Lezione 7

Bioinformatica

Mauro Ceccanti[‡] e Alberto Paoluzzi[†]

[†]Dip. Informatica e Automazione – Università “Roma Tre”
[‡]Dip. Medicina Clinica – Università “La Sapienza”

Contents

BioPython
Installing and exploration
Tutorial

First Course Project
First Start
First Start with Biopython

BioPython

Biopython is a set of freely available tools for biological computation written in Python by an international team of developers.

[<+->]

- The web site provides an online resource for modules, scripts, and web links for developers of Python-based software for life science
- BioPython makes it as easy as possible to use Python for bioinformatics by creating high-quality, reusable modules and scripts

BioPython

Biopython is a set of freely available tools for biological computation written in Python by an international team of developers.

This wiki will help you download and install Biopython, and start using the libraries and tools

Download

Current Release – 1.52

http://biopython.org/wiki/Download

Biopython installation

Short version

[<+->]

- installer for windows: download Python-2.6.2.msi
- standard install on MacOSX, Linux and Unix:
  - download the source
  - from command line in a terminal:

```
1 > python setup.py build
2 > python setup.py test
3 > sudo python setup.py install
```

Biopython installation

Long version

- http://biopython.org/DIST/docs/install/Installation.html

Biopython installation

Best version ;-)

- from a terminal, with easy_install package already installed:

```
1 > easy_install -f http://biopython.org/DIST/biopython
```

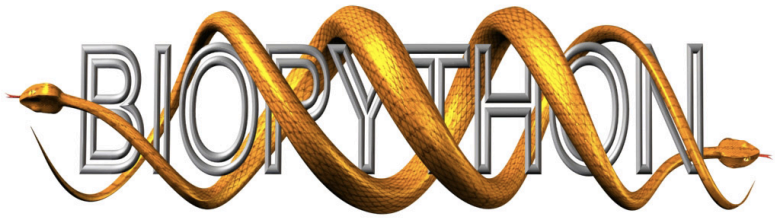
Contents

Biopython

Tutorial and cookbook

Biopython Tutorial and Cookbook

by Jeff Chang, Brad Chapman, Iddo Friedberg, Thomas Hamelryck, Michiel de Hoon, Peter Cock, and Tiago Antão



Biopython Tutorial and Cookbook

Biopython Package contents

The main Biopython releases have lots of functionality, including:

1. The ability to parse bioinformatics files into Python utilizable data structures, including support for the following formats:
   - Blast output – both from standalone and WWW Blast
   - Clustalw
   - FASTA
   - GenBank
   - PubMed and Medline
   - ExPASy files, like Enzyme and Prosite
   - SCOP, including 'dom' and 'lin' files
   - UniGene
   - SwissProt

Biopython Package contents

The main Biopython releases have lots of functionality, including:

1. Files in the supported formats can be iterated over record by record or indexed and accessed via a Dictionary interface.
2. Code to deal with popular on-line bioinformatics destinations such as:
   - NCBI – Blast, Entrez and PubMed services
   - ExPASy – Swiss-Prot and Prosite entries, as well as Prosite searches
3. Interfaces to common bioinformatics programs such as:
   - Standalone Blast from NCBI
   - Clustalw alignment program
   - EMBOSS command line tools
4. A standard sequence class that deals with sequences, ids on sequences, and sequence features.
5. Tools for performing common operations on sequences, such as translation, transcription and weight calculations.

Biopython Package contents

The main Biopython releases have lots of functionality, including:

1. Code to perform classification of data using k Nearest Neighbors, Naive Bayes or Support Vector Machines.
2. Code for dealing with alignments, including a standard way to create and deal with substitution matrices.
3. Code making it easy to split up parallelizable tasks into separate processes.
4. GUI-based programs to do basic sequence manipulations, translations, BLASTing, etc.
5. Extensive documentation and help with using the modules, including this file, on-line wiki documentation,
6. the web site, and the mailing list.
7. Integration with BioSQL, a sequence database schema also supported by the BioPerl and BioJava projects.

PDB: Atomic Coordinate Entry Format Description

Learn to parse PDB files, locally and on the web

Protein Data Bank Contents Guide

- Introduction
- Title Section
- Primary Structure Section
- Heterogen Section
- Secondary Structure Section
- Connectivity Annotation Section
- Miscellaneous Features Section
- Crystallographic and Coordinate Transformation Section
- Coordinate Section
- Connectivity Section
- Bookkeeping Section

Contents

First project

Curation of records of PDB files of aminoacids

Start from:

Amino Acids web page

and

Library of 3-D Molecular Structures, in particular from Amino Acids Section

Contents

Start using Biopython

Importing the package

```
1 Python 2.6.3 (r263:75184, Oct  2 2009, 07:56:03)
2 [GCC 4.0.1 (Apple Inc. build 5493)] on darwin
3 Type "copyright", "credits" or "license()" for more
      information.
4
5 IDLE 2.6.3
6 >>> import Bio
7 >>> print Bio.__version__
8 1.51
9 >>>
```

Start using Biopython

`help()` on the package

```
>>> help(Bio)
Help on package Bio:

NAME
    Bio - Collection of modules for dealing with
        biological data in Python.

FILE
    /Library/Frameworks/Python.framework/Versions/2.6/
        lib/python2.6/site-packages/biopython-1.51-py2.6-
        macosx-10.3-fat.egg/Bio/__init__.py

DESCRIPTION
    The Biopython Project is an international
        association of developers
    of freely available Python tools for computational
        molecular biology.

    http://biopython.org
```

Start using Biopython

`import` the PDB package

```
>>> from Bio.PDB import *
>>> dir()
['AbstractPropertyMap', 'Atom', 'Bio', 'CaPPBuilder', '
    Chain', 'DSSP', 'Dice', 'Entity', 'ExposureCN', '
    FragmentMapper', 'HSExposure', 'HSExposureCA', '
    HSExposureCB', 'Model', 'NeighborSearch', '
    PDBExceptions', 'PDBIO', 'PDBList', 'PDBParser', '
    PPBuilder', 'Polypeptide', 'Residue', 'ResidueDepth',
     'Select', 'Selection', 'Structure', '
    StructureAlignment', 'StructureBuilder', '
    Superimposer', 'Vector', '__builtins__', '__doc__', '
    __name__', '__package__', 'calc_angle', '
    calc_dihedral', 'extract', 'get_surface', 'is_aa', '
    m2rotaxis', 'make_dssp_dict', 'mmCIF', '
    parse_pdb_header', 'refmat', 'rotaxis', 'rotaxis2m',
    'rotmat', 'standard_aa_names', 'to_one_letter_code',
    'vector_to_axis']
```

Start using Biopython

`help()` on the package

```
>>> help(Bio.PDB)
Help on package Bio.PDB in Bio:

NAME
    Bio.PDB

FILE
    /Library/Frameworks/Python.framework/Versions/2.6/
        lib/python2.6/site-packages/biopython-1.51-py2.6-
        macosx-10.3-fat.egg/Bio/PDB/__init__.py

DESCRIPTION
    Classes that deal with macromolecular crystal
        structures. (eg.
    PDB and mmCIF parsers, a Structure class, a module
        to keep
    a local copy of the PDB up-to-date, selective IO of
        PDB files,
    etc.). Author: Thomas Hamelryck.  Additional code by
```

Start using Biopython

`help()` on the module `Atom`

```
>>> Atom
<module 'Bio.PDB.Atom' from '/Library/Frameworks/Python.
    framework/Versions/2.6/lib/python2.6/site-packages/
    biopython-1.51-py2.6-macosx-10.3-fat.egg/Bio/PDB/Atom
    .pyc'>
>>> from Bio.PDB.Atom import *
>>> help(Bio.PDB.Atom)
Help on module Bio.PDB.Atom in Bio.PDB:

NAME
    Bio.PDB.Atom - Atom class, used in Structure objects
        .

FILE
    /Library/Frameworks/Python.framework/Versions/2.6/
        lib/python2.6/site-packages/biopython-1.51-py2.6-
        macosx-10.3-fat.egg/Bio/PDB/Atom.py

CLASSES
```

Start using Biopython

Look the atom.py file ...

```
# Copyright (C) 2002, Thomas Hamelryck (thamelry@binf.ku.dk)
# This code is part of the Biopython distribution and governed by its
# license.  Please see the LICENSE file that should have been included
# as part of this package.

# Python stuff
import numpy

# My stuff
from Entity import DisorderedEntityWrapper
from Vector import Vector

__doc__="Atom class, used in Structure objects."

class Atom:
    def __init__(self, name, coord, bfactor, occupancy,
```