

Comparative Sequence Analysis of the p53 Response Elements Associated with Apoptosis and Cell Cycle Arrest

Feng Cui, Michael Sirotin and Victor B. Zhurkin

Laboratory of Cell Biology, Center for Cancer Research,
National Cancer Institute,
National Institutes of Health, Bethesda, MD 20892

Correspondence should be addressed to: cuif@mail.nih.gov or zhurkin@helix.nih.gov

Abstract

The p53 protein exhibits high binding affinity to its response elements (RE) associated with cell cycle arrest (CCA-sites), but relatively low affinity to the RE associated with apoptosis (Apo-sites). To examine whether the long-range genomic environment of the RE is relevant to the differential p53-DNA binding, we analyzed 22 experimentally validated RE including 15 Apo-sites and 7 CCA-sites. The results show that 90% of the Apo-sites are clustered around the transcription start sites (TSS) of the target genes and located within high GC regions (~65% GC), while 100% of the CCA-sites are scattered further away from the corresponding TSS and located within moderate GC regions (~45% GC). Importantly, putative p53 binding sites that we found using position weight matrix are distributed around the TSS in a manner similar to the corresponding RE. We also observed the difference in distribution of the conformationally anomalous $A_n:T_n$ -runs and $G_n:C_n$ -runs in the vicinity of the Apo-sites and the CCA-sites: the apoptosis-related sites are characterized by a 5-fold higher occurrence of the $G_n:C_n$ -runs ($n \geq 4$) and a lower occurrence of the $A_n:T_n$ -runs. Comparing these distributions with those for the nucleosome positioning sequences, we conclude that the stable positioned nucleosomes are likely to form near the CCA-sites, but not near the Apo-sites. Thus, the different nucleosomal organization of the two sets of p53 response elements may be a key factor affecting the p53-DNA binding (increasing, in agreement with experiment, the p53 affinity to the CCA-sites and decreasing its affinity to the Apo-sites). Moreover, putative Sp1 and STAT binding sites occur with high frequency around the Apo-sites and the CCA-sites respectively, suggesting that neighboring transcription factors may also contribute to the selectivity of the p53-DNA binding to its cognate response elements.

Keywords: p53, p53-DNA binding, response element, apoptosis, cell cycle arrest

Introduction

The tumor suppressor protein p53 is a transcription factor that can induce apoptosis and cell cycle arrest, once activated (1). Intriguingly, after p53 activation, some cells turn into cell cycle arrest (2), while the others undergo apoptosis (3). Many different factors may be involved in governing the critical decision of survival versus death of a cell, which include post-translational modifications of p53 such as acetylation and phosphorylation, cellular p53 protein level, co-activator or co-repressor recruitment, and involvement of the family members like p63 and p73 (4, 5).

p53 trans-activates its target genes through binding to the p53 response elements (RE) connected to these genes (4-6). A typical p53 RE consists of two tandem decameric elements separated by a spacer 0-14 bp. Each decamer follows the consensus pattern RRRC(A/T)(T/A)GYYY, where R and Y are purines and pyrimidines, respectively (7). It is generally believed that p53 binds with high affinity to the RE associated with cell cycle arrest (CCA-sites), and with relatively low affinity to the ones associated with apoptosis (Apo-sites). This model was originally inferred from the observation how a cell responds to different cellular levels of the p53 protein: low levels result in cell cycle arrest, while high levels result in apoptosis (6). Recent data on the p53 binding affinity to its RE also demonstrated the influence of the RE themselves or their flanking regions on the p53-DNA binding and the p53-mediated apoptosis and cell cycle arrest (8-12).

The p53-DNA binding affinity has been studied for pure DNA (12, 13) and for the RE within chromatin (8-11). Earlier data on pure DNA showed that p53 binds stronger to the RE exhibiting high bendability (13). The DNA bending angles directly correlate with the p53 binding affinity: the more the DNA is bent by p53, the stronger the DNA binds p53 (13). The 3D model of the tetrameric p53-DNA complex suggested that p53 could bind to nucleosomal DNA (14, 15). Indeed, it was shown recently that p53 binds to the p21 RE within chromatin with a higher affinity than to pure DNA, indicating that nucleosomes might be involved in regulating the p53 binding affinity (8, 16). Two additional lines of evidence support this model: (i) the amount of acetylated histones H3 and H4 bound to the p21 RE is increased dramatically in response to p53 activation (17); (ii) the p21 RE reside within DNase-I resistant chromatin regions before and after p53 activation, similar to the other CCA-sites such as 14-3-3 σ , KARP-1 (18), and GADD45A (19). Interestingly, all the RE coupled with nucleosomes reported to date are among the CCA-sites.

To examine whether the genomic environment of the p53 RE is relevant to the p53-DNA differential binding, we investigated 15 well characterized Apo-sites and 7 CCA-sites from *Homo sapiens*. Our analysis showed that 90% of the Apo-sites are located very close to the TSS of the target genes (\pm 1 kb), where the positioned nucleosomes are unlikely to form. As a result, p53 tends to bind to the RE in the highly dynamic nucleosome environment, which may affect the binding affinity. In contrast, 100% of the CCA-sites are located at least 2-3 kb away from the TSS where the DNA sequence is more favorable for the nucleosome positioning. The stable well-localized nucleosomes are expected to form here, thereby altering the RE's conformation and facilitating the p53 binding.

Data Sets and Methods

Twenty-three experimentally validated p53 RE are collected from 16 genes known to be transcriptional targets of p53 associated with apoptosis and cell cycle arrest. Accordingly, the RE and genes are classified into two groups (Table 1).

To find putative p53 binding sites (BS) in genome, we are using the approach called position weight matrix (PWM), which calculates probability patterns of all the known BS for any given transcription factor (TF), and then predicts occurrences of the putative BS (20). Here, to evaluate affinity to p53 of an arbitrary DNA fragment, we used the weight matrix PWM-2005, developed by us earlier based on 37 known p53 RE (21). First, each DNA sequence is characterized by 'initial' score – the sum of the weights that each base of the sequence has in the matrix. Each score is then converted to a 'normalized' score from 0 to 100%, equivalent to its relative position between the smallest 'initial' score and the largest one. A cutoff value of 70% is used for this study, which corresponds to the distributions of the score values for the known p53 RE. As mentioned above, each p53 site is assumed to consist of two decamers separated by a spacer 0-14 bp.

Results and Discussion

PWM-2005 Scores Correlate with p53-DNA Binding Affinity. The PWM-2005 procedure is applied to 20 p53 RE whose binding affinities with p53 were measured by fluorescence anisotropy (12), see Table 2. Of the 20 RE, four contain non-canonical decameric motifs RRRDWWHYYY, where D is not a cytosine or H is not a guanine. These imperfect RE are omitted from further analysis. The PWM-2005 scores of the remaining 16 RE show high correlation with their p53-binding affinity ($r \sim 0.65$), thereby demonstrating that the PWM-2005 is a reliable algorithm for predicting putative p53 binding sites. In general, the CCA-sites showing high p53 binding affinity have high PWM-2005 scores while the Apo-sites showing low p53 binding affinity have low PWM-2005 scores (Figure 1).

Distinctive GC Background for the Apoptosis and CCA-sites. For a given DNA element, a slight change in its distance from TSS could result in a dramatic change in its GC background, depending on how close it is from the TSS. For this reason, we checked the GC content of the flanking regions for the two sets of RE (± 250 bp from RE). The results show that the Apo-sites, which are close to the TSS, have relatively high GC background (65% GC on average). In contrast, the CCA-sites, which are remote to the TSS, have moderate or low GC background (45% GC on average). The correlation between the distance from TSS and GC content is considerably high ($r \sim 0.75$) (Figures 2a-b). This is consistent with the observation on human genome: the GC content, in general, is relatively higher in the vicinity of the TSS of a gene. For conclusion, the Apo-sites and the CCA-sites are well separated in terms of distance to TSS and GC content.

Positioning of Putative p53 Binding Sites Correlate with RE Positioning. In addition to the well known Apo- and CCA-sites, the novel p53 RE regulating these genes are being identified continuously. Thus, one cannot be sure that the set of known p53 Apo- and CCA-sites is complete – the existence of certain, still undetected functional p53 sites regulating these genes, remains quite probable. Therefore, we compared positioning of the experimentally established p53 sites relative to TSS with analogous positioning of the hypothetical p53 binding sites.

To this aim, fragments of 20 kb in length (± 10 kb relative to TSS) are aligned so that the TSS are placed at the center. The occurrence of putative BS found by the PWM-2005 in a given window, say, from -3 kb to -2.5 kb from TSS, is averaged across all the fragments in each set. The results show that for the target genes associated with apoptosis, putative p53 BS are enriched near the TSS, peaking at $\pm 1\sim 2$ kb relative to TSS (Figure 3a), while for the target genes associated with cell cycle arrest, putative p53 BS are enriched within ± 2 kb to ± 4 kb relative to the TSS with a deep gap in the vicinity of the TSS (Figure 3b).

The resemblance between the known p53 RE and putative p53 BS with regard to distance from TSS implies that the distinctive distribution of the Apo-sites and the CCA-sites observed here may be biologically meaningful. The p53 protein clearly distinguishes between these two sets of RE, binding to them with differential affinity. In our opinion, this can be explained, at least to some extent, by the difference in genomic environment of the RE. For example, positioned stable nucleosomes in the vicinity of the RE can affect the p53 binding affinity. To check this hypothesis, we analyzed distribution of various sequence elements known to be critical for forming positioned stable nucleosomes. The best known among them are the long A- and G-tracts (22, 23).

A-tracts, G-tracts and Positioned Nucleosomes. Initially, the poly(dA):poly(dT) and poly(dG):poly(dC) homopolymers were shown to be unfavorable for the nucleosome formation in the chromatin reconstitution experiments (22, 23). Recent histone exchange experiments indicated, however, that the homopolymeric nucleosomes can be formed under the low salt condition (24, 25). On the other hand, the long A- and G-tracts are represented quite differently in the nucleosome positioning sequences from chicken and yeast: while the A-tracts frequently appear close to the termini of nucleosomal DNA, the long G-tracts ($G_n:C_n$ with $n \geq 4$) are very rarely observed. We interpret these data as an indication of high flexibility of the GC-rich DNA compared to mixed sequences. Indeed, conformational variability the CG:CG and GG:CC dimeric steps in the protein-DNA complexes by far exceeds variability of the AT-containing dimeric steps (26). The long G-tracts make DNA structures flexible in such a way that nucleosomes can easily slide from one position to another – therefore, these tracts are not observed in the stable positioned nucleosomes.

We collected the flanking regions of each RE (± 100 bp from RE) and 276 nucleosome positioning sequences from chicken (27) and yeast (J. Widom, personal communication), and then compared the occurrences of $A_n:T_n$ (A-tract, $n \leq 6$) and $G_n:C_n$ (G-tract, $n \leq 6$) in these sequences (Figures 4a-b). Surprisingly, the flanking regions of the Apo-sites exhibit at least 5-fold higher occurrence of the $G_n:C_n$ ($n \geq 4$) than the flanking regions of the CCA-sites and the nucleosome

positioning sequences. At the same time, the occurrence of $A_n \cdot T_n$ ($n \geq 4$) in the Apo-site flank regions is 2-fold lower than in the other sequences (Figure 4b).

Due to many unanswered questions in this field, one has to exert caution when structurally interpreting these results. We hypothesize that the long G-tracts frequently appearing in the vicinity of the apoptosis RE indicate that the local chromatin environment of these RE is different from the stable positioned nucleosomes. Instead, these GC-rich sequences provide highly dynamic, sliding-prone regions for nucleosomes, which may affect the DNA binding affinity of p53 *in vivo*.

By contrast, the flanking regions of the CCA-sites are similar to the nucleosome positioning sequences in terms of occurrences of the A- and G-tracts (Figures 4a-b). This probably indicates that stable positioned nucleosomes are likely to form in the vicinity of the CCA-sites. As was shown earlier (13-14), the wild type p53 tetramer bends DNA up to 60° , which is comparable to bending of the 20 bp long DNA fragment in a nucleosome. Therefore, energetically, p53 would favor binding to the REs within nucleosomes with little or no additional penalty for DNA bending. This energetical preference may be one of the reasons for an increased affinity of p53 for the CCA-sites (10).

There is also a stereochemical consideration that may be important: if the p53 binding site is phased in the nucleosomal DNA in such an orientation that the minor grooves of the C(A/T)(T/A)G elements are exposed outside, this would facilitate the p53 binding to its cognate DNA sequence (14, 15). In short, a stable and properly phased DNA wrapping in nucleosome fibril can accelerate the process of recognition of the target sites by p53 tetramer.

Difference in TF Frequency in the Vicinity of p53 RE. In addition to nucleosomes, neighboring TF may also contribute to the p53-DNA binding. To reveal those TF sites which are close to the p53 response elements, we apply the Genomatix software (28, 29) to the flanking sequences of the p53 RE (± 1 kb).

For each of the two sets of RE, the frequency of p53 BS appearing in all fragments of the set is calculated as a reference point (Figures 5a, 5d). Similarly, we calculated the frequency values for more than 50 various TF. The most pronounced among them are Sp1 and STAT proteins, which have been shown to play important roles in p53 binding to the Apo-sites and the CCA-sites (30-33).

For the flanking sequences of the Apo-sites, the Sp1 and STAT frequencies reach 70% and 30% respectively (Figures 5b-c). The Sp1 consensus binding sequence (ACCCCGCCCA) is highly GC-rich, which explains why the Sp1 binding sites are so frequent in the GC-rich flanking regions of the Apo-sites. Thus, Sp1 can act as a common co-activator of p53 to induce its apoptosis-associated target genes. One supporting example is that the Sp1 binding sites near the p53 RE are critical for the p53-mediated trans-activation of PUMA, an apoptosis-associated gene (Figure 5b). Deletion of the Sp1 sites abolished the trans-activation (34). On the other hand, the STAT involvement in apoptosis is independent of the DNA binding, meaning that the STAT sites close to the p53 RE are not obligatory (33). This is also consistent with our observation that the STAT sites are under-represented in the vicinity of the p53 Apo-sites (Figure 5c).

The tendency is reverse in the flanking sequences of the CCA-sites: here, these are the STAT sites that are twice as frequent as the Sp1 sites (their frequencies are 60% and 30% respectively, see Figures 5e-f). The lack of Sp1 sites is in accord with the moderate GC background in the vicinity of the CCA-sites (Figure 5e), whereas the abundance of putative STAT sites implies that this protein may be directly involved in the p53-mediated CCA-pathway regulation. For example, p21, a key protein for the CCA initiation, can be induced in two opposite ways: one way is p53-independent and STAT-dependent (35); the other is p53-dependent and STAT-independent (36). In other words, there is a direct competition between p53 and STAT in this case. Whether the STAT *versus* p53 competition for their binding sites is a common feature for other CCA-associated genes remains unclear. Further efforts along this line will help delineating distinctive CCA pathways.

Conclusion

We propose a model for the p53 differential affinity to the Apo-sites and CCA-sites, based on the different nucleosomal organization of the two sets of p53 response elements. Several experiments have shown that the p53 CCA-sites such as p21, 14-3-3 σ , KARP-1 (18), and GADD45A (19) are located within the DNase-I resistant chromatin regions, and the pattern of DNase-I hypersensitivity remains the same before and after the DNA damage. Therefore, stable positioned

nucleosomes or higher-order chromatin structures are likely to form in the vicinity of these RE. Our finding that the occurrences of the long A- and G-tracts in the flanking regions of known CCA-sites are similar to those in the nucleosome positioning sequences is consistent with this idea. In such a case, the proper orientation of the DNA nucleosomal loops would increase accessibility of the CCA-related response elements and thus would enhance the p53 binding affinity.

In contrast, the Apo-sites are mostly located close to the transcription start sites known to be “open” and nucleosome depleted (37-39). These are highly dynamic regions of chromatin, where the sliding-prone nucleosomes may occupy numerous positions. This kind of dynamic nucleosomal organization would slow down the process of recognition of the target sites by the activated p53 tetramer. Further testing of this model would provide new insights on how the apoptosis and cell cycle arrest are fine-tuned by p53.

References

1. Levine, A. J. (1997) *Cell* **88**, 323-331
2. Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K. & Vogelstein, B. (1990) *Science* **249**, 912-915.
3. Bates, S. & Vousden, K. H. (1999) *Cell. Mol. Life Sci.* **55**, 28-37.
4. Vousden, K. H. & Lu X. (2002) *Nat. Rev. Cancer* **2**, 594-604.
5. Vousden, K. H. (2000) *Cell* **103**, 691-694.
6. Chen, X., Ko, L. J., Jayaraman, L. & Prives, C. (1996) *Genes Dev.* **10**, 2438-2451.
7. El-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W. & Vogelstein, B. (1992) *Nat. Genet.* **1**, 45-49.
8. Espinosa, J. M. & Emerson, B. M. (2001) *Mol. Cell* **8**, 57-69.
9. Kaeser, M. D. & Iggo, R. D. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 95-100.
10. Qian, H., Wang, T., Naumovski, L., Charles, D. L. & Brachmann, R. K. (2002) *Oncogene* **21**, 7901-7911.
11. Szak, S. T., Mays, D. & Pietenpol, J. A. (2001) *Mol. Cell. Biol.* **21**, 3375-3386.
12. Weinberg, R. L., Veprintsev, D. B., Bycroft, M & Fersht, A. R. (2005) *J. Mol. Biol.* **348**, 589-596.
13. Nagaich, A. K., Appella, E. & Harrington, R. E. (1997) *J. Biol. Chem.* **272**, 14842-14849.
14. Nagaich, A. K., Zhurkin, V. B., Durell S. R., Jernigan, R. L., Appella, E. & Harrington, R. E. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 1875-1880.
15. Durell, S. R., Jernigan, R. L., Appella, E., Nagaich, A. K., Harrington, R. E. & Zhurkin, V. B. (1998) in *Structure, Motion, Interaction and Expression of Biological Macromolecules. Proceedings of the Tenth Conversation*, eds. Sarma, R. H. & Sarma, M. H. (Adenine, Schenectady, NY), Vol. **2**, pp. 277-295.
16. Espinosa, J. M., Verdun, R. E. & Emerson, B. M. (2003) *Mol. Cell* **12**, 1015-1027
17. Kaeser, M. D. & Iggo, R. D. (2004) *Oncogene* **23**, 4007-4013.
18. Braastad, C. D., Han, Z. & Hendrickson, E. A. (2003) *J. Biol. Chem.* **278**, 8261-8268
19. Graunke, D. M., Fornace, A. J. Jr. & Pieper, R. O. (1999) *Nucleic Acids Res.* **27**, 3881-3890.
20. Stormo, G. D. & Fields, D. S. (1998) *Trends Biochem. Sci.* **23**, 109-113.
21. Staib, F., Robles, A. I., Varticovski, L., Wang, X. W., Zeeberg, B. R., Sirotnin, M., Zhurkin, V. B., Hofseth, L. J., Hussain, S. P., Weinstein, J. N., Galle, P. R. & Harris, C. C. (2005) *Cancer Res.* **65**, 10255-10264.
22. Simpson, T. E. & Künzler, P. (1979) *Nucl. Acids Res.* **6**, 1387-1415.
23. Rhodes, D. (1979) *Nucl. Acids Res.* **6**, 1805-1816.
24. Puhl, H. L., Gudibande, S. R. & Behe, M. J. (1991) *J. Mol. Biol.* **222**, 1149-1160.
25. Jayasena, S. D. & Behe, M. J. (1989) *J. Mol. Biol.* **208**, 297-306.
26. Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M. & Zhurkin, V. B. (1998) *Proc. Natl. Acad. Sci USA* **95**, 11163-11168.
27. Satchwell, S. C., Drew, H. R. & Travers, A. A. (1986) *J. Mol. Biol.* **191**, 659-675.
28. Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M. Bayerlein, M. & Werner, T. (2005) *Bioinformatics* **21**, 2933-2942.
29. Quandt, K., Frech, K., Karas, H. & Werner, T. (1995) *Nucleic Acids Res.* **23**, 4878-4884.
30. Koutsodontis, G., Vasilaki, E., Chou, W. C., Papakosta, P. & Kardassis, D. (2005) *Biochem. J.* **389**, 443-455.
31. Koutsodontis, G., Tentis, I., Papakosta, P., Moustakas, A. & Kardassis, D. (2001) *J. Biol. Chem.* **276**, 29116-29125.
32. Townsend, P. A., Cragg, M. S., Davidson, S. M., McCormick, J., Barry, S., Lawrence, K. M., Knight, R. A., Hubank, M., Chen, P. L., Latchman, D. S. & Stephanou, A. (2005) *J. Cell Sci.* **118**, 1629-1639.

33. Townsend, P. A., Scarabelli, T. M., Davidson, S. M., Knight, R. A., Latchman, D. S. & Stephanou, A. (2004) *J. Biol. Chem.* **279**, 5811-5820.
34. Koutsodontis, G., Vasilaki, E., Chou, W. C., Papakosta, P. & Kardassis, D. (2005) *Biochem. J.* **389**, 443-455.
35. Agrawal, S., Agarwal, M., Chatterjee-Kishore, M., Stark, G. R. & Chisolm, G. M. (2002) *Mol. Cell. Biol.* **22**, 1981-1992.
36. Nelson, V., Davis, G. E. & Maxwell, S. A. (2001) *Apoptosis* **6**, 221-234.
37. Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D. & Leib, J. D. (2004) *Nat. Genet.* **36**, 900-905.
38. Bernstein, B. E., Liu, C. L., Humphrey, E. L., Perlstein, E. O. & Schreiber, S. L. (2004) *Genome Biol.* **5**, R62.
39. Yuan, G. C., Liu, Y. J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J. & Rando, O. J. (2005) *Science* **309**, 626-630.
40. Nakamura, Y. (2004) *Cancer Sci.* **95**, 7-11.

Table 1. Functional Classification of Human p53 RE and Associated Genes ^a

Function	Number of RE	Number of Genes	Gene Names
Cell Cycle Arrest	7	5	14-3-3 sigma, CCNK, GADD45A, p21, p53R2
Apoptosis	15	11	AFAP1, BAX, FAS, FDXR, IGFBP3, KILLER/DR5, P53AIP1, PIDD, PIG3, PTEN, PUMA

^a Functional classification of the genes is based on recent reviews (4, 40).

Table 2. Correlation Between PWM-2005 (21) and p53 Binding Affinity (12)

Response element	Function	ln(K _D)	PWM-2005 Score (%)
p21	CCA	1.59	84.8
p53R2	CCA	1.74	94.6
PUMA BS2	Apoptosis	1.96	81.9
GADD45	CCA	2.04	86.7
Cyclin G	CCA	2.05	88.2
Noxa	Apoptosis	2.15	89.5
14-3-3σ	CCA	2.29	77.4
p53AIP1	Apoptosis	2.40	78.4
MDM2	Other	2.51	73.3
PIDD	Apoptosis	2.69	90.0
KAI 1	Other	3.50	74.8
Bax	Apoptosis	4.29	75.3
IGF-BP3	Apoptosis	4.39	83.3
p53DINP1	Apoptosis	4.58	74.3
P2XM	Other	5.40	79.6
PUMA BS1	Apoptosis	5.56	63.5

CCA: Cell Cycle Arrest

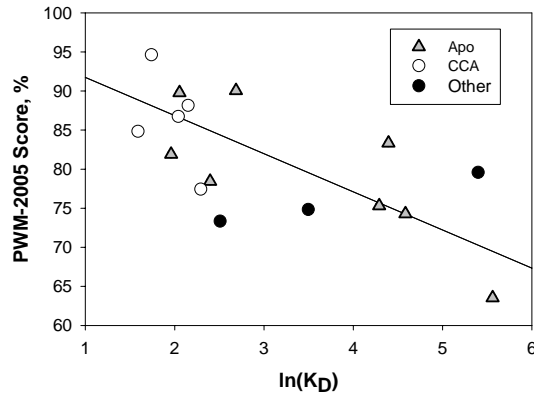


Figure 1. Correlation between PWM-2005 scores and p53-DNA binding affinity of the p53 RE (Table 1). The natural logarithms of dissociation constants (12) for 16 known p53 RE are plotted against their PWM-2005 scores (21). The correlation coefficient, $r = 0.65$. CCA refers to cell cycle arrest.

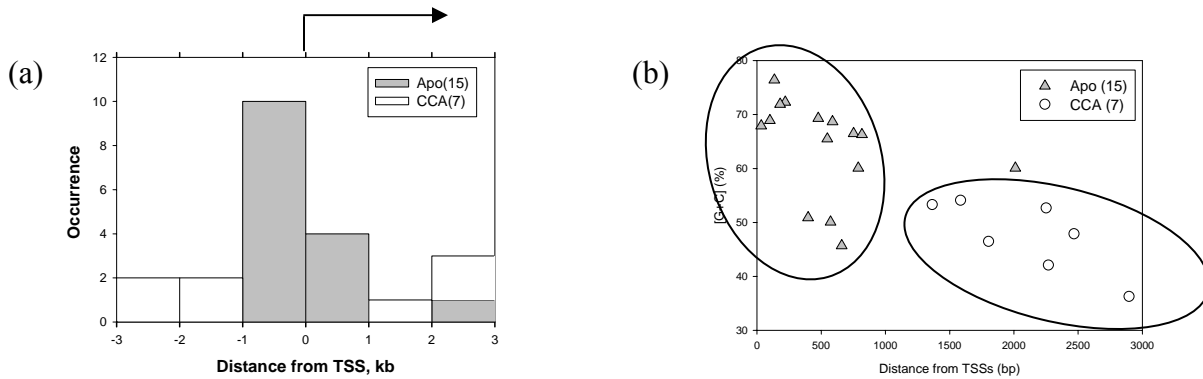


Figure 2. Distinct distribution of p53 RE associated with cell cycle arrest and apoptosis. (a) Distribution of 57 p53 RE in terms of their positions relative to TSS; (b) Correlation between distances of the RE from TSS and GC content of flanking regions ± 250 bp relative to the RE. The correlation coefficient, $r = 0.75$. The arrow in (a) indicates TSS.

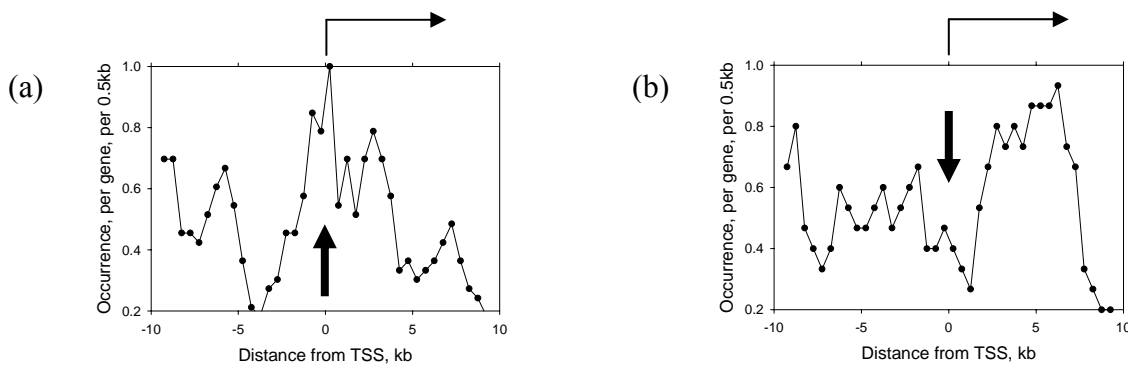


Figure 3. Occurrence of putative p53 BS for the genes associated with (a) apoptosis and (b) cell cycle arrest. The fragments are 20 kb in length (± 10 kb relative to TSS). The putative p53 BS are computed by the PWM-2005 algorithm (21). The occurrence of BS in a window is averaged across all the fragments in a given set. The window size is 500 bp and the increment is 250 bp. The value for each window is a three-point average. The thin arrows indicate TSS, while the thick arrows indicate the distinctive distribution of p53 putative BS in the vicinity of TSS.

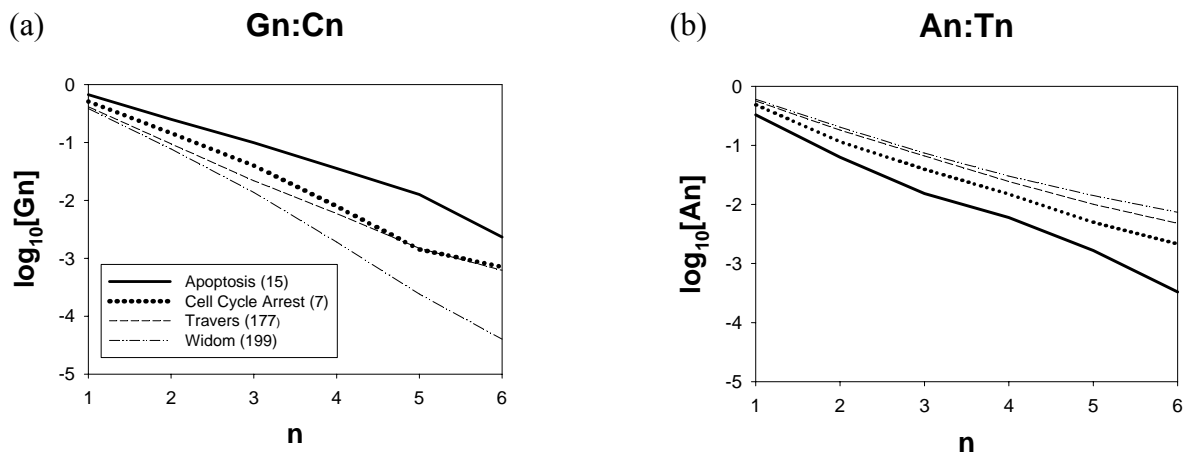


Figure 4. The G-tract and A-tract occurrences in the flanking regions of p53 RE. The fragments are 200 bp in length (+/-100 bp from RE). The numbers of G-tracts and A-tracts are calculated and averaged over all sequences in a given set. The same procedures are applied to 177 nucleosome positioning sequences from chicken (27) and 199 ones from yeast (J. Widom, personal communication).

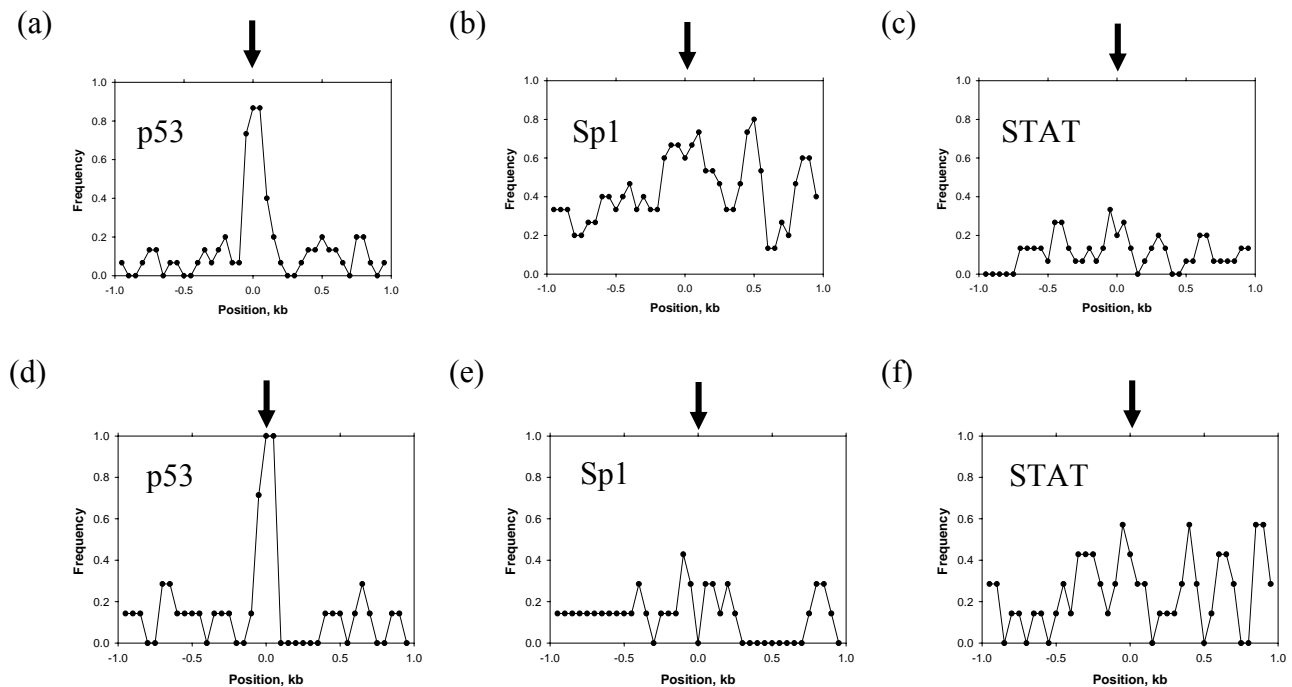


Figure 5. Transcription factor BS for p53, Sp1 and STAT proteins found by Genomatix in the flanking regions of the known p53 RE associated with apoptosis (a-c) and cell cycle arrest (d-e). The fragments are 2 kb in length (+/-1000 bp relative to the p53 RE). The p53 BS detected by Genomatix are used as a positive control (a, d). Ideally, the frequency of p53 BS should be 1.0, meaning that p53 BS is found in all sequences. Several p53 sites in (a) are not found due to a poor resemblance to the consensus sequence. The window size is 100 bp and the increment is 50 bp. The frequency refers to the fraction of fragments that bear the TF within a given window. The black arrows indicate the p53 RE positions.