Foundations in Biomedical Sciences – Lecture 2b

Gordon Rule
Department of Biological Sciences
Carnegie Mellon University

rule@andrew.cmu.edu

Quaternary Structure

Protein Purification Methods

- Affinity chromatography
- Ion exchange
- Size exclusion

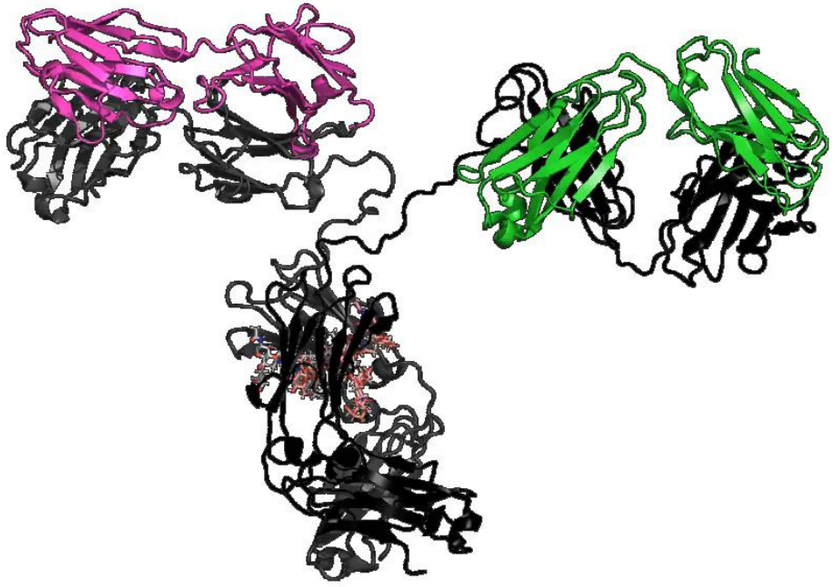
Expression of Recombinant Proteins

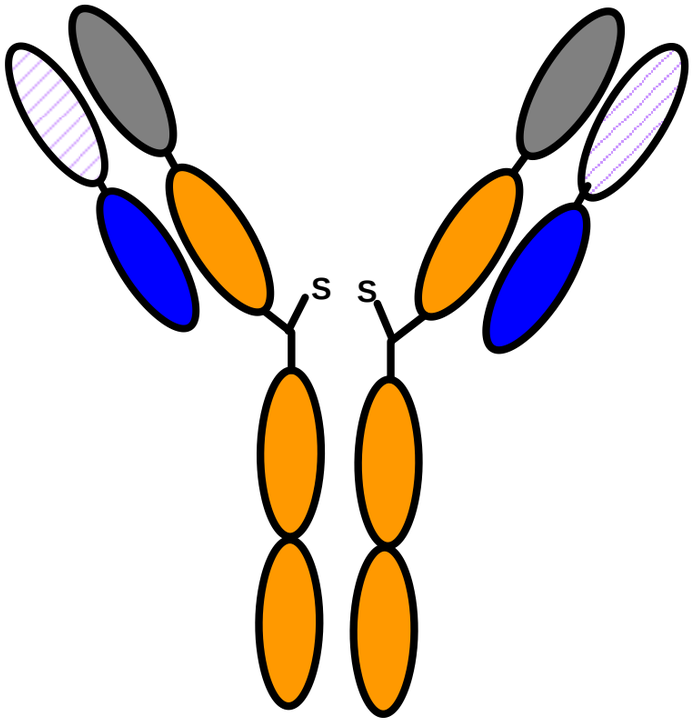
- Transcription and Translation
- E. Coli and T7 RNA Polymerase
- Lac operon as a control system
- pET vectors
- Restriction enzymes & DNA Ligase

Quaternary Structure

- Combinations of polypeptide subunits (combinations of tertiary structures).
- Chains associate by ***non-covalent interactions*** between R groups on the different chains.
- *Additional* stabilization can be provided by disulfide bonds, usually in extra-cellular proteins.
- Proteins can be a dimer, trimer, tetramer, pentamer, etc.

- If the chains are the same, called homo_____. If chains are different, hetero_____





Quaternary structure of Antibody

- two light chains are identical
- two heavy chains are identical

- heterotetramer held together by non-covalent forces & disulfide bonds

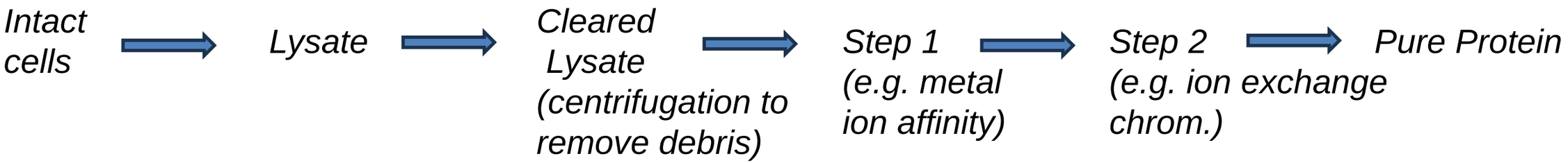
- Antibodies bind two antigens at the top of the Y.

Protein Purification.

Overall Expectations:

1. Describe overall process of column chromatography
2. Understand separation process of:
   a. Size exclusion chromatography (gel filtration) (***Native*** MW)
   b. Cation exchange (+ charged proteins)
   c. Anion exchange (- charged proteins)
   d. Affinity chromatography (ligand, antibody)
   e. Tags (his-tag), genetically encoded.
3. Understand how different components of SDS-PAGE work:
   - Electrophoretic migration: v = q/m
   - Role of SDS: uniform charge to mass ratio -> separation by size only.

Multi-step Purification:



Typical Purification Methods & Their Prevalence

1. Selective precipitation of proteins by ammonium sulfate:
   - Advantage was being able to handle very large volumes of lysate from natural sources
   - Disadvantage is relatively poor resolution (separation of similar proteins from another)
   - Seldom used in the era of recombinant DNA expression of target protein

2. Column Chromatography – separation by size, charge, tags, selective affinity
   - Very scalable (mg to gms)
   - High capacity
   - High resolution (easier to separate similar proteins)

Types of Column Chromatography & Their Frequency of Use

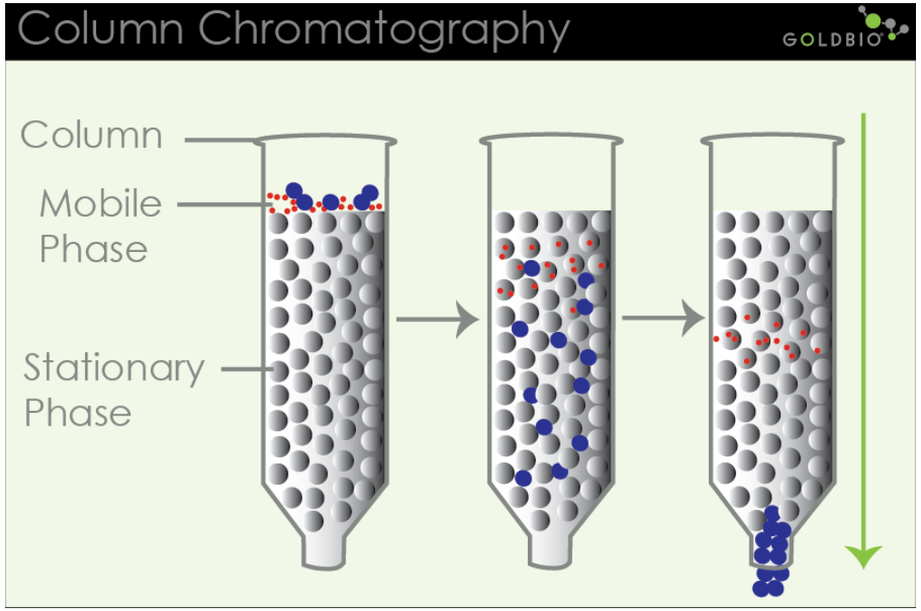
Method	Separation by	Frequency of Use
Metal-ion affinity	Tagged with6 His residues = HisTag	Extremely common (95%)
Maltose and glutathione (GSH) affinity chromatography	Tagged with maltose binding protein or glutathione transferase	Somewhat common (4%)
Ion-exchange chromatography	Charge	Very common 2nd step (70%)
Size exclusion chromatography	Size	Somewhat common 2nd step (30%)
Antibody affinity chromatography	Binding to specific antibody	Relatively rare
Ligand based affinity chromatography	Binding to specific ligand	Targeted to particular classes of enzymes
Hydrophobic affinity chromatography	Polarity of protein surface	Rare

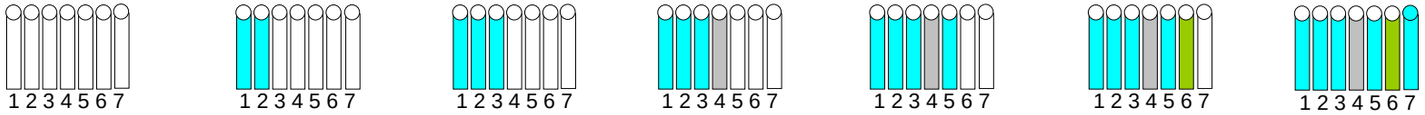
Column Chromatography

- Chromatography is performed in long glass tubes filled with a matrix or resin (particle size similar to fine sand).
- The liquid is a buffered salt solution to keep the proteins in their native (folded) form.
- The mixture of proteins is added to the top of this column and buffer is allowed to flow through the column.
- As the buffer flows through the column the mixture of proteins is drawn down through the column and interacts with the resin. The actual mode of separation depends on the nature of the resin.
- The protein ***elutes*** from the column at a specific volume (V_e) at a specific time.
- The elution volume is divided into separate **fractions** (e.g. separate test-tubes)
- Usually, several different chromatographic steps are performed with different resins, sequentially, during a purification scheme.

Different resins allow separation by:

- Affinity (Metal ions for HisTag)
- Charge (ion exchange)
- Size (size exclusion/gel filtration)





His-Tags – The most common Affinity Tag

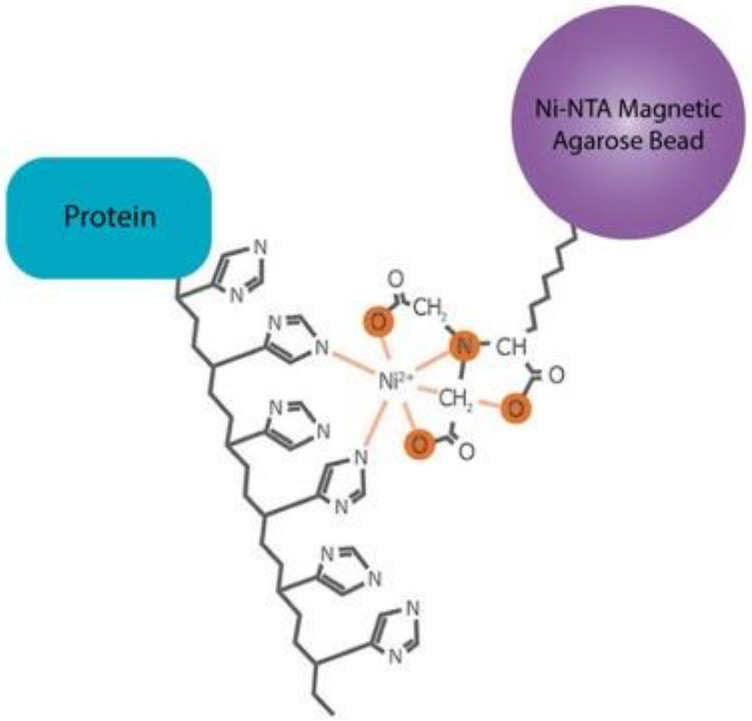
- A His-tag is the addition of 6-10 Histidine residues to either the N-terminus or the C-terminus of the target protein.
- The tag has to be added by changing the DNA sequence of the target protein.
- The His residues will cause the tagged protein to stick to immobilized Ni ions on a resin bead.
- The impurities can be washed away, and the tagged protein released by adding a high concentration of imidazole. The imidazole will compete for the His and bind to the Ni ions.

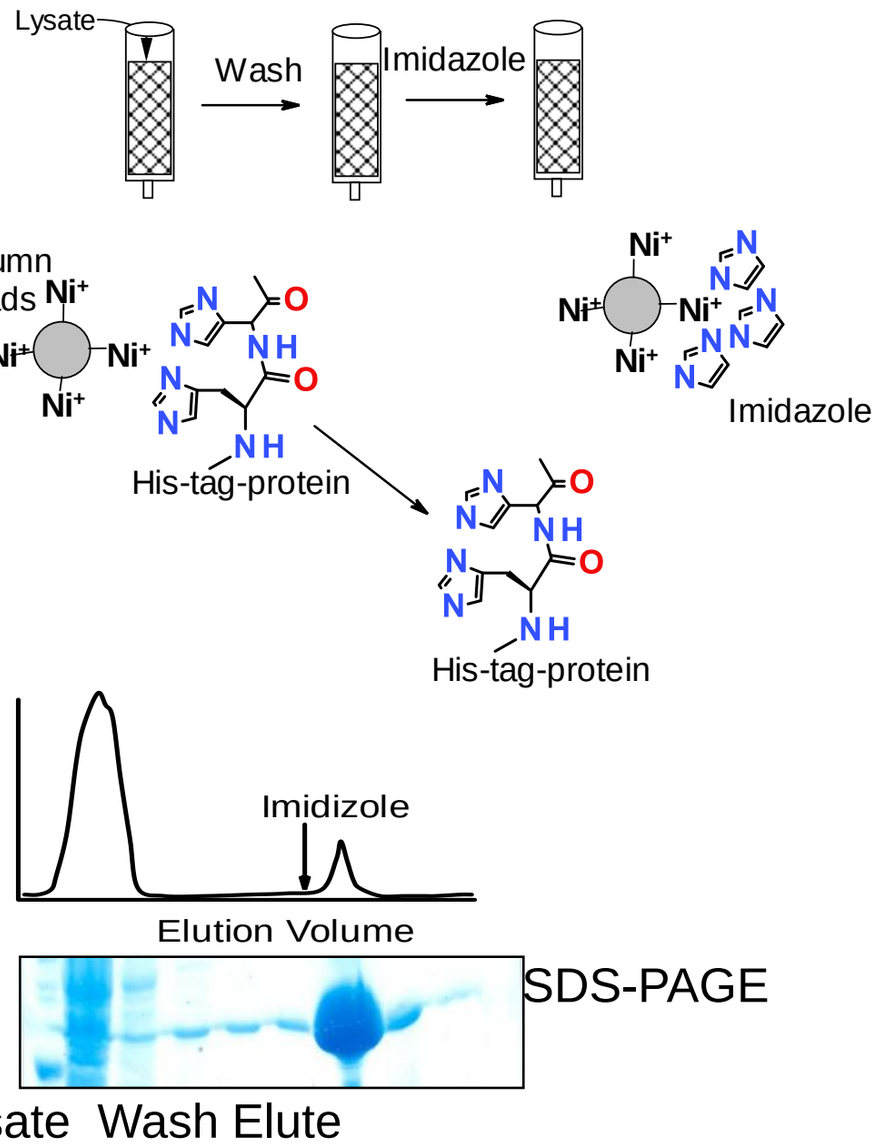
Original DNA/protein sequence:

```
-ATGCCCGGC--
 MetProGly--
```

Modified DNA/tagged protein sequence:

-ATG**CACCACCACCACCACCAC**CCCGGC--
 Met**HisHisHisHisHisHis**ProGly--





Ion Exchange - Effect of pH on Net Charge of Proteins

pI (Isoelectic pH)- pH at which proteins have no *net* charge:

- Proteins become positively charged for pH < pI
- Proteins have a negative charged if pH > pI.

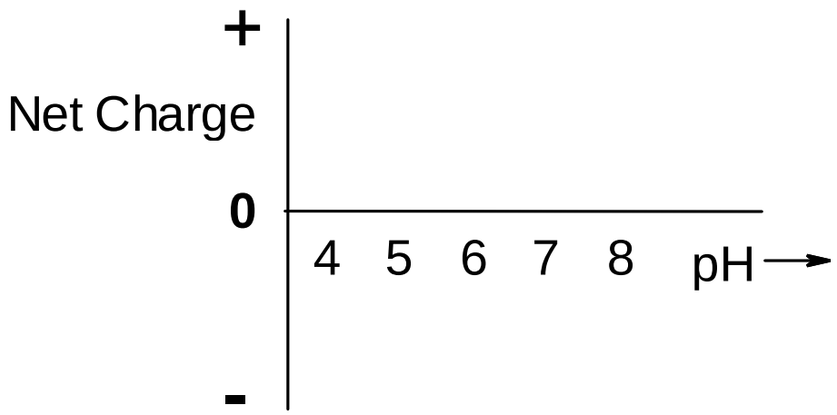
Therefore, the charge on a protein can be changed by altering the pH, increasing or decreasing the binding to ion exchange resins.

The pI can be predicted from the amino acid sequence using the same charge calculation that we used earlier:

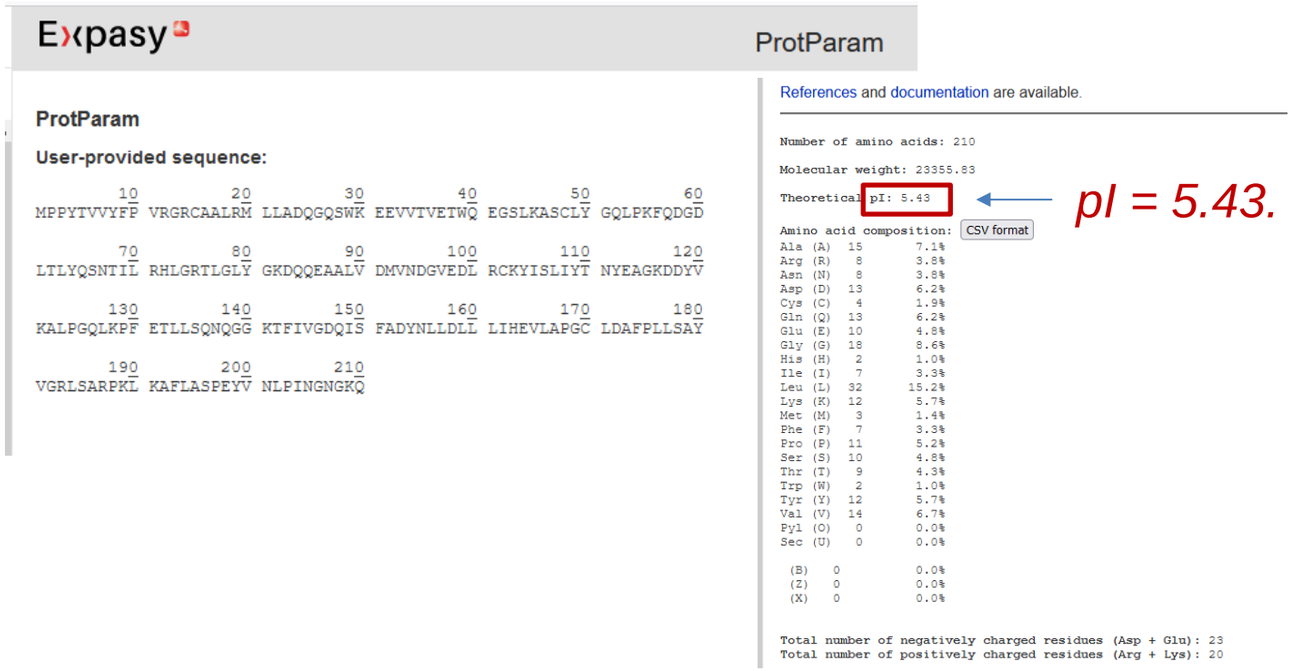
$$\langle q \rangle = \sum_{i=1}^{n} f_A \times q_A + f_{HA} \times q_{HA}$$

and finding the pH value where the charge is zero.

This calculation can be done by web-servers.



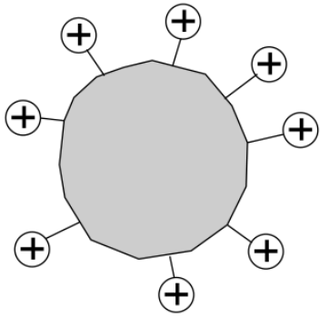
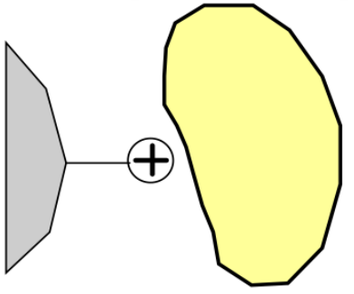
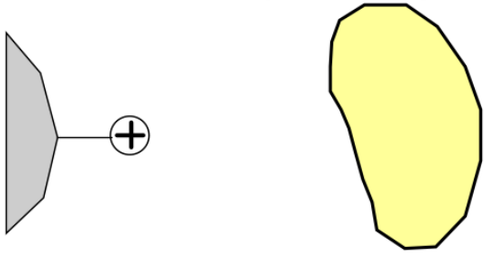
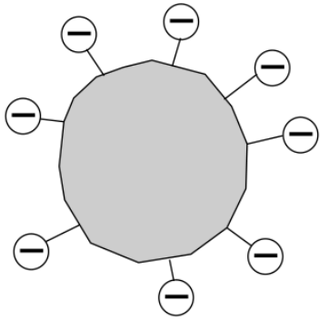
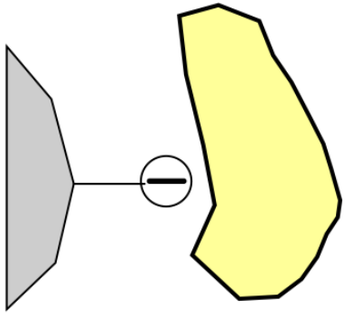
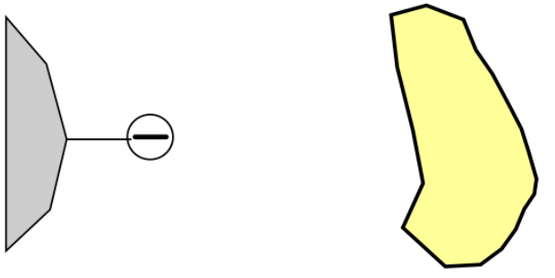
https://web.expasy.org/protparam/



pI = 5.43.

1. *What is the charge on this protein at pH=5.43?*
2. *What is the sign of the charge on this protein at pH=7?*
   negative or positive

Types of Resins – Ion Exchange

Type	Resin	Principal of Separation	How to Elute the Protein
Anion Exchange *Separation by charge.*	• Beads with a **positive** charge 	Protein stick to resin because of: • Overall **negative** charge (*anions*) • Proteins have patches of negative charge 	• Wash, bound proteins **may** elute. • Increase salt concentration to weaken electrostatic interaction. • Change of pH to pH < pI (protein becomes positively charged) 
Cation Exchange *Separation by charge.*	• Beads with a **negative** charge 	Protein stick to resin because of: • Overall **positive** charge (*cations*) • Proteins have patches of positive charge 	• Wash, bound proteins **may** elute. • Increase salt concentration to weaken electrostatic interactions. • Change of pH to pH > pI (protein becomes negatively charged) 

Ion Exchange - Column Chromatography

Cation exchange resin, binds + charge ions.

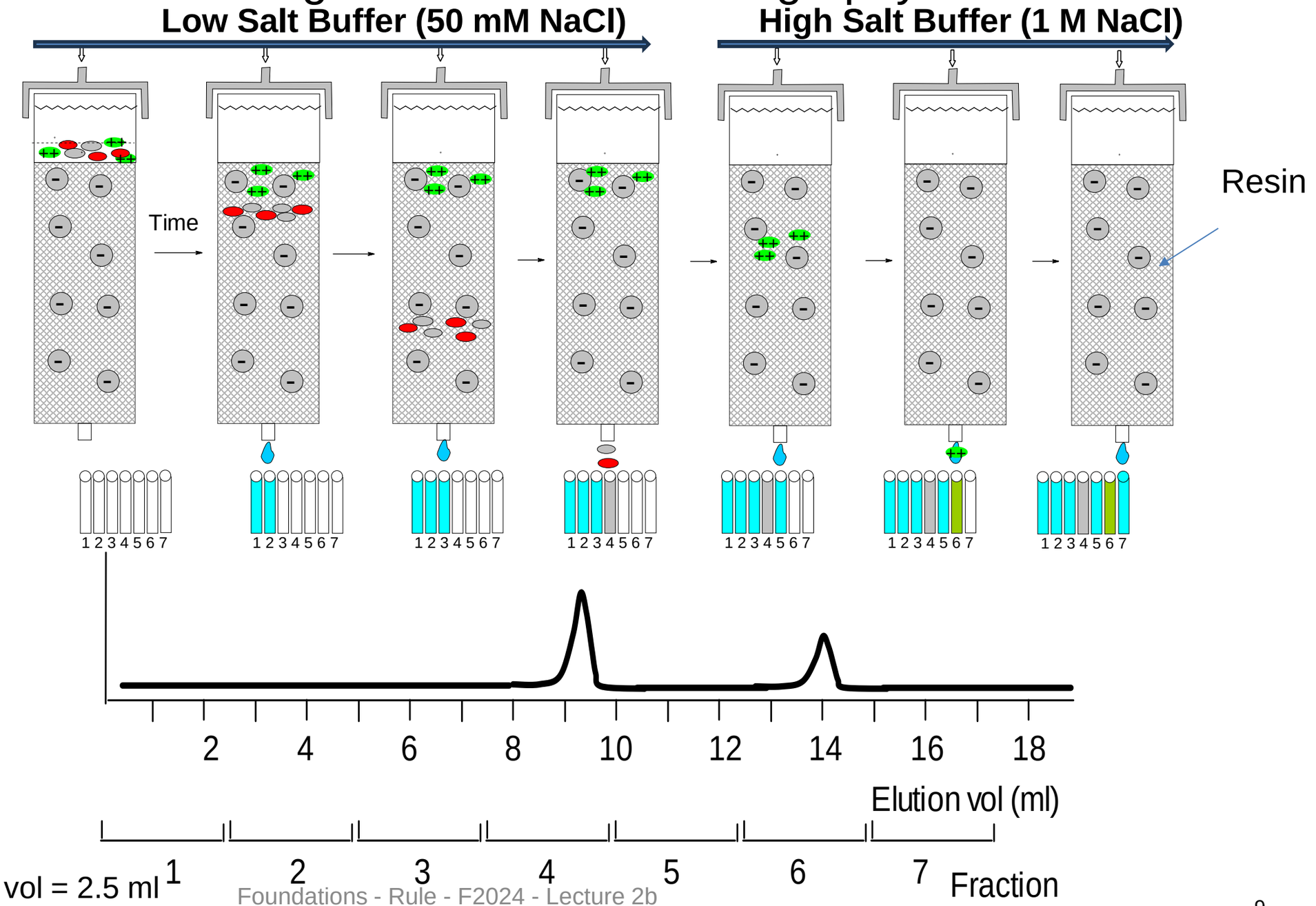
Mixture contains:

A. + Charge (green)
B. Neutral (gray)
C. - Charge (red)

Elution Volume: Volume that has dripped out of the column.

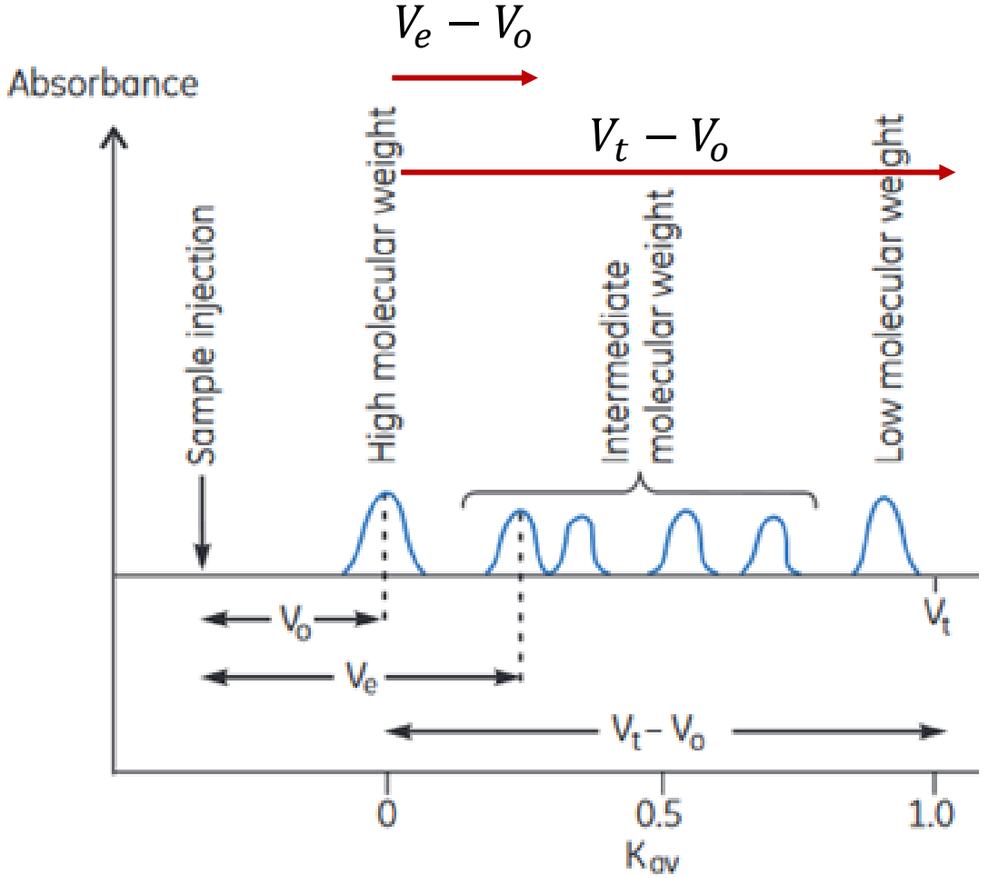
Fraction: Collection of liquid that drips out the bottom of the column – i.e. the total amount of liquid is *fractionated* into separate fractions.

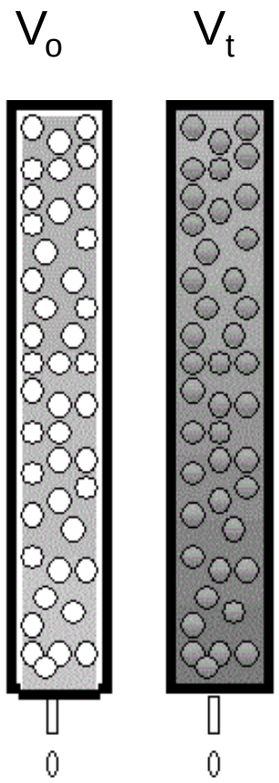
Fractions that contain the protein of interest are pooled (combined) before next step.



Separation by Size - Size Exclusion (aka Gel Filtration)

Size Exclusion (gel filtration) *Separation by size*		• Proteins don't "stick". • Small proteins enter the interior of the beads, and therefore take longer to wash off of the column	• Simply washing the column with buffer will eventually wash the proteins out of the column. • Smaller proteins elute last.





Fractional Elution $K_{AV} = \frac{(V_e - V_o)}{(V_t - V_0)}$

V_e = volume that peak elutes (exits column)

V_t = total column volume, ***accessible to small proteins.***

V_o = void volume, the volume ***outside*** the beads, this amount of liquid has to come through the column before any proteins can elute - ***volume accessible to large proteins***. (Typically ~30% of the total column volume)

V_t - V_0 = volume available for separation

The volume available depends on the MW:
Huge protein – V_o is the available volume, $V_e = V_o$
Tiny protein – V_t is the available volume, $V_e = V_t$

Evaluating Final Purity

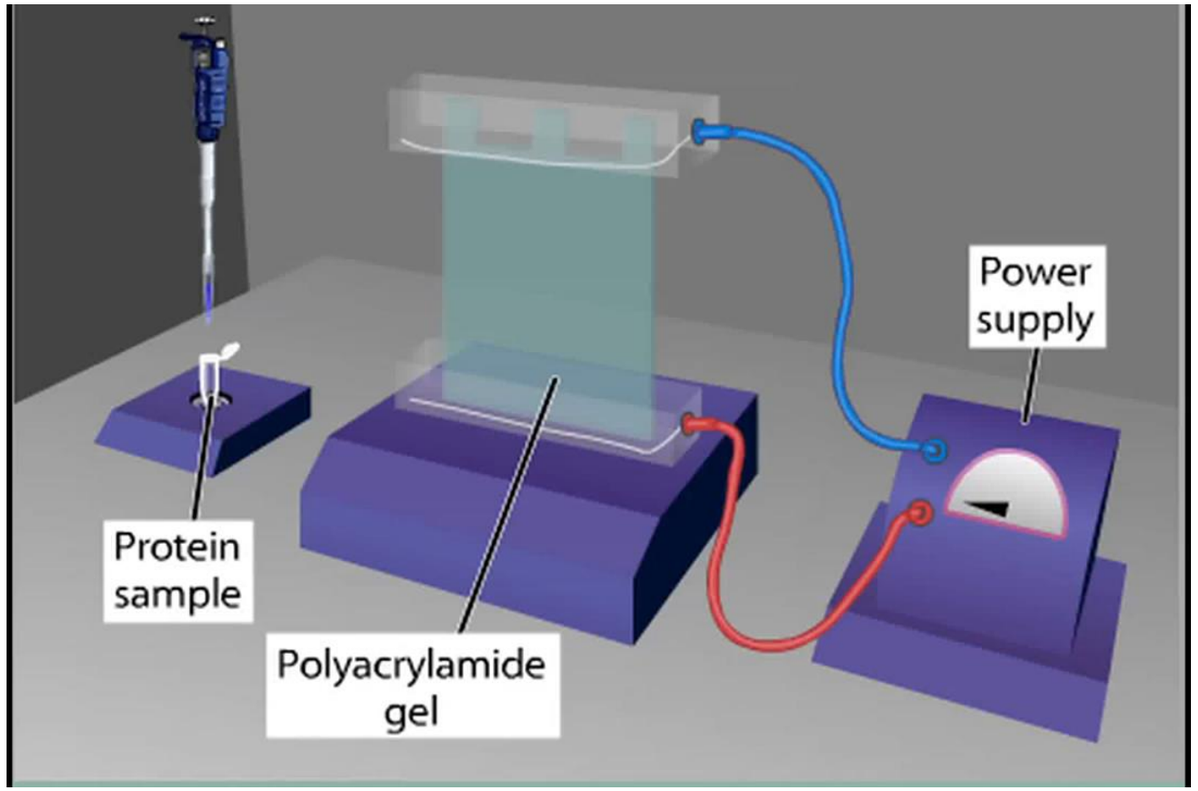
After the protein is purified, its purity can be monitored by:

a) SDS-page gel electrophoresis (most common)
b) Mass spectrometry (sometimes)
c) Amino terminal sequencing (rare)
d) Isoelectric focusing (separation by pI) (rare)

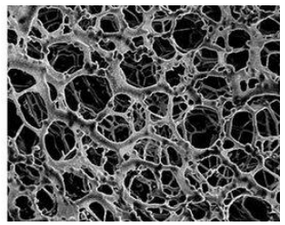
SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Electrophoresis = movement of charged molecules in an electric field: *velocity* $\propto$ *(q/M) x V* (V, volts)

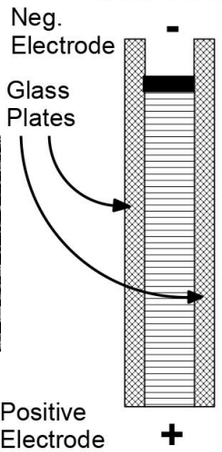
Denatured proteins are pulled through a polymer gel by electrophoresis, separating them solely by size (ideally).



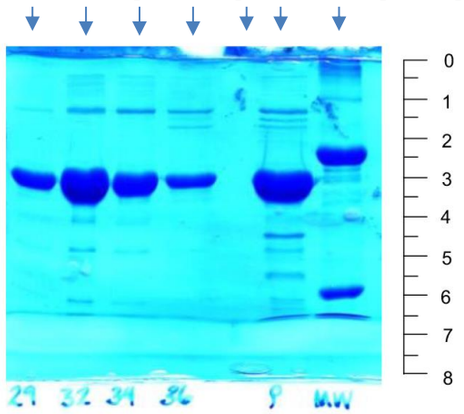
Top View of Gel



Cross-section

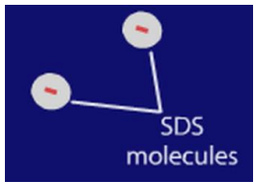


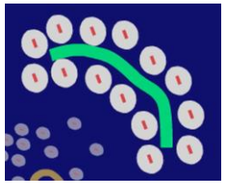
Lanes (multiple samples)

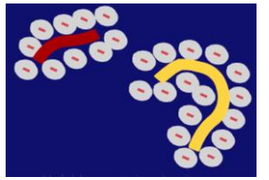


Based on the video:

1. *What happened to all proteins when interacting with SDS?*
2. *What happened to the protein with two chains, after interacting with SDS?*





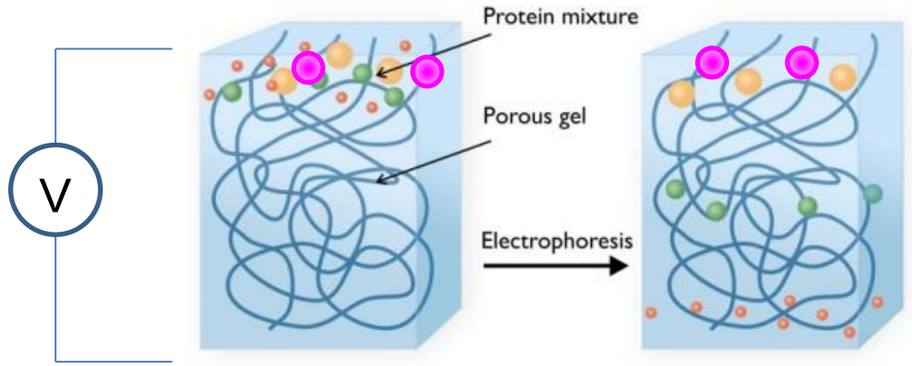


SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

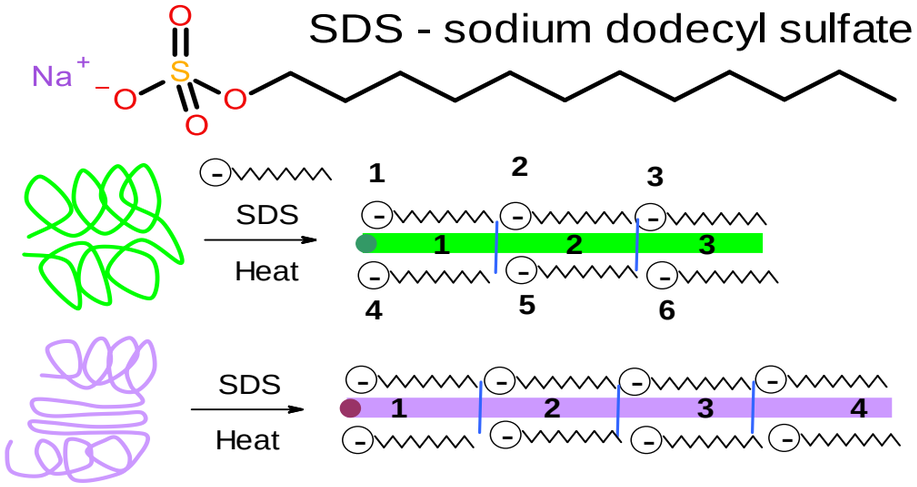
SDS – denatures proteins, giving them:

- a negative charge
- a uniform charge to mass ratio

- All proteins migrate in the same direction.
- Separate only by size as they pass through the gel (Smaller proteins move faster)

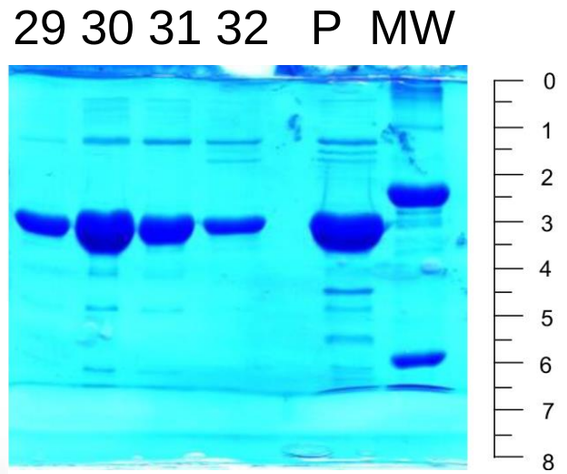


Yellow, green, orange + SDS = - charge
Magenta + SDS = - charge



Charge (q)	Mass (M)	v = q/M
6	3	2 = 6/3

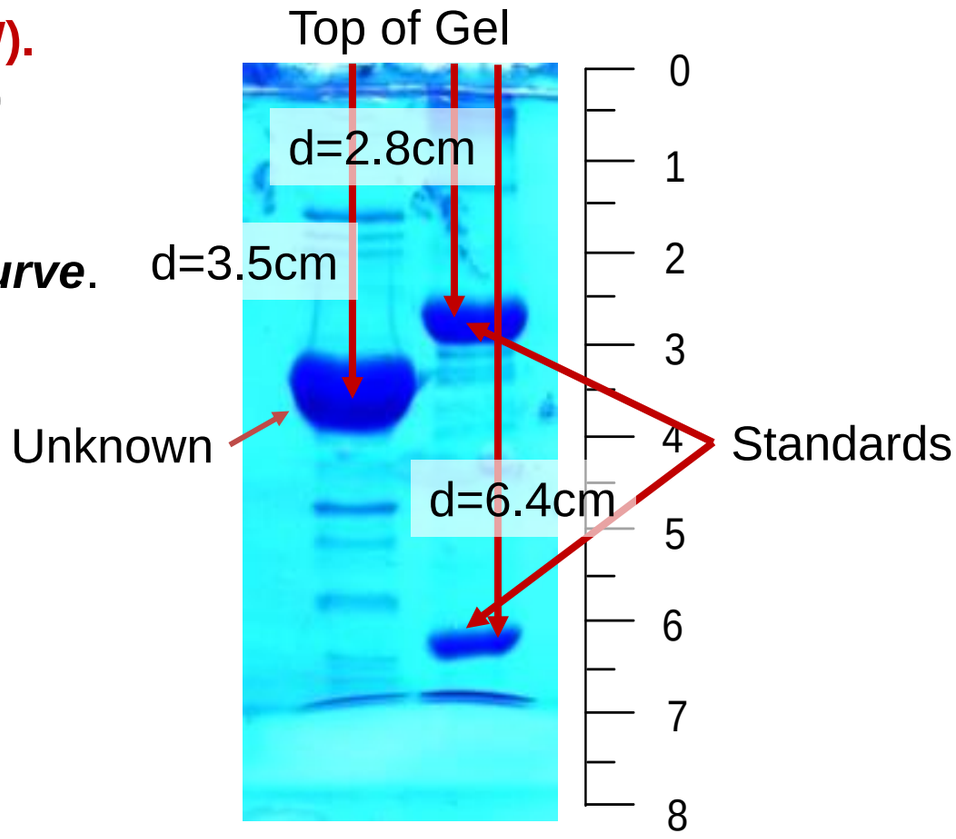
Gels are stained with a stain that is specific for protein. Most commonly used stain is Coomassie blue. Gel shows four fractions (#29, #30, #31, #32) obtained from chromatography, an empty lane, pooled fractions (P), molecular weight standards (MW).

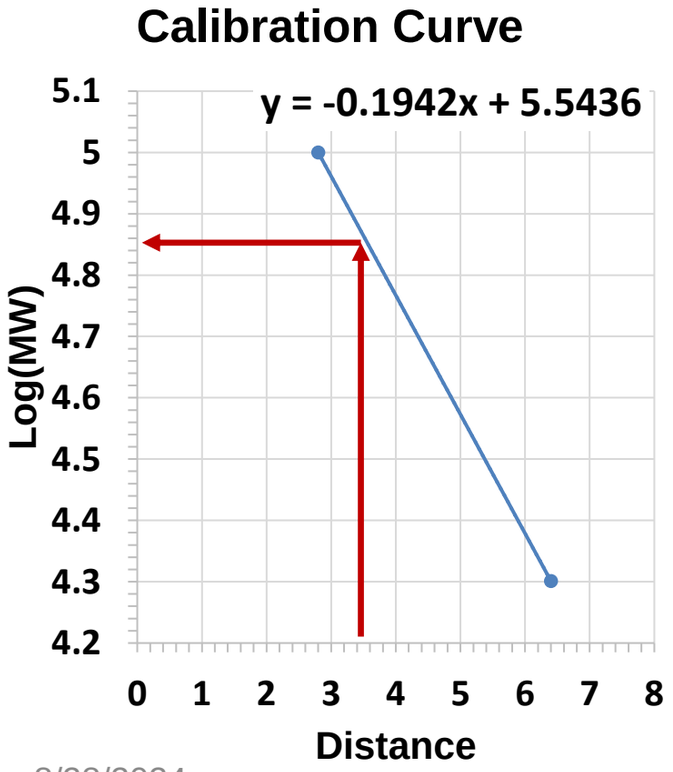


Measuring Denatured Molecular Weight With SDS-PAGE.

A. Molecular Weight Determination – distance migrated ~ ∝ log(MW).

i) Load molecular weight standards on one lane (right lane in above gel)
ii) load unknown(s) in other lane (left lane), apply voltage.
iii) Measure the distance migrated of each band.
iv) Plot log(MW) versus *d* for standards, generating a ***calibration curve***.
This is fit to a polynomial function e.g. $log(MW) = \alpha d^2 + \beta d + \gamma$.
v) Use calibration curve to get logMW of unknown.





Obtaining MW

1. Analytical Solution (better)
Log(MW) = -0.1942 x d + 5.5436
d = 3.5 for unknown
Log(MW) = 4.86
MW = $10^{4.86}$ = ~ 73,100 Da

2. Graphical solution
i. Sketch calibration curve using known MW (standards)
ii. Find intersection of line for d=3.5
iii. Read log(MW) from y-axis

Standards
S1: MW 100,000, distance = 2.8 cm
S2: MW 20,000, distance =6.4

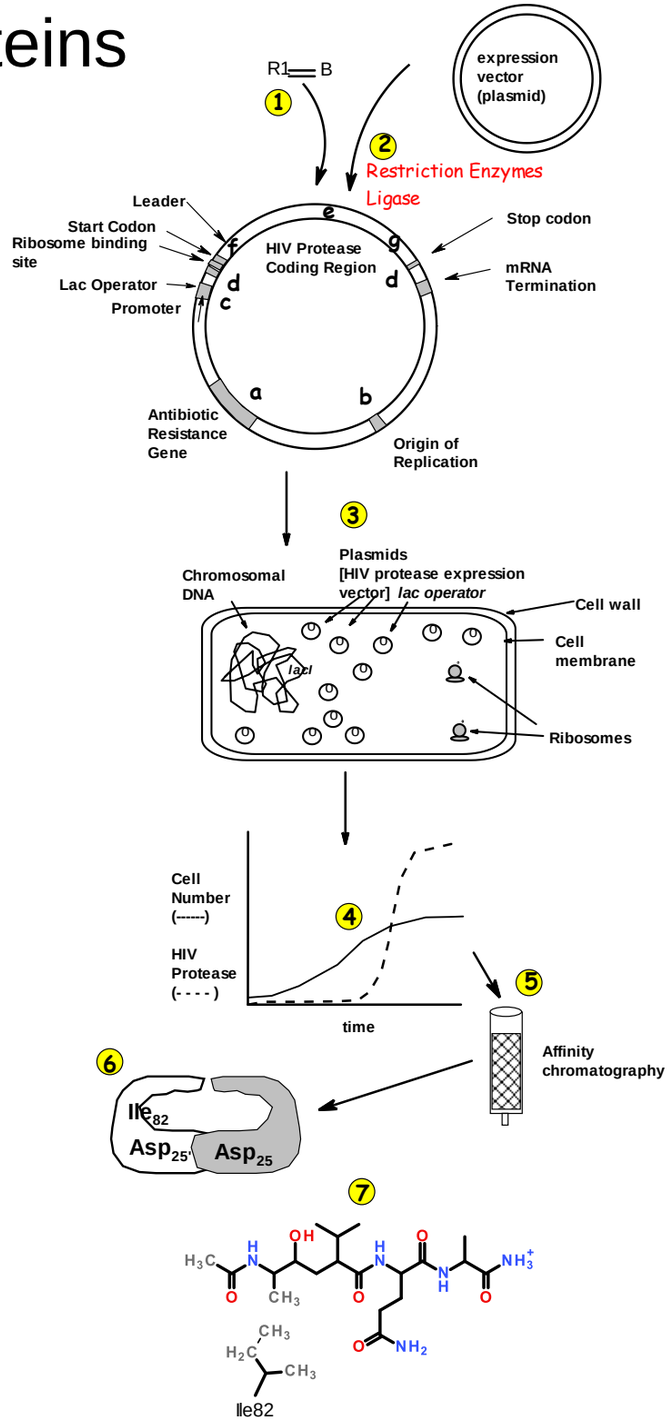
Unknown: Distance =3.5

Production of Recombinant Proteins

The Goal: To produce any desired protein in in E. coli (a widely used bacteria) using recombinant DNA methods.

The overall procedure is as follows:

1) Obtain the codons for the desired protein:
   - PCR from native source
   - Chemical synthesis of gene.
2) Insert the codons into a bacterial expression vector (a specialized form of a DNA plasmid) using restriction endonucleases and DNA ligase.
3) Transformation of the bacteria with the plasmid.
4) Growth of the transformed bacteria, production of the recombinant protein.
5) Purification of the recombinant protein.
6) Structure determination
7) Biochemical studies, e.g. rational drug design, enzyme mechanism



Properties of Expression Vectors

Common Required Features:

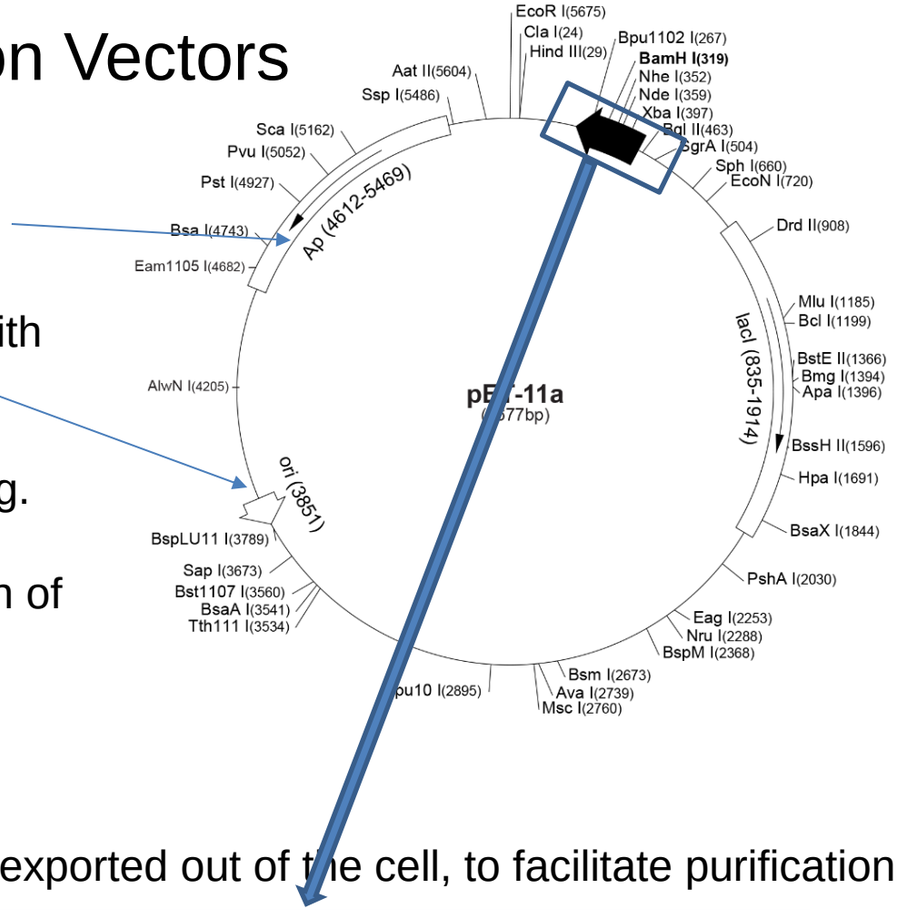
a) Provide antibiotic resistance to the host bacteria, such that only bacteria containing the plasmid will grow in the presence of the antibiotic.

b) An origin of DNA replication so that the plasmid will be replicated with the bacterial DNA

c) DNA sequences that cause the production of mRNA, copying the information in the DNA to mRNA, including a regulated promotor (e.g. lac operator).

d) Sequences in the resultant mRNA that start and stop the production of the recombinant protein.

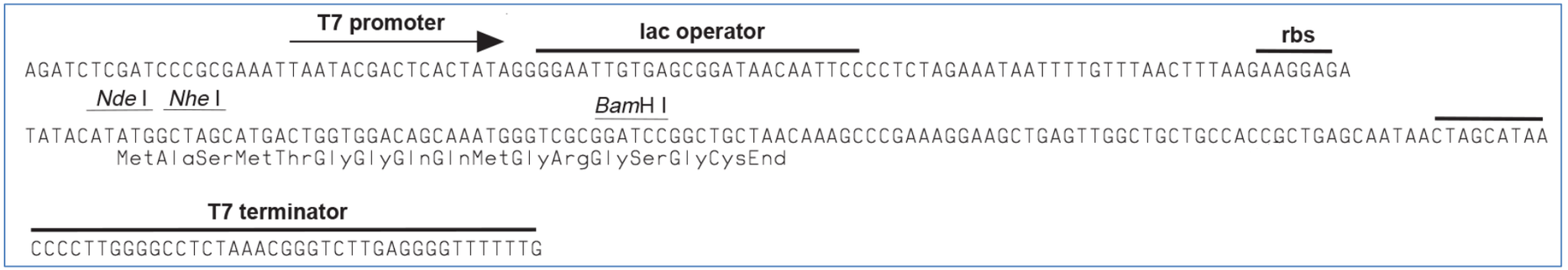
e) Coding region for protein to be expressed.



Optional features on some vectors:

a) His6 tag for affinity purification on a Ni column.

b) Sequences (pelB leader) that cause the recombinant protein to be exported out of the cell, to facilitate purification.

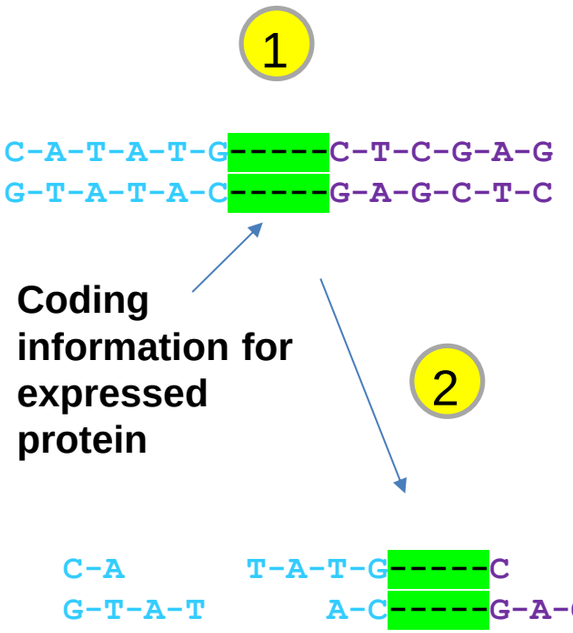


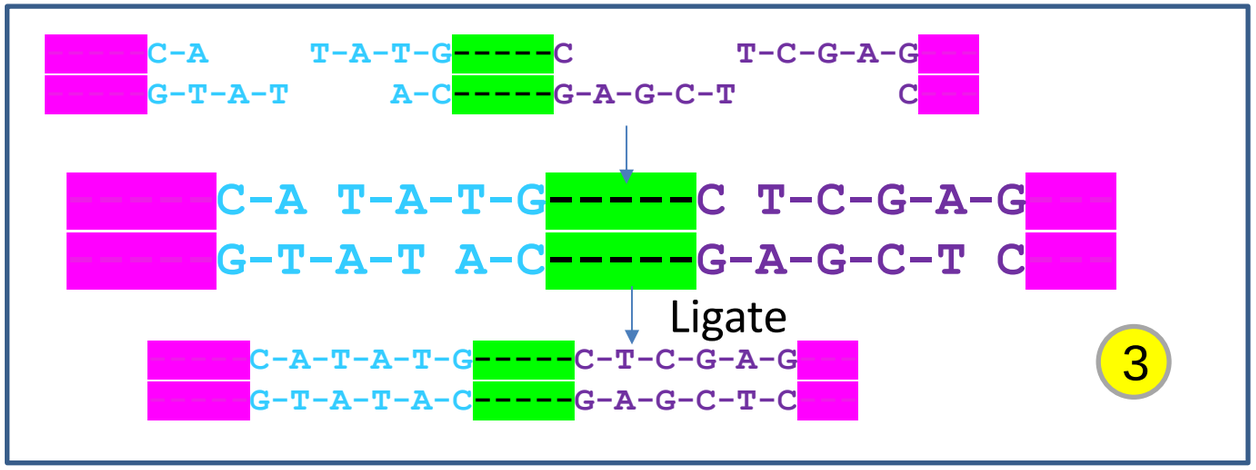
Gene for protein to be expressed inserted by restriction digest of insert & plasmid, mix & ligate.

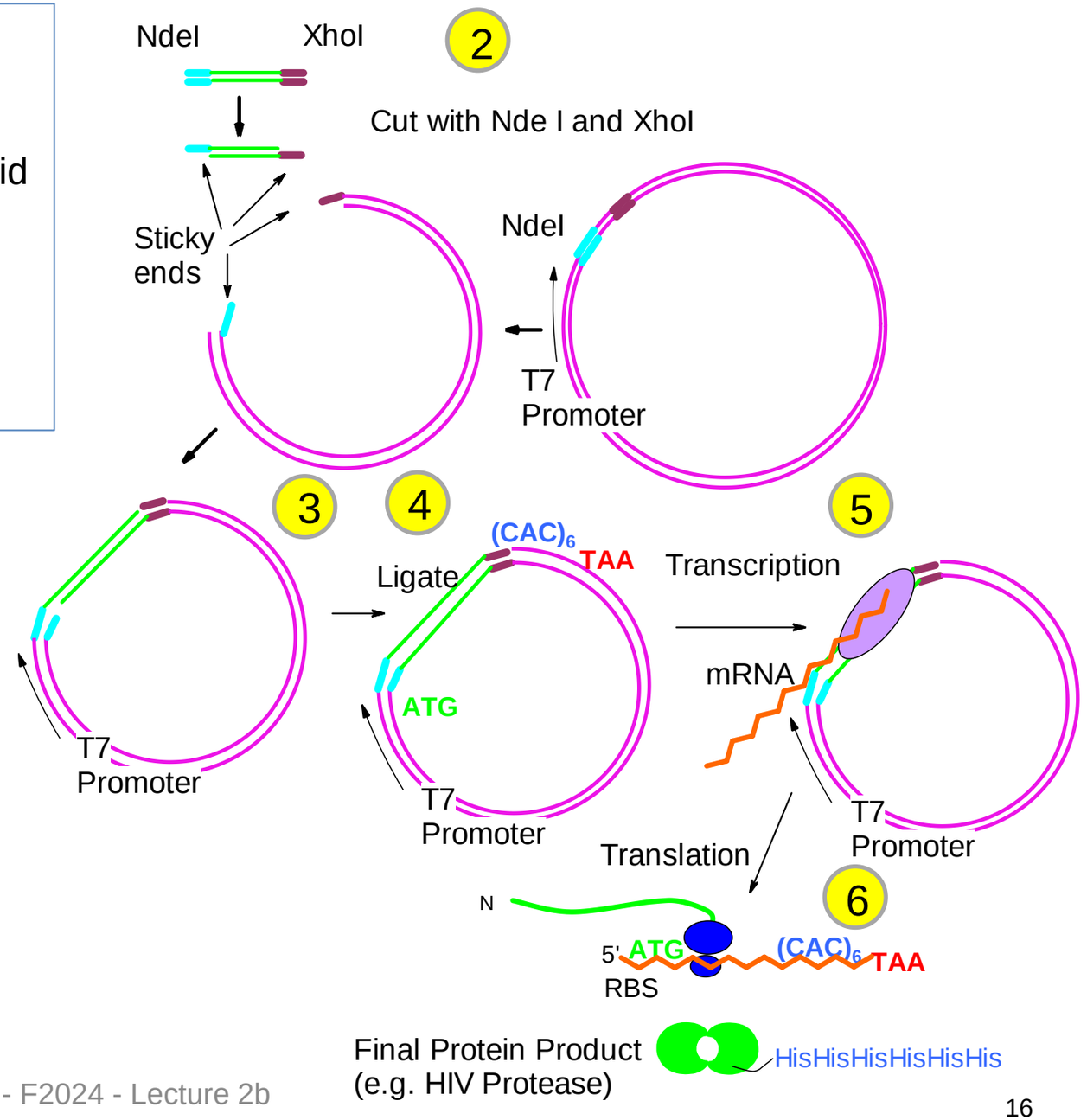
Insertion of DNA into Expression Plasmid

STEPS:

1. Obtain DNA segment that codes for desired protein
2. Cut DNA segment & plasmid with restriction enzymes
3. Ligate (join DNA together)
4. Transform
5. Transcription
6. Translation







Information Transfer In Biology

(DNA Sequence) .ATATGCCCATGTGGTAA..

RNA Polymerase (Transcription)

(mRNA Sequence) .AUAUGCCCAUGUGGUAA..

Translation Ribosome

.U–**AUG**–CCC–AUG–UGG–**UAA**

Punctuation into 3-base codons – Selecting Reading Frame

Met-Pro-Met-Trp (Protein Sequence)

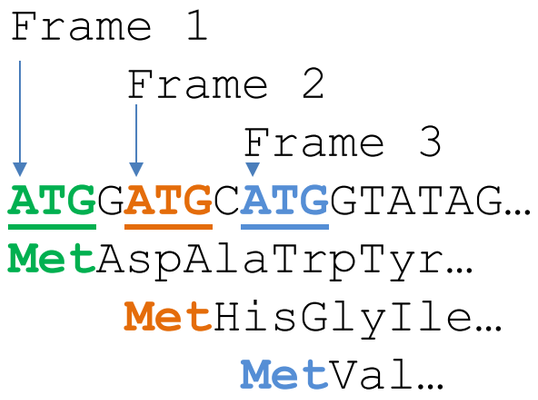
Codon = 3 bases that code for an amino acid.

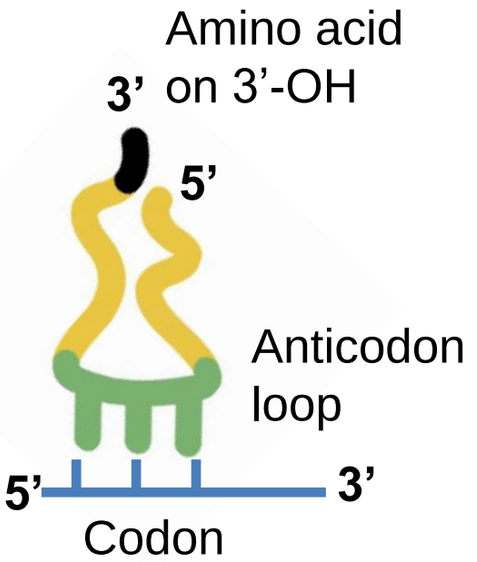
DNA/RNA sequence to Amino Acid Sequence – Codon Table

5' Base	Middle Base				3'
	T	C	A	G	
T	Phe	Ser	Tyr	Cys	T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Term	Term	A
	Leu	Ser	Term	Trp	G
C	Leu	Pro	His	Arg	T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

The codon TGG = ____?

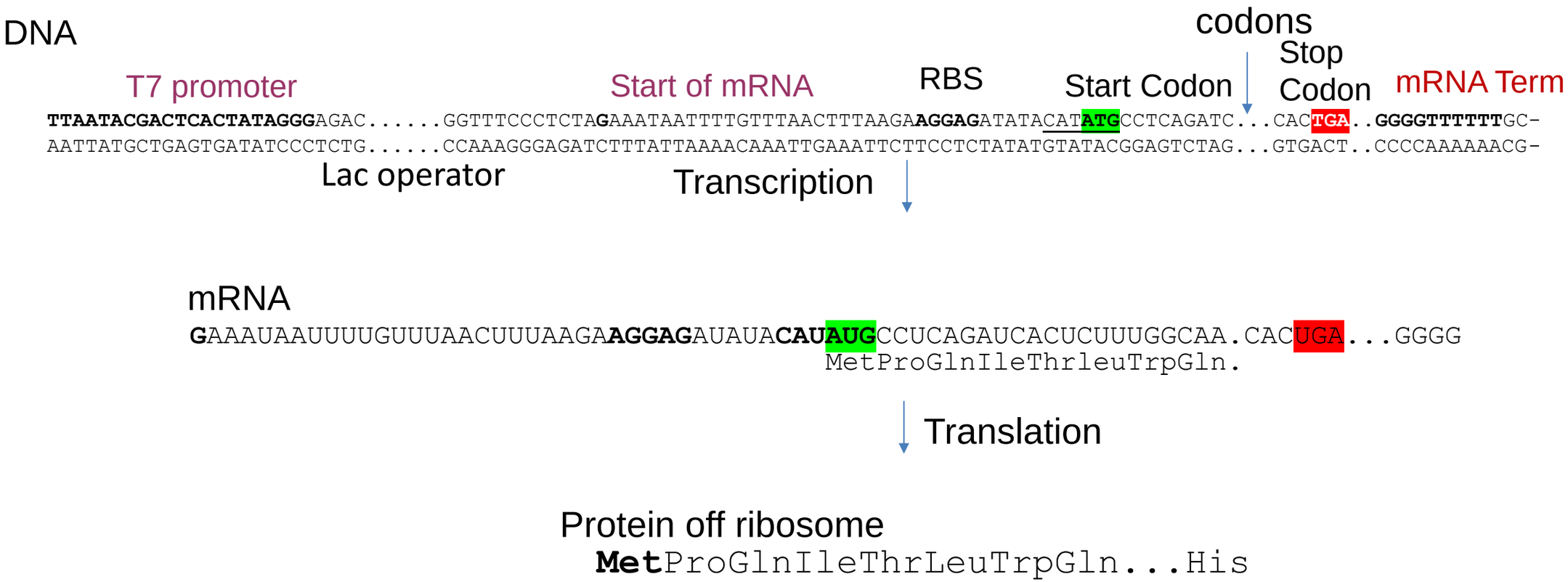
Start codon selection sets the correct reading frame





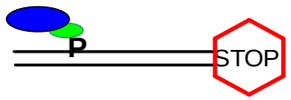
Expectations:
- Concept of reading frame
- Converting codons to AA

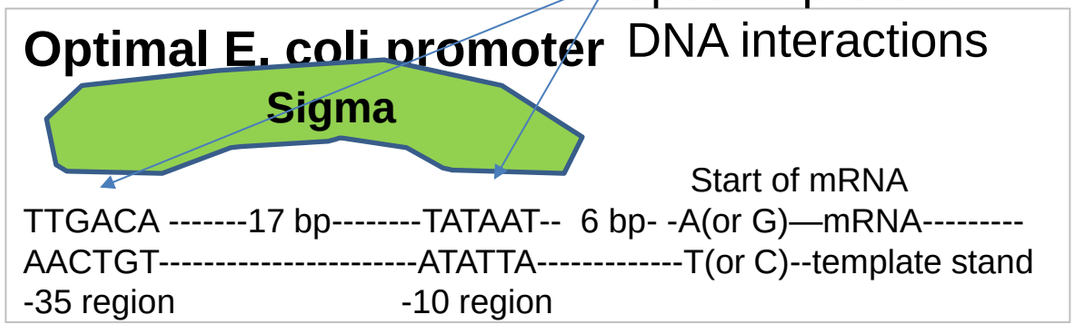
mRNA Synthesis (Transcription) – Important Signals



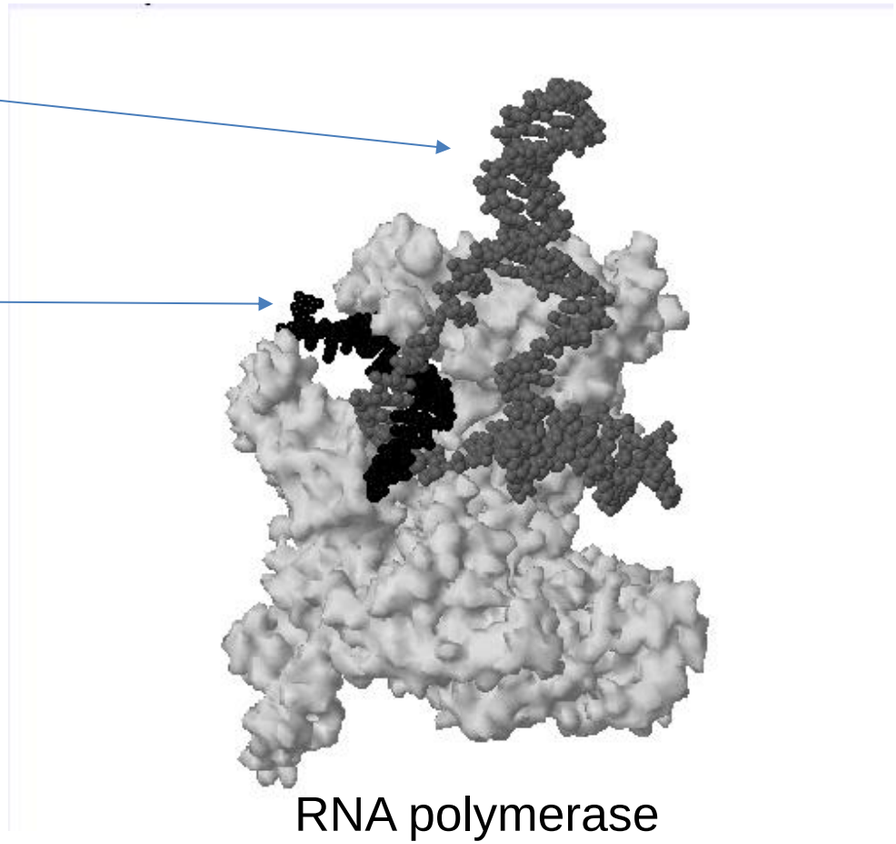
E. Coli RNA Polymerase:

- Holoenzyme: $\sigma + \alpha_2\beta\beta'$
- Binds to promoter (P) sequence in a base specific manner **via σ subunit**.
- Core polymerase = $\alpha_2\beta\beta'$ – sufficient for elongation (sigma factor falls off core enzyme)





-35 = 35 bases from start site



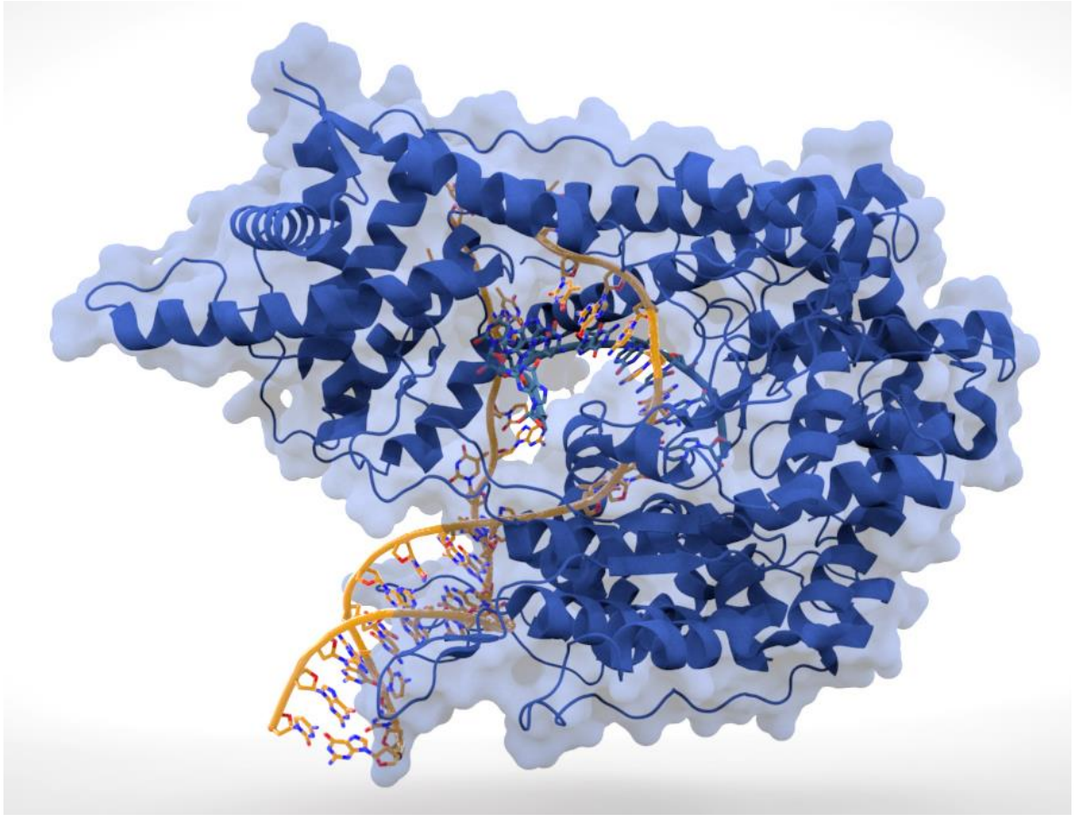
T7 RNA Polymerase

- Isolated from a bacterial virus called T7
- A single chain polymerase
- Recognizes a different promoter sequence – will only transcribe genes with a T7 promoter – selective synthesis of one gene (the one on the plasmid)

T7 is a more robust RNA polymerase than E. coli, seldom stalls during transcription – gives more mRNA and therefore more protein. It is typically used when expressing recombinant proteins from plasmids.

T7 Promoter: TAATACGACTCACTATAGGGAGA

E. Coli Prom: TTGACA -------17 bp----------TATAAT-



Protein Synthesis

(DNA Sequence) .ATATGCCCATGTGGTAA..

RNA Polymerase (Transcription)

(mRNA Sequence) .AUAUGCCCAUGUGGUAA..

.U–**AUG**–CCC–AUG–UGG–**UAA**

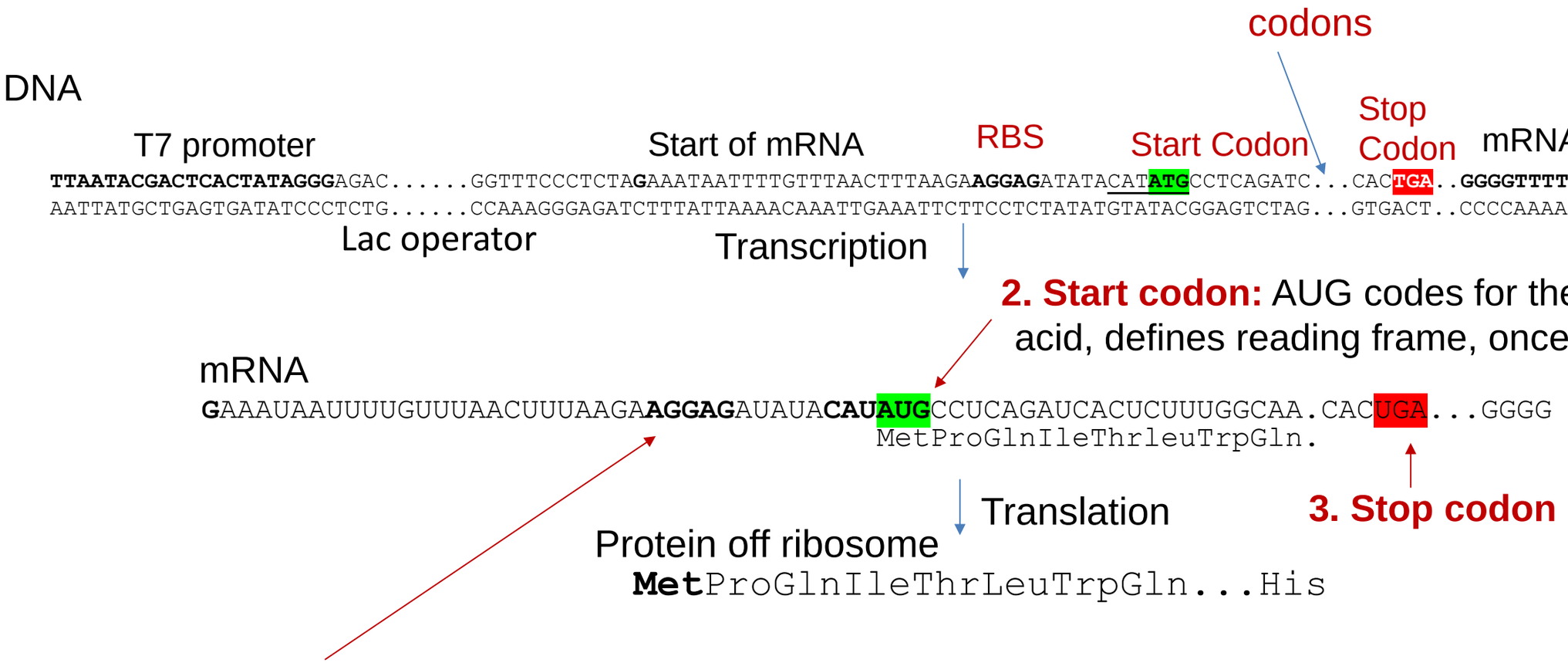
Translation
Ribosome

Punctuation into 3-base codons – Selecting Reading Frame

Met-Pro-Met-Trp (Protein Sequence)

Codon = 3 bases that code for an amino acid.

Protein Synthesis – Important Signals



1. Ribosome Binding Site (RBS): (Shine-Dalgarno [SD] sequence)[Prok only]

- *Positions mRNA on the ribosome so that the correct start codon is used.*
- The optimal spacing between the SD sequence and the AUG is 6-9 bases.

Regulation of Transcription by Lac Repressor

The constitutive expression of high levels of almost any protein is toxic to the bacteria:

- Protein may be toxic (e.g protease)
- Production completes for valuable cellular resources.

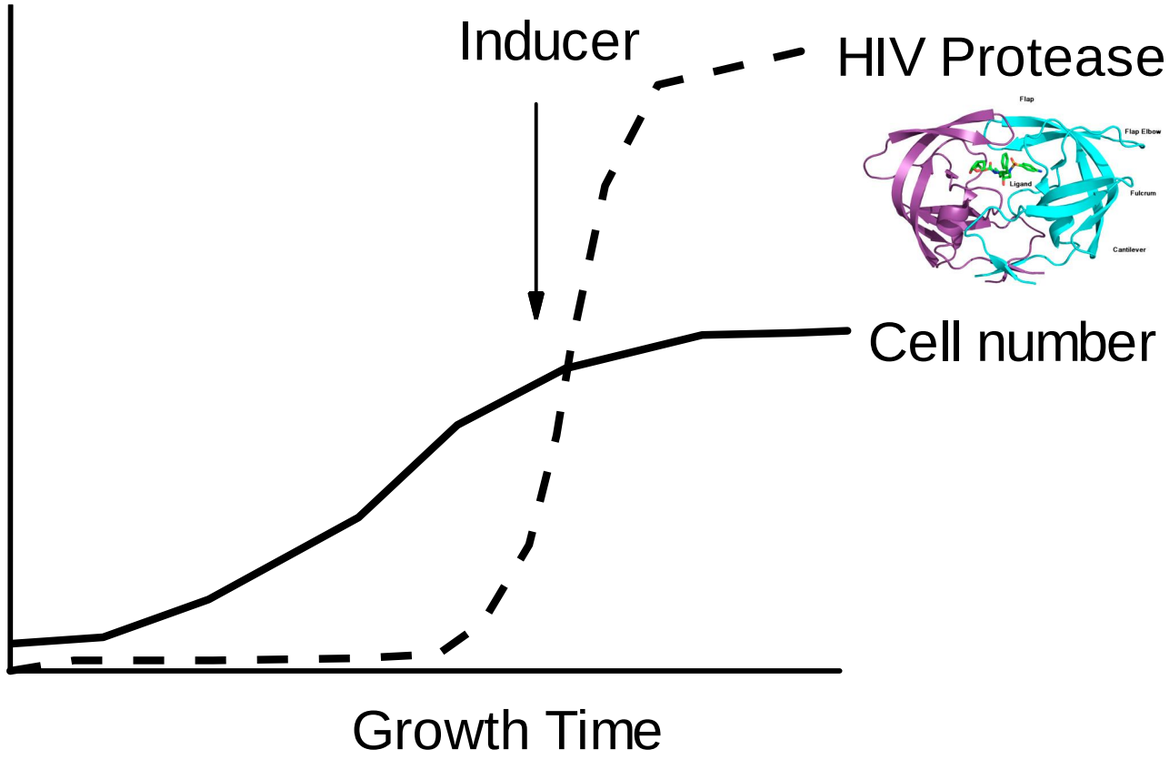
Cell death can occur due to metabolic stress.

An ON/OFF switch is essential.

Expression of the recombinant protein occurs when desired, i.e. when sufficient cell growth has been attained.

The most common ON/Off switch is the **lac repressor-operator** system.

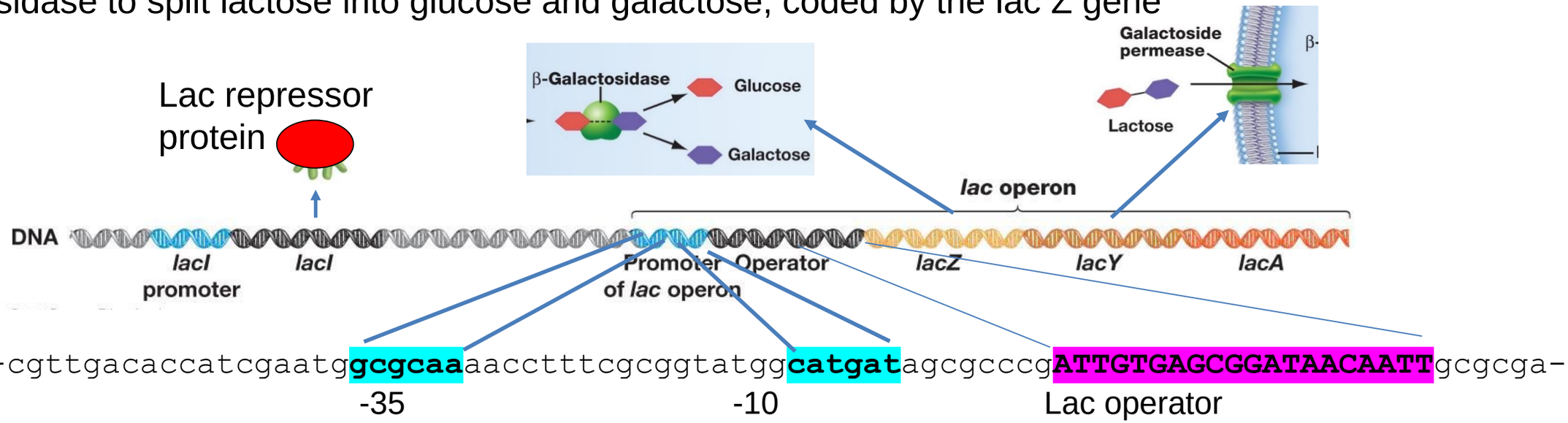
The addition of an **inducer** (IPTG) will start the production of the protein on the plasmid



Transcriptional Regulation: The *lac* Operon – A Natural ON/OFF Switch

Lactose metabolism in bacteria requires 2 proteins – produced from the lac operon (operon is a cluster of protein coding sequences that are produced from a single mRNA, generated from a single promoter)

- Galactoside permease to allow the lactose to enter the cell, coded by the lac Y gene
- Galactosidase to split lactose into glucose and galactose, coded by the lac Z gene

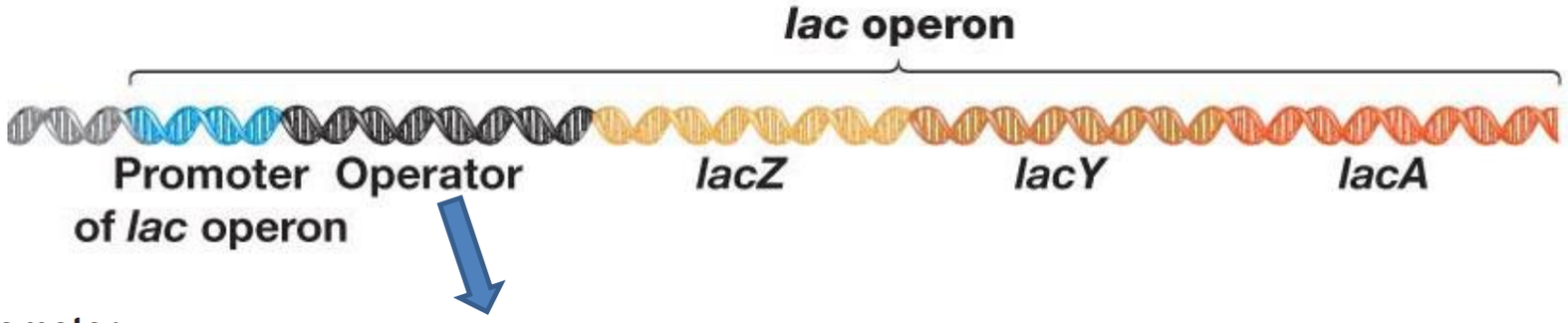


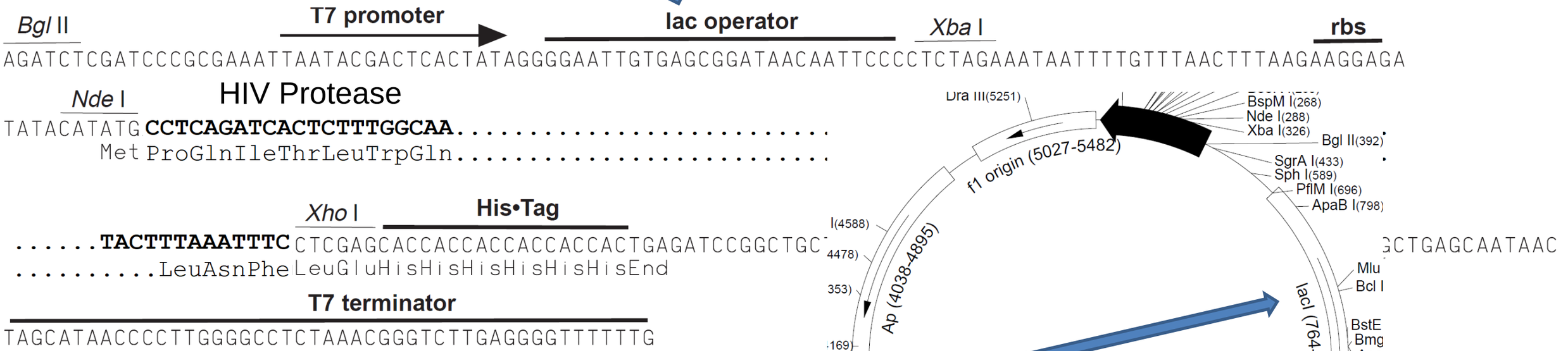
- The production of these proteins is regulated by the **lac repressor**.
- The repressor is a protein that binds to a specific DNA sequence (**lac operator**) and prevents transcription of lacZ and lac Y by blocking E. coli RNA polymerase from moving off the promoter.
- The lac repressor is produced from a separate **lac I gene (not part of the operon)**, which has its own promoter.
- The lac repressor is made all of the time.

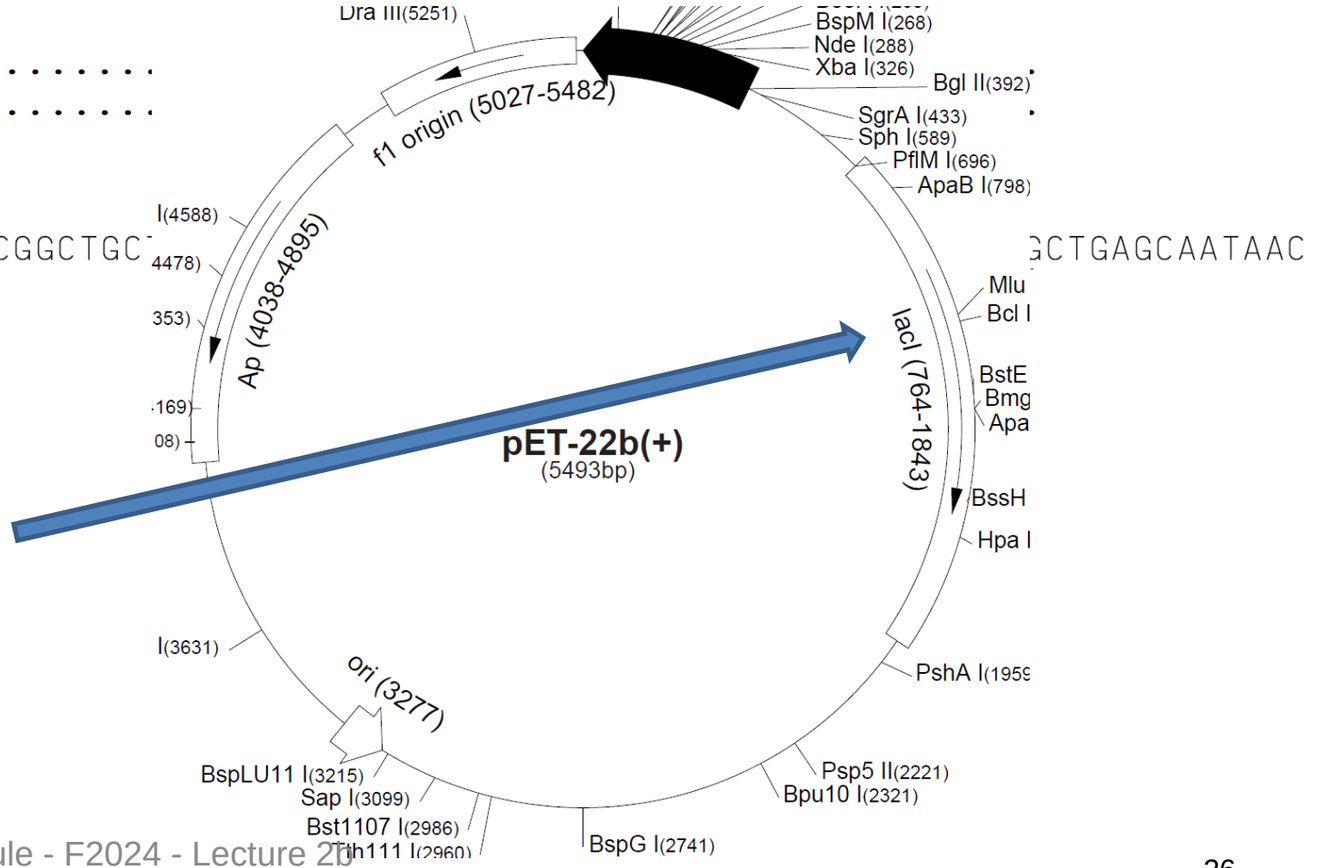
Why? The costly synthesis of enzymes that are required for lactose metabolism are only made when lactose is present in the growth media.

How do we use this control machinery?

Place the lac operator sequence between the T7 promotor and the gene we want to control.

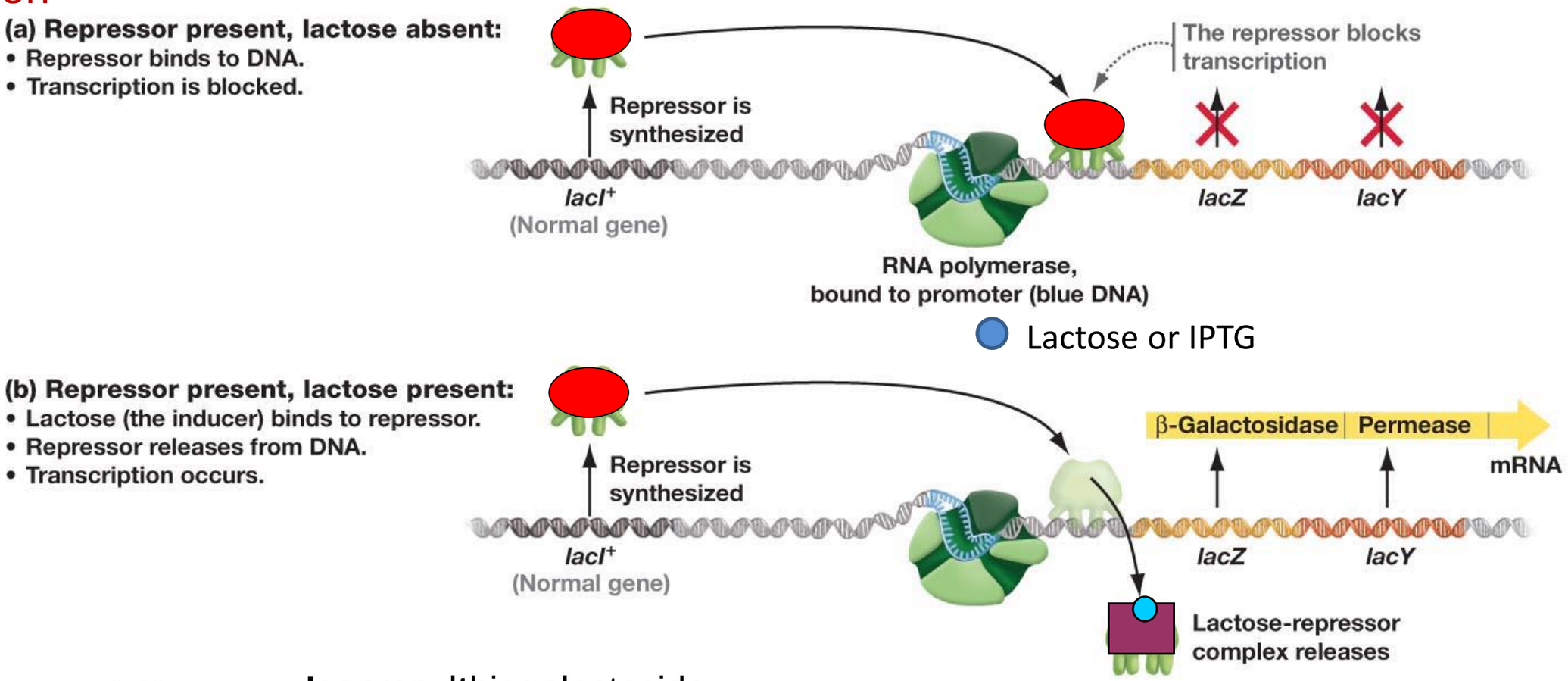




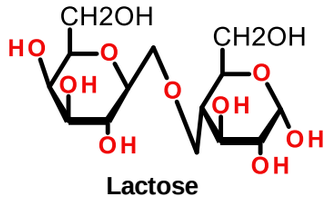


Since there are multiple copies of the plasmid, and not enough lac repressor to go around, the lac I gene is also placed on the plasmid.

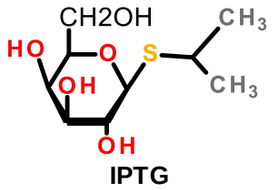
The *lac* Operon



Lactose is the Inducer in nature



Lactose



IPTG

IPTG

Isopropyl**thio**-**g**alactoside (IPTG) is used in the lab.

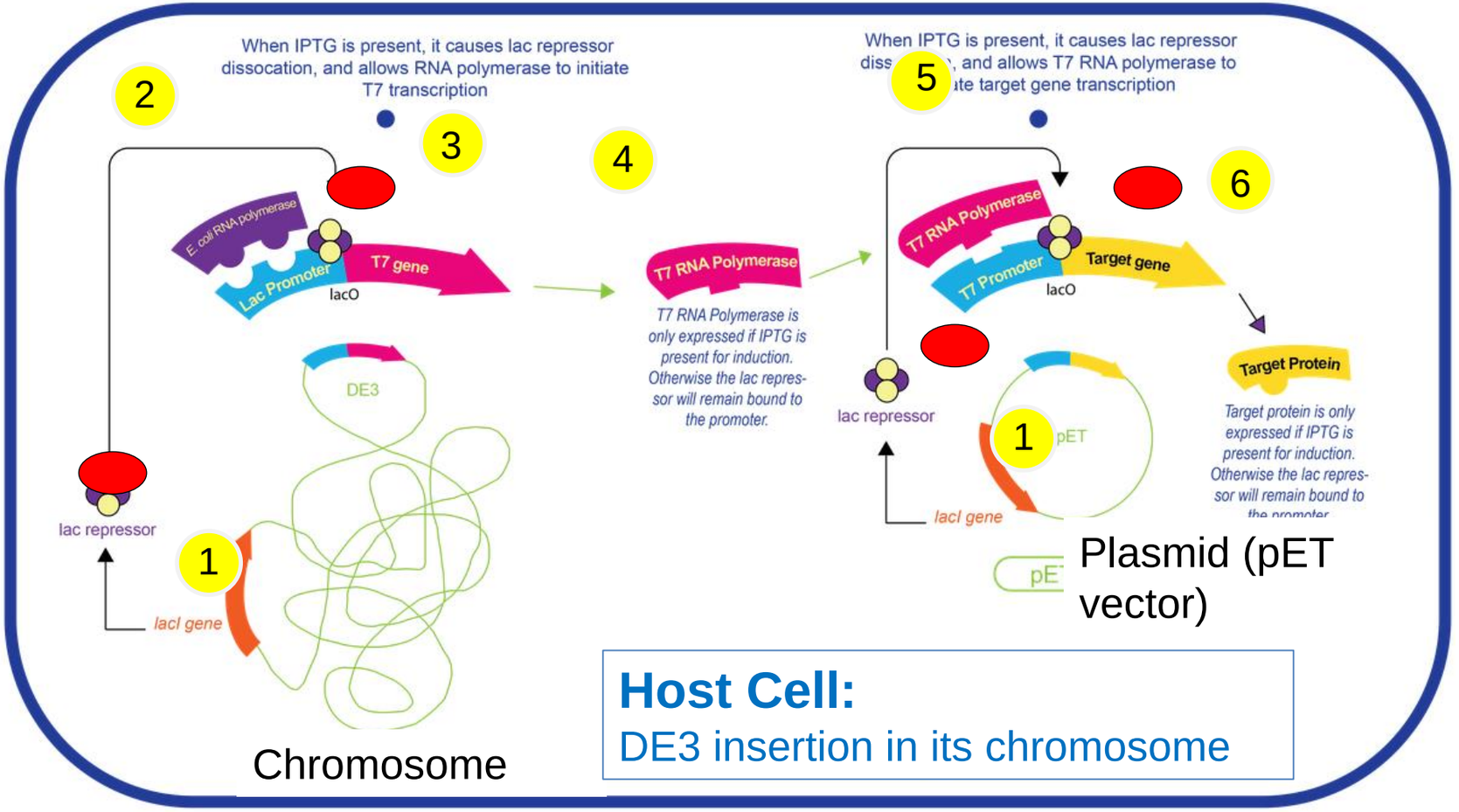
It is a non-hydrolysable inducer that will not be degraded by the cell.

IPTG (and lactose) induce an allosteric change in the lac repressor protein.
R (high DNA affinity) → T (low DNA affinity)

Expression of Recombinant Proteins utilizing the T7 Promotor/Lac Operon

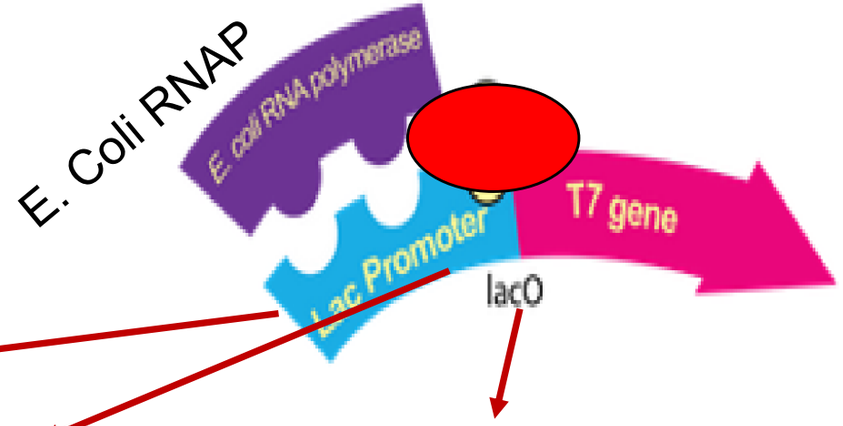
- **T7 RNA polymerase gene** – on the **chromosome** and under control of lac promoter and repressor.
- **T7 promoter** – on **plasmid**, upstream from the gene to be expressed. Usually regulated by lac (shown here).

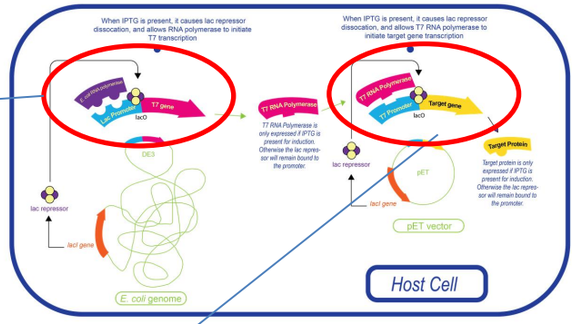
1. The lac repressor is always produced from the lacI gene (chromosome) and from the lacI gene on the plasmid.
2. The lac repressor is initially bound to operator sites, on the chromosome (and the plasmid if regulated)
3. Addition of IPTG causes lac repressor to leave lac operator, upstream from the T7 gene on the chromosome.
4. E. coli RNA polymerase now makes T7 RNA polymerase, using a standard E. coli promoter.
5. T7 RNA polymerase binds to the T7 promoter on the plasmid.



6. T7 RNA polymerase transcribes gene on plasmid, using the T7 promoter. HIV protease mRNA is produced, which is then used by the ribosome to make the mutant HIV protease.

DE3 Sequence Element – Chromosomal: Synthesis of T7 RNA polymerase under lac control

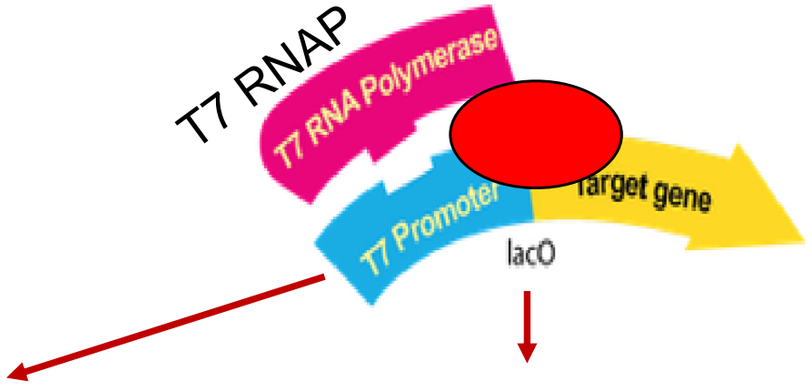




AAA**TTGACA**xxxxxxxxxxxxxxxx**TATAAT**GG**GGAATTGTGAGCGGATAACAATTCC**….**AAGGAG**ATATACAT**ATG**..T7 RNA Poly…

|-35 -E.coli Lac Promoter -10 | |-------Lac operator----------| RBS

T7 – Plasmid: Synthesis of the target protein by T7 RNA polymerase. Usually under lac control

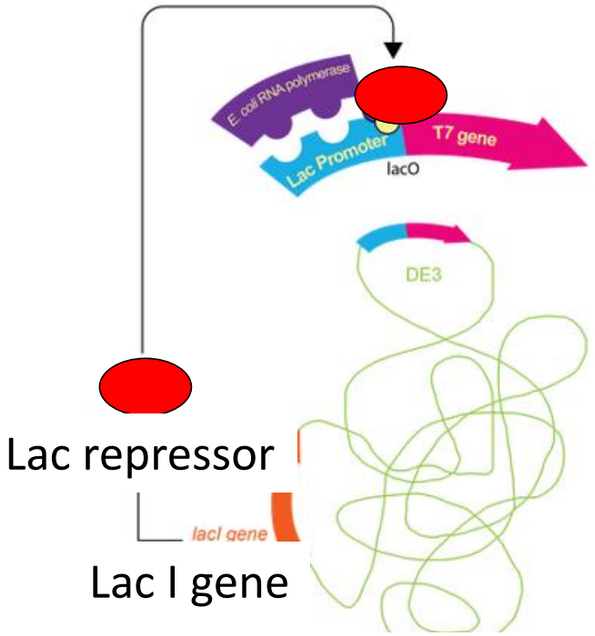


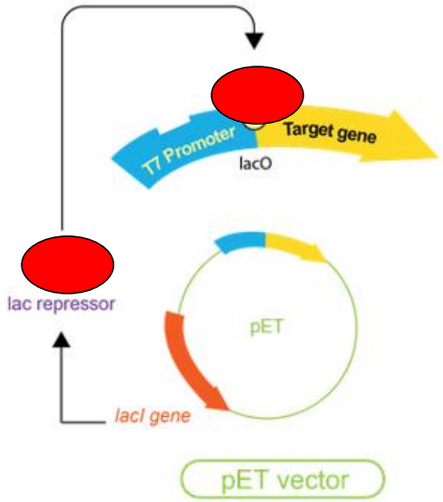
AAA**TTAATACGACTCACTATA**GG**GGAATTGTGAGCGGATAACAATTCC**….**AAGGAG**ATATACAT**ATG**..HIV Prot…

|---T7 promoter--------| |-------Lac operator----------| RBS

No Inducer

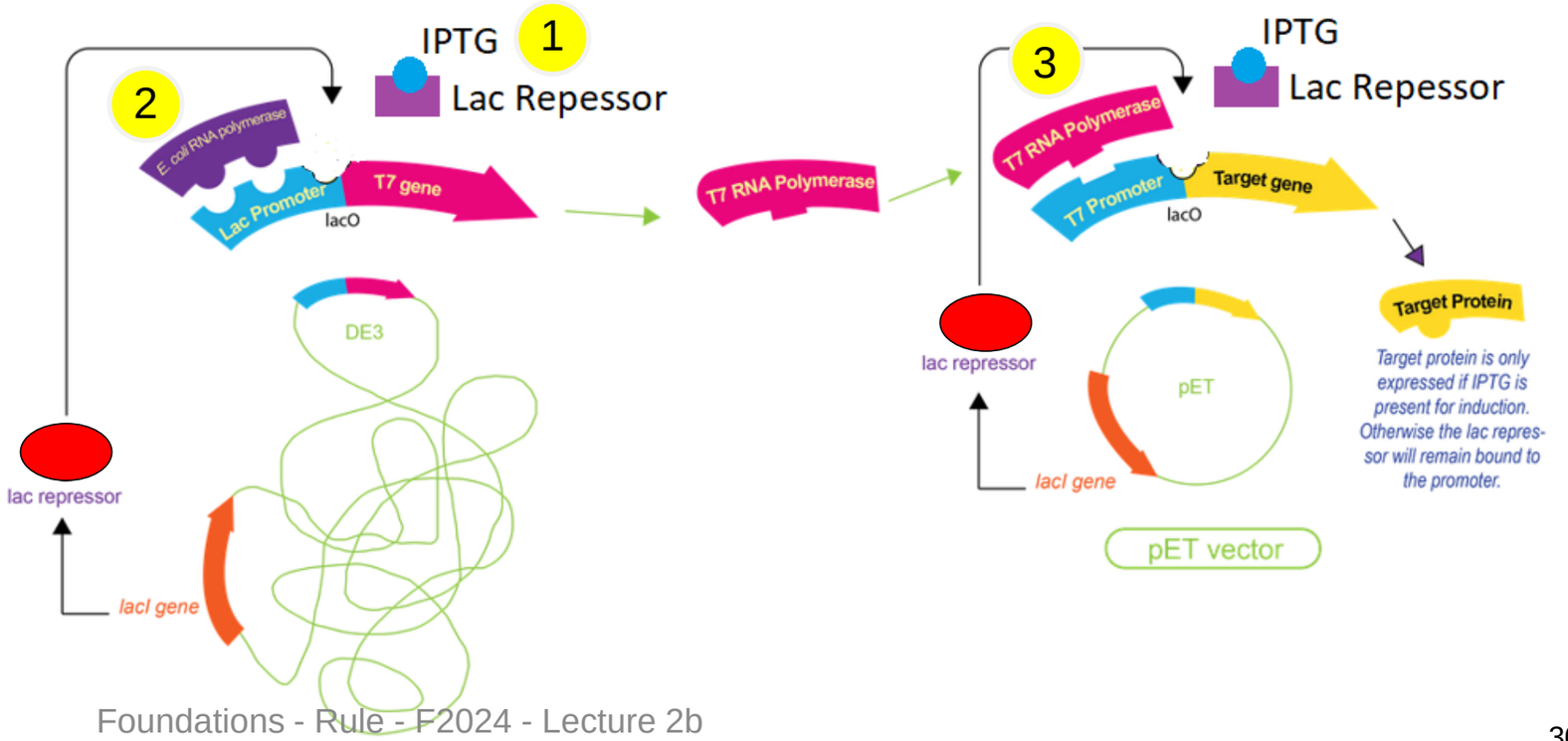
- T7 RNA Pol is not made – its transcription is blocked by lac repressor
- No expression of gene on the plasmid since its promotor is a T7 promoter ***and*** there is no polymerase.





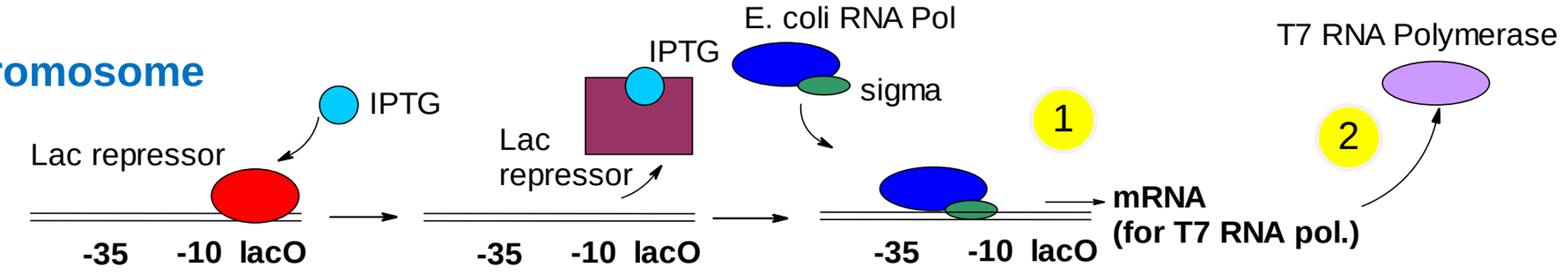
With Inducer (IPTG or Lactose)

1. Lac repressor binds IPTG, dissociates from its operator, on *both* the chromosome & the plasmid.
2. E. Coli RNA polymerase makes mRNA for T7 RNA polymerase, producing T7 RNAP
3. T7 RNAP will generate a mRNA of the gene on the plasmid, producing the target protein.



Regulation of T7 Transcription by Lac Repressor

Events on Chromosome



Addition of IPTG causes:

1. Production of mRNA for T7 RNA polymerase *from the chromosome.*
2. Translation of mRNA to produce T7 RNA polymerase.
3. T7 RNA polymerase produces mRNA *from gene on the plasmid.*
4. mRNA is translated to final target protein.

Events on Plasmid