**DNA to Free Energy**

**Practice**

Promoter1: “ACT**G**AA”

Promoter2: “CAT**T**AT”

Promoter3: “CGT**A**CG”

The bold underlined letters represent the position that represents the transcription site. The length-6 sequence is the promoter region that is 3 nucleotides upstream and 2 nucleotides downstream of that position.

1. List the dinucleotides for each sequence; there should be five of them:
   1. Promoter1: AC, CT,
   2. Promoter2:
   3. Promoter3:
2. Find the free energy value for each dinucleotide. See table on the next page. Look at the first column, the dinucleotides to the left of the backslash. Each dinucleotide is either on that list, or its reverse complement is. Write down the first value that appears in the last column of the table.

Promoter1:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Position | -3 | -2 | -1 | 1 | 2 |
| dinucleotide | AC | CT |  |  |  |
| (reverse complement) | GT |  |  |  |  |
| free energy value | -1.44 | -1.28 |  |  |  |

Promoter2:

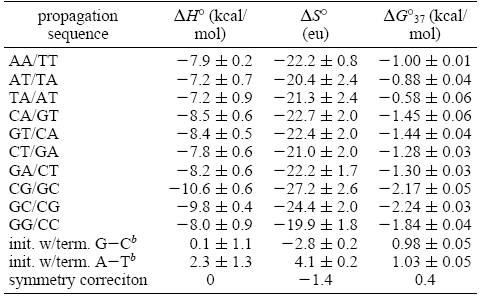
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Position | -3 | -2 | -1 | 1 | 2 |
| dinucleotide |  |  |  |  |  |
| (reverse complement) |  |  |  |  |  |
| free energy value |  |  |  |  |  |

Promoter3:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Position | -3 | -2 | -1 | 1 | 2 |
| dinucleotide |  |  |  |  |  |
| (reverse complement) |  |  |  |  |  |
| free energy value |  |  |  |  |  |

Calculate the average free energy for each column.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Avg(-3) | Avg(-2) | Avg(-1) | Avg(1) | Avg(2) |
|  |  |  |  |  |



*Goal*: Use the method describe in the referenced papers [1, 2] to compute the average free energy for each position in the E. coli genome. Write results to a separate file. (See Figure 1 below for a listing of free energy values.)

Input: datafile

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Position -3* | *Position -2* | *Position -1* | *Position 1* | *Position 2* |
| -2.30 | -5.61 | -1.00 | -2.82 | -4.32 |
| -1.43 | -7.56 | -0.53 | -1.44 | -3.89 |
| -0.56 | -4.83 | -0.98 | -2.09 | -5.01 |

Output: Array of free energy values for each position

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Avg(energy1) | Avg(energy2) | Avg(energy3) | Avg(energy4) | Avg(energy5) |

**References (both found in my Github account under folder “Articles”)**

1. Allawi, H. T., SantaLucia, J. Jr. (1997). Thermodynamics and NMR of internal G.T mismatches in DNA. Biochemistry, 36, 10581–10594.
2. Kanhere, A., Bansal, M. (2005). A novel method for prokaryotic promoter prediction based on DNA stability. BMC Bioinformatics, 6, 1471–2105.