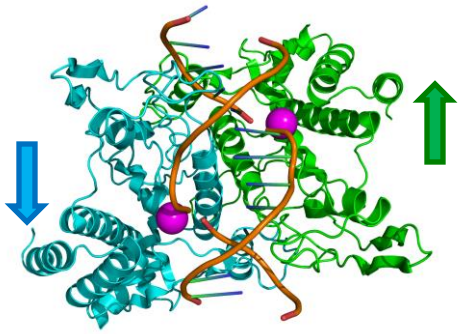
BNFO301: Introduction to Bioinformatics

Restriction of M13 DNA by the restriction enzyme TaqI

If restriction enzymes are mysterious to you, it may help to make them more tangible. To the right is a drawing of a restriction enzyme acting on DNA. Most commonly used restriction enzymes are composed of two identical subunits arranged so that one subunit is upside down with respect to the other. Each subunit recognizes a specific sequence on the DNA. Since the subunits are identical and antiparallel, so are the recognition sequences. As a result, recognition sites of commonly used restriction enzymes are palindromes. For example:

EcoRI recognizes
```
5'-GAATTC-3'
3'-CTTAAG-5'
```

TaqI recognizes
```
5'-TCGA-3'
3'-AGCT-5'
```



The two-subunit protein, EcoRI, binding and cutting DNA at its recognition site. The arrows indicate that the identical subunits are oriented antiparallel with respect to one another. The brown noodles are the two DNA backbones. The pink balls represent atoms of magnesium, required for enzyme activity. Picture from Wikipedia, Restriction Enzyme.

See TaqI in action as it acts on M13 phage by doing the following:

- Go to Phantome/BioBIKE
- Log in (no prior registration required), either as a guest or by registering, if you haven't already
- Bring down the SEQUENCE-OF function from the alphabetical ALL menu
- Type in the *entity* box `ent-m13` (the nickname of enterobacteriophage M13) and press the **Enter** key
- Execute the function by either double-clicking SEQUENCE-OF or clicking **Execute** on the menu found by mousing over the action icon (the green wedge)

You should now be looking at the very small genome sequence of M13. Now to highlight the sites recognized by restriction enzyme TaqI.

- Type in the **Search** box `TCGA` and click Go (or simply press the **Enter** key)

You should see the instances of TCGA highlighted in gray. The DNA in between the highlighted sites (plus a bit of the gray DNA) are the fragments you'll get by cutting with TaqI.

SQR1. What is the coordinate of the first TaqI site?

How big are these fragments? You could calculate from the coordinates of the gray-highlighted sequences, but it's much easier to let BioBIKE do the calculation. Go back to the green screen and:

- Bring down the DIGESTION-OF function from the alphabetical ALL menu
- Type in the *sequence* box `ent-m13` and press the **Tab** key

- Type in the *enzyme* box 'TaqI (the single quote is necessary to tell the function that you are referring to the enzyme TaqI and not a variable named TaqI that might contain anything at all) and press the **Enter** key
- Execute the function by either double-clicking SEQUENCE-OF or executing from the action icon

You should get a pop-up window containing the coordinates of each TaqI site and the distances between each site, i.e. the sizes of the fragments produced by digesting M13 with TaqI.

SQR2. How many TaqI sites are there? How many fragments are produced? Why that number and not one more?

SQR3. Compare the coordinates with those of the gray-highlighted sites in the sequence you displayed earlier. Are they the same, more or less?

SQR4. Compare the coordinates and lengths with Fig. 1 from Fuller et al (1984) [Cell 38:889-900]. What discrepancies do you find? Why?

Figure 1 from Fuller et al (1984) doesn't show M13 but a derivative of M13. If you understand how that derivative differs from M13, you may be able to explain some of the differences between the digestion you virtually performed in BioBIKE and the digestion shown in the figure.

But there's still one more complication. E. coli modifies its DNA at GATC sites as a way to distinguish newly synthesized (unmodified) DNA from old (modified) DNA. It does so by putting a methyl group on the A of GATC. But TaqI does not recognize its recognition sequence of TCGA if one or both of the two A's are methylated.

SQR5. Why did I say two A's?

SQR6. What are two 6-nucleotide sequences containing TCGA that TaqI will not recognize? Are there any such sites in M13?

Indeed, there ***are*** such sites in the M13 derivative. So some of the TCGA sequences in its DNA are not cut by TaqI. This is acknowledged in the legend to Fig. 1 of Fuller et al (1984).