Published online November 1, 2004

5834–5840 *Nucleic Acids Research, 2004, Vol. 32, No. 19*
doi:10.1093/nar/gkh905

Core promoter elements of eukaryotic genes have a highly distinctive mechanical property

Yoshiro Fukue, Noriyuki Sumida, Jun-ichi Nishikawa and Takashi Ohyama*

Department of Biology, Faculty of Science and Engineering, Konan University, 8-9-1 Okamoto, Higashinada-ku, Kobe 658-8501, Japan

Received August 24, 2004; Revised and Accepted October 6, 2004

ABSTRACT

In spite of the abundant data on DNA sequence, the mechanical aspects of promoter DNA remain poorly understood. We classified 1871 human and 196 mouse RNA polymerase II promoters and investigated average flexibility profiles of the human promoters containing either a TATA box or an initiator (Inr) sequence only. Here, we show that TATA boxes and Inr sequences have a common anomalous mechanical property: they are comprised of distinctively flexible and rigid sequences, compared with the other parts of the promoter region. The +2 position in the Inr consensus sequence does not favor adenine to keep the high flexibility and thus this position is more accurately represented as 'T, G, C $\gg$ A'. Additionally, it was also found that DNA region upstream of TATA box or Inr sequence is more rigid than region downstream of each element. These properties may function as a marker for recognition by TATA-binding protein and Inr-binding protein.

INTRODUCTION

Key genetic processes, such as DNA replication, transcription and recombination are regulated by DNA–protein interactions. The human genome in somatic cells is ~2 m long, whereas the length of the TATA box (1,2) and initiator (Inr) elements (3,4) is only 2.4 nm. How can a general transcription factor TATA-binding protein (TBP) (5,6) and Inr-binding protein (7) rapidly find these elements? Intrinsically curved DNA structures may be raised as a possible signal for target-site selection by these proteins. Indeed, they have often been found in TATA box regions (8,9). However, not all TATA-containing promoters contain curved DNA structure. Furthermore, there is no report describing the presence of curved DNA structure in the Inr regions. Thus, DNA curvature cannot be a 'general' signal for the target-site selection. Generally, DNA-binding proteins seem to move from random to specific sites by multiple dissociation/reassociation events within a single DNA molecule (10). An attractive hypothesis is that during these events, some mechanical property of DNA may assist DNA-binding proteins in finding the target sites efficiently.

Flexibility, rigidity, anisotropic bendability (flexibility) and deformability are thought to be representative mechanical properties of DNA. The anisotropic bendability of DNA is well known to play a major role in establishing the overall organization of DNA–protein complexes required for the regulation of transcription (11–13). However, its role in target-site selection by transcription factors is unknown. The easier inherent deformability of the width of the minor groove in A•T-rich sequences seems to be one reason as to why A•T base pairs are required in the TATA box (14). This deformability, however, does not seem to be used in target-site selection by TBP, because there are a large number of A•T-rich sequences in eukaryotic genomes. The flexibility of DNA may play a key role in target-site selection. Indeed, sequence-dependent variations in the flexibility of DNA are suggested to be recognized in target-site selection by bacterial integration host factor, HU and eukaryotic HMG1 (15). In addition, it is also suggested that flexibility may be implicated in the interaction between yeast tDNA upstream regions and TFIIIB (16). However, as a whole, little attention has been paid to the role of mechanical properties of promoter DNA sequences in target-site selection by transcription factors, and the study of yeast tDNA described above stimulated us to analyze the flexibility of human and mouse class II gene (transcribed by RNA polymerase II) promoters.

The TATA box, the Inr element and the downstream promoter element (DPE) are generally known as core promoter elements of class II genes. The TATA box has a consensus TATA(A/T)A(A/T) sequence and is usually located ~25–31 bp upstream of the transcription start site, +1 (1,2). The Inr, which has a consensus PyPyAN(T/A)PyPy sequence (Py, pyrimidine; N, any nucleotide), is located around position +1 (3,4), and the DPE, with consensus sequence PuG(A/T)CGTG (Pu, purine), is centered around position +30 (17). Here, we report average flexibility profiles of human TATA-containing and Inr-containing promoters.

MATERIALS AND METHODS

As a source of human and mouse promoter sequences, we used the Eukaryotic Promoter Database (EPD) (18,19) (http://www.epd.isb-sib.ch). The EPD releases 76 and 78 contained 1871 human and 196 mouse promoters, respectively. We

*To whom correspondence should be addressed. Tel: +81 78 435 2547; Fax: +81 78 435 2539; Email: ohyama@konan-u.ac.jp

Nucleic Acids Research, Vol. 32 No. 19 © Oxford University Press 2004; all rights reserved

Table 1. Element-based sorting of 1871 human and 196 mouse class II gene promoters[a]

Element	Region	Human Detected (%)	Total (%)	Mouse Detected (%)	Total (%)
TATA only	−50 to −44	0		0	
	−49 to −43	0		0	
	−48 to −42	0		0	
	−47 to −41	3 (0.16%)		0	
	−46 to −40	0		0	
	−45 to −39	1 (0.05%)		0	
	−44 to −38	2 (0.11%)		0	
	−43 to −37	0		0	
	−42 to −36	0		0	
	−41 to −35	0		0	
	−40 to −34	1 (0.05%)		0	
	−39 to −33	1 (0.05%)		0	
	−38 to −32	0		0	
	−37 to −31	0		1 (0.51%)	
	−36 to −30	0		0	
	−35 to −29	2 (0.11%)[b]		0	
	−34 to −28	2 (0.11%)[b]		0	
	−33 to −27	9 (0.48%)[b]		2 (1.02%)	
	−32 to −26	10 (0.53%)[b]		6 (3.06%)	
	−31 to −25	19 (1.02%)[b]		6 (3.06%)	
	−30 to −24	27 (1.44%)[b]		4 (2.04%)	
	−29 to −23	20 (1.07%)[b]		8 (4.08%)	
	−28 to −22	6 (0.32%)[b]		3 (1.53%)	
	−27 to −21	3 (0.16%)[b]		4 (2.04%)	
	−26 to −20	5 (0.27%)[b]		1 (0.51%)	
	−25 to −19	0		1 (0.51%)	
	−24 to −18	0		0	
	−23 to −17	2 (0.11%)		0	
	−22 to −16	1 (0.05%)		0	
	−21 to −15	0		0	
	−20 to −14	0		0	
	−19 to −13	0		0	
	−18 to −12	0		0	
	−17 to −11	0		0	
	−16 to −10	0	114 (6.09%)	0	36 (18.37%)
Inr only	−10 to −4	1 (0.05%)		0	
	−9 to −3	3 (0.16%)		0	
	−8 to −2	4 (0.21%)		1 (0.51%)	
	−7 to −1	11 (0.59%)		3 (1.53%)	
	−6 to +1	5 (0.27%)		0	
	−5 to +2	7 (0.37%)		0	
	−4 to +3	6 (0.32%)		0	
	−3 to +4	3 (0.16%)		4 (2.04%)	
	−2 to +5	101 (5.40%)[b]		5 (2.55%)	
	−1 to +6	3 (0.16%)		1 (0.51%)	
	+1 to +7	5 (0.27%)		0	
	+2 to +8	10 (0.53%)		0	
	+3 to +9	12 (0.64%)		2 (1.02%)	
	+4 to +10	3 (0.16%)	174 (9.30%)	3 (1.53%)	19 (9.69%)
DPE only	+25 to +45	8 (0.43%)[b]		0	
TATA + Inr	−50 to +10	19 (1.02%)[b]		9 (4.59%)	
TATA + DPE	−50 to +45	1 (0.05%)		4 (2.04%)	
Inr + DPE	−10 to +45	1 (0.05%)		0	
TATA + Inr + DPE	−50 to +45	0		0	

[a]Human and mouse promoter sequences are from the current releases 76 and 78 of EPD (18,19), respectively.
[b]Subjected to further analyses.

subjected all of them to the element-based sorting. In the flexibility analysis, we excluded several promoters that contained ambiguous nucleotide symbols in the region from −101 to +61. This caused a slight difference in the corresponding sample numbers between Tables 1 and 2 or between Table 1 and Figures 1 and 2. The flexibility parameters were from Brukner *et al*. (20), Satchwell *et al*. (21) and Packer *et al*. (22).

RESULTS AND DISCUSSION

Population of TATA-containing promoters is low

The EPD is a non-redundant collection of eukaryotic class II gene promoter sequences, experimentally defined by a transcription start site (18,19). We first classified all the 1871 human and 196 mouse promoter sequences in the EPD, according to the element(s) in the promoter (Table 1).

Table 2. Frequency of nucleotides at positions +1 and +2 in the human promoters

Promoter type				Nucleotide (%)								#Samples
				+1				+2				
Name	TATA	Inr	DPE	A	G	C	T	A	G	C	T	
I	+	–	–	48.0	31.4	12.7	7.8	27.5	23.5	28.4	20.6	102[a]
II	–	+	–	100	0	0	0	3.0	35.6	19.8	41.6	101
III	–	–	+	62.5	12.5	25.0	0	0	50.0	12.5	37.5	8
IV	+	+	–	100	0	0	0	25.0	0	25.0	50.0	8[b]

[a]The 103 TATA-only promoters that are indicated with 'b' in Table 1 contain a promoter whose nucleotide at position +2 has not been determined. This promoter was therefore excluded from the analysis.
[b]Only promoters that have Inr at the 'correct' site (−2 to +5) were included in the analysis.

Interestingly, the population of promoters that contain a canonical TATA box was found to be small in each case. Screening of the database using TATA(A/T)A as the TATA box consensus sequence resulted in only a slight increase in this population (data not shown). The population of TATA-box-containing promoters in yeast promoters is also low, and they are used to drive transcription of specific genes (23), which may also be true for human and mouse promoters. The population of promoters that contain a canonical Inr sequence was similarly found to be small in each case.

The promoters containing either a TATA box or an Inr sequence only were subjected to further analysis. The relationship between human promoter elements and the nucleotide at transcription start site or at position +2 is shown in Table 2. For promoters with only a TATA box (type I promoters in Table 2), no novel features were found. However, an interesting feature was observed for promoters containing Inr alone (type II) at a 'correct' position, for which the analysis indicated that adenine is not favored at position +2. Hence, only 3 of the 101 human Inr-only promoters had adenine at this position (3.0%), and all of the five mouse Inr-only promoters contained nucleotides other than A at position +2 (data not shown). The +2 position is described as 'N' in the Inr consensus sequence, but our analysis suggests that this position is more accurately represented as 'T, G, C $\gg$ A', at least for human Inr sequences. We return to a discussion of the reasons for adenine not being favored at this position later in the paper. Concerning the type III and IV promoters in Table 2 (III, containing DPE only; and IV, containing both TATA box and Inr sequences, but not DPE), the sample sizes are too limited to predict nucleotide preferences statistically.

TATA and Inr have a highly distinctive mechanical property

To calculate the average flexibility profiles of human promoters, first, we used the DNase I-derived flexibility parameters reported by Brukner *et al*. (20). They determined flexibilities of the 32 trinucleotide steps, based on the observed DNase I cutting frequencies [DNase I cleavage is strongly correlated with local variation in DNA flexibility (24)]. Each triplet flexibility (ln p, where p is defined as the bending propensity) is given in arbitrary units, ranging from −0.280 to 0.194, where a value of zero or values with negative signs have no particular meaning, since these values are generated from a formula containing a natural logarithm. We excluded from the analysis those promoters that contained undetermined nucleotide(s) in the region between −101 and +61. Concerning the promoters containing TATA boxes only, we first analyzed three major populations in which the TATA boxes are located between −31 and −25 (Figure 1A), between −30 and −24 (Figure 1B) and between −29 and −23 (Figure 1C), respectively. Interestingly, a common feature was observed for three TATA box-only populations: the TATA box is divided into two mechanically different parts. When one nucleotide flanking each end of the TATA box sequence is included and the sequence is then segmented into trinucleotide steps stepwise from upstream (5′) to downstream (3′), seven trinucleotide steps (triplets), NTA, TAT (ln p = 0.182), ATA (0.182), TA(A/T), A(A/T)A, (A/T)A (A/T) and A(A/T)N are generated. As shown in Figure 1A–C, the first four triplets confer high flexibility to the TATA box, whereas, in contrast, the fifth, sixth and seventh triplets are markedly different and confer significant rigidity. This feature is particularly marked for promoters that have the TATA box between −30 and −24 (Figure 1B) and between −29 and −23 (Figure 1C). The rigidities of the fifth and sixth steps were higher than expected. The fifth triplet is either ATA or AAA (ln p = −0.274), and their average ln p value is −0.046. Four triplets can occur at the sixth trinucleotide step, (A/T)A(A/T): i.e. AAA (−0.274), AAT (−0.280), TAA (0.068) and TAT (0.182). If they occur with even frequency, the average flexibility would be −0.076. As shown in Figure 1B and C, the actual mean values obtained for the fifth and sixth triplets were considerably smaller than the average values, −0.046 and −0.076, respectively, indicating that many promoters tend to adopt distinctively rigid triplets in this part of the sequence. Eight triplets can occur at the seventh trinucleotide step: AAA (−0.274), AAG (−0.081), AAC (−0.205), AAT (−0.280), ATA (0.182), ATG (0.134), ATC (−0.110) and ATT (−0.280). These triplets give an average value of −0.114, and the actual data shown in Figure 1B and C show mean values close to this. Compared to the promoters in Figure 1B and C, the promoters in the population shown in Figure 1A have less marked rigidity of the fifth, sixth and seventh triplets. However, the contrast in flexibility between upstream triplets and downstream triplets within the TATA box is also evident in this population. The distance between the TATA box and the transcription initiation site may be implicated in this difference, but whatever be the explanation, we were unable to find any special features common to the genes in the population in Figure 1A.

Another interesting feature emerged upon analysis of the average flexibility for 97 TATA box-only promoters, without distinguishing position of the TATA box (Figure 1D). This analysis showed that the duplex region upstream of the TATA box is more rigid than that downstream of the TATA box, on

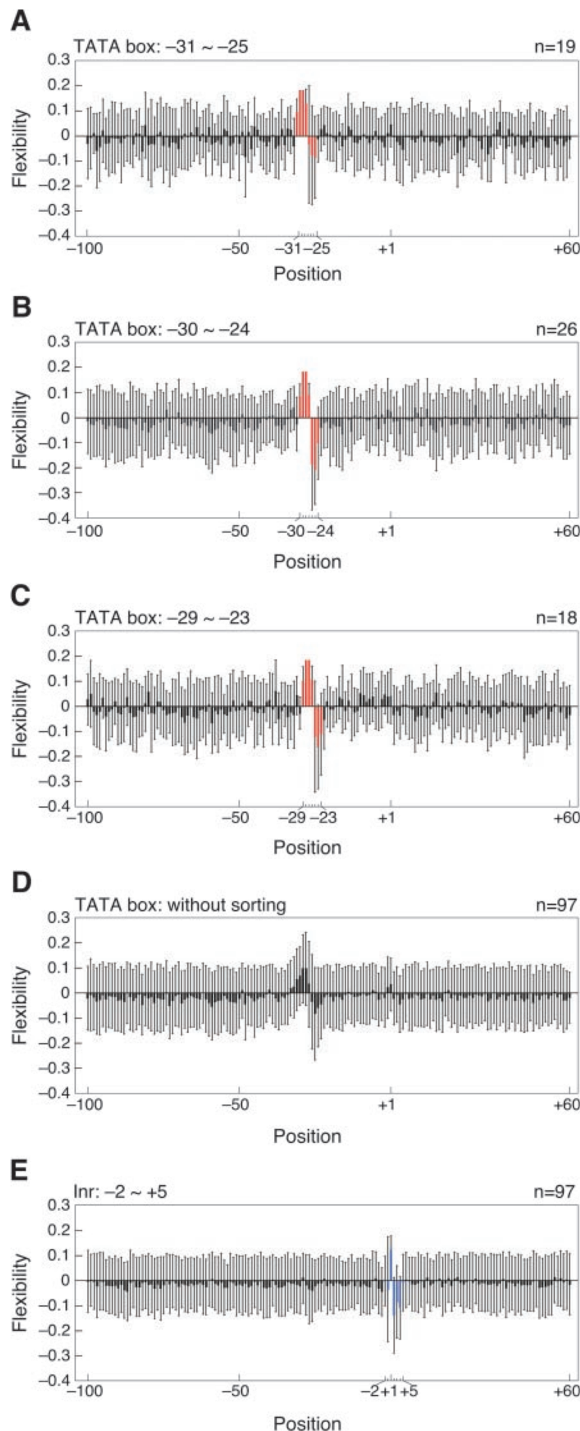

Figure 1. Average flexibility profiles of human promoters containing either a TATA box or an Inr sequence only, as calculated from DNase I-derived flexibility parameters. (**A**), (**B**) and (**C**) show the results for 19, 26 and 18 promoters in which the TATA box is located at positions −31 to −25, −30 to −24 and −29 to −23, respectively. (**D**) 97 promoters in which the TATA box is located between −35 and −20 were analyzed collectively, without location-based grouping. (**E**) Results for 97 promoters in which Inr is located at positions −2 to +5. The flexibility parameters are taken from Brukner *et al*. (20). Flexibility of a given triplet is plotted against the position of the center nucleotide. Values are shown as means ± SD. The flexibilities of TATA box and Inr sequences are colored red (A–C) and blue (E), respectively. This analysis used only promoters for which the sequence between −101 and +61 is completely determined. Thus, sample numbers are slightly different from the data shown in Table 1.

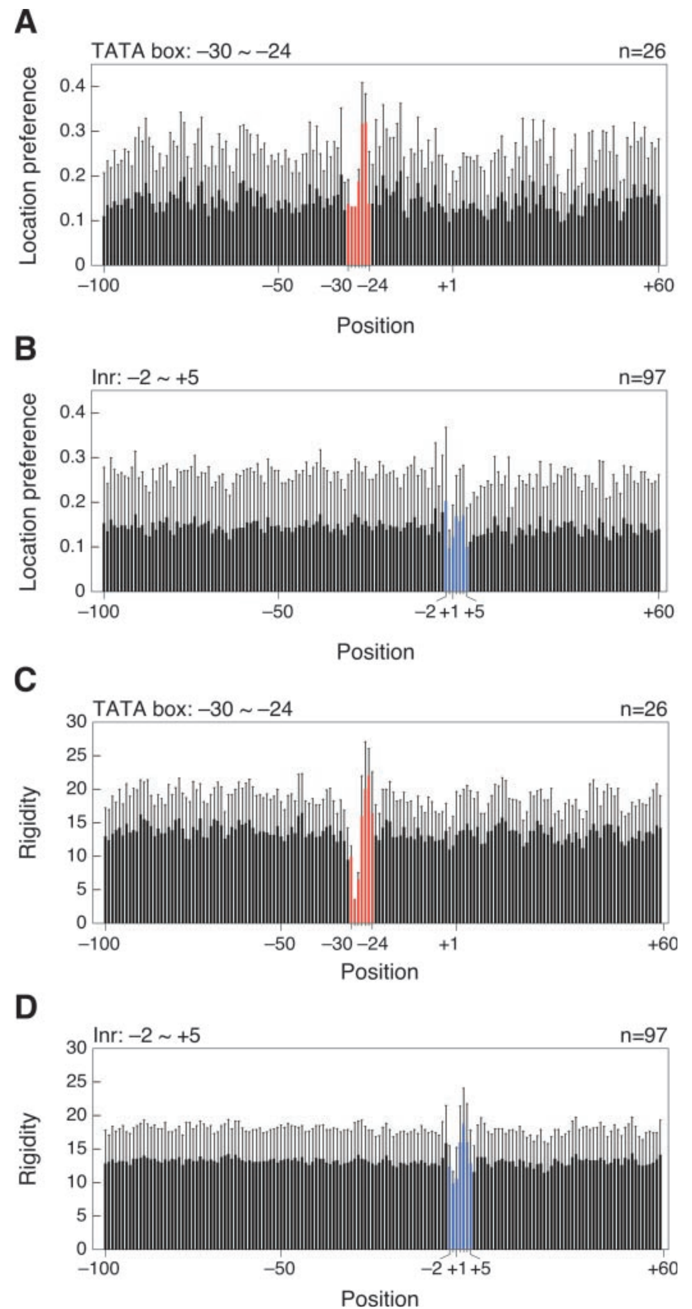

Figure 2. Calculation of average flexibility profiles of promoter sequences, using another two models. (**A**) and (**B**), calculations using a trinucleotide model based on DNA positioning on nucleosomes (21); (**C**) and (**D**), calculations based on a tetranucleotide potential energy surface model [(22); two values are given to a single sequence TGAG and a sequence TCAG is missing in the paper; as the value for TCAG, we used 8.9 (one of the two values)]. The promoters shown in Figure 1B and E were subjected to these analyses, and the resulting data are shown in (A and C), and (B and D), respectively. The value for a given trinucleotide or tetranucleotide is plotted against the position of the central nucleotide or the second nucleotide, respectively. Values are shown as means ± SD. Lower values correspond to more flexible sequences. Colors are used in the same way as that described in Figure 1.

average. Pedersen *et al*. (25) have reported that flexibility is low in the region upstream of the transcription start site and significantly higher downstream. Our results show that for TATA box-only promoters the region between the TATA

box and the transcription start site is also more flexible than that upstream of the box, which Pedersen *et al.* (25) may not have detected because of inclusion of Inr-containing promoters in their analysis. In Figure 1D, the difference in flexibility between the regions upstream and downstream of the TATA box is evident, but is not striking. Importantly, a striking difference is only observed between the TATA box and the other regions. It should also be noted that the flexibility of the transcription initiation site shows a difference from the surrounding regions.

The flexibility profile for 97 human promoters containing an Inr sequence only is shown in Figure 1E. Interestingly, the result is very similar to those in Figure 1A–D, and the Inr region is clearly distinguishable from the other regions. When 1 nt flanking each end of the Inr sequence is included, as for the TATA box analysis, the third trinucleotide step has a remarkable flexibility. As discussed earlier, the data in Table 2 suggest that the nucleotide at position +2 is rarely an A for Inr-only promoters. When an A is present, the sequence from positions −1 to +2 (the third step) could be CAA, which would be the most rigid triplet among the possible eight triplets in this region. Therefore, it is possible that to keep this region flexible, A tends to be not selected. On the other hand, the fourth, fifth and sixth steps are all very rigid, compared to the other steps in the region from −100 to +60. Furthermore, the flexibility profile differs between the regions upstream and downstream of the Inr: the upstream region (−100 ∼ Inr) is slightly more rigid than the downstream region (Inr ∼ +60), on average. Thus, two common features are observed between the promoters containing either a TATA box or an Inr sequence only: (i) both elements contain highly flexible and highly rigid triplet(s) in the upstream and downstream halves, respectively, which are clearly different from the flexibility of other triplets in the promoter; and (ii) the region upstream from each element is more rigid than the downstream region. As described above, Pedersen *et al.* (25) have reported that the flexibility profile upstream of the transcription start site differs from that downstream of the start site. However, the border between the flexible and rigid regions appears not to be the transcription start site but the TATA box or Inr sequence. Furthermore, the difference in flexibility between the upstream and downstream regions is slight.

As described above, the cleavage by DNase I is strongly correlated with local variation in DNA flexibility (24,26,27). However, the enzyme also preferably cuts DNA that is inherently bent toward the major groove (26,27). Therefore, the observed DNase I cutting frequencies do not necessarily reflect local variation in DNA flexibility alone. In order to deduce the ‘pure’ flexibility profiles, we also performed calculations using another set of trinucleotide parameters that were obtained from the data on DNA positioning on nucleosomes (21), or using a set of tetranucleotide parameters obtained from the molecular orbital calculations (22). The results are shown in Figure 2. Although the presence of highly flexible triplets in TATA box is less obvious in Figure 2A, the average flexibility profiles are generally in good accordance with the corresponding data in Figure 1 (note that in Figure 2 lower values correspond to more flexible sequences). The tetranucleotide model incorporates more sequence context information than do dinucleotide- or trinucleotide-based descriptions. Thus, the average flexibility profiles shown in Figure 2C and D may be the most accurate among the three sets of data. Whatever the inference, the average flexibility profiles in Figure 2C and D are strikingly similar to those in Figure 1B and E, respectively.

Figure 1 shows that the region upstream from the TATA box or Inr is slightly more rigid than the downstream region. The same feature is observed in each panel in Figure 2. The mean flexibilities of the upstream and downstream regions are as follows: in panel A, 0.15 versus 0.14 (upstream versus downstream); B, 0.15 versus 0.14; C, 13.89 versus 13.29; and D, 13.19 versus 13.05 (again, lower values correspond to more flexible sequences). Thus, it is safely concluded that three different methods produced essentially the same results. Although the total number of mouse samples is too small to make a conclusion statistically, the same ‘tendency’ to that observed for human promoters was observed (data not shown).

The distinctive flexibility profiles observed for the TATA box and Inr sequences may function in target-site selection by TBP and Inr-binding proteins, such as $TAF_{II}250$ and $TAF_{II}150$ (7). As an analogy, rubber and metal rods of the same appearance can easily be distinguished, even by a light touch, and biological systems may use a similar mechanism. Furthermore, the mechanical properties of the TATA box presumably play a central role in determining the extent and direction of the TBP-induced DNA bend (28), and here we partially confirm this hypothesis.

Using TATA box of the adenovirus major late promoter (AdMLP) and its variants, it has been shown that the extent of DNA bending in the TBP/TATA-box complex is DNA sequence-dependent (29–31). The bend angles of TAAAAAA, CATAAAA, TATAAAC and TATAATA (nucleotide variations from AdMLP are underlined) are considerably smaller than that of the wild-type sequence (TATAAAA). In contrast, a variant TATATAA is more bent than the wild-type sequence (29). Concerning the variants TAAAAAA and TATATAA, changes in the bend angle can be explained in terms of the mechanical properties of DNA. The sequence TAAAAAA shows the least bend among the variants, and this sequence can be shown to be much more rigid than the wild-type and is the most rigid among the variant sequences. The triplet parameters of Brukner *et al.* (20) clearly illustrate this point: i.e. TAA (0.068) versus TAT (0.182); AAA (−0.274) versus ATA (0.182); and AAA (−0.274) versus TAA (0.068). On the other hand, the TATATAA is more flexible than the wild-type sequence, and thus the flexibility of the TATA box correlates well with the bend angle in these cases. In the variants CATAAAA and TATAAAC, however, the altered nucleotides are located at the positions of phenylalanine intercalation (in the wild-type sequence, TATAAAAG, phenylalanine intercalation occurs at either end of the sequence; at T•A:A•T and A•T:G•C base pair steps), where the TATA element is sharply kinked (32,33). Therefore, in these cases, the cause of the reduction in bend angle may mainly be attributable to hindrance of the intercalation. In the case of TATAATA, the triplets ATA and TAG (G is the flanking nucleotide) are more flexible (0.182 and 0.090, respectively) than the corresponding wild-type triplets AAA (−0.274) and AAG (−0.081), respectively. Thus, we cannot explain why this sequence is less bent than the wild-type sequence in the respective TBP complexes, and an unknown mechanism, other than DNA flexibility, may be implicated.

A large change in DNA flexibility that occurs between the upstream and downstream halves (4 and 3 nt, respectively) of the TATA box may contribute to the unwinding of the box. Unwinding of the TATA box enables the DNA to maintain almost symmetrical contact with the under surface of the TBP saddle (32,33), which is very important for stable complex formation. The X-ray crystallography has revealed that the greatest unwinding occurs at the very center of the TATA box (32,33), which is the junction of the flexible and rigid parts of the sequence, as reported in this study. Although protein-induced DNA bending has not been studied for Inr, we predict that it would occur and that the bending trajectory of the Inr sequence in the complex would be similar to the trajectory of the TATA sequence in the TBP/TATA-box complex.

The remaining question is why the upstream regions of TATA box and Inr are slightly more rigid than the downstream regions of these elements. This property may also help transcription factors to find these sequences, or may be utilized in promoter packaging into chromatin. DNA flexibility influences the formation, stability and positioning of nucleosomes (34–36), and thus the lower flexibility of the region upstream of each element may function to make the promoter nucleosome-free for transcription.

Some of the results reported here are experimentally verifiable. For example, the significance of the mechanical properties of TATA box (or Inr) could be assessed by transcription assays using synthetic DNA fragments with 'TATA boxes' that mimic the average flexibility profile of canonical TATA boxes, but have sequences that are completely different from the wild-type sequence. This experiment is now in progress.

ACKNOWLEDGEMENTS

We thank C. Hasegawa and T. Okabe for assistance in the analysis and J. Ohyama for help in preparing the manuscript. This study was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan (T.O.).

REFERENCES

1. Breathnach,R. and Chambon,P. (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.*, **50**, 349–383.
2. Carey,M. and Smale,S.T. (2000) *Transcriptional regulation in eukaryotes: concepts, strategies, and techniques*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3. Javahery,R., Khachi,A., Lo,K., Zenzie-Gregory,B. and Smale,S.T. (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.*, **14**, 116–127.
4. Smale,S.T., Jain,A., Kaufmann,J., Emami,K.H., Lo,K. and Garraway,I.P. (1998) The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes. *Cold Spring Harb. Symp. Quant. Biol.*, **63**, 21–31.
5. Peterson,M.G., Tanese,N., Pugh,B.F. and Tjian,R. (1990) Functional domains and upstream activation properties of cloned human TATA binding protein. *Science*, **248**, 1625–1630.
6. Kao,C.C., Lieberman,P.M., Schmidt,M.C., Zhou,Q., Pei,R. and Berk,A.J. (1990) Cloning of a transcriptionally active human TATA binding factor. *Science*, **248**, 1646–1650.
7. Chalkley,G.E. and Verrijzer,C.P. (1999) DNA binding site selection by RNA polymerase II TAFs: a $TAF_{II}250$-$TAF_{II}150$ complex recognizes the initiator. *EMBO J.*, **18**, 4835–4845.
8. Gabrielian,A.E., Landsman,D. and Bolshoy,A. (1999) Curved DNA in promoter sequences. *In Silico Biol.*, **1**, 183–196.
9. Ohyama,T. (2004) Curved DNA and Transcription in eukaryotes. In Ohyama,T. (ed.), *DNA Conformation and Transcription*. Landes Bioscience, Texas, in press.
10. Gowers,D.M. and Halford,S.E. (2003) Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO J.*, **22**, 1410–1418.
11. van der Vliet,P.C. and Verrijzer,C.P. (1993) Bending of DNA by transcription factors. *Bioessays*, **15**, 25–32.
12. Ozoline,O.N., Deev,A.A. and Trifonov,E.N. (1999) DNA bendability-a novel feature in *E.coli* promoter recognition. *J. Biomol. Struct. Dyn.*, **16**, 825–831.
13. Travers,A.A. (2004) Gene regulation by HMGA and HMGB chromosomal proteins and related architectural DNA-binding proteins. In Ohyama,T. (ed.), *DNA Conformation and Transcription*. Landes Bioscience, Texas, in press.
14. Juo,Z.S., Chiu,T.K., Leiberman,P.M., Baikalov,I., Berk,A.J. and Dickerson,R.E. (1996) How proteins recognize the TATA box. *J. Mol. Biol.*, **261**, 239–254.
15. Grove,A., Galeone,A., Mayol,L. and Geiduschek,E.P. (1996) Localized DNA flexibility contributes to target site selection by DNA-bending proteins. *J. Mol. Biol.*, **260**, 120–125.
16. Giuliodori,S., Percudani,R., Braglia,P., Ferrari,R., Guffanti,E., Ottonello,S. and Dieci,G. (2003) A composite upstream sequence motif potentiates tRNA gene transcription in yeast. *J. Mol. Biol.*, **333**, 1–20.
17. Burke,T.W. and Kadonaga,J.T. (1997) The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by $TAF_{II}60$ of *Drosophila*. *Genes Dev.*, **11**, 3020–3031.
18. Cavin Périer,R., Junier,T. and Bucher,P. (1998) The Eukaryotic Promoter Database EPD. *Nucleic Acids Res.*, **26**, 353–357.
19. Schmid,C.D., Praz,V., Delorenzi,M., Périer,R. and Bucher,P. (2004) The Eukaryotic Promoter Database EPD: the impact of *in silico* primer extension. *Nucleic Acids Res.*, **32**, D82–D85.
20. Brukner,I., Sánchez,R., Suck,D. and Pongor,S. (1995) Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *EMBO J.*, **14**, 1812–1818.
21. Satchwell,S.C., Drew,H.R. and Travers,A.A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.*, **191**, 659–675.
22. Packer,M.J., Dauncey,M.P. and Hunter,C.A. (2000) Sequence-dependent DNA structure: tetranucleotide conformational maps. *J. Mol. Biol.*, **295**, 85–103.
23. Basehoar,A.D., Zanton,S.J. and Pugh,B.F. (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, **116**, 699–709.
24. Hogan,M.E., Roberson,M.W. and Austin,R.H. (1989) DNA flexibility variation may dominate DNase I cleavage. *Proc. Natl Acad. Sci. USA*, **86**, 9273–9277.
25. Pedersen,A.G., Baldi,P., Chauvin,Y. and Brunak,S. (1998) DNA structure in human RNA polymerase II promoters. *J. Mol. Biol.*, **281**, 663–673.
26. Lahm,A. and Suck,D. (1991) DNase I-induced DNA conformation: 2 Å structure of a DNase I-octamer complex. *J. Mol. Biol.*, **222**, 645–667.
27. Weston,S.A., Lahm,A. and Suck,D. (1992) X-ray structure of the DNase I-d$(GGTATACC)_2$ complex at 2.3 Å resolution. *J. Mol. Biol.*, **226**, 1237–1256.
28. Horikoshi,M., Bertuccioli,C., Takada,R., Wang,J., Yamamoto,T. and Roeder,R.G. (1992) Transcription factor TFIID induces DNA bending upon binding to the TATA element. *Proc. Natl Acad. Sci. USA*, **89**, 1060–1064.
29. Starr,D.B., Hoopes,B.C. and Hawley,D.K. (1995) DNA bending is an important component of site-specific recognition by the TATA binding protein. *J. Mol. Biol.*, **250**, 434–446.
30. Wu,J., Parkhurst,K.M., Powell,R.M., Brenowitz,M. and Parkhurst,L.J. (2001) DNA bends in TATA-binding protein-TATA complexes in solution are DNA sequence-dependent. *J. Biol. Chem.*, **276**, 14614–14622.
31. Powell,R.M., Parkhurst,K.M. and Parkhurst,L.J. (2002) Comparison of TATA-binding protein recognition of a variant and consensus DNA promoters. *J. Biol. Chem.*, **277**, 7776–7784.

32. Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.
33. Kim,Y., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
34. Zhurkin,V.B., Lysov,Y.P. and Ivanov,V.I. (1979) Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Res.*, **6**, 1081–1096.
35. Baldi,P., Brunak,S., Chauvin,Y. and Krogh,A. (1996) Naturally occurring nucleosome positioning signals in human exons and introns. *J. Mol. Biol.*, **263**, 503–510.
36. Anselmi,C., Bocchinfuso,G., De Santis,P., Savino,M. and Scipioni,A. (1999) Dual role of DNA intrinsic curvature and flexibility in determining nucleosome stability. *J. Mol. Biol.*, **286**, 1293–1301.