CONCEPT: INDIRECT PROTEIN SEQUENCING VIA GENOMIC ANALYSES

●_____ protein sequencing (via Tandem MS or Edman degradation), is used on *already* extracted/isolated proteins.

□ *Directly* identifies sequence of _____ proteins in a sample; HOWEVER….

●Most protein sequencing data is derived *indirectly* from _____ analyses (translating nucleotide sequences of genes).

Why Obtain Protein Sequences via Genomic Analyses?

●It saves a lot of time! Working with DNA is *easier* than working with proteins.

□ DNA sequencing is significantly _____, cheaper & *more* efficient/informative than direct protein sequencing.

□ Allows us to collect more protein sequencing data faster!

So Why Do We Even Need Direct Protein Sequencing?

●We can't just scrap direct protein sequencing because it has its own advantages!

□ Genomic analyses cannot identify an unknown protein sample on its own.

□ Unlike genomic analyses, Tandem MS can reveal chemically _____ residues (ex. lipoproteins, etc.).

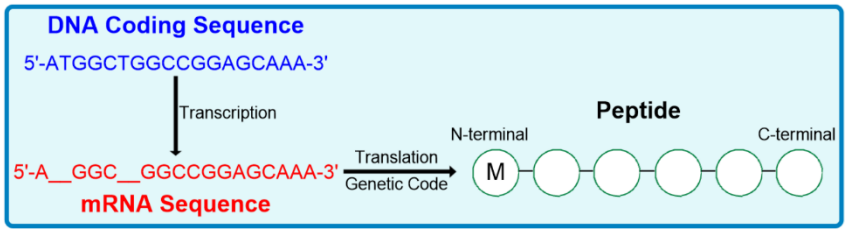
Genetic Code

●Recall: _____ *code* reveals the connection between codons of nucleic acids & amino acids of proteins.

EXAMPLE: Use the Genetic Code to reveal the peptide sequence.

_____ **Base of Codon**

First Base of Codon	U	C	A	G	Third Base of Codon
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr **UAA Stop** **UAG Stop**	UGU Cys UGC Cys **UGA Stop** UGG Trp	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G
A	AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G



Amino Acid Sequence of Peptide Revealed via Genomic Analysis

PRACTICE: Use the genetic code above & the coding DNA sequence below to determine the protein sequence.

5'-ATGGCCTGCGTTCTCAAG-3' ⟶

CONCEPT: INDIRECT PROTEIN SEQUENCING VIA GENOMIC ANALYSES

PRACTICE: Suppose the sequence below is a template DNA sequence. What is the corresponding protein sequence?

5'-ATGGCCTGCGTTCTCAAG-3' 

PRACTICE: Even when the sequence of nucleotides for a gene is available and genomic analyses can be performed, direct chemical techniques on the physical protein are still required to determine:

a) The molecular weight of a simple protein.
b) The N-terminal amino acid residue.
c) The total number of amino acid residues in the protein.
d) The location of disulfide bonds.