Information Theory Approach to Prediction of Novel p53 Response Elements

Ilya G. Lyakhov*[1], Annangarachari Krishnamachari[2], and Thomas D. Schneider[3]

[1]Basic Research Program, SAIC-Frederick, Inc.; [1,3]Center for Cancer Research Nanobiology Program, NCI at Frederick, Frederick, Maryland; [2]Jawaharlal Nehru University, New Delhi, India; [3]NCI.

Abstract

A flexible binding site model for the tumor supressor p53 was created using a molecular information theory approach, based on 66 known p53 decameric sites (Fig 1C, Table 1). This model was used to predict novel p53 binding sites in the human genome. 16 sites were found by scanning genomic sequences around identified promoters in chromosomes 1 and 2. The sites were verified by competitive Electrophoretic Mobility Shift Assays (EMSA, Fig. 2), and more than 80% (13 out of 16 Response Elements) were confirmed (genes: H6PD, FLJ38753, LEPRE1, MGC955, RPS8, CLCA2, RDH14, DQX1, VPS24, PROM2, U5-200KD, BZW1, UGT1A6, and FLJ43374). The unconfirmed sites might be explained by non-B form DNA structures or by binding with p53-related proteins p63 or p73. We observed a p53 Response Element located between the two divergent promoters of the LEPRE1 and MGC955 genes that might allow one p53 tetramer to regulate both genes simultaneously (Fig. 3).

Introduction

p53 is a transcription factor that acts as a tumor suppressor and modulates expression of several genes related to DNA repair, cell-cycle, apoptosis and angiogenesis. p53 is involved in the G1/S and G2/M transition points of the cell-cycle. El-Deiry *et al.* proposed a p53 consensus sequence, 5′-PuPuPuC(A/T)(T/A)GPyPyPy-[0-13 bp spacer]-PuPuPuC(A/T)(T/A)GPyPyPy-3′, that consists of two decameric sequences, separated by a spacer of 0 to 13 bp, that binds a p53 tetramer. Using this consensus, many p53 target genes have been identified and characterized. A few attempts have been made to predict p53 Response Elements (REs) based on arbitrary criteria, rigid weight matrices and asymmetrical models. These approaches have many limitations and assumptions. We propose a model based on Shannon's information theory concepts.

Results and Discussion

The average information content of the flexible p53 model is 12.3 ± 3.1 bits (Fig 1C, Table 1), and 50% of the calculated distances between a p53 RE and a promoter are less than 300 bp, so we scanned the genome sequences with the flexible p53 model around the identified promoters (range -300 to +100) using the mean Ri=12 bits as a cutoff. The Ri cutoff for each decameric site in the flexible model was 5 bits. The sequences of the human chromosomes 1 and 2 were analyzed, and 16 sites were found (Table 2). All 16 sites were verified by competition EMSA (Fig. 2). The results show that the flexible p53 model works and allows us to predict p53 REs with more than 80% precision (13 out of 16). The "mut-cons" oligo (pink bar in Fig. 2) is a negative control that separates the confirmed REs from nonconfirmed. An average p53 site was found between the divergent transcription starts for genes LEPRE1 and MGC955 (Fig. 3).

Table 1. List of the known p53 Response Elements used to create the flexible model

Accession Number	Gene Name
U28935	mdm2-1
U28935	mdm2-2
U51127	Humirf5
U23824	hMSH2-1
U23824	hMSH2-2
J05614	PCNA
AF332559	PUMA
M35878	IGF-1
M35878	IGF-2
J05193	ACTA
J03206	EGFR
AF008303	TGF-beta-1
AF008303	TGF-beta-2
L24498	GADD45
U96098	Collagenase IV
U24170	WAF1-1
U24170	WAF1-2
U17193	BAX
AF257772	MCG10
AF081565	KAI1
apaf-1	APAF-1
maspin	maspin-1
maspin	maspin-2
AF029081	14-3-3
L38719	POLD1
AF274972	PIDD
U75285	Survivin
U48705	Tyrosine
X96438	PRG1
M83094	GPX
L27475	Caspase-1
AY046902	CDC25C
U18300	DDB2
J00277	C-Ha-ras
Y07755	S100A2

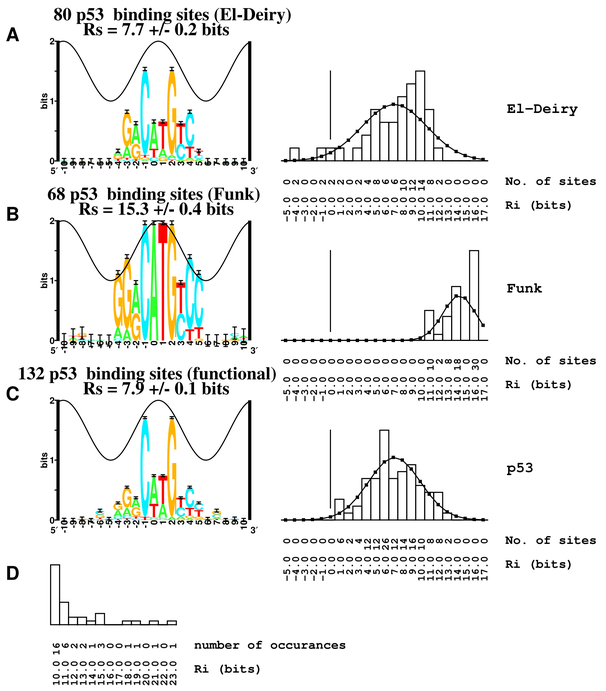

Figure 1. The sequence logos (left) and individual information (Ri) distribution histograms (right) for individual binding sites come from (A) El-Deiry *et al.*, (B) Funk *et al.*, and (C) our collection of proven natural sites. The sine wave (wavelength of 10.6 bp) peak above the logos represents the DNA major groove facing the protein, as determined from X-ray crystallography. A vertical bar indicates 0 bits of information on the distribution histograms. (D) Histogram of the gap between two decameric sites.

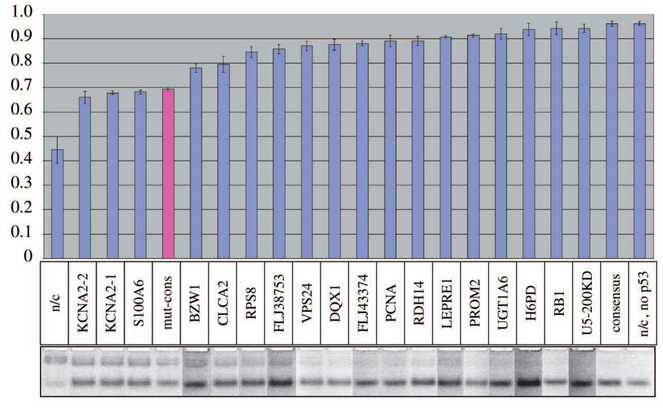

Figure 2. Competition Electrophoretic Mobility Shift Assay with oligos containing the predicted p53 binding sites. The bottom bands are unbound oligos. The top bands are complexes of oligos and p53. The table shows the name of genes that contain predicted p53 Response Elements; "n/c" means that no competitor was added; "n/c, no p53" means that no competitor and no p53 protein was added; "mut-cons" is the oligo without any p53 binding site. Bar height is the intensity of the bottom band divided by the sum of the top and bottom band intensities.

Table 2. List of genes containing the predicted p53 Response Elements

Gene name	Chromosome	Information content, bits	Description
H6PD	1	12.4	hexose-6-phosphate dehydrogenase
FLJ38753	1	12.1	hypothetical protein
LEPRE1	1	12.1	proteoglycan, potential growth suppressor
MGC955	1	12.1	hypothetical protein
RPS8	1	13.1	ribosomal protein S8
CLCA2	1	12.1	calcium-activated ion channel protein
KCNA2	1	14.2, 12.3	potassium channel protein
S100A6	1	12.2	S100 calcium binding protein A6 (calcyclin)
RDH14	2	13.5	retinol dehydrogenase
DQX1	2	13.3	DEAQ box polypeptide 1 (RNA-dependent ATPase)
VPS24	2	14.7	transmembrane protein sorting
PROM2	2	13.2	prominin 2
U5-200KD	2	12.6	U5 snRNP-specific protein, RNA helicase
BZW1	2	13.4	basic leucine zipper protein
UGT1A6	2	14.0	UDP glycosyltransferase
FLJ43374	2	12.8	hypothetical protein

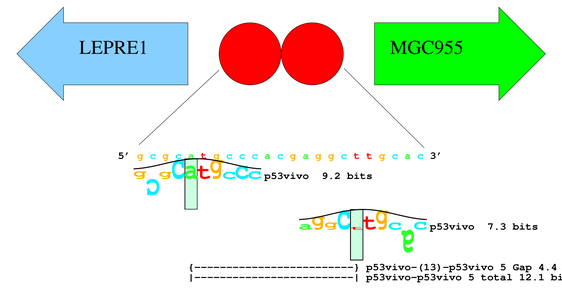

Figure 3. A p53 Response Element located between the two divergent promoters of the LEPRE1 and MGC955 genes that might allow one p53 tetramer to regulate both genes simultaneously.

Conclusions

1. A flexible Response Element model for the p53 protein was created using the molecular information theory approach.

2. 16 Response Elements were found by scanning human chromosomes 1 and 2 using the model. More than 80% of them were confirmed by EMSA. The remaining sites may be specific to p63 or p73.

3. A p53 Response Element controlling a bidirectional promoter was found.

For further information see: `http://www.ccrnp.ncifcrf.gov/~toms/`

Ilya G. Lyakhov*, SAIC-Frederick, Inc., Contract N01-CO-12400, Funded by the National Cancer Institute.
version = 1.12 of p53poster.tex 2006 Jan 05