

# A highly distinctive mechanical property found in the majority of human promoters and its transcriptional relevance

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

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

Received May 11, 2005; Revised and Accepted June 22, 2005

## ABSTRACT

**A recent study revealed that TATA boxes and initiator sequences have a common anomalous mechanical property, i.e. they comprise distinctive flexible and rigid sequences when compared with the other parts of the promoter region. In the present study, using the flexibility parameters from two different models, we calculated the average flexibility profiles of 1004 human promoters that do not contain canonical promoter elements, such as a TATA box, initiator (Inr) sequence, downstream promoter element or a GC box, and those of 382 human promoters that contain the GC box only. Here, we show that they have a common characteristic mechanical property that is strikingly similar to those of the TATA box-containing or Inr-containing promoters. Their most interesting feature is that the TATA- or Inr-corresponding region lies in the several nucleotides around the transcription start site. We have also found that a dinucleotide step from –1 to +1 (transcription start site) has a slight tendency to adopt CA that is known to be flexible. We also demonstrate that certain synthetic DNA fragments designed to mimic the average mechanical property of these 1386 promoters can drive transcription. This distinctive mechanical property may be the hallmark of a promoter.**

## INTRODUCTION

Recognition of promoter and enhancer regions by sequence-specific DNA-binding proteins is a key process in transcription. However, how these proteins find their target sites efficiently is not yet understood. It has been suggested that sequence-specific DNA-binding proteins move from random to specific sites via multiple dissociation/re-association events within a

single DNA molecule (1). In this process, the structural properties of DNA, such as its intrinsic bending and flexibility, may help proteins to locate their target sites. Indeed, several curved DNA structures have been suggested to have such a role (2,3). DNA flexibility is also suggested to play a role in the interaction between yeast tDNA upstream regions and TFIIB (4). DNA flexibility may also be more generally used in binding site recognition by DNA-bending proteins, such as bacterial IHF (integration host factor), HU (nucleoid-associated protein) and eukaryotic HMG1 (high mobility group protein 1) (5).

Recently, we reported that two common features are observed between promoters containing either a TATA box or an Inr sequence: (i) both elements contain highly flexible and highly rigid triplets in their upstream and downstream halves, respectively, and (ii) the upstream region of each element is more rigid than their downstream region (6). These mechanical properties are considered to function as markers for recognition by TATA-binding protein and Inr-binding protein (6). It was also found that of the 1871 human promoter sequences in the eukaryotic promoter database [EPD; (7,8)], the population of promoters that contain a TATA box, Inr sequence or downstream promoter element (DPE) is very small. The EPD is a non-redundant collection of eukaryotic class II gene promoter sequences. In the EPD TATA box-only, Inr-only and DPE-only human promoters account for just 6, 9 and 0.4%, respectively, and the percentage of TATA- and Inr-containing human promoters is only 1% (6). In the present study, we found that GC box-containing promoters account for ~20% of those in the EPD. Therefore, more than half of the human promoters in the EPD do not contain a TATA box, Inr sequence, DPE or a GC box. These promoters will be hereon referred to as 'core-less' promoters. Similarly, the population of TATA box-containing promoters in yeast is also low (9). The proportion of core-less promoters may therefore be high in all eukaryotes.

As described above, we speculate that the mechanical properties common to the TATA-containing and Inr-containing

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

promoters function as markers for recognition by TATA-binding protein and Inr-binding protein. This then raises the question of whether core-less promoters also possess some such specific mechanical property. We calculated the average flexibility profiles of human core-less promoters and GC box-only promoters using parameters from two different models and found that these also have common distinctive mechanical properties in the region containing the transcription start site. Furthermore, we have experimentally demonstrated, for the first time, the significance of this mechanical property of DNA in transcription.

## MATERIALS AND METHODS

### Promoter database and flexibility parameters

We used EPD (7,8) (<http://www.epd.isb-sib.ch>) release 80 as a source of human promoter sequences and subjected all the 1871 sequences to element-based sorting. In the flexibility analysis, we excluded sequences that contained ambiguous nucleotide symbols in the analyzed region. The flexibility parameters were from Brukner *et al.* (10) and Packer *et al.* (11).

### Plasmid construction

We designed the following single-stranded DNA fragments and had them synthesized by Hokkaido System Science (Hokkaido, Japan): template 1, 5'-GCTAACGCGT(AAG)<sub>20</sub>-CAGT(CGT)<sub>10</sub>AGATCTGCTA; template 2, 5'-GCTAACGCGT(AGA)<sub>20</sub>CCAG(CGT)<sub>10</sub>AGATCTGCTA; template 3, 5'-GCTAACGCGT(TCG)<sub>20</sub>CAGT(AAG)<sub>10</sub>AGATCTGCTA; template 4, 5'-GCTAACGCGT(GTC)<sub>20</sub>CCAG(AAG)<sub>10</sub>AGATCTGCTA; template 5, 5'-GCTAACGCGT(AGA)<sub>21</sub>(CGT)<sub>10</sub>AGATCTGCTA; template 6, 5'-GCTAACGCGT(TCG)<sub>21</sub>(AAG)<sub>10</sub>AGATCTGCTA; template 7, 5'-GCTAACGCGT(AAG)<sub>31</sub>AGATCTGCTA; template 8, 5'-GCTAACGCGT(CGT)<sub>31</sub>AGATCTGCTA; template 9, 5'-GCTAACGCGT(TCG)<sub>10</sub>TCCCCGCG(TCG)<sub>7</sub>CAGT(AAG)<sub>10</sub>AGATCTGCTA; primer 1, 5'-TAGCAGATCTACG; and primer 2, 5'-TAGCAGATCTCTT. The templates 1, 2, 5 and 8 were annealed with primer 1 and the templates 3, 4, 6, 7 and 9 were annealed with primer 2. Primer extension reactions were performed according to a standard protocol (12). Each of the resulting double-stranded DNA fragments was then digested with MluI and BglII, and cloned between the corresponding sites in PGV-B (a vector for luciferase assay; TOYO B-Net, Japan). All constructs were confirmed by sequencing.

### Luciferase assay

Simian COS-7 cells were grown in Eagle's minimal essential medium containing 5% fetal bovine serum at 37°C in 5% CO<sub>2</sub>. Each construct was introduced into the cells by electroporation at 300 V/300 μF, in cuvettes containing 1.25 × 10<sup>6</sup> cells and 5 μg of construct in a 260 μl volume. After electroporation, the cells were cultured for 21 h and the luciferase assay performed as reported previously (13).

## RESULTS AND DISCUSSION

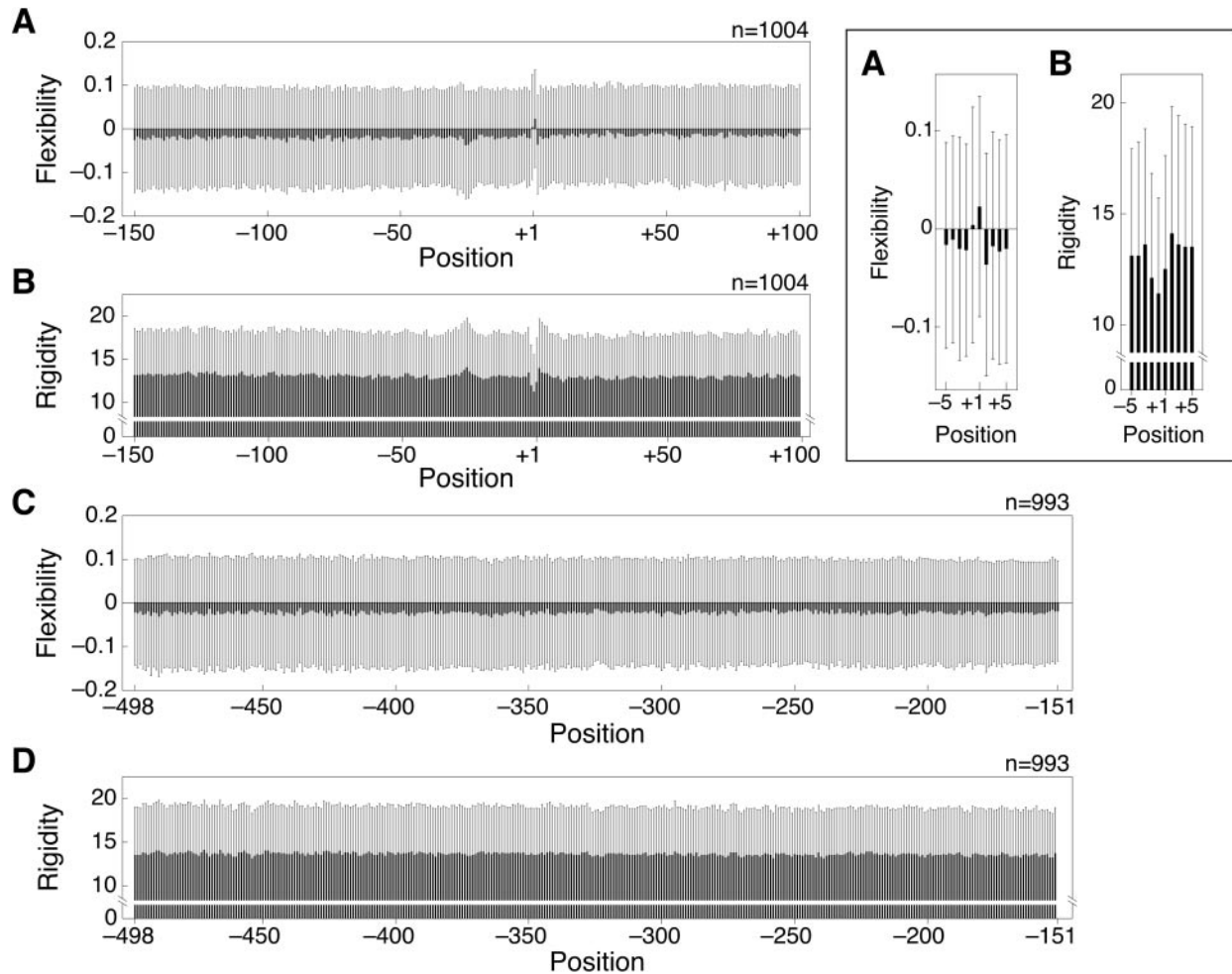
### Population of core-less promoters

The TATA box is usually located ~25–31 bp upstream of the transcription start site, +1 and has a consensus

TATA(A/T)A(A/T) sequence (14,15). The Inr is located around position +1 and has a consensus PyPyAN(T/A)-PyPy (Py, pyrimidine; N, any nucleotide) (16,17). The DPE, with a consensus sequence PuG(A/T)CGTG (Pu, purine), is centered around position +30 (18). There is also a class of RNA polymerase II (pol II) promoters that comprise G/C-rich sequences and contain multiple binding sites for the transcription factor Sp1 instead of the elements described above. The Sp1 binding site has a consensus GGGCGG (19,20) and is often called a GC box. The EPD contains 382 human promoters that harbor a GC box or GC boxes only in the region between –100 and –1. They account for 20.4% of all human promoters in the database. In order to calculate the average flexibility profile of 'core-less' promoters, the promoters that have any one of the sequences described above were excluded [the TATA-, Inr- or DPE-containing promoters had already been subjected to element-based sorting in the previous study and all of the promoters listed in Table 1 in ref. (6) were excluded]. In addition, since the tetranucleotide sequence, TATA, may function as a TATA box in transcription (21–23), we also excluded the promoters that contain this sequence between positions –50 and –10. Finally, 1004 promoters remained.

### Core-less promoters and GC box-only promoters have common distinctive mechanical properties around the transcription start site

In the calculation of an average flexibility profile, we used the DNase I-derived trinucleotide flexibility parameters reported by Brukner *et al.* (10) and a set of tetranucleotide parameters obtained from molecular orbital calculations (11). For the former set of flexibility parameters, each triplet flexibility,  $lnp$  ( $p$  is defined as the bending propensity), is given in arbitrary units, ranging from the lowest flexibility at a value of –0.280 to the highest flexibility at 0.194, where a formula containing a natural logarithm generated values with negative signs or a value of zero. For the latter set of flexibility parameters, lower values correspond to more flexible sequences. Taking these points into consideration, the average flexibility profiles of the 1004 core-less promoters were found to be almost the same (Figure 1A and B). Three characteristic features should be noted, i.e. (i) the region around the transcription start site contains a distinctively flexible sequence and a considerably rigid sequence, compared with other parts of the promoter region, (ii) the upstream region of the transcription start site is slightly more rigid than the downstream region (also see Table 1) and (iii) the region around position –25 is relatively rigid. Characteristics of (i) and (ii) are strikingly similar to the characteristics of the TATA-only promoters and Inr-only promoters, i.e. the TATA boxes and the Inr sequences comprise distinctively flexible and rigid sequences and the DNA region upstream of the TATA box or Inr sequence is more rigid than the region downstream of each element (6). In Figure 1C and D, the average flexibility profiles of a 'control' region, from –498 to –151, of the 993 core-less promoters are shown (the analyses only used promoters for which the sequence of the region was completely determined). As may be seen in Figure 1C and D, we neither detected the flexible triplet or quartet that was observed in sequence including the transcription start site, nor did we detect a



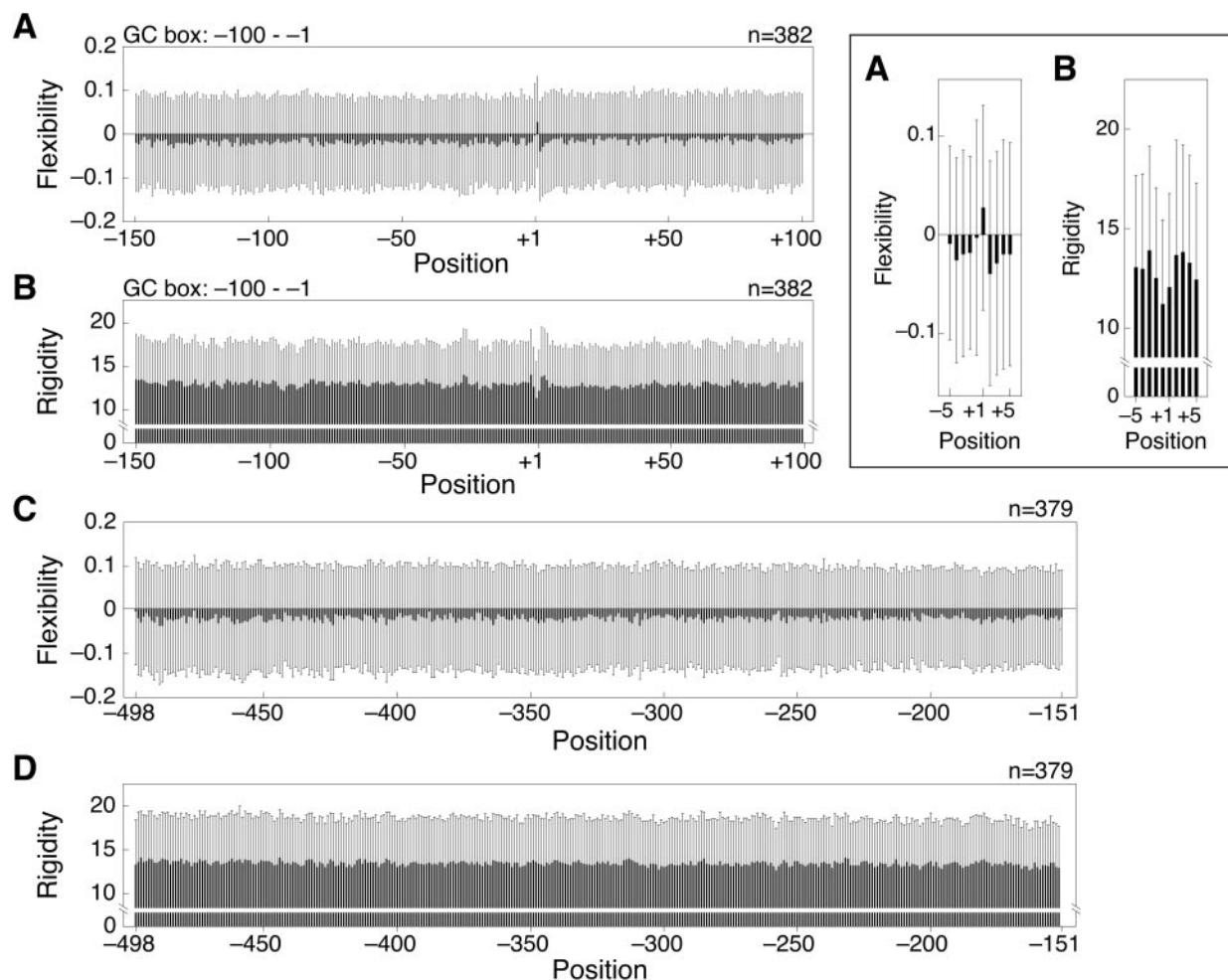
**Figure 1.** Average flexibility profiles of 1004 human promoters that do not contain a TATA box, Inr sequence, DPE or a GC box, as calculated from DNase I-derived flexibility parameters (A) or from the tetranucleotide potential energy surface model (B). The parameters were taken from Brukner *et al.* (10) and Packer *et al.* (11). The '+1' corresponds to the transcription start site. Values are shown as means  $\pm$  SD. Lower values correspond to more flexible sequences in (B). Flexibility is plotted against the position of the center nucleotide (A) or the second nucleotide (B). For example, the triplet from  $-1$  to  $+2$  in (A) is plotted against  $+1$  and the quartet from  $-2$  to  $+2$  (B) is plotted against  $-1$ , respectively (they are the most flexible triplet and quartet in the analyzed region). The data around the transcription start site are magnified on the right. (C and D) show the average flexibility profiles of the region from  $-498$  to  $-151$  as controls and were calculated from DNase I-derived flexibility parameters (C) and the tetranucleotide potential energy surface model (D), respectively. The analyses only used promoters for which the sequence between  $-499$  and  $-150$  was completely determined.

wide flexible region with an average regional flexibility similar to that of the region from  $+3$  to  $+100$ . Interestingly, regional flexibility in the upstream region seems to gradually decrease in accordance with the distance from the transcription start site (Table 1).

Among the core-less promoters in the present study, the most common sequence for the triplet step from  $-1$  to  $+2$  was found to be CAG. However, this triplet accounted for just 13.3% of the total 1004 sequences. The second and the third most common steps were CAT (5.4%) and CAC (5.3%), respectively. These three sequences are known to be flexible (10). For the triplet step from  $-2$  to  $+1$ , GCA was the most common sequence, accounting for 11.0%. This triplet has also been reported to be flexible. For the triplet step from  $+1$  to  $+3$ , AGT was the most common sequence, although it accounted for only 6.3%. This triplet is known to be highly rigid (10). It is therefore clear that the transcription start sites are not defined by the same triplet. Furthermore, it seems that flexible

triplets tend to be selected at the step from  $-1$  to  $+2$ . In particular, the CA sequence, which is known to increase DNA flexibility (24), often seems to be used within the step to provide the transcription start site with flexibility. In contrast, the triplet step from  $+1$  to  $+3$  and the quartet step from  $+1$  to  $+4$  adopt rigid sequences (Figure 1A and B). This triplet is the third most rigid triplet (the most rigid triplet is from  $-27$  to  $-25$ ; the second most rigid triplet is from  $-26$  to  $-24$ ) and the quartet is the fourth most rigid quartet (the most rigid quartet is from  $-27$  to  $-24$ ; the second most rigid quartet is from  $-473$  to  $-470$ ; the third most rigid quartet is from  $-467$  to  $-464$ ) in the whole region.

In the calculation from the triplet parameters, change in the regional flexibility is marked between upstream (from  $-150$  to  $-1$ ) and downstream (from  $+3$  to  $+100$ ) regions of the transcription start site (Table 1). This result is consistent with the results of a study by Pedersen *et al.* (25), in which a set of human pol II promoters were analyzed, without sorting, using



**Figure 2.** Average flexibility profiles of 382 human promoters containing the GC box only, as calculated from DNase I-derived flexibility parameters (A) or from the tetranucleotide potential energy surface model (B). '+1' corresponds to the transcription start site. Values are shown as means  $\pm$  SD. Flexibility is plotted against the position of the center nucleotide (A) or the second nucleotide (B). The data around the transcription start site are magnified on the right. (C and D) show the average flexibility profiles of the region from -498 to -151 as controls and they were calculated from DNase I-derived flexibility parameters (C) or from the tetranucleotide potential energy surface model (D), respectively.

**Table 1.** The regional flexibility of DNA<sup>a</sup>

Calculated from:	Core-less promoter			GC box-only promoter		
	-498 to -151	-150 to -1	+3 to +100	-498 to -151	-150 to -1	+3 to +100
Triplet parameters ( $\times 10^2$ )	$-2.2 \pm 0.38$	$-2.0 \pm 0.46$	$-1.4 \pm 0.39$	$-2.1 \pm 0.65$	$-1.8 \pm 0.59$	$-1.3 \pm 0.56$
Quartet parameters	$13.63 \pm 0.18$	$13.24 \pm 0.28$	$13.07 \pm 0.20$	$13.45 \pm 0.27$	$13.00 \pm 0.33$	$12.85 \pm 0.28$

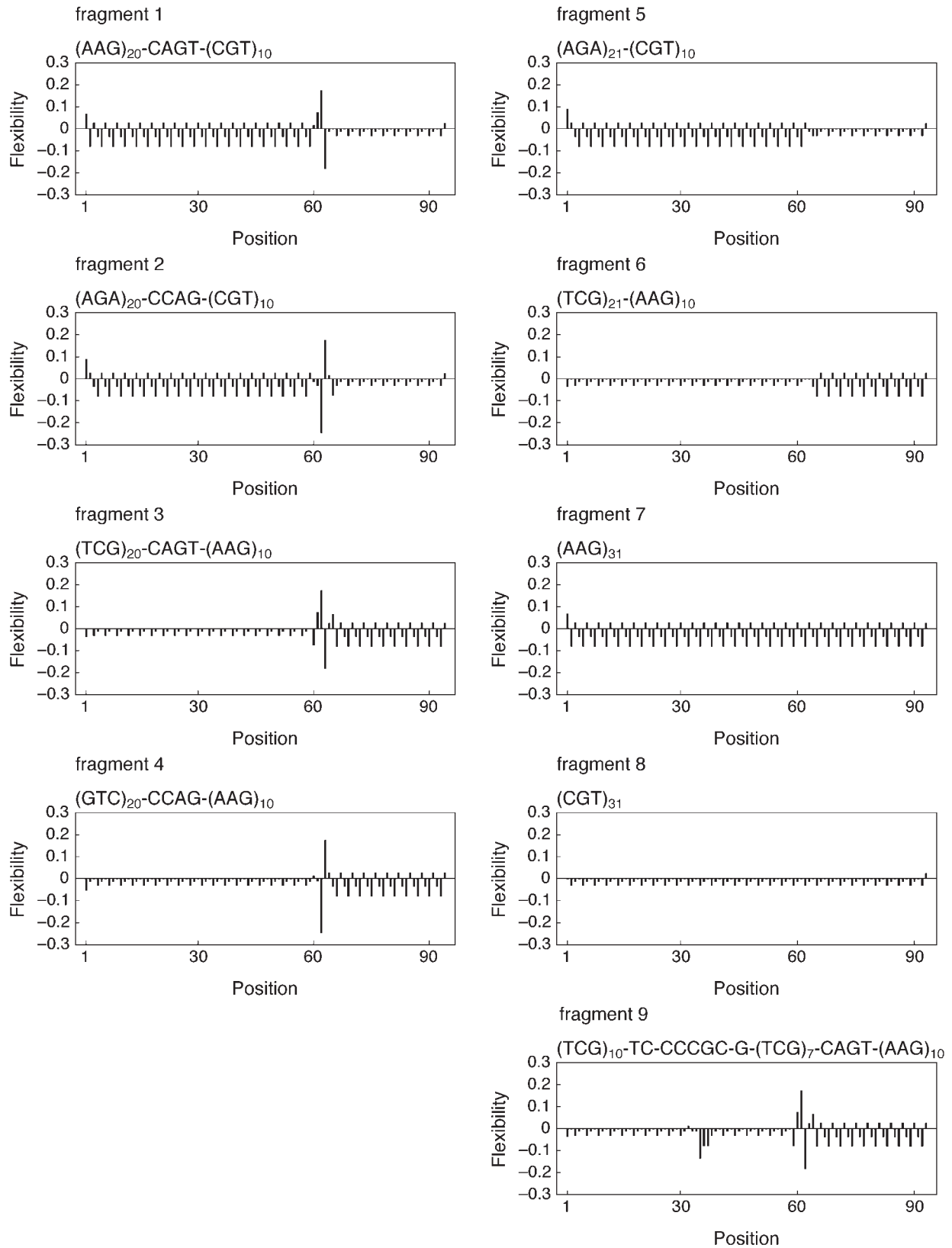
<sup>a</sup>The regional flexibilities were calculated from the mean values of flexibilities of all the triplets or quartets within the indicated regions. Values represent the mean  $\pm$  SD.

a different method. Although in the calculation from the quartet parameters this change is less marked, the flexibility difference between these two regions is also indicated. The average flexibility profiles of the GC box-only promoters were generally very similar to those of the core-less promoters (Figure 2). The most flexible triplet and the most rigid triplet in the whole region are located in the region around the transcription start site. It was also found that the upstream region of the transcription start site is slightly more rigid than the downstream region (also see Table 1). However, the rigid nature of the region around position -25 was less marked. The flexibility profile inherent to the GC box was not

highlighted in the figure. This seems to be because the GC boxes are not confined to a particular position in a promoter.

#### DNA flexibility is a determinant of promoter strength

In order to know whether the mechanical properties of the core-less promoter play some role in transcription, we constructed nine double-stranded DNA fragments (Figure 3), introduced each of them into a vector for luciferase assay, and assayed their promoter activities. According to the DNase I-derived flexibility parameters, the average triplet flexibility of the region from -150 to -1 in Figure 1A was calculated



**Figure 3.** Synthetic double-stranded DNA fragments used as test promoters and their flexibility profiles. Each fragment is indicated by its top strand sequence. The flexibility profiles were calculated using DNase I-derived flexibility parameters. Positions are numbered from the upstream end of each fragment.

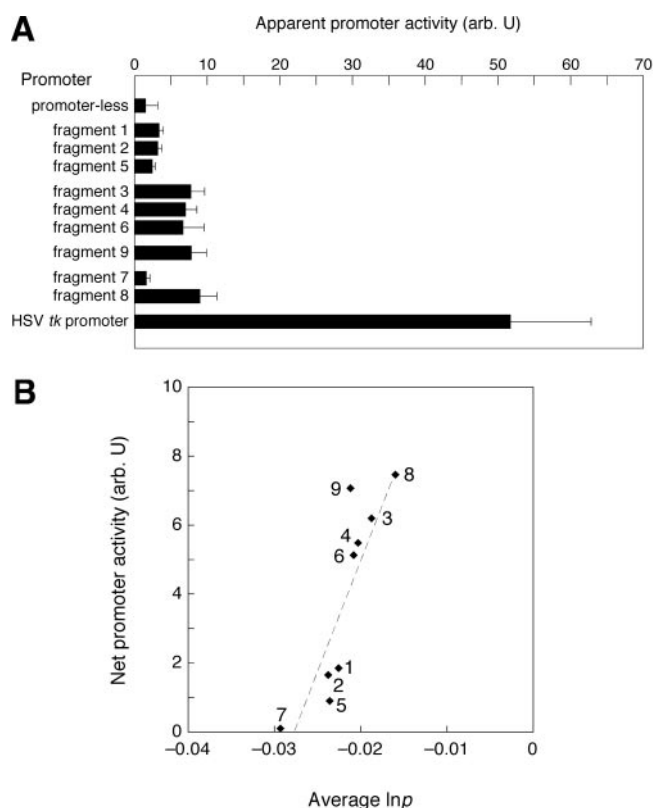
to be  $-0.020$  (the mean value) and that of the region from  $+3$  to  $+100$  was  $-0.014$  (Table 1). In addition, as was described above, the step from  $-1$  to  $+2$  and that from  $+1$  to  $+3$  sometimes adopt CAG and AGT, respectively. Based on this knowledge, we designed test 'promoters'. By inserting the tetranucleotide CAGT, which accounts for 4.3% of the total 1004 sequences (the second most common sequence), between  $(AAG)_n$  and  $(CGT)_n$  sequences, of which average triplet flexibilities are  $-0.030$  and  $-0.016$ , respectively, we prepared fragment 1 which mimics the DNase I-derived average flexibility profile of the core-less promoters (Figures 1A and 3). In fragment 2, CCAG was used instead of CAGT to give the reverse flexibility. The promoter activities of the fragments 1, 2 and 5, which have similar flexibility profiles to each other, were all low (Figure 4A). Although the fragments 3, 4 and 6, in which the upstream 60 bp are on average more flexible than the downstream 30 bp, resemble the core-less promoters in the flexibility profile less than fragments 1, 2 and 5, the former three fragments showed higher activities than the latter three. Furthermore, slight differences could be detected in the former group, i.e. fragments with the tetranucleotide 'partitions' CAGT (fragment 3) and CCAG (fragment 4) generated a slightly positive influence on transcription compared with the fragment with no partition (fragment 6), and it seems

that CAGT was slightly superior to CCAG in this effect. The 'monotonous'  $(AAG)_{31}$  and  $(CGT)_{31}$  sequences gave contrasting results. In particular, the activity of the former (fragment 7) was almost equal to that of the promoter-less construct, indicating that fragment 7 did not function as a promoter. However, the promoter activity of the latter (fragment 8) was calculated to be 14.9% of that of the herpes simplex virus thymidine kinase (HSV *tk*) promoter after correcting for (subtracting) background activity (the activity of the promoter-less construct).

Does the rigid nature of the region around position  $-25$  play a role in transcription? In the TATA-containing promoters, this position is occupied by the TATA box. In order to determine the effect of the slightly rigid nature of this region on transcription, we constructed fragment 9 and its activity was assayed. This fragment was constructed by inserting CCCGC within the TCG-repeats of fragment 3 that showed the second highest activity among the fragments 1–8 and has the 'partition' CAGT. Although fragment 3 mimics the average flexibility profile of the core-less promoters less than fragment 1, the latter was not used to construct fragment 9, as the activity of the latter in promoting transcription was very low. As shown in Figure 4A, modification of the  $-25$  region did not increase transcription levels.

The promoter activities seemed to correlate with the proportion of flexible region in the whole fragment. Thus, the activities were plotted against the average  $\ln p$  [triplet flexibility (10)] in Figure 4B, which clearly showed a correlation. Considering that the activity of fragment 9 was considerably higher than that extrapolated from its flexibility, it is strongly suggested that the rigid sequence around  $-25$  had some positive influence on transcription. Now, we must also discuss the melting ability of each fragment. Generally, the G/C content of DNA is an important factor in its stability, i.e. A/T-rich sequences can melt more easily than G/C-rich sequences. Fragments 1, 2, 5 and 7 (A/T content; 55.3, 54.3, 55.9 and 66.7%, respectively) seem to melt more easily than the fragments 3, 4, 6, 8 and 9 (A/T content; 44.7, 43.6, 44.1, 33.3 and 43.0%, respectively). However, the former fragments were all less effective in transcription than the latter fragments, clearly indicating that the melting potential of DNA did not play a major role in the promotion of transcription. Nevertheless, differences in the regional melting ability of DNA may have a slight effect on transcription, which is discussed later. In conclusion, the transcription assay clarified an important point, i.e. the average flexibility of the whole fragment influences the amount of transcripts produced from a promoter. A general concept is that flexible DNAs wrap around a histone core more easily than rigid DNAs, which usually results in transcriptional repression. Therefore, our data indicate that nucleosomes were not the predominant influence on transcription because more flexible fragments generated a more positive influence on transcription. This raises the question of how more flexible fragments generate more positive effects. One possibility is that formation of the transcription initiation complex may be facilitated more effectively by flexible fragments than by less flexible fragments.

Although the effects generated by CAGT and CCAG were very slight compared with those generated by the whole fragment flexibility, it was suggested that they also played some role in transcription (Figure 4A). Furthermore, Figure 4B



**Figure 4.** Transcription driven by the synthetic DNA fragments. (A) Results of the luciferase assay. The activities of the 'promoter-less' and 'HSV *tk* promoter' were assayed using the vector PGV-B and pST0/TLN (13), respectively. Values represent the mean  $\pm$  SD ( $n = 4$ ). (B) Net promoter activities plotted against average triplet flexibility [ $\ln p$  (10)] of the fragment. The apparent promoter activities shown in (A) were corrected for background activity (the apparent activity of the promoter-less construct). The numerals indicate the fragment number. The broken line is the least squares fit through the mean values of promoter activities of fragments 1–8.

suggested some effect of the rigid sequence around  $-25$ . In an attempt to understand how they affect transcription, we investigated the transcription start site in each construct using various methods. However, we could not detect any definite transcription start site in each fragment (data not shown). For this reason, we speculate that transcription started randomly. Even in the random transcription, the most frequent transcription may have started from a site within the sequences CAGT and CCAG, especially when the rigid  $-25$  sequence was present. However, the low levels of transcription did not allow us to verify this possibility.

The region upstream from the TATA box, Inr sequence or the transcription start site (in the case of core-less promoters and GC box-only promoters) is more rigid than their downstream region [(6) and the present study]. This difference, as well as the distinctive mechanical properties of the TATA box, Inr sequence or the transcription start site, may function as general markers in promoter recognition by transcription factors. The regional features may be recognized at first. Then they may find their target sites by searching the unusual mechanical properties intrinsic to each element or to the transcription start site. After RNA polymerase binds to the promoter,  $\sim 14$  bp of the promoter (between positions  $-11$  and  $+3$ ) melts (26). The difference in the flexibility between the upstream and downstream regions may also be involved in this process. Furthermore, the distinctive mechanical properties given to the 'junction' (the TATA box, Inr sequence or the transcription start site) may facilitate DNA melting. The slight positive effects of the CAGT and CCAG tetranucleotide sequences on transcription (Figure 4) can be explained in terms of this effect. Indeed, in the case of the TATA box, its mechanical properties seem to enable the sequence to unwind (6,27,28).

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions of Teruaki Motoyama, Shinya Inoue and Junko Ohyama. This study was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan to T.O. Funding to pay the Open Access publication charges for this article was provided by MEXT.

*Conflict of interest statement.* None declared.

## REFERENCES

- 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.
- Asayama,M. and Ohyama,T. (2005) Curved DNA and prokaryotic promoters: a mechanism for activation of transcription. In Ohyama,T. (ed.), *DNA Conformation and Transcription*. Landes Bioscience, Texas and Springer, NY, pp. 37–51.
- Ohyama,T. (2005) Curved DNA and transcription in eukaryotes. In Ohyama,T. (ed.), *DNA Conformation and Transcription*. Landes Bioscience, Texas and Springer, NY, pp. 66–74.
- 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.
- 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.
- Fukue,Y., Sumida,N., Nishikawa,J. and Ohyama,T. (2004) Core promoter elements of eukaryotic genes have a highly distinctive mechanical property. *Nucleic Acids Res.*, **32**, 5834–5840.
- Cavin P erier,R., Junier,T. and Bucher,P. (1998) The eukaryotic promoter database EPD. *Nucleic Acids Res.*, **26**, 353–357.
- Schmid,C.D., Praz,V., Delorenzi,M., P erier,R. and Bucher,P. (2004) The eukaryotic promoter database EPD: the impact of *in silico* primer extension. *Nucleic Acids Res.*, **32**, D82–D85.
- 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.
- Br ukner,I., S anchez,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.
- Packer,M.J., Dauncey,M.P. and Hunter,C.A. (2000) Sequence-dependent DNA structure: tetranucleotide conformational maps. *J. Mol. Biol.*, **295**, 85–103.
- Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd edn*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Nishikawa,J., Amano,M., Fukue,Y., Tanaka,S., Kishi,H., Hirota,Y., Yoda,K. and Ohyama,T. (2003) Left-handedly curved DNA regulates accessibility to *cis*-DNA elements in chromatin. *Nucleic Acids Res.*, **31**, 6651–6662.
- Breathnach,R. and Chambon,P. (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.*, **50**, 349–383.
- Carey,M. and Smale,S.T. (2000) *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 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.
- 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.
- 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<sub>1160</sub> of *Drosophila*. *Genes Dev.*, **11**, 3020–3031.
- Kadonaga,J.T., Courey,A.J., Ladika,J. and Tjian,R. (1988) Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science*, **242**, 1566–1570.
- Mitchell,P.J. and Tjian,R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, **245**, 371–378.
- Hahn,S., Buratowski,S., Sharp,P.A. and Guarente,L. (1989) Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc. Natl Acad. Sci. USA*, **86**, 5718–5722.
- 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.
- 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.
- Lyubchenko,Y.L., Shlyakhtenko,L.S., Appella,E. and Harrington,R.E. (1993) CA runs increase DNA flexibility in the complex of lambda cro protein with the O<sub>R</sub>3 site. *Biochemistry*, **32**, 4121–4127.
- 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.
- Ebright,R.H. (1998) RNA polymerase–DNA interaction: structures of intermediate, open, and elongation complexes. *Cold Spring Harb. Symp. Quant. Biol.*, **63**, 11–20.
- 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.
- 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.