Published in final edited form as:

Biochim Biophys Acta. 2013 April ; 1834(4): 725–738. doi:10.1016/j.bbapap.2013.01.012.

Intrinsically disordered regions of p53 family are highly diversified in evolution

Bin Xue¹, Celeste J. Brown², A. Keith Dunker³, and Vladimir N. Uversky^{1,4,5,*}

¹Department of Molecular Medicine, College of Medicine, University of South Florida, Tampa, FL 33612, USA

²Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844-3051, USA

³Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA

⁴Byrd Alzheimer's Research Institute, College of Medicine, University of South Florida, Tampa, FL 33612, USA

⁵Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

Abstract

Proteins of the p53 family are expressed in vertebrates and in some invertebrate species. The main function of these proteins is to control and regulate cell cycle in response to various cellular signals, and therefore to control the organism's development. The regulatory functions of the p53 family members originate mostly from their highly-conserved and well-structured DNA-binding domains. Many human diseases (including various types of cancer) are related to the missense mutations within this domain. The ordered DNA-binding domains of the p53 family members are surrounded by functionally important intrinsically disordered regions. In this study, substitution rates and propensities in different regions of p53 were analyzed. The analyses revealed that the ordered DNA-binding domain is conserved, whereas disordered regions are characterized by high sequence diversity. This diversity was reflected both in the number of substitutions and in the types of substitutions to which each amino acid was prone. These results support the existence of a positive correlation between protein intrinsic disorder and sequence divergence during the evolutionary process. This higher sequence divergence provides strong support for the existence of disordered regions in p53 *in vivo* for if they were structured, they would evolve at similar rates as the rest of the protein.

Keywords

intrinsically disordered proteins; protein evolution; protein-protein interactions; protein-DNA interaction; p53 family

© 2012 Elsevier B.V. All rights reserved.

*To whom correspondence should be addressed: Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs Blvd. MDC07, Tampa, Florida 33612, USA, Phone: 1-317-278-6448; Fax: 1-317-278-9217; vuvsky@health.usf.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Tumor suppressor protein p53 is a crucial transcription factor whose activity is modulated by a wide spectrum of stress signals that potentially affect genome integrity and proper cell proliferation. When activated, p53 coordinates a complex cellular response regulating expression of genes involved in various cellular processes, such as cell cycle progression, apoptosis induction, DNA repair, senescence, response to cellular stress, etc. [1–3]. A p53 deficiency is known to induce developmental abnormalities in animals [4]. Furthermore, when p53 function is lost, either directly through mutation or indirectly through several other mechanisms, the cell often undergoes cancerous transformation [5]. In fact, cancers showing mutations in p53 are found in colon, lung, esophagus, breast, liver, brain, and in hemopoietic and reticuloendothelial tissues [5]. All this makes p53 an extremely important player both in normal organism development and in oncogenesis. Based on its broad range of biological roles, p53 has been recognized as “the guardian of the genome” for its ability to stabilize the genome [6].

Human p53 is a 393 residue-long protein, which can be divided into the N-terminal region, the central DNA Binding Domain (DBD), and the C-terminal region (Figure 1A) [2]. The N-terminal region can be further subdivided into TransActivation Domain 1 (TAD1) (residues 1–40, pink bar), TAD2 (residues 40–60, red bar), and a proline-rich region, PR (residues 64–92, dark pink bar). The C-terminal region contains a tetramerization or oligomerization domain (OD; residues 325–356, cyan bar), and a regulatory C-terminal domain (CTD; residues 356–393, green bar) [2, 7]. Both the N-terminal and C-terminal regions are known to be involved in a multitude of interactions with a wide spectrum of partners. The transactivation region of p53 interacts with TFIID, TFIIF, Mdm2, RPA, CBP/p300 and CSN5/Jab1 [1]. The CTD of p53 interacts with GSK3 β , PARP-1, TAF1, TRRAP, hGcn5, TAF, 14-3-3, S100B($\beta\beta$) and many other proteins [1]. The central DBD domain (residues 94–292, blue bar) of p53 is highly conserved among different species, whereas both N- and C-termini are more susceptible to mutations. The conservation of the p53 DBD is very important for its function. It was estimated that 90% of cancer-related p53 gene mutations are missense mutations in the DBD, resulting in the loss of DNA binding and hence affecting p53 function in cell cycle control [8]. As indicated by the analysis of its crystal and NMR solution structures [9–11], the p53 DBD is an immunoglobulin-like β -sandwich that facilitates DNA binding (Figures 2A and 2B).

In their unbound states, the N- and C-termini of p53 are very flexible and resist crystallization. Proteins and regions without stable 3-dimensional structures under physiological conditions are known as Intrinsically Disordered Proteins (IDPs) and Intrinsically Disordered Protein Regions (IDPRs), respectively [12–14]. The phenomenon of protein intrinsic disorder is highly abundant in nature, and many proteins in various organisms have been predicted to be intrinsically disordered [13, 15–26]. As estimated by computational studies, in eukaryotic proteomes, 45–50% of proteins have long disordered regions with at least 30 consecutive residues, whereas archaea and bacteria have 7–30% such proteins [13, 18, 23, 26, 27]. Even in the Protein Data Bank, which is highly biased toward structured proteins, 70% of the deposited structured proteins have disordered regions of different length [28]. Disordered proteins usually play crucial roles in molecular recognition, regulation, and signaling [12, 15, 29]. Furthermore, such proteins are found to be strongly related to various human diseases, such as cancer, cardiovascular diseases, amyloidoses, neurodegenerative disease, diabetes, etc. [30].

From an evolutionary standpoint, some intrinsically disordered proteins and disordered regions can be highly conserved. For example, human α -synuclein (a canonical neurodegeneration-related intrinsically disordered protein comprised of 140 residues [31,

32]) differs from its mouse counterpart by mere six residues (4%), and there are just 21 residue differences (12%, which include residue differences at 18 positions and 3 insertions/deletions) between the human and canary α -synucleins [33]. Both numbers are similar to those expected for the highly conserved ordered proteins in these species. In flagellin, the ordered central region has greater sequence diversity than its disordered termini [34]. A comprehensive bioinformatics analysis of the InterPro database, a resource integrating eight different protein family and domain databases, revealed that functionally important conserved regions of predicted disorder are rather common and can be found in proteins from all kingdoms of life, including viruses [35, 36]. Furthermore, many functional, evolutionary, and structural units of proteins; i.e., functional domains of a significant size, were shown to be intrinsically disordered, and were grouped into a distinct class of biologically functional protein domains [37]. One should remember, however, that the existence of distinguishable families does not necessarily indicate sequence conservation. Furthermore, sequence conservation is not equivalent to disorder conservation, and regions with conserved disorder are not equivalent to the disordered regions with conserved sequences.

On the other hand, many intrinsically disordered sequences are known to be characterized by high evolutionary rates [38–40]. For example, the analysis of calcineurins [41], topoisomerase [42], ribosomal protein S4 [43], β -subunits of the potassium channel Kv β 1.1 [44], and many other proteins revealed that their disordered regions contain more amino acid substitutions, insertions and deletions than the ordered regions of the same protein [38, 45]. Based on the observation that a significantly higher degree of positive Darwinian selection was observed in intrinsically disordered regions of proteins compared to regions of α -helix, β -sheet or tertiary structures, it was hypothesized that intrinsically disordered regions may be required for the genetic variation with adaptive potential and that these regions may be of “central significance for the evolvability of the organism or cell in which they occur” [46].

Finally, disordered and ordered domains of the same protein can possess similar degrees of conservation and co-evolution, as was shown for the papillomavirus E7 oncoprotein [47]. Overall, a systematic study of several families of proteins having at least one structurally characterized disordered region revealed that the disordered regions are characterized by highly heterogeneous evolutionary rates, with some disordered amino acid sequences evolving slowly, and others evolving more rapidly than ordered sequences [38].

Based on empirical studies of protein flexibility [48] and more recently on large-scale analysis of predicted and experimentally-determined intrinsically disordered regions it has been concluded that peculiarities of the amino acid composition, and not the amino acid sequence might be conserved in some disordered proteins. In fact, a significant number of disordered regions with low sequence conservation between the human proteome and seven other eukaryotes were shown to possess considerable similarity in their chemical composition between orthologs [49]. Furthermore, the amino acid compositions of disordered regions located in different parts of protein (i.e., found in N-terminal, central, or C-terminal regions, where terminal locations correspond to the 40 residues at either end of the protein) were characterized by specific and reproducible biases suggesting that for IDPRs there is location preference within the proteins that is dependent on their chemical composition [49]. Since clustering disordered regions based on their chemical composition resulted in functionally distinct groups, it was concluded that preserving the overall chemical composition can represent one of the ways for disordered regions to keep their flexibility and function through evolution [49].

The p53 molecule is a member of an important protein family, which in addition to p53 includes p63 (see [50, 51]) and p73 [52]. Both p63 and p73 are structurally similar and

functionally related to p53 (Figure 2). The entire p53-family constitutes a unique family-based signaling network that controls such aspects of cell life as proliferation, differentiation, and death [3]. Both human p63 and p73 are almost two-fold longer than p53 and have 680 and 636 residues respectively. The domain organization of the members of p53-family is rather similar (Figure 1). In fact, all three proteins have identifiable TAD, DBD and OD. In p63 and p73, there is an additional C-terminal sterile- α motif (SAM), which is required for p63 and p73 transcriptional activity, but they seem to lack the CTD found in p53 [3]. The various p53 family members have limited overall homology, but strong similarity in the DBD (approximately 60% between p53 and p63/p73 and approximately 85% between p63 and p73) [53, 54]. The DBD and OD of human p53, p63 and p73 possess high structural similarity (Figures 2B and 2C). The SAM domains of p63 and p73 are structurally similar as well (Figure 2D). All the members of the p53 family are expressed in a similar set of alternatively spliced isoforms and their mRNAs are controlled by separate promoters, encoding proteins with alternative N-terminal regions. Furthermore, all three p53-family proteins have similar post-translational modification patterns [3].

Despite their structural similarity, the p53-family members are functionally very diverse. Important differences in their biological roles have been discovered via the analysis of the corresponding knock-out mouse models. These studies revealed that although some mild developmental and fertility defects can be detected in the p53-null mice [4, 55, 56]. These animals were viable and largely normal, but died of cancer at young age, reemphasizing the crucial role of p53 in preventing the spontaneous oncogenesis [57]. The outcomes of the p63 and p73 knock-outs were very different. In fact p73-null mice were born viable but they displayed hydrocephalus, had nervous system abnormalities and immunological problems with chronic inflammation, showed reproductive and behavioral defects, and generally died within the first 2 months. These mice, however, typically did not develop spontaneous tumors [58]. Finally, p63-null mice clearly indicated that this member of the p53-family is involved in maintaining the pool of the proliferating stem cells during development of epithelia [59, 60]. In fact, although these mice were born alive, they died immediately after birth, showing a severe phenotype that included lack of limbs and a wide range of epithelial structures such as skin, prostate, breast, and urothelia [61, 62]. Furthermore, it has been pointed out that the majority of SAM-containing-proteins are involved in regulation of development, supporting the hypothesis that p63 and p73 play a role in differentiation [63–66]. All this clearly suggests that p53, p63 and p73 have gained a high degree of functional specificity since their duplication and divergence during the evolution [53, 67].

Previous evolutionary analysis revealed that p53 family members were widespread among animals [67]. Importantly, these proteins were found in many taxa beyond vertebrates, such as ascidians (sea squirts) [68], cnidarians [69, 70], platyhelminths and others [67]. Furthermore, two distinct genes encoding a p53-like and a p63/p73-like sequences were recently found in the choanoflagellate (*Monosiga brevicollis*) [71], and a p53-like sequence was reported in another protozoan, the amoeba *Entamoeba histolytica* [72]. These findings suggested that the emergence of the p53-family clearly predated the appearance of multicellular animals. Furthermore, it is likely that both p53-like and p63/p73-like genes existed in the common ancestor of the metazoa and protozoa. Since p63/p73-like genes are noticeably absent from several lineages in which p53-like genes are present (such as flies, nematodes and cnidarians) and since the SAM domain was absent from p53, it has been hypothesized that p63 and p73 shared a more recent common ancestor [67]. It was emphasized that the evolutionary persistence of the p53 family, where the amino acid sequences and structures of the DNA binding domains of this family of proteins and even the DNA sequences to which these protein domains bind have been conserved for over a billion years of evolution [73, 74], clearly indicates that this family of genes played important biological roles throughout evolutionary history [75]. Furthermore, a detailed

comparison of the amino acid sequences of three members of the p53 gene family, p53, p63 and p73, in vertebrates revealed that they are evolving at highly distinctive rates, and this variation occurs not only between the three family members but extends to distinct domains in each protein [76].

Even though p53 is one of the most heavily studied of all proteins, the p53 molecule still has some unresolved mysteries. In fact, although the multiple relations between intrinsic disorder and functions of p53 have been broadly explored, the relationship between intrinsic disorder and sequence conservation in this protein is not clear. To resolve this issue, multiple p53 sequences from various species were analyzed to better understand the evolution of this important protein and to evaluate the role of intrinsic disorder in the process of p53 evolution. When viewed in the light of the constraints on evolutionary change provided by protein structure [40, 45, 77–83], the results found herein support the *in vivo* existence of disorder for regions of p53 critical to its function.

Materials and Methods

Datasets

Proteins with sequence similarity to the p53 DBD from humans were identified from UniProt [84] using BLAST [85]. Bacterial sequences were removed from this set because they only had similarity to the DBD and none of the other defining regions of the p53 family. After removing redundant sequences by checking the species name and gene ID, 84 sequences from the original dataset were kept (Table 1). In the cases of multiple sequences with the same species name and gene ID, the longest sequences were selected. Among these 84 sequences, 45 sequences were p53, 29 sequences were p63/73 or predicted p53, and 10 sequences were undefined proteins but with high sequence identity to the DBD domain of p53.

Disorder prediction

The abundance of predicted intrinsic disorder in the members of the p53-family was evaluated by PONDR-FIT due to its accuracy on various datasets [86]. PONDR-FIT is a neural network-based meta-predictor which combines the prediction results of PONDR[®] VL-XT [87, 88], PONDR[®] VSL2 [89, 90], PONDR[®] VL3 [28, 91, 92], FoldInex [93], IUPred [94], and TopIDP [95]. Although it takes the same strategy as CDF-all [96], PONDR-FIT provides disordered prediction on per residue level. All the component predictors of PONDR-FIT have their own specific features: PONDR[®] VL-XT is important for identifying Molecular Recognition Features (MoRFs), which are the structure-prone segments located within a long disordered region and which have been frequently observed in protein-protein interactions [27, 97]. PONDR[®] VSL2 is one of the best disorder predictors with high prediction accuracy of short disordered regions. PONDR[®] VL3 is specially designed for accurate prediction of long disordered regions. IUPred employs pairwise interaction energies obtained from globular proteins and hence is sensitive to the changes in structured part of proteins. FoldIndex is essentially a transformation of Charge-Hydrophathy (CH) plot [14] but giving residue-based prediction. TopIDP is an artificial index which was developed by means of a genetic algorithm to outperform other single indexes on the accuracy of disorder prediction. The advantage in using PONDR-FIT as a tool for the disorder evaluation is in the fact that this algorithm is a meta-predictor that integrates outputs of six individual disorder predictors that are based on rather different logistics and use very different attributes and features. By combining outputs of divergent individual predictors, PONDR-FIT achieves better prediction accuracy. Furthermore, since meta-predictor analysis is based on the integration of the outputs of individual predictors, the

resultant PONDR-FIT scores are generally consistent with the results generated by the individual disorder predictors.

Prediction of α -helical molecular recognition features (α -MoRFs)

Often, intrinsically disordered regions in proteins are involved in protein-protein interactions and molecular recognitions [13, 29, 98–101]. It has been pointed out that many flexible proteins or regions undergo disorder-to-order transitions upon binding, which is crucial for recognition, regulation, and signaling [12–14, 27, 102–105]. A correlation has been established between the specific pattern in the PONDR® VL-XT curve and the ability of a given short disordered regions to undergo disorder-to-order transitions on binding [106]. Based on these specific features seeing in disorder profiles and a set of other attributes an α -MoRF predictor was developed [27]. This predictor focuses on short binding regions within long regions of disorder that are likely to form helical structure upon binding [27]. It uses a stacked architecture, where PONDR® VL-XT is first used to identify short predictions of order within long predictions of disorder and then a second level predictor determines whether the order prediction is likely to be a binding site based on attributes of both the predicted ordered region and the predicted surrounding disordered region. An α -MoRF prediction indicates the presence of a relatively short (20 residues), loosely structured helical region within a largely disordered sequence [27]. Such regions gain functionality upon a disorder-to-order transition induced by binding to partner [104, 105]. Later, a second generation α -MoRF predictor, α -MoRF-II, was developed [97]. The prediction algorithm was improved by including additional α -MoRF examples and their cross species homologues in the positive training set; careful extracting monomer structure chains from PDB as the negative training set; including attributes from recently developed disorder predictors, secondary structure predictions, and amino acid indices as attributes; and constructing neural network based predictors and performing validation [97]. The sensitivity, specificity and accuracy of the resulting predictor, α -MoRF-PredII, were 0.87 ± 0.10 , 0.87 ± 0.11 , and 0.87 ± 0.08 over 10-cross validation, respectively [97]. In this study, α -MoRF regions were predicted by the α -MoRF-II predictor [97].

Sequence divergence

The sequence divergence at each site was evaluated by K2-entropy as follows. First, the sequences were aligned by ClustalW [107]. After the alignment, the frequency of each type of the amino acid was counted for each sequence position as P_i , with gaps being counted as the 21st type of amino acid. The value of the K2-entropy for each position was obtained using the equation $E = -\sum P_i \log_2 P_i$. Finally, the values of E were smoothed over three consecutive residues by taking moving averages on these three residues.

The calculation of entropy and substitution probability per site can be affected by the accuracy of the multiple sequence alignments. After using the default parameters of ClustalW, we checked the alignment manually and adjusted alignments with obvious problems. Although the entropy values for a fraction of the positions are slightly different in various adjusted alignments, the overall trend between entropy and disorder score was not noticeably affected by these adjustments. Therefore, the statistical averages are less affected by the alignments than the per position values. However, in terms of substitution probability, the positions in non-DBD regions may be moderately affected by the alignment accuracy.

The treatment of gaps in multiple sequence alignment may have an unpredictable influence on the calculated entropy because gaps can change the relative frequency of each amino acid at a site with many gapped sequences. In fact, the DBD of the p53-family members are highly conserved in both sequence and structure (Figure 2). After sequence alignment, most sites in the DBDs do not have a high percentage of gaps. However, in less conserved

regions, the gap content can be extremely high. A statistic for the distribution of gaps using human p53 as a template is shown in Figure 3. If human p53 showed a gap after alignment, this gap-containing site was defined as a gap-dominant site. Otherwise, it was defined as a residue-dominant site. By counting the percentage of sequences that have gaps at each site in the aligned sequences, the influence of gaps on alignment can be evaluated. Figure 3A represents the percentage of gaps for gap-dominant and residue-dominant sites. Clearly, almost all residue-dominant sites of human p53 have small percentages of gaps (less than 40%), whereas almost all gap-dominant sites have 60% or more gaps. Although this information was extracted by taking human p53 as a template, the results were not affected when other proteins were used as templates for the alignment. Therefore, these data are informative for subsequent analysis. Figure 3B is a histogram of the distribution of gap percentages. Apparently, the majority of the sites are either residue-dominant, or gap-dominant. This distribution can be easily decomposed into a combination of two single-peak distributions. The saddle point is located at the value of 40 – 50%. For the purpose of focusing on sites with low number of gaps, the value of 40% gaps per site was chosen as a threshold to analyze entropy and all sites with more than 40% gaps were eliminated from the analysis.

Phylogenetic inference

Mega4.0 was used to infer the phylogenetic relationships among the p53 proteins using only the DBD [108]. The maximum likelihood method was used to infer a tree from the manually corrected ClustalW alignment of this region and an automatically inferred starting tree. The Jones-Taylor-Thornton amino acid substitution matrix with uniform rates was used as the evolutionary model. Positions with gaps were excluded from the analysis.

Substitution frequency

To visualize the frequency of different amino acid substitutions based upon the aligned p53-family proteins, the WebLogo tool was used (<http://weblogo.berkeley.edu/>) [109]. WebLogo is a web-based application to generate sequence logos, a graphical representation of a multiple sequence alignment. [110]. Here, each logo represents stacks of symbols, one stack for each position in the sequence or for each type of the amino acid residue. The overall height of the stack indicates the sequence conservation at that position (or for a given type of residue), while the height of symbols within the stack indicates the relative frequency of each amino acid at that position (or substituted for that residue type). In general, a sequence logo provides a better description of functionally important sites than would a consensus sequence [109, 110]. This tool was used to generate logos describing the frequencies of the substitutions of each of the residue type irrespective of their sequence positions. This tool was also used to analyze sequence conservation within the different p53 MoRFs.

Results

Disorder in p53, p63, and p73 proteins

Figure 4 compares the distribution of predicted disorder in human p53, p63, and p73 sequences. The sequences were aligned by overlapping their DBDs since the DBDs are known to be highly conserved. Clearly, the disorder score distributions for these three proteins are highly similar. In fact, the disorder profiles for the DBDs of p53, p63, and p73 (residues 160–360 in the p53 numbering) are almost identical. In addition, two short structure-prone motifs centered at residues 90 and 400 flank the DBDs in these proteins. The positions of predicted α -MoRF regions N, C1, and C2 within the p53 sequence coincide with the TAD1, OD, and CTD discussed above. Being disordered when free in solution, these α -MoRF regions of p53 fold into crystallizable structures after interaction with their

specific binding partners, and therefore they were observed in crystal structures of corresponding complexes [111–113].

Additionally, being 300-residues longer, p63 and p73 have some additional structural features not found in p53. For example, there is a SAM domain located in the C-terminal part of p63 and p73 [114]. Furthermore, both N- and C-termini are noticeably longer in p63 and p73 and their disorder score plots contain several additional dips that correspond to the structure-prone disordered motifs. Actually, motifs N and C of p63 and p73 were predicted to be α -MoRFs. Figure 4 shows also that the disorder profile of the p63 N-terminal region contains three pronounced dips, centered at positions 20, 60 (marked in disorder profile as B1 and B2 fragments respectively), and 90. Only the last of these structure-prone motifs (i.e., motif centered at residue 90) was predicted as an α -MoRF, whereas two other motifs (centered at residues 20 and 60) were not identified as α -MoRFs due to their length. However, the overall appearance of these dips is quite similar to that of typical MoRFs. Similarly, the N-terminal region of the p73 disorder profile contains two dips that correspond to the structure-prone motif G1 and the predicted N MoRF.

Overall, Figure 4 clearly shows that p53, p63, and p73 have a number of similar features. Therefore it was interesting to analyze how these three sequences deviated from each other, assuming that they were homologues. Since the DBDs of the p53-family members were highly conserved (their identity is ~60% [53, 54]), only the predicted MoRFs were taken as check points to compare detailed differences between these three sequences. The sequence alignments for sets of MoRFs from p53, p63, and p73 are shown in Figure 5. Obviously, the N MoRFs of p53, p63, and p73 are rather divergent as it follows from the poor alignment of the corresponding fragments (Figure 5A). However, the N MoRF of p53 and the B2 fragment of p63 (Figure 5B) and the B2 fragment of p63 and the G1 fragment of p73 (Figure 5C) have highly conserved amino acid sequences. Additionally, the C1 and C2 MoRFs of p53 are similar to the M2 and M3 MoRFs of p63 (Figures 5D and 5E, respectively), the M2 fragment of p73 is similar to the C1 MoRFs of p53 (Figure 5F), and the M1 fragments of p63 and p73 also show high similarity (Figure 5G). Hence, in addition to the DBD, all the predicted MoRFs of p53 are also conserved. Assuming that p53 is closer to the ancient prototype, a possible mechanism for the divergence during the p53, p63 and p73 evolution includes the preferential introduction of insertions and substitutions within the disordered regions connecting the ordered domains and order-prone fragments.

Phylogenetic distribution of disorder

Although alignments of the functional regions of human p53, p63, and p73 provided interesting clues about the evolutionary diversification of these proteins, more information can be gained by putting the members of this family into a phylogenetic context. Therefore, we aligned just the DBDs of the proteins in our dataset based upon the region from residue 101 to residue 300 of the human p53 protein. These data then were used to infer the phylogenetic tree shown in Figure 6A. Since the DBD is structured [9–11], it can be used as a starting point for the subsequent analysis of other regions of the p53-family members. Furthermore, by choosing the DBD as this starting point, the evolution of structured and intrinsically disordered regions within the same polypeptide chain can be analyzed. Therefore, the phylogenetic tree generated in our study (Figure 6A) is not based on the alignment of the entire protein sequences, but on the alignment of the DBDs only. Although using only a part of the sequence to infer a phylogeny is not a very popular approach, it is beneficial for the specific purposes of identifying the correlation between the intrinsic disorder and the sequence divergence. Hence, the relative positions of various sequences may not be used as evidence of their evolutionary relationships.

In the phylogenetic tree (Figure 6A), cluster A has early metazoan p53 family sequences, cluster B has mosquito sequences, cluster C includes insects like the fly, and cluster D has aquatic invertebrates. The fourth cluster is not taxonomically cohesive, having lancets, sea anemones, and choanoflagellates. Cluster E has p63/73 proteins in vertebrates. The last cluster G has p53 proteins of vertebrates. Figure 6B shows the percentages of disordered residues in the p53-family of proteins from each species evaluated by the PONDR-FIT algorithm. In this figure, the order of the species is the same as in the phylogenetic tree shown in Figure 6A, with the seven letters indicating seven clusters in this phylogenetic tree. Figure 6B shows that the average amount of disorder in different clusters is different, with ancient clusters (e.g., cluster A) typically showing noticeably less disorder than more recent clusters (e.g., cluster G).

Entropy vs. Disorder

Due to the complexity of visualizing a multiple sequence alignment, entropy is a good indicator of the diversity of amino acids at each site. Figure 7A is a comparison between the K2 entropy and the disorder scores (calculated from the average of all p53 family sequences in the dataset by aligning all other sequences with human p53, P04637) at each position of human p53 evaluated by PONDR-FIT, whereas Figure 7B shows the direct comparison of the disorder score and K2 entropy. These data illustrate that there is agreement between evolutionary divergence and disorder, with disordered residues and disordered regions typically being less conserved than ordered ones (correlation coefficient = 0.676; $t = 18.2$; $df = 392$; $p\text{-value} < 0.001$). However, p53 has a few non-conserved residues which are predicted to be ordered, whereas several highly disordered residues are evolutionary conserved.

Figure 8A further illustrates the relation between the average disorder score for each site in the alignment vs. the K2 score at that site. Circles in Figure 8A represent the data for all the aligned sites in the non-DBD regions, whereas the data for the sites located within the DBD domain are shown by crosses. Although most of the sites with low disorder scores and low entropy in Figure 8A were confirmed to be alignment gaps, the correlation between disorder score and entropy was still obvious. Almost all of the DBD residues have small disorder scores, less than 0.5. The outliers are residues at the C-terminus of the DBD that were predicted to be disordered but were assigned to the DBD. In terms of entropy, despite its broad distribution, the actual values of entropy for most positions are less than 2.5. The overall trend in the Figure 8A is quite clear: the larger the disorder score, the higher the entropy.

An analysis excluding gap-dominant sites is also helpful since the sequences in our analysis have very different sequence lengths, from 340aa to 680aa. The results of this analysis for the sites inside and outside the DBD domain are shown in Figures 8B by crosses and circles, respectively. Interestingly, after removing the gap-dominant sites, the correlation between entropy and disorder score is clearer. The overall patterns in Figure 8B did not change too much in comparison with the patterns shown in Figure 8A, except for the removal of a group of positions characterized by a combination of low disorder scores and low entropy values, suggesting that positions with more than 40% of gaps might have low disorder score. The majority of positions with more than 40% gaps correspond to the gap-dominant sites; i.e., those sites which are gaps in human p53 (which is used as a template in this analysis). Therefore, since p53 is the shortest member of the p53 family, and since both p63 and p73 contain additional ordered SAM domain (see Figure 4) one would expect that the positions with high gap contents would be characterized by low disorder scores. Importantly, almost all positions located outside the DBD domain have larger disorder score (>0.5) and higher entropy values (>2.0) than sites within the DBD. Hence, data shown in Figures 7 and 8

suggest that all the p53 positions can be classified into two groups: either highly disordered and diversified, or mostly structured and conserved.

Substitution frequency analysis

For each amino acid, Figure 9 shows the frequencies of substitutions from Q868M8 to other amino acids in the aligned sequences. These analyses separately provide information for the DBD and other regions on which residue amino acid substitutions are preferred in by evolution. Figures 9A and 9B represent data for all the positions on the aligned sequences, whereas Figures 9C and 9D show results only for positions with 40% or fewer gaps. Because Q868M8 has the shortest distance from the common origin as shown in the phylogenetic tree, it was chosen as the most ancient sequence to test the substitutions of all other sequences. Q868M8 is a p53 protein identified in amoeba. It contains 430 residues, being therefore longer than human p53 but substantially shorter than human p63. Although Figure 9 shows that sites located inside the DBD domain possess similar substitution probabilities, Figures 9C and 9D provide clearer information on the trend of substitutions in regions outside the DBD. Therefore, by taking Figures 9A and 9B as a reference, we will mainly concentrate on the analysis of Figures 9C and 9D.

Figure 9C represents data for the DBD domain and clearly shows that the major structure-promoting residues W, C, F, Y, V, and L are highly conserved. It is seen that W is preferentially substituted by L, with V being the second most common substitution for this residue. C is highly conserved and if mutated, it is preferentially mutated to T, followed by G and S. Gaps in Q868M8 tend to be mutated to A and E.

Compared to Figure 9C, the trend in Figure 9D is noticeably fuzzier, showing that the residues located outside the DBD are less conserved than residues within this domain. Based on the results of other analyses reported above, the conservation in the non-DBD regions is expected to be relatively low. Based on the current sequence alignments, W is preferentially substituted by R, and C seems to be substituted preferentially by K and L, structure-prone residues may mutate to L, whereas all other residues have higher probability of mutation to S. It also appears that the options for amino acids to substitute to other amino acids are less restricted outside the DBD, as indicated by their greater frequency and more even distribution.

We also analyzed the conservation of sequences comprising N- and C-terminal α -MoRFs. Figure 10A represents the position-specific data for the 52 aligned N-terminal MoRFs, whereas data for the 69 C-terminal MoRFs are shown in Figure 10B. This figure clearly illustrates that the sequences of both MoRF regions are essentially more conserved than the sequences of other non-DBD regions. Figure 10 also shows that conservation rules are very different for these two binding regions. In fact, Figure 10A illustrates that the N-terminal α -MoRFs are in general more hydrophobic than the C-terminal α -MoRFs. There is an obvious preference for the structure-promoting F and W residues in the middle of the N-terminal MoRFs (positions 18 and 22), whereas the N- and C-termini of these MoRFs clearly prefer polar and charged residues. On the contrary, N- and C-termini of the C-terminal MoRFs contain noticeable amounts of hydrophobic residues, whereas their central regions tend to have more polar and charged residues.

Finally, we analyzed how the peculiarities of protein-protein interactions are conserved in three different MoRF regions of p53: the N-terminal and two C-terminal MoRFs, N, C1 and C2 (Figure 4). This was assessed by analyzing the correlation between the K2-entropy and changes in the accessible surface area of the MoRFs upon their interaction with the various binding partners. For this purpose, we analyzed 7, 14, and 9 complexes involving the N, C1 and C2 MoRFs, respectively. To determine interface areas between two chains, the solvent

accessible surface area (ASA) of each individual chain was calculated as well as the ASA of the complex. ASA can be calculated analytically [115] or numerically [116] for residues and entire chains from the three dimensional structure of a protein. The amount of the ASA that becomes inaccessible upon complex formation represents the interface area that can be calculated as the change in ASA (Δ ASA); i.e., the sum of the individual chain ASA minus the complex ASA. Δ ASA was estimated by existing methods [117, 118]. Residues directly involved in interactions were identified from molecular structures as residues with a Δ ASA greater than $1 \text{ \AA}^2$ [117, 118]. All calculations here used a probe radius of $1.4 \text{ \AA}$, which roughly corresponds to the size of a water molecule.

Figure 11A represents the 3D-plot illustrating how the ASA of various MoRF residues is affected by binding and how these changes are correlated with the corresponding K2-entropy values. The effects of binding on the ASA of different MoRF residues vary in a wide range, with some residues being unaffected by binding and other residues showing dramatic binding-induced changes in their ASA. Obviously, unaffected residues are likely located on the surface of the complex, whereas residues with large Δ ASA are likely to be involved in the formation of the interface. Figure 11A shows that the burial degree of all the residues in all the MoRFs analyzed in this study is poorly correlated with their conservation. This conclusion follows from a very complex configuration of the surface representing the K2-entropy vs. Δ ASA vs. ASA dependence that is not smoothed and contains a number of local maxima and minima. Of some importance is the fact that residues with large ASA and small Δ ASA are typically characterized by higher entropy (see upper right corner of the plot). On the other hand, MoRF residues with the largest Δ ASA are not characterized by the lowest K2-entropy as one would expect for a set of highly conserved surface residues which become buried at the MoRF binding to a partner. This conclusion is further supported by Figure 11B that represents the K2-entropy vs. Δ ASA for the residues of different MoRFs. Here, each point corresponds to a MoRF residue and is colored according to the residue location in the N-terminal, C1 and C2 MoRFs.

Although there is no correlation between the degree of burial and the conservation degree for the residues within the C2 MoRFs, some weak positive correlation is observed for the C1 MoRFs, whereas burial and conservation of the residues in the N-terminal MoRFs are weakly anti-correlated. The latter observation suggests that the interface residues in the N-terminal MoRFs do have a weak tendency to be conserved.

Discussion

By studying the sequence divergence of the p53-family proteins, structured domains of p53 were observed to be highly conserved while disordered regions were found to be highly divergent. This observation is in agreement with previous findings for disordered proteins [45]. Additionally, disordered regions had a tendency to substitute more diverse amino acids than ordered regions as shown for a broader array of disordered proteins [83]. Upon mutation, structured domains may lose unique structure, hence leading to the loss of specific functions. Therefore, for domains requiring stable structure for function, the number of allowable substitutions will be small. However, such structure-function evolutionary restrictions are not applicable to disordered regions, which might be functional but do not possess unique structure, being very flexible and existing as a highly dynamic ensemble of conformations. It can be expected that some of the conformations may have very important functional roles, while others do not. Hence, as long as the mutations in disordered regions do not affect the sampling of those functionally important conformations, these disordered regions may preserve functionality during their evolution. Furthermore, since disordered regions often function as dynamic linkers between more structured regions, the mutational limitations on the disordered regions are even less strict.

Our analysis showed that all members of the p53 family of proteins are expected to have substantial amounts of disorder. The phylogenetic analysis of the p53 family members from different organisms revealed that the overall amount of intrinsic disorder decreases over time, with more ancient sequences typically showing noticeably less disorder than more recent sequences (Figure 6B). Intriguingly, there is remarkable similarity between the disorder profiles calculated for human p53, p63, and p73 even though these proteins arose prior to the vertebrate radiation [76]. Although the DBDs of these proteins are characterized by the similar shapes of their disorder profiles (Figure 4), all the predicted MoRFs of p53 are also conserved. Of special importance is the observation that the most conserved regions correspond to folded (DBD) or foldable-upon-binding regions (predicted and experimentally verified α -MoRFs). In other words, the divergence during the p53, p63 and p73 evolution was induced by the preferential introduction of insertions and substitutions within the disordered regions connecting the more conserved ordered domains and order-prone functional fragments. This conclusion was further verified by the results of the comparison between the K2 entropy (calculated from the average of all p53 family sequences in the dataset by aligning all other sequences with human p53, P04637) and the PONDR-FIT-based disorder scores evaluated for each position of human p53. This analysis showed that all the positions within the p53 sequence can be classified into two groups, highly disordered and diversified, and mostly structured and conserved. Also, the logo plot analysis clearly illustrated that in addition to the highly conserved DBDs, the sequences of MoRF regions were essentially more conserved than the sequences of other non-DBD regions. Summarizing, it is clear from our analysis that the sequence conservation in the p53 family is characterized by a mosaic architecture with alternating highly conserved and non-conserved regions.

Finally, we used the fact that the ASA of different MoRF residues are differently affected by interactions with different binding partners to evaluate the correlation between the degree of burial and sequence conservation for residues in all p53 MoRFs with known structures. The corresponding analysis of the correlation between the K2-entropy value and the interface area calculated for the residues of different p53 MoRFs co-crystallized with various binding partners as the amount of the ASA that becomes inaccessible upon complex formation (Δ ASA) revealed that the peculiarities of protein-protein interactions are differentially conserved in the three p53 MoRFs. Here, no correlation was found between the degree of burial and the degree of conservation for the residues within the C2 MoRFs, and a weak positive correlation was detected for the C1 MoRFs. There was a weak anti-correlation for residues in the N-terminal MoRFs, suggesting that the interface residues in the N-terminal MoRFs do have a weak tendency to be conserved. This conservation of interface residues in the N-terminal MoRF of p53 indicates the similarity in evolution of folded and foldable-upon-binding regions of this protein, where functionally important residues are more conserved than other residues.

The lack of a similar correlation for the C2 MoRF can be understood based on the observation that this p53 region possesses a unique binding plasticity, being able not only to interact with various binding partners, but also to gain different structures in its bound form [119]. Previously, we performed a detailed analysis of structures of this p53 region bound to four different partners, cyclin A [120], sirtuin [113], CBP [121], and S100 β [122]. We showed that the same MoRF displayed all three major secondary structure types and became an α -helix when bound to S100 β , a β -strand when bound to sirtuin, and an irregular structure with two distinct backbone trajectories when bound to CBP and cyclin A2 [119]. Since the secondary structures of this p53 region were very different in different bound states, it was suggested that p53 is able to utilize different residues within the same region for the interactions with these four partners. In agreement with this hypothesis, the analysis of the buried surface area for each residue in each interaction produced different amino acid

interaction profiles, showing that the same residues are used to different extents in the four interfaces [119]. Therefore, this multitude of bound forms defines the variability of residues used by the C2 MoRF for interaction with different partners and explains the lack of conservation of residues involved in interaction.

While extensive experimental evidence indicates that IDPs/IDPRs clearly exist *in vitro*, there is uncertainty whether they remain disordered *in vivo* even though in-cell NMR data demonstrates the existence and stability of IDPs/IDPRs inside of cells for both prokaryotes and eukaryotes [123–133]. In fact, the skepticism on the existence of IDPs *in vivo* is supported by the reasoning that IDPs/IDPRs cannot exist in cells except transiently since cells have elaborate mechanisms to deal with misfolded proteins (so called unfolded protein response, UPR) and these mechanisms would clear disordered proteins [134]. However, one should remember that UPR seems to be specific for eukaryotes (where it is confined to the endoplasmic reticulum (ER) [135]), and that the prokaryotic cells do not have similar mechanisms. Furthermore, we have recently argued that the reasoning that UPR would clear IDPs is unsupported on several grounds including the differences in the properties of unfolded ordered proteins (which contain high proportions of hydrophobic and aromatic residues) and IDPs (which are highly polar sequences that specifically evolved to be unfolded under physiological conditions) [136]. For example, in yeast, the UPR is triggered by the dimerization of the transmembrane receptor kinase/ endoribonuclease Ire1p [137, 138], subsequent transautophosphorylation of which leads to the activation of its endoribonuclease domain that processes the mRNA encoding the transcription factor Hac1p [139–142] which eventually activates the transcription of a wide array of genes (~5% of the genome), including chaperones, oxidoreductases, phospholipid biosynthetic enzymes, ER-associated protein degradation components, and many other proteins involved in a variety of different cellular processes, including UPR [143–145]. The triggering event of the Ire1p dimerization is, in turn, controlled by the ER-resident Hsp70 chaperone BiP (called Kar2p in *Saccharomyces cerevisiae*) that among several other functions is responsible for sensing the level of unfolded proteins in the ER [144, 146–149]. In Hsp70-type chaperones, such as BiP, a signature binding motif of the unfolded substrate consists of hydrophobic amino acids in every other position. In other words, the substrate motif recognized by BiP resembles a β -strand, one side of which is destined to pack onto the hydrophobic core of a folded protein but has not yet been properly accommodated in the protein fold. Therefore, the sequence properties of the unfolded protein provide means of recognizing unfolded proteins that need chaperone assistance [150]. In addition to BiP, the conserved core region of the Ire1p's ER-luminal domain was recently shown to recognize unfolded proteins too [145]. This Ire1p groove responsible for the recognition of unfolded proteins is lined by a patchwork of conserved hydrophobic and hydrophilic residues [145] and therefore resembles grooves found in class I major histocompatibility complexes (MHCs), which bind peptides with high sequence specificity [145, 151].

The high mutation rates for IDPs/IDPRs observed in this and other studies provides further support for the idea that intrinsic disorder exists as such *in vivo*. If intrinsically disordered protein was generally structured *in vivo*, then we would expect IDPRs to have similar rates and patterns of evolution as other structured regions of the protein. For structured proteins, it has been shown repeatedly that buried residues have fewer substitutions over time while surface residues have higher substitution rates [40, 77, 78, 81]. Lin *et al.* also showed that the presence of IDPRs within a protein increased the overall evolutionary rate [79]. On the other hand, ordered loops with no-regular secondary structure (NORS) are surface residues that do not show higher evolutionary rates than ordered regions of the same protein, indicating that IDPRs are not evolving faster simply because they are on the surface and have no secondary structure to constrain their evolution [79]. Furthermore, although surface residues and IDP residues may have similar compositions [80], their patterns of substitutions

are very different [82, 83]. Additionally, IDPs/IDPRs have insertion/deletion rates that are typically higher than those of ordered proteins [45]. All of these observations provide further support for our assertion that disordered regions evolve differently from ordered surface residues and therefore that IDPs and IDPRs are disordered *in vivo*.

Acknowledgments

This work was supported in part by the grant from the “Molecular and Cellular Biology” Program of the Russian Academy of Sciences (to V.N.U.). C.J.B. was supported by NIH grants 5P20RR16448, 8P20GM103397 to the University of Idaho.

Abbreviations

ASA	solvent accessible surface area
CDF	cumulative distribution function
CH	charge-hydrophathy
CTD	C-terminal domain
DBD	DNA binding domain
ER	endoplasmic reticulum
IDP	intrinsically disordered protein
IDPR	intrinsically disordered protein region
MHC	major histocompatibility complex
MoRFs	molecular recognition features
NORS	no-regular secondary structure
OD	oligomerization domain
PR	proline-rich region
SAM	sterile- α motif
TAD	transactivation domain
UPR	unfolded protein response

References

1. Anderson, CW.; Appella, E. Signaling to the p53 tumor suppressor through pathways activated by genotoxic and nongenotoxic stress. In: Bradshaw, RA.; Dennis, EA., editors. Handbook of Cell Signaling. New York: Academic Press; 2004. p. 237-247.
2. Joerger AC, Fersht AR. Structural biology of the tumor suppressor p53. Annu Rev Biochem. 2008; 77:557–582. [PubMed: 18410249]
3. Collavin L, Lunardi A, Del Sal G. p53-family proteins and their regulators: hubs and spokes in tumor suppression. Cell Death Differ. 2010; 17:901–911. [PubMed: 20379196]
4. Armstrong JF, Kaufman MH, Harrison DJ, Clarke AR. High-frequency developmental abnormalities in p53-deficient mice. Curr Biol. 1995; 5:931–936. [PubMed: 7583151]
5. Hollstein M, Sidransky D, Vogelstein B, Harris CC. p53 mutations in human cancers. Science. 1991; 253:49–53. [PubMed: 1905840]
6. Lane DP. Cancer. p53, guardian of the genome. Nature. 1992; 358:15–16. [PubMed: 1614522]
7. Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, Vucetic S, Iakoucheva LM, Obradovic Z, Dunker AK. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. BMC Genomics. 2009; 10(Suppl 1):S7. [PubMed: 19594884]

8. Soussi T, Beroud C. Assessing TP53 status in human tumours to evaluate clinical outcome. *Nat Rev Cancer*. 2001; 1:233–240. [PubMed: 11902578]
9. Joerger AC, Ang HC, Veprintsev DB, Blair CM, Fersht AR. Structures of p53 cancer mutants and mechanism of rescue by second-site suppressor mutations. *J Biol Chem*. 2005; 280:16030–16037. [PubMed: 15703170]
10. Canadillas JM, Tidow H, Freund SM, Rutherford TJ, Ang HC, Fersht AR. Solution structure of p53 core domain: structural basis for its instability. *Proc Natl Acad Sci U S A*. 2006; 103:2109–2114. [PubMed: 16461916]
11. Wang Y, Rosengarth A, Luecke H. Structure of the human p53 core domain in the absence of DNA. *Acta Crystallogr D Biol Crystallogr*. 2007; 63:276–281. [PubMed: 17327663]
12. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*. 1999; 293:321–331. [PubMed: 10550212]
13. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. *J Mol Graph Model*. 2001; 19:26–59. [PubMed: 11381529]
14. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*. 2000; 41:415–427. [PubMed: 11025552]
15. Dunker AK, Brown CJ, Obradovic Z. Identification and functions of usefully disordered proteins. *Adv Protein Chem*. 2002; 62:25–49. [PubMed: 12418100]
16. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform*. 2000; 11:161–171.
17. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*. 2004; 337:635–645. [PubMed: 15019783]
18. Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK. Comparing and combining predictors of mostly disordered proteins. *Biochemistry*. 2005; 44:1989–2000. [PubMed: 15697224]
19. Feng ZP, Zhang X, Han P, Arora N, Anders RF, Norton RS. Abundance of intrinsically unstructured proteins in *P. falciparum* and other apicomplexan parasite proteomes. *Mol Biochem Parasitol*. 2006; 150:256–267. [PubMed: 17010454]
20. Tompa P, Dosztanyi Z, Simon I. Prevalent structural disorder in *E. coli* and *S. cerevisiae* proteomes. *J Proteome Res*. 2006; 5:1996–2000. [PubMed: 16889422]
21. Galea CA, High AA, Obenauer JC, Mishra A, Park CG, Punta M, Schlessinger A, Ma J, Rost B, Slaughter CA, Kriwacki RW. Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome. *Journal of proteome research*. 2009; 8:211–226. [PubMed: 19067583]
22. Burra PV, Kalmar L, Tompa P. Reduction in structural disorder and functional complexity in the thermal adaptation of prokaryotes. *PloS one*. 2010; 5:e12069. [PubMed: 20711457]
23. Xue B, Williams RW, Oldfield CJ, Dunker AK, Uversky VN. Archaic chaos: intrinsically disordered proteins in Archaea. *BMC Syst Biol*. 2010; 4(Suppl 1):S1. [PubMed: 20522251]
24. Uversky VN, Dunker AK. Understanding protein non-folding. *Biochim Biophys Acta*. 2010; 1804:1231–1264. [PubMed: 20117254]
25. Schad E, Tompa P, Hegyi H. The relationship between proteome size, structural disorder and organism complexity. *Genome Biol*. 2011; 12:R120. [PubMed: 22182830]
26. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: Disorder in thirty five hundred proteomes from viruses and the three domains of life. *Journal of Biomolecular Structure and Dynamics*. 2012 In press.
27. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry*. 2005; 44:12454–12470. [PubMed: 16156658]
28. Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. *Proteins*. 2003; 53(Suppl 6):566–572. [PubMed: 14579347]

29. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. *Biochemistry*. 2002; 41:6573–6582. [PubMed: 12022860]
30. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys*. 2008; 37:215–246. [PubMed: 18573080]
31. Breydo L, Wu JW, Uversky VN. Alpha-synuclein misfolding and Parkinson's disease. *Biochim Biophys Acta*. 2012; 1822:261–285. [PubMed: 22024360]
32. Uversky VN. Neuropathology, biochemistry, and biophysics of alpha-synuclein aggregation. *J Neurochem*. 2007; 103:17–37. [PubMed: 17623039]
33. George JM. The synucleins. *Genome Biol*. 2002; 3 REVIEWS3002.
34. Vonderviszt F, Kanto S, Aizawa S, Namba K. Terminal regions of flagellin are disordered in solution. *J Mol Biol*. 1989; 209:127–133. [PubMed: 2810365]
35. Chen JW, Romero P, Uversky VN, Dunker AK. Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. *J Proteome Res*. 2006; 5:888–898. [PubMed: 16602696]
36. Chen JW, Romero P, Uversky VN, Dunker AK. Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *J Proteome Res*. 2006; 5:879–887. [PubMed: 16602695]
37. Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN. Close encounters of the third kind: disordered domains and the interactions of proteins. *Bioessays*. 2009; 31:328–335. [PubMed: 19260013]
38. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK. Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol*. 2002; 55:104–110. [PubMed: 12165847]
39. Chen SC, Chuang TJ, Li WH. The relationships among microRNA regulation, intrinsically disordered regions, and other indicators of protein evolutionary rate. *Mol Biol Evol*. 2011; 28:2513–2520. [PubMed: 21398349]
40. Lin YS, Hsu WL, Hwang JK, Li WH. Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol Biol Evol*. 2007; 24:1005–1011. [PubMed: 17264066]
41. Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput*. 1998:473–484. [PubMed: 9697205]
42. Shaiu WL, Hu T, Hsieh TS. The hydrophilic, protease-sensitive terminal domains of eucaryotic DNA topoisomerases have essential intracellular functions. *Pac Symp Biocomput*. 1999:578–589. [PubMed: 10380229]
43. Sayers EW, Gerstner RB, Draper DE, Torchia DA. Structural preordering in the N-terminal region of ribosomal protein S4 revealed by heteronuclear NMR spectroscopy. *Biochemistry*. 2000; 39:13602–13613. [PubMed: 11063598]
44. Wissmann R, Baukowitz T, Kalbacher H, Kalbitzer HR, Ruppertsberg JP, Pongs O, Antz C, Fakler B. NMR structure and functional characteristics of the hydrophilic N terminus of the potassium channel beta-subunit Kvbeta1.1. *J Biol Chem*. 1999; 274:35521–35525. [PubMed: 10585425]
45. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. *Curr Opin Struct Biol*. 2011; 21:441–446. [PubMed: 21482101]
46. Nilsson J, Grahm M, Wright AP. Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. *Genome Biol*. 2011; 12:R65. [PubMed: 21771306]
47. Chemes LB, Glavina J, Alonso LG, Marino-Buslje C, de Prat-Gay G, Sanchez IE. Sequence evolution of the intrinsically disordered and globular domains of a model viral oncoprotein. *PLoS One*. 2012; 7:e47661. [PubMed: 23118886]
48. Daughdrill GW, Narayanaswami P, Gilmore SH, Belczyk A, Brown CJ. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *J Mol Evol*. 2007; 65:277–288. [PubMed: 17721672]
49. Moesa HA, Wakabayashi S, Nakai K, Patil A. Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification. *Mol Biosyst*. 2012; 8:3262–3273. [PubMed: 23076520]

50. Osada M, Ohba M, Kawahara C, Ishioka C, Kanamaru R, Katoh I, Ikawa Y, Nimura Y, Nakagawara A, Obinata M, Ikawa S. Cloning and functional analysis of human p51, which structurally and functionally resembles p53. *Nat Med.* 1998; 4:839–843. [PubMed: 9662378]
51. Yang A, Kaghad M, Wang Y, Gillett E, Fleming MD, Dotsch V, Andrews NC, Caput D, McKeon F. p63, a p53 homolog at 3q27–29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol Cell.* 1998; 2:305–316. [PubMed: 9774969]
52. Kaghad M, Bonnet H, Yang A, Creancier L, Biscan JC, Valent A, Minty A, Chalon P, Lelias JM, Dumont X, Ferrara P, McKeon F, Caput D. Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell.* 1997; 90:809–819. [PubMed: 9288759]
53. Yang A, Kaghad M, Caput D, McKeon F. On the shoulders of giants: p63, p73 and the rise of p53. *Trends Genet.* 2002; 18:90–95. [PubMed: 11818141]
54. Murray-Zmijewski F, Lane DP, Bourdon JC. p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. *Cell Death Differ.* 2006; 13:962–972. [PubMed: 16601753]
55. Sah VP, Attardi LD, Mulligan GJ, Williams BO, Bronson RT, Jacks T. A subset of p53-deficient embryos exhibit exencephaly. *Nat Genet.* 1995; 10:175–180. [PubMed: 7663512]
56. Rotter V, Schwartz D, Almon E, Goldfinger N, Kapon A, Meshorer A, Donehower LA, Levine AJ. Mice with reduced levels of p53 protein exhibit the testicular giant-cell degenerative syndrome. *Proc Natl Acad Sci U S A.* 1993; 90:9075–9079. [PubMed: 8415656]
57. Donehower LA, Harvey M, Slagle BL, McArthur MJ, Montgomery CA Jr, Butel JS, Bradley A. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature.* 1992; 356:215–221. [PubMed: 1552940]
58. Yang A, Walker N, Bronson R, Kaghad M, Oosterwegel M, Bonnin J, Vagner C, Bonnet H, Dikkes P, Sharpe A, McKeon F, Caput D. p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. *Nature.* 2000; 404:99–103. [PubMed: 10716451]
59. Senoo M, Pinto F, Crum CP, McKeon F. p63 is essential for the proliferative potential of stem cells in stratified epithelia. *Cell.* 2007; 129:523–536. [PubMed: 17482546]
60. Su X, Paris M, Gi YJ, Tsai KY, Cho MS, Lin YL, Biernaskie JA, Sinha S, Prives C, Pevny LH, Miller FD, Flores ER. TAp63 prevents premature aging by promoting adult stem cell maintenance. *Cell Stem Cell.* 2009; 5:64–75. [PubMed: 19570515]
61. Mills AA, Zheng B, Wang XJ, Vogel H, Roop DR, Bradley A. p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature.* 1999; 398:708–713. [PubMed: 10227293]
62. Yang A, Schweitzer R, Sun D, Kaghad M, Walker N, Bronson RT, Tabin C, Sharpe A, Caput D, Crum C, McKeon F. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature.* 1999; 398:714–718. [PubMed: 10227294]
63. Irwin MS, Kaelin WG. p53 family update: p73 and p63 develop their own identities. *Cell Growth Differ.* 2001; 12:337–349. [PubMed: 11457731]
64. Irwin MS, Kaelin WG Jr. Role of the newer p53 family proteins in malignancy. *Apoptosis.* 2001; 6:17–29. [PubMed: 11321038]
65. Ozaki T, Naka M, Takada N, Tada M, Sakiyama S, Nakagawara A. Deletion of the COOH-terminal region of p73alpha enhances both its transactivation function and DNA-binding activity but inhibits induction of apoptosis in mammalian cells. *Cancer Res.* 1999; 59:5902–5907. [PubMed: 10606232]
66. Harms KL, Chen X. The C terminus of p53 family proteins is a cell fate determinant. *Mol Cell Biol.* 2005; 25:2014–2030. [PubMed: 15713654]
67. Lu WJ, Amatruda JF, Abrams JM. p53 ancestry: gazing through an evolutionary lens. *Nat Rev Cancer.* 2009; 9:758–762. [PubMed: 19776745]
68. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano

- S, Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science*. 2002; 298:2157–2167. [PubMed: 12481130]
69. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*. 2007; 317:86–94. [PubMed: 17615350]
 70. Pankow S, Bamberger C. The p53 tumor suppressor-like protein nvp63 mediates selective germ cell death in the sea anemone *Nematostella vectensis*. *PLoS One*. 2007; 2:e782. [PubMed: 17848985]
 71. King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JG, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature*. 2008; 451:783–788. [PubMed: 18273011]
 72. Mendoza L, Orozco E, Rodriguez MA, Garcia-Rivera G, Sanchez T, Garcia E, Gariglio P. Ehp53, an *Entamoeba histolytica* protein, ancestor of the mammalian tumour suppressor p53. *Microbiology*. 2003; 149:885–893. [PubMed: 12686631]
 73. Belyi VA, Ak P, Markert E, Wang H, Hu W, Puzio-Kuter A, Levine AJ. The origins and evolution of the p53 family of genes. *Cold Spring Harb Perspect Biol*. 2010; 2:a001198. [PubMed: 20516129]
 74. Lane DP, Cheok CF, Brown C, Madhumalar A, Ghadessy FJ, Verma C. Mdm2 and p53 are highly conserved from placozoans to man. *Cell Cycle*. 2010; 9:540–547. [PubMed: 20081368]
 75. Levine AJ. The evolution of the p53 family of genes. *Cell Cycle*. 2012; 11:214–215. [PubMed: 22214668]
 76. Lane DP, Madhumalar A, Lee AP, Tay BH, Verma C, Brenner S, Venkatesh B. Conservation of all three p53 family members and Mdm2 and Mdm4 in the cartilaginous fish. *Cell Cycle*. 2011; 10:4272–4279. [PubMed: 22107961]
 77. Bordo D, Argos P. Evolution of protein cores. Constraints in point mutations as observed in globin tertiary structures. *J Mol Biol*. 1990; 211:975–988. [PubMed: 2313703]
 78. Dean AM, Neuhauser C, Grenier E, Golding GB. The pattern of amino acid replacements in alpha/beta-barrels. *Mol Biol Evol*. 2002; 19:1846–1864. [PubMed: 12411594]
 79. Liu J, Tan H, Rost B. Loopy proteins appear conserved in evolution. *J Mol Biol*. 2002; 322:53–64. [PubMed: 12215414]
 80. Radivojac P, Obradovic Z, Smith DK, Zhu G, Vucetic S, Brown CJ, Lawson JD, Dunker AK. Protein flexibility and intrinsic disorder. *Protein Sci*. 2004; 13:71–80. [PubMed: 14691223]
 81. Jernigan RL, Kloczkowski A. Packing regularities in biological structures relate to their dynamics. *Methods Mol Biol*. 2007; 350:251–276. [PubMed: 16957327]
 82. Schaefer C, Schlessinger A, Rost B. Protein secondary structure appears to be robust under in silico evolution while protein disorder appears not to be. *Bioinformatics*. 2010; 26:625–631. [PubMed: 20081223]
 83. Brown CJ, Johnson AK, Daughdrill GW. Comparing models of evolution for ordered and disordered proteins. *Mol Biol Evol*. 2010; 27:609–621. [PubMed: 19923193]
 84. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res*. 2012; 40:W597–W603. [PubMed: 22661580]

85. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–3402. [PubMed: 9254694]
86. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta.* 2010; 1804:996–1010. [PubMed: 20100603]
87. Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Dunker AK. Identifying Disordered regions in proteins from amino acid sequences. *IEEE Int. Conf. Neural Networks.* 1997; 1:90–95.
88. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins.* 2001; 42:38–48. [PubMed: 11093259]
89. Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins.* 2005; 61(Suppl 7):176–182. [PubMed: 16187360]
90. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *Bmc Bioinformatics.* 2006; 7
91. Vucetic S, Brown CJ, Dunker AK, Obradovic Z. Flavors of protein disorder. *Proteins.* 2003; 52:573–584. [PubMed: 12910457]
92. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinform Comput Biol.* 2005; 3:35–60. [PubMed: 15751111]
93. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics.* 2005; 21:3435–3438. [PubMed: 15955783]
94. Dosztanyi Z, Csizmok V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol.* 2005; 347:827–839. [PubMed: 15769473]
95. Campen A, Williams RM, Brown CJ, Meng J, Uversky VN, Dunker AK. TOP-IDPscale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett.* 2008; 15:956–963. [PubMed: 18991772]
96. Xue B, Oldfield CJ, Dunker AK, Uversky VN. CDF it all: consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions. *FEBS Lett.* 2009; 583:1469–1474. [PubMed: 19351533]
97. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry.* 2007; 46:13468–13477. [PubMed: 17973494]
98. Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci.* 2002; 27:527–533. [PubMed: 12368089]
99. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *Febs J.* 2005; 272:5129–5148. [PubMed: 16218947]
100. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. *Biophys J.* 2007; 92:1439–1456. [PubMed: 17158572]
101. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signalling. *J Mol Recognit.* 2005; 18:343–384. [PubMed: 16094605]
102. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol.* 2002; 12:54–60. [PubMed: 11839490]
103. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol.* 2005; 6:197–208. [PubMed: 15738986]
104. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. Analysis of molecular recognition features (MoRFs). *J Mol Biol.* 2006; 362:1043–1059. [PubMed: 16935303]

105. Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, Uversky VN, Dunker AK. Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res.* 2007; 6:2351–2366. [PubMed: 17488107]
106. Garner E, Romero P, Dunker AK, Brown C, Obradovic Z. Predicting Binding Regions within Disordered Proteins. *Genome Inform Ser Workshop Genome Inform.* 1999; 10:41–50.
107. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23:2947–2948. [PubMed: 17846036]
108. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform.* 2008; 9:299–306. [PubMed: 18417537]
109. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14:1188–1190. [PubMed: 15173120]
110. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 1990; 18:6097–6100. [PubMed: 2172928]
111. Popowicz GM, Czarna A, Holak TA. Structure of the human Mdmx protein bound to the p53 tumor suppressor transactivation domain. *Cell Cycle.* 2008; 7:2441–2443. [PubMed: 18677113]
112. Mittl PR, Chene P, Grutter MG. Crystallization and structure solution of p53 (residues 326–356) by molecular replacement using an NMR model as template. *Acta Crystallogr D Biol Crystallogr.* 1998; 54:86–89. [PubMed: 9761820]
113. Avalos JL, Celic I, Muhammad S, Cosgrove MS, Boeke JD, Wolberger C. Structure of a Sir2 enzyme bound to an acetylated p53 peptide. *Mol Cell.* 2002; 10:523–535. [PubMed: 12408821]
114. Arrowsmith CH. Structure and function in the p53 family. *Cell Death Differ.* 1999; 6:1169–1173. [PubMed: 10637432]
115. Connolly ML. The molecular surface package. *Journal of molecular graphics.* 1993; 11:139–141. [PubMed: 8347567]
116. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M. The Double Cubic Lattice Method - Efficient Approaches to Numerical-Integration of Surface-Area and Volume and to Dot Surface Contouring of Molecular Assemblies. *J Comput Chem.* 1995; 16:273–284.
117. Jones S, Thornton JM. Prediction of protein-protein interaction sites using patch analysis. *J Mol Biol.* 1997; 272:133–143. [PubMed: 9299343]
118. Jones S, Thornton JM. Analysis of protein-protein interaction sites using surface patches. *J Mol Biol.* 1997; 272:121–132. [PubMed: 9299342]
119. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics.* 2008; 9(Suppl 1):S1.
120. Lowe ED, Tews I, Cheng KY, Brown NR, Gul S, Noble ME, Gamblin SJ, Johnson LN. Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. *Biochemistry.* 2002; 41:15625–15634. [PubMed: 12501191]
121. Muftaba S, He Y, Zeng L, Yan S, Plotnikova O, Sachchidanand, Sanchez R, Zeleznik-Le NJ, Ronai Z, Zhou MM. Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol Cell.* 2004; 13:251–263. [PubMed: 14759370]
122. Wu H, Maciejewski MW, Marintchev A, Benashski SE, Mullen GP, King SM. Solution structure of a dynein motor domain associated light chain. *Nat Struct Biol.* 2000; 7:575–579. [PubMed: 10876244]
123. McNulty BC, Young GB, Pielak GJ. Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder. *J Mol Biol.* 2006; 355:893–897. [PubMed: 16343531]
124. Li C, Charlton LM, Lakkavaram A, Seagle C, Wang G, Young GB, Macdonald JM, Pielak GJ. Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: implications for in-cell NMR spectroscopy. *J Am Chem Soc.* 2008; 130:6310–6311. [PubMed: 18419123]
125. Li C, Wang GF, Wang Y, Creager-Allen R, Lutz EA, Scronce H, Slade KM, Ruf RA, Mehl RA, Pielak GJ. Protein (19)F NMR in Escherichia coli. *J Am Chem Soc.* 2010; 132:321–327. [PubMed: 20050707]

126. Schlesinger AP, Wang Y, Tadeo X, Millet O, Pielak GJ. Macromolecular crowding fails to fold a globular protein in cells. *J Am Chem Soc.* 2011; 133:8082–8085. [PubMed: 21534566]
127. Fauvet B, Fares MB, Samuel F, Dikiy I, Tandon A, Eliezer D, Lashuel HA. Characterization of Semisynthetic and Naturally Nalpha-Acetylated alpha-Synuclein in Vitro and in Intact Cells: IMPLICATIONS FOR AGGREGATION AND CELLULAR PROPERTIES OF alpha-SYNUCLEIN. *J Biol Chem.* 2012; 287:28243–28262. [PubMed: 22718772]
128. Binolfi A, Theillet FX, Selenko P. Bacterial in-cell NMR of human alpha-synuclein: a disordered monomer by nature? *Biochem Soc Trans.* 2012; 40:950–954. [PubMed: 22988846]
129. Thongwichian R, Selenko P. In-cell NMR in *Xenopus laevis* oocytes. *Methods Mol Biol.* 2012; 895:33–41. [PubMed: 22760310]
130. Bekei B, Rose HM, Herzig M, Dose A, Schwarzer D, Selenko P. In-cell NMR in mammalian cells: part 1. *Methods Mol Biol.* 2012; 895:43–54. [PubMed: 22760311]
131. Bekei B, Rose HM, Herzig M, Selenko P. In-cell NMR in mammalian cells: part 2. *Methods Mol Biol.* 2012; 895:55–66. [PubMed: 22760312]
132. Bekei B, Rose HM, Herzig M, Stephanowitz H, Krause E, Selenko P. In-cell NMR in mammalian cells: part 3. *Methods Mol Biol.* 2012; 895:67–83. [PubMed: 22760313]
133. Ito Y, Mikawa T, Smith BO. In-cell NMR of intrinsically disordered proteins in prokaryotic cells. *Methods Mol Biol.* 2012; 895:19–31. [PubMed: 22760309]
134. Janin J, Sternberg MJE. Protein flexibility, not disorder, is intrinsic to molecular recognition. *F1000 Biol Rep.* 2013; 5
135. Parmar VM, Schroder M. Sensing endoplasmic reticulum stress. *Adv Exp Med Biol.* 2012; 738:153–168. [PubMed: 22399379]
136. Dunker AK, Uversky VN. The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure. *F1000 Biol Rep.* 2013; 5 In press.
137. Shamu CE, Walter P. Oligomerization and phosphorylation of the Ire1p kinase during intracellular signaling from the endoplasmic reticulum to the nucleus. *EMBO J.* 1996; 15:3028–3039. [PubMed: 8670804]
138. Welihinda AA, Kaufman RJ. The unfolded protein response pathway in *Saccharomyces cerevisiae*. Oligomerization and trans-phosphorylation of Ire1p (Ern1p) are required for kinase activation. *J Biol Chem.* 1996; 271:18181–18187. [PubMed: 8663458]
139. Cox JS, Walter P. A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. *Cell.* 1996; 87:391–404. [PubMed: 8898193]
140. Kawahara T, Yanagi H, Yura T, Mori K. Endoplasmic reticulum stress-induced mRNA splicing permits synthesis of transcription factor Hac1p/Ern4p that activates the unfolded protein response. *Mol Biol Cell.* 1997; 8:1845–1862. [PubMed: 9348528]
141. Sidrauski C, Walter P. The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. *Cell.* 1997; 90:1031–1039. [PubMed: 9323131]
142. Gonzalez TN, Sidrauski C, Dorfler S, Walter P. Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. *EMBO J.* 1999; 18:3119–3132. [PubMed: 10357823]
143. Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, Walter P. Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell.* 2000; 101:249–258. [PubMed: 10847680]
144. Todd-Corlett A, Jones E, Seghers C, Gething MJ. Lobe IB of the ATPase domain of Kar2p/BiP interacts with Ire1p to negatively regulate the unfolded protein response in *Saccharomyces cerevisiae*. *J Mol Biol.* 2007; 367:770–787. [PubMed: 17276461]
145. Credle JJ, Finer-Moore JS, Papa FR, Stroud RM, Walter P. On the mechanism of sensing unfolded protein in the endoplasmic reticulum. *Proc Natl Acad Sci U S A.* 2005; 102:18773–18784. [PubMed: 16365312]
146. Gething MJ. Role and regulation of the ER chaperone BiP. *Semin Cell Dev Biol.* 1999; 10:465–472. [PubMed: 10597629]
147. Kleizen B, Braakman I. Protein folding and quality control in the endoplasmic reticulum. *Curr Opin Cell Biol.* 2004; 16:343–349. [PubMed: 15261665]

148. Ma Y, Hendershot LM. ER chaperone functions during normal and stress conditions. *J Chem Neuroanat.* 2004; 28:51–65. [PubMed: 15363491]
149. Hendershot LM. The ER function BiP is a master regulator of ER function. *Mt Sinai J Med.* 2004; 71:289–297. [PubMed: 15543429]
150. Blond-Elguindi S, Cwirla SE, Dower WJ, Lipshutz RJ, Sprang SR, Sambrook JF, Gething MJ. Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell.* 1993; 75:717–728. [PubMed: 7902213]
151. Bjorkman PJ, Parham P. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu Rev Biochem.* 1990; 59:253–288. [PubMed: 2115762]
152. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics.* 1996; 14:33–38. 27–38. [PubMed: 8744570]

Highlights

- > Evolutionary conservation of the p53 family members is analyzed.
- > The ordered DNA-binding domain of p53 is evolutionary conserved.
- > p53 disordered regions are characterized by high sequence diversity.
- > In evolution, intrinsic disorder and sequence divergence are positively correlated.
- > This provides strong support for the existence of disordered regions in p53 *in vivo*.

Figure 1.

Domain organization of the members of human family of p53 proteins: p53 (A), p63 (B), and p73 (C). In p53, there is Trans Activation Domain 1 (TAD1, residues 1–40, pink bar), TAD2 (residues 40–60, red bar), a proline-rich region (PR, residues 64–92, dark pink bar), the central DNA Bind Domain (DBD, residues 94–292), the C-terminally located a tetramerization or oligomerization domain (OD, residues 325–356, cyan bar) and a regulatory C-terminal domain (CTD, residues 356–393, green bar). In p63, there is a TAD domain (residues 1–107, pink bar), DBD (residues 170–362, blue bar), OD (residues 393–443, cyan bar), and SAM domain (residues 541–607, dark yellow bar). In p73, there is a TAD domain (residues 1–46, pink bar), DBD (residues 131–310, blue bar), OD (residues 345–386, cyan bar), and SAM domain (residues 485–551, dark yellow bar).

Figure 2.

Structural similarity between the ordered domains of the members of p53 family. **A.** NMR solution structure of the DBD of p53 (PDB id: 2FEJ). Representative members of the conformational ensemble are shown as structures of different color. **B.** 3-D structures of the DBDs of p53 (PDB id: 2PCX; blue structure), p63 (PDB id: 2RMN, red structure), and p73 (PDB id: 2XWC; silver structure). Structures of the individual domains are shown together with their structural alignment. **C.** 3-D structures of the oligomerization domains of p53 (PDB id: 1C26; bottom left), p63 (PDB id: 4A9Z; upper left) and p73 (PDB id: 2WQI; upper right). Structurally aligned ODs are shown in the bottom right corner, with blue, red and silver structures corresponding to p53, p63 and p73, respectively. **D.** 3-D structures and structural alignment of the SAM domains of p63 (PDB id: 2Y9U; red structures) and p73 (PDB id: 1DXS; silver structures). Structures were visualized using version 1.8.7 of the visual molecular dynamics tool, VMD (<http://www.ks.uiuc.edu/Research/vmd/>) [152].

Figure 3.

Statistical analysis of the distribution of gaps in the multiple sequence alignment. **A.** The percentage of sequences with gaps at each site in the alignment relative to human p53 (P04637). If a position is a gap for P04637, the percentage is shown as a green square and line, while a site with a residue for P04637, is a brown circle and line. **B.** A histogram of the frequency of sites in the alignment (y-axis) with a given percentage of sequences with gaps (x-axis) at that site

Figure 4.

Disorder prediction for human p53 (UniProt ID: P04637, pink), p63 (UniProt ID: Q9H3D4, black), and p73 (UniProt ID: O15350, green). The prediction was implemented by PONDR-FIT. Residues with score above 0.5 are predicted to be disordered, while those below 0.5 are predicted to be structured. These three sequences were aligned based upon their DBDs. p53 has a structured DBD (PDB ID: 1GZH [9]). In addition to the DBD (PDB ID: 2RMN), p63 and p73 have a second structured SAM domain at the C-terminus (PDB ID: 1RG6). The blue bars are predicted MoRF regions.

A

Q9H3D4-N	97	RMQSDSLSDPMWPQYTNL---	114
O15350-N	54	---LEGMTTSMVAQFNLLSST	71
P04637-N	18	---TFSDLWKLLENVLSPL	35
		. : . *	

B

Q9H3D4-B2	51	LSPEVFQHIWDFLEQ	64	Q9H3D4-B2	40	MSQSTQTNEFLSPEVFQHIWDFLEQ	64
P04637-N	14	LSQETFSDLWKLLPE	27	O15350-G1	1	MAQSTATSPDGG-TTFEHLWSSLEP	24
		** *.*:.*:* :				*:** * . . *:***. **	

D

Q9H3D4-M2	394	RSPDDELLYLVPVRGRETYEMLLKIKES	420	O15350-M2	348	RHGDEDITYYLQVRGRENF	365
P04637-C1	321	KPLDGIFYFTLQIRGRERFEMFRELEA	347	P04637-C1	321	KPLDGIFYFTLQIRGRERF	338
		:: *.* : * :**** :** :***:				: * : : **:*****.	

E

Q9H3D4-M3	438	QQQQQHQLLQKQTSIQS	455	Q9H3D4-M1	363	QVSDSTKNKGDKTRPFQR	380
P04637-C2	375	QSTRHKKLMFKTEGPDS	392	O15350-M1	315	LNESSAKNGAASKRAFKQ	332
		* . .*:*** * . :*				..*:*** .:***.*	

C

F

G

Figure 5.

Sequence alignments for functional segments of human p53, p63, and p73.

- A.** Alignment of N MoRFs of p63 (Q9H3D4), p73 (O15350), and p53 (P04637).
B. Alignment of B2 fragment of p63 (Q9H3D4) and N MoRF of p53 (P04637).
C. Alignment of B2 fragment of p63 (Q9H3D4) and G1 fragment of p73 (O15350).
D. Alignment of M2 MoRF of p63 (Q9H3D4) and C1 MoRF of p53 (P04637).
E. Alignment of M3 MoRF of p63 (Q9H3D4) and C2 MoRF of p53 (P04637).
F. Alignment of M2 MoRF of p73 (O15350) and C1 MoRF of p63 (Q9H3D4).
G. Alignment of M1 MoRF of p63 (Q9H3D4) and M1 MoRF of p73 (O15350).

Figure 6.

Phylogenetic distribution of disorder in the members of the p53-family of proteins.

A. Phylogenetic tree of p53 family sequences in the dataset. The tree was generated by Mega4.0 [108] using the maximum likelihood algorithm. Prior to analysis, the alignment was corrected manually to remove apparent misplaced gaps. The phylogenetic tree was split into 7 clusters indicated by letters from A to G, which are generally based on taxonomic relatedness.

B. Percentage of disordered residues (IDAA%) in the p53 family proteins from each species. Disorder prediction was implemented by PONDR-FIT. The sequential order of the species is

the same as in the phylogenetic tree. The seven letters indicate seven clusters in the phylogenetic tree.

Figure 7.

Evolutionary conservation of disorder distribution within the p53 sequence.

A. Disorder score and K2 entropy at each position of human p53 (P04637). The disorder score is from the POND-R-FIT prediction for P04637. K2 entropy is calculated from the average of all p53 family sequences in the dataset aligned to P04637. Insertions in P04637 were omitted. Avg. entropy is the value of Loess smoothing over the entire sequence.

B. Comparison of the disorder score calculated for human p53 and K2 entropy for the data shown in A.

Figure 8.

The relationship between the average disorder score for each site in the alignment vs. the K2 score at that site.

A. Comparison of disorder score and K2 entropy for all sites in the alignment.

B. Comparison of disorder score and K2 entropy only for sites with less than 40% of gaps.

Red circles are sites from regions out of DBD domains, while green crosses are sites in the DBD domain.

Figure 9.

Bitmaps for amino acid substitutions in p53. After alignment, the substitution frequency for each type of amino acid was counted inside and outside the DBD. Then the frequencies were used to generate an "artificial alignment" represent the corresponding substitution frequencies. Next, the generated alignment was used as input for WebLogo to make the bitmap plot. The positions on the X-axis are for the 21 types of amino acids in the order Trp, Cys, Phe, Ile, Tyr, Val, Leu, His, Met, Ala, Thr, Arg, Gly, Gln, Ser, Asn, Pro, Asp, Glu, Lys, and "X". "X" indicates an alignment gap in Q868M8, which was chosen because it appears to be the most ancient sequence.

A and **B** represent bitmaps for frequency of amino acid substitutions based on all positions in the alignment for sites inside and outside of the DBD regions, respectively.

C and **D** represent bitmaps of substitution frequencies for positions with 40% and less gaps. **C** shows sites in the DBD, while **D** is for sites outside the DBD.

Figure 10.

Logo plots for regions corresponding to the N-terminal (A) and C2 α -MoRFs (B) of p53 proteins. Logo plots are generated based on the alignment of the 52 N-terminal and 69 C-terminal α -MoRFs. Alignments were performed without introducing gaps within the MoRF sequence. In these plots, X indicates amino acids unavailable during the extension of the MoRF region.

Figure 11.

Conservation of the peculiarities of protein-protein interactions for the N-terminal, C1 and C2 MoRF regions of p53. For this analysis, we selected PDB complexes in which one of the binding partners is a fragment of human p53 corresponding to its N, C1 or C2 MoRFs.

A. The 3D-plot representing the K2-entropy vs. Δ ASA vs. ASA dependence. Here, ASA and Δ ASA correspond to the solvent accessible surface area and the binding-change in to the solvent accessible surface area, respectively. The mesh surface is colored according to the K2-entropy values.

B. The K2-entropy *vs.* Δ ASA dependence for the residues in the seven N-terminal MoRFs (green symbols and line), fourteen MoRFs within the C1 MoRFs (blue symbols and line) and nine C2 MoRFs (red symbols and line). Lines represent the results of the linear fits of the corresponding data-points.

Table 1

List of the members of the p53 family analyzed in this study.

SwissProt ID	Species	Length	Gene ID	Molecule
A5JSV4	Japanese flounder	365	-----	p53
A7SFL1	Starlet sea anemone	497	-----	Pred-p53
A8CSR8	Starlet sea anemone	203	-----	p53-like
A8CSR9	Starlet sea anemone	369	-----	P53-like
A8DPD6	Starlet sea anemone	492	-----	p63/73
A9UZX3	Choanoflagellate	523	(5891409)	Pred-p53
A9V4M3	Choanoflagellate	571	-----	p53-like
A9XR54	Mangrove rivulus	365	-----	p53
B0W6L3	Southern house mosquito	436	(6033958)	Pred-p53
B0W6K4	Southern house mosquito	364	-----	Other
B0W6K5	Southern house mosquito	344	(6033950)	Other
B0W6K6	Southern house mosquito	391	(6033951)	Other
B0W6K7	Southern house mosquito	431	(6033952)	p53
B1H2C8	Western clawed frog	588	-----	Pred-p53
B3LZ91	Fruit fly	368	(6501048)	Other
B3P7H3	Fruit fly	508	(6554756)	Pred-p53
B3RZS6	Trichoplax adhaerens	576	-----	Other
B3TLB0	Turquoise killifish	372	-----	p53
B4JSF4	Fruit fly	359	(6567875)	Pred-p53
B4KBB9	Fruit fly	351	(6575374)	Pred-p53
B4HFA1	Fruit fly	489	(6607524)	Other
B4N9F0	Fruit fly	377	(6647791)	Pred-p53
B4PN32	Fruit fly	507	(6537834)	Pred-p53
B4R1C4	Fruit fly	489	(6729186)	Other
B5TJK8	Common whitefish	393	-----	p53
C3XPU2	Lancelet fish-like	649	7222337	p63/73
C3YXH3	Lancelet fish-like	416	7230376	p53-like
C4Q441	Platyhelminthes	410	-----	p53-like
C4Q601	Platyhelminthes	696	-----	p53-like
O09185	Chinese hamster	393	-----	p53
O12946	European flounder	366	-----	p53
O15350	Human	636	Hs.697294	p63/73
O36006	Woodchuck	391	-----	p53
O57538	Green swordtail	342	-----	p53
O88898	Mouse	680	Mm.20894	p63/73
O93379	Channel catfish	376	-----	p53
P04637	Human	393	Hs.654481	p53

SwissProt ID	Species	Length	Gene ID	Molecule
P07193	African clawed frog	363	Xl.394	p53
P10360	Chicken	367	Gga.706	p53
P10361	Rat	391	Rn.54443	p53
P13481	Green monkey	393	-----	p53
P25035	Rainbow trout	396	Omy.11581	p53
P41685	Felis silvestris catus	386	(493847)	p53
P51664	Sheep	382	Oar.489	p53
P56423	Cynomolgus monkey	393	-----	p53
P56424	Rhesus macaque	393	Mmu.3286	p53
P61260	Japanese macaque	393	-----	p53
P67938	Zebu	386	-----	p53
P67939	Bovine	386	Bt.49712	p53
P79820	Japanese ricefish	352	Ola.4812	p53
Q00366	Golden hamster	396	-----	p53
Q0GMA7	Axolotl	387	-----	p53
Q0H3B6	Euprymna scolopes	566	-----	p63/73
Q0JRM9	Green ormer	614	-----	p63/73
Q16JM4	Yellow fever mosquito	239	(5577576)	Other
Q171M1	Yellow fever mosquito	483	Aae.3068(5579967)	Pred-p53
Q171M5	Yellow fever mosquito	459	Aae.8410(5579964)	Pred-p53
Q1AMZ8	Blue mussel	657	-----	p63/73
Q1XDZ3	Pacific oyster	434	-----	p53
Q20646	Caenorhabditis elegans	644	-----	p53-like
Q27937	Northern European squid	564	-----	p53-homo
Q29537	Dog	381	Cfa.3658	p53
Q4H301	Transparent sea squirt	489	Cin.28046	p53
Q4H2Z8	Transparent sea squirt	419	Cin.23005	p53
Q502Q9	Zebrafish	374	Dr.75100	p53
Q68VB0	Blind subterranean mole rat	391	-----	p53
Q6NTF1	Western clawed frog	362	Str.11712	p53
Q6UNX2	Zebrafish	640	Dr.24319	p63/73
Q6WG20	Atlantic surf clam	596	-----	p63/73
Q7QAB9	African malaria mosquito	410	-----	Other
Q7QBX6	African malaria mosquito	325	-----	Pred-p53
Q868M8	Amoeba	430	---	p53-like
Q8IMZ4	Fruit fly	495	Dm.3257(2768677)	p53
Q8JHZ6	Zebrafish	588	Dr.32588	p63/73
Q8SPZ3	Beluga whale	387	-----	p53
Q91XH8	Mouse	391	Mm.222	p53

SwissProt ID	Species	Length	Gene ID	Molecule
Q920Y0	Mongolian gerbil	390	-----	p53
Q92143	Southern platyfish	342	-----	p53
Q95330	Rabbit	391	Ocu.1779	p53
Q98SW0	African clawed frog	365	Xl.762	p53
Q9DEC7	Chicken	582	Gga.4019	p63/73
Q9H3D4	Human	680	Hs.137569	p63/73
Q9JJP2	Mouse	631	Mm.78015	p63/73
Q9JJP6	Rat	680	Rn.42907	p63/73
Q9N252	Pig	387	-----	p53
Q9NGC7	Mya arenaria	621	-----	p63/73
Q9NGC8	Mya arenaria	443	-----	p53-like
Q9TTA1	Common tree shrew	393	-----	p53
Q9TUB2	Pig	386	Ssc.15917(397276)	p53
Q9W664	Barbel	641	-----	p63/73
Q9W678	Barbel	369	-----	p53
Q9W679	Congo puffer	367	-----	p53
Q9WUR6	Guinea pig	391	-----	p53
Q9XSK8	Green monkey	637	-----	p63/73