

# Retrieving and Viewing Protein Structures from the Protein Data Base

7.88J Protein Folding

Prof. David Gossard

September 15, 2003

## PDB Acknowledgements

The **Protein Data Bank** (PDB - <http://www.pdb.org/>) is the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. "The Protein Data Bank." *Nucleic Acids Research* 28 (2000): 235-242.

(PDB Advisory Notice on using materials available in the archive: [http://www.pdb.org/pdb/static.do?p=general\\_information/about\\_pdb/pdb\\_advisory.html](http://www.pdb.org/pdb/static.do?p=general_information/about_pdb/pdb_advisory.html))

## Data used in the "Retrieving, Viewing Protein Structures from the Protein Data Base" Lecture Notes for 7.88J - Protein Folding

PDB site data on page 4 ("PDB Growth") and 7 ("Not all Structures are Different").

PDB screenshots: pages 9 – 14.

Pages 20-22 contain screen shots from ExPASy (SwissPDB):

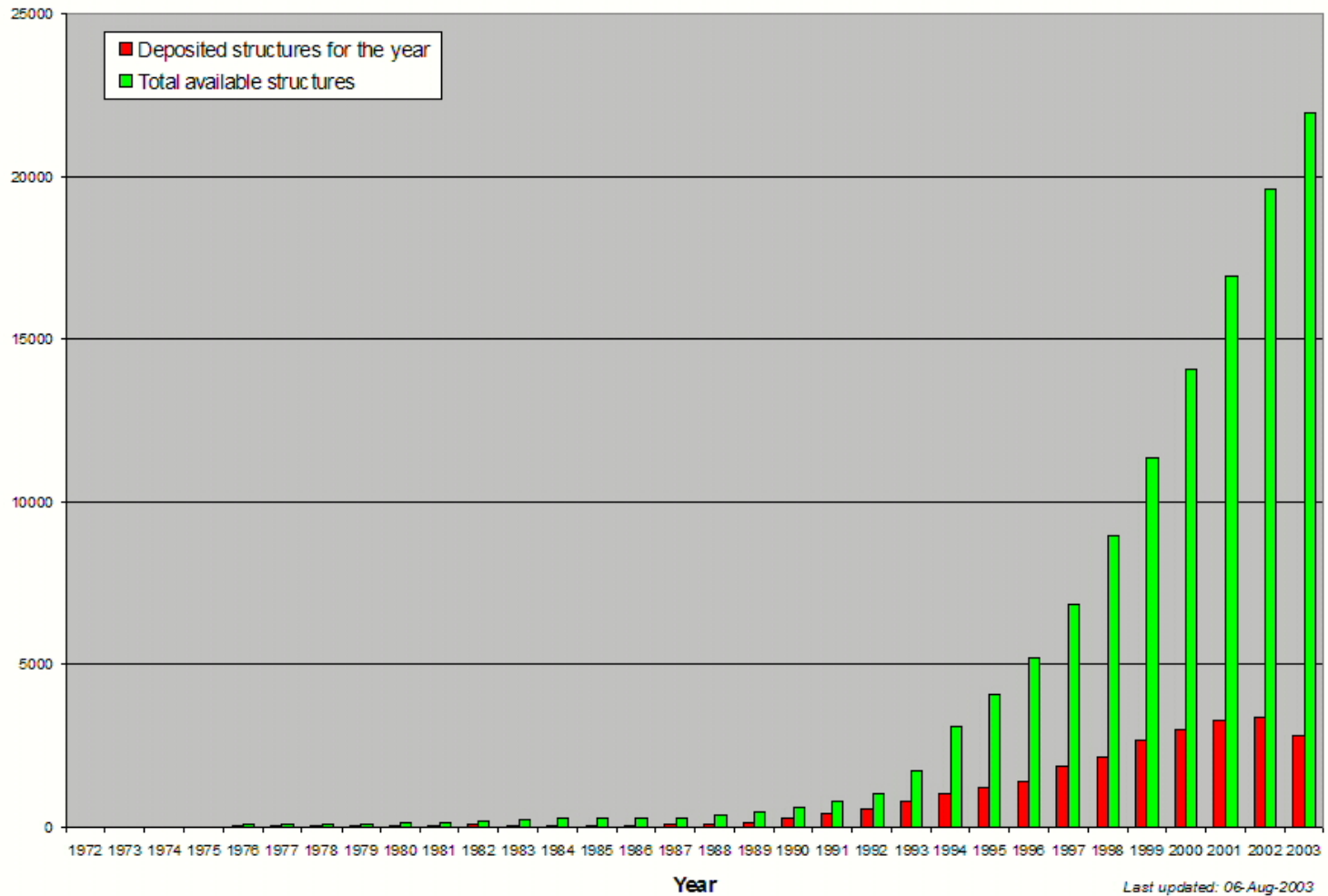
Appel, R. D., A. Bairoch, and D. F. Hochstrasser. "A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. Trends." *Biochem. Sci.* 19 (1994): 258-260.

(ExPASy (**Ex**pert **P**rotein **A**nalysis **S**ystem) proteomics server disclaimer: <http://us.expasy.org/disclaimer.html>)

# Protein Data Base

- Established in 1971
  - Funded by NSF, DOE, NIH
  - Operated by Rutgers, SDSC, NIST
- Purpose: Make protein structure data available to the entire scientific community
- In the beginning: “less than a dozen” protein structures
- Currently has 22,333 protein structures
- Growing at 20% per year
- New structures 50 times larger than those in 1971 are commonplace

# PDB Growth



# Why the “Knee in the Curve”?

- Engineered bacteria as a source of proteins
- Improved crystal-growing conditions
- More intense sources of X-rays
- Cryogenic treatment of crystals
- Improved detectors & data collection
- New method - NMR:
  - Accounts for 15% of new structures in PDB
  - Enables determination of structure of proteins in solution

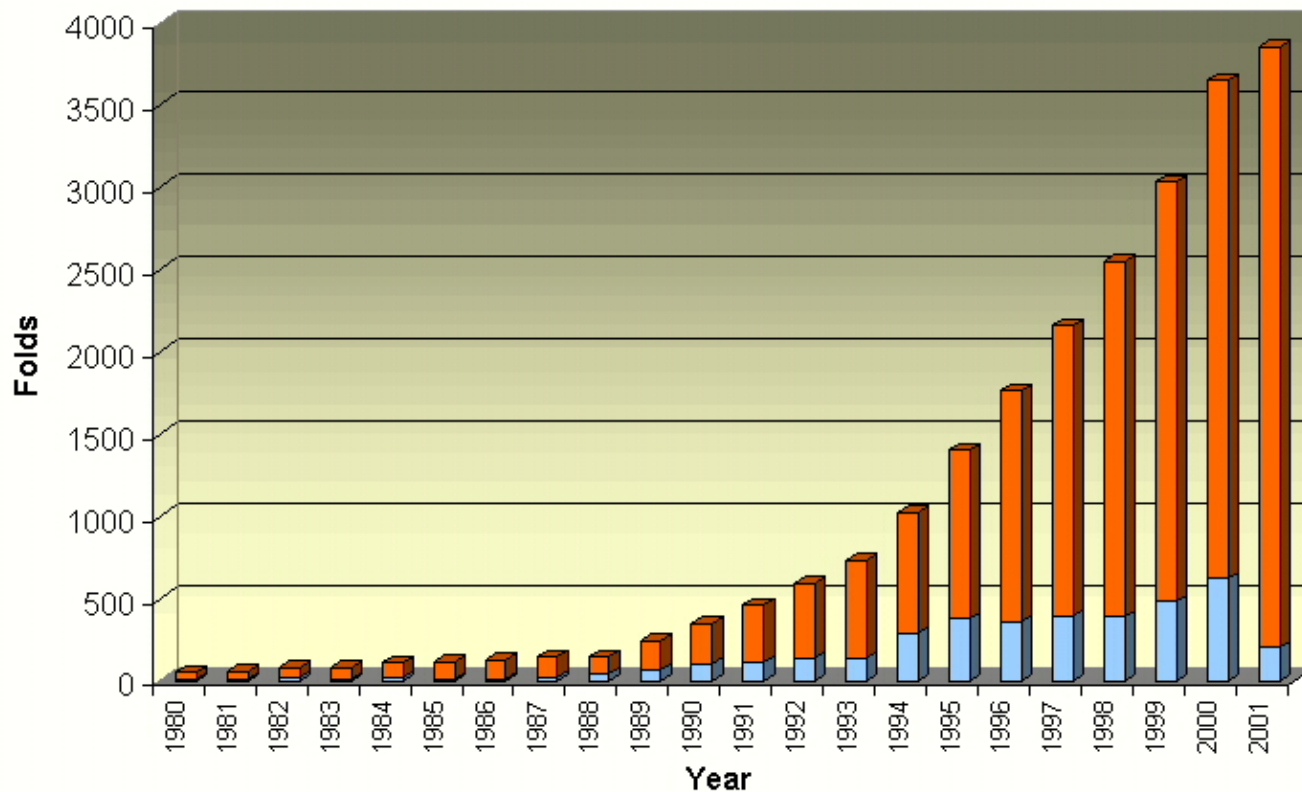
“Protein Structures: From Famine to Feast”, Berman, et.al.  
American Scientist v.90, p.350-359, July-August 2002

# Why is the PDB Important?

- “Collective Leverage”  
for ...
- Understanding molecular machinery
- Rational drug design
- Engineering new molecules
- Structural genomics

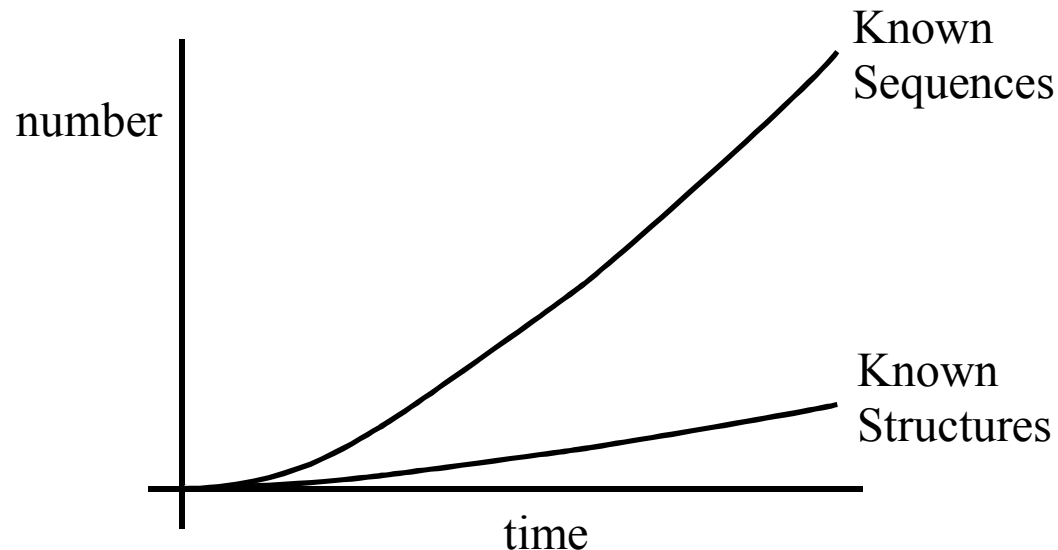
# Not all Structures are Different

- PDB Growth in “New Folds”



# Structure vs Sequence

- New protein structures are being solved more slowly than new protein sequences are being solved
  - Currently, known protein sequences outnumber known protein structures
  - The “sequence-structure” gap continues to widen





# PDB Website

<http://www.rcsb.org/pdb/>

Enter what  
you know...

**DEPOSIT data**  
**DOWNLOAD files**  
**browse LINKS**  
**BETA TEST new features**

**Current Holdings**  
18616 Structures  
Last Update: 03-Sep-2002  
[PDB Statistics](#)

**Molecule of the Month:**  
[Reverse Transcriptase](#)

The Protein Data Bank (PDB) is operated by Rutgers, The State University of New Jersey; the San Diego Supercomputer Center at the University of California, San Diego; and the National Institute of Standards and Technology -- three members of the [Research Collaboratory for Structural Bioinformatics \(RCSB\)](#). The PDB is supported by funds from the [National Science Foundation](#), the [Department of Energy](#), and two units of the National Institutes of Health; the [National Institute of General Medical Sciences](#) and the [National Library of Medicine](#).

**PROTEIN DATA BANK**

Welcome to the PDB, the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data.

[RCSB Home](#) [Contact Us](#) [Help](#)

**Did you find what you wanted?**

[ABOUT PDB](#) | [DATA UNIFORMITY](#) | [RECENT FEATURES](#) | [USER GUIDES](#) | [FILE FORMATS](#) | [EDUCATION](#) | [STRUCTURAL GENOMICS](#) | [PUBLICATIONS](#) | [SOFTWARE](#)

**Search the Archive** ?

Enter a [PDB ID](#) or keyword [Query Tutorial](#)

query by PDB id only  match exact word  
 remove sequence homologues

[SearchLite](#) keyword search form with examples  
[SearchFields](#) customizable search form  
[Status Search](#) find entries awaiting release

**News** [Complete News Newsletter](#) [pdb-I Mailing List Subscribe](#)

3-Sep-2002  
**PDB Focus: Author Searches** Users can query for a particular author of a structure or a primary citation using the SearchFields interface...  
[\[MORE...\]](#)

**PDB Mirrors**

*\*\*Please bookmark a mirror site\*\**

[San Diego Supercomputer Center\\*](#)  
[Rutgers University\\*](#)  
[National Institute of Standards and Technology\\*](#)  
[Cambridge Crystallographic Data Centre, UK](#)  
[National University of Singapore](#)  
[Osaka University, Japan](#)  
[Universidade Federal de Minas Gerais, Brazil](#)

[OTHER SITES](#)

*\*RCSB partner*

In citing the PDB please refer to:

# Query Result Browser

**PDB**  
PROTEIN DATA BANK

## Query Result Browser

Help PDB Home Contact us

Your query found 22 structures in the current PDB release and you have selected 0 structures so far. (There are currently 3 structures being processed or "on hold" matching your query!) You can select specific structures by clicking on the checkbox next to their id. If you do not select any structures, certain options will default to all structures. To examine an individual structure select the Explore link!

Pull down to select option:

◀ ◀ 1-20 ▶ ▶

<input type="checkbox"/>	<b>1A2W</b>	Deposited: 12-Jan-1998 Exp. Method: X-ray Diffraction Resolution: 2.10 Å	{ EXPLORE }
	Title	Crystal Structure Of A 3D Domain-Swapped Dimer Of Bovine Pancreatic Ribonuclease A	
	Classification	Endonuclease	
	Compound	Mol_Id: 1; Molecule: Ribonuclease A; Chain: A, B; Fragment: Swapped Helical Domain; Ec: 3.1.27.5; Biological_Unit: Monomer; Other_Details: Swapped Helical Domain Contains Residues 1-15, Hinge Loops Contain Residues 16-22, Major Domain Contains Residues 23-124	
<input type="checkbox"/>	<b>1A5P</b>	Deposited: 17-Feb-1998 Exp. Method: X-ray Diffraction Resolution: 1.60 Å	{ EXPLORE }
	Title	C[40,95]A Variant Of Bovine Pancreatic Ribonuclease A	
	Classification	Hydrolase	
	Compound	Mol_Id: 1; Molecule: Ribonuclease A; Chain: Null; Ec: 3.1.27.5; Engineered: Yes; Mutation: C40A, C95A	
<input type="checkbox"/>	<b>1A5Q</b>	Deposited: 17-Feb-1998 Exp. Method: X-ray Diffraction Resolution: 2.30 Å	{ EXPLORE }
	Title	P93A Variant Of Bovine Pancreatic Ribonuclease A	
	Classification	Hydrolase	
	Compound	Mol_Id: 1; Molecule: Ribonuclease A; Chain: Null; Ec: 3.1.27.5; Engineered: Yes; Mutation: P93A	
<input type="checkbox"/>	<b>1BEL</b>	Deposited: 21-Dec-1995 Exp. Method: X-ray Diffraction Resolution: 1.60 Å	{ EXPLORE }
	Title	Hydrolase Phosphoric Diester, RNA	
	Classification	Hydrolase	

Which one do I want?

Let's look at this one ...

# Structure Explorer

Yep, that's  
the right  
one...

**PDB**  
PROTEIN DATA BANK

## Structure Explorer - 1FS3

### Summary Information

**Summary Information**

**Title:** Crystal Structure Of Wild-Type Bovine Pancreatic Ribonuclease A

**Compound:** **Mol Id:** 1; **Molecule:** Ribonuclease A; **Chain:** A; **Synonym:** Ribonuclease Pancreatic; **Ec:** 3.1.27.5; **Engineered:** Yes

**Authors:** E. Chatani, R. Hayashi, H. Moriyama, T. Ueki

**Exp. Method:** X-ray Diffraction

**Classification:** Hydrolase

**EC Number:** 3.1.27.5

**Source:** Bos taurus

**Primary Citation:** Chatani, E., Hayashi, R., Moriyama, H., Ueki, T.: Conformational Strictness Required for Maximum Activity and Stability of Bovine Pancreatic Ribonuclease A as Revealed by Crystallographic Study of Three Phe120 Mutants at 1.4 Å Resolution *Protein Sci.* 11 pp. 72 (2002)  
[ [Medline](#) ]

**Deposition Date:** 08-Sep-2000      **Release Date:** 13-Feb-2002

**Resolution [Å]:** 1.40      **R-Value:** 0.217

**Space Group:** P 32 2 1

**Unit Cell:** dim [Å]:    a 64.05    b 64.05    c 63.35  
angles [°]: alpha 90.00 beta 90.00 gamma 120.00

[View Structure](#)  
[Download/Display File](#)  
[Structural Neighbors](#)  
[Geometry](#)  
[Other Sources](#)  
[Sequence Details](#)  
[Structure Factors \(compressed\)](#)

Explore   
[SearchLite](#) [SearchFields](#)

View it...

Download it...

# View Structure



## Structure Explorer - 1FS3

**Title** Crystal Structure Of Wild-Type Bovine Pancreatic Ribonuclease A  
**Classification** Hydrolase  
**Compound** Mol Id: 1; Molecule: Ribonuclease A; Chain: A; Synonym: Ribonuclease Pancreatic; Ec: 3.1.27.5; Engineered: Yes  
**Exp. Method** X-ray Diffraction

### View Structure

[Summary Information](#)  
[View Structure](#)  
[Download/Display File](#)  
[Structural Neighbors](#)  
[Geometry](#)  
[Other Sources](#)  
[Sequence Details](#)  
[Structure Factors \(compressed\)](#)

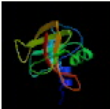
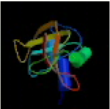
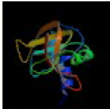
### Interactive 3D Display:

Choose from the following [display options](#) (asymmetric unit only):

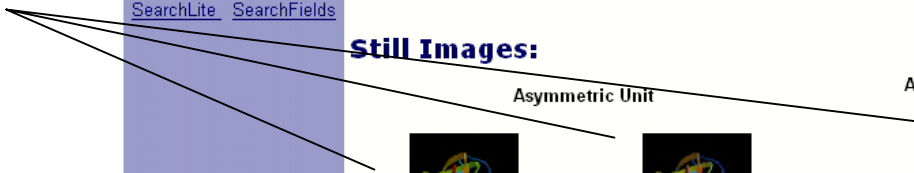
- [VRML \(default options\)](#): Interactive immersive ribbon diagram
- [VRML \(custom options, full screen display\)](#): Interactive immersive ribbon or cylinder diagram with ligands
- [Rasmol](#)
- [Swiss-PdbViewer](#)
- [MICE - Molecular Interactive Collaborative Environment](#) (requires Java Plugin)
- [FirstGlance](#) (needs [Chime](#))
- [Protein Explorer](#) (needs [Chime](#))
- [Sting Millennium](#) (needs [Chime](#) and Java)
- Java (simple interactive sequence/structure/property backbone diagram):

[HELP](#)  
**Download Help**  
[VRML](#)  
[Rasmol](#)  
[Swiss-PdbViewer](#)  
[Chime](#)  
[MICE](#)

### Still Images:

Asymmetric Unit		Assumed Biological Molecule
		
<a href="#">Ribbons (250x250)</a> <a href="#">Ribbons (500x500)</a>	<a href="#">Cylinders (250x250)</a> <a href="#">Cylinders (500x500)</a>	<a href="#">Ribbons (250x250)</a> <a href="#">Ribbons (500x500)</a>

Static Images



# Download/Display

Display the  
file header...

Download  
the file...  
(Select this  
file format)

**PDB**  
PROTEIN DATA BANK

## Structure Explorer - 1FS3

**Title** Crystal Structure Of Wild-Type Bovine Pancreatic Ribonuclease A  
**Classification** Hydrolase  
**Compound** Mol. Id: 1; Molecule: Ribonuclease A; Chain: A; Synonym: Ribonuclease Pancreatic; Ec: 3.1.27.5; Engineered: Yes  
**Exp. Method** X-ray Diffraction

### Download/Display File

Choose from the following data representation formats:

	file format	
	PDB	mmCIF
complete with coordinates	HTML TEXT	TEXT
"header only" (no coordinates)	HTML TEXT	-

### Download the Structure File:

Choose from the following file and compression formats:

compression	file format	
	PDB	mmCIF
none	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Unix compressed	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
GNU zipped ("gzipped")	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

# Header Information

**PDB**  
PROTEIN DATA BANK

## Structure Explorer - 1FS3

**Title** Crystal Structure Of Wild-Type Bovine Pancreatic Ribonuclease A  
**Classification** Hydrolase  
**Compound** Mol\_id: 1; Molecule: Ribonuclease A; Chain: A; Synonym: Ribonuclease Pancreatic; Ec: 3.1.27.5; Engineered: Yes  
**Exp. Method** X-ray Diffraction

**Download/Display File**

**Summary Information**

**View Structure**

**Download/Display File**

**Structural Neighbors**

**Geometry**

**Other Sources**

**Sequence Details**

**Structure Factors**  
(compressed)

Explore

[SearchLite](#) [SearchFields](#)

**Save full entry to disk**

```
HEADER      HYDROLASE                               08-SEP-00   1FS3
TITLE      CRYSTAL STRUCTURE OF WILD-TYPE BOVINE PANCREATIC
TITLE      2 RIBONUCLEASE A
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: RIBONUCLEASE A;
COMPND     3 CHAIN: A;
COMPND     4 SYNONYM: RIBONUCLEASE PANCREATIC;
COMPND     5 EC: 3.1.27.5;
COMPND     6 ENGINEERED: YES
SOURCE     MOL_ID: 1;
SOURCE     2 ORGANISM_SCIENTIFIC: BOS TAURUS;
SOURCE     3 ORGANISM_COMMON: BOVINE;
SOURCE     4 TISSUE: PANCREAS;
SOURCE     5 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE     6 EXPRESSION_SYSTEM_COMMON: BACTERIA;
SOURCE     7 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID
KEYWDS     RIBONUCLEASE, RNASE A, BOVINE PANCREAS, HYDROLASE
EXPDTA     X-RAY DIFFRACTION
AUTHOR     E.CHATANI,R.HAYASHI,H.MORIYAMA,T.UEKI
REVDAT    1   13-FEB-02 1FS3   0
JRNL       AUTH   E.CHATANI,R.HAYASHI,H.MORIYAMA,T.UEKI
JRNL       TITL   CONFORMATIONAL STRICTNESS REQUIRED FOR MAXIMUM
JRNL       TITL 2 ACTIVITY AND STABILITY OF BOVINE PANCREATIC
JRNL       TITL 3 RIBONUCLEASE AS REVEALED BY CRYSTALLOGRAPHY
```

# Visualizing Proteins

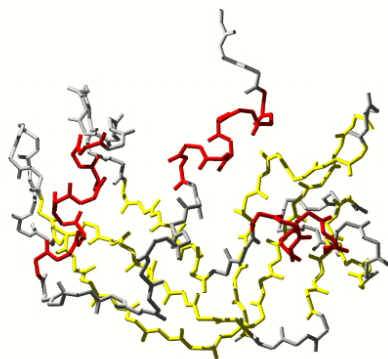
- High complexity
- Multiple levels of structure
- Important properties are “distributed”  
throughout the 3D structure
- No single/simple “point”  
at which to look

# Visualization Objectives

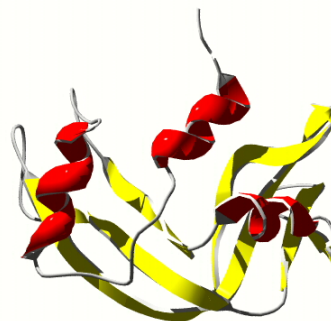
- Structure
  - Backbone; secondary, tertiary & quaternary
- Side chain groups
  - Hydrophobic, charged, polar, acidic/base, etc.
- Cross-links
  - Hydrogen bonds, disulfide bonds
- Surfaces
  - VanderWaals, solvent-accessible
- Charge distributions, distances & angles, etc.



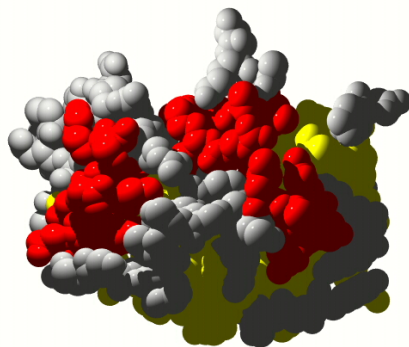
# Display Conventions



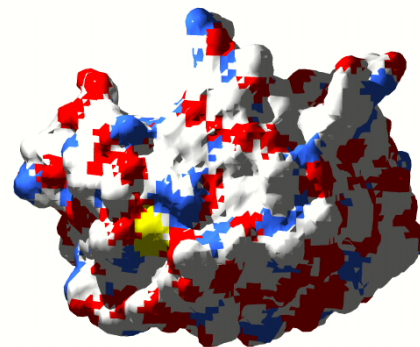
Wireframe



Ribbon



Spacefill



Molecular Surface

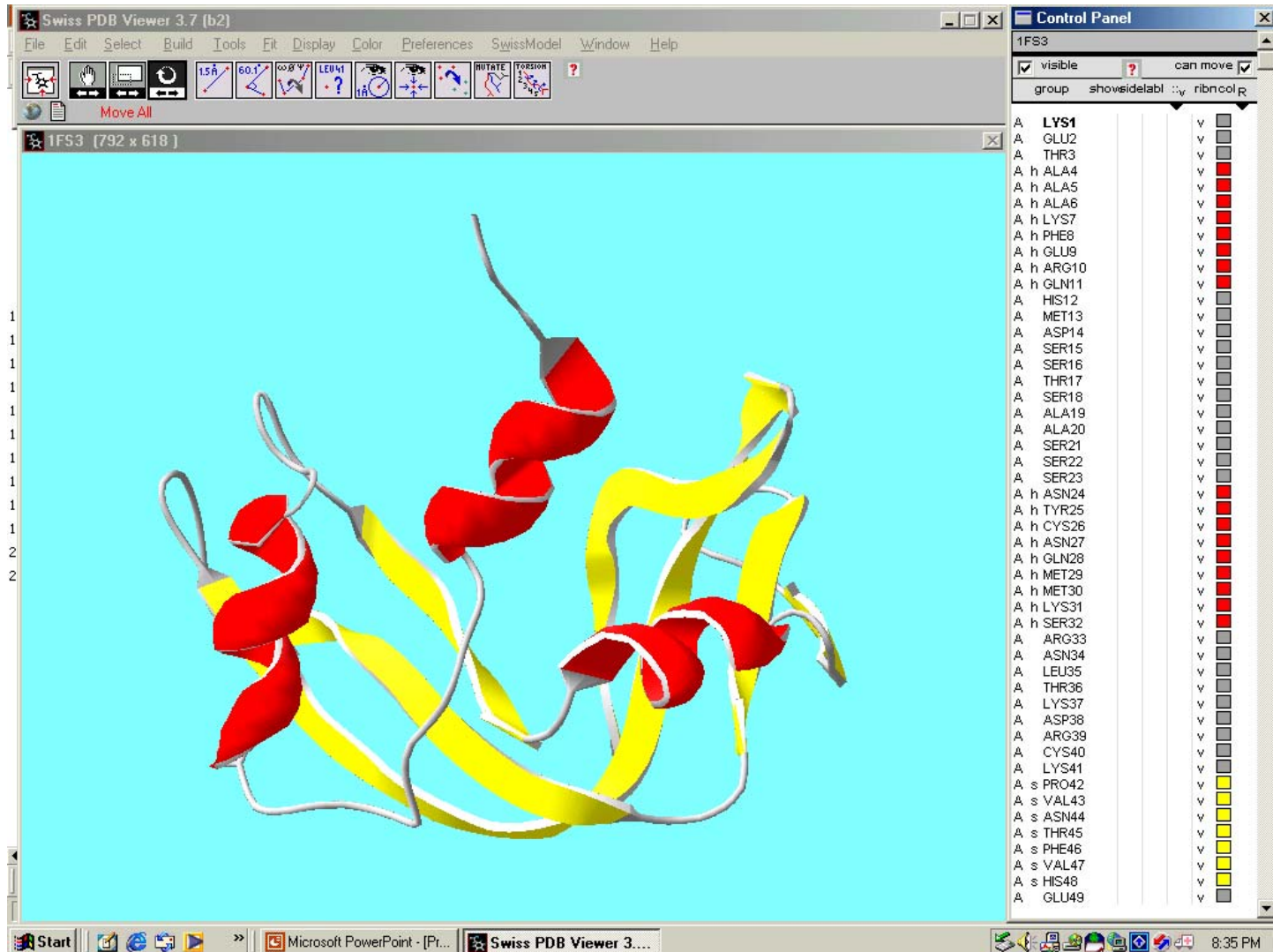
# Visualization Tools

- Viewers (free)
  - 1960's : MAGE, RasMol, Chime
  - 2003 : SwissPDB, Protein Explorer, Cn3D, etc.
- Operating systems – Unix, Windows, Mac
- Our choice (arbitrary) :
  - Chime (plug-in to NETSCAPE)
  - SwissPDB (stand-alone)

# Important URL's

- Protein Data Base
  - <http://www.rcsb.org/pdb/>
- Chime
  - [https://en.wikipedia.org/wiki/MDL\\_Chime](https://en.wikipedia.org/wiki/MDL_Chime)
- SwissPDB
  - <http://www.expasy.ch/spdbv/>
- History of Visualization of Macromolecules
  - <http://www.umass.edu/microbio/rasmol/history.htm>

# SwissPDB



# SwissPDB – Toolbar

Center

Translate

Zoom

Rotate

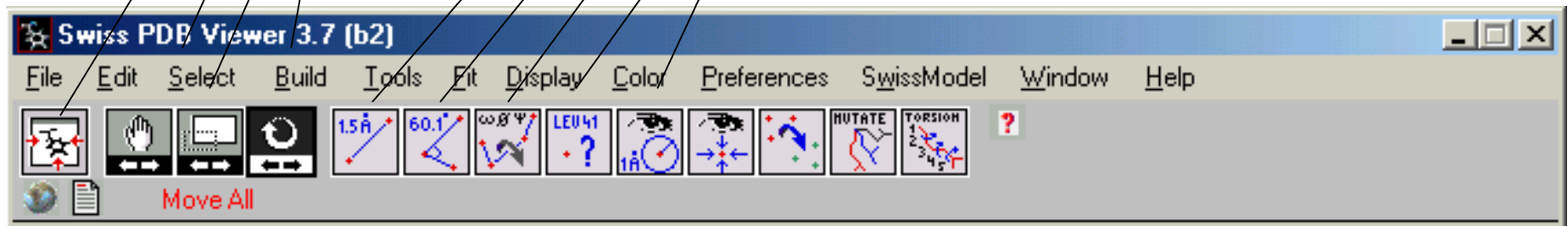
Distance between two atoms

Angle between three atoms

Measure omega, phi and psi angles

Provenance of an atom

Display groups a certain distance from an atom



# Control Panel

Control Panel

1FS3

visible ? can move

group showsidelabl :v ribncol B S

A	<b>LYS1</b>	v	v			-
A	GLU2	v	v			-
A	THR3	v	v			-
A h	ALA4	v	v			-
A h	ALA5	v	v			-
A h	ALA6	v	v			-
A	LYS7	v	v			-

Chain

Helix/sheet

Residue

Main chain

Side chain

Label

Surface Ribbon

Color target

Color

# Point of Information

- Today's lecture material is:
  - a subset of the information available to you in online tutorials
  - presented to “get you started” quickly and to “shorten the learning curve”
  - not exhaustive or even sufficient
    - => should be augmented by actually working through the online tutorials

MIT OpenCourseWare  
<http://ocw.mit.edu>

7.88J / 5.48J / 10.543J Protein Folding and Human Disease  
Spring 2015

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.