STRUCTURAL BIOLOGY

INTRODUCTION

Assist. Prof. Dr. Betul Akcesme
Structural Biology (BIO304)

Monday-Wednesday 10:30-11:45
Classroom: A F1.10

mobile: 061 767 262
e-mail: bakcesme@ius.edu.ba
BOOKS

Protein Structure and Function: From Sequence to Consequence

Gregory A. Petsko and Dagmar Ringe, Brandeis University

Published by New Science Press Ltd and distributed in the United States and Canada by Sinauer Associates, Inc.

AND

Introduction to Protein Structure

Carl Branden & John Tooze
Extra sources


- The Structures of Life-National Institutes of Health (Booklet)

- Lehninger Principles of Biochemistry
Project and Assignments!

- Swiss PDB viewer - visualization of protein 3D structure
Reminders!

• Attendance of lectures are MANDATORY!
  ▫ Up to 30% absence is tolerated! (9 sections out of 28)

• Submission of assignments on time!

• Copy-Past is strictly forbidden for assignments and projects!!
### Course Objectives
The course aims to provide the knowledge of the basic protein molecular architecture and then to connect it to the basic cellular processes like enzymatic activity, transport, and membrane functions.

### Textbook
- **Protein structure and Function**, G. Petsko and D. Ringe, Oxford University Press, 2009

### Learning Outcomes
- **1** Describe the aspects of protein structure and function including protein folding, molecular interactions and recognition.
- **2** Recognize the active sites of proteins in order to correlate structure and function.
- **3** Apply structural knowledge of proteins to the SWISS PDB VIWER to visualize protein structure.
- **4** Analyze a particular protein structure and function through finding, reading, and interpreting relevant scientific literature.

### Teaching Methods
Lecture presentations and class discussions. Visual materials are provided to better explain the concept. Tool, SWISS PDB viewer, is used to visualize protein structures and students are asked to do assignments based on the tool. Project and student presentations.
<table>
<thead>
<tr>
<th>WEEK</th>
<th>TOPIC</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Introduction to Structural Biology Basic structural principles of proteins</td>
<td>Branden Chapter 1</td>
</tr>
<tr>
<td>Week 2</td>
<td>Motifs of protein structure</td>
<td>Branden chapter 2</td>
</tr>
<tr>
<td>Week 3</td>
<td>Alpha domain structures\ Alpha&amp;Beta domain structures</td>
<td>Branden chapter 3-4</td>
</tr>
<tr>
<td>Week 4</td>
<td>Beta domain structures</td>
<td>QUIZ I (WED)</td>
</tr>
<tr>
<td>Week 5</td>
<td>Folding and Flexibility</td>
<td>Branden 5</td>
</tr>
<tr>
<td>Week 6</td>
<td>Folding and Flexibility</td>
<td>SWISS PDB VIWER PRACTICE</td>
</tr>
<tr>
<td>Week 7</td>
<td>SWISS PDB VIWER PRACTICE</td>
<td>Branden 6</td>
</tr>
<tr>
<td>Week 8</td>
<td>DNA structure and DNA recognition in prokaryotes</td>
<td>MIDTERM EXAM (WED)</td>
</tr>
<tr>
<td>Week 9</td>
<td>DNA recognition by eukaryotic transcription factors</td>
<td>Assignment submission of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SWISS PDB VIWER</td>
</tr>
<tr>
<td>Week 10</td>
<td>Membrane Proteins</td>
<td>Branden Chapter 12</td>
</tr>
<tr>
<td>Week 11</td>
<td>Recognition of Foreign Molecules by the Immune system</td>
<td>QUIZ II (WED)</td>
</tr>
<tr>
<td>Week 12</td>
<td>Prediction, Engineering, and Design of Protein Structures</td>
<td>Branden Chapter 15</td>
</tr>
<tr>
<td>Week 13</td>
<td>Determination of Protein structure (NMR and X ray)</td>
<td>Branden Chapter 17</td>
</tr>
<tr>
<td>Week 14</td>
<td>Final Review</td>
<td></td>
</tr>
<tr>
<td>Assessment Methods and Criteria</td>
<td>Evaluation Tool</td>
<td>Quantity</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>Final Exam</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Semester Evaluation Components</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quizzes</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>In-term exam</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Project and presentation(Weekly)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Assignments</td>
<td>2</td>
</tr>
</tbody>
</table>
Why Structure?

"Structures of life may hold the key to developing new medicines, materials, and diagnostic procedures."
“If biology were a car, structural biologists would be looking under the bonnet to find out how the engine works.”

“Put more prosaically, structural biology aims to understand how biology works at the molecular level.”

Ad Bax & Dennis A. Torchia, Nature (Feb. 8th 2007)
Traditional Architecture

Molecular Architecture

Form fits function

Materials
Wood, brick, nails, glass
Amino acids, cofactors

Environmental Factors
Temperature, earthquakes
Temperature, solubility

Population Factors
How many people?
# partner proteins, # reactants

Portals
How many doors and windows?
Passages for substrates and reactants

Motifs/Styles
Spanish, Victorian, 1950's blocky science building
Conserved domains or protein folds

Architects
Julia Morgan, 1950's blocky science building
Evolution
WHAT IS STRUCTURAL BIOLOGY?

**Structural biology** is a branch of molecular biology, biochemistry, and biophysics concerned with the molecular structure of biological macromolecules,

- how they acquire the structures they have,
- how alterations in their structures affect their function.
• Structural biology requires the cooperation of many different scientists, including biochemists, molecular biologists, X-ray crystallographers, and NMR spectroscopists.

• Although these researchers use different techniques and may focus on different molecules, they are united by their desire to better understand biology by studying the detailed structure of biological molecules.
bioinformatics
protein dynamics
protein folding
molecular modeling
computational biology
drug design
Structural Biology
What is the goal of structural biology?

- world-wide initiative aiming at the high-throughput determination (by X-ray or NMR) of protein structures;
- classification of protein in fold families;
- determination of at least one member of each family;
- development of high-throughput techniques (cloning, expression, crystallization, structure solution and refinement);
- development of modeling algorithms (analysis of known structures);
- **long-term goal: prediction and modeling of all protein structures**
HISTORY of STRUCTURAL BIOLOGY

- Very nice timeline!

WHY STUDY THE PROTEINS?

• Structural biologists are mostly interested in proteins, because these molecules do most of the work in the body.

• By studying the structures of proteins, we are better able to understand
  - how they function normally
  - how some proteins with abnormal shapes can cause disease.
A protein called alpha-keratin forms your hair and fingernails, and also is the major component of feathers, wool, claws, scales, horns, and hooves.

The hemoglobin protein carries oxygen in your blood to every part of your body.

Muscle proteins called actin and myosin enable all muscular movement—from blinking to breathing to rollerblading.

Ion channel proteins control brain signaling by allowing small molecules into and out of nerve cells.

Receptor proteins stud the outside of your cells and transmit signals to partner proteins on the inside of the cells.

Enzymes in your saliva, stomach, and small intestine are proteins that help you digest food.

Antibodies are proteins that help defend your body against foreign invaders, such as bacteria and viruses.

Huge clusters of proteins form molecular machines that do your cells’ heavy work, such as copying genes during cell division and making new proteins.
GENOMICS, TRANSCRIPTOMICS PROTEOMICS!

Proteome Complexity

Genome
~20-25,000 genes

Transcriptome
~100,000 transcripts

Proteome
>1,000,000 proteins

Alternative promoters
Alternative splicing
mRNA editing

Post-translational modifications
SCOP and CATH

• SCOP and CATH are the two databases generally accepted as the two main authorities in the world of fold classification.

• Structural Classification of Proteins (SCOP)

• CATH

• http://www.pdb.org/pdb/statistics/contentGrowthChart.do?content=fold-cath
C Class
A Architecture
T Topology (Faltung)
H Homology
-------------------
S Sequence Family (Funktion)

http://www.cathdb.info/
http://www.rcsb.org/pdb/statistics/holdings.do
WHY STUDY to PROTEINS?

- Underpin every aspect of biological activities

- Advances in molecular genetics reveal that many diseases stem from specific protein defects.

  - **Cystic fibrosis (CFTR protein)**
    - to allow chloride ions (a component of table salt) to pass through the outer membranes of cells.
    - 65% case because of deletion aa residue position 508
    - incorrectly folded

  - **Sickle cell anemia**
    - Change of the 6th aa residue from glutamic acid to a valine
• Two most important factors!
  ▫ **CANCER!**
    • P53: ‘switches on’ in response to cellular damage and as a transcription factor controls the cell cycle process.
    • lung, colorectal and skin carcinomas are attributed to molecular defects in p53.
  ▫ **HIV!**
    • In 2003 the World Health Organization (WHO) estimated that over 40 million individuals are infected with this virus in the world today.
    • Affects immune system
    • understanding the structure of HIV proteins and in designing specific inhibitors.
• Others!
  ▫ **Antibody, insulin, drug proteins...**
As a result...

- We will need to understand
  - the structure of proteins,
  - their interaction with other biomolecules,
  - their roles within different biological systems
  - their potential manipulation by genetic or chemical method
The biological diversity of proteins
A selective list of some functional roles for proteins within cells

<table>
<thead>
<tr>
<th>Function</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzymes or catalytic proteins</td>
<td>Trypsin, DNA polymerases and ligases,</td>
</tr>
<tr>
<td>Contractile proteins</td>
<td>Actin, myosin, tubulin, dynein,</td>
</tr>
<tr>
<td>Structural or cytoskeletal proteins</td>
<td>Tropocollagen, keratin,</td>
</tr>
<tr>
<td>Transport proteins</td>
<td>Haemoglobin, myoglobin, serum albumin, ceruloplasmin, transthyretin</td>
</tr>
<tr>
<td>Effector proteins</td>
<td>Insulin, epidermal growth factor, thyroid stimulating hormone,</td>
</tr>
<tr>
<td>Defence proteins</td>
<td>Ricin, immunoglobulins, venoms and toxins, thrombin,</td>
</tr>
<tr>
<td>Electron transfer proteins</td>
<td>Cytochrome oxidase, bacterial photosynthetic reaction centre, plastocyanin, ferredoxin</td>
</tr>
<tr>
<td>Receptors</td>
<td>CD4, acetycholine receptor,</td>
</tr>
<tr>
<td>Repressor proteins</td>
<td>Jun, Fos, Cro,</td>
</tr>
<tr>
<td>Chaperones (accessory folding proteins)</td>
<td>GroEL, DnaK</td>
</tr>
<tr>
<td>Storage proteins</td>
<td>Ferritin, gliadin,</td>
</tr>
</tbody>
</table>
Four examples of biochemical functions performed by proteins

**Binding**

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

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

**Catalysis**

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

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

The GDP-bound (*off*; PDB 1p1l) state of Ras differs significantly from the GTP-bound (*on*; PDB 121p) state. This difference causes the two states to be recognized by different proteins in signal transduction pathways.

Structural proteins

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

Actin fibers are important for muscle contraction and for the cytoskeleton. They are helical assemblies of actin and actin-associated proteins. (Courtesy of Ken Holmes)
Where do extraordinary functional diversity and versatility of proteins come from?

- the chemical diversity of the side chains of their constituent amino acids,
- the flexibility of the polypeptide chain,
- the very large number of ways in which polypeptide chains with different amino acid sequences can fold.
There are four levels of protein structure

(a) Primary

(b) Secondary

(c) Tertiary

(d) Quaternary

alpha helices
beta strands
There are four levels of protein structure
Amino acids

- Proteins are polymers (polypeptides) of 20 different amino acids joined by peptide bonds.
- The chemical characters of the amino-acid side chains have important consequences for the way they participate in the folding and functions of proteins.

Protein stability and function

- Amino acids have properties that are well-suited to carry out a variety of biological functions
  - Capacity to polymerize
  - Useful acid-base properties
  - Varied physical properties
  - Varied chemical functionality
Central carbon atom $C_\alpha$ is attached to amino group, carboxyl, H atom and side chain.

Amino acids share many features, differing only at the R substituent.
Most amino acids are chiral

- The carbon alpha always has four substituents and is tetrahedral

- All (except proline) have:
  - an acidic carboxyl group
  - a basic amino group
  - a hydrogen connected to the carbon alpha

- The fourth substituent (R) is unique
  - In glycine, the fourth substituent is also hydrogen
All amino acids are chiral (except glycine)

nonsuperposable mirror images of each other (enantiomers)
Proteins only contain L amino acids

- to a living system, D and L isomers are as different as the right hand and the left.

- The formation of stable, repeating substructures in proteins generally requires that their constituent amino acids be of one stereochemical series.

Cells are able to specifically synthesize the L isomers of amino acids because the active sites of enzymes are asymmetric, causing the reactions they catalyze to be stereo specific.
<table>
<thead>
<tr>
<th>WEEK</th>
<th>WEEK 3</th>
<th>WEEK 4</th>
<th>WEEK 5</th>
<th>WEEK 6</th>
<th>WEEK 7</th>
<th>WEEK 8</th>
<th>WEEK 9</th>
<th>WEEK 10</th>
<th>WEEK 11</th>
<th>WEEK 12</th>
<th>WEEK 13</th>
<th>WEEK 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Alpha Domain Structure: Hemoglobin structure, sick cell hemoglobin and its relation with malaria</td>
<td>Alpha Beta Domain Structure: Alpha/Beta Tertiary Structure their active site and importance in a particular disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>QUIZ I</td>
<td>Beta domain Structure: Beta-propeller Protein-Associated Neurodegeneration (BPAN)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>What is behind folding of proteins? <a href="http://fold.it/portal/info/about">http://fold.it/portal/info/about</a></td>
<td>Type II diabetes and protein misfolding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Alzheimer and Parkinson disease and protein misfolding</td>
<td>Chemical and pharmacological chaperones</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Swiss PDB: Protein visualization tools</td>
<td>Swiss PDB or any related subject</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>MIDTERM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>DNA recognition in Prokaryotes</td>
<td>DNA recognition in Eukaryotes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Membrane Proteins: Structure of BAM complex</td>
<td>Membrane Proteins: Amphitropic Proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>QUIZ II</td>
<td>Immunology: Immunological Databases and Tools (Computational Immunology) from structural biology perspective</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Prediction engineering Design of the structures: Drug Design and Therapy</td>
<td>Prediction engineering Design of the structures: Computational Methods in Drug Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Determination of protein structure, X-ray NMR</td>
<td>In silico approaches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Protein Engineering Methods Industrial Application of Protein Design Strategies of classification of proteins: SCOP AND CATH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Classification of amino acids

- Common amino acids can be placed in five basic groups depending on their R substituents:
  - Nonpolar, aliphatic (7)
  - Aromatic (3)
  - Polar, uncharged (5)
  - Positively charged (3)
  - Negatively charged (2)

OR

- Nonpolar, aliphatic (9)
- Polar, uncharged (6)
- Positively charged (3)
- Negatively charged (2)

their polarity, or tendency to interact with water at biological pH (near pH 7.0).
The polarity of the R groups varies widely, from nonpolar and hydrophobic (water-insoluble) to highly polar and hydrophilic (water-soluble).
Atom-naming conventions for aminoacids

- Amino acid side chain atoms are assigned Greek letter indices depending on how far removed they are from the carboxylate carbon. Beta (β) atoms are adjacent to the α-carbon, followed by gamma (γ), delta (δ), epsilon (ε), zeta (ζ) and eta (η).
- Identical atoms deriving from a common branch point are distinguished by an index (1 or 2).
NONPOLAR SIDE CHAINS

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

- Smallest amino acid: apolar, achiral, flexible
- Proline is a helix breaker
- Sulfur containing
Hydrophobic amino acids role in shaping proteins

Alanine, valine, isoleucine and leucine are strong helix-favoring residues, while proline is rarely found in helices because its backbone nitrogen is not available for the hydrogen bonding required for helix formation.
Glycine-Proline pairs

Pro-Gly-pairs introduce turns in peptide chains
Proline is a helix breaker

- Proline side chain interferes with the backbone helical packing of residues N-terminal to the proline as well as with the side chains of neighboring residues.
- The secondary amino group of proline residues is held in a rigid conformation that reduces the structural flexibility of polypeptide regions containing proline.
Sulfur containing
Disulfide bond formation

Tyrosine phosphorylated and dephosphorilated (switch on and of many cellular activities)
Aromatic and hydrophilic

These amino acids side chains can form hydrogen bonds.

Cysteine can form disulfide bonds.
Disulfide formation

The disulfide is usually the end product of air oxidation according to the following schematic reaction scheme:

\[ 2 \text{-CH}_2\text{SH} + \frac{1}{2} \text{O}_2 \rightarrow \text{CH}_2\text{-S-S-CH}_2 + \text{H}_2\text{O} \]

Disulfide bonds form between the side chains of two cysteine residues. Two SH groups from cysteine residues, which may be in different parts of the amino acid sequence but adjacent in the three-dimensional structure, are oxidized to form one S—S (disulfide) group.

Disulfides are important in stabilizing the folded 3D structure of a protein.
ELECTRICALLY CHARGED SIDE CHAINS

His coordinates metal atoms
It is often found in active sites of the enzyme
It is a member of CATALYTIC TRIAD, 3 aa are found inside the active site of certain protease enzymes: (serine, aspartate)
These amino acid side chains absorb UV light at 270–280 nm.
Uncommon Amino Acids in Proteins

- **Not incorporated** by ribosomes
  - except for Selenocysteine
  - This rare amino acid residue is introduced *during protein synthesis* rather than created through a postsynthetic modification. It contains selenium rather than the sulfur of cysteine.

- Arise by post-translational modifications of proteins

- The addition of phosphoryl, methyl, acetyl, adenylyl, ADPribosyl, or other groups to particular amino acid residues can increase or decrease a **protein’s activity**

- Reversible modifications, especially phosphorylation, are important in regulation and signaling
Nonstandard amino acids found in proteins: All are derived from standard amino acids. Extra functional groups added by modification reactions are shown in red.
Nonstandard amino acids: Ornithine and Citrulline, which are not found in proteins, are intermediates in the biosynthesis of arginine and in the urea cycle.
pH and amino acids

- pH impacts both **shape** and **enzymatic activity**
- the ionic state (charge) of an amino acid depends on the pH
- at low pH, there will be a **positive charge** (in acid there are extra H+ ions present)
- at high pH conditions, there will be a **negative charge** (in base, there will be a lack of H+ ions)
Ionization of Amino Acids

- **At acidic pH**, the carboxyl group is protonated and the amino acid is in the cationic form.

- **At neutral pH**, the carboxyl group is deprotonated but the amino group is protonated. The net charge is zero; such ions are called Zwitterions.

- **At alkaline pH**, the amino group is neutral –NH₂ and the amino acid is in the anionic form.
Isoelectronic point, pl

- The **isoelectronic point** or **isoionic point** is the pH at which the amino acid **does not migrate in an electric field**.
- This means it is the pH at which the amino acid is neutral, *i.e.* the **zwitterion** form is dominant.
- "Zwitterion" a dipolar molecule (positive and negative areas) but charges cancel out!
Some R-groups can be ionized

The Henderson-Hasselbalch equation allows calculation of the ratio of a weak acid and its conjugate base at any pH.

\[
\text{pH} = pK' - \log \left( \frac{[HB]}{[B^-]} \right) \left( \frac{\text{protonated}}{\text{unprotonated}} \right)
\]
How does this equation help us? (not really necessary for this course)

**dissociation constant,** $K_a$

\[
\text{H} \overset{K_a}{\rightleftharpoons} \text{A} \quad \text{H}^+ + \text{A}^- \\
K_a = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}]} \quad \text{p}K_a = -\log_{10} K_a
\]

• when the pH = pKa then $\log \frac{[\text{HA}]}{[\text{A}^-]} = 0$
  therefore $[\text{HA}] = [\text{A}^-]$ i.e. equal amounts of the two forms, the acid and the conjugate base.

• If we make the solution **more acidic**, i.e. lower the pH,
  so pH < pKa, then $\log \frac{[\text{HA}]}{[\text{A}^-]}$ has to be > 0 so $[\text{HA}] > [\text{A}^-]$.
  This makes sense as it tells us that a **stronger acid** will cause the formation of HA, the protonated form.

• If instead we make the solution **more basic**, i.e. raise the pH,
  so pH > pKa and $\log \frac{[\text{HA}]}{[\text{A}^-]}$ has to be < 0 so $[\text{HA}] < [\text{A}^-]$. This makes sense as it tells us that a **stronger base** will cause the formation of $\text{A}^-$, the deprotonated form.
### Table of pKa and pI values

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>pKa&lt;sub&gt;1&lt;/sub&gt;</th>
<th>pKa&lt;sub&gt;2&lt;/sub&gt;</th>
<th>pKa&lt;sub&gt;3&lt;/sub&gt;</th>
<th>pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>2.34</td>
<td>9.60</td>
<td>---</td>
<td>5.97</td>
</tr>
<tr>
<td>Alanine</td>
<td>2.34</td>
<td>9.69</td>
<td>---</td>
<td>6.00</td>
</tr>
<tr>
<td>Valine</td>
<td>2.32</td>
<td>9.62</td>
<td>---</td>
<td>5.96</td>
</tr>
<tr>
<td>Leucine</td>
<td>2.36</td>
<td>9.60</td>
<td>---</td>
<td>5.98</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>2.36</td>
<td>9.60</td>
<td>---</td>
<td>6.02</td>
</tr>
<tr>
<td>Methionine</td>
<td>2.28</td>
<td>9.21</td>
<td>---</td>
<td>5.74</td>
</tr>
<tr>
<td>Proline</td>
<td>1.99</td>
<td>10.60</td>
<td>---</td>
<td>6.30</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>1.83</td>
<td>9.13</td>
<td>---</td>
<td>5.48</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>2.83</td>
<td>9.39</td>
<td>---</td>
<td>5.89</td>
</tr>
<tr>
<td>Asparagine</td>
<td>2.02</td>
<td>8.80</td>
<td>---</td>
<td>5.41</td>
</tr>
<tr>
<td>Glutamine</td>
<td>2.17</td>
<td>9.13</td>
<td>---</td>
<td>5.65</td>
</tr>
<tr>
<td>Serine</td>
<td>2.21</td>
<td>9.15</td>
<td>---</td>
<td>5.68</td>
</tr>
<tr>
<td>Threonine</td>
<td>2.09</td>
<td>9.10</td>
<td>---</td>
<td>5.60</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>2.20</td>
<td>9.11</td>
<td>---</td>
<td>5.66</td>
</tr>
<tr>
<td>Cysteine</td>
<td>1.96</td>
<td>8.18</td>
<td>---</td>
<td>5.07</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>1.88</td>
<td>9.60</td>
<td>3.65</td>
<td>2.77</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>2.19</td>
<td>9.67</td>
<td>4.25</td>
<td>3.22</td>
</tr>
<tr>
<td>Lysine</td>
<td>2.18</td>
<td>8.95</td>
<td>10.53</td>
<td>9.74</td>
</tr>
<tr>
<td>Arginine</td>
<td>2.17</td>
<td>9.04</td>
<td>12.48</td>
<td>10.76</td>
</tr>
<tr>
<td>Histidine</td>
<td>1.82</td>
<td>9.17</td>
<td>6.00</td>
<td>7.59</td>
</tr>
</tbody>
</table>

The pKa values and the isoelectronic point, pI, are given below for the 20 a-amino acids.

- pKa<sub>1</sub> = **a-carboxyl group**,
- pKa<sub>2</sub> = **a-ammonium ion**, and
- pKa<sub>3</sub> = **side chain group**
**Neutral side chains:** These amino acids are characterized by two pKas: pKa1 and pKa2 for the carboxylic acid and the amine respectively.

\[ \text{pI} = \frac{1}{2} (\text{pKa}_1 + \text{pKa}_2) \]

**Acidic side chains:** The pI will be at a lower pH because the acidic side chain introduces an *extra* negative charge. So the neutral form exists under more acidic conditions when the extra (-) has been neutralized.

\[ \text{pI} = \frac{1}{2} (\text{pKa}_1 + \text{pKa}_3) \]

**Basic side chains:** The pI will be at a higher pH because the basic side chain introduces an *extra* positive charge. So the neutral form exists under more basic conditions when the extra (+) has been neutralized.

\[ \text{pI} = \frac{1}{2} (\text{pKa}_1 + \text{pKa}_3) \]
The organization of the genetic code reflects the chemical grouping of the amino acids

- physical-chemical properties

- single-nucleotide polymorphism in the third position in a codon usually produce same amino acid

- Single-base changes elsewhere in the codon will usually produce a different amino acid, but with the same physical-chemical properties

- conservative substitutions when they are found in comparisons of protein sequences they are taken to indicate conservation of structure between two proteins.

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

<table>
<thead>
<tr>
<th></th>
<th>Gly</th>
<th>Ala</th>
<th>Val</th>
<th>Leu</th>
<th>Ile</th>
<th>Met</th>
<th>Cys</th>
<th>Ser</th>
<th>Thr</th>
<th>Asn</th>
<th>Gln</th>
<th>Asp</th>
<th>Glu</th>
<th>Lys</th>
<th>Arg</th>
<th>His</th>
<th>Phe</th>
<th>Tyr</th>
<th>Trp</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>10</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>2</td>
<td>10</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>45</td>
<td>77</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>21</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>16</td>
<td>11</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>3</td>
<td>9</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>9</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>16</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>11</td>
<td>27</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td>5</td>
<td>35</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>27</td>
<td>7</td>
<td>3</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Glycine & alanine: smallest side chains.

Aspartic acid & glutamic acid (negatively charged)
Primary sequence reveals important clues about a protein

• Evolution conserves amino acids that are important to protein structure and function across species. Sequence comparison of multiple “homologous” of a particular protein reveals highly conserved regions that are important for function.

• Clusters of conserved residues are called “motifs” -- motifs carry out a particular function or form a particular structure that is important for the conserved protein.
Generally only a **limited amount** of a protein’s surface is well conserved

- **Invariant** (the residue is always the same, *e.g.* Asp)
- **Conserved** (the residue is generally similar, *e.g.* negatively charged)
- **Not conserved** (can be many different residues in different species)
Proteins are polypeptide chains

a-Amino acids are preferable to b-amino acids

Covalently joined by peptide bonds (hydrolysis reaction)

Proteins thus have a repeating backbone from which 20 different possible kinds of side chains protrude.
The peptide group is **planar** because the additional electron pair C=O bond delocalized over peptide group such that rotation around the C-N bound is prevented by an energy barrier.
The properties of the peptide bond have important effects on the stability and flexibility of polypeptide chains in water.

- The stability of the peptide bond is due to **resonance**, the delocalization of electrons over several atoms.

- **Electron delocalization** lowers the potential energy of the substance and thus makes it **more stable** than any of the contributing structures.
Each peptide bond is shown in a shaded box. Also shown are the individual dipole moments (arrows) associated with each bond. The dashed lines indicate the resonance of the peptide bond.

**Resonance** has two other important consequences.

**First**, it increases the polarity of the peptide bond: the *dipole moment of each peptide bond*

**Second**, the peptide bond has partial double-bond character, which means that the three non-hydrogen atoms that make up the bond (the carbonyl oxygen O, the carbonyl carbon C and the amide nitrogen N) are coplanar, **and that free rotation about the bond is limited**.
• The angle of the N–Ca bond to the adjacent peptide bond is known as the phi (Φ) torsion angle,

• The angle of the C–Ca bond to the adjacent peptide bond is known as the psi (Ψ) torsion angle

• This combination greatly restricts the number of possible conformations that a polypeptide chain can adopt and makes it possible to determine from simple steric considerations the most likely backbone conformation angles for polypeptide residues other than glycine.
This resonance restricts the number of conformations in proteins -- main chain rotations are restricted to \( \phi \) and \( \psi \).

***

\( \text{Psi (} \psi \text{) involves the } \text{N–Ca–C–N bonds} \) (with the rotation occurring about the \text{Ca–C} bond)

***

\( \text{Phi (} \phi \text{) involves the } \text{C–N–Ca–C} \) bonds (with the rotation occurring about the \text{N–Ca} bond),
Dihedral Angels (Torsion Angels)

- Torsion angles are **dihedral angles**, which are defined by 4 points in space.
- In proteins the two torsion angles phi and psi describe the rotation of the polypeptide chain around the two bonds on both sides of the Ca atom:

The Ramachandran angles in proteins are restricted to certain values, since some angles will result in sterical clashes between **main chain and side chain atoms** in the polypeptide.
Ramachandran plot

- Most combinations of $\phi$ and $\psi$ are not allowed because of Steric Collisions

Ramachandran plot
- To visualize backbone dihedral angles.

Two degrees of freedom:
1. $\phi$ (phi) angle = rotation about N – C$\alpha$
2. $\psi$ (psi) angle = rotation about C$\alpha$ – C
Ramachandran plot

Sterically allowed regions

Right handed alpha helixes
Beta sheet
Left handed alpha helixes (usually found in turn and loop region)

Observed values for all residues except glycine
Glycine residue can adopt many different conformations.

Observed values for glycine residue conformational angles. Notice that the values include combinations of $\phi$ and $\psi$ that are not allowed for other amino acids.

It allows unusual main chain conformation.

• http://guweb2.gonzaga.edu/faculty/cronk/CHEM440pub/L05-index.cfm?L05resource=dihedral

• http://employees.csbsju.edu/hjakubowski/classes/ch331/protstructure/olunderstandconfonfo.html

• http://public.csusm.edu/jayasinghe/BiomolTutorial/PhiPsiTutorial/PhiPsiTutorial.html
Certain side-chain conformations are energetically favorable: ROTAMERS

The staggered conformations are the most energetically favored conformations of two tetrahedrally coordinated carbon atoms. (a) A view along the C—C bond in ethane (CH₃CH₃) showing how the two carbon atoms can rotate. Three indistinguishable staggered conformations are obtained by rotation around the C—C bond. (b—d) Similar views of valine. The conformations are different for valine because the three groups are not equivalent. The first staggered conformation (b) is energetically most favored, less crowded and energetically most favored, the two methyl groups bound to Cβ are both close to small H atom bound to Cα.
Side-chain conformations

• Most side chains have one or few conformations that occur more frequently

> ROTAMERS!

• They can be used in computer programs for modeling proteins
Many proteins contain intrinsic metal atoms

- Excellent ligands: His, Cys, Asp, Glu, H2O
- Common metals: iron, zinc, magnesium, calcium

(a) Redox center of ribonucleotide reductase

(b) Alcohol dehydrogenase

Three zinc ligand
Many proteins contain metal atoms

<table>
<thead>
<tr>
<th>Metal</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium</td>
<td>Charge carrier; osmotic balance</td>
</tr>
<tr>
<td>Potassium</td>
<td>Charge carrier; osmotic balance</td>
</tr>
<tr>
<td>Magnesium</td>
<td>Structure; hydrolase; isomerase</td>
</tr>
<tr>
<td>Calcium</td>
<td>Structure; trigger; charge carrier</td>
</tr>
<tr>
<td>Vanadium</td>
<td>Nitrogen fixation; oxidase</td>
</tr>
<tr>
<td>Chromium</td>
<td>Unknown, possible involvement in glucose tolerance</td>
</tr>
<tr>
<td>Molybdenum</td>
<td>Nitrogen fixation; oxidase; oxo transfer</td>
</tr>
<tr>
<td>Tungsten</td>
<td>Dehydrogenase</td>
</tr>
<tr>
<td>Manganese</td>
<td>Photosynthesis; oxidase; structure</td>
</tr>
<tr>
<td>Iron</td>
<td>Oxidase; dioxygen transport and storage; electron transfer; nitrogen fixation</td>
</tr>
<tr>
<td>Cobalt</td>
<td>Oxidase; alkyl group transfer</td>
</tr>
<tr>
<td>Nickel</td>
<td>Hydrogenase; hydrolase</td>
</tr>
<tr>
<td>Copper</td>
<td>Oxidase; dioxygen transport; electron transfer</td>
</tr>
<tr>
<td>Zinc</td>
<td>Structure; hydrolase</td>
</tr>
</tbody>
</table>
Cys and His coordinate metal atoms

Zn-finger protein (Xenopus Xfin)

His coordinating Fe in Heme-binding protein
Many proteins contain intrinsic metal atoms

E3 ubiquitin-protein ligase UBR1
Choi et al,