

DNA structural alphabet opens ways to understand protein-DNA interactions

Bohdan Schneider¹, Paulína Božíková¹, Petr Cech², Daniel Svozil², Jiří Černý¹

¹Institute Of Biotechnology CAS, Vestec, Czech Republic, ²University of Chemistry and Technology, Prague, Czech Republic
E-mail: bohdan.schneider@gmail.com

We will present an easy-to-use and automated tool providing comprehensive description of DNA structures and demonstrate its usefulness for understanding of protein-DNA interactions. Our description of DNA structure is based on definitions of distinct dinucleotide conformers and their assembly into -to the best of our knowledge- the first structural alphabet of nucleic acids. The alphabet is called CANA (Conformational Alphabet of Nucleic Acids) and consists of thirteen letters for various DNA conformer classes. CANA describes the DNA structural variability way beyond the traditionally used classification schema into A-form, B-form, and Z-form, but at the same time simplifies DNA structural description by transforming complex 3D structural information into the symbolic CANA alphabet. The CANA letters are assigned by an automated procedure available at the web sites dnatco.org [1] and dolbico.org to any PDB formatted structure. The assignment protocol translates 3D DNA structures into textual and/or graphical symbolic language of the CANA letters and streamlines thus human understanding of the complicated molecular structure. By using the CANA alphabet, we describe the structural variability of prototypical DNA structures such as Dickerson-Drew dodecamer and guanine quadruplexes, and critically discuss structural models based on the fiber diffraction. We address the challenge of understanding the specificity of protein-DNA interactions by analyzing CANA words typical for different types of protein-DNA complexes such as various types of transcription factors or histone core particle and compare the patterns of the CANA words in protein-bound and solvent-exposed DNA in these complexes. As an example of the results, Fourier transform of the occurrence of CANA letters of DNA in crystal structures of histone core reveals that one of the letters occurs every 10th step, i. e. every helical turn. Analysis of DNA structure based on the structural alphabet contributed thus to the understanding of the origin of DNA positioning around the histone core proteins.

[1] Černý J et al (2016) "DNATCO: assignment of DNA conformers at dnatco.org" *Nucleic Acids Research* 44, W284-W287.

Keywords: [protein-DNA structure](#), [structural alphabet](#), [histone positioning](#)