# HW 1: Manipulating DNA Data as Strings

The goal here is to do several basic calculations on DNA data, which we will represent using a string.

First you must download the file `dna_strings.ipynb` from the class web page. Start Jupyter, and upload this file. Make the necessary additions to it, and then download it back to your computer. It will be this file that you ultimately submit.

Here are the tasks that you must perform:

1. **Calculate the AT content.** The DNA string consists of four letters: `A`, `C`, `G`, and `T`. You must write a short program that will calculate the fraction of this string that is `A` or `T`.

2. **Complement the DNA.** Calculate the bases that would be on the complementary strand of DNA. To do this you must do two things. You must change every base to its complement (`A` to `T`, `T` to `A`, `C` to `G`, and `G` to `C`), and you must reverse the entire string. For example, if the original string were `ACC`, the complement would be `GGT`.

3. **Calculate the fragment lengths when the DNA is cut with a restriction enzyme.** The EcoRI restriction enzyme will cut the DNA into two fragments, when it finds the data `G*AATTC`. Thus, the first cut fragment will end with `G`, and the second one will begin with `AATTC`. Your program should print the length of each fragment.

4. **Splice out the introns.** An intron is a region of DNA that is skipped over when the code is used to generate a new protein. (More accurately, it is *spliced* out of the RNA.) Exons are regions that *are* read. We assume that the DNA fragment has an intron in the middle: the first 63 characters make up the first exon, and the second exon runs from character 91 (in 1-based counting) to the end. You must print out:

   - Each exon, separately.
   - The fraction of the DNA strand that is part of either exon.
   - The full DNA strand, but the non-coding intron will consist of lower-case letters.

Note that for this homework, these problems are very close to ones in the book. If you get stuck, the book has some answers that might help you.