



Foundation of Bio-Informatics

Lecture 2

Lecture by: Ahmad R. Naghsh-Nilchi, PhD

Department of Artificial Intelligence

Faculty of Computer Engineering

University of Isfahan



DNA to Protein Conversion

- Determining the sequence of a protein is much more difficult than sequencing DNA.
- All proteins that a given organism (whether microbe or human being) can synthesize, are encoded in the DNA sequence of its genome.
- A smart shortcut is to read protein sequences directly at the information source: in DNA sequence!
- This way, we get the amino-acid sequence of a protein that has never been isolated in a test tube.



Genetic Code

- DNA sequence translates into a protein sequence using the genetic code.
- In actual life, the cell itself generates a protein in the same way.
- Genetic code uniquely relates a 4-nucleotide sequence (A, T, G, C) to a suite of 20 amino acids.
- This is done using symbols rather than actual chemicals)
- This is the most brilliant achievements of biologists of 1960s.



How do Genetic Codes work?

- From a given point in your DNA sequence, start reading the sequence 3 nucleotides (or *1 triplet*) at a time.
- Then use the genetic code table to read which amino acid corresponds to this triplet (referred to as *codons*). The protein sequence is achieved!
- Example: for this DNA sequence:

ATGGAAGTATTTAAAGCGCCACCTATTGGGATAG

Triplet: ATG GAA GTA TTT AAA GCG CCA CCT ATT GGG ATA G

Codons: M E V F K A P P I G I stops

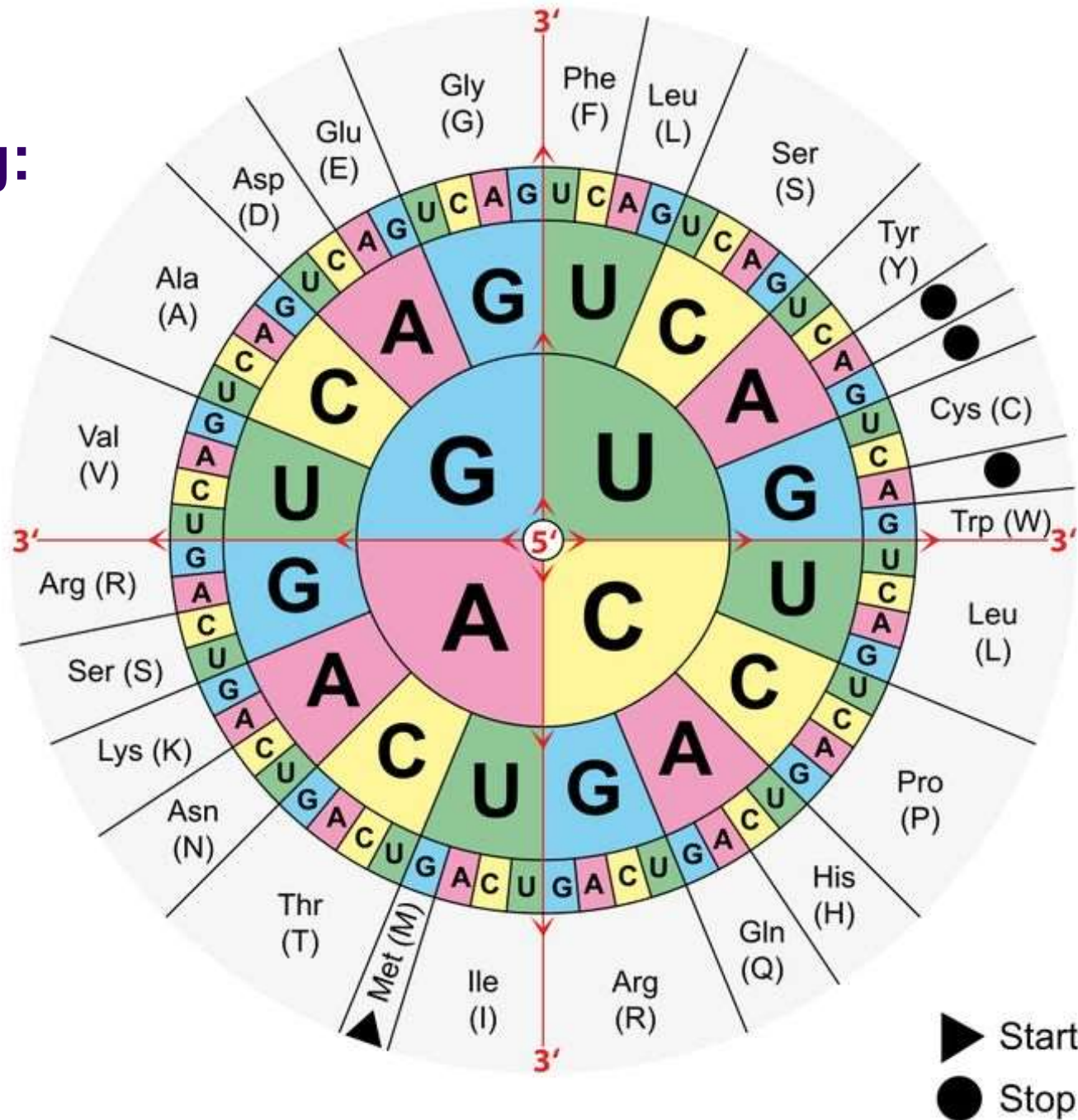


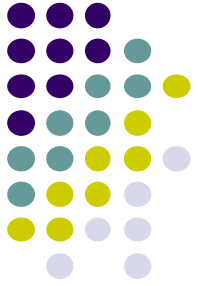
Table of Standard Genetic Code

	T	C	A	G
T	TTT Phe (F)	TCT Ser (S)	TAT Tyr (Y)	TGT Cys (C)
	TTC Phe (F)	TCC Ser (S)	TAC Tyr (Y)	TGC Cys (C)
	TTA Leu (L)	TCA Ser (S)	TAA Stop	TGA Stop
	TTG Leu (L)	TCG Ser (S)	TAG Stop	TGG Trp (W)
C	CTT Leu (L)	CCT Pro (P)	CAT His (H)	CGT Arg (R)
	CTC Leu (L)	CCC Pro (P)	CAC His (H)	CGC Arg (R)
	CTA Leu (L)	CCA Pro (P)	CAA Gln (Q)	CGA Arg (R)
	CTG Leu (L)	CCG Pro (P)	CAG Gln (Q)	CGG Arg (R)
A	ATT Ile (I)	ACT Thr (T)	AAT Asn (N)	AGT Ser (S)
	ATC Ile (I)	ACC Thr (T)	AAC Asn (N)	AGC Ser (S)
	ATA Ile (I)	ACA Thr (T)	AAA Lys (K)	AGA Arg (R)
	ATG Met (M)	ACG Thr (T)	AAG Lys (K)	AGG Arg (R)
G	GTT Val (V)	GCT Ala (A)	GAT Asp (D)	GGT Gly (G)
	GTC Val (V)	GCC Ala (A)	GAC Asp (D)	GGC Gly (G)
	GTA Val (V)	GCA Ala (A)	GAA Glu (E)	GGA Gly (G)
	GTG Val (V)	GCG Ala (A)	GAG Glu (E)	GGG Gly (G)



Codon Ring:





Notes on Genetic Coding:

- If your DNA sequence is correctly listed in 5' to 3' orientation, you generate the protein sequence in the conventional N-to-C terminus, so you do not have to worry about the orientation details ever.
- Your computer can pretend to be a cell and generate the corresponding amino-acid sequence.



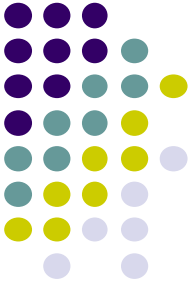
Genetic Coding Challenge

- The resulting protein sequence entirely depends on how you convert the DNA sequence into triplets.
- For instance: starting from the second letter of the previous example leads to a entirely different protein sequence:

ATGGAAGTATTTAAAGCGCCACCTATTGGGATAG

Triplet: A TGG AAG TAT TTA AAG CGC CAC CTA TTG GGA TAG

Codons: W K Y L K R H L L G Y stops



Thank you

for your attention