

## Leveraging Molecular Databases for Variant Interpretation

Part 2:  
Functional Genomics; understanding how pathological missense variants  
affect proteins

Instructor:

Pouria Dasmeh

Center for Human Genetics, Marburg University/UKGM medical center,  
[dasmeh@staff.uni-marburg.de](mailto:dasmeh@staff.uni-marburg.de)  
[www.dasmehlab.com](http://www.dasmehlab.com)



Course material  
(lectures & tutorials)

[https://github.com/dasmeh/UKGM\\_Humangenetik\\_PROTEINS](https://github.com/dasmeh/UKGM_Humangenetik_PROTEINS)

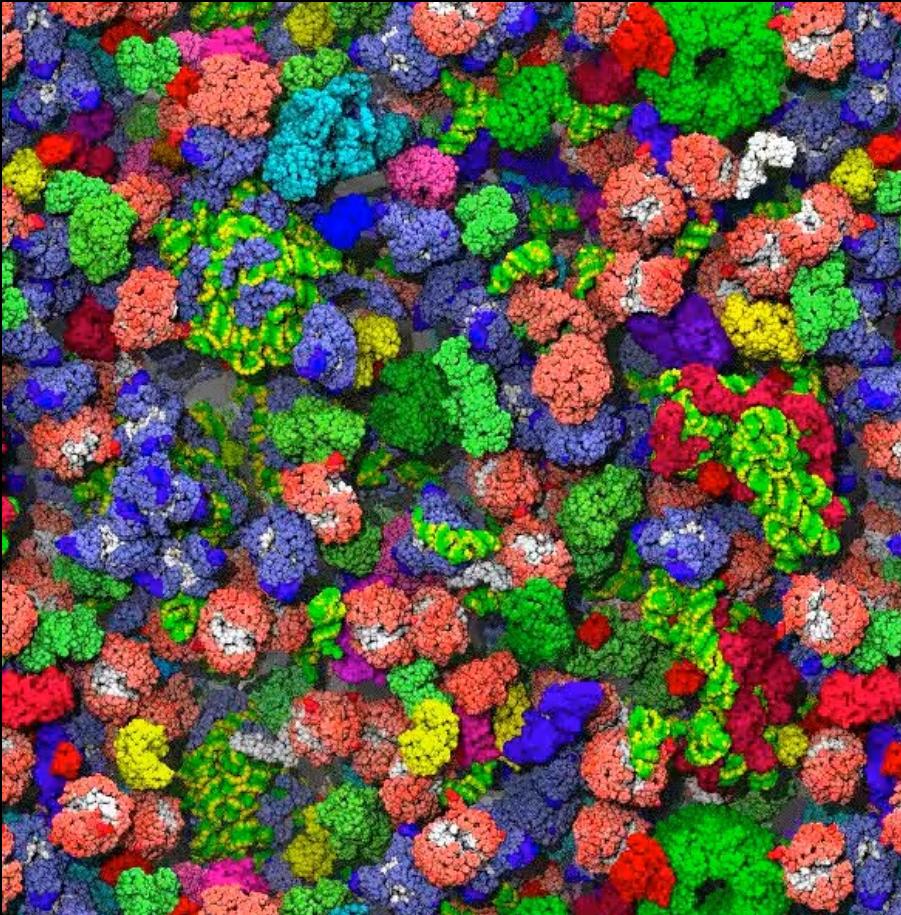
## Course agenda and organizational notes

**Theory lecture (04:30-05:00):** Protein sequence, structure, function

**Practical session 1 (05:00-05:30)**

*Short break*

**Practical session 2 (05:45-06:30)**



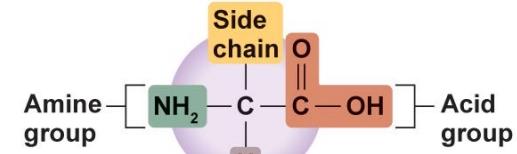
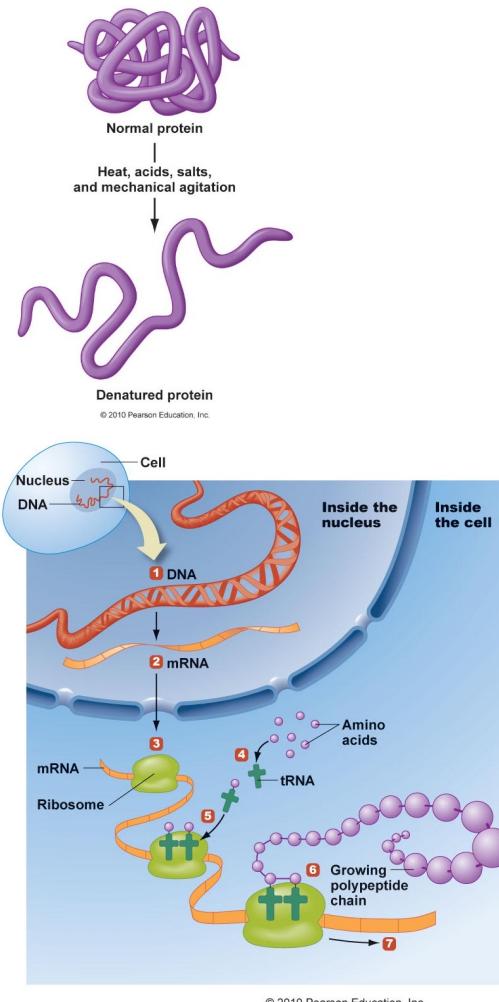
Cytoplasm's viscosity  $> 100 \text{ mPa.s}$

Water  $\sim 1 \text{ mPa.s}$   
Engine oil  $\sim 100 \text{ mPa.s}$

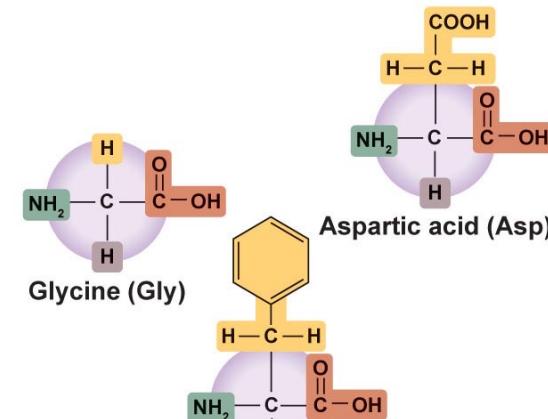
*PLoS computational biology*,  
6(3), e1000694.

# **Part 1: Protein Sequence, Structure, Function**

# Proteins and Amino Acids



**a** Amino acid structure

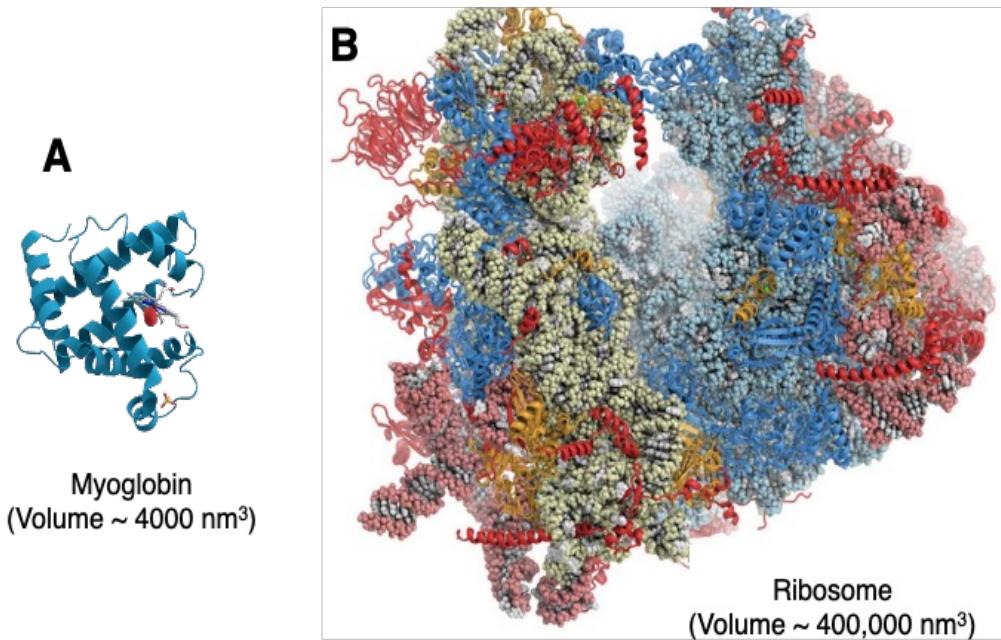


**b** Different amino acids, showing their unique side chains

© 2010 Pearson Education, Inc.

# Proteins

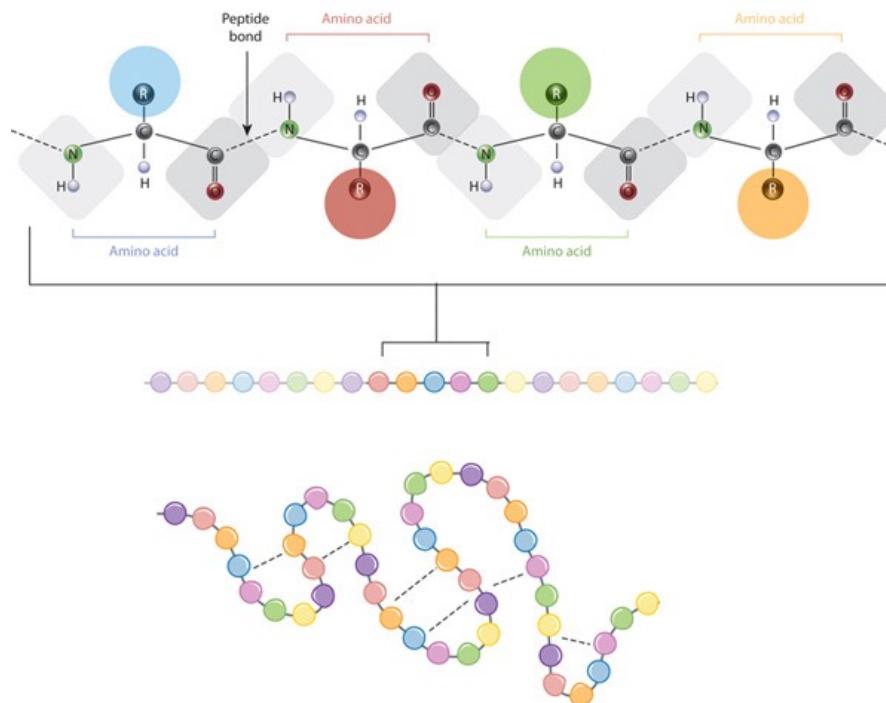
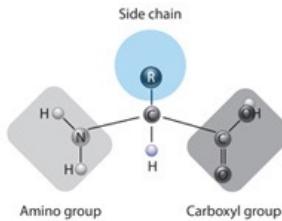
- Made up of chains of amino acids
- Are involved in most of the body's functions and life processes
- The sequence of amino acids is determined by DNA



## Structure of Proteins

- Made up of chains of amino acids; classified by number of amino acids in a chain
  - Peptides: fewer than 50 amino acids
    - Dipeptides: 2 amino acids
    - Tripeptides: 3 amino acids
    - Polypeptides: more than 10 amino acids
  - Proteins: more than 50 amino acids
    - Typically 100 to 10,000 amino acids linked together
- Chains are synthesized based on specific bodily DNA.
- Amino acids are composed of carbon, hydrogen, oxygen, and nitrogen.

# An amino acid



# Amino acids

**TWENTY-ONE  
PROTEINOGENIC  
 $\alpha$ -AMINO ACIDS**

Side chain charge  
at physiological  
pH 7.4

*pK<sub>a</sub>* values shown  
italicized

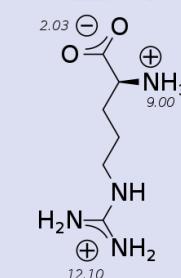
⊕ Positive  
⊖ Negative

*A. Amino Acids with Electrically Charged Side Chains*

Positive

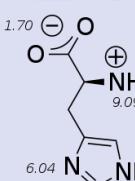
Arginine

Arg R



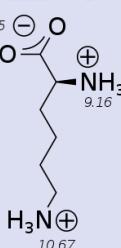
Histidine

His H



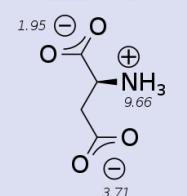
Lysine

Lys K



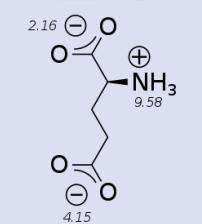
Aspartic Acid

Asp D



Glutamic Acid

Glu E

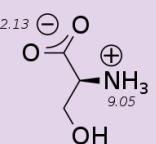


Negative

*B. Amino Acids with Polar Uncharged Side Chains*

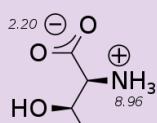
Serine

Ser S



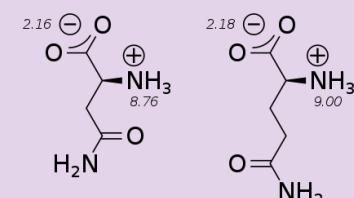
Threonine

Thr T



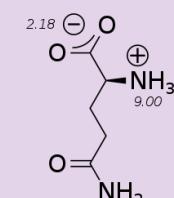
Asparagine

Asn N



Glutamine

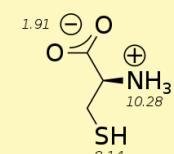
Gln Q



*C. Special Cases*

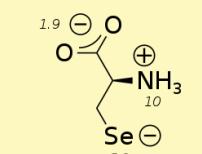
Cysteine

Cys C



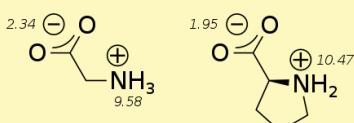
Selenocysteine

Sec U



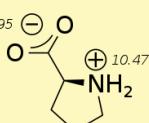
Glycine

Gly G



Proline

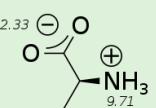
Pro P



*D. Amino Acids with Hydrophobic Side Chains*

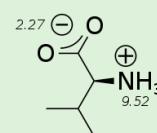
Alanine

Ala A



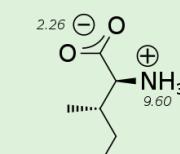
Valine

Val V



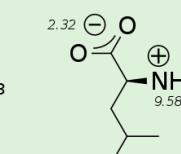
Isoleucine

Ile I



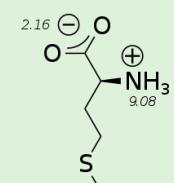
Leucine

Leu L



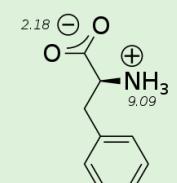
Methionine

Met M



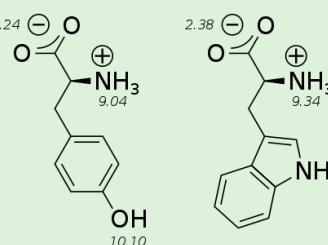
Phenylalanine

Phe F



Tyrosine

Tyr Y

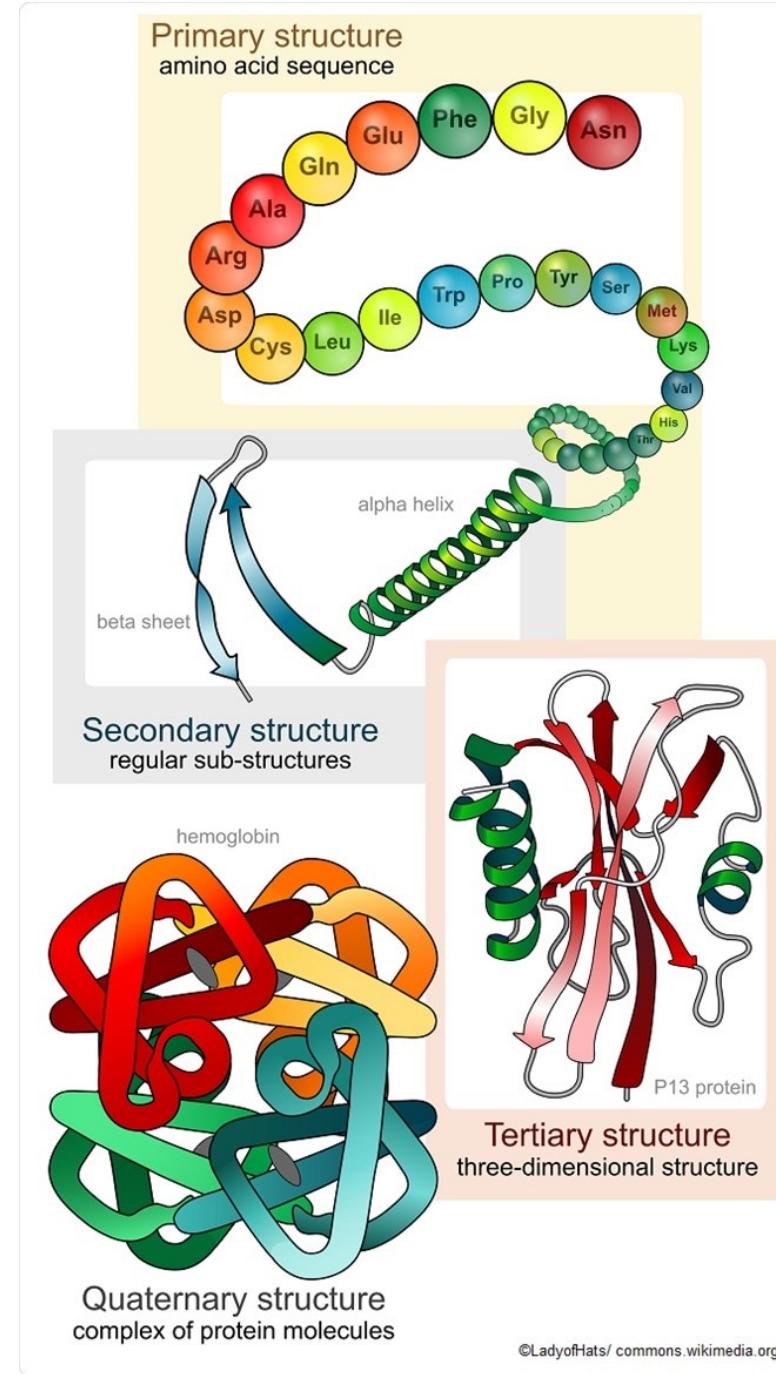


Tryptophan

Trp W



## Structure of the Protein



## Different proteins have different percentages of secondary structure elements

**TABLE 4-4**

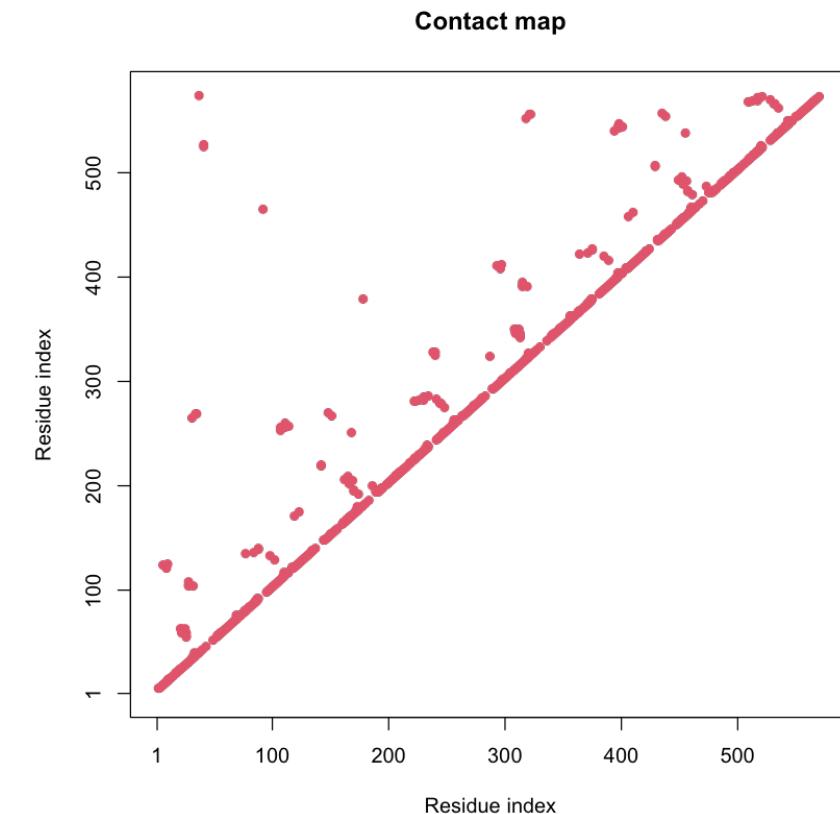
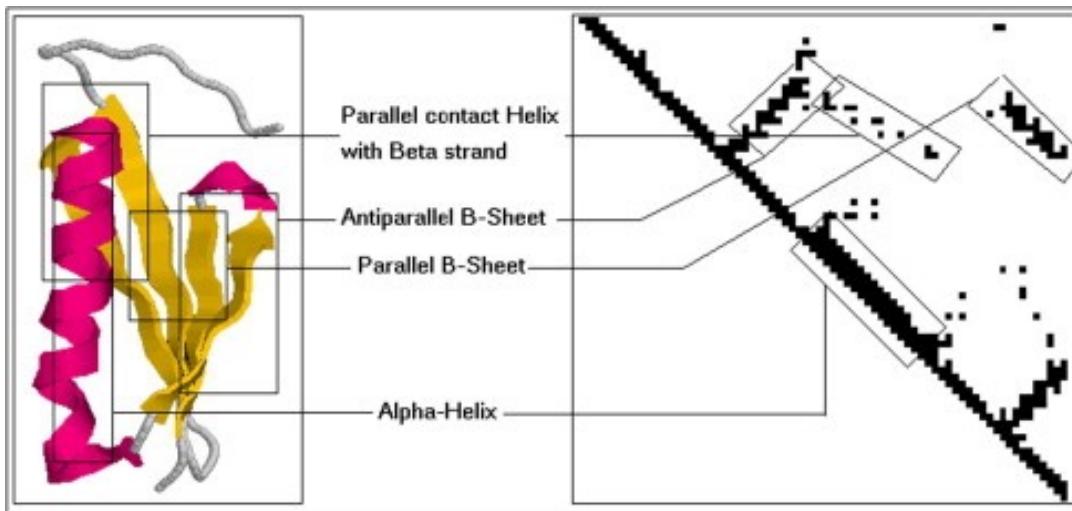
**Approximate Proportion of  $\alpha$  Helix and  
 $\beta$  Conformation in Some Single-Chain  
Proteins**

Protein (total residues)	Residues (%)*	
	$\alpha$ Helix	$\beta$ Conformation
Chymotrypsin (247)	14	45
Ribonuclease (124)	26	35
Carboxypeptidase (307)	38	17
Cytochrome c (104)	39	0
Lysozyme (129)	40	12
Myoglobin (153)	78	0

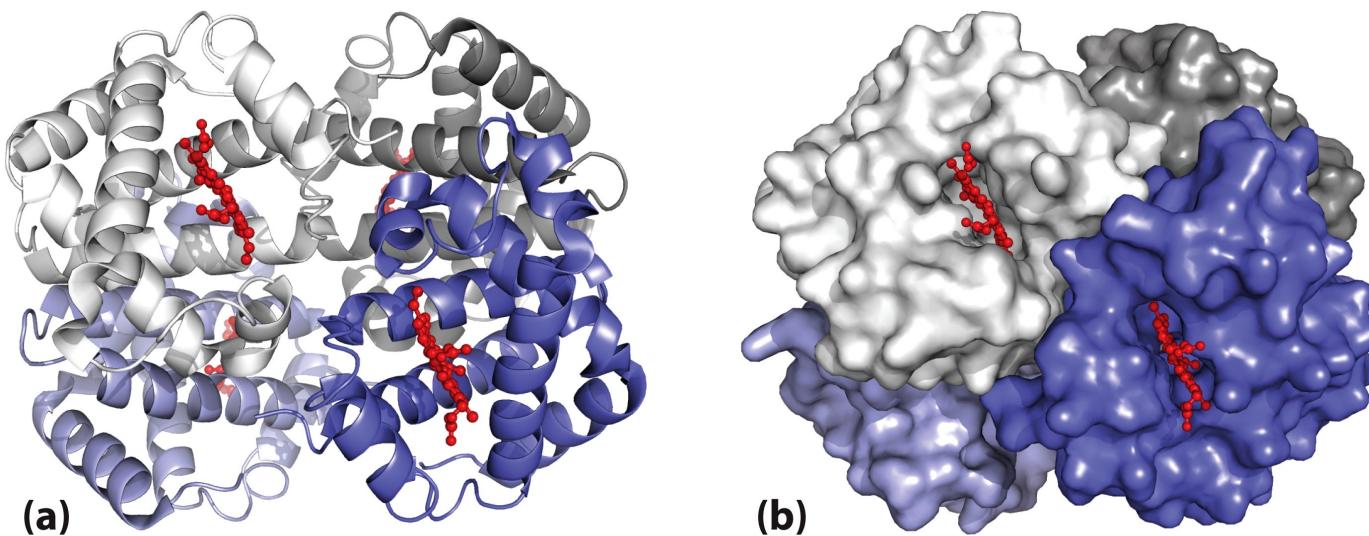
Source: Data from Cantor, C.R. & Schimmel, P.R. (1980) *Biophysical Chemistry, Part I: The Conformation of Biological Macromolecules*, p. 100, W. H. Freeman and Company, New York.

\*Portions of the polypeptide chains not accounted for by  $\alpha$  helix or  $\beta$  conformation consist of bends and irregularly coiled or extended stretches. Segments of  $\alpha$  helix and  $\beta$  conformation sometimes deviate slightly from their normal dimensions and geometry.

# Protein Contact Map



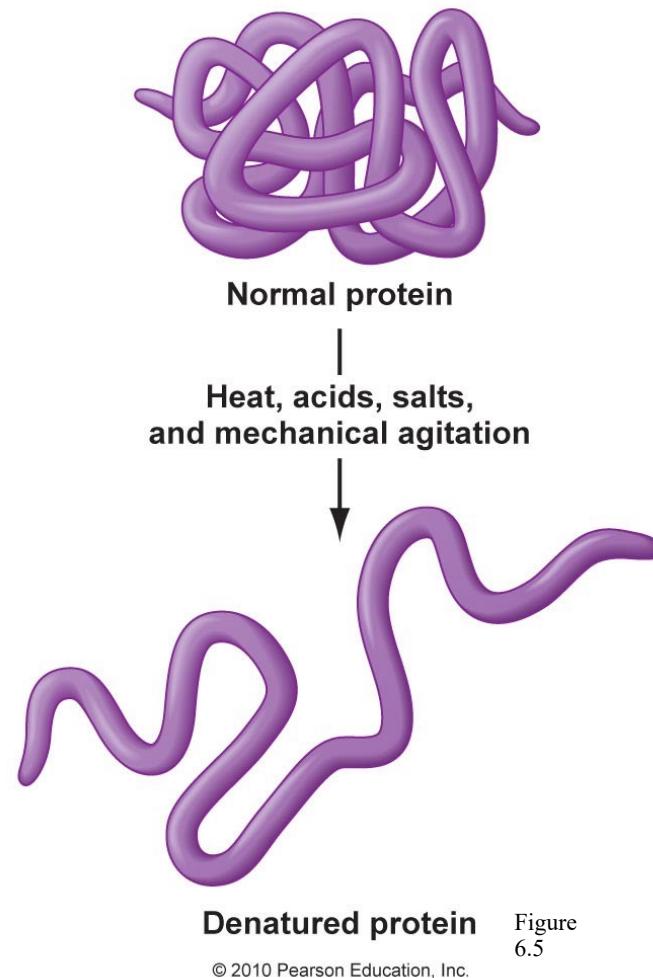
## Quaternary Structure of Hemoglobin



- Hemoglobin ( $M_w$  64,500) contains four polypeptide chains with one heme group each.
- The protein portion, globin, consists of two  $\alpha$  chains (141 residues each) and two  $\beta$  chains (146 residues each).
- The subunits of hemoglobin are arranged in symmetric pairs, each pair having one  $\alpha$  and one  $\beta$  subunit. Hemoglobin can therefore be described either as a tetramer or as a dimer of  $\alpha\beta$  protomers.

# Protein denaturation

- Desaturating agent/factors:
  - Heat
  - Acids
  - Bases
  - Salts
  - Mechanical agitation
  - Mutations  
(interference with folding)

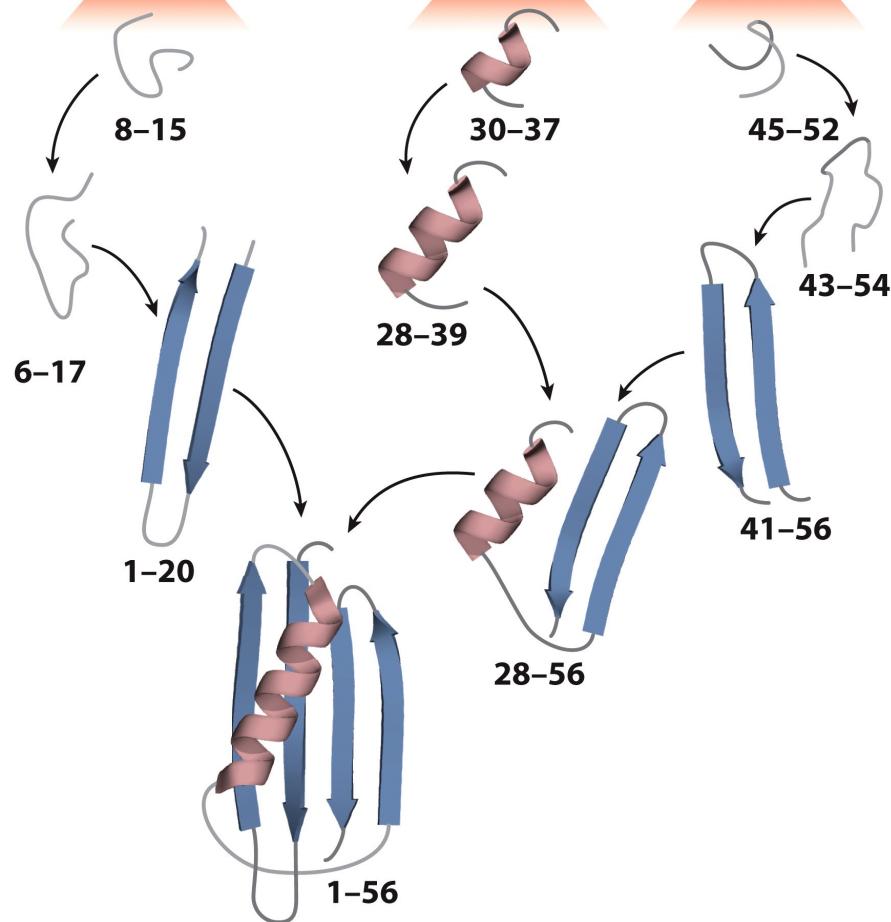


© 2010 Pearson Education, Inc.

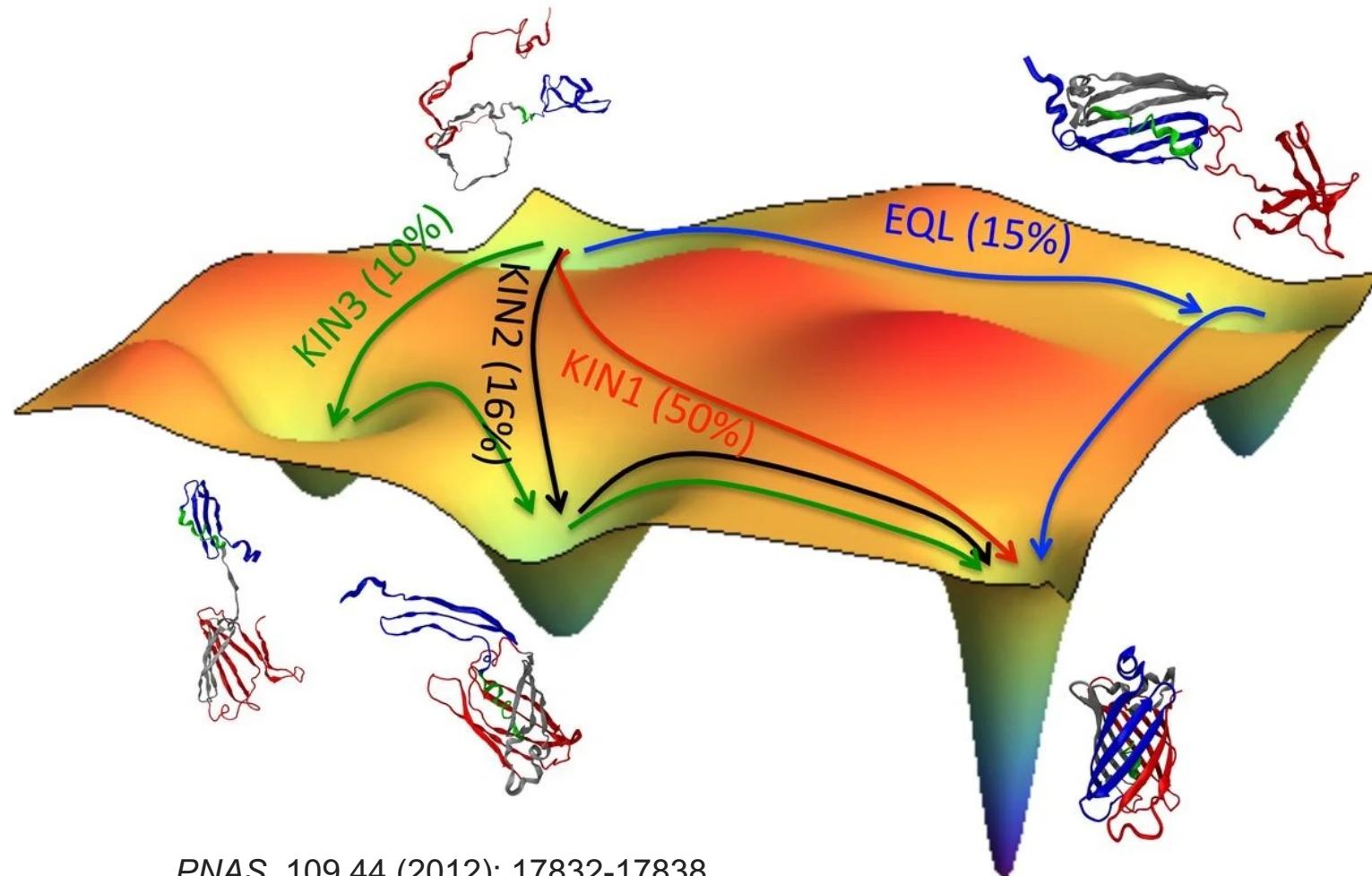
Figure  
6.5

# Overview of Protein Folding

Amino acid sequence of a 56-residue peptide  
MTYKLIL**NGKTLKGE**TTTEAVDAATAEKV **FKQYANDN**GVDGEWT YDDATKTF TVTE



# Protein Folding Intermediates



*PNAS*, 109.44 (2012): 17832-17838.

# Mutational Effects on Protein Stability (FireProt database)

The screenshot shows the FireProtDB homepage. At the top, there's a navigation bar with links for Home, Browse database, Datasets, Use cases, Help, and Acknowledgement. Below the navigation is a search bar with a placeholder "Enter search phrase..." and an "ADVANCED" dropdown. To the right of the search bar is a logo for the Institute of Chemical Technology (IČT).

The main content area has several sections:

- ABOUT**: Describes FireProtDB as a comprehensive, manually curated database of protein stability data for single-point mutants. It highlights the combination of published datasets from ProTherm and ProtaBank, literature search, and own data, leading to data filtering, correction, and validation.
- SOURCES OF DATA**: Shows the flow from various sources like ProTherm, Literature search, ProtaBank, and Own data through Data filtering and Data correction and validation to the final FIREPROT database.
- SOURCES OF ANNOTATIONS**: Details the process of combining Published sets, VariBench, and HOTSPOT WIZARD annotations through Dataset membership and Sequence and structure annotations to reach the FIREPROT database.
- STATISTICS**: Provides metrics such as Number of visitors (10364), Number of experiments (15987), Curated (10352), Uncurated (5635), Number of mutations (6713), Number of proteins (242), and Number of structures (304). It also indicates the last data update on Feb 8, 2022.
- REFERENCES**: Cites Stourac, J., Dubrava, J., Musil, M., Horackova, J., Damborsky, J., Mazurenko, S., Bednar, D. (2020) FireProtDB: Database of Manually Curated Protein Stability Data. Nucleic Acids Research 49: D319-D324.
- DOWNLOADS**: Offers a link to the FireProtDB dump.
- CONTACT**: Provides email (fireprot@sci.muni.cz) and a website (https://loschmidt.chemi.muni.cz).
- OTHER TOOLS**: (not fully visible)

At the bottom, there are two charts:

- Experiments by  $\Delta\Delta G$** : A histogram showing the distribution of experiments based on the change in free energy ( $\Delta\Delta G$ ) in kcal/mol. The x-axis ranges from -12 to 8, and the y-axis ranges from 0 to 6000. The highest frequency is in the (-2, 0] bin.
- Experiments by  $\Delta T_m$** : A histogram showing the distribution of experiments based on the change in melting temperature ( $\Delta T_m$ ) in °C. The x-axis ranges from -15 to > 15, and the y-axis ranges from 0 to 1500. The highest frequency is in the (-5, 0] bin.

What does an average mutation do to a protein ?!

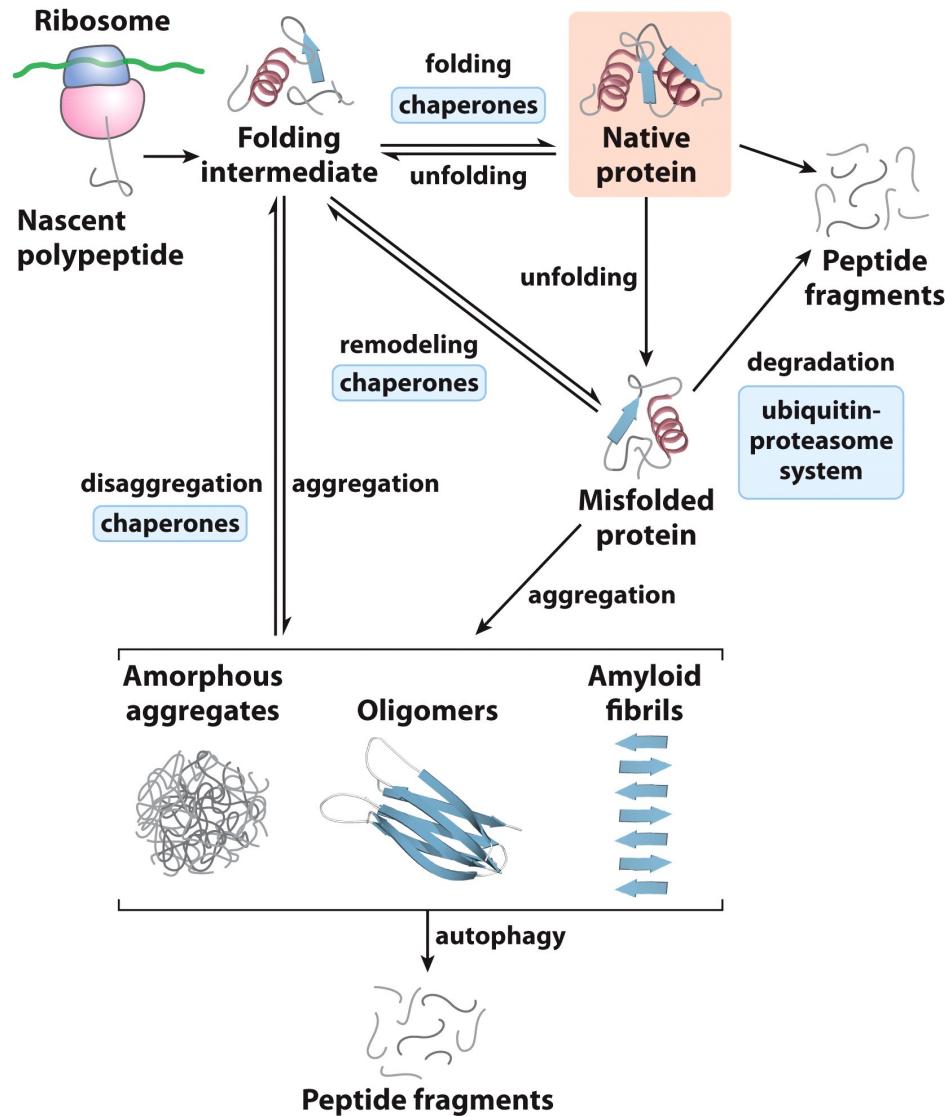
# Proteostasis

The maintenance of an active set of cellular proteins.

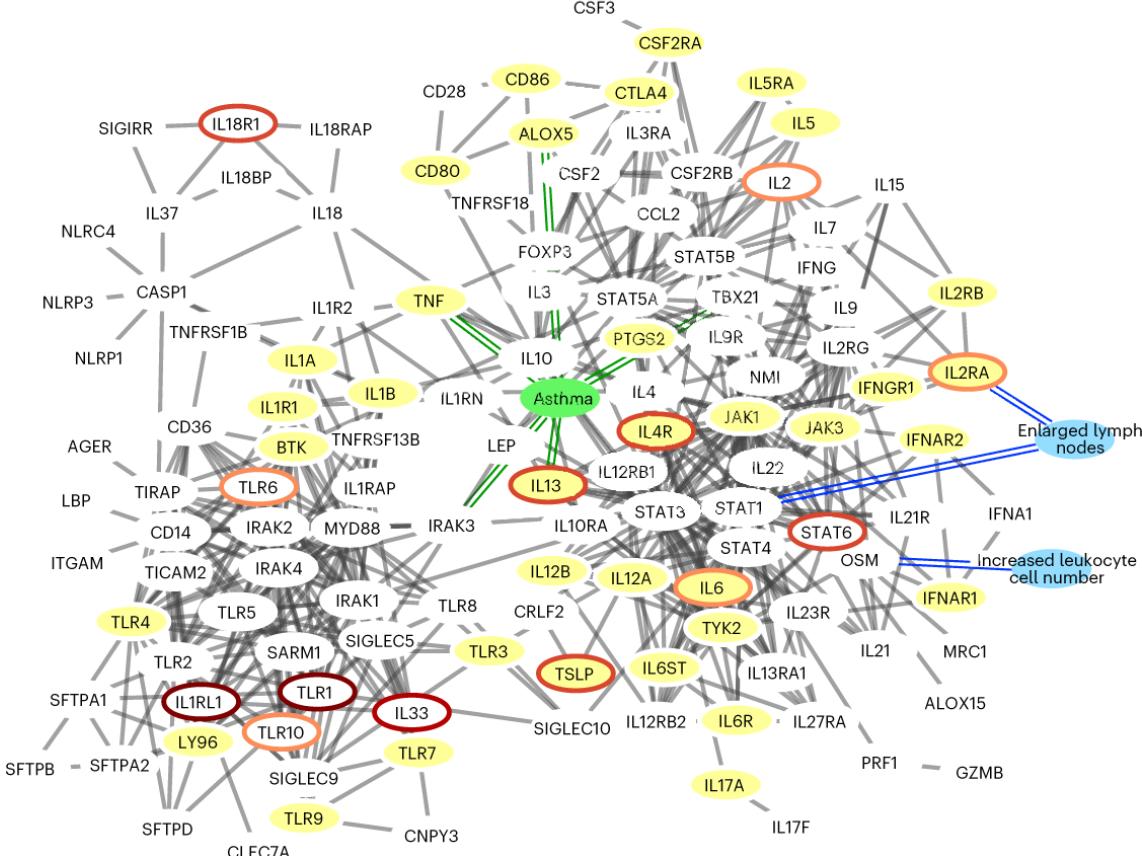
*First*, proteins are synthesized on a ribosome.

*Second*, multiple pathways contribute to protein folding, many of which involve the activity of complexes called chaperones. Chaperones (including chaperonins) also contribute to the refolding of proteins that are partially and transiently unfolded.

*Third*, proteins that are irreversibly unfolded are subject to sequestration and degradation by several additional pathways



# Proteins work together!



Proteins “work together” forming multi complexes to carry out the specific functions

# Protein structure determines interactions with other molecules

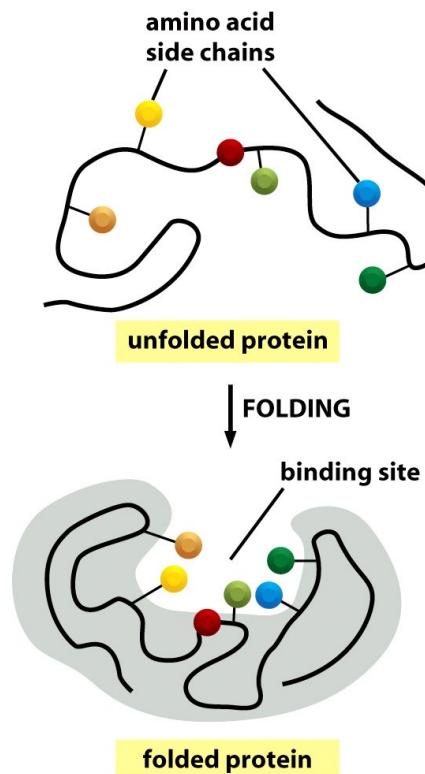
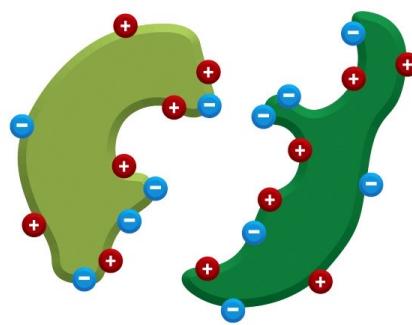


Figure 3-37a Molecular Biology of the Cell 5/e (© Garland Science 2008)

## Protein structure affects protein-protein interactions



# Protein structure affects protein-protein interactions

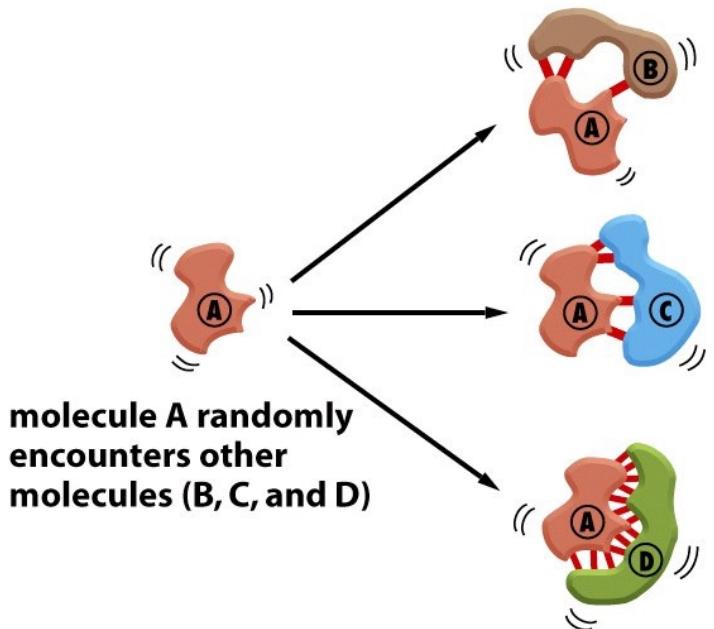
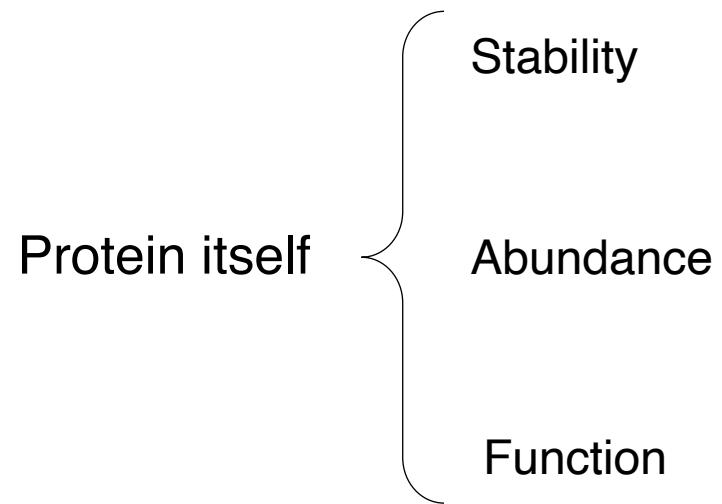


Figure 3-42 Molecular Biology of the Cell 5/e (© Garland Science 2008)

# Disease variant's effect on proteins



Through interactions  
(including regulatory effects)

## **Part 2:**

### **Disease variants of MLH1 (DNA mismatch repair)**



# Uniprot database

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB MLH1 Advanced | List Search Help

Status  
Reviewed (Swiss-Prot) (628)  
Unreviewed (TrEMBL) (45,538)

Popular organisms  
Human (156)  
Mouse (28)  
Rat (20)  
Bovine (14)  
A. thaliana (11)

Taxonomy  
Filter by taxonomy

Group by  
Taxonomy  
Keywords  
Gene Ontology  
Enzyme Class

Feedback

Proteins with  
3D structure (62)  
Active site (18)  
Activity regulation (14)  
Alternative products (isoforms) (62)  
Alternative splicing (62)  
More items

Help

Entry ▲ Entry Name ▲ Protein Names ▲ Gene Names ▲ Organism ▲ Length ▲

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
P40692	MLH1_HUMAN	DNA mismatch repair protein Mlh1[...]	MLH1, COCA2	Homo sapiens (Human)	756 AA
P97679	MLH1_RAT	DNA mismatch repair protein Mlh1[...]	Mlh1	Rattus norvegicus (Rat)	757 AA
Q9JK91	MLH1_MOUSE	DNA mismatch repair protein Mlh1[...]	Mlh1	Mus musculus (Mouse)	760 AA
P38920	MLH1_YEAST	DNA mismatch repair protein MLH1[...]	MLH1, PMS2, YMR167W, YM8520.16	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	769 AA
Q9ZRV4	MLH1_ARATH	DNA mismatch repair protein MLH1[...]	MLH1, At4g09140, F2J3.170, T8A17.9	Arabidopsis thaliana (Mouse-ear cress)	737 AA
Q54KD8	MLH1_DICDI	DNA mismatch repair protein Mlh1[...]	mlh1, DDB_G0287393	Dictyostelium discoideum (Social amoeba)	884 AA
Q9P7W6	MLH1_SCHPO	Putative MutL protein homolog 1[...]	mlh1, SPBC1703.04	Schizosaccharomyces pombe (strain 972 / ATCC 24843) (Fission yeast)	684 AA
A0A2Y9TI72	A0A2Y9TI72_PHYMC	DNA mismatch repair protein Mlh1	MLH1	Physeter macrocephalus (Sperm whale) (Physeter catodon)	758 AA
O95243	MBD4_HUMAN	Methyl-CpG-binding domain protein 4[...]	MBD4, MED1	Homo sapiens (Human)	580 AA
P54278	PMS2_HUMAN	Mismatch repair endonuclease PMS2[...]	PMS2, PMSL2	Homo sapiens (Human)	862 AA
P54277	PMS1_HUMAN	PMS1 protein homolog 1[...]	PMS1, PMSL1	Homo sapiens (Human)	932 AA
P46063	RECOQ1_HUMAN	ATP-dependent DNA helicase Q1[...]	RECQL, RECOQ1, RECQL1	Homo sapiens (Human)	649 AA
Q9UQ84	EXO1_HUMAN	Exonuclease 1[...]	EXO1, EXO1, HEX1	Homo sapiens (Human)	846 AA
Q9UHC1	MLH3_HUMAN	DNA mismatch repair protein Mlh3[...]	MLH3	Homo sapiens (Human)	1,453 AA
I3LT92	I3LT92_PIG	DNA mismatch repair protein Mlh1[...]	MLH1	Sus scrofa (Pig)	757 AA
O49621	MLO1_ARATH	MLO-like protein 1[...]	MLO1, MLO-H1, At4g02600, T10P11.12	Arabidopsis thaliana (Mouse-ear cress)	526 AA
G1QYL9	G1QYL9_NOMLE	MutL homolog 1	MLH1	Nomascus leucogenys (Northern white-cheeked gibbon) (Hylobates leucogenys)	756 AA
G3RJQ3	G3RJQ3_GORGO	MutL homolog 1	MLH1	Gorilla gorilla gorilla (Western lowland gorilla)	756 AA
K7FUJ0	K7FUJ0_PELSI	MutL homolog 1	MLH1	Pelodiscus sinensis (Chinese softshell turtle) (Trionyx sinensis)	727 AA
H0V205	H0V205_CAVPO	MutL homolog 1	MLH1	Cavia porcellus (Guinea pig)	758 AA
I3M618	I3M618_ICTTR	MutL homolog 1	MLH1	Ictidomys tridecemlineatus (Thirteen-lined ground squirrel) (Spermophilus tridecemlineatus)	758 AA
A0A0D9RNU4	A0A0D9RNU4_CHLSB	MutL homolog 1	MLH1	Chlorocebus sabaeus (Green monkey) (Cercopithecus sabaeus)	756 AA
A0A2R9AM85	A0A2R9AM85_PANPA	MutL homolog 1	MLH1	Pan paniscus (Pygmy chimpanzee) (Bonobo)	756 AA

# Uniprot database

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB Advanced | List Search

**P40692 · MLH1\_HUMAN**

**Function**

**Names & Taxonomy**

**Subcellular Location**

**Disease & Variants**

**PTM/Processing**

**Expression**

**Interaction**

**Structure**

**Family & Domains**

**Sequence & Isoforms**

**Similar Proteins**

**Entry** Variant viewer 2,139 Feature viewer Genomic coordinates Publications External links History

BLAST Align Download Add Add a publication Entry feedback

**Function<sup>i</sup>**

Heterodimerizes with PMS2 to form MutL alpha, a component of the post-replicative DNA mismatch repair system (MMR). DNA repair is initiated by MutS alpha (MSH2-MSH6) or MutS beta (MSH2-MSH3) binding to a dsDNA mismatch, then MutL alpha is recruited to the heteroduplex. Assembly of the MutL-MutS-heteroduplex ternary complex in presence of RFC and PCNA is sufficient to activate endonuclease activity of PMS2. It introduces single-strand breaks near the mismatch and thus generates new entry points for the exonuclease EXO1 to degrade the strand containing the mismatch. DNA methylation would prevent cleavage and therefore assure that only the newly mutated DNA strand is going to be corrected. MutL alpha (MLH1-PMS2) interacts physically with the clamp loader subunits of DNA polymerase III, suggesting that it may play a role to recruit the DNA polymerase III to the site of the MMR. Also implicated in DNA damage signaling, a process which induces cell cycle arrest and can lead to apoptosis in case of major DNA damages. Heterodimerizes with MLH3 to form MutL gamma which plays a role in meiosis. [5 Publications]

**Features**

Showing features for binding site<sup>i</sup>.

Download

1 100 200 300 400 500 600 700 756

TYPE	ID	POSITION(S)	DESCRIPTION	BLAST	Add
-- Select --					
▶ Binding site	38	ATP (UniProtKB   ChEBI ⓘ )	[1 Publication] [Combined Sources]		
▶ Binding site	63	ATP (UniProtKB   ChEBI ⓘ )	[1 Publication] [Combined Sources]		
▶ Binding site	82-84	ATP (UniProtKB   ChEBI ⓘ )	[1 Publication] [Combined Sources]	BLAST	Add
▶ Binding site	100-104	ATP (UniProtKB   ChEBI ⓘ )	[1 Publication] [Combined Sources]	BLAST	Add

**GO annotations<sup>i</sup>**

Access the complete set of GO annotations on QuickGO ⓘ

Expand table

## Class activity

Group 1: Function

Group 2: Feature

Group 3: GO annotation

Group 4: Expression

Group 5: Interaction

# HPA database

# THE HUMAN PROTEIN ATLAS

SECTIONS ABOUT NEWS LEARN DATA HELP

The open access resource for human proteins

Search for specific genes/proteins or explore the 12 different sections

e.g. ACE2, GFAP, EGFR

Search Fields »

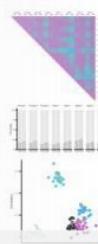
Search help

## News

### The role of trypsin in bottom-up proteomics moving towards clinical application

*The impact of different trypsins on commonly used bottom-up proteomics workflows were investigated across a range of sample types in a publication in Analytical Chemistry. Besides the type of trypsin, different protease concentrations, sample matrices, instrument acquisition techniques, and protease storage times were investigated... [Read more](#)*

*read the latest article - published Wed, 3 Jan 2024*



### The Human Protein Atlas (HPA) has been selected as a Global Core Biodata Resource

We are proud to announce that the Human Protein Atlas is one of the newly appointed Global Core Biodata Resources (GCBR) selected by the Global Biodata Coalition (GBC).

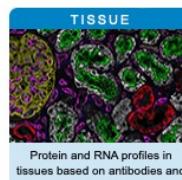


### The role of trypsin in bottom-up proteomics moving towards clinical application

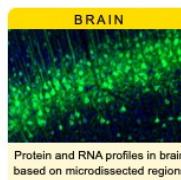
The impact of different trypsins on commonly used bottom-up proteomics workflows were investigated across a range of sample types in a publication in Analytical Chemistry.



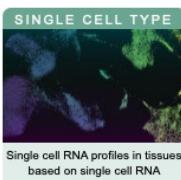
[all news articles](#)



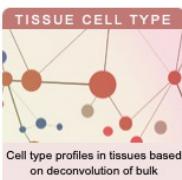
Protein and RNA profiles in tissues based on antibodies and transcriptomics



Protein and RNA profiles in brain based on microdissected regions



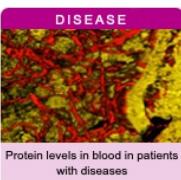
Single cell RNA profiles in tissues based on single cell RNA analysis



Cell type profiles in tissues based on deconvolution of bulk transcriptomics



Protein and RNA profiles in human cancers based on antibodies and transcriptomics



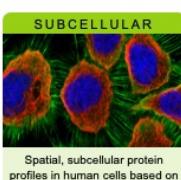
Protein levels in blood in patients with diseases



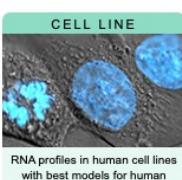
RNA profiles in human immune cells



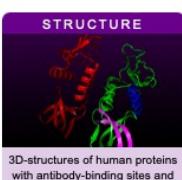
Blood protein levels in healthy individuals and the human secretome



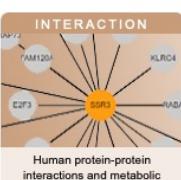
Spatial, subcellular protein profiles in human cells based on antibodies



RNA profiles in human cell lines with best models for human cancers



3D-structures of human proteins with antibody-binding sites and genetic variants



Human protein-protein interactions and metabolic enzyme profiles

<https://www.proteinatlas.org>

# GTEx portal



[About GTEx](#) [Publications](#) [Access Biospecimens](#) [FAQs](#) [Contact](#)

Home Downloads ▾ Expression ▾ Single Cell ▾ QTL ▾ IGV Browser Tissues & Histology ▾ Documentation ▾

Search Gene or SNP ID...

2023-02-23

## GTEx API V2 Released, GTEx API V1 Deprecated

We have released the GTEx API V2. This version of our API provides improved query parameter validation, improved support for querying by tissue ontology identifiers, support for paginated queries, and improved documentation. The documentation for the...

## Resource Overview

### Current Release (V8)

Tissue & Sample Statistics  
Tissue Sampling Info (Anatomogram)  
Access & Download Data  
Release History  
How to cite GTEx?

The Genotype-Tissue Expression (GTEx) project is an ongoing effort to build a comprehensive public resource to study tissue-specific gene expression and regulation. Samples were collected from 54 non-diseased tissue sites across nearly 1000 individuals, primarily for molecular assays including WGS, WES, and RNA-Seq. Remaining samples are available from the GTEx Biobank. The GTEx Portal provides open access to data including gene expression, QTLs, and histology images.



### Browse

#### By gene ID

Browse and search all data by gene

#### By variant or rs ID

Browse and search all data by variant

#### By Tissue

Browse and search all data by tissue

#### Histology Viewer

Browse and search GTEx histology images



### Single Cell

#### Data Overview

Learn more about available single cell data

#### Multi-Gene Single Cell Query

Browse and search single cell expression by gene and tissue

<https://www.gtexportal.org/home/>

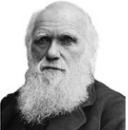
# Paxdb database

 PaxDb 5.0 PaxDb: Protein Abundance Database

All organisms  protein(s) id/name   
ctrl+enter to search

Browse species New species!

All > Eukaryotes (67) > Animals (31) > Vertebrates (22)  
-> Mammals (15)

							
P. troglodytes	H. sapiens	M. mulatta <small>New</small>	C. lupus <small>New</small>	C. familiaris	E. caballus	S. scrofa	B. taurus
							
B. bubalis <small>New</small>	O. aries <small>New</small>	M. musculus	R. norvegicus	C. griseus <small>New</small>	O. cuniculus <small>New</small>	F. catus <small>New</small>	

MLH1

*Homo sapiens* protein

ENSP00000231790 : DNA mismatch repair protein MLH1; Heterodimerizes with PMS2 to form MutL alpha, a component of the post-replicative DNA mismatch repair system (MMR). DNA repair is initiated by MutS alpha (MSH2-MSH6) or MutS beta (MSH2-MSH6) binding to a dsDNA mismatch, then MutL alpha is recruited to the heteroduplex. Assembly of the MutL-MutS- heteroduplex ternary complex in presence of RFC and PCNA is sufficient to activate endonuclease activity of PMS2. It introduces single-strand breaks near the mismatch and thus generates new entry points for the exonuclease EXO1 to degrade the strand containing [...]:

MLH1 in: UniProt

MLH1 abundance information

Dataset	Tissue type	Abundance	Dataset score	IF
H.sapiens - Whole organism, SC (PeptideAtlas,aug,2014)	whole organism	8.37 ppm, 6148. out of 16376	38.1	
H.sapiens - Lymph node (Integrated)	lymph node	5.41 ppm, 4944. out of 10447	37.2	
H.sapiens - Cell line, Cd8, SC (Kim,nature,2014)	cell line	10.4 ppm, 5274. out of 11016	37	
H.sapiens - Female gonad (Integrated)	female gonad	13.2 ppm, 4907. out of 12486	36.9	
H.sapiens - Testis (Integrated)	testis	10.8 ppm, 6260. out of 15560	36.7	

Protein Abundance Histogram

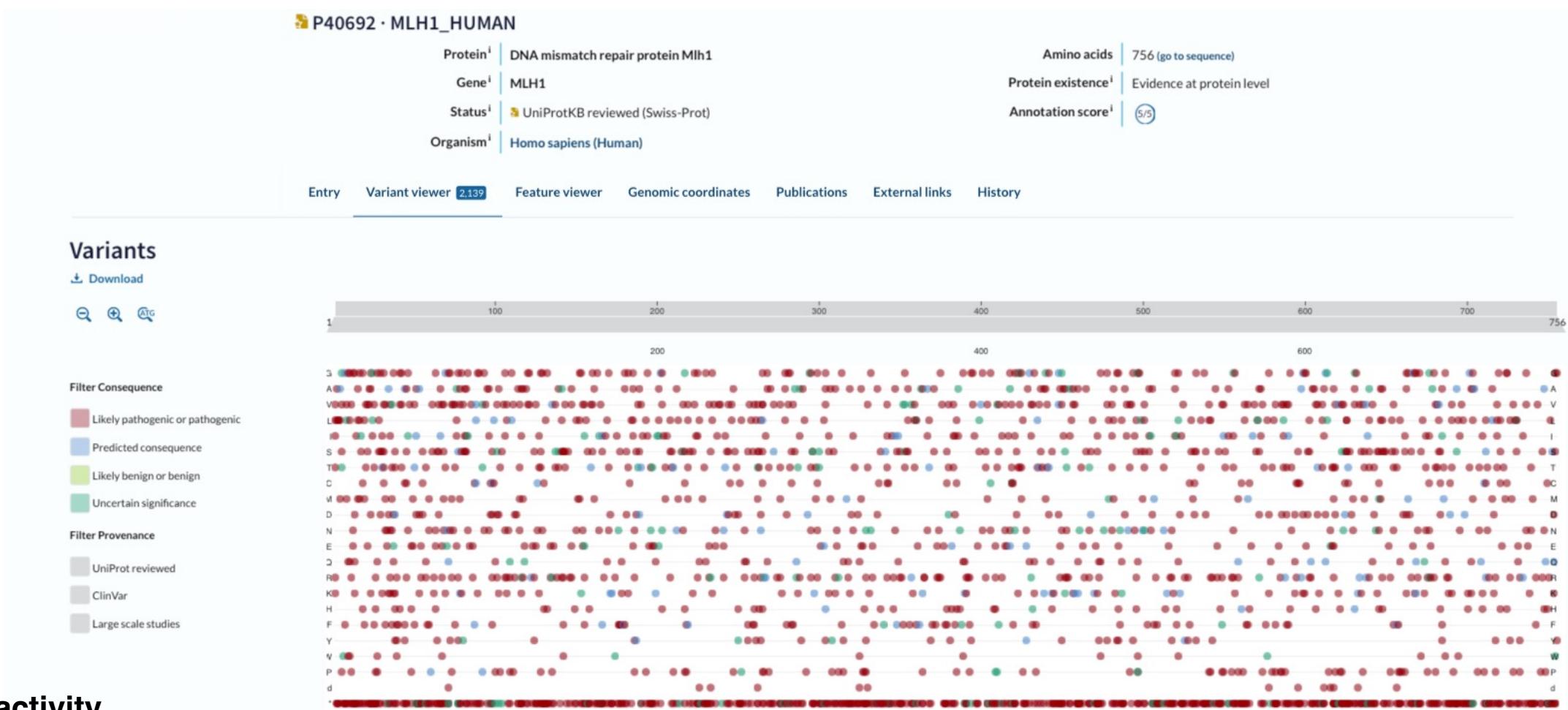
H.sapiens - Whole organism (Integrated)

protein abundance (ppm) [MLH1 marked red]

1 of 50

<https://pax-db.org>

# Uniprot database: variant page



## Class activity

Group 1: 1-60

Group 2: 61-120

Group 3: 121- 180

Group 4: 180:240

Group 5: > 240

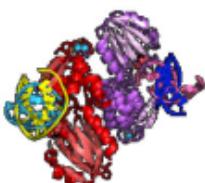
Look at pathological/clinical variants that fall in these position bins

# PDBsum database (EMBL-EBI)

[Services](#)[Research](#)[Training](#)[About us](#)

## PDBsum

Pictorial database of 3D structures in the Protein Data Bank

**Browse options:**[List of PDB codes](#)[Het Groups](#)[Ligands](#)[Drugs](#)[Enzymes](#)**Generate****Figures from Papers**[Gallery](#)[Figure stats](#)**Documentation****Downloads****Acknowledgements****Contact us**

Databases > Structure Databases > PDBsum



PDBsum is a pictorial database that provides an at-a-glance overview of the contents of each 3D structure deposited in the Protein Data Bank ([PDB](#)). It shows the molecule(s) that make up the structure (ie protein chains, DNA, ligands and metal ions) and schematic diagrams of the interactions between them.  
**NEW** Also included are the [AlphaFold](#) predicted models for all human protein for comparison with experimentally determined structures. [Read more ...](#)

**PDB code**

(4 chars)

Example: "[1kfv](#)"**NEW Alpha Fold model (human proteins only)**Enter UniProt accession (or UniProt id), to find Alpha Fold model of given protein. Eg [P00734 \(THR8\\_HUMAN\)](#).see [Analyses](#)**Text search**

Scans all TITLE, HEADER, COMPND, SOURCE and AUTHOR records in the PDB (eg to find a given protein by name).

**Search by sequence**

Perform FASTA search vs all sequences in the PDB to get a list of the closest matches.

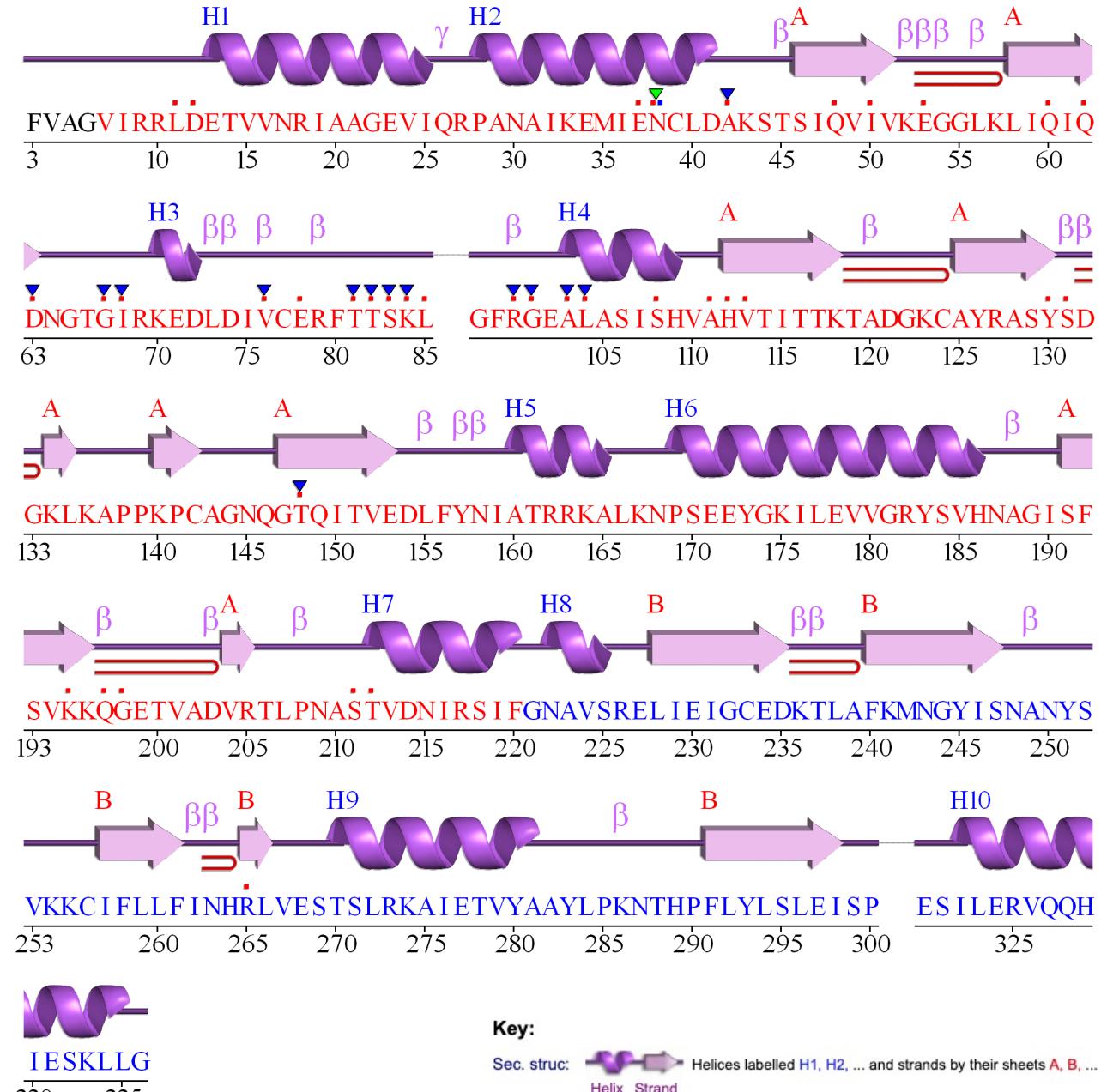
**Search by****UniProt id:**(eg [P03023, LACI\\_ECOLI](#), etc)**Pfam id:**(eg [PF07992](#))**Ensembl id:**(eg [ENSG00000086205, ENST00000256999](#))**Contents**

PDBsum contains

206,449 entries, including  
1,781 superseded

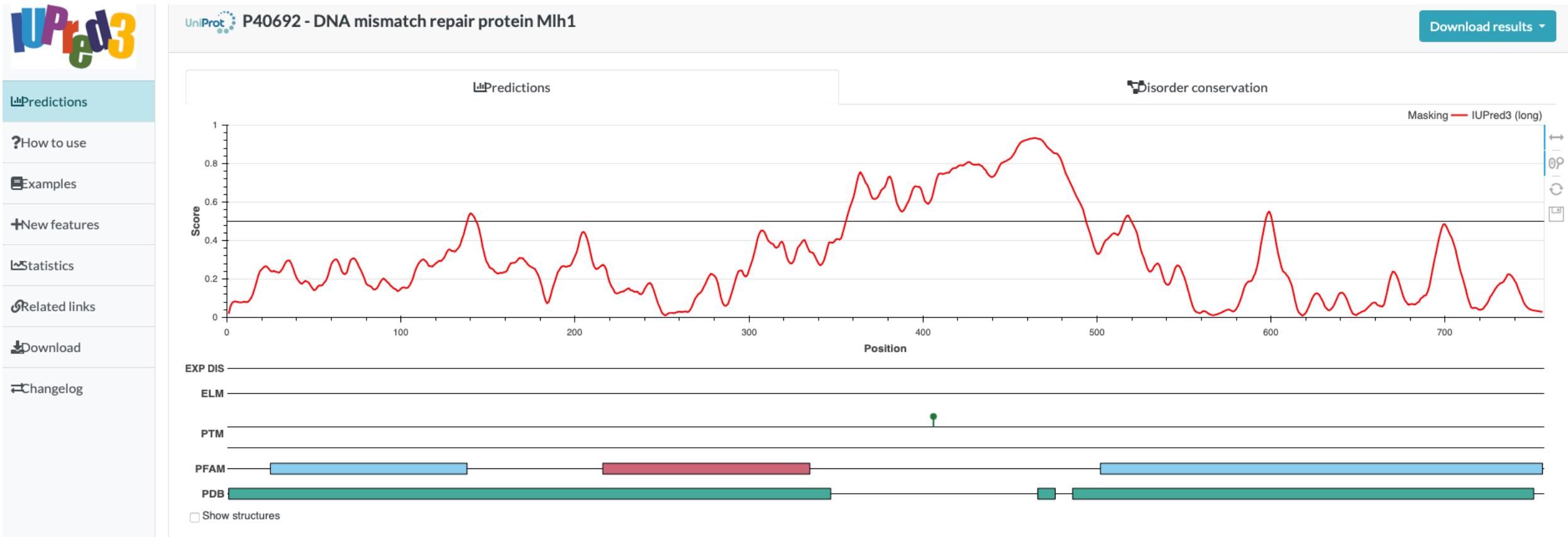
Last update: 10 April, 2023

**In-house version**Download PDBsum 1 to  
process your own  
structures in-house**Related databases**Enzyme 3D structures  
organized by the E.C.  
numbering hierarchy.Structures of drugs and  
their target proteins in the  
PDB.



Did you notice any difference between the protein length (in structure) and the gene length?

# IUPred database



# AlphaFold (EMBL)

AlphaFold Protein Structure Database

Home About FAQs Downloads API

Search for protein, gene, UniProt accession or organism or sequence search BETA Search

Examples: MENFQKVKEKIGEGTYGV... Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli See search help ↗

## DNA mismatch repair protein Mlh1

AlphaFold structure prediction

Download PDB file mmCIF file Predicted aligned error

Share your feedback on structure with DeepMind Looks great Could be improved

### Information

Protein DNA mismatch repair protein Mlh1

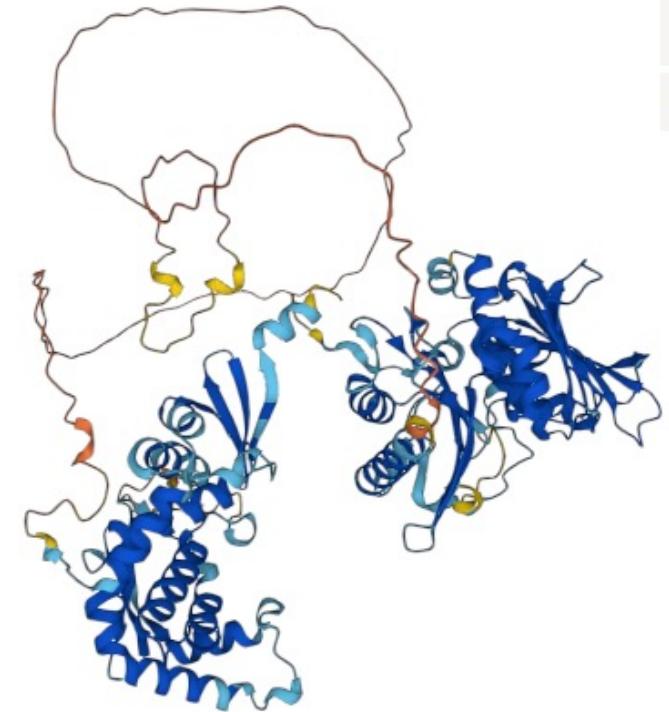
Gene MLH1

Source organism Homo sapiens (Human) [go to search](#) ↗

UniProt P40692 [go to UniProt](#) ↗

Experimental structures 7 structures in PDB for P40692 [go to PDBe-KB](#) ↗

Biological function Heterodimerizes with PMS2 to form MutL alpha, a component of the post-replicative DNA mismatch repair system (MMR). DNA repair is initiated by MutS alpha (MSH2-MSH6) or MutS beta (MSH2-MSH3) binding to a dsDNA mismatch, then MutL alpha is recruited to the heteroduplex. Assembly of the MutL-MutS-heteroduplex ternary complex in presence of RFC and PCNA is sufficient to activate endonuclease activity of PMS2. It introduces single-strand breaks near the mismatch and ... [+ \[show more\]](#) [go to UniProt](#) ↗



<https://alphafold.ebi.ac.uk>



What clinical variants fall in the disordered regions?



214,226 Structures from the PDB  
1,068,577 Computed Structure Models (CSM)

▼ 3D Structures [?](#)

Enter search term(s), Entry ID(s), or sequence

Include CSM [?](#)



[Advanced Search](#) | [Browse Annotations](#)

[Help](#)

PDB-101

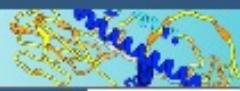
OPDB

EMDataResource

NAKB

wwwPDB

PDB-Dev



Access Computed Structure Models (CSMs) of all available model organisms

[Learn more](#)

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

Experimentally-determined 3D structures from the [Protein Data Bank \(PDB\)](#) archive

[Computed Structure Models \(CSM\)](#) from AlphaFold DB and ModelArchive

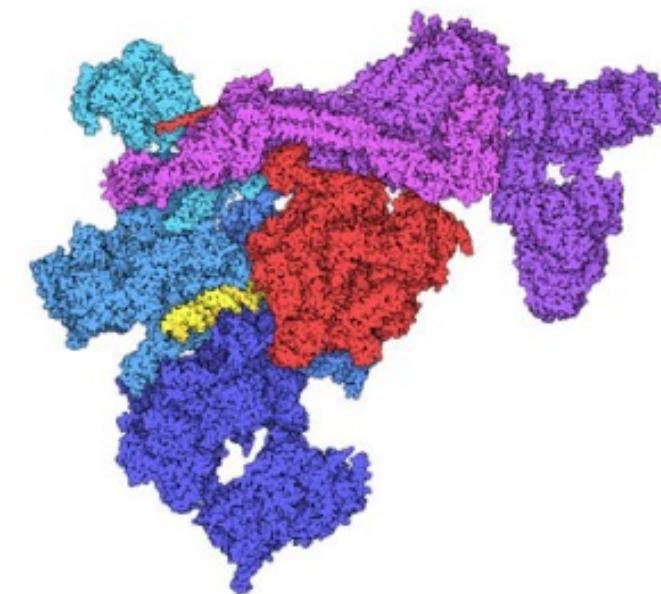
These data can be explored in context of external annotations providing a structural view of biology.

Explore  
NEW  
Features



PDB-101  
Training  
Resources

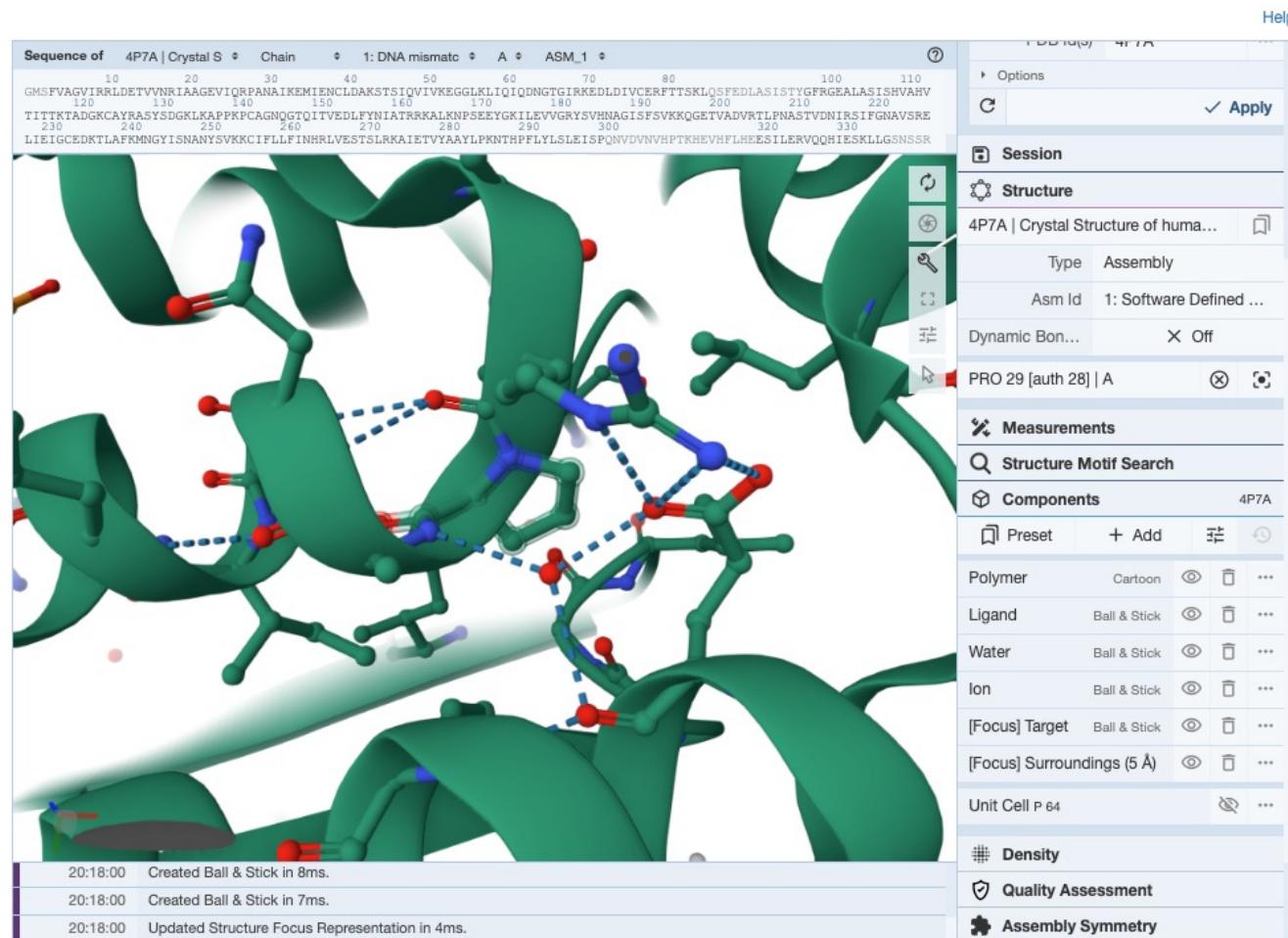
## January Molecule of the Month



Mediator

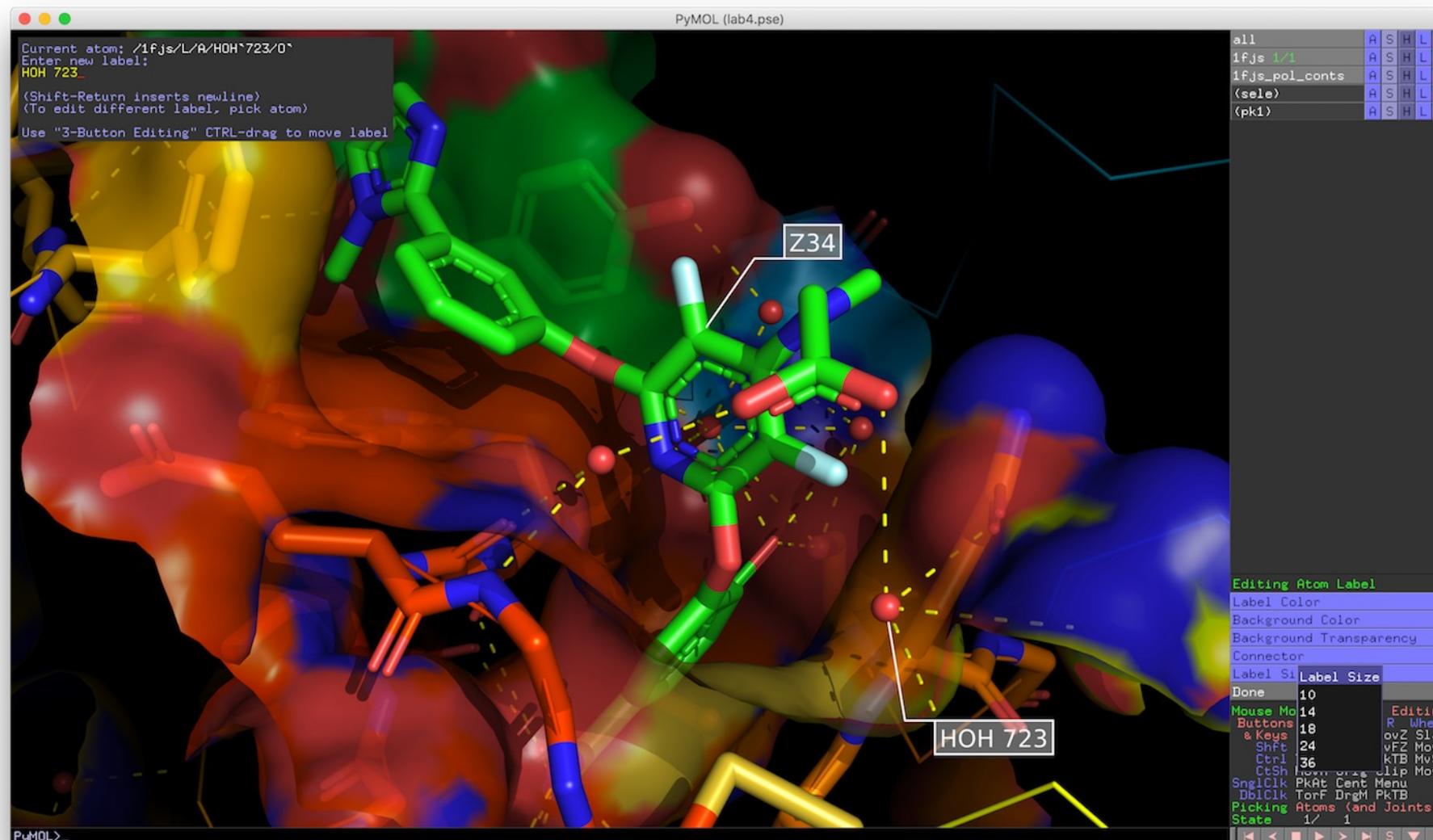
## Mol\* 3D Viewer

This version of Mol\* can be used to upload single or multiple data files and align structures.

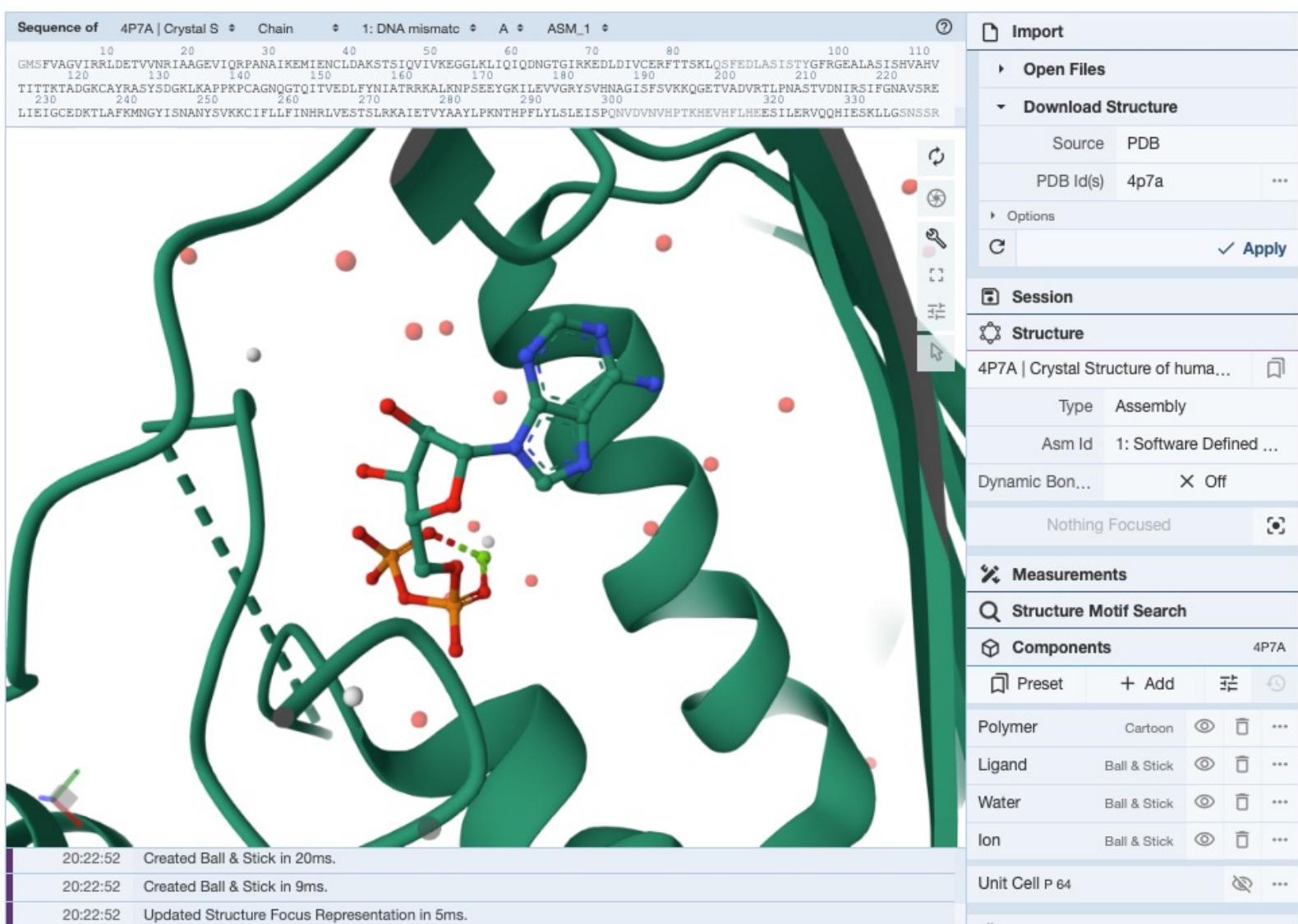


Disease variant:  
P28L

## Pymol: Protein and Molecular Visualization



## Which variants might affect the ATP binding affinity?



Cross validate your results with PDBsum (EMBL) database