

Biomedical Informatics: Lecture 1

M. Ceccanti and A. Paoluzzi

Tue, Mar 4, 2014

Outline: Syllabus, GitHub

- 1 Syllabus
- 2 Student projects
- 3 Grading
- 4 GitHub
- 5 Assignments
- 6 Introduction to bioinformatics

Syllabus

Biomedical Informatics 2014

- General information

Biomedical Informatics 2014

- General information
- Course notes and student home

Biomedical Informatics 2014

- General information
- Course notes and student home
- Programming tools

Biomedical Informatics 2014

- General information
- Course notes and student home
- Programming tools
 - Python

Biomedical Informatics 2014

- General information
- Course notes and student home
- Programming tools
 - Python
 - `pyplasm` (PLaSM for Python)

Biomedical Informatics 2014

- General information
- Course notes and student home
- Programming tools
 - Python
 - pyplasm (PLaSM for Python)
 - Javascript

Biomedical Informatics 2014

- General information
- Course notes and student home
- Programming tools
 - Python
 - pyplasm (PLaSM for Python)
 - Javascript
 - plasm.js (PLaSM for JavaScript)

Student projects

Student projects

Scheduling

- 1 **Project design**: due Fri, May 2, 2014

Student projects

Scheduling

- 1 **Project design**: due Fri, May 2, 2014
- 2 **Working prototype**: due Fri, June 6, 2014

Student projects

Scheduling

- 1 **Project design**: due Fri, May 2, 2014
- 2 **Working prototype**: due Fri, June 6, 2014
- 3 **Project deadline**: Before final exam

Grading

Exam requirements

- 1 Programming project (informatics: prof. Paoluzzi) (≤ 15)

Exam requirements

- 1 Programming project (informatics: prof. Paoluzzi) (≤ 15)
- 2 Term Paper (biomedical: prof. Ceccanti) (≤ 10)

Exam requirements

- 1 Programming project (informatics: prof. Paoluzzi) (≤ 15)
- 2 Term Paper (biomedical: prof. Ceccanti) (≤ 10)
- 3 Project design and prototype presentations (≤ 5)

GitHub

SignOn GitHub

GitHub is a **web-based hosting service** for **software development projects** that **use the Git** revision control system.

GitHub offers both paid plans for **private** repositories, and **free accounts** for **open source projects**.

- configure **git** and **github** (**guide**)

Assignments

Enroll to the course !!



To:

Cc:

Bcc:

Reply To:

Subject:

From:

Cognome Nome
 primo anno laurea magistrale (oppure: secondo ...)
 ingegneria informatica (oppure: altro)
 matricola: xxxxxx
 email: account@provider
 grafica computazionale: SI (oppure: NO)
 interessato a tesi di laurea: SI (oppure: NO)

send an email to me ... NOW!

Projects subjects of 2014

- 1 Essay (Pre-thesis) on bioinformatics topics at the **Mendel Institute** with **Dr. Valerio Mazza**

Projects subjects of 2014

- 1 Essay (Pre-thesis) on bioinformatics topics at the **Mendel Institute** with **Dr. Valerio Mazza**
- 2 Import/export of image/graphics/CAD formats to/from **LAR format**

Projects subjects of 2014

- 1 Essay (Pre-thesis) on bioinformatics topics at the **Mendel Institute** with **Dr. Valerio Mazza**
- 2 Import/export of image/graphics/CAD formats to/from **LAR format**
- 3 Design and implementation of a dedicated CMS (content management system) for **CRARL (Centro Riferimento Alcológico Regione Lazio)**, c/o Policlinico Umberto I

Projects subjects of 2014

- 1 Essay (Pre-thesis) on bioinformatics topics at the **Mendel Institute** with **Dr. Valerio Mazza**
- 2 Import/export of image/graphics/CAD formats to/from **LAR format**
- 3 Design and implementation of a dedicated CMS (content management system) for **CRARL (Centro Riferimento Alcolico Regione Lazio)**, c/o Policlinico Umberto I

Projects subjects of 2014

- 1 Essay (Pre-thesis) on bioinformatics topics at the **Mendel Institute** with **Dr. Valerio Mazza**
- 2 Import/export of image/graphics/CAD formats to/from **LAR format**
- 3 Design and implementation of a dedicated CMS (content management system) for **CRARL (Centro Riferimento Alcológico Regione Lazio)**, c/o Policlinico Umberto I

Your choice **to be discussed with me** (after April, 8)

References

Course syllabus

Principles of Biomedical Informatics, Second Edition

Handbook of Biomedical Informatics

Introduction to bioinformatics

Bioinformatics

Oxford English Dictionary, 2001

*(Molecular) bio – informatics: bioinformatics is conceptualising biology in terms of molecules (in the sense of physical chemistry) and **applying "informatics techniques"** (derived from disciplines such as applied maths, computer science and statistics) to understand and organise the information associated with these molecules, on a large scale. In short, bioinformatics is a management information system for molecular biology and has many practical applications.*

Bioinformatics Central Dogma

states that DNA is transcribed into RNA, which is then translated into proteins

Bioinformatics Central Dogma

states that DNA is transcribed into RNA, which is then translated into proteins

- DNA is composed of 4 **nucleotides** (guanine, cytosine, adenine, and thymine).

Bioinformatics Central Dogma

states that DNA is transcribed into RNA, which is then translated into proteins

- DNA is composed of 4 **nucleotides** (guanine, cytosine, adenine, and thymine).
- Sets of 3 nucleotides in a string of DNA are called **codons**.

Bioinformatics Central Dogma

states that DNA is transcribed into RNA, which is then translated into proteins

- DNA is composed of 4 **nucleotides** (guanine, cytosine, adenine, and thymine).
- Sets of 3 nucleotides in a string of DNA are called **codons**.
- The 20 **amino acids** are specified by these codons.

Bioinformatics Central Dogma

states that DNA is transcribed into RNA, which is then translated into proteins

- DNA is composed of 4 **nucleotides** (guanine, cytosine, adenine, and thymine).
- Sets of 3 nucleotides in a string of DNA are called **codons**.
- The 20 **amino acids** are specified by these codons.
- The **genetic code** is the mapping from codons to amino acids

Bioinformatics Central Dogma

states that DNA is transcribed into RNA, which is then translated into proteins

- DNA is composed of 4 **nucleotides** (guanine, cytosine, adenine, and thymine).
- Sets of 3 nucleotides in a string of DNA are called **codons**.
- The 20 **amino acids** are specified by these codons.
- The **genetic code** is the mapping from codons to amino acids
- A **gene** is a sequence of codons and "non-coding" sequences that determine when the gene is active (expressed).

Bioinformatics Central Dogma

states that DNA is transcribed into RNA, which is then translated into proteins

- DNA is composed of 4 **nucleotides** (guanine, cytosine, adenine, and thymine).
- Sets of 3 nucleotides in a string of DNA are called **codons**.
- The 20 **amino acids** are specified by these codons.
- The **genetic code** is the mapping from codons to amino acids
- A **gene** is a sequence of codons and "non-coding" sequences that determine when the gene is active (expressed).
- There are **stop and start codons** which signal the beginning and end of the protein coding sequence

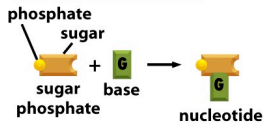
Bioinformatics Central Dogma

states that DNA is transcribed into RNA, which is then translated into proteins

- DNA is composed of 4 **nucleotides** (guanine, cytosine, adenine, and thymine).
- Sets of 3 nucleotides in a string of DNA are called **codons**.
- The 20 **amino acids** are specified by these codons.
- The **genetic code** is the mapping from codons to amino acids
- A **gene** is a sequence of codons and "non-coding" sequences that determine when the gene is active (expressed).
- There are **stop and start codons** which signal the beginning and end of the protein coding sequence
- Proteins are **organic compounds** made of amino acids arranged in a linear chain and folded into a globular form.

Bioinformatics Central Dogma

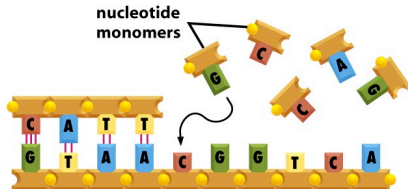
(A) building block of DNA



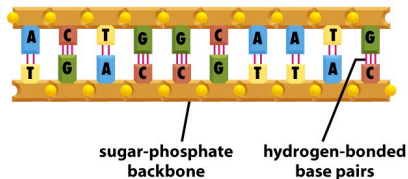
(B) DNA strand



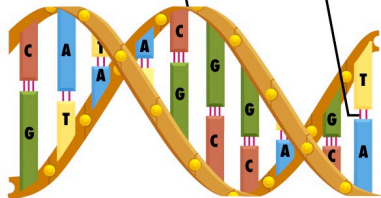
(C) templated polymerization of new strand



(D) double-stranded DNA

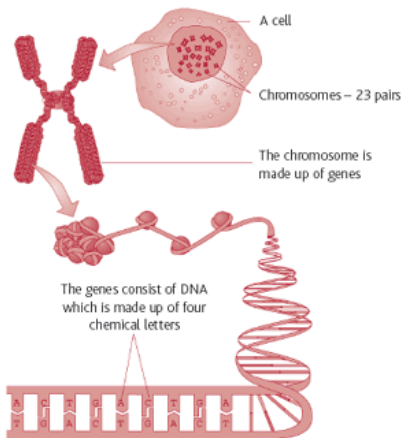


(E) DNA double helix



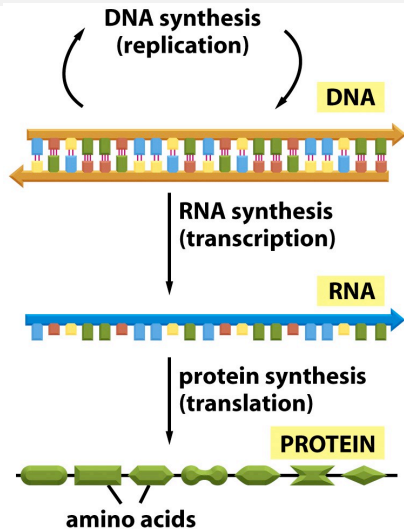
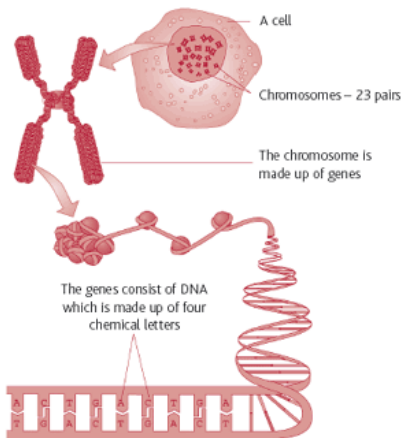
Bioinformatics Central Dogma

DNA structure

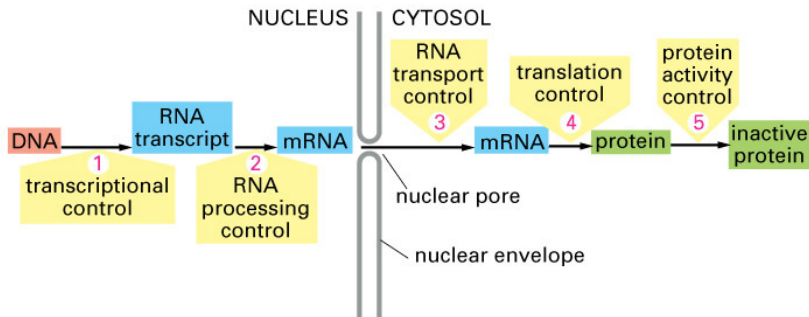


Bioinformatics Central Dogma

DNA structure



Gene expression process



An Information Portal to Biological Macromolecular Structures

As of **Tuesday Sep 29, 2009** there are 60464 Structures | [PDB Statistics](#)

WHAT'S NEW | [HELP](#) | [PRINT](#)

PDB ID or keyword

Search

Advanced Search

Home

News & Publications
Policies
FAQ
Contact
Feedback
About Us

Deposition

All Deposit Services
Electron Microscopy
NMR
Validation Server
BioSync Beamline
Related Tools

Search

Advanced Search
Latest Release
Latest Publications
Sequence Search
Ligand Search
Unreleased Entries
Browse Database
Histograms

Tools

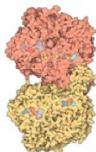
File Downloads
File Formats
Services: RESTful | SOAP

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the **wwPDB**, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Molecule of the Month: Xanthine Oxidoreductase



ultimately yielding uric acid, which is excreted from the body.

■ [Read more ...](#) ■ [Previous Features](#)

Our diet includes a wide variety of different molecules. Many of these molecules are broken down completely and used to generate the metabolic energy that powers our cells. Others are disassembled piece-by-piece and recycled to build our own proteins and nucleic acids. The ones that are left over are broken down and discarded. Xanthine oxidoreductase, shown here from PDB entry **1fo4**, is the last stop for extra purine nucleotides (ATP and GTP) in our cells.

Purines are broken down in several steps,

News

- [Complete News](#)
- [Newsletter](#)
- [Discussion Forum](#)
- [Job Listings](#)

29-September-2009

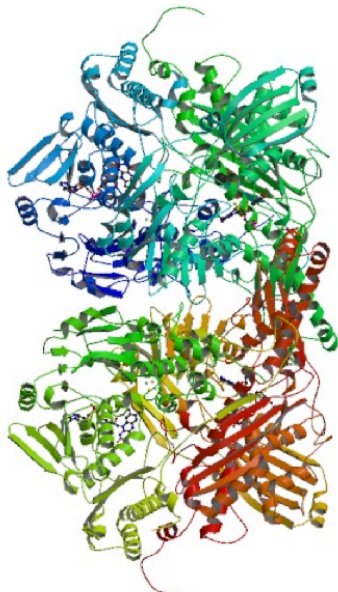
SF-Tool: A Tool for Crystallographic Experimental Data Validation

A streamlined, web-based tool is available for validating crystallographic experimental data. **SF-Tool** can be used to:

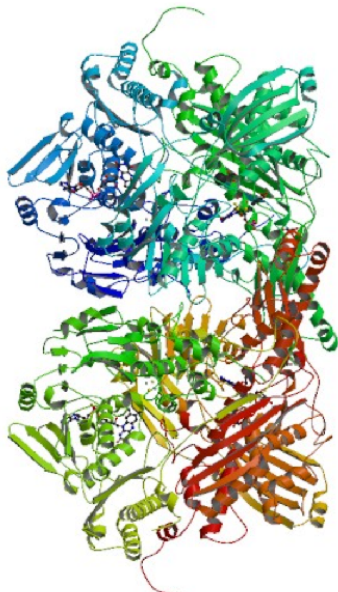
- Validate model coordinates against structure factor data (using SFCheck)
- Easily translate your structure factor file between different formats (mmCIF, MTZ, CNS/CNX, XPLOR, SHELX, TNT, HKL2000, SCALEPACK, D*Trek, SAINT, or OTHER format)

[More >>](#)

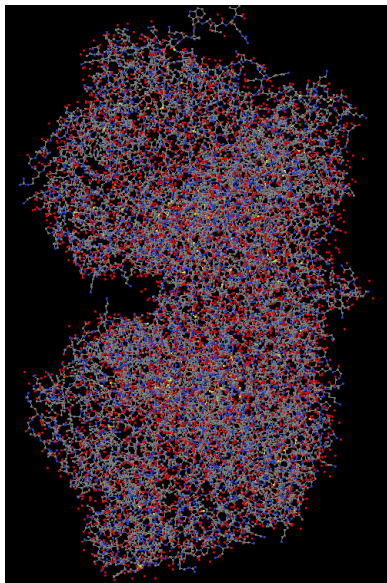
Example of PDB (Protein Data Bank) file



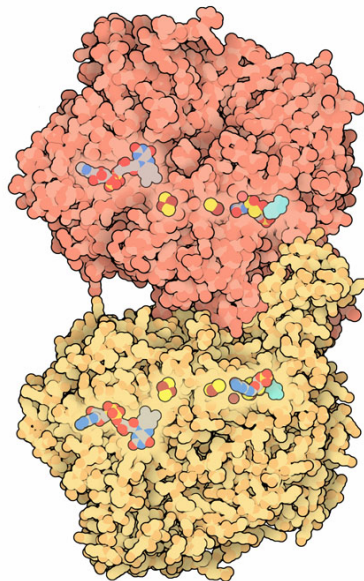
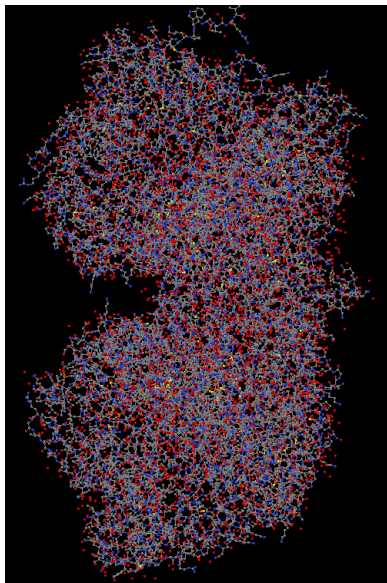
Example of PDB (Protein Data Bank) file



Example of PDB (Protein Data Bank) file



Example of PDB (Protein Data Bank) file



Example of PDB (Protein Data Bank) file

```

HEADER      OXIDOREDUCTASE              24-AUG-00   1FO4
TITLE       CRYSTAL STRUCTURE OF XANTHINE DEHYDROGENASE ISOLATED FROM
TITLE       2 BOVINE MILK
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: XANTHINE DEHYDROGENASE;
COMPND     3 CHAIN: A, B;
COMPND     4 SYNONYM: XD;
COMPND     5 EC: 1.1.1.284
SOURCE     MOL_ID: 1;
SOURCE     2 ORGANISM SCIENTIFIC: BOS TAURUS;
SOURCE     3 ORGANISM COMMON: CATTLE;
SOURCE     4 ORGANISM TAXID: 9913;
SOURCE     5 SECRETION: MILK
KEYWDS     XANTHINE DEHYDROGENASE, FAD, MOLYBDOPTEIN, 2FE-2S IRON
KEYWDS     2 SULFUR CENTERS, SALICYLATE, OXIDOREDUCTASE
EXPDTA     X-RAY DIFFRACTION
AUTHOR     C.ENROTH,B.T.EGER,K.OKAMOTO,T.NISHINO,T.NISHINO,E.F.PAI
REVDAT 1   24-FEB-09 1FO4 1   VERSN
REVDAT 2   01-APR-03 1FO4 1   JRNL
REVDAT 1   25-OCT-00 1FO4 0
JRNL       AUTH  C.ENROTH,B.T.EGER,K.OKAMOTO,T.NISHINO,T.NISHINO,
JRNL       AUTH 2 E.F.PAI
JRNL       TITL  CRYSTAL STRUCTURES OF BOVINE MILK XANTHINE
JRNL       TITL 2 DEHYDROGENASE AND XANTHINE OXIDASE:
JRNL       TITL 3 STRUCTURE-BASED MECHANISM OF CONVERSION.
JRNL       REF   PROC.NATL.ACAD.SCI.USA          V. 97 10723 2000
JRNL       REFIN  ISSN 0027-8424
JRNL       PMID  11005854
JRNL       DOI   10.1073/PNAS.97.20.10723
REMARK 1
REMARK 2
REMARK 3 2 RESOLUTION.      2.10 ANGSTROMS.
REMARK 3
REMARK 3 REFINEMENT.
REMARK 3 PROGRAM      : CNS
REMARK 3 AUTHORS      : BRUNGER,ADAMS,CLORE,DELANO,GROS,GROSSE-
REMARK 3                : KUNSTLIEV,JIANG,KUZNEVSKI,NILGES, PANNU,
REMARK 3                : READ,RICE,SIMONSON,WARREN
REMARK 3
REMARK 3 REFINEMENT TARGET : ENGH & HUBER
REMARK 3
REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 2.10
REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS)  : 25.00
REMARK 3 DATA CUTOFF (SIGMA(F)) : 0.000
REMARK 3 DATA CUTOFF HIGH (ABS(F)) : NULL
REMARK 3 DATA CUTOFF LOW (ABS(F)) : NULL
REMARK 3 COMPLETENESS (WORKING+TEST) (%) : 86.9
REMARK 3 NUMBER OF REFLECTIONS          : 154198
REMARK 3
REMARK 3 FIT TO DATA USED IN REFINEMENT.
REMARK 3 CROSS-VALIDATION METHOD          : NULL
REMARK 3 FREE R VALUE TEST SET SELECTION : RANDOM
ATOM       1  N  ALA  A  3      14.631  27.952  21.732  1.00  44.39  N
ATOM       2  CA  ALA  A  3      15.694  28.727  22.432  1.00  43.95  C
ATOM       3  C  ALA  A  3      16.446  27.796  23.355  1.00  44.12  C
ATOM       4  O  ALA  A  3      16.545  26.597  23.896  1.00  43.86  O
ATOM       5  CB  ALA  A  3      15.875  29.844  23.234  1.00  43.15  C
ATOM       6  N  ASP  A  4      16.986  28.351  24.432  1.00  43.50  N
ATOM       7  CA  ASP  A  4      17.698  27.542  25.402  1.00  42.49  C
ATOM       8  C  ASP  A  4      19.046  27.005  24.915  1.00  40.66  C
ATOM       9  O  ASP  A  4      20.028  27.073  25.645  1.00  41.20  O
ATOM      10  CB  ASP  A  4      16.775  26.399  25.857  1.00  43.83  C
ATOM      11  CG  ASP  A  4      17.522  25.244  26.483  1.00  45.72  C
ATOM      12  OD1  ASP  A  4      18.194  24.486  25.743  1.00  47.22  O
ATOM      13  OD2  ASP  A  4      17.434  25.090  27.719  1.00  47.21  O
ATOM      14  N  GLU  A  5      19.111  26.488  23.693  1.00  38.21  N
ATOM      15  CA  GLU  A  5      20.372  25.944  23.186  1.00  35.72  C
ATOM      16  C  GLU  A  5      21.136  26.883  22.246  1.00  33.85  C
ATOM      17  O  GLU  A  5      20.572  27.426  21.300  1.00  33.09  O
ATOM      18  CB  GLU  A  5      20.116  24.615  22.472  1.00  36.37  C
ATOM      19  CG  GLU  A  5      21.048  23.495  22.893  1.00  36.60  C
ATOM      20  CD  GLU  A  5      20.708  22.163  22.240  1.00  37.83  C
ATOM      21  OE1  GLU  A  5      20.777  22.065  20.997  1.00  37.11  O
ATOM      22  OE2  GLU  A  5      20.376  21.288  22.974  1.00  37.32  O
ATOM      23  N  LEU  A  6      22.425  27.068  22.522  1.00  31.97  N
ATOM      24  CA  LEU  A  6      23.287  27.919  21.704  1.00  30.36  C
ATOM      25  C  LEU  A  6      24.194  27.015  20.878  1.00  29.42  C
ATOM      26  O  LEU  A  6      24.911  26.176  21.424  1.00  29.15  O
ATOM      27  CB  LEU  A  6      24.141  28.837  22.586  1.00  29.72  C
ATOM      28  CG  LEU  A  6      25.255  29.634  21.892  1.00  29.44  C
ATOM      29  CD1  LEU  A  6      24.659  30.626  20.896  1.00  28.91  C
ATOM      30  CD2  LEU  A  6      26.082  30.366  22.937  1.00  29.35  C
ATOM      31  N  VAL  A  7      24.164  27.187  19.563  1.00  28.43  C
ATOM      32  CA  VAL  A  7      24.979  26.356  18.681  1.00  27.74  C
ATOM      33  C  VAL  A  7      25.872  27.163  17.753  1.00  27.10  C
ATOM      34  O  VAL  A  7      25.392  27.988  16.978  1.00  27.83  O
ATOM      35  CB  VAL  A  7      24.091  25.437  17.807  1.00  28.10  C
ATOM      36  CG1  VAL  A  7      24.958  24.616  16.865  1.00  28.34  C
ATOM      37  CG2  VAL  A  7      23.257  24.519  18.692  1.00  27.28  C
ATOM      38  N  PHE  A  8      27.178  26.938  17.842  1.00  26.13  N
ATOM      39  CA  PHE  A  8      28.112  27.629  16.965  1.00  25.66  C
ATOM      40  C  PHE  A  8      29.304  26.731  16.663  1.00  25.33  C
ATOM      41  O  PHE  A  8      29.341  25.576  17.093  1.00  24.51  O
ATOM      42  CB  PHE  A  8      28.546  28.977  17.570  1.00  25.38  C
ATOM      43  CG  PHE  A  8      29.321  28.870  18.850  1.00  25.54  C
ATOM      44  CD1  PHE  A  8      30.712  28.900  18.839  1.00  26.21  C
ATOM      45  CD2  PHE  A  8      28.666  28.084  20.074  1.00  25.64  C
ATOM      46  CE1  PHE  A  8      31.441  28.875  20.032  1.00  25.96  C
ATOM      47  CE2  PHE  A  8      29.385  28.777  21.269  1.00  26.36  C
ATOM      48  CZ  PHE  A  8      30.779  28.815  21.245  1.00  25.99  C
ATOM      49  N  PHE  A  9      30.261  27.240  15.900  1.00  24.87  N
ATOM      50  CA  PHE  A  9      31.429  26.441  15.552  1.00  24.62  C
ATOM      51  C  PHE  A  9      32.740  27.148  15.868  1.00  23.85  C
ATOM      52  O  PHE  A  9      32.836  28.368  15.780  1.00  23.17  O
ATOM      53  CB  PHE  A  9      31.408  26.074  14.862  1.00  24.82  C

```


Beyond central dogma: epigenetics

Genetic Science Learning Center – University of Utah
WHAT IS EPIGENETICS?

Science 29 October 2010:
Vol. 330 no. 6004 p. 611
DOI: 10.1126/science.330.6004.611
INTRODUCTION TO SPECIAL ISSUE
What Is Epigenetics?