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 (viral RNA) from the HIV virus that encodes the HIV protease.
2) Convert this material to DNA using polymerases and then amplify only the DNA bases the code for HIV protease using PCR.
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.

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 in this example).
   f) Sequences in the mRNA that cause the recombinant protein to be exported out of the cell, to facilitate purification (leader peptide).
   g) Histag for affinity purification on a Ni column.

4) Sequence mutant HIV protease gene to determine nature of mutation.
5) Transformation of the bacteria with the plasmid.
6) Growth of the transformed bacteria.
7) Initiation, or induction, of the production of the recombinant protein.
8) Purification of the recombinant protein.
9) Structural and functional studies.
10) Rational drug design.
Lecture 36: Reverse Transcription and PCR

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.

1A. Generate 1st DNA strand

1B. Generate 2nd DNA strand.

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 primers are DNA and synthesized chemically.
- 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 [DNA amplification = 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<T_M) [Anneal] (50-60°C)
iii. Primer extension with DNA Polymerase + dNTPs (T ~ 72-78°C) [Polymerize]
Primer Design:
1. Sufficiently long that it will hybridize (bind via Watson-Crick H-bonds) uniquely to the desired location.
   - Left primer – sequence of the top strand, at the boundary of the region to be amplified. This primer will anneal to the bottom strand.
   - Right primer – sequence of the bottom strand, at the boundary of the region to be amplified. This primer will anneal to the top strand.
2. Sufficiently long that it will give annealing temperatures in the range of 50-60°C, e.g. if \( T_M = 60 \) °C then anneal at 55°C.
   \[ T_M = 81.5 + 0.41(\%GC) - 675/N \]
   (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 sites, often including additional bases to facilitate binding of the res. endonuclease to the DNA.

Example – Simple amplification with no additional 5' sequences.

Left (upstream) primer:

Right (downstream) primer:

Cycle 1:

\[ \text{Denaturation} \]

\[ \text{Primer Annealing} \]

\[ \text{Polymerization} \]

Cycle 2 (after denaturation step) (Focusing on the new DNA strands that were made in cycle 1).

\[ \text{Primer Annealing} \]
Amplification of HIV protease.

The desired PCR product encoding the HIV protease gene is:

5' CGGAATTCCTCACAGCTCCTCCTG6CAA......TTTAAATTCCGATCCCG 3'
3' GCTTTAACGGAGCTCTAGAGAACCCTTT....AAATTTAGCCATGGCC 5'

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 (see next lecture).


C-G-G A-A-T-T-C-HIV---G
G-C-C-T-T-A-A G-Prot-C-C-T-A-G
G-G-C
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 (GAATTC) 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 BamHI-containing right primer would prime synthesis using the upper strand as the template.

Left Primer Sequence: 5'CGGAATTCCTCAAGATCAGACTCTTTTGGC3'
Right Primer Sequence: 5'CGGGATCCGAATTTAPAGT3'

The final product, after PCR.

5'CGGAATTCCTCAAGATCAGACTCTTTTGGC........TTTAAATTCCGGATCCGG 3'
3'GCCTTAAGGAGTCTAGTGAAAGGACGTT........AAATTAAAGCCATGGGC 5'

EcoR1 ProGlnIleThrLeuTrpGln--LeuAsnPhe BamHI

Follow PCR product