

Protein sequence analysis- Hydropathy plot, pI/Mw detection

Protein sequence analysis- Protein sequence

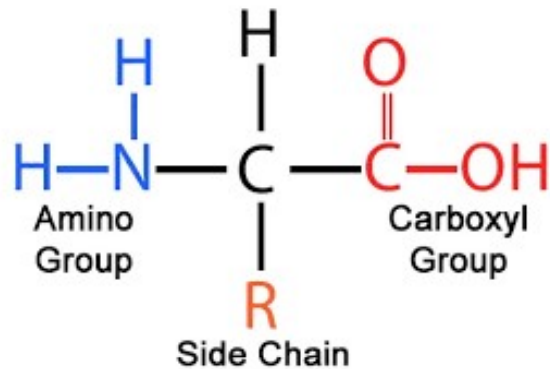
- What is a protein?
- Protein is a macronutrient that is essential to building muscle mass. It is commonly found in animal products, though is also present in other sources, such as nuts and legumes.
- What protein is made of?
- Proteins are made up of hundreds or thousands of smaller units called **amino acids**, which are attached to one another in long chains. There are 20 different types of **amino acids** that can be combined to make a protein.

List of Amino Acids, Abbreviations, and Formula

Amino Acid	Abbreviation	Single letter abbreviation	Linear formula of R-group
Alanine	Ala	A	CH ₃
Arginine	Arg	R	HN=C(NH ₂)-NH-(CH ₂) ₃
Asparagine	Asn	N	H ₂ N-CO-CH ₂
Aspartic acid	Asp	D	HOOC-CH ₂
Cysteine	Cys	C	HS-CH ₂
Glutamine	Gln	Q	H ₂ N-CO-(CH ₂) ₂
Glutamic acid	Glu	E	HOOC-(CH ₂) ₂
Glycine	Gly	G	H
Histidine	His	H	NH-CH=N-CH=C-CH ₂
Isoleucine	Ile	I	CH ₃ -CH ₂ -CH(CH ₃)
Leucine	Leu	L	(CH ₃) ₂ -CH-CH ₂
Lysine	Lys	K	HN ₂ -(CH ₂) ₄
Methionine	Met	M	CH ₃ -S-(CH ₂) ₂
Phenylalanine	Phe	F	Ph-CH ₂
Proline	Pro	P	(CH ₂) ₃
Serine	Ser	S	HO-CH ₂
Threonine	Thr	T	CH ₃ -CH(OH)
Tryptophan	Trp	W	Ph-NH-CH=C-CH ₂
Tyrosine	Tyr	Y	HO-Ph-CH ₂
Valine	Val	V	(CH ₃) ₂ -CH

Amino Acids-Properties

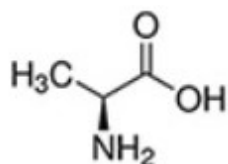
- Each amino acid is structured from an amino group and a carboxyl group bound to a tetrahedral carbon.



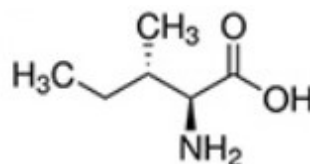
- This carbon is designated as the α -carbon (alpha-carbon).
- Amino acids differ from each other with respect to their side chains, which are referred to as R groups.
- The R group for each of the amino acids will differ in structure, electrical charge, and polarity.

Amino Acids- Groups-based on side chains

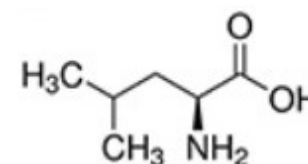
- **Aliphatic amino acids** : An *aliphatic amino acid* is an *amino acid* containing an *aliphatic side chain functional group*. *Aliphatic amino acids* are non-polar and hydrophobic.



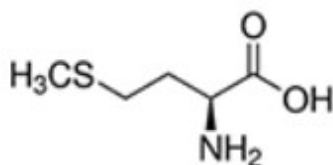
Alanine, Ala, A



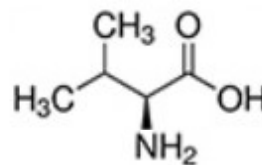
Isoleucine, Ile, I



Leucine, Leu, L

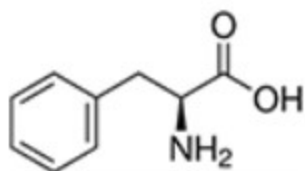


Methionine, Met, M

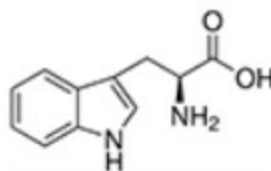


Valine, Val, V

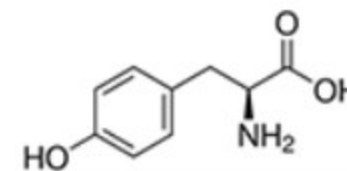
- **Aromatic Amino acids**: An *aromatic amino acid* is an *amino acid* that includes an *aromatic ring*



Phenylalanine, Phe, F



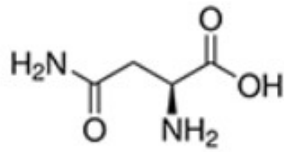
Tryptophan, Trp, W



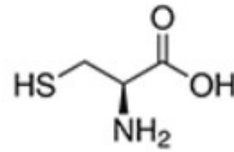
Tyrosine, Tyr, Y

Amino Acids- Groups-based on side chains

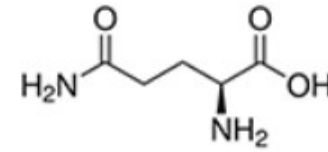
- Polar amino acids :** *Polar amino acids* are those with side-chains that prefer to reside in an aqueous (i.e. water) environment.



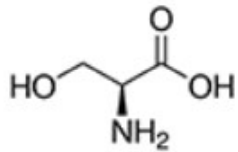
Asparagine, Asn, N



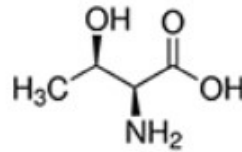
Cysteine, Cys, C



Glutamine, Gln, Q

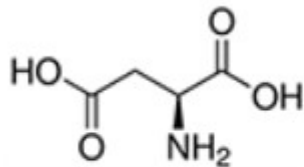


Serine, Ser, S

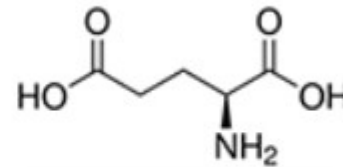


Threonine, Thr, T

- Acidic amino acids:** Those whose side chains can carry a acidic group.



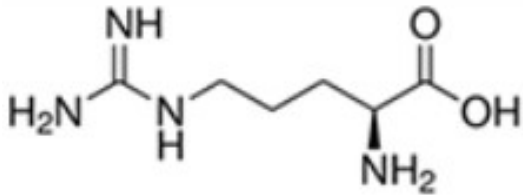
Aspartic acid, Asp, D



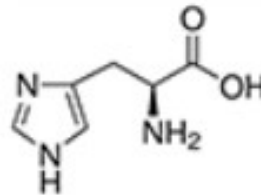
Glutamic acid, Glu, E

Amino Acids- Groups-based on side chains

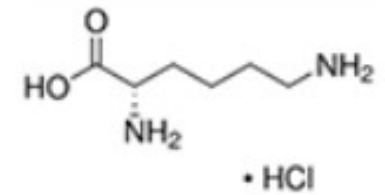
- **Basic amino acids** : *Polar amino acids* are those with side-chains that prefer to have acidic group



Arginine, Arg, R

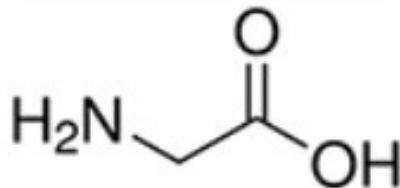


Histidine, His, H

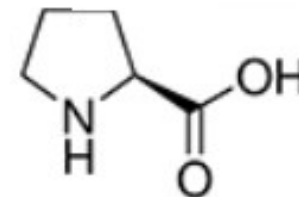


Lysine, Lys, K

- **Unique amino acids**: Proline and Glycine
- Proline is unique in that it is the only amino acid where the side chain is connected to the protein backbone *twice*, forming a five-membered nitrogen-containing ring.
- Strictly speaking, this makes Proline an *imino* acid.
- Glycine is a very unique amino acid in that it contains a hydrogen as its side chain (rather than a carbon as is the case in all other amino acids).



Glycine, Gly, G



Proline, Pro, P

Hydropathy Value or Hydropathy Index

- The **hydropathy** index of an amino acid is a number representing the hydrophobic or hydrophilic properties of its sidechain.
- The **larger** the number is, the more hydrophobic the amino acid.
- In 1982 Jack Kyte and Russell F. Doolittle proposed this theory and gave hydropathy values for all 20 amino acids as follows.

Ala: 1.800	Ile: 4.500
Arg: -4.500	Leu: 3.800
Asn: -3.500	Lys: -3.900
Asp: -3.500	Met: 1.900
Cys: 2.500	Phe: 2.800
Gln: -3.500	Pro: -1.600
Glu: -3.500	Ser: -0.800
Gly: -0.400	Thr: -0.700
His: -3.200	Trp: -0.900
	Tyr: -1.300
	Val: 4.200

>Protein

YLEINVKENLLTVQIKEGRNLIPMDPNGLSDPYVKVKLIPDDKD
QSKKKTRTIKACLNPVWNETLTYDLKPEDKDRRILIEVWDWDR
TSRNDFMGALSFGISEIKNPTNGWF

Hydropathy value for above sequence is -0.712

We get Hydropathy value of the above protein sequence
By adding the hydropathy values given in the slide.

pl or Iso-electric point of the protein

- **Isoelectric point:** *Isoelectric point, also called the **pl** of the **protein**, is the pH at which the net charge of the **protein** is zero.*
- In order to determine the isoelectric point a given protein, we must follow a general rule that consists of two steps
 - (1) Estimate the pH value at which the protein will have a net charge of zero
 - (2) Determine the pKa value right above and right below the estimated pH and find their average.

This corresponds to the isoelectric point (pl value) of the protein.

Molecular Weight (Mw) of the protein

- The **molecular weight (mw)** of an oligopeptide or a **protein** can be determined by summation of the **molecular weights** of its corresponding amino acid sequence.

Amino acid	Symbol	Residue Mass, amu (monoisotopic)
Glycine	G	57.02
Alanine	A	71.04
Serine	S	87.03
Proline	P	97.05
Valine	V	99.07
Threonine	T	101.05
Cysteine	C	103.01
Isoleucine	I	113.08
Leucine	L	113.08
Asparagine	N	114.04
Aspartic Acid	D	115.03
Glutamine	Q	128.06
Lysine	K	128.09
Glutamic Acid	E	129.04
Methionine	M	131.04
Histidine	H	137.06
Phenylalanine	F	147.07
Arginine	R	156.10
Tyrosine	Y	163.06
Tryptophan	W	186.08

>Protein

YLEINVKENLLTVQIKEGRNLIPMDPNGLSDPYVKVKLIPDDKD
 QSKKKTRTIKACLNPVWNETLTYDLKPEDKDRRILIEVWDWDR
 TSRNDFMGALSFGISEIKNPTNGWF

Molecular weight: 13847.92

We get Molecular weight of the above protein sequence
 By adding the individual residue mass of the amino acids
 given in the slide.

For detecting the Hydropathy, pI and Molecular weight of the protein there are few computational Tools such as

- protparam(<https://web.expasy.org/cgi-bin/protparam/protparam>)
- Pepstats(https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/)