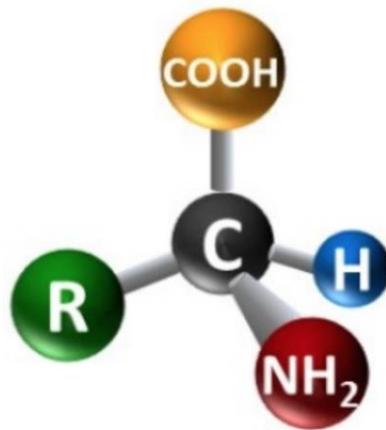




# Amino Acids 1



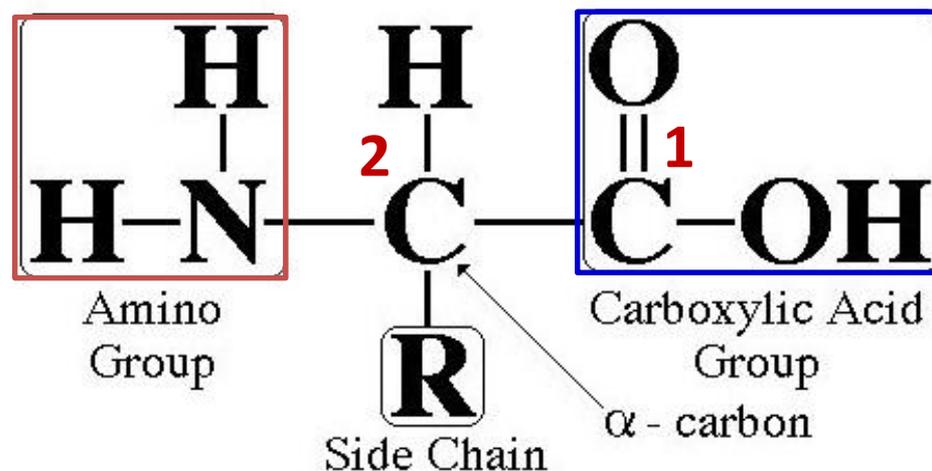
Dr. Nesrin Mwafi

Biochemistry & Molecular Biology Department  
Faculty of Medicine, Mutah University



# Amino Acid Structure

- Amino acids are biologically important organic molecules that contain both **carboxylic acid (-COOH)** as well as **amine (-NH<sub>2</sub>)** groups
- The side-chain also called “**R**” group is specific to each amino acid



- Amino group is attached to  $\alpha$ -carbon (**C2**)
- C, N, O and H are the key elements of amino acids

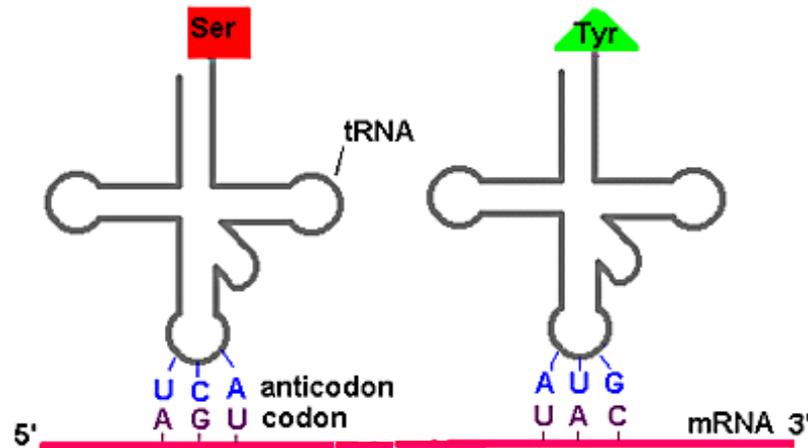
# Biological significance of Amino Acids



1. Amino acids are N-containing molecules
  2. The basic structural building units (monomers) of proteins
  3. Precursors of many biomolecules like neurotransmitters (non-protein role)
  4. They are also utilized as an energy source
- There are **20 standard** (canonical) amino acids which are encoded directly by triplet codons in the universal genetic code during in vivo protein synthesis process (mRNA translation)



# Genetic Code Table



- **The 20 standard** amino acids are known as proteinogenic or natural amino acids

1st base in codon

|   |     | 2nd base in codon |      |      |   |  |
|---|-----|-------------------|------|------|---|--|
|   |     | U                 | C    | A    | G |  |
| U | Phe | Ser               | Tyr  | Cys  | U |  |
|   | Phe | Ser               | Tyr  | Cys  | C |  |
|   | Leu | Ser               | STOP | STOP | A |  |
|   | Leu | Ser               | STOP | Trp  | G |  |
| C | Leu | Pro               | His  | Arg  | U |  |
|   | Leu | Pro               | His  | Arg  | C |  |
|   | Leu | Pro               | Gln  | Arg  | A |  |
|   | Leu | Pro               | Gln  | Arg  | G |  |
| A | Ile | Thr               | Asn  | Ser  | U |  |
|   | Ile | Thr               | Asn  | Ser  | C |  |
|   | Ile | Thr               | Lys  | Arg  | A |  |
|   | Met | Thr               | Lys  | Arg  | G |  |
| G | Val | Ala               | Asp  | Gly  | U |  |
|   | Val | Ala               | Asp  | Gly  | C |  |
|   | Val | Ala               | Glu  | Gly  | A |  |
|   | Val | Ala               | Glu  | Gly  | G |  |

3rd base in codon

# Standard Amino Acids List

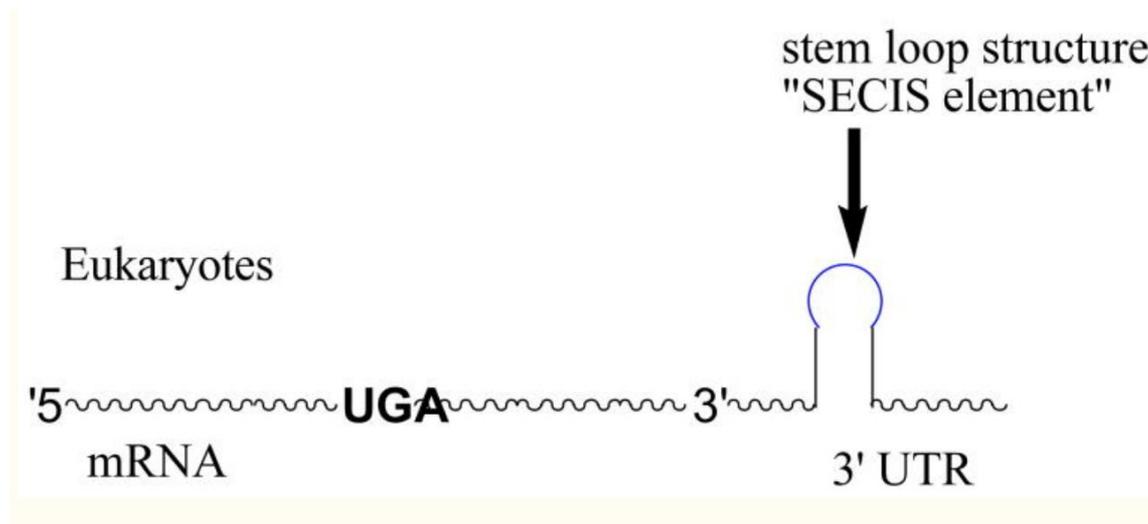


|               |            |            |
|---------------|------------|------------|
| Histidine     | Arginine   | Alanine    |
| Isoleucine    | Asparagine | Asparatate |
| Leucine       | Glutamine  | Cysteine   |
| Methionine    | Glycine    | Glutamate  |
| Phenylalanine | Proline    |            |
| Threonine     | Serine     |            |
| Tryptophan    | Tyrosine   |            |
| Valine        |            |            |
| Lysine        |            |            |

# Standard Amino Acids



1. They are proteinogenic and natural amino acids (the other proteinogenic amino acids N-formyl methionine, pyrrolysine and selenocysteine are called non-standard or non-canonical amino acids)

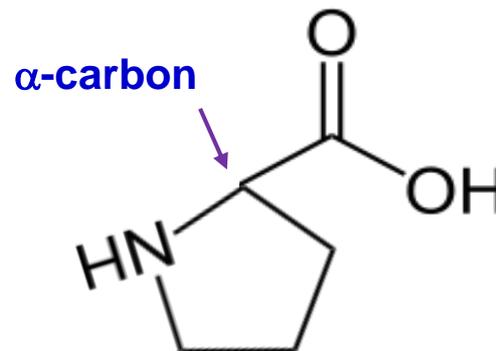
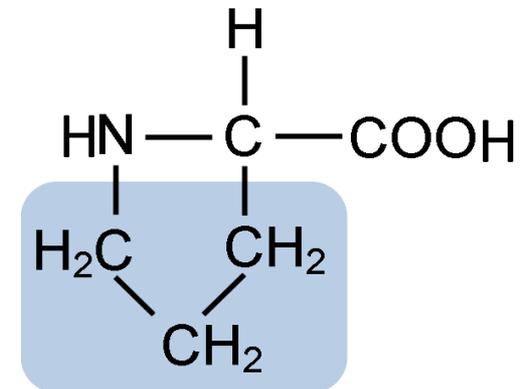
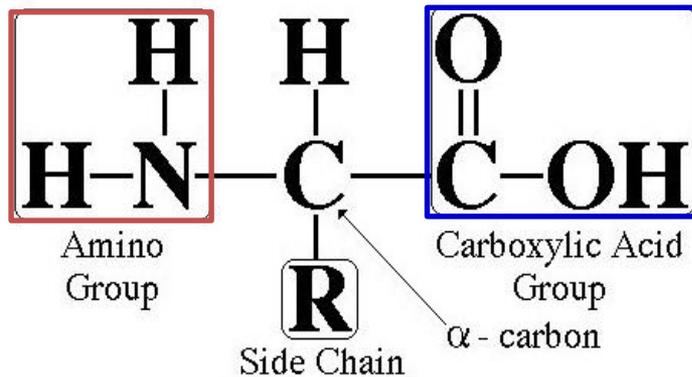


**Incorporation of selenocysteine in protein structure by unique mechanism**



# Standard Amino Acids

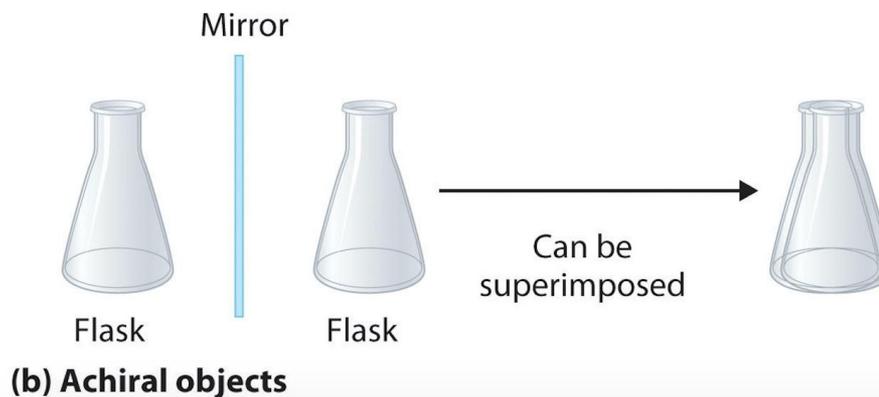
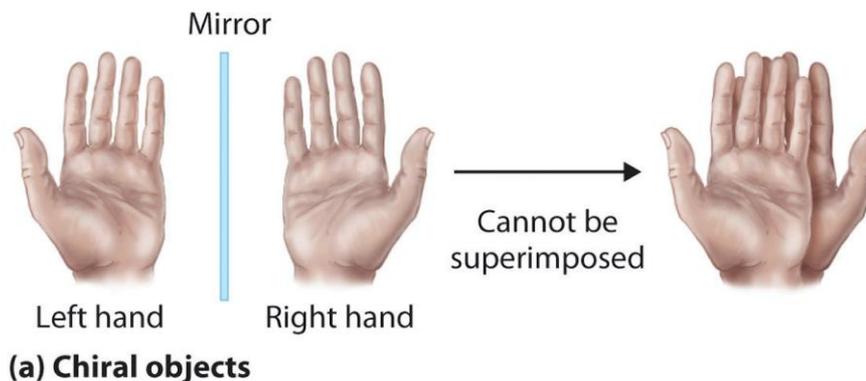
2. Known as 2-, alpha- or  **$\alpha$ -amino acids** as the primary amino group ( $-\text{NH}_2$ ) is attached to  $\alpha$ -carbon (the carbon next to  $-\text{COOH}$  group). Proline is an exception which has a secondary amino group ( $-\text{NH}-$ )



# Standard Amino Acids



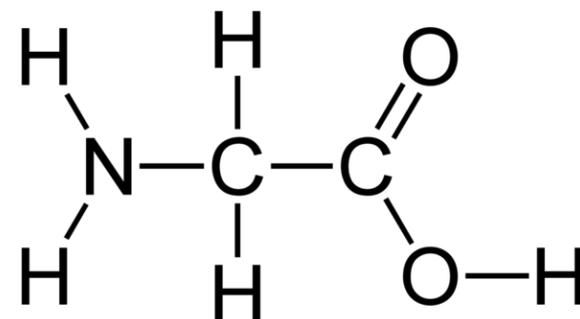
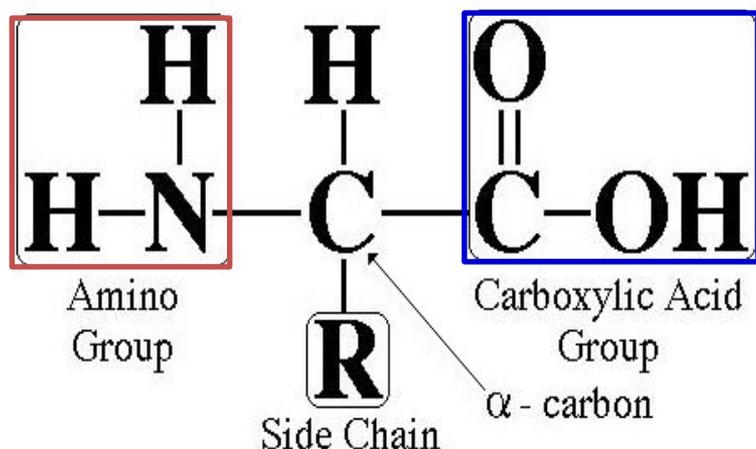
3. They are all chiral molecules (except glycine which has **achiral** C) with **L-** stereochemical configuration (left-handed isomers)





# Standard Amino Acids

- Chiral molecules should contain at least one chiral center (**usually a carbon atom**)
- **Chiral carbon**: asymmetric carbon atom attached to 4 different groups of atoms



**Glycine**

# Isomerization



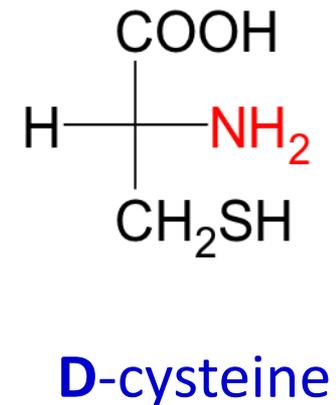
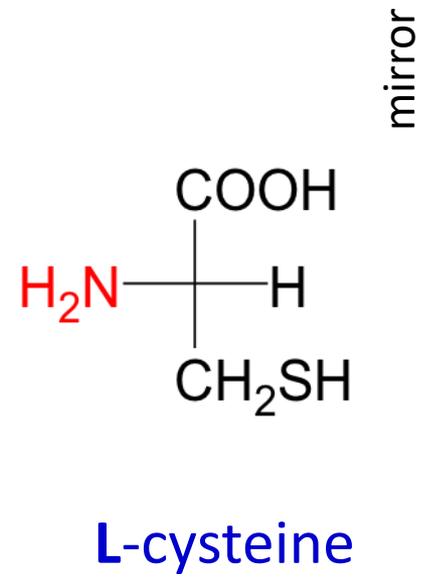
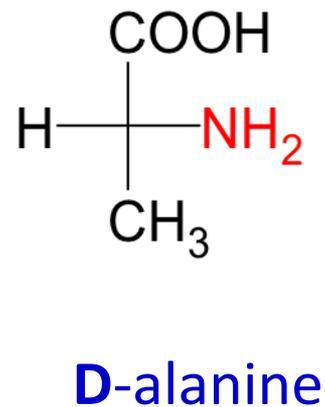
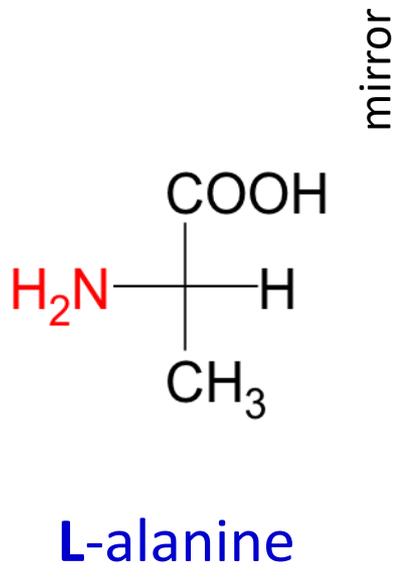
- Isomers: are molecules with same molecular formula but different chemical structures
  1. Constitutional (structural) isomers: atoms and functional groups bind together in different ways
  2. Stereoisomers (spatial isomers): differ in the configuration of atoms rather than the order of atomic connectivity

# D/L Amino Acids



- Enantiomers: are two stereoisomers that are mirror images to each other but not superimposable
- **D-** (dexter)/**L-** (laevus) Nomenclature system: commonly used to assign the configurations in sugars (carbohydrates) and amino acids
- As a rule of thumb: if the amino group is on the right-hand side of  $\alpha$ -carbon at Fisher projection, the configuration is D. If it is on the left-hand side, the configuration is assigned as L.

# Fischer Projections of Amino Acids



**Fisher Projection:** is one way commonly used to represent the structure of chiral molecules like carbohydrates and amino acids

# D/L Amino Acids

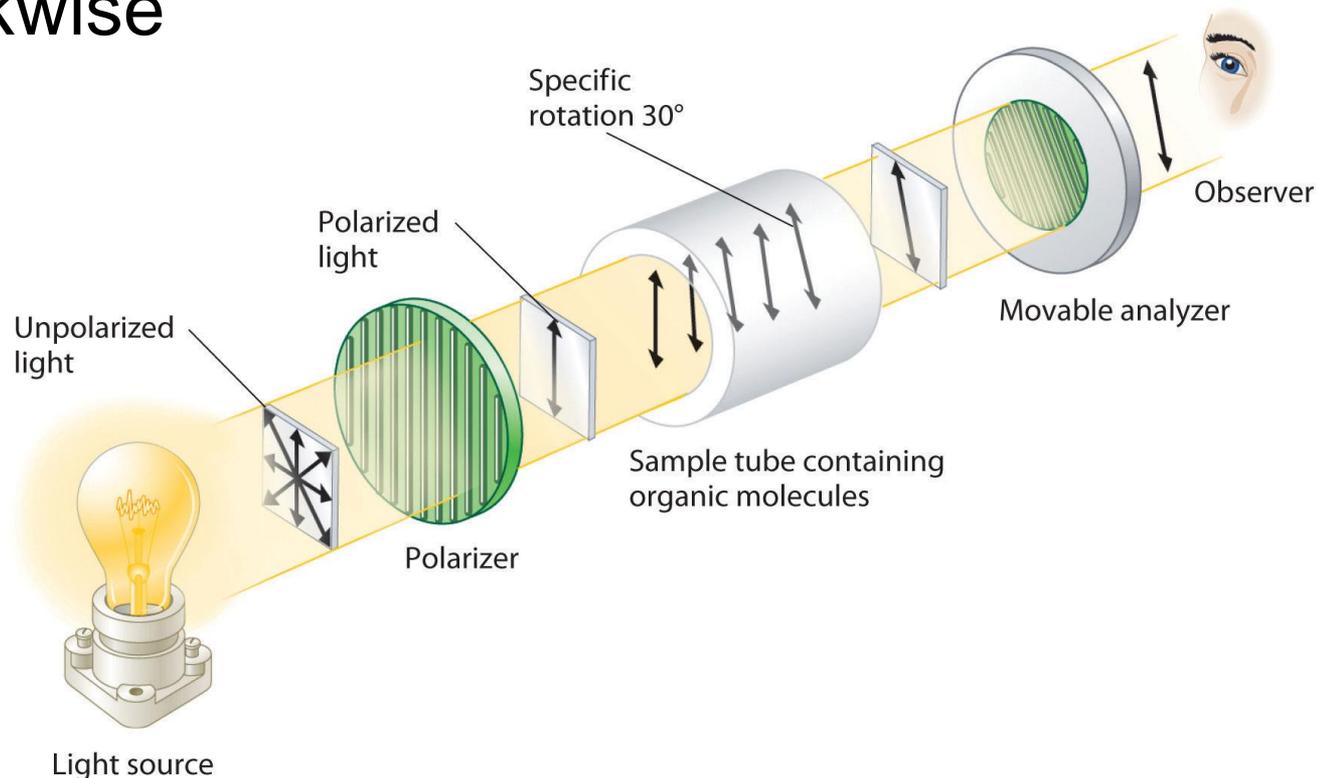


- Most naturally occurring sugars are D-isomers while most naturally occurring amino acids are L-isomers (amino acids of protein)
- D-amino acids polypeptides (right-handed isomers) are components of bacterial cell walls to resist digestion by other organisms



# Optical Activity

- Enantiomers are optically active and can rotate the polarized light plane either clockwise or counterclockwise



**Polarimeter is used to measure optical rotation**

# Optical Activity

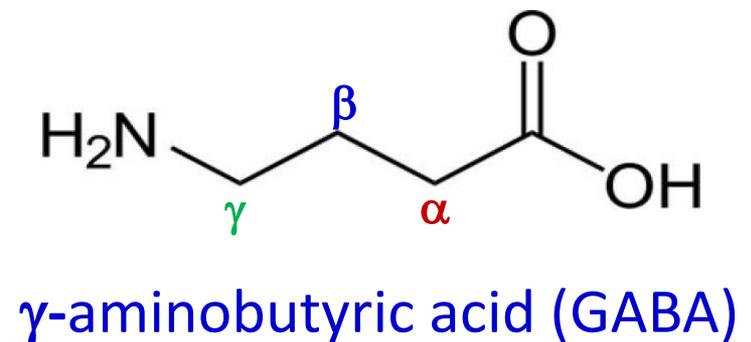
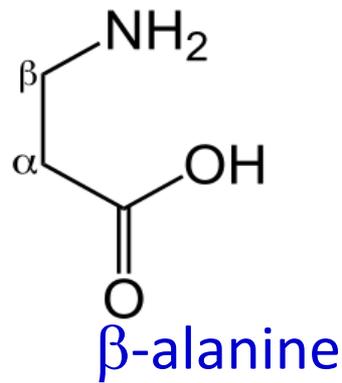
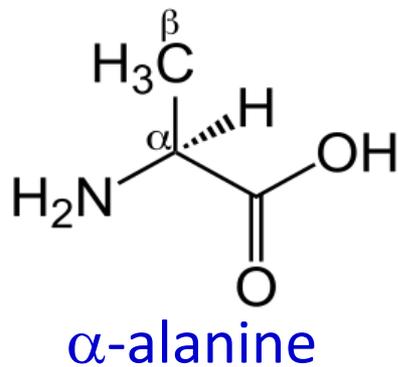


- **(+)/(-) nomenclature system:** if one enantiomer rotates the light clockwise, it is labeled (+) or (*d*) (dextrorotatory). The second mirror image enantiomer is labeled (-) or (*l*) laevorotatory
- D/L system should not be confused with +/- or *d/l* system. For example, D-isomer might be levorotatory
- 9 of 19 L-amino acids commonly found in proteins are dextrorotatory
- **Racemic mixture** contains equal amounts of each enantiomer (net rotation is zero)

# Classification of Amino Acids



- >300 amino acids classified in many ways:
  - 1) Proteinogenic and non-proteinogenic amino acids (either have non-protein role like GABA and carnitine or formed by post-translational modification of protein like hydroxyproline )
  - 2) Standard and non-standard amino acids
  - 3)  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  amino acids

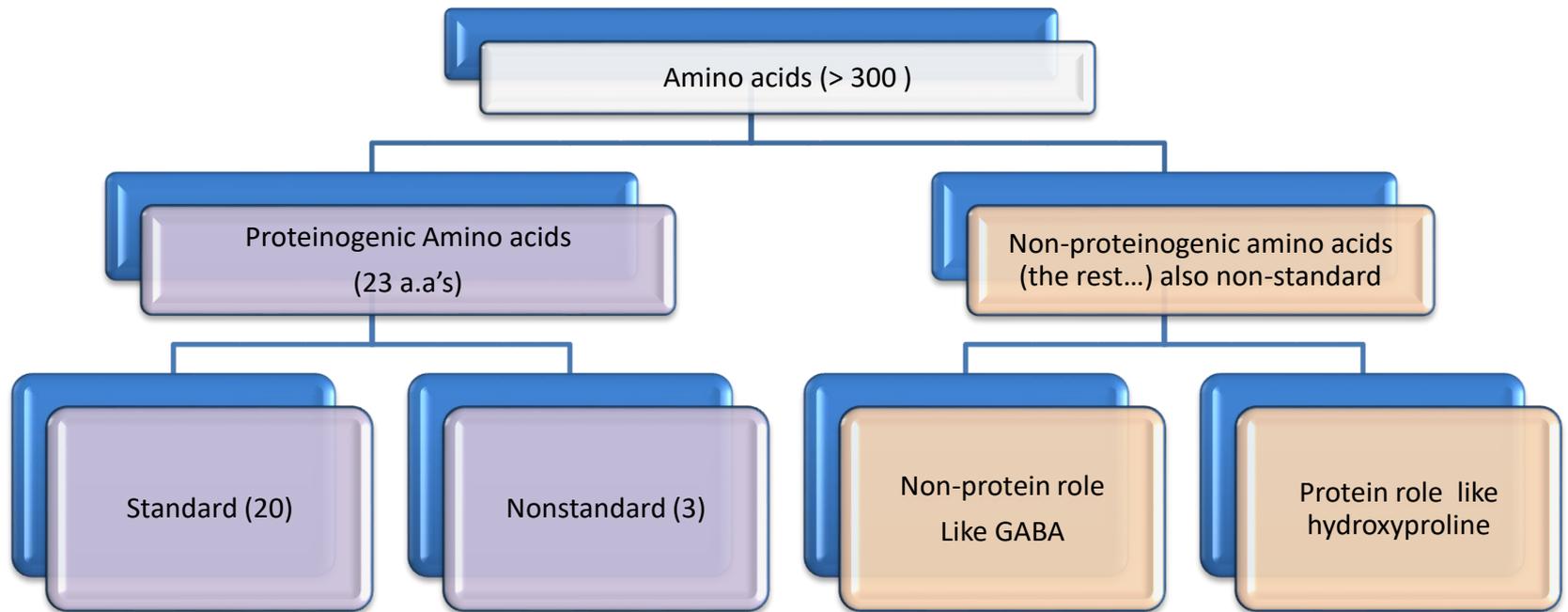


# Classification of Amino Acids



- $\beta$ -amino acids are non-proteinogenic with  $\beta$ -alanine is the only common naturally occurring  $\beta$ -amino acid.  $\beta$ -alanine is used in plants and microorganisms in the synthesis of pantothenic acid (vitamin B<sub>5</sub>)
- Unlike  $\alpha$ -peptides, The  $\beta$ -peptides are artificial peptides used in some antibiotics to counter resistance as they are more stable against proteolytic degradation

# Classification of Amino Acids



# Categories of Standard Amino Acids

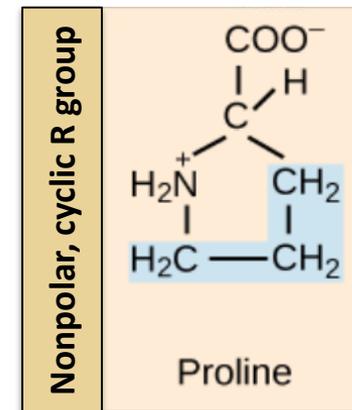
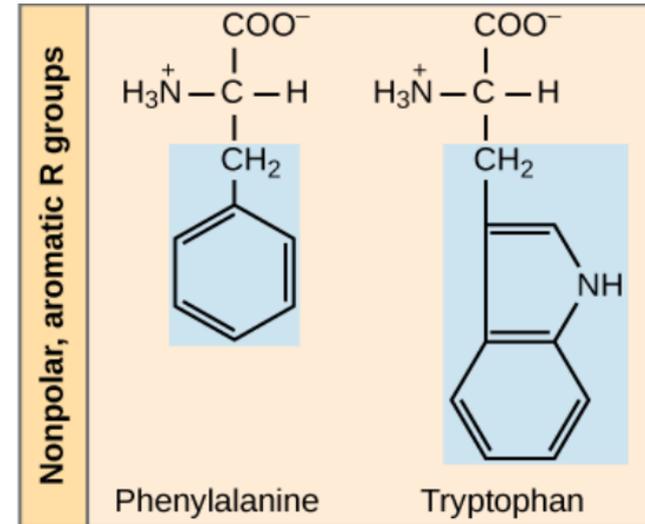
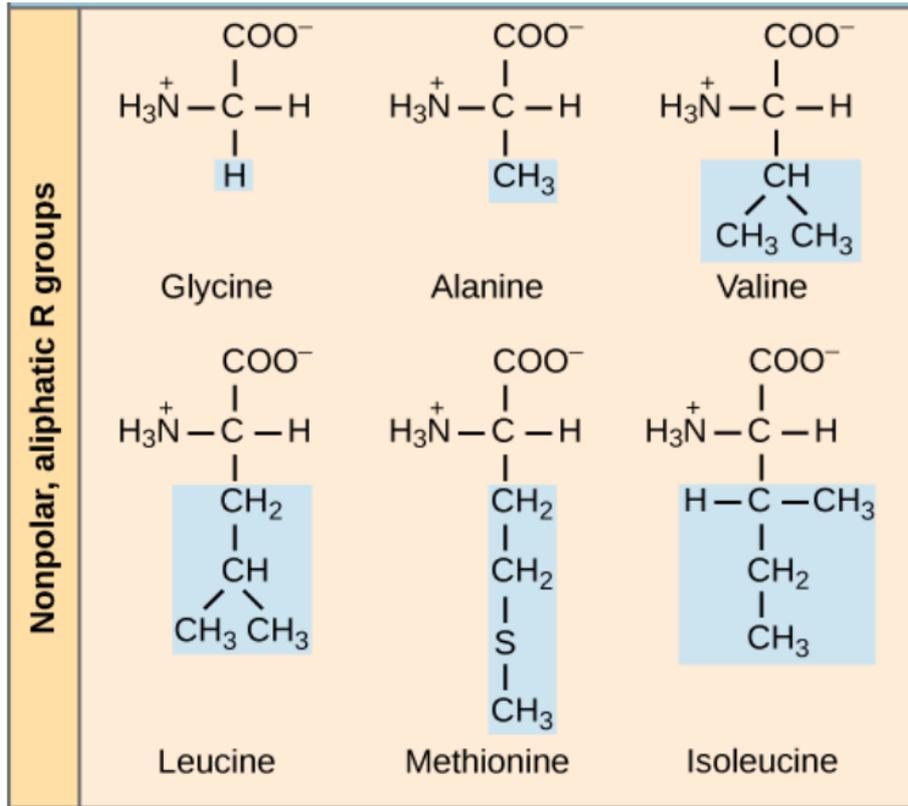


- The 20 standard amino acids are classified into 3 major categories according to the polarities of their “R” groups:
  - 1) **Amino acids with non-polar R groups**
  - 2) **Amino acids with charged polar R groups**
  - 3) **Amino acids with uncharged polar R groups**

# Amino acids with non-polar R groups



- 6 amino acids with aliphatic, 2 with aromatic and one with cyclic side chains



# Amino acids with non-polar R groups

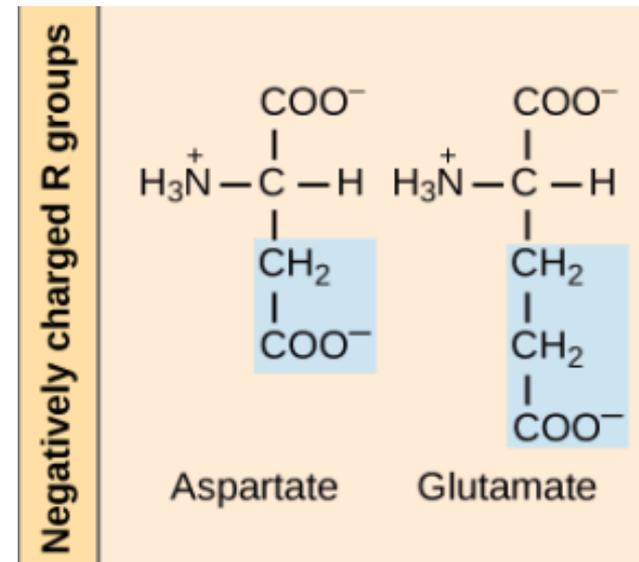
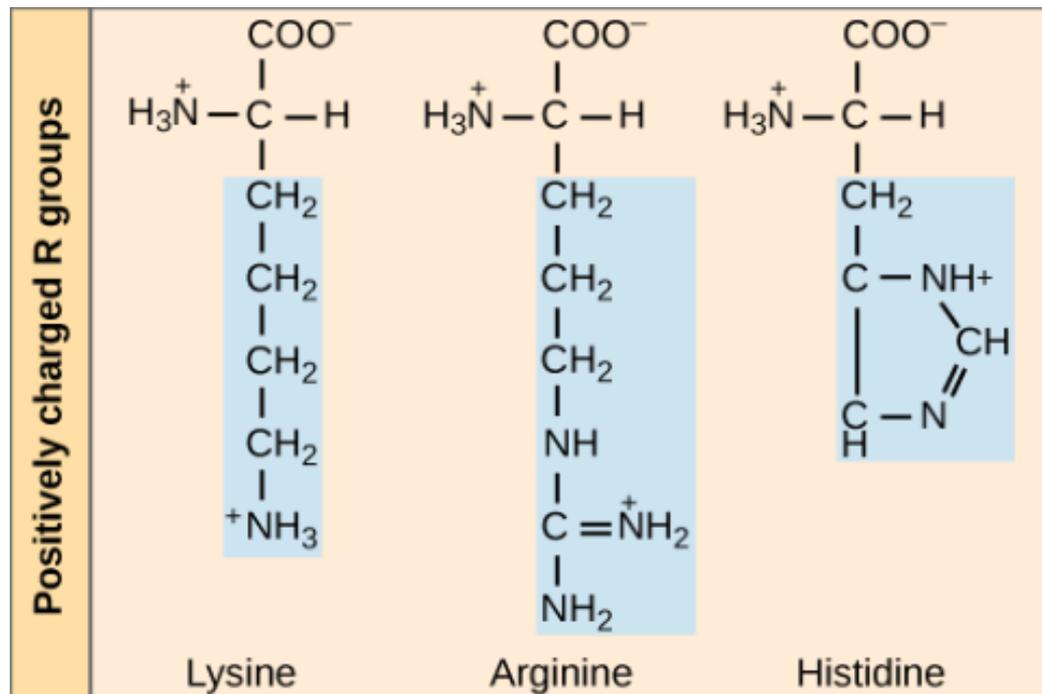


- Glycine has the simplest side chain: H atom
- Alanine, valine, leucine and isoleucine have aliphatic hydrocarbon side chains
- Methionine has a thioether side chain (sulfur atom)
- Proline has a cyclic pyrrolidine side chain
- Phenylalanine has a phenyl moiety
- Tryptophan has an indole group



# Amino acids with charged polar R groups

- 3 amino acids are positively charged (basic) and 2 amino acids are negatively charged (acidic)





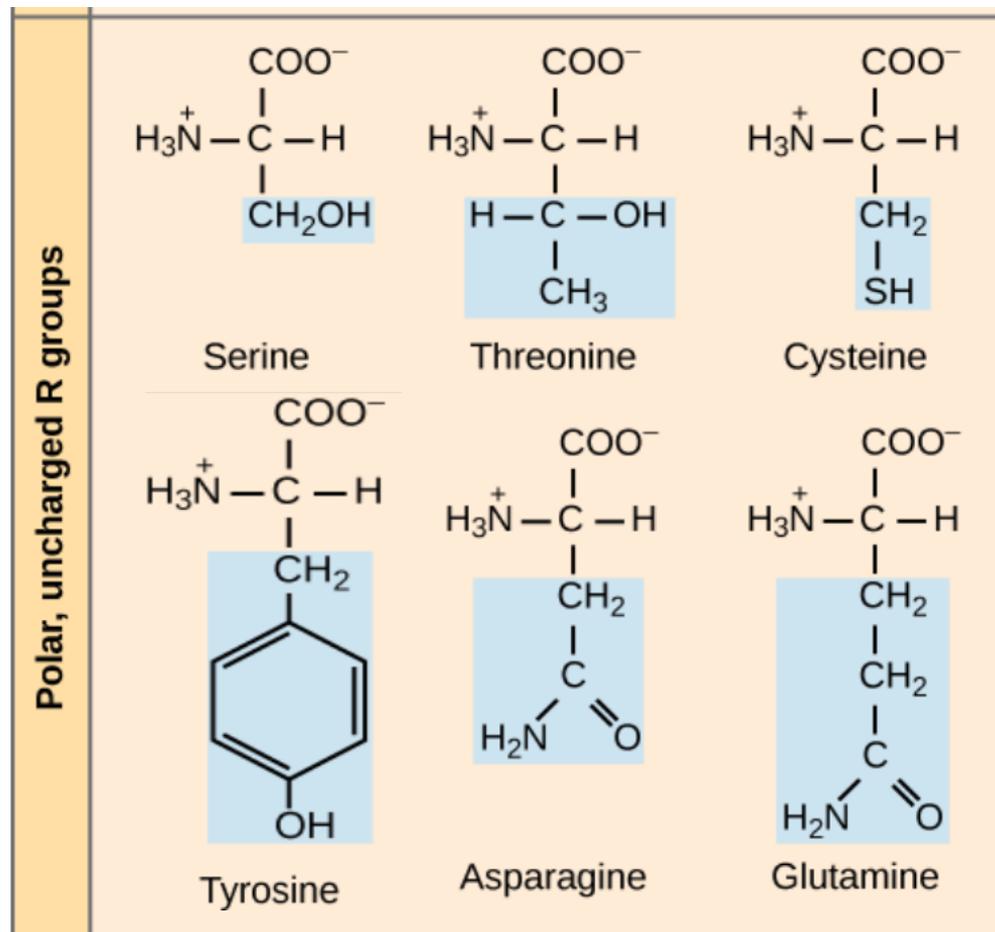
## Amino acids with charged polar R groups

- Arginine has a guanidine group
- Lysine has a butyl ammonium side chain
- Histidine has imidazole group
- Aspartic and glutamic acids in their ionized state are called aspartate and glutamate, respectively

# Amino acids with uncharged polar R groups



- 6 amino acids with hydroxyl, amide or thiol groups





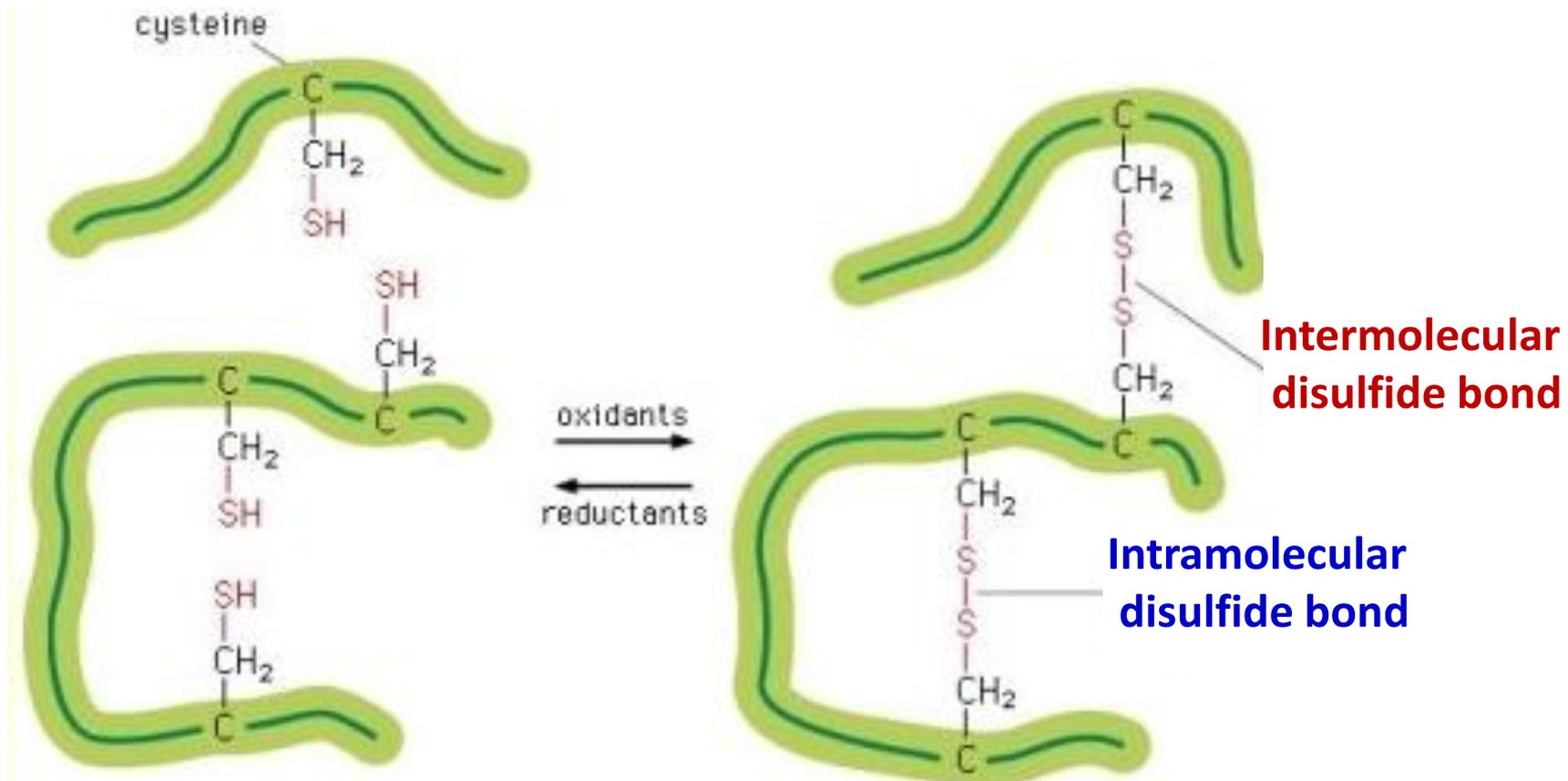
## Amino acids with uncharged polar R groups

- Serine and threonine bear hydroxyl (-OH) R group
- Asparagine and glutamine have amide bearing side chains. They are the amide derivatives of aspartic and glutamic acids
- Tyrosine is aromatic and has a phenolic group
- Cysteine has a thiol group that can form a disulfide bond (-S-S-) with another cysteine through the oxidation of 2 thiol groups (cystine is the oxidized **dimeric** form). The disulfide bridge in proteins contributes to the stability and overall shape of a protein

# Amino acids with uncharged polar R groups



- Disulfide bond** is a covalent linkage formed between the sulfhydryl groups (SH) of **two cysteine residues** (after oxidation) to produce a **cystine** residue



## Amino acids with uncharged polar R groups



- Cysteine residues may be separated from each other by many amino acids in the primary sequence of a polypeptide or may even be located on two different polypeptides. The folding of the polypeptide chain(s) brings the cysteine residues into proximity and permits covalent bonding of their side chains.
- Disulfide bond could be **intramolecular** (2 cysteine residues on the same polypeptide chain) or **intermolecular** (2 cysteine residues on two separate/ different polypeptide chains)

# Amino Acids Abbreviations



| <u>3-letters</u> | <u>1-letter</u> | <u>Amino acid</u>         |
|------------------|-----------------|---------------------------|
| Ala              | A               | <u>A</u> lanine           |
| Arg              | R               | <u>A</u> rginine          |
| Asn              | N               | <u>A</u> sparagine        |
| Asp              | D               | Aspartic acid (Aspartate) |
| Cys              | C               | <u>C</u> ysteine          |
| Gln              | Q               | Glutamine                 |
| Glu              | E               | Glutamic acid (Glutamate) |
| Gly              | G               | <u>G</u> lycine           |
| His              | H               | <u>H</u> istidine         |
| Ile              | I               | <u>I</u> soleucine        |
| Leu              | L               | <u>L</u> eucine           |
| Lys              | K               | Lysine                    |
| Met              | M               | <u>M</u> ethionine        |
| Phe              | F               | Phenylalanine             |
| Pro              | P               | <u>P</u> roline           |
| Ser              | S               | <u>S</u> erine            |
| Thr              | T               | <u>T</u> hreonine         |
| Trp              | W               | Tryptophan                |
| Tyr              | Y               | <u>T</u> yrosine          |
| Val              | V               | <u>V</u> aline            |

# Ensembl Genomic Browser



[BLAST/BLAT](#) | [VEP](#) | [Tools](#) | [BioMart](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)
Login/Register

**Zebrafish (GRCz11)**
Search Zebrafish...

Location: 14:34,490,445-34,494,899    Gene: [wnt8a](#)    Transcript: [wnt8a-201](#)

- Transcript-based displays**
- Summary
  - Sequence
    - Exons
    - cDNA
    - Protein
  - Protein Information
    - Protein summary**
    - Domains & features
    - Variants
  - Genetic Variation
    - Variant table
    - Variant image
    - Population comparison
    - Comparison image
  - External References
    - General identifiers
    - Oligo probes
  - Supporting evidence
  - ID History
    - Transcript history
    - Protein history

**Transcript: wnt8a-201** ENSDART00000132193.3

**Description** wingless-type MMTV integration site family, member 8a [Source:ZFIN;Acc:[ZDB-GENE-980526-332](#)]

**Gene Synonyms** etID309727.14, wnt8, wnt8 ORF1, wnt8 ORF2, wnt8.1, wu:fa20e02, wu:fe05d07

**Location** [Chromosome 14: 34,490,445-34,494,899](#) forward strand.

**About this transcript** This transcript has [7 exons](#), is annotated with [18 domains and features](#), is associated with [261 variations](#) and maps to [228 oligo probes](#).

**Gene** This transcript is a product of gene [ENSDARG00000052910](#) Hide transcript table

Show/hide columns (1 hidden) Filter

| Name                      | Transcript ID                        | bp   | Protein               | Biotype        | UniProt                | RefSeq   | Flags             |
|---------------------------|--------------------------------------|------|-----------------------|----------------|------------------------|--|-------------------|
| <a href="#">wnt8a-201</a> | <a href="#">ENSDART00000132193.3</a> | 1680 | <a href="#">387aa</a> | Protein coding | <a href="#">B8A6B9</a> | -  | APPRIS P1         |
| <a href="#">wnt8a-202</a> | <a href="#">ENSDART00000144301.2</a> | 884  | <a href="#">263aa</a> | Protein coding | <a href="#">B8A6C1</a> | <a href="#">NM_130946</a><br><a href="#">NP_571021</a> | CDS 3' incomplete |
| <a href="#">wnt8a-203</a> | <a href="#">ENSDART00000148044.3</a> | 554  | <a href="#">161aa</a> | Protein coding | <a href="#">B8A6C0</a> | -  | CDS 3' incomplete |

# Ensembl Genomic Browser



- Domains & features
- Variants
- Genetic Variation
  - Variant table
  - Variant image
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history

- Configure this page
- Custom tracks
- Export data
- Share this page
- Bookmark this page

## About this transcript

This transcript has [7 exons](#), is annotated with [18 domains and features](#), is associated with [261 variations](#) and maps to [228 oligo probes](#).

## Gene

This transcript is a product of gene [ENSDARG00000052910](#) [Hide transcript table](#)

| Name      | Transcript ID                        | bp   | Protein               | Biotype        | UniProt                | RefSeq   | Flags             |
|-----------|--------------------------------------|------|-----------------------|----------------|------------------------|--|-------------------|
| wnt8a-201 | <a href="#">ENSDART00000132193.3</a> | 1680 | <a href="#">387aa</a> | Protein coding | <a href="#">B8A6B9</a> | -  | APPRIS P1         |
| wnt8a-202 | <a href="#">ENSDART00000144301.2</a> | 884  | <a href="#">263aa</a> | Protein coding | <a href="#">B8A6C1</a> | <a href="#">NM_130946</a><br><a href="#">NP_571021</a> | CDS 3' incomplete |
| wnt8a-203 | <a href="#">ENSDART00000148044.3</a> | 554  | <a href="#">161aa</a> | Protein coding | <a href="#">B8A6C0</a> | -  | CDS 3' incomplete |

## Protein sequence

- Download sequence
- BLAST this sequence

Exons An exon Another exon Residue overlap splice site

Markup loaded

- Variants are filtered by consequence type

```
MDHLCSKCLCFASQEEGRIEQSLVTHAAMNPCQIFASLVMSICCHILSSTAWSVNNFLMT  
GPKAYLAYTSSVQAGAQSGLIEECKHQFAWDRWNCPEALQLSTHKGLRSATRETAQVHAI  
SAAGVMYTLTKNCMGMDFENCGDDSKI GKMGGRGVWVGGCSDNVNFGDRIAKLFVDALE  
NGHDSRAAVNLHNEAGRLAVKATLKRTCKCHLSGSCSIQTCWMLADFRDIGSYLKI K  
HDQARKLEMDKIRMRAGNSADNRGAIADTFSAVARTELI FMEDSPDYCVKNLSMGLHGTE  
GRECLQSGKNLSQWERRSCRRLCHEGLKVEERRIETVWSCNCKFWCC2TKVCECTQTQT  
TRYFCAKRHRNRRPHNHSRKRQHTRRG
```