First, the program reads a user-inputted DNA sequence as a string. With this string, the program then converts all of the characters to uppercase, checks the string for any invalid characters (any character but A, C, G, or T), and removes them. This string is then processed to calculate the total number of bases as well as the number of each particular base (adenine, cytosine, guanine, and thymine) in the sequence. The percent G~C content is then calculated and, finally, all of the results are printed.

Generally, housekeeping promoters (yellow, in the above graph) have a higher G~C content than tissue specific promoters (green, in the above graph).