7.012

Lecture 1

9/5

- professor has a math PhD???
- other professor does chemistry??? Wanted to study drama too

Important central idea of this class: unifying principles. Not facts only!

Important stuff

- Read the course website.
  - → Course policies, recitation sections, etc.
  - → Office hours, recitations start next week
- Fill out the mock submission survey!
- MITx?
- 4 exams + final. (This is a change!) Also, you can drop an exam.
  - → You DO need to show up.
- Problem sets cannot be late. Collaboration is fine, but not copying.
- Different styles of learning
  - → Paper textbook (Life), series of videos (see MITx), online textbook (!)
  - → Choose what you want for your own learning

What is this class?

- Medicine rapidly developing in the present
- Cancer therapy, etc.
- DNA fingerprinting → then privacy issues, ancestry, etc.
- Constantly changing & driving forward? Things like genome editing!
- "Diversity of life" — what's the common operating system?
- Different "levels" — biospheres, ecosystems, organisms, organs, tissues, cells, organelles, molecules (big to small)
  - → We focus on the small things because they are common.
- Evolution — we won't talk much about it. 60 seconds:
  - → 4.5 billion years — Earth forms
  - 4 billion yr — cools
  - 300 mil / 3.7 billion — life forms, somehow (prokaryotes)
  - ! 2200 mil / 1.5 billion — first nucleated cells (eukaryotes)
  - → 0.5 billion yr — multicellular organisms!
  - 0.005 billion yr — human-like lineage
  - 0.001 billion — Homo sapiens
  - 0.0000002 billion — MIT founded.

- Cell biology — covered a bit.
  - → Important: prokaryotes ~1-2 μm vs eukaryotes ~40 μm.
  - → There is no "the cell," but it doesn't matter that much.
    - General abstract principles!
- Organization of class
  - → Biological functions — biochemists (components w/o organism) vs geneticists (organisms w/o components)
    - ↳ results: look at proteins and genes, respectively.
    - ↳ Then in the 20th century: DNA → molecular biology
      - → Idea: RNA from DNA is sent off to a ribosome and creates proteins!
      - → New scientists: recombinant DNA
      - → Then came genomics: study the entire human genome
      - → Cost of human genome: $3 billion → $500
- Basically, we are always doing more and moving forward
  - → Pertubation, not observation
  - → Data scientists
  - → Continual improvement!

7.012

Lecture 2

9/7

- Change recitations now!
- P-set posted on Monday (see instructions on Stellar)
- Fundamentals of life = fundamentals of matter.

Periodic Table

- Important to life: H, C, O, N, S, P
  - → Also, Se, Si, I
- Cations: K^+, Na^+, Ca^{2+}, Mg^{2+}
- Anions: Cl^-
- d-orbital ("d-block") atoms: Fe, Mn, Co, Ni, Cu, Zn, etc.
  - → ~1/3 all proteins are metalloproteins
  - → Good for catalyzing reactions, etc.

Chemical bonding

- Goal is to have a complete valence shell of (e^-)(electrons)
  - → "Noble gas" configuration
- Who gets the (e^-)? Depends on electronegativity! (χ)
  - → ability of atoms to attract (e^-) from other atoms
  - → [On a scale of ~0–4.] Low near bottom right, high near Fluorine

Ex

Na (1 val.) Cl (7 val.)

→ transfer of electron.
= ionic bond; this gives electrostatic attraction.

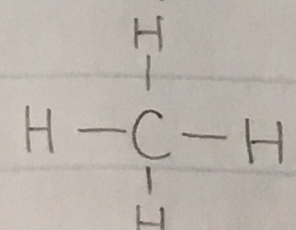
Since $\chi_{Cl} \gg \chi_{Na}$, Cl "takes" an electron from Na.

Ex

H (1 val.) C (4 val.)

⇓

Methane:



Need a lot of electrons → "sharing".

☆ Carbon likes to have 4 bonds.

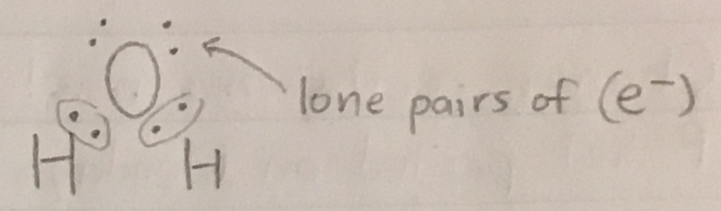
Covalent bond; pairs are shared.

- Nonpolar ≈ equal sharing
- vs
- Polar → unequal sharing

Since $\chi_H = 2.2$, $\chi_C = 2.55$, the difference $0.35 < \underline{0.4}$ (critical value) so nonpolar.

Ex

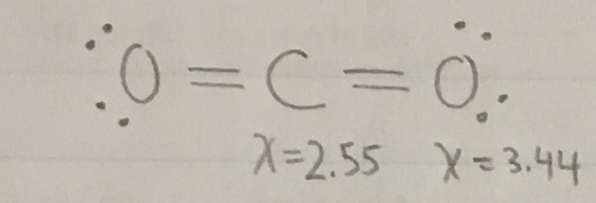
H (1 val.) O (6 val.)



$\chi_H = 2.2$, $\chi_O = 3.44$ → polar.

H_2O = water is polar molecule! (important)

Ex

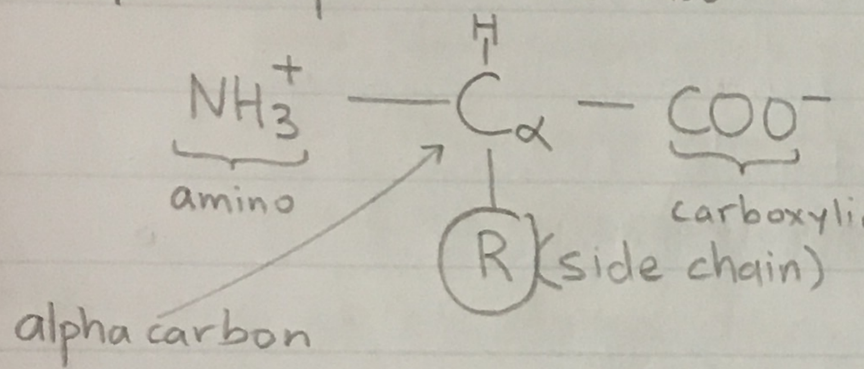


CO_2 = carbon dioxide has double bonds.

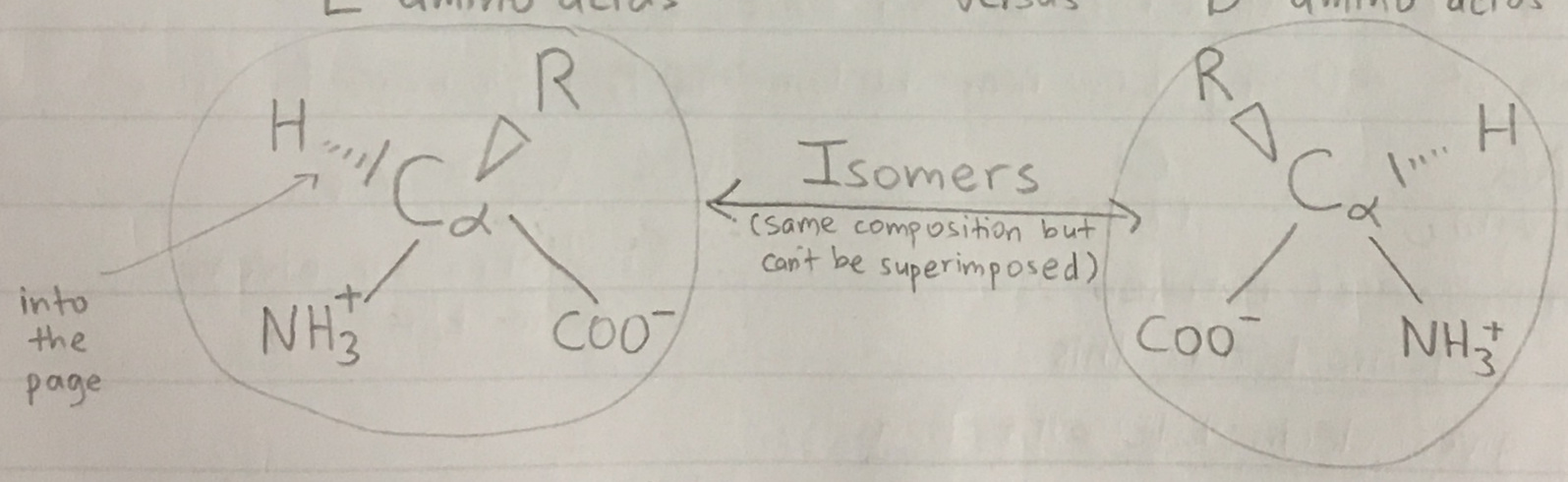
Bonds are polar, but by symmetry, the molecule it is nonpolar.

Molecules of Life

- Peptides / proteins are made of amino acids.



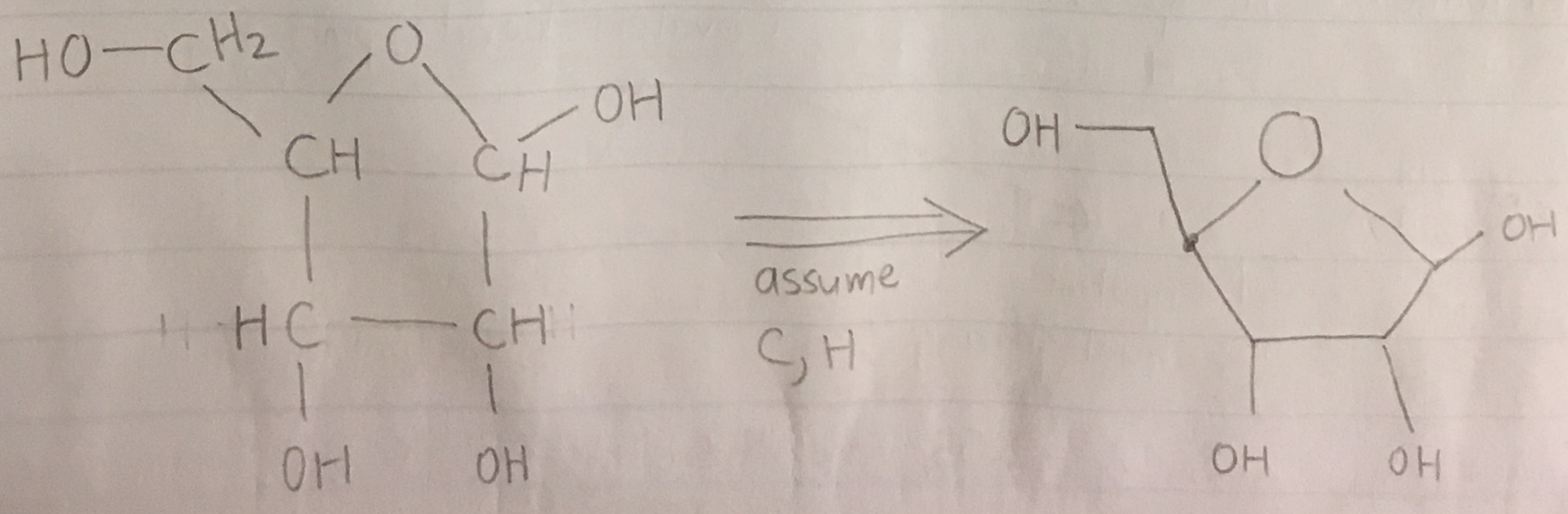
→ L-amino acids versus D-amino acids (mirror image)



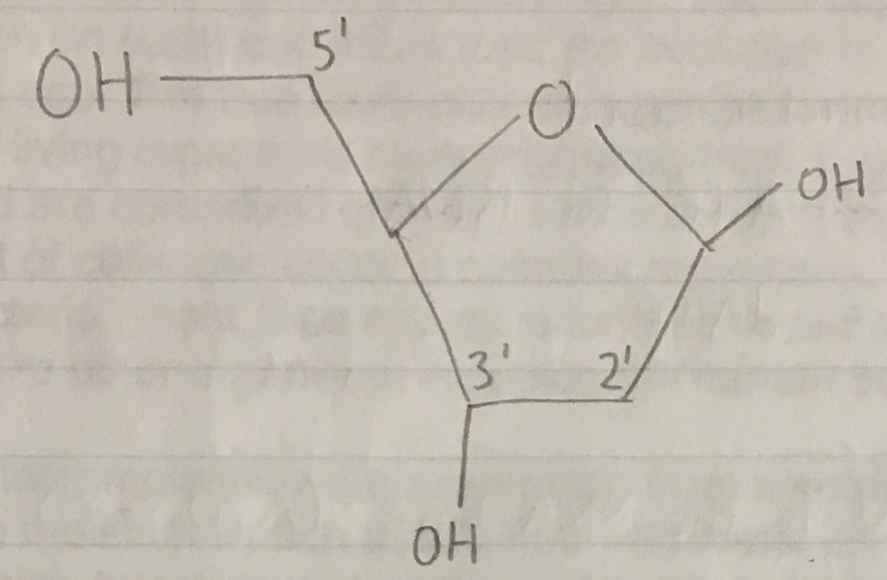
→ Peptide bonds attach amino acids together.

- Carbohydrates (sugars)

Ex Ribose



(Ex) Deoxyribose (missing OH at 2').



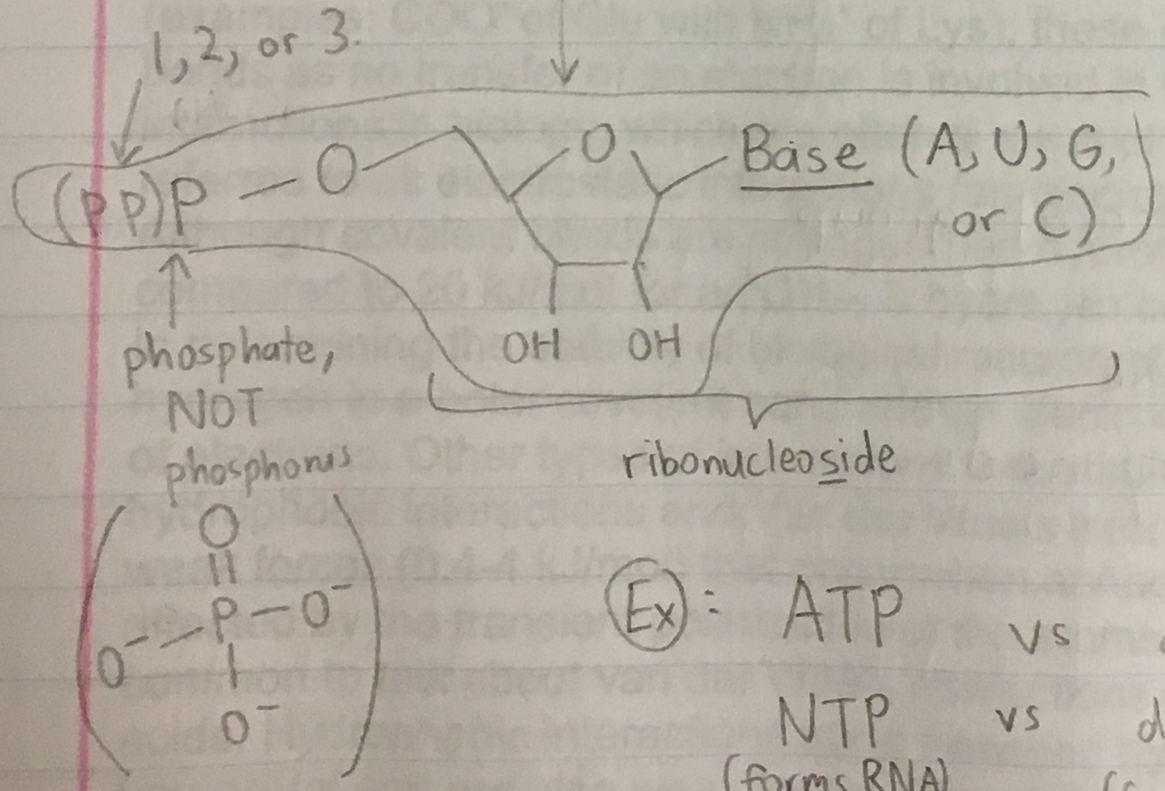
(Ex) Glucose ($C_6H_{12}O_6$)

→ Monosaccharides (single sugars) join and form oligosaccharides;
e.g. starch, cellulose, etc.

→ <u>Glycosidic</u> linkage between sugars.

- Nucleic acids

→ <u>Ribo</u>nucleotides form RNA; <u>Deoxyribo</u>nucleotides form DNA.



Same as RNA, but U becomes T, and missing 2' OH.

(Ex): ATP vs dATP

NTP (forms RNA) vs dNTP (forms DNA)

→ <u>Phosphodiester</u> linkage.

7.012

Lecture 3

- RNA has extra hydroxyl group over DNA
- Mono-, di-, or tri-phosphate

(Ex) If base is A (adenine), the nucleic acid

adenosine triphosphate = ATP for RNA, and

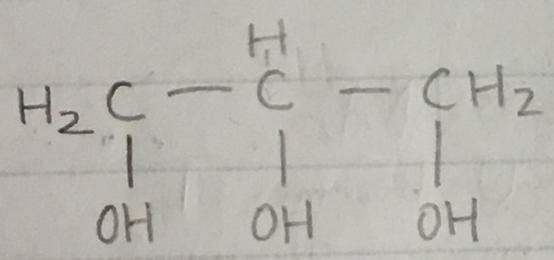
deoxyadenosine " = dATP

- In general, call it NTP where N is the RNA base (A, U, G, C)
  or dNTP where N is A, T, G, or C. (DNA bases)
- NDP is diphosphate, etc.
- DNA stores info; RNA does more stuff
  ⇓ more stable ⇓ catalytic/reactive
- Nucleic acids linked at 5' and 3' carbons by phosphodiester

---

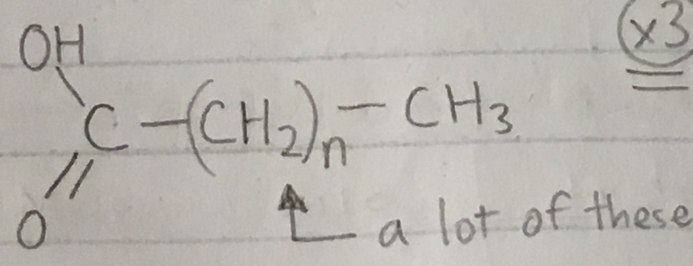
- Lipids (aka fats)

(Ex) triglycerides



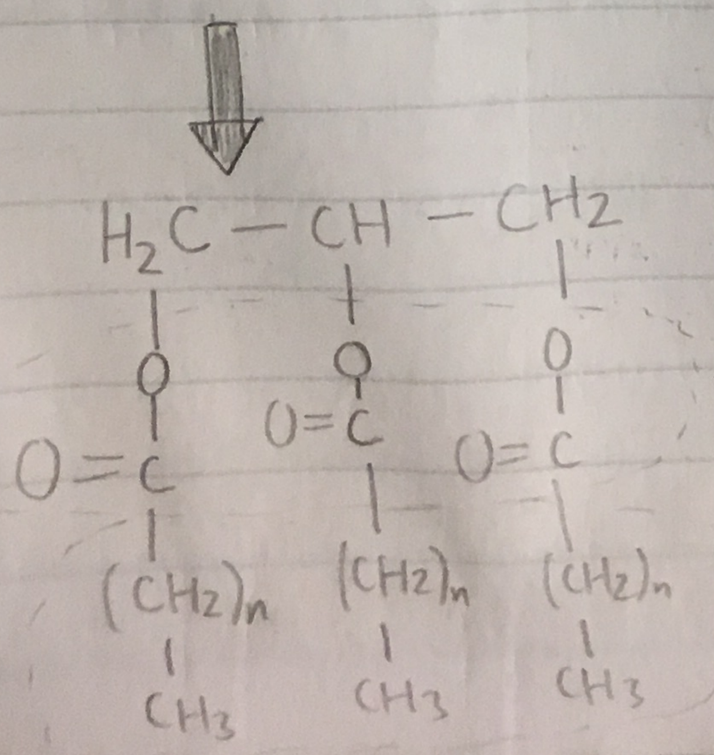
= glycerol group

+



(x3) of these = fatty acids

a lot of these

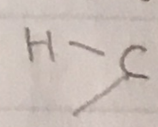


polar

nonpolar; called a hydrocarbon chain

Two kinds

Unsaturated has ≥1 double bond.

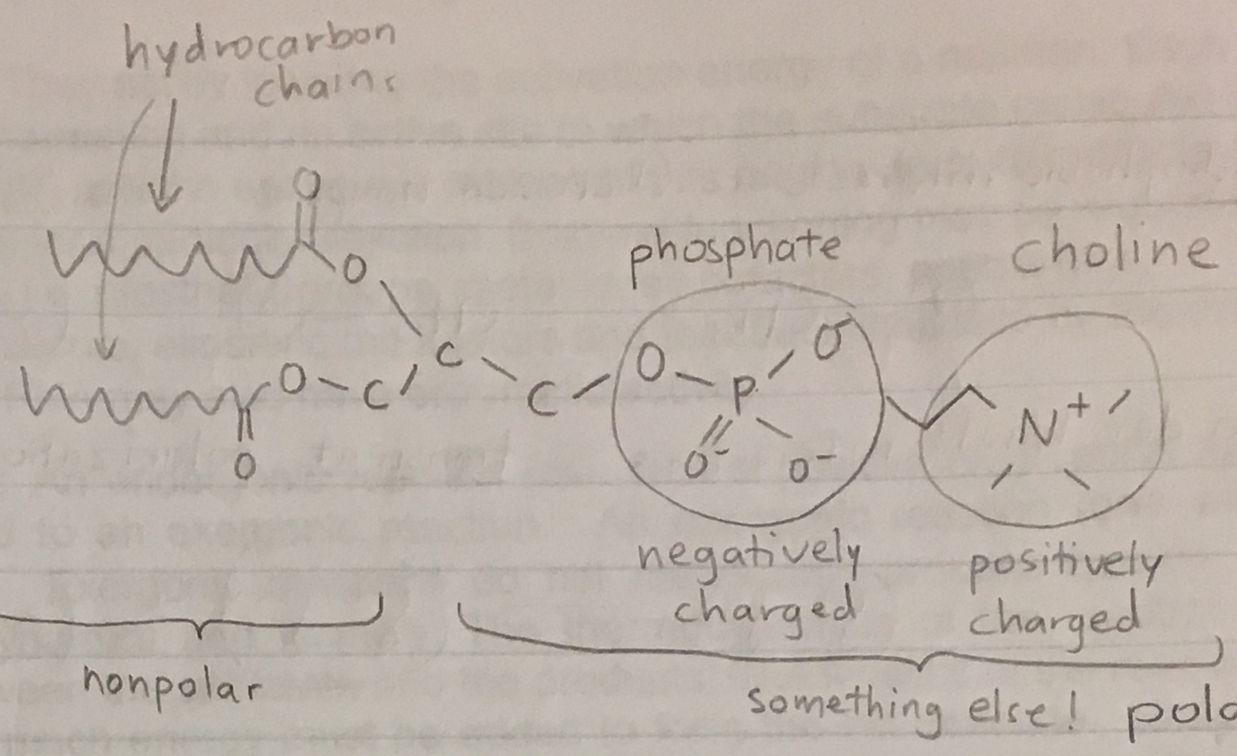


Saturated has none.

Types of Lipids

- Phospholipids.

(Ex) Phosphotidylcholine:



(recall C are not marked)

nonpolar

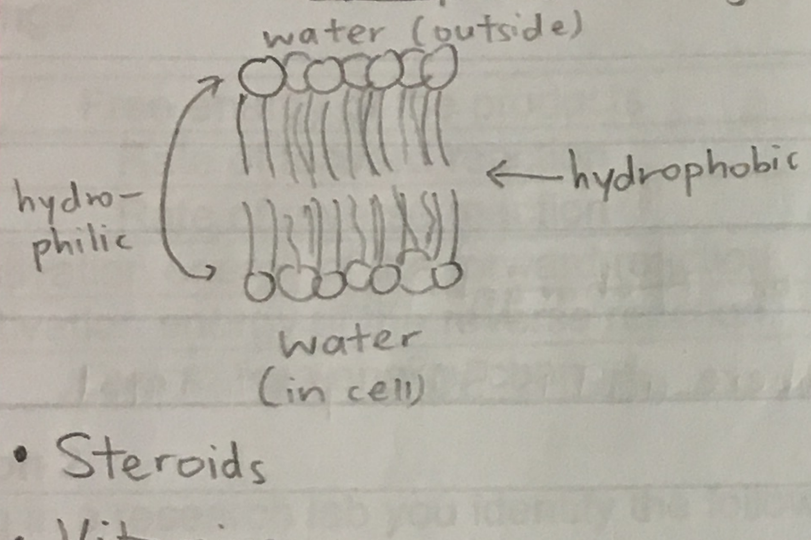
something else! polar.

→ this is called amphipathic.

→ Polar wants to be near water, nonpolar does not.

hydrophilic head group

hydrophobic tail



called the phospholipid bilayer.

This process is driven entropically.
(thermodynamically favorable)

- Steroids
- Vitamins

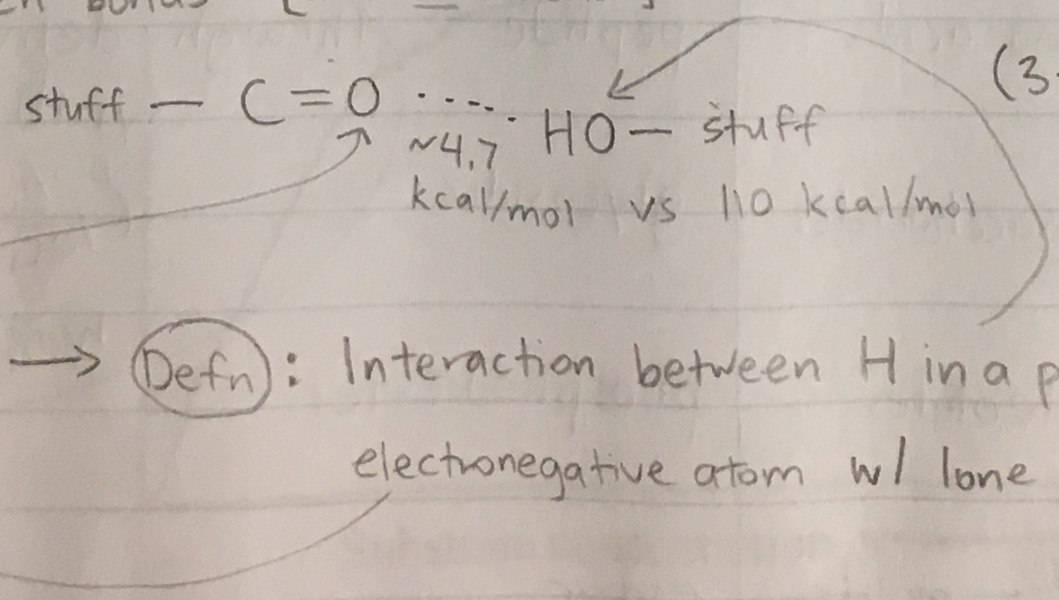
→ (Ex) Vitamin A, D, E, K fat soluble [more nonpolar bonds]
Vitamin B9 water soluble

Interactions that hold molecules together, from strong to weak

- Covalent bonds (C–C, O–H, etc.) (80–110 kcal/mol)
- Electrostatic interactions ($NH_3^+ - COO^-$) (3–8 kcal/mol) [Formal charges]

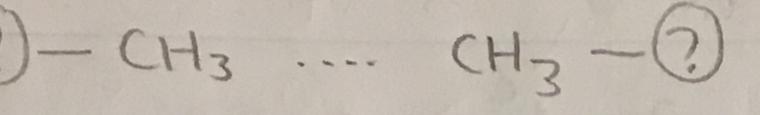
→ Note: $Na^+ - Cl^-$ ionic bonds are much stronger! >100 kcal/mol

- Hydrogen bonds [are not bonds]



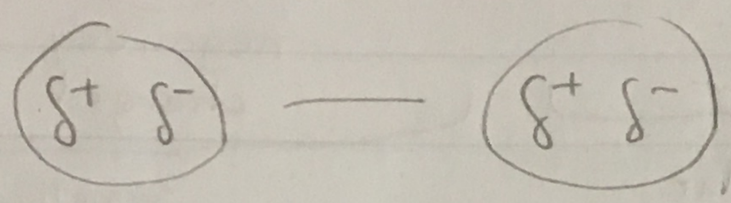
→ (Defn): Interaction between H in a polar bond and some electronegative atom w/ lone pair of (e^-).

- Hydrophobic interactions (between nonpolar substances) in the presence of water



(~1-2 kcal/mol)

- van der Waals interactions — transient polarization.



(~0.1-1 kcal/mol, distance-dependent)

There are also unfavorable interactions:

- Electrostatic repulsions
- Repulsion due to "buried charge" in the middle

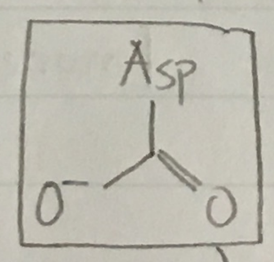
(Ex) Protein structure

→ Amino acids can be polar, nonpolar, + or − charged too.

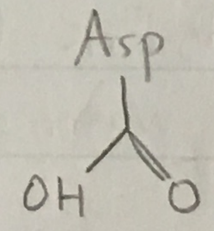
→ Some dependent on pKa → the pH where acid is 50% protonated.

$$\left(pH = pK_a - \log\left(\frac{[\text{protonated}]}{[\text{deprotonated}]}\right)\right)$$

(Ex) Aspartic acid.



or



$pK_a = 3.7$ so if pH < 3.7, more protonated and vice versa.

So at neutral pH ~7, this is dominant species.

Macromolecule	made of	Bond formed
Proteins	amino acids	peptide
Carbohydrates	monosaccharides	glycosidic
Nucleic acids	nucleotides	phosphodiester
Lipids	fatty acids	ester

through dehydration synthesis (loss of water)

7.012

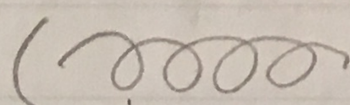
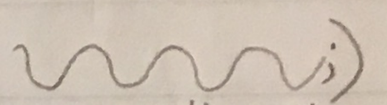
Lecture 4

9/12

Different types of structure

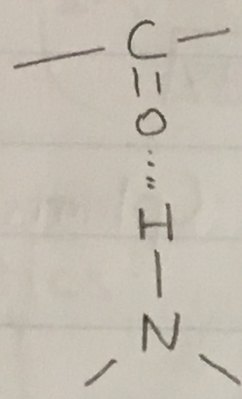
1° (primary) = sequence of amino acids (most proteins have many)

→ connected by peptide

2° (secondary) = α-helix or β-strand ( vs )

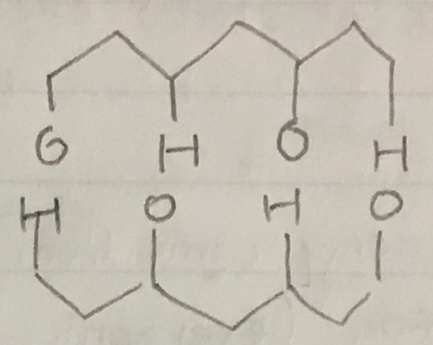
how the amino acids arrange themselves

very stable, hydrogen bonds



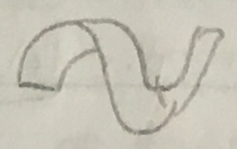
between backbone atoms

hydrogen bonds here too



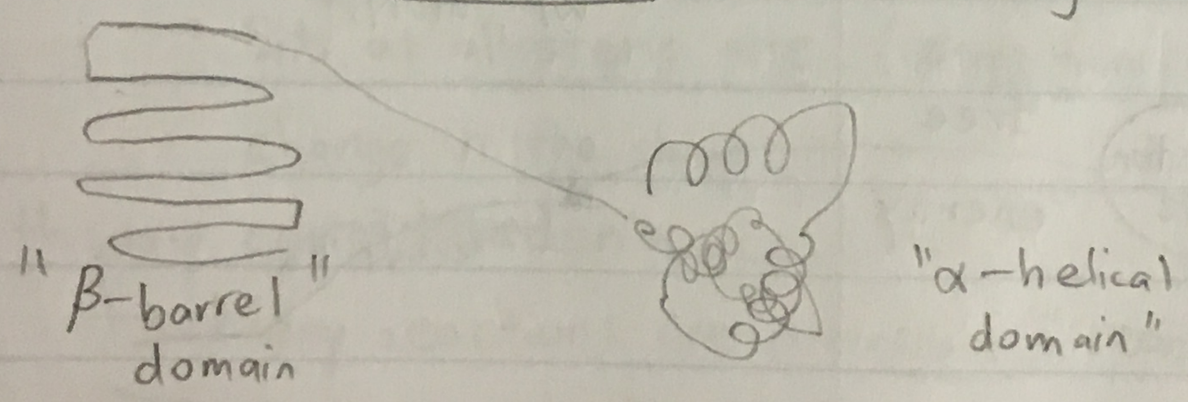
→ parallel or antiparallel

3° (tertiary) = protein fold (the 3D arrangement of the secondary structures)

→ α-helix referred to (ribbon drawing) 

→ β-strands shown as ⇄ or ⇉

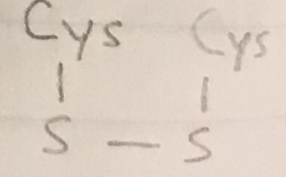
(Ex) chain w/ two domains (self-folding section)



4° (quaternary) = essentially number / arrangements of chains.

→ 1 chain = monomer, 2 chains = dimer, 3 chains = trimer,
4 chains = tetramer. "homo" = same amino acids, "hetero" = different

Factors that stabilize protein structure



covalent bonds → 1° (peptide bond), 3° (disulfide bonds), 4°

electrostatic interactions → 3° (side chains, e.g. Glu, Arg), 4°

hydrogen bonds → 2° (from backbone atoms), 3° (some backbone interaction), 4°

hydrophobic interactions → 3°, 4°

van der Waals → 3°, 4°

Why do proteins fold?

- Hydrophobic effect (entropically driven)
  → "Bury" hydrophobic, nonpolar atoms inside hydrophilic, polar atoms

Chemical reactions of life

$C_6H_{12}O_6$ [has lots of stored energy] + $6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O$ + energy
(glucose)

→ Where does energy come from? Need high availability (plants), favorable reaction (exergonic 👍 vs endergonic 👎)

[spontaneous] [nonspontaneous]

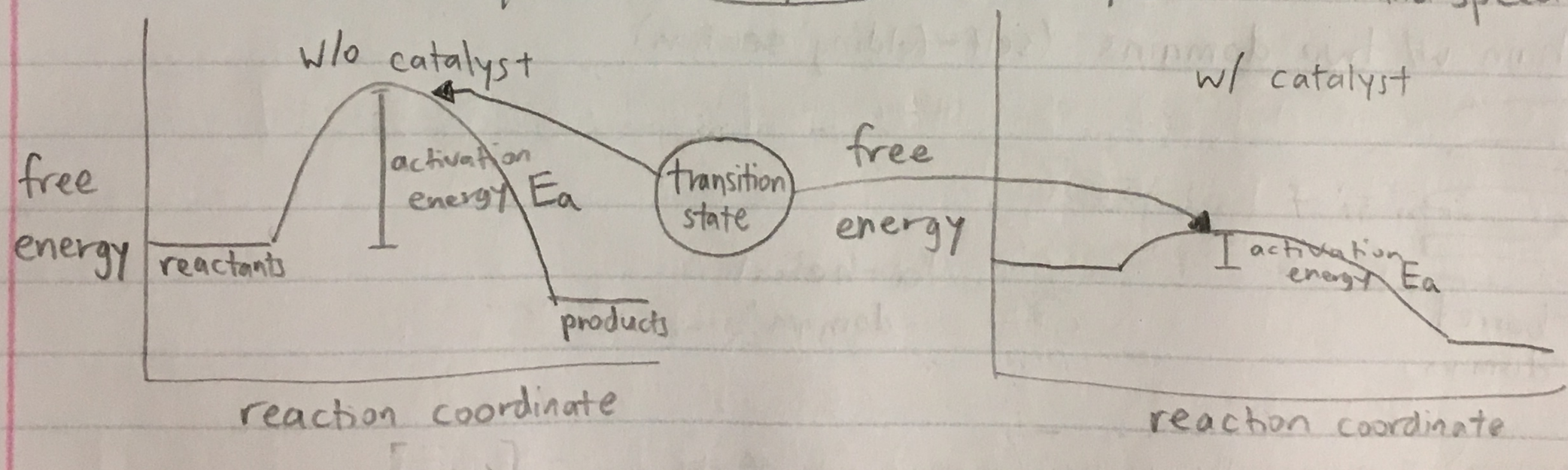
$\Delta G^\circ < 0$ $\Delta G^\circ > 0$

Standard state

→ ΔG° = difference in Gibbs free energy

→ For cellular resp., $\Delta G^\circ = -673$ kcals/mol

But rate of reaction (kinetics) is a factor too. So it is pretty slow unless you use enzymes to catalyze reaction and speed it up



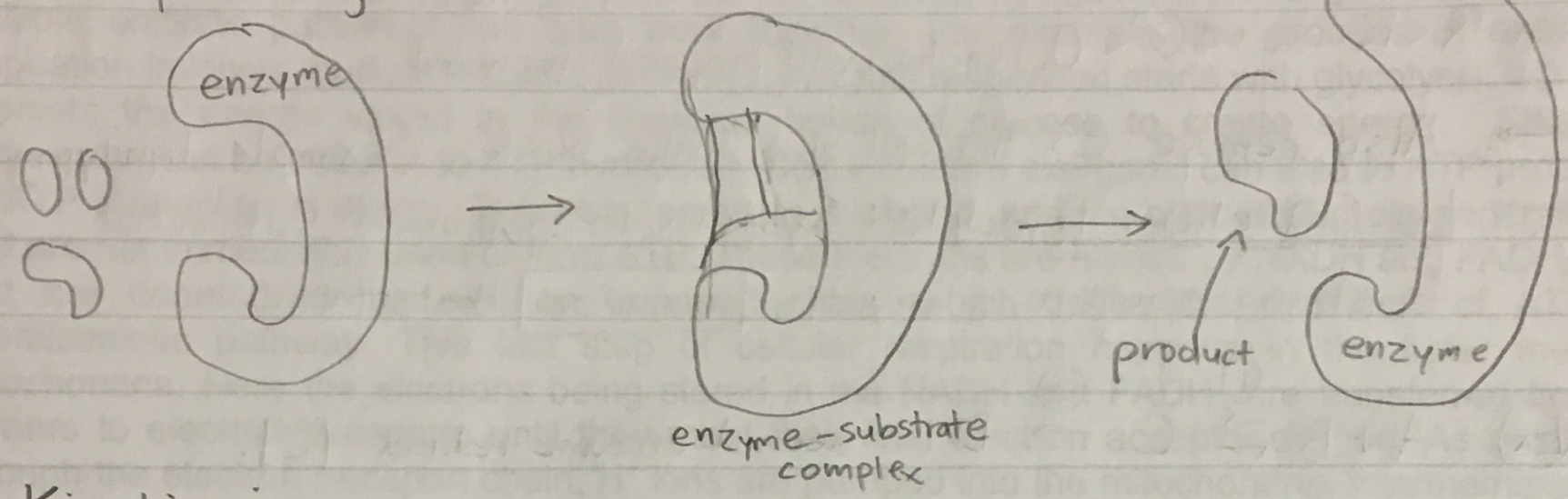
7.012

Lecture 5

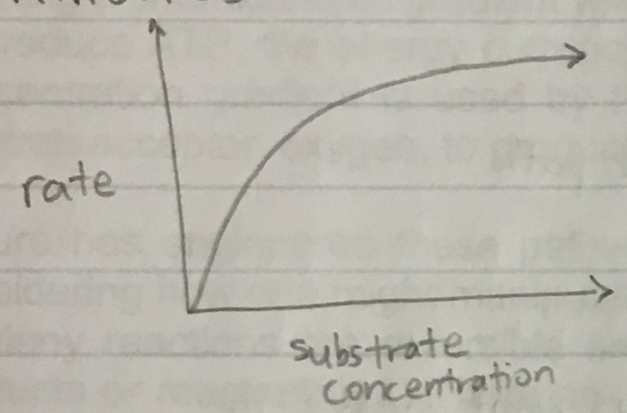
9/14

Enzymes / their pathways

- Bind reactants [= substrates] at an <u>active site</u>
- "Helpful diagram:



- Kinetics:



  - At low [A], rate increases quickly
  - At some point, active sites are full. Known as V_{max}; dependent on amt. of enzyme

- Also bind to <u>inhibitors</u>, which slows reaction
  - → Sits in active site to prevent binding, or
  - → Sits at <u>allosteric site</u> (remote from active site); causes a change in the shape.
- Pathway considerations
  - → Many reactions are <u>reversible</u>; reach equilibrium

$$A + B \rightleftharpoons C$$

↓

$$r_{forward} = r_{back}$$

→ Equilibrium constant $K = \prod_{products} [A]^{a_k} / \prod_{reactants} [B]^{b_k}$ at equilibrium.

where a_k, b_k constants = coeff. in front of chem. molecule

→ If $\Delta G^{o\prime}$ ($pH=7$) < 0 [exergonic, spontaneous], $K > 1$, and vice versa.

- Why? $\Delta G^{o} = -RT \ln K$, where R is the ideal gas constant and T is temperature in Kelvin.

→ In not-standard state,

$\Delta G = RT \ln(Q/K)$, where Q is like K but with arbitrary concentrations.

Shifting a reaction at equilibrium

- $A + B \rightleftharpoons C$; make more C by adding more A or B.

Then $Q < K$ so $\Delta G < 0$, reaction will proceed forward

→ Also remove C; this will also make $Q < K$. (Example: liquifying)

→ Le Châtelier's Principle: System in equilibrium subjected to stress reacts in a way that minimizes that stress.

(Ex) $A + B \overset{①}{\rightleftharpoons} C \overset{②}{\rightarrow} D$ second enzyme reduces [C].

Pathways — rate-limiting step

- Governs overall reaction rate based on slowest path

(Ex) $A + B \underset{①}{\overset{}{\rightleftharpoons}} C \underset{②}{\rightarrow} D \underset{H}{} \overset{③}{\rightarrow} E \overset{④}{\rightarrow} F \underset{⑤}{\overset{}{\rightleftharpoons}} G$

if this is slow, others won't evolve to get faster

⑥ (for example)

Branch points / feedback inhibition / etc.

- Might create J, a branching path for a different product
- For efficiency — no reason to have too many pathways that make the same thing
- Human body does this a lot, but sometimes want to cut off / direct flow
  - → Solution: increase rate. or use inhibitors
  - → Ibuprofen, aspirin examples of inhibitors
- Reagants (e.g. H) can also be inhibiting

(Ex) Aerobic (O_2) cellular respiration

$$\underbrace{C_6H_{12}O_6}_{\text{eat}} + \underbrace{6\ O_2}_{\substack{\text{breathe}\\ \text{in}}} \rightarrow \underbrace{6\ CO_2}_{\substack{\text{breathe}\\ \text{out}}} + \underbrace{6\ H_2O}_{\text{pee}}$$

→ Multiple pathways, each with multiple enzymes

- ↳ Some endergonic, driven by ATP cleavage; others exergonic, generate ATP
- ↳ Some of 32 ATP indirect (NADH transfers electrons; makes ATP later)

Payoff during electron transport / ATP synthesis (28)

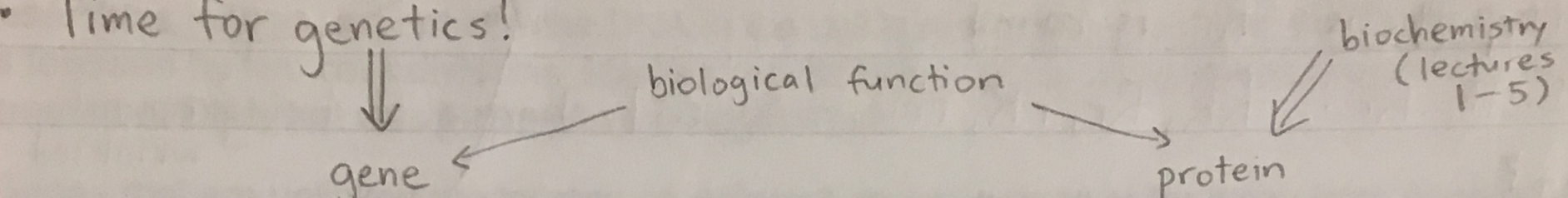
☆ enzyme that makes ATP is ATP synthase

7.012

Lecture 6

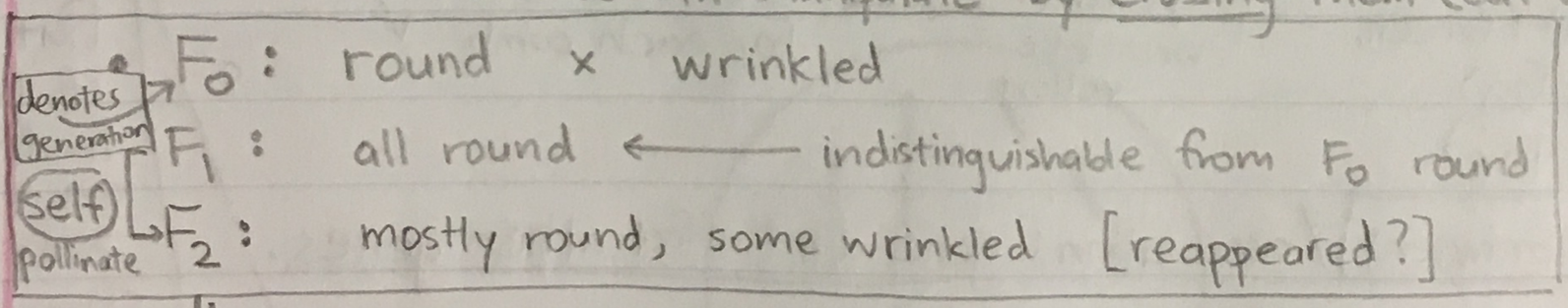
9/17

- 9/20, time TBD: chemistry review
- Time for genetics!



- History
  - → Europe going through changes
  - → Economic changes: better product
  - → How to produce better sheep? (textiles)
    - → Study of heredity (?) in Mauravia (?)
    - → Mendel — monk hired to do this stuff

- Mendel's experiments
  - Pick a plant (strain) — why peas?
    - → Can cross them; reproduce quickly; easily available in market; small
    - → Mendel found 34 varieties; kept breeding to guarantee they were pure (children look like parents).
    - → 2 years → down to 7 pure strains; the controls (1)
  - Plants can produce / self-pollinate (have both pollen / ovial "organs")
    - → But we can manipulate by crossing them (cut off pollen organs, use a brush)! (2)



★ F_1 rounds must be different from F_0. (3)

★ But Mendel counts: he gets 5474 round, 1850 wrinkled. He does this with more and more traits: unfortunately, statistics is invented by biology.

So he just says "it's 3 to 1!" (4)

- He makes a model!
  - → Roundness controlled by some factor: each plant has two. (RR vs rr) are the pure ones.

- Cross = take one from each parent at random. (Mendel's first law)
- Hypothesis: Rr is round.
- Self Rr: RR, Rr, rR, rr ⟶ 3:1 chance of being round: wrinkled

What next?

- Mendel sends a paper ⟶ peer review time!
  - ⟶ Model was made after results
  - ⟶ Need prediction power. (?)

- Why not self F_2?

RR	Rr	Rr	rr
↓	↓	↓	↓
4/4 round	3/4 round	3/4 round	0/4 round

⟶ turns out to be true!

↳ can't tell which is which; we can't see genotypes.

- Backcross / testcross: F_1 (Rr) × (rr) = 50/50 round/wrinkled.
  - ⟶ this works too!

- More complicated crosses (two-factor crosses)

F_0: Round + Green (RRGG) × Wrinkled + yellow (rrgg)

⇓

F_1: Round + Green. (RrGg)

Testcross: wrinkled + yellow ⟶ 25% chance for each comb ✓

(Mendel's second law)

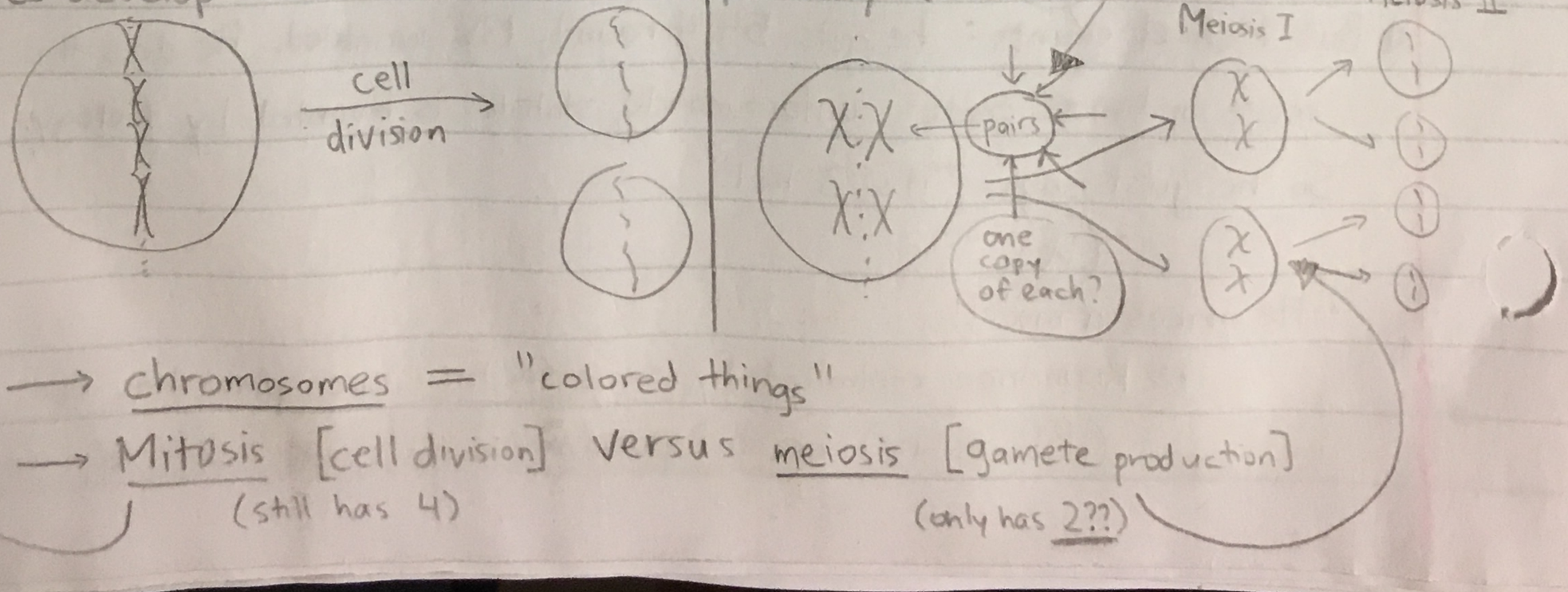
- Independent assortment / association: genes don't affect other genes.
- No one pays much attention, but...

~1900: "interesting..."

But what if two traits are on the same chromosome?

????

- Entirely different field: the cytologist
  - Dyes develop: some cells absorb dyes very well



⟶ chromosomes = "colored things"

⟶ Mitosis [cell division] versus meiosis [gamete production]

(still has 4) (only has 2??)

7.012

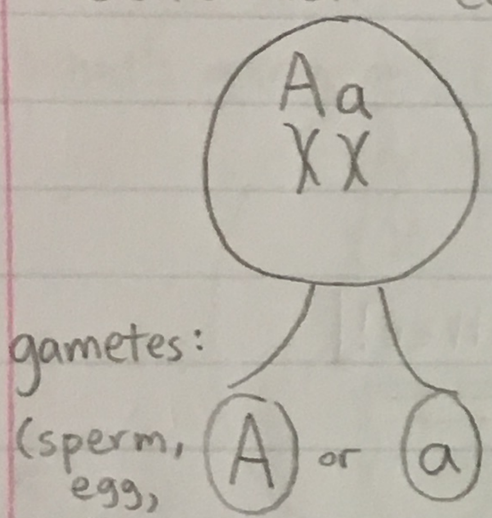
Lecture 7

9/19

- Chem review tomorrow
- Practice test posted!
- Gene = sequence of bases (genotype = RR, Rr, rr, etc.) ← allele
- Phenotype = manifestation (physically)
  ↳ it is these that are dominant or recessive.

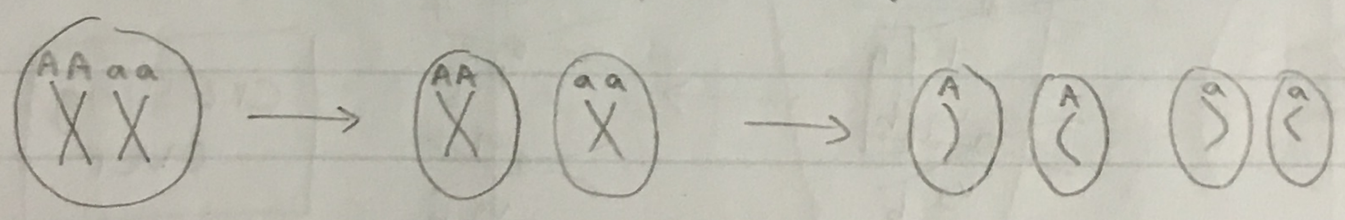
Chromosome theory

- Gene "factors" carried by chromosomes.

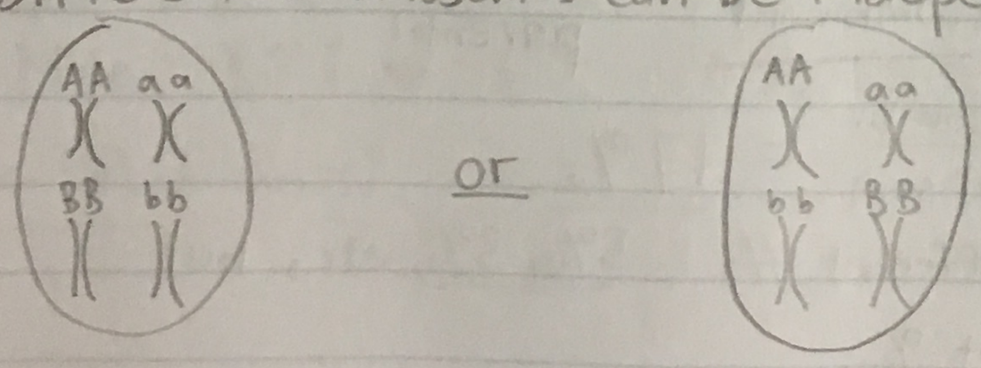


← alleles different = heterozygous
same = homozygous

→ In meiosis, first duplicate your DNA, then do two divisions.



- But... what about multiple factors carried?
  → Different chromosomes can be independent!



→ But it seems that some genes can be on the same chromosome?

- Thomas Hunt Morgan – worked with fruit flies (Drosophila melanogaster)
  dew loving black bodied fly
  → Found mutant flies

- Body: wild type vs black (mutant
- Wing: wild type vs vestigial (mutant)

F_0: body $\frac{+}{+}$ wing $\frac{+}{+}$ ♀ (wild type) × $\frac{b}{b}$ $\frac{vg}{vg}$ (all mutant) ♂

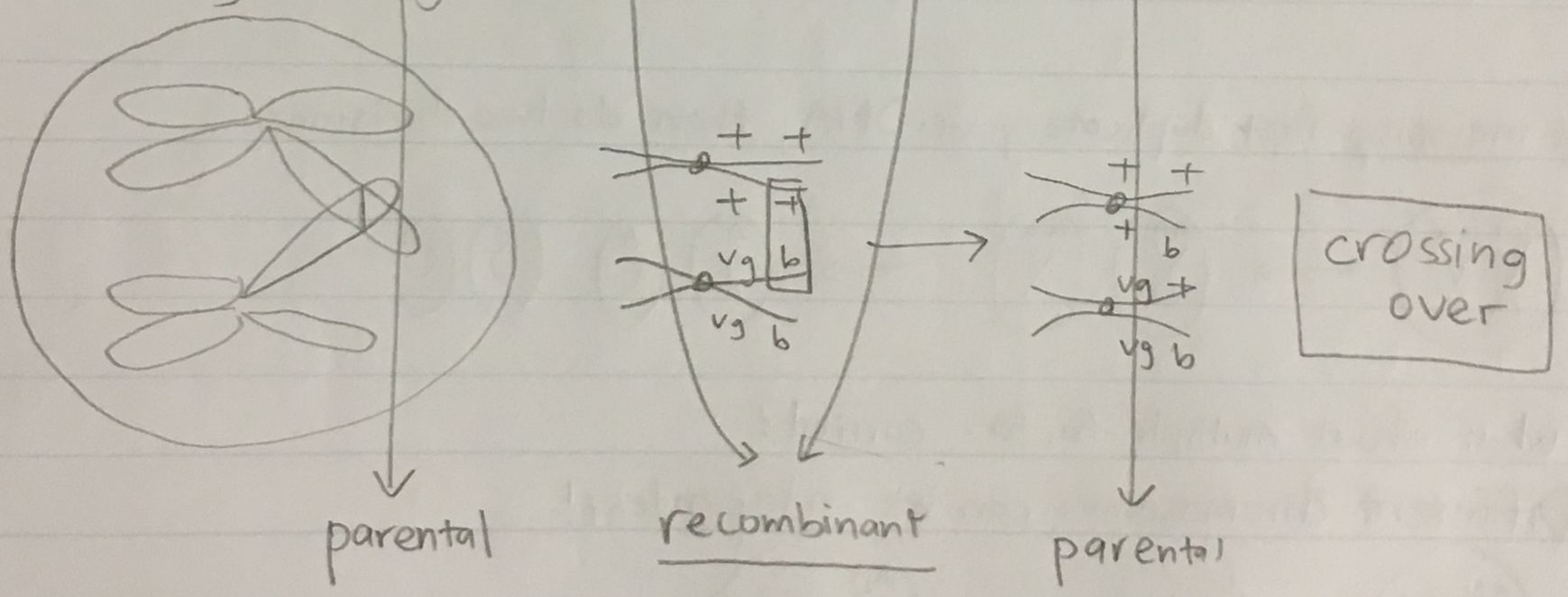
F_1: all $\boxed{\frac{+}{b}\ \frac{+}{vg}}$ (wild type phenotypes)

→ Now testcross F_1 with $\boxed{\frac{b}{b}\ \frac{vg}{vg}}$:

	$\frac{+}{b}$ $\frac{+}{vg}$	:	$\frac{+}{b}$ $\frac{vg}{vg}$	:	$\frac{b}{b}$ $\frac{+}{vg}$	:	b vg / b vg
by Mendel's second law,	1	:	1	:	1	:	1
if same chromosome:	1	:	0	:	0	:	1
actual:	965	:	206	:	185	:	944

[neither!]

Is info being exchanged? Yes!



crossing over

→ Freq of recombination = $\frac{\text{recomb.}}{\text{total sum}}$ = 17%.

→ Other genes had different #s: 5%, 3%, etc, but same pairs = consistent %.

What now?

b —— vg 17% recomb.

cn —— b 8% recomb.

⇒ also, test cn —— vg 9% recomb

→ So cn gene is between b and vg!

→ This is called linkage mapping

↳ disease traits, etc.

But one cross was weird..

F_0: white eye ♂ (recessive) × wild type ♀

→ F_1: wild type. ♀ testcross (×) wild type ♂

⇓

all females wild type
all males: 50% wild, 50% white. ???

What's going on? Look at humans!

XX
XX
⋮
XX
X × }

→ XX versus XY chromosomes!
Same w/ fruit flies.

So,

F_0:

Cross white-eye male w/ normal female:

$X^wY \times X^+X^+ \rightarrow X^wX^+, X^wX^+, YX^+, YX^+$ — F_1: wild type

F_1: $X^wX^+ \times X^+Y$

Female: X^wX^+ or X^+X^+ (all normal)
Male: X^wY or X^+Y (½ white eye!)

☆ sex-linked traits!

7.012

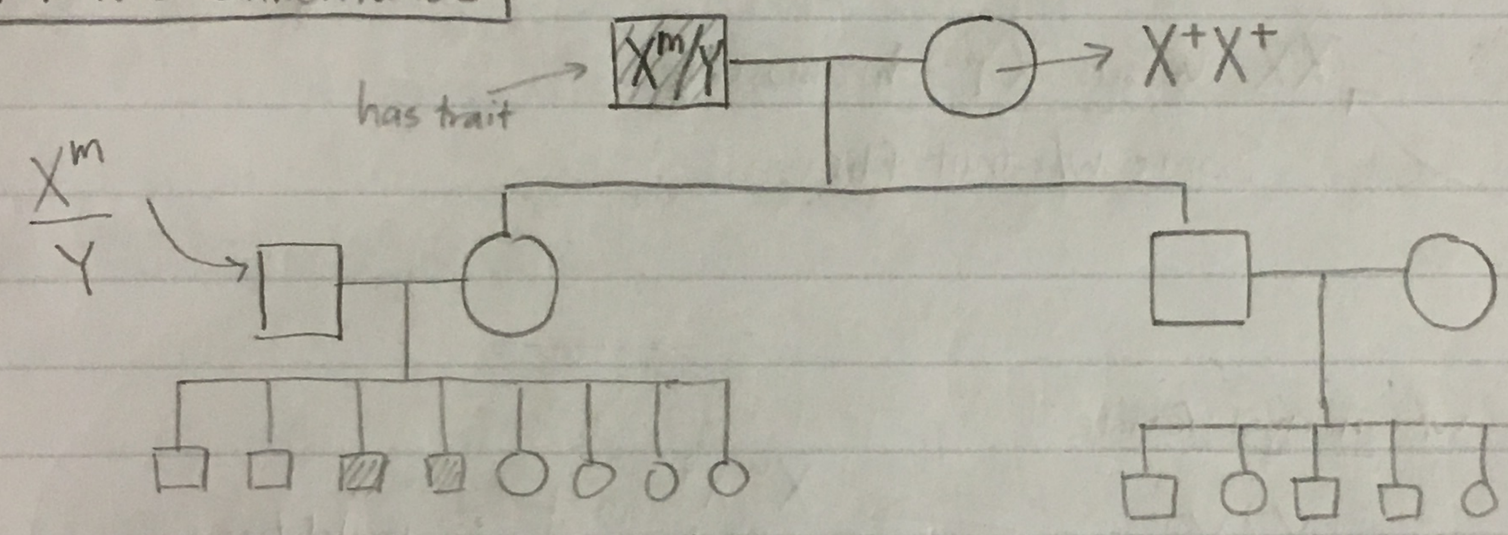
Lecture 8

9/24

- Exam on Wednesday (check whether to go to Walker or 26-100)
  - → Review Monday night
  - → Ice cream social on Wed.?
  - → Pset released Wed.

- Recombinant DNA — map genes before DNA
- Now look at humans!
  - → Charles Davenport → found trait "love of the sea"?!
    - ↳ Sea captains?... maybe there's other reasons...
  - → Genetics no longer the only reason... look at <u>pedigrees</u>!

X-linked Inheritance



→ <u>X-linked</u> recessive disorder?

→ Let X^m be mutant, X^+ "normal". Male grandfather $\frac{X^m}{Y}$, grandmother $\frac{X^+}{X^+}$

→ gen 2 must be $\frac{X^m}{X^+}$, $\frac{X^+}{Y}$.

→ Only colorblind if $\frac{X^m}{Y}$, $\frac{X^m}{X^m}$. $\frac{X^+}{X^m}$ called a <u>carrier</u>.

Rules: ① More common in <u>males</u> than <u>females</u> (p vs p^2)

→ ex: colorblind female has colorblind father

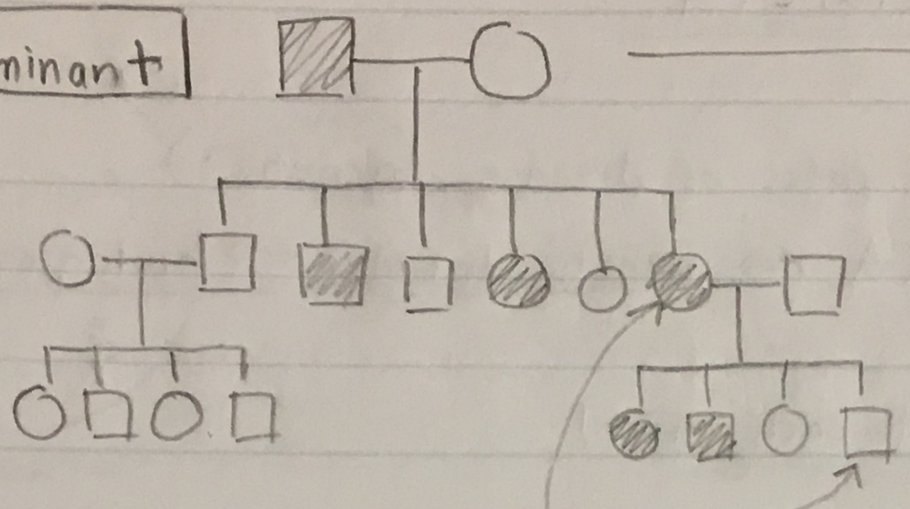
② Carrier ♀ → 50% affected sons

etc.

Fact: ~8% males colorblind. Look at <u>population genetics</u>

→ Square this to get ~0.64% of females colorblind

Autosomal Dominant



$\frac{m}{+} \times \frac{+}{+}$

Rules:
1. Never skips generations (if affected, then a parent has it)
2. ~Equal male/female
3. Transmit ~1/2 of the time.

→ Not X-linked: male here has affected mother

→ Equally affects all independent of sex

→ (Ex) Huntington's disease.

Look at population genetics: p + → p^2 $\frac{+}{+}$

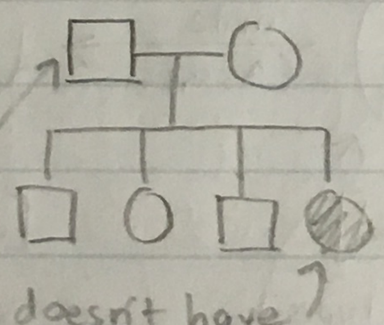
$1-p$ m → $2p(1-p)$ $\frac{+}{m}$

$(1-p)^2$ $\frac{m}{m}$

↳ $(p^2, 2pq, q^2) \approx (1, \sim 2q, \sim 0)$

↳ 1/500 has disease ⇒ ~1/1000 chromosomes carries it

Autosomal Recessive



Rules:
1. Equal male/female
2. Usually parents unaffected
3. 1/4 transmit iff carrier parents

→ Not X-linked: father doesn't have

→ Not autosomal dominant: skipped a generation

But cystic fibrosis → ~1/3?

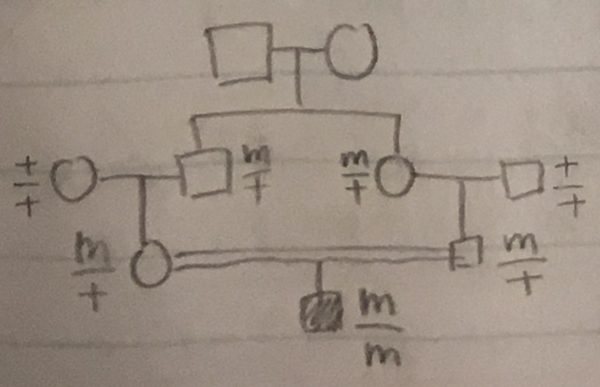
↳ Sometimes never have kids w/ cystic fibrosis! You're only counting those biased already w/ cystic fibrosis.

☆ Human genetics is different b/c you observe

Incomplete penetrance

- Only show trait ~70% of the time (e.g. breast cancer in women)

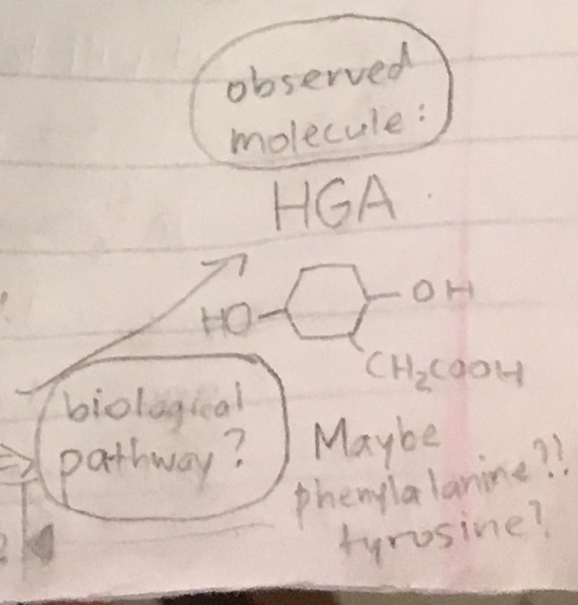
One last case: **what about the biochem?**



"black urine" — alkaptonuria (when washing in highly basic soap)

☆ Majority from first-cousin marriages!

↳ easier to homozygose(?) w/ inbreeding!



Real question: Connecting biochem/genetics

7.012

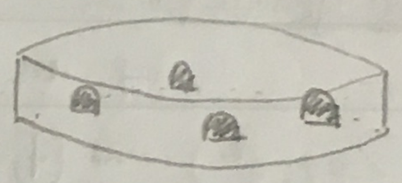
Lecture 9

Recall 1908 — inbreeding = higher rates of disease (genetic?)
↳ Also noticed excess HGA ⇒ genetic disorder ↔ biological pathway

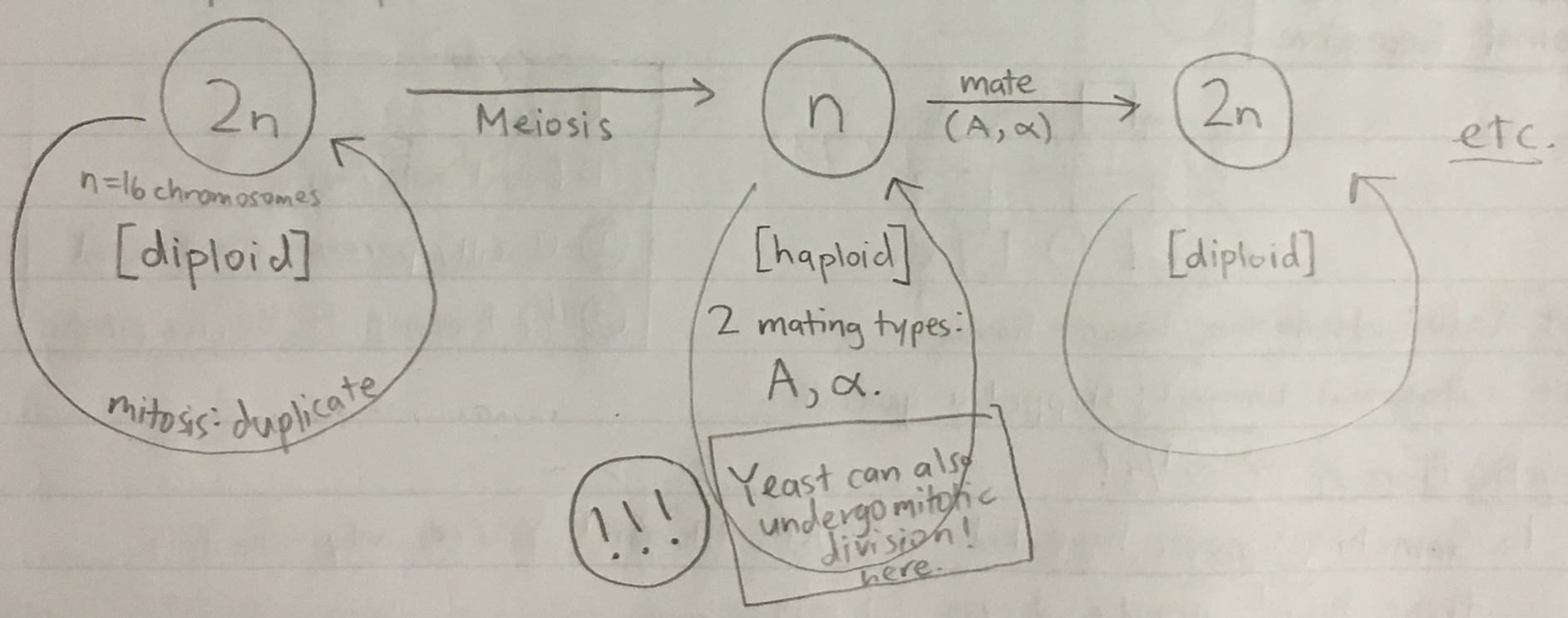
Biochem & genetics started to talk to each other!

Biochemical genetics

- Tried fruit flies
- Instead, used different organism similar (?) to yeast
  → Single-cell eukaryotic fungus
  → To grow it, use a petri dish
  ↳ Mitotic division: colony from yeast each cell.

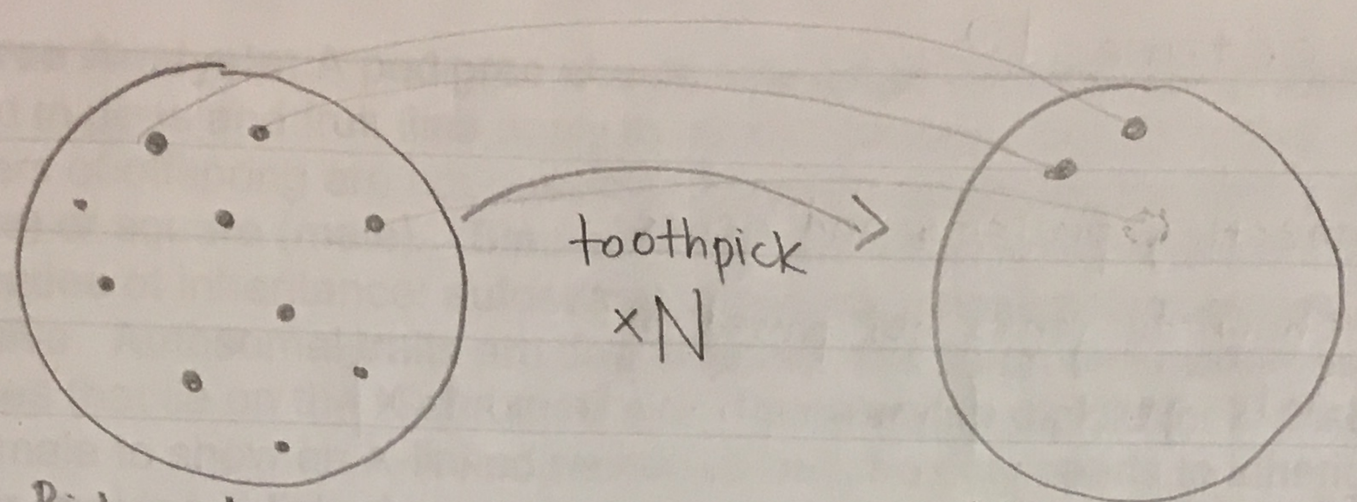


- Life cycle of yeast:

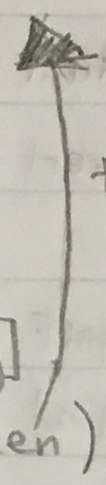


- There are some growth requirements, though... it needs a carbon source (e.g. sugar), + N, P_i salts. Nothing complicated.
  → Yeast doesn't need that much. But if you give it a rich medium, it will take those in, too.
  → regulated biosynthetic enzymes

How do you find mutants?
→ e.g. find yeast that can't make Arg amino acid?



- **Prototrophs** = can grow on minimal media.
- **Auxotrophs** = need something else (e.g. auxotrophs for Arg)



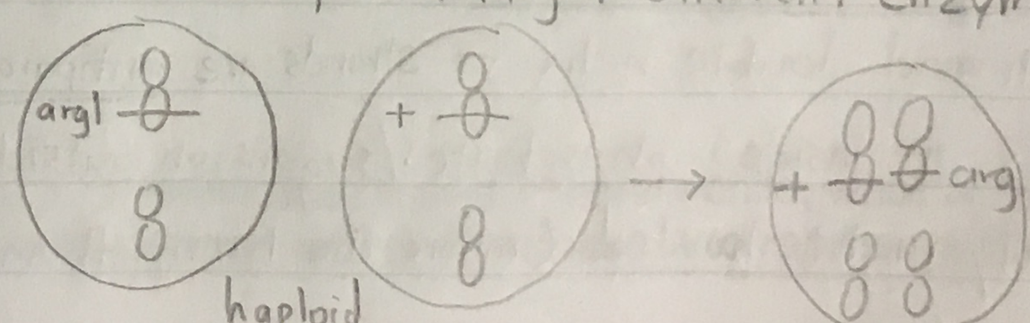
Faster ways? Better transfer instrument?

- Use a piece of velvet! Transfers in the same locations [replica plating]
- Better to have mutants that gain an ability (selection vs screen)
  → The one that grows is (e.g.) antibiotic-resistant
- Diploids or **haploids**? Pick the one w/ only 1 chromosome copy
  → if mutation is recessive, much easier to find those!
  ↳ most mutations likely recessive since 1 working copy is generally enough.

How to characterize these auxotrophs?

- Start with arg1, arg2, ...
  → Probably genes that affect this ability will target different enzymes
- 3 main tests:

→ **Test of recessivity / dominance**
→ If diploid grows, it's recessive. Else, dominant.

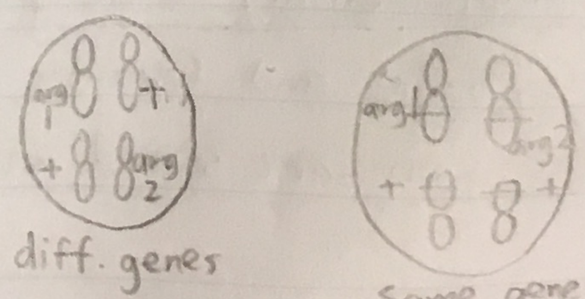


→ **Test of complementation**
→ Recombination says nearby, but not same gene...

	a1	a2	a3	+
a1	X			✓
a2		X		✓
a3			X	✓
+	✓	✓	✓	✓

→ Instead, cross $\frac{arg1}{+}$ with $\frac{arg2}{+}$!
→ Make a table. If (a_i, a_j) = ✓, diff. genes
= X, same genes

Sort into genes!



→ **Test of epistasis** $\alpha \rightarrow \beta \rightarrow \gamma \rightarrow$ Arg (We're given some pathway)

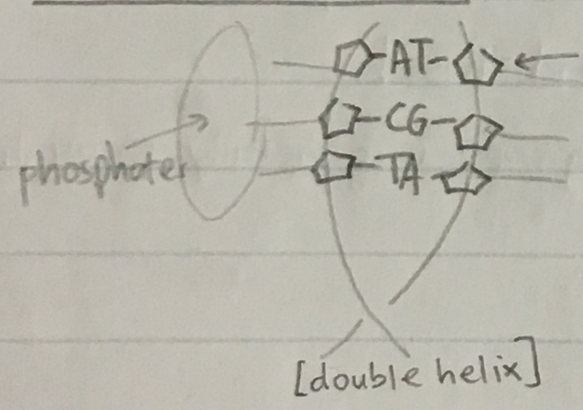
Ask: what if we just give α? Or β? Or γ? **Feed different steps!**

7.012
Lecture 10

Molecular biology — connecting proteins and genes
- → DNA replication (how do genes get passed on?)
- → How to read / translate genes [transcription / translation]
- "The central dogma" ....

Review
- Individual nucleotides connected (3′ → 5′) o w/ phosphodiester linkage
- Different bases

History
- Chargaff rules: A and T = same # bases; same with C and G
- Rosalind Franklin: X-ray diffraction
  - → helical structure of DNA, other measurements
- → Watson and Crick's model — sort of used data w/o permission...



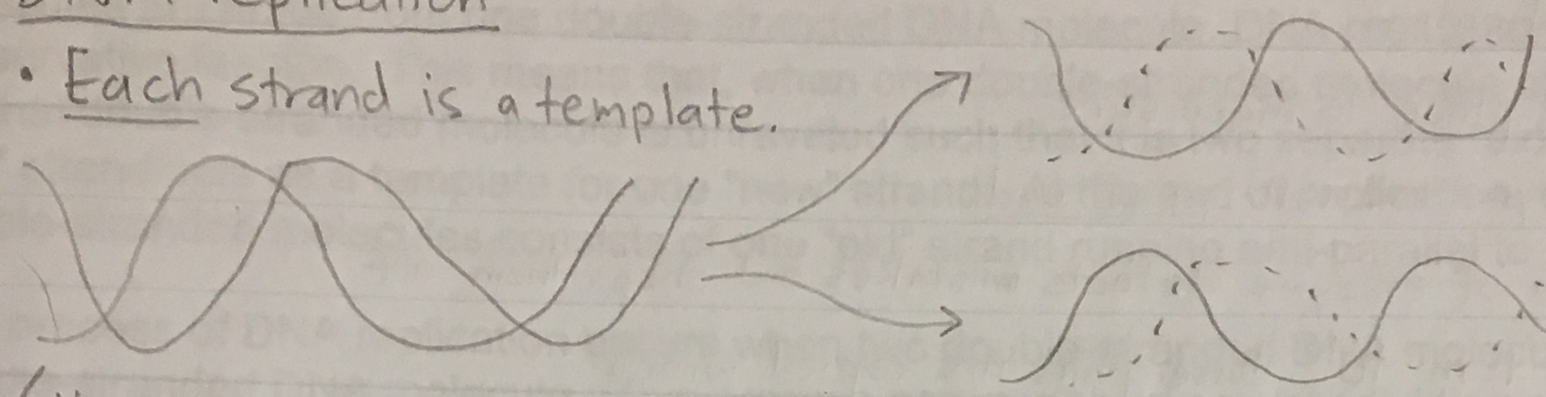
[Model is consistent w/ data.]

Structure of DNA – features
- 2 strand double helix — strands are antiparallel
- Bases on inside; phosphates/sugars on outside
- Usually right-handed (mimic the turn, thumb pointed up)
- Major/minor grooves
- Complementary base pairing — Adenine, Thymine form 2 hydrogen bonds, Cytosine and Guanine form 3 (at standard pH)
- Bases on same strand will stack on each other (van der Waals)

Why is DNA structure essential for function?
- Complementary base pairing = copying mechanism
- DNA can encode lots of info with long chains
- Adaptation — can change bases w/ mutation.

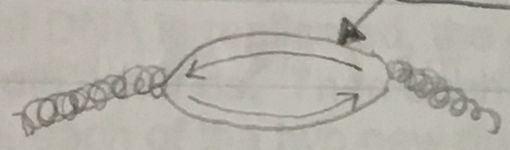
DNA replication

- Each strand is a template.



("semi-conservative" replication. Conservative = preserve original molecule)

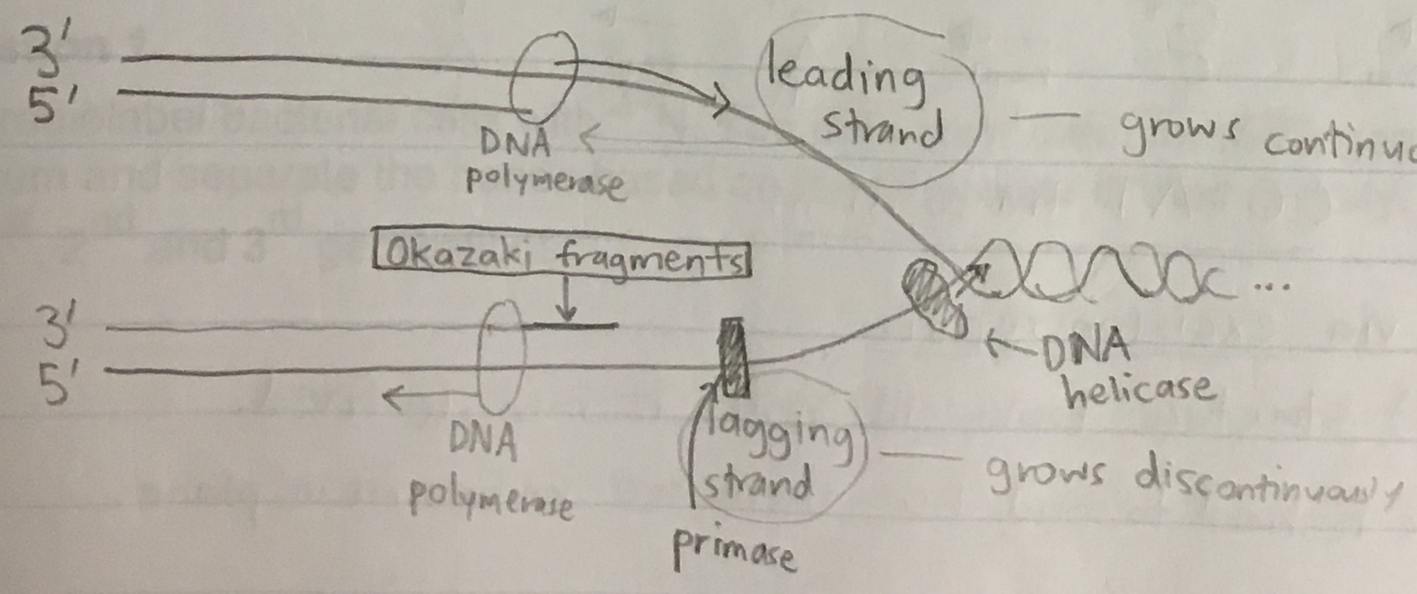
- Start at origin of replication (ori)
  - → DNA unwinded there; this results in a replication fork
  - → Enzyme [DNA helicase] does the job



- Use a primer (a short starter strand) created by [primase]
  - → Complementary to template
  - → Provides 3' OH group so [DNA polymerase] can start.
- Polymerization reaction catalyzed by this ↑
  - → dNTPs added to 3' (new bases added, 5' → 3' direction) for new strand
    - deoxy
    - A, C, G, T
    - triphosphate

★ travel 3' to 5' wrt template

- Pyrophosphate (Ⓟ—Ⓟ) is lost; this cleavage of phosphate groups is favorable
- How to deal w/ opposite directions?



- ~~Lagging strand~~ needs lots of primers — it grows a bit at a time
- [DNA ligase] links those fragments together.

7.012

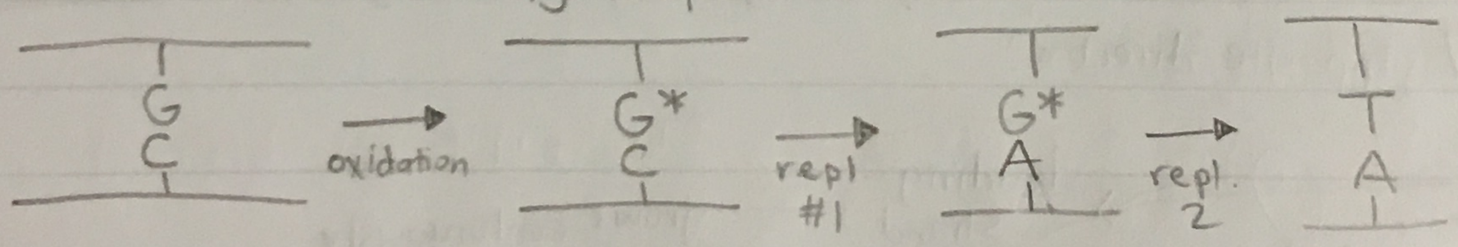
Lecture 11

10/3

- "Fidelity" of DNA replication?
  - Mechanisms in place to make sure.
    - Most DNA polymerase will proofread
      - ↳ Region of enzyme detects mistakes and chops them off
      - → Only 1 in 10^{10} base pairs are copied wrong
    - Others aren't; they're error-prone DNA polymerases

    use: (e.g.) directed evolution
      - ↳ essentially, transform/mutate enzymes & see if it works better?
      - → Nobel Prize in Chemistry awarded today
  - Mismatch repair — repair enzymes scan DNA to look for / fix mistakes.

- How about DNA repair (due to damage?)
  - Bases (A, C, G, T) can be chemically unstable
  - Oxidation [loss of electrons vs actual oxygen atoms?]
    - Oxidative damage = accumulated damage
  - (e.g.) Guanine → 8-oxo-G
  - Double-bonded O means 8-oxo-G can bond to A (?)
  - ROS - reactive oxygen species



this is a CG:AT transversion.

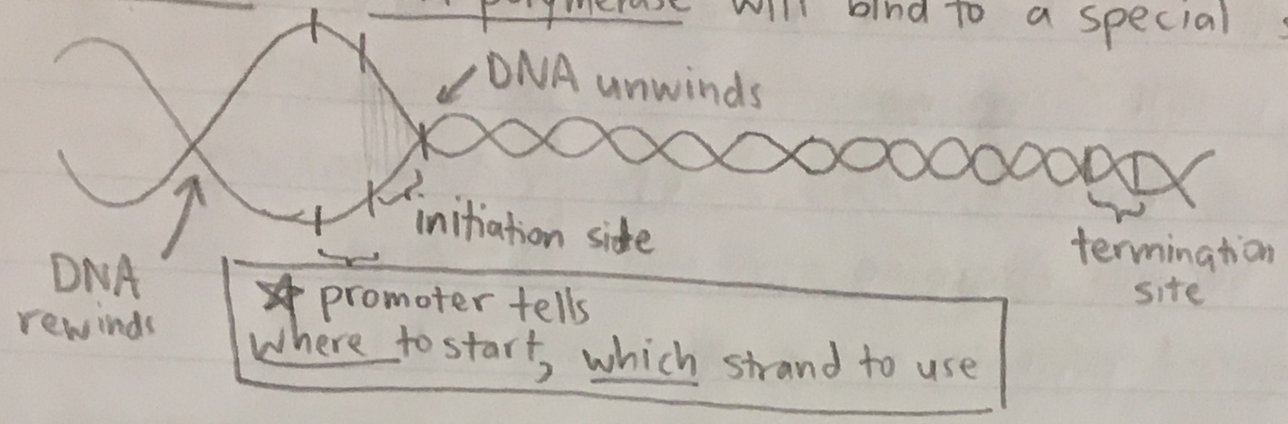
  - Can be repaired via excision repair.
    - → If DNA structure looks different, it can be fixed.
    - ↳ Damage is hard to find, but there are mechanisms in place.

- DNA → Protein sequence?
  - Transcription sends DNA → RNA (DNA copied to complementary seq.)
  - Translation sends RNA → protein

↓
messenger RNA
(mRNA)

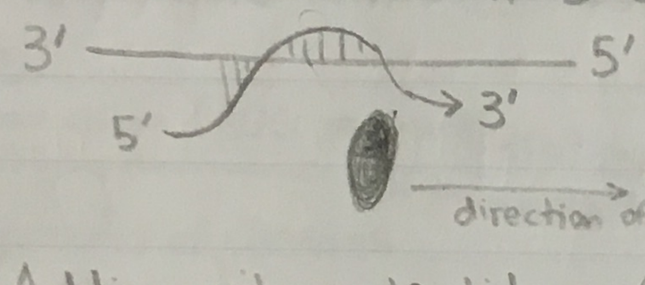
Transcription:

- Initiation: RNA polymerase will bind to a special sequence, the promoter



  - Polymerase binds to promoter w/ help from transcription factors
  - Only ~10 base pairs unwinded

- Elongation: RNA polymerase reads the template from 3′ to 5′
  - Primers are not needed — first base added (nucleotide) is 5′ end of mRNA.
    - ↳ Next ones added at 3′ end of the mRNA



  - Adding ribonucleotides (NTPs) = ATP, CTP, GTP, UTP
    - Uracil
  - 2 phosphates cleaved (just like last time), process becomes favorable.
  - No proofreading here; mRNA has a short lifespan.
  - DNA being rewinded during this as well.

- Termination:
  - RNA polymerase reaches termination site
    - ↳ transcript & polymerase released. mRNA now comp. to DNA template strand
      - → nontemplate strand has same sequence = coding strand.

(e.g.)			
	coding strand	5′	ATG
	template strand	3′	TAC
	mRNA	5′	AUG

→ T → U, otherwise same.

Genetic code — specifies the amino acids used to build protein.

☆ codon = sequence of 3 bases. Each one = amino acid.

↳ ~20 standard bases, $4^2 < 20 < 4^3$. Most amino acids > 1 codon.

- Also special start / stop codons.
- Importantly, no ambiguity, almost universal among all organisms.
  - → Can make proteins in E. coli instead of humans! But viruses...

7.012

Lecture 12

10/5

Single-base mutations (of the genetic code)

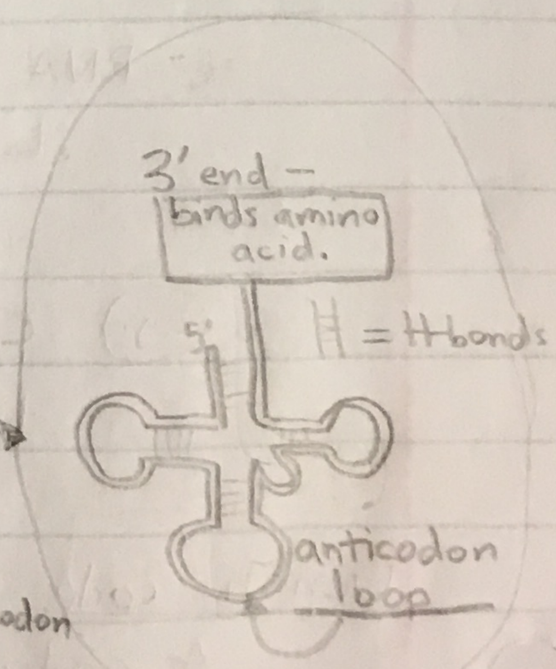
- Silent mutation — change of base does not change amino acid
  - (e.g.) UGU → UGC is still Cysteine. → forms disulfides
- Missense (?) — amino acid is changed (pretty different)
  - (e.g.) UGU → CGU becomes Arginine → positive charge
- Nonsense mutation — truncated protein, since an amino acid is changed to a stop codon.
  - (e.g.) UGU → UGA becomes stop.
- Frameshift mutation — a base is added or deleted, so amino acid gets changed drastically.
  - (e.g.) AUG UGU GAU ··· → AUG GUG UGA U ···
    Met Cys Asp Met Val stop (?)

Translation of the genetic code [RNA → protein]

- We need:
  - → message (mRNA) ①
  - → translator (transfer or tRNA) ②
  - → machinery for translation (the ribosome) ③

  ① ② ③ → 3 types of RNA
  - Ribosome contains ribosomal or rRNA, as well as protein.
- 3 kinds of RNA all made of ribonucleotides, but have different jobs and (3D) structures.

- tRNA — ~70-90 ribonucleotides
  - → structured (sort of like an L) →
  - → H-bonds stabilize it!
  - → represented sometimes w/ cloverleaf representation →



- ★ fun fact: RNA has lots of modifications
  - → Anticodon bases — important for matching/complementing codon
    61 codons, but cells don't have exact #.
    (Depends on organism)
- Wobble position — some times tRNAs can recognize multiple codons.
- Naming: $tRNA_{ala}$ = responsible for ala amino acid, etc.

Charging of tRNA

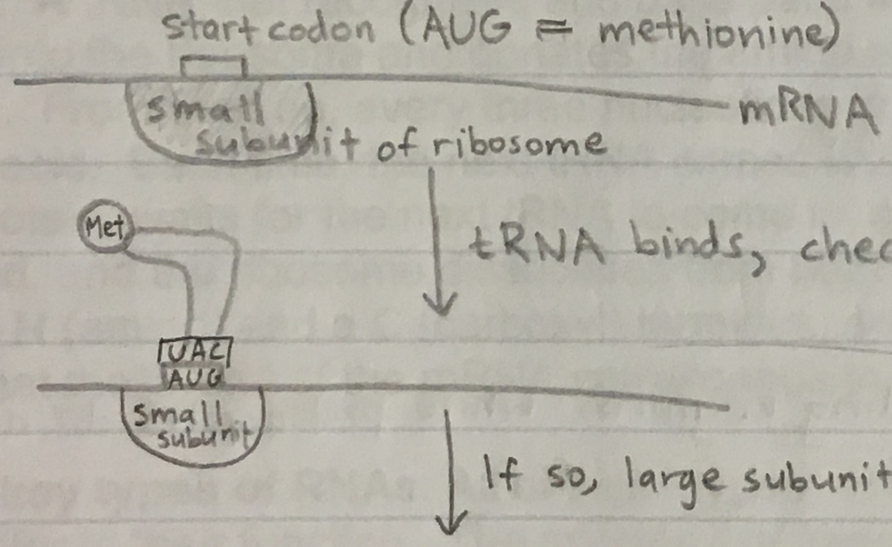
- Enzymes put amino acid onto tRNA (at the 3' end?)
- Form high-energy bond between tRNA, amino acid using ATP
  → Enzymes are specific: $tRNA_{ala}$ activated by ala only, etc.

name: aminoacyl (rel. to amino acids) – tRNA synthetase (req. ATP)

Translation in the ribosome

- Cells have lots of ribosomes that can be reused.
- 2 subunits in the ribosome: large and small.
  ↳ prokaryotes have ~1500–3000 ribonucleotides, ~20–30 proteins
  → More for eukaryotes. Fact: ribosomes are different for pro/eukaryotes, so (e.g.) antibiotics can target bacterial ribosomes only.

- **Initiation**



- **Elongation**
  → Next tRNA binds at the A site
  → break bond b/w tRNA and current amino acid in P site
    ↳ Methionine at N-terminus of chain.
  → Form a peptide bond w/ next amino acid; first tRNA ejected at E site
    ↳ Growing amino acid chain at P site.
  → mRNA read 5' → 3'
  → rRNA is the catalyst for forming these peptide bonds.

repeat until stop

- **Termination**
  → Stop codon → protein release factor enters A site and cleaves polypeptide off.

3 sites: A site – amino acid
P site – peptide
E site – exit

7.012

Lecture 13

10/10

- P-set due Friday
- Exam next Wednesday (10/17)
  → Genetics & molecular biology

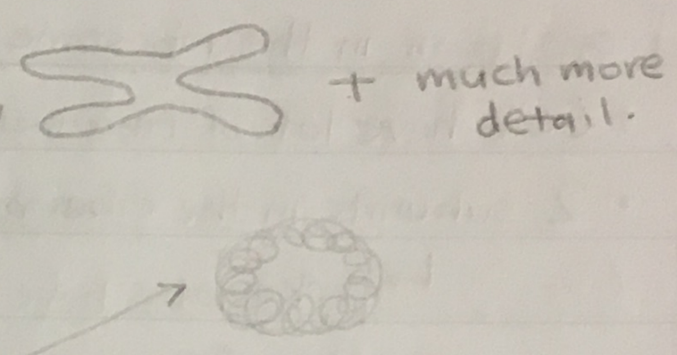
☆ Main idea: lots of variations on the main idea (central dogma)

Differences in genetic material

- Eukaryotic (e.g. human) vs prokaryotic (e.g. bacterial) cells!

Where is the material stored?
→ Nucleus vs cytoplasm (everywhere)

What is the structure?
→ Double-stranded for both (dsDNA)
→ Linear (wrapped around proteins) vs circular + much more detail.
↓
this is highly organized and very complicated



Telomeres?
→ Repeat sequences at ends of chromosomes
→ Only in eukaryotic cells

How compact is the genome?
→ Eukaryotes have introns = non-coding regions, while prokaryotes do not

More about telomeres

- "End parts"
- Repeat sequences at ends of chromosomes: $(TTAGGG)_n$ ← thousands
  → With each replication, chromosomes get shorter
  → Serves as buffer region so essential genes are OK
  → Nothing useful in genetic code
- Enzyme telomerase adds buffer region
  → Active in cancer cell (rapidly dividing)

Transcriptional differences

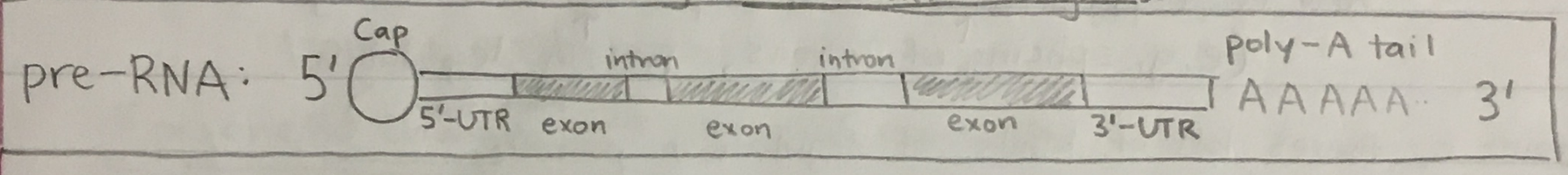
- Location: again occurs in nucleus vs cytoplasm
- mRNA is processed in eukaryotic cells (e.g. introns), not in prokaryotic cells

→ precursor mRNA (the initial transcript) in eukaryotes are not the same as mature RNA.
→ the introns are spliced out

RNA splicing — removal of introns; connection of remaining exons

- Intron — non-coding region between exons, removed by splicing
- Exon — coding regions, remain in transcript
  → both protein-dependent.

☆ Not all non-coding regions are introns!
5′ and 3′ UTR → noncoding untranslated regions



→ catalyst of splicing = spliceosome (-osome = big)
↳ made of many proteins / RNA

- Why splice?
  → If introns were translated, might get different / nonfunctional protein.
  → Also, you can get multiple proteins from the same transcript!
    ↳ alternative splicing; genetic variation
  → So eukaryotic cells = more complicated

Other modifications (in eukaryotes)
- 5′ cap: facilitates mRNA binding to ribosome
  ↳ protects from degradation
- Polyalenine tail for stability of mRNA
  ↳ (?) important for export of mature mRNA (nucleus → cytoplasm). for translation

Translational differences
→ In both types, takes place in ribosome.
  → But in eukaryotes, attached to ER = endoplasmic reticulum. (Free in prokaryotes)
  → Bigger in eukaryotes (good for antibiotics)
  → Ribosome binds to 5′ cap in eukaryotes vs specific seq. in 5′ UTR during initiation
    → 5′ UTR longer in eukaryotes.

→ Eukaryotes — typically monocistronic vs polycistronic in prokaryotes
  1 mRNA, 1 polypeptide
  more starts/stops = multiple polypeptides!
→ More efficient in some scenarios. (e.g. same pathway)

7.012

Lecture 14

10/12

- Review session for exam: Monday (10/15) 7–8:30 in 10-250
- Cryo-EM tour: 10/17 5–7 pm (?)

Gene regulation — related to previous material

↳ regulation of gene transcription / expression

What does it do?

Regulate the amount of active protein!

genes DNA → mRNA → Protein

transcriptional regulation – ↑ vs translational regulation ↑

Also exists post-transcriptional or post-translational regulation

(e.g. splicing of introns, protein degradation / inhibition.)

→ The point is, there's lots of kinds of regulatory processes.

So... why regulate the amount of protein in the cell?

- In eukaryotes, cell differentiation.
  - → Don't need liver cells for other uses, etc.
  - → Prokaryotes are single-cell, don't need to worry about this
- Prokaryotes respond to a change in environment
  - → e.g. availability of food

Kinds of proteins

- Inducible — made only in certain conditions
- Constitutive — always made
  - → e.g. "housekeeping"

Main methods for gene regulation

1. "Reversible modifications of DNA / histone proteins

(this is what DNA is wrapped around)

→ Both prokaryotes & eukaryotes

(e.g.) Methylation of DNA (usually on cytosine), occurs in both types

- DNA methylase puts on a methyl group, DNA demethylase does opposite.
- Heavily methylated genes are NOT expressed (silent)

What are some functions of this?

→ Protection from foreign DNA — cell will methylate it.
  ↳ So it doesn't transcribe data from a virus.
  ↳ Causes some frustration w/ introducing enzymes though...

→ In prokaryotes, identify strand to repair in mismatch repair.
  → Old strand methylated first = template strand
    ↳ Fixes mismatch problem! Only in prokaryotes though

→ But in eukaryotes, we have cell differentiation!
  → Heavily methylated → genes not needed. (eg skin cells have others methylated)

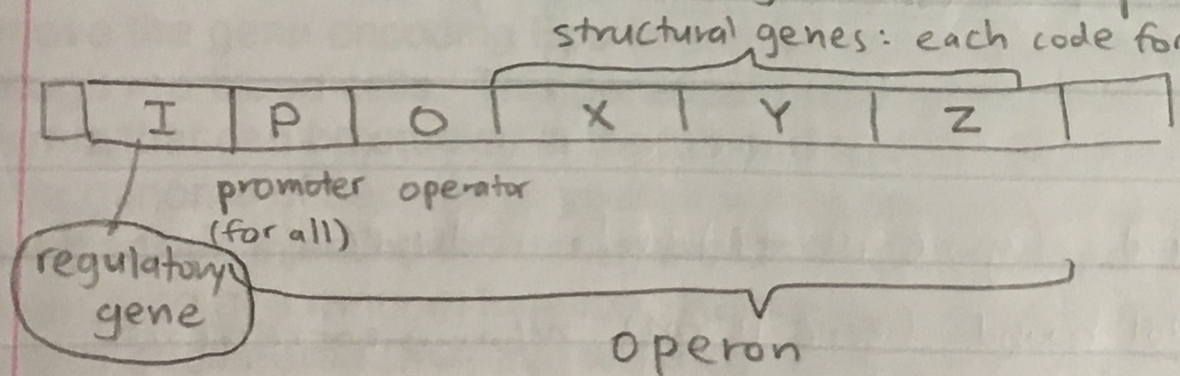
☆ telomerase active for cancer cells — any way to fix regarding methylation?

Epigenetic change → area of research.
does not change in DNA sequence

2. Using DNA binding regulatory proteins

Turn on/off regulatory pathways!

- Organization of genes in prokaryotes → whole pathways can be turned on/off.
- Operon = a cluster of genes w/ one promoter.



→ Promoter is where RNA polymerase binds.
→ Operator is a short sequence b/w promoter & structural genes.
This is where regulatory proteins bind! Sometimes binds to block / allow RNA polymerase from/to binding.

What are types of regulation?

- Negative — a repressor blocks RNA polymerase.
  ↳ if repressor is bound, no transcription.
- Positive — no/weak affinity for RNA polymerase unless in presence of activator.
  ↳ Activator stimulates transcription.

Example of a LAC operon: inducible system for metabolizing lactose

↳ Only need this if lactose is present!

- If lactose absent, operon is repressed by an active repressor.
- When lactose is present, there is induction of gene expression
  → Allolactose, an inducer, binds to the repressor and inactivates it.
  → Now transcription does occur!
- Also, if there is low glucose & lactose is present, further activated transcription w/ activator protein.
  → cAMP increases, which binds to CRP, which helps RNA polymerase bind more efficiently.

- Some repressors only bind w/ a co-repressor. bound to them
  → (e.g.) NikR repressor for nickel uptake in E. Coli only binds in presence of Ni^{2+} ions
  (i.e. avoids too much nickel in cell)

☆ Most co-repressors and inducers work by a conformational change / change in oligomeric state

- Inducible systems — repressor is bound (gene is off) w/o inducer.
- Repressible system — repressor is not bound (gene is on) w/o a co-represior

How to find out how much protein was made?

- Gel electrophoresis! Compare bands to see how much there is w/ or w/o. an inducer.

7.012
Lecture 15

10/15

Exam review: tonight 7-8:30pm

- Molecular biology connected genes ↔ proteins
- But even after all of this, we still can't read any genes?

Why is this hard?

- You can purify enzymes easily - different sizes, other properties
- But DNA all looks the same?!

→ No physical property — only what it encodes.

- Today is the start of Recombinant DNA: how to find genes?

Overview

- 3×10^9 bases in a human genome
- 3×10^3 bases in a gene
- 1 base in a mutation

} ? how?

→ Cloning! tl;dr:

- Cut DNA at specific points
- Paste to vectors: DNA propagating in cell
- Transform into cell
- Select for cells w/ this DNA.

Cutting DNA

- Let's say you want to cut out "GAATTC"

→ Use bacteria! 20 min generation time vs 20 years.

↳ Bacteria have enzymes to cut these out! EcoRI can do this!

- Why would bacteria even have this?

→ Protects it from foreign DNA (e.g. viruses) as an immune system

G|AATTC
CTTAA|G

"reverse palindrome"

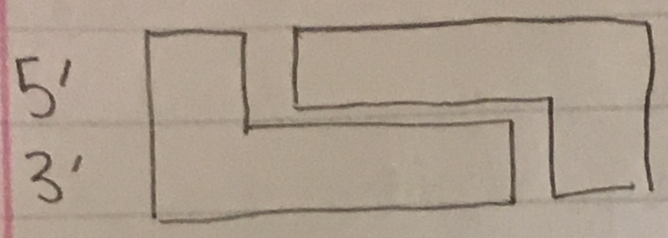
→ What if you cut your own DNA?

↳ Every ~4000 bases, GAATTC should occur. Seems hard to avoid it!

→ But E. Coli can tell the difference between the foreign / own DNA!

☆ Enzyme puts a CH_3 methyl group

↑ restriction enzyme (restricts viruses) is the cutter,
methylation enzyme (puts CH_3) is the protector.

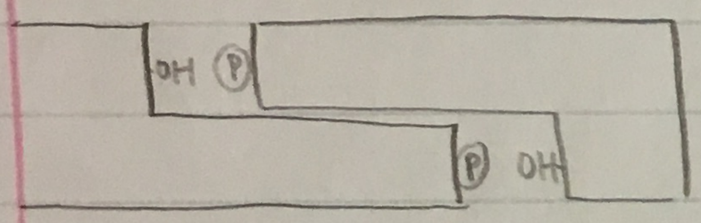


GAATTC, GGATCC, AAGCTT all cut like this

GATC / CTAG , CAG|CTG / GTC|GAC , etc... many restriction enzymes!

- How to find / obtain restriction enzymes?
  - → Used to purify by growing E.coli.
  - → Now order online! e.g. $59 for 10000 units of EcoRI
- So what happens if we put this in human DNA? It'll cut, avg. length ~4000.

Pasting DNA

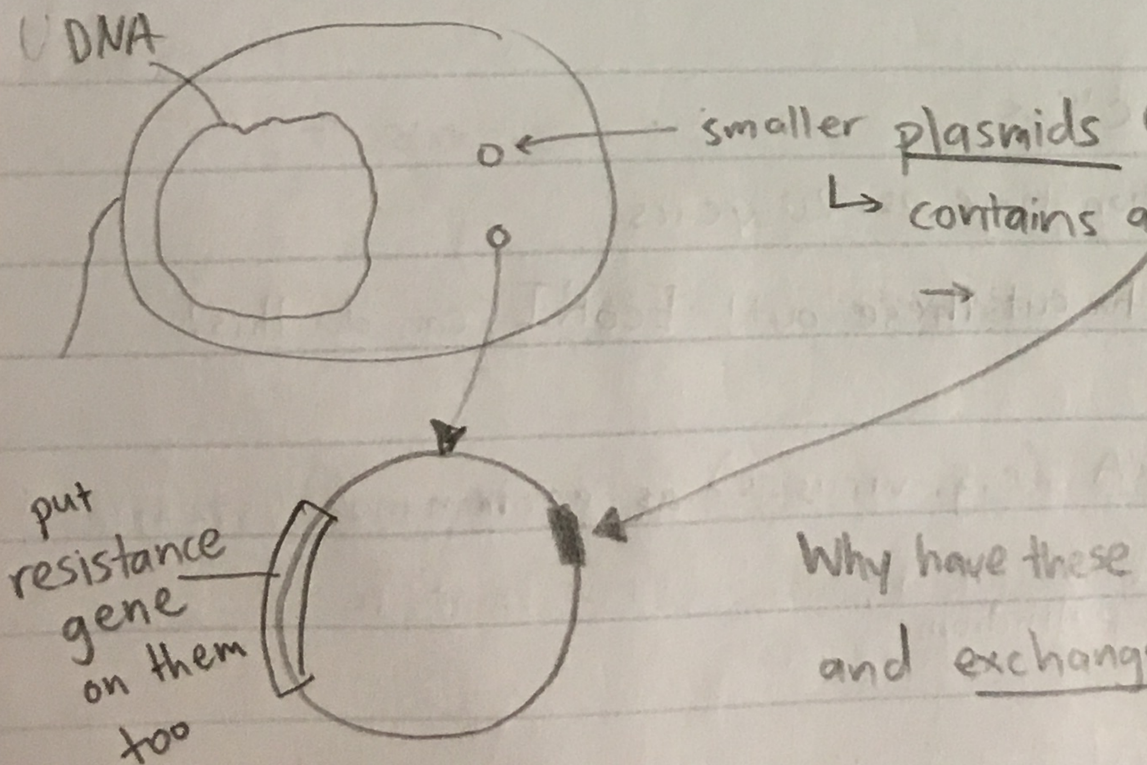


To paste, need to repair sugar-phosphate bonds.

→ that's ligase's job! which you can also buy online

→ You can even glue DNA from different organisms.

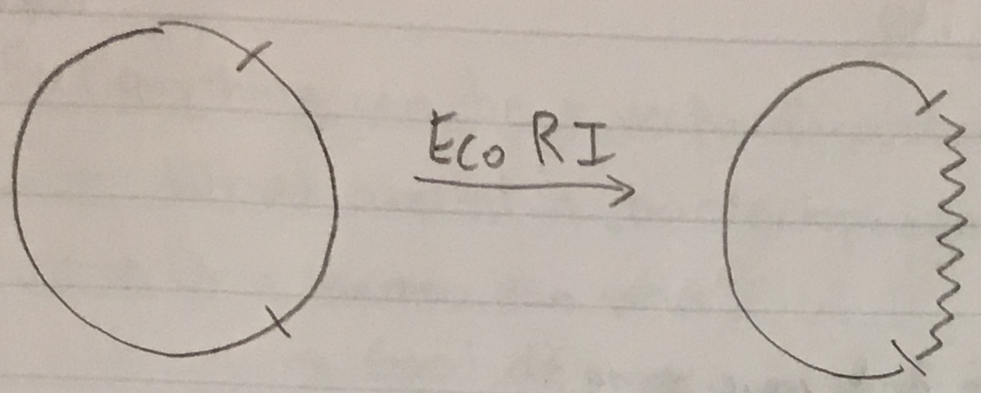
- But we want it to propagate — replicate in E. Coli!
  - ↳ How? Ask the bacteria!



Why have these circles? Transmit information and exchange these! Or do horizontal transfer

(Pen^R, Kan^R)? ≠ ↳ Defense mechanism, again. Antibiotics — penicillin secreted by fungus! So bacteria wants enzyme to degrade it.

- Let's say we use this. Cut plasmid w/ restriction enzyme.



Get a plasmid w/ human DNA!

Transform the DNA

How to get bacteria to take up circles of DNA?

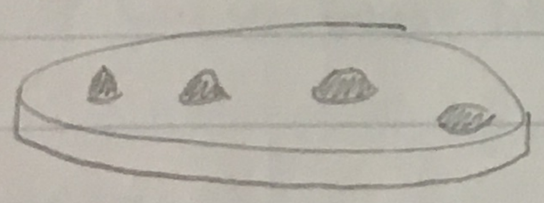
- Do a lot of possible things (e.g. electric shock), and they'll "slurp it up!" by chance.

Select for those with your plasmid

- Pour onto a petri dish; only the ones w/ the resistance gene will survive!

But how do you identify things now? (Random library)

- Many colonies, each w/ single different piece of human DNA



Purified by randomly sticking these in bacteria!

✡ This is only possible b/c DNA goes down to a single molecule & grows it back up!

- But still — how do you find the hemoglobin gene?
  → Transcribe it!
  → Needs plasmid to tell enzymes to do this, but this is possible!
- Early on, though, look at red blood cells, which have lots of RNA containing hemoglobin
  → Reverse transcriptase! Turn back into DNA & clone it
  → Much turns out to be hemoglobin.
- Doesn't work in general though.
  ↳ Try to go from random library to finding your gene

7.012

Lecture 16

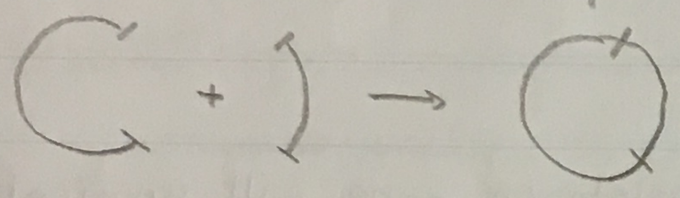
10/19

- Red Sox are in the World Series!
- Pset 4 posted...

Last time — cloning genes. Now — how to find your gene?

Steps

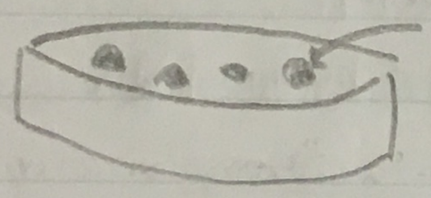
- Cut DNA w/ restriction enzymes @ specific spots
  - ↳ they're an "immune system" for single-celled organisms
  - ↳ end up w/ pieces of DNA
- To make it grow, use vectors to propagate
  - ↳ e.g. bacteria's plasmids (circular dsDNA)



use ligase to stitch back together

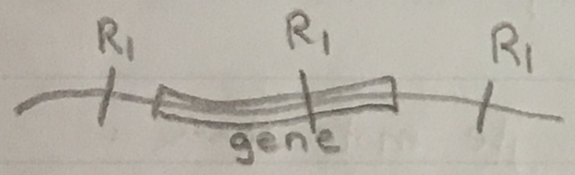
- Select for antibiotic resistance.

So now



different human genes (but purified)

What if...



the gene gets broken in half?

→ Randomly throw in methylase's only cut (on avg) some proportion
↳ Partial digestion (!)

- Alternative: "shred" the DNA randomly w/ sonic waves
  - ↳ Can't match to a plasmid though... they might stick out
  - → Polish the ends by getting rid of single strands
  - → Many variations!

Cloning vectors

- plasmid (5000 base pairs)
- What if we wanted to clone in eukaryotic cells? They have plasmids too!

- But anything can be a vector if it replicates! / propagates
  - → What about a bacteriophage?
  - → Or a mammalian virus?
    - → Good delivery system
  - → Artificial chromosomes? (harder b/c big)

Can we clone RNA? Not quite.. it's single-stranded.

↳ But you can convert it back (reverse transcriptase)

RNA ~~~~ AAAA ← poly-A tail

DNA ~~~~ TTTT ← poly-T starter (called a cDNA)

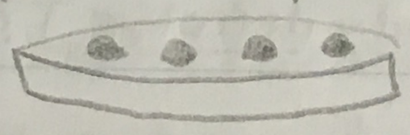
Why is this cDNA cloning good? No introns!

So now, how do we find the gene?

With yeast, we found auxotrophs for arg.

↳ labeled an arg1, arg2, etc. mutant.

- Make a library! (Cloning by complementation)
  - → Cut up yeast DNA, put it in a yeast plasmid. Transform into yeast cells!

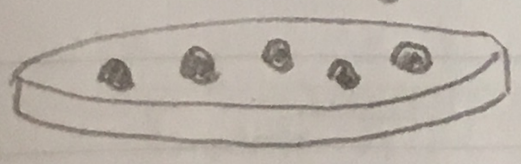


☆ But use wild type DNA, mutant yeast cells. (arg1⁻). Grow on min. media

↳ Only plasmids with arg1⁺ will survive! ↖

- What if I want to clone (human) hemoglobin? not fatal like here...
  - → Way to recognize hemoglobin. Antibodies (grow by injecting mouse)

Steps of (cloning by expression):

→  colonies of human cells (that are defective)

→ use expression plasmids so it has a promoter before the hemoglobin gene

↳ Wash antibodies to see which produce hemoglobin!

- What about diseases that (e.g.) cause brain degeneration?
  - → can't use either of the methods above ↑
  - → 3rd / different way!

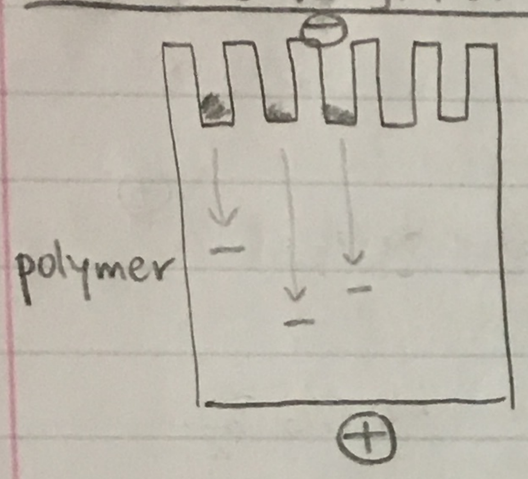
7.012
Lecture 17

Recall

- Chop up DNA, put them in plasmids
- Purify plasmids — now we have a test tube w/ DNA in plasmids
  - ↳ Restriction enzymes → we have 1 piece of DNA

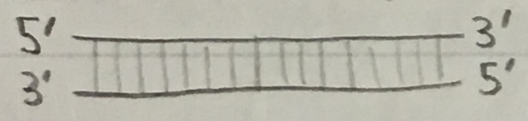
What do we want to know about it?

Measure length of DNA: gel electrophoresis



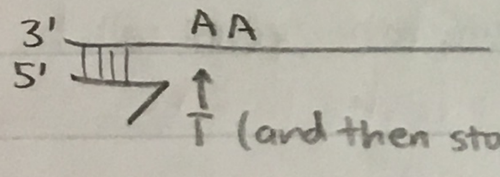
- DNA can wiggle through polymer, but only if there's a current (DNA (−) charged)
- Smaller moves faster
- Compare against known fragments.

Sequencing the base pairs



what is this sequence?

① What if I undo the double helix and give DNA polymerase A, C, G, and a defective T?



↳ this would tell us what the next base is

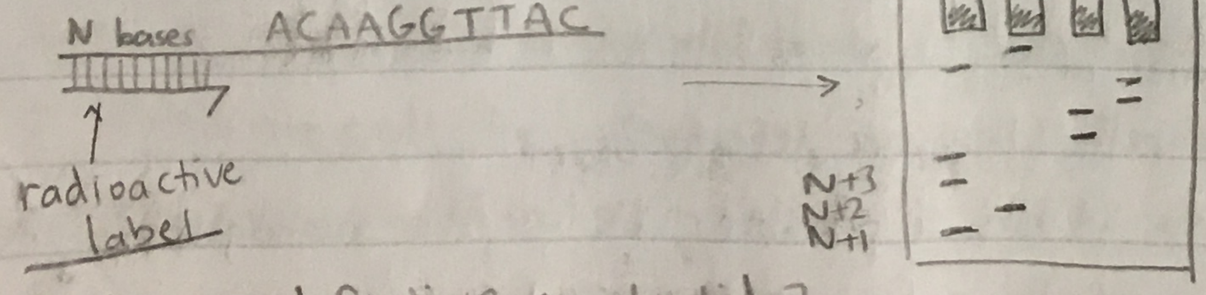
② What if I only add some defective Ts?

↳ we have many copies of the DNA, and each stop happens w/ some probability

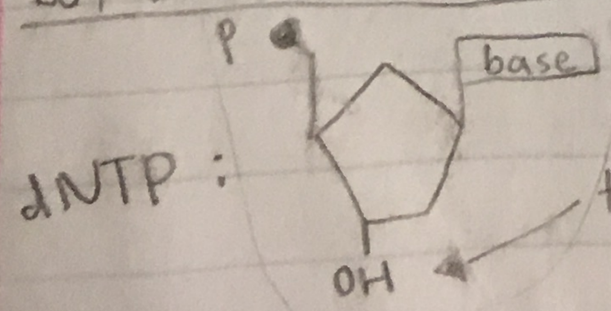
→ Now measure lengths = locations of Ts!

→ Do same w/ A, C, G.

e.g.



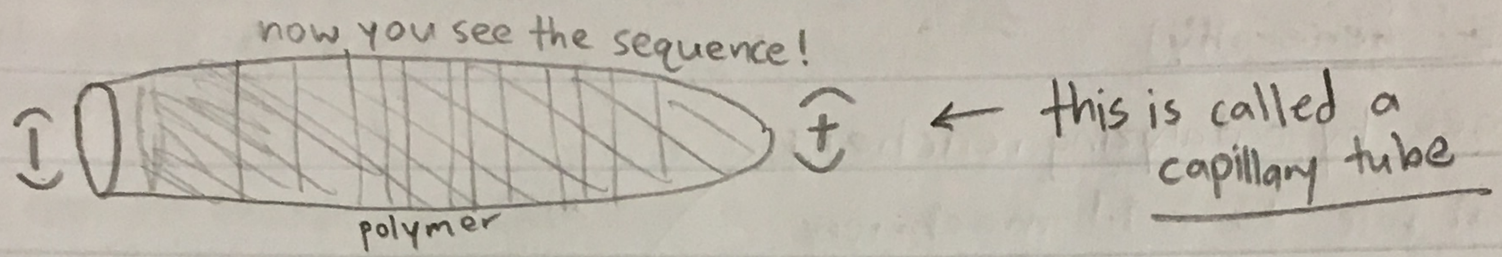
But what is a defective nucleotide?



dNTP:

this 3′ hydroxyl is important — it attaches to the next nucleotide. Remove it!

So this creates 2',3'-dideoxyribonucleotides.
Doing this is tedious though... any way to make it more efficient?
Add a fluorescent dye to defective T! maybe different color w/ A,C,G.
→ Now lanes will light up diff. colors! Now we can throw them all in together, and we can even do the reactions together.

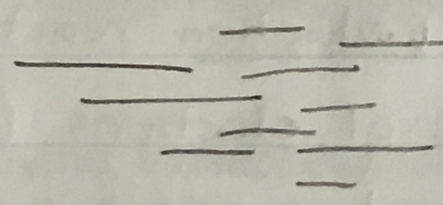


← this is called a capillary tube

Scale this up → machines can do 100000s of base pairs to read DNA.

But how do you assemble this?

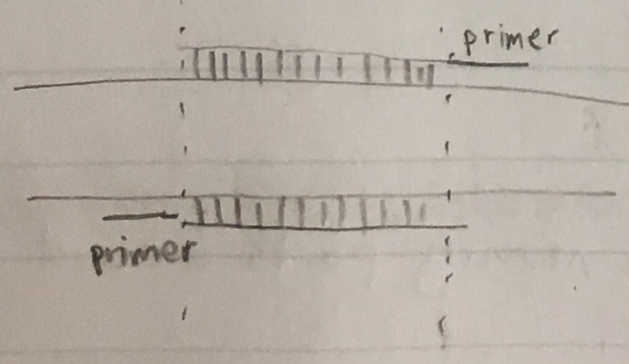
Used to do primer walking - after getting N bases, chop it off and keep going
Instead, start at random points (random subclones)
↳ We can sequence each one individually
→ Now fit the pieces together! Look at overlaps


this works for bacteria

→ BUT humans have really repetitive DNA — it's not random.
↳ handwave handwave it was hard.

So now... how to get your hemoglobin gene if I have mine?

- Throw in a copy of the hemoglobin primer?
  → But we need to amplify this with a polymerase chain reaction



2 copies of the double helix

Keep repeating this!

1 million copies of genes after 20 rounds.

7.012
Lecture 18

10/24

- Another p-set is due on Friday! Make sure you're using the corrected version.

Today we're talking more about proteins — molecular biology tools for studying structure and function

Protein Functions (generally)

→ Act as enzymes by catalyzing reactions
→ Play structural role to hold machinery
→ Transport stuff
  - e.g. hemoglobin
→ Regulation

we'll talk about this now

What are the types of enzymes?

- Oxidoreductases → catalyze oxidation and reduction reactions (usually happen together in redox reaction)
  - ↳ this means loss & gain, respectively, of electrons.
  - (e.g.) Aerobic respiration → pyruvate oxidation!

- Transferases transfer functional groups (where you have to form/break bonds)
  - ↳ CH_3, OH group, etc.
  - (e.g.) DNA methylase adds CH_3 to DNA.

- Hydrolases uses water molecules to cleave bonds
  - (e.g.) Proteases cleave peptide bonds!

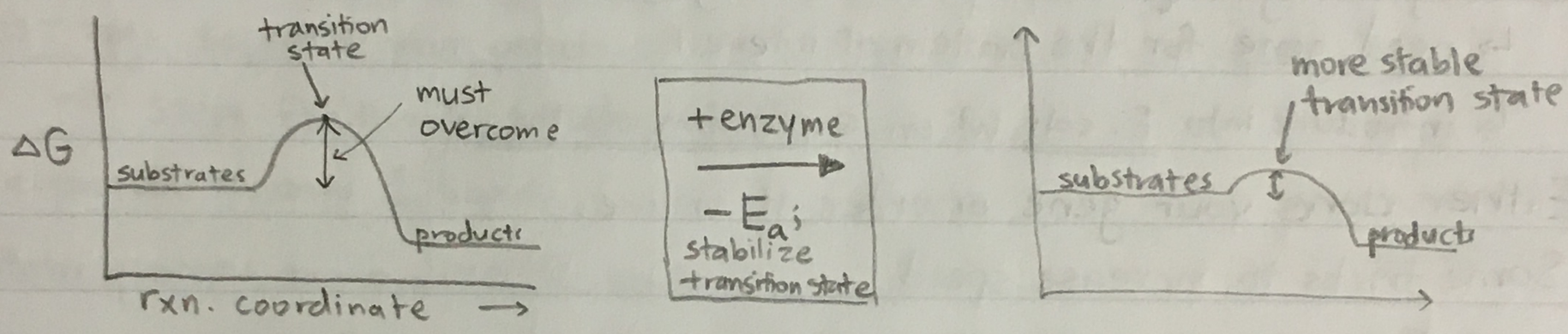
- Isomerases change structure of a molecule by moving parts around
  - (e.g.) TIM — triosephosphate isomerase

- Lyases remove something in an elimination reaction
- Ligases form bonds, but this is "expensive"
  - ↳ so it's coupled to ATP hydrolysis

When we're characterizing enzymes...

- What do we want to know?
  - → Reactions in play — what are the substrates and products?
  - → What is the enzyme mechanism — how many steps? what type?
  - ☆ You can show that mechanisms are inconsistent, but not that they are true — they are just models.
  - → Rate of reaction / enzyme kinetics; how well do molecules bind?

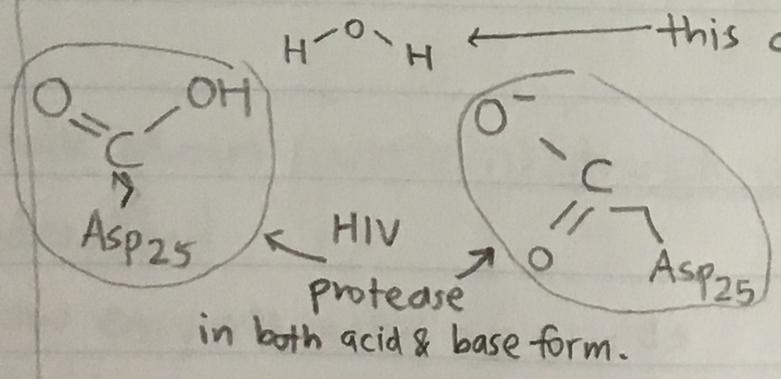
- Common enzyme tricks — a review



→ Acid-base catalysis: stabilize transition state w/ proton acceptance/donation

- Acid donates a proton (H^+), base accepts one

(e.g.) HIV protease: helps water attack carbonyl carbon



→ end result: peptide bond is broken

→ Covalent catalysis — substrate forms a transient covalent bond w/ the enzyme.

(e.g.) Chymotrypsin is a serine protease. Forms a tetrahedral intermediate as well.

→ Metal ion catalysis — metal ions can stabilize transition state by activating H_2O for hydrolysis reactions

(e.g.) zinc protease contains Zn^{2+}.

- **Why do we want to know it?**
  - Design enzyme inhibitors
    - (e.g.) HIV proteins not dangerous in immature form
      - ↳ mature form created by having protease chop up its proteins
    - → So we inhibit HIV protease [to make it not dangerous] → basically, resemble tetrahedral trans. state.
  - See if human enzymes are different from viruses / pathogens
  - Make faster enzymes.

- **How do we do it?**
  - Getting enough "quality" protein—using recombinant DNA methods.
    - ↳ need gene for the protein of interest
    - ↳ introduce into E. coli w/ an expression vector
  - Either clone your gene or order it online.
  - Some tricks to increase speed
    - → Put gene in vector after strong, inducible promoters.
    - → Modify end of gene w/ a purification tag.
      - (e.g.) His-tag → strong affinity for nickel, so now we can separate ones w/ tags from ones without.
    - → Evaluate purity w/ gel.

6 histidine residues at the end

7.012

Lecture 19

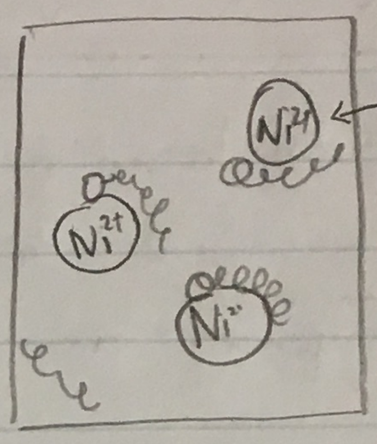
10/26

More about gels from previous lecture!

SDS-PAGE gel electrophoresis

- Separates proteins by mass
  - → confirm expression
  - → evaluate purity of protein
- Proteins not as charged as DNA
  - → SDS covers protein (after first denaturing them)
  - → Uniform charge/mas ratio
  - → So larger proteins travel slower
- His-tags separate your protein of choice
  - → steps: flowthrough, wash, elute
- Pure protein = only 1 band.

☆ Stain proteins to visualize the bands.



Site-directed mutagenesis – test pathways using introduced mutations.

- Use synthesized DNA primer w/ necessary mutation
- E. coli does the rest → now you can see enzyme mechanism w/ your necessary mutation.

Next, we'll talk about (protein) structural biology tools.

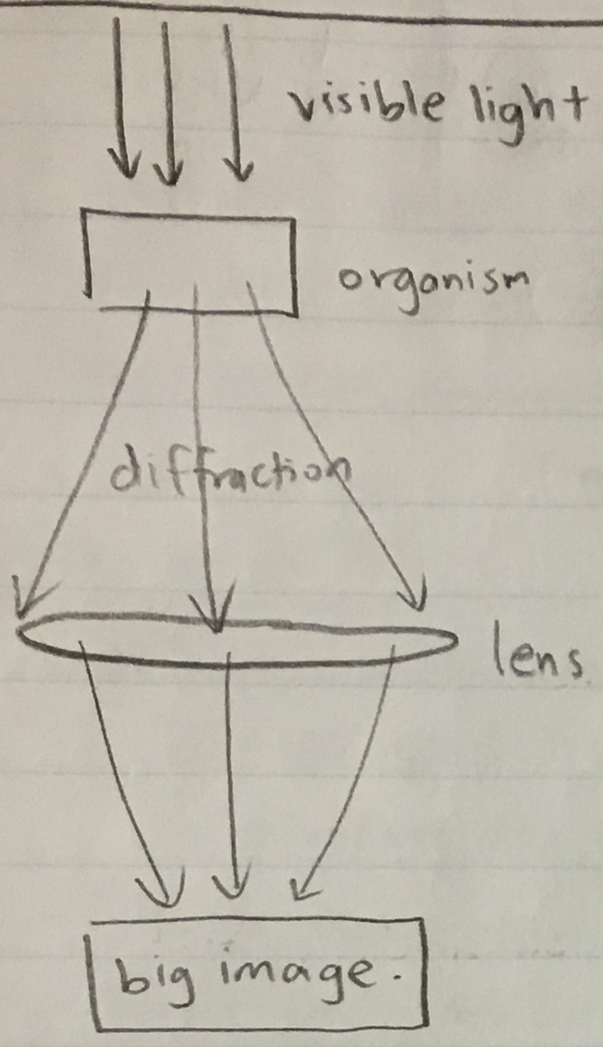
- We can understand
  - → How enzymes make molecules
  - → How to design inhibitors
  - → Basis of genetic disease
  - → Changing enzymes to make new products/chemistry.

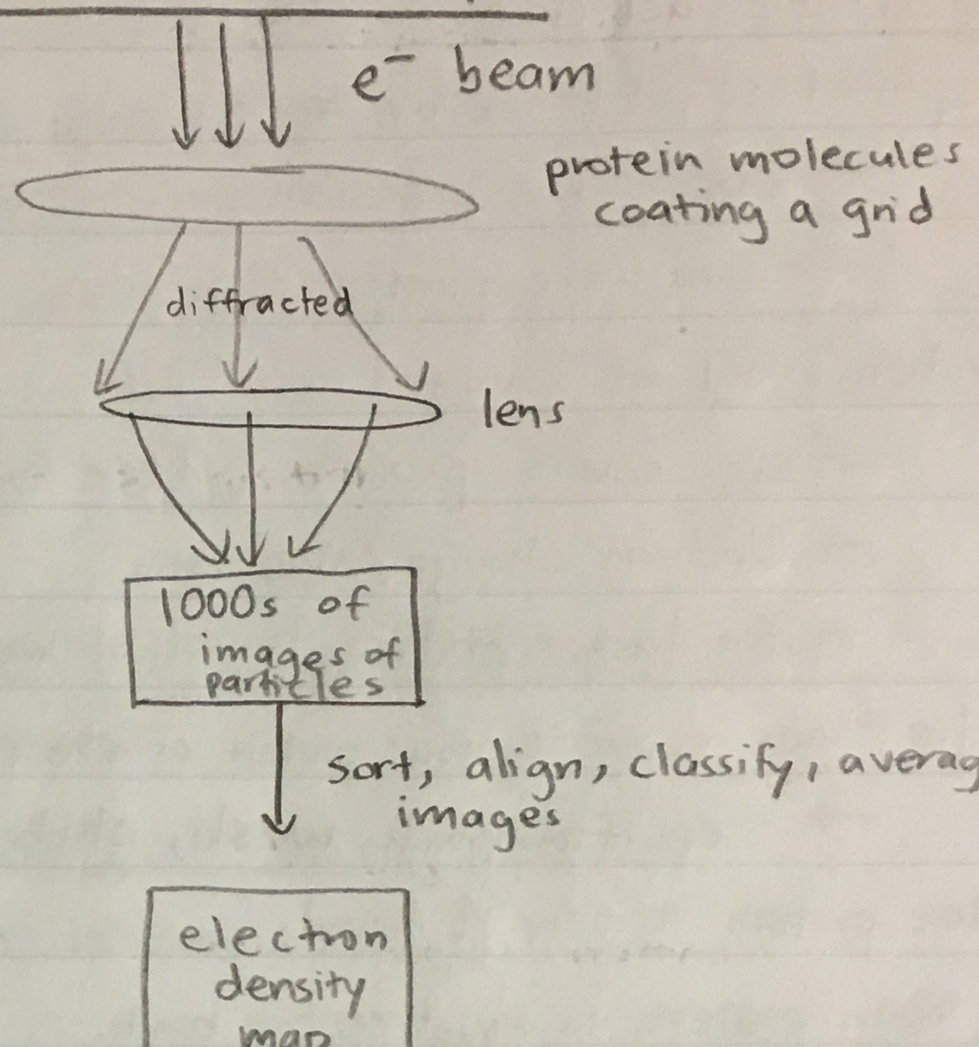
3 main methods for determining structure:

- One of them is NMR (nuclear magnetic resonance spectroscopy), but we're not talking about it.
- Others: X-ray crystallography and electron microscopy

Currently, 145,690 structures on the protein data bank.

Light microscope vs Electron microscope





A few more details...

- Need to preserve specimen in some way
  → Stain specimen or use cryo-EM (frozen-hydrated)
- Have to align all particles and avg. to get detailed info.
- But the particles might have different conformations or orientations.
  → Classify before averaging
- If 3-D, we look at different projections from different angles.
- But it's hard to see: what's a particle? what's noise?
  → wow mAcHiNe LeArNiNg!

Resolution — how precisely we know atomic positions

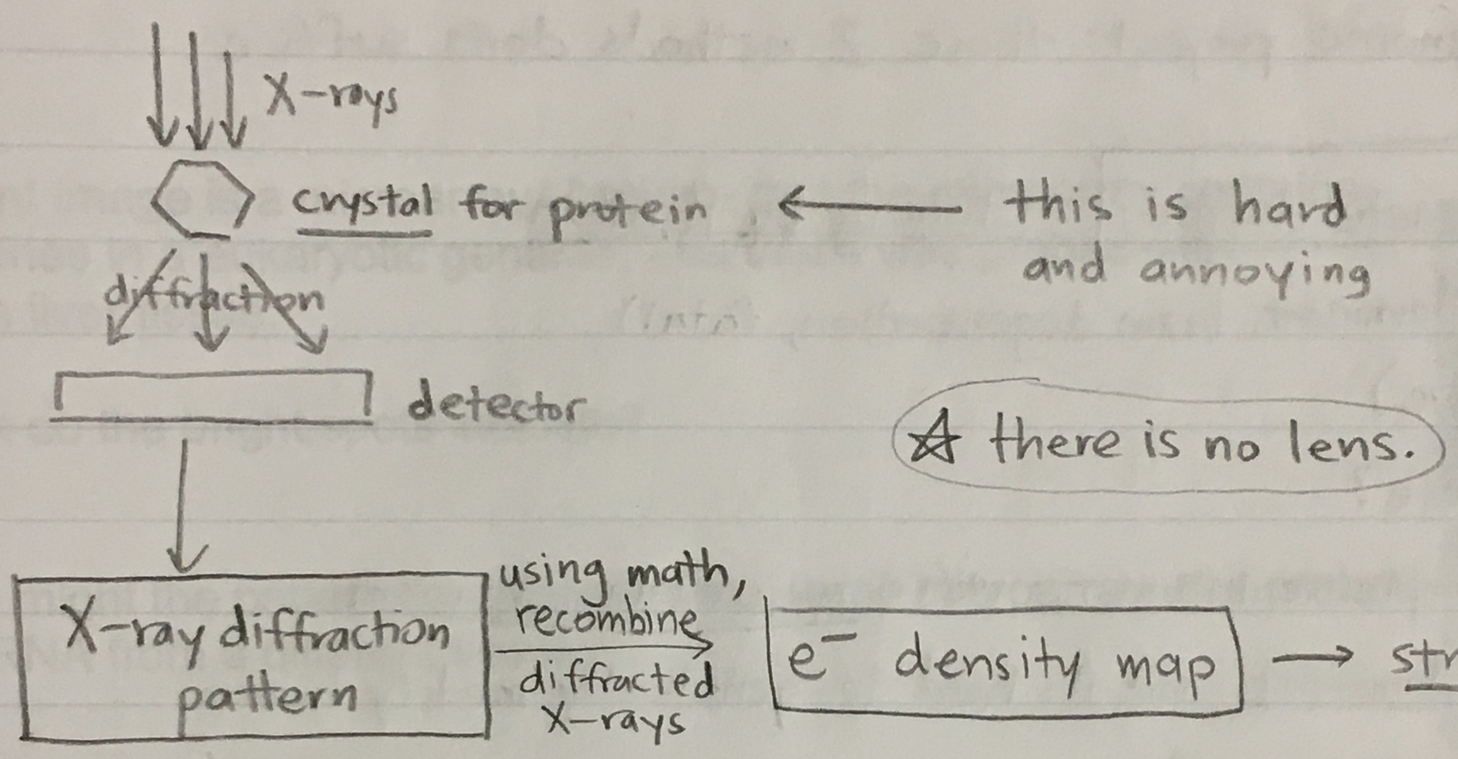
- 15-20 Å (← 1 angstrom = 10^{-10} m) tells us the overall shape.
- 6 Å – start to model α-helices
- 4-5 Å – β-sheets
- ≤3.5-4 Å – side chains.
- ≤2 Å – atomic positions (w/ precision)

* low-resolution can still be accurate, but not precise

Cryo-EM – resolution revolution
→ e^- more easily & efficiently detected w/ aligned / averaged frames,
→ Microscopes going from 19 Å resolution → ~3–4 Å.

compare to...

X-ray crystallography



- Sometimes hard to crystallize proteins in a regular repeating pattern, or only in one conformation.

Pros and Cons

✓	X
• Cryo-EM doesn't need crystals, while X-ray does. ↳ better for membrane proteins (e.g. need lipid bilayer) ✓	
• Cryo-EM can show multiple conformations in one grid ✓, while X-ray may not be able to always do this.	
• In X-ray crystallography, easier to validate crystal structure. (check it back w/ the data).	
• Still higher resolution w/ X-ray at the moment.	

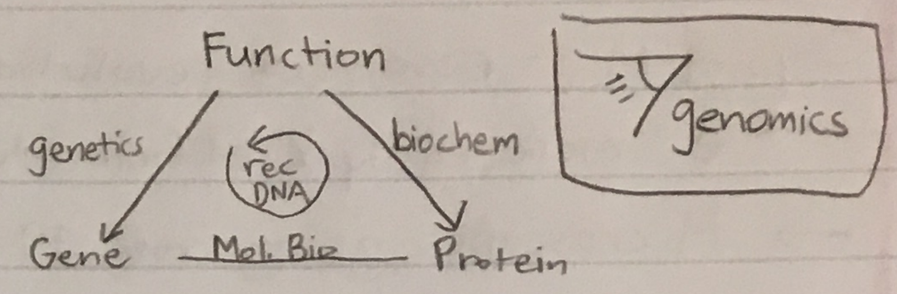
7.012

Lecture 20

10/29

(Missed the first 5 minutes, sorry)

Look at human genome @ larger scale.



Recall: make library by transforming into plasmids

→ To find those

→ Origin of human genome project: those 2 methods don't suffice

How to find gene for a genetic disease? – Cloning by Position

- Huntington's disease (dominant, brain degeneration, fatal)
- Cystic fibrosis (recessive)

→ How to do these?

→ Can't do cloning by complementation — we're multicellular organisms.

Plus, Huntington's is dominant, and it's hard to genetically modify humans

→ What about cloning by expression?

→ We have no idea what proteins caused this! So that doesn't work either

☆ Maybe find linked genes by mapping them.

In flies: (e.g.) cross

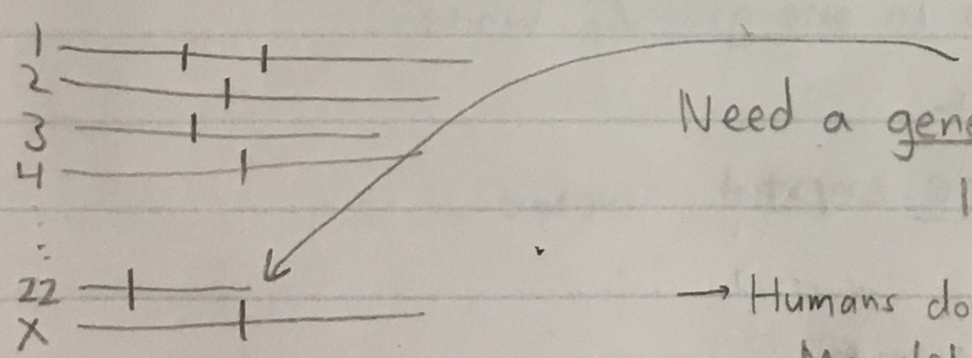
b	cn	vg
+	+	+

x

b	cn	vg
b	cn	vg

and see how linked the genes are.

How to do this in humans?



Need a genetic marker; can't use something like vestigial wings

→ Humans don't have a lot of single-gene Mendelian inheritance, either.

→ Also, we can't cross humans ethically

Key idea

Human DNA has spelling difference b/w 2 copies of the human genome

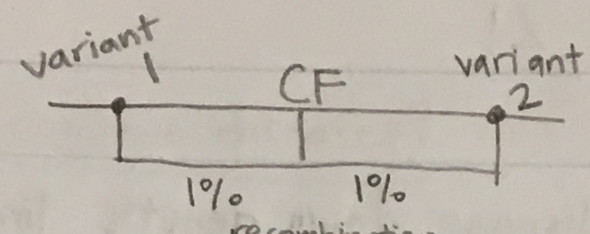
↳ "natural heterozygosty" ~1/1000 bases (e.g. G vs A in one spot)

→ Trace them in families! Keep track of each one. More likely to find specific markers of interest

* Single-nucleotide polymorphism

- Markers near the disease will be very likely inherited (low rec. %)
- Use multiple families across generations.

(e.g.) cystic fibrosis



→ Chromosome walking — put gene over library, find adjacent genes that overlap. → make a primer based on sequence of your marker → sequence ~N base pairs, repeat

→ Takes a very long time

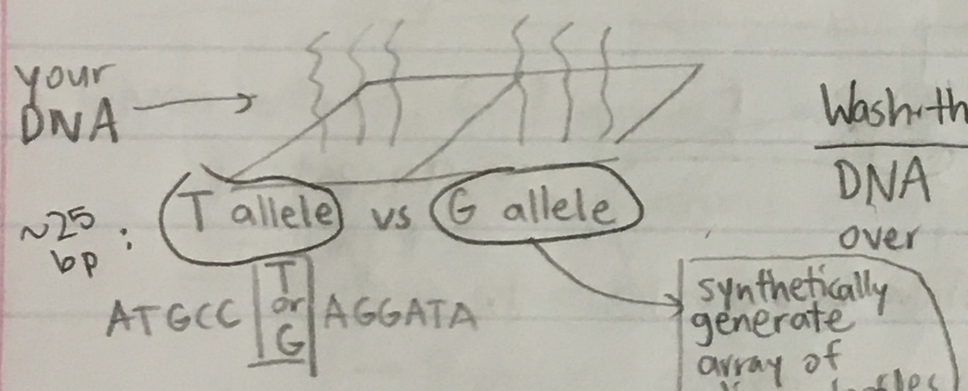
What's a better way? Human Genome Project (flex)

→ Instead of boring genetic maps, find all in parallel.

→ Do lots of sequencing w/ computers

Then what? Improvements?

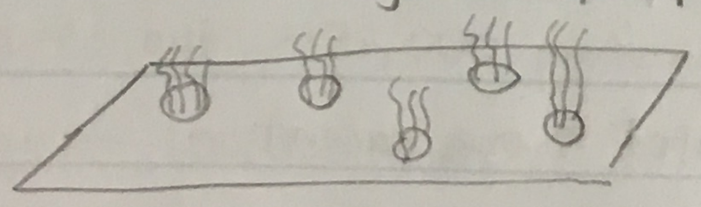
- Genetic variants? Which is more likely?



- How about sequencing?

→ Wash random pieces of DNA of you

→ Do a PCR reaction to get many copies



→ Recall: we used dideoxyribonucleotides and we'd stop.

→ But we can use a chemically reversible reaction!

Reversible blocking groups.

→ we can now do 7 billion spots of this × 150 bases/spot × 2 sides of a chip
~ 2 trillion bases = 24 human genomes! for ~$700./genome

Review

7.012

Lecture 21

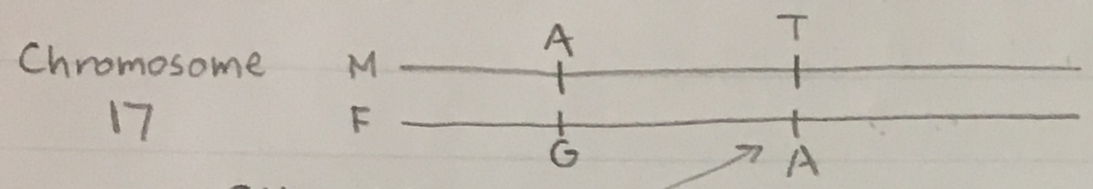
10/31

Last time — how to map strange trait / disease down genetic line

→ Modern biology — how do we do this?

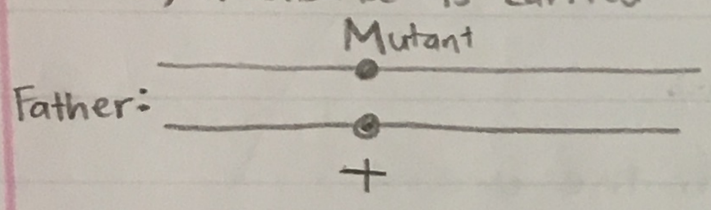
- Human genome finished in April 2003
- Today, all we do is just observe

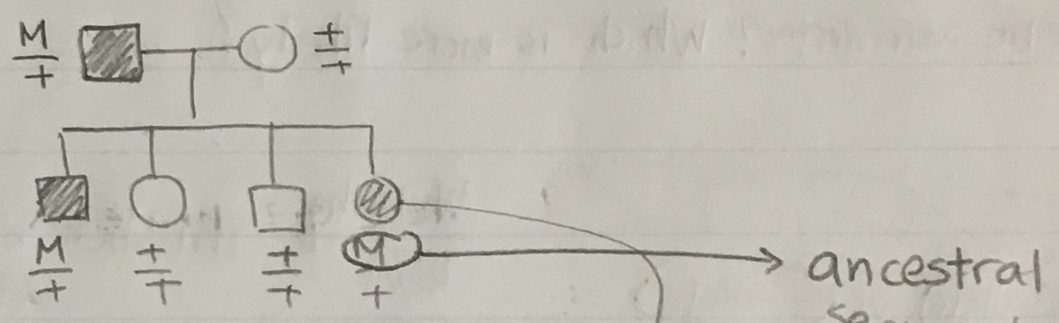
Mapping disease genes



→ Site of heterozygosty b/w the two chromosomes (1 per 1000)

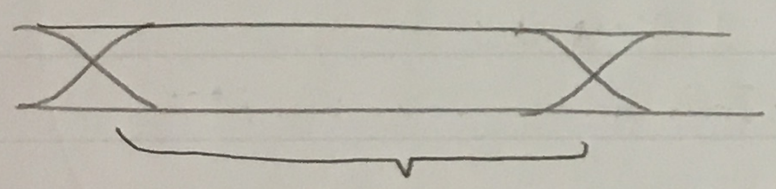
So, if disease is carried:





→ 1cM = 1% crossing-over chance ~ 1 million bases.

↳ on average, ~ 50 cM = first cross-over.

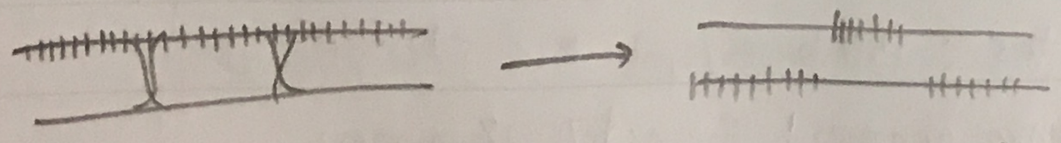


→ 50,000 polymorphisms in dad.

→ Look for ~50k genetic markers attributed to one parent

(e.g.) recessive monogenic traits

→ Inbreeding often leads to homozygosity



segments get smaller:

$\text{in cM} = \frac{50}{N}$ for N generations

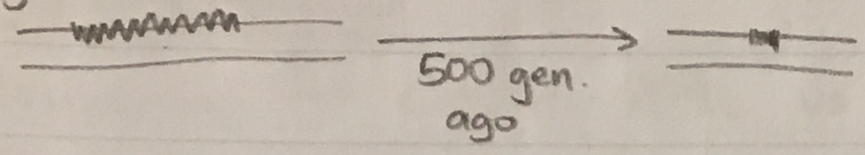
→ But in inbreeding, you don't just inherit mutation:
the whole chunk of mutant DNA is the same!

→ Homozygosity mapping — find large chunk that is identical!

→ (eg) Finland has higher variance in disease alleles
b/c of founder bottleneck. We can find the 1000 common genetic markers from 100 generations ago!

What about polygenic common diseases?

(e.g.) if a gene increases risk by even small factor (2x)



→ You can still see the common region! No need for Mendelian inheritance.

(e.g.) we can map schizophrenia! (at MIT)

→ 3000 w/, 3000 w/o were mapped: no statistical significance

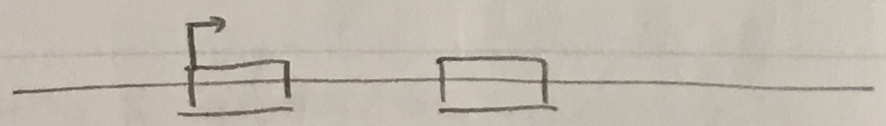
→ increased size to 20,000 → 50,000 → 110,000

↳ Found lots of excess sharing in 100+ spots!

Next, reading genomes

genome ~~~~~~~~~~~~~~~~~~~~ 3 billion nucleotides

- Given letters, find genes? Where are proteins made?
  - → Expect stop codons every ~60 letters on average
  - → If long stretch has no stop codons, highly significant.
  - → Problem: coding sequence has introns; we don't know where they are. (they're ~180 bases)
  - → Solution: read & sequence the cDNA from spliced mRNAs! Now unsplice them!
- → figured out ~21,000 protein-coding regions.
- Transposons — 100,000 copies!



promoter → mRNA → protein = reverse transcriptase copies RNA → DNA.

This one only recognizes itself. So it puts itself somewhere else in the genome!

→ it's been spreading since ~1 billion years ago.

→ We can even map them over time.

- Compare humans & chimpanzees — common ancestor → similar DNA sequences.
  - → genomes ~98.5% identical.
  - → We can still see alignments from millions of years ago!

✱ Important DNA (functionally important) has slower rates of mutation (natural selection!)

↳ so exons = less mutation. Same (ish) with promoters

→ Compare human, mouse DNA

↳ 6x DNA show slower signs of mutation

→ <u>Regulatory</u> systems!

→ Human vs Neanderthals

↳ there was interbreeding, only in Europe/Asia

☆ can find children that are hybrids!

Gene Expression

2 patients with leukemia / blood cancers

→ what if we <u>compare</u> patterns of <u>RNA</u>?

- 20,000 genes in a genome
- To describe, put them in a vector of length 20,000.

Thinking: every cell has some RNA per gene. It's a point in $\mathbb{R}^{20000}$.

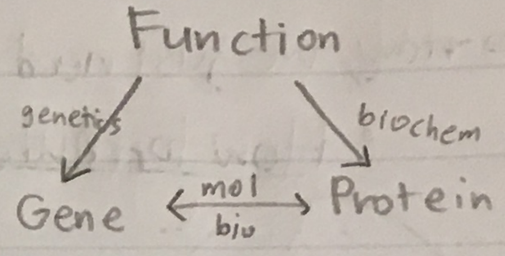
What can you do?

7.012

Lecture 22

11/2

We're going to complete the triangle today:



Global view of RNA Expression

→ Study leukemias → classify AML vs ALL (looked at nuclei, etc.)

↳ different treatments = respond differently

→ Slowly found more & more differences, but it took ~40 yrs to classify

☆ Instead, question: express it as "how much each gene on/off"

↳ tumor 1 = $(x_1, x_2, \dots, x_{21000})$, x_i denotes level of expression

→ Get gene expression by reverse transcription OR a chip

in the last 3-4 years: sequence & count.

oligonucleotides for each gene

→ Now you can easily match by classification: clustering per group!

→ Last 5 years — human cell atlas

↳ Get gene expression cell by cell

- 18 cells → 12.5 million cells!

☆ Discover lots of new cell types by doing this

→ We can't quite finish the triangle yet. How to go gene → function?

→ If we don't have a mutation, make one — interfere with genes!

Adding & subtracting genes

→ Add DNA to bacteria in form of plasmids

→ What about a mouse?

- We could inject a virus
- Or take a fertilized egg and inject DNA there!

→ Hard to be exact, but 1st step to transgenic mice

→ What if we use embryonic stem cells? (for mice)

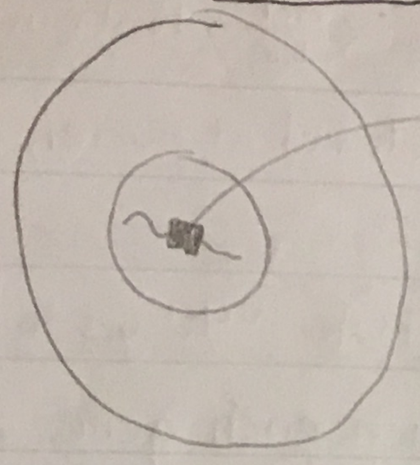
- Early stage of embryo — they have an inner cell mass
- Grow those cells → inject into new embryo

↳ they'll become part of new mouse

→ To recognize, use different fur color! → chimera

→ So now, we just need to be able to change those ES cells! (embryonic stem)

- How to mutate a gene? w/ cells growing on a plate



let's say I want to remove YFG, "your favorite gene"

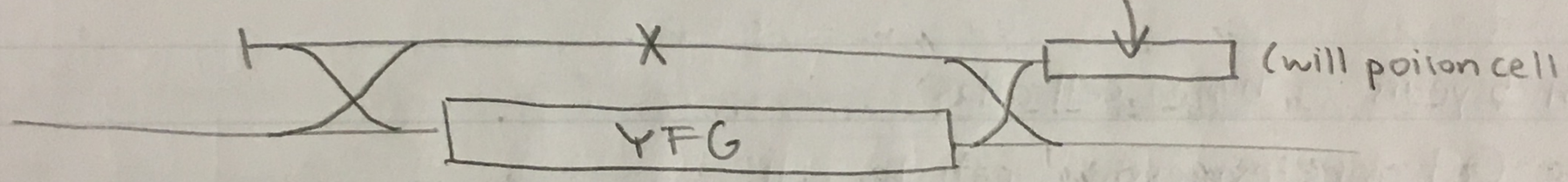
Transfect cell w/ damaged DNA w/ mutation in YFG gene.

→ Then hope for recombination?
only happens 1/million cells

☆ Select for those cells! Give cell an advantage: antibiotic resistance. (positive selection)

→ But we also need negative selection so only rec. at YFG is good.

→ Put negatively selectable marker gene!



→ After ~1 year, could get mouse of choice!

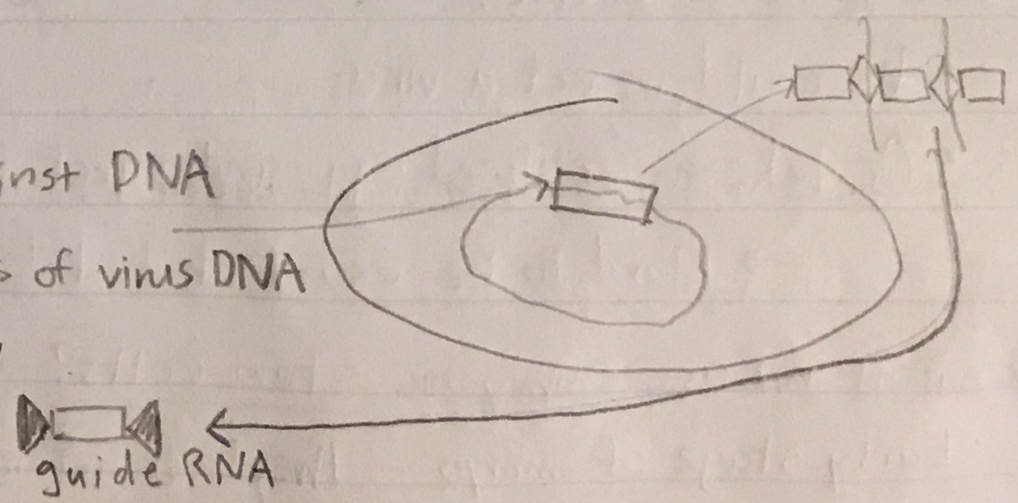
→ There's other modifications we can make (tissue/organ-specific)

Next method: RNA interference

- What if we destroy RNA corresponding to gene?
  - → Cells have sensor system that sees double-stranded RNA
    - ↳ search & destroy: find anything that matches it & destroy it
  - → Defense against dsRNA viruses!

CRISPR

- Bacteria also have bacteria mechanism against DNA
  - → library of memories: 20-30 bases of virus DNA
  - → this is called a CRISPR array
- Bacteria copies into CRISPR RNA
- Cas9 scans DNA/genome and cuts.



→ Works in human cells! (January 2013) → now you can introduce DNA and cut

7.012
Lecture 23

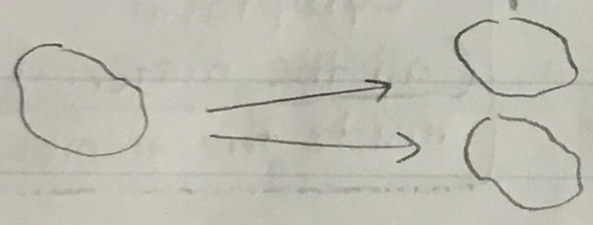
11/5

- Exam on Wednesday
- Pset 6 will be posted a bit earlier
- Vote tomorrow!

Today we'll take a closer look at <u>cell structure</u>

<u>"Cell Theory"</u>

- Fundamental unit of life – (all?) living things are or have cells
  → Cells are made from preexisting cells ; evolved from a common ancestor



cell division

<u>Size of a cell</u>

→ Typically ~1–100 μm ($1\ \mu m = 10^{-6}\ m$)
→ Can't get much larger b/c <u>SA to V</u> ratio $\propto \frac{1}{r}$.
★ Activity in cell scales w/ volume, so we need more and more nutrients / produce more and more waste
★ But transport scales w/ surface area.

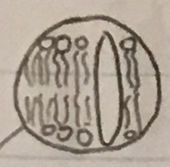
<u>Classifying cells</u>

- Prokaryotic cells – found in form of bacteria, archaea
- vs
- Eukaryotic cells – animals, plants, fungi, <u>protists</u>, other eukaryotes?

Differences:

① Eukaryotic cells have membrane-enclosed internal <u>compartments</u> called organelles (e.g. nucleus)
② Eukaryotic cells ~10–100× bigger than prokaryotic cells

Prokaryotic cell



cytosol (intracellular fluid)

cell membrane (outer boundary)

- consists of phospholipid bilayer, (hydrophobic tails in the middle)
- proteins. Acts as selective barrier (transport in/out), maintains homeostasis, cell signaling, communication.
  - constant internal environment

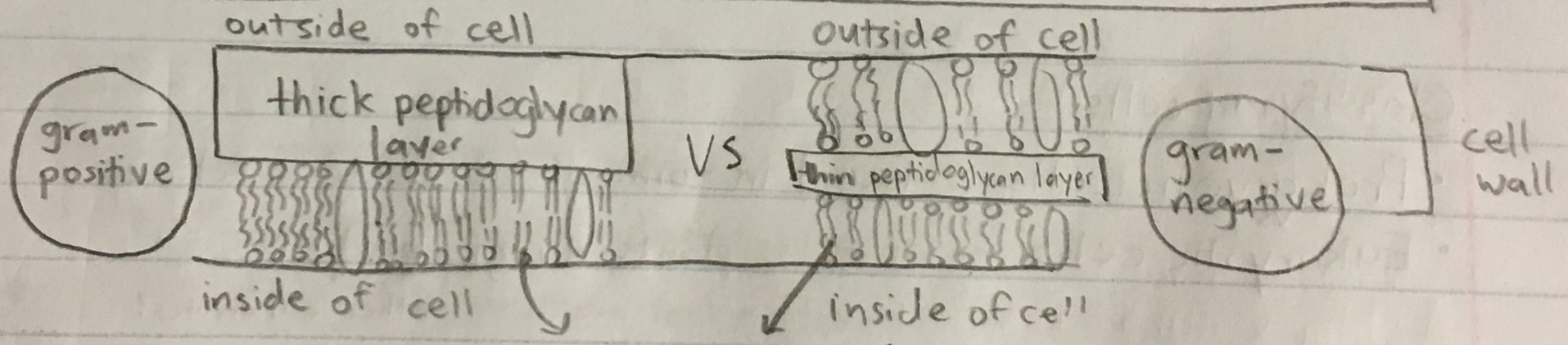
ribosomes (for protein synthesis)

cytoplasm: (all the material inside the membrane)

nucleoid (region in which DNA is located)

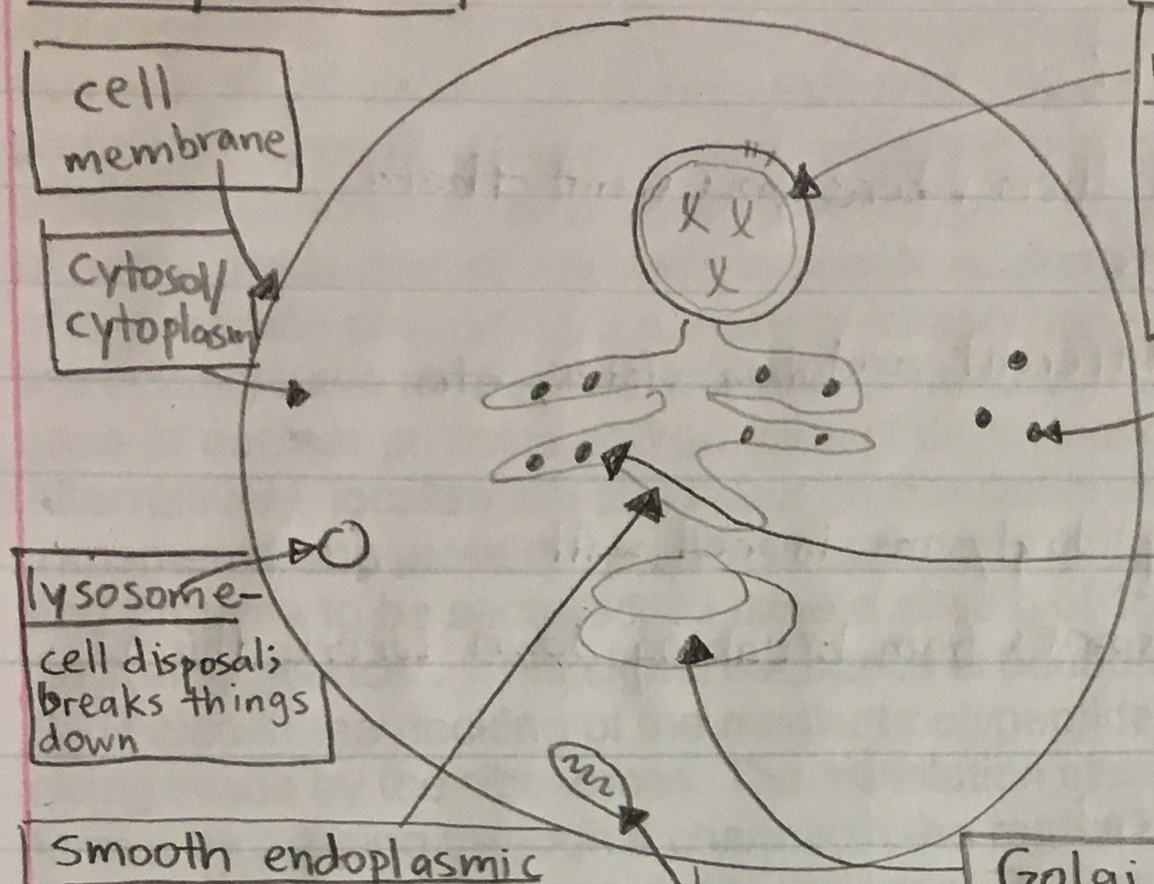
cell wall (found in most prokaryotes)

- consists of peptidoglycans. (peptides, sugars) 2 basic types: (in bacteria)
  - gram-positive vs gram-negative. (retaining stain after washing)



→ antibiotics generally target cell walls - easier for gram-positive

Eukaryotic cells



nucleus w/ **nuclear membrane**

↳ contains most of the genetic material.
site where gene regulation, transcription occur.

free ribosomes – make proteins used in the cell.

rough endoplasmic reticulum (RER)
contains **ribosomes** as well.
→ makes proteins to go **into** membrane, outside cell.

Golgi apparatus – modifies, sorts, and packages proteins that will be used @ membrane / outside cell

mitochondria – powerhouse of the cell, site for ATP synthesis. contains its own DNA, ribosomes, can even make its own proteins. (prokaryotic-like cell inside?)

7.012
Lecture 24

11/9

Protein localization — why is it important?

→ "Everything in its place" — only put them where you want them
(e.g.) DNA polymerase in wrong spots
→ Micro-environments: different pH, different oxidation states, etc.
→ Protein folding, modification
↳ sugars are not flexible — peptidoglycans in cell wall
↳ glycosylation — prevents lysosomes from breaking down wrong things

Natural next question: how does this happen?

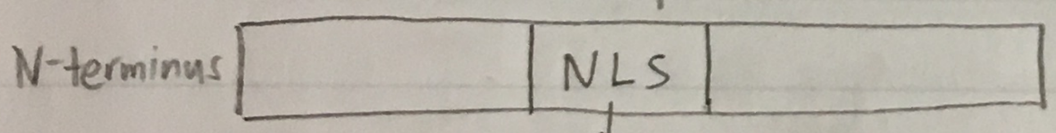
How do proteins get to their cellular location?

- Cytoplasm is where the proteins are made (by the ribosomes)
  ↳ If proteins are cytoplasmic, easy
- Other proteins for lysosome, plasma membrane, Golgi, ER, outside the cell...
  ↳ need a signal — seq. of amino acids that fold so other proteins can bind (for import)

(1) Post-translational import → after proteins are made in cytoplasm

(e.g.) Mitochondria is "given" some proteins aside from its own; also nuclear proteins (in the nucleus)

↓ here's the protein



→ proteins go to nucleus

☆ nuclear localization signal

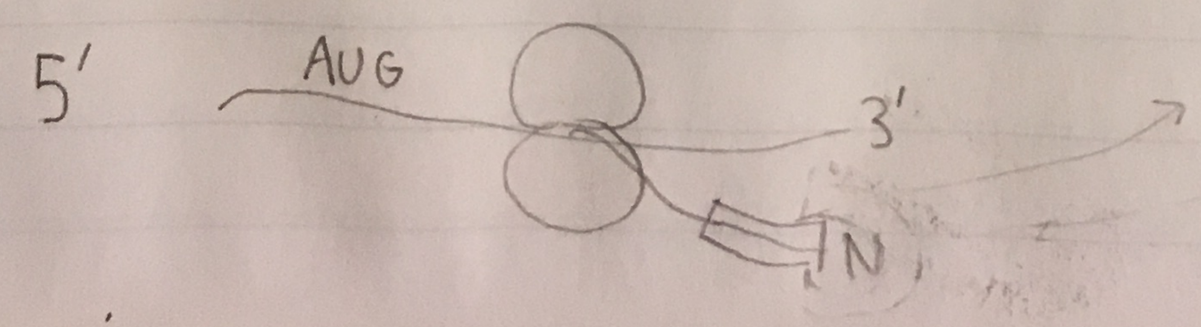
✓ → to check if it is necessary, knockout the gene → proteins go to cytoplasm

✓ → check if sufficient: take cytoplasmic protein, add NLS → went to nucleus

(2) Co-translational translocation → during translation

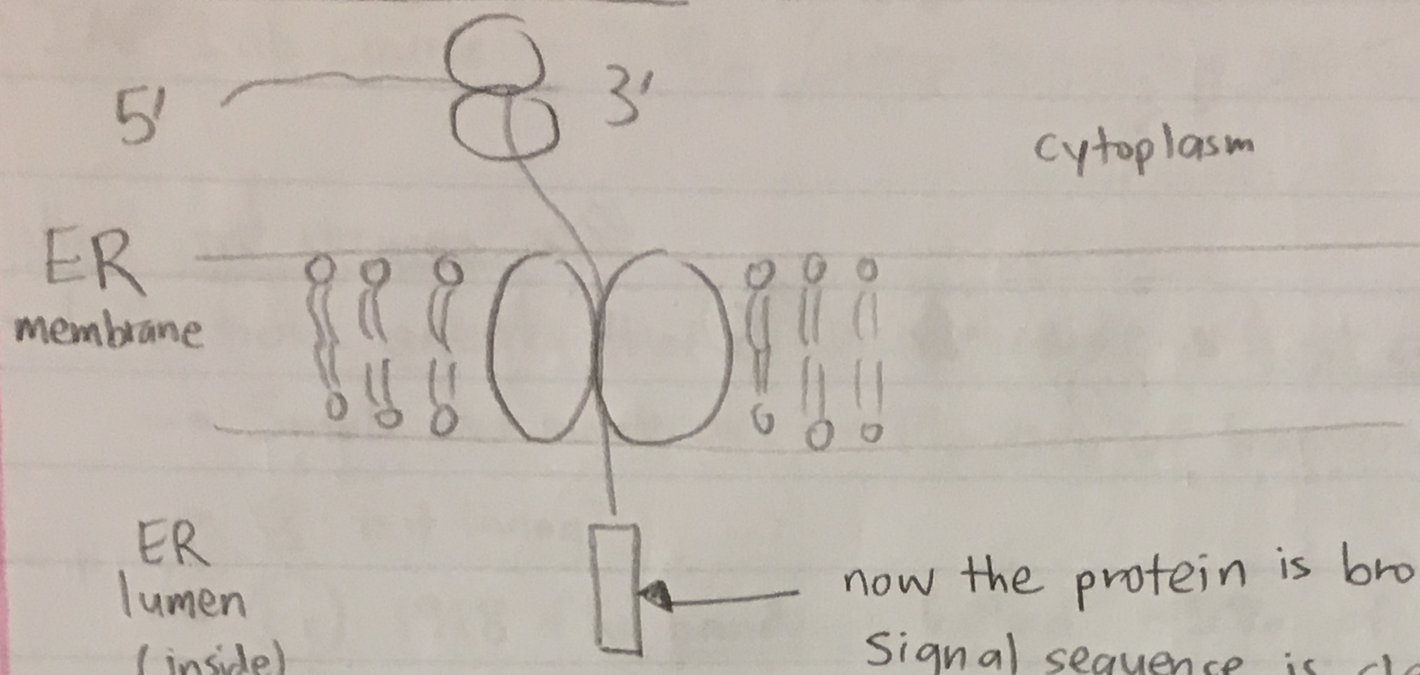
(e.g.) [ER, Golgi, lysosomes, plasma membrane, outside cell] proteins

- All proteins had signal sequence of ~15-30 amino acids at N-terminus only

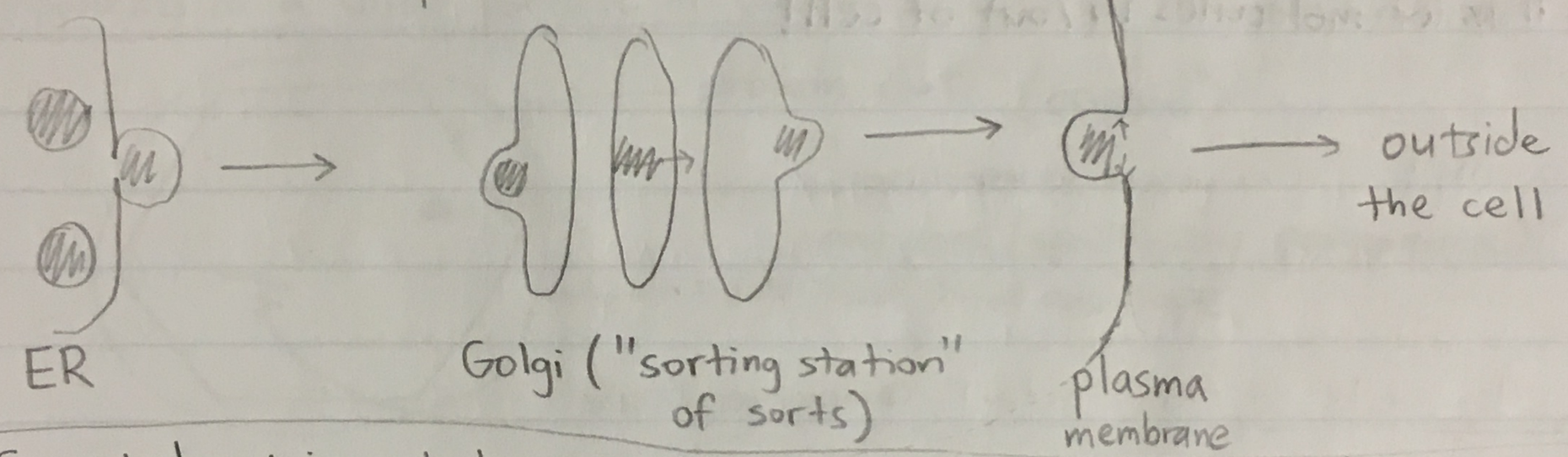


once this is made, protein called Signal Recognition Particle (SRP) binds; stops translation

- Next, SRP brings nascent peptide, mRNA, ribosome to a <u>translocation channel</u> @ ER.



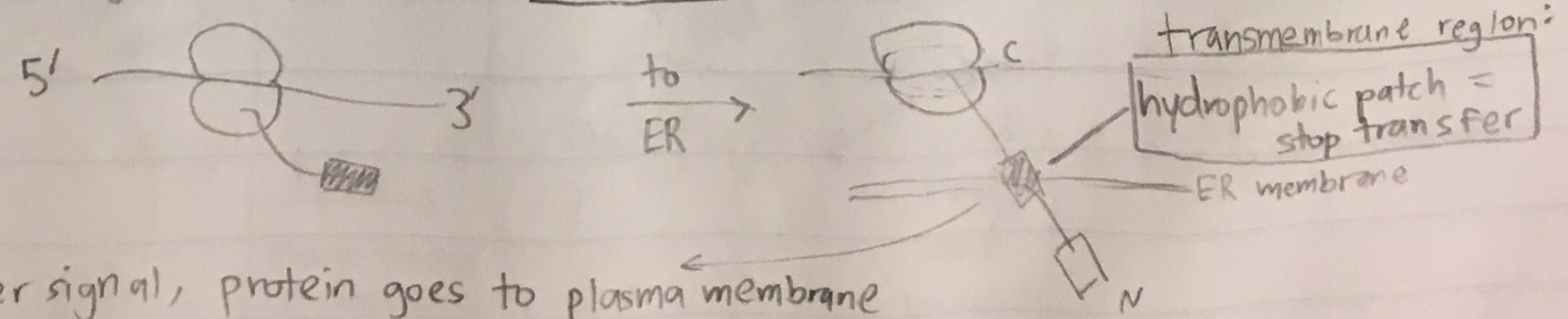
- <u>Vesicle</u> will (often) pinch off from ER



- Secreted proteins only have a signal sequence
- Proteins of <u>ER</u>, <u>Golgi</u>, <u>lysosome</u>, etc. have additional signals
  → ER retention, Golgi localization, Lysosomal localization

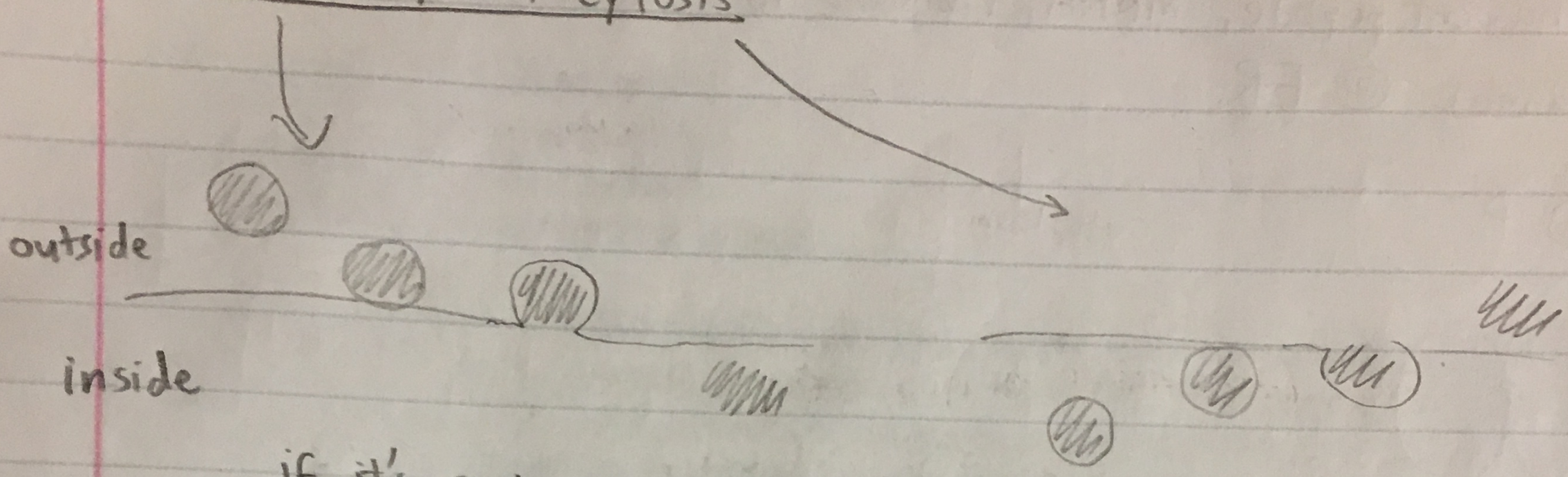
* remember outside of cell is different!
  → so inside of ER, Golgi, vesicles ≅ outside of cell, <u>not</u> inside.
  → inside is a <u>reducing</u> environment, outside is <u>oxidizing</u>

- So now — what about <u>integral membrane proteins</u>?
  → Also have signal sequence + <u>hydrophobic patches</u>



If no other signal, protein goes to plasma membrane
  ↳ N terminus on outside (since ER lumen has N terminus)

Endocytosis / exotocytosis



if it's really big, it's called phagocytosis (e.g. engulfing a bacteria)

→ Helps transfer molecules in / out of cell!

7.012

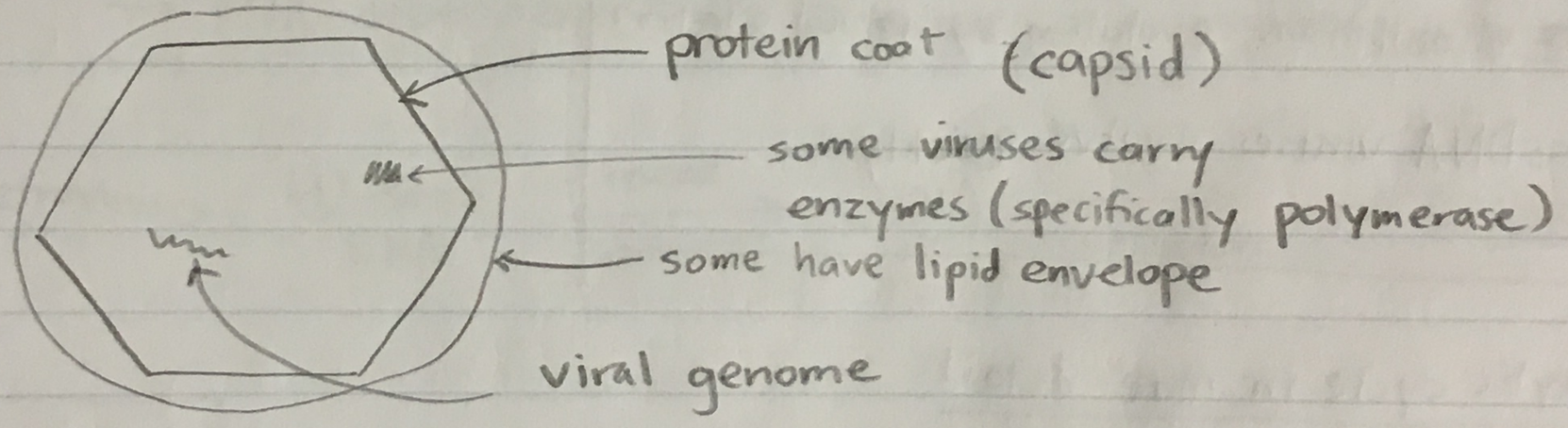
Lecture 25

- P-set 6 posted today — due 11/28
- Exam 3 is posted as well
- IAP Lab Course — 7.102 (Intro to mol. bio techniques) is apply by 12/7

What are viruses?

- Infectious agents that replicate inside a host cell
  - → Small and numerous: ~1% size of bacterial cell, ~10^{31} of them. (10 million / drop of seawater)
  - → Q: is it living?
  - → (Ex) 1918 flu pandemic killed ~5% of world population
  - → (Ex) ~36.7 million worldwide (~2.1 million children) with HIV/AIDS.

Components of a viral particle or virion

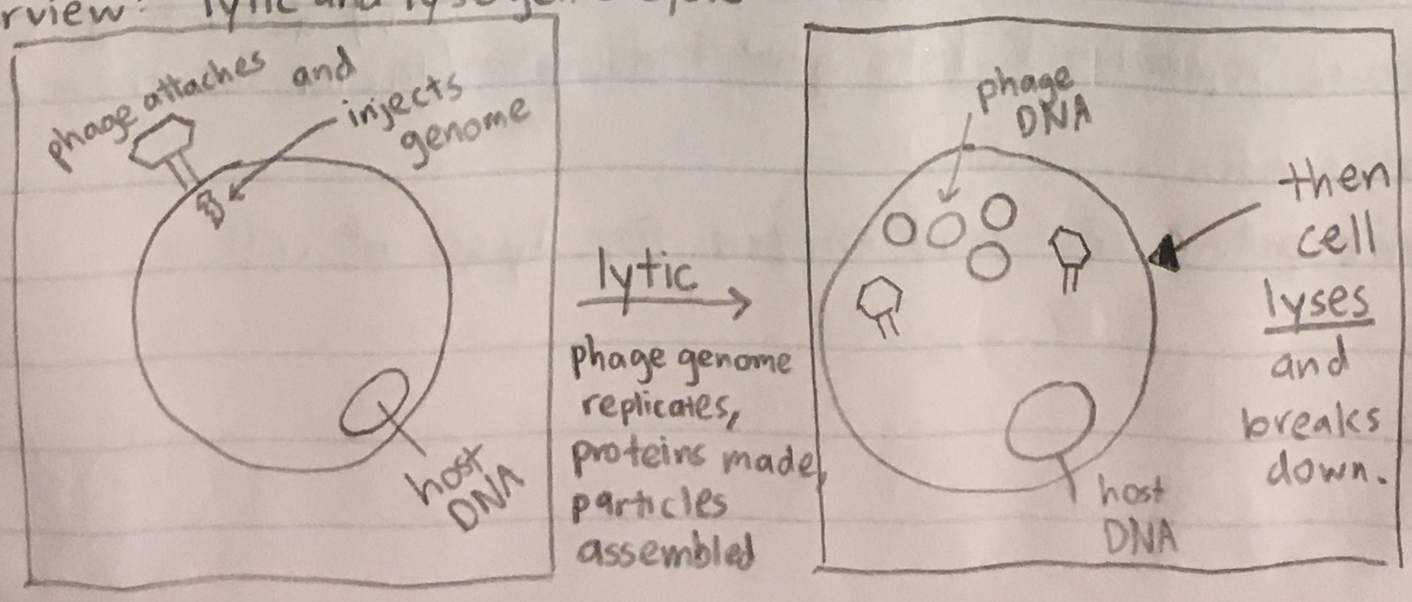


- Viral genome encodes ~2–4 proteins (small, ~2000 bases)
  - → DNA or RNA
  - → sometimes overlapping genes — frameshift to get start, stop, or ordinary codons!

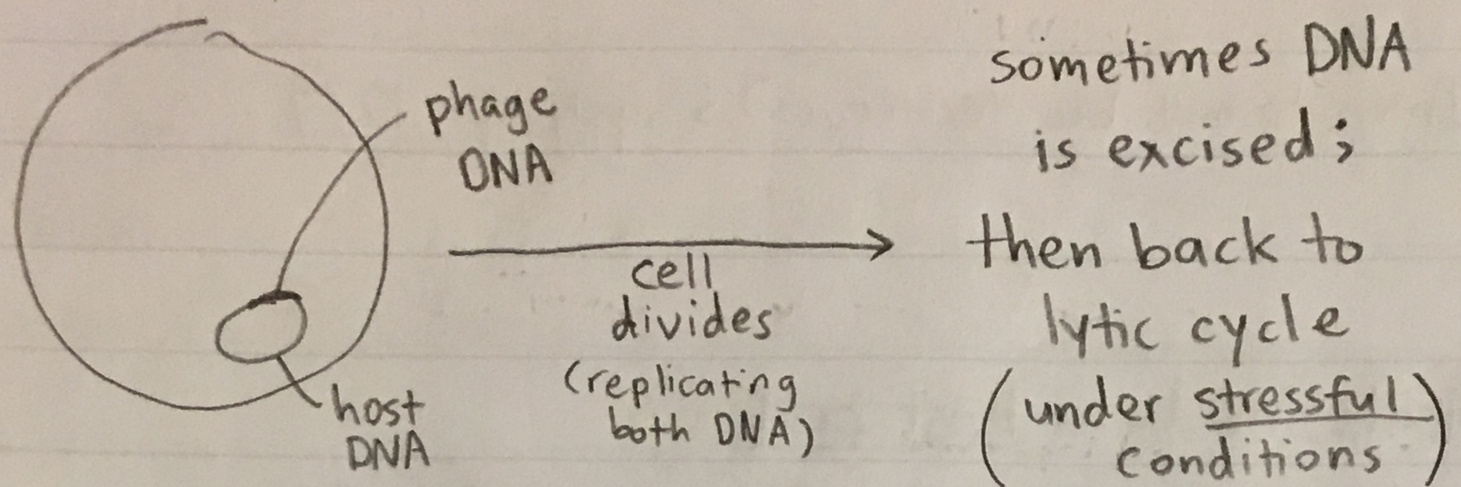
How does a virus replicate?

(Ex) bacteriophage

Overview: lytic and lysogenic cycle



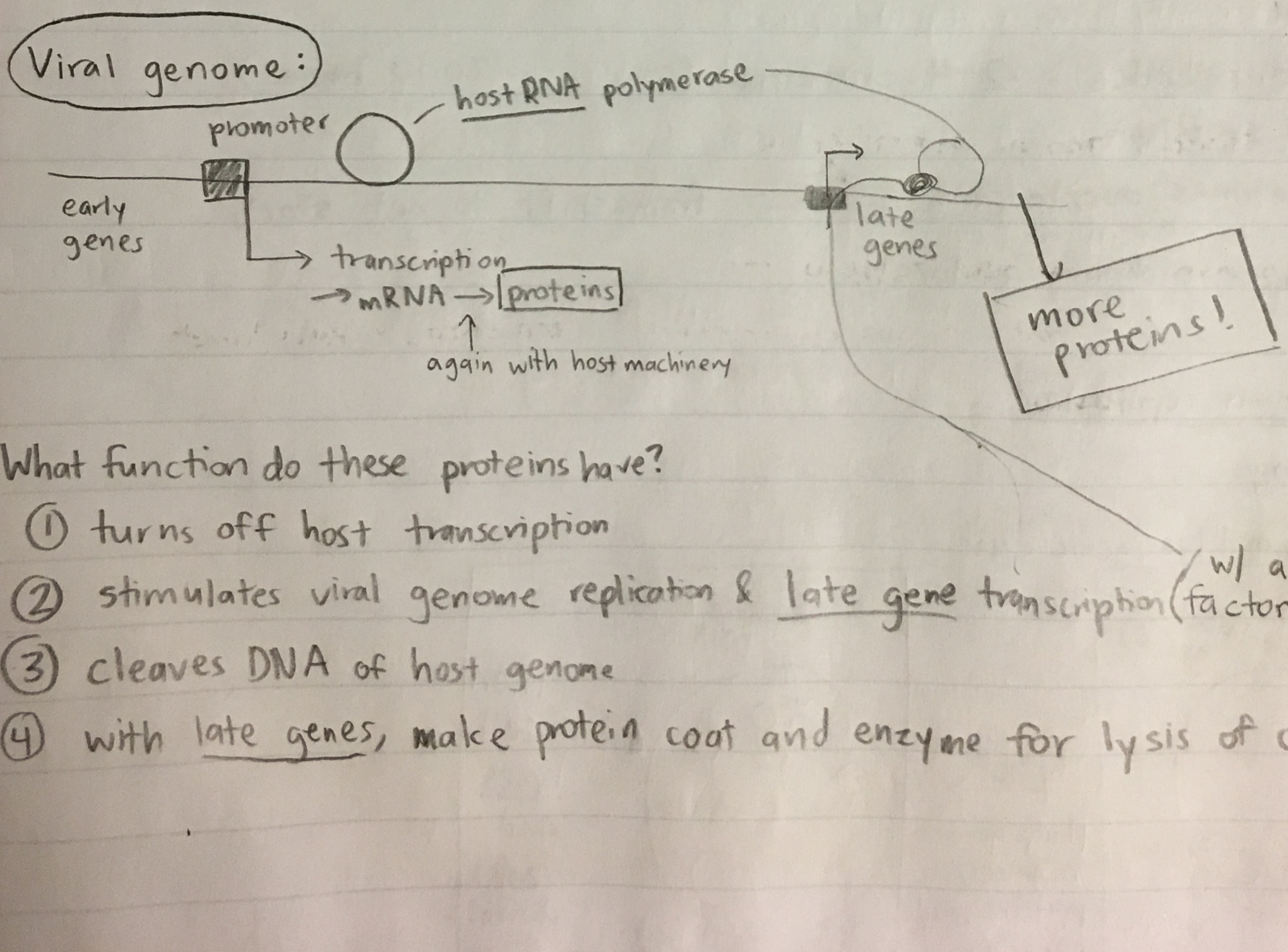
Alternative: combine host and phage DNA.



Key points

- This is a source of horizontal gene transfer
  - ↳ increases genetic diversity
- RNA viruses — replication is more error-prone b/c RNA less stable
  - → lots of mutations, evolution by natural selection
- However, ssDNA viruses evolve quickly

Looking at lytic cycle in more detail



What function do these proteins have?

① turns off host transcription
② stimulates viral genome replication & late gene transcription
③ cleaves DNA of host genome
④ with late genes, make protein coat and enzyme for lysis of cell

Types of viruses

- DNA viruses — ssDNA (parvo) vs dsDNA (herpes)
- RNA viruses — ⊕-sense (cold, SARS) vs ⊖-sense (influenza, rabies, ebola) vs dsRNA (rotavirus)
- Reverse-transcribing viruses — ⊕-sense (retrovirus HIV) vs dsDNA (herpes)

★ ⊕-sense is ready for translation
⊖-sense must be converted to ⊕-sense w/ enzyme first
→ RNA-dependent RNA polymerase - encapsulated in virion

★ Retroviruses: ⊕-sense RNA —reverse transcription→ cDNA —integration→ host DNA

Types of hosts

- Bacterial, archaeal, animal, plant cells
- Some viruses have many hosts — generalists
  and others have only a few — specialists
- High barrier to switching hosts — usually infection is too strong & kills host or too weak and does nothing

7.012

Lecture 26

11/16

What are bacteria?

- Prokaryotic cells
- There are many ; ~10 million / grain of soil
- Estimates vary — ~120k to ~1 trillion species.
  → In particular, genetic diversity mostly unknown
- Nomenclature: Genus species, (e.g.) *Escherichia coli*, *E. coli* for short
  (capitalized ↑ lowercase)
  (e.g.) *Bacillus cereus*, *B. cereus* for short
  ↳ subcategory of species = <u>strain</u>
  (e.g.) *E. coli* K-12 (lab strain), *E. coli* O157-H7 (bad)
- Live basically everywhere in/out of hosts.
  → free living : your gut, lettuce, acidic hot springs, radioactive waste (Uranium-eating)

Friend or foe?

- <u>Symbiotic</u> bacteria — mutually beneficial w/ host
- <u>Commensal</u> bacteria — gets nutrients from host, host not harmed
- <u>Parasitic</u> — (ditto above but) host is harmed.

most fall in this category

☆ Most bacteria aren't parasites, and many parasites are not bacteria!

What do parasites do?

- Produce toxins (poisons)
  → Protein-based or small molecules made by proteins
  → (e.g.) *B. cereus* makes <u>cereulide toxin</u>, causes stomach distress
  → (e.g.) *C. difficile*, *C. botulinum*
  Botox in small amounts, but 400g would kill all people
- Infect parts of body where they shouldn't be
  → <u>normal</u> : gut, mouth, nose, skin
  <u>bad</u> : internal tissue (e.g. *H. pylori*)

Bacteria aren't "born to be bad" though

- **Virulence** — pathogen's ability to damage host
  → Not all pathogens are virulent all the time
  → Controlled by **virulence factors** (proteins / small molecules made by proteins) which increase destructive nature of pathogen.

How to fight bacterial / viral infections?

- **Immune system**:

	viral	bacterial
natural response from body	✓ works	✓ works
to elicit immune response, vaccines can be used	major method !	don't work as well

  → what's in a vaccine?
  → Live-attenuated, killed viruses, viral proteins
  → No vaccine for HIV, Zika yet

- **Drugs**:
  - anti-viral drugs successful sometimes, but not primary method
  - antibiotics — primary treatment for bacterial infections
  → both typically small molecules
  - Anti-viral drugs target viral enzymes (e.g. virus polymerase, reverse transcriptase, HIV protease)
  - Antibiotics target differences between human/bacterial cells
    → e.g. cell wall biosynthesis, ribosome
    → Kill bacteria or keep them from reproducing

☆ Blunt instruments — kill both good / bad bacteria
☆ **Antibiotic resistance!**
  → Bacteria adapt b/c there are not enough new antibiotics
  → Current ones being used too often.

CDC levels of concern.

7.012
Lecture 27

11/19

- P-set due 11/28

"The bacteria within us" — topic for today.

Antibiotic resistance, con't

Defn — when a species or strain becomes resistant to antibiotic.

↳ multi-drug resistance = multiple antibiotics.

- Where did this come from?

→ We keep using the same ones.

↳ Lack of new drugs — discovery void 1986–now.

↳ particularly b/c of cost to bring to market.

→ We use them too often

↳ Humans and animals

↳ allows for natural selection, spread of AR bacteria.

→ Natural selection ⟶ evolution is fast when selection pressure is high.

→ Blunt instrument — kill both good/bad bacteria, untargeted can also develop resistance.

- How big of a problem is this?

2013 US: 2 million sickened by AR bacteria

↳ 23,000 died; cost in this year ~ $20 million.

CDC has threat level for bacteria — urgent, serious, concerning

(ex) N. gonorrhoea — 246,000 drug-resistant infections.

↳ most resistant to tetracyline

(ex) Vancomycin-resistant enterococcus — 20000 drug-resistant / 66000 infections per year.

☆ Resistance has developed for all classes of antibiotics.

- How does resistance develop?

① Inactivation — enzymes able to reduce/inactivate drug

(e.g.) penicillinase — cleaves the molecule

② Modification — alter target enzyme so molecule (antibiotic) cannot bind.

③ Change in ability to get in / stay in cell.
- (e.g.) impermeability — change in cell wall protein
- (e.g.) Active pumping (efflux of drugs)

How is resistance transferred?

① bacteria multiply; take over new hosts?

② bacteria can pick up a resistance gene.

→ Horizontal gene transfer. Take genes and give them to other bacteria. Sometimes involves bacteriophage.

Human Microbiome Project (HMP)

- microbiome — combined genetic info of microbes from particular environment.
- Learn about bacteria that is not easily cultured
  - → culture = grow bacteria under defined conditions
  - → >99% bacteria cannot be easily cultured; we want to know more
    - ↳ instead, culture them in our body! Isolate total DNA from microbes in/on body. (metagenomics)
      - ↑ study of genes from whole communities
      - → over 1000 microbes from diff. body parts
  - New tools developing for DNA sequence analysis

Results of HMP: 10:1 ratio of bacterial : human cells

→ (e.g.) you have 5000 times more bacteria than # people on Earth

☆ Microbiomes are unique for each person!

- 1 million (human) microbiome genes!
  - ↳ 80% not assigned specific function, 50% no annotation.

BLAST search — find against others in data bank.

Alternatively, biochemical characterization or look at operon.

Idea: manipulate microbiome

7.012
Lecture 28

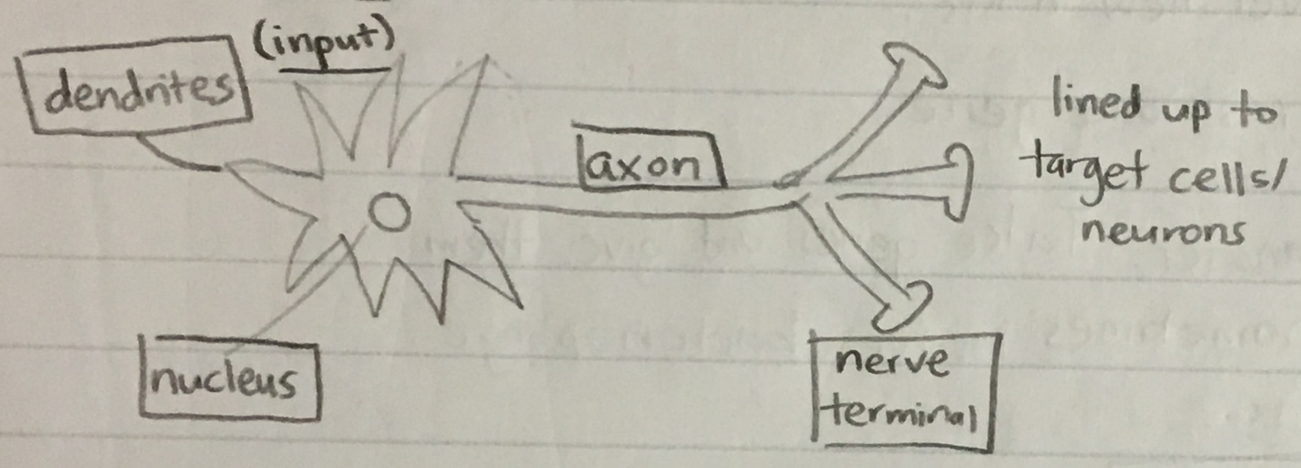
11/21

Neurobiology — the engineering of life

What are neurons?

- Nerve cells ($\sim 10^{12}$ in body)
  → make $\sim 10^3$ connections to other cells (synapses)
  (e.g.) photoreceptor in eye → axon → muscle fiber
- Many types of receptors: light, sound, touch, temperature

Here's one neuron:



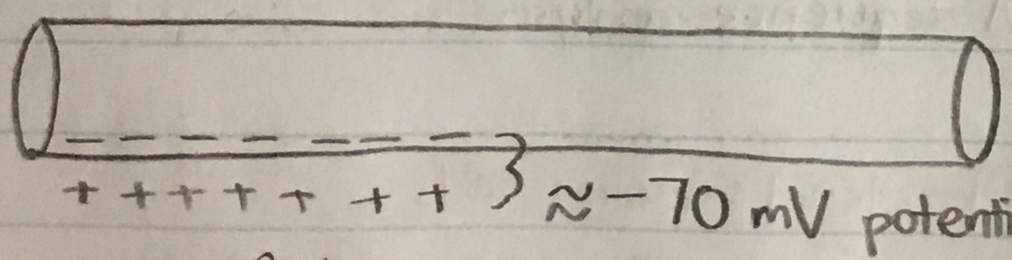
Fact: while the cell body may be ~10 μm, axons can go up to 1 m (brain → spinal cord!)

What are some questions to ask?

- How do receptors transfer signals? How sensitive are they?
- How are signals transmitted along an axon?
- What about across synapses?
- ☆ How do patterns of connections give rise to computations? How do they change for learning? → What about consciousness?

Transmission along an axon

- Take squid axon and cut off the ends — look at electrical properties

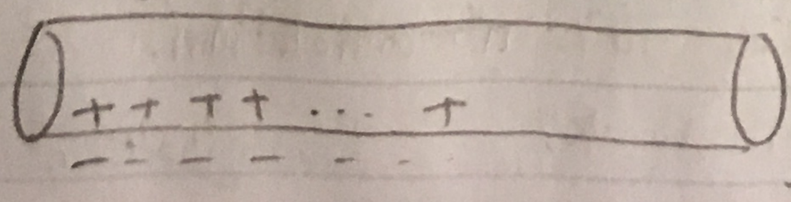


≈ -70 mV potential difference.

But thickness of cell membrane = 3 nm

→ Electric field = $\frac{-70\text{ mV}}{3\text{ nm}}$ = 200000 V/cm

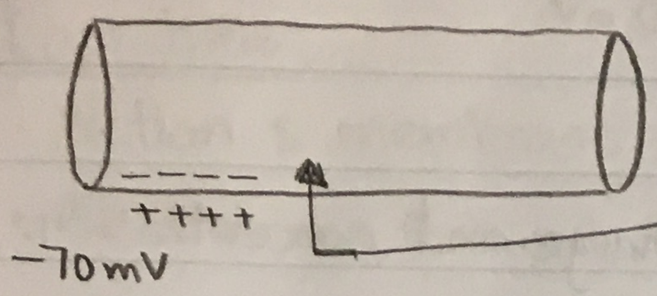
→ This changes when a neuron fires:



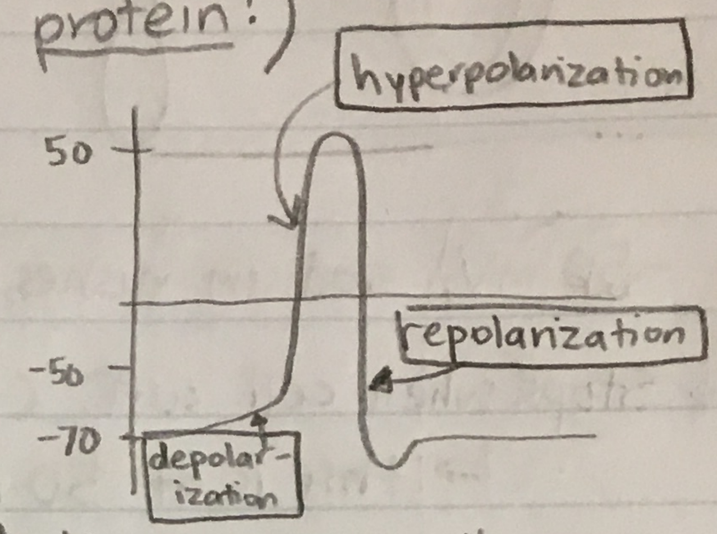
+50 mV.

Large diff. in things w/ dipole moment or charge!

(How to transmit information? Try affecting a protein!)

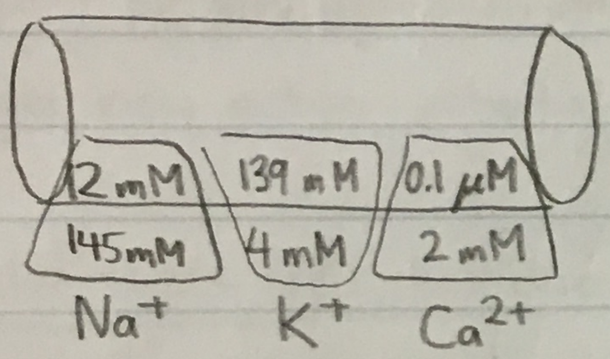


Let's say we gently introduce charge. This happens at −50 mV:



"Action potential"

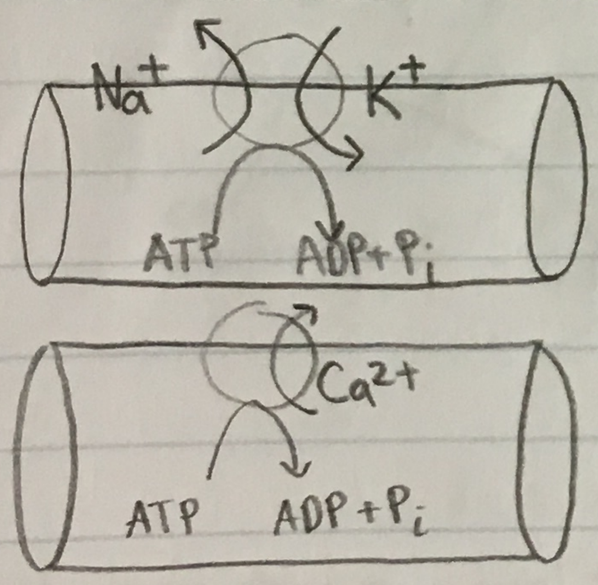
Where do charges come from? Ion concentrations!



Ionic gradients somehow shift!

How?

Let's look at the membrane proteins.

① 

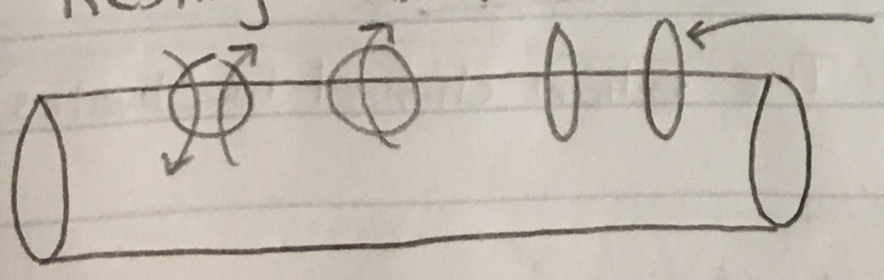
Process requires energy: ATP → ADP + organic phosphate

(Na^+-K^+ ATP-driven pump)

(Ca^{2+} ATP-driven pump)

These are called (obviously) ATP-driven pumps

② Resting Channels



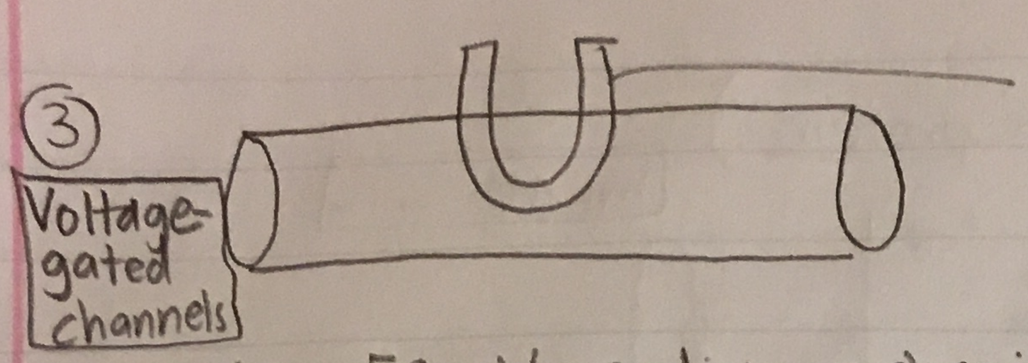
resting K^+ channel; only permeable to Potassium. So general flow back into cell.

Let's say initial membrane potential is 0 V. K^+ flows out, so outside of cell becomes higher potential. Favorable wrt concentration but not charge!

→ concentration / electrical gradients counteract at equilibrium position.

Happens around −70 mV!

Fact: Only 1/100000 of K^+ move, so no veal change in concentrations.



Ion channel for Na^+ closed <u>until</u> cell goes from −70 mV to −50 mV

→ At −50 mV, sodium rushes in. Favorable for both charge and concentration.

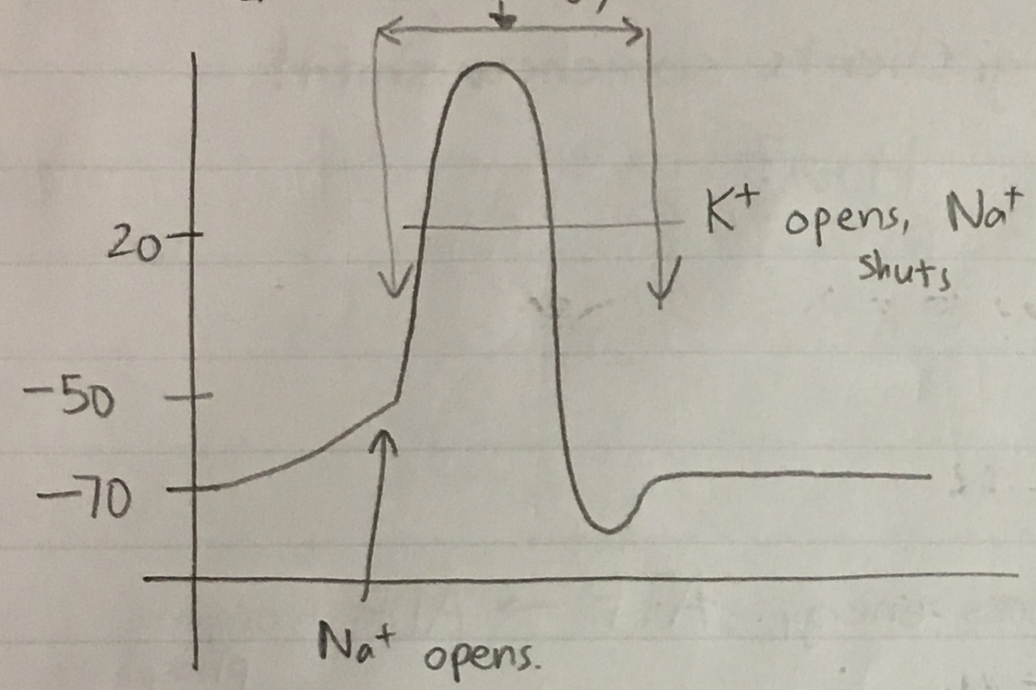
↳ Stops when cell suff. (+) that it's electrically unfavorable

↳ this is at 50 mV!

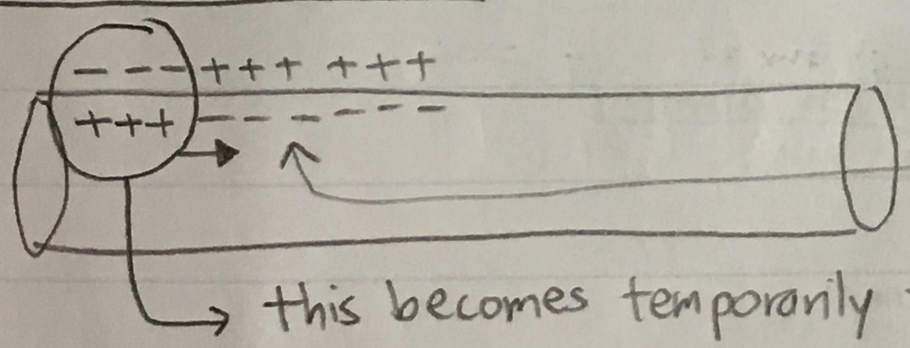
<u>Now what?</u> Around ~20 mV, K^+ voltage-gated channel opens; K^+ rushes out.

→ Na^+ shuts, remains inactive for ~1 ms.

So in 1 millisecond,



<u>So how is this transmitted?</u>



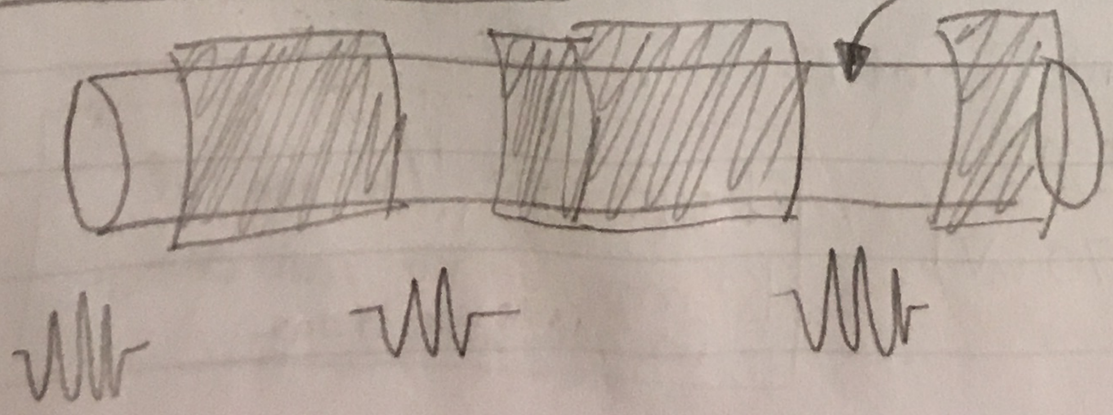
→ this becomes temporarily +50 mV. That moves this, which gets the charge up on the next patch. → <u>action potential!</u>

<u>Question</u> - why doesn't this happen in reverse? The sodium channel inactivates!

→ <u>Unidirectional propagation.</u>

Often axon has an insulator → <u>myelin</u> wraps itself around.

<u>faster propagation!</u>



Transmission leaps ~100 times faster!

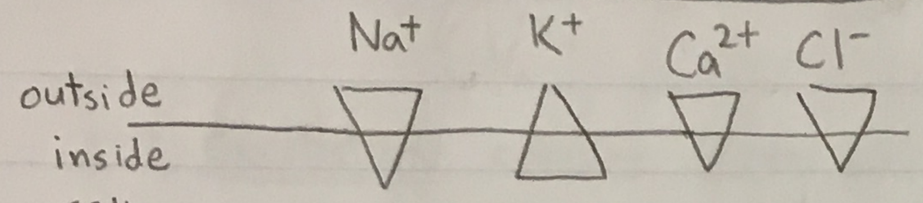
7.012

Lecture 29

11/26

Last time:

- Action & membrane potential
- To balance this out, ionic concentration gradients

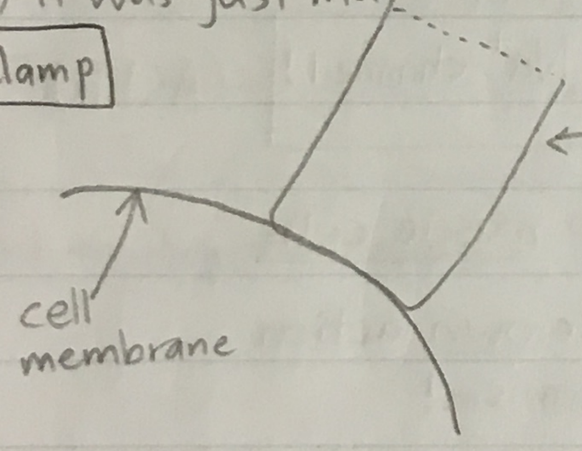


- Resting K^+ channel
- To start action potential, take voltage-gated Na^+, K^+ channels
- Trigger new action potential by dragging charge.

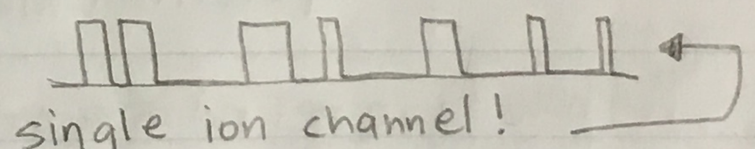
How did these first get discovered?

At first, it was just math, but... what's a way to actually observe the channels?

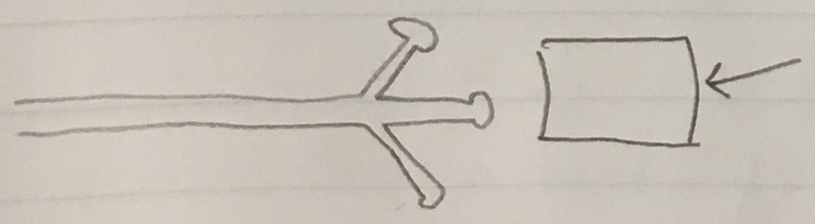
Patch clamp

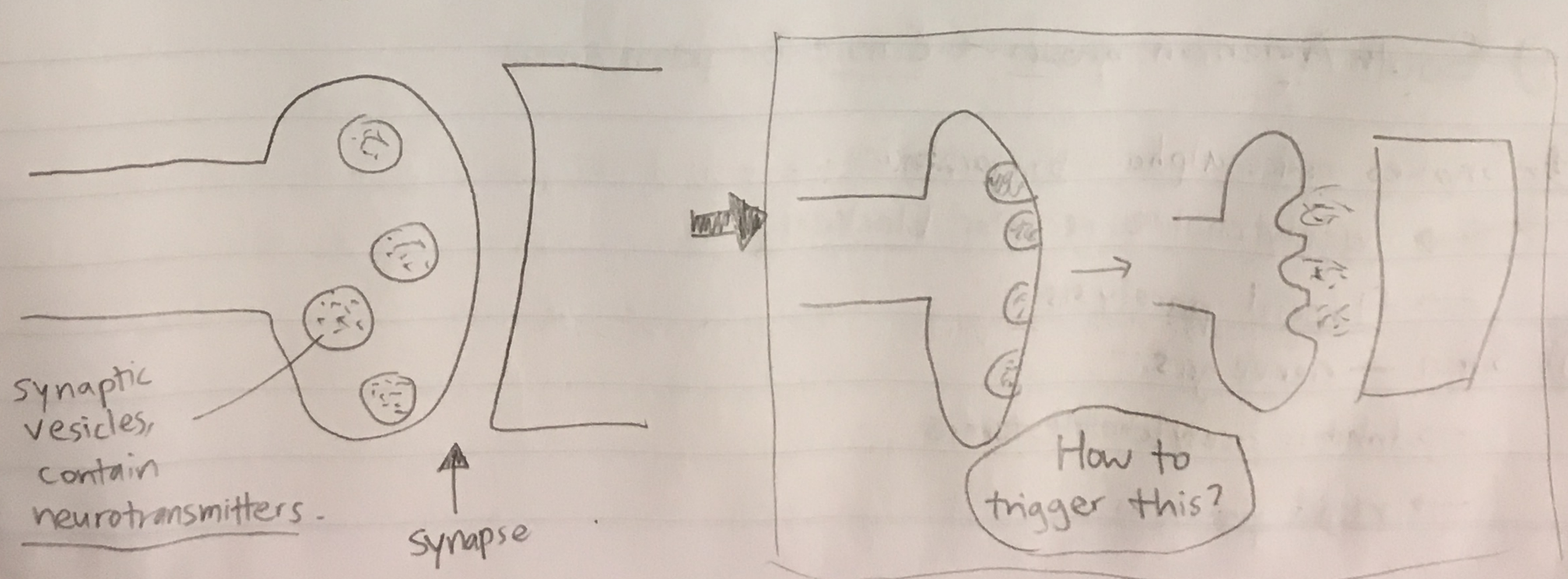


★ rip it off — see current through this piece.



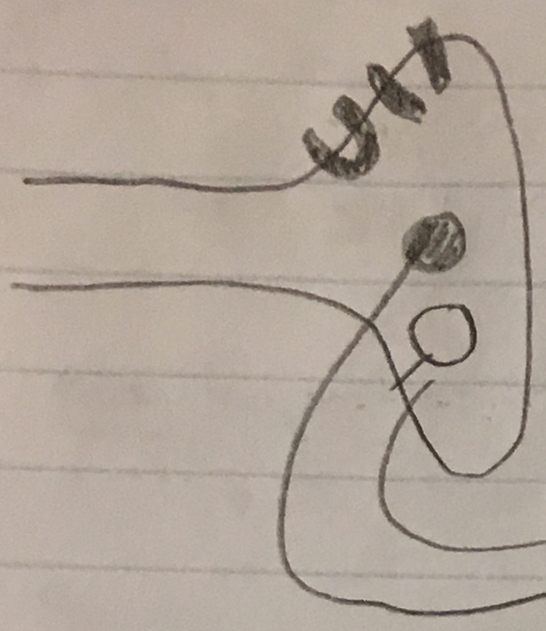
Next, how do we signal to the next cell?





- Use a voltage-gated channel.

At +50 mV:



★ Ca^{2+} has 2000x higher concentration on the outside.

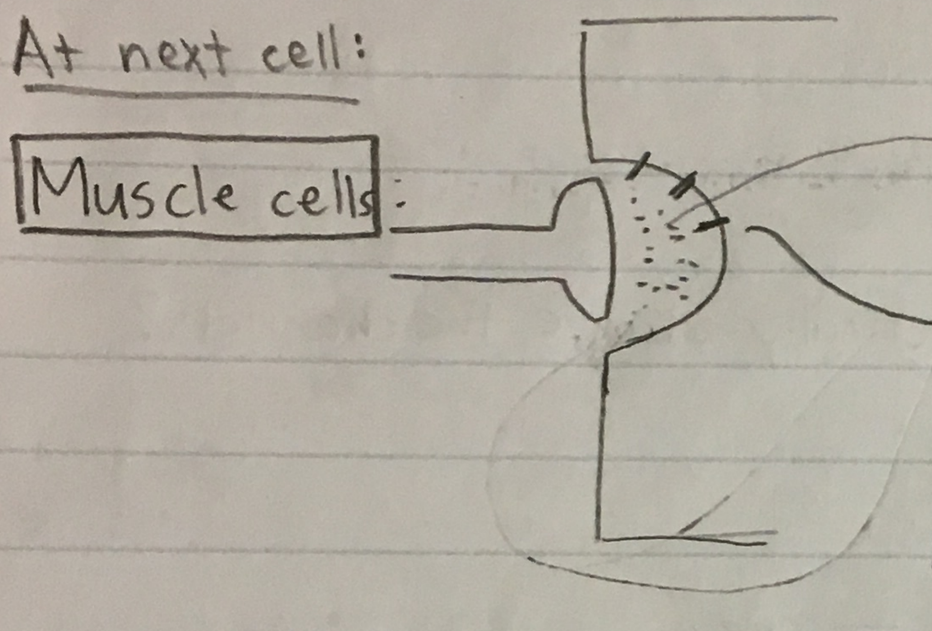
→ this vesicle is tethered by synapsin

→ here's a Ca^{2+}-dependent protein kinase

once Ca^{2+} binds, that tether is let free!

At next cell:

Muscle cells:



the neurotransmitter is acetylcholine (ACh).

→ cell has ACh receptors.

→ ligand-gated Na^+ channel! (AChR)
(ACh)

So Na^+ flows into muscle cells.

→ these have their own action potential mechanism!

But how do we get rid of a molecule?

Break it down w/ enzyme acetylcholinesterase.
(AChE)

What about toxins?

(e.g) Puffer fish → tetrodotoxin.

→ blocks voltage-gated Na^+ channels.

(e.g) South American groups - Curare for poison arrows

(eg) Snakes make alpha-bungarotoxin

→ acetylcholine receptor blockers

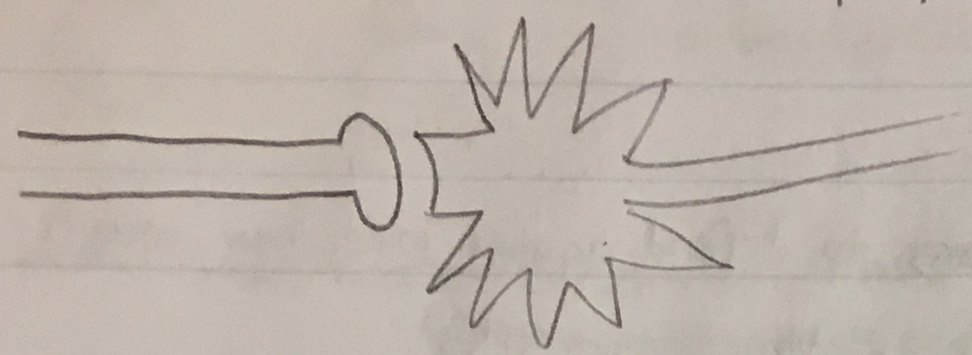
→ flaccid paralysis

(eg) Sarin — nerve gas.

→ Inhibits acetylcholinesterase

→ rigid paralysis

What about neuron–neuron synapses?



Different neurotransmitters this time.

→ <u>excitatory</u> causes V to go up, <u>inhibitory</u> causes V to go down.

↓ e.g. glutamate (Na^+ flows in)

↓ other chemicals (Cl^- flows in)

→ Causes V to change by ±3 mV, but there's no action potential mechanism in the dendrites.

→ Instead, our goal is to shift axon to -50.

→ Tree of dendrites is integrated

↳ charge changes in some way. Analog computer of thousands of signals!

Hard to describe even 1 dendritic tree.

7.012
Lecture 30

11/28

- Exam 4 on 12/3 (Monday)
- P-set 7 due next Friday

Immune system — network of cells, tissues, & substances to defend against infectious agents
→ pathogenic bacteria — cause disease, viruses, parasites.

- Goals: ① keep out, ② destroy, ③ neutralize, (④ coexist)

Lines of defense

in order

- Skin and mucosal barrier
  → both physical and chemical
  → (eg.) stomach acid, enzyme (lysozyme) in bodily fluids

- Innate immunity — cells that initiate early, generalized response.

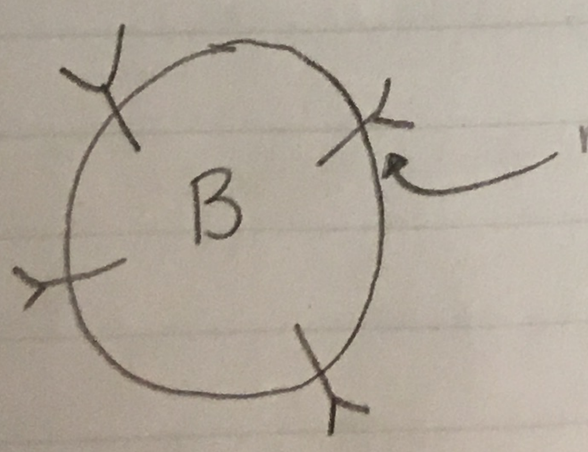
- Adaptive immunity — cells that initiate stronger and more specific response
  → Humoral response — consists of soluble substance → **B cells**
  → Cellular response — T cells

★ Bone marrow contains hematopoietic stem cells
  ↳ turn into all kinds of blood cells!

main focus

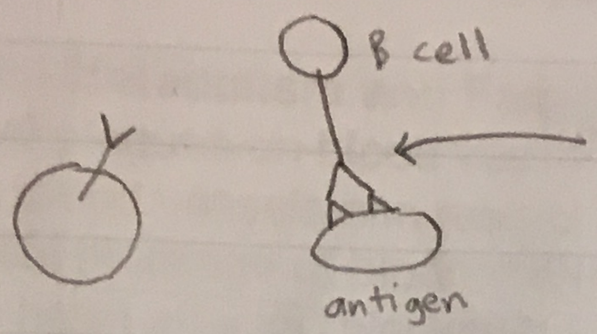
B cells

- make antibodies that bind to antigens. ← foreign substance



Each B cell makes one type of antigen.

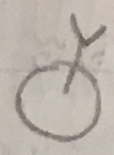
- Blood cells have lymph nodes — where cells meet to activate.



activation when antibody binds to a bacterial protein

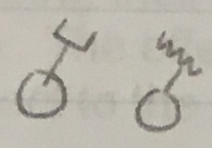
→ Make more B cells of this type (division, proliferation =

↳ lots of ER, secrete soluble antibodies



differentiate into **plasma cells**.

↳ or into **memory cells** stays for years in bone marrow

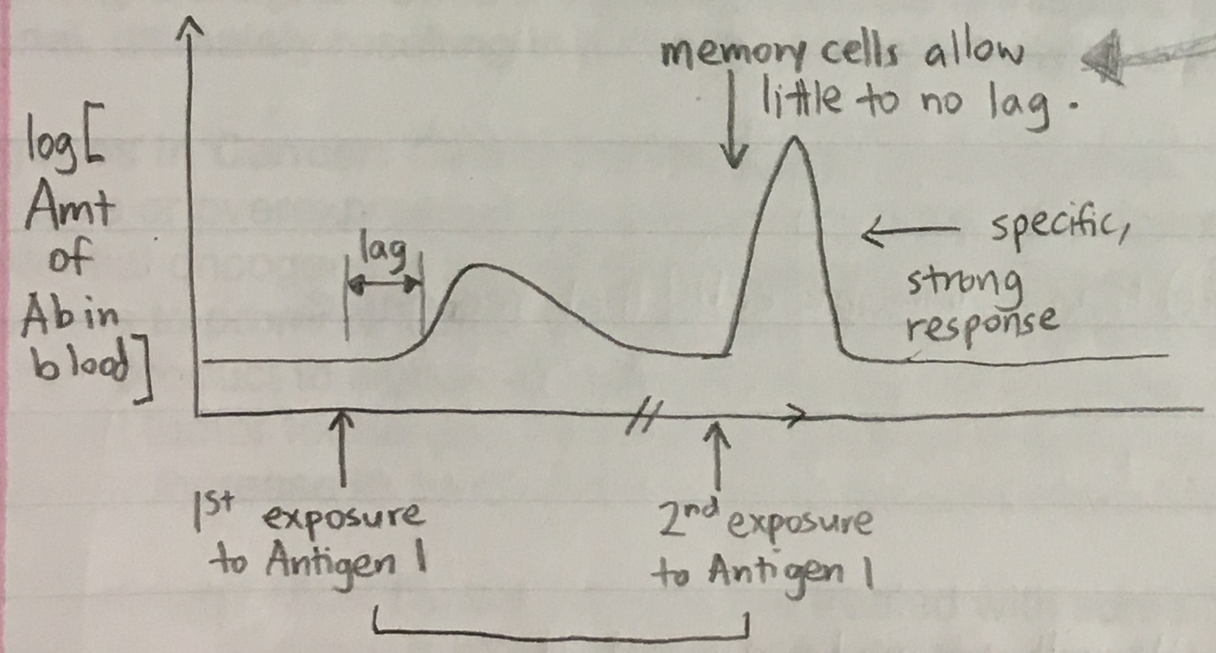
→ Other types  called naive B cells.

What happens when antibody (Ab) binds?

→ Bacteria can't bind, so target is neutralized

→ Alerts phagocytic cells to engulf bacteria

→ Macrophage recognizes antibody.

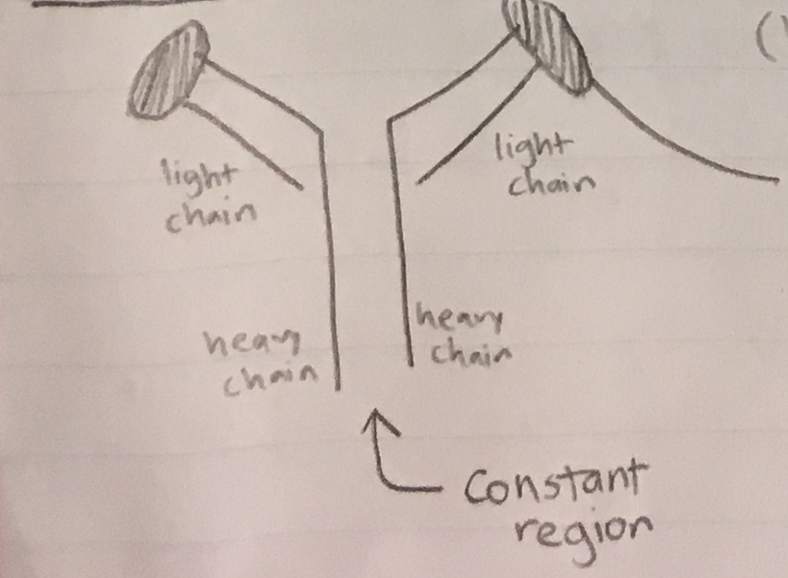


this is why we use vaccines — injecting dead or attenuated virus/protein to get memory cells.

* Why flu vaccine every year?

→ it mutates!

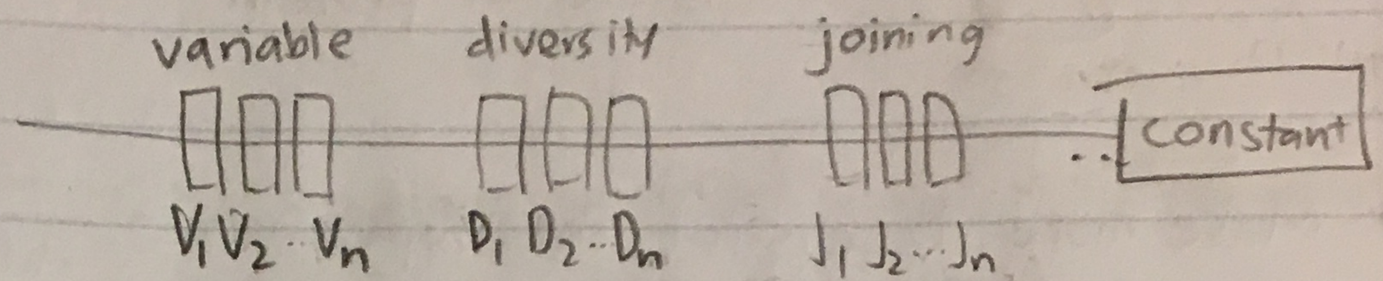
Structure of antibody — 2 heavy, 2 light chains (heterotetramer)



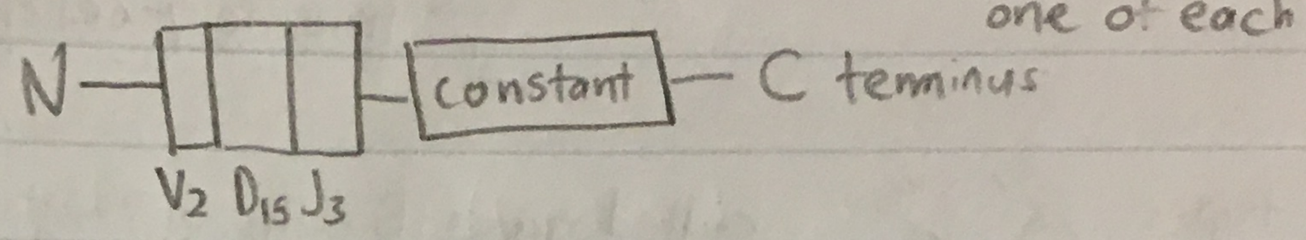
this is where antigen binds: variable region

Antibody diversity — only 20000 genes? How to get mill

(e.g) heavy chain locus (in genome)



① During development, DNA rearrangement — bring together one of each



★ similar happens w/ light chain (VJ)

② Also junctional diversity. — sequence variations introduced.

③ Somatic hypermutation — enzyme does more mutation

→ $\underline{10^{14}}$ combinations — recognize molecule of <u>any</u> type.

What about self vs non-self?

- During early development, <u>clonal selection</u> - immune cells that recognize self are eliminated.

In research and medicine

- Inject any protein into organism (rabbit or mice) → T cells will be made against it.
  - → collect blood serum, contains polyclonal antibodies
    - ↳ monoclonal antibodies from B cells fused to immortalized B cell = hybridoma

7.012

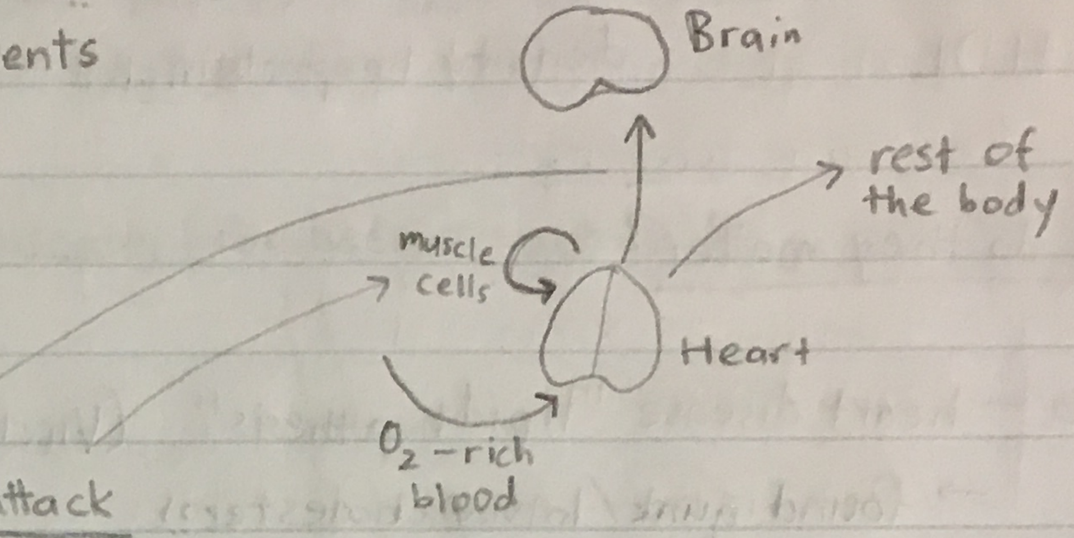
Lecture 31

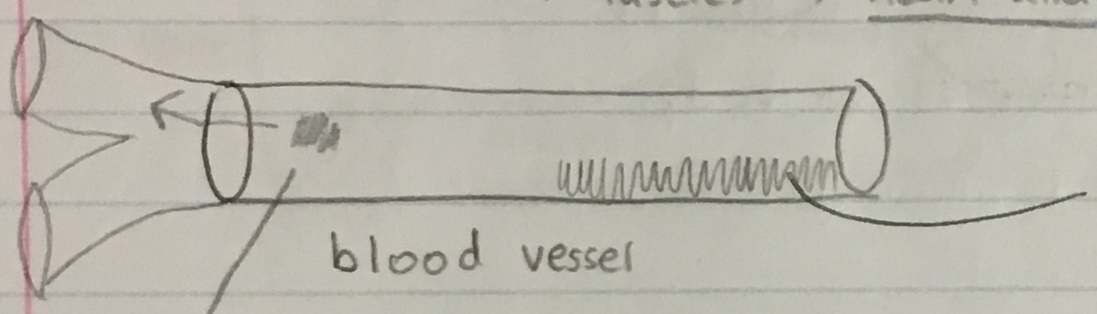
11/30

Title: Rational Medicine: Familial Hypercholesterolemia

Heart Disease

- Purpose of heart: to pump blood
  - → Provides O_2, hormones, nutrients
  - → Pumps red/white blood cells
  - → Pumps away waste
- If blood vessels are blocked, bad.
  - → @ brain → called a stroke
  - → If @ heart muscles → heart attack

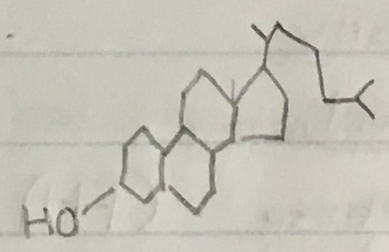




Atherosclerotic plaque:
contains cholesterol / proteins / lipids

→ bad if chunk breaks off and clogs a smaller vessel (or if main one clogs)

What is cholesterol?



- Hydrophobic molecule
- Often attached to phospholipid tail
  - → cholesterol ester (CE)
- Purpose:
  - → Makes up half of cell membrane!
    - ↳ Provides fluidity
  - → Biochemical precursor
    - → Steroid hormone
    - → Vitamin D
    - → Bile acids – emulsify fats
- Source:
  - → Partially from diet.
  - → Also synthesis in body!

general pathway:

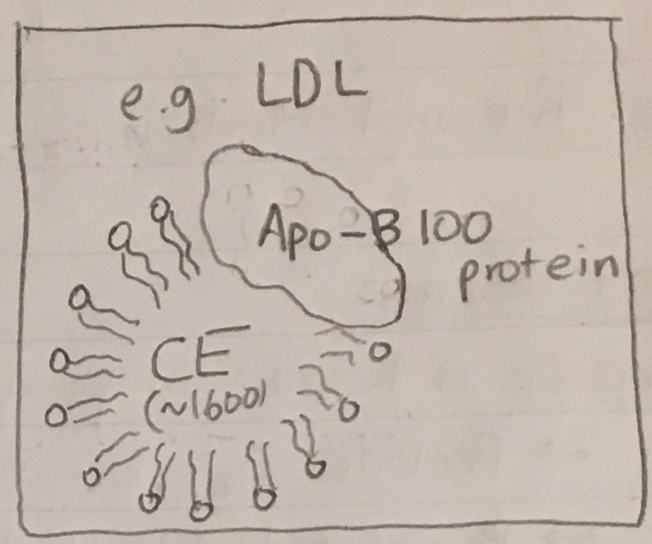
acetic acid

$CH_3COOH \longrightarrow$ [HMGCoA → Mevalonate] → Cholesterol

HMGCoA Reductase
(first committed step, regulation point)

Lipoprotein particles

VLDL
LDL
IDL → high, intermediate, low, very low density lipoprotein particles
HDL



transport system

Why do they matter?

1856 – heart disease "lipid hypothesis" (Virchow)
→ found gunk/lots of cholesterol
1913 – fed rabbits high cholesterol
→ symptoms ~ atherosclerosis
1950s – LDL vs HDL tracked
↓ correlated w/ risk of heart attack
↓ negatively correlated!

Genetics of Cholesterol Levels

Familial hyper (lots of) cholesterol emia (in blood), (FH)

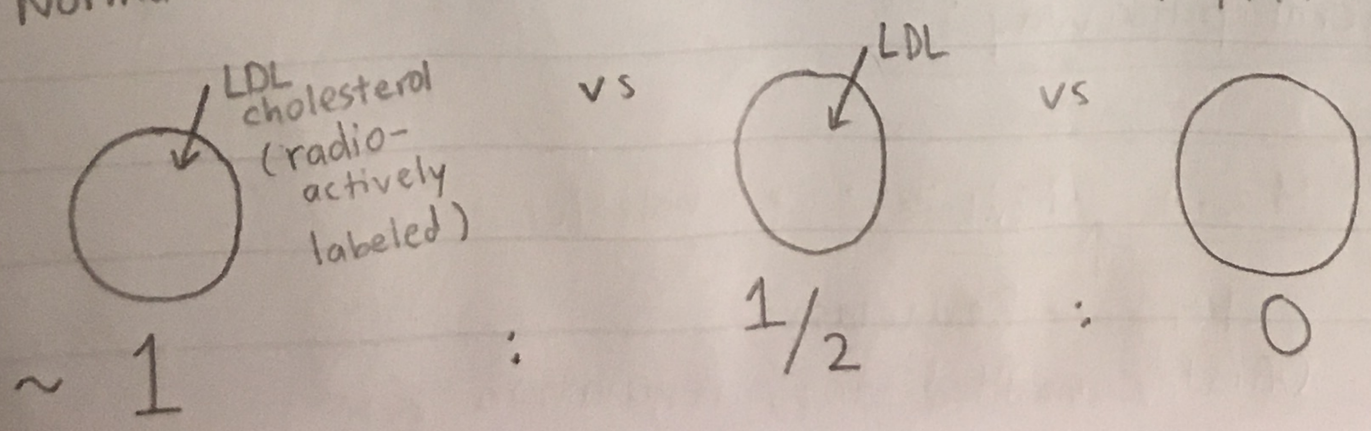
Look at	normal (+/+)	heterozygotes (FH/+)	homozygotes (FH/FH)
LDL	100 mg/dL	250 mg/dL	>600 mg/dL
Age of heart attacks	normal age	10–20 years earlier	teenagers

1 in a million people

so allele freq. is 1/1000
→ freq of heterozygotes = 1/500.

Fibroblasts:
Look at skin cells in petri plate.

Normal cells vs FH/+ vs FH/FH

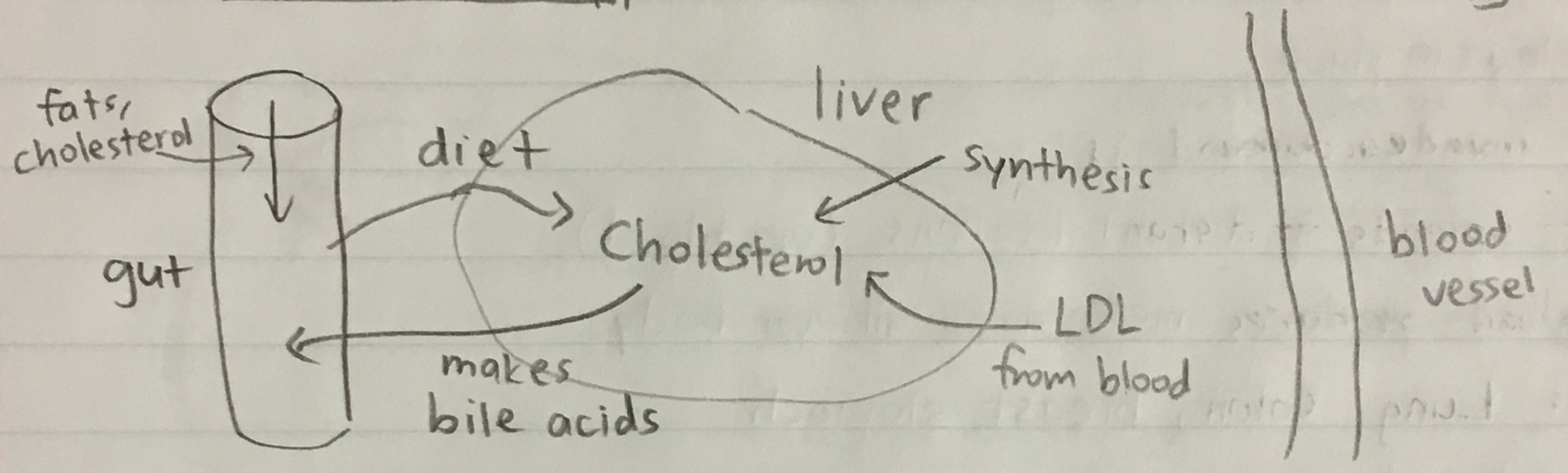


there exists
↓
Hypothesis: ∃ LDL receptor. Helps take up LDL (e.g. liver)!

Biochemical Regulation

If LDL levels are down: → make more HMGCoA reductase
→ more LDL Receptors.

→ So rational therapy for heart disease (for heterozygotes):



Strategy 1 Eat less cholesterol.
→ Has some effect: LDL ↓ ~10%

Strategy 2 Try to deplete cholesterol.
→ Bile acid binding resin → forces body to make more bile acid
→ Still only ↓ 20%

Strategy 3 HMGCoA reductase inhibitor → called statins
→ Decreased LDL by 60%!
→ Taken by 35 million people.

7.012

Lecture 32

11/5

Title: Rational Medicine: Cancer Targeted Therapy

(By the way, exam 4 posted, p-set due Friday)

What is cancer?

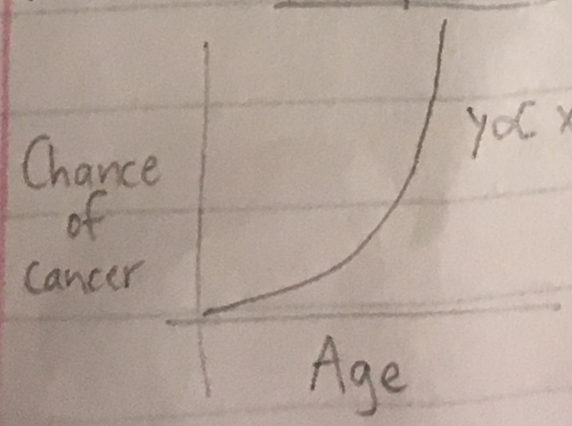
- Cells that grow and divide <u>withaut limit</u>
  - ↳ create tumors
  - → "Neoplasm" = name for abnormal growth of tissue
    - → "Benign" – stays in place
    - → "Malignant" – invades surroundings
    - → "Metastatic" – spreads to distant locations (via blood)
      - ↳ this is bad: produces <u>metastases</u> all over body.
- <u>Common cancers</u>: Lung, Colon, breast, stomach
  - ↳ also skin, prostate, ovarian, etc.
  - → ~1.2 million new cases per year; half cause death.
    - ↳ account for ~1/4 of all deaths
- <u>Cancer grows exponentially</u>
  - ↳ first time you can see on X-ray: 10^8 cells
  - → Palpable: 10^9 cells
  - → Death ~ 10^{12} cells

~1/2 the time it took to get to first diagnosis

Why does cancer happen?

- Mutations.
  - → 10^{16} cells, $3 \cdot 10^9$ bases in genome, 10^{-8} mut/base
  - ↳ 30 mutations / base
- Increase chance by UV (sunlight), smoking, charred beef, working at Chernoby!
  (skin) (lung)

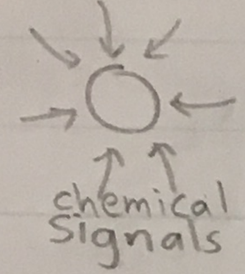
★ Takes <u>multiple mutations</u> to become cancer cell.



↳ <u>Also</u>, can inherit alleles that are likely to lead to cancer

Why don't all cells grow like this?

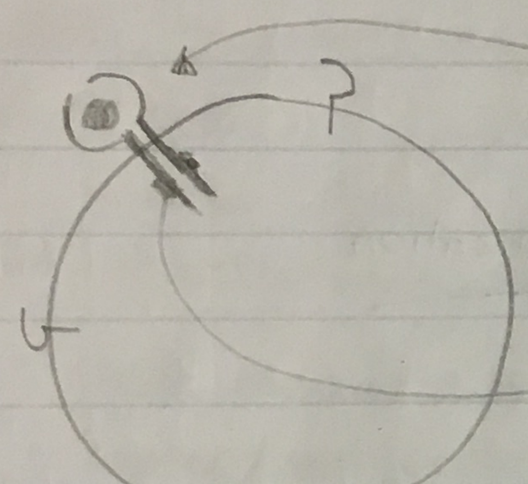
Regulation



growth factors tell a cell to grow

(EGF, FGF, NGF, PDGF, etc.)

How does a cell detect these? Receptors!



presence of growth factor makes it favorable to dimerize.

↳ they're close to each other inside the cell, too.

→ enzyme activity:

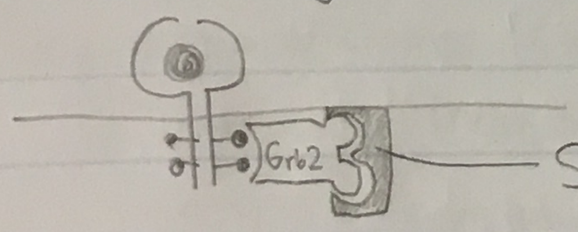
protein kinase phosphorylates Tyr residues.

↓

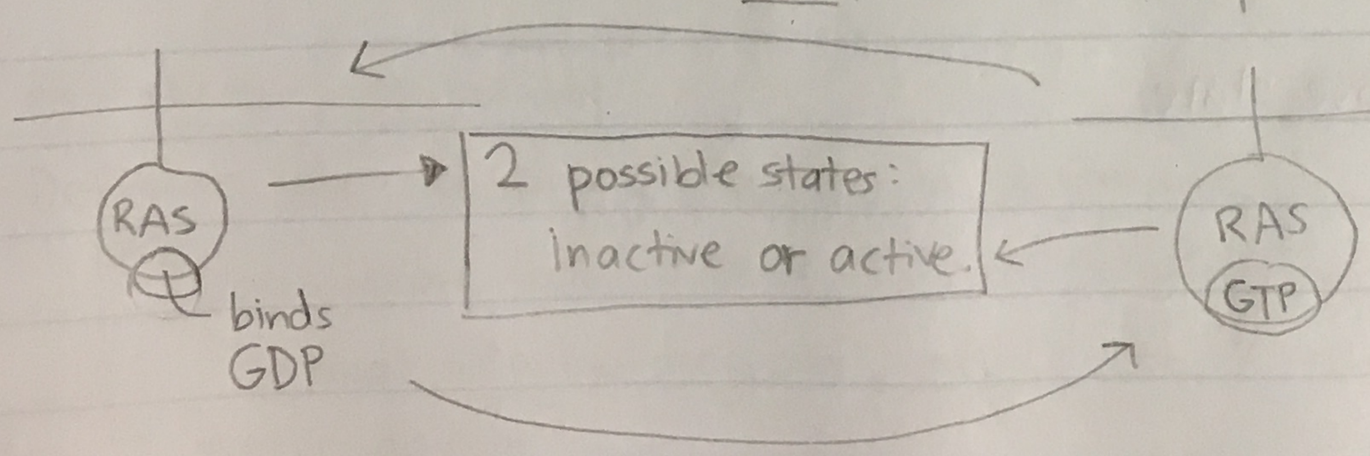
they can phosphorylate each other!

→ Phosphate groups change shape of protein.

↳ changes the binding patterns — adapter protein Grb2.



Sos protein → then binds to / affects RAS, a membrane protein.



RAS is a GTP-ase: catalyzes reaction to make itself inactive.

☆ There are proteins that can speed up GDP → GTP.

↳ GEF (guanine exchange factor); e.g. Sos!

(Also GAPs: GTPase activating proteins competing)

So active RAS undergoes conformational changes

↳ binds to RAF, which becomes phosphorylated.

→ RAF also protein kinase → phosphorylates MEK.

↳ MEK activates ERK, which phosphorylates transcription factors, other things

↳ cell growth!

☆ Reason for this: amplification.

So what mutations cause cancer?

- Dimer stays together w/o GF (growth factor) — constitutive dimerization
- Cell produces too much GF

☐ Knocking out RAS or Sos won't work

☆ But we could have RAS stuck in active state. Very common!

↳ unable to hydrolyze

- Kinases become constitutively active

How to cure?

- Inhibitors for enzymes that are always on. (e.g. RAF)
- Why not inhibitor for RAS?

↳ It would be stuck in active state.

7.012

Lecture 33

12/7

- Final is on Wednesday 12/19 at 9am
- 5.302 – Intro to Experimental Chemistry during IAP!

Title: Using Biology to Save the Planet

Superbugs – today, helping save the planet

What does the planet need to be saved by?

- Climate change / greenhouse gases
  - Risk by 2040
  - ~100000x weight of Empire State in warming emissions
    - Oil in Alaska → spills
    - Water / soil contamination

Microbes can "eat" / metabolize greenhouse gases!

- Also hydrocarbons from crude oil (e.g. toluene)
- Hazardous waste (e.g. perchloroethylene = PCEs) → manmade dry-cleaner solution.

- Key idea – microbes / their enzymes replace chemicals in industrial processes / replace energy-consuming processes!

Let's review **metabolism**.

- Defn – life-sustaining processes in organisms.
  - (eg) conversion: food → energy to run cellular processes
  - Also made into building blocks of life
  - Waste disposal.

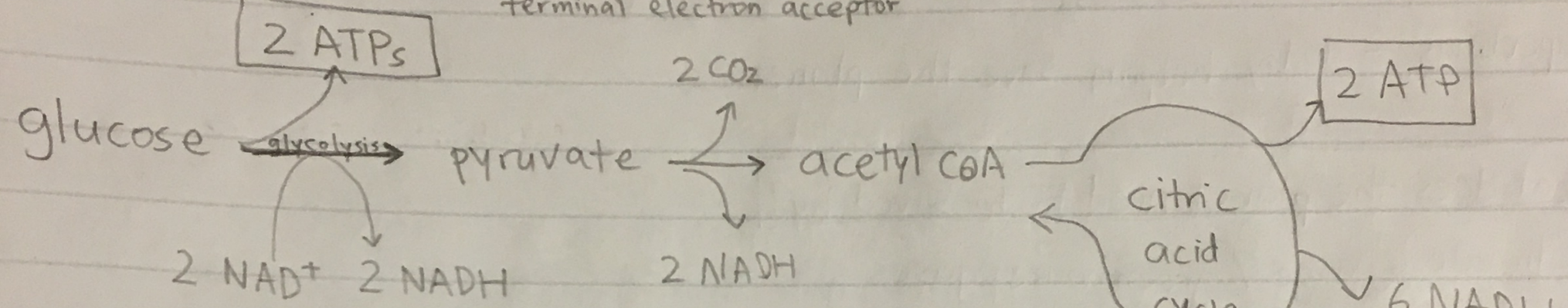
2 types:

Catabolic = breaking down (to make energy),

Anabolic = synthesis, makes energy

Review: Metabolism of glucose with O_2 to make ATP

- Glucose has stored energy.
  - ↳ Redox reaction: O_2 reduced, glucose oxidized
    - O_2 ↑ terminal electron acceptor

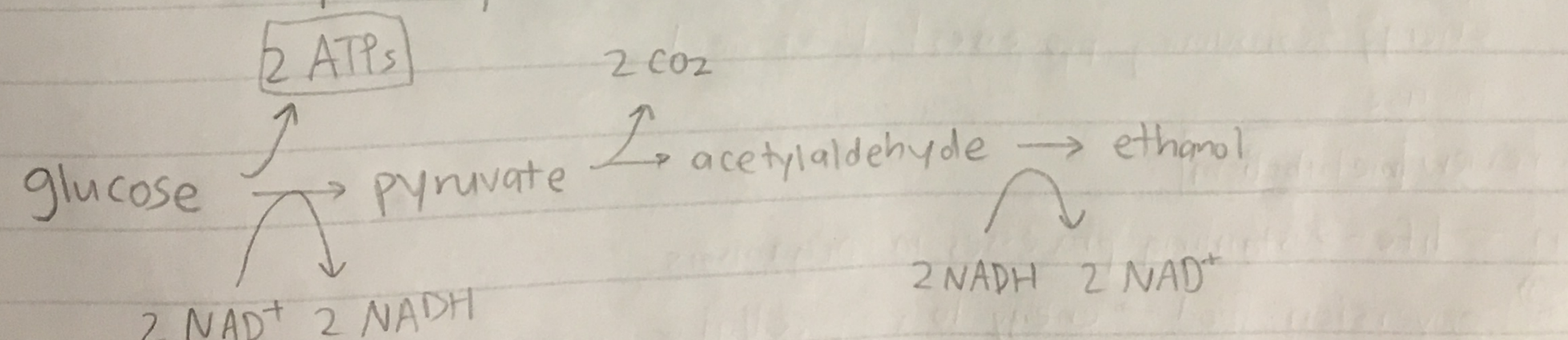


- There is a charged gradient!
- What if there is no mitochondria?
  - ↳ use cell membrane!

⟹ electron transport chain w/ proton gradient make ATP (×28), $O_2 \rightarrow H_2O$.

What if we don't have O_2 as terminal electron acceptor?

- Fermentation to ethanol
  - ↳ use acetylaldehyde.



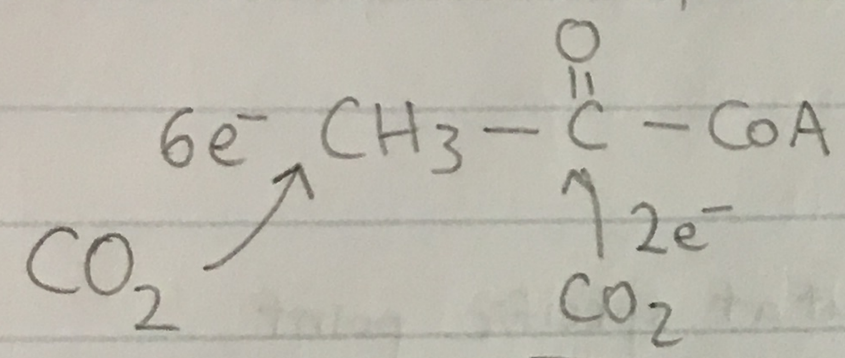
Net of 2 ATPs instead of 32!

Remember: NADH is an electron carrier!

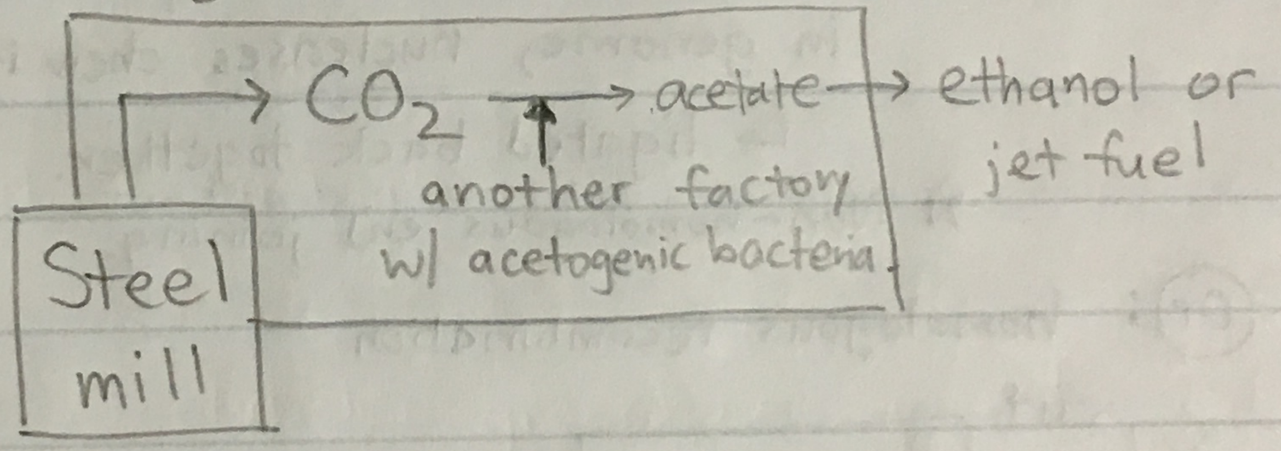
* Glucose also source of carbon!
  → stop at acetyl CoA, make fats instead
  → Can also make nucleic acids, etc.

Absence of glucose + O_2

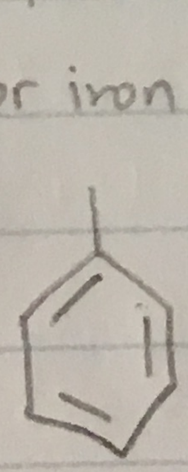
- Acetogenic bacteria (e.g. M. thermoacetica)
  → oxidize H_2 gas; use CO_2 as terminal electron acceptor.
→ Make acetate or acetyl CoA (~10^{11} tons removed / year)!



- LanzaTech



- Anaerobic hydrocarbon metabolizing bacteria (e.g. T aromatica)
  → oxidizes hydrocarbons. (e.g. toluene)
  ↳ use NO_3^- or iron or sulfur as terminal electron acceptors.



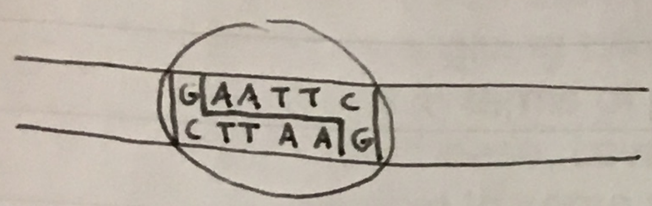
← how to metabolize this?
"unprecedented C-C metabolism"?

7.012

Lecture 34

Today and Wednesday: Science in Society

Recall: Restriction Enzymes

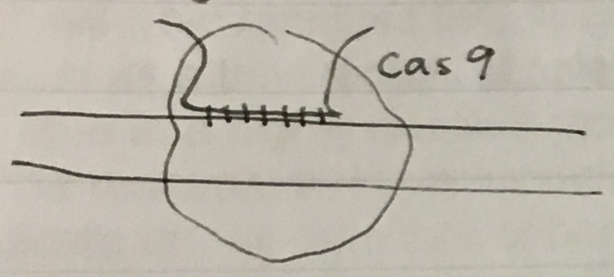


- EcoRI – cuts specific sequence
- come from bacteria
- defend against viruses

→ To protect itself, methylation.

Recall: CRISPR is a programmable restriction enzyme!

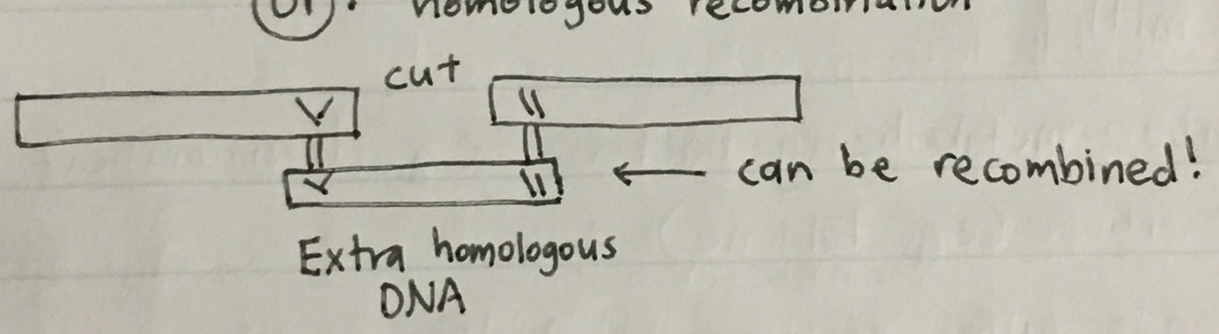
- Uses guide RNA to cut @ ~20 bases.



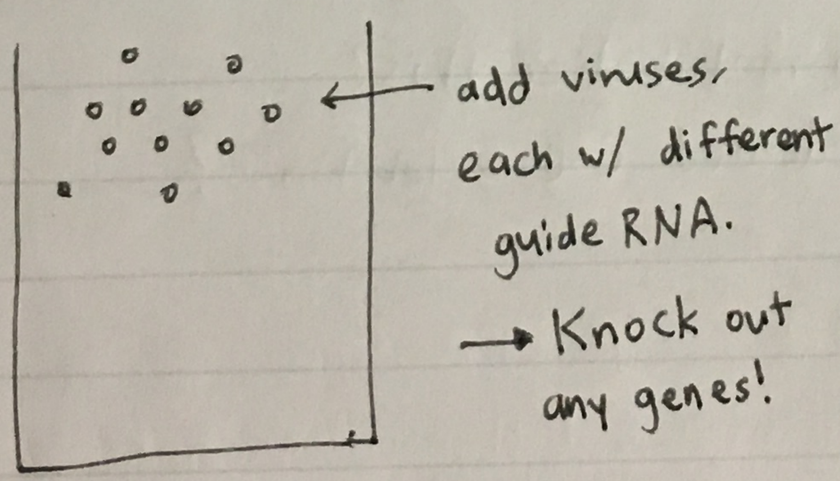
- Useful – if cut at specific point in genome, nucleases chew it back
  ↳ ligated back together.

★ Non-homologous end joining.

(Or): homologous recombination



★ This allows us to add any sequence we want!



Medical applications?

① Somatic gene editing

e.g. Dominant Retinitis Pigmentosa

→ if chew back is a multiple of 3 bases, won't knock out the gene.

→ But could we have an off-target cut? Cancer? etc

② Germline editing — done to fertilized embryo.

e.g. ~~2%~~ of Europeans have a cystic fibrosis allele
≈4%

↳ ~8% of pregnancies?

e.g. norovirus has receptor. If you're mutant for Fut2, you don't get it.

↳ Is this OK? does it do anything else?

→ increases risk of inflammatory bowel disease

7.012
Lecture 35

- Recommendation @ CRISPR babies.

Facts – what happened?

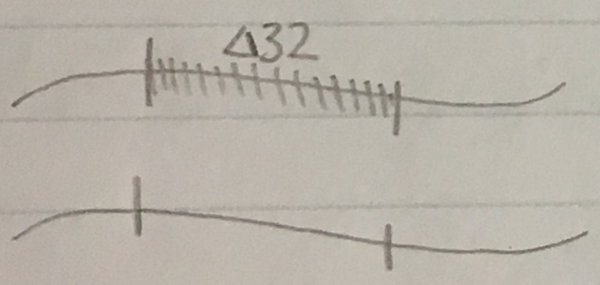
- Jiankui He – took fertilized eggs, tried to edit CCR5 gene.
  - ↳ Lulu & Nana, 2 twin girls – result of reimplanted embryo
- CCR5 – relevant to immune system (cell surface protein). HIV can use as receptor.
  - ↳ Some have Δ32 mutation (frequent in pop.)
    - Fact) HIV can't use homozygote mutatied CCR5, so it's pretty strong. Not absolute protection though
- HIV can transmit w/ father's sperm, but you can wash it
  - ↳ if mother has HIV, there are also ways to drive down HIV infection.

→ So why?

- ↳ Other effects of this gene deletion? Enhancement of some kind?
  - → Lack of CCR5 increases severity of inflammation in brain.

→ Better genes could have been chosen

How well did he do, anyway?



Lulu – heterozygote
wild type / 15 base deletion

Nana – heterozygote
1 base / 4 base deletion

This doesn't even give protection against HIV. Also 15 bp is in frame!

→ Parents were not properly informed (medical jargon)

This is a +1 instead of a -1 frameshift. No way to know effect.

- Done out of own pocket, unclear @ IRB

Another gene: PCSK9. Degrades LDL receptor

↳ Mutations in this gene are present @ Afr.-Am. population

ΔPCSK9 – homozygote → very low LDL levels

⟶ ∃ antibodies to deplete PCSK9 gene.

- Jiankui He has also made deletions in this gene. Also enhancement

Broad societal consensus?

The thing about 7.012 – gets us to the frontier of biology!