Structure Based Sequence Dependent Stiffness Scale for Trinucleotides: A Direct Method

M. MICHAEL GROMIHA

*Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN),
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan*

Abstract. A new set of stiffness parameters for all the 32 trinucleotide units has been set up directly from the three dimensional structures of DNA molecules. It was observed that GAC/GTC is the stiffest trinucleotide and ACC/GGT is the most flexible one. The average stiffness values computed for a set of operator sequences using the new parameters correlate very well with the protein-DNA binding specificity and binding free energy change of 434 repressor and Cro repressor, respectively. The new structure based stiffness scale can explain the protein-DNA binding specificity to the level of 0.92.

Key words: DNA structure, protein-DNA binding, sequence dependent, stiffness scale, Young's modulus

1. Introduction

The concept of sequence dependent DNA structure is important in packaging, transcription, recombination and DNA-replication [1, 2]. Flexibility and curvature of DNA are known to be sequence dependent [3, 4] and the impact of base sequence on the geometry of DNA double helix has been studied by Lavery and Hartmann [5]. Recently, Parbhane et al. [6] analyzed the sequence dependent DNA curvature using artificial neural networks.

Several methods have been proposed to understand the structure and stability of DNA [7–11] and the elastic rod model is one of the simplest models to describe DNA as a long, isotropically elastic rod under stretching, torsional and bending stress [12–14]. Recently, we have developed sequence dependent anisotropic bendability (SDAB) models of DNA in terms of DNA elasticity [15, 16] and applied the model successfully to understand the role of DNA bending in Cro protein-DNA interactions [17].

Hogan and Austin [18] proposed that the elastic properties of DNA may be of great importance to an understanding of protein-DNA binding specificity. Recently, several experiments have been carried out to measure the elasticity of DNA [19–21]. Although the average stiffness of entire DNA can be directly measured [19–21], sequence dependent stiffness values could be derived only by theoretical methods. Hogan and Austin [18] proposed a set of only 3 different stiffness values

for aa, gg and ag. Recently, we have developed a stiffness scale [15] using the flexibility data obtained from the cutting frequency of deoxyribonuclease I [22, 23]. In this work, a new sequence dependent stiffness scale was developed directly from the three dimensional structures of nucleic acids. The average stiffness values of DNA operator sequences obtained with the new scale correlates well with the DNA binding specificity of 434 repressor and the binding free energy change of Cro protein.

2. Materials and Methods

DATABASE

The crystallographic data of 70 DNA molecules form the source for the present study. The DNA molecules have been selected with the following criteria: (i) free DNA oligomers, (ii) no mismatches and (iii) no redundant molecules. The atomic coordinates of all the 70 molecules were taken from the recent release of the Protein Data Bank of Brookhaven National Laboratory [24]. A set of trinucleotides has been considered in the present study to develop the stiffness scale rather than di- or tetranucleotides due to two reasons: (i) the trinucleotide stiffness values can explain the properties of DNA molecules more accurately than that of dinucleotides [22] and (ii) the number of tetranucleotide samples in the database are not sufficient to develop the stiffness scale. The resulting dataset contains 450 trinucleotide samples with significant representations for most of the triplets.

DEVELOPMENT OF STIFFNESS SCALE

The nucleotides in a DNA molecule are represented by their 5' P-atoms similar to representing amino acid residues in protein molecules by their C_α atoms [25–28]. The centre is fixed at the P-atom of the first nucleotide and the distance between this atom (i) and the third atom ($i + 2$) is computed to determine the length of a trinucleotide. Considering a trinucleotide ACT, the distance between the P-atoms in A and T was computed, which represent the length of the trinucleotide, ACT. This procedure is repeated each time by moving the centre to the successive P-atoms along the nucleotide chain to compute the distance between i and $i + 2$ residues (first and last base of each trinucleotide) for all the nucleotides in a given chain. The length for all the trinucleotides in a set of 70 DNA molecules were computed and an average parameter set for the 32 trinucleotide units (l_1) was obtained.

DNA can be considered as an elastic rod with circular cross section, of radius r , and each segments consist of 3.4 Å thick discs corresponding to each base pairs [7, 15, 16, 29]. This yields a length of 6.8 Å (l) for each trinucleotide unit. The increase in length (d/l) for all the trinucleotides can be obtained from the difference between the actual lengths obtained from the crystal structures (l_1) and the normal length (l) known from physical data (6.8 Å).

The Young's modulus for all the trinucleotides can be computed using the classical formula,

$$E = \frac{F \cdot l}{A \cdot dl}, \quad (1)$$

where F is the maximum longitudinal force at which a DNA molecule can be elongated, and is estimated to be 480 pico-Newtons [30], A is the area of cross-section, given by, $A = \pi r^2$, where, r is the radius. The radius r is taken as 10 Å based on physical measurable size [29, 31]; l is the length of each trinucleotide (6.8 Å) and dl is the increase in length, given by, $dl = l_1 - l$, where l_1 is the length computed from the three dimensional structures of DNA molecules. The scale was normalized based on the observation that the average Young's modulus is 3.4×10^8 N m⁻² [20]. The computed Young's modulus values for all the 32 trinucleotides is called '*structure based stiffness scale*'. This normalized scale provides better insight to compare with other scales in literature [32–35].

COMPUTATION OF AVERAGE STIFFNESS VALUE AND CORRELATION COEFFICIENT

Each DNA sequences have been represented by overlapping segments of trinucleotide units. The stiffness values for each trinucleotide was assigned appropriately from the structure based stiffness scale and the average Young's modulus was computed using the equation

$$E = \Sigma E_i / n, \quad (2)$$

where E_i is the Young's modulus for the i^{th} trinucleotide and n is the total number of trinucleotide units.

The correlation coefficient between average stiffness (E) and protein-DNA binding specificity was computed using the familiar formula

$$r = \frac{N \Sigma XY - (\Sigma X \Sigma Y)}{\{[N \Sigma X^2 - (\Sigma X)^2][N \Sigma Y^2 - (\Sigma Y)^2]\}^{1/2}}, \quad (3)$$

where r is the correlation coefficient, N , X and Y , respectively, are the number of data, average stiffness value and experimental protein-DNA binding data.

3. Results and Discussions

STRUCTURE BASED STIFFNESS SCALE FOR TRINUCLEOTIDES

The sequence dependent stiffness parameters for all the 32 trinucleotide units, computed from three dimensional structures of DNA molecules along with their number of occurrences and standard deviations are given in Table I. From this table, we observe that the stiffest trinucleotide is GAC/GTC and the most flexible one is ACC/GGT. It is also observed that CCC/GGG is one among the topmost

Table I. Structure based DNA stiffness (Young's modulus) scale for trinucleotides

Trinucleotide	N_s	Young's modulus E (10^8 N m^{-2})	Trinucleotide	N_s	Young's modulus E (10^8 N m^{-2})
AAA/TTT	27	4.80 (0.10)	CAG/CTG	6	2.40 (0.03)
AAC/GTT	2	3.90 (0.06)	CCA/TGG	12	3.25 (0.17)
AAG/CTT	7	1.91 (0.09)	CCC/GGG	1	6.07 (0.00)
AAT/ATT	49	2.96 (0.07)	CCG/CGG	1	2.40 (0.00)
ACA/TGT	1	4.70 (0.00)	CGA/TCG	40	2.82 (0.06)
ACC/GGT	2	1.57 (0.01)	CGC/GCG	58	3.33 (0.12)
ACG/CGT	12	7.09 (0.32)	CTA/TAG	9	4.75 (0.22)
ACT/AGT	2	3.63 (0.06)	CTC/GAG	8	4.03 (0.26)
AGA/TCT	12	4.03 (0.13)	GAA/TTC	33	2.70 (0.05)
AGC/GCT	9	4.58 (0.23)	GAC/GTC	1	7.83 (0.00)
AGG/CCT	6	4.34 (0.15)	GCA/TGC	18	3.75 (0.07)
ATA/TAT	24	2.36 (0.08)	GCC/GGC	12	3.16 (0.09)
ATC/GAT	28	1.83 (0.06)	GGA/TCC	1	3.69 (0.00)
ATG/CAT	13	3.19 (0.16)	GTA/TAC	17	2.19 (0.07)
CAA/TTG	17	2.53 (0.05)	TAA/TTA	7	2.72 (0.02)
CAC/GTG	6	3.36 (0.08)	TCA/TGA	9	2.97 (0.07)

N_s – Number of samples in each trinucleotide; the deviations are given in brackets.

three stiffest trinucleotides. AAA/TTT is more flexible than CCC/GGG in the newly developed structure based stiffness scale which is the stiffest one in our previously derived scale [15] based on DNase I bendability [22, 23] and other scales, consensus bendability [32, 33] and nucleosomal bendability [34, 35]. There is no correlation between the structure based stiffness scale and previously derived scales.

In Figure 1, we show the comparison between structure based stiffness scale and other scales derived from, DNase bendability data [15, 22, 23], consensus bendability [32, 33] and nucleosomal bendability [34, 35]. Interestingly, several nucleotide units have different behavior in the present scale compared with other scales. A detailed analysis shows that AAG/CTT, ACC/GGT, ATC/GAT, CAA/TTG, CCG/CGG, GAA/TTC and GTA/TAC are more flexible in the present scale than all other three scales. ACG/CGT, CCC/GGG and GAC/GTC are the three trinucleotides that show higher stiffness values in the structure based stiffness scale than other scales. Further, several trinucleotides (AGA/TCT, AGG/CCT, ATG/CAT, CAC/GTG, CAG/CTG, CGA/TCG, CTC/GAG, GAC/GTC, GCA/TGC, GGA/TCC) have the similar behavior to that of other scales.

Figure 1. Comparison of different stiffness scales. The horizontal line indicates the average stiffness value ($3.4 \times 10^8 \text{ N m}^{-2}$). ●: structure based scale (present work); ○: DNase I scale [15, 22]; Δ: consensus scale [32, 33]; ◇: nucleosomal scale [34, 35].

Table II. Computed Young's modulus and experimental binding data for 434 and Cro repressors

No.	Code	Sequence	E (10^8 N m^{-2})	$\ln(k_1/k_2)^*$	ΔG^{**} (kcal mol^{-1})
434 repressor					
1	14	ATATATAT	2.36	0.00	
2	6T	ATTAAAT	2.83	0.40	
3	6C	ATCTAGAT	3.54	1.60	
4	6G	ATGTACAT	3.36	1.60	
5	6G/2	ATGTATAT	2.86	0.90	
6	7A	ATAATTAT	2.68	0.00	
7	7A/2	ATAAATAT	2.59	-1.20	
8	7C	ATACGTAT	3.88	3.90	
9	7G	ATAGCTAT	3.90	3.90	
10	7G/2	ATAGATAT	2.95	1.90	
Cro repressor					
1	OR1	ATATACCTCTGGCGGTGATATAT	2.71		-15.4
2	OR2	ACACAACACGCACGGTGTACAC	3.85		-14.1
3	OR3	ATATATCCCTTGCGGTGATATAT	2.70		-16.1
4	OL1	ATATACCACTGGCGGTGATATAT	2.61		-15.5
5	OL2	GTGTATCTCTGGCGGTGTTGTGT	3.33		-14.9
6	OL3	AAAAACCATCTGCGGTGATAAAA	3.26		-14.4

* $\ln(k_1/k_2)$ is the binding data [18, 36] for DNA binding specificity of 434 repressor.

** ΔG is the binding free energy of Cro repressor to operators [17, 37].

Figure 2a. Relationship between the average stiffness (Young's modulus) and protein-DNA binding specificity of 10 operator sequences with 434 repressor. The solid line shows the regression line and the correlation coefficient is 0.92.

Figure 2b. Relationship between the average stiffness (Young's modulus) and free energy change of 6 operator DNA sequences with Cro repressor. The solid line shows the regression line and the correlation coefficient is 0.90.

CORRELATION OF AVERAGE STIFFNESS WITH PROTEIN-DNA BINDING SPECIFICITY

Koudelka et al. [36] measured the changes in binding affinity of 434 repressor with a set of 10 DNA operator sequences. We computed the average stiffness values using Equation 2 for all the ten sequences with the structure based stiffness scale and the results are presented in Table II. In this table, the DNA sequences and the experimental affinity values are also included. Figure 2a shows the relationship between the average Young's modulus and experimentally measured protein-DNA binding data [18, 36]. We observed an excellent agreement and the correlation coefficient is 0.92. This observation is in agreement with Hogan and Austin [18] that the DNA stiffness is important for protein-DNA binding specificity.

RELATIONSHIP BETWEEN AVERAGE STIFFNESS AND BINDING FREE ENERGY CHANGE OF CRO REPRESSOR

Takeda et al. [37] measured the free energy change (ΔG) due to binding of Cro repressor to the six naturally occurring operators. The sequences for all the 6 operators and their ΔG values are given in Table II. We computed the average stiffness values for all the sequences and the values are included in Table II. The relationship between average stiffness and free energy change for all the six operators are shown in Figure 2b. We observed a good agreement between E and ΔG and the correlation coefficient is 0.90.

In conclusion, a new structure based stiffness scale was developed using the three dimensional structures of DNA molecules and physical concepts. This scale explains the protein-DNA binding specificity and binding free energy change of 434 repressor and Cro repressor, respectively, with operator sequences significantly.

Acknowledgements

The author wishes to thank Professor Sandor Pongor for helpful discussions and Dr Jacob for critical reading of the manuscript.

References

1. Travers, A.A. and Klug, A.: In: N.R. Cozzarelli and J.C. Wang (eds.), *DNA topology and its biological effects*, Cold Spring Harbour Laboratory press, NY, 1990, pp. 57–106.
2. Travers, A.A.: *Curr. Biol.* **4** (1994), 659–661.
3. Calladine, C.R. and Drew, H.R.: *J. Mol. Biol.* **192** (1986), 907–918.
4. Haran, T.E., Kahn, J.D. and Crothers, D.M.: *J. Mol. Biol.* **244** (1994), 135–143.
5. Lavery, R. and Hartmann, B.: *Biophys. Chem.* **50** (1994), 33–45.
6. Parbhane, R.V., Thambe, S.S. and Kulkarni, B.D.: *Bioinformatics* **14** (1998), 131–138.
7. Barkley, M.D. and Zimm, B.H.: *J. Chem. Phys.* **70** (1979), 2991–3007.

8. Zhurkin, V.B., Ulyanov, N.B., Gorin, A.A. and Jernigan, R.L.: *Proc. Natl. Acad. Sci. U.S.A.* **88** (1991), 7046–7050.
9. Yakushevich: *Physica D* **79** (1994), 77–86.
10. Ponnuswamy, P.K. and Gromiha, M.M.: *J. Theor. Biol.* **169** (1994), 419–432.
11. Sanghani, S., Zakrzewska, K., Harvey, S.C. and Lavery, R.: *Nucl. Acid. Res.* **24** (1996), 1632–1637.
12. Vologodskii, A.V. and Frank-Kamenetskii, M.D.: *Methods Enzymol.* **211** (1992), 467–480.
13. Yang, Y., Westcorr, T.P., Pederson, S.C., Tobias, I. and Olson, W.K.: *Trends Biochem. Sci.* **20** (1995), 313–319.
14. Olson, W.K.: *Curr. Opin. Str. Biol.* **6** (1996), 242–256.
15. Gromiha, M.M., Munteanu, M.G., Gabrielian, A. and Pongor, S.: *J. Biol. Phys.* **22** (1996), 227–243.
16. Munteanu, M.G., Gromiha, M.M. and Pongor, S.: *Nonlinear Science Today*, URL (<http://www.springer-ny.com/nst>), 1998.
17. Gromiha, M.M., Munteanu, M.G., Simon, I. and Pongor, S.: *Biophys. Chem.* **69** (1997), 153–160.
18. Hogan, M.E. and Austin, R.H.: *Nature* **329** (1987), 263–266.
19. Smith, S.B., Finzi, L. and Bustamante, C.: *Science* **258** (1992), 1122–1126.
20. Smith, S.B., Cui, Y. and Bustamante, C.: *Science* **271** (1996), 795–799.
21. Strick, T.R., Allemand, J.F., Bensimon, D., Bensimon, A. and Croquette, V.: *Science* **271** (1996), 1835–1837.
22. Brukner, I., Sanchez, R., Suck, D. and Pongor, S.: *EMBO J.* **14** (1995), 1812–1818.
23. Brukner, I., Sanchez, R., Suck, D. and Pongor, S.: *J. Biomol. Str. Dyn.* **13** (1995), 309–317.
24. Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.E., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M.: *J. Mol. Biol.* **112** (1977), 535–542.
25. Manavalan, P. and Ponnuswamy, P.K.: *Nature* **275** (1978), 673–674.
26. Ponnuswamy, P.K. and Gromiha, M.M.: *Int. J. Pept. Protein Res.* **42** (1993), 326–341.
27. Selvaraj, S. and Gromiha, M.M.: *J. Prot. Chem.* **17** (1998), 407–415.
28. Gromiha, M.M. and Selvaraj, S.: *Protein Eng.* **11** (1998), 249–251.
29. Bauer, W.R., Lund, R.L. and White, J.H.: *Proc. Natl. Acad. Sci. U.S.A.* **90** (1993), 833–837.
30. Maddox, J.: *Nature* **375** (1995), 627.
31. Lehninger, A.: *Principles of Biochemistry*, Worth publishers, NY, 1982.
32. Gabrielian, A. and Pongor, S.: *FEBS Lett.* **393** (1996), 65–68.
33. Gabrielian, A., Vlahovicek, K., Munteanu, M.G., Gromiha, M.M., Brukner, I., Sanchez, R. and Pongor, S.: in: R.H. Sharma and M.H. Sharma (eds.), *Structure, Motion, Interaction and Expression of Biological Macromolecules* **1**, Adenine Press, NY, 1998, pp. 117–132.
34. Satchwell, S.C., Drew, H.R. and Travers, A.A.: *J. Mol. Biol.* **191** (1986), 659–675.
35. Goodsell, D.S. and Dickerson, R.E.: *Nucl. Acids Res.* **22** (1994), 5497–5503.
36. Koudelka, G.B., Harrison, S.B. and Ptashne, M.: *Nature* **326** (1987), 886–888.
37. Takeda, Y., Sarai, A. and Rivera, V.M.: *Proc. Natl. Acad. Sci. U.S.A.* **86** (1989), 439–443.