Molecular Biology

Natarajan Meghanathan
Associate Professor of Computer Science
Jackson State University
E-mail: natarajan.meghanathan@jsums.edu
Introduction

- Our body consists of a number of organs.
- Each organ is composed of a number of tissues.
- Each tissue is a collection of similar cells that group together to perform similar functions.
- The individual cell is the minimal self-reproducing unit in all living species.
  - A cell performs two types of functions:
    - (i) Stores and passes the genetic information for maintaining life from generation to generation
      - the cells store the genetic information in the form of double-stranded DNA
    - (ii) Performs chemical reactions necessary to maintain our life.
      - Portions of the DNA called genes are transcribed into closely related molecules called RNAs that guide the synthesis of proteins.
- Proteins are the main catalysts for all the chemical reactions in the cell; they are also involved in transportation, signaling, cell-membrane formation, etc.
- We will look in detail about proteins, DNA and RNA.
Chromosome and Gene

• In most multi-cell organisms, every cell contains the same complete set of genome. (May have some small difference due to mutation)
• The total information stored in all chromosomes constitute a genome.
• Example: Human Genome: has 3G base pairs, organized in 23 pairs of chromosomes.
• Each chromosome is composed of two complementary DNA strands (each strand is a polynucleotide chain).
• A gene is a sequence of DNA that encodes a protein or an RNA molecule.
• In human genome, it is expected there are 30,000 – 35,000 genes.
• For gene that encodes protein,
  – In Prokaryotic genome, one gene corresponds to one protein
  – In Eukaryotic genome, one gene can correspond to more than one protein because of the process “alternative splicing”
DNA

- DNA stores the instruction needed by the cell to perform daily life function.
- It consists of two strands which interwoven together and form a double helix.
  - Each strand is a chain of some small molecules called nucleotides.
- A gene is a sequence of nucleotides encoding specific information.
- DNA can be thought of as a large cookbook with recipes for making every protein in the cell.
- The information in the DNA is used like a library.
  - The info in the gene is read, perhaps a million times during the life of a cell; but the DNA itself is never used up.
- Nucleotide consists of three parts:
  - Deoxyribose
  - Phosphate (bound to the 5’ carbon)
  - Base (bound to the 1’ carbon)
Nucleotide Structure and Phosphates

- Each nucleotide can have one, two or three phosphates.
  - Monophosphate nucleotides have only one phosphate group, and are the building blocks of DNA.
  - Diphosphate and Triphosphate nucleotides (ADP and ATP) have 2 and 3 phosphate groups respectively; used to transport energy in the cell.
  - ATP is like a battery. Whenever power is required, ATP loses one phosphate group to release energy and form ADP.
  - By using the energy in the food, mitochondria in the cell converts ADP back to ATP.
Nucleotide for RNA

- Nucleotide consists of three parts:
  - Ribose Sugar (has an extra OH group at 2’)
  - Phosphate (bound to the 5’ carbon)
  - Base (bound to the 1’ carbon)
Nucleotide Bases

• There are 5 different nucleotides: adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U).
  – A, G are called purines. They have a 2-ring structure.
  – C, T, U are called pyrimidines. They have a 1-ring structure.
• DNA only uses A, C, G, and T.

The CH3 methyl group in Thymine stabilizes the DNA and helps in bonding with other bases.
Nucleotide Pairing: Watson-Crick Rules

- Complementary bases:
  - A pairs with T (two hydrogen bonds)
  - C pairs with G (three hydrogen bonds)
DNA: Polynucleotide Chain

- A strand of DNA is formed by chaining together the nucleotides, forming a sugar-phosphate backbone.

- The sugar-phosphate bond connects the phosphate (at the 5’ carbon of one nucleotide) to the sugar (at the 3’ carbon) of another nucleotide.

Adenine

Cytosine
DNA: Polynucleotide Chain

- A strand of DNA is formed by chaining together the nucleotides, forming a sugar-phosphate backbone.
- The sugar-phosphate bond connects the phosphate (at the 5’ carbon of one nucleotide) to the sugar (at the 3’ carbon) of another nucleotide.
The polynucleotide chain has direction:
Conventionally, we write the sequence from the 5’ end to the 3’ end (a.k.a. the upstream)
Hence, the above sequence is written ACGTA

Downstream: 3’ end to the 5’ end (for the complementary strand)
Double Stranded DNA

- DNA is double stranded within a cell. The two strands are anti-parallel. One strand is the reverse complement of the other.
- The double strands are interwoven together and form a double helix.
- One reason for the double strand for DNA is to facilitate replication.
Forms and Locations of DNA

• **Forms**
  - DNA usually exists in linear form (when located in the nucleus).
    - E.g., in human, yeast
  - In some simple organisms (like E. coli) and in mitochondria and chloroplasts (only in plants), DNA exists in circular form.

• **Locations**
  - In prokaryotes (single-celled organisms with no nuclei, e.g., bacteria), DNA swims within the cell.
  - In eukaryotes (organisms with single or multiple cells, each of which have a nucleus; e.g., plant and animal), DNA is located within the nucleus.
Complexity of the Organism vs. Genome Size vs. Number of Genes

- Genome Size referred in terms of the number of base pairs.
- Human Genome: 3G base pairs
- Amoeba Dubia (a single cell organism): 670G base pairs.
- Genome size has no relationship with the complexity of the organism.

- Prokaryotic genome: E.g., E. coli
  - # base pairs: 5M  # genes: 4K
  - Avg. length of a gene: 1000 base pairs (bp)
  - This implies on average 80% of the E. coli genome consists of useful information (typically called coding regions)

- Eukaryotic genome: E.g., Human
  - # base pairs: 3G  # genes: 20K – 30K
  - Avg. length of a gene: 1000 – 2000 bp
  - This implies less than 3% of human genome is believed to be coding regions. The rest is called junk DNA
  - There is no relationship between genome size and length of coding region (the sequence of nucleotides in the DNA corresponding to a gene).
RNA (Ribo Nucleic Acid)

• RNA is the nucleic acid produced during the transcription process (i.e., from DNA to RNA)
• However, in certain organisms like viruses, RNA is used as genetic material instead of DNA

• RNA has both the properties of DNA and protein
  – Similar to DNA, it can store and transfer information. But, RNA is not as stable as DNA.
  – Similar to protein, it can form complex 3-dimensional structure and perform some functions.
    • But, RNA is not as versatile as protein (that exhibit several functionalities)
RNA vs. DNA

• RNA is single stranded.

• The nucleotides of RNA are quite similar to that of DNA, except that it has an extra OH at position 2’.
  – Due to this extra OH, it can form more hydrogen bonds than DNA. Thus, RNA can form complex 3-dimensional structure.

• RNA use the base U instead of T.
  – U is chemically similar to T. In particular, U is also complementary to A.
Different Types of RNA

- Two broad types: Messenger RNAs (mRNAs) and Non-coding RNAs
- mRNAs: carry the encoded info required to make proteins of all types (DNA transcribed to mRNAs)
- Non-coding RNAs
  - transfer RNA (tRNA): serve as a molecular dictionary used to translate the nucleic acid code to the corresponding amino acid code for the protein
  - ribosomal RNA (rRNA): form part of ribosomes that help to translate mRNAs to proteins.
  - Short ncRNAs help to regulate the process of generating proteins from genes.
  - Long ncRNAs: insignificant; functionality not known
### 20 Different Amino Acids

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>1-Letter</th>
<th>3-Letter</th>
<th>Avg. Mass (Da)</th>
<th>volume</th>
<th>Side chain polarity</th>
<th>Side chain acidity or basicity</th>
<th>Hydropathy index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>A</td>
<td>Ala</td>
<td>89.09404</td>
<td>67</td>
<td>non-polar</td>
<td>Neutral</td>
<td>1.8</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C</td>
<td>Cys</td>
<td>121.15404</td>
<td>86</td>
<td>polar</td>
<td>basic (strongly)</td>
<td>-4.5</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>D</td>
<td>Asp</td>
<td>133.10384</td>
<td>91</td>
<td>polar</td>
<td>Neutral</td>
<td>-3.5</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>E</td>
<td>Glu</td>
<td>147.13074</td>
<td>109</td>
<td>polar</td>
<td>acidic</td>
<td>-3.5</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
<td>Phe</td>
<td>165.19184</td>
<td>135</td>
<td>polar</td>
<td>neutral</td>
<td>2.5</td>
</tr>
<tr>
<td>Glycine</td>
<td>G</td>
<td>Gly</td>
<td>75.06714</td>
<td>48</td>
<td>polar</td>
<td>acidic</td>
<td>-3.5</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
<td>His</td>
<td>155.15634</td>
<td>118</td>
<td>polar</td>
<td>neutral</td>
<td>-3.5</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
<td>Ile</td>
<td>131.17464</td>
<td>124</td>
<td>non-polar</td>
<td>neutral</td>
<td>-0.4</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
<td>Lys</td>
<td>146.18934</td>
<td>135</td>
<td>polar</td>
<td>basic (weakly)</td>
<td>-3.2</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
<td>Leu</td>
<td>131.17464</td>
<td>124</td>
<td>non-polar</td>
<td>neutral</td>
<td>4.5</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
<td>Met</td>
<td>149.20784</td>
<td>124</td>
<td>non-polar</td>
<td>neutral</td>
<td>3.8</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>Asn</td>
<td>132.11904</td>
<td>96</td>
<td>polar</td>
<td>basic</td>
<td>-3.9</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
<td>Pro</td>
<td>115.13194</td>
<td>90</td>
<td>non-polar</td>
<td>neutral</td>
<td>1.9</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
<td>Gln</td>
<td>146.14594</td>
<td>114</td>
<td>non-polar</td>
<td>neutral</td>
<td>2.8</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
<td>Arg</td>
<td>174.20274</td>
<td>148</td>
<td>non-polar</td>
<td>neutral</td>
<td>-1.6</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
<td>Ser</td>
<td>105.09344</td>
<td>73</td>
<td>polar</td>
<td>neutral</td>
<td>-0.8</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
<td>Thr</td>
<td>119.12034</td>
<td>93</td>
<td>polar</td>
<td>neutral</td>
<td>-0.7</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
<td>Val</td>
<td>117.14784</td>
<td>105</td>
<td>non-polar</td>
<td>neutral</td>
<td>-0.9</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
<td>Trp</td>
<td>204.22844</td>
<td>163</td>
<td>polar</td>
<td>neutral</td>
<td>-1.3</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
<td>Tyr</td>
<td>181.19124</td>
<td>141</td>
<td>non-polar</td>
<td>neutral</td>
<td>4.2</td>
</tr>
</tbody>
</table>
Replication and Repair of DNA

• DNA is double stranded.

• When the cells divide,
  – DNA needs to be duplicated and passed on to the two daughter cells.
  – With the help of RNA polymerase, the two strands of DNA serve as template for the synthesis of another complementary strands, generating two identical double stranded DNAs for the two daughter cells.

• When one strand is damaged,
  – it is repaired with the information of another strand.
Mutation

• Despite the near-perfect replication, infrequent unrepaired mistakes are still possible.
  – Those mistakes are called mutations.

• Mutations are changes in the DNA sequence of genes (nucleotide interchanges): can be of two types:
  – Transitions: Alternative Pyrimidines [ C $\leftrightarrow$ T ] or Purines [ A $\leftrightarrow$ G]
  – Transversions: Purine $\leftrightarrow$ Pyrimidine [ C/T $\leftrightarrow$ A/G ]

• Occasionally, some mutations make the cells or organisms survive better in the environment.
  – The selection of the fittest individuals to survive is called natural selection.

• Mutation and natural selection have resulted in the evolution of a diversified organisms.
Transitions vs. Transversions

• **Transitions** are interchanges of two-ring **purines** (A ↔ G) or of one-ring **pyrimidines** (C ↔ T): they therefore involve bases of similar shape.

• **Transversions** are interchanges of **purine** for **pyrimidine bases**, which therefore involve exchange of **one-ring** and **two-ring** structures.

• Although there are twice as many *possible* transversions, because of the molecular mechanisms by which they are generated, transition mutations are generated at higher frequency than transversions.
  – Also transversions are mostly due to externally induced (due to chemical or high-energy radiation) or due to environmental challenges (e.g., lack of a particular amino acid); whereas, transitions are more spontaneous due to transient factors (say, heat).

• As well, transitions are less likely to result in amino acid substitutions. Hence, transitions are mostly not harmful.
Central Dogma (1)

- Central Dogma describes the process of transferring information from DNA to RNA to protein.
- It states that the info from DNA is transferred sequentially to RNA to protein and that the info cannot be transferred back (to DNA).
- The information transfer occurs in two steps:
  - Transcription: DNA to mRNA: An mRNA is synthesized from a DNA template
  - Translation: mRNA to protein: the mRNA is translated into an amino acid sequence by stitching the amino acids one by one.
  - Post-translation modification: protein to modified protein.
Central Dogma (2)

Prokaryotes

DNA → transcription → translation → modification

Cytoplasm

Eukaryotes

DNA → transcription → Add 5’ cap and poly A tail → RNA splicing → export → translation → modification

Nucleus → Cytoplasm
Transcription (Prokaryotes)

• The RNA polymerase enzyme temporarily separates the double-stranded DNA.
• The enzyme locates the transcription start site (a marker denoting the start of a gene).
• The enzyme synthesizes an mRNA following two rules:
  – The bases A, C and G are copied exactly from DNA to mRNA.
  – The base T is copied as U.
• Once the enzyme reaches the transcription stop site (a marker denoting the end of a gene), the transcription process is stopped and an mRNA is synthesized.
• Note that a prokaryotic gene is completely transcribed into an mRNA by RNA polymerase.
• A particular prokaryotic gene transcribes to only one mRNA sequence.
Transcription in Eukaryotes (1)

• Different from transcription in prokaryotes.
• For an eukaryotic gene, only part of it (called the ‘exon’ segments) are finally transcribed into an mRNA.
• Introns are segments of a gene situated between the exon segments
  – An Intron starts with GT ..... and ends with AG.
• Introns can be considerably longer than the exons
  – E.g., for the gene associated with the disease cystic fibrosis in humans, there are 24 introns of total length approx. 1 million bases; whereas the total length of the exons is only 1 kilo bases.
• Splicesomes recognize the start and end of the introns in a gene and splice them out.
Transcription in Eukaryotes (2): Details

• First, the RNA polymerase produces a pre-mRNA that contains both introns and exons.

• Then, a cap is added to the 5’ end and a poly-A tail (150 to 200 Adenines) is added to the 3’ end of the pre-mRNA.
  – The poly(A) tail is important for the nuclear export, translation, and stability of mRNA. Likewise, the 5’ cap adds stability.

• After that, with the help of splicesomes, the introns are removed and an mRNA is produced.

• The final mRNA is transported out of the nucleus and the translation process starts.

• Note that in different tissues or different conditions, the introns that will be removed are not the same: Alternative splicing.
  – Some exons will be grouped with introns and spliced out.
  – Hence, one eukaryotic gene can be transcribed into many different mRNA sequences.
The methyl group in the 5’ cap adds to the stability of the mRNA
Alternative Splicing

DNA

RNA

Alternative Splicing

mRNA

Translation

Protein A

Protein B

Protein C
Translation: Protein Synthesis (1)

- The translation process comprises of generating a protein from the mRNA.
- This is facilitated through a molecular complex called the ribosome that parses the mRNA three nucleotides (called codons) at a time.
- The starting codon is always AUG;
- The ending codons are: UAA, UAG and UGA.
- For every codon read, the ribosome sends out instructions to the cytoplasm looking for a transfer RNA (tRNA) that carries a complementary anti-codon and a corresponding amino acid (for the codon read).
- The anti-codon segment of the first tRNA binds to the codon in the mRNA and the same process is repeated for the adjacent codon by a second tRNA carrying a corresponding amino acid.
  - There are 61 tRNAs, one for each amino acid.
Translation: Protein Synthesis (2)

• The first tRNA then releases its amino acid to the second tRNA and leaves the site. The two amino acids start to form a peptide chain using ATP as the energy source.
  – The binding of an amino acid to a tRNA also consumes energy (from the ATP).

• The third tRNA carrying an amino acid (corresponding to the third codon read) binds to the mRNA; the second tRNA releases its peptide chain that later binds to the amino acid on the third tRNA and the peptide chain is extended.

• This process is repeated until one of the three stop codons is encountered. At this time a polypeptide chain of the amino acids would have been generated and it is released to the cytoplasm. The linear chain of polypeptide amino acids later undergoes folding to form a 3-dimensional protein complex.

• Multiple copies of the protein complexes could be generated from the same mRNA, through a sequence of ribosomes parsing the mRNA, depending on the amount required and the energy available.

• Note that the gene picked up by the RNA polymerase (initially during transcription) depends on the protein that needs to be eventually synthesized.
tRNA

binding site for an amino acid

Anticodon

Translation: Protein Synthesis (1)

m-RNA

Ribosome

tRNAs with the anti-codon and amino acid corresponding for the codon

m-RNA

Ribosome

ATP
Translation: Protein Synthesis (2)

Polypeptide chain folding to a protein
Translation: Protein Synthesis (3)

Multiple copies of proteins are made as other ribosomes follow the first ribosome.

Genetic Code (1)

• During the translation process, the translation of a 3-nucleotide codon to an amino acid is specified by a look-up table called the “genetic code”
  – The genetic code is the same for all organisms.
• Since there are 20 amino acids, we need at least 20 unique codons. Hence, with 4 different bases (A, T/U, C, G), we need \( \log_4(20) = 3 \) nucleotides per codon.
• With 4 different bases (A, T/U, C, G), we can have \( 4^3 = 64 \) different codons, among which three codons are used to indicate the end of the translation site on an m-RNA.
  – Hence, there are \( 64 - 3 = 61 \) unique codons mapped to 20 amino acids. Thus, some amino acids are mapped to more than one codon.
• The genetic code is not a random assignment of the codons to the amino acids.
  – Amino acids that share the same biosynthetic pathway tend to have the same first base in their codons.
  – Amino acids with similar physical properties tend to have similar codons.
<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>TTT Phe [F]</td>
<td>TCT Ser [S]</td>
<td>TAT Tyr [Y]</td>
<td>TGT Cys [C]</td>
</tr>
<tr>
<td></td>
<td>TTC Phe [F]</td>
<td>TCC Ser [S]</td>
<td>TAC Tyr [Y]</td>
<td>TGC Cys [C]</td>
</tr>
<tr>
<td></td>
<td>TTA Leu [L]</td>
<td>TCA Ser [S]</td>
<td>TAA Ter [end]</td>
<td>TGA Ter [end]</td>
</tr>
<tr>
<td>C</td>
<td>CTT Leu [L]</td>
<td>CCT Pro [P]</td>
<td>CAT His [H]</td>
<td>CGT Arg [R]</td>
</tr>
<tr>
<td></td>
<td>CTC Leu [L]</td>
<td>CCC Pro [P]</td>
<td>CAC His [H]</td>
<td>CGC Arg [R]</td>
</tr>
<tr>
<td>A</td>
<td>ATT Ile [I]</td>
<td>ACT Thr [T]</td>
<td>AAT Asn [N]</td>
<td>AGT Ser [S]</td>
</tr>
<tr>
<td></td>
<td>ATC Ile [I]</td>
<td>ACC Thr [T]</td>
<td>AAC Asn [N]</td>
<td>AGC Ser [S]</td>
</tr>
<tr>
<td></td>
<td>ATA Ile [I]</td>
<td>ACA Thr [T]</td>
<td>AAA Lys [K]</td>
<td>AGA Arg [R]</td>
</tr>
</tbody>
</table>
• A gene consists of the three regions: the 5’ untranslated region, the coding region and the 3’ untranslated region.
  – These three regions are together also called the “mRNA transcript” as the transcription step exactly copies these three regions to form a mRNA.

• The coding region contains the codons for proteins.
  – It is also called the open reading frame.
  – It is composed of codons that occur as multiples of three nucleotides.
  – The coding region usually begins with a start codon, and must end with a stop codon; the rest of its codons are not stop codons.

• Before the 5’ untranslated region, we have the regulatory region (a.k.a. the promoter) which regulates the transcription process.
  – The promoter is a DNA segment to which the RNA polymerase binds initiating the transcription of mRNA.
Single Nucleotide Polymorphism (SNP)

• Given the genome of two individuals of the same species, if there exists a position (called loci) where the single nucleotides between the two individuals are different, we call it a single nucleotide polymorphism (SNP).

• For human, we expect SNPs are responsible for over 80% of the variation between two individuals.

• Hence, understanding SNPs can help us to understand the differences within a population.

• For example, in human, SNPs control the color of hair, the blood type, etc of different individual. Also, many diseases like cancer are related to SNPs.
Proteins

- Proteins constitute most of a cell’s dry mass (i.e., the non-watery portion).
- Proteins are the building blocks for cells as well as execute nearly all the cell functions.
- A protein is made from a long chain of amino acids (linked together through a covalent peptide bond).
  - The chain is called a polypeptide whose length can range from 20 to 5000 amino acids (avg. length is 350 amino acids).
  - There are 20 different amino acids – each with different properties.
- Proteins fold into 3-d shapes with the aid of the weak interactions (like hydrogen bonds, hydrophobic interactions, ionic bonds, van derWaal’s forces, etc) between the adjacent and non-adjacent amino acid residues.
  - These interactions influence the shape of a protein, which in turn is vital to its functionality.
Protein Folding

Amino acid

Polypeptide
Amino Acids

- Amino acids are the building blocks of proteins.
- Each amino acid consists of:
  - An amino group (-NH2)
  - Carboxyl group (-COOH)
  - R group (side chain: which determines the amino acid type)
- All the three groups are attached to a single carbon atom (α-Carbon)
- The R group could indicate the nature of charge distribution: basic, acidic or neutral: Basic: positively charged; Acidic: Negatively charged
- Amino acids that have basic or acidic R groups are polar (i.e., the charge distribution within the R group is uneven).
- Neutral amino acids are non-polar (overall even charge distribution).
- Hydrophilic – affinity to form hydrogen bond with water (typically, polar)
- Hydrophobic – cannot form hydrogen bond with water (typically, non-polar)
Polypeptide: Protein Sequence

- Protein or polypeptide chain is formed by joining the amino acids together via a peptide bond.

- One end of the polypeptide is the amino group, which is called N-terminus. The other end of the polypeptide is the carboxyl group, which is called C-terminus.
<table>
<thead>
<tr>
<th></th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Glycine (Gly, G)</strong></td>
<td>(H_2N-\text{C}-\text{C}-\text{OH})</td>
</tr>
<tr>
<td><strong>Histidine (His, H)</strong></td>
<td>(H_2N-\text{CH-C-CH-OH})</td>
</tr>
<tr>
<td><strong>Isoleucine (Ile, I)</strong></td>
<td>(H_2N-\text{C}-\text{C}-\text{OH})</td>
</tr>
<tr>
<td><strong>Leucine (Leu, L)</strong></td>
<td>(H_2N-\text{C}-\text{C}-\text{CH-CH-CH}_{3})</td>
</tr>
<tr>
<td><strong>Lysine (Lys, L)</strong></td>
<td>(H_2N-\text{C}-\text{C}-\text{OH})</td>
</tr>
<tr>
<td><strong>Methionine (Met, M)</strong></td>
<td>(H_2N-\text{C}-\text{C}-\text{S-CH}_{3})</td>
</tr>
<tr>
<td>Amino Acid</td>
<td>Chemical Structure</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Phenylalanine (Phe, F)</td>
<td><img src="image" alt="Phenylalanine" /></td>
</tr>
<tr>
<td>Proline (Pro, P)</td>
<td><img src="image" alt="Proline" /></td>
</tr>
<tr>
<td>Serine (Ser, S)</td>
<td><img src="image" alt="Serine" /></td>
</tr>
<tr>
<td>Threonine (Thr, T)</td>
<td><img src="image" alt="Threonine" /></td>
</tr>
<tr>
<td>Tryptophan (Trp, W)</td>
<td><img src="image" alt="Tryptophan" /></td>
</tr>
<tr>
<td>Tyrosine (Tyr, Y)</td>
<td><img src="image" alt="Tyrosine" /></td>
</tr>
<tr>
<td>Amino Acid</td>
<td>1-Letter</td>
</tr>
<tr>
<td>----------------</td>
<td>----------</td>
</tr>
<tr>
<td>Alanine</td>
<td>A</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>D</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>E</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
</tr>
<tr>
<td>Glycine</td>
<td>G</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
</tr>
<tr>
<td>Asparagnine</td>
<td>N</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
</tr>
</tbody>
</table>
Amino Acids and Protein Structure

- The hydrophilic amino acids are found on the outer surface of a folded protein
  - N, C, Q, G, S, T, Y
- The hydrophobic amino acids are found on the inner surface of a folded protein
  - A, I, L, M, F, P, W, V

- Primary structure
  - The amino acid sequence

- Secondary structure
  - The local structure formed by hydrogen bonding: α-helices and β-sheets.

- Tertiary structure
  - The interaction of α-helices and β-sheets due to hydrophobic effect

- Quaternary structure
  - The interaction of more than one protein to form protein complex