Glossary

Input sequences: the sequences provided for the analysis by user.

Structural template: Set of dinucleotide structural parameters for the DNA from the nucleosome core particle structure used by the program as a template for the calculation of the threading scores.

DNA deformation energy: the energy score that characterizes the easiness with which the structure of the threading template can be imposed on a given input sequence.

Nucleosome positioning score: the statistical score that characterizes the significance of the deviation of the energy score at a given position of the template on the sequence from the mean score for the neighboring positions.

Random sequences: if requested, for each input sequence a set of the random sequences of the same dinucleotide composition is generated to enable calculation of the expected values of the energy and positing scores.

Options and controls

A view of the *nuScore* interface is show in Figure 1 and options and controls are explained below.

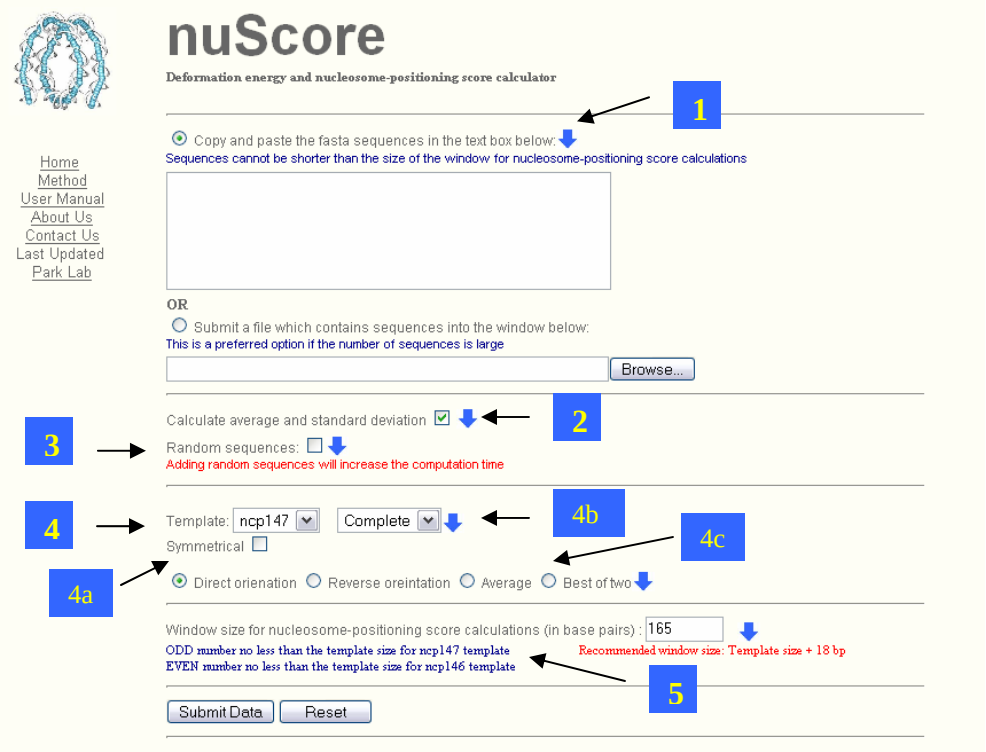


Figure 1. A snapshot of the *nuScore* interface.

1. **Copy-and-paste DNA sequence(s) or download sequence file**
   The sequences are expected to be in plain text format. If multiple sequences are provided they should be placed each on a separate line. Lines starting with ">" are ignored by the program. The program would recognize A, C, G, T letters in lower or upper case as Adenines, Cytosines, Guanines, and Thymines respectively; all other symbols, including blanks, will be treated as unknown nucleotides in the sequence, N and treated as "mixed-sequence" DNA (see Methods for detail).

2. **Average and standard deviation**
   If this option is selected, the program will calculate position-wise averages and standard deviations for all the sequences provided. This option should be selected if aligned sequences are provided. The default value is "selected".

3. **Adding random sequences to the analysis**
   If this option is selected, for each of the sequences provided for the analysis, the program will generate a specified number of random sequences of the same content. The DNA deformation energy and nucleosome positioning scores for the random sequences are reported along with the scores for the input sequences and can be used for reference. Usually more than 10 random sequences are required to produce statistically sound results. The default value is "not selected".

4. **Template selection**
   Currently two structural templates are available. The first is based on the NCP147 structure (PDB_ID 1kx5) and the second one is based on the NCP146 structure (PDB_ID 2cv5).
   4a. Using symmetrical template, i.e. the template with the structural parameters relative to the dyad. E.g. for the NCP147 template $Twist(Step_1) = Twist(Step_{146})$, $Twist(Step_2) = Twist(Step_{145})$, etc (see Methods for detail).
   4b. Using partial template. If this option is selected, only a fragment of the whole template will be used for the threading calculations. The fragment is located symmetrical relative to the dyad of the corresponding template and therefore, the lengths of the partial template should be odd for the 147-bp NCP147 structure and even for the 146-bp NCP146 structure. The default values for partial template is 129 and 128 bp for NCP147 and NCP146 structures respectively.
   4c. Template orientation relative to sequence. 'Direct' orientation corresponds to the orientation of the structural template as it appears in published pdb-file. 'Reverse' orientation corresponds to the template rotated by 180 degrees relative to its structural dyad (see Methods for detail). If options 'Average' or 'best of two' are selected, the average or the lowest energy score for the direct and reverse template orientations will be reported.

5. **Window size for the nucleosome positioning score calculations**
   This number specifies how many neighboring positions of the template on the sequence will be used to calculate the nucleosome positioning score for a given position (see Methods for detail). The default value is determined as template size + 18 bp.

Output files

All output files can be viewed as tables or downloaded in plain-text format.

DNA deformation energy file. The file contains the energy score values for each possible position of the threading template on the DNA sequence. The score is assigned to the position of the nucleosome dyad for each template setting on an input sequence. The numbering of the positions on an input sequence starts from the first base pair.

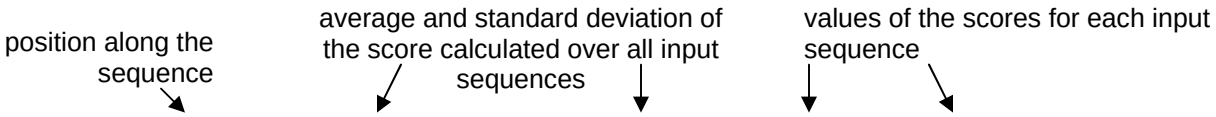
The energy score profile is expected to correlate with the overall nucleosome occupancy measured experimentally (see Supplementary Material).

Nucleosome positioning score file. The file contains the positioning score values calculated in a window of specified size for each possible position of the threading template on the DNA sequence. The score is assigned to the dyad position for each template setting. For example, for the NCP147 template (147 bp), the recommended window size is 165 bp. There are 19 possible positions of the template within the window and, thus, the energy score at the central position can be evaluated against the 18 neighboring score values. Then, the statistical score is reported as nucleosome positioning score for the central position of the window. In such a way the energy and positioning scores will be reported for all positions in the sequence starting from position 83.

The nucleosome positioning score is designed to predict the nucleosome attracting and nucleosome repelling sites and should be compared with the experimental positions of nucleosomes mapped to base pair resolution (see Supplementary material).

GC-content file. Additionally, GC-content is calculated in a sliding window of the same size as the threading template used, and is reported in a separate file.

If selected, the **average and standard deviations** of the scores will be calculated position-wise over all the input sequences. These values are reported in the first two columns of the



Position	Average	Standard Deviation	Sequence 1	Sequence 2
83	-0.427	1.839	-1.728	0.873
84	-0.781	0.266	-0.969	-0.593
…	…	…	…	…

Figure 1. Example of the output file for two input sequences. (Nucleosome-positioning score table is shown; DNA deformation energy and GC-content tables are organized in a similar way)

Results for random sequences. If selected, for each of the input sequences a set of random sequences of the same dinucleotide composition will be generated. User can specify the number of the random sequences in the sets to be generated. Then, the position-wise mean and standard deviations of the energy and positioning scores will be calculated for each set of the random sequences. These values are reported in the corresponding output file. Additionally, average and standard deviations of the mean scores over random sequences will be reported in the first two columns, if the option “calculate average and standard deviation” is selected.

position along the sequence

average and standard deviation of the mean scores calculated for each set of random sequences

mean scores and corresponding standard deviations calculated over all random sequences in the sets generated for each input sequence

Position	Average	Standard Deviation	Sequence 1		Sequence 2	
			mean	r.m.s.d.	mean	r.m.s.d.
83	893.535	32.526	916.534	27.431	870.536	32.934
84	906.363	60.701	949.285	38.462	863.441	48.988
…	…	…	…	…	…	…

Figure 2. Example of the output table of DNA deformation energy score for random sequences when two input sequences were uploaded.

Output graphs

If only one sequence is provided or the option “Calculate average and standard deviation” is selected, the graphs representing DNA deformation energy and nucleosome-positioning scores are shown as functions of the position of the template on the sequence. The user has an option to change the range of the positions included in the plot.

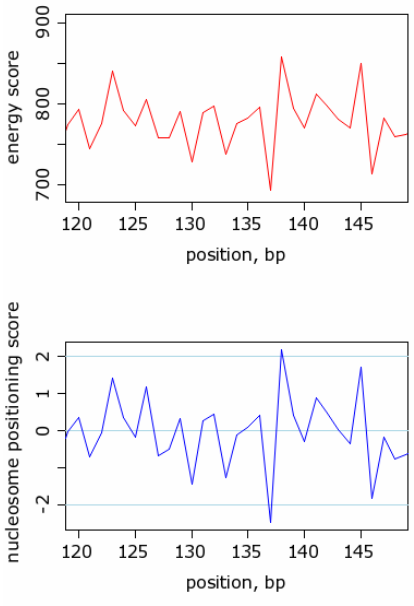


Figure 3. A snapshot of the output graphs (a fragment 601 positioning sequence is shown; the strongest dips in the energy and positioning score profiles correspond to the experimental nucleosome site with 3-bp accuracy.).