1 Illustration of the ThreaDNA implementation

Figure S1: Illustration of ThreaDNA program implementation. (top) Graphical user interface. (bottom) Visualisation of results in the UCSC genome browser (microbes.ucsc.edu) [1]. Caution: the employed fasta sequence file must have the same chromosome name (usually chr) as in the browser, otherwise the import is refused by the server.
2 Nucleosome positioning

Figure S2: Correlation plots for prediction of nucleosome occupation patterns along the *Saccharomyces cerevisiae* genome, with the ABC parameter set for DNA flexibility. The results are comparable to those obtained with the NP parameter set (Fig. 2) but the Pearson correlation coefficients are lower: 0.38 (in vitro) and 0.22 (in vivo), against 0.58 and 0.38 with NP. The effective temperature parameter used was slightly lower (0.25 instead of 0.35 for NP), and both occupancy patterns were computed with a nucleosome occupation size of 147 basepairs. Note that for genomic densities of nucleosomes, the employed algorithm (Boltzmann inversion) is not precise, since it neglects the interaction between adjacent nucleosomes.

Figure S3: Prediction of wrapping energies for high-affinity sequences, based on different structural templates. Each structural template predicts a low energy for its particular sequence, and overestimates that of other high-affinity sequences (same legend as in Fig. 2). The combination of structures used by ThreaDNA incorporates the plasticity of the complex, and successfully predicts the low energy of all sequences showed.
3 Predictions of Fis-DNA affinities

Figure S4: Comparison of Fis deformation energies on different crystallisation sequences, obtained with all existing structural templates. For each sequence, we expect the associated structure to be associated to a weaker interaction energy than other conformations, and we use this feature to test ThreaDNA's ability to compare alternative structural models. Each cell of the table contains the position of minimal deformation energy, compared to the experimentally observed position (0), for the considered sequence (row), based on the considered structural template (column). While some structures often miss the crystallisation position observed on other sequences, all of them correctly predict their own favourite binding position. ThreaDNA thus captures at least qualitatively the relation between sequence and structure in this experiment.

Note that here the positioning of the protein at the centre of the DNA oligomer might be also favoured by crystal forces, which are not taken into account in the model. The prediction above was that each sequence must favour (blue colour) the binding in the conformation observed experimentally (diagonal of the table), versus alternate structures. This indeed happens in most rows, while most exceptions (SJRB, 3JRD, SJRF, 3JRH) are lower-affinity complexes with also lower resolution. This example clearly shows that using a single structure in ThreaDNA, say 3IV5, results in overestimating the energy associated to sequences such as 4IHV, whereas using the whole combination partially solves these problems. This may explain why the predicted affinities of these sequences (Fig. S5) correlate better with the experimental affinities based on the combination ($r^2 = 0.44$) than on 3IV5 only ($r^2 = 0.30$).
Figure S5: Correlation of experimental vs predicted affinities between Fis and the set of sequences used in crystallisation experiments [2]. (A) The predictions are based on a combination of direct readout (PWM obtained from the list of RegulonDB sites) and indirect readout (ThreaDNA). The best result includes 30% direct readout. Correlation p-values of the PWM, ThreaDNA and combined predictions are $P = \times 10^{-2}$, $P = \times 10^{-3}$ and $P = \times 10^{-4}$ respectively. (B) Correlation plot. Note that the PWM lower correlation value is mostly due to two sequences (behind the legend), which bind strongly despite missing G/C bases at positions ±7, whose presence seems overweighted in the PWM.

4 Robustness of CRP affinity prediction with position weight-matrices

Figure S6: Test of robustness of binding predictions from position weight matrices. The correlation between predicted and observed affinities (for the Lindemose et al. data, Fig. 5) was computed with randomly generated subsets of CRP binding sites of increasing size (x-axis), based on the whole dataset of RegulonDB. The predictive power of the PWM is variable for low numbers of experimental binding sites, and becomes stable for around 100-150 sites.

References
