Discrimination of Single Amino Acid Mutations of the p53 Protein by Means of Deterministic Singularities of Recurrence Quantification Analysis

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · MAY 2004

Impact Factor: 2.63 · DOI: 10.1002/prot.20075 · Source: PubMed

5 AUTHORS, INCLUDING:

Alessandro Porrello
University of North Carolina at Chapel Hill
27 PUBLICATIONS **1,161** CITATIONS
SEE PROFILE

Silvia Soddu
Istituto Regina Elena - Istituti Fisioterapici Os…
119 PUBLICATIONS **3,874** CITATIONS
SEE PROFILE

Marco Crescenzi
Istituto Superiore di Sanità
121 PUBLICATIONS **3,642** CITATIONS
SEE PROFILE

Alessandro Giuliani
Istituto Superiore di Sanità
362 PUBLICATIONS **4,461** CITATIONS
SEE PROFILE

PROTEINS: Structure, Function, and Bioinformatics 55:743–755 (2004)

Discrimination of Single Amino Acid Mutations of the p53 Protein by Means of Deterministic Singularities of Recurrence Quantification Analysis

Alessandro Porrello,[1,*] Silvia Soddu,[1] Joseph P. Zbilut,[2] Marco Crescenzi,[3] and Alessandro Giuliani[3]
[1]*Department of Experimental Oncology, Regina Elena Cancer Institute, Via delle Messi d'Oro, Rome, Italy*
[2]*Department of Molecular Biophysics and Physiology, Rush University, Chicago, Illinois*
[3]*Comparative Toxicology and Ecotoxicology Laboratory, Istituto Superiore di Sanitá, Viale Regina Elena, Rome, Italy*

ABSTRACT **p53 is mutated in roughly 50% of all human tumors, predominantly in the DNA-binding domain codons. Structural, biochemical, and functional studies have reported that the different p53 mutants possess a broad range of behaviors that include the elimination of the tumor-suppression function of wild-type protein, the acquisition of dominant-negative function over the wild-type form, and the establishment of gain-of-function activities. The contribution of each of these types of mutations to tumor progression, grade of malignancy, and response to anticancer treatments has been so far analyzed only for a few "hot-spots." In an attempt to identify new approaches to systematically characterize the complete spectrum of p53 mutations, we applied recurrence quantification analysis (RQA), a non-linear signal analysis technique, to p53 primary structure. Moving from the study of the p53 hydrophobicity pattern, which revealed important similarities with the singular deterministic structuring of prions, we could statistically discriminate, on a pure amino acid sequence basis, between experimentally characterized DNA-contact defective and conformational p53 mutants with a very high percentage of success. This result indicates that RQA is a mathematical tool particularly advantageous for the development of a database of p53 mutations that integrates epidemiological data with structural and functional categorizations. Proteins 2004;55:743–755.**

Key words: database; discriminant analysis; hydrophobicity; mutation; polymorphism; prion; recurrence quantification analysis (RQA)

INTRODUCTION

The product of the *p53* tumor-suppressor gene (*TP53*) is a transcription factor, functioning as a homotetramer, largely involved in the regulation of cell response to multiple stressing conditions and in the prevention of cancer development.[1–3] Its tumor-suppressing activities are lost in the large majority of human tumors. In a number of cases this inactivation is due to alterations of different mediators or regulators of p53 pathways (e.g. MDM2 overexpression, $p19^{ARF}$ inactivation, p53 cytoplasmic retention).[4–6] Alternatively, in about 50% of all human cancers, the cause can be a direct mutation of the *TP53* gene.[7,8] Remarkably, these mutations are missense in approximately 90% of cases.[9] In particular, amino acid substitutions have been observed for almost two-thirds of the 393 codons of the p53 protein, including all the residues of the DNA-binding domain (DBD) (amino acids 102–292), except codons 103 and 107, which are the targets of other kinds of mutations.[10,11] In contrast with this large variety of tumor mutations is the underrepresented number of functionally neutral polymorphisms at the protein level in the normal population.[10,11] Indeed, only one polymorphism at codon 47[12] and one at codon 72,[13,14] whose nature is very dubious due to evidence of contribution to gain-of-function activities (see below)[15] and to cancer chemotherapy modulation,[16] alter the amino acid sequence of the p53 protein. Altogether, these observations point to a very peculiar p53 organizational code, characterized by a very low level of redundancy.

At a clinical-epidemiological level, the presence of p53 mutations usually results in a prognosis and therapeutic response worse than in tumors expressing the wild-type p53 protein.[17–21] It has been reported that certain p53 mutations not only inactivate the wild-type tumor-suppressive functions, but also confer new protein activities. These new activities include the dominant negative function over the persisting product of the wild-type allele,[22] and the development of new, oncogenic functions, the so-called gain-of-functions.[23–26]A quite common criterion of classification is dependent on the subdivision of the p53 mutants affecting the DNA binding affinity into two classes. The first group is composed of mutants failing to bind the DNA due to the involvement of residues critical for DNA-protein contacts (e.g., at Arg^{248} and Arg^{273}). The

Grant sponsor: European Union; Grant number: QLG1-1999-00273; Grant sponsors: Associazione Italiana per la Ricerca sul Cancro, Fondo per gli Investimenti della Ricerca di Base, and Ministero della Salute of Italy; Grant sponsor: the National Science Foundation and National Institutes of Health; Grant number: DMS #0240230.

*Correspondence to: Alessandro Porrello, Department of Experimental Oncology, Regina Elena Cancer Institute, Via delle Messi d'Oro, 156, 00158 Rome, Italy. E-mail: porrello@ifo.it

Received 8 August 2003; Accepted 25 November 2003

Published online 5 March 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20075

second group is formed by mutants undergoing an impairment of the stable structure of the DBD (e.g., at Arg^{175} and His^{179}). These two classes were defined based on crystal structure[27] and biochemical studies.[28–34] Interestingly, *in vitro* cell resistance to anticancer drugs was found to correlate with the presence of the conformational but not of the DNA-contact defective mutants in the case of etoposide, and with mutants of both groups in the case of cisplatin.[25] Therefore, this group attribution is particularly relevant, since it can be linked to the cellular behavior of mutants, at least in specific cases, and has been proposed as a basis for distinct clinical approaches.[35] Overall, the complex behaviors of the numerous p53 mutants suggest that a systematic analysis should be undertaken to specify their role in the etiology, pathogenesis, prognosis, and therapeutic response with respect to cancer. Nevertheless, an extensive evaluation of the structural and functional characteristics of all the mutations found in human cancers cannot be actually proposed for the very expensive and demanding work required. To overcome this problem, we asked whether it was possible to categorize the p53 mutations by means of mathematical models based on pure sequence information. From an information theory standpoint proteins can be viewed as "slightly edited strings",[36] with their sequential amino acid arrangements departing not much from randomness. Randomness, in this case, is intended in terms of substantial equivalence between the native protein sequences and their scrambled counterparts whose amino acid ordering is specified by a random-number generator algorithm. Remarkably, the only chemico-physical property showing a statistically significant non-random order is hydrophobicity. This point has been determined by means of mathematical tools such as the measure of mean-square fluctuation of hydrophobicity along the chain and of the distribution of total hydrophobicity.[37] The uniqueness of hydrophobicity is probably due to the crucial role played by the peculiar patterns of juxtaposition of hydrophobic/hydrophilic residues in protein folding.[38] Moreover, a considerable body of research has evidenced the importance played in protein structures of repeating amino acid motifs,[39–41] as well as of complex sequence periodicities.[42] Thus, we hypothesized that the unusual "fragility" of the p53 phenotype to mutations comes from a much more regular—thus more prone to be altered—pattern of hydrophobicity distribution with respect to other proteins.

In order to quantitatively assess the amount of the above mentioned regularity, use was made of recurrence quantification analysis (RQA), a computational technique that has given important results in the characterization of proteins.[43–45] Here we show that the p53 hydrophobicity pattern is very similar to the prion protein (PrP) in terms of marked departure from randomness, at odds with the majority of proteins but consistently with the low p53 polymorphism. Moving from these findings, we asked whether it was possible to approach the problem of defining the characteristics of tumor-associated p53 mutants by means of RQA. This technique was able to statistically discriminate, with a very high percentage of success, the previously mentioned p53 mutations of the residues directly involved in the DNA-protein contacts and those affecting the correct folding of the p53 DBD.[27] This result suggests that RQA is an appropriate tool for the development of a comprehensive p53 database, characterized by information about the structure and function of listed mutations.

THEORY AND COMPUTATION

The Model

RQA is a relatively new non-linear signal analysis technique, especially suited for short, non-stationary numerical sequences,[46–48] successfully applied to the sequence of the amino acid hydrophobicities of proteins.[43–45] (Hydrophobicity values are coded by the logarithm of the octanol-water partition coefficient).[49] The numerical sequence corresponding to the protein primary structure is initially submitted to an n-dimensional embedding by the delay method.[50] The embedding procedure consists in building an n-column embedding matrix (EM) out of the original numerical string, by shifting along the sequence of a fixed lag. Hence, the EM rows correspond to subsequent windows of length n along the sequence, and for the RQA applications must be considered like vectors of an n-dimensional space. The number of columns of the EM corresponds to the embedding dimension (emb) employed. For example, given the sequence 10, 11, 21, 32, 41, 35, 40, 19, . . . and choosing a shifting-lag (lag) = 1 (an integer number compatible with its discrete character), the corresponding four-dimensional (i.e., emb = 4) EM is:

10	11	21	32
11	21	32	41
21	32	41	35
32	41	35	40
41	35	40	19
35	40	19	.
40	19	.	.
19	.	.	.

In particular, the choice of using emb = 4 for general RQA applied to proteins is guided by our previous work on this field[45] and by the calculation of 4 as the minimum value to extract the maximum information content encoded in the entire protein sequences.[51]

From the EM, RQA works out the distance matrix (DM), which reports the Euclidean distances between all the above defined vectors.[46] In particular, every element $\mathbf{X}_{i,j}$ of the DM, which is a square matrix, corresponds to the Euclidean distance between the i-th and j-th EM rows/vectors. At this processing stage, it is highlighted the maximum Euclidean distance (maxdist) among all the calculated distances. Finally, the DM is transformed into a recurrence plot (RP), constituted only by black and white points, located at specific $\mathbf{X}_{i,j}$ coordinates.[52] RP pixels are darkened when the corresponding distance values of DM are lower than a pre-defined radius (rad) and remain white otherwise. The features of the distance function make both the DM and the RP symmetric ($\mathbf{X}_{i,j} = \mathbf{X}_{j,i}$) with the latter showing a darkened main diagonal, corresponding to the

identity line (in the DM, $\mathbf{X}_{i,j} = 0$, when $i = j$) [see Fig. 1 (c)]. The darkened points (i.e., the recurrent points) single out recurrences within the sequence and the RP can be considered as a global picture of the self-correlation structure of the hydrophobicity distribution along the chain.[45] Though the same graphic appearance of the RP is often informative, as assessed by the classification of different RPs made by a set of unbiased observers,[53] a quantitative description has been developed by Webber and Zbilut.[46] They defined a collection of variables for RQA that includes the four indices selected for the computations of this article:

1. Recurrence percentage (REC): corresponding to the percentage of recurrent points over all the RP points or, equivalently, to the percentage of row couples of the EM whose Euclidean distance is below the chosen similarity threshold (rad).
2. Determinism percentage (DET): the percentage of recurrent points that form diagonal line structures, of two points or more, in the RP. The smallest number of points necessary to define one of these structures is imposed by the value of the variable line. DET is a measurer of the "patch" amount of similar hydrophobic/hydrophilic characteristics along the sequence.
3. Entropy (ENT): is defined as the output of the Shannon-Weaver formula for the information entropy, calculated from the histogram of line lengths, and measures the richness of deterministic structure inside the RP.[46]
4. TREND: this index corresponds to the paling of the recurrence structure going away from the main RP diagonal (identity line) and is the average of linear regression coefficients expressing REC as a function of the distance from every point of the identity line. This descriptor is linked to the relative non-stationary character of the sequence.[54]

DET Cross-Recurrence

While the above guidelines allow measuring all the possible similarities between residues (self-correlation), this procedure compares the DET of a single short sequence with that of all the equally long windows of consecutive amino acids of the full protein. At the nth step, the EMs both of the selected sequence and of the nth window are compared, the mutual distances between all their rows are calculated, the RP is created, and the corresponding DET is computed. Finally, every DET obtained is represented on the Y-axis of a graph, where each position on X-axis corresponds to the last amino acid of the selected window of the protein (see Fig. 2).[55] Therefore, this graph shows the amount of resemblance (in terms of hydrophobicity patterns) between the selected sequence and the entire protein.

Two-Dimensional and Three-Dimensional DET Scaling

The DET of a sequence is dependent on the cut-off threshold imposed by means of rad. The two-dimensional DET scaling is the graphical representation of the DET of the full-length sequence as a function of every possible rad, going from its minimum (= 0) to its maximum (= maxdist) level (see Fig. 4). The two-dimensional DET scaling usually consists of a curve in the DET/rad plane, which converges asymptotically and relatively smoothly to 100%, owing to the increase of recurrent points proportional to rad. In contrast, for metastable proteins and peptides,[56–58] it was shown that, at low rad, there is an unexpected clear divergence, with a DET spike, followed by a sudden fall and the subsequent beginning of the expected scaling.[59] This phenomenon is called a singularity, while the interval of maxdist values in which this behavior is observed is the singularity range.[45]

The three-dimensional DET scaling assumes a window of predefined number of consecutive amino acids (in our case 36), which we call an epoch, and repeats the two-dimensional DET scaling for all the possible epochs. It is synthetically represented by a three-dimensional graph having as axes rad, amino acid, and DET (see Fig. 5), and allows for a localization of the sequence portions of amino acids more probably involved in generating the two-dimensional DET scaling peak.[59] Additionally, the subtraction of a three-dimensional graph from another permits to estimate how big is the mutual difference between their deterministic profiles [see Fig. 6 (a, b)].[59,60]

Local ("Sliding Window") RQA Protocol for Mutant Discrimination

An epoch of 36 amino acids moves along the hydrophobicity sequence of the wild type protein with lag = 1. For each of these epochs, the four RQA variables are computed and their graphs are drawn, where on the X-axis are all the epoch positions and on the Y-axis the corresponding RQA variables [see Fig. 7 (A)]. The same procedure is repeated for each mutant protein [see Fig. 7 (B)]. For each mutation the difference between homologous (i.e., representing the same RQA variable) graphs is computed [= (wild type protein) − (mutant protein)] [see Fig. 7 (C)], as well as the point by point sums of the 36 discrete values composing each difference graph.[43] Since these graphs depend on the settled threshold (rad), this procedure is repeatedly performed in the full range 0–50% maxdist. Finally, a matrix is generated with as many rows as mutants and with four columns, corresponding to the four selected RQA variables (REC, DET, ENT, and TREND). These are tabulated as the integrals of the functions associated to the previously calculated values, in the interval 0–50% maxdist.

All the matrix values are submitted to statistical analysis, as described below.

Statistical Analysis

The first statistical analysis to be performed was the computation of the relative "distance from randomness" of examined proteins. This point was approached by generating a collection of twenty different scrambled versions for each protein and comparing basic RQA variables (REC and DET) of both native and scrambled sequences. The dependence of REC and DET on the amino acid order along

the chain was assessed by fixing, as a criterion of significance, a threshold of 3 standard deviations (SDs) of the difference between (D) of REC [REC D = (REC of native protein) − (average value of REC of scrambled proteins)] and DET [DET D = (DET of native protein) − (average value of DET of scrambled proteins)].

The second statistical analysis was directed toward the generation of a quantitative model discriminating p53 mutations acting by i) directly modifying residues involved in DNA-protein contacts and ii) affecting DNA binding by a general rearrangement of DBD structure. Consequently, a discriminant analysis[61] was performed, based on the first principal component[61] of the four selected RQA variables (REC, DET, ENT, TREND)[60] [see Table II (a), columns 2–5]. Discriminant analysis is a statistical procedure commonly used to classify cases into the values of a categorical dependent variable by means of multiple regression. For all the mutations, the normalized difference between mutant and wild type protein was obtained to generate principal component scores [see Table II (a), column 6]. Then, for each mutant a group prediction resulted [see Table II (a), column 8] based on a discriminant score corresponding to the Mahalanobis distance from the "center of gravity" of the complete set of mutants [see Table II (a), column 9].[61] Finally, Wilks's lambda,[61] a statistical test of multivariate Analysis of Variance, was applied to test the significance of the obtained discrimination between the two groups [see Table II (b)] .

The statistical analysis was performed by means of *StatView* (SAS Institute, Inc., Cary, North Carolina).

General Set-Up, Graphical Conventions and Software for RQA

Values for RQA variables were as follows: Figures 1, 2, 3: emb = 4, line = 3, and rad = 3. The same values of emb and line were used in Figure 4, where rad varied as a percentage of maxdist, in the range 0-100. Figures 5, 6, and 7: emb = 3 and line = 2, with rad = 0–10% maxdist (Figs. 5 and 6) and = 0–50% maxdist (Fig. 7), respectively. The values chosen are dictated by the general (Figs. 1, 2, 3, and 4) and local (Figs. 5, 6, and 7) analyses, respectively.[60] For purposes of amino acid indexing of "sliding windows," the values obtained for the amino acids of the entire window are attributed to its last amino acid, in the graphic representation. The software for RQA is available at the web site http://homepages.luc.edu/~cwebber/.

Selection of p53 Mutants From p53 Database

A full and up-to-date p53 mutation list (final version of the database used: R8) was downloaded from the International Agency for Research on Cancer, Lyon, France (http://www-p53.iarc.fr/p53DataBase.htm). The mutations of the p53 DBD to be processed by RQA were selected on the basis of their frequency in tumors[10,11] and of their proper characterization, as assessed by structural[27] and biochemical[28–34] studies.

RESULTS

The Hydrophobicity Pattern of p53 Reveals a Strong Deterministic Structure that Matches the p53 Lack of Polymorphisms

An analysis of the hundreds of independent mutations of p53[10,11] has evidenced the virtual lack of any physiologically silent mutations, at the protein level, in contrast with the very high rate of tumor-associated ones. Therefore, we have hypothesized that p53 possesses a self-organization characterized by a conspicuously deterministic structure, despite the apparent lack of motif regularity in its amino acid sequence, with the exception of the P(X)nP repetitions of the proline rich domain (PRD) and of some amino acid couplets and triplets. To test this hypothesis, the p53 sequence of 393 amino acids was converted into a sequence of 393 hydrophobicity values and represented by a graph [Fig. 1 (a)]. This operation evidenced one region, approximately corresponding to the PRD (amino acids 58–98) [Fig. 1 (b)], where the values are included in a tight hydrophobicity interval. As expected, no further information could be obtained and no evident correlation was directly deducible among p53 regions or sub-domains by this approach. At the same time, the p53 RP [Fig. 1 (c)] showed an abundance of point aggregations (i.e., deterministic islands), which was over the mean levels of a standard set of proteins (our unpublished results). In particular, the largest deterministic island roughly corresponded to the amino acid positions of the PRD. To evaluate whether the minor deterministic islands are in relation with the PRD, we used the DET cross-correlation algorithm, by assuming the PRD as a reference-island. This computation, obtained by shifting along the chain a representative PRD subsequence (amino acids 68–79), showed that six out of the about twenty main deterministic islands have a pattern similar to the p53 PRD (Fig. 2).

Notably, the distribution of the deterministic islands of p53 has a high probability to be altered by mutations, compared to more common protein patterns with fewer, smaller, and less interrelated islands. In fact, whereas periodic structures of the RPs have many chances to be disorganized by amino acid substitutions, their aperiodic structures tend to remain aperiodic after analogous changes. Therefore, from a theoretical perspective, the point distribution of the p53 RP is associated with a marked susceptibility to mutations and consequently matches the p53 lack of polymorphisms.

p53, But Not its Family Members, Has a DET Comparable to Proteins With Diverse Kinds of Structural Complexity

Secondary and super-secondary structures have been demonstrated to have a direct correlation with the structuring of the hydrophobicity pattern.[62,63] Nonetheless, loosely folded proteins displaying a low degree of structural complexity, exhibit markedly ordered patterns of hydrophobicity distribution.[64,65] This point was demonstrated by an analysis of 1,141 proteins randomly selected from a Swiss-Prot repository (http://www.expasy.org/sprot/), which showed a prominent correlation between low structural

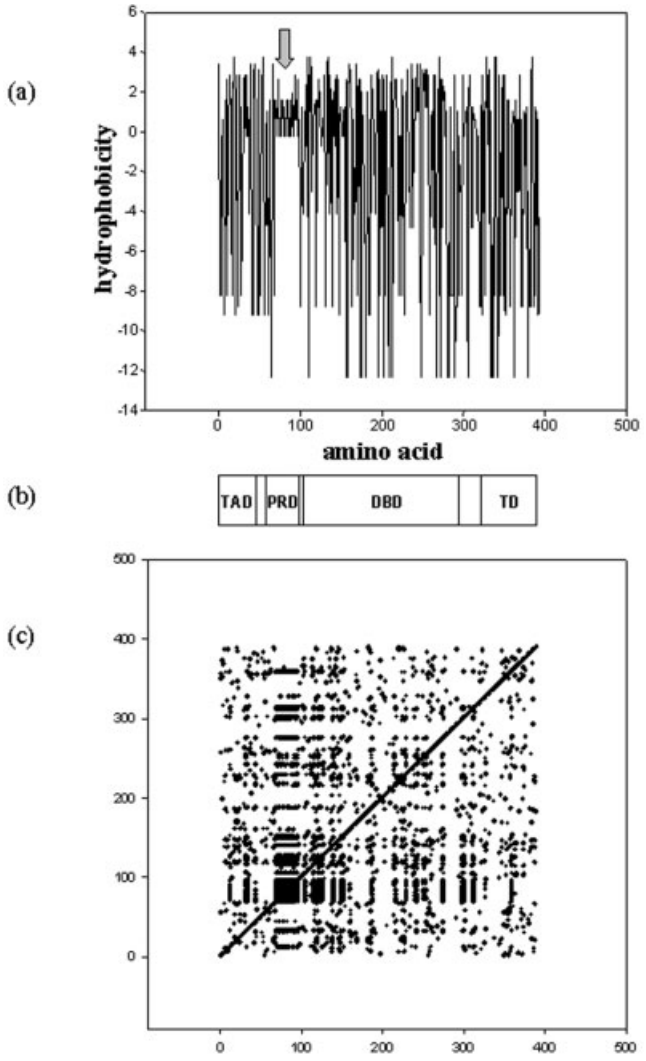

Fig. 1. Hydrophobicity graph and recurrence plot of p53. The p53 primary structure is converted into the corresponding sequence of hydrophobicity values and represented by a graph. On the X-axis are the amino acid positions and on the Y-axis are the relative hydrophobicity values. The arrow points to the PRD, corresponding to a deterministic maximum (**a**). A rectangle that illustrates the principal domains of the p53 protein (**b**). Recurrence plot of the p53 sequence (**c**). DBD, DNA-binding domain; PRD, proline rich domain; TAD, transactivation domain; TD, tetramerization domain.

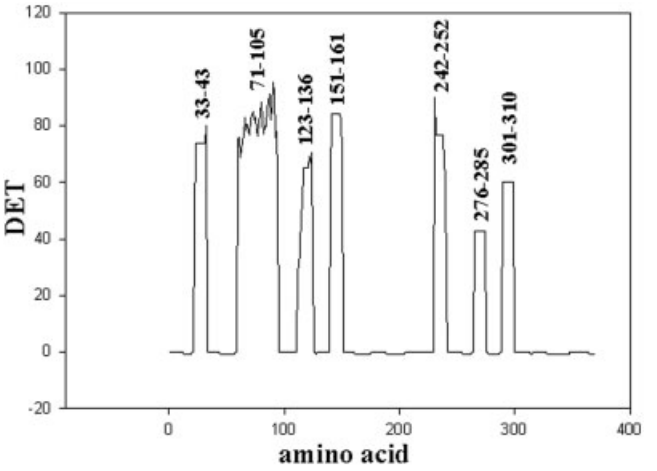

Fig. 2. DET cross-recurrence of the 68–79 "patch" of p53. Extraction of deterministic sequences of the p53 protein, (highlighted vertically), compared to a reference "patch" (amino acids 68–79) of the p53 PRD, by shifting a single position at a time along the complete p53 hydrophobicity sequence. On the X-axis are the amino acid positions and on the Y-axis are the values of DET. DET, determinism percentage.

order and high DET (our unpublished results). The p53 protein is in a critical position, because it was described as intrinsically disordered[65,66] with large unstructured regions, but also with a non-marginal presence of α-helixes and β-sheets in its secondary structure.[67] In order to compare the DET of p53 (= 42.14) with some reference examples, we chose: the alpha chain of hemoglobin (Hb-α), the hemagglutinin neuraminidase (HN) of Sendai virus, and the PrP of the Syrian hamster. Hb-α is a globular protein with a rich and homogeneous secondary structure, and with a well-defined tertiary structuring.[68] HN of Sendai virus is a membrane protein with a constraining hydrophobicity pattern due to its location, partially intramembrane (hydrophobic phase) and partially intracellular (hydrophilic phase).[69] PrP is a cellular protein that can be responsible for either genetic or acquired diseases. In particular, the pathogenic protein (PrP-Sc) derives from an irreversible conformational change of the normal isoform (PrP-C).[70] PrP was reported to belong to the group of loosely folded proteins[64,65] and has a zone of high deterministic structuring that is crucial for protein–protein aggregation.[59] The PrP sequence examined (amino acids 90–231) by RQA did not include the two regions physiologically removed in the cell (amino acids 1–22 and 232–254) and the five copper-binding octarepeat sequence.[71] This choice took into account both the documented importance of this protein fragment[72,73] and the necessity to eliminate the technical over-weighting of PrP DET, due to the presence of the octarepeats, without losing primary information on the sequence. For the sake of brevity, we will refer to this truncated form simply as PrP. As shown in Table I, column 5, the computed value of p53 DET is comparable to these three controls.

Recently, p53 was found to be member of a protein family that includes a series of different spliced forms of *p63* and *p73* genes. These members present sequence homologies and a certain degree of functional parallelism.[74,75] Thus, we asked whether also the largest isoforms of p63 (i.e., p63α) and p73 (i.e., p73α) present a high DET like p53. Notably, the DET values give a description of p53 as a member apart (the most deterministic one) in its homology family, while the other two members have a good mutual affinity (Table I, column 5), in accordance with the experimental evidence.[76]

The RQA Demonstrates a Strong Parallelism Between the Hydrophobicity Patterns of p53 and PrP

To clarify whether the source of high p53 DET was prevalently the presence of secondary and super-second-

TABLE I. REC and DET General Statistics[†]

1 Protein	2 Native REC	3 Average REC	4 REC SD	5 Native DET	6 Average DET	7 DET SD
HN	3.58	3.64	0.22	40.54	37.75	2.88
Hb-α	4.37	4.60	0.34	44.79	40.24	6.73
p63α	2.37	2.05	0.25	35.93	29.54	2.52
p73α	3.24	2.85	0.20	34.64	33.14	1.46
p53 (*)	2.44	1.99	0.17	42.14	26.86	2.63
PrP (*)	4.37	3.48	0.34	47.36	31.84	3.03

[†]The values of REC and DET variables of the native proteins (columns 2 and 5) are flanked by the corresponding averages (columns 3 and 6) and standard deviations (SD) (columns 4 and 7) of their scrambled counterparts. HN, hemagglutinin neuraminidase; Hb-α, alpha chain of hemoglobin; PrP, prion protein. While the difference in REC between native and scrambled sequences is under our significance threshold (= 3 SDs) for all the tabulated proteins, the analogous difference in DET reaches the statistical significance for both p53 and PrP (*).

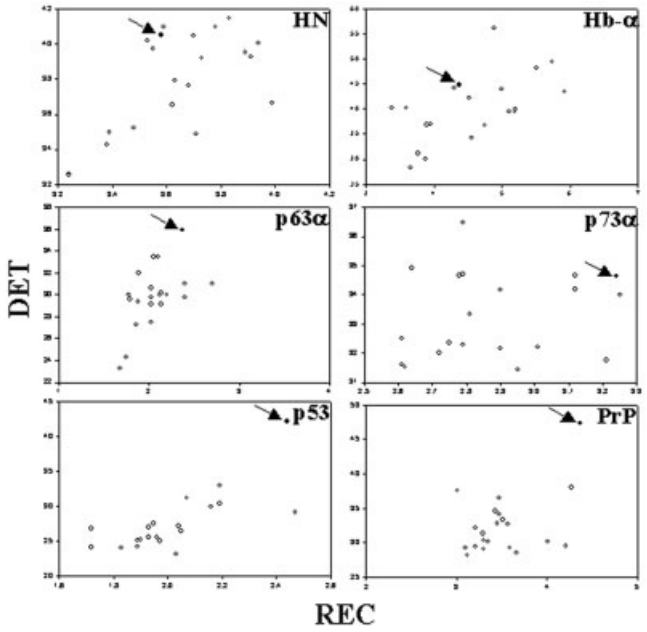

Fig. 3. Exploration of the REC/DET plane. The filled circles correspond to the sequences of native structures, while the empty circles represent 20 scrambled counterparts for each original sequence. X-axis: REC; Y-axis: DET. The arrows help to localize the circles corresponding to native structures. Hb-α, alpha chain of hemoglobin; HN, hemagglutinin neuraminidase; PrP, prion protein.

ary structure (as in the case of Hb-α and HN), or of sequence motifs that create a complex DET peak (like for PrP), we tested REC and DET properties of p53 and controls. Specifically, to quantify the amount of order-dependent information conveyed by the REC and DET variables, we compared their absolute values, relatively to the original sequences, with the average REC and DET of series of twenty scrambled counterparts. Figure 3 gives a graphical appreciation of the difference between native and scrambled counterparts in the REC/DET plane, while the average values gathered from raw data are summarized in Table I. The REC D was under the significance threshold of 3 SDs for all the tabulated proteins ($0 \leq$ REC D ≤ 3 SDs) (Table I, columns 2, 3, 4), thus showing that their REC is scarcely sequence-related. Similar results were obtained for the DET D in the case of Hb-α, HN, p63α and p73α ($0 \leq$ DET D ≤ 3 SDs) (Table I, columns 5, 6, 7), as well as for a long list of other proteins (unpublished results), due to the poor structuring generally borne by protein sequences.[36] In contrast, p53 and PrP have a DET D over the very selective threshold imposed (DET D ≥ 3 SDs) (Table I, columns 5, 6, 7). Altogether, these results demonstrate that the variable DET of both p53 and PrP is markedly dependent on the amino acid sequence. Hence, the variable DET is a sensitive marker for a separation of these two proteins from other controls, based only on the deterministic disposition of their hydrophobicity patterns.

In previous works, we demonstrated that the presence of irreversible conformational switches was related to deterministic singularities in the RQA profile.[59,60] These switches probably depend on the elevated flexibility of highly deterministic (repetitive) patches.[39–42] PrP satisfies this deterministic requirement as well as computer-drawn peptides, designed to optimize their ability to make conformational changes.[58] Remarkably, the two-dimensional DET scaling curve of p53 [Fig. 4 (a)] shows the same short-range peak singularity of PrP [Fig. 4 (b)], at odds with the other controls [Fig. 4 (c)]. Overall, these outcomes support the hypothesis that p53 and PrP deterministic patterns depend on similar structural causes,[77] and are in accordance with the PrP poverty of "true" polymorphisms[78–80] (see Discussion).

The Deterministic Pattern of p53 is Mainly Dependent on its Proline Rich Domain

To localize the protein sites with the best probabilities for producing such a two-dimensional DET scaling singularity, we performed a three-dimensional DET scaling of the p53 protein (Fig. 5). Even considering the indeterminacy linked to the epoch length of 36 residues, the presence of a rugged landscape evidences the existence of "hot" regions for hydrophobicity structuring. By analogy with our previous observations on PrP, the definition of these multiple sites allows for a delimitation of protein portions putatively connected to metastability.[59] Specifically, in the first half of the protein we found a sequence (amino acids 81–122), which overlapped rather well with the p53 PRD. Since PRD is also the main deterministic island, extensively interrelated with the minor ones, we concluded that PRD plays a primary role in raising the overall periodicity

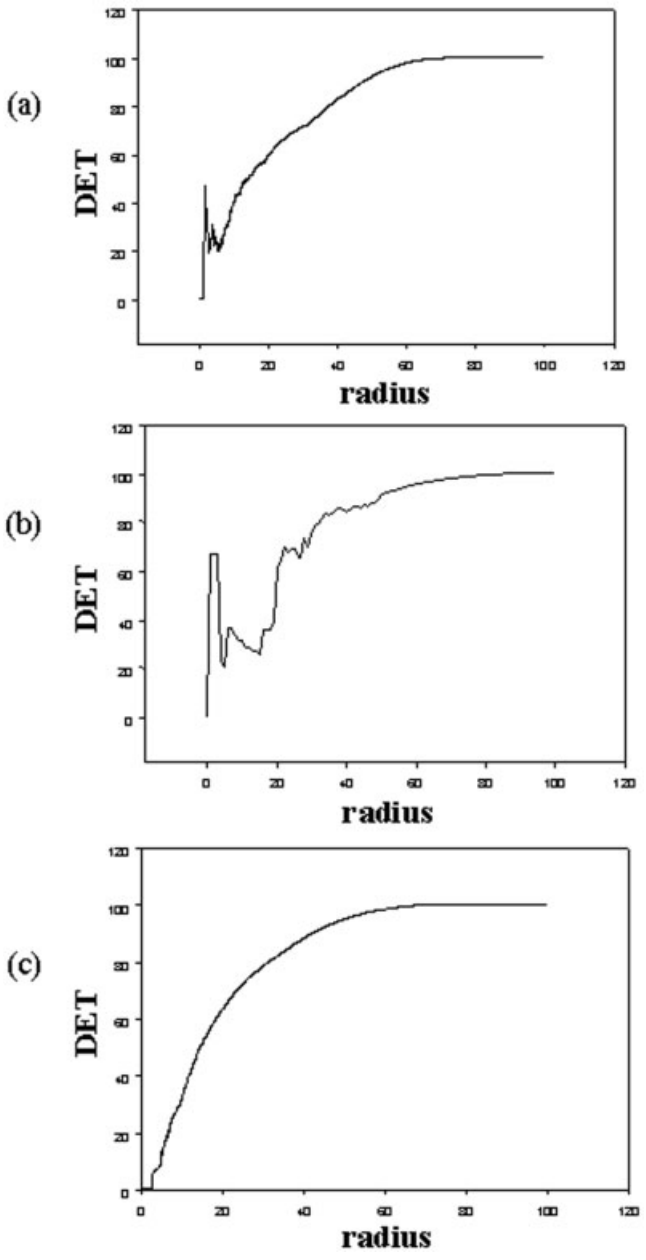

Fig. 4. Two-dimensional DET scaling of p53. The rad, expressed as percentage of maxdist (X-axis), is progressively increased to achieve the saturation of DET (Y-axis). Note the DET peak (singularity) at low rad of p53 (**a**) and PrP (**b**), compared with the regular scaling of Hb-α (**c**), here reported as representative of the other controls. DET, determinism percentage; maxdist, maximum Euclidean distance; rad, radius.

of the p53 RP, thus greatly contributing in settling the properties of its deterministic pattern. The implications of this centrality are examined in the Discussion.

Mutational "Reactivity" and Mutant Discrimination of the p53 Protein

The above analogies between p53 and PrP suggested an examination of the systematic effects played by p53 mutations. Indeed, one of the PrP peculiarities is the ability to significantly modify the three-dimensional DET scaling even for single amino acid substitutions.[59] The amount of the perturbation induced by each mutation can be graphically appreciated by subtracting the three-dimensional DET scaling graph of the mutated from that of the wild-type protein and evaluating the pseudo-volumetric change. Significantly, this procedure has evidenced a p53 situation resembling that of PrP [Fig. 6 (a, b)], because they both show important variations produced by mutations.

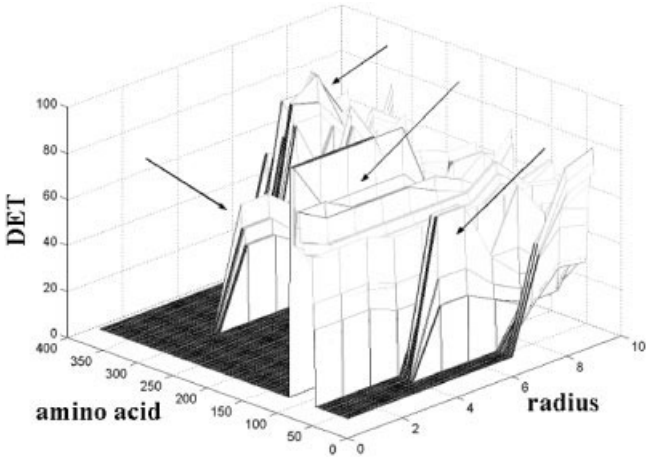

Fig. 5. Three-dimensional DET scaling of p53. DET (Z-axis) is plotted for the different amino acidic residues (X-axis) and for the low levels (= 0–10% maxdist) of rad (Y-axis). Arrows, pointing to examples of singularities where a "shelf" of maximum DET diverges downward, show that the principal source of the p53 singularity is PRD. DET, determinism percentage; rad, radius.

The holistic effects of mutations, whereby even single amino acid substitutions can interfere with the entire energy "landscape" of proteins, were investigated in studies on the structural distribution of cooperative interactions in proteins and on relationships between point mutations and protein sequence variability.[81] In particular, it was shown that the broad range effect of amino acid substitutions is not due to the creation of new conformers, but rather to the setting up of a new dynamic equilibrium among the pre-existing ones.[82] Consequently, we have hypothesized that some proteins, like p53 and PrP, are more prone than others to these conformer rearrangements and that the amount of three-dimensional DET scaling perturbation is a good indicator of the protein mutational "reactivity." Furthermore, we have postulated that RQA is intrinsically able to single out when a protein receives, by single or multiple mutations, a weak or strong impulse to the redistribution of the sub-states in the space of conformers. To test these hypotheses, we decided to study p53 mutations by using the "sliding window" RQA protocol for mutant discrimination (Fig. 7).

The p53 mutant collection subjected to our analysis consisted of 32 mutants, whose sites are located in the DBD. By crystallographic and biochemical analyses, mutants have been subdivided into two groups, failing to bind DNA owing to: i) the interference with the amino acids involved in direct contacts with DNA (group 1); ii) the

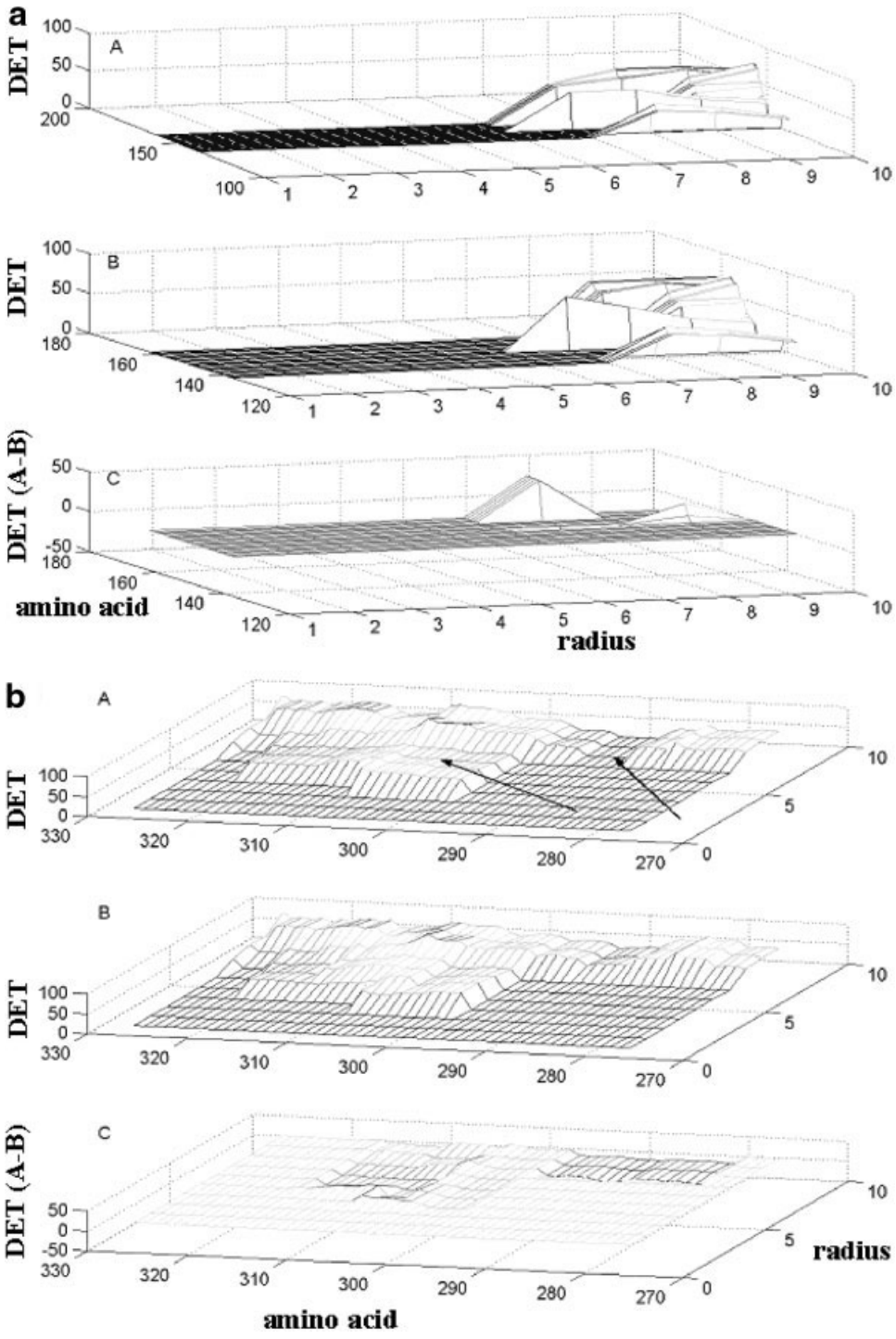

Fig. 6. Variations of p53 three-dimensional DET scaling induced by single amino acid mutations. Representation of mutation V143L. Note the appearance of a positive pseudo-volumetric variation (**a**). Representation of mutation R280S. It is appreciable the development of a complex process, involving the simultaneous appearance of a positive plus negative pseudo-volumetric variation, indicated by arrows (**b**). A = wild type p53; B = mutated p53; C = relative change. X-axis: amino acidic residues; Y-axis: radius (rad); Z-axis (panels A and B): DET; Z-axis (panel C): DET difference (A-B). Both the representations are focused in the neighborhood of the mutated position.

impairing of a stable folding of the DBD (group 2).[27–34] We preliminarily observed that: i) the site of the examined mutations is a unique domain; this imposition makes improbable a successful separation by RQA only due to potential differences among domains; ii) this domain is the best known and the richest in cancer-associated mutations;[10,11] iii) all the sites selected have a high mutation frequency in cancer;[10,11] iv) the two groups are biologically well characterized; indeed, this classification is universally recognized by the researchers in this field;[27] v) the total number (= 32) and the group division of the examined mutations allow to perform a valid statistical study. The local RQA was able to discriminate (see Statistical Analysis of Theory and Computation) between the two groups with only two minor misclassifications (R248W and R249M) [Table II (a)]. Moreover, the absolute values obtained for RQA variables allowed for the determination that global perturbations (group 2) are estimated to be

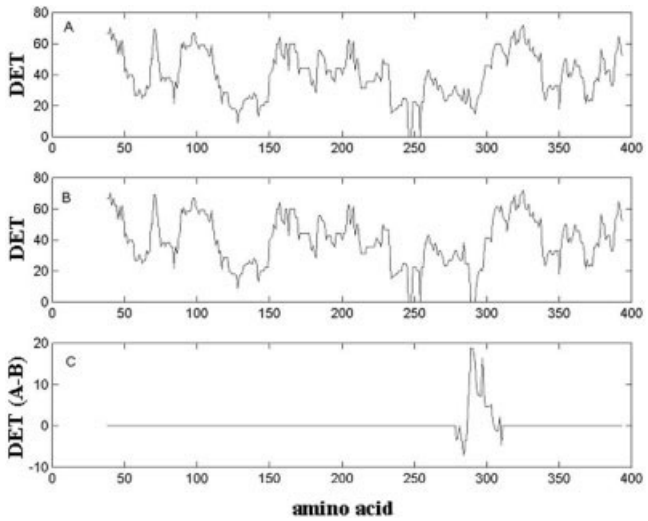

Fig. 7. Calculation of the single values to be processed by the discriminant analysis for local RQA protocol. A "sliding window" of 36 amino acids, with lag = 1, emb = 3, line = 2, and rad varying in the range 0–50% maxdist, was utilized to calculate every RQA variable, for both wild-type (**A**) and mutated p53 (**B**), showing the resulting change (**C**). X-axis: amino acidic residues; Y-axis (A and B): DET; Y-axis (C): DET difference (A-B). The three graphs reported are completely descriptive of this procedure, lacking only in the final step of integration in the interval 0–50% maxdist. In particular, the values shown are referred to DET of the mutation R273H, with rad = 10% maxdist.

larger than local perturbations (group 1), as results from the distribution of normalized global differences between mutant and wild type p53s [Table II (a), column 6]. Remarkably, the discriminant analysis performed (overall percentage of success: 94%) was associated with a high statistical significance ($p = 0.0127$), as assessed by Wilks's lambda statistics [Table II (b)].

Therefore, RQA is able to quantify the amount of perturbation given to p53 by its amino acid mutations, gathering the mutants with similar properties together. This result suggests that even particularly complex information on structure and biochemical functions of the p53 protein can be deduced by RQA, like in this case, from the mere knowledge of its amino acid sequence.

DISCUSSION

The very numerous, different mutations of the p53 protein found in a large number of human cancers have raised a series of questions concerning their biological role, clinical relevance and systematic classification. Consequently, RQA was here employed as an algorithmic tool to deal with two relevant related issues: the lack of p53 polymorphisms at the protein level and the high relevance of a global understanding of its mutants' behavior. At the same time, the strong similarity of their deterministic patterns (see Results) supported, by statistical analysis, suggests a parallelism between the p53 protein and PrP. This result strengthens a theoretical prion-tailored model of p53 describing some implications of this analogy in the set of p53 molecules present within a cell,[83] and is in accordance with the inclusion in the same group of disordered proteins (subjected to disorder to order transitions) for p53 and PrP.[64–66] Moreover, this finding links up with a work describing the capability of the p53 tetramerization domain to form amyloid-like fibrils *in vitro*.[84]

The only two p53 polymorphisms observed in human beings are at amino acid 47 (P47S) and at amino acid 72 (P72R), with a population incidence low (1.5%) and variable in a wide range, respectively.[12–14] Moreover, this very small number of two cases probably should be reduced to only one since it was shown that p53 mutants with a P72R polymorphism have an enhanced tumorigenic potential[15] and a lower response rate to cancer chemotherapy.[16] All this is in strong contrast with Chasman and Adams, who studied the polymorphisms affecting protein function[85] and whose work allows for a demonstration that the ratio polymorphisms/total mutations is generally in the range 0.68–0.74, while for p53 is ≤ 0.01.[10,11] Even taking into account recent calls for careful evaluation of the linkage with cancer of some p53 mutations,[9] the peculiarity of this ratio still remains. Notably, besides the closeness of their hydrophobicity patterns, the poverty of polymorphisms seems to be an additional contact point between p53 and PrP. Indeed, the same neutrality of PrP polymorphisms of the most studied animal species is extremely questionable, since it was described that polymorph PrPs are more susceptible to the transformation PrP-C→PrP-Sc.[78–80] As shown (see first paragraph of the Results), the RQA interpretation of this phenomenon, relative to p53, is that it necessarily comes from the elevated structuring of its hydrophobicity pattern (highly deterministic), whose organization can be easily perturbed by amino acid substitutions,[60] leading to pathological states. It is conjecturable that the vulnerable structure of the p53 protein is at least partially dependent on its peculiarity as a transcription factor, in which the integrity of the whole DBD, the p53 domain that holds the majority of mutations,[10,11] may be fundamental to achieving a proper function. Indeed, the p53 DBD does not possess specific short motifs which are present in other transcription factors, such as basic domains (e.g., Jun, Fos, Myc, E2A, E2F), helix-turn-helix domains (e.g., NK2, HNF-1, HOXs), or zinc-coordinating domains (e.g., the various steroid hormone receptors), but founds its activity on the presence of a wide β-scaffold structure.[27] Interestingly, this is also consistent with the model recently sketched by Harrison et al.[86] of the structural requirement for a β-sheet rich organization to display a prion-like behavior in terms of progressive aggregation of different protein molecules in supra-molecular assemblies. Specifically, we make the hypothesis that, as the β-structures of PrP are essential for the propagation of the pathogenic conformation, the β-structures of p53 DBD may exert a crucial role to determine the assembly of active p53 tetramers, in addition to the functions already attributed to the p53 tetramerization domain.[87] and references therein

The research on p53 structure is intrinsically complex because this protein is quite resistant to crystallization,[88] probably due to its loosely folded character.[65,66] Since the p53 PRD is at the NH_2 end, which has been only partially described at the structural level,[89,90] an interesting path of

TABLE II (a). Synopsis of the Discriminant Analysis†

1 Examined mutation	2 REC	3 DET	4 ENT	5 TREND	6 NORM	7 Group attribution	8 Predicted group	9 Score
R273H	597.002	3603.801	170.430	−11704.6	6.036497	1	1	1.625618
R273C	593.671	3636.181	170.430	−11493.1	6.124909	1	1	1.175096
R273L	568.585	3484.831	165.893	−9973.65	6.128953	1	1	−0.783510
R273P	593.036	3690.668	177.694	−12362.1	6.223346	1	1	1.986002
R248Q	571.758	3467.867	165.893	−10200.4	6.065271	1	1	−0.360440
R248W	563.985	3332.220	152.893	−10366.3	5.908349	1	2	−0.439170
R248L	572.872	3321.537	164.893	−10989.5	5.798044	1	1	1.120357
R280K	560.012	3449.469	164.397	−10426.9	6.159634	1	1	−0.653370
R280S	578.746	3191.738	154.912	−12858.2	5.514920	1	1	2.899235
R280T	576.362	3231.554	154.912	−12308.3	5.606813	1	1	2.177023
R280I	591.126	3346.853	150.198	−11894.6	5.661827	1	1	1.534451
R280G	577.154	3242.741	158.392	−12193.8	5.618502	1	1	2.250293
R175H	574.777	3691.499	159.016	−10549.8	6.422489	2	2	−1.236240
R175S	579.701	4251.989	171.499	−12364.3	7.334797	2	2	−1.239130
R175G	579.701	4251.989	171.499	−12364.3	7.334797	2	2	−1.239130
R175L	572.083	3932.348	170.499	−12050.4	6.873737	2	2	−0.669310
R175C	578.111	3952.224	159.499	−11209.6	6.836445	2	2	−1.556450
R175P	581.605	4224.913	171.499	−12089.4	7.264231	2	2	−1.266740
V143A	570.493	3459.155	169.244	−10886.5	6.063449	2	2	0.321786
V143L	569.380	3515.258	169.244	−10298.1	6.173835	2	2	−0.453680
V143M	570.493	3459.155	169.244	−10886.5	6.063449	2	2	0.321786
H179Y	560.174	3117.434	159.394	−8030.84	5.565117	2	2	−1.937550
H179Q	570.014	3468.329	165.893	−10303.6	6.084638	2	2	−0.383180
H179L	570.968	2827.386	138.699	−9808.70	4.951917	2	2	0.839041
H179N	570.014	3468.329	165.893	−10303.6	6.084638	2	2	−0.383180
H179R	580.650	3346.772	170.647	−10615.2	5.763837	2	2	1.012197
H179D	570.014	3468.329	165.893	−10303.6	6.084638	2	2	−0.383180
R249S	565.416	3546.519	154.893	−10272.3	6.272407	2	2	−1.130820
R249W	564.147	3555.407	154.893	−10614.4	6.302271	2	2	−0.941780
R249M	576.682	3548.548	166.893	−10439.6	6.153388	2	1	−0.115650
R249G	567.478	3521.965	155.893	−10202.3	6.206346	2	2	−0.959550
R249T	565.416	3546.519	154.893	−10272.3	6.272407	2	2	−1.130820

†Columns: type of amino acid substitution (1); RQA differential descriptors (2–5); normalized global difference (NORM) from wild-type p53 (6); experimentally attributed group (7); RQA predicted group (8); score of the discriminant analysis (9). The score utilized is the Mahalanobis distance from the "center of gravity" of the full set of mutants and is normalized to 0 (corresponding to "indeterminate") with variance = 1.

TABLE II (b). Percentages of Correct Group Attribution and Statistical Significance†

	Cases attributed to group 1	Cases attributed to group 2	Cases correctly attributed
DNA-protein contact defective mutants (group 1)	11	1	92%
Conformational defective mutants (group 2)	1	19	95%
Total cases	12	20	94%

†Statistical significance of the classification reported in Table II (a), as measured by Analysis of Variance: Wilks's lambda statistics $F = 3.6294$ ($p = 0.0127$).

research is expected to be the study of the implications descending from the PRD cross-correlation with the sequences corresponding to the minor deterministic islands. Furthermore, the RQA finding that the high values of DET are directly related to intrinsically disordered sequences, which are important candidates to play a role in protein-protein interactions,[64,65] suggests that p53 PRD may be crucial for the binding between p53 and other proteins. In particular, a relevant open question could be to verify experimentally if this domain is involved in the direct interaction of p53 with p300/CBP complexes,[91,92] in analogy with what happens to SAD, the PRD of the tumor-suppressor Smad4.[93]

CONCLUSIONS

The biologists and physicians who are trying to define the behavior of the single p53 mutants must face relevant difficulties put by both the large mutation number and the complex experiments required in vitro and in vivo. This research points, in the long term, to the creation of a

database, which associates each mutant with its biological peculiarities, potentially helpful for the selection of different therapy protocols for oncological patients, based on the *TP53* gene status. Clearly, this effort would be lessened if it were possible to cluster "a priori" the hundreds of p53 mutations recorded. Therefore, this target has become a challenging test for bio-mathematical methods such as RQA, moving from the rich public databases of p53 (http://p53.curie.fr/; http://www-p53.iarc.fr/p53DataBase.htm). In particular, our discrimination study was directed towards the DBD because this domain holds the greatest number of cancer-related mutations. The results obtained encourage applying the local RQA protocol to the entire collection of mutants, in order to analyze the grade of correlation between biological/clinical response and grade/magnitude of RQA perturbation. Regarding this proposal, the peculiar RQA profile of p53 allowed for improving the applications of our method. Indeed, the "sliding window" RQA, originally conceived as an instrument to separate allowed from non-allowed mutations of TEM-1 β-lactamases,[43] was here exploited for a discrimination between two classes of mutations, obtaining a very high percentage of correct classification. As a matter of fact this is, to our knowledge, the first study on p53 that was able to perform statistically significant predictions of the biological activity of mutants based only on amino acid sequence information. It is interesting to observe that the two misclassifications (R248W and R249M) are both situated at the COOH end of the large loop three sub-domain, (one of the three large loops of this domain), whose side chains determine its whole correct spatial orientation. Therefore, we hypothesize that the prominence of these two mutations on "hot spot" codons is linked to the ability of interfering with the proper protein function even with perturbations that stay under the separation threshold of our method, at least as currently performed.

It was recently demonstrated that it is possible to counteract the action of mutations by second-site suppressor mutations carried by peptides[94,95] or by direct rescue of the single site mutated proteins.[96] Nonetheless, it was shown in the Results that RQA variables not only strongly correlate with energetic rearrangements, as statistically demonstrated in this work, but also that this is a positive correlation. Therefore, the heavier the structure perturbation, the bigger the modification of RQA variables. Altogether, these data support the idea that the best "normalizing" peptides or additive mutations might be planned by minimizing the difference between the RQA variables of wild type and double-mutated sequences, as assessed by our local protocol. Remarkably, the relative simplicity of this approach, potentially applicable to personalized protocols in cancer therapies, might favor a direct translation of RQA theoretical models into clinical investigation.

ACKNOWLEDGMENTS

This work was partially supported by European Union grant QLG1-1999-00273, Associazione Italiana per la Ricerca sul Cancro, Fondo per gli Investimenti della Ricerca di Base, and Ministero della Salute of Italy. This material is also based upon a joint DMS/DGMS initiative to support mathematical biology, from the National Science Foundation and National Institutes of Health, under Grant Number NSF DMS #0240230 (A.G. and J.P.Z.). A.P. is recipient of a fellowship from Fondazione Italiana per la Ricerca sul Cancro. The authors thank A. Sacchi of the Regina Elena Cancer Institute (Rome) for her kind contribution in the revision of the manuscript.

REFERENCES

1. Levine AJ. p53, the cellular gatekeeper for growth and division. Cell 1997;88:323–331.
2. Agarwal ML, Taylor WR, Chernov MV, Chernova OB, Stark GR. The p53 network. J Biol Chem 1998;273:1–4.
3. Balint EE, Vousden KH. Activation and activities of the p53 tumour suppressor protein. Br J Cancer 2001;85:1813–1823.
4. Cordon-Cardo C, Latres E, Drobnjak M, Oliva MR, Pollak D, Woodruff JM, Marechal V, Chen J, Brennan MF, Levine AJ. Molecular abnormalities of mdm2 and p53 genes in adult soft tissue sarcomas. Cancer Res 1994;54:794–799.
5. Sherr CJ. The INK4a/ARF network in tumour suppression. Nat Rev Mol Cell Biol 2001;2:731–737.
6. Moll UM, Ostermeyer AG, Haladay R, Winkfield B, Frazier M, Zambetti G. Cytoplasmic sequestration of wild-type p53 protein impairs the G1 checkpoint after DNA damage. Mol Cell Biol 1996;16:1126–1137.
7. Harris CC. 1995 Deichmann Lecture—p53 tumor suppressor gene: at the crossroads of molecular carcinogenesis, molecular epidemiology and cancer risk assessment. Toxicol Lett 1995;82–83: 1–7.
8. Oren M. Regulation of the p53 tumor suppressor protein. J Biol Chem1999;274:36031–36034.
9. Soussi T, Dehouche K, Béroud C. p53 website and analysis of p53 gene mutations in human cancer: forging a link between epidemiology and carcinogenesis. Hum Mutat 2000;15:105–113.
10. Hernandez-Boussard T, Rodriguez-Tome P, Montesano R, Hainaut P. IARC p53 mutation database: a relational database to compile and analyze p53 mutations in human tumors and cell lines. Hum Mutat 1999;14:1–8.
11. Olivier M, Eeles R, Hollstein M, Khan MA, Harris C, Hainaut P. The IARC TP53 database: new online mutation analysis and recommendations to users. Hum Mutat 2002;19:607–614.
12. Felley-Bosco E, Weston A, Cawley HM, Bennet WP, Harris CC. Functional studies of a germ-line polymorphism at codon 47 within the p53 gene. Am J Hum Genet 1993;53:752–759.
13. Harris N, Brill E, Shohat O, Prokocimer M, Wolf D, Arai N, Rotter V. Molecular basis for heterogeneity of the human p53 protein. Mol Cell Biol 1986;6:4650–4656.
14. Beckman G, Birgander R, Sjalander A, Saha N, Holmberg PA, Kivela A, Beckman L. Is p53 polymorphism maintained by natural selection? Hum Hered 1994;44:266–270.
15. Marin MC, Jost CA, Brooks LA, Irwin MS, O'Nions J, Tidy JA, James N, McGregor JM, Harwood CA, Yulug IG, Vousden KH, Allday MJ, Gusterson B, Ikawa S, Hinds PW, Crook T, Kaelin WG Jr. A common polymorphism acts as an intragenic modifier of mutant p53 behaviour. Nat Genet 2000;25:47–54.
16. Bergamaschi D, Gasco M, Hiller L, Sullivan A, Syed N, Trigiante G, Yulug I, Merlano M, Numico G, Comino A, Attard M, Reelfs O, Gusterson B, Bell AK, Heath V, Tavassoli M, Farrell PJ, Smith P, Lu X, Crook T. p53 polymorphism influences response in cancer chemotherapy via modulation of p73-dependent apoptosis. Cancer Cell 2003;3:387–402.
17. Wattel E, Preudhomme C, Hecquet B, Vanrumbeke M, Quesnel B, Dervite I, Morel P, Fenaux P. p53 mutations are associated with resistance to chemotherapy and short survival in hematologic malignancies. Blood 1994;84:3148–3157.
18. Lowe SW. Cancer therapy and p53. Curr Opin Oncol 1995;7:547–553.
19. Taubert H, Meye A, Wurl P. Prognosis is correlated with p53 mutation type for soft tissue sarcoma patients. Cancer Res 1996;56:4134–4136.
20. Powell B, Soong R, Iacopetta B, Seshadri R, Smith DR. Prognostic significance of mutations to different structural and functional

regions of the p53 gene in breast cancer. Clin Cancer Res 2000;6:443–451.
21. Liu MC, Gelmann EP. P53 gene mutations: case study of a clinical marker for solid tumors. Semin Oncol 2002;29:246–257.
22. Oren M. p53: the ultimate tumor suppressor gene? FASEB J 1992;6:3169–3176.
23. Dittmer D, Pati S, Zambetti G, Chu S, Teresky AK, Moore M, Finlay C, Levine AJ. Gain of function mutations in p53. Nat Genet 1993;4:42–46.
24. Zambetti GP, Levine AJ. A comparison of the biological activities of wild-type and mutant p53. FASEB J 1993;7:855–865.
25. Blandino G, Levine AJ, Oren M. Mutant p53 gain of function: differential effects of different p53 mutants on resistance of cultured cells to chemotherapy. Oncogene 1999;18:477–485.
26. Pugacheva EN, Ivanov AV, Kravchenko JE, Kopnin BP, Levine AJ, Chumakov PM. Novel gain of function activity of p53 mutants: activation of the dUTPase gene expression leading to resistance to 5-fluorouracil. Oncogene 2002;21:4595–4600.
27. Cho Y, Gorina S, Jeffrey PD, Pavletich NP. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 1994;265:346–355.
28. Milner J, Cook A, Sheldon M. A new anti-p53 monoclonal antibody, previously reported to be directed against the large T antigen of simian virus 40. Oncogene 1987;1:453–455.
29. Bartek J, Iggo R, Gannon J, Lane DP. Genetic and immunochemical analysis of mutant p53 in human breast cancer cell lines. Oncogene 1990;5:893–899.
30. Gannon JV, Greaves R, Iggo R, Lane DP. Activating mutations in p53 produce a common conformational effect. A monoclonal antibody specific for the mutant form. EMBO J 1990;9:1595–1602.
31. Hinds PW, Finlay CA, Quartin RS, Baker SJ, Fearon ER, Vogelstein B, Levine AJ. Mutant p53 DNA clones from human colon carcinomas cooperate with ras in transforming primary rat cells: a comparison of the "hot spot" mutant phenotypes. Cell Growth Differ 1990;1:571–580.
32. Raycroft L, Schmidt JR, Yoas K, Hao MM, Lozano G. Analysis of p53 mutants for transcriptional activity. Mol Cell Biol 1991;11: 6067–6074.
33. Stephen CW, Lane DP. Mutant conformation of p53. Precise epitope mapping using a filamentous phage epitope library. J Mol Biol 1992;225:577–583.
34. Hainaut P, Milner J. A structural role for metal ions in the "wild-type" conformation of the tumor suppressor protein p53. Cancer Res 1993;53:1739–1742.
35. Rolley N, Butcher S, Milner J. Specific DNA binding by different classes of human p53 mutants. Oncogene 1995;11:763–770.
36. Weiss O, Herzel H. Correlations in protein sequences and property codes. J Theor Biol 1998;190:341–353.
37. Irback A, Sandelin E. On hydrophobicity correlations in protein chains. Biophys J 2000;79:2252–2258.
38. Li H, Tang C, Wingreen NS. Nature of driving force for protein folding: a result from analyzing the statistical potential. Phys Rev Lett 1997;79:765–768.
39. Sedgwick SC, Smerdon SJ. The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem Sci 1999;24:311–316.
40. Kajava AV. Review: proteins with repeated sequence—structural prediction and modeling. J Struct Biol 2001;134:132–144.
41. Sudol M, Sliwa K, Russo T. Functions of WW domains in the nucleus. FEBS Lett 2001;490:190–195.
42. Rackovsky, S. "Hidden" sequence periodicities and protein architecture. Proc Natl Acad Sci USA 1998;95:8580–8584.
43. Zbilut JP, Giuliani A, Webber CL Jr, Colosimo A. Recurrence quantification analysis in structure-function relationships of proteins: an overview of a general methodology applied to the case of TEM-1 beta-lactamase. Protein Eng 1998;11:87–93.
44. Giuliani A, Benigni R, Sirabella P, Zbilut JP, Colosimo A. Nonlinear methods in the analysis of protein sequences: a case study in rubredoxins. Biophys J 2000;78:136–149.
45. Giuliani A, Benigni R, Zbilut JP, Webber CL Jr, Sirabella P, Colosimo A. Nonlinear signal analysis methods in the elucidation of protein sequence-structure relationships. Chem Rev 2002;102: 1471–1492.
46. Webber CL Jr, Zbilut JP. Dynamical assessment of physiological systems and states using recurrence plot strategies. J Appl Physiol 1994;76:965–973.
47. Filligoi G, Felici F. Detection of hidden rhythms in surface EMG signals with a non-linear time-series tool. Med Eng Phys 1999;21: 439–448.
48. Rustici M, Caravati C, Petretto E, Branca M, Marchettini N. Transition scenarios during the evolution of the Belousov-Zhabotinsky reaction in an unstirred batch reactor. J Phys Chem A 1999;103:6564–6570.
49. Palliser CC, Parry DA. Quantitative comparison of the ability of hydropathy scales to recognize surface beta-strands in proteins. Proteins 2001;42:243–255.
50. Broomhead DS, King GP. Extracting qualitative dynamics from experimental data. Physica D 1986;20:217–236.
51. Weiss O, Jimenez-Montano MA, Herzel H. Information content of protein sequences. J Theor Biol 2000;206:379–386.
52. Trulla LL, Giuliani A, Zbilut JP, Webber CL Jr. Recurrence quantification analysis of the logistic equation with transients. Physics Lett A 1996;223:255–260.
53. Giuliani A, Sirabella P, Benigni R, Colosimo A. Mapping protein sequence spaces by recurrence quantification analysis: a case study on chimeric structures. Protein Eng 2000;13:671–678.
54. Giuliani A, Colafranceschi M, Webber CL Jr, Zbilut JP. A complexity score derived from principal components analysis of nonlinear order measures. Physica A 2001;301:567–588.
55. Zbilut JP, Giuliani A, Webber CL Jr. Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. Physics Lett A 1998;246:122–128.
56. Micheletti C, Seno F, Maritan A, Banavar JR. Protein design in a lattice model of hydrophobic and polar amino acids. Phys Rev Lett 1998;80:2237–2240.
57. Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, Ramponi G, Dobson CM. Kinetic partitioning of protein folding and aggregation. Nat Struct Biol 2002;9:137–143.
58. Abkevich VI, Gutin AM, Shakhnovich EI. Theory of kinetic partitioning in protein folding with possible applications to prions. Proteins 1998;31:335–344.
59. Zbilut JP, Webber CL Jr, Colosimo A, Giuliani A. The role of hydrophobicity patterns in prion folding as revealed by recurrence quantification analysis of primary structure. Protein Eng 2000;13: 99–104.
60. Zbilut JP, Sirabella P, Giuliani A, Manetti C, Colosimo A, Webber CL Jr. Review of nonlinear analysis of proteins through recurrence quantification. Cell Biochem Biophys 2002;36:67–87.
61. Steel RGD, Torrie JH, Dickey, DA. Principles and procedures of statistics: A biometrical approach, 3rd edition. New York: Mc Graw-Hill; 1996. 672 p.
62. Selz KA, Mandell AJ, Shlesinger MF. Hydrophobic free energy eigenfunctions of pore, channel, and transporter proteins contain beta-burst patterns. Biophys J 1998;75:2332–2342.
63. Grigoriev IV, Kim SH. Detection of protein fold similarity based on correlation of amino acid properties. Proc Natl Acad Sci USA 1999;96:14318–14323.
64. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42:38–48.
65. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002;41: 6573–6582.
66. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002;323:573–584.
67. Bell S, Klein C, Müller L, Hansen S, Buchner J. p53 contains large unstructured regions in its native state. J Mol Biol 2002;322:917–927.
68. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT. Structure of haemoglobin. A three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. Nature 1960;185:416–422.
69. Bowen HA, Lyles DS. Structure of Sendai viral proteins in plasma membranes of virus-infected cells. J Virol 1981;37:1079–1082.
70. Prusiner SB. Prions. Proc Natl Acad Sci USA 1998;95:13363–13383.
71. Liao YC, Lebo RV, Clawson GA, Smuckler EA. Human prion protein cDNA: molecular cloning, chromosomal mapping, and biological implications. Science 1986;233:364–367.
72. James TL, Liu H, Ulyanov NB, Farr-Jones S, Zhang H, Donne DG, Kaneko K, Groth D, Mehlhorn I, Prusiner SB, Cohen FE. Solution structure of a 142-residue recombinant prion protein correspond-

ing to the infectious fragment of the scrapie isoform. Proc Natl Acad Sci USA 1997;94:10086–10091.
73. Baskakov IV, Legname G, Baldwin MA, Prusiner SB, Cohen FE. Pathway complexity of prion protein assembly into amyloid. J Biol Chem 2002;277:21140–21148.
74. Irwin MS, Kaelin WG. p53 family update: p73 and p63 develop their own identities. Cell Growth Differ 2001;12:337–349.
75. Strano S, Rossi M, Fontemaggi G, Munarriz E, Soddu S, Sacchi A, Blandino G. From p63 to p53 across p73. FEBS Lett 2001;490:163–170.
76. De Laurenzi V, Melino G. Evolution of functions within the p53/p63/p73 family. Ann NY Acad Sci 2000;926:90–100.
77. Lee C, Park SH, Lee MY, Yu MH. Regulation of protein function by native metastability. Proc Natl Acad Sci USA 2000;97:7727–7731.
78. O'Doherty E, Aherne M, Ennis S, Weavers E, Roche JF & Sweeney T. Prion protein gene polymorphisms in pedigree sheep in Ireland. Res Vet Sci 2001;70:51–56.
79. Schroder B, Franz B, Hempfling P, Selbert M, Jurgens T, Kretzschmar HA, Bodemer M, Poser S, Zerr I. Polymorphisms within the prion-like protein gene (Prnd) and their implications in human prion diseases, Alzheimer's disease and other neurological disorders. Hum Genet 2001;109:319–325.
80. Billinis C, Panagiotidis CH, Psychas V, Argyroudis S, Nicolaou A, Leontides S, Papadopoulos O, Sklaviadis T. Prion protein gene polymorphisms in natural goat scrapie. J Gen Virol 2002;83 (Pt. 3):713–721.
81. Hilser VJ, Dowdy D, Oas TG, Freire E. The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. Proc Natl Acad Sci USA 1998;95:9903–9908.
82. Sinha N, Nussinov R. Point mutations and sequence variability in proteins: redistributions of preexisting populations. Proc Natl Acad Sci USA 2001;98:3139–3144.
83. Blagosklonny MV. p53 from complexity to simplicity: mutant p53 stabilization, gain-of-function, and dominant-negative effect. FASEB J 2000;14:1901–1907.
84. Lee AS, Galea C, DiGiammarino EL, Jun B, Murti G, Ribeiro RC, Zambetti G, Schultz CP, Kriwacki RW. Reversible amyloid formation by the p53 tetramerization domain and a cancer-associated mutant. J Mol Biol 2003;327:699–709.
85. Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol 2001;307:683–706.
86. Harrison PM, Chan HS, Prusiner SB, Cohen FE. Conformational propagation with prion-like characteristics in a simple model of protein folding. Protein Sci 2001;10:819–835.
87. Brokx RD, Bolewska-Pedyczak E, Gariepy J. A stable human p53 heterotetramer based on constructive charge interactions within the tetramerization domain. J Biol Chem 2003;278:2327–2332.
88. Friend S. p53: a glimpse at the puppet behind the shadow play. Science 1994;265:334–335.
89. Botuyan MV, Momand J, Chen Y. Solution conformation of an essential region of the p53 transactivation domain. Fold Des 1997;2:331–342.
90. Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, Chang J, Sung YC, Choi KY, Han KH. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. J Biol Chem 2000;275:29426–29432.
91. Grossman SR, Perez M, Kung AL, Joseph M, Mansur C, Xiao ZX, Kumar S, Howley PM, Livingston DM. p300/MDM2 complexes participate in MDM2-mediated p53 degradation. Mol Cell 1998;2:405–415.
92. Wadgaonkar R, Collins T. Murine double minute (MDM2) blocks p53-coactivator interaction, a new mechanism for inhibition of p53-dependent gene expression. J Biol Chem 1999;274:13760–13767.
93. de Caestecker MP, Yahata T, Wang D, Parks WT, Huang S, Hill CS, Shioda T, Roberts AB, Lechleider RJ. The Smad4 activation domain (SAD) is a proline-rich, p300-dependent transcriptional activation domain. J Biol Chem 2000;275:2115–2122.
94. Selivanova G, Iotsova V, Okan I, Fritsche M, Strom M, Groner B, Grafstrom RC, Wiman KG. Restoration of the growth suppression function of mutant p53 by a synthetic peptide derived from the p53 C-terminal domain. Nat Med 1997;3:632–638.
95. Selivanova G, Ryabchenko L, Jansson E, Iotsova V, Wiman KG. Reactivation of mutant p53 through interaction of a C-terminal peptide with the core domain. Mol Cell Biol 1999;19:3395–3402.
96. Nikolova PV, Wong KB, DeDecker B, Henckel J, Fersht, AR. Mechanism of rescue of common p53 cancer mutations by second-site suppressor mutations. EMBO J 2000;19:370–378.