Falling out of the fold: tumorigenic mutations and p53

New structural information on p53, the protein product of a tumor suppressor gene, gives insight into how mutations in the gene lead to loss of protein function.

Chemistry & Biology October 1994, **1**:79–84

Cells grow and eventually divide through a regimented process termed the cell cycle, which is tightly controlled by cues from the extracellular and intracellular environments [1]. The cell cycle is regulated at checkpoints, progression through which can be interrupted when cells fail to meet certain fitness criteria. For example, cells that have experienced global DNA damage fail to progress through a checkpoint that precedes initiation of DNA synthesis; this temporary blockade gives the cell time to repair the genome before replication. When the damage is too severe, the cells undergo programmed self-destruction through a process termed apoptosis. Central to the cellular response to DNA damage is the protein p53 (see [2,3] for review). Inactivation of p53 removes an important restraint on cell growth and thus predisposes cells toward becoming cancerous; *p53* is therefore known as a tumor suppressor gene (see [4,5] for reviews).

Wild-type p53 is a sequence-specific DNA-binding protein that functions as both a transcriptional activator and a repressor [2,3]. Mutations in the *p53* genes that are found in tumor cells invariably correlate with loss of these activities in the mutant protein. Mutant *p53* genes have been sequenced from roughly 2 000 clinical cancers, with the vast majority of the mutations mapping to a region of the protein identified as the sequence-specific DNA-binding domain (Fig. 1a) The mutations are not randomly distributed throughout this domain, but cluster into several 'hotspots', the locations of which correspond roughly to regions of the protein sequence that are conserved across species. So far, much of the biochemical characterization of the mutant proteins has relied on their binding to monoclonal antibodies. Many of the mutant proteins fail to bind antibodies specific for the folded form of p53, but instead bind antibodies specific for denatured p53. Together with

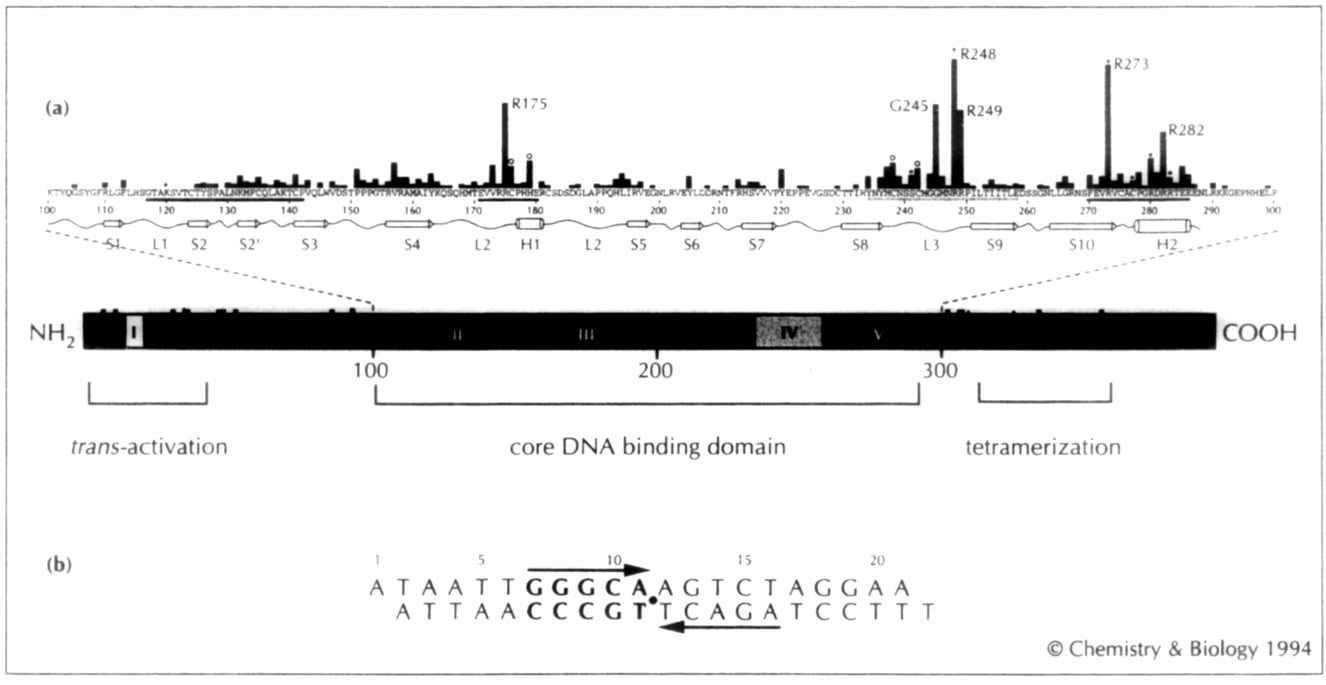

Fig. 1. The p53 protein: domain organization, conserved regions, sequence and secondary structure of the core DNA-binding domain, and DNA half-site. **(a)** Schematic representation of the p53 protein, denoting the five regions (I–V) that are highly conserved evolutionarily and (underneath) the known functional domains; colors of conserved regions II–V correspond to those in Fig. 2. Expansion above the core DNA-binding domain shows the amino-acid sequence in single letter code; diagram directly underneath the sequence illustrates the secondary structure and designations according to Cho *et al.* [8]. Vertical bars denote the frequency of mutations at a particular position in primary human cancer isolates; bars above the schematic sequence represent mutations outside the core DNA-binding domain, and bars above the explicit sequence represent mutations within the core DNA-binding domain; all are scaled proportionally. The fragment structurally characterized by Cho et al. [8] comprises amino acid residues 94–312. **(b)** Sequence of the oligonucleotide used by Cho *et al.* [8] in co-crystallization and structure determination of the p53 core domain. Arrows refer to orientation of consensus pentanucleotide binding sites, of which only one (bold) is occupied by p53 in the protein–DNA co-complex. Numbering of base-pairs corresponds to that referred to in the text.

other factors, including the identification of temperature-sensitive mutants, this evidence gave rise to the hypothesis that some mutant proteins may have a 'mutant conformation', a well-ordered fold that differs from that of the native protein [6,7]. This concept has led many groups to attempt to identify small molecules that would cause the 'mutant conformation' of p53 to revert to that of the wild-type, thus restoring the function of some p53 mutants (for review, see [4]).

Despite impressive advances in understanding the biochemistry of p53, very little was known about the structure of the protein or its specific interactions with DNA. Recently, this situation changed dramatically; Cho *et al.* [8] reported the X-ray structure of the core DNA-binding domain of p53 bound to an oligodeoxynucleotide, and Clore *et al.* [9] reported the NMR structure of the tetramerization domain of p53. These studies give important insights into the relationship between p53 structure and function, and especially into the basis of mutational inactivation; they also have important implications for future drug design efforts. Here we focus on how the p53 core domain recognizes DNA, and how this is affected by tumorigenic mutations.

Native p53 exists as a tetramer in solution and binds to DNA sites that have four copies of the consensus sequence 5'-d[RRRC(A/T)]-3' (R = A or G) arranged in alternating orientation, to give a palindrome of palindromes (→•←X→•←; • represents the outer dyad axes and X is a spacer of variable length) [2,3]. Cho *et al.* [8] crystallized and solved the structure of a monomeric form

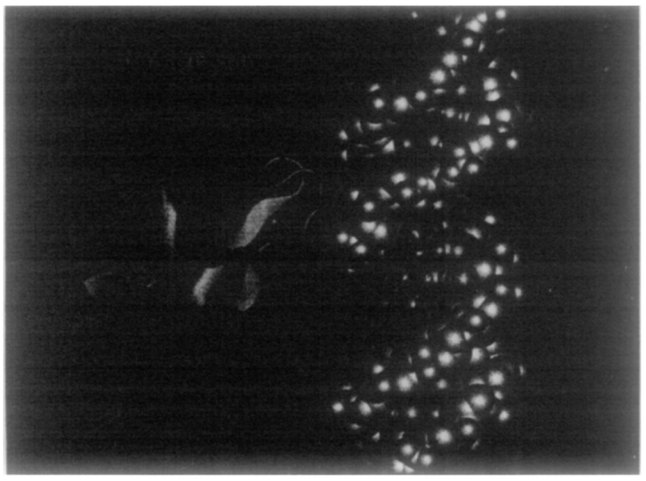

Fig. 2. Backbone ribbon trace of the p53 core domain structure bound to a 21-bp oligonucleotide. Of the three protamers in the unit cell, only the one bound specifically to an intact pentanucleotide site (bold in Fig. 1b) is shown. The DNA is shown in a van der Waals representation, with the deoxyribose phosphate backbone in gray and the bases in blue. The conserved regions of the protein are colored as noted, and the remainder of the protein is colored green. Several features of secondary structure that make up the protein–DNA interface are denoted (see Fig. 1a), with the lettering color-coded to match that of the conserved region of which that feature forms a part.

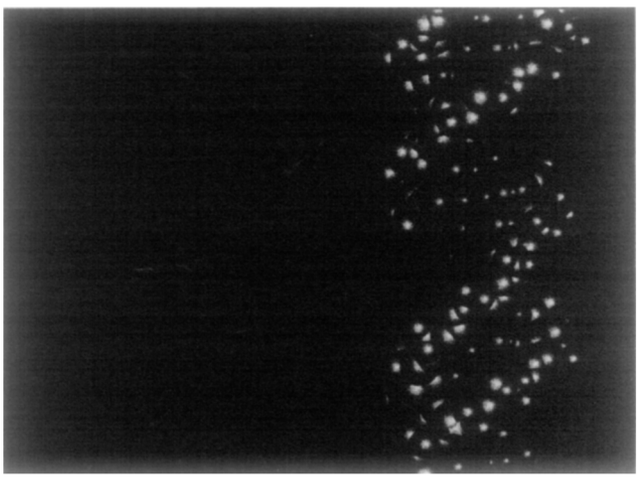

Fig. 3. The loop–sheet–helix motif employed in DNA binding by the core domain of p53. The loop–sheet–helix motif is shown in purple, and the remainder of the protein structure is in green. Secondary structure features are denoted according to Fig. 1a.

of p53 in complex with a duplex 21-base-pair oligomer containing a half-site (→•←; Fig. 1b). Interestingly, only one of the two pentanucleotide sub-sites is bound to p53 in the complex (Fig. 2), suggesting that any possible interaction between monomers is not strong enough to enforce crystallization as a DNA-bound dimer.

The overall structure

The bulk of the p53 core DNA-binding domain is a nine-stranded antiparallel β sandwich decorated with loops. One strand (S10) extends from the end of the β sandwich, where it pairs with a β hairpin (S2 and S2') to form a three-stranded β sheet; this serves as a platform for an α helix (H2), which packs at its amino-terminal end against a long loop (L1) that interconnects the β sandwich and the β sheet (Fig. 3). The module comprising the β sheet, H2, and L1, dubbed the loop–sheet–helix motif by Cho *et al.* [8], contains most of the residues that contact DNA directly. Another prominent feature of the structure is its zinc-binding site. Two zinc ligands are contributed by a short α helix (H1), the other two by a large loop (L3) projecting from the β sandwich (Figs 2 and 4); L3 also contains the two DNA-contact residues that lie outside the loop–sheet–helix motif.

The four conserved sequences within the core domain are shown in Fig. 2. Interestingly, all four highly conserved segments of the core domain are either directly involved in DNA binding (II, IV, V) or serve as a scaffold for DNA-binding elements (II–V). Thus, there is strong selective pressure against mutations that alter the DNA-binding surface of p53. One explanation for this might be that the surfaces of the protein and DNA interact so perfectly that the interface cannot tolerate even conservative changes; however, this interpretation is inconsistent with the fact that p53 binds to a wide variety of DNA sequences. It seems more likely that the DNA-binding surface of p53 is formed by highly cooperative interactions and that the folded structure has only

Fig. 4. Zinc-binding module of the p53 core domain. Residues that ligate zinc are shown in licorice bonds, with Cys ligands in yellow, His in blue. The steel sphere represents Zn^{2+}.

marginal stability. Thus, even conservative amino-acid changes could cause a catastrophic collapse of the DNA-contact surface. The zinc coordination site is close to the DNA-contact surface, indicating that metal chelation may also be important in forming and stabilizing the DNA-binding conformation of p53.

Interactions with DNA

The residues that are in direct contact with DNA are all contained in the loop–sheet–helix motif and L3 (Fig. 5). The loop–sheet–helix motif partially inserts into the major groove, establishing a number of contacts with the edges of DNA bases. Specificity for a C•G base-pair at position 10 (see Fig. 1a), the only invariant position of the consensus sequence, is conferred by a bidentate hydrogen bond between G10 and residue Arg280 of the α-helix H2. The guanidinium head-group of Arg280 is steered by bidentate hydrogen bonding to a carboxylate-containing residue, Asp281. A similar Asp–Arg–G triad has been observed in the co-crystal structure of TFIIIA-type zinc fingers [10,11]; this may not only represent a particularly favorable type of G-contact, but may also enforce selectivity by forcing the Arg to lie in the plane of the base-pair rather than orthogonal to it, which could otherwise allow interaction with an adjacent base-pair [12]. Specificity for position 9 seems to result from hydrogen bonding of C9 to Cys277; this residue is part of the short loop between S10 and H2. Sulfhydryls are generally considered to be poor hydrogen-bond participants, but Cys-SH has been observed to act both as a donor [8,13] and as an acceptor [13] of hydrogen bonds in protein–DNA complexes, and such interactions can clearly make important contributions to sequence-specific recognition. The only other direct amino acid–DNA base contact in the complex is between G8 and Lys120, a residue found in the large loop, L1.

The majority of the DNA–protein interactions take place in the major groove, but the one minor groove interaction is noteworthy: Arg248 inserts itself into the minor groove, establishing a number of hydrogen-bonding interactions with the phosphodiester backbone, and a water-mediated contact to G13. In the region of DNA that interacts with Arg248, the minor groove is compressed by ~ 2 Å and the bases are buckled and highly propeller-twisted. The unusual base-pairing at position 12, together with the water-mediated contact between Arg248 and G13, provide evidence that a single p53 core domain 'sees' the sequence of both pentamers in the p53 half-site (→•←). The helical aberrations appear essential for tight interactions with Arg248, and are of the type that occur more readily in A/T-rich than in G/C-rich DNA. This may be the reason for the selection of A/T base-pairs at the 3'-end of the consensus site. Arg248 is clearly important for DNA binding and p53 function; it is more commonly mutated in human cancers than any other residue in p53.

Several aspects of the protein–DNA interactions seen in the structure of Cho *et al.* [8] are unusual. First, although the p53 core domain, like most sequence-specific DNA-binding proteins thus far characterized, has an α helix that inserts itself into the major groove, the majority of the DNA contacts are made by residues on loops. Loops have previously been found to be involved in DNA recognition [14,15], but they are certainly more important in the p53 core domain than in any other case so far. If loops can be efficient DNA-recognition motifs, it should theoretically be possible to generate antibodies that bind to specific DNA sequences; it is even possible that β-loop folds such as the immunoglobulin and fibronectin domains may one day be found in sequence-specific DNA-binding proteins. Second, in most well-studied cases in which a protein binds DNA stably as a monomer, the structure has multiple independent

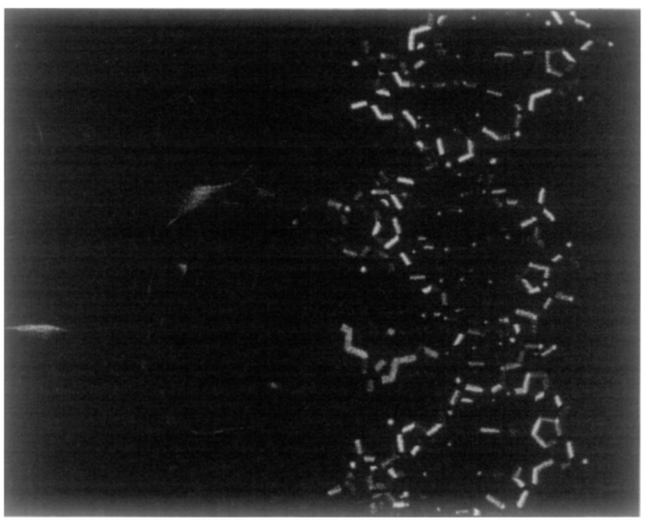

Fig. 5. DNA-contact residues employed by the p53 core domain. Contact residues are shown in licorice bonds, with frequently mutated positions in red and less-frequently mutated positions in yellow. Asp281, which bridges Arg273 and Arg280, is not shown (but can be seen in Fig. 6d of [8]). The DNA is represented in licorice bonds, with the backbone in gray and the bases in blue. Numbering of base-pairs is as given in Fig. 1b and in the text.

modules that interact cooperatively. For example, TFIIIA-type zinc-finger proteins contain multiple tandemly repeated zinc fingers [16], and homeodomain proteins [16], POU-homeodomain proteins [17], Hin recombinase [18], and GATA-1 [14] all have globular domains that bind in the major groove and peptide extensions that make multiple interactions in the minor groove. The single residue of p53 that interacts in the minor groove, Arg248, can hardly be described as being part of an independent structural module, as L3 interacts extensively with the remainder of the core DNA-binding domain. Apart from the p53 core domain, perhaps the only other well-documented case of a single structural module that binds DNA specifically is an engineered monomer of bacteriophage λ Cro, which binds DNA several orders of magnitude less tightly than the native dimeric form of Cro. Unlike Cro, however, the p53 core domain binds a typical response element only five- to six-fold less tightly than the intact p53 tetramer [19].

The p53 core domain thus represents a solution to the problem of binding DNA as a single structural module; the increase in binding energy of the tetramer relative to the monomer is unusually small. It is worth noting, however, that p53 sites vary widely; typically, at least one of the four sub-sites departs considerably from the consensus. Thus, in comparing the binding of the monomer to a typical p53 response element with that of the tetramer, one is actually comparing the binding of the monomer to the 'best' DNA pentanucleotide site (as the monomer will select the best of the four sites available to it) with the binding of the tetramer to all four pentanucleotides, some of which are sub-optimal. The function of oligomerization in the case of p53 may therefore be to expand the range of sequences that can be bound stably. The high degree of degeneracy in the consensus sequence, the variability in spacing on the internal dyad axis, and the use of oligomerization, [8] all

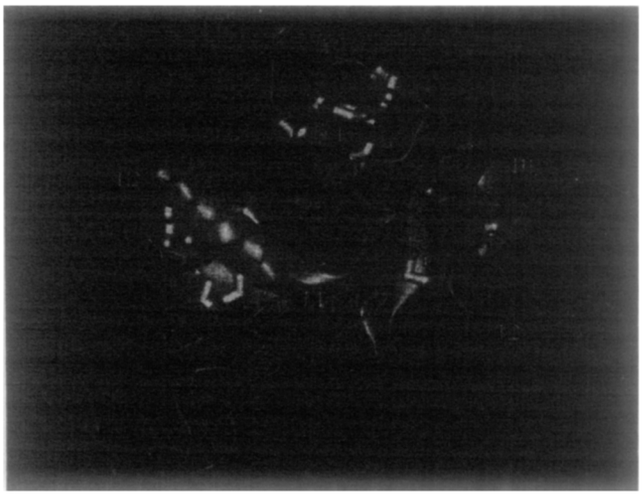

Fig. 6. Positions of mutations in p53 that affect the folding of the protein. Residues in red licorice bonds represent frequently mutated positions, and the other residues with which these interact directly are in yellow. Gly245, which is important in determining the course of the polypeptide backbone, is shown as a red stretch in the ribbon trace. For more detailed views, see Figs. 6e–h of [8].

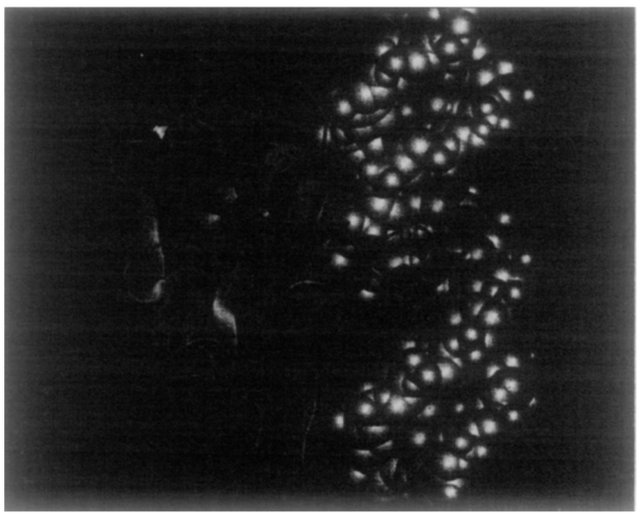

Fig.7. Epitope (red) for the antibody PAb240, which recognizes a variety of p53 mutants but not the wild-type protein.

provide mechanisms for targeting p53 to a wide variety of sites. It is probable that p53 has a wide range of affinities for its various target sequences, providing a mechanism for differential gene regulation at multiple promoters.

Many of the aspects of the promiscuity of p53–DNA binding can be rationalized on the basis of the Cho *et al.* structure. For example, Lys120 could recognize an A instead of a G at base pair 8 by contacting only the *N*7 atom in A instead of *N*7 and O^6 in G; in this configuration, the Lys-NH_3^+ head-group might lie between the purines at base-pairs 7 and 8 (to avoid a steric clash between the A-NH_2 and Lys-NH_3^+ groups), allowing explicit recognition of a purine at base pair 7 [13,15,20]. It is also easy to see how Cys277 could donate a hydrogen bond to a T at base pair 9 (and perhaps make a non-polar contact with the methyl group of T) instead of accepting a hydrogen bond from C. Lastly, recognition at base pair 11 seems to arise from induced DNA distortion of a type that requires only an A/T-rich region.

Even though we may be able to rationalize promiscuous binding on the basis of a given structure, our ability to predict it is woefully inadequate. The arguments used here to rationalize the ability of p53 to bind to different sequences might seem equally pertinent to other proteins on the sole basis of our current knowledge of their structures, but in most of these cases the resulting predictions for variability in binding sites would probably be wrong. Despite the fact that roughly 40 structures of protein–DNA complexes have been solved, our understanding of the structural and energetic origins of sequence specificity remains primitive.

The structural basis of mutational inactivation of p53

One of the most exciting aspects of the structure of the core domain is that it provides clear insight into how changes in the amino-acid sequence inactivate p53. The mutational hotspots correspond closely to the conserved

regions, with the exception that the stretch from 151 to 166 is not designated as a highly conserved region but is a mutational hotspot. On the basis of the structure, the mutations can be divided into two classes: (i) those that directly affect functionality required for interaction with DNA, and (ii) those that affect residues not in direct contact with DNA, but which appear to stabilize the folded structure of the protein. The two residues of p53 most commonly mutated in human cancer cells, Arg248 and Arg273 (see Figs 1a and 5), are both in the first class. Arg248 makes an unusual minor-groove contact that is associated with groove constriction. Arg273 not only makes a phosphate contact, but also buttresses Arg280, a DNA contact residue, through the intermediacy of Asp281. Of the remaining commonly mutated positions, four appear to be involved in maintenance of the folded structure: Arg175 and Arg249 serve to gather together L2, L3 and H1; Arg282 appears to stabilize the loop–sheet–helix motif; and Gly245, instead of supplying an essential functional group, adopts an unusual conformation that allows an essential chain reversal in L3 (Fig. 6). Other positions that are commonly mutated in human cancer include the four ligands for the zinc co-factor (Figs. 1a and 4). The importance of zinc ligation for p53 activity *in vivo* is consistent with the observations that zinc-chelating agents abrogate DNA-binding *in vitro*, presumably by causing unfolding of the protein [19].

The nature of the 'mutant conformation' and its implications for rescue of p53 through small-molecule intervention

Wild-type p53 can be distinguished from many mutant p53 proteins using structure-specific antibodies. This finding led to the notion that p53 can adopt two distinct conformations, one of which (the wild-type conformation) is functional and the other (the 'mutant conformation') is not [21]. This hypothesis became the cornerstone of worldwide drug-discovery efforts aimed at obtaining a small molecule that would confer the wild-type conformation on mutant p53 proteins. Perhaps it is here that the structure of Cho *et al.* [8] will have its greatest practical impact.

The antibody PAb240 is commonly used to detect p53 proteins in the 'mutant conformation'. The epitope to which it binds encompasses most of S7 [22], and is largely buried within the interior of the protein (Fig. 7). For this epitope to become accessible to an antibody, a large part of the core DNA-binding domain would have to refold. Dramatic changes in protein structure have previously been observed, but these involve segmental movements along pivot points, with some resculpting of secondary structure [23,24]; complete remodeling of the secondary and tertiary structure of a protein is without precedent. Moreover, it is highly unlikely that a wide variety of p53 mutants with chemically dissimilar alterations interspersed throughout the sequence could adopt a common mutant conformation. Taking all of this in to account, Cho *et al.* [8] propose the model that satisfies Occam's razor: the mutant conformation is simply an unfolded or partially folded core domain, not a well-ordered fold that is distinct from wild-type. This model, elegant in its simplicity, is consistent with the properties of p53 mutants — including their notorious aggregation *in vitro* and association with heat-shock proteins *in vivo* — and also squares nicely with well-established principles of protein folding. In the post-Cho *et al.* [8] view of the p53 world, the core domain is barely stable and is readily inactivated by a wide variety of mutations. This hypothesis can easily be tested by differential scanning calorimetry or guanidine titration measurements on p53 core domains.

Although the determination of this structure has huge benefits for our understanding of the function of p53, it appears to move the goal-post for drug discovery still farther into the distance. Instead of seeking a molecule that will nudge some sort of reluctant hinge over a pesky conformational barrier, we now need a molecule that will interact tightly with the folded form of the p53 core domain and make up for the interactions lost through mutation. This is an enormous task; many mutations replace interactions that are attractive in the folded protein with repulsive ones. The sum of the resulting energy differences could easily be too large for the binding energy of a small molecule to overcome.

This is not the first time a major drug discovery effort has turned out to center on a flawed premise, nor will it be the last. Another recent example involved the development of molecules to inhibit the rotamase activity of cyclophilin or the FK506-binding protein, efforts that were shown to be fundamentally flawed by the discovery that the catalytic activity of the protein is essentially irrelevant to suppression of T-cell activation [25]. Such are the hazards of attempting to attack the most challenging and potentially beneficial drug targets without having the luxury to wait for basic science to clarify important issues, such as whether the target is truly a valid one.

Acknowledgements: We thank Nikola Pavletich for sharing with us the coordinates of the p53 core domain, and members of the Verdine lab, especially Scot Wolfe, for a critical reading of the manuscript and for assistance with graphics.

References

1. Murray, A. & Hunt, T. (1993). *The Cell Cycle.* Oxford University Press, New York.
2. Prives, C. & Manfredi, J.J. (1993) The p53 tumor suppressor protein: meeting review *Genes Dev* **7**, 529–534.
3. Levine, A., Momand, J. & Finlay, C.A. (1993). The p53 tumour suppressor gene. *Nature* **351** 453–456.
4. McCormick, F. & Myers, P. (1994). From genetics to chemistry: tumor suppressor genes and drug discovery. *Chemistry & Biology* **1**, 7–9.
5. Weinberg, R.A. (1991) Tumor suppressor genes. *Science* **254** 1138–1145.
6. Gannon, J.V., Greaves, R., Iggo, R. & Lane, D.P. (1990). Activating mutations in p53 produce a common conformational effect. A monoclonal antibody specific for the mutant form. *EMBO J.* **9**, 1595–1602.
7. Milner, J. & Metcalf, E.A. (1990). Temperature-dependent switching between 'wild-type' and 'mutant' forms of p53-Val135. *J. Mol. Biol.* **216**, 481–484.
8. Cho, Y., Gorina, S., Jeffrey, P.D. & Pavletich, N.P. (1994). Crystal structure of a p53 tumor suppressor–DNA complex: understanding tumorigenic mutations. *Science* **265**, 346–355.
9. Clore, G.M., *et al.*, & Gronenborn, A.M. (1994). High-resolution structure of the oligomerization domain of p53 by multidimensional NMR. *Science* **265**, 386–391.
10. Pavletich, N.P. & Pabo, C.O. (1991). Zinc finger–DNA recognition: Crystal structure of a Zif268–DNA complex at 2.1 Å.

Science **252**, 809–817.
11. Fairall, L., Schwabe, J.W.R., Chapman, L., Finch, J.T. & Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc finger/DNA recognition. *Nature* **366**, 483–487.
12. Schultz, S.C., Shields, G.C. & Steitz, T.A. (1991). Crystal structure of a CAP–DNA complex: the DNA is bent by 90°. *Science* **253**, 1001–1007.
13. Hegde, R.S., Grossman, S.R., Laimins, L.A. & Sigler, P.B. (1992). Crystal structure at 1.7Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359**, 505–512.
14. Omichinski, J.G., *et al.*, & Gronenborn A.M. (1993). NMR structure of a specific DNA complex of Zn-containing DNA-binding domain of GATA-1. *Science* **261**, 438–446.
15. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S.C. (1992). DNA recognition by Gal4: Structure of a protein–DNA complex. *Nature* **356**, 408–414.
16. Pabo, C.O. & Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* **61** 1053–1095.
17. Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. & Pabo, C.O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* **77**, 21–32.
18. Feng, J., Johnson, R.C. & Dickerson, R.E. (1994). Hin recombinase bound to DNA: The origin of specificity in major and minor groove interactions. *Science* **263**, 348–355.
19. Pavletich, N.P., Chambers, K.A. & Pabo, C.O. (1993). The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots. *Genes Dev.* **7**, 2556–2564.
20. Beamer, L.J. & Pabo, C.O. (1992). Refined 1.8 Å crystal structure of the λ repressor–operator complex. *J. Mol. Biol.* **227**, 177–196.
21. Donehower, L.A., *et al.*, & Bradley, A. (1994). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221.
22. Stephen, C.W. & Lane, D.P. (1992). Mutant conformations of p53: Precise epitope mapping using a filamentous phage epitope library. *J. Mol. Biol.* **225**, 577–583.
23. Ikura, M., Clore, G.M., Gronenborn, A.M., Zhu, G., Klee, C.B. & Bax, A. (1992). Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256**, 632–638.
24. Bullough, P.A., Hughson, F.M., Skehel, J.J. & Wiley, D.C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371** 37-43.
25. Schreiber, S.L. (1991). Chemistry and biology of the immunophilins and their immunosuppressive ligands. *Science* **251**, 283–287.

Daniel A Erlanson and Gregory L Verdine, Department of Chemistry, Harvard University, Cambridge, MA 02138, USA.