

In the Name of God, the Merciful, the Compassionate

Introduction to Bioinformatics

11 - Protein Structure Basics

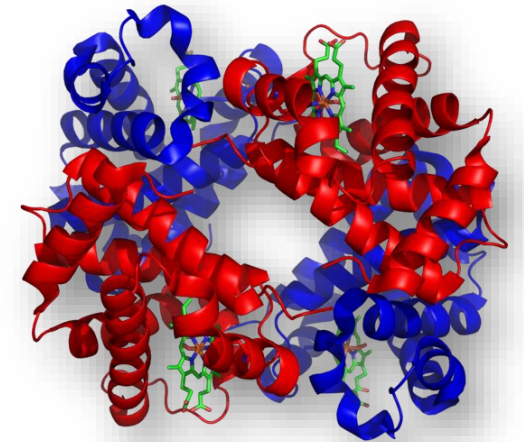
Instructor: Hossein Zeinali

Amirkabir University of Technology

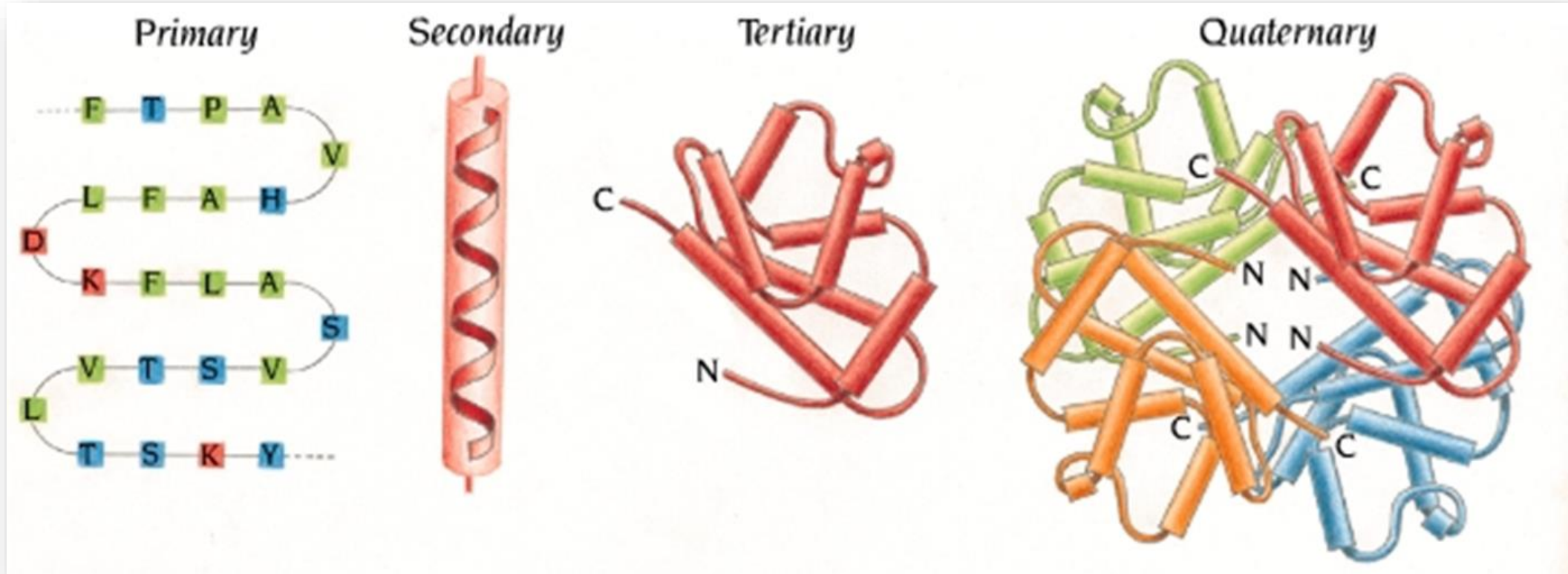


Protein Structure & Function

- Proteins perform most essential biological and chemical functions in a cell.
 - **Protein structure** - primarily determined by sequence
 - **Protein function** - primarily determined by structure
- Globular proteins:
 - compact hydrophobic core & hydrophilic surface
- Membrane proteins: special hydrophobic surfaces
- Folded proteins are only marginally stable
- Predicting protein structure and function can be very hard



4 Basic Levels of Protein Structure



- **Primary**

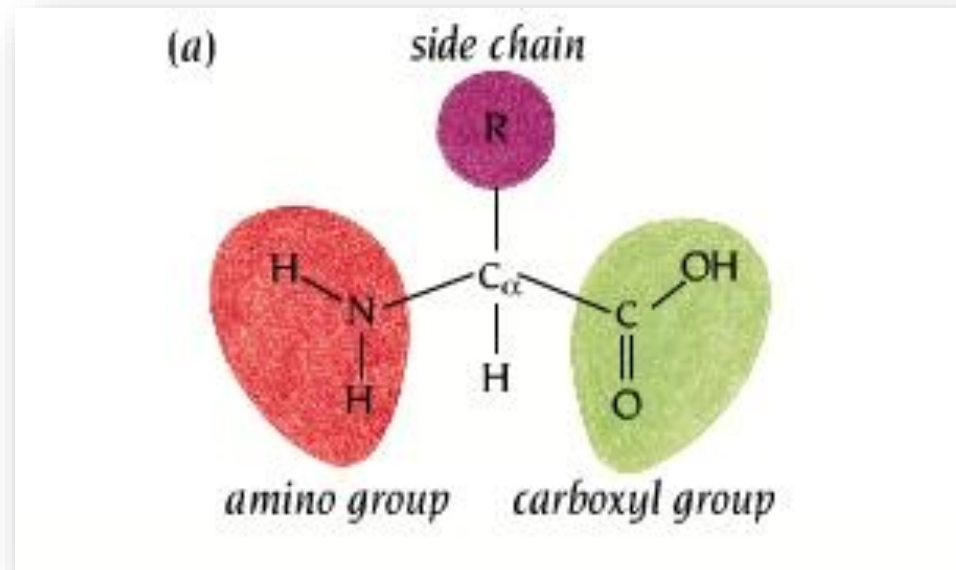
- Linear sequence of amino acids
- Description of **covalent bonds** linking aa's

Amirkabir University of Technology
(Tehran Polytechnic)

4 Basic Levels of Protein Structure

- **Secondary**
 - *Local* spatial arrangement of amino acids
 - Description of *short-range non-covalent* interactions
 - Periodic structural patterns: α -helix, β -sheet
- **Tertiary**
 - Overall 3-D "fold" of a single polypeptide chain
 - Spatial arrangement of 2' structural elements; packing of these into compact "domains"
 - Description of *long-range non-covalent* interactions (plus disulfide bonds)
- **Quaternary**
 - In proteins with > 1 polypeptide chain, spatial arrangement of subunits

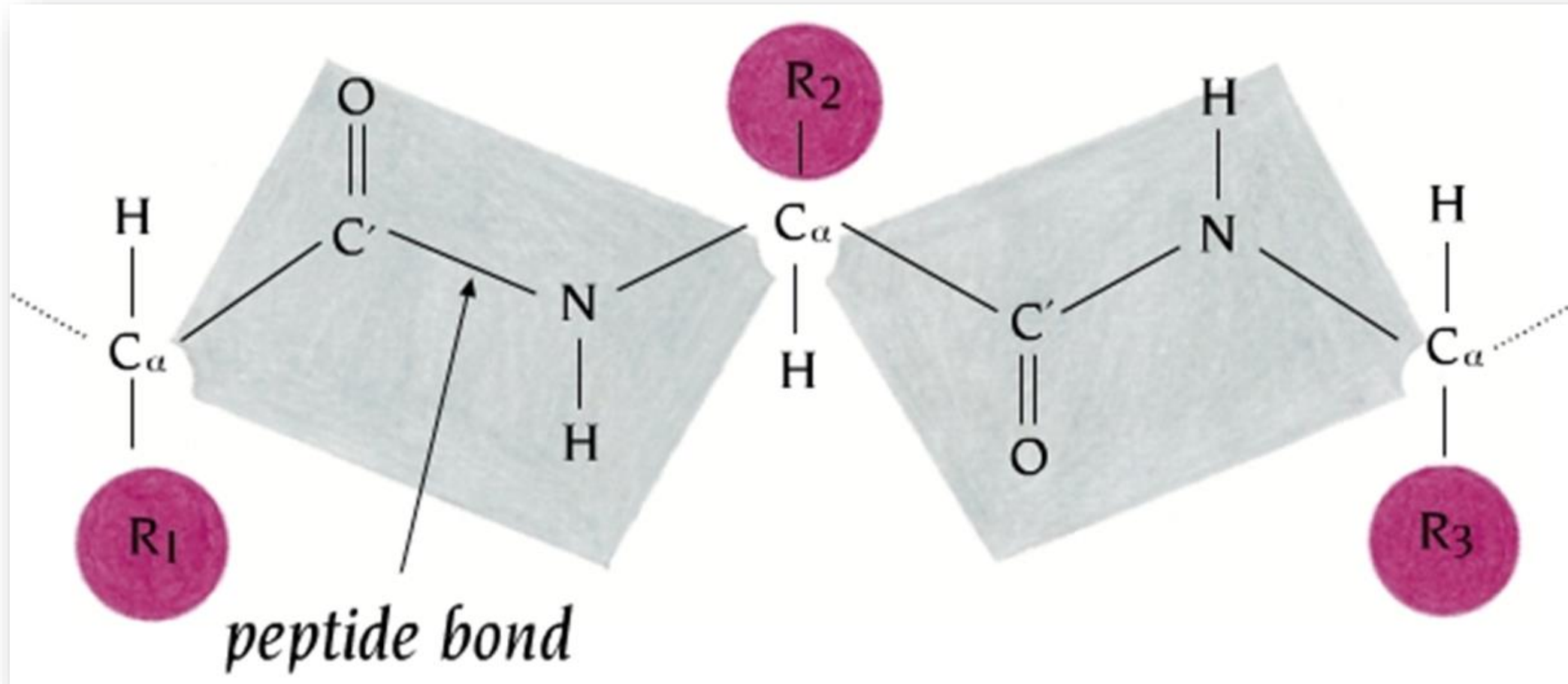
Amino Acids



- Each of 20 different amino acids has different "R-Group" or side chain attached to C_{α}

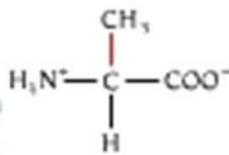
University of Technology
(Tehran Polytechnic)

Peptide Bond is Rigid and Planar

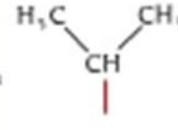


Hydrophobic Amino Acids

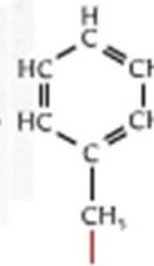
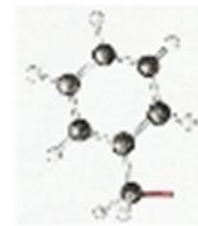
(a) Hydrophobic amino acids



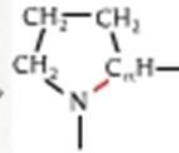
A Ala, Alanine



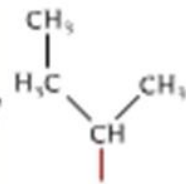
V Val, Valine



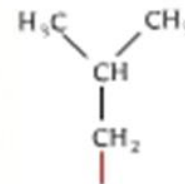
F Phe, Phenylalanine



P Pro, Proline



I Ile, Isoleucine



L Leu, Leucine



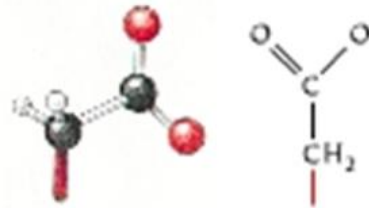
M Met, Methionine

- *Aliphatic side chains* are linear hydrocarbon chains and *aromatic side chains* are cyclic rings.

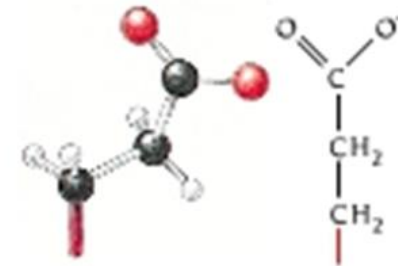
Charged Amino Acids

- The hydrophilic amino acids can be subdivided into **polar** and **charged**.
- *Charged amino acids* can be either positively charged (basic) or negatively charged (acidic).

(b) Charged amino acids



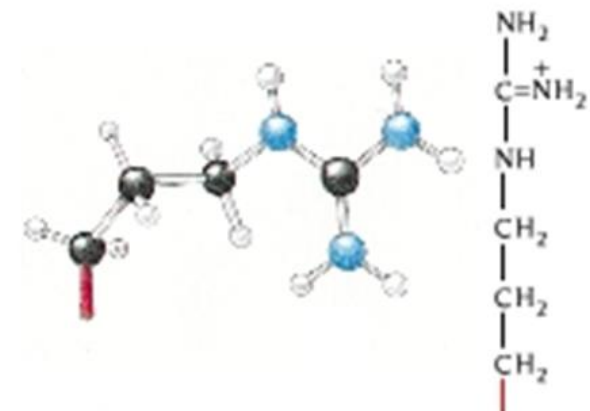
D Asp, Aspartic acid



E Glu, Glutamic acid



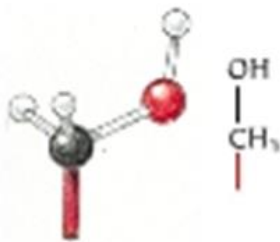
K Lys, Lysine



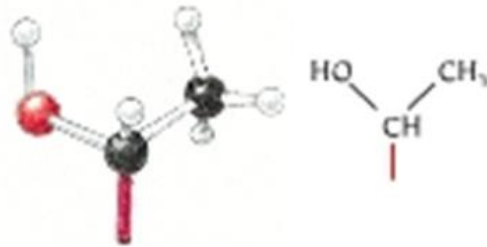
R Arg, Arginine

Polar Amino Acids

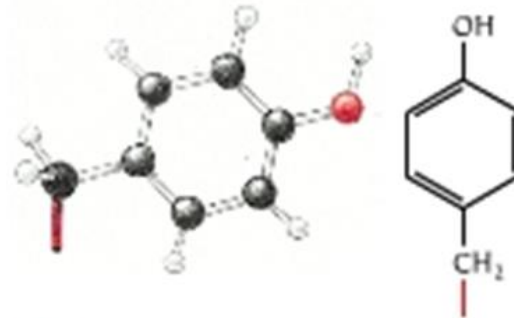
(C) Polar amino acids



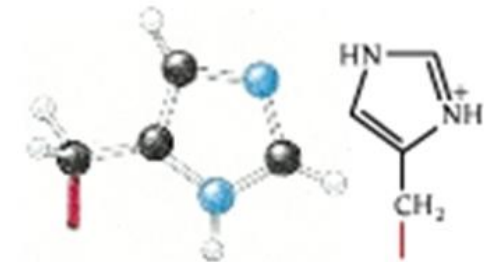
S Ser, Serine



T Thr, Threonine



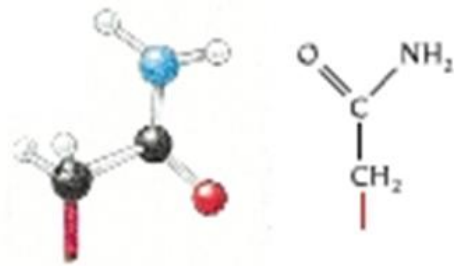
Y Tyr, Tyrosine



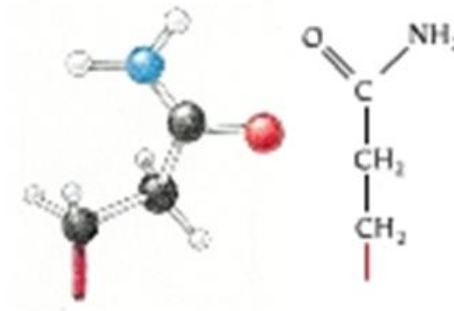
H His, Histidine



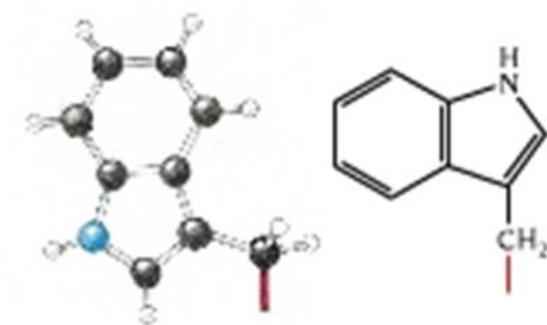
C Cys, Cysteine



N Asn, Asparagine



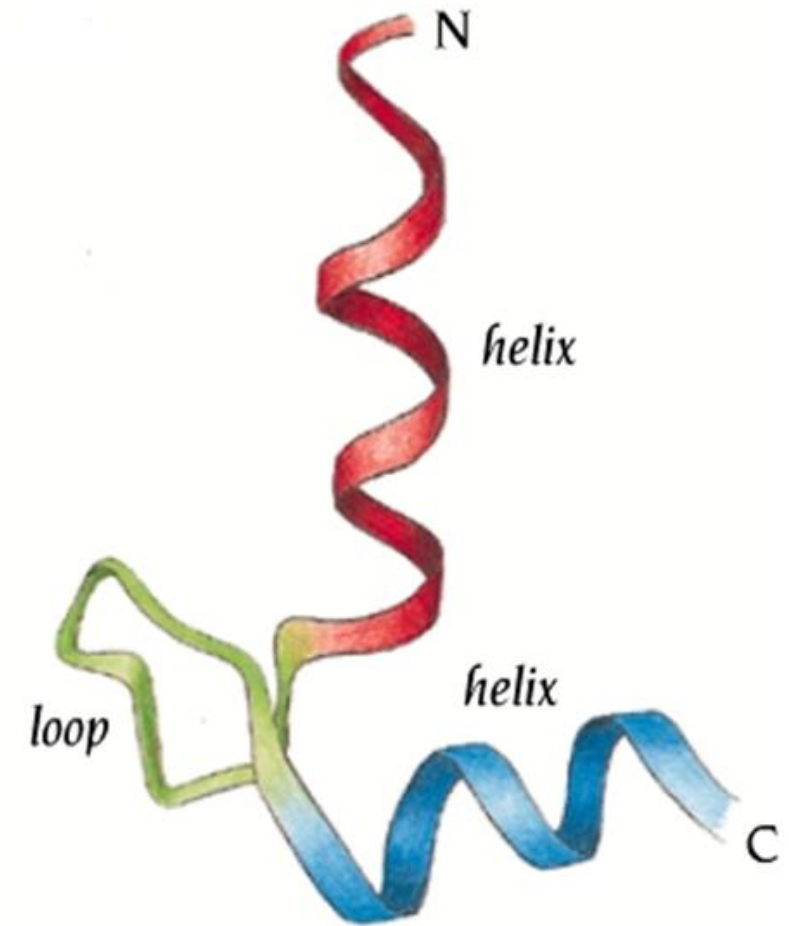
Q Gln, Glutamine



W Trp, Tryptophan

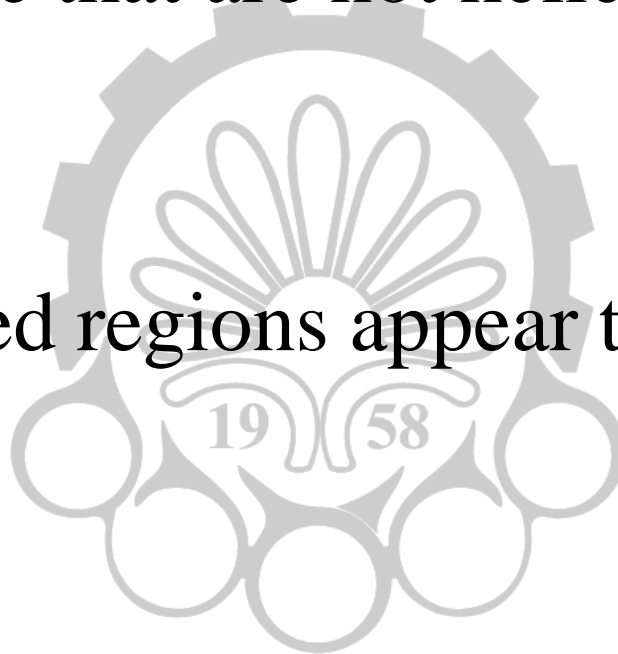
Loops

- Connect helices and sheets
- Vary in length and 3-D configurations
- Are located on surface of structure
- Are more "tolerant" of mutations
- Are more flexible and can adopt multiple conformations
- Tend to have charged and polar amino acids
- Some fall into distinct structural families (e.g., hairpin loops, reverse turns)



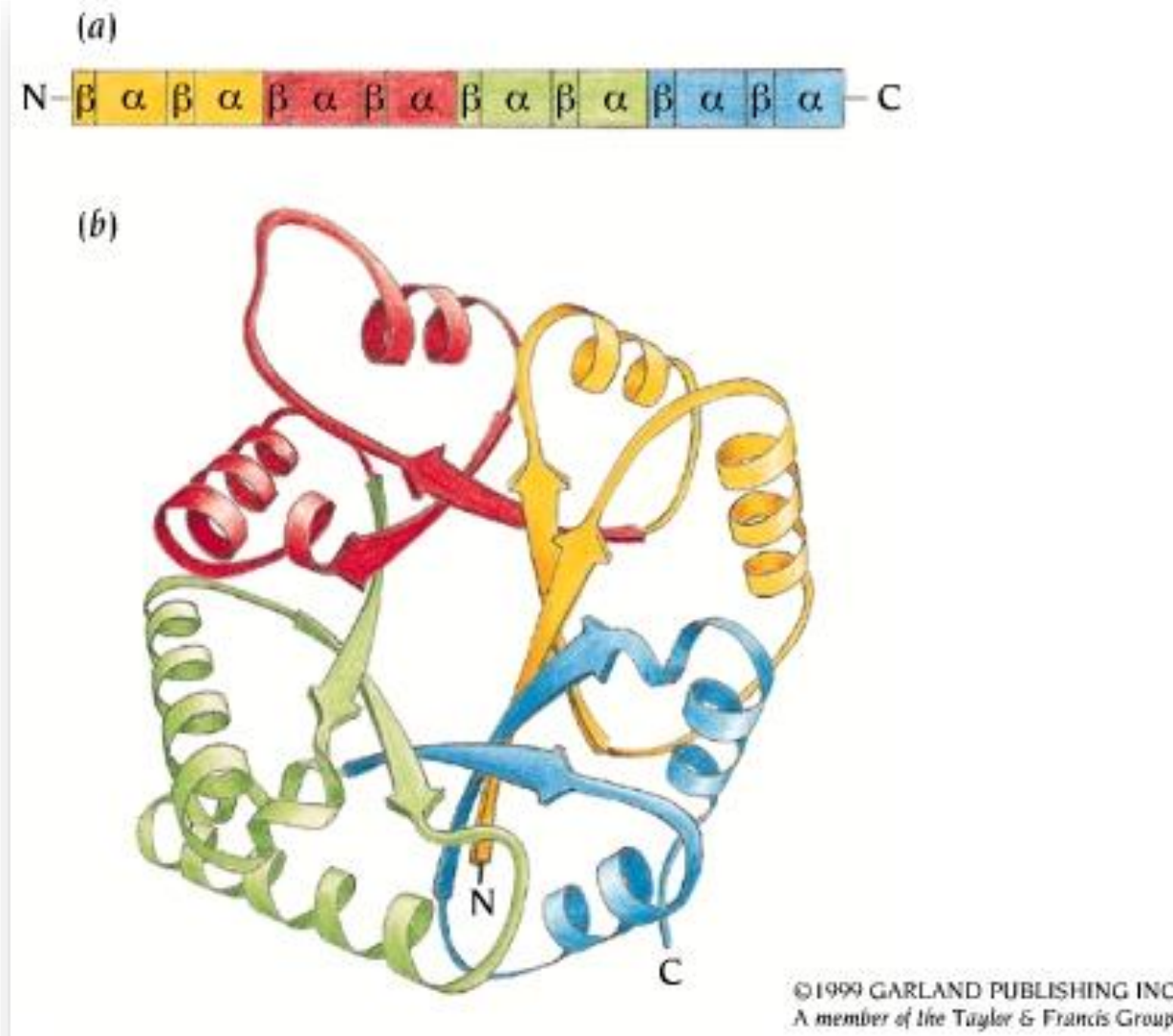
Coils

- Regions of 2' structure that are not helices, sheets, or recognizable turns
- Intrinsically disordered regions appear to play important functional roles



Amirkabir University of Technology
(Tehran Polytechnic)

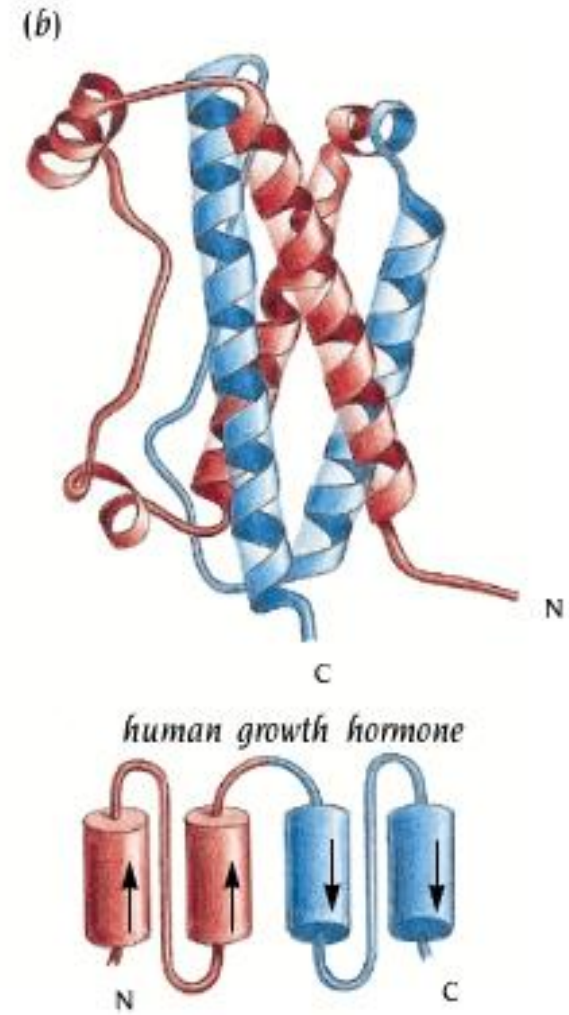
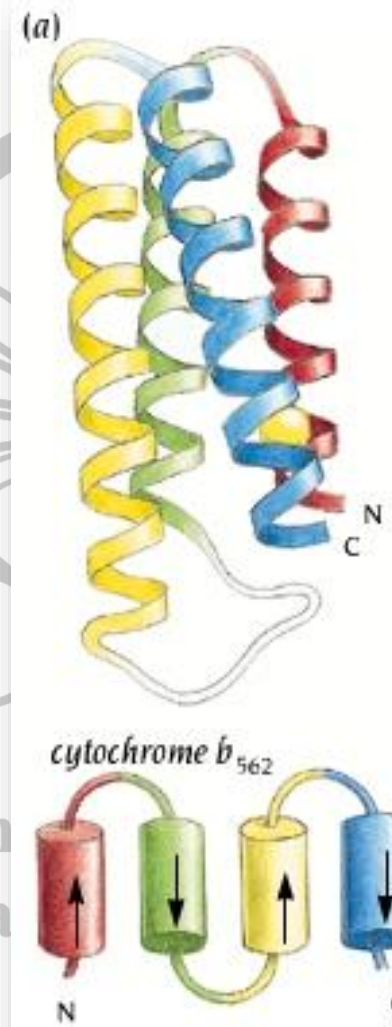
Simple Motifs Combine to Form Domains



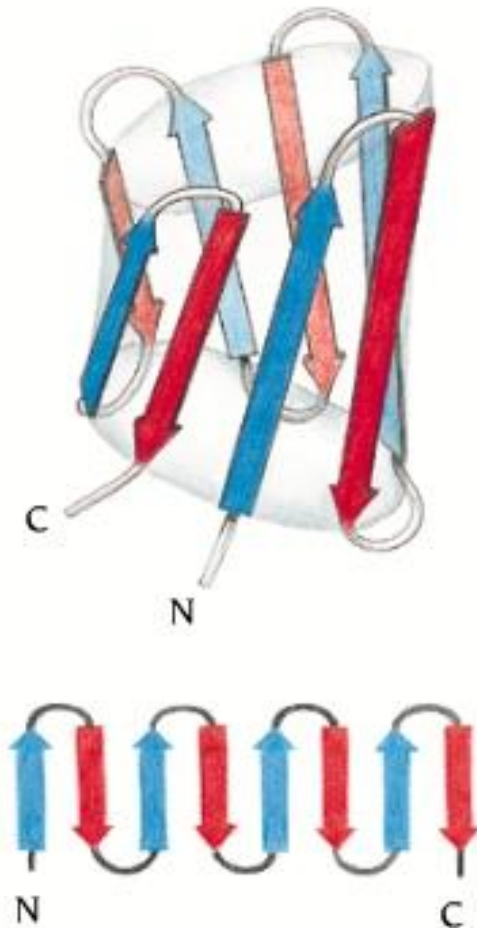
6 Main Classes of Protein Structure

1. α -Domains
 - Bundles of helices connected by loops
2. β -Domains
 - Mainly antiparallel sheets, usually with 2 sheets forming sandwich
3. α/β Domains
 - Mainly parallel sheets with intervening helices, also mixed sheets
4. $\alpha + \beta$ Domains
 - Mainly segregated helices and sheets
5. Multi-domain ($\alpha + \beta$)
 - Containing domains from more than one class
6. Membrane & cell-surface proteins

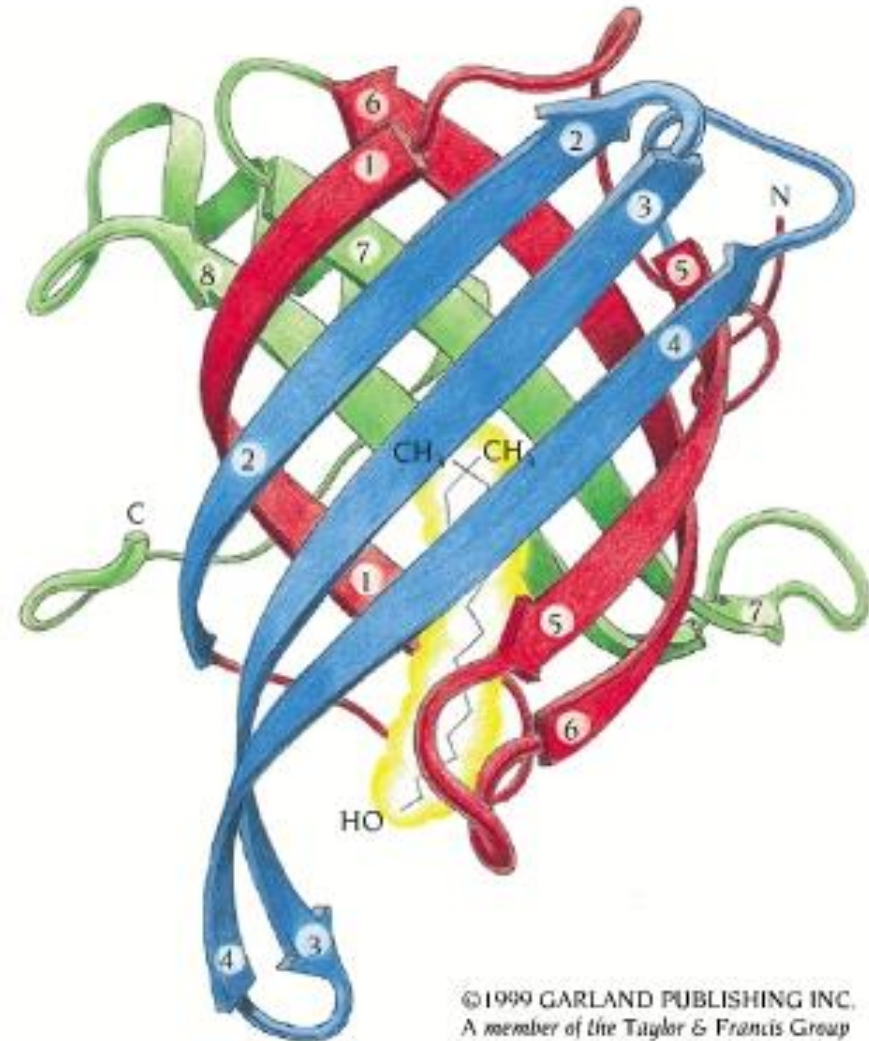
α -Domain Structures: 4-Helix Bundles



β -Sheets: Up-and-Down Sheets & Barrels

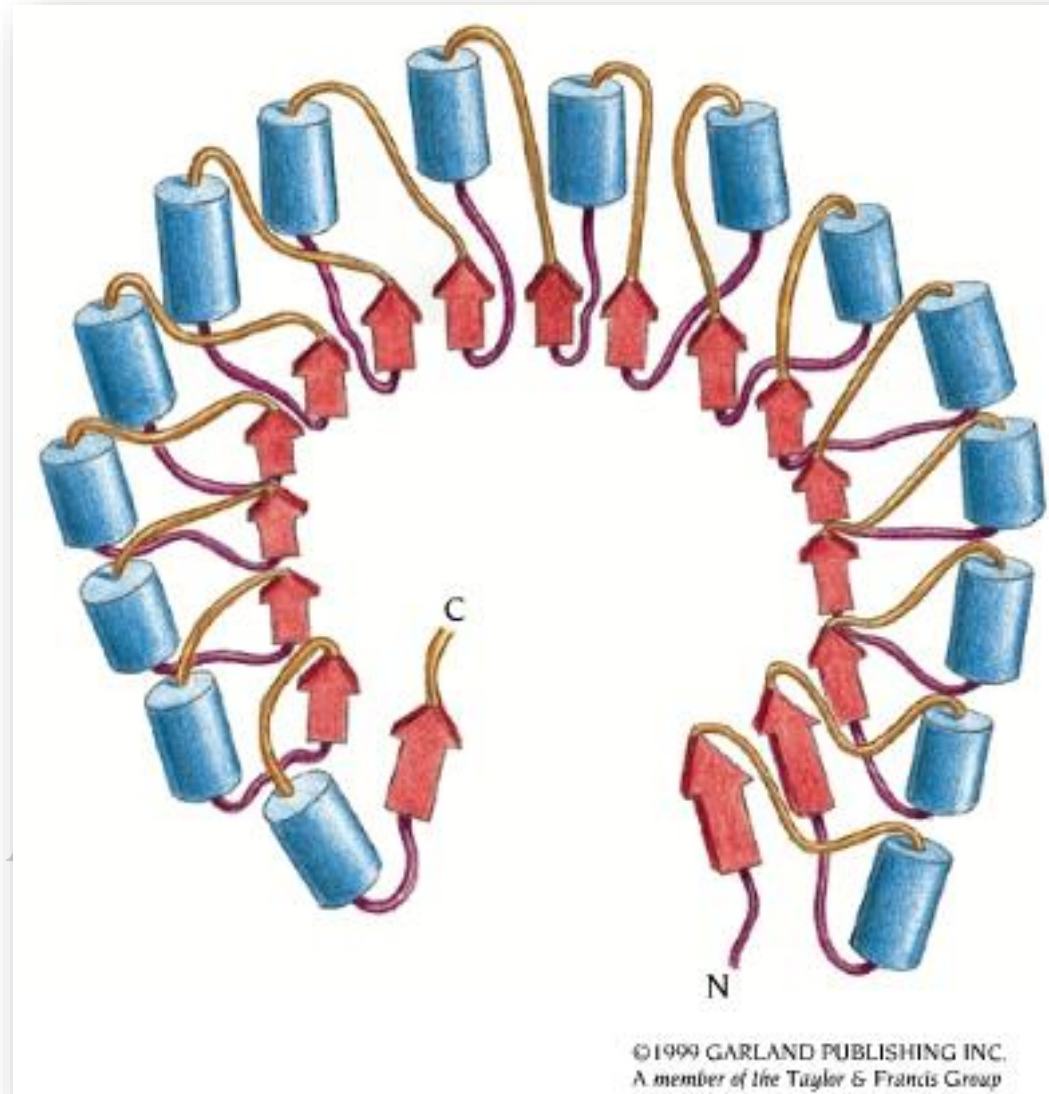


©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group



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

α/β Domains



Protein Structure Databases

- **PDB** - Protein Data Bank
 - <http://www.rcsb.org/pdb/>
 - (RCSB) - THE protein structure database
- **MMDB** - Molecular Modeling Database
 - <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Structure>
 - (NCBI Entrez) - has "added" value
- **MSD** - Molecular Structure Database
 - <http://www.ebi.ac.uk/msd>
 - Especially good for interactions & binding sites

PDB (RCSB)

RCSB Protein Data Bank

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

GDCB Webmail ISU BCB Dobbs Lab BCB 444/544 Schedule of Classes Entrez PSB 2008 Conferences Seminars FadiWiki PPIDB Google Image

RCSB PDB
PROTEIN DATA BANK

A MEMBER OF THE **wwPDB**

An Information Portal to Biological Macromolecular Structures

As of Tuesday Oct 02, 2007 there are 46377 Structures | PDB Statistics

CONTACT US | HELP | PRINT PAGE

PDB ID or keyword Author Site Search Advanced Search

Home Search

- Home
- Getting Started
 - Download Files
 - Deposit and Validate
 - Structural Genomics
 - Dictionaries & File Formats
 - Software Tools
 - General Education
 - Site Tutorials
- BioSync
 - General Information
 - Acknowledgements
 - Frequently Asked Questions
 - Report Bugs/Comments

Quick Tips: Click [here](#) for classroom activities using PDB data and tools.

Are you missing data updates? The PDB archive has moved to <ftp://ftp.wwpdb.org>. For more information click [here](#).

Welcome to the RCSB PDB

The **RCSB** PDB provides a variety of tools and resources for studying the structures of biological macromolecules and their relationships to sequence, function, and disease.

The RCSB is a member of the **wwPDB** whose mission is to ensure that the PDB archive remains an international resource with uniform data.


This site offers tools for browsing, searching, and reporting that utilize the data resulting from ongoing efforts to create a more consistent and comprehensive archive.

Information about compatible browsers can be found [here](#).

A **narrated tutorial** illustrates how to search, navigate, browse, generate reports and visualize structures using this new site. [This requires the Macromedia [Flash player download](#).]

Comments? info@rcsb.org

Molecule of the Month: Superoxide Dismutase



We can't live without oxygen. Our cells rely on oxygen as the final acceptor of electrons in respiration, allowing us to extract far more energy from food than would be possible without oxygen. But oxygen is also a dangerous compound. Reactive forms of oxygen, such as superoxide (oxygen with an extra electron), leak from the respiratory enzymes and wreak havoc on the cell. This superoxide can then cause mutations in DNA or attack enzymes that make amino acids and other essential molecules. This is a significant problem: one study showed that for every 10,000 electrons transferred down the respiratory pathway in *Escherichia coli* cells, about 3 electrons end up on superoxide instead of the proper place. To combat this potential danger, most cells make superoxide dismutase (SOD, shown here from PDB entry **2sod**), an enzyme that detoxifies superoxide.

- [More ...](#)
- [Previous Features](#)

News

- [Complete News](#)
- [Newsletter](#)
- [Discussion Forum](#)

2-October-2007

Structure Deposition Checklist

It is recommended to have the following items on hand when depositing a structure:

- Contact authors name (including PI), e-mail address, postal address, phone and fax numbers
- Title of the deposited structure and any relevant keywords
- Citation information: author names, title, and journal details if these are available
- Macromolecule names
- Biological assembly information
- Ligand names and chemical diagrams
- Sequence and chain ID for each macromolecule, including his tags or cloning artifacts that were not cleaved and any residues not visible due to disorder
- Source information: scientific names for source organisms, expression systems, or details about synthetically produced

Structure at NCBI

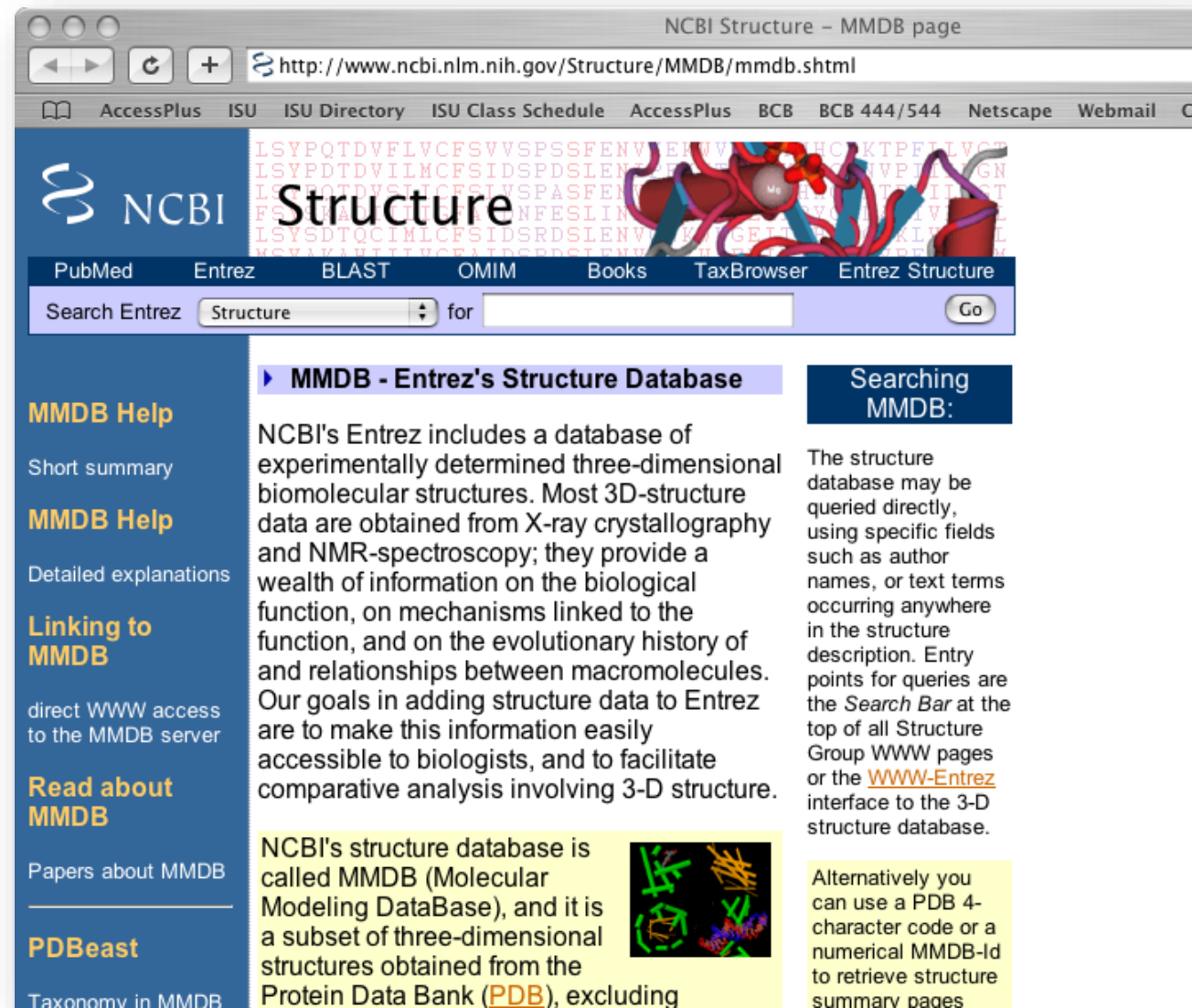
The screenshot shows the NCBI Structure Group website in a Netscape browser window. The address bar displays <http://www.ncbi.nlm.nih.gov/Structure/>. The page features a navigation bar with links to PubMed, Entrez, BLAST, OMIM, Books, TaxBrowser, and Entrez Structure. A search bar labeled "Search Entrez" is set to "Structure" with a "Go" button. The main content area is divided into several sections:

- What's New?**
 - MMDB**: NCBI's structure database
 - Cn3D v4.1**: 3D-structure viewer
 - CDD**: Conserved Domain Database
 - VAST**: Structure comparisons
 - VAST Search**: Submit structure database searches
 - PDBeast**: Taxonomy in MMDB
 - Research**
- The NCBI Structure Group**: .. maintains MMDB, a database of macromolecular 3D structures, as well as tools for their visualization and comparative analysis. MMDB, the Molecular Modeling Database, contains experimentally determined biopolymer structures obtained from the Protein Data Bank ([PDB](#)).
- Structure highlights**:
 - Conserved Domain Database**: A collection of sequence alignments and profiles, representing protein domains conserved in molecular evolution, is now available via the CD-search service. More ..
- try also:**
 - Structure summary via PDB/MMDB-Id:** [input field] [Go]
 - Read more about:** [MMDB](#), [WWW-Entrez](#), [VAST](#)
 - Resources:**
 - [MMDB's FTP-site](#) (including the MMDB database)
 - [NCBI C++ Toolkit](#) containing Cn3D source code.
 - [Publications](#) in Entrez
 - Research software:** [PKB and Threading](#) (requires Splus).

Text searches in MMDB (use the search toolbar at the top of this page) will yield Structure Query pages, providing access to entries that matched the keywords. Structure Summary pages for one/several/all of these may be retrieved. From the Structure Summary pages one may:

- access amino acid and nucleic acid sequences
- retrieve PubMed documents
- get taxonomy information
- view sequence neighbors
- view structure neighbors

MMDB at NCBI



The screenshot shows a Netscape browser window titled "NCBI Structure - MMDB page". The address bar displays the URL <http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml>. The browser's menu bar includes "AccessPlus", "ISU", "ISU Directory", "ISU Class Schedule", "AccessPlus", "BCB", "BCB 444/544", "Netscape", "Webmail", and "Ci".

The main content area features the NCBI logo and the word "Structure" in a large font. Below this is a navigation bar with links to "PubMed", "Entrez", "BLAST", "OMIM", "Books", "TaxBrowser", and "Entrez Structure". A search bar labeled "Search Entrez" contains the text "Structure" and a "Go" button.

The page is divided into several sections:

- MMDB Help**: Includes links for "Short summary" and "Detailed explanations".
- Linking to MMDB**: Provides "direct WWW access to the MMDB server".
- Read about MMDB**: Includes a link for "Papers about MMDB".
- PDBeast**: Includes a link for "Taxonomy in MMDB".

The central text area, titled "MMDB - Entrez's Structure Database", describes the database as containing experimentally determined three-dimensional biomolecular structures. It states that most 3D-structure data are obtained from X-ray crystallography and NMR-spectroscopy, and provides a wealth of information on biological function, mechanisms, and evolutionary history. The goals are to make this information easily accessible to biologists and to facilitate comparative analysis involving 3-D structure.

To the right of this text is a section titled "Searching MMDB:", which explains that the structure database may be queried directly using specific fields such as author names or text terms occurring anywhere in the structure description. Entry points for queries are the "Search Bar" at the top of all Structure Group WWW pages or the [WWW-Entrez](#) interface to the 3-D structure database.

At the bottom right, a yellow box notes that alternatively, one can use a PDB 4-character code or a numerical MMDB-Id to retrieve structure summary pages.

Decorative elements include a 3D ribbon diagram of a protein structure in the top right and a small graphic of a virus-like particle in the bottom right.

MSD: Molecular Structure Database

The screenshot shows the Macromolecular Structure Database (MSD) website. The browser window title is "Macromolecular Structure Database Group" and the address bar shows "http://www.ebi.ac.uk/msd/". The page features a navigation bar with links to GDCB, Webmail, ISU, BCB, Dobbs Lab, BCB 444/544, Schedule of Classes, Entrez, PSB 2008, Conferences, Seminars, and FadiWi. Below this is a search bar with "All Databases" selected and a "Enter Text Here" input field. A sidebar on the left contains links to MSD Home, Services, Resources, Documentation, Submission, FTP/PDB Access, Contact MSD, and Latest Annual Report. The main content area has a breadcrumb trail "EBI > Databases > Structure Databases > MSD" and a "contact msd" link. A "Welcome" section follows, with a paragraph stating: "Welcome to the EBI Macromolecular Structure Database - the European project for the collection, management and distribution of data about macromolecular structures, derived in part from the Protein Data Bank (PDB)." To the right of this text is the PDB logo. Below the welcome message is a grid of links categorized under "Submission", "Documentation", "Resources", and "Services".

Macromolecular Structure Database Group

http://www.ebi.ac.uk/msd/

EMBL-EBI EB-eye Search All Databases Enter Text Here

Macromolecular Structure Database Group

EBI > Databases > Structure Databases > MSD contact msd

Welcome

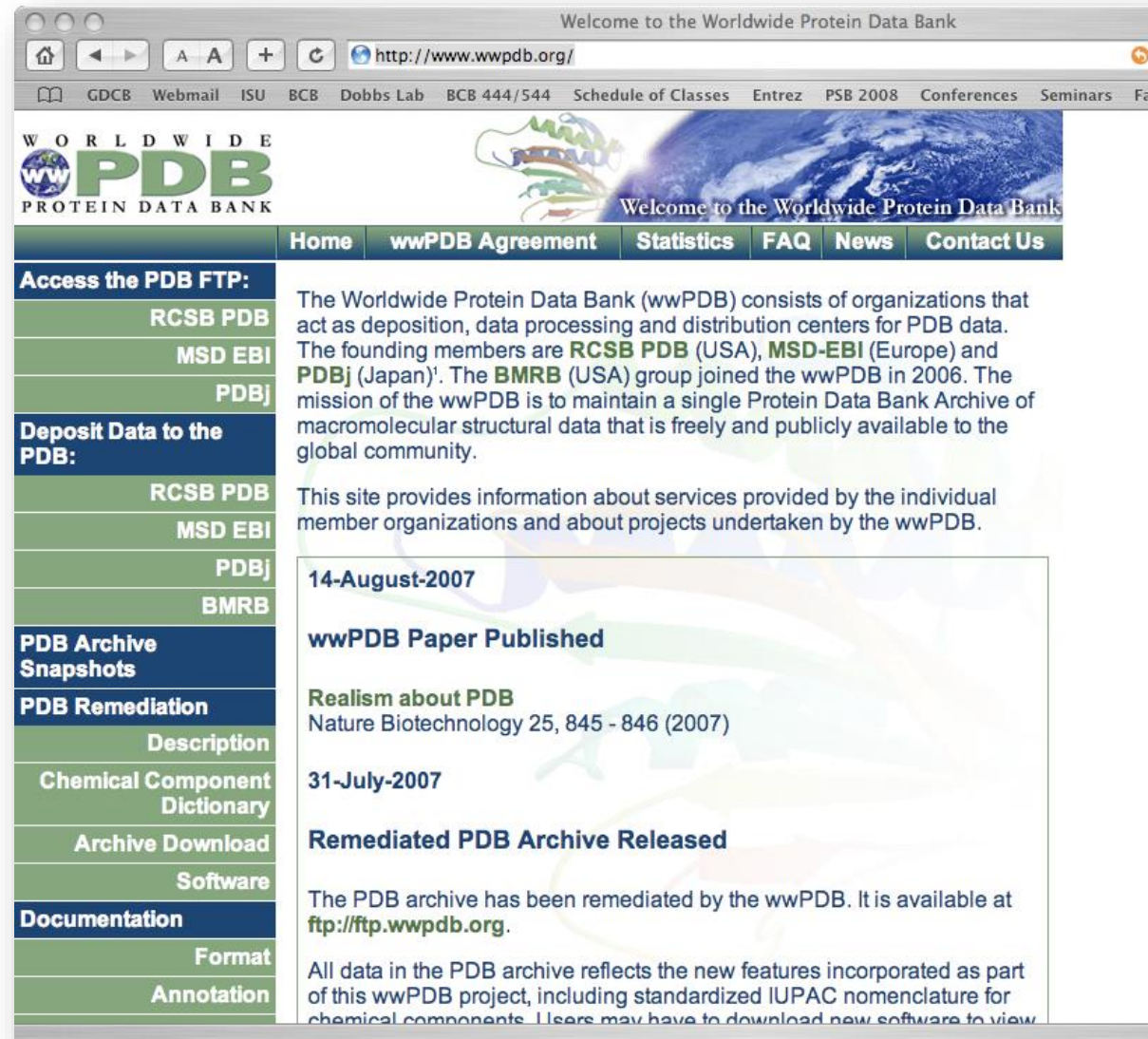
Welcome to the EBI Macromolecular Structure Database - the European project for the collection, management and distribution of data about macromolecular structures, derived in part from the Protein Data Bank (PDB).

WORLDWIDE PDB PROTEIN DATA BANK

Submission Documentation Resources Services

- PDB AutoDep
- MSDSD guide
- SIFTS @ EBI
- MSDlite
- EMDep
- MSDSD schema
- SPINE @ EBI
- MSDpro
- Software
- RECOORD @ EBI
- MSDmotif
- API
- NMR @ EBI
- MSDtemplate
- LIMS
- Funding
- MSDpisa
- mmCIF
- Education
- MSDchem
- XML
- Projects
- MSDmine
- PDB
- CAPRI
- MSDsite
- DB_io_tools
- Publications
- MSDfold
- 3DEM
- About MSD
- MSDtarget
- Conventions
- Past Workshops
- MSDanalysis
- Future Events
- EMsearch
- MSDbar

wwPDB: World Wide PDB



The screenshot shows the wwPDB website in a web browser. The browser's address bar displays <http://www.wwpdb.org/>. The website's header includes the "WORLDWIDE PDB PROTEIN DATA BANK" logo and a navigation menu with links: Home, wwPDB Agreement, Statistics, FAQ, News, and Contact Us. A sidebar on the left contains links for "Access the PDB FTP:" (RCSB PDB, MSD EBI, PDBj), "Deposit Data to the PDB:" (RCSB PDB, MSD EBI, PDBj, BMRB), "PDB Archive Snapshots", "PDB Remediation" (Description, Chemical Component Dictionary, Archive Download, Software), "Documentation" (Format, Annotation), and "PDBj". The main content area features a welcome message, a paragraph about the wwPDB's mission, and a news section with two entries: "14-August-2007 wwPDB Paper Published" and "31-July-2007 Remediated PDB Archive Released".

Welcome to the Worldwide Protein Data Bank

[http://www.wwpdb.org/](#)

GDCB Webmail ISU BCB Dobbs Lab BCB 444/544 Schedule of Classes Entrez PSB 2008 Conferences Seminars Fa

WORLDWIDE PDB PROTEIN DATA BANK

Welcome to the Worldwide Protein Data Bank

[Home](#) [wwPDB Agreement](#) [Statistics](#) [FAQ](#) [News](#) [Contact Us](#)

Access the PDB FTP:

- [RCSB PDB](#)
- [MSD EBI](#)
- [PDBj](#)

Deposit Data to the PDB:

- [RCSB PDB](#)
- [MSD EBI](#)
- [PDBj](#)
- [BMRB](#)

PDB Archive Snapshots

PDB Remediation

- [Description](#)
- [Chemical Component Dictionary](#)
- [Archive Download](#)
- [Software](#)

Documentation

- [Format](#)
- [Annotation](#)

The Worldwide Protein Data Bank (wwPDB) consists of organizations that act as deposition, data processing and distribution centers for PDB data. The founding members are **RCSB PDB** (USA), **MSD-EBI** (Europe) and **PDBj** (Japan)¹. The **BMRB** (USA) group joined the wwPDB in 2006. The mission of the wwPDB is to maintain a single Protein Data Bank Archive of macromolecular structural data that is freely and publicly available to the global community.

This site provides information about services provided by the individual member organizations and about projects undertaken by the wwPDB.

14-August-2007

wwPDB Paper Published

Realism about PDB
Nature Biotechnology 25, 845 - 846 (2007)

31-July-2007

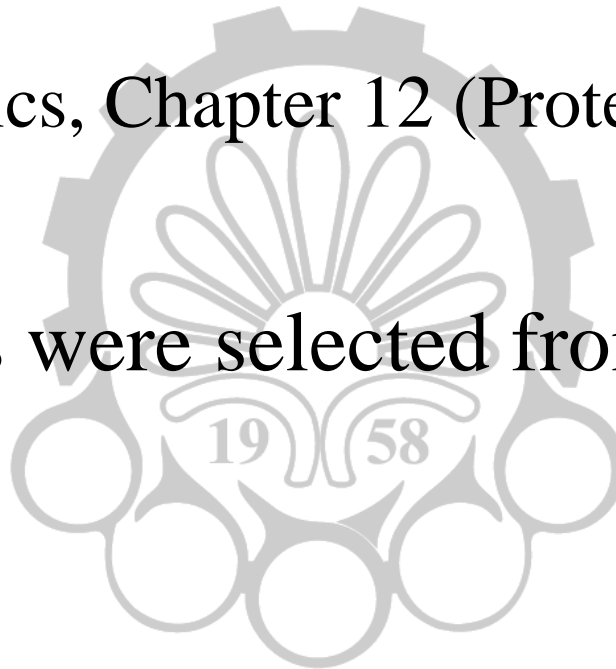
Remediated PDB Archive Released

The PDB archive has been remediated by the wwPDB. It is available at <ftp://ftp.wwpdb.org>.

All data in the PDB archive reflects the new features incorporated as part of this wwPDB project, including standardized IUPAC nomenclature for chemical components. Users may have to download new software to view

References

- Mostly used:
 - Essential bioinformatics, Chapter 12 (Protein Structure Basics)
- IP notice: some slides were selected from Drena Dobbs' slides.



Amirkabir University of Technology
(Tehran Polytechnic)

Thanks for your attention

