7.012 L10

9/28

Molecular Biology I

(2 min late)

Function
Genetics — Biochemistrie
Gene — Protine

2 ways at looking at chemistry + genetics
↑missed his word

How do you seperate out + grind up a unit of heredity?

Turn it from abstract concept to actual thing?

Discovering the "Transforming Principle"

↑word from middle ages
life substance or something like that

F. Griffiths studied Pneumococcus

1. Smooth-coat virlent
   ↳ die
2. Rough-coat nonvirulent
   ↳ no die, harmless

Due to a ? polymer present in 1
that prevents immune system from recognizing

Injects each into a mouse

Smooth → Dead
Rough → Alive

Heat kills smooth bacteria
↳ mouse lives

So rough non virulent + dead smooth virulent
↳ mouse dies
- both harmless individual
- can grow living bacteria out of mouse
Smooth virulent

Called this the transforming principle
(This will turn out to be DNA)

Starts fractioning
Which fraction of harmless bacteria will kill the mouse
trying to get pure molecule
This takes 13 years

Avery, McCarty, McLeod

find it in 1943

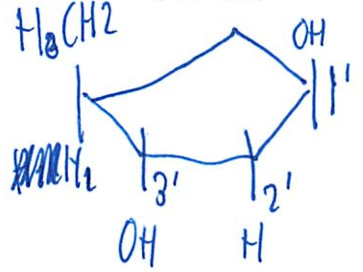
have a process that doesn't need mouse
 └ so faster

material was highly enriced for DNA

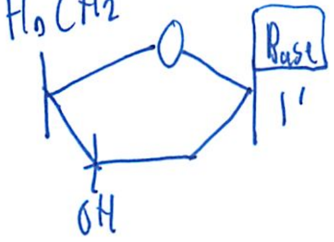
it is Thought at the time that DNA is boring

Structure of DNA

(can't really read

a) 
H₂CH2
OH
1'
H₂
3'
2'
OH
H

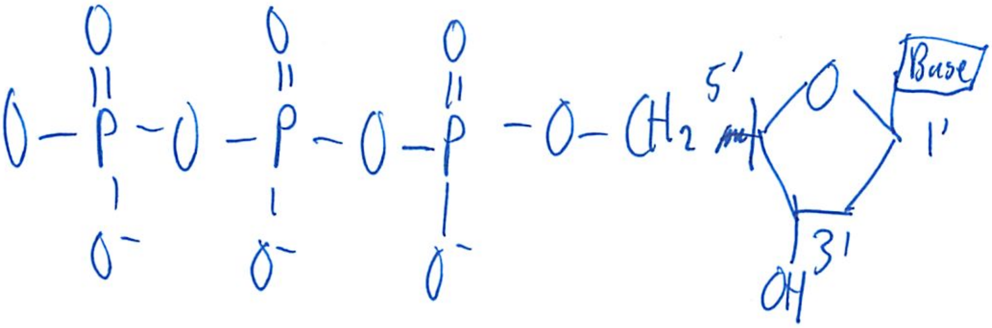
2'-deoxyribose

b) Bases 
H₂CH₂
O
Base
1'
OH

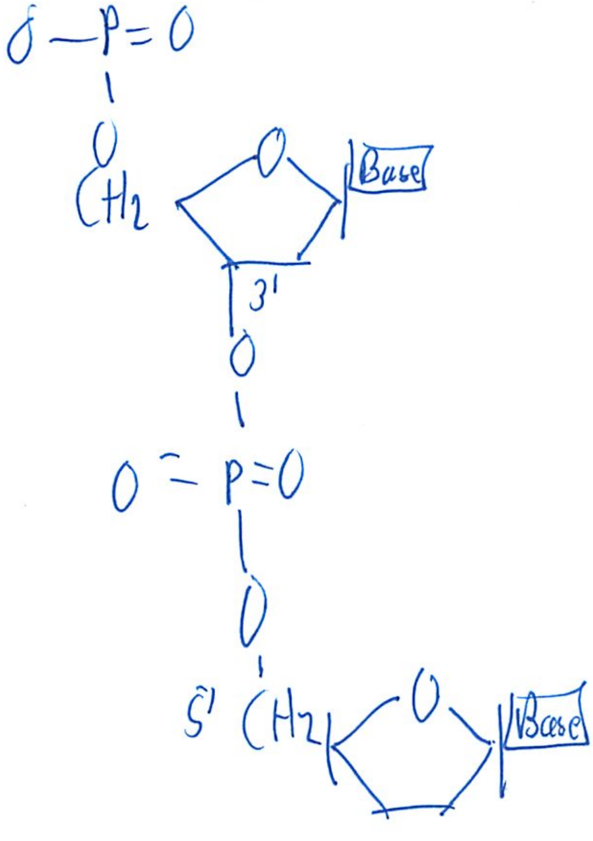
adenne (A)
guanne (G)
thymine (T)
cytosine (C)

④

c) Triphosphate   good at suppling energy

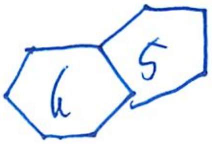

O
O
O
O — P — O — P — O — P — O — CH2
O⁻
O⁻
O⁻
5'
O
Base
1'
3'
OH

d) Sugar-phospate back bone

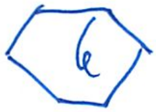

O — P = O
O
CH2
O
Base
3'
O
O = P = O
O
5' CH2
O
Base

e) Bases

Purines A
G



Pyrimdines T
C



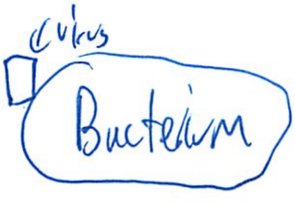
C, N, O, H atoms

Everyone thought DNA was boring
just structuring
Same thing over + over

Plus during World War 2
But some people noticed

Bacterial Viruses

Bacteria get viruses too!



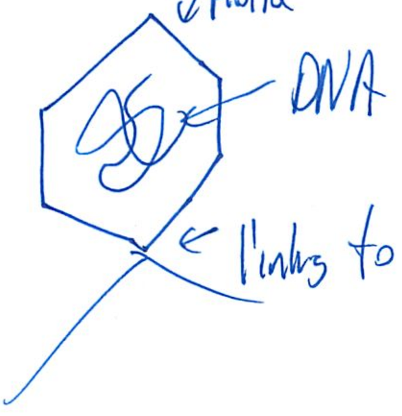
$\text{Bacteriao}\underbrace{\text{phage}}_{\text{eat}}$

After 20-60 min

Bacteria burst open + die

More bacteria viruses spill out

Somehow the Bacteriaphage injects something
which has the instructions for more viruses



Hershy - Chase

radioactivly lable viruses
- └ grow cells nearby radioactive element
- Want something only in DNA
  - O - no, both
  - N - amino acids has - no
  - C - certainly not
  - H - no
  - P - no amino acid has - Perfect for DNA!

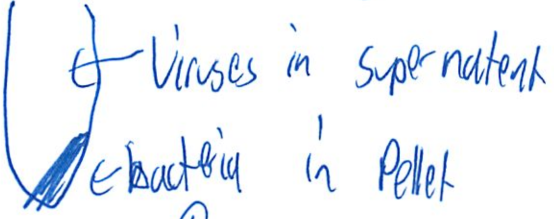
Systine - has sulpher

So ^{32}P or ^{35}S to radio lable

Now need to mix
- shake very hard
- kitchen blender

→ to knock the viruses off the outside of the cell

Cells more dense
so use centefruge


Viruses in supernatent
bacteria in Pellet

See a lot of phosperous and a bit of sulpher
1952 - people more ready

DNA must be transforming principle

Linus Pauling looks at it

Also amatures: Watson + Crick

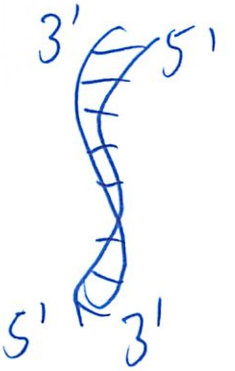
The Double Helix

Read the book *The Helix* by Watson

↳long + complicated story

~~Frank~~

Pauling publishes something but its wrong



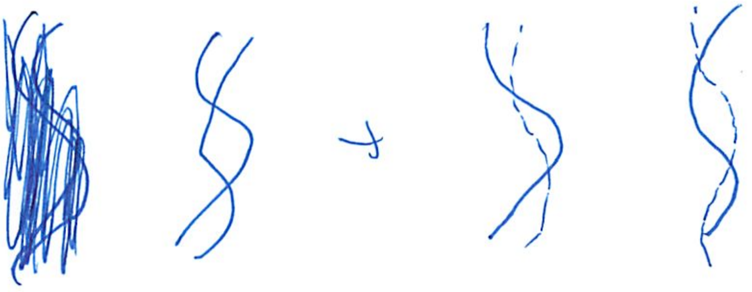
in every textbook

A-T
C-G fit

"It has not escaped our notice"

But it has not been tested

Testing the Double Helix



Semi conservative replication

Must prove 1 strand new + 1 strand old

So make Phosporous - radioactive
See only in old strand

They grow bacteria w/ heavy nitrogen
↑ ^{15}N

Meselson + Stahl

Then swamp it in ^{14}N - ~~abundant~~ common nitrogen
Let it grow
isolate the DNA w/ density centrifugation
find 1x14 and 1x15

Then next generation 15/14 14/14

So replication method proved

* Now used biochemistry to understand genetics

7012
Molecular Bio 2

10/3

Unifying molecular bio + genetics

Watson + Crick: DNA

2 redundent strands

Today: Genes and Protiens

Central Dogma

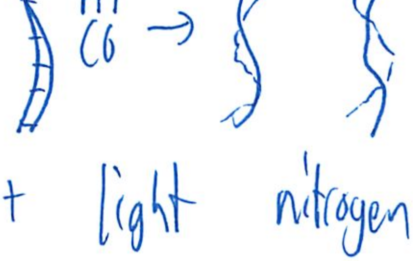
$$\begin{tikzcd} DNA \arrow[loop left] \arrow[r, "transcibed"] & RNA \arrow[r, "translated"] & Protien \end{tikzcd}$$

RNA └ similar to DNA
but diff in important way

Called dogma for historical reasons

Replication

$$\begin{tikzcd} DNA \arrow[loop left] \end{tikzcd}$$



heavy + light nitrogen
Centrafuge → measure densites

②

Some enzyme fills in the dotted line

↳ polymerizes DNA

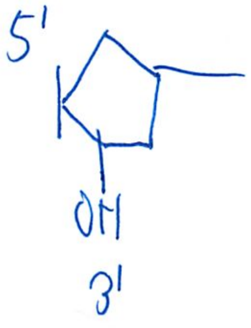
Primer strand

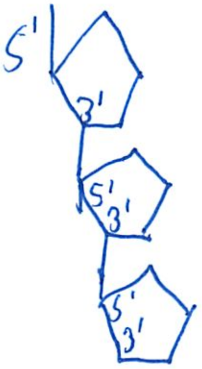
5' $pT_pA_pC_pG_pT$ 3' ← same but opposite dir

(p = sugar phosphate bond)

3' $A_pT_pG_pG_pC_pA_pT_pT_pG_pG_pC_pA_p$ 5'

Template Strand





Very structured

So he wanted to see if he could build this

needs

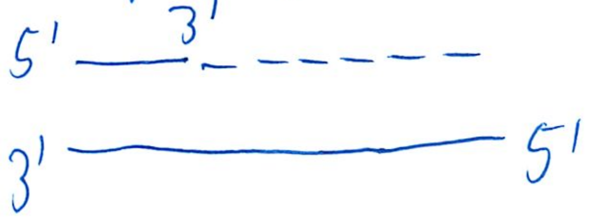
ppp A ppp T free
ppp C ppp G nucleotides

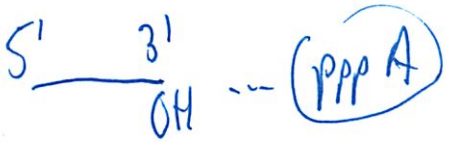
Found enzyme that could extend primer string

↳ DNA polymerase

③

DNA Polymerase



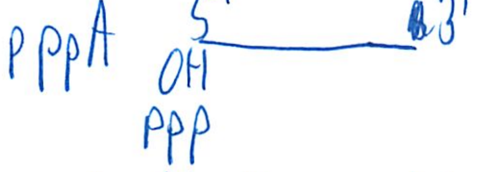


Cleaves off 2 phospahates

joins it on

it always goes 5' → 3'

So why not the other way?

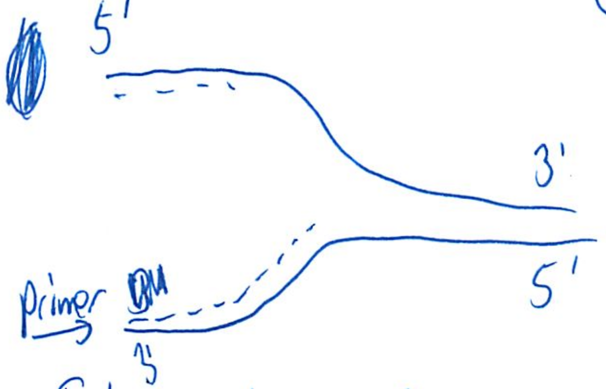

pppA
5'
OH
ppp
3'

So join it up w/ the triphosphate

(missed exactly why)

④

Arthur Kornberg

① 

↑ he used a primer

in real life need sep enzyme to make primer

↳ primase

top Must keep adding primers

bottom just keeps extending

tie together → Lygate

↳ enzyme: lygase

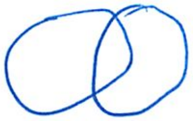
---

②

DNA is very tangled

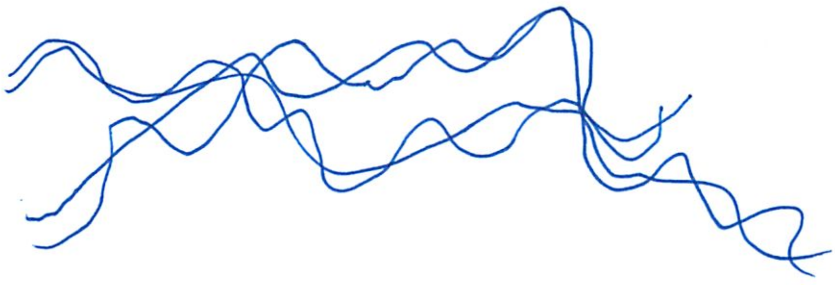
2 double helices wrapped around each other

Topologically → if 2 circles wrapped around each other → can't seperate



Need enzyme to cut the DNA
└ topoisomerase
pass it around
and close it back up

topoisomers same structure wrapped up differently



Actual biochemistry is very difficult

③ Fidelity of Replication

not impossible, but energetically less good

right guy has best H-bonds Δg

but wrong can be squeezed in Δg

Δg for wrong base $\rightarrow k_{eq} = 10^3$

↑ a thousand times less likely

- So 1/1000 mistakes
  But billions of pairs
  So still a lot of errors

Need something better

5' → 3' Polymerase

Can also run background

3' → 5' exo-nuclease

So can put in and take out bases

More likely to run backwards if last base = error

"Proofreading"

Now $1/10^5 - 1/10^6$ mistakes

Better

but still thousands of mistakes!

Then someone else checks

enzyme

-detection

- repair

anything that is mismatched → I'll clip it out

⑧

$1/10^8$ errors

but genome 3 billion $(3 \cdot 1/10^9)$

So 30 differences from dad and mom each

This is how people ~~evolve~~ evlove

2 people in this room are ~~hetro~~ ~~homo~~ hetro zygas
for polymerase

What if they lose good copy

High risk for heritiary colon cancer

④ Speed DNA Polymerase → 2000 nth/sec
??

Turns out DNA Polymerase is not used by ecoli

Genetics → what happens inside cell

Org – mutation

See if can't replicate

Found 1 ecoli strand w/o DNA polymerase

turns out that polymerase was a repair

actual one we call #3

So look at a mutant

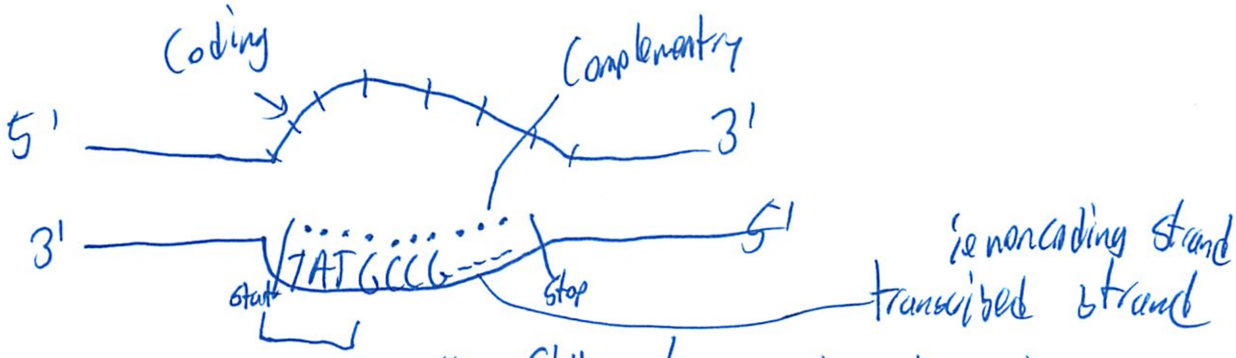
Transcription

DNA → RNA

What does it mean gene for something?

portion or region of chromosome

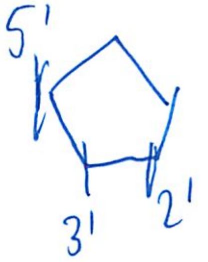
(10)


Coding
Complementary
5'
3'
3'
5'
TATGCCG
start
stop
ie noncoding strand
transcribed strand

enzyme will fill in complementary bases

Does not Use DNA → uses RNA

RNA · Sugar DNA 2' deoxyribose
RNA ribose



Uses 2' → not much of a differences

· Base DNA GCAT
RNA GCAU
↑ exact like T
except for methyl group
does not matter in this class

How does it do this?

Looks like DNA replication

But

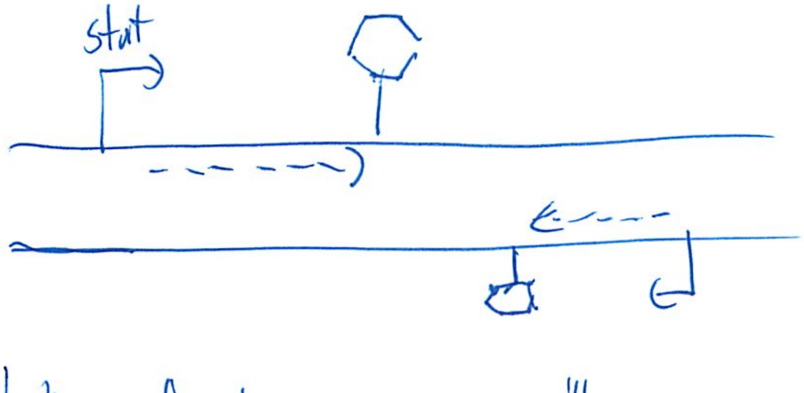
- Single strand copying
  not both strand

- no primer needed

- RNA Polymerase enzyme
  - does not need primer

(missed something he wispired)

How does RNA polymerase know where to start + stop?



both directions are possible

⑫

Lots of regulation/guides/frames that help shape it

10/4

Samin Houshyar
R27 & R28
Recitation 7

I. Experiments

   a. Griffths

   b. Hershy-Chase

   c. Avery

   d. Measolson-Stahl

II. DNA Replication

   a. Substrates & Molecules Involved

   b. Enzymes
      i. Helicase

      ii. DNA Polymerase

      iii. Ligase

iv. Topoisomerases

v. Primase

vi. Telomerases

III. Direction DNA Replication

a. In what direction does DNA polymerase work and WHY?

b. Leading vs. Lagging strands (continuous vs. discontinuous)

IV. Okazaki fragments

V. Proof-reading

7.012 Recitation

10/4

Exam 1

(Before hand mentioned cold 6^3 exam
~~Avg~~ She said then it would not count)

Avg 76

Rec avg 79

- Re add your points
- Go over exam w/ solutions
- Regrade appoint her, MM or DS

OH is W 7:30-9:30 pm
Stata Cafe
No sun

Resources

- Old OCW + Stellar
  - inc 7.013, 7.014 - core of course same
- Course tutors
- OME tutoring
- Seminar XL
- OH
  - any TA
- Class + Rec

She took exam once w/ nose issue

- Got a C-
- But got A in class
- Stuff will be back!

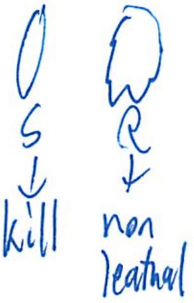
③

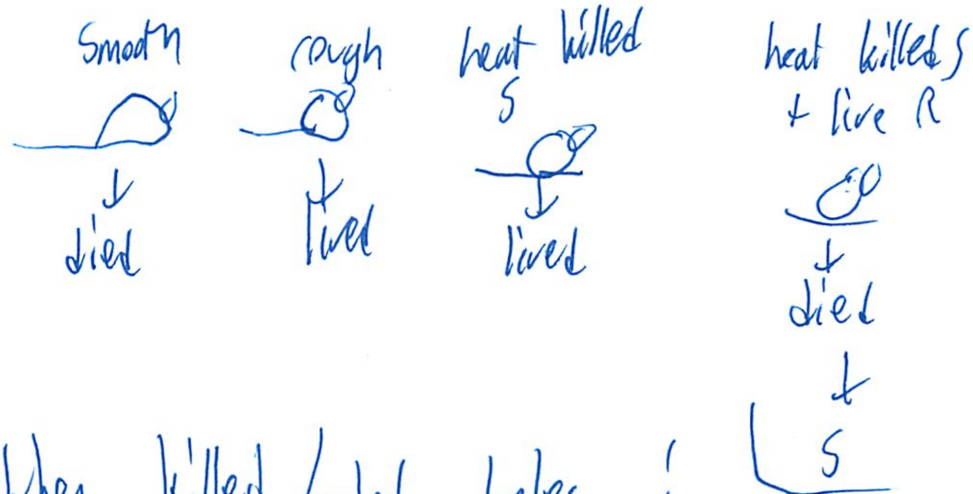
DNA Replication

Griffith

penumica + mice

↑ sp


S
R
kill
non leathal


Smooth
died
rough
lived
heat killed S
lived
heat killed S + live R
died
S

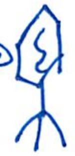
When killed/poked holes in bacteria
it released something → transforming principle
rough took it up + became smooth

Smooth has lots of polysachrides
that protect it from immune system

knew about nucleic acids
but didn't know what DNA did

Hershey - Chase

Used a virus phage


Capsid
(protien)
*S
has DNA
*P

label radioactivity w/ Phospher $^*P$

Sulfer is in ~~DNA but not phage~~
protien $^*S$

If protien transforming → $^*S$ should be found
" DNA " → $^*P$ " " "

End up w/ tube

virus
bacteria

⑤

Look at where the radioactivity

Saw *P in the bacteria

So knew it was DNA

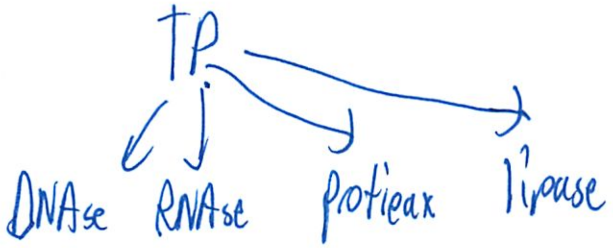
Saw *S in supernatant

This is how can distinguish ~~DNA~~ DNA, protein

Avery

have transforming principal

will take it + treat differently



← enzymes that break these down

adds live R (rough strain from Griffith's)

looks at mice

from the heat killed S

So

↓ ↓ ↓ ↓

live dead dead dead

since broke down the DNA that
coded for Smooth

note when kill DNA → its DNA gets out
when live it's protecting

DNA when heat up → will recombine

Measelson + Stahl

How DNA divides

[Will come back to]

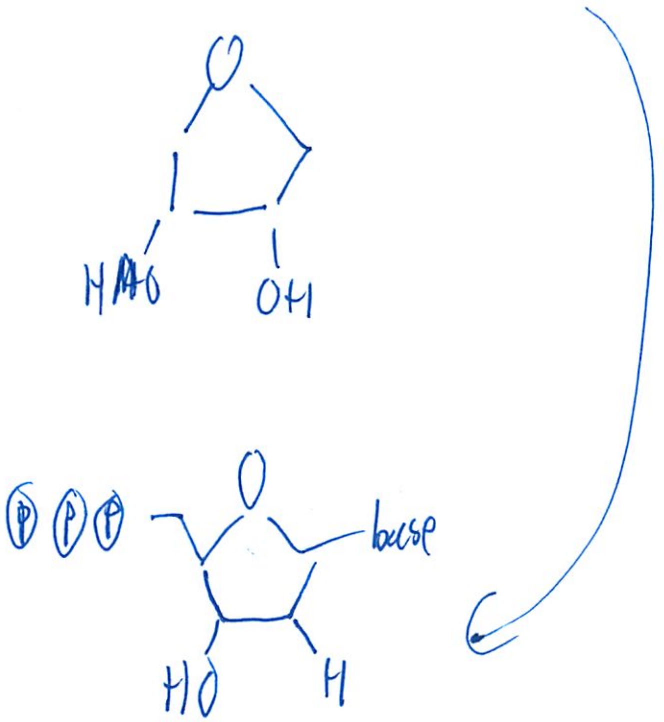
⑦

Replication Substrates

Enzymes

DNA Template

dNTP - deoxy neucleotide



Primers (RNA)

Short strands of nucleotides

Starts w/ RNA primers

Needed to initate replication

Can't start synthasis unless existing strand

Enzymes

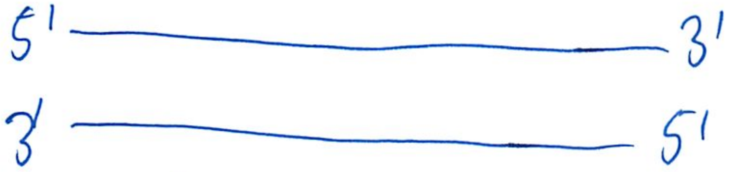
DNA Polymerase

Helicase

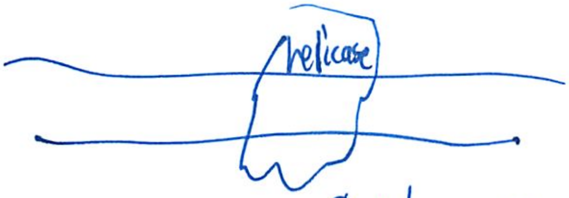
(I don't remember any of this)

- opens DNA at origin of replication

DNA double strand



anti parallel

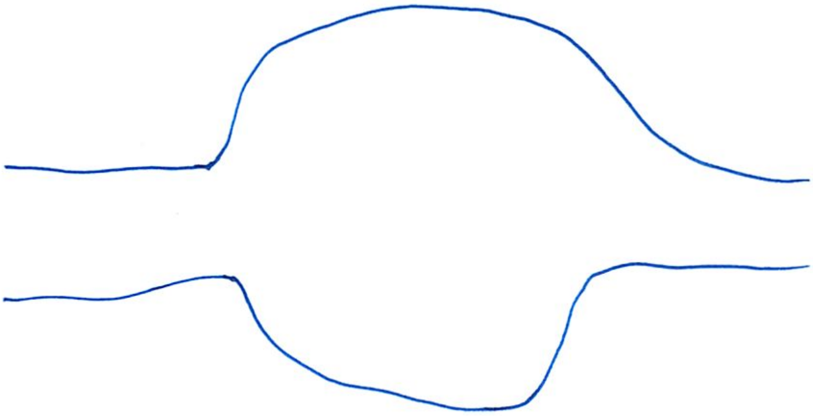


? at very specific location
= origin of replication
= ORI

Where 2 H bonds A-T righ
easier to break

⑨

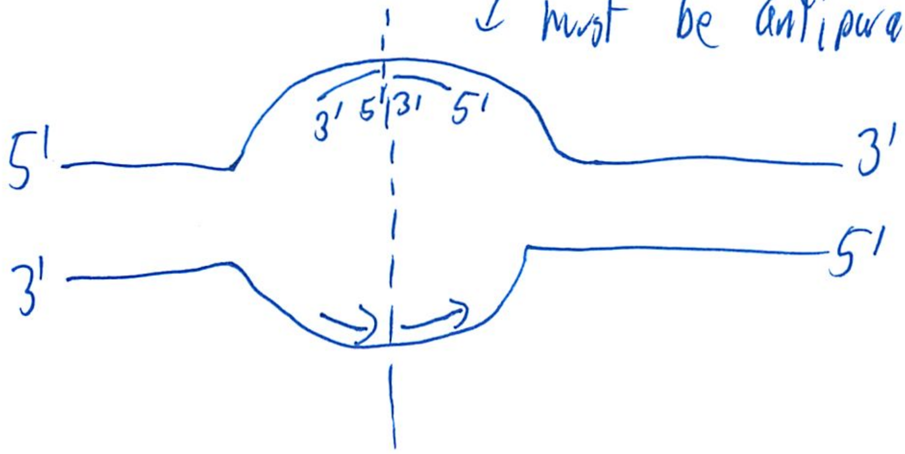
Get replication bubble



Can't start unless primers

- Primase

(RNA polymerase)

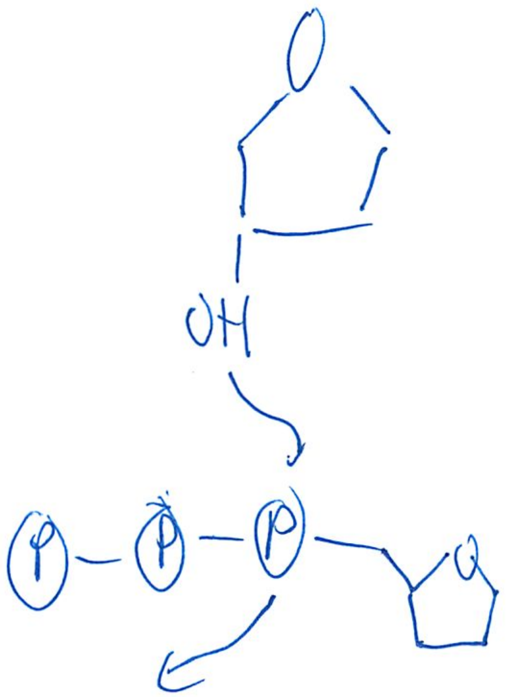


↙ must be antiparallel to parent strand

just put arrow ← 3' 5'

⑩

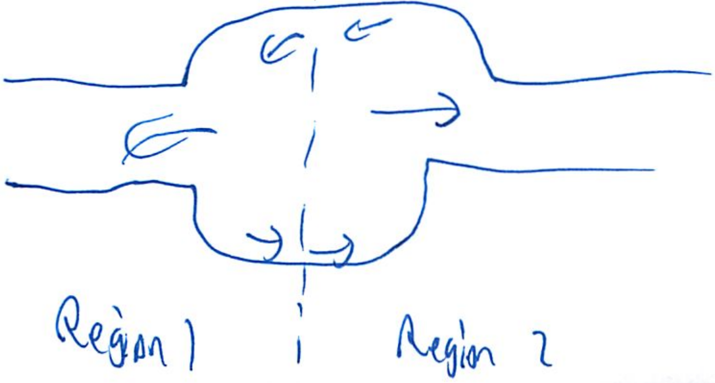
3' end means have open hydroxyl



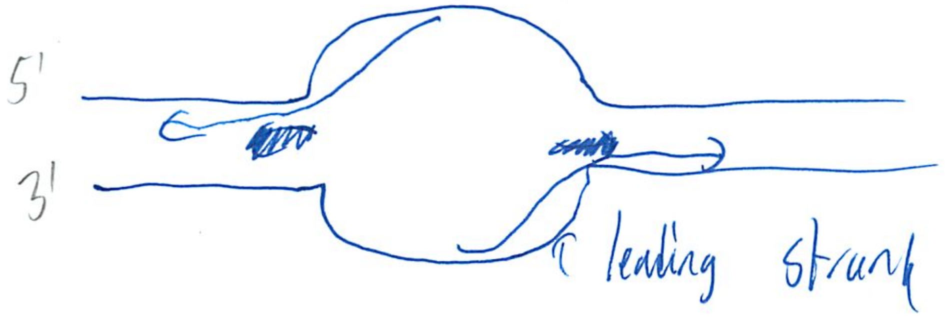
Can' only extend there

No DNA primers in natural DNA replication

Enzymes (DNA polymerase) are in order of opening strand

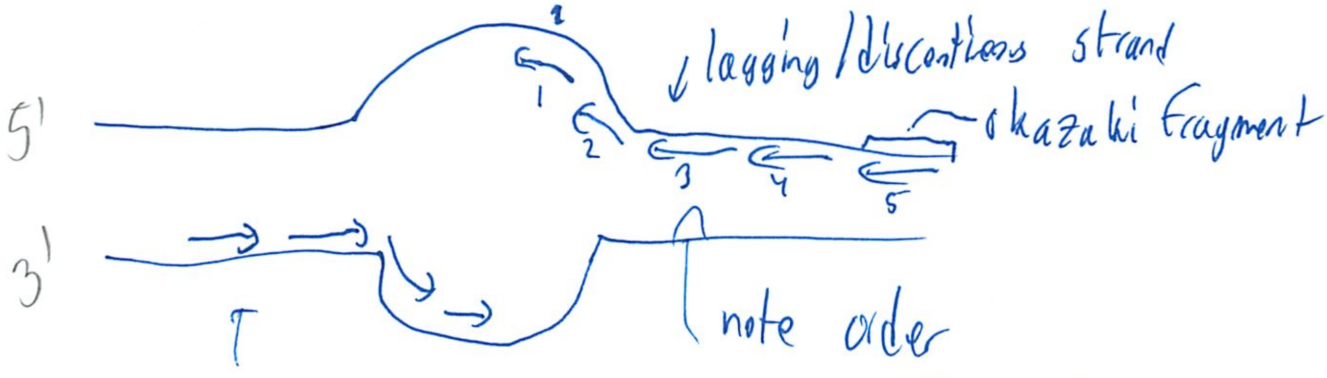

Region 1
Region 2

So only 1 group extents


5'
3'
leading strand

The other direction keeps jumping off

↳ she will send animation


5'
3'
2
1
2
3
4
5
lagging / discontinuous strand
okazaki fragment
note order

note they are not connected to each other

need ligase to combine them

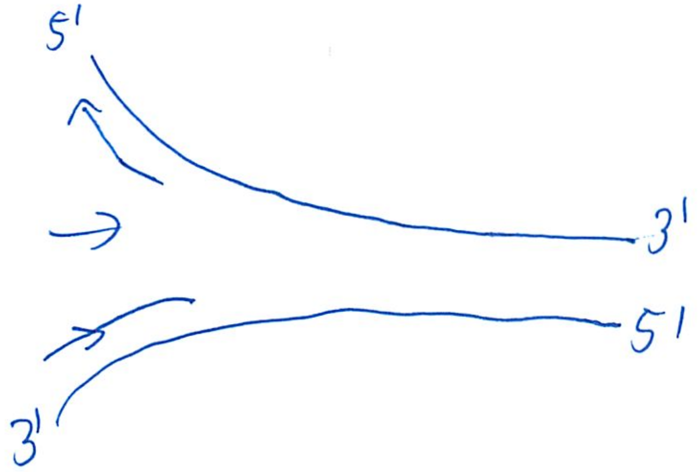
Won't worry about details

So always figure out direction

So draw arrows

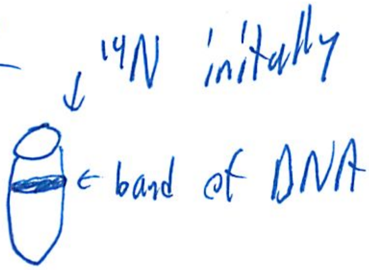
So know which is leading / lagging

⑫



Questions

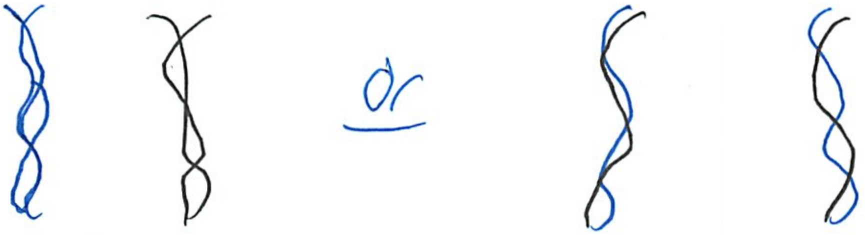
1.


¹⁴N initally
← band of DNA

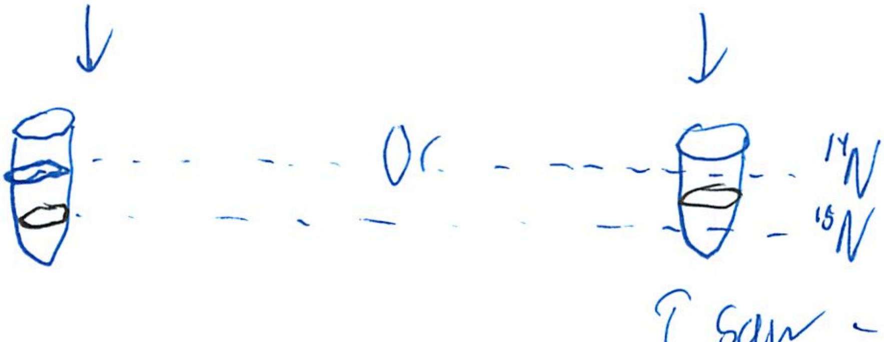
] density/weight gradiant

then switch to ^{13}N

Is it



So



I saw - in the middle

2012 7.012 Recitation 7

Summary of Lectures 10 & 11:

The transforming material (Frederick Grifith's Experiment): In 1928, Frederick Griffith performed an experiment using pneumonia bacteria and mice to confirm that DNA was the genetic code material. He used two strains of *Streptococcus pneumoniae*: a "**smooth**" strain which has a polysaccharide coating around it that makes it look smooth under the microscope, and a "**rough**" strain which doesn't have the coating. When he injected live S strain into mice, the mice contracted pneumonia and died. When he injected live R strain, which typically does not cause illness, into mice, as predicted the mice survived. Griffith then used heat to kill some of the S strain bacteria and injected those dead bacteria into mice. This failed to infect/kill the mice. He then injected another group of mice with a mixture of heat-killed S and live R, and the mice died! The dead mice had contained the live S strain bacteria. Griffith concluded that the live R strain bacteria must have absorbed genetic material from the dead S strain bacteria, and since heat denatures protein, the protein in the bacterial chromosomes was not the genetic material.

DNA as the transforming principal (Hershey and Chase Experiment): In 1952, Alfred Hershey and Martha Chase did an experiment using T2 bacteriophage viruses that infect E. coli bacteria. At that time, people knew that viruses were composed of DNA (or RNA) inside a protein coat/shell called a **capsid** and they replicate by taking over the host cell's metabolic functions to make more virus. They radio labeled the capsid proteins using S^{35} and DNA/RNA using P^{32}. They grew two batches of T2 and *E. coli*: one with radioactive sulfur and one with radioactive phosphorus to get batches of T2 "labeled" with either radioactive S^{35}or P^{32} and used them to infect E. coli. In the next step, still in separate batches, the mixtures were agitated in a kitchen blender to knock loose any viral parts not inside the *E. coli* but perhaps stuck on the outer surface, centrifuged to pellet the E. coli bacteria and tested them for the presence of radioactivity. The S^{35} was found in the supernatant, indicating that the viral protein did not go into the bacteria and P^{32} was found in the bacterial pellet, indicating that viral DNA did go into the bacteria and was the transforming principal.

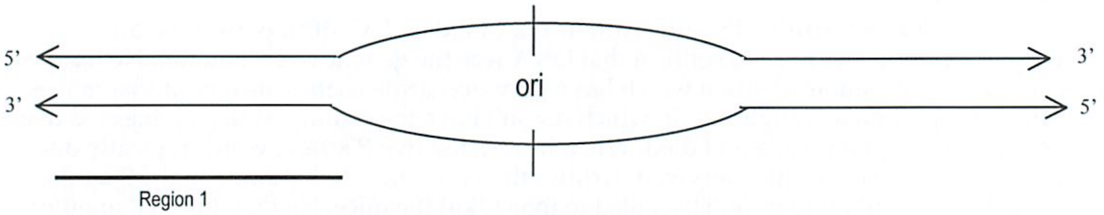
DNA Replication: The process of DNA replication occurs when two double-stranded DNA molecules are made from one double-stranded DNA molecule. This process occurs only in actively dividing cells because DNA replication always precedes cell division. The enzyme that catalyzes DNA replication is the protein DNA polymerase. DNA polymerase catalyzes the reaction of forming a phosphodiester bond between two deoxyribonucleotides. The start signal for DNA polymerase is an origin of replication, which is a site on DNA that may or may not be inside a gene. DNA polymerase proceeds down a piece of DNA until the entire genome is replicated. The new strands are synthesized in the 5′ to 3′ direction. DNA replication occurs in a semi-conservative fashion. This means that, when one double-stranded molecule of DNA is replicated, the original double-stranded molecule is unraveled such that it is two separate "old" strands. Then each "old" strand acts as a template for one "new" strand. At the end of replication, each of the two new double-stranded molecules consists of one "old" strand running antiparallel to one "new" strand.
Besides DNA polymerase, two other enzymes play critical roles in DNA replication. Helicase unwinds the two strands of DNA such that they can be replicated. DNA ligase catalyzes the formation of a covalent bond between two adjacent nucleotides, and is used to seal the gap between the replicated fragments of DNA that are produced on the lagging strands of a replication fork. The leading strand is the strand that is being replicated in the same direction that the fork is being unraveled by helicase; thus DNA polymerase can just replicate this strand is one long continuous piece. The lagging strand is being replicated in the opposite direction that the fork is being unraveled, so DNA polymerase has to constantly hop on, replicate a piece, and then fall off and hop back on again. This leads to the strand being replicated in pieces that must be joined together by DNA ligase.

Transcription: The process of transcription occurs when a double-stranded DNA is unwound and one strand is transcribed to make RNA. The RNA that is transcribed from a gene can be one of three types of RNA: mRNA, tRNA or rRNA. All of these types of RNA are identical in their chemical composition, but they differ in their function. The mRNAs are transcribed from genes that encode proteins; these mRNAs will go on to be translated by ribosomes. The rRNAs and tRNAs are transcribed directly from their respective genes, but they are never translated. Instead, they participate in the process of translating mRNAs. The rRNAs complex with ribosomal proteins to form the ribosome. Each tRNA becomes covalently attached to the correct amino acid, and then donates that amino acid to the ribosome whenever that amino acid is needed for translation of a protein.

Questions:

1) You radiolabel the bacterial cells with N^{15}. You then grow them for three generations in N^{14} containing medium and separate the bands based on the difference in their density. Draw the band profile after the 1st, 2nd and 3rd generations.

2) Consider the following segment of the DNA that is a part of a much larger molecule constituting a chromosome. The sequence of region 1 is shown below.



Region 1: 5′...ATTCGTACGATCGACTGACTGACAGTC...3′
3′...TAAGCATGCTAGCTGACTGACTGTCAG...5′

a) If we assume that a fragment of the lagging strand is made from region 1, what will be its sequence? Label the 5′ and the 3′ ends.

b) Why is DNA synthesis continuous at one strand and discontinuous at the other strand?

3) The following is a partial sequence from the hypothetical gene, gene X. The boxed region is the promoter, and the direction of transcription is indicated by the arrow. Transcription begins at and includes the first G/C base pair after the box.

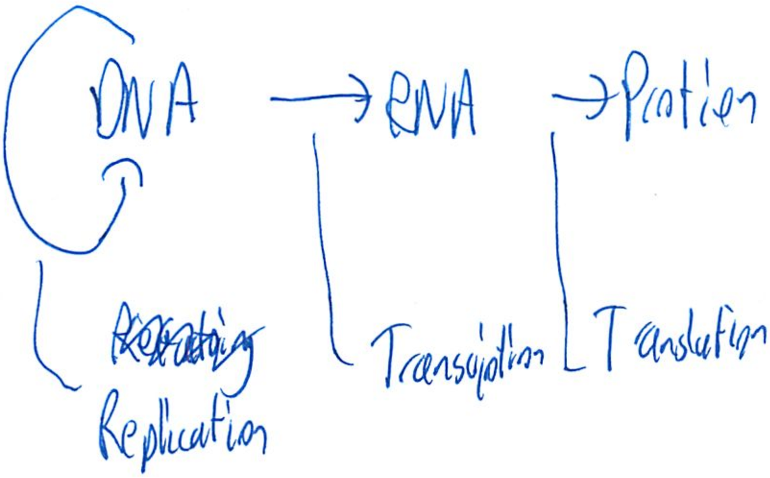
```
5′ GGACCGCGGGGCAGGATTGCTCCGGGCTGTTTCATGACTTGTCAGGTGGGATGACTTGGATGGAAAAGTAGAAGGTCATG
   +---------+---------+---------+---------+---------+---------+---------+---------
3′ CCTGGCGCCCCGTCCTAACGAGGCCCGACAAAGTACTGAACAGTCCACCCTACTGAACCTACCTTTTCATCTTCCAGTAC

5′ GGGTGGCCAACTTGGGCGAGAAAAGGTATATAAAGGTCTCTTGCTCCCATCTACTGCCCCATTTGTAGGTATTCCAGCAG
   +---------+---------+---------+---------+---------+---------+---------+---------
3′ CCCACCGGTTGAACCCGCTCTTTTCCATATATTTCCAGAGAACGAGGGTAGATGACGGGGTAAACATCCATAAGGTCGTC

5′ ATCAGACAACGTCTCATGGGAGTACTTGGATGGAAGAGTAGAAGGTCATGACCAACCTCTTCCAATCCAACCACAAACAG
   +---------+---------+---------+----←----+-------+---------+---------+---------
3′ TAGTCTGTTGCAGAGTACCCTCATGAACCTACCTTCTCATCTTCCAGTACTGGTTGGAGAAGGTTAGGTTGGTGTTTGTC

5  AAAATCAGCCAATATGTCCGACTTCGAGAACAAGAACCCCAACAACGTCCTTGGCGGACACAAGGCCACCCTTCACAACC
   +---------+---------+---------+---------+---------+---------+---------+---------
3′ TTTTAGTCGGTTATACAGGCTGAAGCTCTTGTTCTTGGGGTTGTTGCAGGAACCGCCTGTGTTCCGGTGGGAAGTGTTGG

5′ CTAGTATGTATCCTCCTCAGAGCCTCCAGCTTCCGTCCCTCGTCGACATTTCCTTTTTTTTCATATTACATCCATCCAAG
   +---------+---------+---------+---------+---------+---------+---------+---------
3′ GATCATACATAGGAGGAGTCTCGGAGGTCGAAGGCAGGGAGCAGCTGTAAAGGAAAAAAAAGTATAATGTAGGTAGGTTC
```

What are the first 10 nucleotides of the mRNA produced from gene X?

7.012

10/5

Molecular Bio 3


DNA
RNA
Protien
~~Repeating~~
Replication
Transcription
Translation

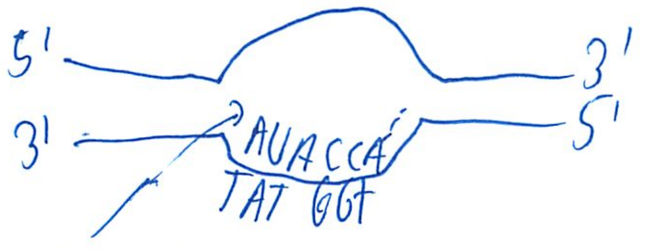
Last time Replication

very complicated

Central to medicine

Transcription

taking info out of chromosome

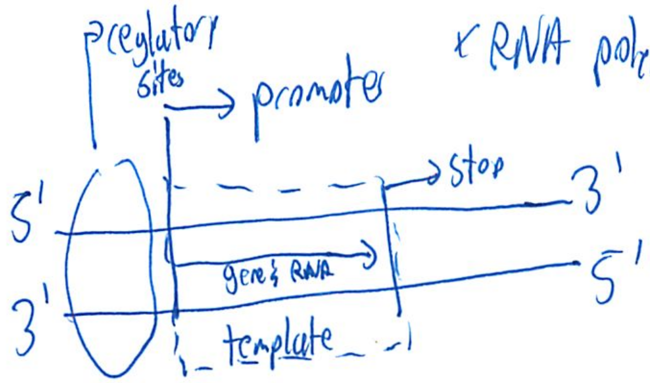

5'
3'
3'
5'
AUACCA
TAT GGT

putting in RNA 5' → 3'

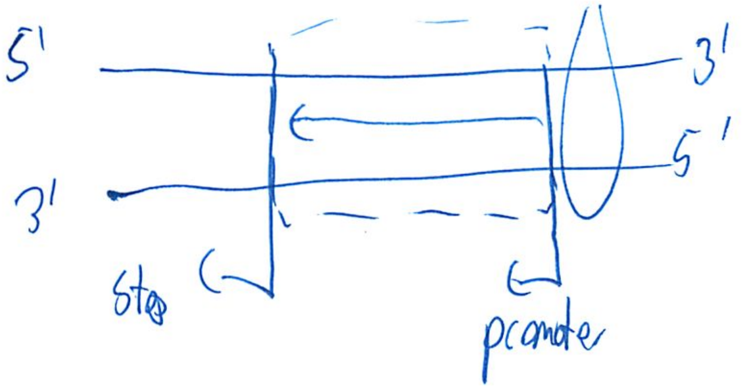
messenger RNA

mRNA

②


regulatory sites
promoter
stop
5'
3'
3'
5'
gene & RNA
template

× RNA polymerase starts
plus helpers → other protiens


5'
3'
5'
3'
stop
promoter

Regulation needed to get right RNA at right time

Translation

5' GUAACCGUA - - - - 3'

Now one alphabet to a totally diff lang
RNA → amino acids
4 letters → 20 letters

Need $4^{\textcircled{3}}$ letters
≥ 20

③

The answer was not obvious

? RNA recognize by shape of amino acid

- loopy codes

? all wrong

Francis Crick

Look up table

3 letters RNA → amino acid

w/ adaptor table

= adaptor hypethois

Is a simple look up table
(projected)

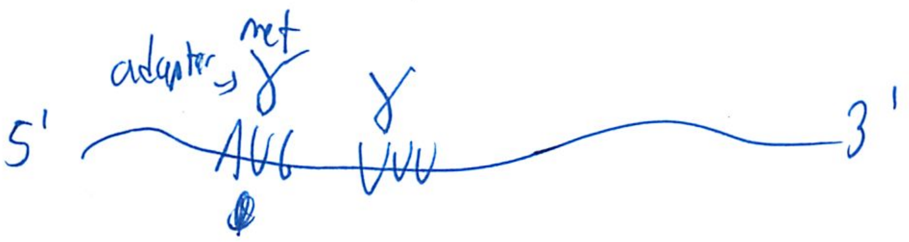
④ 5' 3'

ACC ↓ | ~~[illegible]~~ GUA ↓ Valine | UUU ↓ phe | UAG ↓ Stop

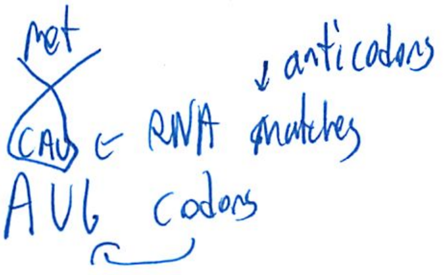
~~[illegible]~~ Like a 2 tape Turing machine

- RNA → tape 1
- Amino acid → tape 2

So we have our RNA


met
adapter →
5'
AUG
UUU
3'

What kind of adapter?


met
CAU
← RNA matches
↓ anticodons
AUG codons

= transfer RNAs

"charged" w/ amino acid connected w/ an enzyme

Functionally translation starts at $\overbrace{AUG}^{\text{met}}$

└ always

well almost always

but for 7,012 always

met - phe - pro ⟶ until stop

How do you know?

Experiments!

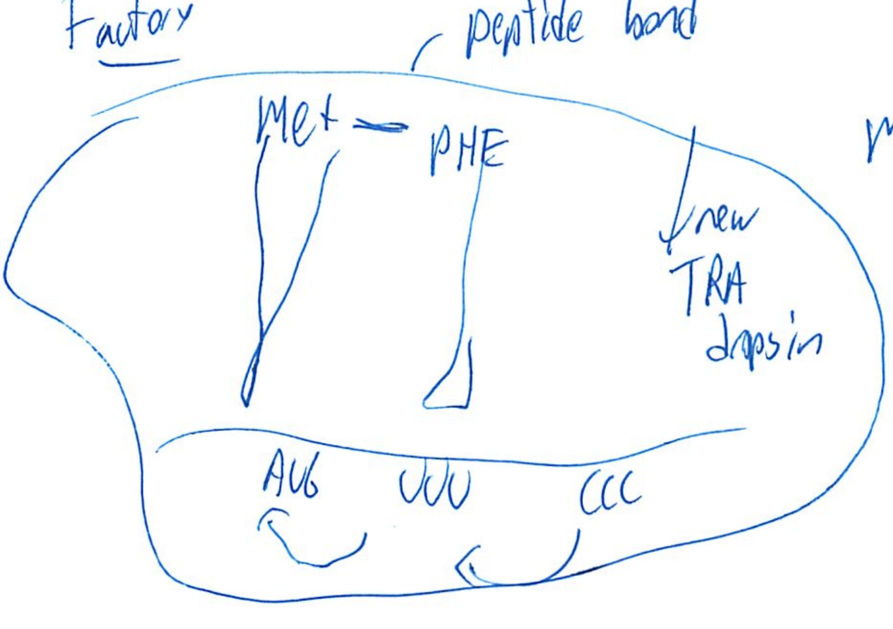
So gave it synthetic RNA

Took a few years in 1960s

Pretty hard to make RNA

note no AUG required here!

Factory


peptide bond
Met
PHE
new TRA drops in
AUG
UUU
CCC

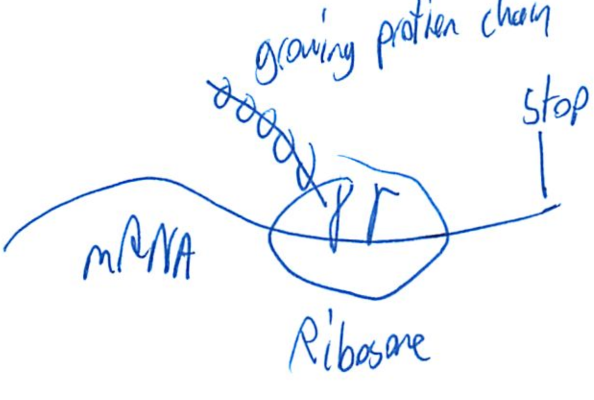
Makes a Ribosome

- protiens + RNA

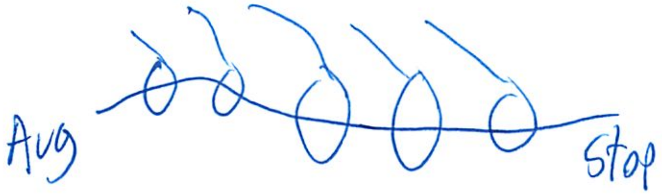
Messenger, transfer

---

after a while


growing protien chain
stop
mRNA
Ribosome

multiple ribosomes can be on message
- same protien
- but a diff molecule of it



How many enzyme molecules needed?
Must regulate!

dynamic balance b/w degradation + making RNA

ribosome translated ribosomal protien
can't boot genome w/o ribosomes
so need continuity of life

all cells have ribosomes

RNA came 1st - since can combine protiens
- but now mostly protines do it
- but RNA does some work in some areas
- Shadow of an ancient world

Prof: Life is about instability
must live on the edge

Dynamic how much is created
DNA is static storage process

No ~~TRN~~ TRNA ~~that~~ that matches stop
in-stead protien recognizes
Comes from ~~a~~ genes

Variations on the Theme

DNA → RNA → Protien

Simple!

- kinda universal across life but not totally

1. Eukarots people
2. Prokariots
3. Virus

⑨

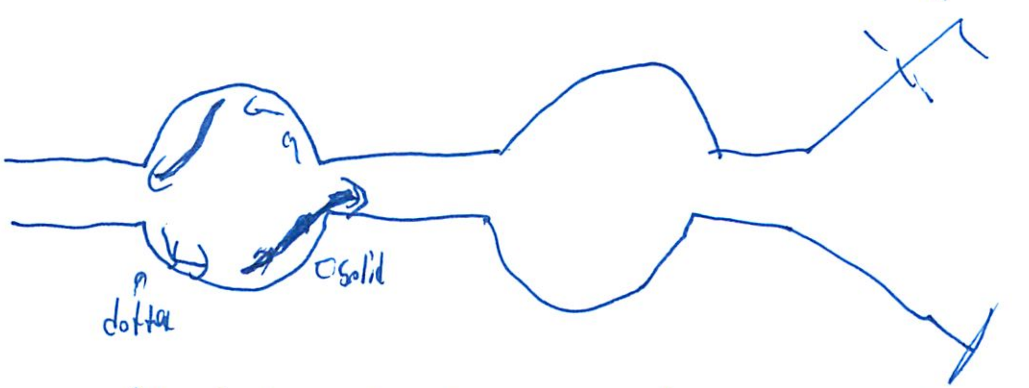
a) replication

- Eukarots - DNA in genome varies tremendously

Humans	23 pairs	3×10^9
mouse	20 pairs	2.7×10^9
tomatoe	12 "	5×10^9
fly	4 "	$2 \cdot 10^8$
yeast	16 "	$1.3 \cdot 10^7$

but are all linear

How do we know to start at top



just stops

can't just cut off the end!

end = telomere

(TTAGGG) → repeats

__Teleomerase__ adds telomeres to the end
TTAGGG

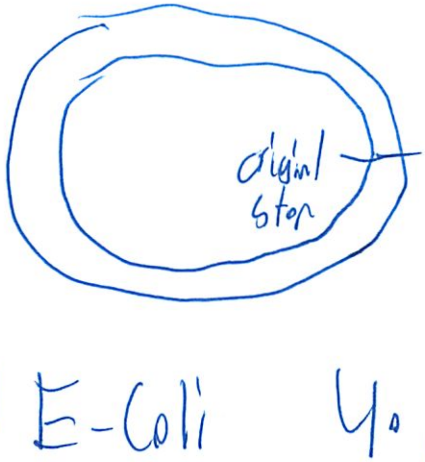
90% of cancers turn on teleomerase

So can use this knowledge as a tool

---

b) __Prokaryotes__

Circular chromosome, DNA
no ends!

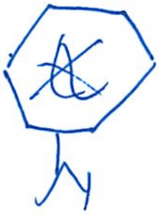


double standed

E-Coli $4 \cdot 10^6$

⑪

c) Viruses



← just small package w/ DNA

10^5-10^6 in size

Want linears or circles

Some linear) double stranded Chromosomes/DNA

Some circular

Some single stranded DNA

how does it replicate?

right before it is used

"travels light"

Some RNA ?? ← next time

7.012 Study

10/8

(Review of new material)
L Molecular Biology

2 ways to look at: chemistry + genetics

Discovering the actual "transforming principle"

Griffiths/: smooth vs virulent
rough/nonvirulent + dead smooth/virulent
Somehow smooth virulent!

DNA
2' deoxyribose

Assay/ method to determine potency of bio activity by testing its effects on living organisms and comparing to known standards

---

Key enzymes - enzymes that destroyed 1 specific function
- RNase
- Protase
- DNase

②

Hershey + Chase rule out proteins

T2 virus → bacteriophage

E-coli

Sulfer - only in protiens

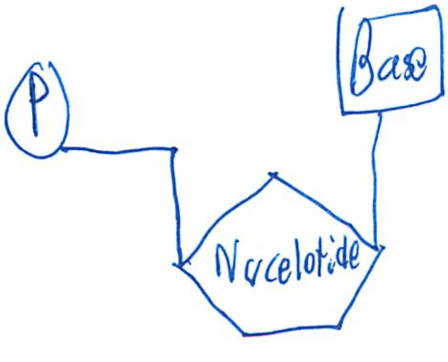
Phosphor - only in DNA

added Radio isomers

blend it up

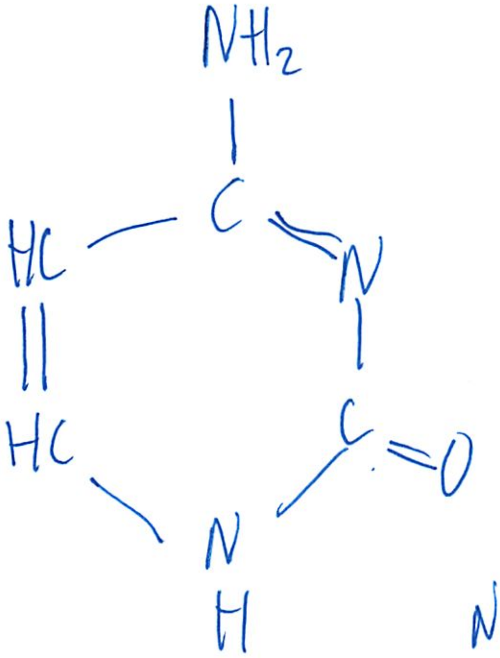
measure pellet + liquid

Structure of DNA

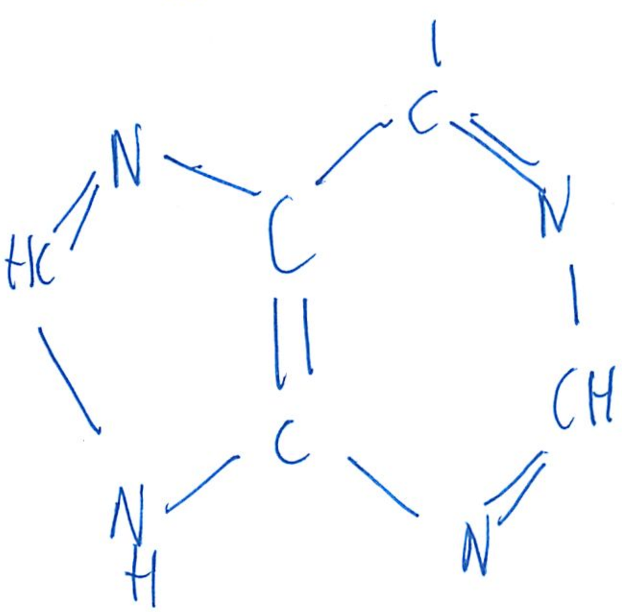


③

different bases


NH2
HC
HC
C
N
C
O
N
H

Cytosine
C


NH2
N
HC
C
C
N
CH
C
N
N
H

Adenine

(ask TA about the lines - sometimes drawn sometimes not)

④

(reading from book)

Chargaff's Rules

C = T

A = G

but otherwise it varies b/w species

double-helix model

2 sugar-phosphate backbones

helixes hold together w/ H bonds

anti-parallel

5' → free phosphate

3' → free hydroxyl

DNA Replication

Semiconservative replication - each offspring gets 1 complete strand from each parent

Meselson + Stahl

Tracked DNA w/ density isotope label

~~& Meselson & Stahl~~

Used N - since in base

^{15}N ← grow E coli in

Then moved to ^{14}N

So when bacteria divided here, it picked up the lighter N

*Centrifuge → heaviest on bottom (^{15}N)

⑥

Replication

Starts at origin of replication

↳ each bacteria chromosome contains a single origin of replication

H-bonds breaks and strands split

helix unwinds at replication forks

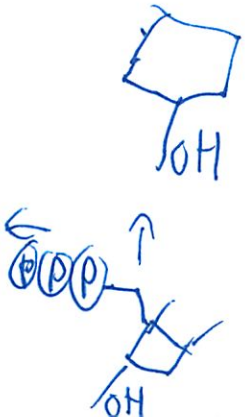
↳ So this is where it unwinds

according to base pairing rules

3' → 5' original

5' → 3' new strand

nucleotides added to 3' end of growing daughter strands



⑦

Several protiens involed in process

- Origin binding proteins - binds the origin of the replication sequence + seperates the DNA strands
- DNA helicase - binds each replication fork to the dable helix + contines to break the H-bonds
- Single strand DNA building protiens SSBs - prevent the seperate strands from joing togeter by binding to seperate strands and stabilizing them
- Topoisomerase binds to ~~DNA~~ dable helix ahead of replication fork + releaves stain as helix unwinds

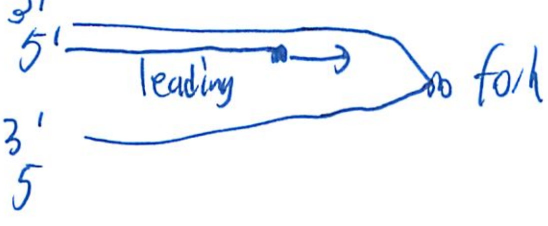
✱ always 5' → 3'

DNA polymease - contines syntasis from existing 3' OH

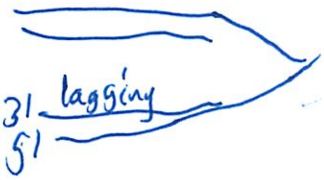
primase - Syntesizes a primer for each template
strand, one nucleotide at a time
usually a short strand of RNA
needs a template strand
- DNA primase falls off, DNA polymerase
adds onto 3'OH of RNA primer then
RNA is removed shortly

DNA Polymerase III - recognizes primer
adds nucleotides to 3' end of primer
- both daughter strands created at same time

leading strand - grows continuosly
3' end faces fork

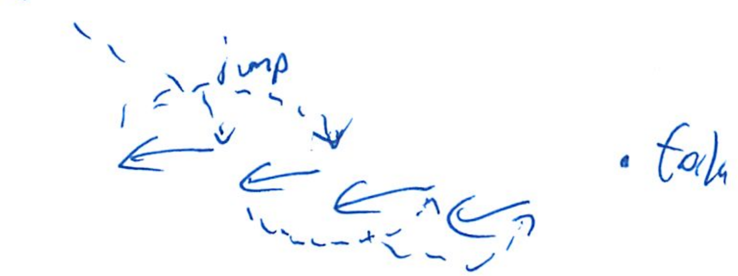


lagging strand – much more complicated



DNA polymerase moves in same direction as helicase

Must periodically jump
- moves **closer** to fork



needs a new primer

~~and~~ forms fragments of 1000-2000 nucleotides
= **Okazaki fragments**

DNA polymerase I - replaces each RNA primer w/ corresponding DNA nucleotide

DNA ligase joins the fragments together to form a single strand

(Red lines in video = RNA fragments)

Proofreading + repair

Mismatch repair - error in base pairing

enzymes cut out incorrect nucleotides + replace it w/ appropriate nucleotides

Nucleotide excision repair - replace damaged DNA from env

Nuclease cuts out DNA strands

Some may be unrepaired
- evolution
- or disease/cancer

⑪ 11p

DNA Replication Eukarotic

billions of base pairs

thousands of origins of replications

linear

L so polymerase III can't add final seq of DNA

end would get chopped off

(don't get why)

No where to attach RNA primer for final fragment

So **telomeres** at end

Telomerase adds more to template strand ↗ [illegible]

further elongating it

Then DNA polymerase + ligase can extend it

(it doesn't say what happens w/ extra lenght...)

⑫

Mutations

heritable changes due to DNA

Germline → in cells that produce gametes

Somatic → non gamete cells
when divides

Can take place in coding + $\underbrace{\text{non coding}}_{97\%}$

point mutations → 1 changed base pair

Can up to 1 mill change missing

transition purine ⇔ purine
pyrimdine ⇔ pyrimidine

transversion pyrmidine ⇔ purine

missense mutation - altered codon

nonsense mutation - stop codon added

(I don't think we have to know this stuff...)

Chromosome

Contain the genes

made of DNA + protines

46 in humans

day to day → uncondensed state = cromatin

dna - to protine mass aprox =

Small → circlur
large → liner

karyotype visualizing chromosome w/ dye stains

Shows a banding pattern
↳ density

drawn w/ short end at top

(lots of mapping techniques...)

Notes. prokarots very little non coding
eukarots very much ״
many more repeids

packaged very condensed
(skipping rest)

Gene Expression

how do genes lead to visible changes?

pathways
(something I was confused about earlier)

$$\begin{tikzcd}
\text{Precursor} \arrow[r] & \text{Orn} \arrow[r] & \text{Cit} \arrow[r] & \text{Arg} \\
\text{Enz 1} \arrow[u] & \text{Enz 2} \arrow[u] & \text{Enz 3} \arrow[u] &
\end{tikzcd}$$

↑ if mutation here - must add Orn to medium

↑ but could add Cit and bypass these steps!

↑ ahh I forgot about this (why did I not realize/figure it out)

One-gene hyp One gene controlled 1 protien
One-enzyme

polypeptide chain of amino acids protines are made from

So instead one gene, one polypeptide

Central Dogma

DNA → RNA → Protien

DNA $\xrightarrow[\text{transcribed}]{}$ mRNA

RNA = single stranded
1 sugar

2' 2'
OH H
RNA DNA

---

mRNA = messenger
tRNA = transfer
rRNA = ribosomal

Transcription Creating the RNA molecule

template strand - used

non template strand - not used

So when does DNA replicate vs RNA transcription happen ?

? cell division

2 daughter cells

?? So both mitosis and meiosis ?

~~mitosis~~ yes both

? makes RNA

So can ~~use~~ make amino acids

Occures 5' → 3' direction

Uses enzyme RNA polymerase

~~pler~~

promoter specific area of DNA where RNA polymerase binds to

~100 nucleotides up stream from transcription site

usually contains a TATA box

transcription factors = bind to the promoter
RNA polymerase notices this

some higher level stuff w/ histones
tight structures to which DNA is wound around

translation takes RNA and turns into amino acids
Occurs at ribosomes
takes place in cytoplasm
20 diff amino acids
triplet code = codons
table of these

reading frame start → stop

tRNA brings a amino acid

a single strand of nucleic acid

folds back on itself for L-shaped structure

One end has an anti-codon

matches to codon on mRNA

←

E P A

site site site

A = acceptor

P = peptidyl

entry of tRNA = req. hydrolysis of 1 GTP molecule

↑

cleveage of chem bonds by adding water

GTP Guanosine 5' triphosphate

Source of energy - like ATP

Regulated

Can turn genes on + off

allows cells to conserve energy +
adapt to surrounding

Often during transcription

(I prob should be doing qu...)

Transcription (why is this now a sep chap?)

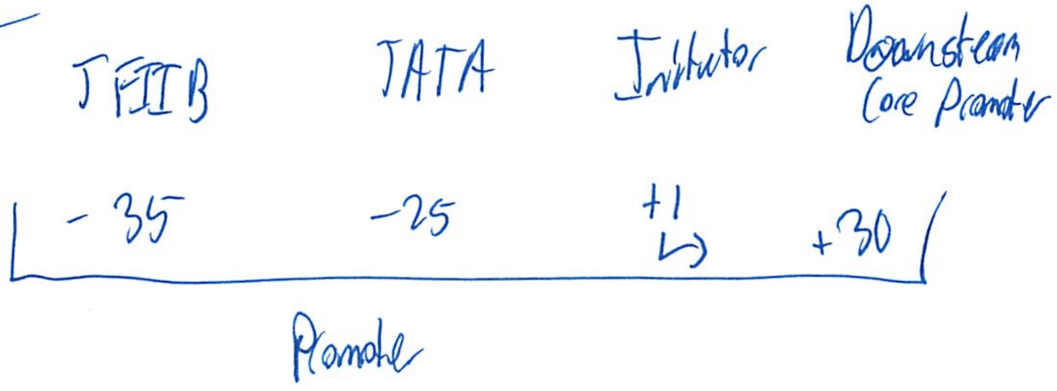
diff	DNA repl	+ trans
	2 strands used	1strand
		moves along 3' → 5'
		generates RNA 5' → 3'

3 phases
initiation
enlongation
termination

req promoter

㉑

Promoter


TFIIB
TATA
Initiator
Downstream Core Promoter
-35
-25
+1
+30
Promoter

(don't think we need these details)

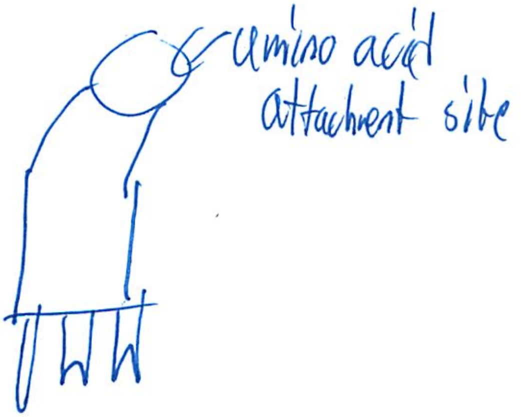
Can make many copies at same time

Translation

DNA -> protiens

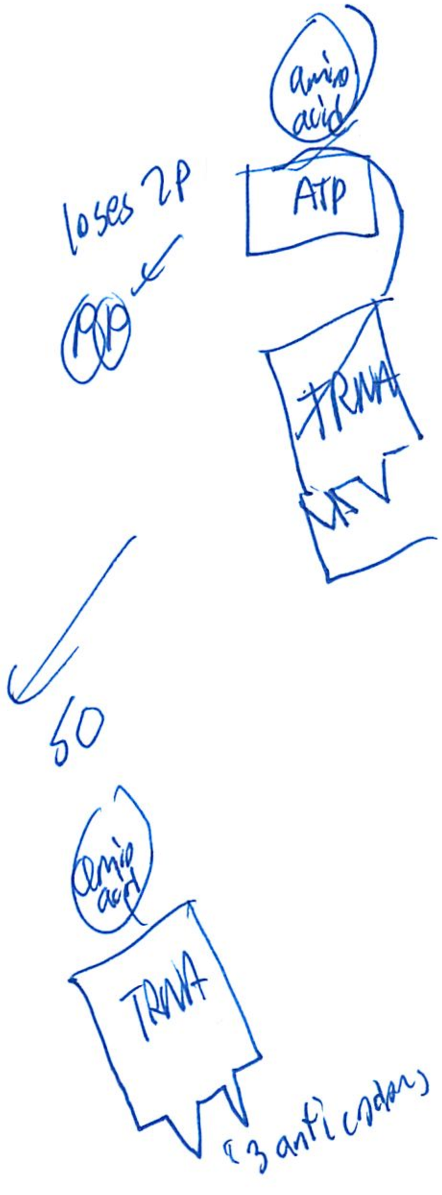
uses adaptor molecules = tRNA

(why the repeat in more detail?)


amino acid attachment site

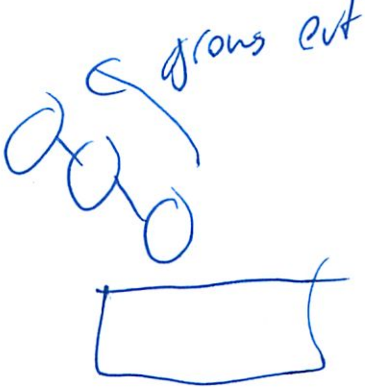
㉒

aminoacyl- tRNA synthease helps w/ matching amino acid to tRNA


amino acid
ATP
loses 2P
PP
tRNA
so
amino acid
tRNA
(3 anticodons)

Long chain of polypeptides

Pops off when stop signal



Reviewing P-set 2

10/4

(so when I turned this in was pretty sure I did well)

2) So why did I get this wrong?

How do we know

$\frac{1}{4}$ AD Ad aD ad

So it comes from $\underbrace{A\overset{a}{A}}_{\uparrow\text{pick } 1}\ \underbrace{D\overset{d}{D}}_{\uparrow\text{pick one}}$

but how do we know that case

Look above F_0 AAdd ~~A~~ aaDD

F_1 AaDd

and Then F_1 is both of those

bitter/smooth not included

(How did I make this mistake - stupid...)

3) a) Expect 9:3:3:1 → but not really

and plus some EEWW
Ee WW
EEWw
Ee Ww etc

So can't tell ? why?

So male is ee ww

is

female can be EEWW
Ee Ww

So when cross....

recombinant genotype thats when crossover - happens right?

③

3ciii) RRaa red wrinkled

rrAA yellow smooth

"so did I just not understand who the parents were"

5) Build up

I was never good at

"did we ever learn"

Yeah reversed it

but I see now

Can bypass

- like if mixing (milk powder + water) ~~+ flour~~
  for milk + flour
- if no milk powder or water - but have
  milk → good enough

④

#6 ¿ What did I mess up here?
I thought I got it

¿ Dom + Recessive ?

Or did I confuse which one they were talking about ?

didn't draw X Y for the X-linked one
which caused me to screw up iii) daughter
gets $X^b X^B$

~~and for some really stupid reason~~

for affected from 2 $\underbrace{aa}_{\text{has it so lower}}$, $\underbrace{X^B X^B}_{\text{why cap?}}$

¿ X-linked recessive
so cap = unaffected ?

2 → all lower a → disease
3 → all lower X^b = disease

(So basically just logic errors
- tricky + I got tricked)

Q4 5

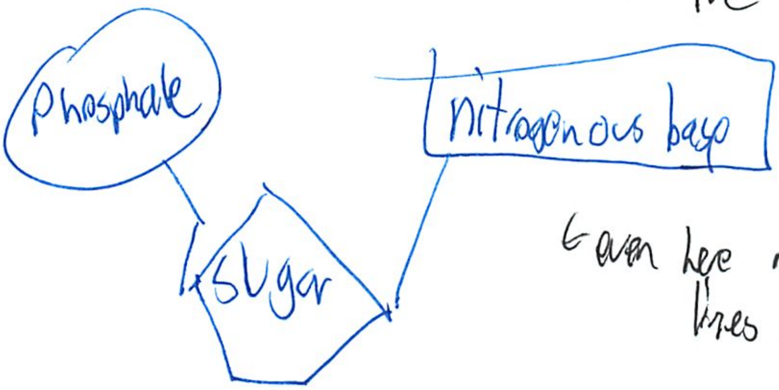
c) [illegible] letters are wrong again...

That dom/recessive thing again...

Quiz Review

/10/0

1a.) ¿Did I not memorize the sections



Phosphate
nitrogenous base
sugar

← even here not drawing lines right

b) energy
ask about

c) ask

d) ¡Measured activation energy wrong!
How stupid!

2) I actually got much of this right!
but it is something I think I know now

B②

3.) This qu is a disaster!

a) any hydrophobic

b) Phosphate group ⊖ charged

So ⊕ charged amino acids lys arg found here

So I don't know where to start here

What bg info am I missing?

c) S-S bond

Cys

again no clue!

d) Carbon + hydrogen

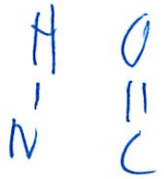
? remember structure of phospholipids

e) a) H-bonds

b) backbone ← sounds really obvious

why did I pick otherwise?

c) functional group



f)

③

f) Add Hydroxyl

Oh I actually got this one

3) Skip

4) ?, same letters issue as before

X chromosome not Y

↳ stuff always on X?

e) What did I screw up here?

oh F_1 FeBb × eebb

or my original ratios?

5) Pretty good

Overall doesn't seem like 50%...

Questions for OH 10/9

How to tell which bonds?

How to tell if mol polar?

Explain H-bonding

Pset 2 #3b

PSet 3 #2c why U?

OH

10/9

Pset 2

3b - bad qv

even though seems to be fitting

but not really

need χ^2 test

See how probable

c) Parental genotypes → get in higher #s

Crossover chance lower

exact cross over is rare

dominate = Capital

not has disease

pay more attention

② Quiz 1

1b) A, C both sugar

- know structures

B, D Nucleotides

(ATP

One is ribose, deoxribose

know ATP = ribose

So B

D: dATP → used in DNA synthasis

c) is protien

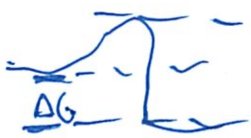
E

any amino acid

d) end. → activation energy inc ΔG

exerg. → " " would not ΔG

products lower



③

Work w/ a tutor

- Understood what going on
- do past exams
  - one a day
  - then have them check

3, a look at R groups if non polar

~~A~~ Arg has +, so ionic, so polar

O, N - most electroneg

ionic, H bond = polar

NH, O → H-bonding ~~they~~ possibles

depends on other molecules

④

→ are ⊖ so want

phosphate groups ~~are~~ ⊕ charged to ionic

a

Can do polar, no charge

H - C not polar

OH polar) since they pull electrons
NH polar) towards them

---

c) Colvant bond
Cys
only one that can do di-sulfide bridge
mentioned in class

d) H, C
no O - since polar

⑤

H-bonding on backbone

iii) $\underbrace{\text{Carboxly or amino}}_{\text{come together to make bond}}$ groups

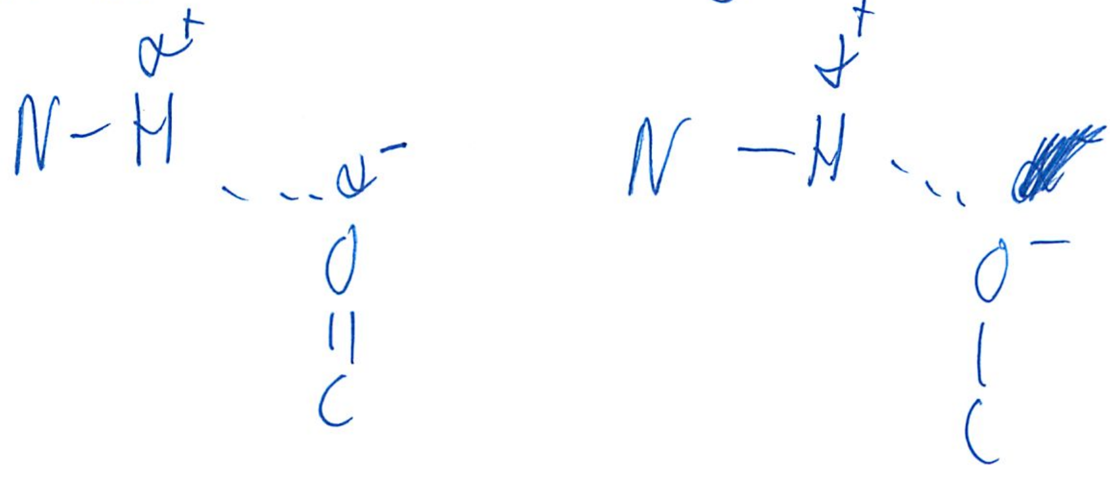
3h) polar w/ nopolar → WdW

ionic: polar + polar −

hydrogen: polar polar

charge or uncharge

H must be on electroneg



δ^+
N – H
δ^-
O
C
δ^+
N – H
O$^-$
C

both H bond

⑥ 4) For our purposes always on X

No recombination X, Y

Females can't show y-linked

4ii) Parental are the big ones

What I did → cross 2 F_1s then 9 3 3 1 (dihybrid)

Here → cross F_1 to true bread homo (test cross)

Have it set up right

but 1:1:1:1 - all =ly likely

if no recomb - all top 2 rows

When not linked

$\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$

⑦

but is - linked

So parental more likely

---

5) but 2 can't be AA if autosonal recessive

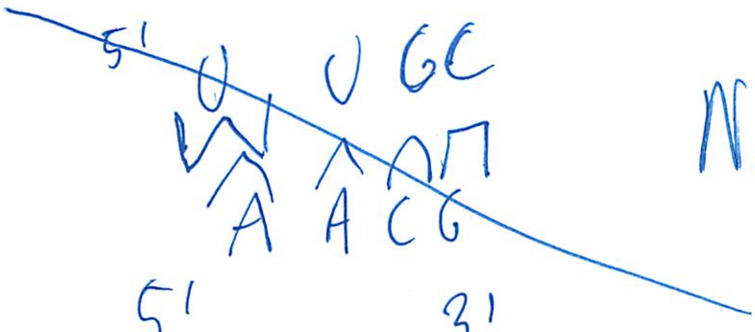
2 must be carrier

but all ~~[illegible]~~ or none

P-Set 3

2c) At the start have RNA primers

Direction matches too!



No!

#ing convention

e) Primase - don't need since

Primase is only for RNA

- makes RNA primers

need in natural DNA rep

but here → DNA primer already provided (please read!)

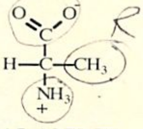
TA 10/9

Name or initials MEP

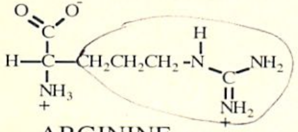
STRUCTURES OF AMINO ACIDS at pH 7.0

carboxyl [

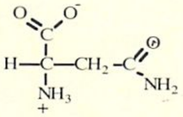
amine [



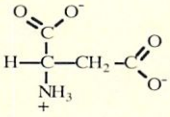
ALANINE
(ala) NP



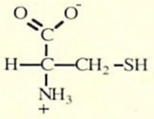
ARGININE
(arg) polar P +



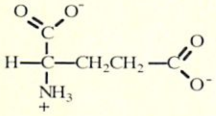
ASPARAGINE
(asn) polar



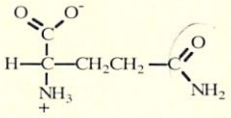
ASPARTIC ACID
(asp) polar −



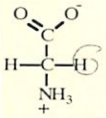
CYSTEINE
(cys) NP



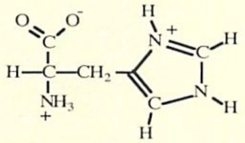
GLUTAMIC ACID
(glu) P −



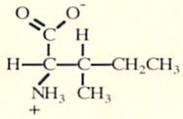
GLUTAMINE
(gln) polar



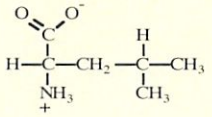
GLYCINE
(gly) NP



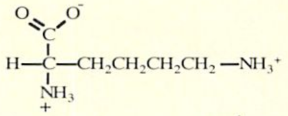
HISTIDINE
(his) polar +



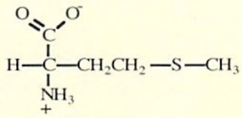
ISOLEUCINE
(ile) NP



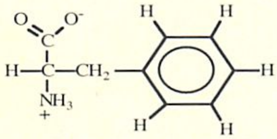
LEUCINE
(leu) NP



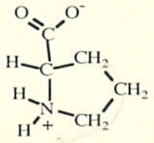
LYSINE
(lys) polar +



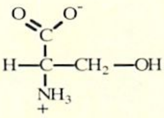
METHIONINE
(met) NP



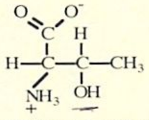
PHENYLALANINE
(phe) NP



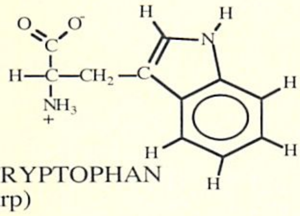
PROLINE
(pro) NP



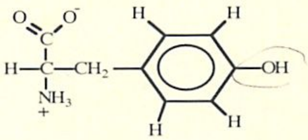
SERINE
(ser) polar



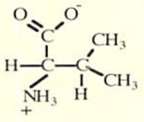
THREONINE
(thr)
polar



TRYPTOPHAN
(trp)
NP



TYROSINE
(tyr)
semi polar



VALINE
(val)
NP

? polar uncharged

10/10

Lecture

Molecular Bio 4

(2 min late)

DNA → RNA → Protien

Replication transciption translation

Variation of Theme

Eukarot

Prokarot

Virus

Slight, but sig differences

Replication

linear ds DNA

Then ends w/ tolomerase

②

b) Bacteria Circular dna ds

c) Viruses linear ds DNA

circular ds DNA

circular single stranded DNA

travel light w/ only 1 strand

RNA?

[illegible]

RNA viruses

a) RNA

how do you replicate?

Can you have RNA in
RNA out?

RNA directed, RNA polymerase

③

Does this exist in the world? Yes

But must do twice

- make complementary strand
- then make template strand

Why will host cell provide the proper RNA directed RNA polymerase? No

Does the original cell have it?

a) Call ~~The RNA be~~ The something be both the genome and mRNA?

? ⊕ stranded virus BYOM

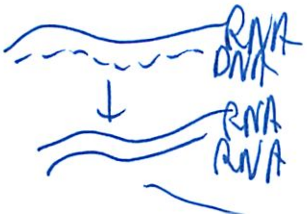
b) ⊖ stranded virus

protien coat of virus includes RNA + directed RNA polyme

BYOP

c) Makes DNA copy from RNA?

Need RNA-directed DNA polymerase



→ Chromosome, becomes part of genome

Prof: insideous

So called reverse transciptase

HIV does this
- Why it is particularly insideous
- HIV/AIDS
- Retrovirus - since runs backward

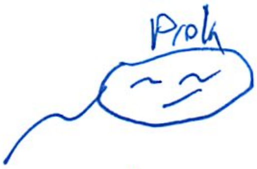
Suppose want drugs againts HIV
    If inhibit reverse transciptase
        ↱ Drugs, chemicals that do this

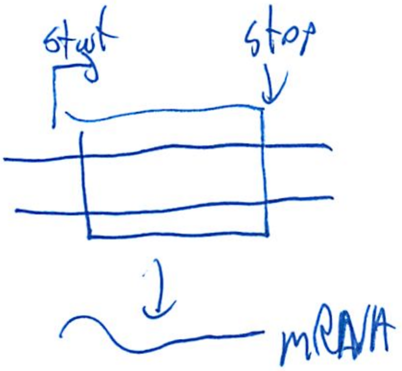
    There is a particular reverse transciptase that
        just inhibits HIV, not other stuff
    Why hard to cure anyone
        Since the DNA is in your chromosome
    Which cells?
        immune cells
        but not egg + sperm cells
        need to ask which cells

⑤

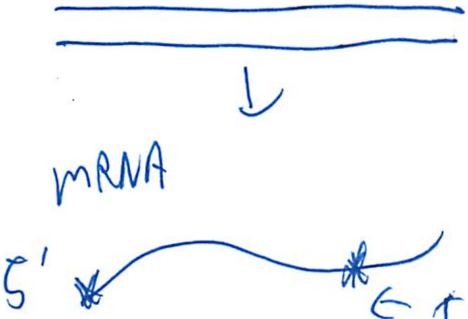
Transciption

a)




Start
stop
mRNA

b)



✱ = modifications

← TAIL

AAAAAA ... (30-200)

Poly-A-tail

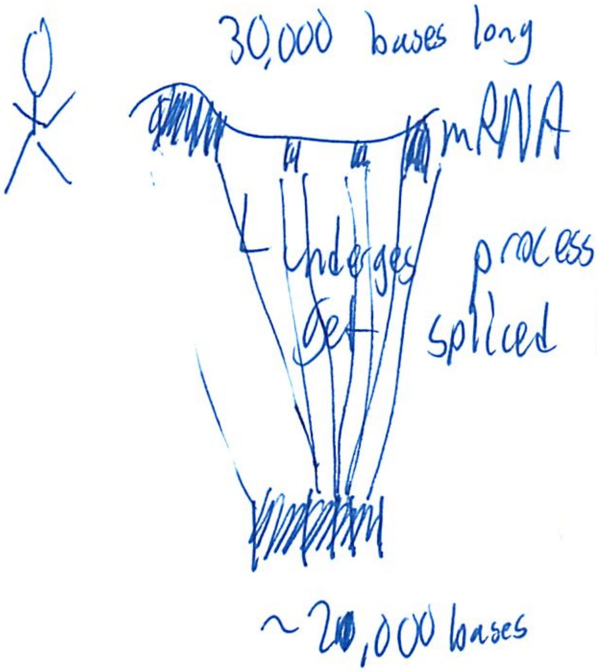
CAP

triphosphate backwards

Gppp

important for mRNA to see

⑥


30,000 bases long
mRNA
~20,000 bases

Undergoes process where only certain bits retained get spliced together

Somehow splices out messages in b/w

spliceosome - splices

exons - parts retained

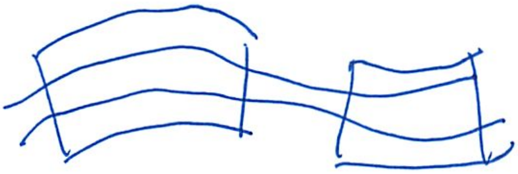
introns - parts spliced out

Factor 8 gene - 200 k bases

spliced down to 10k

Duchenne Muscular Dystrophy 200 k bases → 16k

So why bother



So a single gene might ~~make a single~~ have several alts.

Often multiple splice formats of gene

Will not be spliced multiple way on exam

Plus evolutionary benefit

lots of genes involved w/ exon splitting

Why don't bacteria do it?

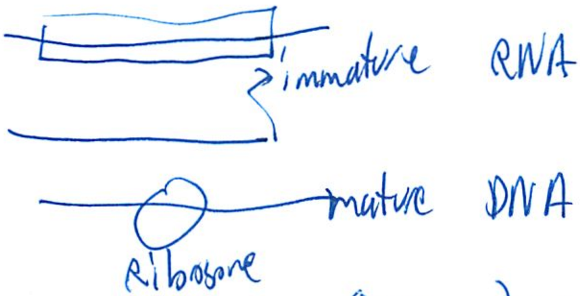
Bacteria needs to replicate quickly

~every 20 min

does not matter in humans

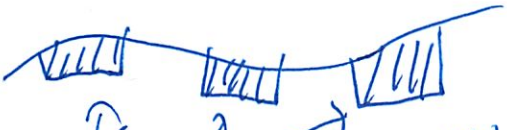
Translation

a) Eukarots


immature RNA
mature DNA
Ribosome

Once been slicd, makes one prothe

b) Prokaryots



Can produce multiple protiens

Poly cistronic message

multiples to start

Why not sep genes?
against fast replication scheme
Why not in humans?
timing does not matter
Only involve if selective advantage

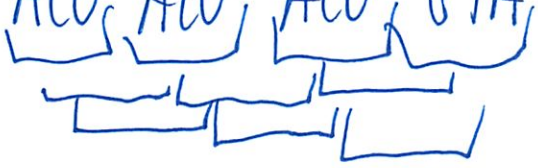
Only 1% of your genome is used

No advantage

c) Viruses

really need fast replication

5' ACU ACU ACU GTA 3'



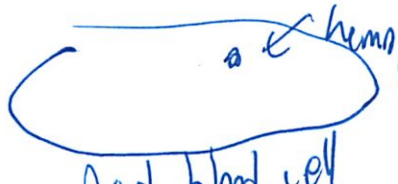
) can have diff reading frames!

So can viruses use multiple reading frames?

Yes + can happen!

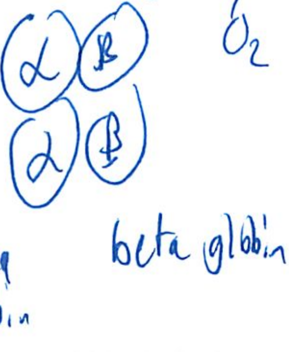
Ribosomes generally scan in order so 1st AUG

Details of 1 Human Gene → Beta Globin


hemoglobin
Red blood cell

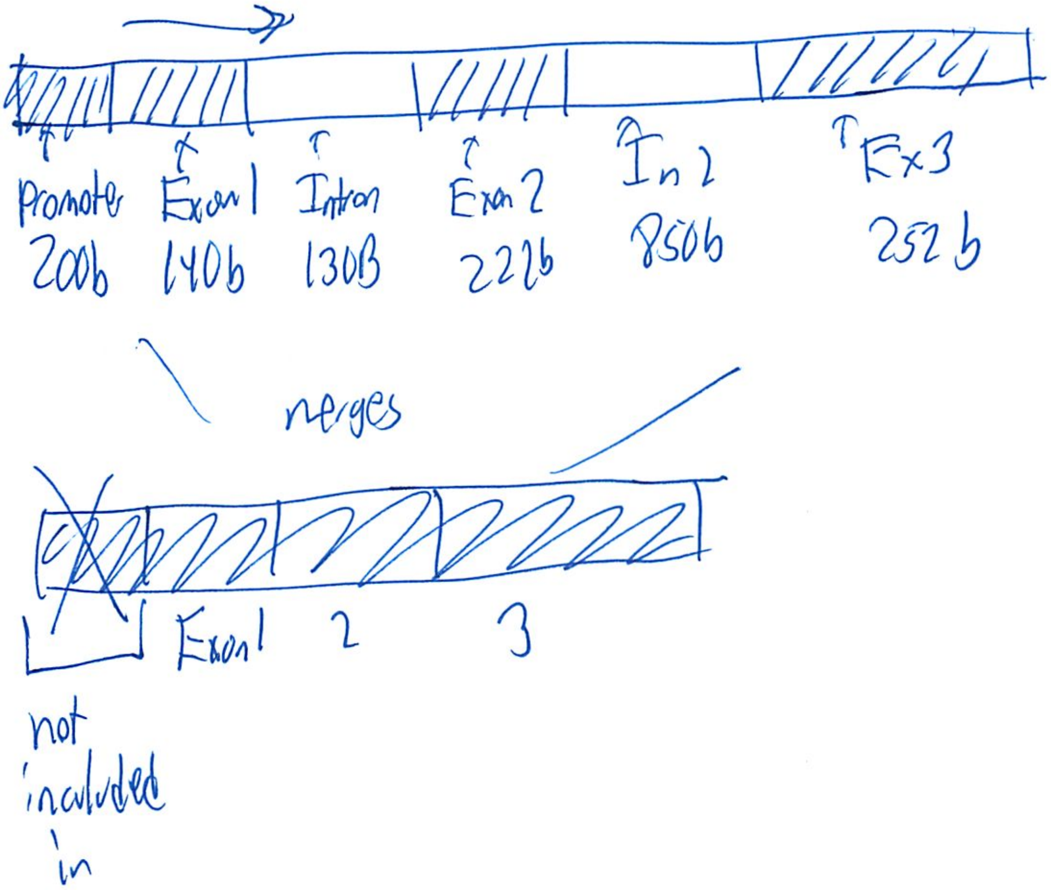
hemoglobin - transports oxy around

tetramer

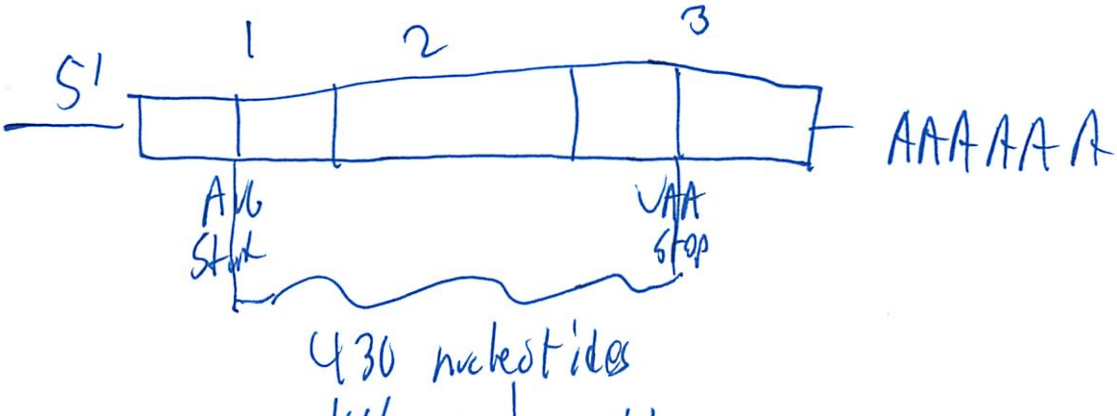

Heme
O_2
α
β
α
β
alpha globin
beta globin

(10)

β - Globin


Promote
Exon 1
Intron
Exon 2
In 2
Ex3
200b
140b
130B
222b
850b
252 b
merges
Exon 1
2
3
not
included
in

β-globin


1
2
3
5'
AAAAAA
AUG
Start
UAA
stop

430 nucleotides
146 amino acids

⑪

What could go wrong?

Single base change in DNA
- Could create a stop
- Could change amino acid
  - Could leave it the same as wt

Single base can be deleted
- Frame shift
- everything else wrong

Start codon could screw up

Entire gene could be deleted

β- globin has been studied for every possible variation

Prof: Be sure you register to vote

10/11

Samin Houshyar
R27&R28
Recitation 8

1. DNA Repair Mechanisms

2. Transcription and Translation

	Start Site	Nature of start site	Catalytic machinery	Product	Modifications
Replication					
Transcription					
Translation					

3. Post-transcriptional and post-translational modifications

4. Mutations

7.012 Recitation 8

10/11

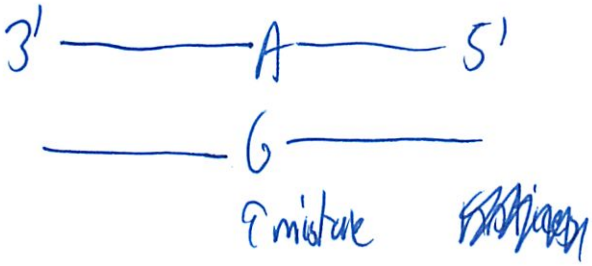
DNA Repair

Several methods

1. During replication

5' -> 3' pol activity

3' -> 5' exonuclease



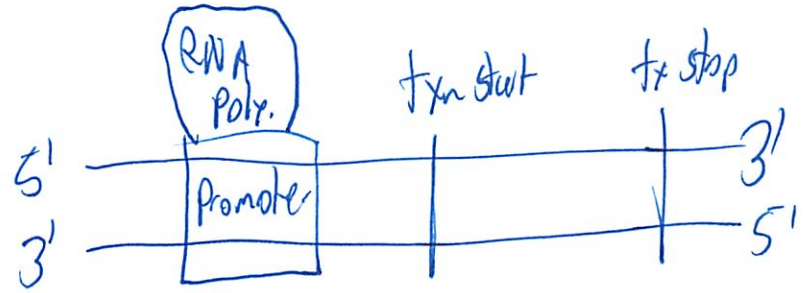
DNA polymerase recognizes kink in chain

So goes backwards

and cuts at piece and puts in proper thing

Actual mechanics + when it happens
are complicated

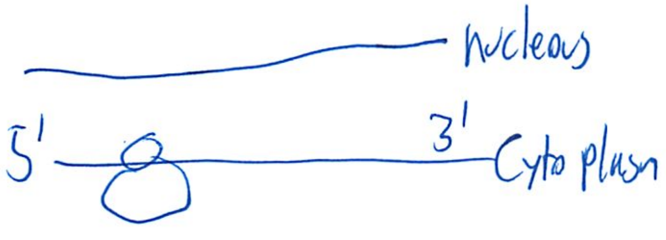
②


RNA Poly.
txn start
tx stop
5'
3'
Promoter
3'
5'

↑
start exactly
at start

RNA is made 5' → 3'

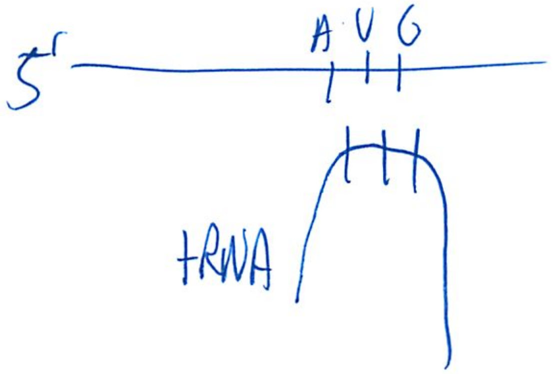
top → coding strand
bottom → ~~the~~ template strand


nucleous
5'
3'
Cyto plasm

↑ ribosome binding pt
travels down until AUG
-est frame

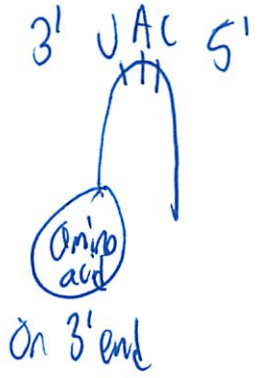
Do not start from begining!

③



* must really pay attention to direction

Codon always read 5' → 3'


3' UAC 5'
amino acid
on 3' end

Every 3 base pairs matched w/ an incoming base pair

Protien made

N → C

3' template

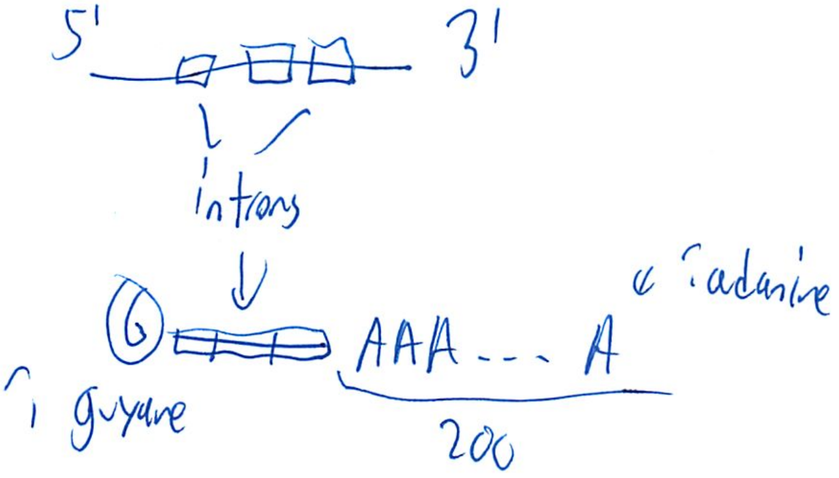
5' mRNA/coding

amino

Caboxyl

④

	Start site	Nature of Start	Catalytic Machinery	Substrates	Product	mod. of prod
Replication	origin of replication	DNA seq recognized by Helicase	DNA polymerase ↳ minor primase	dNTPs	DNA	Methylation Acetylation
Transcription	promoter	DNA seq	RNA polymerase	NTPs ↳ Genic A,U,C	RNA	Splicing ↳ Euk Capen 5' & PolyA tail 3'
Translation	AUG/ Start codon	RNA seq	Ribosame	tRNA Amino acids	Protines/ Peptide	Folding Phosphorylation Clevege Glycosylation etc



Treview

Phosphazation - activates or deactivates protein

⑤

Question 1

~~Looked at wrong one~~

1a) 90 + 222 + 126 ~~+1~~

immature (w/o splicing) 2600

mature; ~~130 + 880~~ 622

The 1 are included

90 + 222 + 126

↑ ATG inc

↑ don't include TAA

b) protiens: ~~1146~~ ~~148~~ 146

↑stop does not include

Mature should always be $\overline{3}$

3. (missed)

Coding = top

5' GUCAUGACCAACC 3'

⑥

5' UUCCAUCCAA... 3'

5' CUUCCAUCCAA 3'

(missed specifics)

They must give you a direction
Or sometimes only 1 direction

Mutation

→ DNA/RNA

- point mutation

  1 base pair changed

  substitutions

  additions

  deletions

  small change

  will focus on small mutations for now

- diff terms for what happens to protein

  Silent - still codes for same amino acids

  Missense Arg → Lys

  both ⊕ charged so hard to tell what effect - likely slight

⑦

Nonsense - change that leads to immature
Stop Codon
protien short
depends where it is how
problematic

but generally the worst

frame shift
— Caused by deletion + addition
i changes everything
most happen in non-sense mutation
will be very bad
(so what is the worst for the
P-set?)

---

3. TGA TAA (missed)
TGG TAA

G$\overset{\downarrow}{\text{C}}$T AAG
GCN TAA

2012 7.012 Recitation 8

Summary of Lectures 12 & 13:

Splicing: Splicing (which occurs in eukaryotic cells) is the process by which pieces of an mRNA initially transcribed from a gene are removed from the initial transcript to form the final, shorter transcript. These pieces, called introns, are regions of mRNA that do not code for protein. The final mRNA transcript after splicing consists only of the exons (i.e. the protein-coding regions) sandwiched between the 5′ and 3′ untranslated regions.

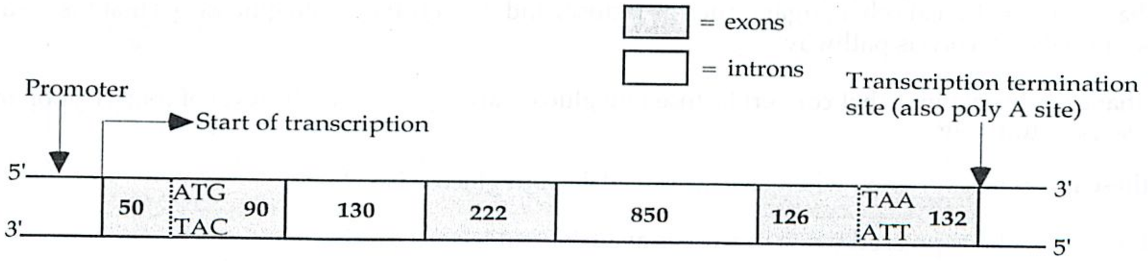
Translation: Translation begins when the ribosome orients itself at the start codon 5′-AUG-3′ on an mRNA. A tRNA that recognizes and base pairs with that codon (i.e. a tRNA with the anticodon 3′-UAC-5′) fits into the ribosome and donates the amino acid to which it is bound (methionine) to the new protein. From then on, every three nucleotides in the mRNA are read by another tRNA as a single amino acid. Each time, the next tRNA comes and fits into the ribosome, donates its amino acid to the new protein, waits for the next tRNA to come in, and then exits. This proceeds until a stop codon is reached, and the ribosome dissociates from both the mRNA and the newly made protein. Every protein has an N (amino) and a C (carboxyl) terminus, and proteins are synthesized in the N to C direction, such that the 5′ end of the mRNA corresponds to the N terminus of the protein.

Mutations: Most mutations that geneticists study are single nucleotide mutations that cause phenotypic changes. The four types of single nucleotide mutations are silent, misense, nonsense, and frameshift. A silent mutation changes a codon but does not change the amino acid encoded by that codon. A misense mutation changes the identity of the amino acid at one position. A nonsense mutation causes a protein to be truncated because a codon is changed to a stop codon. A single nucleotide frameshift is a change that either inserts or deletes a single nucleotide from the coding region of a gene, leading to a change in the reading frame of that gene.

Gene regulation: Although different cell types in your body appear and function very differently from each other, each cell type has the same DNA. The differences in appearance and function are because each cell type is expressing different RNAs and proteins; not all genes in your genome are transcribed and translated in all cells at all times. Gene regulation is the process by which the production of the final, functional product of a gene is regulated. Gene regulation is also critical for single-celled organisms, which mainly regulate their genes based on the environmental conditions under which they are growing. Organisms only produce the set of RNAs and proteins that are necessary given the temperature, salinity, and oxygen and nutrient availability of their environment. The production of the final, functional protein from a gene can be regulated at many steps. First, the gene may or may not begin to be transcribed, depending on whether activator and repressor proteins are bound to the regulatory sites of the gene. The example presented in class was the regulation of the lac operon. Second, if the gene is transcribed, RNA polymerase may or may not transcribe all the way through till the end of the gene. Third, the introns may or may not be spliced out and the message may or may not be transported from the nucleus to the cytoplasm. Fourth, the mRNA may or may not be stable. Fifth, translation of the message by the ribosome may or may not occur. Sixth, if the protein is made from the message, that protein may or may not be active. (Many proteins require different kinds of covalent modifications, such as phosphorylation, in order to be active.) Seventh, the protein may or may not be stable. Eighth, the protein may or may not need to be transported to a specific subcellular location in order to access its substrate and perform its function.

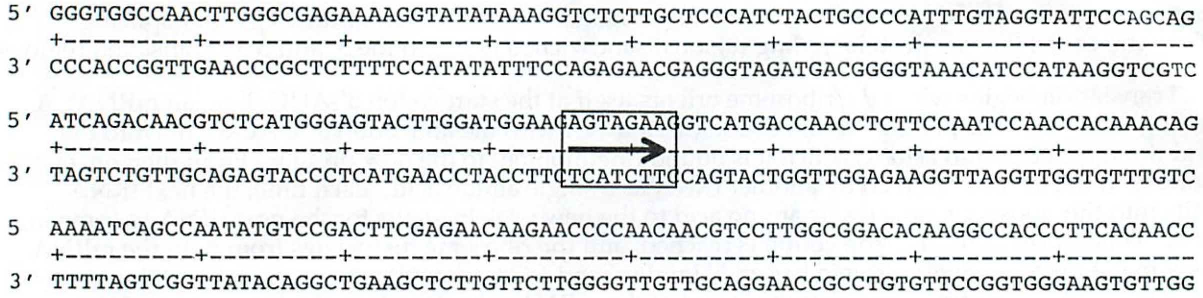
Questions:

1) Shown below is the genomic structure of the human β-globin gene. The numbers within the boxes indicate the length (in nucleotides) of each region. The DNA sequences corresponding to the start codon and the stop codon are indicated.



What is the length (in nucleotides) of the mature, processed β-globin mRNA?

2) The following is a partial sequence from the hypothetical gene, gene X. The boxed region is the promoter, and the direction of transcription is indicated by the arrow. Transcription begins at and includes the first G/C base pair after the box.



```
5' GGGTGGCCAACTTGGGCGAGAAAAGGTATATAAAGGTCTCTTGCTCCCATCTACTGCCCCATTTGTAGGTATTCCAGCAG
   +---------+---------+---------+---------+---------+---------+---------+---------
3' CCCACCGGTTGAACCCGCTCTTTTCCATATATTTCCAGAGAACGAGGGTAGATGACGGGGTAAACATCCATAAGGTCGTC

5' ATCAGACAACGTCTCATGGGAGTACTTGGATGGAAGAGTAGAAGGTCATGACCAACCTCTTCCAATCCAACCACAAACAG
   +---------+---------+---------+---------+---------+---------+---------+---------
3' TAGTCTGTTGCAGAGTACCCTCATGAACCTACCTTCTCATCTTCCAGTACTGGTTGGAGAAGGTTAGGTTGGTGTTTGTC

5  AAAATCAGCCAATATGTCCGACTTCGAGAACAAGAACCCCAACAACGTCCTTGGCGGACACAAGGCCACCCTTCACAACC
   +---------+---------+---------+---------+---------+---------+---------+---------
3' TTTTAGTCGGTTATACAGGCTGAAGCTCTTGTTCTTGGGGTTGTTGCAGGAACCGCCTGTGTTCCGGTGGGAAGTGTTGG
```

- What are the first 15 nucleotides of the mRNA produced from gene X?

- If the direction of the arrow were reversed, what would be the first 15 nucleotides of the mRNA produced from gene X?

3) Drawn below is part of a wild-type gene. The DNA sequence shown encodes the last amino acids of a protein that is normally 380 amino acids long. The **bold & underlined** codon indicates the correct reading frame of this gene. The lower strand of the gene is used as the template during the transcription of mRNA from this gene.

...**<u>GCT</u>**AAGTATTGCTCAAGATTAGGATGATAAATAACTGG–3'
...CGATTCATAACGAGTTCTAATCCTACTATTTATTGACC–5'

a) In the copy of the sequence drawn below, circle one base pair that you could change to make a mutant form of the gene that produces a protein that is now 381 amino acids long. Indicate the identity of one new base pair that could take its place.

b) In the copy of the sequence drawn below, draw a slash between two base pairs where you could add one extra base pair in order to make a single mutant form of the gene that produces a protein that is 373 amino acids long. Indicate the identity of the one new base pair you are adding.

4) For each of the following types of mutations, state how they impact the final gene product.

a) Silent: b) Missense c) Frameshift: d) Nonsense:

5) The bacterium *E. coli* eats sugar for energy. The easiest sugar for bacteria to eat is glucose, because glucose can be fed directly into the pathway of glycolysis, which extracts energy (in the form of ATP) from glucose. However, bacteria can also eat other sugars (such as lactose) and convert them into glucose, so that the resulting glucose goes into the glycolysis pathway.

The genes that encode enzymes that convert lactose into glucose are regulated at the level of transcription in way that makes sense intuitively.

a) Should these genes be expressed when lactose is available but glucose isn't?

b) Should these genes be expressed when glucose is available but lactose isn't?

c) Should these genes be expressed when both lactose and glucose are present?

d) Should these genes be expressed when neither lactose nor glucose are present?

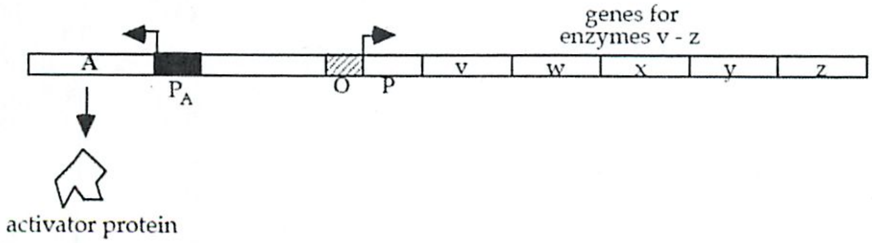
e) Do you think that a lactose conversion enzyme would be expressed from its gene in cells with the following mutant properties if the cells are grown under these conditions:

	In the presence of neither glucose nor lactose	In the presence of lactose only
The repressor protein cannot bind to DNA		
The promoter of the gene is mutated so that RNA polymerase cannot bind to it *		
The site on the DNA where the repressor protein binds is mutated		
The repressor protein can no longer bind lactose		
The repressor is locked into the conformation it adopts when lactose is present		
RNA polymerase is not functional **		

i. What do you think the phenotype of the * cell would be?

ii. What do you think the phenotype of the ** cell would be?

6) You also discover that the genes involved in the synthesis of a compound 3 are organized into an operon. A schematic is shown below. Assume that all the enzymes v - z are needed to produce the compound 3. The regulatory protein is an activator and compound 3 can bind to the activator to influence the association of the activator protein to the O region.



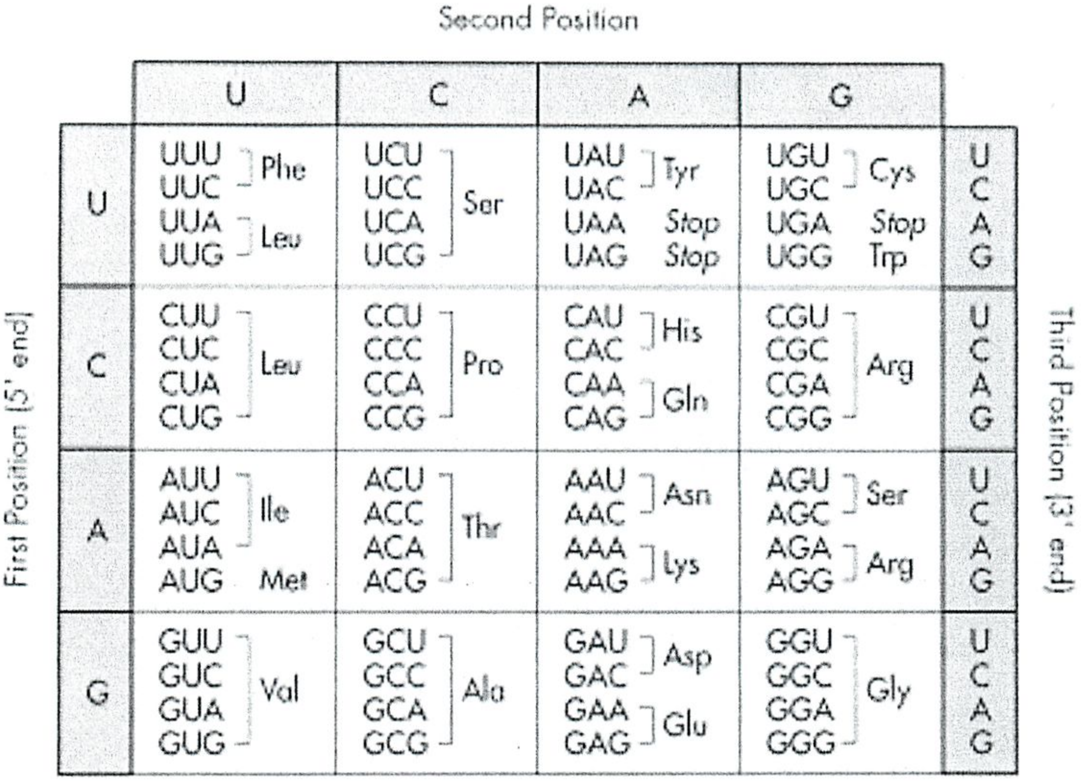
You have a strain carrying a mutation in P ($A^+ P_A^+ P^- O^+ v^+ w^+ x^+ y^+ z^+$). You make partial diploids with the following plasmids. In each case, indicate if the plasmid restores the red pigment. **Explain** your answer.

Plasmid

a) $A^+ P_A^+ P^+ O^+ v^- w^+ x^+ y^+ z^+$

b) $A^+ P_A^- P^+ O^+ v^+ w^+ x^+ y^+ z^+$

Each codon of an mRNA represents an amino acid or a stop codon as shown by the Codon Chart below.



Second Position

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

Doing P-set 3

10/9

2c) UUGC?

~~CCAT~~

AACG

A=T (U)
C=G

¿why is it U?

thought that was only RNA

And this is replication, not transcription

d) RNA primase

Starts each template strand

RNA/DNA builds from

↷?

3a) Direction of Transcription

1 strand

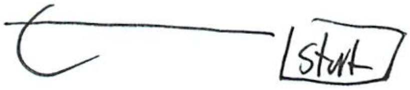
5' → 3' output

moves along 3' → 5'

(I don't get the picture...)

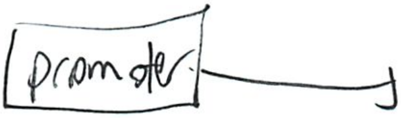
②

top must be template if moving



if right

Otherwise



Coding strand the non-template strand
Called that since it matches the RNA
except U for T

What is N or C?
amino or carboxyl?

tRNA section
thats the attachment thing

amino acids

⑧

Placement

←

E P A

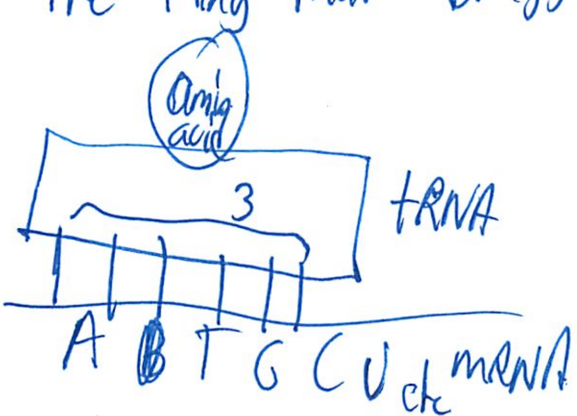
5' 3'

I'm not too sure w/ this diagram

---

9. mRNA = made from DNA

tRNA = the thing that brings the amino acid


amino acid
3
tRNA
A B T G C U
ch
mRNA

read table from mRNA?

So the E P A

5' ← 3'

P has the stack

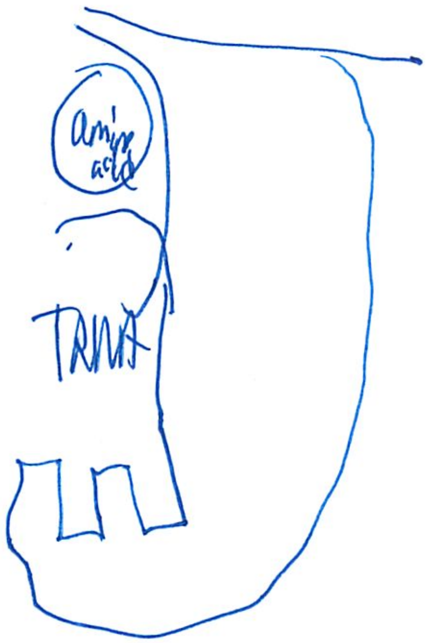
transfers stack to A

A moves to P

⑨

4c) ~~*~~ table shows mRNA 5' → 3'

4f) amino acyl- tRNA synthetase



5' 3'

~~tRNA~~ trp CCA 3' GCA anti

3' GGU CGU cod

UGG 3' UGC

no it also attaches UGC instead of Cys

isn't that Cys

wp: Semi essential amino acids

└how are we supposed to know?

⑤

5a) Since coding seq → just match

* except U for T

↳ saw when looked at b

c) Type of mutation

missence - new nucleotide alters codon to change amino acid

nonsense one that changes to stop

silent = same amino acid

frame shift - insert/delete when not power of 3

deletion not on there → i[illegible] delete when 3

didn't read the other types

(6)

(6) Where does transcription start rel to tl site?

don't want a +1 error

very confused

i do we start?

Ah can work back from stop!

? look for ATG

⑦ w/ Anji 10/10
10:20

b) to right of

c) ~ 10 in

AUG on the coding strand

(always AUG → at least for pset

e) Untill TAA - again on coding seq

f) Same

3b) write top

4f) 10^3 - 10^4 protiens

if evenly distributed 100 - 1000

8

10/20
10:46

Wp: AUC most common

(: not lining up?

Ask folks in recitation

10/11

3a) MRNA

a lot of ribosomes at once

4f) 1/20 of amio acids

~~[illegible]~~ =ly distibuted

Name Michael Plasmeier Section 27 TA Horshyar

2012 7.012 Problem Set 3

Please print out this problem set and answer the questions on the printout. Answers to this problem set are to be turned in at the box outside 68-120 before 4:00 PM, Thursday October 11th.

9/17

Question 1

Briefly describe the experiments performed by each of the followings researchers, and in one sentence summarize the important findings of each experiment.

a) Frederick Griffith, 1928:

A dead virulent bacteria and a living non virulent bacteria produce a live virulent bacteria – indicating there is an underlying inherited "transforming principles."

b) Oswald Avery, *et. al*, 1943-44

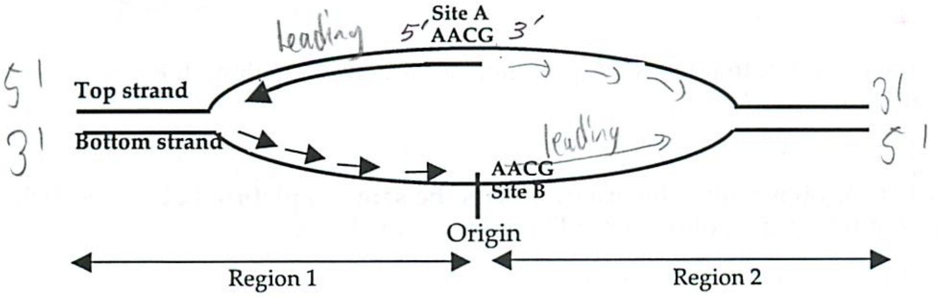
Avery showed that only DNase was able to prevent the "transforming" of having the virulent form continue to show up. – So transforming principle was DNA

c) Alfred Hershey and Martha Chase, 1952

Hershey + Chase colored the Phosphor in DNA in bacteriaphages and saw that it was present in the pellet of bacteria cells when blended – showing again the hereditary info was in DNA

Question 2

Shown below is a schematic of replicating DNA in a bacterial cell.



a) On the diagram, label the 5' and the 3' ends of the parental DNA strands. ✓

b) Which parental DNA strand (*top or bottom*) serves as a template for the synthesis of the leading strand in Region 2?

bottom

0/1

c) To which site (*A, B, or both*) can the primer **5' UUGC 3'** bind?

both

↑ I don't think I looked at direction

Name________________________________ Section______ TA___________

Question 2, continued

d) The replication of which strand (*top, bottom, or both*) in Region 2 would be affected in the absence of RNA primase? Assume that replication has not yet initiated on either strand. Explain.

1/1

both - need the primase to get started

e) You perform DNA replication in a test tube (in vitro) using a **single-stranded linear DNA** as the template and the **appropriate DNA primer.** From the list below, circle the proteins that are required for **one round** of replication.

→ ~~**Primase**~~ [✓ **DNA polymerase**] ~~**Ribonuclease**~~ ~~**Topoisomerase**~~ ~~**Ligase**~~

DNA repl

single strand = no helix

f) Why does the DNA of a eukaryotic cell require multiple origins of replication when some prokaryotic cell genomes have only one origin of replication?

Multiple strands/chromosomes
much longer

g) While studying replication you find a mutant in which the fidelity of replication has decreased by a factor of 100. You suspect that this is due to a mutation in the DNA polymerase enzyme. What specific enzymatic activity of the DNA polymerase allows it to proofread the newly replicated DNA molecule?

0/1

normally it reads the DNA - if a mistake it can back up and go backwards

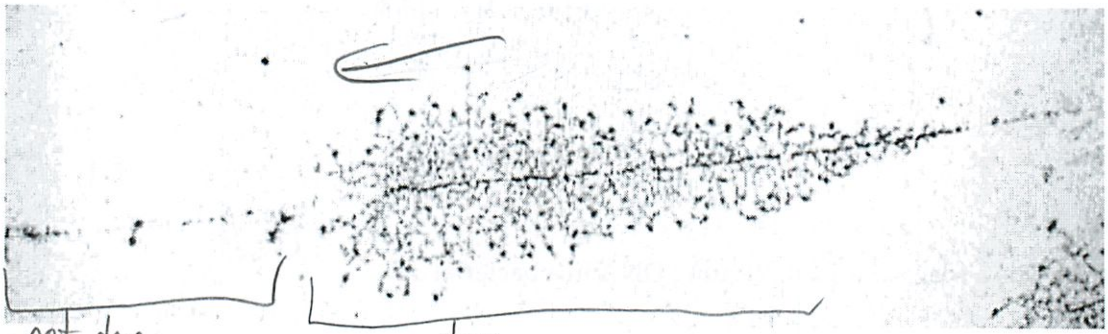
3'→5' exonuclease

looking for specific ans

Question 3

Below is an electron micrograph of a single gene being transcribed. The DNA strand runs horizontally with RNA transcripts extending vertically outward.

1/1



a) Draw an arrow indicating the direction that the RNA polymerases are moving along the DNA strand. Why did you choose this direction?

Since the more particles on the right - it looks like that side is already "done"

b) Below is a partial sequence of the above gene. **Its orientation is the same as pictured above.** Which strand is the template strand, the top or the bottom strand? Explain your choice.

top

$^{5'}$ACTCGATGCTAG$^{3'}$ ← template
$^{3'}$TGAGCTACGATC$^{5'}$

Since we go 3'→5' along the template, while we produce RNA 5'→3'

c) What would be the mRNA sequence transcribed from the above sequence? Be sure to label the 5' and 3' ends.

5' CUAGCAUCGAU 3'
→

1/1

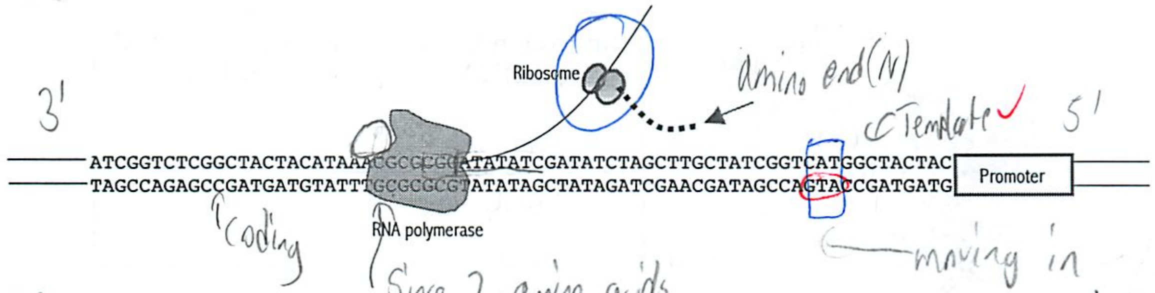
Name______________________________ Section_______ TA___________

Question 3 continued

d) Complete the diagram below by… Transcription

- Labeling 5′ and 3′ on the mRNA.
- Labeling the arrow with either the N or the C to indicate the terminus of the protein.
- Boxing the 3 bases encoding the first amino acid of the protein being made.
- Labeling the template strand for transcription.
- Circling the part of the schematic where tRNAs would bind.

0/2


3'
Ribosome
amino end(N)
Template ✓
5'
ATCGGTCTCGGCTACTACATAAACGCGCGCATATATCGATATCTAGCTTGCTATCGGTCATGGCTACTAC
TAGCCAGAGCCGATGATGTATTTGCGCGCGTATATAGCTATAGATCGAACGATAGCCAGTACCGATGATG
Promoter
coding
RNA polymerase
moving in

Since 2 amino acids goes over the 3rd

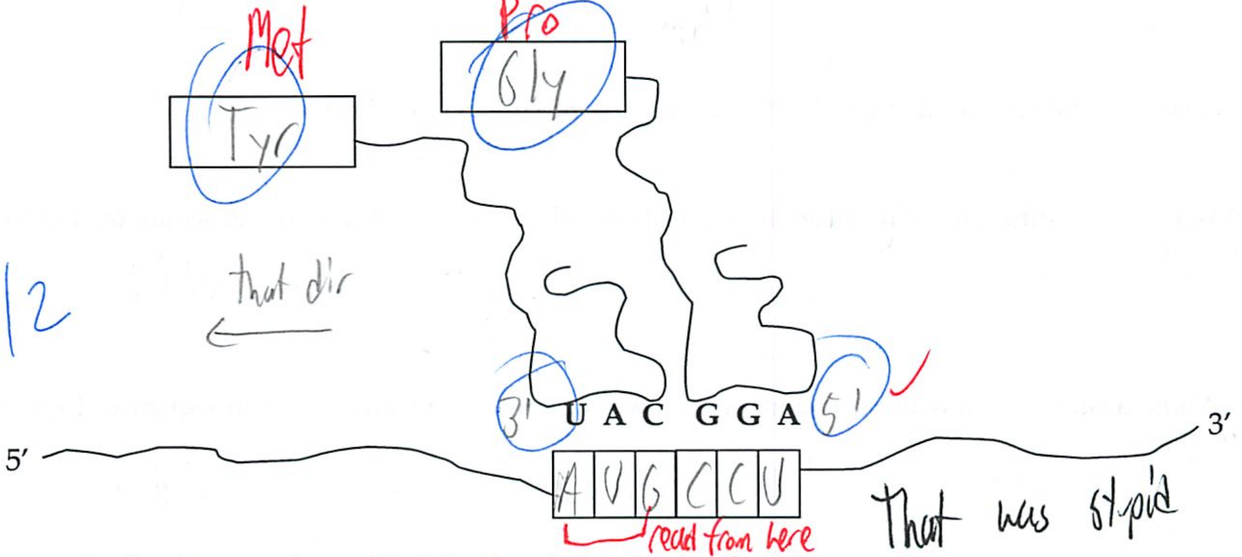
was confused by

Question 4

a) Below is a diagram of two tRNAs and an mRNA in the active site of the ribosome during translation of the mRNA into protein. Three nucleotides from the sequence of each tRNA are shown for you.

0/2

that dir


Met
Tyr
Pro
Gly
3'
U A C G G A
5'
5′
3′
A U G C C U
read from here

That was stupid

- In the diagram above, label the 5′ and 3′ ends of each tRNA.

always opposite

- In the diagram above, fill in the boxes in the mRNA with the 6 nucleotides that would be present there. ✓

- In the diagram above, fill in the box attached to one end of each tRNA with the name of the amino acid that would be attached there.

look at mRNA?

- Which tRNA is about to transfer its attached amino acid over to the other tRNA: the tRNA on the left or the tRNA on the right?

The one at P transfers strand to A
Tyr Gly

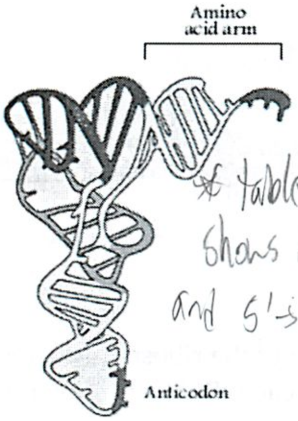
the left ✓

Name______________________ Section_______ TA____________

Question 4, continued

A tRNA molecule is composed of an RNA chain that folds into a 3-D shape like that shown below. At one end it has an anti-codon that base pairs with the appropriate codon on the mRNA and at the other end it has an amino acid arm that binds to a specific amino acid.

b) Below are three anti-codon sequences for three tRNAs, fill in the corresponding amino acid on the blanks.



* table shows mRNA and 5'→3'

anticodon found on tRNA	amino acid attached to tRNA
5′ AGU 3′ ACU 3' UCA 5'	Thr
5′ AUG 3′ CAU UAC	His
5′ CUG 3′ CAG GAC 5	Gln

c) Give the anticodon used in the tRNA encoding trp. Be sure to label the 5′ and 3′.

5' UGG 3' mRNA
3' ACC 5' tRNA ←

d) Would a substitution within a codon for trp always change the resulting protein sequence? Explain your answer.

Yes since only 1 codon codes for TRP (UGG)

e) Would a substitution within a codon for thr always change the resulting protein sequence? Explain your answer.

No, some times there are multiple valid codons for an amino acid → CCC vs CCA

f) An aminoacyl tRNA synthetase is an enzyme that attaches a specific amino acid to the appropriate tRNAs to form an aminoacyl-tRNA. This is sometimes called "charging" the tRNA with the amino acid. Assume you have a cell with a mutation in the gene for the tryptophan aminoacyl tRNA synthetase. This mutant enzyme attaches tryptophan to tRNAs with the anticodons 5′ CCA 3′ and 5′GCA 3′. Explain how protein production in this cell will be altered and estimate how many different types of proteins would be affected inthis cell. Choose from: >10, 10-100, 100-1000, all or the proteins in the cell.

This would bind 5' UGC 3' codon to Trp instead of Cys. This would break the protines that have UGC → Cys ↳ alter Function

It would depend on how many protiens rely on Cys. From WP, it seems that few do >10

Name________________________________ Section_______ TA____________

Question 5

Below is a partial sequence of a coding region, base pairs 61-102 (read left to right) of a 600 base pair open reading frame. The underlined codon indicates the correct reading frame of this gene.

5′ ATCTGGGCTAATACCGCCAACTATATAAACACCCACATTTCG 3′
3′ TAGACCCGATTATGGCGGTTGATATATTTGTGGGTGTAAAGC 5′

a) What is the mRNA sequence encoded by base pairs 61-71?

AUC UGG GCU AA

b) What is the amino acid sequence of the peptide encoded by base pairs 61-69?

Ile Trp Ala

c) How does the resulting peptide change if the sequence is altered as shown below? Also identify the type of mutation, choose from missense, nonsense, silent, frame-shift, or deletion.

i)	original:	5′ ATCTGGGCTAA**T**ACCGCCAACTATATAAACACCCACATTTCG 3′
	altered:	5′ ATCTGGGCTAA**C**ACCGCCAACTATATAAACACCCACATTTCG 3′

AAT = Asn
AAC = Asn
Silent

ii)	original:	5′ ATCTGGGCTAATACCGCCAACTATA**T**AAACACCCACATTTCG 3′
	altered:	5′ ATCTGGGCTAATACCGCCAACTATT**A**AAACACCCACATTTCG 3′

T altered too

AUA = start
UAA - Stop
nonsense

iii)	original:	5′ ATCTGGGCTAA**T**ACCGCCAACTATATAAACACCCACATTTCG 3′
	altered:	5′ ATCTGGGCTAA**A**ACCGCCAACTATATAAACACCCACATTTCG 3′

AAU = Asn
AAA = Lys
missence

iv)	original:	5′ ATCTGGGCTAATACCGCCAACTATATAAACACCCACATTTCG 3′
	altered:	5′ ATCTGGGCTAATACC------TATATAAACACCCACATTTCC 3′

(delete 6 base pairs)

missing Ala Asn deletion

v)	original:	5′ ATC--TGGGCTAATACCGCCAACTATATAAACACCCACATTTCG 3′
	altered:	5′ ATCATTGGGCTAATACCGCCAACTATATAAACACCCACATTTCC 3′

(insert 2 base pairs)

redo rest as not a power of 3 - major changes
frame shift

d) Of the various mutations given above, which the one(s) would most dramatically affect the **function** of the protein encoded by this gene? Explain your answer.

V - Since it is not a multiple of 3 - it would cause everything to shift

Name_________________________ Section_______ TA___________

Question 6

Shown below is a double-stranded bacterial (*E. coli*) DNA sequence coding for a hypothetical protein. Both strands are shown; the top strand reads 5′ to 3′ left to right, while the bottom strand reads 5′ to 3′ right to left. The nucleotides are numbered from 1 to 100. For this problem, transcription begins with and includes the underlined A/T base-pair (indicated with an "a") and proceeds left to right.

```
                  1                  20   ↓Start               40
                  |         a  11    |                          |
Coding    5'-GTGTCCGTATAATATTGTGAGATGTTATATCCCGCCGTCAACACCATCAA-3'
             ---------+---------+---------+---------+---------+
template  3'-CACAGGCATATTATAACACTCTACAATATAGGGCGGCAGTTGTGGTAGTT-5'

                  60                  80                  100
                  |      c            |                    |
          5'-ACAGGATAATCGCCTGCTGGGCAAAGGCGGTGAAGGTAAAGGTGTTGCC-3'
             ---------+---------+---------+---------+---------+
          3'-TGTCCTATTAGCGGACGACCCGTTTCCGCCACTTCCATTTCCACAACGG-5'
```

a) Which strand is used as a template for transcription, the top or the bottom?

botom - since told left to right - must be 3' to 5'

b) Where would the promoter be relative to base pair #1?

about 8 pairs in to the right   Always at the 5' end

c) What are the first 10 nucleotides of the resulting mRNA? Indicate the 5' and 3' ends of the mRNA.

0/1

5' UUA UAU CCC G 3'   starts at start → not ATG

d) What are the first 5 amino acids translated from the resulting mRNA? Indicate the amino (NH_3^+) and carboxy (COO^-) termini of the protein.

0/1

Met 1st NH_3^+ Met Leu Tyr Pro Ala ~~Val~~ COO^-

e) Do the underlined nucleotides TAA (indicated with the letter "**c**") encode a stop codon for this protein? Briefly explain your answer.

No - it is not a power of 3, A later TAA is used

Consider the situations in parts (f-h) independently.

f) A mutation occurs which results in the insertion of an extra G/C (top strand/bottom strand) base-pair immediately after base pair 11 (shown in bold). What effect will this insertion mutation have on transcription and translation?

transcription - none other than obvious

translation - ruins start seq - so protein not made   always starts ATG (that starts the align by 3)

g) A different mutation results in the substitution of the T/A base pair at position 30 (shown in bold and underlined) with a G/C base pair. How would this mutation affect the sequence of the protein that is produced?

TAT → TAG   Tyr → Stop so amino acid chain would only be half produced - no protien

h) A third mutation occurs which results in the substitution of the C/G base pair at position 42 (shown in bold italics) to a T/A base pair. How would this mutation affect the sequence of the protein that is produced?

AAC → AAT

Asn → Asn   Silent mutation, no change in protien

I think they take too much off here...

Name______________________________ Section______ TA__________

Each codon of an mRNA represents an amino acid or a stop codon as shown by the Codon Chart below.

Second Position

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

Solutions for 2012 7.012 Problem Set 3

Question 1

Briefly describe the experiments performed by each of the followings researchers, and in one sentence summarize the important findings of each experiment.

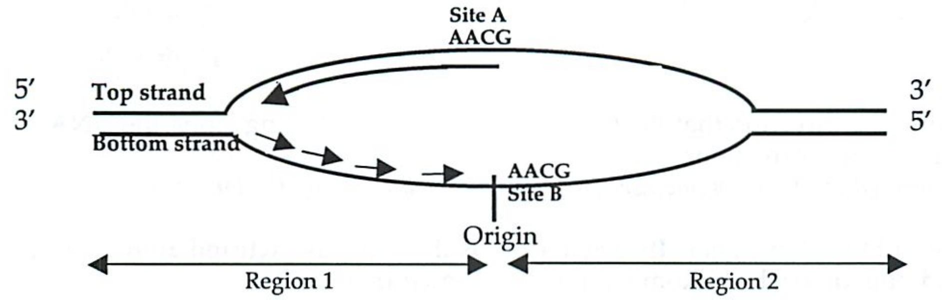
a) Frederick Griffith, 1928:
Griffith used two strains of S. pneumoniae: A smooth and virulent strain and a rough and non-virulent strain. When he injected the smooth strain alone into mice, the mice got pneumonia and died. When he injected the rough strain alone into mice, the mice lived. When he injected heat killed smooth bacteria, the mice lived. However, when he injected heat killed smooth bacteria and live rough bacteria together into the mice, the mice died of pneumonia! The dead mice were found to contain live smooth bacteria.
Important finding: There is some non-protein genetic material that can transform the rough bacteria into the smooth bacteria.

b) Oswald Avery, *et. al*, 1943-44
Avery found that when he put heat killed smooth bacteria and live rough bacteria on the same petri dish, he was able to get smooth colonies. He also found that when he made extract from the dead smooth bacteria and separated the extract into fractions, it was the DNA fraction that was able to transform the live rough bacteria into smooth bacteria.
Important finding: DNA is the genetic material that transforms the rough bacteria into the smooth bacteria.

c) Alfred Hershey and Martha Chase, 1952
Hershey and Chase grew two batches of T2 bacteriophage, one with radioactive sulfur and one with radioactive phosphate. They used these two batches of phage to infect E. coli separately. The separate batches were then agitated in a kitchen blender to separate any viral parts that were not inside the E. coli from the bacteria. The mixtures were centrifuged and the sulfur labeled batch only showed radioactivity in the supernatant while the phosphate labeled batch only showed radioactivity in the pellet.
Important finding: It was the DNA that went into the bacteria and is the transforming principal.

Question 2

Shown below is a schematic of replicating DNA in a bacterial cell.



a) On the diagram, label the 5′ and the 3′ ends of the parental DNA strands.

b) Which parental DNA strand (*top or bottom*) serves as a template for the synthesis of the leading strand in Region 2?
The *Bottom* strand.

c) To which site (*A, B, or both*) can the primer **5′ UUGC 3′** bind?
Only site B.

Question 2, continued

d) The replication of which strand (*top, bottom, or both*) in Region 2 would be affected in the absence of RNA primase? Assume that replication has not yet initiated on either strand. Explain.
Both strands will be affected, as primers are needed to initiate replication of both the leading and the lagging strands.

e) You perform DNA replication in a test tube (in vitro) using a **single-stranded linear DNA** as the template and the **appropriate DNA primer.** From the list below, circle the proteins that are required for **one round** of replication.

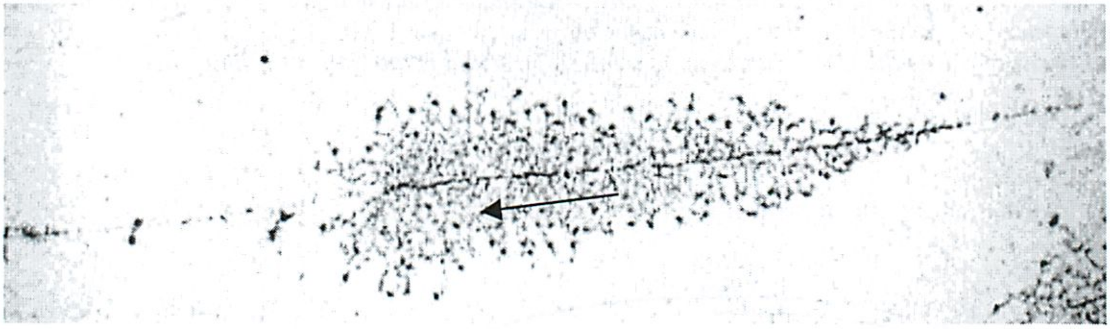
Primase [✓ **DNA polymerase**] **Ribonuclease** **Topoisomerase** **Ligase**

f) Why does the DNA of a eukaryotic cell require multiple origins of replication when some prokaryotic cell genomes have only one origin of replication?
Prokaryotic genomes are often a single circular chromosome, and as such, a single ORI is sufficient. Eukaryotic genomes tend to be much bigger, and are often divided into multiple chromosomes. Every chromosome would require at least one ORI, and generally each chromosome has many ORIs, which decreases the time it takes to copy the genome.

g) While studying replication you find a mutant in which the fidelity of replication has decreased by a factor of 100. You suspect that this is due to a mutation in the DNA polymerase enzyme. What specific enzymatic activity of the DNA polymerase allows it to proofread the newly replicated DNA molecule?
The 3′ → 5′ exonuclease.

Question 3

Below is an electron micrograph of a single gene being transcribed. The DNA strand runs horizontally with RNA transcripts extending vertically outward.



a) Draw an arrow indicating the direction that the RNA polymerases are moving along the DNA strand. Why did you choose this direction?
The longest mRNA will be associated with the polymerases that have been transcribing the longest time.

b) Below is a partial sequence of the above gene. **Its orientation is the same as pictured above.** Which strand is the template strand, the top or the bottom strand? Explain your choice.

$^{5'}$ACTCGATGCTAG$^{3'}$
$^{3'}$TGAGCTACGATC$^{5'}$

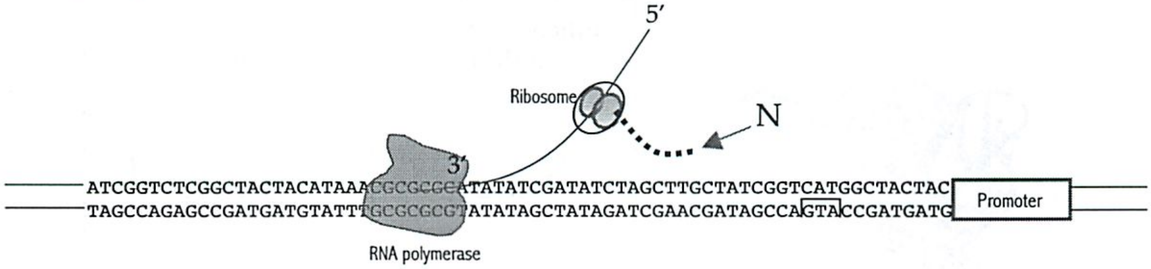
The top strand is the template strand. We know from part a) that the RNA polymerase has to go from right to left, and we also know that mRNA synthesis has to happen from 5′—3′. Thus, the template strand should go 3′—5′ from right to left. The top strand satisfies this requirement.

c) What would be the mRNA sequence transcribed from the above sequence? Be sure to label the 5′ and 3′ ends.
5′-CUAGCAUCGUCGAGU-3′

Question 3 continued

d) Complete the diagram below by…

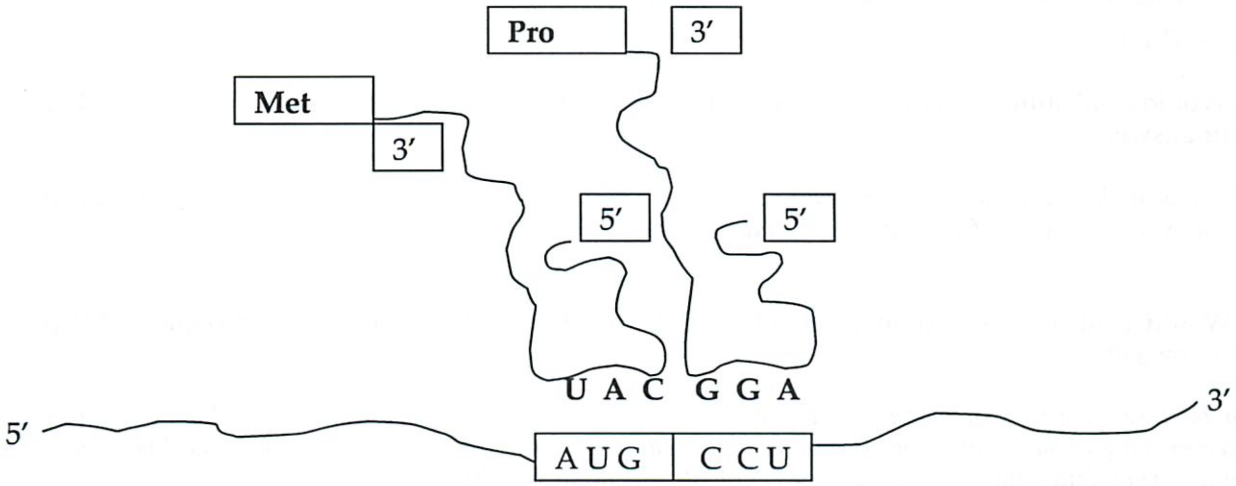
- Labeling 5′ and 3′ on the mRNA.
- Labeling the arrow with either the N or the C to indicate the terminus of the protein.
- Boxing the 3 bases encoding the first amino acid of the protein being made.
- Labeling the template strand for transcription.
- Circling the part of the schematic where tRNAs would bind.


5′
Ribosome
N
3′
ATCGGTCTCGGCTACTACATAAACGCGCGCATATATCGATATCTAGCTTGCTATCGGTCATGGCTACTAC
TAGCCAGAGCCGATGATGTATTTGCGCGCGTATATAGCTATAGATCGAACGATAGCCAGTACCGATGATG
Promoter
RNA polymerase

The top strand is the template strand.

Question 4

a) Below is a diagram of two tRNAs and an mRNA in the active site of the ribosome during translation of the mRNA into protein. Three nucleotides from the sequence of each tRNA are shown for you.

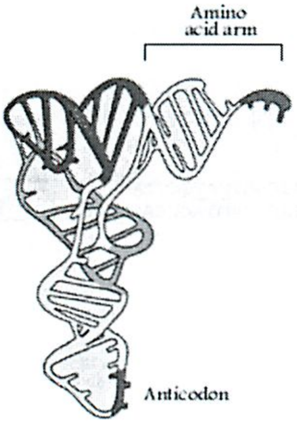

Pro
3′
Met
3′
5′
5′
U A C
G G A
5′
3′
AUG
C CU

- In the diagram above, label the 5′ and 3′ ends of each tRNA.
- In the diagram above, fill in the boxes in the mRNA with the 6 nucleotides that would be present there.
- In the diagram above, fill in the box attached to one end of each tRNA with the name of the amino acid that would be attached there.
- Which tRNA is about to transfer its attached amino acid over to the other tRNA: the tRNA on the left or the tRNA on the right?
  The tRNA on the left*. This allows the tRNA to the left, to leave the ribosome and the peptide to continue being synthesized on the ribosome.*

Question 4, continued

A tRNA molecule is composed of an RNA chain that folds into a 3-D shape like that shown below. At one end it has an anti-codon that base pairs with the appropriate codon on the mRNA and at the other end it has an amino acid arm that binds to a specific amino acid.

b) Below are three anti-codon sequences for three tRNAs, fill in the corresponding amino acid on the blanks.



anticodon found on tRNA	amino acid attached to tRNA
$^{5'}$AGU$^{3'}$	*Threonine*
$^{5'}$AUG$^{3'}$	*Histidine*
$^{5'}$CUG$^{3'}$	*Glutamine*

c) Give the anticodon used in the tRNA encoding trp. Be sure to label the 5′ and 3′.

5′-CCA-3′

d) Would a substitution within a codon for trp always change the resulting protein sequence? Explain your answer.

***Yes** because there is only one codon that encodes Trp, a base substitution in either the 1,2, or 3 position of the codon would result in either a different amino acid or a stop codon.*

e) Would a substitution within a codon for thr always change the resulting protein sequence? Explain your answer.

***No**, because there are four codons that encode for Thr, all four codons have the same base in the 1 and 2 position, however, they all have a different base in the 3 position. Therefore a substitution of a base at the 3 postion will not change the protein sequence because the codon will still encode for Thr.*

f) An aminoacyl tRNA synthetase is an enzyme that attaches a specific amino acid to the appropriate tRNAs to form an aminoacyl-tRNA. This is sometimes called "charging" the tRNA with the amino acid. Assume you have a cell with a mutation in the gene for the tryptophan aminoacyl tRNA synthetase. This mutant enzyme attaches tryptophan to tRNAs with the anticodons 5′ CCA 3′ and 5′GCA 3′. Explain how protein production in this cell will be altered and estimate how many different types of proteins would be affected in this cell. Choose from: >10, 10-100, 100-1000, or all the proteins in the cell.

In this cell, a tRNA with the anticodon 5′GCA 3′ will be charged with either the normal cys or with trp. Therefore any protein in the cell that has a codon recognized by the tRNA with the anticodon 5′GCA 3′ could have a trp in place of a cysteine. Thus any protein containing cysteine will be affected.

Question 5

Below is a partial sequence of a coding region, base pairs 61-102 (read left to right) of a 600 base pair open reading frame. The underlined codon indicates the correct reading frame of this gene.

5′ ATCTGGGCTAATACCGCCAACTATATAAACACCCACATTTCG 3′
3′ TAGACCCGATTATGGCGGTTGATATATTTGTGGGTGTAAAGC 5′

a) What is the mRNA sequence encoded by base pairs 61-71?
5′ AUCUGGGCUAA 3′

b) What is the amino acid sequence of the peptide encoded by base pairs 61-69?
Ile-trp-ala

c) How does the resulting peptide change if the sequence is altered as shown below? Also identify the type of mutation, choose from missense, nonsense, silent, frame-shift, or deletion.

i) original: 5′ ATCTGGGCTAA**T**ACCGCCAACTATATAAACACCCACATTTCG 3′
altered: 5′ ATCTGGGCTAA**C**ACCGCCAACTATATAAACACCCACATTTCG 3′

In this case both AAT and AAC will encode asn, so the resulting peptide is the same. This is a silent mutation.

ii) original: 5′ ATCTGGGCTAATACCGCCAACTATA**T**AAACACCCACATTTCG 3′
altered: 5′ ATCTGGGCTAATACCGCCAACTATT**A**AAACACCCACATTTCG 3′

In this case ATA (ile) has been changed to TAA, which encodes a stop codon. The resulting peptide will terminate after amino acid 28. This is a nonsense mutation.

iii) original: 5′ ATCTGGGCTAA**T**ACCGCCAACTATATAAACACCCACATTTCG 3′
altered: 5′ ATCTGGGCTAA**A**ACCGCCAACTATATAAACACCCACATTTCG 3′

In this case both AAT (asn) has been changed to AAA (Lys). The resulting peptide will be different by one amino acid at amino acid position 24. This is a missense mutation

iv) original: 5′ ATCTGGGCTAATACCGCCAACTATATAAACACCCACATTTCG 3′
altered: 5′ ATCTGGGCTAATACC------TATATAAACACCCACATTTCC 3′
(delete 6 base pairs)

In this case two codons have been deleted CGG (ala) and AAC (asn). The resulting peptide will be shorter by two amino acids. This is a deletion mutation, however the reading frame remains the same. The last codon has also been changed from UCG to UCC, but this does not change the protein.

v) original: 5′ ATC--TGGGCTAATACCGCCAACTATATAAACACCCACATTTCG 3′
altered: 5′ ATCATTGGGCTAATACCGCCAACTATATAAACACCCACATTTCC 3′
(insert 2 base pairs)

In this case two base pairs have been deleted and the reading frame of the protein has been shifted. The resulting peptide will be normal through amino acid 20, but will have a different amino acid sequence from that point on. This frame shift introduces a stop codon at amino acid 29.

d) Of the various mutations given above, which the one(s) would most dramatically affect the **function** of the protein encoded by this gene? Explain your answer.

Mutations ii) and v) both cause greatly shortened polypeptides, neither of which would likely function. Introduction of a stop codon will cause the protein to be truncated wherever the stop codon is introduced. Introduction of a frameshift will result in every amino acid after the insertion or deletion to be different than what it is supposed to be. Furthermore, frameshift mutations often result in an introduction of a stop codon as we saw in the above example. Silent, missence, and even small deletion that keep frame, are usually tolerated better than nonsence and frameshift mutations.

Question 6

Shown below is a double-stranded bacterial (*E. coli*) DNA sequence coding for a hypothetical protein. Both strands are shown; the top strand reads 5′ to 3′ left to right, while the bottom strand reads 5′ to 3′ right to left. The nucleotides are numbered from 1 to 100. For this problem, transcription begins with and includes the underlined A/T base-pair (indicated with an "a") and proceeds left to right.

```
   1                 20                  40
   |       a          |                   |
5'-GTGTCCGTATAATATTGTGAGATGTTATATCCCGCCGTCAACACCATCAA-3'
   ---------+---------+---------+---------+---------+
3'-CACAGGCATATTATAACACTCTACAATATAGGGCGGCAGTTGTGGTAGTT-5'

   60                80                  100
   |      c           |                   |
5'-ACAGGATAATCGCCTGCTGGGGCAAAGGCGGTGAAGGTAAAGGTGTTGCC-3'
   ---------+---------+---------+---------+---------+
3'-TGTCCTATTAGCGGACGACCCCGTTTCCGCCACTTCCATTTCCACAACGG-5'
```

a) Which strand is used as a template for transcription, the top or the bottom?
The bottom strand. Because the question tells us that transcription proceeds left to right and RNA Polymerase goes 5′ to 3′ the bottom strand must be the template.

b) Where would the promoter be relative to base pair #1?
The promoter would be to the left of basepair #1. Promoters are always 5′ to the start of transcription.

c) What are the first 10 nucleotides of the resulting mRNA? Indicate the 5' and 3' ends of the mRNA.
5′ AUAAUAUUGU 3′

d) What are the first 5 amino acids translated from the resulting mRNA? Indicate the amino (NH_3^+) and carboxy (COO^-) termini of the protein.
N- Met-leu-tyr-pro-ala-C

e) Do the underlined nucleotides TAA (indicated with the letter "**c**") encode a stop codon for this protein? Briefly explain your answer.
No, these codons are read (GAT) (AAT), so the TAA is out of frame.

Consider the situations in parts (f-h) independently.

f) A mutation occurs which results in the insertion of an extra G/C (top strand/bottom strand) base-pair immediately after base pair 11 (shown in bold). What effect will this insertion mutation have on transcription and translation?

The mRNA transcript will be longer by 1 nucleotide, but the resulting protein from this mRNA transcript will be the same, this is because the insertion is before the ATG on the mRNA transcript and translation does not start until the first ATG. Therefore this one base pair insertion will not effect the protein sequence because it is considered to be in the 5' untranslated region of the mRNA

g) A different mutation results in the substitution of the T/A base pair at position 30 (shown in bold and underlined) with a G/C base pair. How would this mutation affect the sequence of the protein that is produced?

This results in a premature stop codon, so the protein will be shorter. Insertion of this stop codon causes the protein to be truncated because translation of the mRNA will terminate at this stop codon.

h) A third mutation occurs which results in the substitution of the C/G base pair at position 42 (shown in bold italics) to a T/A base pair. How would this mutation affect the sequence of the protein that is produced?

This is a conservative substitution, both AAC and AAU code for asn, so the resulting protein will be the same.

7.012

10/12

Control of Transcription

Diff b/w mammals is which genes are turned on + off

Nacent = new

in bacteria transcription + translation can happen together at same time

but in eukaryots - segregation of processing

why are some genes turned on while others are turned off

Promoter
RNA polymerase bound
DNA unwinds
RNA Polymerase expands
bubble of unwound DNA

life time of mRNA is very short in bacteria
while one end still being written, other end is done

②

Try operon

Often a given mRNA can make 1
esp in mammels

Promoter - governs transciption control

Operator - controls ___ - importent seq

Some unlinked reglatory gene
- makes Try repressor
- binds to operator
- prevents active transciption

Co-transcribed as a single unit

(He is not making clear what is euk or prok)

If bacteria has a lot of try
then don't make more!
Shut down the bio synthetic pathway
tryp could bind w/ the repressor

③

"allosteric"

now repressor can bind to operator
prevents forward processing
A way bacteria can regulate

We can only make 8 amino acids
eat the rest

homostasis regulates body so fairly steady

lactose milk
lactase enzyme

lac operon -
repression can stick on operator + shut it down
can be removed as well (missed)

④

Functionally very important discovery

Understand how eukaryots operate - didn't in 60s

its much more complicated than bacteria!

lots of ribosoma RNA

- since very stable

mRNA has a limited life span

once exported into cytoplasm

cells wants to turn this on + off

RNA polymerase 2 operates in nucleoplasm

Cell wants constant, synthetic RNA

Regulate RNA polymerase 2

which is pre RNA

post splicing

hetrocromotin - when contensed

euchromatin - dispersed

when being trascribed

So we can look at the density

300-400 ribosomal genes

Crank out a lot of ribosome

When cells divide - each need their own

Each gets larger + larger

~ 5000 ribosomal

40-50 RNA polymerase

[may misheard #s]

black dots = ribosomal protiens

assembly line mass production

How Chromatin is really put together

lots of compaction 2m → 5 microns

So need to compact!

histones form these nucleosomes

⑥

Coil around
which coil arounds

which goes to chromatid
this is not that well understood

if prevent → get crazyness

---

polyribosomes only in cytoplasm
Bacteria life span could only be sev min

---

will gene be transcribed or not?

Unwound chromatid
(This prof is hard to follow)
nucleosome = primary packing device

①

DNA rapped around nucleosomes

Then for more complex structure

Cartoon rep

Optomer of 4 type of Histone molecules

2xH2A
2xH2B
2x H3
2x H4

Very ~~diff~~ similar in plants
evolved diff
but histones didn't change
was no way to improve!

Post translational structure matters

Gene may be several killobases long

Post translational structure - governs if transcribed

4 histone molecules on inside

⑧

Note the entire histone is within

N-terminal tails stick at

Sites for covalent modifications

Can see N-terminal tails can be modified in multiple ways

Change is not perminant

Methlyase may attach /detatch

Turning on + off govered by this n-terminal modifications

Focus mostly on histone H3

↳ most critical ones here

2 diff kinds of protiens interact

Writers - modify histone

Readers - read the changes

erasers - remove the marks by the writers

On + off

⑨

Where in a gene is a particular modification?

Can break up chromatin mechanically

But still set up w/ its nucleosome

Then make a anti body againt a modifiation

└ trimethalated on Hlys 4

How does DNA fit in extent of the gene

Add complexity

upstream ; enhansers

Say if genes transcibed

Could be very far upstream

determine if promoter actually transcibed

---

activators can bine to enhansers

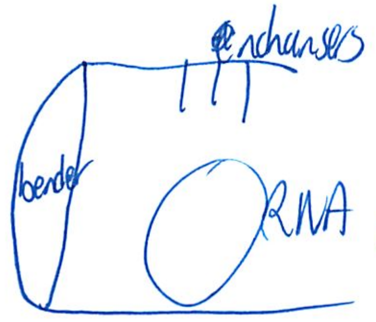
└ 2000 NFAT or AP1

Upstream lots of enhansers

So protine determined combinatorily by all these activators

affect if transcribtions actually happen

DNA bends around

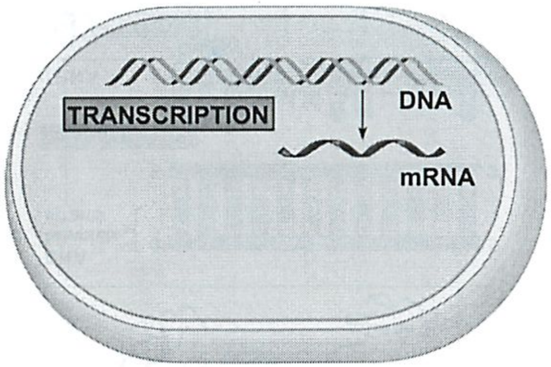

enchansers
bender
RNA polymerase

but covalent structure of histones must allow

What protiens look like in a cell

7.012 2012
Control of Transcription

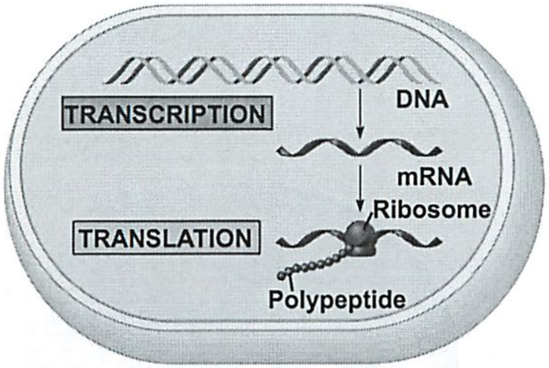
Fig. 17-3a-1


TRANSCRIPTION
DNA
mRNA

(a) Bacterial cell

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

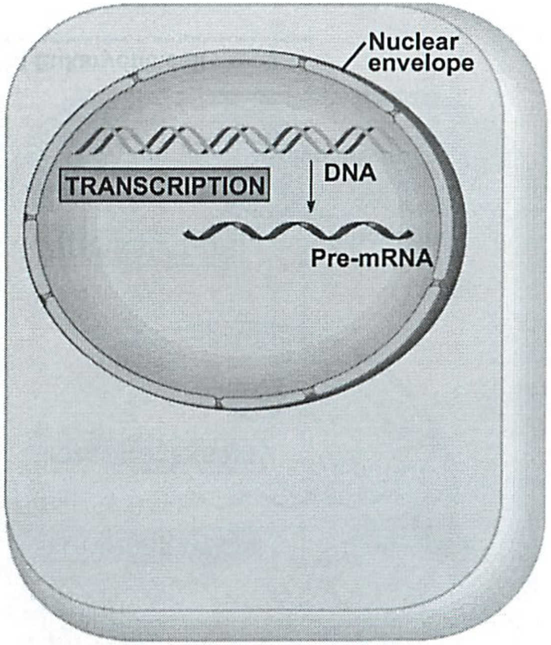
Fig. 17-3a-2


TRANSCRIPTION
DNA
mRNA
Ribosome
TRANSLATION
Polypeptide

(a) Bacterial cell

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Fig. 17-3b-1

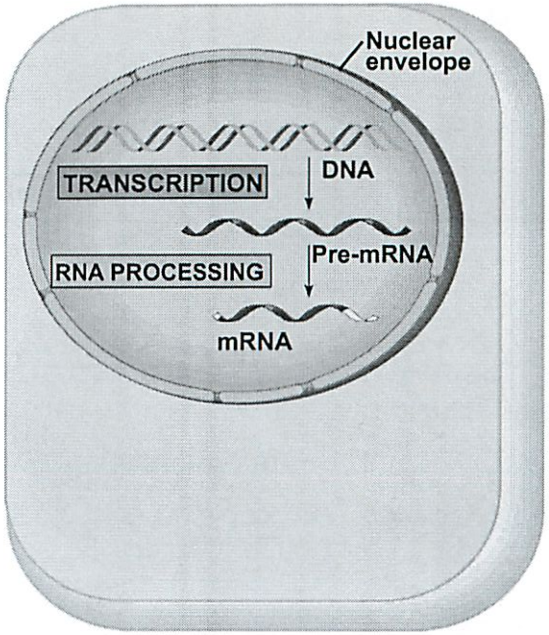

Nuclear envelope
TRANSCRIPTION
DNA
Pre-mRNA

(b) Eukaryotic cell

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

10/12

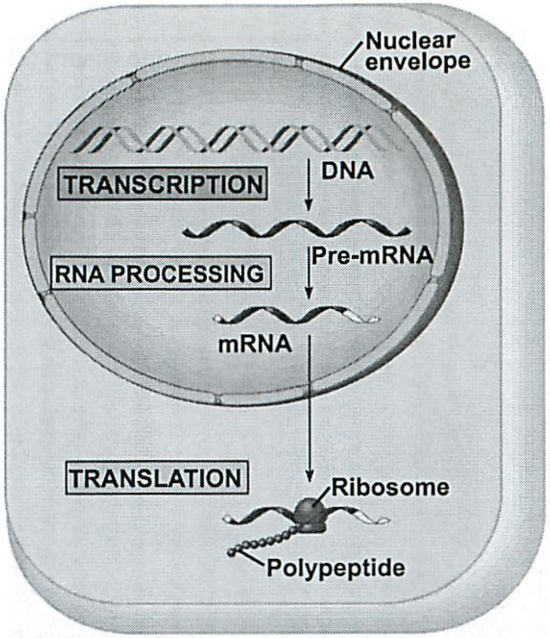
Fig. 17-3b-2


Nuclear envelope
TRANSCRIPTION
DNA
Pre-mRNA
RNA PROCESSING
mRNA

(b) Eukaryotic cell

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

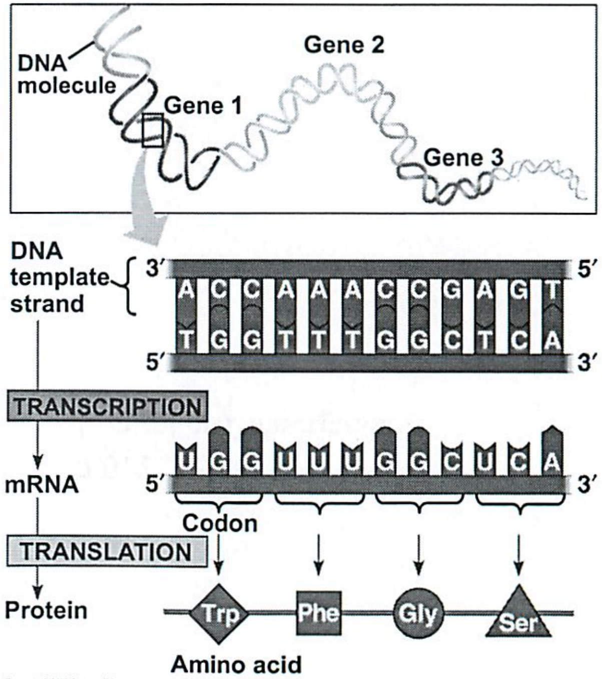
Fig. 17-3b-3


Nuclear envelope
TRANSCRIPTION
DNA
Pre-mRNA
RNA PROCESSING
mRNA
TRANSLATION
Ribosome
Polypeptide

(b) Eukaryotic cell

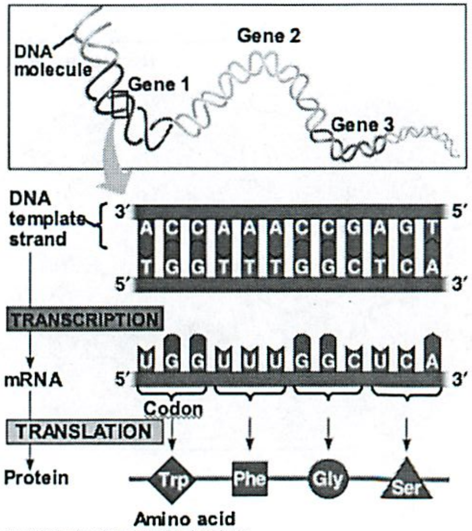
Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Fig. 17-4


DNA molecule
Gene 1
Gene 2
Gene 3
DNA template strand
3′
A C C A A A C C G A G T
T G G T T T G G C T C A
5′
5′
3′
TRANSCRIPTION
mRNA
5′
U G G U U U G G C U C A
3′
Codon
TRANSLATION
Protein
Trp
Phe
Gly
Ser
Amino acid

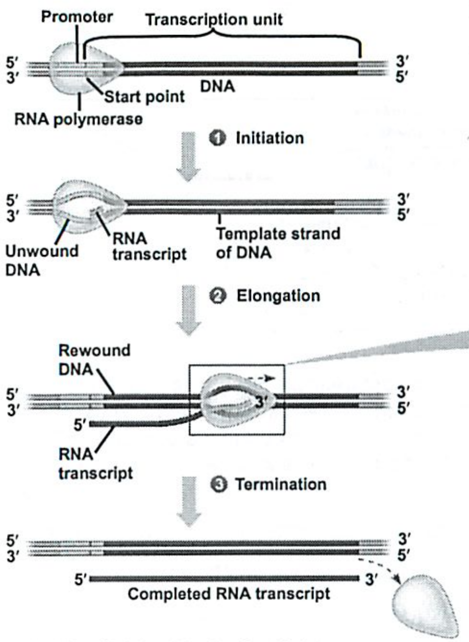
Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Independent of the detailed mechanisms of transcription, it is important to realize that the spectrum of proteins in a cell is determined by whether or not a gene is transcribed. e.g., here Gene 1 is transcribed while Genes 2 & 3 are not.


DNA molecule
Gene 1
Gene 2
Gene 3
DNA template strand
3′
A C C A A A C C G A G T
T G G T T T G G C T C A
5′
5′
3′
TRANSCRIPTION
mRNA
5′
U G G U U U G G C U C A
3′
Codon
TRANSLATION
Protein
Trp
Phe
Gly
Ser
Amino acid

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Fig. 17-7


Promoter
Transcription unit
5′
3′
3′
5′
Start point
DNA
RNA polymerase
❶ Initiation
Unwound DNA
RNA transcript
Template strand of DNA
❷ Elongation
Rewound DNA
5′
RNA transcript
❸ Termination
Completed RNA transcript

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

In more detail (prokaryotic or eukaryotic)

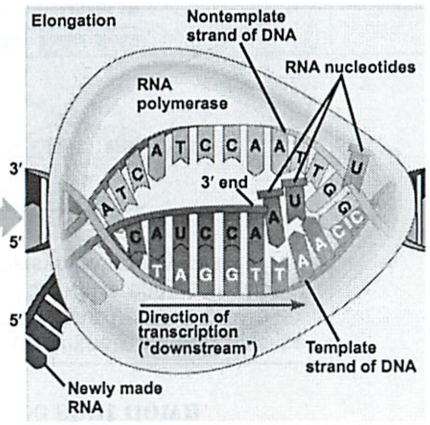
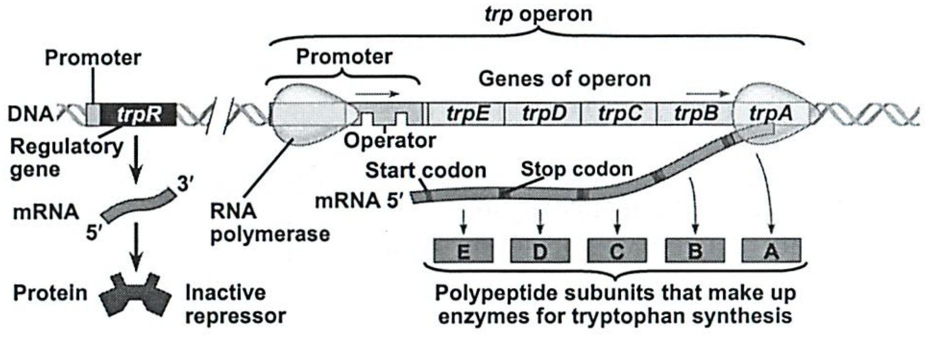

Elongation
Nontemplate strand of DNA
RNA nucleotides
RNA polymerase
3′
5′
5′
3′ end
A T C C A A
T G G
U
C A U C C A
A U
T A G G T T A A C C
Direction of transcription ("downstream")
Template strand of DNA
Newly made RNA

Fig. 18-3a

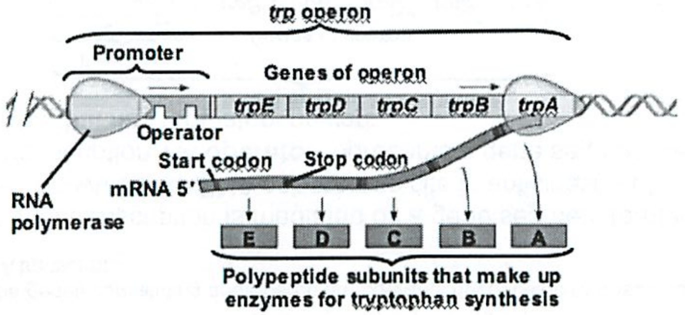
The tryptophan operon


trp operon
Promoter
Promoter
Genes of operon
DNA
trpR
trpE
trpD
trpC
trpB
trpA
Regulatory gene
Operator
Start codon
Stop codon
mRNA
3′
5′
RNA polymerase
mRNA 5′
E
D
C
B
A
Protein
Inactive repressor
Polypeptide subunits that make up enzymes for tryptophan synthesis

(a) Tryptophan absent, repressor inactive, operon on

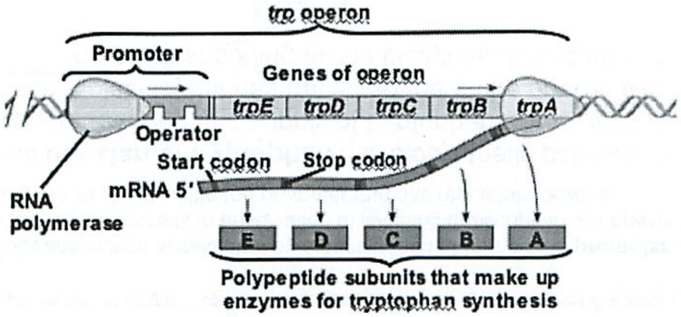
Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Let' s see how a bacterium manages control of transcription, in this case the enzymes that enable it to make its own tryptophan (i.e., an amino acid). In the absence of externally provided tryptophan, the bacterium makes a series of enzymes that in aggregate enable the multiple steps of tryptophan biosynthesis to occur. Not only are these enzymes linked functionally (as part of a common biosynthetic pathway), but cleverly the genes encoding them are linked in a larger genetic unit termed an "**operon**".


trp operon
Promoter
Genes of operon
trpE
trpD
trpC
trpB
trpA
Operator
Start codon
Stop codon
mRNA 5′
RNA polymerase
E
D
C
B
A
Polypeptide subunits that make up enzymes for tryptophan synthesis

Let' s see how a bacterium manages control of transcription, in this case the enzymes that enable it to make its own tryptophan (i.e., an amino acid). In the absence of externally provided tryptophan, the bacterium makes a series of enzymes that in aggregate enable the multiple steps of tryptophan biosynthesis to occur. Not only are these enzymes linked functionally (as part of a common biosynthetic pathway), but cleverly the genes encoding them are linked in a larger genetic unit termed an "**operon**".

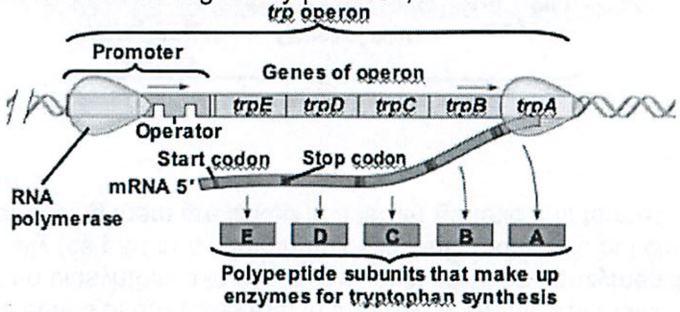
Moreover, the genes constituting and operon are co-transcribed, being represented in a single mRNA transcript.


trp operon
Promoter
Genes of operon
trpE
trpD
trpC
trpB
trpA
Operator
Start codon
Stop codon
mRNA 5′
RNA polymerase
E
D
C
B
A
Polypeptide subunits that make up enzymes for tryptophan synthesis

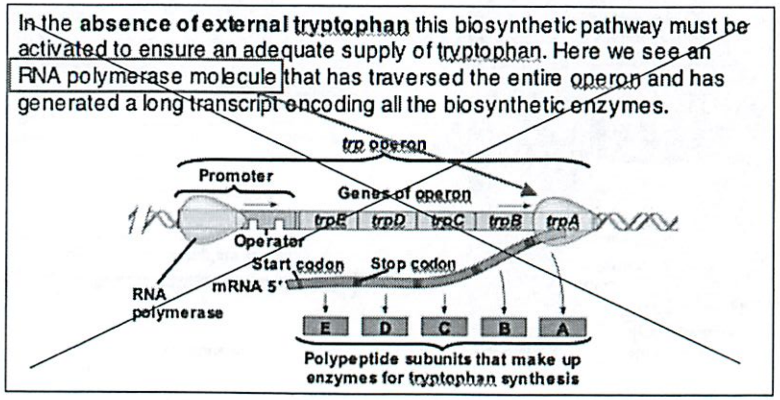
Let's see how a bacterium manages control of transcription, in this case the enzymes that enable it to make its own tryptophan (i.e., an amino acid). In the absence of externally provided tryptophan, the bacterium makes a series of enzymes that in aggregate enable the multiple steps of tryptophan biosynthesis to occur. Not only are these enzymes linked functionally (as part of a common biosynthetic pathway), but cleverly the genes encoding them are linked in a larger genetic unit termed an **"operon"**.

Moreover, the genes constituting and operon are co-transcribed, being represented in a single mRNA transcript.

The control of transcription is mediated by a gene segment termed the **promoter** , on which the RNA polymerase sits in anticipation of initiating transcription. An **operator** represents a gene segment within the promoter that has regulatory powers. ,


trp operon
Promoter
Genes of operon
trpE
trpD
trpC
trpB
trpA
Operator
Start codon
Stop codon
mRNA 5'
RNA polymerase
E
D
C
B
A
Polypeptide subunits that make up enzymes for tryptophan synthesis

What happens when there is abundant tryptophan in the environment? The bacterium no longer needs to waste energy and metabolites to makes its own tryptophan. Hence, the expression (i.e., the transcription) of this operon should be **shut down**.


~~In the **absence of external tryptophan** this biosynthetic pathway must be activated to ensure an adequate supply of tryptophan. Here we see an RNA polymerase molecule that has traversed the entire operon and has generated a long transcript encoding all the biosynthetic enzymes.~~
trp operon
Promoter
Genes of operon
trpE
trpD
trpC
trpB
trpA
Operator
Start codon
Stop codon
mRNA 5'
RNA polymerase
E
D
C
B
A
Polypeptide subunits that make up enzymes for tryptophan synthesis

Let's see how a bacterium manages control of transcription, in this case the enzymes that enable it to make its own tryptophan (i.e., an amino acid). In the absence of externally provided tryptophan, the bacterium makes a series of enzymes that in aggregate enable the multiple steps of tryptophan biosynthesis to occur. Not only are these enzymes linked functionally (as part of a common biosynthetic pathway), but cleverly the genes encoding them are linked in a larger genetic unit termed an **"operon"**.

Moreover, the genes constituting and operon are co-transcribed, being represented in a single mRNA transcript.

The control of transcription is mediated by a gene segment termed the **promoter** , on which the RNA polymerase sits in anticipation of initiating transcription. An **operator** represents a gene segment within the promoter that has regulatory powers. ,

In the **absence of external tryptophan** this biosynthetic pathway must be activated to ensure an adequate supply of tryptophan. Here we see an RNA polymerase molecule that has traversed the entire operon and has generated a long transcript encoding all the biosynthetic enzymes.

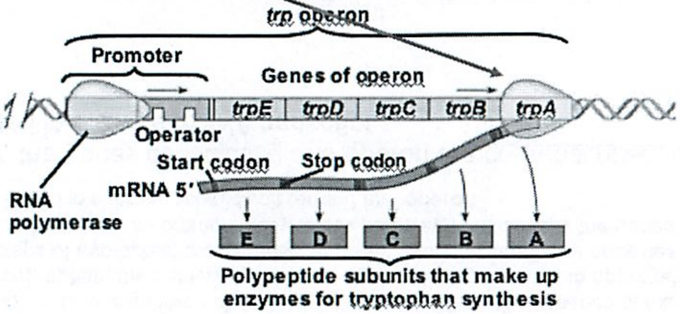
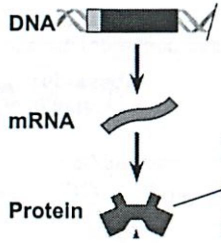

trp operon
Promoter
Genes of operon
trpE
trpD
trpC
trpB
trpA
Operator
Start codon
Stop codon
mRNA 5'
RNA polymerase
E
D
C
B
A
Polypeptide subunits that make up enzymes for tryptophan synthesis

Fig. 18-3b-1

The bacterium accomplishes this shutdown by making a protein Encoded by an unlinked gene. The protein is termed a **repressor.**

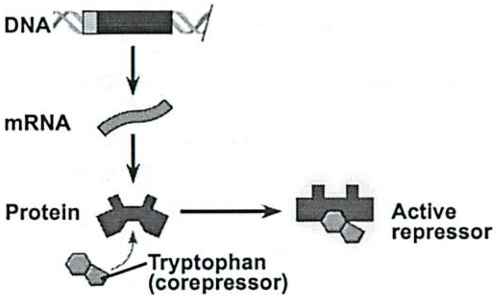

DNA
mRNA
Protein

In the absence of tryptophan, this **repressor** is made but is functionally inactive.

Fig. 18-3b-1

The bacterium accomplishes this shutdown by making a protein Encoded by an unlinked gene. The protein is termed a **repressor.**

However, in the presence of tryptophan, a molecule of this amino will bind directly to the repressor, changing the configuration of the latter.


DNA
mRNA
Protein
Active repressor
Tryptophan (corepressor)

(b) Tryptophan present, repressor active, operon off

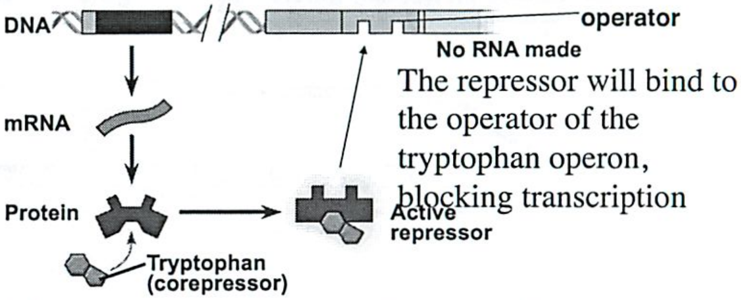
Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings

---

Fig. 18-3b-1

The bacterium accomplishes this shutdown by making a protein Encoded by an unlinked gene. The protein is termed a **repressor.**

However, in the presence of tryptophan, a molecule of this amino will bind directly to the repressor, changing the configuration of the latter.


DNA
operator
No RNA made
mRNA
Protein
Active repressor
Tryptophan (corepressor)

The repressor will bind to the operator of the tryptophan operon, blocking transcription

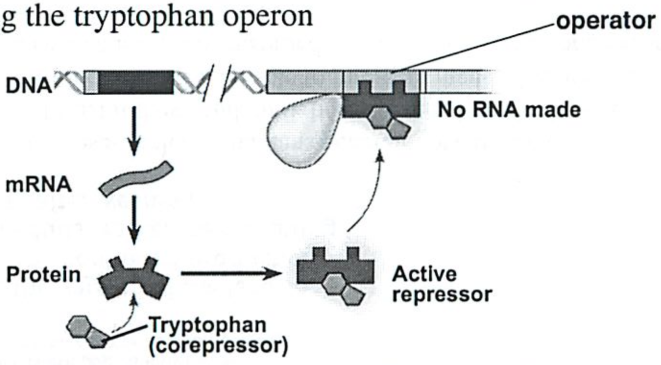
(b) Tryptophan present, repressor active, operon off

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings

---

Fig. 18-3b-2

By binding to the operator, the repressor (activated by having bound a molecule of tryptophan)prevents RNA polymerase from transcribing the tryptophan operon

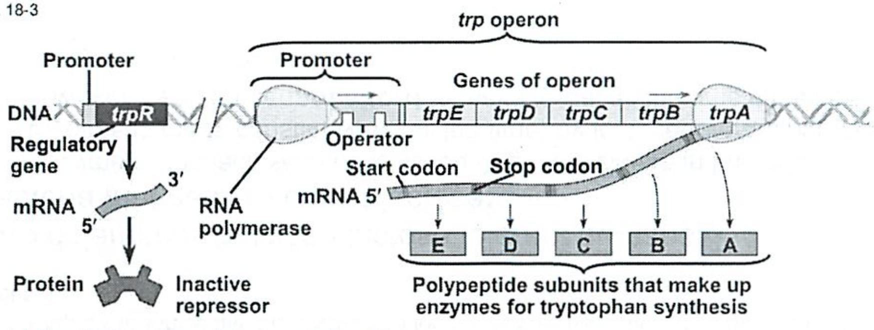

operator
DNA
No RNA made
mRNA
Protein
Active repressor
Tryptophan (corepressor)

(b) Tryptophan present, repressor active, operon off

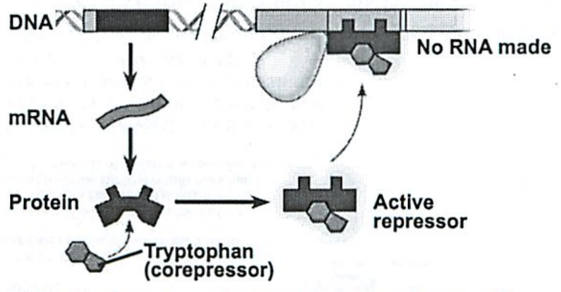
Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings

(note that the gene encoding the repressor is unlinked to the tryp operon)

Fig. 18-3


trp operon
Promoter
Promoter
Genes of operon
DNA
trpR
trpE
trpD
trpC
trpB
trpA
Regulatory gene
Operator
Start codon
Stop codon
3′
mRNA
5′
RNA polymerase
mRNA 5′
E
D
C
B
A
Protein
Inactive repressor
Polypeptide subunits that make up enzymes for tryptophan synthesis

(a) Tryptophan absent, repressor inactive, operon on

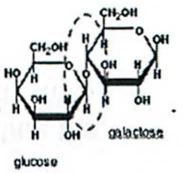

DNA
No RNA made
mRNA
Protein
Active repressor
Tryptophan (corepressor)

(b) Tryptophan present, repressor active, operon off

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings

To summarize, by going from a state where tryptophan is not available to one where it available in abundance inactivates expression of the tryp operon.

Researchers have found the first direct evidence that early Europeans were unable to digest milk: the gene that controls our ability to digest milk (= lactase gene) was missing from Neolithic skeletons dating to between 5840 and 5000 BC.



lactose

However, through exposure to milk, lactose tolerance evolved extremely rapidly, in evolutionary terms.

Today, it is present in over ninety per cent of the population of northern Europe and is also found in some African and Middle Eastern populations but is missing from the majority of the adult population globally.

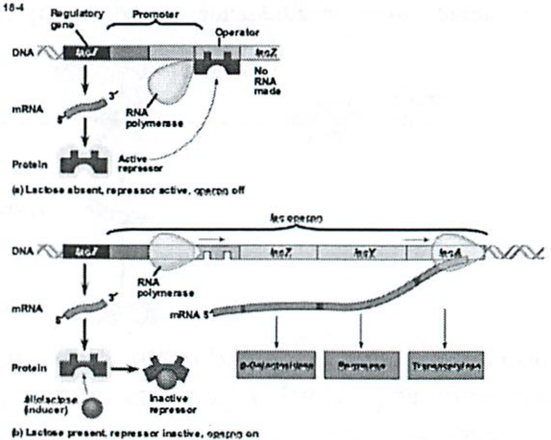
More specifically, the version of the lactase gene that permits the continued production of this enzyme in adults (rather than shutting it down early in life after weaning).

Remember this? (Lactose-intolerant humans generate lots of gas, etc
Etc in their lower GI tract if they drink milk because they lack the lactase in their upper GI tract that would otherwise be able to intercept ingested lactose and break it down to glucose (which is rapidly absorbed into the circulation) long before the bacteria in the lower GI tract can get to it.

A very different situation operates in the "lac operon".

1. In this case, the bacterial cell is reacting to the presence of a nutrient of exogenous origin -- lactose, to whose presence it must respond.
2. Normally the operon is shut off by a repressor that must be removed when the inducer -- lactose, becomes available

Fig. 18-4



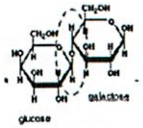
(a) Lactose absent, repressor active, operon off

(b) Lactose present, repressor inactive, operon on

Researchers have found the first direct evidence that early Europeans were unable to digest milk: the gene that controls our ability to digest milk (= lactase gene) was missing from Neolithic skeletons dating to between 5840 and 5000 BC.

However, through exposure to milk, lactose tolerance evolved extremely rapidly, in evolutionary terms.

Today, it is present in over ninety per cent of the population of northern Europe and is also found in some African and Middle Eastern populations but is missing from the majority of the adult population globally.



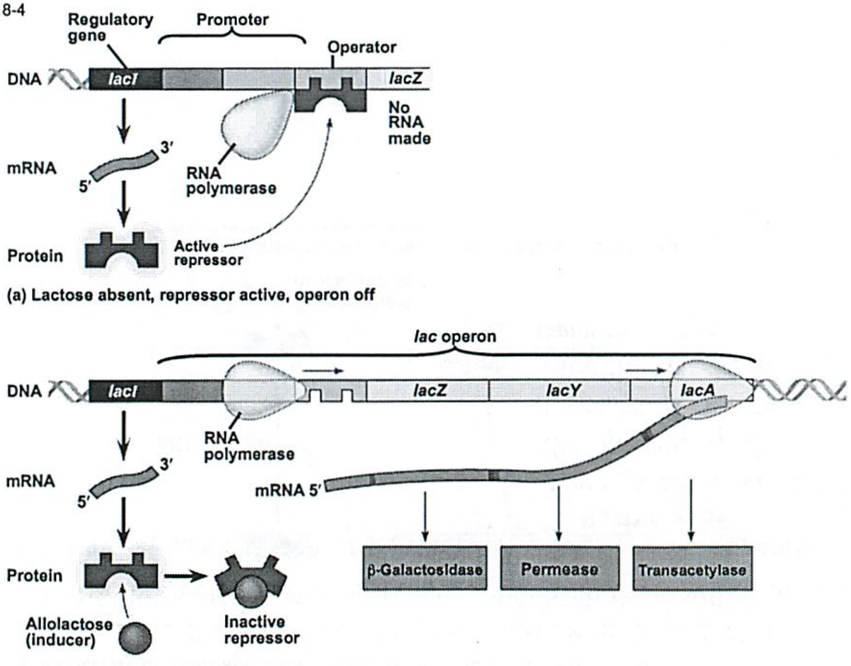
More specifically, the version of the lactase gene that permits the continued production of this enzyme in adults (rather than shutting it down early in life after weaning).

Remember this? (Lactose-intolerant humans generate lots of gas, etc
Etc in their lower GI tract if they drink milk because they lack the lactase in their upper GI tract that would otherwise be able to intercept ingested lactose and break it down to glucose (which is rapidly absorbed into the circulation) long before the bacteria in the lower GI tract can get to it.

Two differences between lactose tolerant humans & bacteria that need to digest lactose:

1. Enzyme called lactase in us is called β -galactosidase in bacteria
2. Enzyme is made constitutively (all the time) by lactose-tolerant humans; however, bacteria **cannot afford** to make β -gal all the time if they only rarely encounter lactose.

Fig. 18-4



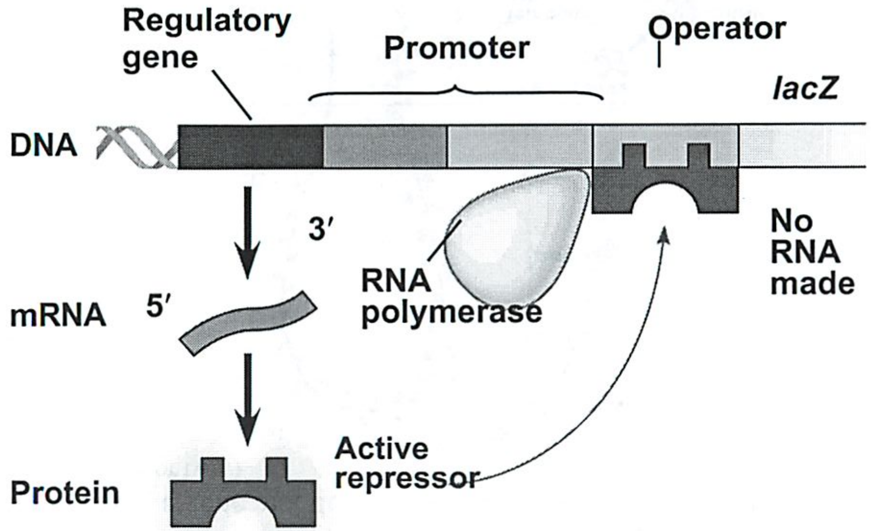
(a) Lactose absent, repressor active, operon off

(b) Lactose present, repressor inactive, operon on

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Fig. 18-4a

Normally, the repressor, made by the *lacI* gene, blocks RNA polymerase. (lactose is not available)

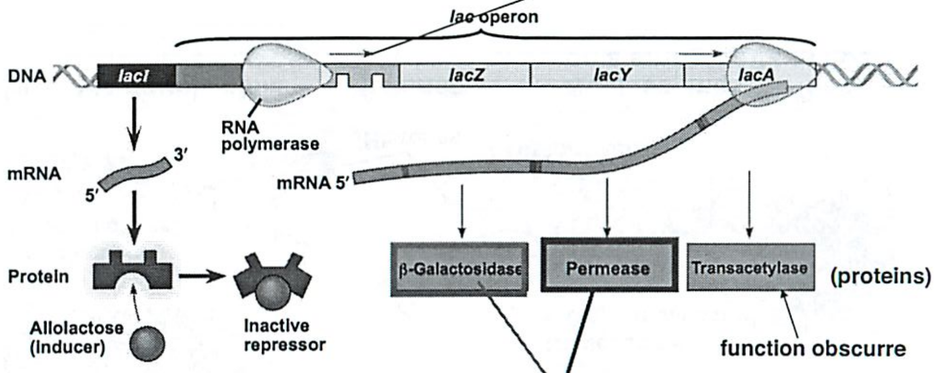

Regulatory gene
Promoter
Operator
lacZ
DNA
3′
mRNA
5′
RNA polymerase
No RNA made
Protein
Active repressor

(a) Lactose absent, repressor active, operon off

Copyright © 2006 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Fig. 18-4b

However, is lactose is present, it binds to the repressor and causes the repressor to lose its grip on the operator DNA


lac operon
DNA
lacI
lacZ
lacY
lacA
RNA polymerase
mRNA
3′
5′
mRNA 5′
Protein
Allolactose (inducer)
Inactive repressor
β-Galactosidase
Permease
Transacetylase
(proteins)
function obscurre

(b) Lactose present, repressor inactive, operon on

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Now, the progress of the RNA polymerase is unimpeded, RNA pol advances and makes a "polycistronic" mRNA I.e., an mRNA that encodes multiple distinct proteins, each with its own reading frame.

The resulting proteins help to **import** and **degrade** lactose.

What about **eukaryotic** genes and their transcription? (us)

Pol I works in the nucleolus cranking out enormous amounts of ribosomal RNA

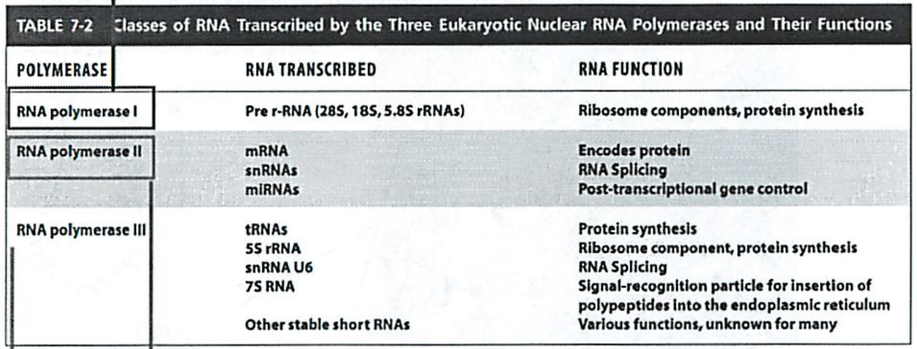


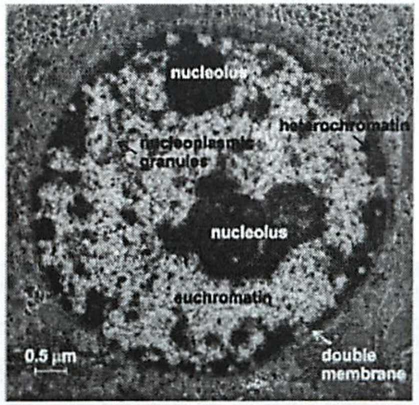
TABLE 7-2 Classes of RNA Transcribed by the Three Eukaryotic Nuclear RNA Polymerases and Their Functions

POLYMERASE	RNA TRANSCRIBED	RNA FUNCTION
RNA polymerase I	Pre r-RNA (28S, 18S, 5.8S rRNAs)	Ribosome components, protein synthesis
RNA polymerase II	mRNA	Encodes protein
	snRNAs	RNA Splicing
	miRNAs	Post-transcriptional gene control
RNA polymerase III	tRNAs	Protein synthesis
	5S rRNA	Ribosome component, protein synthesis
	snRNA U6	RNA Splicing
	7S RNA	Signal-recognition particle for insertion of polypeptides into the endoplasmic reticulum
	Other stable short RNAs	Various functions, unknown for many

Table 7-2
Molecular Cell Biology, Sixth Edition
© 2008 W.H. Freeman and Company

Pol II works in the nucleoplasm making the pre-mRNA of protein-coding genes. (nucleoplasm = nuclear space outside of nucleolus

Pol III works in the nucleoplasm making small RNAs


nucleolus
heterochromatin
nucleoplasmic granules
nucleolus
euchromatin
0.5 µm
double membrane

You can actually see here the RNA molecules being elongated.

Pol I works in the nucleolus cranking out enormous amounts of ribosomal RNA

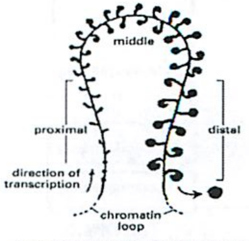
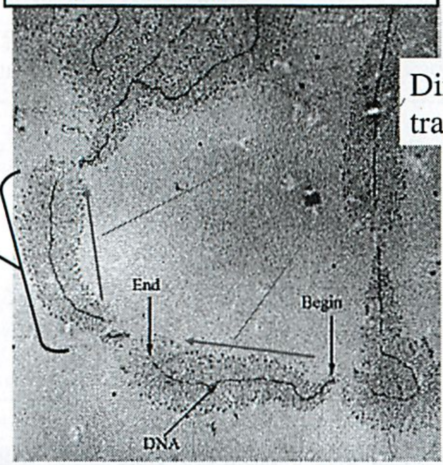

middle
proximal
distal
direction of transcription
chromatin loop

Figure 4–43 part 2 of 2. Molecular Biology of the Cell, 4th Edition.


A single rRNA gene
Direction of transcription
End
Begin
DNA

Here's a famous electron micrograph of ribosomal RNA being transcribed in a (eukaryotic) nucleolus. There are dozens of rRNA genes located in tandem along the chromosome. (The RNA Polymerase molecules responsible for transcription are not visible here.) Each rRNA gene has several dozen RNA polymerase molecules that are moving down the gene, one after another, elongating their rRNA products.

What about the protein-coding genes being transcribed by RNA Pol II in the nucleoplasm? They probably look the same, except that at any point in time there may only be a small number of pol II molecules moving along a gene, not dozens. (Some pol II genes may only be transcribed once every hour or two.)

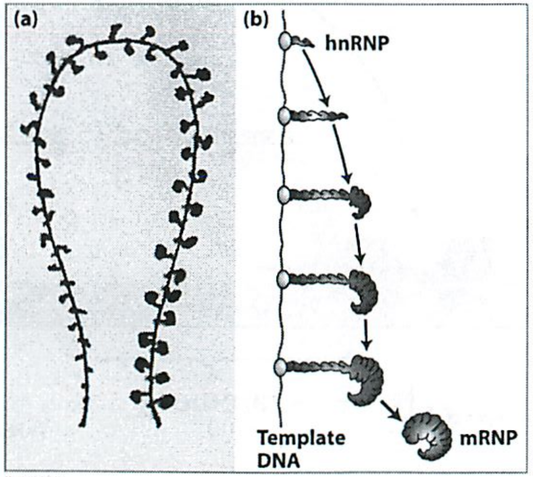

(a)
(b)
hnRNP
Template DNA
mRNP

Figure 8-23ab
Molecular Cell Biology, Sixth Edition
© 2008 W.H. Freeman and Company

Pol I works in the nucleolus cranking out enormous amounts of ribosomal RNA

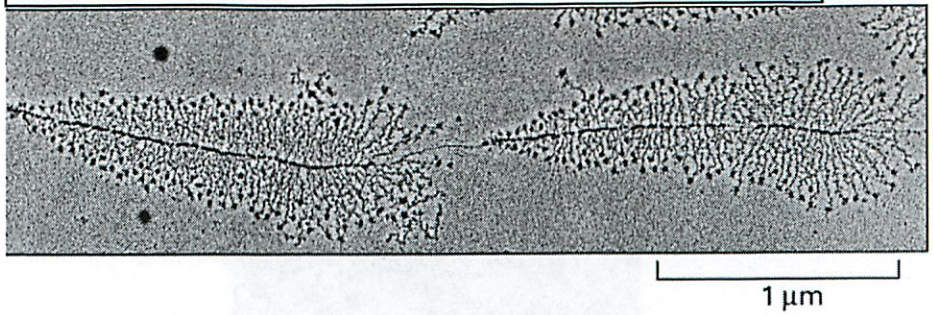
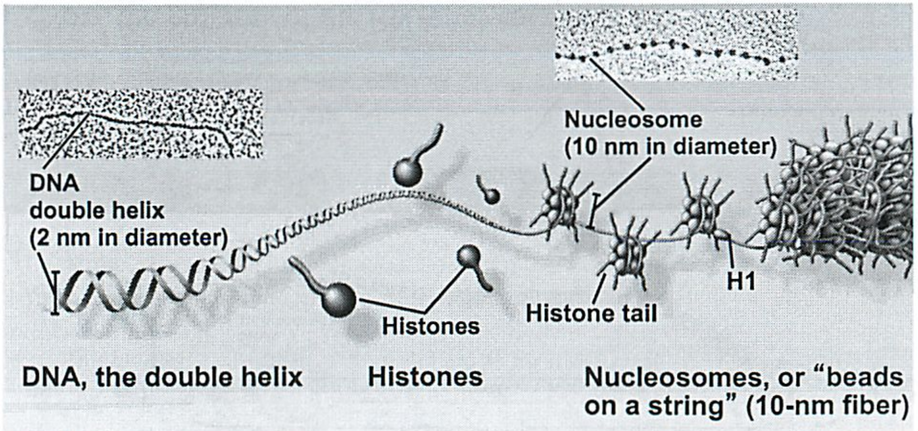


Figure 6–9. Molecular Biology of the Cell, 4th Edition.

The rRNA genes in the nucleolus are an **exception** however because (i) There are several hundred rRNA genes largely arranged in tandem arrays rather than the single-copy genes that encode most proteins (ii) they are transcribe by RNA pol I rather than the pol II for most single-copy genes; (iii) no splicing and polyadenylation; (iv) the nascent RNA molecules are coated with ribosomal proteins in order to assemble ribosome subunits; (v) a single rRNA gene may have several dozen RNA pol I molecules advancing, one after another, down the gene, leading to these "feathers".

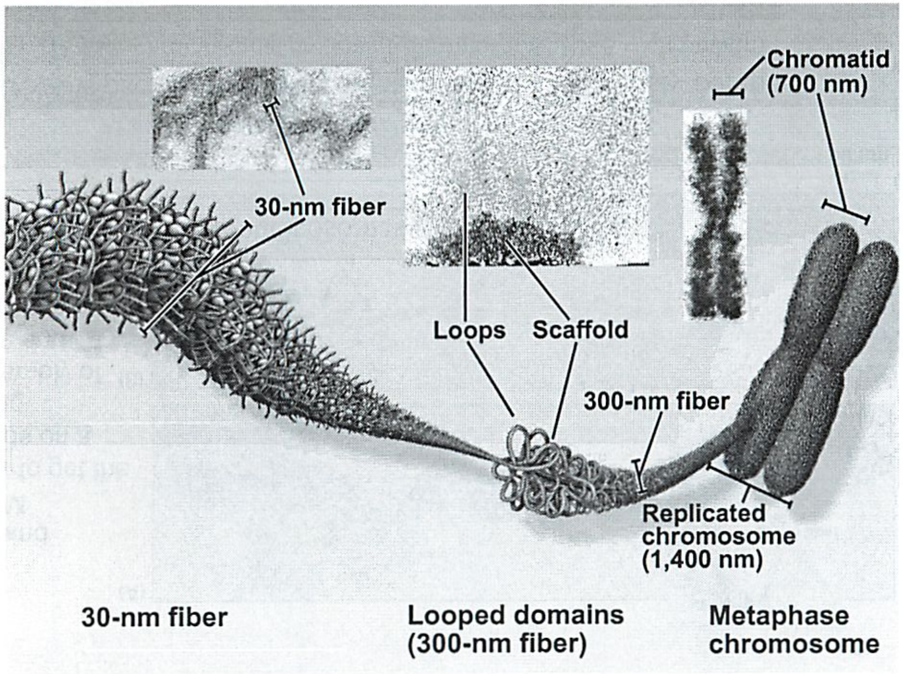
Fig. 16-21a

What about **packaging** of eukaryotic genes? (us)

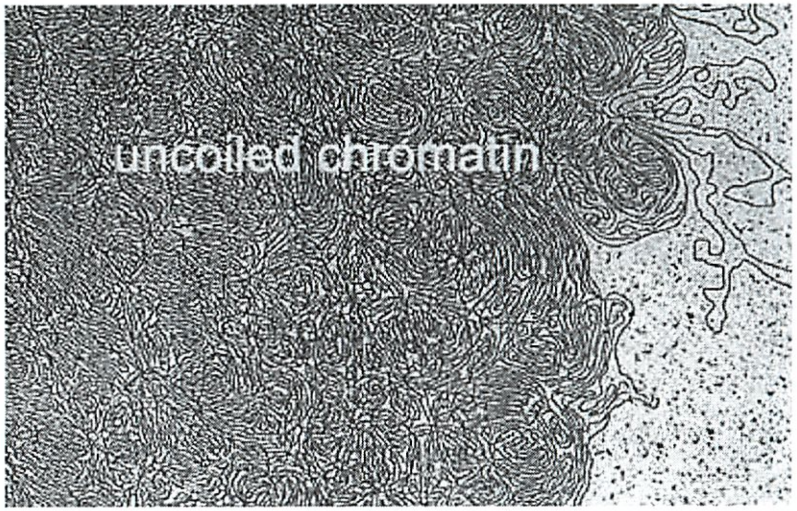

DNA double helix (2 nm in diameter)
Nucleosome (10 nm in diameter)
Histones
Histone tail
H1
DNA, the double helix
Histones
Nucleosomes, or "beads on a string" (10-nm fiber)

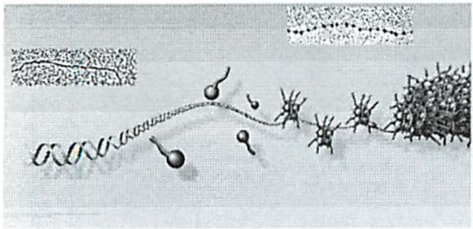
Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

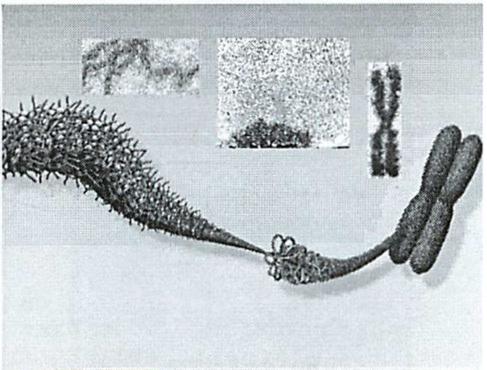
Fig. 16-21b


Chromatid
(700 nm)
30-nm fiber
Loops
Scaffold
300-nm fiber
Replicated
chromosome
(1,400 nm)
30-nm fiber
Looped domains
(300-nm fiber)
Metaphase
chromosome

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

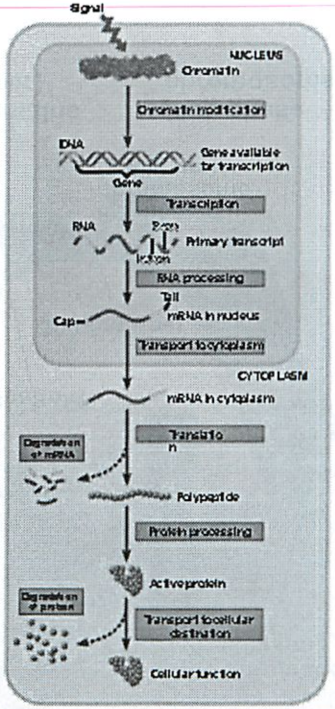

uncoiled chromatin



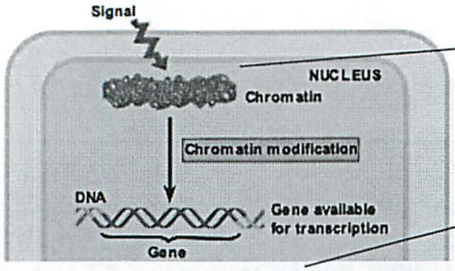


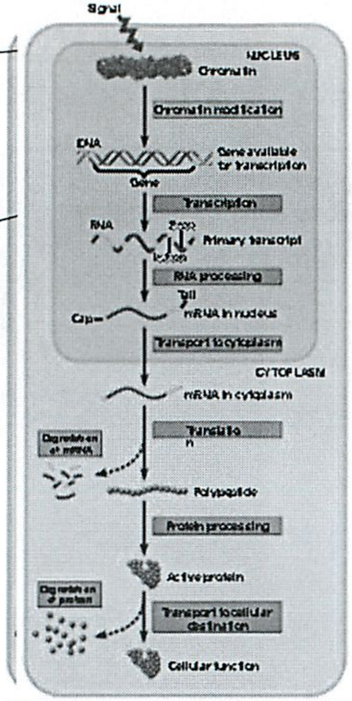
The situation in
eukaryotic cells
Is more complicated because:

1. nuclear/cytoplasmic segregation
2. Splicing
3. Post-translational protein modification
4. Post-translational diversion to specific target sites within the cell


Signal
NUCLEUS
Chromatin
Chromatin modification
DNA
Gene available for transcription
Gene
Transcription
RNA
Exon
Primary transcript
Intron
RNA processing
Tail
Cap
mRNA in nucleus
Transport to cytoplasm
CYTOPLASM
mRNA in cytoplasm
Translation
Degradation of mRNA
Polypeptide
Protein processing
Active protein
Degradation of protein
Transport to cellular destination
Cellular function

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings


Signal
NUCLEUS
Chromatin
Chromatin modification
DNA
Gene available for transcription
Gene

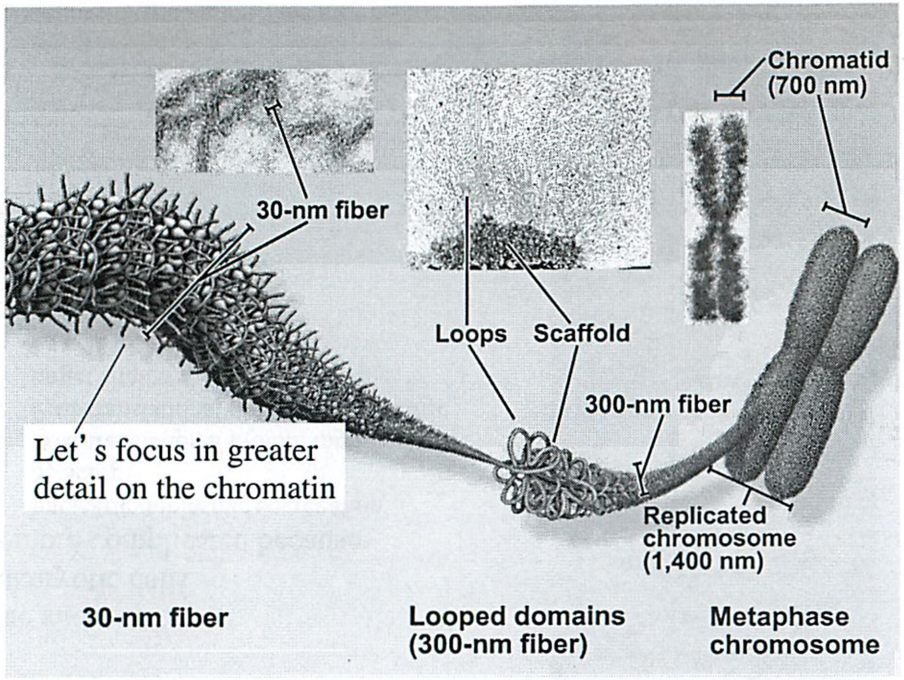

Signal
NUCLEUS
Chromatin
Chromatin modification
DNA
Gene available for transcription
Gene
Transcription
RNA
Exon
Primary transcript
Intron
RNA processing
Tail
Cap
mRNA in nucleus
Transport to cytoplasm
CYTOPLASM
mRNA in cytoplasm
Translation
Degradation of mRNA
Polypeptide
Protein processing
Active protein
Degradation of protein
Transport to cellular destination
Cellular function

Co Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings

Note that the **first step** in this complex cascade is the modification of chromatin to make it susceptible/available to transcription

This image focuses on the products of RNA Pol II, which makes pre-mRNAs in the nucleoplasm

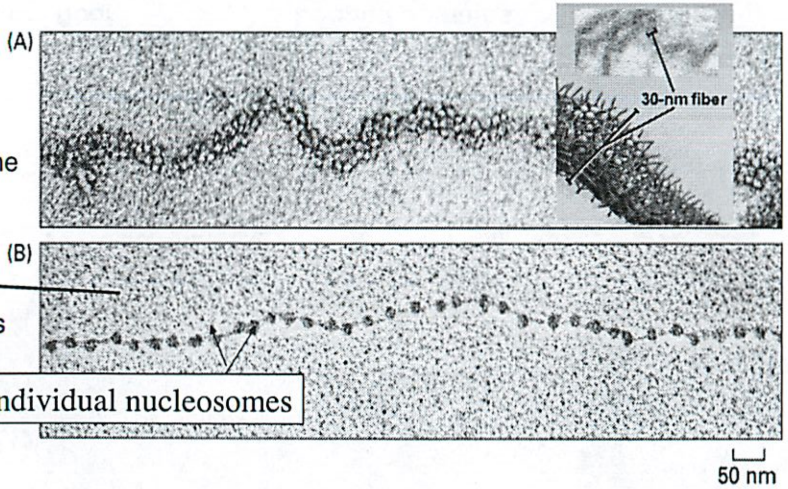
Fig. 16-21b


Chromatid (700 nm)
30-nm fiber
Loops
Scaffold
300-nm fiber
Replicated chromosome (1,400 nm)
30-nm fiber
Looped domains (300-nm fiber)
Metaphase chromosome

Let's focus in greater detail on the chromatin

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Unwind 30nM fiber to get the beads on a string assembly of DNA and nucleosomes


(A)
30-nm fiber
(B)
50 nm

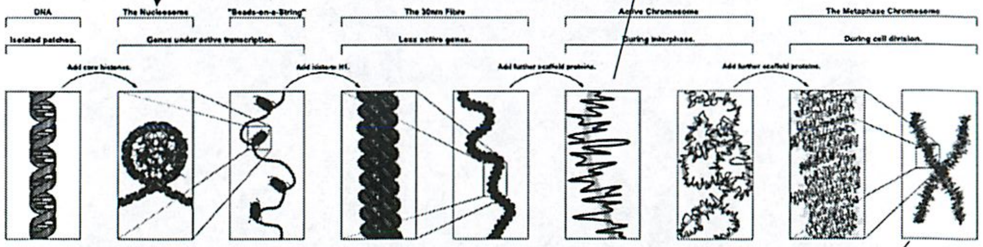
Individual nucleosomes

Figure 4-23. Molecular Biology of the Cell, 4th Edition.

DNA is packaged into chromatin, which enormously compacts it. (DNA + chromosomal proteins + chromosomal RNA = chromatin)

nucleosomes

Stretches of local decompaction associated with transcription

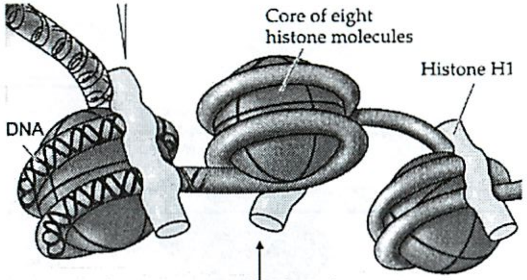

DNA
The Nucleosome
"Beads-on-a-String"
The 30nm Fibre
Active Chromosome
The Metaphase Chromosome
Isolated patches.
Genes under active transcription.
Less active genes.
During interphase.
During cell division.
Add core histones.
Add histone H1.
Add further scaffold proteins.
Add further scaffold proteins.

Increasing compaction ⟶

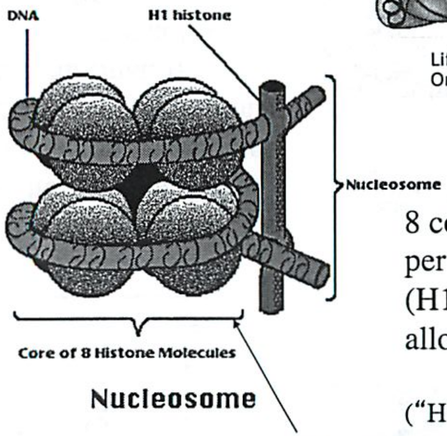
Ultimate compaction during mitosis

2 x H2A
2 x H2B
2 x H3
2 x H4

} core nucleosome


Core of eight histone molecules
Histone H1
DNA

Life, the Science of Biology, by Purves, Orians, & Heller, 5th ed., 1997


DNA
H1 histone
Nucleosome
Core of 8 Histone Molecules
Nucleosome

8 core histone molecules per nucleosome plus a 9th (H1) holding things together allowing compactions

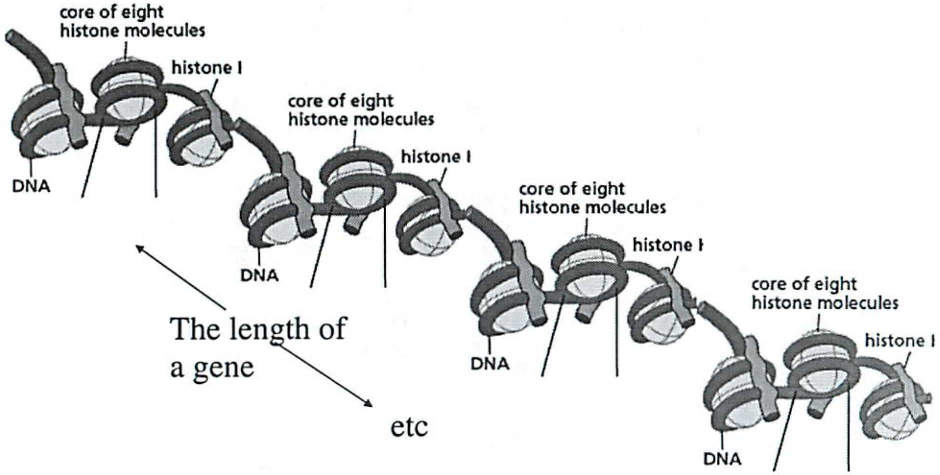
("H" = histone)

http://www.bio.davidson.edu/courses/Molbio/MolStudents/spring2000/lamar/nucleosome.gif

---

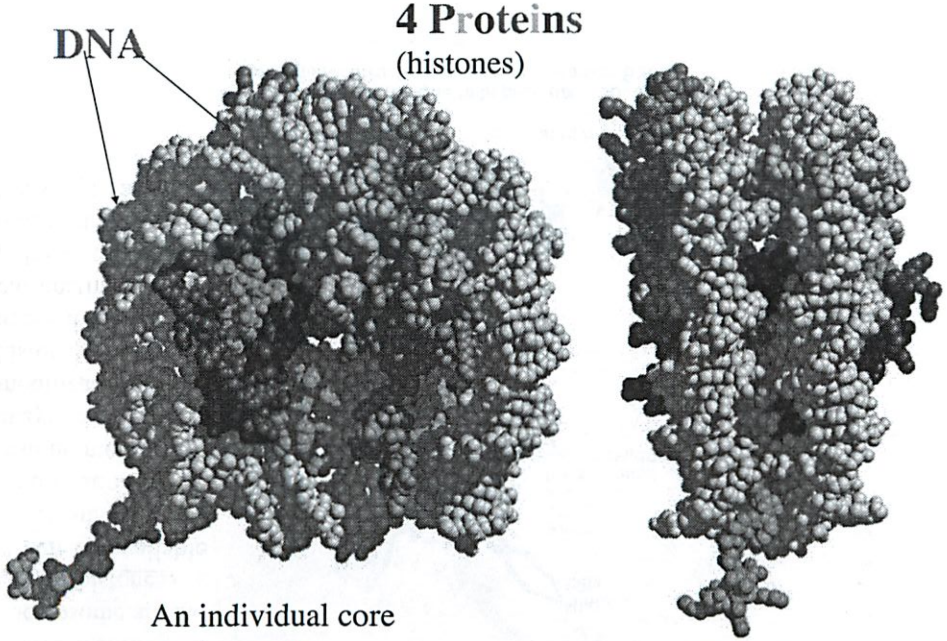
2 x H2A
2 x H2B
2 x H3
2 x H4

} core nucleosome


core of eight histone molecules
histone I
DNA
core of eight histone molecules
histone I
DNA
core of eight histone molecules
histone I
DNA
core of eight histone molecules
histone I
DNA

The length of a gene

etc

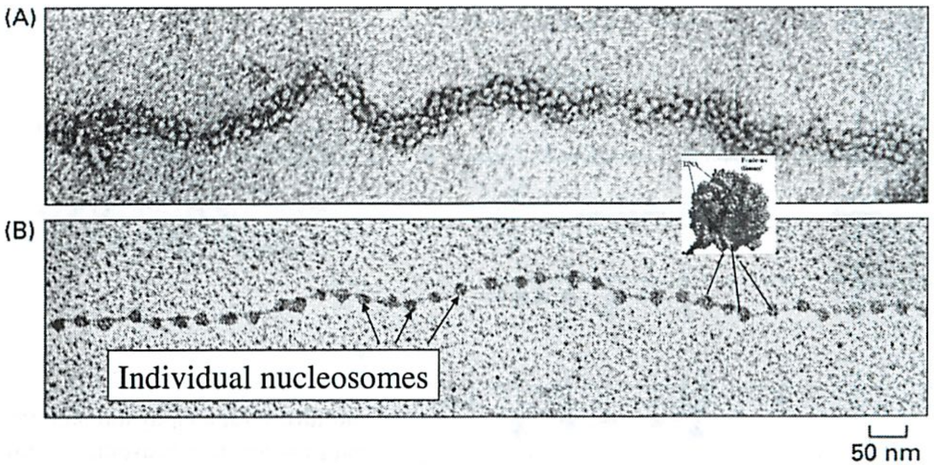


4 Proteins
(histones)

An individual core nucleosome

Side view

Figure 4–25. Molecular Biology of the Cell, 4th Edition.



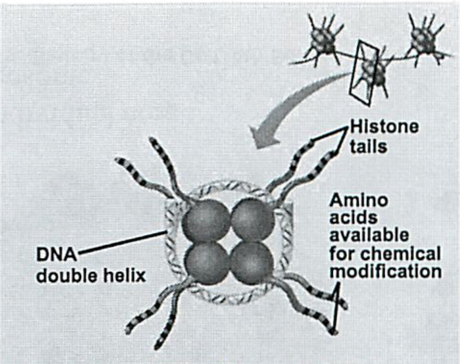
Individual nucleosomes

Figure 4–23. Molecular Biology of the Cell, 4th Edition.

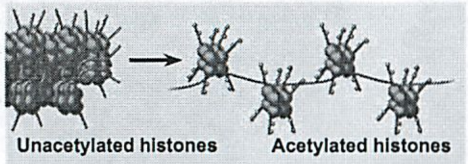
DNA is packaged into chromatin, which enormously compacts it. (DNA + chromosomal proteins + chromosomal RNA = chromatin)

Fig. 18-7

In fact, hanging out from the core nucleosome are the tails of histones, which are available for chemical modification by histone-modifying enzymes. The modification of histones affects the accessibility of the chromatin to RNA polymerases and thus regulates transcription.


Histone tails
Amino acids available for chemical modification
DNA double helix

(a) Histone tails protrude outward from a nucleosome


Unacetylated histones
Acetylated histones

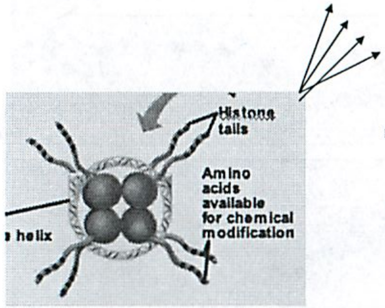
(b) Acetylation of histone tails promotes loose chromatin structure that permits transcription

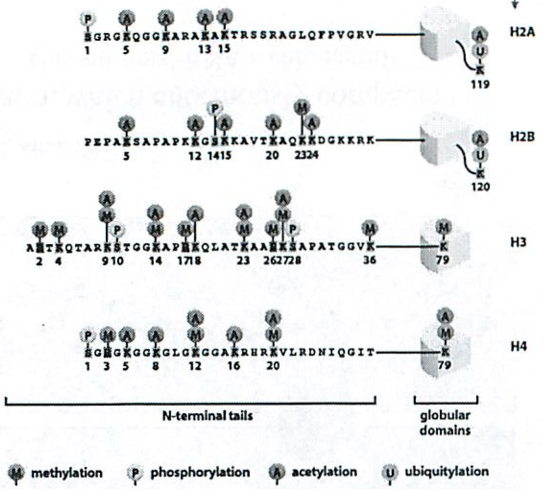
Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Here's a preliminary listing of the modifications of the N-terminal tails of the core nucleosomes (the C=termini are buried in the core nucleosome).

e.g., some of the methylations repress transcription, some of the acetylations induce transcription

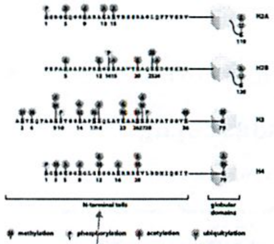
The 4 histones

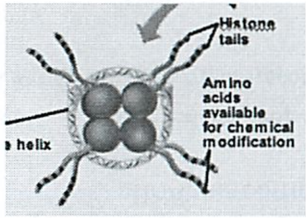

Histone tails
Amino acids available for chemical modification
helix


SGRGKQGGKARAKAKTRSSRAGLQFPVGRV
H2A
119
PEPAKSAPAPKKGSKKAVTKAQKKDGKKRK
H2B
120
ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVK
H3
79
SGRGKGGKGLGKGGAKRHRKVLRDNIQGIT
H4
79
N-terminal tails
globular domains
methylation
phosphorylation
acetylation
ubiquitylation

The histone code

Figure 4-39b *Molecular Biology of the Cell* (© Garland Science 2008)




Histone tails
Amino acids available for chemical modification
helix

To summarize: modifications of the tails of the histone molecules affect the configuration of the chromatin, allowing or disallowing transcription.

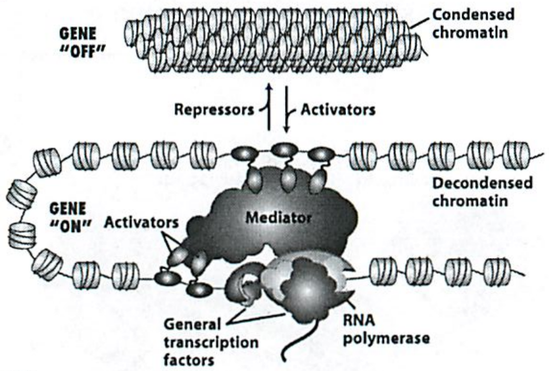
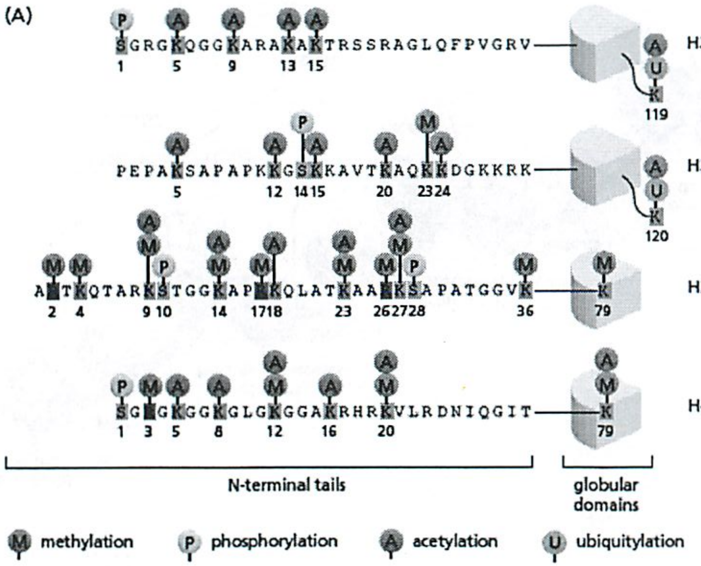
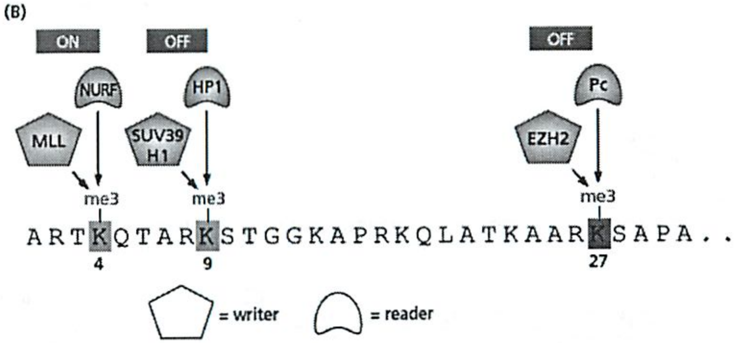

GENE "OFF"
Condensed chromatin
Repressors
Activators
GENE "ON"
Activators
Mediator
Decondensed chromatin
General transcription factors
RNA polymerase

Figure 7-1
Molecular Cell Biology, Sixth Edition
© 2008 W. H. Freeman and Company

Let's revisit the histone modifications


(A)
SGRGKQGGKARAKAKTRSSRAGLQFPVGRV
1 5 9 13 15
H2A
119
PEPAKSAPAPKKGSKKAVTKAQKKDGKKRK
5 12 1415 20 2324
H2B
120
ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVK
2 4 9 10 14 1718 23 262728 36
H3
79
SGRGKGGKGLGKGGAKRHRKVLRDNIQGIT
1 3 5 8 12 16 20
H4
79
N-terminal tails
globular domains
methylation
phosphorylation
acetylation
ubiquitylation

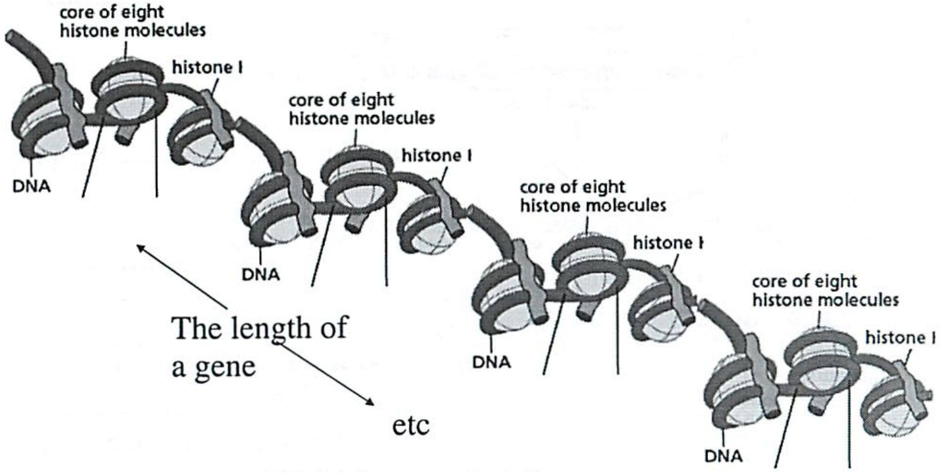

(B)
ON
OFF
OFF
NURF
HP1
Pc
MLL
SUV39 H1
EZH2
me3
me3
me3
ARTKQTARKSTGGKAPRKQLATKAARKSAPA..
4
9
27
= writer
= reader

Histone H3

me3 = trimethylation

2 x H2A
2 x H2B
2 x H3
2 x H4

core nucleosome

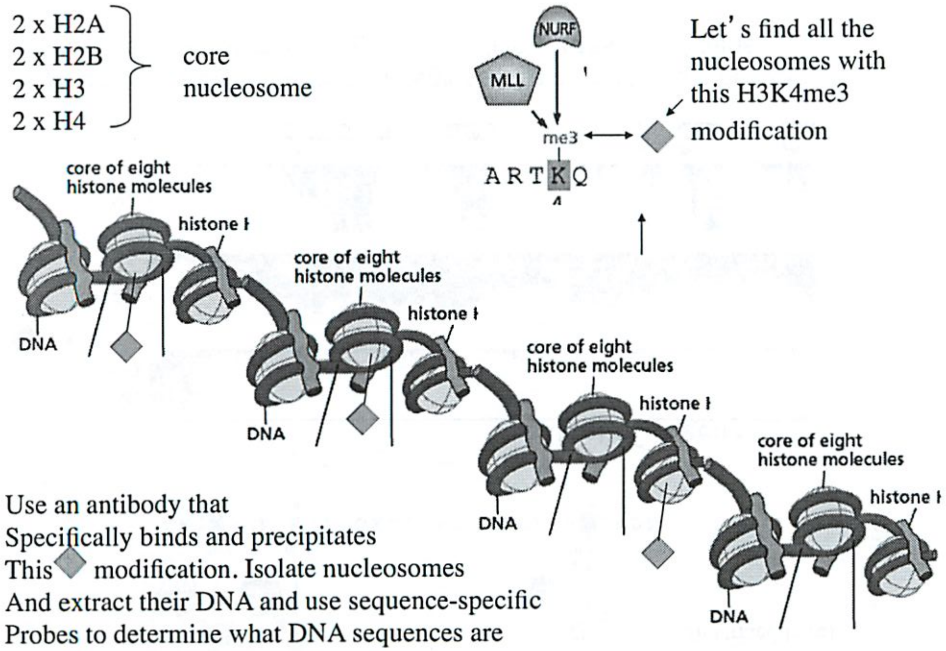

core of eight histone molecules
histone I
DNA
core of eight histone molecules
histone I
DNA
core of eight histone molecules
histone I
DNA
core of eight histone molecules
histone I
DNA

The length of a gene

etc

2 x H2A
2 x H2B
2 x H3
2 x H4

core nucleosome

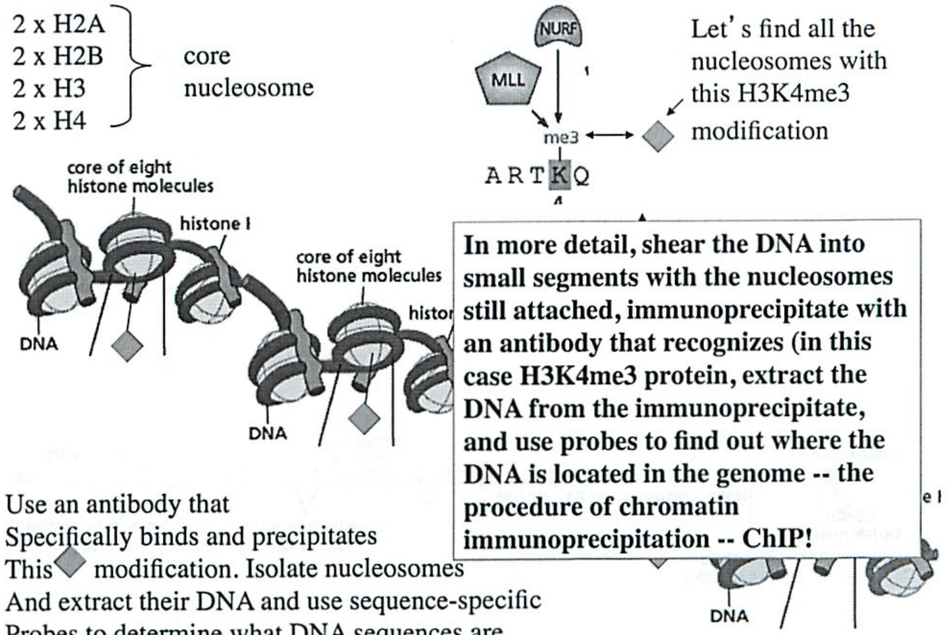

NURF
MLL
me3
ARTKQ
4
core of eight histone molecules
histone I
DNA
core of eight histone molecules
histone I
DNA
core of eight histone molecules
histone I
DNA
core of eight histone molecules
histone I
DNA

Let' s find all the nucleosomes with this H3K4me3 modification

Use an antibody that Specifically binds and precipitates This ◆ modification. Isolate nucleosomes And extract their DNA and use sequence-specific Probes to determine what DNA sequences are In the precipitated nucleosomes.

2 x H2A
2 x H2B
2 x H3
2 x H4

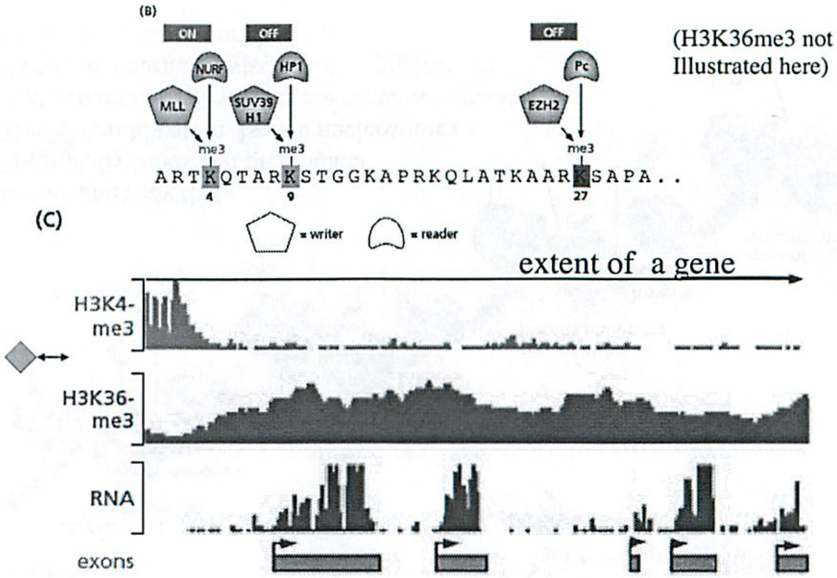
core nucleosome


NURF
MLL
me3
ARTKQ
4
core of eight histone molecules
histone I
DNA
core of eight histone molecules
DNA
DNA

Let' s find all the nucleosomes with this H3K4me3 modification

In more detail, shear the DNA into small segments with the nucleosomes still attached, immunoprecipitate with an antibody that recognizes (in this case H3K4me3 protein, extract the DNA from the immunoprecipitate, and use probes to find out where the DNA is located in the genome -- the procedure of chromatin immunoprecipitation -- ChIP!

Use an antibody that Specifically binds and precipitates This ◆ modification. Isolate nucleosomes And extract their DNA and use sequence-specific Probes to determine what DNA sequences are In the precipitated nucleosomes.

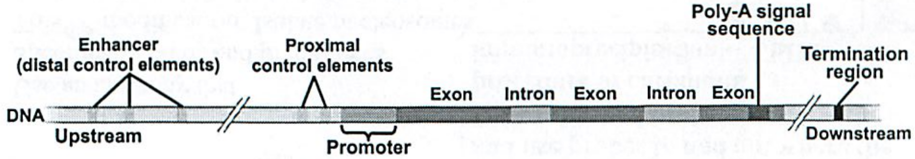


(H3K36me3 not Illustrated here)

H3K4me3 -- enriched in control regions of actively transcribed genes
H3K36me3 -- enriched in regions of actively transcribed gene

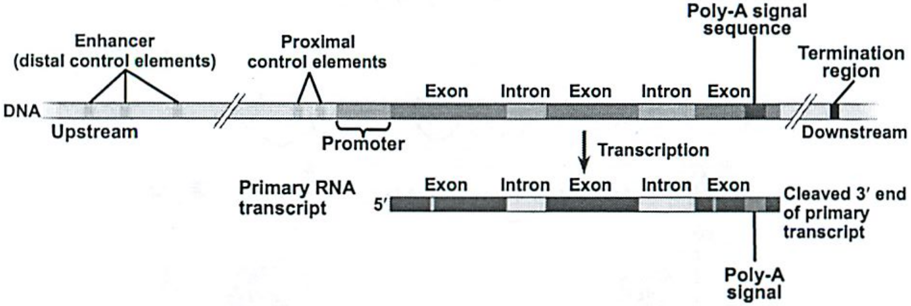
Fig. 18-8-1

The structure of a eukaryotic gene transcribed by pol II


Enhancer
(distal control elements)
Proximal
control elements
Poly-A signal
sequence
Termination
region
Exon
Intron
Exon
Intron
Exon
DNA
Upstream
Promoter
Downstream

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

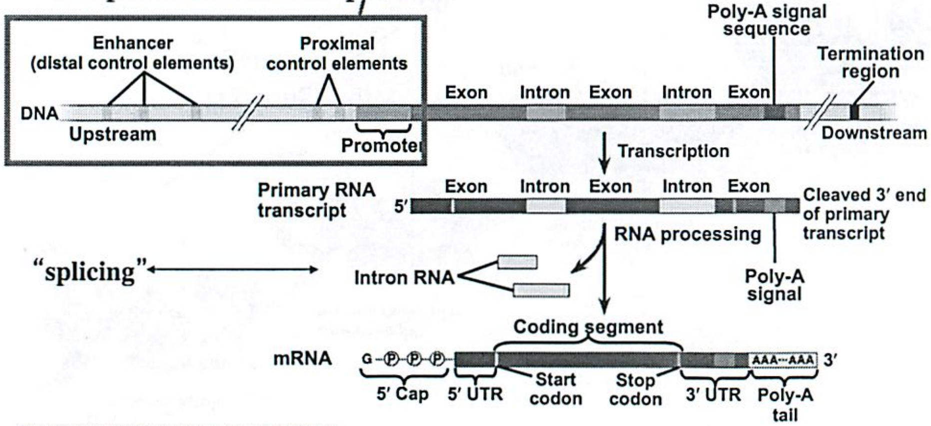
Fig. 18-8-2


Enhancer
(distal control elements)
Proximal
control elements
Poly-A signal
sequence
Termination
region
Exon
Intron
Exon
Intron
Exon
DNA
Upstream
Promoter
Downstream
Transcription
Primary RNA
transcript
5′
Exon
Intron
Exon
Intron
Exon
Cleaved 3′ end
of primary
transcript
Poly-A
signal

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

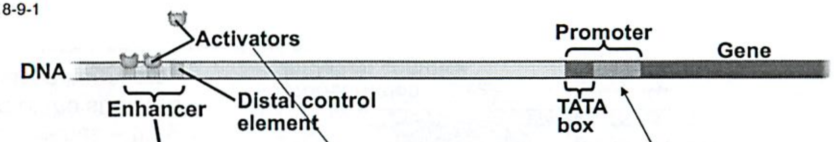
Fig. 18-8-3

Let's focus on the **control elements** of the DNA itself. Directly or indirectly they influence the structure of the histones and thus the process of transcription


Enhancer
(distal control elements)
Proximal
control elements
Poly-A signal
sequence
Termination
region
Exon
Intron
Exon
Intron
Exon
DNA
Upstream
Promoter
Downstream
Transcription
Primary RNA
transcript
5′
Exon
Intron
Exon
Intron
Exon
Cleaved 3′ end
of primary
transcript
RNA processing
"splicing"
Intron RNA
Poly-A
signal
Coding segment
mRNA
G-Ⓟ-Ⓟ-Ⓟ
AAA···AAA
3′
5′ Cap
5′ UTR
Start
codon
Stop
codon
3′ UTR
Poly-A
tail

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Fig. 18-9-1

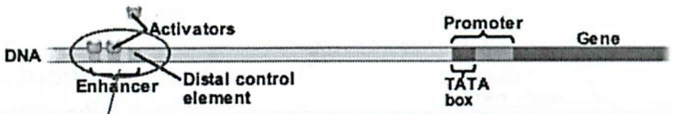

Activators
Promoter
Gene
DNA
Enhancer
Distal control element
TATA box

These **enhancer** sequences may be many kilobases upstream of the promoter

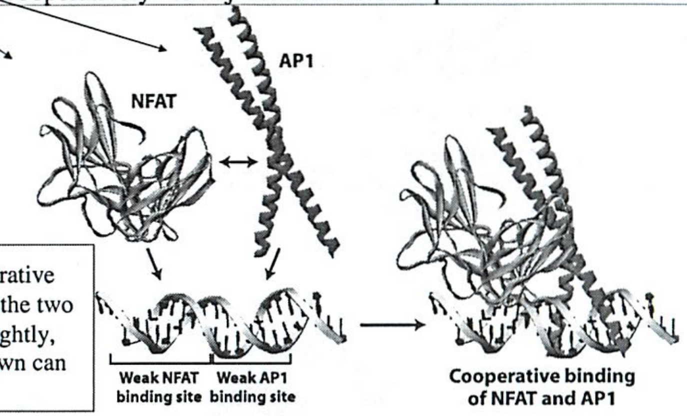
These **activator** proteins bind to the enhancer sequences and regulate the use of the promoter by RNA polymerase II.

Transcription starts here

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.


Activators
Promoter
Gene
DNA
Enhancer
Distal control element
TATA box

What do **activators** actually look like? Here is a pair of transcription factors (NFAT & AP1), each specialized to recognize a **specific nucleotide sequence** (enhancer) in the DNA, binding cooperatively* to adjacent enhancer sequences.

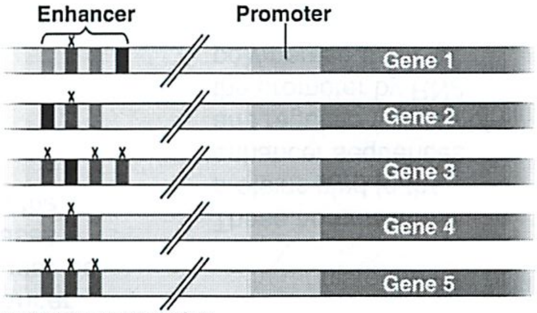

AP1
NFAT
Weak NFAT binding site
Weak AP1 binding site
Cooperative binding of NFAT and AP1

*In this case, cooperative binding means that the two can bind the DNA tightly, while either on its own can bind only weakly.

Figure 7-29
Molecular Cell Biology, Sixth Edition
© 2008 W. H. Freeman and Company

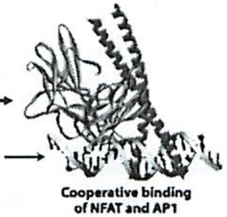
Fig. 18-UN7

There are hundreds, perhaps more, of distinct enhancer sequences. Each of these is recognized and bound by one or several DNA sequence-specific proteins (**activators**) that function **combinatorially** to govern gene transcription.


Enhancer
Promoter
Gene 1
Gene 2
Gene 3
Gene 4
Gene 5

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

This was one example of such combinatorial action


Cooperative binding of NFAT and AP1

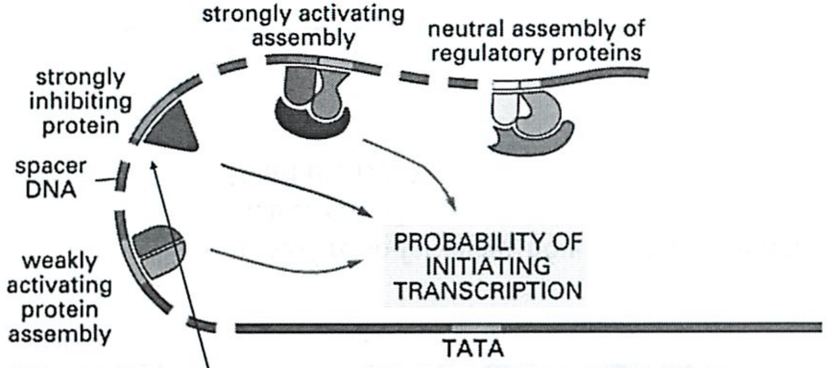
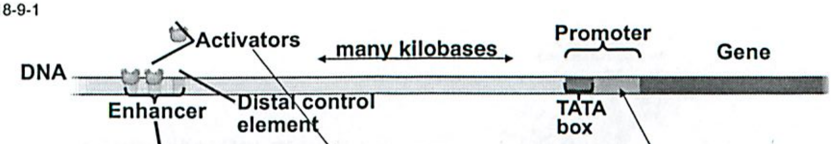

strongly activating assembly
neutral assembly of regulatory proteins
strongly inhibiting protein
spacer DNA
weakly activating protein assembly
PROBABILITY OF INITIATING TRANSCRIPTION
TATA

Figure 7–57. Molecular Biology of the Cell, 4th Edition.

In fact, some of these "activator" proteins may actually function **negatively** to repress transcription

Fig. 18-9-1


Activators
many kilobases
Promoter
Gene
DNA
Enhancer
Distal control element
TATA box

These **enhancer** sequences may be many kilobases upstream of the promoter

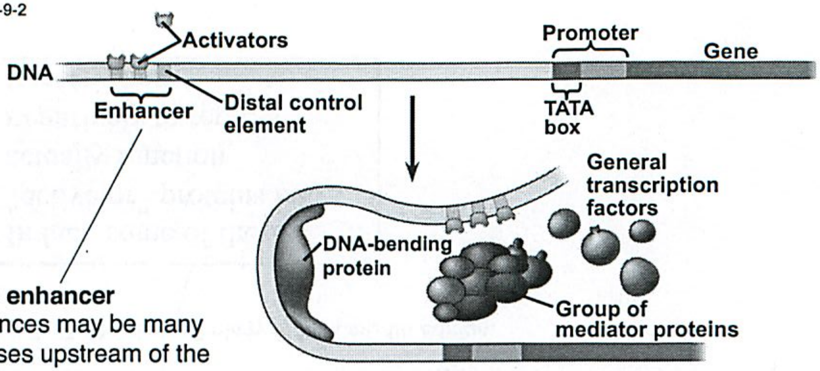
Transcription starts here

These **activator** proteins bind to the enhancer sequences and regulate the use of the promoter by RNA polymerase II.

If they' re so far away, how can these activators influence transcription?

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Fig. 18-9-2

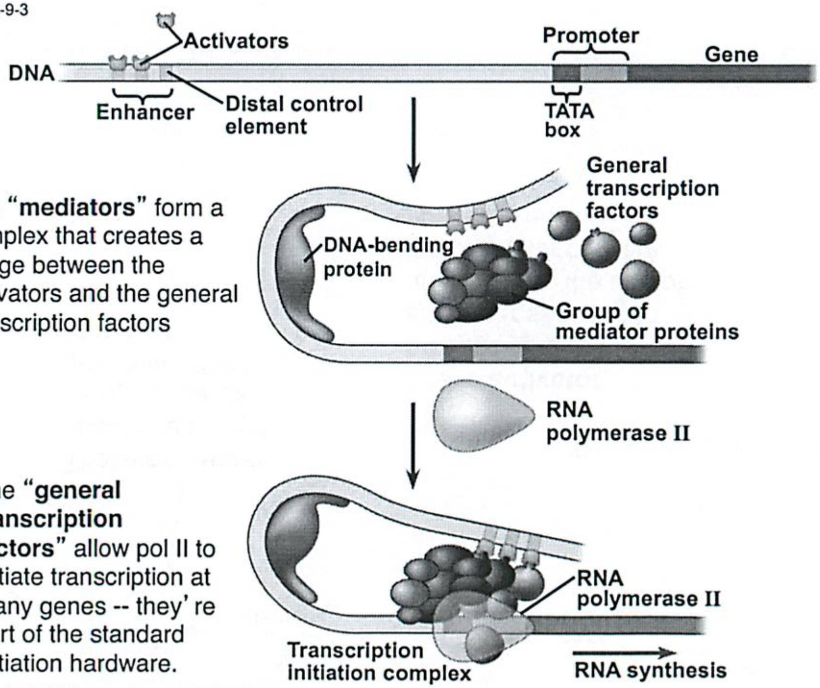

Activators
Promoter
Gene
DNA
Enhancer
Distal control element
TATA box
General transcription factors
DNA-bending protein
Group of mediator proteins

These **enhancer** sequences may be many kilobases upstream of the promoter

If they' re so far away, how can they influence transcription?
Bend the DNA!

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

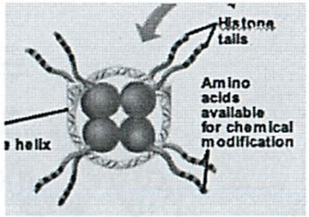
Fig. 18-9-3


Activators
Promoter
Gene
DNA
Enhancer
Distal control element
TATA box
General transcription factors
DNA-bending protein
Group of mediator proteins
RNA polymerase II
RNA polymerase II
Transcription initiation complex
RNA synthesis

The "**mediators**" form a complex that creates a bridge between the activators and the general transcription factors

The "**general transcription factors**" allow pol II to initiate transcription at many genes -- they' re part of the standard initiation hardware.

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.


Histone tails
Amino acids available for chemical modification
helix

Now we can depict the activators (and repressors) as agents that affect in complex ways histone modifications, and in turn the configuration of the chromatin, allowing or disallowing transcription.

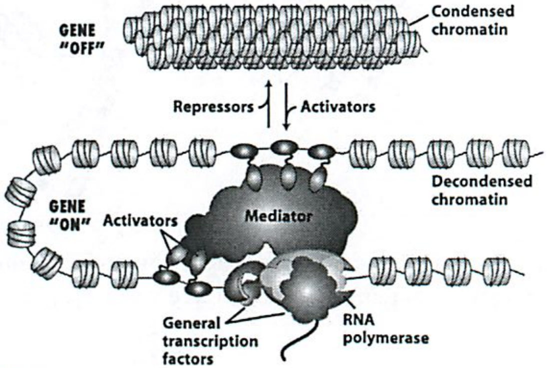

GENE "OFF"
Condensed chromatin
Repressors
Activators
Decondensed chromatin
GENE "ON"
Activators
Mediator
General transcription factors
RNA polymerase

Figure 7-1
Molecular Cell Biology, Sixth Edition
© 2008 W.H. Freeman and Company

Different cell types makes different complex mixtures of proteins

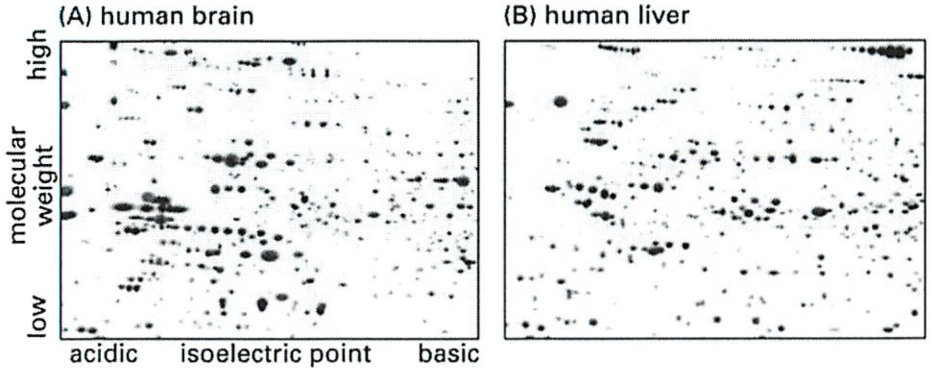


Figure 7–4. Molecular Biology of the Cell, 4th Edition.

Why is it important to be able to control thousands of genes in different ways? (Remember: protein-coding genes are transcribed by Pol II)

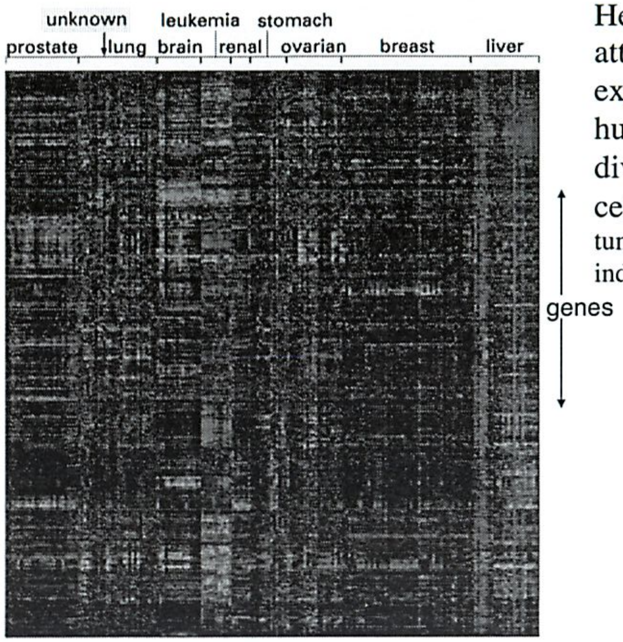


Figure 7–3. Molecular Biology of the Cell, 4th Edition.

Here is an example of an attempt to survey the expression of several hundred genes from a diverse array of human cell types (in this case the tumors arising in each of the indicated tissues).

Color key:

▨ High expression

■ Average expression

▨ Low expression

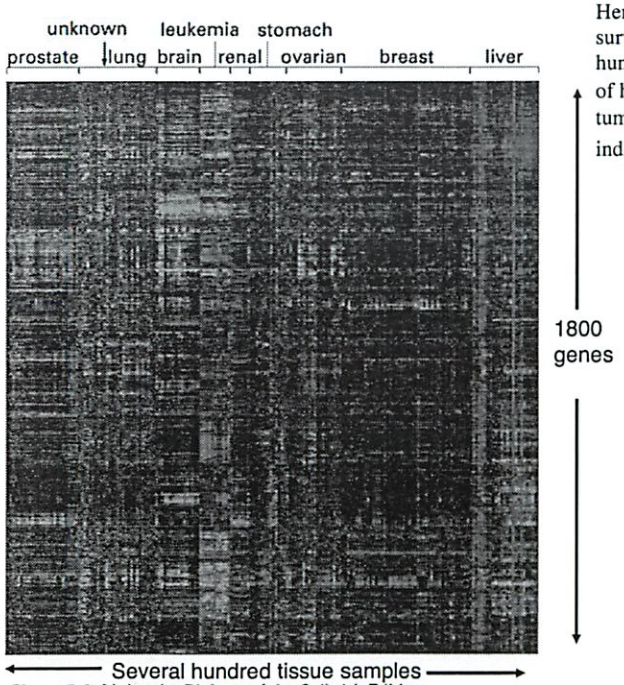


Figure 7–3. Molecular Biology of the Cell, 4th Edition.

Here is an example of an attempt to survey the expression of several hundred genes from a diverse array of human cell types (in this case the tumors arising in each of the indicated tissues).

▨ High expression

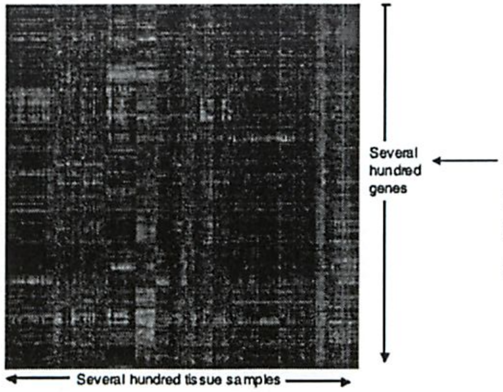
■ Average expression

▨ Low expression

High expression

Imagine

1. That 3-5 thousand genes are expressed at comparable levels in all cells because they are commonly used "housekeeping genes"
2. That there are ~15 thousand genes that are expressed at different levels in different cell types throughout the body (non-housekeeping genes).
3. That there are several hundred distinct cell types in the body, each with its own distinct pattern of expression of "non-housekeeping" genes.
4. Now you can appreciate why/how the combinatorial actions of activators and repressors can regulate different levels of gene expression in different cell types throughout the body.

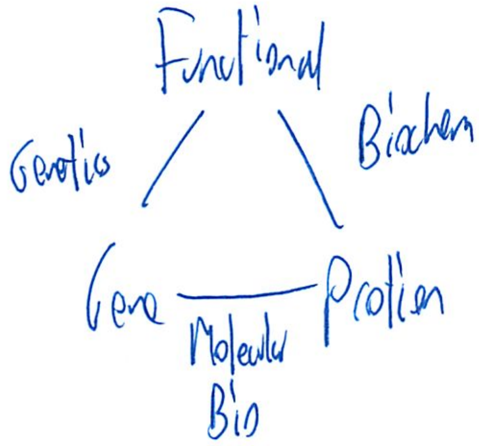


This gene expression array only measures the expression levels of several hundred "non-housekeeping" genes.

7.012
Recombenant
DNA 1

10/15

(1 min late)


Functional
Genetics
Biochem
Gene
Protien
Molecular Bio

So cells we made some artifical DNA
+ got protiens

But what DNA represents what thing?

Purify individual gene

but genome is 3×10^9 base pair

typical gene 3×10^4

typical base pair 1

~~[illegible]~~ 10^5 purification

10^4

Some wanted to give up
But others wanted to know which gene for which function
↳ recombinant DNA movement
→ 1970

Cloning - molecular cloning of pieces of DNA

find, purify, and propagate

DNA spreads out + breaks up (cut - at defined site)
Some random DNA is amplified
Creates a random DNA library

More Formally
1. Cut at defined sites

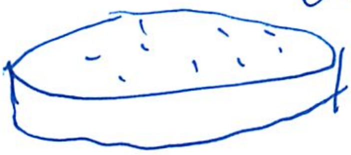
2. Paste into vectors that propagate

But must be in a cell

3. Transform

4. Select for cells that have that DNA

~~DNA~~ colonies of DNA



each has its own piece of DNA

a colony has lots of copies

Wed: How we select the right library

Cutting up DNA

Want to cut up at specific sites

5' GAATTC 3'
CTTAAG

I could cut everytime at this seq

How to engineer?

④

5' AG~~G~~C TA | G AATTC | TTAC 3'

TC GAT | C TTAAG | AATG

← cuts not even

turns at certain e-coli cuts like that

enzyme ⇒ EcoRI

~~the~~ why does E Coli do this?

; protect againts ~~DNA~~ Bacteria Phage

But will cut up its own DNA

(He said each combo is random

4·4·4·4·4·4 chance

we have GAATTC)

E coli protects its own chromosome

puts a methyl mer on the seq

(CH_3)

GAATTC

CTTAAG

(CH_3)

Likelyhood cut before it gets methylated
new DNA

So right most of the time
Which is good enough

So called restriction enzymes - restricts a virus's ability to grow

So today you can order in a catalog
Other restriction enzymes

~~BAM~~

EcoRI G|AATTC
CTTAA|G

BamHI G|GATCC
CCTAG|G

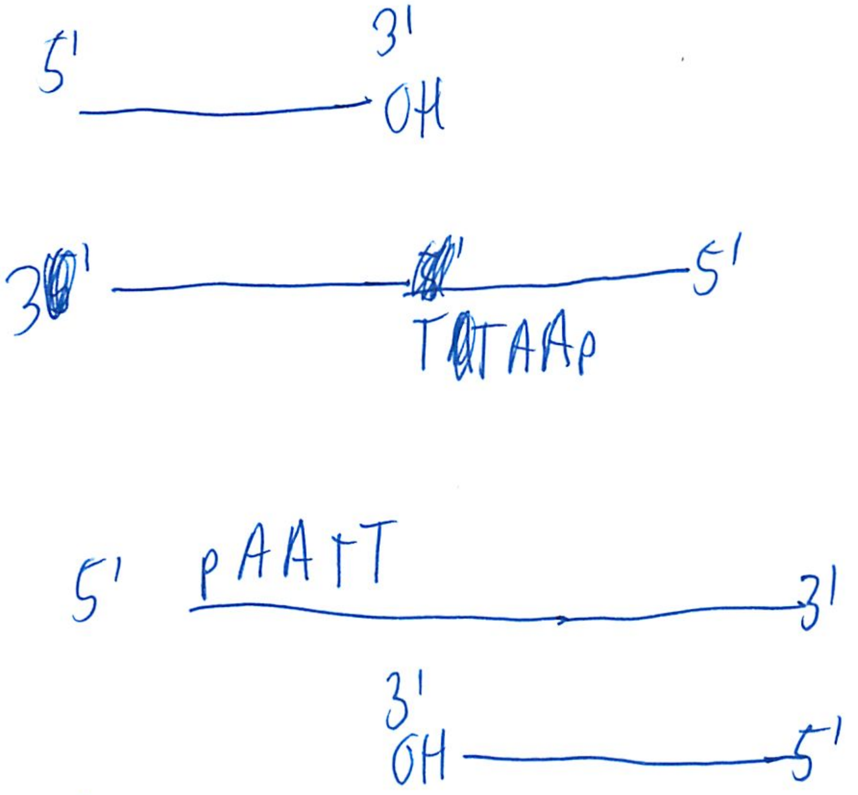
MboI |GATC
CTAG|

PvuII CAG|CTG
GTCGAC|

And thousands more! Order today!

⑥

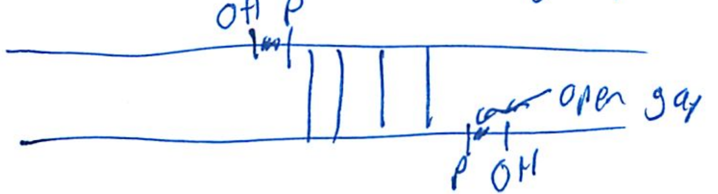
Paste DNA


5'
3'
OH
3'
5'
TATAAp
5' pAATT
3'
3'
OH
5'

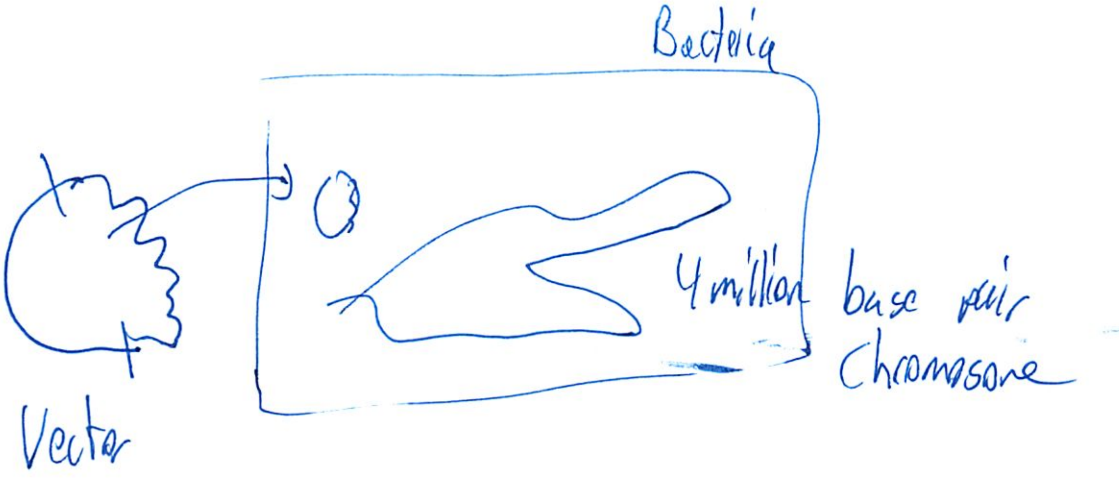
(kinda missed desciption)

Could hydrogen bond

- Have holes in double helix
- Should join fragments to close helix
  - ligate
- So use DNA ligase
  - another thing you can just order


OH p
open gap
p OH

Vectors

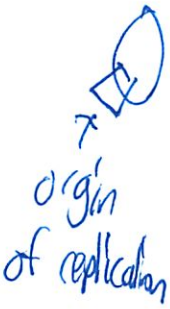

Bacteria
4 million base pair Chromosome
Vector

(missed)

So how do we build this vector?

Again look at Ecoli

Ecoli already has little mini chromosomes


origin of replication

~~a~~ ~4,000 - 10,000 bases
get replicated

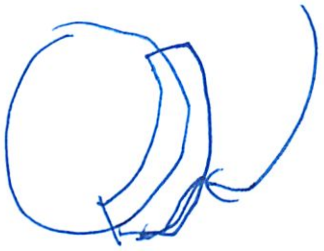
= "plasmids" aka "$\underbrace{\text{epi}}_{\text{on top of}}\text{somes}$"

Why do Bacteria have?

They carry little interesting fns

Prof: like apps in the App Store

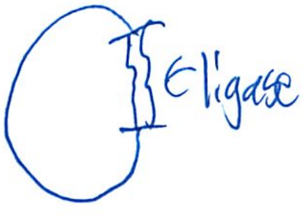
- for immunity againts antibodies
- bacteria are trying to become resistent to attacking chemicals
- so antibiotic resistence gene



How will we purify DNA?

not hard to purify little circles away from long strand of DNA

Can cut w/ restiction enzyme

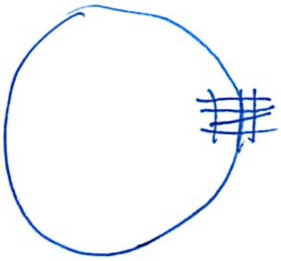


Whole section of plasmids in catalog

What if it didn't have pefect Ecoll section

Might use diff enzume

Or company might tweak the seq



~~[illegible]~~ poly linker

can use a bunch of diff enzyms

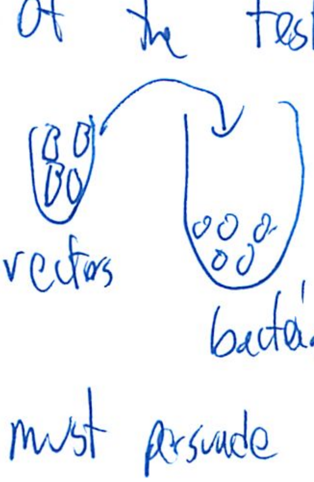
now engineered in many plasmids

sold in the catalog

Transforming Bacteria

Now must get into bacteria

Out of the test tube

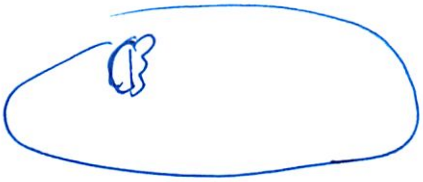


must persuade bacteria to take up DNA

bacteria is constantly picking up DNA from env
never know what ya might get
aka transforming the bacteria

Similar to what Griffiths did

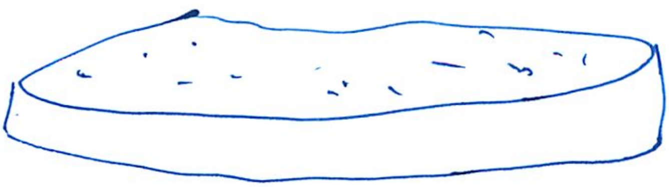
Some will get a vector



But only some of them get it
("small fraction"

Which ones?

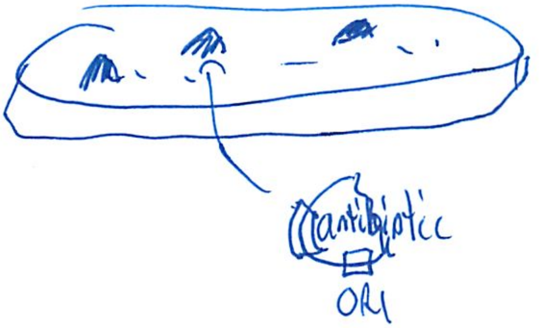
Select for Presence of Vector



Only want one w/ Vector to grow up

Vectors have naturally occuring antibiotic resistence ~~vector~~ marker

Then when plate them out



(missed)

Now gets antibiotic resistance

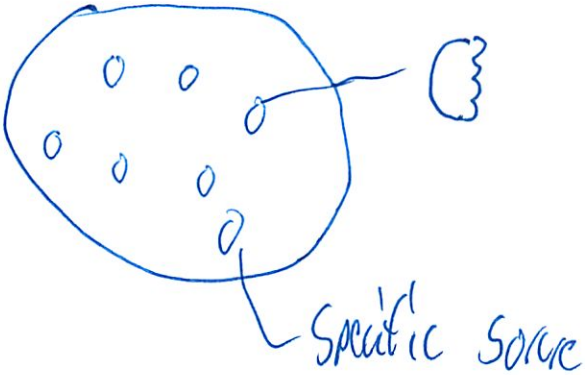
don't want strand w/ already antibiotic resistance

So get the right one out of the catalog

This is 70s + 80s developed

↳ So good tools

Library



Who library of books w/o cover

if enough statistically gurentee whole human genome is there

Vectors

We had been talking as if only thing can transform is human DNA w/ ~~[illegible]~~ ecoli

But are other things

~~A~~ Vector

- Bacterial plasmids	Circles	1-5000 BP	grow in Bacteria
- Yeast plasmids	Circles	"	" yeast

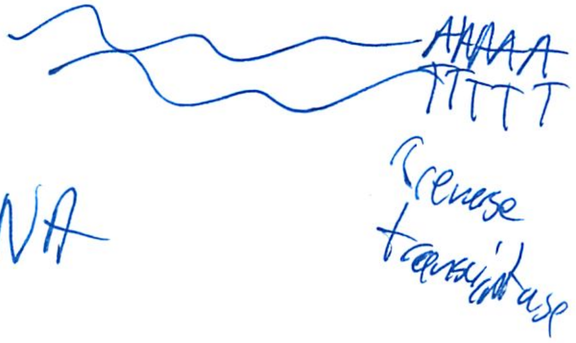
⑬

Basically anything that could grow in a cell
└ that can replicate

- Bacterial virus linear 40,000 BP grow in Bacteria
- Mamelian virus " " grow in mamelian cells
- Artifical chromsomes
- Human DNA
- Zebra DNA } any DNA attached to any vector

Libraries

- Genomic DNA
- RNA - reverse transciptase



turn RNA into DNA
└ called cDNA

- very general
- can purify anything - make a lib of indv pieces

Next lecture: How you get your book out of that library?

10/16
10/18

Samin Houshyar
R27 & R28
Recitation 9 & 10

1. Gene Expression – How do we get different cell types if all of our cells have the same DNA?

    a. Acetylation

    b. Methylation

2. Gene Regulation:
    a. *lac operon*

        i. What is the signal?
        ii. What is the response?
        iii. Schematic of the *lac operon*

        iv. Components of the *lac operon*

        v. Mutations in the *lac operon*. The WT *lac operon* is ___________.

            1. Constitutive

            2. Non-inducible

        vi. Genetic tool: X-gal

- b. *Trp operon*
    - i. What is the signal?
    - ii. What is the response?

3. Recombinant DNA

    - a. Cloning
        - i. Restriction enzymes and sites
            1. Origin
            2. Specificity
            3. Sticky and blunt ends
        - ii. Vectors
            1. Properties of a vector
                - a.
                - b.
                - c.
            2. Types of vectors
                - a.
                - b.
                - c.
        - iii. Steps of cloning:
            1. Cut
            2. Ligate

3. Transform

iv. Ensuring that the cloning process worked:

1. Did the cell take up the plasmid?

2. Did plasmid take up the gene?

3. Did the gene enter in the right orientation?

b. Gel Electrophoresis (can be used for RNA and Protein as well)

i. Principles

ii. What is the gel measuring?

iii. How many bands would you see?

1. Linear DNA cut once unevenly:

2. Linear DNA cut once in the middle:

3. Circular DNA cut once:

c. PCR –What is the basic principle?

i. What are the basic components?

1.

2.

3.

ii. Cycles of PCR

1. Denature

2. Anneal

3. Extend

d. Sequencing (Sanger Method) - What is the basic principle?

i. What are the basic components?

1.

2.

3.

4. Genomes
    a. Libraries
        i. Genomic

        ii. cDNA

    b. Reading the library: Cloning by complementation

c. SNP analyses
   i. What are SNPs?

   ii. How do we use them to predict inheritance of a disease?

d. Microarrays

   i. What are they trying to measure?

   ii. Why is this useful?

7.012

Recitation R4

10/16

(Got PSet 3 back

9/17

I was thinking more 15-16

Perhaps I do need much more help

Since I thought I understood this)

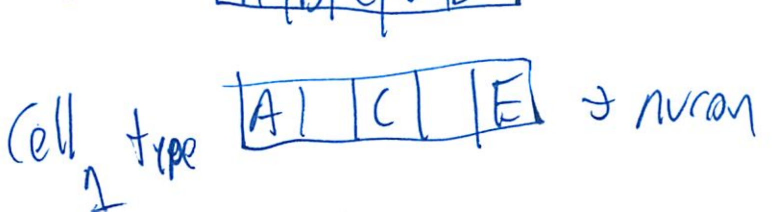
Next exam i 2 weeks

Gene Expression Regulation

All of our cells have same DNA

But lots of diff types of cells

differential expression

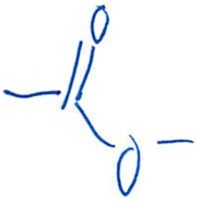

DNA
A
B
C
D
E
Cell type 1
A
C
E
→ nuron

②

Cell type 2 → skin cell

: Control of gene expression

methylation = add methyl groups ($-CH_3$)
compacts DNA
wraps around histones
genes not accessible for transciption
Since packed up + put away

acetylation = add 

has negtive charge
causes it to dissociate from histone molecules
makes DNA avilable for transciption

③

- transcibe ?
- mRNA
  - stable or unstable ?
- inhibit translation
- degrade protien

diff levels of how/what expressing

can modify things at any point in process

---

eaisest is stopping transcition
since each level requires energy
stopping at top = least energy

---

looking at for prok

---

lac operon

need a ~~[illegible]~~ signal + a response

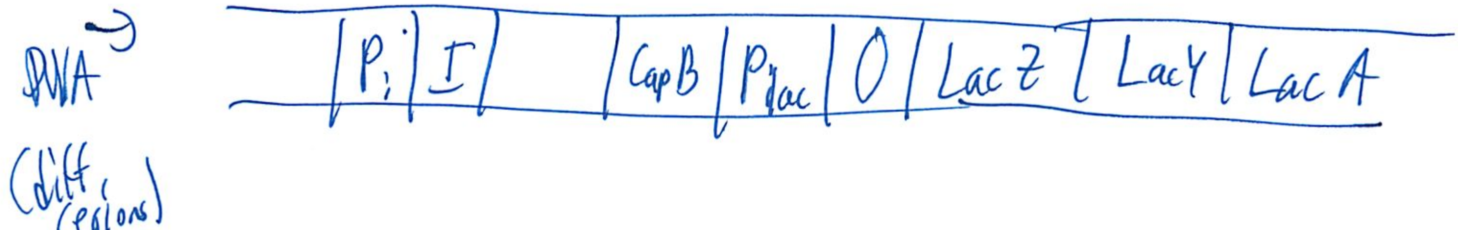
~~signal~~

④

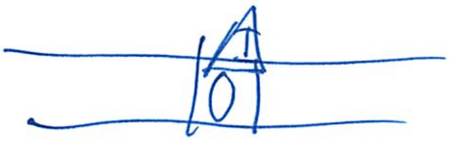
Signal + lactose
− glucose

lactose → glucose + galactose

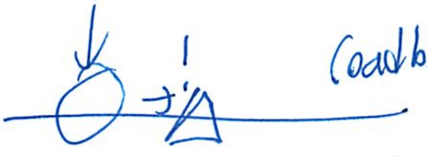
Response Break down lactose


DNA
(diff. regions)
P_i
I
CapB
P_lac
O
Lac Z
Lac Y
Lac A

normally when glucose present
have repressor on operator
└ before actual gene



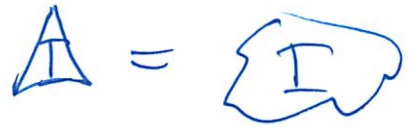
RNA polymerase can't get through!



roadblock!

⑤

When lactose is present

lactose O + 

repressor
inactive
falls off from operator

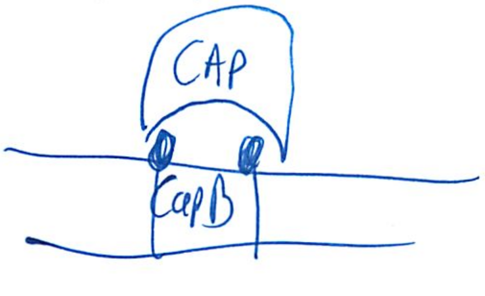
transciption + translation can happen

How do we sense absense of glucose?

↓ glucose = ↑ cyclic AMP

CAMP will bind to CAP

CAP binds to CAP B
↳ which is a transciptional enhanser



⑥

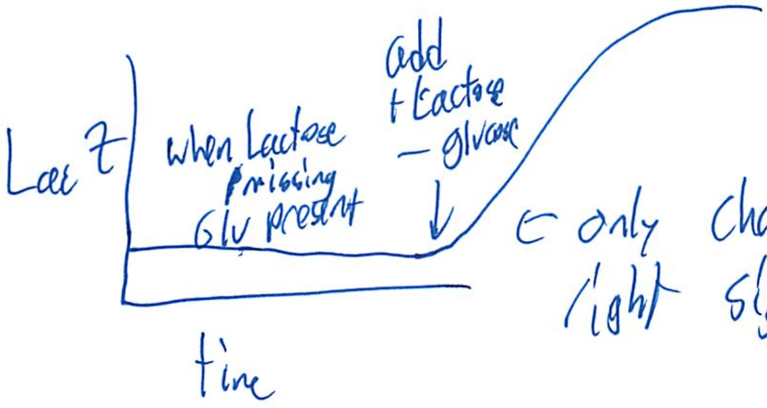
CAMP is a nucleotide

Lac Z → β-galactosidase → breaks down Lactose

Lac Y → Permease → brings in lactose

Lac A → People have no idea

Inducable system



← only changes when light signals present

1. Constitutive - always making LacZ
   no control one

2. Non Inducable - no matter what signal will
   never make LacZ

⑦ ✓ mutations that make

1. Constitutive mutations ✓ must be specific
   - Repressor can't bind to operator
   - P_i^-
   - O doesn't bind Repressor

2. Non-inducable
   - Repressor can't bind lactose
   - Lac Z^- ← non functional
   - P_{lac}^-

And plenty more

Recitation 7 Qu

#5. +lactose − glucose ✓

− lac + glu X

+ lac + glu X

− lac − glu X

based on lac operon system

have lac

and no glu

e) Set of mutations we have

w/o lac w/o glu

w/ Lac

Mutation	− lac − glu	+ lac
repressor can't bind to DNA	Constitutive - always expressed	Constitutive
P_{lac}^-	non-inducable	non-inducable

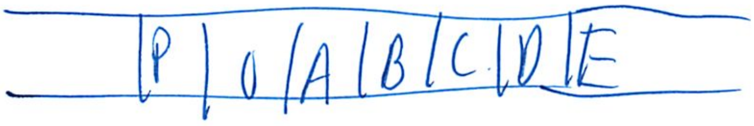
⑨

O^-	Const	Const
R can't bind lac	Non-inducable	Non-inducable
RNA $Poly^-$	Cell Dead	Cell Dead
Repressor locked as if lactose present	Inducable	Constitutive

← I asked isn't that only its children, Didn't seem to matter

---

Triptofan



System is always on

When tryp ↑

repressor becomes active and goes and binds to operator

⑩

Recombinant DNA

Can take DNA from diff org. and put it togeter to make new DNA

1. Cut DNA
2. Ligate into new piece of DNA
3. Transform

So 1st thing we need:

Cestriction enzymes

← cut off the virus

= enzyme

- like ammunal

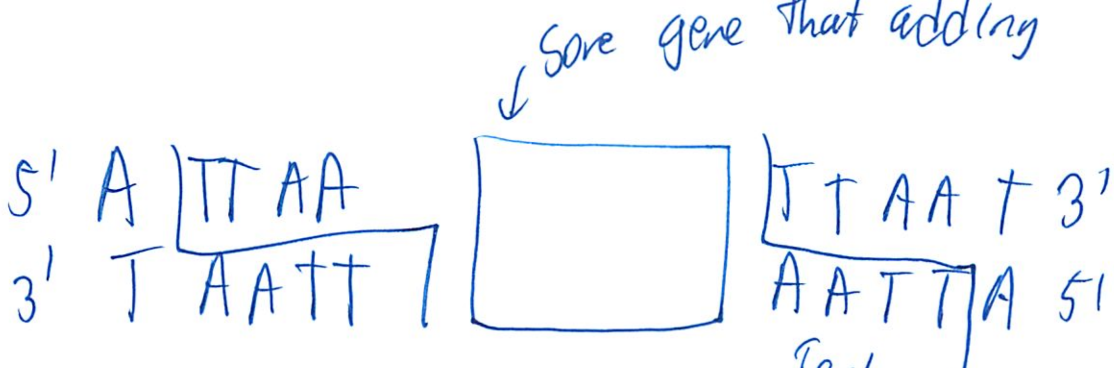
So we use them to cut DNA

~~Usually~~ always → palendromes

don't need to menoize seq

5' A|TTAAT 3'
3' TAATT|A 5'

⑪

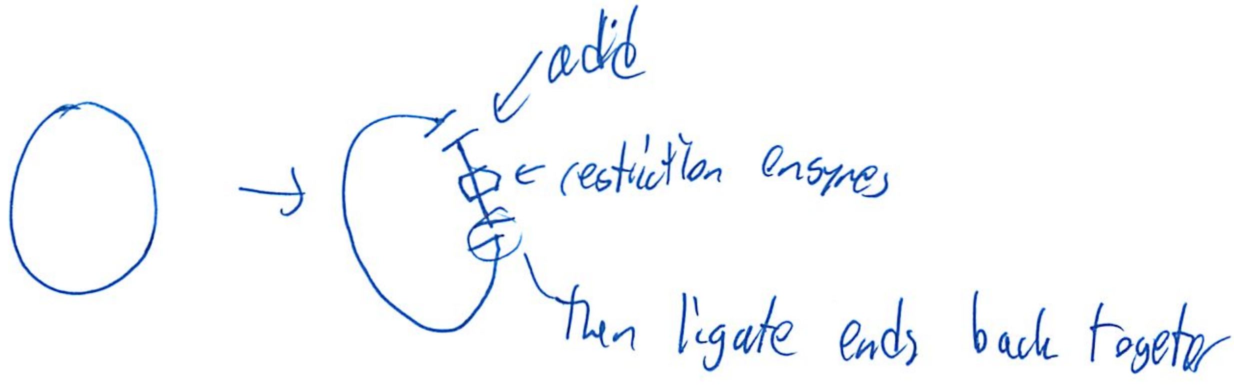

Sore gene that adding
5' A TT AA
3' T AATT
TTAAT 3'
AATTA 5'

ends must match
pad to get it to math

Vector = holde of DNA

= Plasmid
viruses
artifical chromosmes

Plasmid → opens up


add
restiction ensynes
Then ligate ends back togeter

A Few problems may occur?

Did the Plasmid take up the DNA
or did it close back on itself?

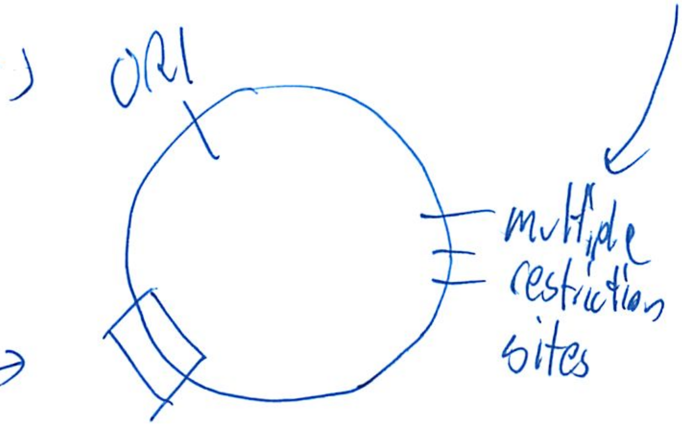
~~[illegible]~~ ⑫

- Bacteria might not take up plasmid
- Could go in backwards

Min, Properties of Vector

need piece of DNA to replicate ind of ~~cell~~ bacteria

need multiple restriction sites


ORI
multiple restriction sites

Selectable marker
usually antibiotic resistance
if have that gene, will survive

2012 7.012 Recitation 9

Summary of Lectures 14 & 15:

Recombinant DNA is DNA that contains fragments of DNA from two different organisms. The standard reagents necessary for recombinant DNA technology are restriction enzymes, the enzyme DNA ligase, and vectors. Restriction enzymes have been co-opted by scientists from bacteria, which use these enzymes as a form of their immune system. These enzymes cut up DNA so that one can isolate any piece of DNA from a genome that one wants. DNA ligase is the enzyme used by cells in DNA replication to join together the fragments that get replicated on the lagging strand. In recombinant DNA technology, DNA ligase is used to join together pieces of DNA that have been cut by the same restriction enzymes. Vectors are pieces of DNA that contain qualities that allow any piece of DNA to be taken up by cells and replicated. The most well-known example of a vector is a plasmid, which is a small circular double-stranded DNA molecule that bacteria are capable of taking up and replicating.

Plasmids used as vectors to get DNA into bacteria have three important features – restriction enzyme cut sites, an origin of replication, and a gene that encodes a protein that makes a bacterium resistant to a certain antibiotic. The restriction enzyme cut sites are present such that you can cut the plasmid open and insert in the piece of DNA in which you are interested. The origin of replication is present so that the plasmid can actually be replicated once it is in the bacterial cell. The antibiotic resistance gene is there so that you know that the bacterial cell took up the plasmid from the environment.

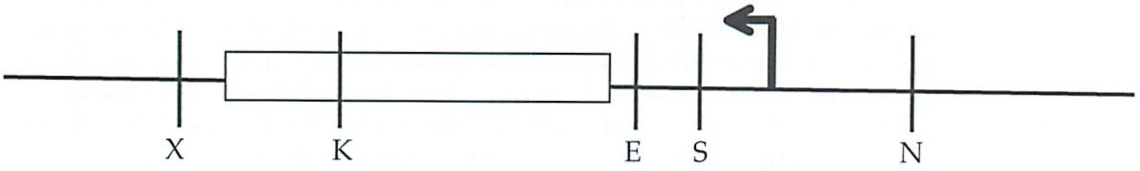
Once you have your gene of interest inserted into your plasmid, you transform the plasmid into the bacteria (i.e. change the growth conditions to encourage the bacteria to take up DNA from the environment). You then grow the transformed bacteria on plates that contain the specific antibiotic to which the gene on the plasmid confers resistance. Any cell that took up the plasmid will grow on medium containing this antibiotic compound, and any cell that did not take in a plasmid will die on this medium.

Recombinant DNA technology can be used to make a "library". A library is a collection of different recombinant DNA molecules (often stored in bacterial cells or phage), the set of which represents all of the genetic material of an organism. A mouse genomic library, for example, would be a population of host bacteria, each of which carries a piece of mouse DNA that was inserted into a cloning vector, such that the collection of cloned DNA molecules represents the entire genome of the mouse. An alternative to a genomic library would be a cDNA library. A cDNA library represents, not the entire genome, but only the DNA that is transcribed. cDNA is complementary DNA, which is DNA that was made in the laboratory by isolating total mRNA from the host organism and copying each mRNA molecule into a double-stranded DNA molecule. Each cDNA is then cloned into an appropriate vector, and the set of recombinant molecules is referred to as the cDNA library.

Gels are slabs of materials such as agarose and acrylamide that form gelatinous matrices when polymerized. DNA, RNA, and proteins can be inserted into such gels, the gels can be immersed in liquid, and then a current can be applied to the gel such that these macromolecules move towards the pole to which they are attracted. Very long macromolecules will move slowly as they attempt to weave their way through the pores in the gel. Smaller macromolecules will move faster. This technique of gel electrophoresis allows macromolecules to be separated by size

Questions:

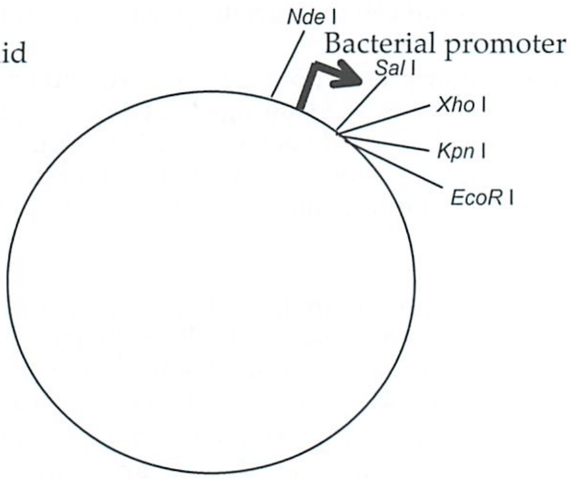
1) You want to insert a specific yeast gene into a specific bacterial plasmid such that the yeast gene will be transcribed in the bacterial cell. Below is a restriction map of a portion of the yeast chromosome that contains the yeast gene in which you are interested. The box indicates the open reading frame of the gene. The ↱ represents the promoter



Below are the enzymes you can use, with their specific cut sites shown as `5'-XXXXXX-3'`
`3'-XXXXXX-5'`

Nde I:	*Sal* I:	*EcoR* I:	*Xho* I:	*Kpn* I:
`CA↓TATG` `GTAT↑AC`	`G↓TCGAC` `CAGCT↑G`	`G↓AATTC` `CTTAA↑G`	`C↓TCGAG` `GAGCT↑C`	`GGTAC↓C` `C↑CATGG`

Below is the map of the plasmid



a) Your task is to design a strategy to insert the yeast gene into the bacterial plasmid. With which one set of enzymes would you choose to cut the yeast genomic DNA and the plasmid, out of the following choices?

1. NdeI and XhoI

2. SalI and KpnI

3. SalI and XhoI

4. XhoI and EcoRI

b) For each of the pairs you didn't choose, explain why you didn't choose them.

1. NdeI and XhoI

2. SalI and KpnI

3. SalI and XhoI

4. XhoI and EcoRI

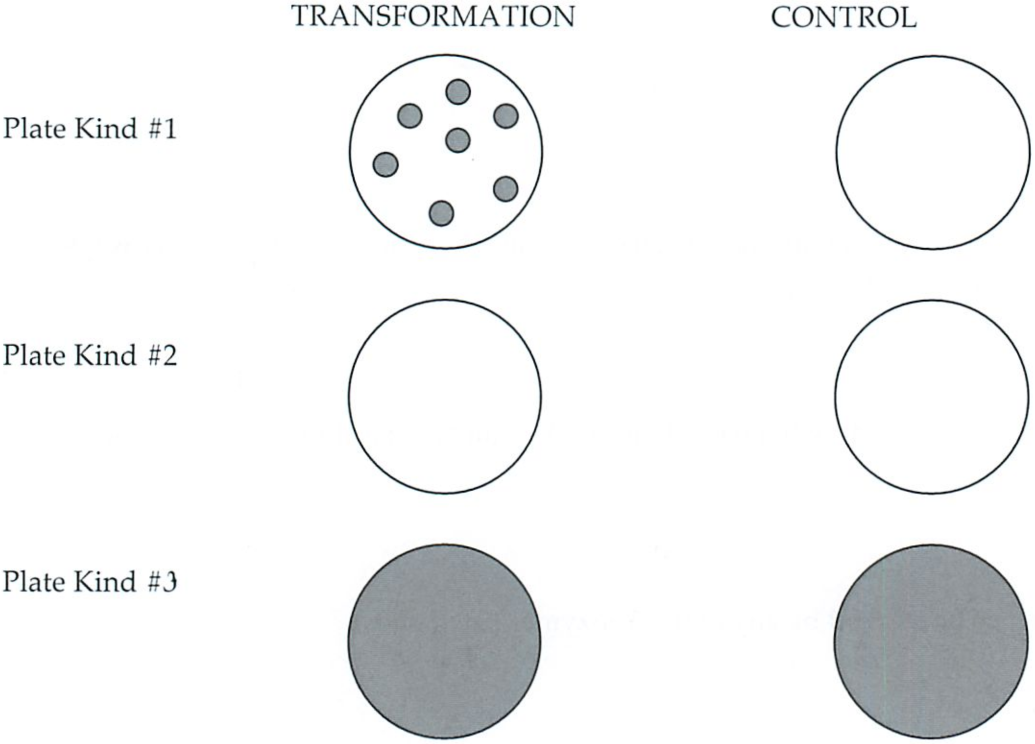
c) If you did the digestion and ligation with the two enzymes you chose above, in how many ways could the insert be inserted into the vector?

d) If the insert was inserted backwards, what would the DNA sequences be at the two sites where ligation happened?

e) Could the above sequence be cleaved by any of the 5 enzymes listed above?

f) Draw a map of the two different plasmids that would result from cutting the yeast genomic DNA and the plasmid and ligating them together. Indicate all promoters, restriction enzyme sites, and open reading frames in your map. Star the one you want.

2) You transform a plasmid that contains a gene that confers resistance to ampicillin into bacteria that are not resistant to any antibiotic. You then divide the transformation into thirds and plate it on three different kinds of plates. You also divide a control tube of bacteria that have not been transformed into thirds and plate it on the three kinds of plates you used. You incubate the plates overnight and come back the next morning to find the following results. Grey indicates bacterial growth.



a) Which set of plates is which? Your choices are: plates with no antibiotic added, plates with ampicillin added, plates with kanamycin (another antibiotic) added.

Plate Kind #1 =

Plate Kind #2 =

Plate Kind #3 =

b) What is an antibiotic?

c) Where do antibiotics come from?

d) How do antibiotics work?

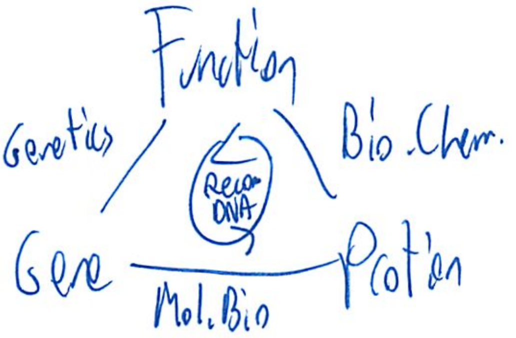
e) How might a gene confer resistance to an antibiotic?

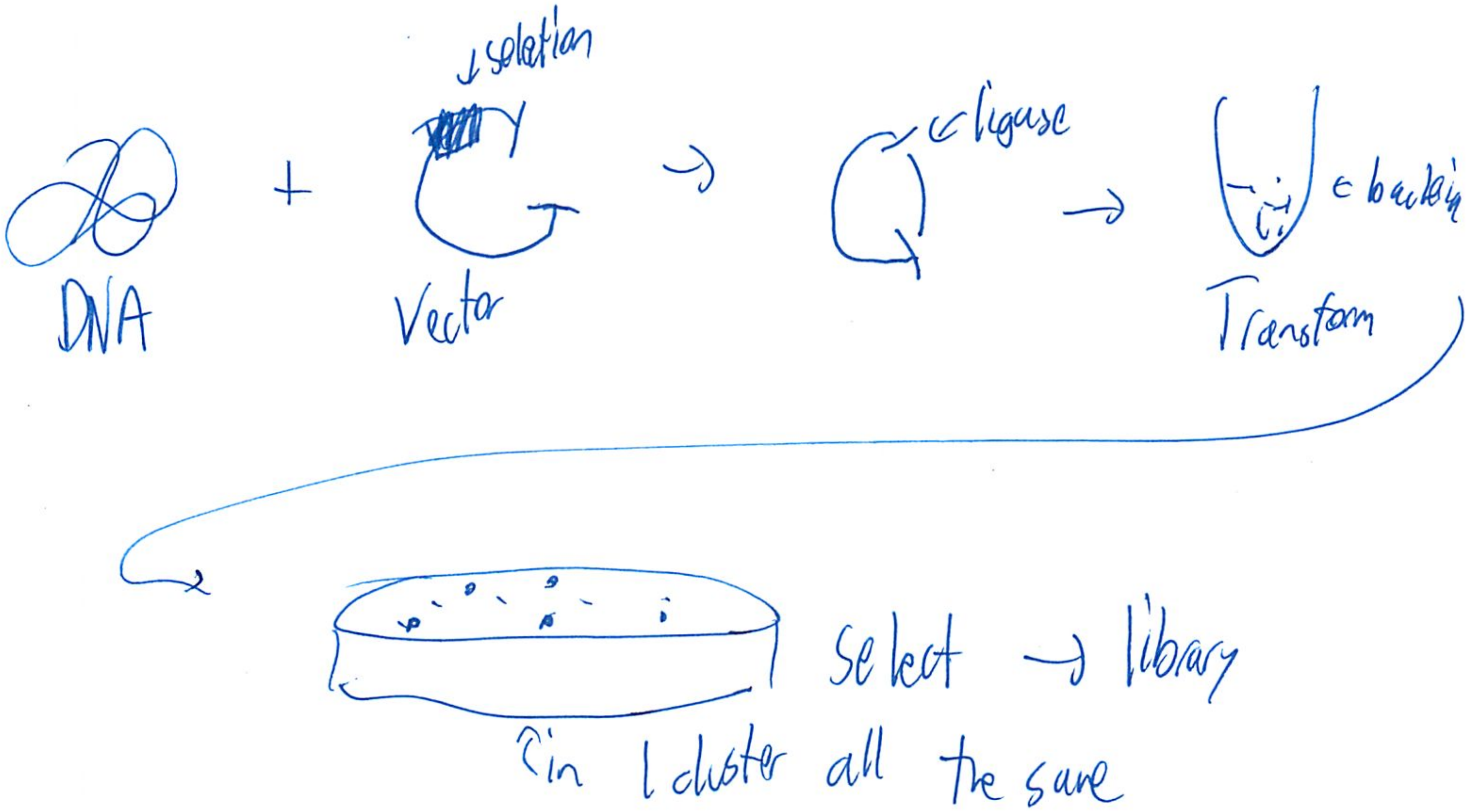
7.012

Recombinant DNA #2

10/17

(5 min late)

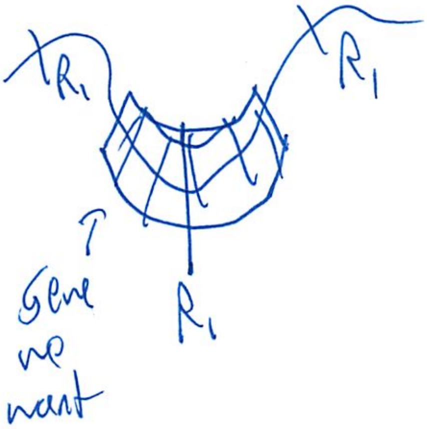

Function
Genetics
Bio.Chem.
Recom DNA
Gene
Protein
Mol.Bio


selection
ligase
bacteria
DNA
+
Vector
Transform
select → library
in 1 cluster all the same

②

But what if

R_1 = restriction enzyme



Can't cut w/ R_1 → since middle

Could use other restriction enzyme

Or control how much restriction
so only cuts half
at random

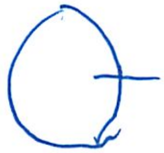
or add methylization
play w/ cation

Could shear w/ physical force
then cut

→ lots of engineering tricks

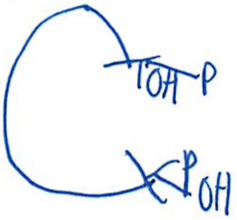
③

ligase could just close up vector
w/o an inset



Could it grow
Yes → since still has antibiotic restricter

Cut from 2 enzymes,
so ends are incompatable

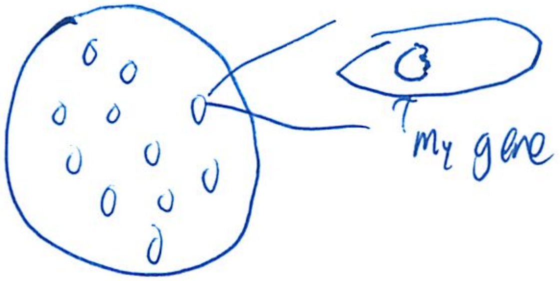


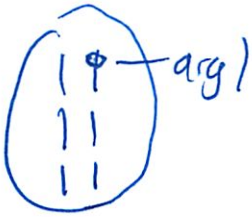
Could take off the phosphates
↳ w/ phosphatase

~~Vect~~ and inset piece of DNA supplies phosphates
Still a little gap on each side
Those don't matter
Cell repairs it for you

④

How to find your gene in a library



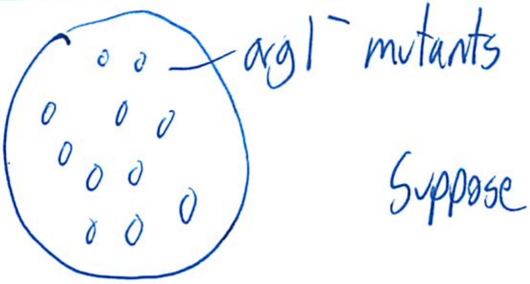


Problem: arg1 mutant
Can't grow on minimal medium

So make (missed)

What is special about arg1 gene
Complementation test
recessive autosomal (? heard correct)
Only need 1 good copy
add back wild type

a) Finding arg1 gene

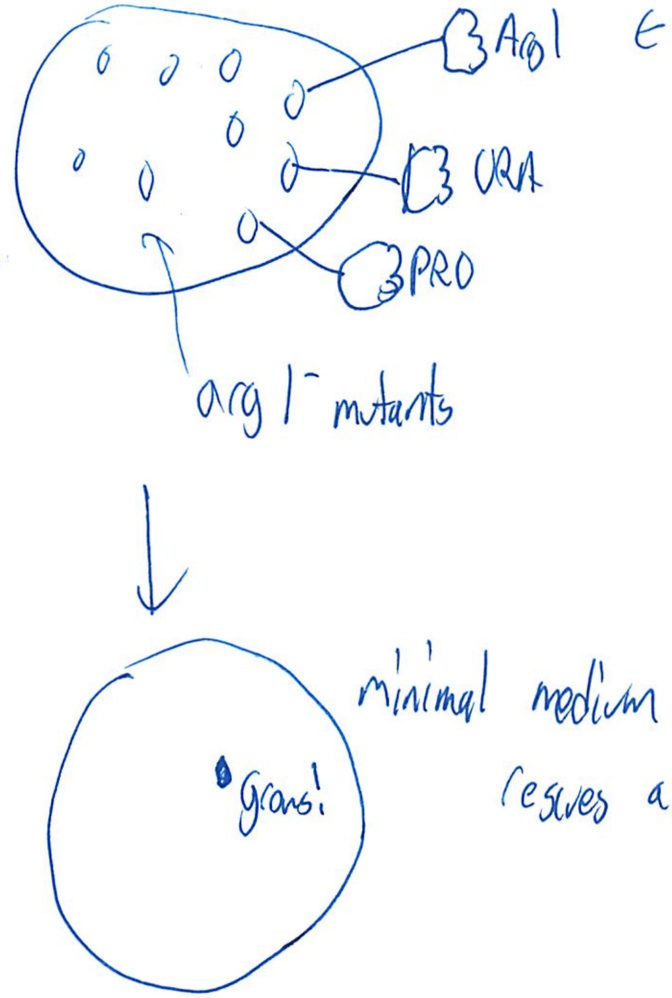


Suppose made a lib in these cells

⑤

each would carry various genes

Wild type DNA w/ mutant cell


Arg1
URA
PRO
arg1⁻ mutants
minimal medium
grows!

← only cell that can grow on minimal medium

rescues a phenotype

Cloning by complementation

What DNA did we start w/
wild type DNA
into mutant cells
to rescue the defect

* Each cell picks some (1 part) of the wild type DNA

engineering tricks so less duplicates

if other defects
↳ same trick
cloning by complementation easy

beta globin | 0

finding the human beta globin gene

but can't clone by complementation in humans

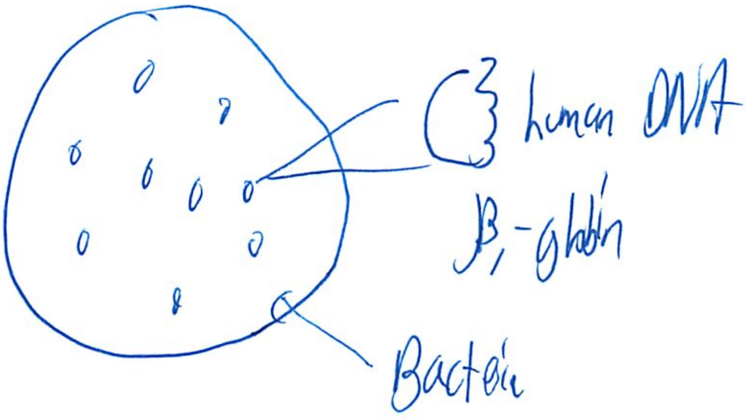
can grow a bunch of them after transforming them

but what if we don't what gene for beta globin
can't do by function

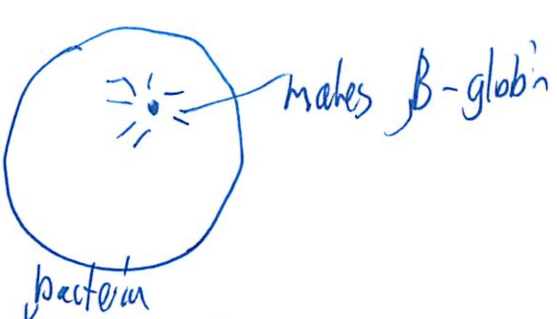
humans can make antibodies that recognize
certain protiens

①

Suppose make a library
w/ lots of cells
each w/ a piece of human DNA
- chopped up, but not too much


human DNA
β-globin
Bacteria

if cell producing β-globin the anti body will recognize it!


makes β-globin

But is bacteria cell carrying gene making the human protein?
bacteria would not recognize promoter of β-globin

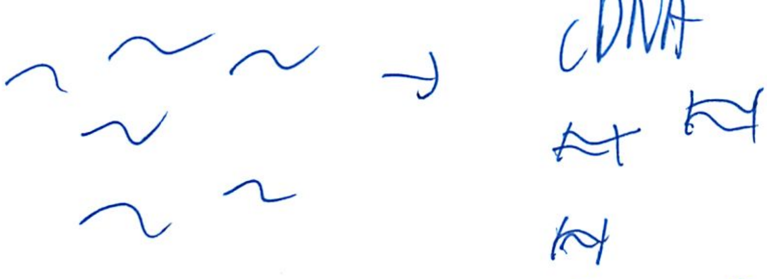
but e-coli might

plus has exons + introns

L bacteria would not know how

¿so make DNA w/ mRNA?

Human RNA   Reverse Transcription



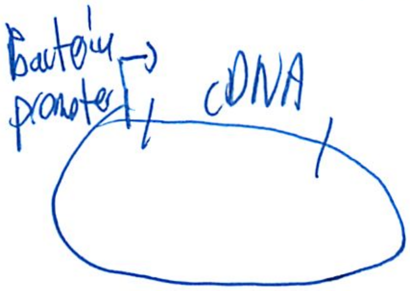
↑ lots of copied DNA pre-spliced

make cDNA library

but promoter still a problem

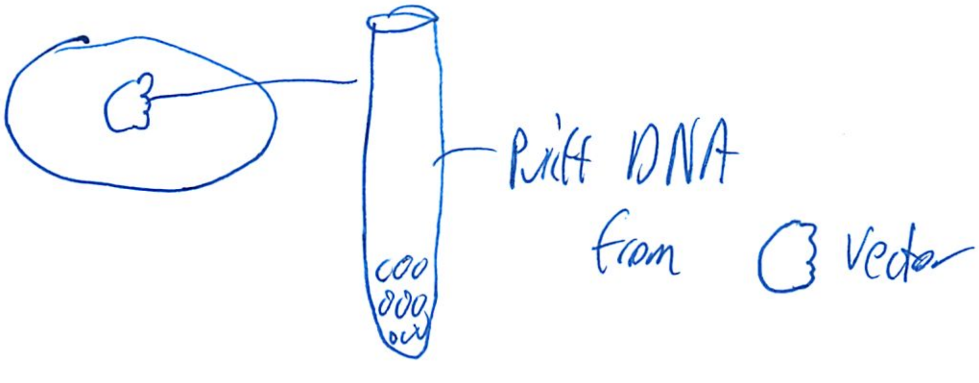
⑨

Can add a bacteria DNA promoter


Bacteria promoter
cDNA

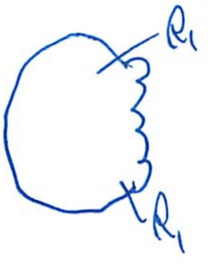
This is the bio tech industry!

Analyze Your Clone


Purify DNA
from vector

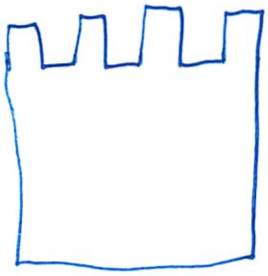
how to see how big insert is

1. Measure size of insert



⑩

to measure size: jello /agerose



Complex, polimer network

add gene

turn on electic field

└ since DNA ⊖ charge

DNA will go ~~[illegible]~~ to ⊕

DNA stats wiggling to gello

smaller pieces move faster

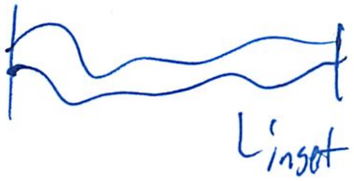
~~[illegible]~~ dye the DNA

Molecular weight standard to tell how big they are

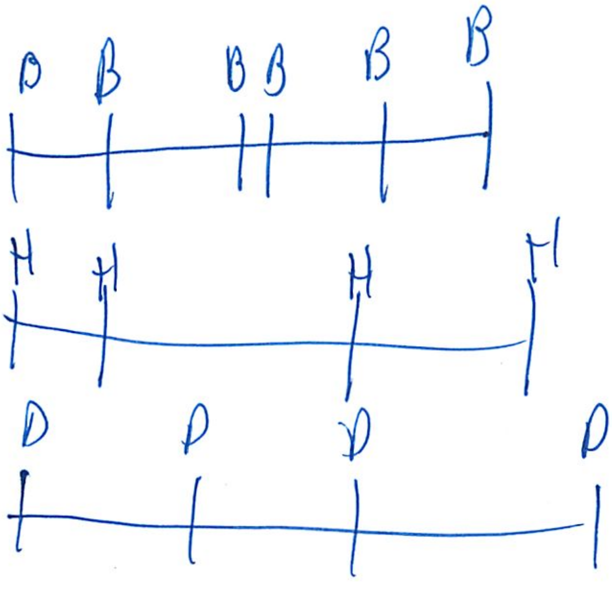
└ so get those from cate log

Can see 2-3 diff frag sites
if multiple genes that can rescue

2, What is DNA seq



have test tube w/ just that fragment
now how to read the seq of DNA
restriction enzymes cut it up

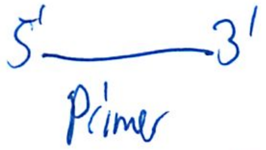

B
B
B
B
B
B
H
H
H
H
D
D
D
D

then do a bunch of math
but not always unique sol
and not always precise enough

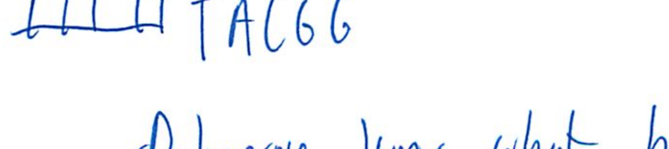
Sequence DNA

have single strand of DNA ~~that~~
want its seq

best way: using tools of replication



Add polymerase -


3'
5'
ATGCC
TACGG

Polymerase knows what bases go in

The Trick (Nobel Prize)

Suppose gave DNA polymerase no G
  Could not go further at G

Or a defective G
  Could put it in G - but not extend it

So we know where last G is by how far
  it got before ~~[illegible]~~ we start

So put in some good + bad Gs
  So it picks at random

Then it stops at diff place each time

Then measure lenght of each

Then do trick for A, T, C

Bio Recitation R10

10/18

(35 min late)

(totally not figureing it out - since missed qu)

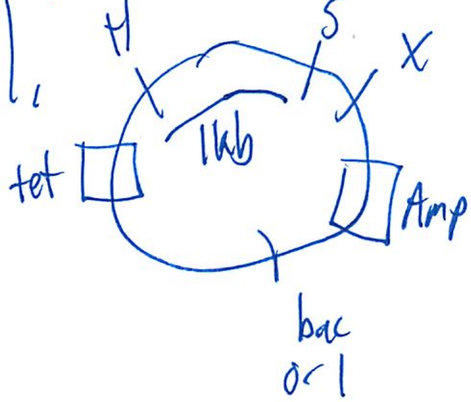
Where inserting the gene

2 diff orientations

practice them

TA: Best thing for Recombinant is just do problems

Rec 10

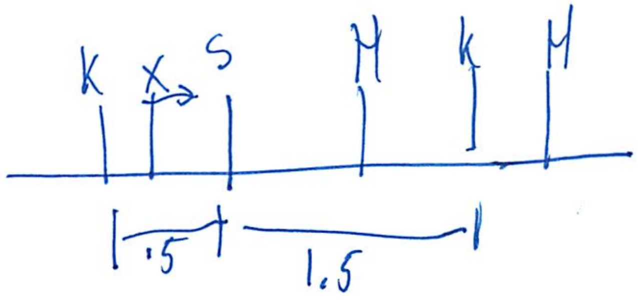


a) ~~the~~ origin

markers

restriction enzyme

②



Xho + kpn

kpn

c) Growth medium

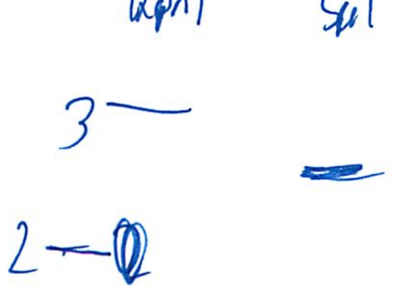
On the one that contains Ampisilin

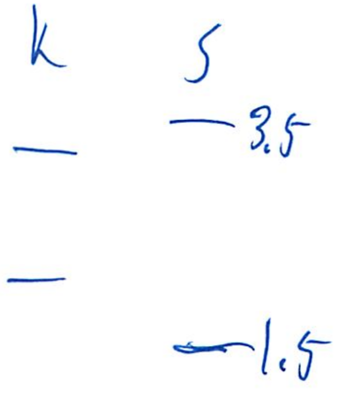
At kpn - lose something ? don't know if recombinant or not

d) kpn + kpn I

(not following closly)

e) Cutting at kpn I gives 2 bands


kpnI
SalI
3
2



2012 7.012 Recitation 10

Summary of Lectures 15 & 16:

Cloning by complementation: This is a way to identify the gene that is mutated in a mutant you have isolated. For example, you might start with a mutant yeast that is an Arg mutant (i.e. can't make its own arginine). You allow this yeast to grow and divide into 6000 identical mutant yeasts, none of which can grow on medium lacking arginine. You transform these yeast with a wild-type yeast genomic library, which is a collection of 6000 plasmids, each one of which contains the wild-type version of a different one of the 6000 genes in yeast. You do the transformation such that each cell receives one plasmid. Only one of the cells will receive the plasmid that had the wild-type version of the gene that was mutated. Only this one cell will now be able to grow on plates lacking arginine.

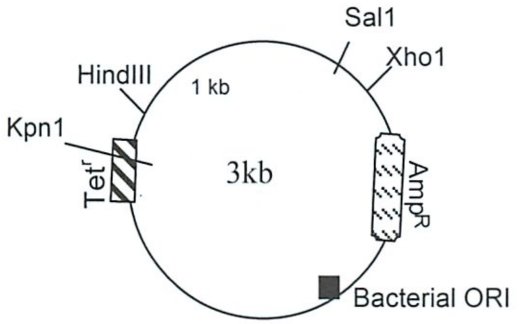
Recombinant DNA technology Application: This technology allows biologists to change the genome of any organism (microbe, plant, or animal) in any way. Such technology allows us to add any gene to a genome or delete any gene from a genome. For instance, we can add the gene for human insulin to a bacterium, such that it will produce insulin for us that we can give to diabetic patients. We can remove the gene encoding Epo from mice to make a mouse that is anemic because it cannot make enough red blood cells. This genetically engineered mouse can then be used as a model for human anemia that can be studied in the lab and can be used to test new potential therapies. Adding a gene to the genome of bacteria or yeast involves putting it on a plasmid. Adding a gene to the genome of mice involves injecting the transgene DNA into a fertilized egg; the transgene DNA will then insert into the genome at a random location. Removing a gene from a genome involves targeting that gene with a piece of engineered DNA that will recombine with the normal gene and replace that gene with a non-functional version of the gene.

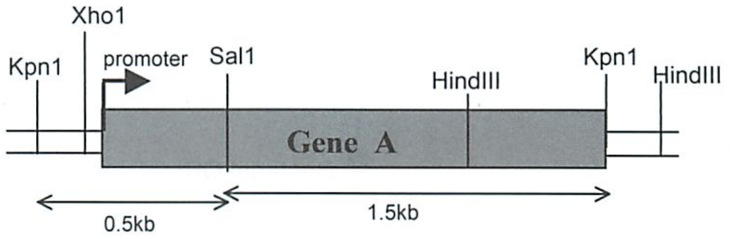
Questions:

1) You purify a protein from a plant cell that can act as a potential appetite suppressant. Owing to its possible commercial application you decide to clone the gene, Gene A, that encodes this protein. You isolate this gene from the plant cell, clone it into a plasmid vector and amplify it in the bacterial cells.

a) List three features that a plasmid must have to allow the cloning and amplification of Gene A in bacterial cells.

b) You decide to use the following plasmid to clone Gene A. To achieve this you digest both the genomic DNA and plasmid DNA using a restriction enzyme. You then ligate the Gene A DNA into the cut plasmids. Finally, you transform the E. coli bacterial cells with the ligation mix (the recombinant plasmids). Note: The recognition sites for Kpn 1 and Sal1 on plasmid are 1 kb apart.



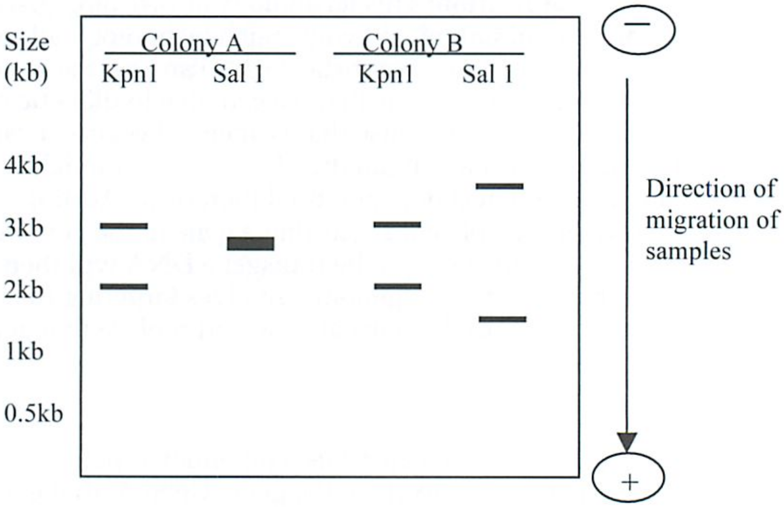


- Which restriction enzyme (*Kpn* I, *Hind* III, *Sal* I or *Xho* I) did you use to digest Gene A for insertion in to the plasmid?

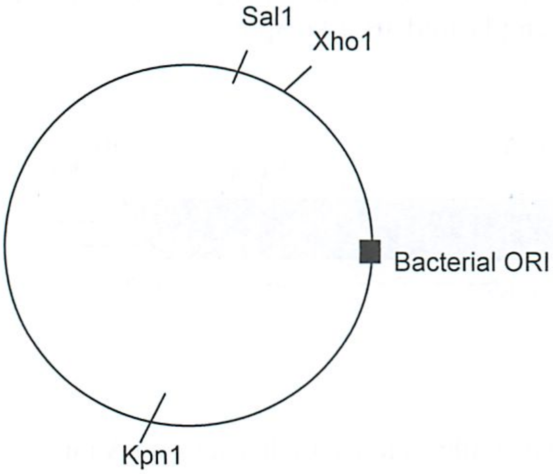
- Which restriction enzyme (*Kpn* I, *Hind* III, *Sal* I or *Xho* I) did you use to digest the plasmid before insertion of Gene A? Briefly explain why.

c) You then plate these transformed bacterial cells onto media that will allow you to distinguish between bacterial cells that obtained the plasmid and those that did not. Onto what type of growth medium will you plate your transformation mix? Explain your answer.

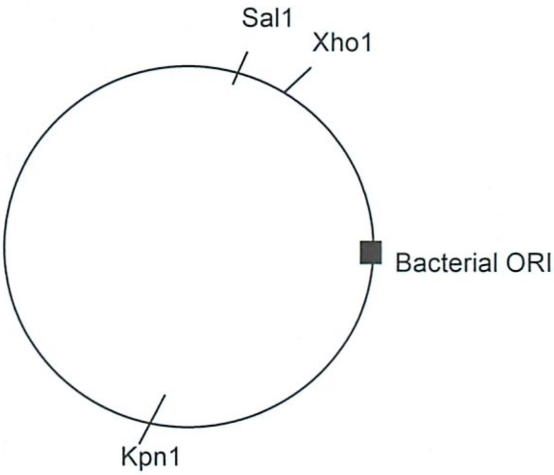
After plating the transformation mix onto selective media, you find several different colonies. You isolate the recombinant plasmids from two colonies, perform a restriction digestion using either Kpn 1 or Sal 1, resolve the digested fragments by gel electrophoresis, and obtain the profile as shown in the following schematic.


Size
(kb)
Colony A
Kpn1
Sal 1
Colony B
Kpn1
Sal 1
4kb
3kb
2kb
1kb
0.5kb
−
+
Direction of migration of samples

d) Complete a sketch of each recombinant plasmid below. Be sure to include Gene A, the Gene A promoter, and all the *Kpn* I, *Hind* III, *Sal* I and *Xho* I restriction enzyme sites found within the recombinant plasmid.


Sal1
Xho1
Bacterial ORI
Kpn1

Recombinant plasmid
from colony A


Sal1
Xho1
Bacterial ORI
Kpn1

Recombinant plasmid
from colony B

2) You are interested in what makes a certain bacterial species that lives in the deep sea produce light. You mutagenize cells and perform a genetic screen for mutants that don't produce light. You find two mutant colonies, mut1 and mut2. You transform the mut1 cells with a plasmid library (which is made up of 5000 plasmids, each of which contain one gene from the wild-type bacterial genome), so that every cell gets one plasmid. You screen the transformed cells and find one colony that now can produce light.

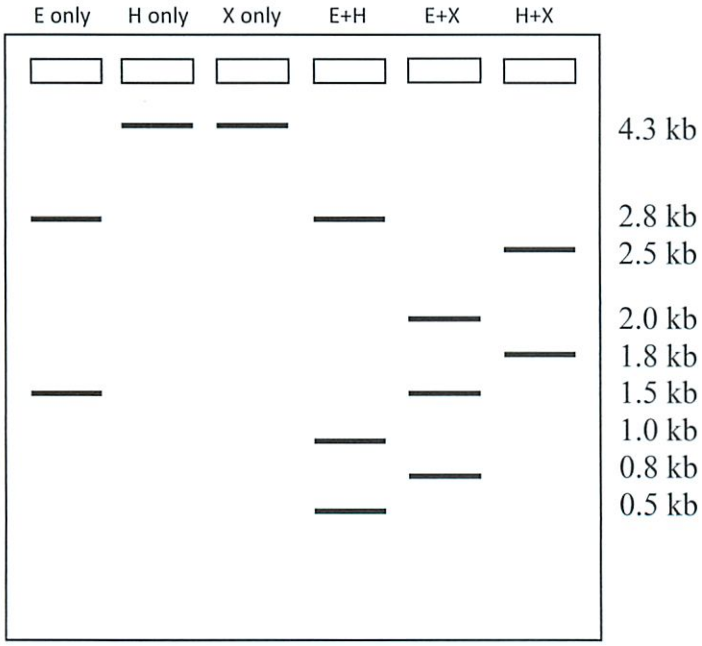
a) Explain why this colony of mut1 cells can produce light.

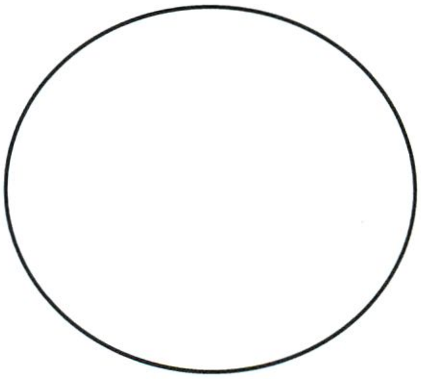
b) How would you identify your gene of interest (which you name "*lyeT*") now that you have this colony?

c) Why couldn't you just determine the sequence the whole genome of mut1 and find the one base-pair that differs between its genome and the wild-type bacterial genome in order to find the mutation?

d) You try to identify the gene mutated in "mut2" by complementation and it doesn't work, but you know that your library contains every gene in the genome. You sequence the *lyeT* gene in mut2 cells and find that it contains a mutation. Propose two possible explanations for your inability to identify that there was a mutation in *lyeT* in the mut2 bacterium using this strategy.

3) Three restriction enzymes have recognition sites in a plasmid: EcoRI ("E"), HindIII ("H"), and XbaI ("X"). You digest the plasmid with each of the following combinations of enzymes and see the following gel.





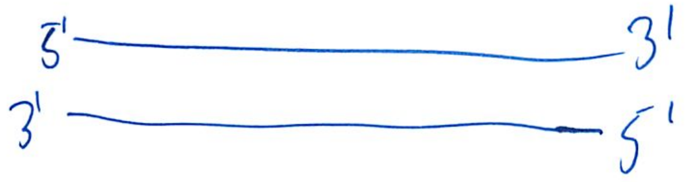
Complete the map of the plasmid indicating where each restriction enzyme cut site is, which restriction enzyme cuts at each site, and how far apart each cut site is.

Recombinant DNA 3

10/19

(5 min late)

DNA Sequencing



Primer

~20 bases

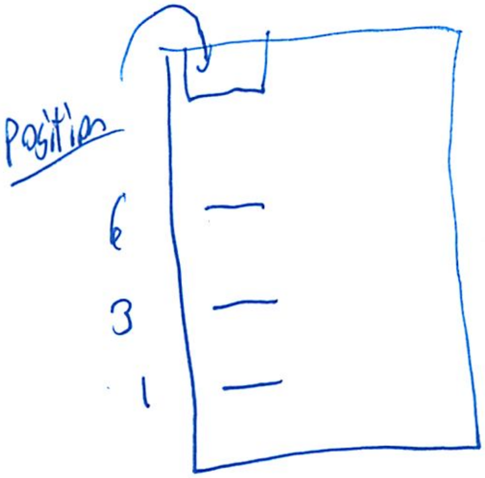
Polymrs

3' TAATGC CGTAATGCC 5'
5' ATTACG ......
GCA* ← defective
GCA TTA GCCA*
↑ good ↑ good

So if 1% A defective
stop 1% of the time

dNTPs

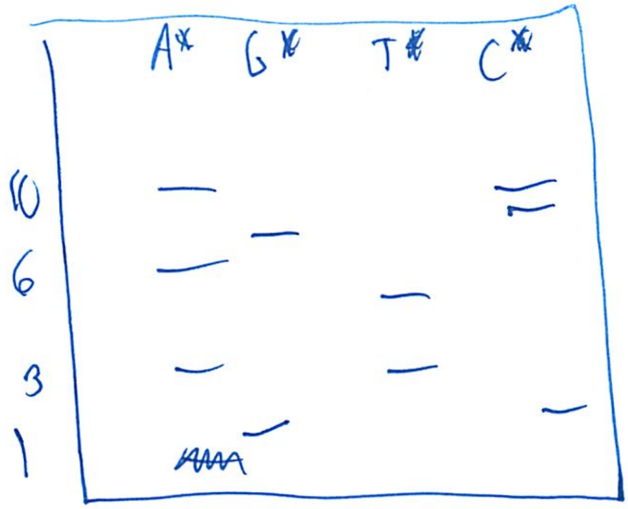
②


Position
6
3
1

gello agerose
measure lenght

W/ Gs
G*
GCATTAG*

So get table



Can tell from table what is where!

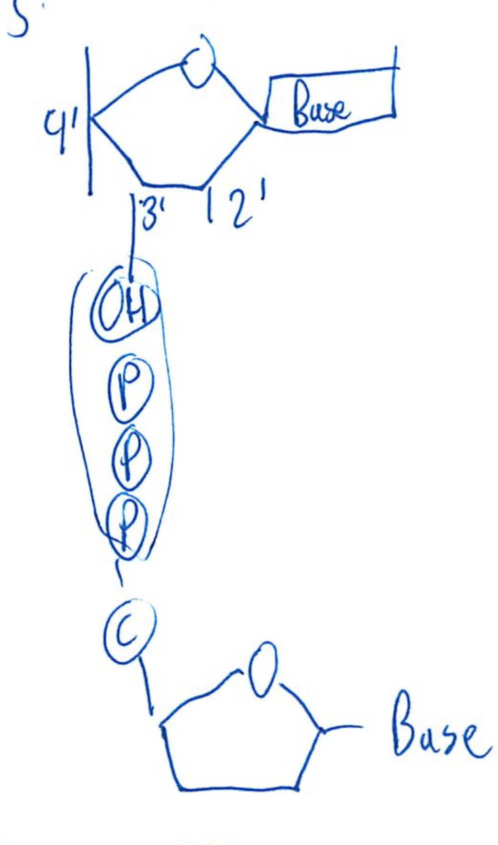
Could radioactive label the primer

Or w/ x-ray film

very long + tedious

Mid -1980s → Better way

What do we mean by deffective A


5'
O
Base
4'
3'
2'
OH
P
P
P
C
O
Base

defective = no hydroxyl

2' deoxy NTP dNTPs

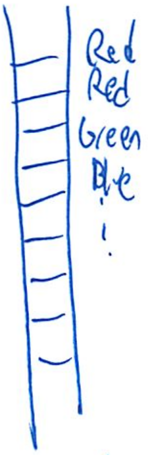
2' 3' dideoxy NTPs

ddNTP

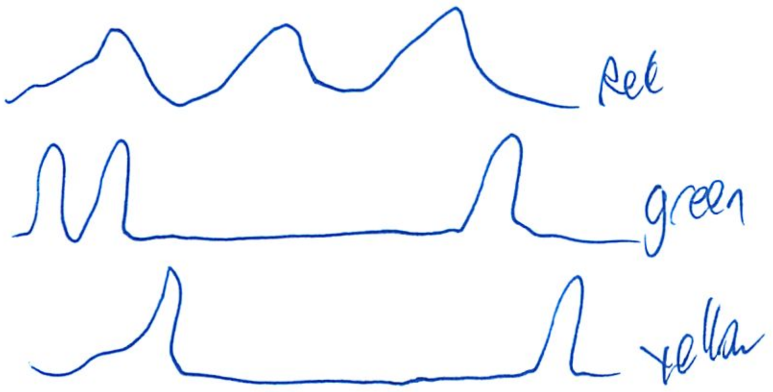
This was painfully slow
max 400-500 letters/week

Or flurecent dye
A* - Blue
T* - Red
C* - Yellow
G* Green

Now can run w/ single lane
Dif Color



Scan w/ laser scanner
As each band migrates past scanner
electricity is on
over time see it


Red
green
yellow

Seperation is log of lenght
So get up to ~700 bases

But can use more than 1 caplary
Up to 96 capilaroiers

1 million in 20 runs/day in 2000

700-1000 bases max one could do at once

What if we had a longer fragment

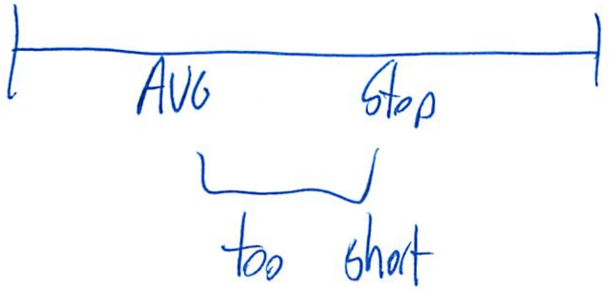
10000 bases

⑥

Could cut up each one
and put in own vector

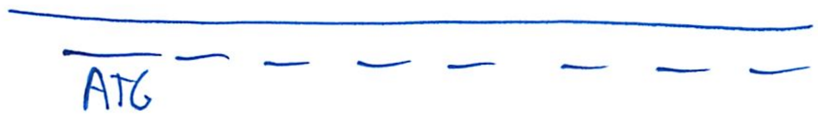
Or sheer up into many small fragments
randomly
make lib of
Sequence those
lots of 200 base pairs
but how to put them back together
look for ones that overlap enough
Long repeating strings screws it up

Computer code to read AUG, Stop



Stop is every ~20 codons

Have the computer look for a long open reading frame (ORF)
lots of codons in row



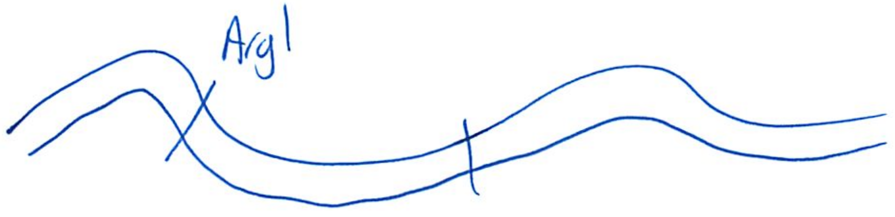
Translate to protien
Compare to DB

life reuses a lot of protiens
So just search the db

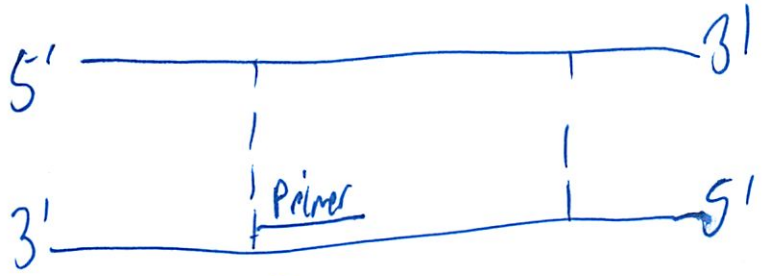
Suppose want to get strand from other places
Can you get same gene from other strand
w/o making libraries

⑧

Finding Your Gene Again



Melt DNA
└ so single strands

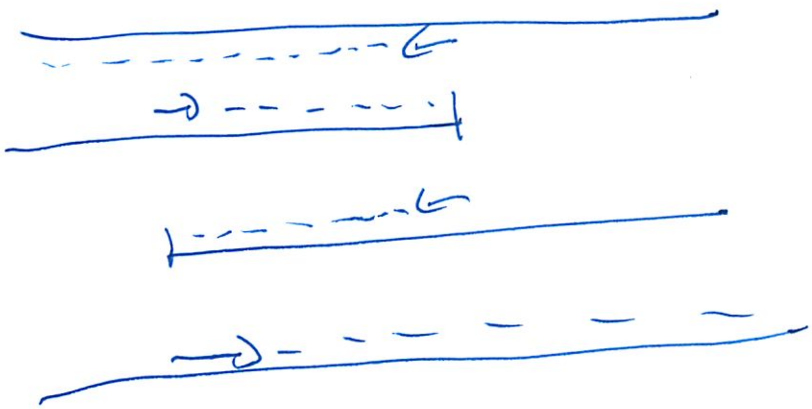


~~[illegible]~~

right primer will bounce around
till it finds the right spot
then it H-bonds

Started w/ 1 ~~[illegible]~~ helix
now have 2

So do it again!
└ iterate!



Now 4

Again → 8
Again → 16
→ 24

30 times ~1 billion

Want primer long enough so only 1 match

Called the polymerase chain reaction (PCR)

Very powerful

If DNA polymerase that can withstand top temp

So look at vent Bacteria → thermus aquaties (Taq)

How get primers?

Machine adds 1 base after another

Order it

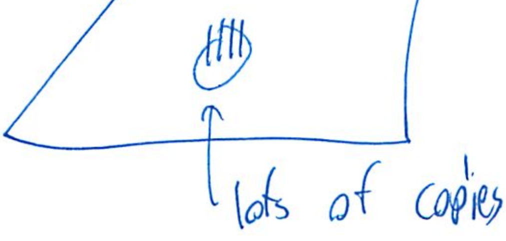
Can order your own primer + polymerase

If had single cell, could do PCR on it

Modern DNA Sequencing

1 million day is prehistoric

Use a glass slide



add DNA polymerase

⑪

Di-deoxoid

Can use camera to see green

But can go futhr

¿If reversable chem blocking goup?

~~T~~ Take Picture, Remove blocker, picture, etc

Can use a bunch of spots

~~2,000,000 sep spots~~

1,000,000,000 sep spots at ~~a~~ a time

~100 letters

= 100,000,000,000 bases / day

100 bill bases / day

So a million times more data for the price

7.012 Studying

10/21

Textbook

Prokaroyt Gene Regulation

all in same space

e coli

needs Tryp

Can make it if doesn exist

Operator specific seq that binds to DNA transciption factors to turn transciption on or off

Promoter region of DNA where RNA polymerase binds to

operon - cluster of genes that are coregulated

repressor protien that binds to operator inhibits RNA polymerase from binding to promoter

allosteric protiens - change shape when bound many of the regulatory protien are ie when lactose present it binds to repressor and prevents repressor from binding to DNA

oh the lac repressor is in DNA
(I guess that makes sense)

repressible operons - repressed when repressor is bound by specific mol (like Tryp example)

inducable operon when repressor is not bound by Specific molecule (the LacI example)

(this seems backwards)

Positive gene regulation CAP interacts w/ cAMP
when CAP-cAMP is bound RNA polymerase has a higher afinity
cAMP levels fall when cell glucose levels increase

(oh the top diagram is here)

③

trp option - presense of trp binds to operator + stops RNA polymerase

lac option - presence of lac binds to repressor and prevents Repressor from attaching

RNA Polymerase sigma factor

without it RNA polymerase starts at wrong site

Something that binds to RNA polymerase

released after transcription starts

Riboswitches

(did we talk about these in class?)

Part of mRNA molecule that bind ligands

Switch translation on + off

Often found in the 5' untranslated region

RNA can easily fold into secondary + tertiary structures

aptamer -

(skipping)

④

Whole-Genome Shotgun Sequencing

1995

Cut into many smaller + overlapping fragments

Seq each fragment

Calc where tey overlap

(not looking at details)

Eukaryote Gene Regulation

Cell-specific differentiation

all human cells have the same DNA inside

Some sort of regulatory mechanism

Usually wrapped around histone protein core

---

2 diff type of chromatin

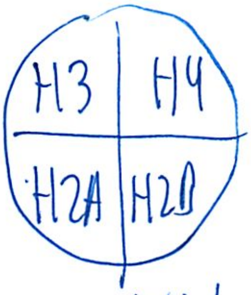
euchromatin - lightly packed

heterochromatin - mostly inactive DNA being inhibited/repressed

⑤

acetylation ↓ net ⊕ charge of histones
↑ transciption factor binding

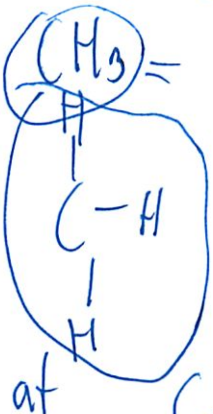
methylation ↑ binding of histones to DNA
↓ DNA avilable for transciption



↑ histone octamers

Methalation preserved during DNA replication

WPi) adding a methyl group

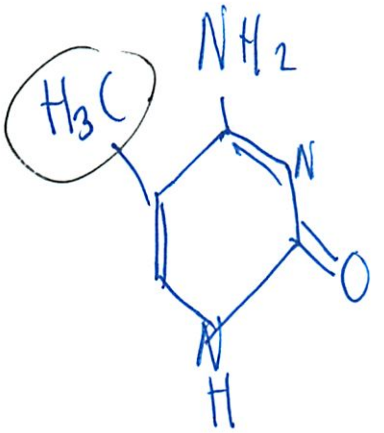


often at CpG sites |CG| ie seq

About 80-90% human CpG sites methylated

5' - Methylcytosine

Methyl grap on 5'



transciption factors - needed to initate trasciption

Some bind to DNA, others to protien

Control elements - specific DNA sequences 4-8 base pairs long

may be grouped as **enhancers**

Can be thousands of nucleotides from the gene

Since it bends over

transciption factories looks lig genes on diff chromosomes

brought together to coordinate transcriptions

during interphase

⑦

(I don't seem to have the long term memory for this class...)

alt. RNA splicing

diff exons/introns

mRNA sticks around longer in eukaryotes

Non Coding DNA

Only 1.5% of Genome codes for specific protiens

Used to think of rest as junk

now see a use for it

microRNA - 22 nucleotides long

(we didn't talk about this either in class)

⑧

Cell Differentation

Change from unspecified cell to specific function

Once cell determined → can't reverse

Sent signals

Cytoplasmic determinants inc proteins, RNAs, + organells

extrinsic signals from outside the cell

Then induction occurs

Special membranes receptors that detect substances

Then promotes or blocks gene transcription
in the target cell

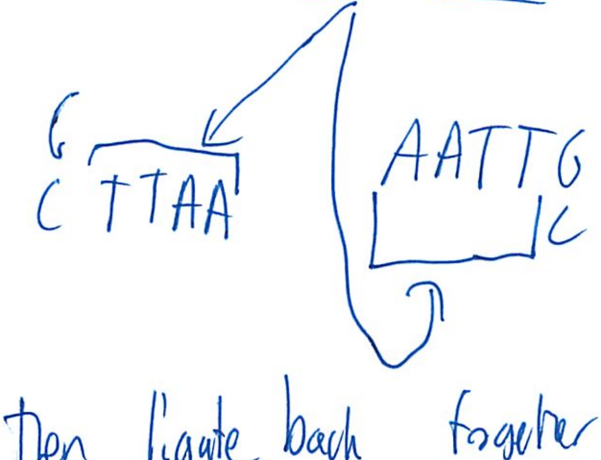
(this section of year does not seem well represented
in the book)

DNA Cloning

~~Small~~ plasmids small circular pieces of DNA
cut open
insert DNA seq
Can reproduce a bunch of times then
ie recombinant DNA

restiction enzymes - cut DNA
at certain restriction sites

Cut straight through s blunt end
but usually → sticky ends



Then ligate back together

(U)

has antibodintic restiance gene for drug resistance selection

? to find bacteria carrying plasmid w/ gene inserted

Can color

So inserting gene breaks coloring

then visually scan for the white ones

Make libraries to study genes of org enmasse

L of bacteria artifical chromosomes

* May consist of DNA pieces that make up
a complete genome

cDNA lib make DNA from mRNA w/
reverse transciptase

note it only includes the important stuff

L ie no introns

⑪

Use flureocent nucleic acid probes
- flood w/ solution
- specifically designed to stick to a target sequence

Expressing cloned genes

requires inserting the gene into an expression vector

contains a bacterial promoter upstream of the insert

PCR Polymerase Chain Reaction

Generates multiple copies of gene seq in large quantaties

1. denature - split double standed DNA w/ heat
2. annealing - primers bind to DNA, giving starting pt
3. extension - Taq polymerase extends seq

Can repeat to get 2^n # of double-stranded DNA sequences

⑫

Gel Eletrophoresis

Seperates molecules based on size

Large fragments towards top

Small " " bottom

UV light show

but supercoiling screws it up

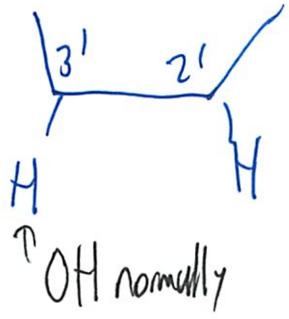
do it after PCR

Use a housekeeping gene to show same amt in each lane

di-deoxynucleotide

Once included in nucleotide polymer

Can't add any more nucleotides



terminates chain

Color each diode - differently

have a bunch, each 1 longer than before

Then measure lenght

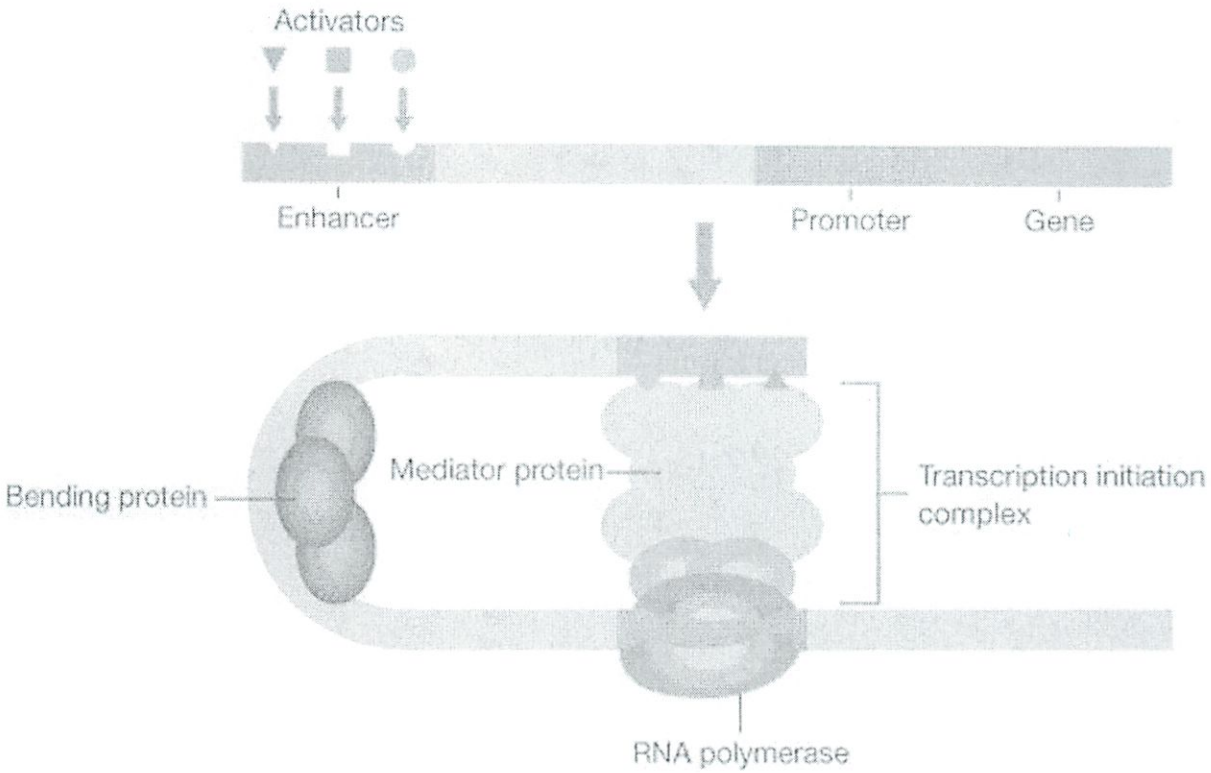
↳ ¿using the gel method?

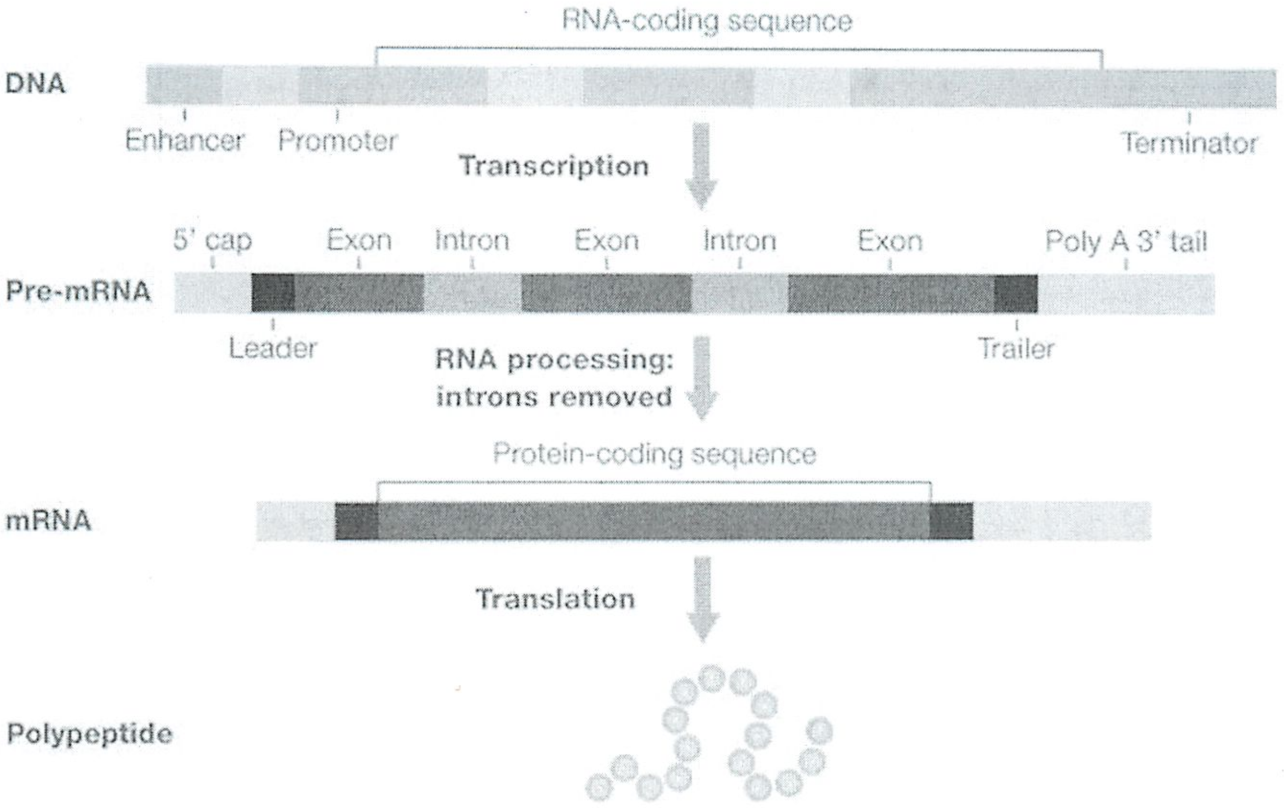
Or now flurocent detector

functions on the spot!

microarrays lots of species-specific DNA sequences

heat map


Activators
Enhancer
Promoter
Gene
Bending protein
Mediator protein
Transcription initiation complex
RNA polymerase

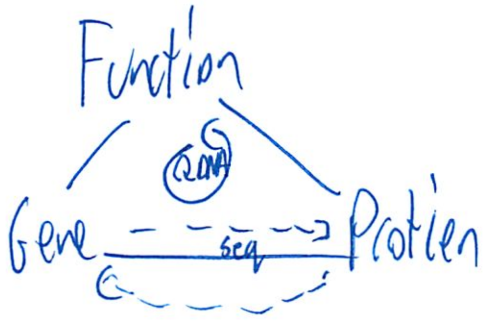

RNA-coding sequence
DNA
Enhancer
Promoter
Terminator
Transcription
5' cap
Exon
Intron
Exon
Intron
Exon
Poly A 3' tail
Pre-mRNA
Leader
Trailer
RNA processing: introns removed
Protein-coding sequence
mRNA
Translation
Polypeptide

Lecture

Recombinant DNA 4

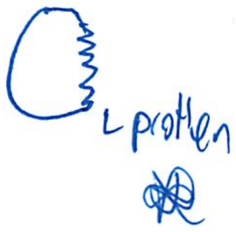
10/22

Recombinant DNA makes this operational
- how to clone genes
- how to infer protien


Function
cDNA
Gene
seq
Protien

Clone by expression

cDNA = copied RNA



Clone by Complementation
Rescue a mutant
(what was this again)
put gene back in lib
See what grows
hard to do in humans for ethical reasons

How do we find the function?

Find a mutant

But how to make a mutant?

Human Genome Project

Goal: sequence Human Genome

Proposed 1985

Launched 1990

Rough Draft 2001

Finished 2003

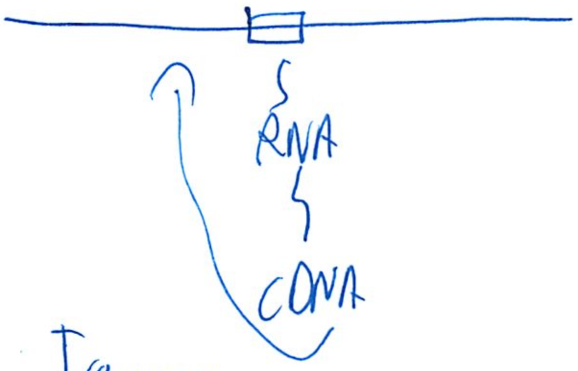
Similar to what we talked about before

If long enough word size (100 base pairs)
+ no repeated sequences
Could put together 3 billion pairs

But there are all these repetive sequences

③

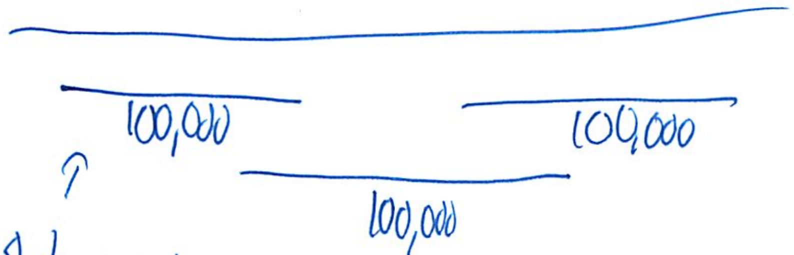
Same/Similar seq many times!



Transposons
Retroviruses
Slam back into genome

Are 50% of human's DNA!
Parasites - There for them
Been around for a billion years
This made HGP hard

So Hierachical mapping


100,000
100,000
100,000

shotgun each piece

? large + smaller bits of overlap

④

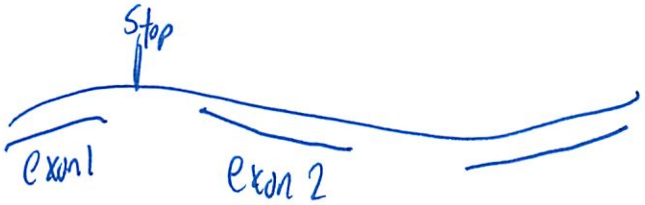
Annotate where the genes are

Could look ATG → stop

to find an open read frame

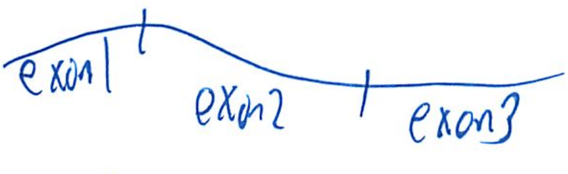
But not very efficent when 3 billion long

Hard when exons


stop
exon1
exon 2

stop does not matter since in intron

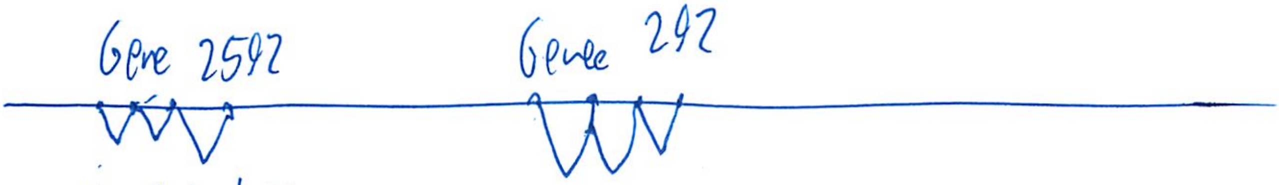
RNA cDNA library


exon1
exon2
exon3

So w/ cDNA library

Can see where exons came from

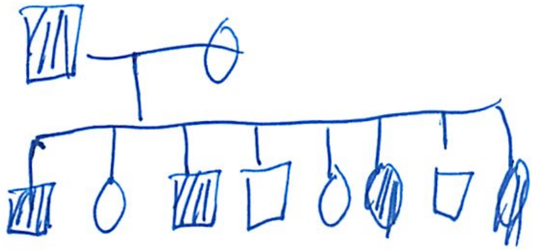
This is most effective way to spot



Huntington's Disease

▨ Dominate disease

Shows up ~50



So where is gene for Huntington's disease

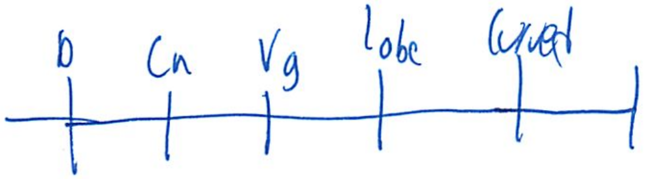
Problem: lots of genetic differences $\frac{1}{1000}$
b/w people

b/w siblings $\frac{1}{2000}$

Today: compare lots of people
but getting ahead

Or compare children w/ and w/o
Looking for linkage
Sternanvat made a map

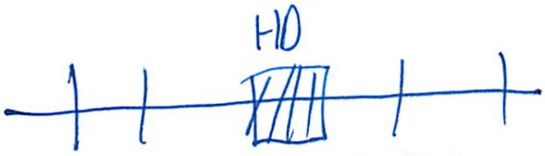
⑥



Set up a Cross

See which allels are inherited correlatedly

1984 Huntington Disease



↑ we don't know this yet

Set up crosses
  └ but can't do in humans

and need other genetic markers
  — Can't to wing shape / eyes

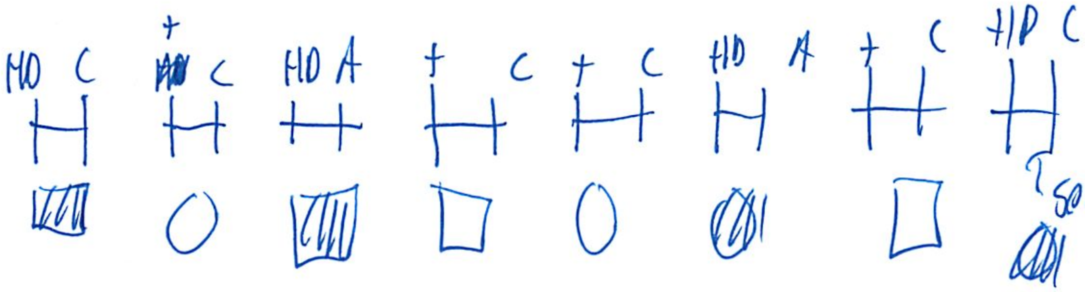
- gender
- eye color
- hair color
- height
  └ 180 genes

) but too many genes

⑦

Remember every $\frac{1}{1000}$ base pairs different $\underbrace{\text{b/w identical twins}}$

DNA spelling differences are just like DNA markers
like Cinabor eyes
└ codominant genetic markers



↑ some recombination

look at all of them →

↑ Corelated ~ certain genetic markers and letters

↑ linkage mapping of diseases

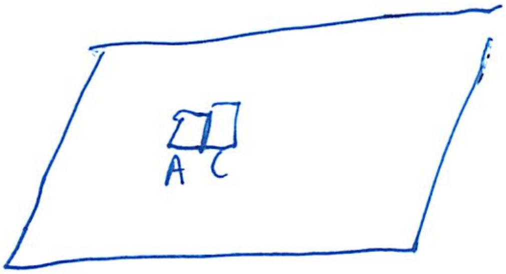
but many things is not a simple gene
We are just over simplifying / picking an easy one

1980 ~6 diseases
2012 3000 diseases

Cool tricks to follow spelling differences
└ can make an array

⑧

Chip w/ detectors for A, C



laser scan to look for differes
w/ florcent marker

Soped up Sturevant method

Gene → Function

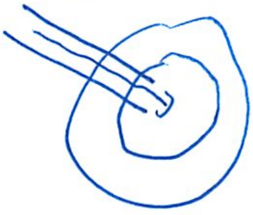
L encodes interesting protien
what potien does it encode

Could knock it out w/ mutase
but need to mutate a bunch of mice
hope it mutates our gene

How can we target our gene?
How do we do that?

w/ fertilized egg

~~ae~~
step 1 getting ~~DNA~~ DNA into egg cell
needle into egg
Squirt into DNA



Cells are nice enough to ligate it
to a Chromosome

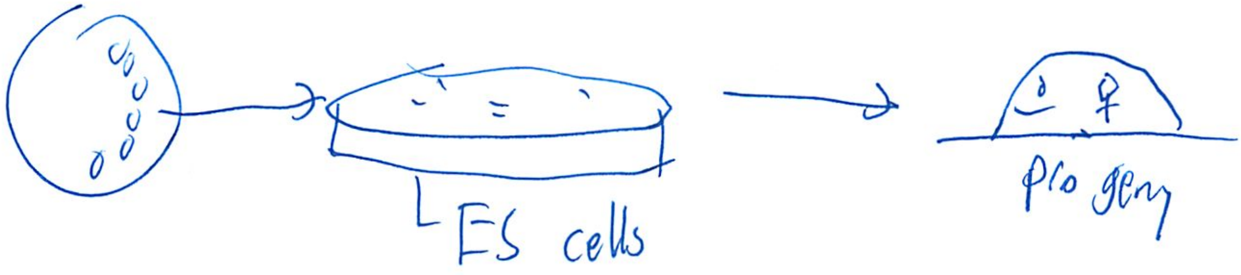
DNA addition → to make transgenic cell
but ligated at random
↳ useful if just want to add
Can't really replace

⑩

Must signal to target gene

Inexact → need a bunch of cells

Better → Embryonic Stem Cells



made modifications
- while outside

put back in mouse

Step 2 Target/knock out specific gene

Usually goes somewhere
but recombination mechanism is active
homologeous recombination

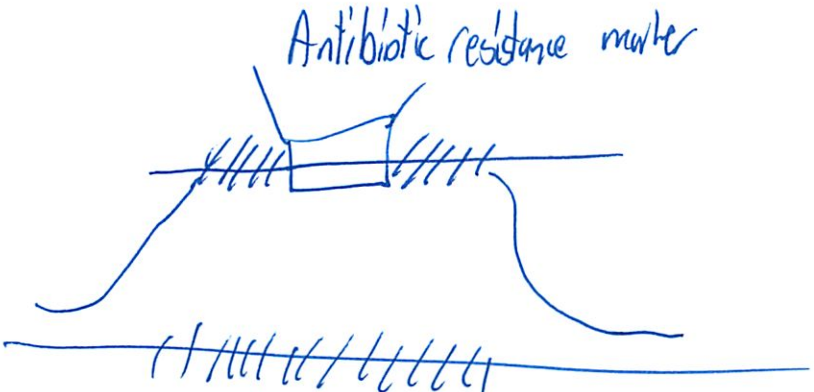
Sometimes genetic recombinatation
occurs at exact right spot

⑪

Small prob $\frac{1}{1000}$ in right location

Need to find it (screen)
bt that is slow

Instead want to select



Only cells w/ marker will grow

Instead add negitive selection marker nearby
Any cell w/ this dies

Slammed in usually to both markers

Or. just $\boxed{+}$ marker not $\boxed{-}$ marker

(don't get his diagram)

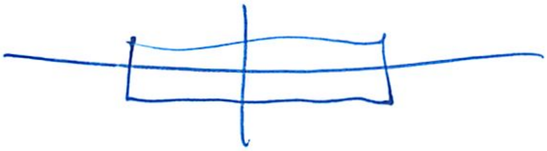
Can treat like bacteria

Since have large population for

How people have been doing for last 15 years

New targeting method (not in book)

Suppose just restriction enzyme



? can't get unique

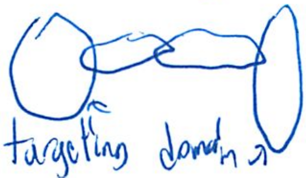
usually 6, 8 base pairs

What if only cuts at 1 place

need to target to 1 spot

restriction enzymes → dymer dimer?

attach a targeting domain



⑬

look at how restiction enzymes do this in ecoli

↳ no logic

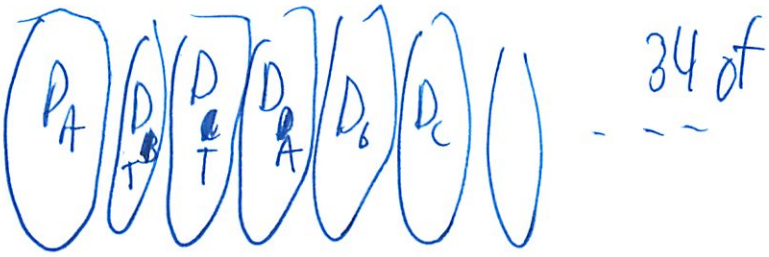
but last few years → plant pathogen that targets seq

has modular repeats

has targeting domains

34 amino acids

that differ only slightly



Give those

and it looks for it

So binds that seq

w/ restiction enzyme it cuts that seq

= Tal affectors

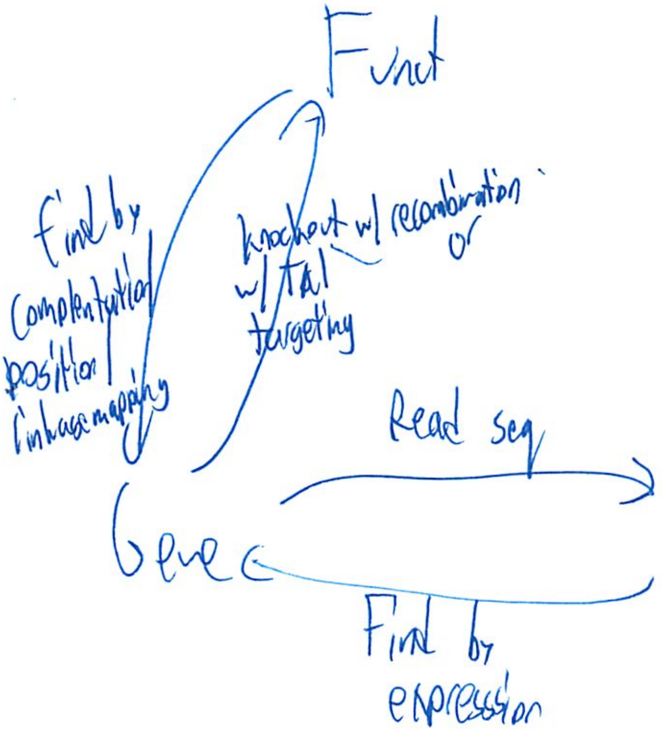
---

need 2 [illegible] folds

Can also attach an activator

Could target turn-on

Big picture


Funct
find by
complentation
position
linkage mapping
knockout w/ recombination
or
w/ TAL
targeting
Read seq
Gene
Protien
Find by
expression

now fully operational

7012 Bio
Recitation

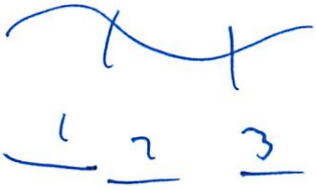
10/23

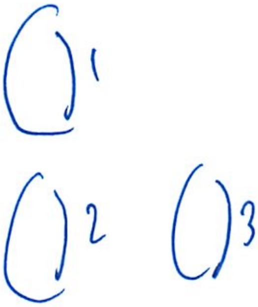
(3 min late)

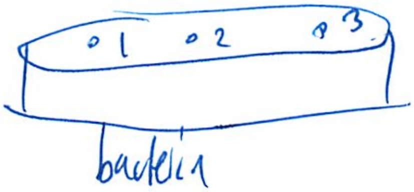
Libraries

Genomic

Yeast DNA





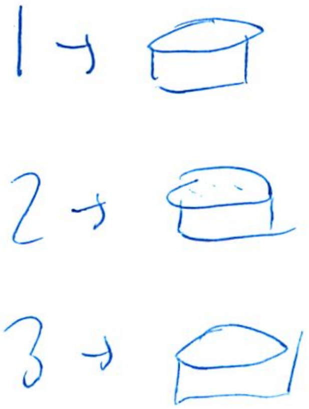


Cloning by complementation

mutant $\xrightarrow{\text{rescued by}}$ wild type

gene from WT lib

yeast Tyr⁻ –



minimal media

minimal media

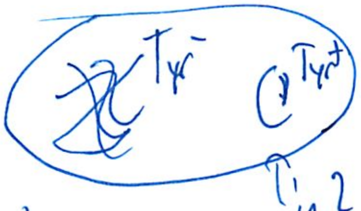
restore function w/ wild type copy

Shows Tyr is (2)

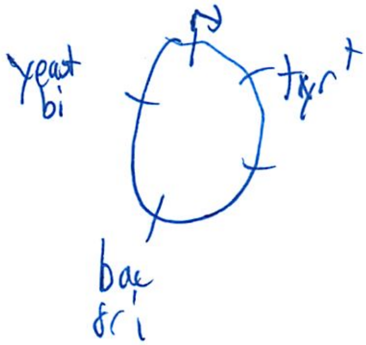
Since able to rescue it w/ 2

②

x all the same mutant


Tyr⁻
Tyr⁺
Yeast
in 2 add
wild type
Gene 2

Grow plasmid in Bacteria


yeast ori
tyr⁺
bac ori

This is how we make a lib and how we use it

plasmid 2 → bacteria
if no rescue

1. Different tyr syntesis gene/pathway
2. No bacteria promoter on plasmid
3. Splicing might be needed
   └ bacteria can't do it

③

If rescue was needed → cDNA

$$\text{mRNA} \xrightarrow[\text{transcriptase}]{\text{reverse}} \text{cDNA}$$

No promoter (problem)

Do same thing as rest of experiment

Diff cells express diff genes

liver cell → mutant skin cell

↓

mRNA

↓

cDNA

only get genes being expressed at that time

Where is mutation? - needed in all cells

Or specific for use

Libraries: Genomic and cDNA

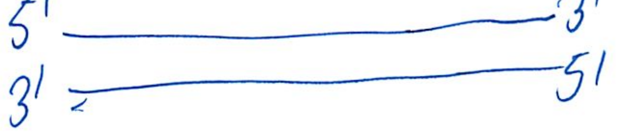
PCR

Goal: Amplify DNA

Need: DNA polymerase
Template
DNA Primers
dNTPs ← nucleotides

Reaction that happens in cycles

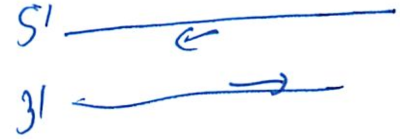
1. Denature
   90-100°C
   depends on what it is exactly



Opens up
H bonds broken

2. Annealing
   50 to 60°
   cool enough so H bonding again

Primers will base pair match



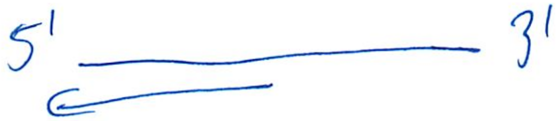
3. Extension

70-75°C

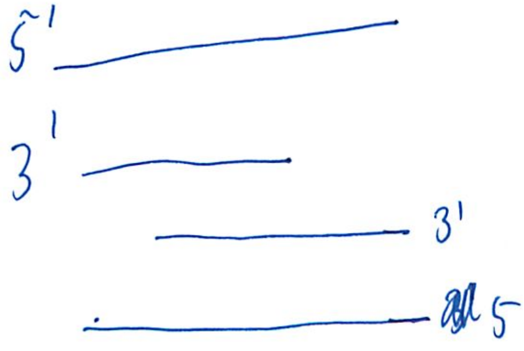
ideal temp for DNA polymerase

taq polymerase - heat stable

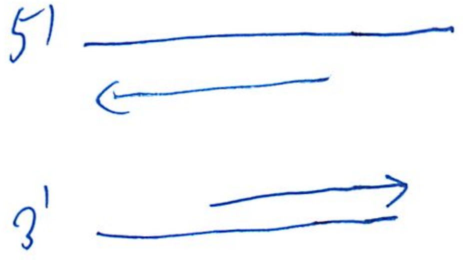
since from hot springs

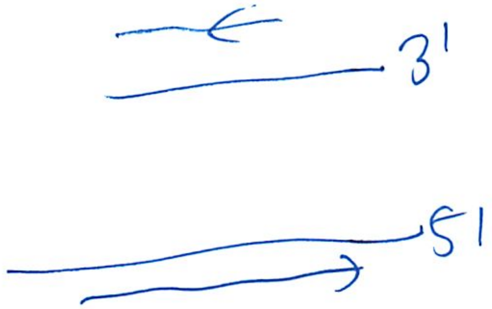


Now have several pieces



4. Will Anneal + extend again





Overall 2^n ↑ # of cycles

Primers don't attach at end
So can only go so far

Only interested in shorter region

Short pieces grow exponential
large linear

①

large pieces might not be complete

Sequencing

if missed lecture → should watch it

this is classic/traditional way

Sanger Method

Same as PCR reaction

Have gene know partial seq of

3' ATCCCC GCC GTAC TAT

5' TAGG

Can guess from prob

don't know

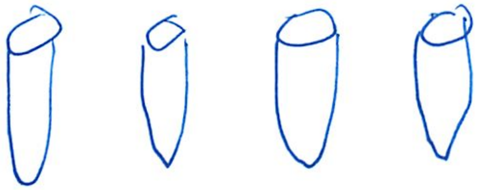
Can add nucleotides to 3' end

Put [illegible] dNTPs - normal

ddNTPs - missing 3' OH

⑧

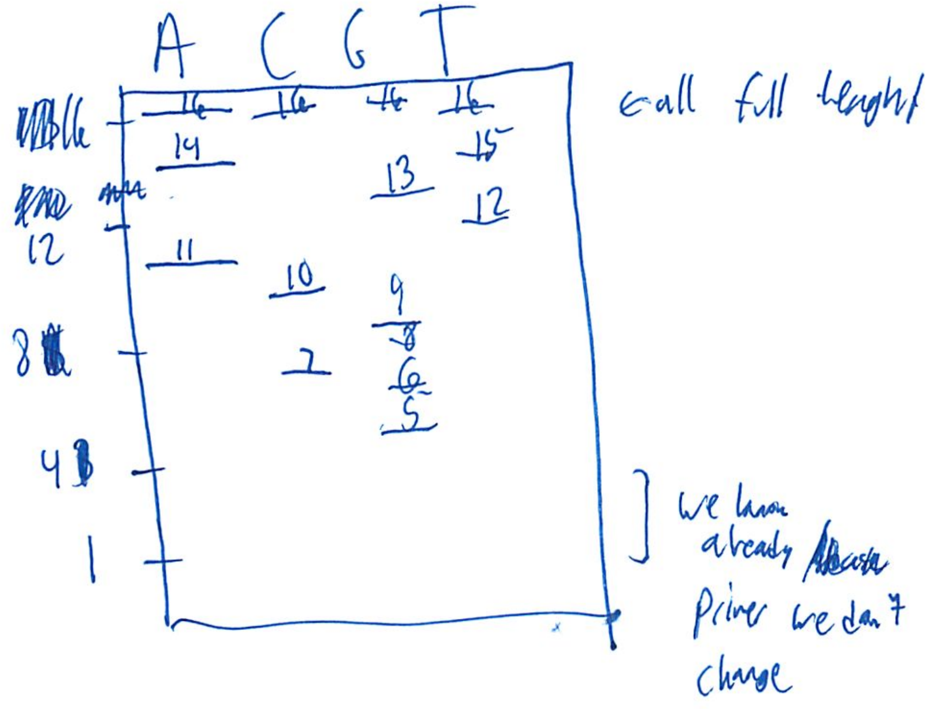
Put in 4 tubes



Everything in each tube
but diff ddNTPs
ddATP ddCTP ddGTP ddTTP

ddATP tube → all the other bases present
but may get a A* (dd) insted of A
Perhaps 1% A*
So prob of making 3 diff type of fragments
↳ based on example
any possible # = # of A

Then sort by size w/ Gel electrophisis


A
C
G
T
16
12
8
4
1
16
16
16
16
14
15
13
12
11
10
9
8
7
6
5

call full lenght

we know already primer we don't change

Should be linear spaced

So then count it up

5' GGCGGCATGAT

know original is complement

So either draw Gel or read Gell

Go bottom ~~and~~ up 5' -> 3'

Usually write seq from gel

└ unless ask for complement

Next Snip arrays

⑩

Recitation 11 Handout

#1 a) 3

b) 1

c) 2

d) 1 is ATP

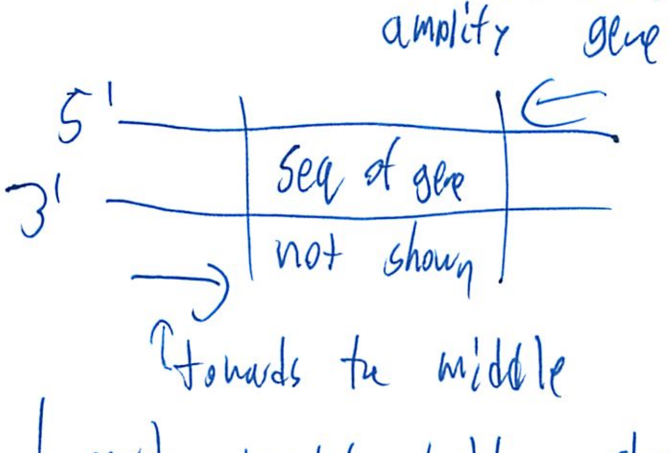
2 is dATP — not a source of energy

#2 Note primers written 5' → 3'

Can only extend 3'

So might need to rotate before extending

Must extend in direction of gene


amplify gene
5'
3'
seq of gene
not shown
towards the middle

Primer 1 must match bottom strand - same as top strand

⑪

Nothing works for 1

Can only choose in sets

So primer 2 must come along

b) doesn't work

c) does work

must flip primer 2

Methylation

Histone protiens

DNA + histone

Can add methyl groups

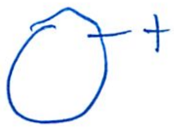
↓ transciptions

↓ expression

Acytlate

Add ⊖ group

histone



Since DNA ⊖ charge will repel ⊖ charge of acytlate ion

expression = transciption + translation

2012 7.012 Recitation 11

Summary of Lectures 17 & 18:

PCR: This is the technique of Polymerase Chain Reaction, which is used to make lots of copies of a chosen piece of DNA. PCR is simply the process of doing about 30 rounds of DNA replication in a row. This technique requires you to combine in a test tube: the original DNA, nucleotides, thermostable DNA polymerase, and primers. Then, one puts the reaction in a machine that performs 30 cycles in a row of three temperatures. The first temperature denatures the double-stranded DNA molecule, the second allows the primers to base-pair with the template, and the third allows the DNA polymerase to generate the newly synthesized strand 5'→ 3' starting from the 3'OH of the primer.

DNA sequencing: DNA sequencing is the technique by which the nucleotide sequence of DNA is determined. DNA sequencing is simply the process of doing DNA replication in the presence of all four normal nucleotides, and a small percentage of some "bad" nucleotides that do not allow for replication to proceed any farther. This technique simply requires you to combine in a test tube: the original template DNA, the four good nucleotides, DNA polymerase, a primer, and then a small concentration of "bad A" (colored red), "bad T" (colored blue), "bad C" (colored yellow), and "bad G" (colored green). Most of the time, the good nucleotides will be added and the template will be extended normally. Occasionally, a bad C (for example) will be put in across from a G in the template. This will stop replication and label the fragment yellow. You allow the reaction to proceed and then load all of the replication products onto a gel that separates them by size. The sequence can then be read from the result of that gel.

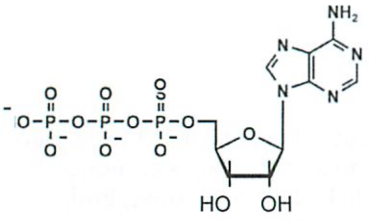
The way that genomes are sequenced is that many copies of the genome are randomly broken into pieces, and then those pieces are sequenced by sequencing machines. A computer then analyzes those sequences and reassembles them into one long continuous piece (for a bacterium with a single circular chromosome) or several linear pieces (for a eukaryote with multiple chromosomes). This technique is called shotgun sequencing. Once the genome is sequenced, one can predict where the genes are in the genome by looking for the signatures of genes, like start codons and stop codons, long open reading frames, promoters, and splice sites (if applicable). Each gene in the genome can then be compared to all other sequences that have ever been determined using the program BLAST. If a new gene shows a large amount of homology to a previously studied gene, the most likely possibility is that the new gene encodes a protein with a similar function to the protein encoded by the previously studied gene.

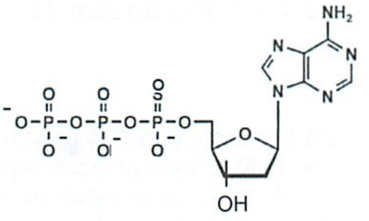
SNPs: A single nucleotide polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide in the genome differs between members of a species or between paired chromosomes in an individual. SNPs are found throughout the genome. A SNP can result in different alleles of a gene, where one of those alleles may be associated with a disease or trait. SNPs are also found in non-coding regions of genes or intergenic regions between genes. SNP mapping is used in forensics and a genetic tool / marker for a trait or disease.

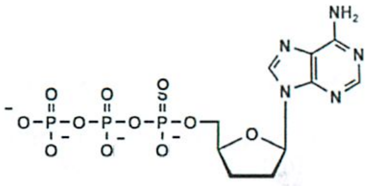
Microarrays: Not all cancers respond equally well to all treatments, so knowing the specific type or subtype of cancer is important to successful treatment. Some cancers can be identified by histological tools, but other are best characterized by determining the gene expression profile. A DNA microarray is a multiplex technology that allows comparison of cells based upon the expression of many different genes. A DNA microarray consists of an arrayed series of thousands of microscopic spots of DNA, each spot representing a gene. When a DNA microarray is probed with mRNA isolated from cells, individual mRNA molecules will hybridize to the appropriate DNA spot. If a gene is highly expressed, more mRNA will be made from that gene, so more mRNA will hybridize to the corresponding DNA spot, and the signal from that spot will be greater.

Questions:

1)


#1


#2


#3

a) Which of the above molecules is non-physiological and is only used in DNA sequencing?

b) Which one of the above molecules is used in RNA?

c) Which one of the above molecules is used in DNA?

d) Which one of the above molecules is used as the major source of energy in cells?

2) You are interested in making many copies of a specific DNA sequence. The sequence that you want to amplify is flanked by regions with the sequence given below:

```
Primer 1 should bind in this region                Primer 2 should bind in this region

5' CGCGCGAATTCGATCGA______________________________TTAATACGTACTAG 3'
3' GCGCGCTTAAGCTAGCT______________________________AATTATGCATGATC 5'
```

Circle the set(s) of primers that will amplify this region.

	Primer 1		Primer 2
Set A:	5'TCGATCGAATTC 3'	AND	5'TAATACGTACTA 3'
Set B:	5'GCTTAAGCTAGC 3'	AND	5'GATCATGCATAA 3'
[✓ Set C:]	5'GAATTCGATCGA 3'	AND	5'CTAGTACGTATT 3'

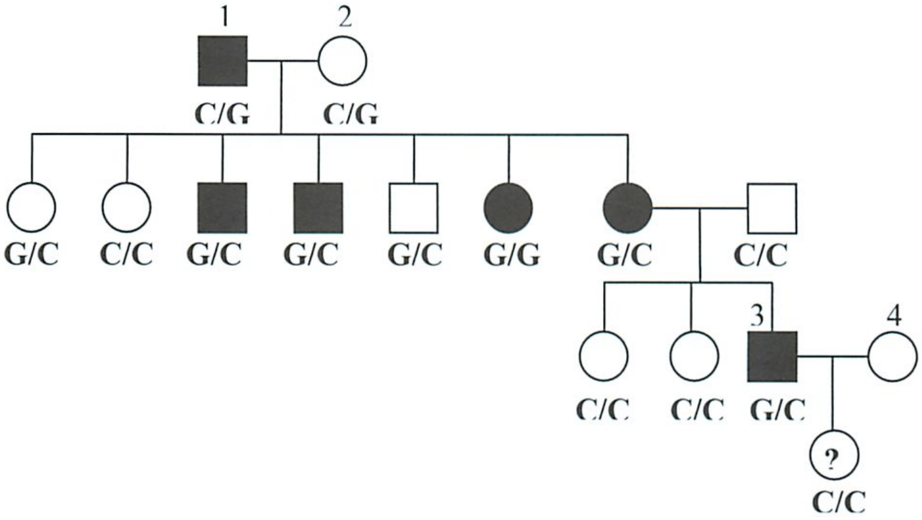
3) a) *Salmonella* is a bacterial genus that is highly related to *E. coli*. You study a *Salmonella* protein that is 100% identical to an *E. coli* protein. At the nucleotide level, the *Salmonella* gene and the *E. coli* gene are only 87% identical. Explain how this is possible.

b) Less than 5% of the human genome is made up of genes. What is the other 95% made of?

c) If you had the entire sequence of a genome of a new bacterium, how would you predict where the genes were in the genome?

d) Why would the same strategy for predicted the genes you proposed for part (c) be more difficult if you had the sequence of a genome of a eukaryote?

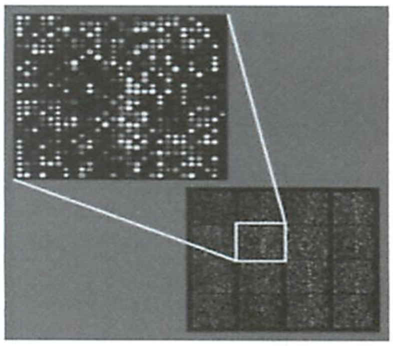
4) You identify a hypothetical Gene R in humans that encodes for protein "R" which is involved in maintaining low blood cholesterol level. This gene shows an autosomal dominant mode of inheritance and the affected individuals are at a higher risk of developing a cardiac disorder (CD). You come across a SNP (Single nucleotide polymorphism) that is tightly linked to Gene R. You decide to use this SNP as a marker for CD. The two alleles of SNP (C and G) are shown for each individual in the following pedigree. *Individuals affected by CD are shaded in black.*



a) Assuming no recombination between this SNP and Gene R, which allele of the SNP is linked to the disease associated **R**⁻ allele in individual #1?

b) Assuming no recombination between this SNP and Gene R, what is the probability that the offspring of individual #3 and individual #4 has the disease?

5) The adjacent image is a microarray. Assume that the microarray contains all of the genes in a eukaryotic genome, and that it was probed with mRNA from liver tissue.



a) What do the bright spots indicate?

b) How might the pattern be different if the same microarray was probed with mRNA from a different tissue?

From: Samin Houshyar <samin@MIT.EDU>
Sent: Tuesday, October 23, 2012 3:20 PM
To: Samin Houshyar
Subject: r27 announcement: Methylation & Acetylation

Note: This mail was sent to all students in the stellar class Introductory Biology Section 27

Methylation & Acetylation

Hi Everyone -

I didn't get a chance to finish the discussion on Q1 of the pset today. To summarize:

1. Acetylation adds acetyl groups which are negatively charged and therefore bind the positive residues (Lys, Arg) of the histones. The negative charges will repel the negative charge on DNA and DNA unwraps making it available for transcription.

2. Methylation
if methylation is on the DNA it will cause DNA to wrap around the histones and reduce gene expression.
if methylation is on the histones, it will cause the DNA to unwrap from the histones (by occluding the positive charges) and therefore gene expression is increased.

If you still have questions, fee free to shoot me an email or ask about it in OH or next recitation.

Best,
Samin

This announcement was made in Stellar on 2012 October 23 by Samin Houshyar

The announcement is also posted on the class website:
https://stellar.mit.edu/S/course/7/fa12/7.012/r27/

Bio P-set 4

10/24

Not much time to do this one unfortunately

But do really need to study over the weekend

This is on recombinant DNA

I always do way worse on these than I thought

1a) inducable

typically repressed when repressor not bound

Can be turned on when regulator protein interacts w/ a specific molecule

lac

repressible repressed when repressor bound

trp

Ok the 2 examples I looked at

Constitutive - always active

trp example → in absence allows

in abundance stops

└ puts inhibitor in

②

Where as w/ lac

in presence of lac → allows polymerase

trp - repressor bound = inhibited

not in presence → makes more

repressable

lac : repressor not bound = inhibited

inducible

So in presence → makes more

What is table?

LAmt enzyme A, B

---

b) Chromatin

Histones - spools around which DNA winds

: simple spools

I never looked at the diff types

The names are weird

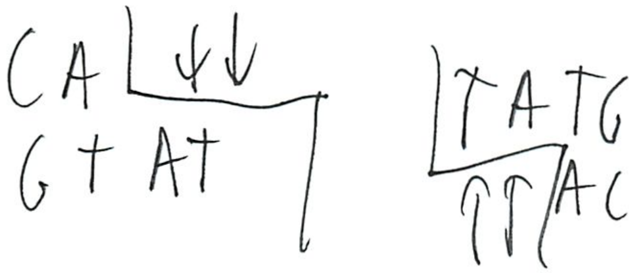
protiens : H1 H2A H2B H3 H4

8 subunit core, 2 each of H2A H2B H3 H4

③

WP: Two each of core histones assemble to form one octameric nucleosome core particle

2. Want it to match



A=T
C=G

(How both?

Sticky end

How directional?

no clue really on this

c) This is not like qv we saw before!

④

transform rapid replication

3d) Should know – slipped my mind

Rec notes 16 – top of pg

Makes perfect sense!

4) Any thing specifically out of book / notes

Jeff

CATG

CATG

↑ forgot other end

BamH + KPN → diff order

⑤

2c) This is diff

Plasmid
Restriction enzymes
Cut open circle

will cut ~~all the~~ most of the BamHI seq

2 bands → 2 pieces of DNA
2 areas where restriction enzymes cut

1 cut → just a line

restriction digest = act of restriction enzyme application

Vector ~ plasmid
basically same

? Where is small
↳ Puzzle ya need to figure out

⑥

Should always add to 5000 bP

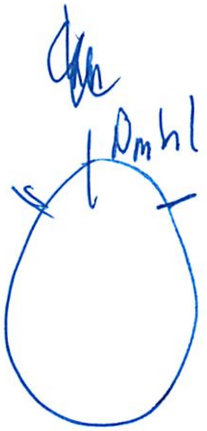
└ otherwise 2 pieces of same size

Cut locations

On a fixed loc of circle

Some of which are given

Vector does not apply



2 loc for Sma I

~~So only two.~~

Also know only 2

Just pick clockwise or CCW

Kpn I present twice

①

4. Where put primer?

Depends which want to start amplifying

Depending on what want to amplify

Does it copy primer?

Yes - has two or gets shorter

So just the start

Primer is complementary

Verify the primer template/copy and direction

---

3. Antibiotic resistance gene

if plasmid goes into bacteria cell

kill cell w/o plasmid

where just standard gene - wild type

⑧

Amp^r – Amplicine the antibiotic
r = resistance
So gene resistance to Ampicilin

must be in bacteria genone
So thinks 5

Ecoli ori
~~easiser~~ grows faster
easier to transform

to transform is to integrate
└ into genome of bacteria
Grow fast into colonies
~~Make 2 identical plasmid~~
~~for one~~ add amplicin
remaining cells – "libary"

Ad negitive ⇒ need 2
plates to compare

⑨

3c) Why would it not cut most of them

One or other doesn't get cut

→ Gene seq looking for not present

So not the same

---

d) Yeast vs Bacteria

So Amp resistance

Take Yeast Gene into Bacteria

~~Hopefully identical /compatible~~

Bacteria → seragate carrier

replicates fast

Could use yeast as well

but slow

Always can use bacteria

⑩

Clone by complementation

Only diploid – One from mom + dad

are mutations in same gene?

Same gene → 2 defective copy of same gene
doesn't grow

diff gene → one wild type copy
if that dom
will grow

here one from A, α

Point where yeast haploid

Can merge to make diploid
└ Must do before able to divide

Evoltion: want mixing

3f) If wild type dominant
Arg⁻ is dom or in diff gene ← 2 reason

(ii)

If dam

L any fixing mechanism won't show up

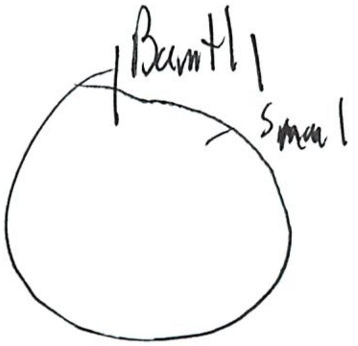
g ii) (stuck)

introns?

5' cap 3' tail

Now finish up!

2c)



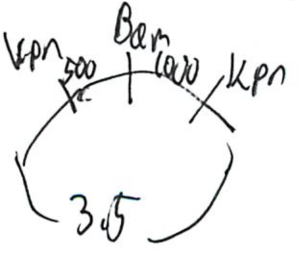
kpn twice - 1.5, 3.5 away

(diff than before)

~~don't need that~~

Bgl 1 is 1.5 or 3.5 from Smal

~~kpn is~~ Bam is 1.0 → .5 b/w kpn


kpn
500
Bam
1000
kpn
3.5

⑫

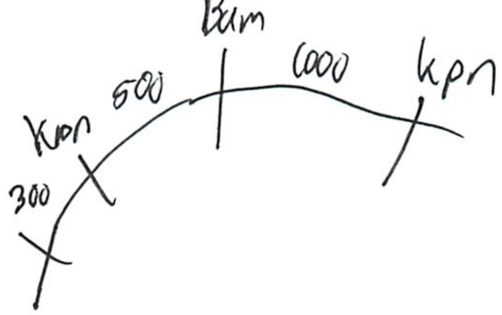
Still unsure - can I just pick a dir

but then Sma Kpn 3.2 1.5 300

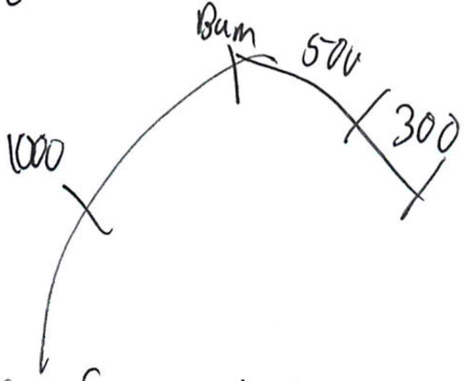
So 300 from Kpn
3.2

and 800

So


Bam
1000
Kpn
500
Kvn
300
Sma

? could it be other way


Bam
500
300
1000

sure? ~~[illegible]~~ circle
but flipped
↳ same or diff ? same I think...

Now test Bgl

between 2.5 and 1.0 on kpn

and 1.5, 3.5 from BAM

↑ 300+300+700

← 2.5+1.0

Verify ✓

Pre graded copy

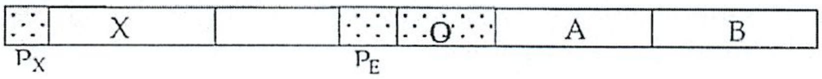
Name Michael Plasmeier Section 27 TA Houshyar

2012 7.012 Problem Set 4

Please print out this problem set and answer the questions on the printout.
Answers to this problem set are to be turned in at the box outside 68-120 before 4:00 PM, Thursday October 25th.

Question 1

The following is a diagram of an inducible operon in *E. coli* and its regulatory region. Enzymes A and B are both required for the breakdown of the sugar maltose. The wild-type operon is regulated by protein X, which is continuously produced at low levels.



P_X	promoter for the regulatory protein
X	gene for the regulatory protein of the AB operon
P_E	promoter for the A and B genes
O	sequence shown to be important for transcriptional regulation by X
A	structural gene for enzyme A
B	structural gene for enzyme B

You have three different mutants (m1, m2, and m3), each one is the result of a **loss-of-function** mutation in a single component shown in the diagram. The mutants m1, m2, and m3 exhibit the following phenotypes when grown with or without maltose in the medium.

	without maltose		with maltose	
Cell	Amount of Enzyme A	Amount of Enzyme B	Amount of Enzyme A	Amount of Enzyme B
WT	low	low	high	high
m1	high	high	high	high
m2	low	low	low	low
m3	high	high	high	high

a) Given the data from the table, label the expression in each cell type as inducible, uninducible or constitutive.

WT: inducable ✓ m1: Constitutive ✓

m2: Uninducable ✓ m3: Constitutive ✓

b) Based on the data shown above, does the regulatory protein X act as a repressor or an activator of the maltose operon? Explain your reasoning.

like lactose, it is an activator. When in presence of maltose, the regulator is busy and is not preventing the creation of enzymes

repressor

c) A single loss-of-function mutation in which component(s) [P_X, X, P_E, O, A or B] could produce the phenotype seen in the m2 mutant? Why?

O or (A and B) prevents all production of A,B no matter the outside world

P_E ← low not none

d) A single loss-of-function mutation in which component(s) [P_X, X, P_E, O, A or B] could produce the phenotype seen in m1 and m3. Explain.

✓ P_X or X since then no regulator is produced, allowing the creation of A,B no matter the amt of maltose.

Name Plasmider Section______ TA__________

Question 1, continued

Chromatin is a term used to describe a combination of DNA and protein. Chromatin functions to package DNA into a small volume, to prevent DNA damage, and to control gene expression. Proteins generally called histones are an important part of chromatin.

- Given that histones have an important role in chromatin, explain why they are usually basic proteins. They are simple spools that just wind up DNA - no need to be fancy.

DNA acidic + ⊖ charge

⊕ Charged

- When examining the histones associated with the DNA at the promoter of active genes, you find that histone H3 is trimethylated on the fourth lysine. Given this observation, describe how might histone H3 be involved in gene regulation.

Methylation on histone causes DNA to unwrap allowing DNA to be translated

Trimethylation on 4th lysine of H3 ← ?

Question 2

A schematic of the vector p7012 is shown. The restriction enzymes listed cut only where indicated; they do not cut anywhere else in the vector or insert.

Nde I:

5′ CATATG
3′ GTATAC

Sal I:

5′ GTCGAC
3′ CAGCTG

EcoR I:

5′ GAATTC
3′ CTTAAG

*BamH*I:

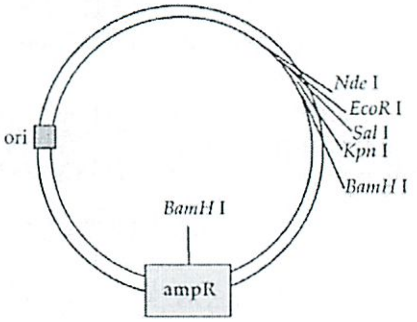
5′ GGATCC
3′ CCTAGG

Kpn I:

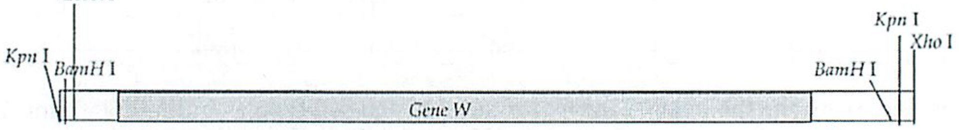
5′ GGTACC
3′ CCATGG

Xho I:

5′ CTCGAG
3′ GAGCTC



a) A schematic of gene W is below. You want to clone all of gene W into the vector p7.012. There are three different strategies that you could use to clone gene W into p7012.



- Strategy 1 uses the restriction enzyme Kpn to cut the vector and restriction enzyme Kpn to cut Gene W. ✓
- Strategy 2 uses the restriction enzyme(s) EcoR and BamH to cut the vector and restriction enzyme(s) EcoR and BamH to cut Gene W.
- Strategy 3 uses the restriction enzyme(s) BAMH and KpnI to cut the vector and restriction enzyme(s) BAMH and KpnI to cut Gene W.

Eco Sal
Eco Xho

Kpn EcoR
Kpn EcoR

b) Which strategies would allow for directional cloning?

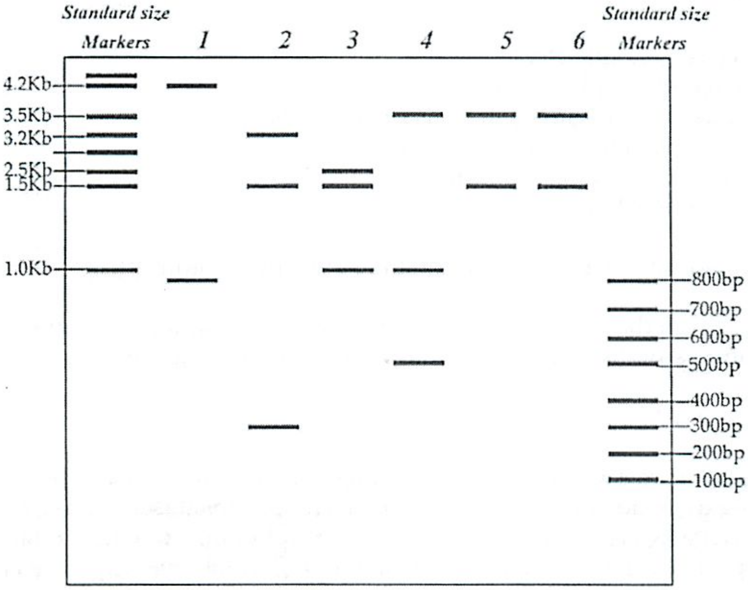
Having the "sticky ends" be a-symmetrical

Strategy 2 directional ↻ clockwise ✓
" 3 " ↺ Counter clockwise

Name____ Section_______ TA_____________

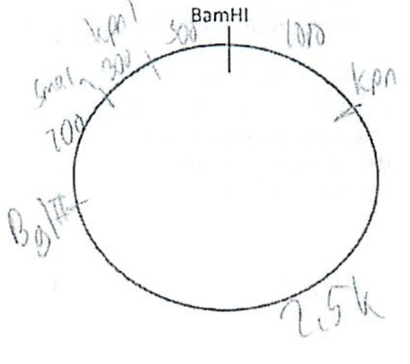
Question 2, continued

c) You are given the plasmid pSET. In order to map this plasmid you set up a series of restriction digests and obtain the following results using agarose gel electrophoresis. Assume that all restriction digests were complete, i.e., each site for each restriction enzyme on each molecule of DNA was cut.


Standard size Markers
1
2
3
4
5
6
Standard size Markers
4.2Kb
3.5Kb
3.2Kb
2.5Kb
1.5Kb
1.0Kb
800bp
700bp
600bp
500bp
400bp
300bp
200bp
100bp

Lane	*Digest*	*Size of fragments in bp*
1	Bam*HI* and Sma*I*	4,7k, 800
2	Sma*I* and Kpn*I*	3,2k, 1.5k, 900
3	Kpn*I* and Bgl*II*	2,5k 1,5k 1.0k ✓
4	Bam*HI* and Kpn*I*	3,5k 1,0k 500
5	Kpn*I*	1,5k 3,5k
6	Bgl*II* and Bam*HI*	1,5k, 3,5k

- Fill in the table above, using the information from the agarose gel to determine the approximate sizes of the fragments produced in digests 1-6.
- Use your answers to determine the approximate size of pSET. pSET = 5000 ✓ base pairs
- Use your answers to add the *Sma*I, *Kpn*I, *Bgl*II sites to plasmid map of pSET. On your map give the distances between each of the restriction sites.


BamHI
Kpn1
500
1000
Kpn
SmaI
300
700
BglII
2,5k

either way ✓

Name________________________ Section______ TA__________

Question 3

You have isolated two different yeast strains, strain 1 and strain 2. Each strain has a single mutation in a different gene such that neither strain 1 or strain 2 can grow in the absence of arginine. You want to clone the wild type copy of the gene or genes that are mutated in strain 1 and strain 2. To do so you plan to:

1) Obtain fragments of the entire yeast genomic DNA
2) Cut chosen vector and ligate each fragment into a vector
3) Use this pool of vectors and recombinant plasmids to transform *E. coli* cells
4) Select for *E. coli* cells that have obtained any vector or plasmid
5) Screen for *E. coli* transformed with a recombinant plasmid
6) Obtain recombinant plasmids from the library
7) transform yeast
8) Plate transformation mix ontp media and select for cells that are arginine prototrophs.

a) To construct a yeast genomic library in *E. coli* that will allow you to successfully complete the steps outlined above, what would be the phenotype of the yeast you would choose as the donor for the genomic DNA?

Wild Type ✓

b) You choose the vector pBluescript II, shown below. Note that the cloning site lies within lacZ, the coding region of the gene that encodes β-galactosidase. A cell that expresses β-galactosidase can take a substrate called X-gal and cleave the β-1,6 linkage to form a product that is bright blue. For each of the following sequences found on pBluescript II, list the step or steps (1-8 above) for which that sequence is needed and explain the role that sequence plays.

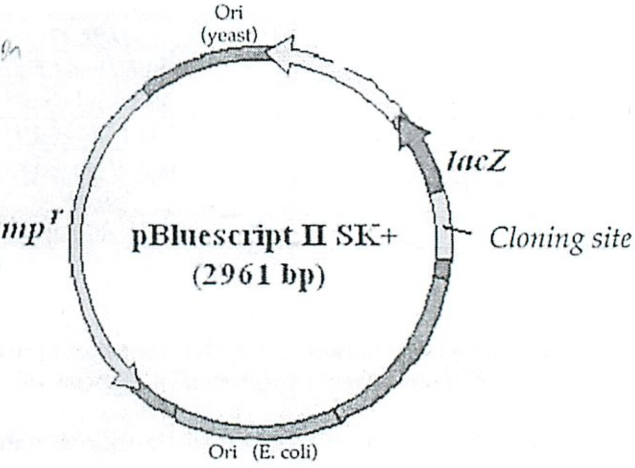
Yeast ori: 7- allow for quick replication
7 and 8

Amp^r: 4-✓ use the antibiotic resistance gene to filter for plasmids

E. coli ori:

2- are the vectors

4-6



c) You digest both the yeast genomic DNA and many copies of the vector with the BamH1 restriction enzyme. You mix the genomic fragments with the cut vectors and add DNA ligase. You then transform *E. coli* cells with the ligation mix and plate on solid agar medium.

i) If one of the many vector molecules is NOT cut with BamH1, or religates without an insert, the lacZ gene remain intact. A cell that carries this plasmid will always express the lacZ gene at high levels, independent of glucose and lactose levels. Do you expect the promoter and regulatory regions associated with this copy of the lacZ gene is the same as the promoter and regulatory regions associated with the lacZ gene in the lac operon? Explain your thoughts.

No✓ we know it must be somehow different because it always produces high levels ind of glucose + lactose levels

requires an activator

Question 3, continued

c) ii) Describe what medium you could use to distinguish the bacterial colonies that carry a non-recombinant vector from the ones that carry a new recombinant plasmid. Explain how this media would allow you to distinguish the bacterial colonies that carry a non-recombinant vector from the ones that carry a new recombinant plasmid.

A medium where any grow the recombined plasmid will grow - basically will only grow if lacZ not present

d) You successfully create a yeast genomic library in E. coli cells, and obtain a pool that represents a complete set of recombinant plasmids from the library. Briefly describe how you would use this complete set of recombinant plasmids to clone by complementation the gene that can restore the yeast of strain 1 to arginine prototrophy.

One of these in the set will have the wild type gene for what ever mutated. This would let us know this is the gene responsible for arg.

e) Would it be possible to use the same library to clone by complementation the gene that can restore the yeast of strain 2 to arginine prototrophy? Explain.

Yes. It will likely be a different gene from the library.

f) You successfully identify a recombinant vector that restores yeast strain 1 to arginine prototrophy (clone 1). You are curious as to whether this gene can also rescue a bacterial cell that is arg– (i.e., it is also an arginine auxotroph). Give 2 reasons why clone 1 *may not* work to rescue the arg– bacterial cell.

Arg- uses a different gene or Arg$^-$ is dominant

g) Your friend suggests that you use her yeast cDNA library to attempt to restore an arg– bacterial cell to arginine prototrophy.

i) Briefly describe how a cDNA library is different from a genomic library.

cDNA library has been reversed transcriptased back into DNA - removing the introns and disaggregating similar genes

ii) You transform arg– bacterial cells with your friend's yeast cDNA library and find a clone, clone 2, that restores the cells to arginine prototrophy. What sequence NOT found on pBluescript II would have been present on the vector that your friend used to create this library? Explain why this sequence is required.

oligo-dT to allow a free OH 3' end to allow reverse transcriptase

Name________________________________ Section_______ TA____________

Question 4

a) Design primers, each 16 nucleotides long, which would allow you to amplify the 80 base pairs of sequence below using PCR. Label the 5′ and 3′ ends.

```
5' GGACCGCGGGGCAGGATTGCTCCGGGCTGTTTCATGACTTGTCAGGTGGGATGACTTGGATGGAAAAGTAGAAGGTCATG   3'
   +---------+---------+---------+---------+---------+---------+---------+---------
3' CCTGGCGCCCCGTCCTAACGAGGCCCGACAAAGTACTGAACAGTCCACCCTACTGAACCTACCTTTTCATCTTCCAGTAC  5'
```

primer is the opposite

Primer 1: 3′ TTTCATCTTCCAGTAC

Primer 2: 5′ GGACCGCGGGGCAGGA

so flipped (is so works?)

b) PCR consists of a series of 20-40 repeated temperature changes, called cycles. Each cycle of PCR involves three different steps.

- To begin, the reaction mixture is prepared. List the components that must be present in the reaction mixture for successful PCR to occur.

  target DNA, primers, taq polymerase, nucleotides (dNTPs) ACTG

  template

- In the first of the regular cycling events, the reaction is heated to 94–98 °C for 20–30 seconds. What occurs during this step?

  denaturing - Double helix of DNA separates

  breaks H bond

- The reaction temperature is then lowered to 50–65 °C for 20–40 seconds. What occurs during this step? When choosing the appropriate temperature for this step, what should you be considering?

  Annealing - primers will match w/ base pair

  Cool enough so H bonds again 50-60°C

  melting temp of primers

- The reaction temperature is then raised to a temperature of 68–80 °C. What occurs during this step? When choosing the appropriate temperature for this step, what should you be considering?

  extension - DNA polymerase extends DNA

  want ideal temp for DNA polymerase 70-75°C

- If you started with a double stranded template molecule, at the completion of your PCR reaction, you will still have the original double stranded template molecule and many copies of the target DNA molecule. Will there be any other types of DNA molecules in your PCR tube. Explain.

  You have longer parts of the DNA seq besides the target seq

c) DNA sequencing using the Sanger method once required four different reaction mixes, but can now be carried out as single reaction.

- List the components needed for DNA sequencing using the Sanger method.

  dNTPs and ddNTPs, fluorescent dye

  DNA polymerase Primer

- Assume you are sequencing a single-stranded template that is 800 bp long, and your primer is 20 nucleotides long (i.e., your primer binds to nucleotides 1-20 of your template). How many different sized DNA molecules will you have when your successful sequencing reaction is complete?

  800-20 = 780 diff sized DNA

  don't code for polymer

  always first 2

Asking Jeff paid off

Name Michael Plasmeier ~~11/18~~ Section 27 TA Houshyar

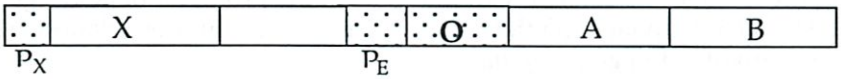
2012 7.012 Problem Set 4

Please print out this problem set and answer the questions on the printout.
Answers to this problem set are to be turned in at the box outside 68-120 before 4:00 PM, Thursday October 25^{4h}.

10/17

Question 1

The following is a diagram of an inducible operon in *E. coli* and its regulatory region. Enzymes A and B are both required for the breakdown of the sugar maltose. The wild-type operon is regulated by protein X, which is continuously produced at low levels.



P_X	promoter for the regulatory protein
X	gene for the regulatory protein of the AB operon
P_E	promoter for the A and B genes
O	sequence shown to be important for transcriptional regulation by X
A	structural gene for enzyme A
B	structural gene for enzyme B

You have three different mutants (m1, m2, and m3), each one is the result of a **loss-of-function** mutation in a single component shown in the diagram. The mutants m1, m2, and m3 exhibit the following phenotypes when grown with or without maltose in the medium.

	without maltose		with maltose	
Cell	Amount of Enzyme A	Amount of Enzyme B	Amount of Enzyme A	Amount of Enzyme B
WT	low	low	high	high
m1	high	high	high	high
m2	low	low	low	low
m3	high	high	high	high

a) Given the data from the table, label the expression in each cell type as inducible, uninducible or constitutive.

1/1 WT: inducable m1: Constitutive 1/1
m2: Uninducable m3: Constitutive

b) Based on the data shown above, does the regulatory protein X act as a repressor or an activator of the maltose operon? Explain your reasoning.

0/1 like lactose, it is an activator. When in presence of maltose, the regulator is busy and is not preventing the creation of enzymes

c) A single loss-of-function mutation in which component(s) [P_X, X, P_E, O, A or B] could produce the phenotype seen in the m2 mutant? Why?

O or (A and B) prevents all production of A, B no matter the outside world

d) A single loss-of-function mutation in which component(s) [P_X, X, P_E, O, A or B] could produce the phenotype seen in m1 and m3. Explain.

Px or X since then no regulator is produced, allowing the creation of A, B no matter the amt of maltose,

Name Plasmaler Section______ TA__________

Question 1, continued

Chromatin is a term used to describe a combination of DNA and protein. Chromatin functions to package DNA into a small volume, to prevent DNA damage, and to control gene expression. Proteins generally called histones are an important part of chromatin.

- Given that histones have an important role in chromatin, explain why they are usually basic proteins. They are simple spools that just wind up DNA – no need to be fancy.

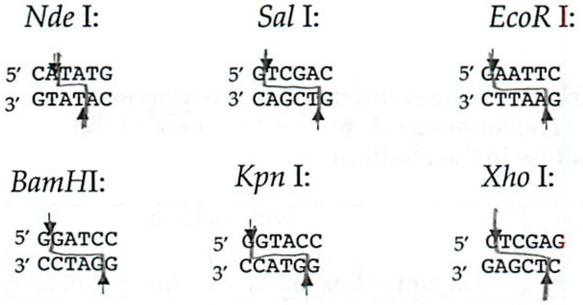
0/1

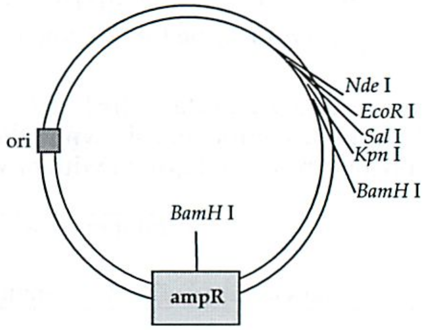
- When examining the histones associated with the DNA at the promoter of active genes, you find that histone H3 is trimethylated on the fourth lysine. Given this observation, describe how might histone H3 be involved in gene regulation.

Methylation on histone causes DNA to unwrap allowing DNA to be translated

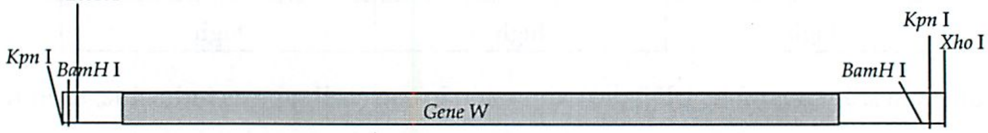
Question 2

A schematic of the vector p7012 is shown. The restriction enzymes listed cut only where indicated; they do not cut anywhere else in the vector or insert.


Nde I:
5′ CATATG
3′ GTATAC
Sal I:
5′ GTCGAC
3′ CAGCTG
EcoR I:
5′ GAATTC
3′ CTTAAG
*BamH*I:
5′ GGATCC
3′ CCTAGG
Kpn I:
5′ GGTACC
3′ CCATGG
Xho I:
5′ CTCGAG
3′ GAGCTC


Nde I
EcoR I
Sal I
Kpn I
BamH I
ori
BamH I
ampR

a) A schematic of gene W is below. You want to clone all of gene W into the vector p7.012. There are three different strategies that you could use to clone gene W into p7012.


EcoR I
Kpn I
BamH I
Gene W
Kpn I
Xho I
BamH I

- Strategy 1 uses the restriction enzyme Kpn to cut the vector and restriction enzyme Kpn to cut Gene W.
- Strategy 2 uses the restriction enzyme(s) EcoR *and* BamH to cut the vector and restriction enzyme(s) EcoR *and* BamH to cut Gene W.
- Strategy 3 uses the restriction enzyme(s) BAMH *and* Kpn I to cut the vector and restriction enzyme(s) BAMH *and* Kpn I to cut Gene W.

0/1 X

X

b) Which strategies would allow for directional cloning?

Having the "sticky ends" be a-symmetrical

Strategy 2 directional ↓ clockwise

" 3 " ↺ counter clockwise

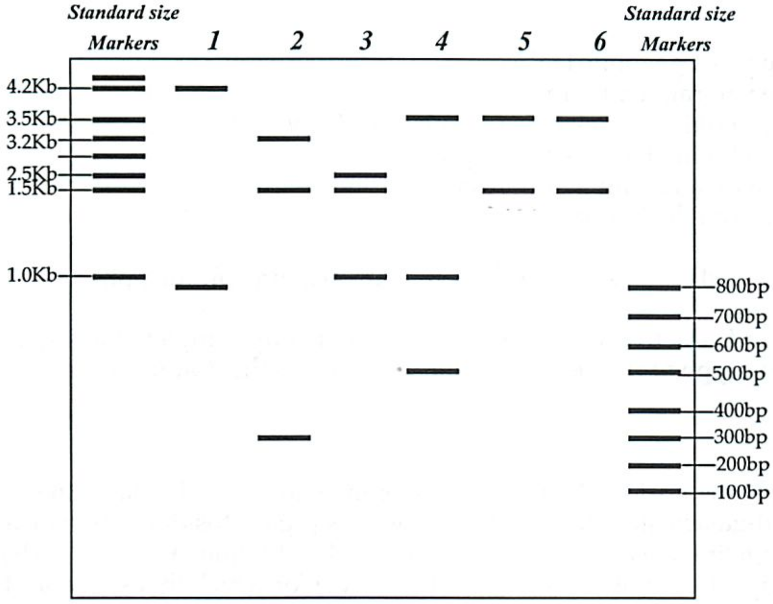
1/1

Right idea.

Name________________________ Section______ TA__________

Question 2, continued

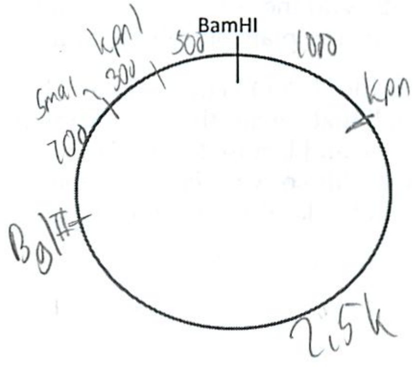
c) You are given the plasmid pSET. In order to map this plasmid you set up a series of restriction digests and obtain the following results using agarose gel electrophoresis. Assume that all restriction digests were complete, i.e., each site for each restriction enzyme on each molecule of DNA was cut.


Standard size Markers
1
2
3
4
5
6
Standard size Markers
4.2Kb
3.5Kb
3.2Kb
2.5Kb
1.5Kb
1.0Kb
800bp
700bp
600bp
500bp
400bp
300bp
200bp
100bp

Lane	*Digest*	*Size of fragments in bp*
1	Bam*HI* and Sma*I*	4.7k, 800
2	Sma*I* and Kpn*I*	3.2k, 1.5k, 300
3	Kpn*I* and Bgl*II*	2.5k 1.5k 1.0k
4	Bam*HI* and Kpn*I*	3.5k 1.0k 500
5	Kpn*I*	1.5k, 3.5k
6	Bgl*II* and Bam*HI*	1.5k; 3.5k

1/1

- Fill in the table above, using the information from the agarose gel to determine the approximate sizes of the fragments produced in digests 1-6.
- Use your answers to determine the approximate size of pSET. pSET = 5000 base pairs 1/1
- Use your answers to add the *Sma*I, *Kpn*I, *Bgl*II sites to plasmid map of pSET. On your map give the distances between each of the restriction sites.


BamHI
kpnI
500
1000
SmaI
300
Kpn
700
BglII
2.5k

Name________________________________ Section______ TA___________

Question 3

You have isolated two different yeast strains, strain 1 and strain 2. Each strain has a single mutation in a different gene such that neither strain 1 or strain 2 can grow in the absence of arginine. You want to clone the wild type copy of the gene or genes that are mutated in strain 1 and strain 2. To do so you plan to:

1) Obtain fragments of the entire yeast genomic DNA
2) Cut chosen vector and ligate each fragment into a vector
3) Use this pool of vectors and recombinant plasmids to transform *E. coli* cells
4) Select for *E. coli* cells that have obtained any vector or plasmid
5) Screen for *E. coli* transformed with a recombinant plasmid
6) Obtain recombinant plasmids from the library
7) transform yeast
8) Plate transformation mix ontp media and select for cells that are arginine prototrophs.

a) To construct a yeast genomic library in *E. coli* that will allow you to successfully complete the steps outlined above, what would be the **phenotype** of the yeast you would choose as the donor for the genomic DNA?

1/1

Wild Type

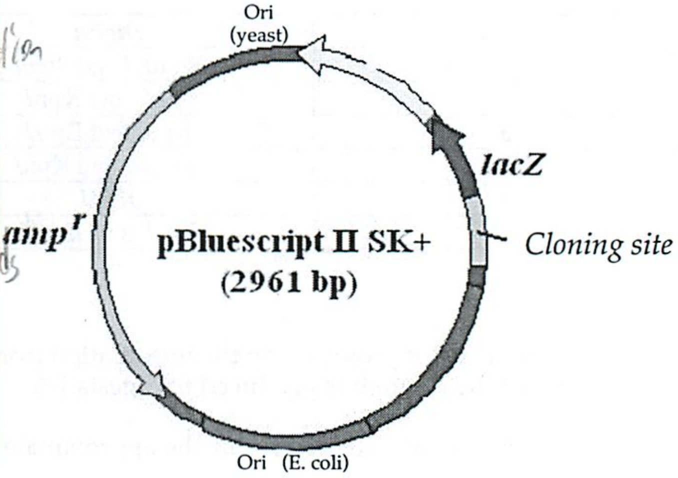
b) You choose the vector pBluescript II, shown below. Note that the cloning site lies within lacZ, the coding region of the gene that encodes β-galactosidase. A cell that expresses β-galactosidase can take a substrate called X-gal and cleave the β-1,6 linkage to form a product that is bright blue. For each of the following sequences found on pBluescript II, list the step or steps (1-8 above) for which that sequence is needed and explain the role that sequence plays.

Yeast ori: 7- allow for quick replication

Amp^r: 4- use the antibodic resistance gene to filter for plasmids

E. coli ori:

2- are the vectors



c) You digest both the yeast genomic DNA and many copies of the vector with the BamH1 restriction enzyme. You mix the genomic fragments with the cut vectors and add DNA ligase. You then transform *E. coli* cells with the ligation mix and plate on solid agar medium.

i) If one of the many vector molecules is NOT cut with BamH1, or religates without an insert, the lacZ gene remain intact. A cell that carries this plasmid will always express the lacZ gene at high levels, independent of glucose and lactose levels. Do you expect the promoter and regulatory regions associated with this copy of the lacZ gene is the same as the promoter and regulatory regions associated with the lacZ gene in the lac operon? Explain your thoughts.

No we know it must be somehow different because it always produces high levels, ind of glucose + lactose levels.

Name______________________________ Section______ TA__________

Question 3, continued

c) ii) Describe what medium you could use to distinguish the bacterial colonies that carry a non-recombinant vector from the ones that carry a new recombinant plasmid. Explain how this media would allow you to distinguish the bacterial colonies that carry a non-recombinant vector from the ones that carry a new recombinant plasmid.

0/1 see sol'n

A meadium where ony grow the recombined plasmid will grow - basically will only grow if lacZ not present

d) You successfully create a yeast genomic library in E. coli cells, and obtain a pool that represents a complete set of recombinant plasmids from the library. Briefly describe how you would use this complete set of recombinant plasmids to clone by complementation the gene that can restore the yeast of strain 1 to arginine prototrophy.

One of these in the set will have the wild type gene for what ever mutated. This would let us know this is the gene responsible for arg.

e) Would it be possible to use the same library to clone by complementation the gene that can restore the yeast of strain 2 to arginine prototrophy? Explain.

0/1

Yes. It will likely be a different gene from the library.

f) You successfully identify a recombinant vector that restores yeast strain 1 to arginine prototrophy (clone 1). You are curious as to whether this gene can also rescue a bacterial cell that is arg– (i.e., it is also an arginine auxotroph). Give 2 reasons why clone 1 *may not* work to rescue the arg– bacterial cell.

Arg- uses a different gene or Arg⁻ is dominant

g) Your friend suggests that you use her yeast cDNA library to attempt to restore an arg– bacterial cell to arginine prototrophy.

i) Briefly describe how a cDNA library is different from a genomic library.

0/1

cDNA library has been reversed transciptased back into DNA - removing the introns and disaggregating similar genes

ii) You transform arg– bacterial cells with your friend's yeast cDNA library and find a clone, clone 2, that restores the cells to arginine prototrophy. What sequence NOT found on pBluescript II would have been present on the vector that your friend used to create this library? Explain why this sequence is required.

See sol'n

0/1

oligo-dT to allow a free OH 3' end to allow reverse transciptase

Name______________________________ Section______ TA__________

Question 4

a) Design primers, each 16 nucleotides long, which would allow you to amplify the 80 base pairs of sequence below using PCR. Label the 5′ and 3′ ends.

```
5' GGACCGCGGGGCAGGATTGCTCCGGGCTGTTTCATGACTTGTCAGGTGGGATGACTTGGATGGAAAAGTAGAAGGTCATG   3'
   +---------+---------+---------+---------+---------+---------+---------+---------
3' CCTGGCGCCCCGTCCTAACGAGGCCCGACAAAGTACTGAACAGTCCACCCTACTGAACCTACCTTTTCATCTTCCAGTAC   5'
```

primer is the opposite

1/1

Primer 1: 3' TTTCATCTTCCAGTAC

Primer 2: 5' GGACCGCGGGGCAGGA

b) PCR consists of a series of 20-40 repeated temperature changes, called cycles. Each cycle of PCR involves three different steps.

- To begin, the reaction mixture is prepared. List the components that must be present in the reaction mixture for successful PCR to occur.

1/1 target DNA, primers, taq polymerase, nucleotides (dNTPs) ACTG

- In the first of the regular cycling events, the reaction is heated to 94–98 °C for 20–30 seconds. What occurs during this step?

1/1 denaturing - Double helix of DNA separates

- The reaction temperature is then lowered to 50–65 °C for 20–40 seconds. What occurs during this step? When choosing the appropriate temperature for this step, what should you be considering?

1/1 Anealing – Primers will match w/ base pair, Cool enough so H bonds again 50-60°C

- The reaction temperature is then raised to a temperature of 68–80 °C. What occurs during this step? When choosing the appropriate temperature for this step, what should you be considering?

extension – DNA polymerase extends DNA 70-75°C want ideal temp for DNA polymerase

- If you started with a double stranded template molecule, at the completion of your PCR reaction, you will still have the original double stranded template molecule and many copies of the target DNA molecule. Will there be any other types of DNA molecules in your PCR tube. Explain.

You have longer parts of the DNA seq besides the target seq

c) DNA sequencing using the Sanger method once required four different reaction mixes, but can now be carried out as single reaction.

- List the components needed for DNA sequencing using the Sanger method.

dNTPs and ddNTPs, florescent dye?

- Assume you are sequencing a single-stranded template that is 800 bp long, and your primer is 20 nucleotides long (i.e., your primer binds to nucleotides 1-20 of your template). How many different sized DNA molecules will you have when your successful sequencing reaction is complete?

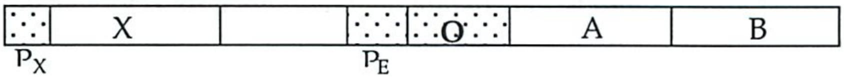
1/1 800-20 = 780 diff sized DNA

? don't code for polymer

Solutions to 2012 7.012 Problem Set 4

Question 1

The following is a diagram of an inducible operon in *E. coli* and its regulatory region. Enzymes A and B are both required for the breakdown of the sugar maltose. The wild-type operon is regulated by protein X, which is continuously produced at low levels.



P_X	promoter for the regulatory protein
X	gene for the regulatory protein of the AB operon
P_E	promoter for the A and B genes
O	sequence shown to be important for transcriptional regulation by X
A	structural gene for enzyme A
B	structural gene for enzyme B

You have three different mutants (m1, m2, and m3), each one is the result of a **loss-of-function** mutation in a single component shown in the diagram. The mutants m1, m2, and m3 exhibit the following phenotypes when grown with or without maltose in the medium.

	without maltose		with maltose	
Cell	Amount of Enzyme A	Amount of Enzyme B	Amount of Enzyme A	Amount of Enzyme B
WT	low	low	high	high
m1	high	high	high	high
m2	low	low	low	low
m3	high	high	high	high

a) Given the data from the table, label the expression in each cell type as inducible, uninducible or constitutive.

WT: ***inducible*** m1: ***constitutive***

m2: ***uninducible*** m3: ***constitutive***

b) Based on the data shown above, does the regulatory protein X act as a repressor or an activator of the maltose operon? Explain your reasoning.
Protein X is acting as a repressor. If it were an activator, there would not be any loss-of-function mutations that caused constitutive expression.

c) A single loss-of-function mutation in which component(s) [P_X, X, P_E, O, A or B] could produce the phenotype seen in the m2 mutant? Why?
A loss-of-function mutation in P_E could cause weaker expression of A and B even in the presence of maltose. It is also possible that X is mutated in a way that disrupts its ability to bind maltose.

d) A single loss-of-function mutation in which component(s) [P_X, X, P_E, O, A or B] could produce the phenotype seen in m1 and m3. Explain.
Loss-of-function in P_x or X that prevented X from being expressed or being able to bind O could prevent A and B from ever being repressed by X. A mutation in O that prevented it from being bound by X would also result in this phenotype.

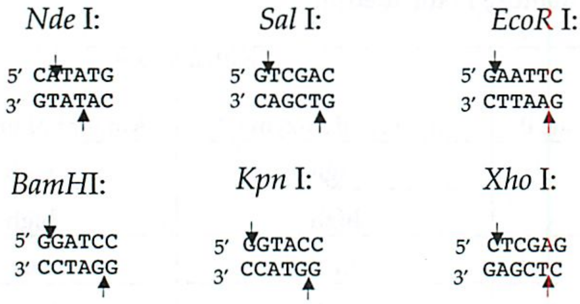
Chromatin is a term used to describe a combination of DNA and protein. Chromatin functions to package DNA into a small volume, to prevent DNA damage, and to control gene expression. Proteins generally called histones are an important part of chromatin.

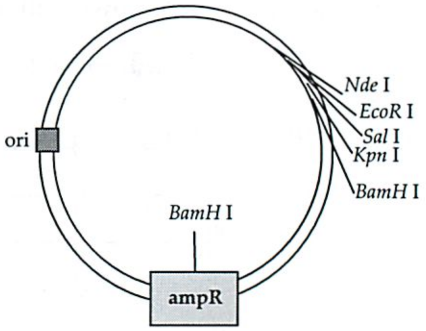
- Given that histones have an important role in chromatin, explain why they are usually basic proteins.
  Histones must interact with DNA, which is acidic and negatively charged. Therefore, basic, positively-charged proteins will be able to interact favorably with DNA and allow for its compaction.

- When examining the histones associated with the DNA at the promoter of active genes, you find that histone H3 is trimethylated on the fourth lysine. Given this observation, describe how might histone H3 be involved in gene regulation.
  Trimethylation on the fourth lysine of H3 must cause DNA to be less tightly wrapped. By opening up the chromatin, the promoters of genes in this area are more accessible to RNA polymerase.

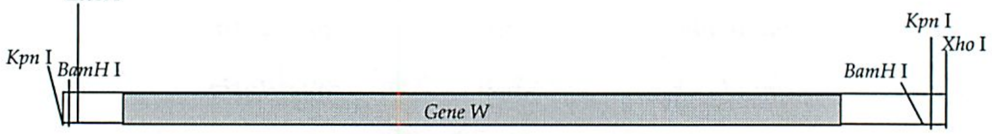
Question 2

A schematic of the vector p7012 is shown. The restriction enzymes listed cut only where indicated; they do not cut anywhere else in the vector or insert.





a) A schematic of gene W is below. You want to clone all of gene W into the vector p7.012. There are three different strategies that you could use to clone gene W into p7012.



- Strategy 1 uses the restriction enzyme *Kpn1* to cut the vector and restriction enzyme *Kpn1* to cut Gene W.

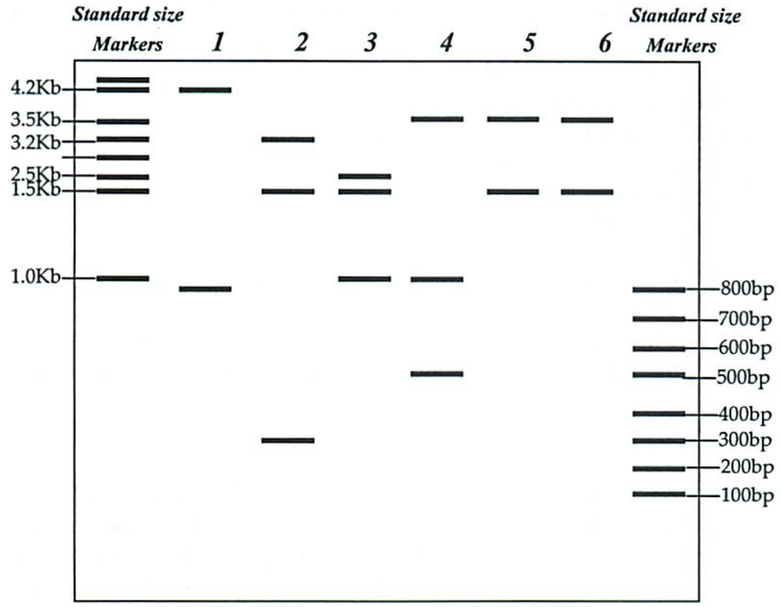
- Strategy 2 uses the restriction enzyme(s) *EcoRI and SalI* to cut the vector and restriction enzyme(s) *EcoRI and XhoI* to cut Gene W.

- Strategy 3 uses the restriction enzyme(s) *Kpn1 and EcoRI* to cut the vector and restriction enzyme(s) *Kpn1 and EcoRI* to cut Gene W.

b) Which strategies would allow for directional cloning?

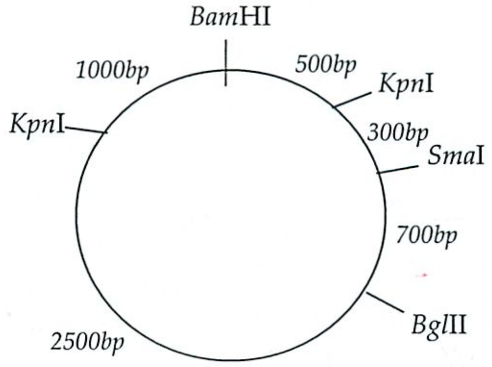
Both Strategy 2 and Strategy 3 would allow for directional cloning, because the two ends of Gene W fragment will have different overhangs left by the different restriction enzymes. The overhangs on the ends of the linearized plasmid will also be different from one another, and so there will be only one way for the Gene W fragment to fit into the plasmid.

c) You are given the plasmid pSET. In order to map this plasmid you set up a series of restriction digests and obtain the following results using agarose gel electrophoresis. Assume that all restriction digests were complete, i.e., each site for each restriction enzyme on each molecule of DNA was cut.


Standard size Markers
1
2
3
4
5
6
Standard size Markers
4.2Kb
3.5Kb
3.2Kb
2.5Kb
1.5Kb
1.0Kb
800bp
700bp
600bp
500bp
400bp
300bp
200bp
100bp

Lane	*Digest*	*Size of fragments in bp*
1	Bam*HI* and Sma*I*	***4200, 800***
2	Sma*I* and Kpn*I*	***3200, 1500, 300***
3	Kpn*I* and Bgl*II*	***2500, 1500, 1000***
4	Bam*HI* and Kpn*I*	***3500, 1000, 500***
5	Kpn*I*	***3500, 1500***
6	Bgl*II* and Bam*HI*	***3500, 1500***

- Fill in the table above, using the information from the agarose gel to determine the approximate sizes of the fragments produced in digests 1-6.
- Use your answers to determine the approximate size of pSET. pSET = *5,000* base pairs
- Use your answers to add the *Sma*I, *Kpn*I, *Bgl*II sites to plasmid map of pSET. On your map give the distances between each of the restriction sites.


*Bam*HI
1000bp
500bp
*Kpn*I
*Kpn*I
300bp
*Sma*I
700bp
*Bgl*II
2500bp

Question 3

You have isolated two different yeast strains, strain 1 and strain 2. Each strain has a single mutation in a different gene such that neither strain 1 or strain 2 can grow in the absence of arginine. You want to clone the wild type copy of the gene or genes that are mutated in strain 1 and strain 2. To do so you plan to:

1) Obtain fragments of the entire yeast genomic DNA
2) Cut chosen vector and ligate each fragment into a vector
3) Use this pool of vectors and recombinant plasmids to transform *E. coli* cells
4) Select for *E. coli* cells that have obtained any vector or plasmid
5) Screen for *E. coli* transformed with a recombinant plasmid
6) Obtain recombinant plasmids from the library
7) Transform yeast
8) Plate transformation mix onto media and select for cells that are arginine prototrophs.

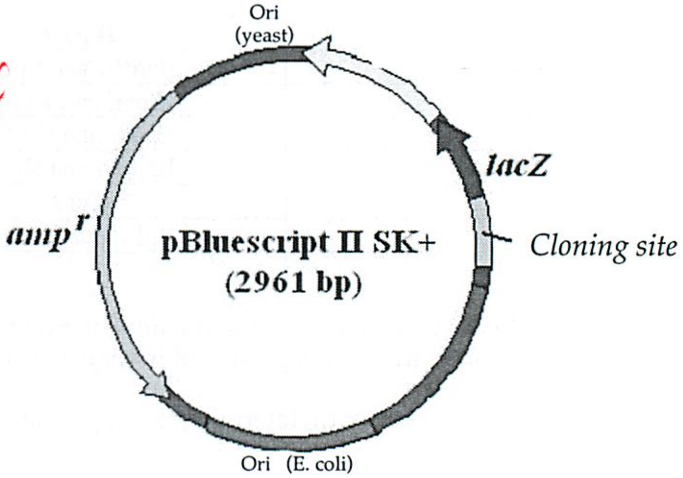
a) To construct a yeast genomic library in *E. coli* that will allow you to successfully complete the steps outlined above, what would be the **phenotype** of the yeast you would choose as the donor for the genomic DNA?
The donor should be wild type (arginine prototroph).

b) You choose the vector pBluescript II, shown below. Note that the cloning site lies within lacZ, the coding region of the gene that encodes β-galactosidase. A cell that expresses β-galactosidase can take a substrate called X-gal and cleave the β-1,6 linkage to form a product that is bright blue. For each of the following sequences found on pBluescript II, list the step or steps (1-8 above) for which that sequence is needed and explain the role that sequence plays.

Yeast ori:
The yeast ori is needed for step 7 and 8. Once the yeast are growing and dividing, they need to copy the plasmid so that each cell has a plasmid.

Amp^r:
Amp^r is needed for step 4. **E. coli** *cells that obtain a plasmid will grow on media with ampicillin, those that did not will die.*

E. coli ori:
The **E.coli** *ori is needed for steps 4 - 6.* **E. coli** *cells that obtain a plasmid will grow on media with ampicillin. As they grow and divide, they need to copy the plasmid so that each cell has a plasmid.*



c) You digest both the yeast genomic DNA and many copies of the vector with the BamH1 restriction enzyme. You mix the genomic fragments with the cut vectors and add DNA ligase. You then transform *E. coli* cells with the ligation mix and plate on solid agar medium.

i) If one of the many vector molecules is NOT cut with BamH1, or religates without an insert, the lacZ gene remains intact. A cell that carries this plasmid will always express the lacZ gene at high levels, independent of glucose and lactose levels. Do you expect the promoter and regulatory regions associated with this copy of the lacZ gene are the same as the promoter and regulatory regions associated with the lacZ gene in the lac operon? Explain your thoughts.
No, the LacZ operon promoter is weak and requires an activator. So the fact that lacZ is always expressed from this plasmid would indicate that a different promoter/regulatory regions have been chosen for use on this vector. Furthermore, expression from the endogenous lac operon is conditional (i.e., dependent upon the presence of lactose) and the expression of lacZ from the vector is independent of lactose, thus there must be a different promoter in the vector.

¿ or a diff operator ¿
Or non functioning?

ii) Describe what medium you could use to distinguish the bacterial colonies that carry a non-recombinant vector from the ones that carry a new recombinant plasmid. Explain how this media would allow you to distinguish the bacterial colonies that carry a non-recombinant vector from the ones that carry a new recombinant plasmid.

from book

You would add X-gal to the growth medium. Cells that express β-galactosidase (from the LacZ gene) can take X-gal and form a blue product, so these cells will be blue. Blue cells have an intact gene for β –galactosidase, so they carry the non-recombinant vector. White cells carry a recombinant vector because the LacZ gene is disrupted by the inserted gene.

d) You successfully create a yeast genomic library in E. coli cells, and obtain a pool that represents a complete set of recombinant plasmids from the library. Briefly describe how you would use this complete set of recombinant plasmids to clone by complementation the gene that can restore the yeast of strain 1 to arginine prototrophy.

You would grow strain 1 yeast cells and then transform them with the pool of recombinant plasmids such that some yeast cell obtain one plasmid, and others don't get a plasmid. You would plate the transformed yeast cells on media that lacks arginine. Strain 1 yeast cells that received a recombinant plasmid that carried the wild-type copy of the mutated gene are restored to arginine prototrophy and will now grow without supplemental arginine. Anything else will not grow without supplemental arginine.

i. cloning by Complentation

e) Would it be possible to use the same library to clone by complementation the gene that can restore the yeast of strain 2 to arginine prototrophy? Explain.

Yes, the yeast from which you made your library were wild-type and thus the pool of plasmids in the library should also contain the wild-type copy of the gene mutated in strain 2.

f) You successfully identify a recombinant vector that restores yeast strain 1 to arginine prototrophy (clone 1). You are curious as to whether this gene can also rescue a bacterial cell that is arg– (i.e., it is also an arginine auxotroph). Give 2 reasons why clone 1 *may not* work to rescue the arg– bacterial cell.

You used yeast genomic DNA which means that the recombinant plasmid has a yeast promoter driving expression of the arg gene, but **E. coli** ***will not recognize the yeast promoter. Genomic DNA from eukaryotes may have introns, but bacteria cannot splice a eukaryotic gene. It is also possible that the arg- bacterial cells are lacking a different gene (maybe there are many enzymes involved in the synthesis of arginine and the arg- bacterial cell is missing a different gene in the pathway). Finally, bacterial cells may synthesize arginine using a completely different pathway so the recombinant vector found will not complement the defect.***

g) Your friend suggests that you use her yeast cDNA library to attempt to restore an arg– bacterial cell to arginine prototrophy.

i) Briefly describe how a cDNA library is different from a genomic library.

A cDNA library is made of DNA fragments that represent the mature mRNA's of a cell. As such, the yeast DNA fragments in this library will not have an endogenous promoter and will not have introns.

ii) You transform arg– bacterial cells with your friend's yeast cDNA library and find a clone, clone 2, that restores the cells to arginine prototrophy. What sequence NOT found on pBluescript II would have been present on the vector that your friend used to create this library? Explain why this sequence is required.

The yeast cDNA fragments in this library will not have promoters. Because you are looking to express the gene in **E. coli*****, your plasmid must contain an*** **E. coli** ***promoter adjacent to the cloning site.***

Question 4

a) Design primers, each 16 nucleotides long, which would allow you to amplify the 80 base pairs of sequence below using PCR. Label the 5′ and 3′ ends.

```
5' GGACCGCGGGGCAGGATTGCTCCGGGCTGTTTCATGACTTGTCAGGTGGGATGACTTGGATGGAAAAGTAGAAGGTCATG  3'
   +---------+---------+---------+---------+---------+---------+---------+---------
3' CCTGGCGCCCCGTCCTAACGAGGCCCGACAAAGTACTGAACAGTCCACCCTACTGAACCTACCTTTTCATCTTCCAGTAC  5'
```

Primer 1: ***5′-GGACCGCGGGGCAGGA-3′***

Primer 2: ***5′-CATGACCTTCTACTTT-3′***

b) PCR consists of a series of 20-40 repeated temperature changes, called cycles. Each cycle of PCR involves three different steps.

- To begin, the reaction mixture is prepared. List the components that must be present in the reaction mixture for successful PCR to occur.
  Template DNA, a pair of DNA primers, dNTPs (nucleotides), and DNA polymerase are the critical components. Appropriate buffers and divalent cations are also needed.

- In the first of the regular cycling events, the reaction is heated to 94–98 °C for 20–30 seconds. What occurs during this step?
  The double-stranded template DNA is denatured into single strands – the heat breaks the hydrogen bonds between the two strands.

- The reaction temperature is then lowered to 50–65 °C for 20–40 seconds. What occurs during this step? When choosing the appropriate temperature for this step, what should you be considering?
  The primers anneal to the single stranded template DNA. The melting temperature of the primers (a function of their GC content) is a key factor to consider when determining the appropriate temperature for this PCR step.

- The reaction temperature is then raised to a temperature of 68–80 °C. What occurs during this step? When choosing the appropriate temperature for this step, what should you be considering?
  This is the extension or elongation step. DNA polymerase adds nucleotides to synthesize new DNA strands. Considerations include the specific DNA polymerase enzyme you are using (they each have different optimal temperatures), and the primer melting temperature (you don't want to be too high above this temperature, or your primers will leave the template before you can begin replication).

- If you started with a double stranded template molecule, at the completion of your PCR reaction, you will still have the original double stranded template molecule and many copies of the target DNA molecule. Will there be any other types of DNA molecules in your PCR tube. Explain.
  Yes, there will be other fragments. Any reaction that uses the original template molecule as its template could extend past the binding site of the other primer and create a longer DNA molecule. There will be excess DNA primers left over at the end of the PCR as well.

c) DNA sequencing using the Sanger method once required four different reaction mixes, but can now be carried out as single reaction.

- List the components needed for DNA sequencing using the Sanger method.
  DNA polymerase, a primer, and all four types of dNTPs/deoxyribonucleotides are necessary in all cases (along with buffer, divalent cations, etc.) Each separate "lane" in a Sanger sequencing experiment will also require one of the four types of ddNTPs (dideoxyribonucleotides) – ddATP, ddCTP, ddGTP, OR ddTTP. This was later updated to utilize fluorescently tagged dideoxyribonucleotides and laser-based monitoring of chain termination, allowing the procedure to be automated and done in a single reaction instead of four separate ones.

- Assume you are sequencing a single-stranded template that is 800 bp long, and your primer is 20 nucleotides long (i.e., your primer binds to nucleotides 1-20 of your template). How many different sized DNA molecules will you have when your successful sequencing reaction is complete?
  780. The primers will always bind the first twenty nucleotides, but then the reaction could be stopped at any (and in reality, will be stopped at each) subsequent nucleotide. For example, there will be a 21 bp molecule, a 22 bp molecule, etc.

Bio Recitation

10/25

P-set Due Today

└ Make sure to turn in!

Exam 2 Wed! – up to Mon's Lecture Recombinant DNA

– Neurobiology → next exam

Review session 10 - 12 Tuesday

am pm

QnA - non structured

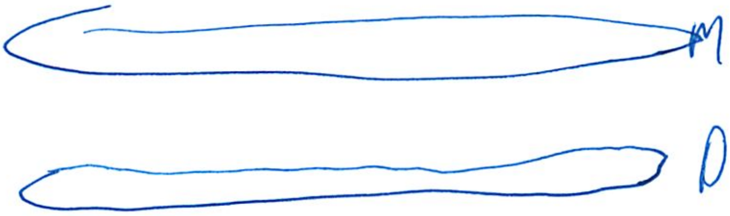
will post questions

Review Handout

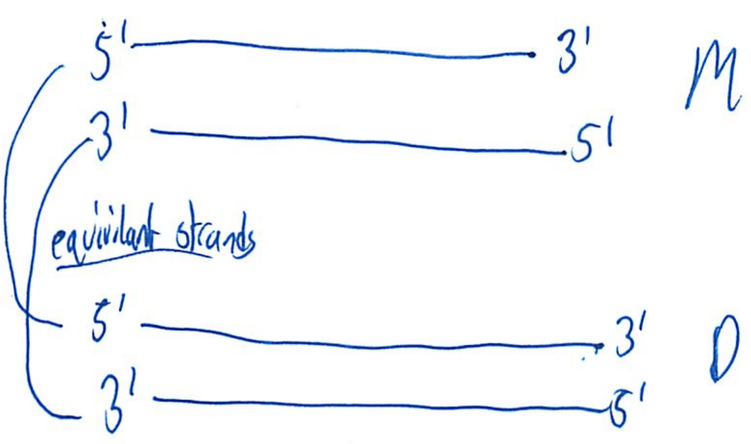
SNP Analysis

Single nucleotide poly-morphism

└ one word



②


5'
3'
M
3'
5'
equivilant strands
5'
3'
D
3'
5'

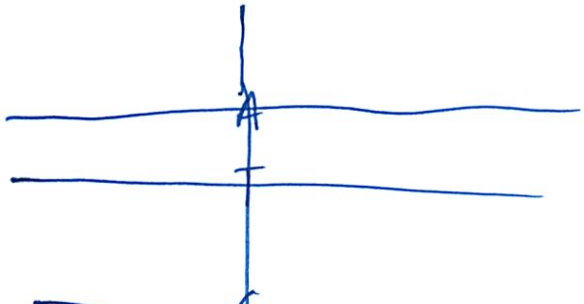
equivilent ⇒ very similar, code for same stuff

both coding strands

Watson + Crick strandss

call one Watson, other crick

← strands overall



are specific locations horizontally

└ SNP regions

single nucleotide

many forms

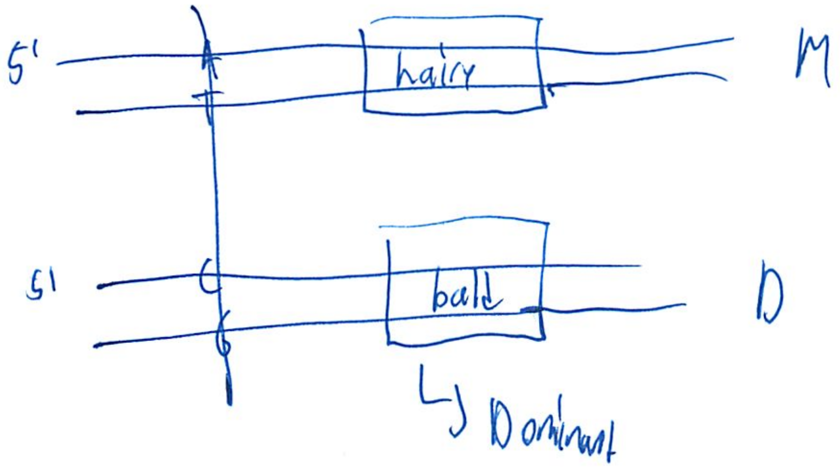
A/C

↑ don't say other ones

③

Snips usually in intron → not gene

↙ One gene - not true in real life


5'
hairy
M
5'
bald
D
↳ Dominant

Don't know where gene is

Or pattern of inheritance (confused)
└ will be given

But know C at line → Bald

Family tree

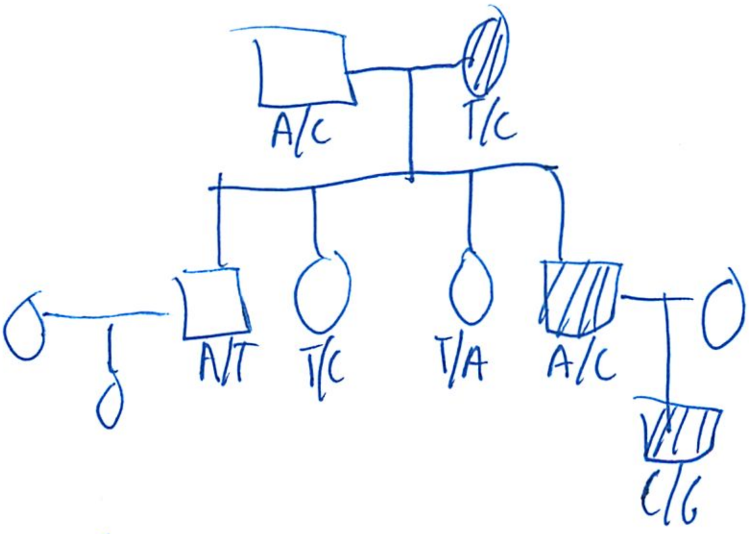

A/C
T/C
A/T
T/C
T/A
A/C
C/G

figure out who else is bald

④

Note order T/C does not matter

Both bald have C
We also know they are physically close
So very low chance of recombination

But not all Cs linked to gene
└ polymorphic

So ● T/C* ← C is associated with disease

□ A/C ← C is not associated

○ T/C ← has to be from dad, since not associated

▨ C/G ← G has to come from Mom
So C* has to come from Dad
So associated

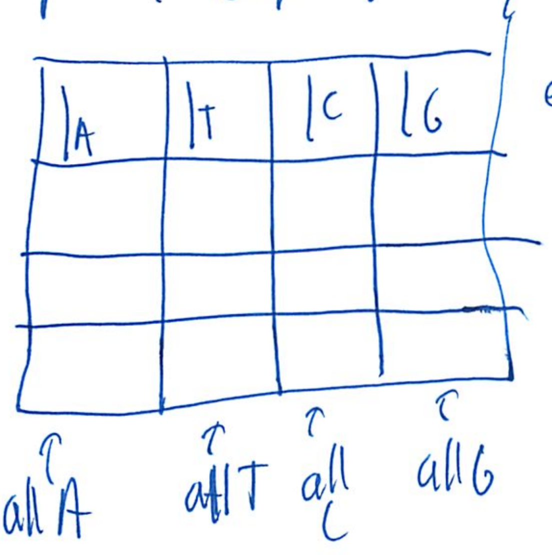
⑤

1. Determine mode of inheritence
2. Find associated SNP
3. Track

Handout talks about MicroArrays but was not covered in class, so skip it

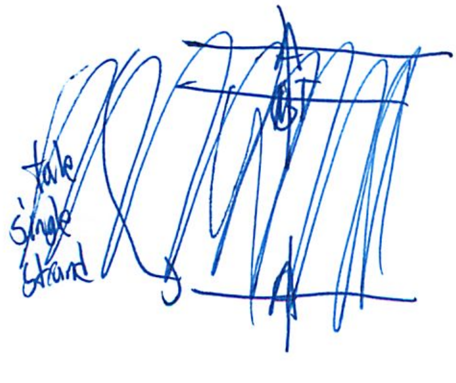
SNP Array

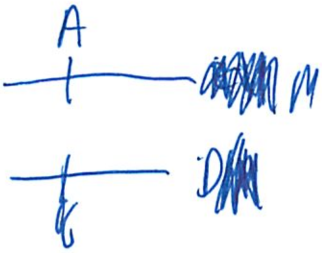
Used to quickly figure out what SNP you have at certain position. Chip w/ many spots


A
T
C
G
all A
all T
all C
all G

← where polymorphism happening

seq same - except 1 place where snip


take single strand
A
T
A



everything that matches in sequence will bind

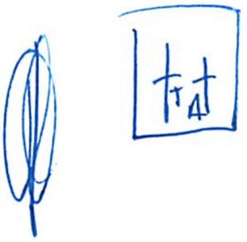
⑥

Special flurecent

bound

└ if binding happens

Some binds

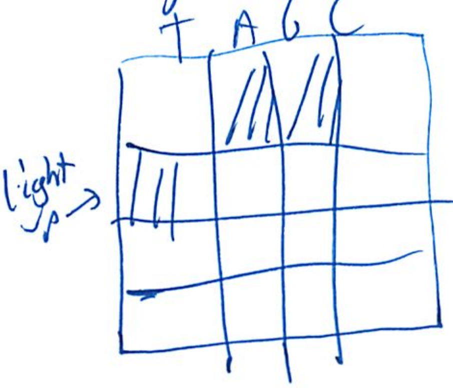


complement binds
rest of strand must match

Wash

So anything unbound washes away

PCR-
amplify
to high amts



A/G both will show up

T/T homozygous - still an SNP
since can vary against family

Could do 1 person per row

Or 1 SNP per row

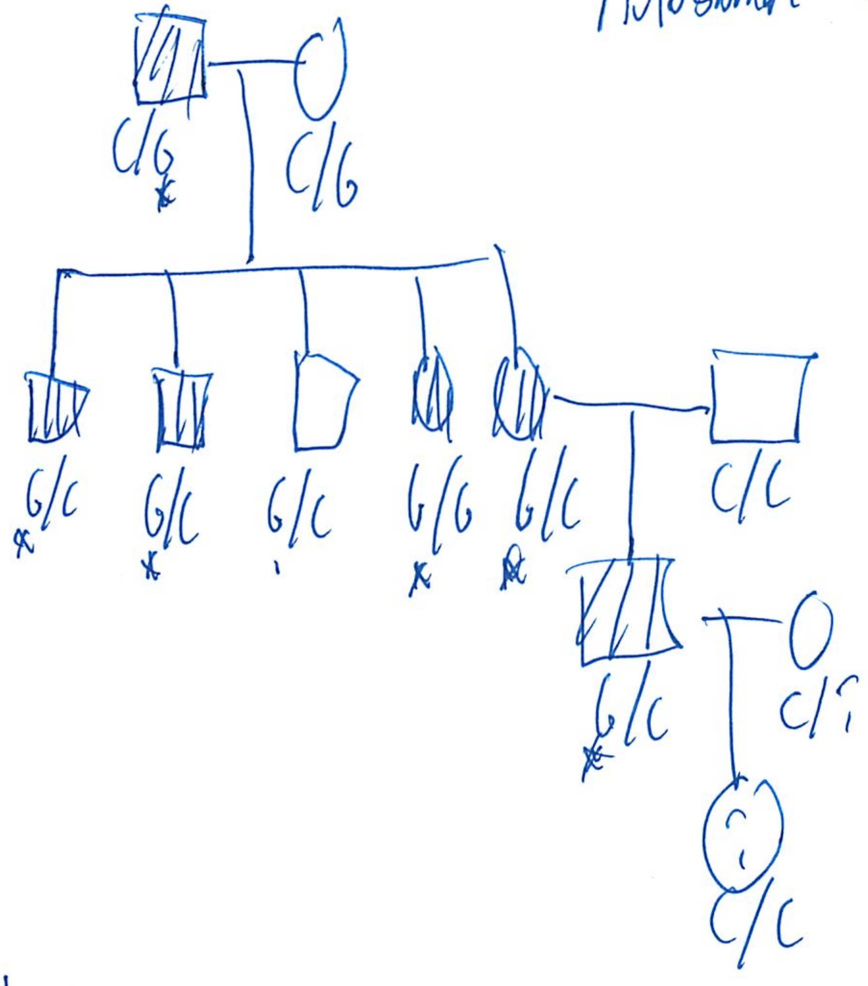
└ custom designed or out of catalog

23+me does a lot of SNP arrays
Certain ones associated w/ certain traits
Whole genome sequencing too expensive

Recitation 11 Q✓

#3

Autosomal Dominant


C/G
C/G
G/C
G/C
G/C
G/G
G/C
C/C
G/C
C/?
?
C/C

Look at top of tree
either C, G associated
know one G not associated
So one must be associated
C as well?
If homozygout → then all children would have

⑧

Whole point of SNP analysis is using tree

In real life → could have recombination

So person not affected → only C

For ○ we can't say anything
C/?

it's not linked to disease
but we don't know
can be any of the 4 bases (associated)
don't know her tree
must look at her parents

Could fill in letter
But don't know if can put *

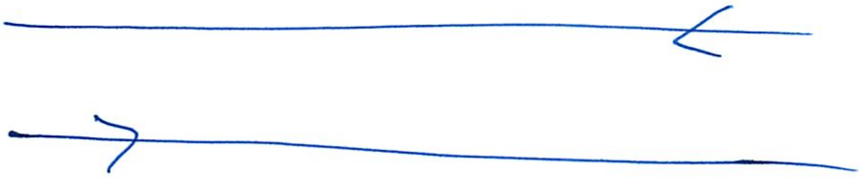
* does not change G → just that associated w/

Other familes may have diff things associated
Could or could not be pop associated

⑨ But class assumes could not

P-Set 4

#4 Primer must be 3'



Not the DNA itself

If have yeast DNA

#2c What does it mean to select / screen

1. If took up plasmid
2. If took up recombinant plasmid

Read 10/30

Samin Houshyar
R27 & R28

Review session: Tuesday 10-12 in regular recitation meeting room. Bring questions. Additional office hours Tuesday 12-2 36-372.

↓ go

I cannot emphasize the importance of doing ***many many many*** PROBLEMS in order to do well in this exam. Go back and redo your psets, old exams, OCW material, recitation problems, etc. Once you are done with those take the practice test timed. I have also posted additional practice problems under section materials.

I. DNA Replication
- A. Enzymes
  - 1) Helicase
  - 2) DNA Polymerase
  - 3) Ligase
  - 4) Topoisomerases
  - 5) Nucleases
  - 6) Telomerases
- B. In what direction does DNA polymerase work and WHY?
- C. What is the substrate for DNA polymerase?
- D. Leading vs. Lagging strands (continuous vs. discontinuous)
- E. Okazaki fragments
- F. Proof-reading

II. Transcription
- A. What are the components required for transcription?
  - 1) P__________
  - 2) R__________
  - 3) Transcription F_________
  - 4)
- B. Where does transcription start?
- C. Where does transcription stop?
- D. How is transcription controlled?
- E. In which direction does transcription go?
- F. mRNA processing
  - 1) What happens to introns and exons?
  - 2) What are other types of modifications?
  - 3) In what organisms do these modification take place?

III. Translation
- A. What are the components necessary for translation?
  - 1)
  - 2)
  - 3)
  - 4)
- B. What is the genetic code? What is a major property of the genetic code?
- C. Where do the products of translation go?
  - 1) Secreted. How?
  - 2) Membrane. How? Where are the N and C termini?
  - 3) Cytosol. How?
  - 4) Organelles

Samin Houshyar
R27 & R28

IV. Gene Regulation
  A. Operons
    1) What organisms?
    2) Components, know the functions of each
      a. Promoters
      b. Repressors
      c. Activators
      d. Operators
      e. Structural genes
    3) For these two operons know the following: ***signal, response, protein components, regulation, positive/negative feedback***
      a. Lac operon
      b. Trop operon
  B. Transcription Regulation
    1) Acetylation
    2) Methylation
      a. DNA
      b. Histones

V. Recombinant DNA
  A. Basic principles of cloning. What are we trying to achieve? Cut, Ligate, Transform & Select
    1) Restriction enzymes – where do they come from? What is their specificity?
      a. What are sticky ends?
    2) Restriction sites
    3) What are vectors? What are properties of vectors?
      a. Know antibiotic resistances: AmpR, KanR, etc. NOTE: neomycin is a mammalian antibiotic, not a bacterial one.
        a. B-galactosidase
        b. Lactose
        c. X-gal
  B. Gel electrophoresis – Know how to read and analyze gels
    1) Linear DNA cut once in the middle. How many bands do you see?
    2) Circular DNA cut once
    3) Linear DNA cut unequally
  C. PCR – What are the components? What is the basic principle?
    1) NTPs, dNTPs, ddNTPs. Which is which? Which ones do you use for PCR? Why?
    2) Cycles of PCR
      a. Denature
      b. Anneal
      c. Extend
    3) What special enzyme do we use in PCR? Why?
    4) What are the primers we use in PCR? DNA or RNA?
    5) How do you pick your primer ***sets***?
    6) How many strands are produced from one piece of DNA after n cycles?
  D. Sequencing – What are the components? Basic principle?
    1) Know how to read sequencing gels
    2) Know where to put 5' and 3' ends

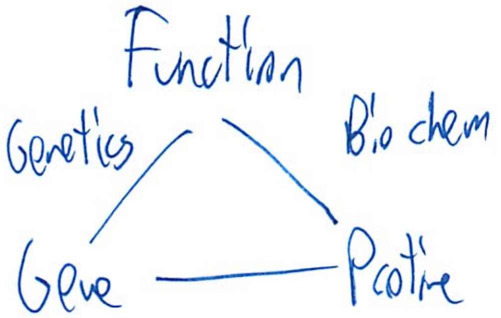
VI. Genomes
  A. Libraries. How do we make them? How do we read them?

    1) Genomic
    2) cDNA
B. Cloning by complementation
    1) What are you trying to accomplish in such an experiment?
    2) What is the experimental setup?
    3) When would you use a mutant or wild type organism?
    4) How do you establish the phenotype?

Bio Exam 2
Study

10/30

Molecular Bio 1 9/28


Function
Genetics
Biochem
Gene
Protine

Transforming Principle
↳ What causes heredity
→ today we know DNA

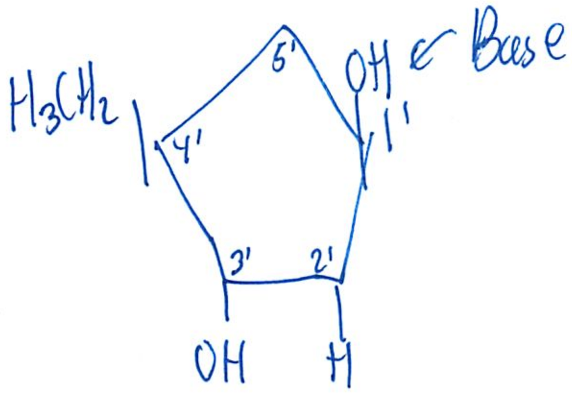
Giffits → ~~[illegible]~~ live nonvirulent + dead ~~(smooth~~ virulent
both individualy harmless
but harmful together
So state must be in non living
section

②

Structure of DNA

A C
T(U) G

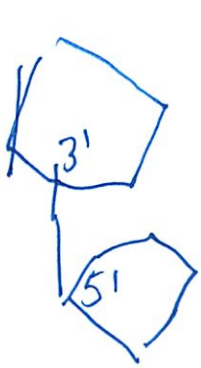
2' dioxy ribose


H3CH2
5'
OH ← Base
4'
1'
3'
2'
OH
H

Triphospate → good at energy

Sugar Phosphate backbone

3' → 5'



(didi blocks 3' addition)

③

Bacteria viruses → bacteria phage

Hershey -Chase → radio labled each

~~radia~~ ~~B~~ ^{32}P or ^{35}S

bacteria in pellet

do twice

See ~~that~~ which one was

bacteria radio active

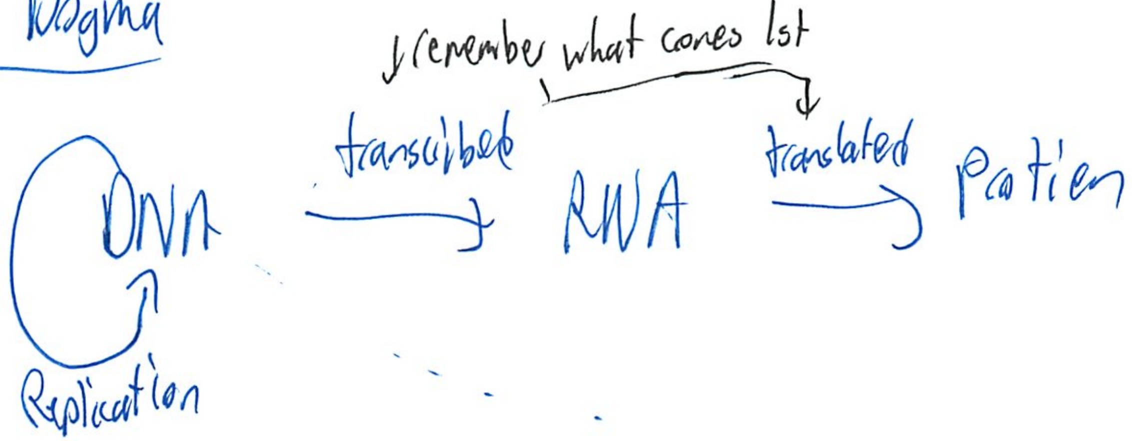
Double helix semi conservative replication

{ one strand from each parent

Watson + Crick

Meselson + Stahl

Central Dogma


remember what comes 1st
DNA
transcribed
RNA
translated
Protien
Replication

DNA polymerase → extends primer string

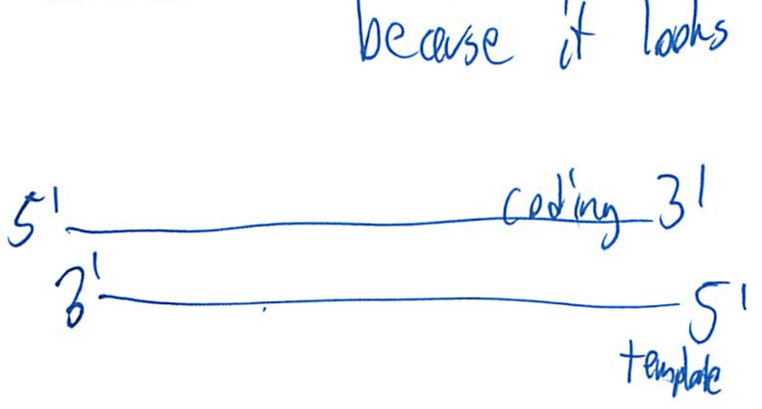
have free nucleotides pppA pppT
pppC pppG

↳ cleaves off 2 phosphates to join it on goes 5'→3'

Template vs Coding strand

template seq of DNA copied

Coding is the mRNA like strand
because it looks like it


5'
coding
3'
3'
5'
template

4b

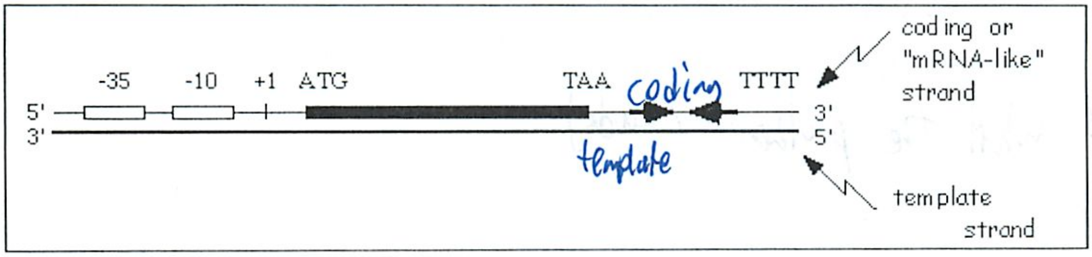
Template vs Coding Strands

It is often useful to distinguish the two strands of DNA -- the strand that is copied into mRNA and subsequently translated has the complementary sequence to the mRNA, while the base sequence of the opposite strand directly corresponds to the codons in the mRNA.

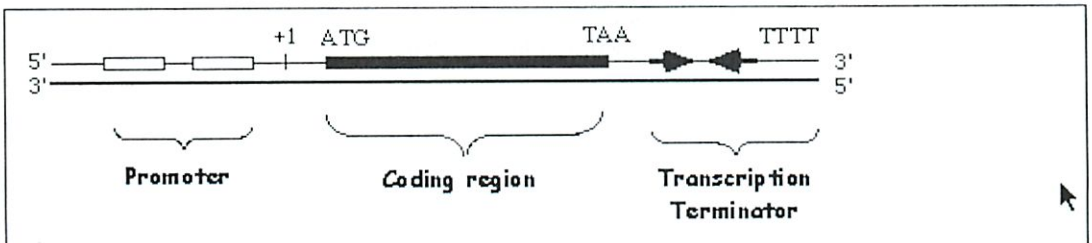
The terms template strand, sense strand, and coding strand are commonly used to describe one of the two strands of DNA, however the nomenclature is quite confusing because different authors have used these terms to describe both strands -- one school argues that the strand copied into mRNA should be considered the template strand, but the other school argues that the opposite strand which reflects the sequence in the mRNA should be considered the template because the corresponding codons are copied into protein. The first definition is used in the figures below, however, to avoid confusion, when using the words template, sense, or coding, it is essential to explicitly define how you are using the terms. I believe that these terms are best defined as described below.

The term **template strand** refers to the sequence of DNA that is copied during the synthesis of mRNA.

The opposite strand (that is, the strand with a base sequence directly corresponding to the mRNA sequence) is called the **coding strand** or the **mRNA-like strand** because the sequence corresponds to the codons that are translated into protein.

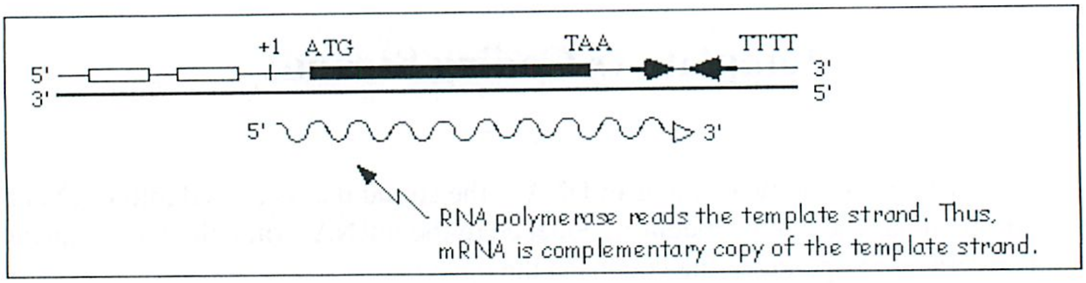


Although RNA polymerase must recognize sequences on the template strand, by convention we draw the DNA sequence and regulatory signals on the "mRNA-like" strand. (This makes it simpler to directly determine the sequence of the resulting RNA.) The following cartoon shows this concept for a hypothetical gene.



Note that features of DNA are encoded in the template strand but usually shown on the corresponding position on the "mRNA-like" strand.

Read from 3' → 5' on template



RNA polymerase reads the template strand. Thus, mRNA is complementary copy of the template strand.

It may be useful to consider a real gene as well. The DNA sequence of the phage P22 *arc* gene and some important regulatory sites is shown below. The upper strand of DNA is the "mRNA-like" strand. The lower strand is the strand that is complementary to the mRNA. The -35 region (TTGACA) and -10 region (TATATT) of the promoter sequence and the transcriptional start site (the **A**) is indicated on the coding strand. Also note that the DNA sequence of the coding strand corresponding to the RNA codons is shown in bold (of course, the T is a U in the RNA) -- the first codon is **ATG** the translational start site (fMet) and the last codon is **TAA** (Ochre) the translational stop codon.

```
                 -35                   -10            +1
5'  AAGTTAGTGT ATTGACATGA TAGAAGCACT CTACTATATT CTCAATAGGT CCACGGTGGA
3'  TTCAATCACA TAACTGTACT ATCTTCGTGA GATGATATAA GAGTTATCCA GGTGCCACCT

    CCTGTATTGT GAGGTGAATA TGAAAGGAAT GAGCAAAATG CCGCAGTTCA ATTTGCGGTG
    GGACATAACA CTCCACTTAT ACTTTCCTTA CTCGTTTTAC GGCGTCAAGT TAAACGCCAC

    GCCTAGAGAA GTATTGGATT TGGTACGCAA GGTAGCGGAA GAGAATGGTC GGTCTGTTAA
    CGGATCTCTT CATAACCTAA ACCATGCGTT CCATCGCCTT CTCTTACCAG CCAGACAATT

    TTCTGAGATT TATCAGCGAG TAATGGAAAG CTTTAAGAAG GAAGGGCGCA TTGGCGCGTA
    AAGACTCTAA ATAGTCGCTC ATTACCTTTC GAAATTCTTC CTTCCCGCGT AACCGCGCAT

    AAGTTGAAGC 3'
    TTCAACTTCG 5'
```

note ATG on coding strand (since that is what the protien reads)

References:

- Arc DNA sequence from Genbank
- King, R., and W. Stansfield. 1985. A dictionary of genetics. Oxford University Press, NY.

Return to Microbial Genetics supplement.

Please send comments, suggestions, or questions to smaloy@sciences.sdsu.edu
Last modified July 12, 2002

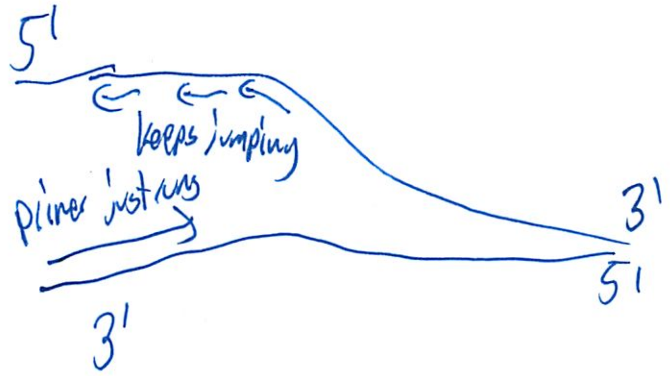
⑤

RNA polymerase recognizes on template strand, by convention draw DNA seq + regulator signals on MRNA like strand

See sheet!

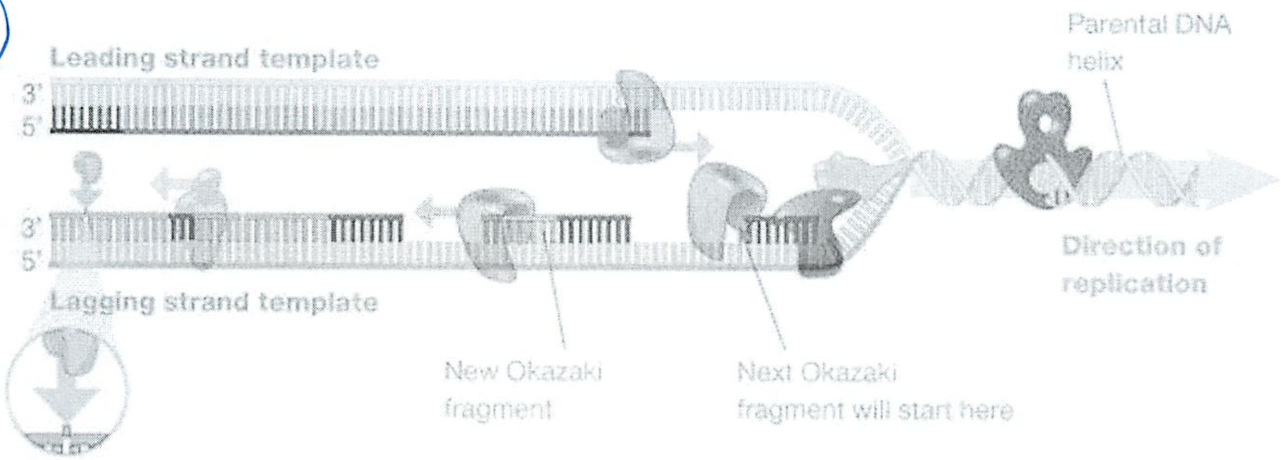
* So template 3' → 5'
coding 5' → 3'

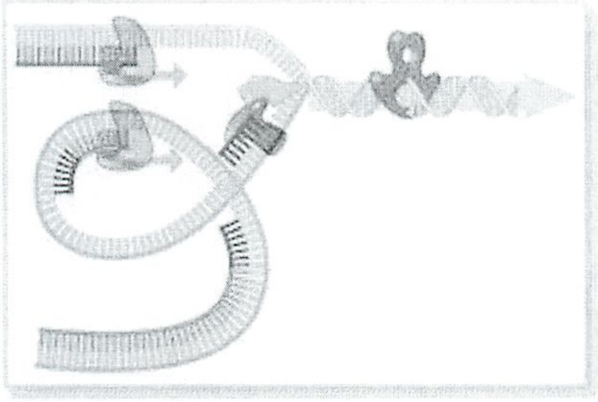
this is what can be confusing!

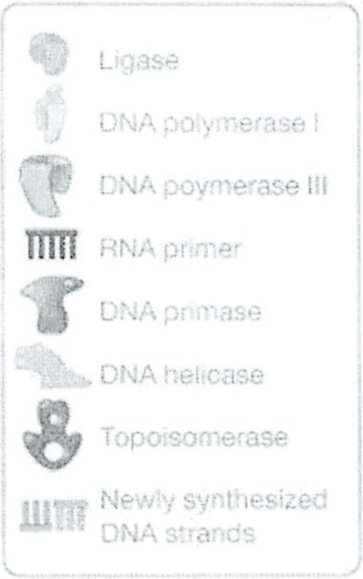

5'
keeps jumping
primer just runs
3'
5'
3'

ligate = ~~thet~~ ties together

⑥

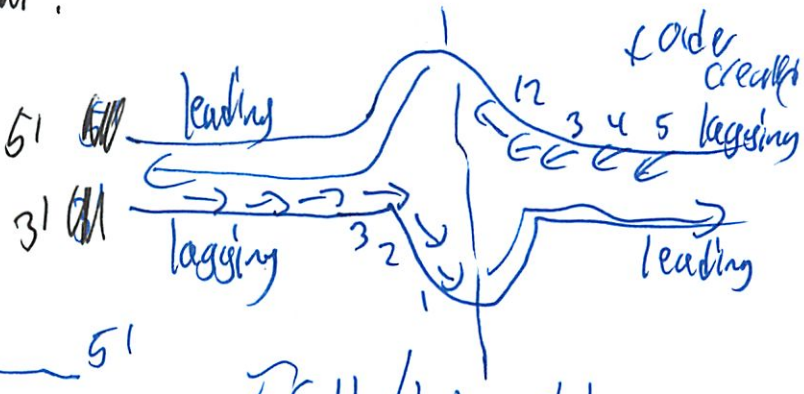

Leading strand template
3'
5'
3'
5'
Lagging strand template
Parental DNA helix
Direction of replication
New Okazaki fragment
Next Okazaki fragment will start here




Ligase
DNA polymerase I
DNA poymerase III
RNA primer
DNA primase
DNA helicase
Topoisomerase
Newly synthesized DNA strands

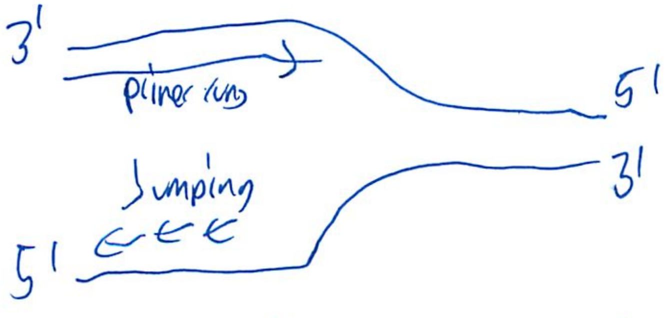
Upside down from how drawn in notes

∴ I hate that!


5'
leading
3'
lagging
3 2
1
12 3 4 5
order created
lagging
leading

↗ full / both sides

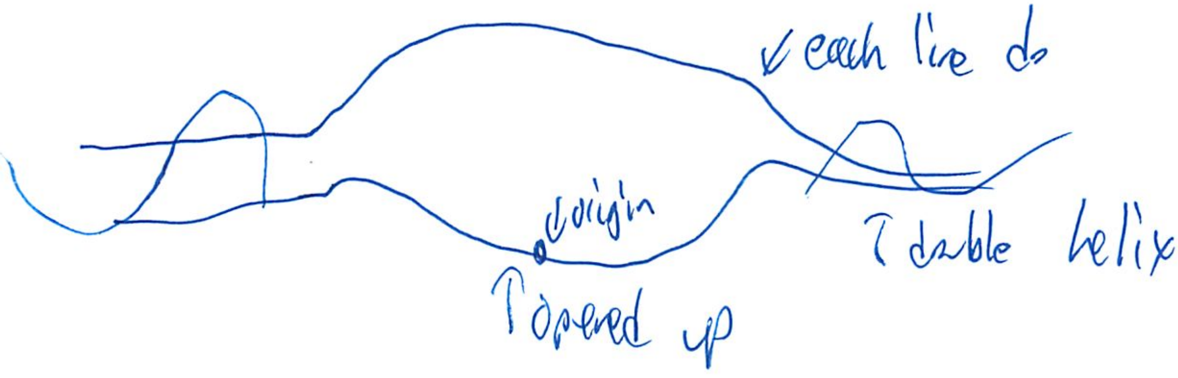
So


3'
primer runs
5'
3'
jumping
5'

Okazaki fragments = 1000-2000 nucleotides

⑦

topoisomerase - cuts the DNA open + closes it back up


each line do
origin
double helix
opened up

Fidelty

mistake every 10^3

but that's a lot

need to be able to fix

proofreading can run backwards + take out bases

$1/10^6$ mistakes

Plus some other enzyme checks

So $1/10^8$ errors

⑧

Transciption

DNA → RNA

gene = region of chromosome

¿What is it exactly?

codes for a polypeptide or RNA chain
that has a function in an organism

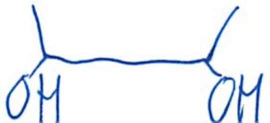
Corresponds w/ unit of inheritance

allele = one variant of
└ red hair allele

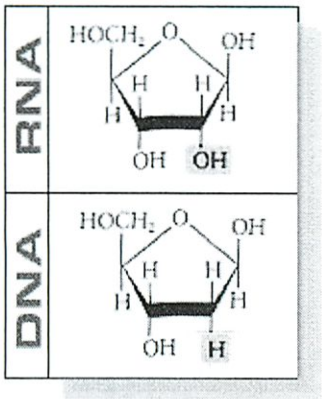
So defn is more about heridity impact
than a # of base pairs...

RNA Sugar is ribose, not deoxyribose

So more reactive C-HO vs C-H bonds
└ aka less stable

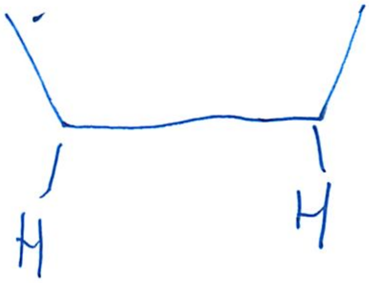


⑨


RNA
HOCH₂
O
OH
H
H
H
H
OH
OH
DNA
HOCH₂
O
OH
H
H
H
H
OH
H

OH = hydroxyl

di - DNA



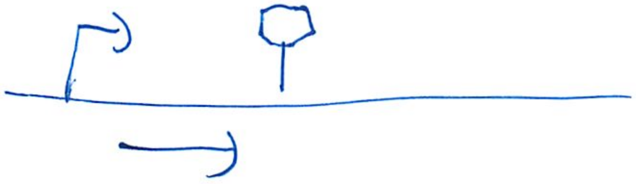
Chain can't extend

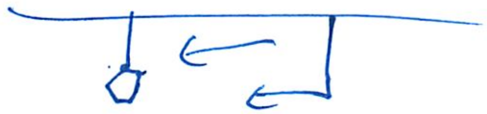
⑩

* Only single strand copying

RNA polymerase

both directions possible





↳ but only 1 at a time

? I think this is actually where coding/template strand matters

↳ or really do both work?

Helicase – opens DNA at origin of replication

A-T only 2 bonds → easier to break

(prob should read qu ~)

Do I locally optimize on q/

Or try to get really broad understanding

so can problem solve?

mRNA → RNA created from transcription

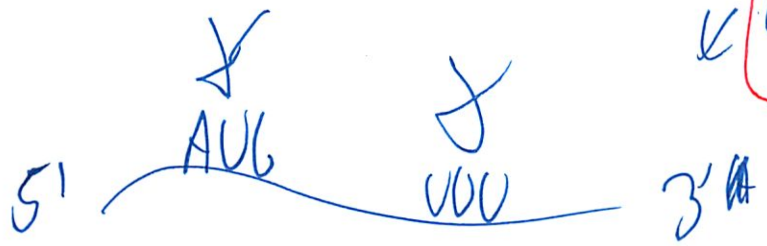
Translation

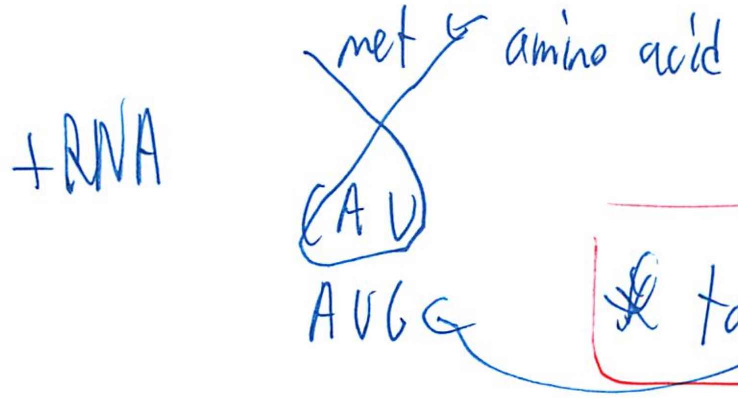
Francis Crick's look up table

3 base pairs each

We have our RNA

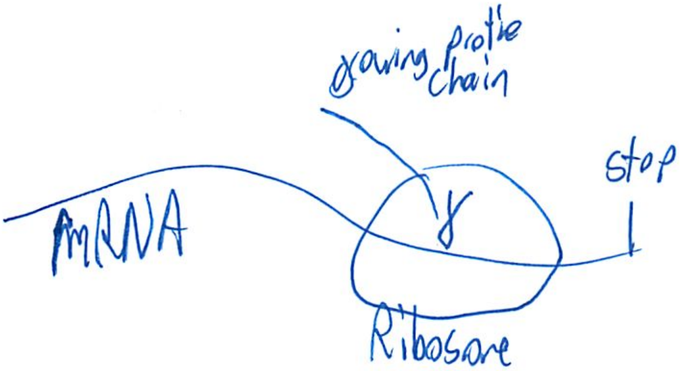
✓ Read 5' → 3' coding/mRNA strand




met ← amino acid
tRNA
CAU
AUG

✱ table based off of

factory called a **ribosome**


growing protie chain
stop
mRNA
Ribosome

Can have multiple on RNA

Variations

Eukarots

of ~~gene~~ DNA varies tremendously
but all linear

end → **telomere** (TTA GGG)$_{\text{repeats}}$

telomerase adds the telomere that got cut off at end

Prokarots Circular!

Assay → method to determine potency of bio activity

rRNA – ribosomal RNA

Combines w/ protine to form ribosome)

coding strand (non template) → ~~not~~ used in transcription

promoter specific area of DNA where RNA binds to

transcription factors = bind to promoter

(I am still unclear on the specifics of
−20
−10
+1
etc

entry of tRNA req hydrolysis of 1 GTP molecule

└ source of energy

(so requires energy?

(14)

TFIIB　　TATA　　Initiator　　Downstream Core Promoter

-35　　-25　　+1　　+30

alth TATA box is just a bunch of Ts + As
not specifically TATA

generally 5' TATAAA 3'

but could be a varient

~+25

Old material OH review (for bonding)

O, N most electroneg

ionic, H bond - polar

Phosophate group ⊕ charged to ionic

H-C not polar
OH polar
NH polar

(15)

ionic: polar + w/ polar −

hydrogen: polar w/ polar charged or uncharged

H must be electro neg

(still should rethink a lot for final)

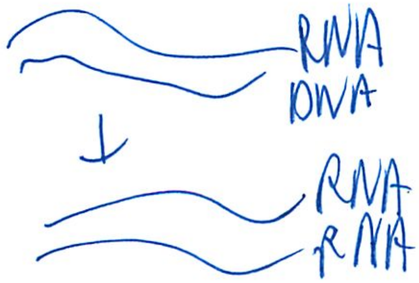
Viruses can be linear or circular

Can have RNA viruses

- make complementary strand
- then make template strand

(this is non core - should prob skip)

Reverse transcriptase make DNA from RNA



} : did we mean DNA?

Poly A tails $(AAA)_{30-500}$ often

Sometimes bits spliced together

exons parts retaind

introns parts spliced out

a ~~single gene mi~~

often multiple ways a gene can be spliced

Viruses can have diff reading frames

lots can go wrong

DNA repair

DNA polymerase recognizes kink in chain

⑰

Control of Transcription

Which genes turned on/off

Trp

repressible operons

↘

repressed when protine bound by molecule

So the presence of Trp inhibits the creation of more

Lac operon

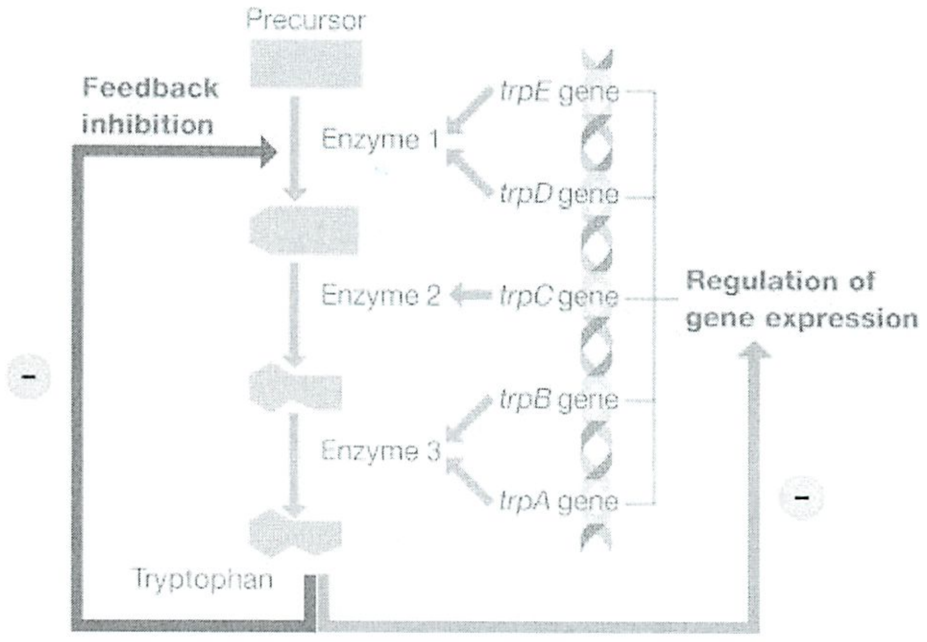
Inducible operons

↓

repressed when repressor **not** bound to specific molecule

18

Trp

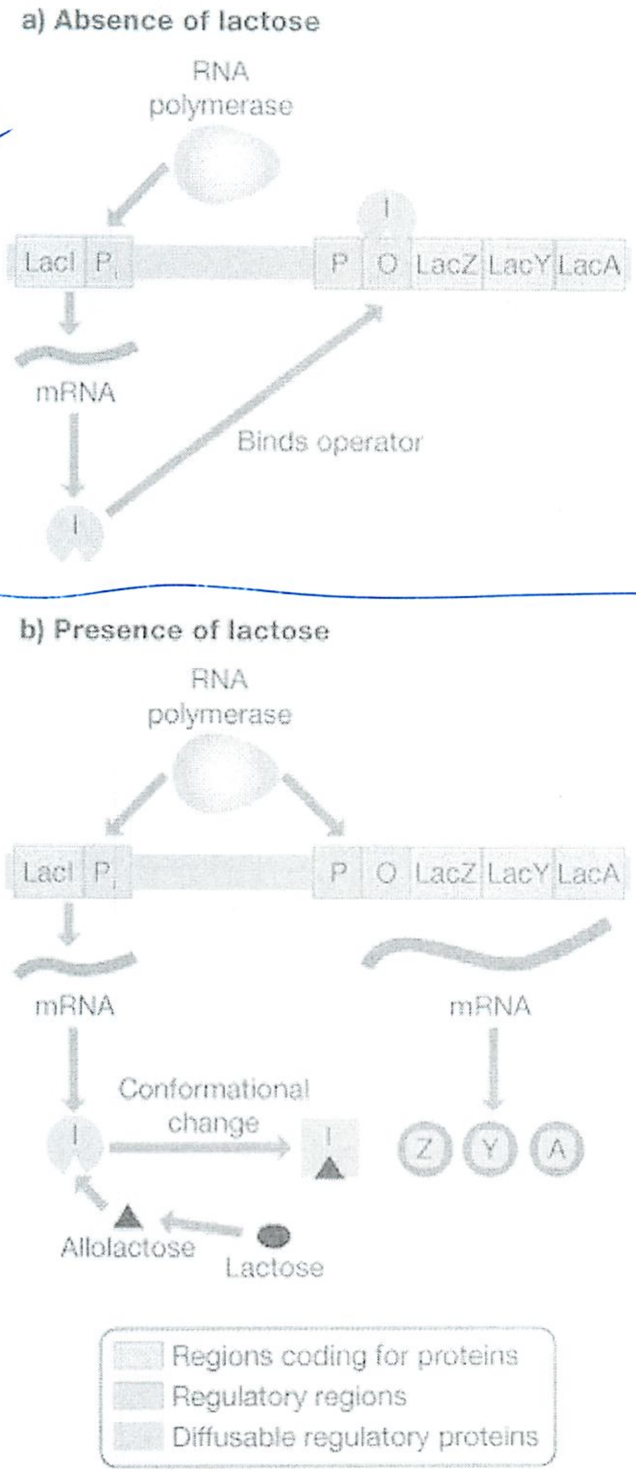

Precursor
Feedback inhibition
Enzyme 1
trpE gene
trpD gene
Enzyme 2
trpC gene
Regulation of gene expression
trpB gene
Enzyme 3
trpA gene
-
-
Tryptophan

repressible

19

Lac

inducable


a) Absence of lactose
RNA polymerase
I
LacI
P
P
O
LacZ
LacY
LacA
mRNA
Binds operator
I
b) Presence of lactose
RNA polymerase
LacI
P
P
O
LacZ
LacY
LacA
mRNA
mRNA
Conformational change
I
I
Z
Y
A
Allolactose
Lactose
Regions coding for proteins
Regulatory regions
Diffusable regulatory proteins

presence of lactose prevents it from binding

⑧

this is how eukarots operate themself

---

mRNA limited life span

---

heterocromatin - when condensed

eucromatin - dispersed
when being transcribed

- Pol I works in nucleous cranking out lots of ~~[illegible]~~ rRNA

rRNA polymerase I

Coils into Chromatin

individual nucleosomes

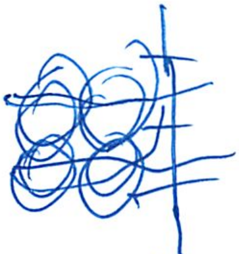
(this lecture was bad)

㉑

Coils around nuclesomes

8 core histone molecules + a 9th

↑ 1 core histone



↑ H_1 histone

nucleosome

2x H2A
2x H2B
2x H3
2x H4

tails of histones hanging out

N-termi tails modified

Some allow/disallow transcription

methylation ↑ binding of ~~DNA~~ ~~Hb~~ hica histones to DNA

↓ ~~M~~ avability of DNA for transcription

22

Several gene expression mechanisms involve chromatin structure.

When eukaryotic cells aren't dividing, chromosomes exist in an uncondensed state called chromatin. Chromatin consists of DNA wrapped around a **histone protein** core. The wrapped DNA isn't as available for transcription as the DNA of prokaryotes, and as we'll discuss, mechanisms exist to relieve this repression. Also in eukaryotes, the RNA polymerase doesn't bind directly to the DNA, but instead binds via a set of proteins: the transcription initiation complex.

Two different types of chromatin can be seen during interphase: euchromatin and heterochromatin. **Euchromatin**, which is a lightly packed form, contains areas of DNA that are undergoing active gene transcription. Not all of the euchromatin is undergoing gene transcription, however. **Heterochromatin**, in contrast, is mostly inactive DNA that is being actively inhibited or repressed in a region-specific manner. The chromatin state can change in response to cellular signals and gene activity. This is facilitated by enzymes that modify histones by adding methyl and acetyl groups to their N-terminal tails. **Acetylation** reduces the net positive charge of the histones, loosening their affinity for DNA, and increasing transcription factor binding. **Methylation**, in contrast, leads to increased binding of histones to DNA, and decreases the availability of DNA for transcription. Figure 2 shows an example of how acetylation and methylation of histones may affect transcriptional activity in a normal cell compared to a cancer cell.

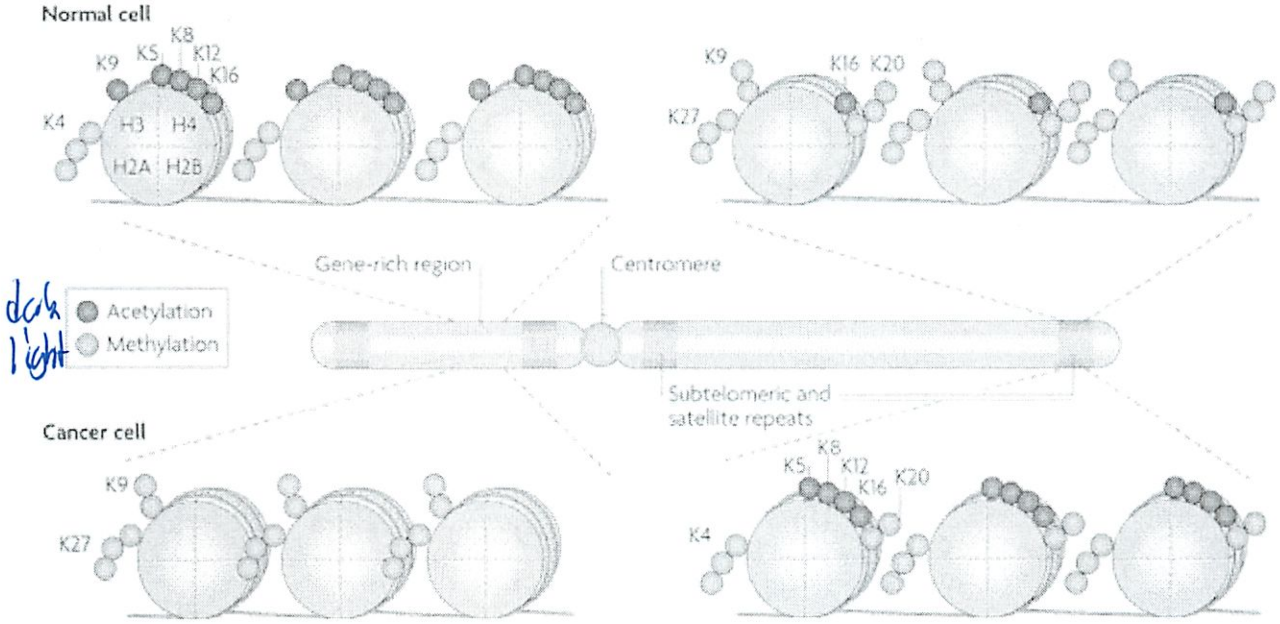


Figure 2: Modifications of methyl and acetyl groups in histones affect transcriptional activity. The grey cylinders represent histone octamers. Acetylation (blue circles) and methylation (green circles) of histone subunits are shown. In normal cells, the promoters of tumor-suppressor genes show acetylation of histone subunits, associated with active transcription. In contrast, in cancer cells, the promoters of tumor-suppressor genes are not acetylated, and they are not actively transcribed. In normal cells, the heterochromatic regions at the ends of the chromosomes do not show acetylation, and the genes are not actively transcribed. In cancer cells, the heterochromatic regions at chromosome ends are acetylated and transcriptionally active.

dark

light

㉓

Acetylation ↓ net ⊕ charge of histones
↓ afinity for DNA
so ↑ DNA avilability

Remember → menomic ⌣̈
Ace is lucky, ↑ trasciption
Meth is bad ↓ trancsiption

Control elements

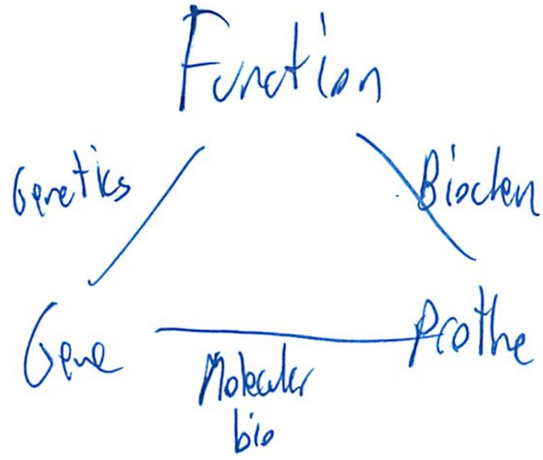
enhancers - hundreds of
activators bind to enhansers
a specific activator for a specific enhanser

Since fold over can control even though 1000s of base pairs away
↳ w/ DNA bending protine
and enhanced by mediator protines

㉔

It's complicated

Recombinant DNA


Function
Genetics
Biochem
Gene
Protein
Molecular bio

What DNA represents what thing?

Cloning pieces of DNA
find, purify, propagate

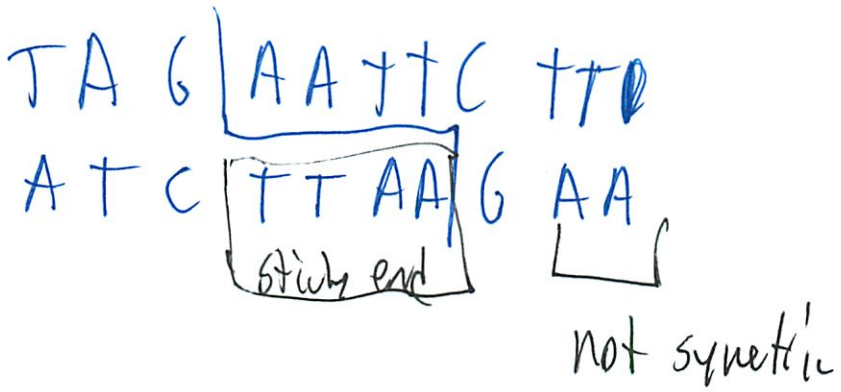
1. Cut up at defined site
   └ with or restriction enzymes?
2. Paste into vectors that propagate
3. Put into cell

4. Transform

forget this again

5. Select for cells w/ that DNA

Cut up
↳ w/ restriction enzyme


TAG AATTC TTC
ATC TTAA G AA
sticky end
not synetic

Paste in DNA

Use DNA Ligase
w/ open Gap

Vectors

⌶

⌶ remember this section now

have origin of replication

antibiotic restance gene - so can pick the
bacteria w/ any plasmid

learned from Jeff ☺

Transforming

get bacteria to take up the DNA

Select for presence of vector w/ bacteria

then only vectors (~~that~~ which have the
antibiotic restance gene) left

Have a library

if lots of them whole human genome
there

like millions or billions of them!

㉗

Reverse transcriptase
↳ turn RNA into DNA
called cDNA

---

if cuts in middle
↳ use diff restriction enzyme
or add right ratio of methylation

or cuts from 2 enzymes so ends incompatable
or take off phosphates
↳ how does this work?

How to find gene in lib?
find the one that grows on minimal media
Called cloning by complentation

㉘

Start w/ wild type + mutant cells

don't get exactly

appears to be that arg^- mutant
↑ {one of many from lib
+ wild type
grows in minimal medium
where it didn't before

(not ~~worth~~ written up well anywhere)

Put wild type DNA into mutant cells
↑ one specific mutant?

Works well when recessive autosomal
So only need 1 good type

---

cDNA library → pre computed

(very confused
This is the recitation I skipped as well..)

Recitation handout : Cloning by Complementation

Way to identify gene thats mutated
in a mutant you have isolated
↑ so have a mutant already

have Arg mutant
(so randomly we just get one of these
- and we are wondering what's broken w/ it
?... is that right?

have a wild-type yeast library ~6000
each w/ ~~6000~~ diff plasmid + diff gene

each cell gets 1 plasmid

One mutant cell will be fixed w/ wild type DNA
like the Griffiths mice

So that is the one that can grow
in cell w/o arginine
?(ah ok I didn't get the problem we were trying to solve)

(30)

Recombinant DNA Application allows biologists to change (+/-) genes from genome

? In what context did we talk about this in class?

DNA sequencing

1% defective

Stop 1% of the time

dNTPs

which I called diDNA earlier

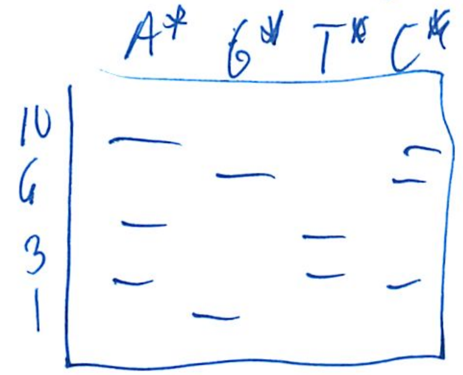
polymers extend 5' 3'

3' A TATG U 5'

5 T A T*

TATAC* etc

Then measure lenght w/ gello agerose



Or flurcent die in 1 column
Scan w/ lsaser scanner

---

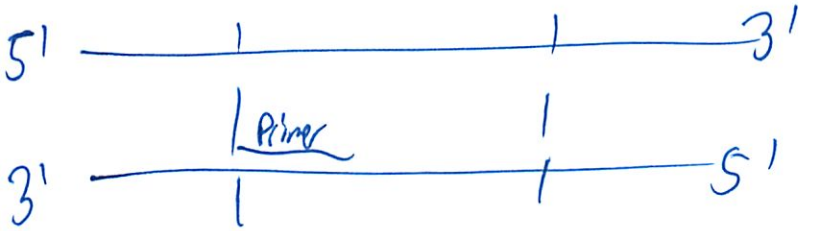
bt what if longer DNA
randomly sheer
look for over lap

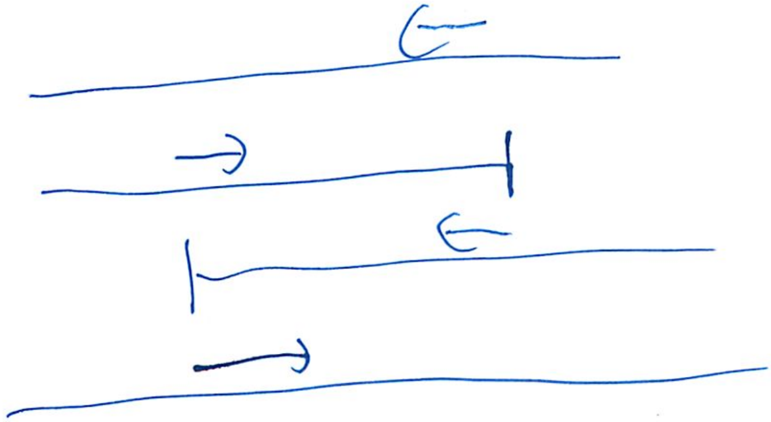
translate to prothen + search the db
(I hae feeling thae is so. much to discover
here)
Llike for science

Primer bounces around till it finds the right spot w/ H bonds

Polymerase Chain Reaction (PCR)

- repeat + repeat
  get 2^x of them





(don't get when this happens)

㉝

1. **denature** - split double stranded DNA w/ heat
2. **annealing** - primers bind to DNA give starting pt
3. **extension** - Taq polymerase extends seq

modern DNA sequencing

↳ camera

microarrays lots of species specific DNA seq

heat map

?what is the difference?

Recombinant DNA allows Gene $\rightleftarrows$ Protein

how do we find the function if can't do

Clone by complementation

Human Genome Project

look for repeat patterns

bt lots of repeated DNA

viruses have actually slammed their way into your genome

hirarchical mapping

annotate where genes are

∟most efficent in cDNA lib

looking for linkages/map

set up a cross

see which allels are inheritted correlated

(I don't get this...)

Some recombinations

produce linkage map of disease

(B.5)

Gene + Function

↑ (can knock out w/ mutase

Can target gene w/ fertilized egg
but ligates at random
Signal target gene
w/ embryonic stem cells

progeny = offspring
also brood
(can also mean students)

Then target/knock out specific gene
Usually goes somewhere
but sometimes recomb. is active

36

Small prob its in the ight place

So need to screen for and select

don't get ~~all~~ at all what we are trying to do

Instead add negitive selection marker nearby

any cell w/ this dies

(was never covered in Recitation)

New version of that

restriction enzyme that only cuts at 1 place

need to target 1 spot

~~dymer~~ dimer?

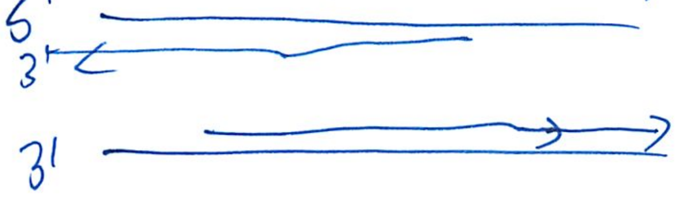
attach a targeting domain

Tal affectors → plant patagens that target seq

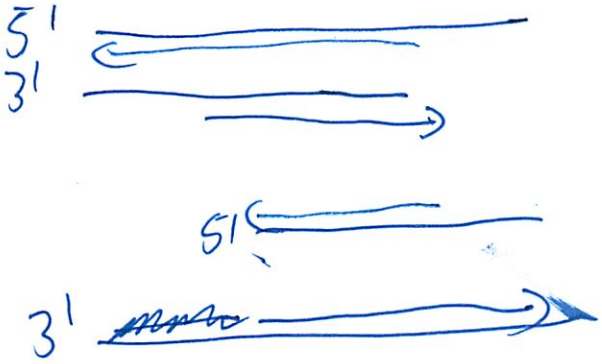
This is not in book, recitation, or online!

PCR steps

primers base pair match



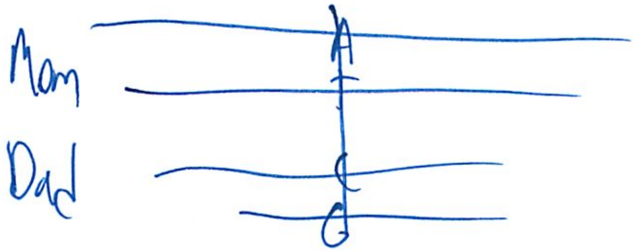
how/why?



SNP analysis

Single nvcleot'ide poly mortasm

have a strad from both Mom + Dad


Mom
Dad
A
T
C
G

(38)

Don't know where gene is
or pattern of inheritance

but have family tree

See what is assaiated/correlated w/ disease

Can set up SNP Array
Chip w/ many spot
the seq will be the same except 1 place w/ SNP
flurecent if binds

(So is that the same as was in lecture?

Skip learning about Microarrays

(watching lecture)

So back to lecture ; how to find genes

— open reading frames → ATG till stop

So I get we are basically looking for correlation

Affected vs not affected siblings

Sterteromore — 1911 Map linkage

looking for recombination

which inherited correlated

└ thus link

— as opposed to 50-50 random

; not perfect since cross-over right ;

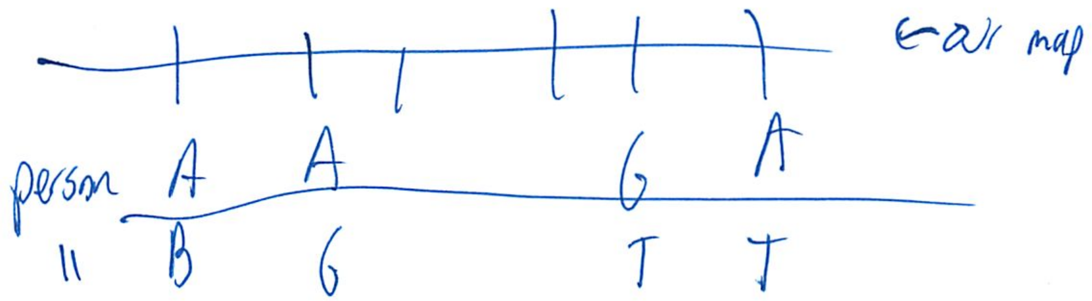
Look for something correlated w/ it

but can't set up crosses w/ humans

So look at family trees

look at a bunch of genetic marker

but need something simple → look at identical twins



Mom + dad -> 3 million spelling diff

These are genetic traits

Can use DNA markers

↳ just like Cinnabar eyes

(but how can we read?

Can seq I guess

Our problem here is function ↔ gene)

So look for correlation

known is close

(so this is SNP — but he never called it that!)

look at lots of spelling diff

---

in real life many things not single genes

(41)

Jed: Array of diff SNP checking

Try to find correlation

Looking for sig differences

Basically just ass

Gen → Fn

Jed: I guess its something about how to find out what fn a gene expresses

Somehow make DNA

Need to target to specific genes

random DNA usually goes to random location

but can se ^ recombination

homogelogous

looks for matches b/w seq *

㊷

* not random but based on seq similarity

Genetic recomb will occur

Grow cell in antibiotic w/ ⊕ marker

So only cells w/ markers grow

but that just gets it somewhere

Put a ⊖ selection marker

any cell w/ that dies

extreamly nearby

if randomly slammed in

Only ones that have my DNA smack in the middle

Only ones w/ homologous recombin

Then grows mouse that has that

And I assume then look what is wrong w/ the mouse

(43)

New targeting (not in book)

string together domains ~~th~~ that bind that seq

+ attach to restricon enzyme → cuts at that seq

(i.e. so can make any arbitrary targeter much easier?

(44)

Questions I have

10/30

Why ~~DNA~~ PCR symetrical ?

-20 -10 +1 etc

tRNA req energy?

TATA box specifics?

Reverse transcriptase

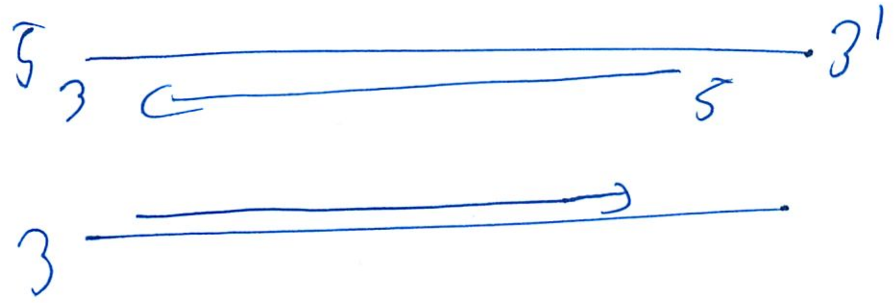
Restriction, enzyme taking off phosphate

microarrays + modern DNA seq

㊺

Ask Jeff

10/30



diff primers

Just coincidence

Make same - but opposite seq

Complement

TATA box

beacon

Anything w/ lots of Ts + As

Pre transcription

+1

One start at

Usually give line

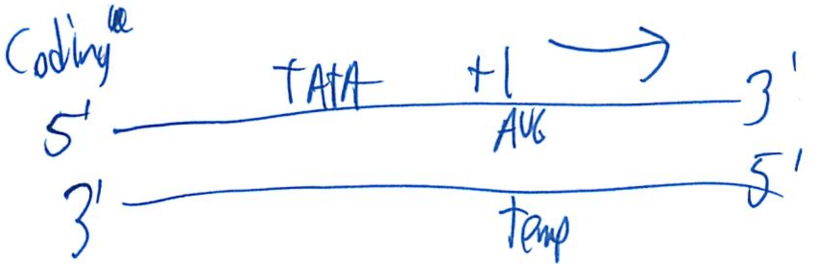
but look for AUG/ATG

(4.6)

+1 is always A?

└ 90% sure

AUG on ~~template~~ strand
Coding



Coding
TATA
+1
3'
5'
AUG
3'
5'
temp

-35, -10 could be site of transcription factor

10/30

Bio 7.012
FALL 2012
EXAM 2 REVIEW
Material taken from MIT OCW

Question 1

Transcription

i. is the process that transfers information from DNA ✓ to mRNA ✓.

ii. in eukaryotic organisms, transcription occurs in the

[✓ Nucleus] ✓ Ribosome Membrane

Translation

i. is the process that transfers information from RNA ✓ to protiens ✓.

ii. in eukaryotic organisms, translation occurs in the

Nucleus [✓ Ribosome] ✓ Membrane

Question 2

The following sequence of DNA encodes a hypothetical polypeptide called Playdo in a hypothetical bacteria *E. hypotheticus*. Transcription starts at and includes the C/G base pair in bold. The underlined T/A base pair indicates the terminator.

write

5'-TTCCCCTATGGATGGTCATCTACGATGCCCCCATCACTAAAGCTTG-3'
3'-AAGGGGATACCTACCAGTAGATGCTACGGGGGTAGTGATTTCGAAC-5'

read

c) What are the first 6 bases of the transcribed RNA? Be sure to label the 5′ and 3′ ends.

5' CCC CUA 3'

d) What are the first 3 amino acids of the Gir subsequent polypeptide? Be sure to label the N- and C- termini.

forget which is which

N Pro Leu Trp C

don't get

Met Asp Gly

We get the table right?

e) How many total amino acids are encoded in this polypeptide?

~~14~~ 10

oh starts at ATG

So ATG ~~ATC TAC CGAT~~

GTC ATC

T include AUG

Met Asp

Jeff has no clue why wrong

A-T(U)
C-G

You identify a strain of bacteria containing a mutant tRNA that is capable of adding a tryptophan residue when it recognizes the codon UAG in the mRNA.

trp

f) What is the sequence of the anticodon of the mutant tRNA? Be sure to label the 5′ and 3′ ends.

3′ AUC 5′ ✓
5′ UAG 3′

g) The Playdo polypeptide would be

longer
[✓ the same length] ✓
shorter

in the presence of the mutant tRNA.

Why? never encounters UAG ✓

Intrigued by Playdo, you search for a similar protein in mice by looking for similar DNA sequences in the mouse genome. You find a gene that matches bacterial Playdo almost perfectly but contains a 36 DNA base pair insertion in the center of it.

When you purify the Playdo polypetide from mouse cells you are shocked to learn that mouse Playdo is the same length in amino acids as bacterial Playdo.

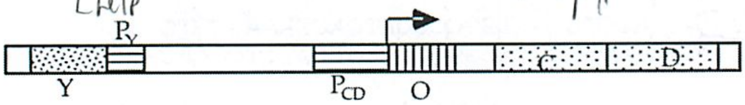
j) Explain how is it possible that mouse Playdo and bacterial Playdo are the same polypeptide length even though they have substantially different gene lengths.

exons + introns ✓
(removed)

Question 3

A scientist has discovered a bacterium that can metabolize alcohol to pyruvate when stimulated by the presence of ethanol. Hoping to use this bacteria to eliminate the harmful effects of excessive alcohol consumption, he finds that this bacterium produces the enzymes C and D which catalyze the ethanol → pyruvate reactions. The genes for enzymes C and D are part of an operon as diagrammed below. ↑; alcohol → pyr ;

L help



a) You find two mutants, mutant 1 and mutant 2 that each carry a loss-of-function mutation in one of the components of the operon. These mutants constitutively produce enzymes C and D. Given this information, Is Y an activator or repressor? Explain.

repressor
~~Activator~~, – when loses function it produces C + D
(but confused)

loss of components (Y + O) allow transcription wether or not ethanol is present so Y is a repressor

don't get – Oh I see | loss of Y allows C productio

↙ like the lac examp

You want to determine what component is missing in each of the mutants. You examine production of enzymes C and D in wild-type, mutant 1 and mutant 2 in the presence and absence of ethanol. You also create partial diploids and measure production of the enzymes.

cell type	No ethanol Production of C and D	+ ethanol Production of C and D
Wild type	low	high
Mutant 1	high	high
Mutant 2	high	high
Mutant 1 + $Y^+P_Y^+P_{CD}^+O^+$	high	high
Mutant 2 + $Y^+P_Y^+P_{CD}^+O^+$	low	high

(→

b) Mutant 1 has a loss-of function mutation in which component of the operon? Explain.

~~The repressor code~~ Y, O actually wrong

Operator O - repressor can't bind

Plus can't add seperatly

((

sure

c) Mutant 2 has a loss-of function mutation in which component of the operon? Explain.

~~The repressor~~ Coding for creating Y O

For repressor Y or Promoter of repressor P_Y

So Y is not made

works if manually added

I see now

Question 4

a) The genes needed to synthesize tryptophan are found in the *trp* operon. The presence of tryptophan controls the transciption of these genes.

i) Assume that the regulatory protein for this operon is a repressor. When do you expect the repressor to be associated with the operator? Cirlce one.

When tryptophan concentration is low. [✓ When tryptophan concentration is high.]

Explain your choice.

inhibit enzym production if enough trp is present

ii) Assume that the regulatory protein for this operon is an activator. When do you expect the activator to be associated with the operator? Cirlce one.

[✓ When tryptophan concentration is low.] When tryptophan concentration is high.

Explain your choice.

Would produce enz to make try if trp is low

Question 5

So this is straight forward

(not repressible/inducible)

Since notes; when lac low → ~~#~~ repressor binds to operator

↓ not the website :)

On your trip to the Amazon you were introduced to a new plant that the indigenous people use as an anti-viral treatment. You took samples home to your lab and you found that this plant makes a protein (the PV protein) that prevents viral replication. Excited by the possible anti-AIDS applications, you construct a genomic DNA library from this plant in the hope of cloning the *PV* gene.

a) What is a genomic DNA library?

a collection of fragments ~~from~~ of ~~all~~ DNA of plant in bacteria cells

b) You get a DNA sample from cells, digest it with a restriction enzyme, and clone it into a vector. List 3 features of the vector that are absolutely required for your library construction.

restriction enzyme, binding site for enzyme ✓

origin ✓ ~~anti body resistance~~ Cloning site / selectable marker

c) Circle on the following lists ALL you would need in order to construct the genomic DNA library. Assume you start with intact plant genomic DNA.

Enzymes	Reagents
[✓ Restriction enzyme]	Size separation gel
[✓ Ligase]	Okasaki fragments
DNA Polymerase	ATP, [✓ TTP, CTP], GTP
[✓ RNA Polymerase]	ddATP, ddTTP, ddCTP, ddGTP
Transcriptase	[✓ Primers]
Reverse Transcriptase	Replication fork
3' to 5' exonuclease	[✓ *E. coli* (bacteria)] ← acts as a host that receives plasmid
[✓ Cloning vector]	Human cells
	Virus

Briefly describe the function of each item circled.

? what do we need each of these

? are we trying to seq?

Question 6

↑

took it too far too far

↳ but I think I am confused on all the steps

making the lib vs sequencing it

vs cloning by complementation

Restriction enzymes are extensively used in molecular biology. Below are the recognition sites of two of these enzymes, BamHI and BclI.

a) BamHI, cleaves after the first G:

5’ GGATCC 3’
3’ CCTAGG 5’

Does cleavage by BamHI result in a 5′ or 3′ overhang? What is the sequence of this overhang?

3' CTAG 5' and 5' GATC 3'

same thing

b) BclI cleaves after the first T:

5’ TGATCA 3’
3’ ACTAGT 5’

what does that mean? → I get it now

Does cleavage by BclI result in a 5′ or 3′ overhang? What is the sequence of this overhang?

5' GATC 3'

or 3' CTAG 5'

c) Given the DNA shown below …

```
5’  ATTGAGGATCCGTAATGTGTCCTGATCACGCTCCACG  3’
3’  TAACTCCTAGGCATTACACAGGACTAGTGCGAGGTGC  5’
```

i) If this DNA was cut with BamHI, how many DNA fragment would you expect? Write out the sequence of these double-stranded DNA fragments.

2 ✓ ATTGAGGATC

don't reconnect / extend

c) ii) If the DNA shown on the previous page in (c) was cut with BclI, how many DNA fragment would you expect? Write out the sequence of these double-stranded DNA fragments.

2 ✓

Bio 7.012
FALL 2012
EXAM 2 REVIEW – KEY
Material taken from MIT OCW

Question 1

Transcription

i. is the process that transfers information from ___*DNA*___ to ___*RNA*___.

ii. in eukaryotic organisms, transcription occurs in the

[✓ Nucleus] Ribosome Membrane

Translation

i. is the process that transfers information from ___*RNA*___ to ___*protein*___.

ii. in eukaryotic organisms, translation occurs in the

Nucleus [✓ Ribosome] Membrane

Question 2

The following sequence of DNA encodes a hypothetical polypeptide called Playdo in a hypothetical bacteria *E. hypotheticus*. Transcription starts at and includes the C/G base pair in bold. The underlined T/A base pair indicates the terminator.

5'-TT**C**CCCTATGGATGGTCATCTACGATGCCCCCATCACTAAAGCTTG-3'
3'-AA**G**GGGATACCTACCAGTAGATGCTACGGGGGTAGTGATTTCGAAC-5'

c) What are the first 6 bases of the transcribed RNA? Be sure to label the 5′ and 3′ ends.

5′-CCCCUA-3′

d) What are the first 3 amino acids of the subsequent polypeptide? Be sure to label the N- and C- termini.

N-Met-Asp-Gly-C

e) How many total amino acids are encoded in this polypeptide?
The gene encodes 10 amino acids. The 11th in-frame codon is the stop codon UAA.

You identify a strain of bacteria containing a mutant tRNA that is capable of adding a tryptophan residue when it recognizes the codon UAG in the mRNA.

f) What is the sequence of the anticodon of the mutant tRNA? Be sure to label the 5′ and 3′ ends.

3′-AUC-5′ or 5′-CUA-3′

g) The Playdo polypeptide would be longer / [✓ the same length] / shorter in the presence of the mutant tRNA.

Why?

The length of Playdo would be the same in the presense of the mutant tRNA because the Playdo gene sequence does not include the TAG, so the mutant tRNA would never be used in translating Playdo

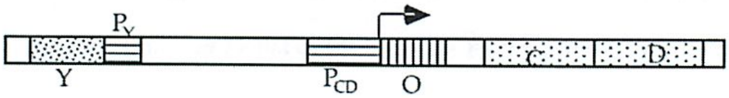
When you purify the Playdo polypetide from mouse cells you are shocked to learn that mouse Playdo is the same length in amino acids as bacterial Playdo.

j) Explain how is it possible that mouse Playdo and bacterial Playdo are the same polypeptide length even though they have substantially different gene lengths.

Mouse genes have introns – regions of DNA within the coding sequence of a gene that do not get translated. These regions are spliced out of the initial transcript when mRNA is prepared. The 36 extra base pairs in the mouse gene are such an intron.

Question 3

A scientist has discovered a bacterium that can metabolize alcohol to pyruvate when stimulated by the presence of ethanol. Hoping to use this bacteria to eliminate the harmful effects of excessive alcohol consumption, he finds that this bacterium produces the enzymes C and D which catalyze the ethanol → pyruvate reactions. The genes for enzymes C and D are part of an operon as diagrammed below.



a) You find two mutants, mutant 1 and mutant 2 that each carry a loss-of-function mutation in one of the components of the operon. These mutants constitutively produce enzymes C and D. Given this information, Is Y an activator or repressor? Explain.

Loss of these components, likely Y and O, allows transcription of C and D whether or not ethanol is present. Thus Y is a repressor.

You want to determine what component is missing in each of the mutants. You examine production of enzymes C and D in wild-type, mutant 1 and mutant 2 in the presence and absence of ethanol. You also create partial diploids and measure production of the enzymes.

cell type	No ethanol Production of C and D	+ ethanol Production of C and D
Wild type	low	high
Mutant 1	high	high
Mutant 2	high	high
Mutant 1 + $Y^+ P_Y^+ P_{CD}^+ O^+$	high	high
Mutant 2 + $Y^+ P_Y^+ P_{CD}^+ O^+$	low	high

b) Mutant 1 has a loss-of function mutation in which component of the operon? Explain.
Mutant 1 has lost the function of the operator, O, such that repressor cannot bind. Operators act at the level of DNA and can only control the genes to which they are physically attached. Even when a wild type copy of all components is added, the O^- *operon will always produce C and D.*

c) Mutant 2 has a loss-of function mutation in which component of the operon? Explain.
Mutant 2 has lost the function of the repressor (Y) or the promoter of the repressor (P_Y) *such that the Y protein is not made, or if it is made it can not bind to the operator. When a wild type copy of all components is added, the Y repressor is made and appropriately regulates transcription of C and D.*

Question 4

a) The genes needed to synthesize tryptophan are found in the *trp* operon. The presence of tryptophan controls the transciption of these genes.

i) Assume that the regulatory protein for this operon is a repressor. When do you expect the repressor to be associated with the operator? Cirlce one.

When tryptophan concentration is low. [✓ When tryptophan concentration is high.]

Explain your choice.
You would inhibit the prodution of the enzymes needed to make tryptophan if enough tryptophan is present.

ii) Assume that the regulatory protein for this operon is an activator. When do you expect the activator to be associated with the operator? Cirlce one.

[✓ When tryptophan concentration is low.] When tryptophan concentration is high.

Explain your choice.
You would produce the enzymes needed to make tryptophan if tryptophan is low.

Question 5

On your trip to the Amazon you were introduced to a new plant that the indigenous peopl use as an anti-viral treatment. You took samples home to your lab and you found that this plant makes a protein (the PV protein) that prevents viral replication. Excited by the possit anti-AIDS applications, you construct a genomic DNA library from this plant in the hope o cloning the *PV* gene.

a) What is a genomic DNA library?
A collection of bacterial cells (or phage particles) each of which contain a different small piece of the genomic DNA that you are interested in.

b) You get a DNA sample from cells, digest it with a restriction enzyme, and clone it into a vector. List 3 features of the vector that are absolutely required for your library constructio

An origin of replication (ori)
A cloning site
A selectable marker

c) Circle on the following lists ALL you would need in order to construct the genomic DN library. Assume you start with intact plant genomic DNA.

Enzymes

- [✓ Restriction enzyme]
- [✓ Ligase]
- DNA Polymerase
- RNA Polymerase
- Transcriptase
- Reverse Transcriptase
- 3' to 5' exonuclease
- [✓ Cloning vector]

Reagents

- Size separation gel
- Okasaki fragments
- ATP, TTP, CTP, GTP
- ddATP, ddTTP, ddCTP, ddGTP
- Primers
- Replication fork
- [✓ *E. coli* (bacteria)]
- Human cells
- Virus

Briefly describe the function of each item circled.
Restriction enzyme: cut the vetor and the genomic DNA
Ligase: join together the cut DNA fragments
Cloning vector: receives the cut genomic DNA and allows propagation in the bacterial cell host.
E. coli: Acts as a host that receives the recombiant plasmids and replicated this new DNA

Question 6

Restriction enzymes are extensively used in molecular biology. Below are the recognition sit of two of these enzymes, BamHI and BclI.

a) BamHI, cleaves after the first G:

5' GGATCC 3'
3' CCTAGG 5'

Does cleavage by BamHI result in a 5' or 3' overhang? What is the sequence of this overhang

5' G 3' *3' **CCTAG** 5'*	*Cleavage by BamHI leaves a 5' overhang*

b) BclI cleaves after the first T:

5' TGATCA 3'
3' ACTAGT 5'

Does cleavage by BclI result in a 5' or 3' overhang? What is the sequence of this overhang?

5' T 3' *3' **ACTAG** 5'*	*Cleavage by BclI leaves a 5' overhang*

c) Given the DNA shown below …

```
5' ATTGAGGATCCGTAATGTGTCCTGATCACGCTCCACG 3'
3' TAACTCCTAGGCATTACACAGGACTAGTGCGAGGTGC 5'
```

i) If this DNA was cut with BamHI, how many DNA fragments would you expect? Write out the sequence of these double-stranded DNA fragments.

2 fragments:

A	*B*
5' ATTGAG 3'	*5' GATCCGTAATGTGTCCTGATCACGCTCCACG 3'*
3' TAACTCCTAG 5'	*3' GCATTACACAGGACTAGTGCGAGGTGC 5'*

c) ii) If the DNA shown on the previous page in (c) was cut with BclI, how many DNA fragment would you expect? Write out the sequence of these double-stranded DNA fragments.

2 fragments.

C	*D*
5' ATTGAGGATCCGTAATGTGTCCT 3'	*5' GATCACGCTCCACG 3'*
3' TAACTCCTAGGCATTACACAGGACTAG 5'	*3' TGCGAGGTGC 5'*

㊼

As Doing Pratice exams

N-terminus

Start of protine

~~N~~ amine grp

$-NH_2$

Order ↓

C terminus

end

Carboxyl grp

$-COOH$

---

table is 5' → 3' on coding strand

(48)

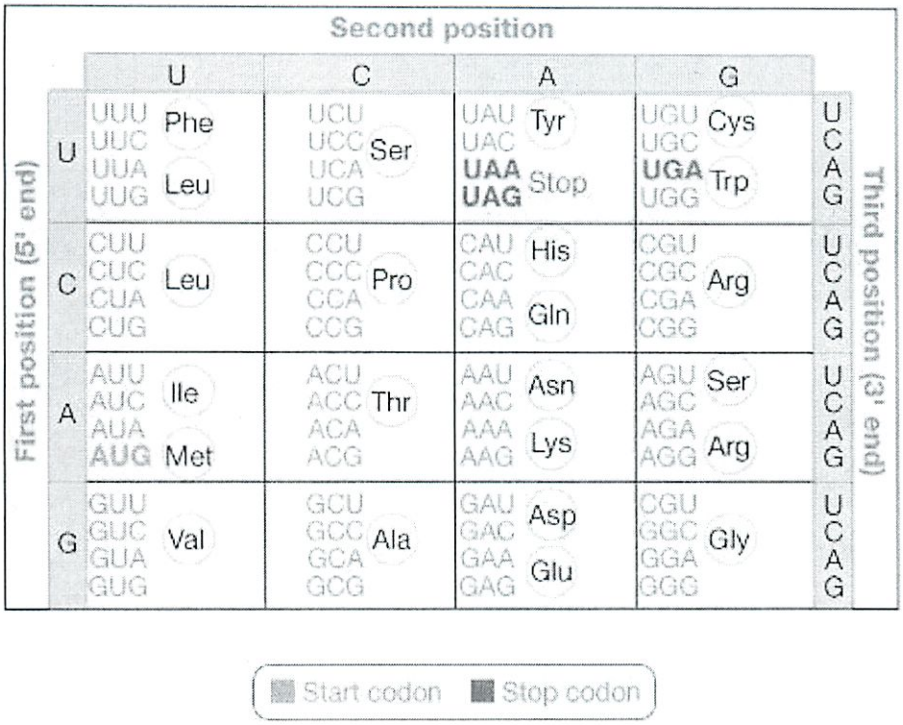
Repressor - binds and inhibits RNA polymerase from binding

[illegible] have a specific binding site on promoter

prevents from transcribing

Operon - regulates ability of RNA polymerase to access the promoter

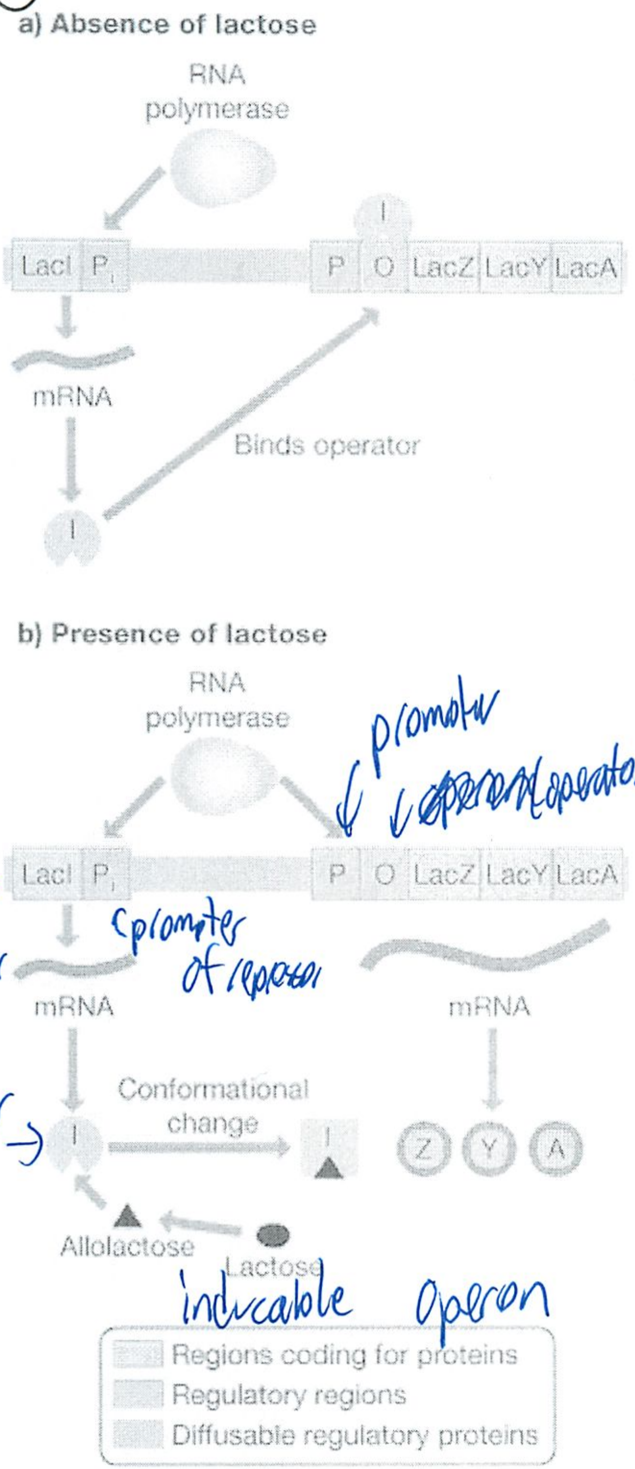
㊾



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

■ Start codon ■ Stop codon

(50)



a) Absence of lactose

b) Presence of lactose

RNA polymerase
LacI
P_I
P
O
LacZ
LacY
LacA
I
mRNA
Binds operator
RNA polymerase
LacI
P_I
P
O
LacZ
LacY
LacA
mRNA
mRNA
Conformational change
I
Z
Y
A
Allolactose
Lactose
▨ Regions coding for proteins
▨ Regulatory regions
▨ Diffusable regulatory proteins

promoter

~~operon~~ operator

gene of repressor

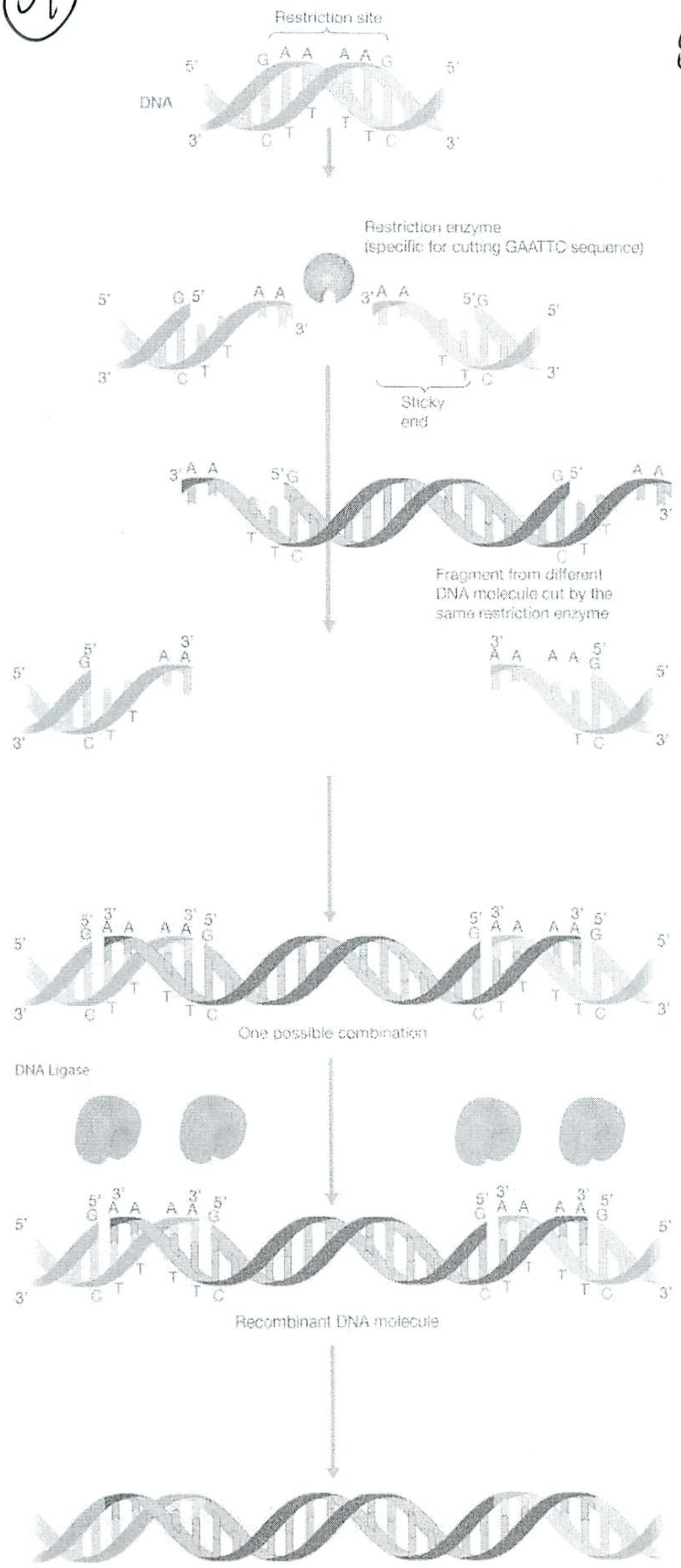
promoter of repressor

repressor →

inducible operon

Operon = Operator + promoter + coding regions of the gene

㊿ (51)


Restriction site
5'
G A A A A G
5'
DNA
3'
C T T T T C
3'
Restriction enzyme
(specific for cutting GAATTC sequence)
5'
G 5'
A A
3'
3' A A
5' G
5'
3'
C T T
T T C
3'
Sticky
end
3' A A
5' G
G 5'
A A
3'
T T C
C T
Fragment from different
DNA molecule cut by the
same restriction enzyme
5'
5' G
3' A A
3'
C T T
3' A A
A A 5' G
5'
3'
T C
3'
5' G 3' A A A A 3' 5' G
5' G 3' A A A A 3' 5' G
5'
3'
C T T T T C
C T T T T C
3'
One possible combination
DNA Ligase
5' G 3' A A A A 3' 5' G
5' G 3' A A A A 3' 5' G
5'
3'
C T T T T C
C T T T T C
3'
Recombinant DNA molecule

So this is

3' G|AATTC A 5'

5' CTTGA|A A 3'

ah so 3' sticks out on both ends

Exam #2

missence - altered content

nonsense - alters codon so stop

frame shift - adding/subtracting not 3

Silent - same amino acid

guessed correctly ...

---

if ~~both~~ a females child has it
and is x-linked recessive
father must have + show (only 1 x linked)

10+11 both carriers
must get one from each

~~Even though no~~ (could be directly linked
but are not here

(53)

but shows 6 must also have mutant

C,6

So never assume C ~~a~~ auto means has

└ well looks for exceptions

They will give you facts that support your conclusion

So if no exceptions shown likely (in problem world)

that ~~there~~ it is perfectly corelated w/

---

for 23 to not be a Carrier

└ since not shaded → we know 0 or 1

to be carier → 20 must be carier

We know 10, 11 both cariers

and

(54)

every ~~pare~~ child C from parents is mutated

→ So 10 c is mutated
10 g not mutated

11 → know 10 G not mutated from
but since 10 shaded, know 11 G mutated
11 C never mutated
(or else 11 shaded)

20 to be not mutated must take
10 g
11 C } both not mutated

Practice for 7.102 Quiz II

Question 1

For the following questions, answer each in the space provided.

a) All DNA polymerases can make a new DNA strand in the 5′ to 3′ direction. To accomplish this, DNA polymerases require what three non-protein components?

neucleotides energy GTP ? ?

b) The type of genomes found in viruses can vary. Some viruses have single (ss) or double-stranded (ds) DNA genomes, others have single (ss) or double-stranded (ds) RNA genomes. Given the data below, identify **all possible** types of genomes for each virus.

Corrected via ~~don't get~~ email →

Virus	% A	% T	% U	% C	% G	Type of genome, i.e., ds-DNA
1	18	18	0	32	32	ds - DNA ds-DNA ✓
2	25	0	25	25	25	~~ss ds RNA~~ ds ss RNA ✓
3	35	15	0	15	35	ss DNA ✓

no RNA the doub strand
no DNA
no RNA sing

c) When a virus infects a cell with its single-stranded (ss) RNA genome, the host cell does not have the ability to replicate the viral genome.

- Explain why the host cell can't replicate the viral (ss) RNA genome.

  RNA is usually not replicated needs RNA dep RNA polymerase

- What approach can a (ss) RNA virus use to replicate its genome in a host cell?

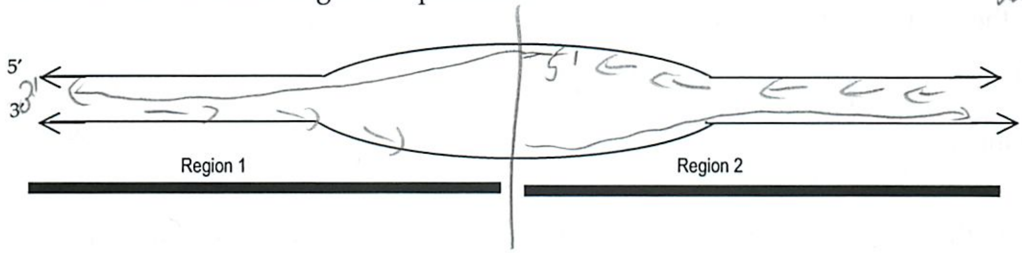
  reverse trancriptase to DNA must deliver it

d) In the first step of a PCR cycle, the mix is heated to denature the template DNA. Which template DNA would require a higher temperature for this first step, a 1000 base pair template with 35% cytosine or a 1000 base pair template with 25% cytosine? Explain your choice.

first more heat, ✓ since bond stronger 3 vs 2 ✓ A T C G ? bonds stronge

e) Below is a schematic of an origin of replication.

Order!



- Would replication of the **top** strand in region 1 be [✓ continuous] or discontinuous?

- Would replication of the **bottom** strand in region 2 be [✓ continuous] or discontinuous?

- Explain why replication of the genome would fail in a cell that lacks DNA ligase?

  can't put the Okazaki fragments back together ✓

trans trans

Question 2

The following is a *partial* sequence of a double stranded **bacterial DNA** that encodes a short peptide. The promoter sequence is shown as XXXXX. Assume transcription begins at the first C/G base pair after the promoter.

Wild-type DNA sequence:

stop

5' XXXXXXXXXXXXCTGCTTCAATATGAACCAGTGGAGTGCCTTAAAGATCTGACGAAACGTCACGGAATCTCTAGACTGCTTCAAT
3' XXXXXXXXXXXXGACGAAGTTATACTTGGTCACCTCACGGAATTTCTAGACTGCTTTGCAGTGCCTTAGAGATCTGACGAAGTTA 5

; ATG sets frame ?

a) For the sequence above,

- Circle the template strand for transcription.
- Label the 5′ and the 3′ ends of each strand.
- Indicate the direction of transcription by an arrow. forgot - but was thinking

b) Give the sequence of the first 10 nucleotides of the mRNA transcript and label its 5′ and 3′ ends.

5' CUG CUUCAA UA3'UG 3'

c) The peptide produced from this mRNA transcript and label its N and C ends.

N Met Asn Gln Trp Ser - ala leu lys ile Stop got lazy

c) Give the base sequence of the **anti-codon** that inserts the fourth amino acid into the peptide and label its 5′ and the 3′ ends.

3' ACC 5' ✓
5' UGG 3'

d) The following are two mutant versions of the wild-type DNA sequence that is shown above. The mutated base pair in both versions is **bold and underlined.**

Mutant 1:

CTGCTTCAATATGAAC**T**AGTGGAGTGCCTTAAAGATCTGACGAAACGTCACGGAATCTCTAGACTGCTTCAAT
GACGAAGTTATACTTG**A**TCACCTCACGGAATTTCTAGACTGCTTTGCAGTGCCTTAGAGATCTGACGAAGTTA

Mutant 2:

CTGCTTCAATATGAA**T**CAGTGGAGTGCCTTAAAGATCTGACGAAACGTCACGGAATCTCTAGACTGCTTCAAT
GACGAAGTTATACTT**A**GTCACCTCACGGAATTTCTAGACTGCTTTGCAGTGCCTTAGAGATCTGACGAAGTTA

For each mutant version,

i) Write the sequence of the peptide that is produced. Label N and C termini.

Mutant 1: N' Met Asn C' stop ✓

Mutant 2: N' Met Asn Gln ... C' ✓

ii) Identify the type of point mutation. *Choose from silent/ missense/ nonsense/ frameshift.* ✓ didn't study

Mutant 1: nonsense ? ✓

Mutant 2: Silent ✓

Question 2 continued

e) Would the substitution of a base that is a part of the 4th codon in the given wild-type DNA sequence always change the resulting peptide sequence? **Explain** your answer.

Yes - since UGG is the only one for Trp ✓

f) Would the substitution of a base that is a part of the 3rd codon in the given wild-type DNA sequence always change the resulting peptide sequence? **Explain** your answer.

No multiple choices for Gln ✓

Question 3

After agarose gel electrophoresis and staining, state how many bands you would see in a lane in which you loaded the following. Consider each independently.

a) A circular plasmid that was cleaved with a restriction enzyme that cuts at one. Assume every molecule is cut. ?

?1 line ✓

b) A circular plasmid of 5000 base pairs that was cleaved with a restriction enzyme that cuts at two sites, at position 1 and position 2500. Assume every molecule is cut.

lenght → 2 × 2500 - would only see 1 ✓

c) A linear piece of DNA that was cleaved with a restriction enzyme that cuts at one site in the center of the molecule. Assume every molecule is cut.

2 halves exactly same lenght → 1 ✓

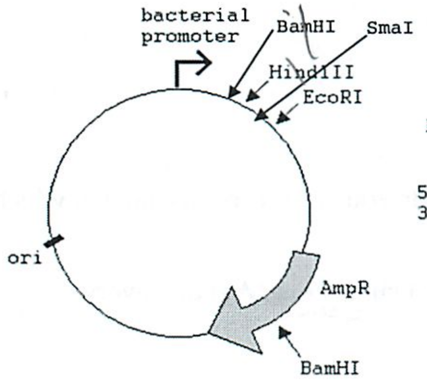
Question 3, continued

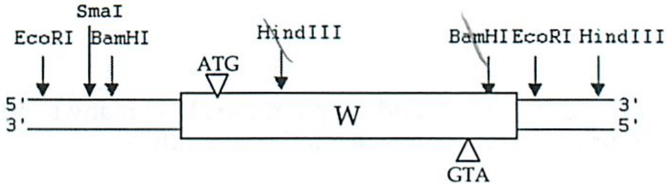
You want to clone gene W into the vector shown below and transform the resulting recombinant plasmid into *E. coli* cells such that the protein encoded by gene W will be expressed in the host bacterial cells. Note the fragment of gene W in the figure does **not** have a promoter.

? what is diff w/ promoter?

* indicates the cleavage site ?

EcoRI:	*BamHI:*	*HindIII:*	*SmaI*
G*AATTC CTTAA*G	G*GATCC CCTAG*G	A*AGCTT TTCGA*A	CCC*GGG GGG*CCC


bacterial promoter
BamHI
SmaI
HindIII
EcoRI
ori
AmpR
BamHI


SmaI
EcoRI
BamHI
HindIII
ATG
BamHI
EcoRI
HindIII
5'
3'
W
3'
5'
GTA

d) There are at least two options for cloning gene W into the vector.

Option #1:
What restriction enzyme(s) would you use to cut the vector?
EcoRI ✓
What restriction enzyme(s) would you use to cut gene W DNA?
EcoRI ✓

Option #2:
What restriction enzyme(s) would you use to cut the vector?
EcoRI + SmaI
What restriction enzyme(s) would you use to cut gene W DNA?
EcoRI + SmaI

but no sticky end
← has sticky end
forgot possibility

e) Which option is better for creating a recombinant plasmid that will express the gene W protein? Explain your answer.

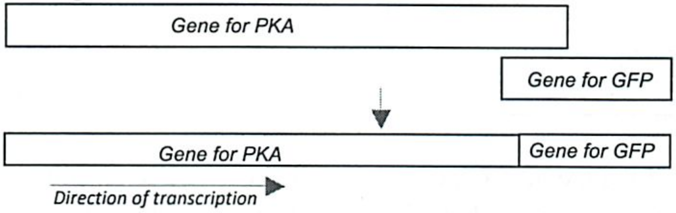
~~↑ since no sticky end in 2~~

don't know where bacteria promoter is (to left or right of)
so keep both options open

Question 4

known is end :) use the test to take the test

You are interested in purifying and characterizing a specific plant protein (PKA) that acts as a potent anti-depressant. You start by fusing the DNA encoding GFP (green fluorescent protein) to the DNA encoding the C terminus of PKA gene. See below.



The following is the partial **cDNA sequence** encoding the C terminus of the PKA protein. The sequence encoding the stop codon is shown in bold. The bars above the sequence show restriction enzyme recognition sites.

Z Y

5′-TCAAGAGGATCCCCGGTACCGAATTCCATGTTA**TAG**CAAGCTCGGAATTAACCCTCAC-3′
3′-AGTTCTCCTAGGGGCCATGGCTTAAGGTACAAT**ATC**GTTCGAGCCTTAATTGGGAGTG-5′

The following is the partial cDNA sequence encoding the N terminus of GFP. The codons are underlined. The bars above the sequence show the restriction enzyme recognition sites.

Z ✓ this moved? Y

5′-TCTAGAGGTACCGGGATCCTGAATTCC ATG CCA AGC GGC-3′
3′-AGATCTCCATGGCCCTAGGACTTAAGG TAC GGT TCG CCG-5′

The recognition sequences and the cleavage sites (indicated by /) for each enzyme are given below.

Enzyme Z	**Enzyme Y**
5′ G/GATC C 3′	**5′ G/AATT C 3′**
3′ C CTAG/G 5′	**3′ C TTAA/G 5′**

a) Choose the restriction enzyme that you will use to cut the two genes before ligating them to make a fusion gene. **Explain** why you chose this restriction enzyme.

Would cut w/ Y since only one that maintains the reading frame

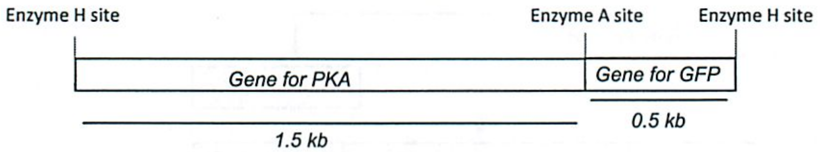
not seeing

~~→ actually in test~~

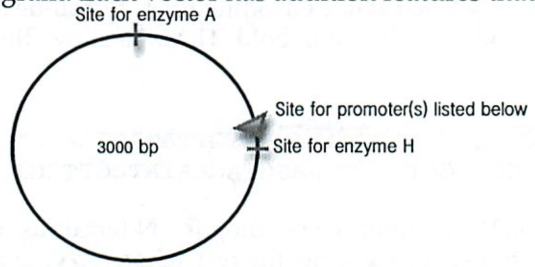
Oh those are the restriction enzyme sensing sites not totally clear on the whole restriction enzyme thing

Question 4, continued

You successfully create a DNA fragment that encodes the PKA-GFP fusion protein as shown below. *(Note: recognition sites for two new restriction enzymes, enzyme A and H are labeled in the schematic below).*



b) You plan to clone this PKA-GFP gene into a vector that will allow you to amplify and express PKA - GFP fusion gene in bacterial as well as mammalian cells. You have the choice of 4 vectors each with the general features shown in the diagram. Each vector has addition features that are listed below.



Vector 1 contains: 1) ampicillin resistence gene, 2) bacterial origin of replication and 3) bacterial promoter   no mamal parts

Vector 2 contains: 1) ampicillin resistence gene, 2) bacterial origin of replication, and 3) mammalian origin of replication, 4) bacterial promoter and 5) a mammalian promoter   all

Vector 3 contains: 1) ampicillin resistence gene, 2) bacterial origin of replication, and 3) mammalian origin of replication, 4) a mammalian promoter   Not clear on what we need   no bac promoter

Which of the above vectors would allow you to clone and express the fusion gene in both bacterial cells and mammalian cells? **Explain** why you selected this vector.
Only vector 2 will allow expression of the fusion gene in both bacterial cells and mammalian cells because only vector 2 has both types of ORIs and both types of promoters. ← answer
? but good to see left

c) Based on the plasmid that you selected, what should be the phenotype of the bacterial cells **prior to transformation**?
The cells prior to transformation must be sensitive to ampicillin. ✓
subject to ampacil

d) Following bacterial transformation, you want to identify the bacterial cells that received a plasmid. Onto what type of media would you plate your transformation mix?
You would plate the cells on media containing ampicillin.

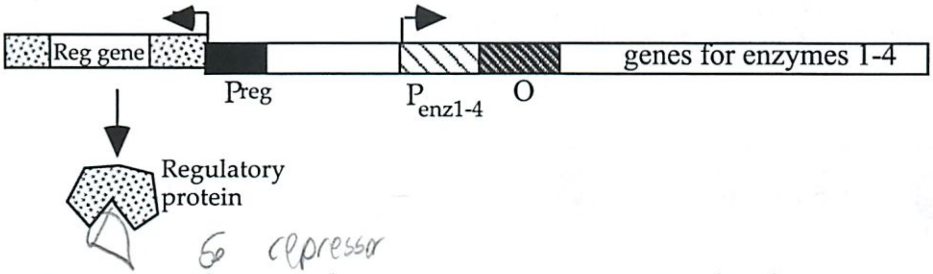
✓

Ampacilin

the one I always miss

Question 5

In bacterial cells, the genes encoding the enzymes for tyrosine production organized into an operon. You find that enzyme production is influenced by the presence of tyrosine, and you know that tyrosine binds to the regulatory protein. A diagram of this operon is shown below.



So repressor

a) You make a mutant that lacks the regulatory protein and find that these cells always produce tyrosine. Does the gene for the regulatory protein encode an activator or a repressor?

repressor

b) Complete the following table.

Strain	Genotype					Enzyme 1-4 Activity	
	reg	P_{reg}	O	P_{enz1-4}	Enz1-4	- Tyrosine	+ Tyrosine
WT	+	+	+	+	+	+	-
A	-	+	+	+	+	+	+
B	+	+	-	+	+	+	+
C	+	+	+	-	+	-	-
D	+	+	+	+	-	-	-

no regulator
no operator

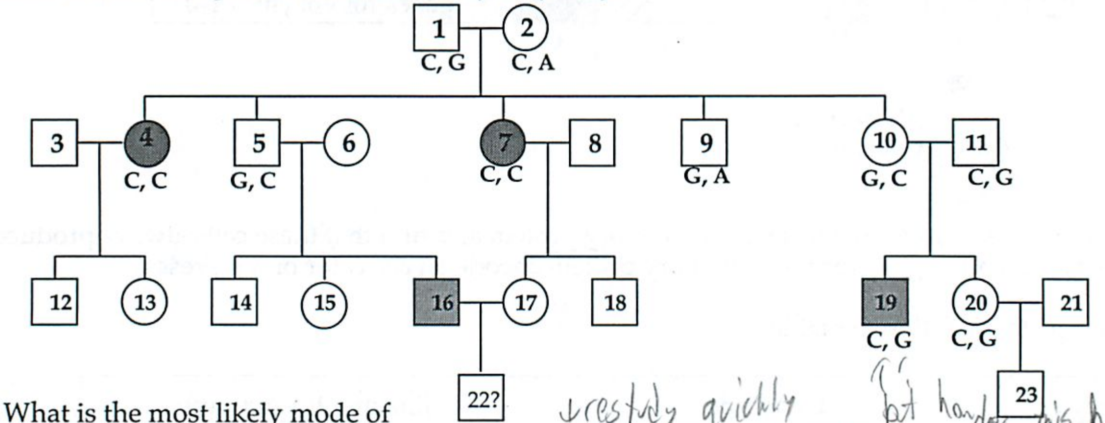
No Operator I guess it still makes it

No ability to bind → no regulating

have gene + is expressed

Question 6

Below is the pedigree of a family with a disease that shows 100% penetrance. *All the individuals that show the disease phenotype are shaded. The two letters identify the two alleles of a SNP that is tightly linked to Gene Z, the gene that is associated with this disease. For example G, A indicates that on one of the chromosomes you would find a G (a G/C base pair) and on other chromosome you would find an A (an A/T base pair). Please note that some of the individuals marrying into this family may be carriers. Also assume that no recombination occurs between the SNP and gene Z in the members of this family.*



↓restudy quickly

but how does this happen?

i) What is the most likely mode of inheritance (*autosomal dominant/ autosomal recessive/ X linked dominant/ X linked recessive*) of this disease?

autosomal recessive ✓

ii) What allele of the SNP is **tightly linked** with the disease allele in individuals 1,2,4,5,and 7?

C ✓

d) List **all** possible genotypes at the Z locus of following individuals in this pedigree? *Note: Use the symbol* X^D, X^d, *D or d where appropriate. In each case, use the letter "D" to represent the allele associated with the dominant phenotype and 'd" to represent the allele associated with the recessive phenotype.* disaster on last exam

Individuals	Genotype at the Z locus
#2	~~C,A C,G C,T~~ Dd
#5	~~"~~ Dd
#9	~~"~~ DD

must be carriers

must not have it

e) What is the probability of Individual #22 being affected?

well 17 has at least one C
16 has both Cs

1/2 ✓

f) Assume that individual #21 is not a carrier of the disease allele. What is the chance that individual #23 is a **carrier** of the disease allele? **Show your work.**

are one C

20 is C, G 21 is not C in both

got C from 11, G from 10 - not linked so not carrier

0%

	√	x
x	C	x
x	x	x

1/4

so not all C are carrier

Answers for Practice for 7.102 Quiz II

Question 1
For the following questions, answer each in the space provided.

a) All DNA polymerases can make a new DNA strand in the 5′ to 3′ direction. To accomplish this, DNA polymerases require what three non-protein components?
Template DNA, dNTPs, primer

b) The type of genomes found in viruses can vary. Some viruses have single (ss) or double-stranded (ds) DNA genomes, others have single (ss) or double-stranded (ds) RNA genomes.
Given the data below, identify **all possible** types of genomes for each virus.

Virus	% A	% T	% U	% C	% G	Type of genome, i.e., ds-DNA
1	18	18	0	32	32	*ds or ss DNA*
2	25	0	25	25	25	*ds or ss* ~~*DNA*~~ RNA
3	35	15	0	15	35	*ss DNA*

emailed update

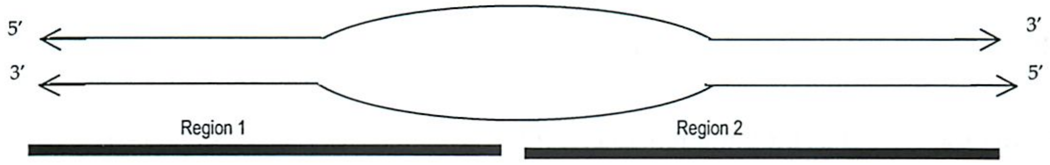
c) When a virus infects a cell with its single-stranded (ss) RNA genome, the host cell does not have the ability to replicate the viral genome.

- Explain why the host cell can't replicate the viral (ss) RNA genome.
  The host cell does not have an RNA dependent RNA polymerase.

- What approach can a (ss) RNA virus use to replicate its genome in a host cell?
  The virus must either deliver the RNA dependent RNA polymerase when it infects the host cell or the virus must encode the RNA dependent RNA polymerase on the RNA such that the host cell can translate it into the needed protein

d) In the first step of a PCR cycle, the mix is heated to denature the template DNA. Which template DNA would require a higher temperature for this first step, a 1000 base pair template with 35% cytosine or a 1000 base pair template with 25% cytosine? Explain your choice.
The greater the G-C content the higher the melting temperature, therefore the 1000 base pair template with 35% cytosine would require a higher melting temperature. This is because the C-G base pair forms three hydrogen bands as compared to two for an A-T base pair.

e) Below is a schematic of an origin of replication.



- Would replication of the **top** strand in region 1 be continuous or discontinuous?
  Continuous

- Would replication of the **bottom** strand in region 2 be continuous or discontinuous?
  Continuous

- Explain why replication of the genome would fail in a cell that lacks DNA ligase?
  Without DNA ligase, the Okazaki fragments could not be joined together.

Question 2

The following is a *partial* sequence of a double stranded **bacterial DNA** that encodes a short peptide. The promoter sequence is shown as XXXXX. Assume transcription begins at the first C/G base pair after the promoter.

Wild-type DNA sequence:

5′ XXXXXXXXXXXXCTGCTTCAATATGAACCAGTGGAGTGCCTTAAAGATCTGACGAAACGTCACGGAATCTCTAGACTGCTTCAAT 3′
3′ XXXXXXXXXXXXGACGAAGTTATACTTGGTCACCTCACGGAATTTCTAGACTGCTTTGCAGTGCCTTAGAGATCTGACGAAGTTA 5′

→

a) For the sequence above,

- Circle the template strand for transcription.
- Label the 5′ and the 3′ ends of each strand.
- Indicate the direction of transcription by an arrow.

b) Give the sequence of the first 10 nucleotides of the mRNA transcript and label its 5′ and 3′ ends.
5′CUGCUUCAAUAUG 3′

c) The peptide produced from this mRNA transcript and label its N and C ends.
N-met-asn-gln-trp-ser-ala-leu-lys-ile-C

c) Give the base sequence of the **anti-codon** that inserts the fourth amino acid into the peptide and label its 5′ and the 3′ ends.
3′ACC5′

d) The following are two mutant versions of the wild-type DNA sequence that is shown above. The mutated base pair in both versions is **bold and underlined.**

Mutant 1:

CTGCTTCAATATGAAC**T**AGTGGAGTGCCTTAAAGATCTGACGAAACGTCACGGAATCTCTAGACTGCTTCAAT
GACGAAGTTATACTTG**A**TCACCTCACGGAATTTCTAGACTGCTTTGCAGTGCCTTAGAGATCTGACGAAGTTA

Mutant 2:

CTGCTTCAATATGAA**T**CAGTGGAGTGCCTTAAAGATCTGACGAAACGTCACGGAATCTCTAGACTGCTTCAAT
GACGAAGTTATACTT**A**GTCACCTCACGGAATTTCTAGACTGCTTTGCAGTGCCTTAGAGATCTGACGAAGTTA

For each mutant version,

i) Write the sequence of the peptide that is produced. Label N and C termini.

Mutant 1: *N-met-asn-C*

Mutant 2: *N-met-asn-gln-trp-ser-ala-leu-lys-ile-C*

ii) Identify the type of point mutation. *Choose from silent/ missense/ nonsense/ frameshift.*

Mutant 1: *Nonsense mutation since a premature stop codon is inserted.*

Mutant 2: *Silent mutation, since the peptide sequence is not changed by this mutation.*

Question 2 continued

e) Would the substitution of a base that is a part of the 4th codon in the given wild-type DNA sequence always change the resulting peptide sequence? **Explain** your answer.
There is only one codon for amino acid tryptophan. Therefore any substitution in this codon will change the amino acid and the resulting protein sequence.

f) Would the substitution of a base that is a part of the 3rd codon in the given wild-type DNA sequence always change the resulting peptide sequence? **Explain** your answer.
There are two codons for asn – 5'AAU3' (codon 1) and 5'AAC3' (codon 2). If the 3rd base of codon 2 is changed to "T" you will see the insertion of asn at the same position in the peptide. However if you change the third base of codon 1 to any other base, the asn at this position in the protein will be replaced by another amino acid. Hence the protein structure and function may change.

Question 3

After agarose gel electrophoresis and staining, state how many bands you would see in a lane in which you loaded the following. Consider each independently.

a) A circular plasmid that was cleaved with a restriction enzyme that cuts at one. Assume every molecule is cut.
One

b) A circular plasmid of 5000 base pairs that was cleaved with a restriction enzyme that cuts at two sites, at position 1 and position 2500. Assume every molecule is cut.
One

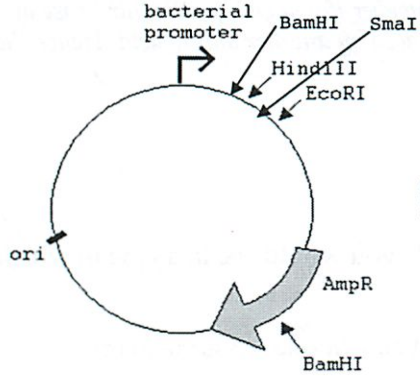
c) A linear piece of DNA that was cleaved with a restriction enzyme that cuts at one site in the center of the molecule. Assume every molecule is cut.
One

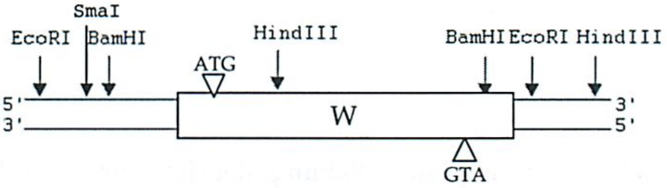
Question 3, continued

You want to clone gene W into the vector shown below and transform the resulting recombinant plasmid into *E. coli* cells such that the protein encoded by gene W will be expressed in the host bacterial cells. Note the fragment of gene W in the figure does **not** have a promoter.

* indicates the cleavage site

EcoRI:	*BamHI:*	*HindIII:*	*SmaI*
G*AATTC	G*GATCC	A*AGCTT	CCC*GGG
CTTAA*G	CCTAG*G	TTCGA*A	GGG*CCC





d) There are at least two options for cloning gene W into the vector.

Option #1:
What restriction enzyme(s) would you use to cut the vector? *EcoRI*

What restriction enzyme(s) would you use to cut gene W DNA? *EcoRI*

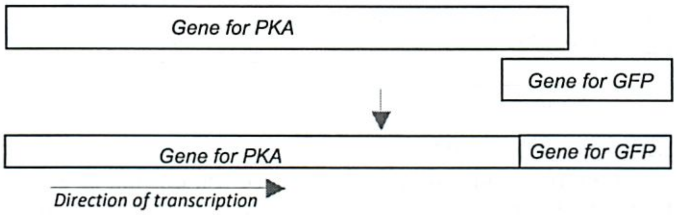
Option #2:
What restriction enzyme(s) would you use to cut the vector? *EcoRI and SmaI*

What restriction enzyme(s) would you use to cut gene W DNA? *EcoRI and SmaI*

e) Which option is better for creating a recombinant plasmid that will express the gene W protein? Explain your answer.
EcoRI only. This option gives both possible orientations of the insert with respect to the bacterial promoter. Since we do not know where the endogenous promoter for gene W would be (to the left or right as diagramed above), we need to keep both options available.

Question 4

You are interested in purifying and characterizing a specific plant protein (PKA) that acts as a potent anti-depressant. You start by fusing the DNA encoding GFP (green fluorescent protein) to the DNA encoding the C terminus of PKA gene. See below.



The following is the partial **cDNA sequence** encoding the C terminus of the PKA protein. The sequence encoding the stop codon is shown in bold. The bars above the sequence show restriction enzyme recognition sites.

```
           Z              Y
         ‾‾‾‾‾‾         ‾‾‾‾‾‾
5'-TCAAGAGGATCCCCCGGTACCGAATTCCATGTTATAGCAAGCTCGGAATTAACCCTCAC-3'
3'-AGTTCTCCTAGGGGCCATGGCTTAAGGTACAATATCGTTCGAGCCTTAATTGGGAGTG-5'
```

The following is the partial cDNA sequence encoding the N terminus of GFP. The codons are underlined. The bars above the sequence show the restriction enzyme recognition sites.

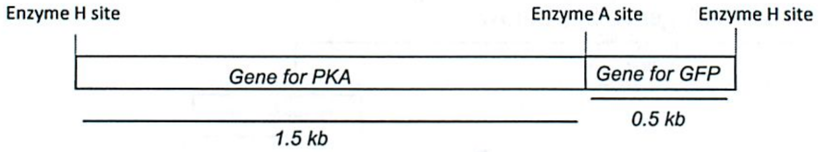
```
                  Z      Y
                ‾‾‾‾‾‾ ‾‾‾‾‾‾
5'-TCTAGAGGTACCGGGATCCTGAATTCC ATG CCA AGC GGC-3'
3'-AGATCTCCATGGCCCTAGGACTTAAGG TAC GGT TCG CCG-5'
                               ‾‾‾ ‾‾‾ ‾‾‾ ‾‾‾
```

The recognition sequences and the cleavage sites (indicated by /) for each enzyme are given below.

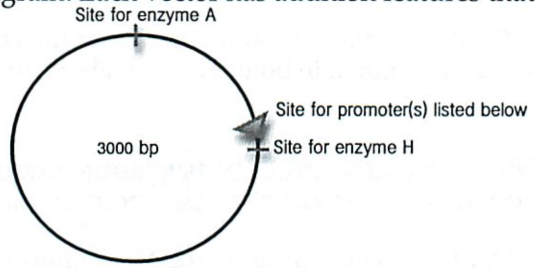
Enzyme Z	Enzyme Y
5′ G/GATC C 3′	5′ G/AATT C 3′
3′ C CTAG/G 5′	3′ C TTAA/G 5′

a) Choose the restriction enzyme that you will use to cut the two genes before ligating them to make a fusion gene. **Explain** why you chose this restriction enzyme.
You would cut with restriction enzyme Y, because it is the only one that maintains the reading frame needed to express GFP.

You successfully create a DNA fragment that encodes the PKA-GFP fusion protein as shown below. *(Note: recognition sites for two new restriction enzymes, enzyme A and H are labeled in the schematic below).*



b) You plan to clone this PKA-GFP gene into a vector that will allow you to amplify and express PKA - GFP fusion gene in bacterial as well as mammalian cells. You have the choice of 4 vectors each with the general features shown in the diagram. Each vector has addition features that are listed below.



Vector 1 contains: 1) ampicillin resistence gene, 2) bacterial origin of replication and 3) bacterial promoter

Vector 2 contains: 1) ampicillin resistence gene, 2) bacterial origin of replication, and 3) mammalian origin of replication, 4) bacterial promoter and 5) a mammalian promoter

Vector 3 contains: 1) ampicillin resistence gene, 2) bacterial origin of replication, and 3) mammalian origin of replication, 4) a mammalian promoter

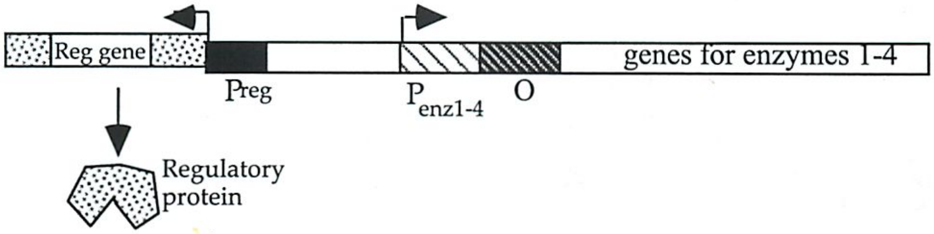
Which of the above vectors would allow you to clone and express the fusion gene in both bacterial cells and mammalian cells? **Explain** why you selected this vector.
Only vector 2 will allow expression of the fusion gene in both bacterial cells and mammalian cells because only vector 2 has both types of ORIs and both types of promoters.

c) Based on the plasmid that you selected, what should be the phenotype of the bacterial cells **prior to transformation**?
The cells prior to transformation must be sensitive to ampicillin.

d) Following bacterial transformation, you want to identify the bacterial cells that received a plasmid. Onto what type of media would you plate your transformation mix?
You would plate the cells on media containing ampicillin.

Question 5

In bacterial cells, the genes encoding the enzymes for tyrosine production organized into an operon. You find that enzyme production is influenced by the presence of tyrosine, and you know that tyrosine binds to the regulatory protein. A diagram of this operon is shown below.



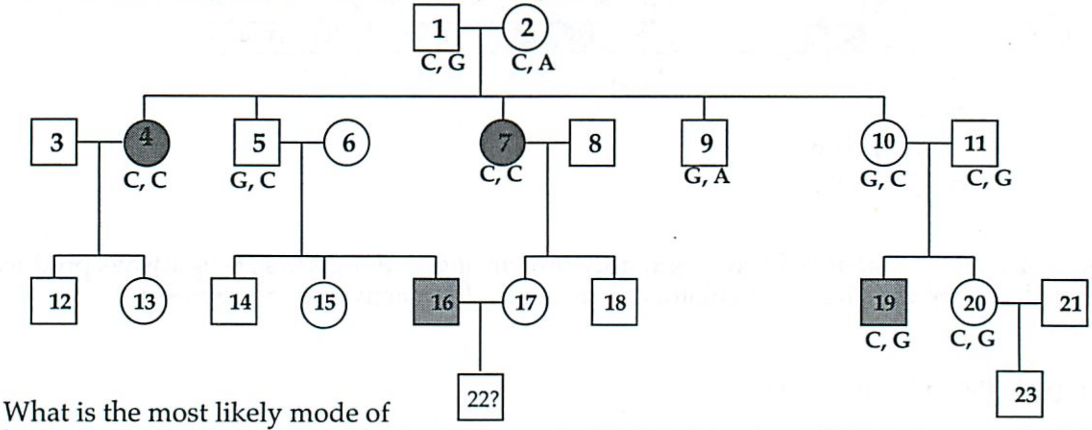
a) You make a mutant that lacks the regulatory protein and find that these cells always produce tyrosine. Does the gene for the regulatory protein encode an activator or a repressor?
Repressor

b) Complete the following table.

Strain	Genotype					Enzyme 1-4 Activity	
	reg	P_{reg}	O	$P_{enz1\text{-}4}$	Enz1-4	- Tyrosine	+ Tyrosine
WT	+	+	+	+	+	+	-
A	-	+	+	+	+	+	+
B	+	+	-	+	+	+	+
C	+	+	+	-	+	-	-
D	+	+	+	+	-	-	-

Question 6

Below is the pedigree of a family with a disease that shows 100% penetrance. *All the individuals that show the disease phenotype are shaded. The two letters identify the two alleles of a SNP that is tightly linked to Gene Z, the gene that is associated with this disease. For example G, A indicates that on one of the chromosomes you would find a G (a G/C base pair) and on other chromosome you would find an A (an A/T base pair). Please note that some of the individuals marrying into this family may be carriers. Also assume that no recombination occurs between the SNP and gene Z in the members of this family.*



i) What is the most likely mode of inheritance (*autosomal dominant/ autosomal recessive/ X linked dominant/ X linked recessive*) of this disease? *Autosomal recessive.*

ii) What allele of the SNP is **tightly linked** with the disease allele in individuals 1,2,4,5,and 7?
C allele.

d) List **all** possible genotypes at the Z locus of following individuals in this pedigree? *Note: Use the symbol X^D, X^d, D or d where appropriate. In each case, use the letter "D" to represent the allele associated with the dominant phenotype and 'd" to represent the allele associated with the recessive phenotype.*

Individuals	Genotype at the Z locus
#2	*Dd*
#5	*Dd*
#9	*DD*

e) What is the probability of Individual #22 being affected?
*50%. P= P(inheritance d from #16) * P(inheritance d from #17) = 1* 0.5 = 0.5 = 50%*

f) Assume that individual #21 is not a carrier of the disease allele. What is the chance that individual #23 is a **carrier** of the disease allele? **Show your work.**
0%. For #23 to be a carrier, #20 or #21 must be a carrier. We know that #21 is not a carrier. We also know that #20 is not affected, so it must have received the C allele from #11 and the G allele from #10. These two alleles are not linked to the disease allele, therefore individual #20 is not a carrier.

Constitutive - always making Lac Z

no control over

- repressor can't bind to Operator
- promoter for repressor protein

Non-inducable - no matter the signal → never make Lac Z

- repressor can't bind to lactose
- in Lacz non operation
  - P_{lac} -

and plenty more

look at w/ + w/o extra lactose

trp ~~opposite~~ when trp ↑

repressor becomes active + binds to operator

tr, so still called repressor?

7.012 Quiz II, Version A

10/31/12

Name: Michael Plasmeier **TA:** Hoschyer **Section #:** 27

Write your name on this page and your initials on all the other pages in the space provided.

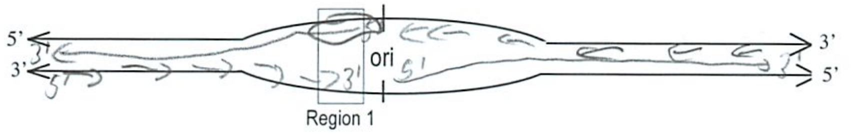
This exam has 9 pages including this coversheet. Check that you have pages 1-9. Page 9 contains the structures of the amino acids and a codon chart.

Question	Value	Score
1	17	11
2	16	14
3	13	4
4	12	7
5	14	10
6	28	25
TOTAL:	100	71

Name or initials MEP

Question 1 (17 points)

Consider the following segment of DNA (that is a part of a much larger molecule constituting a chromosome).



a) The DNA sequence in region 1 is :

```
5′ GCCATG 3′
3′ CGGTAC 5′
```

don't know which is template
yes we do 5' 3'

Give the sequence of a 6 nucleotide **RNA** primer that would occur at region 1 during replication and be elongated to form the leading strand.

5′ C A U G G C 3′ +3

3' ⟵ 3'

b) If primase activity is absent, you would expect which of the following? Put a check next to the **best** answer.

X A decrease mRNA production.

___A decrease tRNA production. X

✓ A decrease in both lagging and leading strand production.

___A decrease in lagging strand production. rep

___A decrease in leading strand production.

same ~ ~~no primase part~~ of ~~rep or trans~~

+0

c) If DNA polymerase lost its 3′ → 5′ exonuclease activity, you would expect which of the following? Put a check next to the **best** answer. repair

___An increase in the fidelity of both lagging and leading strand production.

X A decrease in the fidelity of both lagging and leading strand production.

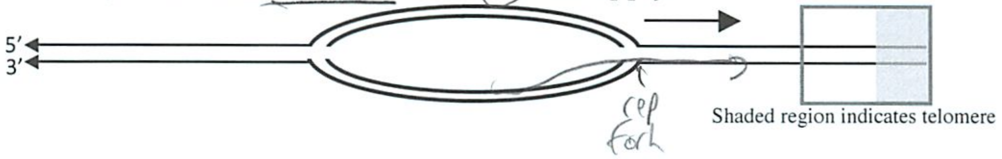
___A decrease in the speed of both lagging and leading strand production.

___A decrease in the fidelity of lagging strand production.

___A decrease in the fidelity of leading strand production.

+4

d) Consider the following schematic of a chromosome. Imagine the replication fork expanding to the right and approaching the end of the chromosome. To generate a complete copy of the original chromosome in the boxed region, which of the following statements is correct? Check **all** that apply.



+4

___DNA polymerase but not telomerase is needed when using the top strand as a template

✓ DNA polymerase but not telomerase is needed when using the bottom strand as a template

___The enzyme telomerase but not DNA polymerase is needed when using the top strand as a template

___The enzyme telomerase but not DNA polymerase is needed when using the bottom strand as a template

X Both DNA polymerase and the enzyme telomerase are needed when using the top strand as a template

X Both DNA polymerase and the enzyme telomerase are needed when using the bottom strand as a template

answers in terms of needed everywhere

Where are we saying it is needed?

11

Name or initials MEP

Question 2 (16 points)

Shown below is a double-stranded bacterial (*E. coli*) DNA sequence coding for the beginning of a hypothetical protein. Both strands are shown. The nucleotides are arbitrarily numbered from 1 to 90. The promoter sequence is italicized and underlined. **For this problem, transcription begins at the 10th base pair after the end of the promoter (i.e., at either base pair 20 or base pair 70, shown in bold).**

```
             10        20        30        40        50        60        70 →      80        90
              |         |         |         |         |         |         |         |         |
5' ...GTGTCGGAAACATATTGTGGGATGTTATATTGACAACGTCAAGGCACCATCAAATATGATCAGGATCATATTCTGCTGGGGCATAGGCGG... template
      ---------+---------+---------+---------+---------+---------+---------+---------+---------+
3' ...CACAGCCTTTGTATAACACCCTACAATATAACTGTTGCAGTTCCGTGGTAGTTTATACTAGTCCTAGTATAAGACGACCCCGTATCCGCC... Coding 5'
```

1/1

a) The mRNA transcribed from this sequence encodes the first few amino acids of the protein. Which strand is used as a **template for transcription**, the top or the bottom? top

reads 3' → 5'
does 5' → 3' | doesn't help

4/4

b) What are the first 10 nucleotides of the resulting mRNA?

5' C C A C A A U A U G 3'

3/3

c) Give the first three amino acids of the peptide encoded by this gene. Label the N and C termini.

N- Met Phe Pro C

The normal protein encoded by a different gene is 500 amino acids long. Below is sequence of the coding region of this new gene beginning at amino acid 100. The bottom strand is used as the template for transcription, and the underlined nucleotides represent the codon for amino acid 100.

```
5' GCTAATACCTGCAACTATATAAACACCCAC 3'
3' CGATTATGGACGTTGATATATTTGTGGGTG 5'
```

d) What is the primary sequence of amino acids 100-103?

3/3

	100	101	102	103
Amino acid:	ala	Asn	Thr	Cys

e) You have discovered mutant 1 that carries a mutation within this region as shown below in bold.

original: 5' GCTAATACCTGCAACTATATAAACACCCAC 3'
Mutant 1: 5' GCTAATACCTG**A**AACTATATAAACACCCAC 3'

Would mutant 1 produce a protein that functions like the normal protein? Explain why you made this choice.

No. UGA is now a stop codon, This would mean only part of the protein is produced - breaking it

2/2

f) You have discovered a different mutant, mutant 2, that carries a different mutation that deletes three base-pairs, as shown below.

deletion

original: 5' GCTAATACCTGCAACTATATAAACACCCAC 3'
Mutant 2: 5' GCTA---CCTGCAACTATATAAACACCCAC 3'

Mutant 2 produces a protein that functions like the normal protein. Explain why this mutation does not alter protein function.

Instead of Asn, Thr, we get just Thr, It appears that the Asn was not essential to protien function.

1/3

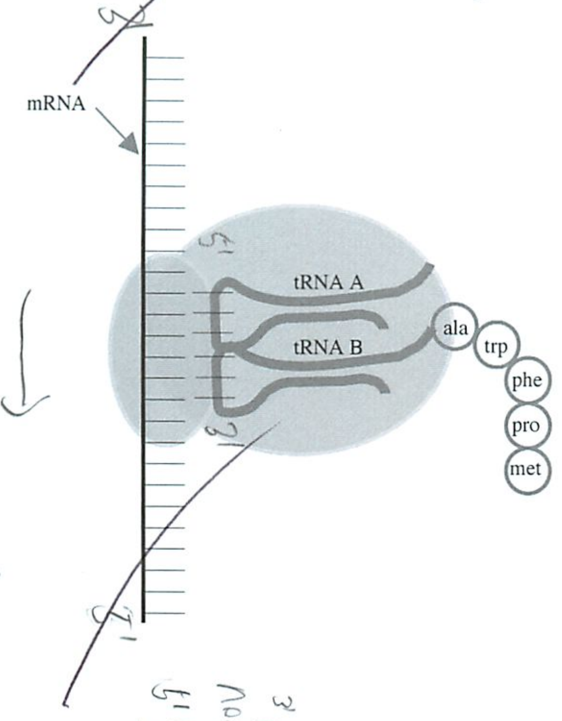
frame?

14/16

Name or initials MEP

Question 3 (13 points) 4 transc → transl

Below is a schematic of a ribosome actively translating an mRNA. The vertical line represents the mRNA, and each small horizontal line represents a ribonucleoside, and two tRNA molecules are labeled.


5'
mRNA
5'
tRNA A
tRNA B
3'
ala
trp
phe
pro
met
3'

reverse of usual

3' 5'
normal
5' B A 3'

a) Label the 3′ and 5′ ends of the mRNA. 0|3

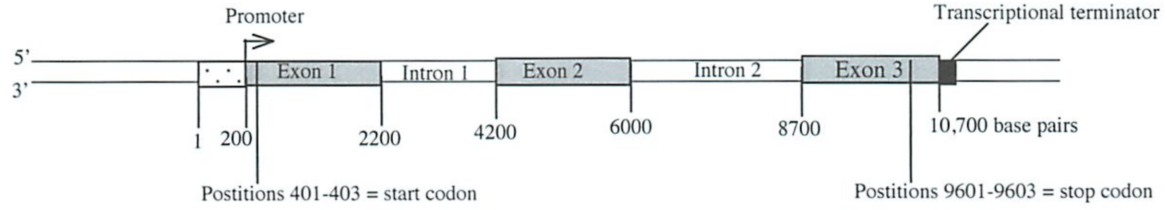
b) Indicate a sequence that could be the anticodon of tRNA A.

anticodon 5′ A G C 3′

codon 3' U C G 5' = Ala

0|4

c) Below is a schematic of gene Y, which encodes protein Y. The promoter region is indicated by the dotted box. Transcription begins immediately following the promoter.


Promoter
Transcriptional terminator
5'
3'
Exon 1
Intron 1
Exon 2
Intron 2
Exon 3
1
200
2200
4200
6000
8700
10,700 base pairs
Postitions 401-403 = start codon
Postitions 9601-9603 = stop codon

- The pre-mRNA produced by this gene would be approximately how many nucleotides long?

10,700 − 200 = 10500 3|3

- Two mature mRNAs of different lengths are produced from this gene. Name or describe the process by which this gene could produce two different mature mRNAs each that encode a different protein.

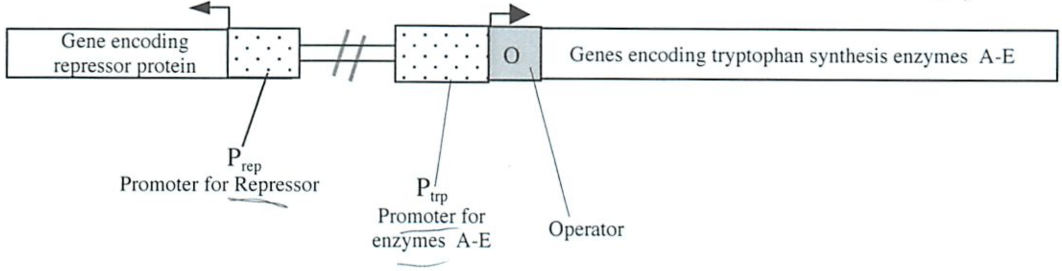
introns/exon process 1|3

Each version of the gene removes different introns
This results in different sizes of protiens created.

Name or initials MFP

Question 4 (12 points)

You design a summer class where you recreate experiments studying the *trp* operon in *E. coli* (see schematic below). Mutants missing one of the enzymes A-E are tryptophan auxotrophs.


Gene encoding repressor protein
O
Genes encoding tryptophan synthesis enzymes A-E
P_{rep}
Promoter for Repressor
P_{trp}
Promoter for enzymes A-E
Operator

a) When would you expect the production of enzymes A-E in a wild-type cell. Check all that apply.

___ When the levels of tryptophan in the cell are high

X When the levels of tryptophan in the cell are low

X When the cell is grown on minimal media.

from book - produces when levels are low

1/3

b) To which of the following would the trp repressor bind? Check all that apply.

___ Lactose

___ P_{rep}

X Operator

X Tryptophan

___ glucose

Prevents creation of more - when high trp blocks creation

1/3

c) A loss of which component or components would produce a cell unable to grow on minimal media? Check all that apply.

___ Gene for repressor

___ P_{rep}

~~X~~ Operator

X P_{trp}

X gene for enzyme A

X gene for enzyme E

loss of fn would produce ∞ Trp

Unable to produce trp at all

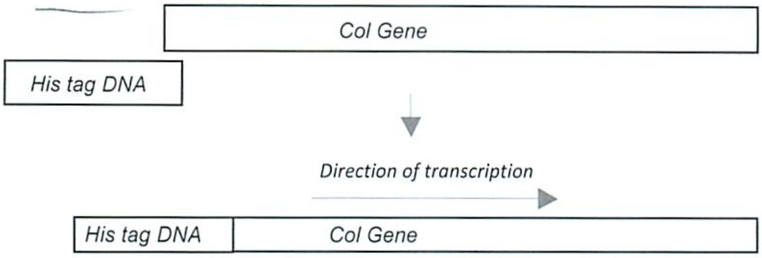
5/6

7

Name or initials MEP

Question 5 (14 points)

You have discovered a human gene Col, encoding the COL protein and associated with a colon disease. You plan to create a recombinant DNA sequence, by ligating a DNA sequence that encodes 5 histidines followed by a few additional amino acids in front of the protein-coding sequence for the Col gene. This will produce a slightly longer protein (see diagram below). You want to be sure that the combined sequence beginning with the five histidines maintains the correct reading frame to allow proper translation of the Col protein.



10/14

Below is the sequence that encodes the **histidine tag.** The bars above the sequence show the restriction enzyme recognition sites. The sequence encoding the start codon is shown in bold.

```
                  1   2   3   4   5   BglII     KpnI     EcoRI
*E. coli*   ...TTATG|CAT|CAT|CAT|CAT|CATA|GATCT|TGGTACC|GAATTC|TTAAG|ATC|C ... 3'
Promoter       ---------+---------+---------+---------+--------
            ...AATACGTAGTAGTAGTAGTATCTAGAACCATGGCTTAAGAATTCTAGG ... 5'
                                       1/3      2/3     2/3
```

Below is part of the cDNA sequence that encodes the Col gene. The sequence encoding the start codon is shown in bold. The bars above the sequence show the restriction enzyme recognition sites.

```
            BglII       KpnI       EcoRI
5'  ... TC|AAGATC|TCC|GCGGTACC|AAGAATTC|GGATC|ATG|TTA|TAG|CAA|GCT|CGG|AAT... 3'
        +---------+---------+---------+---------+---------+-------
3'  ... AGTTCTAGAGGCGCCATGGTTCTTAAGCCTAGTACAATATCGTTCGAGCCTTA... 5'
            2/3          0          2/3
```

The recognition sequences and the cleavage sites (indicated by /) for each enzyme are given below.

```
EcoRI:              KpnI:               BglII:
5' G/AATT C 3'      5' G/GTAC C 3'      5' A/GATC T 3'
3' C TTAA/G 5'      3' C CATG/G 5'      3' T CTAG/A 5'
```

a) You want to ligate these two pieces of DNA together to create a version of the COL protein that has 5 histidines attached to the N terminus in a manner that maintains the reading frame. Which enzyme or enzymes can you use to…

- cut the DNA encoding the **Histidine tag**. List all that apply.

  EcoRI ~~KpnI~~

- cut the **Col gene**? List all that apply.

  EcoRI ~~BglII~~

all cut 2/3 into codon −4

b) Give the DNA sequence of the top strand (as seen in diagram above) that can be found at the junction site between the His tag and the Col gene.

EcoRI
EcoRI

5' G G | T A C | C   G A | A T T | C   G G | A T C | A 3' ✓

6 nucleotides of HIS tag before RE site — 6 nucleotides of the restriction enzyme (RE) site — 6 nucleotides of Col gene after RE site

cut – not a codon line

c) In the sequence of part (b) draw vertical lines between each codon that corresponds to the amino acids of the His tag-COL fusion protein.

✓

Name or initials MFP

Question 6 (28 points)

You successfully create a DNA fragment that encodes the histidine-tagged human COL protein. You plan to clone this fragment into a vector that will allow you to express it in bacterial cells.

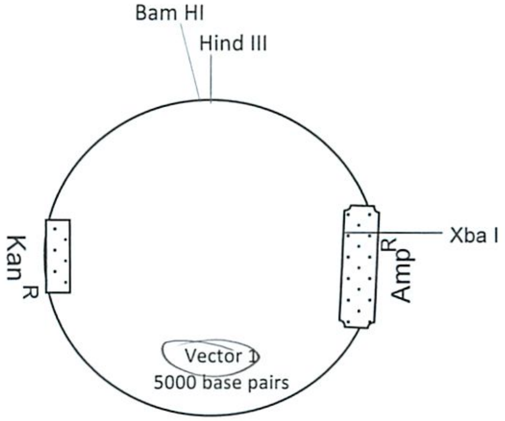
Your plan is to:

1) Cut an appropriate vector and the DNA fragment with *Xba I*.
2) Ligate the cut vector and the fragment together.
3) Transform *E. coli* cells with the ligation mix.
4) Select for *E. coli* cells that have a plasmid.
5) Identify the *E. coli* cells carrying a recombinant plasmid containing the inserted fragment by screening.

The following is a partial schematic of vector 1 that will allow you to complete the plan outlined above.

The Kan^R gene confers resistance to the drug kanamycin.

The Amp^R gene confers resistance to the drug ampicillin.



a) To allow selection for *E. coli* cells that have any plasmid (step 4), and screening for *E. coli* cells with the recombinant plasmid containing the inserted fragment (step 5), you will transform a particular strain of *E. coli* cells. What will be the phenotype of this strain prior to transformation?

0/3 resistant to kanamycin and ampicillin

b) To allow selection for *E. coli* cells that have either vector 1 or a recombinant plasmid, you will plate the transformation mix on media that contains which the following drugs?

3/3

Ampicillin [✓ Kanamycin] Both ampicillin and kanamycin Neither ampicillin or kanamycin

if put something at Xba I ← we do - we break Amp resistance

c) After selection you need to distinguish the colonies that contain cells with a recombinant plasmid from the colonies that contain cells with the original vector 1. To do this, you would replica plate from your selection plate onto media that contained amp.

6/6

Colonies that contain cells with a recombinant plasmid will die on this media. (Live or Die)

Colonies that contain cells with the original vector 1 will live on this media. (Live or Die)

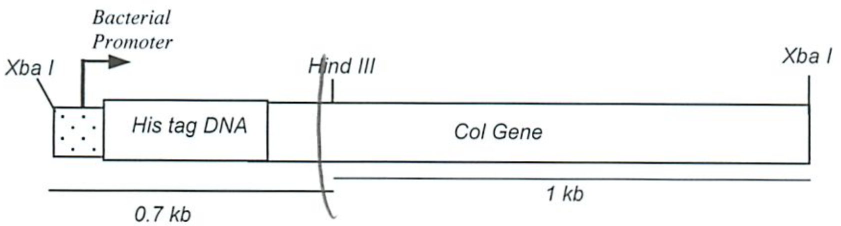
9

Name or initials MEP

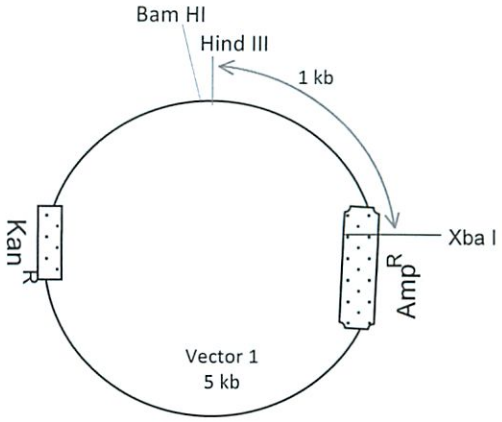
Question 6, continued

You plan to insert he histidine-tagged COL gene into vector 1 as an Xba I fragment (See diagram below). You will obtain two different recombinant plasmids.

from earlie


Bacterial Promoter
Xba I
Hind III
Xba I
His tag DNA
Col Gene
0.7 kb
1 kb

each direction


Bam HI
Hind III
1 kb
Xba I
Kan^R
Amp^R
Vector 1
5 kb

d) To differentiate between the two possible recombinant plasmids, you can cut both with restriction enzyme(s) and separate the resulting fragments by gel electrophoresis.

6/6

- What restriction enzyme or enzymes would you use to differentiate between the two possible recombinant plasmids? Hind III

4/4

- Given your answer above, what sized DNA fragments should you see from the restriction enzyme digestion of the two plasmids?

one recombinant plasmid would give: 1k + .7k, 5k + 1k = 1.7k, 6k

the other recombinant plasmid would give: 1k + 1k, 5k + .7k = 2k, 5.7k

6/6

- Which of these recombinant plasmids would allow expression of the histidine-tagged COL protein in bacterial cells? Explain your answer.

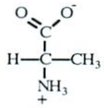
They both would because the bacteria promoter is inserted – it will be inserted in eiter orientation

16

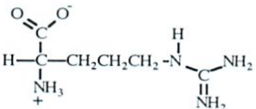
Name or initials__________ 10/31

Quiz 2

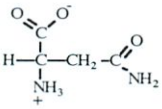
STRUCTURES OF AMINO ACIDS at pH 7.0



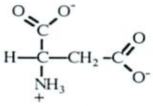
ALANINE
(ala)



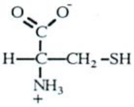
ARGININE
(arg)



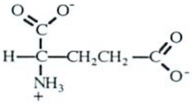
ASPARAGINE
(asn)



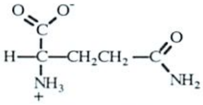
ASPARTIC ACID
(asp)



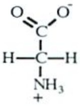
CYSTEINE
(cys)



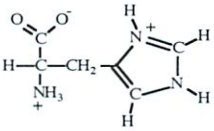
GLUTAMIC ACID
(glu)



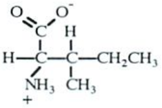
GLUTAMINE hydrop
(gln)



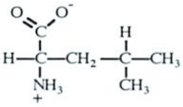
GLYCINE
(gly)



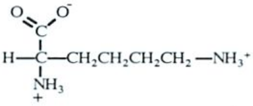
HISTIDINE
(his)



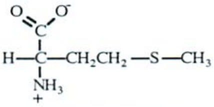
ISOLEUCINE
(ile)



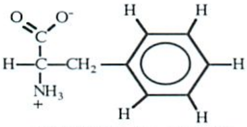
LEUCINE
(leu)



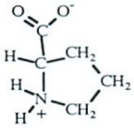
LYSINE
(lys)



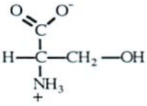
METHIONINE
(met)



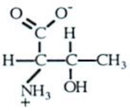
PHENYLALANINE
(phe)



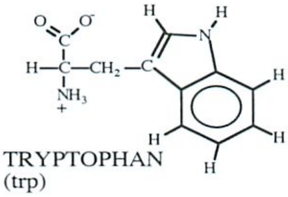
PROLINE
(pro)



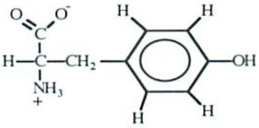
SERINE hydrop
(ser)



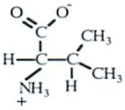
THREONINE
(thr)



TRYPTOPHAN
(trp)



TYROSINE
(tyr)



VALINE
(val)

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

10/31

Think ~~we~~ I did much better at this

Doing pratice problems helped a lot

Figured one out at the end

Prediction 80 B+

Some specific bio keywords / defn I forgot

Didn't focus on studying

But think I've shown I can do it

7.012 Quiz II

Mei 71

Class Average = 69
Median = 18 (?!

Score on exam	Approximate letter grade	% of class with this grade
85-100	A	23
67-84	B	37 ←
50-66	C	22
35-49	D	14.8
0-34	F	3.2

) F was much lower

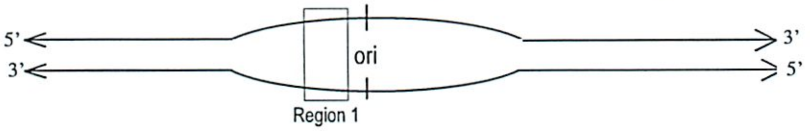
Solutions to 7.012 Quiz II, Version A

Class Average = 69
SD = 18

Score on exam	Approximate letter grade	% of class with this grade
85-100	A	23
67-84	B	37
50-66	C	22
35-49	D	14.8
0-34	F	3.2

Question 1 (17 points)

Consider the following segment of DNA (that is a part of a much larger molecule constituting a chromosome).



a) The DNA sequence in region 1 is :

```
5' GCCATG 3'
3' CGGTAC 5'
```

Give the sequence of a 6 nucleotide **RNA** primer that would occur at <u>region 1</u> during replication and be elongated to form the leading strand.

5′ C A U G G C 3′

b) If primase activity is absent, you would expect which of the following? Put a check next to the **best** answer.

___A decrease mRNA production.

___A decrease tRNA production.

_√_A decrease in both lagging and leading strand production.

___A decrease in lagging strand production.

___A decrease in leading strand production.

c) If DNA polymerase lost its 3′ → 5′ exonuclease activity, you would expect which of the following? Put a check next to the **best** answer.

___An increase in the fidelity of both lagging and leading strand production.

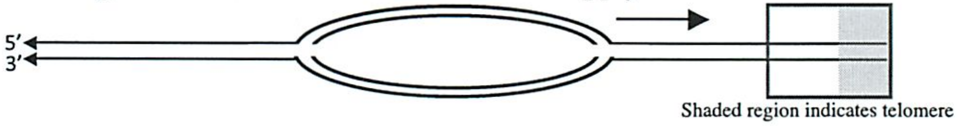
_√_A decrease in the fidelity of both lagging and leading strand production.

___A decrease in the speed of both lagging and leading strand production.

___A decrease in the fidelity of lagging strand production.

___A decrease in the fidelity of leading strand production.

d) Consider the following schematic of a chromosome. Imagine the replication fork expanding to the right and approaching the end of the chromosome. To generate a complete copy of the original chromosome in the boxed region, which of the following statements is correct? Check **all** that apply.



___ DNA polymerase but not telomerase is needed when using the top strand as a template

_√_ DNA polymerase but not telomerase is needed when using the bottom strand as a template

___ The enzyme telomerase but not DNA polymerase is needed when using the top strand as a template

___ The enzyme telomerase but not DNA polymerase is needed when using the bottom strand as a template

_√_ Both DNA polymerase and the enzyme telomerase are needed when using the top strand as a template

___ Both DNA polymerase and the enzyme telomerase are needed when using the bottom strand as a template

Question 2 (16 points)

Shown below is a double-stranded bacterial (*E. coli*) DNA sequence coding for the beginning of a hypothetical protein. Both strands are shown. The nucleotides are arbitrarily numbered from 1 to 90. The promoter sequence is italicized and underlined. **For this problem, transcription begins at the 10th base pair after the end of the promoter (i.e., at either base pair 20 or base pair 70, shown in bold).**

```
              10        20        30        40        50        60        70        80        90
               |         |         |         |         |         |         |         |         |
5' ...GTGTCGGAAACATATTGTGGGATGTTATATTGACAACGTCAAGGCACCATCAAATATGATCAGGATCATATTCTGCTGGGGCATAGGCGG...
      ---------+---------+---------+---------+---------+---------+---------+---------+---------+
3' ...CACAGCCTTTGTATAACACCCTACAATATAACTGTTGCAGTTCCGTGGTAGTTTATACTAGTCCTAGTATAAGACGACCCCGTATCCGCC...
```

a) The mRNA transcribed from this sequence encodes the first few amino acids of the protein. Which strand is used as a **template for transcription**, the top or the bottom? *TOP*

b) What are the first 10 nucleotides of the resulting mRNA?

5′ C C A C A A U A U G 3′

c) Give the first three amino acids of the peptide encoded by this gene. Label the N and C termini.
N- Met – phe – pro - C

The normal protein encoded by a different gene is 500 amino acids long. Below is sequence of the coding region of this new gene beginning at amino acid 100. The bottom strand is used as the template for transcription, and the underlined nucleotides represent the codon for amino acid 100.

```
5′ GCTAATACCTGCAACTATATAAACACCCAC 3′
3′ CGATTATGGACGTTGATATATTTGTGGGTG 5′
```

d) What is the primary sequence of amino acids 100-103?

	100	*101*	*102*	*103*
Amino acid:	*ala*	*asn*	*thr*	*cys*

e) You have discovered mutant 1 that carries a mutation within this region as shown below in bold.

original: 5′ GCTAATACCTGCAACTATATAAACACCCAC 3′
Mutant 1: 5′ GCTAATACCTG**A**AACTATATAAACACCCAC 3′

Would mutant 1 produce a protein that functions like the normal protein? Explain why you made this choice.
No, the substitution UGC → UGA replaces a cys with a stop codon. The protein will only be 102 amino acids long instead of the normal 500 amino acids long. This shorte vesion of the protein will not have the structure to support the normal function of the protein.

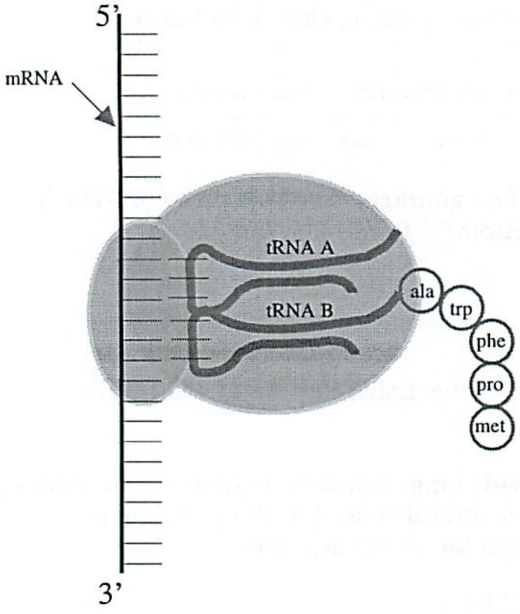
f) You have discovered a different mutant, mutant 2, that carries a different mutation that deletes three base-pairs, as shown below.

original: 5′ GCTAATACCTGCAACTATATAAACACCCAC 3′
Mutant 2: 5′ GCTA---CCTGCAACTATATAAACACCCAC 3′

Mutant 2 produces a protein that functions like the normal protein. Explain why this mutation does not alter protein function.
This three nucleotide deletion removes only 1 of the 500 amino acids, but because we have removed 3, the deletion does not change the reading frame, so the remaining amino acids remain the same. The single missing amino acid must not play a crucial role in the overall structure of the protein and thus does not change the function of the protein.

Question 3 (13 points)

Below is a schematic of a ribosome actively translating an mRNA. The vertical line represents the mRNA, and each small horizontal line represents a ribonucleoside, and two tRNA molecules are labeled.

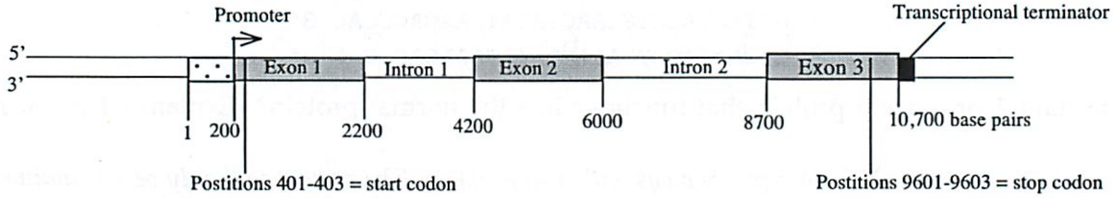

5'
mRNA
tRNA A
tRNA B
ala
trp
phe
pro
met
3'

a) Label the 3′ and 5′ ends of the mRNA.

b) Indicate a sequence that could be the anticodon of tRNA A.

5′ C C *A* 3′

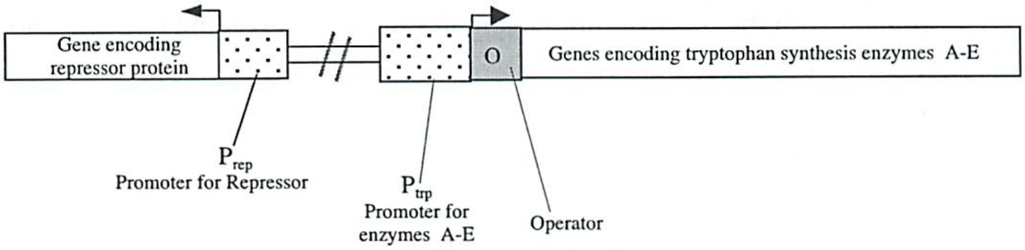
c) Below is a schematic of gene Y, which encodes protein Y. The promoter region is indicated by the dotted box. Transcription begins immediately following the promoter.


Promoter
Transcriptional terminator
5'
3'
Exon 1
Intron 1
Exon 2
Intron 2
Exon 3
1
200
2200
4200
6000
8700
10,700 base pairs
Postitions 401-403 = start codon
Postitions 9601-9603 = stop codon

- The pre-mRNA produced by this gene would be approximately how many nucleotides long?
  10,500
- Two mature mRNAs of different lengths are produced from this gene. Name or describe the process by which this gene could produce two different mature mRNAs each that encode a different protein.
  In generating the mature mRNA, the pre-mRNA could be spliced to include all three exons or alternatively, it could be spliced to include only exons 1 and 3.

Question 4 (12 points)

You design a summer class where you recreate experiments studying the *trp* operon in *E. coli* (see schematic below). Mutants missing one of the enzymes A-E are tryptophan auxotrophs.


Gene encoding repressor protein
O
Genes encoding tryptophan synthesis enzymes A-E
P_{rep}
Promoter for Repressor
P_{trp}
Promoter for enzymes A-E
Operator

a) When would you expect the production of enzymes A-E in a wild-type cell. Check all that apply.

___ When the levels of tryptophan in the cell are high
_√_ When the levels of tryptophan in the cell are low
_√_ When the cell is grown on minimal media.

b) To which of the following would the trp repressor bind? Check all that apply.

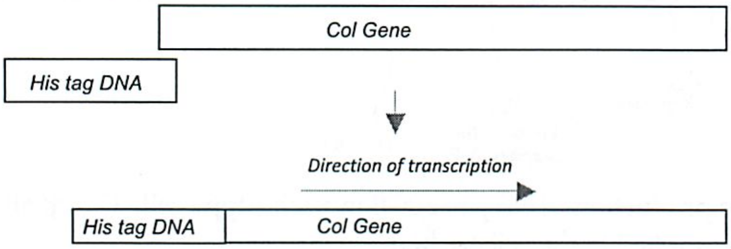
___ Lactose
___ P_{rep}
_√_ Operator
_√_ Tryptophan
___ glucose

c) A loss of which component or components would produce a cell unable to grow on minimal media? Check all that apply.

___ Gene for repressor
___ P_{rep}
___ Operator
_√_ P_{trp}
_√_ gene for enzyme A
_√_ gene for enzyme E

Question 5 (14 points)

You have discovered a human gene Col, encoding the COL protein and associated with a colon disease. You plan to create a recombinant DNA sequence, by ligating a DNA sequence that encodes 5 histidines followed by a few additional amino acids in front of the protein-coding sequence for the Col gene. This will produce a slightly longer protein (see diagram below). You want to be sure that the combined sequence beginning with the five histidines maintains the correct reading frame to allow proper translation of the Col protein.



Below is the sequence that encodes the **histidine tag.** The bars above the sequence show the restriction enzyme recognition sites. The sequence encoding the start codon is shown in bold.

BglII *KpnI* *EcoRI*

E. coli Promoter

..TT**ATG**CATCATCATCATCATAGATCTTGGTACCGAATTCTTAAGATCC ... 3′

---------+---------+---------+---------+--------

..AA**TAC**GTAGTAGTAGTAGTATCTAGAACCATGGCTTAAGAATTCTAGG ... 5′

Below is part of the cDNA sequence that encodes the Col gene. The sequence encoding the start codon is shown in bold. The bars above the sequence show the restriction enzyme recognition sites.

BglII *KpnI* *EcoRI*

5′ ... TCAAGATCTCCGCGGTACCAAGAATTCGGATC**ATG**TTATAGCAAGCTCGGAAT... 3′

+---------+---------+---------+---------+---------+--------

3′ ... AGTTCTAGAGGCGCCATGGTTCTTAAGCCTAG**TAC**AATATCGTTCGAGCCTTA... 5′

The recognition sequences and the cleavage sites (indicated by /) for each enzyme are given below.

```
EcoRI:
5' G/AATT C 3'
3' C TTAA/G 5'

KpnI:
5' G/GTAC C 3'
3' C CATG/G 5'

BglII:
5' A/GATC T 3'
3' T CTAG/A 5'
```

a) You want to ligate these two pieces of DNA together to create a version of the COL protein that has 5 histidines attached to the N terminus in a manner that maintains the reading frame. Which enzyme or enzymes can you use to…

- cut the DNA encoding the **Histidine tag.** List all that apply. *Eco RI*
- cut the **Col gene**? List all that apply. *Eco RI*

b) Give the DNA sequence of the top strand (as seen in diagram above) that can be found at the junction site between the His tag and the Col gene.

5' *G G|T A C|C* *G A|A T T|C* *G G|A T C|A* 3'

6 nucleotides of HIS tag before RE site — 6 nucleotides of the restriction enzyme (RE) site — 6 nucleotides of Col gene after RE site

c) In the sequence of part (b) draw vertical lines between each codon that corresponds to the amino acids of the His tag-COL fusion protein.

Question 6 (28 points)

You successfully create a DNA fragment that encodes the histidine-tagged human COL protein. You plan to clone this fragment into a vector that will allow you to express it in bacterial cells.

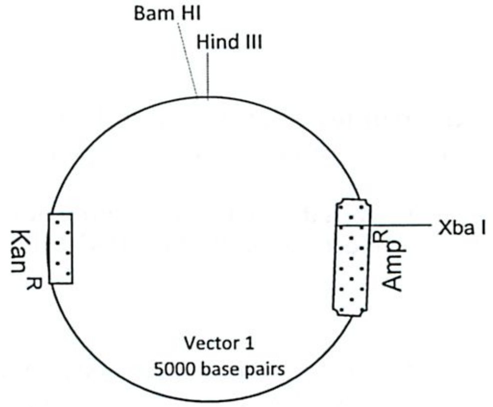
Your plan is to:

1) Cut an appropriate vector and the DNA fragment with *Xba I*.
2) Ligate the cut vector and the fragment together.
3) Transform *E. coli* cells with the ligation mix.
4) Select for *E. coli* cells that have a plasmid.
5) Identify the *E. coli* cells carrying a recombinant plasmid containing the inserted fragment by screening.

The following is a partial schematic of vector 1 that will allow you to complete the plan outlined above.

The Kan^R gene confers resistance to the drug kanamycin.

The Amp^R gene confers resistance to the drug ampicillin.



a) To allow selection for *E. coli* cells that have any plasmid (step 4), and screening for *E. coli* cells with the recombinant plasmid containing the inserted fragment (step 5), you will transform a particular strain of *E. coli* cells. What will be the phenotype of this strain prior to transformation?
The phenotype of this strain prior to transformation should be sensitive to both ampicillin and kanamycin.

b) To allow selection for *E. coli* cells that have either vector 1 or a recombinant plasmid, you will plate the transformation mix on media that contains which the following drugs?

Ampicillin [✓ Kanamycin] Both ampicillin and kanamycin Neither ampicillin or kanamycin

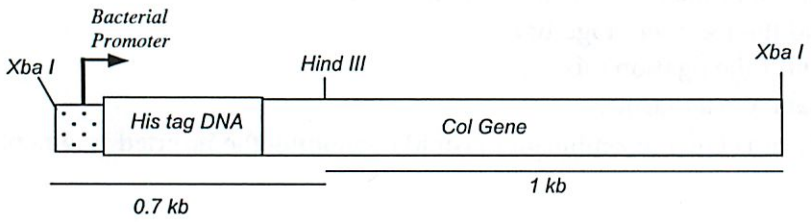
c) After selection you need to distinguish the colonies that contain cells with a recombinant plasmid from the colonies that contain cells with the original vector 1. To do this, you would replica plate from your selection plate onto media that contained *Ampicillin.*

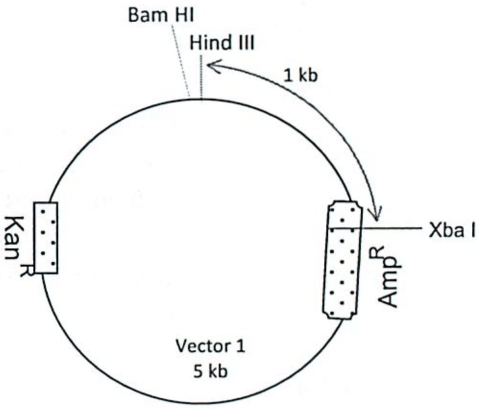
Colonies that contain cells with a recombinant plasmid will *die* on this media. (Live or Die)

Colonies that contain cells with the original vector 1 will *live* on this media. (Live or Die)

Question 6, continued

You plan to insert he histidine-tagged COL gene into vector 1 as an Xba I fragment (See diagram below). You will obtain two different recombinant plasmids.





d) To differentiate between the two possible recombinant plasmids, you can cut both with restriction enzyme(s) and separate the resulting fragments by gel electrophoresis.

- What restriction enzyme or enzymes would you use to differentiate between the two possible recombinant plasmids? *Hind III*

- Given your answer above, what sized DNA fragments should you see from the restriction enzyme digestion of the two plasmids?

  one recombinant plasmid would give: *1.7 + 5*

  the other recombinant plasmid would give: *2 + 4.7*

- Which of these recombinant plasmids would allow expression of the histidine-tagged COL protein in bacterial cells? Explain your answer.
  Either of these recombinant plasmids would allow expression of the histidine-tagged COL protein in bacterial cells because the inserted fusion gene carries an appropriate promoter.