



INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD

BIOINFORMATICS

Lab 1

February 16, 2024

FASTA Sequence

A FASTA sequence is a text-based format for representing biological sequences, such as DNA, RNA, or protein sequences. It is widely used in bioinformatics for storing and exchanging sequence data.

FASTA sequences are often used for

- Storing and sharing sequence data
- Performing sequence similarity searches
- Aligning sequences

FASTA Sequence of Protein CX43

The single-letter amino acid code of the protein CX43 is

```
sp|P17302|CXA1_HUMAN Gap junction alpha-1 protein OS=Homo sapiens
MGDWSALGKLLDKVQAYSTAGGKVWLSVLFIFRILLTGTAVESAWGDEQSAFRCNTQQPG
CENVCYDKSFPISHVRFWVLQIIFVSVPTLLYLAHVYVMRKEEKLNKKEEELKVAQTDG
VNVDMHLKQIEIKKFKYGIEEHGKVKMRGGLLRTYIISILFKSIFEVAFLLIQWYIYGFS
LSAVYTCKRDPCPHQVDCFLSRPTEKTIFIIIFMLVVSLVSLALNIELFYVFFKGVKDRV
KGKSDPYHATSGALSPAKDCGSQKYAYFNGCSSPTAPLSPMSPPGYKLVTGDRNNSCRN
YNKQASEQNWANYSAEQNRMGQAGSTISNSHAQPFDFPDDNQNSKKLAAGHELQPLAIVD
QRPSSRASSRASSRPRPDDLEI
```

A FASTA sequence consists of two parts:

- **Header line:** This line starts with a greater-than symbol (>) This identifier can be alphanumeric and includes information such as the organism, gene
- **Sequence data:** This is the sequence of nucleotides or amino acids, represented by single-letter codes

1 *Corresponding Amino Acids*

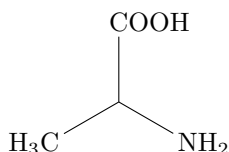
- Alanine
- Arginine
- Asparagine
- Aspartic Acid
- Cysteine
- Glutamine
- Glutamic Acid
- Glycine
- Histidine
- Isoleucine
- Leucine
- Lysine
- Methionine
- Phenylalanine
- Proline
- Serine
- Threonine
- Tryptophan
- Tyrosine
- Valine

Amino Acids

Amino acids are the building blocks of proteins, essential molecules for various biological processes.

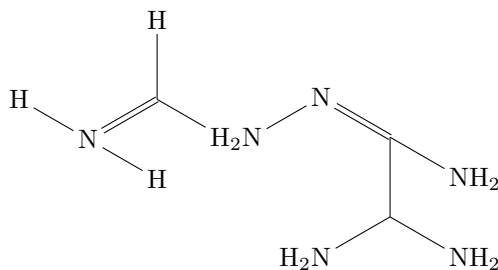
Alanine (Ala)

Alanine is a nonpolar, aliphatic amino acid. It is a non-essential amino acid, meaning the body can synthesize it. Alanine plays a crucial role in energy metabolism, serving as a substrate for gluconeogenesis. Its side chain consists of a simple methyl group, contributing to its hydrophobic nature.



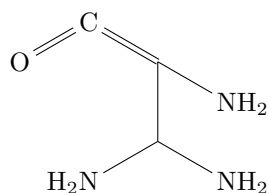
Arginine (Arg)

Arginine is a positively charged, polar amino acid. It is semi-essential, as the body may not produce enough in certain conditions. Arginine is involved in various physiological processes, including protein synthesis, immune response, and the urea cycle. Its side chain contains a guanidinium group, conferring a positive charge under physiological conditions.



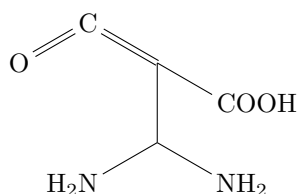
Asparagine (Asn)

Asparagine is a polar, uncharged amino acid. It is a non-essential amino acid synthesized from aspartic acid. Asparagine is crucial for protein structure and function, often found in protein loops and turns. Its side chain contains an amide group, contributing to its polar nature.



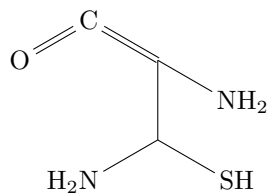
Aspartic Acid (Asp)

Aspartic Acid is a negatively charged, polar amino acid. It is a non-essential amino acid involved in the urea cycle and in the synthesis of purine nucleotides. Aspartic Acid plays a role in protein structure, particularly in ion binding. Its side chain contains a carboxyl group, conferring a negative charge under physiological conditions.



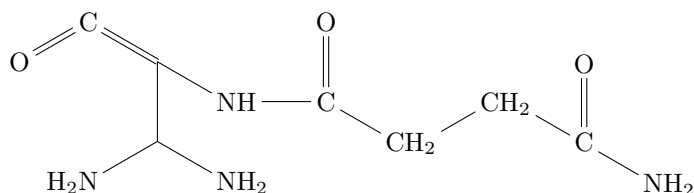
Cysteine (Cys)

Cysteine is a polar, uncharged amino acid. It is semi-essential and can be derived from methionine. Cysteine is a key component in the formation of disulfide bonds, contributing to protein structure. Its side chain contains a thiol group, allowing for the formation of covalent bonds.



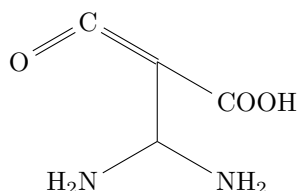
Glutamine (Gln)

Glutamine is a polar, uncharged amino acid. It is a conditionally essential amino acid, meaning the body may require more under certain conditions. Glutamine is essential for various cellular functions, including nitrogen transport and nucleotide synthesis. Its side chain contains an amide group, contributing to its polar nature.



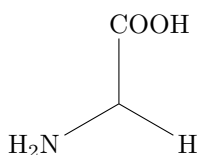
Glutamic Acid (Glu)

Glutamic Acid is a negatively charged, polar amino acid. It is a non-essential amino acid and plays a crucial role in neurotransmission and energy metabolism. Glutamic Acid is a key component of the glutamate neurotransmitter. Its side chain contains a carboxyl group, conferring a negative charge under physiological conditions.



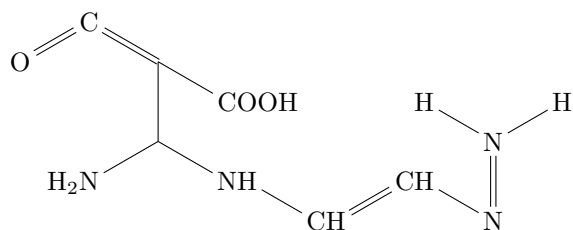
Glycine (Gly)

Glycine is the simplest amino acid, being the only one without a chiral center. It is a nonpolar, aliphatic amino acid and is the smallest of the 20 common amino acids. Glycine is essential for the synthesis of proteins and other important biomolecules. Its side chain consists of a single hydrogen atom, contributing to its simplicity.



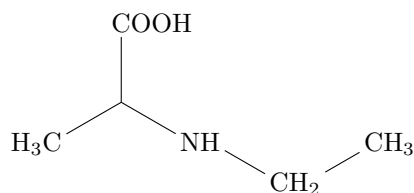
Histidine (His)

Histidine is a positively charged, polar amino acid. It is semi-essential and plays a crucial role in enzyme catalysis and as a component of hemoglobin. Histidine has an imidazole group in its side chain, allowing it to act as a buffer in biochemical reactions.



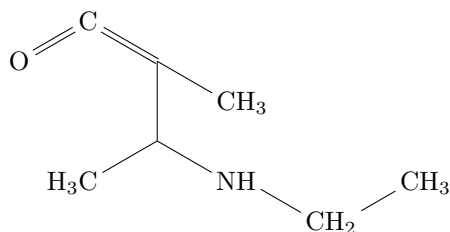
Isoleucine (Ile)

Isoleucine is a nonpolar, aliphatic amino acid. It is an essential amino acid that the body cannot synthesize, and it must be obtained through the diet. Isoleucine is important for protein synthesis and energy regulation. Its side chain consists of a branched-chain, contributing to its hydrophobic nature.



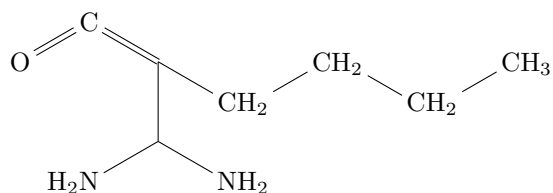
Leucine (Leu)

Leucine is a nonpolar, aliphatic amino acid. It is an essential amino acid that the body cannot produce, and it must be obtained through the diet. Leucine is important for protein synthesis, muscle growth, and energy regulation. Its side chain consists of a branched-chain, contributing to its hydrophobic nature.



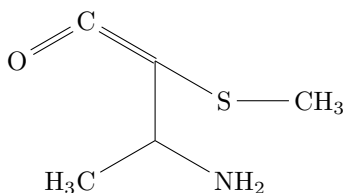
Lysine (Lys)

Lysine is a positively charged, polar amino acid. It is an essential amino acid involved in protein synthesis, collagen formation, and calcium absorption. Lysine plays a crucial role in enzyme activity and hormone production. Its side chain contains an amino group, conferring a positive charge under physiological conditions.



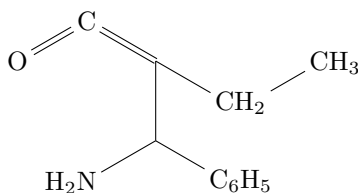
Methionine (Met)

Methionine is a nonpolar, aliphatic amino acid. It is an essential amino acid that plays a crucial role in protein synthesis and is a precursor for various important molecules. Methionine contains a sulfur-containing thioether group in its side chain.



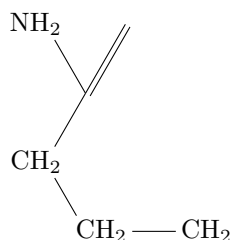
Phenylalanine (Phe)

Phenylalanine is an aromatic amino acid. It is an essential amino acid and serves as a precursor for the synthesis of tyrosine and various neurotransmitters. Phenylalanine has a phenyl ring in its side chain, contributing to its aromatic nature.



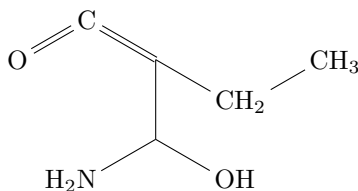
Proline (Pro)

Proline is a nonpolar, aliphatic amino acid. It is unique among amino acids because its side chain forms a cyclic structure. Proline is involved in protein structure and is often found in turns and loops. Its side chain forms a pyrrolidine ring.



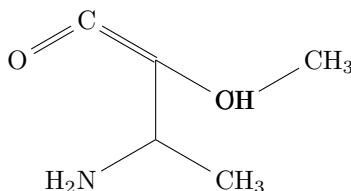
Serine (Ser)

Serine is a polar, uncharged amino acid. It is a non-essential amino acid that plays a crucial role in various biological processes, including the synthesis of proteins and neurotransmitters. Serine contains a hydroxyl group in its side chain, contributing to its polar nature.



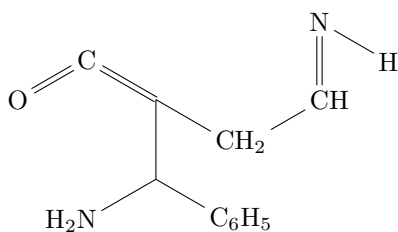
Threonine (Thr)

Threonine is a polar, uncharged amino acid. It is an essential amino acid that the body cannot produce and must be obtained through the diet. Threonine is important for protein synthesis and is a precursor for the synthesis of other amino acids. It contains a hydroxyl group in its side chain.



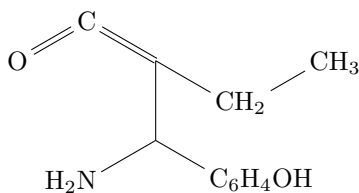
Tryptophan (Trp)

Tryptophan is an aromatic amino acid. It is an essential amino acid and is a precursor for the synthesis of serotonin and melatonin. Tryptophan has an indole ring in its side chain, contributing to its aromatic nature.



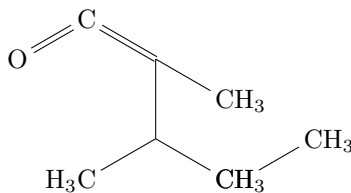
Tyrosine (Tyr)

Tyrosine is an aromatic amino acid. It is a conditionally essential amino acid and is a precursor for the synthesis of neurotransmitters and thyroid hormones. Tyrosine has a phenolic hydroxyl group in its side chain, contributing to its aromatic nature.



Valine (Val)

Valine is a nonpolar, aliphatic amino acid. It is an essential amino acid that the body cannot produce and must be obtained through the diet. Valine is important for protein synthesis and is a branched-chain amino acid. Its side chain consists of a branched-chain, contributing to its hydrophobic nature.



Basic Amino Acids

- Lysine (K)
- Arginine (R)
- Histidine (H)

Neutral Amino Acids

- Methionine (M)
- Glycine (G)
- Alanine (A)
- Leucine (L)
- Glutamine (Q)
- Tyrosine (Y)
- Threonine (T)
- Asparagine (N)
- Phenylalanine (F)
- Cysteine (C)
- Isoleucine (I)
- Proline (P)
- Serine (S)
- Tryptophan (W)
- Valine (V)

Acidic Amino Acids

- Aspartic Acid (D)
- Glutamic Acid (E)

Aromatic Amino Acids

- Tryptophan (W)
- Tyrosine (Y)
- Phenylalanine (F)

Non-aromatic Amino Acids

- Methionine (M)
- Glycine (G)
- Alanine (A)
- Leucine (L)
- Glutamine (Q)
- Threonine (T)
- Asparagine (N)
- Serine (S)
- Cysteine (C)
- Proline (P)
- Isoleucine (I)
- Valine (V)
- Histidine (H)
- Lysine (K)
- Arginine (R)
- Aspartic Acid (D)
- Glutamic Acid (E)

Frequency of Amino Acid

```
1 amino_acids = [  
2     "A", "R", "N", "D", "C",  
3     "Q", "E", "G", "H", "I",  
4     "L", "K", "M", "F", "P",  
5     "S", "T", "W", "Y", "V"  
6 ]  
7  
8 protein = "<FASTA SEQUENCE>"  
9  
10 amino_acid_counts = {aa: 0 for aa in amino_acids}  
11  
12 # Count the occurrences of each amino acid in the protein  
13 for amino_acid in amino_acids:  
14     count = protein.count(amino_acid)  
15     amino_acid_counts[amino_acid] = count  
16  
17 for amino_acid, count in amino_acid_counts.items():  
18     print(f"{amino_acid}: {count}")
```

Listing 1: Code for frequency of each Amino Acid

Amino Acid	Count
Alanine (A)	26
Arginine (R)	17
Asparagine (N)	16
Aspartic Acid (D)	18
Cysteine (C)	9
Glutamine (Q)	18
Glutamic Acid (E)	18
Glycine (G)	23
Histidine (H)	8
Isoleucine (I)	22
Leucine (L)	34
Lysine (K)	28
Methionine (M)	7
Phenylalanine (F)	21
Proline (P)	19
Serine (S)	37
Threonine (T)	13
Tryptophan (W)	6
Tyrosine (Y)	16
Valine (V)	26
Total	382

2 Tissue protein is expressed in

Connexin 43 (CX43) is a protein that forms gap junctions, allowing direct communication between neighboring cells. It is expressed in various tissues throughout the body, including:

1. **Heart:** CX43 plays a crucial role in the electrical coupling of cardiac cells, facilitating synchronized contraction.
2. **Brain:** CX43 is present in astrocytes, contributing to intercellular communication in the central nervous system.
3. **Bone:** CX43 is expressed in osteocytes, participating in the regulation of bone remodeling and mineralization.
4. **Skin:** CX43 is found in epidermal cells and plays a role in wound healing and skin homeostasis.
5. **Kidneys:** CX43 is expressed in renal cells, contributing to the coordination of tubular function.
6. **Liver:** CX43 is present in hepatocytes, contributing to liver function and regeneration.