General Chemistry (CH101): Chemistry around Us

Department of Chemistry

KAIST

Genes and Life

Chapter 13

Diabetes

= have difficulty regulating blood sugar

→ The patients have an imbalance in a hormone, **insulin**

<u>Type 1 diabetes</u>: born with the disease
<u>Type 2 diabetes</u>: develop it at an older age

<Treatment>

Until 1990s a starvation diet = didn't help to extend the life of the children

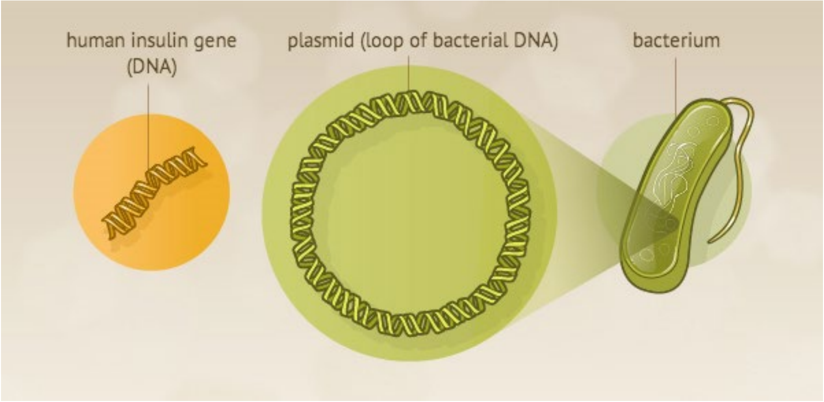
In 1920s Canadian researchers confirmed the role of the insulin
→ insulin was extracted from animal (calf, bovine)
→ 8000 pounds of animal pancreas glands = 1 pound of insulin
→ animal insulin ≠ human insulin (risk of inflammations)

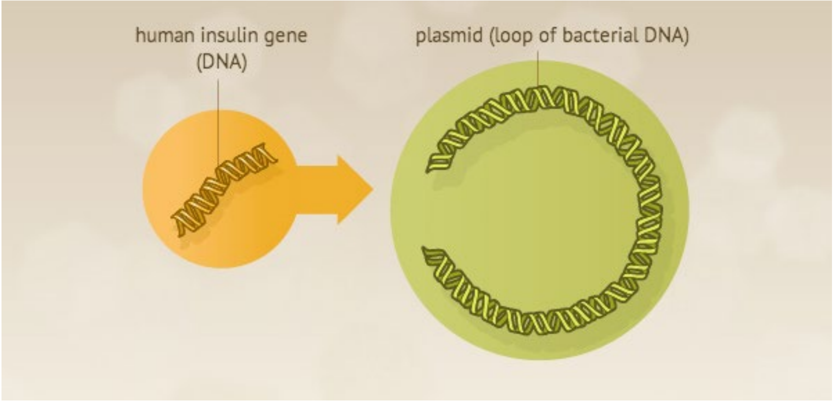
In 1980s Genentech → synthesized mass quantities of human insulin by using bacteria.

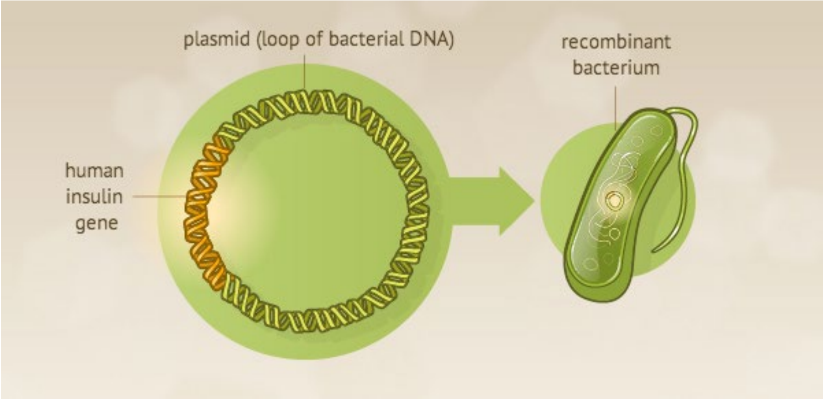
Bacteria produce human insulin

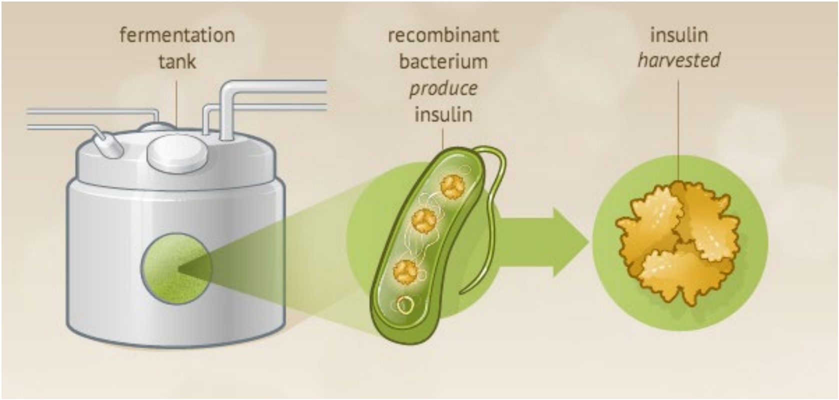
"Teaching" bacteria using a guidebook (genome)

Genome = the whole guide book
Gene = a section of the genome for the production of proteins







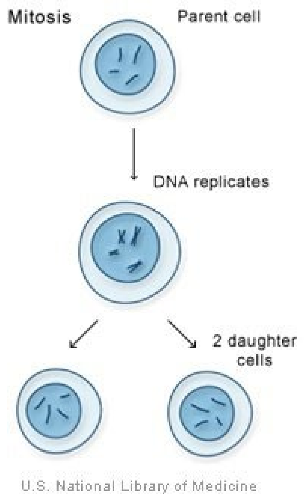


The most common strain of bacteria
= *Escherichia coli (E. coli)*

Genetically engineered E. coli
gives utmost purity

DNA

As cells grow and multiply, “the information” must replicate itself without error

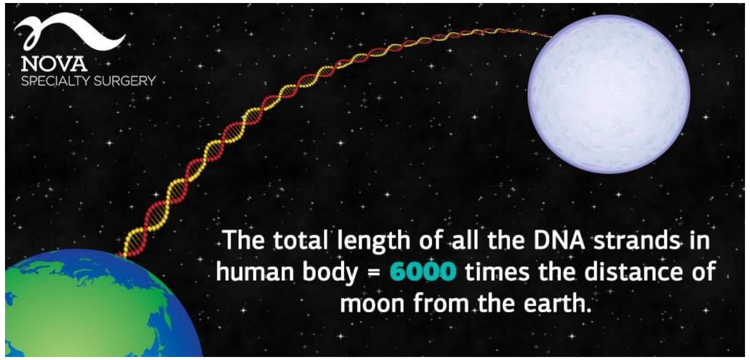


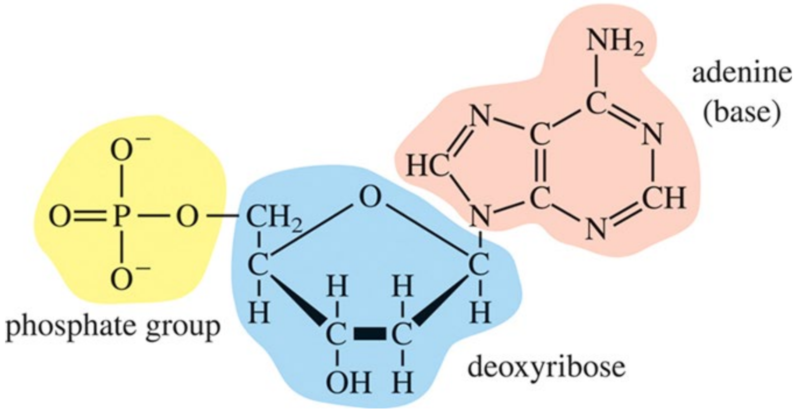
“The information”
= context-sensitive / delicately controllable
= highly advanced database in chemical form

Deoxyribonucleic acid

DNA
= replicate, transfer information, and respond to feedback
= biological polymer that **carries genetic information**
= The monomer of the polymer is **‘nucleotide’**

DNA in each cell (=2 meter)

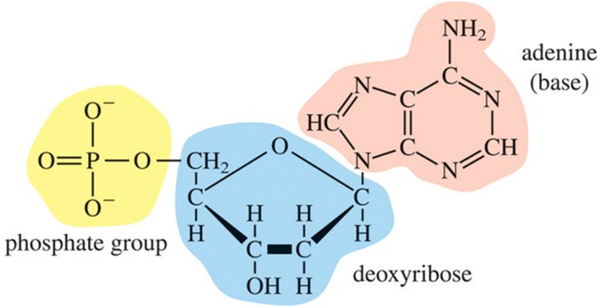




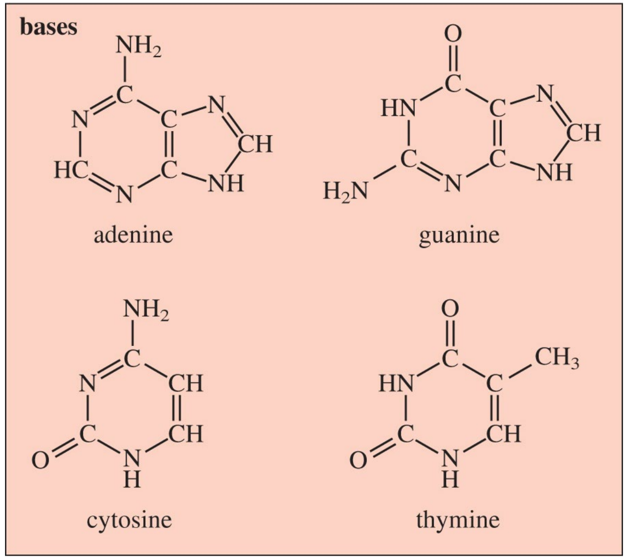
Nucleotide

https://youtu.be/utEQA1T1_LI

Nucleotide



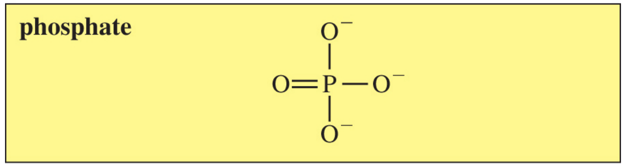
Nitrogen-containing bases

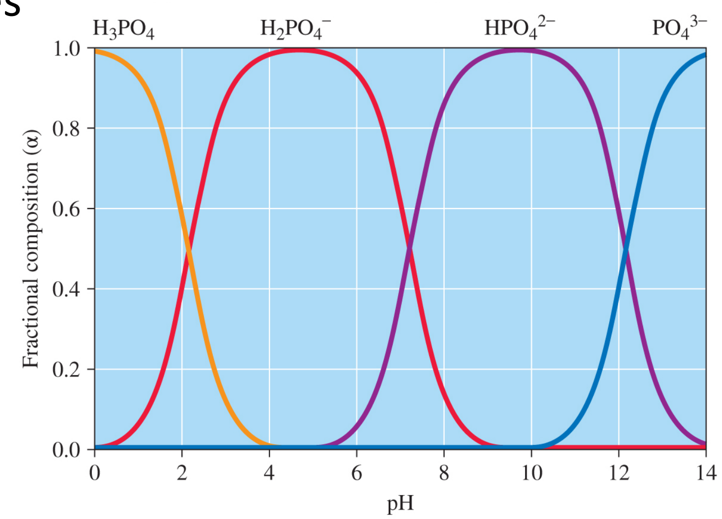


The larger bases

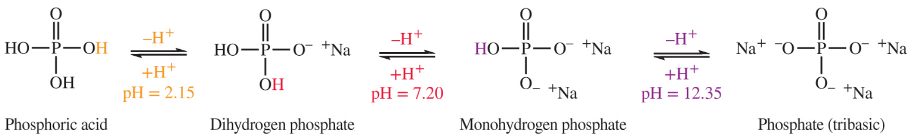
The smaller bases

Phosphate



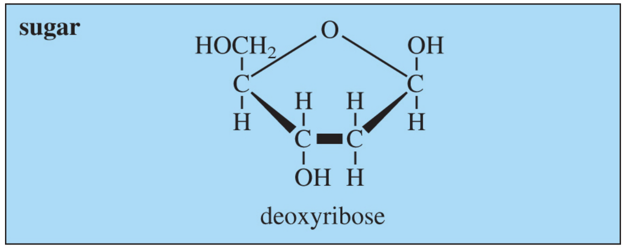


The data presented in the graph are related to the following acid–base equilibria:

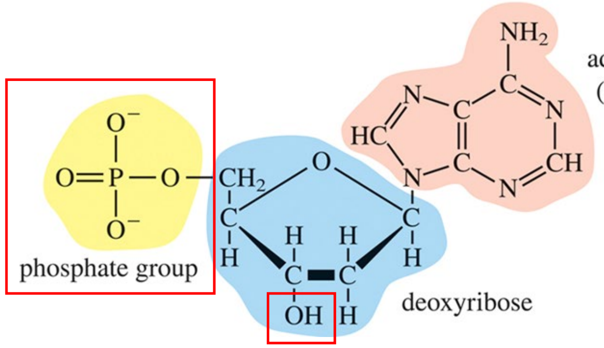


DNA is also pH dependent!

Deoxyribose sugars =monosaccharide

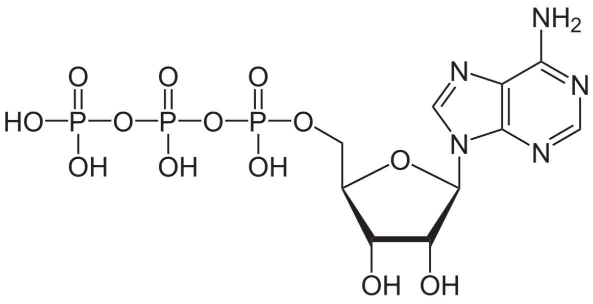


DNA

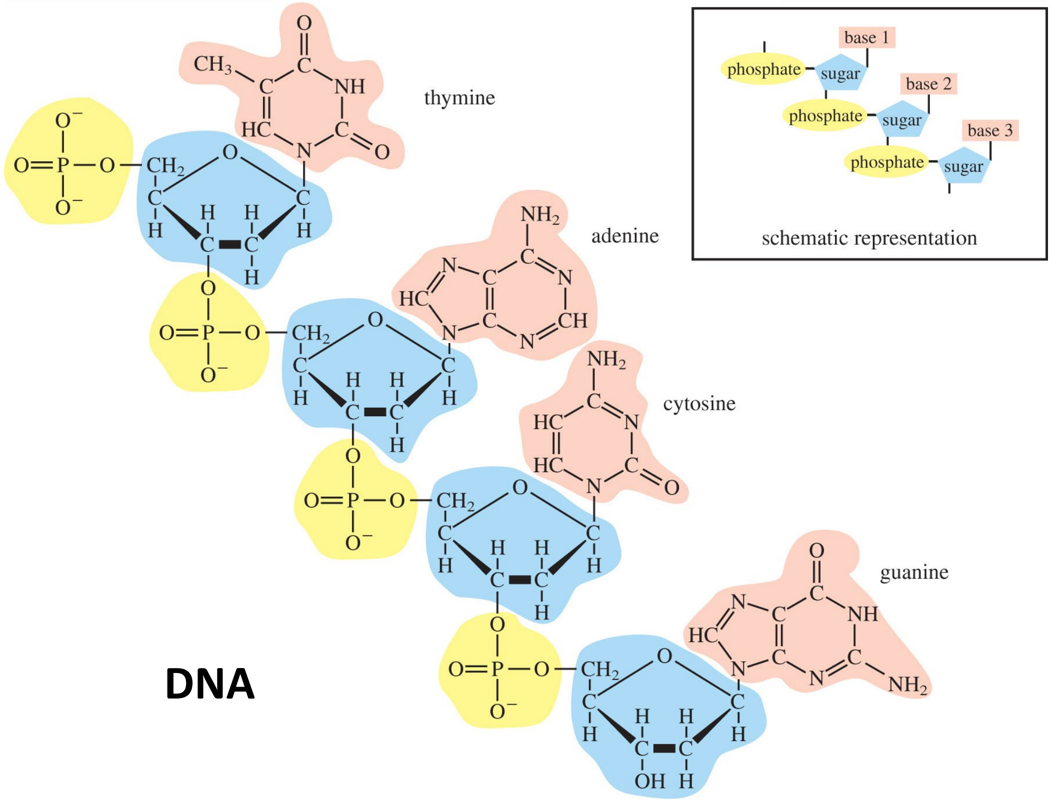


Adenine monophosphate

Nucleotide

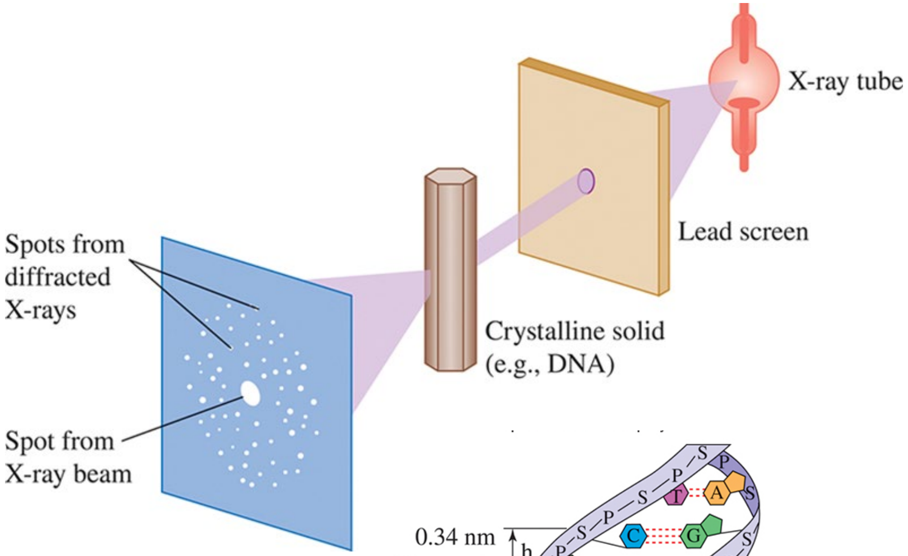


Adenine triphosphate
(ATP)



DNA

X-ray diffraction to see the shape of DNA

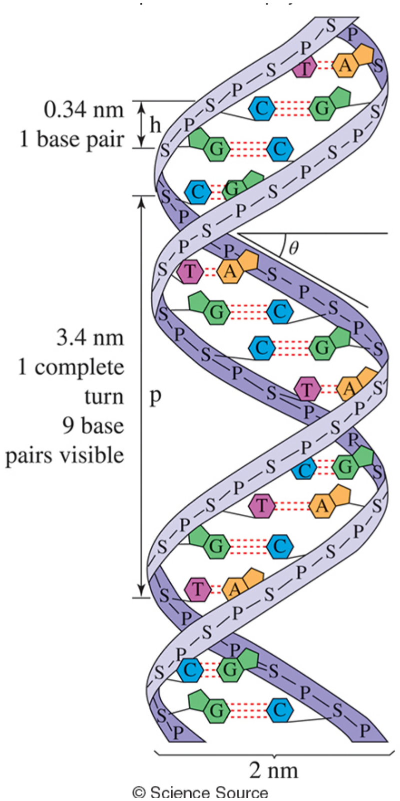


X-ray photons interact with **the atoms** in the crystal (DNA)

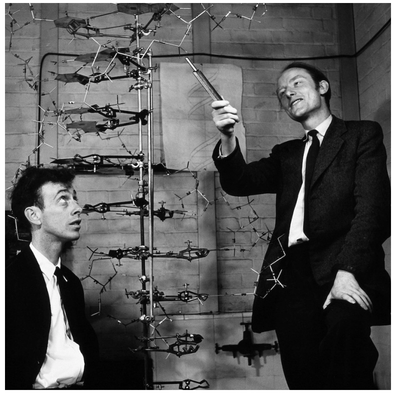
X-rays are **scattered** at certain angles that are **related to the distance between** atoms

The **scattering information** provides the **structural information** of the analyte

X-ray data
(**Rosalind Franklin**)



DNA structure
(**Watson/Crick**)



1 base pair = 0.34 nm

One turn = 3.4 nm (10 base pairs)

Between the strands = 2 nm

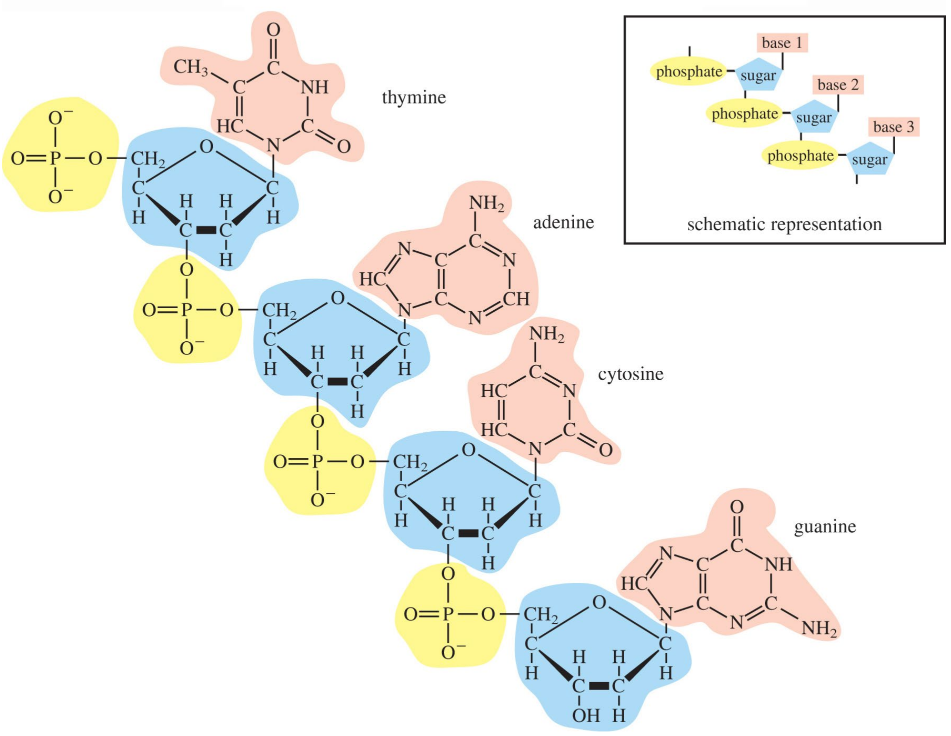
Complementary base pairs

“ward” ≠ “draw”

the words have the same letters but different meaning due to the order

T-A-C ≠ C-A-T

The sequence matters and is **direction sensitive**

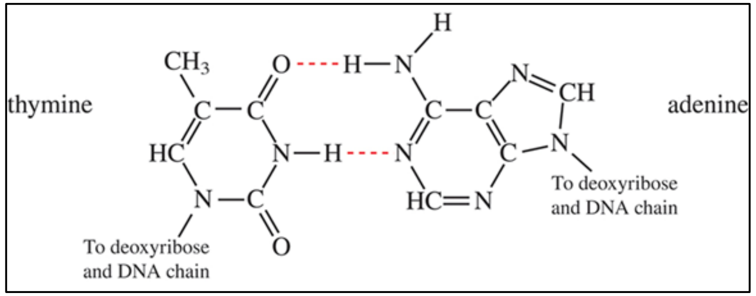


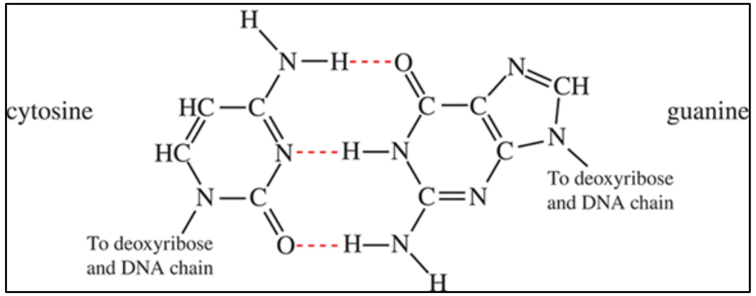
The deoxyribose ring has two different types of chemical bonds with the phosphate backbone

→ 5’ to 3’ is the standard direction

Complementary base pairs

Higher AT?



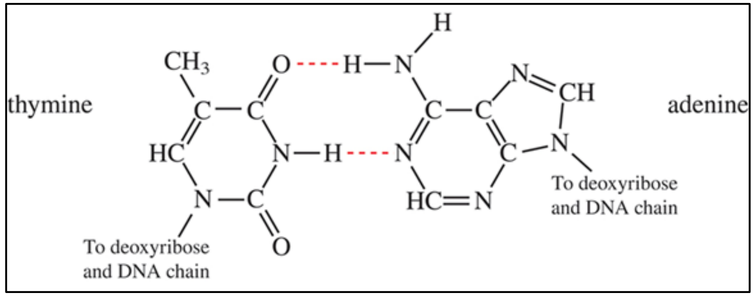


Bases in DNA come in pairs;
(A-T) and (G-C)

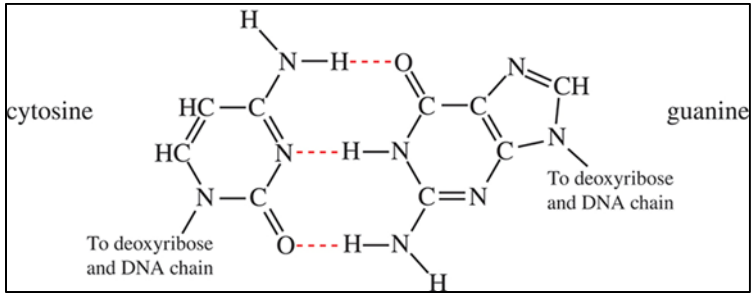
Specific Name	Common Name	Adenine	Thymine	Guanine	Cytosine
Homo sapiens	human	31.0	31.5	19.1	18.4
Drosophila melanogaster	fruit fly	27.3	27.6	22.5	22.5
Zea mays	corn	25.6	25.3	24.5	24.6
Neurospora crassa	mold	23.0	23.3	27.1	26.6
Escherichia coli	bacterium	24.6	24.3	25.5	25.6
Bacillus subtilis	bacterium	28.4	29.0	21.0	21.6

Chargaff's rule → %A = %T and %G = %C

Complementary base pairs



(A-T) has two H-bonding



(G-C) has three H-bonding

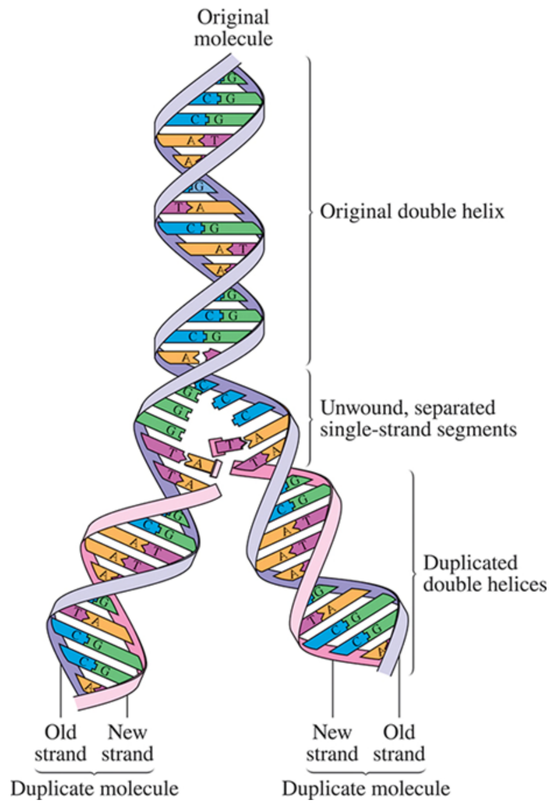
These pairs fit almost perfectly together like a jigsaw puzzle

(leading strand)
ATC TCG GGC ATC

TAG AGC CCG TAG
(complementary strand)
Almost always! (vice versa)

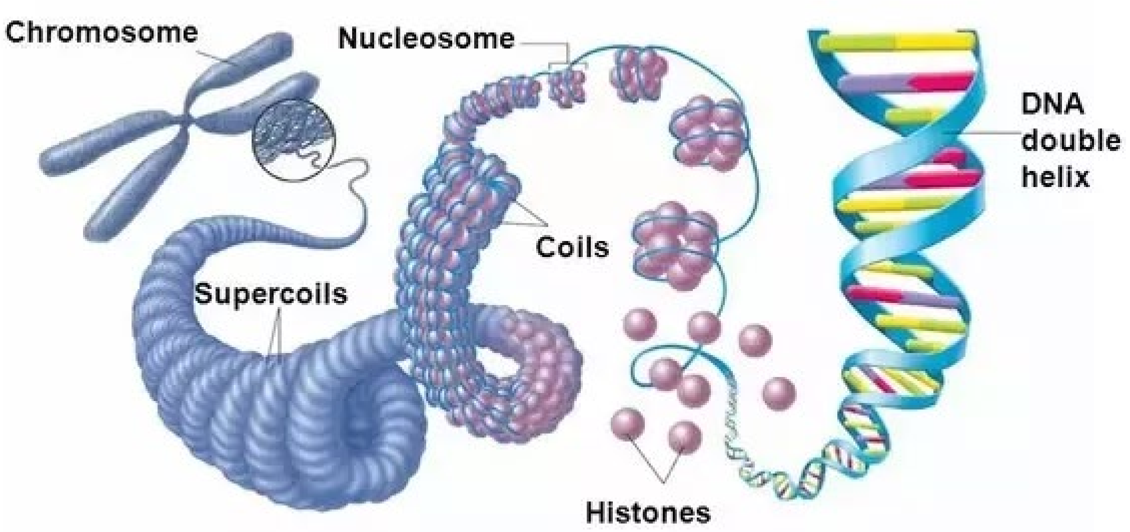
One side of the DNA strand contains all the information required to generate its partner strand

DNA Replication



For a cell division, we need TWO sets of DNA = **replication**

The DNA strands separate → complementary strands are made
Original strands = Newly synthesized strands



23 Chromosomes of the 3 billion-bp human genome
= 1 m of contour length = size of each chromosome → 1.3~10 µm

Q: how such vast quantities of genetic information can be scanned and decoded in a reasonable time while stored in a small portion of the cell's volume?

https://youtu.be/OjPcT1uUZiE

Your Turn 13.7 DNA Sequence Repair

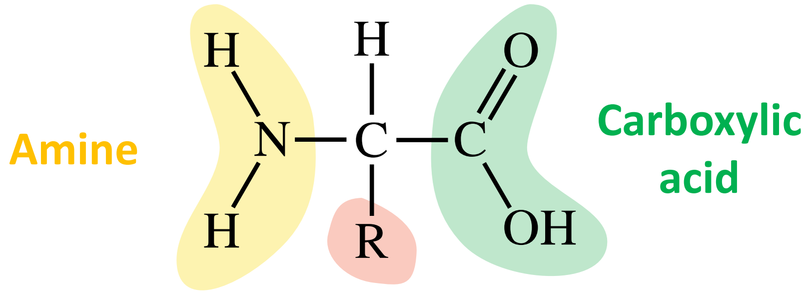
- DNA must be copied with utmost perfection. After replication, enzymes scan over strands of DNA to identify and correct errors in base pairing.
  - ATGCCATGAA
  - TACGGTATTT

a. Find the error in base pairing for this set of DNA strands and circle it.

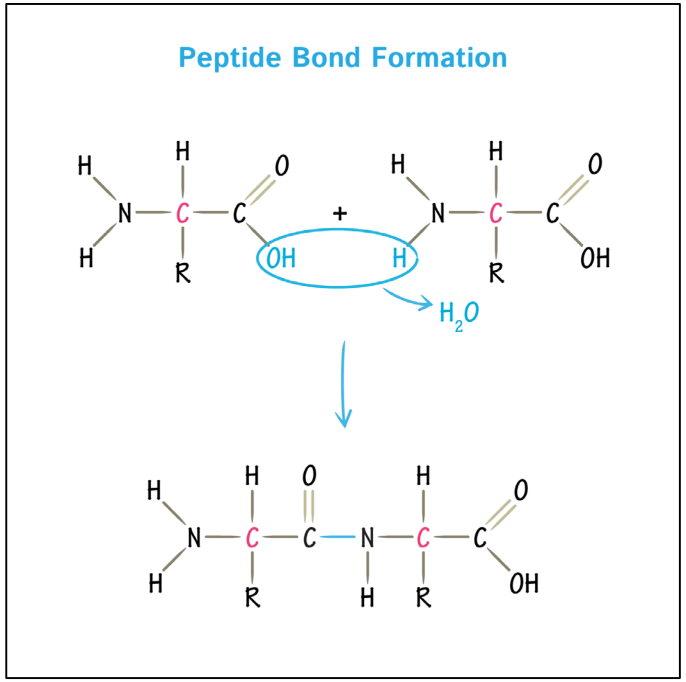
b. Do you expect the mismatched strands to be more or less stable than a correct pair? Explain your reasoning.

Amino acids to proteins

The information is carried in DNA, but it is expressed in other molecules (**proteins**)



= amino acids



- Proteins are everywhere in our body
- Enzymes regulates our body

DNA → Protein

A, T, G, C → 20 amino acids

$4 = 4$

$4^2 = 16$

$4^3 = 64$

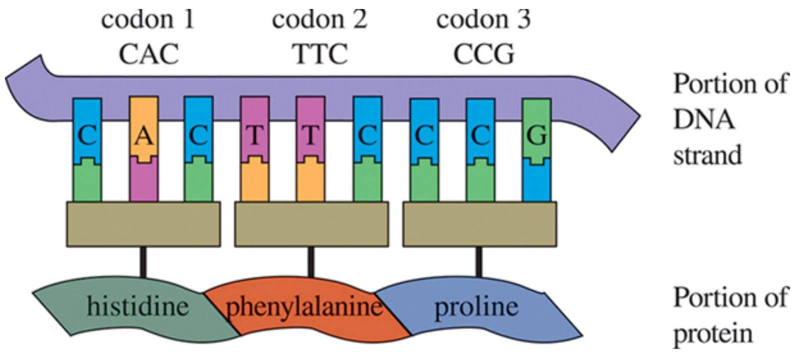
Connected amino acids = **proteins**

Codon

Adv?

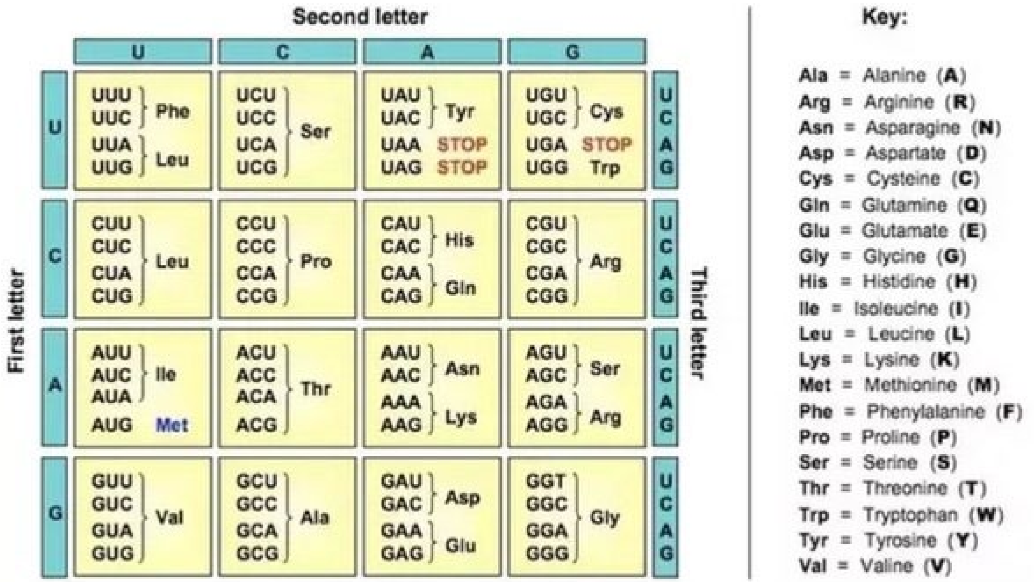
The three-letter groupings of nucleotides

= the basis of the information transfer from DNA to protein≈



64 possible combinations of DNA, but only 20 amino acids?

→ **Redundancy**!!



Second letter

First letter	U	C	A	G	Third letter
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	GGT Gly GGC Gly GGA Gly GGG Gly	U C A G

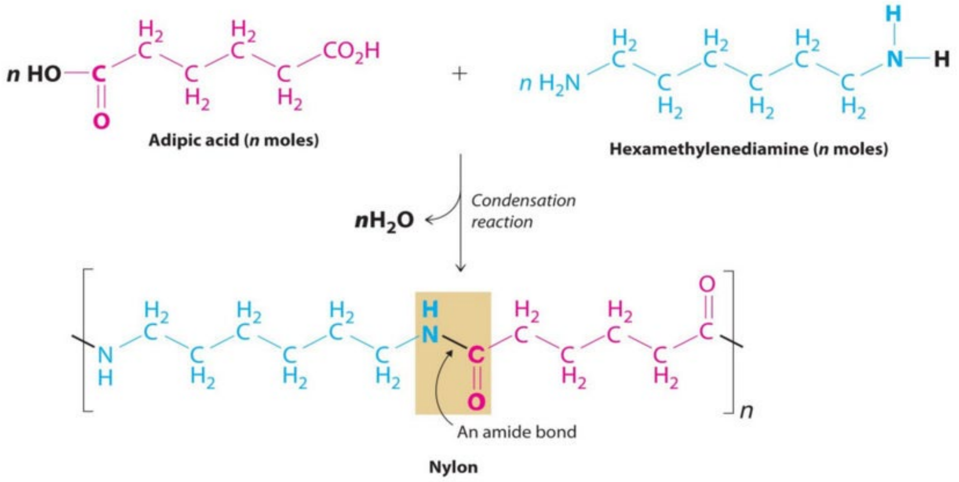
Key:

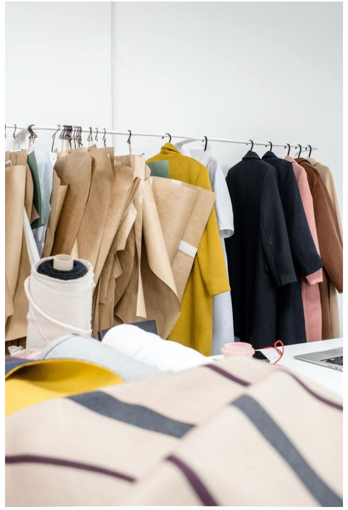
Ala = Alanine (**A**)
Arg = Arginine (**R**)
Asn = Asparagine (**N**)
Asp = Aspartate (**D**)
Cys = Cysteine (**C**)
Gln = Glutamine (**Q**)
Glu = Glutamate (**E**)
Gly = Glycine (**G**)
His = Histidine (**H**)
Ile = Isoleucine (**I**)
Leu = Leucine (**L**)
Lys = Lysine (**K**)
Met = Methionine (**M**)
Phe = Phenylalanine (**F**)
Pro = Proline (**P**)
Ser = Serine (**S**)
Thr = Threonine (**T**)
Trp = Tryptophan (**W**)
Tyr = Tyrosine (**Y**)
Val = Valine (**V**)

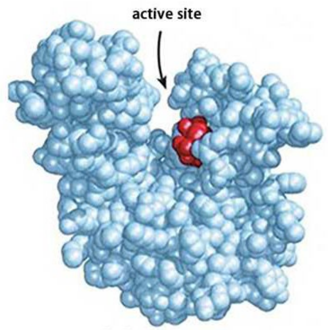
Protein structure

Protein vs Nylon

= they are very different!







20 amino acids vs **2 monomers**

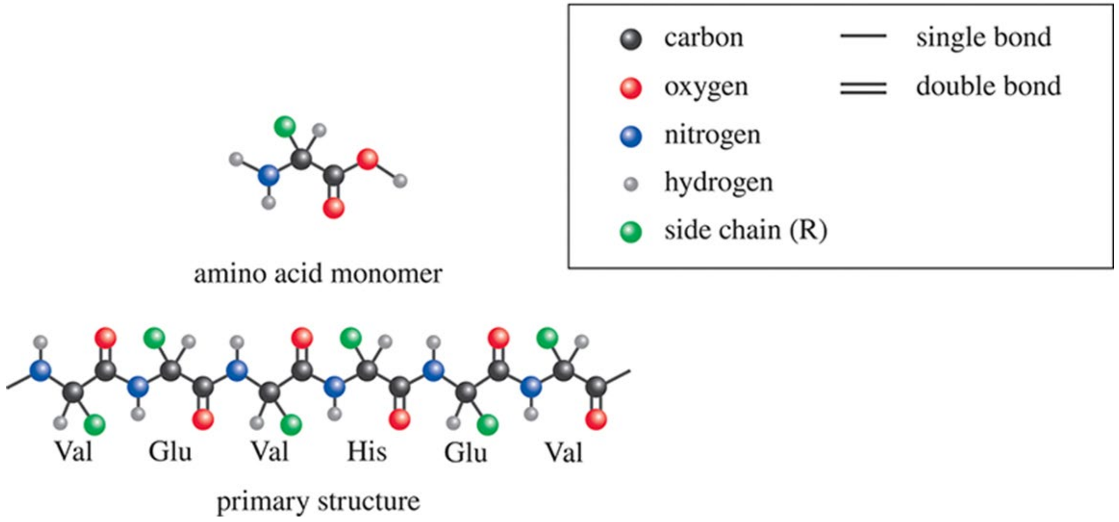
→ **Different structures**

Proteins are not long strands, but collapse into a unique 3D structure

→ The shape is necessary for the function performed in the organism!

Protein structure

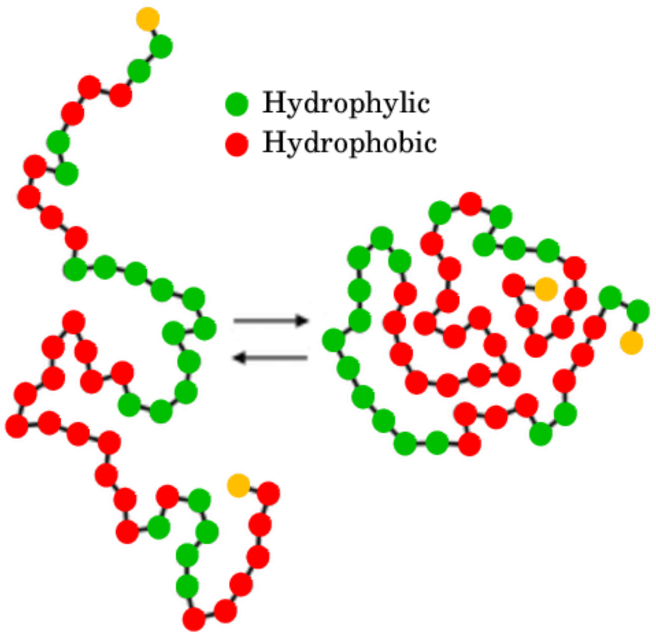
Primary structure: The unique sequence of the amino acids
(most basic identifier of a protein)



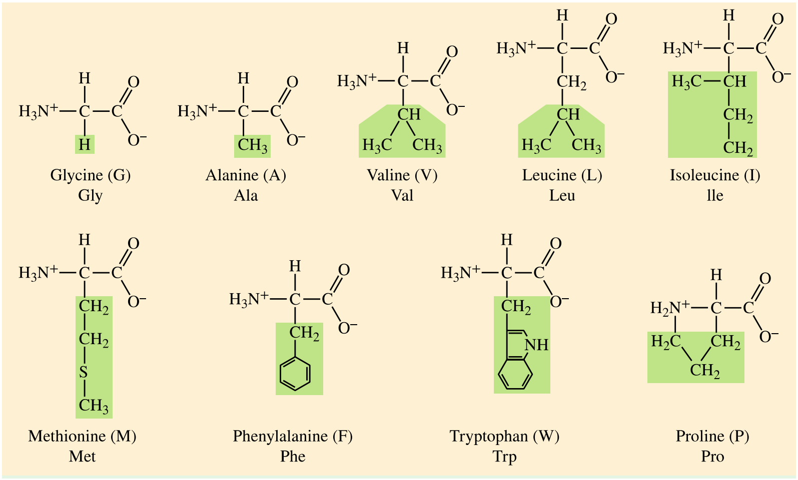
→provides size + a few other details

→ but not sufficient to specify a protein
And explore its shape and function

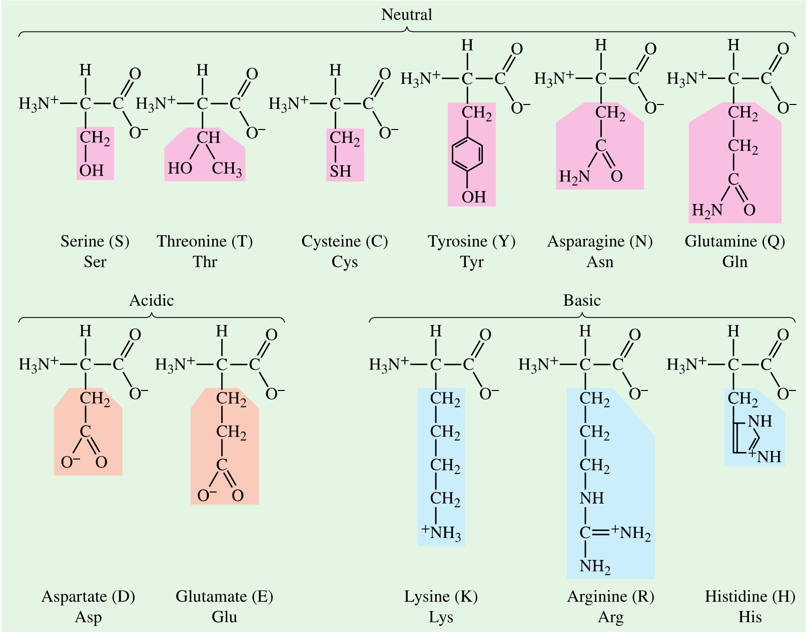
Three-dimensional protein structure



Nonpolar



Polar



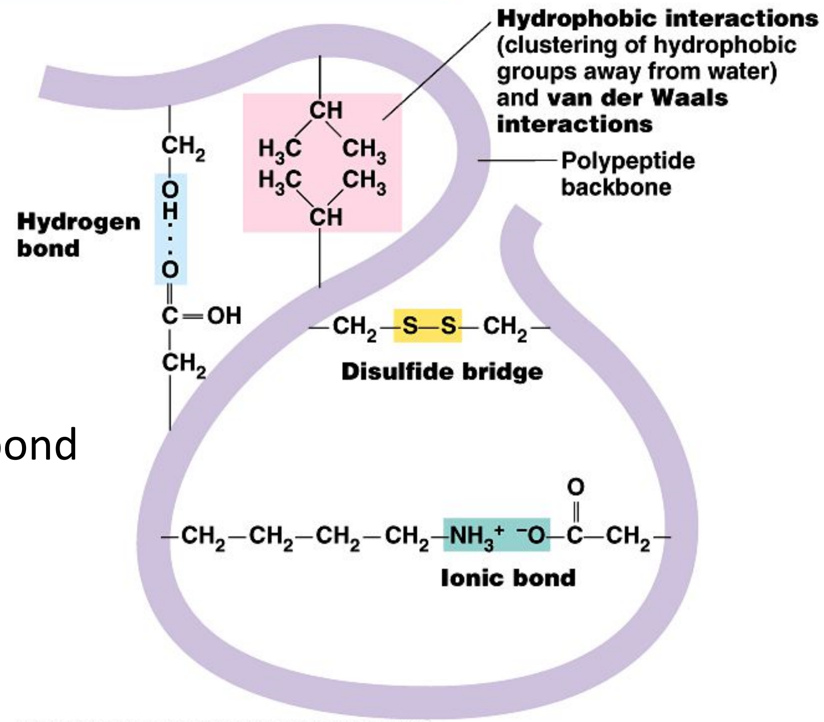
Interactions between side chains hold a protein in a particular overall shape

- **polar**: tends to stay in the water
- **nonpolar**: avoid interactions with water

→ Like oil and water, they tend to separate

https://lab.concord.org/embeddable.html#interactives/samples/5-amino-acids.json

Types of interactions in protein



Polar side chain's interactions

- Ionic attractions
  → between charged side chains
  (i.e., carboxylic acid, amine….)

- Hydrogen bonds
  → Uncharged, yet attract one another by H- bond
  (i.e., hydroxyl, amide ….)

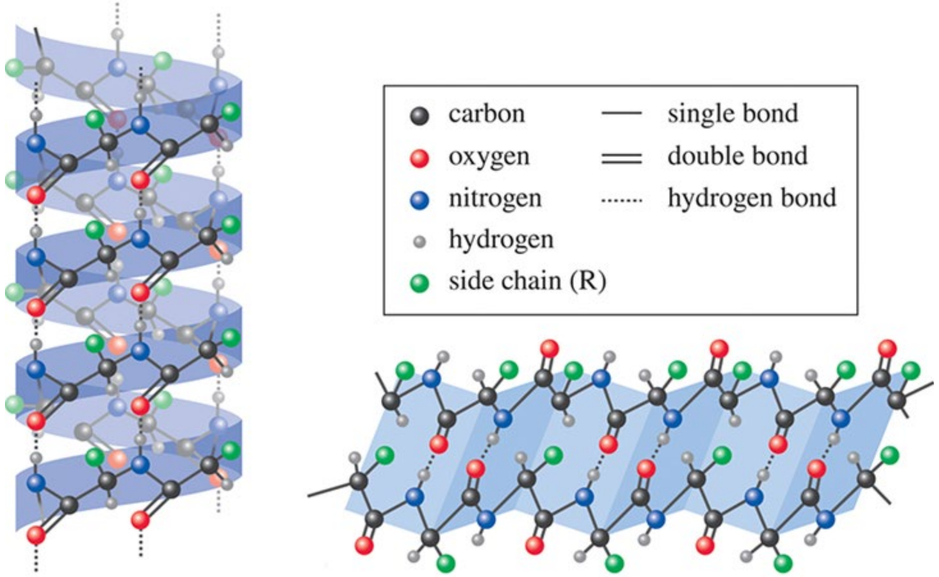
Nonpolar side chain's interactions

- Van der Waals attractions
  → Clustering of hydrophobic groups away from water (inside of proteins)

Thiol groups interaction

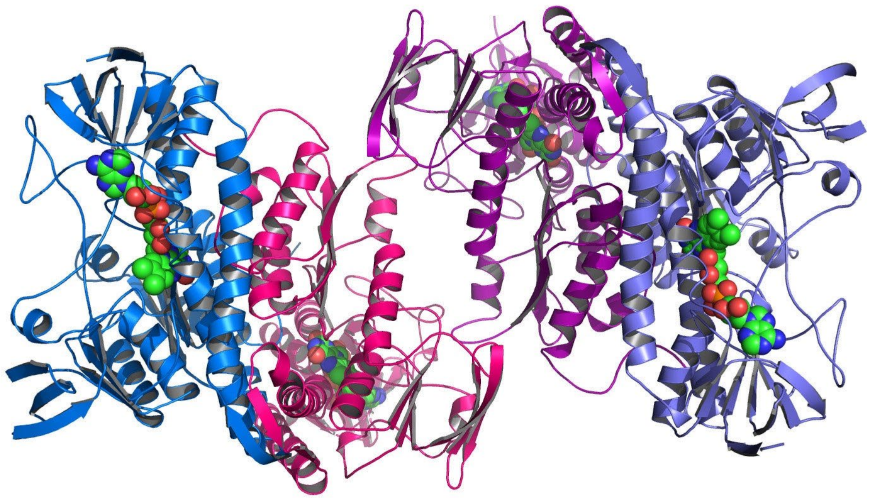
- Important and highly specialized
- Two thiol groups form **disulfide (S-S) bonds** between the adjacent sulfur atoms
  → Unique structure

Three-dimensional protein structure



α-helix

β-pleated sheet



Many protein chains form regular, repeating structures (<u>α-helix, β-pleated sheet</u>)

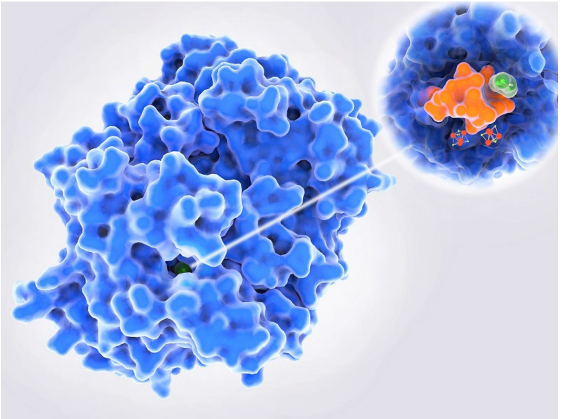
= secondary structure

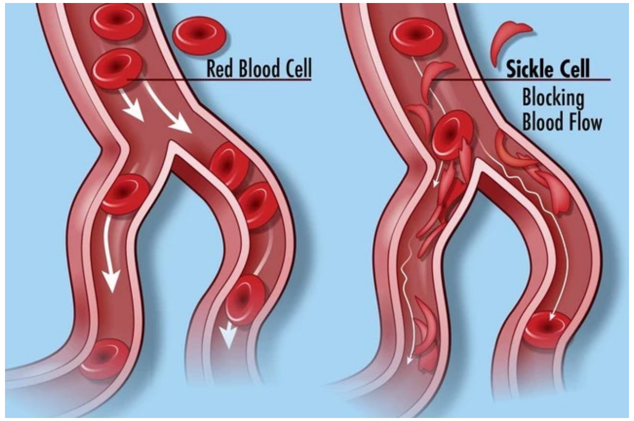
<u>***α*-helix**</u>: a spiraling strand

<u>***β*-pleated sheet**</u>: extended strands stretching alongside each other with a slight zigzag

→ Both structure depend on the intramolecular hydrogen bonds in protein backbone

Three-dimensional protein structure





Tertiary structure
= “global” description of their shape

From only 20 amino acids → a variety of shapes and **a large array of functions**

<u>3D structure dictate specific functionalities</u>
=a subtle change in the primary structure → a profound effect on its function

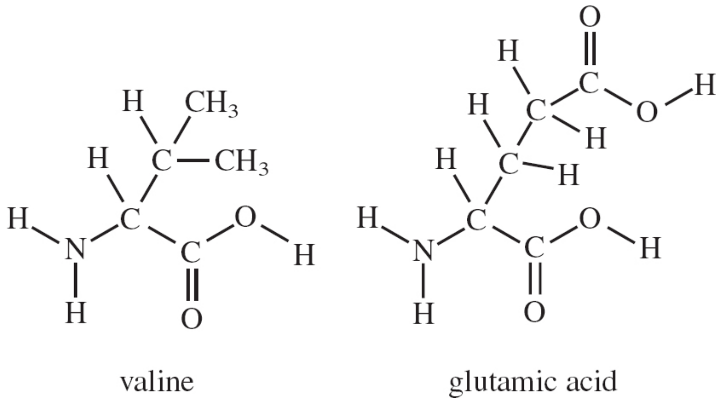
(i.e.) Active site = it binds only specific reactants and accelerates the desire reaction

-Leu → Val: the protein may a little less stable but on the whole the same
- Glu → Val: “sickle-cell disease”

Your Turn: Structure versus Properties of Proteins

Your Turn 13.13 Function Follows Form

- In sickle-cell anemia, a glutamic acid residue in the sequence of hemoglobin is replaced with a val ine residue.



a. Describe the structural difference between these two amino acids: valine and glutamic acid.

b. Predict the solubility for each of these amino acids in water.

c. Explain how these differences could give rise to the deformed cells typified by sickle-cell anemia.

Genetic modification

Genetic modification

<Domesticating a plant>

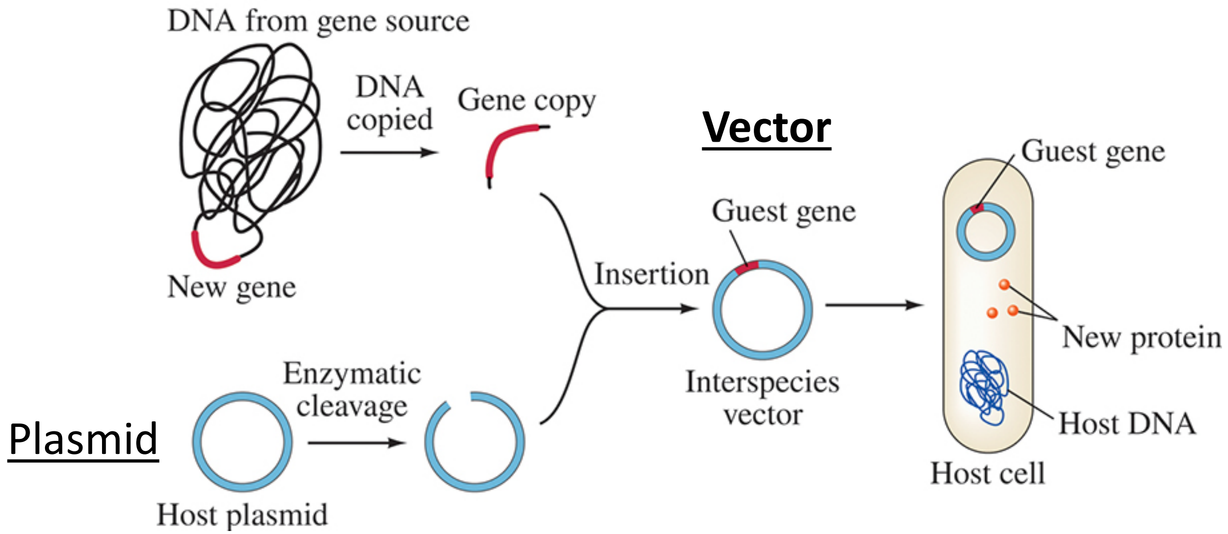
= selecting plants with certain traits (sequence of DNA) and rejecting others

→ slowly over time, we encouraged changes in the DNA, allowing one DNA to carry on, spread, and survive

<A new virulent bacteria from a mutation>

→ Over time, the new bacteria will become prevalent

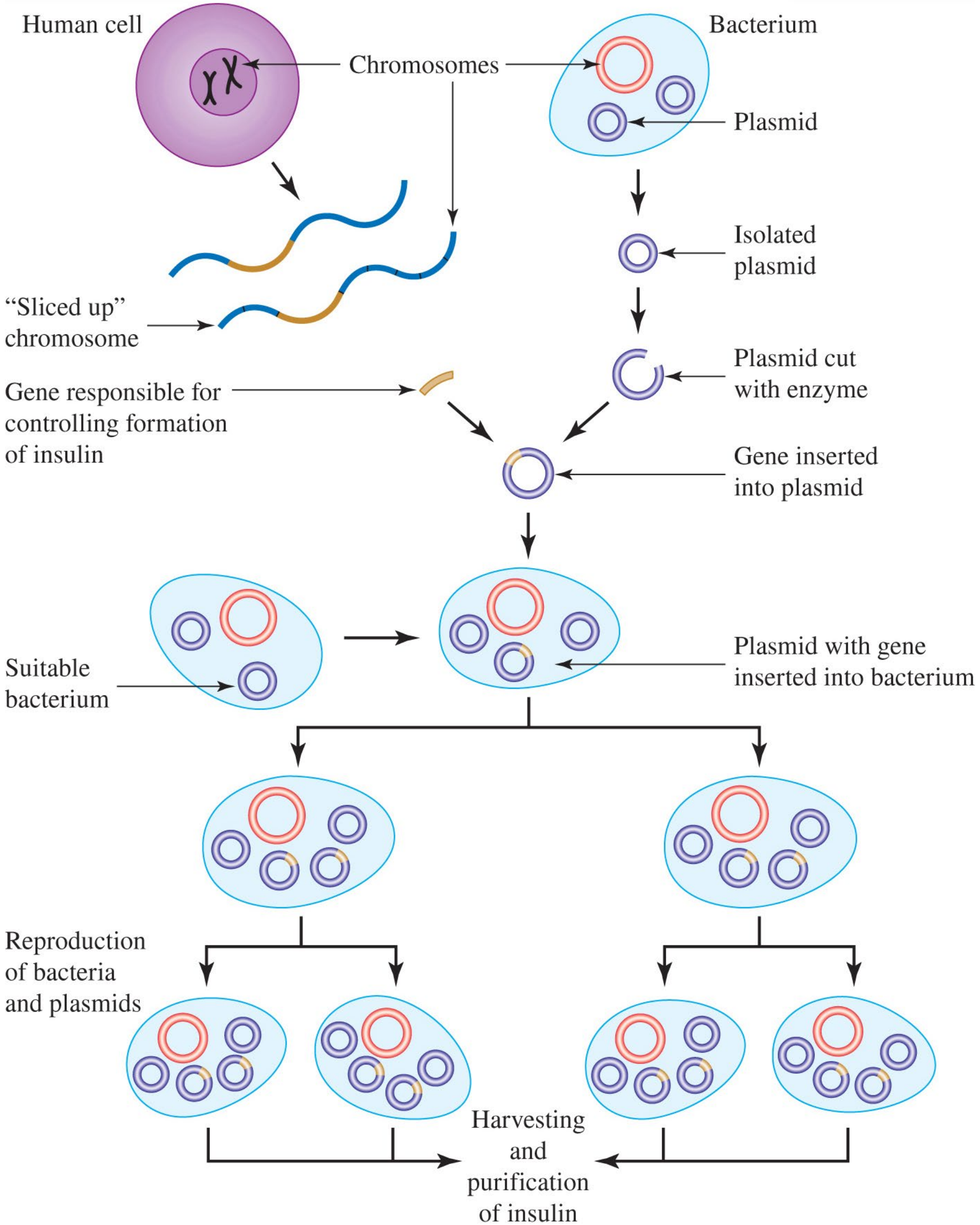
Genetic engineering = direct manipulation of the DNA in an organism



If we change the genes, we change the proteins synthesized by these genes

→ **Bacterial cells can generate human insulin!**

Genetic modification



Genetic engineering

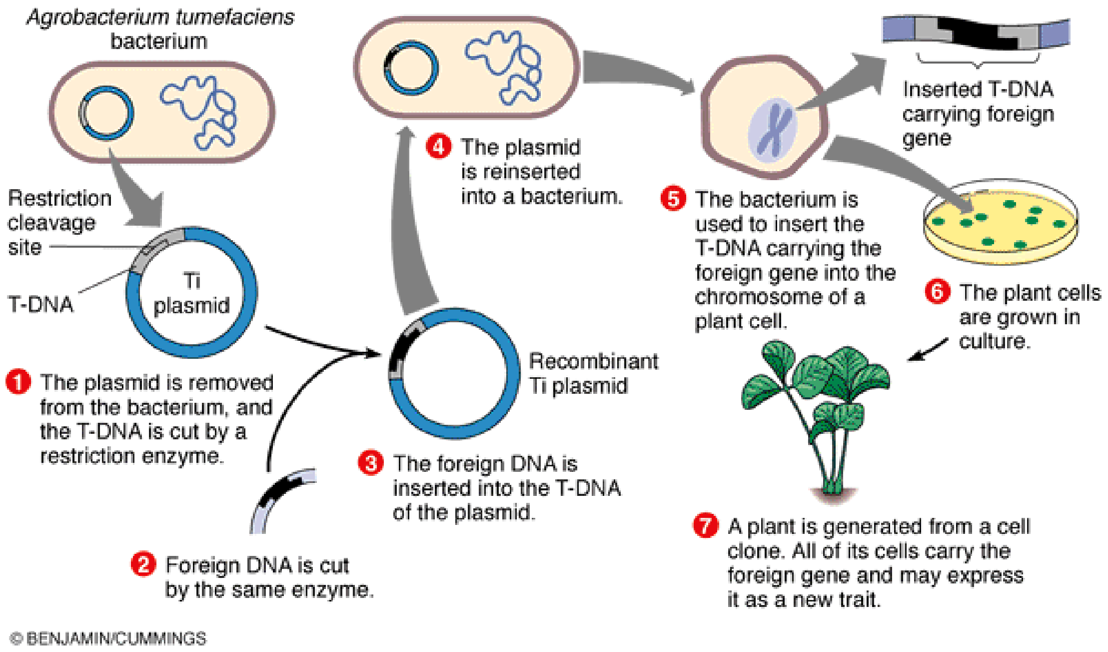
©Nigel Cattlin/Science Source

Genetic engineering in plants?

Inserting foreign DNA becomes trickier as one climbs the evolutionary ladder from bacteria to plants

→They are better at protecting themselves = destroy foreign DNA!!

Genetic engineering



A special soil bacterium has the ability to "infect" plants

= The bacterium creates a bridge into the plant cell

= it transfers its own DNA into the genome of the plant

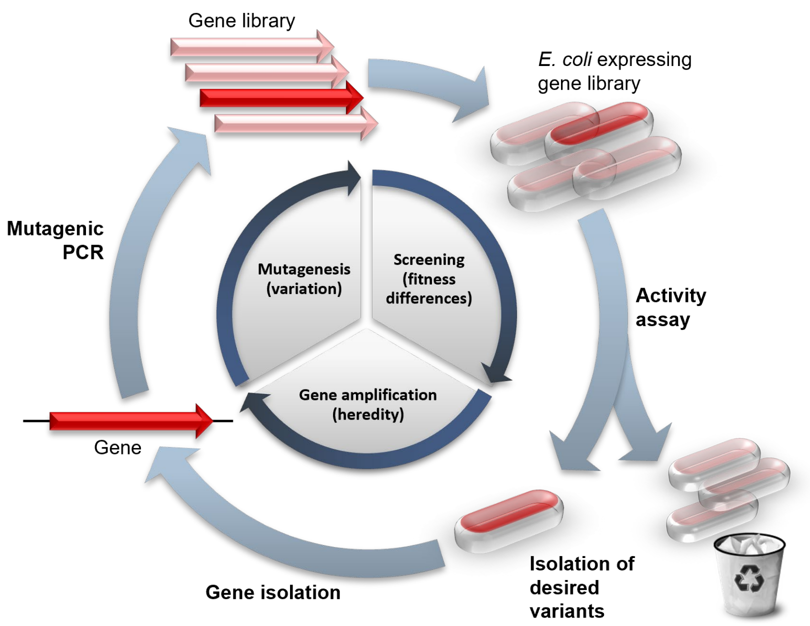
The bacterium = ***Bacillus thuringiensis* (Bt)**

→ Can be engineered to produces a **Bt-toxin**

→ **Bt-plants can produce toxins against crop-destroying pests!**

(="transgenic plant")

Genetic engineering



Let's say that you have a trait in mind, but how can we achieve the trait?

Directed evolution

→ **Random mutations to bacteria (by chemicals or ionizing irradiation)**

→ Some seeds do not grow, other grow but with no change

→ **Still others show unique traits!! >> breeding >> isolate the gene**

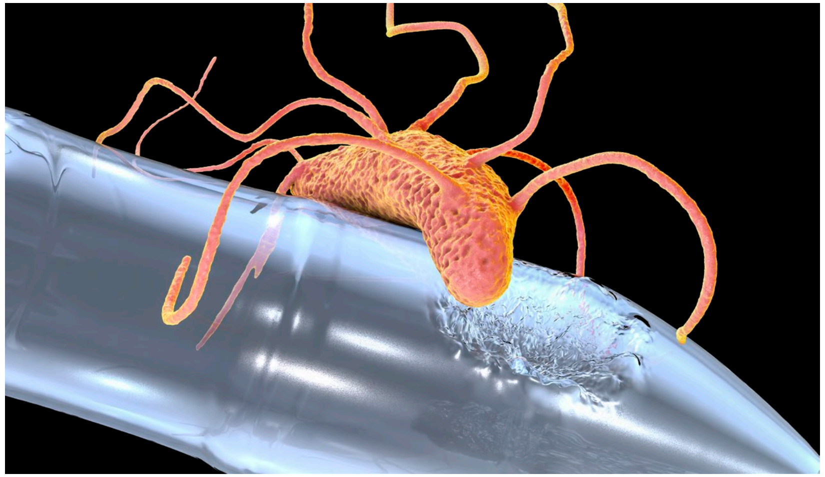
Genetic engineering

Engineered organisms can even produce **plastics**

The synthesis of plastics consume large amount of chemical reagents and energy
(often derived from petroleum)

But these organisms produce monomers from corn, sugarcane, and vegetable oils,
then catalyze the polymerization reaction
(i.e., **PHB = polyhydroxybutyrate**)

PHB is similar to PP but is biodegradable



Some bacteria even EAT plastics for our nature!

“Greener” chemistry

Genetic engineering = “Greener” chemistry

= minimal waste, reusable reagents, low toxicity

What can we do with genetic engineering?

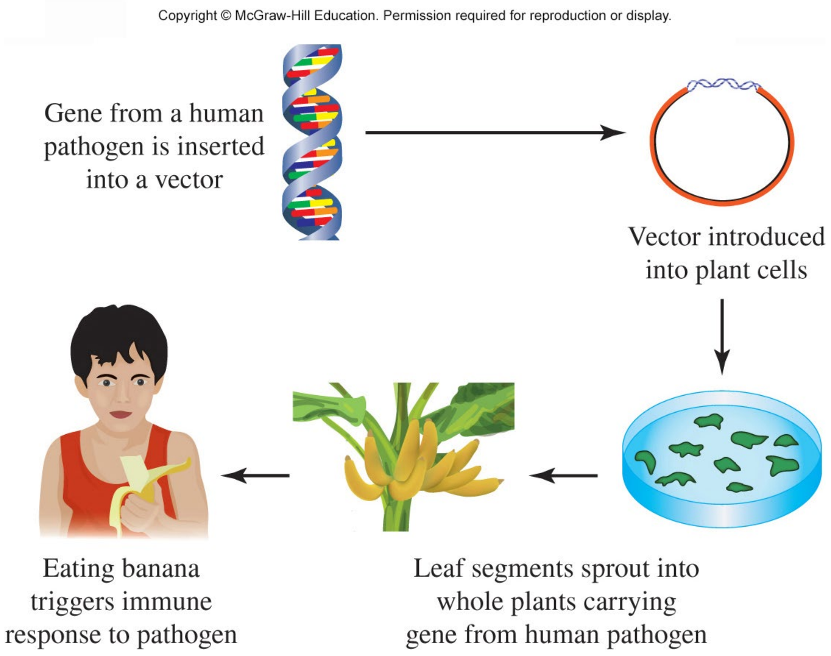
Example 1: Plants that absorb toxic metals from contaminated soil
Example 2: Soybeans that produce high yields of biofuel
Example 3: bacteria that detect and remediate radioactive contaminations

A synthetic route to a desired chemistry may require toxic compounds and large amounts of solvents, and produce a staggering amount of waste *(**up to 100 times the weight of the desired compound**!)*

These “biological machines” perform reactions faster and safer with fewer toxic reagent and less waste.
Also enzymes can be used over and over again!

Downside?

What if we can "farm" vaccines?



Gene-engineered banana that produce vaccines against infectious diseases of the intestinal tract

→ **Eating banana**
→ **Immunity!**

Most vaccines requires refrigeration, and special handling together with trained professionals to administer them

→**Edible vaccine would be easy to transfer and administer**
(Beneficial in developing countries)

Genetically modified organisms (GMO)

Created for beneficial purpose, producing/reducing chemicals

Papaya ringspot virus took a severe toll on production (1990s, Hawaii)

University of Hawaii
→ Developed GMO papayas resistant to the disease
→ Now > 80% papaya consists of the tolerance

250,000~500,000 children become blind and ~667,000 children dies every year from vitamin A deficiency

Swiss Federal Institute of Technology
→ Developed a strain of rice that provides vitamin A
= "golden rice"

Genetically modified organisms (GMO)

“Genetic engineering is not natural”

Counterpoint: The introduction of foreign DNA into a plant’s genome is also found in nature. For instance, several species of sweet potato now exist due to bacterial DNA being inserted into sweet potatoes at least 8000 years ago.

“GMOs cause health problems such as allergies and cancer”

Counterpoint: It is possible for a new gene to express a protein that is an allergen or toxin, this is quite rare. Published safety reports on GMOs show no evidence that the foods are dangerous.

“GMOs cause farmers to overuse pesticides and herbicides”

Counterpoint: Some types of crops have been developed that actually reduce reliance on pesticides. Glyphosphate is often used, which is less toxic than other herbicides & doesn’t require as much per acre.

https://youtu.be/0sR5GHi19HU

- What is the difference between human DNA and bacterial DNA?
- Discuss techniques to study the structure of DNA other than X-ray
- What are the major driving forces stabilizing the DNA double helix?
- Discuss how to add unnatural amino acids
- Discuss the advantages and limitations of AlphaFold
  (AI-based protein structure analyzers)