BCHS 6229
Protein Structure and Function

Lecture 1 (October 11, 2011)

Introduction

Basic Structural Principles
PDB

Overview

Main Goals:

- Carry out a rapid review of the essentials of protein structure & function
- Provide a basis for evaluating current structural biology literature
- Cover selected important topics in protein science
- Include literature as much as possible

Suggested Textbooks:

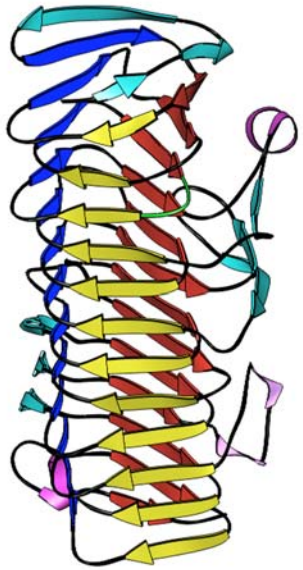
- *Introduction to* ***Protein Structure****, 2nd Ed.*, Branden and Tooze, 1999
- ***Protein Structure*** *and* ***Function***, Petsko and Ringe, 2004
- *Fundamentals of Biochemistry, Voet, 2nd Ed, 2005*

Grade:

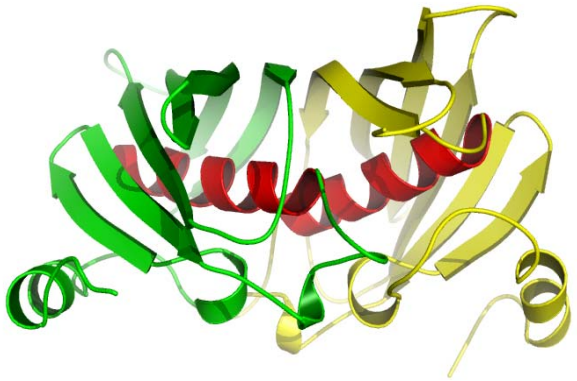
- One short in-class exam (**20** %)
- Two homework assignments (**40** %)
- One project presentation (**40** %)

Yeo Laboratory – Research Interests

- **Type V Secretion in *Haemophilus influenzae***



- **Virulence factors of *Campylobacter jejuni***

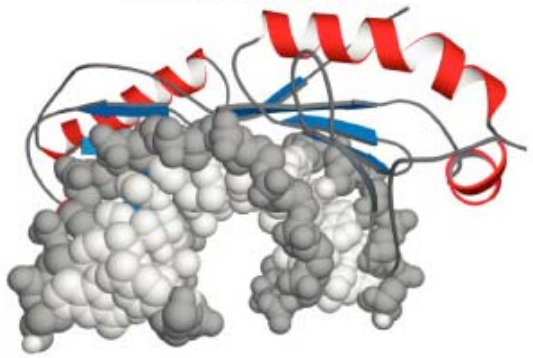


Overview: Basic Structural Principles

Examples of biochemical functions performed by proteins (I)

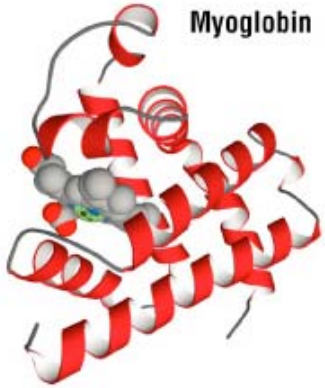
Binding

TATA binding protein



The TATA binding protein binds a specific DNA sequence and serves as the platform for a complex that initiates transcription of genetic information. (PDB 1tgh)

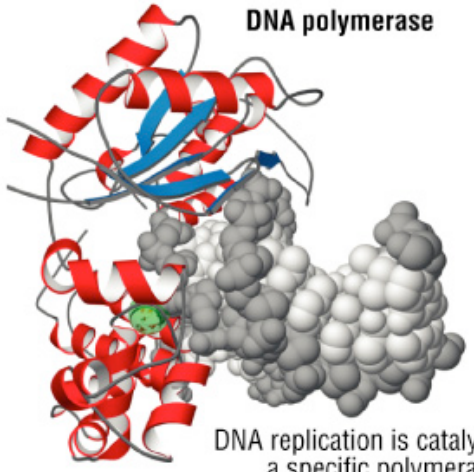
Myoglobin



Myoglobin binds a molecule of oxygen reversibly to the iron atom in its heme group (shown in grey with the iron in green). It stores oxygen for use in muscle tissues. (PDB 1a6k)

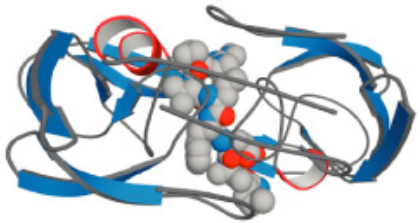
Catalysis

DNA polymerase



DNA replication is catalyzed by a specific polymerase that copies the genetic material and edits the product for errors in the copy. (PDB 1pbx)

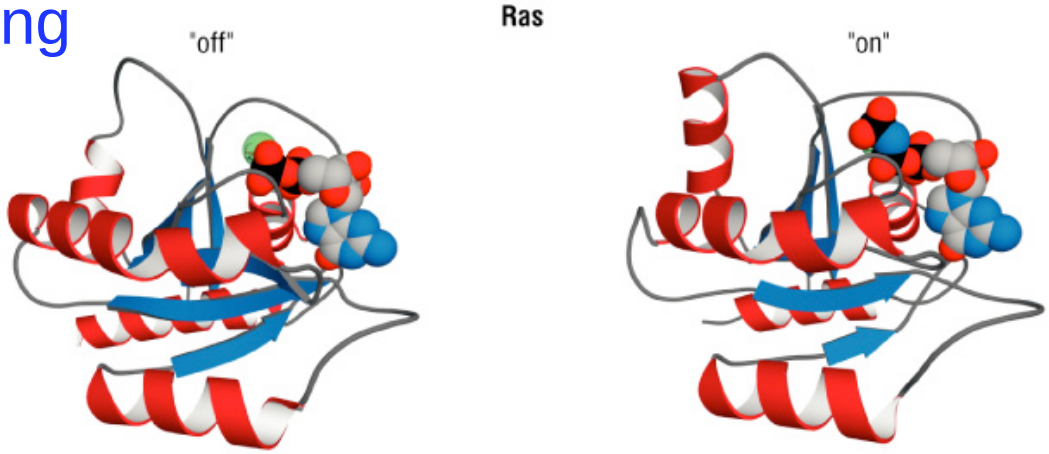
HIV protease



Replication of the AIDS virus HIV depends on the action of a protein-cleaving enzyme called HIV protease. This enzyme is the target for protease-inhibitor drugs (shown in grey). (PDB 1a8k)

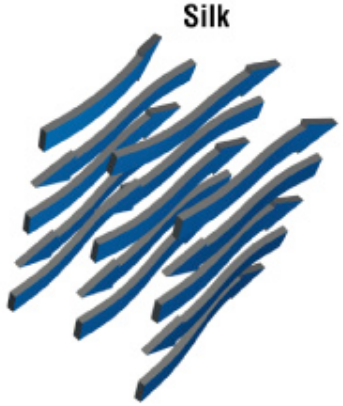
Examples of biochemical functions performed by proteins (II)

Switching

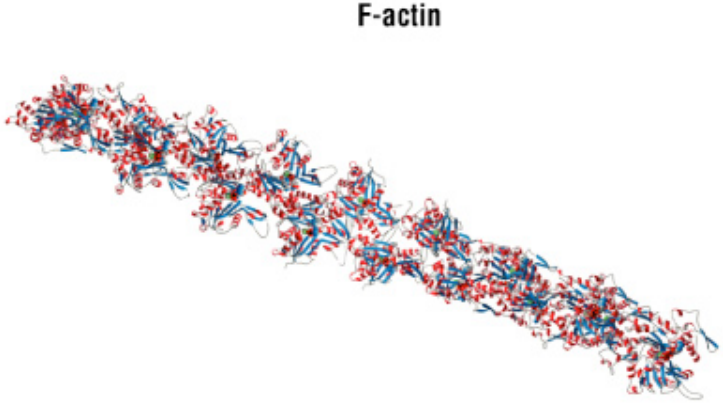


The GDP-bound ("off"; PDB 1pll) state of Ras differs significantly from the GTP-bound ("on"; PDB 121p) state. This difference causes the two states to be recognized by different proteins in signal transduction pathways.

Structural Proteins

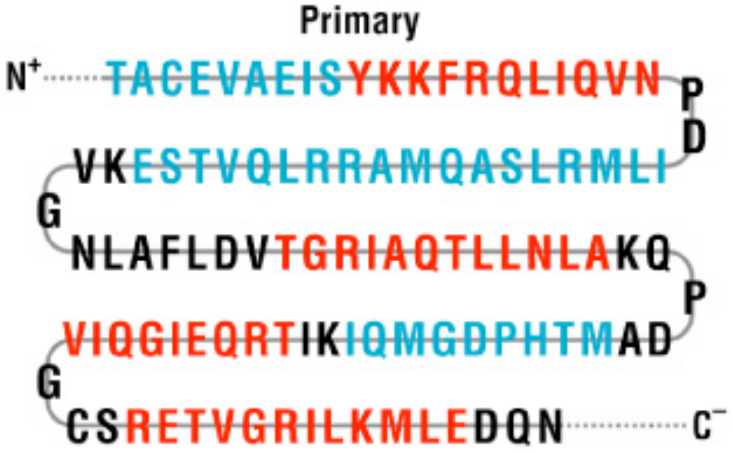


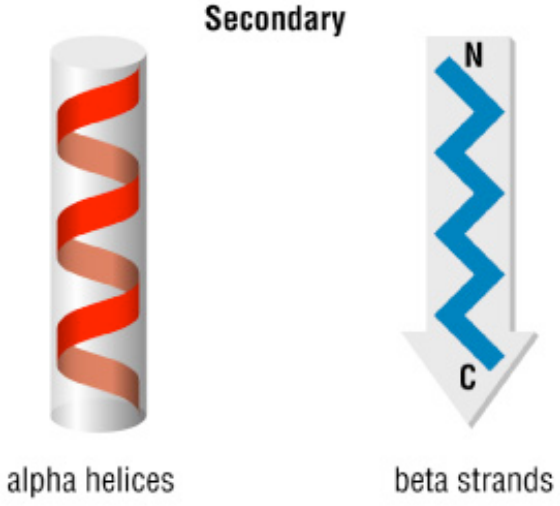
Silk derives its strength and flexibility from its structure: it is a giant stack of antiparallel beta sheets. Its strength comes from the covalent and hydrogen bonds within each sheet; the flexibility from the van der Waals interactions that hold the sheets together. (PDB 1slk)

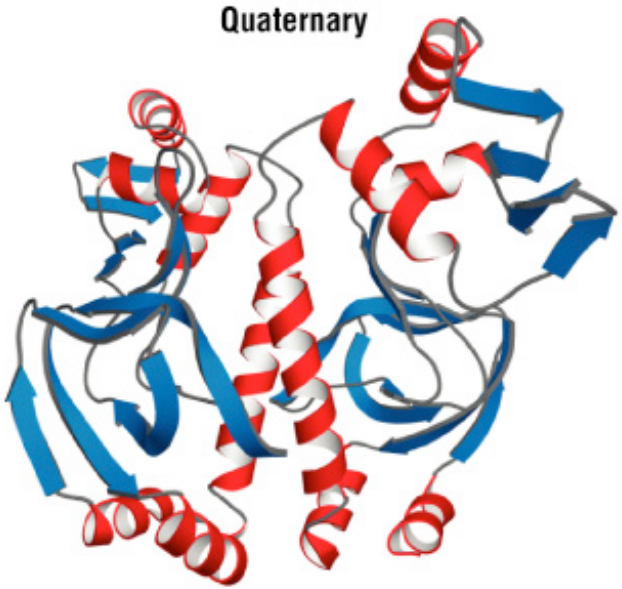


Actin fibers are important for muscle contraction and for the cytoskeleton. They are helical assemblies of actin and actin-associated proteins. (Courtesy of Ken Holmes)

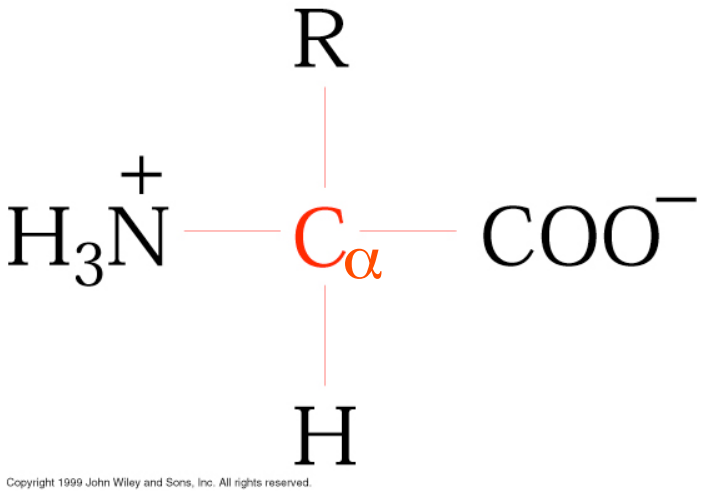
There are four levels of protein structure







Amino acids



Please carefully note the charged ends (termini). At physiological pH, the ends of an amino acid are charged. Certain "R" groups will also be charged at pH 7.

- Backbone of an amino acid is composed of the N, Cα, and C
- Amino acid structures and sequences are written from left to right, starting with the N-terminus (amino) and finishing with the C-terminus (carboxyl)
- The thing that differentiates each amino acid is the "R" group
- Cα is chiral, except in Gly

General properties

- The backbone of individual amino acids are zwitterionic (i.e. has both a positively charged and a negatively charged group)
- In addition, some amino acids have ionizable (i.e. charged) side chains
- Because of these ionizable groups (backbone and some side chains), amino acids can have a number of different charge states
- The “R” group in an amino acid is called the side chain
- An amino acid is often called a “residue” (i.e. an amino acid residue)
- There are 20 standard amino acids - they all differ in “R”

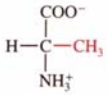
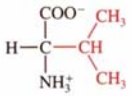
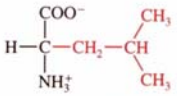
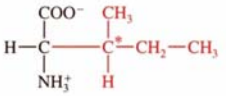
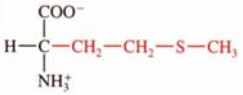
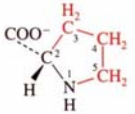
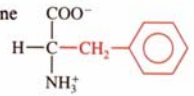
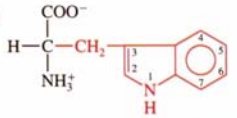
Classification

- Non-polar (9 aa)
  - Glycine (Gly, G), Alanine (Ala, A), Valine (Val, V), Leucine (Leu, L), Isoleucine (Ile, I), Methionine (Met, M), Proline (Pro, P), Phenylalanine (Phe, F), Tryptophan (Trp, W)

- Polar (6 aa)
  - Serine (Ser, S), Threonine (Thr, T), Asparagine (Asn, N), Glutamine (Gln, Q), Tyrosine (Tyr, Y), Cysteine (Cys, C)

- Charged (5 aa)
  - Aspartic acid (Asp, D, -1); Glutamic acid (Glu, E, -1)
  - Lysine (Lys, K, +1); Arginine (Arg, R, +1), Histidine (His, H, +1)

TABLE 4-1. COVALENT STRUCTURES AND ABBREVIATIONS OF THE "STANDARD" AMINO ACIDS OF PROTEINS, THEIR OCCURANCE, AND THE p*K* VALUES OF THEIR IONIZING GROUPS

Name, Three-letter Symbol, and One-letter Symbol	Structural Formula[a]	Residue Mass (D)[b]	Average Occurence in Proteins (%)[c]	pK_1 α-COOH[d]	pK_2 α-NH_3^+ [d]	pK_R Side chain[d]
Amino acids with nonpolar side chains						
Glycine Gly G		57.0	7.2	2.35	9.78	
Alanine Ala A		71.1	7.8	2.35	9.87	
Valine Val V		99.1	6.6	2.29	9.74	
Leucine Leu L		113.2	9.1	2.33	9.74	
Isoleucine Ile I		113.2	5.3	2.32	9.76	
Methionine Met M		131.2	2.2	2.13	9.28	
Proline Pro P		97.1	5.2	1.95	10.64	
Phenylalanine Phe F		147.2	3.9	2.20	9.31	
Tryptophan Trp W		186.2	1.4	2.46	9.41	

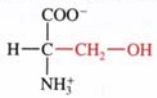
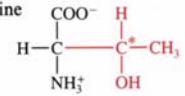
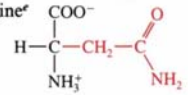
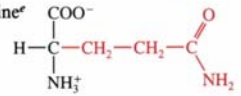
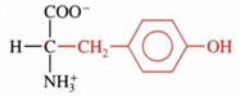
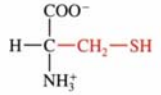
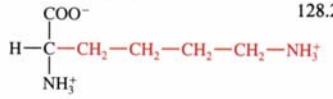
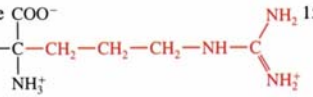
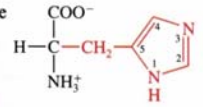
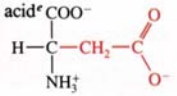
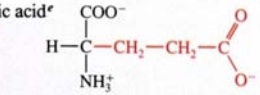
[a] The ionic forms shown are those predominating at pH 7.0 although residue mass is given for the neutral compound. The C_α atoms, as well as those atoms marked with an asterisk, are chiral centers with configurations as indicated according to Fischer projection formulas. The standard organic numbering system is provided for heterocycles.

[b] The residue masses are given for the neutral residues. For the molecular masses of the parent amino acids, add 18.0 D, the molecular mass of H_2O, to the residue masses. For side chain masses, subtract 56.0 D, the formula mass of a peptide group, from the residue masses.

[c] Calculated from a database of nonredundant proteins containing 300,688 residues as compiled by Doolittle, R. F. *in* Fasman, G, D. (Ed.), *Predictions of Protein Structure and the Principles of Protein Conformation*, Plenum Press (1989).

[d] *Source:* Dawson, R.M.C., Elliott, D.C., Elliott, W.H. and Jones, K.M., *Data for Biochemical Research* (3rd ed.), *pp.* 1–31, Oxford Science Publications (1986).

[e] The three- and one-letter symbols for asparagine *or* aspartic acid are Asx and B, whereas for glutamine *or* glutamic acid they are Glx and Z. The one-letter symbol for an undetermined or "nonstandard" amino acid is X.

Name, Three-letter Symbol, and One-letter Symbol	Structural Formula[a]	Residue Mass (D)[b]	Average Occurence in Proteins (%)[c]	pK_1 α-COOH[d]	pK_2 α-NH_3^+ [d]	pK_R Side chain[d]
Amino acids with uncharged polar side chains						
Serine Ser S		87.1	6.8	2.19	9.21	
Threonine Thr T		101.1	5.9	2.09	9.10	
Asparagine[e] Asn N		114.1	4.3	2.14	8.72	
Glutamine[e] Gln Q		128.1	4.3	2.17	9.13	
Tyrosine Tyr Y		163.2	3.2	2.20	9.21	10.46 (phenol)
Cysteine Cys C		103.1	1.9	1.92	10.70	8.37 (sulfhydryl)
Amino acids with charged polar side chains						
Lysine Lys K		128.2	5.9	2.16	9.06	10.54 (ε-NH_3^+)
Arginine Arg R		156.2	5.1	1.82	8.99	12.48 (guanidino)
Histidine His H		137.1	2.3	1.80	9.33	6.04 (imidazole)
Aspartic acid[e] Asp D		115.1	5.3	1.99	9.90	3.90 (β-COOH)
Glutamic acid[e] Glu E		129.1	6.3	2.10	9.47	4.07 (γ-COOH)

Amino acid structure and the chemical characters of the amino acid side chains

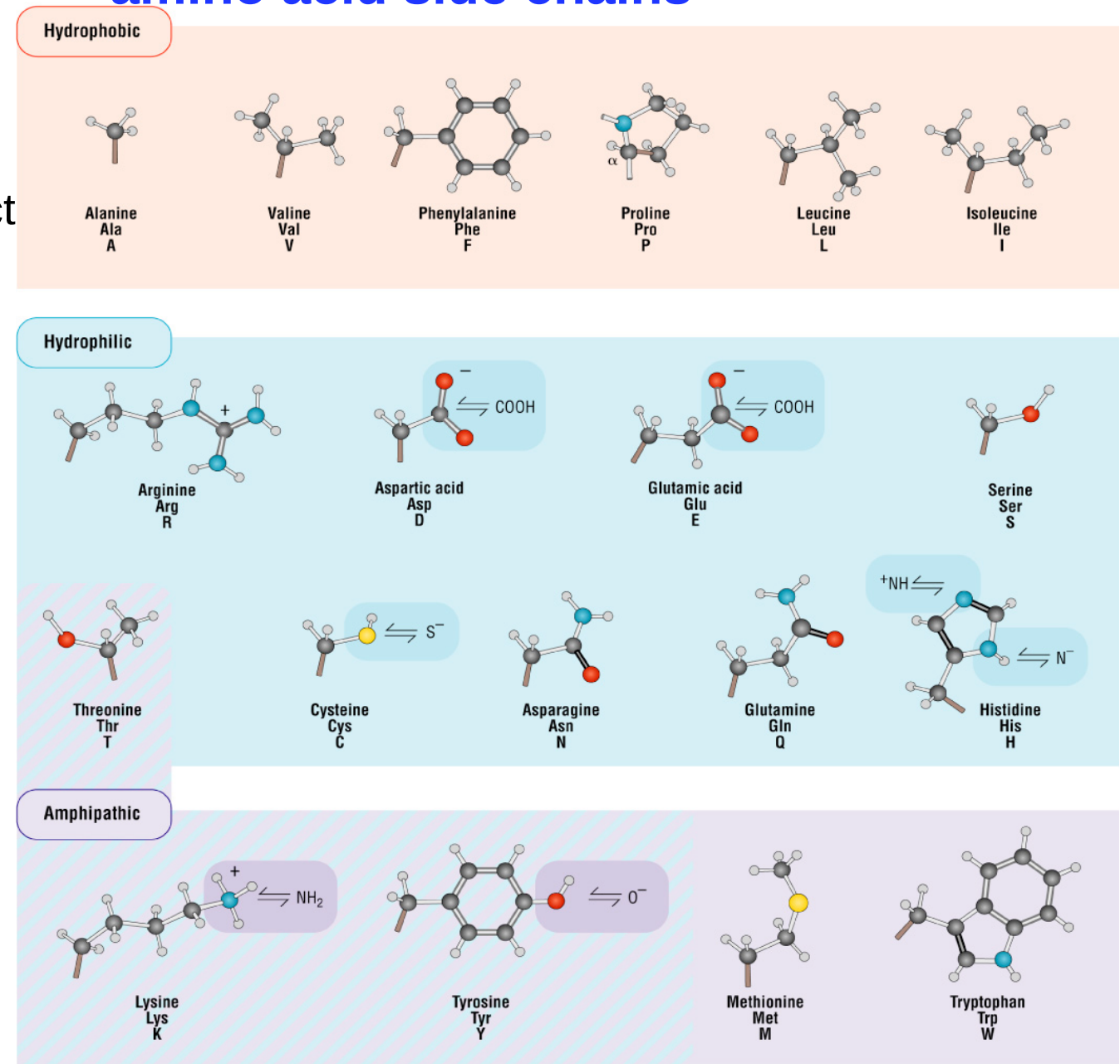
van der Waals

Hydrophobic effect

H - bond

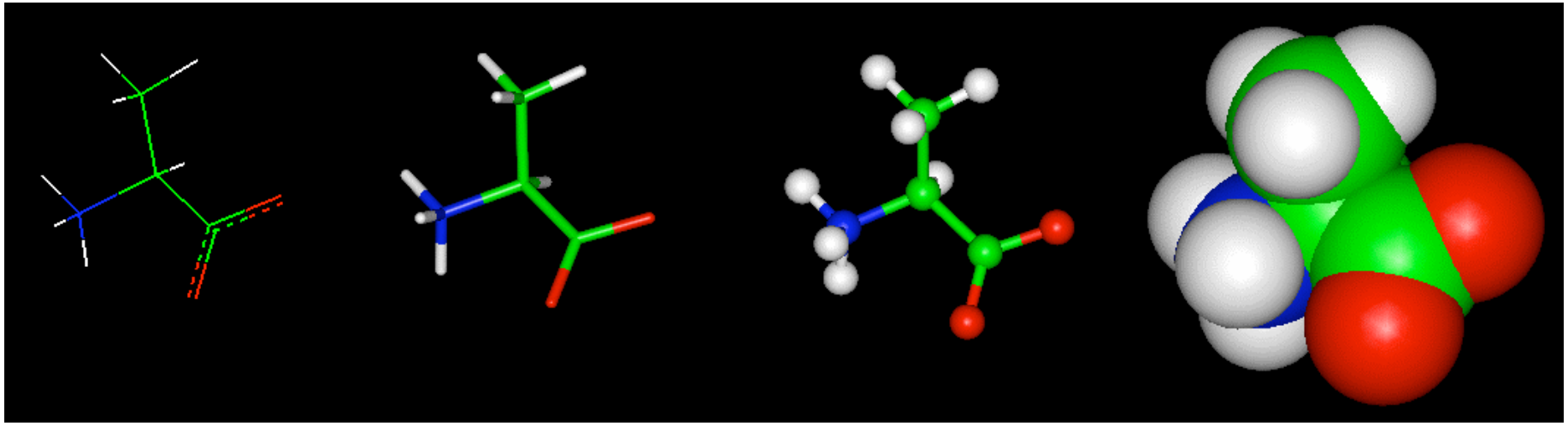
Polar & non-polar

interface



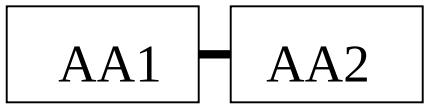
Amino acids

- The amino acid, Alanine (Ala, A) is shown below in line, stick, ball and stick, and CPK (space filling) representations.



Linear arrays (polymers) of amino acids can make a huge number of molecules

Consider a peptide with two amino acids;
There are 20 possibilities at each site



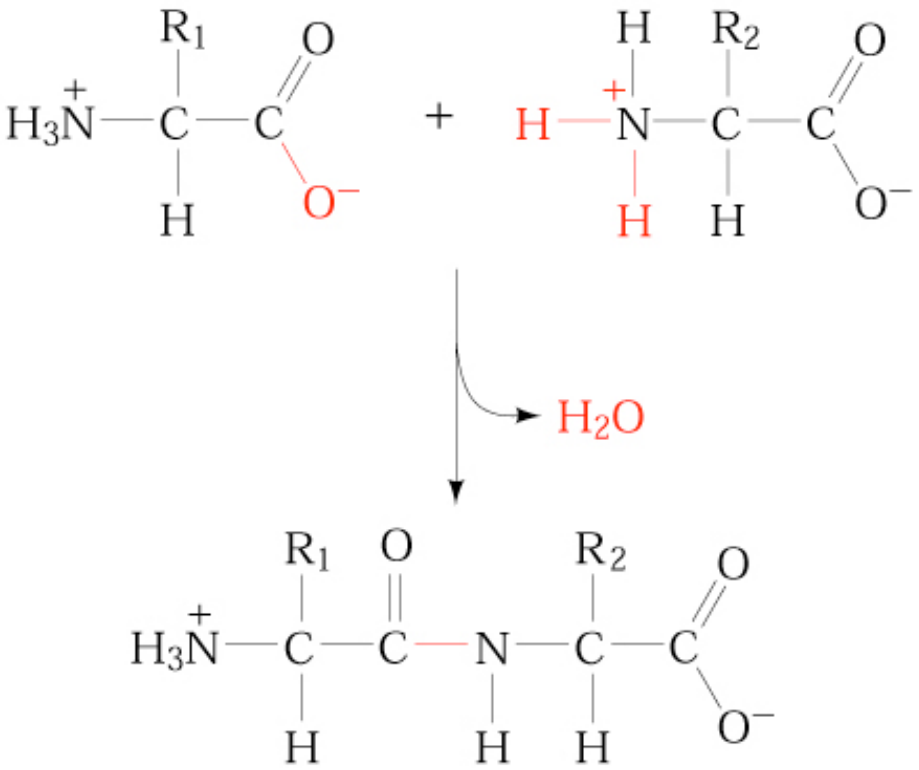
20 x 20 = 400 different molecules

20 x 20 x 20 = 8000 different molecules

For 100 amino acid protein the # of possibilities are:

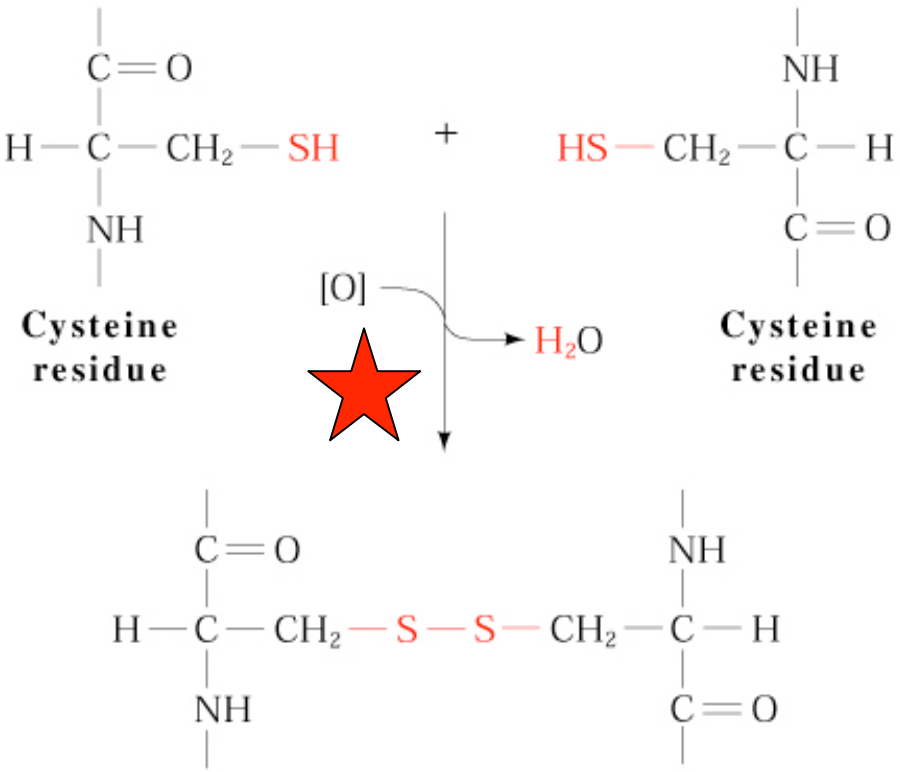
$$20^{100} = 1.27x10^{130}$$

Peptide bonds



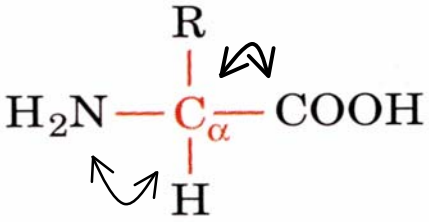
- As mentioned previously, amino acids can be connected together (i.e. condensed) to form a bigger molecule, now containing two amino acids
- The bond formed is a “peptide bond” and the molecule is a dipeptide.
- If we add another amino acid, then we would have a tripeptide

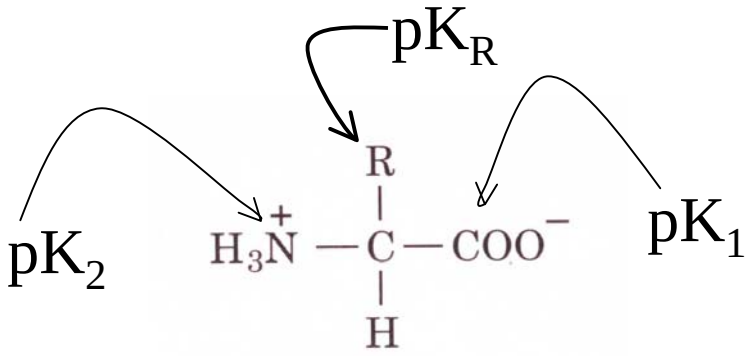
Disulfide bond formation



- Amino acids in a poly-peptide chain can also be cross-linked via two Cys residues
- Cys residues have “SH” groups at the end of their side chains. Two of these groups can be oxidized to form an S-S (disulfide) bond.
- Disulfide bonds can provide stability to a protein structure

Amino Acids:
The building blocks of proteins





α-amino acids because of the α-carboxylic and α-amino groups

pK_1 and pK_2 respectively pK_R is for R group pK's

$pK_1 \approx 2.2$ while $pK_2 \approx 9.4$

Remember these values for the pKa's of the termini for ALL AA's

In the physiological pH range, both carboxylic and amino groups are completely ionized!!

Hint: draw the structures of an amino acid at several pH values

Acid - Base properties of amino acids

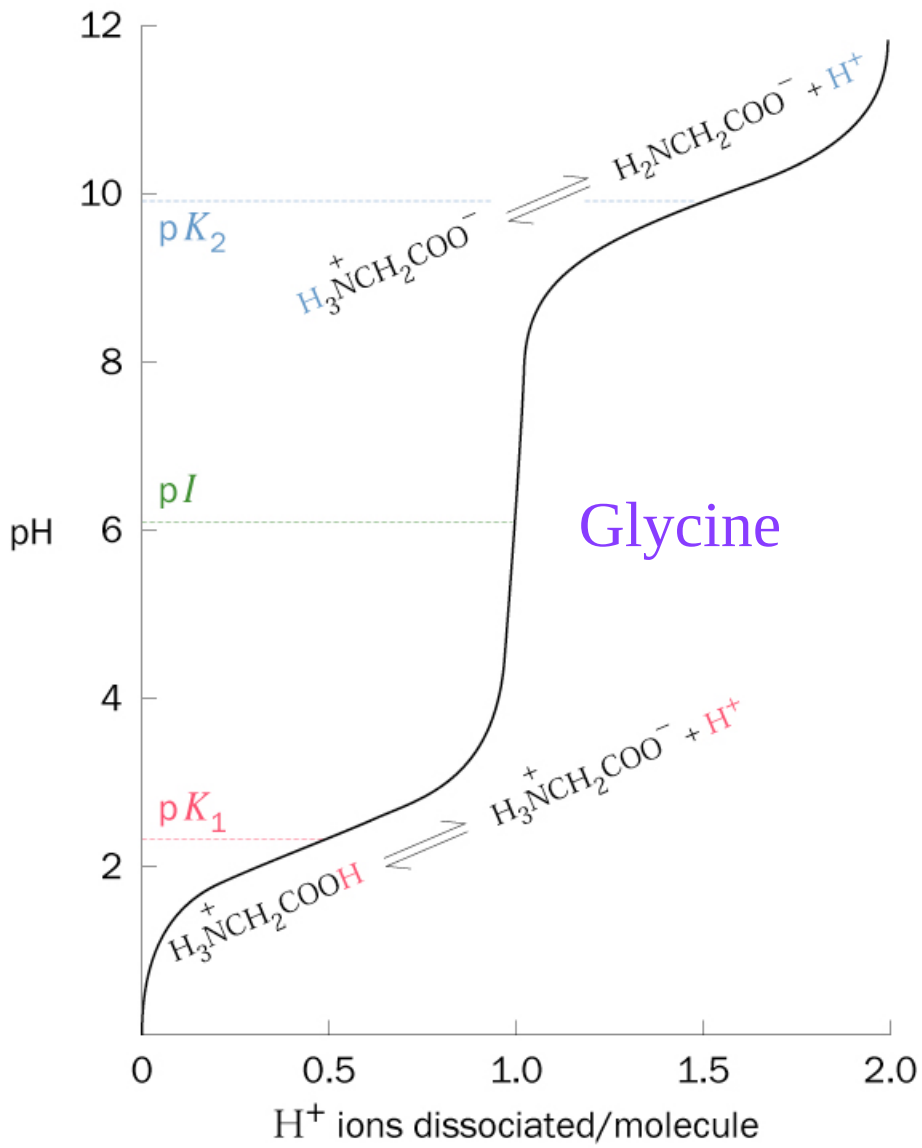
$$pH = pK + \log\left(\frac{[A^-]}{[HA]}\right)$$

Henderson-Hasselbalch Eq.

Isoelectric point: the pH where a protein carries no net electrical charge

$$pI = \frac{1}{2}\left(pK_i + pK_j\right)$$

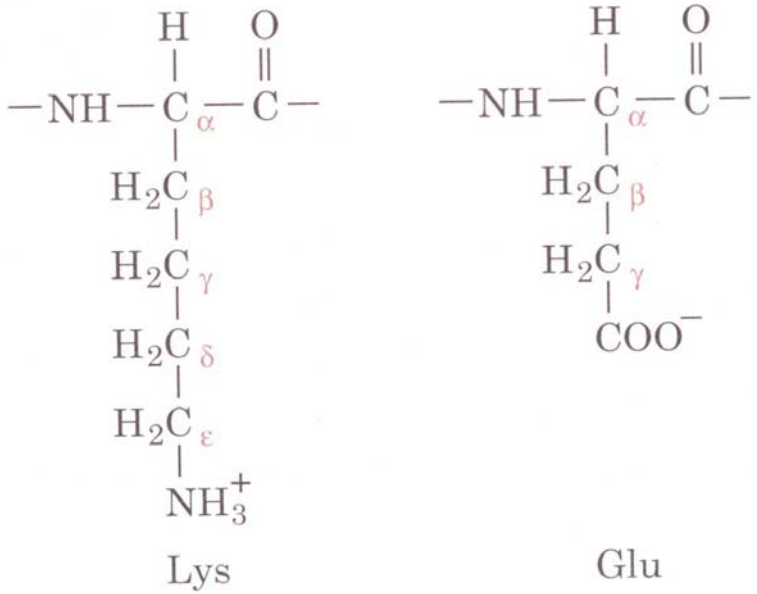
The observed pKa of an amino acid side chain is dependent on its environment in the protein - standard pKa's can be substantially shifted by the protein environment



Isoelectric point

- pI = 0.5(pK_i + pK_j); for 2 ionizable groups
- If amino acid has ionizable side chain, then it must be taken into account when computing pI
- If the side chain is negatively charged when ionized (Asp, Glu), then pI = 0.5(pK_1 + pK_R) (remember pK_1 is the pK_a of the C-terminus, -COOH)
  - e.g., pI of Asp = 0.5(2.20 + 3.90) = 3.05 (the total charge from the side chain and C-term at pH=3.05 is –1 which balances with the +1 charge of the N-term to give a total charge of 0)
- If the side chain is positively charged when ionized (Arg, Lys, His), then pI = 0.5(pK_R + pK_2) (remember pK_2 is the pK_a of the N-terminus, -NH_2)
  - *e.g.*, pI of Lys = 0.5(10.54 + 9.4) = 9.97

Amino acid nomenclature



- Greek lettering used to identify atoms in all amino acid side chains - lysine and glutamate are shown as examples
- Naming is for Carbon atoms - anything attached to the carbon has the same Greek letter
- For example, the NH_3+ at the end of the Lys side chain is Nε

Nomenclature

Glx means either Gln or Glu; same for Asx (Asn or Asp)

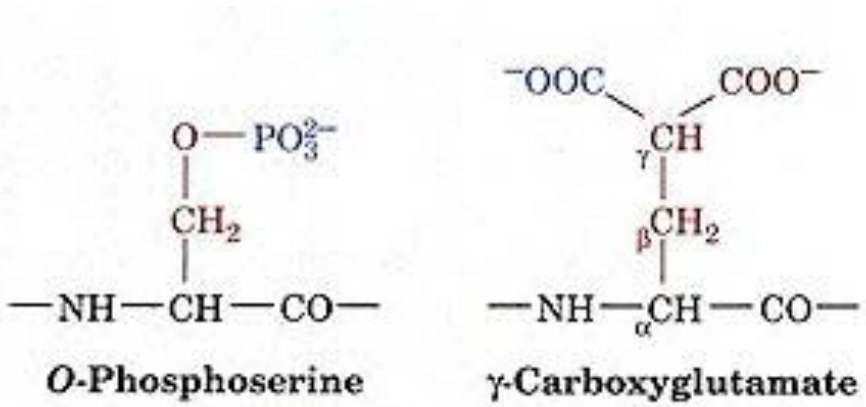
Long name - drop -ine and add -yl and put amino acids in order (e.g. alanine – alanyl, lysine – lysyl, etc.)
The standard method to write an amino acid sequence is from the N-terminus to the C-terminus

N-terminus-AA_1-AA_2-AA_3-AA_4-…AA_n-C-terminus

The protein is different, if named backwards!!
KCAT (Lys-Cys-Ala-Thr) is different from TACK
Order DOES count

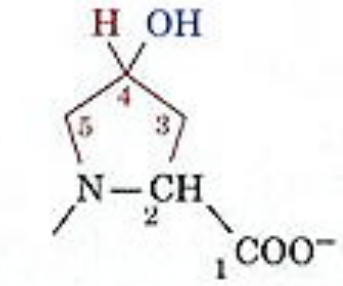
Non-standard amino acids

- Post-translationally modified amino acids
- These transformations are made after the amino acids are already incorporated into a protein
- Typical alterations include: hydroxylation, methylation, acetylation, carboxylation, and phosphorylation
- Addition of PO_3^{2-} to a Ser, Thr, or Tyr is a common theme in signal transduction

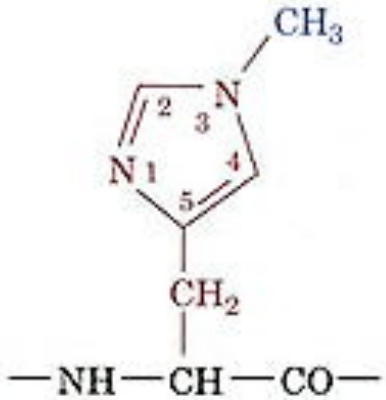


O-Phosphoserine

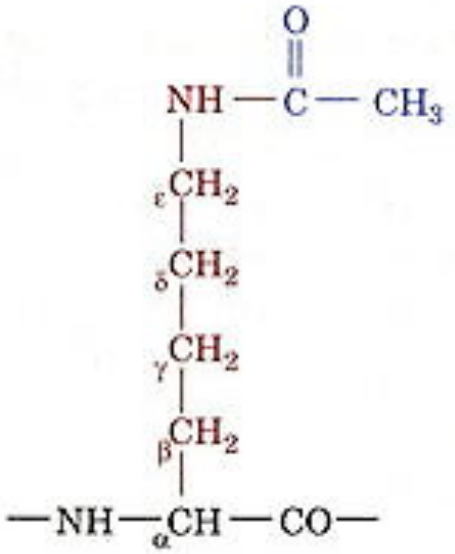
γ-Carboxyglutamate



4-Hydroxyproline



3-Methylhistidine

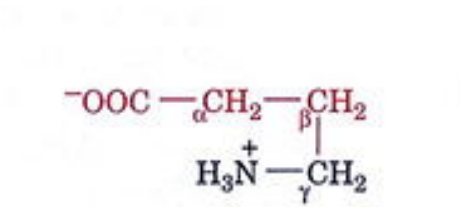


ε-*N*-Acetyllysine

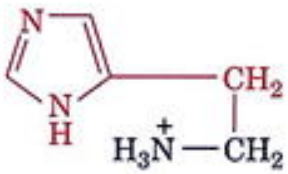
Non-standard amino acids

- Neurotransmitters
  - GABA: glutamine decarboxylation product
  - Dopamine: tyrosine derivative
- Local mediator of allergic reactions
  - Histamine: histidine decarboxylation product
- Thyroid hormone that stimulates vertebrate metabolism
  - Thyroxine: tyrosine derivative

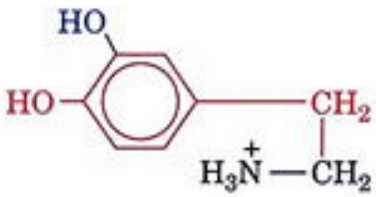
- About 250 amino acids have been found in various plants and fungi



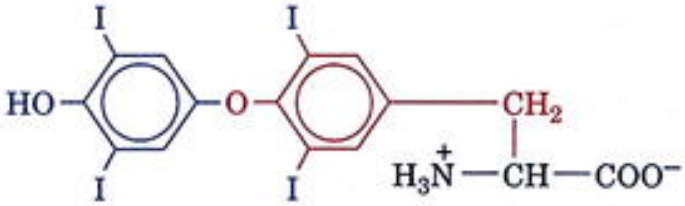
γ-Aminobutyric acid (GABA)



Histamine



Dopamine

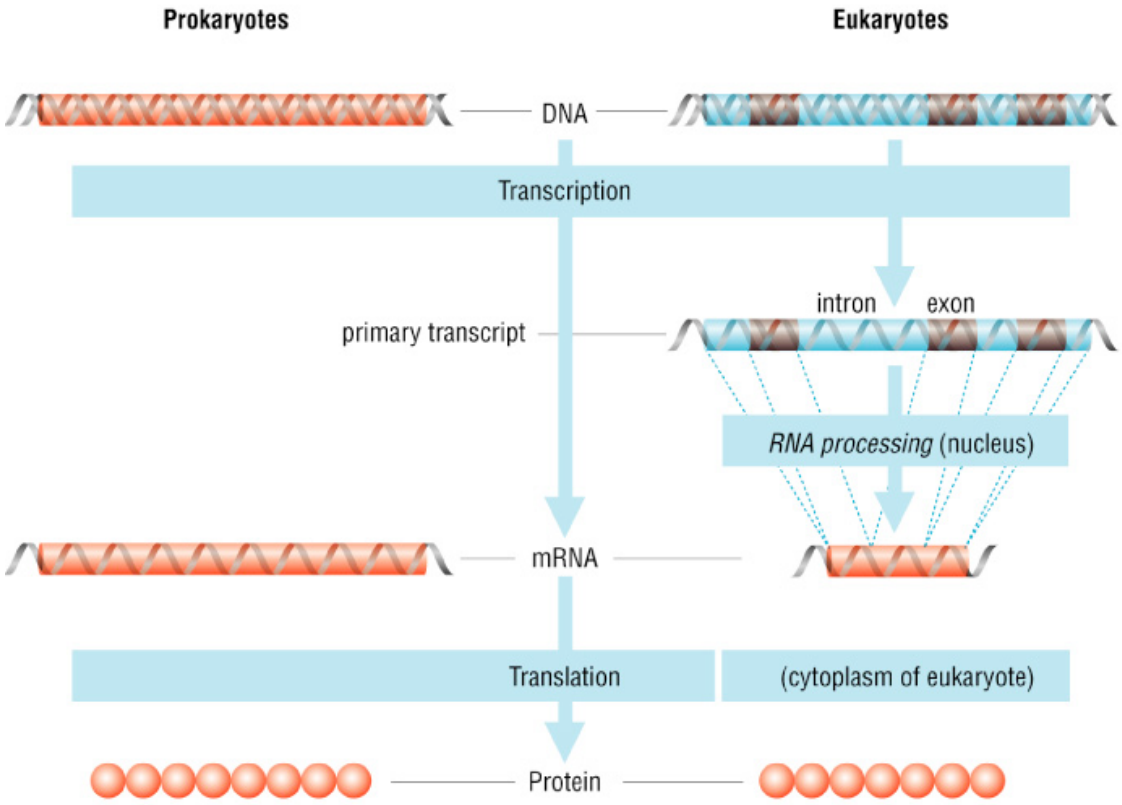


Thyroxine

Gene and Proteins

1st position (5' end)	2nd position U	C	A	G	3rd position (3' end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Amino acids	Abbreviations		Codons
Alanine	Ala	A	GCA GCC GCG GCU
Cysteine	Cys	C	UGC UGU
Aspartic acid	Asp	D	GAC GAU
Glutamic acid	Glu	E	GAA GAG
Phenylalanine	Phe	F	UUC UUU
Glycine	Gly	G	GGA GGC GGG GGU
Histidine	His	H	CAC CAU
Isoleucine	Ile	I	AUA AUC AUU
Lysine	Lys	K	AAA AAG
Leucine	Leu	L	UUA UUG CUA CUC CUG CUU
Methionine	Met	M	AUG
Asparagine	Asn	N	AAC AAU
Proline	Pro	P	CCA CCC CCG CCU
Glutamine	Gln	Q	CAA CAG
Arginine	Arg	R	AGA AGG CGA CGC CGG CGU
Serine	Ser	S	AGC AGU UCA UCC UCG UCU
Threonine	Thr	T	ACA ACC ACG ACU
Valine	Val	V	GUA GUC GUG GUU
Tryptophan	Trp	W	UGG
Tyrosine	Tyr	Y	UAC UAU



Linear relationship, single-nt polymorphism

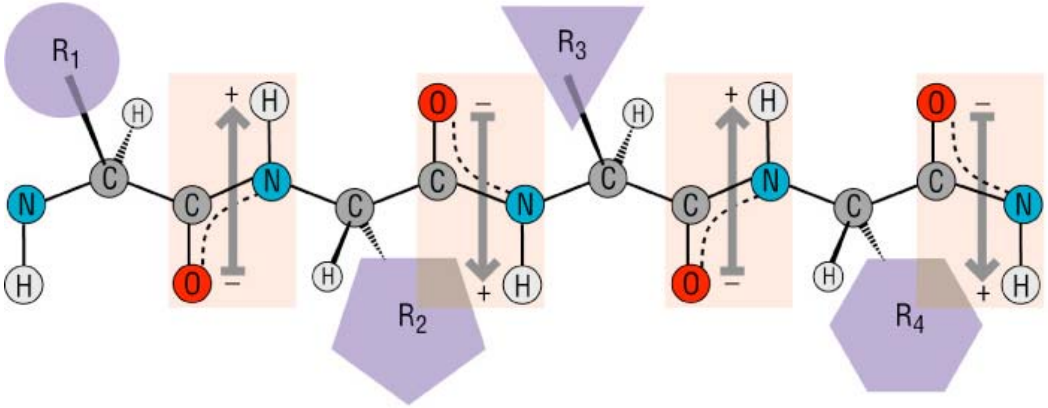
Conservative substitutions

Table of the frequency with which one amino acid is replaced by others in amino-acid sequences of the same protein from different organisms

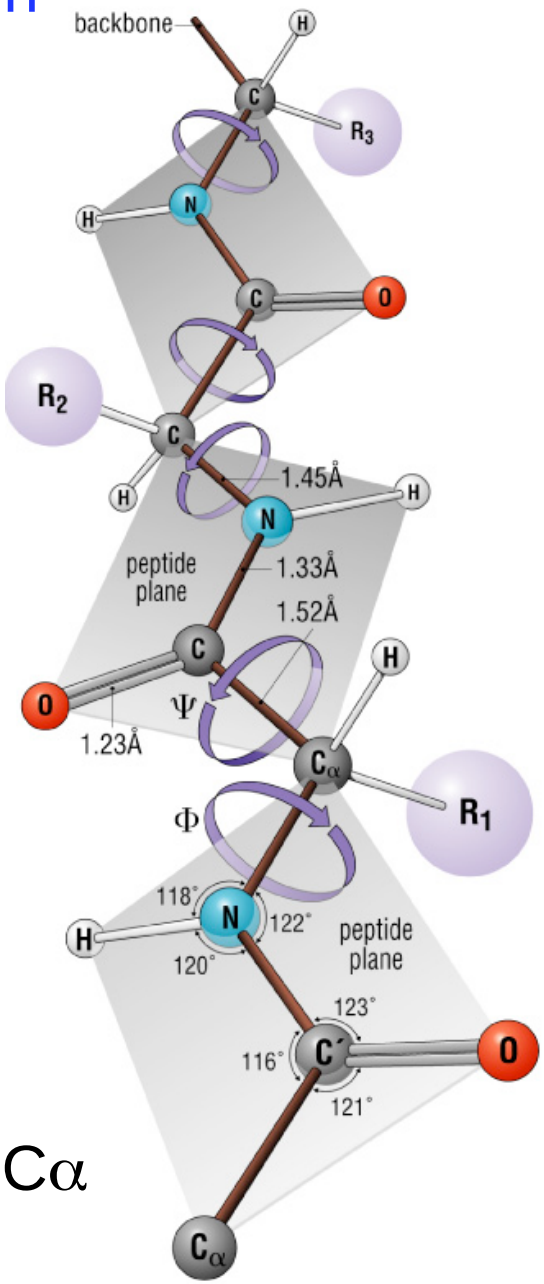
	Gly	Ala	Val	Leu	Ile	Met	Cys	Ser	Thr	Asn	Gln	Asp	Glu	Lys	Arg	His	Phe	Tyr	Trp	Pro
Gly																				
Ala	58																			
Val	10	37																		
Leu	2	10	30																	
Ile		7	66	25																
Met	1	3	8	21	6															
Cys	1	3	3		2															
Ser	45	77	4	3	2	2	12													
Thr	5	59	19	5	13	3	1	70												
Asn	16	11	1	4	4			43	17											
Gln	3	9	3	8	1	2		5	4	5										
Asp	16	15	2		1			10	6	53	8									
Glu	11	27	4	2	4	1		9	3	9	42	83								
Lys	6	6	2	4	4	9		17	20	32	15		10							
Arg	1	3	2	2	3	2	1	14	2	2	12	9		48						
His	1	2	3	4			1	3	1	23	24	4	2	2	10					
Phe	2	2	1	17	9	2		4	1	1					1	2				
Tyr		2	2	2	1		3	2	2	4			1	1		4	26			
Trp				1				2							3		1	1		
Pro	5	35	5	4	1		1	27	7	3	9	1	4	4	7	5	1			

From Protein Structure & Function (Petsko & Ringe)

Diagram of an extended polypeptide chain



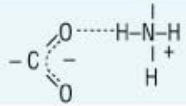
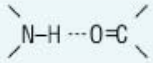
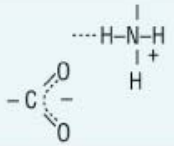
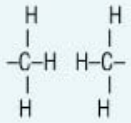
N N+1



Biochemist way - residue

Structural purpose - peptide unit around Cα

Chemical Interactions that Stabilize Polypeptides

Interaction	Example	Distance dependence	Typical distance	Free energy (bond dissociation enthalpies for the covalent bonds)
Covalent bond	$-C_\alpha-C-$	-	1.5 Å	356 kJ/mole (610 kJ/mole for a C=C bond)
Disulfide bond	–Cys–S–S–Cys–	-	2.2 Å	167 kJ/mole
Salt bridge		Donor (here N), and acceptor (here O) atoms <3.5 Å	2.8 Å	12.5–17 kJ/mole; may be as high as 30 kJ/mole for fully or partially buried salt bridges (see text), less if the salt bridge is external
Hydrogen bond		Donor (here N), and acceptor (here O) atoms <3.5 Å	3.0 Å	2–6 kJ/mole in water; 12.5–21 kJ/mole if either donor or acceptor is charged
Long-range electrostatic interaction		Depends on dielectric constant of medium. Screened by water. 1/r dependence	Variable	Depends on distance and environment. Can be very strong in nonpolar region but very weak in water
Van der Waals interaction		Short range. Falls off rapidly beyond 4 Å separation. $1/r^6$ dependence	3.5 Å	4 kJ/mole (4–17 in protein interior) depending on the size of the group (for comparison, the average thermal energy of molecules at room temperature is 2.5 kJ/mole)

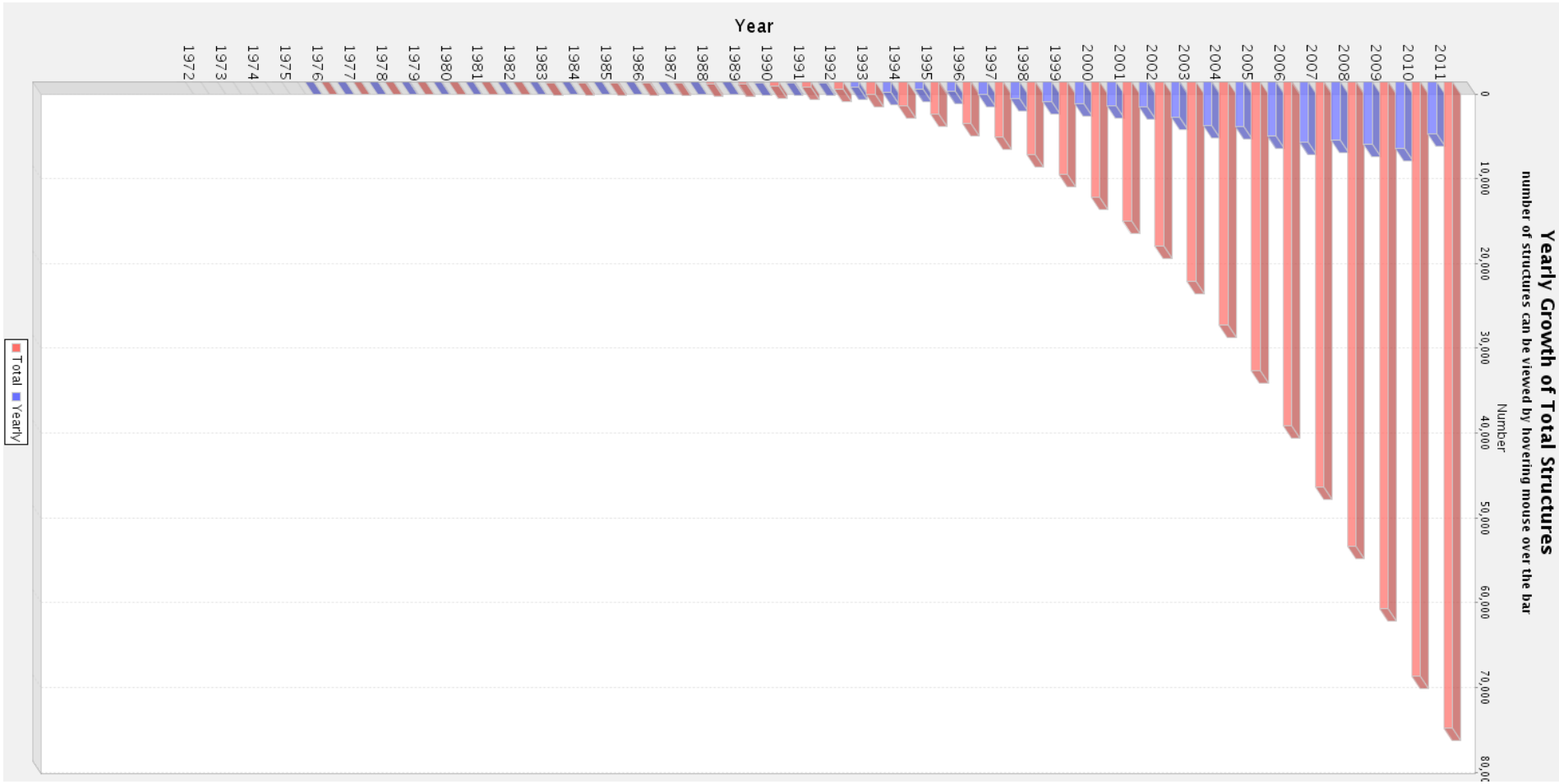
Any specific number is highly dependent on the context in which the interaction is found!!

Retrieving and Viewing Protein Structures from the Protein Data Bank (PDB)

Protein Data Bank

- 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 xx,xxx protein structures
- Growing at 20% per year
- New structures 50 times larger than those in 1971 are common place

PDB Growth



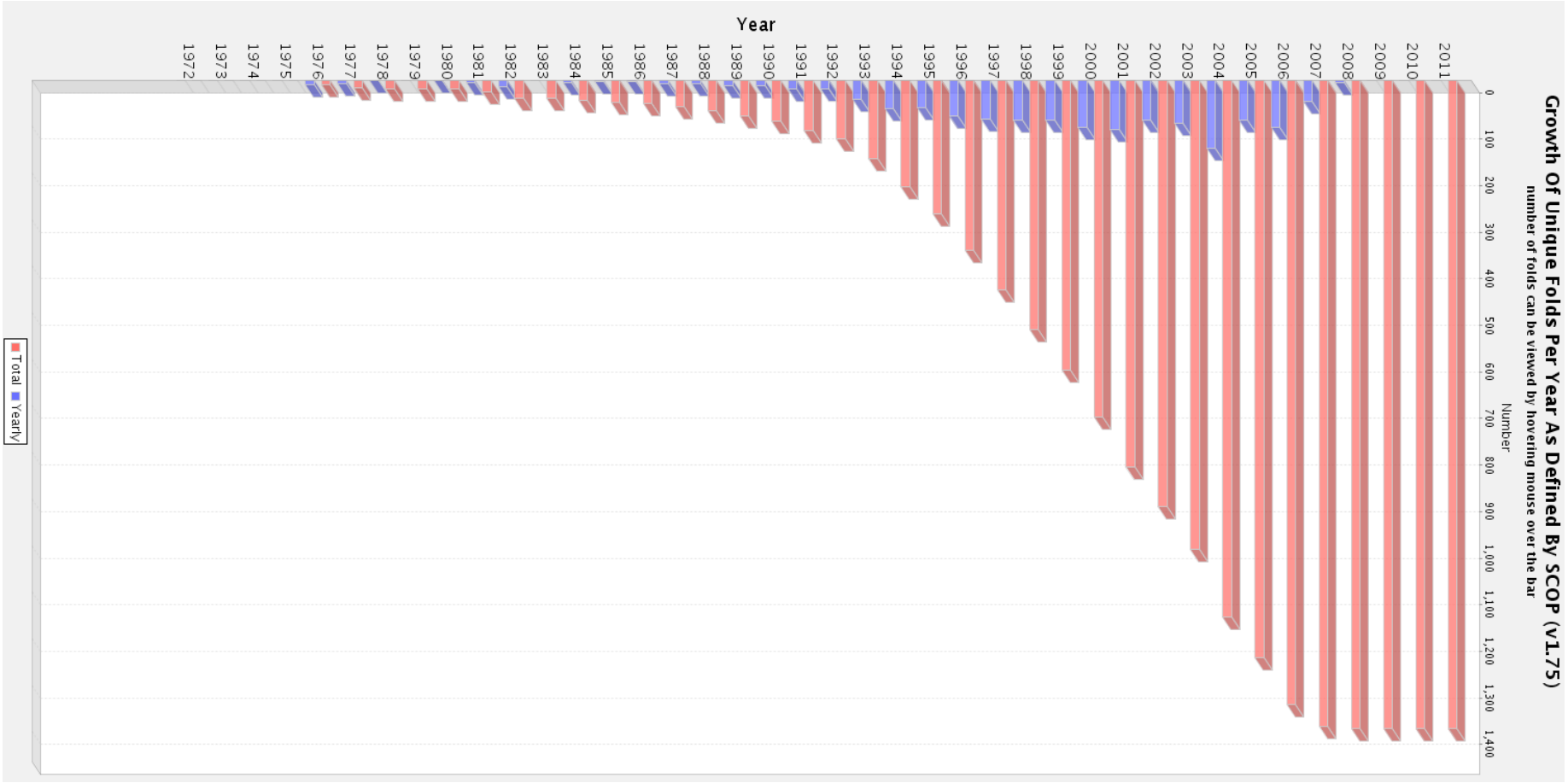
As of October 4, 2011, 76288

- 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

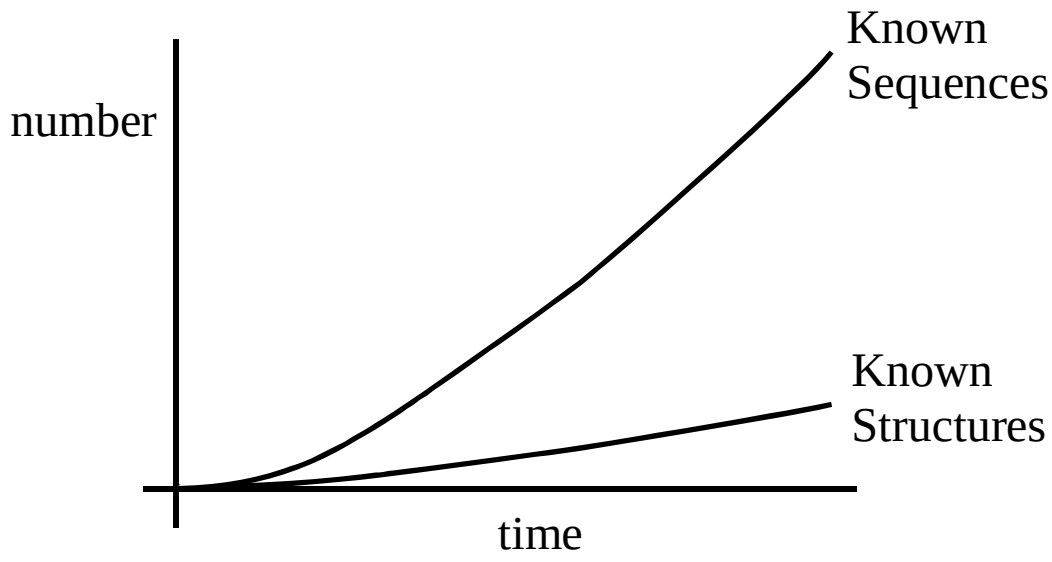
Not all Structures are Different

PDB Growth in “New Folds”



Structure vs. Sequence

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



Point of Information

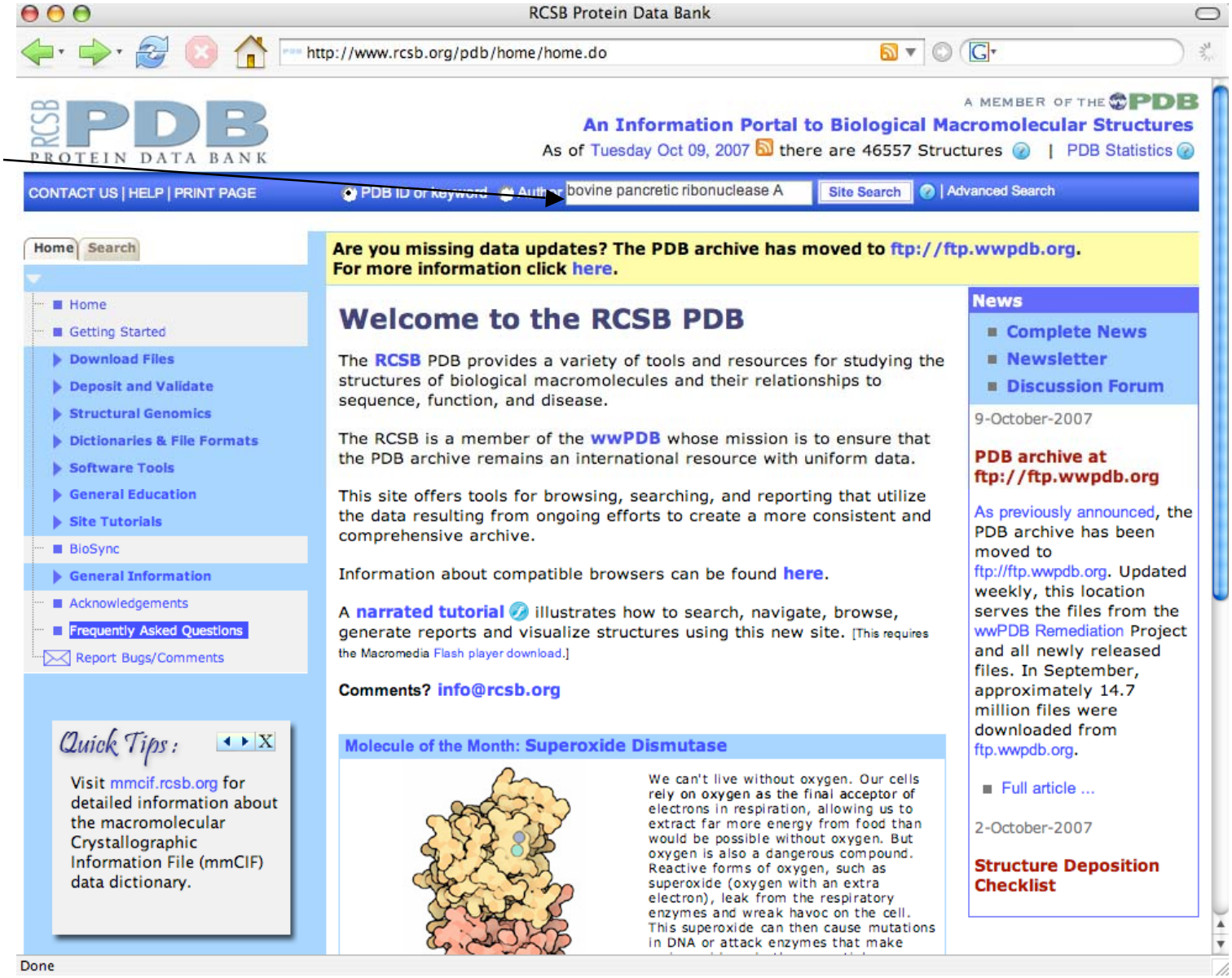
- Today's material is:
  - a subset of the information available to you in online tutorials
  - presented to "get you started" quickly and to "shorten the learning curve"
  - not exhaustive or even sufficient

    => should be augmented by actually working through the online tutorials

PDB Website

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

Enter what you know (names, id codes..)



RCSB PDB PROTEIN DATA BANK | RCSB PDB-101

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

As of **Tuesday Oct 04, 2011 at 5 PM PDT** there are **76288** Structures | **PDB Statistics**

Search | All Categories: All Categories | Author | Macromolecule | Sequence | Ligand

e.g., PDB ID, molecule name, author

Browse | Advanced

Customize This Page

MyPDB Hide
- Login to your Account
- Register a New Account

Home Hide
- News & Publications
- Usage/Reference Policies
- Deposition Policies
- Website FAQ
- Deposition FAQ
- Contact Us
- About Us
- Careers
- External Links
- Sitemap
- New Website Features

Deposition Hide
- All Deposit Services
- Electron Microscopy
- X-ray | NMR
- Validation Server
- BioSync Beamlines/Facilities
- Related Tools

Tools Hide
- Download Files
- Compare Structures
- File Formats
- Services: RESTful | SOAP
- Widgets

PDB-101 Hide
- Structural View of Biology
- Understanding PDB Data
- Molecule of the Month
- Educational Resources

Help Hide

Biological Macromolecular Resource

Full Description

Featured Molecules Hide

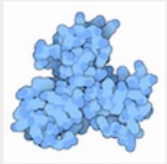
Structural View of Biology List View of Archive By: **Title** | **Date** | **Category**

Molecule of the Month

PDB Pioneers

Structural biology was born in 1958 with John Kendrew's atomic structure of myoglobin, and in the following decade, the field grew rapidly. By the early 1970's, there were a dozen atomic structures of proteins, and researchers were discovering that they had a goldmine of information.

Full Article



Protein Structure Initiative Featured System

Bacterial Armor

Researchers at MCSG have revealed the inner workings of a surface layer protein, showing how bacteria attach their form-fitting protein coats.

Full Article | **Archive** | **PSI Structural Biology Knowledgebase**

Explore Archive Hide

Organism | Taxonomy | Exp. Method | X-Ray Resolution | Release Date | Polymer Type | Enzyme Classification | SCOP Classification

Show all

Organism
- Homo sapiens (18472)
- Escherichia coli (4485)
- Mus musculus (3297)
- Saccharomyces cerevisiae (2073)
- Bos taurus (1954)
- Rattus norvegicus (1650)
- Escherichia coli K-12 (1202)
- Other (40627)

New Structures Hide
- Latest Release
- New Structure Papers
- Unreleased Entries

New Features Hide

Redesigned Search: Query History and MyPDB

Latest features released:

Website Release Archive:

wwPDB News Hide

2011-09-30

PDB Depositions Can Be Targets for CAPRI

2011-09-30

PDB 40 News: Special Anniversary at CSHL October 28-30, 2011

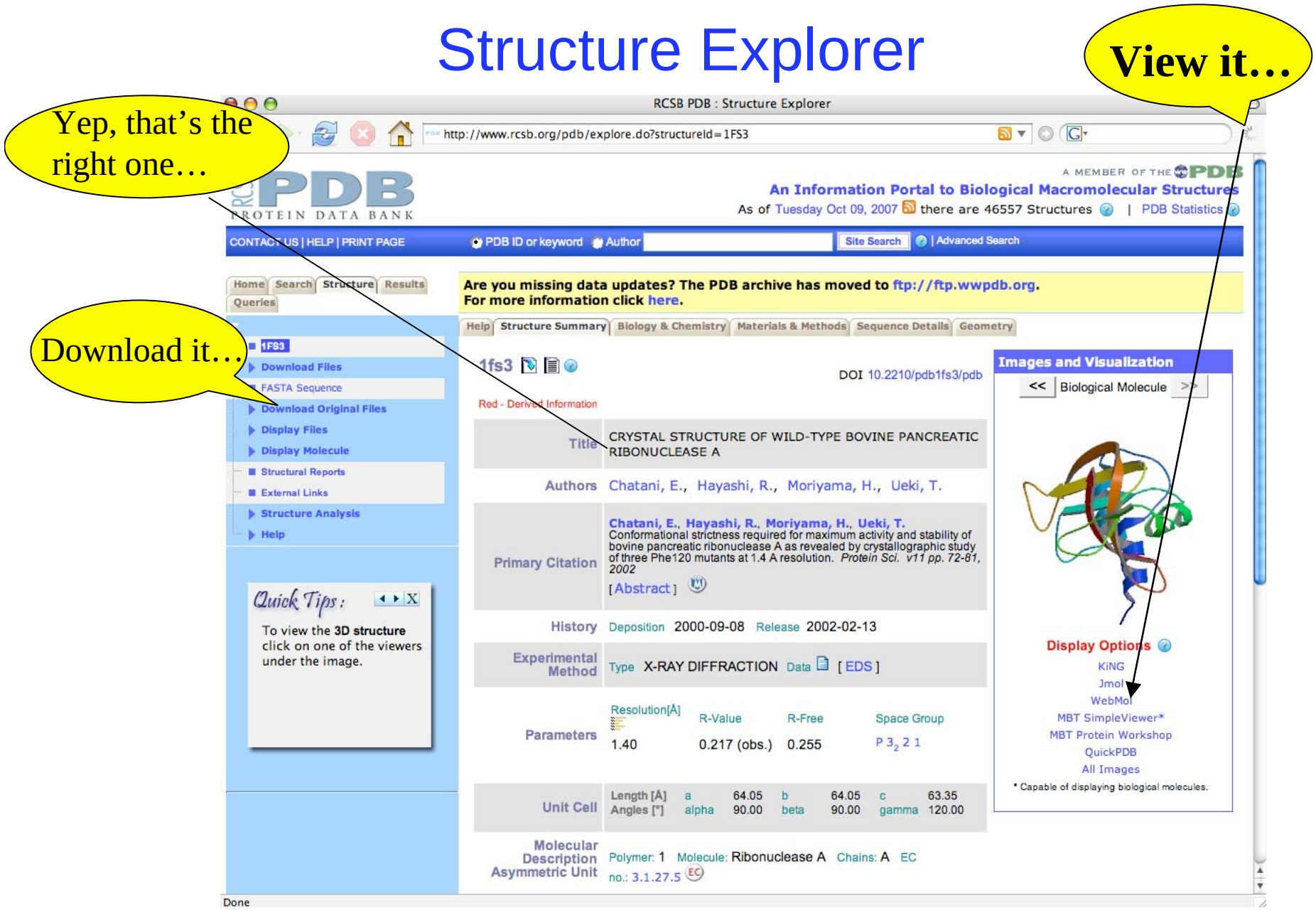
- **IUCr XXII: wwPDB Q&A session, exhibit booth, presentations, posters and more**
- **Full wwPDB News**
- **Statement on Retraction of PDB Entries**

Query Result Browser

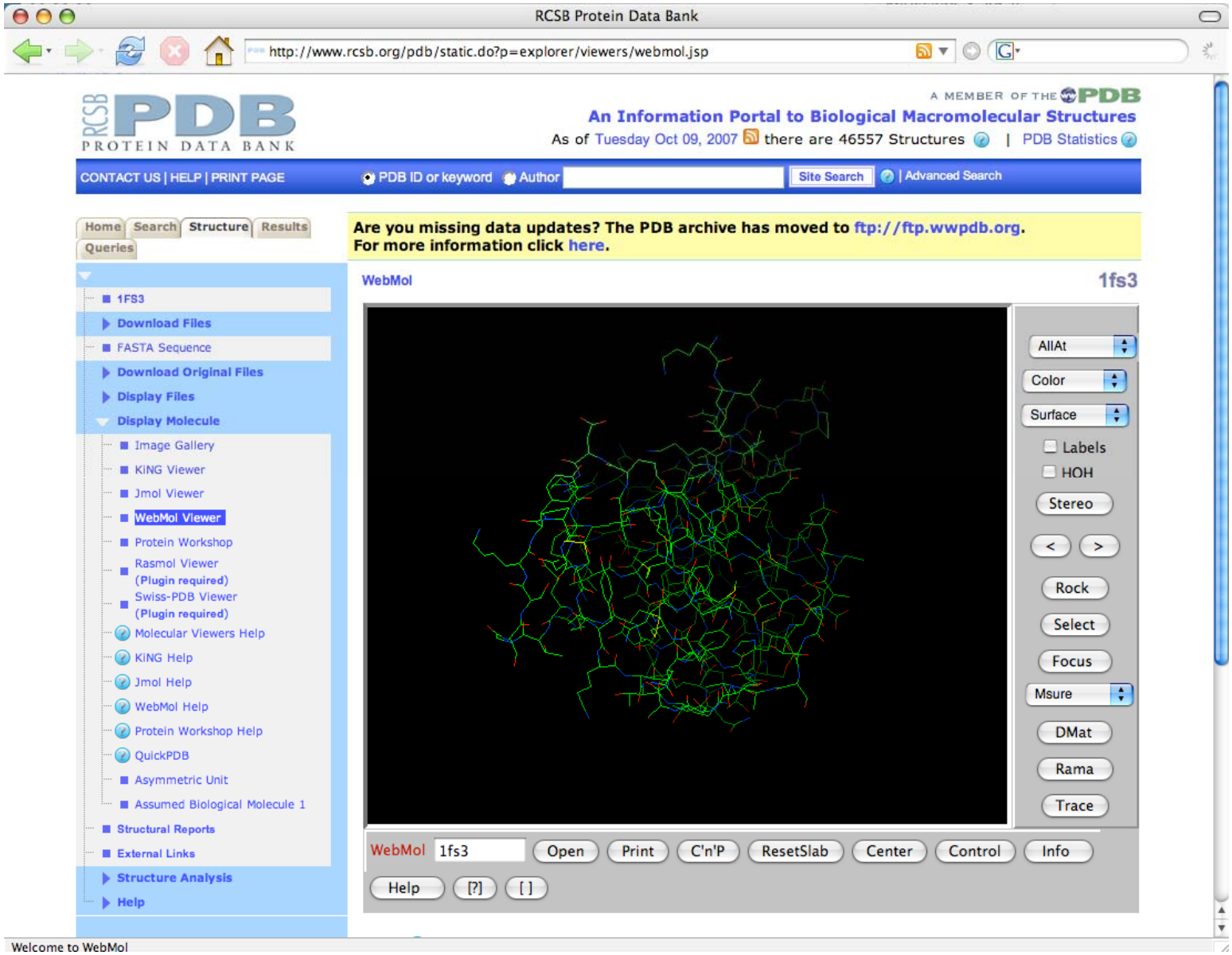
Which one do I want?

Let's look at this one …

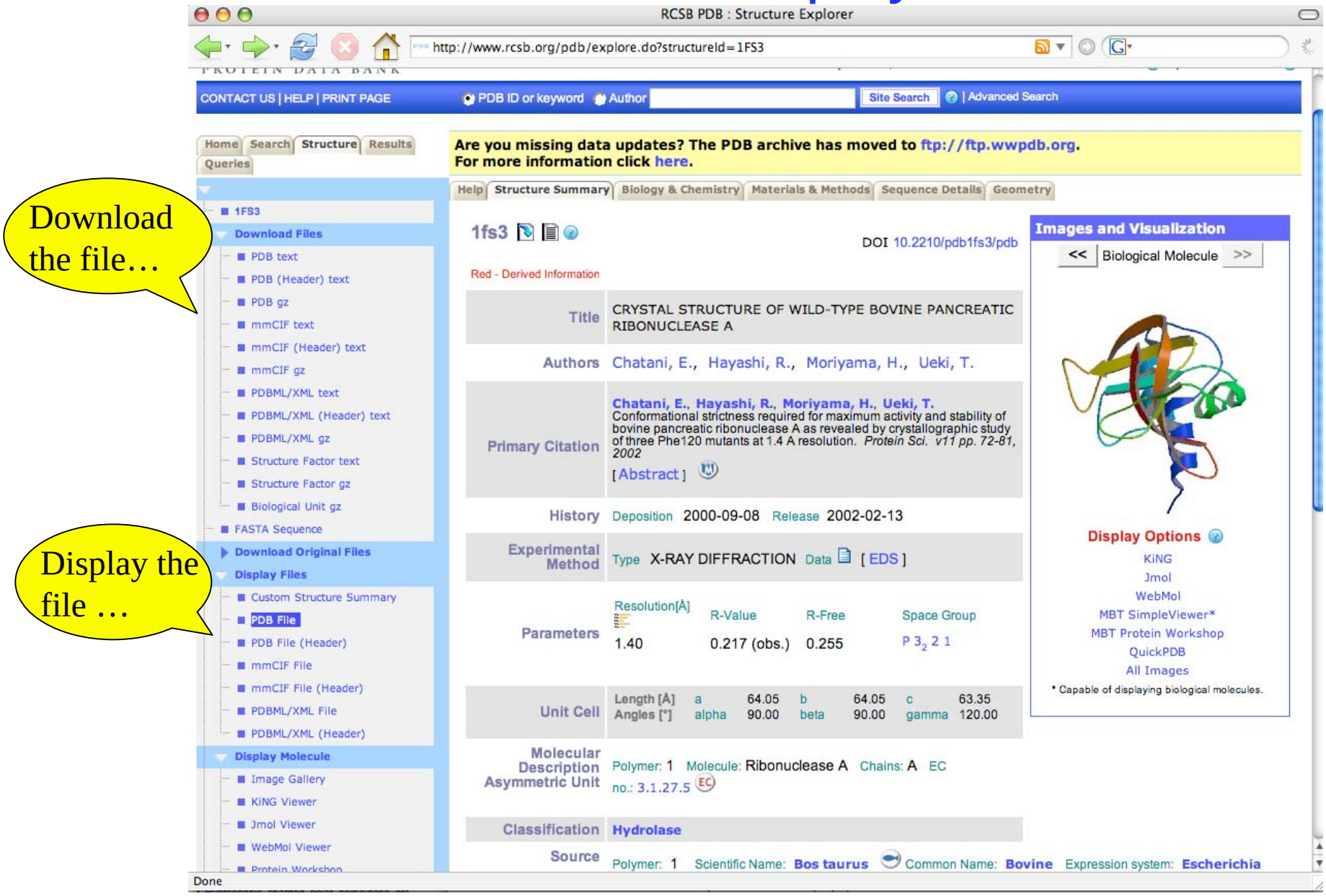
Structure Explorer



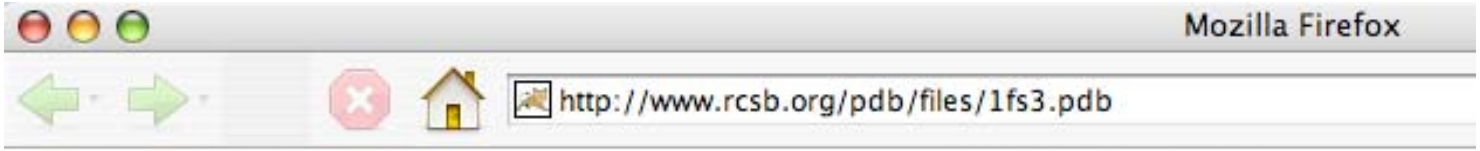
View Structure



Download/Display



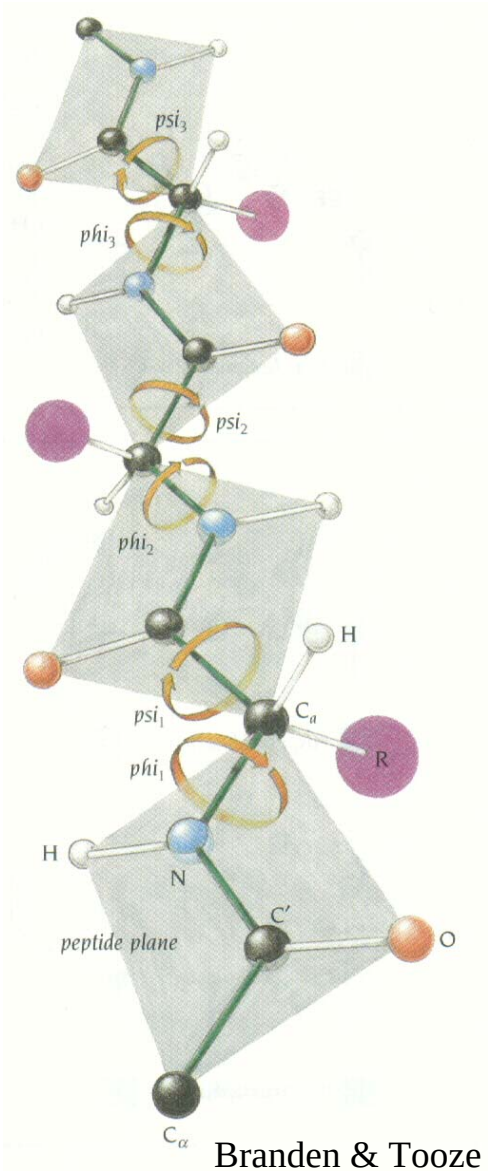
Header Information




```
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   2   31-DEC-02 1FS3    1       REMARK
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 A AS REVEALED BY CRYSTALLOGRAPHIC
JRNL        TITL 4 STUDY OF THREE PHE120 MUTANTS AT 1.4 A RESOLUTION.
JRNL        REF    PROTEIN SCI.                  V.  11    72 2002
JRNL        REFN   ASTM PRCIEI  US ISSN 0961-8368
REMARK   1
REMARK   2
REMARK   2 RESOLUTION. 1.40 ANGSTROMS.
```

Visualizing Proteins

- High complexity
- Multiple levels of structure
- Important properties are “distributed” throughout the 3D structure

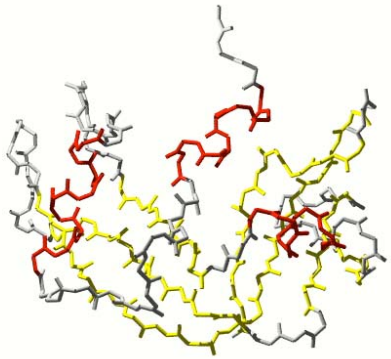


Branden & Tooze

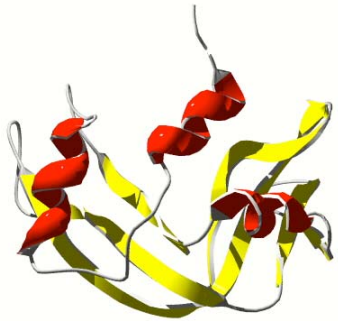
Visualization Objectives

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

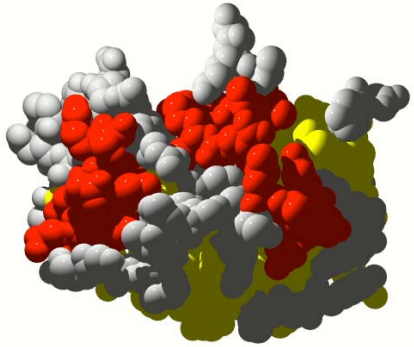
Display Conventions



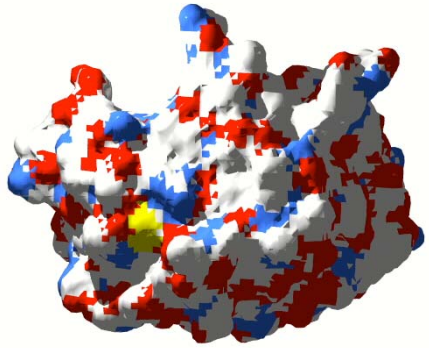
Wireframe



Ribbon



Space filling



Molecular Surface

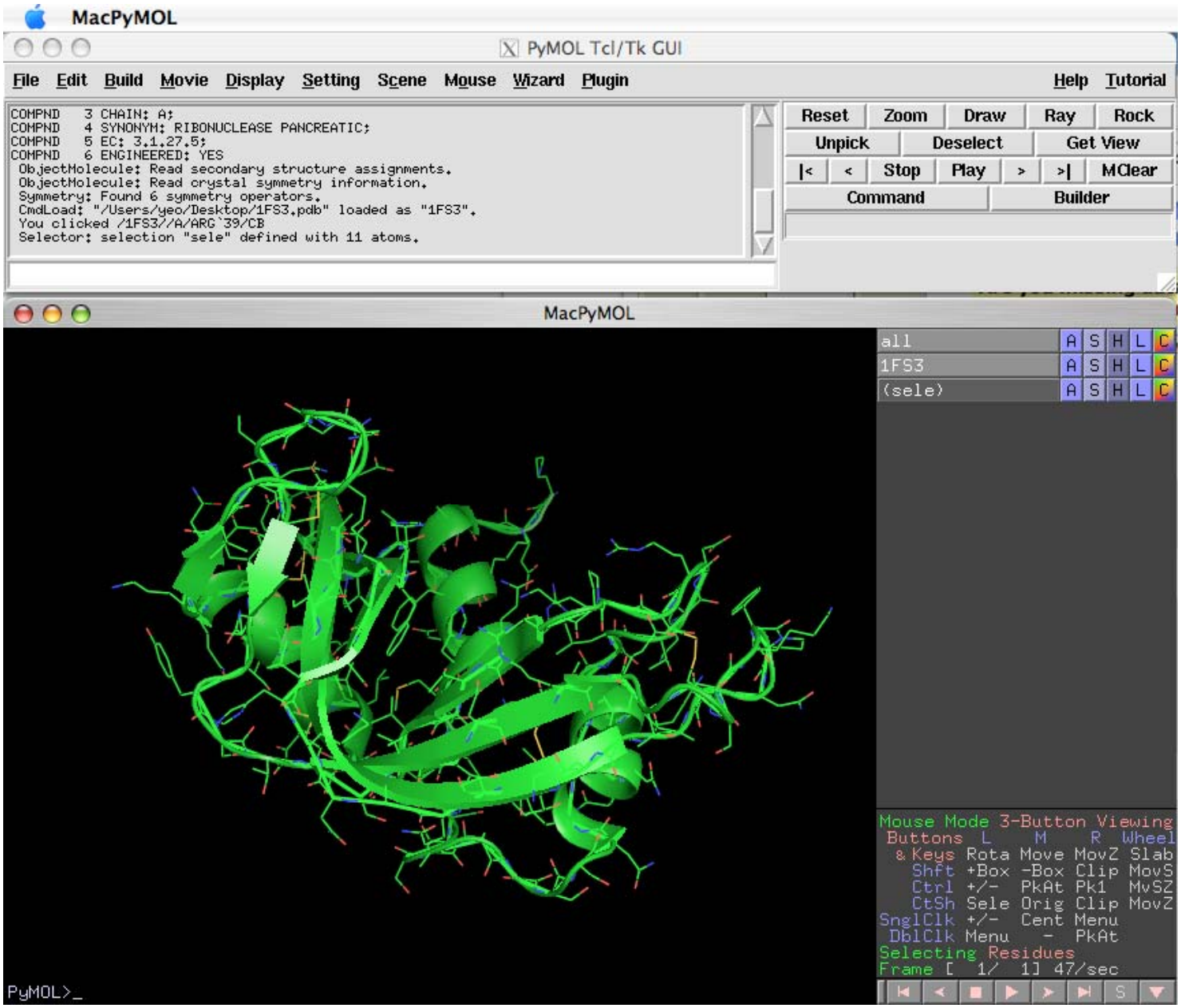
Important URLs & Visualization Tools

PyMol:	http://pymol.sourceforge.net/
Chimera:	http://www.cgl.ucsf.edu/chimera/
VMD:	http://www.ks.uiuc.edu/Research/vmd/
MolMol:	http://hugin.ethz.ch/wuthrich/software/molmol/index.html
Cn3D:	http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml
iMol:	http://www.pirx.com/iMol/
Molview:	http://www.danforthcenter.org/smith/MolView/molview.html
RasMol:	http://www.bernstein-plus-sons.com/software/rasmol/

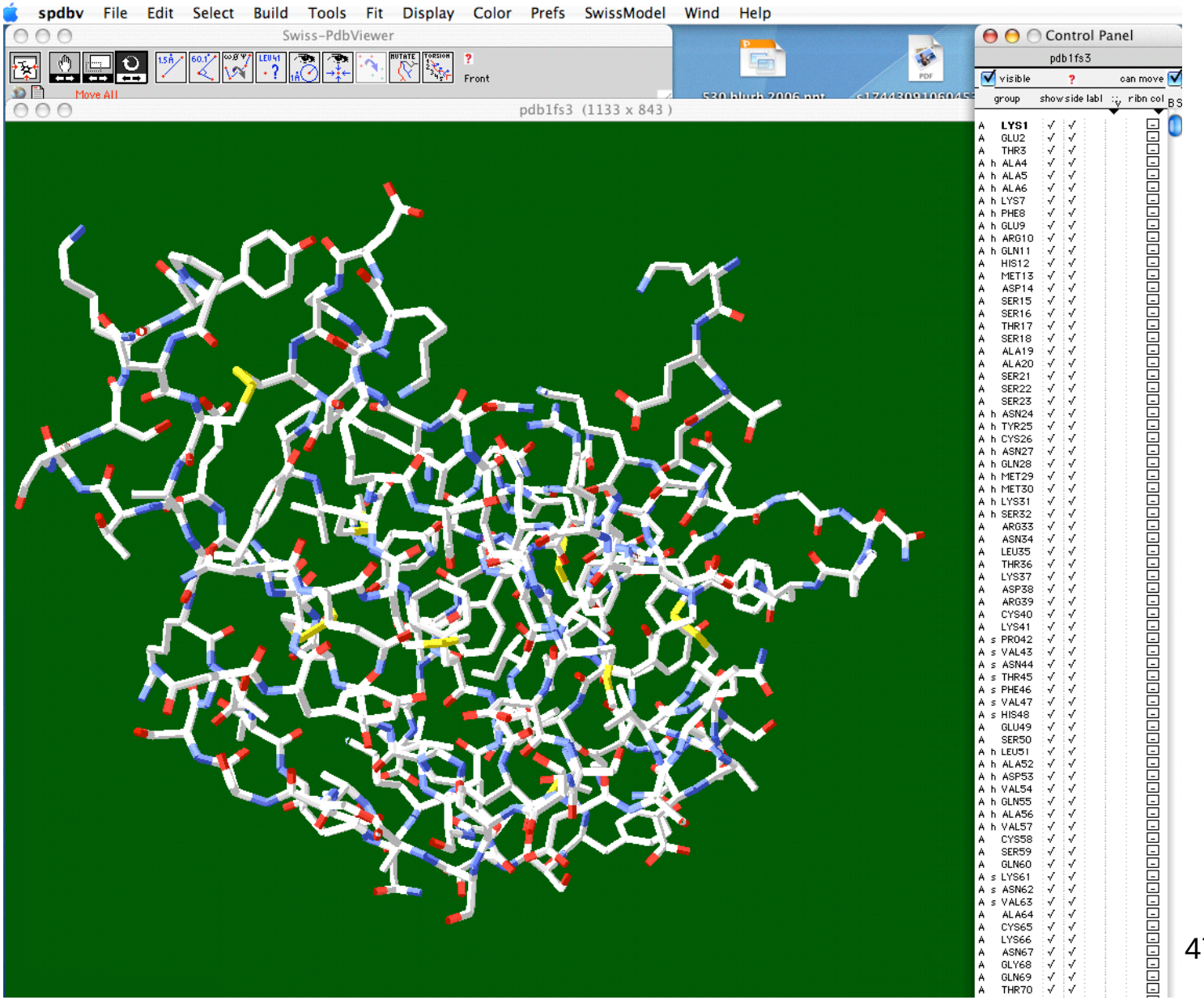
- Operating systems – Unix, Windows, Mac

- Our choice (arbitrary) :
  - –PyMOL
  - –SwissPDB (stand-alone) etc.

PyMOL


MacPyMOL
PyMOL Tcl/Tk GUI
File Edit Build Movie Display Setting Scene Mouse Wizard Plugin
Help Tutorial
COMPND 3 CHAIN: A;
COMPND 4 SYNONYM: RIBONUCLEASE PANCREATIC;
COMPND 5 EC: 3.1.27.5;
COMPND 6 ENGINEERED: YES
ObjectMolecule: Read secondary structure assignments.
ObjectMolecule: Read crystal symmetry information.
Symmetry: Found 6 symmetry operators.
CmdLoad: "/Users/yeo/Desktop/1FS3.pdb" loaded as "1FS3".
You clicked /1FS3//A/ARG`39/CB
Selector: selection "sele" defined with 11 atoms.
Reset Zoom Draw Ray Rock
Unpick Deselect Get View
|< < Stop Play > >| MClear
Command Builder
MacPyMOL
all A S H L C
1FS3 A S H L C
(sele) A S H L C
Mouse Mode 3-Button Viewing
Buttons L M R Wheel
& Keys Rota Move MovZ Slab
Shft +Box -Box Clip MovS
Ctrl +/- PkAt Pk1 MvSZ
CtSh Sele Orig Clip MovZ
SnglClk +/- Cent Menu
DblClk Menu - PkAt
Selecting Residues
Frame [1/ 1] 47/sec
PyMOL>_

SwissPDB


spdbv File Edit Select Build Tools Fit Display Color Prefs SwissModel Wind Help
Swiss-PdbViewer
Front
Move All
pdb1fs3 (1133 x 843)
Control Panel
pdb1fs3
visible
can move
group
show side labl
ribn col
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
A SER50
A h LEU51
A h ALA52
A h ASP53
A h VAL54
A h GLN55
A h ALA56
A h VAL57
A CYS58
A SER59
A GLN60
A s LYS61
A s ASN62
A s VAL63
A ALA64
A CYS65
A LYS66
A ASN67
A GLY68
A GLN69
A THR70

SwissPDB – Toolbar