METHODS

The methodology used in this application is based on the estimation of the energy cost of the structural deformation imposed on DNA in the nucleosome core particle [1]. Such a cost depends on the DNA sequence [2], and since the direct sequence-specific interactions between the histones and DNA bases are essentially absent, the cost of DNA deformation is one of the main factors that determine the affinity of the histone core to DNA.

The cost of DNA deformations is estimated by 'threading' the DNA fragment of a given sequence on the DNA trajectory observed in the structure of a nucleosome core-particle solved experimentally. This approach has been previously applied to the analysis of nucleosome-positioning sequences and a good correspondence between the locations of the minima of the deformation energy score and experimentally mapped nucleosome positions was demonstrated [1].

Energy score calculations

The DNA structure is described at the dinucleotide level, i.e., we use a set of six variables that specify the relative positioning of neighboring base pairs [3] (Figure 1).

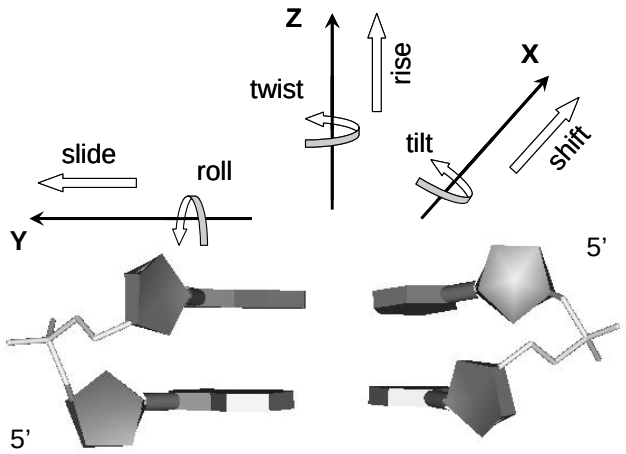

Figure 1. Definition of three translational and three rotational parameters of DNA structure that describe relative orientation of the neighboring base pairs [3]. The minor groove view of a dinucleotide is shown. The x-axis points away from the viewer.

The deformation energy score E of the threaded sequence can be expressed as [2]:

$$E = \sum_{n=1}^{L}\left(\frac{1}{2}\sum_{i=1}^{6}\sum_{j=1}^{6} f_{ij}(\mathrm{MN})\Delta\theta_i^n\Delta\theta_j^n\right), \qquad (1)$$

where $\Delta\theta_i^n = \theta_i^n - \theta_i^0(\mathrm{MN})$ is the imposed deviation of the i-th dinucleotide step parameter θ_i^n at the n-th step of the template from the rest-state value $\theta_i^0(\mathrm{MN})$ of the step MN. The $f_{ij}(\mathrm{MN})$ are DNA stiffness constants that depend on the sequence, and L is the number of base-pair steps in the nucleosome template. The rest-state values and stiffness constants of the 16 dinucleotide steps constitute the DNA flexibility model and were estimated from the analysis of a set of protein-DNA complexes [2]. This approach accounts for the known correlations of the base-pair step parameters [2]. Other models of DNA flexibility, based on the analysis of other sets of the complexes [4] or the molecular dynamics trajectories of DNA [5], can easily be incorporated.

The imposed values of the dinucleotide parameters in Eq. (1) are determined by the structural template used for threading. The program *nuScore* allows selection from eight templates (Table 1).

The templates are based on the structures of the nucleosome core particles solved by crystallography to the resolution 3Å or better. The selected structures contain non-mutant histones from xenopus, chicken, mouse, and human and do not include additional ligands interacting with the histones or DNA. All the structures contain major histone variants except for one structure that comprises the H2A.Z variant of the H2A histone (PDB_ID 1f66). The structural parameters of the nucleosomal DNA, which comprise the templates, were calculated with the program 3DNA [6].

Table 1. Nucleosome core particles used as structural templates

Structure ID	Descriptor	Resolution (Å)	reference
1KX5	NCP147, histones from xenopus	1.9	[7]
1KX3	NCP146, histones from xenopus	2.0	[7]
1KX4	NCP146, histones from xenopus	2.6	[7]
2CV5	NCP146, histones from human	2.5	[8]
1EQZ	NCP146, histones form chicken	2.5	[9]
2NZD	NCP145, histones from xenopus	2.7	[10]
2PYO	NCP147, histones from drosophila	2.4	[11]
1F66	NCP146, histones H3,H4, H2B from xenopus, H2A.Z histone variant from mouse	2.6	[12]

It is possible to use complete or partial templates. In the case of partial template selection, the central fragment of a complete template, which has the specified length, will be used. For example, if 129-bp partial template is selected in the case of NCP147 template, then the fragment from base pair 10 to base pair 138 will be taken for threading. The most pronounced deformations of nucleosomal DNA in NCP147 occur at central 129 bp [13], and 129 bp is used as default size for the partial template (128 bp for NCP146 and 127 bp for NCP145).

Since the threading template is not symmetrical, there are two possible orientations of the template on the sequence. Direct template orientation corresponds to the structure that is reported in the pdb-file. Reverse template orientation corresponds to the template that comprise the dinucleotide structural parameters in the reverse order relative to the original structure, so that $Twist_{reverse}(Step_1) = Twist_{direct}(Step_L)$, $Twist_{reverse}(Step_2) = Twist_{direct}(Step_{L-1})$, etc. Note that the values of Tilt and Shift also change sign in this procedure, so that $Tilt_{reverse}(1) = -Tilt_{direct}(L)$, etc (Figure 2).

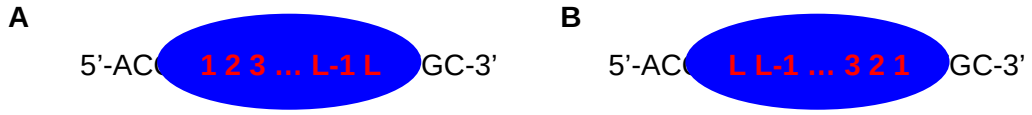

Figure 2. Schematic illustration of the direct (A) and reverse (B) orientations of the structural template (blue oval), consisting of L base-pair steps, on the input sequence.

Additionally user has options to have the average or the lowest energy score for the direct and reverse template orientations reported in the output file. Finally, symmetrical template can be selected. Such a template comprises parameters that were symmetrized relative to the structural dyad, so that $Twist_{sym}(Step_1) = Twist_{sym}(Step_L) = (Twist_{direct}(Step_1) + Twist_{direct}(Step_L))/2$, etc.

Finally, for the sequence steps in the input sequence that contain any symbol other than A, C, G, or T a 'mixed-sequence' energy score [1], which is equal to the mean score on a particular template step for all 16 dinucleotides, will be used in Eq. 1.

Nucleosome-positioning score calculations

A well positioning DNA sequence is expected to show a sharp dip in the energy score for the preferred template setting. The positioning score *P* can be defined by the relative deviation of the energy score calculated for a given position *x* of the template on the sequence with respect to the scores calculated for *n* neighboring positions of the template (Figure 3):

$$P = \left(E(x) - \langle E \rangle_{\pm n/2}\right) \Big/ \sigma_{\pm n/2} \qquad (2)$$

Large negative values of *P* denote nucleosome-attracting sequences (expected to favor nucleosomes) and large positive values point to nucleosome-refractory sequences (expected to repel nucleosomes). Small values of $\sigma_{\pm n/2}$ may result in false positive predictions. To avoid this situation, the condition $|E(0) - \langle E \rangle_{\pm n/2})| \,/\, \langle E \rangle_{\pm n/2} \equiv \delta E > 0.1$ is used in addition to Eq. 2.

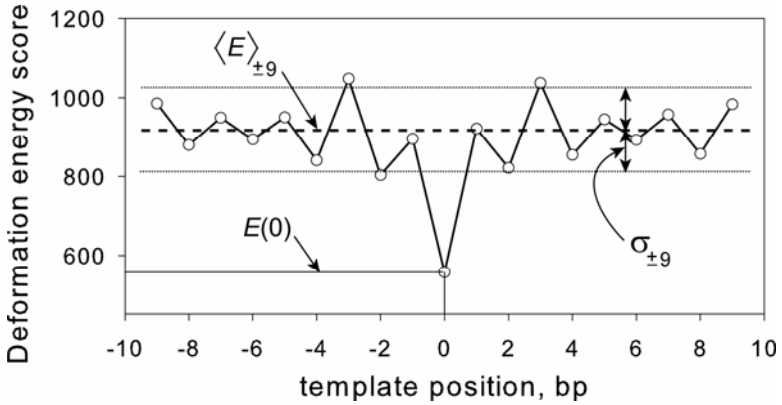

Figure 3. Calculation of the nucleosome-positioning score *P*. Here, the 147 bp sequence from the currently best resolved nucleosome structure (NCP147) [7] is threaded on the template comprised by the central 129 base pairs of this structure. The energy $E(x=0)$ at the position zero corresponds to the original setting with the nucleosome dyad positioned on the central base pair of the sequence. It is compared with the mean energy $\langle E \rangle_{\pm 9}$ of the 18 neighboring settings of the template on the sequence. The positioning score is the ratio of this energy difference and the standard deviation $\sigma_{\pm 9}$ and equals −5 in the case shown.

The user can select the size of the window used to calculate nucleosome-positioning score, *n*, which specifies how many neighboring positions are used to compute the score at a given position (Eq. 2). Considering 18 neighboring positions is a default option in *nuScore*; such a number is chosen because of the following considerations: using fewer neighbors may not be enough to produce statistically sound results and using more positions may interfere with known nucleosome-positioning signals, which often show 10-bp periodicities (e.g. AA:TT signal [14]).

Random sequences

A set of random sequences can be generated for each input sequence, if requested. The generated random sequences have the same dinucleotide composition (equal fractions of 16 possible dimers) as the initial input sequence. The number of sequences in the random set is specified by the user. The deformation energy and positioning scores are calculated for the random sequences and corresponding mean and standard deviation values are calculated over all the sequences in each set.

References

1. Tolstorukov MY, Colasanti AV, McCandlish DM, Olson WK, Zhurkin VB. A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *J. Mol. Biol.*, 2007, **371**, 725-738
2. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 11163–11168.
3. Dickerson RE, Bansal M, Calladine CR, Diekmann S, Hunter WN, Kennard O, von Kitzing E, Lavery R, Nelson HCM, Olson WK, Saenger W, Shakked Z, Sklenar H, Soumpasis DM, Tung C-S, Wang AH-J, Zhurkin VB. Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.*, 1989, **208**, 787–791.
4. Morozov AV, Havranek JJ, Baker D, Siggia ED. Protein-DNA binding specificity predictions with structural models. *Nucleic Acids Res.*, 2005, **33**, 5781-5798.
5. Lankas F, Sponer J, Langowski J, Cheatham III TE. DNA Base-pair step deformability inferred from molecular dynamics simulations. *Biophys. J.*, 2003, **85**, 2872-2883.
6. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, 2003, **31**:5108-5121.
7. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9 A Resolution. *J.Mol.Biol.*, 2002, **319**, 1097-1113.
8. Tsunaka Y, Kajimura N, Tate S, Morikawa K. Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. *Nucleic Acids Res.*, 2005, **33**, 3424-3434.
9. Harp JM, Hanson BL, Timm DE, Bunick GJ. Asymmetries in the nucleosome core particle at 2.5 A resolution. *Acta Crystallogr.*, Sect. D, 2000, **56**, 1513-1534.
10. Ong MS, Richmond TJ, Davey CA. DNA stretching and extreme kinking in the nucleosome core. *J.Mol.Biol.*, 2007, **368**, 1067-1074.
11. Clapier CR, Chakravarthy S, Petosa C, Fernandez-Tornero C, Luger K, Muller CW. Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. *Proteins*, 2007, **71**, 1-7.
12. Suto RK, Clarkson MJ, Tremethick DJ, Luger K. Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat.Struct.Biol.*, 2000, **7**, 1121-1124.
13. Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. *Nature*, 2003, **423**, 145-150.
14. Trifonov EN. Sequence-dependent deformational anisotropy of chromatin DNA. *Nucleic Acids Res.*, 1980, **8**, 4041–4053.