School of Life Sciences
Bachelor Life Sciences Engineering

Laboratoire intégré en sciences de la vie II

Course Manual

BIO-204 Spring 2025

Alexandra Bezler

Laboratory 6:
DNA Sequencing

Objective

- To confirm the presence of the *Amy2* cDNA by Sanger sequencing
- To analyse the cloned sequence (detect mutations)

1. Sequencing Reaction

Sanger sequencing, also referred to as chain termination method, provides information about the nucleotide sequence in a DNA fragment. It is a variation of PCR with a **single primer** using **dideoxynucleotides** (ddNTPs) in the presence of regular dNTPs. The aim is not to produce identical copies of DNA but many copies of variable length. ddNTPS lack the 3' OH group (required for chain growth by linking the 5' phosphate to the 3' OH acceptor): each time a ddNTP is incorporated in the PCR product, the chain stops growing. Since the ddNTPs are present in lower amounts (10-300 less than dNTPs) they only occasionally create a stop of the growing chain. As a result a mixture of DNA strands of each possible length is generated. The four ddNTPs (ddATP, ddGTP, ddCTP, ddTTP) are tagged with different fluorescent dyes, thus the fragments differ in color based on the kind of ddNTP present at the 3' end. After separating the pool of DNA fragments by size by capillary gel electrophoresis (resolved to single nucleotide size difference) the fragments are detected using a laser. The chromatogram shows peaks in four colors, corresponding to each ddNTP generating a read length of approximately 800-1100 bases. The automated sequencing is done by a company (in our case Microsynth). The reaction starts by heating the samples to 95°C (to separate the two DNA strands), followed by a cool down, which allows short oligonucleotide primers (20-30 bp) to bind. DNA polymerase then elongates the primers and synthesizes DNA using regular deoxynucleotides (dNTPs).

Sequencing primers must be designed at a minimum distance of 50 bp from the target sequence. In order to check that pAmy2-His contains the correct coding sequence without any mutations, the insert was sequenced in both directions, using CMV-for or BGH-rev sequencing primers (many standard primers are available from sequencing companies) complementary to sequences upstream and downstream of the multiple cloning site of pcDNA6/*myc*-His.

Sequencing primers used:

- CMV-for: CGCAAATGGGCGGTAGGCGTG (complementary to CMV promoter)
- BGH-rev: TAGAAGGCACAGTCGAGG (complementary to bovine growth hormone polyadenylation signal)

The sequence files on Moodle:

- Sequence1.fasta/.ab1 (reaction with CMV-for sequencing primer)
- Sequence2.fasta/.ab1 (reaction with BGH-rev sequencing primer)

Exercise 1

For sequencing of the recombinant plasmid CMV-for and BGH-rev were used.

a) In Benchling create and attach the standard sequencing primers to your recombinant plasmid pAmy2CDS-myc-HIS cloned last semester (reminder: go to Primers>Create Primers>Manual. If the primer is attached to the wrong location (red nucleotides): first Detach Primer, then go to Primers> Attach Existing, it will scan the sequence for complementary regions). Paste a screenshot of the attached primers into SLIMS.

b) Knowing the distance between the two primers, would sequencing with a single primer be sufficient to verify the sequence of the recombinant Amy2-myc-His?

c) For Sanger sequencing you accidentally used the primer Mm-Amy2-for (used for PCR amplification in lab 3). What is the effect on the sequencing result? Is it sufficient to verify the sequence of the recombinant Amy2-myc-His sequence?

Exercise 2

You want to sequence the **ampicillin resistance gene** in the recombinant plasmid.

a) Create primers using the Benchling 'wizard' tool that suggests suitable primers (standard primer parameters). Choose a suitable primer pair for sequencing in both directions. Paste a screenshot of the attached primers into SLIMS and indicate the nucleotide distance between the primer and the target sequence you want to sequence.

b) Knowing the distance between the two primers, would sequencing with a single primer be sufficient to verify the sequence? Justify your answer.

2. Sequence analysis of recombinant plasmid by alignment

Analyze the sequences using Benchling: align **both** sequencing files (.ab1 files on Moodle) to your recombinant plasmid pAmy2CDS-myc-HIS (open the sequence of the recombinant plasmid (=template) go to Alignment>Create New Alignment, upload the **chromatograms** use standard parameters).
Mismatches between the template and the aligned sequences are highlighted. Note that "Ns" represent peaks that are not readable by the basecalling software; in some cases they can still be identified by eye. Stretches of "Ns" are generally found at the beginning and the end of the sequence where the peaks are not well resolved. Check whether there are any mutations. Mutations are nucleotides that are different from the reference sequence. Silent mutations do not cause changes of amino acids. In case the amino acid changes, the function of the protein may be affected (for example mutation in the active site).

Exercise 3

a) To help you analyze Sanger sequencing data, first read: How to Interpret DNA Chromatograms (Moodle). Then check the beginning and end of the **chromatograms** for

- Baseline noise
- Loss of resolution
- Miscalled bases

b) Check **aligned chromatograms** (.ab1 files) for the presence of the following features:

- Cloning sites (e.g. *Bam*HI and *Xho*I sites)
- Kozak sequence
- Start and stop codons (note that *Amy2* stop codon (TAA) was modified to tyrosine (TAT) upon PCR amplification)
- Myc-epitope
- Histidine-tag (as the sequencing primer is close to the tag, the peaks are not well resolved and the automatic base calling fails. Check histidine sequences by eye)
- Presence of mutations
- Paste relevant parts of the sequence alignments (nucleotide and amino acid sequence) into SLIMS and label the data.

Are the sequences correct? Are there any mutations- if yes what is their effect on the protein sequence? Does your plasmid encode the expected fusion protein and regulatory sequences?

3. Sequence identification by BLAST

Another method to analyse sequence data is known as BLAST (Basic Local Alignment Search Tool). It works with both nucleotide and protein sequences. In our case the nucleotide BLAST compares DNA sequences (.fasta files on Moodle) to sequence databases (for the exercises use 'standard databases'). Sequences producing the most statistical significant alignments are displayed. This technique is typically used to identify unknown sequences or infer evolutionary relationships. Since fasta files are text-based they can be opened with any text editor, note the first line starts with > 'greater than symbol' and contains the description.

Exercise 4

Use nucleotide BLAST to analyse the sequence file of the forward sequencing primer CMV-for (Sequence1.fasta on Moodle). Choose default settings, i.e database: Standard> Nucleotide Collection (nr/nt). Look at the results: by default the table 'Descriptions' is sorted by E value, other criteria may be used depending on the context.

a) Name four distinct top ranked species in column 'scientific name'. What does this tell us about the conservation of the sequenced gene in other rodents (see percent identity)?

b) Click on the top ranked alignment. Can you find the start and stop codon? Justify your answer.

Exercise 5

Use nucleotide BLAST to analyse the sequence file of the reverse sequencing primer BGH-rev (Sequence2.fasta on Moodle).

a) Choose again default settings, i.e. database: Standard> Nucleotide Collection (nr/nt). Click on the top ranked alignment. Can you find the start and stop codon? Justify your answer.

b) Now use nucleotide BLAST with modified search criteria. Choose database: Standard> RefSeq Select RNA sequences (refseq_select) in drop-down menu. What is the percent identity (see column 7 labelled *Per. Ident*) between mouse Amy2 and the closest sequence identified

-in rat?

-in human?

Paste the 'Distance Tree of Results' into SLIMS.