Structural property of regulatory elements in human promoters

Xiao-Qin Cao,* Jia Zeng,† and Hong Yan‡
Department of Electronic Engineering, City University of Hong Kong, Tat Chee Avenue 83, Hong Kong
(Received 19 November 2007; published 15 April 2008)

The capacity of transcription factors to activate gene expression is encoded in the promoter sequences, which are composed of short regulatory motifs that function as transcription factor binding sites (TFBSs) for specific proteins. To the best of our knowledge, the structural property of TFBSs that controls transcription is still poorly understood. Rigidity is one of the important structural properties of DNA, and plays an important role in guiding DNA-binding proteins to the target sites efficiently. After analyzing the rigidity of 2897 TFBSs in 1871 human promoters, we show that TFBSs are generally more flexible than other genomic regions such as exons, introns, 3′ untranslated regions, and TFBS-poor promoter regions. Furthermore, we find that the density of TFBSs is consistent with the average rigidity profile of human promoters upstream of the transcription start site, which implies that TFBSs directly influence the promoter structure. We also examine the local rigid regions probably caused by specific TFBSs such as the DNA sequence TATA(A/T)A(A/T) box, which may inhibit nucleosomes and thereby facilitate the access of transcription factors bound nearby. Our results suggest that the structural property of TFBSs accounts for the promoter structure as well as promoter activity.

DOI: 10.1103/PhysRevE.77.041908 PACS number(s): 87.14.gk, 87.14.gf

I. INTRODUCTION

Based on a cap analysis of gene expression, mammalian promoters can be broadly classified into the DNA sequence TATA(A/T)A(A/T) (TATA)-rich and cytosine and guanine separated by a phosphate (CpG)-rich promoters, and different tissues and families of genes differentially use distinct types of promoters [1]. In addition to sequential motifs such as the TATA box and CpG island, the physical or structural properties of DNA also have important inhibitory or facilitatory roles in regulating gene expression [2–18]. Rigidity is an essential DNA structural property, which previously has been found to be able to guide DNA-binding proteins efficiently to the target sites [12]. Recently, the rigidity of eukaryotic promoters has been extensively examined [7,13–16,18], and it has been suggested that the rigidity may influence DNA looping [3,19], promoter activities [13,14], nucleosome positioning [7,20,21], and transcription factor binding [5,6,13,14,22].

As shown in Fig. 1, the average rigidity profile of human promoters [extracted from the eukaryotic promoter database (EPD) [23]] aligned at the transcription start site (TSS) displays highly distinctive structural properties, which have been summarized as three features [7,13,14,16]: (1) The upstream region of the TSS is slightly more rigid than its downstream region; (2) both TATA and initiator (Inr) boxes contain highly flexible and highly rigid triplets in their upstream and downstream halves, and the core promoters upstream of a TATA or Inr box are slightly more rigid than in its downstream regions; (3) the region around 28 base pairs (bp) upstream of the TSS is relatively rigid in both core and coreless promoters.

*xiaoqcao@cityu.edu.hk
†j.zeng@ieee.org
‡Also at School of Engineering and Information Engineering, University of Sydney, NSW 2006, Australia. h.yan@cityu.edu.hk

The TATA box has a consensus TATA(A/T)A(A/T) sequence located about 25–31 bp upstream of the TSS. The Inr box has a consensus YYANWYY [Y denotes pyrimidine (C, T, or U); N denotes any base (A, C, G, T, or U); W denotes T, U, or A] sequence starting at the TSS. The GC box usually has a consensus GGGCGG sequence as the binding site for the transcription factor Sp1 [24]. These elements are generally known as core promoter elements of class II genes. The EPD [23] contains 1871 nonredundant human promoter sequences, in which TATA-only, Inr-only, and GC-only promoters account for about 6%, 9%, and 20%, respectively [14]. Hence, more than half of the human promoters are coreless promoters.

A general concept is that flexible DNA sequence segments wrap around a histone core more easily than rigid ones, which usually results in transcription repression [21,25,26]. Pedersen *et al.* [7], in an attempt to explain the first feature of the average rigidity profile of promoters, suggested that the more flexible downstream region of the TSS contains periodic flexible triplet pairs CAG-CTG and GGC-GCC, which have been found to correlate with nucleosomes forming to inhibit interactions with DNA-binding proteins. On the other hand, Fukue *et al.* [13,14] found that nucleosomes are not the predominant influence on transcription because more flexible promoter fragments generate a more positive influence on transcription. Furthermore, they experimentally verified that the rigid region around 28 bp upstream of the TSS may have some positive influence on transcription, from experiments on synthetic DNA sequences. Tirosh *et al.* [18] compared DNA rigidity among different yeast species and found that the localized rigid DNA is a general sequence property in yeast promoters, which could influence nucleosome positioning and assist in the assembly of the transcriptional machinery at TATA-less promoters.

Although several characteristics of eukaryotic transcription factor binding sites (TFBSs) have been investigated [24], to the best of our knowledge, the general relationship between the regulatory elements and human promoter structure is still poorly understood. Promoter sequences are usu-

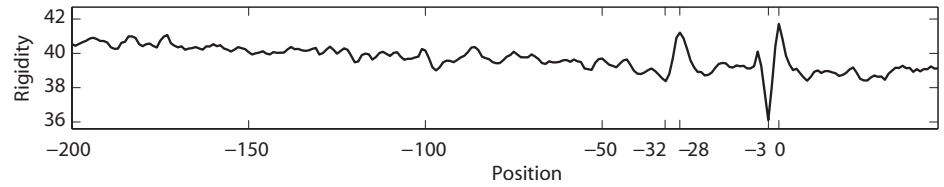

FIG. 1. Average rigidity profile of 1809 human promoters as calculated from the tetranucleotide potential energy surface model by Packer *et al.* [9]. Position 0 corresponds to the TSS. Rigidity is plotted against the position of the start nucleotide. For example, the quartet from −28 to −25 is plotted against −28 (it is the most rigid local quartet in the analyzed region). Four positions, −32, −28, −3, and 0, have a highly distinctive mechanical property in proximal promoters. The upstream side of the TSS is slightly more rigid than its downstream regions.

ally composed of short motifs that function as TFBSs. However, the presence of a particular motif is not sufficient to ensure regulation by the associated transcription factors, which must access and recognize TFBSs within the chromatin. TFBSs should compose *cis*-regulatory modules to enable their interaction with other regulatory proteins. Generally, DNA-binding proteins seem to move from random to specific sites by multiple dissociation and association events within a single DNA molecule [12,27]. During these events, some structural property of DNA may assist DNA-binding proteins in finding the target sites efficiently [28,29]. Therefore, we speculate that the rigidity of TFBSs is one of the most important structural properties to influence human promoter structure and activity. After analyzing the rigidity of 2897 TFBSs in 1871 human promoters, we report that TFBSs are more flexible, which accounts for the promoter structure as well as promoter activity according to the nucleosome positioning model. This notion is further supported by finding certain TFBSs that contain triplet pairs ATA-TAT, TAA-TTA, AAA-TTT, and AAT-ATT, which inhibit nucleosome formation.

II. MATERIALS AND METHODS

A. Promoter, TFBS, and other genomic sequences

All 1871 human promoter sequences from the EPD release 92 [23] are downloaded. We extract the sequence segment from −1000 to +1000 as well as the segment from −200 to +50 relative to the TSS at the position 0. The range $[-1000,+1000]$ provides an overview of the vicinity of the promoter, and the range $[-200,+50]$ is generally enriched with TFBSs characterizing promoter regions.

A collection of 2897 human TFBS sequences from the database for transcription factors and their genomic binding sites (TRANSFAC) [30] is also downloaded. The length of the TFBS varies from 2 to 240. We focus on calculating the rigidity of 841 n-mers $(4 \le n \le 15)$ in TFBSs because the length of most TFBSs is between 4 and 15.

To compare the structure of TFBSs with that of other genomic regions, we randomly extract n-mers $(4 \le n \le 15)$ from 28 272 human exon and 27 068 intron sequences in the exon and intron database (EID) [31]. Similarly, we randomly extract n-mers $(4 \le n \le 15)$ from 30 946 human 3′ untranslated region (UTR) sequences in the databases of sequences and functional elements of 5′ and 3′ UTRs (UTRdb) of eukaryotic mRNAs [32]. In addition, we also randomly extract n-mers $(4 \le n \le 15)$ from promoter regions that are known not to be enriched with TFBSs according to the TRANSFAC database [30]. As a summary, we focus on comparing the structural properties of TFBSs with those of exons, introns, 3′UTRs, and TFBS-poor promoter regions.

B. Tetranucleotide DNA rigidity model

The trinucleotide model that determines the 32 trinucleotide steps based on DNase I cutting frequencies [4] and the tetranucleotide model from the potential energy surface model [9] have been widely used for calculating DNA rigidity profiles. For example, they have been used to analyze the structural properties of human polymerase II (pol II) promoters [7,13,14], mammalian and plant genomes [16], yeast promoters [18], and prokaryotes and eukaryotes [15]. Because the tetranucletide incorporates more genetic context information than do trinucleotide-based descriptions, we adopt it to calculate rigidity profiles of TFBSs and human pol II promoters.

In the tetranucleotide model, slide and shift are the two principal degrees of freedom compared with twist, roll, tilt, and rise. Slide and shift cannot be predicted because they are strongly correlated in neighboring steps, so the conformational energy E_{step} of a dinucleotide step is a function of slide and shift. This function is used in conjunction with the experimental data on the conformations of tetranucleotides to parametrize an energy function E_{junction}, which couples slide and shift in all three steps,

$$E_{\text{junction}} = (\Delta D_y)^2 F_{\Delta D_y} + \left(\sum D_y\right)^2 F_{\Sigma D_y} + (\Delta D_x)^2 \times F_{\Delta D_x} + \left(\sum D_x\right)^2 F_{\Sigma D_x}, \quad (1)$$

where D_y is slide, D_x is shift, and $F_{\Delta D_y}$, $F_{\Sigma D_y}$, $F_{\Delta D_x}$, and $F_{\Sigma D_x}$ are the force constants. The energy of an oligomer of N base pairs is

$$E^N_{\text{oligomer}} = \sum_{n=1}^{N-1} E^n_{\text{step}} + \sum_{n=1}^{N-2} E^n_{\text{junction}}. \quad (2)$$

The rigidity values of all tetranucleotides are calculated from the curvature in the tetranucleotide potential energy surface with respect to slide at the global energy minimum. As a result, a lookup table of 136 tetranucleotide rigidity values is obtained ranging from 1.9 to 27.2. Although the potential energy surface model is a rough approximation for complex, statistically derived properties like conformational preferences, it agrees well with the experimental data from x-ray crystal structures.

At each position of the DNA sequence, we calculate the rigidity value based on the 6-mer (six-base-long sequence). While the 6-mer may reflect only a local pattern, it is a practical choice to characterize the sequence-dependent rigidity demonstrated in the promoter studies [16]. The rigidity of the 6-mer is calculated by summing up the rigidity values of three overlapping component tetranucleotides,

$$R = \sum_{i=1}^{3} r_i, \tag{3}$$

where i is the position index. We calculate the rigidity of the 6-mer against its start position. For example, the 7-mer TATAAAA has rigidity at the first position T

$$R_{\mathrm{T}} = r_{\mathrm{TATA}} + r_{\mathrm{ATAA}} + r_{\mathrm{TAAA}}, \tag{4}$$

and rigidity at the second position A

$$R_{\mathrm{A}} = r_{\mathrm{ATAA}} + r_{\mathrm{TAAA}} + r_{\mathrm{AAAA}}. \tag{5}$$

Therefore, it is possible to calculate rigidity profiles for any given sequences based on 136 unique tetranucleotide parameters. If a sequence is L in length, its rigidity profile is $L-5$ in length according to the above 6-mer rigidity model. Higher values correspond to more rigid sequences, and lower values correspond to more flexible sequences.

Thermodynamic conditions such as the salt concentration (solution environment) may change DNA conformations so as to influence target site localization by site-specific DNA-binding proteins. Schlick *et al.* [33] computationally studied the influence of salt on the structure and energetics of supercoiled DNA. At a high salt concentration, the DNA adopts highly compact and bent interwound states with the bending energy dominating over the other components. At a low salt concentration, the DNA supercoils are much more open and loosely interwound and the electrostatic components are dominant. In the study of DNA-protein binding schemes [27], the test protein stays on the DNA as it travels between sites at low salt when sites are less than 50 bp apart. Transfers greater than 30 bp at *in vivo* salt concentration and over distances greater than 50 bp at any salt concentration always include at least one dissociation step. Thus, the very low salt concentration facilitates long-range sliding by reducing the protein's dissociation rate. The above observations are consistent with the result that lower salt concentration may enhance DNA rigidity [2,34]. Indeed, the tetranucleotide model takes into account cooperative structural transitions induced by the salt concentration because the cooperative nature of the process means that the conformations of the dinucleotide building blocks of the tetranucleotide must be strongly coupled, which has been encoded in Eqs. (1) and (2).

III. FLEXIBLE TFBS

We calculate average rigidity profiles of 841 n-mers ($4 \le n \le 15$) in human TFBSs extracted from TRANSFAC [30]. To compare rigidity with other genomic regions, we calculate the rigidity of randomly selected n-mers ($4 \le n \le 15$) in exons, introns from EID [31], 3′UTRs from UTRdb [32], and other TFBS-poor promoter regions from EPD [23].

TABLE I. Average rigidity of n-mers ($4 \le n \le 15$) in TFBSs, exons, introns, 3′UTRs, and other promoter regions.

n	TFBSs	Exons	Introns	3′UTRs	Other promoter regions
4	12.3	13.5	14.2	13.7	13.6
5	26.0	27.0	28.2	27.7	27.2
6	38.0	40.5	41.9	41.5	40.8
7	53.1	55.1	54.6	54.9	54.3
8	64.5	68.4	68.6	68.7	67.9
9	78.8	81.5	82.3	82.6	81.5
10	98.7	96.3	96.1	96.3	95.1
11	111.0	110.1	109.7	110.3	108.7
12	123.4	124.0	123.5	126.4	122.3
13	137.9	137.7	137.1	138.8	135.9
14	149.5	151.5	150.8	153.1	149.4
15	165.3	165.4	164.5	166.0	163.0

Table I shows the average rigidity of n-mers ($4 \le n \le 15$) in TFBSs, exons, introns, 3′UTRs, and other TFBS-poor promoter regions. We observe that TFBSs are generally more flexible compared to exon, intron, 3′UTR, and TFBS-poor promoter sequences. The n-mers ($4 \le n \le 9$) in TFBSs are significantly more flexible on average than other genomic regions. However, there is only a slight rigidity difference of n-mers ($10 \le n \le 15$) between TFBSs and other genomic regions. Some n-mers ($n=10,11$) in TFBSs are even more rigid than other genomic regions. Nevertheless, taking into consideration that longer TFBSs often occur less frequently than shorter ones, we conclude that TFBSs are generally more flexible regions in human promoters.

The idea of more flexible TFBSs agrees well with the previous study of the DNA structure and DNA-protein interactions, in which Packer *et al.* [9] showed that the top three flexible tetranucleotides involve CA-TG and TA-TA flanking steps. As we know, the TATA box is one of the widely recognized TFBSs, and it contains the most flexible flanking steps TA-TA. Furthermore, the flexible sequences are predicted to be significantly more active than rigid ones, so that they interact easily with transcription factors [5,6,13,14]. Here our results demonstrate that TFBSs are generally more flexible than other genomic regions so that they can be easily accessed and bound by transcription factors. Furthermore, the concept that TFBSs are more flexible supports the view that the promoter activity correlates with the proportion of flexible regions in the whole fragment [14], because the more active promoters may contain enriched TFBSs that in turn increase the overall flexibility of the sequence.

IV. TFBS INFLUENCES PROMOTER STRUCTURE

As shown in Fig. 2, the promoter structure can be reflected by its average rigidity profile, which includes three highly distinctive structural properties at the positions −28, −3, and 0 (the TSS). In contrast to the previous observations [7,13,14] that the upstream region of the TSS is slightly more

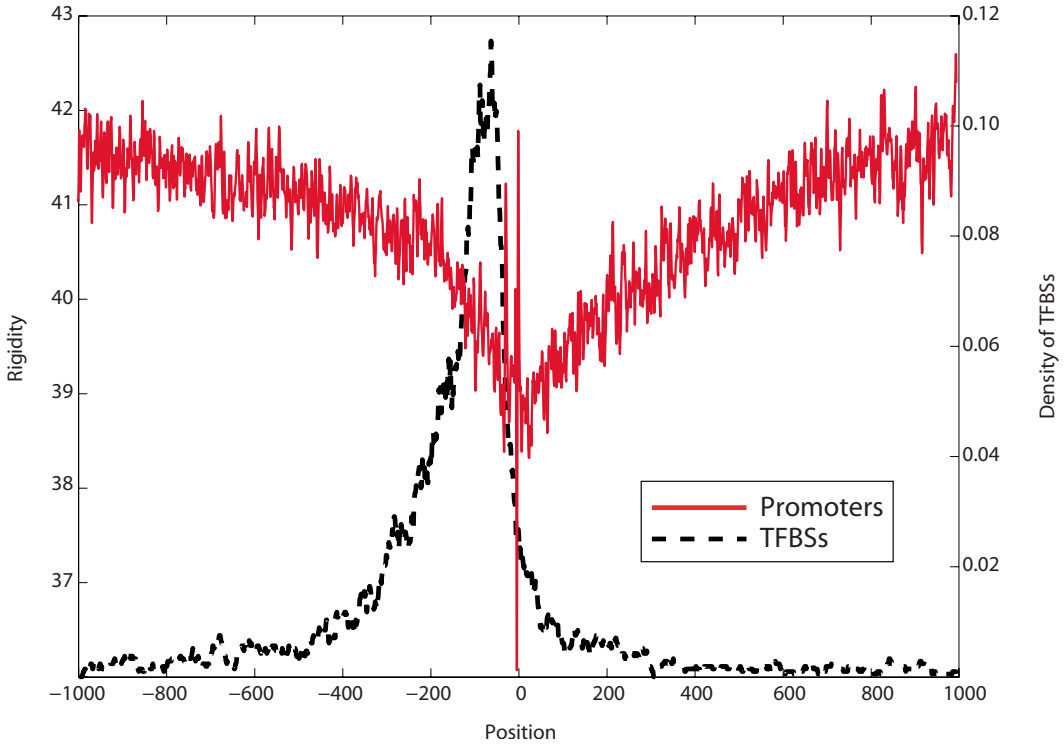

FIG. 2. (Color online) Average rigidity profile of 1809 human promoters as calculated from the tetranucleotide potential energy surface model by Packer *et al.* [9], and the density of TFBSs as calculated based on TRANSFAC including the exact position of 2897 human TFBSs from −1000 to +1000 relative to the TSS at the position 0. Following a highly flexible region, the TSS is the most rigid region, consistent with Fig. 1. The proximal region around the TSS is much more flexible than its upstream and downstream regions such as −1000 and +1000. TFBSs are greatly enriched around the position −75, and this implies that TFBSs may account for the decrease in the upstream rigidity profile around the TSS.

rigid than the downstream one, we observe that both upstream and downstream regions of the TSS are significantly more rigid than the region around the TSS in a broader genetic context from −1000 to +1000 bp relative to the TSS, as shown in Fig. 2. The entire average rigidity profile has a deep V shape with the valley around the TSS. In this valley, we see two significantly rigid regions at the positions −28 and 0 as well as one significantly flexible region at the position −3. Such a deep V shape of an average rigidity profile was also found in yeast promoters [18]. Therefore, the deep V shape of the promoter structure may exist in all eukaryotes.

To investigate the relationship between TFBSs and the promoter structure, we plot the density of TFBSs (black line) in Fig. 2 based on the start and end sites of each TFBS from TRANSFAC [30]. Figure 2 shows that the density of TFBSs is consistent with the average rigidity profile of promoters. In the upstream region of the TSS, the rigidity profile drops when the density profile rises, though the peak of the density profile does not exactly locate the valley of the rigidity profile. Generally, TFBSs are enriched around the position −75 relative to the TSS, where the average rigidity profile has a strong tendency to decrease. Hence we speculate that the structural properties of TFBSs directly influence the average rigidity profile of promoters. According to Table I, TFBSs are more flexible than other genomic regions, which best explains why the rigidity drops upstream of the TSS while the density of TFBSs rises. Therefore, we conclude that the promoter structure is correlated with enrichment of TFBSs. However, the density of TFBSs continues to decrease after position −75, while the rigidity profile still remains at a low rigidity level between the positions −75 and 0. This discrepancy may be caused by some significantly flexible TFBSs even when their density is low. After position 0, the rigidity profile increases with decrease of the density profile. Rather than TFBSs, the flexible region immediately downstream of the TSS is caused mainly by the periodic triplet pairs CAG-CTG and GGC-GCC, which previously have been found to correlate with nucleosome positioning [7].

To understand the relationship between TFBSs and the promoter structure, we illustrate the consensus motifs at four positions −32, −28, −3, and 0 (marked by circles in Fig. 3) in the proximal promoter region $[-200,+50]$. We find that these four positions all correlate with certain TFBSs. In Fig. 3, the consensus motif logos are based on the nucleotide content information. It is the flexible TA-TA and CA-TG flanking steps that result in the local flexible regions at the positions −32 and −3. Similarly, it is the rigid AA-TT and AT-TT flanking steps [9] that lead to the local rigid regions at the positions −28 and 0. Table II shows the top ten most frequent TFBSs and their rigidity values at four positions −32, −28, −3, and 0. We see that the most frequent TFBS at positions −32 and −28 is the canonical TATA box (TATAAAA). Indeed, the TATA box contains half flexible flanking steps TA-TA and half rigid steps AA-AA. The rigidity of the TATA box is slightly lower than the average rigid-

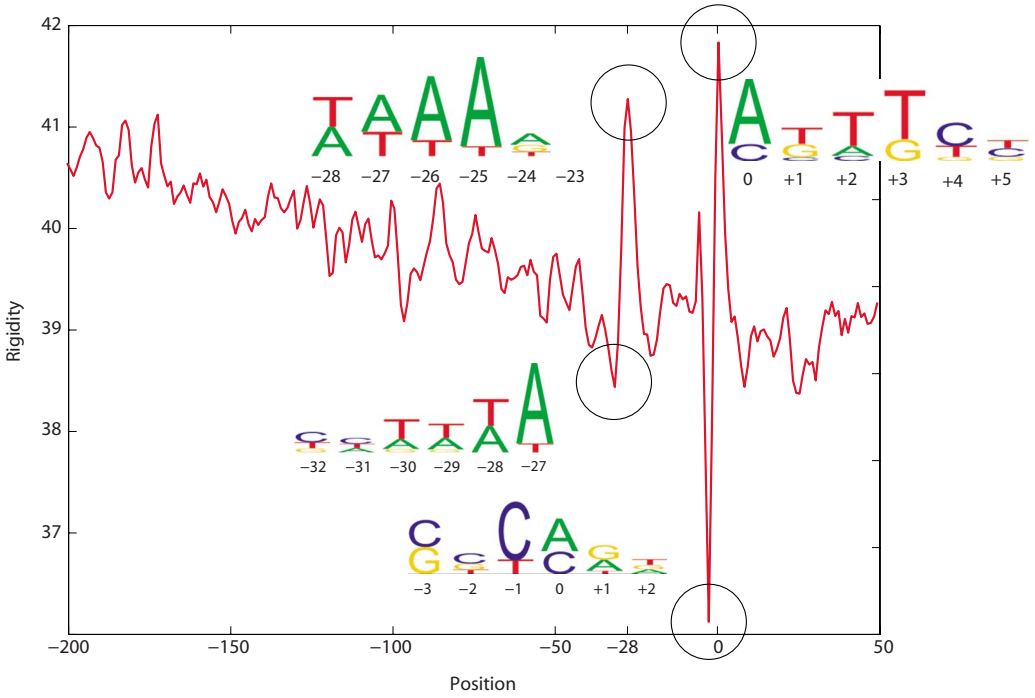

FIG. 3. (Color online) Average rigidity profile of promoters and consensus motifs at the positions −32, −28, −3, and 0. The sequence logo for consensus motifs constructed based on nucleotide information content. The position −32 is a local flexible region correlating with the top flexible TA-TA flanking steps. The position −28 is a local rigid region correlating with the most rigid AA-TT flanking steps. The position −3 is a local flexible region correlating with the top flexible CA-TG flanking steps. The position 0 is a local rigid region correlating with the rigid AT-TT steps.

ity of 7-mers of TFBSs in Table I, which best explains why the entire rigidity profile of promoters has a significantly rigid localized peak while still keeping a low rigidity level around the position −28 in Fig. 1. From Table II we see that there are several more rigid TFBSs on average around the position −28 such as CATTT, GCCCC, and ATTGG. These rigid TFBSs may contribute to the local rigid region, which was suggested to be strongly correlated with promoter activities [14,18]. The most frequent TFBS in the region [−3,0] is CATTT, which is also the consensus motif in Fig. 3. Similarly to the TATA box, half of CATTT is relatively flexible and the other half is relatively rigid, which is consistent with the significant rigidity step in the region [−3,0] in Fig. 3. In addition, among the top three most frequent TFBSs, there are two which have higher rigidity than the average value in Table I. This observation is also consistent with the truth that TSS is the most rigid region at the position 0.

TABLE II. Rigidity of TFBSs at the positions −32, −28, −3, and 0. TFBSs are ranked according to their frequency at these positions.

Position −32∼−28		Position −3∼0	
TATAAAA	52.2	CATTT	34.4
TATAT	12.2	GCGA	7.2
GGCGG	20.0	CAGTTG	44.9
TTCC	10.6	TTCC	10.6
GCGA	7.2	CTGTC	19.6
CGCGG	21.4	CCACC	16.6
CATTT	34.4	CGCGG	21.4
GCCCC	36.0	CATTA	27.4
GGTGG	16.6	GGCGG	20.0
ATTGG	35.4	TTGAA	24.1

In conclusion, our results show that the structural properties of the TFBSs influence both the global and local rigidity profiles of promoters.

V. TFBS FOR PROMOTER ACTIVITY

Pedersen *et al.* [7] suggested that the upstream region is more rigid than the downstream region of the TSS so that the upstream rigidity inhibits nucleosome formation to facilitate transcription factor binding. Our results indicate that the upstream and downstream of the TSS are both significantly more rigid than the region around the TSS with a deep V shape in Fig. 1. The obvious discrepancy between our observation and that of Pedersen *et al.* [7] suggests that it is the local rigidity, such as rigidity at the position −28, rather than global rigidity of the upstream region that plays an important role in destabilizing nucleosomes so as to direct the transcriptional machinery.

To investigate the relationship between the promoter structure and its activity, Fukue *et al.* [14] carried out experiments on nine synthetic promoter fragments and modified the −28 region in the ninth promoter fragment by inserting the rigid sequence CCCGC. They found that the promoter activities seemed to correlate with the proportion of flexible regions in the whole fragment, and, more importantly, the rigid sequence at the position −28 had some positive influ-

ence on transcription. Our results can best support their findings. Obviously, the flexible upstream region of the TSS is enriched with TFBSs that are signals for transcription factors to regulate gene expression. TFBSs are generally flexible sequences, and thus formation of the transcription initiation complex may be easier for flexible than for rigid promoter fragments. We hypothesize that the local rigidity at the position −28 destabilizes nucleosomes to facilitate transcription factor binding; this also supports the finding of [14] of a positive influence of the rigid sequence at the position −28. Therefore, the structural property of TFBSs can account for the promoter activity in terms of nucleosome positioning.

Indeed, Segal *et al.* [26] has verified *in vivo* that low nucleosome occupancy is encoded at functional TFBSs and TSSs. Genomes use their intrinsic nucleosome organization to encode stable nucleosomes over nonfunctional sites, thereby decreasing their accessibility to transcription factors. In addition, the TATA boxes locate in the areas of the genomic sequences that remain unoccupied by nucleosomes; that is, just outside a stably positioned nucleosome. Our results show that the canonical TATAAAA contains half of the flexible tetranucleotide TATA and half of the rigid tetranucleotide AAAA. It is possible that the rigid half AAAA inhibits the nucleosome formation and directs the transcription factor to bind the flexible half region TATA. From Fig. 3, the consensus motifs at positions −32, −28, −3, and 0 contain rigid triple pairs ATA-TAT, TAA-TTA, AAA-TTT, and AAT-ATT that inhibit nucleosomes and thereby facilitate the access of transcription factors bound nearby [25]. Therefore, some TFBSs may play a dual role that on the one hand they inhibit the nucleosome formation with the rigid sequence and on the other hand they interact with transcription factors by the flexible sequence.

VI. CONCLUSIONS

This paper reports on the structural properties of regulatory elements such as TFBSs in human promoters. Generally TFBSs tend to be more flexible than other genomic regions. This structural property can account for the deep V-shape promoter structure. Moreover, the local rigid region at position −28 is enriched by certain rigid TFBSs, which may inhibit nucleosome formation to direct transcription factor binding. We examine the density profile of TFBSs and the average rigidity profile of promoters, and find that TFBSs result in local rigidity or flexibility of the promoter sequence. In particular, we find several rigid TFBSs at the local rigid region at −28, which was previously suggested to correlate closely with the promoter activity according to the nucleosome positioning model [25].

ACKNOWLEDGMENTS

This work is supported by Hong Kong Research Grant Council (Project No. CityU 122607) and City University of Hong Kong (Project No. 9610034).

[1] P. Carninci *et al.*, Nat. Genet. **38**, 626 (2006).
[2] N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, Phys. Rev. Lett. **85**, 2400 (2000).
[3] K. S. Matthews, Microbiol. Rev. **56**, 123 (1992).
[4] I. Brukner, R. Sánchez, D. Suck, and S. Pongor, EMBO J. **14**, 1812 (1995).
[5] D. Starr, B. Hoopes, and D. Hawley, J. Mol. Biol. **250**, 434 (1995).
[6] A. Grove, A. Galeone, L. Mayol, and E. Geiduschek, J. Mol. Biol. **260**, 120 (1996).
[7] A. G. Pedersen, P. Baldi, Y. Chauvin, and S. Brunak, J. Mol. Biol. **281**, 663 (1998).
[8] C. Anselmi, G. Bocchinfuso, P. De Santis, M. Savino, and A. Scipioni, J. Mol. Biol. **286**, 1293 (1999).
[9] M. Packer, M. Dauncey, and C. Hunter, J. Mol. Biol. **295**, 85 (2000).
[10] L. Tsai and L. Luo, J. Theor. Biol. **207**, 177 (2000).
[11] A. Kanhere and M. Bansal, Nucleic Acids Res. **31**, 2647 (2003).
[12] D. M. Gowers and S. E. Halford, EMBO J. **22**, 1410 (2003).
[13] Y. Fukue, N. Sumida, Jun-ichi Nishikawa, and T. Ohyama, Nucleic Acids Res. **32**, 5834 (2004).
[14] Y. Fukue, N. Sumida, Jun-ichi Tanase, and T. Ohyama, Nucleic Acids Res. **33**, 3821 (2005).
[15] A. Kanhere and M. Bansal, Nucleic Acids Res. **33**, 3165 (2005).
[16] K. Florquin, S. Degroeve, Y. Saeys, and Y. Van de Peer, Nucleic Acids Res. **33**, 4255 (2005).
[17] R. Paparcone, S. Morosetti, A. Scipioni, and P. D. Santis, Biophys. Chem. **120**, 71 (2006).
[18] I. Tirosh, J. Berman, and N. Barkai, Trends Genet. **23**, 318 (2007).
[19] J. M. G. Vilar and L. Saiz, Phys. Rev. Lett. **96**, 238103 (2006).
[20] T. Sakaue and H. Löwen, Phys. Rev. E **70**, 021801 (2004).
[21] W. Mobius, R. A. Neher, and U. Gerland, Phys. Rev. Lett. **97**, 208102 (2006).
[22] H. Diamant and D. Andelman, Phys. Rev. E **61**, 6740 (2000).
[23] C. D. Schmid, R. Perier, V. Praz, and P. Bucher, Nucleic Acids Res. **34**, D82 (2006); http://www.epd.isb-sib.ch/
[24] M. Carey and S. T. Smale, *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000).
[25] H. E. Peckham, R. E. Thurman, Y. Fu, J. A. Stamatoyannopoulos, W. S. Noble, K. Struhl, and Z. Weng, Genome Res. **17**, 1170 (2007).
[26] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I. K. Moore, J.-P. Z. Wang, and J. Widom, Nature (London) **442**, 772 (2006).
[27] D. M. Gowers, G. G. Wilson, and S. E. Halford, Proc. Natl. Acad. Sci. U.S.A. **102**, 15883 (2005).
[28] T. Hu and B. I. Shklovskii, Phys. Rev. E **74**, 021903 (2006).
[29] T. Hu and B. I. Shklovskii, Phys. Rev. E **76**, 051909 (2007).

[30] V. Matys *et al.*, Nucleic Acids Res. **34**, D108 (2006).
[31] S. Saxonov, I. Daizadeh, A. Fedorov, and W. Gilbert, Nucleic Acids Res. **28**, 185 (2000).
[32] G. Pesole, S. Liuni, G. Grillo, F. Licciulli, F. Mignone, C. Gissi, and C. Saccone, Nucleic Acids Res. **30**, 335 (2002).
[33] T. Schlick, B. Li, and W. K. Olson, Biophys. J. **67**, 2146 (1994).
[34] K.-K. Kunze and R. R. Netz, Phys. Rev. E **66**, 011918 (2002).