Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2022

Supplementary information: Predicting the configuration and energy of DNA in a nucleosome by coarse-grain modelling

Rasa Giniūnaitė,[1,2] and Daiva Petkevičiūtė-Gerlach,[1,*]

[1] *Department of Applied Mathematics, Kaunas University of Technology, Studentų 50-318, 51368, Kaunas, Lithuania*

[2] *Institute of Mathematics, Vilnius University, Naugarduko 24, 03225, Vilnius, Lithuania*

[*] *daiva.petkeviciute@ktu.lt*

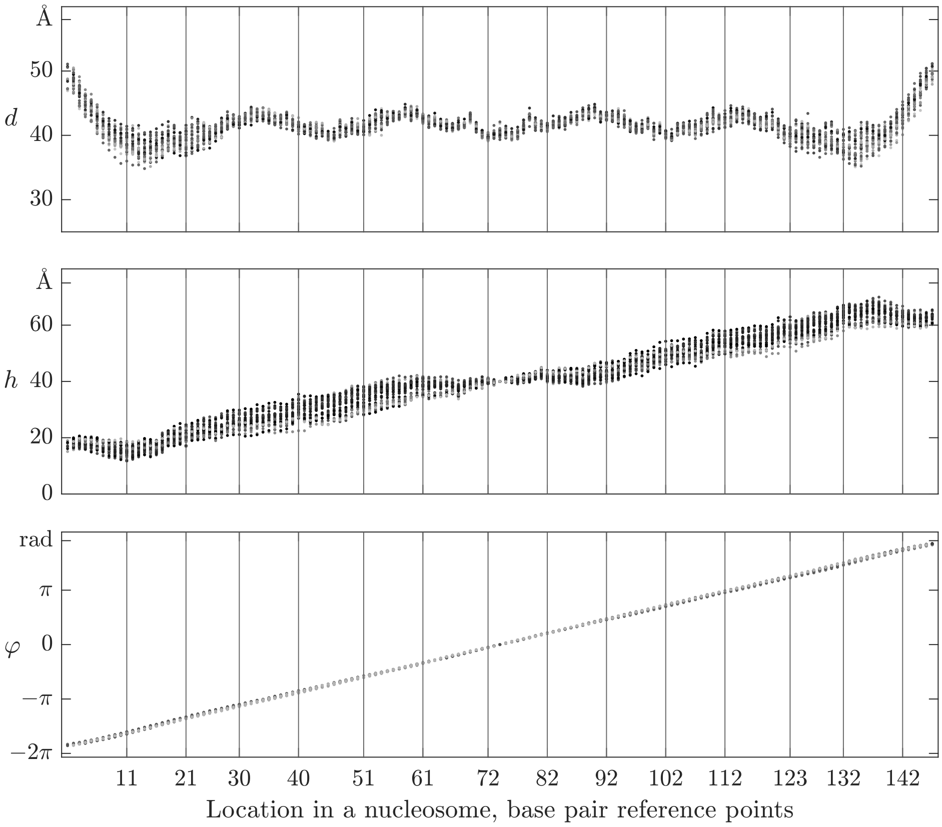

Fig. 1. Cylindrical coordinates of DNA base pair reference points, ordered from left to right along the Watson strands, for 30 experimental X-ray structures of nucleosomes. Coordinate d is the distance from a phosphate to the nucleosome central axis; h is the height, defined as the distance from a plane, perpendicular to the central axis and positioned 40Å away from the dyad point, and φ is the angle between a phosphate and the dyad axis, with respect to the central axis.

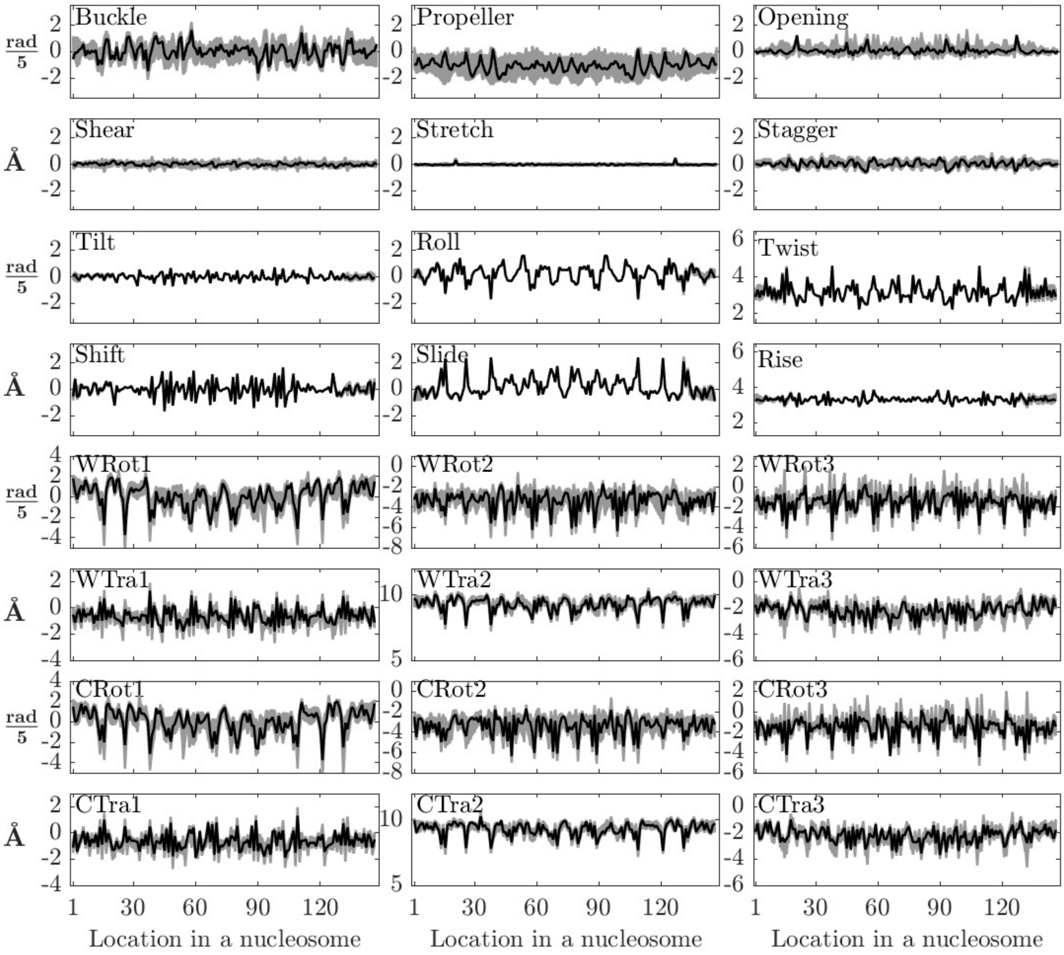

Fig. 2. Elements of the configuration vector w_{opt}, where w_{opt} is the sequence-dependent configuration of DNA on the nucleosome minimising the *cgDNA+* energy, as predicted by our optimisation procedure. Grey lines correspond to configurations predicted for 1000 random DNA sequences of length 147. Black lines correspond to the predicted configuration of the NCP147 sequence. WRot1, WRot2, WRot3 are rotational and WTra1, WTra2, WTra3 are translational coordinates for Watson strand phosphates. Coordinate values are interpolated by piece-wise linear curves for clarity. The spread of values across different sequences is smallest in the inter base pair coordinates, where the biggest variation is towards the ends of the DNA molecule.

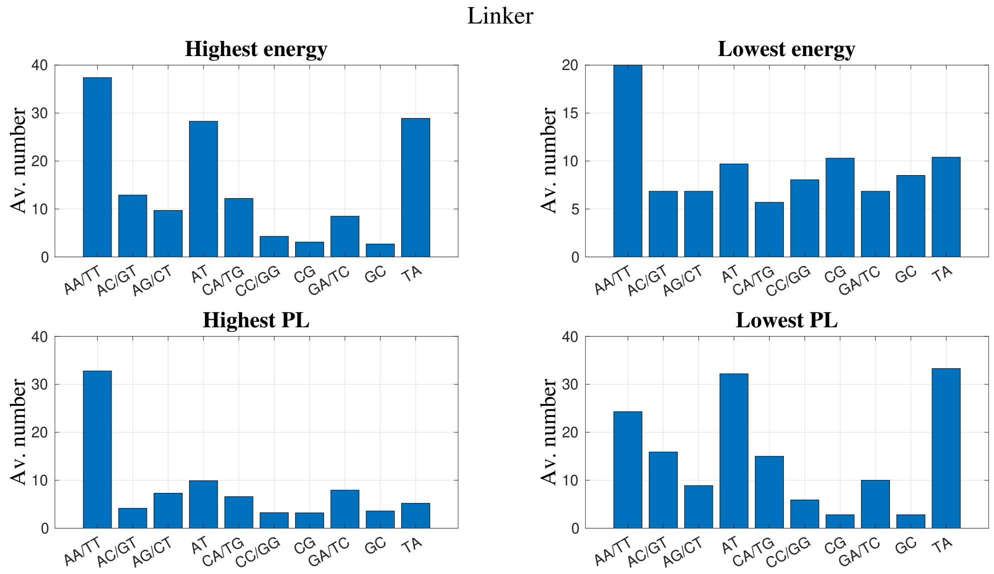

Fig. 3. Comparison of different number of dinucleotides in linker DNA, for the ten sequences with the highest and lowest energy and persistence length (out of 1740 linker sequences from Chen et al. (2016)[1]). We note that we aggregate the non-complementary dinucleotides since our model is centrosymmetric. In the non-aggregated version, the relative number of self-complementary dinucleotides would increase.

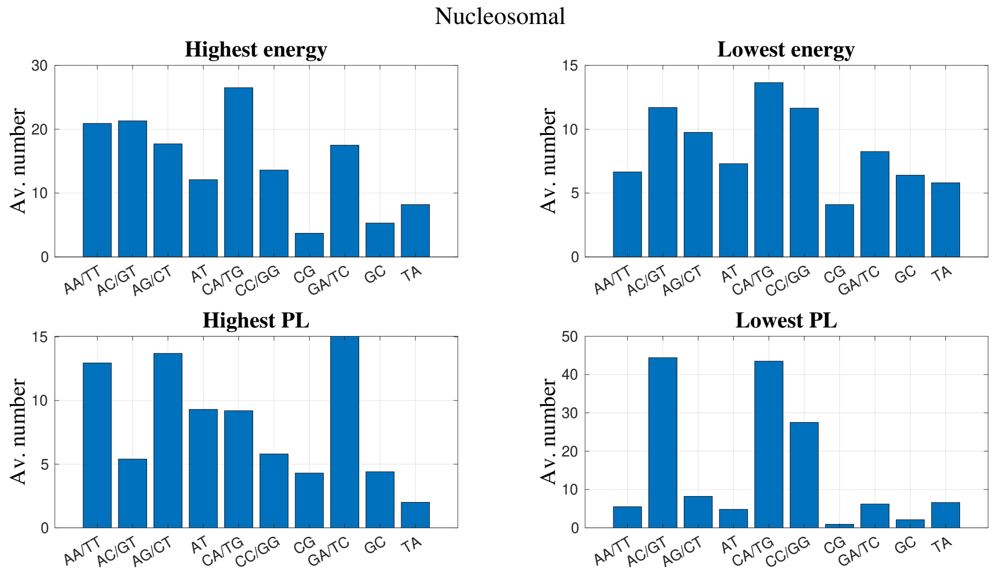

Fig. 4. Comparison of different number of dinucleotides in nucleosomal DNA, for the ten sequences with the highest and lowest energy and persistence length (out of 1880 nucleosomal sequences from Chen et al. (2016)[1]). There is no clear signal in which dinucleotides correspond to extreme values of energy and persistence length. We note that we aggregate the non-complementary dinucleotides since our model is centrosymmetric. In the non-aggregated version, the relative number of self-complementary dinucleotides would increase.

	AT		AA/TT	
Tract length	Energy	Persistence length	Energy Energy	Persistence length
2	0.34	-0.26	0.07	0.17
4	0.33	-0.63	-0.07	0.24
6	0.27	-0.55	-0.15	0.42
8	0.25	-0.52	-0.16	0.45
10	0.23	-0.49	-0.15	0.43

Table 1. Correlation coefficients of energy and persistence length versus the number of poly-AT and poly-AA/poly-TT motifs of various lengths for the linker sequences from Chen et al. (2016)[1].

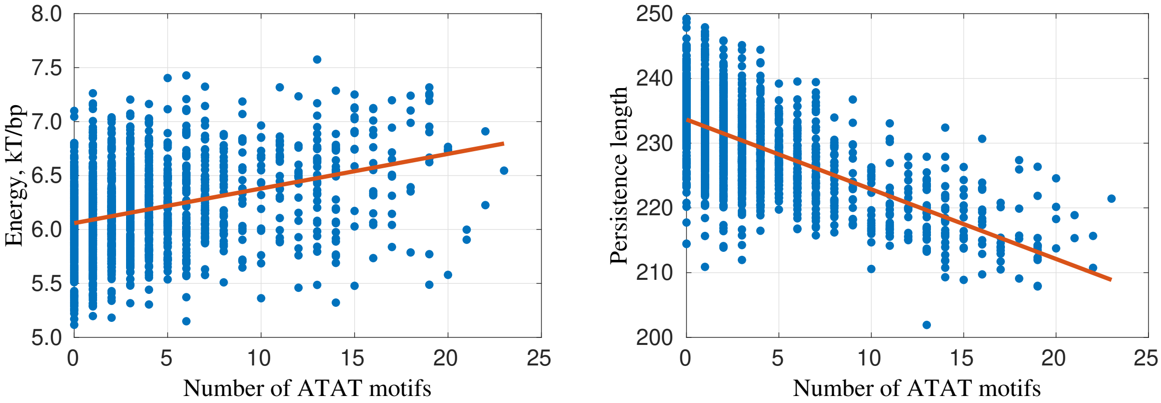

Fig. 5. Scatter plots of (a) energy and (b) persistence length of sequences with different numbers of ATAT tetranucleotides for the linker sequences from Chen et al. (2016)[1]. Red line corresponds to the fitted linear regression line (obtained using MATLAB *polyfit* function). The data suggests a weakly positive relationship between the energy and the number of ATAT tetranucleotides in a sequence. Whereas, there is a negative relationship between the persistence length and the number of ATAT tetranucleotides in a sequence.

	Structure for initial configuration		
Sequence	*1kx5*	*3lel*	*5f99*
NCP147 (as in *1kx5*)	3.86	6.12	8.82
Modified NCP147 (as in *3lel*)	4.25	5.74	8.48
147 bp MMTV-A (as in *5f99*)	5.44	8.15	6.64
Widom 601	5.26	8.13	8.39
Average over 1000 random sequences	5.62	8.63	8.79
Average over 1880 nucleosomal[1] sequences	5.68	8.72	8.83
Average over 1740 linker[1] sequences	6.18	9.12	9.26
Energy of the initial configuration	20.25	64.98	49.26

Table 2. DNA nucleosome wrapping energies in the units of kT/bp, obtained as local minima by our optimisation procedure (except for the bottom row). Rows correspond to input sequences, columns are different initial configurations. The sequence corresponding to the experimental structure used as the initial configuration has the lowest energy among other sequences for the same initial configuration. Averages of the energies of random, nucleosomal and linker sequences are of consistent ordering for different initial conditions. The bottom row contains energy values, corresponding to the initial configurations before the optimisation.

References

[1] W. Chen, P. Feng, H. Ding, H. Lin and K.-C. Chou, *Genomics*, 2016, **107**, 69–75.