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)

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):

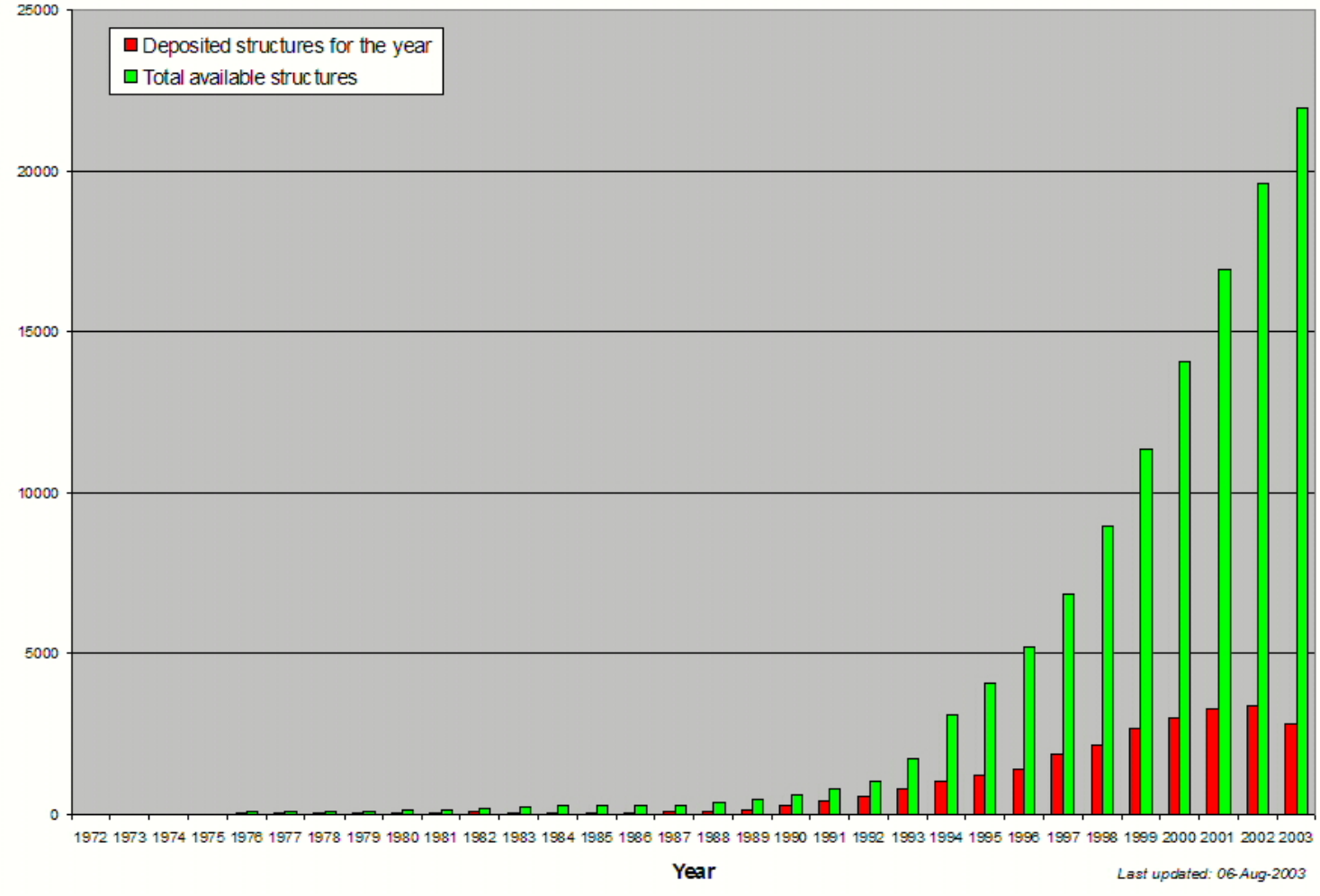
Appel, R. D., A. Bairoch, and D. F. Hochstrasser. "A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. Trends." *Biochem. Sci.* 19 (1994): 258-260.

(ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) 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


Deposited structures for the year
Total available structures
25000
20000
15000
10000
5000
0
1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003
Year
Last updated: 06-Aug-2003

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

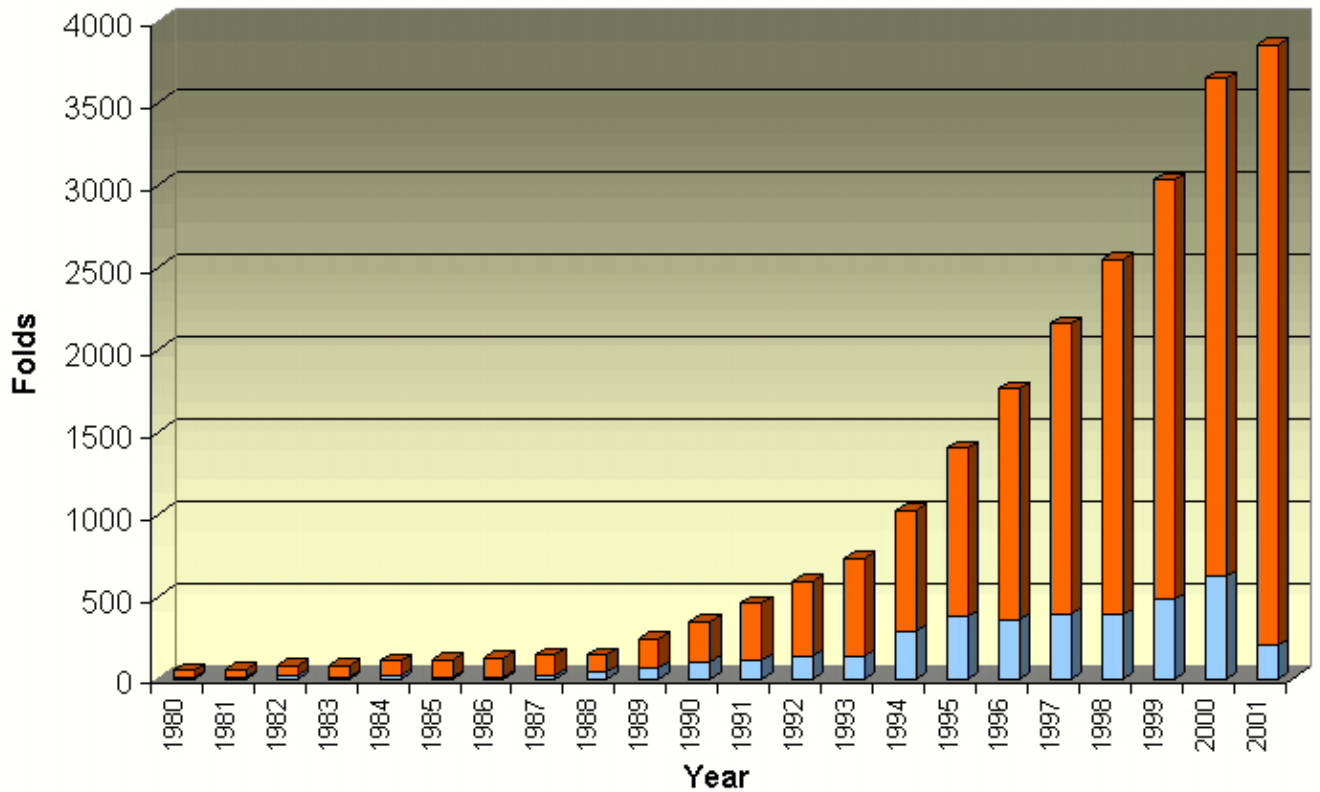
Why is the PDB Important?

- "Collective Leverage"

  for …
- Understanding molecular machinery
- Rational drug design
- Engineering new molecules
- Structural genomics

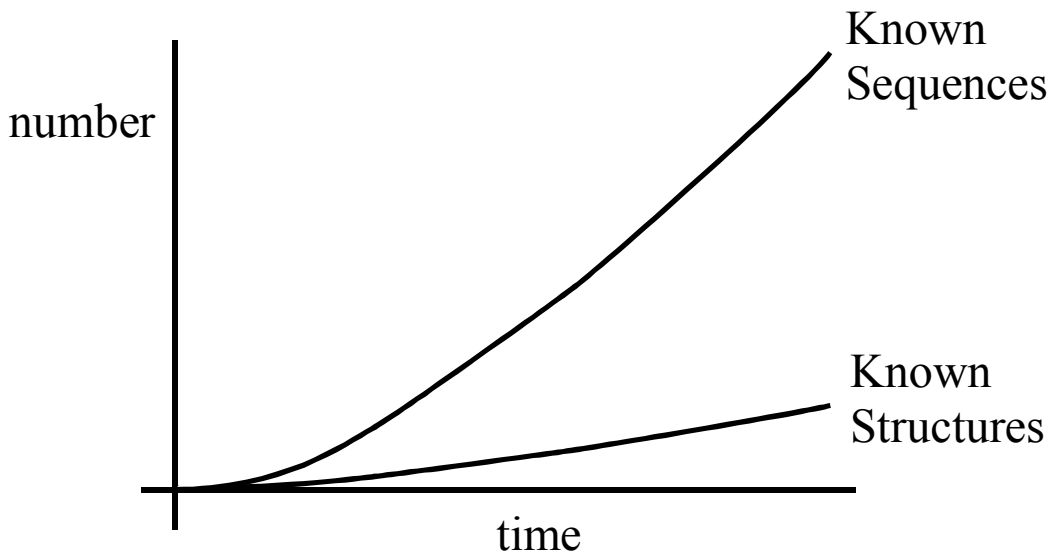
Not all Structures are Different

- PDB Growth in "New Folds"



Structure vs Sequence

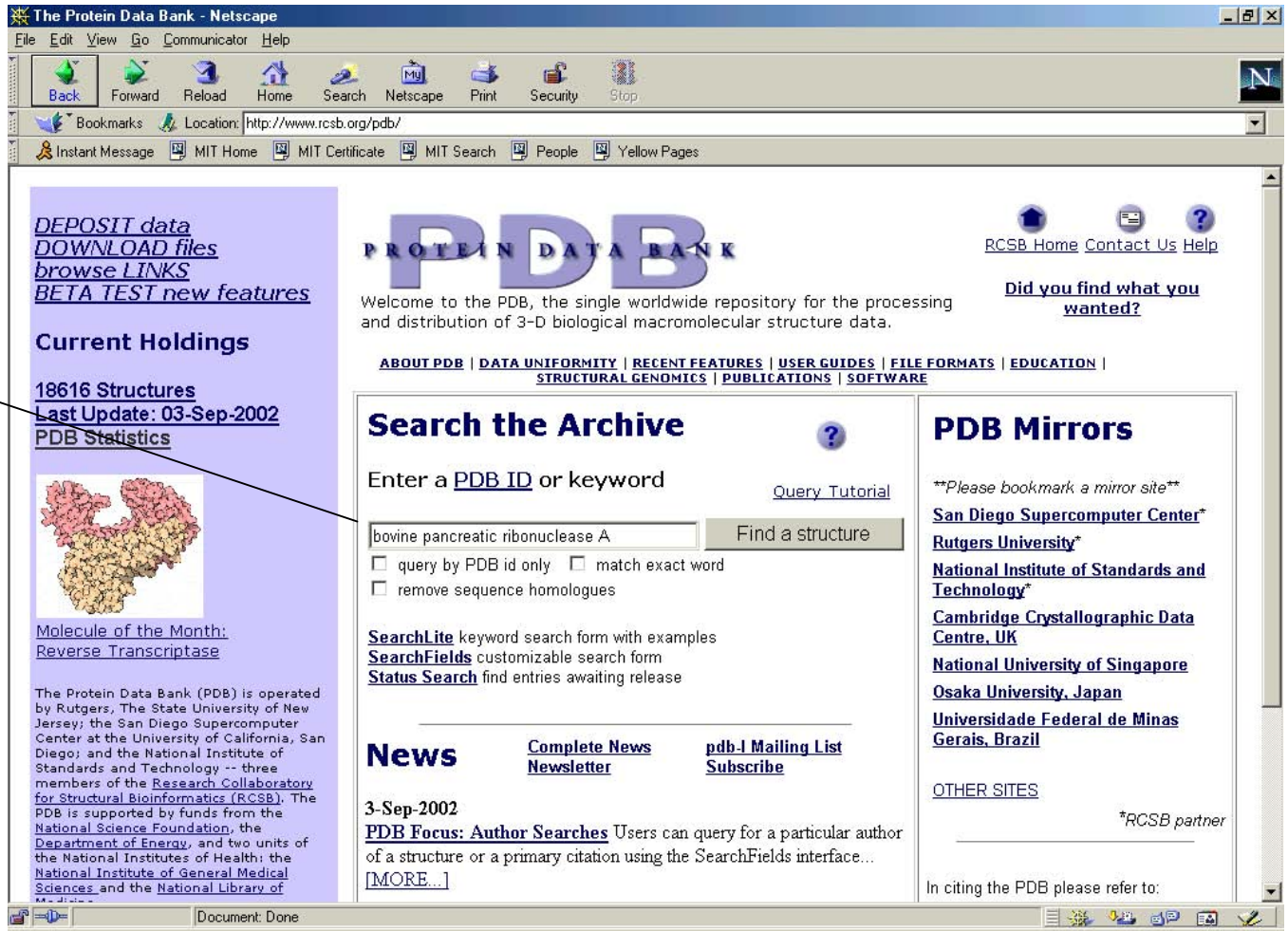
- New protein structures are being solved more slowly than new protein sequences are being solved
  - Currently, known protein sequences outnumber known protein structures
  - The “sequence-structure” gap continues to widen



PDB Website

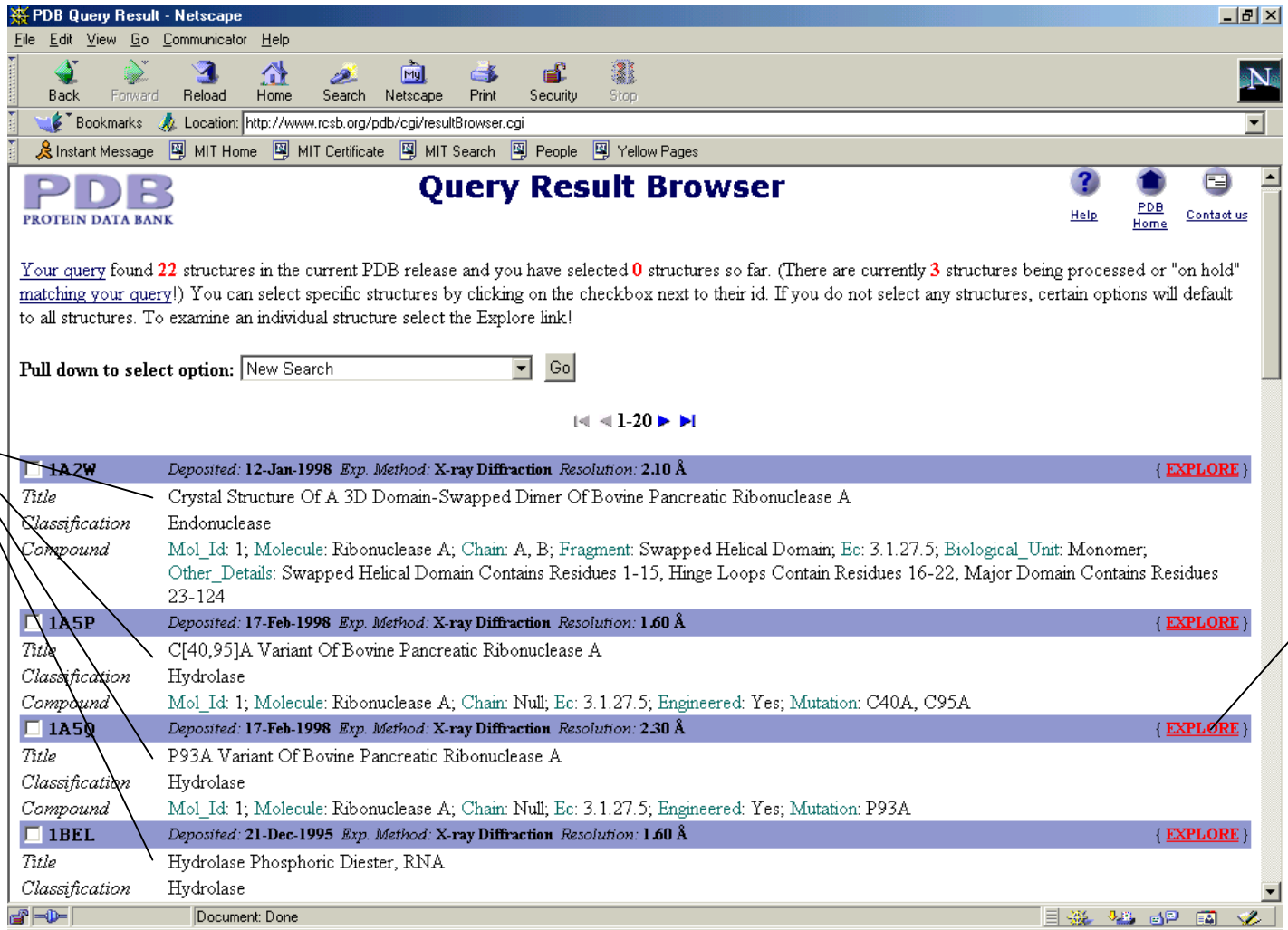
http://www.rcsb.org/pdb/

Enter what you know…



Query Result Browser

Which one do I want?



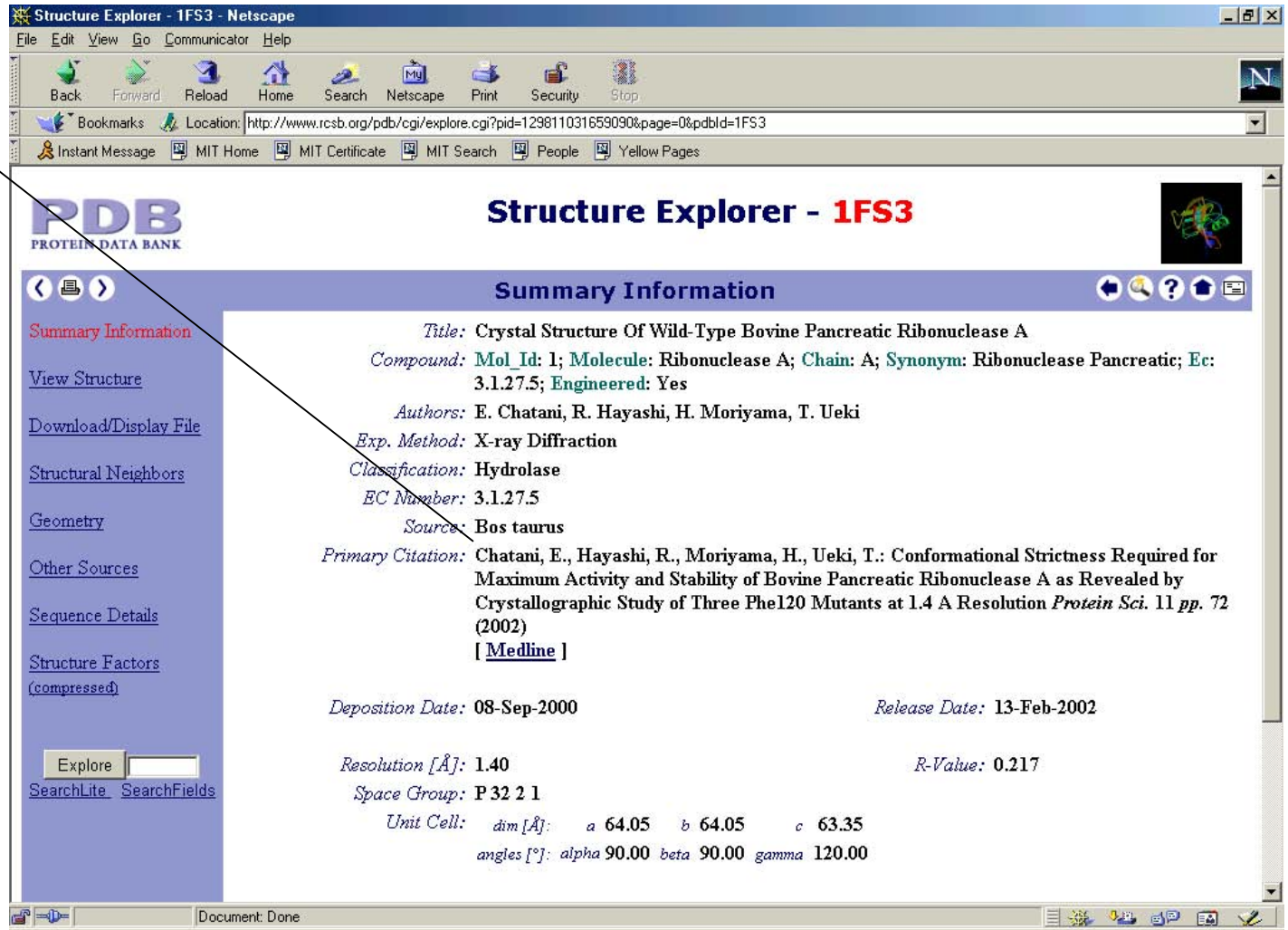
Let's look at this one …

Structure Explorer

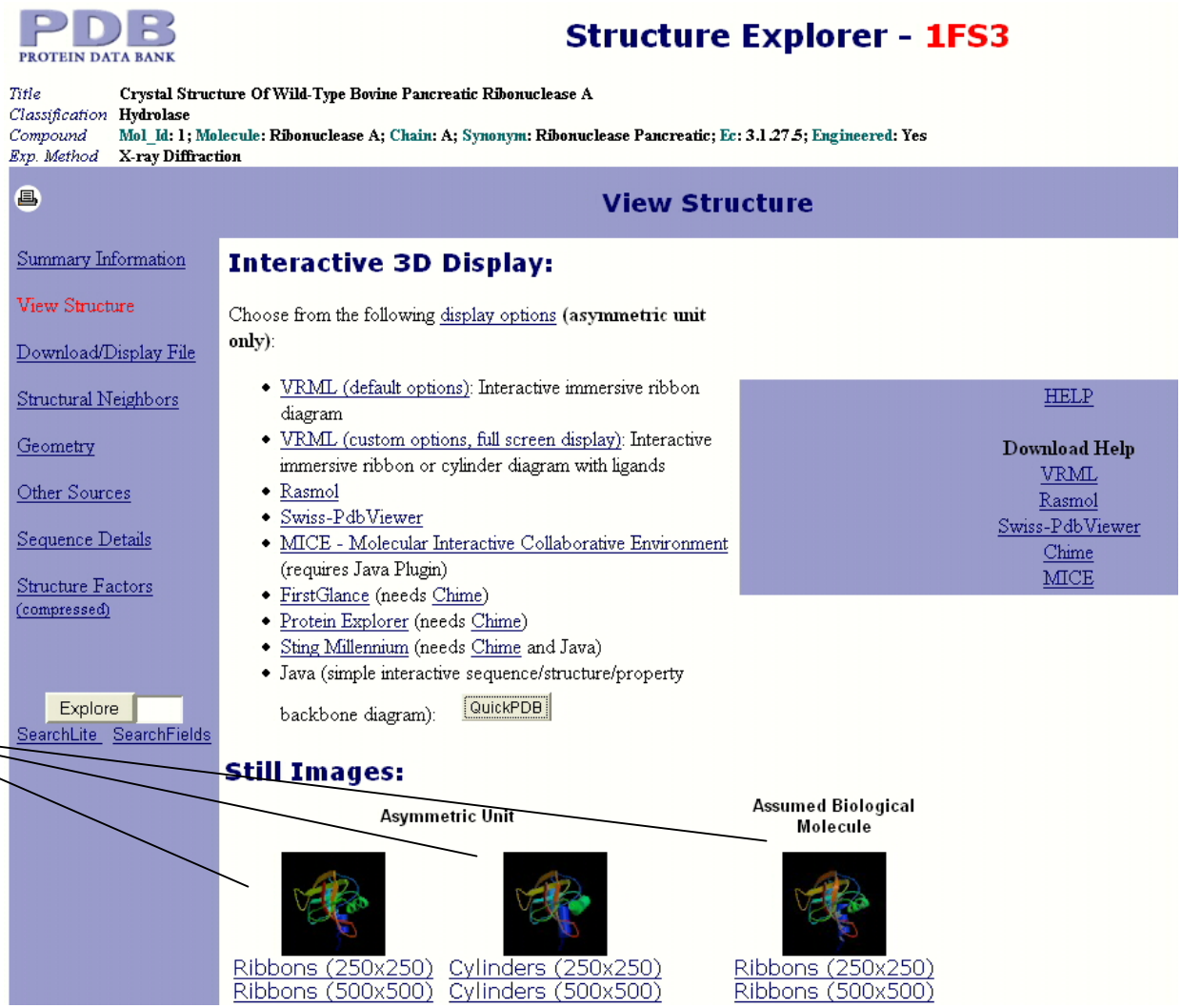
Yep, that's the right one…

View it…

Download it…



View Structure

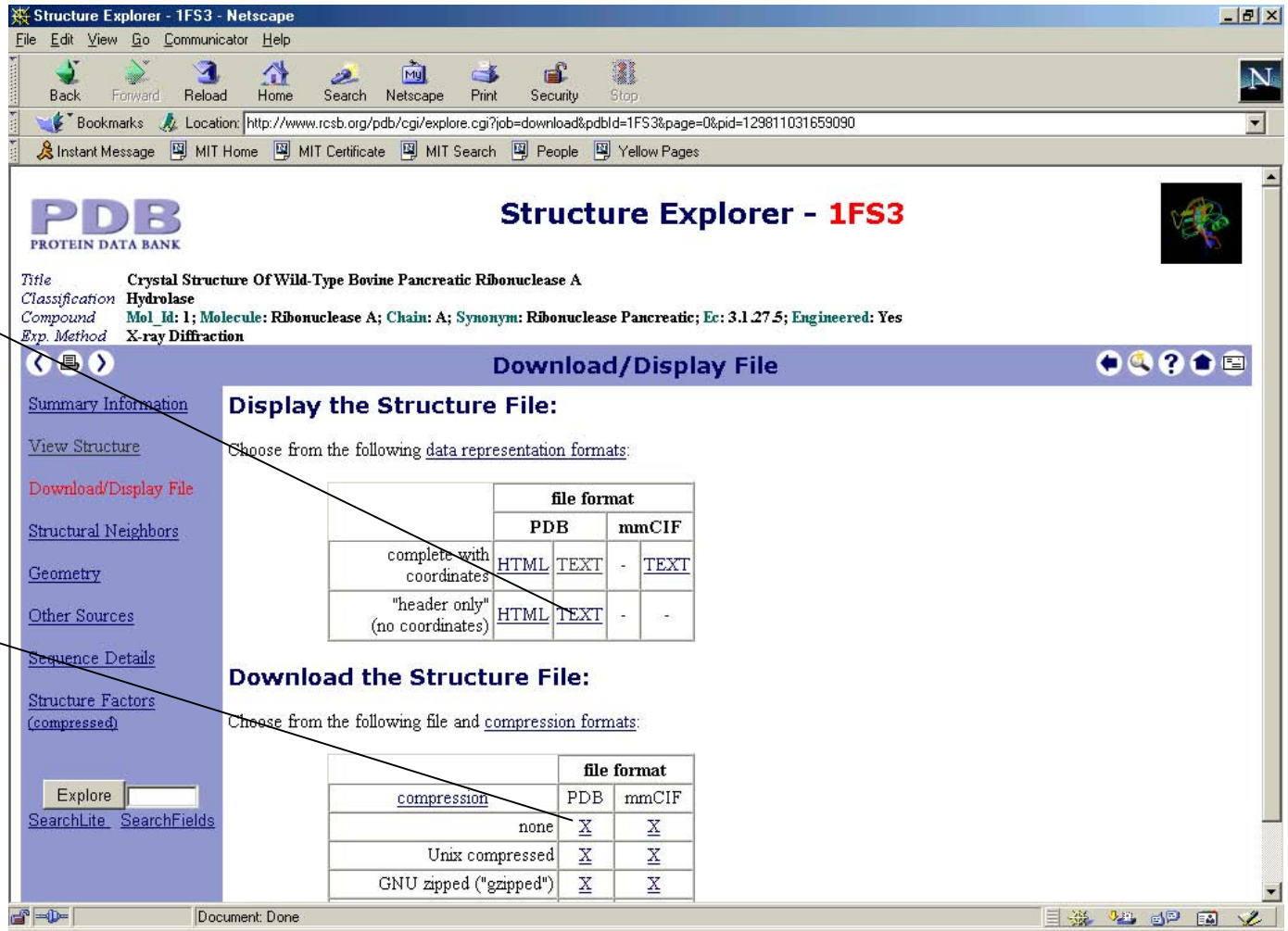


Static Images

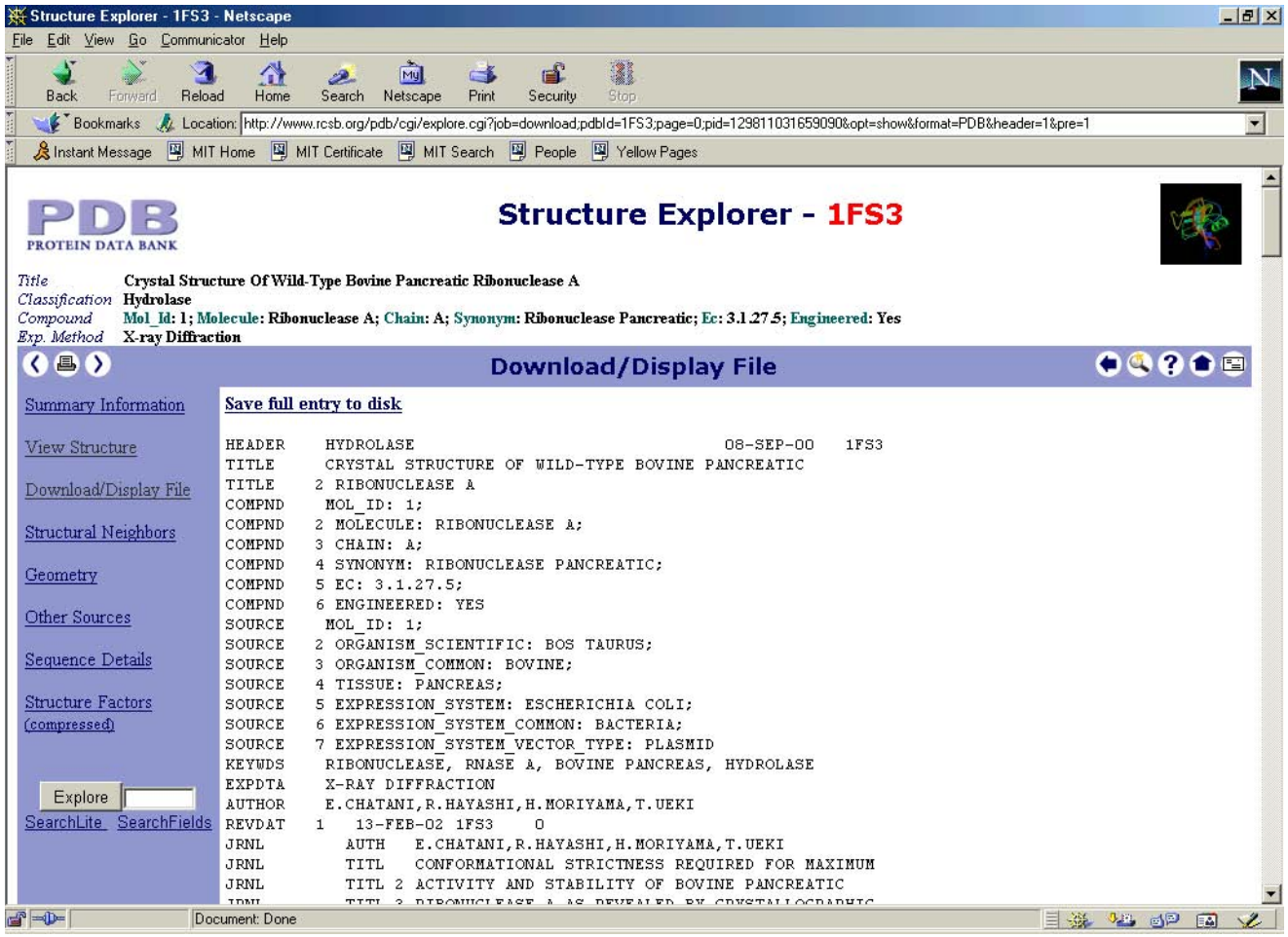
Download/Display

Display the file header…

Download the file… (Select this file format)



Header Information



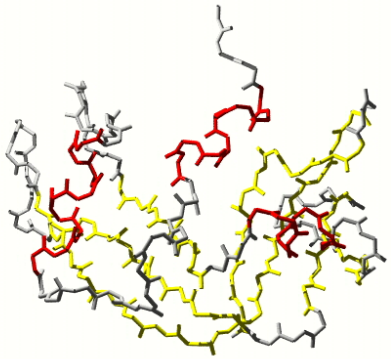
Visualizing Proteins

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

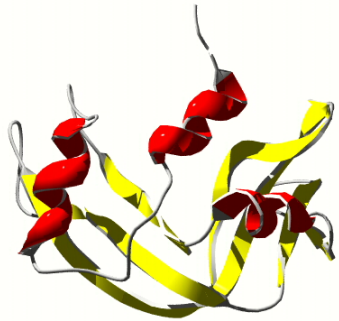
Visualization Objectives

- 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.

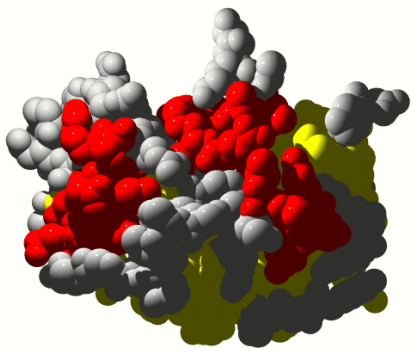
Display Conventions



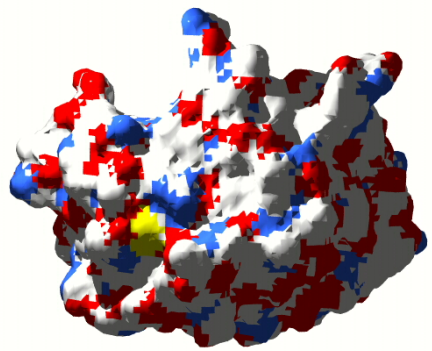
Wireframe



Ribbon



Spacefill



Molecular Surface

Visualization Tools

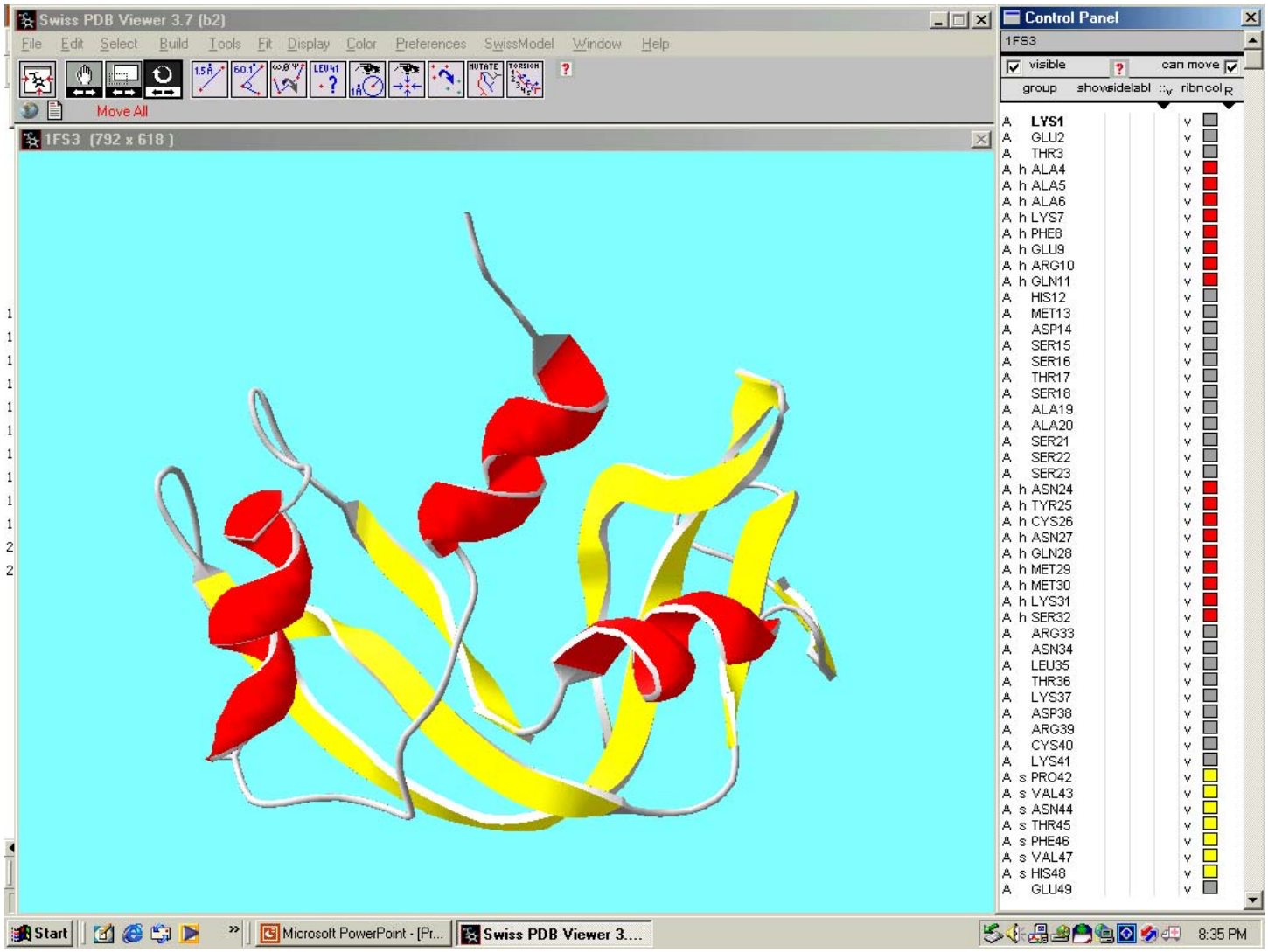
- Viewers (free)
  - 1960's : MAGE, RasMol, Chime
  - 2003 : SwissPDB, Protein Explorer, Cn3D, etc.
- Operating systems – Unix, Windows, Mac

- Our choice (arbitrary) :
  - Chime (plug-in to NETSCAPE)
  - SwissPDB (stand-alone)

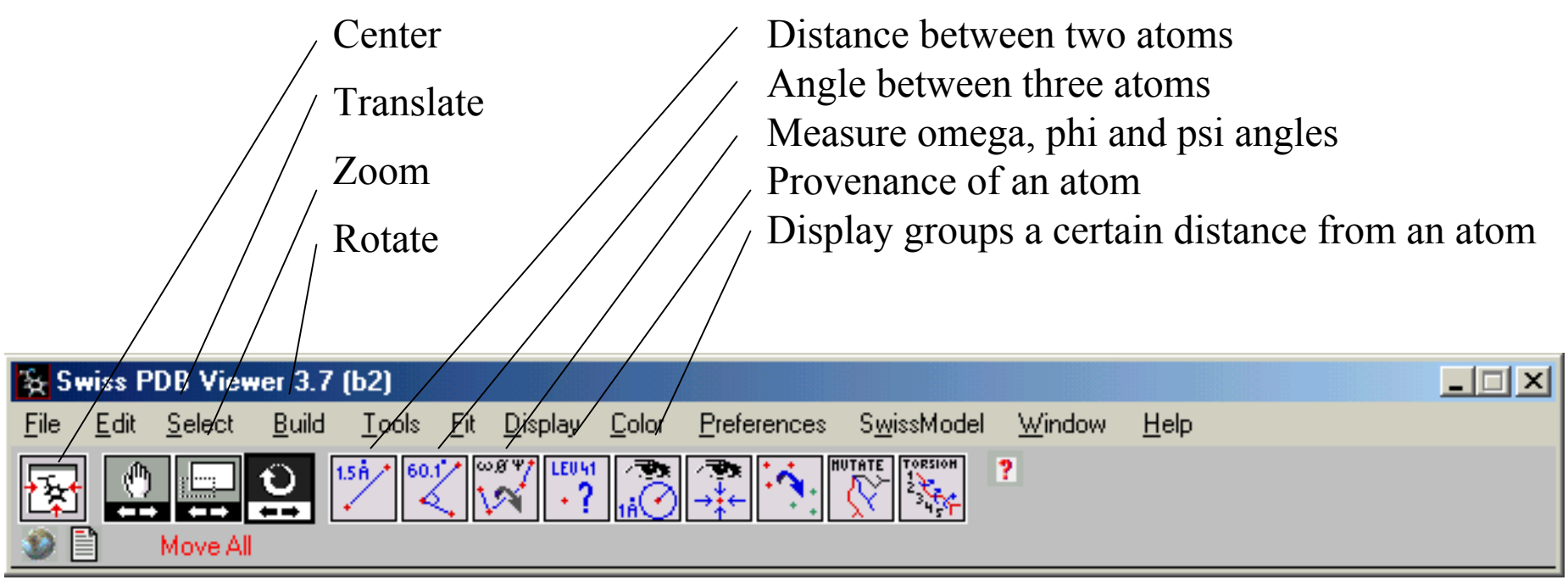
Important URL's

- Protein Data Base
  - http://www.rcsb.org/pdb/
- Chime
  - https://en.wikipedia.org/wiki/MDL_Chime
- SwissPDB
  - http://www.expasy.ch/spdbv/
- History of Visualization of Macromolecules
  - http://www.umass.edu/microbio/rasmol/history.htm

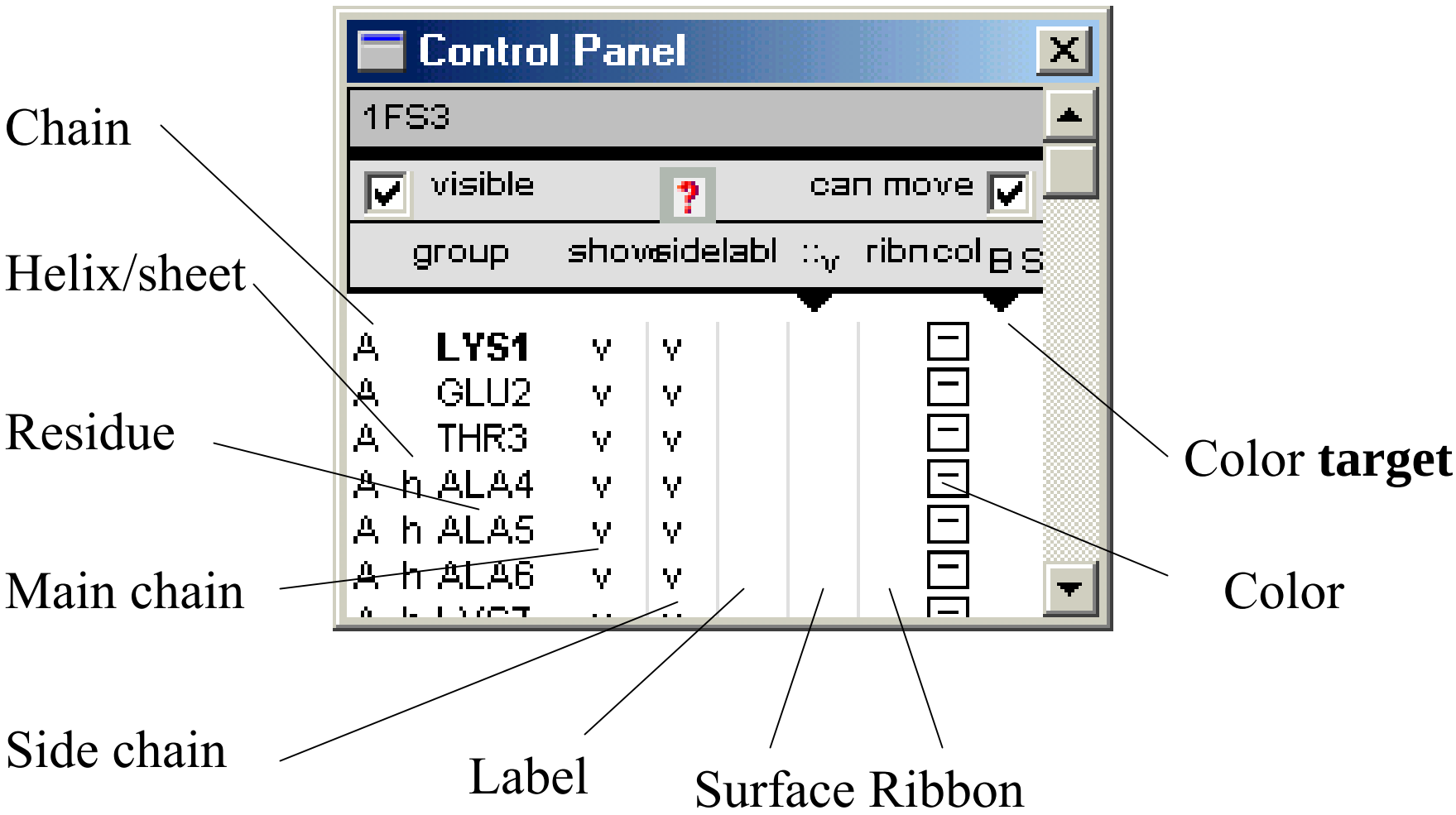
SwissPDB


Swiss PDB Viewer 3.7 (b2)
File Edit Select Build Tools Fit Display Color Preferences SwissModel Window Help
Move All
1FS3 (792 x 618)
Control Panel
1FS3
visible
can move
group showsidelabl ribncol
A LYS1
A GLU2
A THR3
A h ALA4
A h ALA5
A h ALA6
A h LYS7
A h PHE8
A h GLU9
A h ARG10
A h GLN11
A HIS12
A MET13
A ASP14
A SER15
A SER16
A THR17
A SER18
A ALA19
A ALA20
A SER21
A SER22
A SER23
A h ASN24
A h TYR25
A h CYS26
A h ASN27
A h GLN28
A h MET29
A h MET30
A h LYS31
A h SER32
A ARG33
A ASN34
A LEU35
A THR36
A LYS37
A ASP38
A ARG39
A CYS40
A LYS41
A s PRO42
A s VAL43
A s ASN44
A s THR45
A s PHE46
A s VAL47
A s HIS48
A GLU49
Start
Microsoft PowerPoint - [Pr...
Swiss PDB Viewer 3....
8:35 PM

SwissPDB – Toolbar



Control Panel



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

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