

SUPPLEMENTARY INFORMATION

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

Malka Kitayner^{1,3}, Haim Rozenberg^{1,3}, Remo Rohs^{2,3}, Oded Suad¹,
Dov Rabinovich^{1,4}, Barry Honig^{2,5} and Zippora Shakked^{1,5}

¹Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

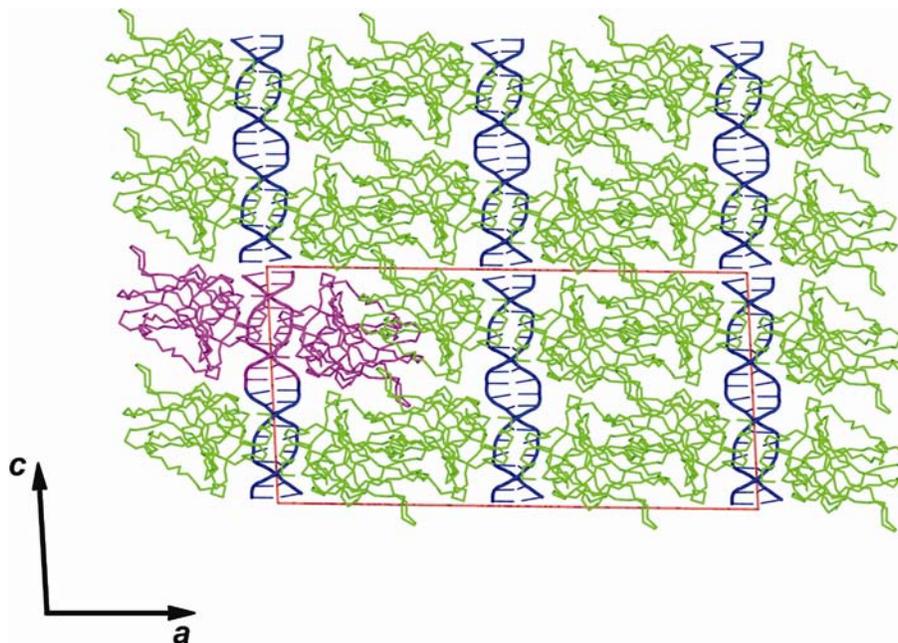
²Howard Hughes Medical Institute, Center for Computational Biology and
Bioinformatics and Department of Biochemistry and Molecular Biophysics, Columbia
University, New York, NY, USA

³These authors contributed equally to this work.

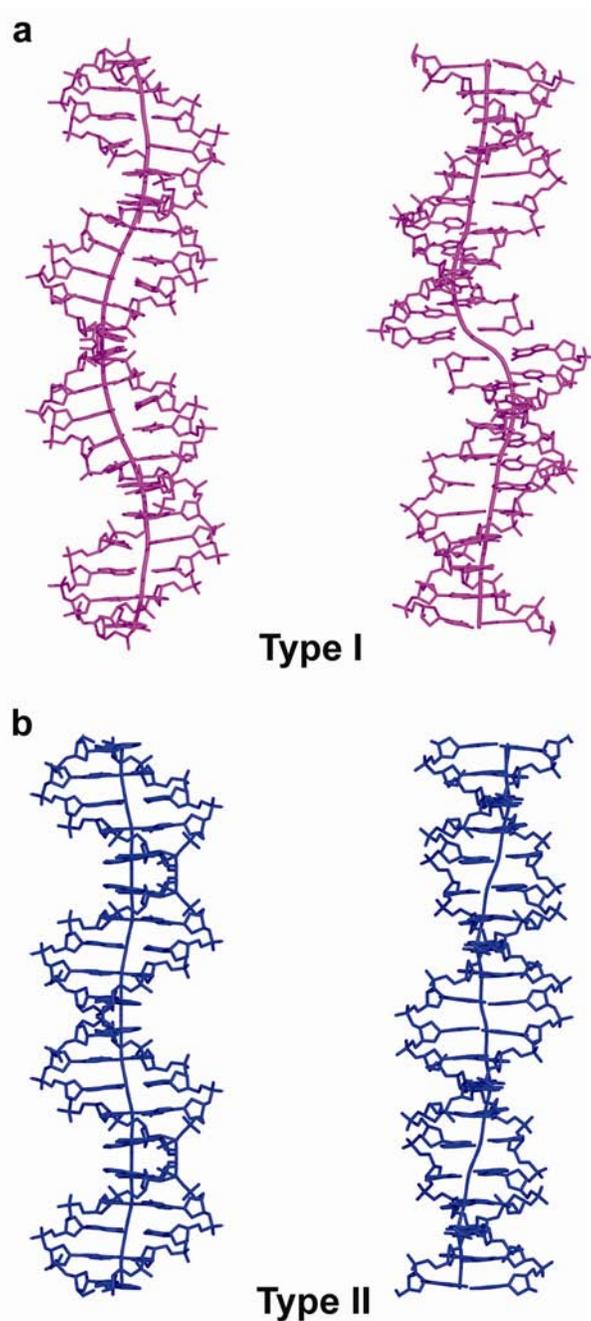
⁴Deceased 6 September 2008

⁵Correspondence should be addressed to:

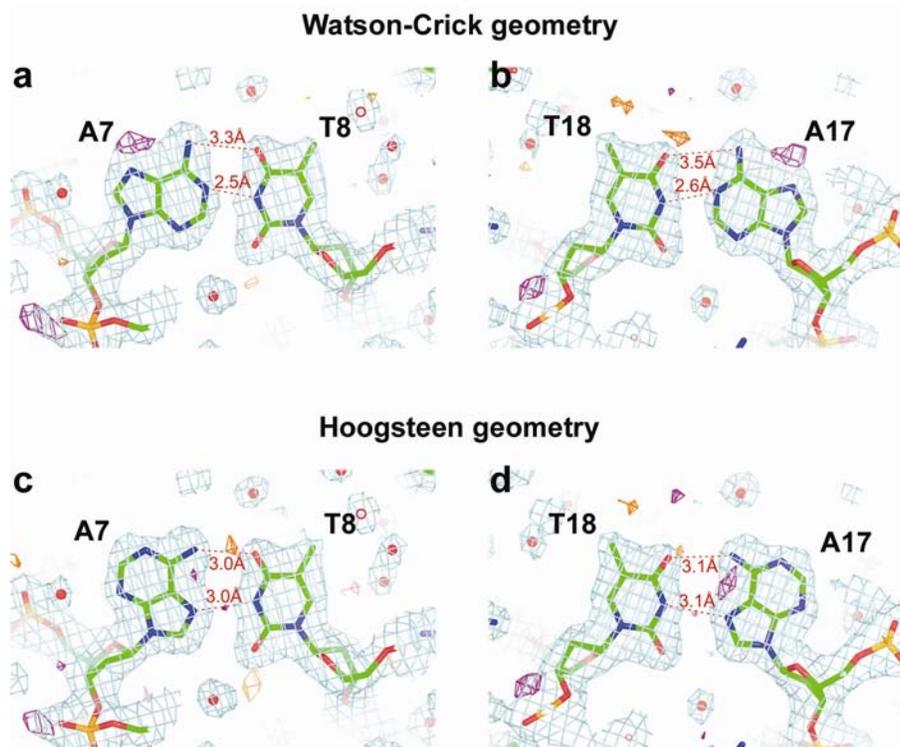
B.H. bh6@columbia.edu or Z.S. zippi.shakked@weizmann.ac.il



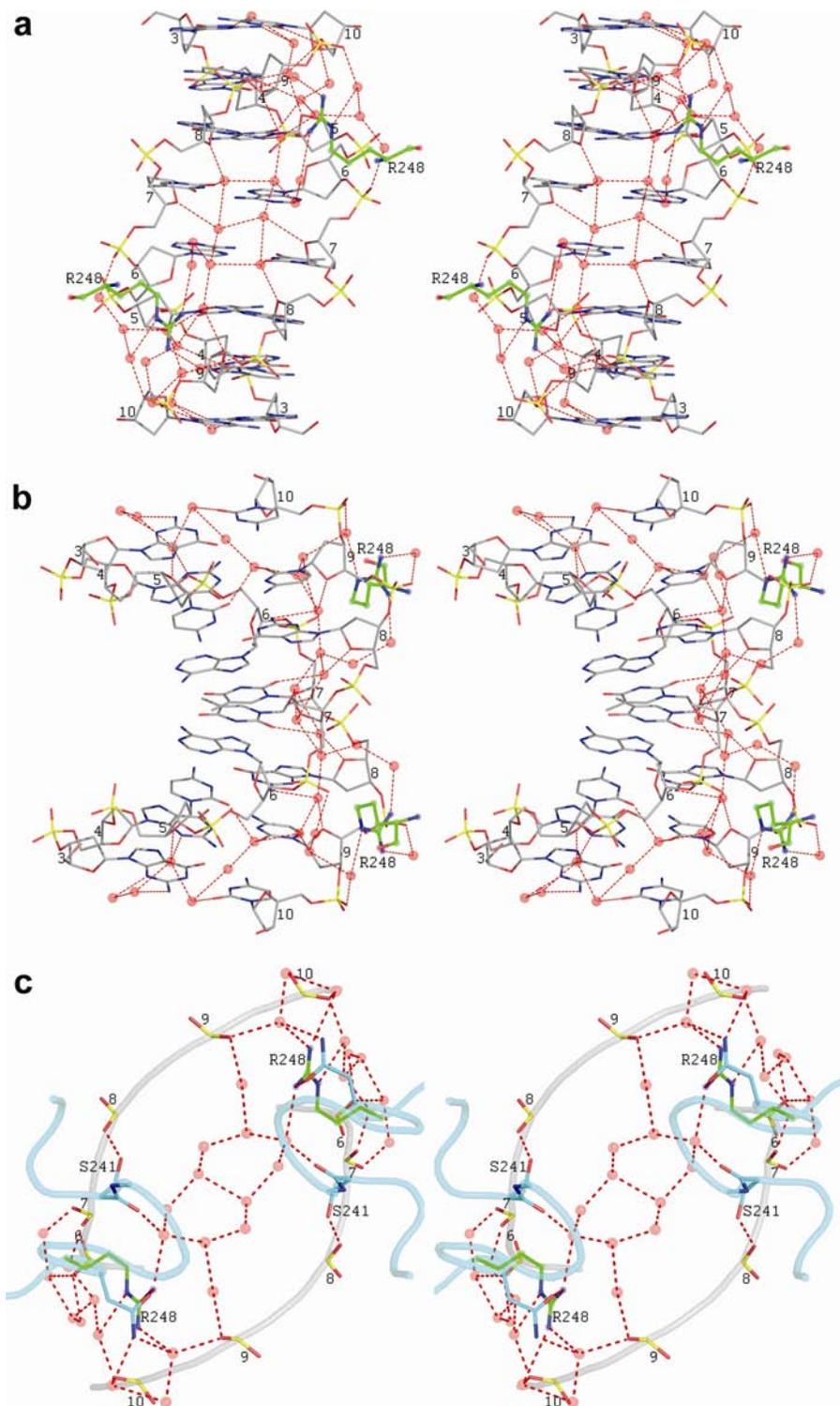
Supplementary Figure 1. Crystal packing of type II p53-DNA complexes. View along the crystallographic *b* axis (based on complex 3). The protein is shown in green, the DNA in blue. The asymmetric unit containing a p53 dimer and a double-stranded DNA half-site is shown in magenta and the unit-cell axes in red. The view highlights the pseudo-translational symmetry along the *c* axis. The packing of complex 3 is very similar to that of complexes 1 and 2 where the asymmetric unit contains a p53 monomer and a single-stranded DNA half-site (see Methods for detailed description).



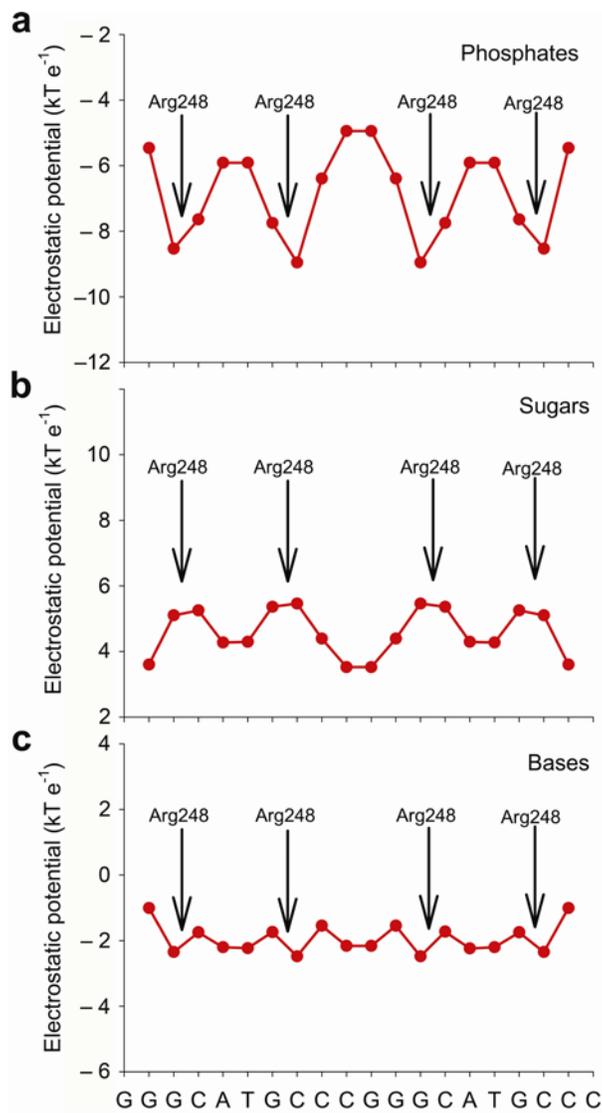
Supplementary Figure 3. DNA conformations in type I and type II complexes. The decameric half-sites are separated by two base pairs in type I complexes (**a**) and are contiguous in type II complexes (**b**). The global helix axis was calculated by Curves¹ (see Methods) for type I complex² and by an in-house version of Curves adapted for Hoogsteen base pairs for the type II complex. The view highlights the large deformation at the junction between half-sites in type I helix.



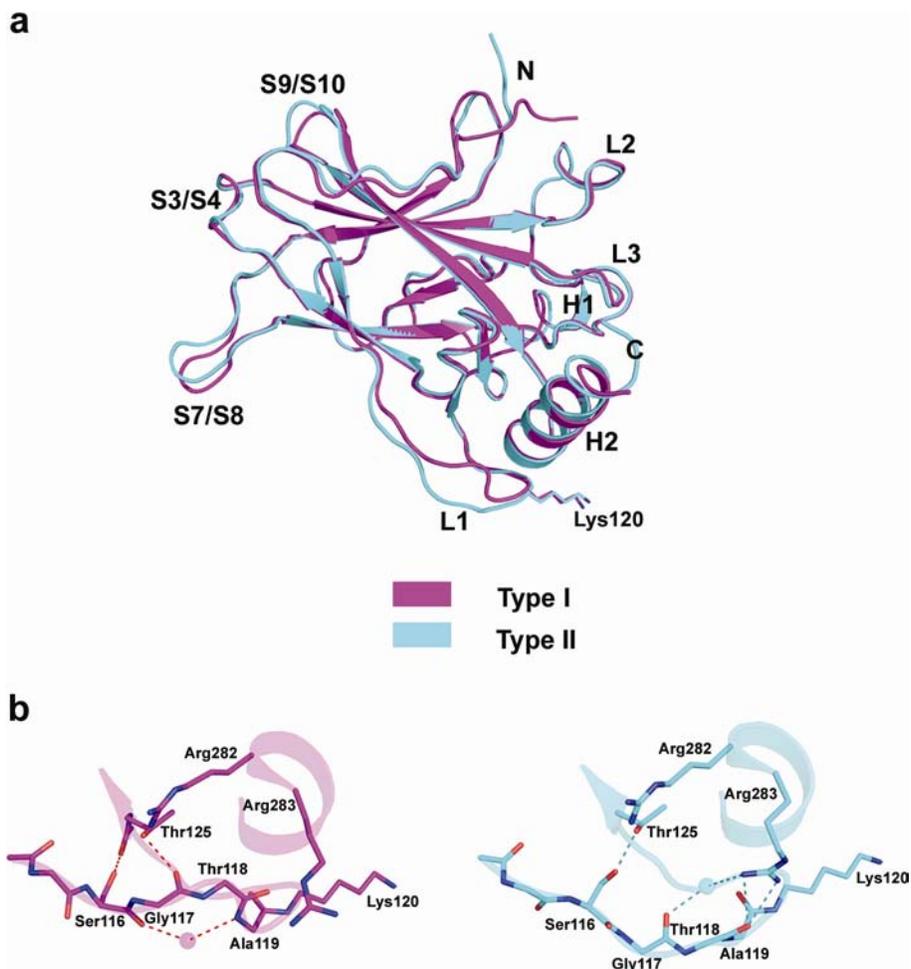
Supplementary Figure 4. Comparison between Watson-Crick (**a,b**) and Hoogsteen base pairs (**c,d**) of the two refined models of the mouse p53 tetramer cross-linked to DNA binding site³. Electron density map, $2F_o - F_c$ at 1σ level, is shown in cyan. Difference map, $F_o - F_c$ at 3σ level, is shown in magenta (positive) and in red (negative). See detailed description in Supplementary Methods.



Supplementary Figure 5. Minor-groove hydration and interactions with Arg248 side chains. **(a)** View down to the minor groove of the DNA half-site and **(b)** View perpendicular to the first one. Only the central eight base pairs of the half-site are shown (GGCATGCC). Nucleotide numbering (5'-3' direction) is from 3 to 10 for each strand. Arg248 residues (shown in green) interact with the minor-groove hydration of the DNA represented by red spheres and dotted lines. Only the first hydration shell and some water molecules from the second hydration shell around Arg248 residues are shown. **(c)** Phosphate backbone of five nucleotides of each strand and protein backbone (residues 240-248) are shown in gray and cyan, respectively. Amino acids Ser241, Asn247 (in cyan) and Arg248 (in green) are shown. Nucleotide numbering (5'-3' direction) is from 6 to 10 for each strand. Only the second hydration shell and some water molecules from the first hydration shell near the phosphate backbone are shown.



Supplementary Figure 6. Decomposition of the electrostatic potential in the minor groove. The electrostatic potential of DNA is caused by charged atoms of **(a)** the phosphates, **(b)** the sugar moieties, and **(c)** the bases. Based on the linear Poisson-Boltzmann equation, the effect of these different chemical entities on the overall potential is estimated. As illustrated in the figure, the phosphates are the main origin of the enhanced negative electrostatic potential at the Arg248 binding sites.



Supplementary Figure 7. Conformation of the L1 recognition loop. **(a)** Superposition of p53 core domains from type I and type II complexes, in magenta (based on PDB ID 2AC0²) and in cyan (based on complex 1), respectively, showing the different conformations of the L1 loops from the two complexes. α -helices and loops are labeled. **(b)** Close-up views of the different stabilizing interactions in each complex, showing the side chains of residues Ser116, Lys120, Thr125, Arg282 and Arg283 and the backbone of residues 114–120 and 124. Water atoms are shown as transparent spheres.

Supplementary Table 1 Direct inter-dimer polar and charged interactions

2d element of 1 st molecule	Contacts				2d element of 2 nd molecule	Distance (Å)
S5/S6	OE2	Glu198	→	OG1 Thr170	L2	2.6
S5/S6	O	Gly199	→	N Ser96	N-ter	2.9
S5/S6	OD1	Asn200	→	N Ser94	N-ter	3.3
S5/S6	N	Leu201	→	O Ser94	N-ter	2.9
S7/S8	OE2	Glu221	→	N Ser95	N-ter	3.1
S7/S8	OE1	Glu224	→	OG Ser99	N-ter	2.7
S7/S8	OE2	Glu224	→	NH1 Arg267	S10	3.5

Supplementary Discussion

A new conformational variant of the L1 recognition loop

The reported crystal structures of the core domain in its free and DNA-bound forms have shown that, upon binding, the L1 loop undergoes a conformational change so as to allow the formation of hydrogen-bond interactions between Lys120 side chains and the second and third base pairs from each end of the decameric half-sites^{2,4-6}. The specific pattern of contacts made by Lys120 is sequence dependent and accommodated by adjustments in the lysine side chain². In several cases, mainly in the free state, the L1 loop was found to be disordered reflecting its high flexibility^{7,8}. In the p53 dimers and tetramers that were cross-linked to their DNA targets, this loop was shown to be either disordered or to display a dramatically different conformation in comparison to self-assembled tetrameric p53–DNA complexes. This feature is probably a result of steric hindrance caused by the covalent linkers between the two molecules³.

In type II complexes, however, the L1 loop is well defined exhibiting a new conformational variant in complexes 1 and 2 whereas complex 3 shows the previously observed L1 conformation². A significant change between the two conformations is observed at the five amino acid region 115–119 that includes a central glycine residue (HSGTA) whereas the L1 regions flanking this pentamer are essentially identical and similarly stabilized by direct and water-mediated hydrogen bonds. Comparisons between the DNA-bound L1 conformations and the corresponding intra-molecular interactions are shown in **Supplementary Fig. 7**.

The role of the L1 loop and its highly conserved K120 residue in recognizing different response elements is uncertain in view of several recent studies. Site-directed mutagenesis studies of L1^{9,10} suggest that the flexibility of this loop and hence its effect on p53 function can be significantly altered by modifications in its sequence. Other studies have shown that K120 acetylation is essential for p53-mediated activation of proapoptotic target genes^{11,12}. Computational analysis indicated that the energy required for desolvation of lysine side chains residues is greater than that of arginines, explaining why lysines are less likely to engage in protein-DNA base contacts compared to arginines¹³. The role of the single lysine residue at the p53-DNA interface may well depend on the specific DNA response element and/or protein co-factors involved in the particular signaling event.

Supplementary Methods

Revised analysis of the published crystal structure of the mouse p53-DNA complex

We reanalyzed the crystal structure of the mouse p53 core-domain tetramer covalently cross-linked to two 23-mer DNA strands from Malecka *et al.*³ using the deposited coordinates and structure factors (PDB code 3EXJ).

In order to check the original refinement, we first used the CNS package^{14,15} and calculated a sigmaa composite annealed omit map using the deposited data. Then, we submitted the model to refinement with CNS including simulated annealing with slow-cooling protocol starting from 4000°C, followed by energy minimization steps, individual *B*-factor minimization and electron density maps calculation (sigma $2mF_o - DF_c$ and $mF_o - DF_c$), without NCS restraints between the two protein molecules in the asymmetric unit. In accord with the maps from the Electron Density Server of Uppsala University (<http://eds.bmc.uu.se/eds/>)¹⁶, the three CNS maps exhibited several errors in the deposited model. Moreover, the composite omit map revealed that the adenine bases of the central A-T doublet of the consensus half-site sequence (GAGCATGCTC) could be better fitted into the map with a *syn* conformation (rotation of 180° around the glycosidic bond) which is compatible with a Hoogsteen geometry. We therefore decided to independently solve and refine the structure. We solved the structure with Phaser¹⁷ using the deposited structure factors and molecule A from the high-resolution structure of

mouse p53 core domain (PDB code 2I0I)¹⁸ as a search model. We then refined the model following the protocol used for the human p53–DNA complexes as described above. In particular, we traced manually the DNA, nucleotide by nucleotide, into the electron density maps. We further refined the whole model with the exclusion of the central two A-T base pairs. At the last stage of the refinement, we refined two independent models, one with Watson-Crick (WC) geometry and the other with Hoogsteen geometry for the central A-T base pairs of each half-site. No geometrical constraints between the bases were used. The final refinement statistics (resolution range 28.7–2.0 Å and the inclusion of 420 water molecules) was similar for the two models with R_{work} and R_{free} being 19 and 26% respectively, and r.m.s. deviations of 0.02 Å and 2° for bond length and bond angles, respectively. The final electron density maps of the A-T base pairs calculated for the two models are shown in **Supplementary Fig. 4**.

The hydrogen-bond distances between the adenine and thymine bases in the WC model differ by more than 0.8 Å from each other as a result of a deviation of the adenine base from a regular WC geometry, whereas the values for the Hoogsteen base pairs are similar to each other. As a result, the relative position of the glycosidic bonds and the attached sugar rings (based on N1-N9 and C1'-C1' distances across strands) in the WC model are on average shorter by 1 Å than the expected values observed in the high-resolution crystal structures of type I complexes² (7.9 and 9.0 Å compared to 8.7 and 10.2 Å, respectively), whereas the corresponding distances in the Hoogsteen model are on average only 0.4 Å longer than the expected values based on the current crystal structures (7.0 and 8.8 Å compared to 6.7 and 8.3 Å, respectively). Also, the electron density maps ($2F_o - F_c$ and $F_o - F_c$) indicate a better fitting to the Hoogsteen model than to the WC one (**Supplementary Fig. 4**). On the basis of the electron density maps as well as the geometry of the two alternative base pairings, we conclude that the dominant form adopted by the central A-T base pairs in this crystal structure is of the Hoogsteen geometry.

The crystal structure of the second published complex of the same components (PDB code 3EXL) could not be used for a revised analysis as the authors used a pseudo unit-cell dimension with $c' = 34$ Å instead of the correct unit cell, $c = 68$ Å, as expected from the packing of 20 base-pair B-DNA helices stacked end-to-end along the c axis.

Supplementary References

1. Lavery, R. & Sklenar, H. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* **6**, 655-67 (1989).
2. Kitayner, M. *et al.* Structural basis of DNA recognition by p53 tetramers. *Mol. Cell* **22**, 741-53 (2006).
3. Malecka, K.A., Ho, W.C. & Marmorstein, R. Crystal structure of a p53 core tetramer bound to DNA. *Oncogene* **28**, 325-33 (2009).
4. 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).
5. Zhao, K., Chai, X., Johnston, K., Clements, A. & Marmorstein, R. Crystal structure of the mouse p53 core DNA-binding domain at 2.7 Å resolution. *J. Biol. Chem.* **276**, 12120-12127. (2001).
6. Joerger, A.C., Allen, M.D. & Fersht, A.R. Crystal structure of a superstable mutant of human p53 core domain. Insights into the mechanism of rescuing oncogenic mutations. *J. Biol. Chem.* **279**, 1291-6 (2004).
7. Joerger, A.C., Ang, H.C., Veprintsev, D.B., Blair, C.M. & Fersht, A.R. Structures of p53 cancer mutants and mechanism of rescue by second-site suppressor mutations. *J. Biol. Chem.* **280**, 16030-7 (2005).
8. Suad, O. *et al.* Structural basis of restoring sequence-specific DNA binding and transactivation to mutant p53 by suppressor mutations. *J. Mol. Biol.* **385**, 249-65 (2009).
9. Resnick, M.A. & Inga, A. Functional mutants of the sequence-specific transcription factor p53 and implications for master genes of diversity. *Proc. Natl. Acad. Sci. USA* **100**, 9934-9939 (2003).
10. Zupnick, A. & Prives, C. Mutational analysis of the p53 core domain L1 loop. *J. Biol. Chem.* **281**, 20464-73 (2006).
11. Sykes, S.M. *et al.* Acetylation of the p53 DNA-binding domain regulates apoptosis induction. *Mol. Cell* **24**, 841-51 (2006).
12. Tang, Y., Luo, J., Zhang, W. & Gu, W. Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. *Mol. Cell* **24**, 827-39 (2006).
13. Rohs, R. *et al.* The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248-53 (2009).
14. 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).
15. Brunger, A.T. Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2**, 2728-2733 (2007).
16. Kleywegt, G.J. *et al.* The Uppsala Electron-Density Server. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2240-2249 (2004).
17. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).

18. Ho, W.C. *et al.* High-resolution structure of the p53 core domain: implications for binding small-molecule stabilizing compounds. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1484-93 (2006).