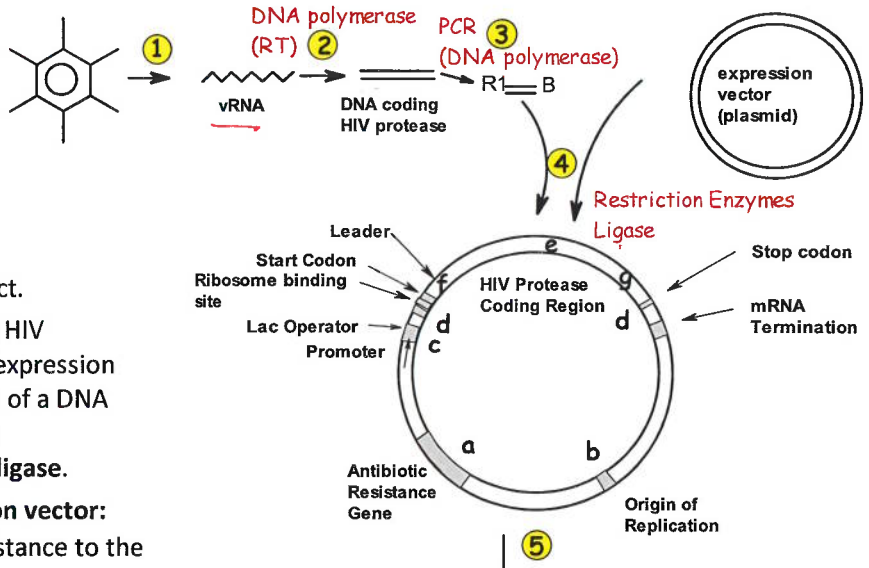
Overall Goal: To produce HIV protease from a drug resistant HIV virus. The protein will be made in E. coli (a widely used bacteria) using recombinant DNA methods. The structure of the purified protein can be determined and new HIV protease inhibitors can be designed that will be effective against the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

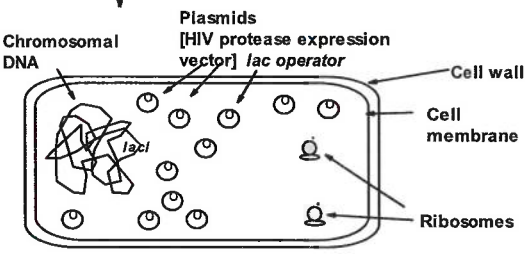
1) Isolate the genetic material (viral RNA) from the HIV virus that encodes the HIV protease.

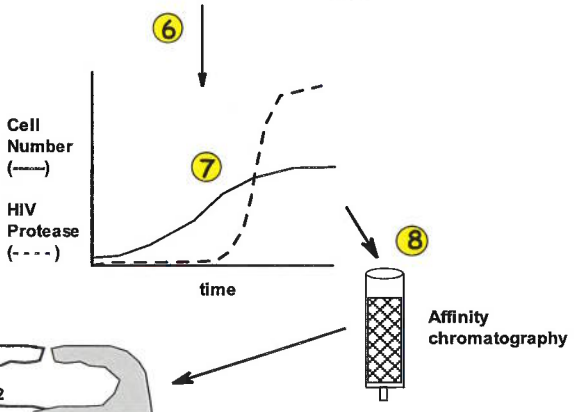
2) Convert this material to DNA using reverse transcriptase **(polymerase).**

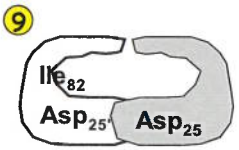
3) Amplify only the DNA bases the code for HIV protease using PCR. Add restriction sites on the end of the PCR product.

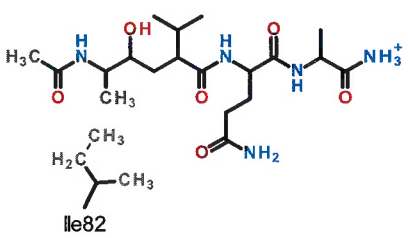
4) Insert the DNA encoding HIV protease into a bacterial expression vector (a specialized form of a DNA plasmid) using **restriction endonucleases** and **DNA ligase.**











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) His_6 tag for affinity purification on a Ni column.

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.

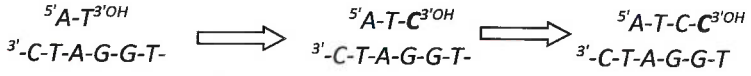
9) Structural and functional studies.

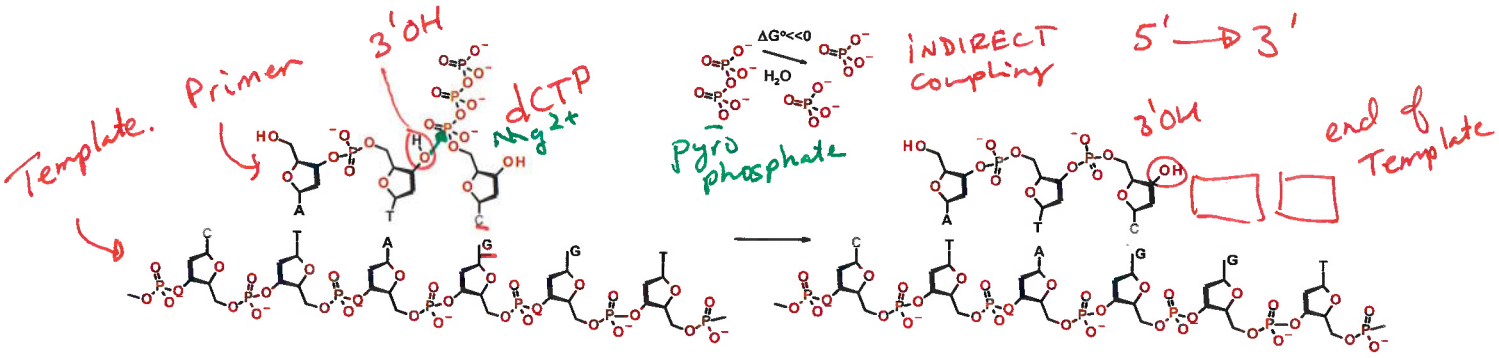
10) Rational drug design.

Lecture 37: DNA Polymerases & Applications

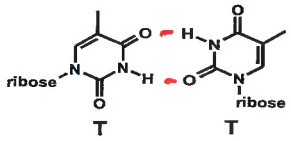
DNA Polymerases:

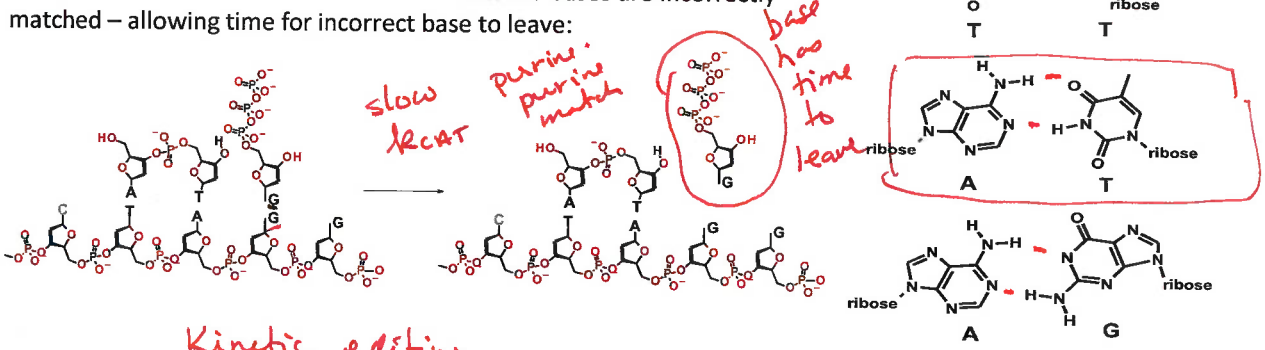
- Utilize a **template** to direct the order of added bases.
  - Most polymerases use a DNA template
  - HIV Reverse transcriptase uses an RNA template (the HIV genome)
- Requires a basepaired **primer** with a 3'OH. Primer can be DNA (biotech applications) or RNA (DNA replication).
- Synthesize chains from **5'→3' direction.** Adding new dNTP to the 3' hydroxyl of the existing polymer until the template ends. Pyrophosphate is released and hydrolyzed to insure a - ΔG.



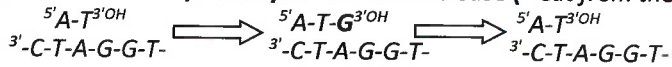


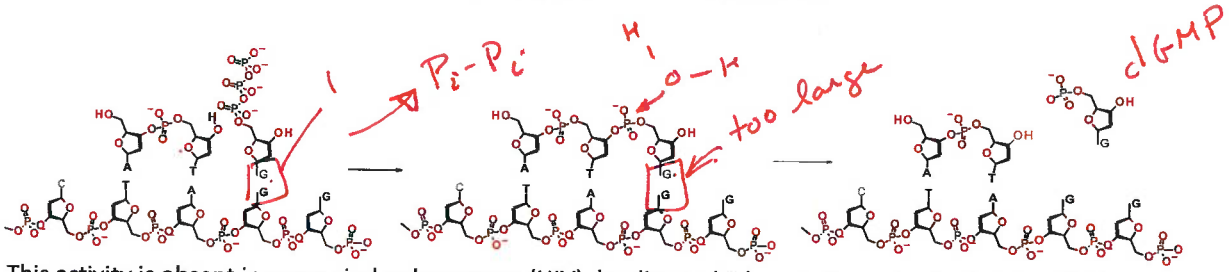
Replication Fidelity of base incorporation is dependent on Watson-Crick base-pairing (A-T, G-C), *plus* purine-pyrimidine (size) matching. Phosphodiester bond formation occurs quickly (1 msec) when the correct match is made. Bond formation is slow when the bases are incorrectly matched – allowing time for incorrect base to leave:





Proofreading (error correction) activity: 3'→5' exonuclease (= *cut from the end*)





This activity is absent in many viral polymerases (HIV), leading to high mutation rates in the virus & drug resistant strains of HIV.

PCR - Polymerase Chain Reaction:

- PCR is a method of *amplifying* a user-selected segment of DNA within a larger DNA molecule, e.g. the segment that codes for HIV protease.
- The location of the amplified segment is *defined* by two primers, they anneal to their templates according to Watson-Crick pairing rules (AT, GC) and initiate polymerization from those sites. The primers are DNA and synthesized chemically.
- The final product (~ 2^n) contains the DNA sequence 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 (lower figure).

The actual reaction mix contains:

i. Double stranded DNA template (pM to nM)

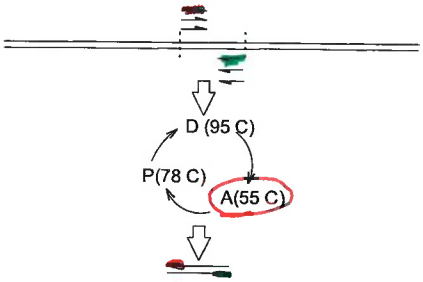
ii. Excess (1 uM) "right" and "left" primers. These will be consumed during each PCR cycle) since they are incorporated into the PCR product.

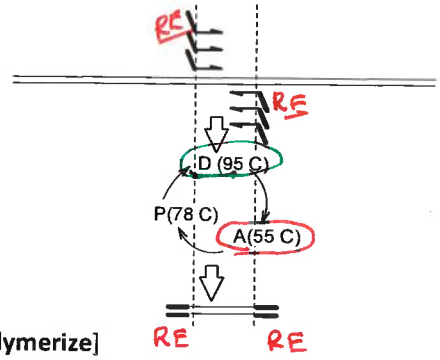
- Left primer = sequence of top strand at left boundary
- Right primer = sequence of the lower strand at the right boundary.

iii. dNTPs (mixture of all four deoxynucleoside triphosphates)

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**]

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

Primer Design:

A. Specificity: Using the following template in PCR:

Amplify → Template

```
5′-CATATGCAGGATCACCACCATCACGGATCCACCTCCACGGATGACAAGAAAAAGGGTAAATTCATTGTGTTCGAGGGTCTGGAT-
3′-GTATACGTCCTAGTGGTGGTAGTGCCTAGGTGGAGGTGCCTACTGTTCTTTTTCCCATTTAAGTAACACAAGCTCCCAGACCTA-
```

With primers 5′-CACCATC and 5′-AGACCC, the above template gives the following PCR product:

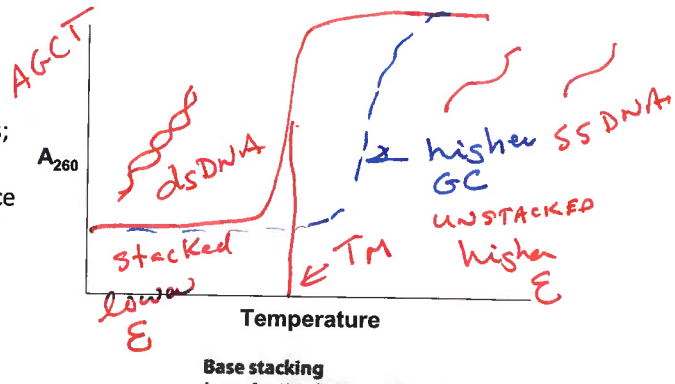
```
5′ CACCATCACGGATCCACCTCCACGGATGACAAGAAAAAGGGTAAATTCATTGTGTTCGAGGGTCT
3′ GTGGTAGTGCCTAGGTGGAGGTGCCTACTGTTCTTTTTCCCATTTAAGTAACACAAGCTCCCAGA
```

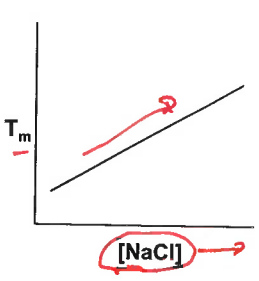
```
5′-CATATGCAGGATCACCACCATCACGGATCCACCTCCACGGATGACAAGAAAAAGGGTAAATTCATTGTGTTCGAGGGTCTGGAT-
                                                                              CCCAGA
          5′ CACCATC 3′
3′-GTATACGTCCTAGTGGTGGTAGTGCCTAGGTGGAGGTGCCTACTGTTCTTTTTCCCATTTAAGTAACACAAGCTCCCAGACCTA-
                                    ⇩
5′-CATATGCAGGATCACCACCATCACGGATCCACCTCCACGGATGACAAGAAAAAGGGTAAATTCATTGTGTTCGAGGGTCTGGAT-
  -GTATACGTCCTAGTGGTGGTAGTGCCTAGGTGGAGGTGCCTACTGTTCTTTTTCCCATTTAAGTAACACAAGCTCCCAGA

                CACCATCACGGATCCACCTCCACGGATGACAAGAAAAAGGGTAAATTCATTGTGTTCGAGGGTCTGGAT-
3′-GTATACGTCCTAGTGGTGGTAGTGCCTAGGTGGAGGTGCCTACTGTTCTTTTTCCCATTTAAGTAACACAAGCTCCCAGACCTA-
                                    ⇩
5′-CATATGCAGGATCACCACCATCACGGATCCACCTCCACGGATGACAAGAAAAAGGGTAAATTCATTGTGTTCGAGGGTCTGGAT-

                CACCATCACGGATCCACCTCCACGGATGACAAGAAAAAGGGTAAATTCATTGTGTTCGAGGGTCT
  -GTATACGTCCTAGTGGTGGTAGTGCCTAGGTGGAGGTGCCTACTGTTCTTTTTCCCATTTAAGTAACACAAGCTCCCAGA

                CACCATCACGGATCCACCTCCACGGATGACAAGAAAAAGGGTAAATTCATTGTGTTCGAGGGTCTGGAT-
                GTGGTAGTGCCTAGGTGGAGGTGCCTACTGTTCTTTTTCCCATTTAAGTAACACAAGCTCCCAGA
3′-GTATACGTCCTAGTGGTGGTAGTGCCTAGGTGGAGGTGCCTACTGTTCTTTTTCCCATTTAAGTAACACAAGCTCCCAGACCTA-
```

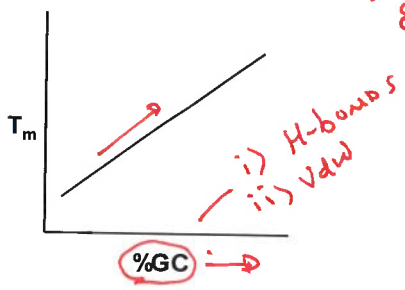
B. Melting Temperature

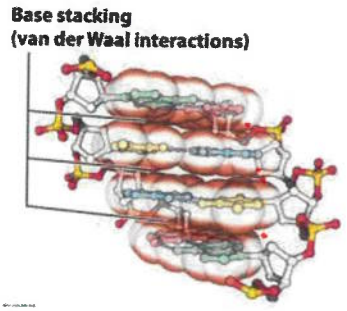
Forces Stabilizing Nucleic Acid Structures.

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 (A_{260}) increases when the DNA melts (**hyperchromatic effect**) due to disruption of base stacking when DNA becomes single stranded.

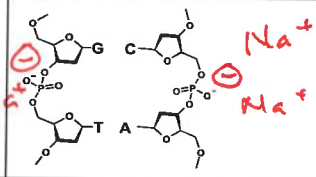
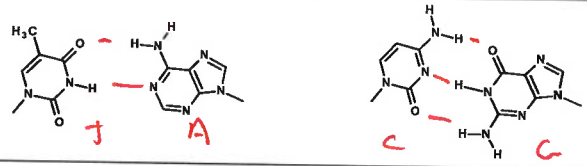
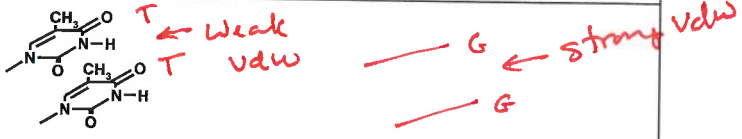
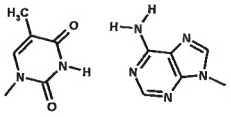








Comparison of Dominate Forces in DNA and Protein Stability:

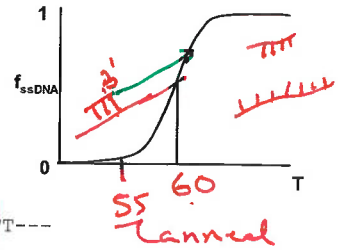
	Energetic Term	Protein Stability	dsDNA stability	Molecular Description of Energetic Terms in DNA Stability.
Δ H	Electrostatic interactions	0	unfav --	
	Hydrogen Bonds	+	++	
	Van der Waals	++	++++++	
Δ S	Conformational Entropy	--------	--------	Unfolded chain is higher in entropy, protein **and** DNA.
	Hydrophobic Effect	++++	+	 Bases are polar and exposed to solvent, minor hydrophobic effect in DNA.

T_M ≈ **81.5 +0.41(%GC) - 675/N** (675/N accounts for the dependence of T_M on length).
This formula does not take into account local sequence effects, which can affect the melting temperature because of different van der Waals interactions between bases.

C. Summary (1+2) and Other Factors.

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 = 60C then anneal at 55C.



3. Has a one or two G/C at the 3' end to ensure the 3' end of the primer is paired to the template. *Which of the following are better primers?*

```
      CTCAGATCACTCTTTGGCA                    CTCAGATCACTCTTTGGC
---AAGGGAGTCTAGTGAGAAACCGTT---         ---AAGGGAGTCTAGTGAGAAACCGTT---
```

better

4. Primer pairs are not complementary to themselves or each other in a way that produces a recessed 3'-OH (often referred to as primer dimers). Otherwise, all of the dNTPs would be consumed by the following reaction:

```
Left primer:     5'ATATGATCGTACTGGATCC3'OH  →  ATATGATCGTACTGGATCCTACGA
Right primer:      3'OHCTAGCATGACCTAGGATGCT     TATACTAGCATGACCTAGGATGCT
```

5. Primer should not form double stranded hairpin, which would reduce its ability to anneal to the template.

Optional Feature - Additional sequences at the 5' ends of the primers will be "added" to the ends of the PCR product. This provides a method of adding restriction sites to the ends of the amplified DNA segment. Often additional bases are added to the 5' end of the restriction site to facilitate binding of the res. endonuclease to the DNA (see example below, CG has been added).

Example - Primer Design to amplify the HIV protease coding Region: 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 additional CG at the 5' end of the primers will facilitate cleavage by the restriction enzymes.

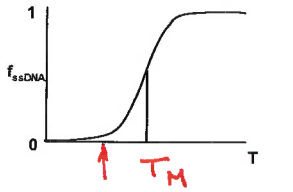
```
LP → 5'CGGAATTC
          EcoR1
                 CCTCAGATCAC3'
5'--------TTCCCTCAGATCACTCTTTGGCAA-------TACTTTAAATTTCCCCATTAGTCCTATT-----3'[U]   ← HIV genome
          ProGlnIleThrLeuTrpGln-----------LeuAsnPhe  ← amino acid sequence of HIV protease
3'--------AAGGGAGTCTAGTGAGAAACCGTT------ATGAAATTTAAAGGGGTAATCAGGATAA-----5'[L]
                                        3'ATGAAATTTAAAG
                                                      \
                                                       CCTAGGGC5'  ← RP
                                                       BamHI
```

Left Primer Sequence: 5'CG**GAATTC**CCTCAGATCAC

T_M = 81.5 +0.41(%GC) - 675/N = 81.5 +0.41(100*10/19)-675/19 = 67.6

Right Primer Sequence: 5'CG**GGATCC**GAAATTTAAAGTA

T_M = 81.5 +0.41(%GC) - 675/N = 81.5 + 0.41(100*8/21)-675/21 = 65.0



The final PCR product is (see pg 5 for diagram of PCR cycles.

```
5' CGGAATTCCCTCAGATCACTCTTTGGCAA.......TTTAAATTTCGGATCCG 3'
3' GCCTTAAGGGAGTCTAGTGAGAAACCGTT.......AAATTTAAAGCCATGGGC 5'
    EcoR1 ProGlnIleThrLeuTrpGln--------LeuAsnPhe BamH1
```

There are a number of sites that can be used to calculate primer properties.
https://www.idtdna.com/calc/analyzer
(Gives more accurate T_M values).

OligoAnalyzer

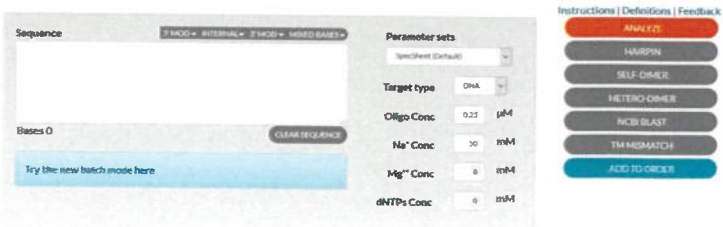


Illustration - Amplification of HIV Protease using PCR: This shows how the 5' sequences on the end of the primers are incorporated into the PCR product (note that only selected templates are followed in this diagram, as indicated by the curved arrows in the margins).