Dynamic flexibility in the *Escherichia coli* genome

Lu Tsai[a,b], Zhirong Sun[a,*]

[a] *Department of Biological Science and Technology, Tsinghua University, Beijing 100084, PR China*
[b] *Department of Biological and Chemical Engineering, Baotou University of Iron and Steel Technology, Baotou 014010, PR China*

Received 7 June 2001; revised 25 September 2001; accepted 26 September 2001

First published online 11 October 2001

Edited by Takashi Gojobori

Abstract Empirical rules based on tetranucleotide parameters were presented to predict the structural parameters twist (Ω), roll (ρ), tilt (τ) and slide (D_y). A statistical mechanical model was used to analyze the flexibility of the *Escherichia coli* genome. The replication terminus region displayed a low level of flexibility. A strong correlation can be seen between G+C content and flexibility. Average flexibilities in the coding regions were found to be significantly larger than those in non-coding regions. The flexible characteristics in the 5′-neighborhood of the coding regions and in three class sigma promoter sequences in the *E. coli* genome were also analyzed.

Key words: Base-pair step interaction; DNA flexibility; Statistical mechanical model; Whole-genome analysis

1. Introduction

Sequence-dependent DNA flexibility may play a significant role in a number of biological processes such as replication, transcription and recombination. It was reported that the formation of many protein–DNA complexes could involve DNA flexibility [1–3], and promoter activity may be affected by DNA flexibility [4]. Flexible DNA fragments were frequently located near regions functionally important for replication initiation, site-specific combination and transcription initiation [5,6]. A series of theoretical approaches have been used to investigate DNA flexibility: (1) the bendability model [7,8], a trinucleotide model, where higher values correspond to great levels of flexibility, (2) the position preference model [9], a trinucleotide model, where lower values correspond to great flexibility, (3) the protein-induced deformability model [10], a dinucleotide model, where higher values correspond to great flexibility and (4) the stacking energy model [11], a dinucleotide model, where more negative values correspond to great stability. Researches investigating DNA flexibility have used these models [12–14]. However these models are not perfect or systematic since they have been derived from mutually different and indirect investigations. Any single model emphasized only particular aspects of DNA flexibility. Historically, several quantities have been introduced to describe DNA flexibility such as persistence length, torsional rigidity and cyclization probability [15–17]. Although these quantities can partially describe DNA flexibility, the essence of flexibility that is determined by the dynamic feature of DNA structure itself has not been emphasized. As an alternative approach, it is suggested here that the flexibility is mainly related to the thermal fluctuation of DNA structure and can be calculated based on a statistical mechanical model for any given sequence.

A series of complete genome sequences have recently become available and some tentative works investigating DNA curvature and flexibility at the genome scale have been done recently [12,18]. This paper gives a new description of DNA flexibility. A statistical mechanical model was used to analyze DNA flexibility. A unified calculation for both the curvature and the flexibility of DNA is given using this model. The first problem (curvature) was discussed in detail in a previous dimer model [19]. The tetranucleotide model is now applied to analyze the flexibility of the *Escherichia coli* K12 genome. Similar analyses can be made for other genomes.

2. Materials and methods

2.1. Crystal structures and genome sequences

The structural parameters of the DNA double helix used in this paper included the angular parameters (Ω, ρ and τ) and the translation parameter (D_y). They were calculated based on data obtained from the nucleic acid database (NDB) [20]. For the sake of simplification, we restrict ourselves to the B-type double helix. The data did not contain any mismatched base pairs, unusual bases or non-Watson–Crick base pairs. The complete sequence of the *E. coli* K12 genome was published by Blattner et al. (GenBank accession number U00096) [21].

2.2. Parameter set for inner base pairs in tetramers

The 'Cambridge Accord' is used for the definitions and nomenclature of the DNA structural parameters [22] in this work. The origin of the base-pair coordinate set is the midpoint of the line connecting C_8 of a purine and C_6 of a pyrimidine. The y-axis (long axis) is defined by the RC_8–YC_6 line, where R = purine, Y = pyrimidine. Its positive direction points from strand II to strand I. The component of the base-pair normal (perpendicular to the y-axis) is used as the z-axis. The x-axis (short axis) completes a right-handed orthogonal axial set with the y-axis and z-axis. The positive x-axis direction points from the major groove to the minor groove. Figs. 1 and 2 illustrate the definitions of the base-pair coordinate set and the various parameters describing the base-pair step geometry in DNA respectively.

In general, the base-pair step geometry cannot be correlated with only two bases that define the step in question; the two flanking steps must also be taken into account [23]. The nearest neighbor interaction between base-pair steps was analyzed by studying tetramers instead of dimers as discussed previously [9,24–26]. Owing to the insufficiency of experimental data, R and Y at the left and right site of each tetramer were used to replace four possible bases. The complementary base sequences in two strands of DNA were considered to classify the independent tetramers into 36 types. The parameters were defined

*Corresponding author. Fax: (86)-10-62772237.
E-mail address: sunzhr@mail.tsinghua.edu.cn (Z. Sun).

Abbreviations: NDB, nucleic acid database; CDS, coding region

0014-5793/01/$20.00
PII: S0014-5793(01)02978-7

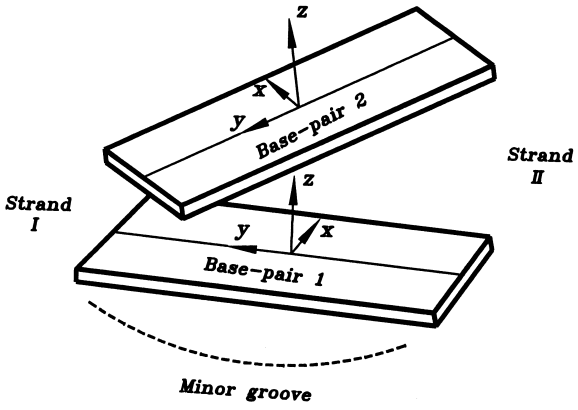

Fig. 1. Illustration of the base-pair coordinate set.

as follows: $\Delta\Omega=\langle\Omega\rangle-\Omega_0$, $\Delta\rho=\langle\rho\rangle-\rho_0$, $\Delta\tau=\langle\tau\rangle-\tau_0$ and $\Delta D_y=\langle D_y\rangle-D_{y0}$. The averaged helical parameters of the base-pair steps obtained in the NDB were $\Omega_0=36°$, $\rho_0=0.4°$, $\tau_0=0°$ and $D_{y0}=0.02$ nm while the averaged values of the helical parameters of the inner base pair for a given tetramer were $\langle\Omega\rangle$, $\langle\rho\rangle$, $\langle\tau\rangle$ and $\langle D_y\rangle$. The parameters $\Delta\Omega$, $\Delta\rho$, $\Delta\tau$ and ΔD_y for most tetramers were calculated directly from the statistical average of the data in the NDB. Only the corresponding values for RGGR and RACY, which are missing in NDB at present, were given with reference to the statistical data of those tetramers possessing the same inner base pair. The parameter set ($\Delta\Omega$, $\Delta\rho$, $\Delta\tau$ and ΔD_y) for the inner base pairs in the tetramers is listed in Table 1.

2.3. Empirical rule predicting local structural parameters of B-DNA

If the experimental twists of the base-pair steps in the jth sequence are $T_j(1)$, $T_j(2)$, ...$T_j(i)$, ..., the corresponding predicted relative twists of the jth sequence are denoted by $\mathrm{Tw}_j(1)$, $\mathrm{Tw}_j(2)$, ..., $\mathrm{Tw}_j(i)$.... Define

$$P=\sum_{j=1}^{N}\sum_{i=1}^{L_j-1}\mathrm{sgn}\left(\frac{\mathrm{Tw}_j(i+1)-\mathrm{Tw}_j(i)}{T_j(i+1)-T_j(i)}+1\right)\Bigg/\sum_{j=1}^{N}(L_j-1) \quad (1)$$

Here N is the number of sequences in the NDB and L_j is the length of the jth sequence. P describes the consistency between the experimental twist curve $T_j(i)$ versus i and the predicted one, $\mathrm{Tw}_j(i)$ versus i. $\mathrm{Sgn}(X)=X/|X|$. The sequences of DNA strands should be read in the 5′-to-3′ direction. The parameters of the inner base pair for each tetramer were taken from Table 1. The predicted relative twists for the ith step in the jth sequence are: $\mathrm{Tw}_j(i)=\Delta\Omega_j(i)-f\Delta\Omega_j(i+1)-f\Delta\Omega_j(i-1)$. The adjustable factor, f ($0\le f\le 1$), implying the modulation of two end base pairs to the structure of the dimer in the center, was determined by maximizing P throughout all the sequences in the NDB. The twist angle of the ith step in the jth sequence was predicted as $\Omega_j(i)=\Omega_0+\mathrm{Tw}_j(i)$. The predictions for rolling, tilting and sliding are similar to those for twisting. The adjustable factor, f, is equal to 0.125, 0.125, 0.075 and −0.125 for twisting, rolling, tilting and sliding predictions, respectively.

2.4. Conformational energy model

Following is a brief review of the statistical mechanical model used in this work [19]. Assume that the thermal fluctuation of twist, roll, tilt and slide for each base-pair step is independent. As done by Olson et al. [26], a simple elastic energy model is used to describe the twisting, rolling, tilting and sliding fluctuations of individual base-pair steps in the local base-pair geometry. The probability of finding the twist value for a base-pair step i in the interval Ω_i and $\Omega_i+\mathrm{d}\Omega_i$ is $P(\Omega_i)\mathrm{d}\Omega_i$. $P(\rho_i)$, $P(\tau_i)$ and $P(D_{yi})$ can be introduced in the same way.

$$P(\Omega_i)=\frac{\exp\left(-\frac{k_{\Omega_i}}{2RT}(\Omega_i-\Omega_{i0})^2\right)}{\int_{-\infty}^{\infty}\exp\left(-\frac{k_{\Omega_i}}{2RT}(\Omega_i-\Omega_{i0})^2\right)\mathrm{d}\Omega_i}=$$

$$\sqrt{\frac{k_{\Omega_i}}{2\pi RT}}\exp\left(-\frac{k_{\Omega_i}}{2RT}(\Omega_i-\Omega_{i0})^2\right) \quad (2)$$

$$P(\Omega_i,\rho_i,\tau_i,D_{yi})=P(\Omega_i)P(\rho_i)P(\tau_i)P(D_{yi}) \quad (3)$$

Here, R is the universal gas constant and T is the temperature. All calculations were done at room temperature. The force constants used in this work are listed in Table 2. (The units of k_Ω, k_ρ and k_τ are kJ/mol rad^2, and for $k_{D_{yi}}$ are kJ/mol nm^2, $R=0.00831$ kJ/mol K.)

The B-DNA structure can be investigated by observing the behavior of the virtual bonds from one base pair to another along the helix. Assume 0 nm and 0.34 nm for the shift (D_x) and the rise (D_z) in each base-pair step. Then, $\mathbf{V}_j=(0,\ D_{yj},\ 0.34\ \mathrm{nm})$ in the local frame. The corresponding vector in the overall frame is $\mathbf{a}_j=\Re_{12}\Re_{23}...\Re_{j-1j}\mathbf{V}_j$. $\Re_{j-1j}$, which is the transformation matrix relating the coordinate frames of the $(j-1)$th and jth base pair, and is the same as used earlier [19]. Let $\langle\mathbf{r}_j\rangle$ be the expectation value of the position vector $\mathbf{r}_j$ of the jth base-pair step. Suppose that the structural fluctuations of base-pair steps are independent. Then,

$$\langle\mathbf{r}_j\rangle=\langle\mathbf{a}_1\rangle+\langle\mathbf{a}_2\rangle+\langle\mathbf{a}_3\rangle+...+\langle\mathbf{a}_j\rangle=$$

$$\langle\mathbf{V}_1\rangle+\langle\Re_{12}\rangle\mathbf{V}_2+\langle\Re_{12}\rangle\langle\Re_{23}\rangle\mathbf{V}_3+...+\langle\Re_{12}\rangle\langle\Re_{23}\rangle...\langle\Re_{j-1j}\rangle\mathbf{V}_j \quad (4)$$

Using Eq. 3, the average transformation matrix $\langle\Re\rangle$ was obtained as a function of average sines and cosines of the twist, roll and tilt angles [19]. We define the flexibility, f_L, for a given DNA sequence of length L as

$$f_L=\sqrt{\langle\Delta r_L^2\rangle}=\sqrt{\langle r_L^2\rangle-\langle r_L\rangle^2} \quad (5)$$

The unit of f_L is nm. With this definition, one can easily calculate the flexibility for a given DNA sequence.

3. Results

3.1. Flexibility of the E. coli K12 genome

The real and shuffled *E. coli* K12 genomes were first divided

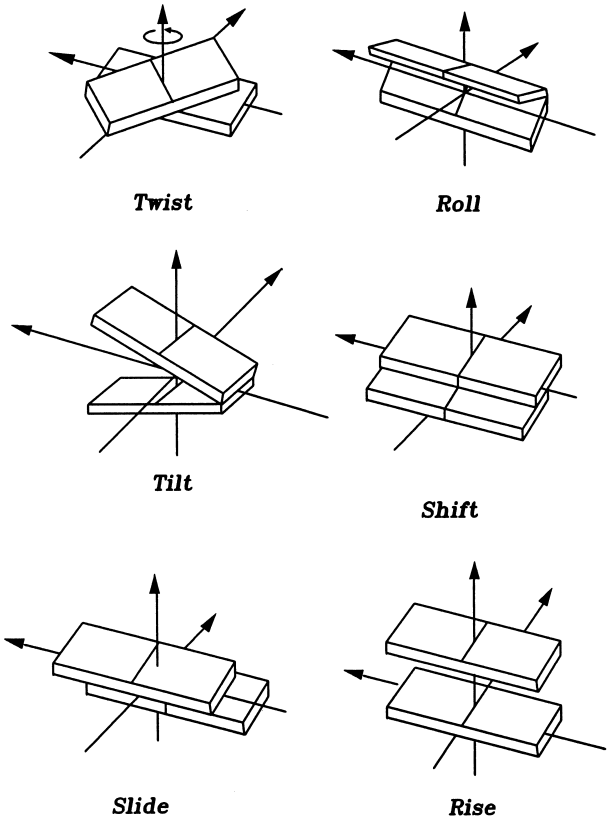

Fig. 2. Definition of the various parameters describing the DNA base-pair step geometry. Images illustrate positive values of the designated parameters.

Table 1
Local helix parameters of inner pairs in 36 tetramers of B-DNA

Tetramer	RAAR	RAAY	YAAR	YAAY	RAGR	RAGY	YAGR	YAGY	RGAR	RGAY	YGAR	YGAY
$\Delta\Omega$ (°)	0.0	1.0	−2.0	−5.0	−2.0	−2.0	−10.0	2.0	3.0	3.0	3.0	2.0
$\Delta\rho$ (°)	0.2	−0.7	3.9	4.1	−2.0	−1.5	2.0	4.6	2.3	1.1	1.4	1.7
$\Delta\tau$ (°)	0.0	0.0	−1.1	−0.4	1.5	1.4	−1.2	3.0	2.2	0.0	−0.5	−1.0
ΔD_y (nm)	−0.02	−0.02	0.02	0.0	0.0	0.02	−0.03	0.08	0.02	−0.04	−0.04	−0.03

Tetramer	RGGR	RGGY	YGGR	YGGY	RCAR	RCAY	YCAR	YCAY	RCGR	RCGY	YCGR	RTAR
$\Delta\Omega$ (°)	10.0	4.0	0.0	−1.0	−3.0	−7.0	14.0	−1.0	−8.0	0.0	−3.0	1.0
$\Delta\rho$ (°)	0.9	1.8	−1.2	2.5	3.8	6.5	−8.2	1.8	5.1	3.5	4.8	5.4
$\Delta\tau$ (°)	−2.0	−4.0	0.8	−1.0	0.2	−1.4	0.1	0.4	−0.2	1.9	1.4	−2.3
ΔD_y (nm)	−0.02	−0.03	−0.08	0.01	0.08	0.06	0.26	0.04	0.05	0.04	0.08	0.16

Tetramer	RTAY	YTAR	RATR	RATY	YATR	RACR	RACY	YACR	YACY	RGCR	RGCY	YGCR
$\Delta\Omega$ (°)	6.0	5.0	−4.0	−3.0	−6.0	−5.0	−2.0	−1.0	−2.0	1.0	1.0	4.0
$\Delta\rho$ (°)	−0.7	0.9	−1.3	−2.2	0.3	−1.6	−1.6	−3.3	−0.8	−3.7	2.1	−8.6
$\Delta\tau$ (°)	−0.5	1.4	1.8	0.3	−0.3	0.2	−1.0	−1.0	0.5	−0.5	1.8	−0.9
ΔD_y (nm)	0.0	0.03	−0.08	−0.08	−0.08	−0.05	−0.05	−0.01	−0.05	0.0	−0.08	0.02

into a series of non-overlapping 1000 bp windows and then the present model was used to calculate the flexibility at each position. Next, a diagram showing the flexibility in any region of the genome was constructed. Most of the flexibility values for a shuffled *E. coli* K12 sequence lie in the range from 110 to 118 nm, but those in the real *E. coli* K12 genome are more spread out in the range from 108 to 120 nm (figure not shown). Therefore, the shuffled *E. coli* K12 sequence displays much less extreme flexibility than the real *E. coli* K12 genome.

The profiles of the smoothed flexibility with 100 kb sliding windows for the *E. coli* K12 genome in Fig. 3a show two strong valleys at 1.26 Mbp and 1.61 Mbp respectively. The valley and the peak of the flexibility profile for the *E. coli* K12 genome coincide exactly with the positions of terminus (1.28 Mbp for TerD and 1.61 Mbp for TerC) and the origin of replication (about 4 Mbp), respectively. Also, the locations of the extremely flexible regions in this work were consistent with those in previous works, while the valleys of flexibility in this work correspond to the peaks in curvature profiles from Pederson et al. [12]. To understand the mechanism behind flexibility extremes in genomes, the G+C content was calculated at each position using 100 kb non-overlapping windows. Then, a profile was determined for the G+C content (Fig. 3b). A strong correlation exists between the G+C content and the flexibility (correlation coefficient $R = 0.922$). G+C-rich regions are usually more flexible.

The sliding window size was selected arbitrarily, however results calculated using other sliding windows were basically the same as those calculated using 1000 bp sliding windows.

3.2. Flexibility in coding regions, non-coding regions and shuffled non-coding regions

To investigate the difference between coding regions (CDS) and non-coding regions in a genome, CDS and non-coding regions were extracted from *E. coli* K12 genomes based on annotations supplied in the GenBank and combined to form two subsequences. Next, the flexibility diagram was constructed for these subsequences using 1000 bp non-overlapping sliding windows. The difference between the average flexibility distributions for non-coding, shuffled non-coding and CDS in the *E. coli* K12 genome differed greatly (Fig. 4). The average flexibility in CDS was larger than that in non-coding subsequences. Similar results are shown in Table 3 for other microbial genomes. The largest values are shown in boldface type.

3.3. Flexibility in the 5′-neighborhood of the CDS

The flexibility in the 5′-neighborhood of the CDS in *E. coli* K12 genome was investigated next. Sequences in the 5′-neighborhood of the CDS in *E. coli* K12 genome were first extracted from the GenBank file based on its annotation and were then aligned with the translation initiation point. The flexibilities of these sequences were calculated for all positions in the range from −500 bp to +500 bp using 147 bp sliding windows (representing 14 multiples of the helical repeat). The averaged flexibilities at each position in the alignment were then calculated. As seen in Fig. 5 valley exists between −126 bp and −51 bp, with a peak between 24 bp and 124 bp with a maximum at 74 bp.

Table 2
Force constants of 10 dimers in the model

Step	AA	AC	AG	AT	CA	CG	GA	GC	GG	TA
K_Ω	109.5	111.6	192.3	94.5	141.7	139.2	99.5	116.2	244.5	166.8
$K_{\rho 1}$	120	405	360	330	165	135	210	360	555	90
$K_{\rho 2}$	60	75	120	60	195	285	90	90	75	210
K_τ	180	180	348	132	372	339	228	345	180	240
$K_{D_{yi}}$	760	900	1950	1110	1850	1760	1380	580	1930	790

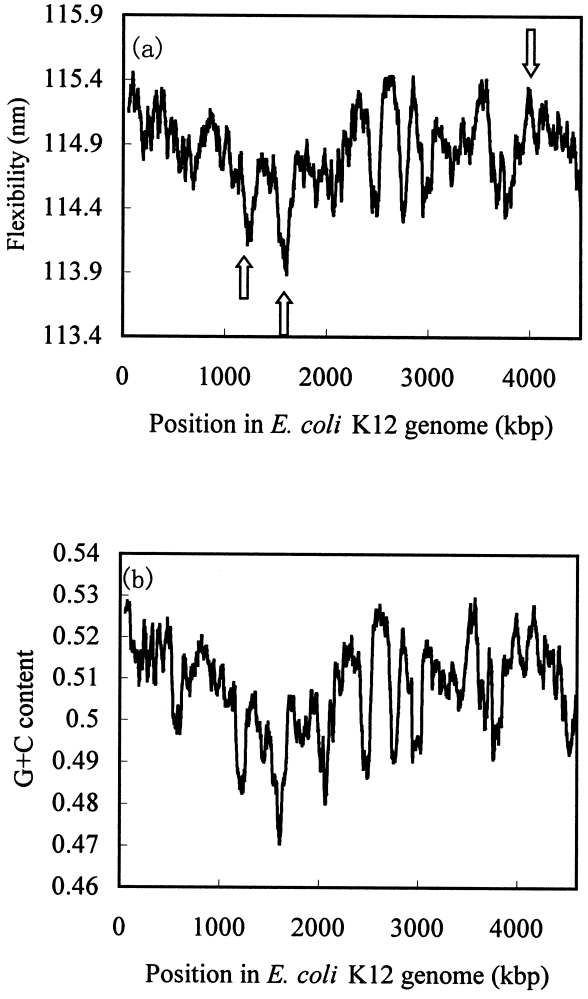

Fig. 3. a: Flexibility profiles smoothed using 100 kb sliding windows for *E. coli* K12 genome. Arrow (↓) identifying predicted origin of replication. Other two arrows (↑) indicate predicted terminus sites for TerD and TerC respectively. b: G+C content profiles using 100 kb sliding windows for *E. coli* K12 genome.

3.4. Promoter flexibility profiles

Three sigma class promoter sequences (sigma-70, sigma-54 and sigma-32) were also extracted from the *E. coli* K12 genome using the annotation in the GenBank files. The promoters were aligned with the transcription initiation point. The averaged flexibility of these sequences was calculated for all positions from −500 bp to +500 bp using 147 bp windows, representing 14 multiples of the helical repeat and the size of putative DNA loops involved in transcription initiation. The sliding windows were placed at every 25 bp. The computational results in the present work showed that an obvious non-average structure exists near the transcription starting point (Fig. 6). Note that the sigma-54 case is different from the two sigma classes mentioned above.

4. Discussion

The valley and peak in the flexibility profile for the *E. coli* K12 genome coincide exactly with the positions of the terminus and the origin of replication, respectively. Pederson et al. [12] also found a trend for higher degrees of curvature, with lower levels of flexibility near the terminus. Initiation of replication at *OriC* in vitro starts with the formation of a complex that requires a series of proteins. The subtle structural adjustment between the DNA and binding protein probably leads to stronger hydrogen-bonding, electrostatic contacts and van der Waals interactions thereby facilitating special DNA–protein recognition. We speculate that the great levels of flexibility in the origin of replication may promote the formation of a DNA–protein complex to start replication. The lower flexibility in the terminus may help terminate the replication. Extremely flexible or rigid regions in the real *E. coli* K12 genome may be related to necessary biological functions and to the evolutionary pressure at the DNA structural level. The specific patterns of DNA flexibility may provide new information at the structural level for genome studies.

The results in previous works [12,14] indicated that the predicted flexibilities in the protein-induced deformability model and in the stacking energy model correlate to G+C content, but there are no correlations to G+C content in two trinucleotide models. The G+C content at the terminus region was low while the G+C content at the origin of replication was high in the present model. The strong correlation between flexibility and G+C content at the level of 1000 bp non-overlapping windows and on the whole-genome scale (Table 3) may be related to the fact that extremely flexible dinucleotides CA, AG, GC and CG in this model were GC-rich. We do not know whether the essence of flexibility is due to GC-rich, but GC-rich regions indeed have lower flexibility.

The average flexibility in CDS is larger than that in non-coding regions for all the genomes. This result is consistent with the conclusion in Pederson's work [12]. The low flexibility of a DNA sequence in intergenic regions may be a universal feature and may be needed to regulation of expression in prokaryote. This conclusion supports the hypothesis that the DNA structure was one of the driving forces behind the evolution of the genome sequence. The fact that CDS are generally more flexible than intergenic regions may be useful for finding new genes in genomes.

The upstream and downstream regions from the translation initiation point play an important role in the translation process. The obvious rigid region from −126 bp to −51 bp and the flexible region from 24 bp to 124 bp relative to the trans-

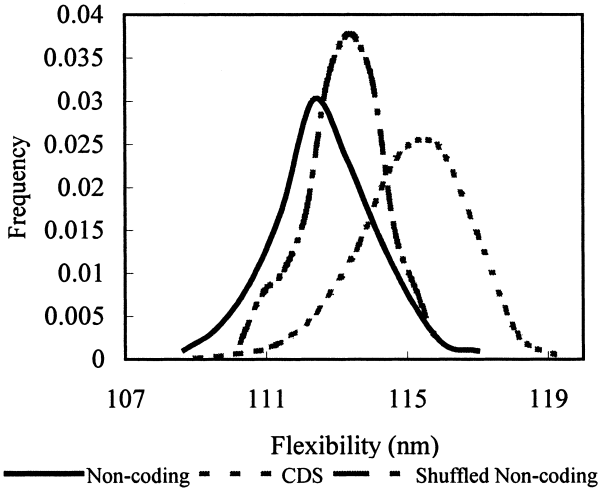

Fig. 4. Flexibility distribution in non-coding, shuffled non-coding regions and CDS for *E. coli* (using 1 kb non-overlapping windows).

Table 3
Average flexibility in coding and non-coding regions using 1 kb non-overlapping sliding windows for 31 microbial genomes

Genome	GenBank acc. no.	GC (%)	f_L in CDS	f_L in non-coding
H. sp.	AE004437	68	**118**	116.9
A. pernix	Aero_p	56	**115.1**	114.5
A fulgidus	NC000917	49	**113.7**	112.8
M. thermo	AE000666	50	**113.2**	111.3
P. abyssi	AL096836	45	**112.8**	111.6
P. horikoshii	Pyro_h	42	**112.2**	111.5
M. janaschii	L77117	31	**110.1**	106.6
D. radiodurans	AE000513	67	**118.1**	114.5
C. crescentus	AE005673	67	**117.6**	117.1
M. tuberculosis	AL123456	66	**117.6**	117.1
T. pallidum	AE000520	53	**115.4**	115.2
X. fastidiosa	AE003849	53	**115.5**	114.1
E. coli	U00096	51	**115.1**	112.7
N. meningitidis	AE002098	52	**114.6**	112.6
T. maritima	AE000512	46	**113.6**	112.3
B. subtilis	AL009126	44	**113.3**	111.6
B. halodurans	BA000004	44	**113.1**	112.2
C. trachomatis	AE001273	41	**113**	111.6
C. pneumoniae	AE002161	40	**112.8**	111.1
S. pcc6803	AB001339	48	**112.7**	111.7
C. muridrum	AE002160	40	**112.5**	111.2
P multocida	AE004439	40	**112.5**	110.9
M. pneumoniae	NC000912	40	**112.2**	110.6
H. pylori	AE000511	39	**111.6**	110.1
A. aeolicus	AE000657	44	**111.3**	110.5
R. prowazekii	AJ235269	29	**110.4**	109.0
M. genitalium	L43967	32	**110.3**	109.9
C. jejuni	AL111168	30	**110.2**	108.3
B. burgdorferi	AE000783	29	**109.4**	108.3
B. sp.	AP000398	26	**109.1**	106.8
U. urealyticum	AF222894	25	**108.9**	107.8

lation start point (Fig. 5) may be useful for binding of ribosome in the translation process. The computational results in the present model showed that an obvious non-average structure exists near the transcription starting point. Sigma-54 promoters, which are employed when ammonia is absent from the medium, were found to be very different from sigma-70 and sigma-32 promoters. These findings are consistent with the fact that sigma-54 is different from the sigma-70 family both structurally and functionally [27,28]. Pederson et al. [12] found that sigma-70 and sigma-32 are more curved and less flexible than the genomic average. The sigma-54 profile is very different from those of sigma-70 and sigma-32. Many experimental works have indicated that the formation of many protein–DNA complexes involved in the transcriptional control of gene expression, such as GCN-ATF/CREB complex [1], the binding of RNA polymerase holoenzyme or its α subunit to UP element in *E. coli* [4], TBP/TATA complex [29], RNA polymerase-σ^{54}-DNA complex [30] and the binding of FIS and H-NS to DNA [31] could involve DNA flexibility. The precise action mechanism of these rigid or flexible regions remains unknown. A detailed biological explanation cannot

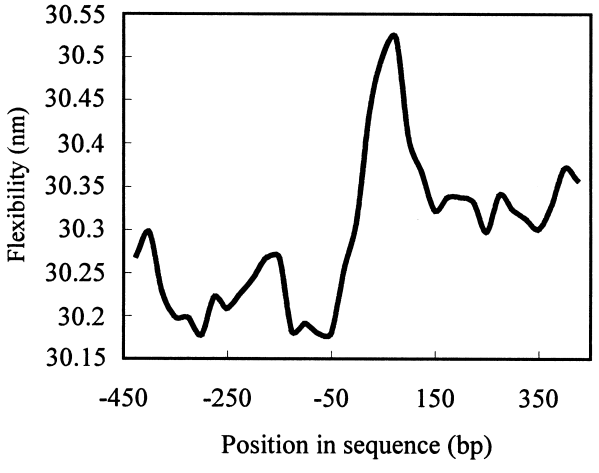

Fig. 5. Flexibility distribution in the 5′-neighborhood of the CDS. The window size is 147 bp.

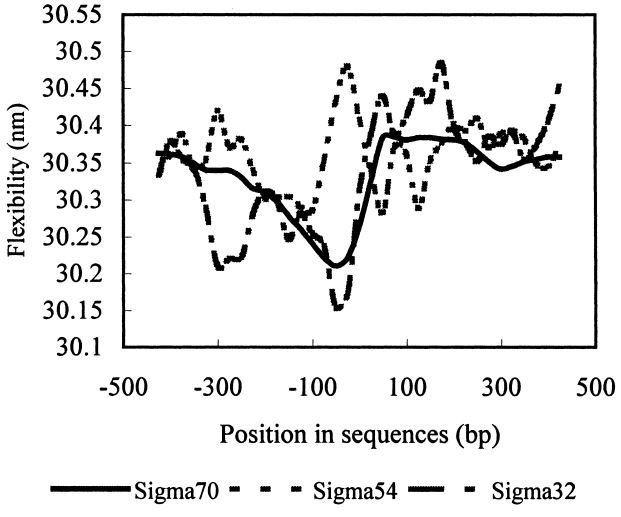

Fig. 6. Averaged flexibility profiles for three sigma classes in the region from 500 bp upstream to 500 bp downstream from the transcription initiation point. The sliding window size is 147 bp.

be given due to the lack of experimental information directly related to our results. Directed biological experiments are expected.

Acknowledgements: L.T. is thankful to L.F. Luo for his helpful discussion. This work was supported by the National Science Foundation of China (199470006), China Postdoctoral Science Foundation and the National Science Foundation of the Inner Mongolia of China (20001302).

References

[1] König, P. and Richmond, T.J. (1993) J. Mol. Biol. 233, 139–154.
[2] Grove, A., Galeone, A., Mayol, L. and Geiduschek, E.P. (1996) J. Mol. Biol. 260, 120–125.
[3] Nagaich, A.K., Zhurkin, V.B., Sakamoto, H., Gorin, A.A., Clore, G.M., Gronenborn, A.M., Appella, E. and Harrington, R.E. (1997) J. Biol. Chem. 272, 14830–14841.
[4] Nègre, D., Bonod-Bidaud, C., Oudot, C., Prost, J-F., Kolb, A., Ishihama, A., Cozzone, A.J. and Cortay, J-C. (1997) Nucleic Acids Res. 25, 713–718.
[5] Grosschedl, R. (1995) Curr. Opin. Cell Biol. 7, 362–370.
[6] SjØttem, E., Anderson, C. and Johansen, T. (1997) J. Mol. Biol. 267, 490–504.
[7] Brukner, I., Jurukovski, V. and Savic, A. (1990) Nucleic Acids Res. 18, 891–894.
[8] Brukner, I., Sanchez, R., Suck, D. and Pongor, S. (1995) EMBO J. 14, 1812–1818.
[9] Satchwell, S.C., Drew, H.R. and Travers, A.A. (1986) J. Mol. Biol. 191, 659–675.
[10] Olson, W.K., Gorin, A.A., Lu, X-J.L., Hock, M. and Zhurkin, V.B. (1998) Proc. Natl. Acad. Sci. USA 95, 11163–11168.
[11] Ornstein, R., Rein, R., Breen, D. and MacElroy, R. (1978) Biopolymers 17, 2341–2360.
[12] Pedersen, A.G., Jensen, L.J., Brunak, S., Stærfeldt, H-H. and Ussery, D.W. (2000) J. Mol. Biol. 299, 907–930.
[13] Baldi, P., Brunak, S., Chauvin, Y. and Pederson, A.G. (1999) Bioinformatics 15, 918–929.
[14] Baldi, P. and Baisnée, P-F. (2000) Bioinformatics 16, 865–889.
[15] Hagerman, P.J. and Ramadevi, V.A. (1990) J. Mol. Biol. 212, 351–362.
[16] Cognet, J.A.H., Pakleza, C., Cherny, D., Delain, E. and Cam, E.L. (1999) J. Mol. Biol. 285, 997–1009.
[17] Bhattacharyya, D., Kundu, S., Thakur, A.R. and Majumdar, R. (1999) J. Biomol. Struct. Dynam. 17, 289–300.
[18] Bolshoy, A. and Nevo, E. (2000) Genome Res. 10, 1185–1193.
[19] Tsai, L. and Luo, L.F. (2000) J. Theor. Biol. 207, 177–194.
[20] Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A.R. and Schneider, B. (1992) Biophys. J. 63, 751–759.
[21] Blattner, F.R., Plunkett III, G. and Bloch, C.A. et al. (1997) Science 277, 1453–1474.
[22] Diekmann, S. (1989) J. Mol. Biol. 205, 787–791.
[23] Yanagi, K., Privé, G.G. and Dickerson, R.E. (1991) J. Mol. Biol. 217, 201–214.
[24] Calladine, C.R., Drew, H.R. and McCall, M.J. (1988) J. Mol. Biol. 201, 127–137.
[25] Bolshoy, A., McNamara, P., Harrington, R.E. and Trifonov, E.N. (1991) Proc. Natl. Acad. Sci. USA 88, 2312–2316.
[26] Olson, W.K., Marky, N.L., Jernigan, R.L. and Zhurkin, V.B. (1993) J. Mol. Biol. 232, 530–554.
[27] Gross, C.A., Lonetto, M. and Losick, R. (1992) In: Transcriptional Regulation (Mcknight, S. and Yamamoto, Y., Eds.), pp. 129–176, Cold spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[28] Merrick, M.J. (1993) Mol. Microbiol. 10, 903–909.
[29] Davis, N., Majee, S.S. and Kahn, J.D. (1999) J. Mol. Biol. 291, 249–265.
[30] Jeltsch, A. (1998) Biophys. Chem. 74, 53–57.
[31] Afflerbach, H., Schröder, O. and Wagner, R. (1999) J. Mol. Biol. 286, 339–353.