Biochemistry Notes: Topics 14-24

Topic 14: Nucleic Acid Structure

Major Functions of DNA and RNA
- mRNA, transcribed from genes
- miRNA, bits of RNA involved in regulation of mRNA
- snRNA, tRNA, DNA

A ribonucleoside
- a sugar
- a carbohydrate with a nitrogenous base attached at the anomeric carbon
- all in beta conformation
- eg. Cytidine

A ribonucleotide
- eg. Cytidine monophosphate= a nucleotide = a nucleoside monophosphate
- the base is cytosine, but the nucleoside is cytidine
- there is a N glycosidic bond
- two rings: one gets number 1, 2, 3.. and the other 1', 2' ,3'..
- When the phosphate it attached we call it a nucleotide
- Mono-1 phosphate, Di-2, Tri-3

A deoxyribonucleotide
- Remove oxygen from the 2' carbon
- Cannot assume that when you say nucleotide whether it is deoxynucleotide or ribonucleotide
- Deoxycytidine triphosphate= a deoxynucleotide= a deoxynucleoside triphosphate

Nitrogenous Bases
- Base: Adenine
  - Nucleoside: Adenosine
  - Abbreviation: A
  - Purine
  - Double ring structure
- Base: Guanine
  - Nucleoside: Guanosine
  - Abbreviation: G
  - Purine
  - Double ring structure
- Base: Cytosine
  - Nucleoside: Cytidine
  - Abbreviation: C
  - Pyrimidines
  - single ring structure
- Base: Thymine
  - Nucleoside: Thymidine
  - Abbreviation: T
  - Pyrimidine
  - single ring structure
- Base: Uracil
  - Nucleoside: Uridine
  - Abbreviation: U
  - Pyrimidine
  - single ring structure
  - only RNA
- all pretty planar structures
- fairly hydrophobic

- pi electrons behave similar to aromatics

Polynucleotides:

- Add together through phosphodiester linkage
- Alago, a short sequence
- Phosphate group joins the 3' carbon to the 5' of the next sugar
- It has polarity to it
- 3' end OH group 3' carbon
- 5' end 5' carbon there (maybe there is a phosphate there)
- Read the sequence from 5' to 3' for naming/indicating bases
- How many sequences can u make that are 100 nucleotides long, 4^{100}

1. Chargaffs Rule
   - If you take any organism and you count how much G and C are in there cells its about the same, and how much A and T is about the same
   - Over a group of a bunch of different organisms
   - It doesn't't matter which tissues you get you always get the same ratio
2. X-ray Diffraction Pattern
   - Tells you stuff about 3D structure
   - Diffraction pattern was generated not by Watson and Crick
   - Rose Franklin,
   - This showed about the period of the helix,
3. Correct tautomeric forms of Bases
   - Ketone or Enol form
   - Watson and crick were able to figure out how base pairing works

Double Helix: "B-DNA"

- The helix was double stranded and those were antiparallel
- 5'-3' up and 3' to 5' down
- Bases were in the middle
- Tend to want to go away from water and go in the middle
- Phosphates that are negatively charged, and this structure allows them to be spaced out
- Focus only on B-DNA, right handed

Forces that Stabilize the double Helix:

- 1.Base stacking
  - Bases are fairly planar, flats parts of the molecule can stack onto of each other, bases themselves are flat and in the middle, bases are perpendicular
  - London dispersion forces, pi electrons attract one another
- 2. Base pairing
  - Bases contribute to base pair

Watson-Crick base pairing

- A with T, 2 hydrogen bonds
- G with C, 3 hydrogen bonds
- They flipped the bases 90deg, because they are perpendicular to the backbone
- Hydrogen bonds between the bases
- When you know the sequence of one strand you automatically know the sequence of the other strand because they are complementary
- Palindrome, reads the same sequence on each strand

Features of DNA double helix (B-DNA)

- Space filling model
- Water is excluded from the middle
- Bases are in direct contact; water is not a competitor for hydrogen bonds
- 2nm in diameter
- Rise is the distance from one base to another 0.34nm
- 10 bases to do a complete turn (full revolution)

- Major groove, more space, major surface through with proteins interact with DNA
- Most of the time proteins interact with DNA in this groove, and you can tell which bases are there just by interacting there
- Just about the right size for an alpha helix to fit
- Minor groove, narrower, sometimes used for interacting's
- Only 1 major groove and 1 minor groove all along the strand

Reason for major and minor grooves

- Down the axis of double helix
- Larger space on one side and a smaller space on the other side, they are asymmetrical
- They are not 180deg from eachother, inequal

Differences between DNA and RNA

- 1)DNA uses deoxyribose as the sugar
    - RNA has a hydroxyl group on the 2' Carbon
- 2) DNA uses T, with a methyl group
    - RNA uses U with a hydrogen in the same spot
    - But they both base pair identically
- 3)DNA is longer than RNA
- 4)DNA is double stranded, while RNA is single stranded
    - RNA can form a stem loop structure with itself
- 5) DNA can be modified but not as often as RNA
    - RNA is more diverse with its chemical structure
    - In general RNA has many more post-transcriptional modifications

RNA Structure

- tRNA:
    - A clover leaf structure
    - D's and Greek letters, modifications
    - 75-80 nucleotides in length
    - Regions of self complementarity
    - Antiparallel regions that form hydrogen bonds, this helps determine the 3D structure
    - Forms hydrogen bonds
    - the folding is determined by the sequence of the nucleotides
    - Forms hairpin loops, or stem loop structure
    - RNA cannot for BDNA structure but it still forms short helices
    - Anticodon loop recognizes mRNA
- rRNA
    - makes a quite complicated secondary structure
    - still have domains that the RNA is folded up into

Human DNA must be condensed

- If you take double helices and line them up end to end, our genome is about 102cm, and we have two copies per cell, so that is 2 meters that have to fit in a single cell
- Long thing in small space, we need a way of condensing

DNA packaging is highly organized

- DNA must still be accessible
- Chromatin= DNA and all associated packaging (eg.histones)
- Have to be able to access various points on the DNA (cant package into a single ball)
- Specialized proteins in a way so that it is condensed but is still accessible at various points
- Histones:
    - Histones are rich in arginine and lysine (amino acids)
    - Every third residue is Arg or Lys
    - Histones are under 200 amino acids in length
    - Question-Why are these proteins rich in lysine and arginine?
        - They are positively charged amino acids K and R

        - To favor the reactions between the protein and the DNA
    - Negatively charged DNA (from phosphate group)
    - They are conserved through evolution
    - This level of conservation is extreme

Nucleosome Core Particles

- Fundamental unit of packaging DNA
- There are two copies of each protein to be 8 total
- Histone H2A, H2B, H3 and H4
- DNA goes around like a spool, approximately less than 2 turns, 146bp
- Histone tails stick out

Nucleosomes

- Nucleosome = NCP + one linker
- Linkers in between, that can vary in length (3-80bp)
- Nucleosome consists of a Nucleosome core particle and a linker
- In total there is ~200bp per nucleosome
- Associate with eachother

30-nm chromatin fibers

- NCPs packed together
- Requires histone H1
- Structure still controversial
- H1 binds to the core particle and controls the angle at which the linker comes off
- Make a ~30nm structure
- Helix type structure, pack together in regular fashion

Higher Levels of DNA packing

- 30nm fibers form loops of 30-200kbp
- loops anchored to nuclear matrix, composed of proteins that are not histones

Chromosome Remodeling Complexes

- change nucleosome structure to allow transcription
- big proteins (ATP-dependent chromatin-remolding complex) that use ATP that can change the structure of the nucleosome core particle
- this can expose different parts of DNA, and to expose a specific sequence or to block a specific sequence from being transcribed
- influenced by covalent modification of histones
    - Tails of histones sticking out
    - These are sites which you can post transitionally modify the proteins
    - Add acetyl groups, methyl groups
    - Sites at which H3 can be modified
    - This makes ‘histone code’, number, types and locations that make a code that signal to the rest of a cell when DNA should be transcribed or expressed or silenced
    - Change affinity of DNA,
    - Or recruit other proteins to come in
    - Control locations on NCP

<u>Topic 15: DNA Replication</u>

Overview of DNA replication

- Replication is semi-conservative
- There are complementary strands
- Half old DNA, half is new

Steps in DNA replication

- 1)Initiation
- 2)Priming

- 3)DNA synthesis (includes proofreading)
- 4)Ligation

Initiation

- DNA is double stranded, needs to make an additional copy so each daughter cell will have the genetic material
- There are specific spots where this starts
- Have a defined sequence (replication origin)
    - Bacteria, only have one replication origin
    - Eukaryotes, have multiple replication origins
    - much longer DNA, thousands of origins
- No specific sequence as replication origin eukaryotes
    - Have many of them
    - Easy to pull apart
    - Must separate the strands to expose nucleotides
- AT rich, at base pairs are less stable
- Initiator proteins bind, to create this replication bubble
- Proteins bind to physically prevent the DNA from reanneling (called single stranded binding proteins)
- Protein called helicase at each end of the bubble, helicase binds to replication forks,
- Helicase unwinds the DNA

Priming

- DNA polymerase can't synthesize DNA without a primer
- Primase synthesizes a short (10-20nt) strand of RNA
    - Antiparallel
    - This is temporary so that DNA polymerase can have something to work with
    - Transcription always happens 5' to 3'

DNA synthesis

- DNA polymerase extends primer from its 3' end
- Binding proteins get pushed out of the way by DNAP
- DNAP synthesizes from 5'-3'
- DNA polymerase takes deoxytriphosphates
- And it try's them one at a time in the active site
- If the base pairing is not correct, the reaction is not favored and it releases the nucleotide
- When its complementary, it fits perfectly
- And allows the OH to displace the phosphate from the incoming nucleotide
- This is an energetically favored reaction
- Each time you do this you hydrolyze a triphosphate to a monophosphate
- Always 5'-3'
- Continuous synthesis, on the leading strand
- Discontinuous synthesis, DNA synthesized in chunks (okazaki fragments), lagging strand
- Lagging strand on one side is the leading strand on the other side

Replication Fork

- Leading strand on the top, Lagging strand on the bottom
- DNA polymerase and helicase create his complex together
- The sliding clamp, goes up and down the strands, job is to keep DNAP on task (processivity)
- Continuously make 5 x 10^5 bases per each primer (because of sliding clamp protein)
- Single stranded binding proteins are still present, which keeps the DNA stabilized
- Primase lays down a primer, that is going to be the primer for the next fragment
- Once this fragment is finished, the polymerase dissociates and sliding clamp comes off and DNAP binds to the next primer
- DNA strand bends around and the lagging strand polymerase associates with the helicase

- The DNA can dissociate onto the next primer

Proofreading

- Error 1 in 10^5 bases
- Mistakes: put in the wrong base, skip a base or add an extra base
- Double helix will not form the right structure, and base pair
- The two strand will not hold as tightly together
- DNAP pauses, and the strand will dissociate
- Exconuclease activity within the DNAP, (exo=outside) cuts DNA starting from the end 3'-5'
- Latches on to the 3' end and chews away nucleotides, catches 99% of mistakes that are made
- Therefore, error rate is 1X10^7 (mistake 1 every 10,000,000 bases)
- Endonuclease cuts from somewhere in the middle

Ligation

- DNA fragments must be joined together
- Ligation in Bacteria:
    - At some point the DNA runs into the RNA primer
    - The polymerase (DNAP3) finishes and dissociates
    - Separate polymerase (DNAP1) comes in to finish the job
        - Chews up the 5' end, degrades primer and synthesizes replacement DNA
        - Process is called nick translation
        - DNAP1 cannot deal with the nick
    - Complete double helix, all engaged in base pairing, except in one spot there is a break in the sugar phosphate back bone (nick)
    - There is base pairing at that nick, but the backbone isn't connected, ligation is done by ligase
        - Which seals the Okazaki fragments
        - Ligase joins the backbones together
        - Uses ATP, lose pyrophosphate
        - Attach adenosine monophosphate on the one side, OH does an attack and removes that
        - Coupling, hydrolysis of ATP to AMP, and pyrophosphate to the joining of this strand
        - Use energy from ATP to attack the DNA strand,
        - Intact strand is complete
    - Where would you find DNA ligase in a replication bubble?
        - On the two lagging strands

Telomeres

- Protein-DNA structures at ends of chromosomes
- 5,000 to 15,000 bp, many short nt repeats
- Linear chromosomes,
    - Structures at the end of the chromosome that indicate the end of the chromosome (telomeres)
    - Structure that forms that signal the end of the chromosome, and allows the cell to not go into crisis mode and try to fix the ends
- Composed on protein and DNA
- Many proteins associates with this
- DNA structure: series of repeats of nucleotides
- TTAGGG, six repeated bases from thousands in humans
- Non coding DNA
- Repeat sequence
- Called t-loop
- 3' overhang: 3' end is longer
- Single stranded

- Displaces and earlier part of itself
- No coding genes

Telomere Replication Problem

- Problem with having a linear chromosome, when you get to the end you cannot completely replicate the end of the lagging strand
- Helicase falls off, and you still have a section that a primer wasn't set down so helicase can't work, and fill in the spot
- Happens on both end of DNA
- We have to make the template longer
    - Enzyme telomerase that has an internal template
    - Telomerase can bind to the repeat sequence and will synthesis those 6 bases
    - It shifts and synthesizes it again
    - And allows you to extend the 3' end
    - A mechanism by which you can put the helicase back on
    - Always have 3' overhang
    - This allows you to maintain length of chromosome, so you don't lose DNA
- Only works in an embryo or a sex cell
- After you are born, you stop producing telomerase in all tissues
- Chromosome are at a permenant length, and every time your cells divide you chromosomes actually get shorter
- Would be a problem if there were genes in telomeres, lose the repeats it does't matter
- At some point you do start cutting into important DNA, the cell has a mechanism to detect when the chromosome will be diving too short and the cell just stops reproducing and enters "senesance" (cell old age)
- No longer going to divide
- Thought to be part of the aging process
- Cancer cells: re-express telomerase such that they can reproduce indefinitely

Some antivirals Target DNA synthesis

- Anitvirals that target DNA replication of viral DNA
- AZT, looks like sugar with base
    - OH is replaced by azide group (n3)
    - Cells recognize as nucleoside and will phosphorylate 5' carbon of sugar
    - And this looks like deoxytriphosphate molecule
    - When this is incorporated in DNA strand you get chain termination because you cannot attach to n3
    - Our polymerase recognizes not to incorporate this, but retroviral DNA doesn't
    - AIDs virus cannot then copy its DNA
    - Doesn't eliminate AIDS but it prevents it from spreading
- Acyclovir
    - Halts replication and stops from replicated

<u>Topic 16: DNA Repair</u>

Basic Mechanism of DNA repair

- DNA damage= any unintended physical or chemical change in DNA
- Over general mechanism:
    - Step1) remove damage part
    - Step2) Use undamaged part as a template, makes a nick
    - Step3) Ligase seals the nick

Causes of DNA damage

- 1)Copying Mistakes: DNA polymerase makes an erroe once every ~10^7 nucleotides
    - mismatch
    - insertion
    - deletion

    - these are types of mistakes DNAP can leave behind
    - Result of unrepaired replication error:
        - Distortion in double helix, whatever reason the problem isnt fixed and replication occurs again
        - Bottom strand is fine
        - But the top strand calls for a different base and now we have a mutation
        - No way for the cell to realize that an issue occurred before
        - Now part of the lineage
        - And this is unfixable
        - Cell has to fix this mistake before it gets too that stage
        - Cell wants to elimate any change
        - A lot of mutations are deleterious and cause damage
- 2)Depurination: los of A or G base
    - blocks DNA replication (overcome by translesion DNAP)
    - Losing purines
    - Happens spontaneously, the base is lost
    - Lose nitrogenous base but you do not disrupt the backbone
    - Breaking of glycosidic bond and you get a
    - A-BASIC SITE, A means not because it has no base
    - This is a problem because it blocks DNA replication
    - Don't know which base to put in during replication
    - When DNAP comes to absaic site, it stops, can cause uneven splitting of daughter cells
    - Cause chromosome rearrangements and cell death
    - Translesion DNAP come in and replicate past the site of damage
    - Result of unrepaired depurination:
        - Bottom strand is fine
        - Translesion polymerase comes in and either ignores the base or also incorporates some base
        - Bulge, or put in wrong base
        - Finishing replication is a higher priority, rather risk changing genome than cell death
        - Why translesion doesn't synthesize the whole strand, translesion go past the site of damage because their active sites are a lot less specific, don't do as good of job as copying what's on the template
        - Only call in when needed
        - Translesion only synthesizes a few bases and then gets kicked out and regular DNAP comes in and finishes
- 3)Deamination: conversion of amine to carbonyl
    - Most common at C
    - Result in a substitution of U for C
    - used to be a C now its a U (deamination)
    - polymerase comes and adds an opposite to U (post replication)
    - U is still identified as a problem
    - if repair happens again, we move damage and look at opposite strand so see what should be filled with
- 4) Pyrimidine dimers: UV light causes ring formation between adjacent pyrimidines
    - Blocks DNA replication (overcome by translesion polymerases)
    - Have to use translesion polymerases, they have to accommodate this type of damage and move past it
    - when UV light shines into body and causes ring formation between pyrimidines
    - bases connected together covalently
    - DNA polymerase is not able to synthesize past site of damage (dimer)

    - stalls replication because polymerase can't get past it
    - this is how asking gets damaged from the sun
- 5) Environmental Factors
    - Ionizing Radiation
        - Causes single and double bond breaks
    - Chemical mutagens
        - Chemically modifying bases, causing problems and mutations
    - Mechanical stress
        - Physically pulling on DNA
    - Result: chemical changes in bases, or breakage of sugar-phosphate backbone
    - Leads to chromosome instability

DNA repair systems

- 1)Proofreading during DNA replication
- 2)Mismatch repair: repair of replication mistakes
    - specifically designed to fix mistakes by DNA polymerase
    - DNA mismatch as a result of replication
    - how does the cell know which should be fixed top or bottom?
        - in newly synthesized strand we leave behind nicks
        - on leading strand, there is something that comes and introduces nicks
        - nicks aren't permanent, eventually they get fixed by ligase
        - these allow us to determine which is the newly synthesized strand and the problem is assumed to be on that one
        - exonucelase comes in and chews up the strand
        - nick is found and DNA is being chewed away from this nick to the sit of the mismatch
        - you get a big gap here and all the bases get removed — even though the ones that were done properly
        - the gap has to then be filled in, ligase fixes the gap
    - this costs quite a bit of energy
        - cell will go to great lengths to fix these mistakes
    - this catches about 99% of mistakes made by DNA polymerase
    - this means we get 1 mistake per every 10^9
- 3)Base excision repair (BER)
    - Repair of modified bases, absaic sites, single strand breaks
    - Short-patch or long-patch
    - Minor sorts of damage

Short patch base excision repair

- Deaminaion event
- first thing that happens is the removal of the base
    - cleave glycosidic bond
- sugar phosphate backbone remains intact
- then we get an endonuclease making a nick in the backbone
- base is inserted by DNA polymerase
- second endonuclease comes in and cuts off flap structure to yield a clean nick
- then DNA ligase repairs the nick
- "BER" because base removal is first step
- shot patch because its only one base
- will repair things that affect 1 base

Long patch base excision repair

- if there is a change to sugar phosphate backbone changing the chemistry
- base is correct
- still wants to fix backbone so base gets removed
- nick is made
- if endonuclease cant cut — we need to move to a place where it can

- polymerase comes back and synthesizes more cases
- displaces whatever bases are there
- long patch because you have so synthesize multiple bases

DNA repair Systems

- 4) Nucleotide excision repair (NER)
  - repair of pyrimidine dimers and damage that disrupts the double helix
  - larger types of damage to the DNA
  - endonuclease cuts on two sides at the site of damage
    - we get endonuclease cuts in 2 spots on either side of damage
    - helicase unwinds the DNA strand, the strand then floats away and gets degraded
    - this leaves us with a large gap, a single stranded section
    - DNA polymerase comes in and fills gap, ligase resolves the nick
    - we remove a bunch of nucleotides compared to just a base
- 5) Double stranded breaks
  - causes problems for replication
  - 2 major pathways
  - A) Non Homologous Ends joining
    - 2 strands stay together but get cut
    - jam the ends together
    - cell sees this and an emergency and would rather just fix quickly
    - cell looks for anyway it can to join the strands somehow
    - end up with deletion of sequence and lose some nucleotides
    - used in organisms with large genomes
    - Break repaired with some loss of nucleotides at the repair site
  - B) Homologous Recombination
    - using a copy and taking the exact info
    - difficult and time consuming
    - have another copy close by
    - cell allows us to use info on one strand on other
    - don't lose any information
    - fixed but haven't lost any nucleotides
    - copies a separate copy to ensure its right
    - break repaired with no loss of nucleotides at repair site

Topic 17 & 18: Prokaryotic and Eukaryotic Transcription

- Gene Expression
  - Humans genome encode 21000 genes
  - In any cell type expresses 10,000
  - Genes are expressed in different levels
  - The expression of the correct genes is essential for growth and differentiation
- Differential Gene Expression
  - Similar in brain and liver tissues
  - But there are also brain or liver specific proteins
  - That are highly or poorly expressed in each tissue
  - Some proteins are found in all, but some aren't
- Significance
  - Most diseases are due to altered expression of one or more genes
  - If you can manipulate gene expression back to normal, you can cure the disease

The Central Dogma

- Defines genetic flow of information from DNA-> RNA->protein
- Transcription and translation

Ways to control regulation:

- Transcription:

    - 1)Initiation
    - 2)Elongation
    - 3)Termination
- RNA Processing:
    - Pre-mRNA becomes mRNA
    - 4)RNA editing
    - 5)5'capping
    - 6)splicing
    - 7)3' polyadenlyation
    - 8)mRNA export from the nucleus to the cytoplasm
    - 9)mRNA degradation, the amount of mRNA found in the cell depends on BOTH its rate of synthesis and its rate of decay
- Translation: (RNA protein)
    - 10)Initiation
    - 11)Elongation
    - 12)Termination
- Protein Modification:
    - 13)Phosphorylation
    - 14)Acetylation
    - 15)Cleavage (insulin)
    - 16)Protein degradation- the amount of protein in a cell depends on BOTH its rate of synthesis and its rate of decay

Why is there an RNA step in Gene expression?

- 1)The RNA step provides an amplification which contributes to differential gene expression
    - all you to have a lot of RNA from one gene and little of the other
- 2) Since RNA can be degraded fast, expression of a gene can be stopped quickly
    - allows you to have an on off switch for gene expression
- 3) provides different opportunities to regulate expression
    - RNA processing
    - RNA export from the nucleus

Prokaryotic Transcription:

- Gene and promoter structure
- RNA polymerase
- Mechanisms of Transcription
- Regulation of transcription
    - Trp operon
    - Lac operon

Transcription:

- transcription is the first step in gene expression
- often most regulated step in gene expression

Terminology

- Promoter: the DNA sequence required to initiate transcription of a gene or operon
- Termination: the DNA sequence required to stop transcription
- Operon: set of genes transcribed form a single promoter an thus expressed from a common RNA

Consensus Sequence:

- The most frequent base at each position in a group of functionally related DNA elements
- Apply to DNA elements, RNA elements, protein elements
- The most frequent base at that specific position
- The consensus sequence is derived, and the consensus sequence does actually have to be an element sequence

Bacterial Operon Structure:

- Promoter sequence, 100 base pairs in length
- 3 elements in the bacterial promoter
- +1 is where transcription begins
- the next base is +2 and so on, and down stream is -1, -2, there is no zero
- there are consensus elements at -10 and -35
- consensus sequences at -10 : TATAAT, -35: TTGACA
- these sequences where found by comparing the sequences of many E coli promoters
    - common elements that suggested functional importance
- took a group of Ecoli promoters, and all aligned them around +1, found
- -10 is called the Pribnow box
- not always exactly at -10, but on average at -10
- -35, all bases had a high percentage of commonality
- How do you test that these regions are important?
    - Way to test is to make a mutation to the -35 region and look at the amount of gene expression of that gene

Bacterial RNA Polymerase

- Enzyme that converts DNA to RNA
- Makes RNA transcripts using DNA as a temple and nucleoside triphosphates (NTP's) as substrates
- It is a multi-subunit enzyme
- Core enzyme can make RNA in a test tube, but does not recognize promoters
- The promoters sequence specificity of RNA polymerase is determined by the sigma subunit
- Core enzyme cannot recognize promoters
- Polymerase Core + Sigma subunit = RNA polymerase Holoenzyme
- Another subunit is recognized from promoter recognition (sigma subunit)
    - Responsible for making contact with the -10 and -35 sequences

Steps in the Initiation of Transcription

- 1)RNAP holoenzyme binds to the promoter forming a **closed complex**
- 2) RNAP unwinds the DNA strands at the start site forming an open complex
    - pull the strands apart on its own
    - forming a bubble around the start site
- 3)The first NTP is brought to the template-no primer for RNA synthesis is required
- 4)Using NTP's (ATP, GTP, CTP and UTP) as substrates, chain elongation proceeds in 5'-3' direction, following base pairing rules
    - forming phosphodiester bonds and pyrophosphate is released
    - the RNAP holoenzyme is stuck in the same spot (like a wheel spinning in the snow)
- 5)After the addition of 5-10 nucleotides, sigma falls off the holoenzyme
- 6) the transcription bubble moves down stream (5'-3') with the template DNA reannealing behind
- 7)RNA synthesis continues until terminator is reached and RNAP falls off
- 8)Sigma rebinds the core, and the cycle can repeat again

Regulation of Gene Expression

- How are different bacterial promoters transcribed at different levels?
- 1)Some genes have -10 and -35 sequences that are better at recruiting RNA pol
    - not sensitive to environmental change
- 2)There is more than one sigma factor- each recognizes different promoter sequences
    - The most prevalent is sigma70 "housekeeping"
    - Others Include:
        - sigma54 (nitrogen metabolism)
        - sigma38 (starvation)
        - sigma32 (heat shock)

    - cells have different sigma factors that will induce genes under specific circumstances
- 3)Gene specific regulatory proteins
    - responsive to environmental change (dynamic)
    - **Negative regulation**-proteins factors repress transcription (eg, Trp and Lac repressor)
    - **Positive regulation**- factors activate transcription (eg. Catabolic Activator Protein)

Negative Regulation: Trp operon of E.coli

- **Trp operon**, encodes 5 genes required for **tryptophan biosynthesis** in E.coli and are transcribed for a common promoter
- Trp operon is only expressed when there is little/no tryptophan in the cellular environment
- If there is tryptophan all ready in the environment they don't need to make it
- Only turn trp on when the cell has to

How is the trp operon regulated?

- Between -10 and -35, there is a DNA sequence called the trp operator that binds a protein called the trp repressor (protein)
- Trp operator is the DNA element
- Trp repressor is the protein that binds to the DNA
- Trp repressor binds tryptophan where there is lots of tryptophan in the cell
- The trp repressor-tryptophan complex binds the operator DNA
- Trp repressor-tryptophan complex, block RNAP from the promoter (sits and blocks RNA pol so no transcription can occur)
- When the [tryptophan] is low, tryptophan no longer binds to the trp repressor and dissociates
- Trp repressor no longer binds to trp operator
- RNA pol binds the promoter and transcription results
- High [tryptophan] genes are off
- Low [tryptophan] genes are on

What is the structural basis for the regulation by tryptophan?

- Trp repressor is 107 residues (small protein)
- Trp repressor-monomer, the protein is made up of 6 alpha helices with unstructured loops in between
    - Alpha helicies 4 and 5 and the loop in between, make up a helix-turn-helix motif
    - Which interacts with dna
    - Common DNA binding sequence element in DNA binding proteins
- Trp repressor dimer, protein dime has 2-fold symmetry
    - (many dna binding proteins work as dimers)
- therefore there is two sites with helix 4 and 5
    - two helix 5s recognize adjacent major grooves in the operator DNA
    - the recognition sequence is highly specific
    - alpha helices are just the right size to fit into the major grooves
- Tryptophan causes a conformational shape change in Trp repressor that allows it to bind DNA
- When tryptophan isn't there, there is conformational change in the protein, where the helix 5 tilt inwards and cannot make contact with the grooves
- Shape change allows it to bind to DNA
- Protein interacts with the bases

General Themes to take from the trp operon

- 1) trp repressor is a site-specific DNA binding protein
- 2) there is a binding site for trp repressor within he trp promoter
- 3) trp repressor inhibits transcription by blocking RNA pol from the promoter

- 4) trp repressor is responsive to an environmental signal (ie. [tryptophan])
    - Question-Mutation in trp operator such that it doesn't allow the trp repressor to bind.
    - This results in the Trp operon is always transcribed
    - RNAP will always have access to the promoter

Lac operon-expression is required for the metabolism of lactose

- Negative control: lac repressor
- Positive control: catabolite activator protein
- Question- In which medium are the genes required for lactose metabolism expressed in E.coli?
    - In a medium containing lactose but not glucose
    - Only when lactose is present, the genes are turned on
    - But all cells prefer to use glucose as carbon source
    - If the cells have glucose and lactose, they wont turn on the lac gene (rather use glucose)
    - Catabolite repression

Players in regulation of the lac operon

- Lac repressor: DNA binding protein
- Catabolite activate protein: transcriptional activator
- Lac operator: DNA element that binds the repressor
- CAP binding site: DNA element that bind CAP
- Lac repressor binds DNA
- Lactose-Lac repressor does not bind DNA
- CAP binding to DNA is regulated by the concentration of glucose in the cells (cAMP is the second messenger)

Mechanisms:

- Lac operator is downstream of the -10
- Upstream of -35 is the CAP binding site
- Glucose is high, lactose is high
    - Operon is off
    - CAP not bound
- No lactose, Glucose present
    - Operon off
    - Lac repressor bound
    - CAP not bound
- No glucose, No lactose
    - Operon off
    - Lac repressor bound
    - CAP is bound
- No glucose, Lactose is present
    - Operon is on
    - CAP interacts with RNA pol and helps it come to promoter
    - RNA transcript is made
- Question-Why is the Lac operon not expressed when CAP is not bound?
    - The -10 and -35 sequences have low affinity for RNA pol
    - Do not have strong affinity for RNAP on its own
    - Having CAP stabilizes that RNA pol to bind
    - Not all -10 and -35 are equal, or needs assistance to work well

Six Differences in Transcription in Prokaryotes and Eukaryotes

- 1) Three RNA polmerases in eukaryotes
    - pol I: most rRNA genes
    - pol II: mRNAs
    - pol III: tRNA, 5S rRNA (of the ribosome)
- 2) No operons in eukaryotes

    - each genes is transcribed as a single unit
    - Prokayrotes: polycistronic
    - Eukaryotes: monocistronic
- 3) Promoter Structure
    - eukaryotic promoters do not have -10 and -35 sequences- there is no sigma factor
    - Promoter recognition is determined by a set of proteins, one of which recognizes a TATA-element
        - Recruits TATA binding protein
        - A lot of factors
- 4) In eukaryotes regulatory proteins often bind DNA several thousand base pairs from the start site
    - often many kb away
    - often factors bound way upstream
    - Works through DNA LOOPING:
        - Factors that are far away can actually interact with the transcription machinery
        - An activator protein can bind to an enhancer binding site and a mediator joins to allow transcription to begin
- 5) Combination Control: groups of proteins work together to determine the expression of a gene
    - you can have on or off, or anywhere in between
    - high expression, medium, low
    - broad range of gene expression through this control
- 6) Nucleosomes are higher order chromatin structure regulate transcription
    - Generally, nucleosomes repress transcription by blocking access of RNAP and regulatory transcription factors
    - nucleosome impedes transcription
    - nucleosome can in fact move (shift), making the promoter more accessible
    - can regulate gene expression on and off just by changing their structure
    - Nucleosome positioning is influenced by many proteins
        - Modifying nucleosome to change chromatin structure

Regulation in Eukaryotic cells required for galactose metabolism in yeast

- When do yeast express the genes required to metabolize galactose?
    - In the presence of galactose and the absence of glucose
    - Genes required for galatose metabolism
    - No glucose present because cells would prefer to use glucose over galactose
- Gal10 is highly expressed in the presence of galactose, 1000 units
    - Rafinose is present, 100 units are present
    - About 1 unit in glucose
    - Glucose and Galactose, just about one unit of expression
- Regulatory protein is called Gal4 activator protein
    - It induces expression of GAL10 in galactose containing medium
    - Positively induces transcription
- Gal4 is a DNA binding protein that recognizes sites in the GAL10 promoter
    - Acts like a dimer
    - And fits in the major groove
- There are four Gal4 binding sites in the GAL10 promoter
    - As a group there are called UAS
- In a galatose medium, Gal4 protein helps to recruit RNAP2 to the promoter, and activates gene expression
- The GAL10 promoter is not expressed in the absence of galactose (eg. When Rafinose is the carbon source)

    - In this case, there is a protein called Gal 80 that binds to Gal4, and blocks Gal 4 functioning
    - This stops it from recruiting RNApol
    - Gal80 inhibits Gal4 function
    - Sensitive to the medium (environment)
    - Don't get expression
    - Gal80 can account for 100fold less of gene expression
- Gal 10 promoter is not expressed in medium containing glucose: catabolite repression
    - The Mig1 protein represses GAL10 in glucose
    - Binds to the promoter and block RNA pol from coming in, and this sterically inhibits expression
    - No gal4 is bound to the Gal10 promoter even if galatose is presence
    - Mig1 also blocks expression of the Gal4 gene,
    - This also blocks the Gal4 promoter, so Gal4 cannot be expressed and never bind to the Gal10 promoter

Key Points

- 1) the GAL genes are subject to positive and negative regulation (combination control)
- 2) Activation of galactose regulated genes requires a site-specific DNA binding protein (GAL4)
- 3)Gal4 helps recruit RNApol to the promoter
- 4)Activity of Gal4 can be regulated-by Gal80 protein and by Mig1-regulatory proteins are themselves subject to regulation
- 5) The system is sensitive to environmental signals (glucose and galactose)

<u>Topic 19: RNA Processing in Eukaryotes</u>

Formation of mRNAs: pre-mRNA-> mRNA

- RNA Processing in Eukaryotic cells
    - 5' Capping
    - 3' polyadenylation
    - Splicing

5' Capping:

- At the 5'end of eukaryotic mRNA is a 7-methylguanosine that is added

Importance of the 5' Cap

- Marks the 5' end of the mRNA as being intact (like a signal)
- Required for:
    - mRNA export from the nucleus
    - translation of the mRNA

3' Polyadneylation:

- almost all eukaryotic mRNAs end with a long (~300 base) PolyA tail
- how its added:
    - there is a signal that recruits an enzyme that cuts the end of the DNA
    - and adds a polyAtail (-----AAAAAAAAA)
- Importance:
    - Marks the 3' end of the mRNA as being intact
    - Protects the mRNA from degradation
    - Required for:
        - mRNA export
        - translation

RNA Splicing

- protein encoding sequence (exons) are interrupted by noncoding sequences (introns)
- bacteria don't have these things, only found in eukaryotes
- introns are 'spliced' out of the primary transcript to give the mature mRNA
- Exons=expressed, Introns=interruptions
- Human beta-globin gene 3 exons, 3 intron

- Humans factor VIII gene, has 26 exons and 24 introns
  - This gene is over 200,000 nucleotides
  - Most of this gene is introns
- Why bother splicing:
  - Splicing increases, the coding capacity of the genome
  - **Differential Splicing**: An RNA can be spliced in different ways to create related but distinct proteins
- In problems:
  - Include all combination of just the exons, not introns included
- Splicing Pattern is often tissue specific (muscle protein, alpha-tropomyosin)

Mechanism of Splicing

- There are 3 specific sequences required for splicing in the RNA:
- 1) 5' splice junction
  - beginning of the intron
- 2) 3' splice junction
  - end of the intron
- 3) Branch point
  - middle region of intron
- Introns are removed in two consecutive phosphodiester transesterification reactions
  - 1)Attack by the 2'-OH of the adenine at the branch point with the 5' splice junction
    - You get a cool 2'-5' linkage in a lariat form
  - 2)Attack by the 3' OH of the free 5" splice junction with the 3' splice junction
    - lariat (loop) gets degraded

The Splieosome is the enzymatic machinery for splicing

- composed of snRNPs which contain both RNA and protein U1, U2, U4, U5 and U6
- Through RNA-RNA and protein-protein interactions, the spliceosome positions the RNA for splicing
- U1 snRNP at the 5' junction and the U2 snRNP at the 3' junction
- The rest of the snRNP molecules come in and assemble together to preform the function
- Self-Splicing RNA (pg253-256 in text)
  - Able to self catalyze its own conversion into mRNA

Splicing and Human disease:

- Abnormal splicing of the bet- globin RNA can result in hemoglobin deficiency
- Mutations can result in problems in the splicing in Beta-Thalassemia
  - Mutation in the 3' splice junction of the second intron
  - Cant recognize the 3' junction and use a Cryptic site 3' within the intron, and you get intron RNA in the mRNA
  - OR, skip the second exon, and misses the 3' junction of the first intron

RNA Processing occurs when Transcription is ongoing

- mRNA export from the nucleus requires proteins that interact with the 5' cap and the pol A tail, and with specific protein carriers
- transport occurs through the nuclear pore
- Nuclear Pores are selective gates

Topic 20: Translation

Translation: mRNA to Protein

- Genetic Code
- tRNA
- Aminoacyl tRNA synthases
- Ribosome
- Mechanism of translation

- Initiation
- Termination antibiotics

Genetic Code

- The genetic code “spells” out the amino acid sequence in 3 “letter” “words” called codons
- Why is each codon 3 bases?
    - Question-how many different codons does the cell need?
        - 20
        - because of the 20 amino acids
    - Question-Given that any nucleotide can be A, U, G or C how many base codons are there?
        - 64
        - one letter code with A, G, U C would provide 4 codons
        - a two letter code would provide 4x4=16 codons
        - a three letter code would provide 16x4+64 codons
        - As there are 20 amino acids, a 3 letter codons is sufficient

Key features of the genetic code

- It is (almost) universal
- The code has evolved once
- Non overlapping
    - Overlapping would place significant restrictions on what amino acid residues could follow each other
- No Gaps
- Redundancy- some codons specify the same amino acid
    - 64 possible ‘words’ but only 20 amino acids
    - often occurs in the third position of the codon
- Functionally related amino acids have similar codons
- Question-why do functionally related amino acids have similar codons?
    - Increases the change of a functional protein in the case of a single base mutation

Types of Mutations

- Missense Mutation
    - Results in a single amino acids change
- Frame shift mutation
    - Insertion or deletion of bases can change in the reading frame of the protein from the point of the mutation onward
- Question-which is least deleterious to the function of a protein?
    - Insertion of 3 consecutive bases to the gene
    - Doesn’t cause a frame shift but just 1 new amino acid
- Nonsense mutation
    - A nonsense mutation results in premature termination of the protein

tRNA

- Bring the amino acid to the growing polypeptide chain
- Similar structure for all tRNA
- ~80 bases
- attached to an amino acid at the top
- and they have 3 loops
- stem loop structure that is due to base pairing with itself
- there are two key single stranded regions
    - 3’ acceptor site
    - anticodon

Wobble

- most organisims have fewer than 45 different tRNAs
- how can all 61 codons be used?

- Some tRNA species must pair with more than one codon
- For some tRNA base pairing between the anticodon and the codon only requires matching two positions

Aminoacyl tRNA Synthases

- Couples the 3' end of a tRNA to its correct amino acid
- Generally only one for each amino acid
- Charging of tRNAs
  - Store energy from ATP in a high energy ester linkage
  - Many have proofreading function

Ribosomes

- Catalyze protein synthesis
- Ribosomes have two subunits each composed of RNA and protein
- They have three sites for binding tRNA on the ribosome
  - A for the site that binds aminoacyl tRNA
  - P for the site that binds peptidyl tRNA
  - E for the site from which tRNA exits
- mRNA is bound close to the A and P sites

Steps in Translation

- A peptidyl-tRNA is in the P site of the ribosome
- Step1)
  - An aminoacyl-tRNA binds to the A site of the ribosome
  - Requires base-pairing between the tRNA and the codon
- Step2)
  - The energy from the ester bond of the peptidyl-tRNA in the P site is used to form a new peptide bond between the amino acids in the A and P sites
  - Ribosome RNA catalyzes this
- Step3)
  - Peptide bond formation is coupled to a conformation change in the ribosome that shifts the large subunit "forward"
  - Towards the 3' end of the RNA
  - This conformational change repositions the tRNAs
  - P into E and A into P
- Step4)
  - The small subunit moves forward (5' to the 3' end) exactly 3 bases
  - The tRNA leave the E site
  - Step1 is repeated

Specificity of translation comes from?

- Correct charging of the aa-tRNA synthetase
- And base pairing of the anticodon and codon in the A-site of the ribosome

Key Points for Translation

- 1) Translation occurs in the cytoplasm
- 2) Occurs in a 5'-3' direction along the RNA making a protein from N to C terminus
- 3) The mRNA is decoded one codon at a time
- 4)Energy for peptide bond synthesis comes from the high energy aa-tRNA ester bond-indirectly from ATP
- 5) Complex reaction involving:
  - a) both RNA and protein molecules
  - b) conformational changes in the ribosome
- 6)Specificity comes from the:
  - a)Aminoacyl tRNA synetases (putting the right amino acid on each tRNA)
  - b)Requirement for base pairing in the A site of the ribosome

Initiating Translation

- determining the reading frame is important

Eukaryotic Translational Initiation:

- Small ribosomal subunit with transitional initiation factors bound, has initiator tRNA bound to it (Met)
- This complex, (subunit and initiator tRNA) binds to the 5' cap of the mRNA
- Ribosome moves from the 5' to scan for the first AUG, Initiator tRNA moves along RNA searching for the first AUG
- Initiation factors dissociate
- The large ribosomal subunit binds
- Aminoacyl- tRNA binds and step 1 begins
- Question-What might happen if there was a hairpin loop in the mRNA between the 5' end and the AUG?
    - Might decrease translation
    - Because the ribosome might have a hard time getting to the AUG
    - Might stall and slow down translation
    - Example of regulation of translation, of secondary structure in RNA

Prokaryote Translation Initiation

- Question- Why must translation initiation in bacteria use a mechanism other than 5' scanning?
    - There are multiple proteins encoding from a common mRNA in bacteria
    - Multiple genes, polycistronic
- Ribosomes recognize internal ribosome binding sites found just upstream from each functional AUG
- Just before AUG is a ribsosomal binding site,
- Internal ribosome binding site, can be regulated and have a different affinity so you can get different amounts of proteins made from a single mRNA strand

Termination translation

- Requires one of the three stop codons
- Specific termination factors
- Each stop codon has a specific termination factor
- Mechanism:
    - When there is a stop codon in the A site of the ribosome
    - This signal for the termination factor to come into this site
    - Termination factor enters, and this triggers the ribosome to catalyse the addition of water into the peptide chain
    - This releases the peptide
    - And the ribosome dissociates from the RNA

Antibiotics and Translation

- Many antibiotics block bacterial translation
- Translations importance and intricate molecular interaction make it a prime therapeutic target
    - Tetracycline
    - Chloamphenicol
    - Puromycin
    - Streptomycin
    - See table 7.3, and understand mechanisms

Catastrophes facing humankind

- Cant be fixed:
    - Celestial collision
    - Nuclear catastrophe
- Can be fixed:
    - Global warming
    - Food crisis
    - Water crisis
    - Global pandemic (ebola, SARS, avian flu)
    - Antibiotic resistant bacteria

Recombinant DNA Technologies

- Technique which recombine DNA fragments from different sources into new molecules with unique features

Synthetic biology

- Engineering of biological systems, with the goal of carrying out novel functions or preforming existing functions more efficiently
- Manipulating genomes of organisms is not new
- Humans been doing it for 1000's of years through classical genetic selection
- Plant/crop and livestock breeding are examples of humans manipulating the DNA of other organisms

Comparison of classical genetic techniques and RDT

- Classical Genetics:
  - Slow: limited by the breeding time of the organism and chance genetic events
  - Exchange of genetic materials is limited to breeding species
- Recombinant DNA TECHNOLOGY
  - Rapid as quick as a few days in some organisms
  - No limitations

Significance of Recombinant DNA technology

- 1) Research-key tool to understand cell and molecular structure function
- 2) Biotechnology-major impact on society

Medicine

- Drug production and design
- Human insulin
- Human growth hormone
- Vaccines
- Artemisinin (malaria)
- Interleukins
- Diagnosis of Disease- detect pathogens and disease causing genes through their DNA signature
- Genetic counseling- does an individual carry a disease related allele
- Gene Therapy
  - Many diseases result from a defective gene
    - We can add back functional genes to cure disease (CRISPR-Cas9 gene editing)

Agriculture

- Producing crops with unique features
- Vitamin A enhanced rice
- Cold drought resistant crops
- Pest resistant corn
- Non allergenic peanuts
- Production of novel molecules in crops (like drugs)
- Manufacturing, biofuels, spider silk

Forensics and Law

- DNA fingerprinting

Question-In a DNA sequencing reaction you generally add ~100ng of DNA. Assuming the DNA is 3000bp and one base pair has a mass of ~600Daltons, approximately how many molecules of template are added?

- 1 mole of DNA of 3000bp has a mass of: 3000bp x 600g/bp =18 x 10^5 or 2x10^6g
- 1mole-> 2x10^6g
- xmole->100ng (1x10^-7 g)
- X= 1X10^-7 / 2x10^6
- X=5X10^-14 moles
- 1 mole = 6.02X10^23 molecules
- avogadros number x 5X10^-14= 3X10^10 molecules

Topic 21 & 22: Recombinant DNA Technology and Genetic Engineering

Recombinant DNA Technology

- DNA hybridization- Northern blotting
- cDNA synthesis
- PCR
- Gel electrophoresis
- Restriction Enzymes
- Plasmids, cloning, transformation
- DNA sequencing
- Site direct mutagenesis

Recombinant DNA technology Case Study

- you have identified a human protein that will be of big value
- goal is to mass produce this protein (YFP)

Why is this important

- many proteins are difficult to get from the native source
- RDT can dramatically increase expression and facilitate purification

The process of expressing YFP will require:

- 1)cloning YFG
- 2) introducing YFG into E.coli
- 3) purifying the protein from E.coli

Ecoli are often used to express YFP

- grow quickly and inexpensively
- genetic engineering is simple
- mutlicopy plasmids and strong promoters can drive expression
- protein extracts are easily made

The first issue we need to think about

- the clone the gene, we need a source first
- Express our favorite human gene in ecoli, Can genomic DNA be used as the source?
    - Only for the few genes that lack introns
    - Bacteria don't know how to handle introns
- Introns are a problem in expressing eukaryotic genomic DNA in E.coli because the bacteria can not process them
- We will use mRNA (cDNA) to clone YFG
- mRNA by definition does not contain introns
- each mRNA encodes one gene
- Do all tissues express YFG as mRNA?
    - No

Nucleic Acid Hybridization

- Denaturation/melting
- Renaturation/ annealing/ hybridization

DNA Denaturation

- Duplex DNA -> Single stranded DNA
- Can covert using heat or hydroxide
- Melting temperature (Tm)
    - Temperature at which 50% of the molecules are single stranded, given a DNA at a defined concentration

Factors that determine the Tm

- Intrinsic Factors:
    - Which has higher tm?
        - G:C rich DNA
        - More G,C the higher the melting temp
        - More base stacking and more Hydrogen bonding
        - Therefore higher tm

        - Which has higher tm, long DNA and short DNA or both?
            - A long DNA
            - The longer the DNA, the more interactions the higher the TM
            - Tm~ 4deg(G:C) + 2deg(A:T)
        - Which has a higher tm, DNA with a mismatch, or perfect sequence
            - DNA that is perfect, because the mismatch has less hydrogen bonding
- Extrinsic factors that affect the tm:
    - Salt, raise the tm of the DNA
        - Salt concentration, the higher the salt concentration, the higher the tm
        - Positively charged ions shield the repulsive interactions of the negatively charged phosphate backbone
    - Molecule that interferes with hydrogen bonding, (urea) lower the tm
        - Cause DNA to fall apart more easily
        - Solvents that disrupt hydrogen bonding interfere with base pairing and lower the Tm

Melting is reversible

- Process is called renaturation, hybridization or annealing
- Hybridization requires complementary stands but some mismatched can be tolerated
- Can form a DNA/RNA hybrid
    - Some mismatch can be tolerated
- Hybridization:
    - Hybridization will detect complementary nucleic acid sequences
    - Hybridizations are often done after transferring nucleic acids to membranes
    - Southern blot: DNA probe to DNA on membrane
    - Northern blot: DNA probe to RNA on the membrane
- Example of a Northern Blot: Goal is to identify which tissue expresses YFG
    - Step1) Isolate RNA from a group of tissues
        - Tissue-> lyse cells and purify mRNA
    - Step2) Separate the RNA on the basis of size by agarose gel electrophoresis
    - Step3) Transfer the RNA to a nitrocellulose membrane (blotting)
    - Step4) Label the Nucleic acid probe
        - Radioactively label DNA that encodes a portion of YFG
    - Step5) Incubate the HOT PROBE with the filter to allow complementary strands to anneal
    - Step6) wash away the non-specifically bond probe
    - Step7) Expose the nitrocellulose filter to an X-ray film to determine where the “hot” probe has annealed
        - The radioactive probe will appear as bands in the lanes in which YFG is expressed
- Southern blot: hybridize DNA to DNA
    - Commonly used to address whether YFG is expressed in different organisms
    - Microarrays: DNA to DNA on glass slides
    - Used for profiling RNA in different cell types

Synthetic Oligonucleotides

- Fragments of single stranded DNA with defined sequence that are made synthetically
- When they are short (20-30 bases) they are referred to as primers or oligos

mRNA can’t be Cloned

- We have to convert it to cDNA
- cDNA is complementary to DNA copy of the mRNA

Reverse Transcriptase

- enzyme that makes DNA using RNA as a template
- Isolated from RNA viruses

Making cDNA

- Use the purified mRNA

- Anneal a primer to the polyA tail, by using a repeating T primer
- Make DNA copy with reverse transcriptase
- Then treat with an Alkali to degrade the RNA, leaving a single strand of cDNA

Polymerase Chain Reaction (PCR)

- Invented by Kerry Mullis
- Became a standard technique in 1990
- Exponential amplification of any DNA from a source in which it is found as little as once
- Reagents Required:
    - 1) template DNA
    - 2) two oligonucleotide primers which flank YFG
    - 3) dNTPs
    - 4) DNA polymerase
- PCR cycle:
    - Double stranded DNA is denatured at 95deg, and the strands separate
    - The temperature is lowered to 50deg and the primers are hybridized to the strands
    - DNA polymerase synthesises the strands at 72C
    - The number of copies produced after n cycles is 2^n

Two Key Technical Advances

- Discovery of thermostable DNA polymerase
    - Taq polymerase
- Thermocyclers that oscillate between the 3 required temperatures (50C, 94C, 70C)

How do we purify the PCR product

- Gel Electrophoresis
    - Separation o DNA is bases on their differential mobility in a matrix with very small pores
- Agarose(gel)
    - Polysaccharide
    - After melting it hardens into a jello like form
    - It has small pores through with DNA can pass
- Electrophoresis
    - The movement of molecules in an electric field
    - Question-when placed in an electric field at pH7 DNA will migrate toward which pole?
        - Positive
- Gel Electrophoresis
    - DNA is sieved through a matrix of agarose being pulled by an electric field
    - The agarose gel is in a chamber, submerged in buffer
    - An electric field is the applied
    - DNA fragments are pulled through the agarose matrix and move toward the positive pole
    - Question-which migrates faster in an agarose gel?
        - Small DNA fragments
        - Mobility of a DNA fragment is inversely related to the log of its length
- Detection of the DNA in the gel
    - 1) When stained with ethidium bromide DNA fluoresces red under UV light
    - 2) Autoradiography:
        - radioactively label the 5’ end of your DNA fragments using polynucleotide kinase and [32P}-ATP
        - expose your gel to an Xray film

To get YFG into Ecoli, insert it into a plasmid

- Plasmids:
    - Foreign DNA introduced into E.coli with plasmids

    - A plasmid is a fragment of DNA that replicates independently from the host chromosome
    - Can be from 1-100 copies per cell
    - ~3000bp
    - ciricular
- Key Features required to make a plasmid useful for cloning:
    - 1) Origin of replication
    - 2) A selectable marker: usually a gene encoding resistance to an antibiotic
    - 3) A site(s) into which YFG can be inserted

Restriction Enzymes

- DNA is a long, relatively homogenous polymer
- Restriction enzymes cut into DNA into defined workable units
- Site specific DNA binding proteins that recognize and cleave palindromic DNA sequences usually 4, 6 or 8 base pairs
- Palindromic DNA sequence:
    - Sequences that have 2-fold symmetry
    - Question-which of the following DNA sequences is not a palindrome?
        - TAAAAT -> compliment to ATTTTA

Enzymes

- EcoRI:
    - Leaves sticky ends with 5' overhangs
    - G/AATTC
- KpnI:
    - Cleaves to leva sticky ends with 3' overhangs
    - GGTAC/C
- SspI:
    - Cleaves to blunt ends
    - AAT/ATT
- Some restriction enzymes cut identical sequences, but leave different overhang ends
- Some restriction enzymes cut different recognition sequences but leave identical over hang ends
- Question-What is the probability that any random 4bp sequence is a Hae111 site (GGCC)?
    - 1 in 256, 4different base pairs, and you need ^4
    - or (1/4)(1/4)(1/4)(1/4)= (1/256)
- Question-What is the probability of any bp sequence being Hinc11 is ~1/1000, how many Hinc11 site are found in the human genome (3billion bp)
    - 3 million, 3X10^9/ 10^3 =3X10^6

Restriction Enzymes are Enzymes

- like any other enzyme EACH restriction enzymes has preferred conditions in which it functions
- Temperature, pH
- They also have reaction kinetics (takes time)

DNA Ligase

- DNA ligase=glue for DNA
- Will reseal compatible sticky ends and much less efficiently than blunt ends
- Requires an energy source (ATP)
- Sticky ends, line up side to side and group together
- Ligase sees this, and comes in a seals these ends
- And reform the phosophodiester back bone
- Ligase seals the nicks
- And give intact DNA
- Ligase requires phosphates on the 5' end

Steps in Cloning

- Digest ~100ng of YFG and a plasmid vector with the same restriction enzyme or one that gives compatible sticky ends
- Purify the DNA, often by Electrophoresis
- Incubate vector(plasmid) + insert (YFG) in the presence of DNA ligase and ATP
    - The sticky ends anneal (hybridize)
    - Ligase seals the ends with covalent phospodiester bond

Transformation

- The process by which cells take up DNA from their environment
- Natural property of some bacteria
- E.coli must be treated with chemicals to do it
- Mix, the Ecoli with ligated DNA (heatshock them)
- Some of the Ecoli take up the DNA
- Plate the cells onto agar-plates containing antibiotic
- Cells with the plasmid will divide forming visible colonies
- E.coli colonies containing YFG
- Question-Not all the bacteria on the plate contain a plasmid with YFG inserted what is the reason for this?
    - Contaminating ampicillin resistant bacteria may appear
    - Contaminating DNA may be ligated into the vector
    - The plasmid is able to reseal with itself than the vector
    - Cant tell recirculating colonies that have resistance from the colonies with the insert

Checking that the clones contain YFG

- Amplifying bacteria in individual colonies
- Isolate the plasmid DNA
- Verify which ‘clones’ contain YFG
    - Restriction map
    - Hybridization
    - PCR
    - sequence

Restriction Mapping

- A DNA molecule can be defined by the positions of its restriction enzyme cut sites

Gibson Assembly

- Allows construction of recombinant molecules without using restriction enzymes
- Based on annealing of complementary ends
- Very important in synthetic biology

DNA Sequencing

- To be sure that YFG is correct, it should be sequenced
- Two methods were devolved in early 1970s to sequence
    - Chemical sequencing (Maxam and Gilbert)
    - Dideoxy or chain termination sequencing (Sanger)

Chain termination sequencing:

- Involves the enzymatic synthesis of a DNA strand in the presence of base specific chain terminators
- 2’,3’ Dideoxynucleotides (missing the hydroxyl group on the 3rd Carbon)
- If you remove the 3’ OH, it gets inserted into the chain, but you can-t add on past that point
- act as chain terminators

Materials required to sequence a DNA

- Dna to be sequenced (template)
- Oligonucleotide primer that anneals to the template
- DNA polymerase

- dNTPs (G, A, T, C)
- ddNTPs (small amount, 1/500 of dNTP)

Extension Rx In the presence of ddATP

- terminated that strand
- At each vase where a dA should be inserted into the growing chain, there is a 1/500 chance a ddA will be inserted and the chain terminated
- Do 4 separate reactions, each with different ddNTP
- Denature DNA molecules
- Analyse by gel electrophoresis
- Read the DNA sequence as a ladder from 5'-3'
- Autoradiography of sequencing experiment

Question-Which of the following need to be considered when expressing your favorite human protein in E.coli?

- Human and Ecoli promoter structures are different
- Introns are found in human genes
- They differ in the mechanism of translational initiation
- Different post translational modification (different in prokaryotes and eukaryotes)

Producing YFP

- Purify protein from Ecoli
- Check protein for purity and activity
- Now ready to use

Site directed Mutagenesis

- Creating a gene with altered sequence and in turn altered function
- Converting a codon for Asp to a codon for Ala
- Process:
    - Make synthetic oligonucleotide primer with the mutation
    - Primer anneals to strand with the mutation (single stranded)
    - Add DNAP and add ligase to seal the nick
    - Mismatch base at this position
    - Transfer into ecoli
    - Some Ecoli will have wild type and some Ecoli will have the mutant gene

Transgenic Organisms (GMO)

- An organism that has its genome permanently altered by genetic engineering
- You create a GMO by integrating DNA into the organism's genome through the process of **homologous recombination**
- There types of genetic change are possible
    - Gene knockouts-remove the gene entirely
        - No active gene present
    - Gene replacement-site directed change into the gene and then into the genome
        - Only mutant gene is active
    - Gene addition
        - Both genes are active

Uses of PCR

- 1) Rapid isolation of YFG
    - Use genomic DNA as source
    - Use PCR to get a bunch of clonal copies
- 2) Analysis of bacteria or viruses in clinical environmental samples
    - isolate blood, remove cells
    - purify the RNA or DNA
    - use PCR to test with a non infected person
    - PCR product, doesn't have it, Product not formed they have it
- 3)Diagnosis, DNA fingerprinting
    - Short tandem repeats (STRs)

        - In many regions of our genome there are repetitive sequences
        - Ex)CACACACACA X40
        - If a PRC reaction is preformd with genomic DNA, and a pair of primers flanking on STR locus, you get PCR products of the maternal gene and the paternal gene (genome is diploid)
    - Question- Is there a chance that any two individuals will have the same two bands when any one STR locus is examined by PCR?
        - True
        - Two indidduals might have the same two bands simply by chance
        - For this reason, multiple STR loci are analyzed in a DNA fingerprint
        - Each STR locus requires a distinct pair of primers

CSI London:

- You are a me member of a CSI unit investigating a robbery, you collect DNA from the crime scene, is it the culprits?
- Points to consider:
    - Have to be very clean, may contaminate the sample
    - There is a chance that two people can have the same pattern (1 and 5billion)
    - Based on probability
- In the lab you must isolate DNA from the sample and the suspects sample
- You PCR with the 3 primer pairs, using each of the suspects DNAs and forensic DNA sample as templates

Sequencing Genomes

- Traditional genome sequencing involves:
    - 1) Creating a Genomic DNA library
    - 2) Many independent sequencing reactions
    - 3) aligning all the independent sequencing into a continuous sequence
- Genomic DNA library
    - A collection of cloned DNA fragments that represents all of the DNA in a organism genome
- Constructing a Genomic Library:
    - Take human DNA, you cleave it with a restriction enzyme and you get millions of fragments
    - Clone all these fragments, put them into plasmids
    - Introduce of the plasmids into the bacteria, you get millions of clones

H Influenza was the first genome sequenced:

- Gram negative
- Circular genome
- 2 million bp in size
- Question- The H.flu genome is 2 million base pairs. How many independent clones must be in the library to cover the genome at least once assuming an average insert of 2000bp:
    - 1000
    - 2mil bp / 2000bp/clone =1000
    - that's the absolute minimum
    - have many fold excess to insure full coverage of the genome
- Creating the Genomic Library:
    - Start with millions of cells
    - Extracted DNA, sonicate randomly (use blunt ends for cloning)
    - Purify the DNA fragments of ~2000bp, size selected on agarose gel
    - Transfer this into Ecoli, and prepare a clone library
    - 20,000 clones (20fold excess) and each clone represents a fragment of the genome
    - isolate the plasmids, add a primer that anneals to the plasmid
    - they sequence the insert, (can only read 500bp through the 2000bp insert)

- 25000 sequence runs from 20000 resulting in 12 mil bp of sequence (6fold excess)
- some were sequenced from one side, a few from both sides
- They then put all these sequences together in the correct order, (ordering of pieces)
  - Sequence CONTIG:
    a continuous DNA sequence that represents a portion of the genome
  - Contings are not physical entities. They are sequences (computer makes by aligning sequences)
- Computer search's for overlaps over all 25000 sequences
- Overlaps are arranged into contigs, individual sequence run (not a cloned fragment)
- 20,000 clones (25000 sequences)->computer aligned them into 140 contigs
  - lower the number the better
- Question-If you have identified 140 contigs and the organism's genome is circular how many gaps do you have to fill
  - 140
  - with 140 contigs you will have 140 gaps in a circular genome
  - you need to fil the gaps to complete the genome

How do you fill the Gaps:

- two types:
  - 1) Sequence Gaps
  - 2) Physical Gaps
- sequence gaps, can be closed by completing the sequence of the clones in the library (sequencing on both sides)
- The computer scans for clones that are found in 2 different contigs
- Cloned DNA fragment that is found on two different contigs
- Physical gaps, are not represented by clones in the library
- They occur because some DNA sequences are not easily cloned, (probably because they produce something toxic in the ecoli)
- Physical gaps can be filled using PCR, with genomic DNA as the template (don't use the library but use the genomic DNA)
- Make 2 primers for each end contig. These are to prime in the direction "out of "the contig (primers point outwards towards the gap)
- Get a product when you use PCR between gaps (cant have massive products over a lot of kp)
- Have to have primers on opposite strands? Primers positioned side by side on contigs will give you a product
- Closing a physical gap, you do PCR with with all the primer (1+3, 1+4, 1+5… so on) and genomic DNA as the template
- Question-If 2 contigs are adjacent they will yield a PCR product?
  - To get the sequence of the gap, you cut the DNA out from the gel and sequence the gap

Annotating the Sequenced Genome

- Expressed Regions:
  - Protein coding genes
  - tRNAS
  - rRNAs
  - other functional RNAS
  - small regulatory RNA
- Not expressed regions:
  - Regulatory sequences: promoters, terminators
  - Positioning elements
  - Nucleosome positioning elements origins of replication

- Telomeres

Open Reading Frames (ORF)

- A series of codons starting with an initiation codon and ending with a termination codon
- The First step in annotating a genome is for a computer program to scan the genomic sequence, searching for ORFs
- Question- In any given (double stranded) DNA sequence, how many fames must the computer scan through?
  - 6
  - because ATG/TAG or AT/GTA/G or A/TGTAG
  - any stretch of DNA has 6 potential open reading frames
  - 3 on each strand
- Having identified all of the ORFs, what criteria will we use to decided if an ORF actually encodes a protein?

ORFs that encode Proteins

- 1)most contain 100 codons or greater
- 2)Show a codon usage typical for the organism (codon bias)
  - different organisms can have different relative usages of codons
- Codon Bias:
  - Certain codons are more frequently used to encode each amino acid
  - The codon bias is organism specific
  - An ORF that displays the codon bias typical of the organism is more likely to be a protein

Functional ORFs

- 3)Related sequences, encoding similar proteins are usually found in other species
  - BLAST searches can reveal related sequences
- 4) Are found expressed as mRNA
  - northern blot
  - microarray
  - RNAseq
- 5) Contain appropriate regulatory sequences
  - -10 and -35 for bacteria
  - TATA for eukaryotes
- 6) Contain chromatin signatures of expressed genes
  - nucleosome acetylation and methylation patterns
- 7)Are found as proteins
  - western blot
  - mass spectrometry

Topic 24: Molecular Biology of Cancer

Molecular Basis of Cancer

- The leading cause of death in North America
- Strikes victims of all ages
- Is becoming more prevalent as the population ages
- Cancer is characterized by genetic and biochemical defects
- Biochemistry and molecular biology provide avenues for cures

Tumour Types

- Adenoma (Benign)
- Adenocarcinoma (malignant)

Metastasis

- Malignant tumours can metastasize, that is spread from their primary site to one or mores secondary sites
- Different cancers tend to spread to different sites

- Common sites of spread are to the bones, liver and lungs

How do we know that Cancer is a Genetic Disease?

- Cancer can be inherited
- Retinoblastoma (cancer in th eye)
- Xeroderma pigmentosa (a cancer of the skin)
- Some forms of breast prostate

DNA damaging agents cause cancer

- Radiation: UV, xrays
- Chemical Mutages

What are some of the Avoiadble Causes of Cancer?

- Agents that cause DNA damage
- Chemicals that cause cells to divide
    - Tumor promoters
- Certain viruses
    - Hepatitis
    - HPV
- Question- What is the single most preventable cause of cancer?
    - Smoking

A single mutation is not sufficient to cause cancer

- Tumour progression involves successive rounds of mutation and selection
- At each round the descendent acquires another mutation allowing it to grow faster or in abnormal places
- Most cancers derive from a single cell
- First mutation may allow the cell to grow more quickly
- Second mutation may allow the cells to grow in the absence of the basil lamina
- A third mutation may allow the cells to penetrate the basil lamina

Properties of Cancer Cells

- 1) Divide in the absence of growth factors
- 2) Are immortal
    - Do not respond to signal that normally trigger cell death
- 3) Have lost cell cycle control
    - Checkpoints in the Cell cycle:
    - G2: check for cell size and DNA replication
    - Spindle Assembly Checkpoint: check for chromosome attachment to spindle
    - G1: check for cell size, nutrients, growth factors, DNA damage
- 4) Are genetically unstable
    - more point mutations
    - copy number variation
    - major chromosome abnormalities
- 5) Cancer cells can multiply in abnormal places (metastasis)

Types of Cancer Causing Genes

- 1) Oncogenes
    - a mutant form of a normal genes whose presence causes cancer
    - Dominant-gain of function
    - Overactivity Mutation (gain of function)
        - single mutation event
        - creates oncogene
        - activating mutation enables oncogene to stimulate cell proliferation
    - Mutation in protein encoding region
        - Hyperactive protein made in normal amounts
    - Gene Amplification: normal protein greatly overproduced
    - Chromosomal rearrangement: nearby regulatory DNA sequence causes normal protein to be over produced
        - Fusion to actively transcribed gene produces hyperactive fusion protein

- 2) Tumour suppressors
    - A gene whose absence causes cancer (recessive, loss of function
    - Underactivity mutation (loss of function)
        - Mutation event inactivates tumor suppressor gene
        - Second mutation event inactivates second gene copy
        - Two inactivating mutations functionally eliminate the tumor suppressor gene, stimulating cell proliferation

Functions of Cancer Causing Genes

- Most oncogenes and tumor suppressors code for protein that act in or regulate cell division or cell differentiation
- Oncogenes are the accelerator pedal
- Tumor suppressors are the brakes

Function of Cancer causing genes

- 1)growth factors and cellular receptors from groth factors that are stuck in the on position
    - EGFR (epidermal growth factor receptor
    - Drives cell division
    - Constantly divide
- 2)Molecules involved in cell-cell interactions
    - the cell-cell interactions that control growth and differentiation are lost
    - normal cell wont divide because it has neighbour
    - cancer cell doesn't recognize that it has neighbouring cells and divides anyways
    - don't care if there is basil lamina there or not
- 3)Regulators of normal/programmed cell death (apoptosis)-p53
    - cells do not respond to the normal signals
    - cancer cells express telomerase
- 4) Transcription factors (myc, jun, fos)
    - cancer can be caused by too much or too little expression of the genes that regulate cell growth, differentiation or cell death
- 5) DNA repair proteins—(Xeroderma pigmentosa XPC)
    - cells accumulate more and more DNA damage and thus mutations to key genes

Classic Cancer Treatments

- surgery
- radiation
    - stop cells from replicating by damaging DNA
- chemotherapy
    - stop cells from replicating by damaging DNA or interfering with the mitotic machinery or reducing replication substrates
    - not specific enough
- specificity if the treatment
    - radiation and classic chemotherapy
- different cancers have different underlying causes
    - different genes involved
    - different originating tissues
    - different locations
    - not one cure all
- Heterogeneity of the tumor
    - Not all the cells within the tumor are identical
    - The tumor cells are always "evolving"
- Cancer cells develop resistance to drugs
    - Drugs target acquires mutations
    - Cells start to overexpress transporters that pump out the drugs
- Inaccessibility of the tumor to drug

    - Blood brain barrier

Goals of Modern cancer treatment

- Earlier detection
- Greater drug specificity (antibodies*)
- Combination therapies
- Identify drugs for the “undrugable” targets
- Prevention

Gleevec as a treatment from CML

- CML
    - Accounts for 20% of all leukemia’s
    - Results from abnormal proliferation of hemotopc stemcells
    - 90% of adults with CML have the Ph chromosome (Ph+)
    - fusion of two chromosomes (9 and 22) Bcr with Abl
- Abl is a protein kinase that phosphorylates protein requires in cell differentiation cell division and cell adhesion
- Abl is wild type
- Changes substrate specificity
- BrcAbl phosphorylates targets (proteins it should phosphorylate) that end up activating cell division and enabling apoptosis
    - Cells keep dividing
- Gleevac fits into ATP pocket of BrcAbl and inhibits it ability to phosphorylate
- Abl with Gleevac, knowing the structure of Abl allowed people to design the drug to do this