

## Nucleosome DNA Bendability Matrix (*C. elegans*)

<http://www.jbsdonline.com>

**I. Gabdank<sup>1</sup>**  
**D. Barash<sup>1</sup>**  
**E. N. Trifonov<sup>2,3,\*</sup>**

### Abstract

An original signal extraction procedure is applied to database of 146 base nucleosome core DNA sequences from *C. elegans* (S. M. Johnson *et al. Genome Research* 16, 1505-1516, 2006). The positional preferences of various dinucleotides within the 10.4 base nucleosome DNA repeat are calculated, resulting in derivation of the nucleosome DNA bendability matrix of 16×10 elements. A simplified one-line presentation of the matrix ("consensus" repeat) is ...A(TTTCCGGAAA)T.... All 6 chromosomes of *C. elegans* conform to the bendability pattern. The strongest affinity to their respective positions is displayed by dinucleotides AT and CG, separated within the repeat by 5 bases. The derived pattern makes a basis for sequence-directed mapping of nucleosome positions in the genome of *C. elegans*. As the first complete matrix of bendability available the pattern may serve for iterative calculations of the species-specific matrices of bendability applicable to other genomic sequences.

Key words: Nucleosome; Nucleosome DNA; Nucleosome positioning; Dinucleotides; Bendability matrix.

### Introduction

Since discovery of weak ~10.5 base periodicity of AA/TT in chromatin DNA sequences, which is the second major message in eukaryotic DNA that overlaps with 3-base periodical protein-coding message (1, 2), some progress has been made in the unraveling of the sequence structure of the nucleosome DNA and, specifically, sequence structure of the hidden 10.4 base repeat of the nucleosome DNA. Two major schools are competing today in the attempts to derive the complete sequence pattern characteristic for the nucleosome DNA. The "counter-phase" school claims that the RR and YY dinucleotides, in particular, AA and TT dinucleotides, are distributed along nucleosome DNA periodically, in alternating RR/YY (AA/TT) fashion (2-8). The "in-phase" school maintains that the periodical AA and TT dinucleotides are in the same phase within the repeat unit (9-11). Lack of sufficiently accurate experimental data on nucleosome positioning, and exceptionally weak sequence signal are to blame for the longevity of the controversy.

One of currently used well performing nucleosome sequence patterns, AA and TT dinucleotides only (counter-phase type), has been derived by multiple alignment of about 200 experimental nucleosome sequences, determined with various positional accuracies (4). Calculation of a complete nucleosome sequence pattern, with all important details, would be possible only with substantially larger collection of experimental sequences. Very large databases of nucleosome DNA fragments and of full 146 base long nucleosome core DNA sequences of *C. elegans* became recently available (12). With this resource it was possible to partially reconstruct the sequence structure of the 10.4 base repeat in [R, Y] alphabet: (YYYYYRRRRR)<sub>n</sub>, and in [A, T] alphabet: (TTTYTARAAA)<sub>n</sub>, by analysis of preferred distances between various dinucleotides (8). In this work we describe full reconstruction of the 10.4 base repeat of nucleosome DNA of *C. elegans* from the large collection of 146

<sup>1</sup>Department of Computer Science  
Ben Gurion University of the Negev  
P.O.B 653

Be'er Sheva 84105, Israel

<sup>2</sup>Genome Diversity Center  
Institute of Evolution, University of Haifa  
Mount Carmel, Haifa 31905, Israel

<sup>3</sup>Division of Functional  
Genomics and Proteomics  
Faculty of Science  
Masaryk University, Kamenice 5  
Brno CZ-62500, Czech Republic

\*Phone: +972 4 828 8096  
Fax: +972 4 824 6554  
Email: trifonov@research.haifa.ac.il

base nucleosome core DNA sequences (12). Distributions of all 16 dinucleotides within the 10.4 base repeat are derived. An original computational procedure is introduced for that purpose – determination of preferred positions within the period for different dinucleotides by alignment of 12-base sequence fragments with the same dinucleotides at the ends; thus, deriving matrix of nucleosome DNA bendability (2, 3) in its modern version.

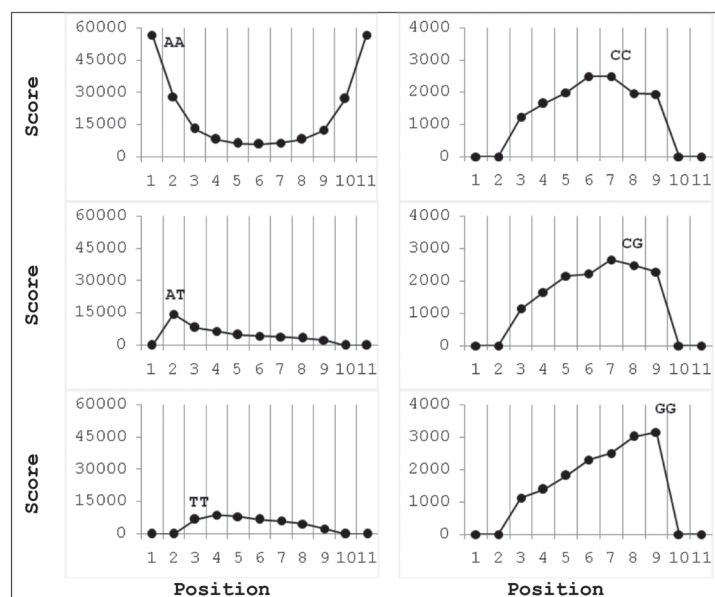
### Methods

#### Database of the 146 Base Nucleosome Core Sequences

The sequences are taken from (12), database UUPc (Unique Unambiguous Pyrocore). It contains 28,230 core sequences from Chromosome I of *C. elegans*, 30,310 sequences from Chromosome II, 26,111 – Chromosome III, 30,177 – Chromosome IV, 39,547 – Chromosome V, and 33,488 sequences from Chromosome X. The database is available at the website [www.genome.org](http://www.genome.org), in section of Supplements.

#### Matrices of Optimal Positions

For calculation of the AA8AA matrix all occurrences of the motif in the database of 146 base core sequences (12) are aligned, and scores of every dinucleotide in all 10 positions of the motif are placed in the 16×10 matrix. Total occurrence of the AA8AA motif in the nucleosome cores of Chromosome I is about 57.000 (Fig. 1), that is approximately two encounters per every nucleosome. Since the most preferred positions within the period are often uncertain, 5 highest score positions of 10 are taken for each dinucleotide to indicate the range where the optimal positions may be located. For example (Fig. 1), in the AA8AA matrix the highest score positions for AA are 1-3, 9 and 10, for AT – positions 2-6, TT – 3-7, CC, CG, and GG – positions 5-9. The cumulative matrix of highest score positions (Fig. 2) is calculated by summing results obtained from matrices AA8AA, AT8AT, TT8TT, and CG8CG, aligned by CG (position 1 in the cumulative matrix), which is located at (maximal) position 7 in AA8AA matrix (Fig. 1), at position 6 in AT8AT matrix, and at 5 – in TT8TT matrix (not shown).



**Figure 1:** Distribution of dinucleotides AT, TT, CC, CG, and GG within AA8AA segments extracted from nucleosome core DNA of Chromosome I. Positions 1 and 11 are occupied by AA.

#### Matrices of Bendability

The optimal position matrices, with all symmetrical elements, taken with equal weights (Figs. 2 and 4), are used for detection of best matches to the positional matrix in the core sequences, one best match per each nucleosome core sequence. Detected

11 base long sequence segments are aligned and absolute (not shown) and relative scores (ratio of occurrence in given position to average score over 10 positions) are calculated for the occurrences of all 16 dinucleotides in 10 positions of the matrix.

### **Results and Discussion**

#### *10.4 Base Repeat in the Nucleosomes of Chromosome I*

In the analysis below it is assumed as in (2, 3) that certain proportion of the dinucleotide stacks in the nucleosome DNA are oriented relative to the surface of the histone octamer in an optimal way, to make easier the deformation of DNA in the direction towards the surface. If the same stack is placed one DNA helical repeat away, it has the same orientation. The problem of reconstruction of the nucleosome positioning pattern is, thus, reduced to finding out what are optimal orientations of 16 different base-pair stacks within one period of the nucleosome DNA. The different optimal orientations, of course, correspond to different positions within the period, from 1 to 10 (or from 1 to 11, as the non-integer period 10.4 bases would suggest). Since there are 16 dinucleotides and only 10 or 11 positions within the period, more than one of different optimally oriented stacks may prefer the same position within the period. In this case the solution would have a form of a 16×10 matrix, with the 16 dinucleotides as elements, located in their respective optimal positions. Such matrix would only indicate the positions. In order to fully describe the 10.4 base repeat unit, the “strengths” of the positions – the deformational affinities – have to be calculated for every dinucleotide element and every position within the period. That would make the matrix of bendability (2, 3). The original version of the matrix has been used 25 years ago for the first successful comparisons of the calculations with experimentally derived nucleosome positions in several sequences available at that time (2, 3).

There are two axes of dyad symmetry for the 10 base-pair long segment of DNA bent in the nucleosome. The axes pass through minor grooves – one facing the histone octamer, and another one, 5 bases away, – facing outwards. The positions that are left and right from the minor grooves, symmetrically, are identical in terms of deformability of respective (complementary) base-pair stacks. In the consensus sequence (matrix) of the 10-base repeat there are also two respective positions, points of complementary symmetry. Most likely, these positions would be occupied by some of four self-complementary dinucleotides (AT, CG, GC, TA), at distance 5 from one another.

Since the method of the calculations is new and it is implemented for the first time, some non-technical description of the method would be appropriate. The idea is that if the signal (pattern) is known only very approximately, the first approximation pattern (few signal elements properly positioned) may help to extract from the sequences the whole pattern, *i.e.*, other signal elements and their positions within the pattern. In other words, one would expect to regenerate the signal from only few known parts of it. For this, all occurrences of good match to the approximate pattern should be collected and analyzed. If the signal is strong enough, at certain positions within the collected sequences higher occurrences of some elements (presumably, signal elements) will be observed.

We have chosen for the first round of calculations the minimal pattern of repeating AA dinucleotides, that is, the sequence AAxxxxxxxAA (AA8AA) where AA appears again after 10 base steps. As the 10-11 base periodicity of the AA dinucleotides in the nucleosome DNA is well established fact, encounter of the motif AA8AA in a given nucleosome DNA sequence is quite likely. Moreover, it would be expected to occur there more often than in random sequence. When large amount of nucleosome sequences is screened and the AA8AA motifs collected, the proportion of those AA8AA motifs that actually participate in the nucleosome

	1	2	3	4	5	6	7	8	9	10	11
AA	2	2	<b>4</b>	<b>4</b>	<b>4</b>	2	2	0	0	0	2
TT	2	0	0	0	2	2	<b>4</b>	<b>4</b>	<b>4</b>	2	2
AG	3	3	<b>4</b>	3	2	2	1	1	0	1	3
CT	3	1	0	1	1	2	2	3	<b>4</b>	3	3
GA	3	<b>4</b>	<b>4</b>	<b>4</b>	2	0	0	0	1	2	3
TC	3	2	0	0	1	0	2	<b>4</b>	<b>4</b>	<b>4</b>	3
GG	3	<b>4</b>	3	1	1	1	0	1	<b>4</b>	2	3
CC	3	2	3	1	0	1	1	2	<b>4</b>	<b>4</b>	3
AC	2	1	3	2	2	2	2	1	2	3	2
GT	2	3	2	1	1	2	2	2	3	2	2
CA	3	3	3	3	1	1	1	1	2	2	3
TG	3	2	0	1	1	1	1	<b>4</b>	<b>4</b>	3	3
TA	1	1	<b>4</b>	3	2	0	1	3	<b>4</b>	2	1
AT	1	2	3	2	1	<b>4</b>	2	1	2	2	1
CG	<b>4</b>	2	2	2	1	1	0	2	3	3	<b>4</b>
GC	2	3	<b>4</b>	1	0	0	1	1	<b>4</b>	<b>4</b>	2

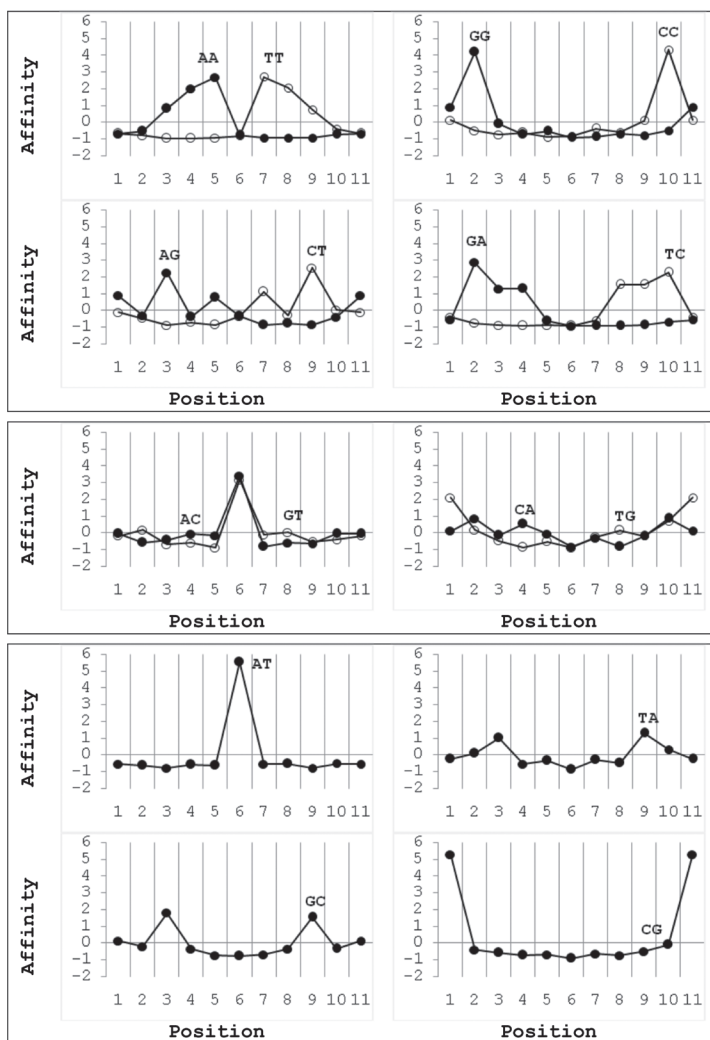
**Figure 2:** Optimal positions of all dinucleotides within 10-base repeats. Combined matrix derived by alignment of respective AA8AA, AT8AT, TT8TT, and CG8CG matrices, with common start at CG (positions 1 and 11). The values from 0 to 4 indicate how often given position in the combined matrix corresponds to excessive scores in the four matrices. Complementarily symmetrical pairs of consensus preferential positions are shadowed. The basic 16×10 matrix is boxed. Column 11 (beginning of next 10-base repeat) is added to illustrate the symmetry of the repeat. The axes of symmetry are indicated by arrows.

positioning may be sufficient to reveal how the dinucleotides located in the inner positions of the motif, are distributed. Total 56,699 occurrences of AA8AA found exceed by  $8,850 \pm 220$  their amount (47,850) in respective shuffled sequences with the same proportions of dinucleotides (data not shown), which is very significant excess. Some of the dinucleotides may prefer specific positions within the period. All such elements together, with their specific positions will represent one period of the nucleosome positioning pattern. Thus, by collecting the AA8AA motifs from all 146 base nucleosome core sequences of Chromosome 1 and counting occurrences of all 16 dinucleotides within the repeat, the distributions of the 16 dinucleotides in 9 available positions (including positions Ax and xA) are calculated. This simple signal processing procedure is based on the accumulation of signal (preferential occurrence of a given dinucleotide in given position within the period). In the large collection of the 12-base long sequences even weak preference of a given dinucleotide to a given position within the repeat may result in detectable excess in the score. Ideally, one would expect that every dinucleotide would show its preferred position. In reality, due to inherent weakness of the signal, the sequence noise components may obscure the pattern. Fortunately, due to large size of the core DNA database used (12) the signal turned out to be well detectable. In the Figure 1 distributions of some of the dinucleotides within the repeat AA8AA are shown, with clear preferences to their respective optimal positions: AA at positions 1, 2 and 10; AT at positions 2 and 3; TT at 4 and 5; CC at 6 and 7; CG at 7 and 8; and GG at 8 and 9. As expected, the self-complementary dinucleotides, AT and CG in this case, have their maxima separated by 5 bases. Similar calculations with motifs TT8TT, AT8AT, and CG8CG revealed the same 5-base distance between preferred positions of AT and CG, and relative positioning of other dinucleotides very similar to what is observed in case of AA8AA (not shown). All four sets of distributions, aligned by CG (placed at position 1) are combined together in the Figure 2. Every line of this matrix shows in how many of the four sets the respective dinucleotide appears at given position as excessive (see *Methods*). The numbers 4 (in bold), thus, correspond to consensus positions of higher occurrences in all four cases. Zeros, on the other hand, mark the consensus positions where the respective dinucleotides are avoided. Similar calculations with other 12-base motifs show that 10 of total 16 motifs confirm the matrix in the Figure 2 (AG8AG, CT8CT, GA8GA, TC8TC, GG8GG, CC8CC, in addition to four motifs on which the matrix is built), while remaining 6 motifs (AC8AC, GT8GT, CA8CA, TG8TG, GC8GC, and TA8TA) do not make any clear pattern of preferences, due to their weak contributions to the overall pattern (see below).

One striking feature of the combined positional matrix in the Figure 2 is its nearly ideal complementary symmetry, as expected. For example, AA occupies positions 3-5, while TT – symmetrical positions 7-9, AG – position 3 and complementary CT – position 9, and so on. The preferred dinucleotide positions and their symmetrical counterparts are shadowed.

With the optimal positions of various dinucleotides determined, one can now estimate the relative affinities of the dinucleotides to these positions, that is, to establish the matrix of bendability. For this the matrix of positions is first used to locate best fit 10-base segments in the collection of the 146 base core sequences. Summing the respective occurrences of various dinucleotides within the 10-base period one derives the 1<sup>st</sup> approximation matrix of scores (not shown) that, on one hand, confirms the optimal positions (maximal scores). On the other hand, if any of the optimal positions have been missed in the calculation of the positional matrix (Fig. 2), they may now appear in the 1<sup>st</sup> approximation matrix of scores as a result of regeneration of missing signal components. Indeed, no consensus

positions have been indicated, for example, by the positional analysis for AC and GT dinucleotides, while the matrix of scores suggested the position 6 as clear preference for both these dinucleotides (not shown, see also Fig. 3). These elements, thus, are added to the initial matrix of positions, and the 2<sup>nd</sup> approximation matrix of scores calculated. Now all positions are confirmed, and final matrix of bendability can be calculated as relative deviations from average scores for every line (dinucleotide) separately. The maximal values of the deviation can be considered as the measure of the preferences – affinity of a given dinucleotide to the position. The matrix of bendability for nucleosomes of Chromosome 1 is presented in Figure 3 in form of affinity functions for all 16 dinucleotides. As the figure demonstrates, the most sensitive to the position within the period are dinucleotides CG and AT. In the positions 1 and 6, respectively, they appear about 6.5 times more often than their average occurrence in this 10-base interval. Least sensitive to the positions are dinucleotides CA, GC, TG, and TA (affinities 2.0-2.2). AA and TT dinucleotides are rather modestly selective (affinity ~3.6). It is worth noting, however, that the AA and TT dinucleotides, though of moderate positional affinity, are primary contributors to the overall periodical pattern in terms of actual occurrences, due to their high share in the dinucleotide composition of the *C. elegans* DNA. The CG dinucleotides, on the other hand, are rather rare. However, when CG is present, it has a high tendency to be found in appropriate position (position 1) within the hidden 10.4 base repeats.



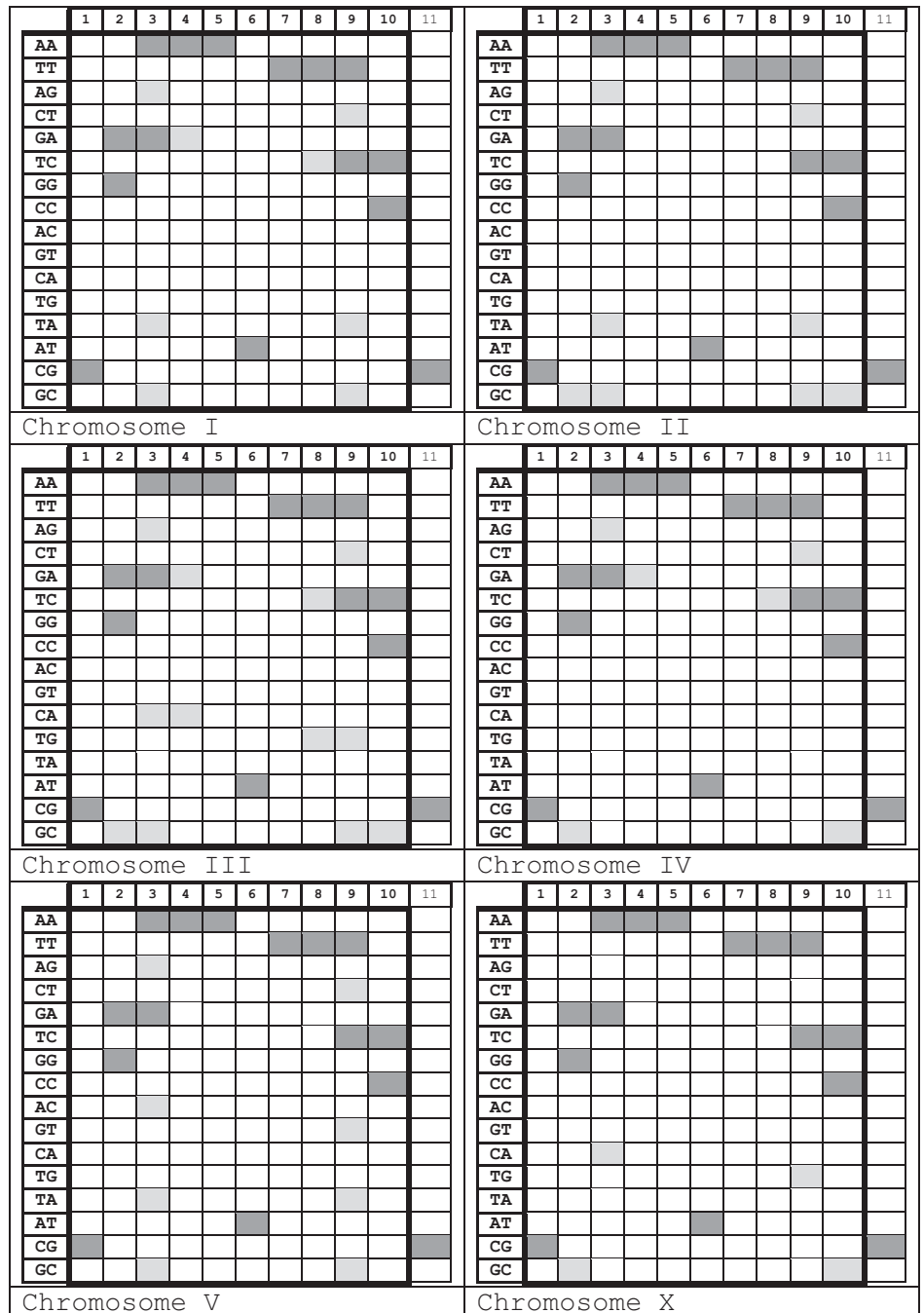
**Figure 3:** Matrix of bendability presented in form of relative scores of respective dinucleotides along the 10-base repeat. Position 11 (beginning of the next 10-base repeat) is added. The self-complementary dinucleotides CG and AT occupy symmetry positions, 1 (and 11) and 6, respectively. The amplitudes shown correspond to affinities of respective dinucleotides to various positions within the pattern (actual scores divided by their average in 10 positions).

Very similar positional matrices are obtained for the motifs AA9AA and others (not shown). To satisfy the periodicity of 10.4 bases (13) for derivation of the full length nucleosome DNA bendability matrix (e.g., 16×125) the matrices 16×10 should be

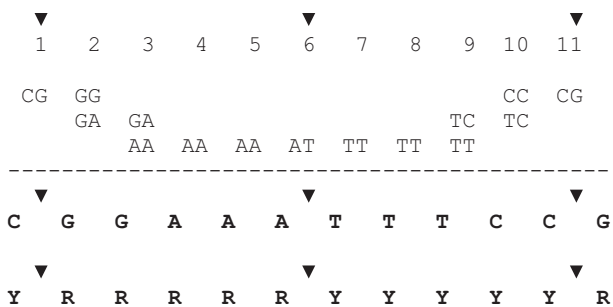
transformed in such a way that the positions within the period 10 would correspond to (non-integer) positions within the period 10.4 (work in progress).

*Positional and Bendability Matrices of other Chromosomes of C. elegans*

The calculations described in the previous section have been applied to the 146 base core sequences of other chromosomes of *C. elegans*, from the same database (12). In the Figure 4 six optimal position matrices (as in Fig. 2) are displayed together, revealing their close resemblance to one another. The matrix elements that are common to all of them are shaded dark. All the common positional preferences can be described by one-line consensus CGGAAATTTCCG, the same for all six chromosomes, or in more general form (8) YRRRRYYYYYR (Fig. 5). The lack of exact equality of the matrices can be, perhaps, explained by small differences in dinucleotide composition of the nucleosome DNA from different chromosomes. The obvious adherence of the matrices to a common pattern is a good



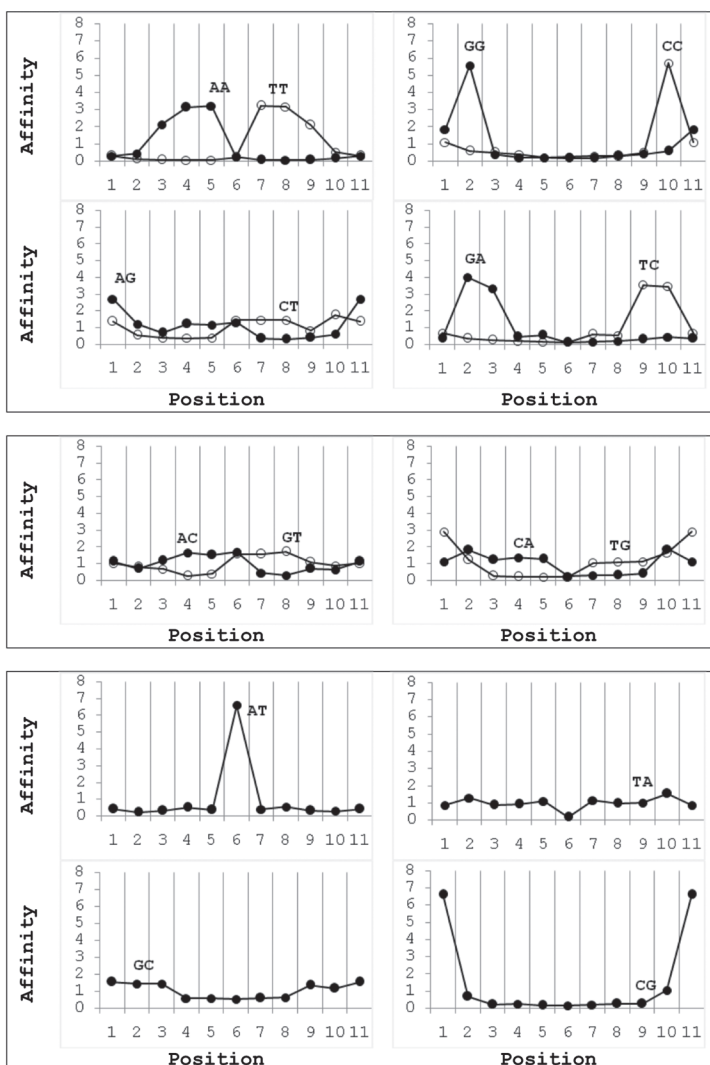
**Figure 4:** Matrices of optimal positions within the 10.4 base repeat for chromosomes I-V and X of *C. elegans*. Symmetrical elements of the individual matrices are shadowed light. Elements common for all 6 matrices are dark shadowed.



**Figure 5:** Compact presentation of the consensus positions of highest positional affinity dinucleotides within the 10 base period. The symmetry axes are shown by triangles.

reason to consider one combined matrix of bendability representing all *C. elegans* nucleosomes, shown in the Figure 6.

The highest selectivity to the positions within the period is displayed by CG, AT, GG/CC, GA/TC, and AA/TT dinucleotides. In a recent study (14) the CG dinucleotide is considered, indeed, as highly selective. Supporting experimental data (15 and references therein) indicate that the (non-methylated) CG dinucleotides favor the minor grooves in contact with the octamer surface. That puts AT dinucleotides in the outward position, while purines of the RRRRR runs, essentially, are oriented towards the histones. Such orientation of the RR·YY tetranucleotides is suggested by weaker stacking interaction between the pyrimidines, making the outwardly positioned YY stacks more deformable (*e.g.*, 16). The TA dinucleotides are the least selective. In the Chromosome 1 they are preferentially positioned at 4 or 6 bases from one another (Fig. 2). In accordance with the se-



**Figure 6:** Combined matrix of bendability of the nucleosome DNA of *C. elegans*, calculated for segments from all nucleosome core sequences (six chromosomes) that match the common positional matrix (dark shadowed elements in Fig. 4). For details see the legend to Figure 3.

quence inclusion effect (see below) the TA dinucleotides often follow TTT motif and precede AAA motif of the 10.4 base repeat. This would explain the apparent weak ~5 base periodicity of TA elements observed earlier (8). The central location of the TA element in the TTTYTARAAA pattern derived in that earlier paper by distance analysis, is inconsistent with the results of our current study where more advanced approach (regeneration of the pattern from its parts) is used. The TTT and AAA positions, however, are the same as in (8).

*The Bendability is not the Sole Reason for Positional Preferences of the Dinucleotides*

There are several reasons for a given dinucleotide to occupy specific position within the hidden 10.4 base repeat:

- I. Physical preference. Every dinucleotide stack has its individual deformational preferences towards one or another direction. To be in accord with the direction of DNA bending in the nucleosome, the dinucleotide stack would take appropriate orientation, *i.e.*, position within the 10.4 base repeat, wherever other messages (codes) overlapping in the same nucleosome DNA sequence (17) would allow it.
- II. Sequence linkage (inclusion effect). Every strong dinucleotide AB, well positioned within the repeat, is accompanied by neighboring dinucleotides NA and BN. For example, strong element TT would be flanked by elements AT, CT, and GT immediately upstream and by elements TA, TC, and TG – downstream, irrespective of their physical affinity to these positions.
- III. Exclusion effect. If several different dinucleotide elements have strong affinities to specific positions within the repeat, then other less committed elements would occupy vacant positions, by positional exclusion, if not excluded by immediate neighbors.
- IV. Some dinucleotides in some genomes or parts thereof are more frequent than others. An obvious consequence would be higher occurrence of these dinucleotides at their respective physically preferred positions, at the expense of other dinucleotides that otherwise would claim the same positions.
- V. Simultaneous presence of several different codes in the nucleosome DNA sequence inevitably causes non-optimal placement of some dinucleotides that have to fulfill other important sequence functions that would suggest the dinucleotide positioning inconsistent with the chromatin code.
- VI. Finally, due to the same reason, some of the dinucleotide positions may be exaggerated. For example, the sequences encoding aliphatic alpha-helices, have the periodicity 10.5 bases (18), indistinguishable from the nucleosome 10.4 base periodicity within the span of the encoded helix (30-40 bases). Any pattern characteristic of these DNA segments would admix to the bendability pattern and distort it.

All six factors combine in an intricate pattern, which is influenced by the overall dinucleotide composition of DNA and is, thus, species-specific (6, 19). Only due to degeneracy of various codes overlapping on the same sequence (17) the respective encoded functions are simultaneously implemented. The nucleosome DNA sequences carry messages responsible for protein coding (1, 2), gene splicing (19, 20), and formation of alpha-helices in proteins (13, 18). The chromatin code is one

of the most degenerate, *i.e.*, only small proportion of the dinucleotides present in any given nucleosome DNA sequence are actually found in their optimal positions (4). This is, perhaps, justified by necessity of unfolding the nucleosomes during replication, transcription, and remodeling. If all the nucleosomes would be very strong (large proportion of dinucleotide stacks optimally oriented) their unfolding would be energetically expensive for the cell. On the other hand, marginal stability of natural nucleosomes and weakness of the sequence positioning signal make the problem of extraction and characterization of the signal so difficult.

#### *Comparison with Other Patterns Suggested in Literature*

One of the earliest nucleosome phasing patterns has been introduced by Zhurkin (21): YRxxxRYxxxYR, with emphasis to YR and RY elements. The xxRRRxxYYYxx pattern (3, 7), on the other hand, is based on RR and YY elements. Both patterns are good complement to each other and perfect fit to the recently developed general YRRRRRYYYYYR pattern (8 and this work).

The positional preferences suggested in (11) significantly differ from the pattern derived in this work. Segal *et al.* place AA, TT, and TA dinucleotides in one common location within the period, while GC dinucleotides are placed 5 bases away. In our pattern there is no clear positional preference neither for TA nor for GC elements. That is to say, actually, these elements do not contribute to the nucleosome pattern. The AA and TT elements (and central AT), on the other hand, span together 7 positions of 10 (see Figs. 4 and 5) and, thus, can not be viewed as all belonging to the same phase.

#### **References and Footnotes**

1. E. N. Trifonov, J. L. Sussman. *Proc Natl Acad Sci USA* 77, 3816-3820 (1980).
2. E. N. Trifonov. *Nucl Acids Res* 8, 4041-4053 (1980).
3. G. Mengeritsky, E. N. Trifonov. *Nucl Acids Res* 11, 3833-3851 (1983).
4. I. Ioshikhes, A. Bolshoy, K. Derenshteyn, M. Borodovsky, E. N. Trifonov. *J Molec Biol* 262, 129-139 (1996).
5. A. B. Cohanin, Y. Kashi, E. N. Trifonov. *J Biomol Str Dyn* 22, 687-694 (2005).
6. S. B. Kogan, M. Kato, R. Kiyama, E. N. Trifonov. *J Biomol Str Dyn* 24, 43-48 (2006).
7. F. Salih, B. Salih, E. N. Trifonov. *J Biomol Str Dyn* 24, 489-493 (2007).
8. F. Salih, B. Salih, E. N. Trifonov. *J Biomol Str Dyn* 26, 273-282, (2008).
9. S. C. Satchwell, H. R. Drew, A. A. Travers. *J Mol Biol* 191, 659-675 (1986).
10. J. Widom. *J Mol Biol* 259, 579-588 (1996).
11. E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I.K. Moore, J.Z. Wang, J. Widom. *Nature*, 442, 772-778 (2006).
12. S. M. Johnson, F. J. Tan, H. L. McCullough, D. P. Riordan, A. Z. Fire. *Genome Research* 16, 1505-1516 (2006).
13. A. B. Cohanin, Y. Kashi, E. N. Trifonov. *J Biomol Str Dyn* 23, 559-566 (2006).
14. F. Salih, B. Salih, S. Kogan, E. N. Trifonov. *J Biomol Str Dyn* 26, 9-16 (2008).
15. S. Pennings, J. Allan, C. S. Davey. *Briefings Funct Genomics Proteomics* 3, 351-361 (2005).
16. E. N. Trifonov. *CRC Crit Rev Biochemistry* 19, 89-106 (1985).
17. E. N. Trifonov. *Bull Math Biol* 51, 417-432 (1989).
18. V. B. Zhurkin. *Nucl Acids Res* 9, 1963-1971, (1981).
19. S. Kogan, E. N. Trifonov. *Gene* 352, 57-62 (2005).
20. D. A. Denisov, E. S. Shpigelman, E. N. Trifonov. *Gene* 205, 145-149 (1997).
21. V. B. Zhurkin. *FEBS Letters* 158, 293-297 (1983).

*Date Received: September 4, 2008*

**Communicated by the Editor Ramaswamy H. Sarma**

