

# Lezione 7

## Bioinformatica

Mauro Ceccanti<sup>‡</sup> e Alberto Paoluzzi<sup>†</sup>

<sup>†</sup>Dip. Informatica e Automazione – Università “Roma Tre”

<sup>‡</sup>Dip. Medicina Clinica – Università “La Sapienza”



## BioPython

Installing and exploration  
Tutorial

## First Course Project

First Start  
First Start with Biopython



# 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.

[<+>]

- ▶ 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

## 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

## 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

## 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

Installing and exploration

**Tutorial**

## First Course Project

First Start

First Start with Biopython

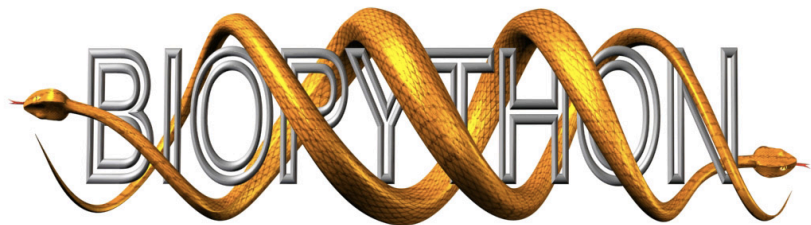


# Biopython

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

- Chapter 1 Introduction
- Chapter 2 [Quick Start – What can you do with Biopython?](#)
- Chapter 3 [Sequence objects](#)
- Chapter 4 Sequence Record objects
- Chapter 5 Sequence Input/Output
- Chapter 6 Sequence Alignment Input/Output, and Alignment Tools
- Chapter 7 [BLAST](#)
- Chapter 8 Accessing NCBI's Entrez databases
- Chapter 9 [Swiss-Prot and ExPASy](#)
- Chapter 10 [Going 3D: The PDB module](#)
- Chapter 11 Bio.PopGen: Population genetics
- Chapter 12 Supervised learning methods
- Chapter 13 Graphics including GenomeDiagram
- Chapter 14 Cookbook – Cool things to do with it
- Chapter 15 The Biopython testing framework
- Chapter 16 Advanced
- Chapter 17 Where to go from here – contributing to Biopython
- Chapter 18 Appendix: Useful stuff about Python





# 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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.



# 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.



# 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.



# 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.





# 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.



# 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.



# 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



# 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



# 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



# 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



# 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





# 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



# 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



# 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



# 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



# 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



# 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

## BioPython

Installing and exploration

Tutorial

## First Course Project

**First Start**

First Start with Biopython



# 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

## BioPython

Installing and exploration

Tutorial

## First Course Project

First Start

First Start with Biopython



# 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
4     information.
5
6 IDLE 2.6.3
7 >>> import Bio
8 >>> print Bio.__version__
9 1.51
10 >>>
```



# Start using Biopython

`help()` on the package

```
1 >>> help(Bio)
2 Help on package Bio:
3
4 NAME
5     Bio - Collection of modules for dealing with
6         biological data in Python.
7
8 FILE
9     /Library/Frameworks/Python.framework/Versions/2.6/
10    lib/python2.6/site-packages/biopython-1.51-py2.6-
11    macosx-10.3-fat.egg/Bio/__init__.py
12
13 DESCRIPTION
14     The Biopython Project is an international
15     association of developers
16     of freely available Python tools for computational
17     molecular biology.
18
19     http://biopython.org
```



# Start using Biopython

import the PDB package

```
1 >>> from Bio.PDB import *
2 >>> dir()
3 ['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

```
1 >>> help(Bio.PDB)
2 Help on package Bio.PDB in Bio:
3
4 NAME
5     Bio.PDB
6
7 FILE
8     /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
9
10 DESCRIPTION
11     Classes that deal with macromolecular crystal
        structures. (eg.
12     PDB and mmCIF parsers, a Structure class, a module
        to keep
13     a local copy of the PDB up-to-date, selective IO of
        PDB files,
14     etc.). Author: Thomas Hamelryck. Additional code by
        Kristian
```



# Start using Biopython

help() on the module Atom

```
1 >>> Atom
2 <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' >
3 >>> from Bio.PDB.Atom import *
4 >>> help(Bio.PDB.Atom)
5 Help on module Bio.PDB.Atom in Bio.PDB:
6
7 NAME
8     Bio.PDB.Atom - Atom class, used in Structure objects
9     .
10 FILE
11     /Library/Frameworks/Python.framework/Versions/2.6/
12     lib/python2.6/site-packages/biopython-1.51-py2.6-
13     macosx-10.3-fat.egg/Bio/PDB/Atom.py
14
15 CLASSES
16     Atom
```



# Start using Biopython

Look the atom.py file ...

```
1  # Copyright (C) 2002, Thomas Hamelryck (thamelry@binf.ku
    .dk)
2  # This code is part of the Biopython distribution and
    governed by its
3  # license. Please see the LICENSE file that should have
    been included
4  # as part of this package.
5
6  # Python stuff
7  import numpy
8
9  # My stuff
10 from Entity import DisorderedEntityWrapper
11 from Vector import Vector
12
13 __doc__="Atom_class,_used_in_Structure_objects."
14
15 class Atom:
16     def __init__(self, name, coord, bfactor, occupancy,
```

