines	245
Chapter 15 Writing Good Programs	263
Part 4 Data Visualization	285
Chapter 16 Creating Scientific Diagrams	287
Chapter 17 Creating Molecule Images with PyMOL	301
Chapter 18 Manipulating Images.....	323
Part 5 Biopython	341
Chapter 19 Working with Sequence Data.....	347
Chapter 20 Retrieving Data from Web Resources.....	363
Chapter 21 Working with 3D Structure Data	375
Part 6 Cookbok	535
Recipe 1: The PyCogent Library	395
Recipe 2: Reversing and Randomizing a Sequence.....	399
Recipe 3: Creating a Random Sequence with Probabilities.....	403
Recipe 4: Parsing Multiple Sequence Alignments Using Biopython	405
Recipe 5: Calculating a Consensus Sequence from a Multiple Sequence Alignment.....	409
Recipe 6: Calculating the Distance between Phylogenetic Tree Nodes	413
Recipe 7: Codon Frequencies in a Nucleotide Sequence	417
Recipe 8: Parsing RNA 2D Structures in the Vienna Format.....	421
Recipe 9: Parsing BLAST XML Output.....	425
Recipe 10: Parsing SBML Files	427
Recipe 11: Running BLAST.....	431
Recipe 12: Accessing, Downloading, and Reading Web Pages in Python.....	437
Recipe 13: Parsing HTML Files	441
Recipe 14: Split a PDB File into PDB Chain Files.....	445
Recipe 15: Find the Two Closest C α Atoms in a PDB Structure	447
Recipe 16: Extract the Interface between Two PDB Chains	451
Recipe 17: Building Homology Models Using Modeller	455
Recipe 18: RNA 3D Homology Modeling with ModeRNA	459
Recipe 19: Calculating RNA Base Pairs from a 3D Structure.....	463
Recipe 20: A Real Case of Structural Superimposition: The Serine Protease Catalytic Triad.....	467
APPENDIX A: COMMAND OVERVIEW.....	471
APPENDIX B: PYTHON RESOURCES	495
APPENDIX C: RECORD SAMPLES	499
APPENDIX D: HANDLING DIRECTORIES AND PROGRAMS WITH UNIX	507
INDEX	525