FEBS Letters 489 (2001) 166–170 FEBS 24549

Analysis of sequence-dependent curvature in matrix attachment regions

Jun Yamamura, Koji Nomura*

Institute of Agriculture and Forestry, University of Tsukuba, Tsukuba, Ibaraki 305-8572, Japan

Received 6 December 2000; accepted 26 December 2000

First published online 15 January 2001

Edited by Lev Kisselev

Abstract Sequence-dependent DNA conformations of matrix attachment regions (MARs) available in a database were calculated using the wedge model, and compared with randomly chosen genes, promoters, enhancers and transposons. The MARs had a longer bent part and higher angle/helical turn than the other regions. It is known that some MAR sequences have A-tracts that cause DNA bending, and we also found many A-tracts in examined MARs. Furthermore, non-random and clustered distribution of A-tracts shown here gave further evidence of the importance of A-tracts for MAR conformations. These results suggest that DNAs of MARs have a characteristic conformation instead of conserved sequence.

Key words: Matrix attachment region; Wedge model; Bent DNA; DNA curvature

1. Introduction

Chromosomes consist of many regions having specific functions [1]. Genes contain genetic information. The expression of genes is controlled by promoters, enhancers, and related regions on the same genome. Centromere and telomere are regions that are important for the formation and separation of chromosomes. Repetitive DNA and many kinds of transposons may be important in evolution [2]. A small part of chromosomal DNA maintains DNA conformation such as loops in the interphase nucleus [3–5]. Those chromosome regions bind to nuclear matrix, the residual nuclear structure after the extraction of histone, nuclear membrane and most of the chromatin [6]. The region had been called the matrix-associated region [7] or scaffold-associated region [8,9], but is now called the matrix attachment region (MAR).

It is reported that many MARs play important roles in enhancing gene expression [10–15] and DNA replication [16,17]. Although the importance of MARs in the nucleus has been given considerable attention, the mode of interaction between nuclear matrix and MARs is still obscure. DNA sequences of the MARs have usually high AT content [18] and often harbor specific motifs, especially A-box and T-box [9] and the consensus sequence for topoisomerase II [7]. Putative origins of replication (ARS elements) [19] have also been observed in many MARs. There are several reports on the conservation of binding manner between MAR and binding proteins in the nuclear matrix. A MAR sequence derived from the immunoglobulin gene of the kappa light chain of mouse has the potential to bind with nuclear matrices prepared from *Saccharomyces cerevisiae* [20], while the MAR from *Drosophila* fushitarazu gene can attach to tobacco nuclear matrix [21]. This interspecific compatibility between MARs and their binding proteins seems to be supported by nucleotide sequences of MARs and peptide sequences of binding proteins. Thus, nucleotide sequences of MARs may be conserved not only in the motives described above but also in whole lengths to some extent. However, nucleotide sequences of reported MARs are low in homology [22]. For example, eight sequenced MARs on human chromosome 19 had no mutual homologies [23]. Thus, it seems that the nucleotide sequence provides only limited information on the nature of MARs.

Double strand DNA can form three-dimensional (3-D) structures in nuclear space. The formation of specific structures induced by proteins has been observed in nucleosomes, which are formed when DNA wraps the histone octamer [1]. Double strand DNA can also form a bent conformation without interactions with other molecules [24]. Intrinsic bending of DNA modulates transcription in cyanobacteria [25]. Bending in the promoter region of a Gram-negative bacterium is essential for its activity [26,27]. A bending site is also present in the replication origin of bacteriophage lambda and is required for protein binding [28]. Furthermore, the cleavage site by topoisomerase II might have bending in *Drosophila* [29]. MARs of chicken, rat and *Drosophila* are believed to be bent [30–33]. In some cases, bent DNA may contribute to the binding property of MARs with nuclear matrices [31,32]. Furthermore, nuclear matrix proteins that bind to bent DNA have been identified [33,34]. Therefore, it seems that the bending of DNA is important for MARs.

Bending of double strand DNA is mainly caused by homopolymeric dA of at least 4 bp, called A-tract [35,36]. These A-tracts apparently support the bending of MARs, because many known MARs have A-tracts [21,30,31,37–40].

Among the models predicting DNA conformation, the wedge model proposed by Ulanovsky and Trifonov [41] is superior in its ability to predict DNA conformation [42]. In this model, DNA conformation is given by wedge angles that are between stacked base pairs. Bolshoy et al. [43] proposed wedge angles between all 16 combinations of nearest neighbor base pairs by statistical analysis of DNA migration. Using a program based on this model [44], researchers have predicted DNA bending in many chromosomal regions [45–50]. Thus the model is a powerful tool to calculate the 3-D structure of a double strand DNA. Unfortunately, there have been no programs based on the wedge model which run on Microsoft Windows-based PCs. Therefore, we composed a program based on the wedge model that works on PCs which features,

*Corresponding author. Fax: (81)-298-53 6692.
E-mail: daucus@sakura.cc.tsukuba.ac.jp

PII: S0014-5793(01)02098-1

among other things, a search function for discretional nucleotide sequence.

In this study, we predicted the 3-D structures of all available MARs in public database by using the wedge model. The aim of the study is to know whether MARs have characteristic and/or common DNA structures. Our comparison of the structures of MARs with randomly chosen sequences of genes, promoter regions, enhancer regions, transposons showed that many MARs tended to harbor bent DNA rather than genes, promoters and enhancers. Non-random distribution of A-tracts in MARs was also shown.

2. Materials and methods

2.1. Program

The program to calculate 3-D structures of given DNA sequences was written in Microsoft Excel VBA. We used an algorithm proposed by Manning and Maddocks [51], and the table of specific angles between base pairs presented by Bolshoy et al. [43].

To indicate curvature, we used the angle between two vectors 31 bp (~3 helical turns) apart, as per Goodsell and Dickerson [52]. We used the angle between base pair$^{(i)}$ and base pair$^{(i+31)}$ as the indicator angle for base pair$^{(i)}$. For simplification, the angle per 31 bp was converted to the angle per helical turn (=10.5 bp).

2.2. DNA sequences

We use the MAR sequences in the S/MARt database (http://transfac.gbf.de/SMARtDB). There are 123 available MARs out of 193 collections (accession numbers SM0000001–3, SM0000005, SM0000006, SM0000008–48, SM0000050, SM0000052–78, SM0000082–99, SM0000107, SM0000114, SM0000116, SM0000119, SM0000137, SM0000142, SM0000150, SM0000158–160, SM0000164, SM0000165, SM0000178), consisting of 56 human, 19 mouse, four rat, five hamster, four chicken, four *Drosophila*, seven *Arabidopsis*, five sorghum, six rice, two maize, one pea, one soybean, one potato, two tobacco, two petunia, one French bean and two yeast MARs. As controls, we obtained 30 gene sequences from the EMBL database. Ten sequences were randomly chosen from *Arabidopsis*, 10 from human and 10 from mouse. We also obtained 30 promoter regions, 30 enhancer regions and 30 transposons from unspecified eukaryotes (Table 1). Many of the genes contained intron and many of the transposons contained genes.

Table 1
Accession numbers of randomly chosen sequences

Gene	Promoter	Transposon	Enhancer
AF115474	AJ249888	AJ009237	AF044976
AF233753	AF250147	U75360	AJ271782
AF265547	U19862	AJ289698	AF225952
AF268093	AF254075	AF136221	X54929
AF233752	Y16044	AJ242983	AF013722
AB006078	AF059314	X56231	AF047373
AB023448	A31609	Z69893	AF085245
AB028469	AF233737	AF260904	AF099475
AF017056	AF262063	U37228	AF152113
X80127	AF252566	AF033098	AF218259
AF133627	AJ276488	L38255	AF222994
D89501	AF086710	L25911	U21227
U03493	AB041231	L38254	Y13401
U10360	AF207862	L12220	L07488
AJ277365	X70232	U28041	X55006
U03735	AJ242953	M16478	X81476
U08997	AF193838	U51228	M57451
U50133	AF246992	D42062	X54926
U17894	AJ277875	L41171	Y09580
U32672	AJ251807	L48685	Z46773
L29190	AF227990	U38613	AF042709
D31952	AJ131727	U38614	AF060506
L20420	AJ243070	M80343	AF132808
J03482	AB042835	AF108961	X16802
K00971	AF060887	U55049	X81977
L27842	AB033127	X73309	M26696
M23453	AF233613	X59156	AF173831
M58564	AJ243067	M17424	L19326
D38077	AF272889	AF010445	Z98207
M63649	A19296	Z30334	M31033

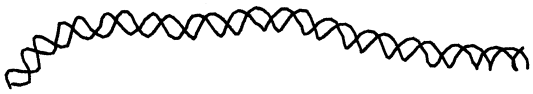

Fig. 1. Conformation of the *E. coli* galactose operon by our program. Two-dimensional display of 3-D coordinates.

3. Result and discussion

3.1. Graphical presentation of MARs

To confirm the accuracy of our program, we calculated the 3-D conformation of the *gal* control region of the *Escherichia coli* galactose operon. Our conformation of the sequence (Fig. 1) coincides with that previously reported using the same parameters [44]. This suggests that our program provided accurate 3-D structures of given sequences.

We calculated the 3-D coordinates of MAR conformations using our program. Fig. 2 shows the most effective way to represent DNA conformations. As we can see, there are various conformations of MARs. For example, in four MARs of *Drosophila*, SM0000110 showed the largest curvature, SM0000037 and SM0000118 had some strong bent regions, while SM0000205 was almost straight. These results suggest that bending and conformation are different in each MAR.

3.2. Curved region

Succession of bending is also important for DNA conformation because the degree of curvature varies by the length of

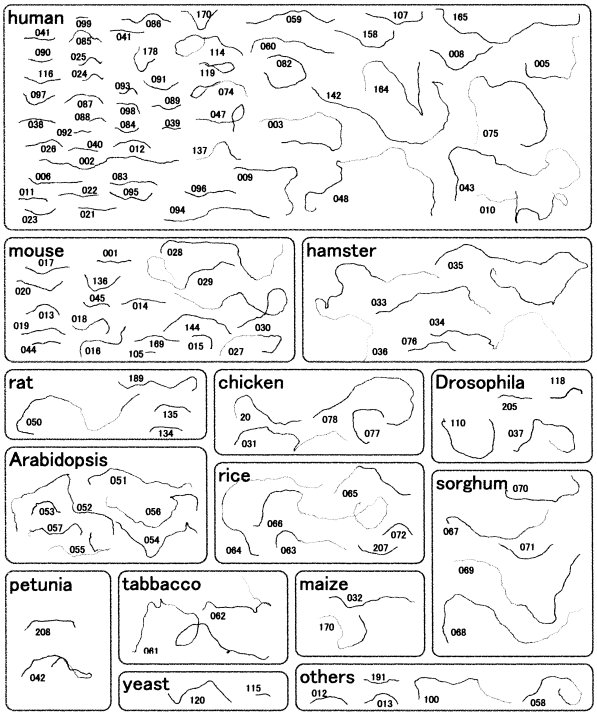

Fig. 2. Conformation of 123 MARs. The number next to each graphic is a three-digit accession number.

(a)

AAAGACTTTT AKAGACATTA ATCGACATCT TATATCAAAT
GTTATTTTTT TGCAACATGT AGTATCAATC GTTGTAGTTT
TCACAACATT TCTTCATAAA TTTGCTGTTT TTACGATATT
TCTTCATAAA TGTTATTGTT TTTACGATAT TTAGTATCAA
ATGTCATTTT TTATAACATT TAATATCAAA TGTCATTATT
TTTATTACAT TTTGTATCAA ATGTCGTAAT TTTGACGATA
TTTCTTCATT AATGTCGTTA TTATTACGAC ATTTTGTATC
AAATGTCATA TTTTGATGAC ACTTAGTATC AAATGTCGTA
ATGTTTGCGA CACTTTGTAT CAAATGTCGT AGTATTTACG
ACATTTATTC ATAAATGACA TGAATTTCAT ATTTTTCTG

(b)

TGAAGATTTT GAACATTGTT TTTGTTATTG TTGGTTAATC
CCACCATTTT TGTAACTTTT TAATCAAATT AGTATATTTT
TGTGTTGTAA GGTGACACTA ATGAATTGAT TTGTTATCCT

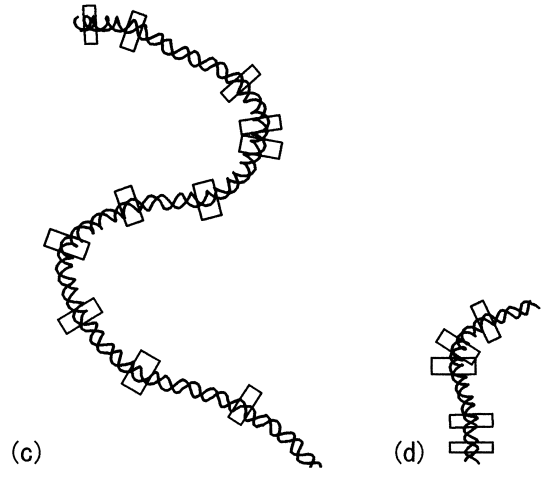

Fig. 3. DNA bending in a region where many A-tracts appeared. (a) Partial sequence of SM0000061 tobacco MAR. The fragment contains a roughly three helical turn interval of T-tract. (b) Partial sequence of SM0000051 *Arabidopsis* MAR. The fragment contains a roughly one helical turn interval of T-tract. (c, d) Conformation of partial sequences of SM0000061 and SM0000051 predicted by the wedge model.

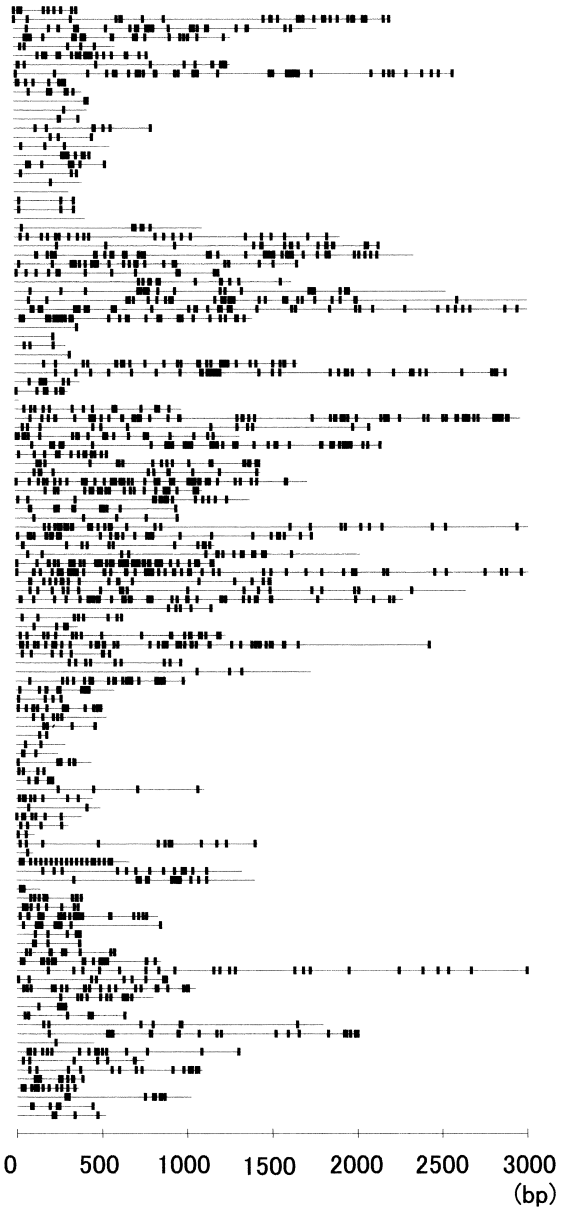

Fig. 4. The distribution of A-tracts in MARs. The black squares represent the start point of either A-tracts or T-tracts in each fragment. The regions longer than 3000 bp have been omitted. MARs were lined up according to the order described in Section 2.

it. We calculated the indicator angles at all base pairs except the first and last 15 bp in each MAR fragment. To know whether a fragment harbored a curved region or not, we checked the presence of strong bending succession. Based on Gabrielian [53], we considered 15° per helical turn to be strong bending. We regarded over 10 bp as successive and divided them into two groups: 11–20 bp and over 20 bp.

We counted the number of fragments harboring long successive bent regions in all available MARs, as well as 30 promoters, 30 enhancers, 30 genes, and 30 transposons that were randomly chosen (Table 2). Twenty-eight of the 123 MARs harbored 21 bp or longer successive bent ($>15°$) regions, 37 harbored 10–20 bp long regions and 58 had no such region. The ratios of each fragment to all 123

Table 2
Number of fragments harboring curved region

	Max length of successive bent parts			
	–10 bp	11–20 bp	21 bp–	Total
MAR	58 (47.2%)	37 (30.1%)	28 (22.8%)	123 (100%)
Genes	17 (57.7%)	4 (13.3%)	9 (30.0%)	30 (100%)
Promoter regions	17 (56.7%)	7 (23.3%)	6 (20.0%)	30 (100%)
Enhancer regions	22 (73.3%)	5 (16.7%)	3 (10.0%)	30 (100%)
Transposons	6 (20.0%)	15 (50.0%)	9 (30.0%)	30 (100%)

Table 3
Distribution of indicator angle/helical turn (%)

	< 9°	9–15°	15° <	Length of samples (average ± S.D.)
MARs	70.4	25.1	4.5	1083 ± 1019
Genes	75.8	21.1	3.1	3012 ± 1893
Promoter regions	75.3	21.9	2.8	1226 ± 1032
Enhancer regions	76.7	20.7	2.6	578 ± 457
Transposons	72.3	23.7	4.0	1851 ± 1740

MARs were 22.8%, 30.1% and 47.2%, respectively. The ratios of fragments that had no curved region in MARs (47.2%) were smaller than those of promoters (56.7%), enhancers (73.3%) and genes (57.7%). Nine of 30 (30.0%) transposons had 21 bp or longer bent regions and 15 (50.0%) had 11–20 bp long bent regions. The transposons showed the highest score in both the 11–20 bp and 21 bp or longer groups. The interaction between transposase and transposon DNA needs DNA bending [54–57], but it is not clear whether the protein causes bending or the bend is intrinsically present in the DNA [56].

It should be noted that the lengths of DNA fragments used in this study varied. The average length of the 30 enhancers was 578 ± 457, whereas that of the genes was 3012 ± 1893. The length of fragments may have affected the percentage of fragments harboring curved regions. Longer fragments were more likely to harbor such regions.

3.3. Distribution of angles/helical turn

To eliminate the influence of fragment length, we calculated the distribution of indicator angles in each region (Table 3) by counting the numbers of base pairs whose indicator angle/helical turns were 9° or less, between 9° and 15°, and over 15° [53]. We also calculated their respective ratios to total fragment length. The total length was the sum of each fragment length minus the number of base pairs that were not used to calculate the angle. In total MARs, the percentage of base pairs in 9–15° was 25.1% and that of over 15° was 4.5%. These values were higher than the values for genes, promoters and enhancers, which were similar. In these three regions, the percentage of base pairs in 9–15° was about 20% and about 30% in the over 15° group.

The distribution of angles/helical turn in enhancers was similar to those for genes and promoters although the number of fragments having curved regions was smallest in enhancer regions (Tables 2 and 3). It seems that the short average length of enhancers (578 bp) might have decreased the opportunity to have curved regions. The percentage of base pairs under 9° in MARs and transposons was smaller than in the other regions. The ratio of fragments showing curved regions in Table 2 suggests that MARs and transposons tend to bend more than the other three regions. Thus, it seems likely that bent DNA in MARs is involved in mechanisms like transposon and transposase.

3.4. A-tract in MARs

It is known that MARs tend to harbor A-tracts, which often exist as clusters [21,40]. Since poly dA is not bent per se [36], DNA bending can be caused by A-tracts when they are close to each other and when the interval is a multiple of a DNA helical turn (= 10.5 bp) [36,37]. For instance, the region of a tobacco MAR (SM0000061) and *Arabidopsis* MAR (SM0000051) showed strong bending (Fig. 3). The sequence of the region contained many A-tracts, which appeared at intervals of about three helical turns (≒31 bp) and one helical turn (≒10.5 bp), respectively. Our search of A-tracts in all MARs (Fig. 4) showed they had non-random distribution and clustered closely together in many MARs.

Izaurralde et al. reported that poly dA binds to histone H1 protein [58]. Distamycin, a drug that recognizes poly dA, suppresses binding between MAR and either nuclear matrix or histone H1 [59]. This result suggests that MAR controls chromatin conformation via interaction between poly dA and histone H1 [58]. Furthermore, A-tracts in MAR appear to have an important role in chromosome assembly [60]. Our results indicate that the sequence of clustered A-tracts and/or DNA bending by them is important for MARs to form a higher order structure of DNA in nuclei.

This study found various conformations in MARs, as well as high degrees of sequence-dependent curvature and clustered A-tracts in MARs. These results suggest that MARs can be distinguished from other DNA regions and classified by DNA conformation. With conventional molecular biology, computer-aided analysis of DNA conformation may be a useful tool for examining the interaction between DNA and proteins.

Acknowledgements: We thank Dr. R. Osawa, University of Tsukuba, for his helpful suggestions. This work was supported by Grant-in-Aid from the Ministry of Education of Japan.

References

[1] Brown, T.A. (1999) GENOMES, BIOS Scientific Publishers, Magdalen Road, Oxford.
[2] Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994) Molecular Biology of the Cell, Garland Publishing, 717 Fifth Avenue, New York.
[3] Vogelstein, B., Pardoll, D.M. and Coffey, D.S. (1980) Cell 22, 79–85.
[4] Masuda, K., Takahashi, S., Nomura, K., Arimoto, M. and Inoue, M. (1993) Planta 191, 532–540.
[5] Nomura, K., Saito, W., Ono, K., Moriyama, H., Takahashi, S., Inoue, M. and Masuda, K. (1997) Plant Cell Physiol. 38, 1060–1068.
[6] Berezney, R. and Coffey, D.S. (1974) Biochem. Biophys. Res. Commun. 60, 1410–1417.
[7] Cockerill, P.N. and Garrard, W.T. (1986) Cell 44, 273–282.
[8] Mirkovitch, J., Mirault, M.E. and Laemmli, U.K. (1984) Cell 39, 223–232.
[9] Gasser, S.M. and Laemmli, U.K. (1986) Cell 46, 521–530.
[10] Phi-van, L., von Kries, J.P., Ostertag, W. and Stratling, W.H. (1990) Mol. Cell Biol. 10, 2302–2307.
[11] Klehr, D., Maass, K. and Bode, J. (1991) Biochemistry 30, 1264–1270.
[12] Schubeler, D., Mielke, C., Maass, K. and Bode, J. (1996) Biochemistry 35, 11160–11169.
[13] Namciu, S.J., Blochlinger, K.B. and Fournier, R.E. (1998) Mol. Cell Biol. 18, 2382–2391.
[14] Wiersma, E.J., Ronai, D., Berru, M., Tsui, F.W.L. and Shulman, M.J. (1999) J. Biol. Chem. 274, 4858–4862.
[15] Whitelaw, C.B.A., Grolli, S., Accornero, P., Donofrio, G., Farini, E. and Webster, J. (2000) Gene 244, 73–80.
[16] Mullenders, L.H.F., van Zeeland, A.A. and Natarajan, A.T. (1983) Biochim. Biophys. Acta 740, 428–435.
[17] McCready, S.J. and Cook, P.R. (1984) J. Cell Sci. 70, 189–196.
[18] Boulikas, T. (1993) J. Cell Biochem. 52, 14–22.
[19] Amati, B.B. and Gasser, S.M. (1988) Cell 54, 967–978.
[20] Cockerill, P.N. and Garrard, W.T. (1986) FEBS Lett. 204, 5–7.
[21] Breyne, P., Montagu, M.V., Depicker, A. and Gheysen, G. (1992) Plant Cell 4, 463–471.

[22] Avramova, Z., Tikhonov, A., Chen, M. and Bennetzen, J.L. (1998) Nucleic Acids Res. 26, 761–767.
[23] Nikolaev, L.G., Tsevegiyn, T., Akopov, S.B., Ashworth, L.K. and Sverdlov, E.D. (1996) Nucleic Acids Res. 24, 1330–1336.
[24] Challberg, S.S. and Englund, P.T. (1980) J. Mol. Biol. 138, 447–472.
[25] Asayama, M., Hayasaka, Y., Kabasawa, M., Shirai, M. and Ohyama, T. (1999) J. Biochem. 125, 460–468.
[26] Bageshwar, U.K., Raina, R., Choudhury, N.R. and Das, H.K. (1998) Can. J. Microbiol. 44, 405–415.
[27] Cheema, A.K., Choudhury, N.R. and Das, H.K. (1999) J. Bacteriol. 181, 5296–5302.
[28] Zahn, K. and Blattner, F.R. (1987) Science 236, 416–422.
[29] Bechert, T., Diekman, S. and Arndt-Jovin, D.J. (1994) J. Biol. Struct. Dye 12, 605–623.
[30] Krajewski, W.A. and Razin, S.V. (1992) Mol. Gen. Genet. 235, 381–388.
[31] von Kries, J.P., Phi-van, L., Diekmann, S. and Stratling, W.H. (1990) Nucleic Acids Res. 18, 3881–3885.
[32] Homberger, H.P. (1989) Chromosoma 98, 99–104.
[33] Hibino, Y., Nakamura, K., Tsukada, S. and Sugano, N. (1993) Biochim. Biophys. Acta 1174, 162–170.
[34] Hibino, Y., Tsukuda, S. and Sugano, N. (1997) Carcinogenesis 18, 707–713.
[35] Koo, H.S., Wu, H.M. and Crothers, D.M. (1986) Nature 320, 501–506.
[36] Haran, T.E. and Crothers, D.M. (1989) Biochemistry 28, 2763–2767.
[37] Mielke, C., Kohwi, Y., Kowhi-Shigematsu, T. and Bode, J. (1990) Biochemistry 29, 7475–7485.
[38] Kas, E. and Laemmli, U.K. (1992) EMBO J. 11, 705–716.
[39] Nabirochkin, S., Ossokina, M. and Heidmann, T. (1998) J. Biol. Chem. 273, 2473–2479.
[40] Strissel, P.L., Dann, H.A., Pomykala, H.M., Diaz, M.O., Rowley, J.D. and Olopade, O.I. (1998) Genomics 47, 217–229.
[41] Ulanovsky, L.E. and Trifonov, E.N. (1987) Nature 326, 720–722.
[42] Haran, T.E., Kahn, J.D. and Crothers, D.M. (1994) 244, 135–143.
[43] Bolshoy, A., McNamara, P., Harrington, R.E. and Trifonov, E.N. (1991) Proc. Natl. Acad. Sci. USA 88, 2312–2316.
[44] Shpigelman, E.S., Trifonov, E.N. and Bolshoy, A. (1993) CABIOS 9, 435–440.
[45] Boulikas, T. (1994) J. Cell Biochem. 55, 513–529.
[46] Schatz, T. and Langowski, J. (1997) J. Biomol. Struct. Dyn. 15, 265–275.
[47] Nair, T.M. (1998) FEBS Lett. 422, 94–98.
[48] Rodley, P., Nishikawa, N.S., Hirota, M., Wada-Kiyama, Y., Noguchi, C.T., Oishi, M. and Kiyama, R. (1998) DNA Res. 5, 349–354.
[49] Wada-Kiyama, Y., Kuwabara, K., Sakuma, Y., Oishi, Y., Trifonov, E.N. and Kiyama, R. (1999) FEBS Lett. 444, 117–124.
[50] Bechert, T., Heck, S., Fleig, U., Diekmann, S. and Hegemann, J.H. (1999) Nucleic Acids Res. 27, 1444–1449.
[51] Manning, R.S. and Maddocks, J.H. (1996) J. Chem. Phys. 105, 5626–5646.
[52] Goodsell, D.S. and Dickerson, R.E. (1994) Nucleic Acids Res. 22, 5497–5503.
[53] Gabrielian, A., Vlahovicek, K. and Ponger, S. (1997) FEBS Lett. 406, 69–74.
[54] Arciszewska, L.K. and Craig, N.L. (1991) Nucleic Acids Res. 19, 5021–5029.
[55] Derbyshire, K.M. and Grindley, N.D.F. (1992) EMBO J. 11, 3449–3455.
[56] van Pouderoyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H.A. and Sixma, T.K. (1997) EMBO J. 16, 6044–6054.
[57] York, D. and Reznikoff, W.S. (1997) Nucleic Acids Res. 25, 2153–2160.
[58] Izaurralde, E., Kas, E. and Laemmli, U.K. (1989) J. Mol. Biol. 210, 573–585.
[59] Kas, E., Izaurralde, E. and Laemmli, U.K. (1989) J. Mol. Biol. 210, 587–599.
[60] Strick, R. and Laemmli, U.K. (1995) Cell 83, 1137–1148.