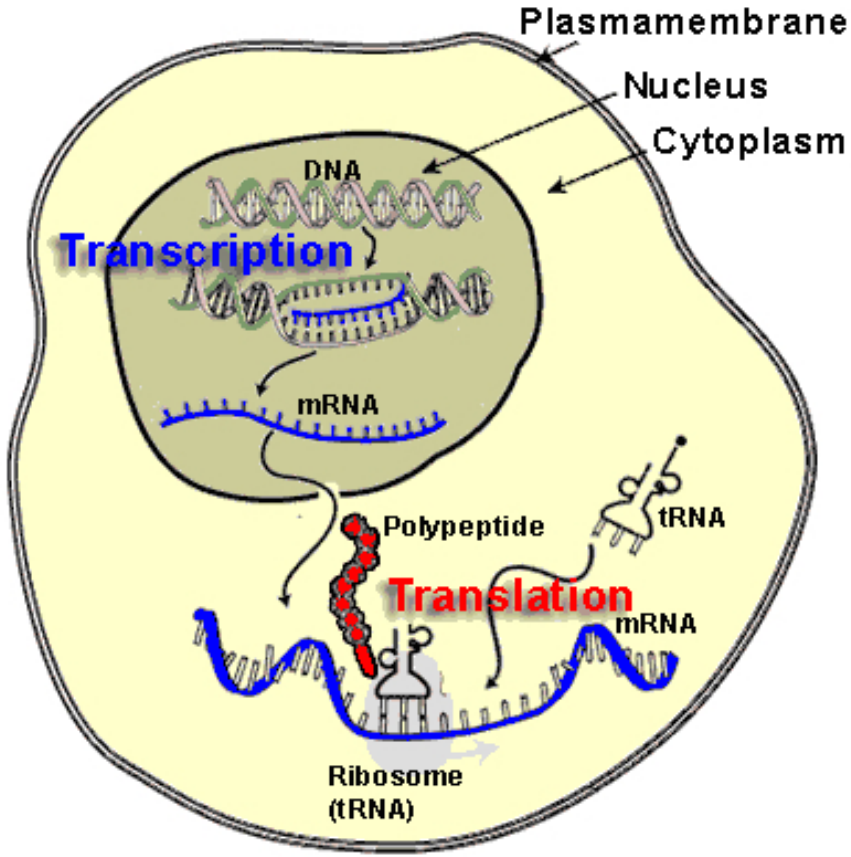
DNA and the Genetic Code

June 14, 2011

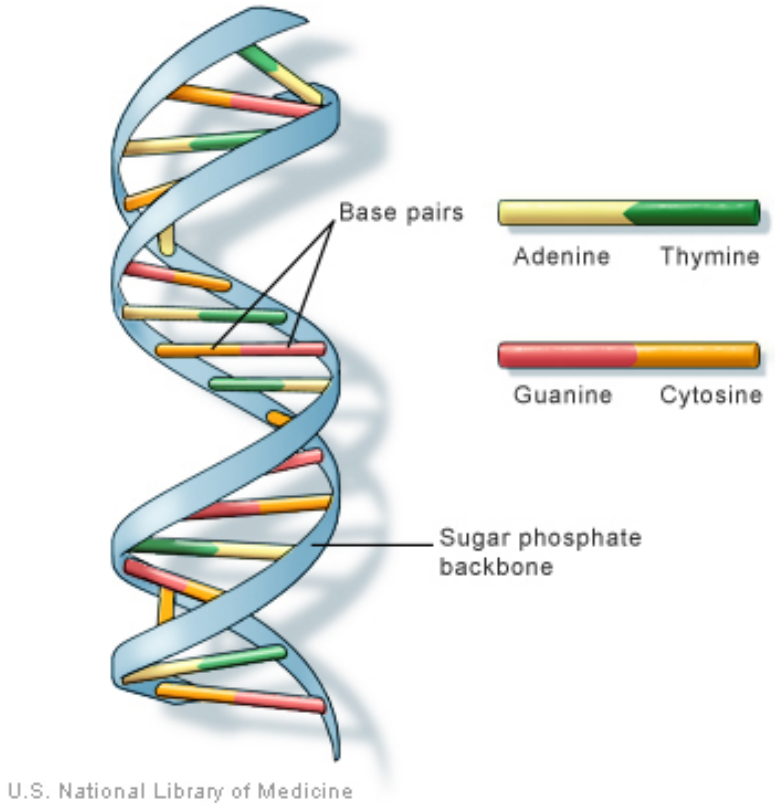
Protein synthesis

- The genetic code is used to make proteins.
- Proteins, which are biochemical compounds, are used everywhere!!
- Protein synthesis requires two steps: transcription and translation.

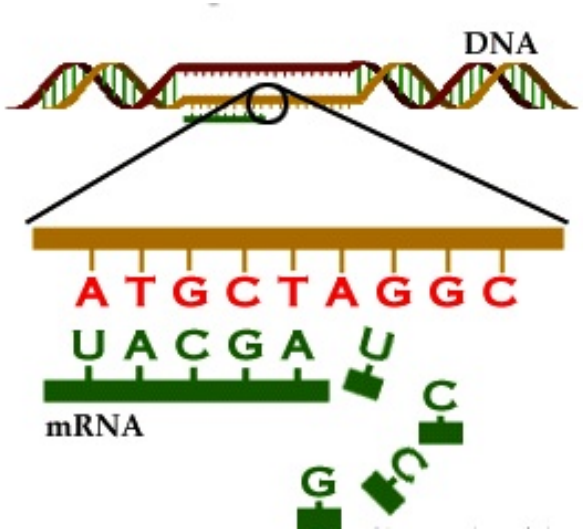


DNA

- DNA, Deoxyribonucleic acid, is a two-stranded, helical polymer that contains genetic instructions used during development.
- DNA is comprised of 4 bases: guanine (G), adenine (A), cytosine (C), and thymine (T).
- The bases are paired, G on one strand pairs with C on the other strand. Similarly, A pairs with T.
- Therefore, we call the two strands of DNA complement strands



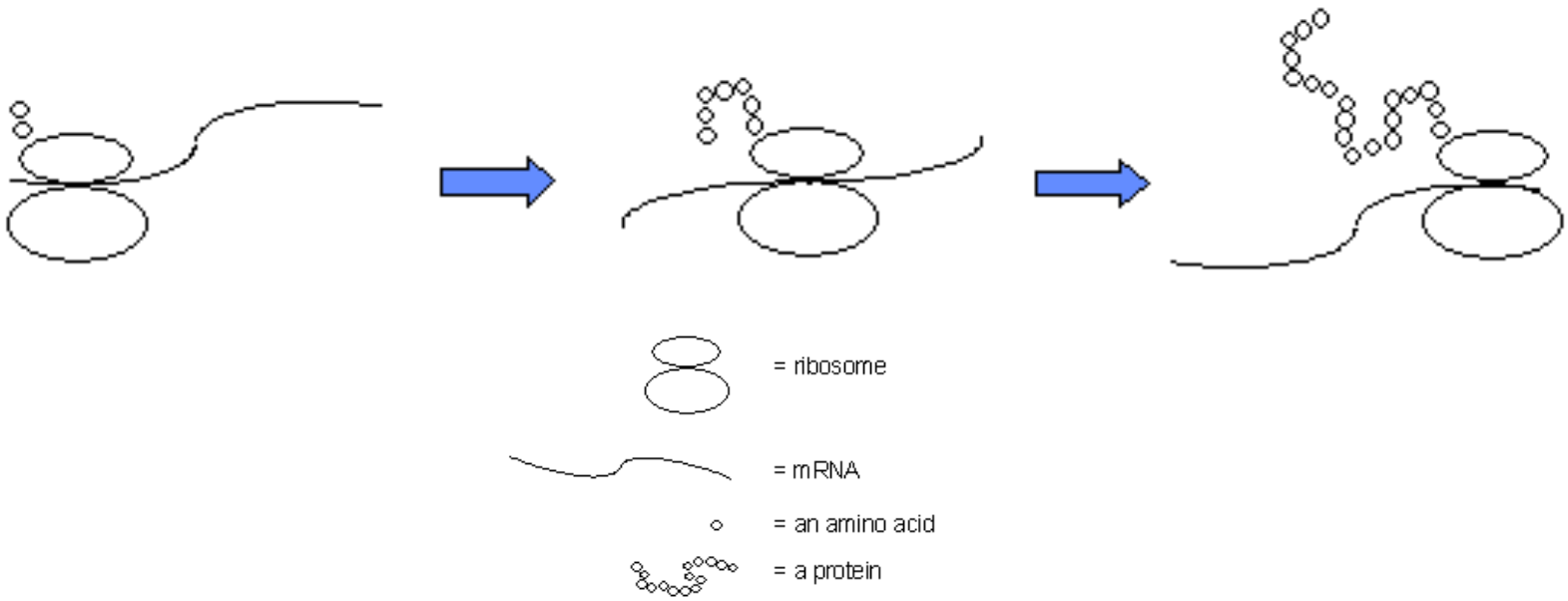
Transcription



- Transcription is the synthesis of mRNA (messenger RNA) from DNA
- The two DNA strands separate from one another. One strand is used for replication. The other is used for transcription.
- mRNA is the complement strand of the original DNA except with thymine replaced with uracil (U).
- mRNA carries genetic information from the gene out of the nucleus, into the cytoplasm of the cell where it is translated to produce protein.

Translation

- Translation is the process where ribosomes decode mRNA to produce amino acids.
- mRNA is decoded in three-base sections called codons.
- The codons code for one of 20 amino acids.
- There are 64 different codons (4^3) so several different codons can specify the same amino acid, or none at all.



Decoding mRNA

- Scientist knew that amino-acid sequences were specified by sequences of bases, but they did not know how the mRNA sequence was "read" to make proteins.
- Marshall Nirenberg conducted experiments using E. coli. He inserted an RNA strand made up of only uracil chain (UUUUU...) which produced an amino acid chain made up of only phenylalanine (Phe-Phe-Phe...).

Decoding mRNA (continues)

This was the first step in understanding the genetic code.

Nirenberg's mRNA sequence: UUUUUUUUUUUUUUUU

Sample mRNA sequence: AUGGCCUUAGGUACUAAAU

Questions he couldn't answer with this experiment are:

- How long are codons ("words")? Are they the same length?
- Do the codons overlap?
- Are codons consecutive bases? Every other?
- Is there "punctuation" between codons?

Decoding mRNA (continued)

- H. Gobind Khorana performed a similar experiment to answer the previous questions.

- He used more complex mRNA sequences to obtain the following data:

mRNA	Polypeptide(s) synthesized
$(UC)_n$	(Ser-Leu)
$(UG)_n$	(Val-Cys)
$(AC)_n$	(Thr-His)
$(AG)_n$	(Arg-Glu)
$(UUC)_n$	(Ser-Ser) and (Leu-Leu) and (Phe-Phe)
$(UUG)_n$	(Leu-Leu) and (Val-Val) and (Cys-Cys)
$(AAG)_n$	(Arg-Arg) and (Lys-Lys) and (Glu-Glu)
$(CAA)_n$	(Thr-Thr) and (Asn-Asn) and (Gln-Gln)
$(UAC)_n$	(Thr-Thr) and (Leu-Leu) and (Tyr-Tyr)
$(AUC)_n$	(Ile-Ile) and (Ser-Ser) and (His-His)
$(GUA)_n$	(Ser-Ser) and (Val-Val)
$(GAU)_n$	(Asp-Asp) and (Met-Met)
$(UAUC)_n$	(Tyr-Leu-Ser-Ile)
$(UUAC)_n$	(Leu-Leu-Thr-Tyr)
$(GAUA)_n$	None
$(GUAA)_n$	None

Your Task...should you accept

Use the previous data from the two experiments to figure out the genetic code. In particular, show that the code:

- is consecutive, non-overlapping triplets.
- has no punctuation.
- often has a degenerate third position.
- contains some codons meaning "stop".

Also, explain why

- coding with doublets is impossible
- the only possible reading restart interval less than 9 is 3.

Sample mRNA sequence:

AUGGCCUUAGGUACUAAAU