Homework 3 Name_____

Handed out: 9/22/08
Due: 9/29/08

To submit this assignment I prefer you use the 'submit' command on either glue.umd.edu or linux.grace.umd.edu. Alternatively, you can either email me the completed assignment or hand it in class. The assignment is due at midnight.

1. What are the sp(i) values for the string below (think pattern in the KMP algorithm):

```
P =  A C T T A C T A A C T C T A A T A G G A C T T A C T C T A
sp = 0 0 0 0 1 2 3 1 1 2 3 0 0 1 1 0 1 0 0 1 2 3 4 5 6 7 0 0 1
```

2. Assume you are matching the string above against a long text, e.g. the human genome. Assume you've matched all characters up to position 10 (starting the numbering from 0, string ending at AACT), but mismatched on the 4th C of the pattern.
   How far will you have to shift the pattern before attempting to match the string again?
   Which is the next character in the pattern that will be compared to the text?

   (simply draw the next position of the pattern below and circle or underline the character that will be compared to the text)

```
P =  A C T T A C T A A C T C T A A T A G G A C T T A C T C T A
                    A C T T A C T A A C T C T A A T A G G A C T T ...
```

3. With the help of the example from part 1, can you figure out some of the following relationships? Please explain.
   1. Define one possible relationship between sp(i) and sp(i – 1).

      sp(i) <= sp(i - 1) + 1

   2. Can you define a similar relationship between Z[i] and Z[i – 1]?

      no

   3. Assuming you already know all the Z-values for a string, could you use them to compute the sp(i) values? In other words, what is the relationship between the sp values and the Z values?

      Z[i] = sp(i + Z[i])

      You can compute the sp values by moving right to left through the Z array and filling in the sp values using the formula above. why right to left? For you to figure out.