Protein structure and function

Proteins

- ***Extremely versatile in their function***

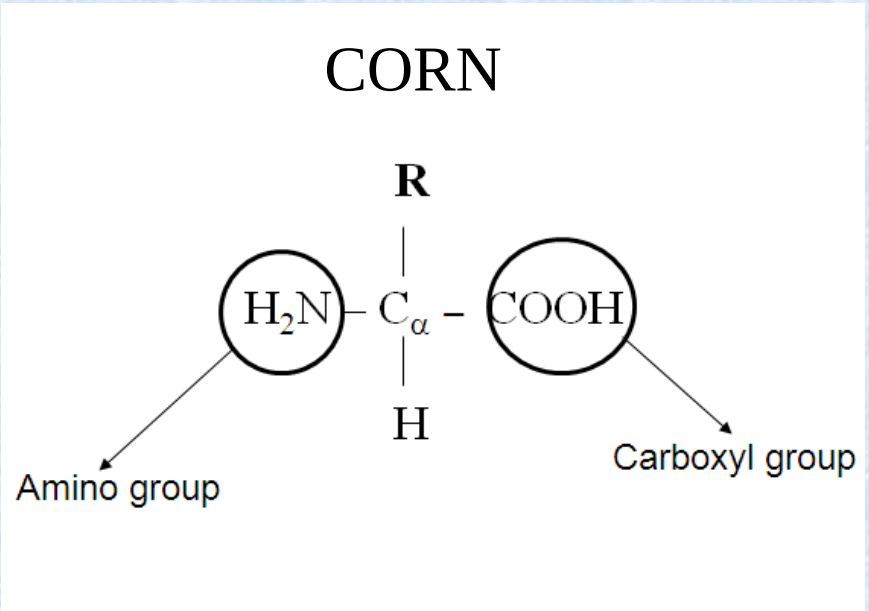
Nucleic acids

- **Carry genetic information**

Carbohydrates

- **Staple of human diet**
- **Structural and protective elements**

Building blocks



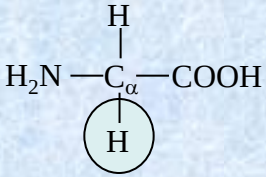
3 and 1 letter codes

Alanine	Ala	A	Methionine	Met	M
Cysteine	Cys	C	Asparagine	Asn	N
Aspartic acid	Asp	D	Proline	Pro	P
Glutamic acid	Glu	E	Glutamine	Gln	Q
Phenylalanine	Phe	F	Arginine	Arg	R
Glycine	Gly	G	Serine	Ser	S
Histidine	His	H	Threonine	Thr	T
Isoleucine	Ile	I	Valine	Val	V
Lysine	Lys	K	Tryptophan	Trp	W
Leucine	Leu	L	Tyrosine	Tyr	Y

B, J, O, U, X, Z

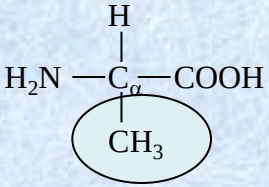
Hydrophobic residues

Hydrogen

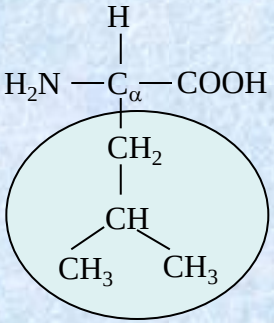


Glycine
Gly, G

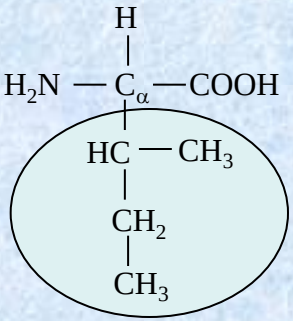
Aliphatic



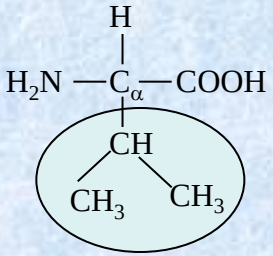
Alanine
Ala, A



Leucine
Leu, L

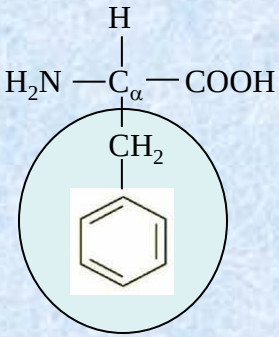


Isoleucine
Ile, I

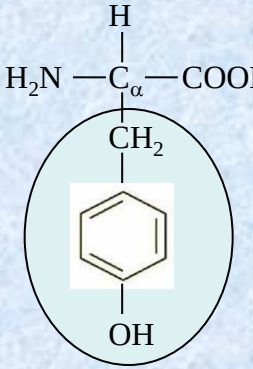


Valine
Val, V

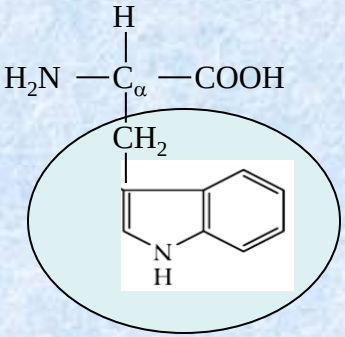
Aromatic



Phenylalanine
Phe, F

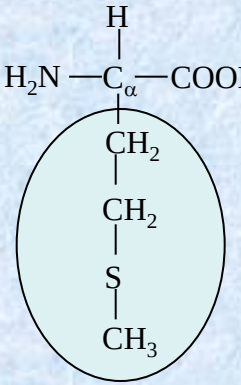


Tyrosine
Tyr, Y

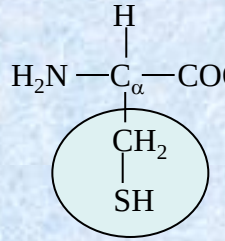


Tryptophan
Trp, W

Sulfur containing



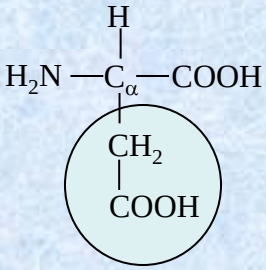
Methionine
Met, M



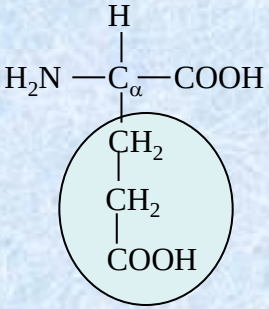
Cysteine
Cys, C

Hydrophilic residues

Negative charged

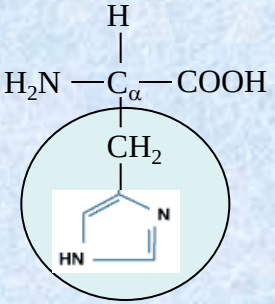


Aspartic acid
Asp, D

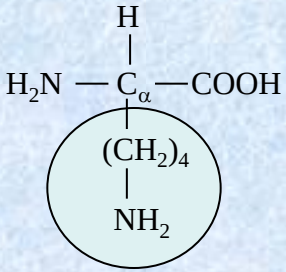


Glutamic acid
Glu, E

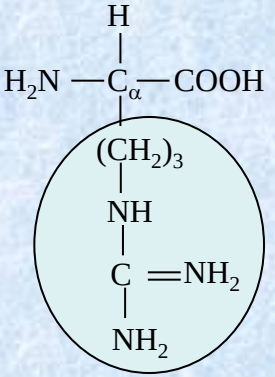
Positive charged



Histidine
His, H

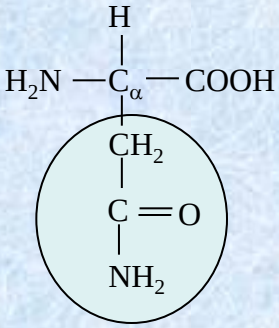


Lysine
Lys, K

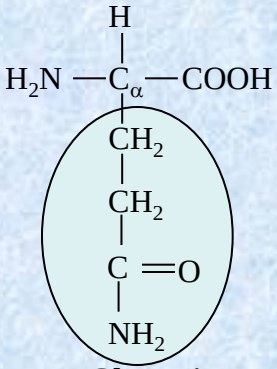


Arginine
Arg, R

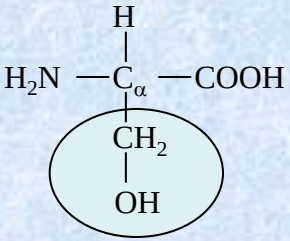
Polar



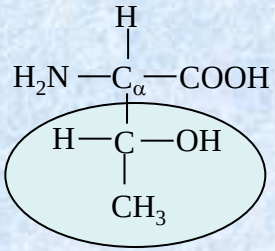
Asparagine
Asn, N



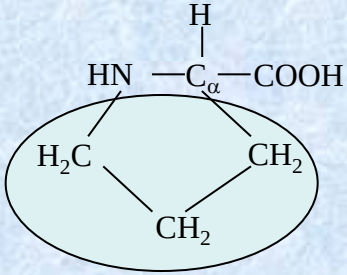
Glutamine
Gln, Q



Serine
Ser, S

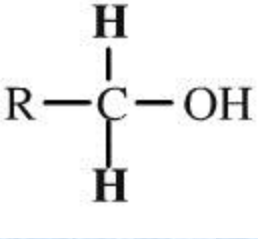


Threonine
Thr, T

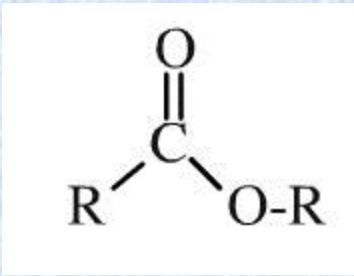


Proline
Pro, P

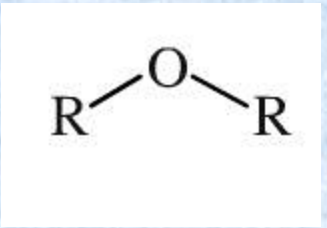
Chemical groups



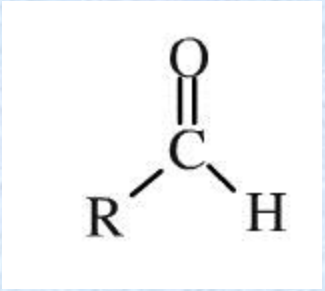
Alcohol



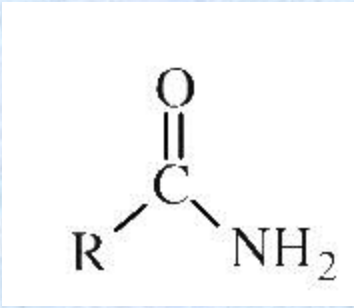
Ester



Ether

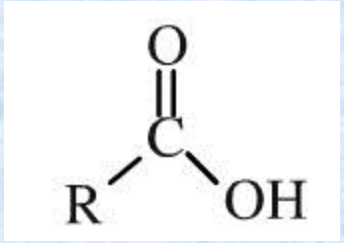


Aldehyde

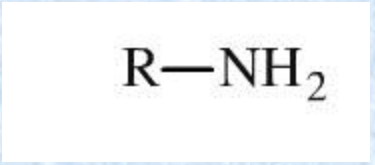


Amide

Alkane



Acid

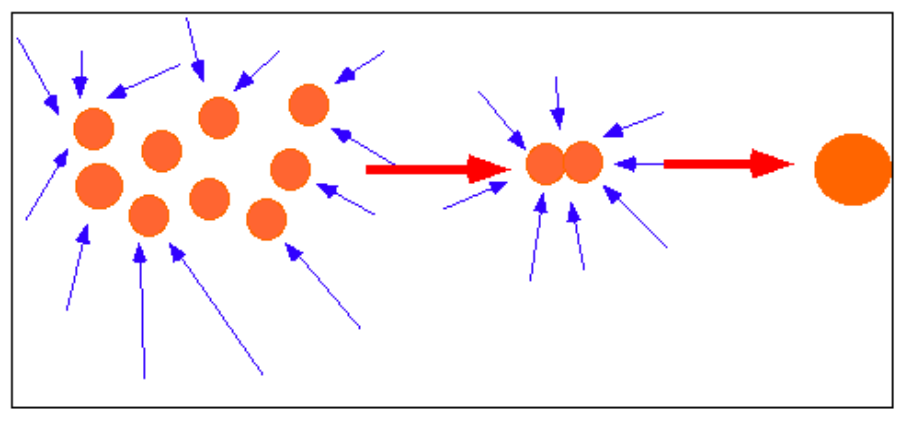


Amine

Aliphatic
Aromatic

Amino acids: properties

1. Nonpolar (Hydrophobic; water hating; hydrocarbon alkyl groups (alkane branches) or aromatic (benzene rings)



2. Polar (hydrophilic; water loving): functional groups such as acids, amides, alcohols, and amines

Amide > Acid > Alcohol > Amine > Ether > Alkane

Neutral (glutamine, asparagine)
Acidic (glutamic acid, aspartic acid)
Basic (lysine, arginine)

Polarity

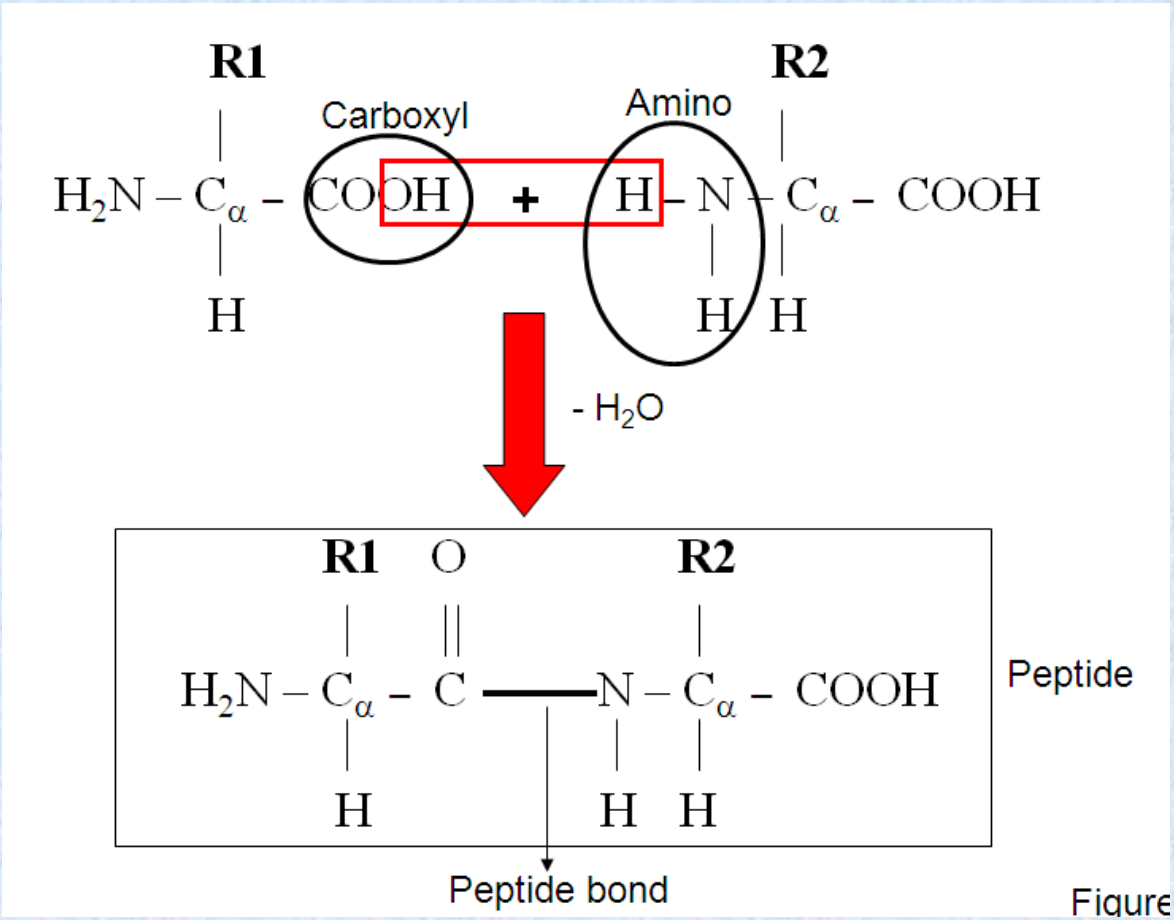
1. Rank the following according to increasing non-polarity i.e. 1 = least non-polar, 4 = most non-polar; leu; phe; val; ala (A, V, L, F)

2. Rank the following amino acids by increasing polarity. i.e.
   1 = more non-polar.
   ser ; glu ; asp ; lys ; ala ; gln (A, K, S, E, D, Q)

3. Which amino acid is most insoluble in water: isoleucine or alanine ? (I)

4. Which amino acid is most soluble in water: lys or ser? (S)

Peptide bond

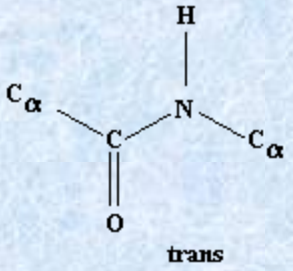


Figure

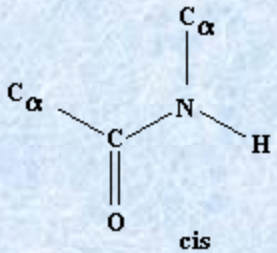
Combination of amino acids

Elimination of water

Peptide bond formation; strong



trans



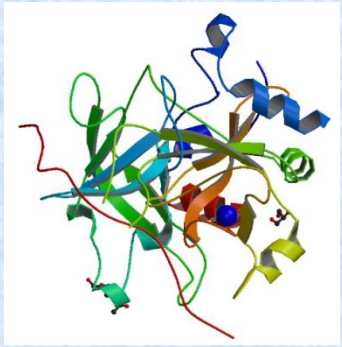
cis

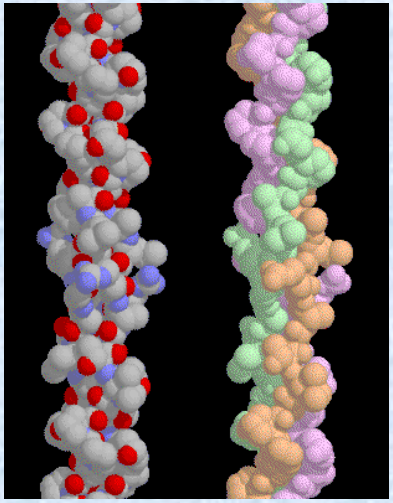
Proteins

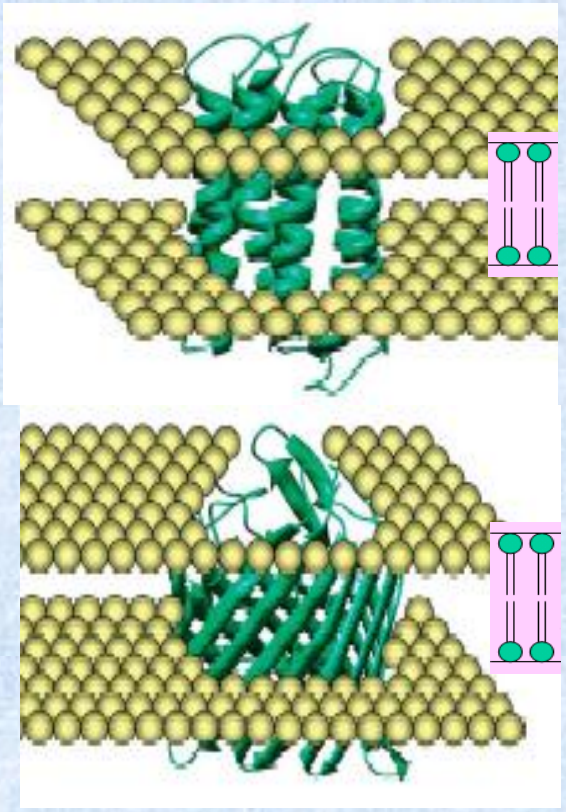
Globular proteins

Fibrous proteins

Membrane proteins







Cytoplasm

Inner membrane

α-helical

Periplasm

Outer membrane

β-barrel (TMβ)

Outer space

Proteins: Functions

Enzymes

Specialized proteins with catalytic activity (increase the rate of chemical reactions).

Only a small portion of the enzyme (3–4 amino acids) is directly involved in catalysis. The region that contains these catalytic residues, binds the substrate, and then carries out the reaction is known as the active site.

All the chemical reactions in cell are catalyzed by enzymes.

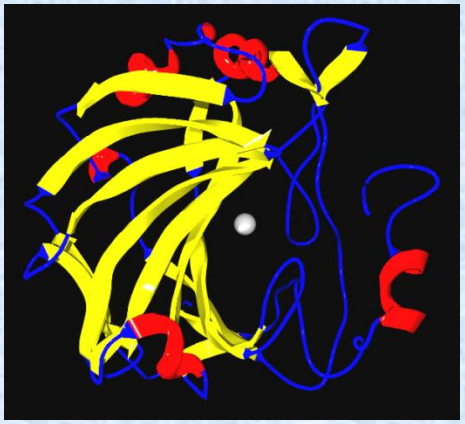
The **carbonic anhydrases** form a family of enzymes that catalyze the rapid interconversion of carbon dioxide and water to bicarbonate.

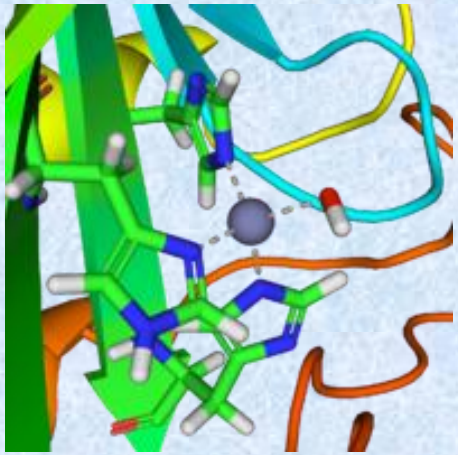
$$CO_2 + H_2O \xrightarrow{\text{Carbonic anhydrase}} H_2CO_3$$

(in tissues; high CO_2 concentration)

$$H_2CO_3 \xrightarrow{\text{Carbonic anhydrase}} CO_2 + H_2O$$

(in lungs; low CO_2 concentration)





three histidine residues and a hydroxide group coordinating (dashed lines) the zinc on at center.

The active site of most carbonic anhydrases contains a zinc ion; they are therefore classified as metalloenzymes.

Proteins: Functions

In enzymatic reactions, the **molecules** at the beginning of the process are called **substrates**, and they are converted into different molecules, called the **products**

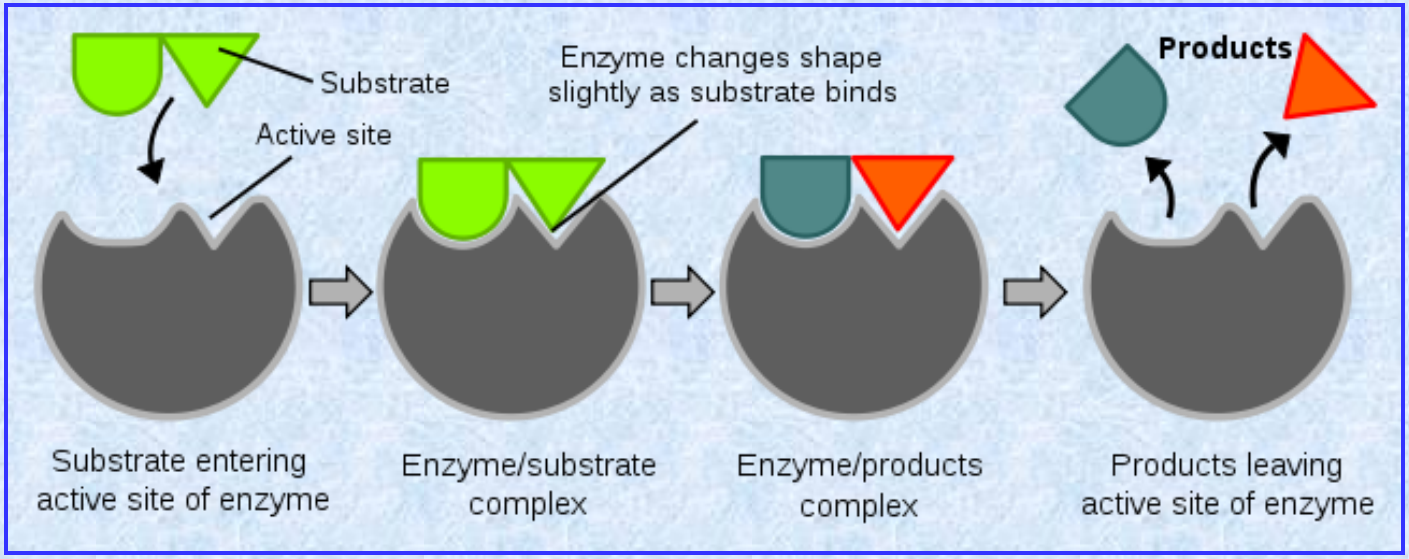
$$CO_2 + H_2O \xrightarrow{\text{Carbonic anhydrase}} H_2CO_3$$

(in tissues; high CO_2 concentration)

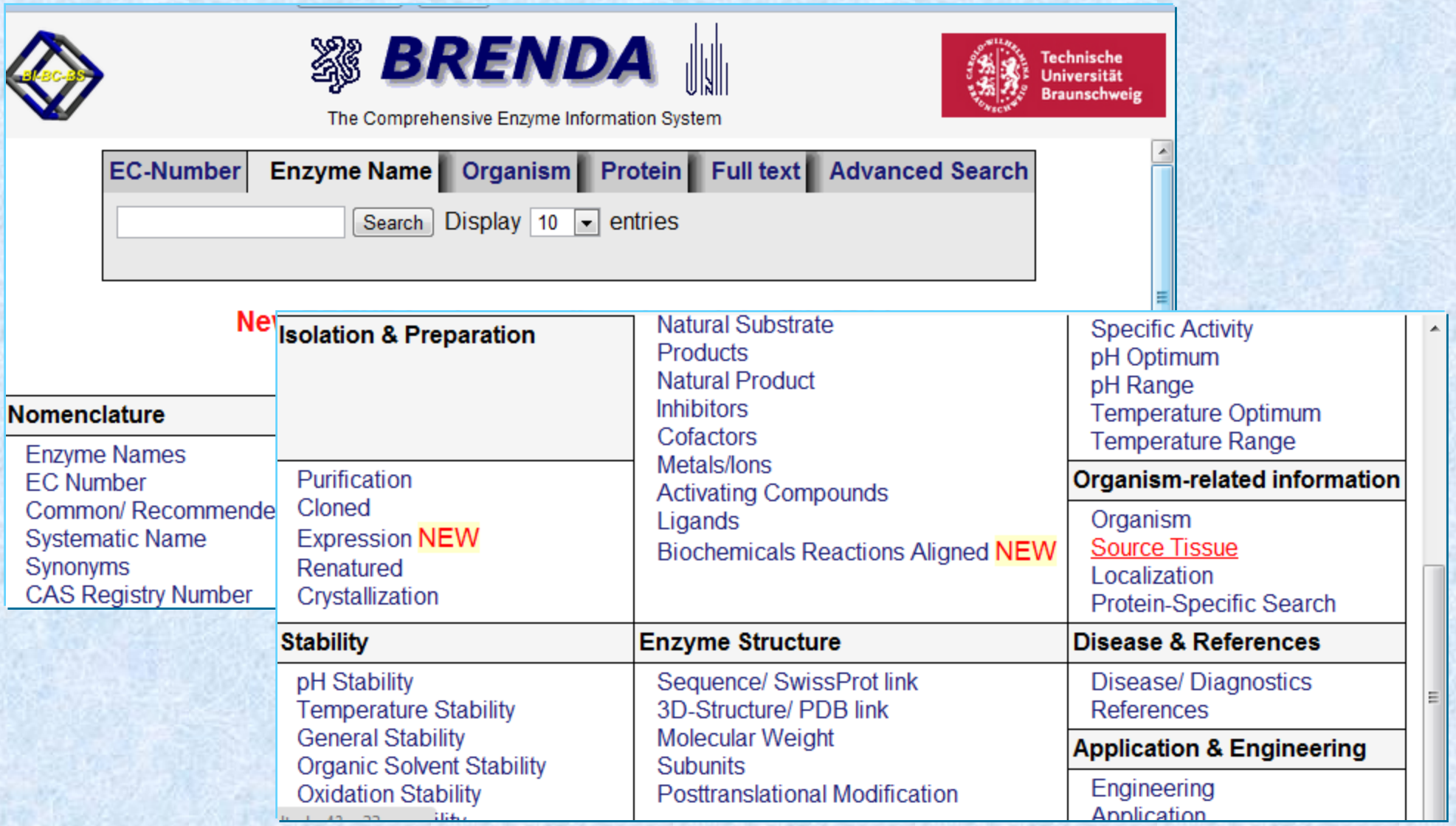
$$H_2CO_3 \xrightarrow{\text{Carbonic anhydrase}} CO_2 + H_2O$$

(in lungs; low CO_2 concentration)

Mechanism



Alcohol dehydrogenase: facilitate to break down alcohol (toxic) to others (aldehade/ketone)
Catalyzing oxidation



http://www.brenda-enzymes.org/

EC number: format

Every enzyme code consists of the letters "EC" followed by four numbers separated by periods. Those numbers represent a progressively finer classification of the enzyme.

- **For example, the tripeptide aminopeptidases have the code "EC 3.4.11.4", whose components indicate the following groups of enzymes:**

EC 3 **enzymes are hydrolases (enzymes that use water to break up some other molecule)**

EC 3.4 **are hydrolases that act on peptide bonds**

EC 3.4.11 **are those hydrolases that cleave off the amino-terminal amino acid from a polypeptide**

EC 3.4.11.4 **are those that cleave off the amino-terminal end from a tripeptide**

Top level codes

Top-level EC numbers[4]

Group	Reaction catalyzed	Typical reaction	Enzyme example(s) with trivial name
EC 1 *Oxidoreductases*	To catalyze oxidation/reduction reactions; transfer of H and O atoms or electrons from one substance to another	AH + B → A + BH (**reduced**) A + O → AO (**oxidized**)	Dehydrogenase, oxidase
EC 2 *Transferases*	Transfer of a functional group from one substance to another. The group may be methyl-, acyl-, amino- or phosphate group	AB + C → A + BC	Transaminase, kinase
EC 3 *Hydrolases*	Formation of two products from a substrate by hydrolysis	AB + H_2O → AOH + BH	Lipase, amylase, peptidase
EC 4 *Lyases*	Non-hydrolytic addition or removal of groups from substrates. C-C, C-N, C-O or C-S bonds may be cleaved	RCOCOOH → RCOH + CO_2 or [X-A-B-Y] → [A=B + X-Y]	Decarboxylase
EC 5 *Isomerases*	Intramolecule rearrangement, i.e. isomerization changes within a single molecule	AB → BA	Isomerase, mutase
EC 6 *Ligases*	Join together two molecules by synthesis of new C-O, C-S, C-N or C-C bonds with simultaneous breakdown of ATP	X + Y+ ATP → XY + ADP + Pi	Synthetase

BRENDA

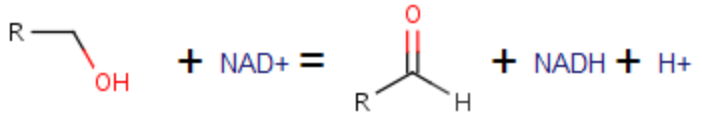
The Comprehensive Enzyme Information System

EC 1.1.1.1 - alcohol dehydrogenase

EC NUMBER	COMMENTARY
1.1.1.1	-

RECOMMENDED NAME	GeneOntology No.
alcohol dehydrogenase	GO:0004025

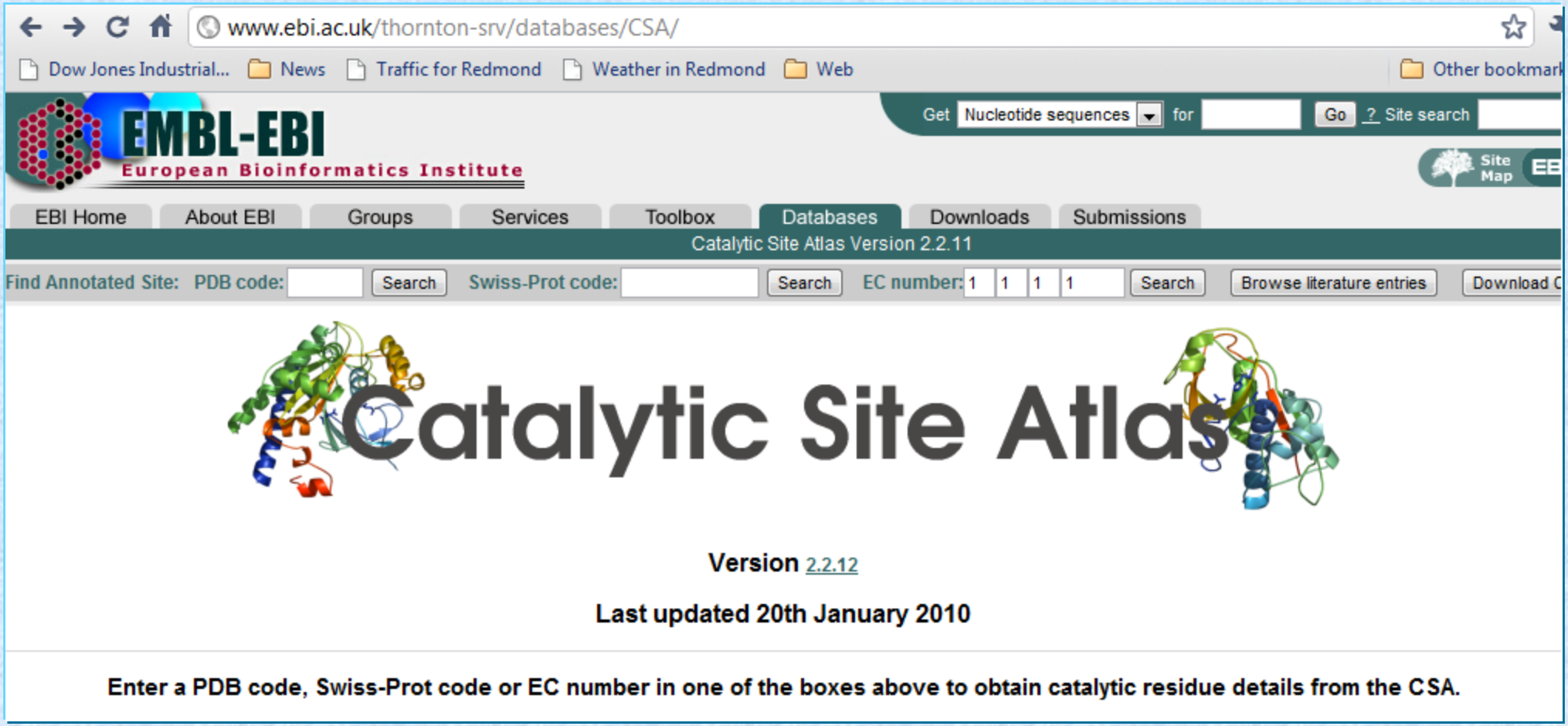
Reaction catalyzed by alcohol dehydrogenase (1.1.1.1)



A primary alcohol + NAD+ = an aldehyde + NADH + H+

REACTION	REACTION DIAGRAM	COMMENTARY	ORGANISM	LITERATURE
A primary alcohol + NAD+ = an aldehyde + NADH + H+		ordered bi bi mechanism with cofactor adding first to form a binary enzyme complex	Homo sapiens	285578
A primary alcohol + NAD+ = an aldehyde + NADH + H+		Ser48 is involved in catalysis, isozyme gamma(2)gamma(2)	Homo sapiens	654727

Catalytic site atlas



http://www.ebi.ac.uk/thornton-srv/databases/CSA/

EC: 1.1.1.1

CSA Entries:

1a4u, 1a71, 1a72, 1adb, 1adc,
1adf, 1adg, 1adh, 1agn, 1axe,
1axg, 1b14, 1b15, 1b16, 1b21,
1bto, 1cdo, 1d1s, 1d1t, 1dda,
1deh, 1e3e, 1e3i, 1e3l, 1ee2,
1h2b, 1hdx, 1hdy, 1hdz, 1het,
1heu, 1hf3, 1hld, 1hso, 1hsz,
1ht0, 1htb, 1j5r, 1ju9, 1jvb,
1lde, 1ldy, 1llu, 1m6h, 1m6w,
1ma0, 1mc5, 1mg0, 1mg5, 1mgo,
1mp0, 1n8k, 1n92, 1nto, 1nvg,
1o2d, 1otq, 1p1r, 1qlh, 1qlj,
1qv6, 1qv7, 1r37, 1rjw, 1sby,
1teh, 1u3t, 1u3u, 1u3v, 1u3w,
1vj0, 1ye3, 2adh, 2eer, 2fze,
2fzw, 2hcy, 2jhf, 2jhg, 2ohx,
2oxi, 3adh, 3bto, 3hud, 3i4c,
5adh, 6adh, 7adh, 8adh

CSA entry for 8adh

Homologous Entry

Title:	Oxidoreductase(nad(a)-choh(d))		
Compound:	Apo-liver alcohol dehydrogenase (e.c.1.1.99.8)		
Mutant:	No		
UniProt/Swiss-Prot:	P00327-ADHE_HORSE	**EC Class:**	1.1.1.1
Other CSA Entries:	Homologues of 8adh Entries for UniProt/Swiss-Prot: P00327 Entries for EC: 1.1.1.1	**Other Databases:**	PDB entry: 8adh PDBsum entry: 8adh UniProt/Swiss-Prot: P00327 IntEnz entry: 1.1.1.1 KEGG entry: 1.1.1.1 EzCatDB entry: D00001

Sites:

☑ Catalytic Site (Get help with this section)

Found by: PsiBLAST alignment on 1guf
1guf has EC code 0... whereas 8adh has EC code 1.1.1.1. The difference in function suggests that the transfer of annotation from 1guf to 8adh may be incorrect.
In addition to these residues, this site requires the cofactor(s) with PDB three-letter code(s) NDP for its catalytic activity.

Residue	Chain	Number	UniProt number	Functional part
LEU	A	57	58	Sidechain

☑ Catalytic Site (Get help with this section)

Found by: PsiBLAST alignment on 1qlh
In addition to these residues, this site requires the cofactor(s) with PDB three-letter code(s) ZN, NAD for its catalytic activity.

Residue	Chain	Number	UniProt number	Functional part
SER	A	48	49	Sidechain
HIS	A	51	52	Sidechain

Use the check-boxes to select site(s) to view on the 3D structure in RasMol, and press Display Protein

Catalytic residues are indicated in red.

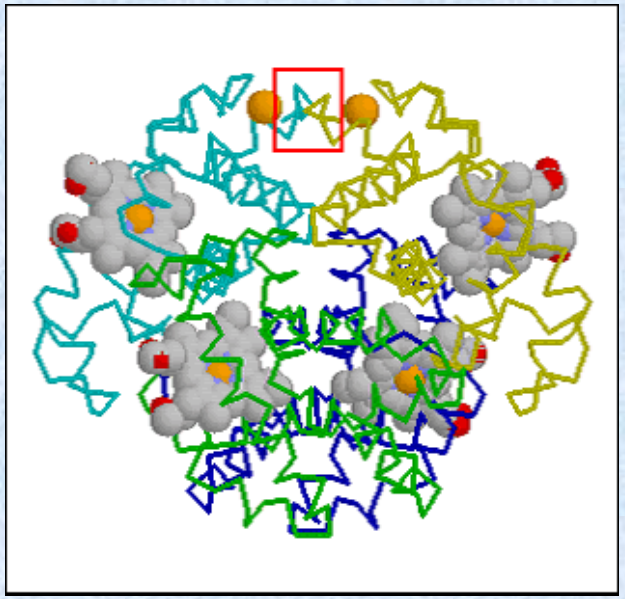
Chain: A

```
       10        20        30        40        50        60        70        80        90       100
        |         |         |         |         |         |         |         |         |         |
STAGKVIKCKAAVLWEEKKPFSIEEVEVAPPKAHEVRIKMVATGICRSDDHVVSGTLVTPLPVIAGHEAAGIVESIGEGVTTVRPGDKVIPLFTPQCGKC

       110       120       130       140       150       160       170       180       190       200
        |         |         |         |         |         |         |         |         |         |
RVCKHPEGNFCLKNDLSMPRGTMQDGTSRFTCRGKPIHHFLGTSTFSQYTVVDEISVAKIDAASPLEKVCLIGCGFSTGYGSAVKVAKVTQGSTCAVFGL

       210       220       230       240       250       260       270       280       290       300
        |         |         |         |         |         |         |         |         |         |
GGVGLSVIMGCKAAGAARIIGVDINKDKFAKAKEVGATECVNPQDYKKPIQEVLTEMSNGGVDFSFEVIGRLDTMVTALSCCQEAYGVSVIVGVPPDSQN

       310       320       330       340       350       360       370
        |         |         |         |         |         |         |
LSMNPMLLLSGRTWKGAIFGGFKSKDSVPKLVADFMAKKFALDPLITHVLPFEKINEGFDLLRSGESIRTILTF
```

Proteins: Functions

Transport proteins

- Binds and carry specific molecules or ions from one organ to another
- Hemoglobin is present in red blood cells, which efficiently carries oxygen from the lungs to the tissues of the body. Hemoglobin also helps in the transportation of carbon dioxide and hydrogen ions back to the lungs.



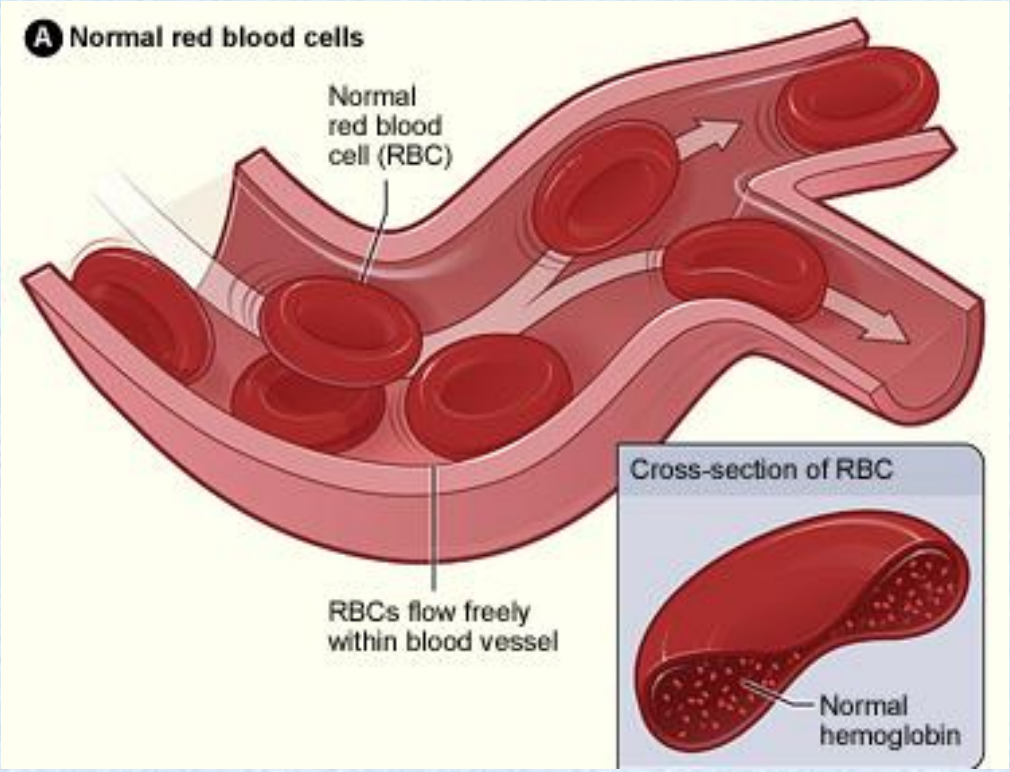
HBB Sequence in Normal Adult Hemoglobin (Hb A):

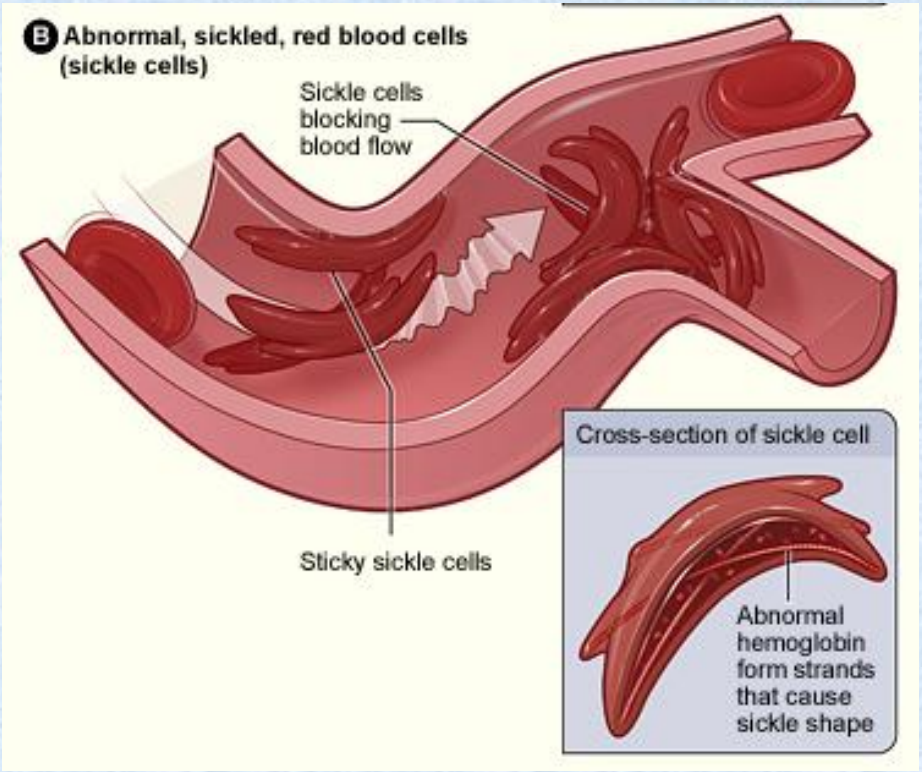
Nucleotide	CTG	ACT	CCT	GAG	GAG	AAG	TCT
Amino Acid	Leu	Thr	Pro	Glu	Glu	Lys	Ser
	3			6			9

HBB Sequence in Mutant Adult Hemoglobin (Hb S):

Nucleotide	CTG	ACT	CCT	GTG	GAG	AAG	TCT
Amino Acid	Leu	Thr	Pro	Val	Glu	Lys	Ser
	3			6			9

Proteins: Functions

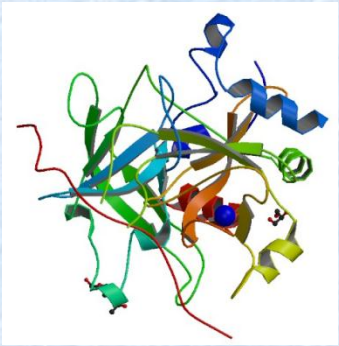




Proteins: Functions

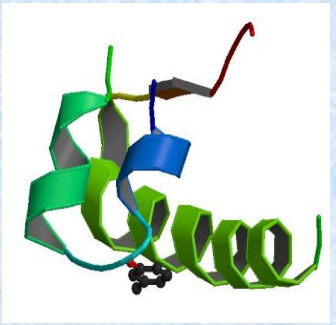
Defense proteins

- **Antibodies** (immunoglobulins) can recognize and precipitate or neutralize invading bacteria or viruses from other species;
- Thrombin is a **blood clotting protein** that prevent loss of blood when a vascular system is injured.



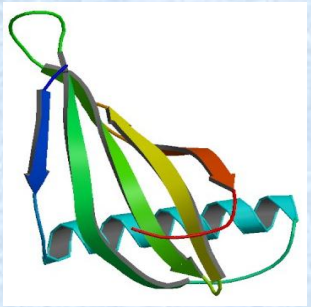
Regulatory proteins

- Regulate cellular or physiological activity.
- Eg. insulin, which regulates **sugar metabolism**.



Others

- Monellin has a intensely sweet taste, which is used to be a **sweetener**.



Proteins: Functions

Nutrient and strorage proteins

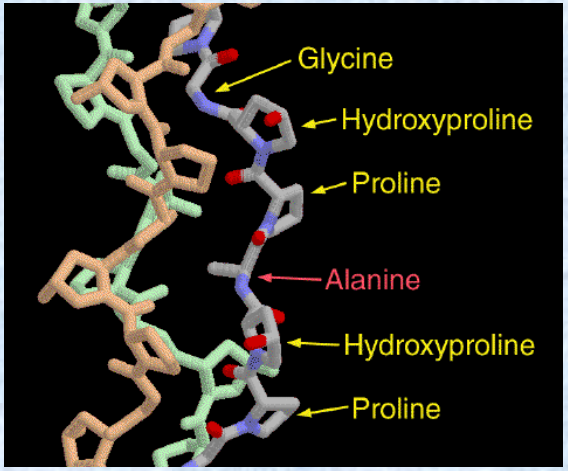
- Ovalbumin, major protein of egg white, caesin in milk

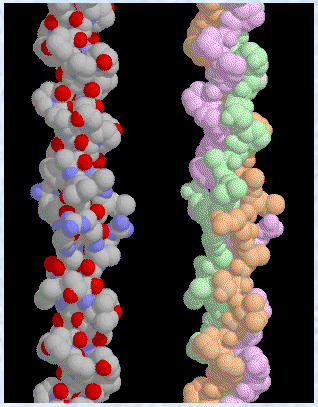
Contractile or Motile proteins

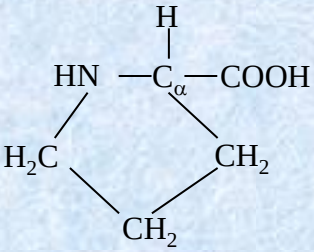
- Actin and myosin function in the contractile system of skeletal muscle and many non-muscle cells.

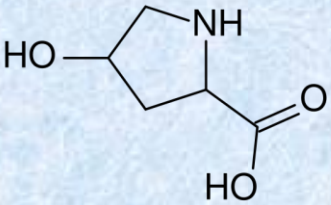
Structural proteins

- Give biological **strength/ protection**. Major component of tendons and cartilage is the fibrous protein, collagen. Hairs, finger nails and feathers consists of keratin.





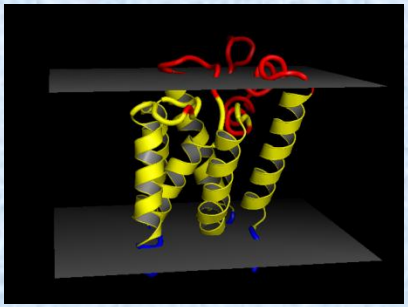


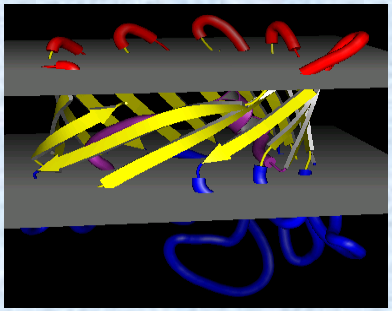


Hydroxyproline is produced by hydroxylation of the amino acid proline by the enzyme prolyl hydroxylase following protein synthesis

Membrane proteins

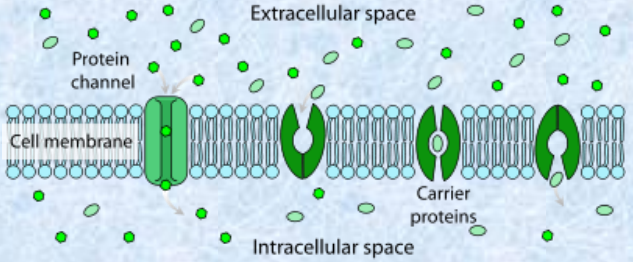
- Proteins that are embedded into membranes are called membrane proteins
- Structure: α-helical and β-barrel





- Function: Transporters, receptors, channels

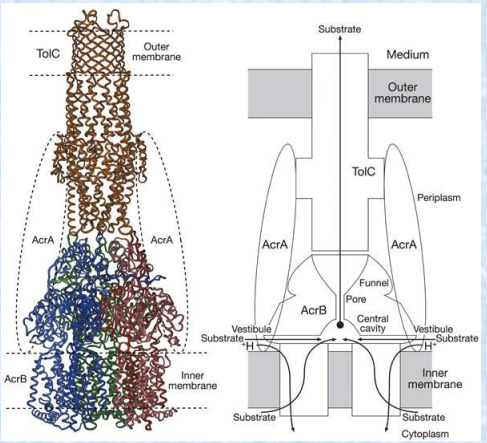
Transporters



A membrane transport protein (transporter) is a membrane protein involved in the movement of ions, small molecules, or macromolecules such as another protein across a biological membrane.

The proteins may assist in the movement of substances by facilitated diffusion or active transport.

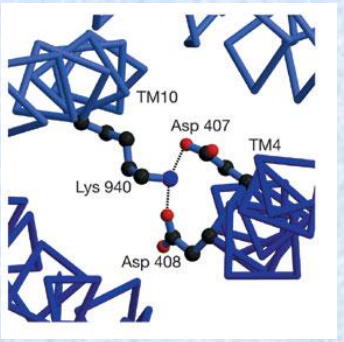
E.g. Multidrug efflux transporter AcrB



When substrate is transported AcR might recruit TolC to form direct transit pathway from the cytoplasm to extracellular.

Two pathways (through cytoplasm; central cavity; central trimer hole plays a role in substrate translocation)

Ion pairs are possible candidates for transmembrane proton translocation site



Asp are protonated, ion pairs are disturbed; helices TM4 and TM10: conformation change -> transduced to the pore region by remote conformational coupling and open the pore.

Murakami et al. (2002) Nature 419, 587

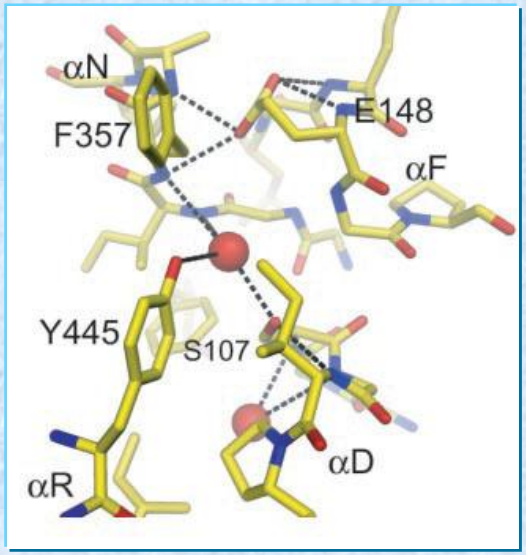
Channels

Ion channels are membrane proteins that selectively allow a given species of ion to pass through them.

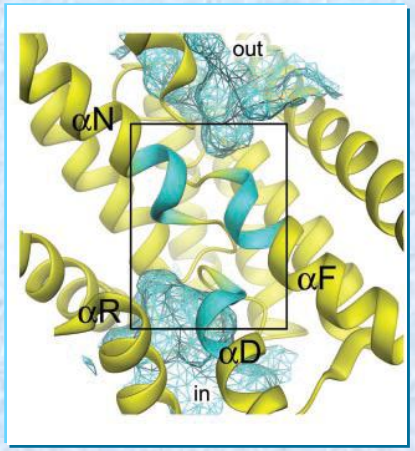
The two key properties are **selective ion conduction** and **gating**.

Selective conduction refers to a channel's ability to **select one ion species** among those present in the cellular environment.

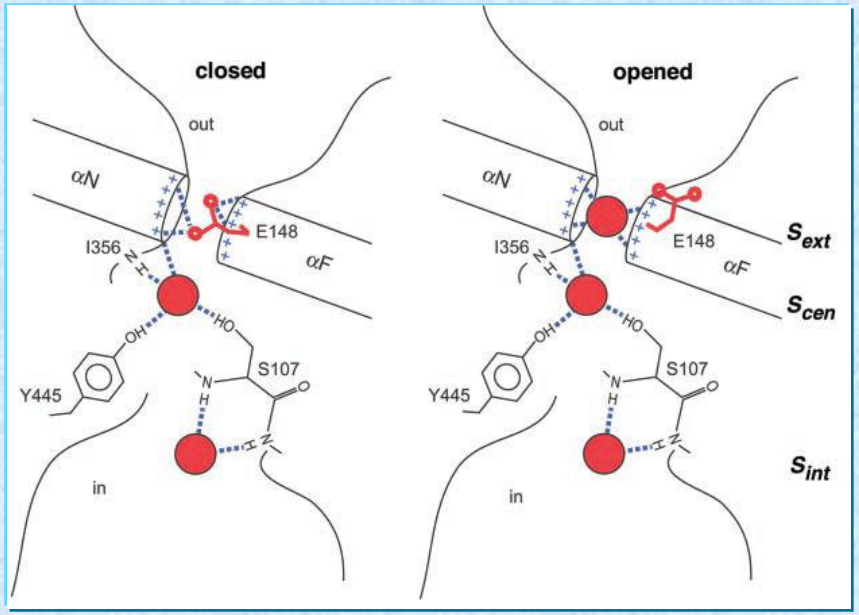
Gating refers to **opening and closing** the pore, the process by which ion conduction is turned on or off



Residues in the vicinity of chloride ions



Chloride channel
In: intracellular
Out: extracellular



Dutzler et al. (2003) Science 300, 108

Olfactory receptors

Olfaction is one of the senses involved in the perception of chemo-signals in the external environment.

The detection of odorant molecules involves specific binding to specialized receptor in olfactory system.

OR proteins belong to the G Protein-Coupled Receptor superfamily, which is characterized by the presence of hydrophobic transmembrane domains.

Each OR can recognize several chemically related molecules, and a specific odorant may bind to several ORs

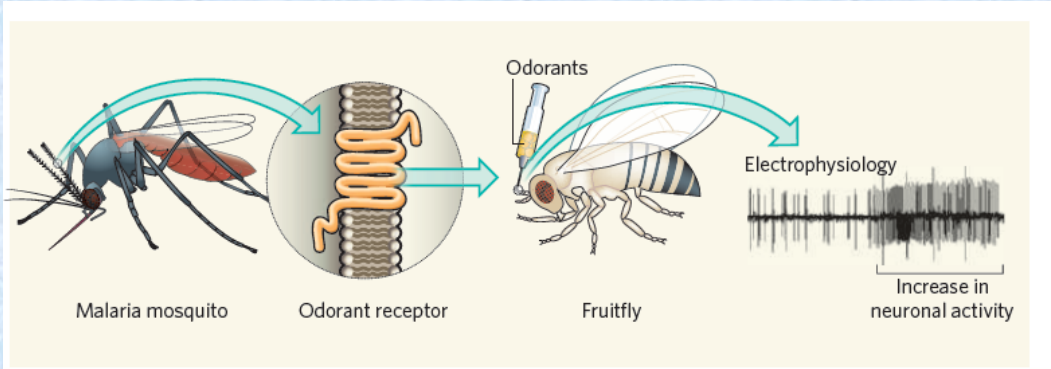


Figure 1 | What's that smell? Odorant receptors (ORs) in the antennas of the malaria mosquito *Anopheles gambiae* detect odorants produced by humans. The genes that might encode ORs have been identified from the insect's genome[4]. Carlson and colleagues[3] expressed 72 of these genes, one kind at a time, in the 'empty neurons' of the fruitfly *Drosophila melanogaster*. Of the putative ORs integrated into empty neurons, 50 turned out to be functional in the resulting mutant flies. The authors exposed these flies to a panel of 110 odorants, and measured the electrical activity of the OR-containing neurons. Odorants that bind to and activate ORs cause an increase in the number of spikes in the neuron's electrical activity. In this way, the authors identified which odorants activate (or, in some cases, inhibit) which ORs in *A. gambiae*.

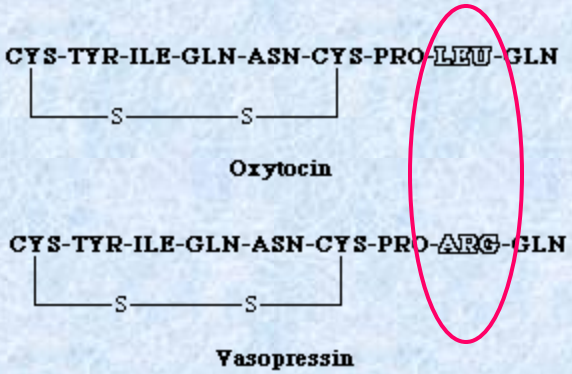
Structure-Function Relationship

Protein structure → Protein function

Function of a protein is dictated by its structure

Hormones

Simple change cause high blood pressure

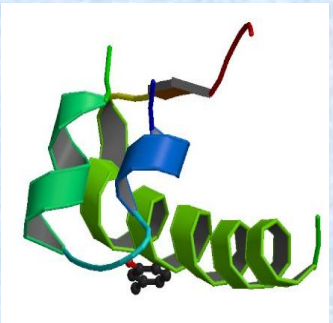


HBB Sequence in Normal Adult Hemoglobin (Hb A):

Nucleotide	CTG	ACT	CCT	GAG	GAG	AAG	TCT
Amino Acid	Leu	Thr	Pro	Glu	Glu	Lys	Ser
	3			6			9

HBB Sequence in Mutant Adult Hemoglobin (Hb S):

Nucleotide	CTG	ACT	CCT	GTG	GAG	AAG	TCT
Amino Acid	Leu	Thr	Pro	Val	Glu	Lys	Ser
	3			6			9



MUTATION			PROTEIN	GENE_NAME	DISEASE
Phe	24	Ser	INSULIN	INS	DIABETES, MODY
His	10	Asp	INSULIN	INS	HYPERPROINSULINAEMIA
Phe	24	Leu	INSULIN	INS	HYPERPROINSULINAEMIA
Arg	39	His	INSULIN	INS	HYPERPROINSULINAEMIA
Arg	39	Leu	INSULIN	INS	HYPERPROINSULINAEMIA
Arg	39	Pro	INSULIN	INS	HYPERPROINSULINAEMIA
Val	42	Leu	INSULIN	INS	HYPERPROINSULINAEMIA

http://wiz2.pharm.wayne.edu/biochem/prot.html

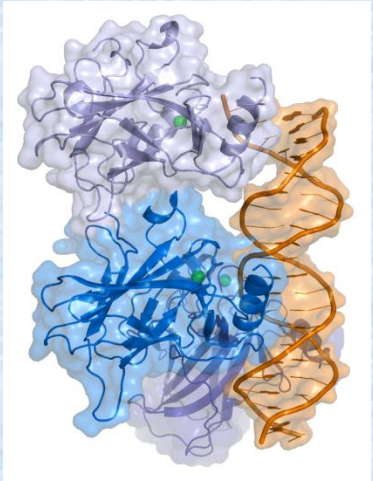
Structure-Function Relationship

Mutations in the p53 causes human cancer.

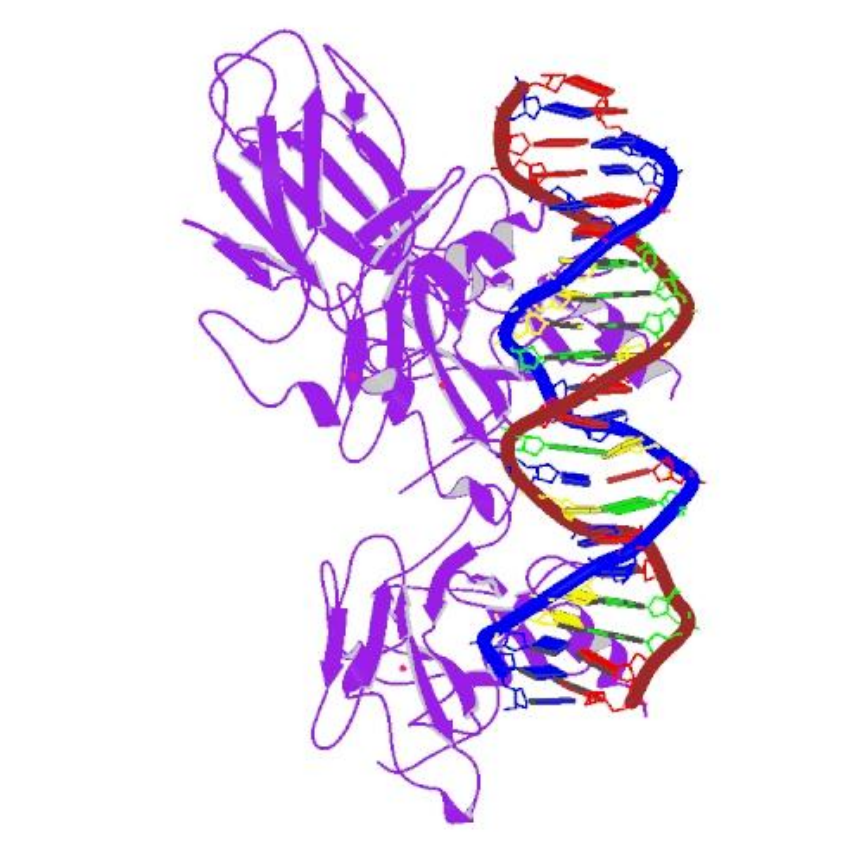
The majority of the mutations occur in the core domain (sequence-specific DNA binding activity of the p53 protein; residues 102-292)

Result in loss of DNA binding.

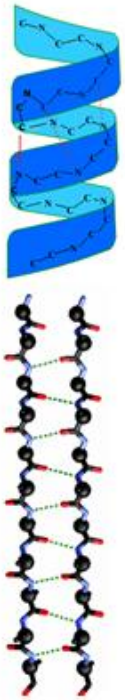
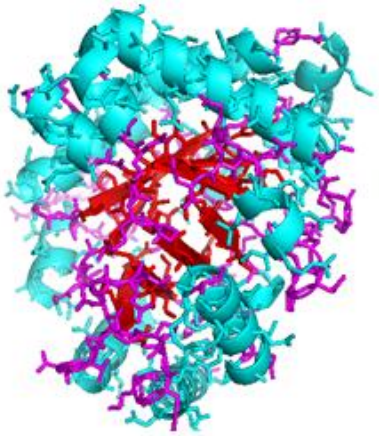
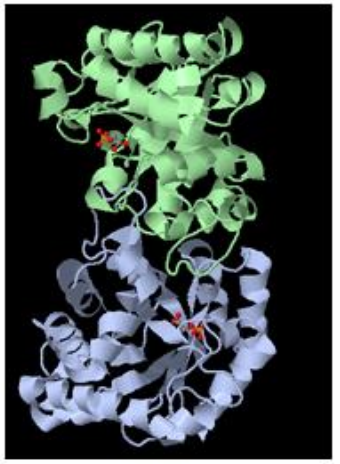
The structure supports the hypothesis that DNA binding is critical for the biological activity of p53



TUMOR SUPPRESSOR P53 COMPLEXED WITH DNA



Protein Structure

Primary structure	Secondary structure	Tertiary structure	Quaternary structure
Met Ile Lys Ala Asn Ile			
Amino acid sequence	α-helix, β-strand	3D structure with atomic details	Assembled subunits

Primary structure: human hemoglobin

>sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens

VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPD
AVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFR
LLGNVLVCVLAHHFG KEFTPPVQAAYQKVVAGVANALAHKYH

Primary structure describes the linear sequence of amino acid residues in a protein.

It includes all covalent bonds between amino acids.

The relative arrangement of the linked amino acids is not specified.

Databases for protein sequences

EXProt

Munich Information Center for Protein Sequences (MIPS)

NCBI Protein database

PIR - Protein Information Resource (Georgetown University)

PIR-NREF

PRF

SWISS-PROT (Swiss Institute of Bioinformatics)

TrEMBL

UniProt - The Universal Protein Knowledgebase