BCMB30010 – Advanced Techniques in Molecular Science

Lecture 1 – DNA Technology

Overview – Recombinant DNA Technology:

- Recombinant DNA involves the ligation of a vector with a DNA fragment
    - DNA fragment can be genomic DNA or cDNA
- This recombinant DNA is inserted into a host
- It propagates/replicates in the host as the host replicates

cDNA libraries:

A cDNA library is a collection of all genes expressed (mRNA) in a cell/tissue/organism at a given time (the time of harvesting) cloned into vectors

- cDNA = reverse transcribed mRNA
    - This is more stable than mRNA
    - mRNA is unstable, susceptible to degradation by RNases and difficult to store
    - Has been spliced, capped and tailed

Making a cDNA library:

1. Harvest mRNA from your chosen source using oligo-dT primers linked to agarose (wash well) and elute with low salt buffer.
    - Oligo-dT primers target the poly-A tail in order to enrich mRNA exclusively
2. Use oligo-dT primers to produce one DNA strand using reverse transcriptase
    - Results in tightly bound RNA:DNA hybrid
3. Use RNase H enzyme to remove the RNA from the RNA:DNA hybrid
4. Use terminal transferase to add nucleotides to the 3’ end
    - Use only G nucleotides so that we have a 3’GGG end and a 5’TTT end
5. Use oligo-dC primer to synthesise the complementary strand using DNA polymerase
    - Yields one double-stranded (ds) DNA for each mRNA strand
6. Use EcoRI methylase to methylate every EcoRI site (to protect these sites)
7. Add an EcoRI linker (a dsDNA sequence that contains an EcoRI site) using DNA ligase
8. Cleave with EcoRI and ligate into a vector using ligase
    - This will only cut the linker, as other sites are protected
    - Phages are usually a good vector
9. Introduce vector into the host
    - If using phage, infect bacteria with virus to get a good transfection rate

Enzymes:

Terminal transferase:

- Transfers nucleotides to the 3' end of single-stranded (ss) or dsDNA
- Used to be used for cloning (added C's to the cDNA insert and G's to the vector)

Reverse transcriptase:

- An RNA-dependant (RNA is the template) DNA polymerase
    - However, the lab version can work on either RNA or DNA
- From RNA/retroviruses

RNase H:

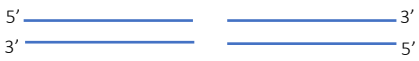
- Cuts phosphodiester bond of RNA, but ONLY in RNA:DNA hybrids
- Exonuclease (cuts ends) and endonuclease (cuts internally) activity
- Sometimes activity is with reverse transcriptase (not ideal)
- Can be used to remove poly-A tail
    - Hybridise poly-A tail with an oligo-dT, then RNase H will only cut the hybrid

EcoRI methylase:

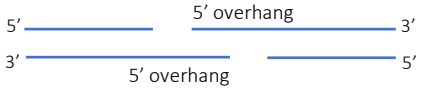
- Specifically methylates A in the EcoRI site 'GAATTC'
- Results in EcoRI being unable to cut at this methylated site

Restriction endonucleases/enzymes:

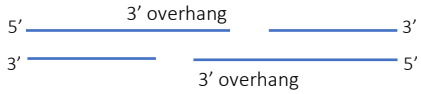
- 'molecular scissors' that cut DNA at specific sites
    - Cuts the phosphodiester bonds within DNA
- Bacterial restriction enzymes
    - Used as a defence mechanism against foreign (unmethylated) DNA
- Have many uses in DNA technology
- Type II (most commonly used)
    - Binds DNA as homodimers and recognise palindromes
    - Require Mg^{2+}
- Type I and III
    - Both larger and have modifying activities
    - Less useful in research
- The enzyme's recognition loops generally bind the major groove of DNA
- Different restriction enzymes recognise different sequences
- Different restriction enzymes generate different types of ends
    - Blunt ends



- Cohesive/Sticky ends
  - 5’ overhang



  - 3’ overhang



- Must be used under standard conditions, otherwise you risk:
  - Star activity – loss of specificity (more cuts than expected)
  - Partial activity – does NOT cut at all sites (less cuts than expected)

Lecture 2 – DNA Technology (continued)

Restriction endonucleases/enzymes (continued)

Reaction requirements (prevents partial or star activity):

- Buffer
- DNA
- Restriction enzyme
- Water

Total volume for a single digest = 20 μL

Incubate for 1-2 hrs at the optimum temperature for the restriction enzyme

Using two restriction enzymes:

- A more difficult process but is required to perform directional cloning

There are two options for how to perform a reaction with two restriction enzymes:

- DOUBLE DIGEST – if the two enzymes have the same requirements, the reaction can be performed simultaneously
    - NOTE: if they require different temperatures, perform at lower temperature
- SEQUENTIAL DIGEST – if the two enzymes have different requirements, perform one reaction and then the other
    - Must evaluate which to perform first

SEQUENTIAL DIGEST:

Guidelines:

- Enzyme volume should never be more than 10% of the total volume (e.g. not more than 2 μL in 20 μL)
- Perform first digest in 20 μL and second in 40 μL
- Perform each digest for 1 hour at 37 °C
- Heat inactivate between digests for 20 minutes at 65 °C
- mM = nmole/μL

Estimations/assumptions

- 6mM Tris – can use up to 10 mM, require a minimum of 5 mM
- pH – must be within 0.5 pH units of the specified value
- Assume enzymes are in 50% (v/v) glycerol and include reducing agents
- If the enzyme requires KCl, it won't work with NaCl
- If the enzyme requires NaCl it will work with a mixture of NaCl and KCl
- The final concentration of salts must be exact

DNA Ligase

- Binds to dsDNA and joins one strand at a time
- Physiological function in DNA repair, replication and recombination
- Can ligate blunt or cohesive ends
    - However easier to ligate cohesive ends rather than blunt ends
    - For cohesive ends there is a transient interaction and only joins compatible ends – must be either both 3' or both 5' overhang fragments
    - Blunt ends can ligate to anything

Mechanism:

- ATP-dependent in eukaryotes and T4 bacteriophage
- NADH-dependent in prokaryotes
- Lysine is critical at the active site
- Optimum temperature = 25 °C
- Involves an AMP-intermediate and results in a covalent bond

Cloning:

- Amount of insert DNA must in excess of the vector
- Cohesive ends: Km = 0.6 μM
    - Incubate for approximately 30 minutes – 3 hours
    - Use a temperature less than the melting temperature
- Blunt ends: Km = 50 μM
    - Incubate for approximately 18 hours (longer than for cohesive)
    - Require more enzyme and salt or monovalent ions

DNA Polymerase

- Template dependent 5' to 3' polymerase
- May have exonuclease activity (depends on which one)
- Requires Mg^{2+} and a primer
- Highly conserved in structure and function

DNA polymerase I

- Three different domains:
    - Hydrophobic centre for synthesis
    - Proofreading function – can remove mismatches through 3' to 5' exonuclease
    - 5' to 3' exonuclease activity – removes nucleotides ahead of synthesis
- Synthesis and proofreading work together – polymerase slides back to correct mistake
- 5' to 3' exonuclease is bad for experiments as it can't be controlled

Mechanism

1. Nucleotide base pairs and enzyme binds to dsDNA
2. Covalent reaction between 5' phosphate of incoming nucleotide and 3' OH of DNA
3. Release of pyrophosphate

Klenow fragment

= DNA polymerase I without its 5' to 3' exonuclease domain

- This removes the uncontrollable 5' to 3' exonuclease activity
- Can convert cohesive ends to blunt ends by
    - Filling in the gap for 5' overhang (requires dNTPs)
    - Digesting/cutting back the 3' overhang (does NOT require dNTPs)

Exo-Klenow fragment = DNA polymerase I with ONLY its polymerase activity (no exonuclease)

Polymerase Chain Reaction (PCR)

Standard PCR cycle

1. Denaturation of DNA (90-95°C for 30-60 seconds)
2. Annealing of primers (T_{ann} ~50-55°C for 60 seconds)
3. Extension from primers (68-72°C for 60-120 seconds)

- Repeat steps 1-3 several times
- NOTE: after a few cycles, the product is in excess of the original template

DNA polymerases:

- requires a thermostable DNA polymerase

Polymerase	Proofreading	Error rate	Processivity	Extension rate
Taq	No proofreading	10^{-4} (bad)	5-20	3600
Pfu	3' to 5' proofreading	10^{-7}	5-20	600
KOD	3' to 5' proofreading	10^{-6}	300 (up to 10kB)	6000

*Processivity = how long polymerase stays on DNA, higher processivity allows longer product

- Taq is cheap, thus preferred for short amplification
- NOTE: must check resultant sequence if using PCR products for cloning
- NOTE: "KOD hot start" has mutations to improve processivity to amplify ~21 kB
- Can mix polymerases: e.g. Taq:Pfu at a 15:1 ratio allows proofreading
- Can improve processivity by:
    - Adding accessory proteins to promote DNA polymerase binding to DNA
    - Add a non-specific dsDNA binding domain (DNA pol remains bound)

Lectures 3 and 4 – DNA Technology and Sequencing

Optimising PCR

- Test a range of Mg^{2+} concentrations
- Test a range of T_{ann} temperatures

Primer design for PCR

Melting temperature, $T_{melting}$ (T_m):

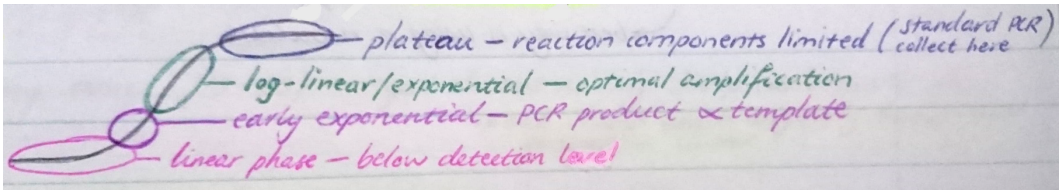
- Should be more than 50 °C
- Want annealing temperature to be similar for both primers
- Should be equal to 50% of annealing of a strand

Sequence

- Approximately 18-30 bases, best is 20-25 bases
- Optimise specificity
- Avoid runs of internal G's and C's as they can form hairpins
- Avoid primers self-annealing
    - Very problematic as PCR favours short amplification
    - To avoid this, make sure that the 3' ends of primers are NOT complementary, otherwise they form 'primer dimers'
- Aim for 50% G/C content
    - Consider including GC-rich 'clamps' at the 3' ends (improves amplification)
- May include a restriction site at 5' end (helps with next steps)
- Can be checked/aided by web-based software

Quantitative PCR (qPCR)

- PCR products are detected by fluorescent dyes which only bind dsDNA
- PHASES:



- Allows determination of how much template was in the original sample
    - mRNA is proportional to cDNA
- Quantitate during early exponential phase
    - Ct (cycle threshold) value: cycle at which the sample crosses the threshold
    - Lower Ct value indicates more template and thus higher expression
    - Plot Ct against $\log_{10}$(starting copy number) to form a standard curve
        - Gradient of this linear line = efficiency (want >90)

SEQUENCING

Knowledge of sequences provides:

- Ultimate characterisation of DNA
- All information encoded in an organism
- A platform with which to conduct evolutionary studies

Sanger Sequencing (dideoxy sequencing)

- Uses dideoxy nucleotides (ddNTPs) at LOW concentration
    - These nucleotides have no 3'OH, thus extension is terminated at this base
    - Randomly incorporated into a strand
- Obtain a family of products (primer, primer +1, primer +2, etc.)
- Order this family of products by size using capillary gel electrophoresis
- Each ddNTP has a different fluorescent dye attached that a laser detects (e.g. C=blue)
- Uses a single primer (start extension from one end only
    - Must know some of the sequence to design this primer
- Can be automated, now read length ≈800, with ~40 runs per day

Next-generation Sequencing

- Greatly reducing cost of sequencing (dropped away from Moore's law)
- Needed to re-sequence genomes as well as to perform de novo (new) sequencing
- Includes Illumina/Solexa, Ion torrents, 454/Roche, and others
- Sequence is obtained as the polymerase is adding the nucleotides

ILLUMINA SEQUENCING

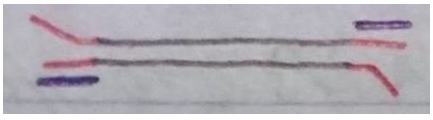
- About $^{2}/_{3}$ of all sequencing data is obtained via this technique
- Obtain approximately 30 reads of each nucleotide (allows double-checking)

Preparation of DNA:

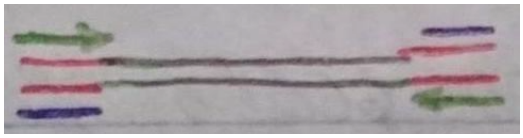
- Gather genomic DNA and break DNA into short sequences/fragments (~300-400 bp)
    - Usually use a nebuliser to do this
- Polish fragments with Klenow and dNTPs
- Add an A nucleotide to the 3' end of DNA using Taq polymerase
- Ligate each end of DNA to Y-adaptors (match halfway)

Process:

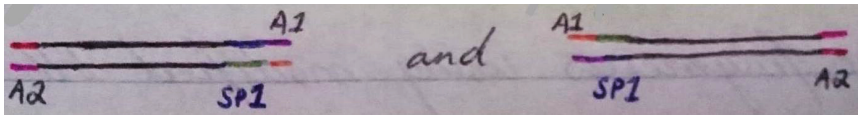
1. Add short and long primers, but only short primers can anneal
    - Long cannot anneal due to the Y-adaptor



2. After one cycle, both the long and short primers anneal



KEY: obtain identical DNA fragments with different ends:



3. Check DNA to ensure the template is the correct size and has the correct ends
4. Denature DNA and add it to a flow cell covered with primers SP1A1 and A2
    - Primers are covalently linked to the flow cell
    - DNA binds to partner adaptor/primer
5. Synthesise a new complementary strand
    - The new strand is covalently linked through primer/adaptor to the flow cell
6. Denature dsDNA and wash out the free template
7. Free ends anneal to complementary primers on the flow cell and synthesise strands
    - Termed 'bridging PCR'
8. Repeat many times (10-20 times) to obtain clusters of identical DNA sequence
    - Fragment is ~300bp long and so cannot reach far, thus you get a cluster
9. Wash away strands bound to A1 primer
    - A1 primer has a dU in its sequence
        i. Behaves as a dT
        ii. Sensitive to cutting by 'USER' enzyme (mixture of two enzymes)
    - Now we have all DNA single stranded and in the same orientation
10. Add SP1 primer at T_{ann} and **READ THE SEQUENCE** using reversible terminators
    - Reversible terminators are nucleotides that block the 3' OH allowing only one base addition
    - Each dNTP has a different dye attached
    - Uses a mutated DNA polymerase that can accommodate modified bases
11. After each addition, remove fluorescent dye using TCEP reducing agent
    - This allows the next nucleotide to be added
    - TCEP removes fluorophore and azidomethyl blocking group, leaving a propargyl group

Read Pair Mates

= two sequence reads that are separated by a defined distance

- Can greatly reduce the size of reads needed to assign sequences to genome
    - Normally would need 100 bp to place accurately in non-repetitive region
    - Using read pair mates, need only 50 bp
- Really useful for determining repeat regions
- Really useful for creating new and/or reference sequences/genomes

Using Sanger sequencing:

- Make a library with 2 kB inserts
- Sequence a section from each end = read pair mate
- Can do this with different size inserts

Using Next-generation sequencing:

- Fractionate genomic DNA into 2 kB sequences
- Add Klenow and biotin-dNTPs to fill the ends
- Ligate two ends together (2 kB circles with biotin at the join)
- Fragment circles into 400-600 bp fragments
- Use streptavidin to isolate the piece labelled with biotin (has two ends)
- Sequence the fragments using a modified Illumina protocol
    - A1 has dU as normal, A2 has an 8-oxo-G (cut by Fpg)
    - Generate clusters, add USER, sequence one end using SP1
    - Remove strand generate clusters (~2 cycles), add Fpg, then sequence with SP2
    - The position on the flow cell will be the same for the first and second sequences, allowing you to get a read pair mate

Bioinformatics

= application of information technology to allow storage, analysis and management of biological information

Databases

- Store information with specific ways to add and retrieve data
- Includes primary data, annotations, research citations/links, etc.

Sequence alignment:

- Can compare data using sequence alignment
- Provides scores (based on substitution matrix) to quantitate alignments
    - Matrices differ between databases, they have different scores for matches/substitutions, some have gap penalties etc.

- Different alignment types
    - Global alignment – best alignment over entire length
    - Local alignment – align parts of a sequence
    - Multiple sequence alignment – compares 3 or more sequences at the same time to provide a consensus sequence

BLAST (Basic Local Alignment Search Tool)

- Approximates the results of the Smith-Waterman algorithm (the best one)
- Computationally breaks sequence into bits to compare, then builds up from there to provide a final comparison
- Scores and e-values are provided, taking into account:
    - Query sequence length
    - Database size
    - Scoring system/substitution matrix