DNA and genetic information

- DNA carries plans for the primary structure of nucleic acids (DNA, RNA) and proteins.
- DNA of single cell has capacity over 1 million pages of text (900 copies of our textbook!)
- however, only about 1% of DNA ever gets translated into proteins- equivalent to about 1 large book.

DNA is not all the information necessary in an organism (consider “Jurassic Park”)

Information in binary computers

- two basic symbols :1 or 0 (“bits”)
- physical form of bits varies with medium
- 8 bit words = bytes (e.g 10110111)
- 2^8 = 256 unique bytes are possible
- This is a large enough set of symbols to represent characters in language

Information in binary computers

- ASCII code- American Standard Code for Information Interchange
- Each byte was assigned to a particular symbol (letter, numeral, etc.)
- Therefore, series of bytes can represent English sentences, math equations, etc.

Hierarchies of symbols

English	computer	genetics
letter (26)	bit (2)	nucleotide (4)
word (1-28 letters)	byte (8 bits)	codon (3 nucleotides)
sentence	line	gene
book	program	genome

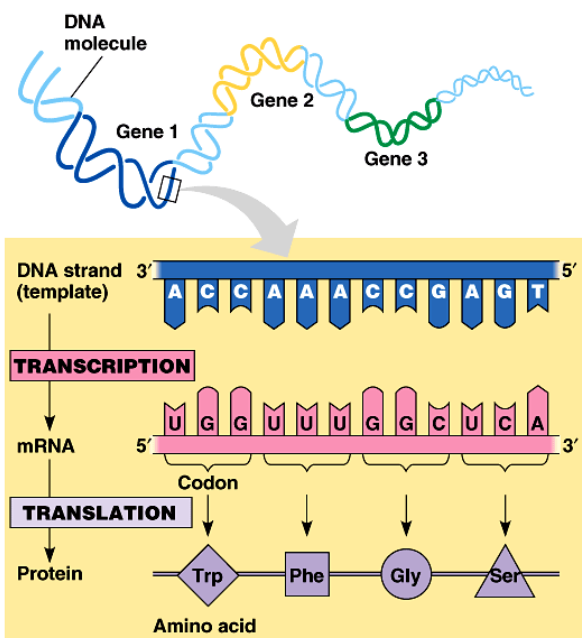
Information in organisms

- DNA and RNA- polymers of nucleotides
- 4 letter alphabet = 4 kinds of nucleotides
  DNA: A, T, G, C RNA: A, U, G, C
- sequence of nucleotides in a gene specifies the sequence of amino acids in a protein.

Genetic code

- "words" (codons or triplets) are 3 letters long in genetic code
- each group of 3 nucleotides corresponds to one amino acid.
- A nucleotide sequence (sequence of codons) can be “translated” into an amino acid sequence, i.e., a peptide or protein

The triplet code



Genetic code

- the <u>Genetic Code</u> is the correspondence between triplets and amino acids
- deciphered in early 60's by Marshall Nirenberg
- He used synthetic polynucleotides and a cell-free translation system (e.g. poly-A gave poly-phenylalanine)

The dictionary of the genetic code

First base (5′ end)	Second base: U	Second base: C	Second base: A	Second base: G	Third base (3′ end)
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
U	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
U	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
U	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
C	CUC Leu	CCC Pro	CAC His	CGC Arg	C
C	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
C	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
A	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
A	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
A	AUG Met or start	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
G	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
G	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
G	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

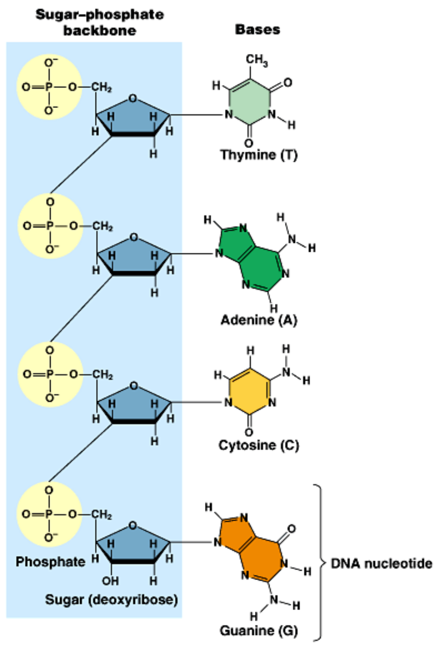
The genetic code is universal

- This fact indicates a single origin for all living things.
- A human gene placed in a bacterium or yeast cell can yield the same protein.
- Biotechnology

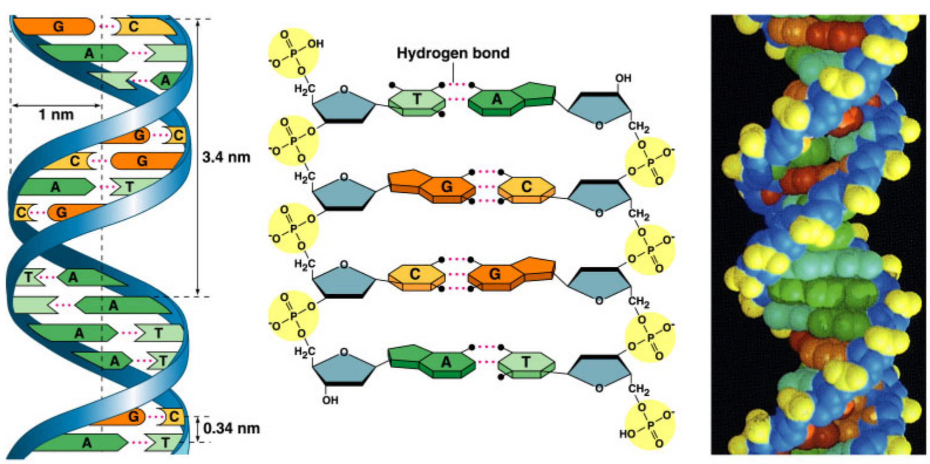
Genetic information undergoes 3 processes:

- Replication to copy the DNA for cell reproduction
- Transcription of genes into messenger RNA molecules for translation by ribosomes
- Translation of mRNA by ribosomes into protein

The structure of a DNA stand

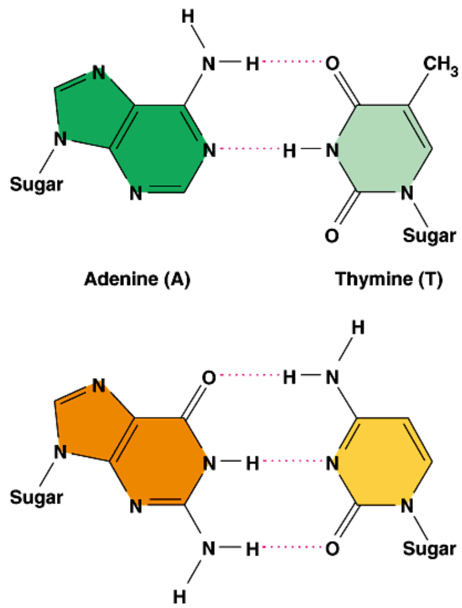


The double helix

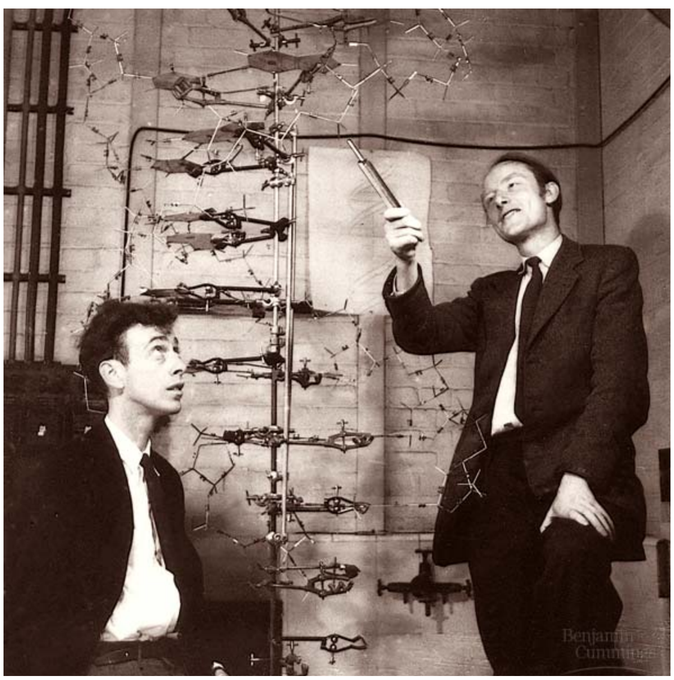


(a) Key features of DNA structure (b) Partial chemical structure (c) Space-filling model

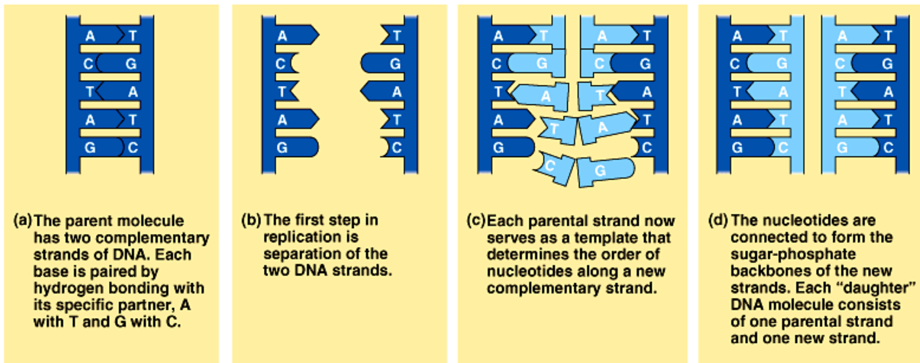
Base pairing in DNA



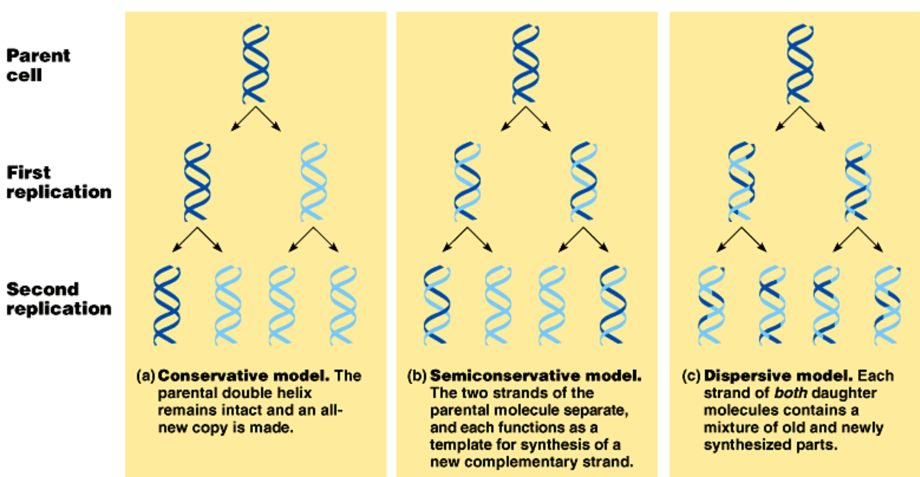
James Watson and Francis Crick 1953



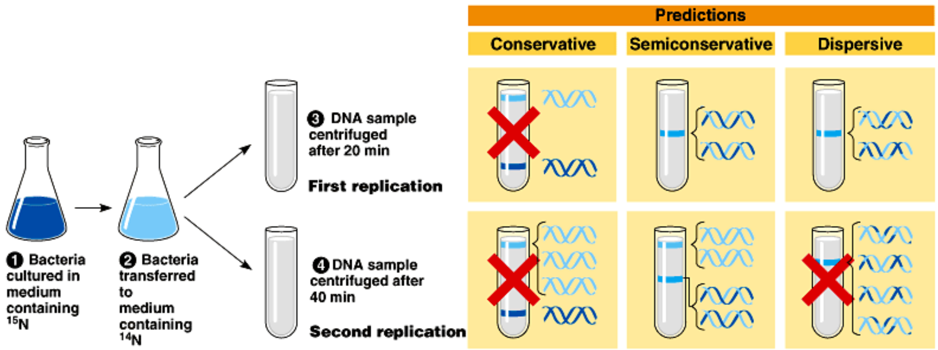
A model for DNA replication: the basic concept



Three alternative models of DNA replication



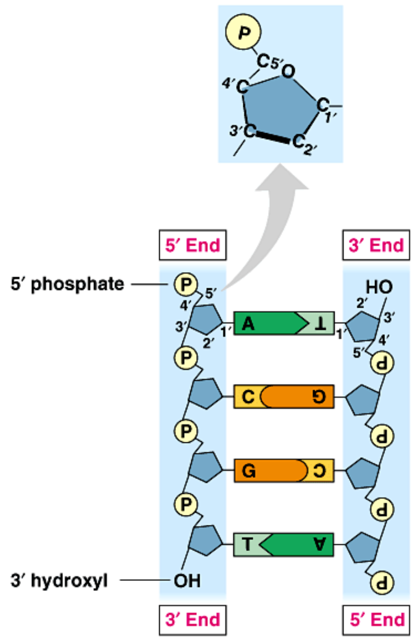
The Meselson-Stahl experiment tested three models of DNA replication (1958)



Structure of DNA

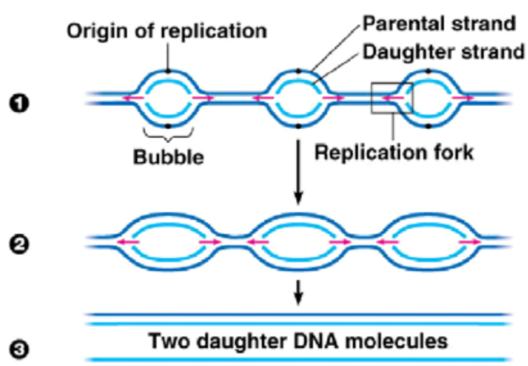
- Base pairing
- Complementary strands
- 5' and 3' ends, antiparallel
- Leading and lagging strands
- Template and complementary strands (=sense and antisense)

The two strands of DNA are antiparallel

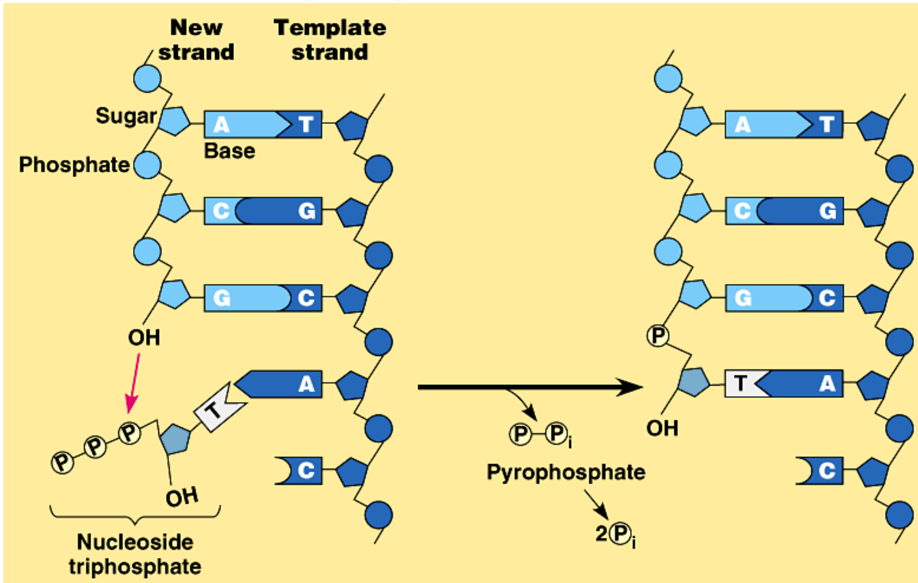


Replication

- Replication “origins” and “bubbles” -single in prokaryotes, multiple in eukaryotes



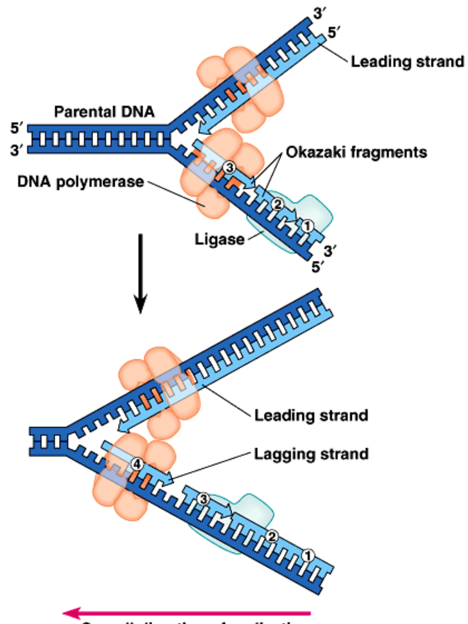
Nucleoside triphosphates added



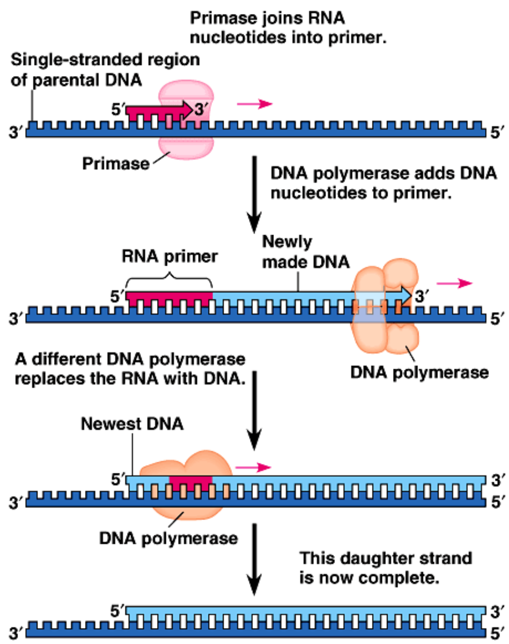
Replication

- DNA synthesis always starts with RNA primer (10 bases) laid first, later replaced with DNA
- Leading strand is continuous
- Lagging strand is discontinuous, in Okasaki fragments (100-200 bases long)
- Ligase joins the fragments

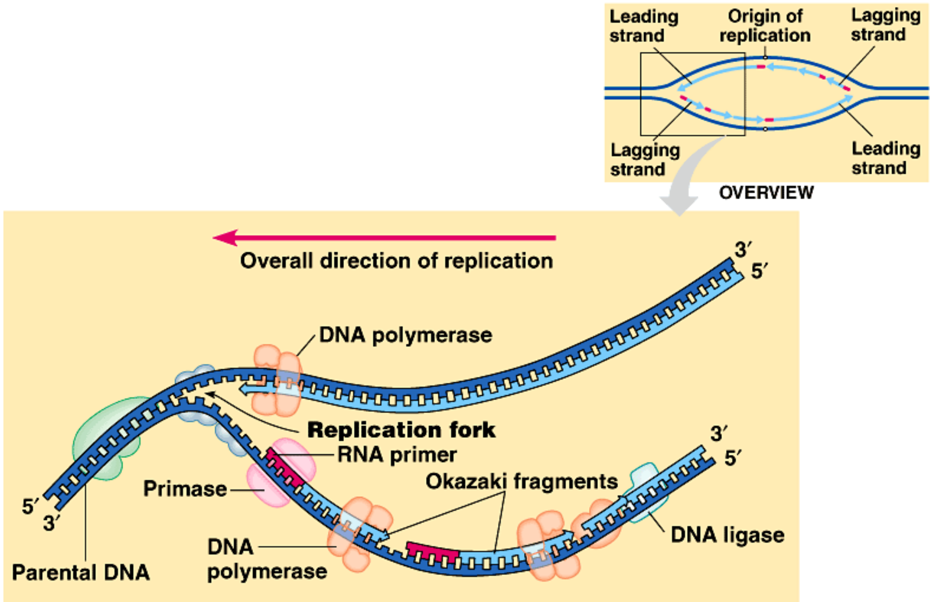
Synthesis of leading and lagging strands during DNA replication



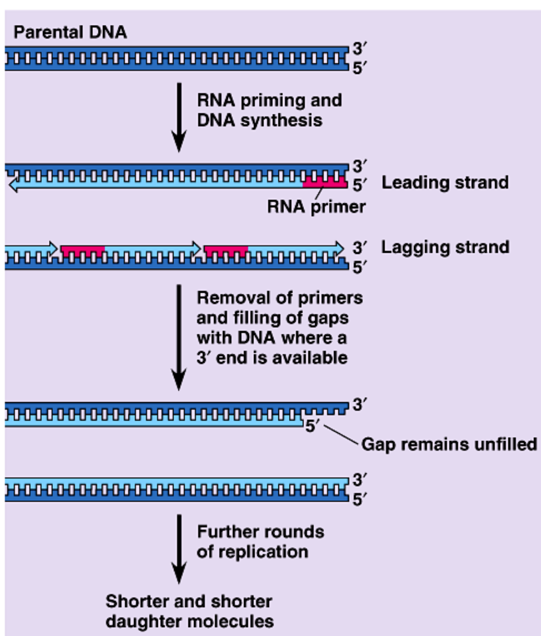
Priming DNA synthesis with RNA

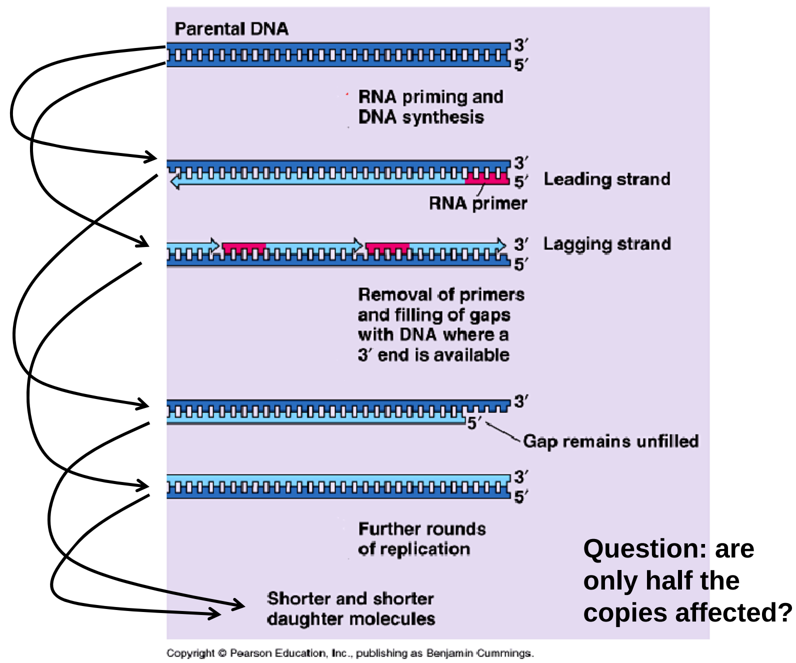


A summary of DNA replication



The end-replication problem

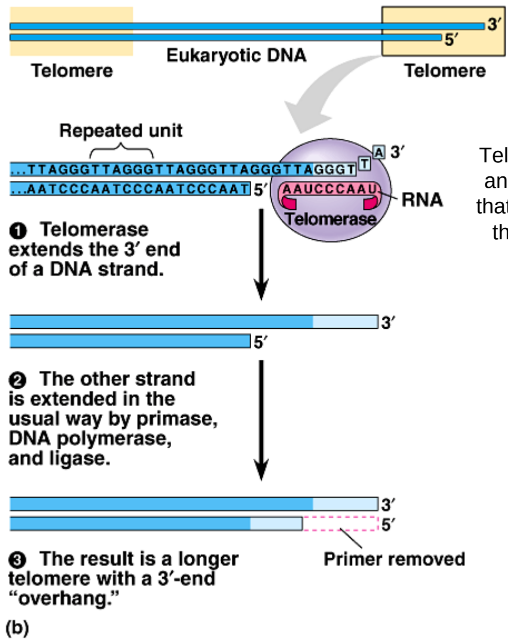




End-replication & telomeres

- DNA polymerase can't finish the 5' ends
- DNA would get 10 bases shorter each time replicated
- <u>Telomeres</u> are "expendable" repetitive (TTAGGG) sequences at end of chromosomes
- <u>Telomerase</u> restores telomers, present in germ-line cells and cancerous cells

Telomeres and telomerase



Telomerase includes an RNA component that is the pattern for the telomere DNA

Telomeres and telomerase: Telomeres of mouse chromosomes