Description of Graphs

Find below the detailed description of each graph. Note that characters A, C, G and T in the formulas denote the number of corresponding nucleotide in a window.

- *DNA Flexibility* — searches for regions of high DNA helix flexibility in a DNA sequence. The average *Threshold* in a window is calculated by the following formula:

```
(sum of flexibility angles in the window) / (the window size - 1)
```

For more detailed information see *DNA Flexibility* paragraph.

- *GC Content (%)* — shows the percentage of nitrogenous bases (either guanine or cytosine) on a DNA molecule. It is calculated by the following formula:

```
(G+C)/(A+G+C+T)*100
```

- *AG Content (%)* — shows the percentage of nitrogenous bases (either adenine or guanine) on a DNA molecule. It is calculated by the following formula:

```
(A+G)/(A+G+C+T)*100
```

- *GC Frame Plot* — this graph is similar to the GC content graph but shows the GC content of the first, second and third position independently. It is most effective in organisms with GC rich genomic sequence but it also works on all microbial sequences.
- *GC Deviation (G-C)/(G+C)* — shows the difference between the “G” content of the forward strand and the reverse strand. *GC Deviation* is calculated by the following formula:

```
(G-C)/(G+C)
```

- *AT Deviation (A-T)/(A+T)* — shows the difference between the “A” content of the forward strand and the reverse strand. *AT Deviation* is calculated by the following formula:

```
(A-T)/(A+T)
```

- *Karlin Signature Difference* — dinucleotide absolute relative abundance difference between the whole sequence and a sliding window. Let:

```
f(XY) = frequency of the dinucleotide XY
f(X)  = frequency of the nucleotide X

p(XY) = f(XY) / f(X) * f(Y)

p_seq(XY) = p(XY) for the whole sequence
p_win(XY) = p(XY) for a window
```

The *Karlin Signature Difference* for a window is calculated by the following formula:

```
sum(p_seq(XY) - p_win(XY)) / 16
```

- *Informational Entropy* — is calculated from a table of overlapping DNA triplet frequencies. The use of overlapping triplets smooths the frame effect. *Informational Entropy* is calculated by the following formula:

```
-(triplet frequency)*log10(triplet frequency)/log10(2)
```