

Introduction to Bioinformatics

Problem Set 5: DNA Structure

1. You work at the Center for Disease Control and have been given to analyze a suspected bioterror agent that may be a type of Variola, a double-stranded DNA virus that is the causative agent of smallpox. Smallpox was one of humanities major scourges, killing as many as 500 million people in the 20th century. It is believed to have been wiped out in 1979 with an aggressive vaccination campaign. Since few alive today are resistant to smallpox, a live virus could wreak havoc.

To analyze the virus, you obtain DNA from the precious sample of purified virus and (as a control) from canine parvovirus, a single-stranded DNA virus you happen to have on hand. Unfortunately, you forget to label the two tubes! What else to do but to continue,... so you analyze the two DNA samples for nucleotide composition (in the same way as Chargaff), giving the results shown below. Which tube contains viral DNA, and why?

	G	A	T	C	Recovery
Tube 1	15.7	33.4	32.7	11.4	93.2
Tube 2	18.7	34.4	26.9	14.1	94.1

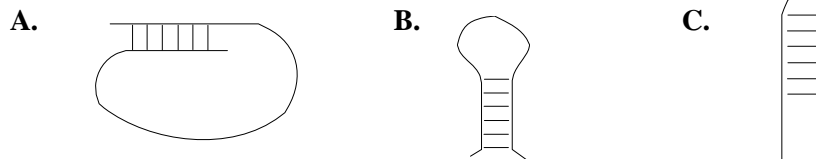
2. You have good reason to believe that an ATG start codon lies somewhere in the double-stranded DNA whose sequence is represented below. List all the sites where that codon may lie.

10 20 30
 GGCAGAATGTGAAGCTAGGCATTGTACCTAG

3. Consider the sequence below of a single-stranded DNA fragment.

5' -AGAGAGAGCTAAGGTCTCTCC-3'

Which of the following is a likely structure for the single-stranded fragment to assume?



4. You isolate a new restriction enzyme EcoRX (an enzyme that cuts DNA at a specific position). You find that the enzyme cuts the sequence ATGGTATACTGAACGAA once. Like many proteins that bind to DNA, restriction enzymes generally recognize palindromic sequences, usually 4 to 8 nucleotides in length.
- Presuming that EcoRX does so as well, what is the sequence it recognizes and cuts?
 - About how often would you expect to find this sequence in a random piece of DNA? (form of answer should be "one site every X nucleotides")
 - What assumptions did you make in determining your answer to Part b.?
 - The cyanobacterium *Anabaena* has a genome 7.2 megabases in size and about 21% of its bases are guanines. Predict the number of EcoRX sites in the *Anabaena* genome.