Bioinformatics

CS300

Chap 2

Computational Manipulation of DNA

Spring 2021

Oliver BONHAM-CARTER

Genes and Alleles

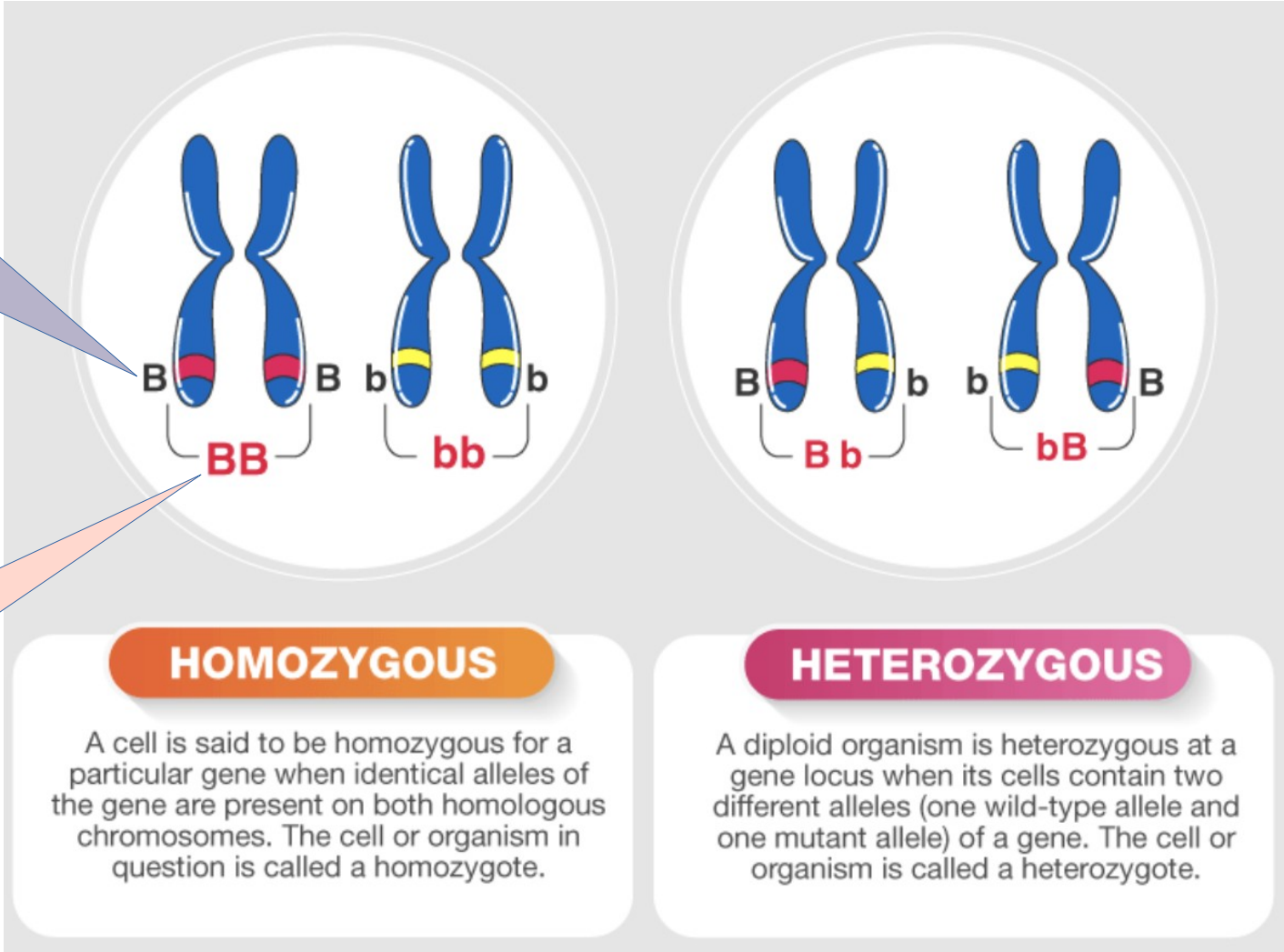
- **Gene**: A distinct sequence of nucleotides forming a piece of a chromosome. In biology, a gene is a sequence of nucleotides in DNA or RNA that codes for a molecule (a *protein*) that has a function. During gene expression, the DNA is first copied into RNA which is then transcribed into protein.

- **Allele**: One of two or more *alternative* forms of a gene that arise by mutation and are found at the same place on a chromosome.

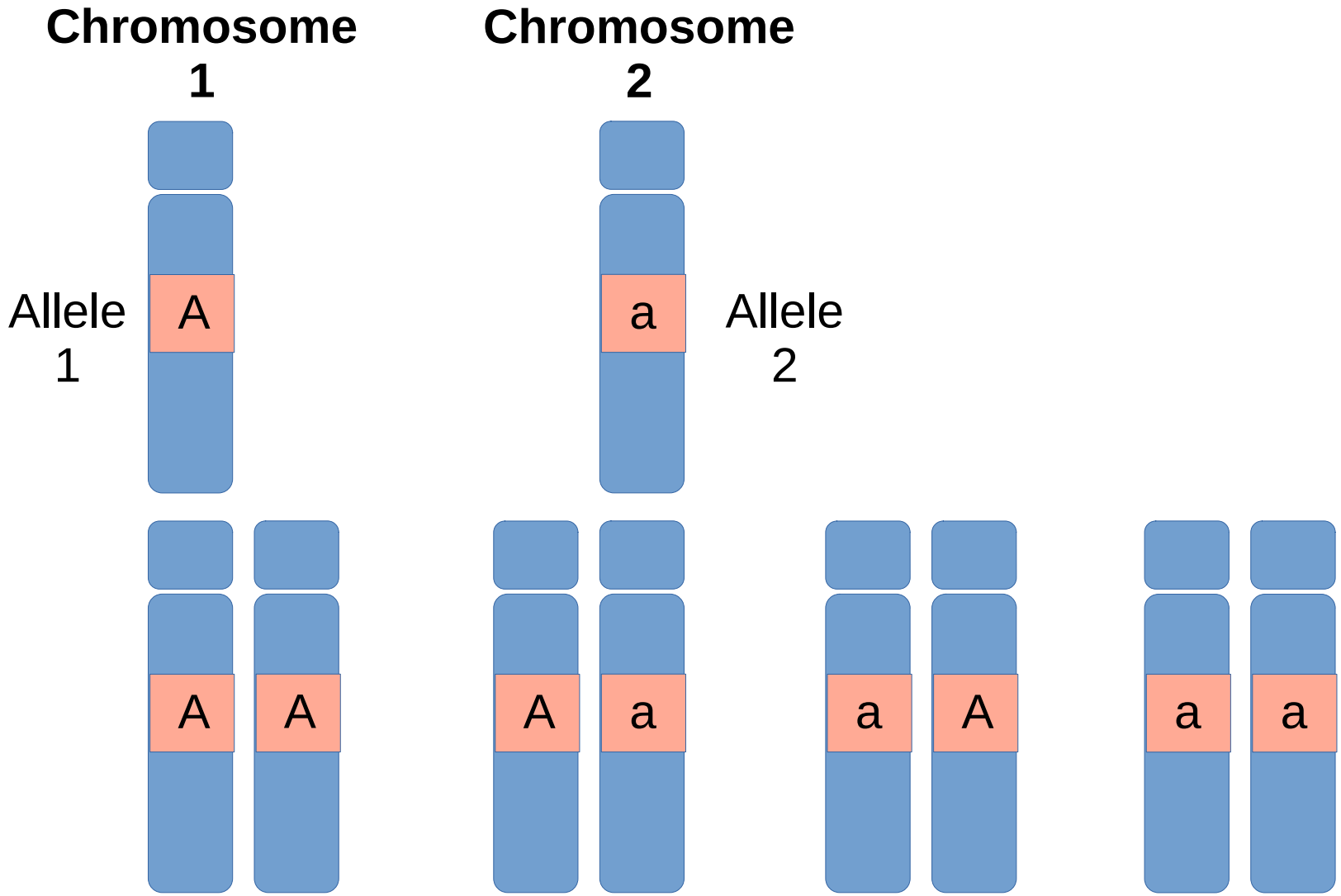
Genes Versus Alleles

Alleles: B and b (two alternative forms of gene)

Genes: BB, Bb, BB and bb

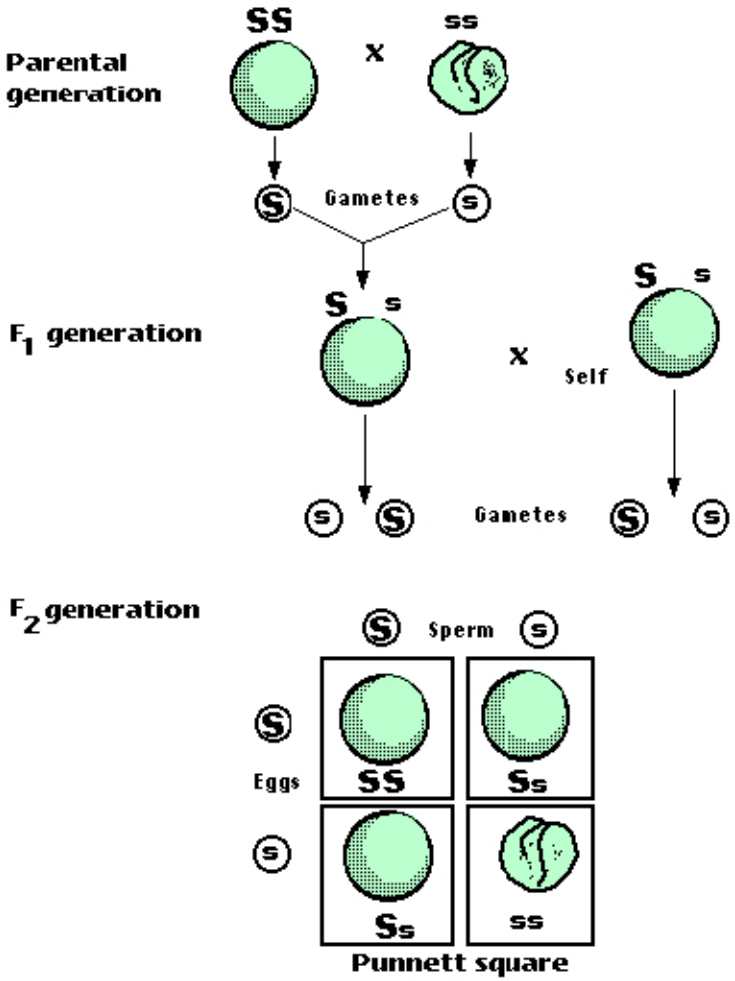


Genes And Alleles

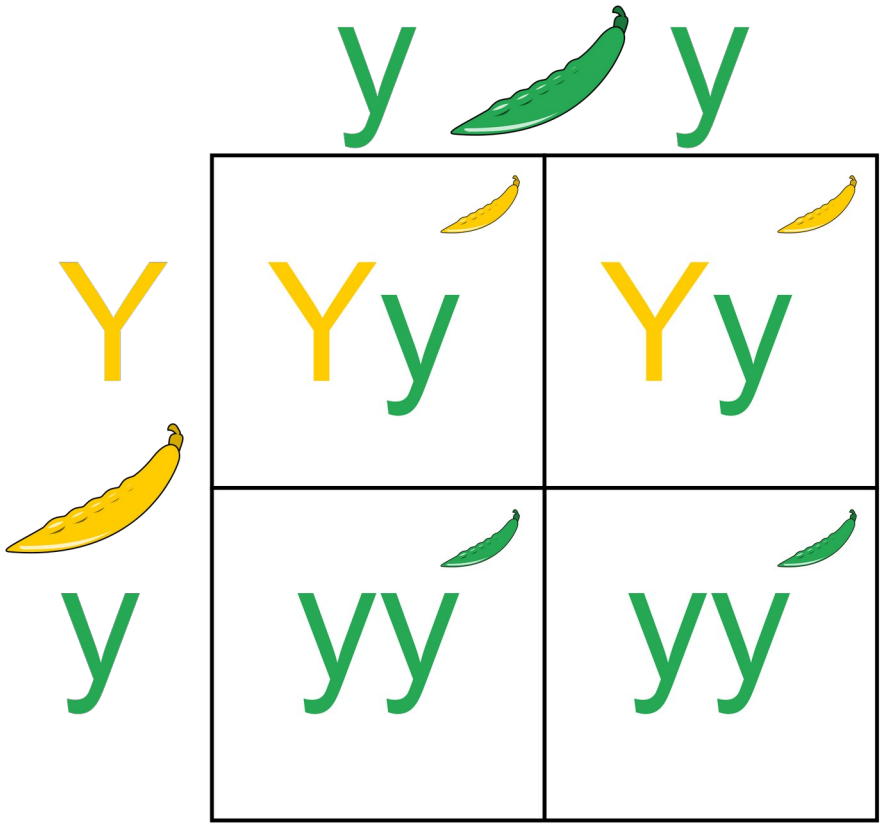


Mendelian Genetics

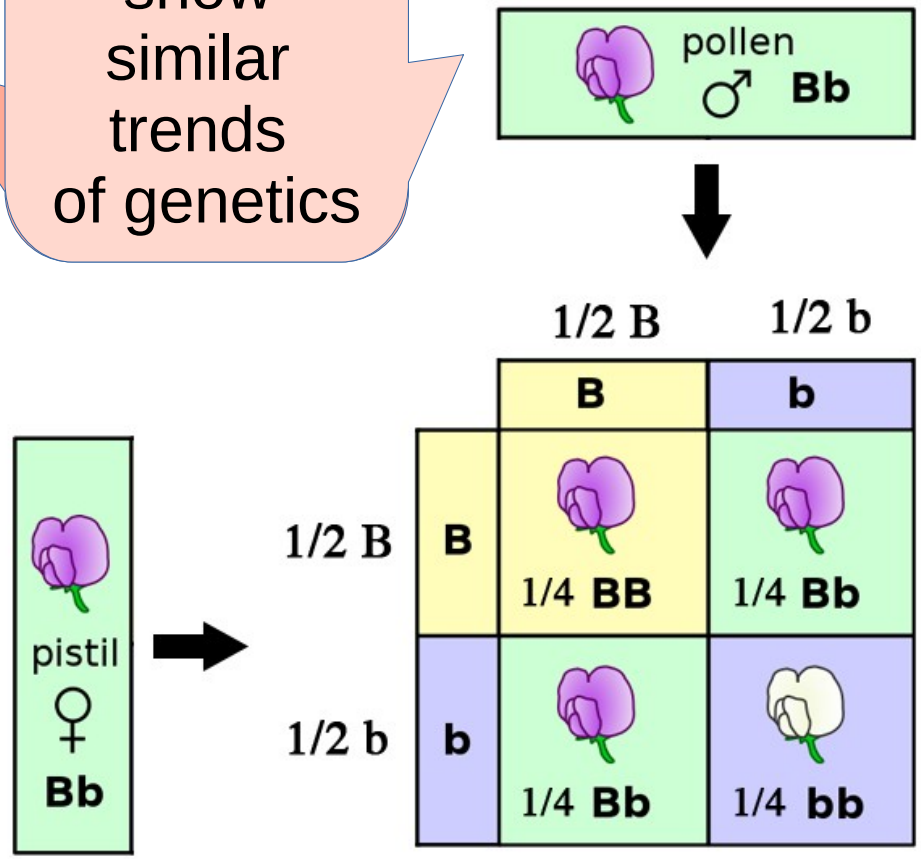
- Mendelian genetics
- Breeding experiments with *Pisum sativum* plants.
- Crossing large pees (SS genes) with small pees (ss genes), over several generations
- First generation: individuals exhibiting only one of the two traits (large and small)
- Inbred generations showed distributions of ¾ large pees (SS, Ss, sS), ¼ small pees (ss)
- S – *dominant* trait, s – *recessive* trait



Mendelian Genetics



Pees and flowers show similar trends of genetics

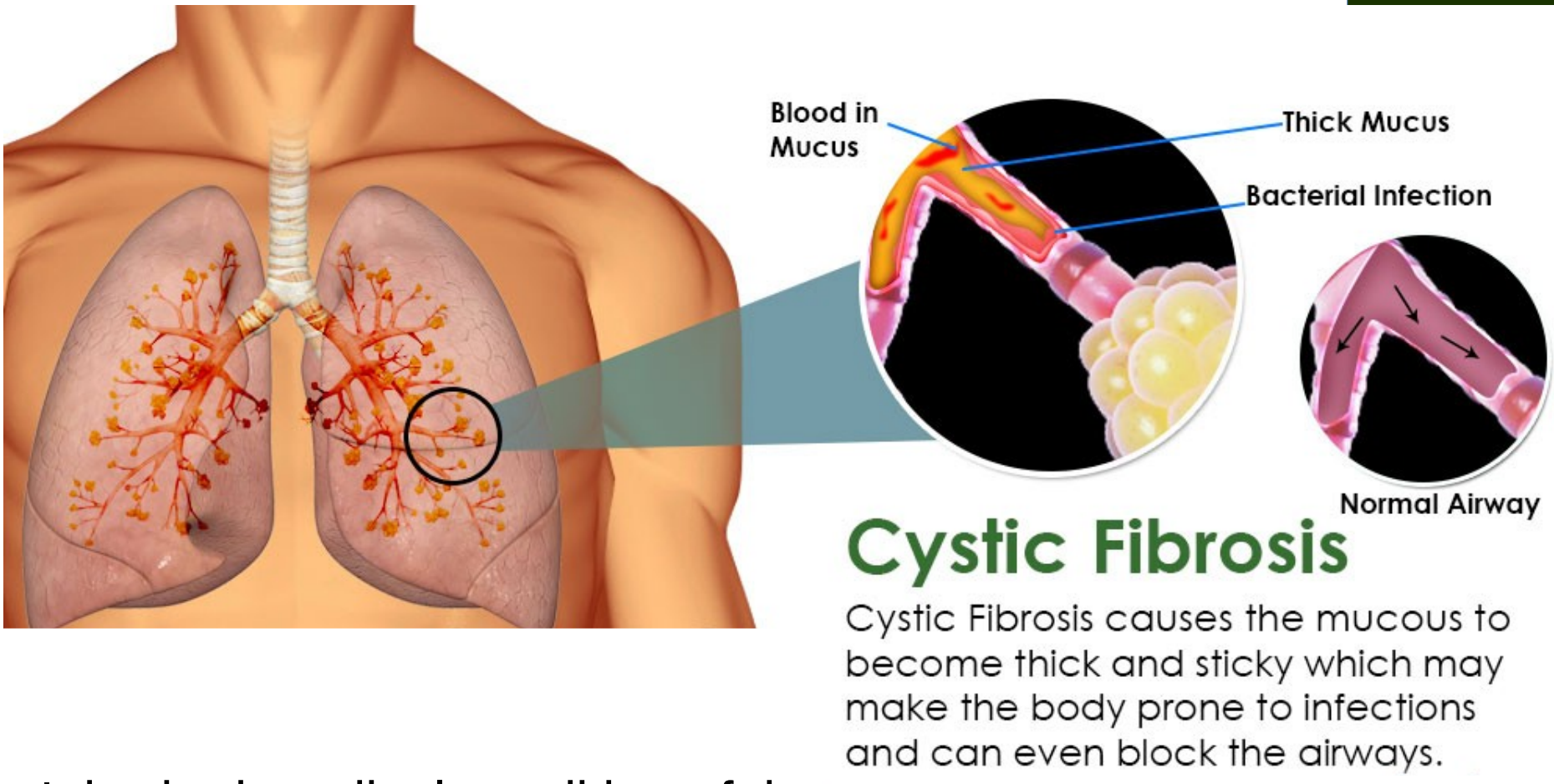


The Cystic Fibrosis Gene

- Cystic Fibrosis Transmembrane conductance: **CFTR**
- Gene product (protein) is bad regulator;
  - Fails to move water after displacing chloride ions in epithelial (thin tissue) cells
- Water follows chloride ions by osmosis.

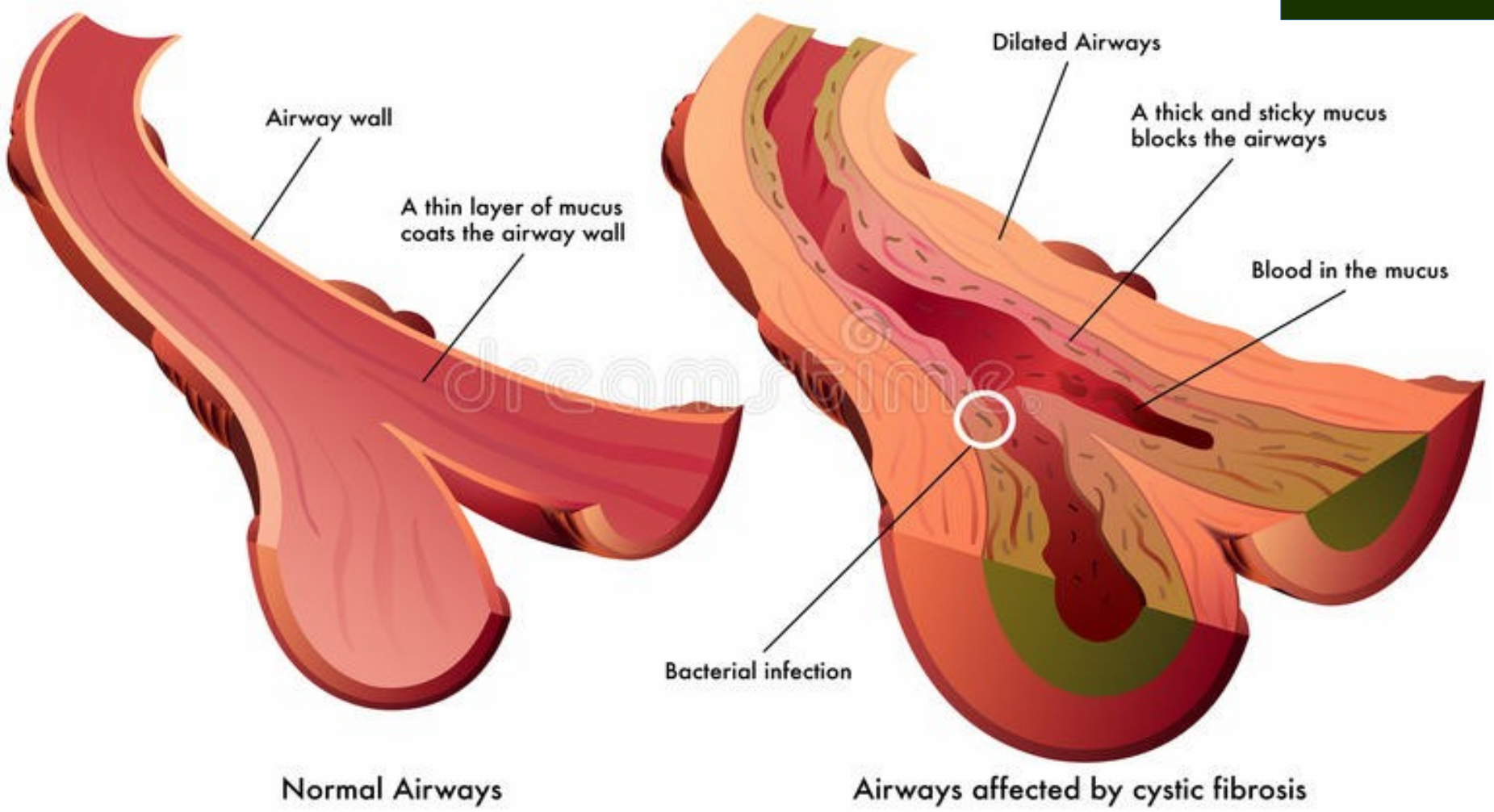
- **What if water regulation were not possible in the cells and organs? How much of your body is made up of water??**

Cystic Fibrosis



- Inherited medical condition of the secretory glands (producers of mucous and sweat)

Cystic Fibrosis: Symptoms



- Restricted flow in airways from mucous build-ups.
- Suffocation

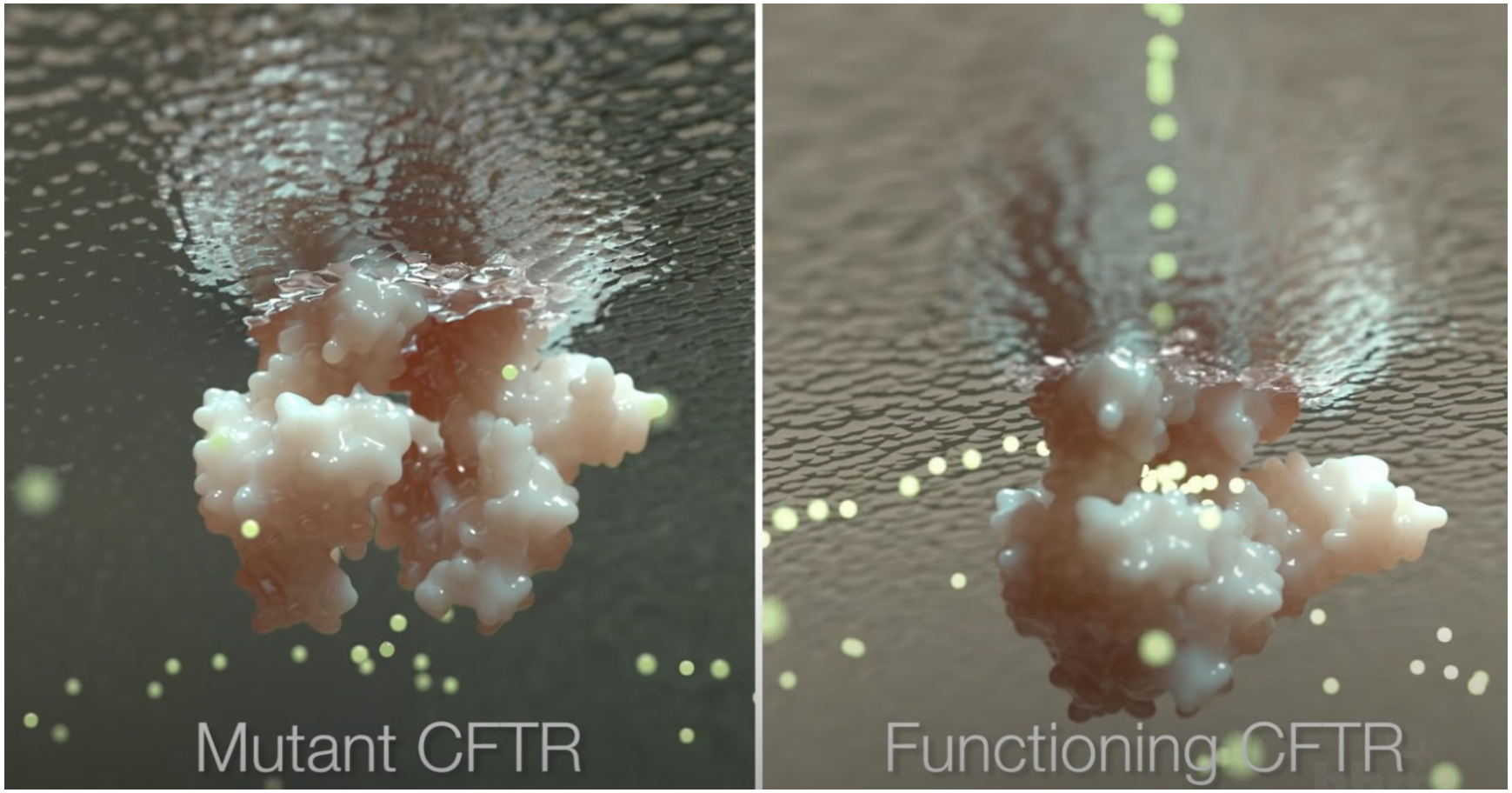
A Build-Up of Anything is Bad

- If garbage collection crews stop removing waste, then *things quickly get messy*.

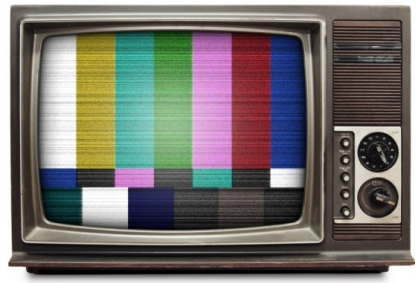
Cystic Fibrosis: Symptoms

- Clubbed fingers: occurs in heart and lung diseases that reduce the amount of oxygen in the blood

Blocked Ion Channels

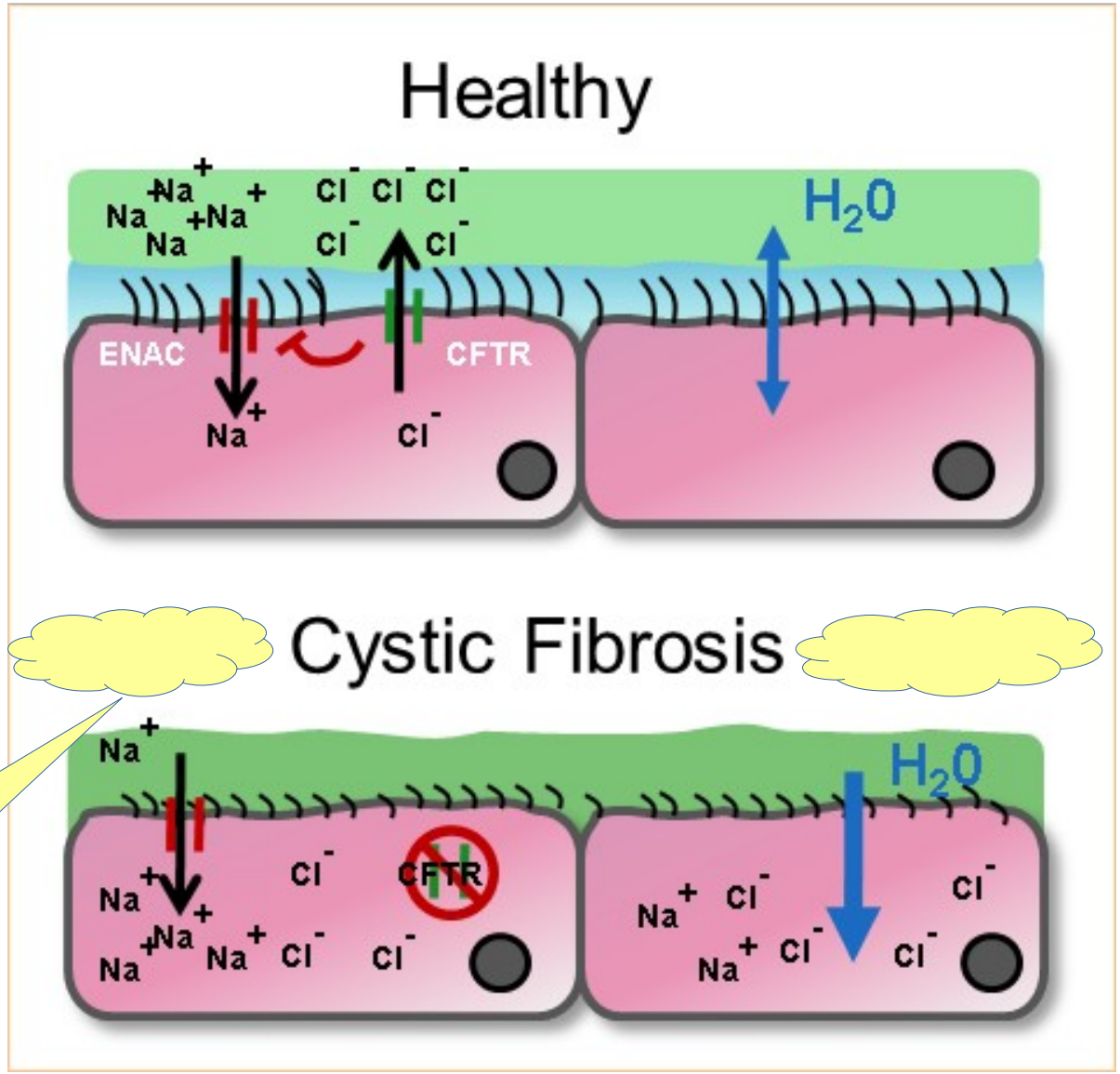


- Short videos of membrane transport proteins: https://www.youtube.com/watch?v=EuLVCYrurok
- Mechanism and Treatment https://www.youtube.com/watch?v=6IbP1ASGv9w



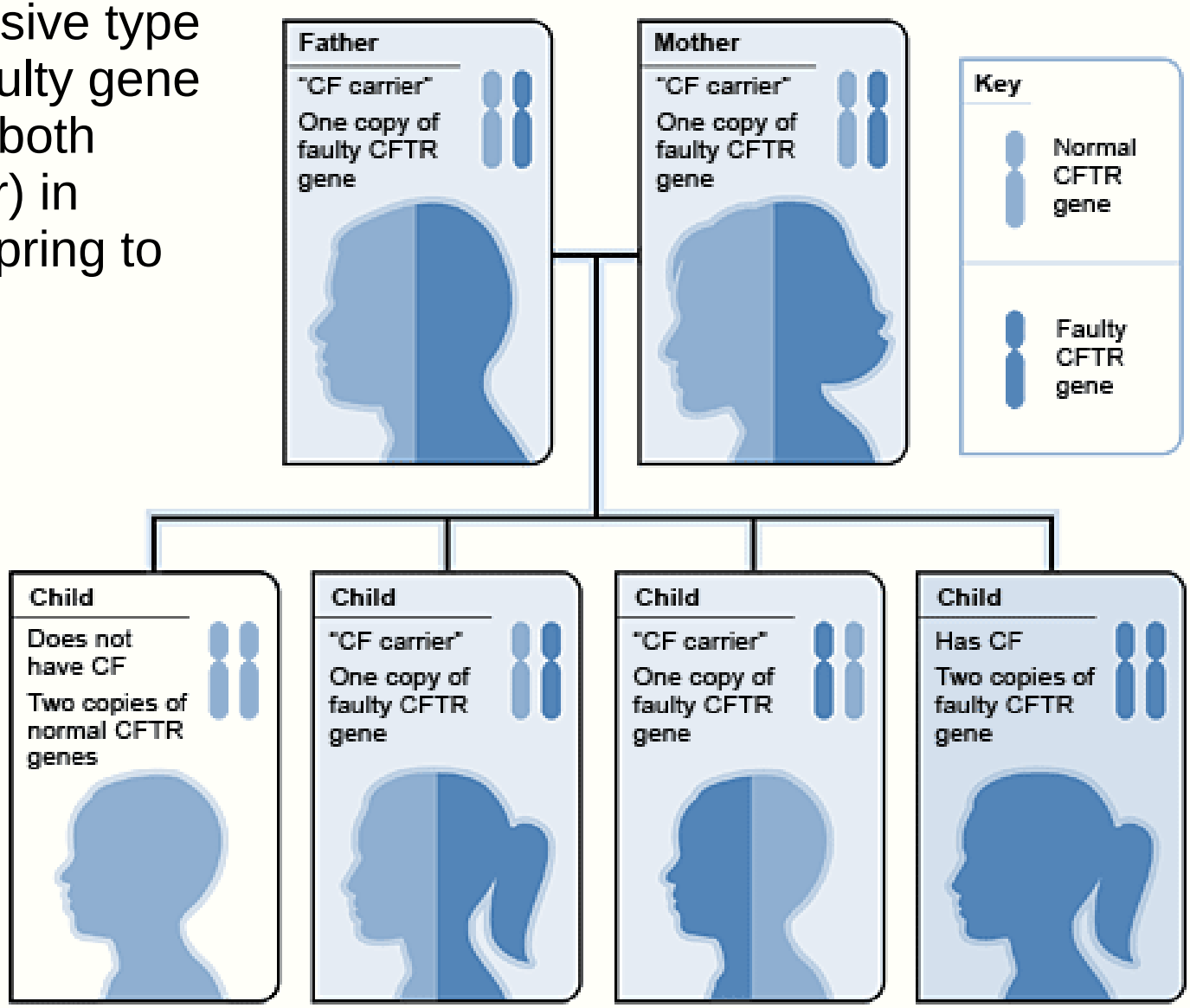
The Cystic Fibrosis Gene

- Gene codes for four different proteins: only one working type to move chloride ions and enable water displacement.



Cystic Fibrosis: Inheritance

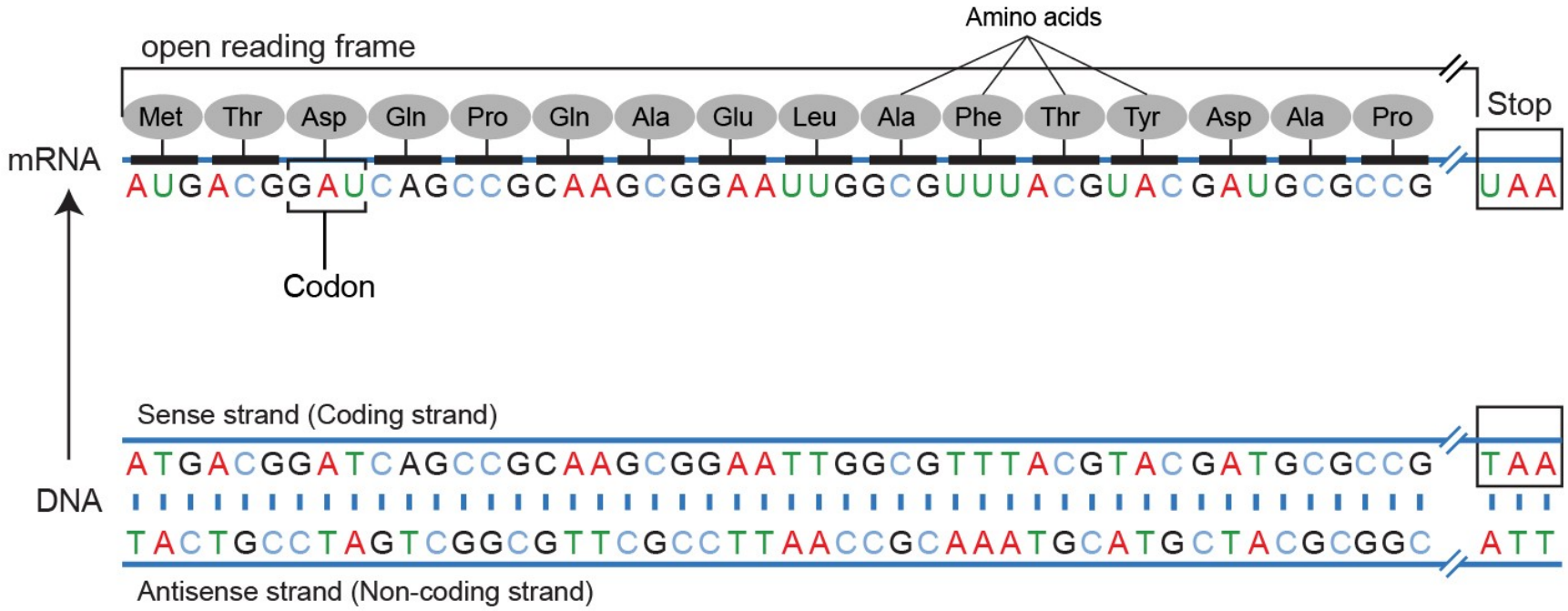
- Autosomal recessive type condition: one faulty gene is inherited from both parents (together) in order for the offspring to get this condition



- Modeled via Mendelian Genetics
- Impossible to know that someone is sure to get a condition.

Open Reading Frames

- An open reading frame (ORF) is the part of a reading frame that has the ability to be translated into protein.
- An ORF is a continuous stretch of codons that begins with a **start** codon (usually AUG) and ends at a **stop** codon (usually UAA, UAG or UGA).



Cite:
https://www.genome.gov/genetics-glossary/Open-Reading-Frame

Open Reading Frames: Simple Example

- Pam Can See The Man and Dog
- Frame shift by one letter!
- ~~P~~ amC anS eeT heM ana ndD ~~og~~
- Frame shift by two letters!
- ~~Pa~~ mCa nSe eTh eMa nan dDo ~~g~~
- Frame shift by three letters
- ~~Pam~~ Can See The Man and Dog

Reading by triplets

Notice how the code changes depending on where you start reading? (That is a *frameshift*.)

Open Reading Frames: DNA Example

Note: RF means *reading frame, where you start reading the words.*

Original: **CAATGGCGAATCGACGTGTATAAA**

RF1 - 5' - CAA TGG CGA ATC GAC GTG TAT AAA – 3'

RF2 - 5' - C AAT GGC GAA TCG ACG TGT ATA AA – 3'

RF 3 - 5' - CA **ATG** GCG AAT CGA CGT GTA **TAA** A – 3'

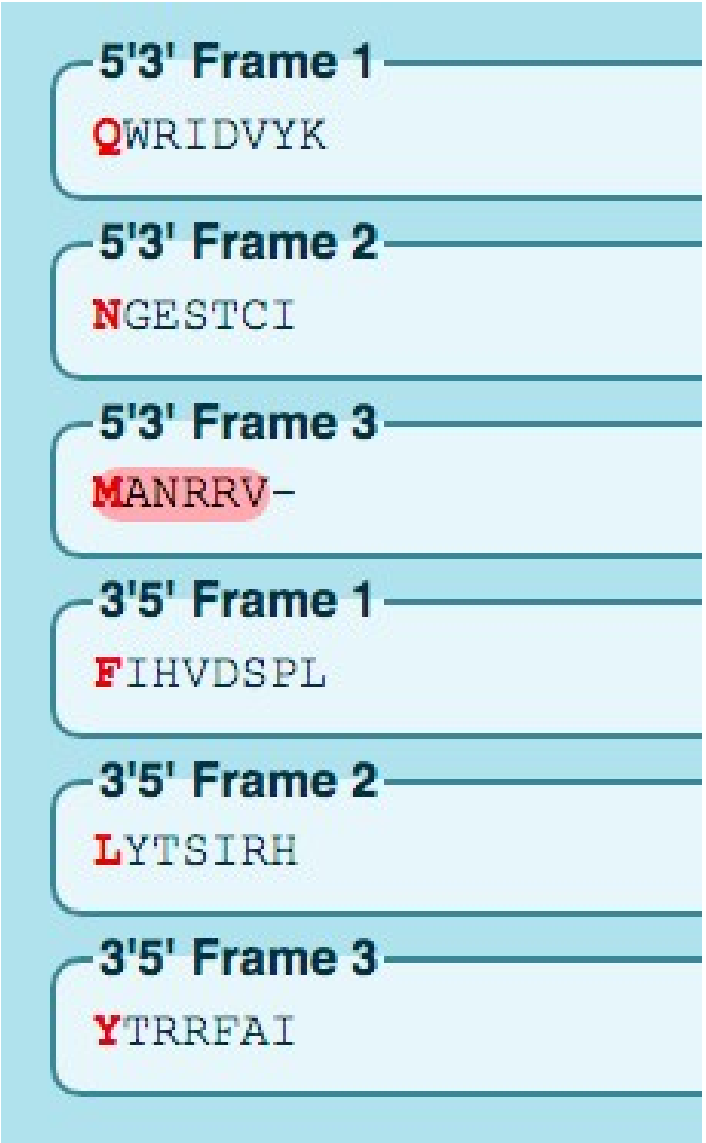
3' - CAA TGG CGA ATC GAC GTG TAT AAA - 5' – RF 4

3' - C AAT GGC GAA TCG ACG TGT ATA AA - 5' – RF 5

3' - CA ATG GCG **AAT** CGA CGT **GTA** TAA A - 5' – RF 6

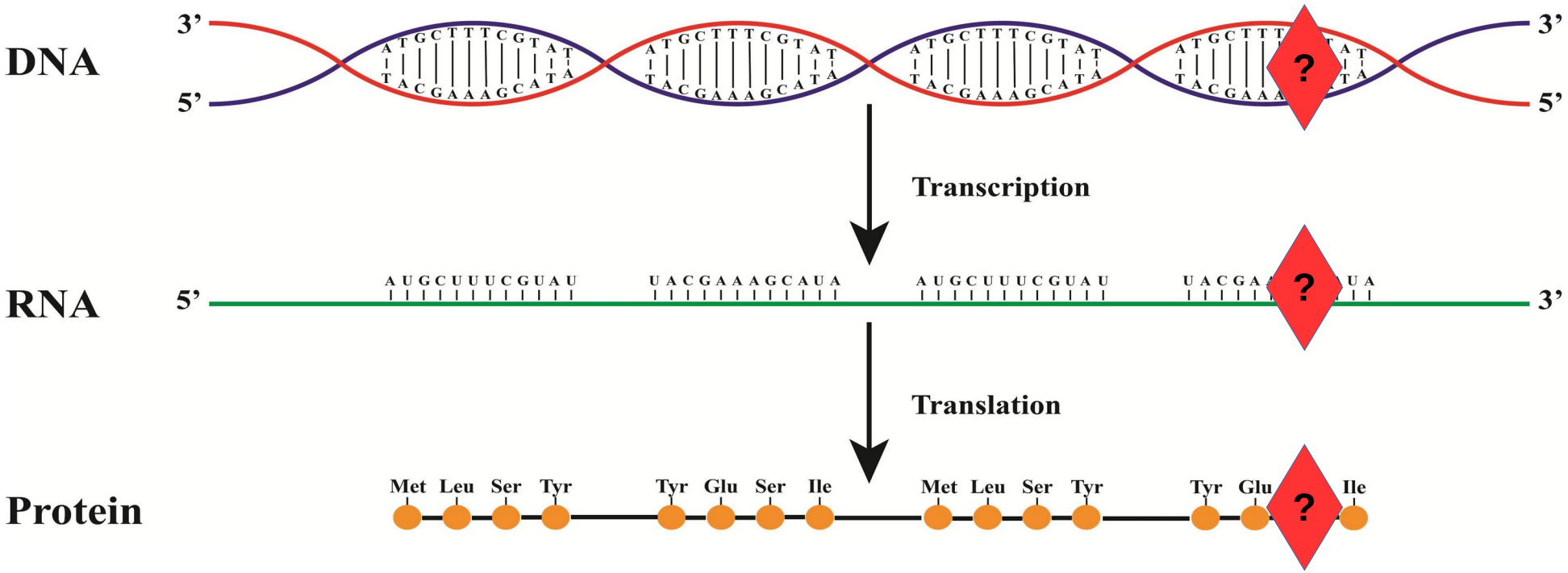
Open Reading Frames: Online

- Original: **CAATGGCGAATCGACGTGTATAAA**
- Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.
  - https://web.expasy.org/translate/



Sequence is Carrier?

- How do we determine if a sequence carries the Cystic Fibrosis allele?
- Get DNA sample and translate into protein. Then compare product protein sequence to that of a “working protein”
- Is there a difference (structure or function) between of the protein sequences?



Remember the Codon Table?

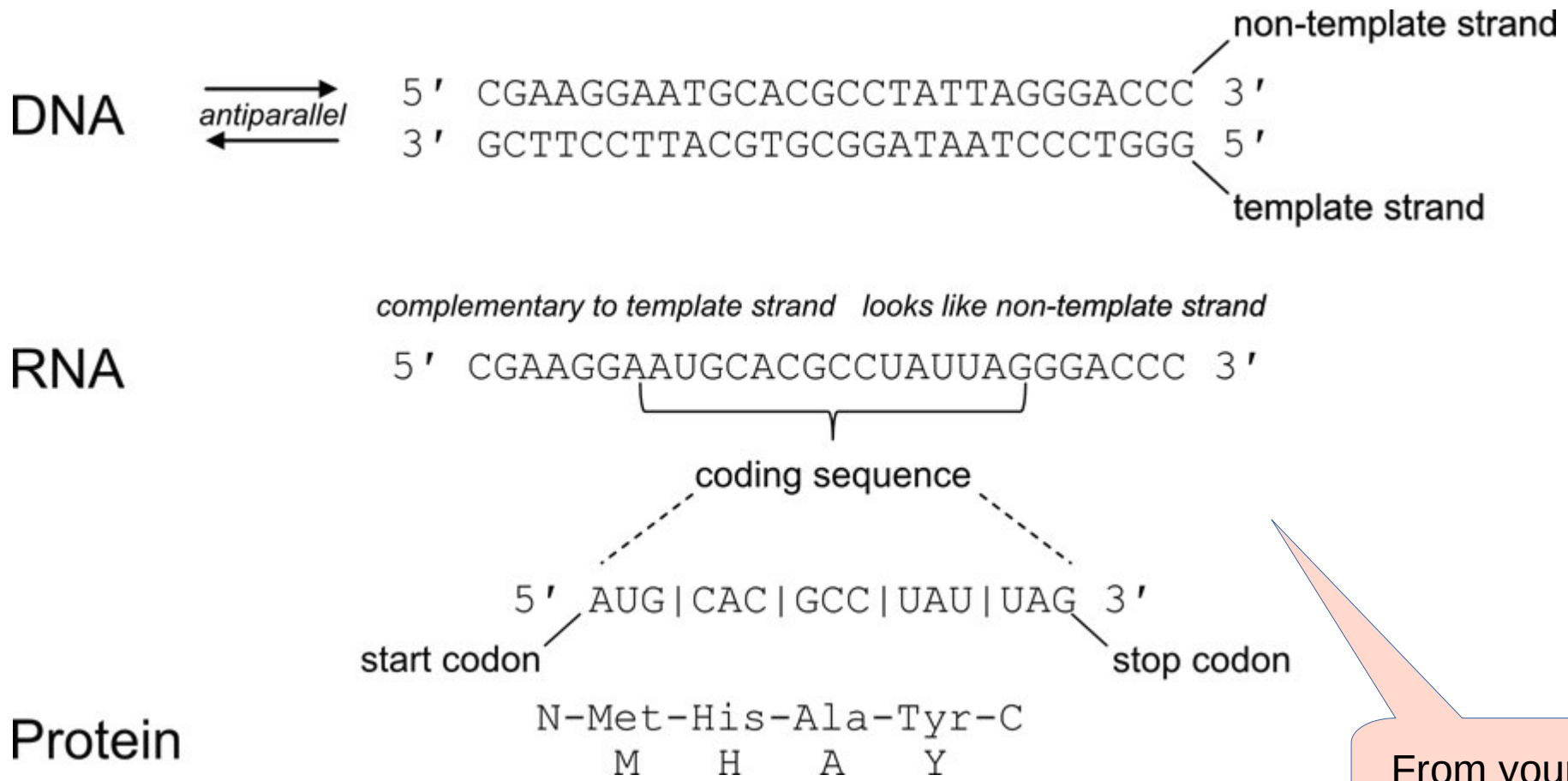
- DNA triplets called *codons*, translate into amino acids
- T's from DNA are read as U's as RNA after transcription

Standard genetic code

1st base	2nd base								3rd base
	T		C		A		G		
T	TTT	(Phe/F) Phenylalanine	TCT	(Ser/S) Serine	TAT	(Tyr/Y) Tyrosine	TGT	(Cys/C) Cysteine	T
	TTC		TCC		TAC		TGC		C
	TTA	(Leu/L) Leucine	TCA		TAA[B]	Stop (*Ochre*)	TGA[B]	Stop (*Opal*)	A
	TTG		TCG		TAG[B]	Stop (*Amber*)	TGG	(Trp/W) Tryptophan	G
C	CTT		CCT	(Pro/P) Proline	CAT	(His/H) Histidine	CGT	(Arg/R) Arginine	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	(Ile/I) Isoleucine	ACT	(Thr/T) Threonine	AAT	(Asn/N) Asparagine	AGT	(Ser/S) Serine	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	ATG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GTT	(Val/V) Valine	GCT	(Ala/A) Alanine	GAT	(Asp/D) Aspartic acid	GGT	(Gly/G) Glycine	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GTG		GCG		GAG		GGG		G

Summary:
The Steps to Study Protein

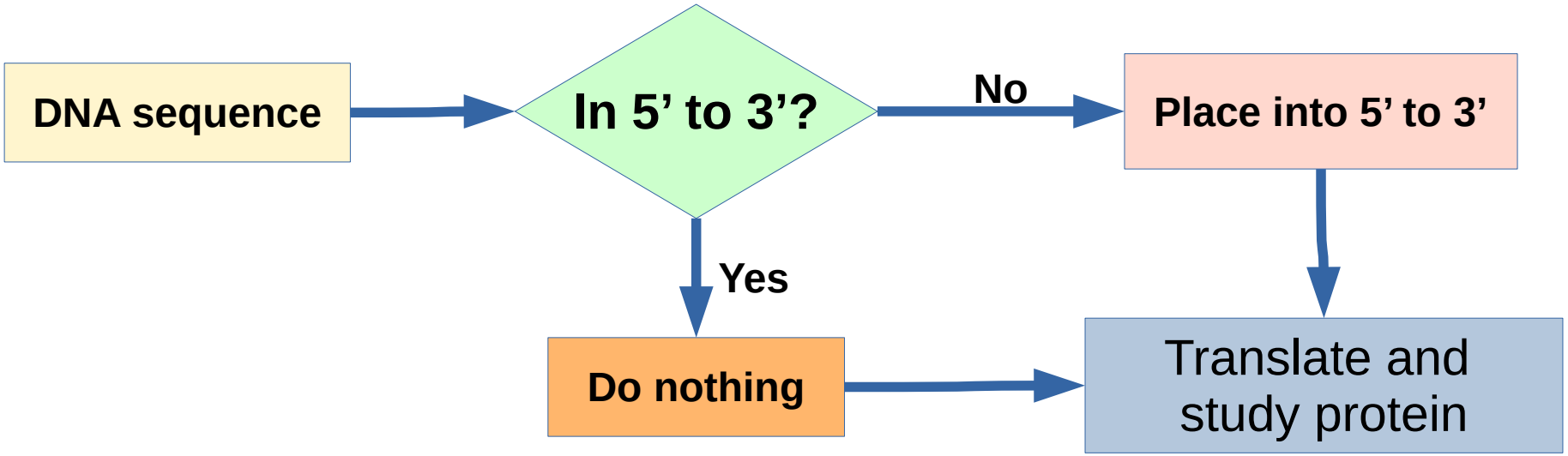
- Translating DNA to find defects in the protein



From your textbook!

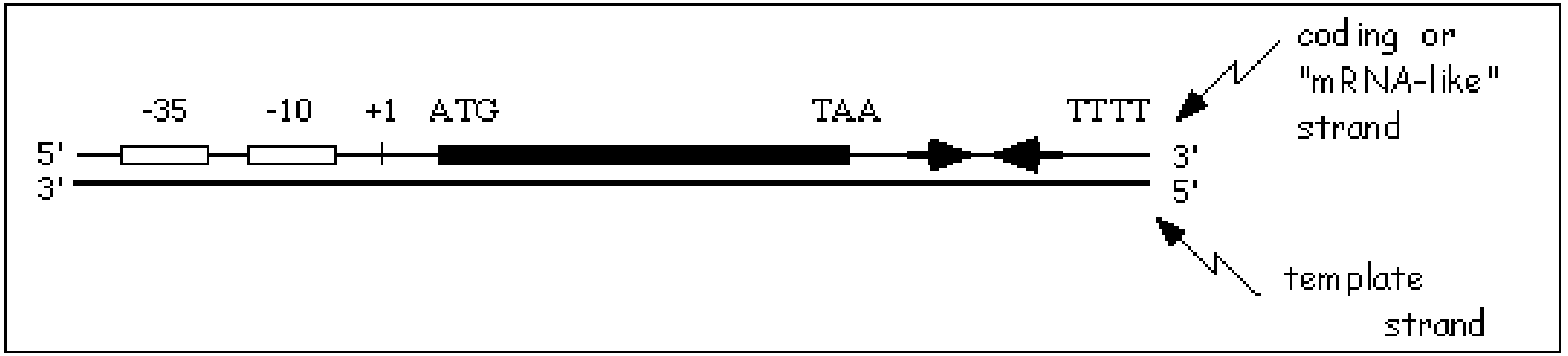
Remember: DNA Must Be In 5’ to 3’ Direction for Protein Translation

- Unlabeled strands of DNA are assumed to be in the 5’ to 3’, (left to right) direction.
- What are the steps to place a sequence into a format for translation simulation with bioinformatics tools?



Template vs nonTemplate

- Input:
  - DNA sequence: AGCAT
  - Strands:
  - **template** (used to make mRNA), 3’ -> 5’
  - **non-template** (the compliment of this strand that looks like mRNA), 5’ -> 3’
- Output:
  - Template strand in 3’ -> 5’ orientation ready for transcription



Transcription Algorithm

- Input: **template** strand in the **3’ → 5’** orientation
- Output: mRNA strand in the **5’ → 3’** orientation
  - Traverse the string from left to right
    - add complementary base to the output string
    - (note T is now U)

DNA Seq = 3’ - ATGCGT - 5’

mRNA Seq = 5’ - AUGCGC - 3’

Alternative Transcription Algorithm

- Input: **non-template** strand in the **5’ → 3’** orientation
- Output: mRNA strand in the **5’ → 3’** orientation
  - Traverse the string from left to right
    - Replace all the T’s with U’s

DNA Seq = 5’ - ATGCGT - 3’

mRNA Seq = 5’ - AUGCGC - 3’

Translation Algorithm

- Input: mRNA strand in the **5’ → 3’** orientation
- Output: amino acid sequence
  - Traverse the string codon by codon
  - Add one amino acid to protein, corresponding to the protein sequence.

WAIT! Why is the 5’ to 3’ direction so important?! Remember the carbon atoms on DNA?