**Lecture 37: Vector Construction and DNA Sequencing**

**Step 3 : Insertion of HIV protease Gene into the Expression Vector.**

The final product, after PCR.

5' CG**GAATTC**cctcagatcactctttggcaa....TTAAATTTC**GGATCC**CG3'

3' GC**CTTAAG**ggagtctagtgagaaaccgtt....aatttaaag**CCATGG**GC5'

 **EcoR1 ProGlnIleThrLeuTrpGln----LeuAsnPhe BamH1**

A detailed view of the expression vector is shown below:

-Promoter for---Regulatory element---Ribosome--Start--Leader--**GAATTC**-------**GGATCC**--Stop---mRNA--------

-mRNA Syn.------for mRNA Syn---------binding---Codon--Peptide-**CTTAAG**-------**CCTAGG**--Codon—-Termination-

 site EcoR1 BamH1

After insertion of the gene that encodes the HIV protease, the DNA will look like the following:

-Promoter for---Regulatory element---Ribosome--Start--Leader--**GAATTC**--**HIV**------**GGATCC**--Stop---mRNA---

-mRNA Syn.------for mRNA Syn---------binding---Codon--Peptide-**CTTAAG**--**Protease**-**CCTAGG**--Codon--Term.--

 site EcoR1 BamH1

**Step A**. Cut HIV protease DNA (PCR product) with EcoR1 and BamHI

*C-G****-G-A-A-T-T-C****-HIV--****G-G-A-T-C-C-****C-G C-G-G* ***A-A-T-T-C****-HIV--****G G-A-T-C-C****-C-G*

*G-C****-C-T-T-A-A-G****-Prot-****C-C-T-A-G-G****-G-C G-C-C-T-T-A-A* ***G****-Prot-****C-C-T-A-G G****-G-C*

 EcoR1 BamH1

 *C-G-****G******A-A-T-T-C****-HIV--****G G-A-T-C-C****-C-G*

 *G-C-****C-T-T-A-A******G****-Prot-****C-C-T-A-G G****-G-C*

**Step B**. Cut Expression vector with EcoR1 and BamH1, discarding small segment of DNA between the EcoR1 and BamH1 sites:

 -Start--Leader--**G-A-A-T-T-C**-------**G-G-A-T-C-C**--Stop---mRNA--

 -Codon--Peptide-**C-T-T-A-A-G**-------**C-C-T-A-G-G**--Codon—-Term.-

 EcoR1 BamH1

 -Start--Leader--**G G-A-T-C-C**--Stop---mRNA--

 -Codon--Peptide-**C-T-T-A-A G**--Codon—-Term.-

 EcoR1 BamH1

**Step C.** Mix HIV protease fragment and expression vector:

 ***A-A-T-T-C****--HIV------****G***

 ***G****--Protease-****C-C-T-A-G***

 -Start--Leader--**GG-A-T-C-C**--Stop---mRNA--

 -Codon--Peptide-**C-T-T-A-A G**--Codon—-Term.-

**Step D**. Cool to allow H-bonds to form due to cohesive (sticky) ends.

 -Start--Leader--**G *A-A-T-T-C****--HIV------****G* G-A-T-C-C**--Stop--mRNA--

 -Codon--Peptide-**C-T-T-A-A *G****--Protease-****C-C-T-A-G* G**--Codon—Term.-

**Step E**. Use **DNA ligase**, to rejoin DNA fragments.

--**G-*A-A-T-T-C****--HIV------****G-*G-A-T-C-C**--Stop

--**C-T-T-A-A-*G****--Protease-****C-C-T-A-G-*G**--Codon

 EcoR1 BamH1

**Step F**. Transform mixture of ligated plasmids into bacteria and select for bacteria using the antibiotic resistance gene on the plasmid; bacteria containing the plasmid will produce a protein that will make them resistant to antibiotic present in the culture. Typically one bacteria will be transformed with a single plasmid.

**Step G**. Isolate plasmid DNA from candidate bacteria and digest with restriction endonucleases to **verify** that the plasmid is correct. Verify fragment sizes using agarose gel electrophoresis (DNA fragments separate based on the same principal as protein separation in SDS-PAGE).

**DNA Sequencing:**

*Key Concept*: If we know the distance of each type of base from a common origin, then it is possible to deduce the sequence of the DNA.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Distance from end (bases) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| A |  |  |  |  | X |  | X |  |  |
| G |  |  |  |  |  | X |  |  |  |
| C | X | X |  | X |  |  |  |  | X |
| T |  |  | X |  |  |  |  | X |  |
| Sequence | C | C | T | C | A | G | A | T | C |

Obtaining this information is conceptually quite simple. The idea is to generate a collection of **all** possible DNA chains that:

*1. All begin at the same location,*

*2. End, or terminate, with a known base,*

*3. Measure the position of termination by measuring the length of the fragment using gel electrophoresis.*

**1. Priming**: Beginning at the same location is accomplished by priming DNA synthesis with a DNA primer. A primer is selected that will provide sequencing information 3’ to the primer. In order to generate a primer it is necessary to know the DNA sequence of the region that is adjacent to the region to be sequenced – this is known because the HIV protease gene is now inserted into the expression vector, with the following sequence:

5'--ACCAAGCGCGC**GAATTC**cctcagatcactctttggcaa----TACtttaaatttc**GGATCC**TAA-----3'[U]

3'--TGGTTCGCGCGCTTAAGggagtctagtgagaaaccgtt----ATGaaatttaaag**CCTAGG**ATT-----5'[L]

The primer that will be used is: 5’-ACCAAGCGCGC**GAATTC-3’**, generating the following template-primer junction:

 **1 17**

 **ACCAAGCGCGCGAATTC**

3'--TGGTTCGCGCGCTTAAGggagtctagtgagaaaccgtt----ATGaaatttaaag**CCTAGG**ATT-----5'[L]

**2. Termination:** Termination of a growing DNA chain at a *known base* (A, G, C or T) is achieved by the inclusion of a *small* amount (1%) of a single dideoxynucleoside base in with the mixture of *all four* normal bases. For example a reaction would contain dATP, dTTP, dCTP, dGTP and 1% ddATP. The small amount of ddATP would cause chain termination whenever it would be incorporated into the DNA in place of dATP. The incorporation of ddATP would be random and thus all possible chains that end in 'A' will exist in the products of the reaction.

Four *separate* reactions are performed, each with a different dideoxynucleotide (ddATP, ddGTP, ddTTP, or ddCTP), generating four separate collections of DNA molecules, each terminated at a *known* base. *Each* of these reaction mixtures contains:

i) DNA template & primer

ii) All four nucleoside triphosphates (dNTPs)

iii) DNA polymerase

iv) A small amount (1%) of *only* *one* of the four dideoxy NTP

**Example**: Products in ddC Reaction:

 **ACCAAGCGCGCGAATTC**

3'--TGGTTCGCGCGCTTAAGggagtctagtg…

 **ACCAAGCGCGCGAATTC** **ACCAAGCGCGCGAATTC**

3'--TGGTTCGCGCGCTTAAGggagtctagtg… 3'--TGGTTCGCGCGCTTAAGggagtctagtg…

 **ACCAAGCGCGCGAATTC**C **ACCAAGCGCGCGAATTC**C

3'--TGGTTCGCGCGCTTAAGggagtctagtg… 3'--TGGTTCGCGCGCTTAAGggagtctagtg…

**3. Determining the position of termination:** The position of the termination point is determined by measuring the sizes of the DNA molecules with gel electrophoresis. Since *DNA synthesis is initiated from a common point using a primer, all DNA molecules begin at the same point on the sequence. Consequently, the length of the DNA fragment indicates the position of the termination by the dideoxyNTP.*

The possible DNA fragments that would be obtained in each of the four reactions described above are listed below. Note that all if the fragments shown below end in a dideoxy base (3’-H). The full length product can also be seen in some cases, especially if the template is short, such that some of the chains never terminate with a dideoxy base.



|  |  |  |
| --- | --- | --- |
| Dideoxy |  Length | Sequence of Fragment |
| ddATP | 22 | ACCAAGCGCGCGAATTC**CCTCA**3'-H  |
|  | 24 | ACCAAGCGCGCGAATTC**CCTCAGA**3'-H |
|  | 27 | ACCAAGCGCGCGAATTC**CCTCAGATCA**3'-H |
| ddGTP | 23 | ACCAAGCGCGCGAATTC**CCTCAG**3'-H |
| ddCTP | 18 | ACCAAGCGCGCGAATTC**C**3'-H |
|  | 19 | ACCAAGCGCGCGAATTC**CC**3'-H |
|  | 21 | ACCAAGCGCGCGAATTC**CCTC**3'-H |
|  | 26 | ACCAAGCGCGCGAATTC**CCTCAGATC**3'-H |
|  | 28 | ACCAAGCGCGCGAATTC**CCTCAGATCAC**3'-H |
| ddTTP | 20 | ACCAAGCGCGCGAATTC**CCT**3'-H |
|  | 25 | ACCAAGCGCGCGAATTC**CCTCAGAT**3'-H |

The 17 bases from the primer contribute to the molecular weight, hence the first fragment has a length of 18.

The sequence is simply read from **bottom to top of the gel**, giving the sequence that is *complementary* to the *template* strand, or equivalently, the DNA sequence beginning at the end of the primer, in the 5'-3' direction:

 **ACCAAGCGCGCGAATTC**

3'--TGGTTCGCGCGCTTAAGggagtctagtg……