Amino acids and primary structure of the protein
What Are Proteins?

- Large molecules
- Made up of **chains of amino acids**
- Are found in every cell in the body
- Are involved in most of the body’s functions and life processes
- The sequence of amino acids is determined by DNA
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.
A total of 20 amino acids are the basic building blocks of proteins:

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>abbreviations</th>
<th>Amino acid</th>
<th>abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>alanine</td>
<td>Ala A</td>
<td>phenylalanine</td>
<td>Phe F</td>
</tr>
<tr>
<td>arginine</td>
<td>Arg R</td>
<td>proline</td>
<td>Pro P</td>
</tr>
<tr>
<td>asparagine</td>
<td>Asn N</td>
<td>serine</td>
<td>Ser S</td>
</tr>
<tr>
<td>aspartic acid</td>
<td>Asp D</td>
<td>threonine</td>
<td>Thr T</td>
</tr>
<tr>
<td>cysteine</td>
<td>Cys C</td>
<td>tryptophan</td>
<td>Trp W</td>
</tr>
<tr>
<td>glutamine</td>
<td>Gln Q</td>
<td>tyrosine</td>
<td>Tyr Y</td>
</tr>
<tr>
<td>glutamic acid</td>
<td>Glu E</td>
<td>valine</td>
<td>Val V</td>
</tr>
<tr>
<td>glycine</td>
<td>Gly G</td>
<td>leucine</td>
<td>Leu L</td>
</tr>
<tr>
<td>histidine</td>
<td>His H</td>
<td>lysine</td>
<td>Lys K</td>
</tr>
<tr>
<td>isoleucine</td>
<td>Ile I</td>
<td>methionine</td>
<td>Met M</td>
</tr>
</tbody>
</table>

"three letter code" "one letter code"
Common structural features of amino acids

- All amino acids found in proteins are alpha-amino acids

- All amino acids found in proteins have the L-configuration

- Enantiomeric D-configuration does not occur in proteins

- Present in other macromolecular structures (cell wall of bacteria)
Absolute configuration of chiral centers

- Absolute configuration of a chiral centre can be determined by the Cahn-Ingold-Prelog system.
- Substituents are ranked according to their atomic number.
- The order of priority for some common functional groups is: SH > OH > NH₂ > COOH > CHO > CH₂OH > C₆H₅ > CH₃ > ²H > ¹H.

- If the order of substituents is counterclockwise it is S configuration (Latin: sinistrus, left) and if the order is clockwise it is the R configuration (Latin: rectus, right).
The anatomy of amino acids
Peptide Bonds Link Amino Acids

- Form when the acid group (COOH) of one amino acid joins with the amine group (NH$_2$) of a second amino acid
- Formed through condensation
- Broken through hydrolysis
Condensation and Hydrolytic Reactions

A peptide bond forms by condensation when the acid group (COOH) and amine group of two different amino acids join and release a molecule of water.

When peptide bonds are broken by hydrolysis, the hydroxyl group (OH) and hydrogen (H) from water are added.
Properties of amino acids

- Capacity to polymerize
- Useful acid-base properties
- Varied physical properties
- Varied chemical functionality
Essential, Nonessential, and Conditional AA

- *Essential* — must be consumed in the diet
- *Nonessential* — can be synthesized in the body
- *Conditionally essential* — cannot be synthesized due to illness or lack of necessary precursors
  - Premature infants lack sufficient enzymes needed to create arginine
Classification of amino acids

- According to the properties of their side chain amino acids can be grouped into five main classes:

**Nonpolar, aliphatic:** alanine, valine, leucine, isoleucine, methionine, proline and glycine

**Aromatic:** phenylalanine, tyrosine and tryptophan

**Polar, uncharged:** serine, threonine, cysteine, asparagine and glutamine

**Positively charged:** arginine, lysine and histidine

**Negatively charged:** aspartic and glutamic acid
A GUIDE TO THE TWENTY COMMON AMINO ACIDS

AMINO ACIDS ARE THE BUILDING BLOCKS OF PROTEINS IN LIVING ORGANISMS. THERE ARE OVER 500 AMINO ACIDS FOUND IN NATURE - HOWEVER, THE HUMAN GENETIC CODE ONLY DIRECTLY ENCODES 20. ‘ESSENTIAL’ AMINO ACIDS MUST BE OBTAINED FROM THE DIET, WHILST NON-ESSENTIAL AMINO ACIDS CAN BE SYNTHESISED IN THE BODY.

Chart Key:  
- ALIPHATIC 
- AROMATIC 
- ACIDIC 
- BASIC 
- HYDROXYLIC 
- SULFUR-CONTAINING 
- AMIDIC 
- NON-ESSENTIAL 
- ESSENTIAL

### Chemical Structure

- NAME A
- three letter code DNA codons

### Chart

- **ALANINE** (A)
  - Ala
  - GCT, GCC, GCA, GGG

- **GLYCINE** (G)
  - Gly
  - GCT, GCC, GCA, GGG

- **ISOLEUCINE** (I)
  - Ile
  - ATT, ATC, AIA

- **LEUCINE** (L)
  - Leu
  - CTT, CTC, CTA, CTG, TTA, TTG

- **PROLINE** (P)
  - Pro
  - CCT, CCG, CCA, CGG

- **VALINE** (V)
  - Val
  - GTT, GTC, GCA, GAG

- **PHENYLALANINE** (F)
  - Phe
  - TTT, TAC

- **TRYPTOPHAN** (W)
  - Trp
  - TGG

- **TYROSINE** (Y)
  - Tyr
  - TAT, TAC

- **ASPARTIC ACID** (D)
  - Asp
  - GAT, GAC

- **GLUTAMIC ACID** (E)
  - Glu
  - GAA, GAG

- **ARGININE** (R)
  - Arg
  - CGT, CGC, CGA, CGG, AGA, AGG

- **HISTIDINE** (H)
  - His
  - CAT, CAC

- **LYSINE** (K)
  - Lys
  - AAA, AAG

- **SERINE** (S)
  - Ser
  - TCT, TCC, TCA, TCG, AGT, AGC

- **THREONINE** (T)
  - Thr
  - ACT, ACC, ACA, ACG

- **CYSTEINE** (C)
  - Cys
  - TGT, TGC

- **METHIONINE** (M)
  - Met
  - ATG

- **ASPARAGINE** (N)
  - Asn
  - AAT, AAC

- **GLUTAMINE** (Q)
  - Gln
  - CA, CAG

### Note

This chart only shows those amino acids for which the human genetic code directly codes for. Selenocysteine is often referred to as the 21st amino acid, but is encoded in a special manner. In some cases, distinguishing between asparagine/aspartic acid and glutamine/glutamic acid is difficult. In these cases, the codes asx (B) and glx (Z) are respectively used.

© COMPOUND INTEREST 2014 - WWW.COMPOUNDCHEM.COM | Twitter: @compoundchem | Facebook: www.facebook.com/compoundchem
Shared under a Creative Commons Attribution-NonCommercial-NoDerivatives licence.
Glycine is the smallest amino acid

- Glycine is the only achiral standard amino acid
Apolar (hydrophobic) amino acids

<table>
<thead>
<tr>
<th></th>
<th>Structure</th>
<th>Name, Code, Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td><img src="#" alt="Alanine Structure" /></td>
<td>Alanine, Ala, A</td>
</tr>
<tr>
<td>Valine</td>
<td><img src="#" alt="Valine Structure" /></td>
<td>Valine, Val, V</td>
</tr>
<tr>
<td>Leucine</td>
<td><img src="#" alt="Leucine Structure" /></td>
<td>Leucine, Leu, L</td>
</tr>
<tr>
<td>Isoleucine</td>
<td><img src="#" alt="Isoleucine Structure" /></td>
<td>Isoleucine, Ile, I</td>
</tr>
<tr>
<td>Proline</td>
<td><img src="#" alt="Proline Structure" /></td>
<td>Proline, Pro, P</td>
</tr>
</tbody>
</table>
The planar hydrophobic amino acids

- Phenylalanine, Phe, F
- Tyrosine, Tyr, Y
- Tryptophan, Trp, W
Serine and Threonine have a hydroxylated side chain
The uncharged derivatives of aspartic and glutamic acid
The positively charged amino acids

- Lysine, Lys, K
- Arginine, Arg, R
- Histidine, His, H
Acidic amino acids

- Aspartic and glutamic acid have a carboxyl group in their side chains and are therefore acidic
Sulfur-containing amino acids

- In methionine, the sulfur atom is buried in the interior of a hydrocarbon side chain.
- In cysteine it is part of an exposed, chemically reactive thiol group.
Formation of a disulfide-bridge stabilizes protein structure
If it is easier for you... 

- Non-polar side chains
• Polar side chains
Electrically charged side chains

- **Acidic**
  - Aspartate (D) Asp
  - Glutamate (E) Glu

- **Basic**
  - Lysine (K) Lys
  - Arginine (R) Arg
  - Histidine (H) His
Aromatic side chains

Aromatic R groups

Phenylalanine

Tyrosine

Tryptophan
Uncommon Amino Acids in Proteins

- Not incorporated by ribosomes
- 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.
Amino acid derivatives in proteins

• 4-Hydroxyproline and 5-hydroxylysine are present in the structural protein collagen which is the most abundant protein in mammals

• Gama-Carboxyglutamate is a constituent of proteins involved in blood clotting

• Other modifications are phosphorylations, acetylations, methylations and glycosylations.
Average abundance of the 20 standard amino acids in proteins

<table>
<thead>
<tr>
<th>above average (%)</th>
<th>below average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>alanine</td>
<td>arginine</td>
</tr>
<tr>
<td>glycine</td>
<td>proline</td>
</tr>
<tr>
<td>leucine</td>
<td>isoleucine</td>
</tr>
<tr>
<td>serine</td>
<td>asparagine</td>
</tr>
<tr>
<td>lysine</td>
<td>glutamine</td>
</tr>
<tr>
<td>valine</td>
<td>tyrosine</td>
</tr>
<tr>
<td>glutamate</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>threonine</td>
<td>cysteine</td>
</tr>
<tr>
<td>aspartate</td>
<td>histidine</td>
</tr>
<tr>
<td></td>
<td>methionine</td>
</tr>
<tr>
<td></td>
<td>tryptophan</td>
</tr>
</tbody>
</table>

Small, apolar amino acids make up 1/4 of proteins. Aromatic (8.1%) and sulfur-containing (4.5%) amino acids are rare.
Non-standard amino acids not found in proteins

- Citrulline
- Ornithine
- Thyroxine

These two amino acids are important intermediates in the nitrogen metabolism (urea cycle).

Iodine-containing thyroid hormone that generally stimulates vertebrate metabolism.
Summary

- Proteins: definition
- Amino acids: anatomy, classification and structure
- Modifications
- Derivatives
- Average abundance
- Non-standard AA
Small peptides have important biological roles

- Enkephalins are pentapeptides for control of pain:
  - Leucine enkephalin: Tyr-Gly-Gly-Phe-Leu
  - Methionine enkephalin: Tyr-Gly-Gly-Phe-Met
- Bradykinin, vasopressin and oxytocin are peptide hormones. They are all nonapeptides.
  - Bradykinin inhibits the inflammation of tissues
  - Vasopressin controls the blood pressure.
  - Oxytocin induces labor in pregnant women and stimulates milk production in nursing mothers.
Relative hydropathy of amino acid side chains

- hydropathy index combines hydrophobicity and hydrophilicity of the side chains
- regarded as a measure of the tendency of an amino acid to seek an aqueous environment

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Hydropathy Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>isoleucine</td>
<td>4.5*</td>
</tr>
<tr>
<td>valine</td>
<td>4.2</td>
</tr>
<tr>
<td>leucine</td>
<td>3.8</td>
</tr>
<tr>
<td>phenylalanine</td>
<td>2.8</td>
</tr>
<tr>
<td>cysteine</td>
<td>2.5</td>
</tr>
<tr>
<td>methionine</td>
<td>1.9</td>
</tr>
<tr>
<td>alanine</td>
<td>1.8</td>
</tr>
<tr>
<td>proline</td>
<td>1.6</td>
</tr>
<tr>
<td>glycine</td>
<td>-0.4</td>
</tr>
<tr>
<td>threonine</td>
<td>-0.7</td>
</tr>
<tr>
<td>serine</td>
<td>-0.8</td>
</tr>
<tr>
<td>tryptophan</td>
<td>-0.9</td>
</tr>
<tr>
<td>tyrosine</td>
<td>-1.3</td>
</tr>
<tr>
<td>histidine</td>
<td>-3.2</td>
</tr>
<tr>
<td>asparagine</td>
<td>-3.5</td>
</tr>
<tr>
<td>glutamine</td>
<td>-3.5</td>
</tr>
<tr>
<td>aspartate</td>
<td>-3.5</td>
</tr>
<tr>
<td>glutamate</td>
<td>-3.5</td>
</tr>
<tr>
<td>lysine</td>
<td>-3.9</td>
</tr>
<tr>
<td>arginine</td>
<td>-4.5</td>
</tr>
</tbody>
</table>
Acid-base properties of amino acids

- amino acid can never exist as an uncharged compound in solution
- However, at a certain pH the negative charge of the carboxyl group exactly balances the positive charge of the amino group
- *Isoelectric point (pI)* - amino acid has no net charge
- \( pI = pK_a(COOH) + pK_a(NH_2) / 2 \)
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)*
## pKa values of amino acids

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>$pK_a$(COOH)</th>
<th>$pK_a$(NH$_2$)</th>
<th>$pK_a$(side chain)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>2.34</td>
<td>9.69</td>
<td>-</td>
</tr>
<tr>
<td>Arginine</td>
<td>2.17</td>
<td>9.04</td>
<td>12.48</td>
</tr>
<tr>
<td>Asparagine</td>
<td>2.02</td>
<td>8.84</td>
<td>-</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>2.09</td>
<td>9.82</td>
<td>3.86</td>
</tr>
<tr>
<td>Cysteine</td>
<td>1.71</td>
<td>10.78</td>
<td>8.33</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>2.19</td>
<td>9.67</td>
<td>4.25</td>
</tr>
<tr>
<td>Glutamine</td>
<td>2.17</td>
<td>9.13</td>
<td>-</td>
</tr>
<tr>
<td>Glycine</td>
<td>2.34</td>
<td>9.6</td>
<td>-</td>
</tr>
<tr>
<td>Histidine</td>
<td>1.82</td>
<td>9.17</td>
<td>6.04</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>2.36</td>
<td>9.68</td>
<td>-</td>
</tr>
<tr>
<td>Leucine</td>
<td>2.36</td>
<td>9.6</td>
<td>-</td>
</tr>
<tr>
<td>Lysine</td>
<td>2.18</td>
<td>8.95</td>
<td>10.8</td>
</tr>
<tr>
<td>Methionine</td>
<td>2.28</td>
<td>9.21</td>
<td>-</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>1.83</td>
<td>9.13</td>
<td>-</td>
</tr>
<tr>
<td>Proline</td>
<td>1.99</td>
<td>10.6</td>
<td>-</td>
</tr>
<tr>
<td>Serine</td>
<td>2.21</td>
<td>9.15</td>
<td>-</td>
</tr>
<tr>
<td>Threonine</td>
<td>2.63</td>
<td>9.1</td>
<td>-</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>2.38</td>
<td>9.39</td>
<td>-</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>2.2</td>
<td>9.11</td>
<td>10.07</td>
</tr>
<tr>
<td>Valine</td>
<td>2.32</td>
<td>9.62</td>
<td>-</td>
</tr>
</tbody>
</table>
Some R-groups can be ionized

Henderson-Hasselbalch

\[ \text{pH} = pK' - \log \frac{[HB]}{[B^-]} \]

Allows calculation of the ratio of a weak acid and its conjugate base at any pH
**dissociation constant,** $K_a$

$$
\ce{H + A <=> K_a \rightarrow H^+ + A^-}
$$

$$
K_a = \frac{[H^+][A^-]}{[HA]} \quad \text{pK}_a = -\log_{10} K_a
$$

**Henderson-Hasselbalch**

$$
pH = pK' - \log \left( \frac{[HA]}{[A^-]} \right)
$$

when the pH = pKa then $\log \frac{[HA]}{[A^-]} = 0$

therefore $[HA] = [A^-]$ 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 $pH < pK_a$, then $\log \frac{[HA]}{[A^-]}$ has to be $> 0$ so $[HA] > [A^-]$. 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 $pH > pK_a$ and $\log \frac{[HA]}{[A^-]}$ has to be $< 0$ so $[HA] < [A^-]$. This makes sense as it tells us that a *stronger base* will cause the formation of $A^-$, the deprotonated form.
Neutral side chains: These amino acids are characterized by two pKas: pKa1 and pKa2 for the carboxylic acid and the amine respectively. \( pI = \frac{1}{2} (pKa_1 + pKa_2) \)

Acidic side chains: The pH 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. \( pI = \frac{1}{2} (pKa_1 + pKa_3) \)

Basic side chains: The pH 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. \( pI = \frac{1}{2} (pKa_1 + pKa_3) \)
Chemical grouping of the amino acids

- organization of the genetic code reflects the chemical grouping
- 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
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 “homologs”
- reveals highly conserved regions that are important for function
- “motifs” - Clusters of conserved residues
- carry out a particular function or form a particular structure that is important for the protein
Structure of the Protein

• Four levels of structure
  • Primary structure
  • Secondary structure
  • Tertiary structure
  • Quaternary structure
Primary structure

- Is the unique sequence of amino acids in a polypeptide
- Like order of the letters in a very long word
- Determined by inherited genetic information
- Transcription/translation

Figure 5.20
Secondary structure

- Is the folding or coiling of the polypeptide into a repeating configuration
- Includes the $\alpha$ helix and the $\beta$ pleated sheet
α helix

- Coil
- Held together by hydrogen bonding between every fourth amino acid
β-sheet

- Two or more regions of polypeptide chain are lying side by side
- Connected by hydrogen bonds between parts of the two parallel polypeptide backbones
- Make up the core of many globular proteins and in some fibrous proteins (spider web)
Tertiary structure

- Is the overall three-dimensional shape of a polypeptide
- Results from interactions between amino acids and R groups
Different types of interactions

**Hydrophobic interactions**: upon protein folding, AA with hydrophobic side chains end up in clusters at the core of the protein (out of water contact)

**Disulfide bridges**: where 2 cysteine monomers are brought close together by the folding. $-S-S-$
Quaternary structure

- Is the overall protein structure that results from the aggregation of two or more polypeptide subunits
Levels of protein organization

Primary protein structure
is sequence of a chain of amino acids

Secondary protein structure
occurs when the sequence of amino acids are linked by hydrogen bonds

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

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

• Alteration of the protein’s shape and thus functions through the use of

  • Heat
  • Acids
  • Bases
  • Salts
  • Mechanical agitation

• Primary structure is unchanged by denaturing
# Roles of proteins

## Table 6.3

### The Many Roles of Proteins

<table>
<thead>
<tr>
<th>Role of Protein</th>
<th>How It Works</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural and mechanical support and</td>
<td>Proteins are the body’s building materials, providing strength and flexibility to tissues, tendons, ligaments, muscles, organs, bones, nails, hair, and skin. Proteins are also needed for the ongoing maintenance of the body.</td>
</tr>
<tr>
<td>maintenance</td>
<td></td>
</tr>
<tr>
<td>Enzymes and hormones</td>
<td>Proteins are needed to make most enzymes that speed up reactions in the body and many hormones that direct specific activities, such as regulating blood glucose levels.</td>
</tr>
<tr>
<td>Fluid balance</td>
<td>Proteins play a major role in ensuring that body fluids are evenly dispersed in the blood and inside and outside cells.</td>
</tr>
<tr>
<td>Acid-base balance</td>
<td>Proteins act as buffers to help keep the pH of body fluids within a tight range. A drop in pH will cause body fluids to become too acidic, whereas a rise in pH can make them too basic.</td>
</tr>
<tr>
<td>Transport</td>
<td>Proteins shuttle substances such as oxygen, waste products, and nutrients (such as sodium and potassium) through the blood and into and out of cells.</td>
</tr>
<tr>
<td>Antibodies and the immune response</td>
<td>Proteins create specialized antibodies that attack pathogens that may cause illness.</td>
</tr>
<tr>
<td>Energy</td>
<td>Because proteins provide 4 calories per gram, they can be used as fuel or energy.</td>
</tr>
</tbody>
</table>
Summary

- Amino acids
- Classification
- pI and pKa
- Levels of protein structure
- Protein forms and roles
Quiz 1

- Next week Thursday 24.10.2019
- From the first lecture
- 4 questions