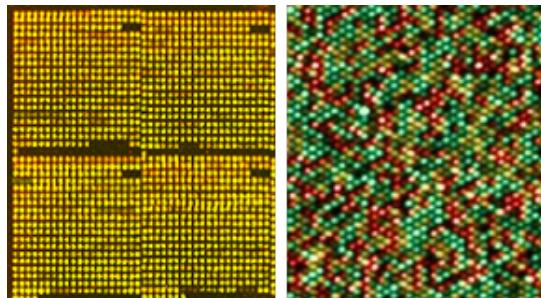


Effect of variations on protein structure

Variation

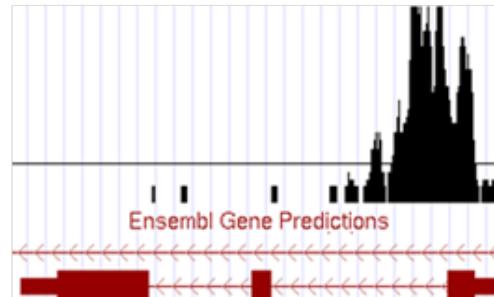
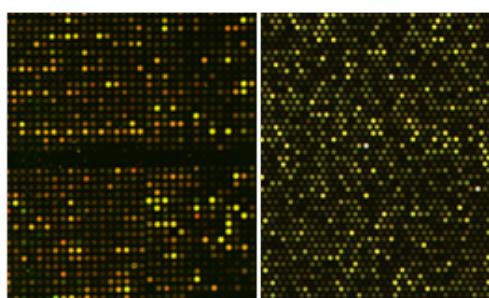
Omics views of genomes

Genetic variation

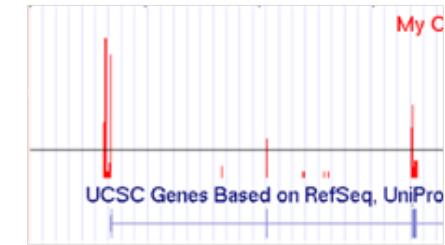
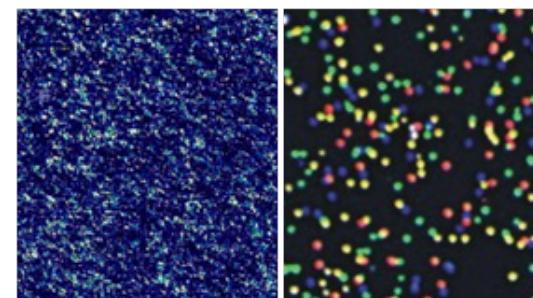


TCCTTCGGACTCTCGA~~C~~TGAAACCTTTAGGTG X 1
CTTTCGGACTCTCGA~~C~~TGAAACCTTTAGGTG X 1
CTTTCGG~~A~~CTCTCGG~~C~~CTGAAACCTTTAGGTG X 1
TTTCGGACTCTCGA~~C~~CTCGAACCTTTAGGTGTA X 2
TTTCGGACTCTCGG~~C~~CTGAAACCTTTAGGTGTA X 1
TTTCGGACTCTCGA~~C~~CTCGAACCTTTAGGTGTA X 2
TCGGACTCTCGA~~C~~CTCGAACCTTTAGGTGTAAA X 2
CGGACTCTCGA~~C~~CTCGAACCTTTAGGTGTAAA X 3
CGGACTCTCGG~~C~~CTCGAACCTTTAGGTGTAAA X 1
CGGACTCTCGG~~C~~CTCGAACCTTTAGGTGTAAA X 1
GGA~~A~~CTCGG~~C~~CTCGAACCTTTAGGTGTAAAAG X 1
GACTCTCGG~~C~~CTCGAACCTTTAGGTGTAAAAGAG X 1
ACTCTCGA~~C~~CTCGAACCTTTAGGTGTAAAAGAG X 1
CTCTCGG~~C~~CTCGAACCTTTAGGTGTAAAAGAG X 1
CTCTCGA~~C~~CTCGAACCTTTAGGTGTAAAAGAG X 1
CTCGAACCTTTAGGTGTAAAAGAGCC X 1
TCGAACTCGAACCTTTAGGTGTAAAAGAGCC X 2
TCGGCTCGAACCTTTAGGTGTAAAAGAGCC X 1
CGGCTCGAACCTTTAGGTGTAAAAGAGCC X 2
AGAAAGCCTGAGAGCCGAGCTTGAAATCCACATTCTCTGGCTGC

Epigenetic variation



Expression variation



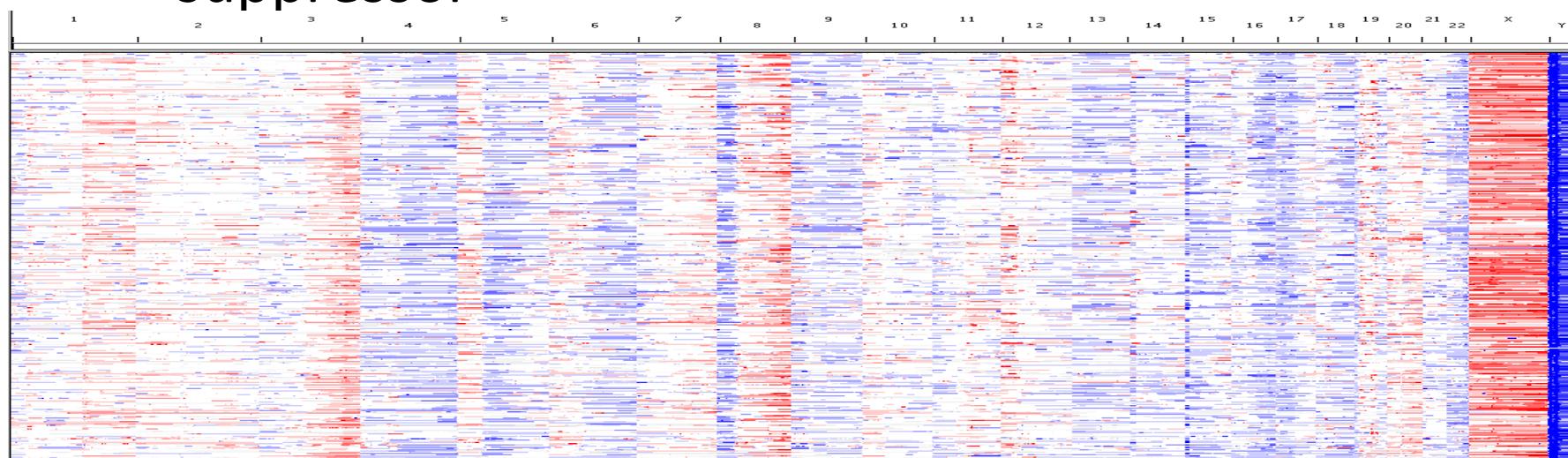
SNPs, loss-of-heterozygosity
Copy number variants

DNA methylation
Chromatin

RNA expression
Gene structure

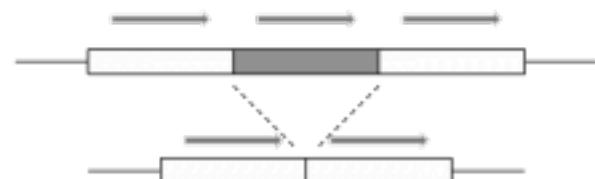
DNA Copy Number Alterations

- Chunks of the genome can be amplified
 - Leading to many copies of an oncogene
 - Which leads to overexpression of the gene
- Chunks can also be lost (deleted)
 - And that is one mechanism to lose a tumor suppressor

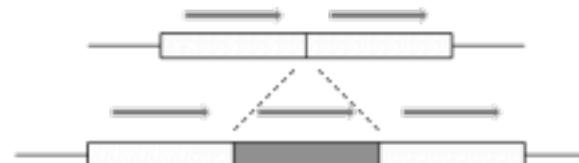


Structural variation

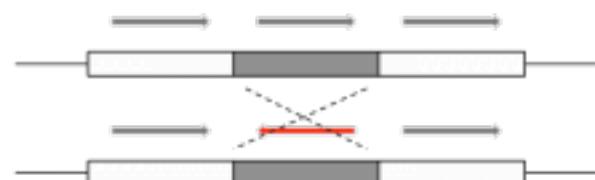
Structural Variation



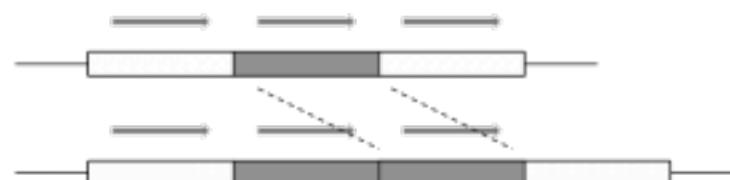
Deletion



Insertion



Inversion



Duplication



Copy Number Variation

Point mutations

- Nucleotide change can lead to:
 - An early stop codon → protein cannot fold
 - Changes in protein structure
 - Create a constitutively active protein

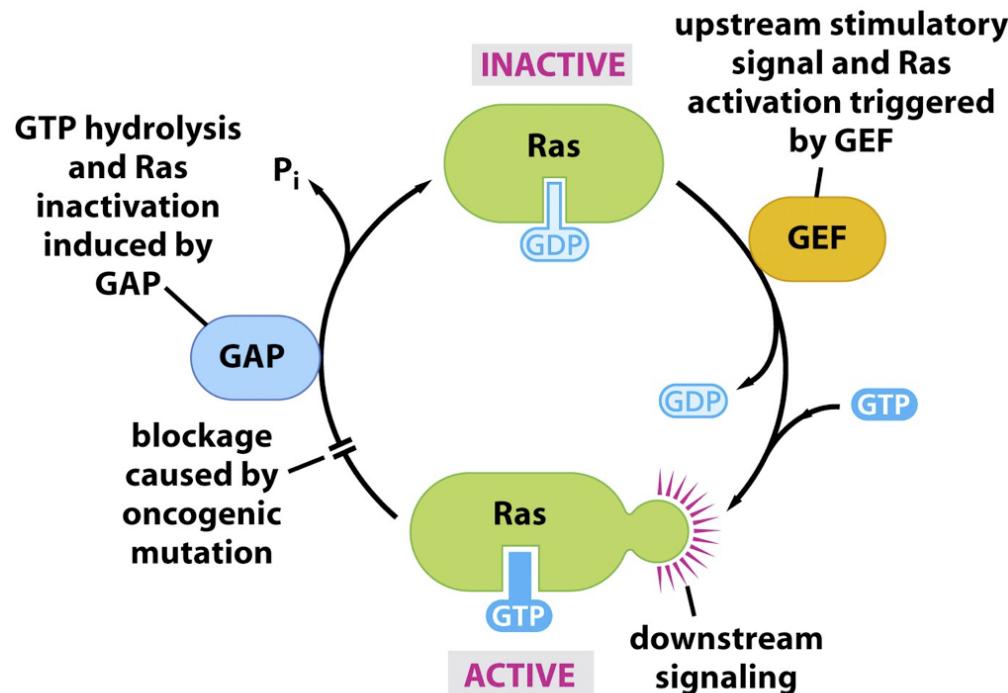
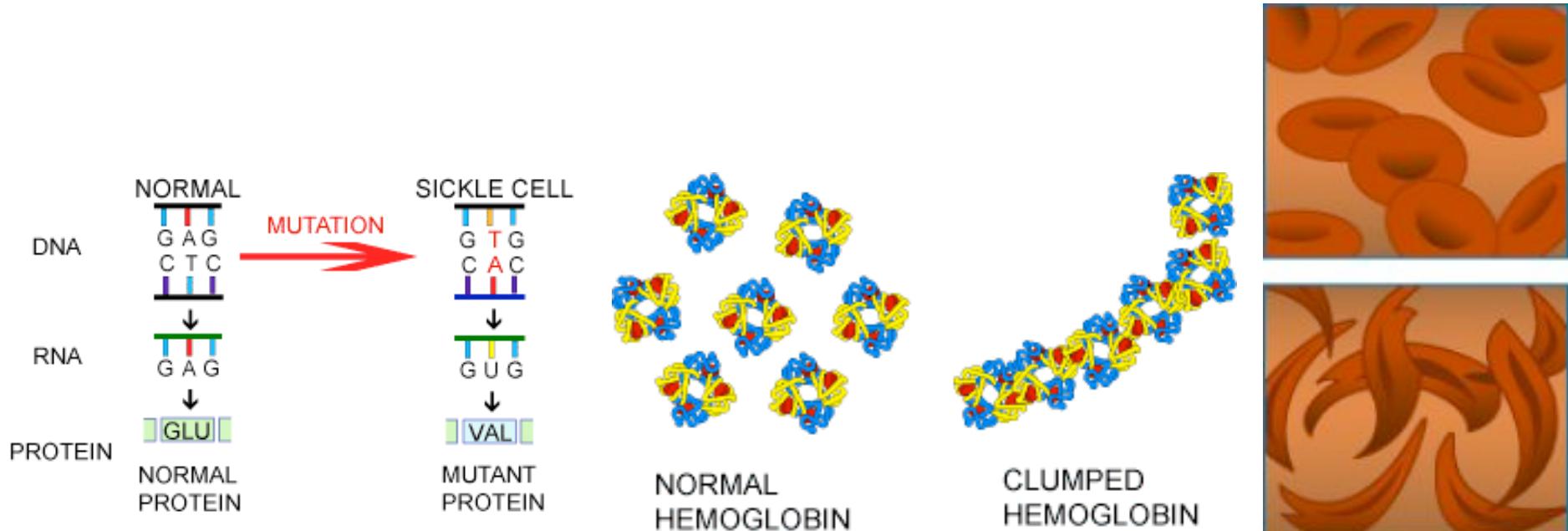


Figure 5-30 The Biology of Cancer (© Garland Science 2007)

Sickle cell anemia

Genetic disease with severe symptoms
Caused by a mutation in hemoglobin gene

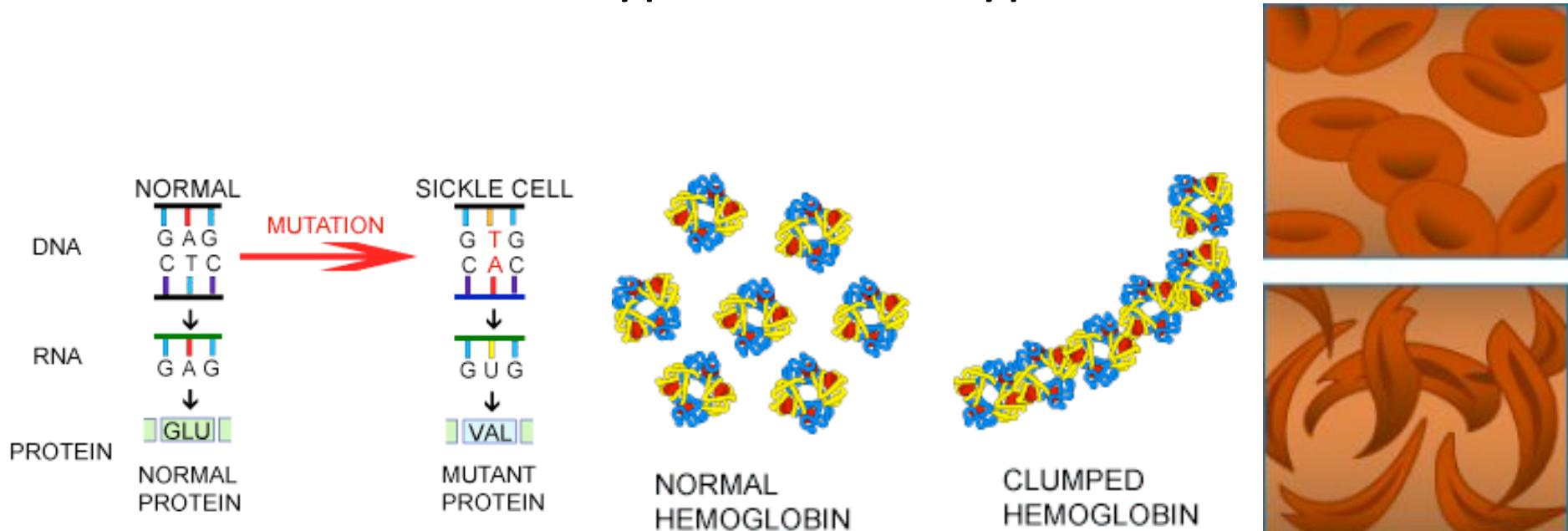
Genotype → Phenotype



Sickle cell anemia

Genetic disease with severe symptoms
Caused by a mutation in hemoglobin gene
Positive side effect: Provides resistance to malaria

Genotype → Phenotype



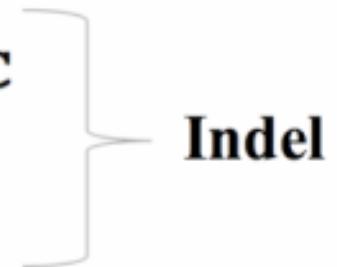
Single nucleotide polymorphism

Reference	ACTGACGCATGCATCATGCATGC
SNP	ACTGACGCATGCATCATT C TATGC

SNPs result from the substitution of a single base-pair.
Transversion event substituting a
Thymine nucleic acid in place of a Guanine.

Indels

Reference	ACTGACGCATGCATCATGCATGC
Insertion	ACTGACGCATG GTA CATCATGCATGC
Deletion	ACTGACG--TGCATCATGCATGC

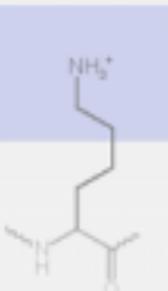
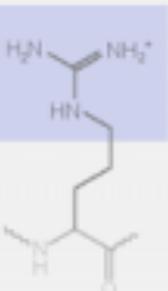
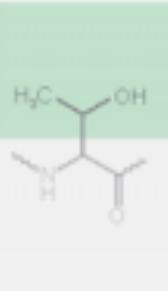


Indels affect a string of base-pairs.
GTA has been inserted
CA has been deleted

Vocabulary

- Synonymous change – no change in aa seq
 - GCT and GCC code for alanine
 - May affect translation due to different codon usage and rarity of some tRNAs
- Nonsense – coding codon to stop codon
 - GGA glycine to TGA stop codon
 - Truncated protein – usually nonsense mediated decay
- Missense – change in aa seq
 - ACC threonine to AAC asparagine

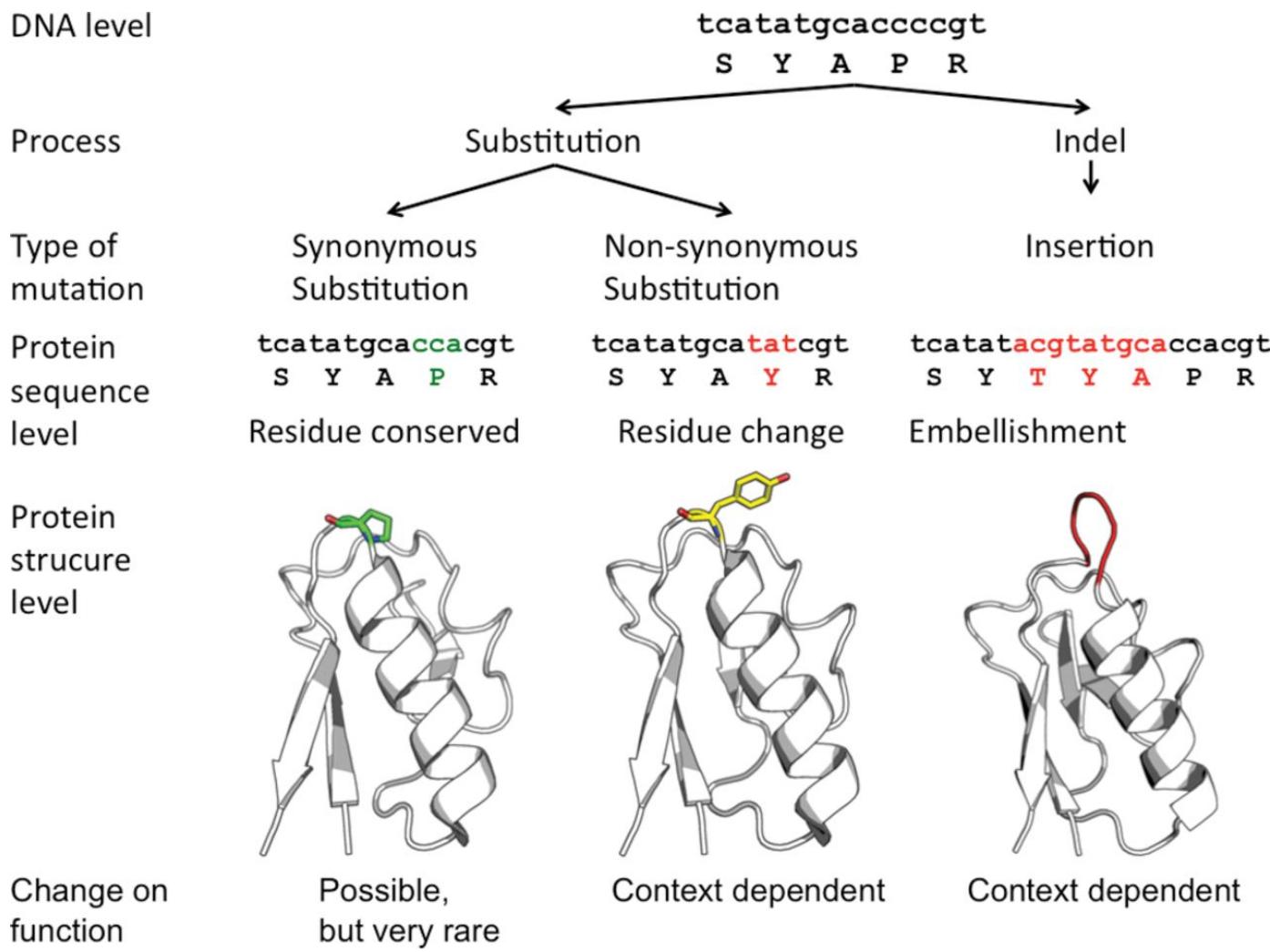
Point mutations

No mutation	Point mutations				
	Silent	Nonsense	Missense		
			conservative	non-conservative	
DNA level	TTC	TTT	ATC	TCC	TGC
mRNA level	AAG	AAA	UAG	AGG	ACG
protein level	Lys	Lys	STOP	Arg	Thr
					

Proteins must

- Fold correctly
- Be stable (but not too stable)
- Bind to other proteins or ligands in the cell
- Have a functional catalytic site
- Assemble correctly (oligomerization)

Possible effects of mutations on proteins.



Romain A. Studer et al. Biochem. J. 2013;449:581-594

Effect of variations on protein structure

Protein Structures

Protein Structure

- Why protein structure?
- The basics of protein sequence structure
- Levels of protein structure
- Sequence and structure
- Conservation and variation
- Classification of protein structures

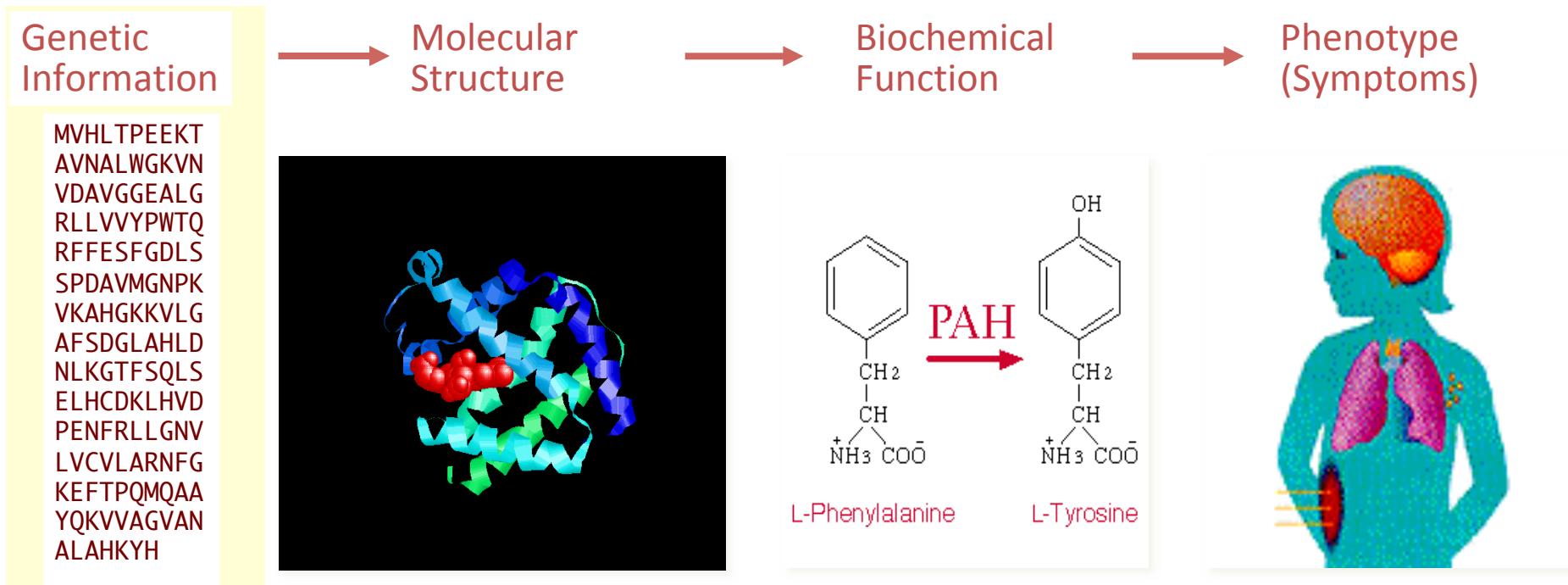
Why protein structure?

- In the factory of living cells, proteins are the workers, performing a variety of biological tasks.
- Each protein has a particular 3-D structure that determines its function.

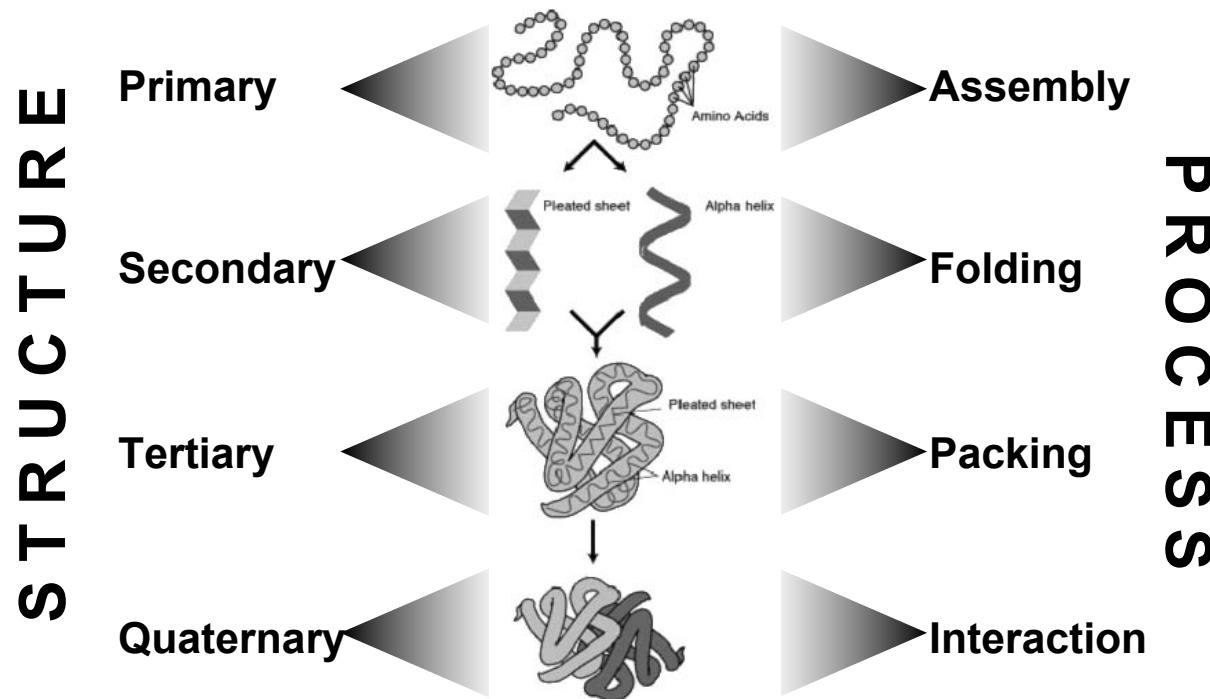
Sequence → Structure → Function

- Protein structure is more conserved than protein sequence, and more closely related to function.

Central Paradigm of Bioinformatics



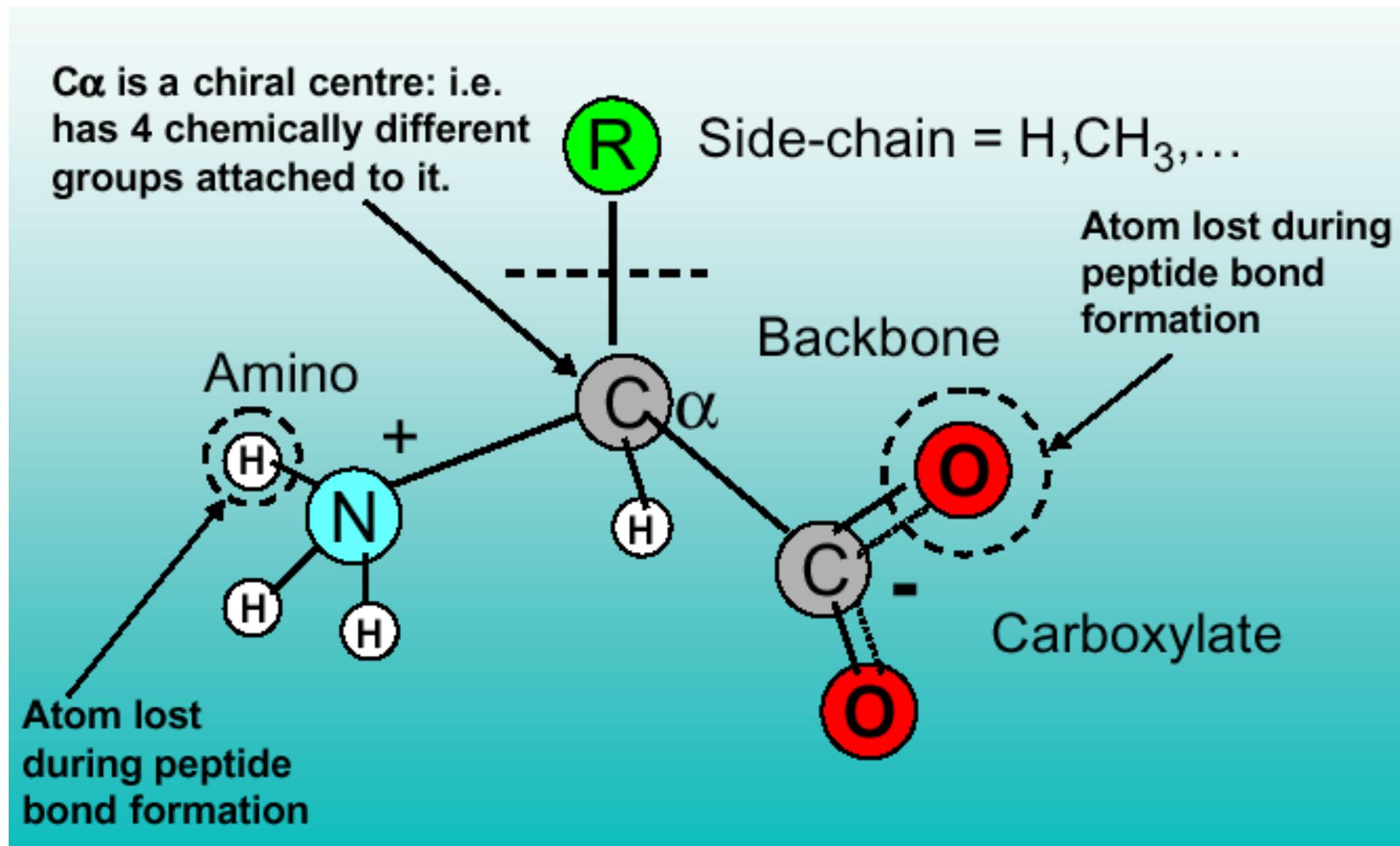
Biology/Chemistry of Protein Structure



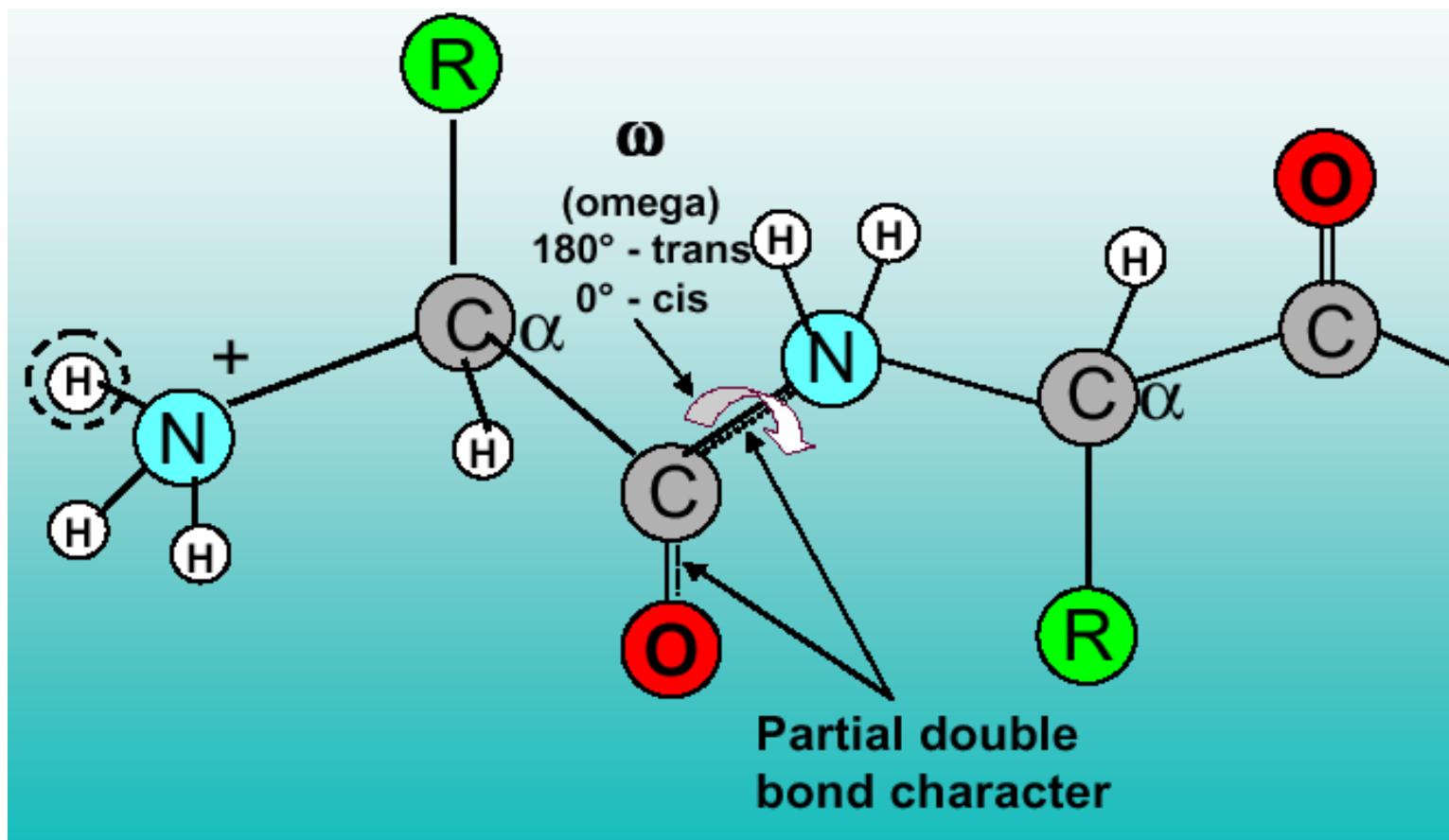
Some basic information

- Proteins are linear heteropolymers: one or more polypeptide chains
- Building blocks: 20 types of amino acids.
- Range from a few 10s-1000s aa
- Three-dimensional shapes (“fold”) adopted vary enormously.

Common structure of Amino Acid



Formation of polypeptide chain

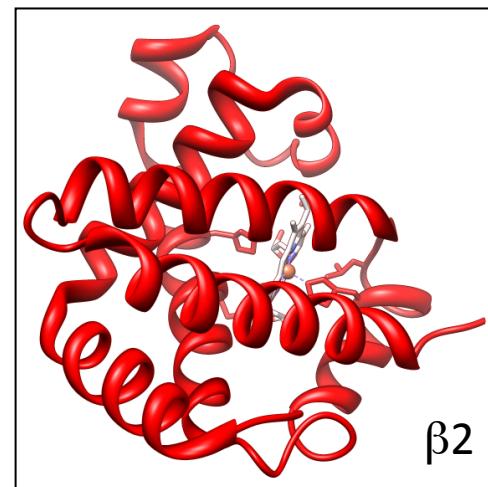


Four Levels of Protein Structure

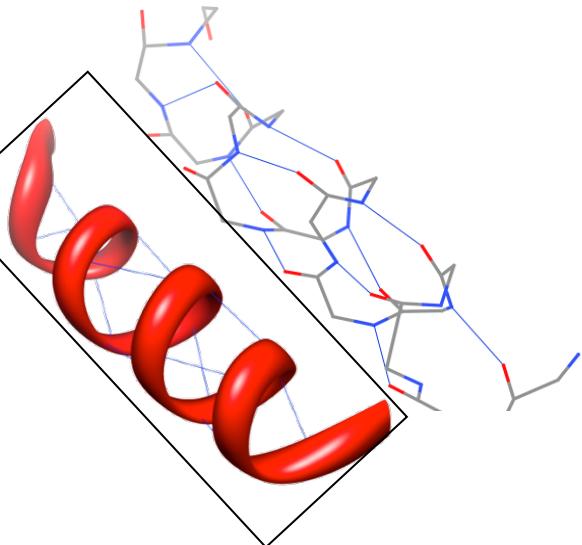
- *Primary, 1°*

TPEEKSAVTALWGKV

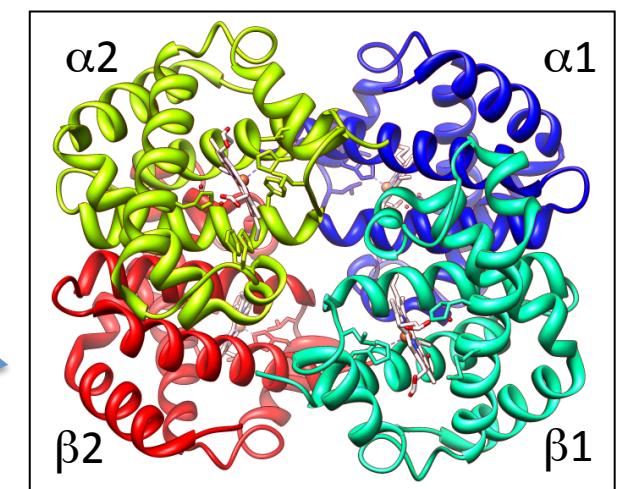
- *Secondary, 2°*



- *Tertiary, 3°*



- *Quaternary, 4°*



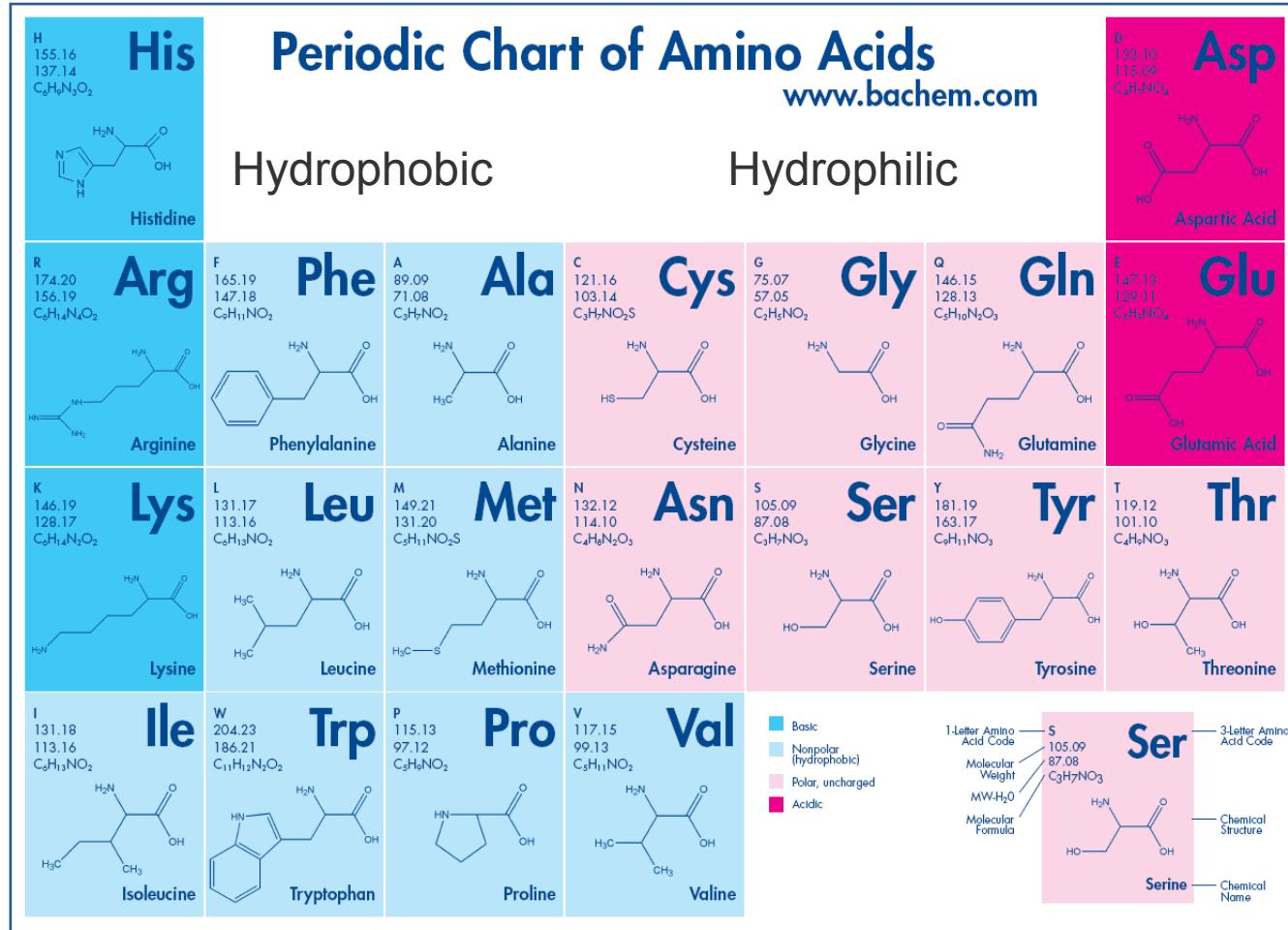
Primary structure

- This is simply the amino acid sequences of polypeptides chains (proteins).

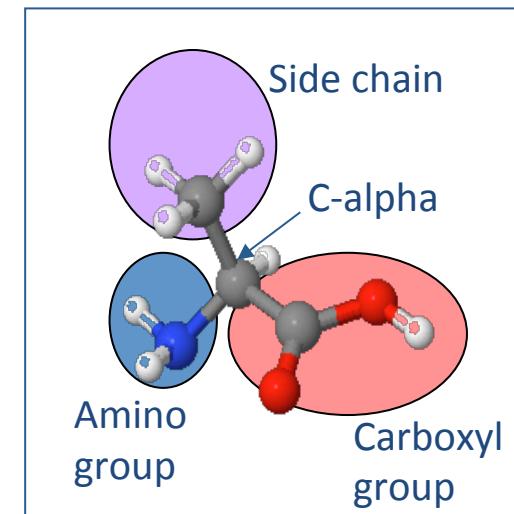
MHGA_YRTPRSKT_DA_YG_CQ_IL_ET_RA_S

Protein Building Blocks: Amino Acids

Basic

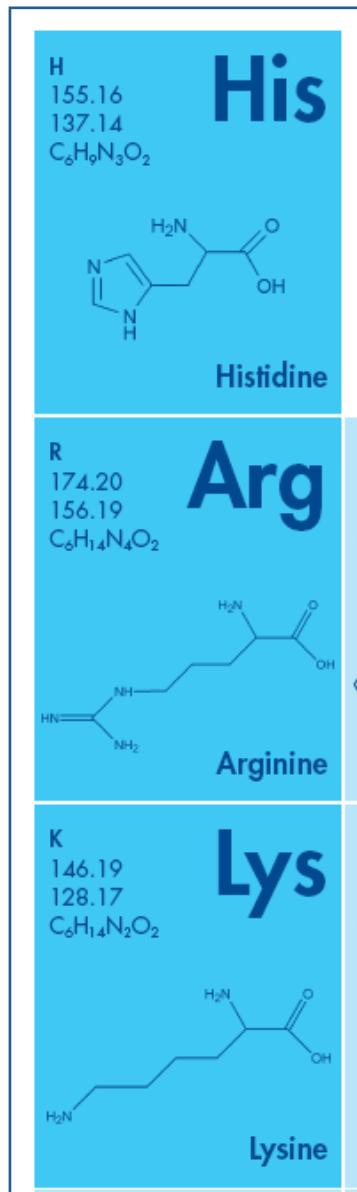


Acidic



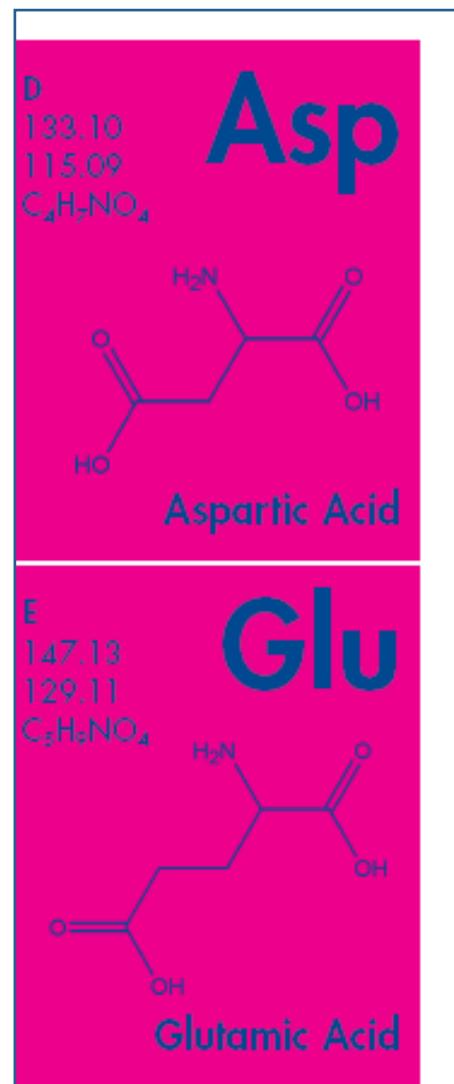
Protein Building Blocks: Amino Acids

Basic



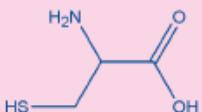
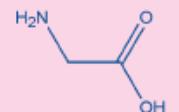
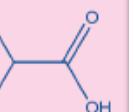
Protein Building Blocks: Amino Acids

Acidic



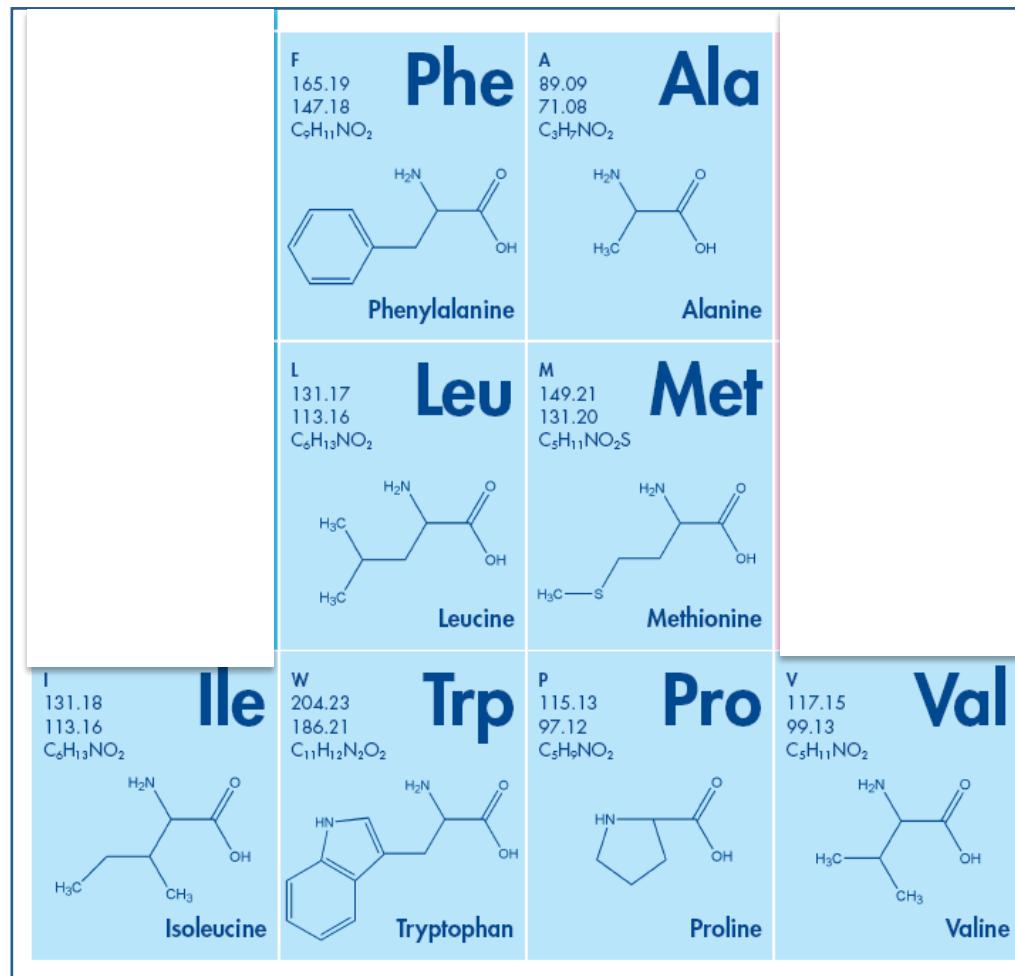
Protein Building Blocks: Amino Acids

Hydrophilic

Cys C 121.16 103.14 $\text{C}_3\text{H}_7\text{NO}_2\text{S}$  Cysteine	Gly G 75.07 57.05 $\text{C}_2\text{H}_5\text{NO}_2$  Glycine	Gln Q 146.15 128.13 $\text{C}_5\text{H}_{10}\text{N}_2\text{O}_3$  Glutamine	
Asn N 132.12 114.10 $\text{C}_4\text{H}_6\text{N}_2\text{O}_3$  Asparagine	Ser S 105.09 87.08 $\text{C}_3\text{H}_7\text{NO}_3$  Serine	Tyr Y 181.19 163.17 $\text{C}_9\text{H}_{11}\text{NO}_3$  Tyrosine	Thr T 119.12 101.10 $\text{C}_4\text{H}_9\text{NO}_3$  Threonine

Protein Building Blocks: Amino Acids

Hydrophobic

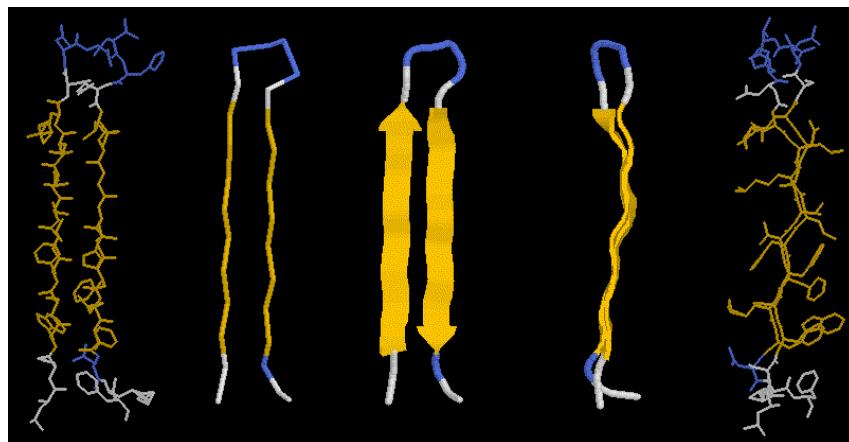


Secondary structure

- Local organization of protein backbone: α -helix, β -strand (groups of β -strands assemble into β -sheet), turn and interconnecting loop.



an α -helix



various representations and orientations of a two stranded β -sheet.

β -Sheet (parallel)

All strands run in the same direction

Catechol O-Methyltransferase



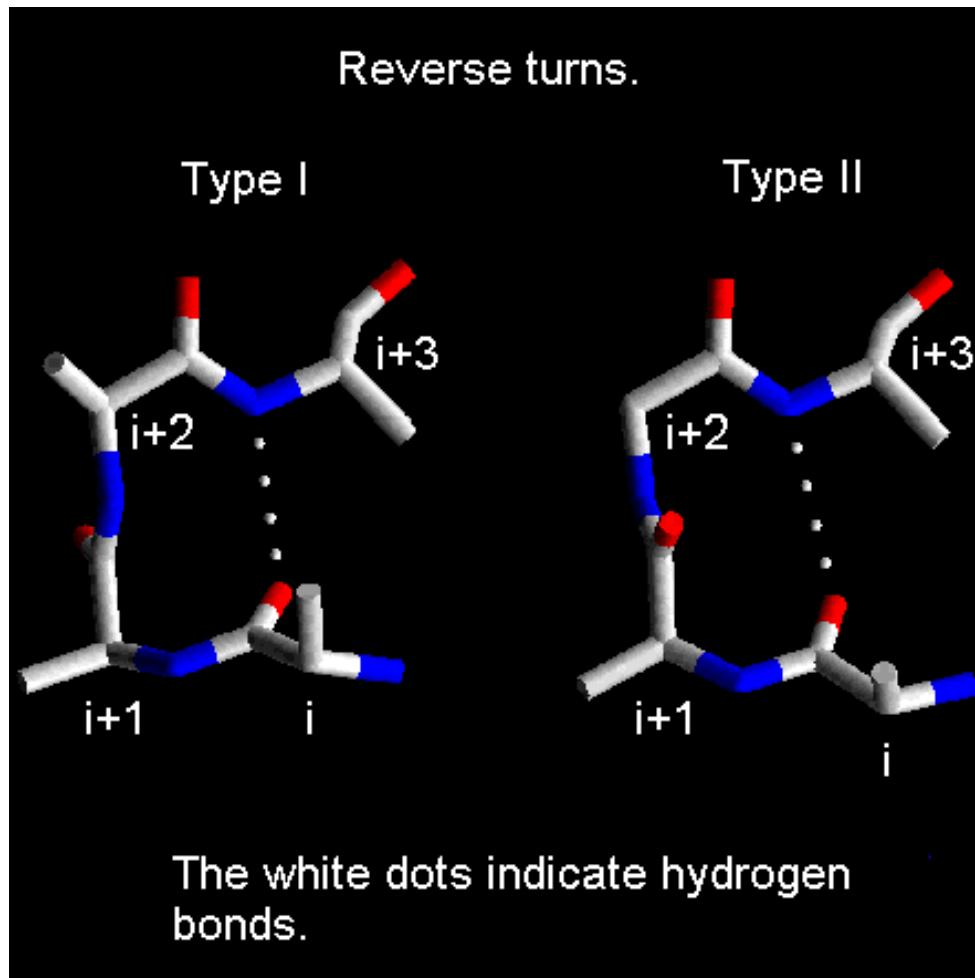
β -Sheet (antiparallel)

All strands run in
the opposite
direction, more
stable

Urate oxidase



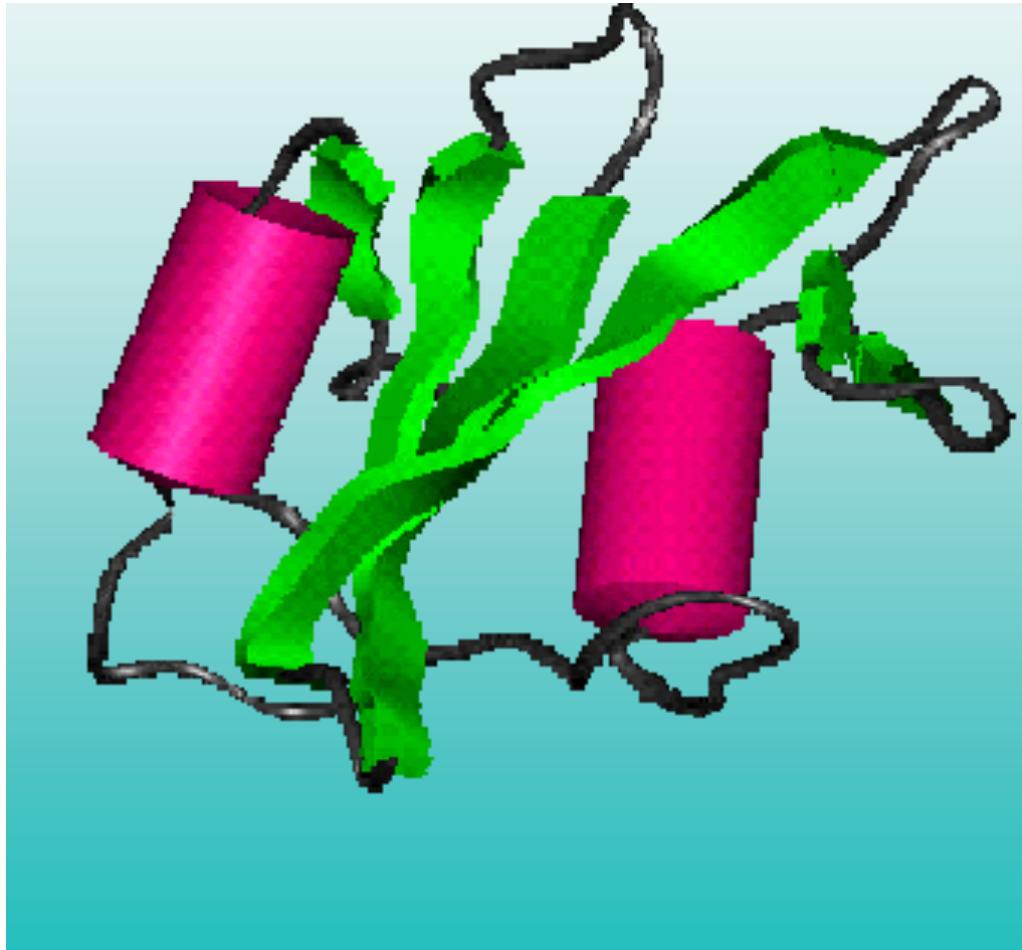
Loops and Turns



Loops: often contain hydrophilic residue on the surface of proteins

Turns: loops with less than 5 residues and often contain G, P

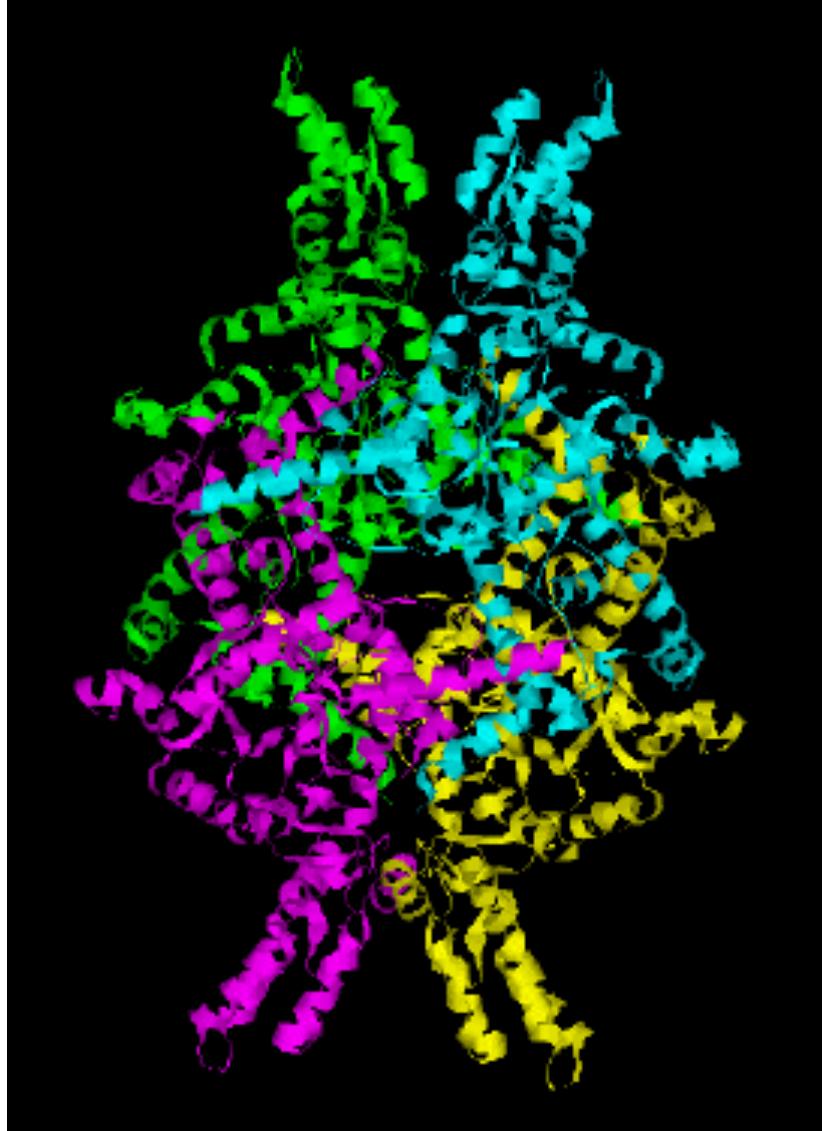
Tertiary structure



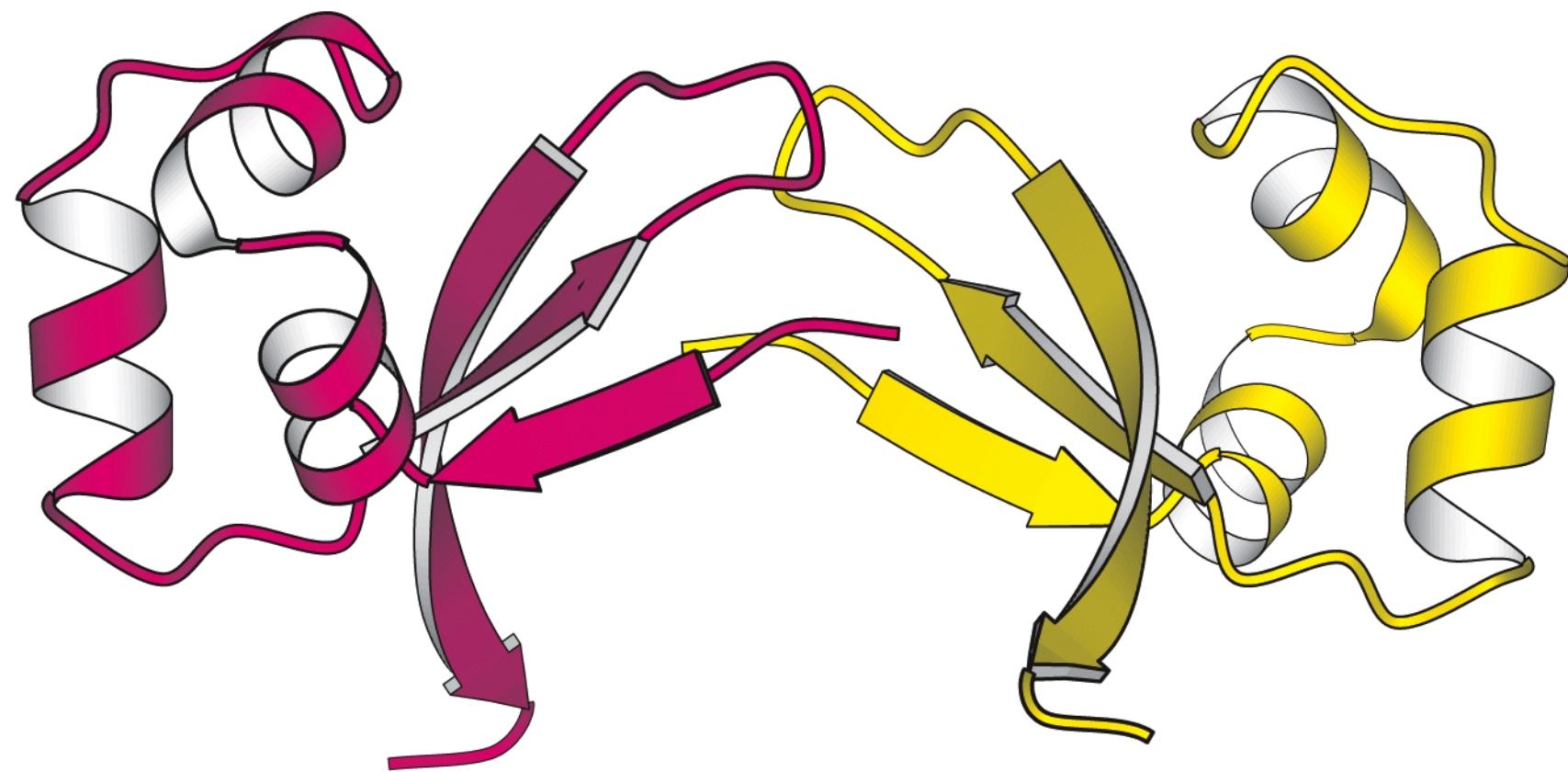
Packing the secondary structure elements into a compact spatial unit

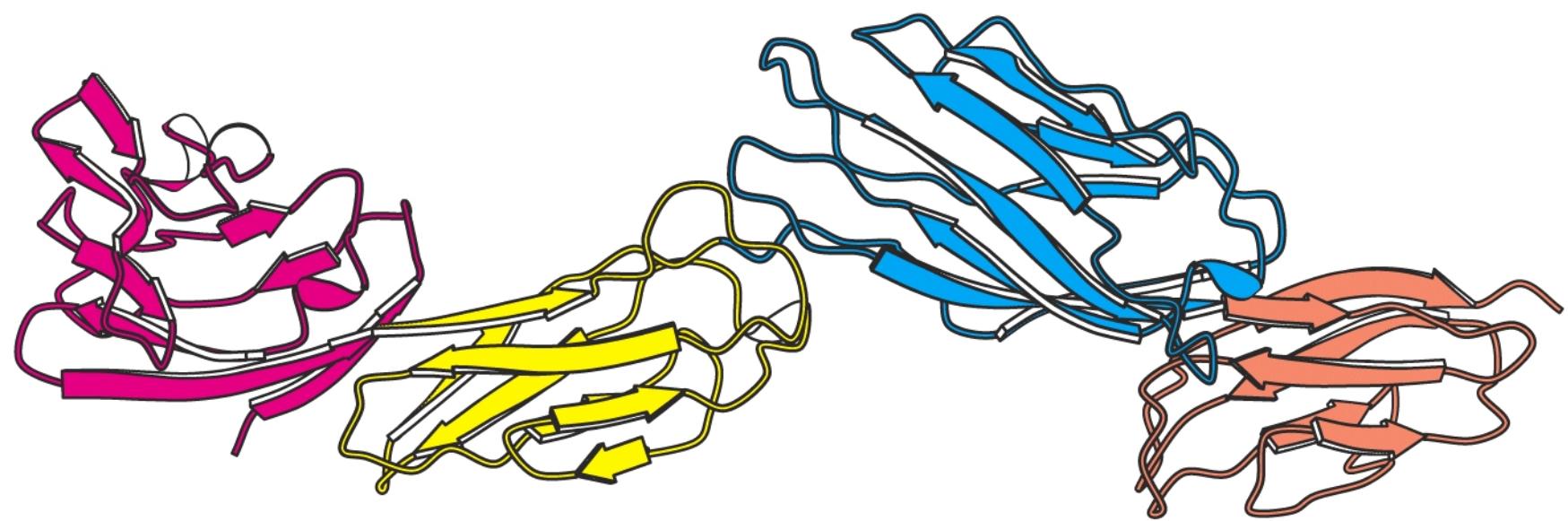
“Fold” or domain– this is the level to which structure prediction is currently possible.

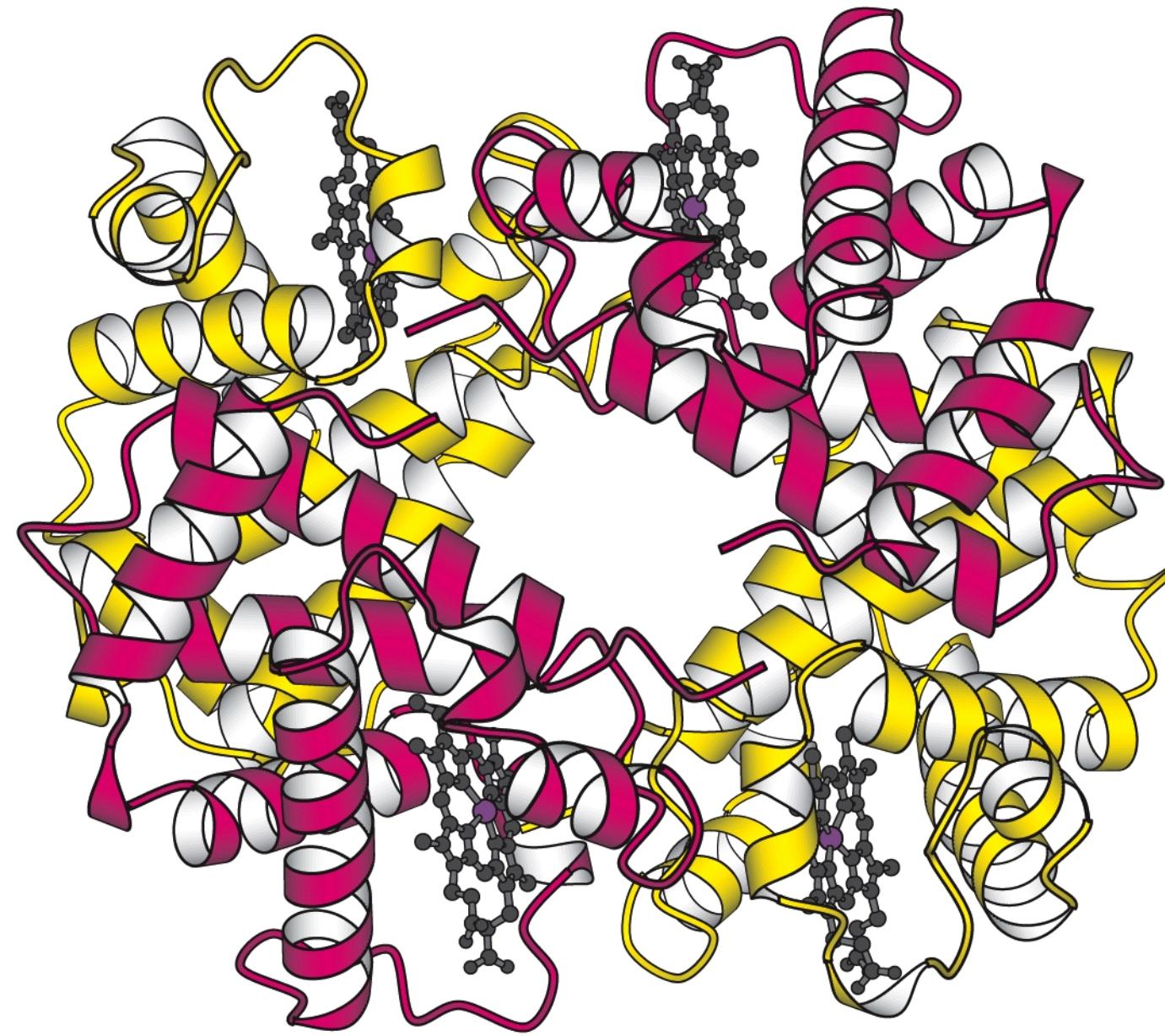
Quaternary structure



- Assembly of homo or heteromeric protein chains.
- Usually the functional unit of a protein, especially for enzymes



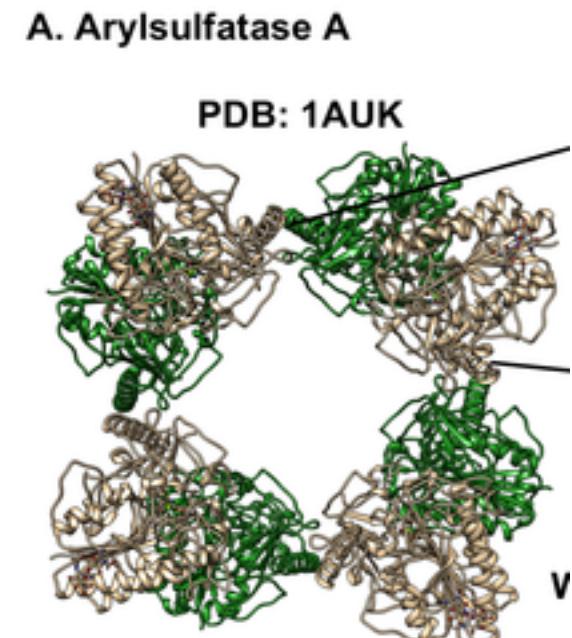




Protein Structure-Function:

Vocabulary check

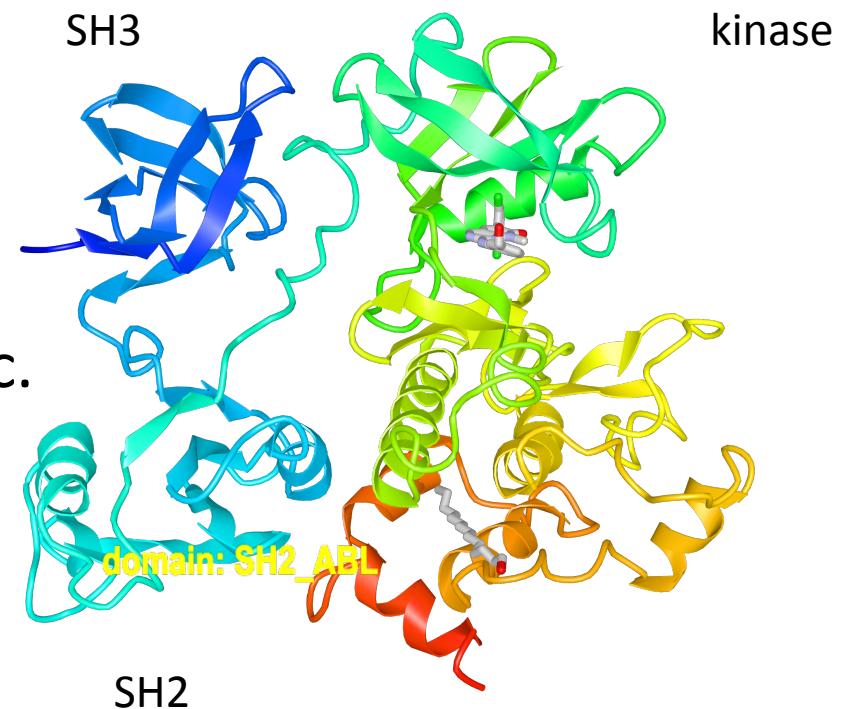
- Protein structure:
 - Chain
 - Motif
 - Domain
 - Oligomer – dimer, trimer etc.
- Protein function:
 - Binding
 - Catalysis
 - Regulation
 - Structural support



Protein Structure-Function:

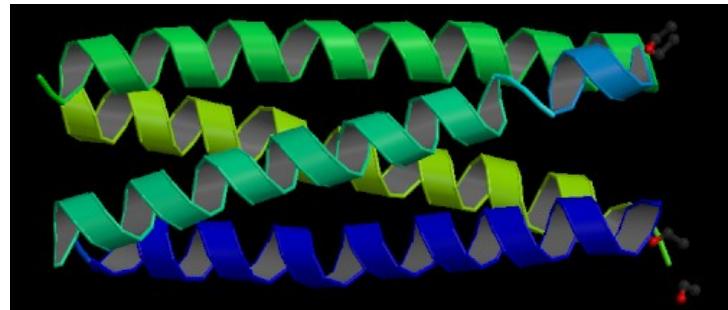
Vocabulary check

- Protein structure:
 - Chain
 - Motif
 - Domain
 - Oligomer – dimer, trimer etc.
- Protein function:
 - Binding
 - Catalysis
 - Regulation
 - Structural support

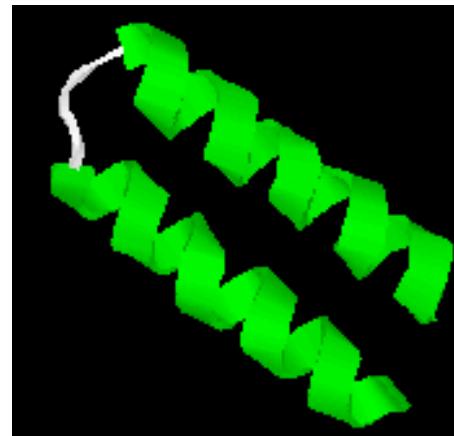


Abl kinase – a multidomain protein

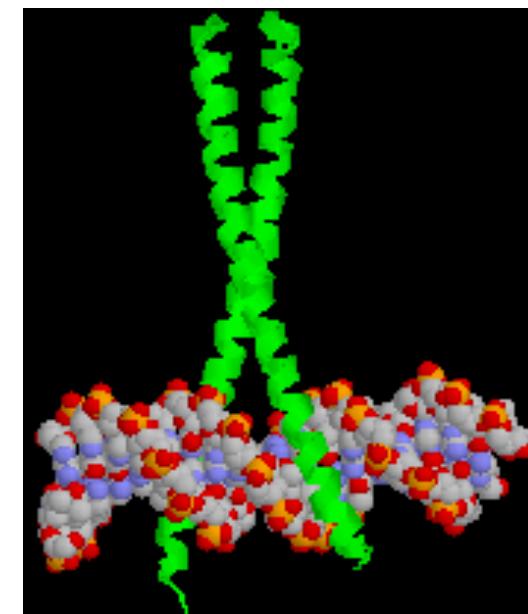
Motifs



Four helix bundle



Helix-loop-helix



Coiled coil

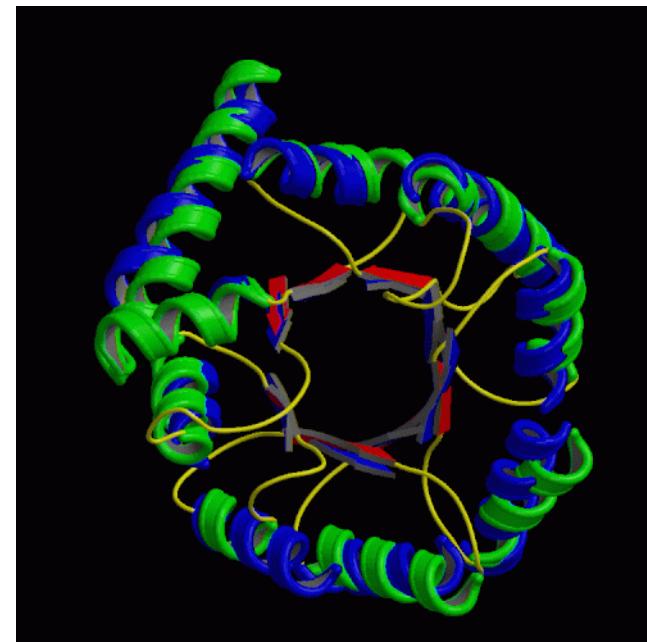
“structure” usually refers to
3-D structure of protein.

Structure is better conserved than sequence

Structure can tolerate a wide range of mutations.

Physical forces favor certain structures.

Number of folds is limited.
Currently ~700
Total: 1,000 ~10,000

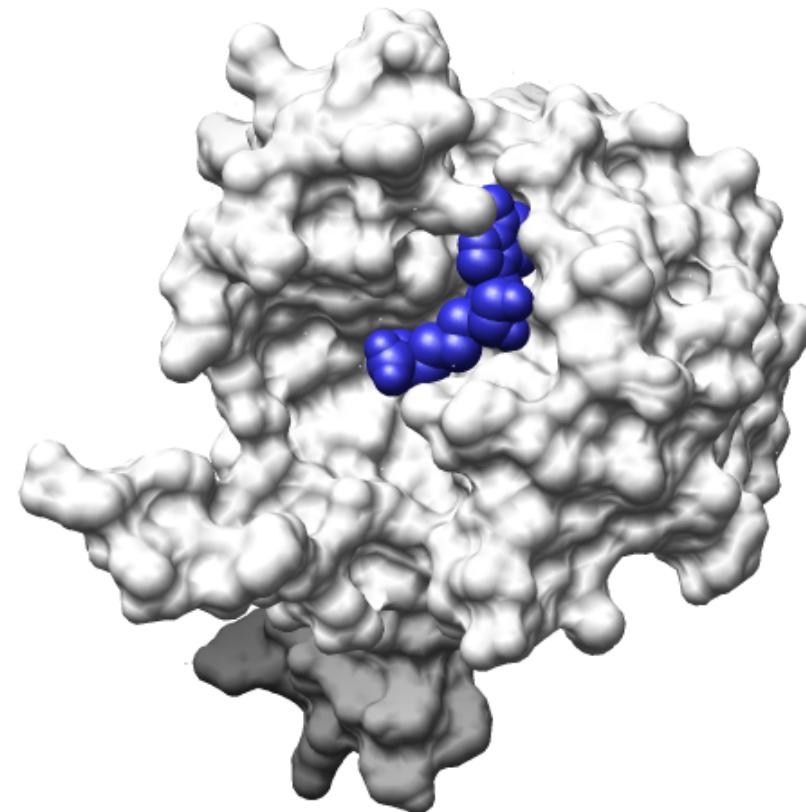


Investigating & visualizing protein structures

Protein 3D Structures

Structure – shape guides function

Binding pockets

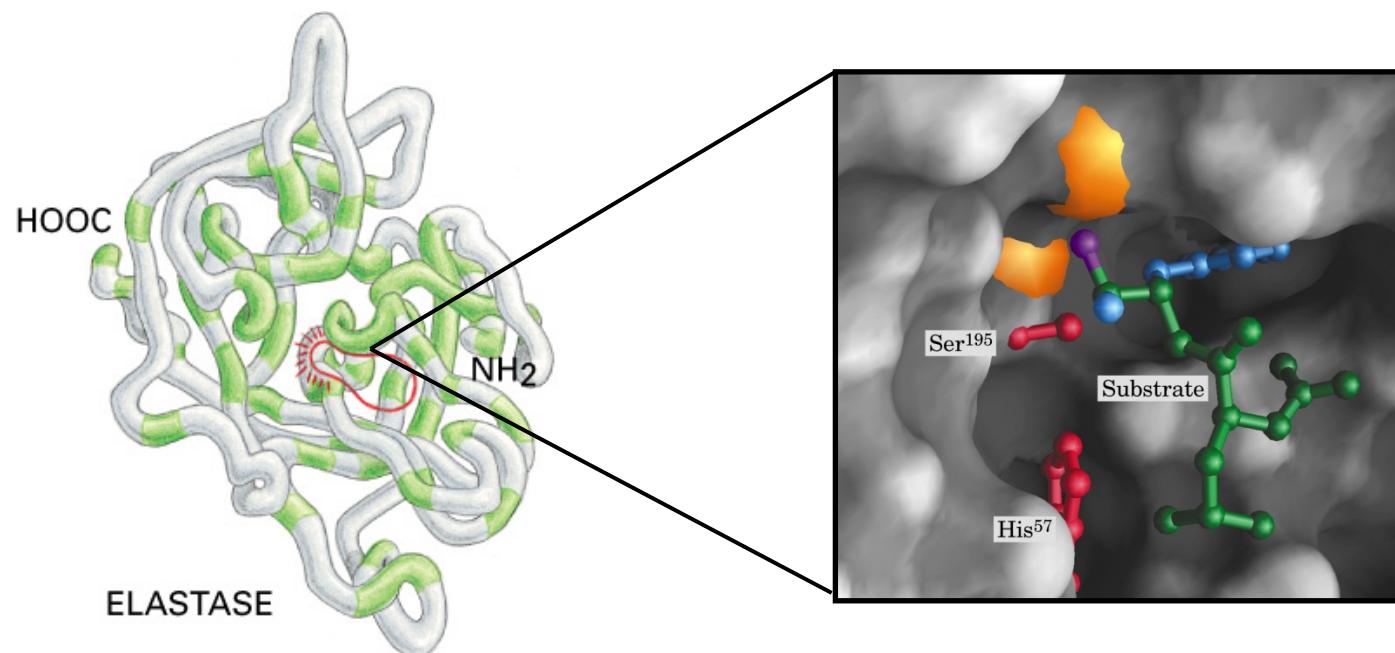


PDB ID 1nw7

Investigating & visualizing protein structures

Protein 3D Structures

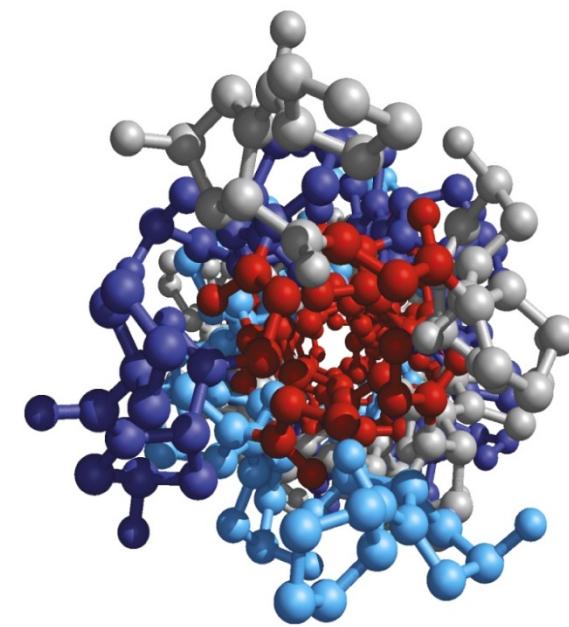
Structure – chemical properties guide function



Investigating & visualizing protein structures

Protein 3D Structures

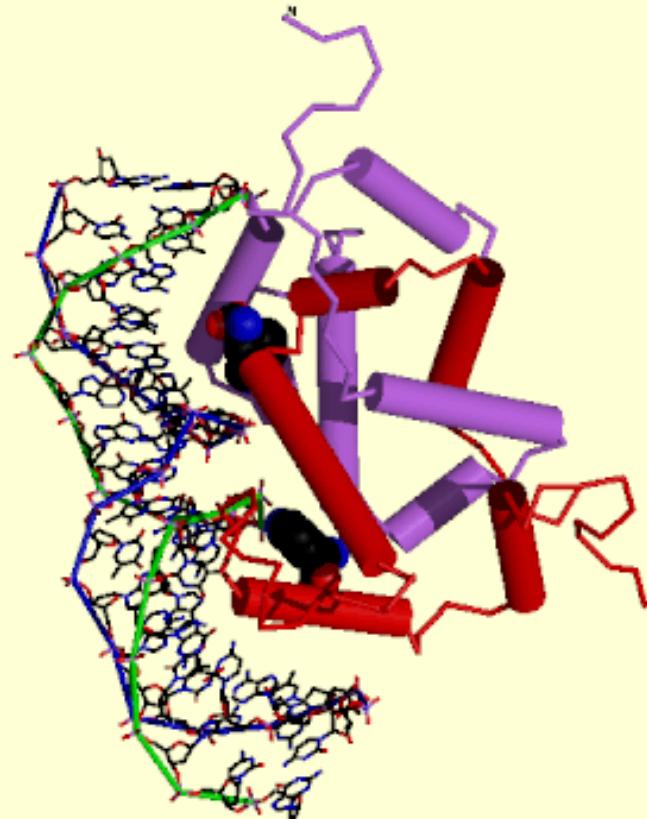
Structure – global fold guides function



Structure-Function Relationship

Certain level of function can be found without structure. But the structure is a key to understand the detailed mechanism.

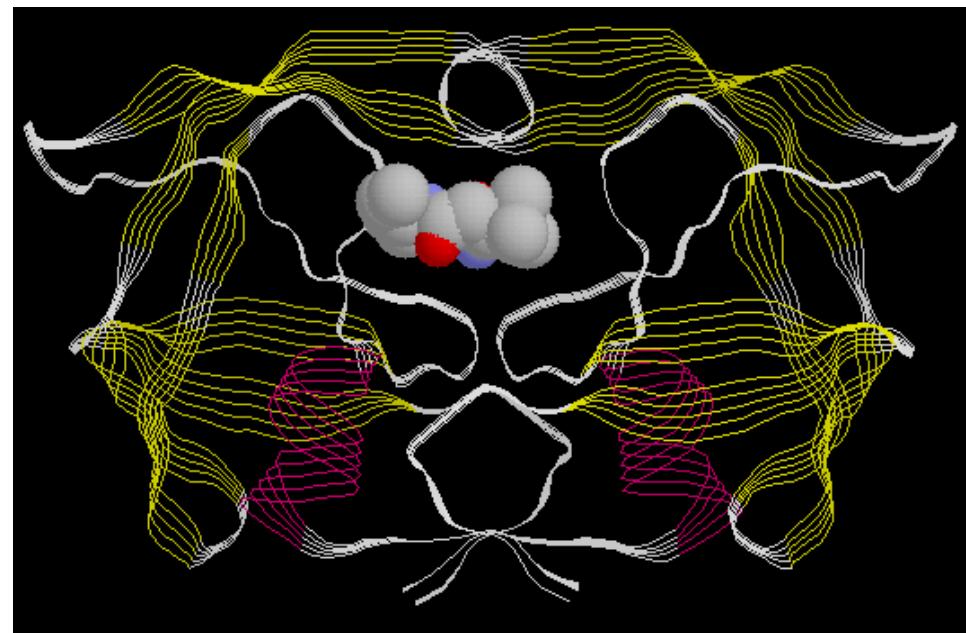
structure is a powerful tool for function inference.



Trp repressor as a function switch

Structure-Based Drug Design

**Structure-based rational
drug design is a major
method for drug discovery.**

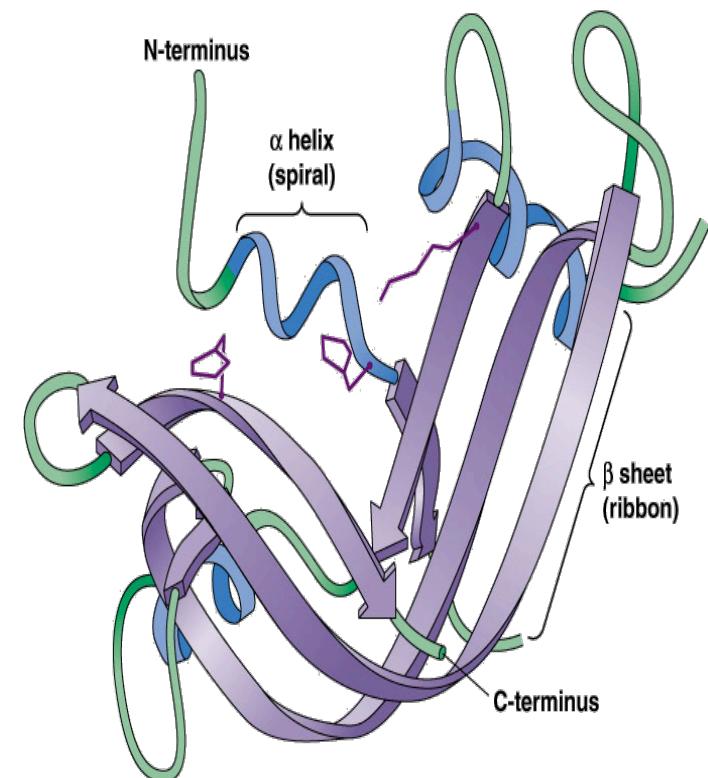


HIV protease inhibitor

Investigating & visualizing protein structures

Tertiary structure = protein fold

- Structure is more **conserved**
- Structure can be used to detect distant **evolutionary** relationships
- Structure can give information on protein **function**
- Structure can be used for **drug design**



(b) Spiral-and-ribbon model

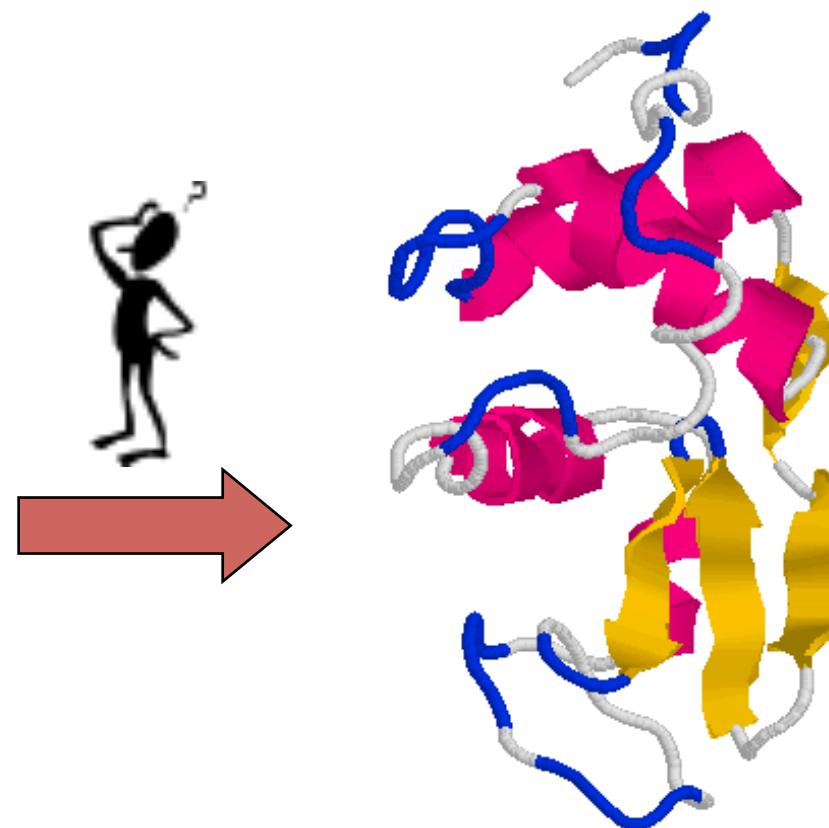
Copyright © 2003 Pearson Education, Inc., publishing as Benjamin Cummings.

Protein Folding Problem

A protein folds into a unique 3D structure under the physiological condition: determine this structure

Lysozyme sequence:

```
KVFGRCELAA AMKRHGLDNY  
RGYSLGNWVC AAKFESNFNT  
QATNRNTDGS TDYGILQINS  
RWWCNDGRTP GSRNLNCNIPC  
SALLSSDITA SVNCAKKIVS  
DGNGMNAWVA WRNRCKGTDV  
QAWIRGCRL
```



Levinthal's paradox

- Consider a 100 residue protein. If each residue can take only 3 positions, there are $3^{100} = 5 \times 10^{47}$ possible conformations.
 - If it takes 10^{-13} s to convert from 1 structure to another, exhaustive search would take 1.6×10^{27} years!
- Folding must proceed by progressive stabilization of intermediates.

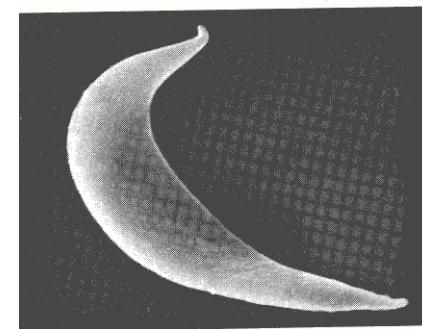
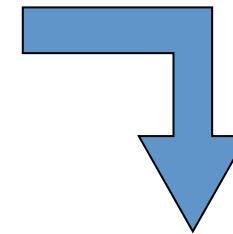
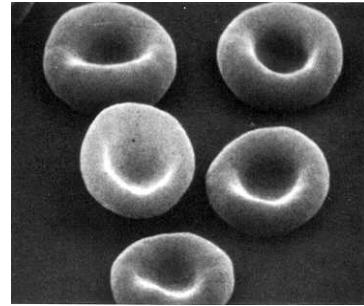
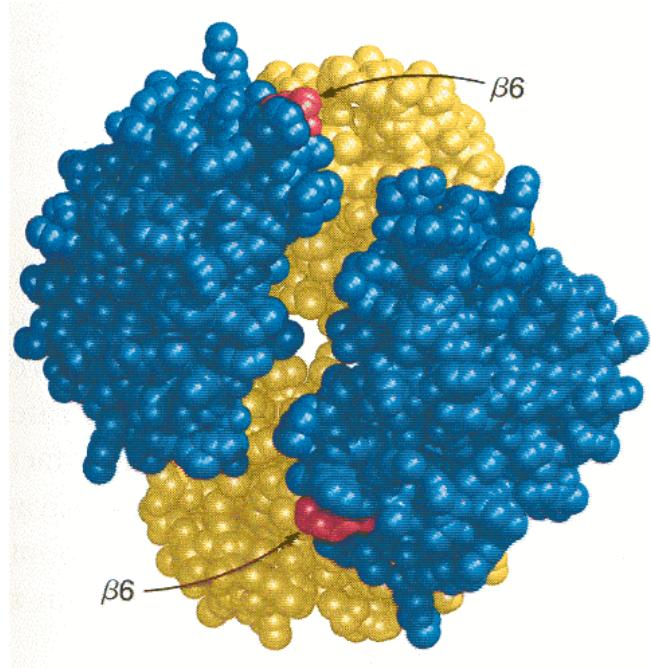
Forces driving protein folding

- It is believed that *hydrophobic collapse* is a key driving force for protein folding
 - Hydrophobic core
 - Polar surface interacting with solvent
- Minimum volume (no cavities)
- Disulfide bond formation stabilizes
- Hydrogen bonds
- Polar and electrostatic interactions

Effect of a single mutation

- Hemoglobin is the protein in red blood cells (erythrocytes) responsible for binding oxygen.
- The mutation E→V in the β chain replaces a charged Glu by a hydrophobic Val on the surface of hemoglobin
- The resulting “sticky patch” causes hemoglobin to agglutinate (stick together) and form fibers which deform the red blood cell and do not carry oxygen efficiently
- Sickle cell anemia was the first identified molecular disease

Sickle Cell Anemia



Sequestering hydrophobic residues in the protein core protects proteins from hydrophobic agglutination.

Structural information

- Protein Data Bank: maintained by the Research Collaboratory of Structural Bioinformatics(RCSB)
 - <http://www.rcsb.org/pdb/>
 - As of Sunday Sep 30, 2018 there are 144682 Structures
 - including structures of Protein/Nucleic Acid Complexes, Nucleic Acids, Carbohydrates
- Each structure has a **PDB ID**: a 4 character unique identifier
- Most structures are determined by X-ray crystallography (about 130000 entries). Some other methods are NMR (about 12000 entries) and electron microscopy(EM) (about 2500 entries). Theoretically predicted structures were removed from PDB a few years ago.

Structural information

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

RCSB PDB PROTEIN DATA BANK 144682 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

PDB-101 Worldwide Protein Data Bank EMDDataBank Nucleic Acid Database Worldwide Protein Data Bank Foundation

Search by PDB ID, author, macromolecule, sequence, or ligands Go Advanced Search | Browse by Annotations

f t y q

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

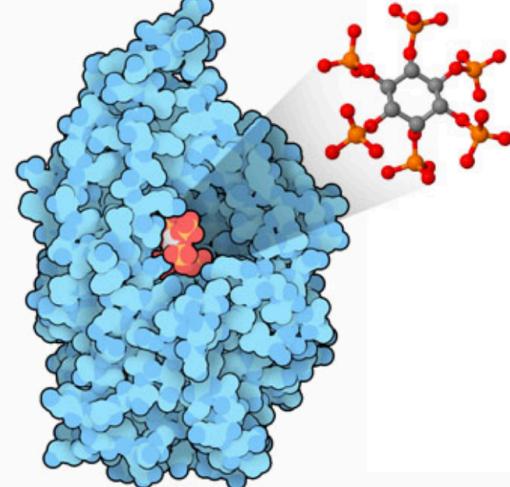
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

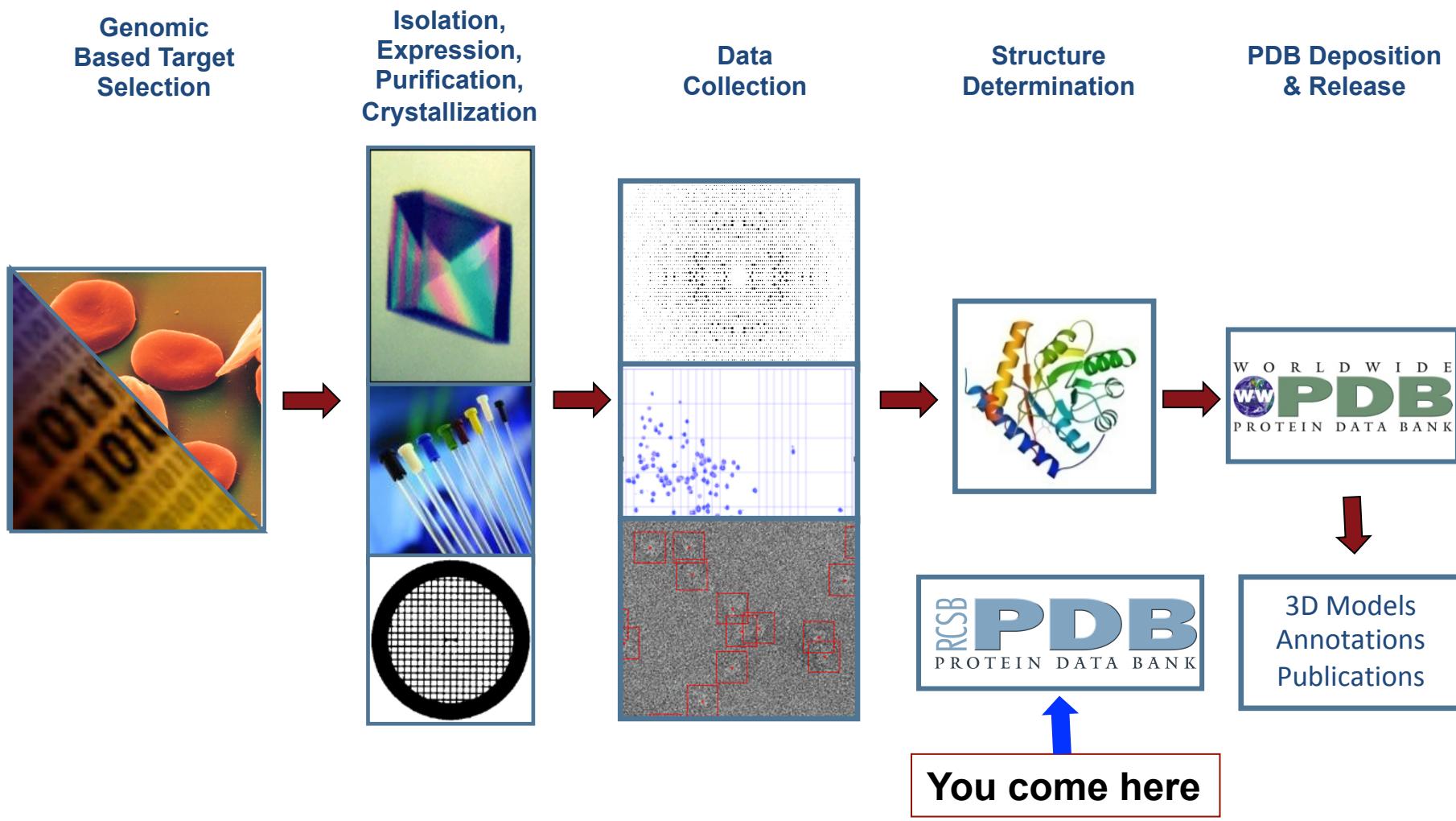
Openings with RCSB PDB at UCSD

JOIN OUR TEAM

September Molecule of the Month



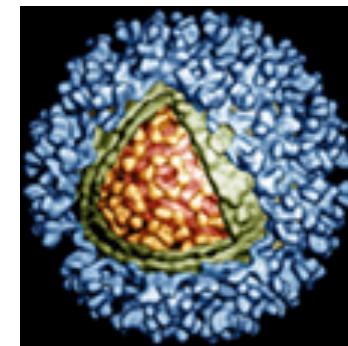
The Data Pipeline



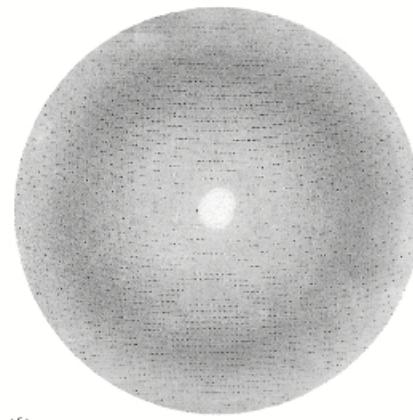
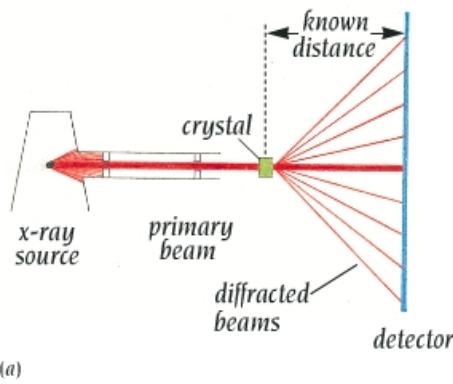
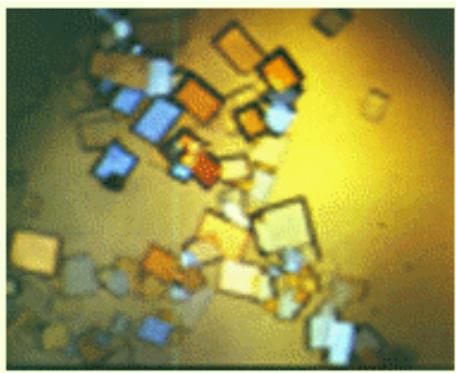
www.rcsb.org • info@rcsb.org

Experimental techniques for structure determination

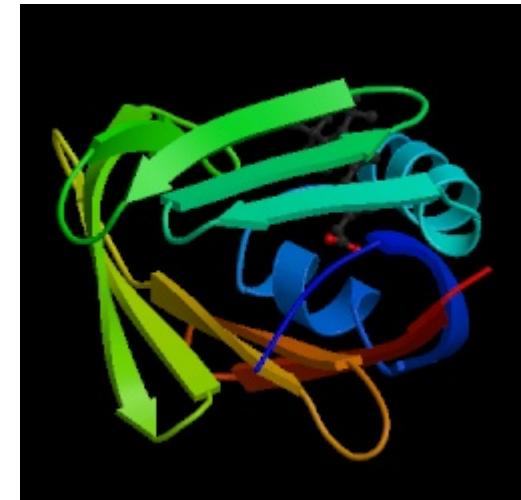
- X-ray Crystallography
- Nuclear Magnetic Resonance spectroscopy (NMR)
- Electron Microscopy/Diffraction
- Free electron lasers ?



X-ray Crystallography



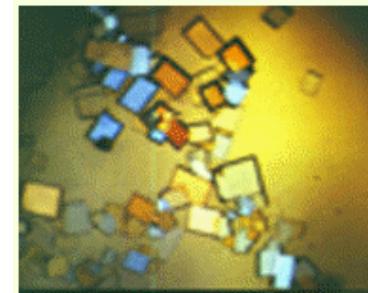
(b)



©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

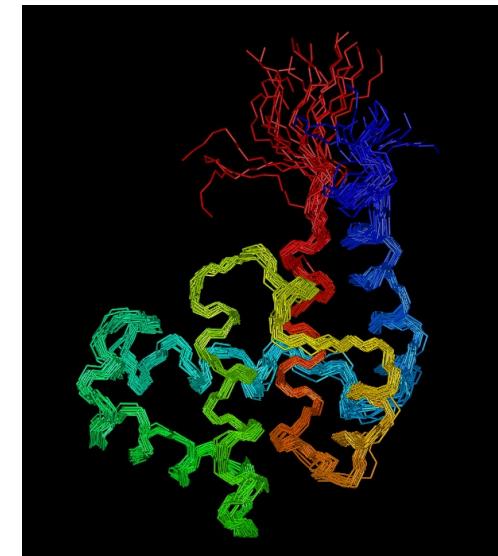
X-ray Crystallography..

- From small molecules to viruses
- Information about the positions of individual atoms
- Limited information about dynamics
- Requires crystals



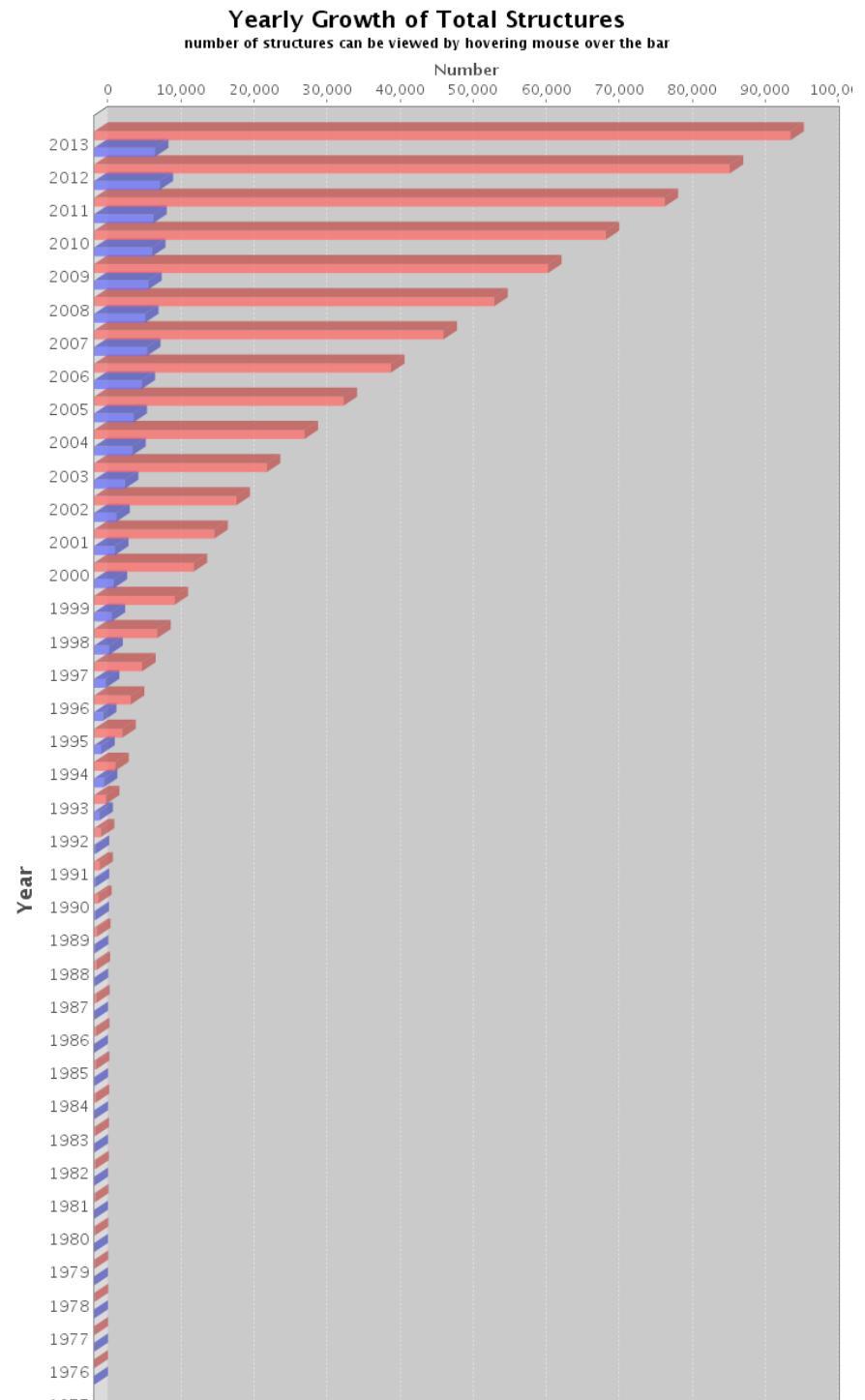
NMR

- Limited to molecules up to ~50kDa (good quality up to 30 kDa)
- Information about distances between pairs of atoms
 - A 2-d resonance spectrum with off-diagonal peaks
- Requires soluble, non-aggregating material



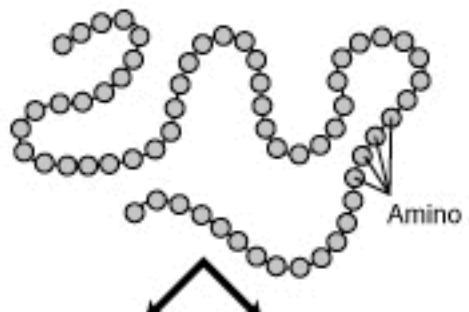
PDB Growth

Red: Total
Blue: Yearly



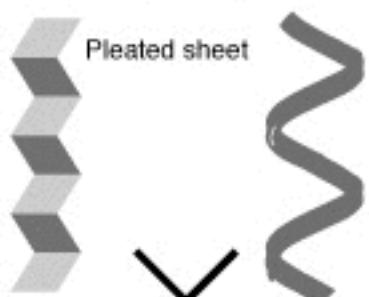
What is in the PDB?

- Coordinate and experimental data files
 - Details about sample preparation, data collection and structure solution
 - Sequence(s) of polymers (proteins and nucleic acids) in the structure
 - Information about ligands in the structure
-
- Links to various resources that describe the sequence, function and other properties of the molecule.
 - Classification of structures by sequence, structure, function and other criteria

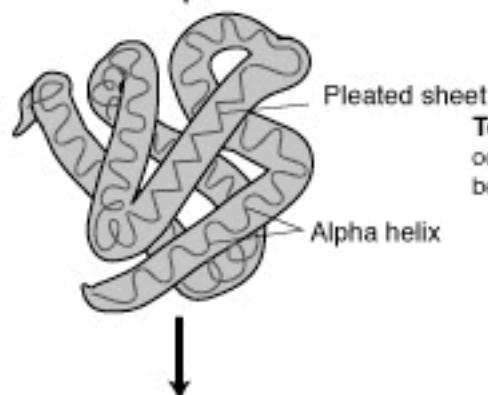


Primary protein structure
is sequence of a chain of amino acids

Amino Acids



Alpha helix



Pleated sheet

Tertiary protein structure
occurs when certain attractions are present
between alpha helices and pleated sheets.

Alpha helix



Quaternary protein structure
is a protein consisting of more than one
amino acid chain.

[Structure Summary](#)[3D View](#)[Annotations](#)[Sequence](#)[Sequence Similarity](#)[Structure Similarity](#)[Experiment](#)

Biological Assembly 1

Display structure

CD View: Structure | Electron Density | Ligand Interaction

Standalone Viewers

Protein Workshop | Ligand Explorer

Global Symmetry: Asymmetric - C1

Global Stoichiometry: Monomer - A

Biological assembly 1 assigned by authors.

Macromolecule Content

- Total Structure Weight: 123122.28
- Atom Count: 6655
- Residue Count: 1074
- Unique protein chains: 1

1OPL

Structural basis for the auto-inhibition of c-Abl tyrosine kinase

DOI: 10.2210/pdb1OPL/pdb

Classification: [TRANSFERASE](#)

Organism(s): [Homo sapiens](#)

Expression System: [Spodoptera frugiperda](#)

Mutation(s): 3

Deposited: 2003-03-06 Released: 2003-04-08

Deposition Author(s): [Nagar, B., Hantschel, O., Young, M.A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., Kuriyan, J.](#)

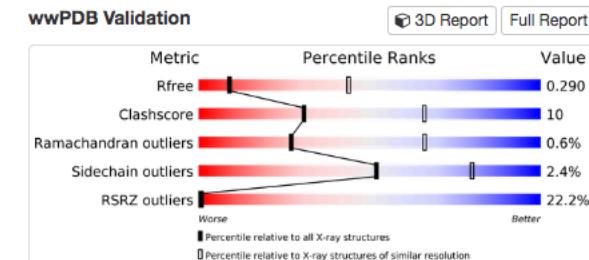
Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 3.42 Å

R-Value Free: 0.315

R-Value Work: 0.306

wwPDB Validation

This is version 1.2 of the entry. See complete history.

Literature[Download Primary Citation](#)

Structural basis for the autoinhibition of c-Abl tyrosine kinase

[Nagar, B., Hantschel, O., Young, M.A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., Kuriyan, J.](#)
(2003) Cell 112: 859-871

PubMed: [12654251](#) [Search on PubMed](#)

Primary Citation of Related Structures:
1OPK, 1OPJ

PubMed Abstract:

c-Abl is normally regulated by an autoinhibitory mechanism, the disruption of which leads to chronic myelogenous leukemia. The details of this mechanism have been elusive because c-Abl lacks a phosphotyrosine residue that triggers the assembly of the ...

Download structure

The paper describing
the structure

Investigating & visualizing protein structures

RCSB- The Protein Data Bank

PDB files have a specific format :

- TITLE
- REMARK
- COMPND
- SEQRES- the original sequence
- HELIX, BETA- secondary structure
- ATOM – The actual protein/DNA/RNA chain
- HETATM- additional atoms such as ligands, water etc.
- ...

PDB Files: the “header”

HEADER OXIDOREDUCTASE (SUPEROXIDE ACCEPTOR) 13-JUL-94
COMPND MANGANESE SUPEROXIDE DISMUTASE (E.C.1.15.1.1) COMPLEXED
COMPND 2 WITH AZIDE
OURCE (THERMUS THERMOPHILUS, HB8)
AUTHOR M.S.LAH, M.DIXON, K.A.PATTRIDGE, W.C.STALLINGS, J.A.FEE,
AUTHOR 2 M.L.LUDWIG
REVDAT 2 15-MAY-95
REVDAT 1 15-OCT-94
JRNL AUTH M.S.LAH, M.DIXON, K.A.PATTRIDGE, W.C.STALLINGS,
JRNL AUTH 2 J.A.FEE, M.L.LUDWIG
JRNL TITL STRUCTURE-FUNCTION IN E. COLI IRON SUPEROXIDE
JRNL TITL 2 DISMUTASE: COMPARISONS WITH THE MANGANESE ENZYME
JRNL TITL 3 FROM T. THERMOPHILUS
JRNL REF TO BE PUBLISHED
REMARK 1 AUTH M.L.LUDWIG, A.L.METZGER, K.A.PATTRIDGE, W.C.STALLINGS
REMARK 1 TITL MANGANESE SUPEROXIDE DISMUTASE FROM THERMUS
REMARK 1 TITL 2 THERMOPHILUS. A STRUCTURAL MODEL REFINED AT 1.8
REMARK 1 TITL 3 ANGSTROMS RESOLUTION
REMARK 1 REF J.MOL.BIOL. V. 219 335 1991
REMARK 1 REFN ASTM JMOBAK UK ISSN 0022-2836
REMARK 1 REFERENCE 2
REMARK 1 AUTH W.C.STALLINGS, C.BULL, J.A.FEE, M.S.LAH, M.L.LUDWIG
REMARK 1 TITL IRON AND MANGANESE SUPEROXIDE DISMUTASES:
REMARK 1 TITL 2 CATALYTIC INFERENCES FROM THE STRUCTURES

- JRNL- reference

PDB files - coordinates

Atom & Residue					
ATOM	7	SD	MET	A	1
ATOM	8	CE	MET	A	1
ATOM	9	N	ILE	A	2
ATOM	10	CA	ILE	A	2
HETATM	3139	C6	SAH		328
HETATM	3140	N6	SAH		328
HETATM	3141	N1	SAH		328
HETATM	3142	C2	SAH		328
HETATM	3143	N3	SAH		328
HETATM	3144	C4	SAH		328
HETATM	3145	O	HOH		329
HETATM	3146	O	HOH		330
HETATM	3147	O	HOH		331

Atom Number

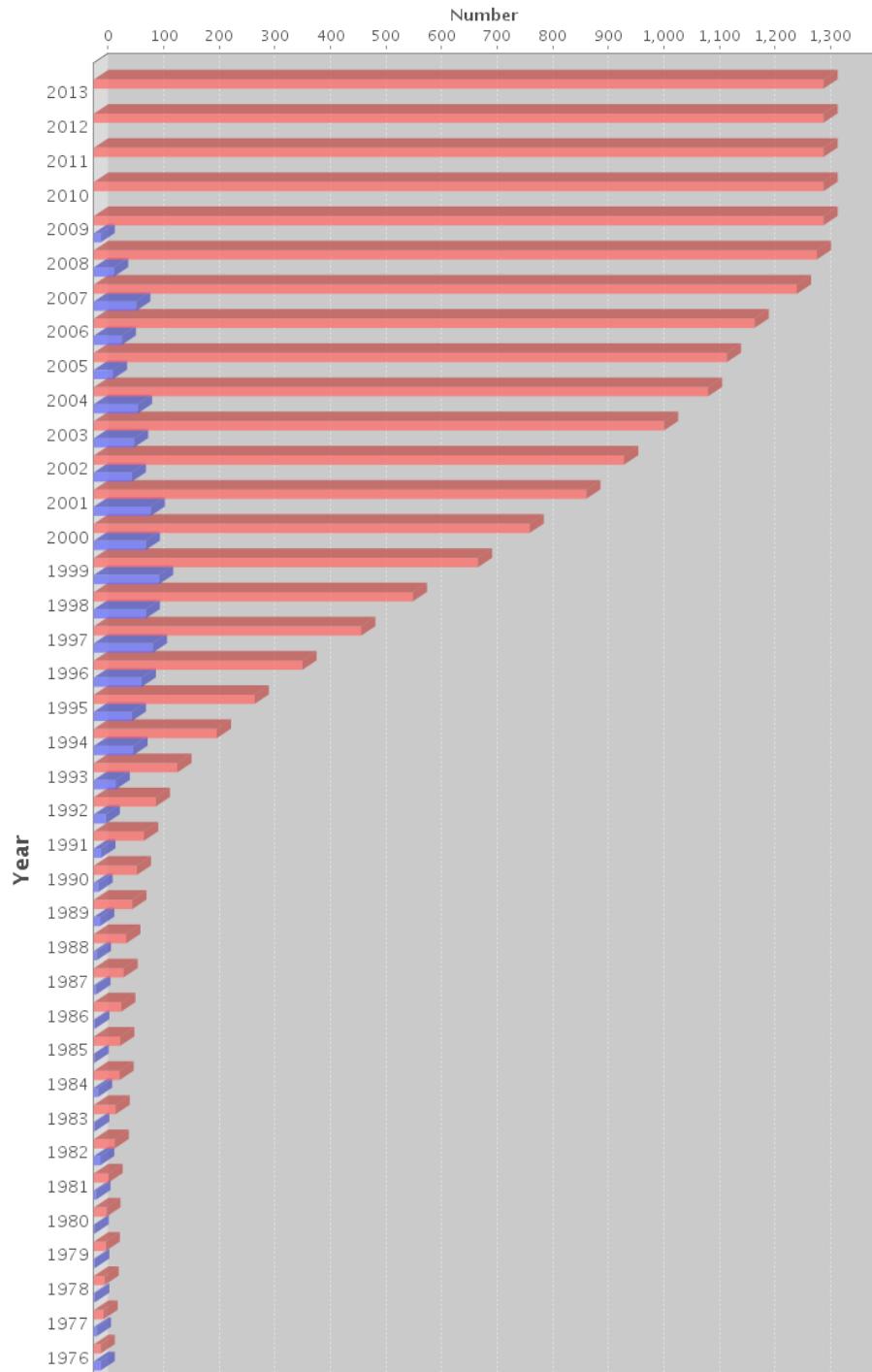
Atom, residue or molecule

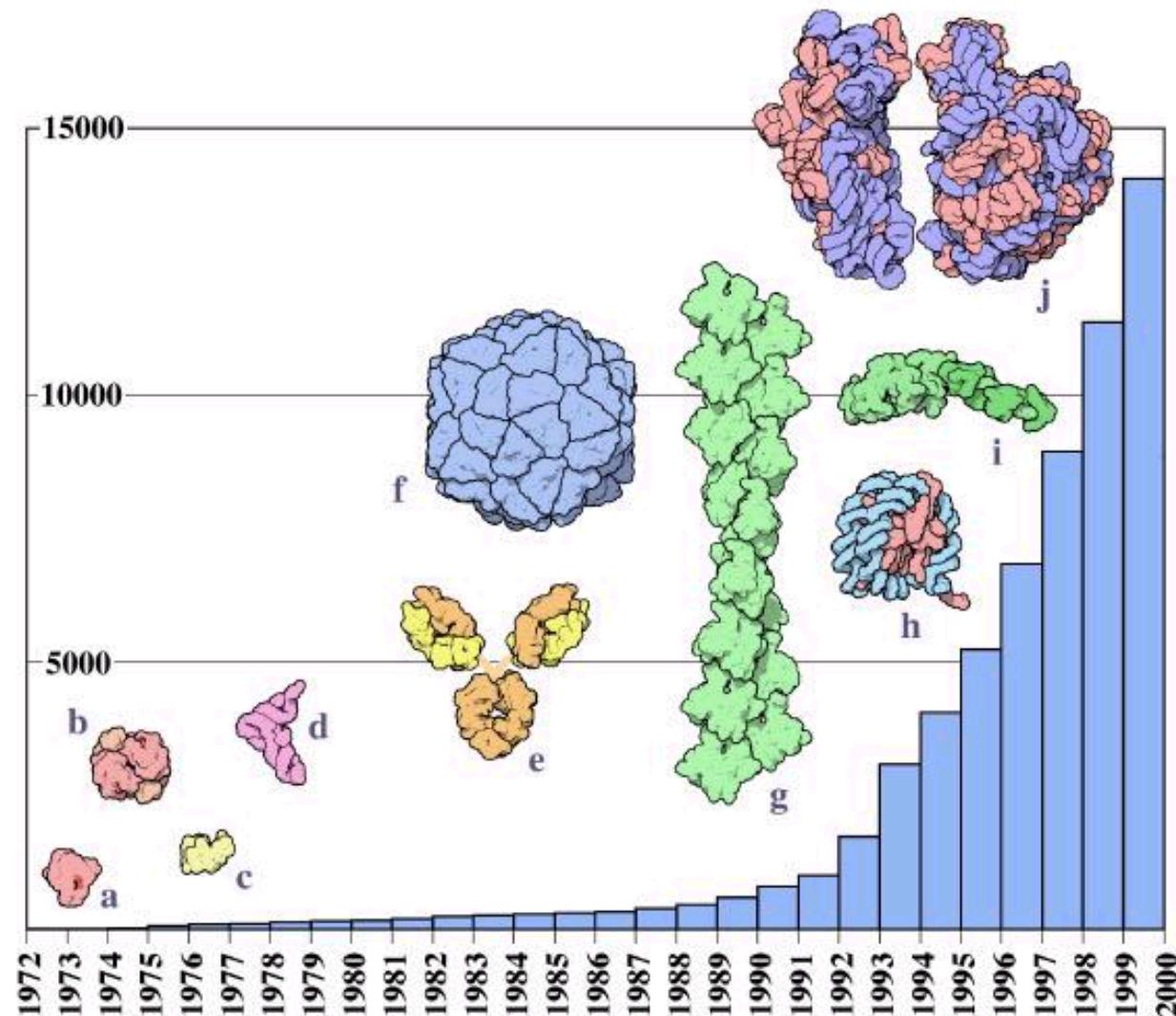
Chain

Coordinates: X, Y, Z

Growth of unique folds (topologies) per year

Number of folds is limited.
Currently ~700
Total: 1,000 ~10,000

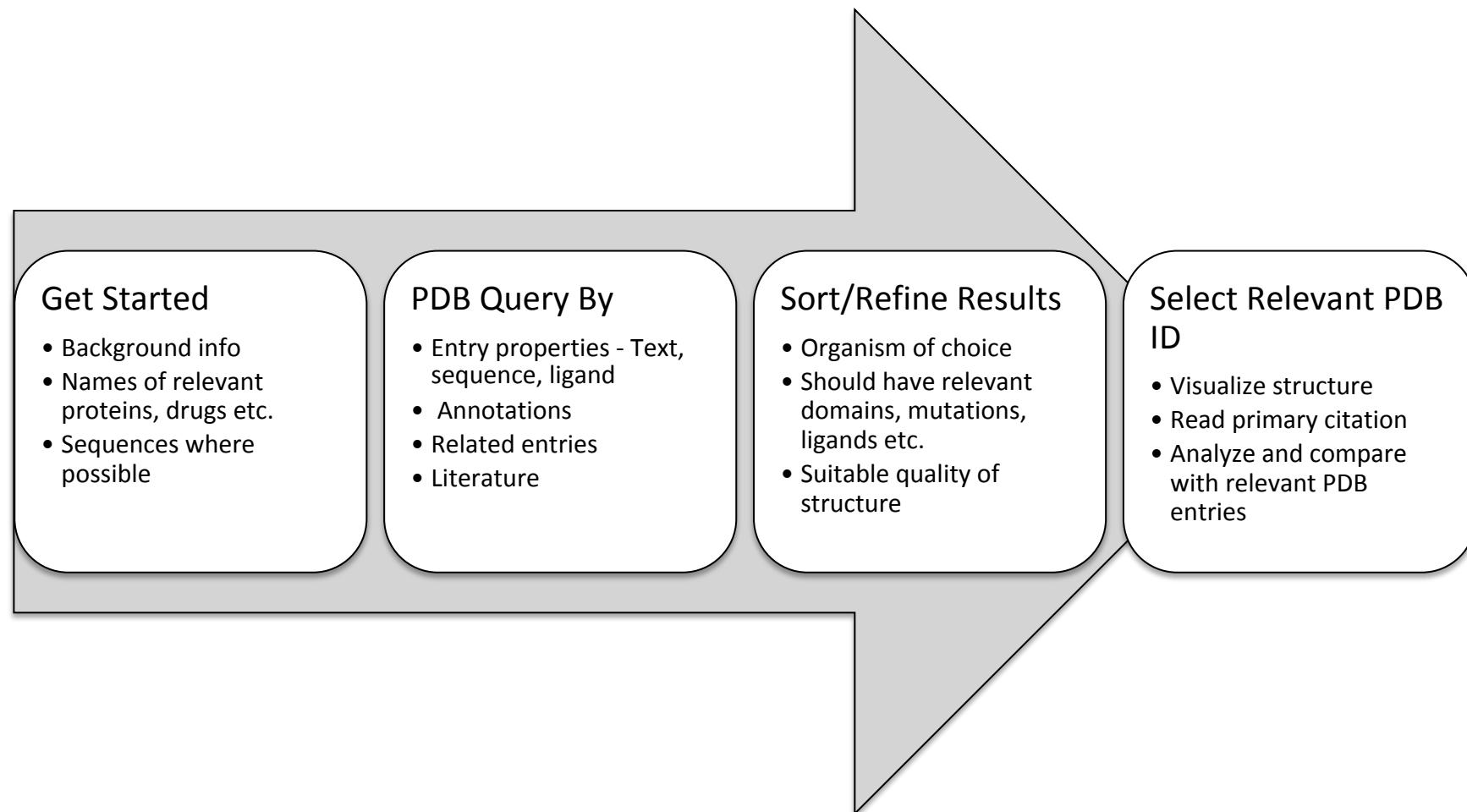




- (a) myoglobin (b) hemoglobin (c) lysozyme (d) transfer RNA
(e) antibodies (f) viruses (g) actin (h) the nucleosome
(i) myosin (j) ribosome

Courtesy of David Goodsell, TSRI

Finding PDB Structures Related to a Topic of Interest



Protein structure databases

- [PDB](#)
 - 3D structures
- [SCOP](#)
 - Murzin, Brenner, Hubbard, Chothia
 - Classification
 - Class (mostly alpha, mostly beta, alpha/beta (interspersed), alpha+beta (segregated), multi-domain, membrane)
 - Fold (similar structure)
 - Superfamily (homology, distant sequence similarity)
 - Family (homology and close sequence similarity)

The SCOP Database

Structural Classification Of Proteins

FAMILY: proteins that are >30% similar, or >15% similar and have similar known structure/function

SUPERFAMILY: proteins whose families have some sequence and function/structure similarity suggesting a common evolutionary origin

COMMON FOLD: superfamilies that have same secondary structures in same arrangement

CLASS: alpha, beta, alpha–beta, alpha+beta, multidomain

Protein databases

- [CATH](#)
 - Orengo et al
 - Class (alpha, beta, alpha/beta, few SSEs)
 - Architecture (orientation of SSEs but ignoring connectivity)
 - Topology (orientation and connectivity, based on SSAP = fold of SCOP)
 - Homology (sequence similarity = superfamily of SCOP)
 - S level (high sequence similarity = family of SCOP)
 - SSAP alignment tool (dynamic programming)

Some databases/tools

- **PubMed**—PubMed contains more than 23 million abstracts for biomedical literature from MEDLINE, life science journals, and online books.
 - <https://www.ncbi.nlm.nih.gov/pubmed/>
- **dbSNP**—The Single Nucleotide Polymorphism Database (dbSNP) is an archive for genetic variation within and across different species developed and hosted by National Center for Biotechnology Information (NCBI) in collaboration with National Human Genome Research Institute (NHGRI). The database contains information about SNPs, short deletion and insertional polymorphisms (indels/DIPs), microsatellite markers and short tandem repeats (STRs), multi nucleotide Polymorphisms (MNPs), heterozygous Sequences, and named variants.
 - <https://www.ncbi.nlm.nih.gov/projects/SNP/>
- **NCBI Variation Viewer** - Variation Viewer allows you to view, search, and navigate variations in genomic context. You can review data from dbSNP, dbVar and ClinVar, or upload your own data.
 - <https://www.ncbi.nlm.nih.gov/variation/view/overview/>

Proteins must

- Fold correctly
- Be stable (but not too stable)
- Bind to other proteins or ligands in the cell
- Have a functional catalytic site
- Assemble correctly (oligomerization)

How does variation affect proteins?

Activity—The variation causes increase, decrease, or complete loss of protein activity

Src kinase – mutant is constitutively active

Aggregation—Variation renders the protein aggregation prone

Cystatin, Cu, Zn superoxide dismutase

Stability—Variation may make the protein susceptible to proteolytic cleavage, or cause a change in thermal inactivation temperature, or cause a change in the energy of stabilization of the protein. It can also lead to destabilization of a protein oligomer, loss of packing or hydrophobic interactions, or change a mode(s) of protein-protein interaction.

DJ-1

Binding/Dissociation—changes in affinity for a known binding partner, or alterations in association or dissociation kinetics. It can also cause structural changes in the binding site or affect specificity for a binding partner(s).

Ras

Assembly—The SNV affects the oligomeric assembly properties of the protein.

arylsulfatase A, DJ-1

Rearrangement—Variation causes conformational changes in the neighborhood of the mutation.

Abl kinase, von Willebrand factor

Examples for each SNV related effect category.

Activity	rs137852646	Glycyl-tRNA synthetase	2PMF	2ZT5	G526R	Loss of activity	Charcot-Marie-Tooth disease	[50]
Aggregation	rs121912442	Cu, Zn superoxide dismutase [HSOD]	1N19	4FF9	A4V	Destabilization of protein and formation of aggregates.	Lou Gehrig's disease	[51]
Stability	rs74315351	DJ-1	2RK4	1P5F	M26I	Leads to decrease thermal stability and inactivation.	Rare forms of familial Parkinsonism	[52,54]
Binding	rs104894227	HRAS	2QUZ	2CE2	K117R	Increases the rate of nucleotide dissociation and results in constitutive activation of HRAS.	Costello Syndrome	[55]
Assembly	rs1141718	Manganese superoxide dismutase	1VAR	1MSD	I58T	The packing defects due to the mutation disrupt the dimer-tetramer equilibrium and favor the dimer over tetramer in solution.	Amyotrophic Lateral Sclerosis	[56]
Rearrangement	rs61749389	von Willebrand factor	1IJK	1OAK	I546V	The mutation causes a "Gain of Function" effect and produces a phenotype in which regulation is lost	von Willebrand disease	[57]

<https://doi.org/10.1371/journal.pone.0171355.t003>

Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355. <https://doi.org/10.1371/journal.pone.0171355>

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

Activity

Download the following papers

Bhattacharya R, Rose PW, Burley SK, Prlić A (2017) Impact of genetic variation on three dimensional structure and function of proteins. PLOS ONE 12(3): e0171355.
<https://doi.org/10.1371/journal.pone.0171355>

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171355>

Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nature Medicine*, 23(6), 703–713.

<http://doi.org/10.1038/nm.4333>

Please read through the papers and answer the following questions about the papers.

1. What is the biological question?
2. What is the method?
3. What significant scientific contribution does the paper make?

We will go over these papers in more detail soon.

Activity

Go through the papers and identify one protein whose mutation (remember we focus on missense mutations) is associated with a disease state.

Go to the PDB and search for the protein using its name.

Give a brief summary of the available information in the PDB:

How many entries are there for this protein?

Which method(s) were used in characterization of the structure?

For the most recent entry with a publication record:

What is the structure determination method?

What is the Uniprot ID?

How many chains does it have?

What is the sequence length?

Which ligands (if any) are present?

Examine the 3D structure using 3D view. Color by chain.

Play with all of the options and prepare a structure figure.