Dynamics of p53 DNA binding sites contributes to functional selectivity of p53-driven gene expression

Jessy Safieh[1], Ariel Chazan[1], Pratik Vyas[1,2], Hanna Saleem[1], Yael Danin-Poleg[1], Dina Ron[1]*, and Tali E. Haran[1]*

[1]Faculty of Biology, Technion, Technion City, Haifa, Israel

*To whom correspondence may be addressed.
Email: bitali@technion.ac.il; Phone: 972-4-8293767
Email: dinar@technion.ac.il; Phone: 972-4-8294217
Fax: 972-4-8225153

[2] Present address: Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel

Abstract

The tumor suppressor protein p53 is situated in the midst of a complex network that is activated in response to cellular stress. An unresolved question is how p53 activates its myriad target genes in response to the severity of the stress signal and consequently coordinates the functional outcome in a timely manner. We have previously shown that DNA torsional flexibility distinguishes among p53 response elements (REs). Here we calculated the flexibility of over 200 p53 REs. By connecting functional pathways of p53-dependent genes to the calculated flexibility of their REs, we show that genes belonging to pathways activated rapidly upon stress (e.g., cell-cycle arrest, energy metabolism and innate immunity) contain REs that are significantly more flexible relative to REs of genes involved in pathways that need to be more strictly regulated or are activated later in the response to stress (e.g., intrinsic apoptosis and p53 negative regulation). The global structural properties of several p53 REs belonging to the different pathways were experimentally validated. Additionally, reporter gene expression driven by flexible p53 REs occurred at lower p53 levels and with faster rates than expression from rigid REs. Moreover, analysis of published endogenous mRNA levels of p53 target genes as a function of the flexibility of their REs support our hypothesis. Overall, we demonstrate that DNA flexibility of p53 REs contributes significantly to the timely expression of p53 target genes and thereby plays an important role in cell-faith decisions in the p53 circuity.

Keywords

Protein-DNA interactions; DNA torsional flexibility; p53 target genes; transcription factor; tumor suppression

Introduction

Gene regulation depends on the correct spatiotemporal arrangements of transcription factors and their specific DNA response elements (REs), which leads to a productive interaction between them. It is now commonly accepted that recognition of DNA by regulatory proteins occurs through forming hydrogen bonds from the protein side chains to the DNA bases (termed direct or base readout) [1], and is supplemented by hydrogen bonds to the DNA backbone (termed indirect readout) [2], and by recognition of structural and dynamical (flexibility) properties of the DNA double helix at the transcription factor REs (termed shape readout) [3].

The tumor suppressor protein p53 is part of a complex network that is activated in response to various cellular stress signals. Activated p53 functions mainly as a transcription factor, regulating the expression of numerous genes, involved not only in various cellular pathways critical for preventing cancer, but also in pathways unrelated to cancer surveillance [4, 5]. p53 is currently viewed more as the "guardian of homeostasis", rather than the "guardian of the genome" [6], because it is currently appreciated that p53 responds both to acute DNA damage and to more "routine" activities, such as energy metabolism [7-10] and embryonic development [11]. p53 in normal unstressed cells is present at low levels, caused by the association of p53 with the MDM2 protein [12]. Upon exposure of cells to stress signals, cellular levels of p53 increase due to disruption of the p53-MDM2 interaction [12].

p53 has a multi-domain modular structure [13, 14]. The N-terminus (NTD) contains a transactivation domain, the core domain contains a sequence-specific DNA binding domain (DBD), and the C-terminal domain (CTD) includes a tetramerization domain (TD) and a basic regulatory domain. The DBD of p53 binds to REs that consists of two decameric repeats of the form RRRCWWGYYY (R=A, G; W=A, T; Y=C, T), separated by 0-18 base pairs (bp) [15, 16]. Base-pairs separating p53 half-sites ("spacer sequences") are present in ~50% of all validated p53 REs [15-17]. Full-length p53, as well as shorter protein constructs that incorporate the DBD, bind to their DNA targets cooperatively as tetramers, composed of a dimer of dimers [13, 18-20].

Precise discrimination of p53 multitude target genes is crucial for p53 to trigger correct cell-fate decisions. It is now clear that p53 responses to cellular stress signals are quite heterogenous, and vary with respect to p53 levels needed for target-gene activation [21-23] and the time elapsed since p53 induction (early versus late response) [22]. However, the molecular mechanisms of p53 timely responses remain unclear. Two alternative models were proposed to explain differential gene expression upon p53 induction [24, 25]. *The selective binding model* focuses on direct binding of p53 to its REs as the fundamental step of gene regulation by p53, and inherent differences in p53 REs as an essential driving force for discrimination between p53 target genes. *The selective context model*, suggests that p53 is

not the crucial force behind target gene selectivity, but that discrimination is achieved by other events at the cellular and genomic levels. Within the selective binding model, base readout as well as indirect and shape readouts were all shown to contribute to the interaction of p53 with its target promoters [26-34]. In particular, we have previously demonstrated experimentally that DNA torsional (twist) flexibility changes significantly between p53 REs that differ in the uncontacted WW (W = A, T) motif at the center of each p53 half-site [27], or at the RRR (YYY) bases abutting the central motif from either side [34]. Moreover, we have shown that when p53 levels are low, the torsional flexibility of p53 REs positively correlates with transactivation levels, whereas when p53 levels are highly induced, transactivation levels positively correlate with p53 binding affinity [28].

The energy parameter deformability (V(B) in units of °^{3}Å^3) is a quantitative measure for the overall flexibility of B-DNA base-pair steps, that is their dynamic ability to deform (flex) under an external force, such as that arising from the binding of protein molecules [35, 36]. We have previously determined that this rigorous measure of helical flexibility, that is dependent on all DNA helical degrees of freedom, has a strong and significant correlation with the torsional (twist) flexibility of p53 natural and consensus-like REs measured using cyclization kinetics experiments [34], and can thus be used to estimate the flexibility properties of p53 REs.

Here we demonstrate that upon classifying p53 target genes by the flexibility properties of their REs, we obtain a range of REs' flexibilities – from very flexible to very rigid REs. Moreover, we show that genes that are expressed immediately upon p53 induction (when p53 protein levels are low) have flexible REs. On the other hand, we find that genes that are activated only later in the response to stress (when p53 levels are high) harbor very rigid REs. Using combined structural and functional analyses, in vitro and in cells, we establish here that DNA torsional flexibility of p53 REs has an important role in singling out the appropriate target sites for timely activation among the numerous p53-dependent promoters.

Results and Discussion

Classification of p53 REs by the functional categories of the nearby gene

For this classification we analyzed a set of 235 genes, previously studied by us [34], that abide by the following criterions: (a) the genes are directly activated by p53; (b) the REs of these genes lack spacer sequences between half sites; (c) their REs are located within the promoter of the first intron/exon of the gene. The specific function of each gene, the biological process in which it participates, and the consequent cellular outcome are summarized in Table S1. For statistical analysis of possible relationships between functional outcome and DNA flexibility these genes were organized into supergroups based on the

cellular outcome of their action (Table 1). For example, cell migration involves cytoskeleton rearrangements, ECM degradation, cell shape changes etc. Moreover, these same processes are also part of axon guidance and wound healing. We grouped all these biological processes into a supergroup termed “Cytoskeleton/cell migration”. This functional categorization was applicable to 208 out of the 211 genes for which a p53-dependent function is currently known. The remaining three genes (PRDM1, PTP4A1, XPC), have two main functions one of which is p53 negative regulation that is common to all three genes [37-39] (Table S1 and Table S2). In this p53-centric study, we chose to assign the above three genes to the p53-negative-regulation group.

The set of 211 p53 target genes analyzed here harbor 228 20-bp binding sites. This is due to the observation that p53 target sites frequently contain more than two decameric repeats, and instead contain clusters of decamers, either abutting or distantly spaced. Cluster sites are known to confer additional binding affinity and increase transactivation levels from the nearby gene [15, 40]. In order not to bias the analyses of the relationship between DNA flexibility and functional outcome from p53-dependent gene activation, by multiple REs of the same p53 target gene, we took only one 20-bp RE per gene for structural and sequence analyses. This was done both when the cluster sites were separated by a long intervening region (13 sites, Table S1), and when the sites were composed of more than two abutting decamers (seven sites, Table S1). In both cases we chose for analyses the more flexible full-site RE, based on the assumption that the more flexible RE is the primary and dominant RE of the cluster [34].

Functional outcome categories of p53 activated genes have REs with a distinct structural signature

We calculated the DNA flexibility of REs belonging to the assembled set of the 211 p53-dependent genes, using the energy parameter deformability V(B) [34, 36]. Next, we analyzed the relationship between the calculated REs deformability and functional outcome groups that contain at least five genes each, to minimize noise from small groups (total 18 groups, 187 genes). The variation in REs deformability between functional-outcome groups shows significant differences in their mean deformability values (F = 19.71, P = <0.0001, Figure 1, Table S3). The results show that p53 target genes that have REs with deformability values that are significantly above the mean are those that are needed for establishing a fast and timely p53-dependent response to cellular needs or acute stress: early DNA damage response steps, innate immunity, development, and energy metabolism (colored blue in Figure 1 and termed collectively as the “flexible REs category”). In contrast, p53 target genes that have REs with deformability values significantly below the mean belong to target genes that need to be strictly regulated, or that contribute to a later stage in

the response to stress: intrinsic apoptosis, p53 negative regulation, cellular stress response and cytoskeleton/cell motility (colored red in Figure 1 and termed collectively as the "rigid REs category"). A post-hoc test established that each group within the flexible REs category is significantly (P <0.0063) and solely different from each group within the rigid REs category but not from most of the remaining groups (Table S3). The largest difference is between development and cellular stress response (Z = 6.385, P <0.0001). Of note to p53 functions as a tumor suppressor protein, are the significant differences between early DNA damage response (DDR) steps and cellular stress response (Z = 5.03, P <0.0001), early DDR steps and intrinsic apoptosis (Z = 4.37, P =0.0019), and early DDR steps and p53 negative regulation (Z = 4.10, P = 0.0063). Concerning the groups with median flexibility, one can rationalize the mean deformability of several groups but not all. Lipid metabolism, like energy metabolism, is an early event, though it does not have as extremely flexible REs as does the energy metabolism group. The extrinsic (death receptor) apoptosis pathway has above-average mean deformability, although it is a death-related functional-outcome group, and hence expected by the our analysis to have deformability values that are below the mean. This may be related to the observations that extrinsic apoptosis is initiated by signals originating from outside the cell, such as natural-killer lymphocytes, or cytotoxic-T lymphocytes, and thus is intimately connected with the innate- and adaptive-immune systems [41], that by our analysis have above mean deformability values. A third example is the transition between cell-cycle arrest, DNA repair, and apoptosis that is known to be gradual and to depend on the accumulated level of DNA damage [42]. Moreover, the regulation of DNA repair should be stringent, since erroneous DNA repair can lead to mutations and other chromosomal defects leading to cancer development [43]. Thus, we expect DNA repair to have mean deformability between those of early DDR steps and intrinsic apoptosis, as can be observed from Figure 1.

Interestingly, there were no difference in classifications to categories, between genes that have one full site REs and those that constitute a cluster of REs having two or more full site REs. In the flexible category there are four genes with clusters of REs (BTG2, p21, COL18A1 and PANK1), in the rigid category there are also four genes with REs clusters (BAX, MDM2, PMAIP1 and RFX7), and in the interim category there are five such genes (FHL2, GPC1, LIMK2, TGFA and TNFAIP8). Finally, the analysis was unaltered when excluding all sites containing RE clusters.

Functional outcome categories of p53-activated genes have REs with a unique sequence signature

Sequence logos are a convenient way to depict base-pair preferences and information content of DNA binding sites [44-46]. The sequence logo of the whole set of 211 REs of p53-

dependent genes with known functional outcome (Figure 2A) shows a degenerate sequence pattern, similar to the pattern from previous studies [15, 16, 34, 47]. Separating this overall sequence pattern to that of the two extreme categories (Figure 2B-C), one can observe increase and decrease in nucleotide preferences in various positions. To better evaluate quantitatively the differences in sequence preferences, we calculated the information content per nucleotide position ($Iseq_i$), of the set of aligned REs sequences per category (Figure 2D, see Materials and Methods for equations). Information content is related to the thermodynamics of protein-DNA interactions because it is a measure of the discrimination, by a regulatory protein, between binding to a specific binding site versus binding to a random DNA sequence [48-51]. Comparison of the two extreme-functional-outcome categories (Figure 2D) shows that in the 5' CWW position and the 3' WWG position (CWW on the opposite strand) of the flexible REs category have higher information content (i.e., the sequence space is more constrained and thus sequences vary less within this category in these positions) relative to the rigid REs category (~47% higher, respectively). On the other hand, the category with rigid REs is more constrained at the 9th nucleotide, in the middle of the inner YYY tract, and in the 11th position. The inner YYYRRR region of the rigid REs category has 33% higher information content than that of the flexible REs category.

The observation of high information content at the two CWWG motifs of the flexible REs category is due to the almost exclusive preference for the flexible CATG sequences in both half-sites. p53 does not directly contact the WW doublet in the complex with its REs [26, 29-31, 33, 52, 53], and thus DNA recognition in this region is by shape readout. We have previously shown that CATG-containing p53 sites are more torsionally flexible than CAAG and CTAG containing sites [27]. This is due to the flexible CA and TG dinucleotides making this motif. The high flexibility of CATG-containing REs is the reason for the high kinetic stability of p53 dimers on these sites, which facilitates the swift assembly of functional p53 tetramers from DNA-bound p53 dimers [28]. The CAAG and the CTAG motifs are rigid, and hence require higher p53 levels for forming functional tetramers, as the dimeric species are not stably bound to such sites [28]. However, it is likely that this destabilization of the CAAG and CTAG motifs is partly compensated for by the stabilization of the p53-DNA interface via hydrogen bonds between Lys120 and the central G base in the opposite strand to the internal YYY tract [26, 29-31, 33, 52, 53]. This may explain the stringent requirement for a C_9 in the middle of the internal YYY tract in the rigid REs category. Thus, our results suggest that various forces optimize p53-REs interactions selectively. In the flexible REs category, high information content (binding free-energy contributions of each base) in shape-readout positions comes together with high deformability (structural property of regions within the DNA double helix). These features of the DNA double helix within the flexible REs category all contribute to a

possibility for a fast response upon stress. In the rigid REs category, a direct-readout contact (Lys120 to G base) compensates for low binding affinity in the shape-readout positions.

It has been suggested that the distinction between pro-survival and pro-apoptotic genes resides at least in part in the increased frequency of A/T versus G/C bps at positions 3, 8, 13, and 18 of p53 REs, respectively, regardless of the sequence of the CWWG core motif [54]. This suggestion was based on the authors study of DNA binding by the Lys120Arg p53 mutant, and on an analysis of p53 REs of 16 pro-survival versus 16 pro-apoptotic genes [54]. Indeed, it is known that when the sequence is $G_1G_2A_3$, the binding affinity to p53 is higher than when the sequence is $G_1G_2G_3$ or $A_1G_2G_3$, where in the last two cases Lys120 binds to both G_2 and G_3 bases, and in the first case to G_2 and the T base on the opposite strand to A_3 [26]. However, our results using 42 flexible REs pro-survival genes, and 47 rigid REs pro-death genes show that other signals are more important for the distinction between survival and death. First, results from this study, as well as several previous studies [15, 27, 55] strongly suggest that the central CWWG vary between pro-survival genes, such as cell-cycle arrest, where it is CATG, shown to be more torsionally flexible [27], to pro-death genes, such as intrinsic apoptosis, where it is CAAG, shown to be more torsionally rigid [27]. Here we add that a highly conserve C base at position 9 is important for optimal direct contact of Lys120 to the G base on the opposite strand, especially in the rigid REs category, where the contact to the central CAAG motif is sub-optimal.

Experimental validation of the flexibility of five natural p53 REs by cyclization kinetics

Cyclization kinetics of DNA minicircles is a robust method to measure the global structural and dynamical properties of DNA molecules [56, 57], and it is carried out by following the rate of ligase-catalyzed closure of DNA molecules with cohesive ends into small rings. The REs chosen for the current study belong to genes that function at different cellular pathways (Table 2), and they were chosen by their deformability values (Table 2, rightmost column). As in our previous studies on p53 binding sites [27, 34, 58], the only significant change in global-structural parameters of these sites is in the twist (torsional) flexibility of the DNA binding sites (Table 2, Figure S1). Moreover, the results of the experimentally-derived torsional flexibility values showed similar trend to the values of the calculated deformability (Table 2). Thus, rigorous experimentally-derived torsional-flexibility values, show that CCNG1, RRM2B, and p21-5' are the most flexible sites in this set, whereas PMAIP1 and TP53AIP1 are the most rigid sites. This experimentally corroborates our hypothesis that p53 REs that belong to the early DDR steps group are more flexible than those that belong to the intrinsic apoptosis pathway.

p53 target gene activation as a function of p53 levels

To further validate our hypothesis in a cellular context we carried out luciferase reporter gene assays in human non-small lung carcinoma cell line (H1299), ectopically expressing p53 along with reporter plasmids, each carrying a distinct RE. First, we ascertained that p53 ectopic expression is linearly correlated with the amount of transfected p53 expression plasmid pC53-SN3. This was done by transfecting cells with increasing amounts of p53 expression vector and immunoblotting the cell lysate (Figure S2A). Densitometry analysis of the observed p53 band, normalized to the β-actin band, confirmed the linear relationship between p53 protein level and pC53-SN3 amount (Figure S2B, R = 0.998). We then carried out reporter-gene assays for the five p53 binding sites described above at five different p53 levels (Figure 3). We analyzed transactivation measured by luciferase activity 48 h after transfection, to enable observing a significant signal also in cells transfected with the lowest p53 plasmid amount (1 ng p53 plasmid). Results show that transactivation rises as a function of p53 levels (Figure 3). One-way ANOVA, followed by mean comparisons, revealed that transactivation levels of the five genes are significantly different from each other at 1, 2, 5,10, 25ng of pC53-SN3 plasmid (Figure S3A-E, $P < 0.005$), but are similar at 75 ng of pC53-SN3 (Figure 2A, Figure S3F).

Focusing on cells expressing very low levels of p53 (1 ng p53 plasmid, Figure 3C), results show that transactivation from two flexible REs (RRM2B, and p21-5') was distinctly higher compared to transactivation from the two rigid REs (PMAIP1 and TP53AIP1, $F = 4.96$, $P = 0.0044$, Figure S3A). When 2 ng p53 plasmid were used (Figure 3C), transactivation levels from each of the three flexible REs were significantly higher than that from the two rigid REs ($F = 22.9$, $P < 0.0001$, Figure S3B). At a higher p53 plasmid level the PMAIP1 and p21-5' REs are not monotonously increasing in transactivation levels, as the other REs are. Thus, individual pair-wise differences between REs are significantly different only when comparing the two most flexible REs - CCNG1 and RRM2B, to the two rigid REs – PMAIP1 and TP53AIP1 (Figure 3B, Figure S3 C-E). The transactivation from cells transfected with all tested p53 plasmid amounts, except 75 ng, was correlated with the experimental torsional (twist) flexibility of the five studied REs (Pearson $r\ (P)$ = 0.35 (0.03), 0.74 (<0.0001), 0.64 (0.004), 0.84 (<0.0001) at 1, 5, 10, and 25 ng p53 plasmid, respectively), being most strongly at 2 ng plasmid (Pearson $r = 0.91$, $P < 0.0001$). This is consistent with our previous results [28], that at basal levels of p53 that is a strong and significant correlation between transactivation and DNA flexibility of p53 REs.

We compared known amounts of recombinant full-length p53 protein to transfected levels of p53 plasmid in H1299 cells (Figure 4). In cells harboring 1 and 2 ng p53 plasmid, we estimated that p53 protein levels are approximately 3.7×10^{-4} and 7.7×10^{-4} pmole/μg of whole-cell extract (WCE), respectively, which is similar to basal p53 levels previously found

in HCT116 cells ~5 X 10-4 pmole/microg WCE [59]. At 75 ng p53 plasmid there are 3.6 X 10^{-3} pmole/μg WCE, which is similar to endogenous p53 levels found after stress induced by camptothecin treatment in HCT116 cells 11 X 10-4 pmole/microg WCE [59]. Thus, the results ascertain that the amount of p53 plasmid used in this transiently-transfected plasmid study corresponds to known levels of endogenous p53 protein that range from basal levels to those induced by cellular stress.

p53 target gene activation as a function of time

The higher transactivation levels observed 48 h post-transfection from flexible REs, relative to rigid REs, could be due to an earlier start of transactivation, or alternatively due to faster transactivation rates. To distinguish between these possibilities we examined transactivation levels, at 25 ng p53 plasmid (to enable observation of a significant signal also at shorter times), at additional time points (12, 18, and 24 h post-transfection, Figure 3D). One-way ANOVA analysis at 12 h post-transfection revealed a significant difference between the transactivation level of the five REs ($F = 4.34$, $P = 0.008$), which remains significant at 18 h ($F = 4.36$, $P = 0.008$) and 24 h ($F = 79$, $P < 0.0001$). A post-hoc analysis indicated that at 12 h post-transfection, transactivation from CCNG1 and p21-5' was significantly different from all rigid REs (Figure S4A). At 18 h post-transfection, post-hoc analysis indicated that transactivation from RRM2B was significantly different from all rigid REs (Figure S4B). At 24 h post-transfection, transactivation from all flexible REs studied here was higher than transactivation by all rigid REs (Figure S4C). Even though p53 levels in our study are due to ectopic expression of p53, the time frame for observation of post-transfection transactivation activity are in line with studies probing p53 activation due to induction of cellular stress events [60, 61]. A positive significant correlation is observed between p53-dependent transactivation level from the five studied REs, at all studied post-transfection time points, and their experimental twist flexibility ($r = 0.50$, 0.53, and 0.77, and $P = 0.004$, 0.001 and $P < 0.0001$, for 12, 18, and 24 h, respectively).

Hafner's time-dependent RNA-seq study of endogenous mRNA levels of p53 target genes support the division to functional outcome categories based on extreme REs flexibility

To further support our results and to be able to mimic a more natural environment, where transcription is affected by multiple other factors, we analyzed the time-dependent RNA-seq data from the study by Hafner et al. [62]. Hafner et al. studied the relationship between p53 dynamics and p53-dependent gene expression following DNA damage caused by γ-irradiation at low p53 levels, which triggers a series of p53 pulses [62]. The measurements were taken every hour between 1 h and 12 h, and another measurement at 24 h. Their

results showed that the RNA-seq measurements can be clustered into five distinct clusters according to gene expression dynamics [62]. We analyzed RNA-seq measurements of p53 target genes that belong to cluster 1 (the cluster with the most enriched p53 bound genes) and for genes that appear in our set (18 of 32 genes in cluster 1). Since the fold-change of mRNA expression (relative to the expression level at time zero) changes as a function of time, we normalized fold-change values to the average fold-change of all selected genes at a particular time point. Likewise, we used normalized deformability values (Table S1). We then plotted the normalized fold-change of gene expression as a function of normalized deformability values of each specific gene. Values are ranked relative to the average value of the parameter (i.e., the lines drawn across both axes, Figure 5). From Figure 5 it can be observed that early DDR steps genes (colored blue) have above average flexible REs, and are expressed at above average level (>2 normalized fold change) at every tested time point. On the other hand, genes belonging to the extremely rigid REs category (colored with various shades of red) are all below average rigidity and are expressed at below average level. The only exceptions are the two genes from the cellular stress response group, SESN1 and SESN2, as well as MDM2 from p53 negative regulation group. Nonetheless, these genes also express rather low levels of mRNA (<2 normalized fold change, Figure 5). MDM2 is known to have a second full site, 17-bp away from the first one, and as discussed above, additional p53 REs can lead to enhanced transactivation levels [15, 40]. This could perhaps be the case also for SESN1 and SESN2, as additional p53 REs (or p53-like REs) clusters can be observed in the UCSC Genome Browser for these genes. Analysis of such small data set is problematic, yet, differences in expression levels were seen only between the flexible REs group (early DDR steps) and the rigid REs group (intrinsic apoptosis, cellular stress response, cytoskeleton, and p53 negative regulation), at all the studied time points ($P < 0.004$). All in all, analysis of the results of Hafner at al. [62] support our hypothesis that DNA flexibility is an effective indicator for the level of p53-activated gene expression at low levels of p53, and that genes that are expressed early after stress induction have flexible REs and are mostly expressed at higher levels than those that are expressed at a later stage in the response to stress. At present, no conclusions can be drawn for genes belonging to groups with intermediate flexibility REs with regard to their level of mRNA expression. Studies of a larger set of p53-dependent genes at high temporal and positional resolution will help to further substantiate these observations in natural cellular environment.

Summary and Conclusions

Orchestrated expression of p53 target genes governs cell-fate decisions. An important question in the field is how p53 discriminates between the myriad p53-dependent target genes. The importance of differential binding affinities of p53 to its REs [55], the dependence of

cell-fate decisions on p53 levels [21-23], and the elapsed time after p53 induction [22] have been widely acknowledged. However, the mechanistic basis for timely p53-dependent gene expression is currently unknown. In our previous studies, we showed that variation in torsional flexibility as a function of the RE sequence is the major structural feature that differentiates between p53 REs [27, 34]. Moreover, at high p53 levels, transactivation from a given RE, is regulated by p53 binding affinity whereas at basal p53 levels it is regulated by p53 binding kinetics, governed by the DNA flexibility of the RE [28]. These mechanistic observations support selectivity in p53 binding for understanding target-gene discrimination in the p53 system. Yet, it does not explain how p53 can coordinate its myriad target genes in an orderly fashion. Here we show that DNA flexibility is a key factor in determining p53 functional outcome. Transcription of genes required for fast response, at low p53 levels, is associated with REs characterized by high DNA flexibility. As we have shown previously [28], DNA torsional flexibility enables a stable binding of p53 dimers to p53-dependent REs also when p53 levels are low. Under these conditions, p53 tetramers can be formed sequentially (i.e., at low binding cooperativity) without the need for p53 accumulation over an extended time span. Therefore, our results, obtained in-vitro and in cells, indicate that the information for the selective binding of p53 to DNA is intrinsically encoded within the base sequences of the p53 REs. Hence, in addition to the statement that “p53 is smart” [25], the current data demonstrate that the DNA of p53 response elements is “smart” - the REs base sequences are information-rich by virtue of their structural properties. Consequently, the timely expression of genes’ activation is another fundamental code within the layers of codes that co-exist within the DNA double helix.

Our findings that REs flexibility contributes significantly to the orderly expression of p53-activated genes is in line with the observations that the presence of a single p53 full-site RE is sufficient for activating p53-dependent gene expression, without the need for cooperation from additional TFs [63, 64]. This does not preclude combinatorial control by co-activating TFs at some genomic binding sites [65], as well as effects originating from chromatin context [66, 67], or from post-transcriptional events, such as mRNA stability of p53 target genes [62].

Materials and Methods

Functional annotation of p53 target genes

We used here a previously assembled set of 235 p53 target genes that are all upregulated by p53, are validated for binding in cellular context and for p53-dependent gene expression, and reside in the promoter of the first exon/intron region [34]. This set of genes harbor 250 p53 REs that are all without spacer sequences between half sites, because p53 REs with long spacer sequences between half-sites (>9bp) can be bound by p53 in hemi-specific mode, a

binding mode that nonetheless has a comparable binding affinity and transactivation levels to that of the full-specific complex [58]. We modified the previous dataset by deleting the DDIT4 REs, because this gene was shown to be directly upregulated by RFX7 and not by p53 [68, 69] and added two genes RhoC and Scrib (Table S1). We located each gene in the Genecards suite (https://www.genecards.org/) by their Entrez ID, gene symbol, and gene name. We then looked for functional annotations of these genes in the Gene Ontology (GO) project (under the 'Biological process' and 'Molecular function' categories) through UniProtKB/Swiss-Prot (https://www.uniprot.org). GO annotations were supplemented by publications retrieved through individual literature searches for each gene product, and especially those centering on the role of these genes in the p53 system.

Cells, plasmids, and reagents

H1299 human non-small lung carcinoma cells (p53-null cells) were obtained from ATTC (CRL-5803). They were thawed and propagated in RPMI-1640 (Biological industries, Beth Haemek, Cat # 01-101-1A) with 10% Fetal Bovine Serum (Biological industries, Beth Haemek, Cat # 04-127-1A), 2 mM L-glutamine, and 1% Pen-Strep solution (Biological Industries, Beth Haemek, Cat # 31-031-1C). All experiments were performed in this medium. Cells were maintained in a humidified incubator at 37°C and 5% CO_2. Cells were routinely checked for negative mycoplasma infection.

The reporter gene vectors pCLuc Mini- TK 2, encoding Cypridina secreted luciferase under the control of a minimal promoter, and pCMV-GLuc, encoding constitutively expressed Gaussia secreted luciferase, were purchased from Thermo-Fisher. pC53-SN3 construct (where p53 expression is driven by the CMV promoter) was used for wild-type p53 expression in H1299 cells. p53 expression plasmid was kindly provided by Varda Rotter, Weizmann Institute of Science, Rehovot, Israel.

Generation of reporter plasmids containing p53 REs

All DNA sequences were synthesized by Sigma Genosys (Israel), purified by standard desalting, and are described in Table S4. The sequences corresponding to p53 REs were incorporated in pCluc Mini-TK plasmids using site-directed mutagenesis with KAPA HiFi™ DNA Polymerase kit (Gamidor Diagnostics), and transformed into competent DH5α cells, according to manufacturer recommendations. Overexpressed plasmids were extracted using QIAprep® Spin Miniprep Kit (Qiagen), according to the manufacturer's protocol. The identity of incorporated sequences was verified using Sanger sequencing.

Reporter gene assays

H1299 cells were seeded in 24-well plates at a density of $5X10^4$ cells/well and incubated for 24 h. Then cells were co-transfected with 800 ng of either pCLuc Mini-TK 2 harboring p53 RE or a corresponding empty vector along with 50 ng of pCMV-GLuc and 0 to 75 ng (0, 1, 2, 5, 10, 25, 75 ng) of p53 expression vector (pC53-SN3). The total DNA amount was adjusted using pcDNA 3.1 plasmid. Transfection was performed with jetPEI (Polyplus) reagent according to the manufacturer's recommendation. Transfection efficiency was monitored by transfecting separate wells with 0.1 μg eGFP along with Gluc and pcDNA plasmids. An aliquot of growth medium (20 µl) was taken at various time points post-transfection and luminescence was measured using Pierce™ Cypridina Luciferase Flash Assay Kit and Pierce™ Gaussia Luciferase Flash Assay Kit (both from Thermo Fisher) according to the manufacturer's protocol. Signal was detected by CLARIOstar 96 well plate reader (BMG LABTECH). Four to eight independent experiments were carried out, each containing two technical replicates.

Analysis of transactivation from p53 REs

To correct for transfection variability, all results from Cypridina Luciferase luminescence were normalized to Gaussia Luciferase luminescence. Next, the results were normalized in two steps. First, the results obtained with pCLuc Mini-TK 2 harboring p53RE ($Cluc^{+RE}$) were divided by the results obtained with the corresponding empty vector ($Cluc^{-RE}$). Second, the results obtained with p53 (1-75ng of pC53-SN3) were divided by the results obtained without p53 (0 ng of pC53-SN3). Fold increase in transactivation is thus given by the equation:

Fold increase in transactivation= $(Cluc^{+RE}/Cluc^{-RE})^{+p53}/(Cluc^{+RE}/Cluc^{-RE})^{-p53}$

p53 protein quantitation and detection

Cells were lysed in RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% Na deoxycholate, 0.1% SDS, 25 mM Tris-Cl, pH 7.5) containing 1/100 (Vol/Vol) Protease inhibitor cocktail (Sigma, Cat # P8340). An equal amount of protein from the different samples were resolved on 10% SDS PAGE, and transferred to PVDF membrane. The blots were subsequently incubated overnight at 4°C with primary antibodies against p53 (DO-7 mouse mAb, Cell Signaling, Cat # 48818) and β actin (13E5 rabbit mAb, Cell Signaling, Cat # 4970). Membranes were then reacted with a secondary antibody (Anti-mouse IgG, and Anti-rabbit IgG conjugated to Horseradish Peroxidase (both from Cell signaling, Cat # 7076 and # 7074, respectively), developed using Western Bright™ enhanced chemiluminescence reagent (Advansta), air-dried, and exposed using Fusion Pulse (Vilber) detection system, followed by quantitation using Evolution-Capt software. For p53 serial-dilution curve, full-length recombinant p53 protein was resolved on gels together with the transfection samples.

Cyclization kinetics and simulation of cyclization data

DNA constructs for cyclization kinetics were synthesized using the top library/bottom test sequence PCR scheme, described previously [27, 34, 57, 58], except that here we used as test sequences DNA sequences containing the full 20-bp long p53 sites plus 5-bp flanking sequences on either side, specific for each natural p53 site (see Table S5 for sequences), because we have previously shown that sequences flanking p53 REs modulate p53 binding and transactivation [34]. However, the global structural characteristics of the sites (Table 2, Figure S1) are those of the central 20-bp p53 sites without flanks, to concur with our previous studies [27, 34, 58]. The DNA test sequences for cyclization experiments (bottom PCR templates) were synthesized by Sigma Genosys (Israel), whereas the library DNA sequences (top PCR templates) and the fluorescein- and tetramethylrhodamine (TAMRA)-labeled oligonucleotide primers were synthesized by the Keck foundation Laboratory at Yale University. Cyclization kinetics measurements were carried out as described previously [27, 34, 58]. PCR reactions (50 µl) contained 1× PrimeSTAR buffer, 0.2 mM dNTP mix, 0.3 µM of each of the primers, 100 nM of the top and bottom templates, and 0.04 units of PrimeSTAR DNA polymerase (Takara, Japan, Cat # R010B). Ligase concentration was varied as a function of phasing length (0.08 U/µl for the in-phase 156L14 and 156L16 constructs and 1.0 U/µl for all other phasing constructs) and total length (from 0.08 U/µl for the 157 constructs, to 0.8 U/µl for the 154–155 and 158–160 constructs and 2.5 U/µl for the 150–153 constructs). We derived quantitative data on the conformational properties of the test sequences by simulating the cyclization data as previously described [27, 34, 58, 70], using the simulation program developed by Zhang and Crothers [70]. The outcome of the simulations are the bend angle, twist angle, roll and tilt flexibility and twist flexibility per DNA sequence.

Information content analysis

Information content per DNA binding site, Iseq [48, 49], was calculated separately for the whole set of REs as well as for the flexible and the rigid functional outcome categories, on the mononucleotide level, using the following equations:

(1) $Iseq\ (mono) = \sum_{l=1}^{s} Iseq_l(mono) = \sum_{l=1}^{s}\sum_{B=0}^{3} f_{lb} * \ln\frac{f_{lb}}{P_B} = \sum_{l=1}^{s}\sum_{B=0}^{3} f_{lb} * \ln(4 * f_{lb})$

Where $Iseq_l$(mono) is the mononucleotide information content per position, f_{lB} is the frequency of occurrence of base B in position l in the population of all possible binding sites, and P_B of base B is the frequency in the whole genome, which is usually taken to be 0.25 for each base. To correct for small sample size, we replaced equation (1) with the following equation, shown to be the best estimate for the sequence information when f_{lB} is unknown precisely [49]:

(2) $Iseq\ (mono) = \sum_{l=1}^{s} Iseq_l(mono) = \sum_{l=1}^{s} \sum_{B=0}^{3} \left(\frac{n_{lb}+1}{N+4}\right) * \ln\left(4 * \frac{n_{lb}+1.5}{N+4.5}\right)$

Where N is the sample size, and n_{lb} is the number of occurrences of base-pair B (B = 0, I, 2, 3) at position l. The expected variance in assigning f_{lB} values was calculated from [48]:

(3) $\sigma_{lB}^2 = \langle f_{lB}^2 \rangle - \langle f_{lB} \rangle^2 = \frac{(n_{lB}+1)(N-n_{lB}+1)}{(N+4)^2(N+5)}$

Other computational and statistical tests

The values for the deformability (V(B) in units of $°^3Å^3$) of each DNA base-pair step were calculated as previously described [34], using the values of [36]. Statistically significant differences in DNA deformability between p53-dependent functional-outcome groups, as well as analysis of the differences in transactivation level of the studied REs, were tested using one-way analysis of variance (ANOVA). We used the Dunn method for joint ranking as a post-hoc test, to determine the groups that are significantly different from each other by their deformability, since this method is adjusted for multiple non-parametric comparisons. We used student's t-test as a post-hoc test for the differences between transactivation levels. All *P* values used in this study are two-sided. *Z* (Table S3) denotes the standardized test statistic, which has an asymptotic standard normal distribution under the null hypothesis of no difference in means. In testing the strength of the relationship between transactivation assays from reporter gene assays (RGA) and deformability (V(B)), we used Pearson correlation, using all data points from the biologically independent RGA experiments. All analyzes were conducted using JMP®, Version 16 (SAS Institute Inc., Cary, NC, 2021).

CRediT authorship contribution statement:

Jessy Safieh: Methodology, Software, Validation, Formal analysis, Investigation. **Ariel Chazan:** Methodology, Validation, Investigation. **Pratik Vyas:** Investigation. **Hanna Saleem:** Resources. **Yael Danin-Poleg:** Formal analysis; Writing - Review & Editing. **Dina Ron:** Supervision, Resources, Writing - Review & Editing. **Tali E. Haran:** Conceptualization, Supervision, Funding acquisition, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Varda Rotter (Weizmann Institute of Science, Rehovot, Israel) for the p53 expression plasmid, pC53-SN3. We thank Alberto Inga (CIBIO, Trento, Italy) for stimulating discussions of the manuscript. We thank the biomedical core facility (BCF, Technion) for Sanger sequencing. This work was supported by the Israel Science Foundation (Grant #1517/14 to T.E.H.).

References

1. Seeman, N.C., Rosenberg, J.M., Rich, A., (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA*, **73**, 804-808.
2. Otwinowski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F., Sigler, P.B., (1988). Crystal structure of *trp* repressor/operator complex at atomic resolution [published erratum appears in Nature 1988 Oct 27;335(6193):837]. *Nature*, **335**, 321-329.
3. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., Mann, R.S., (2010). Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, **79**, 233-269.
4. Vousden, K.H., Prives, C., (2009). Blinded by the Light: The Growing Complexity of p53. *Cell*, **137**, 413-431.
5. Kastenhuber, E.R., Lowe, S.W., (2017). Putting p53 in Context. *Cell*, **170**, 1062-1078.
6. Aylon, Y., Oren, M., (2016). The Paradox of p53: What, How, and Why? *Cold Spring Harb. Perspect. Med.*, **6**, a026328.
7. Gottlieb, E., Vousden, K.H., (2010). p53 regulation of metabolic pathways. *Cold Spring Harb. Perspect. Biol.*, **2**, a001040.
8. Liang, Y., Liu, J., Feng, Z., (2013). The regulation of cellular metabolism by tumor suppressor p53. *Cell Biosci.*, **3**, 9.
9. Lacroix, M., Riscal, R., Arena, G., Linares, L.K., Le Cam, L., (2020). Metabolic functions of the tumor suppressor p53: Implications in normal physiology, metabolic disorders, and cancer. *Mol. Metab.*, **33**, 2-22.
10. Lahalle, A., Lacroix, M., De Blasio, C., Cisse, M.Y., Linares, L.K., Le Cam, L., (2021). The p53 Pathway and Metabolism: The Tree That Hides the Forest. *Cancers (Basel)*, **13**, 133.
11. Koifman, G., Aloni-Grinstein, R., Rotter, V., (2019). p53 balances between tissue hierarchy and anarchy. *J. Mol. Cell. Biol.*, **11**, 553-563.
12. Michael, D., Oren, M., (2003). The p53-Mdm2 module and the ubiquitin system. *Semin. Cancer Biol.*, **13**, 49-58.
13. Joerger, A.C., Fersht, A.R., (2008). Structural biology of the tumor suppressor p53. *Annu. Rev. Biochem.*, **77**, 557-582.
14. Joerger, A.C., Fersht, A.R., (2016). The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches. *Annu. Rev. Biochem.*, **85**, 375-404.
15. Riley, T., Sontag, E., Chen, P., Levine, A., (2008). Transcriptional control of human p53-regulated genes. *Nat. Rev. Mol. Cell Biol.*, **9**, 402-412.
16. Menendez, D., Inga, A., Resnick, M.A., (2009). The expanding universe of p53 targets. *Nat. Rev. Cancer.*, **9**, 724-737.
17. Nguyen, T.T., Grimm, S.A., Bushel, P.R., Li, J., Li, Y., Bennett, B.D., Lavender, C.A., Ward, J.M., Fargo, D.C., Anderson, C.W., Li, L., Resnick, M.A., Menendez, D., (2018). Revealing a human p53 universe. *Nucleic Acids Res.*, **46**, 8153–8167.
18. Clore, G.M., Ernst, J., Clubb, R., Omichinski, J.G., Kennedy, W.M., Sakaguchi, K., Appella, E., Gronenborn, A.M., (1995). Refined solution structure of the oligomerization domain of the tumour suppressor p53. *Nat. Struct. Biol.*, **2**, 321-333.

19. Jeffrey, P.D., Gorina, S., Pavletich, N.P., (1995). Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science*, **267**, 1498-1502.
20. Lee, W., Harvey, T.S., Yin, Y., Yau, P., Litchfield, D., Arrowsmith, C.H., (1994). Solution structure of the tetrameric minimum transforming domain of p53. *Nat. Struct. Biol.*, **1**, 877-890.
21. Chen, X., Ko, L.J., Jayaraman, L., Prives, C., (1996). p53 levels, functional domains, and DNA damage determine the extent of the apoptotic response of tumor cells. *Genes Dev*, **10**, 2438-2451.
22. Zhao, R., Gish, K., Murphy, M., Yin, Y., Notterman, D., Hoffman, W.H., Tom, E., Mack, D.H., Levine, A.J., (2000). Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. *Genes Dev.*, **14**, 981-993.
23. Kracikova, M., Akiri, G., George, A., Sachidanandam, R., Aaronson, S.A., (2013). A threshold mechanism mediates p53 cell fate decision between growth arrest and apoptosis. *Cell Death Differ.*, **20**, 576-588.
24. Espinosa, J.M., (2008). Mechanisms of regulatory diversity within the p53 transcriptional network. *Oncogene*, **27**, 4013-4023.
25. Vousden, K.H., (2000). p53: death star. *Cell*, **103**, 691-694.
26. Kitayner, M., Rozenberg, H., Kessler, N., Rabinovich, D., Shaulov, L., Haran, T.E., Shakked, Z., (2006). Structural basis of DNA recognition by p53 tetramers. *Mol. Cell*, **22** 741-753.
27. Beno, I., Rosenthal, K., Levitine, M., Shaulov, L., Haran, T.E., (2011). Sequence-dependent cooperative binding of p53 to DNA targets and its relationship to the structural properties of the DNA targets. *Nucleic Acids Res.*, **39**, 1919-1932.
28. Jordan, J.J., Menendez, D., Sharav, J., Beno, I., Rosenthal, K., Resnick, M.A., Haran, T.E., (2012). Low-level p53 expression changes transactivation rules and reveals superactivating sequences. *Proc. Natl. Acad. Sci. USA*, **109**, 14387-14392.
29. Kitayner, M., Rozenberg, H., Rohs, R., Suad, O., Rabinovich, D., Honig, B., Shakked, Z., (2010). Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat. Struct. Mol. Biol.*, **17**, 423-429.
30. Chen, Y., Dey, R., Chen, L., (2010). Crystal structure of the p53 core domain bound to a full consensus site as a self-assembled tetramer. *Structure*, **18**, 246-256.
31. Chen, Y., Zhang, X., Dantas Machado, A.C., Ding, Y., Chen, Z., Qin, P.Z., Rohs, R., Chen, L., (2013). Structure of p53 binding to the BAX response element reveals DNA unwinding and compression to accommodate base-pair insertion. *Nucleic Acids Res.*, **41**, 8368-8376.
32. Zhang, X., Dantas Machado, A.C., Ding, Y., Chen, Y., Lu, Y., Duan, Y., Tham, K., Chen, L., Rohs, R., Qin, P.Z., (2014). Conformations of p53 Response Elements in Solution Deduced Using Site-Directed Spin Labeling and Monte Carlo Sampling. *Nucleic Acids Res.*, **42**, 2789–2797.
33. Golovenko, D., Brauning, B., Vyas, P., Haran, T.E., Rozenberg, H., Shakked, Z., (2018). New Insights into the Role of DNA Shape on Its Recognition by p53 Proteins. *Structure*, **26**, 1-14.
34. Senitzki, A., Safieh, J., Sarma, V., Golovenko, D., Danin-Poleg, Y., Inga, A., Haran, T.E., (2021). The complex architecture of p53 binding sites. *Nucleic Acids Res.*, **49**, 1364–1382.
35. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M., Zhurkin, V.B., (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA*, **95**, 11163-11168.
36. Balasubramanian, S., Xu, F., Olson, W.K., (2009). DNA sequence-directed organization of chromatin: structure-based computational analysis of nucleosome-binding sequences. *Biophys. J.*, **96**, 2245-2260.
37. Zhou, X., Cao, B., Lu, H., (2017). Negative auto-regulators trap p53 in their web. *J. Mol. Cell Biol.*, **9**, 62-68.

38. Min, S.H., Kim, D.M., Heo, Y.S., Kim, Y.I., Kim, H.M., Kim, J., Han, Y.M., Kim, I.C., Yoo, O.J., (2009). New p53 target, phosphatase of regenerating liver 1 (PRL-1) downregulates p53. *Oncogene*, **28**, 545-554.
39. Krzeszinski, J.Y., Choe, V., Shao, J., Bao, X., Cheng, H., Luo, S., Huo, K., Rao, H., (2014). XPC promotes MDM2-mediated degradation of the p53 tumor suppressor. *Mol. Biol. Cell*, **25**, 213-221.
40. Bourdon, J.C., Deguin-Chambon, V., Lelong, J.C., Dessen, P., May, P., Debuire, B., May, E., (1997). Further characterisation of the p53 responsive element--identification of new candidate genes for trans-activation by p53. *Oncogene*, **14**, 85-94.
41. Ricci, M.S., El-Deiry, W.S., (2007). *The Extrinsic Pathway of Apoptosis*. In D.A. Gewirtz, S.E. Holt, S. Grant (eds.), *Apoptosis, Senescence, and Cancer. Cancer Drug Discovery and Development series*, Humana Press.
42. Offer, H., Erez, N., Zurer, I., Tang, X., Milyavsky, M., Goldfinger, N., Rotter, V., (2002). The onset of p53-dependent DNA repair or apoptosis is determined by the level of accumulated damaged DNA. *Carcinogenesis*, **23**, 1025-1032.
43. Williams, A.B., Schumacher, B., (2016). p53 in the DNA-Damage-Repair Process. *Cold Spring Harb. Perspect. Med.*, **6**.
44. Schneider, T.D., Stormo, G.D., Gold, L., Ehrenfeucht, A., (1986). Information content of binding sites on nucleotide sequences. *J. Mol. Biol.*, **188**, 415-31.
45. Schneider, T.D., Stephens, R.M., (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097-6100.
46. Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., (2004). WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188-1190.
47. Hafner, A., Bulyk, M.L., Jambhekar, A., Lahav, G., (2019). The multiple mechanisms that regulate p53 activity and cell fate. *Nat. Rev. Mol. Cell. Biol.*, **20**, 199-210.
48. Berg, O.G., von Hippel, P.H., (1987). Selection of DNA binding sites by regulatory proteins. Statistical- mechanical theory and application to operators and promoters. *J. Mol. Biol.*, **193**, 723-750.
49. Berg, O.G., von Hippel, P.H., (1988). Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. *J. Mol. Biol.*, **200**, 709-723.
50. Stormo, G.D., (1991). Probing information content of DNA-binding sites. *Methods Enzymol.*, **208**, 458-68.
51. Hertz, G.Z., Stormo, G.D., (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics*, **15**, 563-577.
52. Ho, W.C., Fitzgerald, M.X., Marmorstein, R., (2006). Structure of the p53 core domain dimer bound to DNA. *J. Biol. Chem.*, **281**, 20494-20502.
53. Malecka, K.A., Ho, W.C., Marmorstein, R., (2009). Crystal structure of a p53 core tetramer bound to DNA. *Oncogene*, **28**, 325-333.
54. Farkas, M., Hashimoto, H., Bi, Y., Davuluri, R.V., Resnick-Silverman, L., Manfredi, J.J., Debler, E.W., McMahon, S.B., (2021). Distinct mechanisms control genome recognition by p53 at its target genes linked to different cell fates. *Nat. Commun.*, **12**, 484.
55. Weinberg, R.L., Veprintsev, D.B., Bycroft, M., Fersht, A.R., (2005). Comparative binding of p53 to its promoter and DNA recognition elements. *J. Mol. Biol.*, **348**, 589-596.
56. Crothers, D.M., Drak, J., Kahn, J.D., Levene, S.D., (1992). DNA bending, flexibility, and helical repeat by cyclization kinetics. *Methods Enzymol.*, **212**, 3-29.
57. Zhang, Y., Crothers, D.M., (2003). High-throughput approach for detection of DNA bending and flexibility based on cyclization. *Proc. Natl. Acad. Sci. USA*, **100**, 3161-3166.
58. Vyas, P., Beno, I., Xi, Z., Stein, Y., Golovenko, D., Kessler, N., Rotter, V., Shakked, Z., Haran, T.E., (2017). Diverse p53/DNA binding modes expand the repertoire of p53 response elements. *Proc. Natl. Acad. Sci. USA*, **114**, 10624-10629.

59. Lokshin, M., Tanaka, T., Prives, C., (2005). Transcriptional regulation by p53 and p73. *Cold Spring Harb. Symp. Quant. Biol.*, **70**, 121-128.
60. Boggs, K., Reisman, D., (2006). Increased p53 transcription prior to DNA synthesis is regulated through a novel regulatory element within the p53 promoter. *Oncogene*, **25**, 555-565.
61. Reisman, D., Takahashi, P., Polson, A., Boggs, K., (2012). Transcriptional Regulation of the p53 Tumor Suppressor Gene in S-Phase of the Cell-Cycle and the Cellular Response to DNA Damage. *Biochem. Res. Int.*, **2012**, 808934.
62. Hafner, A., Stewart-Ornstein, J., Purvis, J.E., Forrester, W.C., Bulyk, M.L., Lahav, G., (2017). p53 pulses lead to distinct patterns of gene expression albeit similar DNA-binding dynamics. *Nat. Struct. Mo.l Biol.*, **24**, 840-847.
63. Verfaillie, A., Svetlichnyy, D., Imrichova, H., Davie, K., Fiers, M., Kalender Atak, Z., Hulselmans, G., Christiaens, V., Aerts, S., (2016). Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic. *Genome Res.*, **26**, 882-895.
64. Younger, S.T., Rinn, J.L., (2017). p53 regulates enhancer accessibility and activity in response to DNA damage. *Nucleic Acids Res.*, **45**, 9889-9900.
65. Catizone, A.N., Uzunbas, G.K., Celadova, P., Kuang, S., Bose, D., Sammons, M.A., (2020). Locally acting transcription factors regulate p53-dependent cis-regulatory element activity. *Nucleic Acids Res.*, **48**, 4195-4213.
66. Su, D., Wang, X., Campbell, M.R., Song, L., Safi, A., Crawford, G.E., Bell, D.A., (2015). Interactions of chromatin context, binding site sequence content, and sequence evolution in stress-induced p53 occupancy and transactivation. *PLoS Genet.*, **11**, e1004885.
67. Freewoman, J.M., Snape, R., Cui, F., (2021). Temporal gene regulation by p53 is associated with the rotational setting of its binding sites in nucleosomes. *Cell Cycle*, **20**, 792-807.
68. Coronel, L., Hackes, D., Schwab, K., Riege, K., Hoffmann, S., Fischer, M., (2022). p53-mediated AKT and mTOR inhibition requires RFX7 and DDIT4 and depends on nutrient abundance. *Oncogene*, **41**, 1063-1069.
69. Coronel, L., Riege, K., Schwab, K., Forste, S., Hackes, D., Semerau, L., Bernhart, S.H., Siebert, R., Hoffmann, S., Fischer, M., (2021). Transcription factor RFX7 governs a tumor suppressor network in response to p53 and stress. *Nucleic Acids Res.*, **49**, 7437-7456.
70. Zhang, Y., Crothers, D.M., (2003). Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization. *Biophys. J.*, **84**, 136-153.
71. Zhang, Y., Xi, Z., Hegde, R.S., Shakked, Z., Crothers, D.M., (2004). Predicting indirect readout effects in protein-DNA interactions. *Proc. Natl. Acad. Sci. USA*, **101**, 8337-8341.
72. Spitzer, M., Wildenhain, J., Rappsilber, J., Tyers, M., (2014). BoxPlotR: a web tool for generation of box plots. *Nat. Methods*, **11**, 121-122.

Table 1. End functional-outcome supergroups with at least five genes each (related to Figure 1)

Supergroup	Functional outcomes included in the group
Cell adhesion/anti-cell-motility	Increase in cell-matrix adhesion; increase in cell-cell adhesion; anti-metastasis; anti-migration; anti-invasion; anti-angiogenesis
Cell growth	Cell division machinery; mitotic cell cycle; mitotic spindle organization; regulation of cytokinesis; symmetric cell division; cell proliferation
Cellular stress response	Response to: ER stress; metabolic stress, hypoxia, heat stress, cold stress; osmotic stress; unfolded protein response; ROS; oxidative stress; oxidative DNA damage; stress granules formation
Cytoskeleton/cell motility	Cytoskeleton organization; actin polymerization; ECM organization and degradation; muscle contraction; neuronal migration; movement along microtubules; directional migration; apical-basal polarity; cytoskeleton re-organization; axon guidance; migration; invasion; wound healing; EMT; platelet aggregation; angiogenesis; decrease in cell adhesion
Development	Embryonic development; morphogenesis; neurogenesis; synaptogenesis; muscle development
DNA repair	Damage repair; translesion DNA synthesis; synthesis of dNTPs
Early DNA damage response (DDR) steps	DNA damage response activation; cell-cycle arrest (G1/S and G2/M); p53 positive regulation; other events closely following DNA damage; immediate early genes for DDR
Energy metabolism	Energy metabolism; glucose metabolism; oxidative phosphorylation; ATP synthesis
Lipid metabolism	Lipid metabolism; lipid droplets; cholesterol metabolism; steroids metabolism; steroidogenesis
Extrinsic apoptosis	Death receptor pathway
Innate immunity	Innate immune response; inflammation; pyroptosis
Intrinsic apoptosis	Intrinsic apoptosis machinery and effectors
Ion homeostasis	Ion channels; ion transport; transport of small molecules
Membrane trafficking	Endocytosis; exocytosis; vesicle transport; secretory pathways; Golgi; membrane transport; vesicular lipid transport
p53 negative regulation	Targeting p53 for destruction; p53 down regulation after damage repair
Cell signaling	Signal transduction by various pathways, when the signaling event is followed by multiple possible functional-outcome results
Stem cells and differentiation	Stem cell biology; stem cells self-renewal; pluripotency maintenance and regulation; stem cells differentiation
Survival	Anti-apoptosis; anti-necrosis; anti-pyroptosis; anti other kinds of deaths; cell survival

Table 2. Best-fit parameters from cyclization kinetics measurements for the sequences studied here

Functional outcome	Gene symbol	Sequence	Bend angle (°) [a, b]	Twist angle (°) [b]	Roll and tilt flexibility (°) [b, c]	Twist flexibility (°) [b,d]	Deformability (V(B), °^{3}Å^3) [e]
Cell cycle	CCNG1	GCACAAGCCCAGGCTAGTCC	-0.72 (0.09)	34.52 (0.08)	4.54 (0.09)	5.77 (0.09)	2.326
DNA Repair	RRM2B	TGACATGCCCAGGCATGTCT	-1.49 (0.10)	34.14 (0.10)	4.20 (0.09)	5.65 (0.09)	2.495
Cell cycle	p21-5’	GAACATGTCCCAACATGTTG	-0.06 (0.08)	34.04 (0.09)	4.45 (0.08)	5.34 (0.08)	2.263
Apoptosis	PMAIP1	GAGCGTGTCCGGGCAGGTCG	0.83 (0.12)	34.30 (0.15)	4.33 (0.11)	5.07 (0.12)	2.042
Apoptosis	TP53AIP1	TCTCTTGCCCGGGCTTGTCG	-1.02 (0.11)	33.96 (0.11)	4.20 (0.12)	4.91 (0.10)	1.926
Range of values observed [f]			−10.0 to 2.4	33.96 to 34.85	4.20 to 6.40	3.9 to 5.80	

[a] Bending is by roll and its center is located at the fourth step of all sequences, i.e. at the C-A, C-G, or C-T step.
[b] Numbers in parenthesis are the simulation errors, calculated as described in ref. [70].
[c] Roll and tilt flexibility are the thermal fluctuations of the roll and tilt angle between adjacent base pairs [57, 70].
[d] Twist flexibility is the thermal fluctuation of the twist angle between adjacent bases [57, 70].
[e] The deformability values are those from Table S1, and were calculated as previously described [34], using the values of [36].
[f] These values denote the range of observed values for these parameters in other DNA sequences, determined by the same method and the same kind of constructs used here [27, 34, 58, 71].

Figure legends:

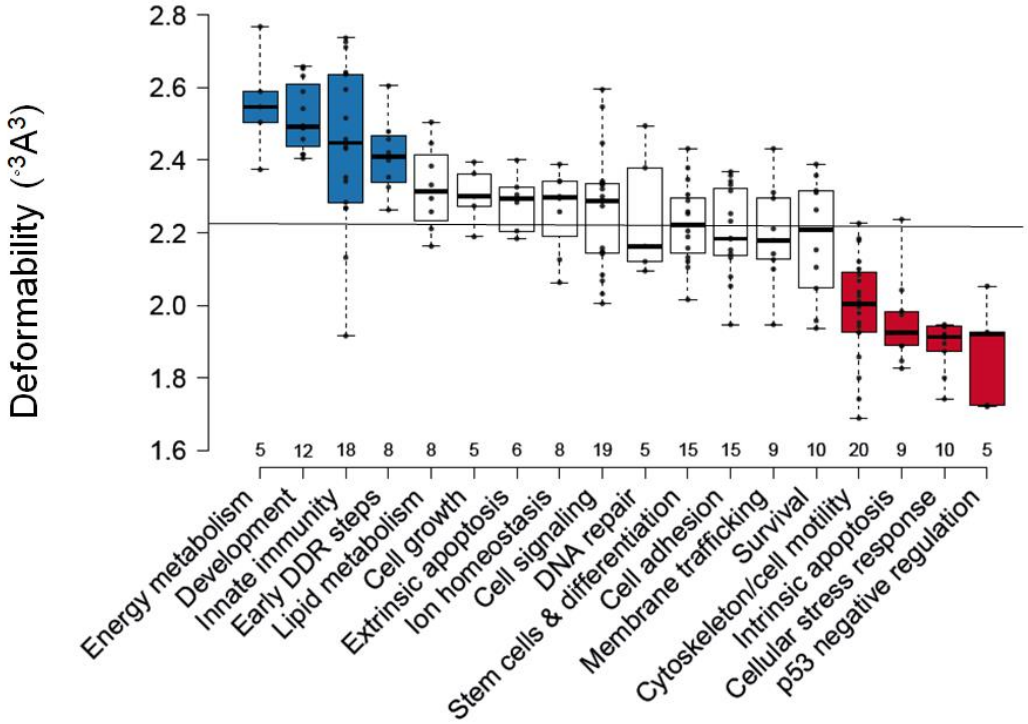

Figure 1. p53 REs belonging to functional-outcome groups that are needed early in the response to stress are more deformable than those needed later in the response to stress. Shown are functional-outcome groups with at least five genes each. All data points are shown as dots. The line across all groups marks the mean of means of the deformability of all analyzed groups. Center lines show the medians of each group. The figure was created by BoxPlotR [72]. Box limits indicate the 25th and 75th percentiles, as determined by the R software. Whiskers extend to maximum and minimum values. Boxes are arranged by decreasing group average values from left to right. Blue boxes denote the “flexible REs category”, and red boxes signify the “rigid REs category”. For a list of group average values see Table S2, and for significant differences between the groups see Table S3. DDR = DNA damage response.

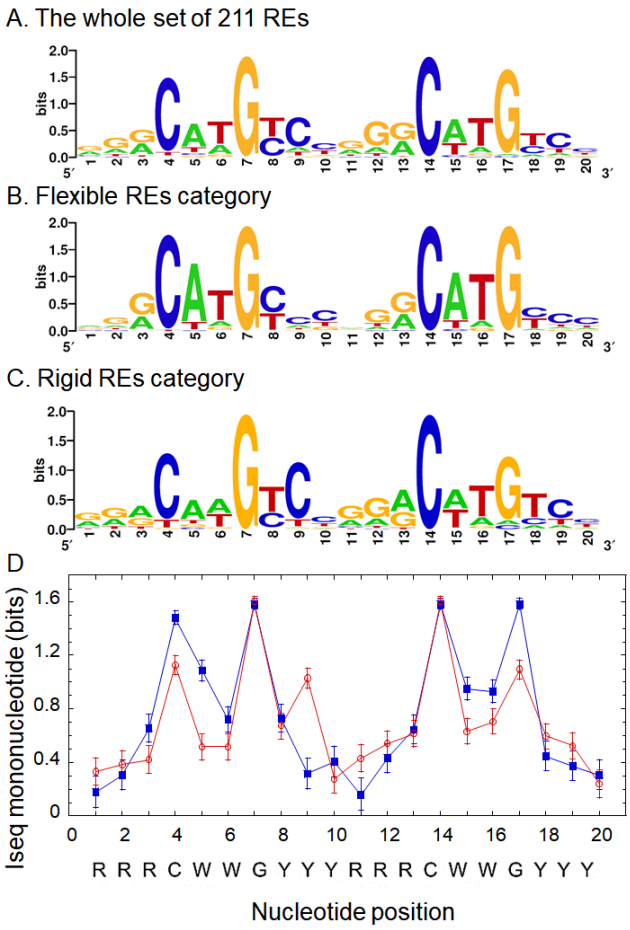

Figure 2. Sequence logos and information content of the studied REs. (A) All 211 p53 REs without spacers and with known functional outcome of p53 activation studied here. (B) REs belonging to the flexible category. (C) REs belonging to the rigid category. All sequences were aligned by transcription direction before generating the sequence logos. Sequence logos were generated by weblogo [45]. (D) Information content per position ($Iseq_l$) on the mononucleotide level for the REs belonging to the flexible category (blue squares) versus those belonging to the rigid category (red circles).

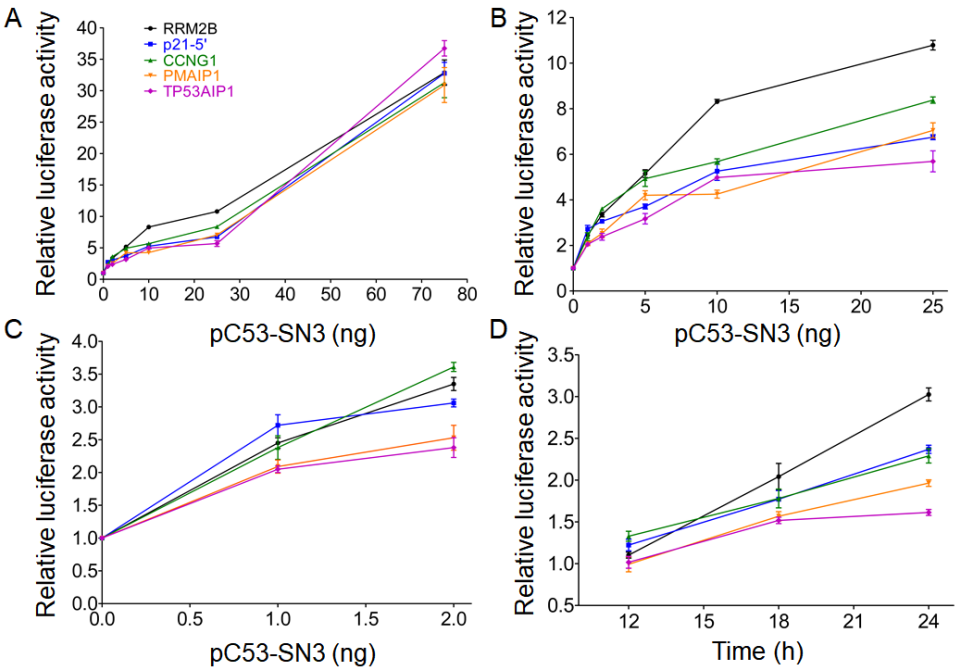

Figure 3. Transactivation from p53 REs as a function of p53 plasmid levels and of time. Fold increase in transactivation level was measured from three flexible p53 REs (CCNG1, RRM2B, and p21-5') and two rigid REs (PMAIP1, and TP53AIP1) as a function of six different p53 expression levels, 48 h post-transfection (A-C), and additional three time points, using 25 ng p53 plasmid (D). Luminescence values were normalized to the transfection efficiency of co-transfected, constitutively expressed Gaussia luciferase (Gluc). Results were further normalized first to the empty Cypridina secreted luciferase (CLuc) vector, and then to results obtained without p53. Error bars represent the mean ± SD of four to eight independent experiments, each containing two technical replicas. (A) The whole range, 1-75 ng p53 plasmid. (B)-(C) for clarity these panels show the 1-25 ng and 1-2 ng range, respectively.

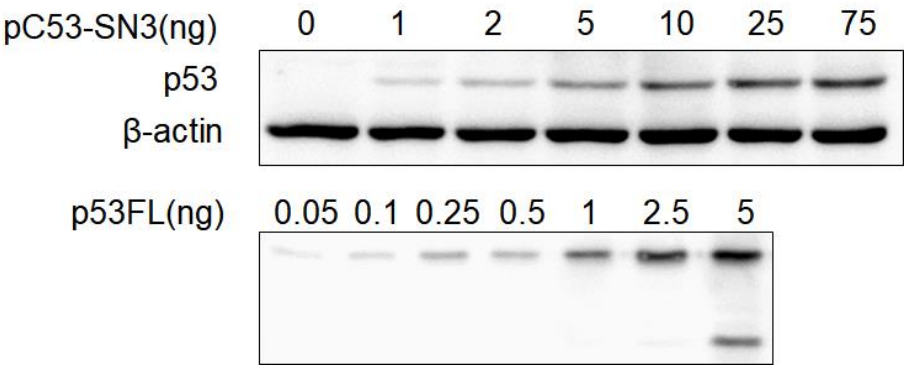

Figure 4: Estimation of the amount of p53 protein in H1299 cells upon transfection with various amounts of p53 plasmid. Upper panel shows H1299 cells transfected with increasing amounts of p53 expression vector, as described in Figure S2. Lower panel represents a serial dilution of purified full-length p53 recombinant protein. The lower band in the 5-ng lane is a degradation product, which was included in the quantification.

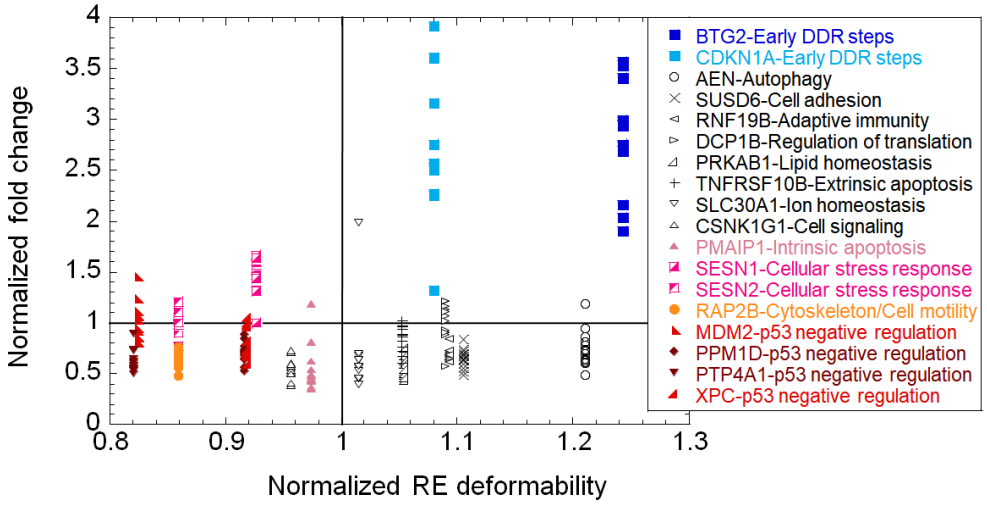

Figure 5. Endogenous mRNA levels from p53-dependent genes are grouped by their RE deformability. The data shown is based on the RNA-seq results from Hafner et al. [62]. Fragments Per Kilobase Million (FPKM) values of each gene (i), at each time point (j) were normalized to their FPKM value at T = 0, arriving at fold-change values. Then these values were further normalized to the average fold-change of all genes at each particular time point (j). That is, each point is calculated from the equation: $(FPKM_iT_j/FPKM_iT_0)/\frac{1}{n}\sum_{i=1}^{n}(FPKM_iT_j/FPKM_iT_0)$. Deformability of genes' REs was normalized in the same manner. Plotted are the normalized fold-change values as a function of normalized deformability values. The lines crossing each axis denote values that are above/below the average value of the parameter shown in each specific axis.