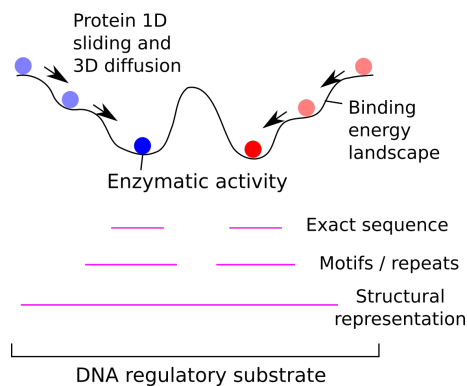


**Table 1: Overview of DNA structural properties and representative variables in protein-DNA interaction.**

DNA structural properties	Facilitated protein-DNA interactions	Representative structural variables	References
DNA stability and propensity for destabilizations and melting bubble formation	Enzymatic processing of substrates, e.g. relaxase nicking of transfer regions leads to secondary structure formation	Duplex stability, Thermally induced duplex destabilization (TIDD)	[11, 17, 19, 22]
Major and minor groove properties	Readout of chemical information, e.g. transcription factors in promoters	DNAShape, ORChID2	[2, 4, 16, 20]
Intrinsically curved or flexible regions	Binding and topological changes, e.g. IHF binding in promoters	DNAseI cleavage frequency, Persistence length	[3, 5, 13]
DNA twist and supercoiling	Topological changes recognized by proteins, e.g. histones, and affect multiple other properties	Twist and other conformational variables	[8, 14, 15, 20]
Differences in DNA spacing and orientation in binding and enzymatic sites	Affect binding with multiple contact points and protein complex formation	Helical repeats	[5, 20, 21]
Propensity for transitions between DNA forms B-DNA, A-DNA, Z-DNA	Affect overall features recognized by proteins and their accessibility	B-A and B-Z transition propensities	[1, 6, 7, 9]



**Figure 1: Possibilities for combined sequence structure algorithms based on how proteins recognize and bind their active sites in the regulatory DNA [10, 12, 16, 18]. Interactions of lower specificity with the surrounding DNA (corresponding to DNA with less conserved nucleotide sequence but defined structural properties) guide the proteins towards their specific binding sites (highly conserved or exact sequence).**

## REFERENCES

- [1] Aida M 1988 *Journal of Theoretical Biology* **130**(3), 327–335.
- [2] Bishop E P, Rohs R, Parker S C J, West S M, Liu P, Mann R S, Honig B and Tullius T D 2011 *ACS Chem. Biol.* **6**(12), 1314–1320.
- [3] Brukner I, Sanchez R, Suck D and Pongor S 1995 *EMBO J.* **14**(8), 1812–1818.
- [4] Chiu T P, Comoglio F, Zhou T, Yang L, Paro R and Rohs R 2016 *Bioinformatics* **32**(8), 1211–1213.
- [5] Geggier S and Vologodskii A 2010 *Proc. Natl. Acad. Sci. U. S. A.* **107**(35), 15421–15426.
- [6] Hartmann B, Malfroy B and Lavery R 1989 *J. Mol. Biol.* **207**(2), 433–444.
- [7] Ho P S, Ellison M J, Quigley G J and Rich A 1986 *EMBO J.* **5**(10), 2737–2744.
- [8] Karas H, Knüppel R, Schulz W, Sklenar H and Wingender E 1996 *Comput. Appl. Biosci.* **12**(5), 441–446.
- [9] Kulkarni M and Mukherjee A 2013 *J. Chem. Phys.* **139**(15), 155102.
- [10] Levo M, Zalckvar E, Sharon E, Dantas Machado A C, Kalma Y, Lotam-Pompan M, Weinberger A, Yakhini Z, Rohs R and Segal E 2015 *Genome Res.* **25**(7), 1018–1029.
- [11] Lucas M, González-Pérez B, Cabezas M, Moncalian G, Rivas G and de la Cruz F 2010 *J. Biol. Chem.* **285**(12), 8918–8926.
- [12] Marcovitz A and Levy Y 2013 *The Journal of Physical Chemistry B* **117**(42), 13005–13014.
- [13] Moncalian G, Valle M, Valpuesta J M and de la Cruz F 1999 *Molecular Microbiology* **31**(6), 1643–1652.
- [14] Olson W K, Gorin A A, Lu X J, Hock L M and Zhurkin V B 1998 *Proc. Natl. Acad. Sci. U. S. A.* **95**(19), 11163–11168.
- [15] Perez A, Noy A, Lankas F, Luque F J and Orozco M 2004 *Nucleic Acids Res.* **32**(20), 6144–6151.
- [16] Rohs R, West S M, Sosinsky A, Liu P, Mann R S and Honig B 2009 *Nature* **461**(7268), 1248–1253.
- [17] SantaLucia, Jr J 1998 *Proc. Natl. Acad. Sci. U. S. A.* **95**(4), 1460–1465.
- [18] Slattery M, Zhou T, Yang L, Dantas Machado A C, Gordán R and Rohs R 2014 *Trends Biochem. Sci.* **39**(9), 381–399.
- [19] Sut M V, Mihajlovic S, Lang S, Gruber C J and Zechner E L 2009 *J. Bacteriol.* **191**(22), 6888–6899.
- [20] Watson J D, Baker T A, Bell S P, Gann A, Levine M K and Losick R 2008 *Molecular Biology of the Gene. 6th. ed* Pearson/Benjamin Cummings.
- [21] Williams S L and Schildbach J F 2007 *J. Bacteriol.* **189**(10), 3813–3823.
- [22] Zrimec J and Lapanje A 2015 *IEEE/ACM Trans. Comput. Biol. Bioinform.* **12**(5), 1137–1145.