

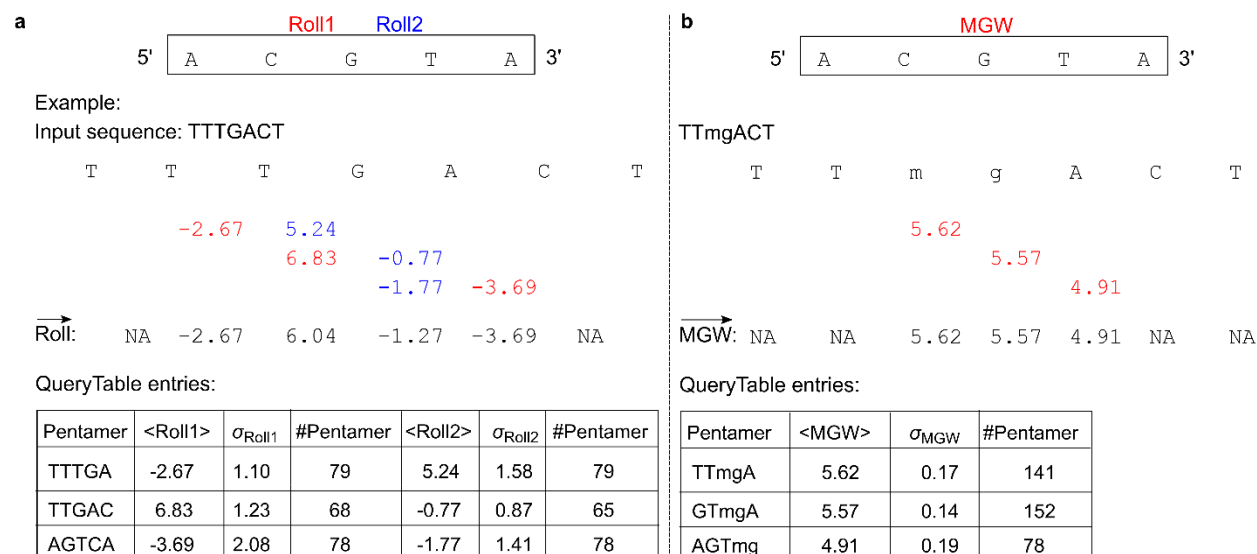


Fig. S1. Shape vector calculation. **a** Two Roll values, Roll1 and Roll2, were assigned to a given pentamer by using the query table for bp steps 2-3 and 3-4, respectively (illustrated at the top). The PQT lookup procedure is explained for calculation of the Roll feature vector for DNA sequence 5'-TTTGACT-3' as an example. Retrieval of the Roll feature vector for this sequence queries the lookup table three times for listed pentamers in the table. Because the third query pentamer, 5'-TGACT-3', finds its reverse complement 5'-AGTCA-3' in the table, the search resulted in the reversal of Roll1 and Roll2 values of 5'-AGTCA-3'. The same process was adopted for the base-pair step feature HelT. **b** Illustration of MGW feature vector calculation. The process is simplified in this case because the search returns a single value at the central bp for a given pentamer. MGW values for two flanking nucleotides are undefined because values at these positions cannot be calculated as per definition of minor groove.

< . >: average; σ : standard deviation.