Molecular Biology Primer

Selected Topics in Computer Intelligence - 2015

Bioinformatics Programming

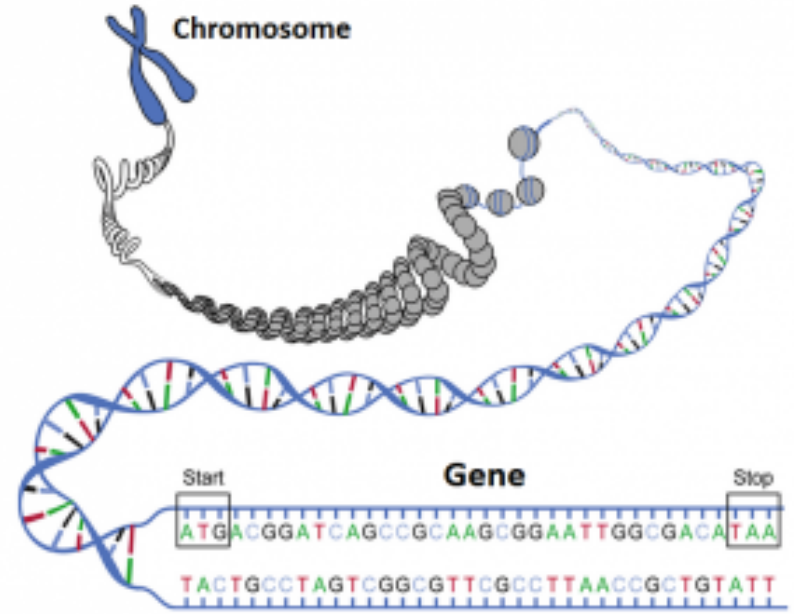
Computer Engineering, Chiang Mai University

What is Life made of?

- All cells have a life cycle: born, eat, replicate, and die
  - Cells have no brain – How to decide when to do something.?
  - Pathways – complex networks of chemical reactions
- All life depends on 3 types of molecules:
  - DNA – vast library describing how the cell works
  - RNA – transfer certain short pieces of library to different places
  - Proteins – form enzymes that perform various activity in body
- DNA and RNA are strings of *4 alphabets* (nucleotides)
- Protein is strings of *20 letters* (amino acids)

What is the Genetic Material?

- Genetic information was organized into genes (or traits) that resided on chromosomes
- Certain genes are inherited together
  - They are located closer on a chromosome
- Genetic map
  - order of genes on chromosomes

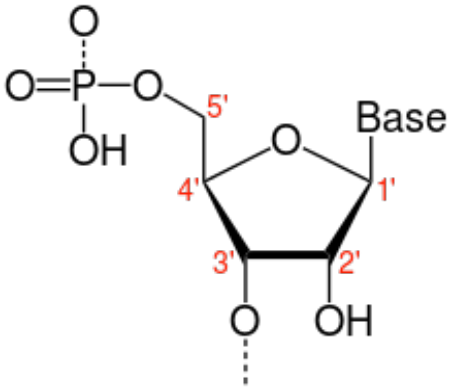


What Molecule Codes for Genes?

- DNA was discovered in 1869 by *Johann Friedrich Miescher* – isolated "*nuclein*" from the nuclei of white blood cells

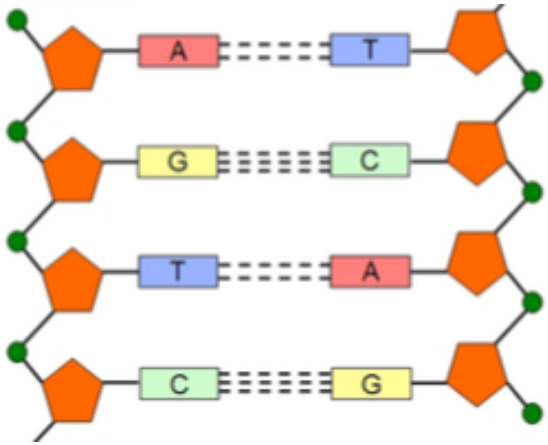
- 4 bases: **A**denine, **T**hymine, **G**uanine, and **C**ytosine
  - Originally found the 5th base – **U**racil

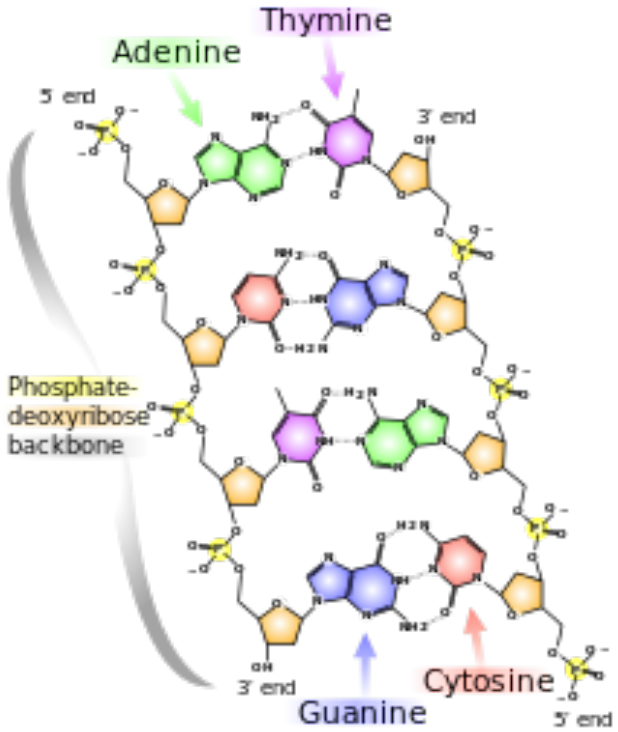
- Later on, nucleic acids ware grouped into 2 classes
  - DNA (Deoxyribonucleic acid)
    - Phosphate Group + Sugar ring + Bases (A,T,C,G)
  - RNA (Ribonucleic acid)



What is the Structure of DNA?

- The modern DNA era began in the 1953
  - *James Watson & Francis Crick* found out that the structure of DNA is "double helix strands"
  - Two strands were held together by hydrogen bond between specific base paring – Chargaff rule
    - A-T and C-G
    - The two strands has complementary relationship





What Carries Information between DNA and Protein?

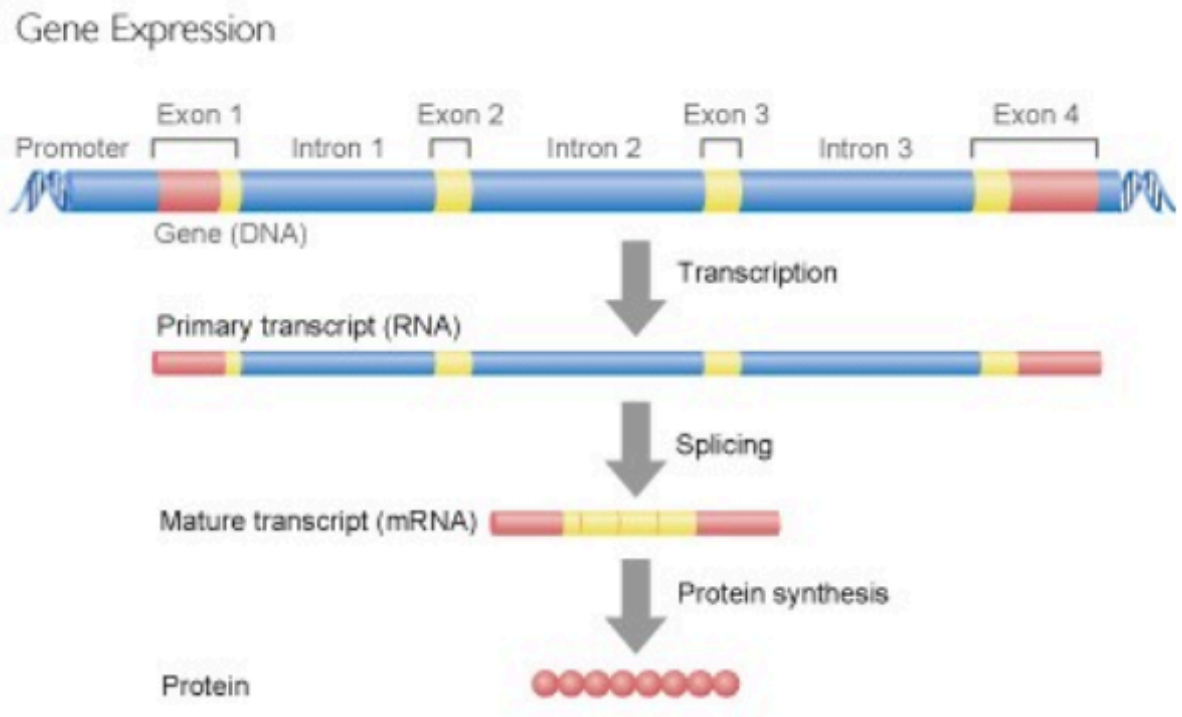
- DNA content of a cell does not change over time, but the concentrations of different proteins do
- How DNA generates an enormous variety of different Proteins?
- What translate texts written in a 4-letter alphabet into text written in 20-letter alphabets? and How?
- Proteins could not be made directly from DNA
  - DNA resides within the nucleus
  - Protein synthesis had been observed to happen outside the nucleus
    - The Cytoplasm

What Carries Information between DNA and Protein?

- In 1960, biologist proved that RNA is complementary to the DNA segment (i.e., the gene) that codes for protein
- DNA serves as a template used to copy a gene into messenger RNA (mRNA) - transcription
- mRNA then carries the gene's genetic info to the ribosome to make a particular protein - translation
  - RNA is more active than DNA - usually lives in single-strand form
- The transcription of a gene into mRNA is tightly controlled – not all genes produce proteins at all time

What Carries Information between DNA and Protein?

- In eukaryote, a gene is typically broken into many pieces
  - Cut out *intron* and concatenate all the *exons* together
  - Happens before the mRNA entering the ribosome
  - Ribosomes are molecular factories where proteins are assembled



How are Proteins Made?

- The amino acids were linked together into linear chains to form proteins
  - Properties of proteins were defined by the composition and arrangement of their amino acids
- Triplets of consecutive letters in DNA (codons) were responsible for the amino acid sequence in a protein
  - Some different triplets of nucleotides may code for the same amino acid
- The 3D structure of protein determines what role a protein plays in the cell
- It is possible that one gene to code for many proteins
  - Results from alternative splicing

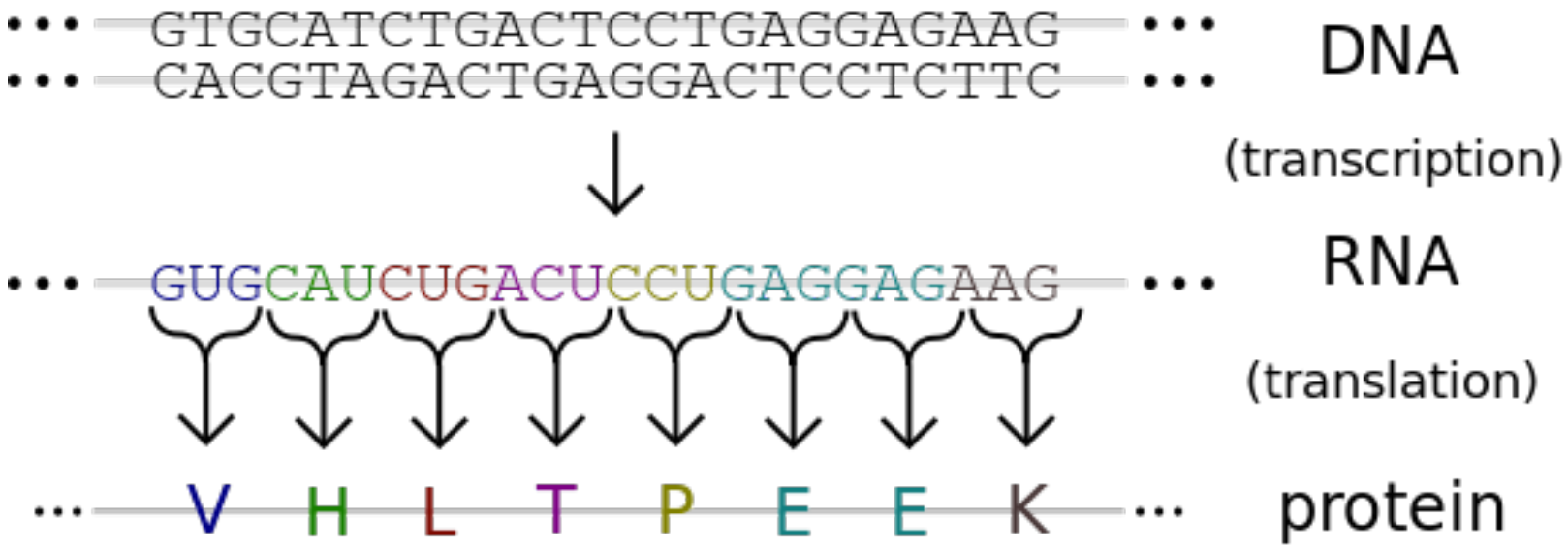
How are Protein Made?

- RNA Codon Table

1st base	2nd base								3rd base
	U		C		A		G		
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L) Leucine	UCA		UAA	Stop (*Ochre*)	UGA	Stop (*Opal*)	A
	UUG		UCG		UAG	Stop (*Amber*)	UGG	(Trp/W) Tryptophan	G
C	CUU		CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	AUG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

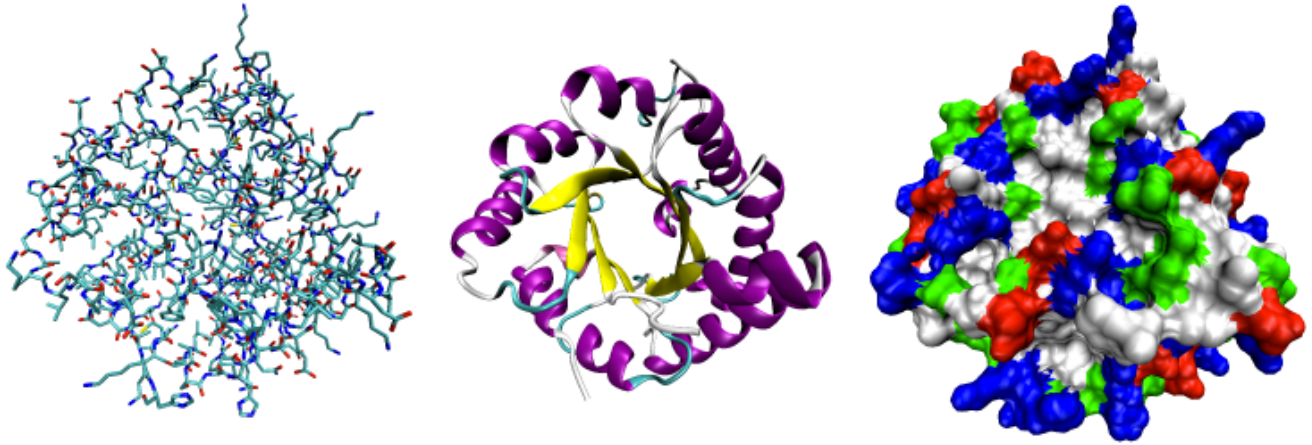
How are Proteins Made?

- The DNA sequence of a gene encodes the amino acid sequence of protein

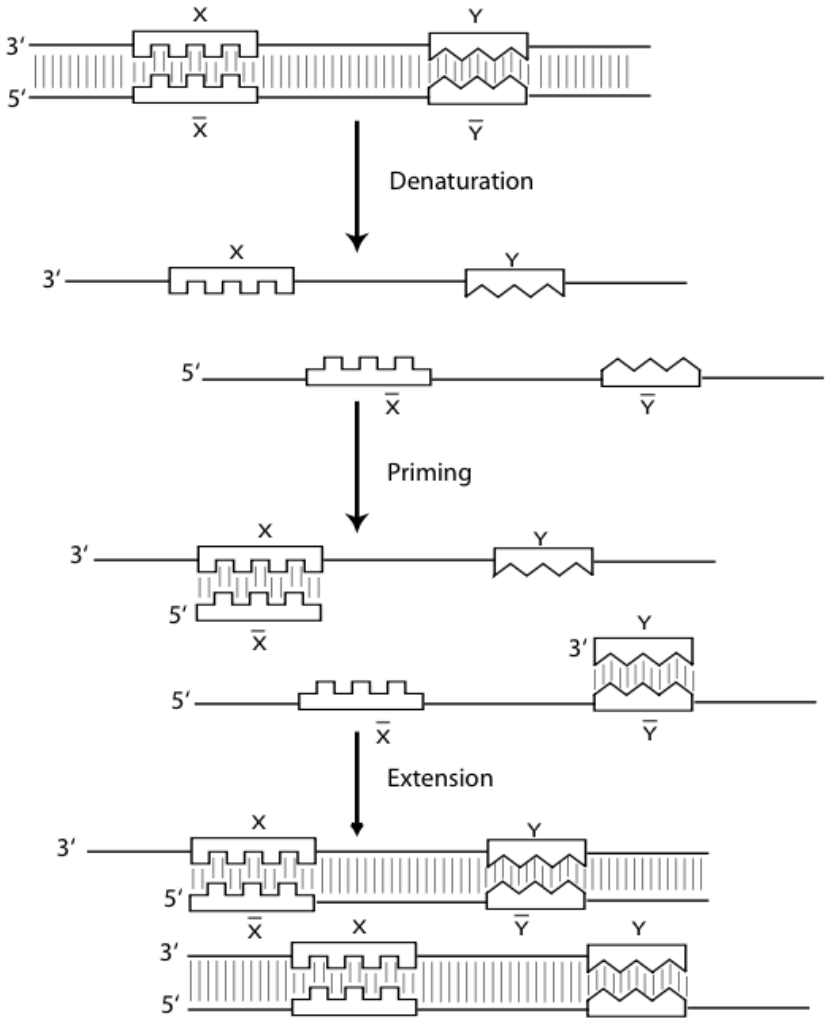


Protein Structure

- Most proteins fold into unique 3D structures
  - Many can fold unassisted, simply through the chemical properties of their amino acids, others require the aid of molecular chaperones
  - Primary structure: the amino acid sequence
  - Secondary structure: regularly repeating local structure
    - Alpha helix, beta sheet, turn
  - Tertiary structure: overall shape of a single protein molecule (aka. Fold)
    - Controls the basic function of the protein



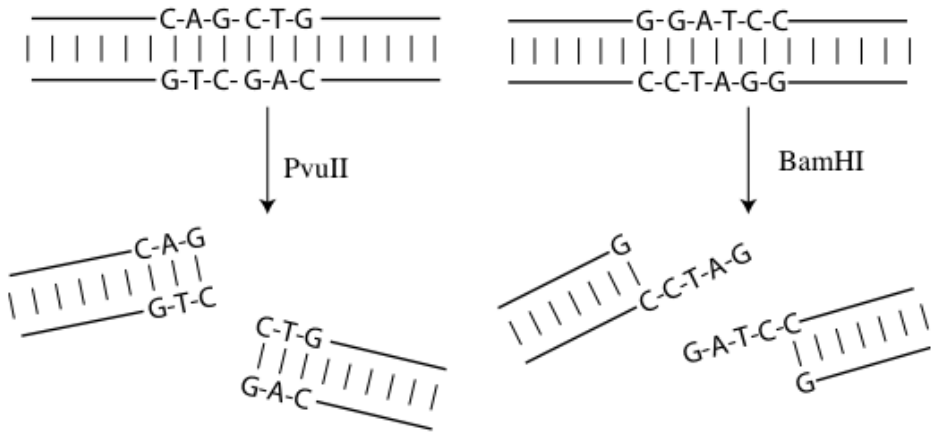
How can we Analyze DNA?



- Copying
  - Create many copies of the DNA fragment – millions to billions
  - PCR – Polymerase Chain Reaction
    - Repetitive process with exponential grow
    - Three operations:
      - denaturation,
      - priming,
      - extension

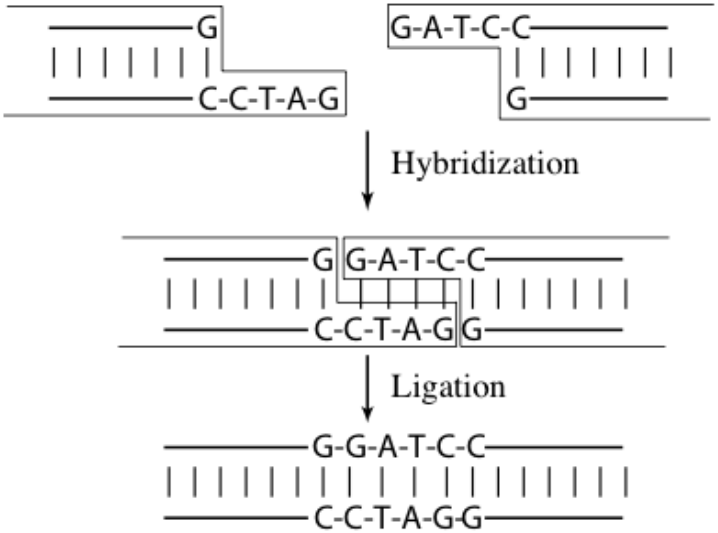
Cutting and Pasting DNA

- To study a gene of interest, sometimes necessary to cut it out of an organism's genome and reintroduce into some host organism that is easy to grow
  - We need some kind of "scissors" to do the task
- Restriction enzymes – proteins that act as molecular scissor that cut DNA at every occurrence of a certain string (restriction site)
  - The enzymes cut DNA into restriction fragments
  - Blunt cut, or stick cut



Cutting and Pasting DNA

- Many ways to use two pieces of DNA together
- Hybridization – based on complementary base-paring
- Ligation – fixing bonds within single strands



Measuring DNA Length

- Gel Electrophoresis
  - Measuring size without actually finding the exact sequence
  - DNA is a negative charged molecule moving toward positive pole
  - The gel acts as a molecular brake
  - The speed of migration of a fragment is related to the size
  - The migration distance can be used to estimate the size of a fragment

Probing DNA

- Test whether a particular DNA fragment is present in a given DNA strand
  - Often done using hybridization
- Probes – single-stranded DNA fragment
  - 20-30 nucleotides long
  - Known sequence
  - Fluorescent tag
- DNA array – composed of spots
  - Each spot are many copies of complement of one gene's mRNA transcript

How do Individuals of a Species Differ?

- Traits are caused by variations in gene
- Genome: the complete set of nucleic acid sequence for an organism
  - Protein-coding DNA genes
  - Noncoding DNA
- Only 0.1% of the 3 billion nucleotide human genome are different between any two individuals
- The master genome represent all the possible genomes that an individual of that species could have

How Do Different Species Differ?

- The genomes of different organisms may be vastly different and similar
  - Many genes in humans and flies are similar
- A species is a collection of individuals whose genomes are "compatible"
- What parts of the fruit fly genome are similar and dissimilar to human genome?
  - Bioinformatics helps answering these kinds of questions
  - Comparative Genomic
  - Alignment Algorithm - **BLAST**