HUMAN MUTATION 19:149–164 (2002)

RESEARCH ARTICLE

Integrating Mutation Data and Structural Analysis of the TP53 Tumor-Suppressor Protein

Andrew C.R. Martin,[1]* Angelo M. Facchiano,[2] Alison L. Cuff,[1] Tina Hernandez-Boussard,[3] Magali Olivier,[3] Pierre Hainaut,[3] and Janet M. Thornton[4,5]

[1]*School of Animal and Microbial Sciences, University of Reading, Reading, UK*
[2]*CRISCEB-Research Center of Computational and Biotechnological Sciences, Second University of Naples, Naples, Italy*
[3]*International Agency for Research on Cancer, Lyon, France*
[4]*Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College London, London, UK*
[5]*Department of Crystallography, Birkbeck College, London, UK*

Communicated by Anne-Lise Børreson-Dale

TP53 encodes p53, which is a nuclear phosphoprotein with cancer-inhibiting properties. In response to DNA damage, p53 is activated and mediates a set of antiproliferative responses including cell-cycle arrest and apoptosis. Mutations in the TP53 gene are associated with more than 50% of human cancers, and 90% of these affect p53-DNA interactions, resulting in a partial or complete loss of transactivation functions. These mutations affect the structural integrity and/or p53-DNA interactions, leading to the partial or complete loss of the protein's function. We report here the results of a systematic automated analysis of the effects of p53 mutations on the structure of the core domain of the protein. We found that 304 of the 882 (34.4%) distinct mutations reported in the core domain can be explained in structural terms by their predicted effects on protein folding or on protein-DNA contacts. The proportion of "explained" mutations increased to 55.6% when substitutions of evolutionary conserved amino acids were included. The automated method of structural analysis developed here may be applied to other frequently mutated gene mutations such as dystrophin, BRCA1, and G6PD. Hum Mutat 19:149–164, 2002. © 2002 Wiley-Liss, Inc.

KEY WORDS: p53; TP53; relational database; structural analysis; SNP; DNA binding; transcription factor; tumor suppressor

DATABASES:
TP53 – OMIM:191170; GDB:120445; GenBank: U94788; PDB:1TSR; http://www.iarc.fr/p53/ (IARC p53 Page); http://perso.curie.fr/Thierry.Soussi/polymorphism.html (p53 Polymorphisms)
Full results of the present analysis can be found at: http://www.bioinf.org.uk/p53/ or http://www.rubic.rdg.ac.uk/p53/

INTRODUCTION

From the discovery of p53 (TP53; MIM# 191170) in 1979, to the elucidation of its roles in the cell, the interest in this protein has increased continuously [Matlashewski, 1999; May and May, 1999]. The p53 tumor suppressor protein is a nuclear phosphoprotein with cancer-inhibiting properties [Crawford, 1983; Culotta and Koshland, 1993; Harris, 1993; Levine, 1997]. This ubiquitous factor is kept in a repressed state in normal cells, but is activated by post-translational modifications in response to multiple

Received 2 July 2001; accepted revised manuscript 22 October 2001.

*Correspondence to: Dr. Andrew C.R. Martin, School of Animal and Microbial Sciences, University of Reading, Whiteknights, PO Box 228, Reading RG6 6AJ, UK.

Contract grant sponsor: U.K. Medical Research Council; Contract grant sponsor: EC; Contract grant number: QLG1-1999-00273.

Current address for Angelo M. Facchiano: Istituto di Scienze dell'Alimentazione, CNR, via Roma 52 A/C, 83100 Avellino, Italy.

Current address for Tina Hernandez-Boussard: Department of Genetics, Stanford University, Stanford, CA 94305-5120.

©2002 WILEY-LISS, INC.
DOI 10.1002/humu.10032

forms of stress, both genotoxic (such as irradiation, chemical carcinogens, or cytotoxic agents used in cancer therapy) or non-genotoxic (such as hypoxia, depletion of ribonucleotides, and oncogenic activation of growth signaling cascades) [North and Hainaut, 2000]. When active, the p53 protein accumulates to high levels in the nucleus and acts as a multi-functional transcription factor to enhance or repress the expression of several sets of genes involved in cell cycle progression, apoptosis, adaptive response to stress, differentiation, and DNA repair [Vogelstein et al., 2000]. Thus, p53 controls and coordinates anti-proliferative responses to prevent DNA replication from occurring when cells are exposed to adverse conditions. The mechanism of the p53 mediated suppression of cell cycle progression involves arrest within the G1 phase [Levine, 1997; Ko and Prives, 1996] as a consequence of the p53 induced synthesis of p21, an inhibitor of cyclin E/cdk2 and cyclin A/cdk2 kinases. In this way, p53 gives DNA repair mechanisms time to correct damage before the genome is replicated. If damage to the cell is too severe, p53 initiates apoptosis by inducing transcription of genes encoding proapoptotic factors [Lakin and Jackson, 1999; Chao et al., 2000]. p53 also enhances or represses the expression of genes involved in the adaptive response to stress, differentiation, and the DNA repair process. These properties have led to the concept that p53 plays a central role in carcinogenesis.

Tumor-specific p53 mutations were first identified in 1989 [Romano et al., 1989]. Loss of p53 function is the most common event in human cancer, with more than half of all invasive tumors involving the decrease or total loss of p53 function. Mutations in one allele assert a dominant–negative effect over the remaining wild-type allele, resulting in genetic instability, loss of hetrozygosity, and a detrimental effect on the function of p53 [Brachmann et al., 1998]. Some may also exert their own oncogenic activity [Ko and Prives, 1996]. The use of increasingly precise functional assays has shown that some mutants retain wild-type activities toward a subset of promoters, whereas others have not only lost wild-type function, but have gained pro-oncogenic functions that are still poorly understood [Sigal and Rotter, 2000]. Recent evidence suggests that these pro-oncogenic functions might result from complex formation between certain mutant p53 proteins and p73 or p63, two newly discovered proteins with homologies to p53, and with functions in normal differentiation and development [Wiederschain et al., 2001; Irwin and Kaelin, 2001]. Correct functioning of p53 is critical to radiation and chemotherapy since both rely on causing DNA damage which triggers apoptosis via p53 [Brachmann et al., 1998].

Raw mutation data have been collected over a number of years by groups in Germany and France. To date, over 15,000 individual patients with TP53 gene mutations have been reported in the world literature. These mutations are compiled in a database maintained at the International Agency for Research on Cancer (IARC TP53 database, www.iarc.fr/p53). This is the largest dataset available on the variations and mutations of any human gene [Hainaut et al., 1998]. Release 4 of the database consists of more than 14,000 mutations affecting over 300 residues linked with more than 60 different tumors. This collection of data is now being expanded with information on the pathology and clinical outcome of different mutations and tumors.

In contrast to many other tumor suppressors, which are often inactivated by deletion or frameshift mutations, most of the mutations in TP53 are point mutations (missense mutations: 75%; nonsense mutations: 8%). These mutations are exceptionally diverse in their nature and position. Thus, it is possible to draw tumor-specific mutation spectra that show significant differences from one type of cancer to the other. This observation has two very important implications: first, the spectrum of mutations reveals information on the mutagenic processes that cause human cancers, and second, the whole set of mutations observed in cancer can be analyzed as an immense, in vivo, random mutagenesis experiment aimed at identifying residues which are important in the maintenance of the tumor suppressive function of the protein.

The open reading frame of human p53 codes for 393 amino acids. consisting of three major structural domains: an N-terminal domain which contains a strong transcription activation signal [Vogelstein and Kinzler, 1994], a DNA-binding core domain, and C-terminal domain which mediates oligomerization. Comparisons of p53 sequences from different species indicate five blocks of highly conserved residues which coin-

cide with mutation clusters found in p53 in human cancers. A detailed analysis of p53 mutations based on evolutionary conservation has been performed by Walker et al. [1999]. The vast majority of the mutations in p53 cluster in conserved regions that encode the sequence specific DNA-binding "core" domain (residues 96–292). This domain has been crystallized in the form of a complex with its target DNA by Cho et al. [1994], as shown in Figure 1, and in the form of a protein complex with the protein BP2 (p53-binding protein 2) by Gorina and Pavletich [1996]. Several structures of the C-terminal oligomerization domain have also been solved [Jeffrey et al., 1995; Clore et al., 1995; Mittl et al., 1998, for example]. Some 20% of all mutations are concentrated at five "hotspot" codons in the core domain: 175, 245, 248, 249, and 273 (standard human numbering scheme as used in Protein Databank file 1tsr [Cho et al., 1994] and throughout this work).

The core domain consists of a large β-sandwich of two anti-parallel sheets of four and five strands, respectively. This acts as a scaffold supporting three loop-based regions—a loop/β-strand/α-helix motif (L1), and two large loops (L2 and L3). L2 and L3 are stabilized by zinc coordination and side-chain interactions [Cho et al., 1994; Wong et al., 1999]. DNA is bound by L1 and L3—the L1 helix slots into the major groove and L3 binds in the minor groove. The L2 loop stabilizes L3 by packing against it. It has been proposed that p53 binds as a tetramer [Vogelstein and Kinzler, 1992] and Pavletich et al. [1993] stated that oligomerization interactions occur through the C-terminal domain (residues 325–356) [Jeffrey et al., 1995].

The notion that several categories of mutants may exist has received a lot of attention since it was realized that not all mutants are functionally equivalent and substitutions in p53 have been classified in a number of ways. Michalovitz et al. [1991] suggested a genetic classification of mutations based on the dominance of their ac-

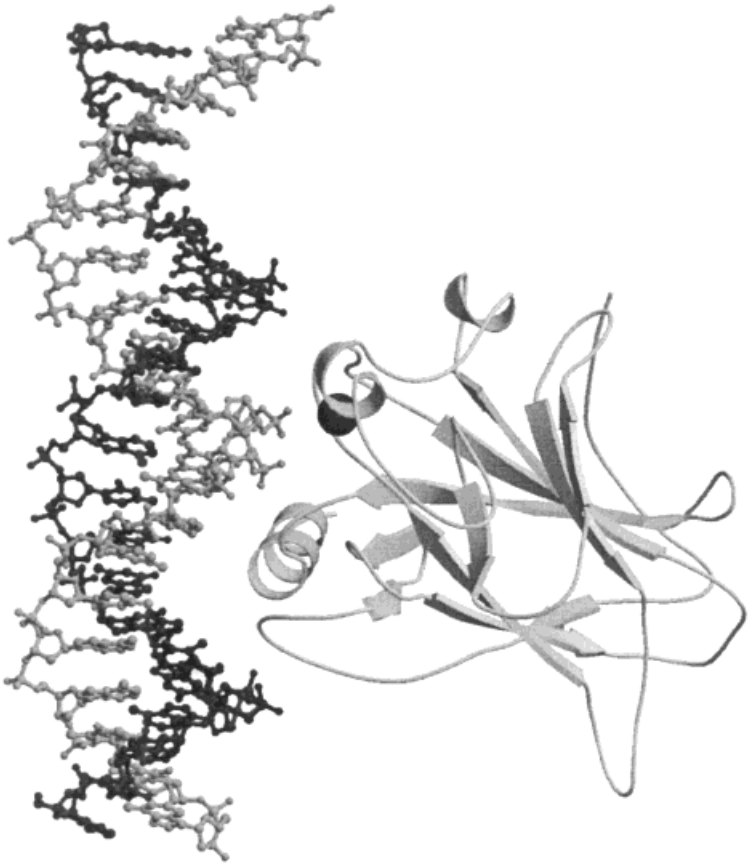

FIGURE 1. **Crystal structure of the core domain of p53 bound to DNA as solved by Cho *et al.* [1994] rendered using Molscript [Kraulis, 1991] and Raster3D [Merritt and Bacon, 1997].**

tivity. More commonly, the substitutions found in cancer have been classified into two broad categories according to their position and type: 1) contact mutations, located at the protein-DNA interface and breaking crucial contacts between the protein and DNA, or between p53 and other proteins [Brachmann et al., 1998; Nikolova et al., 2000]; 2) conformational mutations, located in the protein skeleton and thought to prevent its correct folding, thus preventing high-affinity DNA binding. Indeed, p53 has been shown to be only marginally stable at body temperature [Nikolova et al., 2000], so any mutation which further reduces stability is likely to lead to unfolding/misfolding in vivo.

Although the percentage of occurrence quoted varies considerably, inherited p53 mutations are rare. Li et al. [1992] suggest 0.01% in the normal population and 0.1–1% in various cancer patients, while Guinn and Padua [1995] state that 5% of p53 mutations are inherited. Germ line mutations in the p53 gene have been observed in several families with Li-Fraumeni syndrome [Malkin et al., 1990; Srivastava et al., 1990]. This results in an inherited predisposition to a broad spectrum of cancers including breast cancer, osteosarcomas, soft tissue sarcoma, melanoma, adenocortical carcinomas, and leukemias all of which appear at an early age. The GenBank entry for the human p53 gene (TP53, U94788) lists six sequence variations (four single nucleotide polymorphisms (SNPs), one single base deletion, and one 16-base repeat). Of these, three of the SNPs occur in coding regions (at bases 12032, 12139, and 13399). Only one of these, 12139G→C, results in an amino acid change (arginine→proline) at residue 72 [Ara et al., 1990] in the N-terminal domain. Soussi's p53 web site (http://perso.curie.fr/Thierry.Soussi/polymorphism.html) lists an additional polymorphism resulting in an amino acid substition: proline-47→serine [Felley-Bosco et al., 1993].

Endogenous processes, including methylation and deamination of cytosine at CpG residues, free radical damage, and errors that may occur during the synthesis or repair of DNA can result in p53 mutations [Hainaut et al., 1998]. Mutations can also occur via DNA damage induced by exogenous, physical, or chemical carcinogens. In some cases "mutagen fingerprints" have been identified where certain carcinogens are responsible for specific mutations [Greenblatt et al., 1994; Harris, 1996]. For example, cigarette smoke causes G:C to T:A transversions in lung cancers [Chiba et al., 1990] while aflatoxin B1 (AFB1) in the diet, particularly in China and Africa, causes G:C to T:A transversions specifically at the third base pair of codon 249 (AGG→AGT) and is associated with liver cancers. Similarly, UVB exposure is associated with CC:GG to TT:AA dipyrimadine transitions in skin cancers [Brash et al., 1991].

There is growing evidence that biological variations between mutants may have clinical relevance. Recent studies have shown that the response of individual patients to cytotoxic therapy may be influenced by the nature and position of the mutations in p53. In particular, in breast cancer, patients with mutations within conserved residues of the DNA-binding surface appear to have increased resistance to doxorubicin-based therapy, but not to taxol [Aas et al., 1996; Geisler et al., 2001]. If confirmed, these differences are bound to have an enormous impact in molecular prognosis and in selection of appropriate cancer treatments.

In theory, it should be possible to restore at least some functional activity to tumor-derived p53 mutants by 1) enhancing the stability of the protein in its folded state and/or 2) providing additional DNA contacts [Brachmann et al., 1998; Nikolova et al., 2000; Sigal and Rotter, 2000]. It is possible to rescue some p53 mutations using second-site suppressor mutations. For example, the "hotspot" mutation G245S causes structural changes in the L2 and L3 loops, suggestive of distortion of the conformation necessary for DNA binding. Nikolova et al. [2000] found that the suppressor mutant N239Y restored the stability of G245S and resulted in an improvement in DNA binding. They observed similar results using other second-site suppressors to restore some degree of normal function to p53. The marginal stability of p53 suggests that it may be possible to restore wild-type activity through design of drugs which bind the correctly folded form, thus moving the equilibrium through simple mass action [Brachmann et al., 1998; Wong et al., 1999; Nikolova et al., 2000]. However, so far the exploitation of the available clinical data has been hampered by our limited understanding of the structural and functional characteristics of each of the individual p53 mutant proteins.

The work presented here is an attempt to exploit the knowledge available on p53 protein structure and the power of current structural analysis methods to classify the various types of p53 mutants into specific categories. We performed a structural analysis of the p53 crystal structure, calculating secondary structure, backbone torsion angles, solvent accessibility, and hydrogen bonding parameters, and stored these data in a relational database. We have then interconnected this database with the IARC TP53 mutation database to correlate structural effects with mutations in an automated fashion. We find we are able to rationalize the effects of 34.5% of distinct mutations on purely structural grounds. These mutations represent 50% of all mutations described so far in human cancers. If we also consider residues which are 100% conserved across a range of species (and therefore likely to be important for the function of p53), this percentage rises to 55.6% (representing 77.8% of the observed mutations). Thus, a not insignificant number of mutations cannot be explained on a structural basis. We propose that these unexpected mutations may fall into one of three classes: 1) those which are not involved in cancer, are essentially nonpathogenic and are detected only by virtue of their accidental presence in a clone of cancer cells; 2) those which we have genuinely failed to identify, possibly because they have only a subtle destabilizing effect on the marginally stable p53; 3) those which are on the surface of the p53 core domain and are involved in interactions with the other p53 domains or with other proteins.

While the biological conclusions of this analysis do not differ significantly from the analysis performed by Cho et al. [1994], this is the first example of a fully automated protocol being applied to survey the effects of all known mutations and is of general applicability.

MATERIALS AND METHODS

Mutation Data

The raw mutation data are maintained by Hainaut in a relational database [Hernandez-Boussard et al., 1999] and made available as a flat file from ftp://ftp.ebi.ac.uk/pub/databases/p53/. The latter format was imported into a PostgreSQL relational database (www.PostgreSQL.org/) using a script written in Perl. The raw data [Hainaut et al., 1998] contain p53 mutations associated with human cancers identified by sequencing and published in the literature. These data include mutations found in normal, pre-neoplastic, and neoplastic tissues, including metastases, as well as cell lines derived from such tissues. It should be noted, however, that some mutants may simply be an indication of DNA damage and may not be tumorigenic. The data file contains 34 columns and includes data on cell-line, codon, DNA base and amino acid substitution, International Classification of Diseases for Oncology (ICD-O) tumor-site, tumor morphology and histology, tumor grade or stage, and risk factors (sex, country of origin, smoking status, and alcohol consumption).

We considered both in-frame and out-of-frame insertions in the same manner; in both cases it is clear that the function of p53 could be disrupted. We also flagged silent-point mutations. Earlier versions of the p53 data required considerable clean-up during this procedure; the current dataset required minimal clean-up (some frameshift mutants classified as "point" rather than "del" or "ins," minor changes to the page numbering format of references, etc.).

Structural Data

Our structural analysis was based on the Human p53 crystal structure PDB file 1tsr chain B [Cho et al., 1994]. The numbering used in this file adopts the standard scheme used throughout the literature. The parameters calculated were: secondary structure using DSSP [Kabsch and Sander, 1983], hydrogen bonding using HBPlus [McDonald and Thornton, 1994], backbone torsion angles, and solvent accessibility [Lee and Richards, 1971] using NAccess (Simon Hubbard, unpublished). These data were imported into a database table keyed by residue (codon) number.

Sequence Variability

Sequences used in this analysis came from human, cat, golden hamster, bovine, sheep, mouse, rainbow trout, rat, chicken, North European squid, dog, green monkey (Cercopithecus aethiops), Macaca mulatta, Xenopus laevis, and Spermophilus beecheyi.

We considered sequence variability on the basis that residues which are 100% conserved across such a diverse selection of species must be conserved for functional reasons. Thus we

may not have direct structural explanations of why mutations to these residues might affect the function of p53, but we know that these residues are critical to the function of p53 and this is likely to be as a result of interactions with other proteins.

At each residue position the sequence variability was assessed using a score based on the PET91 mutation matrix [Jones et al., 1994], normalized such that all scores on the diagonal are maximal and equal. The score is calculated as the average pairwise sum of the matrix scores normalized by the maximum score in the matrix:

$$S_n = \left[\frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} S_{ij}}{{}_N C_2} \right] / S_{\max}$$

where n is the position in the sequence, N is the number of sequences, S is a score from the mutation matrix, and ${}_NC_2$ is the number of combinations of two elements from the set of N elements (${}_nC_r = n!/((n-r)!r!)$). In this scheme, complete conservation scores 1.0; lower levels of conservation score values down to 0.0, depending not only on the raw variability (as is the case with statistical entropy-based scores [Shenkin et al., 1991]), but also on the nature of the mutation. The sequence variability scores are stored in the structural data table and are illustrated in Figure 2.

FIGURE 2. **Conservation of the p53 core domain, as based on the sequence variability score. The non-conserved regions are shown in light grey, low conservation regions in grey, and the most highly conserved regions in black. Picture rendered using QTree (ACRM, unpublished, www.bioinf.org.uk/software/qtree/).**

Assessing Sidechain Replacements

For the present study, very simple assessments of the effects of changes in the structural properties were used. For example, if a residue was involved in donating a sidechain hydrogen bond and is replaced by a residue without hydrogen-bond donor potential we claim to have explained the structural effect of the mutation. If the replacement sidechain is also able to donate a hydrogen bond, the geometry of the new hydrogen bond is not tested, it is assumed that small changes in the structure can be accommodated. We thus take a cautious approach and do not classify such mutants as explained even though they may, in fact, be explained in this way.

Each unique sidechain replacement is also assessed on the basis of steric acceptability. The current procedure is again very simple; we adopt a minimum perturbation protocol (MPP) [Shih et al., 1985; Snow and Amzel, 1986] to model the new sidechain into the 3D crystal structure of the p53 core domain and then count any bad clashes with the substituted sidechain. MPP proceeds as follows:

1. Perform a maximum overlap protocol (MOP) [Snow and Amzel, 1986, for definition] replacement of the sidechain where torsion angles are inherited from the parent sidechain where possible.
2. Build a near-neighbors list using a cutoff of 8Å (this is greater than the longest sidechain, tryptophan).
3. Spin the sidechain about $\chi 1$ and $\chi 2$ torsion angles in 30 steps flagging each position as either making bad contacts or not.
4. If the parent conformation (resulting from MOP) makes zero or one bad contacts then that conformation is accepted.
5. If that fails, then for all the conformations with zero or one bad contacts, a choice is made from allowed rotamers.

6. If that fails, the first conformation with a minimal number of clashes is selected.

A bad contact is defined as two atoms whose centers are closer than 2.5Å—this is a simple good/bad assessment; no degree of bad contact is calculated. We take three clashes as being indicative of a sidechain replacement which cannot be accommodated. Again this is a conservative decision; it appears that two clashes are sufficient to disrupt the structure in many cases.

By using the ability of PostgreSQL to allow user-defined functions, the clash assessment can be performed on the fly. In practice, for speed reasons, it is useful to cache the results of all unique sidechain replacements into a column in another database table. This can be achieved by performing a single SQL query on the database. These data were stored in a database table keyed by residue (codon) number and replacement residue type.

Analyzing the Data

Analysis of the data was performed using a set of Perl routines which query the database and extract and format the data. The procedure has been completely automated such that it can be repeated on new datasets as these become available.

RESULTS AND DISCUSSION

Summary of Data

Table 1 summarizes the mutation data from Release 4 of the p53 mutation databank. For the purposes of this investigation, we have concentrated on analyzing the distinct mutations which result in a simple amino acid substitution in the core domain for which a crystal structure is available. As the table shows, there are 882 of these. This is approximately 51% of the total number of the distinct mutations; the remaining 49% are either more complex mutations, insertions, deletions, or occur outside the core domain. These simple substitution mutations in the core domain represent 69.8% of the total number of observed mutations. The full results of this analysis are available on the Web (www.bioinf.org.uk/p53/ or www.rubic.rdg.ac.uk/p53/).

TABLE 1. Summary of p53 Mutation Data*

	Total observed	Distinct
Total number of mutations	14,050	1,729
Complex mutations	69	60
Deletions	1,250	253
Insertions	357	152
Tandem mutations	200	67
Silent mutations	5	1
Point mutations	12,138	1,363
Of these		
Tandem/Point mutations resulting in an amino acid substitution	10,204	1,083
Tandem/Point mutations resulting in an amino acid substitution in the core domain	9,812	882

*Many mutations are observed more than once as they may occur in different patients or different tumours. We therefore also consider "distinct" mutations rather than the total number of observed mutations.

Mutations Affecting Hydrogen Bonding

Hydrogen bonds stabilize the structure of a protein and hydrogen bonding ability must be satisfied throughout.

As described by Baker and Hubbard [1984] the following residues are classified as able to donate a hydrogen bond: H, K, N, Q, R, S, T, W, Y, while the following residues can accept a hydrogen bond: D, E, H, N, Q, S, T, Y. If a residue involved in a hydrogen bond is substituted by another residue unable to form the hydrogen bond, the protein is likely to be destabilized. As well as replacement by hydrophobic residues, this includes replacement of a residue donating a hydrogen bond with one only capable of accepting hydrogen bonds and vice versa. The amino acids K, R, and W are only able to donate, while E and D are only able to accept hydrogen bonds. Other amino acids able to form hydrogen bonds are both donors and acceptors (H, N, Q, S T, Y). Thus if one of K, R, or W is substituted by E or D (or vice versa) then we will suggest that the disruption of hydrogen bonding has disrupted the structure. However, if the hydrogen bonding pair is on the surface of the protein, then a substitution with a hydrogen-bond-capable residue which cannot maintain the native hydrogen bond will have little effect on stability since the sidechains can instead form hydrogen bonds with water. There is a total of 4,703 substitution mutations (309 distinct mutations) involving hydrogen bonding residues. Using our assessment, (see Methods section) where we do not consider the precise geometry and assume that a small local rearrangement can be accommodated, we find that we can explain 43.2% of observed mutations to hydrogen bonding residues (52.5% of distinct mutations).

Of the 309 distinct mutations which appear to disrupt hydrogen bonding, 17 involve mutation to a donor-only (R, K, H) or acceptor-only (E, D) sidechain and have at least 10% relative accessibility with the sidechain itself being accessible. The partner amino acid is also accessible in 16 of the 17 cases. Thus for 16 of these, it is possible for the hydrogen bonding requirements to be satisfied by water. Therefore, we may have only justified 48.1% of distinct mutations to hydrogen-bonding residues (see Table 2).

Mutations to Proline

Owing to the cyclic sidechain of proline, the backbone is more restricted in the conformations which it can adopt. Thus mutations from other residues to proline may result in distortion of the structure if the parent amino acid did not adopt a backbone conformation permitted for proline. In addition, proline residues will break an α-helix or disrupt a β-sheet in all but edge β-strands since cyclization of the sidechain prevents the regular backbone H-bond formation. Thus mutations to proline in these circumstances will lead to an incorrectly folded protein. A total of 332 point/tandem mutations (58 distinct mutations) result in a mutation to proline (in addition there are 77 silent mutations at 20 distinct sites involving proline, of which 62 occur in the core at 12 distinct sites).

Of the 332 mutations resulting in a substitution by proline, 320 occur in the core at 50 distinct sites. Figure 3 shows these core domain substitutions in the form of a Ramachandran plot. Those combinations which are disallowed regions for proline are indicated. We define the allowed regions for proline as $-70° \le \phi \le -50°$ and ($-70° \le \Psi \le -50°$ or $110° \le \Psi \le 130°$). A total of 47 of the 50 mutations (94%) are disallowed and will thus result in disruption of the structure. Note however, that these ranges for the allowed regions are rather strict, so this may be a slight over-prediction. Some, indeed, are borderline and may be accommodated by a very small rearrangement (e.g. L137P). The 47 disallowed proline mutations sites are illustrated in Figure 4.

TABLE 2. **Mutation to Residues Involved in Hydrogen-Bonding***

	Total mutations involving H-bonding residues		H-bonding potential not conserved	
	Observed	Distinct	Observed	Distinct
Donor	3856	205	1422	104
Acceptor	1479	155	881	85

*The first pair of colums shows the numbers of mutations and the second pair of columns shows the numbers of mutations where hydrogen-bonding potential is lost.

Mutations From Glycine

A total of 809 mutations (70 distinct mutations) are observed from a native glycine to another residue (in addition there are 71 silent mutations of glycine at 14 distinct sites). Of these, 771 (53 distinct) occur in the core region.

Because it has no sidechain, glycine is able to adopt conformations which are sterically hindered for other amino acids. Substitution of any native glycine residues which adopt one of these conformations will thus result in disruption of the structure resulting in an incorrectly folded protein.

The allowed regions of the Ramachandran plot for non-glycine/non-proline residues are, for this purpose, defined as: ($-180.0° \le \phi \le -30.0°$ / $60.0° \le \Psi \le 180.0°$) or ($-155.0° \le \phi \le -15.0°$ / $-90.0° \le \Psi \le 60.0°$) or ($-180.0° \le \phi \le -45.0°$ / $-180.0° \le \Psi \le -120.0°$) or ($30.0° \le \phi \le 90.0°$ / $-20.0° \le \Psi \le 105.0°$). All non-glycine residues in the p53 crystal structure fall within these limits.

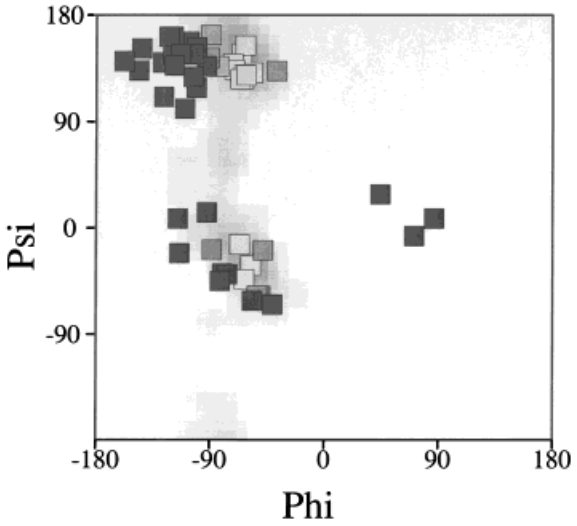

FIGURE 3. **Mutations to proline occurring in the DNA binding domain of p53. The darker background color of this Ramachandran plot (produced by ProCheck [Laskowski et al., 1993]) indicates favored regions for proline. Mutations to proline occurring in allowed regions are shown as light squares, generously allowed regions as grey squares, and disallowed regions as black squares.**

FIGURE 4. **The 47 sites where disallowed proline substitutions are observed are indicated in grey.**

FIGURE 5. **The six sites at which glycines adopt backbone conformations disallowed for other residues and where substitutions occur are indicated in grey.**

With the exception of the glycine residues at codons 117, 154, 187, 244, 245, and 262, all the others fall in regions allowed for other amino acids. Therefore, only mutations to these six glycines will result in disruption of the structure. These sites are illustrated in Figure 5. Figure 6 shows a Ramachandran plot of the p53 core domain indicating glycine residues which are substituted by other amino acids. Thirty-two of 53 core region distinct mutations from glycine (60.4%) are disallowed. Full data are on the Web at www.rubic.rdg.ac.uk/p53/ or www.bioinf.org.uk/p53/.

Residue Clashes

If a substituted residue is too large for the available space it will lead to distortion of the structure and may result in the protein folding incorrectly. Of the 882 distinct substitution mutations in the core, 24 (2.7%) result in a bad clash (three or more bad contacts with surrounding atoms in the best sidechain orientation, see Fig. 7). If we consider that any number of bad contacts will disrupt the structure, we can include a further 44 distinct mutations, resulting in a total of 68 (7.7%) distinct mutations resulting in bad clashes.

Mutations Involving DNA Binding

The most common mutations observed in p53 are involved in binding DNA. These mutations result in the protein either being unable to bind to p53 or losing specificity of interactions. We define DNA binding residues as those in which the relative accessibility changes by at least 5% between the complexed form of p53 observed in the crystal structure and the same structure of p53 but with the DNA removed. This identifies 14 residues (alanine-119, lysine-120, serine-121, asparagine-239, serine-241, methionine-243, asparagine-247, arginine-248, arginine-273, cysteine-275, alanine-276, cysteine-277, arginine-280, arginine-283) all of which are seen to have mutations.

At these 14 sites, a total of 2,383 mutations is observed, 74 of which are distinct. While mutations at the more peripheral of these sites may, in some circumstances, allow DNA still to bind, the stability of the complex and the specificity of DNA binding is likely to be affected.

Mutations Involving Zinc Binding

Zinc binding is essential for the function of p53—presumably it does not adopt the correct conformation in the absence of zinc binding.

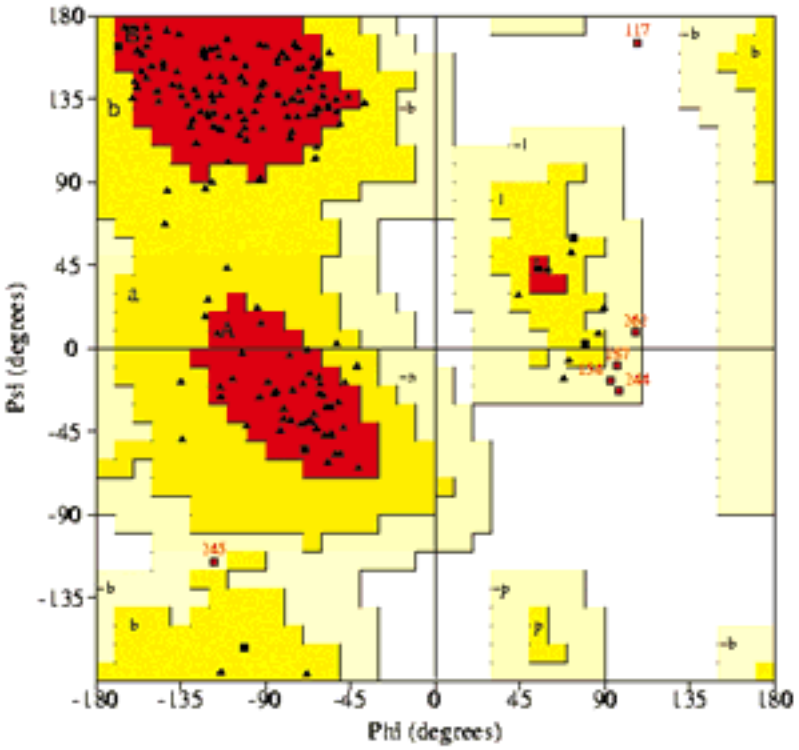

FIGURE 6. **Ramachandran plot of the p53 core domain generated using Procheck. All native non-glycine residues are shown as triangles while native glycine residues are shown as squares. Those which are in disallowed conformations for non-glycine residues (and which undergo mutation) are labeled.**

Thus mutations to the residues involved in interaction with zinc will result in p53 being nonfunctional. Examination of the crystal structure shows that cysteine-176, histidine-179, cysteine-238, and cysteine-242 are all involved in zinc binding. A total of 611 mutations are observed at the four sites, 29 of which are distinct. Any mutation to these residues is likely to prevent or destabilize zinc binding, destabilizing the structure and resulting in loss of function.

Mutations to Conserved Regions

A total of 63 residues in the core domain are 100% conserved across all species for which p53 sequences were analyzed (see the Methods section). These are residues 98, 113, 120,121, 122, 125, 127, 130, 132, 137, 139, 142, 151, 152, 158, 159, 164, 172, 173, 175, 177, 178, 179, 196, 198, 199, 205, 208, 215, 216, 218, 219, 220, 221, 223, 230, 239, 240, 241, 242, 243, 244, 245, 247, 249, 251, 253, 257, 262, 265, 266, 267, 270, 271, 272, 275, 276, 277, 278, 279, 280, 281, and 282.

While we cannot currently offer a direct structural explanation for many of these, one can assume that they are conserved throughout evolution for a good reason and, in the case of surface residues, this is likely to be that the amino acid is critical for interactions with other proteins. 4,767 mutations resulting in amino acid substitutions occur (335 distinct) to these 63 conserved residues.

CONCLUSIONS

Although p53 is one of the most studied proteins, we still do not have a full understanding of the role of individual p53 mutants in carcinogenesis. Both functional studies and analysis of mutations in human cancers provide overwhelming evidence that the main biochemical activity targeted in cancer cells is sequence-specific DNA-binding. However, each mutation damages the structure of the DNA-binding domain in a specific way, leading to various functional consequences. For example, it has long been recognized that the R273H mutant, which affects a residue involved in contacts with the phosphate backbone of DNA, has "milder" biological effects than R175H, which affects a residue thought to play an important role in the stabilization of interactions between the L2 and L3

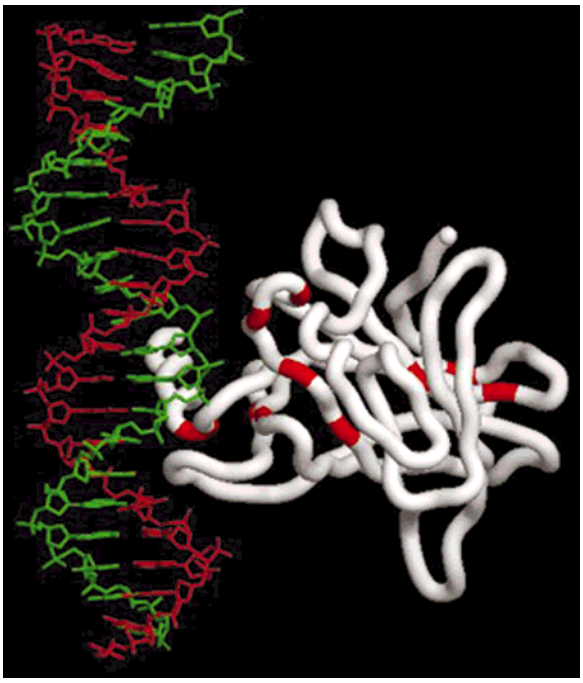

FIGURE 7. **The 11 sites at which substitutions result in bad clashes are indicated in grey.**

loops. More recently, it has been shown that several relatively rare mutants could retain partial DNA-binding activity with, for example, the capacity to transactivate high-affinity binding promoters such WAF-1, but not low-affinity ones such as BAX. These differences may reflect the extent of the structural perturbation induced by the mutation into the protein structure.

This is the first time this type of overview analysis has been performed. Other work on p53 mutations has tended to concentrate on individual mutants of interest rather than attempting to automate the classification of structural

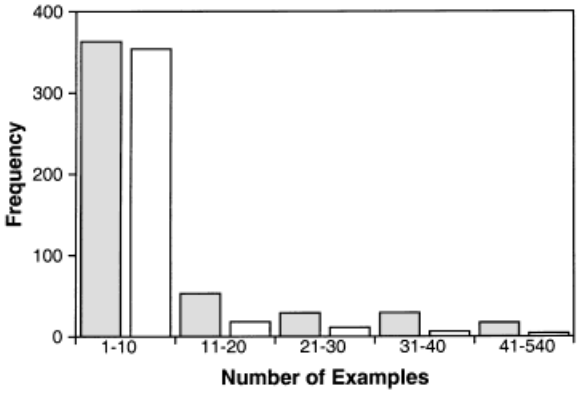

FIGURE 8. **Frequency distributions of explained mutations (grey bars) and unexplained mutations (white bars).**

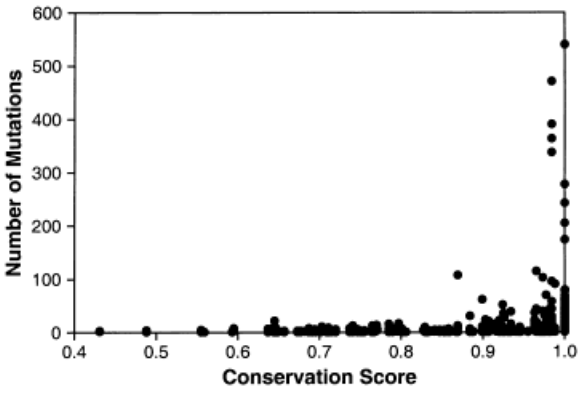

FIGURE 9. **Plot of number of mutations observed in the IARC TP53 database against conservation score calculated across a diverse range of species.**

effects. Some of the mutations can be explained in multiple ways as shown in detail on the web site (www.bioinf.org.uk/p53/ or www.rubic.rdg.ac.uk/p53/).

In this report, we have used a combination of standard structural criteria to evaluate the impact of missense mutations in the DNA-binding core domain in an automated manner. It should of course be noted that mutations do occur outside the core domain, and indeed two polymorphisms (P47S and R72P) have been identified in the N-terminal transcription domain. Overall, this strategy allows us to provide a realistic, purely structural explanation for loss of DNA binding activity for 304 of the 882 distinct mutations (34.5%) reported in the Release 4 version of the IARC TP53 mutation database. Together, these 304 mutations correspond to 4,916 of the 9,824 entries for missense mutations in the DNA-binding domain in the database (50%). However, 16 of the structural mutations were involved in surface H-bonds and, as already stated, it is possible that their hydrogen-bonding requirements are satisfied by water. Seven of these mutations have other likely explanations for their effects (100% conservation, clashes, or involvement in zinc binding). However, this is not the case for the remaining nine mutations which may therefore have no predictable structural effect. If we take these mutations into account, 33.4% of distinct mutations in the core domain are explained on purely structural grounds and 54.5% if 100% conservation is also considered.

This figure is surprisingly low and indicates that over 65% of the distinct mutations, repre-

senting about half of all missense mutations described in cancer, cannot be evaluated with the simple structural criteria used here. However, it is important to note that p53 is only marginally stable at body temperature [Nikolova et al., 2000], so mutations which lead to only a small decrease in stability may lead to unfolding/misfolding in vivo. The list of the mutations "explained" by our structural criteria includes most common mutation hotspots, with two major exceptions: R175H and R249S. Interestingly, R175H and R249S are among the best functionally characterized p53 mutants. They both show total loss of transcriptional activity in functional yeast assays and in reporter cell systems. They adopt a conformation recognized by the monoclonal antibody PAb240, which is directed toward an epitope (residues 213–217) which is cryptic in the correctly folded, wild-type protein [Gannon et al., 1990].

The structural consequences of these two substitutions have been described by Cho et al. [1994]. Arginine-175 is located in the L2 loop, adjacent to cysteine-176, one of the zinc binding sites, and is involved in bridging the L2 and L3 loops. It donates a pair of hydrogen bonds to the backbone carbonyl of methionine 237 on the L3 loop and a bifurcated hydrogen bond to the backbone of proline 191 and the sidechain of serine-183 on the L2 loop. Hydrogen bonding to the carbonyl groups of the polypeptide backbone is probably essential for correct folding. Substitution of arginine by histidine is thought to perturb the geometry of these hydrogen bonds, although histidine is, in theory, capable of donating two hydrogen bonds and does not induce major clashes with neighboring residues. The functional consequences of two other substitutions at codon 175, R175P and R175G, have been evaluated in vitro. R175P has an intermediate DNA-binding activity (binding to WAF-1 but not BAX promoters), and R175G is apparently functionally equivalent to wild-type p53. Both of these substitutions are exceedingly rare in cancer, in agreement with the idea that they do not induce a strong mutant phenotype. However, it should be noted that both of these rare substitutions are predicted as "explained" according to the criteria used here (both of them break hydrogen bonding; in addition, R175P introduces a disallowed proline). In pure structural terms, substitution of arginine to histidine would be expected to be the mildest of the three mutations. Therefore, we can speculate that the functional impact of R175H is not so much a consequence of the loss of arginine than of the new structural constraints induced by the presence of a histidine at this position. These constraints might be related to the perturbation of zinc coordination by histidine, a good potential metal ligand.

The interpretation of the effects of R249S is more complex. According to Cho et al. [1994], this arginine, located within the L3 loop, makes at least four distinct contacts with other residues, including a van der Waals contact with the side chain of histidine-162, two hydrogen bonds to the backbones of methionine 246 and glycine-245, and a hydrogen bond with the sidechain of glutamine-171. Substitution of arginine by serine may maintain one of these hydrogen bonds, but this is probably not sufficient for the stabilization of the protein structure. It should be noted that this mutation is rare in cancers other than hepatocellular carcinoma, in which R249S mutations are thought to occur as a direct consequence of mutagenesis by aflatoxin B1, a potent dietary hepatocarcinogen. The other frequent mutation at codon 249, R249M, which is common in lung cancers, is "explained" according to the criteria used in the present study (disruption of hydrogen bonding). Thus, it cannot be ruled out that the effects of R249S reside more in the specific properties of this particular mutation in the context of liver cancer, than in the extent of the structural damage induced by this particular substitution.

Overall, these interpretations indicate that several important mutations in p53 cannot be "explained" using the simple purely structural criteria, and that the biological effect of a substitution is not necessarily proportional to the predicted extent of structural perturbation. A closer analysis of the biological properties of common mutations that escape structural "explanation" is likely to reveal interesting information for the understanding of the exact functions of these mutants.

In a second round of analysis, we have included in our explanatory criteria the degree of conservation of amino acids among species. While we were able to explain 34.5% of mutations resulting in substitutions in the core domain on purely structural grounds, if mutations

to 100% conserved amino acids are also included, we are able to "explain" 490 of the 882 distinct mutations (55.6%), representing 7,641 of the 9,824 missense mutations described (77.8%). These "explained" mutations now include all major mutations hotspots.

Of the unexplained mutations, it might be expected that the majority of these will be on the surface. Using a cutoff of 10% relative accessibility to classify a residue as exposed, we find that 236 of the 367 unexplained distinct mutations (64.31%) are exposed. These are prime candidates for interaction with other proteins or other domains of p53.

Note that our criteria for classifying a mutation as explained are fairly strict. For example we assume that any hydrogen-bonding sidechain substitution will be able to maintain the hydrogen bond if it has donor or acceptor capabilities the same as the parent; in practice, a structural change may be necessary.

In general, mutations explained on the basis of structure or conservation are ones which occur frequently in the p53 database while unexplained mutations are relatively rare events. A comparison of the distribution frequencies of the unexplained and explained mutations and of the unexplained mutations with mutations explained solely on conservation terms (see Fig. 8) shows a significant difference in these distributions (χ^2 test, $p < 10^{-7}$). As the figure shows, for rare mutations (one to 10 examples), approximately equal numbers are explained and unexplained. For more frequent mutations (11 or more examples), the number of explained mutations is always much larger than the unexplained. This confirms that we are better able to explain more frequently occurring mutations in structural terms. We also examined the frequency of mutation as a function of conservation between species. As can be seen in Figure 9 all the frequently mutated sites have high conservation scores; mutations at all sites with conservation scores < 0.85 are rare events.

Thus unexplained mutations fall into two categories. The first category includes mutations occurring very rarely in cancer. Of the 392 substitutions "not explained," 353 occur 10 times or fewer and 295 occur five times or fewer in the IARC TP53 database (each representing less than 0.1% of reported substitutions, including 295 substitutions occurring between one and five times). It is likely that these substitutions have no major functional impact, and that they may be detected in cancer solely as "passenger" mutations occurring in cell clones whose expansion is driven by mutations in other oncogenes and tumor suppressor genes. In their effect, these "passenger mutations" may be similar to silent mutations. The second category, comprising 39 substitutions occurring more than 10 times in the IARC database, are more common mutations likely either to be on the surface and involved in domain:domain or protein:protein interactions or are residues the structural importance of which we have not been able to identify automatically. The more common substitutions include V143A (15 occurrences), P151S (53 occurrences), R157F (108 occurrences), and M237I (103 occurrences). V143A is a temperature-sensitive mutant, which by definition has a very mild effect on protein structure, since the protein has a mutant phenotype at 37°C only and is capable of folding into the correct, wild-type conformation at 32°C. P151S and R157F are located in the L1 loop at the surface of the protein opposite the DNA binding side (P151 in the loop connecting beta strands S3 and S4, R157 in S4). S237I is located in loop L3, adjacent to C238, one of the zinc ligands. S237 is bridged by a hydrogen bond to R175. As for the substitution of the latter residue (see above), S237I may affect tertiary structure by preventing correct metal coordination.

Of the 353 rare unexplained mutations that occur 10 times or fewer and the 295 that occur five times or fewer, the majority (232 and 203 respectively) are exposed. Many of these may be of silent phenotype and are simply indicators of DNA damage. Only 39 of the 392 unexplained mutations are common events (occurring more than 10 times in the IARC p53 databank) and of these, just 10 are exposed: A138V (26 occurrences, 98.2% conservation), R156H (11 occurrences, 64.5% conservation), S166T (13 occurrences, 74.1% conservation), H168R (16 occurrences, 78.5% conservation), M169I (17 occurrences, 79.7% conservation), R181C (18 occurrences, 90.3% conservation), R181H (24 occurrences, 90.3% conservation), P190L (23 occurrences, 92.0% conservation), N235S (13 occurrences, 76.5% conservation), and P250L (25 occurrences, 98.0% conservation). It is tempting to speculate that this small set of residues may affect p53 function

through their effects on protein:protein interactions or are involved in contacts with other domains of p53 such as the C-terminal domain, which is important for the regulation of DNA-binding activities. This is particularly true for residues 138, 181, 190, and 250, all of which have conservation scores above 90%. The effect of the disruption of such interactions cannot be predicted on the basis of the crystal structures currently available. Indeed, arginine-181 in the R181H mutation lies at the outer surface of the protein and is known to be involved in interchain contacts between p53 monomers in the tetrameric structure which is required for high-affinity DNA binding.

Clearly the sidechain replacement assessment could be made much more sophisticated and this will be addressed in future work. A minimization procedure could be incorporated into the sidechain replacement together with a measure of the degree of bad contact rather than a simple yes/no assessment of clashes. Current work is investigating the effect of voids opened in the structure through substitutions of a large sidechain with a smaller one. In addition we could use X-Site scores [Laskowski et al., 1996] to assess the acceptability of sidechain replacements. Similarly, rather than simply assessing residues on the basis of ability to donate or accept hydrogen bonds, it would be possible to assess the geometry of replacements. Despite these limitations, the analysis presented here may help in future attempts to classify mutants in categories to evaluate their potential functional impact in cancer, as well as to identify mutations which may serve as clues for protein-protein interactions involved in the regulation of DNA-binding. In the future, we intend to apply the patch-analysis methodology of Jones and Thornton [1996, 1997] to identify regions of the protein surface likely to be involved in such interactions. In the long term, it is hoped that properties of p53 mutations, such as dominant negative activity, oncogenic potential, and temperature-sensitivity may be explained by classification of p53 mutations into structural groups whose molecular basis may then be analyzed.

A major role of this type of analysis is to suggest hypotheses that can be tested experimentally. We have suggested that 10 common mutations that occur on the surface of the p53 core domain may be involved in functions of p53 other than DNA binding through interactions with other domains or proteins and four of these are highly conserved between species. Patch analysis [Jones and Thornton, 1996; Jones and Thornton, 1997] could be used to strengthen this proposal before performing experiments to determine the effects of these mutations on DNA binding. We also propose that a number of rare mutations to surface residues have silent phenotypes and this can also be tested by in vitro assays. The structural effects of a number of these mutations could also be assayed through relatively rapid screening methods such as circular dichroism. From a functional perspective, the use of yeast-based assays [Maurici et al., 2001] has shown its usefulness as an experimental system to analyze the DNA-binding and transcriptional activities of various p53 mutants. Using such assays, it has been possible to identify mutants that have only partial loss of transcriptional activity. These assays may represent a very good experimental approach to test some of the predictions of the present study, in particular to detect mutants with transcriptionally silent phenotype.

We see this approach not only as a useful tool in examination of p53 mutations, but also as a paradigm for the study of many other diseases caused by point mutations such as favism caused by mutations in G6PD. In the near future, when structural data become available, it will become possible to apply the same forms of analysis to dystrophin, BRCA1—in all cases mutation databanks are available.

ACKNOWLEDGMENTS

This work was supported by a grant from the EC (QLG1-1999-00273 to T.H.-B., M.O., and P.H.) and by the UK Medical Research Council (Bioinformatics Studentship to A.L.C.).

REFERENCES

Aas T, Børresen AL, Geisler S, Smith-Sorensen B, Johnsen H, Varhaug J, Akslen LA, Lonning P. 1996. Specific p53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. Nat Med 2:811–814.

Ara S, Lee P, Hansen M, Saya H. 1990. Codon-72 polymorphism of the Tp53 gene. Nucleic Acids Res 18:4961.

Baker EN, Hubbard RE. 1984. Hydrogen bonding in globular proteins. Progr Biophy Mol Biol 44:97–179.

Brachmann RK, Eby KYY, Pavletich NP, Boeke JD. 1998. Genetic selection of intragenic suppressor mutations that

reverse the effects of common p53 cancer mutations. EMBO J 17:1847–1859.

Brash DE, Rudolph JA, Simon JA, Lin A, McKenna GJ, Baden HP, Halperin AJ, Ponten J. 1991. A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. Proc Natl Acad Sci USA 88:10124–10128.

Chao C, Saito S, Kang J, Anderson C, Appella E, Xu Y. 2000. p53 transcriptional activity is essential for p53-dependent apoptosis following DNA damage. EMBO J 19:4967–4975.

Chiba I, Takahashi T, Nau MM, D'Amico D, Curiel DT, Mitsudomi T, Buchhagen DL, Carbone D, Piantadosi S, Koga H, Reissman P, Slamon DJ, Holmes EC, Minna JD. 1990. Mutations in the p53 gene are frequent in primary, resected non-small cell lung cancer. Oncogene 5:1603–1610.

Cho Y, Gorina S, Jeffrey PD, Pavletich NP. 1994. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 265:346–355.

Clore GM, Ernst J, Clubb R, Omichinski JG, Kennedy WM, Sakaguchi K, Appella E, Gronenborn AM. 1995. Refined solution structure of the oligomerization domain of the tumour suppressor p53. Nat Struct Biol 2:321–333.

Crawford L. 1983. The 53,000-dalton cellular protein and its role in transformation. Int Rev Exp Pathol 25:1–50.

Culotta E, Koshland Jr DE. 1993. p53 sweeps through cancer research. Science 262:1958–1959.

Felley-Bosco E, Weston A, Cawley H, Bennett W, Harris C. 1993. Functional studies of a germ-line polymorphism at codon-47 within the p53 gene. Am J Hum Genet 53: 752–759.

Gannon JV, Greaves R, Iggo R, Lane DP. 1990. Activating mutations in p53 produce a common conformational effect. A monoclonal antibody specific for the mutant form. EMBO J 9:1595–1602.

Geisler S, Lonning PE, Aas T, Johnsen H, Fluge O, Haugen D, Lillehaug J, Akslen L, Børresen-Dale A. 2001. Influence of TP53 gene alterations and c-erbB-2 expression on the response to treatment with doxorubicin in locally advanced breast cancer. Cancer Res 61:2505–2512.

Gorina S, Pavletich NP. 1996. Structure of the p53 tumour suppressor bound to the ankyrin and SH3 domains of 53BP2. Science 274:1001–1005.

Greenblatt MS, Bennett WP, Hollstein M, Harris CC. 1994. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. Cancer Res 54:4855–4878.

Guinn BA, Padua RA. 1995. p53: a role in the initiation and progression of leukaemia? Cancer J 8:195–200.

Hainaut P, Hernandez T, Robinson A, Rodriguez-Tome P, Flores T, Hollstein M, Harris CC, Montesano R. 1998. IARC database of p53 gene mutations in human tumors and cell lines: updated compilation, revised formats and new visualisation tools. Nucleic Acids Res 26:205–213.

Harris CC. 1993. p53: at the crossroads of molecular carcinogenesis and risk assessment. Science 262:1980–1981.

Harris CC. 1996. p53 tumor suppressor gene: from the basic research laboratory to the clinic — an abridged historical perspective. Carcinogenesis 17:1187–1198.

Hernandez-Boussard T, Rodriguez-Tome P, Montesano R, Hainaut P. 1999. IARC p53 mutation database: a relational database to compile and analyze p53 mutations in human tumors and cell lines. International Agency for Research on Cancer. Hum Mutat 14:1–8.

Irwin MS, Kaelin WGJ. 2001. Role of the newer p53 family proteins in malignancy. Apoptosis 6:17–29.

Jeffrey PD, Gorina S, Pavletich NP. 1995. Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7Ångströms. Science 267:1498–1502.

Jones DT, Taylor WR, Thornton JM. 1994. A mutation data matrix for transmembrane proteins. FEBS Lett 339:269–275.

Jones S, Thornton JM. 1996. Principles of protein–protein interactions. Proc Natl Acad Sci USA 93:13–20.

Jones S, Thornton JM. 1997. Prediction of protein–protein interaction sites using patch analysis. J Mol Biol 272:133–143.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure. Biopolymers 22:2577–2637.

Ko L, Prives C. 1996. p53: puzzle and paradigm. Genes Dev 10:1054–1072.

Kraulis PJ. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Cryst 24:946–950.

Lakin N, Jackson S. 1999. Regulation of p53 in response to DNA damage. Oncogene 18:7644–7655.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291.

Laskowski R, Thornton JM, Humblet C, Singh J. 1996. X-SITE: Use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins. J Mol Biol 259:175–201.

Lee BK, Richards FM. 1971. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 55:379–400.

Levine AJ. 1997. p53, the cellular gatekeeper for growth and differentiation. Cell 88:323–331.

Li FP, Garber JE, Friend SH, Strong LC, Patenaude AF, Juengst ET, Reilly PR, Correa P, Fraumeni Jr JF. 1992. Recommendations on predictive testing for germ line p53 mutations among cancer-prone individuals. J Natl Cancer Instit 84:1156–1160.

Malkin D, Li FP, Strong LC, Fraumeni JF, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA, Friend SH. 1990. Germ line p53 mutations in a familial syndrome of breast-cancer, sarcomas, and other neoplasms. Science 250:1233–1238.

Matlashewski G. 1999. p53 twenty years on, meeting review. Oncogene 18:7618–7620.

Maurici D, Monti P, Campomenosi P, North S, Frebourg T,

Fronza G, Hainaut P. 2001. Amifostine (WR2721) restores transcriptional activity of specific p53 mutant proteins in a yeast functional assay. Oncogene 20:3533–3540.

May P, May E. 1999. Twenty years of p53 research: structural and functional aspects of the p53 protein. Oncogene 18:7621–7636.

McDonald IK, Thornton JM. 1994. Satisfying hydrogen-bonding potential in proteins. J Mol Biol 238:777–793.

Merritt EA, Bacon DJ. 1997. Raster3D: photorealistic molecular graphics. Meth Enzymol 277:505–524.

Michalovitz D, Halevy O, Oren M. 1991. p53 mutations — gains or losses. J Cell Biochem 45:22–29.

Mittl PR, Chene P, Grutter MG. 1998. Crystallization and structure solution of p53 (residues 326-356) by molecular replacement using an NMR model as template. Acta Crystallogr D54:86–89.

Nikolova PV, Wong KB, DeDecker B, Henckel J, Fersht AR. 2000. Mechanism of rescue of common p53 cancer mutations by second-site suppressor mutations. EMBO J 19:370–378.

North S, Hainaut P. 2000. P53 and cell-cycle: a finger in every pie. Pathol Biol 48:255–270.

Pavletich NP, Chambers KA, Pabo CO. 1993. The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots. Genes Dev 7:2556–2564.

Romano JW, Ehrhart JC, Duthu A, Kim CM, Appella E, May P. 1989. Identification and characterization of a p53 gene in a human osteosarcoma cell line. Oncogene 4:1483–1488.

Shenkin PS, Erman B, Mastrandrea LD. 1991. Information-theoretical entropy as a measure of sequence variability. Proteins: Struct Funct Genet 11:297–313.

Shih HL, Brady J, Karplus M. 1985. Structure of proteins with single-site mutations: a minimum perturbation approach. Proc Natl Acad Sci USA 82:1697–1700.

Sigal A, Rotter V. 2000. Oncogenic mutations of the p53 tumor suppressor: the demons of the guardian of the genome. Cancer Res 60:6788–6793.

Snow ME, Amzel LM. 1986. Calculating three-dimensional changes in protein structure due to amino acid substitutions: the variable domain of immunoglobulins. Proteins: Struct Funct Genet 1:276–279.

Srivastava S, Zou ZQ, Pirollo K, Blattner W, Chang EH. 1990. Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. Nature (London) 348:747–749.

Vogelstein B, Kinzler KW. 1992. p53 and dysfunction. Cell 70:523–526.

Vogelstein B, Kinzler KW. 1994. X-rays strike p53 again. Nature (London) 370:174–175.

Vogelstein B, Lane D, Levine A. 2000. Surfing the p53 network. Nature (London) 408:307–310.

Walker D, Bond J, Tarone R, Harris C, Makalowski W, Boguski M, Greenblatt M. 1999. Evolutionary conservation and somatic mutation hotspot maps of p53: correlation with p53 protein structural and functional features. Oncogene 18:211–218.

Wiederschain D, Gu J, Yuan ZM. 2001. Evidence for a distinct inhibitory factor in the regulation of p53 functional activity. J Biol Chem 276:27999–28005.

Wong KB, DeDecker BS, Freund SM, Proctor MR, Bycroft M, Fersht AR. 1999. Hot-spot mutants of p53 core domain evince characteristic local structural changes. Proc Natl Acad Sci USA 96:8438–8442.