Review of Molecular Biology
What is life?

Energy

Self-replicating chemical system
Can we decipher the biological hardware and software?

(Andrianantoandro, Basu, Karig, Weiss (2006))
Some Overarching Organizational Rules:

• Biological parts self-assemble

• Biological parts are actively transported

• Biological parts interact specifically

These interactions lead to complex spatial and temporal patterns that produce the variety of life

complex behaviour arising within complex interacting systems = the language of physics
The nature and size of things in biology
Two Generic Cell Types:

- **Bacteria**
  - Prokaryote
    - DNA organized in a single chromosome.
    - No nucleus. No mitosis.

- **Eukaryotes**
  - DNA organized in multiple chromosomes inside a nucleus.
  - Mitotic division.

**Higher Organisms**

**Us**
Prokaryotic Cells:

Thick Walled
- Staphylococcus
- Streptococcus
- Anthrax

Thin Walled
- (The plague, salmonella, meningitis, E. coli, cholera)

Bad pathogens

Penicillin

Thick Walled
- (Staphylococcus, Streptococcus, anthrax)

DNA
Cytoplasm
Plasmid
Ribosome
Pilus
Plasma membrane
Cell wall
Capsule
Bacterial flagellum

Gramp +
Peptidoglycan
Membrane

Gramp -
Peptidoglycan
Membrane
Periplasm
Lipopolysaccharide & protein
Eukaryotic Cells:

The mitochondria make energy – very important
E. Coli – our biological ruler
Huge variety in cells:
Setting some length scales

Figure 2.15 (part 1 of 2) Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Figure 2.15 (part 2 of 2) Physical Biology of the Cell, 2ed. (© Garland Science 2013)
The stuff inside: small molecules

- **water** – we’re 70% H2O

- **ions**: H+, Na+, Ca2+, K+, etc. used to drive transport, pumps

- **sugars** – glucose, ribose, sucrose – substrates for many metabolic reactions

- **nucleotides** – form DNA & RNA
  - ATP (adenosine triphosphate – the big energy source of cells)

- **amino acids** – form proteins

- **fatty acids** – chain like molecules that are the basis of lipids that make the membrane
Lipids & Membranes: The packaging

- Lipids are molecules which form membrane bi-layers
- consist of two fatty acids joined by a ‘head’ molecule of glycerol
- head group likes water, tail does not like water
- Lipids spontaneously self-assemble into membranes
Four different nucleotides: A, C, T, G

Watson-Crick Base Pairing

A = T
G = C
Figure 1.3b Physical Biology of the Cell, 2ed. (© Garland Science 2013)
The Structure of DNA in Cells:

- How do you pack 1 m of DNA into a nucleus that is 2 μm in size?

2 nm DNA gets wrapped around histone complexes forming what are called nucleosomes (DNA negatively charged & histones are positively charged)

These then wrap up and pack further and condense
RNA:

Alphabet = G, C, A, U (instead of T)

- carry the genetic information around as mRNA
- can carry out chemical functions
Proteins: Amino Acids

- Proteins are polymers built from 20 amino acids
Proteins: Structure

Secondary Structure
Helices & Strands

Tertiary Structure
Densely packed hydrophobic core
### Table 1.1: Rules of thumb for biological estimates.

<table>
<thead>
<tr>
<th>Quantity of interest</th>
<th>Symbol</th>
<th>Rule of thumb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell volume</td>
<td>$V_{E.coli}$</td>
<td>$\approx$ 1 $\mu$m³</td>
</tr>
<tr>
<td>Cell mass</td>
<td>$m_{E.coli}$</td>
<td>$\approx$ 1 pg</td>
</tr>
<tr>
<td>Cell cycle time</td>
<td>$t_{E.coli}$</td>
<td>$\approx$ 3000 s</td>
</tr>
<tr>
<td>Cell surface area</td>
<td>$A_{E.coli}$</td>
<td>$\approx$ 6 $\mu$m²</td>
</tr>
<tr>
<td>Macromolecule concentration in cytoplasm</td>
<td>$c_{\text{macromol}}^{E.coli}$</td>
<td>$\approx$ 300 mg/mL</td>
</tr>
<tr>
<td>Genome length</td>
<td>$N_{E.coli}^{bp}$</td>
<td>$\approx$ 5 $\times$ 10⁶ bp</td>
</tr>
<tr>
<td>Swimming speed</td>
<td>$v_{E.coli}$</td>
<td>$\approx$ 20 $\mu$m/s</td>
</tr>
<tr>
<td><strong>Yeast</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume of cell</td>
<td>$V_{\text{yeast}}$</td>
<td>$\approx$ 60 $\mu$m³</td>
</tr>
<tr>
<td>Mass of cell</td>
<td>$m_{\text{yeast}}$</td>
<td>$\approx$ 60 pg</td>
</tr>
<tr>
<td>Diameter of cell</td>
<td>$d_{\text{yeast}}$</td>
<td>$\approx$ 5 $\mu$m</td>
</tr>
<tr>
<td>Cell cycle time</td>
<td>$t_{\text{yeast}}$</td>
<td>$\approx$ 200 min</td>
</tr>
<tr>
<td>Genome length</td>
<td>$N_{\text{yeast}}^{bp}$</td>
<td>$\approx$ 10⁷ bp</td>
</tr>
<tr>
<td><strong>Organelles</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diameter of nucleus</td>
<td>$d_{\text{nucleus}}$</td>
<td>$\approx$ 5 $\mu$m</td>
</tr>
<tr>
<td>Length of mitochondrion</td>
<td>$l_{\text{mito}}$</td>
<td>$\approx$ 2 $\mu$m</td>
</tr>
<tr>
<td>Diameter of transport vesicles</td>
<td>$d_{\text{vesicle}}$</td>
<td>$\approx$ 50 nm</td>
</tr>
<tr>
<td><strong>Water</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume of molecule</td>
<td>$V_{\text{H}_2\text{O}}$</td>
<td>$\approx$ 10⁻² nm³</td>
</tr>
<tr>
<td>Density of water</td>
<td>$\rho$</td>
<td>1 g/cm³</td>
</tr>
<tr>
<td>Viscosity of water</td>
<td>$\eta$</td>
<td>$\approx$ 1 centipoise (10⁻² g/(cm·s))</td>
</tr>
<tr>
<td>Hydrophobic embedding energy</td>
<td>$E_{\text{hydr}}$</td>
<td>2500 cal/(mol nm²)</td>
</tr>
</tbody>
</table>

Table 1.1 (part 1 of 2) Physical Biology of the Cell, 2ed. (© Garland Science 2013)
### Table 1.1: Rules of thumb for biological estimates.

<table>
<thead>
<tr>
<th>Quantity of interest</th>
<th>Symbol</th>
<th>Rule of thumb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DNA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length per base pair</td>
<td>$l_{bp}$</td>
<td>$\approx 1/3 \text{ nm}$</td>
</tr>
<tr>
<td>Volume per base pair</td>
<td>$V_{bp}$</td>
<td>$\approx 1 \text{ nm}^3$</td>
</tr>
<tr>
<td>Charge density</td>
<td>$\lambda_{\text{DNA}}$</td>
<td>$2\ \text{e}/0.34 \text{ nm}$</td>
</tr>
<tr>
<td>Persistence length</td>
<td>$\xi_p$</td>
<td>$50 \text{ nm}$</td>
</tr>
<tr>
<td><strong>Amino acids and proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radius of “average” protein</td>
<td>$r_{\text{protein}}$</td>
<td>$\approx 2 \text{ nm}$</td>
</tr>
<tr>
<td>Volume of “average” protein</td>
<td>$V_{\text{protein}}$</td>
<td>$\approx 25 \text{ nm}^3$</td>
</tr>
<tr>
<td>Mass of “average” amino acid</td>
<td>$M_{\text{aa}}$</td>
<td>$\approx 100 \text{ Da}$</td>
</tr>
<tr>
<td>Mass of “average” protein</td>
<td>$M_{\text{protein}}$</td>
<td>$\approx 30,000 \text{ Da}$</td>
</tr>
<tr>
<td>Protein concentration in cytoplasm</td>
<td>$c_{\text{protein}}$</td>
<td>$\approx 150 \text{ mg/mL}$</td>
</tr>
<tr>
<td>Characteristic force of protein motor</td>
<td>$F_{\text{motor}}$</td>
<td>$\approx 5 \text{ pN}$</td>
</tr>
<tr>
<td>Characteristic speed of protein motor</td>
<td>$v_{\text{motor}}$</td>
<td>$\approx 200 \text{ nm/s}$</td>
</tr>
<tr>
<td>Diffusion constant of “average” protein in cytoplasm</td>
<td>$D_{\text{protein}}$</td>
<td>$\approx 10 \mu\text{m}^2/\text{s}$</td>
</tr>
<tr>
<td><strong>Lipid bilayers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thickness of lipid bilayer</td>
<td>$d$</td>
<td>$\approx 5 \text{ nm}$</td>
</tr>
<tr>
<td>Area per molecule</td>
<td>$A_{\text{lipid}}$</td>
<td>$\approx \frac{1}{2} \text{ nm}^2$</td>
</tr>
<tr>
<td>Mass of lipid molecule</td>
<td>$m_{\text{lipid}}$</td>
<td>$\approx 800 \text{ Da}$</td>
</tr>
</tbody>
</table>
Information Storage, Replication & Processing:

What does the cell use all these molecules/structure for?

1) Storing Chemical Information
2) Passing On/Replicating this information
3) Processing and calculating with this information

We’ll look at each of these steps now
Cellular “states”:

- Cellular state is determined by which genes=proteins are “ON”

- Different cell types = different genetic programs that are being run

- Same DNA

- BUT

- Different genes are “ON”
Information Storage:

• The information for running the cellular program is stored in the sequence of DNA.

• DNA is like the hard drive of a computer.

• It stores information, contains programs that get executed at specific times.

• Each protein has a specific sequence in the DNA called a ‘gene’, and this DNA is called ‘coding’ DNA, as it codes for messages that make proteins.

• The majority of DNA sequence does not code for ‘genes’ and is called non-coding DNA.

• It is the non-coding DNA which contains the ‘programs’ which determine what genes will get made at a given time or place.

• Different cells are running different programs and thus making different sets of genes.
The Sizes of the Hard-drives

Bacteria (usually) have single circular chromosome – have many small plasmids

Eukaryotes have multiple linear chromosomes

- **haploid** = single copy of every chromosome
- **diploid** = two copies of every chromosome
- **polytoid** = multiple copies of every chromosome

<table>
<thead>
<tr>
<th>organism</th>
<th>#</th>
<th>T</th>
<th>size</th>
<th>genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>virus</td>
<td>1</td>
<td>H</td>
<td>5kb</td>
<td>10</td>
</tr>
<tr>
<td>E. coli</td>
<td>1</td>
<td>H</td>
<td>5Mb</td>
<td>4,377</td>
</tr>
<tr>
<td>S. Cervisiae</td>
<td>16</td>
<td>H</td>
<td>12Mb</td>
<td>5,570</td>
</tr>
<tr>
<td>C. elegans</td>
<td>16</td>
<td>D</td>
<td>100Mb</td>
<td>19,000</td>
</tr>
<tr>
<td>A. thaliana</td>
<td>5</td>
<td>D</td>
<td>115Mb</td>
<td>25,498</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>4</td>
<td>D</td>
<td>122Mb</td>
<td>14,000</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>23</td>
<td>D</td>
<td>3Gb</td>
<td>25,000</td>
</tr>
</tbody>
</table>
Figure 3.12 Physical Biology of the Cell, 2ed. (© Garland Science 2013)
Information Replication: Cell Division:

Asexual and normal cell division

Sex cells = gametes = haploids
DNA replication:

Bacteria have single origin of replication

Replication Times:

bacteria ~ 40 mins (1 origin)
1000 nucleotides/s

humans ~ few hours (many origins)
100 nucleotides/s
Information Processing: Central Dogma:

Replication → DNA → Transcription → RNA → Translation → Protein
DNA to Proteins? Genetic Code

- 64 codons, with highly non-uniform mapping

<table>
<thead>
<tr>
<th>DNA</th>
<th>mRNA</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATGCAGCCGATATAA</td>
<td>AUG CAG CCG AUA UAA</td>
<td>MEPI(STOP)</td>
</tr>
</tbody>
</table>

![Table of genetic code]
Gene Transcription:

- Bacteria = ON
- Eukaryotes = OFF, requires lots of other help

Promoter

RNA Polymerase

mRNA

Pushes polymerase off of DNA

● Bacteria = ON
● Eukaryotes = OFF, requires lots of other help
Messages are delivered differently:

**Prokaryotes (Simple):**

- Genes in bacteria are organized in operons – one contiguous message RNA is generated

**Eukaryotes (Complex):**

- Genes contain coding (exons) and non-coding (introns) sequence
- Introns are spliced out and exons are assembled to make final mRNA
- final mRNA is exported out of nucleus and made into protein
Translation:

- Ribosome is big machine that translates mRNA into protein
- In bacteria, Ribosome binds to specific site on mRNA
- In eukaryotes, Ribosome binds to 5' cap
- Translation ends when 'STOP' codon is reached
Figure 3.13 Physical Biology of the Cell, 2ed. (© Garland Science 2013)
How cells get energy

- Oxidation of food $\rightarrow$ CO2 and H2O and energy
- **Metabolism** = processes of converting/synthesizing molecules in cells
  - **Catabolism** = breaking down of molecules
  - **Anabolism** = synthesizing of molecules
- Two major energy reserves: sugars, fatty acids
Most chemical reactions in cells are mediated by proteins called enzymes.

Enzymes act as catalysts, by reducing the activation barrier between two states.
Energy Storage in Cells

- Some chemical reactions in cells are energetically unfavourable (e.g., DNA & protein synthesis).
- Couple energetically bad pathway with favourable pathway using enzyme.
- Favourable pathway is driven by taking energy from energy carriers.
- Most pervasive energy carrier is ATP.

\[
\text{ATP} \rightarrow \text{ADP} + \text{P} + \text{H}_2\text{O} + (13 \text{ kcal/mol} = 25 \text{ kT})
\]

- DNA synthesis burns 2 ATP.

HEAT

Energy from burning of food

store in ATP

burn ATP to drive reactions
Converting Food to Useful Energy:

**Glycolosis**
- anaerobic
- converts glucose → 2 ATP

**Citric Acid Cycle = Krebs Cycle:**
- aerobic
- occurs in the mitochondria
- generates about 30 ATP from a single glucose!!!
- convert about 50% of possible energy (c.f. car ~ 20%)

![Diagram of glucose conversion between liver and muscle](image)
Numbers of things in cell biology
Why knowing numbers in biology matters?

- Given the vast range of scales and the large variety of parts, it's important to have a sense of the numbers of things.

- Why? (1) Being able to estimate numbers of parts etc builds intuition.
  (2) Allows one to assess what aspects of a given biological process will be important quantitatively.
  - In complex systems, the things one ignores are as important as the things that are considered.
  - Models are always incomplete.

Roadmap to Quantitative Biology:

\[ F = ma \]
\[ D^2 \theta = 0 \]

Experiment \rightarrow Quantitative Data \rightarrow Intuition & model building
Figure 2.4 Physical Biology of the Cell, 2ed. (© Garland Science 2013)
Some estimates of parts:

- **Volume**: $V_{cell} \approx 1 \text{ \mu m}^3 = 1 \text{ fL}$
- **Area**: $A_{cell} \approx 6 \text{ \mu m}^2$

**Cells are crowded:**

**Mass**: $Mass = density \times Volume$

$\propto density \text{ H}_2\text{O} \times V_{cell}$

$\propto 1 \text{ g/mL} \times 1 \text{ fL} = 1 \text{ pg}$

**Dry mass**: Experiment $\Rightarrow M_{dry} \approx 0.2 \text{ pg} = 30\%$ Mass

**Protein mass**: avg. protein = 300 amino acids (AA)

$1 \text{ AA} = 100 \text{ Da}$

$\therefore 1 \text{ Da} = \frac{M_{protein}}{1.6 \times 10^{-24} \text{ kg}}$

so average protein mass,

$M_{protein} = (300) \times (100 \text{ Da}) \times (1.6 \times 10^{-24} \text{ g}) \approx 5 \times 10^{-20} \text{ g}$

**Number of proteins in E. coli**: 50% of dry mass is protein from experiment

$N_{protein} = \frac{\text{protein mass}}{\text{mass per protein}} = \frac{0.15 \text{ pg}}{5 \times 10^{-20} \text{ g}} \approx 3 \times 10^6$
Number of Ribosomes:

Experiment: 20% of protein mass is contributed by ribosomes

\[ M_{\text{ribosome}} = 2.5 \text{ MDa} \]

but 1/3 Ribosome is protein & 2/3 is rRNA

\[ N_{\text{ribosome}} = \frac{(0.2)(0.15 \text{ pg})}{830,000 \text{ Da}} = 20,000 \]

How much volume do ribosomes occupy?

\[ V_{\text{ribosome}} = \frac{4}{3} \pi (10 \text{ nm})^3 \]

\[ V_{\text{tot}} = (20,000) \frac{4}{3} \pi (10 \text{ nm})^3 \approx 10^8 \text{ nm}^3 \]

\[ \approx 10\% \ V_{\text{cell}} \]
<table>
<thead>
<tr>
<th>Substance</th>
<th>% of total dry weight</th>
<th>Number of molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macromolecules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>55.0</td>
<td>$2.4 \times 10^6$</td>
</tr>
<tr>
<td>RNA</td>
<td>20.4</td>
<td></td>
</tr>
<tr>
<td>23S RNA</td>
<td>10.6</td>
<td>19,000</td>
</tr>
<tr>
<td>16S RNA</td>
<td>5.5</td>
<td>19,000</td>
</tr>
<tr>
<td>5S RNA</td>
<td>0.4</td>
<td>19,000</td>
</tr>
<tr>
<td>Transfer RNA (4S)</td>
<td>2.9</td>
<td>200,000</td>
</tr>
<tr>
<td>Messenger RNA</td>
<td>0.8</td>
<td>1,400</td>
</tr>
<tr>
<td>Phospholipid</td>
<td>9.1</td>
<td>$22 \times 10^6$</td>
</tr>
<tr>
<td>Lipopolysaccharide (outer membrane)</td>
<td>3.4</td>
<td>$1.2 \times 10^6$</td>
</tr>
<tr>
<td>DNA</td>
<td>3.1</td>
<td>2</td>
</tr>
<tr>
<td>Murein (cell wall)</td>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>Glycogen (sugar storage)</td>
<td>2.5</td>
<td>4,360</td>
</tr>
<tr>
<td><strong>Total macromolecules</strong></td>
<td><strong>96.1</strong></td>
<td></td>
</tr>
<tr>
<td>Small molecules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolites, building blocks, etc.</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>Inorganic ions</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td><strong>Total small molecules</strong></td>
<td><strong>3.9</strong></td>
<td></td>
</tr>
</tbody>
</table>
What is the concentration of 1 molecule in a E. coli?

Concentrations and #’s of molecules

concentration = \frac{\#}{volume}

units: molar = [M] = \frac{\text{mol}}{\text{liter}} = \frac{6.02 \times 10^{23}}{1 \text{ L}}

cellular concentrations range from 1 nM \rightarrow 1 \mu M

What is concentration of 1 molecule in E. coli?

C_i = \frac{1}{1 \text{ fl}} = \frac{1 \text{ molecule}}{1 \times 10^{-15} \text{ L}} \approx \frac{1 \text{ mol}}{6 \times 10^{23} \text{ molecule}} \approx 2 \text{ nM}

so a concentration of 2 \mu M \approx 1000 molecules in the cell

1 nM \sim 1 molecule inside the volume of an E. coli cell
Cells are crowded places

Distance between molecules:

- Assume molecules sit on a square lattice in a square cell of volume $V$

\[ c = \frac{N}{V} = \frac{N}{(N \cdot d^3)} = d^{-3} \]

so spacing = $d = c^{-1/3}$

for $c = 2 \mu M \Rightarrow d \approx 150 \text{ nm}$

for $c = (1 \times 10^6) \text{ M} \Rightarrow d \approx 1 \text{ nm} \leftarrow$ hardly any space & best fit proteins
The timing of things: timescales in biology
Why timescales matter?

Biology is not static, it’s dynamic!

It has dynamics over a range of timescales

Q: How do we treat physical systems that have processes that are operating over many different timescales

A: We choose a timescale of interest and only consider processes that are of the same scale. Faster processes will be considered to be at steady state - we’ll define this later.
Figure 3.1 Physical Biology of the Cell, 2ed. (© Garland Science 2013)
Development of *Drosophila*

1 day

- Egg → Larva → Pupa → Adult fly

Early development of *Drosophila*

- Eggs → Blastoderm → L3 → Pupa → Adult fly

Bacterial cell division

- Minutes

Figure 3.2a Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Figure 3.2b Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Figure 3.2c Physical Biology of the Cell, 2ed. (© Garland Science 2013)
Figure 3.2f Physical Biology of the Cell, 2nd ed. (© Garland Science 2013)
Gating of ion channels

Enzyme catalysis
The lives of molecules: RNA lifetimes

(A) E. coli

(B) S. cerevisiae

Figure 3.14 Physical Biology of the Cell, 2ed. (© Garland Science 2013)
The lives of molecules: Protein lifetimes

![Histogram of protein half-lives](image)

*S. cerevisiae*

Figure 3.15a *Physical Biology of the Cell, 2nd ed.* (© Garland Science 2013)
Timing estimates

How fast is the replication machinery?

- There are $5 \times 10^6$ bp in E. coli
- Thus $rate_{bp} = \frac{5 \times 10^6 \text{ bp}}{3000 \text{ s}} = 2000 \text{ bp/s}$

- In E. coli, 2 DNA polymerases replicate the DNA from 1 origin in opposite directions
- A single DNA polymerase in E. coli has a rate of 1000 bp/s

Aside: In higher organisms, DNA polymerase rate ~ 100bp/s and the genomes are much larger

Q: how does the DNA ever get replicated in time?
Timing estimates:

What is the rate of protein synthesis?

Previously, $\approx 3 \times 10^6$ proteins in an E. coli cell

\[
\text{rate protein} = \frac{3 \times 10^6}{3000} = 1000 \text{ proteins/s}
\]

Q: what are the mechanical properties of these bio machines (DNA polymerase, ribosomes) that allow them to operate at these rates?

Q: given the # of ribosomes in E. coli, how many proteins per second is ribosome making?
In many biological processes, time is relative.

All that matters is that a task gets completed.

So in the cell cycle there are many checkpoints.
Building an oscillator

General Design of Biological Oscillators

- negative feedback + time delay can produce oscillations.

The production and degradation of cyclins, depends upon the abundance of cellular resources. The absolute time of the period, T, will vary depending on the environment.
Biophysical Experimental Techniques:
Figure 2.13c Physical Biology of the Cell, 2ed. (© Garland Science 2013)
Figure 4.12 Physical Biology of the Cell, 2ed. (© Garland Science 2013)
Molecular Biology Experimental Methods:
Chromatography:

- Chromatography is used to separate and purify molecules from a complex mixture.

**Types:**
- **Charge** = ion-exchange chrom.
- **Hydrophobicity** = hydrophobic chrom.
- **Size** = gel-filtration chrom.
- **Binding** = affinity chrom.

**Affinity chromatography** can be used to elute very pure protein samples:
- DNA for DNA binding proteins
- Antibodies for specific protein
- Protein binding partners
- Engineered tags (HIS, GST)

= Co-immune precipitation = CoIP
Gel Electrophoresis

- Used to determine size of biomolecules
- -vely charged molecules move in applied field through gel

- For proteins, SDS (detergent) is used to denature proteins and give them net negative charge = SDS-PAGE

- 2D SDS-PAGE can be used to sort proteins by both size and native charge
DNA manipulation

- Use restriction enzymes to cut specific nucleotide sequences (e.g. HpaI = GTTAAC; EcoRI = GAATTC)
- Some make blunt ends, some make dangling 'sticky' ends

![DNA sequences](image)

- Used to chop up large DNA segments (i.e. Chromosome)
- Used to ligate (join) different DNA fragments -> genetic engineering
Northern, Southern, and Westerns

- Used to detect quantitative amounts of specific DNA (southern), RNA (northern) or proteins (Western) from a complex mixture (i.e. cellular extract).

- Labelled DNA probe 'hybridizes' with matching partner in sample

- In westerns, a labelled antibody is used for probe
DNA Cloning

- Replicate specific DNA fragment in large quantities using bacteria
- Fragment is inserted into circular plasmid DNA and transfected into bacteria
- Large collection of bacteria colonies each with different fragment = library

- Types of libraries = genomic, cDNA
  - cDNA library comes only from mRNA
  - Use cDNA plasmid to express lots of specific protein = expression vector
Dideoxy method: use special nucleotides that stop growth of DNA
--> produces DNA that stops at specific letter positions
Polymerase Chain Reaction (PCR):

- Use 2 DNA primers to pull out and amplify specific region of DNA in a sample.
Our chromosomes contain regions of 'satellite' DNA that vary in length – each individual has specific length = fingerprint

PCR amplify regions and then use gels to measure the lengths -> yields fingerprint
Imaging the action

- Reporter

- Green fluorescent protein (GFP), lacZ, labelled antibodies

![Diagram of gene expression with GFP fusion and enhancer](image)
Imaging II

- **In-situ hybridization** measure the presence of mRNA of specific gene
- Takes a snapshot of gene pattern – use fluorescently labelled RNA probe
- Use confocal microscope to image different 2D layers ☀️☀️☀️ build 3D image
Detecting biomolecular interactions

FRET (Flourescence Resonance Energy Transfer)

Yeast Two Hybrid

BAIT

PREY
Interactions II:

Phage Display:
- Phage = virus that infects bacteria. Has an external coat of protein
- Can fuse foreign proteins to coat proteins of phage
- Used to screen libraries of proteins for specific interaction drug design