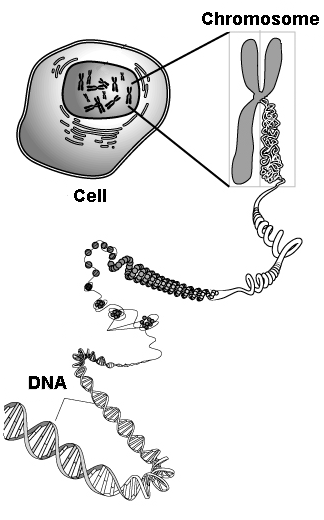
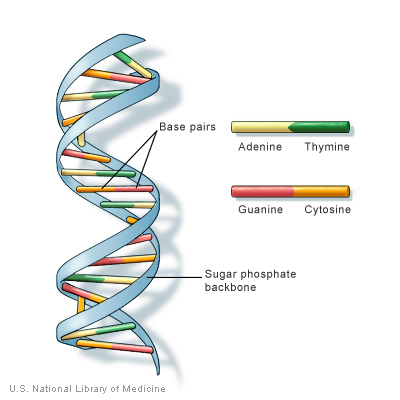
Biology Review

Proteins determine what a cell is, what it can do and how it interacts with its environment. Proteins are composed of amino acids linked together in long chains called peptides. The sequence of the amino acids in a peptide is determined by the base sequence in DNA molecules found in chromosomes in the nucleus. DNA is a double stranded helix with a phosphate-deoxyribose (a five-carbon sugar) backbone attached to four bases with adenine (A) pairing with thymine (T) and guanine (G) pairing with cytosine (C).

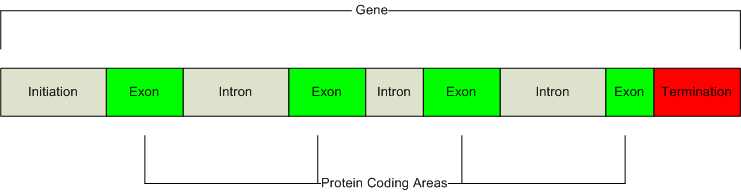
DNA structure

1. 2.



A gene is a segment of a chromosome (DNA) that contains the instructions for the amino acid sequence of a single protein. A chromosome may contain thousands of genes along its DNA length. A gene may be several hundred to several hundred thousand base pairs long.

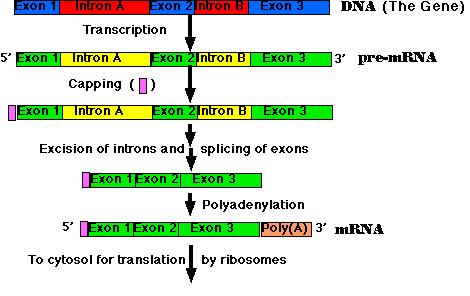
3.



In eukaryotic cells (those that have a nucleus) a gene has the elements seen above. The initiation site (often called the promoter site) regulates whether the gene is active or not. Transcription factors regulate gene activation by binding to this site. An active gene undergoes transcription—the process of making mRNA. The mRNA made from the gene contains all the elements except the initiation segment. Exons are sequences of DNA in the gene that are expressed in the final mRNA molecule. Introns are sequences that are removed during mRNA processing and do not end up as part of the final mRNA molecule. Genes may have many exons and introns—up to 40 or more—and many have more intron DNA than exon DNA.

Making the mRNA message: RNA polymerase binds to the initiation site and creates a molecule of single-stranded mRNA that contains all the elements of the DNA from which it is copied. This is called pre-mRNA. A 7-methylquanosine cap is added to the 5’ end (the 5’—“five prime”— and 3’—“three prime”— labeled ends of the mRNA are derived from the orientation of the ribose sugar molecule in the sugar-phosphate backbone of the RNA). The introns are cut out and removed, the exons are spliced together and a poly-A tail (which may contain several hundred adenine bases) is added to the 3’end.

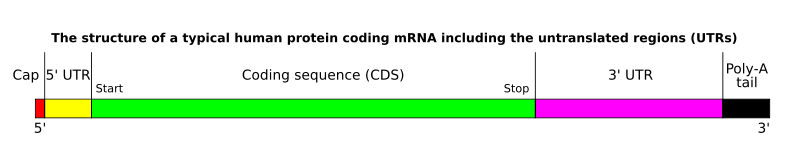
4.



The final mRNA molecule contains the following features seen below:

* The 5’ 7-methylquanosine cap (red).
* The 5’ **u**n**t**ranslated **r**egion (yellow). A UTR is a non-coding sequence of bases in the mRNA. The 5’ UTR contains the ribosome binding site.
* The coding sequence—CDS (green), which begins with the sequence AUG, the start codon. Codons are sequences three bases long and code for one amino acid. The start codon also sets the reading frame for the coding sequence—which set of bases, taken three at a time, will be used. The end of the CDS is one of three stop codons, UAA, UAG or UGA.
* The 3’ UTR (purple), which contains the poly A polymerase binding site, the enzyme that creates the poly A tail. UTRs can be hundreds of bases long.
* The poly-A tail (black), a string of adenine bases (which may be several hundred bases long). Only mRNA molecules have this feature.
* Note: uracil (U) substitutes for thymine (T) in RNA.

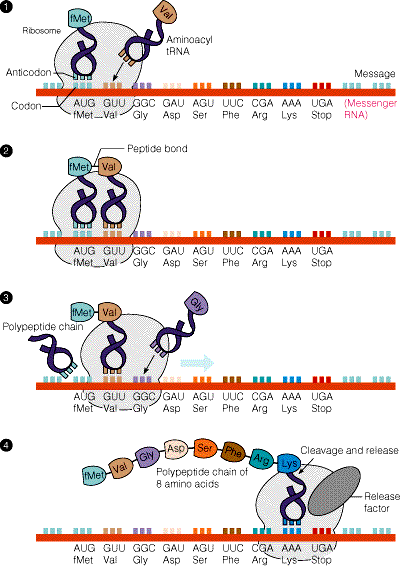
5.



The mature mRNA leaves the nucleus and binds to a ribosome in the endoplasmic reticulum.

The amino acid sequence of a protein is determined by the base sequence of the CDS of the mRNA. Starting with the start codon (AUG), the sequence of the mRNA is read three bases at a time (one codon). tRNA molecules bring in specific amino acids based on the tRNA’s anti-codon, the part of the tRNA molecule that binds to the mRNA. Each type of tRNA can bind only one specific amino acid. Amino acids form a peptide bond with each other to form the protein.

6.



Proteins fold into three dimensions based on the chemical interactions of the side chain structure of the amino acids. Peptides may be hundreds of amino acids long. A protein may be just one peptide or a combination of many peptides. Signal peptides, short sequences of amino acids at the beginning of a protein, help guide a protein to its final cellular location. Signal peptides are usually cleaved to yield the mature protein once it has reached its correct cellular destination. Some amino acids may undergo post-translational modification. Glycosylation (addition of sugar molecules) and phosphorylation (the addition of phosphate molecules) are two common modifications.

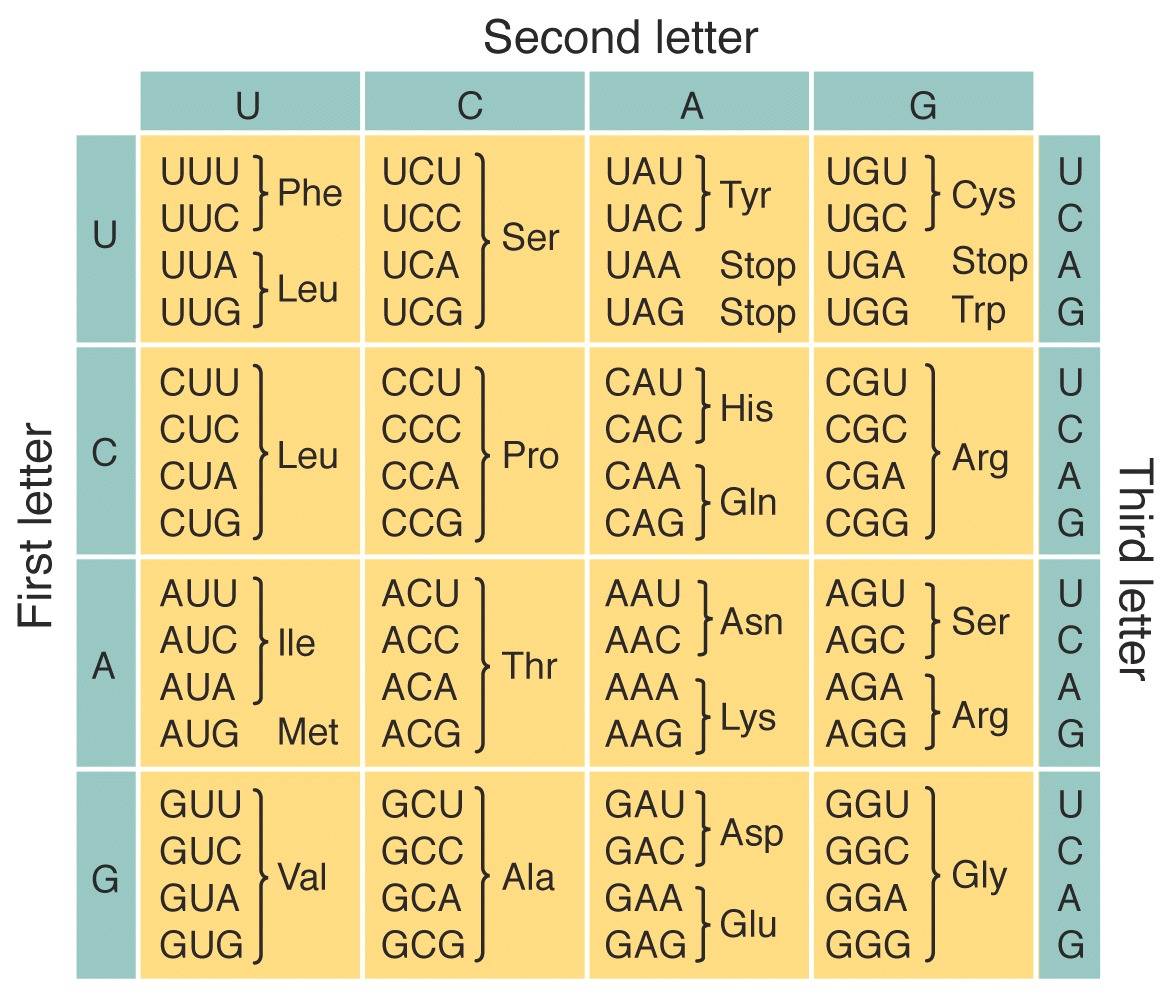
The twenty amino acids that make up proteins may have more than one codon, the sequence of bases taken three at a time. The genetic code shows these sequences. Note that AUG, the start codon, codes for the amino acid methionine.

7.

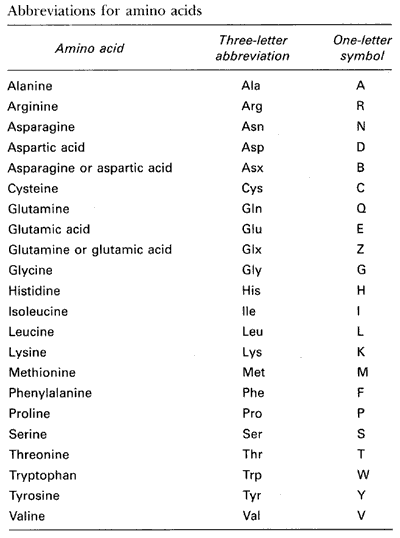
start

Readng the genetic code:

* The first base of the codon is read on the left
* The second base of the codon is read on the top
* The third base of the codon is read on the right
* Only methionine and tryptophan have one codon. All the others have redundant codes. Some organisms may use some codons preferentially over others for the same amino acid.



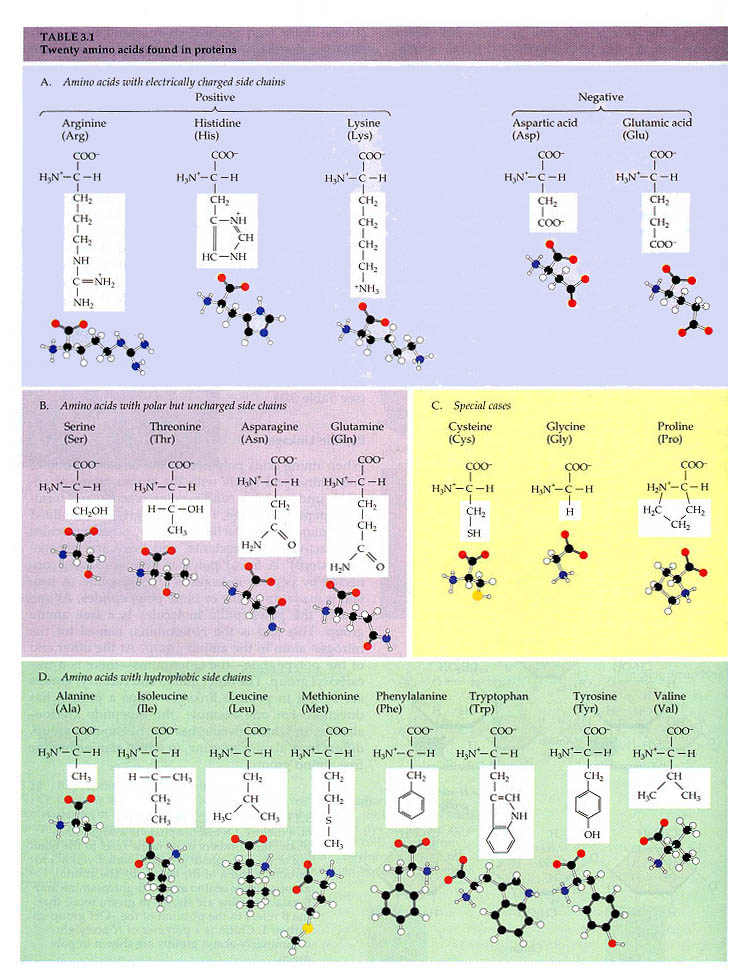
8.



Note: The one-letter amino acid symbol is used most often in biotechnology applications.

When incorporated into a peptide, the most important reactive element of an amino acid is its side chain structure. Some have positive or negative charges—some are polar—some are hydrophobic—some are hydrophilic—some are large—some are small. The sequence of amino acids in a protein and the interaction of their side chain components greatly determine the folding of the protein and its interactions with itself and its environment—what it is and what it can do.

9.

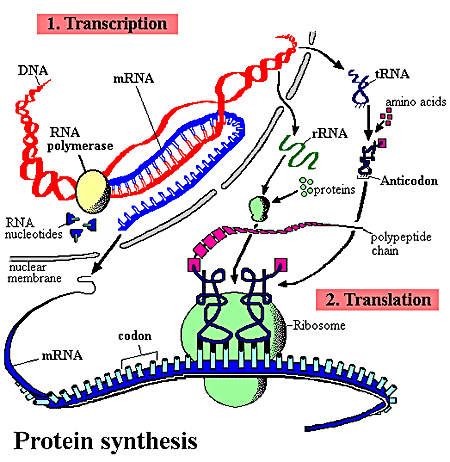


Protein synthesis put together

Transcription refers to the information in the DNA base sequence changing to the base sequence in an RNA molecule. Note: As well as the genes that code for mRNA required for the synthesis of new proteins, there are also genes that code for the RNA sequences found in tRNA and rRNA molecules. Other genes code for RNA sequences found in iRNA, molecules of RNA that inhibit the translation of mRNA, thus providing a level of regulation on translation that has only recently been recognized.

Translation refers to the information in the RNA base sequence changing to the amino acid sequence in the protein. Processing of mRNA, migrating it physically from the nucleus to the endoplasmic reticulum and binding it to the ribosome all offer points of regulation in the process. Folding a protein into its final and functional form is aided by chaperon proteins (not shown here).

10.

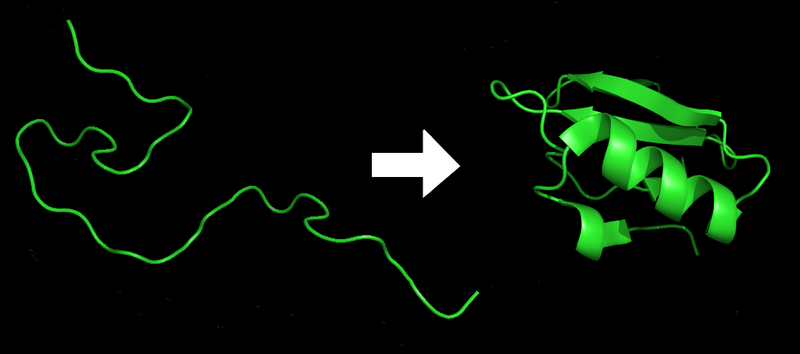


Protein folding: the three dimensional structure of a working molecule

The **primary structure** of a protein is the sequence of amino acids dictated by the CDS of the mRNA. From this sequence, the protein folds to minimize the net energy of the protein molecule in its environment.

Protein folding

11.



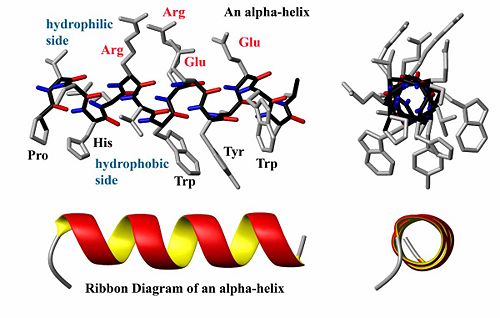
Since most proteins are in an aqueous environment, hydrophilic amino acids are generally found on the exterior of the molecule and hydrophobic ones in the interior.

**Secondary** structural folding involves local interactions of amino acids that are independent of their side chain structures. Two such folding forms are **alpha helix** and **beta sheets**.

An **alpha helix** is a structure formed by hydrogen bonds between the backbone carbonyl of one amino acid and the backbone NH of the amino acid four residues away causing the peptide to twist into a spiral or helix. The side chain structures stick out from the structure in a spiral arrangement.

12.

Top view of an alpha helix showing the side chain structures sticking out from the spiral.

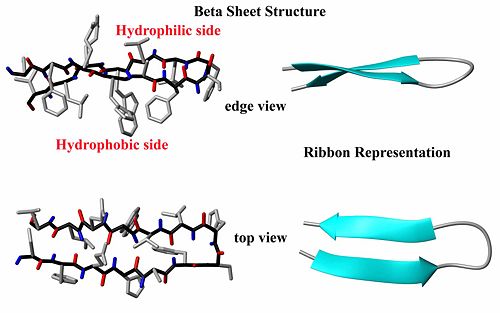


Ribbon diagrams are often used in images of proteins

A **beta sheet** is a structure formed from two or more straight amino acid chains that are hydrogen bonded side by side. Beta sheets may be formed from a single chain if it contains a beta turn, which contains a hairpin loop structure and runs back anti-parallel to itself or they can be formed from two separate chains. Beta sheets have “sides”, where the side chain structures of the amino acids stick out above or below the plane of the sheet.

13.

beta loop



A single protein can have many helices and beta sheets in its final folded form. The final folding of a single peptide, with all its helices and beta sheets, is called its **tertiary** structure.

ribbon diagram of the enzyme alcohol dehydrogenase

14.



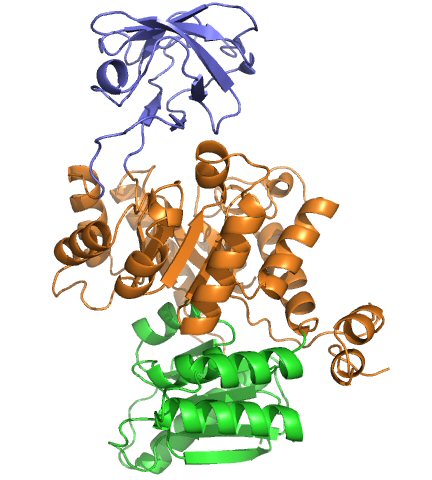
Protein domains

Many proteins contain regions or domains that fold into specific forms and perform specific functions within a protein. An example is shown below for pyruvate kinase, an enzyme of the glycolytic pathway. The enzyme converts phosphoenolpyruvate (PEP) to pyruvate as it generates ATP from ADP. The catalytic site is formed between the brown and green domains where the green domain binds ADP.

15.

domains are often joined by short amino acid segments called linkers

pyruvate kinase domains



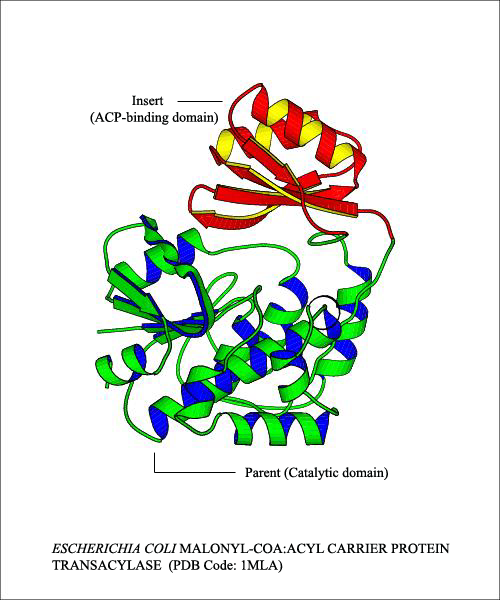
Similar functional domains may be shared by different proteins for the same function. The shared calcium binding domain found in many enzymes of the coagulation pathway is an example. The antigen binding domain found in immunoglobulin proteins is found in many other proteins where protein binding is part of their function. DNA binding proteins share similar domains for binding to DNA. ATP binding domains are common in proteins that utilize ATP in their reactions.

16.

Malonyl-CoA:Acyl Carrier Protein transferase has two domains with different functions:

ACP domain binds the carrier protein

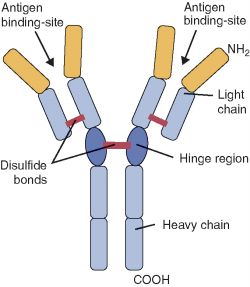
The catalytic domain binds malonyl Co-A and attaches it to the carrier protein



**Quaternary** structured proteins are made from more than one peptide. Immunoglobulin molecules (antibody proteins that protect us from disease organisms) are made from two heavy peptide chains and two light peptide chains joined together by disulfide bridges between peptides. There are several domains within each peptide.

17.

Structure of an immunoglobulin molecule showing its peptides and major domains



Hemoglobin, the protein that carries oxygen in red blood cells is a tetramer composed of two sets of peptides, alpha (red) and beta (blue)peptides. Each peptide binds a non-protein organic molecule, heme (green). The heme contains an iron atom that binds reversibly to oxygen.

Hemoglobin

18.

The alpha and beta peptides are held together by the interaction of the side chain structures of the amino acids in each peptide. The binding of oxygen to a heme on one peptide causes a ripple effect through the molecule making binding of oxygen to the other three hemes on the other peptides more likely.

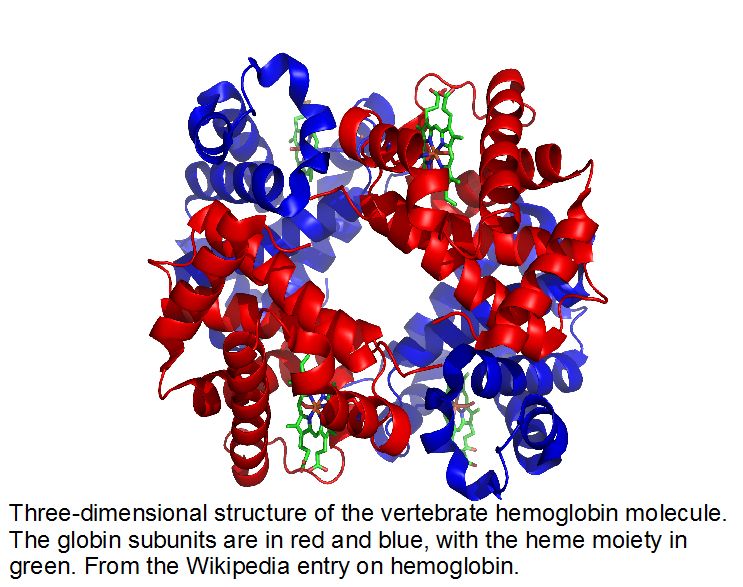


Image references

1. <http://www.ams.org/featurecolumn/images/chromosome.gif>

2. <http://static.howstuffworks.com/gif/dna-2.jpg>

3. <http://www.medicalcomputing.net/help/images/gene_structure.png>

4. <http://www.uic.edu/classes/bios/bios100/lecturesf04am/Pre-mRNA.gif>

5. <http://www.molecularstation.com/molecular-biology-images/data/503/MRNA-structure.png>

6. <http://oregonstate.edu/instruct/bb331/lecture12/fi5p20.gif>

7. <http://www.horstmann.com/sjsu/cs46a/lab6/genetic-code.jpg>

8. <http://www.carolguze.com/images/biomolecules/AminoAcidJargonStryerBio3.gif>

9. <http://www.dur.ac.uk/stat.web/Bioinformatics/aminostructure2.JPG>

10. <http://www.utpa.edu/faculty/materon/3401/mainimages/protein_synthesis.gif>

11. <http://upload.wikimedia.org/wikipedia/commons/a/a9/Protein_folding.png>

12. <http://en.citizendium.org/images/thumb/2/26/AlphaHelixByDEVolk.jpg/500px-AlphaHelixByDEVolk.jpg>

13. <http://en.citizendium.org/images/thumb/2/29/BetaSheetByDEVolk.jpg/500px-BetaSheetByDEVolk.jpg>

14. <http://www.cogs.susx.ac.uk/courses/gc/alcohol-dehydrogenase-protein.jpg>

15. <http://chemistry.umeche.maine.edu/CHY431/1pkm-1.jpg>

16. <http://www.domins.org/images/1mla.png>

17. <http://content.answers.com/main/content/img/elsevier/vet/gr222.jpg>

18. <http://www.sciencemusings.com/blog/uploaded_images/Protein2-732879.jpg>