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) His6 tag 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 35: DNA Stability, Protein/DNA Interactions, Polymerases

Forces Stabilizing Nucleic Acid Structures.

Double stranded DNA & RNA can be reversibly denatured ("melting"). Cooperative transition from double stranded helix → single stranded random coils; the change in absorbance of the bases at λ=260 nm can be used to monitor this transition. The absorbance (A260) increases when the DNA melts (hyperchromatic effect).

\[ T_M \propto [NaCl] \]
\[ T_M \propto \%GC \]

Comparison of Dominate Forces in DNA and Protein Stability:

<table>
<thead>
<tr>
<th>Energetic Term</th>
<th>Protein Stability</th>
<th>dsDNA stability</th>
<th>Molecular Description of Energetic Terms in DNA Stability.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔH</td>
<td>Electrostatic interactions</td>
<td>0</td>
<td>Higher salt reduces electrostatic repulsion; ( T_M \uparrow )</td>
</tr>
<tr>
<td></td>
<td>Hydrogen Bonds</td>
<td>+</td>
<td>+ +</td>
</tr>
<tr>
<td></td>
<td>Van der Waals</td>
<td>++</td>
<td>+ +</td>
</tr>
<tr>
<td>ΔS</td>
<td>Conformational Entropy</td>
<td>------</td>
<td>Unfolded chain is higher in entropy, protein and DNA.</td>
</tr>
<tr>
<td></td>
<td>Hydrophobic Effect</td>
<td>+ + +</td>
<td>Bases are polar and exposed to solvent, minor hydrophobic effect in DNA.</td>
</tr>
</tbody>
</table>

2. DNA-Protein Interactions

Specificity is often uncorrelated to affinity.
Forces and functional groups involved in recognition.

**A. non-sequence specific**

1. Electrostatic bonding to the backbone.
   a) side chains of Lys and Arg to phosphates.
   b) release of metal ions (e.g. K⁺) favors binding (large increase in ΔS of ions).
   c) Binding affected by NaCl concentration.

2. Van der Waals: Stacking (and intercalation) of Phe, Trp, and Tyr side chains. More prevalent in single stranded (ss) nucleic acid.

**B. Sequence Specific**

3. Hydrophobic interaction with the 5-methyl of T.

4. Non-Watson-Crick Hydrogen bonding to the polar edges of the bases and to sugars, usually in the major groove.
   (Major groove is large arc of a circle drawn through the basepair, minor groove is small arc).
   a) Side chains of Arg, Asn, Gln, Tyr...
   b) Protein mainchain, C=O, NH.

   **Note:** WC hbond acceptor, C=O, can participate in one additional non-WC H-bonds.

---

**Example:**
The curves show the dissociation constants (K_d) for DNA binding for two different proteins (A, B) as a function of salt concentration and RNA composition.

**Protein A:** Binding independent of salt but not phosphate.
Large drop in K_d when U is replaced by C: recognizing base.

**Protein B:** Binding dependent on salt, independent of base.
Likely interacting with phosphate.
Restriction Endonuclease: [endo - cut within, nuclease - cleave nucleic acid].

- Provide a tool to cleave DNA at specific DNA sequences.

Isolated from bacteria, normal biological function is to degrade invading viral DNA. Named after bacterial species the particular enzyme was isolated from.

A. DNA Recognition – “Restriction Site”
1. Restriction enzymes binds to specific recognition sequences with near absolute specificity and high affinity ($K_0 = 10^{-19}$ M).
2. Enzymes usually bind in major groove, forming both specific and non-specific interactions.
3. These 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).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Recognition Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI</td>
<td>-G-A-A-T-T-C-</td>
</tr>
<tr>
<td></td>
<td>-C-T-T-A-A-G-</td>
</tr>
<tr>
<td>PstI</td>
<td>-C-T-G-C-A-G-</td>
</tr>
<tr>
<td></td>
<td>-G-A-C-G-T-C-</td>
</tr>
<tr>
<td>EcoRV</td>
<td>-G-A-T-A-T-C-</td>
</tr>
<tr>
<td></td>
<td>-C-T-A-T-A-G-</td>
</tr>
<tr>
<td>HaeIII</td>
<td>-G-G-C-C-</td>
</tr>
<tr>
<td></td>
<td>-C-C-G-G-</td>
</tr>
</tbody>
</table>

Example: EcoRI: GAATTC

a) Non-specific interactions with DNA phosphates.

b) Specific hydrogen bonds with donor and acceptors at the edge of bases (major groove):

B. DNA Cleavage (Shorthand – give sequence of top strand & indicate cut site ^)
1. Require Mg$^{2+}$ for cleavage. Generate a 3’OH and 5’ phosphate.
2. Cleave both strands at the same location, due to rotational symmetry.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Recognition Sequence</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>(G&quot;AATT&quot;)</td>
<td>-C-T-T-A-A-G-</td>
<td>C-T-T-A-A-G-</td>
</tr>
<tr>
<td>PstI:</td>
<td>-C-T-G-C-A-G-</td>
<td>A-C-G-T-C- 3’ overhang</td>
</tr>
<tr>
<td>(CTGCA&quot;G&quot;)</td>
<td>-G-A-C-G-T-C-</td>
<td>C-T-G-C-A-</td>
</tr>
<tr>
<td>(GAS&quot;ATC&quot;)</td>
<td>-C-T-A-T-A-G-</td>
<td>A-T-C-</td>
</tr>
</tbody>
</table>

3. Generate different types of ends, dep. on cleavage location.
DNA Ligase:
- Provides a tool to join DNA fragments together.
- Uses ATP to join 5′ phosphate to 3′-OH
- Two ends of the DNA have to be held in close proximity to allow reformation of the phosphodiester bond.

Sticky end ligation:

Blunt end ligation:

DNA Polymersases:
- Utilize a template to direct the order of added bases.
- Require a basepaired primer with a 3′OH. Primer can be DNA or RNA.
- Synthesize chains from 5′→3′ direction. Adding new dNTP to the 3′ hydroxyl of the existing polymer. Pyrophosphate (PPi) is released.
- The PPi is hydrolyzed to insure a large negative ΔG for the reaction. What type of energy coupling is this?
- Fidelity of base incorporation is dependent on Watson-Crick base-pairing (A-T, G-C), plus purine-pyrimidine matching.
- Kinetic editing: 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 incorrect base to leave.

Proofreading activity: 3′→5′ exonuclease activity (error correction).
(exo = end, exonuclease = cut from the end)

Note: This activity is absent in many viral polymerases (e.g. HIV reverse transcriptase), leading to high mutation rates in the virus and a high number of viruses that are resistant to antiviral drugs.