Minireview

What positions nucleosomes? – A model

Ryoiti Kiyama[a,*], Edward N. Trifonov[b,c]

[a] *Research Center for Glycoscience, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan*
[b] *Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel*
[c] *Genome Diversity Center, Institute of Evolution, University of Haifa, Haifa 31905, Israel*

Received 16 May 2002; accepted 5 June 2002

First published online 17 June 2002

Edited by Lev Kisselev

Abstract Here we propose a new determinant for localization of nucleosomes along genomic DNA, in addition to sequence-dependent features. The new specific class of chromatin scaling signals involves curved DNA. According to the observed positional distribution of DNA curvature, the new synchronizing signal occurs once per four nucleosomes on average. This new factor in nucleosome positioning should substantially influence the efficiency of biological reactions through regulatory factors microscopically and the entire chromatin structure through the 30 nm fiber structure macroscopically. Allocation of the new type of signals is found to be fixed evolutionarily although they could be shifted in accordance with the hierarchy of functional genomic structures.

Key words: Chromatin; Nucleosome positioning; Genome structure; Signal for phasing; Higher order structure

1. Signals for nucleosome positioning

'To be positioned or not to be' is a fundamental but difficult question in nucleosome studies. The first notions that nucleosomes occupy specific positions [1], and that phasing or positioning of the nucleosomes is mediated by multiple histone–DNA interactions [2], initiated the hunt for specific signals or mechanisms involved in nucleosome positioning (reviewed in [3,4]). Despite the early argument that nucleosomes are randomly distributed and slide along DNA, it soon became obvious that some positioning signals are, indeed, required for chromatin organization as well as for biological reactions such as transcription, recombination and replication [5]. There is one relevant question not asked heretofore: should each nucleosome have its own unique positioning signal, or could it be uniquely positioned on DNA without the signal of its own? Nucleosomes in the packed state naturally have less freedom than those in vitro or in activated regions in vivo. They may occupy alternative positions in fragmented chromatin in vitro, while in vivo their positions may well be uniquely dictated by other nucleosomes, neighbors in the sequence and neighbors in space. In this case, nucleosome positioning in general may involve a strong synchronizing component, for example, special key nucleosomes firmly placed by a special signal.

Table 1 summarizes the potential or proposed signals for nucleosome positioning. These include signals for rotational and/or translational positioning. The respective sequence motifs may be specific or degenerate, periodically dispersed or localized. The dispersed category is represented by short DNA sequences two to four nucleotides long, such as dinucleotide AA, trinucleotides AAA, VWG and CTG, and tetranucleotides NGGR and TGGA. The effects of these short sequences are magnified by repetitive appearance at roughly 10–11 bp intervals along the DNA double helix as well as in the case of longer sequences $(A/T)_3NN(G/C)_3NN$ and $A_5(G/C)_5$. Some of these signal sequences are closely related to specific DNA structures: for example, AA/TT or AAA/TTT for curved DNA, $(CA)_n$ or $(CG)_n$ for Z-DNA, or Pu·Py (polypurine·polypyrimidine) sequences for triplex DNA. In contrast, positioning signals belonging to the localized category are associated with some functions in which the nucleosomes are structurally involved. Enhancers or binding sites for transcription factors can signal the nucleosomal positions. Exclusion of nucleosome cores by some sequences or preferred localization of some ligands within linkers mediated by specific DNA sequences or DNA structures can also enforce the nucleosome positioning. Other signals such as acetylation of histones, methylation of DNA and topological status of chromatin and chromosomal DNA can also influence nucleosome positions.

Curved DNA has been considered a likely ubiquitous signal for nucleosome positioning for many years. On the basis of the results of analysis of nucleosome positioning sequences in the mouse mammary tumor virus, Pina et al. [6] noted that curved DNA could determine the nucleosome positions by mere similarity to the bent DNA within the nucleosome core. By examining 204 nucleosome DNA sequences, Ioshikhes et al. [7] found periodicity of AA and TT dinucleotides at an interval of 10.3 bp within nucleosomes, that may also cause some DNA curvature. Numerous sites of curved DNA were found in many loci [8], and mapping the DNA bend sites in the 66 kb region of the human β-globin locus revealed that they appear at an average interval of 680 bp [9]. A close relationship between the computer-predicted nucleosome positions and the mapped DNA bend sites was observed in the human estrogen receptor α gene [10]. We suggest that these widespread bend sites at an average distance of about 680 bp may be considered the universal signal for group positioning of the nucleosomes.

*Corresponding author. Fax: (81)-298-61 6190.
E-mail address: kiyama.r@aist.go.jp (R. Kiyama).

0014-5793/02/$22.00
PII: S0014-5793(02)02937-X

Table 1
Proposed nucleosome positioning signals

Nucleosome positioning signal	Ref.
Specific nucleotide sequences	
SV40 enhancer	[27]
Satellite DNA	[28]
$(A/T)_3NN(G/C)_3NN$	[29,30]
AAA/TTT	[31]
Albumin enhancer	[32]
NGGR	[33]
AA(inside)/TT(outside)	[7]
Non-T(A/T)G (or VWG)	[34,35]
A+T-rich sequence in 5S rRNA gene	[36]
Cis-elements for TFIIIA	[37]
$[A_5(G/C)_5]_4$	[38]
$(CTG)_n$	[39]
Specific DNA structures	
Curved DNA	[40–42]
Sequences/structures that disrupt nucleosome formation	
Z-DNA (CA or CG repeats)	[43]
Curved DNA at yeast ARS	[44]
Triplex DNA (Pu·Py sequences)	[45,46]
$T_{14}A_{11}$ of *Alu* element	[47]
TGGA	[48]
Binding of proteins/transcription factors	
Histone H1(H5)	[49]
Histone $(H3–H4)_2$	[50]
NF1 of MMTV	[51,52]
Histone H4	[53]
Sp1 and NFκB	[54]
Adf-1 and GAGA factor	[55]
Others	
Histone acetylation	
DNA methylation	
Superhelicity/chromatin folding	

2. Why are nucleosomes specifically positioned?

One of the advantages of nucleosome positioning is efficient formation of higher order chromatin structures and folding these structures into a compact space. Compaction of metaphase chromosomes could be achieved through organized behavior of chromatin, which starts from the lowest level of structure, the nucleosome. The structure of the next level of chromatin organization, the 30 nm fiber, is uncertain. Recent findings suggested that nucleosomes in the fiber are organized in a zig-zag manner with two nucleosomes as a unit [11] and with definite linker lengths of 10*n* or 10*n*+5 bp [12]. The change in the linker length may lead to dramatic changes in the chromatin fiber structure since every extra base pair in the linker between two nucleosomes would cause their relative rotation in 3D by about 35° [13,14]. An accurate nucleosome positioning in the 5S RNA gene region of *Xenopus borealis* demonstrated that the nucleosome-to-nucleosome distances are all different within the region [15], suggesting very specific relative orientations of the nucleosomes in space. A general irregular and yet unique model of the 30 nm fiber is suggested [16]. The fiber may have some aspects of regularity as well. For example, a dinucleosome unit of chromatin structure was suggested in an early work [17,18]. Studies on the distribution of the bend sites in DNA also indicate a tetranucleosome as the unit of the next level of chromatin organization [8]. In all cases, a specific spatial positioning (orientation) of individual nucleosomes in the fiber is expected. The nucleosome positioning would then be an essential determinant of the entire chromatin formation.

Another important role of nucleosome positioning is related to the function of the genome. The information encoded in the genomic DNA in the form of *cis*-acting elements is translated by *trans*-acting factors. When the chromatin is present, it could act as a structural barrier, a modulator of accessibility of the regulatory elements. Pina et al. [19] reported that specific nucleosome positions were needed for the function of the mouse mammary tumor virus promoter. For integration of HIV, specific positioning of nucleosomes at the integration sites was observed as well [20]. As reviewed by Beato and Eisfeld [21], precise positioning of *trans*-acting factors and availability of binding motifs to them are of key importance for their functions. This interaction is realized either by free access or by direct or indirect access (discussed below), but when the locus is closed and the chromatin is tightly packed only the first of several coordinated *trans*-factors can directly access the respective exposed motif. This is why chromatin remodeling is needed for biological reactions, transcription for example, to change the structural chromatin environment from the inactive to the active state. Once the reaction has started, nucleosome positions would be less important because chromatin is now rearranged for the given function and nucleosomes may slide or unfold to be replaced by the complex of *trans*-acting factors and RNA polymerases. An intricate correlation between positions of transcription factor binding sites and of nucleosomes is revealed by recent analysis of the promoter sequences and flanking regions [22]. The distribution of potential nucleosome positions around the promoters suggests at least two different linear and spatial arrangements of the nucleosomes.

3. How are nucleosomes positioned?

This question could be answered in three not necessarily unrelated ways: perfect positioning of each nucleosome, partial positioning or random placement. As it is now clear that there are no specific arrays of nucleotide sequences repeating as massively as nucleosomes, one should search for some less frequent but stronger specific signals, perhaps of a structural nature. Among many unusual DNA structures, the curved DNA is the most structurally attractive and widespread to serve as the potential additional positioning signal. DNA bend sites show a wide variety of nucleotide sequences ([23], for example). Most known non-B DNA structures are formed by simple repetition of nucleotide sequences, while in the nucleotide sequences that can form curved DNA it is the combination of different wedge angles that causes deflection of the axis of the double helix, and the sequences for the curved DNA, generally, are not repetitive. Among 10 different dinucleotide steps, the step AG (CT) provides the highest wedge angle of 8.4°. The dinucleotides AA (TT), frequent in the genomic DNA of higher eukaryotes, provide 7.2° and contribute most to the curvature [24].

As described above, searching for sites of DNA curvature, or DNA bend sites, revealed the nearly periodic appearance of the mapped bend sites in many genes. A thorough mapping of the sites in the human β-globin locus revealed 98 such sites with an average nearly regular spacing of 680 bp [9]. The minor peaks observed in the histogram of distribution of

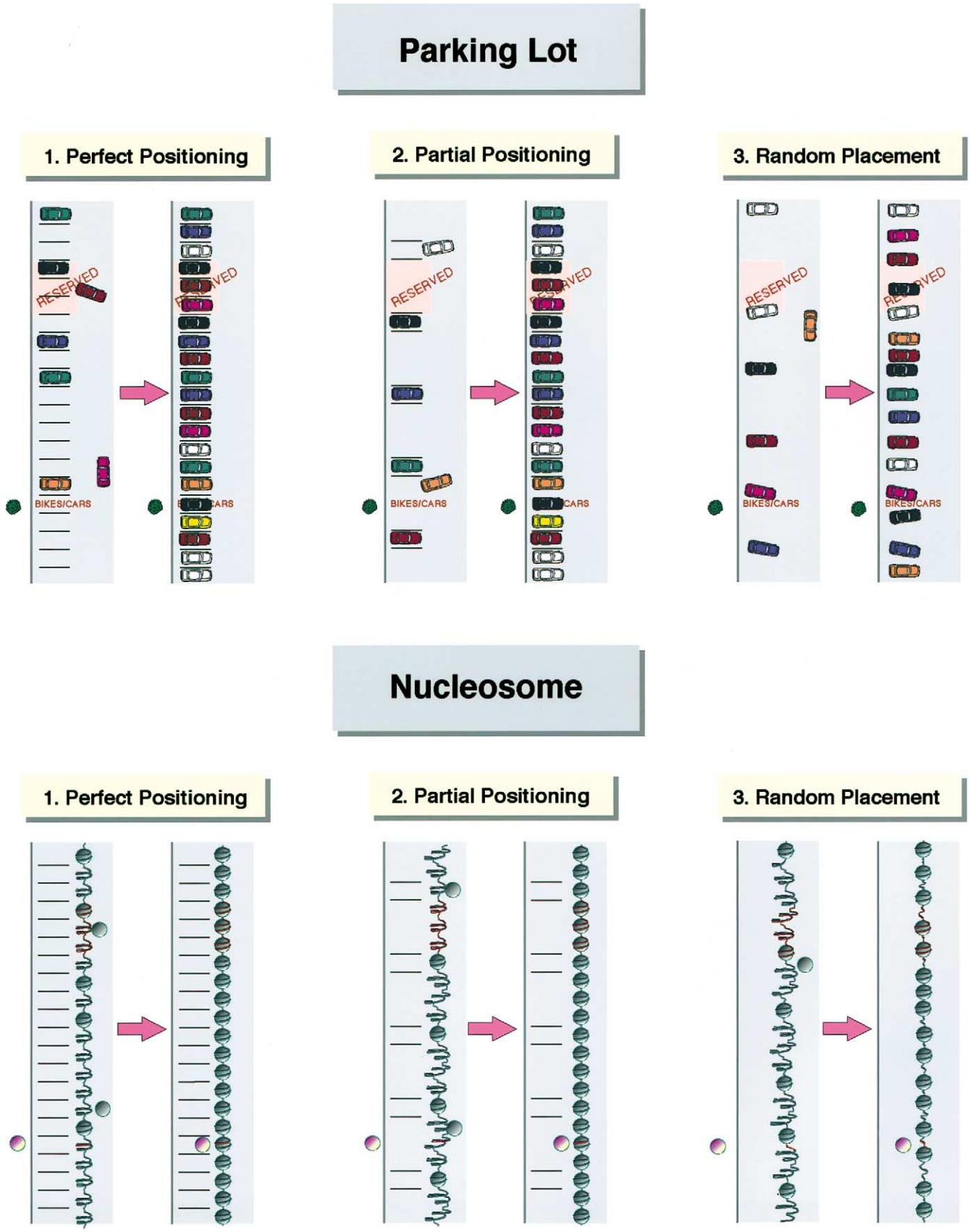

Fig. 1. Parking lot model of nucleosome positioning. Positioning of cars in a parking lot is used to explain how nucleosomes (cars) are positioned by specific signals (lines on the lot). In the case of perfect and partial positioning, parking of cars starts from the designated spaces and, subsequently, empty spaces are filled uniquely. Likewise, nucleosome positioning starts at the key nucleosomes, which are marked by specific signals. In random placement, parking efficiency is expected to be less. The figure also shows that the signals are included or excluded from the coding region (the RESERVED parking space painted in red) and that a transcription factor (tree) interacts with nucleosomes at a specific position 'BIKES/CARS'.

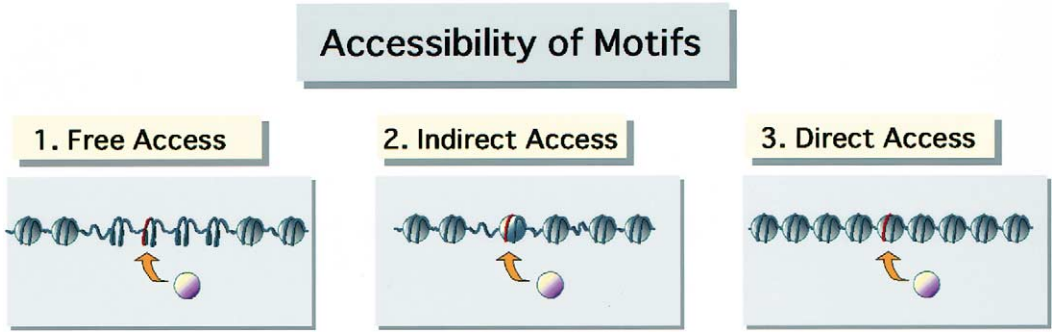

Fig. 2. Accessibility of transcription factors to the cognate binding motifs on nucleosomes (shown in red). A transcription factor can access the binding motif freely (free access), indirectly mediated by another factor (highlighted) already bound to the motif (indirect access), or by direct access to nucleosome DNA in the presence of a nucleosome (direct access). The controlled targeting of the motif could be achieved through positioning of the nucleosome that contains the motif or through nucleosome positioning from the neighboring key nucleosome (see 'BIKES/CARS' in Fig. 1).

the sites correspond to multiples of 170 bp, and the 680 bp average distance, the highest peak, corresponds to the tetramer of the 170 bp unit. This result indicated that one of four nucleosomes on average is associated with curved DNA. In support of this, we provided evidence that nucleosomes are positioned uniquely in vivo as well as in vitro at these bend sites [25]. Removing these sites abolished the positioning not only at the bend sites but also at those in the direct vicinity. This suggested that although the curvature signal involves only one nucleosome, the positions of the other three nucleosomes of the tetranucleosome unit are indirectly determined by the curvature signal as well. This gives the basis for the idea of secondary externally dictated nucleosome positioning.

A model of such secondary nucleosome positioning, the parking lot model, is shown in Fig. 1. This analogy is driven by the notion that perfect positioning of cars in the lot can be achieved not only by a full set of individual prescribed borders but also by the 'no-choice' positioning if the distance between only scarcely marked borders is not far enough for cars to position randomly. The unmarked space with a distance sufficient for three cars, as in the figure, may accommodate these three cars, so that the positions of all cars become uniquely determined. This can be applied to nucleosomes (Fig. 1, bottom). One of the disadvantages of perfect positioning is that the strong signal for every nucleosome would pose a rather strong constraint on the nucleotide sequence. This would be especially disadvantageous within the coding regions. The described scheme of partial positioning where only one of few nucleosomes is positioned firmly by a specific (in this case) DNA curvature signal suggests a hierarchy of nucleosome signals in the genome. The passively positioned nucleosomes between strong synchronizing signals may or may not be additionally secured by weak sequence preference that would appear in in vitro experiments as one of several weak alternative positions. The alternatives are excluded after the 'cars' all park cooperatively. Such a hierarchy of the nucleosome signals would be necessary when the coding regions are involved. For example, the curvature periodicity is disrupted in the coding regions of the globin genes [9]. No apparent DNA bend sites were found within the exons of all genes that were analyzed to date.

4. When are the nucleosomes positioned?

The presence of a localized strong signal for cooperative nucleosome positioning suggests that pre-fixed positioning could be disrupted at specific times in development or differentiation, changing the accessibility of the otherwise hidden target motifs for binding proteins. Such an epigenetic change of the genome can be seen in globin gene switching. The locus control region of the human β-globin locus is located 6 to more than 20 kb upstream of the first gene to be expressed, the ε-globin gene, and is represented by a collection of DNase I hypersensitive sites, HS1 to HS5. The accessibility of NF-E2, which confers the enhancer activity, to its binding site in HS2 is mediated by remodeling factors in the presence of ATP [25]. This can be achieved in three possible ways (Fig. 2). The obvious case is when the respective motif is not in contact with histones (free access). If the chromatin structure is present, accessibility should be much higher if the motif is efficiently exposed to the incoming factor by facing a specific direction due to specific nucleosome positioning and, thus, orientation (direct access). This would also be true if there is an additional factor that helps this recognition process (indirect access). Note that replacing the nucleosome with such a factor requires specific recognition of the binding site through indirect or direct access. The NF-E2 site alone has no nucleosome positioning activity and removing the neighboring positioning DNA bend site reduces the enhancer activity [26], suggesting that proper nucleosome location is essential for its function.

One could imagine that selection pressures would introduce or abolish these sites over long periods of time (Fig. 3). After a sequence is inserted and the distance between the signals is

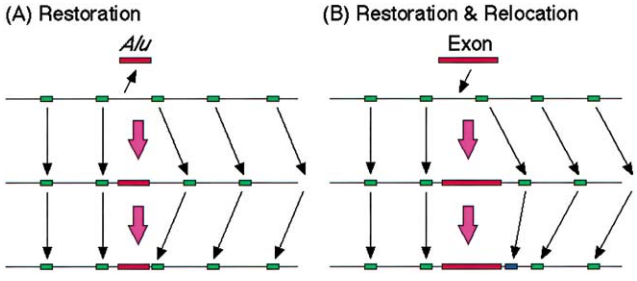

Fig. 3. Mechanisms of evolutionary adjustment and conservation of DNA bend sites. Periodicity of DNA bend sites is restored to the original by deletion of a short sequence with the identical length to that of the inserted *Alu* element (A) but some sites could be permanently relocated after insertion of a longer sequence (B).

changed, these signals can be returned to the original position by shortening the sequence. This may be illustrated by the insertion of *Alu* sequences found upstream of the human ε-globin gene. The curvature signal could be shifted permanently when a long sequence, such as an exon, is inserted. Thus, one may conclude that the nucleosomes have to be positioned at least for the specific modulation of the accessibility of regulatory sites. Quite likely, the positioning is important as well for the unique folding of the chromatin, its overall architecture.

5. Is it universal?

So far, only over 100 examples of mapped DNA bend sites are available, all of them located in genomes of higher animals. Although the chromatin structure is universal, genomes show a very wide degree of diversification, which can be seen in the genome size, gene density, base composition, frequency of specific nucleotide arrays, and mutation or evolution rates. Many types of nucleotide sequences could be used as signals for cooperative nucleosome positioning. Since the nucleosome structure is conserved from lower to higher eukaryotes, we believe that the DNA structural design for the synchronizing positioning signal is likely to be conserved as well. That is, the curved DNA may well be, indeed, a universal primary nucleosome positioning signal.

Acknowledgements: We thank J. Beckmann for early discussions on the parking lot model and M. Oishi and T. Ikemura for valuable comments.

References

[1] Lohr, D., Tatchell, K. and van Holde, K.E. (1977) Cell 12, 829–836.
[2] Sussman, J.L. and Trifonov, E.N. (1978) Proc. Natl. Acad. Sci. USA 75, 103–107.
[3] Thoma, F. (1992) Biochim. Biophys. Acta 1130, 1–19.
[4] Wolffe, A.P. (1994) Trends Biol. Sci. 19, 240–244.
[5] Wolffe, A.P. (1995) Chromatin: Structure and Function, Academic Press, New York.
[6] Pina, B., Barettino, D., Truss, M. and Beato, M. (1990a) J. Mol. Biol. 216, 975–990.
[7] Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M. and Trifonov, E.N. (1996) J. Mol. Biol. 262, 129–139.
[8] Kiyama, R. (1998) Gene Ther. Mol. Biol. 1, 641–647.
[9] Wada-Kiyama, Y., Suzuki, K. and Kiyama, R. (1999) Mol. Biol. Evol. 16, 922–930.
[10] Wada-Kiyama, Y., Kuwabara, K., Sakuma, Y., Onishi, Y., Trifonov, E.N. and Kiyama, R. (1999) FEBS Lett. 444, 117–124.
[11] Rydberg, B., Holley, W.R., Mian, I.S. and Chatterjee, A. (1998) J. Mol. Biol. 284, 71–84.
[12] Widom, J. (1992) Proc. Natl. Acad. Sci. USA 89, 1095–1099.
[13] Noll, M., Zimmer, S., Engel, A. and Dubochet, J. (1980) Nucleic Acids Res. 8, 21–42.
[14] Ulanovsky, L.E. and Trifonov, E.N. (1986) in: Biomolecular Stereodynamics III (Sarma, R.H. and Sarma, M.H., Eds.), pp. 35–44, Adenine Press, New York.
[15] Drew, H.R. and Calladine, C.R. (1987) J. Mol. Biol. 195, 143–173.
[16] Trifonov, E.N. (1991) in: Theoretical Biochemistry and Molecular Biophysics, Vol. 1: DNA (Beveridge, D.L. and Lavery, R., Eds.), pp. 377–388, Adenine Press, New York.
[17] Burgoyne, L.A. and Skinner, J.D. (1981) Biochem. Biophys. Res. Commun. 99, 893–899.
[18] Khachatrian, A.T., Pospelov, V.A., Svetlikova, S.B. and Vorobiev, V.I. (1981) FEBS Lett. 128, 90–92.
[19] Pina, B., Brüggemeier, U. and Beato, M. (1990b) Cell 60, 719–731.
[20] Pruss, D., Bushman, F.D. and Wolffe, A.P. (1994) Proc. Natl. Acad. Sci. USA 91, 5913–5917.
[21] Beato, M. and Eisfeld, K. (1997) Nucleic Acids Res. 25, 3559–3563.
[22] Ioshikhes, I., Trifonov, E.N. and Zhang, M.Q. (1999) Proc. Natl. Acad. Sci. USA 96, 2891–2895.
[23] Koo, H.-S., Wu, H.-M. and Crothers, D.M. (1986) Nature 320, 501–506.
[24] Bolshoy, A., McNamara, P., Harrington, R.E. and Trifonov, E.N. (1991) Proc. Natl. Acad. Sci. USA 88, 2312–2316.
[25] Onishi, Y., Wada-Kiyama, Y. and Kiyama, R. (1998) J. Mol. Biol. 284, 989–1004.
[26] Onishi, Y. and Kiyama, R. (2001) Nucleic Acids Res. 29, 3448–3457.
[27] Clarke, M.F., FitzGerald, P.C., Brubaker, J.M. and Simpson, R.T. (1985) J. Biol. Chem. 260, 12394–12397.
[28] Linxweller, W. and Horz, W. (1985) Cell 42, 281–290.
[29] Shrader, T.E. and Crothers, D.M. (1989) Proc. Natl. Acad. Sci. USA 86, 7418–7422.
[30] Tanaka, S., Zatchej, M. and Thoma, F. (1992) EMBO J. 11, 1187–1193.
[31] Muyldermans, S.V. and Travers, A.A. (1994) J. Mol. Biol. 235, 855–870.
[32] McPherson, C.E., Horowitz, R., Woodcock, C.L., Jiang, C. and Zaret, K.S. (1996) Nucleic Acids Res. 24, 397–404.
[33] Travers, A.A. and Muyldermans, S.V. (1996) J. Mol. Biol. 257, 486–491.
[34] Baldi, P., Brunak, S., Chauvin, Y. and Krogh, A. (1996) J. Mol. Biol. 263, 503–510.
[35] Stein, A. and Bina, M. (1999) Nucleic Acids Res. 27, 848–853.
[36] Tomaszewski, R. and Jerzmanowski, A. (1997) Nucleic Acids Res. 25, 458–466.
[37] Pfaff, S.L. and Taylor, W.L. (1998) Mol. Cell. Biol. 18, 3811–3818.
[38] Fitzgerald, D.J. and Anderson, J.N. (1998) Nucleic Acids Res. 26, 2526–2535.
[39] Godde, J.S. and Wolffe, A.P. (1996) J. Biol. Chem. 271, 15222–15229.
[40] Drew, H.R. and Travers, A.A. (1985) J. Mol. Biol. 186, 773–790.
[41] Fitzgerald, D.J., Dryden, G.L., Bronson, E.C., Williams, J.S. and Anderson, J.N. (1994) J. Biol. Chem. 269, 21303–21314.
[42] Wada-Kiyama, Y. and Kiyama, R. (1994) J. Biol. Chem. 269, 22238–22244.
[43] Nickol, J., Behe, M. and Felsenfeld, G. (1982) Proc. Natl. Acad. Sci. USA 79, 1771–1775.
[44] Snyder, M., Buchman, A.R. and Davis, R.W. (1986) Nature 324, 87–89.
[45] Westin, L., Blomquist, P., Milligan, J.F. and Wrange, O. (1995) Nucleic Acids Res. 23, 2184–2191.
[46] Espinas, M.L., Jimenez-Garcia, E., Martinez-Balbas, A. and Azorin, F. (1996) J. Biol. Chem. 271, 31807–31812.
[47] Englander, E.W. and Howard, B.H. (1996) J. Biol. Chem. 271, 5819–5823.
[48] Cao, H., Widlund, H.R., Simonsson, T. and Kubista, M. (1998) J. Mol. Biol. 281, 253–260.
[49] Stein, A. and Künzler, P. (1983) Nature 302, 548–550.
[50] Dong, F. and van Holde, K.E. (1991) Proc. Natl. Acad. Sci. USA 88, 10596–10600.
[51] Archer, T.K., Cordingley, M.G., Wolford, R.G. and Hager, G.L. (1991) Mol. Cell. Biol. 11, 688–698.
[52] Eisfeld, K., Candau, R., Truss, M. and Beato, M. (1997) Nucleic Acids Res. 25, 3733–3742.
[53] Roth, S.Y., Shimizu, M., Johnson, L., Grunstein, M. and Simpson, R.T. (1992) Genes Dev. 6, 411–425.
[54] Widlak, P., Gaynor, R.B. and Garrard, W.T. (1997) J. Biol. Chem. 272, 17654–17661.
[55] Gao, J. and Benyajati, C. (1998) Nucleic Acids Res. 26, 5394–5401.