CS262

A Zero-Knowledge Based Introduction to Biology

Biology

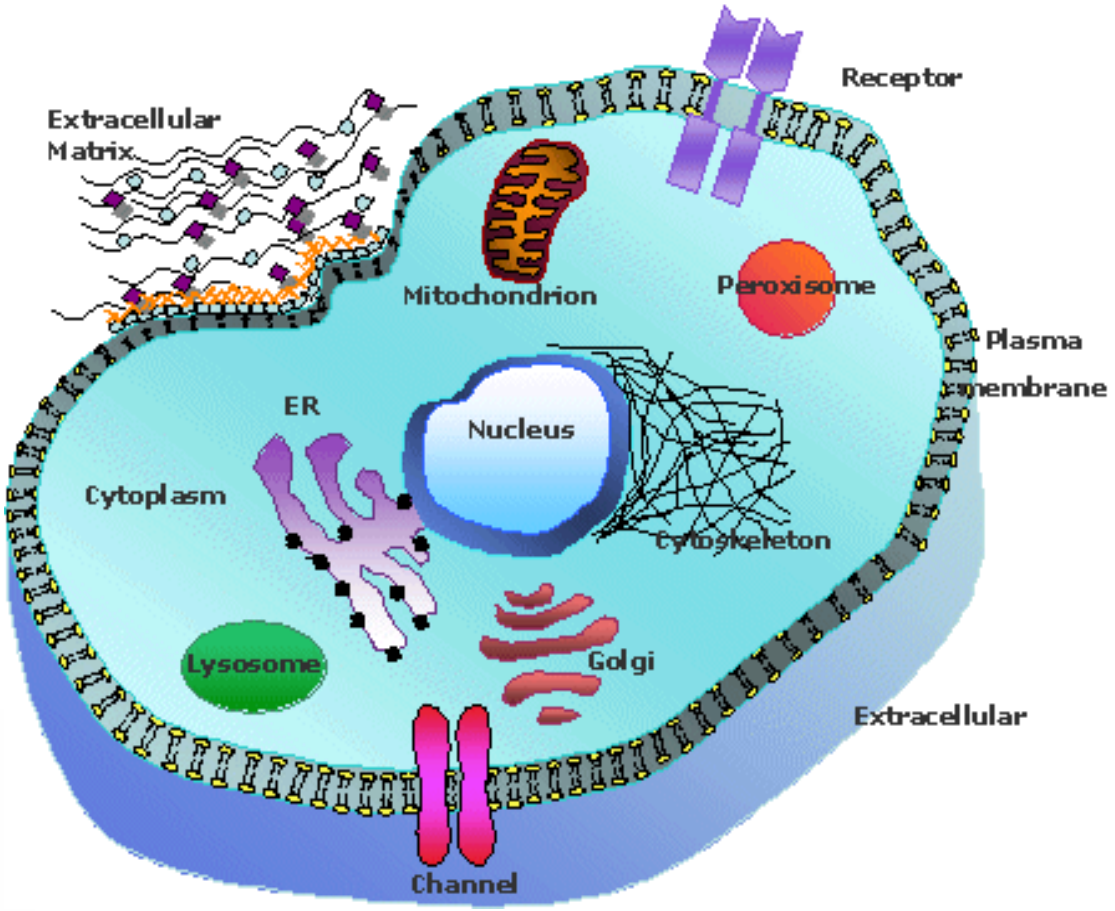
- From the greek word *βίος* = life
- Timeline:
  - 1683 – discovery of **bacteria**
  - 1858 – **Darwin**'s natural selection
  - 1865 – **Mendel**'s laws
  - 1953 – **double helix** suggested by Watson-Crick
  - 1955 – discovery of DNA and RNA **polymerase**
  - 1978 – sequencing of first genome (5kb virus)
  - 1983 – invention of **PCR**
  - 1990 – discovery of **RNAi**
  - 2000 – human genome (draft)

How to learn some?

- Online sources
  - Wikipedia
    - http://www.wikipedia.org/
  - John Kimball's Biology Pages
    - http://biology-pages.info/
- Cold Spring Harbor Meetings
  - CSHL Biology of Genomes
  - CSHL Genome Informatics
- Hang out with biologists

The Cell

cell, nucleus, cytoplasm, mitochondrion



How many?

- Cells in the human body:

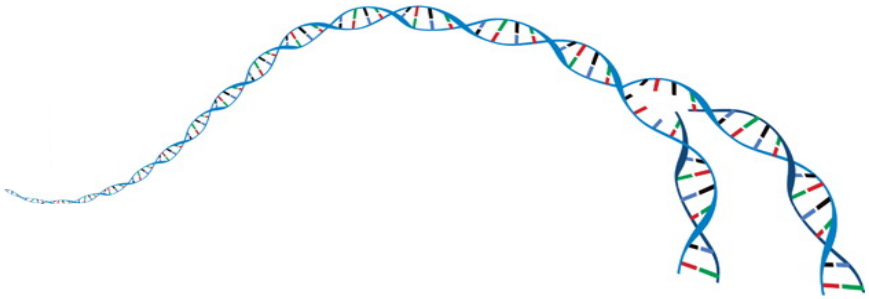
  $\sim 10^{14}$ (100 trillion)

  $\sim 10^{15}$ bacterial cells!

Chromosomes

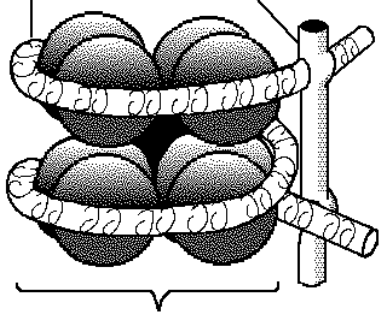
histone, nucleosome, chromatin, chromosome, centromere, telomere

ome



DNA H1

chromatin



~146bp

H2A, H2B, H3, H4

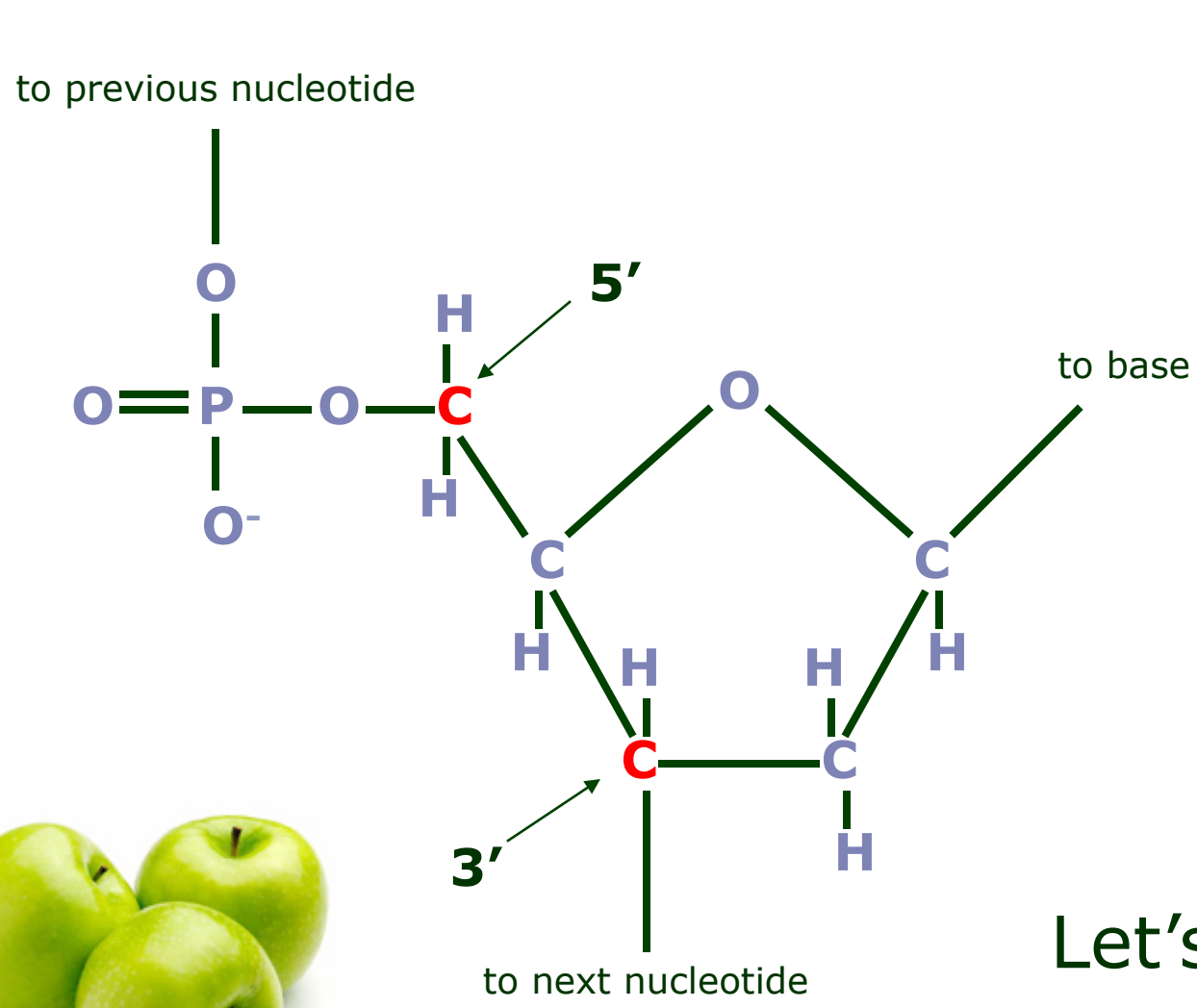
How many?

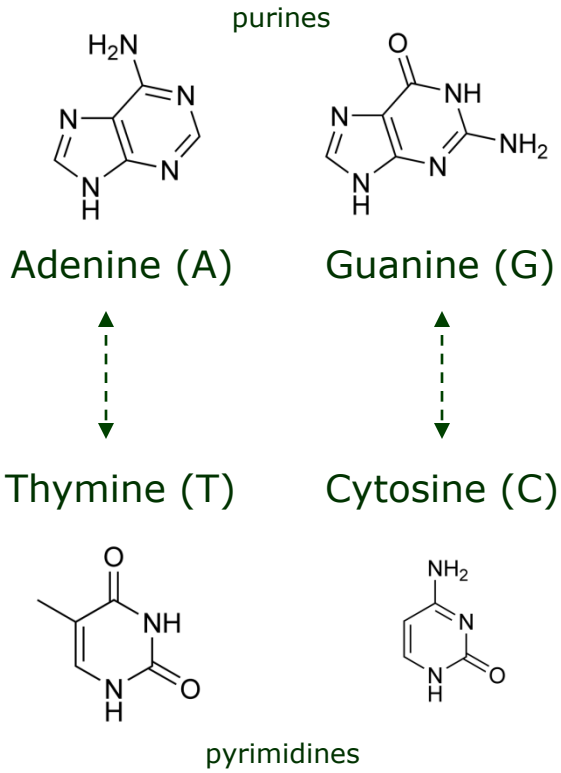
- Chromosomes in a human cell:

  46 (2x22 + X/Y)

Nucleotide

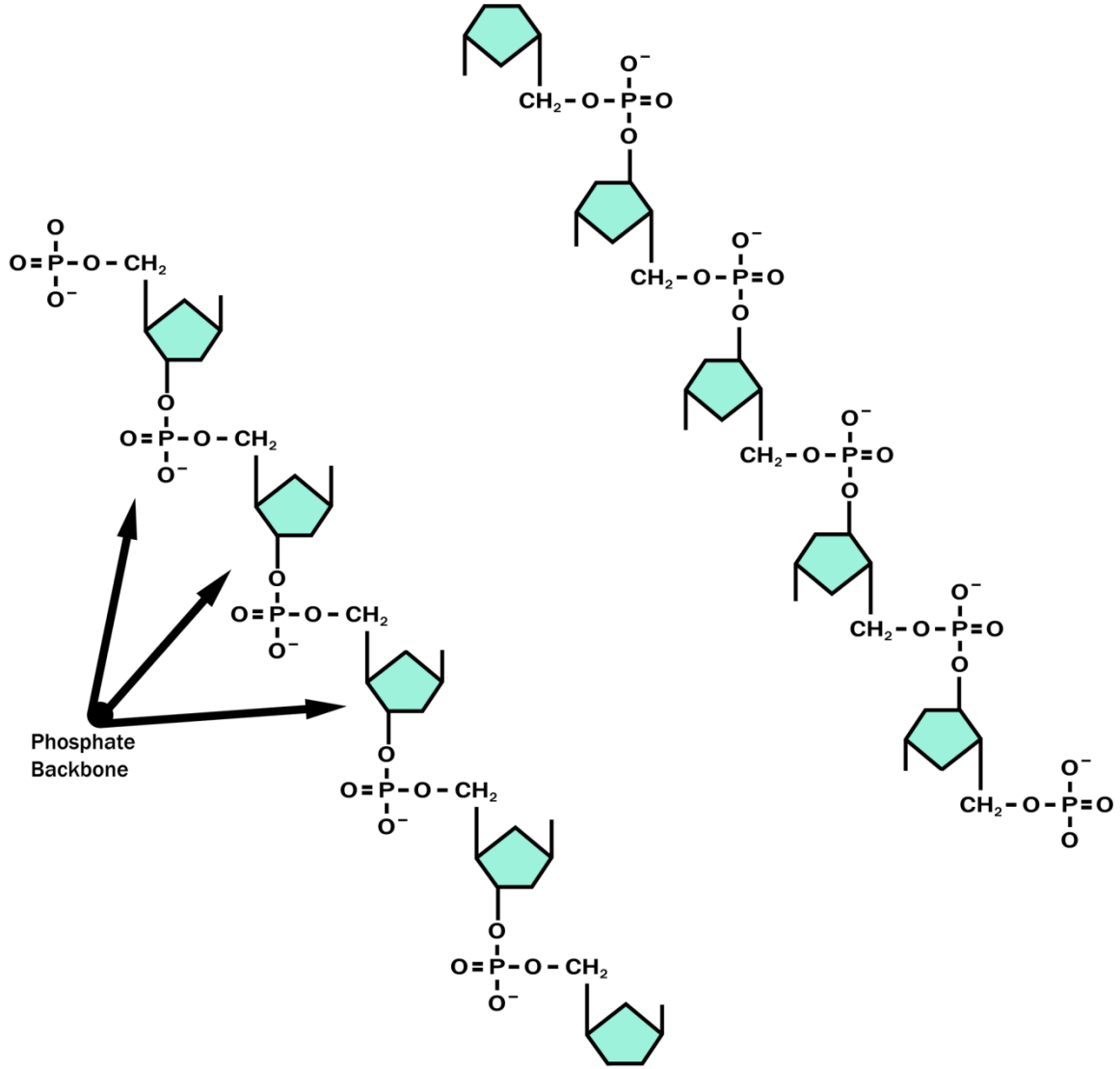
deoxyribose, nucleotide, base, A, C, G, T, purine, pyrimidine, 3’, 5’





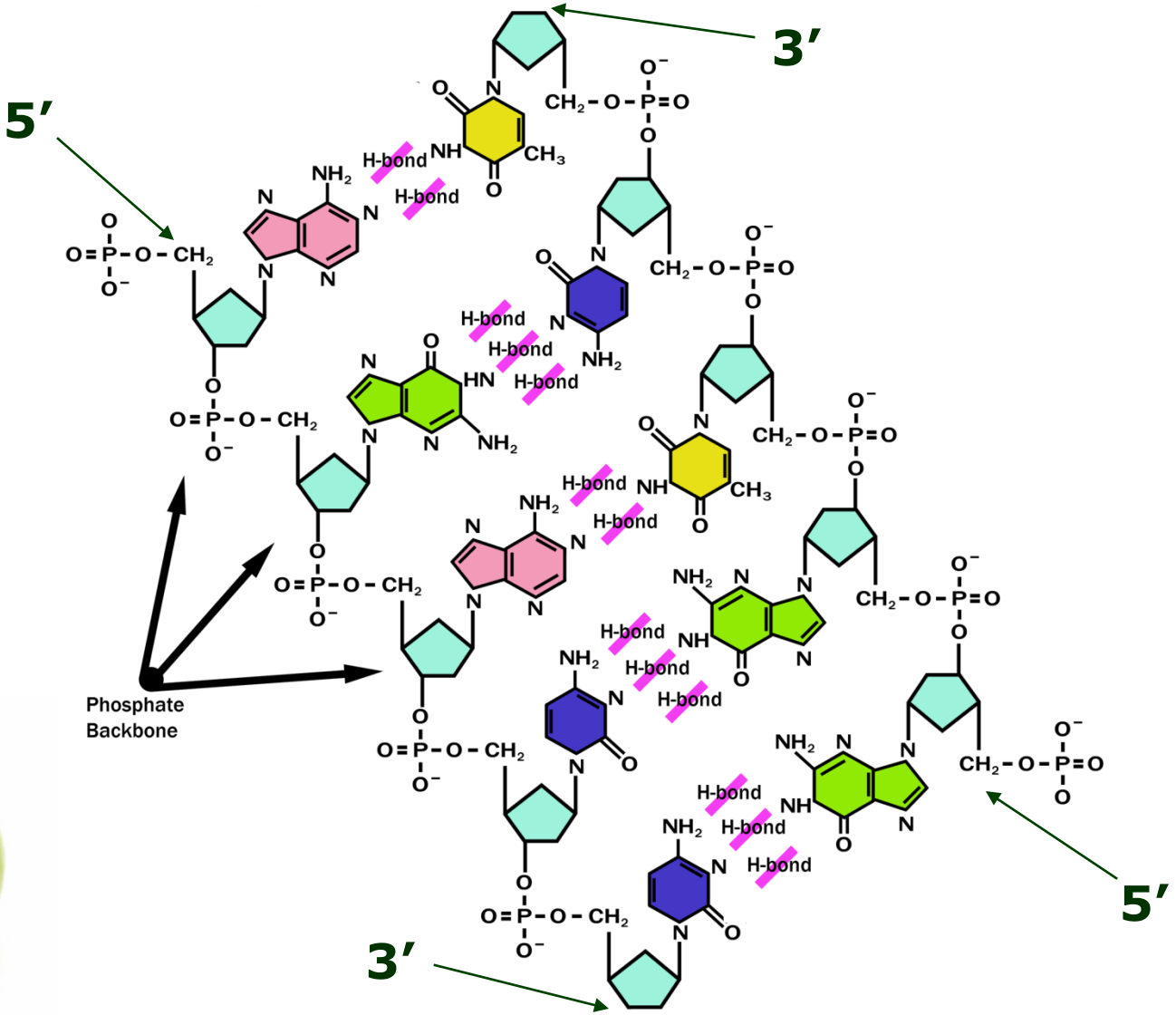
Let’s write “AGACC”!

“AGACC” (backbone)



“AGACC” (DNA)

deoxyribonucleic acid (DNA)

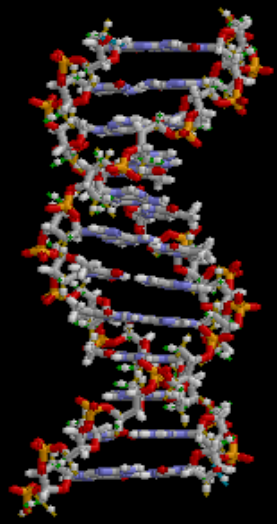


DNA is double stranded

strand, reverse complement

5’ — AGACC — 3’

3’ — TCTGG — 5’

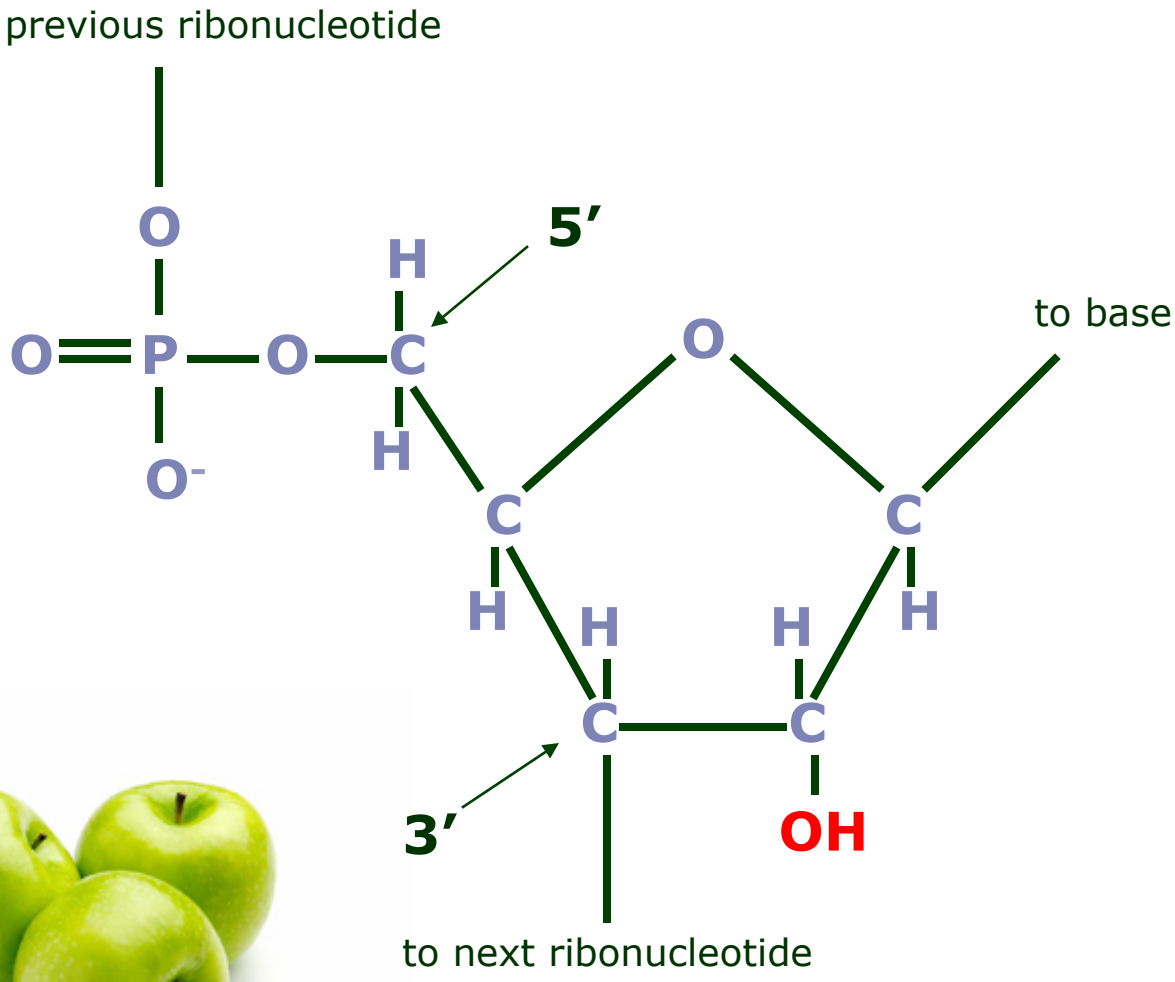


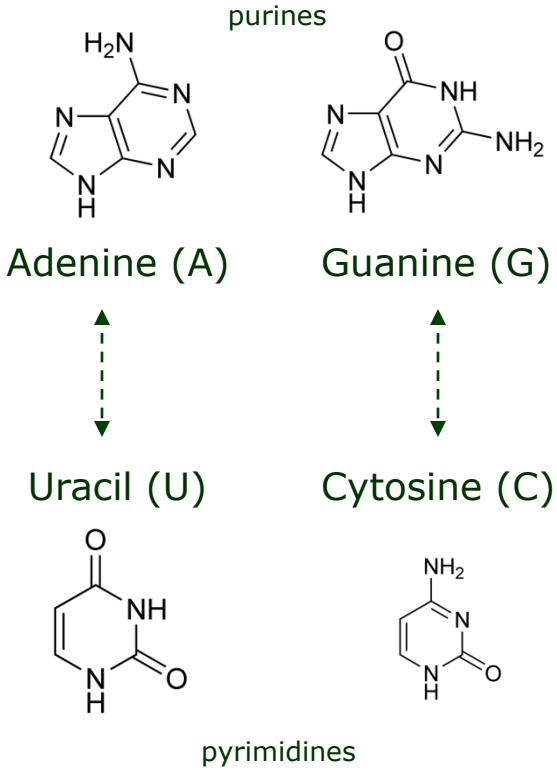
DNA is always written 5’ to 3’

AGACC or GGTCT

RNA

ribose, ribonucleotide, U





How many?

- Nucleotides in the human genome:

  ~ 3 billion

Genes & Proteins

gene, transcription, translation, protein

Double-stranded DNA

5′ TAGGATCGACTATATGGGATTACAAAGCATTTAGGGA...TCACCCTCTCTAGACTAGCATCTATATAAAACAGAA 3′

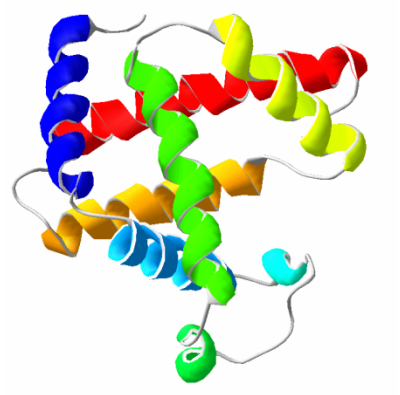
3′ ATCCTAGCTGATATACCCTAATGTTTCGTAAATCCCT...AGTGGGAGAGATCTGATCGTAGATATATTTTGTCTT 5′

(transcription)

Single-stranded RNA AUGGGAUUACAAAGCAUUUAGGGA...UCACCCUCUCUAGACUAGCAUCUAUAUAA

(translation)

protein



How many?

- Genes in the human genome:

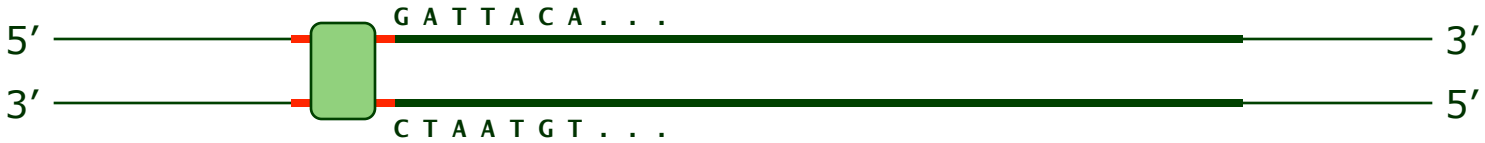
  ~ 20,000 – 25,000

Gene Transcription

promoter

Gene Transcription

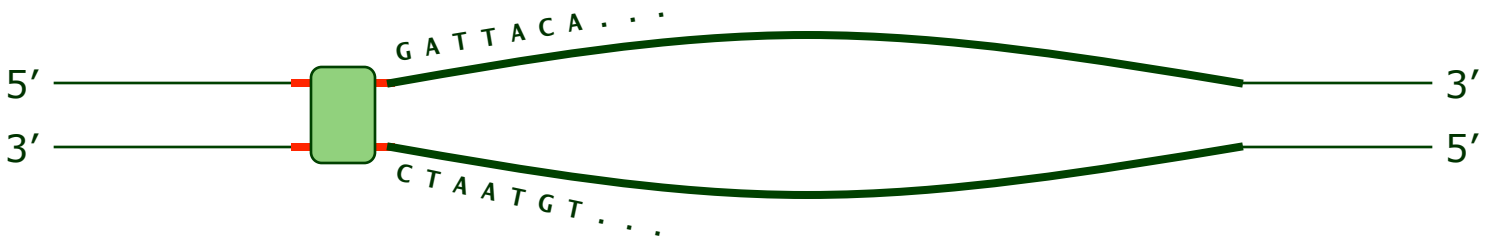
transcription factor, binding site, RNA polymerase



Transcription factors recognize **transcription factor binding sites** and bind to them, forming a complex.

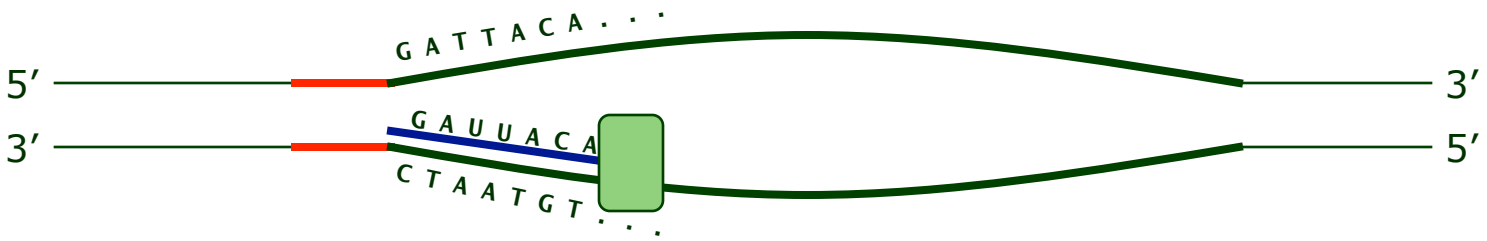
RNA polymerase binds the complex.

Gene Transcription



The two strands are separated

Gene Transcription

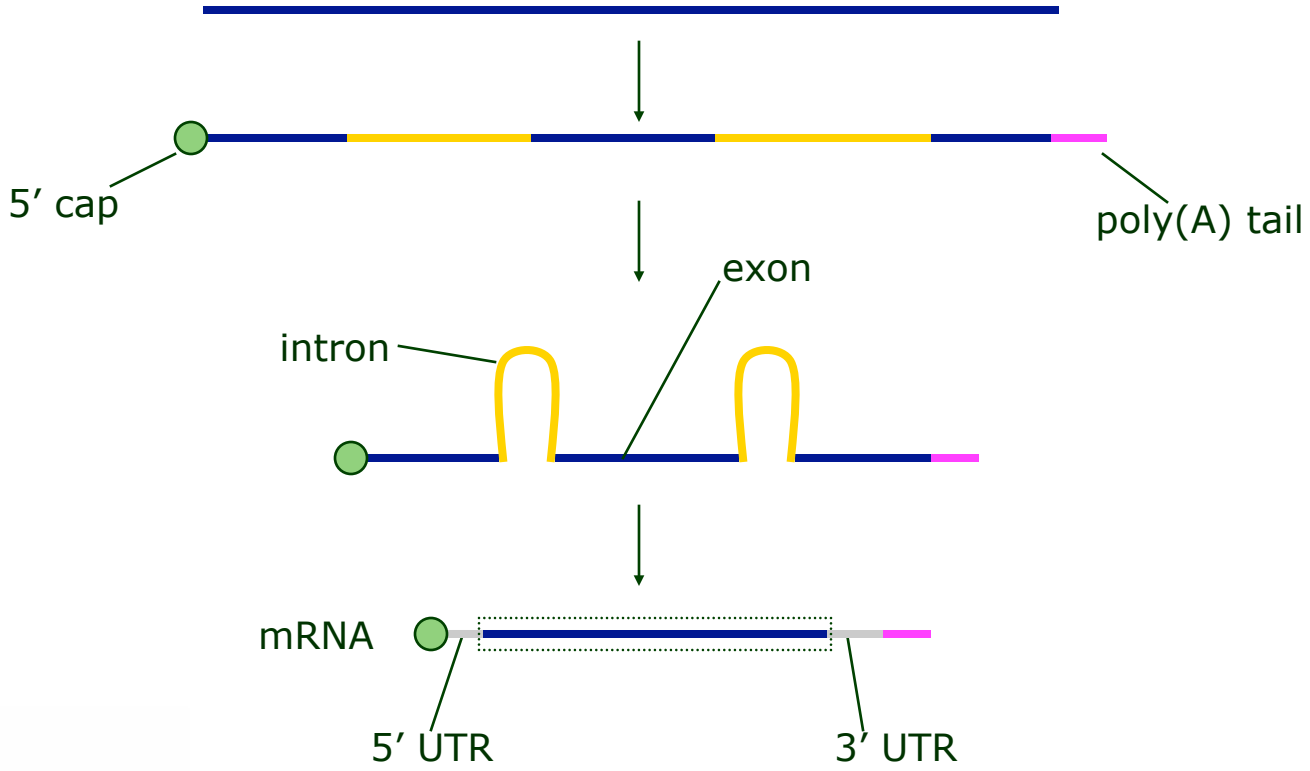


An RNA copy of the 5′→3′ sequence is created from the 3′→5′ template

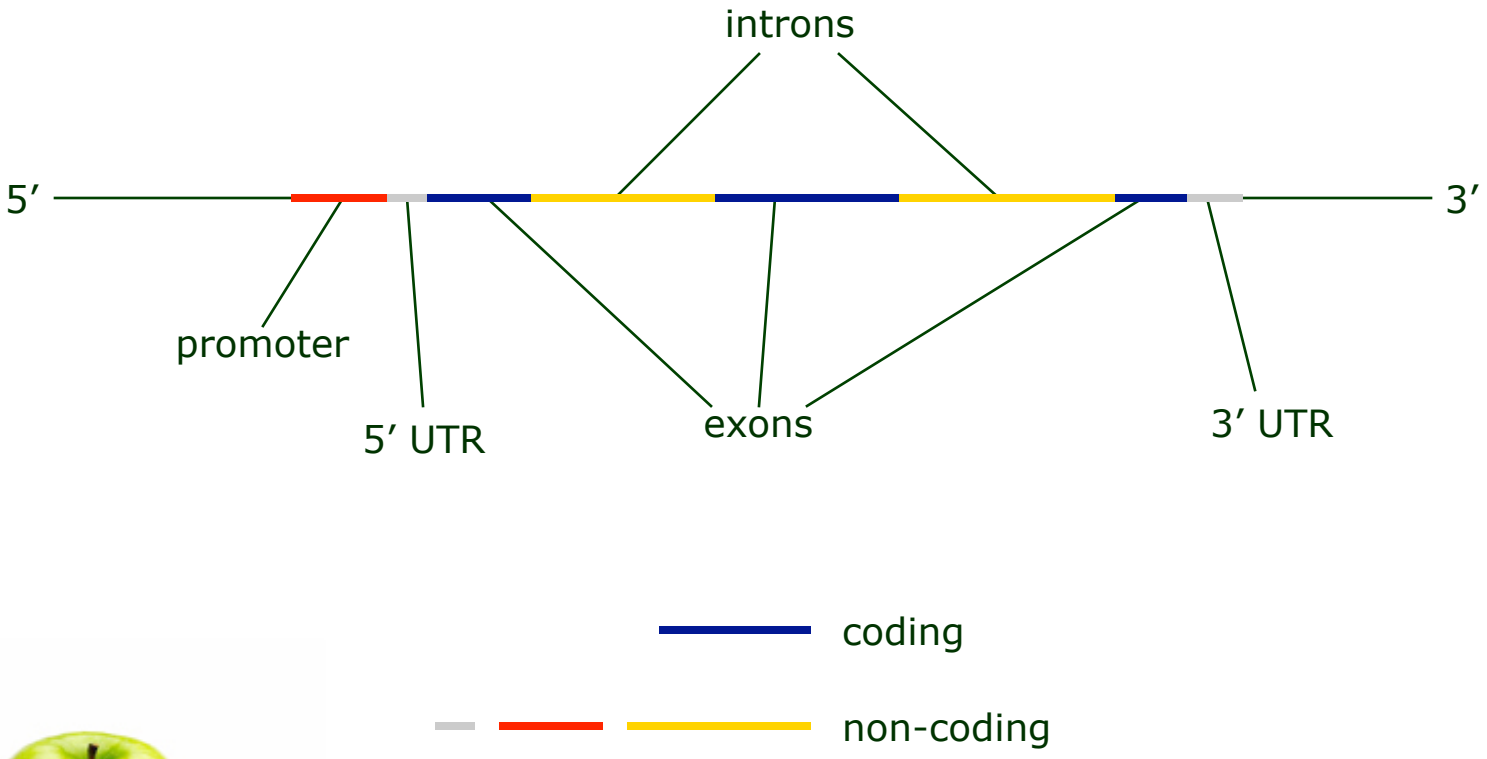
Gene Transcription

RNA Processing

5’ cap, polyadenylation, exon, intron, splicing, UTR, mRNA



Gene Structure

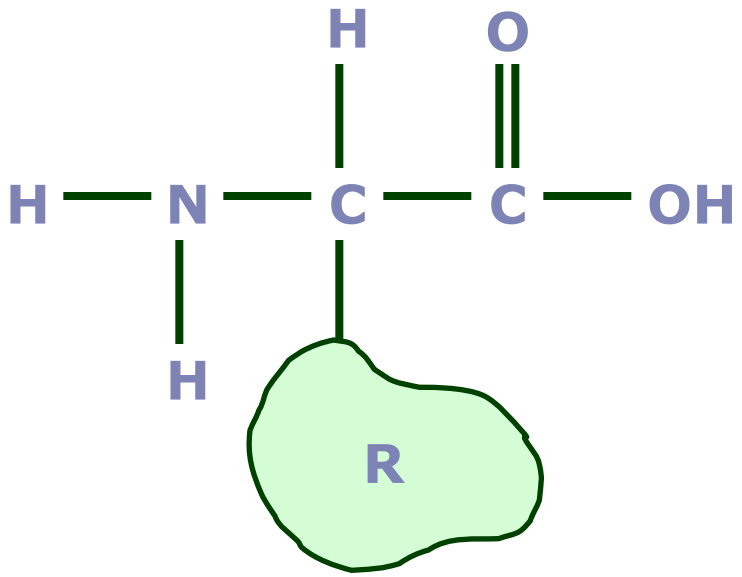


How many?

- Exons per gene:
  ~ 8 on average (max: 148)
- Nucleotides per exon:
  170 on average (max: 12k)
- Nucleotides per intron:
  5,500 on average (max: 500k)
- Nucleotides per gene:
  45k on average (max: 2,2M)

Amino acid

amino acid

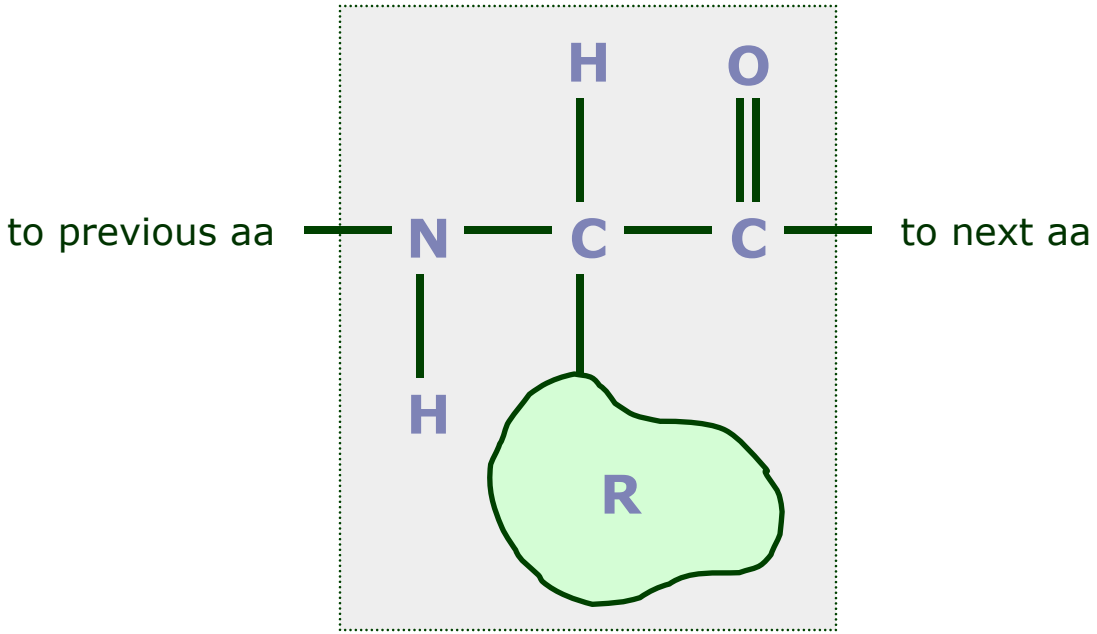


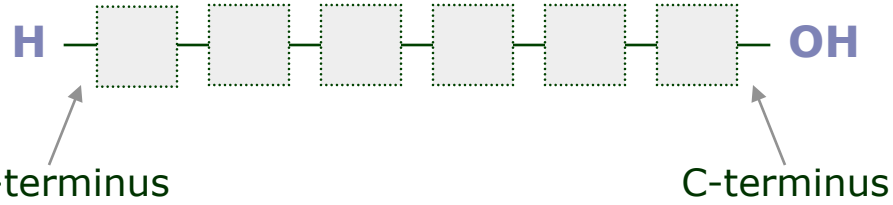
Alanine
Arginine
Asparagine
Aspartate
Cysteine
Glutamate
Glutamine
Glycine
Histidine
Isoleucine
Leucine
Lysine
Methionine
Phenylalanine
Proline
Serine
Threonine
Tryptophan
Tyrosine
Valine

There are 20 standard amino acids

Proteins

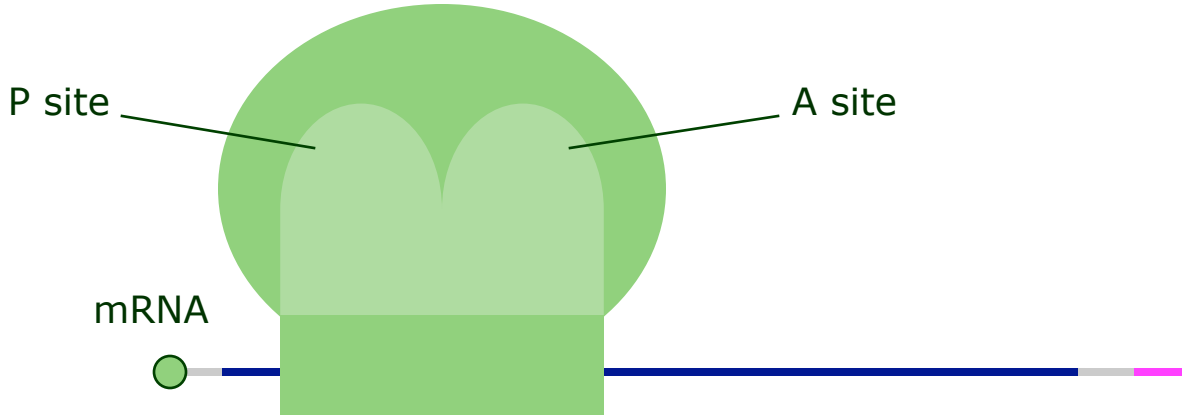
N-terminus, C-terminus





Translation

ribosome, codon



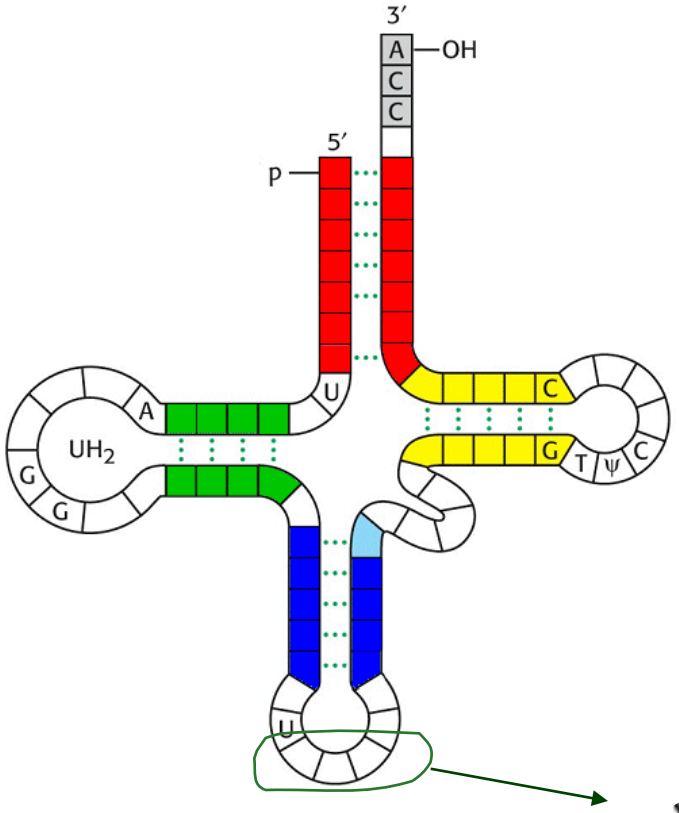
The ribosome synthesizes a protein by reading the mRNA in triplets (codons). Each codon is *translated* to an amino acid.

The Genetic Code

	U	C	A	G	
U	UUU **Phenylalanine** (Phe)	UCU **Serine** (Ser)	UAU **Tyrosine** (Tyr)	UGU **Cysteine** (Cys)	**U**
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	**C**
	UUA **Leucine** (Leu)	UCA Ser	UAA **STOP**	UGA **STOP**	**A**
	UUG Leu	UCG Ser	UAG **STOP**	UGG **Tryptophan** (Trp)	**G**
C	CUU **Leucine** (Leu)	CCU **Proline** (Pro)	CAU **Histidine** (His)	CGU **Arginine** (Arg)	**U**
	CUC Leu	CCC Pro	CAC His	CGC Arg	**C**
	CUA Leu	CCA Pro	CAA **Glutamine** (Gln)	CGA Arg	**A**
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	**G**
A	AUU **Isoleucine** (Ile)	ACU **Threonine** (Thr)	AAU **Asparagine** (Asn)	AGU **Serine** (Ser)	**U**
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	**C**
	AUA Ile	ACA Thr	AAA **Lysine** (Lys)	AGA **Arginine** (Arg)	**A**
	AUG **Methionine** (Met) or **START**	ACG Thr	AAG Lys	AGG Arg	**G**
G	GUU **Valine** (Val)	GCU **Alanine** (Ala)	GAU **Aspartic acid** (Asp)	GGU **Glycine** (Gly)	**U**
	GUC Val	GCC Ala	GAC Asp	GGC Gly	**C**
	GUA Val	GCA Ala	GAA **Glutamic acid** (Glu)	GGA Gly	**A**
	GUG Val	GCG Ala	GAG Glu	GGG Gly	**G**

Translation (tRNA)

tRNA, anticodon

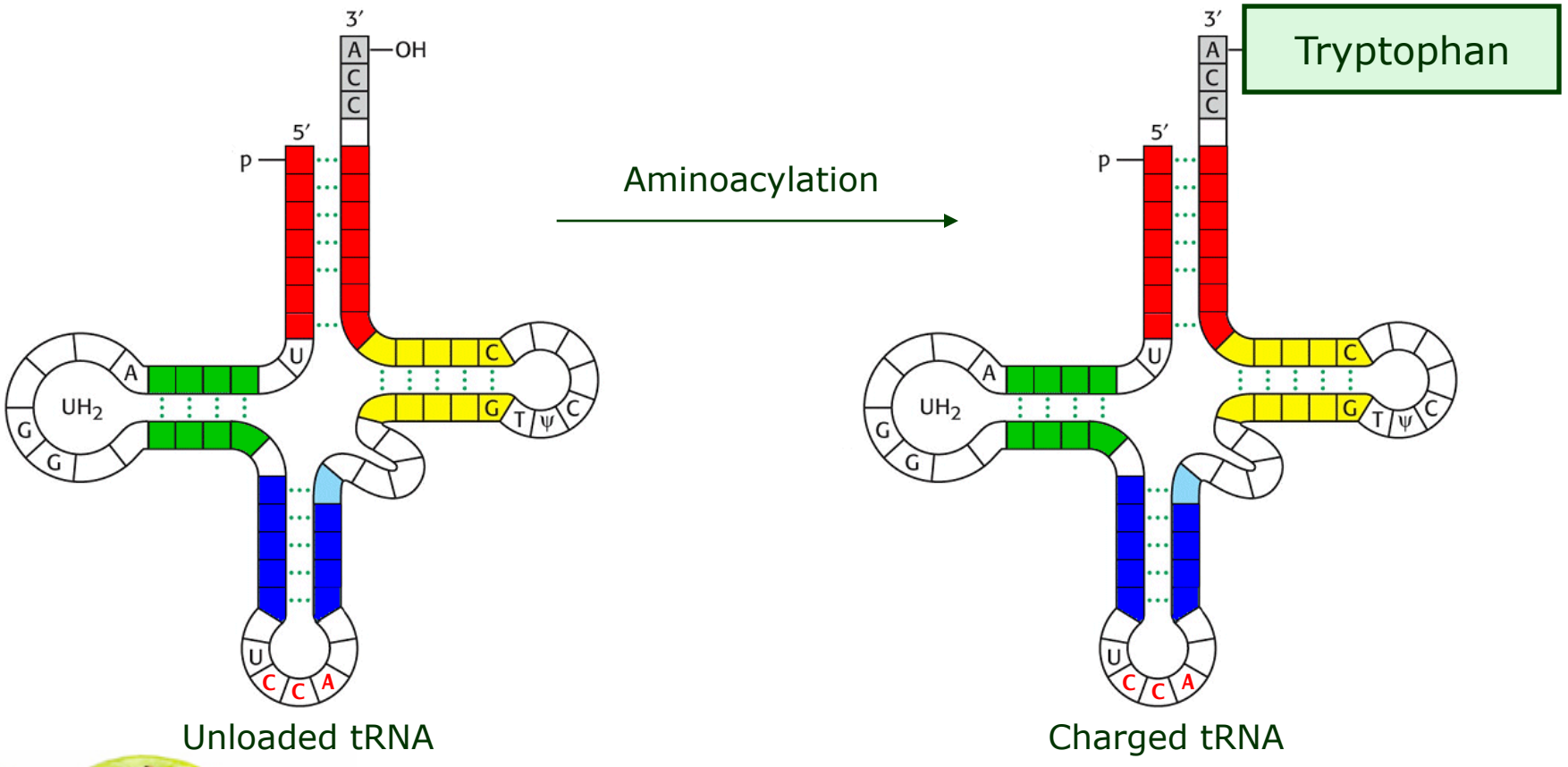


(Tryptophan codon: UGG)

Tryptophan anticodon

Translation (tRNA)

aminoacylation



Unloaded tRNA

Charged tRNA

Translation

5′ . . . A U U A U G G C C U G G A C U U G A . . . 3′

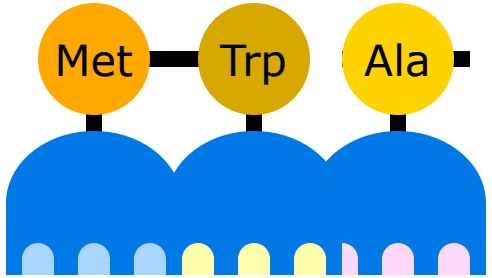
UTR | Met | Ala | Trp | Thr

Start Codon

Translation

5′ . . . A U U A U G G C C U G G A C U U G A . . . 3′

Translation



5′ . . . A U U A U G G C C U G G A C U U G A . . . 3′

Errors?

mutation

- What if the transcription / translation machinery makes mistakes?
- What is the effect of **mutations** in coding regions?

Reading Frames

reading frame

G C U U G U U U A C G A A U U A G

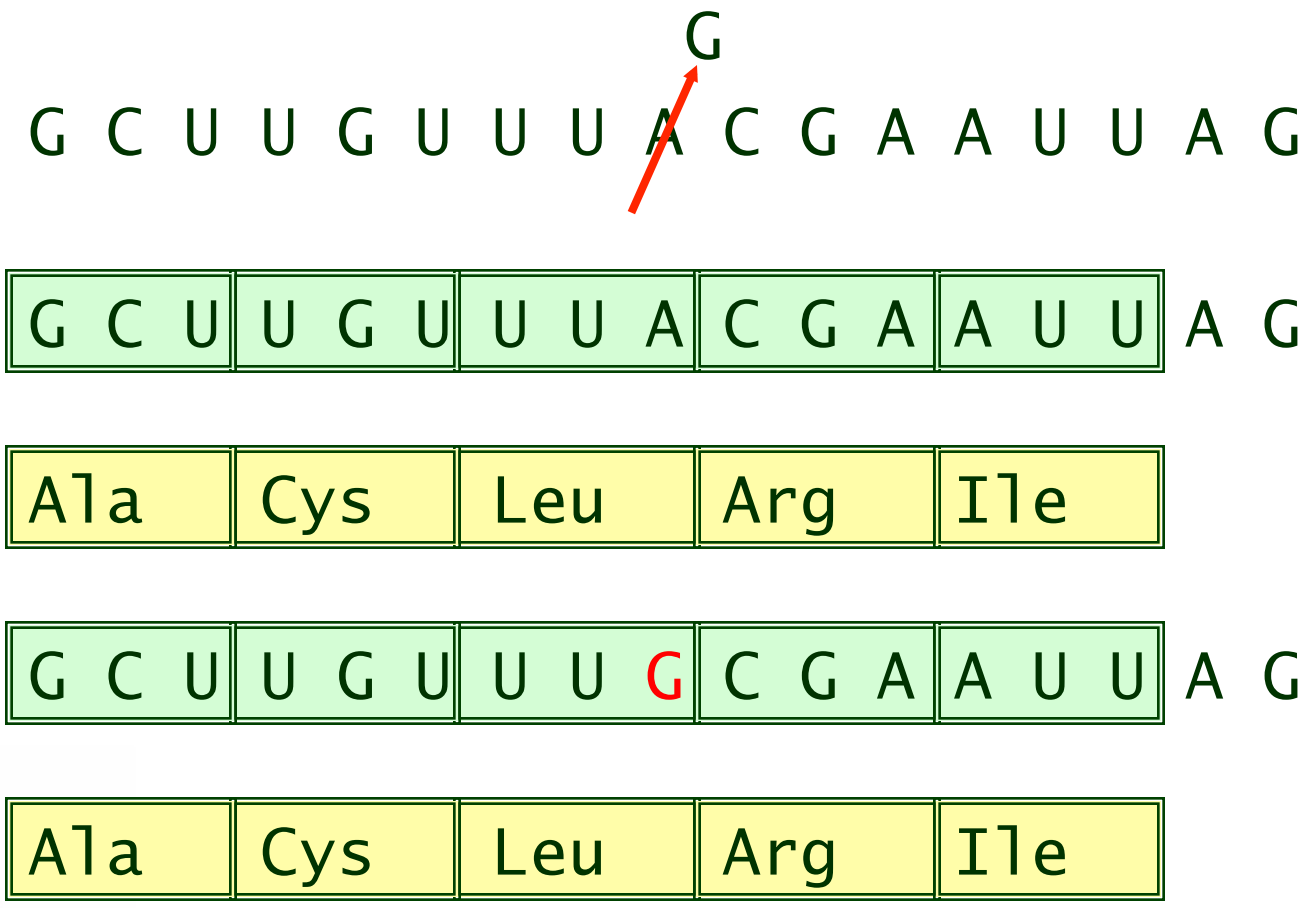
G C U U G U U U A C G A A U U A G

G C U U G U U U A C G A A U U A G

G C U U G U U U A C G A A U U A G

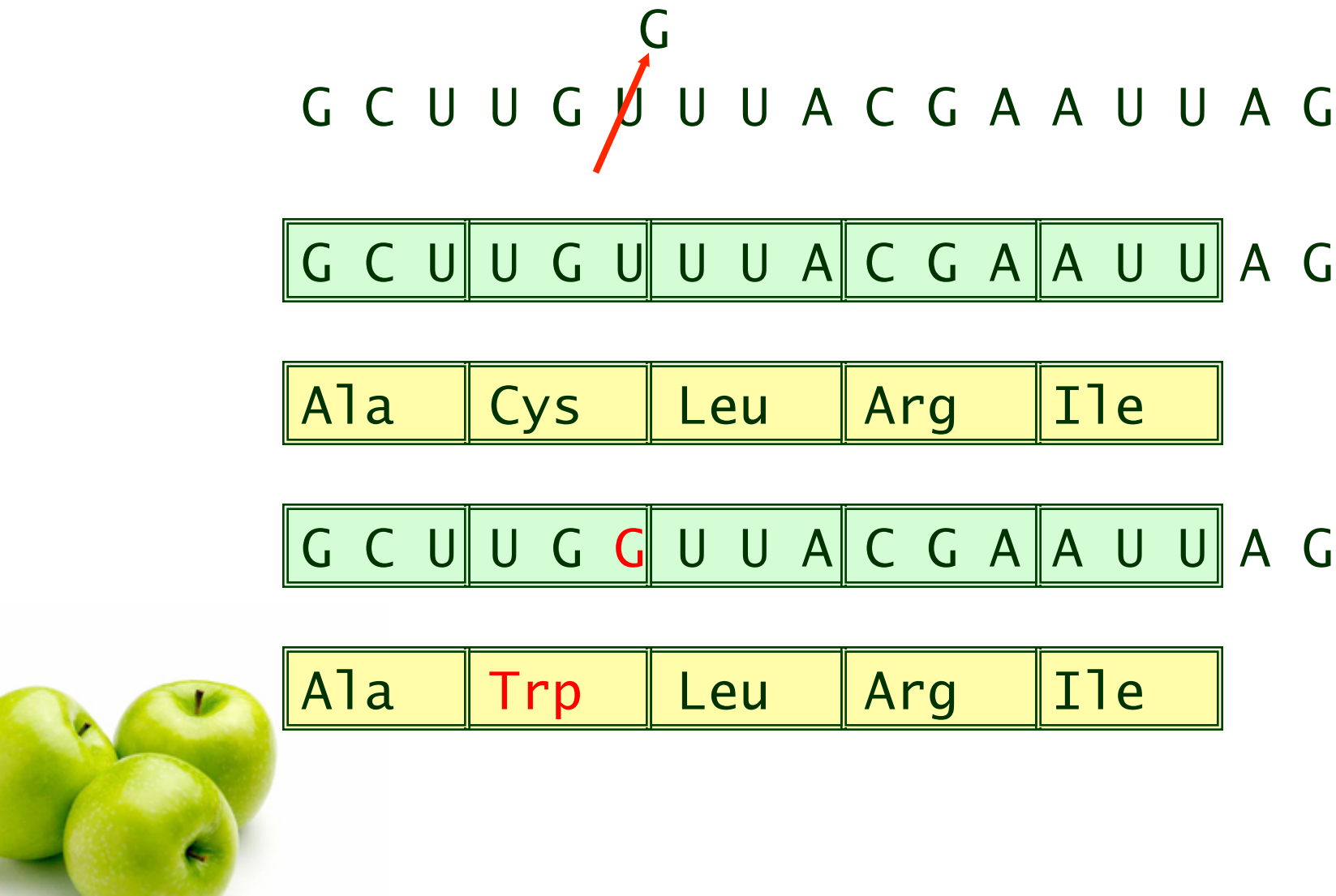
Synonymous Mutation

synonymous (silent) mutation, fourfold site



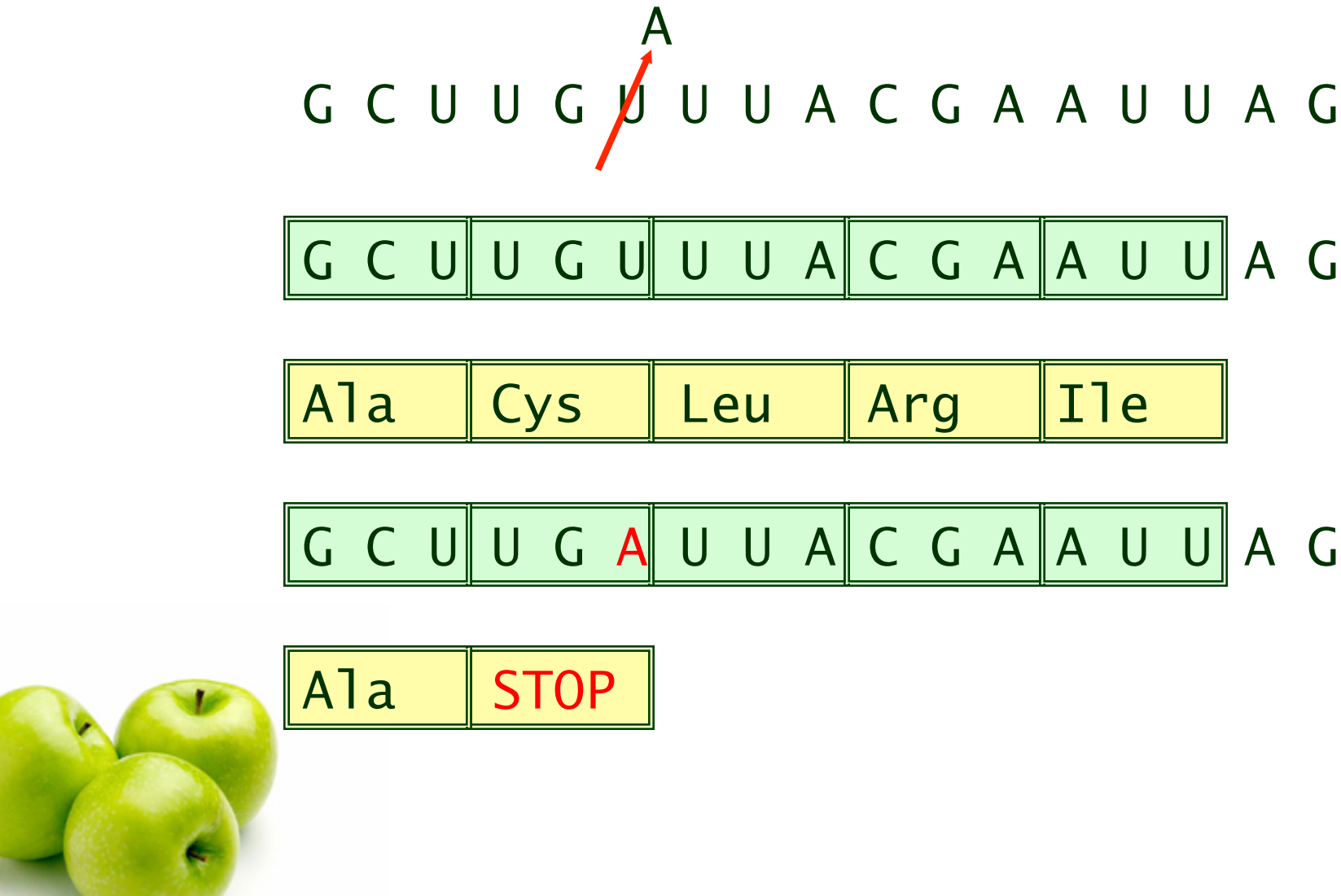
Missense Mutation

missense mutation



Nonsense Mutation

nonsense mutation



Frameshift

frameshift

G C U U G U U U A C G A A U U A G

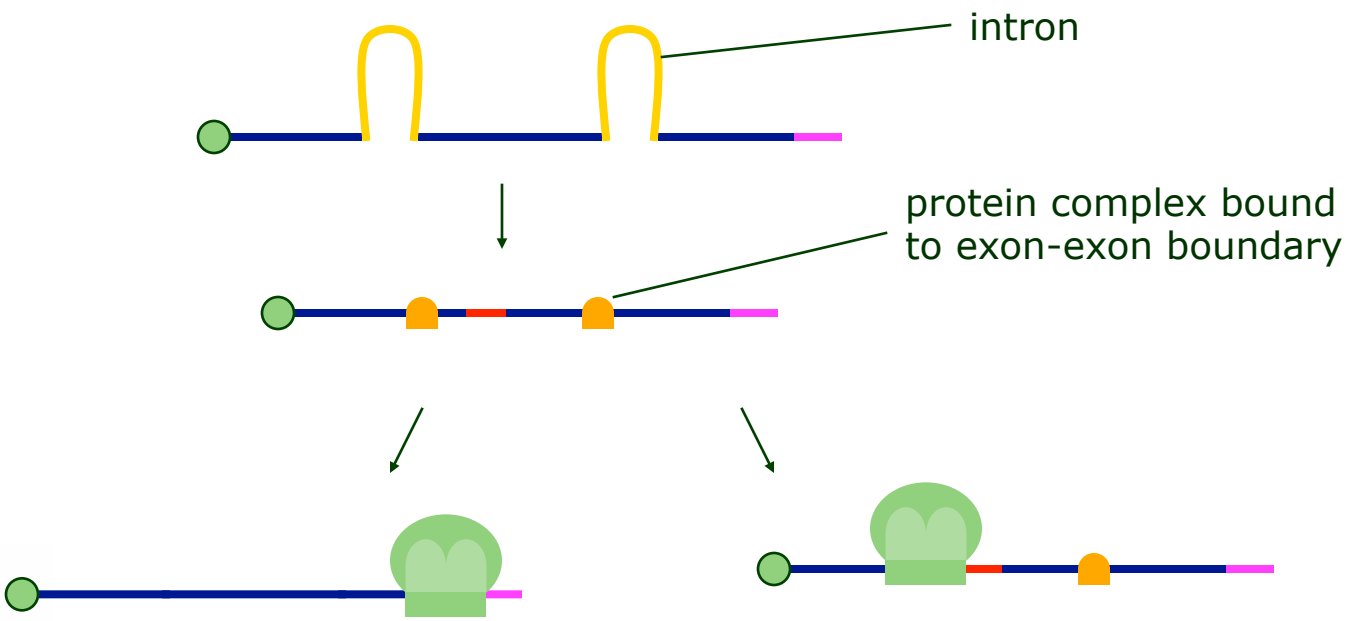
G C U	U G U	U U A	C G A	A U U	A G
Ala	Cys	Leu	Arg	Ile	

G C U	U G U	U A C	G A A	U U A	G
Ala	Cys	Tyr	Glu	Leu	

Quality Control

nonsense-mediated decay

- Nonsense-Mediated mRNA Decay (NMD)
  (Destroy mRNA with premature STOP codon)



Gene Expression Regulation

regulation

- When should each gene be expressed?
- **Regulate** gene expression

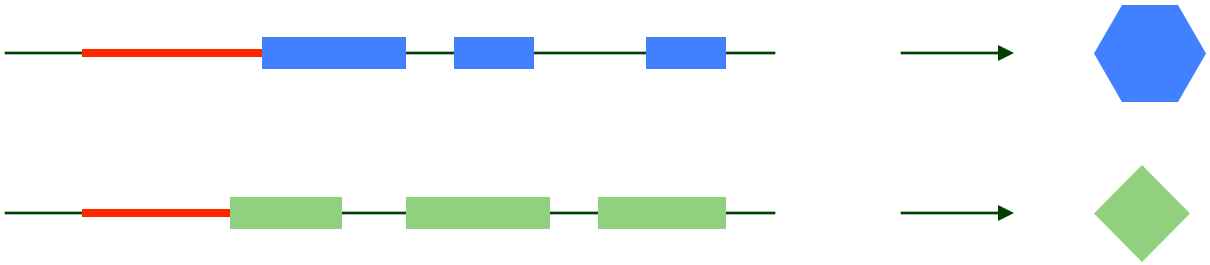
Examples:

- Make more of gene A when substance X is present
- Stop making gene B once you have enough
- Make genes C_1, C_2, C_3 simultaneously

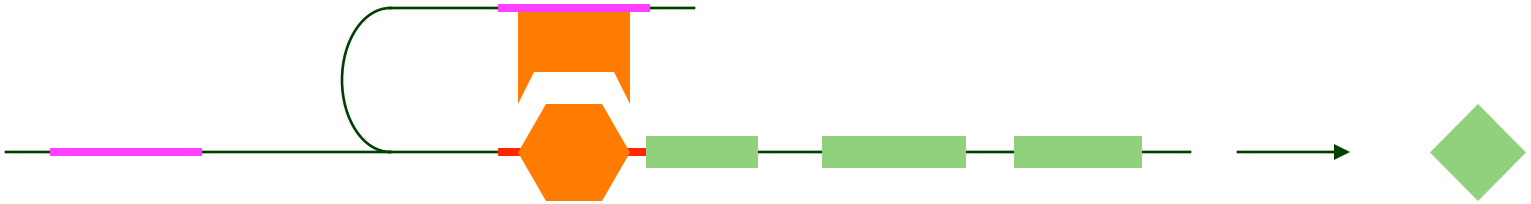
Regulatory Mechanisms

enhancer, silencer

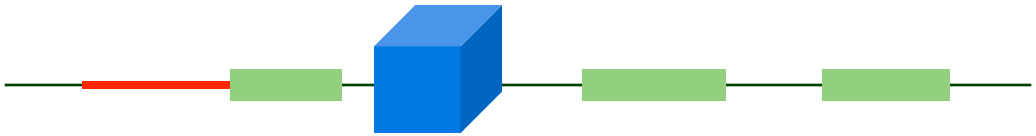
Transcription Factor Specificity:



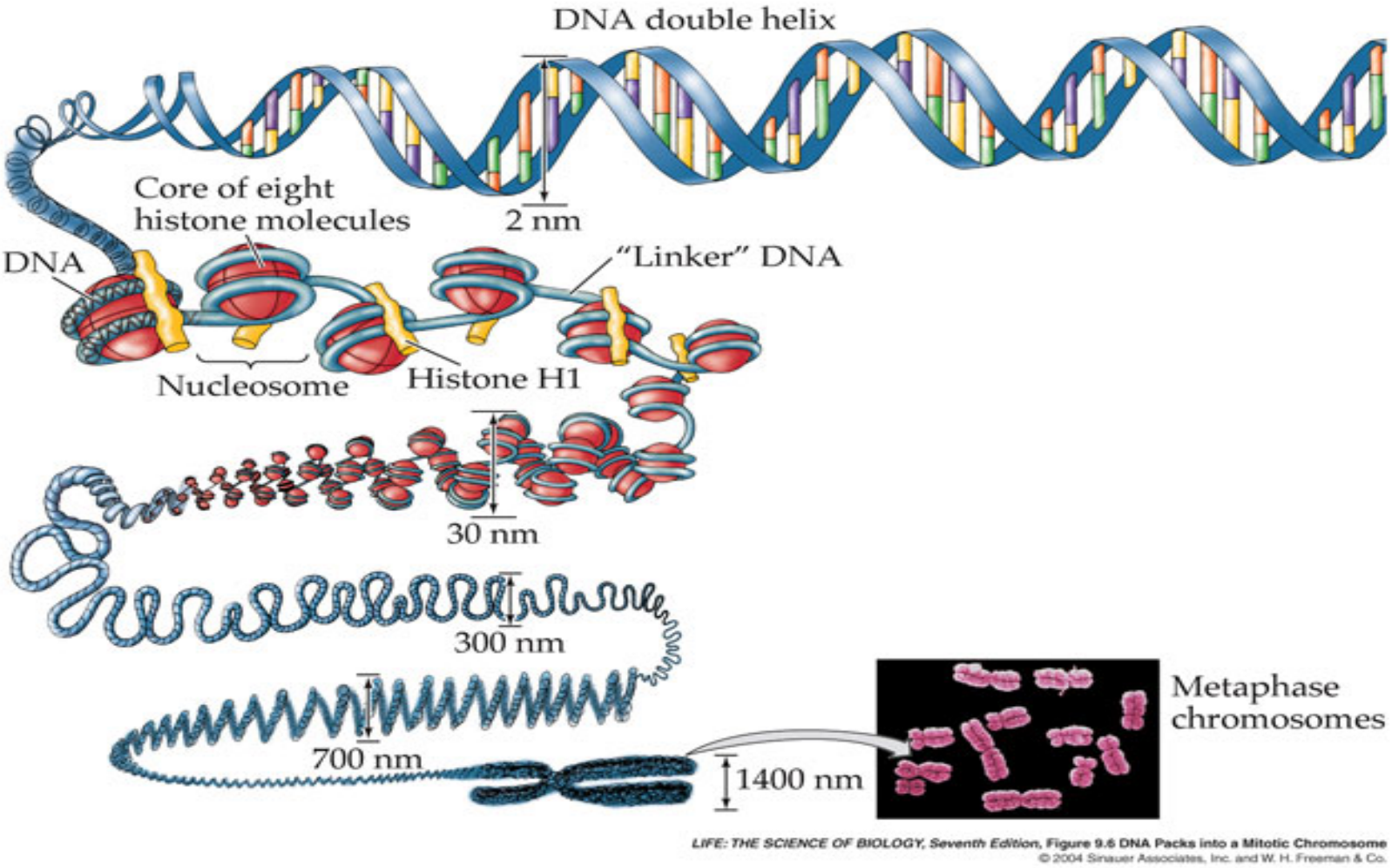
Enhancer:



Silencer:



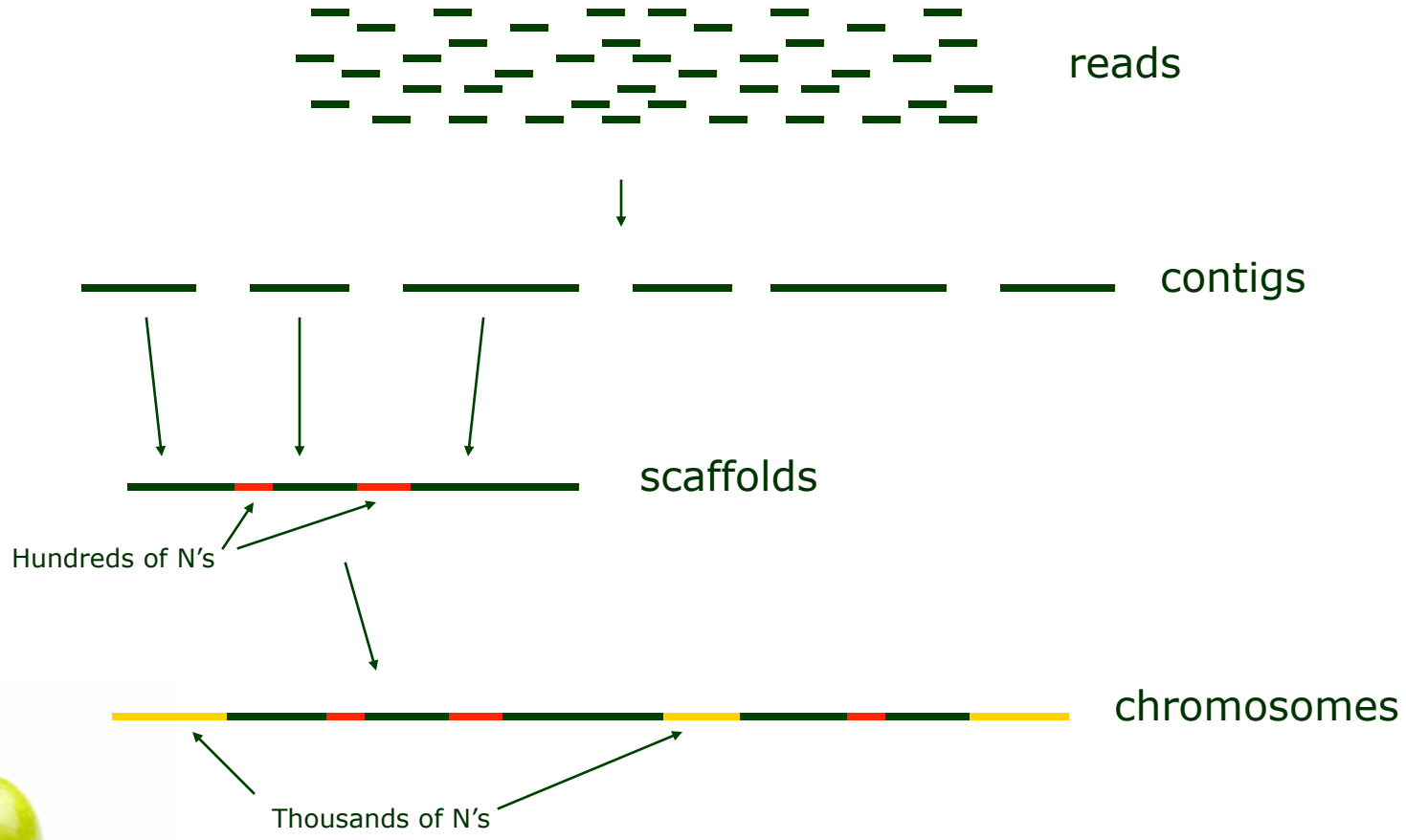
Chromatin



LIFE: THE SCIENCE OF BIOLOGY, Seventh Edition, Figure 9.6 DNA Packs into a Mitotic Chromosome

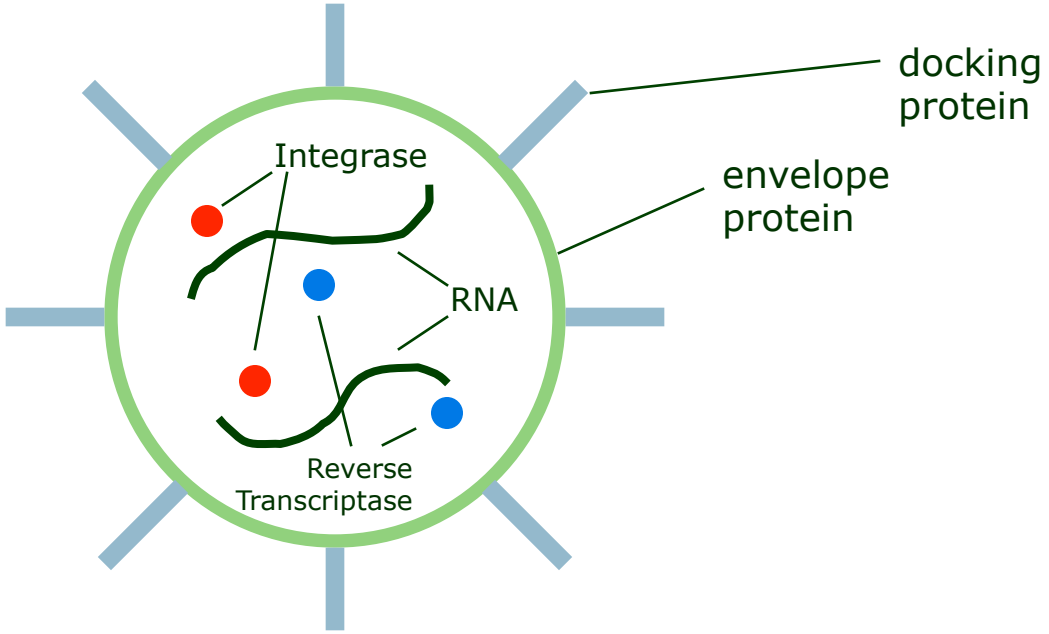
Assemblies

read, contig, scaffold, sequencing gaps, assembly

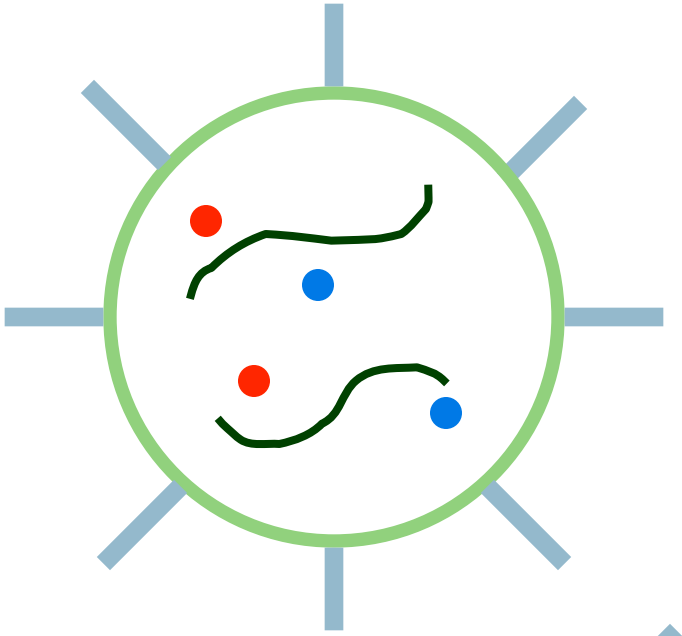


Retrovirus

virus, reverse transcriptase, integrase

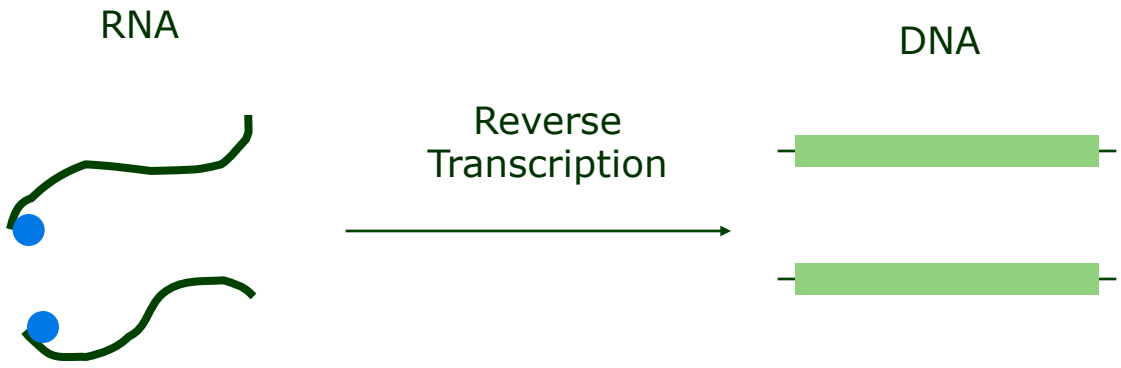


Infection

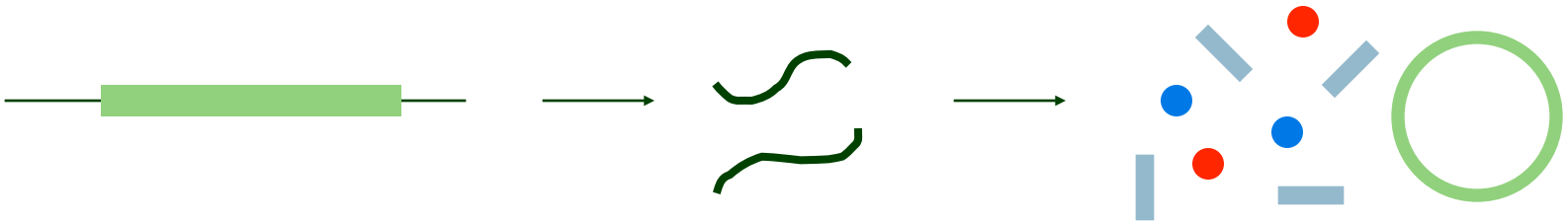


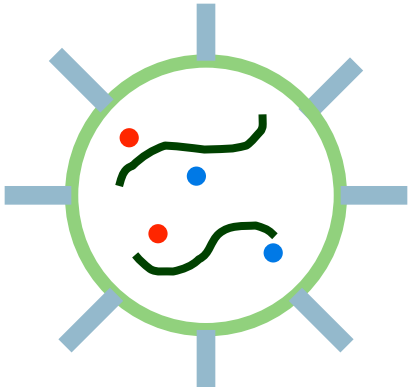
Infection

Replication cycle



Replication cycle





Are they alive?

- Polio virus made from scratch ($300,000 DARPA project – 2002)

> “ The first part of the sequence was painstakingly pieced together by hand and took over a year. The researchers then hired a commercial laboratory, Integrated DNA Technologies, to synthesise the remaining two thirds of the sequence mechanically. This took an additional two months.”

Are they alive?

- Polio virus made from scratch ($300,000 DARPA project – 2002)

> “ Once the entire sequence was replicated, it was reconverted into RNA by enzymatic means. Viral propagation and replication were accomplished by throwing the virus into a predesigned protein soup that contained all the polymerases and other enzymatic ingredients necessary for RNA transcription and translation. The synthetic virus was able to successfully replicate itself from this mixture.”

Are they alive?

- Polio virus made from scratch ($300,000 DARPA project – 2002)

> “ The viral copies were then injected into the brains of mice, which subsequently developed paralysis indistinguishable from polio. ”

The end?


READING THE ENTIRE GENOME CODE...
Listen to this part: 'AGGCTAATC CGCATAACTG'
Wow!

Keywords

cell, nucleus, cytoplasm, mitochondrion, histone, nucleosome, chromatin, chromosome, centromere, telomere, deoxyribose, nucleotide, base, A, C, G, T, purine, pyrimidine, 3′, 5′, deoxyribonucleic acid (DNA), strand, reverse complement, ribose, ribonucleotide, U, gene, transcription, translation, protein, promoter, transcription factor, binding site, RNA polymerase, 5′ cap, polyadenylation, exon, intron, splicing, UTR, mRNA, amino acid, N terminus, C terminus, ribosome, codon, tRNA, anticodon, aminoacylation, mutation, reading frame, synonymous (silent) mutation, fourfold site, missense mutation, nonsense mutation, frameshift, nonsense-mediated decay, regulation, enhancer, silencer, read, contig, scaffold, sequencing gaps, assembly, virus, reverse transcriptase, integrase