

**Supplementary Material for:**

**Structure-based Analysis of DNA Sequence Patterns  
Guiding Nucleosome Positioning *in vitro***

Feng Cui and Victor B. Zhurkin \*

Laboratory of Cell Biology, National Cancer Institute, NIH, Bethesda, MD 20892, USA

\*Corresponding author

**Supplementary Figures**

Supplementary **Figure S1**: The Roll values for nucleosomal DNA *versus* base-pair step positions.

Supplementary **Figure S2**: Profiles of the positioning scores combining individual sequence patterns.

Supplementary **Figure S3**: Positioning score profiles combining sequence patterns with the weights optimized for the two positions in Group III.

Supplementary **Figure S4**: Positioning score profiles for the two halves of the oriented nucleosome sequences with the optimized weights for Groups I and II combined.

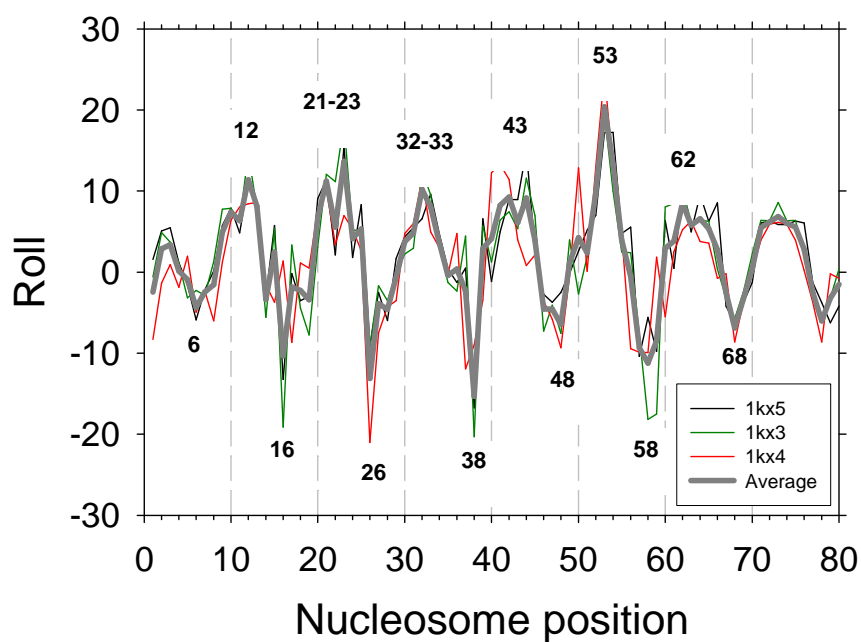
Supplementary **Figure S5**: The positioning score profiles for the 147-bp long  $\alpha$ -satellite sequence with the weights optimized for Group I sequences (Set 3-I), for Group II (Set 3-II), and for Groups I and II combined (Set 3; see Table SIII for the weight values).

**Supplementary Tables**

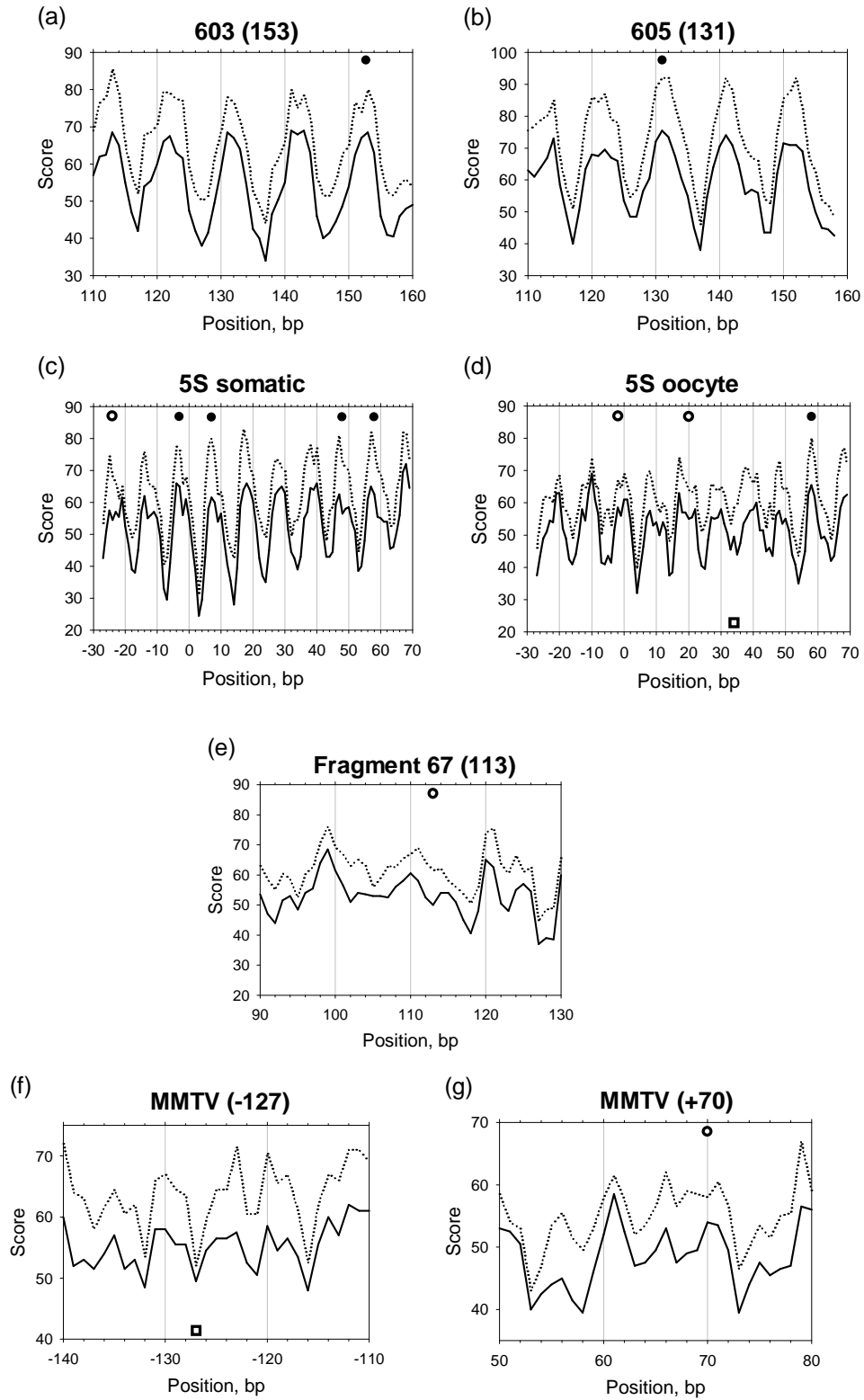
Supplementary **Table S1**: Nucleosome positioning DNA sequences.

Supplementary **Table S2**: Coefficients (counts) for sequence motifs.

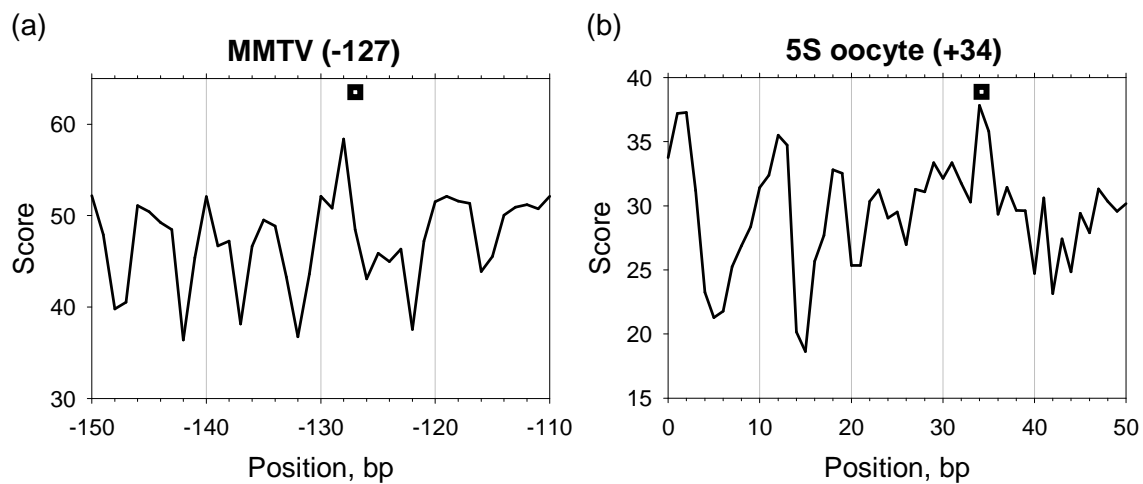
Supplementary **Table S3**: Weights of the 28 sequence patterns used in this study.



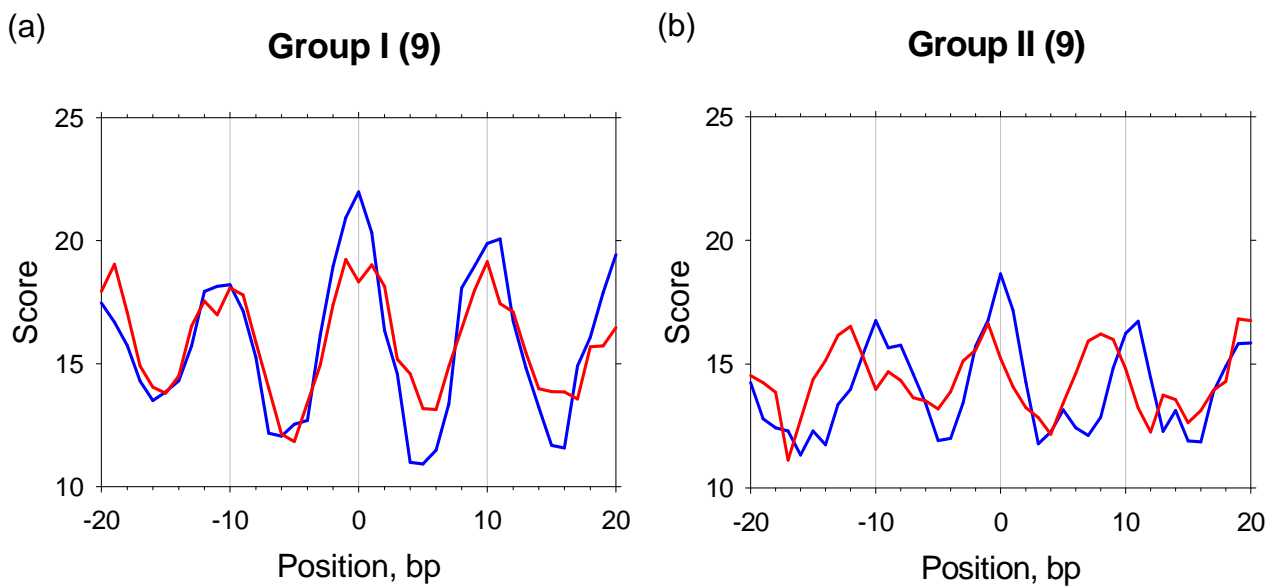
**Figure S1:** The Roll values for nucleosomal DNA *versus* base-pair step positions (54). The core DNA fragments in the three nucleosomal structures, 1KX5 (147 bp, black), 1KX4 (146 bp, red) and 1KX3 (146 bp, green) were used to calculate the Roll angles with CompDNA/3DNA software (26, 59). The values obtained for the three structures were averaged for each base-pair step and shown as a thick line. The peak and ‘dip’ positions for the averaged profile are indicated; these positions define the centers of the major and minor sites presented in Figure 2.



**Figure S2:** Profiles of the positioning scores combining individual sequence patterns. Notations are the same as in Figure 4.

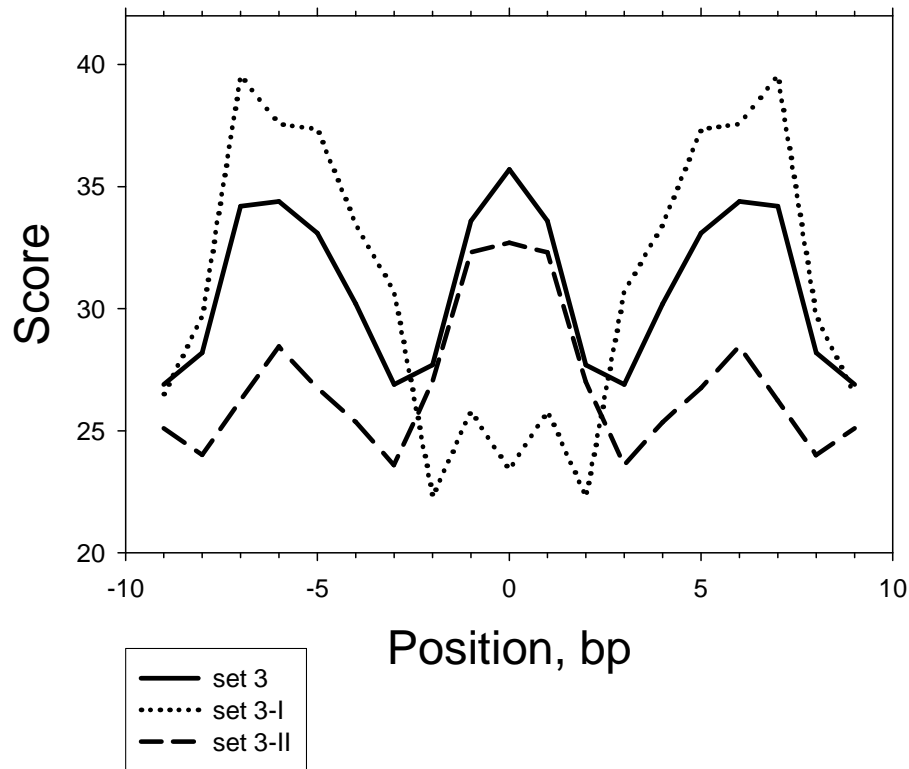


**Figure S3:** Positioning score profiles combining sequence patterns with the weights optimized for the two positions in Group III (see Table SIII for the weight values). Squares represent the dyad positions given in Table I.



**Figure S4.** Positioning score profiles for the two halves of the oriented nucleosome sequences with the optimized weights for Groups I and II combined (Table II and Set 3 in Table SIII). The extended nucleosomal sequences in Groups I (a) and II (b) are oriented from the ‘strong’ to ‘weak’ direction, with the experimental dyads placed at position 0. For each position of the 147-bp fragment, the scores for the ‘strong’ half (blue) and the ‘weak’ half (red) are averaged across the Groups I and II.

## NCP-147



**Figure S5.** The positioning score profiles for the 147-bp long  $\alpha$ -satellite sequence (54) with the weights optimized for Group I sequences (Set 3-I), for Group II (Set 3-II), and for Groups I and II combined (Set 3; see Table SIII for the weight values). The nucleosomal template is shortened by removing 9 bp at both ends. (The sites SHL  $\pm 6.5$  are removed from the minor site list (Figure 2); all the major sites are kept intact.) The experimental dyad is denoted as position 0.



Fragment\_67      1            113    CACAGGACCCAGCTATGACCATGATTACGAATTCCGGCAAAAACCGGCAAAAACGCGGAAAAAGCCGCAAAAACCTCGAGGGTACCCGGGGATCCTCTAGAGTC  
TAGATCTGT<sup>a</sup>GACTGCAGTCGACCTGCAGGCCAAAAACGCGCAAAAACCCGCAAAAAGGCCGAAAAATGATCACAAAGCTTGGCACTGGCCGTCGTTTACAAC  
GTCGTGACTGGGACCCCC

<sup>a</sup> Position of the 1<sup>st</sup> nucleotide of the sequence.

<sup>b</sup> Position of the nucleosome dyad determined experimentally.

<sup>c</sup> The nucleosome dyads are highlighted in red.

<sup>d</sup> The 5S sea urchin sequence (180 bp) contains two parts: -91 to -1 and +1 to +89, with no nucleotide at position 0.



**Table SII**  
Coefficients (counts) for sequence motifs.

pattern	motif	minor-groove sites		Major-groove sites	
		non-critical SHL $\pm 6.5, \pm 2.5, \pm 0.5$	critical SHL $\pm 5.5, \pm 4.5, \pm 3.5, \pm 1.5$	non-critical SHL $\pm 6, \pm 5, \pm 4, \pm 3, \pm 1$	critical SHL $\pm 2$
YR	TA	1	2	0.5	1
	CA	1	2	1	2
	CG	0.5	1	1	2
WW	AA, AT, TA	1	2	0	0
WWW	AAA, AAT, TAA, TAT	1	2	0	0
SS	GG, GC, CG	0	0	1	2
SSS	GGG, GGC, CGG, CGC	0	0	1	2
GC	GC	if no YA:TR, then 1	if no YA:TR, then 2	0	0
YTAR	CTAA, CTAG, TTAA, TTAG	1	2	0	0
YCAR	CCAA, CCAG, TCAA, TCAG	1	2	0	0
YCGR	CCGA, CCGG, TCGA, TCGG	0.5	1	0	0
RTAY	ATAC, GTAC, ATAT, GTAT	0	0	0.5	1
RCAY	ACAC, GCAC, ACAT, GCAT	0	0	1	2
RCGY	ACGC, GCGC, ACGT, GCGT	0	0	1	2

**Table SIII**  
Weights of the 28 sequence patterns used in this study.

Pattern (SHL)	Set 1	Set 2	Set 3	Set 3-I	Set 3-II	Set 3-III	
			Groups I & II	Group I	Group II	MMTV (-127)	5S oocyte (+34)
minor YR ( $\pm 5.5, \pm 3.5$ )	2	2	2	0.5	2	0.9	0
minor YR ( $\pm 4.5$ )	2	2	1.5	2	0.1	1	1.5
minor YR ( $\pm 1.5$ )	2	2	2	2	0.8	0	0
minor YR (other)	1	1	0.1	0	0.7	1.8	0
minor YYRR ( $\pm 5.5, \pm 3.5$ )	0	2	2	0.5	2	0.9	0
minor YYRR ( $\pm 4.5$ )	0	2	1.5	2	0.1	1	1.5
minor YYRR ( $\pm 1.5$ )	0	2	2	2	0.8	0	0
minor YYRR (other)	0	1	0.1	0	0.7	1.8	0
minor WW ( $\pm 5.5, \pm 3.5$ )	2	2	0	0.1	0	1	0.1
minor WW ( $\pm 4.5$ )	2	2	0.8	0.7	0.5	0	1.6
minor WW ( $\pm 1.5$ )	2	2	1.5	2	0.2	1.8	1.4
minor WW (other)	1	1	0.5	0.5	0.5	1.3	0.4
minor WWW ( $\pm 5.5, \pm 3.5$ )	2	2	0	0.1	0	1	0.1
minor WWW ( $\pm 4.5$ )	2	2	0.8	0.7	0.5	0	1.6
minor WWW ( $\pm 1.5$ )	2	2	1.5	2	0.2	1.8	1.4
minor WWW (other)	1	1	0.5	0.5	0.5	1.3	0.4
minor GC ( $\pm 5.5, \pm 3.5$ )	0	2	0.8	0	0.6	0.5	0.9
minor GC ( $\pm 4.5$ )	0	2	1.9	1.7	1.2	0.3	0.3
minor GC ( $\pm 1.5$ )	0	2	0.4	0.2	0	0.2	0.9
minor GC (other)	0	1	1.6	0.2	1.9	0.3	0.9
Major YR ( $\pm 2$ )	2	2	0	2	0	0.7	2
Major YR (other)	1	1	0	0	0.6	2	2
Major RYRY ( $\pm 2$ )	0	2	0.2	2	0	0.5	2
Major RYRY (other)	0	1	2	0.3	1.9	1.4	0.4
Major SS ( $\pm 2$ )	2	2	1.2	1.8	0.8	2	1.6
Major SS (other)	1	1	0.1	0.6	0.2	0	0
Major SSS ( $\pm 2$ )	2	2	1.2	1.8	0.8	2	1.6
Major SSS (other)	1	1	0.1	0.6	0.2	0	0

Pattern (SHL): The ‘critical’ positions of the minor and major-groove sites are shown in parentheses (see Figure 2 for details).

Sets 1 and 2: Pre-determined weights used in the ‘canonical’ and ‘alternative’ sequence patterns (see Figures 4 and S2 for details).

Set 3: Optimized weights of the sequence patterns combined in 18 independent subsets (see Table II).