



Element V2023 command line tool

[Introduction](#)

[Installation instructions](#)

[Launching Element](#)

[Configuration settings](#)

Introduction

Element is an interactive tool (implemented in Perl) for rapid and convenient detection of core promoter elements within a given sequence. The basic algorithm and verifications are described in [this paper](#).

The table below summarizes the analyzed elements and their parameters.

The cutoff is the only parameter that can be controlled by the user.

| element | default cutoff | offset relative to TSS | Length (bp) | Reference for PWM |
|---------|----------------|------------------------|-------------|---------------------------------|
| GAGA | 6.7 | -80 | 10 | CISBP motif M5247, version 1.02 |
| BREu | 5.8 | -37 | 7 | PMID: 26226151 |
| TATA | 6.0 | -30 | 8 | PMID: 26226151 |
| BREd | 4.2 | -24 | 7 | PMID: 26226151 |
| XCPE1 | 4.0 | -8 | 10 | PMID: 26226151 |
| XCPE2 | 6.0 | -9 | 11 | PMID: 26226151 |
| Motif 1 | 11.0 | -7 | 11 | PMID: 12537576 |
| dTCT | 9.5 | -2 | 8 | PMID: 26226151 |
| hTCT | 6.0 | -1 | 7 | PMID: 26226151 |
| BBCABW | 3.6 | -3 | 6 | PMID: 28108474 |
| hInr | 0.7 | -2 | 7 | PMID: 26226151 |
| dInr | 2.3 | -2 | 6 | PMID: 26226151 |
| MTE | 8.5 | Inr+19 | 12 | PMID: 26226151 |
| bridge | 0.78 | Inr+19 | 5+4 | PMID: 26226151 |
| DPE | 0.7 | Inr+29 | 6 | PMID: 26226151 |
| PB | 7.6 | 25 | 7 | PMID: 18505835 |



Installation instructions

Install ElemeNT by extracting the ElemeNT_V2023.zip file into your chosen destination folder.

Content of the zip package:

- Software executables:
ElemeNT_V2023.exe - for Windows OS (built under Windows 64bit).
After extracting the zip file, add .exe extension to the *ElemeNT_V2023.exe* file (currently without any extension), *i.e.*, rename the file *ElemeNT_V2023* *ElemeNT_V2023.exe*
ElemeNT_V2023_binary - for Unix OS (built under Red Hat Enterprise Linux Server release 7.9 (Maipo))
To allow execution, the file should be granted with execution permission.
Use `sed -i 's/\r//g' Sample_input.txt` to replace between windows/UNIX end-of-line characters.
- The *Elements* folder, which contains the elements' position weight matrices (PWM) files. The nucleotide order in the PWM files (top to bottom) is A C G T.
Note – if needed, each PWM file's contents can be modified to search for position weight matrices other than the defaults.
- Config.txt file- contains the parameter default values for the ElemeNT run.
The file name should not be changed. The file content is case sensitive.
Make sure you leave one blank space after each ":" character in the config.txt file.
See "Configuration Settings" section for detailed explanations.
- Sample_input.txt is an example of input file from which the sequence/s will be read

Launching ElemeNT

After setting the parameters in the config.txt, open a command prompt, change the path to the installation directory and run the relevant executable (*i.e.*, *ElemeNT_V2023.exe* for windows or *ElemeNT_V2023_binary* for Linux). The executable will read the parameters for the run from Config.txt.

Output of the run:

For every run, a log file will be created. The log will include the input and output file names, parameters and cutoffs that were set for the run. The numbers and names of the processed sequences will be printed at the end of the file.

The output file, containing the information of the elements that were found, will be created in the same folder.

A match will be printed only if the motif's score is above the specified cutoff AND the position is within the allowed range, calculated using the formula: $\text{start} + \text{offset} \pm \text{smooth}$.



Configuration settings

InputFileName (txt file name)

This parameter points to the input file that contains the sequence(s) to be analyzed (default value is set to Sample_input.txt). The input sequences should be provided in FASTA format. The length of each individual sequence is limited to 1000bp.

If full path is **not** indicated, the input file should be located under the installation directory (at the same folder as the Element executable).

OutputFileName (txt file name)

This parameter determines the name of the Element output file. If full path is **not** indicated, the file will be created under the installation directory (file format is tab-delimited). This file will include information about the predicted elements on the input sequences.

If a putative element exists, its relative position and score are denoted (position, score). If more than one putative element exists, all the possible elements are indicated separated by a semicolon. **For MTE, DPE and Bridge elements the denoted position is of the associated Inr, as these elements are only functional at the precise position relative to the Inr.**

Note: Although it can result in redundant information, the MTE, DPE and Bridge elements will be printed for each initiator separately (dInr/hInr/BBCABW). Each match is followed by the relevant initiator in parentheses.

ReverseSeq (yes/no)

By default, the Element search is run on the input sequence (ReverseSeq set to 'no'). If the ReverseSeq is set to 'yes' then the Element search will be performed both on the provided input sequence and on its reverse complement strand.

When parameter is set to 'yes', the reverse complement sequences are labeled "revcom" and appear just below the original input sequence.

Start/Smooth constants (numerical values)

The parameter StartConstant is used to specify the expected TSS position within the provided sequence. The parameter SmoothConstant determine the allowed smooth window for TSS position. *i.e.*, Element will refer to TSS position as: $\text{StartConstant} \pm \text{SmoothConstant}$ and will search for putative elements around the TSS, based on their expected offset relative to it (the elements expected offsets relative to TSS are listed in the table included in the introduction section above).

Note that the start position of the predicted elements will be displayed relative to the specified TSS position (StartConstant). The "smooth" constant will be used to broaden the range of the allowed positions relative to the defined offset, listed in the table included in the introduction section. The range will be expanded in both directions by the value specified in SmoothConstant.

If no TSS data is available, it is recommended to use StartConstant = 0 and SmoothConstant => sequence length. With this setting, Element search will be performed on input sequence as if TSS can be located at any position within the sequence.



ChrData (yes/no)

Specify whether to extract location data from the input FASTA line.

Possible parameters (case sensitive)- yes/no.

“no” will display the names extracted from the FASTA lines (all characters after the ‘>’ until a space character).

“yes” will display the chromosome, start and end positions, name and strand. Choosing this option requires a specific FASTA line structure (adjusted for UCSC output lines). For example:

>**AnyText(NAME)** range=**chr1:120905983-120906083** 5’pad=0 3’pad=0 **strand=+** repeatMasking=none

The bold data is the important structure. Note that the space after the name is mandatory for the detection of the chromosomal location data.

Order

Specifies which elements you want to check. The current list includes all the available elements. Change this line **only** if you wish to search for a subset of the elements. hlnr and dlnc elements must appear first in order for the DPE, MTE and bridge elements to be printed. Changing the order of the elements may result in an unexpected behavior.

Note: the DPE, MTE, and Bridge elements are treated as "lnr-dependent", *i.e.*, these motifs are only scored at the precise location relative to each detected initiator, based on the known strict spacing requirement.

Therefore, an initiator within the order parameter will (by default) search for these motifs as well. **The MTE, DPE and Bridge elements can be managed by the cutoff score parameters** - to omit an element from the analysis, you should remove the relevant cutoff line (importantly, both of the Bridge cutoffs (*i.e.*, bridge1Cutoff & bridge2Cutoff) should be removed in order to omit it).

GC (1 < numeric value < 100)

The percentage of the GC-content as defined by the use. The motif search will be normalized to the provided GC-content.

Cutoffs

Specify the cutoff for each element separately. There is no need to erase the cutoff for elements that are not specified in the “order” line.

Support contact:

Orit Adato (orit.adato@biu.ac.il)

Prof. Tamar Juven-Gershon (Tamar.Gershon@biu.ac.il)