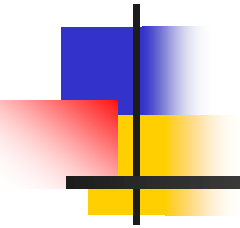


Bioinformatics

Molecular Biology Primer

A Recap of the last class
What is Bioinformatics?

- The marriage between computer science and molecular biology
  - The algorithm and techniques of computer science are being used to solve the problems faced by molecular biologists
- 'Information technology applied to the management and analysis of biological data'
  - Storage and Analysis are two of the important functions – bioinformaticians build tools for each

A Recap of the last class
What is Bioinformatics?

- This is the age of the Information Technology
- Information to the volume of Britannica Encyclopedia is stored in each of our cells
- 'Bioinformatics tries to determine what info is biologically important'

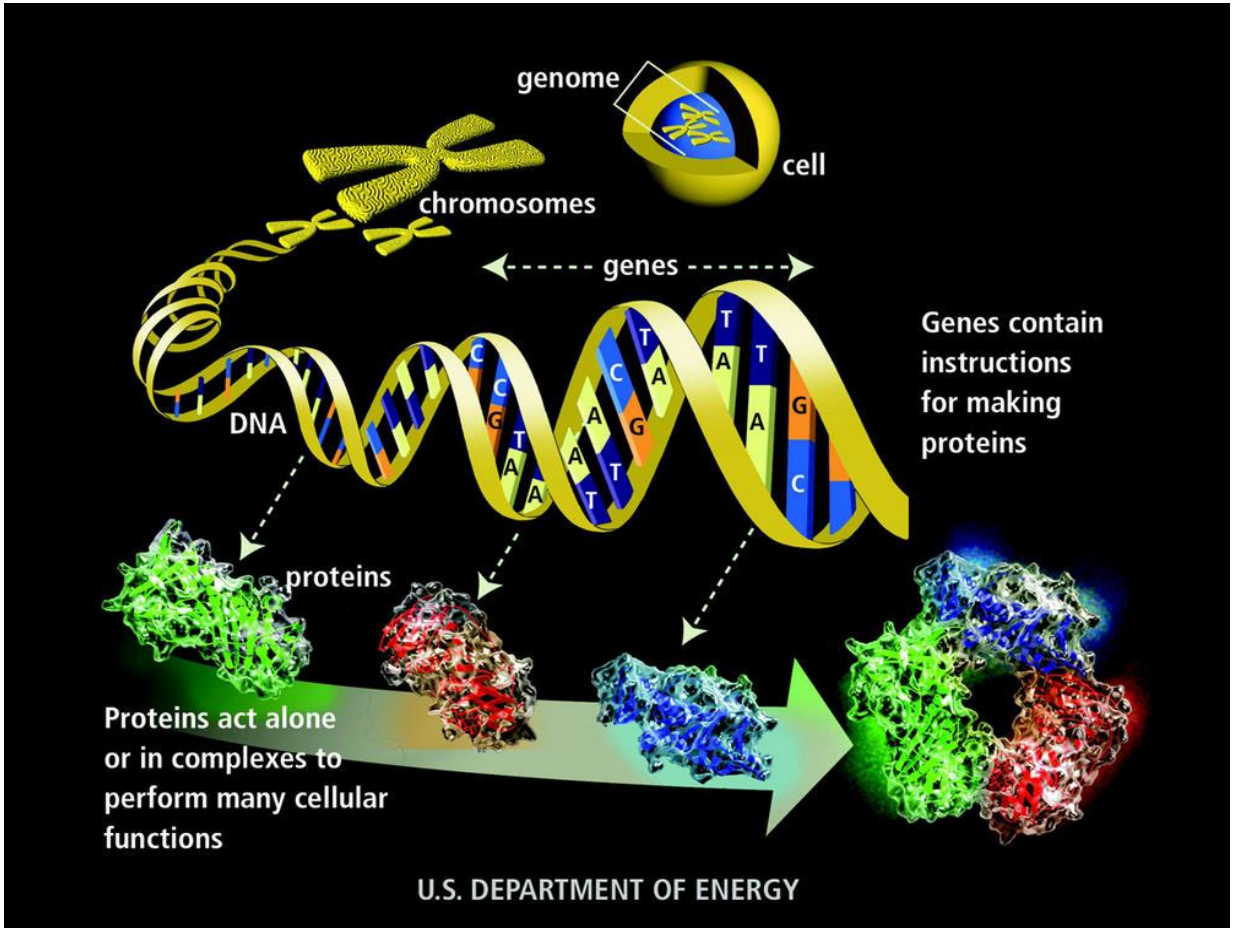
DNA and Genes

- DNA is where the genetic information is stored
- Blonde hair and blue eyes are inherited by this
- Gene - The basic unit of heredity
  - There are genes for characteristics i.e. a gene for blond hair etc
- Genes contain the information as a sequence of nucleotides
- Genes are abstract concepts – like longitude and latitudes in the sense that you cannot see them separately
- Genes are made up of nucleotides

Genome

- A genome is the full set of instructions needed to make every cell, tissue, and organ in your body.
- four-letter language of DNA (A, C, T, and G).
- The human genome contains 3 billion of these "letters" or bases.
- 23 pair of chromosomes
- Each chromosome contains genes
- 32,185 genes in human body
- These genes are the deciding factors for the development of protein

Genome


genome
cell
chromosomes
genes
DNA
Genes contain instructions for making proteins
proteins
Proteins act alone or in complexes to perform many cellular functions
U.S. DEPARTMENT OF ENERGY

Nucleotide (nt)

- Each nt is made up of
  - Sugar
  - Phospate group
  - Base
- The base it contains makes the only difference between one nt and the other
- There are 4 different bases
  - G(uanine),A(denine),T(hymine),C(ytosine)
- The information is in the order of nucleotide and the order is the info
- Genes can be many thousands of nt long
- The complete set of genetic instructions is called genomes

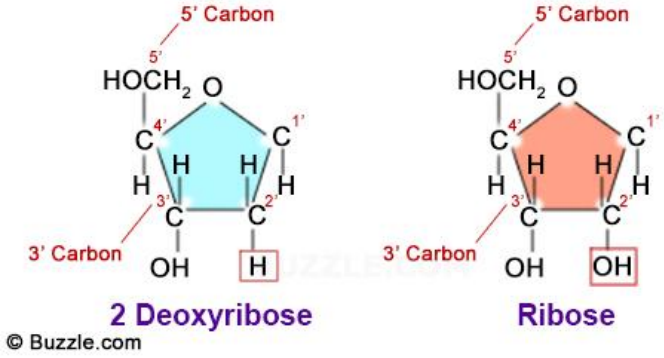
Chromosomes

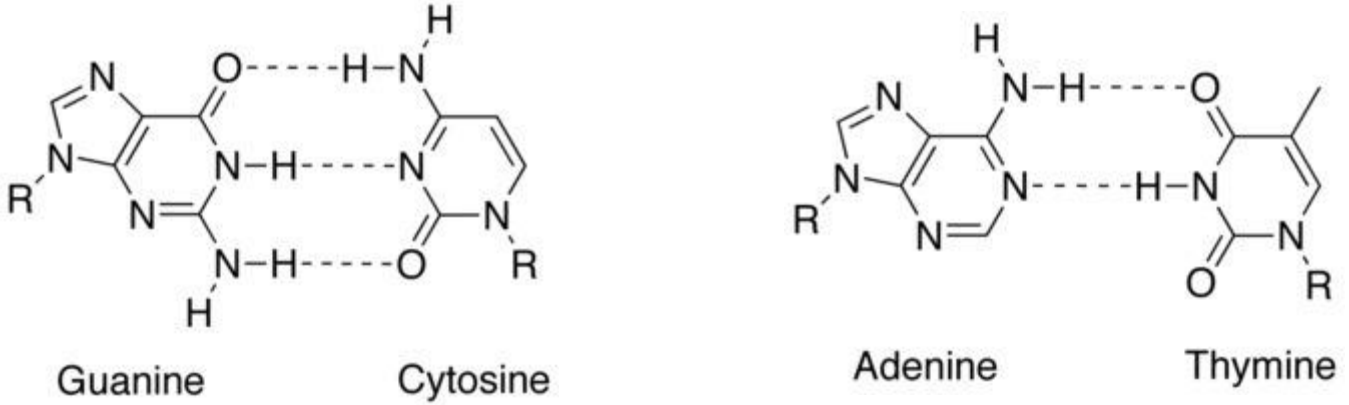
- DNA strings make chromosomes
- Analogy
  - Letters - nt
  - Sentences – genes
  - Individual volumes of Britannica encyclopedia – chromosomes
  - All voles together - Genome

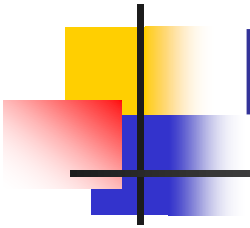
Double Helix

- The DNA is a double helix
- Each strand has complementary information
- Each particular base in one strand is bonded with another particular base in the next strand
  - G - C
  - A - T
- For example -
  - AATGC one strand
  - TTACG other strand

Double Helix







Proteins

- Proteins are very important biological feature
- Amino Acids make up the proteins
- 20 different amino acids are there
- The function of a protein is dependant on the order of the amino acids

Proteins…

- The information required to make aa is stored in DNA
- DNA sequence determines amino acid sequence
- Amino Acid sequence determines protein structure
- Protein structure determines protein function
- A Substance called RNA is used to carry the Info stored in the DNA that in turn is used to make proteins
- Storage - DNA
- Information Transfer – RNA
- RNA is the message boy!

Biological Nomenclature

- **Cell**

  Basic Unit of life

- **DNA (Deoxy Ribonucleic Acid)**
  Ladder-like molecule that stores genetic information, including information needed to build proteins

Biological Nomenclature

- **Nucleotide**

The basic unit of DNA consisting of a base (A, C, T, or G), a sugar, and a phosphate

- **Chromosome**

A structure within a cell, made up of DNA, that directs the activity within the cell and passes on genetic information to new cells

Biological Nomenclature

- Protein
  Long-chained molecules used to build and repair cells; made of one or more chains of amino acids

- Amino acid
  The building block of proteins

Biological Nomenclature

- Polypeptide

a chain of amino acids joined together by peptide bonds; proteins are made up of one or more polypeptides

- Enzyme

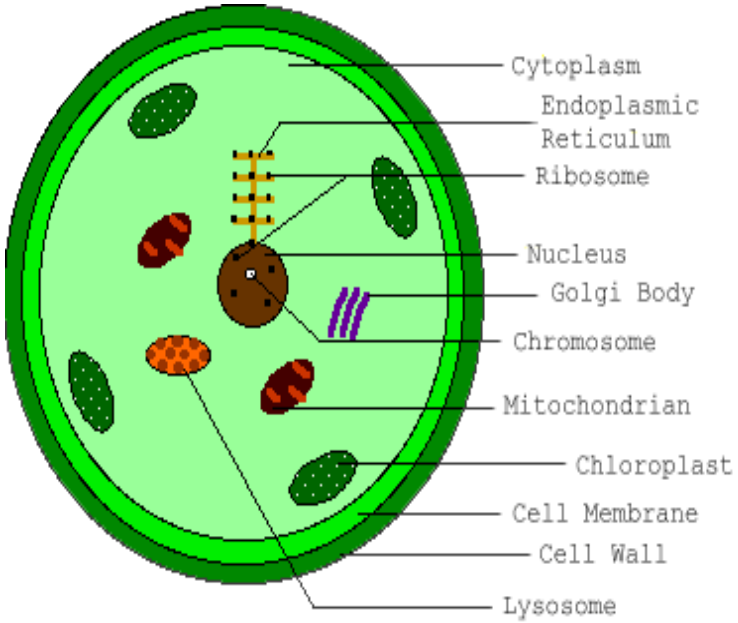
A protein that helps control chemical reactions

Biological Nomenclature

- **Ribosomes**
  Protein-making sites within cells

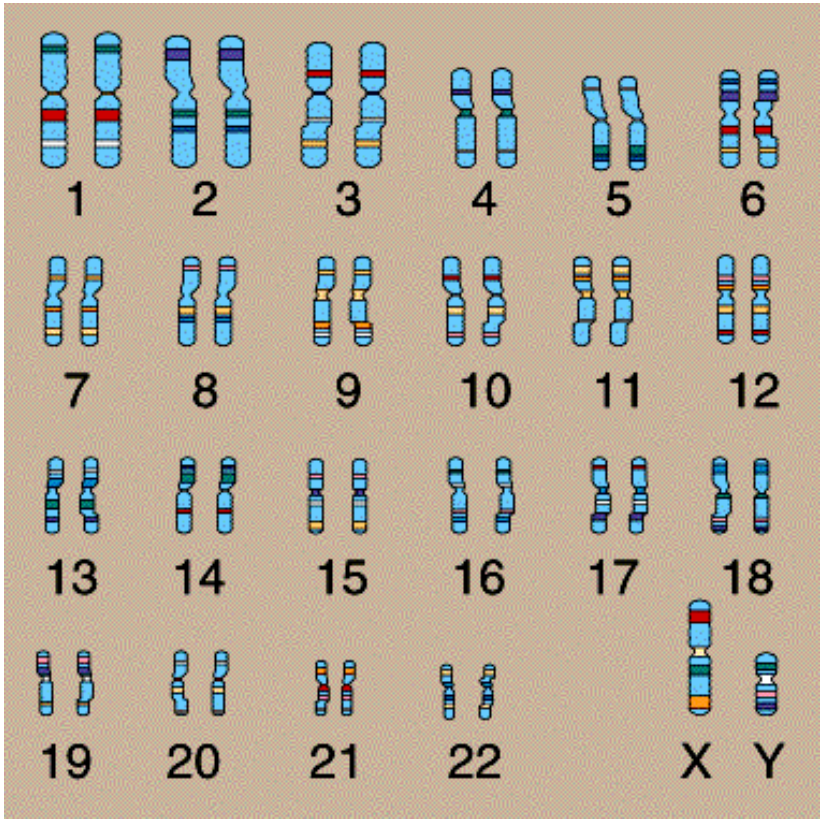
- **RNA (ribonucleic acid)**
  Substance that reads the genetic information carried by DNA; the same as DNA except that it is made up of only one strand, contains the base uracil (U) instead of thymine (T), and contains the sugar ribose instead of deoxyribose

Cells

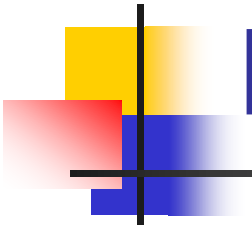


- Basic unit of life
- Different types of cell:
  - Skin, brain, red/white blood
  - Different biological function
- Cells produced by cells
  - Cell division (mitosis)
  - 2 daughter cells
- Eukaryotic cells
  - Have a nucleus

Nucleus and Chromosomes



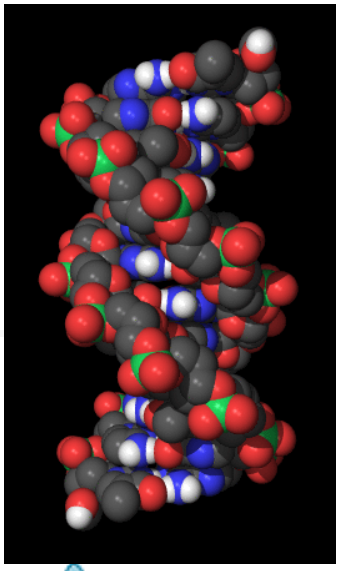
- Each cell has nucleus
- Rod-shaped particles inside
  - Are chromosomes
  - Which we think of in pairs
- Different number for species
  - Human(46),tobacco(48)
  - Goldfish(94),chimp(48)
  - Usually paired up
- X & Y Chromosomes
  - Humans: Male(xy), Female(xx)
  - Birds: Male(xx), Female(xy)



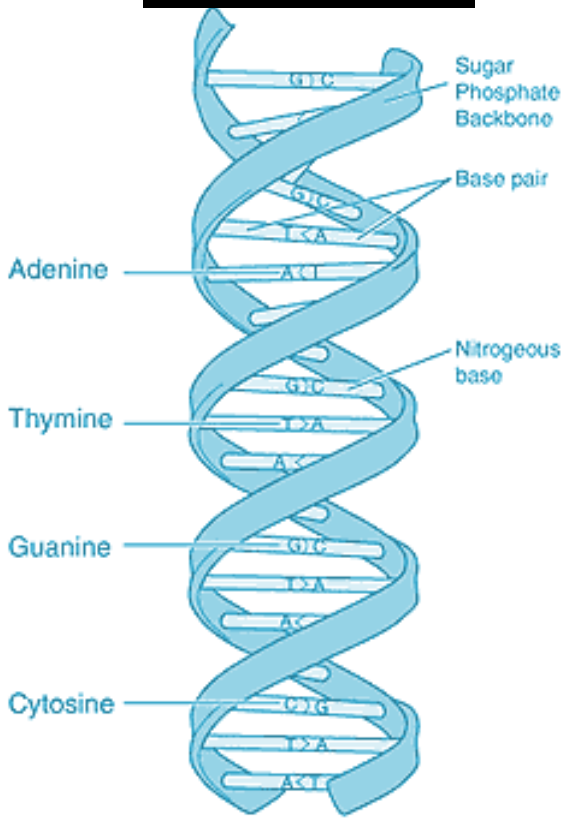
DNA Strands

- Chromosomes are same in every cell of organism
  - Supercoiled DNA (Deoxyribonucleic acid)
- Take a human, take one cell
  - Determine the structure of all chromosomal DNA
  - You've just read the human genome (for 1 person)
  - Human genome project
    - 13 years, 3.2 billion chemicals (bases) in human genome
- Other genomes being/been decoded:
  - Pufferfish, fruit fly, mouse, chicken, yeast, bacteria

DNA Structure

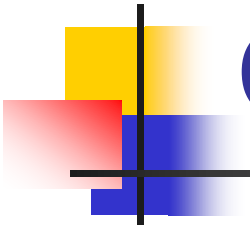


- Double Helix (Crick & Watson)
  - 2 coiled matching strands
  - Backbone of sugar phosphate pairs
- Nitrogenous Base Pairs
  - Roughly 20 atoms in a base
  - Adenine ⇔ Thymine [A,T]
  - Cytosine ⇔ Guanine [C,G]
  - Weak bonds (can be broken)
  - Form long chains called polymers
- Read the sequence on 1 strand
  - GATTCATCATGGATCATACTAAC



Differences in DNA

- DNA differentiates:
  - Species/race/gender
  - Individuals
- We share DNA with
  - Primates,mammals
  - Fish, plants, bacteria
- Genotype
  - DNA of an individual
    - Genetic constitution
- Phenotype
  - Characteristics of the resulting organism
    - Nature and nurture



Genes

- Chunks of DNA sequence
  - Between 600 and 1200 bases long
  - 32,000 human genes, 100,000 genes in tulips
- Large percentage of human genome
  - Is “junk”: does not code for proteins
- “Simpler” organisms such as bacteria
  - Are much more evolved (have hardly any junk)
  - Viruses have overlapping genes (zipped/compressed)
- Often the active part of a gene is split into exons
  - Separated by introns

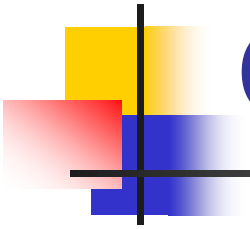
The Synthesis of Proteins

- Instructions for generating Amino Acid sequences
  - (i) DNA double helix is unzipped
  - (ii) One strand is transcribed to messenger RNA
  - (iii) RNA acts as a template
    - ribosomes translate the RNA into the sequence of amino acids
- Amino acid sequences fold into a 3d molecule
- Gene expression
  - Every cell has every gene in it (has all chromosomes)
  - Which ones produce proteins (are expressed) & when?

Transcription

- Take one strand of DNA
- Write out the counterparts to each base
  - G becomes C (and vice versa)
  - A becomes T (and vice versa)
- Change Thymine [T] to Uracil [U]
- You have transcribed DNA into messenger RNA
- Example:

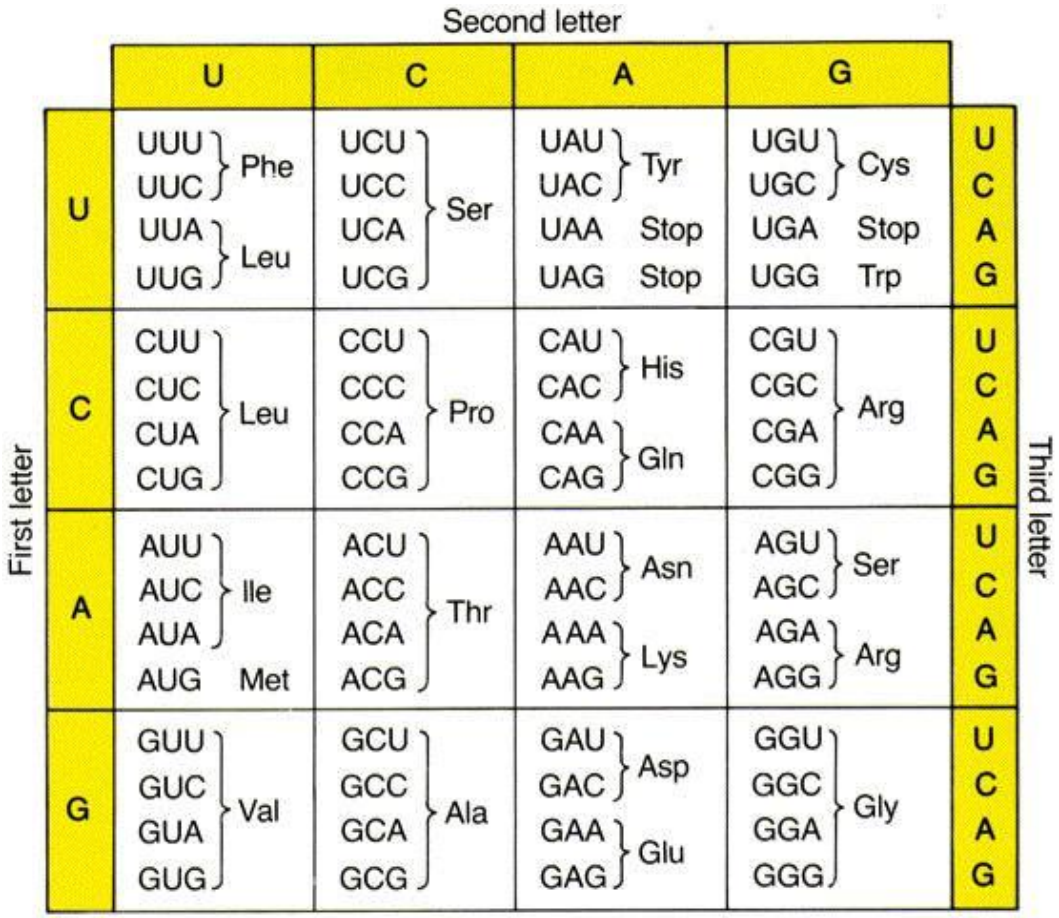
```
Start:          GGATGCCAATG
Intermediate:   CCTACGGTTAC
Transcribed:    CCUACGGUUAC
```



Genetic Code

- How the translation occurs
- Think of this as a function:
  - Input: triples of three base letters (Codons)
  - Output: amino acid
  - Example: ACC becomes threonine (T)
- Gene sequences end with:
  - TAA, TAG or TGA

Genetic Code



A=Ala=Alanine

C=Cys=Cysteine

D=Asp=Aspartic acid

E=Glu=Glutamic acid

F=Phe=Phenylalanine

G=Gly=Glycine

H=His=Histidine

I=Ile=Isoleucine

K=Lys=Lysine

L=Leu=Leucine

M=Met=Methionine

N=Asn=Asparagine

P=Pro=Proline

Q=Gln=Glutamine

R=Arg=Arginine

S=Ser=Serine

T=Thr=Threonine

V=Val=Valine

W=Trp=Tryptophan

Y=Tyr=Tyrosine

Example Synthesis

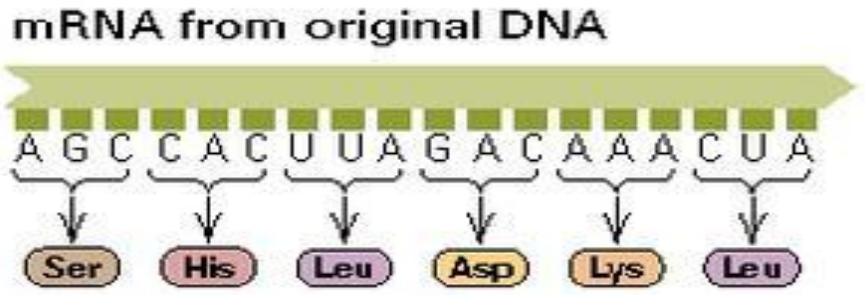
- TCGGTGAATCTGTTTGAT

  Transcribed to:

- AGCCACUUAGACAAACUA

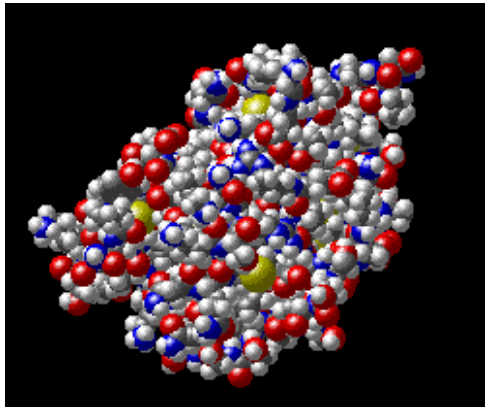
  Translated to:

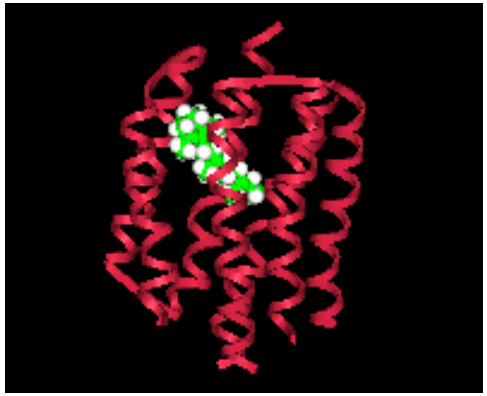
- SHLDKL



Proteins

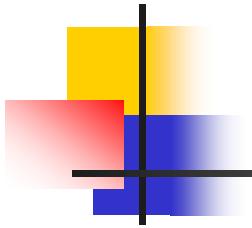
- DNA codes for
  - strings of amino acids
- Amino acids strings
  - Fold up into complex 3d molecule
  - 3d structures:conformations
  - Between 200 & 400 “residues”
  - Folds are proteins
- Residue sequences
  - Always fold to same conformation
- Proteins play a part
  - In almost every biological process





Further Reading

- **Structure of Cells**
  - http://www.wiley.com/legacy/college/boyer/0470003790/animations/animations.htm
- **Human Genome Project at Sanger Centre**
  - http://www.sanger.ac.uk/HGP/
- **Talking glossary of genetic terms**
  - http://www.genome.gov/glossary.cfm
- **Primer on molecular genetics**
  - http://www.ornl.gov/TechResources/Human_Genome/publicat/primer/toc.html



Thank You