

Table S1. Types of DNA fragments and their counts. Summary of types of sequences considered for all-atom Monte Carlo (MC) simulations. Most sequences were designed to cover different flanking sequences. Sequences between “-” symbols in column 1 are “core sequences”. Other sequences are regarded as flanks. “N” in designed sequences represents general DNA alphabet letters {A, C, G, T}. Methylated cytosine (“m”) and subsequent guanine (“g”) bases are underlined.

Fragments	Number of MC simulations	Selected from
Human HOXA9 binding sequences	3	[1]
Human HOXA5 binding sequences	84	[1]
CGNN-5mer-NNCG	1054	Designed
CGNN-NNNN <u>mg</u> -NNCG	1298	Designed
CGCG-NN <u>mg</u> NN-CGCG	253	Designed
CGCG-NNN <u>mg</u> N-CGCG	256	Designed
CGNN- <u>mg</u> NNN <u>mg</u> -NNCG	496	Designed
CGNN-poly[A/T] ₄ <u>mg</u> -NNCG	74	Designed
Total	3518	

[1] Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, et al. DNA methylation presents distinct binding sites for human transcription factors. eLife 2013;2:e00726.