

# Information about Sequence

Statistics about an opened sequence can be found on the *Statistics* tab of the *Options Panel*. When a region is selected in the sequence, the statistics is interactively re-calculated for this region only. The following information is available:

- Common statistics (length, molecular weight, etc.) - see the detailed description below
- Characters occurrence
- Dinucleotides occurrence - available for nucleotide sequences only

Note that all data, displayed on the *Statistics* tab, can be selected with the mouse and copied. Use the copy item in the context menu or a shortcut - Ctrl+C on Windows or Linux, Cmd+C on macOS.

## Nucleotide sequence common statistics

The following common statistical information is calculated for a nucleotide sequence:

- *Length*
- *GC content*
- *Molecular weight*
- *Extinction coefficient*
- *Melting temperature*
- *nmole/OD260*
- *g/OD260*

### GC content

The percentage of guanine (G) and cytosine (C) bases within the sequence or its selected region, for example:

$$\text{GC-content("ACGTAC")} = ((0 + 1 + 1 + 0 + 0 + 1) / 6) * 100\% = 50\%$$

If the sequence contains degenerate base characters, average values are used, for example:

$$\text{GC-content("ACGNBCT")} = ((0 + 1 + 1 + 1/2 + 2/3 + 1 + 0) / 7) * 100\% \approx 59.52\%$$

In this example "1/2" is used for "N" (any nucleotide), "2/3" is used for "B" (that means "C", "G", or "T" according to the IUPAC nucleotide code).

### Molecular weight

*Molecular weight* for a single-stranded molecule is calculated as a sum of the atomic masses of the molecule compounds:

$$\text{DNA molecular weight} = nA * 251.24 + nT * 242.23 + nC * 227.22 + nG * 267.24 + (n-1) * 61.97$$

$$\text{RNA molecular weight} = nA * 267.24 + nU * 244.20 + nC * 243.22 + nG * 283.24 + (n-1) * 61.97$$

Here "nA", "nT", "nC", "nG", "nU" denote the number of the corresponding nucleotide in the molecule, "n" is the number of all bases (61.97 is the weight of an internal phosphate).

Note that for degenerate base characters average value of nucleotide weight is used, for example, if the sequence also contain "Y" characters (that is "C" or "T"), the sum will include one more summand - "nY\*(242.23 + 227.22)/2".

*Molecular weight* for a double-stranded molecule is calculated as the sum of the single strands molecular weights.

To calculate the *Extinction coefficient*, an approach proposed by Richard Owczarzy is used: <http://www.owczarzy.net/extinctionDNA.htm>. That is for a single-stranded molecule:

$$\text{Extinction coefficient} = \text{sum}(\text{extinction coefficients of all dinucleotides}) - \text{sum}(\text{extinction coefficients of inner mononucleotides})$$

For example, let's calculate the molar extinction coefficient (") for "ATGCA":

$$\begin{aligned} \text{(ATGCA)} &= (\text{AT}) + (\text{TG}) + (\text{GC}) + (\text{CA}) - (\text{T}) - (\text{G}) - (\text{C}) = \\ &= 22800 + 19000 + 17600 + 21200 - 8700 - 11500 - 7400 = \\ &= 53000 \end{aligned}$$

As for the other statistics, average values are used in case of degenerate base characters.

*Extinction coefficient* for a double-stranded molecule is calculated as a sum of the extinction coefficients of the two single strands ( $e_{s1} + e_{s2}$ ) multiplied by coefficient of (1 - hypochromicity  $h_{260nm}$ ). The hypochromicity effect can be taken into account as follows:

$$h_{260nm} = (0.287f_{AT} + 0.059f_{GC})$$

where  $f_{AT}$  and  $f_{GC}$  are fractions of AT and GC base pairs, respectively.

## Melting temperature

The melting temperature is calculated as follows. For sequences of length 15 or longer:

$$T_m = 64.9 + 41 * (nG + nC - 16.4) / (nA + nT + nG + nC)$$

For shorter sequences:

$$T_m = (nA + nT) * 2 + (nG + nC) * 4$$

Here "nA", "nT", "nC", "nG" denote the number of the corresponding nucleotide.

## nmole/OD260

The amount of DNA or RNA represented in nanomoles per 1 unit of absorbance at 260 nm dissolved in 1 ml cuvette with 1 cm pathlength.

$$nmole/OD260 = 1000000 / molarExtCoef$$

## g/OD<sub>260</sub>

The amount of DNA or RNA represented in microgrames per 1 unit of absorbance at 260 nm dissolved in 1 ml cuvette with 1 cm pathlength.

$$g/OD260 = nmoleOD260 * molarWeight * 0.001$$

## Amino acid sequence common statistics

The following common statistical information is calculated for an amino acid sequence:

- *Length*
- *Molecular weight*
- *Isoelectric point*