nature structural & molecular biology

Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs

Malka Kitayner[1,5], Haim Rozenberg[1,5], Remo Rohs[2–5], Oded Suad[1], Dov Rabinovich[1,6], Barry Honig[2–4] & Zippora Shakked[1]

p53 binds as a tetramer to DNA targets consisting of two decameric half-sites separated by a variable spacer. Here we present high-resolution crystal structures of complexes between p53 core-domain tetramers and DNA targets consisting of contiguous half-sites. In contrast to previously reported p53–DNA complexes that show standard Watson-Crick base pairs, the newly reported structures show noncanonical Hoogsteen base-pairing geometry at the central A-T doublet of each half-site. Structural and computational analyses show that the Hoogsteen geometry distinctly modulates the B-DNA helix in terms of local shape and electrostatic potential, which, together with the contiguous DNA configuration, results in enhanced protein-DNA and protein-protein interactions compared to noncontiguous half-sites. Our results suggest a mechanism relating spacer length to protein-DNA binding affinity. Our findings also expand the current understanding of protein-DNA recognition and establish the structural and chemical properties of Hoogsteen base pairs as the basis for a novel mode of sequence readout.

The tumor suppressor protein p53 acts as a transcription factor in response to a wide variety of cellular stress signals, regulating the expression of an array of different genes that mediate a variety of growth-inhibitory events, including cell-cycle arrest and apoptosis[1–4]. p53 binds in a sequence-specific manner to DNA binding sites that consist of two copies of the 10–base pair (bp) motif RRRCWWGYYY (R = A,G; W = A,T; Y = C,T) separated by a variable number of base pairs[5–7]. Upon binding to DNA targets, p53 forms tetramers, its basic functional unit[8,9]. Human p53 is a 393-residue protein that contains three major functional domains[10–12]: a transactivation domain at the N terminus, a core domain that contains the sequence-specific DNA binding domain, and a C-terminal domain that comprises an oligomerization domain followed by a regulatory domain. The core domain is the main target for mutations, as 80–90% of the missense mutations identified in human tumors are found in this region[13]. Knowing how p53 chooses which genes to activate is critical in understanding its role as a tumor suppressor. A number of variables, such as cofactor recruitment, post-translational modification, differential binding affinity of p53 for DNA and p53-mediated gene activation and repression, all have a role in dictating p53 response[1–4].

Several studies have shown that most of the DNA sequences involved in cell-cycle arrest, DNA repair and negative regulation bind p53 with high affinity[14,15]. In contrast, sequences from genes involved in apoptosis show large variations in affinity, with some binding as tightly as cell cycle–arrest sequences and others up to 100 times more weakly[15]. It was shown that cell cycle–related sequences with high p53 binding affinities do not have 'spacers' interspersed between the two half-sites and have two or fewer mismatches within the consensus DNA binding site. In contrast, apoptosis-related sequences with lower affinity for p53 binding have interspersed sequences between the two half-sites and/or three or more mismatches[14–16], suggesting that intrinsic structural properties of the DNA sequence might help to modulate the affinity of p53 for its binding sites[14,15].

The binding of p53 molecules to DNA has been studied using a variety of biophysical methods (reviewed in ref. 17). The first structural information was for the core domain of human p53 bound as a monomer to a DNA duplex containing a single decameric motif[18]. It was found that the core domain adopts an immunoglobulin-like β-sandwich structure that provides a scaffold for a DNA binding surface consisting of a loop-sheet-helix motif and two loops stabilized by a zinc ion[18]. We presented the first structural data on human p53 tetramers bound to DNA—specifically, on complexes between the p53 core domain and DNA dodecamers incorporating consensus half-sites of different sequences[19]. We showed that two p53 molecules bind each DNA half-site in a sequence-specific manner and that two such dimers assemble into tetramers. In all these structures, the two decameric half-sites were separated by a 2-bp spacer. DNA binding studies have shown that such DNA targets have lower binding affinity compared to those where the two half-sites are contiguous[19]. Additional structural data on p53-DNA interactions were obtained from crystal structures of a dimer and a tetramer of the mouse p53 core domain, but in this case, the protein was covalently linked to one DNA half-site in the dimer and to two contiguous half-sites in the tetramer via cysteine side chains and cytosine bases[20,21]. Until recently, no structural information has been available on naturally assembled p53 tetramers on full-length DNA targets with contiguous

[1]Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel. [2]Howard Hughes Medical Institute, [3]Center for Computational Biology and Bioinformatics and [4]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York, USA. [5]These authors contributed equally to this work. [6]Deceased 6 September 2008. Correspondence should be addressed to B.H. (bh6@columbia.edu) or Z.S. (zippi.shakked@weizmann.ac.il).

Received 6 February; accepted 5 March; published online 4 April 2010; doi:10.1038/nsmb.1800

Table 1 Data collection and refinement statistics

	p53–DNA[a] complex 1	p53–DNA complex 2	p53–DNA complex 3
Data collection			
Space group	*C*2	*C*2	*C*2
Cell dimensions			
a, *b*, *c* (Å)	137.2, 49.8, 34.0	137.3, 49.6, 33.9	138.4, 49.8, 68.1
α, *β*, *γ* (°)	90.0, 94.0, 90.0	90.0, 94.7, 90.0	90.0, 93.5, 90.0
Resolution (Å)	1.8 (1.83–1.8)[b]	1.7 (1.73–1.7)	1.88 (1.91–1.88)
R_{sym}	4.9 (22.5)	5.9 (20.3)	9.9 (48.6)
$I/\sigma I$	32.1 (6.3)	21.0 (5.0)	25.3 (4.7)
Completeness (%)	95.0 (91.6)	98.9 (96.9)	99.9 (100)
Redundancy	5.1 (5.0)	4.0 (3.7)	6.9 (7.2)
Refinement			
Resolution (Å)	24.9–1.8	22.8–1.7	34.5–1.9
No. of unique reflections	20,253	24,859	34,235
Protein/DNA duplexes in asymmetric unit	1/0.5	1/0.5	2/0.5
R_{work} / R_{free}	14.8 / 21.2	15.0 / 21.4	22.1 / 26.6
No. atoms			
Protein	1,560	1,576	3,058
DNA	225	225	410
Ligand/ion	9	1	4
Water	343	344	312
B-factors (Å^2)			
Protein	27.7	25.6	27.2
DNA	26.9	28.4	26.0
Ligand/ion	39.6	20.8	31.2
Water	36.7	35.3	33.2
R.m.s. deviations			
Bond lengths (Å)	0.026	0.027	0.031
Bond angles (°)	2.34	2.41	2.82

Each dataset was collected from a single crystal.
[a]The DNA used is cGGGCATGCCCg for complexes 1 and 2 and tGGGCATGCCCGGGCATGCCC for complex 3. [b]Values in parentheses are for the highest-resolution shell.

half-sites. Very recently, the crystal structure of a p53 core-domain tetramer bound to a different DNA sequence with contiguous half-sites was reported[22]. Differences between that structure and the one reported here are discussed below.

Here, we present the high-resolution crystal structures of self-assembled human p53 core-domain tetramers bound to DNA targets with contiguous decameric half-sites, each showing Hoogsteen base pairs[23] at the central A-T doublet of each half-site. A comparative analysis of the newly determined structures and our previously reported structures shows that the three-dimensional architecture of p53–DNA complexes depends on the specific sequence and shape of the DNA half-sites and the DNA spacer between them. Our results provide new insights into the complexity and variability of sequence-specific DNA recognition by p53.

RESULTS

Overall structure of p53 bound to contiguous DNA half-sites

The core domain of human p53 used in the present study spans residues 94–293 and is referred to as the core or p53 DNA binding domain, p53DBD. For crystallization experiments, we used the wild-type (WT) core domain and a 'structurally restored' double-mutant core, R249S H168R, that was shown to bind DNA[24]. We cocrystallized each protein with the dodecameric sequence, cGGGCATGCCCg, used previously[19]. We also cocrystallized the WT p53 with a 21-mer DNA that contains the same DNA half-sites and a 5′-overhung thymidine, tGGGCATGCCCGGGCATGCCC. We crystallized the present p53–DNA complexes in the *C*2 space group, where the asymmetric unit contains either one protein molecule and a single DNA strand that contains one half-site (complexes 1 and 2 in **Table 1**) or two protein molecules and a single DNA strand that contains two half-sites (complex 3 in **Table 1**) that, via symmetry operations, generate continuous stacks of p53–DNA tetramers (**Supplementary Fig. 1**). Complexes 1 and 2 are identical apart from the two mutations (R249S H168R) in the core domain of complex 2.

The three crystallized complexes show a new organization of p53 molecules on the DNA helix. Although we used the same 12-mer oligomer DNA for complexes 1 and 2 as in our previously determined crystal structures[19,24], here the bound B-DNA organizes the two decameric half-sites in a different way. In the previously reported structures, the 3′-end nucleotides and the 5′-end nucleosides of two double-helical dodecamers remain part of each helix and act as a 2-bp spacer between the two half-sites (**Fig. 1a**). In the structures reported here, the last two nucleotides at the 3′ end of each dodecameric strand are extrahelical such that the tenth base (C10) is stacked on the first base (C1) of an adjacent molecule, thus creating a continuous DNA helix with a 10-bp helical repeat (**Fig. 1b**). Whereas in complexes 1 and 2 the two half-sites are not covalently linked, the structure of complex 3 contains covalently linked decameric DNA half-sites (**Fig. 1c**) but is nevertheless very similar to that of the other two complexes. This finding, taken together with our previously reported structural data on p53–DNA complexes[19], shows that backbone gaps in p53 binding sites do not interfere with the self-assembly of p53–DNA tetramers, provided that the base pairs in such regions are properly arranged to form a continuous DNA binding site that is recognized by p53. The analysis presented here is based on the higher-resolution WT structure of complex 1, unless stated otherwise. The complexes from our previous study[19] and the present one are referred to as type I and type II, respectively.

In both types of complexes, four p53 molecules interact with two DNA half-sites to form a tetramer that is a dimer of dimers. However, the two types of complexes differ in the relative orientation of the two half-sites and hence in the three-dimensional architecture of the corresponding p53–DNA tetramers (shown here in **Fig. 2** and in Fig. 1 from ref. 19). In the type II complexes, the two dimers are

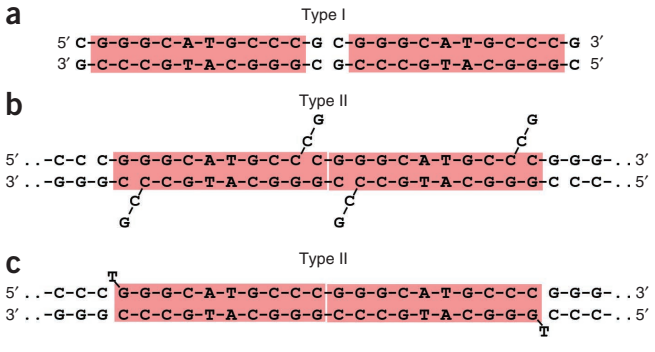

Figure 1 DNA binding sites of type I and type II complexes. (**a**) Type I complex formed by DNA dodecamers. The two half-sites are separated by a 2-bp spacer[19]. (**b**) Type II complexes formed by DNA dodecamers. The two half-sites are contiguous, as the two nucleotides at each 3′ end are extrahelical and the decameric duplexes are stacked end to end (complexes 1 and 2 in **Table 1**). (**c**) Type II complex formed by 21-mer DNA oligomers: a 20-bp duplex with contiguous half-sites and extrahelical thymidine nucleosides at each 5′ end (complex 3 in **Table 1**). Complex 3 is essentially identical in structure to complexes 1 and 2.

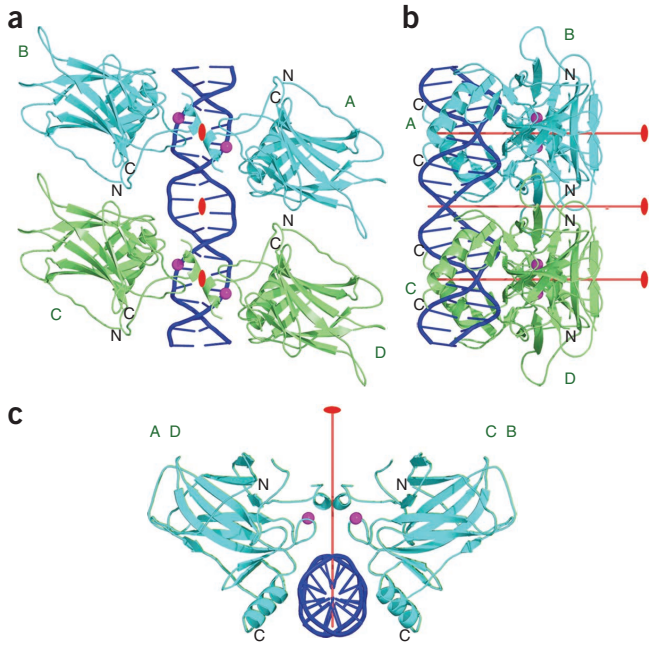

Figure 2 Overall structure of the p53 core-domain tetramer bound to DNA with contiguous half-sites. (**a**) Four p53 core domains (designated A, B, C and D) interact with a 20-bp DNA (shown in blue). The core tetramer is a dimer of dimers A-B (cyan) and C-D (in green). The four zinc ions are shown as magenta spheres. View is down the central dyad of the core tetramer. (**b**) View perpendicular to the central dyad and the DNA helix axis. (**c**) View down the DNA helix axis. Shown in red are the central dyad between dimers and the two local dyads within dimers. The figure is based on the coordinates of complex 3. Sequence and secondary structure of the core domain are shown in **Supplementary Figure 2c**.

parallel to each other (**Fig. 2**), whereas in type I complexes, the two decameric half-sites are separated by 2 bp and, as a result, the two DNA-bound p53 dimers are rotated in a clockwise manner by nearly 33° (ref. 19). In type II complexes, each tetramer has a perfect dyad symmetry coinciding with the crystallographic two-fold axes and additional dyad symmetry within each dimer. Hence, the three dyad axes of each complex are parallel to each other, and the overall DNA helix is straight.

Extensive protein-protein interactions stabilize type II complexes

The direct protein-DNA contacts in type II complexes are essentially identical to those in type I complexes. However, the two types of tetramers show distinct differences in their protein-protein interfaces. Two types of protein-protein interfaces were previously characterized[19]. The first one, referred to as the symmetrical or intradimer interface, is located within each core dimer bound to its DNA half-site. The second one, referred to as the translational or interdimer interface, links the two core dimers along the DNA helix.

The intradimer interface formed by the type II complexes shows similar characteristics to that of type I complexes[19]. However, we observed appreciable changes in the orientation of the monomers relative to each other and to the DNA (**Fig. 3a**). We observed a distinct difference at the region distant from the DNA: in type II structures, there is a pair of bidentate salt bridges between Glu180 and Arg181 from each monomer whose existence is further supported by water-mediated hydrogen bonds between the charged residues and the backbone of the H1 helix (**Fig. 3b**). In type I structures, the Arg181 side chains from the two monomers are stacked on each other[19] (**Fig. 3c**). The existence of salt bridges between the arginine and aspartate residues was previously proposed on the basis of site-directed mutagenesis and DNA binding experiments[25]. The charged interactions in type II dimers are facilitated by the further proximity of the monomers in this region relative to that in type I dimers (illustrated in **Fig. 3a**,**c**) and as a result lead to a marked increase in the intradimer buried surface area (850 Å^2 in type II compared with 650 Å^2 in type I complexes).

We found much greater differences in the interdimer interfaces. The buried surface area between monomers along the DNA helix in type II structures is extensive, ~1,200 Å^2, compared to a maximal value of nearly 600 Å^2 observed in type I structures[19], in which the two dimers are further apart because of the 2-bp spacer and their large rotation relative to each other (**Fig. 4**). The type II interface consists mainly of an extensive water-mediated network of polar and charged residues

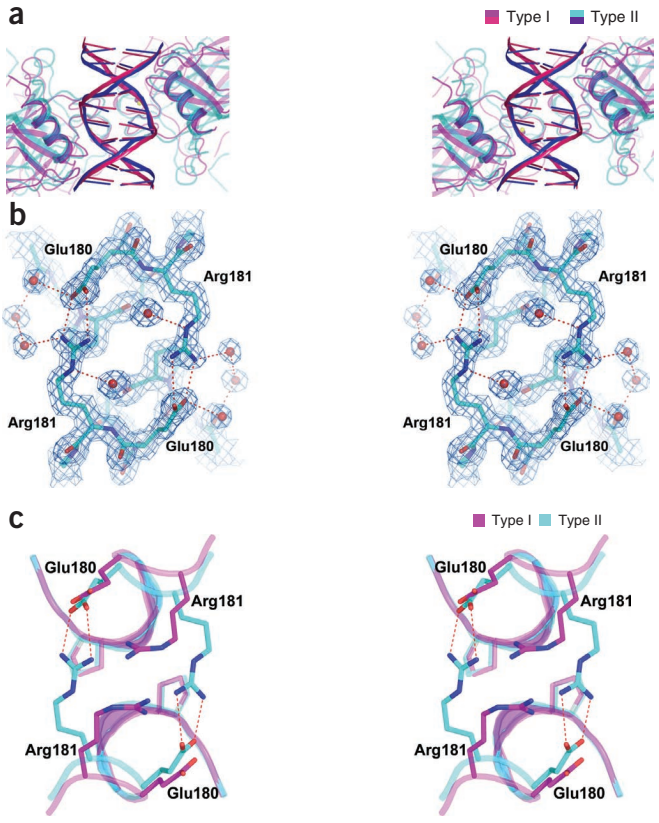

Figure 3 Comparison of p53 core dimers bound to DNA half-sites in type I and type II complexes. (**a**) Superposition of type I structure (DNA in pink, p53 and zinc ion in magenta) onto type II structure (DNA in blue, p53 in cyan, zinc ions in yellow) is based on the DNA backbone atoms. The stereo view is down the dyad of the dimer and into the DNA major groove. The view highlights the different DNA conformations and the different arrangements of the p53 molecules on their DNA in type I and type II complexes. (**b**) Stereo view of the salt bridges between residues Glu180 and Arg181 in type II intradimer interface and the supporting hydration shell, shown within the electron density ($2F_o - F_c$ at 1σ level). Water molecules from the first hydration shell are shown as red spheres (based on complex 2). (**c**) Stereo view of the superposition of the same regions from **b** in type I (magenta) and type II (cyan) complexes, showing the stacking of the guanidinium groups of arginine residues in type I and the salt bridges in type II as well as the further proximity of the protein backbones in type II relative to type I complexes.

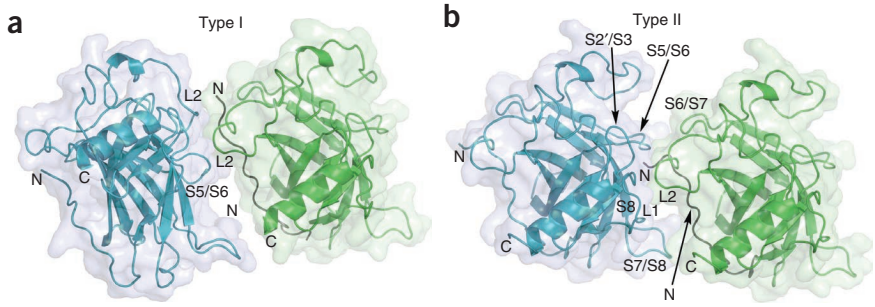

Figure 4 Interdimer interfaces. (**a**,**b**) Type I (**a**) and Type II (**b**) interfaces. Two p53 core domains comprising half of the interdimer interface are shown in cyan and green (the other half is related by dyad symmetry). Secondary elements are labeled (boldface) as defined in **Supplementary Figure 2c**. Also labeled are the N and C termini. The structures of the DNA-bound core domains in type I and type II complexes are very similar except for the flexible L1 recognition loop (see **Supplementary Discussion** and **Supplementary Fig. 7**).

(illustrated in **Supplementary Fig. 2**). Direct interactions between polar and charged residues are listed in **Supplementary Table 1**.

The large increase in protein-protein interactions among the four core domains of type II tetramers provides a structural basis for understanding the different binding affinities measured for p53 core domain and DNA targets that have identical half-sites but are either contiguous or separated by 2 bp[19]. As the protein-DNA contacts are very similar between the two complexes, it appears that the observed higher affinity and stability of p53 complexes with DNA targets with contiguous half-sites is caused by enhanced protein-protein interactions within and between the two dimers observed in type II tetramers. Additional factors relating to DNA shape and electrostatics, discussed below, further contribute to the binding properties of the complexes.

The DNA binding site contains Hoogsteen base pairs

As described above, the DNA binding site in type II complexes is made of two contiguous double-helical decamers stacked end to end in the crystal so as to form a continuous B-DNA helix (**Fig. 1b**,**c**). The DNA trajectory highlighted by the curved helix axis (**Supplementary Fig. 3b**) shows that each of the decameric half-sites is bent slightly toward the major groove at the central A-T doublet and away from the core dimer (**Fig. 2b**) in a manner similar to that shown by type I complexes[19]. However, in type II complexes, the combined two–half-site DNA is straight as a result of a comparable bend toward the major groove at the GC-rich region between half-sites (**Fig. 2b** and **Supplementary Fig. 3b**). A distinctly larger bending at the junction between half-sites and toward the core tetramer is observed in type I complexes, where the two half-sites are separated by a 2-bp spacer[19] (**Supplementary Fig. 3a**). This large deformation results in a reduced distance of nearly 34 Å between the centers of the DNA half-sites in comparison to a straight helix, thereby facilitating the formation of an interdimer interface by the attached core dimers.

A novel feature of the type II p53–DNA complexes is that the A-T base pairs at the center of each half-site show the Hoogsteen geometry, in which hydrogen bonds are formed between the pyrimidine N3 and the purine N7 atoms and either the adenine N6 and the thymine O4 or the guanine O6 and the cytosine N4 atoms. Comparisons between Watson-Crick and Hoogsteen base pairs from type I and type II complexes are shown in **Figure 5a–c**. The effects of the alternative base-pairing geometry on the backbone trajectory and associated groove dimensions are illustrated in **Figure 5d**. A detailed comparison between the Watson-Crick– and Hoogsteen-containing DNA conformations of the same base sequence shows striking changes in several helical parameters, including helix diameter and groove dimensions (**Fig. 6**). A reduction in the helix diameter by 1–2 Å is observed in the vicinity of Hoogsteen base pairs (**Fig. 6a**) and is associated with substantial narrowing of the minor groove in the regions flanking the central CATG elements (**Fig. 6b**). As described above, the DNA oligomer used here for complexes 1 and 2, cGGGCATGCCCg, is the same DNA dodecamer used previously[19,24]. The observation of different structures suggests that type I complexes, with Watson-Crick base pairs and a 2-bp spacer, and type II complexes, with Hoogsteen base pairs, no spacer and extra helical bases, are in equilibrium in solution, and further suggests that changes in crystallization conditions affect the trapping of one form or the other. In the case of complex 3, where the two decameric half-sites are contiguous and covalently linked, only the Hoogsteen-containing structure is obtained, suggesting that this particular form is driven by the enhanced interdimer protein-protein interactions facilitated by the lack of spacer in the DNA sequence.

To further evaluate the importance of Hoogsteen base pairing, we reanalyzed the published data of the mouse p53 core domain

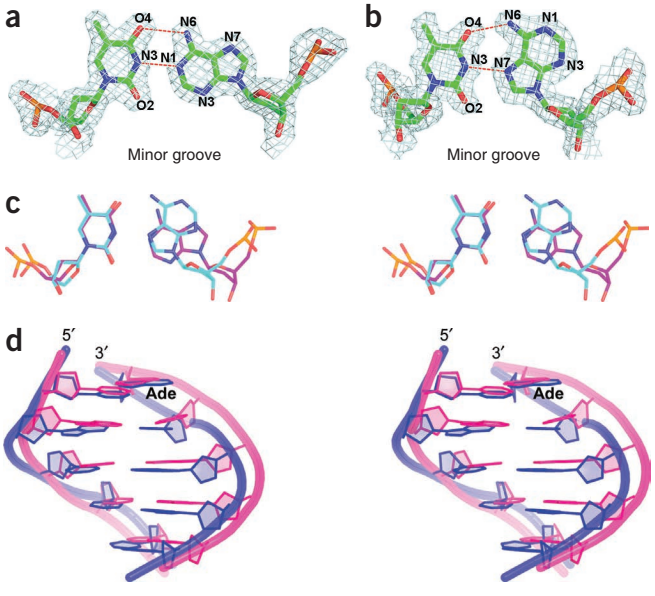

Figure 5 Comparison of Hoogsteen and Watson-Crick base pairs. (**a**,**b**) Watson-Crick A-T base pair from type I complex[19] (PDB 2AC0) (**a**) and Hoogsteen A-T base pair from type II complex (**b**). Both are shown within their respective electron density maps ($2F_o - F_c$ at 1σ level). (**c**) Stereo view of the superposition of Watson-Crick and Hoogsteen base pairs showing the change in the positions of the glycosidic bonds of the adenine bases and the attached backbones. (**d**) Stereo view of DNA quarter sites from type I and type II complexes shown in magenta and blue, respectively, illustrating the change in the backbone trajectory following the Hoogsteen base pair and its effect on the narrowing of the minor groove in type II helix. The superposition is based on the coordinates of the central thymidine nucleotides of the corresponding half-sites. The adenine base is labeled (Ade).

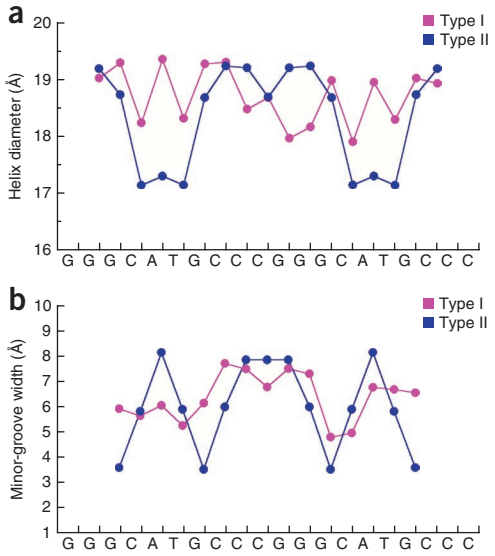

Figure 6 DNA helix parameters in type I and type II complexes. (**a**) Variations in helix diameter along the DNA helix. (**b**) Variations in minor-groove width along the DNA helix. Calculations were performed using our in-house version of Curves[37] adapted for Hoogsteen base pairs. The values for the G-C spacer in type I complex are not shown for clarity.

tetramer covalently linked to a DNA duplex, where the central A-T doublets (underlined) of each of the two contiguous half-sites (GAGCATGCTCGAGCATGCTC) were modeled as Watson-Crick base pairs[21]. Our analysis (described in **Supplementary Methods** and **Supplementary Fig. 4**) shows that a large fraction (greater than 50%) of the A-T base pair doublets at the center of each half-site adopts the Hoogsteen geometry. The trapping of both forms in this crystal structure, compared to a single form in type II structures, is likely to result from changes in the base sequence between the two, as well as from the relatively small interdimer interface in the 'forced' tetramer (~715 Å^2); the latter is probably affected by the conformational deflection of the flexible L1 loop toward the interface imposed by the linker chain[21].

In the recently published crystal structure of a p53 core-domain tetramer bound to DNA with contiguous half-sites, the central A-T base pairs of each half-site are in the standard Watson-Crick geometry[22]. However, there are a number of distinct differences between that structure and the one reported here that appear to be related to the specific DNA sequence used (AGGCATGCCTAGGCATGCCT in the other study[22] and GGGCATGCCCGGGCATGCCC here). For example, it is possible that the difference in sequence affects the base-pairing geometry of the central A-T doublet. The presence of a TpA step between the two half-sites in one sequence and a GC-rich region in the other may allow the DNA to assume alternative conformations that maintain the overall integrity of the complex. In addition, the relative arrangement of the two dimers on the double helix, which depends on the DNA conformation, is different in the two cases. As a result, interdimer interfaces vary between the two complexes in terms of surface area and specific interactions (compare **Fig. 4**, **Supplementary Fig. 2** and **Supplementary Table 1** with Fig. 5 in ref. 22). These differences may have thermodynamic consequences, and indeed we previously found that the protein-DNA binding affinity of the complex used in the other study is considerably lower than that of the complex used here[19]. Thus, small changes in DNA sequence are reflected in different energetic constraints that might well favor one base-pairing geometry over another. This is consistent with the apparent equilibrium between Watson-Crick and Hoogsteen base pairs discussed above for the covalently linked DNA duplex. Another possible source of the geometrical changes between the two structures is that they are influenced by end effects, as only 18 bp were resolved in the crystal structure reported in the other study[22], in comparison to the fully resolved full-length binding sites observed in the present study.

Role of DNA shape and electrostatics in p53-DNA recognition

As described above, the three-dimensional architecture of the p53–DNA tetramer varies from type I to type II, mainly depending on whether there is a spacer sequence between half-sites. Yet the protein-DNA contacts formed by residues from the loop-sheet-helix motif (L1, S10 and H2) and the L3 loop are essentially identical in the two complexes. Even the drastic change in the base-pairing geometry at the central A-T base pairs, from Watson-Crick in type I to Hoogsteen in type II, does not appear to affect the recognition pattern in terms of direct protein-DNA interactions. This is because the two A-T base pairs at the center of each half-site do not participate in direct interactions with the protein. The only protein-DNA contacts involving nucleotides that form Hoogsteen base pairs are made between the symmetrically disposed Arg273 side chains from each core-domain molecule and the thymidine phosphate of each strand. As in type I complexes, these interactions are further buttressed by an extensive network of hydrogen bonds and salt bridges involving the side chains of Arg273, Asp281 and Arg280 as well as a highly ordered water molecule[19].

The minor groove provides the interaction sites for Arg248, the most frequently mutated residue in human cancer[13]. In type I complexes, the Arg248 side chains show variable conformations, and their interactions with the groove are mostly mediated by water molecules[19]. The Arg248 residues in type II complexes show a single conformation and interact via water molecules with the DNA backbone (**Supplementary Fig. 5**). Arg248 is located in narrow regions of the minor groove, which correspond to minima in electrostatic potential (**Fig. 7**). As shown above, the Hoogsteen base pairs are associated with distinct narrowing of the minor groove in the regions flanking the central CATG element of each half-site (**Fig. 6b**). Hoogsteen base pairing leads to a displacement of the DNA backbone toward the minor groove, because the preservation of base-pairing and base-stacking interactions requires a different position of the glycosidic

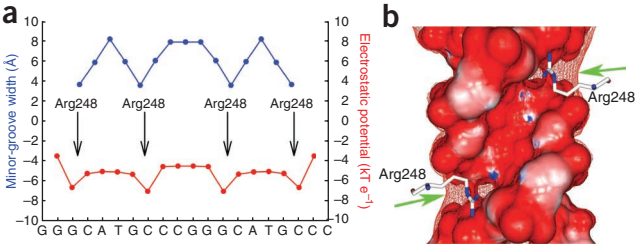

Figure 7 Recognition of DNA shape and electrostatic potential by Arg248 residues in type II complexes. (**a**) The Hoogsteen geometry of the two central A-T base pairs in each half-site leads to narrow minor-groove regions (blue plot) at the ends of the CATG elements. Groove narrowing results in enhanced negative electrostatic potential (red plot) aligned with the binding sites of the Arg248 side chains. (**b**) The electrostatic potential mapped onto the molecular surface of one half-site of the binding site (negative potential in red, positive potential in blue). In addition, a red mesh indicates the isopotential surface at –5 kT e^{-1}. As highlighted by green arrows, the mesh reaches outward the minor groove in the narrow regions, close to the binding sites of the positively charged Arg248 side chains.

bond and the attached backbone than when Watson-Crick base pairs are present (**Fig. 5a–d**). This effect appears to account for the observed minor-groove narrowing, which is then stabilized through interactions with Arg248 side chains.

These findings extend our recent observations that arginines are frequently located at positions of minimal minor-groove width that are associated with the presence of short A tracts[26]. Hoogsteen base pairs apparently offer another means of generating this type of stabilizing interaction, although the source of the narrowing is very different from the one identified previously. What the two mechanisms have in common is the relationship between DNA shape and electrostatic potential. Indeed, a decomposition of the electrostatic potential in the minor groove of type II complexes identifies the geometric arrangement of the phosphate groups as the dominant factor influencing the magnitude of the potential (**Supplementary Fig. 6**).

The presence of local minima in potential where the four Arg248 residues are located suggests that the specific shape induced by the Hoogsteen base-pairing may help to position the p53 monomers on the DNA. In addition to the postulated effects on positioning, it seems clear that Arg248 mutations considerably reduce p53-DNA binding affinities, explaining the frequent observation of Arg248 mutations in human tumors[13].

DISCUSSION

In this paper, we present high-resolution crystal structures of self-assembled complexes between the core domain of human p53 and its consensus DNA targets where the two decameric half-sites are contiguous. In our previous studies, the two half-sites were separated by 2 bp. In both cases, four p53 molecules self-assembled on two B-DNA half-sites to form a tetramer that is a dimer of dimers. However, the three-dimensional architecture of the form reported here, referred to as a type II complex, is distinct from the previously discovered type I form. The differences result from the specific configuration and the particular shape of the DNA binding site. Whereas in type II structures the DNA helix is straight, a large deformation at the junction between half-sites is observed in type I complexes. This deformation brings the attached dimers into close proximity, thereby facilitating interdimer protein-protein interactions that contribute to the stability and cooperativity of the tetrameric complex[19]. However, this interdimer interface is much smaller than that of type II complexes, where the two dimers are parallel to each other due to the contiguous arrangement of the half-sites. Establishing as large an interface in type I complexes as that seen in type II complexes would require a much larger deformation of the DNA at a much greater energy cost, whose magnitude would depend on the intrinsic structure and flexibility of the specific DNA sequence. This provides an explanation of the effect of the DNA spacer in lowering the observed binding affinity and stability of type I complexes in comparison to complexes with a contiguous configuration of the same DNA half-sites[19].

Although the direct contacts between the protein and the DNA are essentially identical in both cases, type II complexes show distinct structural features, of which the most notable is the noncanonical Hoogsteen base-pairing geometry shown by the A-T base pairs at the center of each half-site. The propensity of A-T and G-C base pairs to adopt the Hoogsteen geometry has been shown in other protein–DNA complexes[27–31]. In two complexes, the TATA box bound to TBP[28] and the DNA bound to IHF[27], Hoogsteen base pairs are associated with unusually large DNA deformations. In the case of the replication complex of DNA polymerase ι, the enzyme is 'specialized' for recognizing Hoogsteen base pairs[30,31]. In one of our previously determined type I p53–DNA complexes, two A-T base pairs that are present between the two decameric half-sites, and hence are not involved in DNA binding, showed Hoogsteen geometry[19]. The only example of which we are aware in which a Hoogsteen base pair is observed in undistorted B-DNA is a binding site of the MATα2 homeodomain, where a single A-T base pair has a Hoogsteen geometry[29]. The type II complexes reported here represent the first example, to our knowledge, in which several Hoogsteen base pairs have been observed within a transcription-factor binding site that is part of an essentially undistorted B-DNA helix. Moreover, they are found to have a key role in modulating the DNA shape, and thus the quaternary structure and energetics, of the corresponding complexes.

Hoogsteen base pairs are not frequently observed in B-DNA, although the propensity of A-T base pairs to adopt this geometry has been shown by several crystal structures of free DNA fragments that contain sequences of alternating adenine and thymine bases[32,33]. However, this conformation seemed to be induced by crystal-packing interactions of the free DNA fragments, because NMR studies of $(dA{-}dT)_n$ fragments (n=3,4,5) indicated that the Watson-Crick geometry was the dominant one in solution[32]. Hence, there must be a force driving the A-T base pairs to assume this conformation in their complexes with p53. This force seems to be provided by the large increase in protein-protein interactions among the four core domains of type II tetramers that is enabled by the DNA sequence (no spacer) and conformation formed when the two central base pairs assume a Hoogsteen geometry. Thus, the observed higher affinity and stability of p53 complexes with DNA targets that contain contiguous half-sites relative to those with an inserted 2-bp spacer between the same half-sites[19] seem to be caused by the enhanced protein-protein and protein-DNA interactions observed in type II complexes.

In type II p53–DNA complexes, Hoogsteen base pairing affects the shape of the DNA binding site and, specifically, generates four narrow-minor groove regions that are distributed over the two half-sites. These, in turn, are responsible for the presence of enhanced negative electrostatic potentials, which are recognized by Arg248 side chains. The correlation between DNA shape and minor-groove electrostatics is a newly identified protein-DNA readout mechanism that is used by Hox proteins[34] and other protein families[26,35]. p53 uses a similar mechanism to position its Arg248 side chains and to stabilize its interactions with DNA. However, in contrast to previously described systems, variations in the shape of DNA in the present complexes are driven by the altered base-pairing geometry rather than by sequence-dependent narrowing of the minor groove.

The use of Hoogsteen base pairs in protein-DNA recognition extends the array of base- and shape-recognition mechanisms that are available for the readout of DNA sequence (see review in ref. 35). The extension of the four-letter alphabet of Watson-Crick base pairs by Hoogsteen base pairs is apparently used by p53 in recognizing, in a differential manner, a wide range of DNA-response elements. Based on the crystal structures reported here and on comparisons with other structures described above, the relative predisposition of A-T doublets to adopt the Hoogsteen geometry in the context of CATG, a sequence motif that is abundant in natural p53-response elements[7], is probably higher in DNA targets with contiguous half-sites than in those with noncontiguous half-sites. It is noteworthy that, in a series of binding experiments performed with natural p53 binding sites, several p53-response elements implicated in cell-cycle arrest and DNA repair that contain contiguous half-sites with CATG core elements were among the sites with the highest binding affinity. Conversely, apoptosis-related response elements containing CATG elements or other A/T combinations, but with 1- or 2-bp inserts between half-sites, were

among the sites with the lowest binding affinity[15]. The structural and energetic basis of the CATG preference for the noncanonical base-pairing geometry and the effect of the flanking bases are still unclear. Additionally, it still remains to be seen which pairing schemes are likely to be used by the less common AT-rich doublets, such as AA, TT or TA. It is very likely that the presence of such doublets, or of those mixed with G-C base pairs at the center of each half-site, will lead to diverse DNA shapes and electrostatic patterns, thereby further expanding the repertoire of readout mechanisms used by p53. The growing complexity of p53 function discovered in recent years[36] may thus be associated, at least in part, with the growing diversity in its DNA-recognition modes.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Protein Data Bank: Coordinates and structure factors for the three p53–DNA complexes have been deposited with accession codes 3IGK, 3IGL and 3KZ8.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank our colleagues Y. Halfon, A. Kapitkovsky, A. Eldar and Y. Diskin-Posner for help and the staff at the European Synchrotron Radiation Facility (Grenoble) for facilitating X-ray data collection. The work was supported by grants from the Israel Science Foundation (grant no. 954/08), the Kimmelman Center for Biomolecular Structure and Assembly, the EC (FP6) program, the German-Israeli Foundation for Scientific Research & Development and the Minerva Foundation with funding from the Federal German Ministry of Education and Research (Z.S.) and the US National Institutes of Health grant U54 CA121852 (B.H.). Z.S. holds the Helena Rubinstein Professorial chair in Structural Biology.

AUTHOR CONTRIBUTIONS

M.K., H.R., O.S., D.R. and Z.S. performed the X-ray structure analyses; R.R. performed the computational work; M.K., H.R., R.R., B.H. and Z.S. contributed to the writing of the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Vogelstein, B., Lane, D. & Levine, A.J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
2. Vousden, K.H. & Lu, X. Live or let die: the cell's response to p53. *Nat. Rev. Cancer* **2**, 594–604 (2002).
3. Oren, M. Decision making by p53: life, death and cancer. *Cell Death Differ.* **10**, 431–442 (2003).
4. Laptenko, O. & Prives, C. Transcriptional regulation by p53: one protein, many possibilities. *Cell Death Differ.* **13**, 951–961 (2006).
5. el-Deiry, W.S., Kern, S.E., Pietenpol, J.A., Kinzler, K.W. & Vogelstein, B. Definition of a consensus binding site for p53. *Nat. Genet.* **1**, 45–49 (1992).
6. Funk, W.D., Pak, D.T., Karas, R.H., Wright, W.E. & Shay, J.W. A transcriptionally active DNA binding site for human p53 protein complexes. *Mol. Cell. Biol.* **12**, 2866–2871 (1992).
7. Wei, C.L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).
8. Friedman, P.N., Chen, X., Bargonetti, J. & Prives, C. The p53 protein is an unusually shaped tetramer that binds directly to DNA. *Proc. Natl. Acad. Sci. USA* **90**, 3319–3323 (1993).
9. Weinberg, R.L., Veprintsev, D.B. & Fersht, A.R. Cooperative binding of tetrameric p53 to DNA. *J. Mol. Biol.* **341**, 1145–1159 (2004).
10. Ko, L.J. & Prives, C. p53: puzzle and paradigm. *Genes Dev.* **10**, 1054–1072 (1996).
11. Levine, A.J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).
12. May, P. & May, E. Twenty years of p53 research: structural and functional aspects of the p53 protein. *Oncogene* **18**, 7621–7636 (1999).
13. Olivier, M. *et al.* The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum. Mutat.* **19**, 607–614 (2002).
14. Qian, H., Wang, T., Naumovski, L., Lopez, C.D. & Brachmann, R.K. Groups of p53 target genes involved in specific p53 downstream effects cluster into different classes of DNA binding sites. *Oncogene* **21**, 7901–7911 (2002).
15. Weinberg, R.L., Veprintsev, D.B., Bycroft, M. & Fersht, A.R. Comparative binding of p53 to its promoter and DNA recognition elements. *J. Mol. Biol.* **348**, 589–596 (2005).
16. Tokino, T. *et al.* p53 tagged sites from human genomic DNA. *Hum. Mol. Genet.* **3**, 1537–1542 (1994).
17. Joerger, A.C. & Fersht, A.R. Structural biology of the tumor suppressor p53. *Annu. Rev. Biochem.* **77**, 557–582 (2008).
18. Cho, Y., Gorina, S., Jeffrey, P.D. & Pavletich, N.P. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* **265**, 346–355 (1994).
19. Kitayner, M. *et al.* Structural basis of DNA recognition by p53 tetramers. *Mol. Cell* **22**, 741–753 (2006).
20. Ho, W.C., Fitzgerald, M.X. & Marmorstein, R. Structure of the p53 core domain dimer bound to DNA. *J. Biol. Chem.* **281**, 20494–20502 (2006).
21. Malecka, K.A., Ho, W.C. & Marmorstein, R. Crystal structure of a p53 core tetramer bound to DNA. *Oncogene* **28**, 325–333 (2009).
22. Chen, Y., Dey, R. & Chen, L. Crystal structure of the p53 core domain bound to a full consensus site as a self-assembled tetramer. *Structure* **18**, 246–256 (2010).
23. Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **16**, 907–916 (1963).
24. Suad, O. *et al.* Structural basis of restoring sequence-specific DNA binding and transactivation to mutant p53 by suppressor mutations. *J. Mol. Biol.* **385**, 249–265 (2009).
25. Dehner, A. *et al.* Cooperative binding of p53 to DNA: regulation by protein-protein interactions through a double salt bridge. *Angew. Chem. Int. Edn Engl.* **44**, 5247–5251 (2005).
26. Rohs, R. *et al.* The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248–1253 (2009).
27. Rice, P.A., Yang, S., Mizuuchi, K. & Nash, H.A. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295–1306 (1996).
28. Patikoglou, G.A. *et al.* TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13**, 3217–3230 (1999).
29. Aishima, J. *et al.* A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res.* **30**, 5244–5252 (2002).
30. Nair, D.T., Johnson, R.E., Prakash, S., Prakash, L. & Aggarwal, A.K. Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing. *Nature* **430**, 377–380 (2004).
31. Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. & Aggarwal, A.K. Human DNA polymerase iota incorporates dCTP opposite template Gvia a G.C + Hoogsteen base pair. *Structure* **13**, 1569–1577 (2005).
32. Abrescia, N.G., Gonzalez, C., Gouyette, C. & Subirana, J.A. X-ray and NMR studies of the DNA oligomer d(ATATAT): Hoogsteen base pairing in duplex DNA. *Biochemistry* **43**, 4092–4100 (2004).
33. De Luchi, D., Tereshko, V., Gouyette, C. & Subirana, J.A. Structure of the DNA coiled coil formed by d(CGATATATATAT). *ChemBioChem* **7**, 585–587 (2006).
34. Joshi, R. *et al.* Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* **131**, 530–543 (2007).
35. Rohs, R. *et al.* Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* (in the press) (2010).
36. Vousden, K.H. & Prives, C. Blinded by the light: the growing complexity of p53. *Cell* **137**, 413–431 (2009).
37. Lavery, R. & Sklenar, H. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* **6**, 655–667 (1989).

ONLINE METHODS

Production and purification of protein and DNA samples. We produced human p53 core domain or DNA binding domain (referred to as p53DBD) and a double mutant, containing a hot-spot mutation R249S and second-site suppressor mutation H168R (referred to as p53DBD-RS/HR), in *Escherichia coli* and purified them as described previously[19,24]. We synthesized DNA oligonucleotides of the sequences 5′-cGGGCATGCCCg-3′ (referred to as 12-mer) and 5′-tGGGCATGCCCGGGCATGCCC-3′ (referred to as 21-mer), containing one and two consensus half-sites, respectively, in their duplex forms, at the Weizmann Institute DNA synthesis unit and purified them by ion-exchange chromatography. We then dialyzed them against water, lyophilized them and finally dissolved them in water to obtain a final concentration of 10 mg ml^{-1}.

Crystallization, data collection and structure determination. We mixed p53DBD (5.4 mg ml^{-1}) and p53DBD-RS/HR (6.7 mg ml^{-1}) each with DNA solutions using a range of molar ratios of protein/single-stranded DNA from 1.2 to 2.5. We performed crystallization experiments at 19 °C by the hanging-drop vapor-diffusion technique, using 2 µl protein-DNA solution and 2 µl reservoir solution equilibrated against 0.5 ml reservoir solution. We used crystallization screen kits PEG/Ion I and II (Hampton Research Inc.) for initial screening. We obtained cocrystals of p53DBD or p53DBD-RS/HR with the 12-mer under identical conditions: protein/DNA ratio of 2.4 and reservoir solution of 160 mM sodium acetate and 16% (v/v) PEG 3350. We obtained cocrystals of p53DBD with the 21-mer by using a protein/DNA ratio of 1.2 and a reservoir solution containing 200 mM ammonium iodide and 20% (v/v) PEG 3350. Prior to data collection, we transferred the crystals from their mother liquor either to Paratone-N oil or in a stepwise manner to solutions of the reservoir, with increasing concentration of ethylene glycol or glycerol up to 20% (v/v) used as cryoprotectants.

We measured X-ray diffraction data from crystals of p53DBD–12-mer (complex 1) and p53DBD-RS/HR–12-mer (complex 2), flash-cooled at 100 K using an Oxford Cryostream system, on a Rigaku R-AXIS IV++ image-plate detector mounted on a Rigaku RU-H3R rotating anode generator with Cu K_α radiation focused by Osmic multilayer confocal mirrors. We collected X-ray data from a crystal of p53DBD–21-mer at the European Synchrotron Radiation Facility beamline ID23-1 at a wavelength of 0.9762 Å. We processed the data using the HKL2000 package[38]. Crystal data and intensity statistics are given in **Table 1**.

Crystals of the complexes with the 12-mer DNA (complexes 1 and 2; space group *C*2) were isomorphous, indicating the presence of one p53 molecule and one single-stranded DNA in the asymmetric unit. The space group (*C*2) and the unit-cell dimensions of the third complex (with the 21-mer DNA) were similar to that of the other two except for the *c* axis, which was twice as long. The intensities of *hkl* reflections with $l = 2n + 1$ were much weaker than the other reflections, indicating a pseudotranslational symmetry with $c' = 1/2c$ as was later confirmed by the analysis (see below).

We solved the structures of complexes 1 and 2 by the molecular replacement approach as implemented in the program CNS[39,40], using molecule A of the crystal structure of the core domain of human p53 (ref. 18) (PDB 1TSR) as a search model (protein atoms only). Electron density maps showed the presence of one DNA strand and one p53 molecule in the asymmetric unit that form, via the crystallographic two-fold axes, a p53 tetramer bound to two double-helical decamers, with extrahelical bases at each 3′ end interacting with adjacent molecules in the crystal.

We solved the structure of p53DBD–21mer with MolRep[41] from the CCP4 package[42] using the same search model (a single p53 molecule). The asymmetric unit contained two protein molecules related by a pseudotranslation along the *c* axis. Initial electron density maps showed the presence of a single DNA strand in the asymmetric unit, forming a p53 tetramer bound to a 20-bp DNA via the crystallographic two-fold axis. The overhung 5′ thymine base could not be located.

We conducted refinement of the three crystal structures in a similar manner. We first subjected the structure to a low-resolution refinement (typically 15–3.5 Å), with each protein molecule being treated first as a rigid body. This was followed by simulated annealing with a slow-cooling procedure (starting temperature of $T = 4{,}000$ K) continuing with several cycles of minimization and individual temperature-factor refinement (CNS package[39,40]) in order to remove any bias from the starting model. We further refined the structure using REFMAC5[43] and Coot[44]. We then traced the DNA oligonucleotides from the electron density maps and added water molecules. We traced multiple conformations of several residues. We did not include in the final models disordered protein and DNA residues that could not be modeled in the electron density maps. We used the program PROCHECK[45] for structure analysis and validation. Refinement statistics are given in **Table 1**. All the structure-based figures were drawn with PyMol[46].

Computational methods. We analyzed DNA structures using a modified version of Curves[37], adopted for the inclusion of Hoogsteen base pairs. To build a helix axis regardless of the orientation of a base in a *syn* or *anti* conformation, the program identifies the conformation and rotates the base-fixed coordinate system so that the local helix axis points in the same direction for both the *syn* and *anti* conformations. We symmetrized structural parameters calculated with our in-house version of Curves for type II complexes based on the palindromic DNA sequence and crystal-packing symmetry.

We calculated electrostatic potentials of the DNA with DelPhi[47] by solving the nonlinear Poisson-Boltzmann equation at physiological ionic strength (0.145 M). We took atomic radii and partial charges from the Amber force field[48]. The calculations used a 1.4-Å probe radius, interior dielectric constant of 2, solvent dielectric constant of 80, Debye-Hückel boundary conditions with five focusing steps and a cubic grid size of 165. We derived the electrostatic potential as a function of base sequence based on geometric reference points, which are located for nucleotide *i* in the midpoint between the phosphate atoms of nucleotide $i + 1$ in strand 1 and nucleotide $i - 1$ in strand 2. The definition of the reference point is slightly different than that in previous work[34] so as to prevent clashes of this point with guanine amino groups in the minor groove. The location of the arrows that indicate the Arg248 positions in the electrostatic potential plots is derived from contacts of the guanidinium groups with backbone atoms located between two base pairs. Electrostatic potentials mapped on the molecular surface of the DNA binding site are shown with GRASP2 (ref. 49).

38. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
39. Brunger, A.T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
40. Brunger, A.T. Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733 (2007).
41. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
42. Collaborative Computational Project. Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
43. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
45. Laskowski, R.A., Macarthur, M.W., Moss, D.S. & Thornton, J.M. Procheck—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
46. DeLano, W.L. The PyMOL Molecular Graphics System (San Carlos, California, USA, 2002).
47. Rocchia, W. *et al.* Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **23**, 128–137 (2002).
48. Cornell, W.D. *et al.* A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
49. Petrey, D. & Honig, B. GRASP2: visualization, surface properties, and electrostatics of marcomolecular structures and sequences. *Methods Enzymol.* **374**, 492–509 (2003).

doi:10.1038/nsmb.1800