**Lecture 36: Restriction Enzymes, Ligase, DNA Polymerases, PCR**

**Overall Goal**: To produce HIV protease from a drug resistant HIV virus. The protein will be made in E. coli (a widely used bacteria) using recombinant DNA methods. The structure of the purified protein can be determined and new HIV protease inhibitors can be designed that will be effective against the altered virus. The overall procedure is as follows:

1) Isolate the genetic material (RNA) from the HIV virus that encodes the HIV protease.

2) Convert this material to DNA using polymerases, amplify HIV protease coding region using PCR (polymerase chain reaction)

3) Insert the DNA encoding HIV protease into a bacterial expression vector (a specialized form of a DNA plasmid) using restriction endonucleases and DNA ligase. Verify sequence using DNA sequencing.

**Properties of the expression vector:**

a) Provide antibiotic resistance to the host bacteria, such that only bacteria with the plasmid will grow in the presence of the antibiotic (selective pressure for maintaining the plasmid)

b) An origin of DNA replication so that the plasmid will be replicated with the bacterial DNA

c) DNA sequences that cause the production of mRNA, copying the information in the DNA to mRNA, including a regulated on/off switch (lac operator).

d) Sequences in the mRNA that start and stop the production of the recombinant protein.

e) Coding region for protein to be expressed (HIV protease).

f) Sequences in the mRNA that cause the recombinant protein to be exported out of the cell, to facilitate purification (leader peptide).

g) His6 tag for affinity purification.

4) Transformation of the bacteria with the plasmid.

5) Growth of the transformed bacteria.

6) Initiation, or induction, of the production of the recombinant protein.

7) Purification of the recombinant protein.

8) Structural and functional studies.

 DNA Modifying Enzymes:

**A. Restriction Endonuclease:** [*endo* - cut within, *nuclease* - cleave nucleic acid]. Used by bacteria to degrade invading viral DNA. Named after bacterial species the particular enzyme was isolated from, i.e. Eco = *E. Coli.*

1. Enzyme binds to specific **recognition sequence**s with near absolute specificity and high affinity (KD = 10-10 M).

2. Enzymes usually bind in *major* groove, forming *both* specific and non-specific interactions.

3. Homodimeric enzymes have 180 degree rotational symmetry. Because of the symmetry in the enzyme, the DNA sequence also symmetrical. The sequence is the same on the top and bottom strands (referred to as palindromic sequences).

4. Require Mg2+ for cleavage, usually cleave both strands at the same position.

**Enzyme Recognition Products**

 **Sequence**

**EcoRI(G^AATTC): -G-A-A-T-T-C -G A-A-T-T-C- 5’ overhang = sticky end**

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

**PvuI(CGAT^CG): -C-G-A-T-C-G- -C-G-A-T C-G- 3’ overhang = Sticky end**

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

**EcoRV(GAT^ATC): -G-A-T-A-T-C- -G-A-T A-T-C- blunt end**

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

**HaeIII (GG^CC): -G-G-C-C -G-G C-C- blunt end**

 **-C-C-G-G -C-C G-G-**

**Example: EcoR1: G^AATTC**

|  |  |
| --- | --- |
| a) Non-specific interactions with DNA phosphates. phos | ---G-A-A-T-T-C------C-T-T-A-A-G------G A-A-T-T-C------C-T-T-A-A G------G A-A-T-T-C------C-T-T-A-A G--- |
|  b) Specific hydrogen bonds with donor and acceptors at the edge of bases in the major groove: |
|  |  |

**B. DNA Ligase** – Uses ATP to join 5’phosphate to 3’-OH, provided the two groups are held in close proximity. Complementary overhangs and blunt ends can be ligated. Fragments created by the same restriction enzyme can always be joined to each other.

*Sticky end ligation:*

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

 -C-T-T-A-A G- -C-T-T-A-A G- --C-T-T-A-A**-**G—

*Blunt end ligation:*

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

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

**C. DNA Polymerases:** DNA polymerases play an important role in the replication of DNA.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Polymerases**  | **Template** | **Monomers** | **5'-3' Polymerase** | **3'-5' Exonuclease** |
| **DNA Pol III (***E. coli***)** | DNA | dNTP |  |  |
| **DNA Pol I (***E. coli***)** | DNA | dNTP |  |  |
| **Reverse Transcriptase (**HIV**)** | RNA/DNA | dNTP |  |  |

1. Utilized a **template** to direct the order of added bases.
2. DNA polymerases require a basepaired **primer with a 3’OH**, RNA poly. do not.
3. Synthesize chains from **5'→3' direction. Adding new bases to the 3' hydroxyl of the existing polymer.**
4. Use nucleotside triphosphates, producing a chain one base longer + pyrophosphate (PPi). The PPi is hydrolyzed to insure a large negative ΔG for the reaction. Yet another example of indirect coupling in biochemical reactions!

1. Fidelity of base incorporation is dependent on Watson-Crick base-pairing (A-T, G-C), *plus* purine-pyrimidine matching (see right). Phosphodiester bond formation occurs quickly (1 msec) when the correct match is made. Bond formation is slow when the bases are incorrectly matched – allowing time for proofreading. An example of kinetic control of substrate specificity.
2. Incorrectly incorporated bases are removed by a **proofreading activity: 3'→5' exonuclease activity**.

i) This activity is found in *most* polymerases.

ii) This activity is absent in many viral polymerase (e.g HIV rev. trans.), leading to high mutation rates in the virus and a high number of viruses that are resistant to HIV drugs.

**5'**→**3' Polymerase:**

5'G-C-T3'OH G-C-T-A3'OH G-C-T-A-T3'OH

3'C-G-A-T-A-C-C C-G-A-T-A-C-C C-G-A-T-A-C-C

**3'**→**5' Exonuclease:**

5'G-C-T-T3'OH G-C-T3'OH

3'C-G-A-T-A-C-C C-G-A-T-A-C-C

**Generation of dsDNA from HIV viral RNA (Step 1):**

**Reverse transcriptase** has the following activities:

i. RNA dependent DNA polymerase activity.

ii. RNase activity, degrades RNA template.

iii. DNA dependent DNA polymerase activity.

**PCR = Polymerase Chain Reaction (Step 2):**

* PCR is a method of *amplifying* a user-selected segment of DNA within a larger DNA molecule.
* The location of the amplified segment is *defined* by two primers, they anneal to their templates according to W-C basepairing rules and initiate polymerization from those sites.
* The final product contains the DNA between the two primers **and** the primers themselves.
* Any non-complementary bases at the 5’ end of the primers are also incorporated into the PCR product.

The actual reaction mix contains:

i. Double stranded DNA template

ii. Excess “right” and “left” primers

iii. dNTPs

iv. *Thermostable* DNA polymerase

The last item on the above list is quite important, since each cycle of PCR requires a heating step to denature the newly synthesized DNA and to allow the primers to bind. Therefore the polymerase must be able to withstand high temperature.

The amplification cycle [**D**NA **a**m**p**lification = **DAP**] consists of the following three steps, repeated 15-25 times. After the third cycle, each cycle doubles the amount of PCR product.

i. Denaturation of double stranded DNA (T=95 C) [**Denature**]

ii. Annealing of primers on each strand (T=55 C) [**Anneal**]

iii. Primer extension with DNA Polymerase + dNTPs (T ~ 78 C) [**Polymerize**]

The final PCR product encoding the HIV protease gene is:

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

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

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

The sequences at each end, containing GAATTC and GGATCC are not part of the HIV genetic material. These DNA sequences were added during the PCR reaction and contain restriction endonuclease sites that will be used to precisely insert the HIV gene into the expression plasmid.

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

**Primer Design:**

1. Sufficiently long that it will hybridize (bind via Watson-Crick H-bonds) uniquely to the *desired location*.

2. Sufficiently long that it will hybridize during the annealing step, e.g. TM = 60 C if the annealing T is 55C.

 **TM .= 81.5 +0.41(%GC) - 675/N**

 (675/N accounts for the dependence of melting temperature on length):

3. *May* contain additional sequences at the 5’ ends of the primers that will result in the incorporation of restriction site, plus additional bases (two in this example) to facilitate binding of the res. endonuclease to the DNA.

We would design primers that are the sequence of the start (top strand) and end (bottom strand) of the HIV protease DNA coding region, with an EcoR1 site (GAATCC) on the left primer and a BamHI site (GGATCC) on the right primer. The EcoR1-containing left primer would prime DNA synthesis using the *lower* strand as the template and the BamH1-containing right primer would prime synthesis using the *upper* strand as the template.

 5'CG**GAATTC**

cctcagatcactctttggc3’

5'--------ttccctcagatcactctttggcaa-------TACtttaaatttccccattagtcctatt-----3'[U]

 **ProGlnIleThrLeuTrpGln-----------LeuAsnPhe**

3'--------aagggagtctagtgagaaaccgtt-------ATGaaatttaaaggggtaatcaggataa-----5'[L]

 **3’**TGaaatttaaag

 **CCTAGG**GC5'

**The final product, after PCR.**

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

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

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

